Algorithms and Computation in Mathematics • Volume 20

Johannes Buchmann
Ulrich Vollmer

Binary Quadratic Forms

An Algorithmic Approach

With 17 Figures and 5 Tables

Springer

Authors

Johannes Buchmann
Ulrich Vollmer

Technical University
Department of Computer Science
Hochschulstraße 10
64289 Darmstadt
Germany
E-mail: buchmann@cdc.informatik.tu-darmstadt.de
uvollmer@cdc.informatik.tu-darmstadt.de

Library of Congress Control Number: 2006938722

Mathematics Subject Classification (2000): 11-01, 11Y40, 11E12, 11R29

ISSN 1431-1550

ISBN-10 3-540-46367-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-46367-2 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springer.com

Typesetting: by the authors and techbooks using a Springer LaTeX macro package
Cover design: *design & production* GmbH, Heidelberg

Printed on acid-free paper SPIN: 10521650 46/techbooks 5 4 3 2 1 0

Contents

List of Figures

List of Algorithms

Introduction

This book deals with algorithmic problems concerning binary quadratic forms $f(X,Y) = aX^2 + bXY + cY^2$ with integer coefficients a, b, c, the mathematical theories that permit the solution of these problems, and applications to cryptography. A considerable part of the theory is developed for forms with real coefficients and it is shown that forms with integer coefficients appear in a natural way.

Much of the progress of number theory has been stimulated by the study of concrete computational problems. Deep theories were developed from the classic time of Euler and Gauss onwards to this day that made the solutions of many of these problems possible. Algorithmic solutions and their properties became an object of study in their own right.

This book intertwines the exposition of one very classical strand of number theory with the presentation and analysis of algorithms both classical and modern which solve its motivating problems. This algorithmic approach will lead the reader, we hope, not only to an understanding of theory and solution methods, but also to an appreciation of the efficiency with which solutions can be reached.

The computer age has led to a marked advancement of algorithmic research. On the one hand, computers make it feasible to solve very hard problems such as the solution of Pell equations with large coefficients. On the other, the application of number theory in public-key cryptography increased the urgency for establishing the complexity of several computational problems: many a computer system stays only secure as long as these problems remain intractable.

Thus, number theory has become a research area not only in mathematics but also in computer science. This book tries to combine both worlds. It talks about mathematical theory, algorithms, and complexity.

The material presented is suitable as an introduction to many areas of (algorithmic) number theory for which the theory of binary quadratic forms is a starting point. We illustrate this for the areas *Diophantine equations*, *geometry of numbers*, and *algebraic number theory.*

Diophantine equations are of the form $f(x_1, \dots, x_n) = 0$ where f is a polynomial with integer coefficients and the goal is to find integers $x_1, \dots, x_n$ that satisfy this equation. In 1900 Hilbert proposed the problem of finding an algorithm that decides the solvability of any given Diophantine equation as the tenth of his celebrated problems. Seventy years later Juri V. Matijasevič [Mat70] proved that there cannot be such an algorithm. It is therefore necessary to develop algorithms that solve subclasses of Diophantine equations. The easiest Diophantine equations are linear. They can be solved using the Euclidean algorithm. Next, univariate quadratic Diophantine equations can be solved using algorithms for extracting square roots. Thus, the first really difficult Diophantine equation is bivariate quadratic. Solving those equations can be reduced to solving Diophantine equations $ax^2 + bxy + cy^2 = n$ with integers a, b, c, n, see [Mat61]. This type of Diophantine equation is discussed in this book.

One of the most difficult algorithmic problems in the *geometry of numbers* is the computation of shortest vectors in lattices. The shortest vector problem (SVP) is known to be NP-hard [Ajt98], i.e. an algorithm solving it can be used to solve any problem for all of whose solutions exist polynomial size proofs. The intractability of the shortest vector problem is the security basis for many new cryptographic systems. As explained in Chapter 5, the reduction algorithm for positive definite binary quadratic forms solves the shortest vector problem for two-dimensional lattices. This algorithm is the basis for the LLL algorithm [LLL82] and many other reduction algorithms that find short vectors in higher dimensional lattices.

Key computational problems in *algebraic number theory* are the computation of fundamental units and the class group of a number field. This book solves these problems for quadratic number fields. The most efficient fundamental unit and class group algorithms for higher degree number fields are generalizations of the algorithms described here.

This book can serve as a textbook for mathematics and computer science students. It is based on courses that the first author taught in Saarbrücken and Darmstadt at the mathematics and computer science departments. For the most part only basic mathematical knowledge is required. In some parts, more elaborate mathematics is used which is not treated here, for example results from analytic number theory. In such cases the results are explained and other treatments are referenced. The book emphasizes proofs of the properties of the given algorithms, in particular their complexity. Possible optimizations are sometimes omitted in order to maintain clarity.

This book is also a monograph on modern developments in the algorithmic theory of binary quadratic forms and quadratic number fields. Here are a few highlights. Chapter 9 describes the fastest deterministic algorithm for computing the structure of a finite Abelian group from a generating system. That chapter also explains a fast rigorous deterministic algorithm for computing a generating system of the class group of a quadratic number field. Chapter 11 presents subexponential algorithms for class group and unit computation in quadratic number fields that are fully proved under the assumption of a certain extended Riemann hypothesis (ERH). This chapter also introduces compact representations of quadratic irrationalities which are crucial in many complexity results. For example, the length of the standard representation of a fundamental unit in a real quadratic field is typically exponential in the length of the discriminant of that field. Thus, using the standard representation only exponential fundamental unit algorithms are possible. The compact representation of the fundamental unit, however, requires only polynomial size. This enables us to provide polynomial size proofs for the validity of a given class number and regulator of a real quadratic order (assuming ERH). In other words, the class number problem for quadratic fields is in NP.

Unavoidably, many interesting topics with close connection to the material presented here had to be omitted. The most notable of these omissions are perhaps genus theory and the theory of continued fractions the latter of which could serve as yet another language in which to cast the reduction theory of indefinite integral forms. We also refer the reader to the rich literatur on the algebraic techniques which so profitably can be brought to bear in the context. We mention two examples. Class field theory identifies class groups with Galois groups of field extensions and permits among other things the characterization of the set of primes represented by a given (positive definite) form, see e.g. [Cox89]. Arakelov theory provides a natural and generalizable group-theoretic basis for the study of infrastructure [Schar] which we have chosen to introduce with a focus on its algorithmic usefulness.

We have not tried to trace the historical development of the ideas presented in this book, neither classical nor recent. Representative of many other contributors we would like to mention (in alphabetical order and with no slight intended to anybody omitted) Henri Cohen, James Hafner, Jeff Lagarias, Hendrik Lenstra, Kevin McCurley, Michael Pohst, René Schoof, Martin Seysen, Daniel Shanks, Hugh Williams, and Hans Zassenhaus who have discovered and analyzed fundamental algorithms for binary quadratic forms.

Content

The book starts by introducing basic notions and facts concerning binary quadratic forms, for example the discriminant of a form and its meaning. Also, important algorithmic problems are discussed, for example the *representation problem* and the *minimum problem.* We say that a number r is represented by a form f if there exist integers x and y such that $f(x, y) = r$. The pair (x, y) is called a representation of r by f. Solving the minimum problem means finding the smallest positive or the largest negative real number r that can be represented by a given form f. The representation problem is the following: given a form f with integer coefficients and an integer n, decide whether n can be represented by f. If there is one such representation, find them all. This is the content of Chapter 1. That Chapter also explains naive methods for solving the algorithmic problems, it demonstrates their difficulty, and shows the relevance of the problems in other contexts.

Next, Chapter 2 discusses equivalence of forms. The notion of equivalence is crucial for solving the algorithmic problems addressed in this book. The chapter describes a strategy for solving representation problems $f(x, y) = n$. This strategy consists of four steps. The first step reduces the representation problem to finding all representations (x, y) with $\gcd(x, y) = 1$. The second step constructs all forms $nX^2 + BXY + CY^2$ with the same discriminant as f. The third step decides equivalence between those forms and f. When such an equivalence is discovered, a representation is found. In the fourth step the automorphism group of f is computed and applied to the representations found in the third step. This strategy gives rise to three algorithmic problems: how to construct the forms (n, B, C), how to decide equivalence of forms, and to find appropriate transformations, and how to find the automorphism group of a form. This chapter also partially answers the last question. The connection between the automorphisms of a form and the solution of the Pell equation $x^2 - \Delta y^2 = \pm 4$ is explained and the automorphism group of positive definite forms is determined.

Chapter 3 solves the construction problem that arises in the strategy from the previous chapter. The existence of forms (n, B, C) with given discriminant can be decided by evaluating Kronecker symbols. The algorithm for evaluating Kronecker symbols is based on the law of quadratic reciprocity. The forms are actually constructed using algorithms that extract square roots modulo a positive integer.

The theory of binary quadratic forms can be presented in several ways. This theory can also be explained using point or lattices in the plane or even continued fractions. Chapter 4 describes the correspondence between forms on the one hand and points and lattices in the plane on the other. This correspondence permits the use of geometric arguments in subsequent chapters, for example in the proof of the periodicity of the cycle of reduced forms in Chapter 6 and in the characterization of composable forms in Chapter 7. The lattices that correspond to integral forms will later turn out to be fractional ideals of orders in quadratic number fields. This way of looking at the theory of binary quadratic forms permits the translation into the theory of quadratic number fields in Chapter 8.

Chapter 5 explains reduction theory for positive definite forms. It shows that the reduction algorithm for positive definite forms efficiently solves the problem of deciding equivalence between such forms and finds the minimum of a positive definite form.

Reduction theory for indefinite forms is treated in Chapter 6. This theory also decides equivalence of forms and solves the minimum problem. However, the results in reduction theory for indefinite forms are quite different from those obtained in reduction theory for positive definite forms. Instead of a unique reduced form there may be many reduced forms in each equivalence class of forms. The reduction algorithm finds the so called cycle of reduced forms in an equivalence class which contains all reduced forms in that class. This result is proved using geometric arguments. Also, the automorphism group of an integral indefinite form is determined.

Chapters 3, 5, and 6 solve the main algorithmic problems. However, the efficiency of the solution is not optimal. To obtain more efficient algorithms, additional theory is required. This theory is developed in the next chapters. In Chapter 7 composition of forms is discussed in the language of lattices. The product of two lattices is defined and the class of lattices whose product is again a lattice are characterized. That characterization leads to fundamental notions from the theory of quadratic number fields: rings of multipliers of lattices, quadratic orders, and maximal quadratic orders. It is shown that the product of two lattices is a lattice if and only if their ring of multipliers is in the same maximal order. An efficient algorithm for calculating the product of two such lattices is explained and it is proved that the set of all lattices with the same ring of multipliers forms an Abelian group.

Chapter 8 studies quadratic number fields, the fields of fraction of quadratic orders. It is shown that such a number field F contains a unique maximal order, the ring of algebraic integers in F. The group of lattices whose ring of multipliers is a fixed order $\mathcal{O}$ in F turns out to be the group of invertible $\mathcal{O}$-ideals. Unique factorization of $\mathcal{O}$-ideals into prime ideals is proved and the structure of the unit group of $\mathcal{O}$ is deduced from the structure of the automorphism group of forms whose discriminant is the discriminant of $\mathcal{O}$. The multiplicativity of invertible $\mathcal{O}$-ideals induces a group structure on the equivalence classes of primitive integral forms. The corresponding groups are called class groups.

Chapter 9 introduces those groups in the language of ideals and explains algorithms for computing their structure. Those algorithms consist of two parts. First, a generating system for the class group is computed. Then generic algorithms are developed and analyzed that compute the structure of a finite Abelian group from a generating system. It is shown that assuming the ERH the structure of the class group of an imaginary quadratic order of discriminant Δ can be computed in time $|\Delta|^{1/4+o(1)}$.

To prove this result for real quadratic orders, a faster equivalence test is required that does not need to compute the cycle of all reduced ideals in an equivalence class. Such an algorithm is presented in Chapter 10. The algorithm also computes the fundamental unit of an order from which all units of that order can be obtained. The running time is again $\Delta^{1/4+o(1)}$ both for fundamental unit computation and equivalence testing of reduced forms. This can be proved without reliance on a Riemann hypothesis. The running times of the class group, fundamental unit, and equivalence testing algorithms are faster than those of the algorithms obtained in classical reduction theory. However, they are still exponential.

Chapter 11 presents subexponential algorithms for class group and unit computation and equivalence testing. They are index calculus algorithms which again depend for their correctness and run time bound on the ERH. A key ingredient for unit computation is the compact representation of elements in quadratic fields.

Finally, Chapter 12 shows how the intractability of computational problems in quadratic fields such as the discrete logarithm in the class group of imaginary quadratic orders and equivalence testing of ideals of real quadratic orders can be used as the security basis for many cryptographic algorithms.

The appendix collects results and algorithms for linear algebra problems over the integers needed in Chapters 9 and 11.

Acknowledgments

Many people have directly or indirectly supported the authors in writing this book.

Hans-Joachim Stender, Michael Pohst, and Hans Zassenhaus introduced the first author to algorithmic number theory and the geometry of numbers. The second author owes Reinhard Bölling his first acquaintance with the theory of quadratic number fields. In many conversations, Henri Cohen, Hendrik Lenstra, Hugh Williams, and many other colleagues contributed their ideas. Hugh Williams also contributed the presentation of a number of algorithms and several numerical examples in this book. The books of Eric Bach and Jeffrey Shallit, [BS96], Duncan Buell [Bue89] and Henri Cohen [Coh00] have been most influential for us. Several results that are presented here are from the PhD theses of Christine Abel, Stephan Düllmann, Safuat Hamdy, Detlef Hühnlein, Michael Jacobson, Markus Maurer, Andreas Meyer, Stefan Neis, Renate Scheidler, Arthur Schmidt, Edlyn Teske, Patrick Theobald, Christoph Thiel, and the second author.

The authors would like to thank them all.

Chapter references and further reading

[Ajt98] Miklos Ajtai, *The shortest vector problem in l_2 is NP-hard for randomized reductions (extended abstract)*, Proceedings of the 35th annual ACM Symposium on Theory of Computing (Jeffrey Vitter, ed.), ACM Press, 1998, pp. 10–19.

[BSh66] Zenon I. Borevič and Igor R. Šafarevič, *Number theory*, Academic Press, New York, 1966.

[BS96] Eric Bach and Jeffrey Shallit, *Algorithmic number theory*, MIT Press, Cambridge, Massachusetts and London, England, 1996.

[Buc04] Johannes Buchmann, *Introduction to cryptography*, second ed., Springer-Verlag, 2004, Undergradute Texts in Mathematics.

[Bue89] Duncan A. Buell, *Binary quadratic forms*, Springer-Verlag, New York, 1989.

[Coh00] Henri Cohen, *A course in computational algebraic number theory*, fourth corrected ed., Graduate Texts in Mathematics, no. 138, Springer-Verlag, 2000.

[Cox89] David A. Cox, *Primes of the form $x^2 + ny^2$*, Wiley, New York, 1989.

[Gau86] Carl Friedrich Gauss, *Disquisitiones arithmeticae*, English ed., Springer-Verlag, New York, 1986, Translated by A. Clark.

[Lan66] Edmund Landau, *Elementary number theory*, second edition ed., Chelsea Publishing Company, 1966.

[LLL82] Arjen K. Lenstra, Hendrik W. Lenstra, Jr., and László Lóvasz, *Factoring polynomials with rational coefficients*, Mathematische Annalen **261** (1982), 515–534.

[Mat61] George B. Mathews, *Theory of numbers*, 2 ed., Chelsea, New York, 1961.

[Mat70] Juri Matijasevič, *Enumerable sets are diophantine*, Soviet Mathematics. Doklady **11** (1970), no. 2, 354–358.

[Mol96] Richard A. Mollin, *Quadratics*, CRC Press, 1996.

[PZ89] Michael Pohst and Hans Zassenhaus, *Algorithmic algebraic number theory*, CUP, 1989.

[Schar] René Schoof, *Computing Arakelov class groups*, Surveys in algorithmic number theory (to appear), `http://www.mat.uniroma2.it/~schoof/infranew2.pdf`.

[Zag81] Don B. Zagier, *Zetafunktionen und quadratische Zahlkörper*, Springer-Verlag, 1981.

1

Binary Quadratic Forms

In this book we study *binary quadratic forms*

$$f(X,Y) = aX^2 + bXY + cY^2$$

with integer coefficients a, b, c, or, more generally, with real coefficients a, b, c, where not all three coefficients a, b, and c are zero. We write

$$f = (a, b, c) \ ,$$

and call f a *form*. If the coefficients a, b, c are integers then we call f an *integral form*.

1.1 Computational problems

We describe difficult computational problems concerning the values $f(x, y)$ where x, y are integers. Those problems will be solved in this book. Mostly, the solutions will be given in algorithmic form. For background reading on algorithm complexity and basic number-theoretic algorithms we refer to [BS96].

1.1.1 Finding representations

The first computational problem is to find solutions $(x, y) \in \mathbb{Z}^2$ of the *Diophantine equation*

$$ax^2 + bxy + cy^2 = n \tag{1.1}$$

where n is an integer. Such a solution (x, y) is called a *representation* of n by f. If such a representation exists, then we say that f *represents* n. If $\gcd(x, y) = 1$, then (x, y) is called a *proper representation* of n by f. Otherwise, (x, y) is called an *improper representation* of n by f.

Example 1.1.1. We consider the Diophantine equation

$$x^2 - y^2 = 113 , \tag{1.2}$$

that is, we try to find the representations of 113 as the difference of two integer squares. If (x, y) is such a representation, then so are $(\pm x, \pm y)$. Both x and y are non-zero since 113 is not a square in $\mathbb{Z}$. We may, therefore, assume that $x \geq y > 0$. Now (1.2) can be written as

$$(x - y)(x + y) = 113 .$$

Since 113 is a prime number and because there is unique factorization in $\mathbb{Z}$, it follows that

$$x + y = 113 \quad \text{and} \quad x - y = 1 . \tag{1.3}$$

From (1.3) we obtain $2x = 114$ or $x = 57$ and $y = 113 - 57 = 56$. Hence, the set of all solutions of (1.2) is

$$\Big\{(\pm 57, \pm 56)\Big\} .$$

We were able to find those solutions because the polynomial $X^2 - Y^2$ is the product of two linear polynomials with integer coefficients and because there is unique factorization in $\mathbb{Z}$.

Example 1.1.2. We consider the Diophantine equation

$$x^2 + y^2 = 100049 , \tag{1.4}$$

that is, we try to represent 100049 as the sum of two squares. If (x, y) is a solution of that equation, then $x^2 \leq 100049$ and $y^2 \leq 100049$. Hence, we can determine all solutions of (1.4) by trying a finite number of possibilities. We find that the set of all solutions is

$$\Big\{(\pm 215, \pm 232), (\pm 232, \pm 215)\Big\} .$$

The method for solving (1.4) can also be used to solve any Diophantine equation of the form

$$x^2 + y^2 = n \tag{1.5}$$

with a positive integer n. However, this naive algorithm is very slow since it tests roughly n pairs (x, y). The algorithm receives n as the only input. The length of the binary expansion of n is $\lfloor \log_2 n \rfloor + 1$. Therefore, the running time of the algorithm is exponential.

Example 1.1.3. Consider the Diophantine equation

$$x^2 - 2y^2 = 1 . \tag{1.6}$$

Trying a few small values for x and y, we find that $(\pm 3, \pm 2)$ are solutions of (1.6). But those are not the only solutions. For, if (x, y) is any solution of (1.6), then $1 = (x^2 - 2y^2)^2 = x^4 - 4x^2y^2 + 4y^4 = (x^2 + 2y^2)^2 - 2(2xy)^2$. Hence, $\big(\pm(x^2 + 2y^2), \pm 2xy\big)$ is also a solution of (1.6). This shows that (1.6) has infinitely many solutions.

In Example 1.1.3, no general method for solving a Diophantine equation of the form $x^2 - dy^2 = 1$ for a positive integer d has been presented. Also, it is not clear whether all solutions of the Diophantine equation (1.6) have been found. In Chapter 6 we will present an algorithm that finds all solutions of such a Diophantine equation. To show that this is a difficult problem, we present another example.

Example 1.1.4. Consider the Diophantine equation

$$x^2 - 11101y^2 = -4\ . \tag{1.7}$$

A solution of this equation is $(x, y) = (18253909681995, 173250639377)$. This is the solution with the smallest positive y.

Example 1.1.5. Consider the Diophantine equation

$$x^2 - 5y^2 = 2\ . \tag{1.8}$$

Trying many possible values for x and y we do not find a solution of (1.8). In fact, there is no solution of (1.8). To see this, we consider (1.8) modulo 5. We obtain $x^2 \equiv 2 \pmod 5$. The following table lists the squares modulo 5.

x	0	1	2	3	4
$x^2 \bmod 5$	0	1	-1	-1	1

From this table we see that 2 is not a square mod 5. Therefore, (1.8) cannot have a solution.

Concluding the discussion of this section, we formulate the *representation problem* for integral binary quadratic forms.

Problem 1.1.6 (Representation problem). Given an integer n and an integral form f.

1. Decide whether or not f represents n.
2. Determine the number of representations of n by f.
3. Determine all representations of n by f.

1.1.2 Finding the minimum

Let $f = (a, b, c)$ be a form. We call

$$\lambda_1(f) = \inf\Big\{|f(x, y)|^{1/2} : (x, y) \in \mathbb{Z}^2,\ (x, y) \neq (0, 0)\Big\} \tag{1.9}$$

the *minimum* of f.

Example 1.1.7. Consider the form

$$f(X,Y) = 269X^2 + 164XY + 25Y^2 \; . \tag{1.10}$$

It can be written as

$$f(X,Y) = (10X + 3Y)^2 + (13X + 4Y)^2 \; . \tag{1.11}$$

This is the sum of two squares. If we are able to find integers x, y such that $10x + 3y = 0$ and $13x + 4y = 1$, then $f(x,y) = 1$. For $x = -3$ and $y = 10$ this is true. Hence, the minimum of f is 1.

We formulate the *minimum problem.*

Problem 1.1.8 (Minimum problem). Given an integral form f. Find its minimum.

1.2 Discriminant

1.2.1 Definition

In Example 1.1.7 we were able to find the minimum of f because f can be written as the sum of two squares. In general, we can write

$$4af(X,Y) = (2aX + bY)^2 - \Delta Y^2 \; , \tag{1.12}$$

where

$$\Delta = b^2 - 4ac \; . \tag{1.13}$$

We can also write

$$4cf(X,Y) = (2cY + bX)^2 - \Delta X^2 \; . \tag{1.14}$$

We give a name to the quantity Δ.

Definition 1.2.1. *The* discriminant *of f is $\Delta(f) = b^2 - 4ac$.*

Example 1.2.2. We have $\Delta(X^2 - Y^2) = \Delta(1,0,-1) = 4$, $\Delta(X^2 + Y^2) = \Delta(1,0,1) = -4$, $\Delta(X^2 - 2Y^2) = \Delta(1,0,-2) = 8$, $\Delta(X^2 - 5Y^2) = \Delta(1,0,-5) = 20$, and $\Delta(269X^2 + 164XY + 25Y^2) = -4$.

Proposition 1.2.3.

1. *If f is an integral form, then $\Delta(f)$ is an integer with $\Delta(f) \equiv 0 \pmod 4$ or $\Delta(f) \equiv 1 \pmod 4$.*
2. *Any integer Δ with $\Delta \equiv 0 \pmod 4$ or $\Delta \equiv 1 \pmod 4$ is the discriminant of an integral binary quadratic form.*

Proof. The first assertion follows from (1.13) and from the fact that a square of an integer is either 0 or 1 modulo 4. To prove the second assertion, we note that for any integer Δ with $\Delta \equiv 0, 1 \pmod 4$ the form

$$\big(1, \Delta \bmod 4, (\Delta \bmod 4 - \Delta)/4\big) \tag{1.15}$$

is an integral form of discriminant Δ. □

We also obtain the following statement.

Proposition 1.2.4. *If $f = (a, b, c)$ is integral, then $b \equiv \Delta(f) \pmod 2$.*

Proof. We have $\Delta(f) = b^2 - 4ac$. So b is even if and only if the right hand side of this equation is even. □

1.2.2 The matrix of a form

We introduce the matrix of f. It will be useful to prove properties of the discriminant of f.

Definition 1.2.5. *The* matrix *of $f = (a, b, c)$ is*

$$M(f) = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} .$$

Using its matrix the form f can be written as

$$f(X, Y) = (X, Y)M(f)(X, Y) . \tag{1.16}$$

Also, we have

$$\Delta(f) = -4 \det M(f) . \tag{1.17}$$

Example 1.2.6. The matrix of the form $2X^2 + 6XY + 5Y^2$ is $\begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$.

1.2.3 Solving the representation problem for $\Delta(f) < 0$

Let $\Delta = \Delta(f) < 0$ and let n be an integer. We show how the Diophantine equation

$$ax^2 + bxy + cy^2 = n \tag{1.18}$$

can be solved. Since $b^2 - 4ac = \Delta < 0$ it follows that $ac > 0$. By (1.12) and (1.14) we have to solve one of the Diophantine equations

$$4an = (2ax + by)^2 + |\Delta|y^2 \tag{1.19}$$

or

$$4cn = (2cy + bx)^2 + |\Delta|x^2 \ . \tag{1.20}$$

The right hand side of those equations is ≥ 0. Hence, they can only have a solution if $an \geq 0$ and $cn \geq 0$.

For $n = 0$ we obtain $x = 0$ from (1.20) and $y = 0$ from (1.19). So in this case, the only solution is $(x, y) = (0, 0)$.

Assume that $n \neq 0$. By (1.19) and (1.20) all solutions $(x, y) \in \mathbb{Z}^2$ of the Diophantine equation (1.18) satisfy

$$x^2 \leq 4cn/|\Delta| \qquad \text{and} \qquad y^2 \leq 4an/|\Delta| \ . \tag{1.21}$$

This proves the following result.

Proposition 1.2.7. *If $\Delta(f) < 0$ and if n is a real number, then the Diophantine equation $ax^2 + bxy + cy^2 = n$ has only finitely many solutions.*

To find all solutions of (1.18) we test for every solution (x, y) of (1.21) whether it satisfies (1.18).

Example 1.2.8. We solve the Diophantine equation

$$3x^2 + 2xy + 2y^2 = 28 \ .$$

The discriminant of this form is -20. From (1.21) we obtain $x^2 \leq 4 \cdot 2 \cdot 28/20 = 11.2$. Hence $|x| \leq 3$. Form (1.21) we also obtain $y^2 \leq 4 \cdot 3 \cdot 28/20 = 16.8$. Hence $|y| \leq 4$. Testing all possible pairs (x, y) we find that the representations of 28 by $(3, 2, 2)$ are $(2, 2)$, $(-2, -2)$, $(-2, 4)$, $(2, -4)$.

We analyze this algorithm. The number of pairs (x, y) that satisfy (1.21) is roughly $16\sqrt{ac}|n|/\varDelta$, hence, for fixed f, proportional to n. Since n is an input to the algorithm, this algorithm is exponential. Nevertheless, we have shown how to solve the representation problem for negative discriminants. For positive discriminants this method cannot be applied.

1.2.4 Positive definite, negative definite, and indefinite forms

We introduce a classification of binary quadratic forms and we will show that this classification depends on their discriminant.

Definition 1.2.9.

1. *The form f is called* positive definite *if for any pair $(x, y) \neq (0, 0)$ of real numbers the value $f(x, y)$ is positive.*
2. *The form f is called* positive semidefinite *if for no pair (x, y) of real numbers the value $f(x, y)$ is negative.*
3. *The form f is called* negative definite *if for any pair $(x, y) \neq (0, 0)$ of real numbers the value $f(x, y)$ is negative.*

4. *The form f is called* negative semidefinite *if for no pair (x, y) of real numbers the value $f(x, y)$ is positive.*
5. *The form f is called* indefinite *if there is a pair (x, y) of real numbers with $f(x, y) < 0$ and a pair (x', y') of real numbers with $f(x', y') > 0$.*

If a form is positive definite or negative definite we call it *definite*. If a form is positive semidefinite or negative semidefinite we call it *semidefinite*. The next theorem explains how to distinguish between those classes of forms.

Proposition 1.2.10.
1. *The form f is positive definite if and only if $\Delta(f) < 0$ and $a > 0$.*
2. *The form f is negative definite if and only if $\Delta(f) < 0$ and $a < 0$.*
3. *The form f is positive semidefinite if and only if $\Delta(f) \leq 0$ and ($a > 0$ or $c > 0$).*
4. *The form f is negative semidefinite if and only if $\Delta(f) \leq 0$ and ($a < 0$ or $c < 0$).*
5. *The form f is indefinite if and only if $\Delta(f) > 0$.*

Proof. Let $\Delta = \Delta(f)$. We prove the "if" part of all assertions.

3. and 4. Let $\Delta \leq 0$. If $a > 0$ then (1.12) implies $f(x, y) = \left((2ax + by)^2 + |\Delta| y^2\right)/(4a) \geq 0$ for all $x, y \in \mathbb{R}$. Hence, f is positive semidefinite. Likewise, if $a < 0$, then f is negative semidefinite. Also, (1.14) implies that f is positive semidefinite for $c > 0$ and negative semidefinite for $c < 0$.

1. and 2. We assume that $\Delta < 0$. We show that $f(x, y) \neq 0$ for $(x, y) \neq (0, 0)$. Then 1. and 2. follow from 3. and 4. We have $4ac = b^2 + |\Delta| > 0$. Let $(x, y) \in \mathbb{R}^2$. If $y \neq 0$ then $f(x, y) \neq 0$ by (1.12). If $x \neq 0$ then $f(x, y) \neq 0$ by (1.14).

5. Let $\Delta > 0$. If $a \neq 0$, then $f(1, 0) \cdot f(b, -2a) = -\Delta a^2 < 0$ by (1.12). Hence f is indefinite. If $c \neq 0$ then $f(0, 1)f(-2c, b) = -\Delta c^2 < 0$ by (1.14). Hence, f is indefinite. If $a = c = 0$, then $f = bxy$ and $b \neq 0$, since f is non-zero. Hence, f is indefinite.

The "only if" part of 3., 4., and 5. follows from the fact that a (non-zero) form is either positive semidefinite or negative semidefinite or indefinite. The "only if" parts of 1. and 2. follow from Exercise 1.5.8. □

Example 1.2.11. The form $f(X, Y) = -2X^2 + 3XY - 2Y^2$ is negative definite since $\Delta(f) = -7$ and $a = -2 < 0$.
The form $f(X, Y) = 2X^2 - 3XY + 2Y^2$ is positive definite since $\Delta(f) = -7$ and $a = 2 > 0$.
The form $f(X, Y) = X^2 + 3XY + Y^2$ is indefinite since $\Delta(f) = 5 > 0$.

1.3 Reducible forms with integer coefficients

Let $f = (a, b, c)$ be an integral form and let $\Delta = \Delta(f)$ be the discriminant of f. In Example 1.1.1 we have found the representations of 113 by factoring f

into two linear factors with integer coefficients. In this section we answer the question when this is possible.

A *binary linear form* is a polynomial $l(X,Y) = dX + cY$. If the coefficients c, d are rational numbers then the form is called *rational.* If the coefficients are integers then the form is called *integral.* If the form is integral and $\gcd(d,c) = 1$, then the form is called *primitive.*

In this section we will prove the following theorem:

Theorem 1.3.1. *Let $f = (a,b,c)$ be an integral binary quadratic form. Then the following statements are equivalent:*

1. *The discriminant of f is a square of an integer.*
2. *The form f is the product of two rational binary linear forms.*
3. *The form f is the product of two integral binary linear forms.*
4. *There is $(x,y) \in \mathbb{Z}^2$ with $(x,y) \neq (0,0)$ and $f(x,y) = 0$.*

Proof. If $a = 0$, then $\Delta = b^2$ and $f(X,Y) = Y(bX + cY)$. Hence, f has all the properties described in Theorem 1.3.1. In the remainder of this proof we assume that $a \neq 0$.

If $\Delta = d^2$ with $d \in \mathbb{Z}$, then (1.12) yields the factorization

$$f(X,Y) = \big(2aX + (b+d)Y\big)\big(2aX + (b-d)Y\big)/(4a) \ . \tag{1.22}$$

Hence, f is the product of two rational linear forms.

Now we assume that f is the product of two rational linear forms. Then we can write

$$f(X,Y) = \frac{n}{d} p_1(X,Y) p_2(X,Y)$$

with $n, d \in \mathbb{Z}$, $d > 0$, $\gcd(n,d) = 1$, and two primitive integral binary linear forms p_1 and p_2. This implies

$$df(X,Y) = np_1(X,Y)p_2(X,Y) \ . \tag{1.23}$$

From Lemma A.3.2 we obtain

$$d \operatorname{cont}(f) = n \ .$$

Since $\gcd(n,d) = 1$, this implies $d = 1$. By (1.23) f is the product of two integral binary linear forms.

Assume that $f(X,Y) = (dX + eY)(gX + hY)$ with integers d, e, g, h. Then not all of those integers are zero and $0 = f(e,-d) = f(h,-g)$.

Finally, assume that $(x,y) \in \mathbb{Z}^2$, $(x,y) \neq (0,0)$ and $f(x,y) = 0$. Since $a \neq 0$ we have $y \neq 0$ since $f(x,0) = ax^2$. Hence, (1.12) implies $\Delta = ((2ax + by)/y)^2$. So Δ is an integer which is the square of a rational number. This means, that Δ is the square of an integer. □

Definition 1.3.2. *If f has the properties of Theorem 1.3.1 then we call f* reducible.

We explicitly determine the factorization of primitive reducible forms $f = (a, b, c)$ with $a \neq 0$. (For $a = 0$ the factorization of f is trivial.) From (1.22) we obtain

$$4af(X,Y) = \big(2aX + (b+d)Y\big)\big(2aX + (b-d)Y\big)\ . \tag{1.24}$$

It follows from Lemma A.3.2 that $4a = \gcd(2a, b+d)\gcd(2a, b-d)$. From Proposition 1.2.4 we obtain $\gcd(2a, b+d) = 2\gcd(a, (b+d)/2)$ and $\gcd(2a, b-d) = 2\gcd(a, (b-d)/2)$. Therefore, a factorization of f is

$$f(X,Y) = \frac{aX + (b+d)/2Y}{\gcd\big(a, (b+d)/2\big)}\,\frac{aX + (b-d)/2Y}{\gcd\big(a, (b-d)/2\big)}\ . \tag{1.25}$$

If f is not primitive, then $f/\operatorname{cont}(f)$ is primitive and can be factored as in (1.25). If we multiply this factorization with $\operatorname{cont}(f)$, then we obtain a factorization of f as a product of two integral binary linear forms.

Example 1.3.3. Consider the form $f(X,Y) = 4X^2 + 4XY - 15Y^2$. It is primitive. Its discriminant is $\Delta = 16^2$. Hence $\Delta = d^2$ with $d = 16$. The linear factors of f are

$$\frac{4X + 10Y}{2} = 2X + 5Y\ , \quad \frac{4X - 6Y}{2} = 2X - 3Y\ .$$

Hence, we have found the factorization

$$4X^2 + 4XY - 15Y^2 = (2X + 5Y)(2X - 3Y)\ .$$

1.4 Applications

The computational problems concerning forms that we have described in Section 1.1 are of great importance in many areas of computational mathematics. In this section, we present a few examples.

1.4.1 Lattice vectors of short and given length

We present geometric versions of the representation and minimum problem. For a brief introduction to lattices, see A.4. Let

$$\mathbf{b} = (b_1, b_2), \mathbf{c} = (c_1, c_2) \in \mathbb{R}^2$$

be two linearly independent vectors. Consider the lattice

$$L = \{x\mathbf{b} + y\mathbf{c} : x, y \in \mathbb{Z}\}\ .$$

The square of the Euclidean length of a lattice vector

$$\mathbf{v} = x\mathbf{b} + y\mathbf{c},\ x, y \in \mathbb{Z}$$

is

$$||\mathbf{v}||^2 = (xb_1+yc_1)^2+(xb_2+yc_2)^2 = x^2(b_1^2+b_2^2)+2xy(b_1c_1+b_2c_2)+y^2(c_1^2+c_2^2)\ .$$

If we set

$$a = b_1^2 + b_2^2, b = 2(b_1c_1 + b_2c_2), c = c_1^2 + c_2^2$$

and

$$f = (a, b, c)\ ,$$

then

$$||\mathbf{v}||^2 = f(x, y)\ .$$

Finding a vector of length l in L means finding a representation of l^2 by f. Also, finding the length of a shortest non-zero vector in L means finding the minimum of f. We denote that minimum by $\lambda_1(L)$.

Note that the discriminant of f is

$$\Delta(f) = -4(d(L))^2\ . \qquad (1.26)$$

Here $d(L)$ denotes the determinant of L, that is, the absolute value of the determinant of any basis of L (see Definition A.4.1).

Example 1.4.1. Let $\mathbf{b} = (3, 4)$ and $\mathbf{c} = (5, 6)$. Then $||x\mathbf{b} + y\mathbf{c}||^2 = 25x^2 + 78xy + 61y^2$. Finding the square of the length of a shortest non-zero vector in $L = \{x\mathbf{b} + y\mathbf{c} : x, y \in \mathbb{Z}\}$ means finding the minimum of the form $(25, 78, 61)$.

1.4.2 Lattice packings

Let $\mathbf{b}$, $\mathbf{c}$, L, and f be as in the last section. We cover the plane $\mathbb{R}^2$ with nonintersecting discs of equal radius r. For each lattice point, there is one circle centered in this lattice points. The circles have maximum size. We are looking for a lattice such that the area covered by all circles is as large as possible. The fraction of the plane covered by the spheres is $(\pi/4)\gamma(L)$ with

$$\gamma(L) = \frac{\lambda_1(f)}{d(L)} = \frac{2\lambda_1(f)}{\sqrt{|\Delta(f)|}}. \qquad (1.27)$$

Finding a densest lattice packing means finding a lattice such that $\gamma(L)$ is maximum.

Example 1.4.2. Consider the lattice $L = \mathbb{Z}^2$. We have $M(L) = M(1, 0, 1) = 1$ and $d(L) = 1 = -\Delta(1, 0, 1)/4$. So $\gamma(L) = 1$.

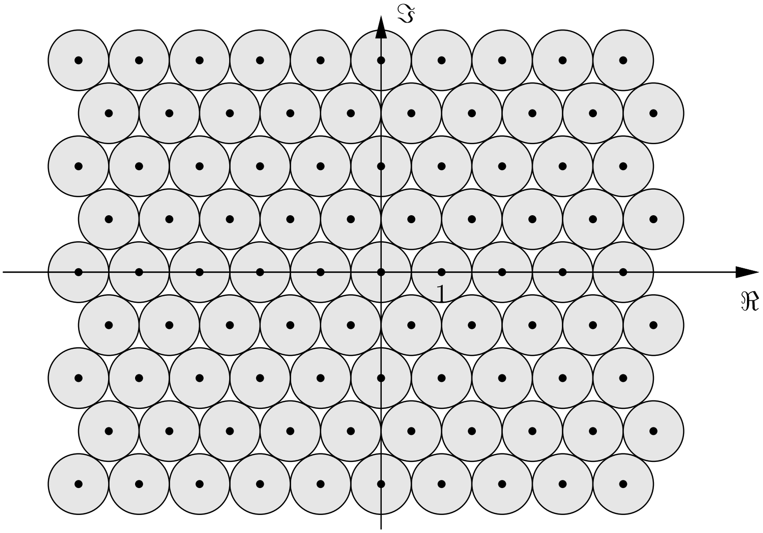

Fig. 1.1. Two-dimensional sphere packing

1.4.3 Factoring with ambiguous forms

We define ambiguous forms which can be used to factor discriminants of integral forms.

Definition 1.4.3. *An integral form f is called* ambiguous *if $f = (a, ka, c)$ with integers a, k, c.*

If $f = (a, ka, c)$ is an ambiguous form, then $\Delta(f) = a(k^2a - 4c)$ which is a factorization of Δ. We obtain a factorization of $\Delta(f)$ which is non-trivial if $a \neq \pm 1$ and $k^2a - 4c \neq \pm 1$.

Example 1.4.4. The form $(a, b, c) = (1, 1, -1)$ is an ambiguous form of discriminant 5 and we have $5 = a(a - 4c) = 1 * 5$. This is a trivial factorization of 5.

The form $(a, b, c) = (1009, 0, 1511)$ is an ambiguous form of discriminant $\Delta = -6098396$ and we have $\Delta = -4ac = -4 \cdot 1009 \cdot 1511$ which is a non-trivial factorization of -6098396.

Example 1.4.4 shows that finding an ambiguous form can mean finding non-trivial factorizations of integers. In fact, there are efficient integer factoring algorithms that use this fact.

1.4.4 Diophantine approximation

For any real number r the form $f(X, Y) = X^2 - r^2Y^2$ can be written as

$$f(X,Y) = (X - rY)(X + rY)\ . \tag{1.28}$$

This implies that we can write

$$\left|\frac{X}{Y} - r\right| = \left|\frac{f(X,Y)}{Y(X+rY)}\right|\ . \tag{1.29}$$

If r is positive and x, y are positive integers such that $f(x,y)$ is small, then x/y is a good rational approximation to r. This means that finding small values of f helps finding good rational approximations to r.

Example 1.4.5. Consider the form $f(X,Y) = X^2 - 2Y^2$. From Exercise 1.5.5 we know that this form represents ±1 infinitely often. Such representations are $(3,2)$, $(17,12)$, and $(577,408)$. Those representations yield the approximations $3/2$, $17/12$, and $577/408$ to $\sqrt{2}$. Note that

$$|\sqrt{2} - \frac{3}{2}| < 10^{-1}\ , |\sqrt{2} - \frac{17}{12}| < 3 * 10^{-3}\ , |\sqrt{2} - \frac{577}{408}| < 3 * 10^{-6}\ .$$

1.5 Exercises

Exercise 1.5.1. Determine the discriminant of all integral positive definite forms and all integral irreducible indefinite forms (a,b,c) with $a,b,c \in \{0,\pm 1\}$.

Exercise 1.5.2. Determine all representations of 199 by the form $f = (3,5,7)$.

Exercise 1.5.3. Prove that the form $f(X,Y) = 6X^2 - 5XY - 6Y^2$ is reducible in $\mathbb{Z}[X,Y]$ and determine its factorizations.

Exercise 1.5.4. Prove that the Diophantine equation $x^2 - 5y^2 = 1$ has infinitely many solutions.

Exercise 1.5.5. Prove that the Diophantine equation $x^2 - 2y^2 = -1$ has infinitely many solutions.

Exercise 1.5.6. Prove that the Diophantine equation $x^2 - 5y^2 = 3$ has no solution.

Exercise 1.5.7. Find a shortest non-zero vector in the lattice $L = \mathbb{Z}(1,0) + \mathbb{Z}(1/2, \sqrt{3}/2)$.

Exercise 1.5.8. Prove that a form is semidefinite and not definite if and only if its discriminant is zero.

Chapter references and further reading

[BS96] Eric Bach and Jeffrey Shallit, *Algorithmic number theory*, MIT Press, Cambridge, Massachusetts and London, England, 1996.

2

Equivalence of Forms

In Example 1.1.7, we were able to find the minimum of the form f using a transformation of variables. In this section, we generalize this approach. We introduce transformations that do not change the minimum of a form. Also, the numbers that can be represented by f remain the same. Those transformations will enable us to simplify the representation problem and the minimum problem.

We let $f = (a, b, c)$ be a form with real coefficients.

2.1 Transformation of forms

For

$$U = \begin{pmatrix} s & t \\ u & v \end{pmatrix} \in \mathbb{R}^{(2,2)} \tag{2.1}$$

we define

$$U(X, Y) = (sX + tY, uX + vY) \ . \tag{2.2}$$

Then

$$f\big(U(X,Y)\big) = f(s,u)X^2 + \big(2(ast + cuv) + b(sv + tu)\big)XY + f(t,v)Y^2 \ . \tag{2.3}$$

If $\det U \neq 0$, then we set

$$(fU)(X, Y) = (\det U) f\big(U(X, Y)\big) \ . \tag{2.4}$$

Note that the matrix of fU is

$$M(fU) = \det(U) U^T M(f) U \ . \tag{2.5}$$

Example 2.1.1. Let $f(X, Y) = -13X^2 - 36XY - 25Y^2$ and $U = \begin{pmatrix} -4 & 3 \\ 3 & -2 \end{pmatrix}$. Then $f\big(U(X,Y)\big) = f(-4,3)X^2 + \big(2((-13)\cdot(-4)\cdot 3 - 25\cdot 3\cdot(-2)) - 36(4\cdot 2 + 3\cdot 3)\big)XY + f(3,-2)Y^2 = -X^2 - Y^2$ and $(fU)(X,Y) = X^2 + Y^2$.

The map that sends a pair (f, U) to fU, where f is a form and U is an element of the group $\mathrm{GL}(2, \mathbb{R})$ of all 2 by 2 matrices with real entries and nonzero determinant, defines a right action of $\mathrm{GL}(2, \mathbb{R})$ on the set of all forms (see Section A.2). That is, we have $fI_2 = f$ and $f(UV) = (fU)V$ for any form f and all matrices $U, V \in \mathrm{GL}(2, \mathbb{R})$.

2.2 Equivalence

Let $U \in \mathbb{Z}^{(2,2)}$ with $\det U \neq 0$, let $g = fU$, and let $n \in \mathbb{R}$. If (x, y) is a representation of $(\det U)n$ by g then $(\det U)f\big(U(x, y)\big) = (\det U)n$. Hence, $U(x, y)$ is a representation of n by f. The linear map

$$\mathbb{Z}^2 \to \mathbb{Z}^2 \,, \quad (x, y) \mapsto U(x, y) \tag{2.6}$$

sends representations of $(\det U)n$ by g to representations of n by f.

Example 2.2.1. We use the notation from Example 2.1.1 and set $g = fU$. The vector $(1, 0)$ is a representation of -1 by g. We have $\det U = -1$. Therefore, $U(1, 0) = (-4, 3)$ is a representation of 1 by f.

If the map (2.6) is a bijection, that is, if U is invertible in $\mathbb{Z}^{(2,2)}$, then all representations of n by f can be found by determining all representations of $(\det U)n$ by g.

We describe the group $\mathrm{GL}(2, \mathbb{Z})$ of matrices in $\mathbb{Z}^{(2,2)}$ that have a multiplicative inverse in $\mathbb{Z}^{(2,2)}$. Let U be as in (2.1). The multiplicative inverse of U is

$$U^{-1} = \frac{1}{\det U} \begin{pmatrix} v & -t \\ -u & s \end{pmatrix} , \tag{2.7}$$

that is, we have $UU^{-1} = U^{-1}U = I_2$ where I_2 is the 2 by 2 identity matrix. In general, U^{-1} has rational entries. The inverse U^{-1} has integral entries if and only if $\det U \in \{\pm 1\}$. Hence $\mathrm{GL}(2, \mathbb{Z})$ consists of the matrices in $\mathbb{Z}^{(2,2)}$ with determinant ± 1 and is a group with respect to matrix multiplication. The subgroup of $\mathrm{GL}(2, \mathbb{Z})$ that contains all matrices of determinant 1 is denoted by $\mathrm{SL}(2, \mathbb{Z})$.

Example 2.2.2.
If $U = \begin{pmatrix} 3 & 4 \\ 4 & 5 \end{pmatrix}$, then $\det U = -1$. Hence $U \in \mathrm{GL}(2, \mathbb{Z}) \setminus \mathrm{SL}(2, \mathbb{Z})$. Also, $U^{-1} = \begin{pmatrix} -5 & 4 \\ 4 & -3 \end{pmatrix}$.

Equation (2.4) defines a right action of the groups $\mathrm{GL}(2, \mathbb{Z})$ and $\mathrm{SL}(2, \mathbb{Z})$ on the set of forms. Therefore, equivalence and proper equivalence of forms, as defined below, are equivalence relations.

Definition 2.2.3.

1. *Two forms f and g are called* equivalent *if $g = fU$ with $U \in GL(2,\mathbb{Z})$. The* GL$(2,\mathbb{Z})$*-orbit of a form is called the* equivalence class *of that form.*
2. *Two forms f and g are called* properly equivalent *if $g = fU$ with $U \in SL(2,\mathbb{Z})$. The* SL$(2,\mathbb{Z})$*-orbit of a form is called the* proper equivalence class *of that form.*

We note that two forms f and g are equivalent if and only if their GL$(2,\mathbb{Z})$-orbit is the same. Likewise, f and g are properly equivalent if and only if if their SL$(2,\mathbb{Z})$-orbit is the same.

Example 2.2.4. As we have seen in Example 2.1.1, the forms $f(X,Y) = -13X^2 - 36XY - 25Y^2$ and $g(X,Y) = X^2 + Y^2$ are equivalent. But is is not clear whether they are properly equivalent.

The *equivalence problem* for binary quadratic forms is an important computational problem.

Problem 2.2.5 (Equivalence problem).

1. Given two forms. Decide whether they are equivalent or even properly equivalent.
2. Given two equivalent forms f and g. Find $U \in$ GL$(2,\mathbb{Z})$ with $g = fU$.
3. Given two properly equivalent forms f and g. Find $U \in$ SL$(2,\mathbb{Z})$ with $g = fU$.

2.3 Invariants of equivalence classes of forms

An *invariant* of an equivalence class of forms is a quantity that is the same for all forms in that equivalence class. We will show in this section that the minimum, the content and the discriminant of a form are invariants of its equivalence class.

We first show that properly equivalent forms represent the same numbers.

Proposition 2.3.1. *Let $U \in$ SL$(2,\mathbb{Z})$, let n be a real number, and let M be the set of all representations of n by fU. Then $UM = \{U(x,y) : (x,y) \in M\}$ is the set of all representations of n by f and the map $M \rightarrow UM$, $(x,y) \mapsto U(x,y)$ is a bijection that maps proper representations to proper representations.*

Proof. Since $U \in$ GL$(2,\mathbb{Z})$, the map $M \rightarrow UM$, $(x,y) \mapsto U(x,y)$ is a bijection. Also , if (x,y) is a representation of n by fU then $f\big(U(x,y)\big) = n$. Hence, $U(x,y)$ is a representation of n by f. So UM is the set of all representations of n by f. Finally, by Exercise 2.9.3 proper representations are mapped to proper representations. □

An immediate consequence of Proposition 2.3.1 is the following result:

Corollary 2.3.2. *Two equivalent forms have the same minimum.*

Proposition 2.3.3. *The content of two equivalent integral forms is the same.*

Proof. Let f be an integral form and let $U \in \mathrm{GL}(2, \mathbb{Z})$. It follows from (2.3) that the gcd of the coefficients of f divides the coefficients of fU. Since $f = (fU)U^{-1}$, it follows that the gcd of the coefficients of fU divides the coefficients of f. This proves the assertion. □

Next, we show that the discriminant is an invariant of the equivalence classes of forms.

Proposition 2.3.4. *The discriminant of two equivalent forms is the same.*

Proof. Let $U \in \mathrm{GL}(2, \mathbb{Z})$ and $g = fU$. It follows from (1.17) and (2.5) that $\Delta(fU) = -4 \det M(fU) = -4 \det\big((\det U)U^T M(f)U\big) = -4(\det U)^4 \det$ $M(f)$. Since $(\det U)^2 = 1$ we have $\Delta(fU) = \Delta(f)$. □

It follows from Proposition 2.3.4 and Theorem 1.2.10 that all forms in the equivalence class of an indefinite form are indefinite and that all forms in the proper equivalence class of a positive definite form are positive definite. More precisely, we obtain the following.

Proposition 2.3.5. *The equivalence class of a positive definite form f is the disjoint union of the proper equivalence class of f and the proper equivalence class of fU where U is any matrix in* $\mathrm{GL}(2, \mathbb{Z})$ *with* $\det U = -1$.

Proof. We have already seen that the forms in the proper equivalence class of a positive definite form are positive definite and the forms in the proper equivalence class of a negative definite form are negative definite. Since $\mathrm{GL}(2, \mathbb{Z})$ contains matrices with determinant -1, there is a negative definite form in the equivalence class of any positive definite form. Hence, this equivalence class has the asserted structure. □

As a consequence of Proposition 2.3.3 and Proposition 2.3.5 we obtain the following equivalence relations.

Corollary 2.3.6.

1. *Proper equivalence is an equivalence relation on the set of all integral primitive positive definite forms.*
2. *Equivalence is an equivalence relation on the set of all integral primitive indefinite forms.*

2.4 Two special transformations

We introduce two special elements of $\mathrm{SL}(2, \mathbb{Z})$ that will be used frequently later.

We set

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} . \tag{2.8}$$

Then, for any form $f = (a, b, c)$, we have

$$(a, b, c)S = (c, -b, a) . \tag{2.9}$$

Also

$$S^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} , \quad S^3 = S^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} , \quad S^4 = I_2 . \tag{2.10}$$

Hence, S generates a subgroup of order 4 in $\mathrm{SL}(2, \mathbb{Z})$. We also set

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} . \tag{2.11}$$

Then, by Exercise 2.9.8, we have

$$T^s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} , \quad s \in \mathbb{Z} . \tag{2.12}$$

Therefore, T generates the infinite subgroup

$$\Gamma = \left\{ \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} : s \in \mathbb{Z} \right\} \tag{2.13}$$

of $\mathrm{SL}(2, \mathbb{Z})$. Multiplying $U \in \mathbb{Z}^{(2,2)}$ from the right by T^s means leaving the first column of U unchanged and adding the first column s times to the second column. In Exercise 2.9.1 we prove that the matrices T and S generate $\mathrm{SL}(2, \mathbb{Z})$. It is easy to find out whether two forms are in the same Γ-orbit, where Γ is the subgroup of $\mathrm{SL}(2, \mathbb{Z})$ from (2.13). We explain how this can be done. For $s \in \mathbb{Z}$ we have

$$fT^s = f \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} = (a, b + 2as, as^2 + bs + c) . \tag{2.14}$$

Therefore, we can check whether $f = (a, b, c)$ and f' are in the same Γ-orbit by testing whether $f' = (a, b + 2as, as^2 + bs + c)$ for some integer s. The next proposition modifies this criterion slightly.

Proposition 2.4.1. *Let $f = (a, b, c)$ with $a \neq 0$. Then $f' = (a', b', c')$ is in the same Γ-orbit as f if and only if*

1. $\Delta(f') = \Delta(f)$,
2. $a' = a$, and

3. $b' = b + 2as$ *with* $s \in Z$*.*

If those conditions are satisfied, then $f' = fT^s$*.*

Proof. If $f\Gamma = f'\Gamma$, then the assertions follow from (2.14).

Conversely, suppose that the three conditions are satisfied. We must show that $c' = as^2 + bs + c$. We have $b^2 - 4ac = \Delta(f) = (b')^2 - 4ac' = (b + 2as)^2 - 4ac' = b^2 + 4asb + 4a^2s^2 - 4ac'$. This implies $4a(as^2 + bs + c) = 4ac'$. Since $a \neq 0$, this proves the assertion. □

Example 2.4.2. Let $f = (253, -110, 12)$ and $f' = (253, 2420, 5787)$. We have $2420 = -110 + 2 \cdot 5 \cdot 253$ and $\Delta(f) = \Delta(f') = -44$. Hence, $f' = fT^5$ by Proposition 2.4.1.

2.5 Automorphisms of forms

In this section we study transformations in $\mathrm{GL}(2, \mathbb{Z})$ that, when applied to a form f, leave that form unchanged.

Definition 2.5.1. *A matrix* $U \in \mathrm{GL}(2, \mathbb{Z})$ *with* $fU = f$ *is called an* automorphism *of* f*. A matrix* $U \in \mathrm{SL}(2, \mathbb{Z})$ *with* $fU = f$ *is called a* proper automorphism *of* f*.*

Example 2.5.2. For any form f we have $fI_2 = f(-I_2) = f$. Therefore, I_2 and $-I_2$ are automorphisms of any form.

The set of automorphisms of f is a subgroup of $\mathrm{GL}(2, \mathbb{Z})$. It is called the *automorphism group* of f and it is denoted by $\mathrm{Aut}(f)$. The set of proper automorphisms of f is a subgroup of $\mathrm{SL}(2, \mathbb{Z})$ and of $\mathrm{Aut}(f)$. That subgroup is called the *proper automorphism group* of f and it is denoted by $\mathrm{Aut}^+(f)$. The proper automorphism group of f always contains the elements I_2 and $-I_2$. If those transformations are the only automorphisms of f, then we say that the automorphism group of f is trivial. If those transformations are the only proper automorphisms of f, then we say that the proper automorphism group of f is trivial. Note that by Proposition 2.3.5, every automorphism of a positive definite form is a proper automorphism of that form.

In order to find all representations of a real number n by f it is useful to determine the proper automorphism group of f. If (x, y) is a (proper) representation of a real number n by f, then for any proper automorphism U of f the vector $U(x, y)$ is also a (proper) representation of n by f. In other words, the proper automorphism group $\mathrm{Aut}^+(f)$ acts on the set of all representations of n by f and also on the set of all proper representations of n by f.

Definition 2.5.3. *Two representations* (x, y) *and* (x', y') *of a real number* n *by* f *are called* equivalent *if there is a proper automorphism* U *of* f *with* $(x', y') = U(x, y)$*.*

Equivalence of representations is an equivalence relation.

2.5.1 Non-integral forms

If r is a non-zero real number then the automorphism group of rf is the same as the automorphism group of f. We give a necessary condition for a form to have a non-trivial automorphism group.

Theorem 2.5.4. *If* $\operatorname{Aut}(f)$ *is nontrivial, then there is a positive integer* r *such that the form* rf *has integer coefficients.*

Proof. Let $U \in \mathrm{GL}(2,\mathbb{Z})$ such that $fU = f$. Then $M(f) = (\det U)U^T M(f)U$. First assume that $\det U = 1$. Then

$$(U^T)^{-1}M(f) = M(f)U \ . \tag{2.15}$$

Let U be as in (2.1). Then

$$M(f)U = \frac{1}{2}\begin{pmatrix} 2as+bu & 2at+bv \\ 2cu+bs & 2cv+bt \end{pmatrix} \tag{2.16}$$

and

$$(U^T)^{-1}M(f) = \frac{1}{2}\begin{pmatrix} 2av-bu & -2cu+bv \\ -2at+bs & 2cs-bt \end{pmatrix} \ . \tag{2.17}$$

From (2.15), (2.16), and (2.17) we obtain

$$at = -cu \ , \quad bu = a(v-s) \ , \quad bt = c(s-v) \ . \tag{2.18}$$

If $U \neq \pm I_2$ then $u \neq 0$ or $t \neq 0$ or $s - v \neq 0$. If $u \neq 0$ then (2.18) implies $(a,b,c) = (a/u)(u,(v-s),-t)$. If $t \neq 0$ then (2.18) implies $(a,b,c) = (c/t)(-u,(s-v),t)$. If $s - v \neq 0$ then $(a,b,c) = (b/(v-s))(u, v-s, -t)$. If $\det U = -1$ the proof is analogous. □

2.5.2 Integral forms

We describe the connection between the automorphism group of a primitive integral form $f = (a,b,c)$ of discriminant Δ and the *Pell equation*

$$x^2 - \Delta y^2 = \pm 4 \ . \tag{2.19}$$

For two real numbers x, y we define the matrix

$$U(f,x,y) = \begin{pmatrix} (x-yb)/2 & -cy \\ ay & (x+yb)/2 \end{pmatrix} \ . \tag{2.20}$$

The connection between the Pell equation and $\operatorname{Aut}(f)$ is described in the next theorem.

Theorem 2.5.5. *Let f be a primitive form.*

1. *The map that sends a solution $(x, y) \in \mathbb{Z}^2$ of the Pell equation $x^2 - \Delta y^2 = \pm 4$ to the matrix $U(f, x, y)$ is a bijection between the set of those solutions and the automorphism group of f.*
2. *The map that sends a solution $(x, y) \in \mathbb{Z}^2$ of the Pell equation $x^2 - \Delta y^2 = 4$ to the matrix $U(f, x, y)$ is a bijection between the set of those solutions and the proper automorphism group of f.*

Proof. By Exercise 2.9.6, the map from the theorem sends solutions of the Pell equation to automorphisms of f. The map is clearly injective. To prove the surjectivity, we let U be an automorphism of f, U as in (2.1). If $a \neq 0$, then $\gcd(a, b, c) = 1$ and (2.18) imply that a divides u. Let $x = s + v$ and $y = u/a$. Then $U = U(f, x, y)$ by (2.18). Hence, $x^2 - \Delta y^2 = 4 \det U \in \{\pm 4\}$ which means that (x, y) is a solution of the Pell equation. The case $a = 0$ and the second assertion are left to the reader as an exercise. □

Note that the proof of Theorem 2.5.5 contains an explicit formula for the inverse maps in the case where $a \neq 0$.

Example 2.5.6. Let $\Delta \equiv 0 \pmod 4$. Then the form $f = (1, 0, -\Delta/4)$ has discriminant Δ. If (x, y) is a solution of the Pell equation (2.19), then the corresponding automorphism is

$$U(f, x, y) = \begin{pmatrix} x/2 & \Delta y/4 \\ y & x/2 \end{pmatrix} . \tag{2.21}$$

Conversely, if

$$\begin{pmatrix} s & t \\ u & v \end{pmatrix} \in \operatorname{Aut}(f) \tag{2.22}$$

then $(x, y) = (2s, u)$ is the corresponding solution of the Pell equation.

Let $\Delta \equiv 1 \pmod 4$. Then the form $f = \big(1, 1, (1 - \Delta)/4\big)$ has discriminant Δ. If (x, y) is a solution of the Pell equation (2.19), then the corresponding automorphism is

$$U(f, x, y) = \begin{pmatrix} (x - y)/2 & (\Delta - 1)y/4 \\ y & (x + y)/2 \end{pmatrix} . \tag{2.23}$$

Conversely, if

$$\begin{pmatrix} s & t \\ u & v \end{pmatrix} \in \operatorname{Aut}(f) \tag{2.24}$$

then $(x, y) = (2s + v, u)$ is the corresponding solution of the Pell equation.

Here is another important observation.

Theorem 2.5.7. *The automorphism groups of all integral primitive forms of a fixed discriminant are isomorphic.*

Proof. Let f and g be two integral primitive forms of the same discriminant Δ. By Theorem 2.5.5, the map that sends the automorphism $U(f,x,y)$ of f to the automorphism $U(g,x,y)$ of g, where (x,y) is a solution of the Pell equation (2.19), is a bijection. Also, if (x',y') is another solution of the Pell equation, then by Exercise 2.9.7 the pair $\big((xx'+yy'\Delta)/2,(xy'+x'y)/2\big)$ is also such a solution and

$$U(f,x,y)U(f,x',y') = U\big(f,(xx'+yy'\Delta)/2,(xy'+x'y)/2\big)\,. \tag{2.25}$$

This proves that our map is a homomorphism. □

2.5.3 Positive definite forms

We can now determine the automorphism group of all integral positive definite forms.

Example 2.5.8. We determine the automorphism group of the integral primitive forms of discriminant -4. By Theorem 2.5.7, it suffices to consider the form $f=(1,0,1)$. The corresponding Pell equation is $x^2+4y^2=4$. The only solutions of that equation are $(\pm 2,0)$ and $(0,\pm 1)$. Theorem 2.5.5 implies

$$\operatorname{Aut}(1,0,1) = \left\{\pm\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \pm\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\right\}\,. \tag{2.26}$$

This automorphism group is the cyclic group of order four generated by S from (2.8).

Example 2.5.9. We determine the automorphism group of the integral primitive forms of discriminant -3. By Theorem 2.5.7, it suffices to consider the form $f=(1,1,1)$. Its automorphism group is the cyclic group generated by

$$\begin{pmatrix}0 & -1\\ 1 & 1\end{pmatrix}\,.$$

It is of order 6. This is shown in Exercise 2.9.9.

Finally, we determine the remaining automorphism groups.

Theorem 2.5.10. *If f is an integral primitive positive definite form of discriminant $\Delta \neq -3,-4$, then the automorphism group of f is trivial.*

Proof. The corresponding Pell equation is $x^2+|\Delta|y^2=4$. Since $|\Delta|\geq 7$ the only solutions of that equation are $(x,y)=(\pm 2,0)$. Hence, the assertion follows from Theorem 2.5.5. □

2.5.4 Indefinite forms

If f is indefinite, that is, if $\Delta > 0$, then the solution of the Pell equation and the computation of the automorphism group of f are more difficult. We will prove in Section 6.12 that the automorphism group of an integral irreducible form of positive discriminant is $\{\pm T^k : k \in \mathbb{Z}\}$, where T is an automorphism of infinite order, the so called *fundamental automorphism* of f. Here, we only give an example.

Example 2.5.11. Consider the indefinite form $(1, 0, -2)$. The corresponding Pell equation is $x^2 - 8y^2 = \pm 4$. A solution of that equation is $(2, 1)$. An automorphism of this form is

$$U = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} .$$

The first few powers of U are

$$U^2 = \begin{pmatrix} 3 & 4 \\ 2 & 3 \end{pmatrix} , \quad U^3 = \begin{pmatrix} 7 & 10 \\ 5 & 7 \end{pmatrix} , \quad U^4 = \begin{pmatrix} 17 & 24 \\ 12 & 17 \end{pmatrix} .$$

This shows that the automorphism group of this form is infinite. We will see in Section 6.12 that U is the fundamental automorphism of the form $(1, 0, -2)$ that is, $\mathrm{Aut}(1, 0, -1) = \{\pm U^k : k \in \mathbb{Z}\}$.

2.6 A strategy for finding proper representations

Let n be a real number. In this section we will show that the proper representations of n by a form f correspond to the forms (n, B, C) in the proper equivalence class of f. Based on this observation, we will explain a strategy for finding all proper representations of n by f.

Let $(s, u) \in \mathbb{Z}^2$ be a proper representation of n by f. By Exercise 2.9.4 there is a matrix $U \in \mathrm{SL}(2, \mathbb{Z})$ with first column (s, u). By (2.3) we have $fU = \big(f(x, y), B, C\big) = (n, B, C)$ with real coefficients B, C.

Example 2.6.1. Consider the form $f(X, Y) = 3X^2 + 4XY + 5Y^2$. The pair $(4, 5)$ is a proper representation of 253 by f. For

$$U = \begin{pmatrix} 4 & -1 \\ 5 & -1 \end{pmatrix} ,$$

we have $\det U = 1$ and $fU = (253, -110, 12)$.

From a proper representation of n by f we have constructed a form (n, B, C) that is properly equivalent to f. This construction is not unique, since for any $V \in \Gamma$ we have $(n, B, C)V = (n, B', C')$. We will now precisely describe the correspondence between proper representations of n by f and the forms (n, B, C).

Proposition 2.6.2. *The map that sends the equivalence class of a proper representation (s, u) of n by f to the Γ-orbit of the form fU, where U is any matrix in* $\mathrm{SL}(2, \mathbb{Z})$ *with first column (s, u), defines a bijection between the equivalence classes of proper representations of n by f and the Γ-orbits of forms (n, B, C) in the proper equivalence class of f.*

Proof. We prove that the map from the proposition is well defined. Let (x, y) and (x', y') be equivalent proper representations of n by f. Then there is a proper automorphism V of f with $V(x', y') = (x, y)$. Let $U, U' \in \mathrm{SL}(2, \mathbb{Z})$ such that the first column of U is (x, y) and the first column of U' is (x', y'). We must show that $fU'\Gamma = fU\Gamma$. The first column of VU' is (x, y). By Exercise 2.9.4, we have $U\Gamma = VU'\Gamma$. Since V is an automorphism of f, we have $fU\Gamma = fVU'\Gamma = fU'\Gamma$. This shows that the map is well defined.

Next, we prove that the map is surjective. Suppose that (n, B, C) is properly equivalent to f. Let $U \in \mathrm{SL}(2, \mathbb{Z})$ such that $fU = (n, B, C)$ and let (x, y) be the first column of U. Then, as we have seen above, the image of (x, y) is the Γ-orbit of f.

Finally, we prove the injectivity. Suppose that (x, y) and (x', y') are proper representations of n by f whose image is the same Γ-orbit of forms. Then there are $U \in \mathrm{SL}(2, \mathbb{Z})$ with first column (x, y) and $U' \in \mathrm{SL}(2, \mathbb{Z})$ with first column (x', y') such that $fU = fU'$. Hence $V = U'U^{-1}$ is a proper automorphism of f of determinant 1 with $V(x', y') = (x, y)$. This shows that (x, y) and (x', y') are equivalent. □

Example 2.6.3. The pair $(2, 3)$ is a proper representation of 13 by $f = X^2 + Y^2$. In Example 2.5.8 we have shown that the proper automorphism group of f is $\{I_2, S, S^2, S^3\}$. Hence, the equivalence class of the representation $(2, 3)$ is $\{\pm(2, 3), \pm(-3, 2)\}$. The map from Proposition 2.6.2 sends this equivalence class to the Γ-orbit of $(13, -36, 25)$.

Theorem 2.6.2 reduces the problem of finding proper representations of n by f to the following three problems:

1. Find all Γ-orbits of forms (n, B, C) of discriminant $\Delta(f)$.
2. For each Γ-orbit of forms (n, B, C) of discriminant $\Delta(f)$, decide whether its elements are properly equivalent to f. If the elements of such a Γ-orbit are properly equivalent to f find $U \in \mathrm{SL}(2, \mathbb{Z})$ such that fU is in that orbit. The first column of U is a proper representation of n by f. Collect all those representations.
3. Determine the proper automorphism group $\mathrm{Aut}^+(f)$ and the $\mathrm{Aut}^+(f)$-orbit $\mathrm{Aut}^+(f)(x, y)$ for each proper representation found in step 2.

If (a, b, c) is an integral form, then in the first step it suffices to find all Γ-orbits of forms (n, B, C) that have the same content as f. In particular, if f is primitive, it suffices to find the primitive forms (n, B, C) of discriminant Δ.

Example 2.6.4. We determine all proper representations of 1 by $f = X^2 + Y^2$. The discriminant of f is $\Delta(f) = -4$. So we compute all Γ-orbits of forms

$(1, B, C)$ of discriminant -4. By (2.14) any such Γ-orbit contains a form $(1, B, C)$ with $0 \leq B \leq 1$. Hence we must find all pairs $(B, C) \in \{0, 1\} \times \mathbb{Z}$ such that $B^2 - 4C = -4$. Any such B must be even. Hence, $B = 0$ and $C = 1$. So $(1, 0, 1)\Gamma = f\Gamma$ is the only Γ-orbit of forms $(1, B, C)$. Since $f = fI_2$ belongs to that Γ-orbit. The first column of I_2 is $(1, 0)$. So $(1, 0)$ is a proper representation of 1 by f. By Example 2.5.8 the automorphism group of f is $\mathrm{Aut}(f) = \{I_2, S, S^2, S^3\}$. The $\mathrm{Aut}(f)$-orbit of $(1, 0)$ is $\{(1, 0), (0, 1), (-1, 0), (0, -1)\}$. These are all proper representations of 1 by f.

2.7 Determining improper representations

In the previous section, we have explained how to find proper representations of a real number n by f. If (x, y) is a representation of n by f and if $d = \gcd(x, y)$ then $(x/d, y/d)$ is a proper representation of n/d^2 by f. If f is an integral form and n is a non-zero integer, this observation can be used to find all representations of n by f as follows:

1. Find all positive integers d whose square divides n.
2. For each d with $d^2 \mid n$ determine all proper representations (x, y) of n/d^2 by f. Collect all (dx, dy).

2.8 Ambiguous classes

We define ambiguous equivalence classes of forms.

Definition 2.8.1. *An equivalence class of integral indefinite irreducible forms is called* ambiguous *if (a, b, c) is equivalent to $(a, -b, c)$ for every form (a, b, c) in that equivalence class.*

We give simple characterizations for ambiguous classes. Let $f = (a, b, c)$ be an integral irreducible indefinite form.

Lemma 2.8.2. *The following statements are equivalent:*

1. *The equivalence class of (a, b, c) is ambiguous.*
2. *The form (a, b, c) is equivalent to $(a, -b, c)$.*
3. *The form (a, b, c) is equivalent to (c, b, a).*

Proof. If the equivalence class of (a, b, c) is ambiguous, then, by definition, (a, b, c) is equivalent to $(a, -b, c)$. Also, if (a, b, c) is equivalent to $(a, -b, c)$ then (a, b, c) is equivalent to (c, b, a) since $(a, -b, c)$ and (c, b, a) are equivalent. Also, for the same reason, if (a, b, c) is equivalent to (c, b, a), then (a, b, c) is equivalent to $(a, -b, c)$.

Finally, let $(a, b, c)U = (c, b, a)$ with $U \in \mathrm{GL}(2, \mathbb{Z})$. We show that the equivalence class of (a, b, c) is ambiguous. For this purpose, let the form $F = (A, B, C)$ be equivalent to f, say $F = fV$ with some $V \in \mathrm{GL}(2, \mathbb{Z})$. We must show that (A, B, C) is equivalent to (C, B, A).

Let

$$W = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

Denote (c, b, a) by f^w and (C, B, A) by F^w. Then $f^w = -fW$, and $F^w = -FW$. Hence

$$F^w = -FW = -fVW = f^w WVW = fUWVW = FV^{-1}UWVW. \qquad \square$$

Proposition 2.8.3. *The equivalence class of an ambiguous form is ambiguous.*

Proof. If a form $f = (a, b, c)$ is ambiguous, then f is equivalent to $(a, 0, C)$ or to (a, a, C) for some integer C. By Lemma 2.8.2 the equivalence classes of those forms are ambiguous since $(a, 0, C)$ is equal to $(a, -0, C)$ and (a, a, C) is in the same Γ-orbit as $(a, -a, C)$. □

2.9 Exercises

Exercise 2.9.1. Prove that for each $U \in \mathrm{SL}(2, \mathbb{Z})$ there exists $k \in \mathbb{Z}_{\geq 0}$ and exponents $s_i, t_i \in \mathbb{Z}$, $1 \leq i \leq k$ such that

$$U = T^{s_1} S^{t_1} T^{s_2} S^{t_2} T^{s_3} \cdots S^{t_k} T^{s_k} .$$

Exercise 2.9.2. Determine the representation of

$$U = \begin{pmatrix} 6 & 11 \\ 13 & 24 \end{pmatrix}$$

as a power product of S and T as in Exercise 2.9.1.

Exercise 2.9.3. Prove that $\gcd(x, y) = \gcd\big(U(x, y)\big)$ for $(x, y) \in \mathbb{Z}^2$ and $U \in \mathrm{GL}(2, \mathbb{Z})$.

Exercise 2.9.4. Show that for any pair (x, y) of integers with $(x, y) \neq (0, 0)$ there is a matrix $U \in \mathbb{Z}^{(2,2)}$ with first column (x, y) and determinant $\gcd(x, y)$. Also show that the set of all such matrices is the Γ-orbit of any such matrix U.

Exercise 2.9.5. Prove that the map that sends a pair (r, f) consisting of a real number r and a form f to their product rf defines an action of the set $\mathbb{R}_{>0}$ of all positive real numbers on the set of positive definite forms. Also show that this map defines an action of $\mathbb{R}$ on the set of all indefinite forms. Show in both cases that the number of orbits is infinite.

Exercise 2.9.6. Let f be an integral form of discriminant Δ and let (x, y) be integers such that $x^2 - \Delta y^2 = \pm 4$. Prove that $U(f, x, y)$ is an automorphism of f.

Exercise 2.9.7. Let Δ be an integer, $\Delta \equiv 0, 1 \pmod 4$. Let $(x, y) \in \mathbb{Z}^2$ and $(x', y') \in \mathbb{Z}^2$ be solutions of the Pell equation $x^2 - y^2\Delta = \pm 4$. Prove that $(xx' + yy'\Delta, xy' + x'y)$ is also such a solution and that $U(f, x, y)U(f, x', y') = U\big(f, (xx' + yy'\Delta)/2, (xy' + x'y)/2\big)$ for any integral primitive form f of discriminant Δ.

Exercise 2.9.8. Prove (2.12)

Exercise 2.9.9. Verify Example 2.5.9.

Exercise 2.9.10. Determine the first three primes that are represented by $f = (1, 1, 1)$. For any of those primes p determine a form (p, b, c) that is equivalent to $(1, 1, 1)$.

Exercise 2.9.11. Determine all proper representations of 3 by $f = (1, 1, 1)$.

Exercise 2.9.12. Prove that for any discriminant there is exactly one Γ-orbit $(1, b, c)\Gamma$ of forms of discriminant Δ. Construct this Γ-orbit

Exercise 2.9.13. Show that the automorphism group of the form $(1, 0, -5)$ is infinite.

3

Constructing Forms

In Chapter 2 we have explained a strategy for finding primitive representations of an integer a by an integral form f. In this strategy, the first step is to determine all integral forms (a, b, c) of discriminant $\Delta(f)$. In this chapter, we explain how those forms can be found.

In this chapter, a form is always an integral form. We fix an integer a and a discriminant Δ. Except for Section 3.2 the integer a is assumed to be positive. We set

$$\mathcal{F}(\Delta, a) = \Big\{(a, b, c)\Gamma : b, c \in \mathbb{Z}, \Delta(a, b, c) = \Delta\Big\} . \tag{3.1}$$

We will see that this set is finite and we will determine its cardinality

$$R(\Delta, a) = \big|\mathcal{F}(\Delta, a)\big| . \tag{3.2}$$

We also set

$$\mathcal{F}^*(\Delta, a) = \Big\{(a, b, c)\Gamma : b, c \in \mathbb{Z}, \gcd(a, b, c) = 1,\ \Delta(a, b, c) = \Delta\Big\} . \tag{3.3}$$

and

$$R^*(\Delta, a) = \big|\mathcal{F}^*(\Delta, a)\big| . \tag{3.4}$$

3.1 Reduction to finding square roots of Δ modulo $4a$

If (a, b, c) is a form of discriminant Δ then $\Delta = b^2 - 4ac$. This implies that

$$\Delta \equiv b^2 \pmod{4a} . \tag{3.5}$$

Hence, Δ is a square modulo $4a$. In the next proposition, we show how to find $\mathcal{F}(\Delta, a)$ by determining the square roots of Δ modulo $4a$.

Proposition 3.1.1. *The map*

$$\mathcal{F}(\Delta, a) \to \{b + 2a\mathbb{Z} : b \in \mathbb{Z}, b^2 \equiv \Delta \pmod{4a}\}\ , \quad (a, b, c)\Gamma \mapsto b + 2a\mathbb{Z}$$

is a bijection.

Proof. By (2.14) and (3.5), the map is well defined and injective. Also, if $b \in \mathbb{Z}$ with $b^2 \equiv \Delta \pmod{4a}$, then we can write $\Delta = b^2 - 4ac$ for some integer c. Hence $b + 2a\mathbb{Z}$ is the image of $(a, b, c)\Gamma$. □

By Proposition 3.1.1, we have

$$R(\Delta, a) = |\{b \in \mathbb{Z} : b^2 \equiv \Delta \pmod{4a}, -a < b \le a\}|. \tag{3.6}$$

Hence, we obtain the following corollary.

Corollary 3.1.2. *The sets $\mathcal{F}(\Delta, a)$ and $\mathcal{F}^*(\Delta, a)$ are finite.*

In the next example, we use Proposition 3.1.1 to determine $\mathcal{F}(\Delta, a)$ and $\mathcal{F}^*(\Delta, a)$.

Example 3.1.3. Let $a = 3$. The squares modulo $4a = 12$ in $\{0, 1, \ldots, 11\}$ are $0, 1, 4, 9$. Hence, a discriminant Δ is a square modulo $4a$ if $\Delta \pmod{12} \in \{0, 1, 4, 9\}$. We determine the square roots of those squares in $M = \{-2, -1, \ldots, 3\}$. The only square root of 0 in M is 0 and the only square root of 9 in M is 3. The square roots of 1 in M are ± 1 and the square roots of 4 in M are ± 2.

If $\Delta = -3$, then $\Delta \equiv 9 \pmod{12}$. Hence Δ is a square modulo 12 and since the only square root of 9 in M is 3 it follows that $\mathcal{F}(-3, 3) = \mathcal{F}^*(-3, 3) = \big\{(3, 3, 1)\Gamma\big\}$ and $R(-3, 3) = R^*(3, -3) = 1$.

If $\Delta = -4$, then $\Delta \equiv 8 \pmod{12}$. Hence, there is no form $(3, b, c)$ of discriminant -4 and $R(-4, 3) = 0$.

If $\Delta = 5$, then again, there is no form $(3, b, c)$ of discriminant Δ and $R(5, 3) = 0$.

If $\Delta = 13$, then $\Delta \equiv 1 \pmod{12}$. Hence, $\mathcal{F}(13, 3) = \mathcal{F}^*(13, 3) = \big\{(3, \pm 1, -1)\Gamma\big\}$ and $R(13, 3) = R^*(13, 3) = 2$.

If $\Delta = -12$, then $\Delta \equiv 4 \pmod{8}$. Hence $\mathcal{F}(-12, 2) = \big\{(2, 2, 2)\big\}$, and $\mathcal{F}^*(-12, 2) = \emptyset$, $R(-12, 2) = 1$, and $R^*(-12, 2) = 0$.

3.2 The case $a < 0$

We have

$$\mathcal{F}(\Delta, |a|) = \{\big(|a|, b, \operatorname{sign}(a)c\big)\Gamma : (a, b, c)\Gamma \in \mathcal{F}(\Delta, a)\}$$

and

$$\mathcal{F}^*(\Delta, |a|) = \{\big(|a|, b, \operatorname{sign}(a)c\big)\Gamma : (a, b, c)\Gamma \in \mathcal{F}^*(\Delta, a)\}.$$

Hence, we have

$$R(\Delta, a) = R(\Delta, |a|)\ ,\ \ R^*(\Delta, a) = R^*(\Delta, |a|)\ .$$

It follows that it suffices to compute $R^*(\Delta, a)$ for non-negative integers a. In the sequel, we assume that $a > 0$. The case $a = 0$ is treated in Exercise 3.7.1.

3.3 Fundamental discriminants and conductor

We introduce a few important notions. By a *discriminant* we mean an integer $\Delta \neq 0$ with $\Delta \equiv 0 \pmod 4$ or $\Delta \equiv 1 \pmod 4$.

Example 3.3.1. The first few discriminants are $1, 4, 5, 8, 9, 12, 13$ and $-3, -4, -7, -8, -11, -12$.

Definition 3.3.2.

1. *The* conductor *of a discriminant* Δ *is the largest positive integer* f *such that* Δ/f^2 *is a discriminant. We denote it by* $f(\Delta)$.
2. *A* fundamental discriminant *is a discriminant* Δ *that has conductor* 1.

We give a more explicit formula for the conductor of a discriminant Δ. Let

$$|\Delta| = \prod_{p|\Delta} p^{e(p)}$$

be the prime factorization of $|\Delta|$. Then

$$f = \prod_{p|\Delta} p^{e'(p)}$$

where

$$2e'(2) = \begin{cases} e(2) - (e(2) \bmod 2) & \text{if } \Delta/2^{e(2)} \equiv 1 \pmod 4\ , \\ e(2) - (e(2) \bmod 2) - 2 & \text{if } \Delta/2^{e(2)} \equiv 3 \pmod 4\ , \end{cases} \tag{3.7}$$

and

$$2e'(p) = e(p) - (e(p) \bmod 2) \tag{3.8}$$

for all odd primes. Note that if Δ is a non-zero discriminant, then $f(\Delta)$ is the uniquely determined positive integer f such that Δ/f^2 is a fundamental discriminant.

Example 3.3.3. We have $f(1) = 1$, $f(4) = 2$, $f(5) = 1$, $f(8) = 1$, $f(9) = 3$, $f(12) = 1$, $f(13) = 1$.

We describe the fundamental discriminants more explicitly. We call a non-zero integer *square free* if it has no multiple prime divisors.

Proposition 3.3.4. *Let Δ be an integer. Then Δ is a fundamental discriminant if and only if*

1. $\Delta \equiv 1 \pmod 4$ *and* Δ *is square free, or*
2. $\Delta \equiv 0 \pmod 4$, $\Delta/4 \equiv 2, 3 \pmod 4$, *and* $\Delta/4$ *is square free.*

Proof. Exercise 3.7.4.

Example 3.3.5. 12 is a fundamental discriminant since the only square that divides 12 is 4 and $12/4 = 3$ is not a discriminant.

-12 is not a fundamental discriminant since 4 divides -12 and $-12/4 = -3$ is a discriminant.

3.4 The case of a prime number

Definition 3.4.1. *A* prime form *is a form (p, b, c) where p is a prime number.*

In this section, we determine all prime forms of discriminant Δ. Let p be a prime number. By Proposition 3.1.1, such a prime form can only exist if Δ is a square modulo $4p$.

We define quadratic residues.

Definition 3.4.2. *Let m be a positive integer and let x be an integer that is coprime to m. Then x is called a* quadratic residue modulo m *if there is a solution $y \in \mathbb{Z}$ of the congruence $y^2 \equiv x \pmod m$ and a* quadratic nonresidue modulo m *otherwise.*

We introduce the Legendre symbol. It will be used to give an easy formula for $R(\Delta, p)$ and for $R^*(\Delta, p)$.

Definition 3.4.3. *For an integer m and an odd prime p, the* Legendre symbol $\left(\frac{m}{p}\right)$ *is defined as follows:*

1. If p divides m, then $\left(\frac{m}{p}\right) = 0$.
2. If p does not divide m, then

$$\left(\frac{m}{p}\right) = \begin{cases} 1 & \textit{if } m \textit{ is a quadratic residue modulo } p, \\ -1 & \textit{if } m \textit{ is a quadratic nonresidue modulo } p. \end{cases}$$

In addition, to treat the case $p = 2$ we set

$$\left(\frac{m}{2}\right) = \begin{cases} 0 & \text{if } m \text{ is even,} \\ (-1)^{(m^2-1)/8} & \text{if } m \text{ is odd.} \end{cases} \tag{3.9}$$

This is a special case of the Kronecker symbol that will be defined in Section 3.4.3.

Note that

$$\left(\frac{m}{2}\right) = \begin{cases} -1 & \text{if } m \equiv 3, 5 \pmod 8, \\ 1 & \text{if } m \equiv 1, 7 \pmod 8. \end{cases}$$

Now we determine the set of square roots of Δ mod $4p$. It is empty if $\left(\frac{\Delta}{p}\right) = -1$. If $\left(\frac{\Delta}{p}\right) = 0$, we define

$$b(\Delta, p) = \begin{cases} 0 & \text{if } p = 2 \text{ and } \Delta \equiv 0 \pmod 8, \\ 2 & \text{if } p = 2 \text{ and } \Delta \equiv 4 \pmod 8, \\ 0 & \text{if } p \text{ is odd and } \Delta \text{ is even}, \\ p & \text{if } p \text{ is odd and } \Delta \text{ is odd}. \end{cases} \tag{3.10}$$

Lemma 3.4.4.

1. *If $\left(\frac{\Delta}{p}\right) = 0$, then the set of square roots of Δ mod $4p$ is $b(\Delta, p) + 2p\mathbb{Z}$.*
2. *If $\left(\frac{\Delta}{p}\right) = 1$, then there is a uniquely determined square root $b(\Delta, p)$ of Δ mod $4p$ in $\{1, \ldots p-1\}$. Also, the set of square roots of Δ mod $4p$ is $\pm b(\Delta, p) + 2p\mathbb{Z}$.*

Proof. If $p = 2$, then $b(\Delta, p) = 1$.

Let p be an odd prime. Since $\left(\frac{\Delta}{p}\right) = 1$, there is a square root of Δ mod p in $\{1, \ldots, p-1\}$. Let r be such a square root. Set

$$b(\Delta, p) = \begin{cases} r & \text{if } r \equiv \Delta \pmod 2, \\ p - r & \text{otherwise}. \end{cases}$$

Then $b(\Delta, p)$ is a square root of Δ mod $4p$. Let b be any square root of Δ mod $4p$. Since $\mathbb{Z}/p\mathbb{Z}$ is a field, the polynomial $x^2 - \Delta$ has exactly two square roots in $\mathbb{Z}/p\mathbb{Z}$. They are $\pm r + p\mathbb{Z}$. Hence $b \equiv \pm r \pmod p$. Also $b \equiv \Delta \pmod 2$. It follows that $b \equiv \pm b(\Delta, p) \pmod{2p}$. □

From the proof of Lemma 3.4.4 we obtain algorithm `sqrtMod4P` below. It reduces the problem of finding $b(\Delta, p)$ to the problem of computing square

Algorithm 3.1 `sqrtMod4P` (Δ, p)

Input: A discriminant Δ and a prime p such that Δ is a square modulo p.
Output: The square root $b(\Delta, p)$ of Δ modulo $4p$.

```
if $p = 2$ then
   if $\Delta$ is even then return $2(\Delta/4 \bmod 2)$
   if $\Delta$ is odd then return 1
else
   $r \leftarrow$ sqrtModP$(\Delta, p)$
   if $r \equiv \Delta \pmod 2$ then return $r$
   else return $p - r$
```

roots modulo odd primes. The algorithm uses `sqrtModP`(Δ, p) which computes a square root of Δ mod p in $\{0, \ldots, p\}$. For example, the algorithm `sqrtModP`, which is explained in Section 3.4.4, can be used for this purpose.

We can now present formulas for $R(\Delta, p)$ and for $R^*(\Delta, p)$. We also explain how to compute $\mathcal{F}(\Delta, p)$ and $\mathcal{F}^*(\Delta, p)$. For $\left(\frac{\Delta}{p}\right) \neq -1$ we define

$$c(\Delta, p) = \frac{b(\Delta, p)^2 - \Delta}{4p}. \tag{3.11}$$

Note, that $c(\Delta, p)$ is an integer.

Proposition 3.4.5. *1. If $\left(\frac{\Delta}{p}\right) = -1$, then $R(\Delta, p) = R^*(\Delta, p) = 0$.*

2. If $\left(\frac{\Delta}{p}\right) = 0$ and p does not divide the conductor of Δ, then $R(\Delta, p) = R^*(\Delta, p) = 1$ and $\mathcal{F}(\Delta, p) = \mathcal{F}^*(\Delta, p) = \{(p, b(\Delta, p), c(\Delta, p))\Gamma\}$.

3. If $\left(\frac{\Delta}{p}\right) = 0$ and p divides the conductor of Δ, then $R(\Delta, p) = 1$, $R^*(\Delta, p) = 0$ and $\mathcal{F}(\Delta, p) = \{(p, b(\Delta, p), c(\Delta, p))\Gamma\}$.

4. If $\left(\frac{\Delta}{p}\right) = 1$, then $R(\Delta, p) = R^*(\Delta, p) = 2$ and $\mathcal{F}(\Delta, p) = \mathcal{F}^*(\Delta, p) = \{(p, \pm b(\Delta, p), c(\Delta, p))\Gamma\}$.

Proof. The formulas for $R(\Delta, p)$ and $\mathcal{F}(\Delta, p)$ follow from Lemma 3.4.4 and Proposition 3.1.1.

We determine the primitive forms. If $\left(\frac{\Delta}{p}\right) = 0$, then the form $(p, b(\Delta, p), c(\Delta, p))$ is primitive if and only if p does not divide the conductor of Δ (see Exercise 3.7.5). Also, if $\left(\frac{\Delta}{p}\right) = 1$, then $b(\Delta, p)$ is not divisible by p. Hence all forms (p, b, c) of discriminant Δ are primitive. □

Note that

$$R(\Delta, p) = \left(\frac{\Delta}{p}\right) + 1. \tag{3.12}$$

Algorithm `numberOfPrimeForms` calculates $R(\Delta, p)$ as described in Proposition 3.4.5. For $\left(\frac{\Delta}{p}\right) \geq 0$, Algorithm `primeForm` calculates the form $(p, b(\Delta, p), c(\Delta, p))$ as described in that proposition.

Example 3.4.6. By Proposition 3.4.5, we have $\mathcal{F}(29, 7) = \mathcal{F}^*(29, 7) = \{(7, \pm 1, -1)\Gamma\}$ since $\left(\frac{29}{7}\right) = 1$ and $b(29, 7) = 1$.

Algorithm 3.2 `numberOfPrimeForms` $(\Delta,\ p)$

Input: A discriminant Δ and a prime number p
Output: $R(\Delta, p)$

```
  return $\left(\frac{\Delta}{p}\right) + 1$
```

Algorithm 3.3 `primeForm` $(\Delta,\ p)$

Input: A discriminant Δ and a prime number p with $R(\Delta, p) > 0$
Output: The form $\bigl(p, b(\Delta, p), c(\Delta, p)\bigr)$

```
$b \leftarrow$ sqrtMod4P$(\Delta, p)$
return $\bigl(p, b, (b^2 - \Delta)/(4p)\bigr)$
```

3.4.1 The Euler criterion

Here is a simple formula for the Legendre symbol. If fast exponentiation techniques are used (see [Buc04]), then this formula can be used to efficiently evaluate the Legendre symbol.

Theorem 3.4.7 (Euler criterion). *If p is an odd prime and if y is an integer which is not divisible by p, then $\left(\frac{y}{p}\right) \equiv y^{(p-1)/2} \pmod p$*

Proof. Let g be a primitive root modulo p and let $u = g^{(p-1)/2}$. Then $u^2 \equiv 1 \pmod p$ and $u \not\equiv 1 \pmod p$. So u is a zero of the polynomial $X^2 - 1 \bmod p$. The only zeros mod p of this polynomial are ± 1. Hence,

$$g^{(p-1)/2} \equiv -1 \pmod p$$

Also, there is $k \in \{0, 1, \ldots, p-1\}$ with

$$y \equiv g^k \pmod p .$$

Now y is a quadratic residue modulo p if and only if k is even. Also

$$y^{(p-1)/2} \equiv g^{k(p-1)/2} \equiv (-1)^k \pmod p .$$

This implies the assertion. □

Example 3.4.8. Let $p = 1237$ and $\Delta = 17$. We compute $\Delta^{(p-1)/2} \bmod p$ by fast exponentiation. We have $p - 1/2 = 618 = 2^9 + 2^6 + 2^5 + 2^3 + 2$. We compute the successive squares and find $17^2 \equiv 289 \pmod{1237}$, $17^{2^3} \equiv 243 \pmod{1237}$, $17^{2^5} \equiv 547 \pmod{1237}$, $17^{2^6} \equiv 1092 \pmod{1237}$, $17^{2^9} \equiv 256 \pmod{1237}$. Also $289 \cdot 243 \cdot 547 \cdot 1092 \cdot 256 \equiv 1 \pmod{1237}$. The computation requires 9 squarings and four multiplications modulo p. The result shows that 17 is a quadratic residue modulo 1237. But we do not know a square root of 17 modulo 1237 yet. From Proposition 3.4.5 we also know that $R^*(17, 1237) = 2$.

Evaluating the Legendre symbol by the Euler criterion requires $\mathrm{O}(\log p)$ multiplications in $\mathbb{Z}/p\mathbb{Z}$. If we use standard arithmetic, then the bit complexity of this algorithm is $\mathrm{O}((\log p)^3)$. Below we will see how the law of quadratic reciprocity can be used to evaluate the Legendre symbol even faster.

3.4.2 The law of quadratic reciprocity

We formulate the law of quadratic reciprocity. For the proof we refer to [IR82].

Theorem 3.4.9. *Let p and q be odd primes. Then the following are true:*

1. $\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$.
2. $\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8}$.
3. $\left(\frac{p}{q}\right) = (-1)^{(p-1)(q-1)/4}\left(\frac{q}{p}\right)$.

Theorem 3.4.9 implies that $\left(\frac{-1}{p}\right)$ and $\left(\frac{2}{p}\right)$ can be evaluated for any odd prime p by checking congruence conditions modulo 4 and 8, respectively.

Corollary 3.4.10. *Let p be an odd prime. Then the following are true.*

1. $\left(\frac{-1}{p}\right) = \begin{cases} 1 & \textit{if } p \equiv 1 \pmod 4, \\ -1 & \textit{if } p \equiv -1 \pmod 4. \end{cases}$
2. $\left(\frac{2}{p}\right) = \begin{cases} 1 & \textit{if } p \equiv \pm 1 \pmod 8, \\ -1 & \textit{if } p \equiv \pm 5 \pmod 8. \end{cases}$

Example 3.4.11. We have $\left(\frac{-1}{1237}\right) = 1$ since $1237 \equiv 1 \pmod 4$. We have $\left(\frac{2}{1237}\right) = -1$ since $1237 \equiv 5 \pmod 8$.

3.4.3 The Kronecker symbol

We introduce the Kronecker symbol. Its properties imply that $R(\Delta, p)$ only depends on the congruence class of p modulo $|\Delta|$. Quadratic reciprocity for the Kronecker symbol will also allow us to compute the Legendre symbol very efficiently. For the proofs of the results presented in this section see [IR82].

Definition 3.4.12. *Let m and n be integers. Then the* Kronecker symbol $\left(\frac{m}{n}\right)$ *is defined as follows.*

1. $\left(\frac{m}{0}\right) = \begin{cases} 1 & \textit{if } m = \pm 1, \\ 0 & \textit{otherwise.} \end{cases}$
2. $\left(\frac{m}{-1}\right) = \begin{cases} 1 & \textit{if } m \geq 0, \\ -1 & \textit{if } m < 0. \end{cases}$
3. If n is a prime number then, $\left(\frac{m}{n}\right)$ is the Legendre symbol.
4. If $n = (-1)^{e(-1)} \prod_{p|n} p^{e(p)}$ is the prime factorization of n, then

$$\left(\frac{m}{n}\right) = \left(\frac{m}{-1}\right)^{e(-1)} \prod_{p|n} \left(\frac{m}{p}\right)^{e(p)}.$$

In particular, $\left(\frac{m}{1}\right) = 1$ for all integers m.

We state important results concerning the Kronecker symbol. They can be proved using the definition of the Kronecker symbol and the law of quadratic reciprocity for the Legendre symbol.

Theorem 3.4.13. *Let m, m', n, n' be integers, n, n' odd and positive. Then the following are true.*

1. $\left(\frac{m}{n}\right) = 0$ *if and only if* $\gcd(m,n) > 1$.
2. *If* $m \equiv m' \pmod{n}$, *then* $\left(\frac{m}{n}\right) = \left(\frac{m'}{n}\right)$.
3. $\left(\frac{m}{n}\right)\left(\frac{m}{n'}\right) = \left(\frac{m}{nn'}\right)$.
4. $\left(\frac{m}{n}\right)\left(\frac{m'}{n}\right) = \left(\frac{mm'}{n}\right)$.
5. $\left(\frac{-1}{n}\right) = (-1)^{(n-1)/2}$.
6. $\left(\frac{2}{n}\right) = (-1)^{(n^2-1)/8}$.
7. *If m, n are odd and positive, then* $\left(\frac{m}{n}\right) = (-1)^{(n-1)(m-1)/4}\left(\frac{n}{m}\right)$.
8. *If* $m \neq 0$, $m \equiv 0, 1 \pmod{4}$ *and* $n \equiv n' \pmod{|m|}$, *then* $\left(\frac{m}{n}\right) = \left(\frac{m}{n'}\right)$.

Proof. The verification of these statements is deferred to Exercise 3.7.12.

Using the last assertion of Theorem 3.4.13 we can evaluate the Legendre symbol $\left(\frac{\Delta}{p}\right)$ for a fixed discriminant Δ and many primes p as follows. We compute the table of all values $\left(\frac{\Delta}{n}\right)$ for $0 \leq n < \Delta$. If we want to compute $\left(\frac{\Delta}{p}\right)$ for a prime number p, then we can find the value $\left(\frac{\Delta}{p}\right) = \left(\frac{\Delta}{p \bmod |\Delta|}\right)$ by a table look-up.

Example 3.4.14. For $\Delta = 5$ we find the values $\left(\frac{5}{0}\right) = 0$, $\left(\frac{5}{1}\right) = 1$, $\left(\frac{5}{2}\right) = -1$, $\left(\frac{5}{3}\right) = -1$, and $\left(\frac{5}{4}\right) = 1$. This implies that $\left(\frac{5}{45691}\right) = \left(\frac{5}{1}\right) = 1$, $\left(\frac{5}{45697}\right) = \left(\frac{5}{2}\right) = -1$, $\left(\frac{5}{45763}\right) = \left(\frac{5}{3}\right) = -1$, and $\left(\frac{5}{45779}\right) = \left(\frac{5}{4}\right) = 1$.

Theorem 3.4.13 can be used to evaluate the Kronecker symbol $\left(\frac{m}{n}\right)$ very efficiently. We explain how this is done. Let m and n be integers.

First, let $m = 0$. We have

$$\left(\frac{0}{n}\right) = \begin{cases} 1 & \text{if } n = \pm 1, \\ 0 & \text{otherwise.} \end{cases}$$

Now let $m \neq 0$. We show how Theorem 3.4.13 can be used to reduce the determination of $\left(\frac{m}{n}\right)$ to the computation of $\left(\frac{0}{n'}\right)$ for some n'. Write

$$n = (-1)^x 2^y n'$$

with an odd positive n' and $x, y \geq 0$. Then the definition of the Kronecker symbol tells us that

$$\left(\frac{m}{n}\right) = (-1)^{x + y(m^2-1)/8}\left(\frac{m}{n'}\right).$$

So it suffices to explain the computation $\left(\frac{m}{n}\right)$ with an odd and positive n. By Theorem 3.4.13, we may replace m by $m \bmod n$. Hence, we assume that

$$0 \leq m < n. \tag{3.13}$$

Now Theorem 3.4.13 is used to reduce the evaluation of the Kronecker symbol $\left(\frac{m}{n}\right)$ to the evaluation of a Kronecker symbol $\left(\frac{m'}{n'}\right)$ with $0 \leq m' < m$. If we iterate this reduction, then we will eventually have $m' = 0$ and we can evaluate the Kronecker symbol as described above. The reduction proceeds in two steps. We write

$$m = 2^y m_0\ ,$$

where $y \geq 0$ and m_0 is odd and positive. Then by Theorem 3.4.13 we have

$$\left(\frac{m}{n}\right) = (-1)^{y(n^2-1)/8}\left(\frac{m_0}{n}\right).$$

It remains to evaluate $\left(\frac{m_0}{n}\right)$. So assume that m is odd and positive. Theorem 3.4.13 yields

$$\left(\frac{m}{n}\right) = (-1)^{(m-1)(n-1)/4}\left(\frac{n \bmod m}{m}\right).$$

Hence, we replace m by $n \bmod m$ and m by n. Note that the new m is nonnegative and less than the old m. This concludes the description of the reduction.

The algorithm `kronecker` that we have just described is very similar to the Euclidean algorithm. The last value of n is the greatest common divisor of the initial values of m and n.

Example 3.4.15. We explain how `kronecker` calculates $\left(\frac{17}{3}\right)$.

$$\left(\frac{17}{3}\right) = \left(\frac{2}{3}\right) = -\left(\frac{1}{3}\right) = -\left(\frac{0}{1}\right) = -1.$$

Proposition 3.4.16. *Algorithm* `kronecker`(m, n) *takes time* $\mathrm{O}\big(\mathrm{size}(m)\,\mathrm{size}$ $(n)\big)$ *to compute* $\left(\frac{m}{n}\right)$.

Proof. Theorem 5.9.3 in [BS96]. □

From Proposition 3.4.16 we also obtain an estimate for the running time of Algorithm `numberOfPrimeForms`.

Corollary 3.4.17. *Algorithm* `numberOfPrimeForms` *has running time* $\mathrm{O}\big(\mathrm{size}(\Delta)\,\mathrm{size}(p)\big)$. □

Algorithm 3.4 `kronecker` (m, n)

Input: Integers m and n
Output: $\left(\frac{m}{n}\right)$

```
if $2 \mid n$ and $2 \mid m$ then return 0.
if $\operatorname{sign}(m) = \operatorname{sign}(n) = -1$ then $j \leftarrow -1$ else $j \leftarrow 1$.
$m \leftarrow |m|$, $n \leftarrow |n|$.
while $n$ is even do
  if $m \equiv 3, 5 \pmod 8$ then $j \leftarrow -j$
  $n \leftarrow n/2$
$m \leftarrow m \bmod n$
while $m \neq 0$ do
  while $m$ is even do
    if $n \equiv 3, 5 \pmod 8$ then $j \leftarrow -j$
    $m \leftarrow m/2$
  if $m \equiv 3 \pmod 4$ and $n \equiv 3 \pmod 4$ then $j \leftarrow -j$
  interchange $m$ and $n$
  $m \leftarrow m \bmod n$
if $n = 1$ then return $j$
else return 0
```

3.4.4 Computing square roots modulo *p*

In order to determine prime forms, algorithm `primeForm` must find a square root of a discriminant modulo a prime number. In this section we describe how this is done.

Here is a first simple case:

Proposition 3.4.18. *Let $p \equiv 3 \pmod 4$ and let r be a quadratic residue modulo p. Then $s = r^{(p+1)/4} \bmod p$ is a square root of r modulo p.*

Proof. Since r is a quadratic residue modulo p Theorem 3.4.7 implies

$$r^{(p-1)/2} \equiv 1 \pmod p .$$

It follows that

$$\left(r^{(p+1)/4}\right)^2 \equiv r^{(p-1)/2} r \equiv r \pmod p . \qquad \square$$

Example 3.4.19. Let $p = 2347$ and $\Delta = 17$. We have $p \equiv 3 \pmod 4$. Using the Euler criterion we verify that Δ is a quadratic residue modulo p. Now $17^{(2347+1)/4} \bmod 2347 = 799$. In fact, $799^2 \equiv 17 \pmod p$.

Using Proposition 3.4.18 we obtain the following complexity result.

Proposition 3.4.20. *Let p be a prime number with $p \equiv 3 \pmod 4$ and let $r \in \{0, \dots, p-1\}$ be a square modulo p. Then a square root of r modulo p can be computed in time $\mathrm{O}((\operatorname{size} p)^3)$.*

If we use this method for extracting square roots modulo p we obtain the following running time estimate for `primeForm`.

Corollary 3.4.21. *If* $p \equiv 3 \pmod 4$*, then* `primeForm` *runs in time* $\mathrm{O}\big(\operatorname{size}(p)\operatorname{size}(\Delta) + (\operatorname{size} p)^3\big)$.

Now we turn to the general case. We explain the algorithm of Tonelli for extracting square roots in a finite cyclic group G of known order $|G|$. If we apply this algorithm to the cyclic group $(\mathbb{Z}/p\mathbb{Z})^*$, then we obtain an algorithm for extracting square roots modulo p.

Let ρ be an element of G that is a square in G. We want to find a square root of ρ. We write

$$|G| = 2^t m\ , \quad m \text{ odd.} \tag{3.14}$$

The algorithm is based on the following result.

Lemma 3.4.22. *Assume that G has even order. Then half of the elements of G are nonsquares in G. If γ is such a nonsquare, then there is an even integer $e \in \big\{0, 1, \dots, 2^t - 1\big\}$ such that $\rho^m = \gamma^{me}$ and*

$$\big(\rho\gamma^{-e}\big)^{(m+1)/2}\gamma^{e/2} \tag{3.15}$$

is a square root of ρ.

Proof. In Exercise 3.7.16 it is shown that half of the elements in G are nonsquares.

Let γ be a nonsquare in G. The set

$$H = \Big\{\alpha^m : \alpha \in G\Big\}$$

is the subgroup of G of order 2^t. We claim that γ^m is not a square in H. Let $\gamma^m = \beta^2$ and let x and y be integers with $xm + y2^t = 1$. The integers x and y exist because m and 2^t are coprime. Then

$$\gamma = \gamma^{xm+y2^t} = \beta^{2x}\gamma^{y2^t} = \left(\beta^x\gamma^{y2^{t-1}}\right)^2 ,$$

a contradiction since γ was assumed to be a nonsquare in G. Exercise 3.7.17 implies that γ^m generates H. It follows that there is $e \in \{0, 1, \dots, 2^t - 1\}$ such that

$$\rho^m = (\gamma^m)^e\ .$$

Since ρ is a square in G it follows that ρ^m is a square in H. This implies that e is even.

Thus

$$(\big(\rho\gamma^{-e}\big)^{(m+1)/2}\gamma^{e/2})^2 = \rho.$$

□

Algorithm `sqrtModP` shown on page 48 uses (3.15) to calculate a square root of ρ. If G has odd order, then the algorithm sets $\gamma = 1$ and $e = 0$. Otherwise the algorithm determines a nonsquare γ in G. For this purpose, the algorithm selects a random element in G and tests whether γ is a square or not. The chance of finding a non square is $1/2$, since the number of squares in G is $|G|/2$. The square test uses, for example, a generalization of Euler's criterion (see Exercise 3.7.16). If $G = (\mathbb{Z}/p\mathbb{Z})^*$ for some prime number p, then algorithm `kronecker` can also be used as a square test.

We explain how Algorithm `sqrtModP` finds e. First, we determine

$$\rho_m = \rho^m \text{ and } \gamma_m = \gamma^m \ .$$

The exponent e is the discrete logarithm of ρ_m to the base γ_m. It can be computed by the Pohlig-Hellman algorithm. We explain how this algorithm works. Since $e \in \{0, 1, \ldots, 2^t - 1\}$ we can write

$$e = b_0 + 2b_1 + 2^2 b_2 + \cdots + 2^{t-1} b_{t-1} \ , \quad b_i \in \{0,1\}, 1 \leq i < t \ .$$

Since e is even we have $b_0 = 0$. We compute the coefficients b_i, $1 \leq i < t$, iteratively. Let $1 \leq i < t$ and assume that we know

$$e_i = b_0 + b_1 \cdot 2 + \cdots + b_{i-1} \cdot 2^{i-1} \ .$$

and

$$\alpha_i = \rho_m \gamma_m^{-e_i} \ .$$

Then

$$\alpha_i = \gamma_m^{b-i2^i + b_{i+1}2^{i+1} + \ldots + b_{t-1}2^{t-1}} \ .$$

Since the order of H is 2^t, we have

$$\alpha_i^{2^{t-i-1}} = \gamma_m^{b_i 2^{t-1}} \ .$$

Therefore, $b_i = 0$ if and only if $\alpha_i^{2^{t-i-1}} = 1$.

Example 3.4.23. We use the algorithm of Tonelli to compute a square root of $a = 2$ modulo $p = 17$. A quadratic nonresidue modulo 17 is $c = 3$. Hence we have the following setup: $m = 1$, $t = 4$, $r_m = r = 2$, $c_m = c^{-1} \bmod 17 = 6$. We now obtain the following intermediate results where each residue classes is represented by an element in that class.

i	1	2	3
r_m	2	4	16
c_m	2	4	16
$r_m^{2^{t-i-1}}$	-1	-1	16
e	2	6	14

Algorithm 3.5 `sqrtModP` (r, p)

Input: A prime p and a square r modulo p
Output: A square root s of r modulo p or `nil`

```
$m \leftarrow p-1$, $t \leftarrow 0$
while $m$ is even do
  $t \leftarrow t+1$
  $m \leftarrow m/2$
Choose a random element $c \in \{1,\dots,p-1\}$
if $\left(\frac{c}{p}\right) = 1$ then return nil
$r_m \leftarrow r^m \bmod p$
$c_m \leftarrow c^{-m} \bmod p$
$e \leftarrow 0$, $i \leftarrow 0$
while $r_m \neq 1$ do
  $i \leftarrow i+1$
  $c_m \leftarrow c_m^2$
  if $r_m^{2^{t-i-1}} \bmod p \neq 1$ then
    $e \leftarrow e + 2^i$
    $r_m \leftarrow r_m c_m \bmod p$
$a \leftarrow rc^{-e} \bmod p$
return $c^{e/2} a^{(m+1)/2} \bmod p$
```

We obtain $a = 1$. The square root is $3^7 \bmod 17 = 11$. In fact, $11^2 \equiv 2 \pmod{17}$.

Proposition 3.4.24. *The algorithm* `sqrtModP` *has the following properties. It fails with probability* $1/2$*. If it does not fail, then it returns a square root of* $r \bmod p$ *provided that* r *is a square mod* p*. Its running time is* $\mathrm{O}((\log p)^4)$*.*

Proof. See [BS96].

No deterministic polynomial time algorithm for extraction square roots mod p is known. Therefore, we formulate the following problem.

Problem 3.4.25. Find a deterministic polynomial time algorithm for extracting square roots modulo p.

It should be noted, however, that René Schoof gave an algorithm using the arithmetic of elliptic curves which does extract square roots of a *fixed* x modulo varying prime p in time $O(\log^9 p)$, cf. [Sch85].

If we use `sqrtModP`(Δ, p) to extract the square root in `primeForm`, then we obtain the following result.

Corollary 3.4.26.
Algorithm `primeForm` *has success probability* $1/2$ *and running time* $\mathrm{O}\big(\mathrm{size}(\Delta)\,\mathrm{size}(p) + (\mathrm{size}(p))^4\big)$*.* □

3.5 The case of a prime power

In this section we determine $\mathcal{F}(\Delta, p^e)$ and $\mathcal{F}^*(\Delta, p^e)$ for prime p and $e > 1$.

Proposition 3.5.1. *Let $e \in \mathbb{N}$, $e \geq 1$.*

1. *If $\left(\frac{\Delta}{p}\right) = -1$, then $R(\Delta, p^e) = R^*(\Delta, p^e) = 0$.*
2. *If $\left(\frac{\Delta}{p}\right) = 0$, $e \geq 2$, and p does not divide the conductor of Δ, then $R(\Delta, p^e) = R^*(\Delta, p^e) = 0$.*
3. *If $\left(\frac{\Delta}{p}\right) = 1$, then $R(\Delta, p^e) = R^*(\Delta, p^e) = 2$. Also there is exactly one square root $b(\Delta, p^e)$ of Δ mod $4p^e$ with $b(\Delta, p^e) \equiv b(\Delta, p^{e-1}) \pmod{2p^{e-1}}$ and $-p^e < b(\Delta, p^e) < p^e$. The square roots of Δ mod $4p^e$ in $\{-p^e+1, \ldots, p^e-1\}$ are $\pm b(\Delta, p^e)$. They are not divisible by p.*

Proof. If $\left(\frac{\Delta}{p}\right) = -1$, then Δ is not a square mod p which implies that Δ is not a square mod $4p^e$. Proposition 3.1.1 implies that $R(\Delta, p^e) = R^*(\Delta, p^e) = 0$.

Let $\left(\frac{\Delta}{p}\right) = 0$. Then p divides Δ. Suppose that $\mathcal{F}(\Delta, p^e)$ is not empty, Δ is a square mod $4p^e$ and that b is an integer with $b^2 \equiv \Delta \pmod{4p^e}$. Then p divides b, hence p^2 divides Δ since $e \geq 2$. This implies $\Delta/p^2 \equiv (b/p)^2 \pmod{4p^{e-2}}$. This shows that Δ/p^2 is a discriminant. Hence p must divide the conductor of Δ.

Let $\left(\frac{\Delta}{p}\right) = 1$. We use induction on e. By Proposition 3.4.5 the assertion is true for $e = 1$. Assume that $e > 1$, and that the assertion holds for $e - 1$.

For any square root b of Δ mod $4p^{e-1}$, we construct a square root b_1 of Δ mod $4p^e$ with $b_1 \equiv b \pmod{2p^{e-1}}$ and $|b_1| < p^e$.

Write

$$b_1 = b + 2kp^{e-1}$$

with some integer k satisfying $|k| \leq (p-1)/2$. We must choose k such that $b_1^2 \equiv \Delta \pmod{4p^e}$. Now

$$b_1^2 = (b + 2kp^{e-1})^2 = b^2 + 4bkp^{e-1} + 4k^2p^{2e-2} .$$

Since $e \geq 2$ we have $2e - 2 \geq e$. Hence

$$b_1^2 \equiv b^2 + 4bkp^{e-1} \pmod{4p^e} .$$

This shows that $b_1^2 \equiv \Delta \pmod{4p^e}$ if and only if

$$k \equiv b^{-1}\frac{\Delta - b^2}{4p^{e-1}} \pmod{p} ,$$

where b^{-1} is the inverse of b mod p. It follows from the construction of k that b_1 is unique modulo $2p^e$.

By assumption, the numbers $\pm b(\Delta, p^{e-1})$ are the only two square roots of Δ mod $4p^{e-1}$ in the interval $\{-p^{e-1}+1, \ldots, p^{e-1}-1\}$.

In the prior construction, we first set $b = b(\Delta, p^{e-1})$. The resulting square root b_1 of $\Delta \bmod 4p^e$ satisfies $|b_1| \leq p^e - 1$. We call it $b(\Delta, p^e)$. If we set $b = -b(\Delta, p^{e-1})$ we obtain $b_1 = -b(\Delta, p^e)$.

Any square root of $\Delta \bmod 4p^e$ must also be a square root of $\Delta \bmod 4p^{e-1}$, and hence be congruent to $\pm b(\Delta, p^{e-1}) \bmod 2p^{e-1}$, by assumption. It follows that $\pm b(\Delta, p^e)$ are the only square roots of $\Delta \bmod 4p^e$ in $\{-p^e+1, \ldots, p^e-1\}$. The claim of the induction is shown. □

From the proof of Lemma 3.5.1 we obtain the part of algorithm `sqrtMod4PE` on page 52 for computing $b(\Delta, p^e)$ which deals with discriminants Δ prime to p. We illustrate this algorithm in the next example.

Example 3.5.2. Let $p = 5$ and $\Delta = 21$. Then $\left(\frac{\Delta}{p}\right) = 1$ since $21 \equiv 1 \equiv (\pm 1)^2 \pmod 5$. Using the construction in the proof of Lemma 3.5.1 we obtain $b(21, 25) = 11$.

Assume that $e \geq 2$ and p divides $f(\Delta)$. This case is still open. It will be discussed now. Let

$$n = \max\{k \in \mathbb{N} : p^k \mid f(\Delta), 2k \leq e\}.$$

Then $n \geq 1$. Let

$$\Delta_0 = \Delta / p^{2n}, \ e_0 = e - 2n.$$

We reduce the problem of finding $R(\Delta, p^e)$ to the task of determining $R(\Delta_0, p^{e_0})$. As we will explain now, this value has already been determined.

If $e_0 = 0$, then $R(\Delta_0, p^{e_0}) = R(\Delta_0, 1) = 1$. If $e_0 = 1$ then $R(\Delta_0, p^{e_0}) = R(\Delta_0, p) = 1 + \left(\frac{\Delta_0}{p}\right)$ by (3.12). If $e_0 > 1$, then n equals the maximal exponent k such that $p^k \mid f(\Delta)$, and, hence, p does not divide the conductor of Δ_0. Hence, by Proposition 3.5.1 we have $R(\Delta_0, p^{e_0}) = 0$ for $\left(\frac{\Delta_0}{p}\right) \in \{0, -1\}$ and $R(\Delta_0, p^{e_0}) = 2$ otherwise.

Proposition 3.5.3. *Assume that p divides the conductor of Δ and let n, e_0, Δ_0 be as described above. Then we have $R(\Delta, p^e) = p^n R(\Delta_0, p^{e_0})$ and*

$$R^*(\Delta, p^e) = R^*(\Delta_0, p^{e_0}) \cdot \begin{cases} p^{n-1}\left(p - \left(\frac{\Delta_0}{p}\right) - 1\right) & \text{if } e_0 = 0, \\ p^{n-1}(p-1) & \text{if } e_0 > 0 \text{ and } p \nmid \Delta, \\ p^n & \text{if } e_0 > 0 \text{ and } p \mid \Delta. \end{cases}$$

Proof. For any form $f = (p^e, b, c)$ with discriminant Δ, we know that p^{2n} divides $\Delta - 4p^e c = b^2$. Thus we have a map

$$\psi : \mathcal{F}(\Delta, p^e) \longrightarrow \mathcal{F}(\Delta_0, p^{e_0}) : (p^e, p^n b, c)\Gamma \longmapsto (p^{e_0}, b, c)\Gamma \ .$$

If $f_i = (p^e, p^n b_i, c_i)$ for $i = 1, 2$, then $\psi(f_1\Gamma) = \psi(f_2\Gamma)$ if and only if

$$b_1 = b_2 + 2kp^{e_0} \ .$$

The forms f_1 and f_2 are Γ-equivalent if and only if $p^n \mid k$. Hence, the map ψ is p^n-to-1. This proves the first assertion.

A form $f = (p^e, p^n b, c)$ is primitive if and only if $p \nmid c$. Let $f' = (p^{e_0}, b', c')$ be a form in $\psi(f\Gamma)$. Then $b = b' + 2kp^{e_0}$ and

$$c = \frac{p^{2n}b^2 - \Delta}{4p^e} = \frac{b^2 - \Delta_0}{4p^{e_0}} = \frac{(b' + 2kp^{e_0})^2 - \Delta_0}{4p^{e_0}}\ . \qquad (*)$$

We determine for given b' the number of k with $0 \le k < p^n$ for which $p \nmid c$.

We begin with the case that $e_0 = 0$. Then $p \,|\, c$ if and only if Δ_0 is a square modulo p (or modulo 8, if $p = 2$) and

$$k \equiv \frac{\pm b(\Delta_0, p) - b(\Delta_0, 1)}{2} \quad (\text{mod } p) \quad [\text{ mod2 if } p = 2.]$$

Thus there are $((\frac{\Delta_0}{p}) + 1)p^{n-1}$ such values of k. Using $R(\Delta_0, 1) = R^*(\Delta_0, 1)$, we obtain the first case of the second assertion of the proposition.

If $e_0 \ge 1$, we can simplify $(*)$. We get that $p \mid c$ if and only if

$$p \mid (c' + kb'). \qquad (**)$$

We see that f is imprimitive if $\psi(f\Gamma)$ is imprimitive. Assume $\psi(f\Gamma)$ is primitive. If $p \nmid \Delta_0$, then also $p \nmid b'$, and hence there are p^{n-1} values of k for which $(**)$ is satisfied. If, however, $p \mid \Delta_0$, then $p \mid b'$ and $p \nmid c'$ so that f is primitive for any value of k. □

With the preceding proposition and its proof we are ready to give an algorithm which computes a form of the type (p^e, b, c) for all discriminants Δ for which it exists. Note that it follows from Proposition 3.5.3 and from Proposition 3.5.5 that for any prime p and any positive integer e the values $R(\Delta, p^e)$ and $R^*(\Delta, p^e)$ can be computed in polynomial time.

Example 3.5.4. Let $\Delta = 117 = 3^2 \cdot 13$, $p = 3$, and $e = 2$. Then $f(\Delta) = 3$, $n = 1$, $\Delta_0 = 13$, $e_0 = 0$, and $R(117, 9) = 3R(13, 1) = 3$. The square roots of 117 mod $4 * 13$ are 3, $3 * 3 = 9$, $3 * 5 = 15$. Hence, the Γ-orbits of forms $(9, b, c)$ of discriminant 117 are $(9, 3, -3)\Gamma$, $(9, 9, -1)\Gamma$, and $(9, 15, 3)\Gamma$. Also, $R^*(117, 9) = 3R(13, 1) - R(13, 3) = 1$. In fact, $(9, 9, -1)\Gamma$ is the only Γ-orbit of primitive forms $(9, b, c)$ of discriminant 117.

If $R(\Delta, p^e) > 0$, then we set

$$c(\Delta, p^e) = \frac{b(\Delta, p^e)^2 - \Delta}{4p^e}.$$

Then it follows from Proposition 3.5.1 that

$$\mathcal{F}(\Delta, p^e) = \mathcal{F}^*(\Delta, p^e) = \{\big(p^e, \pm(b(\Delta, p^e), c(\Delta, p^e)\big)\Gamma\}$$

provided $(\frac{\Delta}{p}) = 1$.

Algorithm `primePowerForm` on page 52 computes the form $\big(p^e, b(\Delta, p^e), c(\Delta, p^e)\big)$.

Algorithm 3.6 `sqrtMod4PE` $(\Delta,\ p\ ,e)$

Input: A discriminant Δ, a prime p with $R(\Delta, p) > 0$, and $e \geq 1$
Output: $b(\Delta, p^e)$

```
if $p \mid \Delta$ then
  for $(n \leftarrow 1,\ 2n+2 \leq e$ and $p^{2n+2} \mid \Delta,\ n \leftarrow n+1)$
  if $p = 2$ and $2^{-2n}\Delta \not\equiv 0, 1 \mod 4$ then
    $n \leftarrow n - 2$
  $\Delta_0 \leftarrow \Delta/p^{2n}$, $e_0 \leftarrow e - 2n$
  if $e_0 = 0$ then
    return $p^n(\Delta_0 \bmod 2)$
    else return $p^n$sqrtMod4PE$(\Delta_0, p^{e_0})$
else
  $b \leftarrow$ sqrtMod4P$(\Delta, p)$
  $f \leftarrow 1$
  while $f < e$ do
    $k \leftarrow b^{-1}\frac{\Delta - b^2}{4p^f} \bmod p$
    $b \leftarrow b + 2kp^f \bmod 2p^{f+1}$
    $f \leftarrow f + 1$
  if $b > p^e$ then
    return $b - 2p^e$
    else return $b$
```

Algorithm 3.7 `primePowerForm` $(\Delta,\ p,\ e)$

Input: A discriminant Δ, a prime p with $R^*(\Delta, p) > 0$, and $e \geq 1$
Output: The form $\big(p^e, b(\Delta, p^e), c(\Delta, p^e)\big)$

```
$b \leftarrow$ sqrtMod4PE$(\Delta, p)$
$c \leftarrow \frac{b^2 - \Delta}{4p^e}$
return $\big(p^e, b, c\big)$
```

Proposition 3.5.5. *Algorithms* `sqrtMod4PE`(Δ, p, e) *and* `primePowerForm`(Δ, p, e) *have success probability* $1/2$ *and running time* $\mathrm{O}\big(\mathrm{size}(\Delta)\,\mathrm{size}(p) + (\mathrm{size}(p))^4 + e^2(\mathrm{size}(p))^2\big)$.

Proof. We analyze Algorithm `sqrtMod4PE`. By Corollary 3.4.26, finding the initial form (p, b, c) has success probability $1/2$ and takes time $\mathrm{O}\big(\mathrm{size}(\Delta)\,\mathrm{size}(p) + \mathrm{size}(p)^4\big)$. Each iteration requires time $\mathrm{O}\big(e(\mathrm{size}(p))^2\big)$. The number of iterations is $e - 1$. This proves the assertion. □

An algorithm for enumerating all elements of $\mathcal{F}(\Delta, p^e)$ in the case that $p \mid f(\Delta)$ can be derived from Algorithm 3.6 and the proof of Proposition 3.5.3. This is left to the reader as exercise 3.7.21.

3.6 The case of a composite integer

We now determine $R(\Delta, a)$ and $R^*(\Delta, a)$ for arbitrary positive integers a. We first prove that the functions $R(\Delta, \cdot)$ and $R^*(\Delta, \cdot)$ are multiplicative.

Lemma 3.6.1. *If a_1 and a_2 are coprime positive integers, then $R(\Delta, a_1a_2) = R(\Delta, a_1)R(\Delta, a_2)$ and $R^*(\Delta, a_1a_2) = R^*(\Delta, a_1)R^*(\Delta, a_2)$.*

Proof. For an integer a let

$$S(\Delta, a) = \{b + 2a\mathbb{Z} : b^2 \equiv \Delta \pmod{4a}\} .$$

The Chinese remainder theorem implies that the map

$$\begin{aligned} S(\Delta, a_1a_2) &\to S(\Delta, a_1) \times S(\Delta, a_2) \\ b + 2a_1a_2\mathbb{Z} &\mapsto (b + 2a_1\mathbb{Z}, b + 2a_2\mathbb{Z}) \end{aligned}$$

is a bijection. Proposition 3.1.1 implies that $R(\Delta, a_1a_2) = R(\Delta, a_1)R(\Delta, a_2)$.

We show that $R^*(\Delta, \cdot)$ is also a multiplicative function. Let b be a square root of Δ mod $4a_1a_2$. Let $c = (b^2 - \Delta)/(4a_1a_2)$, $c_i = (b^2 - \Delta)/(4a_i)$, $i = 1, 2$. Since a_1 and a_2 are coprime, it follows that the form (a_1a_2, b, c) is primitive if and only if both forms (a_1, b, c_1) and (a_2, b, c_2) are primitive. This proves our assertion. □

If a is a positive integer with prime factorization

$$a = \prod_{p|a} p^{e(p)}, \tag{3.17}$$

then by Lemma 3.6.1 we have

$$R(\Delta, a) = \prod_{p|a} R\big(\Delta, p^{e(p)}\big) \tag{3.18}$$

and

$$R^*(\Delta, a) = \prod_{p|a} R^*\big(\Delta, p^{e(p)}\big) . \tag{3.19}$$

Since for each prime factor p of a the values $R(\Delta, p^{e(p)})$ and $R^*(\Delta, p^{e(p)})$ can be computed in polynomial time, the values $R(\Delta, a)$ and $R^*(\Delta, a)$ can be computed in polynomial time, provided that the prime factorization of a is known.

Example 3.6.2. Let $\Delta = 540$ and $a = 126$. The prime factorization of a is

$$a = 2 \cdot 3^2 \cdot 7 .$$

The conductor of Δ is 3. So Proposition 3.4.5 implies $R(\Delta, 2) = R^*(\Delta, 2) = 1$. Proposition 3.5.3 implies $R^*(540, 9) = 2$. Finally, since $\left(\frac{540}{7}\right) = 1$, Proposition 3.4.5 implies $R^*(540, 7) = 2$. So $R^*(540, 126) = 4$.

In general, the computation of the $\mathcal{F}(\Delta, a)$ and $\mathcal{F}^*(\Delta, a)$ takes exponential time since in the worst case, those sets have exponential cardinality. However, if the number of prime factors of a is fixed, then those sets can be computed in probabilistic polynomial time.

Example 3.6.3. Let $a = 3^2 \cdot 5^3$ and $\Delta = 61$. Then $\Delta \equiv 1 \pmod 3$ and $\Delta \equiv 1 \pmod 5$. Hence, $\left(\frac{\Delta}{3}\right) = \left(\frac{\Delta}{5}\right) = 1$. So $R(\Delta, a) = R^*(\Delta, a) = R^*(\Delta, 3^2)R^*(\Delta, 5^3) = 2 \cdot 2 = 4$.

We first construct the forms $(3^2, b, c)$ of discriminant 61. We start from the form $(3, 1, -5)$. Using the method from the proof of Proposition 3.5.1, we obtain the form $(3^2, 13, 3)$. Hence there are two Γ-orbits of forms $(3^2, b, c)$ of discriminant 61, namely $(3^2, \pm 13, 3)\Gamma$.

We construct the forms $(5^3, b, c)$ of discriminant 61. Again, we start from the form $(5, 1, -6)$. From this form we obtain the form $(5^2, 31, 18)$. From this form we obtain the form $(5^3, 81, 26)$. Hence there are two Γ-orbits of forms $(5^3, b, c)$ of discriminant 61, namely $(5^3, \pm 81, 26)\Gamma$.

Now we apply the Chinese remainder theorem. We solve

$$b \equiv 13 \pmod{3^2}, \quad b \equiv 81 \pmod{5^3}, \quad b \equiv 1 \pmod 2 .$$

We obtain the form $(3^2 \cdot 5^3, 1831, 745)$. Next, we solve

$$b \equiv -13 \pmod{3^2}, \quad b \equiv 81 \pmod{5^3}, \quad b \equiv 1 \pmod 2 .$$

We obtain the form $(3^2 \cdot 5^3, 581, 75)$. Hence

$$\mathcal{F}(61, 3^2 \cdot 5^3) = \{(3^2 \cdot 5^3, \pm 581, 75)\Gamma, (3^2 \cdot 5^3, \pm 1831, 745)\Gamma\}.$$

3.7 Exercises

Exercise 3.7.1. Find an algorithm that computes $\mathcal{F}(\Delta, 0)$.

Exercise 3.7.2. Determine $\mathcal{F}(\Delta, 5)$ for all discriminants Δ with $|\Delta| \leq 10$.

Exercise 3.7.3. Show that all integral forms (a, b, c) whose discriminant is a fundamental discriminant are primitive.

Exercise 3.7.4. Prove Proposition 3.3.4.

Exercise 3.7.5. Let p be a prime number with $\left(\frac{\Delta}{p}\right) \geq 0$. Prove that the form $(p, b(\Delta, p), c(\Delta, p))$ is primitive if and only if p does not divide the conductor of Δ.

Exercise 3.7.6. Determine all primitive representations of 3 by a form of discriminant -3 by the same method that was used in Example 3.1.3. .

Exercise 3.7.7. Let $p = 2437$. Use the Euler criterion to compute $R^*(5, p)$ and $R^*(28, p)$.

Exercise 3.7.8. Find $\mathcal{F}^*(21, 125)$.

Exercise 3.7.9. Determine $R^*(588, 49)$, $R^*(392, 3)$, $\mathcal{F}^*(588, 49)$, and $\mathcal{F}^*(392, 3)$.

Exercise 3.7.10. Write a program that computes $R^*(\Delta, a)$ for any quadratic discriminant Δ and any integer a.

Exercise 3.7.11. Determine $R^*(540, 132)$ and $\mathcal{F}^*(540, 132)$.

Exercise 3.7.12. Prove Theorem 3.4.13.

Exercise 3.7.13. Prove that $\left(\frac{m}{n}\right) = 0$ if and only if $\gcd(m, n) \neq 1$.

Exercise 3.7.14. Determine $\left(\frac{133}{257}\right)$ and $\left(\frac{128}{228}\right)$.

Exercise 3.7.15. Let p be an odd prime, $p \equiv 3 \bmod 8$, and r a quadratic residue modulo p.

1. Prove that $r^{(p-1)/4} \equiv \pm 1 \pmod{p}$.
2. Let $r^{(p-1)/4} \equiv 1 \pmod{p}$. Show that $r^{(p+3)/8} \bmod p$ is a square root of r modulo p.
3. Let $r^{(p-1)/4} \equiv -1 \pmod{p}$. Show that $2^{-1}(4r)^{(p+3)/8} \bmod p$ is a square root of r modulo p where 2^{-1} denotes the inverse of 2 modulo p.

Exercise 3.7.16. Let G be a finite cyclic group of even order.

1. Show that the number of squares in G is $|G|/2$.
2. Show that $\gamma \in G$ is a square in G if and only if $\gamma^{|G|/2} = 1$.

Exercise 3.7.17. Let G be a finite cyclic group of order 2^t for some positive integer t. Prove that any non square in G generates G.

Exercise 3.7.18. Let G be a group of odd order. Prove that any element of G is a square in G. Explain, how a square root of an element of G can be found.

Exercise 3.7.19. Use the algorithm of Tonelli to compute a square root of 13 modulo 17.

Exercise 3.7.20. Implement the algorithm of Tonelli. Use the implementation to compute a square root of 2 modulo 12329.

Exercise 3.7.21. Let p be prime, e an integer larger than 1, and Δ a discriminant for which $p^e \mid f(\Delta)$. Use the proof of Proposition 3.5.3 to develop algorithms that enumerate all elements of $\mathcal{F}(\Delta, p^e)$ and $\mathcal{F}^*(\Delta, p^e)$.

Chapter references and further reading

[BS96] Eric Bach and Jeffrey Shallit, *Algorithmic number theory*, MIT Press, Cambridge, Massachusetts and London, England, 1996.

[Buc04] Johannes Buchmann, *Introduction to cryptography*, second ed., Springer-Verlag, 2004, Undergradute Texts in Mathematics.

[IR82] Kenneth Ireland and Michael Rosen, *A classical introduction to modern number theory*, Springer-Verlag, New York, 1982.

[Sch85] René Schoof, *Elliptic curves over finite fields and the computation of square roots mod p*, Mathematics of Computation **44** (1985), 483–494 (English).

4

Forms, Bases, Points, and Lattices

In this chapter we explain the correspondence between binary quadratic forms with real coefficients and points, $\mathbb{R}$-bases, and lattices in the real plane. This correspondence will enable us to use quadratic number fields and the geometry of numbers in the theory of forms.

4.1 Two-dimensional commutative $\mathbb{R}$-algebras

4.1.1 Definition

We introduce commutative $\mathbb{R}$-algebras.

Definition 4.1.1. *A commutative $\mathbb{R}$-algebra is a commutative ring A with unit element which is an $\mathbb{R}$-vector space and which satisfies*

$$(r\alpha + s\beta)\gamma = r(\alpha\gamma) + s(\beta\gamma)$$

for all $r, s \in \mathbb{R}$ and $\alpha, \beta, \gamma \in A$. The dimension *of a commutative $\mathbb{R}$-algebra is its dimension as an $\mathbb{R}$-vector space.*

Example 4.1.2. The field $\mathbb{C}$ of complex numbers is a two-dimensional commutative $\mathbb{R}$-algebra.

If A is a commutative $\mathbb{R}$-algebra with unit element 1, then the map

$$\mathbb{R} \to A\,, \quad r \mapsto r \cdot 1 \tag{4.1}$$

is an injective field-homomorphism. That map is used to embed $\mathbb{R}$ into A. The real number r is identified with the element $r \cdot 1$ of A.

We define isomorphisms between commutative $\mathbb{R}$-algebras.

Definition 4.1.3. *Let A and A' be commutative $\mathbb{R}$-algebras. A map $\varphi : A \to A'$ is called a* homomorphism *between A and A', if*

$$\varphi\big(r(\alpha\beta + \gamma)\big) = r\big(\varphi(\alpha)\varphi(\beta) + \varphi(\gamma)\big)$$

for all $r \in \mathbb{R}$ and all $\alpha, \beta, \gamma \in A$ and if φ maps the unit element of A to the unit element of A'. An isomorphism *between A and A' is a bijective homomorphism between A and A'. An* automorphism *of A is an isomorphism between A and A.*

Example 4.1.4. The map that sends a complex number to its complex conjugate is an automorphism of the commutative $\mathbb{R}$-algebra $\mathbb{C}$.

We determine all two-dimensional commutative $\mathbb{R}$-algebras up to isomorphism.

Lemma 4.1.5. *Let A be a two-dimensional commutative $\mathbb{R}$-algebra. Then exactly one of the following three statements holds.*

1. *There is an $\mathbb{R}$-basis $(1, i)$ of A with $i^2 = 1$.*
2. *There is an $\mathbb{R}$-basis $(1, i)$ of A with $i^2 = 0$.*
3. *There is an $\mathbb{R}$-basis $(1, i)$ of A with $i^2 = -1$.*

Proof. There is an $\mathbb{R}$-basis $(1, \alpha)$ of A and we have $\alpha^2 = x + y\alpha$ with $x, y \in \mathbb{R}$. This implies $\big(\alpha - (y/2)\big)^2 = x + y^2/4$. Set

$$i = \begin{cases} \frac{\alpha - y/2}{\sqrt{|x+y^2/4|}} & \text{if } x + y^2/4 \neq 0, \\ \alpha - y/2 & \text{otherwise.} \end{cases}$$

Then $(1, i)$ is still an $\mathbb{R}$-basis of A and $i^2 \in \{0, \pm 1\}$ as asserted.

Assume that there is an $\mathbb{R}$-basis $(1, \alpha)$ of A with $\alpha^2 = 1$ and another $\mathbb{R}$-basis $(1, \beta)$ of A with $\beta^2 = -1$. Then we can write $\alpha = x + y\beta$ with $x, y \in \mathbb{R}$. Hence,

$$1 = \alpha^2 = x^2 - y^2 + 2xy\beta. \tag{4.2}$$

It follows that $xy = 0$. Since α and 1 are linearly independent, this implies $x = 0$. By (4.2) we have $1 = -y^2$, a contradiction. In a similar way, it can be shown that A cannot have two $\mathbb{R}$-bases $(1, \alpha)$ and $(1, \beta)$ with $\alpha^2 = 1$ and $\beta^2 = 0$ or $\alpha^2 = -1$ and $\beta^2 = 0$. □

We will now explicitly construct commutative $\mathbb{R}$-algebras that have bases as in Lemma 4.1.5.

The commutative $\mathbb{R}$-algebra $\mathbb{C}$ contains a square root of -1. We fix such a square root and denote it by $i(-1) = \sqrt{-1}$. Instead of $\mathbb{C}$ we also write A_{-1}. We note that that $(1, \sqrt{-1})$ is an $\mathbb{R}$-basis of $\mathbb{C}$. So any complex number α can in a unique way be written as $\alpha = x + y\sqrt{-1}$ with $x, y \in \mathbb{R}$. The coefficient x

is called the *real part* of α. It is denoted by $\Re(\alpha) = \Re\alpha$. The coefficient y is called the *imaginary part* of α. It is denoted by $\Im(\alpha) = \Im\alpha$. So we can write

$$\alpha = \Re\alpha + \Im\alpha\sqrt{-1} \, .$$

We describe a two-dimensional commutative $\mathbb{R}$-algebra that has an $\mathbb{R}$-basis $\big(1, i(1)\big)$ where $i(1)^2 = 1$. For $\alpha = (\alpha_1, \alpha_2)$ and $\beta = (\beta_1, \beta_2) \in \mathbb{R}^2$ we define the component-wise product

$$\alpha\beta = \alpha \cdot \beta = (\alpha_1\beta_1, \alpha_2\beta_2) \, . \tag{4.3}$$

With this multiplication, $\mathbb{R}^2$ is a two-dimensional commutative $\mathbb{R}$-algebra with unit element $(1, 1)$. We set

$$i(1) = (1, -1) \, . \tag{4.4}$$

Then $i(1)^2 = 1$. We denote that commutative $\mathbb{R}$-algebra by A_1.

Example 4.1.6. We determine all square roots of 1 in A_1. If $\alpha = (\alpha_1, \alpha_2) \in A_1$ with $\alpha^2 = 1$ then $\alpha_1^2 = 1$ and $\alpha_2^2 = 1$. Hence, $1 = (1, 1)$, $-1 = (-1, -1)$, $(1, -1)$, and $(-1, 1)$ are the square roots of 1 in A_1. Two of those square roots, 1 and -1, are real numbers. The other two, $(1, -1)$ and $(-1, 1)$ are not.

The construction of the two-dimensional commutative $\mathbb{R}$-algebra A_0 that contains an $\mathbb{R}$-basis $(1, i(0))$ with

$$i(0)^2 = 0$$

is left to the reader as Exercise 4.7.1.

Theorem 4.1.7. *Up to isomorphism, A_1, A_0, and A_{-1} are the only two-dimensional commutative $\mathbb{R}$-algebras. They are pairwise non-isomorphic.*

Proof. It follows from Lemma 4.1.5 that the commutative $\mathbb{R}$-algebras A_1, A_{-1}, and A_0 are pairwise non-isomorphic.

We show that up to isomorphism there are no other two-dimensional commutative $\mathbb{R}$-algebras. Let A be a such an $\mathbb{R}$-algebra. By Lemma 4.1.5 there is an element $i \in A$ such that $i^2 \in \{0, 1, -1\}$. Then $(1, i)$ is an $\mathbb{R}$-basis of A. The map

$$A \to A_{i^2} \, , \quad x + iy \mapsto x + yi(i^2), \; x, y \in \mathbb{R}$$

is an isomorphism of commutative $\mathbb{R}$-algebras. □

4.1.2 Notation

We will see, that positive definite forms are oriented norm forms of A_{-1} and that indefinite forms are oriented norm forms of A_1. Since we are mainly concerned with positive definite and indefinite forms, we will concentrate on A_1 and A_{-1}. We let $j \in \{\pm 1\}$, $i = i(j)$, and $A = A_j$. Recall that by virtue of (4.1) we have $\mathbb{R} \subset A$.

4.1.3 Geometry of multiplication

We give a geometric interpretation of the multiplications in $\mathbb{C}$ and in A_1.

Clearly, multiplication by $\alpha = (\alpha_1, \alpha_2) \in A_1$ is a scaling by the factor α_1 in one direction and by the factor α_2 in the other direction.

Next we consider multiplication in $\mathbb{C}$. Any $\alpha \in \mathbb{C}$ can be written as

$$\alpha = |\alpha|(\cos\varphi + \sqrt{-1}\sin\varphi) \tag{4.5}$$

where $|\alpha|$ is the absolute value of α and φ is a real number that is uniquely determined mod 2π. The smallest such non-negative φ is called the *argument* of α and is denoted by $\arg(\alpha) = \arg\alpha$. The argument of α is the counterclockwise angle between 1 and α. If β is another complex number then

$$\alpha\beta = |\alpha||\beta|(\cos(\arg\alpha + \arg\beta) + \sqrt{-1}\sin(\arg\alpha + \arg\beta)) \ . \tag{4.6}$$

This means that multiplication by α is a counterclockwise rotation by the angle $\arg\alpha$ and a scaling by the factor $|\alpha|$.

4.1.4 Units and zero divisors

A *unit* of A is an element α in A that has a multiplicative inverse, that is, there is $\beta \in A$ with $\alpha\beta = 1$. It is easy to verify that the set of units of A is a commutative group. It is called the *unit group* of A and is denoted by A^*. Since $A_{-1} = \mathbb{C}$ is a field, all non-zero elements of A_{-1} are units of A_{-1}.

A *zero divisor* of A is a non-zero element α in A such that there is a non-zero $\beta \in A$ with $\alpha\beta = 0$. Since A_{-1} is a field, A_{-1} contains no zero divisors. We determine the unit group and the set of zero divisors of A_1.

Proposition 4.1.8.

1. *The unit group of A_1 is $A_1^* = \{(\alpha_1, \alpha_2) \in A_1 : \alpha_1\alpha_2 \neq 0\}$.*
2. *The set of zero divisors of A_1 is $\{(\alpha_1, \alpha_2) \in A_1 : (\alpha_1, \alpha_2) \neq 0$ and $\alpha_1\alpha_2 = 0\}$.*
3. *A non-zero element of A_1 is either a unit or a zero divisor of A_1.*

Proof. Exercise 4.7.1.

4.1.5 Automorphisms

We determine the automorphisms of A.

Lemma 4.1.9. *If $\alpha \in A \setminus \mathbb{R}$ with $\alpha^2 = j$, then $\alpha = \pm i$.*

Proof. Write $\alpha = x + iy$, $x, y \in \mathbb{R}$. Then

$$j = \alpha^2 = x^2 + jy^2 + 2xyi \ . \tag{4.7}$$

Since $(1, i)$ is an $\mathbb{R}$-basis of A and since $j \in \mathbb{R}$, (4.7) implies $2xy = 0$. Hence, we have $x = 0$ or $y = 0$. Now $y = 0$ is impossible since $\alpha \notin \mathbb{R}$. Hence $x = 0$. From (4.7) we obtain $jy^2 = j$, which implies $y = \pm 1$. □

We define the conjugation map

$$\sigma : A_j \longrightarrow A_j \; : \; x + i(j)y \longmapsto x - i(j)y \; . \tag{4.8}$$

For a geometric interpretation of σ, see Figures 4.1 and 4.2.

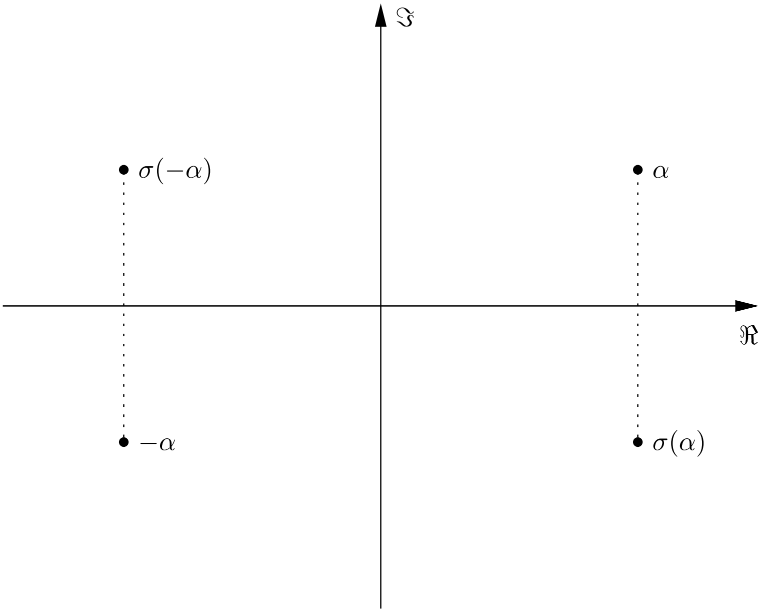

Fig. 4.1. Geometric interpretation of σ in $\mathbb{C}$

Proposition 4.1.10. *Except for the identity, the map $\sigma : A \to A$, $\alpha \mapsto \sigma(\alpha)$ is the only automorphism of A.*

Proof. It is easy to verify that σ is an automorphism of A. Let φ be another automorphism of A. If $x, y \in \mathbb{R}$, then $\varphi(x + iy) = x + y\varphi(i)$. Also, $\varphi(i)^2 = \varphi(i^2) = \varphi(j) = j$. Hence, Lemma 4.1.9 implies $\varphi(i) = \pm i$, as asserted. □

For $\alpha \in \mathbb{C}$ we have $\sigma(\alpha) = \Re\alpha - i\Im\alpha$. So the map $\sigma : \mathbb{C} \to \mathbb{C}$ sends a complex number to its complex conjugate. This map is the reflection with respect to the real axis (see Figure 4.1).

For $\alpha \in A_1$, $\alpha = (\alpha_1, \alpha_2)$, we have $\sigma(\alpha) = (\alpha_2, \alpha_1)$ (see Exercise 4.7.3). Hence, the map $\sigma : A_1 \to A_1$ is the reflection with respect to the real line $\mathbb{R}$ (see Figure 4.2).

We show that the real numbers are the only elements of A that are fixed by σ.

Lemma 4.1.11. *Let $\alpha \in A$. Then $\alpha = \sigma(\alpha)$ if and only if $\alpha \in \mathbb{R}$.*

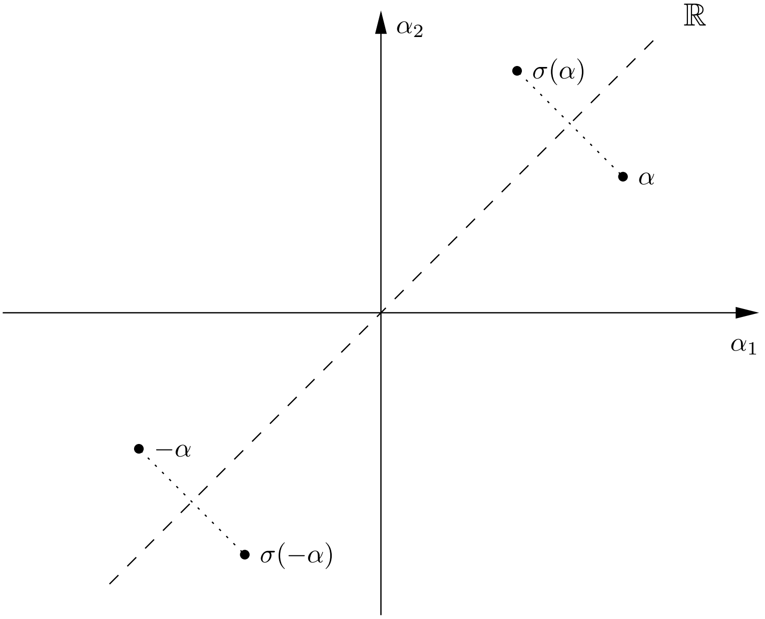

Fig. 4.2. Geometric interpretation of σ in A_1

Proof. Let $x, y \in \mathbb{R}$. Then $x + iy = \sigma(x + iy) = x - iy$ if and only if $y = 0$. This, in turn, is true if and only if $x + iy \in \mathbb{R}$. □

4.1.6 Norm, trace, and characteristic polynomial

We define norm, trace, and characteristic polynomial of elements of A.

Definition 4.1.12. *Let $\alpha \in A$.*

1. *The* norm *of α is $\mathrm{N}(\alpha) = \alpha\sigma(\alpha)$.*
2. *The* trace *of α is $\mathrm{Tr}(\alpha) = \alpha + \sigma(\alpha)$.*
3. *The* characteristic polynomial *of α is $c_\alpha(X) = X^2 - \mathrm{Tr}(\alpha)X + \mathrm{N}(\alpha)$.*

Example 4.1.13. Since $\sigma(i) = -i$, the norm of i is $\mathrm{N}(i) = -i^2 = -j$, the trace of i is $\mathrm{Tr}(i) = i - i = 0$, and the characteristic polynomial of i is $c_i(X) = X^2 - j$.

Proposition 4.1.14. *Norm and trace of $\alpha \in A$ are real numbers.*

Proof. We have $\sigma\bigl(\mathrm{N}(\alpha)\bigr) = \sigma\bigl(\alpha\sigma(a)\bigr) = \sigma(\alpha)\alpha = \mathrm{N}(\alpha)$ and $\sigma\bigl(\mathrm{Tr}(\alpha)\bigr) = \sigma\bigl(\alpha + \sigma(\alpha)\bigr) = \sigma(\alpha) + \alpha = \mathrm{Tr}(\alpha)$. Hence, Lemma 4.1.11 implies the assertion. □

The norm is multiplicative, that is, we have

$$\mathrm{N}(\alpha\beta) = \mathrm{N}(\alpha)\mathrm{N}(\beta)\ , \quad \alpha, \beta \in A\ . \tag{4.9}$$

The trace is additive, that is, we have

$$\mathrm{Tr}(\alpha + \beta) = \mathrm{Tr}(\alpha) + \mathrm{Tr}(\beta)\ , \quad \alpha, \beta \in A\ . \tag{4.10}$$

The characteristic polynomial of $\alpha \in A$ can be written as

$$c_\alpha(X) = (X - \alpha)\big(X - \sigma(\alpha)\big)\ . \tag{4.11}$$

This implies that α and $\sigma(\alpha)$ are zeros of the characteristic polynomial of α. Also note that

$$\mathrm{N}(x + iy) = x^2 - i^2y^2\ , \quad \mathrm{Tr}(x + iy) = 2x\ , \quad x, y \in \mathbb{R}\ . \tag{4.12}$$

For $r \in \mathbb{R}$ we have

$$\mathrm{N}(r) = r^2\ , \quad \mathrm{Tr}(r) = 2r\ , \quad c_r(X) = (X - r)^2\ . \tag{4.13}$$

If $\alpha \in \mathbb{C}$, then we have $\mathrm{N}(\alpha) = (\Re\alpha)^2 + (\Im\alpha)^2$. In this case, the norm of α is the square of the length of the vector α in the complex plane. Also, $\mathrm{Tr}(\alpha) = 2\Re\alpha$. This is twice the length of the projection of α parallel to the real line. This geometric interpretation of the norm and the trace in the complex plane is illustrated in Figure 4.3. Since $\mathbb{C}$ is a field, the characteristic polynomial of α has exactly two zeros, namely α and $\sigma(\alpha)$.

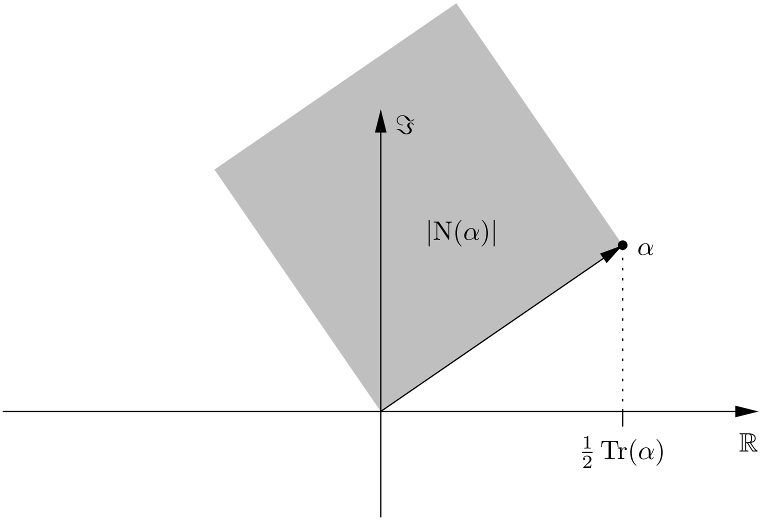

Fig. 4.3. Geometric interpretation of norm and trace in $\mathbb{C}$

Example 4.1.15. Let $\alpha = 1 + \sqrt{-2}$. Then $\mathrm{N}(\alpha) = 1 + 2 = 3$, $\mathrm{Tr}(\alpha) = 2$, $c_\alpha(X) = (X - 1 + \sqrt{-2})(X - 1 - \sqrt{-2}) = X^2 - 2X + 3$.

If $\alpha \in A_1$, $\alpha = (\alpha_1, \alpha_2)$, then $\mathrm{N}(\alpha) = \alpha_1\alpha_2$. So the absolute value of the norm of α is the area of a rectangle (see Figure 4.4). The trace of α is $\mathrm{Tr}(\alpha) = \alpha_1 + \alpha_2$. Also, the characteristic polynomial of α can be written as

$$c_\alpha(X) = (X - \alpha_1)(X - \alpha_2) \ . \tag{4.14}$$

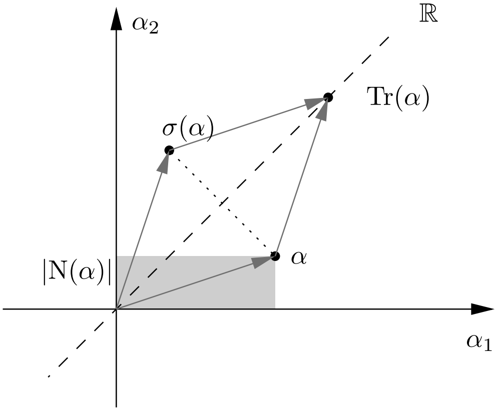

Fig. 4.4. Geometric interpretation of norm and trace in A_1

That characteristic polynomial has four zeros, namely α, $\sigma(\alpha)$, α_1, and α_2. Two of those zeros are real numbers. The other two are not. This is illustrated in Figure 4.5.

Example 4.1.16. Let $\alpha = (1 + \sqrt{2}, 1 - \sqrt{2})$. Then $\mathrm{N}(\alpha) = -1$, $\mathrm{Tr}(\alpha) = 2$ and $c_\alpha(X) = (X - 1 + \sqrt{2})(X - 1 - \sqrt{2}) = X^2 - 2X - 1$.

We characterize the characteristic polynomial of $\alpha \in A \setminus \mathbb{R}$.

Proposition 4.1.17. *Let $\alpha \in A \setminus \mathbb{R}$. Then the characteristic polynomial of α is the only monic polynomial with real coefficients of degree ≤ 2 such that α is a zero of this polynomial.*

Proof. It follows from (4.11) that $c_\alpha(\alpha) = 0$. So α is a zero of $c_\alpha(X)$ and we have

$$\alpha^2 = \mathrm{Tr}(\alpha)\alpha - \mathrm{N}(\alpha) \ . \tag{4.15}$$

Assume that $b, c \in \mathbb{R}$ with $\alpha^2 - b\alpha + c = 0$. Since $(1, \alpha)$ is an $\mathbb{R}$-basis of A, it follows from (4.15) that $b = \mathrm{Tr}(\alpha)$ and $c = \mathrm{N}(\alpha)$. □

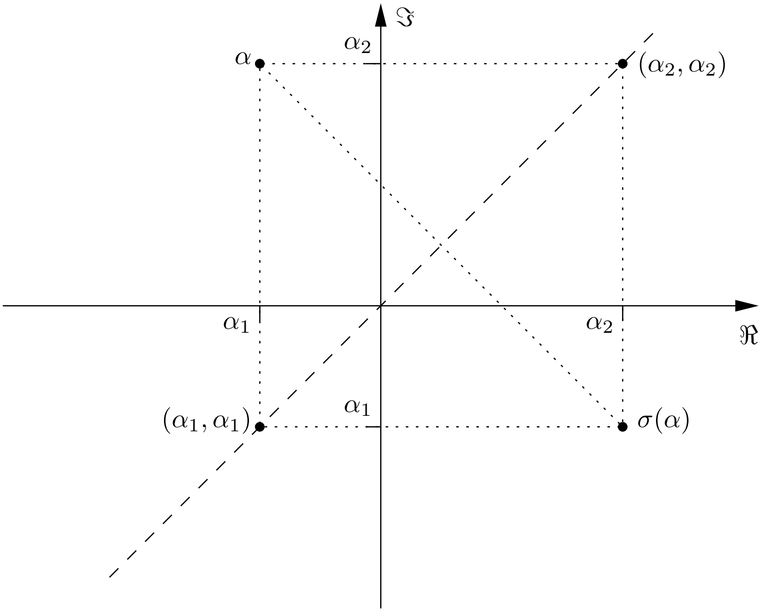

Fig. 4.5. The four zeros of c_α in A_1

4.1.7 Orientation

We introduce the orientation of an ℝ-basis of A. We need the following lemma.

Lemma 4.1.18. *If* $(\alpha, \gamma) = (1, i)T$ *with* $T \in \mathbb{R}^{(2,2)}$, *then* $\sigma(\alpha)\gamma - \alpha\sigma(\gamma) = 2i \det T$.

Proof. Let $T = \begin{pmatrix} s & t \\ u & v \end{pmatrix}$. Then $\sigma(\alpha)\gamma - \alpha\sigma(\gamma) = (s-iu)(t+iv)-(s+iu)(t-iv) = 2i(sv - tu) = 2i \det T$. □

It follows from Lemma 4.1.18 that $\big(\sigma(\alpha)\gamma - \alpha\sigma(\gamma)\big)/i$ is a real number that is non-zero if and only if (α, γ) is an ℝ-basis of A. This justifies the following definition.

Definition 4.1.19. *The* orientation *of an* ℝ*-basis* $B = (\alpha, \gamma)$ *of* A *is the sign of* $\big(\sigma(\alpha)\gamma - \alpha\sigma(\gamma)\big)/i$. *It is denoted by* $\mathrm{o}(B)$. *The orientation of* $\theta \in A \setminus \mathbb{R}$ *is* $\mathrm{o}(\theta) = \mathrm{o}(1, \theta)$.

Example 4.1.20. The orientation of $(1, i)$ is 1. The orientation of $(1, -i)$ is -1. The orientation of $(i, 1)$ is $= -1$.

Let B be an ℝ-basis of A. By Exercise 4.7.8 we have

$$\mathrm{o}(BT) = \mathrm{sign}(\det T)\mathrm{o}(B) \ , \quad T \in \mathrm{GL}(2, \mathbb{R}) \ , \tag{4.16}$$

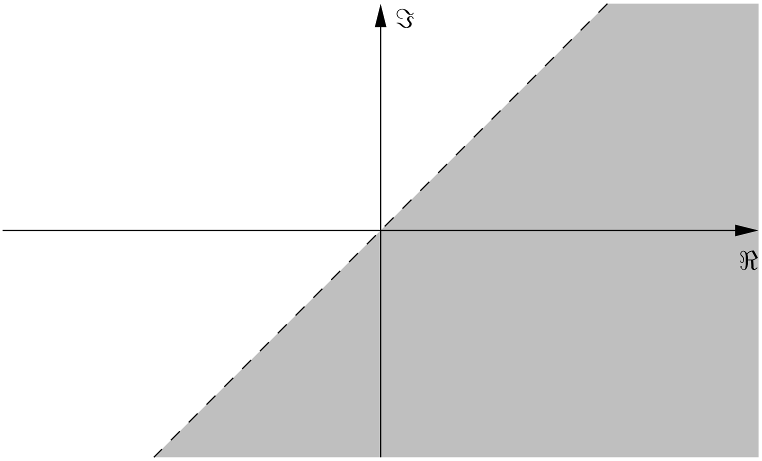

Fig. 4.6. Points of positive orientation in A_1

and

$$\mathrm{o}(\varepsilon B) = \mathrm{sign}\big(\mathrm{N}(\varepsilon)\big)\mathrm{o}(B) \ , \quad \varepsilon \in A^* \ . \tag{4.17}$$

We also have

$$\mathrm{o}(x+iy) = \mathrm{sign}(y) \ , \quad x \in \mathbb{R}, y \in \mathbb{R}\setminus\{0\} \ . \tag{4.18}$$

By (4.18), a complex number has positive orientation if and only if it lies in the *upper half plane*

$$\mathcal{U} = \Big\{\theta \in \mathbb{C} : \Im\theta > 0\Big\} \ . \tag{4.19}$$

It follows from (4.17) and the observations in Section 4.1.3 that the orientation of an $\mathbb{R}$-basis of $\mathbb{C}$ is invariant under rotation and scaling with some real number. Therefore, the bases of $\mathbb{C}$ with positive orientation are exactly the bases (α, γ) such that the counterclockwise angle from α to γ is between 0 and π.

Also, if $\theta = (\theta_1, \theta_2) \in A_1$, then

$$\mathrm{o}(\theta) = \mathrm{sign}(\theta_1 - \theta_2) \ . \tag{4.20}$$

This means that the points of positive orientation in A_1 are below the real line $\big\{(r,r) : r \in \mathbb{R}\big\}$ (see Figure 4.6).

Since the orientation of an $\mathbb{R}$-basis of A_1 is invariant under rotation and scaling, it follows that the bases of A_1 with positive orientation are exactly the bases (α, γ) such that the clockwise angle from α to γ is between 0 and π.

4.1.8 Discriminant

Definition 4.1.21. *The* discriminant *of a pair* $(\alpha, \gamma) \in A^2$ *is* $\Delta(\alpha, \gamma) = \big(\sigma(\alpha)\gamma - \sigma(\gamma)\alpha\big)^2$. *The discriminant of* $\theta \in A$ *is* $\Delta(\theta) = \Delta(1, \theta)$.

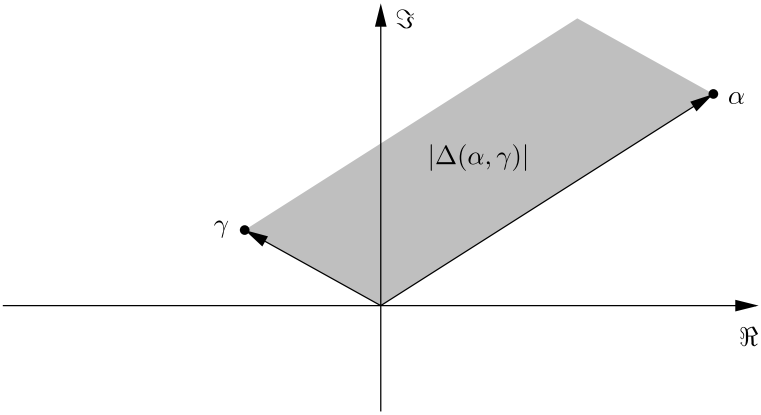

Fig. 4.7. Geometric interpretation of $\Delta(\alpha,\gamma)$

Example 4.1.22. We have $\Delta(i) = \Delta(1,i) = \mathrm{Tr}(i)^2 - 4\mathrm{N}(i) = 4j$. We also have $\Delta(1+\sqrt{-2}, 1-\sqrt{-2}) = -32$.

From Lemma 4.1.18 we obtain the following formula for the discriminant.

$$\Delta((1,i)T) = 4j(\det T)^2 \ , \quad T \in \mathbb{R}^{(2,2)}. \tag{4.21}$$

It follows that the absolute value of the discriminant of an $\mathbb{R}$-basis B of A is four times the square of the area of the parallelotope spanned by B.

Let B be an $\mathbb{R}$-basis of A. Then it follows from (4.21) that

$$\Delta(BT) = (\det T)^2\Delta(B) \ , \quad T \in \mathrm{GL}(2,\mathbb{R}) \ , \tag{4.22}$$

and

$$\Delta(\varepsilon B) = \big(\mathrm{N}(\varepsilon)\big)^2\Delta(B) \ , \quad \varepsilon \in A^* \tag{4.23}$$

(see Exercise 4.7.8). We also have

$$\Delta(x+iy) = 4jy^2 \ , \quad x,y \in \mathbb{R} \ . \tag{4.24}$$

As shown in Exercise 4.7.12 equations (4.24), (4.12), and (4.18) yield

$$\theta = \frac{\mathrm{Tr}(\theta) + o(\theta)i\sqrt{|\Delta(\theta)|}}{2} \ , \quad \theta \in A \ . \tag{4.25}$$

4.2 Irrational forms, bases, points and lattices

We have seen in Example 1.1.1 that finding the representations of an integer n by a reducible integral form f can be reduced to finding the representations of the factors of n by the linear factors of f. Representations of integers by linear forms can be found using the Euclidean algorithm. This is considered elementary. Here, we restrict our attention to integral forms that are irreducible in $\mathbb{Z}[X,Y]$. More generally, we consider irrational forms that are defined now.

Definition 4.2.1. *A form f with real coefficients is called* irrational, *if $f(x, y) \neq 0$ for all $(x, y) \in \mathbb{Z}^2$, $(x, y) \neq (0, 0)$.*

Note that an integral form is irrational if it is irreducible in $\mathbb{Z}[X, Y]$. Also, any positive definite form is irrational.

Example 4.2.2. The form $X^2 + Y^2$ is irrational. The form $F(X, Y) = X^2 + (1 - \pi)XY - \pi Y^2 = (X - \pi Y)(X + Y)$ is not irrational since $f(1, -1) = 0$.

We also define irrational bases, points, and lattices. We will see that irrational forms correspond to those objects.

Definition 4.2.3.

1. *A two-dimensional lattice L in A is called* irrational, *if L contains no zero divisors.*
2. *An $\mathbb{R}$-basis $B = (\alpha, \gamma)$ of A is called* irrational, *if the lattice $L(B) = \{x\alpha + y\gamma : x, y \in \mathbb{Z}\}$ is irrational.*
3. *A point $\theta \in A \setminus \mathbb{R}$ is called* irrational, *if the basis $(1, \theta)$ is irrational.*

Since $\mathbb{C}$ contains no zero divisors, any $\mathbb{R}$-basis, point, or two-dimensional lattice in $\mathbb{C}$ is irrational. We determine the irrational lattices, bases and points in A_1 more explicitly.

Proposition 4.2.4.

1. *A point $\theta = (\theta_1, \theta_2) \in A_1 \setminus \mathbb{R}$ is irrational if and only if both coordinates θ_1 and θ_2 are irrational numbers.*
2. *An $\mathbb{R}$-basis $B = (\alpha, \gamma)$ of A_1 is irrational if and only if α is a unit in A_1 and γ/α is an irrational point in $A_1 \setminus \mathbb{R}$.*
3. *A two-dimensional lattice L in A_1 is irrational if and only if any two different points in L differ in both coordinates.*

Proof. 1. If θ is not irrational, then the lattice $\mathbb{Z} + \theta\mathbb{Z}$ contains a zero divisor. By Proposition 4.1.8 there are integers x and y, $y \neq 0$, such that exactly one of the coordinates of the point $x + y\theta$ is zero. This implies that a coordinate of θ is a rational number.

Conversely, if one of the coordinates of θ is x/y with integers x, y, $y \neq 0$, then $x - y\theta$ is a zero divisor in the lattice $L(\theta)$. Hence, θ is not irrational.

2. Let B be irrational. Since $\alpha \in L(B)$, it follows that α is not a zero divisor of A_1. Proposition 4.1.8 implies that α is a unit in A_1. Also, the lattice $(1/\alpha)L(B)$ is irrational. This means that γ/α is irrational.

Conversely, assume that α is a unit of A_1 and that γ/α is irrational. Then $L(B) = \alpha L(\gamma/\alpha)$ is irrational which means that B is irrational.

3. Suppose that L is not irrational. Then L contains a zero divisor θ. By Proposition 4.1.8 that zero divisor is of the form $(r, 0)$ or $(0, r)$ with a non-zero real number r. Hence, that lattice contains the two points $(0, 0)$ and θ that differ in exactly one coordinate.

Suppose that L contains two points θ and θ' that differ in exactly one coordinate. Then L also contains the point $\theta - \theta'$ which is a zero divisor by Proposition 4.1.8. So L is not irrational. □

4.3 Bases, points, and forms

4.3.1 Oriented norm forms

Let $B = (\alpha, \gamma)$ be an $\mathbb{R}$-basis of A. Then

$$\mathrm{N}(x\alpha + y\gamma) = \mathrm{N}(\alpha)x^2 + \mathrm{Tr}\big(\alpha\sigma(\gamma)\big)xy + \mathrm{N}(\gamma)y^2 \ , \quad x, y \in \mathbb{R} \ . \tag{4.26}$$

This motivates the following definition.

Definition 4.3.1. *Let $B = (\alpha, \gamma)$ be an $\mathbb{R}$-basis of A. The* oriented norm form *of B is*

$$f_B(X, Y) = \mathrm{o}(B)\Big(\mathrm{N}(\alpha)X^2 + \mathrm{Tr}\big(\alpha\sigma(\gamma)\big)XY + \mathrm{N}(\gamma)Y^2\Big) \ .$$

The oriented norm form of a point $\theta \in A \setminus \mathbb{R}$ is

$$f_\theta = f_{(1,\theta)} \ .$$

Example 4.3.2. If $B = (1, i)$, then $f_B(X, Y) = X^2 - i^2Y^2 = X^2 - jY^2$. If $j = 1$, then $f_B(X, Y) = X^2 - Y^2$. If $j = -1$, then $f_B(X, Y) = X^2 + Y^2$.

Example 4.3.3. Let $\theta = (1 + \sqrt{-3})/2$. Then $\mathrm{o}(\theta) = \mathrm{sign}(\Im\theta) = 1$, $\mathrm{Tr}(\theta) = \theta + \sigma(\theta) = 1$, $\mathrm{N}(\theta) = \theta\sigma(\theta) = 1$, hence $f_\theta = (1, 1, 1)$.

Let $\theta = \big((1 + \sqrt{5})/2, (1 - \sqrt{5})/2\big) \in A_1$. Then $\mathrm{o}(\theta) = \mathrm{sign}(\theta_1 - \theta_2) = 1$, $\mathrm{Tr}(\theta) = 1$, $\mathrm{N}(\theta) = -1$, hence $f_\theta = (1, 1, -1)$.

Let B be an $\mathbb{R}$-basis of A. It is shown in Exercise 4.7.10 that for any $\mathbb{R}$-basis of A we have

$$\Delta(f_B) = \Delta(B) \ . \tag{4.27}$$

Also, for a point $\theta \in A \setminus \mathbb{R}$ we have

$$\Delta(f_\theta) = \Delta(\theta) \ . \tag{4.28}$$

4.3.2 Main results

We present the main results that will be proved in this section.

The set

$$A^*(1) = \Big\{\varepsilon \in A : |\mathrm{N}(\varepsilon)| = 1\Big\}.$$

is a subgroup of the unit group A^* of A. It acts on the set of irrational $\mathbb{R}$-bases of A. The $A^*(1)$-orbit of an $\mathbb{R}$-basis B of A is

$$A^*(1)B = \Big\{\varepsilon B : \varepsilon \in A^*(1)\Big\}$$

If B is an $\mathbb{R}$-basis of $\mathbb{C}$, then $\mathbb{C}^*(1)B$ is the set of all bases which are obtained from B by a rotation. If B is an $\mathbb{R}$-basis of A_1, then $A_1^*(1)B$ is the set of all bases which are obtained from B by a scaling with a factor whose norm has absolute value 1.

The first result shows that irrational forms can be identified with $A^*(1)$-orbits of irrational $\mathbb{R}$-bases of A.

Theorem 4.3.4.

1. *The map*

$$\big\{\mathbb{C}^*(1)B : B \ \mathbb{R}\text{-basis of } \mathbb{C}, \mathrm{o}(B) > 0\big\} \to \big\{f : f \text{ positive definite form}\big\},$$
$$\mathbb{C}^*(1)B \mapsto f_B$$

   is a bijection.

2. *The map*

$$\big\{A_1^*(1)B : B \text{ irrational } \mathbb{R}\text{-basis of } A_1\big\} \to \big\{f : f \text{ irrational indefinite form}\big\},$$
$$A_1^*(1)B \mapsto f_B$$

   is a bijection.

3. *The maps from 1. and 2. commute with the action of* $\mathrm{GL}(2,\mathbb{Z})$. *More precisely, for any irrational* $\mathbb{R}$*-basis* B *of* A *and any* $U \in \mathrm{GL}(2,\mathbb{Z})$ *we have* $f_{BU} = f_B U$.

The second result shows that irrational points in $A \setminus \mathbb{R}$ can be identified with $\mathbb{R}_{>0}$-orbits of forms.

Theorem 4.3.5.

1. *Let* $\mathcal{U}$ *be the upper half plane defined in* (4.19)*. The map*

$$\mathcal{U} \to \big\{\mathbb{R}_{>0}f : f \text{ positive definite}\big\}$$
$$\theta \mapsto \mathbb{R}_{>0}f_\theta$$

   is a bijection.

2. *The map*

$$\{\theta \in A_1 \setminus \mathbb{R} : \theta \text{ irrational}\} \to \big\{\mathbb{R}_{>0}f : f \text{ irrational and indefinite}\big\}$$
$$\theta \mapsto \mathbb{R}_{>0}f_\theta$$

   is a bijection.

3. *The maps of 1. and 2. commute with the action of* $\mathrm{GL}(2,\mathbb{Z})$*, that is for any irrational point* θ *in* $A \setminus \mathbb{R}$ *and any* $U \in \mathrm{GL}(2,\mathbb{Z})$ *we have* $\mathbb{R}_{>0}f_{\theta U} = \mathbb{R}_{>0}f_\theta U$.

4.3.3 Properties of oriented norm forms

We present important properties of oriented norm forms.

Proposition 4.3.6.

1. *If* B *is an* $\mathbb{R}$*-basis of* $\mathbb{C}$ *with positive orientation, then the form* f_B *is positive definite.*
2. *If* θ *is a point in the upper half plane, then the form* f_θ *is positive definite.*
3. *If* B *is an* $\mathbb{R}$*-basis of* A_1*, then the form* f_B *is indefinite.*

4. If θ is a point in $A_1 \setminus \mathbb{R}$, then the form f_θ is indefinite.

Proof. This proposition follows from (4.27) and Proposition 1.2.10. □

We show that irrational bases and points correspond to irrational forms.

Proposition 4.3.7.
1. Let B be an $\mathbb{R}$-basis of A. Then B is irrational if and only if f_B is irrational.
2. Let $\theta \in A \setminus \mathbb{R}$. Then θ is irrational if and only if f_θ is irrational.

Proof. All $\mathbb{R}$-bases of $\mathbb{C}$ and all points in $\mathbb{C} \setminus \mathbb{R}$ are irrational, since $\mathbb{C}$ does not contain zero divisors. Also, all positive definite and negative definite forms are irrational. This proves the assertion for $A = \mathbb{C}$ since by Proposition 4.3.6 the oriented norm forms of the $\mathbb{R}$-bases of $\mathbb{C}$ are positive definite or negative definite.

Let B be an $\mathbb{R}$-basis of A. Then $B = (\alpha, \gamma)$ is irrational if and only if the lattice $L(B)$ does not contain zero divisors. The only zero divisors of A_1 are the non-zero points in A_1 of norm zero. Therefore, B is irrational if and only if the oriented norm form $f_B(x, y)$ is non-zero for all $(x, y) \in \mathbb{Z}^2$, $(x, y) \neq (0, 0)$. This means that f_B is irrational. The analogous argument can be used to show that for $\theta \in A_1 \setminus \mathbb{R}$ the form f_θ is irrational if and only if θ is irrational. □

Proposition 4.3.8. *Let B be an $\mathbb{R}$-basis of A. If $\varepsilon \in A^*$, then $f_{\varepsilon B} = |\mathrm{N}(\varepsilon)| f_B$.*

Proof. Let $B = (\alpha, \gamma)$. Using (4.17) we obtain

$$f_{\varepsilon B} = \mathrm{o}(\varepsilon B)\mathrm{N}(X\varepsilon\alpha + Y\varepsilon\gamma) = \mathrm{sign}\big(\mathrm{N}(\varepsilon)\big)\mathrm{N}(\varepsilon)\mathrm{o}(B)\mathrm{N}(X\alpha + Y\gamma) = |\mathrm{N}(\varepsilon)| f_B.$$

□

Proposition 4.3.9.
1. If θ and θ' are irrational points of $A \setminus \mathbb{R}$, then $f_\theta = f_{\theta'}$ if and only if $\theta = \theta'$.
2. If B and B' are irrational $\mathbb{R}$-bases of A, then $f_B = f_{B'}$ if and only if $B' = \varepsilon B$ with $\varepsilon \in A$ such that $|\mathrm{N}(\varepsilon)| = 1$.

Proof. 1. If $f_\theta = f_{\theta'}$, then trace, norm and orientation of θ and θ' are the same. Hence, $\theta = \theta'$ by (4.25).

2. Let $B = (\alpha, \gamma)$ and $B' = (\alpha', \gamma')$. Comparing the coefficients of the forms f_B and $f_{B'}$ we obtain

$$|\mathrm{N}(\alpha')| = |\mathrm{N}(\alpha)|. \tag{4.29}$$

Since B and B' are irrational, the points α and α' are units in A. Set

$$\theta = \gamma/\alpha\ , \quad \theta' = \gamma'/\alpha'\ .$$

Then Proposition 4.3.8 implies

$$|\mathrm{N}(\alpha)| f_\theta = f_B = f_{B'} = |\mathrm{N}(\alpha')| f_{\theta'}. \tag{4.30}$$

It follows from (4.29) and (4.30) that $f_\theta = f_{\theta'}$. Hence by 1. we have $\theta = \theta'$. For $\varepsilon = \alpha'/\alpha$ we have $|\mathrm{N}(\varepsilon)| = 1$ and $B' = \varepsilon B$. □

4.3.4 Action of $\mathbf{GL(2,\mathbb{Z})}$

We define an action of $\mathrm{GL}(2,\mathbb{R})$ on the set of irrational points in $A \setminus \mathbb{R}$.

Definition 4.3.10. *If $\theta \in A \setminus \mathbb{R}$ is irrational and $U = \begin{pmatrix} s & t \\ u & v \end{pmatrix} \in \mathrm{GL}(2,\mathbb{R})$, then we set*

$$\theta U = \frac{t + v\theta}{s + u\theta}.$$

Example 4.3.11. Let $\theta \in \mathbb{C}$. For $T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, we obtain $\theta T = -1/\theta$. For $U = S^s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$, $s \in \mathbb{Z}$, we obtain $\theta U = \theta + s$.

Definition 4.3.12. *Two irrational points in $A \setminus \mathbb{R}$ are called* equivalent *if there is $U \in \mathrm{GL}(2,\mathbb{Z})$ with $\theta' = \theta U$. They are called* properly equivalent *if there is $U \in \mathrm{SL}(2,\mathbb{Z})$ with $\theta' = \theta U$.*

Proposition 4.3.13.
1. If B is an irrational $\mathbb{R}$-basis of A and $U \in \mathrm{GL}(2,\mathbb{Z})$, then $f_{BU} = f_B U$.
2. If θ is an irrational point in $A \setminus \mathbb{R}$ and $U = \begin{pmatrix} s & t \\ u & v \end{pmatrix} \in \mathrm{GL}(2,\mathbb{Z})$, then $f_{\theta U} = \big(1/|\mathrm{N}(s + u\theta|)\big) f_\theta$.

Proof. 1. The matrix of the form f_B can be written as

$$M(f_B) = \frac{\mathrm{o}(B)}{2} \begin{pmatrix} \alpha & \sigma(\alpha) \\ \gamma & \sigma(\gamma) \end{pmatrix} \begin{pmatrix} \sigma(\alpha) & \sigma(\gamma) \\ \alpha & \gamma \end{pmatrix} . \tag{4.31}$$

Since σ is a homomorphism, this implies

$$M(f_{BU}) = \frac{\mathrm{o}(BU)}{2} U^T \begin{pmatrix} \alpha & \sigma(\alpha) \\ \gamma & \sigma(\gamma) \end{pmatrix} \begin{pmatrix} \sigma(\alpha) & \sigma(\gamma) \\ \alpha & \gamma \end{pmatrix} U . \tag{4.32}$$

By (4.16), we have $\mathrm{o}(BU) = \mathrm{sign}(\det U)\mathrm{o}(B)$. Hence (4.31) and (4.32) imply

$$M(f_{BU}) = (\det U) U^T M(f_B) U.$$

But by (2.5), this is the matrix of $M(f_B U)$. So we have proved that $f_{BU} = f_B U$.

2. Using Proposition 4.3.8 and 1. we obtain $f_{\theta U} = f_{\big(1/(s+u\theta)\big)(1,\theta)U} = 1/|\mathrm{N}(s + u\theta)| f_\theta U$. □

4.3.5 Bases and points associated to forms

We also define bases that are associated to irrational forms.

Definition 4.3.14. *Let $f = (a, b, c)$ be an irrational form, $\Delta = \Delta(f)$, $j = \operatorname{sign}\Delta$, and $i = i(j)$. Then $a \neq 0$ and we set*

$$B(f) = \big(a, \frac{b + i\sqrt{|\Delta|}}{2}\big)$$

and

$$\theta(f) = \frac{b + i\sqrt{|\Delta|}}{2a}\ .$$

For an irrational form $f = (a, b, c)$ we have

$$\Delta(B(f)) = a^2\Delta(f)\ , \quad \Delta(\theta(f)) = (1/a^2)\Delta(f). \tag{4.33}$$

Example 4.3.15. For $f = (1, 1, 1)$ we have $\theta(f) = (1 + \sqrt{-3})/2$. For $f = (1, 1, -1)$ we have $\theta(f) = \big((1 + \sqrt{5})/2, (1 - \sqrt{5})/2\big)$.

The following characterization of $\theta(f)$ is easy to verify.

Lemma 4.3.16. *Let $f = (a, b, c)$ be a form with $a \neq 0$. Then the point $\theta(f)$ is the zero of the polynomial $f(X, -1)$ in $A \setminus \mathbb{R}$ whose orientation is the sign of a.*

Proof. Exercise 4.7.13. □

Example 4.3.17. Let $f = 2X^2 + 3XY + 4Y^2$ and $U = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$, $s \in \mathbb{Z}$. Then $\Delta(f) = -23$ and $\theta(f) = (3 + \sqrt{-23})/4$. Now, $\theta(f)U = \theta(f) + s = (3 + 4s + \sqrt{-23})/4$, $fU = 2X^2 + (3 + 4s)XY + (2s^2 + 3s + 4)Y^2$, and $\theta(fU) = (3 + 4s + \sqrt{-23})/4 = \theta(f)U$.

Proposition 4.3.18. *Let $f = (a, b, c)$ be an irrational form, let $B = (\alpha, \gamma)$ be an irrational $\mathbb{R}$-basis of A and let $\theta \in A \setminus \mathbb{R}$ be irrational. Then we have*

1. *$\theta(f_\theta) = \theta$,*
2. *$\theta(f_B) = \gamma/\alpha$,*
3. *$f_{\theta(f)} = (1/|a|)f$,*
4. *$B(f_B) = \frac{(\mathrm{N}(\alpha))^2}{\alpha}B$,*
5. *$f_{B(f)} = |a|f$.*

Proof. 1. We have $f_\theta = \mathrm{o}(\theta)\big(1, \mathrm{Tr}(\theta), \mathrm{N}(\theta)\big)$ and $\Delta(f_\theta) = \Delta(\theta)$. Hence, the definition of $\theta(f)$ and (4.25) imply the assertion.

2. By Proposition 4.3.8 and 1. we have $\theta(f_B) = \theta\big(f_{\alpha(1,\gamma/\alpha)}\big) = \theta(|\mathrm{N}(\alpha)|f_{\gamma/\alpha}) = \gamma/\alpha$.

3. This assertion follows from the fact that $\mathrm{o}\big(\theta(f)\big) = \mathrm{sign}(a)$, $\mathrm{Tr}\big(\theta(f)\big) = b/a$, and $\mathrm{N}\big(\theta(f)\big) = c/a$.

4. Let $B = (1, \theta)$. Then 1. implies $B(f_B) = \big(1, \theta(f_\theta)\big) = (1, \theta) = B$. Let $B = (\alpha, \gamma)$, $\theta = \gamma/\alpha$. We obtain from Proposition 4.3.8 $B(f_B) = B(f_{\alpha(1,\theta)}) = B(|\mathrm{N}(\alpha)| f_\theta) = (\mathrm{N}(\alpha))^2 B(f_\theta) = \big((\mathrm{N}(\alpha))^2/\alpha\big)B$.

5. We have $f_{B(f)} = f_{a(1,\theta(f))} = a^2 f_{\theta(f)} = a^2/|a| f = |a| f$. □

Proposition 4.3.19. *Let f be an irrational form.*

1. *If f is positive definite, then $B(f)$ is an $\mathbb{R}$-basis of $\mathbb{C}$ with positive orientation and $\theta(f)$ is a point in $\mathcal{U}$.*
2. *If f is indefinite, then $B(f)$ is an irrational $\mathbb{R}$-basis of A_1 and $\theta(f)$ is an irrational point in A_1.*

Proof. If a form f is irrational, then both $B(f)$ and $\theta(f)$ are irrational.

If f is positive definite, then $\Delta(f) < 0$ and $a > 0$ by Proposition 1.2.10. By (4.33) this implies $\Delta(B(f)) < 0$ and $\Delta(\theta(f)) < 0$. By (4.21), this shows that $B(f)$ is an $\mathbb{R}$-basis of $\mathbb{C}$ and that $\theta(f)$ is a point in $\mathcal{U}$. The orientation of $B(f)$ and $\theta(f)$ is $\mathrm{sign}(a) > 0$.

If f is indefinite, then $\Delta(f) > 0$ by Proposition 1.2.10. By (4.33) this implies $\Delta(B(f)) > 0$ and $\Delta(\theta(f)) > 0$. By (4.21), this shows that $B(f)$ is an $\mathbb{R}$-basis of A_1 and that $\theta(f)$ is a point in A_1. □

Proposition 4.3.20. *If $f = (a, b, c)$ and $f' = (a', b', c')$ are irrational forms, then*

1. *$B(f) = B(f')$ if and only if $f = f'$ and*
2. *$\theta(f) = \theta(f')$ if and only if $|a'|f = |a|f'$.*

Proof. 1. If $f = f'$, then $B(f) = B(f')$. If $B(f) = B(f')$, $f = (a, b, c)$, and $f' = (a', b', c')$, then $a = a'$ and Proposition 4.3.18 implies $|a|f = f_{B(f)} = f_{B(f')} = |a|f'$, hence $f = f'$.

2. If $|a'|f = |a|f$, then $\theta(f) = \theta(f')$. If $\theta(f) = \theta(f')$, then Proposition 4.3.18 implies $(1/|a|)f = f_{\theta(f)} = f_{\theta(f')} = (1/|a'|)f'$. □

Proposition 4.3.21. *Let $f = (a, b, c)$ be an irrational form and $U = \begin{pmatrix} s & t \\ u & v \end{pmatrix} \in \mathrm{GL}(2, \mathbb{Z})$. Then we have*

1. *$\theta(fU) = \theta(f)U$ and*
2. *$B(fU) = \det U \,\big(s + u\sigma(\theta(f))\big) B(f)U$.*

Proof. 1. Using Propositions 4.3.13 and 4.3.18 we obtain $\theta(fU) = \theta\big(f_{\theta(f)}U\big) = \theta\big(f_{\theta(f)U}\big) = \theta(f)U$.

2. From 1. we deduce

$$\begin{aligned} B(fU) &= a\mathrm{N}\big(s + u\theta(f)\big) \det U \,\big(1, \theta(fU)\big) \\ &= a\mathrm{N}\big(s + u\theta(f)\big) \det U \,\big(1, \theta(f)U\big) \\ &= \det U \,\big(s + u\sigma(\theta(f))\big) B(f)U\ . \end{aligned}$$

□

4.3.6 Proof of the main results

The maps from Theorems 4.3.4 and 4.3.5 are well defined by Proposition 4.3.6. They are injective by Proposition 4.3.9 and surjective by Proposition 4.3.9. The maps commute with the action of $\mathrm{GL}(2, \mathbb{Z})$ by Proposition 4.3.13.

4.4 Lattices and forms

In this section we explain how to translate between the laguages of forms and lattices.

4.4.1 Lattices that correspond to forms

We begin by defining how to associate lattices to bases, points and forms.

Definition 4.4.1.
1. With an $\mathbb{R}$-basis $B = (\alpha, \gamma)$ of A we associate the lattice

$$L(B) = \mathbb{Z}\alpha + \mathbb{Z}\gamma.$$

2. With a point in $A \setminus \mathbb{R}$ we associate the lattice

$$L(\theta) = L(1, \theta) = \mathbb{Z} + \mathbb{Z}\theta.$$

3. With a form f we associate the lattice $L(f) = L(B(f))$.

Example 4.4.2. If $f = (1, 0, 1)$, then $L(f) = \mathbb{Z} + \mathbb{Z}\sqrt{-1}$. This is the lattice of *Gaussian integers*. If $f = (1, 1, 1)$, then $L(f) = \mathbb{Z} + \mathbb{Z}(1 + \sqrt{-3})/2$. This is the hexagonal lattice. Those lattices are illustrated in Figure 4.8 and Figure 4.9.

In Exercise 4.7.14 it is shown that the discriminant of a $\mathbb{Z}$-basis of a two-dimensional lattice L in A is an invariant of L. This justifies the following definition.

Definition 4.4.3. *The* discriminant *of a two dimensional lattice L in A is the discriminant of any $\mathbb{Z}$-basis of L. It is denoted by $\Delta(L)$.*

4.4.2 Main result

Here is our main result.

Theorem 4.4.4.
1. The map

$$\{f\Gamma : f \textit{ positive definite form}\} \mapsto \{L : L \textit{ irrational lattice in } \mathbb{C}, L \cap \mathbb{R} \neq \emptyset\}$$
$$f\Gamma \mapsto L(f)$$

is a bijection.

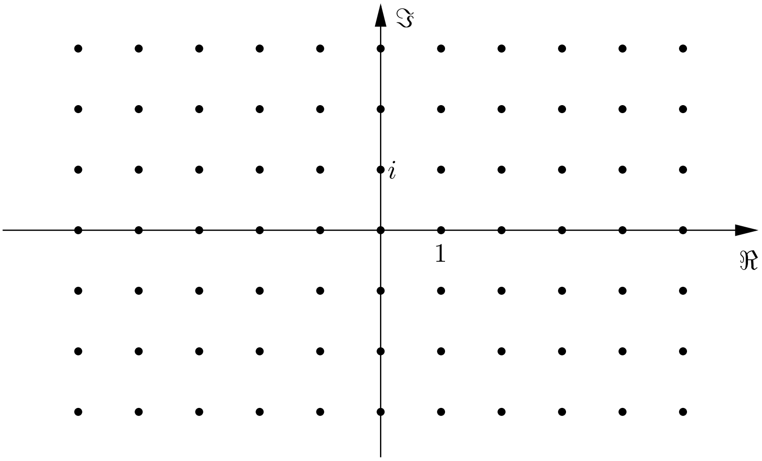

Fig. 4.8. The Gaussian integers

2. The map

$$\begin{aligned} \{f\Gamma : f &= (a,b,c) \textit{ irr. indefinite }, a>0\} \\ &\mapsto \{L : L \textit{ irr. lattice in } A_1, L\cap\mathbb{R}\neq\emptyset\} \\ f\Gamma &\mapsto L(f) \end{aligned}$$

is a bijection.

We will also define equivalence of lattices and we will show that the equivalence of forms implies the equivalence of lattices.

4.4.3 Properties of lattices associated to forms

We prove a few elementary properties of the lattice associated to a form and show how the action of $\mathrm{GL}(2,\mathbb{Z})$ on the set of forms is translated to lattices.

Proposition 4.4.5. *Let $f=(a,b,c)$ be an irrational form and $U=\begin{pmatrix} s & t \\ u & v \end{pmatrix}\in \mathrm{GL}(2,\mathbb{Z})$. Then we have*

1. $L(f)=aL\big(\theta(f)\big)$,
2. $\Delta\big(L(f)\big)=a^2\Delta(f)$,
3. $|a|=\min L(f)\cap\mathbb{R}_{>0}$,
4. $L(fU)=\big(s+u\sigma(\theta(f))\big)L(f)$.

Proof. 1. This is an easy computation.

2. By (4.27) and Proposition 4.3.18 we have $\Delta\big(L(f)\big)=\Delta\big(B(f)\big)=\Delta(f_{B(f)})=\Delta(|a|f)=a^2\Delta(f)$.

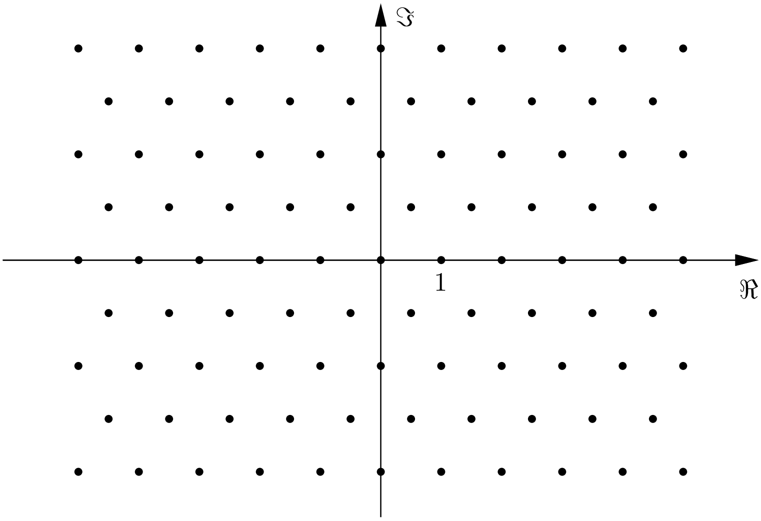

Fig. 4.9. The hexagonal lattice

3. If r is a positive real number in $L(f)$, then we can write $r = xa + ya\theta(f)$ with $x, y \in \mathbb{Z}$. Since $\theta(f) \notin \mathbb{R}$, we have $y = 0$. Hence, the minimal positive r is obtained for $x = \text{sign}(a)$. This r is $r = |a|$.

4. From Proposition 4.3.21 we obtain $L(fU) = L(B(fU)) = L\Big((s + u\sigma\big(\theta(f)\big)B(f)U\Big)$. Since

$$B(f)U$$

is another basis of $L(f)$, this implies the assertion. □

Example 4.4.6. Let $f = (1,0,1)$ and $f' = (2,2,1)$. Since $(2,2,1) = (1,0,1)U$ with $U = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ we have L(f') = $(1 - \sqrt{-1})L(f)$.

Proposition 4.4.7. *Let $f = (a,b,c)$ and $f' = (a',b',c')$ be irrational forms with $aa' > 0$. Then $L(f) = L(f')$ if and only if the Γ-orbits of f and f' are equal.*

Proof. Let $L(f) = L(f')$. Since $aa' > 0$, it follows from Proposition 4.4.5 3. that $a = a'$. Now $B(f) = a\big(1, \theta(f)\big)$. Hence, $\theta(f') = \theta(f) + m$ for some integer m. It follows that $\theta(f') = \theta(f)U$ with some $U \in \Gamma$. Proposition 4.3.21 1. implies $\theta(f') = \theta(fU)$. From Proposition 4.3.18 2. we obtain $fU = |a|f_{\theta(fU)} = |a'|f_{\theta(f')} = f'$.

Conversely, assume that $f' = fU$ with $U \in G$. Then $U \in \Gamma$ and Proposition 4.3.21 2. imply $L(f') = L\big(B(f')\big) = L\big(B(fU)\big) = L\big(B(f)U\big) = L\big(B(f)\big) = L(f)$. □

.

4.4.4 Equivalence of lattices

We define equivalence of lattices.

Definition 4.4.8.

1. *Two lattices L and L' in A are called* equivalent *if there is $\alpha \in A^*$ with $L' = \alpha L$.*
2. *Two lattices L and L' in A are called* properly equivalent *if there is $\alpha \in A^*$ with $\mathrm{N}(\alpha) > 0$ and $L' = \alpha L$.*

We show that this definition is compatible with the equivalence definition for forms.

Proposition 4.4.9. *Let $f = (a, b, c)$ and $f' = (a', b', c')$ be forms. Then the following statements are true:*

1. *If the forms f and f' are equivalent, then the lattices $L(f)$ and $L(f')$ are equivalent.*
2. *If $aa' > 0$ and the forms f and f' are properly equivalent, then the lattices $L(f)$ and $L(f')$ are properly equivalent.*
3. *If the lattices $L(f)$ and $L(f')$ are equivalent, then the points $\theta(f)$ and $\theta(f')$ are equivalent.*
4. *If $aa' > 0$ and the lattices $L(f)$ and $L(f')$ are properly equivalent, then the points $\theta(f)$ and $\theta(f')$ are properly equivalent.*

Proof. 1. Assume that $f' = fU$ with $U = \begin{pmatrix} s & t \\ u & v \end{pmatrix} \in \mathrm{GL}(2, \mathbb{Z})$. Then Proposition 4.4.5 4. implies

$$L(f') = \big(s + u\theta(f)\big)L(f). \tag{4.34}$$

Hence $L(f)$ and $L(f')$ are equivalent.

2. Assume that in the proof of 1. we have $U \in \mathrm{SL}(2, \mathbb{Z})$. Then $B(f)$ and $B(f)U$ have the same orientation. Also, since $aa' > 0$ the bases $B(f)$ and $B(f') = B(fU)$ have the same orientation. It follows that

$$\mathrm{o}\big(B(fU)\big) = \mathrm{o}\big(B(f)U\big). \tag{4.35}$$

From Proposition 4.3.21 2. and (4.17) we obtain $\mathrm{o}\big(B(fU)\big) = \mathrm{sign}\Big(\mathrm{N}\big(s + \sigma(\theta(f))\big)\Big)\mathrm{o}\big(B(fU)\big)$. So (4.35) implies that $\mathrm{N}\big(s + u\theta(f)\big) > 0$. Hence, $L(f)$ and $L(f')$ are properly equivalent by (4.34).

3. Since $\theta(f) = \theta((1/a)f)$, we may assume that $a = a' = 1$, $L(f) = L(\theta(f))$, and $L(f') = L(\theta(f'))$. Assume that $L(f)$ and $L(f')$ are equivalent. Then there is a unit ε in A such that $L(f') = \varepsilon L(f)$. It follows that $\varepsilon(1, \theta(f))$ is a $\mathbb{Z}$-basis of $L(f')$. Hence, there is $U \in \mathrm{GL}(2, \mathbb{Z})$ with $\varepsilon(1, \theta(f)) = (1, \theta(f'))U$. This implies $\theta(f) = \theta(f')U$. Hence, $\theta(f)$ and $\theta(f')$ are equivalent.

4. Assume that $aa' > 0$ and that in the proof of 3. we have $\mathrm{N}(\varepsilon) > 0$. Then from (4.17) we obtain $\mathrm{o}(\varepsilon(1, \theta(f))) = \mathrm{sign}(\mathrm{N}(\varepsilon))\mathrm{o}(\theta(f)) = \mathrm{sign}(a)$. But since $\varepsilon(1, \theta(f)) = (1, \theta(f'))U$, it follows from (4.16) that $\mathrm{sign}(a') = \mathrm{sign}(a) = \mathrm{o}((1, \theta(f'))U) = \mathrm{sign}(\det U)\mathrm{o}(1, \theta(f')) = \mathrm{sign}(\det U)\,\mathrm{sign}(a')$. So $\det U = 1$ which means that $\theta(f)$ and $\theta(f')$ are properly equivalent. □

4.4.5 Forms associated to lattices

We also associate forms with lattices in A.

Lemma 4.4.10. *Let L be an irrational lattice that contains non-zero real numbers. Let $a = \min(L \cap \mathbb{R}_{>0})$. Then there is an irrational form $f = (a, b, c)$, unique up to Γ-equivalence, satisfying $L = L(f)$.*

Proof. There exists a $\theta \in A$ such that $(a, a\theta)$ is a positively oriented basis of L, cf. Exercise A.6.1. Set $f = af_\theta$. Then $f = (a, b, c)$ with $b = a\,\mathrm{Tr}(\theta)$ and $c = aN(\theta)$. Since the lattice L is irrational, so are the basis $(a, a\theta)$ (by Definition 4.2.3), the point θ (by Proposition 4.2.4), and the form f_θ (by Proposition 4.3.7). Hence Proposition 4.3.18 implies $L(f) = L(B(f)) = L\Big((a, a\theta(f_\theta))\Big) = L(a, a\theta) = L$.

Uniqueness follows from Proposition 4.4.7. □

Definition 4.4.11. *Let L be an irrational lattice that contains non-zero real numbers. Then the form from Lemma 4.4.10 is denoted by f_L.*

By definition we have

$$L(f_L) = L \tag{4.36}$$

for any irrational lattice L.

4.4.6 Proof of the main result

We prove Theorem 4.4.4. The maps are well defined by Proposition 4.3.19. The maps are injective by Proposition 4.4.7. The maps are surjective by (4.36).

4.5 Quadratic irrationalities and forms

In this section we characterize the points in $A\backslash\mathbb{R}$ that correspond to irreducible integral forms. Those points are the quadratic irrationalities.

Definition 4.5.1. *A* quadratic irrationality *is an element of A which is a zero of an irreducible quadratic polynomial in $\mathbb{Q}[X]$.*

Note that quadratic irrationalities in $A \setminus \mathbb{R}$ are irrational.

Example 4.5.2. The complex number $\sqrt{-3}$ is a quadratic irrationality. It is a zero of the polynomial $X^2 + 3$. Also, $\sqrt{3}$ is a quadratic irrationality. It is a zero of $X^2 - 3$. Another quadratic irrationality is $(\sqrt{3}, -\sqrt{3})$ which belongs to $A_1 \setminus \mathbb{R}$ and is also a zero of the polynomial $X^2 - 3$.

Lemma 4.5.3. *For any quadratic irrationality θ there is exactly one primitive quadratic polynomial $p \in \mathbb{Z}[X]$ with $p(\theta) = 0$ and positive leading coefficient. This polynomial is irreducible in $\mathbb{Q}[X]$.*

Proof. The existence of a polynomial with the asserted properties follows from the definition of a quadratic irrationality. We prove the uniqueness. Let $p(X) = aX^2 + bX + c$ and $p'(X) = a'X^2 + b'X + c'$ be two such polynomials. Then $q = p/a - p'/a'$ is a linear polynomial in $\mathbb{Q}[X]$ with $q(\theta) = 0$. Since θ is irrational, we must have $q = 0$. This implies that $a'p = ap'$. But $a = \mathrm{cont}(ap') = \mathrm{cont}(a'p) = a'$ since p and p' are primitive. Hence $p = p'$, as asserted. □

We characterize the quadratic irrationalities in $A \setminus \mathbb{R}$.

Proposition 4.5.4. *Let $\theta \in A \setminus \mathbb{R}$. Then θ is a quadratic irrationality if and only if the characteristic polynomial of θ has rational coefficients and no rational zero.*

Proof. If $c_\theta \in \mathbb{Q}[X]$ and c_θ has no rational zero, then c_θ is irreducible in $\mathbb{Q}[X]$. Also, we have $c_\theta(\theta) = 0$ by (4.11). Hence, by definition, θ is a quadratic irrationality.

Conversely, assume that θ is a quadratic irrationality. Then by definition, θ is a zero of a monic polynomial in $\mathbb{Q}[X]$ without rational zeros. By Proposition 4.1.17, this is the characteristic polynomial of θ. □

It follows from Proposition 4.5.4 that norm and trace of quadratic irrationalities in $A \setminus \mathbb{R}$ are rational numbers. In the next two examples we show that Proposition 4.5.4 can be used to decide whether $\alpha \in A \setminus \mathbb{R}$ is a quadratic irrationality.

Example 4.5.5. Let $\theta = (\sqrt{3}+\sqrt{12}, -\sqrt{3}-\sqrt{12}) \in A_1$. We have $\mathrm{Tr}(\theta) = 0$ and $\mathrm{N}(\theta) = -27$. So $c_\theta(X) = X^2 - 27$. We know from (4.14) that the real zeros of that polynomial are $\sqrt{3} + \sqrt{12}$ and $-\sqrt{3} - \sqrt{12}$. They are both irrational. Hence, θ is a quadratic irrationality.

Example 4.5.6. Let $\theta = (\sqrt{3} + 1, -\sqrt{3} - 1) \in A_1$. We have $\mathrm{Tr}(\theta) = 0$ and $\mathrm{N}(\theta) = -4 - 2\sqrt{3}$. Hence, θ is not a quadratic irrationality.

We show that we can identify the quadratic irrationalities in $A_1 \setminus \mathbb{R}$ with the quadratic irrationalities in $\mathbb{R}$.

Corollary 4.5.7. *The map that sends a quadratic irrationality (θ_1, θ_2) in $A_1 \setminus \mathbb{R}$ to its first component is a bijection between the quadratic irrationalities in $A_1 \setminus \mathbb{R}$ and the quadratic irrationalities in $\mathbb{R}$.*

Proof. Clearly, the map is well defined. We show that the map is surjective. So let r be a quadratic irrationality in $\mathbb{R}$. By Lemma 4.5.3, there is an integral primitive irreducible polynomial $p(X) = aX^2 - bX + c$ such that $p(r) = 0$. Set $\Delta = b^2 - 4ac$. Then $r = (b + \sqrt{\Delta})/(2a)$ or $r = (b - \sqrt{\Delta})/(2a)$. Set $\theta = \big((b + \sqrt{\Delta})/(2a), (b - \sqrt{\Delta})/(2a)\big)$. Then r is the image of θ or of $\sigma(\theta)$. Finally, we prove the injectivity. Suppose that $\theta = (\theta_1, \theta_2)$ and $\theta' = (\theta'_1, \theta'_2)$ are two quadratic irrationalities in $A_1 \backslash \mathbb{R}$ which have the same first component. By Lemma 4.5.3, there are two primitive integral quadratic polynomials p and p' with $p(\theta) = p'(\theta') = 0$. Since $\theta_1 = \theta'_1$ we also have $p(\theta_1) = p'(\theta_1) = 0$. Hence Lemma 4.5.3 implies $p = p'$ and therefore, $\theta_2 = \theta'_2$ because θ_2 is the second real zero of p and θ'_2 is the second real zero of p'. □

We prove that quadratic irrationalities in $A \setminus \mathbb{R}$ and primitive integral irreducible forms can be identified.

Theorem 4.5.8.

1. *The map that sends an integral primitive positive definite form f to the point $\theta(f)$ is a bijection between all those forms and the quadratic irrationalities in the upper half plane.*
2. *The map that sends an integral irreducible primitive indefinite form $f = (a, b, c)$ to the point $\theta(f)$ is a bijection between all those forms and the quadratic irrationalities in A_1.*
3. *Let f and f' be integral primitive irreducible forms and let $U \in \mathrm{GL}(2, \mathbb{Z})$. Then $f' = fU$ if and only if $\theta(f') = \theta(f)U$.*

Proof. 1. and 2. Let $f = (a, b, c)$ be an integral irreducible form. Then $a \neq 0$. So $\theta(f)$ is a quadratic irrationality. Theorem 4.3.5 implies that the two maps are well defined and injective.

We prove the surjectivity of the maps. Let θ be a quadratic irrationality in $A \setminus \mathbb{R}$. Let $p(X)$ be the polynomial from Lemma 4.5.3 with $p(\theta) = 0$. Set $f(X, Y) = \mathrm{o}(\theta) Y^2 p(-X/Y)$. Then $f(\theta, -1) = 0$ and Lemma 4.3.16 implies $\theta(f) = \theta$.

3. We prove the third assertion. If $f' = fU$ then $\theta(f') = \theta(f)U$ by Theorem 4.3.5. Conversely, assume that $\theta(f') = \theta(f)U$. Then $\theta(f') = \theta(fU)$ by Theorem 4.3.5. This theorem also implies that there is a positive real number r with $f' = rfU$. But both f' and fU are integral primitive forms. So we have $r = 1$. □

By Theorem 4.5.8, any quadratic irrationality $\theta \in A \setminus \mathbb{R}$ has a unique representation $\theta = \theta(a, b, c)$ with some integral primitive form (a, b, c). We call this the *standard representation* of θ . To determine the standard representation of θ, we compute the orientation o, the trace t, and the norm n of θ. We also compute the least common denominator d of t and n. Then the standard representation of θ is $\theta = \theta(od, td, nd)$.

Example 4.5.9. Consider the complex number $\theta = \sqrt{-3} + \sqrt{-12}$. We show that θ is a quadratic irrationality and determine its standard representation.

The orientation of θ is 1, the trace of θ is 0 and the norm of θ is 27. It follows that the standard representation of θ is $\theta = \theta(1, 0, 27)$.

4.6 Quadratic lattices and forms

We now consider the lattices that are obtained from integral irreducible forms. We will explain the correspondence between such forms and lattices.

Definition 4.6.1. *If f is an integral irreducible form, then the lattice $L(f)$ is called a* quadratic lattice.

The next proposition explains the correspondence between integral primitive forms and lattices.

Proposition 4.6.2. *Let $f = (a, b, c)$ and $f = (a', b', c')$ be integral primitive irreducible forms. Then the following are true:*

1. *If $aa' > 0$, then the lattices $L(f) = L(f')$ are equal if and only if f and f' belong to the same Γ-orbit.*
2. *If $aa' > 0$, then the lattices $L(f)$ and $L(f')$ are equivalent if and only if the forms f and f' are equivalent.*
3. *The lattices $L(f)$ and $L(f')$ are properly equivalent if and only if the forms f and f' are properly equivalent.*

Proof. 1. This is a consequence of Theorem 4.4.4.

2. If f and f' are equivalent, then $L(f)$ and $L(f')$ are equivalent by Proposition 4.4.9. If $L(f)$ and $L(f')$ are equivalent, then again by Proposition 4.4.9 the points $\theta(f)$ and $\theta(f')$ are equivalent. Hence, f and f' are equivalent by Theorem 4.5.8.

3. The proof of the third assertion is left to the reader as Exercise 4.7.15. □

As we see from Proposition 4.6.2 quadratic lattices and Γ-orbits of primitive integral forms (a, b, c) with $a > 0$ can be identified. We also have the following result.

Corollary 4.6.3.

1. *The map that sends the proper equivalence class of an integral primitive positive definite form $f = (a, b, c)$ to the proper equivalence class of the lattice $L(f)$ is a bijection between the proper equivalence classes of integral positive definite forms and the proper equivalence classes of quadratic lattices in $\mathbb{C}$.*
2. *The map that sends the equivalence class of an integral primitive irreducible indefinite form $f = (a, b, c)$ to the equivalence class of the lattice $L(f)$ is a bijection between the equivalence classes of integral primitive irreducible forms and the equivalence classes of quadratic lattices in A_1.*
3. *The map that sends the proper equivalence class of an integral primitive irreducible indefinite form $f = (a, b, c)$ with $a > 0$ to the equivalence class of the lattice $L(f)$ is a bijection between the proper equivalence classes of such forms and the equivalence classes of quadratic lattices in A_1.* □

4.7 Exercises

For all exercises we fix $j \in \{-1, 1\}$ and we set $i = i(j)$ and $A = A_j$.

Exercise 4.7.1. Prove Proposition 4.1.8.

Exercise 4.7.2.

1. Construct A_0. For this purpose, define a product on $\mathbb{R}^2$ such that $\mathbb{R}^2$ becomes a two-dimensional commutative $\mathbb{R}$-algebra that contains an $\mathbb{R}$-basis $(1, i(0))$ with $i(0)^2 = 0$.
2. Determine the group of units and the zero divisors of A_0.

Exercise 4.7.3. Let $\alpha = (\alpha_1, \alpha_2) \in A_1$. Show that $\sigma(\alpha) = (\alpha_2, \alpha_1)$.

Exercise 4.7.4. Determine all complex numbers α with $\Re\alpha^2 = \Re\alpha$ and $\Im\alpha^2 = \Im\alpha$.

Exercise 4.7.5. Prove that A_1 contains no square root of -1.

Exercise 4.7.6. Let φ be an automorphism of A.

1. Prove that $\varphi(-\alpha) = -\varphi(\alpha)$ for all $\alpha \in A$.
2. Let $\alpha \in A^*$. Prove that $\varphi(\alpha) \in A^*$ and that $\varphi(\alpha^{-1}) = \varphi(\alpha)^{-1}$.

Exercise 4.7.7. Which of the following points in $\mathbb{C}$ is a quadratic irrationality: $\sqrt{-8} + \sqrt{-2}$, $\sqrt{-4} + \sqrt{-2}$?

Exercise 4.7.8.

1. Prove that $\mathrm{o}(\varepsilon B) = \mathrm{sign}(\mathrm{N}(\varepsilon))\mathrm{o}(B)$ and $\Delta(\varepsilon B) = \big(\mathrm{N}(\varepsilon)\big)^2\Delta(B)$ for any $\mathbb{R}$-basis B of A and any unit ε of A.
2. Prove that $\mathrm{o}(BT) = \mathrm{sign}(\det T)\mathrm{o}(B)$ and $\Delta(BT) = (\det T)^2\Delta(B)$ for any $\mathbb{R}$-basis B of A and any $T \in \mathrm{GL}(2, \mathbb{R})$.

Exercise 4.7.9. Determine the forms f_B for the bases $B = (\sqrt{-4}, 1 + \sqrt{-1})$ and $B = \big(1, 2 + (\sqrt{5}, -\sqrt{5})\big)$.

Exercise 4.7.10. Prove that $\Delta(f_B) = \Delta(B)$ for any $\mathbb{R}$-basis B of A.

Exercise 4.7.11. Determine the orientation of the $\mathbb{R}$-basis $\big((1,0),(0,1)\big)$ of A_1.

Exercise 4.7.12. Prove the formula (4.25).

Exercise 4.7.13. Prove Lemma 4.3.16.

Exercise 4.7.14. Let L be a two-dimensional lattice in A. Prove that the discriminant of an $\mathbb{R}$-basis of L is an invariant of L.

Exercise 4.7.15. Let $f = (a, b, c)$ and $f = (a', b', c')$ be integral primitive irreducible forms with $aa' > 0$. Prove that the lattices $L(f)$ and $L(f')$ are properly equivalent if and only if the forms f and f' are properly equivalent.

Exercise 4.7.16. Prove that $\sigma\bigl(L(a,b,c)\bigr) = L(a,-b,c)$ for any integral irreducible form (a,b,c).

Exercise 4.7.17. Prove that the non-real elements of a quadratic lattice L are quadratic irrationalities and that the real numbers in L are integers.

5

Reduction of Positive Definite Forms

In this chapter we solve the problems of deciding equivalence and finding the minimum of forms of negative discriminant. First we show, that it suffices to solve those problems for positive definite forms. Then we solve the problems for positive definite forms using reduction theory. We define reduced forms and we show that every proper equivalence class of positive definite forms contains exactly one reduced form. We prove that the coefficient of X^2 in that reduced form is the minimum of any form in the proper equivalence class. We present an efficient algorithm for computing the reduced form that is properly equivalent to a given form. That algorithm can be used to find the minimum of the form as the coefficient of X^2 in the reduced form. Also, proper equivalence of two positive definite forms can be efficiently decided by reducing those forms and then comparing the result. In the whole chapter we assume that Δ is a negative real number and that $f = (a, b, c)$ is a form with real coefficients and discriminant Δ. Except for the first section, we assume that f is positive definite.

5.1 Negative definite forms

By Proposition 2.3.5, an equivalence class of forms of discriminant Δ is the disjoint union of two proper equivalence classes of forms, one containing the positive definite forms in the class and one containing the negative definite forms in the class. The map that sends a form f to the form fV with

$$V = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

is a bijection between those proper equivalence classes. Hence, if f and g are forms of discriminant Δ and if f is positive definite and g is negative definite, then we can decide the equivalence of f and g by deciding the proper equivalence of f and gV. Also, if we find that $fU = gV$ with $U \in \mathrm{SL}(2, \mathbb{Z})$, then $f = gVU^{-1}$. Likewise, the problem of deciding the equivalence of two negative definite forms can be reduced to deciding the equivalence of two positive definite forms. Also, the minimum of a negative definite form f is the minimum of the positive definite form fV. Therefore, in this chapter we only consider positive definite forms and we deal with proper equivalence only.

5.2 Normal forms

In Section 2.4 we have seen that

$$\Gamma = \left\{ \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} : s \in \mathbb{Z} \right\}$$

is a cyclic subgroup of $\mathrm{SL}(2, \mathbb{Z})$ that is generated by

$$S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} .$$

The group Γ acts on the set of positive definite forms. We first determine an element in the Γ-orbit of f for which the absolute value of the coefficient of XY is minimal. For $s \in \mathbb{Z}$ we have

$$fS^s = \big(a, b + 2sa, f(s,1)\big) . \tag{5.1}$$

Hence, the coefficient of X^2 of any form in the Γ-orbit of f is a and the coefficient of XY of such a form is uniquely determined modulo $2a$. To explain the proper choice of s we introduce some notation.

Definition 5.2.1. *Let r be a real number. Then $\lfloor r \rfloor$ is the uniquely determined integer with*

$$0 \le r - \lfloor r \rfloor < 1 . \tag{5.2}$$

Also, $[r]$ is the uniquely determined integer with

$$-1/2 \le r - [r] < 1/2 . \tag{5.3}$$

If we choose

$$s = \left\lfloor \frac{a-b}{2a} \right\rfloor = \left[\frac{-b}{2a} \right] , \tag{5.4}$$

then we have

$$-a < b + 2sa \le a . \tag{5.5}$$

This choice of s minimizes the absolute value of b. Also, since by (1.12) and (5.1) the coefficient of Y^2 in fS^s is

$$f(s,1) = \frac{(2sa+b)^2 + |\Delta|}{4a} , \tag{5.6}$$

this choice of s also minimizes this coefficient.

Definition 5.2.2. *The form f is called* normal *if $-a < b \leq a$.*

As we have seen above, the Γ-orbit of f contains precisely one normal form that can be obtained as fS^s with s from (5.4). The normal form in the Γ-orbit of f is called the *normalization* of f. *Normalizing* f means replacing f by its normalization.

Example 5.2.3. Let

$$f(X,Y) = (195751, 1212121, 1876411)\ .$$

Then the discriminant of f is $\Delta = -3$ and the normalization of f is

$$f(X-3Y,Y) = (195751, 37615, 1807)\ .$$

5.3 Reduced forms and the reduction algorithm

We now explain an efficient reduction procedure that yields a uniquely determined reduced form in the proper equivalence class of f. In this reduction algorithm, the coefficient a is made as small as possible.

We define reduced forms.

Definition 5.3.1. *The positive definite form (a,b,c) is called* reduced *if it is normal, $a \leq c$, and if $b \geq 0$ for $a = c$.*

In the next example we show that for any integer Δ with $\Delta \equiv 0, 1 \pmod 4$ there is an integral reduced form of discriminant Δ.

Example 5.3.2. Let Δ be a negative integer, $\Delta \equiv 0, 1 \pmod 4$ and let $b = \Delta \bmod 2$. Then $f = \bigl(1, b, (b^2 - \Delta)/4\bigr)$ is a reduced form of discriminant Δ, and it is the only reduced form $(1, b, c)$ of discriminant Δ. This is shown in Exercise 5.15.2. For example, $(1,1,1)$ is a reduced form of discriminant -3, $(1,0,1)$ is a reduced form of discriminant -4, and $(1,1,2)$ is a reduced form of discriminant -7.

Definition 5.3.3. *The form f from Example 5.3.2 is called the* principal form *of discriminant Δ. Its equivalence class is called the* principal class *of discriminant Δ.*

In the reduction algorithm we use the reduction operator. It is defined now.

Definition 5.3.4. *By $\rho(f) = \rho(a,b,c)$ we denote the normalization of $(c, -b, a)$. We call ρ the* reduction operator *for positive definite forms.*

Using (5.4) it is easy to check that

$$\rho(f) = (c, -b + 2sc, cs^2 - bs + a) \tag{5.7}$$

with

$$s = s(f) = \left\lfloor \frac{c+b}{2c} \right\rfloor = \left\lceil \frac{b}{2c} \right\rceil \tag{5.8}$$

Also, if we set

$$U(f) = \begin{pmatrix} 0 & -1 \\ 1 & s(f) \end{pmatrix} \tag{5.9}$$

then

$$\rho(f) = fU(f) \ . \tag{5.10}$$

Hence, the forms f and $\rho(f)$ are properly equivalent.

Example 5.3.5. Let $f = (195751, 37615, 1807)$. Then $s(f) = 10$ and $\rho(f) = (1807, -37615 + 20 \cdot 1807, 1807 \cdot 100 - 37615 \cdot 10 + 195751) = (1807, -1475, 301)$.

The reduction algorithm is very simple. First, the form f is normalized. Then the algorithm proceeds iteratively. If f is reduced, then the algorithm returns f. Otherwise, f is replaced by $\rho(f)$. This is called a *reduction step.*

Example 5.3.6. Let $f = (195751, 37615, 1807)$. This form is normal. Applying the reduction operator, we find $\rho(f) = (1807, -1475, 301)$, $\rho^2(f) = (301, 271, 61)$, $\rho^3(f) = (61, -27, 3)$, $\rho^4(f) = (3, 3, 1)$, and $\rho^5(f) = 1, 1, 1)$. The form $(1, 1, 1)$ is reduced.

To prove that the reduction algorithm terminates we need the following lemma.

Lemma 5.3.7. *Let r be a positive real number and let x, y be real numbers. If $f(x, y) \leq r$ then $x^2 \leq 4cr/|\Delta|$ and $y^2 \leq 4ar/|\Delta|$.*

Proof. Since the discriminant Δ of f is negative, the first inequality follows from (1.14) and the second one from (1.12). □

Theorem 5.3.8. *Given a positive definite form f, the reduction algorithm for positive definite forms terminates and returns a reduced form that is properly equivalent to f.*

Proof. If the reduction algorithm terminates, then it returns a reduced form. Also, since normalization and each reduction step are transformations with matrices from $\mathrm{SL}(2, \mathbb{Z})$, that reduced form is properly equivalent to f.

Assume that $f = (a, b, c)$ is an input on which the reduction algorithm does not terminate. Denote by $f_i = (a_i, b_i, c_i)$ the form f computed in the algorithm after the ith reduction step. Then $(a_i)_{i \geq 0}$ is a strictly decreasing sequence of values of f. This is impossible since by Lemma 5.3.7 there are only finitely many real numbers which are less than a and which can be represented by f. □

Corollary 5.3.9. *Every proper equivalence class of positive definite forms contains a reduced form.*

We explain how a transformation matrix $T \in \mathrm{SL}(2, \mathbb{Z})$ can be computed such that fT is reduced. We let $T_0 \in \mathrm{SL}(2, \mathbb{Z})$ such that fT_0 is normal. Then we apply the reduction algorithm to $f_0 = fT_0$, $f_0 = (a_0, b_0, c_0)$. Let k be the number of reduction steps performed by the reduction algorithm. By $f_i = (a_i, b_i, c_i)$ we denote the form obtained after the ith reduction step and we let $s_i = s(f_i)$. In the ith reduction step we also compute the matrix

$$T_i = T_{i-1}U(f_{i-1}) \ , \quad 1 \leq i \leq k.$$

Then

$$f_i = fT_i \ , \quad 0 \leq i \leq k.$$

Hence, with $T = T_k$, the form fT is reduced. We show that the entries of T_i satisfy a rather simple recursion. If we write

$$T_i = \begin{pmatrix} p_i & p_{i+1} \\ q_i & q_{i+1} \end{pmatrix} \ , \quad 0 \leq i \leq k \ , \tag{5.11}$$

then

$$p_{i+2} = s_i p_{i+1} - p_i \ , \quad q_{i+2} = s_i q_{i+1} - q_i \ , \quad 0 \leq i \leq k - 1 \ . \tag{5.12}$$

Example 5.3.10. Let

$$f = (195751, 37615, 1807) \ .$$

This form is normal. We apply the above algorithm to f and $T_0 = I_2$.

We obtain the following table in which the form f_6 is reduced.

i	0	1	2	3	4	5	6
a_i	195751	1807	301	61	3	1	
b_i	37615	−1475	271	−27	3	1	
c_i	1807	301	61	3	1	1	
p_i	1	0	−1	2	5	−22	−49
q_i	0	1	10	−21	−52	229	510
s_i	10	−2	2	−4	2		

We conclude this section by describing the algorithms formally. Algorithm `reduce` on page 90 reduces a form and determines the corresponding transformation. It uses the Algorithms `normalize` and `rho` that implement normalization and the reduction operator.

Algorithm 5.1 `rho` (f, T)

Input: A positive definite form $f = (a, b, c)$, $T = \begin{pmatrix} t_{1,1} & t_{1,2} \\ t_{2,1} & t_{2,2} \end{pmatrix} \in \mathbb{Z}^{(2,2)}$.

Output: $\big(\rho(f), TU(f)\big)$.

```
$s \leftarrow s(f)$
return $\left((c, -b + 2sc, cs^2 - bs + a), \begin{pmatrix} t_{1,2} & t_{1,1} + st_{1,2} \\ t_{2,2} & t_{2,1} + st_{2,2} \end{pmatrix}\right)$
```

Algorithm 5.2 `normalize` (f, T)

Input: A positive definite form $f = (a, b, c)$.
Output: The normalization g of f and $U \in \Gamma$ such that $g = fU$.

```
$s \leftarrow \lfloor (a - b)/(2a) \rfloor$
return $\left((a, b + 2sa, as^2 + bs + c), \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}\right)$
```

Algorithm 5.3 `reduce` (f)

Input: A positive definite form f.
Output: A reduced form g and $T \in \mathrm{GL}(2, \mathbb{Z})$ with $fT = g$.

```
$(g, T) \leftarrow$ normalize$(f)$
while $g$ is not reduced do
    $(g, T) \leftarrow$ rho$(g, T)$
return $(g, T)$
```

5.4 Properties of reduced forms

We prove bounds for the coefficients of reduced forms.

Lemma 5.4.1. *If (a, b, c) is reduced, then $a \leq \sqrt{|\Delta|/3}$.*

Proof. Assume that (a, b, c) is reduced. Then $|\Delta| = 4ac - b^2 \geq 4a^2 - a^2 = 3a^2$ which implies that $a \leq \sqrt{|\Delta|/3}$. □

Corollary 5.4.2.
1. *There are only finitely many integral reduced forms of discriminant Δ.*
2. *There are only finitely many equivalence classes of integral forms of discriminant Δ.*

Proof. By Definition 5.3.1 and Lemma 5.4.1 we have $|b| \leq a \leq \sqrt{|\Delta|/3}$ for any positive definite reduced form (a, b, c). Also, in such a form c is uniquely determined by a, b, and Δ. This implies the first assertion. The second assertion follows from Corollary 5.3.9 and the first one. □

Using Lemma 5.4.1, we are able to determine all integral reduced forms of a given negative discriminant. In Section 5.11 we will present an algorithm for doing so. In Table 5.1 we list all reduced forms for a few small discriminants.

Δ	reduced forms
-3	$(1,1,1)$
-4	$(1,0,1)$
-7	$(1,1,2)$
-8	$(1,0,2)$
-11	$(1,1,3)$
-15	$(1,1,4)$, $(2,1,2)$
-19	$(1,1,5)$
-20	$(1,0,5)$, $(2,2,3)$
-23	$(1,1,6)$, $(2,-1,3)$, $(2,1,3)$
-24	$(1,0,6)$, $(2,0,3)$
-28	$(1,0,7)$, $(2,2,4)$
-31	$(1,1,8)$, $(2,-1,4)$, $(2,1,4)$

Table 5.1. Reduced positive definite forms of small discriminant

5.5 The number of reduction steps

In this section we prove upper bounds on the number of reduction steps performed by the reduction algorithm. We start by giving a sufficient condition for a normal form to be reduced.

Lemma 5.5.1. *If f is normal and $a < \sqrt{|\Delta|}/2$, then f is reduced.*

Proof. Let f be normal and $a < \sqrt{|\Delta|}/2$. Since $\Delta < 0$ we have

$$c = \frac{b^2 + |\Delta|}{4a} \geq \frac{|\Delta|}{4a} > \frac{a^2}{a} = a\ .$$

Thus f is reduced. □

We prove that the reduction steps make the coefficient of X^2 considerably smaller as long as $a \geq \sqrt{|\Delta|}$

Lemma 5.5.2. *If f is normal and $a \geq \sqrt{|\Delta|}$, then $c \leq a/2$.*

Proof. Let f be normal. Then we have $b^2 \leq a^2$. Hence, it follows from $a \geq \sqrt{|\Delta|}$ that

$$c = \frac{b^2 + |\Delta|}{4a} \leq \frac{a^2 + a^2}{4a} = \frac{a}{2}\ .$$

□

We show that at most one application of ρ suffices to make f reduced if $a < \sqrt{|\Delta|}$.

Lemma 5.5.3. *Let f be normal but not reduced. If $a < \sqrt{|\Delta|}$, then $\rho(f)$ is reduced.*

Proof. Assume that $a < \sqrt{|\Delta|}$. Since f is normal but not reduced we have either $a > c$ or ($a = c$ and $b < 0$). In the latter case, $\rho(f)$ is obviously reduced. So assume that $a > c$. If $c < \sqrt{|\Delta|}/2$, then $\rho(f)$ is reduced by Lemma 5.5.1. Assume that $c \geq \sqrt{|\Delta|}/2$, that is, $4c^2 \geq |\Delta|$. Since $|b| \leq a < \sqrt{\Delta}$ it follows that $b^2/(4c^2) < 1$. Hence $|s(f)| \leq 1$ by (5.8). If $s(f) = 0$ then $\rho(f) = (c, -b, a)$ which is reduced since $c < a$. Assume that $|s(f)| = 1$. Then $\operatorname{sign}(s(f)) = \operatorname{sign}(b)$ by (5.8). Now $\rho(f) = \big(c, -b + 2s(f)c, a - |b| + c\big)$. If $a > |b|$, then $a - |b| + c > c$. Hence $\rho(f)$ is reduced. If $a = b$, then $s(f) = 1$ since a is positive. So we have $\rho(f) = (c, -a + 2c, c)$. But $c \geq \sqrt{|D|}/2 > a/2$, so $-a + 2c > 0$. This shows that $\rho(f)$ is reduced. □

Theorem 5.5.4. *The number of reduction steps performed by Algorithm* `reduce` *when applied to a positive definite form $f = (a, b, c)$ is at most $\lfloor \log_2(a/\sqrt{|\Delta|}) \rfloor + 2$.*

Proof. In each reduction step, the form (a, b, c) is replaced by the normalization of $(c, -b, a)$. If $a \geq \sqrt{|\Delta|}$ and if the resulting form is (a', b', c'), then, by Lemma 5.5.2, we have $a' = c \leq a/2$. This shows that after at most $\log_2(a/\sqrt{|\Delta|}) + 1$ reduction steps the reduction algorithm finds a form (a, b, c) with $a < \sqrt{|\Delta|}$. It follows from Lemma 5.5.3 that at most one more reduction step is necessary to determine a reduced form. □

In Example 5.3.10 five reduction steps were necessary to find the reduced form equivalent to the initial form. The theoretical result, yields an upper bound of 13 steps. This shows that the bound in Theorem 5.5.4 is not sharp.

5.6 Bit complexity of the reduction algorithm

We now analyze the bit complexity of the reduction algorithm when applied to reduce an integral form. We first estimate the size of the numbers which are used in Algorithm `reduce`.

Lemma 5.6.1. *If Algorithm* `reduce` *is applied to a positive definite normal form $f = (a, b, c)$, then the entries of the transformation matrix T in Algorithm* `reduce` *are bounded by $2\max\{a, c\}/\sqrt{|\Delta|}$.*

Proof. Let $K = \max\{a, c\}$. Let $T = \begin{pmatrix} s & t \\ u & v \end{pmatrix}$ and $g = (A, B, C)$ be a matrix and a form computed in Algorithm `reduce` (f). Then $A \leq K$, $C \leq K$, and we have $g = fT$. It follows from (1.12) that

$$4aA = 4af(s,u) = (2as+bu)^2 + |\Delta|u^2.$$

This implies

$$u^2 \leq 4aA/|\Delta| \leq 4K^2/|\Delta|.$$

Since $4aC = 4af(t,v)$, the same argument yields the inequality

$$v^2 \leq 4K^2/|\Delta|.$$

Using (1.14) the same bound is proved for s^2 and t^2. □

From Theorem 5.5.4 and Lemma 5.6.1 we immediately obtain the following result.

Proposition 5.6.2. *Given an integral form (a,b,c), Algorithm* `reduce` *performs $\mathrm{O}\big(\log(a/\sqrt{|\Delta|})\big)$ arithmetic operations on numbers of binary length $\mathrm{O}(\operatorname{size} f)$.*

From Proposition 5.6.2 we can deduce that the number of bit operations required by Algorithm `reduce` is at most $\mathrm{O}((\operatorname{size} f)^3)$. But we can do even better. We now show that the bit complexity of the reduction algorithm is $\mathrm{O}((\operatorname{size} f)^2)$.

Lemma 5.6.3. *If r is a real number with $|r| > 1/2$ then $r^2 \geq \big(|[r]| - 1/2\big)^2$.*

Proof. We have $-1/2 \leq r - [r] \leq 1/2$. This shows that

$$[r] - 1/2 \leq r \leq [r] + 1/2\ . \tag{5.13}$$

If $r > 1/2$ then $[r] \geq 1$. So (5.13) implies $|r| = r \geq [r] - 1/2 > 0$. Therefore, $r^2 \geq \big(|[r]| - 1/2\big)^2$. If $r < -1/2$ then $[r] \leq -1$. So (5.13) implies $|r| = -r \geq -[r] - 1/2 = |[r]| - 1/2 > 0$. Therefore, $r^2 \geq \big(|[r]| - 1/2\big)^2$. □

Lemma 5.6.4. *If f is normal but not reduced, then $a \geq |s(f)|c$.*

Proof. Set $s = s(f)$. If $|s| \leq 1$, then the assertion holds since f is normal but not reduced. Let $|s| > 1$. Using the formula (5.8) and Lemma 5.6.3 we obtain

$$a = \frac{b^2 + |\Delta|}{4c} > \frac{b^2}{4c} = c\left(\frac{|b|}{2c}\right)^2 \geq c\left(|s| - \frac{1}{2}\right)^2 > c|s|\big(|s| - 1\big) \geq c|s|\ .$$

□

Lemma 5.6.5. *If f is normal and $|s(f)| \leq 1$, then one of the forms f, $\rho(f)$ or $\rho^2(f)$ is reduced.*

Proof. Set $s = s(f)$. Assume that $|s| \leq 1$ and that neither f nor $\rho(f)$ are reduced. Then $a > c$. Let $\rho(f) = (A,B,C)$. Then $C = cs^2 - bs + a$. Since $|s| \leq 1$ and f is normal, we have $C \geq c = A$. Since $\rho(f)$ is not reduced we have $C = A$ and $B < 0$. Hence, $\rho^2(f)$ is reduced. □

Lemma 5.6.5 tells us that in Algorithm `reduce`, possibly except for the last two reduction steps, we always have $|s(g)| \geq 2$.

Theorem 5.6.6. *The running time of Algorithm* `reduce` *when applied to a positive definite form f is* $\mathrm{O}((\operatorname{size} f)^2)$*.*

Proof. By Lemma 5.6.1 the size of all integers appearing in `reduce` is $\mathrm{O}(\operatorname{size} f)$. The first normalization step requires time $\mathrm{O}((\operatorname{size} f)^2)$. Let k be the number of reduction steps performed by `reduce`. Let $g_i = (a_i, b_i, c_i)$ be the form that is input to the i-th reduction step. The determination of $s_i = s(g_i) = \lfloor (b_i + c_i)/2c_i \rfloor$ takes time $\mathrm{O}\big(\operatorname{size}(c_i)\operatorname{size}(s_i)\big) = \mathrm{O}\big(\operatorname{size}(f)\operatorname{size}(s_i)\big)$. All other computations in the i-th reduction step take time $\mathrm{O}\big(\operatorname{size}(f)\operatorname{size}(s_i)\big)$. Lemma 5.6.5 implies $s_i \geq 2$ for $1 \leq i \leq k-2$. Hence, $\operatorname{size}(s_i) = \mathrm{O}(\log |s_i|)$ for $1 \leq i \leq k-2$ which implies that the time for the first $k-2$ reduction steps is $\mathrm{O}\Big(\operatorname{size}(f) \log \prod_{i=1}^{k-2} |s_i|\Big)$. By Lemma 5.6.4 we have

$$a_1 \geq a_2 s_1 \geq \cdots \geq a_{k-1} \prod_{i=1}^{k-2} |s_i| \geq \prod_{i=1}^{k-2} |s_i| \,.$$

Hence, the time required by the first $k-2$ reduction steps is

$$\mathrm{O}\Big(\operatorname{size}(f) \log \prod_{i=1}^{k-2} |s_i|\Big) = \mathrm{O}\big(\operatorname{size}(f)\log(a_1)\big) = \mathrm{O}((\operatorname{size} f)^2)$$

which concludes the proof. □

Theorem 5.6.6 shows that when using classical algorithms for integer arithmetic, the time for reducing positive definite forms is quadratic, just like the time for the classical integer arithmetic algorithms.

5.7 Uniqueness of reduced forms

In this section we prove that every equivalence class of positive definite forms contains exactly one reduced form. Our proof uses successive minima of forms which we introduce now.

Definition 5.7.1.

1. *The* first successive minimum *of f is $\lambda_1(f)$, the minimum of f (cf. Definition 1.9).*
2. *The second successive minimum of f is the square root of the smallest positive real number such that there are two linearly independent representations of real numbers $\leq r$ by f. It is denoted by $\lambda_2(f)$.*

Example 5.7.2. The first successive minimum of $f(X,Y) = X^2 + Y^2$ is 1 since 1 is the smallest positive real number that can be represented by f. The second successive minimum of $f(X,Y) = X^2 + Y^2$ is also 1 since $(1,0)$ and $(0,1)$ are two linearly independent representations of 1 by f.

The first successive minimum of $f(X,Y) = X^2 + 2Y^2$ is 1 since 1 is the smallest positive real number that can be represented by f. The second successive minimum of $f(X,Y) = X^2 + 2Y^2$ is $\sqrt{2}$. The reason for this is the following: $(\pm 1, 0)$ are the only representations of 1 by f. They are linearly dependent. Also, $(0,1)$ is a representation of 2 by f which is linearly independent of $(1,0)$.

Proposition 5.7.3. *The first and second successive minimum of f are invariants of the proper equivalence class of f.*

Proof. Exercise 5.15.5 □

Lemma 5.7.4. *Let $f = (a,b,c)$ be a reduced positive definite form, $(x,y) \in \mathbb{Z}^2$.*

1. *If $y \neq 0$, then $f(x,y) \geq c$ and if $|y| \geq 2$ then $f(x,y) > c$.*
2. *If $x \neq 0$, then $f(x,y) \geq a$.*
3. *If $a < c$ and $f(x,y) = a$, then $|x| = 1$ and $y = 0$.*
4. *If $a > |b|$, $f(x,y) = c$, and $y \neq 0$, then $x = 0$ and $|y| = 1$.*

Proof. 1. Let $y \neq 0$. Since $f(x,y) = f(-x,-y)$ we may assume that $y \geq 0$. If $x = 0$, then $f(x,y) = f(0,y) = cy^2 \geq c$. If $|x| \geq 1$ and $y = 1$, then

$$f(x,y) = x(ax+b) + c\,. \tag{5.14}$$

Now f is reduced and therefore $a \geq |b|$. Hence $x(ax+b) = |x||ax+b| > 0$. Therefore, (5.14) implies $f(x,y) \geq c$. If $|y| \geq 2$, then $f(x,y) \geq 4|\Delta|/(4a)$ by (1.12). Since f is reduced and therefore $\Delta \geq 3a^2 \geq 3b^2$ by Lemma 5.4.1 this implies that $f(x,y) > (b^2 + |\Delta|)/(4a) = c$.

2. Let $x \neq 0$. If $y = 0$, then $f(x,y) = ax^2 \geq a$. If $y \neq 0$, then $f(x,y) \geq c$, as we have proved in 1.

3. Let $a < c$. We have already seen in 1. that $f(x,y) \geq c > a$ for $y \neq 0$. Hence, $f(x,y) = a$ implies $y = 0$ and $|x| = 1$.

4. Let $a > |b|$. We have already seen in the proof of 1. that $f(x,y) > c$ for $|y| \geq 2$. Let $|y| = 1$ and $x \neq 0$. Then $a > |b|$ implies $f(x,y) \geq |x|\big(a|x| - |b|\big) + c > c$. Hence $f(x,y) = c$ and $y \neq 0$ implies $x = 0$ and $|y| = 1$. □

Definition 5.7.5. *The form $f = (a,b,c)$ is called* semi-reduced *if it is normal and $a \leq c$.*

If the form f is semi-reduced, then by Lemma 5.5.3 one of the forms f or $\rho(f)$ is reduced.

Theorem 5.7.6. *Let $f = (a, b, c)$ be a normal positive definite form. Then the following statements are equivalent.*

1. *f is semi-reduced.*
2. *$\lambda_1(f)^2 = a$.*
3. *$\lambda_1(f)^2 = a$ and $\lambda_2(f)^2 = c$.*

Proof. The third assertion implies the second assertion. It therefore suffices to show that the second assertion implies the first assertion and that the first assertion implies the third assertion.

Assume that $\lambda_1(f)^2 = a$. Then $c = f(0, 1) \leq a$. Hence, f is semi-reduced.

Suppose that f is semi-reduced. Then $a \leq c$. It follows from the first two assertions of Lemma 5.7.4 that $f(x, y) \geq a$ for $(x, y) \neq (0, 0)$. Since $f(1, 0) = a$, this implies that $\lambda_1(f)^2 = a$. Also, the first assertion of Lemma 5.7.4 implies that $\lambda_2(f)^2 \geq c$. Since $f(1, 0) \leq c$ and $f(0, 1) = c$, it follows that $\lambda_2(f)^2 = c$. □

Theorem 5.7.6 shows that reduction theory of positive definite forms solves the problem of computing the successive minima of positive definite forms.

We will now prove the central result of this section.

Theorem 5.7.7. *Every equivalence class of positive definite forms contains precisely one reduced form.*

Proof. Suppose that $f = (a, b, c)$ and $f' = (a', b', c')$ are properly equivalent positive definite reduced forms. Then it follows from Proposition 5.7.3 and Theorem 5.7.6 that $a = a'$ and $c = c'$ and therefore $b = \pm b'$. If $a = c$ or $a = b$, then $b, b' \geq 0$ which implies that $b = b'$. Suppose that $a < c$ and $a > |b|$. Let

$$U = \begin{pmatrix} s & t \\ u & v \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z})$$

with $fU = f'$. Then $f(s, u) = a$ by (2.3). Hence $a < c$ and the third assertion of Lemma 5.7.4 imply that $|s| = 1$ and $u = 0$. Since $\det U = 1$ this means that $sv = 1$, so $v = s$. Then the fourth assertion of Lemma 5.7.4 and $f(t, v) = c$ show that $t = 0$. So $U = \pm I_2$ and $f = f'$. □

5.8 Deciding equivalence

If we want to decide whether two positive definite forms f and f' are equivalent, we can use algorithm `isEquivalent` shown on the facing page.

Theorem 5.8.1. *Algorithm* `isEquivalent` *requires time* $\mathrm{O}\big(\max\{\operatorname{size} f, \operatorname{size} f'\}^2\big)$.

Proof. This theorem follows from Theorem 5.6.6. □

Algorithm 5.4 `isEquivalent` $(f,\ f')$

Input: Two positive definite forms f, f'
Output: $T \in \mathrm{SL}(2,\mathbb{Z})$ with $f' = fT$ if f and f' are equivalent; `nil` if f and f' are not equivalent.

```
if $\Delta(f) \neq \Delta(f')$ then return nil
$(g, U) \leftarrow$ reduce($f$)
$(g', U') \leftarrow$ reduce($f'$)
if $g = g'$ then return $T = U(U')^{-1}$
else return nil
```

5.9 Solving the representation problem

Let f be an integral primitive positive definite form. The primitive representations of n by f, that is, the primitive solutions (x, y) of the Diophantine equation

$$ax^2 + bxy + cy^2 = n \tag{5.15}$$

can be computed as follows.

1. Initially, the solution set S is empty.
2. Compute $\mathrm{Aut}(f)$ as described in Section 2.5.3.
3. Compute the prime factorization $n = \prod_{p|n} p^{e(p)}$.
4. By the method described in Section 3.6 determine $\mathcal{F}^*(\Delta, n)$.
5. From each Γ-orbit in $\mathcal{F}^*(\Delta, n)$ do the following
   a) Choose a representative g.
   b) Apply `isEquivalent` $(f,\ g)$.
   c) If `isEquivalent` (f,g) yields a matrix $U \in \mathrm{SL}(2,\mathbb{Z})$, then extract the first column (x, y) from U and set $S \leftarrow S \cup (x, y)\mathrm{Aut}(f)$.

Example 5.9.1. We determine all representations of 1125 by the form $f = (5, 9, 1)$. We have $\Delta(f) = 61$. The factorization of 1125 is

$$1125 = 3^2 \cdot 5^3.$$

We know from Example 3.6.3 that

$$\mathcal{F}^*(61, 1125) = \{(1125, \pm 581, 75)\Gamma, (1125, \pm 1831, 745)\Gamma\}.$$

Also, we know from Theorem 2.5.10 that $\mathrm{Aut}(f) = \{\pm I_2\}$. The reduction of f yields $(3, 5, -3)$ and we have

$$(3, 5, -3) = f \begin{pmatrix} -2 & 1 \\ 1 & -1 \end{pmatrix} . \tag{5.16}$$

The reduction of $(1125, 1831, 745)$ yields $(1, 7, -3)$. The reduction of $(1125, -1831, 745)$ yields $(3, 5, -3)$ and we have

$$(1125, -1831, 745) = (3, 5, -3)\begin{pmatrix} 16 & -13 \\ 21 & -17 \end{pmatrix} . \tag{5.17}$$

It follows from (5.16) and (5.17) that

$$(1125, -1831, 745) = f\begin{pmatrix} -2 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 16 & -13 \\ 21 & -17 \end{pmatrix} = f\begin{pmatrix} -11 & 9 \\ -5 & 4 \end{pmatrix} .$$

Hence, the representations of 1125 by $(5, 9, 1)$ are $\pm(11, 5)$.

5.10 Solving the minimum problem

We can now also solve the minimum problem. If we want to determine the minimum of a form f, then we apply the reduction algorithm to f. If the resulting reduced form is (a, b, c), then by Theorem 5.7.6 the first successive minimum of f is $\sqrt{a}$ and the second successive minimum of f is $\sqrt{c}$. This implies that the minimum of f is a. This algorithm also works for non-integral forms.

Example 5.10.1. Let $f = (195751, 37615, 1807)$. As we have seen in Example 5.3.6, reducing this form yields $(1, 1, 1)$. This shows that the first and the second successive minimum of f is 1.

5.11 Class number

We already know from Corollary 5.4.2 that the number of equivalence classes of integral positive definite forms of a fixed discriminant is finite. By Theorem 5.7.7, this number is exactly the number of reduced forms of discriminant Δ. For example, there is just one equivalence class of integral positive definite forms of discriminant -3. Hence all positive definite forms of that discriminant are equivalent. In this section, we describe simple methods for finding the number of those equivalence classes. We will only count classes of primitive forms since most questions about imprimitive forms can be reduced to questions about primitive forms. We let Δ be a negative integer, $\Delta \equiv 0, 1 \pmod 4$.

Definition 5.11.1. *The* class number $h(\Delta)$ *is the number of proper equivalence classes of primitive integral forms of discriminant* Δ.

By Theorem 5.7.7 the class number of Δ can be determined by counting the number of reduced forms of discriminant Δ. We explain how this can be done.

We first determine the reduced forms (a, b, c) for which $b \geq 0$. Since $b \equiv \Delta \pmod 2$, this implies $b \geq \Delta \pmod 2$. Also, since f is reduced, we must have

Algorithm 5.5 `classNumber` (Δ)

Input: A negative discriminant Δ.
Output: The class number $h(\Delta)$.

```
$h \leftarrow 0$
for ($b \leftarrow \Delta \bmod 2$, $b \leq \sqrt{|\mathit{\Delta}|/3}$, $b \leftarrow b+2$) do
  $A \leftarrow (b^2 - \Delta)/4$
  for ($a \leftarrow \max\{1, b\}$, $a \leq \sqrt{A}$, $a \leftarrow a+1$) do
    if $A \equiv 0 \pmod{a}$ then
      $c \leftarrow A/a$
      if $\gcd(a, b, c) = 1$ then
        if $b = 0$ or $a = b$ or $a = c$ then $h \leftarrow h+1$
        else $h \leftarrow h+2$
return $h$
```

$b \leq a \leq \sqrt{|\Delta|/3}$ by Lemma 5.4.1. Fix b with $\Delta \pmod{2} \leq b \leq \sqrt{|\Delta|/3}$. Suppose that a and c are integers such that (a, b, c) is a primitive reduced form of discriminant Δ. Then $c = (b^2 - \Delta)/(4a)$ which implies that a divides the integer $A = (b^2 - \Delta)/4$. Also, $a \leq c = A/a$. This shows that $a^2 \leq A$. Finally, $\gcd(a, b, A/a) = 1$. Conversely, if we find an integer a with $a \mid A$, $\max\{1, b\} \leq a$, $a^2 \leq A$, and $\gcd(a, b, A/a) = 1$, then $f = (a, b, c)$ is a reduced form of discriminant Δ.

The primitive reduced forms (a, b, c) of discriminant Δ with negative b can be obtained as the forms $(a, -b, c)$ where (a, b, c) is a primitive reduced form of discriminant Δ with $b > 0$ and $a < c$.

The ideas that were just explained, lead to Algorithm `classNumber` that given a negative discriminant Δ returns the class number $h(\Delta)$. It is easy to modify this algorithm in such a way that it also returns all reduced forms of discriminant $\mathit{\Delta}$.

We prove an upper bound on the running time of Algorithm `classNumber`.

Theorem 5.11.2. *Algorithm* `classNumber` *has running time* $\mathrm{O}(|\Delta|(\operatorname{size}\Delta)^2)$.

Proof. The number of b's, that the algorithm inspects, is $\mathrm{O}(\sqrt{|\Delta|})$. The binary length of each b is $\mathrm{O}(\log|\Delta|)$. For each b, the algorithm inspects $\mathrm{O}(\sqrt{|\Delta|})$ many a's. The binary length of each a is $\mathrm{O}(\log|\Delta|)$. The number of arithmetic operations that the algorithm performs for each pair (a, b) is $\mathrm{O}(1)$. This proves the assertion. □

The next example demonstrates how the algorithm `classNumber` works.

Example 5.11.3. Let $\Delta = -191$. Since $\Delta \equiv 1 \pmod{2}$ we start with $b = 1$ and $A = 48$. We now have to inspect all values for a with $1 \leq a \leq 6$. For $a = 1$ we $c = 48$ and the reduced form $(1, 1, 48)$. For $a = 2$ we $c = 24$ and the reduced forms $(2, 1, 24)$ and $(2, -1, 24)$. For $a = 3$ we $c = 16$ and the reduced forms $(3, 1, 16)$ and $(3, -1, 16)$. For $a = 4$ we $c = 12$ and the reduced forms

Algorithm 5.6 `classNumberList` (D)

Input: A positive integer D.
Output: An array h such that $h[-\Delta]$ is the class number of Δ for any negative discriminant Δ.

```
for ($n \leftarrow 3$, $n \le D$, $n \leftarrow n+1$) do
  $h[n] \leftarrow 0$
for ($a \leftarrow 1$, $a \le \sqrt{D/3}$, $a \leftarrow a+1$) do
  for ($b \leftarrow 0$, $b \le a$, $b \leftarrow b+1$) do
    for ($c \leftarrow a$, $c \le (D+b^2)/(4a)$, $c \leftarrow c+1$) do
      if $\gcd(a,b,c) = 1$ then
        $n \leftarrow 4ac - b^2$
        if $a = b$ or $a = c$ or $b = 0$ then $h[n] \leftarrow h[n]+1$
        else $h[n] \leftarrow h[n]+2$
return $h$
```

$(4,1,12)$ and $(4,-1,12)$. For $a=5$ we find no form. For $a=6$ we $c=8$ and the reduced forms $(6,1,8)$ and $(6,-1,8)$. For $b=3$ we find the forms $(5,3,10)$ and $(5,-3,10)$. For $b=5$ we find the forms $(6,5,9)$ and $(6,-5,9)$. Also, there is no form with $b=7$. Hence, the class number of -191 is 13.

Algorithm `classNumber` is not optimal since the class number of Δ is of the order of magnitude $\sqrt{|\Delta|}$. However, since the O-constant in Theorem 5.11.2 is small (see Exercise 3.7.3), that algorithm works efficiently for small values of Δ.

Now suppose that we want to find the class numbers for all quadratic discriminants Δ with $-D < \Delta < 0$, for some bound $D > 0$. If we apply `classNumber`, then we must spend roughly D^2 arithmetic operations. But using `classNumber` for each Δ means considering each pair (a,b) many times. A much faster way is to determine all reduced forms of all discriminants under consideration simultaneously. For each appropriate form (a,b,c) we check whether it yields a primitive reduced form. If it does, we increment the class number of the corresponding discriminant. This is what happens in Algorithm `classNumberList`. Its complexity is estimated in the next Theorem.

Theorem 5.11.4. *Algorithm* `classNumberList` *has running time* $\mathrm{O}(D^{3/2}$ $(\operatorname{size}\Delta)^2)$.

Proof. For each triple (a,b,c) Algorithm `classNumberList` performs $\mathrm{O}(1)$ arithmetic operations on numbers of binary length $\mathrm{O}(\log D)$. For each pair (a,b) at most $(D+b^2)/(4a) \le 4D/(12a)$ values of c are considered. For each a the algorithm uses $a+1$ values of b. Since the algorithm uses $\mathrm{O}(\sqrt{D})$ values of a, the total number of arithmetic operations is $\mathrm{O}(D^{3/2})$. □

Theorem 5.11.4 implies that Algorithm `classNumberList` spends on average $\mathrm{O}(D^{1/2})$ elementary operations per discriminant which is much less than Algorithm `classNumber`. Also, Algorithm `classNumberList` can be modified such that the algorithm also outputs the list of all reduced forms for each discriminant. If this output is required, then `classNumberList` is optimal since the number of reduced forms that the algorithm outputs is approximately $D^{3/2}$.

The class numbers of a few small discriminants are shown in Table 5.2.

Δ	-3	-4	-7	-8	-11	-15	-19	-20	-23	-24	-28	-31
$h(\Delta)$	1	1	1	1	1	2	1	2	3	2	1	3

Table 5.2. Class numbers

5.12 Reduction of semidefinite forms

In this section we apply the reduction algorithm for positive definite forms to a form $f = (a, b, c)$ that is positive semidefinite but not positive definite. By Proposition 1.2.10 this means that $\Delta(f) = 0$ and $a > 0$ or $c > 0$. For the case where f is an integral form, we will show that the reduction algorithm is a version of the Euclidean algorithm.

Since $\Delta(f) = 0$, it follows from (1.12) and (1.14) that there are real numbers d, e with

$$d = \sqrt{a}\ , \quad e = \sqrt{c}\ , \quad b = 2de.$$

such that

$$f(X, Y) = (dX + eY)^2\ . \tag{5.18}$$

Also, we have

$$b^2 = 4ac. \tag{5.19}$$

If f is a primitive integral form, then a and c are coprime. Hence, both a and c are squares in $\mathbb{Z}$. It follows that d and e are integers. Using (5.8) we obtain

$$s(f) = \left[\frac{b}{2c}\right] = \left[\frac{2de}{2e^2}\right] = \left[\frac{d}{e}\right]\ .$$

Now

$$\rho(f) = \Big(d(-Y) + e\big(X + s(f)Y\big)\Big)^2 = \big(eX + Y(-d + e[d/e])\big)\ .$$

So an application of the reduction operator ρ replaces d by e and e by the representative e' of the residue class of $-d$ mod e with $-e/2 < e' \leq e/2$. This is a version of the Euclidean algorithm which is used to solve linear Diophantine equations. If $e \neq 0$ and d/e is a rational number, then the coefficient of Y is eventually zero and the coefficient of X^2 is $\gcd(d, e)^2$. But if d/e is irrational, then the sequence of absolute values of the coefficients of Y^2 is a strictly decreasing sequence. Also, we always have $|e| \leq |d/2|$ and therefore $c \leq a/4$. Therefore, the reduction algorithm never terminates. From those considerations the following Theorem can be deduced.

Theorem 5.12.1. *Let f be a semidefinite form. Then f can be written as $f(X,Y) = (dX + eY)^2$ with real numbers d, e. Also, the following are true.*

1. *If $e = 0$ or d/e is a rational number, then f is equivalent to a uniquely determined form $(g, 0, 0)$. This form is computed by the reduction algorithm. If d and e are integers, then $|g| = \gcd(d, e)$.*
2. *If $e \neq 0$ and d/e is irrational, then the reduction algorithm does not terminate.*

Proof. Exercise 5.15.9. □

Example 5.12.2. Consider the form $f(X,Y) = (2X + 3Y)^2 = 4X^2 + 12XY + 9Y^2$. Then $\rho(f) = (9, 6, 1)$, $\rho^2(f) = (1, 0, 0) = \big(\gcd(2,3), 0, 0\big)$.

5.13 Geometry of reduction

We use the results of Chapter 4 to give a geometric interpretation of reduction theory for positive definite forms.

5.13.1 Reduced points

In Definition 4.3.1 we have introduced a correspondence between points in the upper half plane and forms and we have shown in Theorem 4.3.5 that those forms are positive definite.

Definition 5.13.1. *A point θ in the upper half plane is called reduced if the form f_θ is reduced.*

Here is a characterization of reduced points in the upper half plane.

Lemma 5.13.2. *An element $\theta \in \mathcal{U}$ is* reduced *if and only if it is normal, $\mathrm{N}(\theta) \geq 1$, and if $\Re\theta \geq 0$ for $\mathrm{N}(\theta) = 1$.*

Proof. Exercise 5.15.8. □

It follows from Theorem 5.7.7 that the set of reduced points in the upper half plane as shown in Figure 5.1 is a fundamental domain under the action of $\mathrm{SL}(2, \mathbb{Z})$ on the upper half plane.

We discuss the geometry of reduction. The reduction operator sends $\theta \in \mathcal{U}$ to

$$\rho(\theta) = \frac{-1}{\theta} + s(\theta)\ , \quad s(\theta) = \left[\Re\left(\frac{1}{\theta}\right)\right]\ . \tag{5.20}$$

This is shown in Figure 5.2.

The reduction algorithm transforms a point $\theta \in \mathcal{U}$ into a properly equivalent reduced point by first normalizing it and then replacing θ by $\rho(\theta)$ until a reduced point is found.

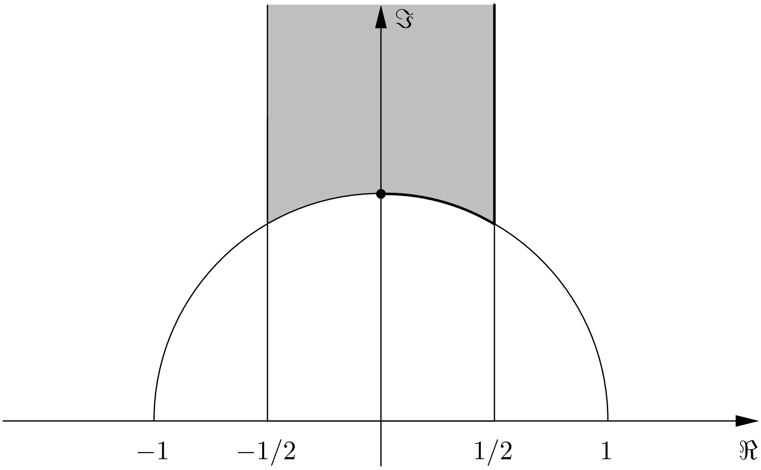

Fig. 5.1. The reduced points

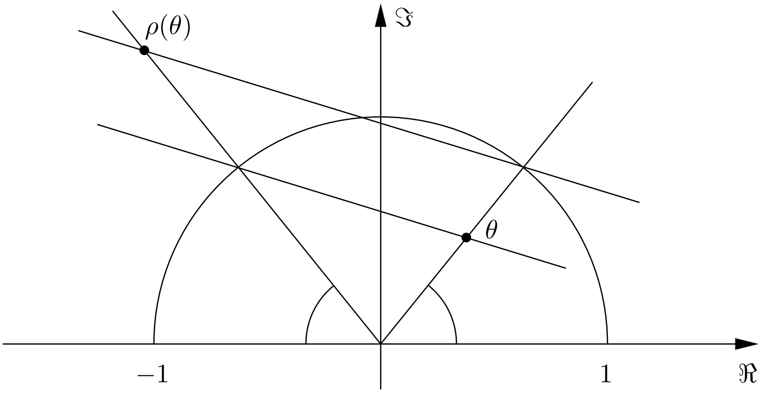

Fig. 5.2. Geometric interpretation of the ρ-operator

Example 5.13.3. The point $\theta = (37615 + \sqrt{-3})/(2 \cdot 195751)$ can be reduced using 6 applications of the reduction operator. In the course of the reduction procedure we obtain the numbers $\theta_i = (b_i + \sqrt{-3})(2a_i)$ with a_i and b_i as in Example 5.3.10.

5.14 The densest two-dimensional lattice packing

In Section 1.4.2 we have described the problem of finding the densest lattice packing. In this section we will solve this problem for two-dimensional lattices. We need to find the maximum of

$$\gamma(f) = \frac{2\lambda_1(f)^2}{\sqrt{|\Delta(f)|}} \tag{5.21}$$

where f ranges over all positive define binary quadratic forms with real coefficients. If f is a form for which $\gamma(f)$ is maximum, then the lattice $L(f)$ admits a densest lattice packing.

Theorem 5.14.1. *If f is a positive definite reduced form with real coefficients then $\gamma(f)^2 \geq 4/3$. Also, we have $\gamma(f)^2 = 4/3$ if and only if $f = r(1,1,1)$ with $r \in \mathbb{R}_{>0}$.*

Proof. Let $f = (a, b, c)$ be a positive definite binary quadratic form with real coefficients. It follows from Lemma 5.4.1 that

$$\gamma(f)^2 \leq 4a^2/|\Delta| \leq 4/3.$$

Now assume that $a^2/|\Delta| = 1/3$. Then

$$a^2 = -\Delta/3 = (4ac - b^2)/3 \geq 4a^2 - a^2)/3 = a^2.$$

It follows that $a = b = c$. Conversely, if $a = b = c$, then $\gamma(f)^2 = 4/3$. □

It follows from Theorem 4.3.4 that up to rotation and scaling with positive real numbers, the lattice that admits the densest two-dimensional packing is

$$L = L(1,1,1) = \mathbb{Z} + \mathbb{Z}\frac{1+\sqrt{-3}}{2}.$$

This is the hexagonal lattice.

5.15 Exercises

Exercise 5.15.1. Determine all integral normal forms of all negative integral discriminants greater than -12.

Exercise 5.15.2. Let Δ be a negative integer, $\Delta \equiv 0, 1 \pmod 4$, and let $b = \Delta \bmod 2$. Show that $f = \big(1, b, (b^2 - \Delta)/4\big)$ is a reduced form of discriminant Δ and that it is the only reduced form $(1, b, c)$ of discriminant Δ.

Exercise 5.15.3. Apply the reduction algorithm to the form $f = (33824333, 889961, 5854)$ and compute the corresponding transformation.

Exercise 5.15.4. Determine all solutions of the Diophantine equation $1260895X^2 + 178438XY + 6313Y^2 = 25$.

Exercise 5.15.5. Prove Proposition 5.7.3.

Exercise 5.15.6. Apply the reduction algorithm to $(7X + 5Y)^2$.

Exercise 5.15.7. Let Δ be a negative discriminant. Show that $h(\Delta) = 1$ if and only if every integral primitive form of discriminant Δ represents 1.

Exercise 5.15.8. Prove Lemma 5.13.2.

Exercise 5.15.9. Prove Theorem 5.12.1.

Chapter references and further reading

[BB99] Ingrid Biehl and Johannes Buchmann, *An analysis of the reduction algorithms for binary quadratic forms*, Voronoi's Impact on Modern Science, Institute of Mathematics Kyiv (Peter Engel and Halyna M. Syta, eds.), National Academy of Sciences of Ukraine, 1999, pp. 71–98.

[JSW06] Michael J. Jacobson, Jr., Reginald E. Sawilla, and Hugh C. Williams, *Efficient ideal reduction in quadratic fields.*, International Journal of Mathematics and Computer Science **1** (2006), no. 1, 83–116 (English).

[Lag80] Jeffrey C. Lagarias, *Worst-case complexity bounds for algorithms in the theory of integral quadratic forms*, Journal of Algorithms **1** (1980), 142–186.

[Sch91] Arnold Schönhage, *Fast reduction and composition of binary quadratic forms*, International Symposium on Symbolic and Algebraic Computation, ISSAC '91 (Stephen M. Watt, ed.), ACM Press, 1991, pp. 128–133.

6

Reduction of Indefinite Forms

In this chapter we explain reduction theory for indefinite forms which is quite different from reduction theory for positive definite forms. Reduced indefinite forms can only be used to decide equivalence of integral indefinite forms and the decision algorithm is much less efficient than in the positive definite case since reduction is no longer unique. Reduction theory also solves the minimum problem for integral indefinite forms.

We assume that Δ is a positive real number and that $f = (a, b, c)$ is an indefinite form of discriminant Δ. We also assume that f is irrational. This makes the reduction theory much easier to explain. If f is an integral form irrationality means that Δ is not a square in $\mathbb{Z}$ (see Theorem 1.3.1).

6.1 Normal forms

As in the case of positive definite forms we start by fixing a representative in the Γ-orbit of f which we call normal.

Definition 6.1.1. *The form f is called* normal *if*

$$-|a| < b \leq |a| \quad \textit{for} \quad |a| \geq \sqrt{\Delta}\,,$$

$$\sqrt{\Delta} - 2|a| < b < \sqrt{\Delta} \quad \textit{for} \quad |a| < \sqrt{\Delta}\,.$$

Example 6.1.2. Consider the form $(a, b, c) = (5, 5, 1)$. Its discriminant is 5. The form f is normal since $|a| = 5 > \sqrt{5}$ and $-|a| = -5 < 5 = b \leq 5 = |a|$.

Example 6.1.3. Consider the form $(a, b, c) = (-3, 5, 4)$. Its discriminant is 73. The form f is normal since $|a| = 3 \leq \lfloor\sqrt{73}\rfloor = 8$ and $\sqrt{73} - 6 < b = 5 < \sqrt{73}$.

We explain, how f can be transformed into a normal form. Define

$$s = \begin{cases} \operatorname{sign}(a) \left\lceil \dfrac{-b}{2|a|} \right\rceil = \operatorname{sign}(a) \left\lfloor \dfrac{|a| - b}{2|a|} \right\rfloor & \text{for } |a| \geq \sqrt{\Delta}\,, \\ \\ \operatorname{sign}(a) \left\lfloor \dfrac{\sqrt{\Delta} - b}{2|a|} \right\rfloor & \text{for } |a| < \sqrt{\Delta}\,. \end{cases} \tag{6.1}$$

Recall that

$$S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

is a generator of the subgroup Γ of $\mathrm{SL}(2, \mathbb{Z})$. Then it is easy to verify that

$$fS^s = \bigl(a, b + 2sa, as^2 + bs + c\bigr)$$

is normal. That form is called the *normalization* of f. *Normalizing* f means replacing f by its normalization.

Example 6.1.4. Consider the form $(a, b, c) = (5, 7, 1)$. Its discriminant is 29. It is not normal. Its normalization is $(a, b + 2sa, as^2 + bs + c) = (5, -3, -1)$ with $s = \operatorname{sign}(a) \bigl\lfloor (\sqrt{\Delta} - b)/(2|a|) \bigr\rfloor = \lfloor (\sqrt{29} - 7)/10 \rfloor = -1$.

If $|a| < \sqrt{\Delta}$, then the computation of the normalizing integer s requires to determine $\sqrt{\Delta}$ to a certain precision. We show that for integral forms we only need to know $\lfloor \sqrt{\Delta} \rfloor$.

Lemma 6.1.5. *Let r be a real number and let n, m be integers, $n > 0$. Then $\bigl\lfloor (r - m)/n \bigr\rfloor = \bigl\lfloor (\lfloor r \rfloor - m)/n \bigr\rfloor$ and $\bigl\lceil (r - m)/n \bigr\rceil = \bigl\lceil (\lceil r \rceil - m)/n \bigr\rceil$.*

Proof. Let $l = \lfloor (r - m)/n \rfloor$. Clearly

$$l \geq \left\lfloor \frac{\lfloor r \rfloor - m}{n} \right\rfloor . \tag{6.2}$$

Moreover, we have $0 \leq (r - m)/n - l = (1/n)\bigl(r - (m + ln)\bigr)$. This shows that $m + ln \leq \lfloor r \rfloor$ or

$$l \leq \left\lfloor \frac{\lfloor r \rfloor - m}{n} \right\rfloor . \tag{6.3}$$

Equations (6.2) and (6.3) prove the first assertion. The proof of the second assertion is Exercise 6.18.1. □

It follows from Lemma 6.1.5 that for an integral form $f = (a, b, c)$ with $|a| < \sqrt{\Delta}$ the normalizing integer is

$$s = \left\lfloor \frac{\lfloor \sqrt{\Delta} \rfloor - b}{2|a|} \right\rfloor . \tag{6.4}$$

This is very important, since in many computations the discriminant Δ is a fixed quantity. So $\lfloor \sqrt{\Delta} \rfloor$ can be computed once and for all.

6.2 Reduced forms

Next, we introduce reduced forms.

Definition 6.2.1. *The form f is called* reduced *if* $\left|\sqrt{\Delta} - 2|a|\right| < b < \sqrt{\Delta}$.

Example 6.2.2. The form $f = (5, 3, -1)$ is not reduced. The discriminant of f is 29 and we have $\left|\sqrt{\Delta} - 2|a|\right| = 4.61.. > b$. The form $f = (1, 5, -1)$ is reduced since $\left|\sqrt{\Delta} - 2|a|\right| = 3.38.. < b = 5 < \sqrt{\Delta}$.

From the definition of a reduced form we obtain the following consequence.

Lemma 6.2.3. *If (a, b, c) is reduced then $(-a, b, -c)$ is also reduced.* □

Example 6.2.4. Let $\Delta \in \mathbb{Z}$. If Δ is odd, then set b to the greatest odd integer less than $\sqrt{\Delta}$. If Δ is even, then set b to the greatest even integer less than $\sqrt{\Delta}$. We claim that $f = \left(1, b, (b^2 - \Delta)/4\right)$ is a reduced form of discriminant Δ. It is easy to verify that $\Delta(f) = \Delta$. To verify that f is reduced, we note that $\Delta \geq 5$. Hence $2a = 2 < \sqrt{\Delta}$. This means that the reduction condition is $\sqrt{\Delta} - 2 < b < \sqrt{\Delta}$. So the above choice of b makes f reduced. Note that f is the only integral reduced form $(1, b, c)$ of discriminant Δ.

Definition 6.2.5. *The form f from Example 6.2.4 is called the* principal form *of discriminant Δ. Its equivalence class is called the* principal class *of discriminant Δ.*

We have seen above that an integral form $f = (a, b, c)$ of discriminant Δ can be normalized using an algorithm which obtains a, b and $\lfloor\sqrt{\Delta}\rfloor$ as inputs. We show in the next lemma that the same input suffices to check whether an integral indefinite form is reduced.

Lemma 6.2.6. *If f is integral and normal, then f is reduced if and only if* $2|a| - b \leq \lfloor\sqrt{\Delta}\rfloor$

Proof. Assume that f is reduced. Then $\left|\sqrt{\Delta} - 2|a|\right| < b < \sqrt{\Delta}$. Hence, we have

$$\sqrt{\Delta} - 2|a| < b < \sqrt{\Delta} \tag{6.5}$$

and

$$2|a| - \sqrt{\Delta} < b < \sqrt{\Delta}\,. \tag{6.6}$$

It follows from (6.6) that $|a| < \sqrt{\Delta}$. So (6.5) implies that f is normal.

Suppose that f is normal and

$$2|a| - b \leq \lfloor\sqrt{\Delta}\rfloor. \tag{6.7}$$

We first show that $|a| < \sqrt{\Delta}$. Assume that $|a| \geq \sqrt{\Delta}$. Since f is normal and Δ is not a square in $\mathbb{Z}$, we have $|a| = 2|a| - |a| \leq 2|a| - b < \sqrt{\Delta}$. This is a contradiction. Therefore, $|a| < \sqrt{\Delta}$. The normalization condition for that case is $\sqrt{\Delta} - 2|a| < b < \sqrt{\Delta}$. Also, (6.7) implies $b > 2|a| - \sqrt{\Delta}$. Hence, we have $\left|\sqrt{\Delta} - 2|a|\right| < b < \sqrt{\Delta}$ which means that f is reduced. □

We show that the coefficients of reduced forms are bounded.

Lemma 6.2.7. *If f is reduced, then $|a|+|c|<\sqrt{\Delta}$ and a and c have opposite sign.*

Proof. Since $0<b<\sqrt{\Delta}$, it follows that $4ac=b^2-\Delta<0$. This shows that a and c are of opposite sign. Also, we have $4|a|\bigl(|a|+|c|-\sqrt{\Delta}\bigr)=4|a|^2-4|a|\sqrt{\Delta}+\Delta-b^2=\bigl(2|a|-\sqrt{\Delta}\bigr)^2-b^2<0$, where the last inequality follows from $\bigl|2|a|-\sqrt{\Delta}\bigr|<b$. Hence $|a|+|c|-\sqrt{\Delta}<0$ as asserted. □

Corollary 6.2.8. *There are only finitely many integral reduced forms of a fixed discriminant.*

6.3 Another characterization of reduced forms

In this section we give a geometric characterization of reduced forms and show that if (a,b,c) is reduced then (c,b,a) is also reduced. In Definition 4.3.14 we have introduced the point

$$\theta(f)=\bigl(\theta_1(f),\theta_2(f)\bigr)=\bigl(\frac{b+\sqrt{\Delta}}{2a},\frac{b-\sqrt{\Delta}}{2a}\bigr)\ . \tag{6.8}$$

We know that

$$f(X,Y)=a\bigl(X+\theta_1(f)Y\bigr)\bigl(X+\theta_2(f)Y\bigr)\ . \tag{6.9}$$

Therefore, $\theta_1(f)$ and $\theta_2(f)$ are the real zeros of the polynomial $f(X,-1)$. Also, we have

$$\theta_1(f)+\theta_2(f)=b/a\ ,\quad \theta_1(f)-\theta_2(f)=\sqrt{\Delta}/a\ ,\quad \theta_1(f)\theta_2(f)=c/a\ . \tag{6.10}$$

Lemma 6.3.1. *The form $f=(a,b,c)$ is normal if and only if either*

1. $|\theta_1-\theta_2|>1$ *and* $0<\operatorname{sign}(\theta_2(f)-\theta_1(f))\,\theta_2(f)<1$, *or*
2. $|\theta_1-\theta_2|<1$ *and* $-1<\operatorname{sign}(\theta_1(f)-\theta_2(f))(\theta_1(f)+\theta_2(f))\le 1$.

Proof. By (6.10), $|\theta_1(f)-\theta_2(f)|>1$ if and only if $|a|<\sqrt{\Delta}$. The form f is normal in this case if and only if

$$\sqrt{\Delta}-2|a|<b<\sqrt{\Delta}\quad\text{or, equivalently,}\quad 0<-\operatorname{sign}(a)\frac{b-\sqrt{\Delta}}{2a}<1\ .$$

The last condition is equivalent to $0<\operatorname{sign}(\theta_2(f)-\theta_1(f))\,\theta_2(f)<1$ since $\operatorname{sign}(\theta_1(f)-\theta_2(f))=\operatorname{sign}(a)$.

Analogously, $|\theta_1(f)-\theta_2(f)|<1$ if and only if $|a|>\sqrt{\Delta}$. The form f is normal in this case if and only if

$$-|a|<b\le|a|\quad\text{or, equivalently}\quad -1<\operatorname{sign}(a)\frac{b}{a}\le 1\ .$$

Using again (6.10), we see that this is equivalent to $-1<\operatorname{sign}(\theta_1(f)-\theta_2(f))(\theta_1(f)+\theta_2(f))\le 1$. □

Lemma 6.3.2. *The form $f = (a, b, c)$ is reduced if and only if $|\theta_1(f)| > 1$, $|\theta_2(f)| < 1$, and $\theta_1(f)\theta_2(f) < 0$.*

Proof. Suppose that f is reduced. Then $0 < b < \sqrt{\Delta}$ and $-b < \sqrt{\Delta} - 2|a| < b$. This implies $|\theta_1(f)| = (b + \sqrt{\Delta})/2|a| > 1$ and $|\theta_2(f)| = (\sqrt{\Delta} - b)/2|a| < 1$. Also, Lemma 6.2.7 implies that $\theta_1(f)\theta_2(f) = c/a < 0$.

Conversely, assume that the inequalities in the Lemma hold. Since $c/a = \theta_1(f)\theta_2(f) < 0$, it follows that $\Delta = b^2 + 4|a||c|$. Hence $|b| < \sqrt{\Delta}$. So $|\theta_1(f)| = (b + \sqrt{\Delta})/(2|a|)$ and $|\theta_2(f)| = (\sqrt{\Delta} - b)/(2|a|)$. So $|\theta_1(f)| > 1$ and $|\theta_2(f)| < 1$ implies that $\left|\sqrt{\Delta} - 2|a|\right| < b$. □

Lemmas 6.3.1 and 6.3.2 are depicted in Figure 6.1.

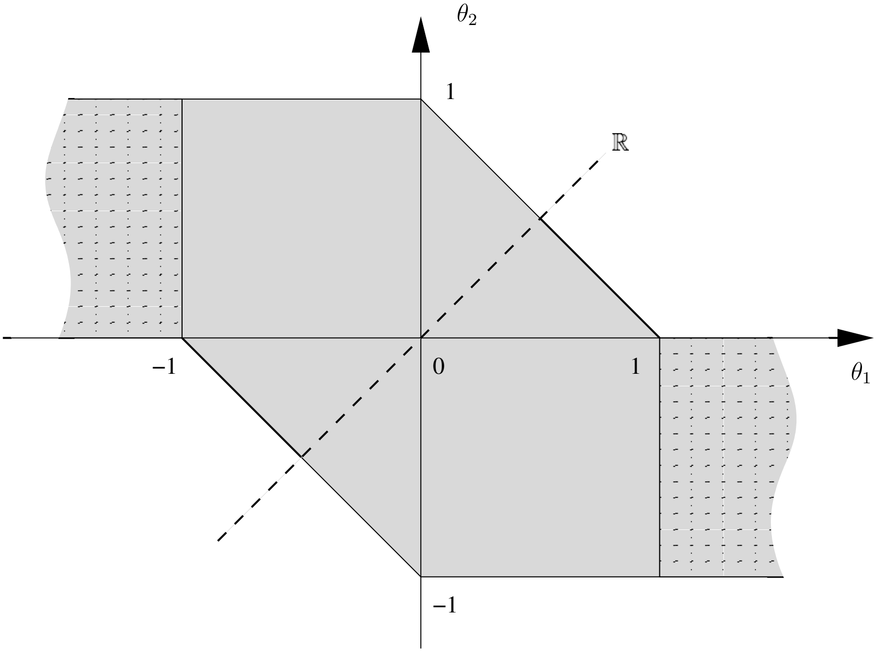

Fig. 6.1. Points corresponding to normal ■ and reduced ▨ forms

Corollary 6.3.3. *If f is normal, but not reduced, then $|\theta_1(f)|, |\theta_2(f)| < 1$.* □

Corollary 6.3.4. *If the form $f = (a, b, c)$ is reduced, then the form (c, b, a) is also reduced.*

Proof. We have $\theta_1(c, b, a) = 1/\theta_2(a, b, c)$ and $\theta_2(c, b, a) = 1/\theta_1(a, b, c)$. Hence, Lemma 6.3.2 implies the assertion. □

6.4 The reduction algorithm

In this section, we describe an algorithm for computing a reduced form in the proper equivalence class of a given indefinite form. As in the case of positive definite forms, we denote by $\rho(f)$ the normalization of the form $(c, -b, a)$. To be more explicit, we set

$$\rho(f) = \big(c, -b + 2sc, cs^2 - bs + a\big) \tag{6.11}$$

with

$$s = s(f) = \begin{cases} \operatorname{sign}(c) \left\lfloor \dfrac{b}{2|c|} \right\rfloor = \operatorname{sign}(c) \left\lfloor \dfrac{|c| + b}{2|c|} \right\rfloor & \text{for } |c| \geq \sqrt{\Delta}\,, \\ \\ \operatorname{sign}(c) \left\lfloor \dfrac{\sqrt{\Delta} + b}{2|c|} \right\rfloor & \text{for } |c| < \sqrt{\Delta}\,. \end{cases} \tag{6.12}$$

For integral forms, we can use (6.4) and replace $\sqrt{\Delta}$ by $\lfloor\sqrt{\Delta}\rfloor$. We also define

$$U(f) = \begin{pmatrix} 0 & -1 \\ 1 & s(f) \end{pmatrix}\,. \tag{6.13}$$

Then

$$\rho(f) = fU(f)\,.$$

Example 6.4.1. Consider the form $f = (5, -3, -1)$. Its discriminant is 29. It is not reduced since b is negative. We apply the reduction operator to that form. We have $s = s(f) = -\lfloor(5-3)/2\rfloor = -1$. Hence, $\rho(f) = (-1, 3+2, 5-3-1) = (-1, 5, 1)$. This form is reduced since $\sqrt{29} - 2 < 5 < \sqrt{29}$. It is easily checked that one more application of ρ yields the reduced form $(1, 5, -1)$. Also, if we apply ρ once more, then we obtain the form $(-1, 5, 1)$ again. This shows that proper equivalence classes of indefinite forms may contain more than one reduced from.

The reduction algorithm `reduce`(f) from Section 5.3 can also be used here. In the next section we prove correctness and termination of this algorithm.

Example 6.4.2. Consider the form $f = (-1360889, -747003, -102509)$. Its discriminant is $\Delta(f) = 5$. We apply the reduction algorithm to that form. Using the notation from Example 5.3.10 we obtain the following table.

i	0	1	2	3	4	5	6
a_i	-1360889	-102509	-13021	-491	-11	-1	
b_i	-747003	-73069	-5057	147	7	1	
c_i	-102509	-13021	-491	-11	-1	1	
p_i	1	0	-1	-3	-14	101	-390
q_i	0	1	4	11	51	-368	1421
s_i	4	3	5	-7	-4		

The form $(-1, 1, 1)$ is reduced.

6.5 The number of reduction steps

In this section, we prove correctness and termination of the reduction algorithm `reduce` by proving an upper bound on the number of reduction steps. We start by giving a sufficient condition for f being reduced.

Lemma 6.5.1. *If f is normal and $|a| \leq \sqrt{\Delta}/2$, then f is reduced.*

Proof. If f is normal and $|a| \leq \sqrt{\Delta}/2$, then we have $\left|\sqrt{\Delta} - 2|a|\right| = \sqrt{\Delta} - 2|a| < b < \sqrt{\Delta}$. □

Next we show that for $|a| \geq \sqrt{\Delta}$ each reduction step reduces the size of $|a|$ by a factor at least $1/4$. Also, if $|a| < \sqrt{\Delta}$, then at most one reduction step makes f reduced.

Lemma 6.5.2. *Let f be normal.*

1. If $|a| \geq \sqrt{\Delta}$, then $|c| \leq |a|/4$.
2. If $|a| < \sqrt{\Delta}$, then $\rho(f)$ is reduced.

Proof. Assume that $|a| \geq \sqrt{\Delta}$. Then we have $\Delta \leq a^2$ and $b^2 \leq a^2$ since f is normal. This implies

$$|c| = \frac{\Delta - b^2}{4|a|} \leq \frac{a^2}{4|a|} = \frac{|a|}{4} .$$

Let $|a| < \sqrt{\Delta}$. If $|c| \leq \sqrt{\Delta}/2$ then $\rho(f)$ is reduced by Lemma 6.5.1. So suppose that $|c| > \sqrt{\Delta}/2$. We have

$$-\sqrt{\Delta} < \sqrt{\Delta} - 2|a| < b < \sqrt{\Delta}$$

since f is normal and $|a| < \sqrt{\Delta}$. This implies that

$$0 < \sqrt{\Delta} - b < 2|a|, \ \sqrt{\Delta} + b > 0.$$

Hence,

$$\frac{\sqrt{\Delta}+b}{2|c|} = \frac{2|a|}{\sqrt{\Delta}-b} > 1$$

and therefore

$$-b + 2|c| < \sqrt{\Delta}. \tag{6.14}$$

It follows from $|b| < \sqrt{\Delta}$ and (6.14) imply

$$|c| < \sqrt{\Delta}.$$

On the other hand, $|b| < \sqrt{\Delta}$ and $|c| > \sqrt{\Delta}/2$ imply

$$-b + 2|c| > 2|c| - \sqrt{\Delta} = \left|\sqrt{\Delta} - 2|c|\right|. \tag{6.15}$$

From (6.14) and (6.15) we see that $s(f) = \operatorname{sign}(c)$ and $\rho(f) = \bigl(c, -b+2|c|, a - \operatorname{sign}(c)b + c\bigr)$ is reduced. □

Lemma 6.5.2 enables us to prove an upper bound on the number of reduction steps performed by Algorithm `reduce` when applied to an indefinite form. This is done in the following proposition.

Theorem 6.5.3. *Given an indefinite form f, Algorithm* `reduce` *terminates with a correct result after at most $\frac{1}{2}\log_2(|a|/\sqrt{\Delta}) + 2$ reduction steps.*

Proof. In each reduction step, the form (a, b, c) is replaced by the normalization of $(c, -b, a)$. If $a \geq \sqrt{\Delta}$ and if the resulting form is (a', b', c'), then, by Lemma 6.5.2, we have $a' = c \leq a/4$. This shows that after at most $(1/2)\log_2(a/\sqrt{|\Delta|}) + 1$ reduction steps the reduction algorithm finds a form (a, b, c) with $a < \sqrt{|\Delta|}$. It follows from Lemma 6.5.2 that at most one more reduction step is necessary to determine a reduced form. □

Theorem 6.5.3 proves that `reduce` terminates. A consequence is the following result.

Corollary 6.5.4. *Each proper equivalence class of indefinite forms contains a reduced form.* □

Corollary 6.5.5. *If n is the smallest positive real number that can be represented by f, then there is a reduced form (n, B, C) in the equivalence class of f.*

Proof. Exercise 6.18.8. □

We obtain another interesting consequence of Lemma 6.5.2

Lemma 6.5.6. *If f is reduced, then $\rho(f)$ is reduced.*

Proof. If f is reduced, then $|a| < \sqrt{\Delta}$ by Lemma 6.2.7. Hence, $\rho(f)$ is reduced by Lemma 6.5.2. □

Note that in general, this statement is not true for positive definite forms. We will show later that ρ is a permutation of the set of all reduced indefinite forms.

6.6 Complexity of reducing integral forms

In this section we prove that the reduction algorithms from Sections 5.3 and 6.4 reduce an integral form f in time quadratic in the size of the coefficients of f. Note that in this section f can be either positive definite or indefinite.

6.6.1 Sizes

Assume that the reduction algorithm is applied to the normal form $f = (a, b, c)$. We let k be the number of reduction steps performed by the reduction algorithm. Also, we define f_i, s_i, T_i, p_i, q_i as in Section 5.3.

Lemma 6.6.1. *Let* $K = \max\{|a|, |b|, |c|\}$*. Then*

$$|p_i|, |q_i| \le 2\frac{K}{\sqrt{|\Delta|}} , \quad 0 \le i \le k-1 .$$

$$|p_k|, |q_k| \le K + \frac{K}{\sqrt{|\Delta|}} .$$

Proof. Let $\Delta > 0$. For a form $g = (A, B, C)$ of discriminant Δ let

$$\theta_1(g) = \frac{B+\sqrt{\Delta}}{2A} , \quad \theta_2(g) = \frac{B-\sqrt{\Delta}}{2A} .$$

Then

$$\theta_j(\rho(g)) = s - \frac{1}{\theta_j(g)} , \quad j = 1, 2 .$$

Thus we have

$$\theta_j(f_{i+1}) = s_i - \frac{1}{\theta_j(f_i)} , \quad 0 \le i \le k-1 , \quad j = 1, 2 . \tag{6.16}$$

Define

$$\mu_{j,-1} = 1 , \quad \mu_{j,i} = \prod_{l=0}^{i} \theta_j(f_l) , \quad 0 \le i \le k , \quad j = 1, 2 .$$

Then we obtain from (6.16) the recursion

$$\begin{pmatrix} \mu_{1,i} & \mu_{1,i+1} \\ \mu_{2,i} & \mu_{2,i+1} \end{pmatrix} = \begin{pmatrix} \mu_{1,i-1} & \mu_{1,i} \\ \mu_{2,i-1} & \mu_{2,i} \end{pmatrix} U(f_i) , \quad 0 \le i \le k-1 .$$

This shows that

$$\begin{pmatrix} \mu_{1,i} & \mu_{1,i+1} \\ \mu_{2,i} & \mu_{2,i+1} \end{pmatrix} = \begin{pmatrix} 1 & \theta_1(f) \\ 1 & \theta_2(f) \end{pmatrix} T_{i+1} , \quad 0 \le i \le k-1 .$$

Hence

$$\begin{pmatrix}\mu_{1,i}\\ \mu_{2,i}\end{pmatrix} = \begin{pmatrix}1 & \theta_1(f)\\ 1 & \theta_2(f)\end{pmatrix}\begin{pmatrix}p_i\\ q_i\end{pmatrix}, \quad 0 \le i \le k .$$

Multiplying the last equation with the inverse of the matrix

$$\begin{pmatrix}1 & \theta_1(f)\\ 1 & \theta_2(f)\end{pmatrix}$$

we obtain

$$p_i = \frac{\theta_2(f)\mu_{1,i} - \theta_1(f)\mu_{2,i}}{\theta_2(f) - \theta_1(f)}, \quad q_i = \frac{\mu_{2,i} - \mu_{1,i}}{\theta_2(f) - \theta_1(f)}, \quad 0 \le i \le k . \tag{6.17}$$

Since f_i is normal for $0 \le i \le k$ it follows from a straightforward calculation that $|\theta_2(f_i)| \le 1$ and thus $|\mu_{2,i}| \le 1$ for $1 \le i \le k$. Moreover for the normal but not reduced forms f_i, $0 \le i \le k-1$ one easily obtains $|\theta_1(f_i)| \le 1$ and $|\mu_{1,i}| \le 1$ for $1 \le i \le k-1$. Therefore the triangle inequality and $\theta_2(f) - \theta_1(f) = \sqrt{\Delta}/a$ yield

$$|p_i|, |q_i| \le 2\frac{|a|}{\sqrt{\Delta}}, \quad 0 \le i \le k-1 .$$

We finally estimate $|p_k|$, $|q_k|$. Since f_k is reduced, it follows that $\theta_1(f_k) \le \sqrt{\Delta}$. Hence (6.17) implies

$$|p_k|, |q_k| \le |a| + \frac{|a|}{\sqrt{\Delta}} .$$

□

Using Theorem 6.5.3 and Lemma 6.6.1 we immediately obtain the following result which was first proved in [Lag80].

Corollary 6.6.2. *Given an integral form* (a, b, c) *Algorithm* `reduce` *performs* $\mathrm{O}\big(\log(|a|/\sqrt{|\Delta|})\big)$ *arithmetic operations on numbers of binary length* $\mathrm{O}(\operatorname{size} f)$*.* □

6.6.2 Quadratic complexity

From Proposition 5.6.2 we can deduce that the number of bit operations required by Algorithm `reduce` is $\mathrm{O}((\operatorname{size} f)^3)$. But indeed we will prove that the bit complexity of the reduction algorithm is $\mathrm{O}((\operatorname{size} f)^2)$. This proof appeared first in [BB99]. Note that the algorithm of Schönhage [Sch91] has only quasi linear complexity. However, for being more efficient in practice, Schönhage's algorithm has to be applied to forms of extremely large size.

We need an auxiliary result.

Lemma 6.6.3. *Let f be a normal positive definite or indefinite integral binary quadratic form. If f, $\rho(f)$, and $\rho^2(f)$ are not reduced then $|s(f)| \ge 2$ and $|a| \ge |s(f)||c|$.*

Proof. Let $f = (a, b, c)$, $s = s(f)$. Assume that neither f nor $\rho(f)$ nor $\rho^2(f)$ are reduced.

Let $\Delta < 0$. Then $a > c$. Let $\rho(f) = (A, B, C)$. Then $C = cs^2 - bs + a$. This shows that for $|s| \leq 1$ we have $C \geq c = A$. Since $\rho(f)$ is not reduced we have $C = A$ and $B < 0$. Then $\rho^2(f)$ is reduced. Thus, $|s| \geq 2$. Next, we prove that $a \geq |s|c$. Since $s = \operatorname{sign}(c)\big[b/2|c|\big]$ one easily verifies that

$$|s| \leq \frac{|b|}{2|c|} + \frac{1}{2} \,. \tag{6.18}$$

This implies

$$a = \frac{b^2 + |\Delta|}{4c} > \frac{b^2}{4c} = c\left(\frac{|b|}{2c}\right)^2 \geq c\left(|s| - \frac{1}{2}\right)^2 > c|s|\big(|s| - 1\big) \geq c|s| \,.$$

This proves the assertion.

Let $\Delta > 0$. If $|a| < \sqrt{\Delta}$ then $\rho(f)$ is reduced by Lemma 6.5.2. Also, if $|c| < \sqrt{\Delta}$ then for the same reason $\rho^2(f)$ is reduced. Let $|a| \geq \sqrt{\Delta}$ and $|c| \geq \sqrt{\Delta}$. Since $0 \leq b^2 = \Delta + 4ac$ and $4|a||c| \geq 4\Delta$ it follows that $4ac > 0$, $b^2 > 4ac$, and $b^2 > \Delta$. Since f is normal and $|a| \geq \sqrt{\Delta}$ we have $|a| \geq |b|$. Therefore, $|a||b| \geq b^2 > 4|a||c|$. This means that $|b|/(2|c|) \geq 2$. But $s(f) = \operatorname{sign}(c)\big[b/(2|c|)\big]$. Thus $|s| \geq 2$. Next, we prove that $|a| \geq |s||c|$. Since $b^2 > \Delta$ and $|c| \geq \sqrt{\Delta}$ we obtain

$$|a| = \frac{b^2 - \Delta}{4|c|} \geq |c|\left(\left(\frac{b}{2|c|}\right)^2 - \frac{1}{4}\right) \,. \tag{6.19}$$

By (6.18) we have $(b/(2c))^2 \geq \big(|s| - 1/2\big)^2$. So we obtain $|a| \geq |c||s|\big(|s| - 1\big)$ from (6.19). Since $|s| \geq 2$ this proves the assertion. □

We are now able to prove our main result.

Theorem 6.6.4. *If f is a positive definite or indefinite quadratic form, then* `reduce` *requires time* $\mathrm{O}((\operatorname{size} f)^2)$.

Proof. By Lemma 6.6.1 the size of all integers appearing in `reduce` is $\mathrm{O}(\operatorname{size} f)$. The first normalization step requires time $\mathrm{O}((\operatorname{size} f)^2)$. Let k be the number of reduction steps performed by `reduce`. Let $f_i = (a_i, b_i, c_i)$ be the form that is input to the i-th reduction step. The determination of $s_i = s(f_i)$ takes time $\mathrm{O}\big(\operatorname{size}(c_i)\operatorname{size}(s_i)\big) = \mathrm{O}\big(\operatorname{size}(f)\operatorname{size}(s_i)\big)$. Also all other computations in the i-th reduction step take time $\mathrm{O}\big(\operatorname{size}(f)\operatorname{size}(s_i)\big)$.

By Lemma 6.6.3 follows $|s_i| \geq 2$ for $1 \leq i \leq k - 2$. Hence, $\operatorname{size}(s_i) = \mathrm{O}(\log|s_i|)$ for $1 \leq i \leq k - 2$ which implies that the time for the first $k - 2$ reduction steps is $\mathrm{O}\Big(\operatorname{size}(f)\log\prod_{i=1}^{k-2}|s_i|\Big)$. Moreover by Lemma 6.6.3 we have

$$|a_1| \geq |s_1||a_2| \geq \cdots \geq |a_{k-1}|\prod_{i=1}^{k-2}|s_i| \geq \prod_{i=1}^{k-2}|s_i| \,.$$

Hence, the time required by the first $k-2$ reduction steps is

$$\mathrm{O}\left(\operatorname{size}(f)\log\prod_{i=1}^{k-2}|s_i|)\right) = \mathrm{O}\bigl(\operatorname{size}(f)(\log|a_1|)\bigr) = \mathrm{O}\bigl((\operatorname{size} f)^2\bigr)\ .$$

Finally the time for the last two reduction steps obviously is bounded by $\mathrm{O}((\operatorname{size} f)^2)$ too. □

6.7 Enumerating integral reduced forms of a given discriminant

In this section we return to restrict ourselves to indefinite forms and explain how all reduced forms with given discriminants can be found.

6.7.1 Fixed discriminant

From the definition of reduced forms we obtain the following characterization which will be used in the enumeration algorithm.

Lemma 6.7.1. *The form $f = (a, b, c)$ is reduced if and only if $0 < b < \sqrt{\Delta}$ and $\sqrt{\Delta} - b < 2|a| < \sqrt{\Delta} + b$.*

Proof. Suppose that f is reduced. We show that f has the asserted properties. Obviously, $0 < b < \sqrt{\Delta}$. Also, $\bigl|2|a| - \sqrt{\Delta}\bigr| < b$ implies $\sqrt{\Delta} - 2|a| < b$ and $2|a| - \sqrt{\Delta} < b$, hence $\sqrt{\Delta} - b < 2|a| < \sqrt{\Delta} + b$.

Conversely, from the properties in the lemma it can be easily deduced that f is reduced. □

Suppose that Δ is an integer which is not a perfect square. From Lemma 6.7.1 we obtain the following method for enumerating all integral reduced forms of discriminant Δ. Fix some b with $0 < b < \sqrt{\Delta}$, $b \equiv \Delta \pmod 2$. Determine $A = (\Delta - b^2)/4$. For all a with $(\sqrt{\Delta} - b) < 2|a| < (\sqrt{\Delta} + b)$ check whether $c = -A/a$ is an integer. If it is, then (a, b, c) is a reduced form of discriminant Δ. After having examined all b with $0 < b < \sqrt{\Delta}$, all integral reduced forms of discriminant Δ are found. This process can be made faster. By Corollary 6.3.4 we obtain with each reduced form (a, b, c) three more reduced forms, namely $(-a, b, -c)$, (c, b, a) and $(-c, b, -a)$ for free. Also, since

$$-ac = A = \frac{\Delta - b^2}{4}$$

it follows that for $a^2 > A$ we have $c^2 < A$. If we, therefore, determine with each form (a, b, c) also the forms (c, b, a), $(-a, b, -c)$, $(-c, b, -a)$, we may just look for values of a with $\lceil(\sqrt{\Delta} - b)/2\rceil \le a \le \lfloor\sqrt{A}\rfloor$.

The detailed description of the algorithm for finding all integral primitive reduced forms of discriminant Δ is left to the reader as Exercise 6.18.4.

Note that if Δ is a fundamental discriminant, then the condition $\gcd(a, b, c) = 1$ is automatically satisfied.

Example 6.7.2. We determine all the reduced forms (a, b, c) of discriminant $\Delta = 73$. We have $\lfloor\sqrt{\Delta}\rfloor = 8$. Hence, $\Delta \equiv 1 \pmod 2$ and $0 < b \leq 8$.

If we put $b = 1$, we get $A = 18$, $\lfloor\sqrt{A}\rfloor = 4$, and $(\lceil\sqrt{\Delta}\rceil - b)/2 = (9-1)/2 = 4$. Hence, there is no reduced form with $b = 1$.

If we put $b = 3$ then we get $A = (\Delta - b^2)/4 = (73-9)/4 = 16$, $\lfloor\sqrt{A}\rfloor = 4$, $(\lceil\sqrt{\Delta}\rceil - b)/2 = 3$. Hence, we get the reduced forms $(4, 3, -4)$, $(-4, 3, 4)$.

If we put $b = 5$ then we get $A = 12$, $\lfloor\sqrt{A}\rfloor = 3$, $(\lceil\sqrt{\Delta}\rceil - b)/2 = 2$. Hence we get the reduced forms $(2, 5, -6)$, $(-2, 5, 6)$, $(6, 5, -2)$, $(-6, 5, 2)$, $(3, 5, -4)$, $(-3, 5, 4)$, $(4, 5, -3)$, $(-4, 5, 3)$.

If we put $b = 7$ then we get $A = 6$, $\lfloor\sqrt{A}\rfloor = 2$, $(\lceil\sqrt{\Delta}\rceil - b)/2 = 1$. So we obtain the reduced forms $(1, 7, -6)$, $(-1, 7, 6)$, $(6, 7, -1)$, $(-6, 7, 1)$, $(2, 7, -3)$, $(-2, 7, 3)$, $(3, 7, -2)$, $(-3, 7, 2)$.

Example 6.7.3. Consider $\Delta = 76$. We have $\lfloor\sqrt{\Delta}\rfloor = 8$, $\lceil\sqrt{\Delta}\rceil = 9$. Hence $b \equiv 0 \pmod 2$ and $1 \leq b \leq 8$.

If we put $b = 2$ we obtain $A = (76 - 4)/4 = 18$, $\lfloor\sqrt{A}\rfloor = 4$, $\lceil(\lceil\sqrt{\Delta}\rceil - b)/2\rceil = 4$. Hence there are no reduced forms with $b = 2$.

If we put $b = 4$, we obtain $A = 15$, $\lfloor\sqrt{A}\rfloor = 3$, $\lceil(\lceil\sqrt{\Delta}\rceil - b)/2\rceil = 3$. Hence, we obtain the reduced forms $(3, 4, -5)$, $(-3, 4, 5)$, $(-5, 4, 3)$, $(5, 4, -3)$.

For $b = 6$ we obtain $A = 10$, $\lfloor\sqrt{A}\rfloor = 3$ and $\lceil(\lceil\sqrt{\Delta}\rceil - b)/2\rceil = 2$ and therefore we get the forms $(2, 6, -5)$, $(-2, 6, 5)$, $(5, 6, -2)$, $(-5, 6, 2)$.

If we put $b = 8$ we obtain $A = 3$, $\lfloor\sqrt{A}\rfloor = 1$,$\lceil(\lceil\sqrt{\Delta}\rceil - b)/2\rceil = 1$. Hence, we find the reduced forms $(1, 8, -3)$, $(-1, 8, 3)$, $(3, 8, -1)$, $(-3, 8, 1)$.

6.7.2 Bounded discriminant

Let D be a positive integer. The following lemma can be used to find all integral reduced forms of discriminant $\Delta \leq D$.

Lemma 6.7.4. *Let D be a positive integer. Then f is indefinite, reduced, $\Delta \leq D$, and $0 < a \leq |c|$ if and only if $0 < a < \sqrt{\Delta}/2$, $a \leq -c < \sqrt{\Delta} - a$, $|c| - a < b \leq \sqrt{D + 4ac}$.*

Proof. Suppose that f is indefinite, and reduced with $\Delta \leq D$ and $0 < a \leq |c|$. By Lemma 6.2.7, c is negative and $a - c < \sqrt{\Delta}$. So $a \leq |c|$ implies $2a \leq a - c < \sqrt{\Delta}$, hence $a < \sqrt{\Delta}/2$. Also, it follows that $a \leq -c < \sqrt{d} - a$. Now $\Delta = b^2 - 4ac$, hence $b^2 \leq D + 4ac$. Also, since f is reduced and $a < \sqrt{\Delta}/2$ it follows that $b > \sqrt{\Delta} - 2a$, hence $b + 2a > \sqrt{\Delta}$ which implies $b^2 + 4ab + 4a^2 > b^2 - 4ac$ and therefore $b > a - c$.

Conversely, assume that $0 < a < \sqrt{d}/2$, $a \leq -c < \sqrt{d} - a$, and $|c| - a < b \leq \sqrt{D + 4ac}$. Then $-c$ is positive, hence $a \leq |c|$. Also, $\Delta = b^2 + 4a|c|$. This inequality together with $a \le |c|$ shows that $a < \sqrt{\Delta}/2$. Hence, f is reduced by Lemma 6.5.1. Since $b \le \sqrt{D + 4ac}$ and $b > 0$ we have $\Delta \le D$. Finally, since Δ is not a perfect square the form f is non degenerate. □

6.8 Reduced forms in an equivalence class

In this section, we we explain how to find all reduced forms in the equivalence class or the proper equivalence class of the given form f.

6.8.1 The reduction operator is bijective

We know from Lemma 6.5.6 that the reduction operator sends a reduced form to another reduced form in the same proper equivalence class. We prove that this map is a permutation of the set $\mathcal{F}^+(f)$ of all reduced forms in the proper equivalence class of f.

Theorem 6.8.1. *Let g be an irrational indefinite form. Then the map $\mathcal{F}^+(g) \to \mathcal{F}^+(g)$, $f \mapsto \rho(f)$ is a bijection.*

Proof. We prove that ρ is injective. Let $f = (a, b, c), f' = (a', b', c') \in \mathcal{F}^+(g)$ with $\rho(f) = \rho(f')$. We have $\rho(f) = \big(c, -b + 2sc, cs^2 - bs + a\big)$ with $s = s(f)$ and $\rho(f') = (c', -b' + 2s'c', c'(s')^2 - b's' + a')$ with $s' = s(f')$. Hence $c = c'$ and $b \equiv b' \pmod{2|c|}$. Since f and f' are reduced, we have $b = b'$ and therefore $a = a'$.

We prove that the map is surjective. We claim that an inverse image of a reduced form $f = (a, b, c)$ is

$$\rho^{-1}(f) = f' = \big(at^2 - bt + c, -b + 2at, a\big) \,. \tag{6.20}$$

with

$$t = t(f) = \operatorname{sign}(a) \left\lfloor \frac{b + \sqrt{\Delta}}{2|a|} \right\rfloor \,. \tag{6.21}$$

The forms f and f' are properly equivalent. Also, since $\rho(f')$ is the uniquely determined normal form in the Γ-orbit of $(a, b - 2at, at^2 - bt + c)$ and since f is a normal form in that Γ-orbit, we have $\rho(f') = f$. We show that f' is reduced. Since (a, b, c) is reduced, it follows from Corollary 6.3.4 that the form (c, b, a) is also reduced. Now $\rho(c, b, a) = (a, -b + 2at, at^2 - bt + c)$. By Lemma 6.5.6 this form is also reduced. Another application of Corollary 6.3.4 implies that f' is reduced. □

If f is an integral form, then in (6.21) we may replace $\sqrt{\Delta}$ by $\lfloor\sqrt{\Delta}\rfloor$ as in (6.4). We can also write

$$\rho^{-1}(f) = fV(f)$$

with

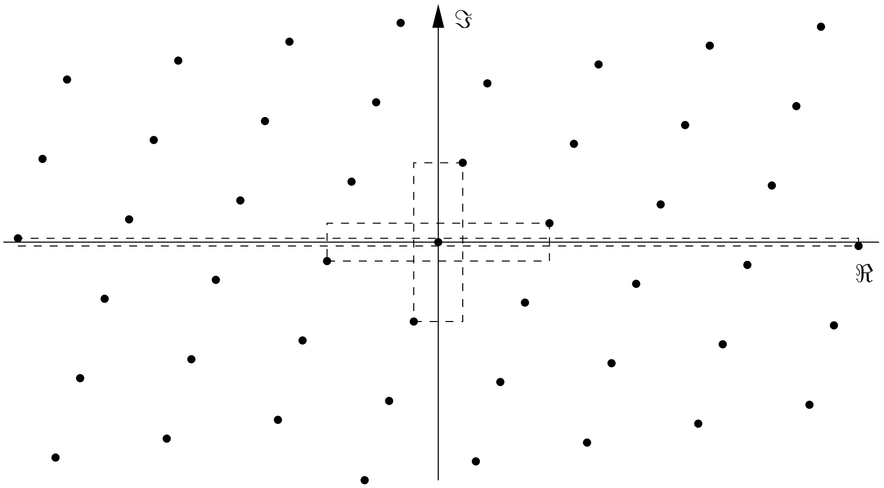

Fig. 6.2. Some minimal points in a 2-dimensional lattice

$$V(f) = \begin{pmatrix} t(f) & 1 \\ -1 & 0 \end{pmatrix} . \tag{6.22}$$

Note that

$$t(f) = s\bigl(\rho^{-1}(f)\bigr) , \quad V(f) = U^{-1}(\rho^{-1}(f)) .$$

Example 6.8.2. Let $f = (1, 7, -6)$. Then $\Delta(f) = 73$, $t(f) = \lfloor (7+8)/2 \rfloor = 7$ and $\rho^{-1}(f) = (49 - 49 - 6, -7 + 14, 1) = (-6, 7, 1)$. We verify this result. Set $g = (-6, 7, 1)$. Then $s(g) = 7$ and $\rho(g) = (1, -7+14, -6-49+49) = (1, 7, -6)$.

Below we will show that the permutation ρ of $\mathcal{F}^+(f)$ is transitive. For the proof we need a geometric interpretation of reduction theory which we give in the next two sections.

6.8.2 Geometric characterization of reduced forms

We introduce the most important notions in the geometric interpretation of reduction theory for indefinite forms. Let L be a two-dimensional irrational lattice in A_1. We define minimal points and bases.

Definition 6.8.3.

1. *A point* $\mu = (\mu_1, \mu_2) \in L$ *is called a* minimal point *of* L *if* $\mu \neq 0$ *and if there is no* $\theta = (\theta_1, \theta_2) \in L$ *different from* 0 *and* $\pm\mu$ *with* $|\theta_1| \leq |\mu_1|$ *and* $|\theta_2| \leq |\mu_2|$.
2. *A basis* $B = (\alpha, \gamma) = \bigl((\alpha_1, \alpha_2), (\gamma_1, \gamma_2)\bigr)$ *of* L *is called a* minimal *basis of* L *if* $|\alpha_1| < |\gamma_1|$, $|\alpha_2| > |\gamma_2$ *and there is no* $\theta = (\theta_1, \theta_2) \in L$ *different from* 0, $\pm\alpha$, *and* $\pm\gamma$ *with* $|\theta_1| \leq |\gamma_1|$ *and* $|\theta_2| \leq |\alpha_2|$.

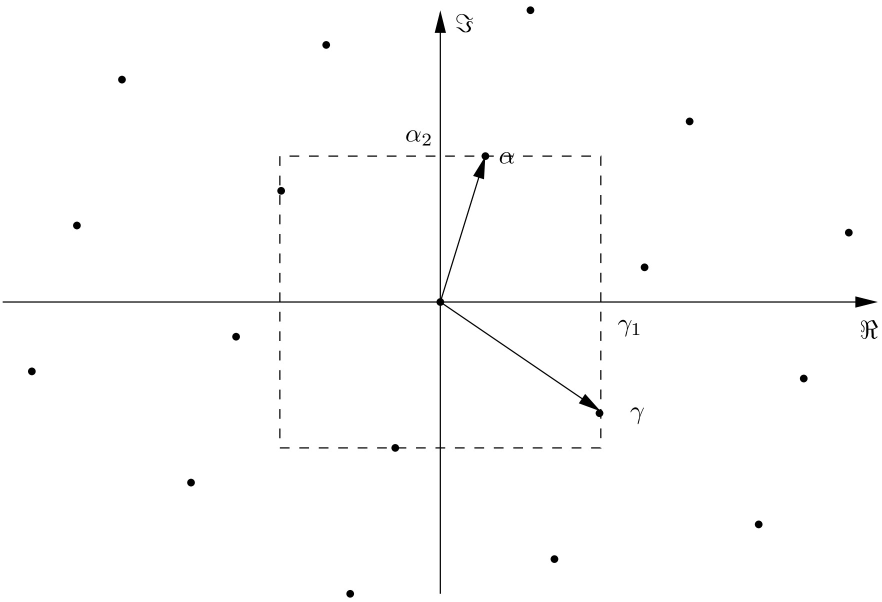

Fig. 6.3. A minimal basis of a 2-dimensional lattice

Geometrically speaking, a minimal point in L is a non-zero point μ such that the smallest rectangle that contains $\pm\mu$ and whose sides are parallel to the axes, does not contain lattice points different from 0 and $\pm\mu$. Likewise, a minimal basis of L is a $\mathbb{Z}$-basis (α, γ) of L such that the smallest rectangle that contains $\pm\alpha$ and $\pm\gamma$ and whose sides are parallel to the axes does not contain lattice points different from 0 and $\pm\alpha$ and $\pm\gamma$. From Proposition 4.2.4 we obtain the following alternative characterization of minimal points and bases. It tells us that for the minimality of a point or a basis of L it is sufficient that the interior of the above rectangles does not contain lattices points different from 0.

Lemma 6.8.4.

1. *A point $\mu = (\mu_1, \mu_2) \in L$ is a* minimal point *of L if and only if $\mu \neq 0$ and if there is no non-zero $\theta = (\theta_1, \theta_2) \in L$ with $|\theta_1| < |\mu_1|$ and $|\theta_2| < |\mu_2|$.*
2. *A basis $B = (\alpha, \gamma) = \big((\alpha_1, \alpha_2), (\gamma_1, \gamma_2)\big)$ of L is a* minimal *basis of L if and only if $|\alpha_1| < |\gamma_1|$, $|\alpha_2| > |\gamma_2$ and there is no non-zero $\theta = (\theta_1, \theta_2) \in L$ with $|\theta_1| < |\gamma_1|$ and $|\theta_2| < |\alpha_2|$.*

Proof. Exercise 6.18.2.

The next lemma will be useful in many proofs.

Lemma 6.8.5.

1. *Let $\varepsilon \in A_1^*$ and let $\theta \in A_1$. Then θ is a minimal point of L if and only if $\varepsilon\theta$ is a minimal point of εL, which in turn, is true if and only if $\sigma(\theta)$ is a minimal point of $\sigma(L)$.*

2. Let $\varepsilon \in A_1^*$ and let $(\alpha, \gamma) \in A_1^2$. Then (α, γ) is a minimal basis of L if and only if $(\varepsilon\alpha, \varepsilon\gamma)$ is a minimal basis of εL, which in turn, is true if and only if $\big(\sigma(\gamma), \sigma(\alpha)\big)$ is a minimal basis of $\sigma(L)$.

Proof. Exercise 6.18.9.

Here is the geometric interpretation of reduced forms.

Theorem 6.8.6. *Let $L = (1/a)L(f)$. Then the following statements are equivalent:*

1. The form f is reduced.
2. The pair $\big(1, \theta(f)\big)$ is a minimal basis of L.
3. The form f is normal and 1 is a minimal point of L.

Proof. We write $\theta = (\theta_1, \theta_2) = \theta(f)$.

Suppose that f is reduced. We show that $(1, \theta)$ is a minimal basis of L. By Lemma 6.3.2 we have $|\theta_1| > 1$, $|\theta_2| < 1$ and $\theta_1\theta_2 < 0$. Suppose that x, y are integers such that

$$|x + y\theta_1| < |\theta_1| \, , \quad |x + y\theta_2| < 1 \, .$$

We must show that $x = y = 0$. If $y = 0$, then the second inequality implies $x = 0$. Suppose that $y \neq 0$. The first inequality implies that $x \neq 0$ and that x and $y\theta_1$ have opposite sign. So x and $y\theta_2$ have the same sign since $\theta_1\theta_2$ have opposite sign by Lemma 6.3.2. However, this contradicts the second inequality.

Let $(1, \theta)$ be a minimal basis of L. We show that f is reduced. We have $|\theta_1| > 1$ and $|\theta_2| < 1$. If $\theta_1\theta_2 > 0$, then $\big|1 - \operatorname{sign}(\theta_1)\theta_1\big| < |\theta_1|$ and $\big|1 - \operatorname{sign}(\theta_1)\theta_2\big| < 1$ which contradicts the minimality of 1. Hence, $\theta_1\theta_2 < 0$. Lemma 6.3.2 implies that f is reduced.

Now assume that the first or the second condition holds. Then f is normal and 1 is a minimal point of L

Finally, assume that f is normal and that 1 is a minimal point of L. We show that f is reduced. If $|a| \geq \sqrt{\Delta}$ then $|\theta_1| \leq 1$ and $|\theta_2| \leq 1$ since f is normal. Hence $\theta \in \{0, \pm 1\}$, since 1 is a minimal point of L. This is impossible. It follows that $|a| < \sqrt{\Delta}$. Since f is normal, we have $\sqrt{\Delta} - 2|a| < b < \sqrt{\Delta}$. This implies that θ_1 and θ_2 have opposite sign and $|\theta_2| < 1$. So $|\theta_1| > 1$ since 1 is a minimal point of L. Lemma 6.3.2 implies that f is reduced. □

Corollary 6.8.7. *Let $B = (\alpha, \gamma)$ be a basis of L. Then the following statements are equivalent:*

1. The form f_B is reduced.
2. B is a minimal basis of L.
3. The form f_B is normal and α is a minimal point of L.

Proof. Let $f = f_B$ and let $\theta = \theta(f)$. By Proposition 4.3.18 we have $\theta = \gamma/\alpha$. By Lemma 6.8.5, the basis B is a minimal basis of L if and only if $(1, \theta)$ is a minimal basis of $L(\theta)$. The same Lemma implies that α is a minimal point of L if and only if θ is a minimal point of $L(\theta)$. Hence, the corollary follows from Theorem 6.8.6. □

6.8.3 The reduction operator is transitive

Let $f = (a, b, c)$ be reduced, $a > 0$. We know from Theorem 6.8.1 that the reduction operator ρ is a permutation of the set $\mathcal{F}^+(f)$ of all reduced forms in the proper equivalence class of f.

In this section we use geometric arguments to show that this permutation is transitive. [1]

Set

$$f_i = \rho^i(f),\ i \in \mathbb{Z}\ .$$

In particular, we have $f_0 = f$. We must prove that $\mathcal{F}^+(f) = \{f_i : i \in \mathbb{Z}\}$. We set

$$B_0 = \frac{1}{a}B(f_0) = (1, \theta(f)) = \left(1, \left(\frac{b+\sqrt{\Delta}}{2a}, \frac{b-\sqrt{\Delta}}{2a}\right)\right)\ .$$

Then $f_{B_0} = (1/a)f$ by Proposition 4.3.18 and the orientation of B_0 is positive. We also set

$$B_{i+1} = B_i U(f_i)\ , \quad i \geq 0\ ,$$
$$B_{i-1} = B_i V(f_i)\ , \quad i \leq 0\ ,$$

with $U(f_i)$ from (6.13) and $V(f_i)$ from (6.22). Since the $U(f_i)$ and $V(f_i)$ have discriminant 1, it follows from (4.16) that the B_i are bases with positive orientation of the lattice

$$L = (1/a)L(f)\ .$$

Since $f_{i+1} = f_i U(f_i)$ and $f_{i-1} = f_i V(f_i)$, $i \in \mathbb{Z}$ Proposition 4.3.18 implies

$$f_{B_i} = (1/a)f_i\ , \quad i \in \mathbb{Z}\ . \tag{6.23}$$

Also, since the f_i are reduced, it follows from Corollary 6.8.7 that the B_i are minimal bases of L.

We will now show that the sequence $(\pm B_i)_{i\in\mathbb{Z}}$ contains all minimal bases of L with positive orientation. From this fact we will deduce that $\mathcal{F}^+(f) = \{\rho^i(f) : i \in \mathbb{Z}\}$.

Because of the special form of the transformations $U(f_i)$ and $V(f_i)$, $i \in \mathbb{Z}$, we can write

$$B_i = (\mu_i, \mu_{i+1})\ , \quad i \in \mathbb{Z}\ . \tag{6.24}$$

Then by Proposition 4.3.18 we have

$$\theta(f_i) = \frac{\mu_{i+1}}{\mu_i},\ i \in \mathbb{Z}\ . \tag{6.25}$$

[1] For a different proof in the language of continued fractions see [Bue89]. Roughly, computing the cycle of a form f is equivalent to computing the partial quotients in the continued fraction expansion of $\theta(f)$, and one shows for two forms f and f' in the same class that the continued fraction expansions of $\theta(f)$ and $\theta(f')$ can be converted into each other.

We also write

$$\mu_i = (\mu_{i,1}, \mu_{i,2}) , \quad i \in \mathbb{Z} .$$

The next Lemma shows that the sequence $(|\mu_{i,1}|)$ is strictly increasing with exponential growth.

Lemma 6.8.8. *We have* $|\mu_{i+1,1}| > |\mu_{i,1}|$ *and* $|\mu_{i+2,1}| > 2|\mu_{i,1}|$ *for* $i \in \mathbb{Z}$.

Proof. The first inequality follows from the fact that B_i is a minimal basis.

We prove the second inequality. Let

$$\alpha = \operatorname{sign}(\mu_{i+2,1})\mu_{i+2} - \operatorname{sign}(\mu_{i,1})\mu_i .$$

Write $\alpha = (\alpha_1, \alpha_2)$.

Assume that $|\mu_{i+2,1}| \le 2|\mu_{i,1}|$. We will deduce that $|\alpha_j| \le |\mu_{i,j}|$, $j = 1, 2$. This contradicts the minimality of μ_i.

We have $|\mu_{i,1}| < |\mu_{i+2,1}| \le 2|\mu_{i,1}|$. Then

$$|\alpha_1| = \Big||\mu_{i+2,1}| - |\mu_{i,1}|\Big| \le |\mu_{i,1}| . \tag{6.26}$$

Now we show that the analogous inequality also holds for the second coordinate. Since f_i is reduced, (6.25) and Lemma 6.3.2 imply that $\mathrm{N}(\mu_{i+1})/\mathrm{N}(\mu_i) < 0$. Therefore, $\mathrm{N}(\mu_i)$ and $\mathrm{N}(\mu_{i+2})$ have the same sign. This implies that $\operatorname{sign}(\mu_{i+2,1})\mu_{i+2,2}$ and $\operatorname{sign}(\mu_{i,1})\mu_{i,1}$ have the same sign. Also, $|\mu_{i+2,2}| < |\mu_{i,2}|$. Hence $|\alpha_2| = \big|\operatorname{sign}(\mu_{i+2,1})\mu_{i+2,2} - \operatorname{sign}(\mu_{i,1})\mu_{i,2}\big| < |\mu_{i,2}|$. □

As a consequence of Lemma 6.8.8 we obtain the following result.

Corollary 6.8.9. *1.* $\lim_{i\to\infty} |\mu_{i,1}| = \infty$. *2.* $\lim_{i\to-\infty} |\mu_{i,1}| = 0$. □

Theorem 6.8.10.
1. The set of all minimal points of L *is* $\big\{\pm\mu_i : i \in \mathbb{Z}\big\}$.
2. The set of all minimal bases of L *with positive orientation is* $\big\{\pm B_i : i \in \mathbb{Z}\big\}$.

Proof. 1. Suppose that $\alpha = (\alpha_1, \alpha_2) \in \mathrm{Min}(L)$. It follows from Corollary 6.8.9 that there is $i \in \mathbb{Z}$ with $|\mu_{i,1}| \le |\alpha_1| < |\mu_{i+1,1}|$. We show that $\alpha = (\alpha_1, \alpha_2) \in \{\pm\mu_i\}$. Otherwise $|\mu_{i,1}| < |\alpha_1| < |\mu_{i+1,1}|$. Since (μ_i, μ_{i+1}) is a minimal basis of L, this implies $|\mu_{i,2}| < |\alpha_2|$ and this contradicts the minimality of α.

2. Let $B = (\alpha, \gamma)$ be a minimal basis of L with positive orientation. Then α is a minimal point of L. So 1. implies that $\alpha = s\mu_i$ for some $s \in \{\pm 1\}$ and $i \in \mathbb{Z}$. Then sB_i and B are minimal bases of positive orientation with the same first element. We show that $sB_i = B$. It suffices to show that $\gamma = (\gamma_1, \gamma_2) = s\mu_{i+1}$. Since B is a minimal basis of L we have $|\gamma_2| < |\alpha_2| = |\mu_{i+1,2}|$. So $|\gamma_1| < |\mu_{i+2,1}|$ contradicts the minimality of B_i. Likewise, $|\gamma_1| > |\mu_{i+2,1}|$ contradicts the minimality of B. It follows from Proposition 4.2.4 that $\gamma = \pm\mu_{i+1}$. Since B and sB_i have the same orientation, it follows that $B = sB_i$. □

The main results of this section follow from Theorem 6.8.10.

Corollary 6.8.11. *The set of reduced forms in the proper equivalence class of the reduced indefinite form f is $\{\rho^i(f) : i \in \mathbb{Z}\}$.*

Proof. Let g be a reduced indefinite form in the proper equivalence class of f. Then $g = fU$ with $U \in \mathrm{SL}(2, \mathbb{Z})$. Set $C = B_0U$. By Proposition 4.3.18 we have $g = fU = af_{B_0}U = af_C$. Corollary 6.8.7 implies that C is a minimal basis of L. Hence, by Theorem 6.8.10 there is an integer i with $B_i = C$. So (6.23) implies $g = af_C = af_{B_i} = f_i$. □

Corollary 6.8.12. *The map $\rho : \mathcal{F}^+(f) \to \mathcal{F}^+(f)$ is a transitive permutation.* □

Example 6.8.13. Let $f_0 = f = (1, 1, -1)$. Then $\Delta(f) = 5$. We compute $\mathcal{F}^0(f)$. We have $f_1 = \rho(f_0) = (-1, 1, 1)$ and $f_2 = \rho(f_1) = (1, 1, -1) = f_0$. We also have $f_{-1} = \rho^{-1}(f_0) = f_1$ or, more generally, $f_i = f_j$ if and only if $i \equiv j \pmod 2$. So the sequence $(f_i)_{i \in \mathbb{Z}}$ is periodic with period length 2.

In Example 6.8.13 we have determined the sequence $\left(\rho^i(f)\right)_{i \in \mathbb{Z}}$ for the integral form $(1, 1, -1)$. We have seen that this sequence is periodic with period length 2. We characterize all forms for which this sequence is periodic.

Theorem 6.8.14. *Let f be reduced. If $\theta(f)$ is a quadratic irrationality, then the sequence $\left(\rho^i(f)\right)_{i \in \mathbb{Z}}$ is periodic. Otherwise the elements of this sequence are pairwise distinct.*

Proof. Assume that $\theta(f)$ is a quadratic irrationality. Then the number of reduced forms in the equivalence class of f is finite by Exercise 6.18.5. Therefore, there are integers i, k, $k > 0$ with $\rho^i(f) = \rho^{i+k}(f)$. Since ρ is a bijection, it follows that $\rho^s(f) = \rho^t(f)$ for all integers s, t with $s \equiv t \pmod k$. This means that the sequence $\left(\rho^i(f)\right)_{i \in \mathbb{Z}}$ is periodic.

Conversely, assume that $f_j = f_i$ for integers i, j with $i < j$. Since f and f_i are equivalent, it suffices to prove that $\theta = \theta(f_i)$ is a quadratic irrationality. Set $T = U(f_i)U(f_{i+1}) \cdots U(f_{j-1})$. Then $T \in \mathrm{SL}(2, \mathbb{Z})$, $f_j = f_iT$, and $B_j = B_iT$. Let $T = \begin{pmatrix} s & t \\ u & v \end{pmatrix}$. Then Proposition 4.3.21 implies $\theta T = \theta(f_i)T = \theta(f_iT) = \theta(f_i) = \theta$, hence $u\theta^2 + (s - v)\theta - t = 0$. If $u \neq 0$, then θ is a quadratic irrationality. Let $u = 0$. Then $s = v$ and $|s| = 1$ since $\det T = 1$. So $t = 0$, hence $T = \pm I_2$ which is impossible since $B_i \neq \pm B_j$. □

Note that $\theta(f)$ is a quadratic irrationality if and only if b/a and c/a are rational numbers which, in turn, is true if and only if $\mathbb{R}_{>0}f$ contains an integral form.

6.9 Enumeration of the reduced forms in an equivalence class

We use the results of the previous section to enumerate all reduced forms in the equivalence class of a given indefinite reduced form. We first discuss the relationship between the proper equivalence class and the equivalence class of f.

For a form $f = (a, b, c)$ set

$$\tau(f) = f \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = (-a, b, -c) \; . \tag{6.27}$$

Then

$$\tau\rho(f) = \rho\tau(f) = \Big(-c, -b + 2s(f)c, -\big(a - bs(f) + cs(f)^2\big)\Big) \; . \tag{6.28}$$

Also, f and $\tau(f)$ are equivalent but not necessarily properly equivalent.

Proposition 6.9.1.

1. *If f and $\tau(f)$ are properly equivalent, then the equivalence class of f is equal to the proper equivalence class of f.*
2. *If f and $\tau(f)$ are not properly equivalent, then the equivalence class of f is the disjoint union of the proper equivalence classes of f and $\tau(f)$.*

Proof. 1. Assume that f and $\tau(f)$ are properly equivalent. Let g be a form in the equivalence class of f. We show that g is properly equivalent to f. If g is improperly equivalent to f, then g is properly equivalent to $\tau(f)$. But $\tau(f)$ is properly equivalent to f. Hence, g is properly equivalent to f.

2. Assume that f and $\tau(f)$ are not properly equivalent. Then the proper equivalence classes of f and $\tau(f)$ are disjoint. In addition, any form in the equivalence class of f belongs either to the proper equivalence class of f or to the proper equivalence class of $\tau(f)$. □

Corollary 6.9.2. *If $\theta(f)$ is not a quadratic irrationality, then the equivalence class of f is the disjoint union of two proper equivalence classes.* □

6.10 Cycles of reduced forms

Let f be an integral form. In order to compute the reduced forms in the equivalence class or the proper equivalence class of f it is more efficient to calculate the sequence $(\tau\rho)^i(f)_{i\in\mathbb{Z}}$ than the sequence $\big(\rho^i(f)\big)$. This is demonstrated in the next example.

Example 6.10.1. Let $f = (1, 3, -2)$. Then the sequence $\big(\rho^i(f)\big)_{i\in\mathbb{Z}}$ is periodic with period $\big((1, 3, -2),\ (-2, 1, 2),\ (2, 3, -1),\ (-1, 3, 2),\ (2, 1, -2),\ (-2, 3, 1)\big)$ while the sequence $\big((\rho\tau)^i(f)\big)_{i\in\mathbb{Z}}$ is periodic with period $\big((1, 3, -2),\ (2, 1, 2)$ $(2, 3, -1)\big)$. The second period is half as long as the first period. Also, since the length of the second period is odd and since we know that in the sequence $\big(\rho^i(f)\big)_{i\in\mathbb{Z}}$ the sign of the coefficient of X^2 alternates we can tell from the second period that the first period contains $\tau(f)$ and is twice as long. By Theorem 6.9.1 the equivalence class and the proper equivalence class of f are equal.

It follows from Theorem 6.8.14 that the sequence $\left((\tau\rho)^i(f)\right)_{i\in\mathbb{Z}}$ is periodic.

Definition 6.10.2. *Let f be integral.*

1. *The* proper cycle *of f is the sequence $\left(\rho^i(g)\right)_{i\in\mathbb{Z}}$ where g is a reduced form which is properly equivalent to f.*
2. *The* cycle *of f is the sequence $\left((\tau\rho)^i(g)\right)_{i\in\mathbb{Z}}$ where $g = (A, B, C)$ is a reduced form with $A > 0$ which is equivalent to f.*

The cycle of f is an invariant of the equivalence class of f. The cycle of the principal form of discriminant Δ is called the *principal cycle* of discriminant Δ. Also, the proper cycle of f is an invariant of the proper equivalence class of f. We represent the cycle or the proper cycle of f by its period $(f_0, \dots, f_{l-1})$ where f_0 is any reduced form in the cycle.

We describe the elements of the cycle and the proper cycle of f.

Proposition 6.10.3.

1. *The proper cycle of f consists of all reduced forms in the proper equivalence class of f.*
2. *The cycle of f consists of all reduced forms (A, B, C) with $A > 0$ in the equivalence class of f.*

Proof. 1. This statement is a consequence of Corollary 6.8.11.

2. The equivalence class of f contains a reduced form (a', b', c') with $a' > 0$. So we assume without loss of generality that $a > 0$. By construction, any form (A, B, C) in the cycle of f is a reduced form in the equivalence class of f with $A > 0$. Conversely, let (A, B, C) be a reduced form in the equivalence class of f with $A > 0$. Then one of the forms (A, B, C) or $\tau(A, B, C)$ is properly equivalent to f. Hence, one of those forms belongs to the proper cycle of f. If (A, B, C) belongs to the proper cycle of f, then $(A, B, C) = \rho^i(f)$ with even i since both a and A are positive and the sign of the coefficient of X^2 of the forms $\rho^j(f)$ is $(-1)^j$, $j \in \mathbb{Z}$. Since τ^2 is the identity and since τ commutes with ρ, it follows that $(A, B, C) = (\tau\rho)^i(f)$. This implies that (A, B, C) belongs to the cycle of f. Assume that $\tau(A, B, C) = (-A, B, -C)$ belongs to the proper cycle of f. Then $\tau(A, B, C) = \rho^i(f)$ with an odd i. Hence, $(A, B, C) = (\tau\rho^i)(f) = (\tau\rho)^i(f)$. So also in this case, (A, B, C) belongs to the cycle of f. □

Example 6.10.4. By Example 6.10.1 the cycle of $f = (1, 3, -2)$ is $\left((1, 3, -2), (2, 1, 2), (2, 3, -1)\right)$. We can also write $\left((2, 1, 2), (2, 3, -1), (1, 3, -2)\right)$ for this cycle. The proper cycle of f is $\left((1, 3, -2), (-2, 1, 2), (2, 3, -1), (-1, 3, 2), (2, 1, -2), (-2, 3, 1)\right)$ which can also be written as $\left((2, 3, -1), (-1, 3, 2), (2, 1, -2), (-2, 3, 1), (1, 3, -2), (-2, 1, 2)\right)$.

We explain how the proper cycle is computed from the cycle of f.

Proposition 6.10.5. *Let f be integral and let $(f_0, f_1, \dots, f_{l-1})$ be the cycle of f.*

1. *If the length l of that cycle is odd, then the proper cycle of f and $\tau(f)$ is $\big(f_0, \tau(f_1), f_2, \dots, f_{l-1}, \tau(f_0), f_1, \dots, \tau(f_{l-1})\big)$. In this case the equivalence class of f is equal to the proper equivalence class of f.*
2. *If the length l of that cycle is even, then the proper cycle of f is $\big(f_0, \tau(f_1), \dots, \tau(f_{l-1})\big)$ and the proper cycle of $\tau(f)$ is $\big(\tau(f_0), f_1, \dots, f_{l-1}\big)$. Also, the equivalence class of f is the disjoint union of the proper equivalence class of f and the proper equivalence class of $\tau(f)$.*

Proof. 1. Suppose that l is odd. Then $\tau(f) = \tau(\rho\tau)^l(f) = \rho^l(f)$. Hence, the proper cycle of f is of the asserted form. Also, f and $\tau(f)$ are properly equivalent. Proposition 6.9.1 implies that the equivalence class of f is equal to the proper equivalence class of f.

2. Suppose that l is odd. Then $f = (\tau\rho)^l(f) = \rho^l(f)$. Hence, the proper cycle of f and $\tau(f)$ have the asserted form. Those cycles are different since the number of reduced forms (A, B, C) with $A > 0$ in the equivalence of f is l. Hence, f and $\tau(f)$ are not properly equivalent. Proposition 6.9.1 implies that the equivalence class of f is the disjoint union of the proper equivalence class of and and the proper equivalence class of $\tau(f)$. □

Let $f = (a, b, c)$ with $a > 0$. We explain how to compute the cycle $(f_0, f_1, \dots, f_{l-1})$ of f. We set $f_0 = f$,

$$f_0 = (a_0, b_0, c_0)$$

and

$$s_i = |s(f_i)| = \left\lfloor \frac{b_i + \lfloor\sqrt{\Delta}\rfloor}{2|c_i|} \right\rfloor . \tag{6.29}$$

Then

$$f_{i+1} = (a_{i+1}, b_{i+1}, c_{i+1}) = \big(|c_i|, -b_i + 2s_i|c_i|, -(a_i + b_i s_i + c_i s_i^2)\big) . \tag{6.30}$$

Another way of writing f_{i+1} is

$$f_{i+1} = f_i \begin{pmatrix} 0 & 1 \\ 1 & |s(f_i)| \end{pmatrix}, \quad i \in \mathbb{Z} . \tag{6.31}$$

To compute matrices

$$T_i = \begin{pmatrix} p_i & p_{i+1} \\ q_i & q_{i+1} \end{pmatrix} \in \mathrm{GL}(2, \mathbb{Z}) \tag{6.32}$$

with

$$f_i = fT_i .$$

we set $T_0 = I_2$ and

$$p_{i+2} = s_i p_{i+1} + p_i \ , \quad q_{i+2} = s_i q_{i+1} + q_i \ , \quad i \in \mathbb{Z} \ . \tag{6.33}$$

We give two examples for cycles of reduced forms.

Example 6.10.6. Let $f = (1, 7, -6)$. The form f is reduced and the discriminant of this form is 73. The cycle computation yields

i	0	1	2	3	4	5	6	7	8	9
a_i	1	6	2	3	4	4	3	2	6	1
b_i	7	5	7	5	3	5	7	5	7	7
$-c_i$	6	2	3	4	4	3	2	6	1	6
p_i	1	0	1	3	7	10	17	44	149	193
q_i	0	1	1	4	9	13	22	57	193	250
s_i	1	3	2	1	1	2	3	1	7	

This means that the cycle of f has period length 9. This is an odd number. Hence, the proper cycle of f has period length 18.

Example 6.10.7. Let $f = (1, 8, -3)$. The form is reduced and the discriminant of this form is 76.

i	0	1	2	3	4	5	6
a_i	1	3	5	2	5	3	1
b_i	8	4	6	6	4	8	8
$-c_i$	3	5	2	5	3	1	3
p_i	1	0	1	1	4	5	14
q_i	0	1	2	3	11	14	39
s_i	2	1	3	1	2	8	

This means that the cycle of f has length 6. This is an even number. Hence the proper cycle of f has also period length 6. That proper cycle is $\big((1, 8, -3), (-3, 4, 5), (5, 6, -2), (-2, 6, 5), (5, 4, -3), (-3, 8, 1)\big)$.

Algorithm `cycle` computes the cycle of f. In this algorithm we use Procedure `rhoTau` which computes the function $\rho\tau$.

Finally, we prove bounds for the entries of T_i.

Lemma 6.10.8. *We have*

1. $p_{i+1} > p_i > 0$ *and* $p_{i+2} \geq 2p_i$ *for* $i \geq 3$*,*
2. $q_{i+1} > q_i > 0$ *and* $q_{i+2} \geq 2q_i$ *for* $i \geq 2$*,*
3. $p_i \geq 2^{\lfloor (i-3)/2 \rfloor}$ *for* $i \geq 3$*,*
4. $q_i \geq 2^{\lfloor (i-2)/2 \rfloor}$ *for* $i \geq 2$*,*

Algorithm 6.1 `cycle` (f)

Input: An integral irreducible indefinite reduced form f
Output: The cycle of f

```
Initialize the cycle $C$ with $f$
$g \leftarrow$ rhoTau$(f)$
while $g \neq f$ do
   Append $g$ to $C$
   $g \leftarrow$ rhoTau$(g)$
return $C$
```

5. $p_{i+2} < (\sqrt{\Delta}+1)^i$ *for* $i \geq 0$*,*
6. $q_{i+1} < (\sqrt{\Delta}+1)^i$ *for* $i \geq 0$*.*

Proof. By Exercise 6.18.15 we have $s_i \geq 1$ for $i \geq 0$. We use the recursion (6.33) and obtain $p_1 = 0$, $p_2 = 1$, $p_3 = s_1$ and $p_{i+2} = s_i p_{i+1} + p_i \geq p_{i+1} + p_i > p_{i+1}$ for $i \geq 2$. This implies $p_{i+2} \geq 2p_i$ for $i \geq 2$. The second assertion is proved analogously. The third and fourth assertion follow from the first and the second assertion. We prove the fifth assertion. We have $p_2 = 1 = (\sqrt{\Delta}+1)^0$, $p_3 = s_1 < \sqrt{\Delta}+1$ by Exercise 6.18.15 and $p_{i+2} = s_i p_{i+1} + p_i < \sqrt{\Delta}(\sqrt{\Delta}+1)^{i-1} + (\sqrt{\Delta}+1)^{i-2} = (\sqrt{\Delta}^2 + \sqrt{\Delta}+1)(\sqrt{\Delta}+1)^{i-2} < (\sqrt{\Delta}+1)^i$. The last assertion is proved analogously. □

Corollary 6.10.9. *We have* $\operatorname{size} T_i = \mathrm{O}(i \operatorname{size} \Delta)$ *and* $i = \mathrm{O}(\operatorname{size} T_i)$ *for* $i \in Z_{>0}$*.* □

6.11 Deciding equivalence

Let f and f' be integral irreducible forms with the same discriminant. Then the equivalence of f and f' can be decided as follows. First, a reduced form $g = (A, B, C)$ with $A > 0$ in the equivalence class of f is computed. Then, the cycle of f' is computed. Each element of that cycle is compared to g. If the form g is found in that cycle, then f and f' are equivalent. Otherwise, they are not. Also, if $g = fU = f'U'$ with $U, U' \in \mathrm{GL}(2, \mathbb{Z})$, then $f' = fU(U')^{-1}$. The implementation details are left to the reader as Exercise 6.18.14.

Example 6.11.1. Let $f = (1, 7, -6)$ and $g = (3, 5, -4)$. Those forms are both reduced and have discriminant 73. The cycle of f was computed in Example 6.10.6. There, we have seen that g belongs to that cycle. Hence, f and g are equivalent. Since the length of the cycle of f is odd, the equivalence class of f is equal to the proper equivalence class of f. Hence, f and g are properly equivalent.

Example 6.11.2. Let $f = (1, 8, -3)$ and $g = (2, 6, -5)$. Those forms are both reduced and have discriminant 76. The cycle of f was computed in Example 6.10.7. There, we have seen that g belongs to that cycle. Hence, f and g

are equivalent. But g does not belong to the proper cycle of f. Hence, f and g are not properly equivalent.

6.12 The automorphism group

In this section we determine the automorphism group of an integral irreducible form $f = (a, b, c)$.

6.12.1 The structure

Let $U \in \mathrm{GL}(2, \mathbb{Z})$ such that $f_0 = fU$ is reduced and $f_0 = (A, B, C)$ with $A > 0$. Let $(f_0, \ldots, f_{l-1})$ be the cycle of f. Let T_l be the transformation from Section 6.10. Then $f_0 T_l = f_0$. Therefore, T_l is an automorphism of f_0. Also, $fUT_lU^{-1} = f$. Hence

$$T = UT_lU^{-1} \tag{6.34}$$

is an automorphism of f. We show that T generates an infinite subgroup of $\mathrm{Aut}(f)$.

Lemma 6.12.1. *The cyclic group* $\langle T \rangle = \left\{T^i : i \in \mathbb{Z}\right\}$ *is infinite and* $\det T = (-1)^l$ *where* l *is the length of the cycle of* f*.*

Proof. We have $T^i = UT_l^iU^{-1}$ for $i \in \mathbb{Z}$, and

$$T_l = \prod_{i=0}^{l-1} \begin{pmatrix} 0 & 1 \\ 1 & |s(f_i)| \end{pmatrix} . \tag{6.35}$$

It follows from (6.35) and from Lemma 6.10.8 that the entries of T possibly except for the entry in the first row and column are positive. This implies that the entries of T^i in the second row and column form a strictly increasing sequence. Hence, the matrices T^i, $i \geq 0$ are pairwise distinct. This implies that $\langle T \rangle$ is infinite. Also, the factors of the right hand side of (6.35) have determinant -1. Therefore, $\det T = (-1)^l$. □

Example 6.12.2. We compute the automorphism T for $f = (1, 7, -6)$ using the results from Example 6.10.6. We obtain

$$T = T_l = \begin{pmatrix} 193 & 1500 \\ 250 & 1943 \end{pmatrix} .$$

Also, $\det T = -1$.

Example 6.12.3. We compute the automorphism T for $f = (1, 8, -3)$ using the results from Example 6.10.7. We obtain

$$T = T_l = \begin{pmatrix} 14 & 117 \\ 39 & 326 \end{pmatrix} .$$

Also, $\det T = 1$.

Now we determine $\operatorname{Aut}(f)$ and $\operatorname{Aut}^+(f)$.

Theorem 6.12.4. *We have*

$$\operatorname{Aut}(f) = \langle -1 \rangle \times \langle T \rangle = \left\{ \pm T^i : i \in \mathbb{Z} \right\} .$$

Also, if the length l of the cycle of f is odd then the narrow automorphism group is $\operatorname{Aut}^+(f) = \langle -1 \rangle \times \langle T^2 \rangle$ *and if l is even, then* $\operatorname{Aut}^+(f) = \operatorname{Aut}(f)$.

Proof. Clearly, we have $\langle -1 \rangle \times \langle T \rangle \subset \operatorname{Aut}(f_0)$.

We prove the converse inclusion. We first assume that $U = I_2$, that is, $f = f_0$ and $T = T_l$. We denote by B_i the $\mathbb{R}$-bases from Section 6.8.3 with $f_{B_i} = af_i$, $i \in \mathbb{Z}$. Let $V \in \operatorname{Aut}(f_0)$. Then $f_{B_0 V} = af_0 V = af$ by Proposition 4.3.13. This form is reduced. Hence $B_0 V$ is a minimal basis of $L(B_0)$ by Corollary 6.8.7. By Theorem 6.8.10 there is $i \in \mathbb{Z}$ such that $B_0 V \in \left\{ \pm B_i \right\}$. Since $B_i = B_0 T_i$, this implies that $V \in \left\{ \pm T_i \right\}$. Also, since $f_i = fT_i = fV = f$, we have $i \equiv 0 \pmod{l}$ because of the minimality of l. So $V \in \left\{ \pm T_l^j \right\}$ for some integer j.

We prove the general case. Let $V \in \operatorname{GL}(2, \mathbb{Z})$. Then V is an automorphism of f if and only if $U^{-1}VU$ is an automorphism of f_0. Hence, $U^{-1}VU \in \left\{ \pm T_l^i : i \in \mathbb{Z} \right\}$. This, in turn, is true if and only if $V \in \left\{ \pm T^i : i \in \mathbb{Z} \right\}$.

Finally, we determine the narrow automorphism group of f. Clearly, we have $\operatorname{Aut}^+(f) \subset \operatorname{Aut}(f)$. The matrix T_l from above is the product of l matrices of norm -1. Hence, if l is even then $\det T = \det T_l = 1$. This implies that $\operatorname{Aut}^+(f) = \operatorname{Aut}(f)$ in this case. Let l be odd. Then $\det T = \det T_l = -1$. Then $\left\{ \pm T^{2i} : i \in \mathbb{Z} \right\} \subset \operatorname{Aut}^+(f)$. To prove the converse inclusion let $V \in \operatorname{Aut}^+(f)$. Then there is $s \in \{\pm 1\}$ and $i \in \mathbb{Z}$ with $V = sT^i$. Since $\det T = -1$ the exponent e must be even. □

Definition 6.12.5. *The automorphism T is called the* fundamental automorphism *of f.*

In Examples 6.12.2 and 6.12.3 we have computed fundamental automorphisms.

6.12.2 Solving the Pell equation

In Section 2.5.2 we have explained that there is a close connection between the automorphism group of a form and the solutions of the Pell equation

$$x^2 - \Delta y^2 = \pm 4 \tag{6.36}$$

It follows from Lemma 6.12.1 and from Theorem 2.5.5 that this Pell equation has infinitely many solutions since $\Delta > 0$. In this section we explain how to find all solutions of (6.36).

Definition 6.12.6. *The* fundamental solution *of the Pell equation* (6.36) *is the solution $(x, y) \in \mathbb{Z}^2$ with $x, y > 0$ and minimal y.*

We show that the fundamental solution of the Pell equation is closely related to the fundamental automorphism.

Proposition 6.12.7. *If (x, y) is the fundamental solution of the Pell equation* (6.36)*, then $U(f, x, y)$ is the fundamental automorphism of f.*

Proof. By Theorem 2.5.5 the automorphisms of f are exactly the matrices

$$\begin{pmatrix} (x - yb)/2 & -cy \\ ay & (x + yb)/2 \end{pmatrix} \tag{6.37}$$

where (x, y) is a solution of (6.36). By Theorem 6.12.4 those automorphisms are of the form $\pm T^i$ where T is the fundamental automorphism of f. If all the entries of T are positive then the entry ay in T^i is strictly increasing. So the fundamental solution of (6.36) yields the fundamental automorphism of f. If not all entries in T are positive, then (6.35) implies that $l = 1$, f is reduced, and

$$T = \begin{pmatrix} 0 & 1 \\ 1 & x \end{pmatrix} .$$

From (6.37) we obtain $f = (a, b, c) = (1, 1, -1)$ and $\Delta = 5$. For $\Delta = 5$ the fundamental solution of (6.36) is $(3, 1)$. That solution yields the fundamental automorphism of f. □

By Proposition 6.12.7, the fundamental solution of the Pell equation (6.36) can be easily computed from the fundamental automorphism of the principal form of discriminant Δ. Also, using Proposition 6.12.7 the fundamental automorphism of any other integral form of discriminant Δ can be computed from the fundamental solution of the Pell equation (6.12.7).

Example 6.12.8. From Example 6.12.2 we see that $(2136, 250)$ is the fundamental solution of the Pell equation $x^2 - 73y^2 = \pm 4$. In fact, we have $2136^2 - 73 \cdot 250^2 = -4$.

From Example 6.12.3 we see that $(340, 39)$ is the fundamental solution of the Pell equation $x^2 - 76y^2 = \pm 4$. In fact, we have $340^2 - 76 \cdot 39^2 = 4$.

The results of this section also show that the Pell equation (2.19) has infinitely many solutions if Δ is a positive discriminant which is not a square. The structure of the set of all those solutions will be studied in Section 8.3.

6.13 Complexity

Let f be an integral indefinite irreducible reduced form and let Δ be the discriminant of f. We analyze the complexity of the algorithms that were presented in the previous sections.

We first discuss the complexity of computing the cycle of f.

Proposition 6.13.1.

1. *If f is reduced, then the algorithm* `cycle`(f) *computes the cycle of f in time* $\mathrm{O}(l(\operatorname{size}\Delta)^2)$ *where l is the length of the cycle of f.*
2. *The equivalence of two reduced integral forms of discriminant Δ can be decided in time* $\mathrm{O}(l(\operatorname{size}\Delta)^2)$*.*

Proof. 1. The elements of the cycle of f are reduced. Lemma 6.2.7 implies that one application of the reduction operator takes time $O\big(\operatorname{size}\Delta)^2\big)$. Hence, the total running time is $\mathrm{O}(l(\operatorname{size}(\Delta)^2)$.

2. To decide the equivalence of two reduced forms it suffices to compute the cycle of one of the forms. □

This complexity of `cycle` is close to optimal since there are l elements in the cycle of f and it requires time $\Omega(l)$ to write them all down. Also, as experiments show, the cycle length l is typically of the order of magnitude $\sqrt{\Delta}$.

Next we study the computation of the fundamental solution of the Pell equation.

Proposition 6.13.2.

1. *The fundamental solution (x, y) of the Pell equation $X^2 - \Delta Y^2 = \pm 4$ can be computed in time* $\mathrm{O}\big((\operatorname{size}(y)\operatorname{size}(\Delta))^2\big)$*.*
2. *The fundamental automorphism of a reduced integral form of discriminant Δ can be computed in time* $\mathrm{O}\big((\operatorname{size}(y)\operatorname{size}(\Delta))^2\big)$*.*
3. *If f and f' are equivalent reduced integral forms, then a transformation $T \in \mathrm{GL}(2, \mathbb{Z})$ with $f' = fT$ can be computed in time* $\mathrm{O}\big((\operatorname{size}(y)\operatorname{size}(\Delta))^2\big)$*.*

Proof. 1. To compute the fundamental solution (x, y) of the Pell equation, we compute the transformation T_l from (6.32) where l is the length of the cycle of f. It follows from Proposition 6.12.7 that $y = q_l$ with q_l from (6.32). Lemma 6.10.8 implies

$$l = \mathrm{O}(\operatorname{size} y)\ . \tag{6.38}$$

Each reduction step takes time $\mathrm{O}((\operatorname{size}\Delta)^2)$. Also by Lemma 6.10.8 the sizes of the matrices T_i are bounded by $l \operatorname{size}\Delta$ and the size of each s_i is bounded by $\operatorname{size}\Delta$. Hence, computing T_{i+1} from T_i requires time $l(\operatorname{size}\Delta)^2$. So (6.38) implies that the computation of T_l and (x, y) takes time $\mathrm{O}\big((\operatorname{size}(y)\operatorname{size}(\Delta))^2\big)$.

2. By Proposition 6.12.7 the fundamental automorphism of a reduced integral form of discriminant Δ can be easily computed from the fundamental automorphism of the corresponding Pell equation. Hence, the assertion follows from 1.

3. The computation of the transformation T requires the computation of the cycle and the corresponding transformations. Therefore, the proof is analogous to the proof of 1. □

Proposition 6.13.2 shows that the fundamental solution of the Pell equation can be computed in quadratic time in the size of the input and output. In general, this output is very large. In order to reduce the running time it is necessary to find a shorter representation of that fundamental solution of the Pell equation.

6.14 Ambiguous cycles

The cycles computed in Examples 6.10.6 and 6.10.7 are symmetric. In this section we show that the reason for this symmetry is the fact that the corresponding equivalence classes are ambiguous. We let $f = (a, b, c)$ be an integral indefinite irreducible form. We define

$$\kappa(a, b, c) = (-c, b, -a) \ . \tag{6.39}$$

We have seen in Lemma 2.8.2 that the equivalence class of f is ambiguous if and only if f and $\kappa(f)$ are equivalent. The next lemma can be used to compute the cycle of $\kappa(f)$ from the cycle of f.

Lemma 6.14.1. *If $i \in \mathbb{Z}$, then $\kappa\rho^i(f) = \rho^{-i}\kappa(f)$.*

Proof. For $i \geq 1$ we prove the assertion by induction. It follows from the formulas for $\rho(f)$ and for $\rho^{-1}(f)$ in (6.11) and (6.20) that $\kappa\rho(f) = \rho^{-1}\kappa(f)$. Assume that the assertion is true for some $i \geq 1$. Then $\kappa\rho^{i+1}(f) = \rho^{-1}\kappa(\rho^i(f)) = \rho^{-i-1}\kappa(f)$.

To prove the assertion for $i < 0$ we use the fact that $\kappa^2(f) = f$. Let $i < 0$ then $\kappa\rho^i(f) = \kappa\rho^i\kappa\kappa(f) = \kappa\kappa\rho^{-i}\kappa(f) = \rho^{-i}\kappa(f)$. □

It follows from Lemma 6.14.1 that from the cycle of the form f the cycle of $\kappa(f)$ can be obtained as follows.

Corollary 6.14.2. *Let f be an integral irreducible indefinite form and let $(f_0, f_1, \dots, f_{l-1})$ be the cycle of f. Then $\big(\kappa(f_l), \kappa(f_{l-1}), \dots, \kappa(f_1)\big)$ is the cycle of $\kappa(f)$.* □

Example 6.14.3. Consider the form $f = (2, 17, -14)$. Its discriminant is 401 and it is reduced. The cycle of f is $\big((2, 17, -14), (14, 11, -5), (5, 19, -2)\big)$. The cycle of $\kappa(f) = (14, 17, -2)$ is $\big((14, 17, -2), (2, 19, -5), (5, 11, -14)\big)$.

If f is ambiguous, then f and $\kappa(f)$ are equivalent. Therefore, the cycle of f and and the cycle of $\kappa(f)$ are the same and this cycle is symmetric. We describe this symmetry more precisely.

Definition 6.14.4. *The form f is called* symmetric *if $\kappa(f) = f$, that is, if $f = (a, b, -a)$.*

If f is symmetric, then

$$\Delta(f) = b^2 + 4a^2 \ .$$

So from a symmetric form we obtain the representation of its discriminant as the sum of two squares. We recall that the form f is ambiguous, if $f = (a, ka, c)$ with an integer k. If f is of this form, then

$$\Delta = a(k^2a - 4c) \ .$$

This is a factorization of the discriminant of f. We characterize ambiguous forms in a cycle.

Lemma 6.14.5. *Let f be reduced. Then f is ambiguous if and only if $\tau\rho^{-1}(f) = \kappa(f)$.*

Proof. Let $t = t(f)$

Suppose that $\tau\rho^{-1}(f) = \kappa(f)$. Then

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

This implies $-b + 2at(f) = b$, hence $b = at(f)$. So f is ambiguous.

Conversely, let $f = (a, ka, c)$ with $k \in \mathbb{Z}$. Then

$$\tau\rho^{-1}(f) = (-at^2 + kat - c, -ka + 2at, -a). \tag{6.40}$$

This form is reduced. By Corollary 6.3.4 the form $(-a, -ka+2at, -at^2+kat-c)$ is also reduced. Hence, we have

$$\left|\sqrt{\Delta} - 2|a|\right| < -ka + 2at < \sqrt{\Delta} \ . \tag{6.41}$$

Since the form f is also reduced, we have

$$|\sqrt{\Delta} - 2|a|| < -ka < \sqrt{\Delta} \ . \tag{6.42}$$

It follows from (6.41) and (6.42) that $t = 0$. From (6.40) we obtain $\tau\rho^{-1}(f) = \kappa(f)$. □

Theorem 6.14.6. *Assume that the equivalence class of f is ambiguous. Let l be the length of the cycle of f and let $(f_0, \dots, f_{l-1})$ be that cycle where $f_i = (\tau\rho)^i(f_0)$, $i \in \mathbb{Z}$.*

1. *If the length l of the cycle of f is odd, then this cycle contains one ambiguous and one symmetric form. If f_0 is the ambiguous form in the cycle and $l = 2k+1$ with $k \geq 0$, then the cycle is $\big(\kappa(f_{k-1}), \dots, \kappa(f_0), f_0, f_1, \dots, f_k\big)$ and f_k is the symmetric form.*
2. *Let the length l of the cycle of f be even, that is, $l = 2k$, $k \geq 1$. Then this cycle contains two symmetric and no ambiguous or two ambiguous and no symmetric form. In the first case, the cycle is $\big(\kappa(f_{k-1}), \dots, \kappa(f_1), f_0, f_1, \dots, f_k\big)$ and f_0 and f_k are the symmetric forms. In the second case, the period is $\big(\kappa(f_{k-1}), \dots, \kappa(f_0), f_0, f_1, \dots, f_{k-1}\big)$ and f_0 and $\kappa(f_{k-1})$ are the ambiguous forms.*

Proof. Exercise 6.18.16. □

Example 6.14.7. The reduced indefinite forms of discriminant $\varDelta = 105$ can be grouped into two cycles, namely $\big((1,9,-6),(6,3,-4),(4,5,-5),(5,5,-4),(4,3,-6),(6,9,-1)\big)$ and $\big((2,7,-7),(7,7,-2),(2,9,-3),(3,9,-2)\big)$. The first cycle has even length and contains the ambiguous forms $(1,9,-6)$ and $(5,5,-4)$. From the second ambiguous form we obtain the divisor 5 of 105. The second cycle has also even length and contains the two ambiguous forms $(7,7,-2)$ and $(3,9,-2)$ from which we obtain the divisors 7 and 3 of 105. The knowledge of the ambiguous forms has given us the complete prime factorization of 105, namely $105 = 3 \cdot 5 \cdot 7$.

Since the Γ-orbit of any ambiguous integral indefinite irreducible form contains a reduced form (see Exercise 6.18.17), we can find the complete factorization of a positive discriminant Δ, that is not a square, by determining all reduced ambiguous forms of that discriminant.

6.15 Solution of the representation problem

We have seen in Section 6.11 how to decide the equivalence of two integral indefinite irreducible forms. Also, in Section 6.12 we have explained how to calculate the automorphism group of an integral indefinite irreducible form. Hence, we can solve the representation problem for such forms. This is illustrated in the next example.

Example 6.15.1. We determine the representations of 19 by $f = (1,7,-6)$. The discriminant of f is 73. Since $73 \equiv 16 = 4^2 \pmod{19}$, it follows that 73 is a quadratic residue mod 19. There are two Γ-orbits of forms $(19,b,c)$ of discriminant 73 namely the Γ-orbits of $(19,23,6)$ and of $(19,-23,6)$.

If we reduce $(19,23,6)$, then we obtain $(2,7,-3)$ which by Example 6.10.6 belongs to the proper cycle of f. So f and $(19,23,6)$ are properly equivalent and we find the transformation

$$(1,7,-6)\begin{pmatrix} 5 & 4 \\ 6 & 5 \end{pmatrix} = (19,23,6)\ .$$

So $\pm(5,4)$ is a representation of 19 by $(1,7,-6)$.

If we reduce $(19,-23,6)$, then we obtain the form $(2,5,-6)$. By Example 6.10.6 this form also belongs to the proper cycle of f and we find the transformation

$$(1,7,-6)\begin{pmatrix} 151703 & -57711 \\ 196506 & -74755 \end{pmatrix} = (19,-23,6)\ .$$

So $(151703,196506)$ is a representation of 19 by $(1,7,-6)$.

In Example 6.12.2 we have computed the fundamental automorphism T of the form $(1, 7, -6)$. That automorphism has determinant -1. By Theorem 6.12.4 the narrow automorphism group of $(1, 7, -6)$ is $\langle -1 \rangle \times \langle T^2 \rangle$ with

$$T^2 = \begin{pmatrix} 412249 & 3204000 \\ 534000 & 4150249 \end{pmatrix}$$

We obtain all representations of 19 by $(1, 7, -6)$ by multiplying the two representations that we found by all proper automorphisms of $(1, 7, -6)$.

6.16 Solving the minimum problem

We prove that reduction theory solves the minimum problem for integral indefinite forms. We use the following Lemma.

Lemma 6.16.1. *Any cycle or proper cycle of reduced indefinite irreducible forms of discriminant Δ contains a form (a, b, c) with $|a| < \sqrt{\Delta}/2$.*

Proof. It follows from Corollary 6.8.11 that a cycle consists of all reduced forms (a, b, c) with positive a in the equivalence class of the forms in that cycle. Let f be an indefinite irrational form of discriminant Δ. Let (a, b, c) be a reduced form in the cycle or the proper cycle of f. By Lemma 6.2.7 we have $|a| + |c| < \sqrt{\Delta}$ for any reduced form (a, b, c). So $|a| < \sqrt{\Delta}/2$ or $|c| < \sqrt{\Delta}/2$. If $|a| < \sqrt{\Delta}/2$, then we are done. Suppose that $|a| \geq \sqrt{\Delta}/2$. Then $|c| < \sqrt{\Delta}/2$. Also, the form $\rho(a, b, c) = (c, B, C)$ with $B, C \in \mathbb{Z}$ is in the proper cycle of f and $\tau\rho(f) = (-c, B, C)$ with $B, C \in \mathbb{Z}$ belongs to the cycle of f. This shows that also in the case $|a| \geq \sqrt{\Delta}/2$ both the cycle of f and the proper cycle of f contain a form (A, B, C) with $|A| < \sqrt{\Delta}/2$. □

Corollary 6.16.2. *The minimum of an integral indefinite irreducible form f is the absolute smallest integer that appears as an a in a form (a, b, c) in the proper cycle of f.*

Proof. Any a in a form (a, b, c) that is properly equivalent to f can be represented by f. Hence, Lemma 6.16.1 shows that the minimum of f is smaller than $\sqrt{\Delta}/2$. Also, if a is the minimum of f, then there is a form (a, b, c) or a form $(-a, b, -c)$ that is properly equivalent to f. That form is reduced by Lemma 6.5.1. Hence, it belongs to the proper cycle of f. □

Example 6.16.3. We determine the minimum of the form $f = (5, 4, -3)$. In Example 6.10.7 we have seen that the proper cycle of f contains the form $(1, 8, -3)$. Hence, the minimum of f is 1.

6.17 Class number

From Corollary 6.5.4 and Corollary 6.2.8 we obtain the following result.

Theorem 6.17.1. *The number of equivalence classes of integral indefinite forms of a fixed discriminant is finite.* □

Definition 6.17.2. *Let Δ be a positive integer. The number of equivalence classes of primitive integral forms of discriminant Δ is called the* class number *of Δ. It is denoted by $h(\Delta)$. The number of proper equivalence classes of integral forms of discriminant Δ is called the* narrow class number *of Δ. It is denoted by $h^+(\Delta)$.*

For negative discriminants the class number $h(\Delta)$ can be computed by counting the number or primitive reduced forms of that discriminant. For positive discriminants we use the following theorem.

Theorem 6.17.3. *The class number $h(\Delta)$ is the number of distinct cycles of integral primitive forms (a, b, c) with $0 < a \leq |c|$ and discriminant Δ. If the length of those cycles is even then $h(\Delta) = 2h^+(\Delta)$. If the length of those cycles is odd then $h(\Delta) = h^+(\Delta)$.*

Proof. It follows from Corollary 6.8.11 that a cycle of reduced forms consists of all reduced forms (a, b, c) with positive a in the equivalence class of the forms in that cycle. By Lemma 6.16.1, every cycle of reduced forms of discriminant Δ contains a form (a, b, c) with $0 < a < \sqrt{\Delta}/2$. To determine the class number $h(\Delta)$ it therefore suffices to count the number of cycles of such forms. The assertion concerning the narrow class number follows from Proposition 6.10.5. □

Based on Theorem 6.17.3, we obtain the following method for computing the class number $h(\Delta)$ and the narrow class number $h^+(\Delta)$. We first list all primitive reduced forms (a, b, c) with $0 < a \leq |c|$ and discriminant Δ. We order this set, for example lexicographically, to make binary search possible. Then we pick the first element f in that ordered set, we determine its cycle and remove its elements from the list if possible. Repeating this process until the list is empty, we find all cycles. By Theorem 6.17.3, their number is $h(\Delta)$. If the cycle lengths are even then the narrow class number is $2h(\Delta)$. If the cycle lengths are odd then the narrow class number is $h(\Delta)$.

Example 6.17.4. Let $\Delta = 73$. By Example 6.7.2 we find the list of reduced forms

$$(1, 7, -6), (2, 5, -6), (2, 7, -3), (4, 3, -4), (3, 5, -4) .$$

Those forms have been written down in lexicographical ordering. The smallest element in this sequence is $(1, 7, -6)$. By Example 6.10.6 the cycle of $(1, 7, -6)$ contains all reduced forms in the list. Hence the class number of 73 is 1. Also, the length of the cycle of L is odd. Therefore the narrow class number of 73 is also 1.

6.18 Exercises

Exercise 6.18.1. Prove the second assertion of Lemma 6.1.5.

Exercise 6.18.2. Prove Lemma 6.8.4.

Exercise 6.18.3. Determine the automorphism group of an integral indefinite reducible form.

Exercise 6.18.4. Find an algorithm that computes all integral primitive reduced forms of a fixed discriminant. Estimate its running time.

Exercise 6.18.5. Let f be an indefinite form such that $\theta(f)$ is a quadratic irrationality. Prove that the number of reduced forms in the equivalence class of f is finite.

Exercise 6.18.6. Find all the reduced integral forms of discriminant $5, 8, 12, 13$.

Exercise 6.18.7. Let (α, γ) be a basis of an irrational lattice L in A and let α be a minimal point of L. Prove that $\Delta(\gamma/\alpha) > 1$.

Exercise 6.18.8. Prove Corollary 6.5.5.

Exercise 6.18.9. Prove Lemma 6.8.5.

Exercise 6.18.10. Show that any minimal point of an irrational lattice L in A can be supplemented to a reduced basis of L.

Exercise 6.18.11. Assume that Δ is a positive integer which is not a perfect square. Show that the number of integral reduced forms of discriminant Δ is even.

Exercise 6.18.12. Use the geometric interpretation of the reduction operator for bases to prove (6.21).

Exercise 6.18.13. Develop an algorithm that determines the primitive integral reduced forms of all non square discriminants Δ with $0 < \Delta \leq D$ where D is some given bound.

Exercise 6.18.14. Find an algorithm that decides the equivalence of two integral indefinite irreducible forms f and f' and that returns $T \in \mathrm{GL}(2, \mathbb{Z})$ with $f' = fT$ if f and f' are equivalent. Also, find an algorithm that decides the proper equivalence of two integral indefinite irreducible forms f and f' and that returns $T \in \mathrm{SL}(2, \mathbb{Z})$ with $f' = fT$ if f and f' properly are equivalent.

Exercise 6.18.15. Let f be an irrational indefinite reduced form. Prove that $1 \leq |s(f)| < \sqrt{\Delta}$.

Exercise 6.18.16. Prove Theorem 6.14.6.

Exercise 6.18.17. Prove that the Γ-orbit of any ambiguous integral indefinite irreducible form contains a reduced form.

Exercise 6.18.18. Use the notation of Section 6.8.3 and prove that we have $|\sigma(\mu_{i+1,1})| > |\sigma(\mu_{i,1})|$ and $|\sigma(\mu_{i+2,1})| < 2|\sigma(\mu_{i,1})|$ for $i \in \mathbb{Z}$.

Chapter references and further reading

[BB99] Ingrid Biehl and Johannes Buchmann, *An analysis of the reduction algorithms for binary quadratic forms*, Voronoi's Impact on Modern Science, Institute of Mathematics Kyiv (Peter Engel and Halyna M. Syta, eds.), National Academy of Sciences of Ukraine, 1999, pp. 71–98.

[Bue89] Duncan A. Buell, *Binary quadratic forms*, Springer, New York, 1989.

[JSW06] Michael J. Jacobson, Jr., Reginald E. Sawilla, and Hugh C. Williams, *Efficient ideal reduction in quadratic fields.*, International Journal of Mathematics and Computer Science **1** (2006), no. 1, 83–116 (English).

[Lag80] Jeffrey C. Lagarias, *Worst-case complexity bounds for algorithms in the theory of integral quadratic forms*, Journal of Algorithms **1** (1980), 142–186.

[Sch91] Arnold Schönhage, *Fast reduction and composition of binary quadratic forms*, International Symposium on Symbolic and Algebraic Computation, ISSAC '91 (Stephen M. Watt, ed.), ACM Press, 1991, pp. 128–133.

7

Multiplicative Lattices

Let $j \in \{\pm 1\}$, $A = A_j$, and $i = i(j)$. In this chapter we define the product of lattices in A and characterize the two-dimensional lattices in A whose product is a lattice. By a form we mean an irrational form with real coefficients and non-zero discriminant. By an *integral discriminant* we mean an integer Δ with $\Delta \equiv 0, 1 \bmod 4$ which is not a square in $\mathbb{Z}$.

7.1 Lattice operations

Let L, M, and K be additive subgroups of A. We define sum, product and quotient of those groups.

Definition 7.1.1.

1. *The* sum *of L and M is $L + M = \{\alpha + \beta : \alpha \in L, \beta \in M\}$.*
2. *The* product *of L and M is the additive subgroup of A generated by all products $\alpha\beta$, $\alpha \in L$, $\beta \in L'$.*
3. *The* quotient *of L and M is $L : M = \{\alpha \in A : \alpha M \subset L\}$.*

Note that $L + M$, LM, and $L : M$ are additive subgroups of A (see Exercise 7.5.1).

We explain why we are particularly interested in the product of lattices. The question of whether a number can be represented by a form can be reduced to the problem whether a lattice contains a point of certain norm. Suppose that we want to know whether a lattice in A contains a point of norm n. If we can factor this lattice as $L = L_1 L_2$ with lattices L_1 and L_2, and if we can show that L_i contains a point of norm n_i, $i = 1, 2$ such that $n = n_1 n_2$, then L contains a point of norm n.

Before we give examples we first mention a few computing rules.

Proposition 7.1.2. *We have*

1. $(L + M) + K = L + (M + K)$

2. $L + M = M + L$,
3. $(LM)K = L(MK)$,
4. $LM = ML$,
5. $L(M + K) = LM + LK$,
6. $(L : M)(M : K) = L : K$,

Proof. Exercise 8.7.1.

We explain how the product of two-dimensional lattices in A is determined.

Proposition 7.1.3. *If L_1 and L_2 are two-dimensional lattices in A and if $B_k = (\alpha_k, \gamma_k)$ is a basis of L_k, $k = 1, 2$ then $L_1L_2 = \mathbb{Z}\alpha_1\alpha_2 + \mathbb{Z}\alpha_1\gamma_2 + \mathbb{Z}\alpha_2\gamma_1 + \mathbb{Z}\gamma_1\gamma_2$.*

Proof. Since the products $\alpha_1\alpha_2, \alpha_1\gamma_2, \alpha_2\gamma_1, \gamma_1\gamma_2$ belong to L_1L_2 we have $M = \mathbb{Z}\alpha_1\alpha_2 + \mathbb{Z}\alpha_1\gamma_2 + \mathbb{Z}\alpha_2\gamma_1 + \mathbb{Z}\gamma_1\gamma_2 \subset L_1L_2$. Conversely, let $\theta \in L_1L_2$. Then $\theta = \sum_{(\delta,\varepsilon)\in S} \delta\varepsilon$ where S is a finite subset of $L_1 \times L_2$. Any element $\delta \in L_1$ can be written as $\delta = x_1\alpha_1 + y_1\gamma_1$ with integers x_1, y_1. Also, any element $\varepsilon \in L_2$ can be written as $\varepsilon = x_2\alpha_2 + y_2\gamma_2$ with integers x_2, y_2. Their product belongs to M. Since M is an additive group, we see that $L_1L_2 \subset M$. □

Example 7.1.4. Let $L = 2\mathbb{Z} + \sqrt{-2}\mathbb{Z}$, $M = 3\mathbb{Z} + \sqrt{-2}\mathbb{Z}$. We claim that $L + M = K = \mathbb{Z} + \sqrt{-2}\mathbb{Z}$. We have $L, M \subset K$ and therefore $L + M \subset K$. To show that $K \subset L + M$ it suffices to prove that $1 \in L + M$. Since $1 = 3 - 2$, this is true.

We present an example of two lattices whose product is a lattice.

Example 7.1.5. Let $L_1 = L(2, 2, 1) = 2\mathbb{Z} + (1 + \sqrt{-1})\mathbb{Z}$ and $L_2 = (5, 4, 1) = 5\mathbb{Z} + (2 + \sqrt{-1})\mathbb{Z}$. Then $L_1L_2 = 10\mathbb{Z} + 2(2 + \sqrt{-1})\mathbb{Z} + 5(1 + \sqrt{-1})\mathbb{Z} + (1 + 3\sqrt{-1})\mathbb{Z} = 10\mathbb{Z} + (3 + \sqrt{-1})\mathbb{Z} = L(10, 6, 1)$. So the product of L_1 and L_2 is the lattice $L = L(10, 6, 1)$.

In Example 7.1.5 we have seen two lattices whose product is a lattice. In general, the product of two lattices is not a lattice, as we will see in the next example.

Example 7.1.6. Let $L_1 = \mathbb{Z} + \sqrt{-1}\mathbb{Z}$ and $L_2 = \mathbb{Z} + \sqrt{-2}\mathbb{Z}$. Then $L_1L_2 = \mathbb{Z} + \sqrt{-1}\mathbb{Z} + \sqrt{-2}\mathbb{Z} + \sqrt{2}\mathbb{Z}$. We show that L_1L_2 is not a lattice. Assume that L_1L_2 is a lattice. Let $a = \min(L_1L_2 \cap \mathbb{R}_{>0})$. Then $1 = xa$ and $\sqrt{2} = ya$ with $x, y \in \mathbb{Z}$. Hence, $\sqrt{2} = y/x$. This contradicts the irrationality of $\sqrt{2}$.

7.2 Quadratic orders

To characterize the irrational lattices in A whose product is a lattice we introduce quadratic orders.

7.2.1 Basics

We introduce quadratic orders.

Definition 7.2.1. *A quadratic order is a two-dimensional lattice in A which is also a unitary subring of A.*

We characterize the quadratic orders. We recall that for an integer Δ, $\Delta \equiv 0, 1 \bmod 4$, there is exactly one reduced form $(1, b, c)$ of discriminant Δ which is called the *principal form* of discriminant Δ.

Theorem 7.2.2. *The quadratic orders are exactly the lattices $L(f)$ where f is a principal form.*

Proof. It is easy to verify that $L(1, b, c)$ is a unitary subring of A for any integral irreducible form $(1, b, c)$.

Conversely, assume that $\mathcal{O}$ is a quadratic order. Then $1 \in \mathcal{O}$. Since $\mathcal{O}$ is a ring and discrete as a point set, it follows that $1 = \min(\mathbb{R}_{>0} \cap \mathcal{O})$. Hence, there is a $\mathbb{Z}$-basis $(1, \theta)$ of $\mathcal{O}$ of positive orientation. Since $\mathcal{O}$ is a ring, we have $\theta^2 = -b\theta - c$ with integers b, c. It follows that $\theta = \theta(1, b, c)$ and $\mathcal{O} = L(1, b, c)$. By Theorem 4.4.4 we have $\mathcal{O} = \mathcal{O}(f)$ with the principal form f of discriminant $\Delta = b^2 - 4c$. □

Definition 7.2.3. *Let $\mathcal{O}$ be a quadratic order, $\mathcal{O} = \mathcal{O}(f)$ with a principal form f. The* discriminant *of $\mathcal{O}$ is $\Delta(\mathcal{O}) = \Delta(f)$. If $\Delta(\mathcal{O}) > 0$, then $\mathcal{O}$ is called a* real quadratic order*. If $\Delta(\mathcal{O}) < 0$, then $\mathcal{O}$ is called an* imaginary quadratic order*.*

If $\Delta \in \mathbb{Z}$, $\Delta \equiv 0, 1 \bmod 4$, then there is exactly one principal form of discriminant Δ. Therefore, there is exactly one quadratic order of discriminant Δ which we denote by $\mathcal{O}_\Delta$ and we have

$$\mathcal{O}_\Delta = \mathbb{Z} + \theta_\Delta \mathbb{Z} \tag{7.1}$$

with

$$\theta_\Delta = \theta(1, \Delta, \frac{\Delta^2 - \Delta}{4}) = \frac{\Delta + i(\operatorname{sign} \Delta)\sqrt{|\Delta|}}{2}. \tag{7.2}$$

For $x, y \in \mathbb{R}$ we have

$$\operatorname{Tr}(x + y\theta_\Delta) = 2x + \Delta y \ , \qquad \operatorname{N}(x + y\theta_\Delta) = x^2 + xy\Delta + y^2\frac{\Delta^2 - \Delta}{4} \tag{7.3}$$

and

$$\operatorname{o}(x + y\theta_\Delta) = \operatorname{sign}(y) \ , \qquad \Delta(x + y\theta_\Delta) = y^2\Delta \ . \tag{7.4}$$

Example 7.2.4. The quadratic order of discriminant -4 is

$$\mathcal{O}_{-4} = \mathbb{Z} + \mathbb{Z}\sqrt{-1} \ .$$

The quadratic order of discriminant 5 is

$$\mathcal{O}_5 = \mathbb{Z} + \mathbb{Z}(\frac{1 + \sqrt{5}}{2}, \frac{1 - \sqrt{5}}{2}) \ .$$

Example 7.2.5. Let Δ be an integral discriminant, let $\mathcal{O} = \mathcal{O}_\Delta$, let $\theta = \theta_\Delta$, and let $i = i(\operatorname{sign}\Delta)$. We show that $\mathcal{O}\mathcal{O} = \mathcal{O}$. By (7.1) both $(1, \theta)$ and $(1, \sigma(\theta))$ are bases of $\mathcal{O}$. Hence, over $\mathbb{Z}$ the product $\mathcal{O}\mathcal{O}$ is generated by 1, θ, and $\theta\sigma(\theta) = \mathrm{N}(\theta) = (\Delta^2 - \Delta)/4$. Therefore, $\mathcal{O}\mathcal{O} = \mathcal{O}$.

7.2.2 Maximal orders

In this section we prove that every quadratic order in A is contained in a unique order which is maximal with respect to inclusion.

Lemma 7.2.6. *Let Δ be an integral discriminant and let d be a positive integer. Then*

$$\mathcal{O}_{d^2\Delta} = \mathbb{Z} + d\mathcal{O}_\Delta = \mathbb{Z} + \mathbb{Z}d\theta_\Delta\ .$$

Proof. We have

$$\mathcal{O}_{d^2\Delta} = \mathbb{Z} + \mathbb{Z}\theta_{d^2\Delta} = \mathbb{Z} + \mathbb{Z}\big(d\theta_\Delta + \frac{\Delta d(d-1)}{2}\big) = \mathbb{Z} + \mathbb{Z}d\theta_\Delta = \mathbb{Z} + d\mathcal{O}_\Delta$$

since

$$\frac{\Delta d(d-1)}{2} \in \mathbb{Z}.$$

□

Proposition 7.2.7. *A quadratic order $\mathcal{O}$ is contained in a quadratic $\mathcal{O}'$ if and only if $\Delta(\mathcal{O}') = d^2\Delta(\mathcal{O})$ with a positive integer d.*

Proof. Write $\Delta = \Delta(\mathcal{O})$ and $\Delta' = \Delta(\mathcal{O}')$. Assume that $\Delta' = d^2\Delta$ with a positive integer d. Then Lemma 7.2.6 implies $\mathcal{O}' \subset \mathcal{O}$.

Conversely, let $\mathcal{O}' \subset \mathcal{O} \subset A$. Then $\theta_{\Delta'} \in \mathcal{O} = \mathbb{Z} + \mathbb{Z}\theta_\Delta$. This implies $\operatorname{sign}(\Delta) = \operatorname{sign}(\Delta')$ and with $i = i(\operatorname{sign}\Delta)$ we have

$$\frac{\Delta' + i\sqrt{|\Delta'|}}{2} = c + d\frac{\Delta + i\sqrt{|\Delta|}}{2}$$

with $c, d \in \mathbb{Z}$. So $\Delta' = d^2\Delta$. □

Example 7.2.8. The order $\mathcal{O}_{20}$ is contained in the order $\mathcal{O}_5$ since $20 = 2^2 * 5$ and 5, being congruent to 1 mod 4, is the discriminant of an order. However, the order $\mathcal{O}_{-20}$ is not contained in a larger order since $-20 = 2^2 * (-5)$ is the only way of writing $-20 = d^2D$ with an integer $d > 0$; but -5 is not the discriminant of an order since $-5 \equiv 3 \bmod 4$.

It follows from Proposition 7.2.7 that a quadratic order $\mathcal{O}$ is not properly contained in any other order if and only if $\Delta(\mathcal{O})$ is a fundamental discriminant (see Definition 3.3.2). This gives rise to the following definition.

Definition 7.2.9.

1. *The* conductor *of a quadratic order is the conductor of its discriminant. It is denoted by $f(\mathcal{O})$.*
2. *A quadratic order is called* maximal *if its discriminant is a fundamental discriminant.*

Theorem 7.2.10. *Any quadratic order $\mathcal{O}$ is contained in a unique maximal order.*

Proof. Let $\mathcal{O}$ be a quadratic order. As we have seen in Section 3.3 there is exactly one fundamental discriminant Δ such that $\Delta(\mathcal{O}) = d^2\Delta$ with a positive integer d, namely $\Delta = \Delta(\mathcal{O})/f(\mathcal{O})^2$. So the assertion follows from Proposition 7.2.7. □

Example 7.2.11. Consider the order $\mathcal{O}_{51005}$. Its discriminant 51005 is not fundamental since $51005 = 101^2 \cdot 5$. Hence, $\mathcal{O}_{51005}$ is contained in the maximal order $\mathcal{O}_5$.

7.3 Multiplicative lattices

In this section we characterize the irrational lattices in A whose product is a lattice in A.

7.3.1 Ring of multipliers

We introduce the ring of multipliers of a lattice L in A. By Exercise 7.5.4, the quotient $L : L$ is a unitary subring of A.

Definition 7.3.1. *The* ring of multipliers *of a lattice L in A is defined as $L : L$. This ring is denoted by $\mathcal{O}(L)$.*

Example 7.3.2. We claim that the ring of multipliers of $\mathbb{Z}$ is $\mathbb{Z}$. Clearly, we have $\mathbb{Z} \subset \mathbb{Z} : \mathbb{Z}$. Conversely, let $\alpha \in \mathbb{Z}$ with $\alpha\mathbb{Z} \subset \mathbb{Z}$. Then we have in particular $\alpha = \alpha \cdot 1 \in \mathbb{Z}$. Hence $\mathbb{Z} : \mathbb{Z} \subset \mathbb{Z}$. So we have shown that $\mathbb{Z} : \mathbb{Z} = \mathbb{Z}$.

Example 7.3.3. Let $\mathcal{O}$ be a quadratic order. We claim that $\mathcal{O}(\mathcal{O}) = \mathcal{O}$. From Example 7.2.5 we know that $\mathcal{O}\mathcal{O} = \mathcal{O}$. Hence, $\mathcal{O} \subset \mathcal{O}(\mathcal{O})$. Conversely, if $\alpha \in \mathcal{O}(\mathcal{O})$ then $\alpha\mathcal{O} \subset \mathcal{O}$. In particular, we have $\alpha = \alpha \cdot 1 \in \mathcal{O}$. So $\mathcal{O}(\mathcal{O}) \subset \mathcal{O}$.

If L is a lattice in A, then we have (see Exercise 7.5.5)

$$\mathcal{O}(L)L = L. \tag{7.5}$$

Also, the ring of multipliers of L does not change, if L is multiplied with a unit in A, that is,

$$\mathcal{O}(L) = \mathcal{O}(\alpha L)\ , \quad \alpha \in A^*. \tag{7.6}$$

In Example 7.3.2 we have seen that the ring of multipliers of the quadratic lattice $L(2,2,1)$ is the quadratic order $\mathcal{O}_{-4}$. We will now show that, in general, the ring of multipliers of an irrational lattice in A is either $\mathbb{Z}$ or a quadratic order.

Theorem 7.3.4. *Let L be an irrational lattice in A. If L is equivalent to a lattice $L(f)$ for some integral primitive irreducible form f then $\mathcal{O}(L) = \mathcal{O}_{\Delta(f)}$. Otherwise, $\mathcal{O}(L) = \mathbb{Z}$.*

Proof. We first prove that $\mathcal{O}(L)$ is a lattice. Choose a $\mathbb{Z}$-basis (α, γ) of L. Since L is irrational we have $\alpha, \gamma \in A^*$. Also, (7.6) implies $\mathcal{O}(L) = \mathcal{O}(L(\theta))$ with $\theta = \gamma/\alpha$. Since $1 \in L(\theta)$ we obtain

$$\mathcal{O}(L) = \mathcal{O}(L(\theta)) \subset \mathcal{O}(L(\theta))L(\theta) = L(\theta). \tag{7.7}$$

Corollary A.4.8 implies that $\mathcal{O}(L)$ is a lattice.

Next, we show that the ring of multipliers of $L = L(a,b,c)$, where $f = (a,b,c)$ is an integral primitive irreducible form of discriminant Δ, is $\mathcal{O}_\Delta$. Write $\theta = \theta(a,b,c)$. By (7.6) we have $\mathcal{O}(L) = \mathcal{O}(L(\theta))$. Since $a\theta^2 = -b\theta - c$, it follows that $a\theta$ is a multiplier of $L(\theta)$. Hence, the lattice $\mathcal{O}(L)$ is two-dimensional. Choose a $\mathbb{Z}$-basis $(1, \omega)$ of $\mathcal{O}(L)$. By (7.7) we can write $\omega = x + y\theta$ with $x, y \in \mathbb{Z}$. Without loss of generality, we assume that $x = 0$. Then $\omega = y\theta$ and y is the smallest positive integer such that $y\theta$ is a multiplier of $L(\theta)$. We know that $y\theta$ is a multiplier of $L(\theta)$ if and only if $y\theta^2 \in L(\theta)$. Now $y\theta^2 = -y(b\theta + c)/a$. Since f is primitive, the smallest y for which this expression is in $L(\theta)$ is $y = |a|$ which implies $\mathcal{O}(L) = \mathcal{O}_\Delta$.

Now we prove that $\mathcal{O}(L) = \mathbb{Z}$ unless L is equivalent to a lattice $L(f)$ with a primitive integral irreducible form f.

We know that $\mathcal{O}(L)$ is a lattice. Assume that $\mathcal{O}(L)$ is two-dimensional. Then by definition, $\mathcal{O}(L)$ is a quadratic order. By (7.7) we have $L(\theta) \subset (1/d)\mathcal{O}(L)$ for some positive integer d. By Theorem 7.2.2 and Exercise 4.7.17 the elements of $\mathcal{O}(L)$ are rational numbers or quadratic irrationalities. In particular, θ is a quadratic irrationality. By Theorem 4.5.8 there is a uniquely determined primitive integral irreducible form $f = (a,b,c)$ with $\theta = \theta(a,b,c)$ and $L = (\alpha/a)L(f)$. Hence, L is equivalent to $L(f)$. It follows that $\mathcal{O}(L)$ is one-dimensional, unless L is equivalent to a lattice $L(f)$ with a primitive integral irreducible form f.

Assume that $\mathcal{O}(L)$ is a one-dimensional lattice, $\mathcal{O}(L) = \alpha\mathbb{Z}$ for some $\alpha \in A^*$. Then $1 = x\alpha$ for some integer x. This implies $\alpha \in \mathbb{Q}$ and, therefore, $\mathcal{O}(L) \subset \mathbb{R}$. Since $\mathcal{O}(L)$ is discrete as a point set, we have $\mathcal{O}(L) = \mathbb{Z}$. □

It follows from the proof of Theorem 7.3.4 that the ring of multipliers of an irrational lattice L can be determined as follows: Find a basis $B = (\alpha, \gamma)$ of L. Set $\theta = \gamma/\alpha$. Then $L = \alpha L(\theta)$. Determine f_θ. If this polynomial has an irrational coefficient, then the ring of multipliers of L is $\mathbb{Z}$. Otherwise, multiply f_θ by the least common multiple of the denominators of its coefficients. The result is the integral primitive irreducible polynomial $f = (a, b, c)$. Compute the discriminant Δ of f. The ring of multipliers of L is $\mathcal{O}_\Delta$.

Example 7.3.5. We determine the ring of multipliers of $L = \mathbb{Z}\sqrt{3} + \mathbb{Z}(\sqrt{3} + \sqrt{-3})/2$. Since $L = (\sqrt{3}/2)L(2,2,1)$ and $\Delta(2,2,1) = -4$, it follows that $\mathcal{O}(L) = \mathcal{O}_{-4} = L(1,0,1) = \mathbb{Z} + \mathbb{Z}\sqrt{-1}$.

7.3.2 Irrational lattices whose product is a lattice

To characterize the lattices whose product is a lattice we need the following auxiliary result.

Lemma 7.3.6. *Let L, L' be irrational lattices in A such that $1 \in L'$. Then the following are true.*

1. $L \subset LL'$.
2. If LL' is a lattice, then there is a positive integer e with $eLL' \subset L$.

Proof. Since $1 \in L'$ we have $L \subset LL'$. Assume that LL' is a lattice. Since $L \subset LL'$, the lattice LL' is two-dimensional. If B is a basis of L and if B' is a basis of LL', then $B = B'U$ with a non-singular matrix $U \in \mathbb{Z}^{(2,2)}$. This implies $B' = (1/\det U)B \operatorname{adj}(U)$. So $(\det U)LL' \subset L$. □

Theorem 7.3.7. *If L, L' are irrational lattices in A, then their product LL' is a lattice if and only if their rings of multipliers are quadratic orders which are contained in the same maximal order.*

Proof. By (7.6) we may, without loss of generality, assume that $1 \in L$ and $1 \in L'$.

Suppose that $\mathcal{O}(L)$ and $\mathcal{O}(L')$ are quadratic orders which are contained in the maximal order $\mathcal{O}$. Then Lemma 7.3.6 implies that $eL = e\mathcal{O}(L)L \subset \mathcal{O}(L) \subset \mathcal{O}$ for some positive integer e. Likewise, we have $e'L' \subset \mathcal{O}$ for some positive integer e'. This implies $LL' \subset (1/ee')\mathcal{O}$. By Corollary A.4.8, the product LL' is a lattice.

Conversely, assume that LL' is a lattice. By Lemma 7.3.6 we have $eLL' \subset L$ for some positive integer e. Therefore, eL' is contained in the ring of multipliers of L. Theorem 7.3.4 implies that $\mathcal{O}(L)$ is a quadratic order. In the same way it is proved that $\mathcal{O}(L')$ is a quadratic order. Since $\mathcal{O}(L)LL' \subset LL'$, the ring of multipliers of L is contained in the ring of multipliers of LL'. Likewise, we have $\mathcal{O}(L') \subset \mathcal{O}(LL')$. Hence, the orders $\mathcal{O}(L)$ and $\mathcal{O}(L')$ are contained in the same maximal order. □

Example 7.3.8. Consider the lattices $L_1 = L(29,24,1845)$, $L_2 = L(17,12,3141)$, and $L_3 = (23,18,1136)$. We have $\Delta(29,24,1845) = -213444 = 231^2 * (-4)$, $\Delta(17,12,3141) = -68644 = 131^2(-4)$, $D(23,18,1136) = 1222 * (-7)$. Hence, $\mathcal{O}(L_1)$ and $\mathcal{O}(L_2)$ are contained in the maximal order $\mathcal{O}_{-4}$ and $\mathcal{O}(L_3)$ is contained in the maximal order $\mathcal{O}_{-7}$. By Theorem 7.3.7 the product L_1L_2 is a lattice but the products L_1L_3 and L_2L_3 are not.

Corollary 7.3.9. *Let $\mathcal{O}$ be a maximal quadratic order in A. Then the lattices in A whose ring of multipliers is $\mathcal{O}$ form a semigroup with respect to multiplication.*

We determine the product of two orders which are contained in the same maximal order.

Proposition 7.3.10. *Let Δ_1 and Δ_2 be discriminants with $\Delta_1/f(\Delta_1)^2 = \Delta_2/f(\Delta_2)^2$. Then $\mathcal{O}_{\Delta_1}\mathcal{O}_{\Delta_2} = \mathcal{O}_{\gcd(\Delta_1,\Delta_2)}$.*

Proof. Assume that $\Delta_1, \Delta_2 \equiv 0 \bmod 4$. Let f_k be the conductor of Δ_k and let $\Delta = \Delta_k/f_k^2$, $k = 1, 2$. Then $\mathcal{O}(\Delta_k) = \mathbb{Z} + \mathbb{Z} i f_k \sqrt{|\Delta|}/2$, $i = i(\operatorname{sign}\Delta)$, $k = 1, 2$ and $\mathcal{O}(\Delta_1)\mathcal{O}(\Delta_2) = \mathbb{Z} + \mathbb{Z} i f_1 \sqrt{|\Delta|}/2 + \mathbb{Z} i f_2 \sqrt{|\Delta|}/2 + \mathbb{Z}(f_1 f_2 \Delta)/4 = \mathbb{Z} + \mathbb{Z} i \gcd(f_1, f_2)\sqrt{|\Delta|}/2 = \mathcal{O}_{\gcd(\Delta_1,\Delta_2)}$. The other cases are proven analogously. □

7.3.3 The group $\mathcal{L}(\mathcal{O})$

Let $\mathcal{O}$ be a quadratic order in A and let Δ be the discriminant of $\mathcal{O}$. Denote by $\mathcal{L}(\mathcal{O})$ the set of all lattices with ring of multipliers $\mathcal{O}$. In this section we prove that $\mathcal{L}(\mathcal{O})$ is a multiplicative group.

In Example 7.2.5 we have seen that $\mathcal{O}\mathcal{O} = \mathcal{O}$. This implies that $\mathcal{O} \in \mathcal{L}(\mathcal{O})$. Also, if $L \in \mathcal{L}(\mathcal{O})$ then $\mathcal{O}L = L\mathcal{O} = L$. Hence, $\mathcal{O}$ acts as a neutral element with respect to multiplication.

Next, we show that any $L \in \mathcal{L}(\mathcal{O})$ has a multiplicative inverse.

Let L and L' be two-dimensional lattices in A. For any basis B of L and B' of L' there is a matrix $T \in \mathrm{GL}(2, \mathbb{R})$ such that $B = B'T$. The set of bases of L is $B\mathrm{GL}(2, \mathbb{Z})$ and the set of bases of L' is $B'\mathrm{GL}(2, \mathbb{Z})$. Therefore, $|\det T|$ is independent of the choice of B and B'. We define

$$[L : L'] = |\det T| \tag{7.8}$$

Note that $[L : L']$ is a positive real number. If L is contained in L' then $[L : L']$ is the index of the additive subgroup L in the additive group L'.

Proposition 7.3.11. *Let $L \in \mathcal{L}(\mathcal{O})$. Then we have $L\sigma(L) = [L : \mathcal{O}(L)]\mathcal{O}$.*

Proof. By Theorem 7.3.4, the lattice L can be written as $L = \alpha L(a, b, c)$ with an integral primitive form $f = (a, b, c)$ of discriminant Δ and $\alpha \in A^*$. Then $\sigma(L) = \sigma(\alpha)L(a, -b, c)$. Hence, $L\sigma(L) = \mathrm{N}(\alpha)L(a, b, c)L(a, -b, c)$. Set $\theta = (b + i\sqrt{|\Delta|})/2$, $i = i(\operatorname{sign}(\Delta))$. Over $\mathbb{Z}$, the lattice $L(a, b, c)L(a, -b, c)$ is generated by a^2, $a\theta$, $a\sigma(\theta)$, and $\theta\sigma(\theta) = ac$. Hence, another generating system is $a^2, a(\theta + \sigma(\theta)) = ab, ac, a\theta$. Since $\gcd(a, b, c) = 1$, we obtain the generating system a, $a\theta$ which is a basis of $a\mathcal{O}_\Delta$. By Exercise 7.5.7 we have $[L : \mathcal{O}] = |\mathrm{N}(\alpha)a|$. Hence, the proposition is proved. □

Proposition 7.3.11 justifies the following definition:

Definition 7.3.12. *The* inverse *of* $L \in \mathcal{L}(\mathcal{O})$ *is* $L^{-1} = (1/[L : \mathcal{O}])\sigma(L)$.

By Proposition 7.3.11 we have

$$LL^{-1} = \mathcal{O}, \quad L \in \mathcal{L}(\mathcal{O}). \tag{7.9}$$

Also, $L^{-1} \in \mathcal{L}(\mathcal{O})$. Hence, every lattice in $\mathcal{L}(\mathcal{O})$ has a multiplicative inverse in $\mathcal{L}(\mathcal{O})$.

If (a, b, c) is an integral primitive form of discriminant Δ then

$$L(a, b, c)^{-1} = (1/a)L(a, -b, c). \tag{7.10}$$

So the inverse of $L(a, b, c)$ can be computed extremely efficiently.

Finally, we prove that $\mathcal{L}(\mathcal{O})$ is closed under multiplication. In fact, we prove a slightly more general result.

Proposition 7.3.13. *Let L_1 and L_2 be irrational lattices and assume that $\mathcal{O}(L_1)$ and $\mathcal{O}(L_2)$ are quadratic orders which are contained in the same maximal order. Then $\mathcal{O}(L_1L_2) = \mathcal{O}(L_1)\mathcal{O}(L_2)$.*

Proof. Let $\alpha \in A^*$. Then $\alpha \in \mathcal{O}(L_1L_2)$ if and only if $\alpha L_1L_2 \subset L_1L_2$. If we multiply this inclusion by $L_1^{-1}L_2^{-1}$, then (7.9) implies that $\alpha \in \mathcal{O}(L_1L_2)$ if and only $\alpha\mathcal{O}(L_1)\mathcal{O}(L_2) \subset \mathcal{O}(L_1)\mathcal{O}(L_2)$. This, in turn, is true if and only if $\alpha \in \mathcal{O}(L_1)\mathcal{O}(L_2)$ since by Proposition 7.3.10 the product $\mathcal{O}(L_1)\mathcal{O}(L_2)$ is a quadratic order and by Example 7.3.3 the ring of multipliers of a quadratic order is that quadratic order. □

So we have proved the following theorem.

Theorem 7.3.14. *Let $\mathcal{O}$ be a quadratic order. Then the set $\mathcal{L}(\mathcal{O})$ of all lattices with ring of multipliers $\mathcal{O}$ is an Abelian group with respect to multiplication. The neutral element of that group is $\mathcal{O}$.*

7.3.4 Computing the product of lattices

Let $f_k = (a_k, b_k, c_k)$ be integral primitive forms with $a_k > 0$, and let $\Delta_k = \Delta(f_k)$, $k = 1, 2$. Assume that the product of $L_1 = L(f_1)$ and $L_2 = L(f_2)$ is a lattice. In this section we explicitly determine the product of L_1 and L_2.

Proposition 7.3.15. *We have* $[L_1L_2 : \mathcal{O}(L_1L_2)] = [L_1 : \mathcal{O}(L_1)][L_2 : \mathcal{O}(L_2)]$

Proof. It follows from Propositions 7.3.11 and 7.3.13 that $[L_1L_2 : \mathcal{O}(L_1L_2)]$ $\mathcal{O}(L_1L_2) = L_1\sigma(L_1)L_2\sigma(L_2) = [L_1 : \mathcal{O}(L_1)][L_2 : \mathcal{O}(L_2)]\mathcal{O}(L_1)\mathcal{O}(L_2) = [L_1 : \mathcal{O}(L_1)][L_2 : \mathcal{O}(L_2)]\mathcal{O}(L_1L_2)$. This proves the assertion. □

In the following theorem we present an explicit formula for $L(g_1)L(g_2)$.

Theorem 7.3.16. *We have*

$$L(a_1,b_1,c_1)L(a_2,b_2,c_2) = mL(a,b,c)$$

where m,a,b,c are determined as follows. Set

$$\Delta_k = \Delta(a_k,b_k,c_k), \quad d_k = \frac{f(\Delta_k)}{\gcd(f(\Delta_1),f(\Delta_2))}, \quad k=1,2.$$

Then

$$m = \gcd\bigl(d_1a_2, d_2a_1, \frac{d_1b_2+d_2b_1}{2}\bigr),$$

and

$$a = \frac{a_1a_2}{m^2}$$

Moreover, let $j,k,\ell \in \mathbb{Z}$ such that

$$jd_1a_2 + kd_2a_1 + l\frac{d_1b_2+d_2b_1}{2} = m.$$

Then $ja_2b_1 + ka_1b_2 + l(b_1b_2 + d_1d_2\Delta)/2$ is divisible by m and

$$b \equiv \frac{ja_2b_1 + ka_1b_2 + l(b_1b_2+d_1d_2\Delta)/2}{m} \mod 2a.$$

Also,

$$c = \frac{b^2-\Delta}{4a}.$$

Proof. By Propositions 7.3.13 and 7.3.10 the ring of multipliers of $L = L_1L_2$ is $\mathcal{O}_{d^2\Delta}$. By Exercise 7.5.6 we can write

$$L = rL(a,b,c) = r\bigl(a\mathbb{Z} + \mathbb{Z}\frac{b+i\sqrt{|\Delta|}}{2}\bigr) \tag{7.11}$$

with an integral primitive irreducible form (a,b,c) of discriminant $\Delta = \Delta_1/d_1^2 = \Delta_2/d_2^2$ and $r \in \mathbb{R}_{>0}$. We determine r, a, and b. Set $i = i(\operatorname{sign}\Delta)$. Over $\mathbb{Z}$, the lattice L is generated by

$$\begin{aligned}
&a_1a_2\\
\theta_1 &= \frac{a_2(b_1+id_1)\sqrt{|\Delta|}}{2}\\
\theta_2 &= \frac{a_1(b_2+id_2)\sqrt{|\Delta|}}{2}\\
\theta_3 &= \frac{(b_1b_2+d_1d_2\Delta)/2 + i((b_1d_2+b_2d_1)/2)\sqrt{|\Delta|}}{2}.
\end{aligned}$$

Now r is the smallest positive real number such that $(s+ir\sqrt{|\Delta|})/2$ belongs to L for some $s \in \mathbb{R}$. In view of the generators we see that $r = m$ with m from the Theorem. Moreover, it follows from (7.11) and Proposition 7.3.15 that $a_1a_2 = [L : \mathcal{O}(L)] = m^2a$. This implies $a = a_1a_2/m^2$. Finally, the formula for b follows from the uniqueness of b mod $2a$. □

We explain two important special cases of Theorem 7.3.16.

Corollary 7.3.17. *Assume that* $\Delta_1 = \Delta_2$*. Then we have*

$$L(g_1)L(g_2) = mL(a, b, c)$$

where m, a, b, c *are determined as follows.*

$$m = \gcd\big(a_2, a_1, \frac{b_1 + b_2}{2}\big),$$

and

$$a = \frac{a_1 a_2}{m^2}$$

Moreover, let $j, k, \ell \in \mathbb{Z}$ *such that*

$$ja_2 + ka_1 + l\frac{b_2 + b_1}{2} = m.$$

Then $ja_2b_1 + ka_1b_2 + l(b_1b_2 + \Delta)/2$ *is divisible by* m *and*

$$b \equiv \frac{ja_2b_1 + ka_1b_2 + l(b_1b_2 + \Delta)/2}{m} \mod 2a.$$

Also,

$$c = \frac{b^2 - \Delta}{4a}.$$

Corollary 7.3.18. *Assume that* $\Delta_1 = \Delta_2$ *and* $\gcd(a_1, a_2) = 1$*. Then we have*

$$L(g_1)L(g_2) = L(a_1a_2, b, c)$$

where b, c *are determined as follows. Let* $j, k \in \mathbb{Z}$ *such that*

$$ja_2 + ka_1 = 1.$$

Then

$$b \equiv ja_2b_1 + ka_1b_2 \mod 2a$$

and

$$c = (b^2 - \Delta)/(4a).$$

7.4 Composition of forms

In Section 4.6 we have seen that quadratic lattices can be identified with Γ-orbits of integral primitive irreducible forms. In the previous sections we have explained how to multiply quadratic lattices whose ring of multipliers is contained in the same maximal order. The corresponding operation on the Γ-orbits of forms is called *composition* . This is explained in this section.

Example 7.4.1. Let $f_1(X_1, Y_1) = X_1^2 + Y_1^2$ and $f_2(X_2, X_2) = X_2^2 + 2X_2Y_2 + 2Y_2^2$. Then $f_1(X_1, Y_1)f_2(X_2, Y_2) = X^2 + Y^2$ with $X = X_1X_2 + Y_2X_1 - Y_1Y_2$ and $Y = Y_1X_2 + Y_2X_1 + Y_1Y_2$.

In Example 7.4.1 we could write

$$f_1(X_1, Y_1)f_2(X_2, Y_2) = f(X, Y) \tag{7.12}$$

where f_1, f_2, f are forms and

$$(X, Y) = (X_1X_2, X_2Y_1, Y_2X_1, Y_1Y_2) \begin{pmatrix} p_0 & q_0 \\ p_1 & q_1 \\ p_2 & q_2 \\ p_3 & q_3 \end{pmatrix} \tag{7.13}$$

with integers p_i, q_i, $0 \leq i \leq 3$. If we have representations (x_i, y_i) of integers n_i by f_i, $i = 1, 2$, then (7.13) can be used to compute a representation of $n = n_1n_2$ by f.

Definition 7.4.2. *Forms f_1 and f_2 for which a transformation* (7.13) *and a form f exists such that* (7.12) *holds are called* composable.

We will now show how the theory of multiplicative lattices developed so far can be used to identify the composable integral irreducible forms We will also explain how to find the transformation in (7.13) and the form f if they exist.

Theorem 7.4.3. *Two integral primitive irreducible forms are composable if and only if their discriminant divided by the square of their conductor is the same fundamental discriminant.*

Proof. Let f_1 and f_2 be two integral primitive irreducible forms. Suppose that f_1 and f_2 are composable. Let f and p_i, q_i, $0 \leq i \leq 3$ be as in (7.12) and (7.13). Assume that the discriminant of f is $d^2\Delta$ with a positive integer d and a fundamental discriminant Δ. Also, let x_2, y_2 be integers such that $f_2(x_2, y_2) \neq 0$. Then

$$n_1 f_1(X_1, Y_1) = f(U_1(X_1, Y_1))$$

where $n_1 = f_2(x_2, y_2)$ and

$$U = \begin{pmatrix} p_0x_2 + p_1y_2, p_2x_2 + p_3y_2 \\ q_0x_2 + q_1y_2, q_2x_2 + q_3y_2 \end{pmatrix}.$$

The matrix U has integer entries. It follows that $\Delta(f_1) = \Delta d^2(\det U)^2/n_1^2$. Since Δ is a fundamental discriminant it follows that n_1 divides $d \det U$. Hence, the discriminant of f_1 divided by the square of its conductor is Δ. By the same argument we see that the discriminant of f_2 divided by the square of its conductor is Δ.

Conversely, assume that the discriminant of f_1 and f_2 divided by the square of their conductor is the same fundamental discriminant. Then by Theorem 7.3.16 we have $L(f_1)L(f_2) = mL(f)$ with $m \in \mathbb{N}$ and an integral primitive form $f = (a, b, c)$. Let $f_i = (a_i, b_i, c_i)$, $i = 1, 2$ and

$$\theta = a\theta(f)\ , \quad \theta_i = a_i\theta(f_i)\ , \quad i = 1, 2.$$

Then $a_1\theta_2$, $a_2\theta_1$, and $\theta_1\theta_2$ are contained in $mL(f)$. Hence, there are representations

$$a_2\theta_1 = m(p_1 a + q_1\theta), a_1\theta_2 = m(p_2 a + q_2\theta), \theta_1\theta_2 = m(p_3 a + q_3\theta) \qquad (7.14)$$

with integers p_i, q_i, $i = 1, 2, 3$. We also set $p_0 = m$, $q_0 = 0$. Then $mf(X, Y) = f_1(X_1, X_2)f_2(X_2, Y_2)$ holds with X, Y as in (7.13). □

The proof of Theorem 7.4.3 contains an algorithm for composing two forms. This is illustrated in the following example.

Example 7.4.4. Let $f_1 = (2, 2, 1)$, $f_2 = (5, 4, 1)$. Then $\Delta(f_1) = \Delta(f_2) = -4$. We use the notation from Theorem 7.3.16 and the proof of Theorem 7.4.3. We obtain $d_1 = d_2 = m = 1$, $a = 10$, $b = -6$, $c = 1$. Hence $f = (10, -6, 1)$. Now

$$\theta_1 = 1 + \sqrt{-1}, \quad \theta_2 = 2 + \sqrt{-1}.$$

and

$$\theta = -3 + \sqrt{-1}.$$

Also,

$$a_1\theta_2 = 4 + 2\sqrt{-1} = a + 2\theta, \quad a_2\theta_1 = 5 + 5\sqrt{-1} = a + 5\theta,$$

and

$$\theta_1\theta_2 = 1 + 3\sqrt{-1} = a + 3\theta.$$

If we set

$$\begin{pmatrix} p_0 & q_0 \\ p_1 & q_1 \\ p_2 & q_2 \\ p_3 & q_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 5 & 3 \end{pmatrix},$$

then (7.12) is satisfied.

Remark 7.4.5. Shanks proposed an efficient algorithm called NUCOMP which combines composing two forms and a (partial) reduction of the result [Sha89]. The basic idea is to compute a transform of the matrix M on the right of (7.13) with smaller entries. This is done by computing first two of the four rows of M, and then performing column operations on M to make its entries smaller, i.e. of size in $O(\sqrt[4]{|\Delta|})$. Such operations correspond to transformations of the resulting form which one expects to end up having coefficients in $O(\sqrt{|\Delta|})$ so that only few reduction steps are required to obtain a reduced form in the class of the composition of the two given forms. For details, see [vdP03].

7.5 Exercises

Exercise 7.5.1. Prove that the sum, product, and quotient of additive subgroups of A are additive subgroups of A.

Exercise 7.5.2. Prove Proposition 7.1.2.

Exercise 7.5.3. Let α be a real number and let a be an integer. Prove that α is an algebraic integer if and only $\alpha - a$ is an algebraic integer.

Exercise 7.5.4. Prove that for any lattice L in A the quotient $L : L$ is a unitary subring of A.

Exercise 7.5.5. Prove that for any lattice L in A we have $\mathcal{O}(L)L = L$ and $\mathcal{O}(\alpha L) = \mathcal{O}(L)$ for any $\alpha \in A^*$.

Exercise 7.5.6. Let $\mathcal{O}$ be a quadratic order and let L be a lattice with ring of multipliers $\mathcal{O}$. Prove that there is exactly an integral primitive irreducible form $f = (a, b, c)$ and a positive real number r such that $L = rL(f)$. Also prove that r is uniquely determined and that the Γ-orbit of f is uniquely determined.

Exercise 7.5.7. Let $f = (a, b, c)$ be an integral primitive irreducible form, let $\alpha \in A_{\operatorname{sign}\Delta(f)}$, and let $L = \alpha L(f)$. Prove that $[L : \mathcal{O}(L)] = |\mathrm{N}(\alpha)a|$.

Chapter references and further reading

[Len82] Hendrik W. Lenstra, Jr., *On the calculation of regulators and class numbers of quadratic fields*, Journées Arithmétiques 1980 (Cambridge) (J. V. Armitage, ed.), London Mathematical Society Lecture Note Series, vol. 56, Cambridge University Press, 1982, pp. 123–150.

[Sha89] Daniel Shanks, *On Gauss and composition I, II*, Number Theory and Applications, Calgary 1988 (Richard A. Mollin, ed.), NATO ASI Series, Series C, vol. 265, Kluwer Academic Publishers, 1989, pp. 163–178, 179–204.

[vdP03] Alfred van der Poorten, *A note on NUCOMP*, Mathematics of Computation **72** (2003), no. 244, 1935–1946 (electronic).

8

Quadratic Number Fields

We have seen in Chapter 7 that the set of all lattices whose ring of multipliers is contained in a fixed maximal order is a monoid with respect to multiplication of lattices. In this chapter we study the algebraic structure of this monoid more closely. We also discuss the properties of the fields of fractions of quadratic orders.

8.1 Basics

We fix a discriminant, that is, an integer Δ that is not a square in $\mathbb{Z}$ and that satisfies $\Delta \equiv 0, 1 \bmod 4$. We also set $i = i(\operatorname{sign}\Delta)$. The field of fractions of the order $\mathcal{O}_\Delta$ is

$$F = \mathbb{Q}(i\sqrt{|\Delta|}) = \mathbb{Q} + \mathbb{Q}i\sqrt{|\Delta|}. \tag{8.1}$$

Let $\Delta > 0$. We simplify F by embedding F into the field $\mathbb{R}$ of real numbers. By Corollary 4.5.7 the map

$$\mathbb{Q}(i\sqrt{\Delta}) \to \mathbb{Q}(\sqrt{\Delta}) = \mathbb{Q} + \mathbb{Q}\sqrt{\Delta}, \quad i\sqrt{\Delta} \mapsto \sqrt{\Delta} \tag{8.2}$$

is an isomorphism of fields. By virtue of this isomorphism, we identify $\mathbb{Q}(i\sqrt{\Delta})$ with $\mathbb{Q}(\sqrt{\Delta})$. The *conjugate* of an element $\alpha = x + y\sqrt{\Delta}$, $x, y \in \mathbb{Q}$ in $\mathbb{Q}(\sqrt{\Delta})$ is $\sigma(\alpha) = x - y\sqrt{\Delta}$. Sometimes it will be convenient to write α^σ for $\sigma(\alpha)$. By the *norm*, *trace*, or *characteristic polynomial* of an element in $\mathbb{Q}(\sqrt{\Delta})$ we mean the norm, trace, or characteristic polynomial of its inverse image in $\mathbb{Q}(i\sqrt{\Delta})$, respectively. The *discriminant* or *orientation* of a pair in $\mathbb{Q}(\sqrt{\Delta})$ that is linearly independent over $\mathbb{Q}$ is the discriminant or orientation of its inverse image in $\mathbb{Q}(i\sqrt{\Delta})$. For $\alpha = x + y\sqrt{\Delta}$, $x, y \in \mathbb{Q}$, we have $\mathrm{N}(\alpha) = x^2 - y^2\Delta$, $\mathrm{Tr}(\alpha) = 2x$, $\mathrm{o}(\alpha) = \mathrm{sign}(y)$, and $\Delta(\alpha) = 4y^2$.

Example 8.1.1. Let $\Delta = 5$. The field of fractions of $\mathcal{O}_5$ is $\mathbb{Q}(\sqrt{5}, -\sqrt{5})$. That field is identified with $\mathbb{Q}(\sqrt{5})$. We have $\mathrm{Tr}(1 + \sqrt{5}) = 2$, $\mathrm{N}(1 + \sqrt{5}) = -4$, $\Delta(\sqrt{5}) = 20$, $\mathrm{o}(\sqrt{5}) = 1$.

For $\Delta < 0$ and $\Delta > 0$ we can now write

$$\theta_\Delta = \frac{\Delta + \sqrt{\Delta}}{2}. \tag{8.3}$$

The field of fractions of $\mathcal{O}_\Delta$ is $\mathbb{Q}(\sqrt{\Delta})$. This field is a subfield of $\mathbb{C}$ and it is a two-dimensional $\mathbb{Q}$-vector space. We give a name to such fields.

Definition 8.1.2. *A* quadratic number field *is a subfield of* $\mathbb{C}$ *that is a two-dimensional* $\mathbb{Q}$*-vector space. It is called a* real quadratic number field *if it is a subfield of* $\mathbb{R}$ *and an* imaginary quadratic number field *otherwise.*

Example 8.1.3. The fields $\mathbb{Q}(\sqrt{5})$, $\mathbb{Q}(\sqrt{45})$, and $\mathbb{Q}(1+\sqrt{20})$ are quadratic number fields. They are, in fact, all equal.

We describe the quadratic orders that are contained in a fixed quadratic number field.

Theorem 8.1.4. *Let F be a quadratic number field. Then the following statements are true.*

1. *The field F contains exactly one maximal order.*
2. *A quadratic order is contained in F if and only if it is contained in the unique maximal order in F.*
3. *The field of fractions of all quadratic orders in F is F.*

Proof. We show that F is the field of fractions of a quadratic order. As a two-dimensional $\mathbb{Q}$-vector space, F has a $\mathbb{Q}$-basis $(1, \theta)$. Since F is a field, θ^2 must belong to F. Hence, there are rational numbers x, y such that $\theta^2 = x\theta + y$. It follows that θ is a quadratic irrationality. Let $\theta = \theta(f)$ be its standard representation with an integral irreducible form f (see Section 4.5). If Δ is the discriminant of f then $F = \mathbb{Q}(\sqrt{\Delta})$.

We prove the first and second assertion. Let $\Delta_0 = \Delta/f(\Delta)^2$. Then $F = \mathbb{Q}(\sqrt{\Delta}) = \mathbb{Q}(\sqrt{\Delta_0})$. Hence, the maximal order $\mathcal{O}_{\Delta_0}$ is contained in F. Let $\mathcal{O}'$ be another order that is contained in F and let Δ' be its discriminant. Then $\sqrt{\Delta'} \in F$. Hence, $\sqrt{\Delta'} = x + y\sqrt{\Delta_0}$ with rational numbers x, y. We square both sides of this equation and obtain $\Delta' = x^2 + y^2\Delta_0 + 2xy\sqrt{\Delta_0}$. This implies $x = 0$. Hence, $\Delta' = y^2\Delta_0$. Write $y = n/d$ with integers n, d, $d > 0$, $\gcd(n, d) = 1$. Then $d^2\Delta' = n^2\Delta_0$ and $n = f(d^2\Delta')$. Therefore, d divides n. So $\gcd(n, d) = 1$ implies $d = 1$ and therefore, $\mathcal{O}' \subset \mathcal{O}_{\Delta_0}$.

The proof of the third assertion is left to the reader as Exercise 8.7.4. □

It follows from Theorem 8.1.4 that the real quadratic number fields are exactly the fields of fractions of the real quadratic orders and the imaginary quadratic number fields are exactly the fields of fractions of the imaginary quadratic orders.

Definition 8.1.5. *Let F be a quadratic number field. We denote the maximal order, that is contained in F, by $\mathcal{O}_F$ and call it the* maximal order *of F. The discriminant of F is the discriminant of $\mathcal{O}_F$. We write $\Delta(F)$ for that discriminant.*

We describe the elements of quadratic number fields.

Proposition 8.1.6. *Any element of a quadratic number field is a quadratic irrationality or a rational number.*

Proof. Let F be a quadratic number field and let Δ be a discriminant such that F is the field of fractions of $\mathcal{O}_\Delta$. Trace and norm of any element θ in F is a rational number. It therefore follows from Lemma 4.5.4 that the element θ is a quadratic irrationality or a rational number. □

8.2 Algebraic integers

To give another characterization of the maximal order that is contained in a quadratic number field we introduce algebraic integers.

Definition 8.2.1. *An* algebraic integer *is a complex number that is a zero of a monic polynomial with integer coefficients.*

Example 8.2.2. Any integer z is an algebraic integer since it is the zero of the polynomial $X - z$. We show that the integers are the only algebraic integers in $\mathbb{Q}$. Let $q \in \mathbb{Q}$ be an algebraic integer and let p be a monic polynomial with integer coefficients such that $p(q) = 0$. Then $p(X) = (X - q)f(X)$ with $f \in \mathbb{Q}[X]$. Write $q = n/d$ with integers n, d, $d > 0$, $\gcd(n, d) = 1$, and let g be the least common multiple of the denominators of the coefficients of f. Then $dgp(X) = (dX - n)gf(X)$. Using Lemma A.3.2 we obtain $dg = \operatorname{cont}(dgp(X)) = \operatorname{cont}(dX - n)\operatorname{cont}(gf(X)) = 1$. Hence, $d = 1$.

We determine the quadratic irrationalities that are algebraic integers.

Proposition 8.2.3. *Let θ be a quadratic irrationality. Then the following statements are equivalent.*

1. *θ is an algebraic integer.*
2. *The trace and norm of θ are integers.*
3. *If $\theta = \theta(a, b, c)$ is the standard representation of θ, then $|a| = 1$.*

Proof. Assume that θ is an algebraic integer. Then there is a monic polynomial $f \in \mathbb{Z}[X]$ such that $f(\theta) = 0$. This implies $f(\sigma(\theta)) = 0$. Therefore, $f = c_\theta g$ with $g \in \mathbb{Q}[X]$. Let d and e be the least common multiples of the denominators of the coefficients of c_θ and g, respectively. Then $def = (dc_\theta)(eg)$. Lemma A.3.2 implies $d = e = 1$. So norm and trace of θ, which are the coefficients of c_θ, are integers.

Conversely, if norm and trace of θ are integers, then c_θ is a monic polynomial with integer coefficients and θ is a zero of that polynomial. This implies that θ is an algebraic integer.

The equivalence of the second and third assertion follows from the definition of the standard representation. □

Example 8.2.4. Consider the quadratic irrationality $\theta = (1+\sqrt{5})/2$. Its standard representation is $\theta = \theta(1, 1, -1)$. Hence, θ is an algebraic integer. Consider the quadratic irrationality $\theta = (1+\sqrt{3})/2$. Its standard representation is $\theta = \theta(2, 2, -1)$. Hence θ is not a quadratic integer.

Here is another characterization of the maximal order that is contained in a quadratic number field.

Theorem 8.2.5. *Let F be a quadratic number field. Then the maximal order of F is the set of all algebraic integers in F.*

Proof. Let Δ be the discriminant of F, let $\mathcal{O}$ be the maximal order of F, and let $\alpha \in F$. We show that α is an algebraic integer if and only if $\alpha \in \mathcal{O}$.

By Proposition 8.1.6 the number α is either a rational number or a quadratic irrationality.

If α is a rational number then by Example 8.2.2 it is an algebraic integer if and only if it is an integer. Since $F \cap \mathbb{Z} = \mathcal{O} \cap \mathbb{Z} = \mathbb{Z}$, it follows that rational numbers in F are algebraic integers if and only if they belong to $\mathcal{O}$.

Now assume that α is a quadratic irrationality. If α is an algebraic integer then, by Proposition 8.2.3 we can write $\alpha = (b + \sqrt{\Delta'})/2$ with an integer b and a discriminant Δ'. So α is contained in an order $\mathcal{O}_{\Delta'}$ that is contained in F. By Theorem 8.1.4 the order $\mathcal{O}_{\Delta'}$ is contained in the maximal order $\mathcal{O}$ of F. So $\alpha \in \mathcal{O}$.

Conversely, assume that $\alpha \in \mathcal{O}$. Then $\alpha = x + y\theta_\Delta$ with $x, y \in \mathbb{Z}$. It follows that norm and trace of α are integers. Hence, Proposition 8.2.3 implies that α is an algebraic integer. □

Theorem 8.2.5 shows that the set of all algebraic integers in a quadratic number field is an integral domain. It is called the *ring of integers* of F. It is is an analogue of the ring $\mathbb{Z}$ of integers in the field $\mathbb{Q}$ of rational numbers.

8.3 Units of orders

Let $\mathcal{O}$ be a quadratic order, let Δ be the discriminant of $\mathcal{O}$, and let F be the field of fractions of $\mathcal{O}$, that is, the quadratic number field in which $\mathcal{O}$ is contained. Then $\mathcal{O}$ is an integral domain. In this section we determine the group $\mathcal{O}^*$ of units of $\mathcal{O}$.

8.3.1 Correspondence to the Pell equation

We show that there is a bijection between the solutions of the Pell equation

$$x^2 - y^2\Delta = \pm 4 \tag{8.4}$$

and the units in $\mathcal{O}$.

Lemma 8.3.1. *Let $\varepsilon \in \mathcal{O}$. Then ε is a unit in $\mathcal{O}$ if and only if* $|\mathrm{N}(\varepsilon)| = 1$.

Proof. If $\varepsilon \in \mathcal{O}^*$ then ε and $1/\varepsilon$ belong to $\mathcal{O}$. Hence, $\mathrm{N}(\varepsilon)$ and $\mathrm{N}(1/\varepsilon) = 1/\mathrm{N}(\varepsilon)$ are integers. This implies $|\mathrm{N}(\varepsilon)| = 1$.

Conversely, assume that $|\mathrm{N}(\varepsilon)| = 1$. Let $\varepsilon = x + y\theta_\Delta$ with $x, y \in \mathbb{Z}$. Then $\sigma(\varepsilon) = x + y\sigma(\theta_\Delta)$. Since $\sigma(\theta_\Delta) \in \mathcal{O}$ we also have $\sigma(\varepsilon) \in \mathcal{O}$. Also $|\mathrm{N}(\varepsilon)| = 1$ implies $1/\varepsilon = \sigma(\varepsilon)/\mathrm{N}(\varepsilon) \in \mathcal{O}$. Hence, $\varepsilon \in \mathcal{O}^*$. □

Proposition 8.3.2. *The units of $\mathcal{O}$ are exactly the numbers $(x + y\sqrt{\Delta})/2$ where x and y are integers which satisfy the Pell equation* (8.3.8).

Proof. Let x, y be integers and assume that they satisfy the Pell equation. We can write $(x+y\sqrt{\Delta})/2 = (x-y\Delta)/2+y\theta_\Delta$. Since x, y satisfy the Pell equation we have $x^2 \equiv y^2\Delta \bmod 4$. So $x \equiv y\Delta \bmod 2$. It follows that $(x - y\Delta)/2$ is an integer and, therefore, $(x + y\sqrt{\Delta})/2 \in \mathcal{O}$. Also, since x, y satisfy the Pell equation we have $|\mathrm{N}(x + y\sqrt{\Delta})| = 1$. Hence, $(x + y\sqrt{\Delta})/2$ is a unit by Lemma 8.3.1.

Conversely, let $\varepsilon \in \mathcal{O}^*$. Then we can write $\varepsilon = (x + y\sqrt{\Delta})/2$ with integers x, y. By Lemma 8.3.1 we have $|\mathrm{N}(\varepsilon)| = 1$. Hence, x, y satisfy the Pell equation. □

From Proposition 8.3.2 we can deduce a correspondence between the automorphism group of a primitive integral form f of discriminant Δ and the unit group $\mathcal{O}^*$. For $\alpha = (x + y\sqrt{\Delta})/2$ we set

$$U(f, \alpha) = \begin{pmatrix} \frac{x-yb}{2} & -cy \\ ay & \frac{x+yb}{2} \end{pmatrix}. \tag{8.5}$$

Then

$$\mathrm{N}(\alpha) = \det(U(f, \alpha)). \tag{8.6}$$

Proposition 8.3.3. *Let f be an integral primitive form of discriminant Δ. The map that sends a unit ε of $\mathcal{O}$ to $U(f, \varepsilon)$ is a group isomorphism between $\mathcal{O}^*$ and* $\mathrm{Aut}(f)$.

Proof. By Theorem 2.5.5 and Proposition 8.3.2 this map is a bijection. We show that the map is a homomorphism. First, the unit 1 of $\mathcal{O}$ is mapped to $U(f, 1) = I_2$. Hence, the identity of $\mathcal{O}^*$ is mapped to the identity of $\mathrm{Aut}(f)$. Let $\alpha = (x + y\sqrt{\Delta})/2$ and $\alpha' = (x' + y'\sqrt{\Delta})/2$, $x, y, x', y' \in \mathbb{Z}$, be two units of $\mathcal{O}$. Their product is $(xx' + yy'\Delta)/2 + (xy' + x'y)\sqrt{\Delta}/2$. By Exercise 2.9.7 it is sent to the automorphism $U(f, \alpha)U(f, \alpha')$. □

Proposition 8.3.3 enables us to determine the unit group $\mathcal{O}^*$ explicitly. This will be done in the following two sections.

8.3.2 Units of imaginary quadratic orders

The next theorem describes the unit group of imaginary quadratic orders. That unit group is almost always *trivial*, that is, it only contains the units ± 1.

Theorem 8.3.4.

1. *The group $\mathcal{O}^*_{-3}$ is the group of sixth roots of unity. It is generated by $(1+\sqrt{-3})/2$.*
2. *The group $\mathcal{O}^*_{-4}$ is the group of fourth roots of unity. It is generated by $\sqrt{-1}$.*
3. *The group $\mathcal{O}^*_{\Delta}$ with $\Delta < -4$ is $\{\pm 1\}$.*

Proof. Let $\Delta = -3$. Then $\mathcal{O} = L(1,1,1)$. By Example 2.5.9 the automorphism group of the form $(1,1,1)$ is cyclic of order 6 generated by $\begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$. By Example 2.5.6, the corresponding solution of the Pell equation is $(1,1)$ which corresponds to the unit $(1+\sqrt{-3})/2$.

Let $\Delta = -4$. Then $\mathcal{O} = L(1,0,1)$. By Example 2.5.8 the automorphism group of the form $(1,1,1)$ is cyclic of order 4 generated by $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. By Example 2.5.6, the corresponding solution of the Pell equation is $(0,1)$ which corresponds to the unit $(\sqrt{-4})/2 = \sqrt{-1}$.

If $\Delta < -4$ then by Theorem 2.5.10 the automorphism group of any primitive integral form of discriminant Δ is trivial. By Proposition 8.3.3 the unit group $\mathcal{O}^*$ is $\{\pm 1\}$. □

8.3.3 Units of real quadratic orders

Let Δ be positive. We recall that the *fundamental solution* of the Pell equation (6.36) is the solution $(x,y) \in \mathbb{Z}^2$ with $x,y > 0$ and minimal y (see Definition 6.12.6). The next theorem describes the unit group of real quadratic orders in terms of the fundamental solution of the Pell equation.

Theorem 8.3.5. *If $\Delta > 0$ then $\mathcal{O}^* = \langle -1 \rangle \times \langle (x+y\sqrt{\Delta})/2 \rangle$ where (x,y) is the fundamental solution of the Pell equation $x^2 - \Delta y^2 = \pm 4$.*

Proof. This theorem follows from Proposition 6.12.7 and Proposition 8.3.3. □

Definition 8.3.6. *Let $\Delta > 0$. If (x,y) is the fundamental solution of the Pell equation $x^2 - \Delta y^2 = \pm 4$, then $(x+y\sqrt{\Delta})/2$ is called the* fundamental unit *of $\mathcal{O}$. It is denoted by ε_Δ. Also, the* regulator *of $\mathcal{O}$ is $R_\Delta = \log \varepsilon_\Delta$.*

By Proposition 6.13.2 the fundamental unit $\varepsilon = (x+y\sqrt{\Delta})/2$ of $\mathcal{O}$ can be computed in polynomial time in the size of the output (x,y). However, the size of the output (x,y) may be exponential in the size of the input Δ.

Proposition 8.3.7. *Let $\Delta > 0$. Let f be a reduced form with discriminant Δ, and $f = f_0, \dots, f_{l-1}$ its cycle. Then*

$$\varepsilon_\Delta = \prod_{i=0}^{l-1} \theta_1(f_i) \ .$$

Proof. Define T_i as in (6.32) and (6.33). Then T_l is the fundamental automorphism of f. Let $T_{l,1} = (p_l, q_l)^{\mathrm{T}}$ be the left column of T_l. Let further (x, y) be the fundamental solution of the Pell equation. Then by Proposition 6.12.7 we have $p_l = (x - yb)/2$ and $q_l = ay$. It follows that

$$(1, \theta_1(f))T_{l,1} = \varepsilon_\Delta. \tag{$*$}$$

Define

$$(\mu_i, \mu_{i+1}) = (1, \theta(f))T_i \ .$$

Then by Proposition 4.3.18

$$\theta(f_i) = \frac{\mu_{i+1}}{\mu_i} \ ,$$

and

$$\prod_{i=0}^{l-1} \theta(f_i) = \mu_l \ .$$

If we take first components on both sides of this equation, and use $\mu_{l,1} = \varepsilon_\Delta$ which is implied by $(*)$ we obtain the asserted equality. □

Table 8.1 on page 164 contains the fundamental units for the first few real quadratic discriminants.

The sign of the norm of ε_Δ is related to the parity of the length of the cycles of forms of discriminant Δ.

Proposition 8.3.8.

1. *If the norm of the fundamental unit of $\mathcal{O}$ is positive, then the cycles of reduced forms of discriminant Δ have even length and $h(\Delta) = 2h^+(\Delta)$.*
2. *If the norm of the fundamental unit of $\mathcal{O}$ is negative, then the cycles of reduced forms of discriminant Δ have odd length and $h(\Delta) = h^+(\Delta)$.*

Proof. By (8.6) the norm of the fundamental unit is the determinant of the fundamental automorphism of any primitive integral form f of discriminant Δ. So the statement follows from Theorem 6.12.4. □

8.4 Ideals of orders

Let F be a quadratic number field and let $\mathcal{O} \subset F$ be a quadratic order of discriminant Δ. In this section we will see that the lattices in F, whose ring of multipliers is $\mathcal{O}$, form a multiplicative group and we will study the algebraic properties of this group.

Δ	fundamental unit ε_Δ
5	$(1+\sqrt{\Delta})/2$
8	$(2+\sqrt{\Delta})/2$
12	$(4+\sqrt{\Delta})/2$
13	$(3+\sqrt{\Delta})/2$
17	$4+\sqrt{\Delta}$
20	$(4+\sqrt{\Delta})/2$
21	$(5+\sqrt{\Delta})/2$
24	$5+\sqrt{\Delta}$
28	$(16+3\sqrt{\Delta})/2$
29	$(5+\sqrt{\Delta})/2$
33	$23+4\sqrt{\Delta}$

Table 8.1. Fundamental units

8.4.1 Fractional $\mathcal{O}$-ideals

We will show that the lattices in F whose ring of multipliers contains $\mathcal{O}$, are exactly the fractional $\mathcal{O}$-ideals that are defined now.

Definition 8.4.1.

1. *An $\mathcal{O}$-ideal is an additive subgroup of $\mathcal{O}$ which is an $\mathcal{O}$-module with respect to multiplication.*
2. *A* primitive *$\mathcal{O}$-ideal is an $\mathcal{O}$-ideal $\mathfrak{a}$ that cannot be written as $\mathfrak{a} = m\mathfrak{b}$ with another $\mathcal{O}$-ideal $\mathfrak{b}$ and $m \in \mathbb{N}$.*
3. *A* fractional $\mathcal{O}$-ideal *is a subset $\mathfrak{b}$ of F such that $d\mathfrak{b}$ is an $\mathcal{O}$-ideal for some positive integer d.*

Any $\mathcal{O}$-ideal is a fractional $\mathcal{O}$-ideal. An $\mathcal{O}$-ideal is also called an *integral $\mathcal{O}$-ideal.*

Example 8.4.2. Let $\alpha \in F$. Then $\alpha\mathcal{O} = \{\alpha\beta : \beta \in \mathcal{O}\}$ is an $\mathcal{O}$-ideal. Any $\mathcal{O}$-ideal of that form is called *principal*. The number α is called a *generator* of $\alpha\mathcal{O}$.

We characterize the two-dimensional lattices in F.

Proposition 8.4.3. *The two-dimensional lattices in F are exactly the fractional $\mathcal{O}'$-ideals of the orders $\mathcal{O}'$ that are contained in F.*

Proof. Let $\mathcal{O}'$ be an order in F and let $\mathfrak{a}$ be a fractional $\mathcal{O}'$-ideal. Then $d\mathfrak{a} \subset \mathcal{O}'$ for some positive integer d. By Corollary A.4.8, $\mathfrak{a}$ is a lattice. Also, since $\mathcal{O}\mathfrak{a} \subset \mathfrak{a}$, the lattice $\mathfrak{a}$ is two-dimensional.

Conversely, assume that L is a two-dimensional lattice in F. Let (α, γ) be a $\mathbb{Z}$-basis of L. Then (α, γ) is also a $\mathbb{Q}$-basis of F. Another $\mathbb{Q}$-basis of F is

$\sqrt{\Delta}(\alpha, \gamma)$. So we can write $\sqrt{\Delta}(\alpha, \gamma) = (\alpha, \gamma)T$ with $T \in \mathrm{GL}(2, \mathbb{Q})$. If d is the denominator of $\det T$, then $d\sqrt{\Delta}L \subset L$. Hence, $d\sqrt{\Delta}$ is a multiplier of L. By Theorem 7.3.4, the ring of multipliers of L is an order $\mathcal{O}'$. Since $LL^{-1} = \mathcal{O}'$, Lemma 7.3.6 implies that $\mathcal{O}' \subset (1/e)L$ for some positive integer e. Hence, the order $\mathcal{O}'$ is contained in F. □

We also characterize the fractional $\mathcal{O}$-ideals.

Proposition 8.4.4. *Let $\mathfrak{b} \subset F$. Then the following statements are equivalent.*

1. *$\mathfrak{b}$ is a fractional $\mathcal{O}$-ideal.*
2. *$\mathfrak{b}$ is a lattice whose ring of multipliers contains $\mathcal{O}$.*
3. *$\mathfrak{b} = qL(f)$ with a positive rational number q and an integral form f of discriminant Δ.*

Proof. If $\mathfrak{b}$ is a fractional $\mathcal{O}$-ideal then $d\mathfrak{b}$ is an additive subgroup of the lattice $\mathcal{O}$ for some positive integer d. By Corollary A.4.8, $\mathfrak{b}$ is a lattice. Also, $\mathfrak{b}$ is an $\mathcal{O}$-module, that is, $\mathcal{O}\mathfrak{b} \subset \mathfrak{b}$. This implies that $\mathcal{O} \subset \mathcal{O}(\mathfrak{b})$.

Assume that $\mathfrak{b}$ is a lattice with $\mathcal{O} \subset \mathcal{O}(\mathfrak{b})$. It follows from Theorem 7.3.4 that $\mathcal{O}(\mathfrak{b})$ is a quadratic order. Let Δ_0 be the discriminant of $\mathcal{O}(\mathfrak{b})$. By Proposition 7.2.7 we have $\Delta = d^2\Delta_0$ with a positive integer d. Also, Exercise 7.5.6 implies that $L = rL(a_0, b_0, c_0)$ with a primitive integral form f_0 of discriminant Δ_0 and a positive real number r. In fact, r is a rational number by Exercise 8.7.2. Set $q = r/d$ and $f = df_0$. Then q is a rational number, the form f is integral, we have $\Delta(f) = \Delta$, and $\mathfrak{b} = qL(f)$.

Finally, let $\mathfrak{b} = qL(f)$ with a positive rational number q and a form f of discriminant Δ. Let $f = (a, b, c)$ and let $d = \gcd(a, b, c)$. Then $(1/d)f$ is a primitive form of discriminant $\Delta_0 = \Delta/d^2$. Theorem 7.3.4 implies that $\mathfrak{b}$ is a lattice and $\mathcal{O}(\mathfrak{b}) = \mathcal{O}_{\Delta_0}$. Also, by Proposition 7.2.7, the order $\mathcal{O}$ is contained in $\mathcal{O}_{\Delta_0}$. Hence, $\mathfrak{b}$ is an $\mathcal{O}$-module. Also, if $q = n/d$ with positive integers n, d, then $d\mathfrak{b}$ is an additive subgroup of $\mathcal{O}$. Hence, $\mathfrak{b}$ is a fractional $\mathcal{O}$-ideal. □

It follows from Proposition 8.4.4 that any fractional $\mathcal{O}$-ideal $\mathfrak{b}$ can be written as

$$\mathfrak{b} = qL(f) = q\left(\mathbb{Z}a + \mathbb{Z}\frac{b + \sqrt{\Delta}}{2}\right) \tag{8.8}$$

with a positive rational number q and an integral form $f = (a, b, c)$ of discriminant Δ with $a > 0$. The next proposition shows in which sense the representation (8.8) of the fractional ideal $\mathfrak{b}$ is unique.

Proposition 8.4.5. *Let $\mathfrak{b}$ be a fractional $\mathcal{O}$-ideal and let $\mathfrak{b} = qL(f)$ be the representation of $\mathfrak{b}$ from (8.8). Then q is uniquely determined and f is unique modulo Γ. Also, $\mathfrak{b}$ is an integral $\mathcal{O}$-ideal if and only if $q \in \mathbb{N}$, and $\mathfrak{b}$ is an integral primitive $\mathcal{O}$-ideal if and only if $q = 1$.*

Proof. Exercise 8.4.5. □

Definition 8.4.6. *If we choose in* (8.8) *the form* (a, b, c) *to be normal, then that representation is called the* standard representation *of* $\mathfrak{b}$. *We write*

$$f_{\mathfrak{b}} = (a, b, c). \tag{8.9}$$

Proposition 8.4.5 implies the following result.

Corollary 8.4.7. *The map that sends the* Γ*-orbit of an integral form* (a, b, c) *of discriminant* Δ *to the lattice* $L(f)$ *is a bijection between all those* Γ*-orbits and the integral primitive* $\mathcal{O}$*-ideals.*

By Corollary 8.4.7 we can identify the integral primitive $\mathcal{O}$-ideals with the Γ-orbits of integral forms of discriminant Δ.

In the next proposition we explain how to compute the inverse of the map in Corollary 8.4.7, that is, how to find a representation (8.8) of a fractional $\mathcal{O}$-ideal that is given in terms of a $\mathbb{Z}$-basis.

Proposition 8.4.8. *Let* $\mathfrak{b}$ *be a fractional* $\mathcal{O}$*-ideal and let* (α_1, α_2) *be a* $\mathbb{Z}$*-basis of* $\mathfrak{b}$. *Write*

$$\alpha_i = \frac{x_i + y_i\sqrt{\Delta}}{2d} \ , \ \ x_i, y_i \in \mathbb{Z} \ , i = 1, 2, \ \ d \in \mathbb{Z}_{>0}. \tag{8.10}$$

Let

$$m = \gcd(y_1, y_2) = u_1 y_1 + u_2 y_2 \ , \ \ u_1, u_2 \in \mathbb{Z}, \tag{8.11}$$

and

$$a = \left| \frac{y_2 x_1 - y_1 x_2}{2m^2} \right| \ , \ \ b = \frac{u_1 x_1 + u_2 x_2}{m} \ , \ \ c = \frac{b^2 - \Delta}{4a} \ , \ \ q = \frac{m}{d} \tag{8.12}$$

Then a*,* b*, and* c *are integers, the form* (a, b, c) *has discriminant* Δ*, and we have*

$$\mathfrak{b} = qL(a, b, c).$$

Proof. Exercise 8.7.5. □

Proposition 8.4.8 yields an algorithm that is used in the next example.

Example 8.4.9. We determine the representation (8.8) of the principal ideal $\mathfrak{b} = \sqrt{5}\mathcal{O}_5$. Since $\big(1, (1 + \sqrt{5})/2\big)$ is a $\mathbb{Z}$-basis of $\mathcal{O}_5$, it follows that $(\alpha_1, \alpha_2) = \big(\sqrt{5}, (5 + \sqrt{5})/2\big)$ is a $\mathbb{Z}$-basis of $\mathfrak{b}$. Using the notation from Proposition 8.4.8 we have $x_1 = 0$, $y_1 = 2$, $x_2 = 5$, $y_2 = 1$. Hence $m = 1$ and we can use $u_1 = 1$ and $u_2 = -1$. So $a = b = 5$, $c = 1$ and $\mathfrak{b} = L(5, 5, 1) = 5\mathbb{Z} + \mathbb{Z}(5 + \sqrt{5})/2$.

8.4.2 Invertible $\mathcal{O}$-ideals

It follows from Proposition 8.4.4 that fractional $\mathcal{O}$-ideals are lattices in F. As lattices, fractional $\mathcal{O}$-ideals can be multiplied (see Section 7.3.4). We now show that the set of fractional $\mathcal{O}$-ideals is closed under this multiplication.

Proposition 8.4.10.

1. *With respect to multiplication, the set of integral $\mathcal{O}$-ideals is a commutative semigroup with neutral element $\mathcal{O}$.*
2. *With respect to multiplication, the set of fractional $\mathcal{O}$-ideals is a commutative semigroup with neutral element $\mathcal{O}$.*

Proof. Let $\mathfrak{a}$ and $\mathfrak{b}$ be integral $\mathcal{O}$-ideals. Their product $\mathfrak{a}\mathfrak{b}$ is an additive subgroup of $\mathcal{O}$ and an $\mathcal{O}$-module. Therefore, $\mathfrak{a}\mathfrak{b}$ is an $\mathcal{O}$-ideal. This proves the first assertion. The second assertion is a consequence of the first one. □

In Section 7.3.3 we have introduced the inverse $\mathfrak{a}^{-1}$ of a fractional $\mathcal{O}$-ideal $\mathfrak{a}$ and we have shown that $\mathfrak{a}\mathfrak{a}^{-1} = \mathcal{O}(\mathfrak{a})$. We will now determine the fractional $\mathcal{O}$-ideals that have a multiplicative inverse in the semigroup of fractional $\mathcal{O}$-ideals.

Definition 8.4.11. *An* invertible $\mathcal{O}$-ideal *is a fractional $\mathcal{O}$-ideal $\mathfrak{a}$ that is invertible in the semigroup of fractional $\mathcal{O}$-ideals, that is, there is a fractional $\mathcal{O}$-ideal $\mathfrak{b}$ with $\mathfrak{a}\mathfrak{b} = \mathcal{O}$.*

Example 8.4.12. All principal $\mathcal{O}$-ideals are invertible $\mathcal{O}$-ideals. In particular, $\mathcal{O}$ is an invertible $\mathcal{O}$-ideal.

The following result is obviously correct.

Theorem 8.4.13. *The set of invertible $\mathcal{O}$-ideals is an Abelian group with respect to ideal multiplication. The neutral element is $\mathcal{O}$.*

We denote the group of invertible $\mathcal{O}$-ideals by $\mathcal{I}(\mathcal{O})$. In the next proposition we characterize the invertible $\mathcal{O}$-ideals.

Proposition 8.4.14. *Let $\mathfrak{a}$ be a fractional $\mathcal{O}$-ideal. Then the following statements are equivalent.*

1. *$\mathfrak{a}$ is an invertible $\mathcal{O}$-ideal.*
2. *The ring of multipliers of $\mathfrak{a}$ is $\mathcal{O}$.*
3. *We have $\mathfrak{a} = qL(f)$ with $q \in \mathbb{Q}_{>0}$ and an integral primitive form f of discriminant Δ.*
4. *We have $\mathfrak{a}\mathfrak{a}^{-1} = \mathcal{O}$.*

Proof. Assume that $\mathfrak{a}\mathfrak{b} = \mathcal{O}$ for some fractional $\mathcal{O}$-ideal $\mathfrak{b}$. Then $\mathfrak{a}^{-1}\mathfrak{b}^{-1} = \mathcal{O}$. Hence, $\mathcal{O} = \mathcal{O}\mathcal{O} = \mathfrak{a}\mathfrak{a}^{-1}\mathfrak{b}\mathfrak{b}^{-1} = \mathcal{O}(\mathfrak{a})\mathcal{O}(\mathfrak{b}) \supset \mathcal{O}(\mathfrak{a}) \supset \mathcal{O}$. Hence, $\mathcal{O}(\mathfrak{a}) = \mathcal{O}$.

Assume that $\mathcal{O}(\mathfrak{a}) = \mathcal{O}$. Write $\mathfrak{a} = qL(f)$ as in (8.8). Then f is primitive by Theorem 7.3.4 and Proposition 8.4.5.

Assume that $\mathfrak{a} = qL(f)$ with $q \in \mathbb{Q}_{>0}$ and an integral primitive form f of discriminant Δ. Then $\mathfrak{a}\mathfrak{a}^{-1} = \mathcal{O}$ by (7.9). Hence, $\mathfrak{a}$ has a multiplicative inverse in the semigroup of fractional $\mathcal{O}$-ideals. □

Proposition 8.4.14 implies the following result.

Corollary 8.4.15. *If $\mathcal{O}$ is a maximal order, then all fractional $\mathcal{O}$-ideals are invertible.*

Proof. Let $\mathcal{O}$ be maximal and let $\mathfrak{a}$ be a maximal $\mathcal{O}$-ideal. Then $\mathcal{O}$ is contained in $\mathcal{O}(\mathfrak{a})$. Since $\mathcal{O}$ is maximal, we have $\mathcal{O}(\mathfrak{a}) = \mathcal{O}$. Hence, $\mathfrak{a}$ is invertible by Proposition 8.4.14. □

8.5 Factorization of ideals

Let $\mathcal{O}$ be a quadratic order. In this section we show that in the semigroup of integral $\mathcal{O}$-ideals, whose index in $\mathcal{O}$ is coprime to the conductor of $\mathcal{O}$, there is unique factorization into so-called prime ideals. We also explain how to compute this factorization.

Let Δ be the discriminant and let F be the field of fractions of $\mathcal{O}$.

8.5.1 Norm

We define the norm of $\mathcal{O}$-ideals.

Definition 8.5.1. *Let $\mathfrak{a}$ be a fractional $\mathcal{O}$-ideal. The* norm *of $\mathfrak{a}$ (as a fractional $\mathcal{O}$-ideal) is defined as*

$$\mathrm{N}_\Delta(\mathfrak{a}) = \mathrm{N}_\mathcal{O}(\mathfrak{a}) = [\mathcal{O} : \mathfrak{a}] \tag{8.13}$$

where $[\mathcal{O} : \mathfrak{a}]$ is defined as in (7.8).

If $\mathfrak{a}$ is a fractional $\mathcal{O}$-ideal written as in (8.8) as

$$\mathfrak{a} = qL(a, b, c)$$

with a form (a, b, c) of discriminant Δ, then

$$\mathrm{N}_\mathcal{O}(\mathfrak{a}) = q^2 a. \tag{8.14}$$

Note that a lattice in F can be a fractional ideal of many orders and that the norm of $\mathfrak{a}$ as a fractional ideal depends on that order. If it is clear, which order we mean, then we write $\mathrm{N}(\mathfrak{a})$ for $\mathrm{N}_\mathcal{O}(\mathfrak{a})$.

Example 8.5.2. Consider the lattice $L = L(2, 2, -2) = 2\mathcal{O}_5$. It is an $\mathcal{O}_{20}$-ideal and an $\mathcal{O}_5$-ideal and we have $\mathrm{N}_5(L) = 4$ and $\mathrm{N}_{20}(L) = 2$.

Lemma 8.5.3. *Let $\mathfrak{a}$ and $\mathfrak{b}$ be fractional $\mathcal{O}$-ideals. Then the following are true.*

1. If $\mathfrak{a}$ and $\mathfrak{b}$ are invertible $\mathcal{O}$-ideals, then $\mathrm{N}_\mathcal{O}(\mathfrak{a}\mathfrak{b}) = \mathrm{N}_\mathcal{O}(\mathfrak{a})\mathrm{N}_\mathcal{O}(\mathfrak{b})$.

2. *If* $\mathfrak{a}$ *and* $\mathfrak{b}$ *are integral* $\mathcal{O}$*-ideals and if* $\mathfrak{a}$ *is a subset of* $\mathfrak{b}$*, then* $\mathrm{N}_{\mathcal{O}}(\mathfrak{b})$ *divides* $\mathrm{N}_{\mathcal{O}}(\mathfrak{a})$*.*

Proof. Let $\mathfrak{a} = qL(a,b,c)$ and $\mathfrak{b} = q'L(a',b',c')$ with rational numbers q, q' and forms (a,b,c), (a',b',c') of discriminant Δ.

1. By Corollary 7.3.17 we have $\mathrm{N}(\mathfrak{ab}) = q^2(q')^2a_1a_2 = \mathrm{N}(\mathfrak{a})\mathrm{N}(\mathfrak{b})$.

2. This assertion follows from the fact that the norm of an integral $\mathcal{O}$-ideal is the index of that ideal in $\mathcal{O}$. □

8.5.2 Divisibility of $\mathcal{O}$-ideals

We discuss divisibility in the semigroup of integral $\mathcal{O}$-ideals. By $\mathfrak{a}$, $\mathfrak{b}$, and $\mathfrak{c}$ we denote integral $\mathcal{O}$-ideals.

Definition 8.5.4. *We say that* $\mathfrak{b}$ divides $\mathfrak{a}$ *(in the semigroup of integral* $\mathcal{O}$*-ideals) if* $\mathfrak{a} = \mathfrak{bc}$ *for some integral* $\mathcal{O}$*-ideal* $\mathfrak{c}$*.*

We prove necessary and sufficient conditions for the divisibility of $\mathcal{O}$-ideals.

Lemma 8.5.5.

1. *If* $\mathfrak{b}$ *divides* $\mathfrak{a}$*, then* $\mathfrak{a}$ *is a subset of* $\mathfrak{b}$ *and the norm of* $\mathfrak{b}$ *divides the norm of* $\mathfrak{a}$*.*
2. *If* $\mathfrak{b}$ *is an invertible* $\mathcal{O}$*-ideal and if* $\mathfrak{a}$ *is a subset of* $\mathfrak{b}$*, then* $\mathfrak{b}$ *divides* $\mathfrak{a}$*.*

Proof. Suppose that $\mathfrak{b}$ divides $\mathfrak{a}$. Then there is an $\mathcal{O}$-ideal $\mathfrak{c}$ with $\mathfrak{a} = \mathfrak{bc}$. Since $\mathcal{O}\mathfrak{b} \subset \mathfrak{b}$ and since $\mathfrak{c} \subset \mathcal{O}$ it follows that $\mathfrak{a} = \mathfrak{bc} \subset \mathfrak{b}\mathcal{O} \subset \mathfrak{b}$. By Lemma 8.5.3, the norm of $\mathfrak{b}$ divides the norm of $\mathfrak{a}$.

Conversely, assume that $\mathfrak{b}$ is invertible and that $\mathfrak{a}$ is contained in $\mathfrak{b}$. Set $\mathfrak{c} = \mathfrak{ab}^{-1}$. Then $\mathfrak{a} = \mathfrak{bc}$. Also, we have $\mathfrak{c} = \mathfrak{ab}^{-1} \subset \mathfrak{bb}^{-1} = \mathcal{O}$. Hence $\mathfrak{c}$ is an integral $\mathcal{O}$-ideal. □

We give a few examples. In the first example we show how to use the sufficient condition in Lemma 8.5.5 to find out that an ideal $\mathfrak{b}$ divides an ideal $\mathfrak{a}$ and how to calculate $\mathfrak{c}$ with $\mathfrak{a} = \mathfrak{bc}$.

Example 8.5.6. Let $\Delta = 5$, $\mathfrak{a} = 11\mathcal{O}_5$, $\mathfrak{b} = L(11,15,5)$. Then $\mathfrak{a} \subset \mathfrak{b}$ and $\mathfrak{b}$ is an invertible $\mathcal{O}_5$-ideal by Proposition 8.4.14. Hence, $\mathfrak{b}$ divides $\mathfrak{a}$ by Lemma 8.5.5. Also, for $\mathfrak{c}$ with $\mathfrak{a} = \mathfrak{bc}$ we must have $\mathfrak{c} = \mathfrak{ab}^{-1}$. So (7.10) shows that $\mathfrak{c} = L(11,-15,5)$.

The next example shows that the divisibility of $\mathrm{N}(\mathfrak{a})$ by $\mathrm{N}(\mathfrak{b})$ is not a sufficient condition for the divisibility of $\mathfrak{a}$ by $\mathfrak{b}$.

Example 8.5.7. Let $\Delta = 5$, $\mathfrak{a} = L(11,15,5)$, $\mathfrak{b} = L(11,-15,5)$. Then $\mathrm{N}(\mathfrak{a}) = \mathrm{N}(\mathfrak{b}) = 11$. We show that $\mathfrak{a}$ is not a subset of $\mathfrak{b}$. If $\mathfrak{a}$ were a subset of $\mathfrak{b}$ then $\mathfrak{b}$ would contain $(15+\sqrt{5})/2$ and $(15-\sqrt{5})/2$. Therefore, $\mathfrak{b}$ would contain 15. This is impossible since 15 is not a multiple of 11. It follows that $\mathfrak{b}$ does not divide $\mathfrak{a}$. It follows from a similar argument that $\mathfrak{a}$ does not divide $\mathfrak{b}$.

Finally we present an example which shows that it is possible that $\mathfrak{a} \subset \mathfrak{b}$ but $\mathfrak{b}$ does not divide $\mathfrak{a}$.

Example 8.5.8. Let $\mathcal{O} = \mathcal{O}_{20}$. Consider the $\mathcal{O}$-ideals $\mathfrak{b} = L(2, 2, -2) = 2\mathcal{O}_5$ and $\mathfrak{a} = 2\mathcal{O}$. Then $\mathfrak{a} \subset \mathfrak{b}$. We show that $\mathfrak{b}$ does not divide $\mathfrak{a}$ in the semigroup of integral $\mathcal{O}$-ideals. Assume that there is an integral $\mathcal{O}$-ideal $\mathfrak{c}$ with $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$. By Lemma 8.5.3 we know that $\mathrm{N}(\mathfrak{c})$ divides $\mathrm{N}(\mathfrak{a}) = 4$. If $\mathrm{N}(\mathfrak{c}) = 1$, then $\mathfrak{c} = \mathcal{O}$ which would imply the falsehood $\mathfrak{a} = \mathfrak{b}$. If $\mathrm{N}(\mathfrak{c}) = 4$, then $\mathfrak{c} = \mathfrak{a}$. This in turn would imply $\mathfrak{a} = \mathfrak{b}\mathfrak{a}$ and $\mathfrak{b} = \mathcal{O}$ since $\mathfrak{a}$ is principal. Hence, $\mathrm{N}(\mathfrak{c}) = 2$. So $\mathfrak{c} = \mathfrak{b}$ because $\mathfrak{b}$ is the only integral $\mathcal{O}$-ideal of norm 2. But $\mathfrak{b}^2 = L(4, 2, -1)$, a contradiction. Note that $L(2, 2, -2)$ is not an invertible $\mathcal{O}$-ideal.

Next, we define coprime ideals. This definition is based on the following proposition. In this proposition we use the notion of the sum of two $\mathcal{O}$-ideals which was introduced in Definition 7.1.1. In Exercise 8.7.9 we explain how to compute this sum. Also, by a *common divisor* of $\mathfrak{a}$ and $\mathfrak{b}$ we mean an integral $\mathcal{O}$-ideal $\mathfrak{c}$ that divides both $\mathfrak{a}$ and $\mathfrak{b}$.

Proposition 8.5.9. *If* $\mathfrak{a}+\mathfrak{b}$ *is an invertible* $\mathcal{O}$*-ideal, then* $\mathfrak{a}+\mathfrak{b}$ *is the uniquely determined common divisor of* $\mathfrak{a}$ *and* $\mathfrak{b}$ *that is divisible by any other common divisor of* $\mathfrak{a}$ *and* $\mathfrak{b}$*.*

Proof. Assume that $\mathfrak{a} + \mathfrak{b}$ is an invertible $\mathcal{O}$-ideal. Since $\mathfrak{a}$ and $\mathfrak{b}$ are both contained in $\mathfrak{a}+\mathfrak{b}$, it follows from Lemma 8.5.5 that $\mathfrak{a}+\mathfrak{b}$ is a common divisor if $\mathfrak{a}$ and $\mathfrak{b}$. Now suppose that there is any other common divisor $\mathfrak{c}$ of $\mathfrak{a}$ and $\mathfrak{b}$. Then $\mathfrak{a}$ and $\mathfrak{b}$ are subsets of $\mathfrak{c}$ by Lemma 8.5.5 and therefore $\mathfrak{a} + \mathfrak{b} \subset \mathfrak{c}$. Since $\mathfrak{a} + \mathfrak{b}$ is invertible, Lemma 8.5.5 implies that $\mathfrak{a} + \mathfrak{b}$ divides $\mathfrak{c}$.

To prove the uniqueness we assume that $\mathfrak{c}$ is a common divisor of $\mathfrak{a}$ and $\mathfrak{b}$ that is divisible by any other common divisor of $\mathfrak{a}$ and $\mathfrak{b}$. Then $\mathfrak{c}$ is in particular divisible by $\mathfrak{a}+\mathfrak{b}$. Also, $\mathfrak{a}+\mathfrak{b}$ is divisible by $\mathfrak{c}$. Lemma 8.5.5 implies that $\mathfrak{a} + \mathfrak{b} \subset \mathfrak{c} \subset \mathfrak{a} + \mathfrak{b}$, hence $\mathfrak{c} = \mathfrak{a} + \mathfrak{b}$. □

It follows from Proposition 8.5.9 and Exercise 8.7.10 that if $\mathfrak{a} + \mathfrak{b} = \mathcal{O}$, then $\mathcal{O}$ is the only common divisor of $\mathfrak{a}$ and $\mathfrak{b}$. This motivates the following definition.

Definition 8.5.10. *The ideals* $\mathfrak{a}$ *and* $\mathfrak{b}$ *are called* coprime *if* $\mathfrak{a} + \mathfrak{b} = \mathcal{O}$*.*

Example 8.5.11. Let $\Delta = 5$, $\mathfrak{a} = L(11, 15, 5)$, $\mathfrak{b} = L(11, -15, 5)$. Then $\mathfrak{a}+\mathfrak{b} = \mathcal{O}$ since $\mathfrak{a}+\mathfrak{b}$ contains 11 and 15 and therefore $1 = 3 * 15 - 4 * 11$. The ideals $\mathfrak{a}$ and $\mathfrak{b}$ are coprime.

8.5.3 Unique factorization into coprime ideals

We show that in a semigroup of $\mathcal{O}$-ideals that is generated by finitely many pairwise coprime invertible $\mathcal{O}$-ideals there is unique factorization. We let $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$ be integral invertible $\mathcal{O}$-ideals.

Lemma 8.5.12. *If* $\mathfrak{a}$ *and* $\mathfrak{b}$ *are coprime and if* $\mathfrak{a}$ *and* $\mathfrak{c}$ *are coprime, then* $\mathfrak{a}$ *and* $\mathfrak{b}\mathfrak{c}$ *are coprime.*

Proof. We have $\mathcal{O} = (\mathfrak{a}+\mathfrak{b})(\mathfrak{a}+\mathfrak{c}) = \mathfrak{a}^2 + \mathfrak{a}\mathfrak{b} + \mathfrak{a}\mathfrak{c} + \mathfrak{c}\mathfrak{b} \subset \mathfrak{a} + \mathfrak{c}\mathfrak{b} \subset \mathcal{O}$. Hence, $\mathfrak{a}$ and $\mathfrak{b}\mathfrak{c}$ are coprime. □

Corollary 8.5.13. *Let* $\mathcal{I}$ *and* $\mathcal{J}$ *be finite sets of integral* $\mathcal{O}$*-ideals. Assume that* $\mathfrak{a}+\mathfrak{b} = \mathcal{O}$ *for all* $\mathfrak{a} \in \mathcal{I}$ *and* $\mathfrak{b} \in \mathcal{J}$. *Then* $\prod_{\mathfrak{a}\in\mathcal{I}} \mathfrak{a} + \prod_{\mathfrak{b}\in\mathcal{J}} \mathfrak{b} = \mathcal{O}$.

Proof. We use induction on $|\mathcal{I}| + |\mathcal{J}|$. If $|\mathcal{I}| + |\mathcal{J}| = 2$, then the assertion is trivial. Now assume that $|\mathcal{I}| + |\mathcal{J}| = n > 2$ and that the assertion is true for $|\mathcal{I}| + |\mathcal{J}| \leq n-1$. Without loss of generality, assume that $|\mathcal{I}| > 1$. Fix $\mathfrak{a}_0 \in \mathcal{I}$ and set $\mathfrak{a}' = \prod_{\mathfrak{a}\in\mathcal{I},\mathfrak{a}\neq\mathfrak{a}_0} \mathfrak{a}$. Also set $\mathfrak{b}' = \prod_{\mathfrak{b}\in\mathcal{J}} \mathfrak{b}$. Then $\mathfrak{a}' + \mathfrak{b}' = \mathcal{O}$ by the induction hypothesis. Also, the induction hypothesis implies $\mathfrak{a}_0 + \mathfrak{b}' = \mathcal{O}$. So Lemma 8.5.12 implies the assertion. □

Theorem 8.5.14. *Let* $\mathcal{I}$ *be a finite set of pairwise coprime integral invertible* $\mathcal{O}$*-ideals different from* $\mathcal{O}$. *Then any integral* $\mathcal{O}$*-ideal in the multiplicative semigroup generated by* $\mathcal{I}$ *can be written as* $\mathfrak{a} = \prod_{\mathfrak{a}\in\mathcal{I}} \mathfrak{a}^{e(\mathfrak{a})}$ *with uniquely determined non-negative integers* $e(\mathfrak{a})$, $\mathfrak{a} \in \mathcal{I}$.

Proof. Let $\mathfrak{a}$ be an ideal in the semigroup generated by $\mathcal{I}$. Then $\mathfrak{a}$ can be written as $\mathfrak{a} = \prod_{\mathfrak{b}\in\mathcal{I}} \mathfrak{b}^{e(\mathfrak{b})}$, $e(\mathfrak{b}) \in \mathbb{Z}$, $\mathfrak{b} \in \mathcal{I}$. We prove the uniqueness. Suppose that there is a second factorization $\mathfrak{a} = \prod_{\mathfrak{b}\in\mathcal{I}} \mathfrak{b}^{f(\mathfrak{b})}$ with $f(\mathfrak{b}) \in \mathbb{Z}$ for all $\mathfrak{b} \in \mathcal{I}$. Let $\mathcal{I}_1, \mathcal{I}_2 \subset \mathcal{I}$ such that $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$, $e(\mathfrak{b}) \geq f(\mathfrak{b})$ for $\mathfrak{b} \in \mathcal{I}_1$ and $e(\mathfrak{b}) < f(\mathfrak{b})$ for $\mathfrak{b} \in \mathcal{I}_2$ Set

$$x(\mathfrak{b}) = |e(\mathfrak{b}) - f(\mathfrak{b})|\ ,\ \ \mathfrak{b} \in \mathcal{I}.$$

Then

$$\prod_{\mathfrak{b}\in\mathcal{I}_1} \mathfrak{b}^{x(\mathfrak{b})} = \prod_{\mathfrak{b}\in\mathcal{I}_2} \mathfrak{b}^{x(\mathfrak{b})} \tag{8.15}$$

and

$$x(\mathfrak{b}) \geq 0\ , \quad \mathfrak{b} \in \mathcal{I}.$$

We prove that $x(\mathfrak{b}) = 0$ for each $\mathfrak{b} \in \mathcal{I}$. Choose $\mathfrak{c} \in \mathcal{I}$ such that $x(\mathfrak{c}) > 0$. Assume w.l.o.g. $\mathfrak{c} \in \mathcal{I}_1$. In (8.15) add $\mathfrak{c}$ on both sides

$$\mathfrak{c} = \mathfrak{c} + \mathfrak{c}^{x(\mathfrak{c})} \cdot \prod_{\mathfrak{b}\in\mathcal{I}_1\setminus\{\mathfrak{c}\}} \mathfrak{b}^{x(\mathfrak{b})} = \mathfrak{c} + \prod_{\mathfrak{b}\in\mathcal{I}_2} \mathfrak{b}^{x(\mathfrak{b})} = \mathcal{O}$$

where the last equality follows from Corollary 8.5.13. This contradicts the assumption $\mathfrak{c} \neq \mathcal{O}$. □

8.6 Unique factorization into prime ideals

Let $\mathcal{O}$ be a quadratic order and let Δ be the discriminant of $\mathcal{O}$. In this section we show that all $\mathcal{O}$-ideals whose norm is coprime to the conductor of $\mathcal{O}$ are products of so called prime ideals. We also show that the factors in those products are uniquely determined. We first define prime ideals and maximal ideals.

8.6.1 Prime ideals

Definition 8.6.1.

1. *A* prime ideal *of $\mathcal{O}$ is an $\mathcal{O}$-ideal $\mathfrak{p}$ different from $\mathcal{O}$ that has the following property. Whenever a product $\alpha\beta$ of elements α, β in $\mathcal{O}$ belongs to $\mathfrak{p}$, then one of the factors belongs to $\mathfrak{p}$.*
2. *A* maximal $\mathcal{O}$-ideal *is an integral $\mathcal{O}$-ideal $\mathfrak{m}$ different from $\mathcal{O}$ that is not contained in another $\mathcal{O}$-ideal different from $\mathcal{O}$ and $\mathfrak{m}$.*

Example 8.6.2. The $\mathcal{O}$-ideal $\{0\}$ is a prime ideal of $\mathcal{O}$.

Here is a first observation.

Proposition 8.6.3. *Any maximal $\mathcal{O}$-ideal is a prime ideal of $\mathcal{O}$.*

Proof. Suppose that $\mathfrak{m}$ is a maximal $\mathcal{O}$-ideal. Let $\alpha, \beta \in \mathcal{O}$ such that $\alpha\beta \in \mathfrak{m}$. Assume that $\beta \notin \mathfrak{m}$. Then $\beta\mathcal{O} + \mathfrak{m}$ is an $\mathcal{O}$-ideal that properly contains $\mathfrak{m}$. Since $\mathfrak{m}$ is maximal, we have $\beta\mathcal{O} + \mathfrak{m} = \mathcal{O}$. This implies $\alpha\beta\mathcal{O} + \alpha\mathfrak{m} = \alpha\mathcal{O}$. But $\alpha\mathcal{O} = \alpha\beta\mathcal{O} + \alpha\mathfrak{m} \subset \mathfrak{m}$, since $\alpha\beta \in \mathfrak{m}$. Hence $\alpha \in \mathfrak{m}$. So $\mathfrak{m}$ is a prime ideal of $\mathcal{O}$. □

In Proposition 8.6.4 we will show that the converse of Proposition 8.6.3 is also true for non-zero prime ideals.

We determine all non-zero prime ideals of $\mathcal{O}$. For a prime number p with $\left(\frac{\Delta}{p}\right) \neq -1$ we define

$$\mathfrak{p}(\Delta, p) = p\mathbb{Z} + \mathbb{Z}\frac{b(\Delta, p) + \sqrt{\Delta}}{2} \tag{8.16}$$

with $b(\Delta, p)$ as defined in Section 3.4.

Proposition 8.6.4.

1. *If p is a prime number with $\left(\frac{\Delta}{p}\right) = -1$, then $p\mathcal{O}$ is an invertible prime ideal of $\mathcal{O}$.*
2. *If p is a prime number with $\left(\frac{\Delta}{p}\right) = 1$, then $\mathfrak{p}(\Delta, p)$ and $\sigma(\mathfrak{p}(\Delta, p))$ are distinct invertible prime ideals of $\mathcal{O}$.*
3. *If p is a prime number with $\left(\frac{\Delta}{p}\right) = 0$, then $\mathfrak{p}(\Delta, p)$ is a prime ideal of $\mathcal{O}$ which is invertible if and only if p does not divide the conductor of $\mathcal{O}$.*

There are no other non-zero prime ideals of $\mathcal{O}$. Also, all those prime ideals are maximal.

Proof. We show that the ideals from the proposition are all maximal $\mathcal{O}$-ideals. Then Proposition 8.6.3 implies that they are prime ideals. Let p be a prime number.

First assume that $\left(\frac{\Delta}{p}\right) \neq -1$. Then $\mathfrak{p}(\Delta, p)$ is defined and the norm of that ideal is p. By Lemma 8.5.3 the norm of any integral $\mathcal{O}$-ideal, which contains $\mathfrak{p}(\Delta, p)$, is a divisor of p. Hence that norm is either 1 or p. Since the norm of an integral $\mathcal{O}$-ideal is the index of that ideal in $\mathcal{O}$ as an additive subgroup, it follows that the only integral $\mathcal{O}$-ideals which contain $\mathfrak{p}(\Delta, p)$ are $\mathcal{O}$ and $\mathfrak{p}(\Delta, p)$ itself. So $\mathfrak{p}(\Delta, \mathcal{O})$ is a maximal $\mathcal{O}$-ideal. It follows that $\sigma(\mathfrak{p}(\Delta, p))$ is also a maximal $\mathcal{O}$-ideal. For $\left(\frac{\Delta}{p}\right) = 0$ we have $\mathfrak{p}(\Delta, p) = \sigma(\mathfrak{p}(\Delta, p))$.

Assume that $\left(\frac{\Delta}{p}\right) = -1$. Then there is no integral $\mathcal{O}$-ideal of norm p since by (8.14) any such $\mathcal{O}$-ideal is of the form $L(p, b, c)$ with a form (p, b, c) and by Proposition 3.4.5 such a form does not exist. As in the case $\left(\frac{\Delta}{p}\right) = 1$, we see that $p\mathcal{O}$ is a maximal invvertible $\mathcal{O}$-ideal.

By Proposition 8.4.14 the ideals $\mathfrak{p}(\Delta, p)$ and $\sigma(\mathfrak{p}(\Delta, p))$ are invertible if and only if the form $(p, b(\Delta, p), c(\Delta, p))$ is primitive. This, in turn, is true if and only if p does not divide the conductor of $\mathcal{O}$. For $\left(\frac{\Delta}{p}\right) = 1$, this is always true.

Now we show that there are no other prime ideals of $\mathcal{O}$. Let $\mathfrak{p}$ be a prime ideal of $\mathcal{O}$. We can write $\mathfrak{p} = mL(a, b, c)$ with a positive integer m and a form (a, b, c) of discriminant Δ with $a > 0$. Then ma is the smallest positive integer in $\mathfrak{p}$. We claim that ma is a prime number. Otherwise, $ma = uv$ with integers $u, v > 1$. Since $\mathfrak{p}$ is a prime ideal, $\mathfrak{p}$ must contain u or v which contradicts the minimality of ma. So there is a prime number p such that $m = p$ and $a = 1$ or $m = 1$ and $a = p$. In the first case, we have $P = p\mathcal{O}$. In the second case we have $P = L(p, b, c)$ with a prime form (p, b, c) of discriminant Δ. It remains to be shown that $p\mathcal{O}$ is not a prime ideal if Δ is a square mod $4p$. So assume that Δ is a square mod $4p$. The ideal $p\mathcal{O}$ is properly contained in $\mathfrak{p}(\Delta, p)$ and in $\sigma(\mathfrak{p}(\Delta, p))$ since $\mathrm{N}(p\mathcal{O}) = p^2$ and $\mathrm{N}(\mathfrak{p}(\Delta, p)) = \mathrm{N}(\sigma(\mathfrak{p}(\Delta, p))) = p$. Hence $p\mathcal{O}$ neither contains $(b(\Delta, p) + \sqrt{\Delta})/2$ nor $(b(\Delta, p) - \sqrt{\Delta})/2$. But $p\mathcal{O}$ contains the product $(b(\Delta, p) + \sqrt{\Delta})(b(\Delta, p) - \sqrt{\Delta})/4 = pc(\Delta, p)$. Hence, $p\mathcal{O}$ is not a prime ideal. □

Note that by Proposition 8.6.4 all non-zero prime ideals of $\mathcal{O}$ are maximal $\mathcal{O}$-ideals.

Example 8.6.5. We determine the first few prime ideals of the order $\mathcal{O} = \mathcal{O}_5$. Since 5 is a fundamental discriminant, the conductor of 5 is 1. By Proposition 8.6.4, all prime ideals of $\mathcal{O}$ are invertible. The discriminant 5 is 5 mod 8. Hence, $2\mathcal{O}$ is a prime ideal. Also, 5 is a quadratic nonresidue mod 3. So $3\mathcal{O}$ is a prime ideal. There is one prime ideal of $\mathcal{O}$ contained in $5\mathcal{O}$, namely $\mathfrak{p}(5, 5) = L(5, 5, 1)$. Also, $7\mathcal{O}$ is a prime ideal. Finally 5 is a quadratic residue mod 11. We have $b(5, 11) = 7$. Hence, we obtain the two prime ideals $\mathfrak{p}(5, 11) = L(11, 7, 1)$ and $\sigma(\mathfrak{p}(5, 11)) = L(11, -7, 1)$.

We determine the powers of the prime ideal $\mathfrak{p}(\Delta, p)$ for a prime number p with $\left(\frac{\Delta}{p}\right) \neq -1, 0$.

Proposition 8.6.6. *Let p be a prime number with $\left(\frac{\Delta}{p}\right) = 1$ and let e be a positive integer then*

$$\mathfrak{p}(\Delta, p)^e = \mathbb{Z}p^e + \mathbb{Z}\frac{b(\Delta, p^e) + \sqrt{\Delta}}{2} .$$

Proof. Exercise 8.7.13.

8.6.2 Unique factorization

We now prove that $\mathcal{O}$ ideals, whose norm is coprime to the conductor of $\mathcal{O}$, admit unique factorization into prime ideals.

Lemma 8.6.7. *Two different prime ideals of $\mathcal{O}$ are coprime.*

Proof. Let $\mathfrak{p}$ and $\mathfrak{q}$ be different prime ideals. Then the $\mathcal{O}$-ideal $\mathfrak{p}+\mathfrak{q}$ properly contains $\mathfrak{p}$. Since by Proposition 8.6.4 the prime ideal $\mathfrak{p}$ is a maximal $\mathcal{O}$-ideal, we must have $\mathfrak{p}+\mathfrak{q}=\mathcal{O}$. □

Theorem 8.6.8. *Any integral $\mathcal{O}$-ideal, whose norm is coprime to the conductor of $\mathcal{O}$, is a product of invertible prime ideals of $\mathcal{O}$. The factors in that product are uniquely determined.*

Proof. Let $\mathfrak{a}$ be an $\mathcal{O}$-ideal whose norm is coprime to the conductor of $\mathcal{O}$. We claim that $\mathfrak{a}$ is divisible by an invertible prime ideal. The index of $\mathfrak{a}$ in $\mathcal{O}$ is finite. Therefore, $\mathfrak{a}$ is contained in a maximal ideal $\mathfrak{p}$ which by Proposition 8.6.3 is a prime ideal. By Proposition 8.5.3 the norm of $\mathfrak{p}$ divides the norm of $\mathfrak{a}$. Therefore, the norm of $\mathfrak{p}$ is coprime to the conductor of $\mathcal{O}$. By Proposition 8.6.4 the prime ideal $\mathfrak{p}$ is invertible. So Lemma 8.5.5 implies that $\mathfrak{p}$ divides $\mathfrak{a}$. Consider $\mathfrak{c}=\mathfrak{a}\mathfrak{p}^{-1}$. If $\mathfrak{c}$ would equal $\mathfrak{a}$, then $\mathfrak{a}=\mathfrak{a}\mathfrak{p}$ and hence $\mathfrak{a}=\mathfrak{a}\mathfrak{p}^n\subseteq\mathfrak{p}^n$ for all n which would imply by Proposition 8.5.3 $\mathrm{N}(\mathfrak{p}^n)\mid\mathrm{N}(\mathfrak{a})$ again for all n. Thus, $\mathfrak{a}$ is properly contained in the integral ideal $\mathfrak{c}=\mathfrak{a}\mathfrak{p}^{-1}$ which has norm coprime to the conductor, and we can apply our argument recursively. We see that $\mathfrak{a}$ is a product of invertible prime ideals. The uniqueness of the decomposition follows from Theorem 8.5.14. □

In the next example, we show that, in general, integral $\mathcal{O}$-ideals whose norm is not coprime to the conductor of $\mathcal{O}$ do not factor into prime ideals.

Example 8.6.9. Let $\mathcal{O}=\mathcal{O}_{20}$. Consider the $\mathcal{O}$-ideal $2\mathcal{O}$. By Proposition 8.6.4 it is not a prime ideal. The only possible prime ideal divisors of $2\mathcal{O}$ is $L(2,2,-2)$. It is the only $\mathcal{O}$-ideal of norm 2. Hence, if $2\mathcal{O}$ is divisible by that prime ideal, then $2\mathcal{O}$ must be the square of that ideal. But $L(2,2,-2)^2=L(4,2,-1)\neq 2\mathcal{O}$.

We explicitly determine the factorization of an integral $\mathcal{O}$-ideal whose norm is coprime to the conductor of $\mathcal{O}$. We first discuss the case of principal ideals generated by integers. For a prime number p and an integer m denote by $e(p,m)$ the largest integer e such that p^e divides m. Thus the prime factorization of m is

$$m=\prod_{p\in\mathbb{P}}p^{e(p,m)}.$$

Proposition 8.6.10. *Let m be an integer which is coprime to the conductor of $\mathcal{O}$. Then*

$$m\mathcal{O} = \prod_{(\Delta/p)=-1} (p\mathcal{O})^{e(p,m)} \prod_{(\Delta/p)=0} \mathfrak{p}(\Delta,p)^{2e(p,m)} \prod_{(\Delta/p)=1} (\mathfrak{p}(\Delta,p)\sigma(\mathfrak{p}(\Delta,p)))^{e(p,m)}$$

is the factorization of $m\mathcal{O}$ into a product of invertible prime ideals of $\mathcal{O}$.

Proof. It follows from Proposition 8.6.4 and Exercise 8.7.14 that the prime ideal factorization given in the Proposition is correct. □

Next, we describe the factorization of primitive $\mathcal{O}$-ideals. We set

$$\mathbb{P}(\Delta) = \{p \in \mathbb{P} : \left(\frac{\Delta}{p}\right) \neq -1\}.$$

Proposition 8.6.11. *Consider an $\mathcal{O}$-ideal*

$$\mathfrak{a} = a\mathbb{Z} + \mathbb{Z}\frac{b+\sqrt{\Delta}}{2}$$

where (a, b, c) is an integral form of discriminant Δ. For any prime number $p \in \mathbb{P}(\Delta)$ let

$$\mathfrak{a}(p) = \begin{cases} \mathfrak{p}(\Delta,p) & \textit{if } b \equiv b(\Delta,p) \mod 2p \\ \sigma(\mathfrak{p}(\Delta,p)) & \textit{if } b \equiv -b(\Delta,p) \mod 2p. \end{cases}$$

Then

$$\mathfrak{a} = \prod_{p\in\mathbb{P}(\Delta)} \mathfrak{a}(p)^{e(p,a)}$$

is the factorization of $\mathfrak{a}$ into a product of invertible prime ideals of $\mathcal{O}$.

Proof. The discriminant Δ is a square mod $4a$ and therefore it is a square mod $4p$ for every prime number p that divides a. Hence, a factors completely into prime numbers in $\mathbb{P}(\Delta)$.

Let p be a prime divisor of a. Since the norm of the product of $\mathcal{O}$-ideals is the product of the norms of the factors, Proposition 8.6.4 implies that there are $e(a,p)$ many prime ideal factors of $\mathfrak{a}$ of norm p.

If $\left(\frac{\Delta}{p}\right) = 0$, then $\mathfrak{p}(\Delta,p)$ is the only prime ideal of norm p and $\mathfrak{p}(\Delta,p)^2 = p\mathcal{O}$. Hence $e(a,p) = 1$ since $\mathfrak{a}$ is a primitive $\mathcal{O}$-ideal.

If $\left(\frac{\Delta}{p}\right) = 1$, then by Proposition 8.6.4 there are two prime ideals of $\mathcal{O}$ of norm p, namely $\mathfrak{p}(\Delta,p)$ and $\sigma(\mathfrak{p}(\Delta,p))$. By Exercise 8.7.14, the product of those ideals is $p\mathcal{O}$. Since $\mathfrak{a}$ is a primitive $\mathcal{O}$-ideal, exactly one of those ideals divides $\mathfrak{a}$. Suppose that $\mathfrak{p}(\Delta,p)$ divides $\mathfrak{a}$. Then by Lemma 8.5.5 the ideal $\mathfrak{a}$ is contained in $\mathfrak{p}(\Delta,p)$. This implies $(b+\sqrt{\Delta})/2 = xp + y(b(\Delta,p)+\sqrt{\Delta})/2$ with integers x, y. So $y = 1$ and $b \equiv b(\Delta,p) \bmod 2p$. Likewise, if $\sigma(\mathfrak{p}(\Delta,p))$ divides $\mathfrak{a}$, then $b \equiv -b(\Delta,p) \bmod 2p$. Hence, the decomposition formula in the proposition is correct. □

Example 8.6.12. We determine the factorization of the $\mathcal{O}_{37}$-ideal $\mathfrak{a} = L(123, 23, 1)$. We have $123 = 3 * 41$. Now $\mathfrak{p}(37, 3) = L(3, 1, 3)$ and $\mathfrak{p}(37, 41) = (41, 23, 3)$. Also, $23 \equiv -1 \bmod 6$. Therefore, $\mathfrak{a} = L(3, -1, 3) * L(41, 23, 3)$.

8.7 Exercises

Exercise 8.7.1. Prove that the field of fractions of the quadratic order of discriminant Δ is given by (8.1).

Exercise 8.7.2. Let F be a quadratic number field. Prove that $F \cap \mathbb{R} = \mathbb{Q}$.

Exercise 8.7.3. Prove Proposition 8.4.5.

Exercise 8.7.4. Prove the third assertion of Theorem 8.1.4.

Exercise 8.7.5. Prove Proposition 8.4.8.

Exercise 8.7.6. Find integral $\mathcal{O}$-ideals $\mathfrak{a}$ and $\mathfrak{b}$ such that $\mathfrak{b}$ does not divide $\mathfrak{a}$ and $\mathfrak{a} \subset \mathfrak{b}$

Exercise 8.7.7. Find a quadratic order $\mathcal{O}$ and two invertible $\mathcal{O}$-ideals whose sum is not invertible. (*Hint:* $L(p^2, p, 1) + L(p^2, -p, 1) = pL(1, 1, 1)$.)

Exercise 8.7.8. Let $\mathfrak{a}$ be an invertible $\mathcal{O}$-ideal and let $\mathrm{N}(\mathfrak{a}) = ab$ be a factorization of $\mathrm{N}(\mathfrak{a})$ where a and b are coprime positive integers. Show that $\mathfrak{a}_a = a\mathcal{O} + \mathfrak{a}$ and $\mathfrak{a}_b = b\mathcal{O} + \mathfrak{a}$ are coprime $\mathcal{O}$-ideals and that $\mathfrak{a} = \mathfrak{a}_a \mathfrak{a}_b$

Exercise 8.7.9. Let $\mathfrak{a} = L(a, b, c)$ and $\mathfrak{b} = L(a', b', c')$ be two fractional $\mathcal{O}$-ideals $\mathfrak{a}$ and $\mathfrak{b}$. Prove that a $\mathbb{Z}$-basis of $\mathfrak{a} + \mathfrak{b}$ is $\bigl(\gcd(a, a', \frac{b+b'}{2}), (b + \sqrt{\Delta})/2\bigr)$.

Exercise 8.7.10. Show that the only integral $\mathcal{O}$-ideal that divides an order $\mathcal{O}$ is $\mathcal{O}$.

Exercise 8.7.11. Let $\mathfrak{a}$ be an invertible $\mathcal{O}$-ideal and let $\alpha, \beta \in F^*$. Show that $\alpha\mathfrak{a} = \beta\mathfrak{a}$ if and only if α/β is a unit in $\mathcal{O}$.

Exercise 8.7.12. Show that two $\mathcal{O}$-ideals are coprime if their norms are coprime.

Exercise 8.7.13. Prove Proposition 8.6.6.

Exercise 8.7.14. Let p be a prime number. Prove that

$$p\mathcal{O} = \begin{cases} \mathfrak{p}(\Delta, p)^2 & \text{if } \left(\frac{\Delta}{p}\right) = 0, \\ \mathfrak{p}(\Delta, p)\sigma(\mathfrak{p}(\Delta, p)) & \text{if } \left(\frac{\Delta}{p}\right) = 1. \end{cases}$$

Exercise 8.7.15. Prove that the regulator R_Δ of $\mathcal{O}_\Delta$ is bounded from below by 0.48.

9

Class Groups

Let Δ be a quadratic discriminant, that is, Δ is an integer that is not a square and $\Delta \equiv 0, 1 \pmod 4$. Let $\mathcal{O}$ be the quadratic order of discriminant Δ and let F be the field of fractions of $\mathcal{O}$. In this chapter we show that the set of equivalence classes of integral primitive forms of discriminant Δ and the set of equivalence classes of invertible $\mathcal{O}$-ideals are finite Abelian groups, and we will discuss computational problems concerning those groups such as extracting roots, computing element orders and discrete logarithms, and determining the group structure.

9.1 Ideal classes

In this section we define equivalence of $\mathcal{O}$-ideals and we explain how to use reduction theory to decide ideal equivalence. Let $\mathfrak{a}$ be a fractional $\mathcal{O}$-ideal,

$$\mathfrak{a} = qL(f_{\mathfrak{a}}), \quad q \in \mathbb{Q}_{>0} \; .$$

Also write

$$f_{\mathfrak{a}} = (a, b, c)$$

which is by Definition 8.4.6 a normal form with first coefficient $a > 0$.

9.1.1 Equivalence

In Definition 4.4.8 we have introduced equivalence of lattices. In the special case of $\mathcal{O}$-ideals this leads to the following definition.

Definition 9.1.1. *Two fractional $\mathcal{O}$-ideals $\mathfrak{a}$ and $\mathfrak{b}$ are called* equivalent *if there is $\alpha \in F^*$ such that $\mathfrak{b} = \alpha\mathfrak{a}$. They are called* properly equivalent *if there is $\alpha \in F^*$ such that $\mathrm{N}(\alpha) > 0$ and $\mathfrak{b} = \alpha\mathfrak{a}$.*

Example 9.1.2. The $\mathcal{O}_5$-ideals $\mathcal{O}_5$ and $L(5, 5, 1)$ are equivalent since $L(5, 5, 1) = \sqrt{5} \cdot \mathcal{O}_5$.

Equivalence and proper equivalence of fractional $\mathcal{O}$-ideals are equivalence relations.

Definition 9.1.3. *An equivalence class of $\mathcal{O}$-ideals is called an* $\mathcal{O}$-ideal class. *A proper equivalence class of $\mathcal{O}$-ideals is called a* proper $\mathcal{O}$-ideal class.

We formulate two computational problems.

Problem 9.1.4 (Equivalence problem for $\mathcal{O}$-ideals). Given two fractional $\mathcal{O}$-ideals $\mathfrak{a}$, $\mathfrak{b}$.

1. Decide whether $\mathfrak{a}$ and $\mathfrak{b}$ are equivalent.
2. If $\mathfrak{a}$ and $\mathfrak{b}$ are equivalent, find $\alpha \in F^*$ with $\mathfrak{b} = \alpha\mathfrak{a}$.

Problem 9.1.5 (Proper equivalence problem for $\mathcal{O}$-ideals). Given two fractional $\mathcal{O}$-ideals $\mathfrak{a}$, $\mathfrak{b}$.

1. Decide whether $\mathfrak{a}$ and $\mathfrak{b}$ are properly equivalent.
2. If $\mathfrak{a}$ and $\mathfrak{b}$ are properly equivalent, find $\alpha \in F^*$ with $\mathrm{N}(\alpha) > 0$ and $\mathfrak{b} = \alpha\mathfrak{a}$.

Those equivalence problems can be solved by means of reduction theory for $\mathcal{O}$-ideals which is explained in the next sections.

9.1.2 Reduced $\mathcal{O}$-ideals

We define reduced $\mathcal{O}$-ideals. We use the correspondence between forms and ideals from Section 8.4.1.

Definition 9.1.6. *The ideal $\mathfrak{a}$ is called* reduced *if $q = 1$ and $f_{\mathfrak{a}}$ is a reduced form.*

We recall a few important results in the language of ideals.

Proposition 9.1.7. *Let $\Delta < 0$.*

1. *The ideal $\mathfrak{a}$ is reduced if and only if $a = \mathrm{N}(\mathfrak{a}) = \min\{\mathrm{N}(\alpha) : \alpha \in \mathfrak{a}, \alpha \neq 0\}$ and $b \geq 0$ for $a = c$.*
2. *If $\mathfrak{a}$ is reduced, then $\mathrm{N}(\mathfrak{a}) < \sqrt{|\Delta|/3}$.*
3. *If $\mathrm{N}(\mathfrak{a}) < \sqrt{|\Delta|}/2$, then $\mathfrak{a}$ is reduced.*

Proof. The assertions follow from Theorem 5.7.6, Lemma 5.4.1, and Lemma 5.5.1. □

Proposition 9.1.8. *Let $\Delta > 0$ and let $\mathfrak{a}$ be an integral primitive $\mathcal{O}$-ideal.*

1. *The ideal $\mathfrak{a}$ is reduced if and only if there is no non-zero α in $\mathfrak{a}$ with $|\alpha| < \mathrm{N}(\mathfrak{a})$ and $\sigma(\alpha) < \mathrm{N}(\mathfrak{a})$.*
2. *If $\mathfrak{a}$ is reduced, then $\mathrm{N}(\mathfrak{a}) < \sqrt{\Delta}$.*
3. *If $\mathrm{N}(\mathfrak{a}) < \sqrt{|\Delta|}/2$, then $\mathfrak{a}$ is reduced.*

Proof. The assertions follow from Theorem 6.8.6, Lemma 6.2.7, and Lemma 6.5.1. □

9.1.3 Reduction of $\mathcal{O}$-ideals

We define the reduction operator for fractional $\mathcal{O}$-ideals. We set

$$\rho(\mathfrak{a}) = L(\rho(f_{\mathfrak{a}})) \; . \tag{9.1}$$

We determine $\rho(\mathfrak{a})$ more explicitly. Define

$$\gamma(\mathfrak{a}) = \frac{-b+\sqrt{\Delta}}{2aq} = \frac{-2c}{q(b+\sqrt{\Delta})} \; . \tag{9.2}$$

Lemma 9.1.9. *We have*

$$\rho(\mathfrak{a}) = \gamma(\mathfrak{a})\mathfrak{a}.$$

Proof. Since for any form g the lattice $L(g)$ is an invariant of the Γ-orbit of g we have $\rho(\mathfrak{a}) = L(c, -b, a)$. Now

$$\frac{-b+\sqrt{\Delta}}{2aq}\mathfrak{a} = \frac{-b+\sqrt{\Delta}}{2a}L(a,b,c) = \mathbb{Z}c + \mathbb{Z}\frac{-b+\sqrt{\Delta}}{2} = L(c,-b,a) = \rho(\mathfrak{a}) \; .$$

□

The following lemma explains the connection between the reducing factor $\gamma(\mathfrak{a})$ of the ideal $\mathfrak{a}$ and the point associated to $f_{\mathfrak{a}}$ in $\mathbb{C}$ or A_1, respectively. For $\Delta > 0$ we use the identification (8.2).

Lemma 9.1.10. *Let $\mathfrak{a}$ be a primitive ideal. If $\Delta < 0$, then $\gamma(\mathfrak{a}) = -\theta(f_{\mathfrak{a}})^{\sigma}$. If $\Delta > 0$, then $\gamma(\mathfrak{a}) = -\theta_2(f_{\mathfrak{a}})$.*

Proof. This follows from Definition 4.3.14 and (9.2). □

Lemma 9.1.11. *If $\Delta > 0$ and $\mathfrak{a}$ is reduced, then $\gamma(\mathfrak{a}) > 0$.*

Proof. By definition $b < \sqrt{\Delta}$, and $a, q > 0$. □

Lemma 9.1.12. *If $\mathfrak{a}$ is primitive, but not reduced, then $|\gamma(\mathfrak{a})|, |\gamma(\mathfrak{a})^{\sigma}| < 1$.*

Proof. For $\Delta < 0$ the assertion follows from Lemmas 9.1.10 and 5.13.2. For $\Delta > 0$ this follows from Lemma 9.1.10 and Corollary 6.3.3. □

We explain an algorithm for reducing $\mathfrak{a}$. If $\Delta < 0$, then we apply the reduction algorithm from Section 5.3 to the form $f_{\mathfrak{a}}$. If $\Delta > 0$, then we apply the reduction algorithm from Section 6.4 to $f_{\mathfrak{a}}$. Let T be the reducing transformation and let (s, u) be its first column. Set

$$\alpha(\mathfrak{a}) = \frac{1}{q}(s + u\frac{b-\sqrt{\Delta}}{2a}). \tag{9.3}$$

Then Proposition 4.4.5 implies

$$L(fT) = \alpha(\mathfrak{a})\mathfrak{a} \,. \tag{9.4}$$

We call $\alpha(\mathfrak{a})$ the *reducing number* for $\mathfrak{a}$. By Lemma 5.6.1 and Lemma 6.6.1 we have

$$|s|, |u| \leq K + \frac{2|K|}{\sqrt{|\Delta|}} \tag{9.5}$$

where $K = \max\{|a|, |b|, |c|\}$. Algorithm `reduce`$(\mathfrak{a})$ receives as input the $\mathcal{O}$-ideal $\mathfrak{a}$ and returns the pair $(\alpha(\mathfrak{a})\mathfrak{a}, \alpha(\mathfrak{a}))$.

Example 9.1.13. We use Example 5.3.10 to illustrate ideal reduction. Let

$$\mathfrak{a} = \mathbb{Z} \cdot 195751 + \mathbb{Z}\frac{37615 + \sqrt{-3}}{2} \,.$$

Then

$$f_{\mathfrak{a}} = (195751, 37615, 1807) \,.$$

Applying the reduction operator to $f_{\mathfrak{a}}$ we obtain

$$(1, 1, 1) = f_{\mathfrak{a}} \begin{pmatrix} -22 & -49 \\ 229 & 510 \end{pmatrix} \,.$$

The form $(1, 1, 1)$ is reduced. In fact, that form is the only reduced form of discriminant -3. Hence, the reducing number for $\mathfrak{a}$ is

$$\alpha(\mathfrak{a}) = \big(-22 + 229\frac{1 - \sqrt{-3}}{2}\big)$$

and we have

$$\mathbb{Z} + \mathbb{Z}\frac{1 + \sqrt{-3}}{2} = \big(-22 + 229\frac{1 - \sqrt{-3}}{2}\big)\Big(\mathbb{Z} \cdot 195751 + \mathbb{Z}\frac{37615 + \sqrt{-3}}{2}\Big) \,.$$

9.1.4 Equivalence testing in imaginary quadratic orders

Let $\Delta < 0$. Then all numbers in F^* have positive norm. Therefore, the equivalence problem and the proper-equivalence problem are identical. Also, by Theorem 5.7.7 every ideal class of $\mathcal{O}$ contains exactly one reduced ideal.

To decide equivalence of two fractional $\mathcal{O}$-ideals $\mathfrak{a}$ and $\mathfrak{b}$ we can proceed as follows. We compute the reduced ideals $\mathfrak{a}'$ and $\mathfrak{b}'$ that are equivalent to $\mathfrak{a}$ and $\mathfrak{b}$, respectively. If $\mathfrak{a}' = \mathfrak{b}'$, then $\mathfrak{a}$ and $\mathfrak{b}$ are properly equivalent. Otherwise, $\mathfrak{a}$ and $\mathfrak{b}$ are not equivalent. Also, if $\mathfrak{a}' = \mathfrak{b}'$, then we have

$$\alpha(\mathfrak{a})\mathfrak{a} = \mathfrak{a}' = \mathfrak{b}' = \alpha(\mathfrak{b})\mathfrak{b} \,.$$

If we set

$$\alpha = \frac{\alpha(\mathfrak{a})}{\alpha(\mathfrak{b})} ,$$

then we have

$$\mathfrak{b} = \alpha\mathfrak{a} .$$

This solves the equivalence problem for ideals in imaginary quadratic orders. Theorem 5.6.6 implies that the running time of this algorithm is quadratic in the length of the input.

Example 9.1.14. Let

$$\mathfrak{a} = \mathbb{Z} \cdot 195751 + \mathbb{Z}\frac{37615 + \sqrt{-3}}{2}$$

and

$$\mathfrak{b} = \mathbb{Z} + \mathbb{Z}\frac{1 + \sqrt{-3}}{2}.$$

As we have seen in Example 9.1.13, the ideal $\mathfrak{b}$ is reduced and reducing $\mathfrak{a}$ yields $\mathfrak{b}$. Also,

$$\alpha(\mathfrak{a}) = \left(-22 + 229\frac{1 - \sqrt{-3}}{2}\right)$$

and

$$\alpha(\mathfrak{b}) = 1.$$

Hence, the $\mathcal{O}$-ideals $\mathfrak{a}$ and $\mathfrak{b}$ are equivalent and we have $\mathfrak{b} = \alpha(\mathfrak{a})\mathfrak{a}$.

9.1.5 Equivalence testing in real quadratic orders

Let $\Delta > 0$. Let $\mathfrak{a}$ be a fractional $\mathcal{O}$-ideal. Then by the results of Section 6.10 the equivalence class of the form $f_{\mathfrak{a}}$ contains a cycle $(f_1, \dots, f_l)$ of reduced forms. We call $(L(f_1), \dots, L(f_l))$ the *cycle of reduced ideals* in the ideal class of $\mathfrak{a}$. The following simple lemma implies that this cycle is enumerated starting from $\mathfrak{b}_1 = L(f_1)$ by successive application of the operator ρ.

Lemma 9.1.15. *If $\mathfrak{a}$ is reduced, then $f_{\rho(\mathfrak{a})} = (\tau\rho)(f_{\mathfrak{a}})$.*

Proof. If $f_{\mathfrak{a}} = (a, b, c)$, then $a > 0$ and by Lemma 6.2.7 $c < 0$. Thus Lemma 9.1.15 follows immediately from the definitions of the operators ρ in (9.1) and τ in (6.27).

Proposition 6.10.3 allows us to deduce the considerably deeper fact that the cycle contains all reduced ideals in the ideal class of $\mathfrak{a}$.

Example 9.1.16. By Example 6.10.1 the cycle of

$$\mathfrak{a} = \mathbb{Z} + \mathbb{Z}\frac{3 + \sqrt{17}}{2}$$

is

$$\left(\mathbb{Z} + \mathbb{Z}\frac{3 + \sqrt{17}}{2}, \mathbb{Z} \cdot 2 + \mathbb{Z}\frac{1 + \sqrt{17}}{2}, \mathbb{Z} \cdot 2 + \mathbb{Z}\frac{3 + \sqrt{17}}{2}\right).$$

To decide equivalence of the fractional $\mathcal{O}$-ideals $\mathfrak{a}$ and $\mathfrak{b}$ we can proceed as follows. We compute the cycle of $\mathfrak{a}$. Also, we apply the reduction algorithm to $\mathfrak{b}$. Let $\mathfrak{b}'$ be the result of that reduction. If $\mathfrak{b}'$ is in the cycle of $\mathfrak{a}$, then $\mathfrak{a}$ and $\mathfrak{b}$ are equivalent. Otherwise, they are not. If $\mathfrak{a}$ and $\mathfrak{b}$ are equivalent, then using the method explained in Section 6.11 we can find a transformation $U \in \mathrm{GL}(2,\mathbb{Z})$ with $f_{\mathfrak{b}} = f_{\mathfrak{a}}U$. Proposition 4.4.5 can be used to determine $\alpha \in F^*$ with $\mathfrak{b} = \alpha\mathfrak{a}$.

Suppose that $\mathfrak{a}$ and $\mathfrak{b}$ are equivalent and that α is found with $\mathfrak{b} = \alpha\mathfrak{a}$. We explain how proper equivalence of $\mathfrak{a}$ and $\mathfrak{b}$ can be decided. If $\mathrm{N}(\alpha) > 0$, then $\mathfrak{a}$ and $\mathfrak{b}$ are properly equivalent. Let $\mathrm{N}(\alpha) < 0$. Then we determine the fundamental unit ε of the ring of multipliers of $\mathfrak{a}$ as explained in Section 8.3. If its norm is negative, then $\mathrm{N}(\varepsilon\alpha) > 0$. Hence $\mathfrak{a}$ and $\mathfrak{b}$ are properly equivalent since $\mathfrak{b} = \varepsilon\alpha\mathfrak{a}$. However, if $\mathrm{N}(\varepsilon) > 0$, then $\mathfrak{a}$ and $\mathfrak{b}$ are not properly equivalent. This can be seen as follows. Suppose that $\mathfrak{b} = \alpha'\mathfrak{a}$ with $\alpha' \in F^*$ and $\mathrm{N}(\alpha') > 0$. Then $\alpha\mathfrak{a} = \mathfrak{b} = \alpha'\mathfrak{a}$. Hence α/α' is a unit of negative norm in the ring of multipliers of $\mathfrak{a}$ which is impossible.

Example 9.1.17. Let

$$\mathfrak{a} = \mathbb{Z} + \mathbb{Z}\frac{7+\sqrt{73}}{2}$$

and

$$\mathfrak{b} = \mathbb{Z}\cdot 4 + \mathbb{Z}\frac{3+\sqrt{73}}{2} \ .$$

Then

$$f_{\mathfrak{a}} = (1,7,-6)\ , \quad f_{\mathfrak{b}} = (4,3,-4)\ .$$

By Example 6.10.6 both forms are reduced and $f_{\mathfrak{b}}$ is in the cycle of $f_{\mathfrak{a}}$. Also, we have

$$f_{\mathfrak{b}} = f_{\mathfrak{a}}\begin{pmatrix} 7 & 10 \\ 9 & 13 \end{pmatrix}$$

It follows from Proposition 4.4.5 that

$$\mathfrak{b} = \left(7 + 9\frac{1-\sqrt{73}}{2}\right)\cdot\mathfrak{a}\ .$$

9.2 Ambiguous ideals and classes

In Sections 1.4.3 and 2.8 we have introduced and characterized ambiguous forms and ambiguous form classes. In Section 6.14 we have seen that the cycle in an ambiguous class of integral irreducible indefinite forms is symmetric. In this section we translate those notions and results into the language of ideals.

Definition 9.2.1.

1. *An $\mathcal{O}$-ideal $\mathfrak{a}$ is called* ambiguous *if the corresponding form $f_{\mathfrak{a}}$ is* ambiguous.
2. *An $\mathcal{O}$-ideal $\mathfrak{a}$ is called* symmetric *if the corresponding form $f_{\mathfrak{a}}$ is* symmetric.
3. *An equivalence class of reduced $\mathcal{O}$-ideals is called* ambiguous *if the corresponding equivalence class of reduced forms is ambiguous.*

We characterize ambiguous and symmetric $\mathcal{O}$-ideals.

Proposition 9.2.2. *1. An $\mathcal{O}$-ideal $\mathfrak{a}$ is ambiguous if and only if $\mathfrak{a} = \sigma(\mathfrak{a})$.*
2. A reduced $\mathcal{O}$-ideal $\mathfrak{a}$ is symmetric if and only if $\rho(\mathfrak{a}) = \sigma(\mathfrak{a})$.

Proof. Let $f = f_{\mathfrak{a}} = (a, b, c)$.

1. Let $\mathfrak{a}$ be ambiguous. Then $f = (a, ka, c)$ with $k \in \mathbb{Z}$. Hence $\mathfrak{a} = a\mathbb{Z} + (ka + \sqrt{\Delta})/2\mathbb{Z} = a\mathbb{Z} + (ka - 2ka + \sqrt{\Delta})/2\mathbb{Z} = a\mathbb{Z} + (-ka + \sqrt{\Delta})/2\mathbb{Z} = a\mathbb{Z} + (ka - \sqrt{\Delta})/2\mathbb{Z} = \sigma(\mathfrak{a})$. Conversely, assume that $\mathfrak{a} = \sigma(\mathfrak{a})$. Then $a\mathbb{Z} + (b + \sqrt{\Delta})/2\mathbb{Z} = a\mathbb{Z} + (b - \sqrt{\Delta})/2\mathbb{Z}$. So $b \in \mathfrak{a}$. But $\mathfrak{a} \cap \mathbb{Z} = a\mathbb{Z}$. Hence $b = ka$ with $k \in \mathbb{Z}$. So $\mathfrak{a}$ is ambiguous.

2. Let $\mathfrak{a}$ be reduced and symmetric. Then $f = (a, b, -a)$. Hence, $\rho(\mathfrak{a}) = a\mathbb{Z} + (b - \sqrt{\Delta})/2\mathbb{Z} = \sigma(\mathfrak{a})$. Conversely, if $\rho(\mathfrak{a}) = \sigma(\mathfrak{a})$, then $c\mathbb{Z} + (b - \sqrt{\Delta})/2\mathbb{Z} = a\mathbb{Z} + (b - \sqrt{\Delta})/2\mathbb{Z}$. Hence $c = \pm a$. But a and c have opposite sign by Lemma 6.2.7. Hence $f = (a, b, -a)$. □

As lattices, ambiguous ideals are symmetric with respect to the real line $\mathbb{R}$.

Let $\Delta > 0$. If the equivalence class of the $\mathcal{O}$-ideal $\mathfrak{a}$ is ambiguous, then by Theorem 6.14.6 the cycle of $\mathfrak{a}$ and the cycle of $\sigma(\mathfrak{a})$ are the same. In the next theorem this is used to prove that the embedding of the reduced forms into the cycle of $\mathfrak{a}$ is symmetric in one of three possible ways.

Theorem 9.2.3. *Assume that the equivalence class of the $\mathcal{O}$-ideal $\mathfrak{a}$ is ambiguous. Let $\mathfrak{a}_i = \rho^i(\mathfrak{a})$, $i \in \mathbb{Z}$.*

1. *If the length l of the cycle of $\mathfrak{a}$ is odd, then this cycle contains one ambiguous and one symmetric ideal. If $\mathfrak{a}_0$ is the ambiguous ideal in the cycle and $l = 2k + 1$ with $k \geq 0$, then the cycle is $\big(\sigma(\mathfrak{a}_k), \dots, \sigma(\mathfrak{a}_1), \mathfrak{a}_0, \mathfrak{a}_1, \dots, \mathfrak{a}_k\big)$ and $\mathfrak{a}_k$ is the symmetric ideal.*
2. *Let the length l of the cycle of $\mathfrak{a}$ be even, that is, $l = 2k$, $k \geq 1$. Then this cycle contains two symmetric and no ambiguous or two ambiguous and no symmetric ideal. In the first case, the cycle is $\big(\sigma(\mathfrak{a}_{k-1}), \dots, \sigma(\mathfrak{a}_1), \sigma(\mathfrak{a}_0), \mathfrak{a}_0, \mathfrak{a}_1, \dots, \mathfrak{a}_{k-1}\big)$ and $\mathfrak{a}_0$ and $\mathfrak{a}_k$ are the symmetric ideals. In the second case, the period is $\big(\sigma(\mathfrak{a}_{k-1}), \dots, \sigma(\mathfrak{a}_1), \mathfrak{a}_0, \mathfrak{a}_1, \dots, \mathfrak{a}_k\big)$ and f_0 and $\mathfrak{a}_k$ are the ambiguous ideals.*

Proof. Exercise 10.4.4. □

9.3 Fundamentals on class groups

9.3.1 Definition

We introduce a few groups.

Definition 9.3.1.
1. *By $\mathcal{I}_\Delta$ we denote the group of all invertible $\mathcal{O}$-ideals.*
2. *By $\mathcal{P}_\Delta$ we denote the group of all principal $\mathcal{O}$-ideals.*
3. *By $\mathcal{P}_\Delta^+$ we denote the group of all principal $\mathcal{O}$-ideals which have a generator of positive norm.*

The group $\mathcal{P}_\Delta^+$ is a subgroup of $\mathcal{P}_\Delta$ which, in turn, is a subgroup of $\mathcal{I}_\Delta$. Also, the coset of a fractional $\mathcal{O}$-ideal $\mathfrak{a}$ in the quotient group $\mathcal{I}_\Delta/\mathcal{P}_\Delta$ is its equivalence class. The coset of $\mathfrak{a}$ in the quotient group $\mathcal{I}_\Delta/\mathcal{P}_\Delta^+$ is its proper equivalence class.

Definition 9.3.2. *The* class group *of $\mathcal{O}$ is the quotient group $\mathcal{I}_\Delta/\mathcal{P}_\Delta$. It is denoted by Cl_Δ. The* narrow class group *of $\mathcal{O}$ is the quotient group $\mathcal{I}_\Delta/\mathcal{P}_\Delta^+$. It is denoted by Cl_Δ^+.*

Theorem 9.3.3. *Class group and narrow class group of $\mathcal{O}$ are finite Abelian groups. The order of the class group is the class number $h(\Delta)$. The order of the narrow class group is the narrow class number $h^+(\Delta)$.*

Proof. The assertion follows from Corollary 5.4.2 and Theorem 6.17.1. □

If $\mathfrak{a}$ is an $\mathcal{O}$-ideal, then we denote by $[\mathfrak{a}]$ the equivalence class of $\mathfrak{a}$. Also, if (a, b, c) is a form of discriminant Δ, then we denote by $[a, b, c]$ the equivalence class of the $\mathcal{O}$-ideal $L(a, b, c)$.

9.3.2 Imaginary quadratic class groups

Let $\Delta < 0$. Then all non-zero elements in $\mathbb{Q}(\sqrt{\Delta})$ have positive norm. Therefore, if two fractional $\mathcal{O}$-ideals are equivalent, then they are properly equivalent. This implies that the class group Cl_Δ is equal to the narrow class group Cl_Δ^+. Also, it follows from Corollary 4.6.3 that the equivalence classes of invertible fractional $\mathcal{O}$-ideals can be identified with the equivalence classes of integral positive definite primitive forms of discriminant Δ. Therefore, the elements of Cl_Δ can also be considered as equivalence classes of integral primitive forms of discriminant Δ. By Theorem 5.7.7, each equivalence class of primitive positive definite quadratic forms contains exactly one reduced form. We represent each class in Cl_Δ by its reduced representative (a, b, c). We have the following estimate for the coefficients of the reduced forms that represent the classes.

Proposition 9.3.4. *If (a, b, c) is an integral positive definite reduced form of discriminant Δ, then $0 < a < \sqrt{|\Delta|/3}$, $|b| \le a$, and $c \le \left(\sqrt{|\Delta|/3} + |\Delta|\right)/4$.*

Proof. Since (a,b,c) is reduced we have $|b| \leq a$ and $a < \sqrt{|\Delta|/3}$ by Lemma 5.4.1. Also, we have $4ac = b^2+|\Delta| \leq a^2+|\Delta|$. Hence $c \leq (a+|\Delta|)/4 \leq \left(\sqrt{|\Delta|/3}+|\Delta|\right)/4$. □

We explain the basic group operations in Cl_Δ and their complexity.

1. (Equality) To decide equality of two elements of the class group, we compare the reduced representatives . If they are equal, then the classes are equal. Otherwise they are not. By Proposition 9.3.4 the binary length of those representatives is $\mathrm{O}(\log|\Delta|)$. Hence, this equality test takes time $\mathrm{O}(\log|\Delta|)$.
2. (Inverse) Let $[a,b,c] \in \mathrm{Cl}_\Delta$. Then $[a,b,c]^{-1} = [a,-b,c]$. The reduced representative in that class is $(a,-b,c)$ if $-a < b < a$ and (a,b,c) if $b = a$. By Proposition 9.3.4 the binary length of this representative is $\mathrm{O}(\log|\Delta|)$. Hence, computing the inverse takes time $\mathrm{O}(\log|\Delta|)$.
3. (Product) Two classes in Cl_Δ are multiplied by multiplying their reduced representative and reducing the result. If we use the product formula from Section 7.3.4 and the reduction algorithm from Section 5.3, then Theorem 5.6.6 implies that computing the product of two classes takes time $\mathrm{O}\left((\log|\Delta|)^2\right)$.

Example 9.3.5. Let $\Delta = -31$. Using algorithm `classNumber` from Section 5.11 we find that $h(-31) = 3$. Hence, the class group Cl_{-31} is cyclic of order 3. The class $[2,1,4]$ is not the identity in the class group. Therefore, its order is 3 and it generates the class group. We have $[2,1,4]^2 = [2,-1,4]$ and $[2,1,4]^3 = [1,1,8]$.

9.3.3 Real quadratic class groups

Let $\Delta > 0$. It follows from Corollary 4.6.3 that the equivalence classes of invertible fractional $\mathcal{O}$-ideals can be identified with the equivalence classes of integral indefinite primitive forms of discriminant Δ. Therefore, the elements of Cl_Δ can also be considered as equivalence classes of integral primitive forms of discriminant Δ. Each equivalence class of integral indefinite forms contains exactly one cycle of reduced forms. By Proposition 6.10.3, that cycle consists of all reduced forms (a,b,c) with $a > 0$ in the equivalence class of f. It follows that the class groups Cl_Δ can also be identified with the set of cycles of reduced forms (a,b,c) of discriminant Δ with $a > 0$. We represent an element of the class group by a reduced form in the corresponding cycle.

We have the following estimate for the coefficients of the reduced forms that represent the classes.

Proposition 9.3.6. *If (a,b,c) is an integral indefinite reduced form of discriminant Δ then $|a|+|c| < \sqrt{\Delta}$ and $0 < b < \sqrt{\Delta}$.*

Proof. The assertion follows from Lemma 6.2.7 and Definition 6.2.1. □

We explain the basic group operations and their complexity.

1. (Equality) Deciding equality of two classes in Cl_Δ is much more complicated than in the imaginary quadratic case. One possibility is to compute the cycle of reduced form that represents the one class and to check whether the reduced form that represents the other class belongs to that cycle. By Proposition 6.13.1 this equality test takes time $\mathrm{O}\big(l(\log\Delta)^2\big)$ where l is the length of the cycle and typically has order of magnitude $\sqrt{\Delta}$.
2. (Inverse) Let $[a,b,c] \in \mathrm{Cl}_\Delta$. Then $[a,b,c]^{-1} = [a,-b,c]$. A reduced representative in the inverse class $[a,b,c]^{-1}$ is $(-c,b,a)$. By Proposition 9.3.6, computing the inverse takes time $\mathrm{O}(\log\Delta)$.
3. (Product) Two classes are multiplied by multiplying their reduced representatives and reducing the result. If we use the product algorithm from Section 7.3.4 and the reduction algorithm from Section 6.4 then Theorem 6.6.4 implies that computing the product of two classes takes time $\mathrm{O}\big((\log\Delta)^2\big)$.

By Corollary 4.6.3, the elements of the proper class group Cl^+_Δ can be identified with the proper equivalence classes of integral forms of discriminant Δ. By Theorem 6.17.3 and Proposition 8.3.8, the narrow class group Cl^+_Δ is equal to the class group Cl_Δ if and only if the fundamental unit of $\mathcal{O}$ has norm -1. If the norm of the fundamental unit of $\mathcal{O}$ has norm 1, then the class group Cl_Δ is isomorphic to a quotient group $\mathrm{Cl}^+_\Delta/\big\langle L(-c,b,1)\big\rangle$ where $(1,b,c)$ is the principal form of discriminant Δ (see Exercise 9.8.3).

Example 9.3.7. Let $\Delta = 21$. Then $h(\Delta) = 1$ and $\big((1,3,-3),(3,3,-1)\big)$ is the only cycle of reduced forms of discriminant 21. The class group $\mathrm{Cl}(21)$ is trivial.

However, since the period length of the above cycle is even, the narrow class number $h^+(12)$ is 2. The narrow class group $\mathrm{Cl}^+(21)$ is $\big\{[-1,1,3]_+, [1,1,3]_+ = [-1,1,3]_+^2\big\}$ where $[\mathfrak{a}]_+$ denotes the narrow equivalence class of $\mathfrak{a}$.

9.3.4 The class number formula

The analytic class number formula connects the class number h_Δ with the value of the series

$$L(1,\chi_\Delta) = \sum_{n=1}^{\infty} \chi_\Delta(n) n^{-s}\,, \quad \text{where } \chi_\Delta(n) = \left(\frac{\Delta}{n}\right) \tag{9.6}$$

at $s = 1$. The series is called a *Dirichlet series* . It was introduced by Dirichlet in his proof that an arithmetic progression contains an infinitude of primes if it does not have a common divisor.

For a proof that the series (9.6) indeed converges for $s = 1$, see [Apo86] Chapter 11.6. There it is proven that in general a series of the form

$$\sum_{n=1}^{\infty} f(n) n^{-s}$$

converges for s with $\Re s > 0$ if the sums

$$\sum_{n=1}^{N} f(n)$$

are bounded independent of N. We prove this boundedness for $f(n) = \left(\frac{\Delta}{n}\right)$. It is true in general for non-trivial homomorphisms from a group G to $\mathbb{C}^*$. Such homomorphisms are called *characters.* The group in question here is $G = (\mathbb{Z}/|\Delta|\mathbb{Z})^*$. Characters on this group are also called *Dirichlet* characters.

Lemma 9.3.8. *Let χ be a character of G. Denote $|G|$ by n. Then*

$$\sum_{g\in G} \chi(g) = \begin{cases} n \text{ if } \chi(g) = 1 \text{ for all } g \in G \\ 0 \text{ else.} \end{cases}$$

Proof. If χ is constant, then we have

$$\sum_{g\in G} \chi(g) = \sum_{g\in G} 1 = n.$$

Let χ not be constant so that there is a $h \in G$ with $\chi(h) \neq 1$. Then we have

$$(1-\chi(h))\sum_{g\in G}\chi(g) = \sum_{g\in G}(\chi(g) - \chi(h)\chi(g)) = \sum_{g\in G}\chi(g) - \sum_{g\in G}\chi(hg). \quad (9.7)$$

The map $G \to G$, $g \mapsto hg$ is a bijection. This implies that if g runs through all the elements of G, then also $g \cdot h$ runs through all the elements of G. Hence (9.7) implies that

$$(1-\chi(h))\sum_{g\in G}\chi(g) = 0.$$

Since $1 - \chi(h) \neq 0$ the assertion $\sum_{g\in G}\chi(g) = 0$ is proved. □

It remains to see that the Kronecker symbol always defines a non-trivial character on $(\mathbb{Z}/|\Delta|\mathbb{Z})^*$.

Lemma 9.3.9. *Let Δ be a discriminant. Then there exists $n \in \mathbb{Z}$ such that* $\left(\frac{\Delta}{n}\right) = -1$.

Proof. Without loss of generality, we may assume that Δ is a fundamental discriminant. First, suppose that Δ is a power of 2. The only such fundamental discriminants are $\Delta = -4$, 8 and -8. From Theorem 3.4.13 we obtain

$$\left(\frac{-4}{3}\right) = -1, \quad \left(\frac{8}{5}\right) = -1, \quad \left(\frac{-8}{5}\right) = -1.$$

Now let p be an odd prime divisor of Δ and let x be a quadratic nonresidue modulo p. Choose n such that

$$n \equiv x \bmod p, \quad n \equiv 1 \bmod |\Delta/p|, \quad n \equiv 1 \bmod 4.$$

Then we obtain from Theorem 3.4.13

$$\left(\frac{\Delta}{n}\right) = \left(\frac{p}{n}\right)\left(\frac{\Delta/p}{n}\right) = \left(\frac{n}{p}\right) = -1. \qquad \square$$

The class number formula involves the *Dirichlet structure constant* κ_Δ which is defined as follows.

Let $\Delta < 0$. Let w be the number of roots of unity in $\mathcal{O}_\Delta$. In general, we have $w = 2$. Only if $\Delta = -4$ we have $w = 4$ and if $\Delta = -3$ we have $w = 6$. Then

$$\kappa_\Delta = \frac{2\pi}{w\sqrt{|\Delta|}} \, .$$

Let $\Delta > 0$. Let R_Δ be the regulator of $\mathcal{O}_\Delta$. Then

$$\kappa_\Delta = \frac{2R_\Delta}{\sqrt{\Delta}} \, .$$

Now let Δ be again an arbitrary discriminant. If we denote by h_Δ the class number of $\mathcal{O}_\Delta$ then analytic class number formula can be stated as follows.

Theorem 9.3.10 (Class number formula). $\kappa_\Delta h_\Delta = L(1, \chi_\Delta)$.

For a proof see Part IV of [Lan66] where this is Theorem 209. This formula can, for example, be used to compute the class number if one determines the values of κ_Δ and $L(1, \chi_\Delta)$ to sufficient precision. This will be explained below.

In the sequel, we will, however, frequently need only an estimate for the size of the class group Cl_Δ. Such estimate can be found by bounding the size of the value of $L(1, \chi_\Delta)$.

Let

$$c_1(\Delta) = \begin{cases} (\frac{1}{2}\log\Delta + 1) & \text{if } \Delta > 0 \\ (\frac{1}{2}\log|\Delta| + \log\log|\Delta| + 1) & \text{if } \Delta < 0. \end{cases} \tag{9.9}$$

We quote the following result from [Hua42].

Theorem 9.3.11.
1. If $\Delta < 0$, then $h_\Delta \le c_1(\Delta)\sqrt{|\Delta|}$.
2. If $\Delta > 0$, then $h_\Delta R_\Delta \le c_1(\Delta)\sqrt{\Delta}$ and $R_\Delta > .48$.

Lower bounds for h_Δ, or $h_\Delta R_\Delta$, respectively, were proven by, among others, Carl Ludwig Siegel (asymptotics only) and Titao Takuzawa (effectively computable bounds). These bounds can be used to establish lists of quadratic orders with small class number and short period length, see e.g. the work or Richard Mollin and Hugh Williams [MW92].

From Theorem 9.3.11, we see that h_Δ grows at most only slightly faster than $\sqrt{|\Delta|}$. The following estimate is useful for small $|\Delta|$.

Corollary 9.3.12. *$h_\Delta < |\Delta|/4$.*

Proof. For $\Delta < -21$ or $\Delta > 74$ this follows directly from Proposition 9.3.11. Note that $R_\Delta > \log(1 + \sqrt{\Delta}/2)$ by Definitions 8.3.6 and 6.12.6. For the remaining discriminants the inequality can be directly verified. □

We will now show how to approximate $L(1, \chi_\Delta)$ to within a factor of 2. The bounds we will give here were proved by Bach in [Bac95] Theorem 9.1. They are based on the hypothesis that all zeros s of the Dirichlet L-series $L(s, \chi_\Delta)$ with $0 < \Re s < 1$ have $\Re s = 1/2$. This is called the *Extended Riemann Hypothesis (ERH)*. A looser unconditional bound can be found in many textbooks on analytical number theory, e.g. in [Lan66].

Bach's result is motivated by another representation for $L(1, \chi_\Delta)$, the so called *Euler product.*

A function $f : \mathbb{N} \to \mathbb{C}$ is called *multiplicative* if $f(mn) = f(m)f(n)$ for any two positive integers m, n with $\gcd(m, n) = 1$. The function f is called *completely multiplicative* if $f(mn) = f(m)f(n)$ for any two positive integer m, n.

Theorem 9.3.13. *Let f be a multiplicative function such that the series $\sum_{n=1}^{\infty} f(n)$ is absolutely convergent. Then the sum of the series can be expressed as an absolutely convergent infinite product*

$$\sum_{n=1}^{\infty} f(n) = \prod_p \{1 + f(p) + f(p^2) + \ldots\}$$

extended over all primes. If f is completely multiplicative the product simplifies and we have

$$\sum_{n=1}^{\infty} f(n) = \prod_p \frac{1}{1 - f(p)}.$$

Proof. A proof can be found in [Apo86] where it is Theorem 11.6.

In each case of Theorem 9.3.13 the product is called the *Euler product* of the series.

The character χ_Δ is a completely multiplicative function. The Dirichlet series (9.6) is absolutely convergent if $\Re s > 1$. This suggests to approximate $L(1, \chi_\Delta)$ by truncated Euler products.

For positive real numbers x set

$$B(x,\chi) = \prod_{p<x} \left(1 - \frac{\chi(p)}{p}\right)^{-1}. \tag{9.10}$$

where the product is taken over all prime numbers $< x$.

Let n be an integer, $n \geq 2$. We introduce weights a_i, defined by

$$a_i(n) = \frac{(n+i)\log(n+i)}{\sum_{j=0}^{n-1}(n+j)\log(n+j)}, 0 \leq i \leq n-1. \tag{9.11}$$

The following theorem which is true under the assumption of the Extended Riemann Hypothesis describes the precision with which the value of the L-series at 1 can be expressed as a weighted sum of the truncated Euler products $B(x,\chi)$.

Theorem 9.3.14. *(ERH) Let n_0, A, B be any triplet from Table 9.1. Then for $n \geq n_0$ we have*

$$|\log L(1,\chi) - \sum_{i=0}^{n-1} a_i(n)\log B(n+i,\chi)| \leq \frac{A\log\Delta + B}{\sqrt{n}\log n}.$$

n_0	A	B
5	16.397	47.183
10	12.170	38.831
50	8.628	29.587
100	7.962	27.145
500	7.106	22.845
1000	6.897	21.528
5000	6.593	19.321
10000	6.510	18.606
50000	6.378	17.397
100000	6.338	17.031
500000	6.269	16.409
1000000	6.246	16.217

Table 9.1. Bach constants

We will now explain how Theorem 9.3.14 can be used to compute an approximation of the class number $h(\Delta)$ when $\Delta < 0$ and of $R_\Delta h(\Delta)$ when $\Delta > 0$.

We need the following auxiliary result.

Lemma 9.3.15. *Let $y \in \mathbb{R}$ with $0 < y < 1$. Then*

$$\max\left(-x, \frac{|x|}{1+y}\right) \le |\log(x+1)| \le \frac{|x|}{1-y}$$

for $x \in \mathbb{R}$ with $|x| \le y$.

Proof. First, consider the function

$$f(x) = \log(1+x) - \frac{x}{1-y}.$$

Its derivative is

$$f'(x) = \frac{1}{1+x} - \frac{1}{1-y}.$$

If $|x| \le y$ then $f'(x) \le 0$. Since $f(0) = 0$ it follows that $f(x) \le 0$ for $0 \le x \le y$ and $f(x) \ge 0$ for $-y \le x \le 0$. Hence

$$|\log(1+x)| = \log(1+x) \le \frac{x}{1-y} = \frac{|x|}{1-y}, \quad 0 \le x \le y$$

and

$$|\log(1+x)| = -\log(1+x) \le \frac{-x}{1-y} = \frac{|x|}{1-y}, \quad -y \le x \le 0.$$

Next, assume $0 \le x \le y$t and consider the function

$$f(x) = \log(1+x) - \frac{x}{1+y}.$$

Its derivative is

$$f'(x) = \frac{1}{1+x} - \frac{1}{1+y}$$

and we have $f'(x) \ge 0$. Since $f(0) = 0$ it follows that $f(x) \ge 0$. Hence

$$|\log(1+x)| = \log(1+x) \ge \frac{x}{1+y} = \frac{|x|}{1+y}, \quad 0 \le x \le y$$

Finally, assume $-1 \le -x < 0$ and consider the function $f(x) = \log(1+x) - x$. Its derivative is

$$f'(x) = \frac{1}{1+x} - 1$$

and we have $f'(x) > 0$ for $-1 < x < 0$. Since $f(0) = 0$ it follows that $f(x) \le 0$. Hence

$$-x \le -\log(1+x) = |\log(1+x)|.$$ □

For a positive integer n set

$$\ell(n, \Delta) = \exp\Bigl(\sum_{i=0}^{n-1} a_i(n) \log B(n+i, \chi_\Delta)\Bigr)\ . \tag{9.12}$$

Also fix any triplet n_0, A, B from Table 9.1 and set

$$C(n) = \frac{A \log|\Delta| + B}{\sqrt{n} \log n} .$$

Then we obtain from Theorem 9.3.14

$$|\log(L(1,\chi_\Delta)/\ell(n,\Delta))| \leq C(n), \quad n \geq n_0. \tag{9.13}$$

Theorem 9.3.16. *If* $C(n) < \log 2$ *then*

$$\left| \frac{L(1,\chi_\Delta)}{\ell(n,\Delta)} - 1 \right| < \frac{C(n)}{1 - C(n)}.$$

This theorem shows that for n in $O(\log|\Delta|)^2$, and $\tilde{h} = \ell(n,\Delta)/\kappa_\Delta$, the quotient $h_\Delta/\tilde{h}$, or respectively, $h_\Delta R_\Delta/\tilde{h}$ is close to 1.

Proof. Fix $n \geq 0$ and set

$$x = L(1,\chi_\Delta)/\ell(n,\Delta) - 1.$$

Then $|\log(1+x)| \leq C(n) < \log 2$. Hence $1/2 < 1 + x < 2$ and this implies $|x| < 1$. Therefore, Lemma 9.3.15 with $y = |x|$ and (9.13) yield

$$\frac{|x|}{1+|x|} < |\log(1+x)| < C(n).$$

An easy computation implies the assertion. □

9.4 Computing in finite Abelian groups

The class group Cl_Δ is a finite Abelian group. In this section we describe some computational problems concerning finite Abelian groups and we present algorithms for solving those problems. Those generic algorithms can be applied to class groups.

Let G be a finite Abelian group, written multiplicatively with neutral element 1. We assume that the following operations are efficiently computable.

- For $a, b \in G$ we can compute $c = a * b$.
- For $a \in G$ we can compute the inverse a^{-1}.
- For $a, b \in G$ we can decide whether $a = b$.

As we have seen in Sections 9.3.2 and 9.3.3, those assumptions are satisfied for class groups. We call these the *group operations*. Note that from every group element a we can determine the neutral element $1 = a * a^{-1}$.

We introduce a few basic notions. Let $M = (g_1, \ldots, g_k)$ be a sequence of elements of G. If $\mathbf{v} \in \mathbb{Z}^k$, $\mathbf{v} = (v_i)_{1 \leq i \leq k}$, then we set

$$M^{\mathbf{v}} = \prod_{i=1}^{k} g_i^{v_i} . \tag{9.14}$$

The *subgroup of G generated by M* is

$$\langle M \rangle = \left\{ M^{\mathbf{v}} : \mathbf{v} \in \mathbb{Z}^k \right\} . \tag{9.15}$$

With respect to inclusion, it is the smallest subgroup of G that contains M. If M has only one element, that is, $M = \{g\}$, then we write $\langle M \rangle = \langle g \rangle$. If $\langle M \rangle = G$, then M is called a *generating system* for G. We also say that G is generated by M. The group G is called *cyclic* if it is generated by one element. Such an element is called a *generator* of G. Finally, if $T \in \mathbb{Z}^{(k,l)}$ with column vectors $\mathbf{t}_1, \dots, \mathbf{t}_l$, then we set

$$M^T = \left(M^{\mathbf{t}_1}, \dots, M^{\mathbf{t}_l} \right) . \tag{9.16}$$

Example 9.4.1. Let $G = \mathrm{Cl}_{-31}$. By Example 9.3.5 the group G is cyclic of order 3 with generator $[2, 1, 4]$. A generating system for G is

$$M = ([2, 1, 4], [2, -1, 4])$$

and we have

$$M^{(2,-1)} = [2, 1, 4]^2 [2, -1, 4]^{-1} = [2, -1, 4][2, 1, 4] = [1, 1, 8] .$$

Also, we have

$$M^{\begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}} = (M^{(1,0)}, M^{(2,-1)}) = \left([2, 1, 4], [1, 1, 8]\right) .$$

9.4.1 Basic problems

We describe a few important computational problems for G.

Problem 9.4.2 (Root problem (RP)). Given $g \in G$ and a positive integer e,

1. decide whether g is the eth power of an element in G, and
2. if g is the eth power of an element in G, find an *eth root* of g; that is, find $h \in G$ with $h^e = g$.

Problem 9.4.3 (Order problem (OP)). Given $g \in G$ compute $\operatorname{order} g$, the *order* of g in G, that is, the least positive integer x such that $g^x = 1$.

Problem 9.4.4 (Discrete logarithm problem (DLP)). Given $g, h \in G$,

1. decide whether h belongs to the cyclic subgroup $\langle g \rangle$ of G generated by g, and
2. if $h \in \langle g \rangle$, find $\log_g h$, the *discrete logarithm* of h to the base g, that is, the least non-negative integer x such that $g^x = h$.

9.4.2 Structure

In this section we present the problem of finding the structure of G. That problem contains the order problem and the discrete logarithm problem as special cases. We start by explaining what we mean by the the structure of G. That notion is based on the following theorem.

Theorem 9.4.5. *There are positive integers $s_1, \dots, s_k$ such that $s_k > 1$, s_i divides s_{i-1}, $1 < i \leq k$, and there are cyclic subgroups G_i of order s_i, $1 \leq i \leq k$ such that*

$$G = G_1 \times G_2 \times \cdots \times G_k \ . \tag{9.17}$$

The sequence $(s_i)_{1 \leq i \leq k}$ is uniquely determined.

We will prove Theorem 9.4.5 in Section 9.7.

Definition 9.4.6. *The sequence of invariants of G is $(s_1, \dots, s_k)$ from Theorem 9.4.5. Each s_i, $1 \leq i \leq k$, is called an* invariant *of G.*

Example 9.4.7. Consider the group $G = \mathrm{Cl}_{-31}$. By Example 9.3.5 we have

$$G = \langle [2, 1, 4] \rangle \ .$$

Hence, there is only one invariant of G and that invariant is 3.

Corollary 9.4.8. *There are prime numbers $p_1, \dots, p_m$, positive integers $e_1, \dots, e_m$, and cyclic subgroups H_i of G of order $p_i^{e_i}$, $1 \leq i \leq m$, such that*

$$G = H_1 \times H_2 \times \cdots \times H_m \ . \tag{9.18}$$

Up to permutation, the sequence $\big((p_i, e_i)\big)$ is uniquely determined.

Proof. Exercise 9.8.4. □

We formulate the structure problem.

Problem 9.4.9 (Structure problem (SP)). Given a generating system for G, compute the sequence $(s_1, \dots, s_k)$ of invariants of G and elements $g_1, \dots, g_k$ such that the subgroup generated by g_i has order s_i for $1 \leq i \leq k$ and $G = \langle g_1 \rangle \times \cdots \times \langle g_k \rangle$.

9.4.3 Connections between the problems

In this section we explain the connection between the root problem, the order problem, the discrete logarithm problem, and the structure problem.

We first discuss a special case of the root problem.

Proposition 9.4.10. *Let $g \in G$, let n be a positive integer multiple of the order of g and let e, f be integers with $ef \equiv 1 \pmod{n}$. Then g^f is an eth root of g in G.*

Proof. Since $ef \equiv 1 \pmod{n}$ we can write $ef = 1 + kn$ with an integer k. Hence

$$(g^f)^e = g^{fe} = g^{1+kn} = g(g^n)^k = g\ .$$

□

Proposition 9.4.10 implies the following. If a small positive integer multiple n of the order of G is known, and if e is an integer that is coprime to n and if $g \in G$, then an eth root of g can be computed using one gcd-computation and $\mathrm{O}(\log |G|)$ group operations. As we have seen in Section 3.4.4, the problem of extracting eth roots in G is much more complicated if e is not coprime to the order of G.

We explain how the order problem can be solved efficiently using a discrete logarithm algorithm.

Proposition 9.4.11. *Let $g \in G$. If l is the discrete logarithm of g^{-1}, then the order of g is $l + 1$.*

Proof. By definition, we have $l = \min\{k \geq 0 : g^k = g^{-1}\}$. It follows that $l = \min\{k \geq 0 : g^{k+1} = 1\}$. This implies the assertion. □

We note that the algorithm from Section 9.7 that solves the structure problem also solves the general root problem, the order problem, and the discrete logarithm problem.

9.5 Generating systems

In this section we present two generating systems for the class group Cl_Δ. The first is big but it can be fully proved to be a generating system for Cl_Δ. The second is only known to be a generating system for Cl_Δ if the truth of an extended Riemann hypothesis is assumed.

We set

$$c_2(\Delta) = \begin{cases} \lfloor\sqrt{|\Delta|/3}\rfloor & \text{for } \Delta < 0\ , \\ \lfloor\sqrt{\Delta}/2\rfloor & \text{for } \Delta > 0\ . \end{cases}$$

We present a first generating system for the class group.

Proposition 9.5.1. *The class group Cl_Δ is generated by the set*

$$\{\,[\mathfrak{a}] \mid \mathfrak{a} \textit{ is an integral primitive } \mathcal{O}\textit{-ideal with } \mathrm{N}(\mathfrak{a}) < c_2(\Delta)\,\}\ .$$

Proof. By Corollary 5.3.9 and Corollary 6.5.4, each $\mathcal{O}$-ideal class contains a reduced ideal. If $\Delta < 0$, then by Lemma 5.4.1 the norm of that ideal is bounded by $c_2(\Delta)$. If $\Delta > 0$ and if the class contains the reduced ideal $L(a, b, c)$, then it also contains the reduced ideal $L(-c, b, -a)$. By Lemma 6.2.7 the norm of one of those ideals is bounded by $c_2(\Delta)$. So each ideal class contains an integral primitive $\mathcal{O}$-ideal of norm $\leq c_2(\Delta)$. □

For a positive real number c we let

$$\mathbb{P}(\Delta, c) = \left\{ p \in \mathbb{P} : \left(\frac{\Delta}{p}\right) \neq -1, p \leq c, p \nmid f(\Delta) \right\} \tag{9.19}$$

and

$$\mathcal{P}(\Delta, c) = \{\mathfrak{p}(\Delta, p) : p \in \mathbb{P}(\Delta, c)\} \ .$$

Now assume that Δ is a fundamental discriminant. For this case, we present a smaller generating system.

Proposition 9.5.2. *The set* $\big\{[\mathfrak{p}] : \mathfrak{p} \in \mathcal{P}\big(\Delta, c_2(\Delta)\big)\big\}$ *generates the class group* Cl_Δ*.*

Proof. Since Δ is a fundamental discriminant it follows from Theorem 8.6.8 that each $\mathcal{O}$-ideal of norm $\leq c_2(\Delta)$ is a product of prime ideals of $\mathcal{O}$ whose norm is in $\mathbb{P}\big(\Delta, c_2(\Delta)\big)$. For any $p \in \mathbb{P}\big(\Delta, c_2(\Delta)\big)$ there are at most two prime ideals of norm p. They are $\mathfrak{p}(\Delta, p)$ and $\sigma(\mathfrak{p}(\Delta, p))$. Also, $\sigma(\mathfrak{p}(\Delta, c))$ is equivalent to $\mathfrak{p}(\Delta, p)^{-1}$. Proposition 9.5.1 implies that any ideal class is a power product of ideal classes of prime ideals in $\mathcal{P}\big(\Delta, c_2(\Delta)\big)$. □

Again, let Δ be a fundamental discriminant. We set

$$c_3(\Delta) = \begin{cases} 6(\log|\Delta|)^2 & \text{for } \Delta < 0\ , \\ 12(\log\Delta)^2 & \text{for } \Delta > 0\ . \end{cases} \tag{9.20}$$

In [Bac90] the following is shown under the assumption of the Extended Riemann Hypothesis (GRH).

Proposition 9.5.3. *(ERH) The set* $\big\{[\mathfrak{p}] : \mathfrak{p} \in \mathcal{P}\big(\Delta, c_3(\Delta)\big)\big\}$ *generates the class group* Cl_Δ*.*

9.6 Computing a generating system in time $|\Delta|^{1/2+o(1)}$

Let Δ be a fundamental discriminant. The smallest fully proved generating system that was presented in Section 9.5 has approximately $|\Delta|^{1/2}$ elements. The computation of each generator requires extracting a square root modulo a prime number. If the fastest deterministic algorithm for extracting square roots modulo primes is used, then the determination of the generating system requires approximately time $|\Delta|^{3/4}$. In this section we describe Algorithm `GeneratingSystem` that computes a generating system for the class group Cl_Δ and the class number h_Δ in time $|\Delta|^{1/2+o(1)}$. It avoids the computation of all generators. The algorithm is based on an idea of H.W. Lenstra Jr. It is the fastest known deterministic class number algorithm.

9.6.1 The idea

Algorithm `GeneratingSystem` proceeds iteratively. In each iteration the algorithm knows a sequence M of elements in Cl_Δ, all elements of the subgroup

$$\mathrm{Cl} = \langle M \rangle$$

of Cl_Δ, and the set

$$P = \{p : p \in \mathbb{P}\big(\Delta, c_2(\Delta)\big), [\mathfrak{p}(\Delta, p)] \notin \mathrm{Cl}\} .$$

The set P contains the norm of all elements of the generating system from Proposition 9.5.2 that do not belong to the subgroup Cl of Cl_Δ found so far.

Algorithm `GeneratingSystem` selects a prime number p in P and computes $\mathfrak{p} = \mathfrak{p}(\Delta, p)$ from (8.16). Algorithm `GeneratingSystem` replaces M by $M \circ ([\mathfrak{p}])$ and Cl by $\langle M \rangle = \langle \mathrm{Cl} \cup \{[\mathfrak{p}]\} \rangle$. Then it deletes all primes q in P such that $[\mathfrak{p}(\Delta, q)] \in \mathrm{Cl}$. We will show that it is possible to decide efficiently whether $[\mathfrak{p}(\Delta, q)] \in \mathrm{Cl}$ without actually computing $\mathfrak{p}(\Delta, q)$.

If P is empty, then $\mathrm{Cl} = \mathrm{Cl}_\Delta$ and M is a generating system for Cl_Δ. Otherwise, `GeneratingSystem` starts a new iteration. Note that the algorithm also calculates the class number $h(\Delta) = |\mathrm{Cl}|$.

The advantage of this approach is the following. In each iteration exactly one prime ideal is computed and the subgroup Cl grows by a factor at least two. Therefore, at most $\lceil \log_2 |\mathrm{Cl}_\Delta| \rceil$ iterations are necessary to compute Cl_Δ and the generating system. By Theorem 9.3.11, the size of the generating system is $\mathrm{O}(\log|\Delta|)$ and the number of prime ideals that are actually computed is also $\mathrm{O}(\log|\Delta|)$.

9.6.2 Updating Cl, S, and P

We explain how Algorithm `GeneratingSystem` represents, initializes, and updates M, Cl, and P.

Algorithm `GeneratingSystem` stores the ideal classes in M and Cl in terms of reduced representatives. In order to be able to decide whether an ideal class represented by a reduced ideal is contained in Cl, Algorithm `GeneratingSystem` also uses the set S of all reduced ideals of the classes in Cl. For $\Delta < 0$, the sets Cl and S are basically the same since the ideal classes are stored in terms of their uniquely determined reduced representatives. For $\Delta > 0$ this is not true since each ideal class contains a cycle of reduced ideals, cf. Section 10.1.2.

Initially, `GeneratingSystem` sets

$$M = () , \quad \mathrm{Cl} = \{[\mathcal{O}]\}, \quad S = \{\mathcal{O}\}$$

and

$$P = \mathbb{P}\big(\Delta, c_2(\Delta)\big) .$$

In each iteration, `GeneratingSystem` selects the smallest prime number in P, and computes the prime ideal $\mathfrak{p} = \mathfrak{p}(\Delta, p)$.

To update Cl, `GeneratingSystem` finds the order e of the coset $[\mathfrak{p}] \cdot \mathrm{Cl}$ in the quotient group $\mathrm{Cl}_\Delta/\mathrm{Cl}$. To find this order, Algorithm `GeneratingSystem` computes the first positive integer e such that a reduced representative of $[\mathfrak{p}]^e$ is in S. When e is found, Algorithm `GeneratingSystem` replaces Cl by

$$\mathrm{Cl} \cup_{1 \le x < e} [\mathfrak{p}]^x \mathrm{Cl} \; .$$

Also, Algorithm `GeneratingSystem` updates S by computing

$$S \cup \{\mathfrak{a} : \mathfrak{a} \text{ reduced ideal in } [\mathfrak{p}]^x C, 1 \le x < e, C \in \mathrm{Cl}\} \; .$$

Now we explain how P is updated.

Let $\Delta > 0$. Suppose that the reduced ideal $\mathfrak{a}$ is inserted into S. If $\mathfrak{a} = \mathfrak{p}(\Delta, p)$ for a prime number $p \in P$, then p is removed from P. Except for this modification, P remains unchanged.

This is justified as follows. Let $p \in P$. Then $p \le \sqrt{\Delta}/2$. Hence, by Lemma 6.5.1 the prime ideal $\mathfrak{p}(\Delta, p)$ is reduced. This means that the ideal class $[\mathfrak{p}(\Delta, p)]$ is in Cl if and only if the ideal $\mathfrak{p}(\Delta, p)$ is in S.

Let $\Delta < 0$. Then the update procedure for P is slightly more complicated since the ideals $\mathfrak{p}(\Delta, p)$ with $p \in P$ are not necessarily reduced. Algorithm `GeneratingSystem` uses the following result.

Lemma 9.6.1. *Let $\Delta < 0$ and let $p \in \mathbb{P}(\Delta, c_2(\Delta))$. Then the class $[\mathfrak{p}(\Delta, p)]$ belongs to Cl if and only S contains $\mathfrak{p}(\Delta, p)$ or $L(a, b, p)$ with $b \le 0$ or $L(a, b, c)$ with $p = a - b + c$ and $0 \le 2a - b \le p$.*

Proof. Set $\mathfrak{p} = \mathfrak{p}(p, \Delta)$.

Assume that $[\mathfrak{p}]$ belongs to Cl. We show that the reduced ideal in the equivalence class of $\mathfrak{p}$ has the asserted form.

If $\mathfrak{p}$ is reduced, then $\mathfrak{p}$ is in S.

Assume that $\mathfrak{p}$ is not reduced. Then $\rho(\mathfrak{p})$ is reduced by Lemma 5.5.3 since $p < \sqrt{|D|/3} < \sqrt{|\Delta|}$. Let

$$f = (p, B, C) = \big(p, b(p, \Delta), c(p, \Delta)\big), \quad s = s(f) \; .$$

It follows from the proof of Lemma 5.5.3 that $|s| \le 1$. Since $B \ge 0$ by the definition of $b(p, \Delta)$, it follows that $s \in \{0, 1\}$. Hence, $\rho(f) = (C, -B, p)$ or $\rho(f) = (C, -B + 2C, p - B + C)$. It follows that the reduced ideal in the class of $\mathfrak{p}(\Delta, p)$ is of the form (a, b, p) with $b \le 0$ or (a, b, c) with $p = a - b + c$ and $B = 2a - b$. So $0 \le 2a - b \le p$.

Now assume that S contains an ideal as described in the lemma. If that ideal is $\mathfrak{p}(\Delta, p)$, then $[\mathfrak{p}(\Delta, p)]$ is in Cl.

Assume that $L(a, b, p)$ with $b \le 0$ is in S. Then b is a square root of Δ mod $4p$ and $0 \le -b < a \le p$. Hence $\mathfrak{p}(\Delta, p) = L(p, -b, a)$. Since $L(a, b, p)$ and $L(p, -b, a)$ are equivalent, it follows that $[\mathfrak{p}(\Delta, p)]$ is in Cl.

Assume that $L(a, b, c)$ with $p = a - b + c$ and $0 \le 2a - b \le b$ is in S . We know that $L(a, b, c)$ is equivalent to $L(a - b + c, 2a - b, a) = L(p, 2a - b, a)$. Since $0 \le 2a - b \le p$, it follows that $L(p, 2a - b, a) = \mathfrak{p}(\Delta, p)$. So $[\mathfrak{p}(\Delta, p)]$ is in Cl. □

Let $\Delta < 0$. Using Lemma 9.6.1 Algorithm `GeneratingSystem` updates P as follows. Suppose that the reduced ideal $\mathfrak{a}$ is inserted into S. If there exists $p \in P$ with $\mathfrak{a} = \mathfrak{p}(\Delta, p)$ or $\mathfrak{a} = L(a, b, p)$ with $b \le 0$ or $\mathfrak{a} = L(a, b, c)$ with $p = a - b + c$ and $0 \le 2a - b \le p$ then Algorithm `GeneratingSystem` removes p from P.

Here is the algorithm.

Algorithm 9.1 `GeneratingSystem` (Δ)

Input: A quadratic discriminant Δ.

Output: A generating system M for Cl_Δ and the class number $h(\Delta)$.

```
$M \leftarrow ()$, $\mathrm{Cl} = S \leftarrow \emptyset$
$P \leftarrow \mathbb{P}(\Delta, c_2(\Delta))$
while $P \ne \emptyset$ do
   Select a prime number $p \in P$
   $\mathfrak{p} \leftarrow \mathfrak{p}(\Delta, p)$
   $M \leftarrow M \circ ([\mathfrak{p}])$
   $e \leftarrow 0$, $C \leftarrow [\mathcal{O}_\Delta]$
   repeat
      $e \leftarrow e + 1$
      $\mathfrak{a} \leftarrow$ reduced ideal in $C \leftarrow C \cdot [\mathfrak{p}]$
   until $\mathfrak{a} \in S$
   $\mathrm{Cl} \leftarrow \mathrm{Cl} \cup \{[\mathfrak{p}]^x C : 1 \le x < e, C \in \mathrm{Cl}\}$
   $S \leftarrow S \cup \{\mathfrak{a} : \mathfrak{a}$ reduced ideal in $[\mathfrak{p}]^x C, 1 \le x < e, C \in \mathrm{Cl}\}$
   Update $P$ as described above
Return $M$ and $h(\Delta) \leftarrow |\mathrm{Cl}|$
```

9.6.3 Examples

Example 9.6.2. Let $\Delta = -227$. Then $-\Delta$ is a prime number. Hence, Δ is a fundamental discriminant. Since $c_2(\Delta) = \lfloor\sqrt{|\Delta|/3}\rfloor = 8$ and $\left(\frac{2}{\Delta}\right) = -1$, $\left(\frac{3}{\Delta}\right) = 1$, $\left(\frac{5}{\Delta}\right) = -1$, and $\left(\frac{7}{\Delta}\right) = 1$, we initially have

$$P = \mathbb{P}(\Delta, c_2(\Delta)) = \{3, 7\} ,$$

$$M = () , \mathrm{Cl} = \{[\mathcal{O}]\}, \quad S = \{\mathcal{O}\} .$$

`GeneratingSystem` calculates $\mathfrak{p}(\Delta, 3) = L(3, 1, 19)$ and determines e. We have $[3, 1, 19]^2 = [7, 5, 9]$, $[3, 1, 19]^3 = [7, -5, 9]$, $[3, -1, 19]^4 = [3, -1, 19]$, and $[3, 1, 19]^5 = [1, 1, 57]$. Hence $e = 5$,

$$\mathrm{Cl} = \Big\{[1, 1, 57], [3, 1, 19], [7, 5, 9], (2), [7, -5, 9], [3, -1, 19]\Big\} .$$

It follows that the new S is

$$S = \Big\{(1, 1, 57), (3, 1, 19), (7, 5, 9), (7, -5, 9), (3, -1, 19)\Big\} ,$$

So GeneratingSystem removes 3 and 7 from P. Then P is empty and GeneratingSystem has found that the class group Cl_{-227} is generated by $[3,1,19]$ and has order 5. In fact, we have even found that Cl_{-227} is cyclic.

Example 9.6.3. Let $\Delta = 229$. Then Δ is a prime number. Hence, Δ is a fundamental discriminant. Since $c_2(\Delta) = \lfloor\sqrt{\Delta}/2\rfloor = 7$ and $\left(\frac{2}{\Delta}\right) = -1$, $\left(\frac{3}{\Delta}\right) = 1$, $\left(\frac{5}{\Delta}\right) = 1$, and $\left(\frac{7}{\Delta}\right) = -1$, we have

$$P = \mathbb{P}\big(\Delta, c_2(\Delta)\big) = \{3,5\} .$$

Also,

$$M = () , \quad \mathrm{Cl} = \{[\mathcal{O}]\}, \quad S = \{\mathcal{O}\} .$$

Algorithm GeneratingSystem calculates $\mathfrak{p}_1 = \mathfrak{p}(3,229) = L(3,1,-19)$ and determines e. We have $[3,1,-19] = [3,13,-5]$, $[3,1,-19]^{-2} = [9,-7,-5]$, $[3,1,-19]^{-3} = [1,15,-1]$. Hence, $e = 3$ and

$$\mathrm{Cl} = \Big\{[1,15,-1],[3,13,-5],[9,7,-5]\Big\} .$$

The cycle of $(1,15,-1)$ is $((1,15,-1))$, of $(3,13,-5)$ it is $\big((3,13,-5)\,(5,7,-9)\,(9,11,-3)\big)$, and of $(9,7,-5)$ it is $(9,7,-5)\,(5,13,-3)\,(3,11,-9)$. So we obtain

$$S = \Big\{(1,15,-1),(3,13,-5),(5,7,-9),(9,11,-3)),(9,7,-5),\\ (5,13,-3),(3,11,-9)\Big\} .$$

Since $(3,13,-5) \in S$ we delete 3 from P. Since $(5,7,-9)$ we also delete 5 from P. Then P is empty. Hence, Algorithm GeneratingSystem has proved that the class group $\mathrm{Cl}(229)$ is of order 3 and generated by $[3,13,-5]$. In fact, we have proved that $\mathrm{Cl}(229)$ is cyclic of order 3.

9.6.4 Analysis

In order to give a bound for the running time of GeneratingSystem, we first need an estimate for the length of the generating system that the algorithm computes.

Proposition 9.6.4. *The length of the generating system computed by Algorithm* GeneratingSystem *is in* $\mathrm{O}(\log|\Delta|)$.

Proof. Assume that the generating system computed by Algorithm GeneratingSystem is of length l. In each iteration the size of Cl is at least doubled. Algorithm GeneratingSystem terminates when the subgroup Cl is equal to Cl_Δ. Hence, $l \le \log_2 h_\Delta$ and the assertion follows from Proposition 9.3.11. □

Theorem 9.6.5. *Algorithm* `GeneratingSystem` *computes a generating system for and the order of the class group Cl_Δ in time $|\Delta|^{1/2+o(1)}$.*

The factor $|\Delta|^{o(1)}$ entering the time bound can in fact be explicitly given as a small constant times the product of $c_1(\Delta)$ and a term in $O(\log|\Delta|)^2$ depending on the arithmetic used for calculating and reducing products of ideals with norm smaller than $\sqrt{|\Delta|}$, cf. sections 5.6 and 6.6.2.

Proof. We use the notation from Sections 9.6.1 and 9.6.2.

We first analyze the computation of the generating system M. To compute an element of that system we extract a square root of Δ mod p for a prime number p with $p < \sqrt{|\Delta|}$. This can be done in time $|\Delta|^{1/2+o(1)}$. So by Proposition 9.6.4, the computation of the elements of the generating system takes time $|\Delta|^{1/2+o(1)}$.

Next we estimate the time for computing the sets Cl and S. The computation of an element of Cl requires one multiplication in the class group which takes time $|\Delta|^{o(1)}$. Using the bound for the number of elements in Cl which follows from Proposition 9.3.11 we see that the time for computing Cl is $|\Delta|^{1/2+o(1)}$. This implies that the computation of S for $\Delta < 0$ also takes time $|\Delta|^{1/2+o(1)}$. If $\Delta > 0$, then determining S requires computing the cycle in all ideal classes of $\mathcal{O}_\Delta$. One element of such a cycle can be computed in time $|\Delta|^{o(1)}$. By Corollary 10.1.7 the number of elements in one cycle is $O(R_\Delta)$. Hence, S contains $O(h_\Delta R_\Delta)$ elements. Using the upper bound for $h_\Delta R_\Delta$ from Proposition 9.3.11, we see that S is computed in time $|\Delta|^{1/2+o(1)}$. □

9.7 Computing the structure of a finite Abelian group

Let G be a finite Abelian group. We explain how to compute the structure of G from a generating system

$$M = (g_1, \dots, g_l)$$

of G.

9.7.1 The basic algorithm

Definition 9.7.1. *A* relation *for M is a vector $\mathbf{v} \in \mathbb{Z}^l$ such that $M^{\mathbf{v}} = 1$.*

Proposition 9.7.2. *The set $L(M)$ of all relations for M is an l-dimensional lattice in $\mathbb{Z}^l$. Its determinant is the order of G.*

Proof. The map

$$\mathbb{Z}^l \longrightarrow G\,, \quad \mathbf{z} \mapsto M^{\mathbf{z}} \;.$$

is a homomorphism of groups which is surjective since M is a generating system for G. The kernel of that map is $L(M)$. The homomorphism theorem implies that $\mathbb{Z}^l/L(M) \cong G$. Therefore, $L(M)$ is of finite index $|G|$ in $\mathbb{Z}^l$ which means that $L(M)$ is a lattice of dimension l. The index of $L(M)$ in $\mathbb{Z}^l$ is the determinant of $L(M)$. Hence, the the determinant of $L(M)$ is the order of G. □

The lattice $L(M)$ is called the *relation lattice* of M.

The algorithm for solving the structure problem is based on the following result which also proves Theorem 9.4.5.

Proposition 9.7.3. *Let $B \in \mathbb{Z}^{(l,l)}$. Assume that the columns of B form a basis of the relation lattice $L(M)$. Let $D \in \mathbb{Z}^{(l,l)}$ be the Smith normal form of B. Let $D = \operatorname{diag}(n_1, \dots, n_k, 1, \dots, 1)$ with $n_k > 1$. Let $U'D = BV$ with $U', V \in \mathrm{GL}(l, \mathbb{Z})$. Let $U \in \mathbb{Z}^{(l,l)}$ with $U \equiv U' \pmod{|G|}$. Let $M^U = (h_1, \dots, h_k, h_{k+1}, \dots, h_l)$. Then the following are true.*

1. The order of G is $|\det B|$.
2. The order of h_i is n_i, $1 \le i \le k$, $h_{k+1} = \ldots = h_l = 1$ and

$$G = \langle h_1 \rangle \times \cdots \times \langle h_k \rangle \,. \tag{9.21}$$

3. If the group G can be written as

$$G = G_1 \times \cdots \times G_m \tag{9.22}$$

where m is a non-negative integer, G_i are nontrivial cyclic subgroups of G, $1 \le i \le m$, with $|G_{i+1}| \mid |G_i|$, $1 \le i < m$, then $m = k$ and $|G_i| = n_i$, $1 \le i \le k$.

Proof. 1. This is a consequence of Proposition 9.7.2

2. and 3. We claim that M^U is a generating system for G. Clearly, we have $(M^U)^{\mathbf{v}} \in G$ for any $\mathbf{v} \in \mathbb{Z}^l$. Conversely, let $g \in G$. Then there is $\mathbf{v} \in \mathbb{Z}^l$ with $g = M^{\mathbf{v}}$. Since $\gcd(\det U, |G|) = 1$, it follows that there is $\tilde{U} \in \mathbb{Z}^{(l,l)}$ such that $U\tilde{U} \equiv I_l \pmod{|G|}$ where I_l is the $l \times l$-identity matrix. Set $\tilde{\mathbf{v}} = \tilde{U}\mathbf{v}$. Then

$$(M^U)^{\tilde{\mathbf{v}}} = (M^U)^{\tilde{U}\mathbf{v}} = M^{U\tilde{U}\mathbf{v}} = M^{\mathbf{v}} = g.$$

This proves our claim.

Next, we show that the columns of D form a basis for the relation lattice of M^U. Since $(M^U)^D = M^{UD} = M^{U'D} = M^{BV}$, it follows that the columns of D are relations for M^U. Let $\mathbf{v}$ be a relation for M^U. Then $1 = (M^U)^{\mathbf{v}} = M^{U'\mathbf{v}}$. It follows that $U'\mathbf{v}$ is a relation for M. Since BV is a basis for $L(M)$, there is $\mathbf{x} \in \mathbb{Z}^l$ with $U'\mathbf{v} = BV\mathbf{x} = U'D\mathbf{x}$. Hence, $\mathbf{v} = D\mathbf{x}$. This proves that D is a basis of $L(M^U)$. It follows that the ith diagonal element d_i of D is the order of h_i, $1 \le i \le l$. In particular, we have $h_{k+1} = \ldots = h_l = 1$. This implies that $(h_1, \dots, h_k)$ is a generating system for G.

Since D is a basis of $L(M^U)$, it follows that if $(h_1, \dots, h_k)^{\mathbf{e}} = 1$ for some $\mathbf{e} \in \mathbb{Z}^k$, then $e_i \equiv 0 \pmod{n_i}$ where e_i is the ith entry in $\mathbf{e}$, $1 \le i \le k$. This proves (9.21).

4. Assume that we have two representation $G = G_1 \times \cdots \times G_m = G'_1 \times \cdots \times G'_{m'}$ as in (9.22), $m' \geq m$. Let $j \in \{1, \dots, m\}$. We will show that $|G_i| = |G'_i|$, $1 \leq i < j$. Then

$$\prod_{i=1}^{j-1} \frac{|G_i|}{|G_j|} = |G^{|G_j|}| = \prod_{i=1}^{j-1} \frac{|G_i|}{|G_j|} \prod_{i=j}^{m'} \frac{|G'_i|}{\gcd(|G'_i|, |G_j|)} . \tag{9.23}$$

It follows that $|G_j| = \gcd(|G_j|, |G'_j|)$. Hence, $|G_j|$ divides $|G'_j|$. The same argument shows that $|G'_j|$ divides $|G_j|$. Hence, $|G_j| = |G'_j|$. Also, we see that $m = m'$. □

By Proposition 9.7.3, the structure of G can be computed as follows. We determine a basis B of the relation lattice $L(M)$. Then we use Algorithm `snfModular` to find the Smith normal form D of B and a matrix $U \in \mathbb{Z}^{(n,n)}$ with the properties from Proposition 9.7.3. The invariants and the representation of G as a product of cyclic groups whose orders are the invariants can be computed as described in Proposition 9.7.3.

We explain a simple algorithm for computing a basis of the relation lattice $L(M)$. We let G_i be the subgroup of G generated by the first i elements of the generating system M, $1 \leq i \leq l$; that is,

$$G_i = \langle (g_1, \dots, g_i) \rangle , \quad 1 \leq j \leq l . \tag{9.24}$$

Also, set

$$G_0 = \{1\} . \tag{9.25}$$

Proposition 9.7.4. *Let $B = (b_{ij}) \in \mathbb{Z}^{(l,l)}$ be a matrix in upper triangular form whose columns are relations for M and such that $b_{j,j}$ is the smallest positive integer e with $g_j^e \in G_{j-1}$ for $1 \leq j \leq l$. Then B is a basis of the relation lattice $L(M)$.*

Proof. Since B is non-singular by definition, it suffices to show that the columns $(\mathbf{b}_1, \dots, \mathbf{b}_l)$ of B form a generating system for $L(M)$. Let $\mathbf{v} = (v_1, \dots, v_l) \in L(M)$. We show by induction that there are integers $x_l, x_{l-1}, \dots, x_1$ such that for $j = l, l-1, \dots, 1, 0$ the last $l - j$ entries of $\mathbf{v} - \sum_{i=j+1}^{l} x_i \mathbf{b}_i$ are zero.

For $j = l$ the assertion is trivial. Assume that the assertion holds for some j, $1 \leq j \leq l$. Set $\mathbf{w} = \mathbf{v} - \sum_{i=j+1}^{l} x_i \mathbf{b}_i$. Then

$$\mathbf{w} = (w_1, \dots, w_j, 0, \dots, 0) .$$

We show that

$$w_j = x_j b_{j,j} , \quad x_j \in \mathbb{Z} . \tag{9.26}$$

Then the last $l - j + 1$ entries of $\mathbf{v} - \sum_{i=j}^{l} x_i \mathbf{b}_i$ are zero. Let x, y be integers with

$$\gcd(b_{j,j}, w_j) = x w_j + y b_{j,j} . \tag{9.27}$$

Then the jth entry of the vector $x\mathbf{w} + y\mathbf{b}_j$ is $\gcd(b_{j,j}, w_j)$ and the last $l - j$ entries of that vector are zero. Since $x\mathbf{w} + y\mathbf{b}_j$ is a relation for M, it follows that

$$g_j^{\gcd(b_{j,j},w_j)} \in G_{j-1} \,. \tag{9.28}$$

So $\gcd(b_{j,j}, w_j) = b_{j,j}$, since $b_{j,j}$ is the smallest positive integer e with $g_j^e \in G_{j-1}$. This implies (9.26). □

We explain how to compute a basis B as in Proposition 9.7.4. Let $j \in \{1, \ldots, k\}$. Assume that we know $\mathbf{b}_1, \ldots, \mathbf{b}_{j-1}$. To compute $\mathbf{b}_j$, we search for the smallest positive integer e such that there is $\mathbf{v} \in \mathbb{Z}^{j-1}$ with $(g_1, \ldots, g_{j-1})^{\mathbf{v}} g_j^e = 1$. Then $\mathbf{b}_j = \mathbf{v} \circ (e) \circ (0, \ldots, 0) \in \mathbb{Z}^l$. Note that e is the order of the coset $g_j G_{j-1}$ in the factor group G/G_{j-1}. We may choose $\mathbf{v} = (v_1, \ldots, v_{j-1})$ such that $0 \le v_i < b_{i,i}$, $1 \le i \le j - 1$. Then the basis computed in this way is in Hermite normal form.

Example 9.7.5. We determine the subgroup G of $\mathrm{Cl}(-1123)$ that is generated by the classes $[7, 5, 41]$ and $[17, 13, 19]$. We have $M = \big([7, 5, 41], [17, 13, 19]\big)$. The neutral element of G is $[1, 1, 281]$. We first determine the smallest positive integer e with $[7, 5, 41]^e = [1, 1, 281]$, that is, the order of $[7, 5, 41]$. We obtain $[7, 5, 41]^0 = [1, 1, 281]$, $[7, 5, 41]^1 = [7, 5, 41]$, $[7, 5, 41]^2 = [17, -13, 19]$, $[7, 5, 41]^3 = [17, 13, 19]$, $[7, 5, 41]^4 = [7, -5, 41]$, $[7, 5, 41]^5 = [1, 1, 281]$. The first element of the basis of $L(M)$ is $(5, 0)$. Next, we compute $[7, 5, 41][17, 13, 19] = [7, -5, 41]$ and $[7, 5, 41]^2[17, 13, 19] = [1, 1, 281]$. The second vector of the basis of $L(M)$ is $(2, 1)$. The matrix, whose columns are the basis vectors, is

$$B = \begin{pmatrix} 5 & 2 \\ 0 & 1 \end{pmatrix} .$$

It is immediately clear that $G = \langle M \rangle$ is cyclic of order 5. The only invariant of G is 5. The representation (9.21) of G is $G = \big\langle [7, 5, 41] \big\rangle$.

We analyze this simple algorithm.

Proposition 9.7.6. *Assume that a generating system $(g_1, \ldots, g_l)$ for G is known. Then the structure problem for G can be solved*

1. *using $\mathrm{O}(l|G|)$ group operations,*
2. *storing $\mathrm{O}(l)$ group elements, and*
3. *computing the Smith normal D form of an upper triangular matrix $B \in \mathbb{Z}^{(l,l)}$ whose entries are bounded by $|G|$ including a matrix $U \in \mathbb{Z}^{(l,l)}$ as in Proposition 9.7.3.*

Proof. For $j \in \{1, \ldots, l\}$ computing $\mathbf{b}_j$ requires $\mathrm{O}(\prod_{i=1}^{j} b_{i,i}) = \mathrm{O}(|G|)$ group operations, $1 \le j \le l$. To find those vectors, it suffices to store the generators and $\mathrm{O}(1)$ additional group elements. Once the basis B of $L(M)$ is known, its Smith normal form and the transformation $U \in \mathbb{Z}^{(l,l)}$ is computed. By construction, the entries of B are bounded by $|G|$. The computation of the generators of the cyclic factors requires $\mathrm{O}(l \log |G|)$ group operations (see Exercise 9.8.5). □

We explain how the structure algorithm presented in this section can be used to solve the order problem and the discrete logarithm problem.

We start with the order problem. Let $g \in G$. The order of g is the single entry in the HNF-basis of the relation lattice $L(g)$.

Next, we discuss the discrete logarithm problem. Let $g, h \in G$. To solve the discrete logarithm problem we determine the HNF-basis

$$\begin{pmatrix} b_{1,1} & b_{1,2} \\ 0 & b_{2,2} \end{pmatrix}$$

of the relation lattice $L(g, h)$. Then $h \in \langle g \rangle$ if and only if $b_{2,2} = 1$ and if $b_{2,2} = 1$, then $b_{1,2}$ is the discrete logarithm of h to the base g (see Exercise 9.8.6).

9.7.2 Terr's algorithm – computing orders

We explain a more efficient algorithm for computing the order of an element in G. The algorithm was invented by Terr [Ter00] based on earlier ideas of Shanks [Sha71]. The algorithm uses the following result.

Lemma 9.7.7. *Let $g \in G$. Then there is $e \in \mathbb{N}$ and $f \in \{0, \ldots, e-1\}$ with $g^{e(e+1)/2} = g^f$. If e is chosen minimal with this property, then $e(e-1)/2 < \operatorname{order}(g) \leq e(e+1)/2$ and $\operatorname{order}(g) = e(e+1)/2 - f$*

Proof. Let $e \in \mathbb{N}$ such that $e(e-1)/2 < \operatorname{order}(g) \leq e(e+1)/2$. Since $e(e-1)/2 + e = e(e+1)/2$, such an e exists. Let $f = e(e+1)/2 - \operatorname{order} g$. Then $f \in \{0, \ldots, e-1\}$. Also, since $g^{e(e+1)/2-f} = g^{\operatorname{order}(g)} = 1$, it follows that $g^{e(e+1)/2} = g^f$. This proves the existence of e and f.

We prove the minimality of e. Let $e' \in \mathbb{N}$, $f' \in \{0, \ldots, e'-1\}$ such that $g^{e'(e'+1)/2-f'} = 1$. Then $e'(e'+1) \geq e'(e'+1)/2 - f' \geq \operatorname{order}(g) = e(e+1)/2 - f > e(e-1)/2$. Since e and e' are integers, this implies $e'(e'-1)/2 \geq e(e-1)/2$. Hence, $e' \geq e$. □

For $e = 1, 2, \ldots$ Terr's algorithm computes the set

$$\mathsf{babySet} = \{(g^f, f) : 0 \leq f < e\} \tag{9.29}$$

and checks whether there exists a pair of the form $(g^{e(e+1)/2}, f)$ in babySet for some f. By Lemma 9.7.7 this will eventually happen. If this happens for the first time, then we have $\operatorname{order}(g) = e(e+1)/2 - f$. In the eth iteration of the algorithm we use

$$\mathsf{babyElement} = g^e, \quad \mathsf{giantElement} = g^{e(e+1)/2} .$$

Here is the algorithm.

Algorithm 9.2 `order` (g)

Input: A group element g.
Output: The order n of g.

```
babySet ← $\{(1,0)\}$.
$e \leftarrow 1$
babyElement ← $g$
giantElement ← $g$
loop
    if babySet contains a pair (giantElement, $f$) then return $n = e(e+1)/2 - f$
    insert (babyElement, $e$) into babySet
    babyElement ← $g \cdot$ babyElement
    $e \leftarrow e + 1$
    giantElement ← giantElement · babyElement
```

We analyze Terr's algorithm.

Theorem 9.7.8. *Let $g \in G$ and let $n = \operatorname{order}(g)$. Given g, algorithm* `order`(g) *terminates and returns n. Algorithm* `order`(g) *executes no more than $\sqrt{2n}+1/2$ iterations, $2\sqrt{2n}-1$ multiplications in G and $\sqrt{2n}+1/2$ table look-ups. Also, algorithm* `order`(g) *stores at most $\sqrt{2n}+1/2$ elements of G.*

Proof. It follows from Lemma 9.7.7 that `order` terminates and upon termination we have $e(e-1)/2 < n \leq e(e+1)/2$. Since e and n are integers we have $(e-1/2)^2 = e(e-1)+1/4 < 2n$ which implies $e < \sqrt{2n}+1/2$. In the first $e-1$ iterations of the while loop, 2 multiplications are executed. In the last iteration no multiplication is performed. Also, table babySet is accessed twice in each iteration, once to test whether $(\mathsf{giantElement}, f) \in \mathsf{babySet}$, and once to store the pair $(\mathsf{babyElement}, e)$ in babySet. Since the number of iterations is at most $\sqrt{2n}+1/2$, this implies the assertion. □

Here is an example.

Example 9.7.9. Let $\Delta = -227$ and $f = (3,1,19)$. Then $\Delta(f) = \Delta$. We determine the order of the equivalence class $C = [3,1,19]$ of f in the class group Cl_{-227}.

Initialization: We have $\mathsf{babySet}_1 = \{([1,1,57],0)\}$, and $\mathsf{babyElement}_1 = \mathsf{giantElement}_1 = [3,1,19]$.

$e = 1$: The set $\mathsf{babySet}_1$ does not contain a pair with first component $\mathsf{giantElement}_2$. $\mathsf{babySet}_2 = \{([1,1,57],0), ([3,1,19],1),\}$, $\mathsf{babyElement}_2 = [7,-5,9]$, $\mathsf{giantElement}_2 = [7,-5,9]$.

$e = 2$: The set $\mathsf{babySet}_2$ does not contain a pair with first component $\mathsf{giantElement}_2$. $\mathsf{babySet}_3 = \{([1,1,57],0), ([3,1,19],1), ([7,5,9],2)\}$, $\mathsf{babyElement}_2 = [7,-5,9]$, $\mathsf{giantElement}_3 = [3,1,19]$.

$e = 3$: The set $\mathsf{babySet}_3$ contains the pair $\big([3,1,19],1\big)$ with first component $\mathsf{giantElement}_3$. Hence, the order of C is $\operatorname{order} C = 3(3+1)/2 - 1 = 5$.

9.7.3 Terr's algorithm – computing the structure

We now explain Algorithm `HNFRelationBasis` that uses algorithm `order` from the previous section to compute the structure of G. Let

$$M = (g_1, \ldots, g_l)\ .$$

be a generating system for G.

Our algorithm computes the matrix $B \in \mathbb{Z}^{(l,l)}$ from Proposition 9.7.3. We write

$$\mathbf{b}_j = (b_{1,j}, \ldots, b_{l,j})\ .$$

Let $j \in \mathbb{Z}$, $1 \le j \le l$ and suppose that we have computed the basis vectors $\mathbf{b}_1, \ldots, \mathbf{b}_{j-1}$. We describe the computation of $\mathbf{b}_j$. The idea is as follows. Let

$$\mathbf{e}_i = (\underbrace{0,\ldots,0}_{i-1}, 1, \underbrace{0,\ldots,0}_{l-i}) \in \mathbb{Z}^l, \quad 1 \le i \le l\ .$$

The subgroup H generated by $g_1, \ldots, g_{j-1}$ is

$$H = \{\prod_{i=1}^{j-1} g_i^{x_i} : 0 \le x_i < b_{i,i}, 1 \le i < j\}\ . \tag{9.30}$$

Note that H depends on j, but for simplicity, we omit the index j. The entry $b_{j,j}$ is the order of the coset $g_j H$ in the factor group G/H. So we can use the order algorithm from the previous section to calculate that entry. We have to look for the smallest e such that

$$g_j^{e(e+1)/2} = g_j^f h \tag{9.31}$$

for some $f \in \{0, \ldots, e-1\}$ and some $h \in H$. As in algorithm `order`, we could store the values on the right hand side and try to find an e that satisfies (9.31). However, H can be as big as the whole group G. This is too large to obtain the complexity that we want. Therefore, we split H into two parts. We use a decomposition

$$\{1, \ldots, j-1\} = I_1 \cup \{m\} \cup I_2 \tag{9.32}$$

where the three sets on the right hand side are pairwise disjoint. Those sets will be specified later. Let

$$H_1 = \{(M^{-\mathbf{v}}, \mathbf{v}) : \mathbf{v} = \sum_{i \in I_1} x_i \mathbf{e}_i, 0 \le x_i < b_{i,i}, i \in I_1\} \tag{9.33}$$

and

$$H_2 = \{(M^{\mathbf{v}}, \mathbf{v}) : \mathbf{v} = \sum_{i \in I_2} x_i \mathbf{e}_i, 0 \le x_i < b_{i,i}, i \in I_2\} . \tag{9.34}$$

The decomposition in (9.32) is chosen such that

$$|H_i| \le \sqrt{|H|}, \quad i = 1, 2 . \tag{9.35}$$

We set

$$s = \left\lceil \sqrt{|H|}/|H_1| \right\rceil \tag{9.36}$$

and

$$t = \left\lceil \sqrt{|H|}/|H_2| \right\rceil . \tag{9.37}$$

Now we have the following result.

Lemma 9.7.10. *Any $h \in H$ can be written as $h = h_1^{-1} g_m^{qs+r} h_2$ where h_i is the first entry of a pair in H_i, $i = 1, 2$ and we have $0 \le q < t$ and $0 \le r < s$.*

Proof. By (9.30) and (9.32) we can write

$$h = h_1^{-1} g_m^n h_2$$

where h_i is the first entry of a pair in H_i, $i = 1, 2$ and $n \in \{0, \ldots, b_{m,m} - 1\}$. Write $n = qs + r$ with $0 \le r < s$. Then $qs < b_{m,m}$. Hence $q < b_{m,m}/s \le b_{m,m}|H_1|/\sqrt{|H|} = |H|/(|H_2|\sqrt{|H|}) \le t$. □

Now we modify (9.31). To find $\mathbf{b}_j$ we look for the smallest e such that

$$g_j^{e(e+1)/2} h_2 g_m^{qs} = g_j^f h_1 g_m^{-r} , \tag{9.38}$$

where $(h_i, \mathbf{v}_i) \in H_i$ for some $\mathbf{v}_i$, $i = 1, 2$, $0 \le r < s$, $0 \le q < t$, and $0 \le f < e$. Then

$$\mathbf{b}_j = \mathbf{v}_1 + \mathbf{v}_2 + (qs + r)\mathbf{e}_m + (e(e+1)/2 - f)\mathbf{e}_j . \tag{9.39}$$

To look for a match of the form (9.38) we use two sets. The first one is

$$\mathsf{babySet} = \{(g_j^f h, \mathbf{v} - f\mathbf{e}_j) : (h, \mathbf{v}) \in \mathsf{auxiliaryBabySet}, 0 \le f < e\}$$

where

$$\mathsf{auxiliaryBabySet} = \{(h_1 g_m^{-r}, \mathbf{v} + r\mathbf{e}_m) : (h_1, \mathbf{v}) \in H_1, 0 \le r < s\} .$$

So in babySet we store the elements from the right hand side of (9.38). The second set is

$$\mathsf{giantSet} = \{(h_2 g_m^{qs}, \mathbf{v} + qs\mathbf{e}_m) : (h_2, \mathbf{v}) \in H_2, 0 \le q < t\} .$$

As in the order algorithm we use

$$\mathsf{babyElement} = g_j^e, \quad \mathsf{giantElement} = g_j^{e(e+1)/2} .$$

In iteration e we multiply giantElement with each element of giantSet and check whether the product is in babySet. If this happens, we have the match we were looking for and can compute $\mathbf{b}_j$. If there is no such match, we increment e, update babySet, babyElement and giantElement and repeat the procedure. If $\mathbf{b}_j$ has been determined and $j = l$, then the algorithm terminates. If $\mathbf{b}_j$ has been determined and $j < l$, then a new decomposition (9.32) is determined and the sets I_1, I_2, H_1, H_2, auxiliaryBabySet, and giantSet are updated. If $b_{j,j} = 1$, then the decomposition (9.32) and the sets I_1, I_2, H_1, H_2, auxiliaryBabySet remain unchanged. The treatment of the other cases can be seen in the algorithm.

We explain the choice of I_1, I_2 and m. Initially, we set $I_1 = I_2 = \emptyset$, $m = 0$, $H_1 = H_2 = H = \{1\}$. Then (9.35) is satisfied. Suppose that $1 \le j < l$ and assume that $\mathbf{b}_j$ has been determined. We explain how to update I_1, I_2, and m for the next iteration. If

$$b_{j,j} \prod_{i \in I_1} b_{i,i} \le \sqrt{\prod_{i=1}^{j} b_{i,i}}\,, \tag{9.40}$$

then we replace I_1 by $I_1 \cup \{j\}$ and (9.35) is still satisfied. If (9.40) is false and if $m > 0$, then

$$\prod_{i \in I_2 \cup \{m\}} b_{i,i} < \sqrt{\prod_{i=1}^{j} b_{i,i}}\,. \tag{9.41}$$

So we replace I_2 by $I_2 \cup \{m\}$ and m by j and (9.35) is satisfied. Finally, if (9.40) is false and if $m = 0$, then we set m equal to j and (9.35) is satisfied.

Here is an example.

Example 9.7.11. Let $\Delta = -227$ and $f_1 = (3, 1, 9)$, $f_2 = (7, 5, 9)$. Then $\Delta(f_i) = \Delta$, $1 \le i \le 2$. We determine the structure of the group generated by the equivalence classes $C_i = [f_i]$, $1 \le i \le 2$. We use the generating system $M = (C_1, C_2)$.

In Example 9.7.9 we have already found that the order of C_1 is 5. Therefore, the first basis vector of the HNF-basis of M is $(5, 0)$.

We explain the computation of $\mathbf{b}_2$.

Initialization: $m = 1$, $H_1, H_2 = ([1, 1, 57], (0, 0))$, $s = t = 3$

$$\mathsf{auxiliaryBabySet} = \left\{ \begin{array}{c} \big(C_1^{0} = [1,1,57], (0,0)\big), \\ \big(C_1^{-1} = [3,-1,19], (1,0)\big), \\ \big(C_1^{-2} = [7,-5,9], (2,0)\big) \end{array} \right\},$$

$$\mathsf{giantSet} = \left\{ \begin{array}{c} \big(C_1^{0} = [1,1,57], (0,0)\big), \\ \big(C_1^{3} = [7,-5,9], (3,0)\big), \\ \big(C_1^{6} = [3,1,19], (6,0)\big) \end{array} \right\}.$$

We also have babyElement = giantElement = C_2, babySet = auxiliaryBabySet.

$e = 1$. We find that (giantElement $* [7, -5, 9] = [1, 1, 57], (0, 0)) \in$ babySet. Hence, $\mathbf{b}_2 = (0, 0) + (3, 0) + (0, 1) = (3, 1)$.

Algorithm 9.3 `HNFRelationBasis` (M)

Input: A sequence $M = (g_1, \ldots, g_l)$ of group elements
Output: The HNF-Basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_l)$ of $L(M)$

```
$H_i \leftarrow \{(1, (0, \ldots, 0))\}$, $i = 1, 2$
$I_i \leftarrow \emptyset$, $i = 1, 2$
$s \leftarrow 0$, $t \leftarrow 0$, $m \leftarrow 0$
auxiliaryBabySet $\leftarrow H_1$, giantSet $\leftarrow H_2$
for $j = 1, \ldots, l$ do
  $e \leftarrow 1$
  babySet $\leftarrow$ auxiliaryBabySet, babyElement $\leftarrow g_j$, giantElement $\leftarrow g_j$
  loop
    for all $(g, \mathbf{v}) \in$ giantSet do
      if babySet contains a pair $(g \cdot \mathsf{giantElement}, \mathbf{w})$ then
        $\mathbf{b}_j \leftarrow \mathbf{v} + \mathbf{w} + (e(e+1)/2)\mathbf{e}_j$
        break
    babySet $\leftarrow$ babySet $\cup \{(g \cdot \mathsf{babyElement}, \mathbf{v} - e\mathbf{e}_j) : (g, \mathbf{v}) \in \mathsf{auxiliaryBabySet}\}$
    $e \leftarrow e + 1$, babyElement $\leftarrow$ babyElement $\cdot g_j$, giantElement $\leftarrow$ giantElement $\cdot$ babyElement
  if $j < l$ and $b_{j,j} > 1$ then
    if $b_{j,j} \prod_{i \in I_1} b_{i,i} \leq \sqrt{\prod_{i=1}^{j} b_{i,i}}$ then
      $I_1 \leftarrow I_1 \cup \{j\}$
      $H_1 \leftarrow H_1 \cup \{(g_j^{-x} g, \mathbf{v} + x\mathbf{e}_j) : (g, \mathbf{v}) \in H_1, 0 \leq x < b_{j,j}\}$
    else
      if $m > 0$ then
        $I_2 \leftarrow I_2 \cup \{m\}$
        $H_2 \leftarrow H_2 \cup \{(g_m^x g, \mathbf{v} + x\mathbf{e}_m) : (g, \mathbf{v}) \in H_2, 0 \leq x < b_{m,m}\}$
      $m \leftarrow j$
    $s \leftarrow \lceil \sqrt{\prod_{i=1}^{j} b_{i,i}} / \prod_{i \in I_1} b_{i,i} \rceil$
    $t \leftarrow \lceil \sqrt{\prod_{i=1}^{j} b_{i,i}} / \prod_{i \in I_2} b_{i,i} \rceil$
    auxiliaryBabySet $\leftarrow \{(h_1 g_m^{-r}, \mathbf{v} + r\mathbf{e}_m) : (h_1, \mathbf{v}) \in H_1, 0 \leq r < s\}$
    giantSet $\leftarrow \{(h_2 g_m^{qs}, \mathbf{v} + qs\mathbf{e}_m) : (h_2, \mathbf{v}) \in H_2, 0 \leq q < t\}$
return $(\mathbf{b}_1, \ldots, \mathbf{b}_k)$
```

9.7.4 Analysis of the structure algorithm `HNFRelationBasis`

In the analysis of the structure algorithm we need the following auxiliary lemma which compares sums and products of arbitrary numbers in $\mathbb{R}_{\geq 1}$.

Lemma 9.7.12. *1. Let $k \in \mathbb{N}$ and $a_1, \ldots, a_k \in \mathbb{R}_{\geq 1}$. Then*
$\sum_{j=1}^{k} a_j \leq \prod_{j=1}^{k} a_j + (k-1)$
2. Let $k \in \mathbb{N}$ and $a_1, \ldots, a_k \in \mathbb{R}_{\geq 2}$. Then
$\sum_{j=1}^{k} \prod_{i=1}^{j} \sqrt{a_j} \leq (2 + \sqrt{2}) \prod_{j=1}^{k} \sqrt{a_j}$

Proof. 1. For any $x, y \in \mathbb{R}_{\geq 1}$ we have

$$x + y - xy - 1 = \underbrace{(x-1)}_{\geq 0}\underbrace{(1-y)}_{\leq 0} \leq 0.$$

Hence

$$x + y \leq xy + 1. \tag{9.42}$$

Now, we prove the first statement of the lemma by induction.

For $k = 1$, the assertion is true.

Assume that the statement is true for $k - 1$. Then (9.42) implies

$$\sum_{j=1}^{k} a_j = \sum_{j=1}^{k-1} a_j + a_k \leq \underbrace{\prod_{j=1}^{k-1} a_j}_{\geq 1} + a_k + (k-2) \leq \prod_{j=1}^{k} a_j + (k-1)$$

2. Let $B = \prod_{j=1}^{k} \sqrt{a_j}$. Then

$$\begin{aligned}\sum_{j=1}^{k}\left(\prod_{i=1}^{j}\sqrt{a_i}\right) &= B\sum_{j=1}^{k}\left(\prod_{i=j+1}^{k}\frac{1}{\sqrt{a_i}}\right) \leq B\sum_{j=1}^{k}\left(\prod_{i=j+1}^{k}\frac{1}{\sqrt{2}}\right)\\ &= B\sum_{j=1}^{k}\left(\frac{1}{\sqrt{2}}\right)^{k-j} = B\sum_{j=0}^{k-1}\left(\frac{1}{\sqrt{2}}\right)^{j} = B\frac{1-(1/\sqrt{2})^k}{1-1/\sqrt{2}}\\ &= B(2+\sqrt{2} - \frac{2+\sqrt{2}}{\sqrt{2}^k}) \leq (2+\sqrt{2})\prod_{j=1}^{k}\sqrt{a_j}\end{aligned}$$

□

We now present the complexity result. By $l(M)$ we denote the number of diagonal entries in the HNF-basis of $L(M)$ that are greater than 1.

Theorem 9.7.13. *Algorithm* `HNFRelationBasis` *computes the HNF-basis of the lattice of relations on M and executes*

- *at most $l = |M|$ inversions,*
- *at most $(48 + 8l - 6l(M))\sqrt{|G|} + 2l(M)\log\sqrt{|G|}$ multiplications in G,*
- *at most $4(2 + \sqrt{2} + l - l(M))\sqrt{|G|}$ table look-ups.*

The algorithm uses

- *two tables of at most $\sqrt{|G|}$,*
- *two tables of at most $2\sqrt{|G|}$,*
- *one table of at most $4\sqrt{|G|}$*

pairs $(g, \mathbf{q}) \in G \times \{0, \ldots, \lfloor\sqrt{|G|}\rfloor\}^{|M|}$.

Proof. Correctness of `HNFRelationBasis` follows from Theorem 9.7.8, Lemma 9.7.10 and the arguments following the proof of this lemma.

We first estimate, the sizes of the sets H_1, H_2, babySet, auxiliaryBabySet, and giantSet. Then we estimate the number of group operations and table look-ups.

Consider the computation of $\mathbf{b}_j$. Let $e(j)$ be the final value for e in the computation of $\mathbf{b}_j$. By Lemma 9.7.8 we have

$$e(j) < \sqrt{2b_{j,j}} + 1/2 \leq 2\sqrt{b_{j,j}}.$$

First, `HNFRelationBasis` computes the order of the coset g_jH in the factor group G/H, and the vector $\mathbf{b}_j$. Then, it updates the sets H_1, H_2, auxiliaryBabySet, and giantSet for the next loop.

We analyze the first step. By (9.35) we have

$$|H_i| \leq \sqrt{|H|} = \sqrt{\prod_{i=1}^{j} b_{i,i}} \leq \sqrt{|G|}, \quad i = 1, 2. \tag{9.43}$$

It follows from (9.36) and (9.37) that

$$|\mathsf{auxiliaryBabySet}| = s|H_1| \leq 2\sqrt{|G|} \text{ and } |\mathsf{giantSet}| = t|H_2| \leq 2\sqrt{|G|}. \tag{9.44}$$

The set babySet is constructed from set auxiliaryBabySet from the $(j-1)$th iteration. Therefore we have

$$|\mathsf{babySet}| \leq e(j)|\mathsf{auxiliaryBabySet}| \leq 4\sqrt{b_{j,j}} \cdot \sqrt{\prod_{i=1}^{j-1} b_{j,j}} \leq 4\sqrt{|G|}.$$

We estimate the number of table look-ups. We consider the cases $e(j) = 1$ and $e(j) > 1$. In the first case, we have $b_{j,j} = 1$. In the second case we have $b_{j,j} \geq 2$. By Lemma 9.7.12 we have at most

$$\begin{aligned}
\sum_{j=1}^{l} e(j)|\mathsf{giantSet}| &\leq \sum_{j=1}^{l} 2e(j)\sqrt{\prod_{i=1}^{j-1} b_{i,i}} \leq \sum_{j=1}^{l} 4\prod_{i=1}^{j}\sqrt{b_{i,i}} \\
&\leq 4 \sum_{j=1,e(j)>1}^{l} \prod_{i=1}^{j}\sqrt{b_{i,i}} + 4 \sum_{j=1,e(j)=1}^{l} \prod_{i=1}^{j}\sqrt{b_{i,i}} \\
&\leq 4(2+\sqrt{2}) \prod_{j=1,e(j)>1}^{l} \sqrt{b_{i,i}} + 4(l - l(M)) \prod_{j=1}^{l}\sqrt{b_{i,i}} \\
&\leq 4(2+\sqrt{2}+l-l(M))\sqrt{|G|}
\end{aligned}$$

table look-ups.

We estimate the number of multiplications. The number of multiplications necessary to multiply all first elements of giantSet by giantElement is

$$M_1 \le \sum_{j=1}^{l} e(j)|\mathsf{giantSet}| \le 4(2+\sqrt{2}+l-l(M))\sqrt{|G|}.$$

The number of multiplications to update babySet is

$$M_2 \le \sum_{j=1}^{l} e(j)|\mathsf{auxiliaryBabySet}| \le 4(2+\sqrt{2}+l-l(M))\sqrt{|G|}.$$

The number of multiplications necessary to update babyElement and giantElement is

$$\begin{aligned}
M_3 &\le \sum_{j=1}^{l}(2e(j)-2) \le 2\sum_{j=1}^{l} e(j) - 2l = 2 \sum_{j=1,e(j)>1}^{l} e(j) + 2 \sum_{j=1,e(j)=1}^{l} e(j) - 2l \\
&\le 2 \sum_{j=1,e(j)>1}^{l} (\sqrt{2b_{j,j}} + 1/2) + 2(l-l(M)) - 2l \\
&\le 2\sqrt{2} \sum_{j=1,e(j)>1}^{l)} \sqrt{b_{j,j}} + l(M) - 2l(M) \\
&\le 2\sqrt{2} \prod_{j=1,e(j)>1}^{l} \sqrt{b_{j,j}} + 2\sqrt{2}(l(M)-1) - l(M) \quad \text{(Lemma 9.7.12)} \\
&\le 2\sqrt{2|G|} + 2l(M).
\end{aligned}$$

Now we analyze the number of multiplications necessary to update H_1, H_2, auxiliaryBabySet, and giantSet. No multiplications are executed, if $\mathbf{b}_{j,j} = 1$. In each loop with $\mathbf{b}_{j,j} > 1$ either $|H_1|$ multiplications are necessary to update H_1 or $|H_2|$ multiplications are necessary to update H_2. Next, $|\mathsf{auxiliaryBabySet}| + |\mathsf{giantSet}|$ multiplications are used to update auxiliaryBabySet and babySet, and finally at most $2\lfloor \log \sqrt{|G|} \rfloor$ multiplications are performed to compute g_m^s during the computation of giantSet. By (9.43), (9.44) and Lemma 9.7.12, the total number of multiplications required for those updates is

$$\begin{aligned}
M_4 &\le \sum_{j=1}^{l}(5\sqrt{\prod_{i=1}^{j} b_{i,i}} + 2\lfloor \log \sqrt{|G|} \rfloor) \\
&\le 5(2+\sqrt{2})\sqrt{|G|} + 2l(M)\log\sqrt{|G|}.
\end{aligned}$$

So the number of multiplications is

$$M \le M_1 + M_2 + M_3 + M_4 \le (48 + 8l - 6l(M))\sqrt{|G|} + 2l(M)\log\sqrt{|G|}$$

Finally the algorithm executes at most l inversions. □

Corollary 9.7.14. *Computing the structure of the finite Abelian group G from the generating system M requires storing $\mathrm{O}(\sqrt{|G|})$ pairs $(g, \mathbf{q}) \in G \times \{0, \ldots, \lfloor\sqrt{|G|}\rfloor\}^{|M|}$, $\mathrm{O}(|M|\sqrt{|G|})$ multiplications and inversions in G, $\mathrm{O}(|M|\sqrt{|G|})$ table look-ups, and $(|M| \log |G|)^{\mathrm{O}(1)}$ bit operations.*

Proof. The number of group operations and table look-ups is estimated in Theorem 9.7.13. We estimate the number of bit operations. When the algorithm updates the sets auxiliaryBabySet, babySet, giantSet, H_1, and H_2 it executes $|M|\sqrt{|G|}(\log |G|)^O(1)$ bit operations. By Propositions A.5.17 and A.5.19, the computation of the Smith normal forms is possible in time $(|M| \log |G|)^{\mathrm{O}(1)}$ The entries of the HNF-basis are in $\{0, \ldots, |G|\}$, and the dimension of that matrix is $|M| \times |M|$. □

9.7.5 Application to class groups

We can now prove the first complexity result for class group structure computation. By $\mathrm{o}(1)$ we mean a function that converges to zero as $|\Delta|$ goes to infinity.

Proposition 9.7.15. *Let Δ be a discriminant. The structure of Cl_Δ can be computed using time and storage $|\Delta|^{1/2+\mathrm{o}(1)}$.*

Proof. Let $\Delta < 0$. Using the algorithm from Section 9.6 a generating system for Cl_Δ can be computed using $|\Delta|^{1/2+\mathrm{o}(1)}$ group operations and storing $|\Delta|^{1/2+\mathrm{o}(1)}$ group elements. Then a generating system for Cl_Δ of length $|\Delta|^{\mathrm{o}(1)}$ is known. An application of Proposition 9.7.6 and Proposition 9.3.11 implies the assertion.

The proof for the case $\Delta > 0$ is left to the reader as Exercise 9.8.8. □

Proposition 9.7.16. *(ERH) Let Δ be a discriminant. Assuming the ERH the structure of the class group Cl_Δ can be computed in time $|\Delta|^{1/4+\mathrm{o}(1)}$.*

Proof. Let $\Delta < 0$. We use the generating system from Proposition 9.5.3. The number of elements in the generating system is $|\Delta|^{\mathrm{o}(1)}$. Also, we have $h_\Delta = |\Delta|^{1/2+\mathrm{o}(1)}$ by Proposition 9.3.11. Hence, Corollary 9.7.14 implies the assertion.

To prove the Proposition for the case $\Delta > 0$ we have to apply the results of Chapter 10. The proof is left to the reader as Exercise 10.4.3. □

9.8 Exercises

Exercise 9.8.1. Verify Example 9.1.2.

Exercise 9.8.2. Prove that equivalence and proper equivalence of $\mathcal{O}$-ideals are equivalence relations.

Exercise 9.8.3. Let $\Delta > 0$ and assume that the norm of the fundamental unit of $\mathcal{O}$ is 1. Show that the class group Cl_Δ is isomorphic to the quotient group $\mathrm{Cl}_\Delta^+/\big\langle L(-c, b, 1)\big\rangle$ where $(1, b, c)$ is the principal form of discriminant Δ.

Exercise 9.8.4. Prove Corollary 9.4.8.

Exercise 9.8.5. Show that the computation of M^U in Proposition 9.7.3 requires $\mathrm{O}(l \log |G|)$ group operations if U is chosen such that the entries in U are bounded by $|G|$.

Exercise 9.8.6. Let G be a finite Abelian group. Let $g, h \in G$. Let

$$\begin{pmatrix} b_{1,1} & b_{1,2} \\ 0 & b_{2,2} \end{pmatrix}$$

be the HNF-basis of the relation lattice $L(g, h)$. Prove that $h \in \langle g \rangle$ if and only if $b_{2,2} = 1$. Also show that if $b_{2,2} = 1$, then $b_{1,2}$ is the discrete logarithm of h to the base g.

Exercise 9.8.7. Let G be a group. Let $g \in G$ and let n be the order of G.

1. Prove the following. There are $e \geq 1$ and $f \in \{1, \ldots, e\}$ such that $g^{e(e-1)/2+f} = 1$. If e is chosen minimal with this property, then $e(e-1)/2 \leq \mathrm{order}(g) < e(e+1)/2$, f is uniquely determined, and $\mathrm{order}(g) = e(e-1)/2 + f$.
2. Describe and analyze an algorithm for computing the order of an element of G that is based on 1.

Exercise 9.8.8. Prove Proposition 9.7.15 for the case $\Delta > 0$.

Exercise 9.8.9. Determine the structure of the class group Cl_{-163}.

Chapter references and further reading

[Apo86] Tom M. Apostol, *Introduction to analytic number theory*, Springer-Verlag, 1986.

[Bac90] Eric Bach, *Explicit bounds for primality testing and related problems*, Mathematics of Computation **55** (1990), 355–380.

[Bac95] ———, *Improved approximations for Euler products*, Fourth Conference of the Canadian Number Theory Association (Karl Dilcher, ed.), CMS Proceedings, vol. 15, Canadian Mathematical Society, 1995, pp. 13–28.

[BJT97] Johannes Buchmann, Michael J. Jacobson, Jr., and Edlyn Teske, *On some computational problems in finite abelian groups*, Mathematics of Computation **66** (1997), no. 220, 1663–1687.

[BP95] Johannes Buchmann and Sachar Paulus, *Algorithms for finite abelian groups*, Proceedings of Number Theoretic and Algebraic Methods in Computer Science (NTAMCS) '93 (Singapor) (van der Poorten, Shparlinski, and Zimmer, eds.), World Scientific, 1995.

[BS05] Johannes Buchmann and Arthur Schmidt, *Computing the structure of a finite abelian group*, Mathematics of Computation **74** (2005), 2017–2026.

[BW91] Johannes Buchmann and Hugh C. Williams, *Some remarks concerning the complexity of computing class groups of quadratic fields*, J. Complexity **7** (1991), no. 3, 311–315 (English).

[CyDO01] Henri Cohen, Franzisco Diaz y Diaz, and Michel Olivier, *Algorithmic methods for finitely generated abelian groups*, Journal of Symbolic Computation **31** (2001), no. 1-2, 133–147.

[HS06] Safuat Hamdy and Filip Saidak, *Arithmetic properties of class numbers of imaginary quadratic fields*, JP Journal of Algebra, Number Theory and Applications **6** (2006), no. 1, 129–148.

[Hua42] Loo-keng Hua, *On the least solution of Pell's equation*, Bulletin of the American Mathematical Society **48** (1942), 731–735.

[Jac98] Michael J. Jacobson, Jr., *Experimental results on class groups of real quadratic fields*, Algorithmic Number Theory, ANTS-III (Joe P. Buhler, ed.), Lecture Notes in Computer Science, vol. 1423, Springer-Verlag, 1998, pp. 463–474.

[JRW06] Michael J. Jacobson, Jr., S. Ramachandran, and H.C. Williams, *Numerical results on class groups of imaginary quadratic fields*, Algorithmic Number Theory, ANTS-VII (Michael Pohst Florian Hess, Sebastian Pauli, ed.), Lecture Notes in Computer Science, vol. 4076, Springer-Verlag, 2006, pp. 87–101.

[Lan66] Edmund Landau, *Elementary number theory*, second edition ed., Chelsea Publishing Company, 1966.

[MW92] Richard A. Mollin and Hugh C. Williams, *On real quadratic fields of class number two*, Mathematics of Computation **59** (1992), no. 200, 625–632.

[PZ85] Michael Pohst and Hans Zassenhaus, *Über die Berechnung von Klassenzahlen und Klassengruppen algebraischer Zahlkörper*, Journal für die Reine und Angewandte Mathematik **361** (1985), 50–72.

[Sch06] Arthur Schmidt, *Quantum algorithm for solving the discrete logarithm problem in the class group of an imaginary quadratic field and security comparison of current cryptosystems at the beginning of quantum computer age*, Emerging Trends in Information and Communication Security (Günter Müller, ed.), Lecture Notes in Computer Science, vol. 3995, Springer-Verlag, 2006, pp. 481–493.

[Sha71] Daniel Shanks, *Class number, a theory of factorization and genera*, 1969 Number Theory Institute (Providence, R.I.), Proceedings of Symposia in Pure Mathematics, vol. 20, AMS, 1971, pp. 415–440.

[Ter00] David C. Terr, *A modification of Shanks' baby-step giant-step algorithm*, Mathematics of Computation **69** (2000), no. 230, 767–773.

[Tes98] Edlyn Teske, *A space efficient algorithm for group structure computation*, Math. Comput. **67** (1998), no. 224, 1637–1663 (English).

[Tes99] ———, *The Pohlig-Hellman method generalized for group structure computation*, J. Symbolic Comput. **27** (1999), no. 6, 521–534. MR 2000f:20090

10

Infrastructure

Let $\mathcal{O}$ be a real quadratic order, let Δ be the discriminant of $\mathcal{O}$, and let R be the regulator of $\mathcal{O}$. In this chapter we describe an algorithm that is based on an idea of Shanks [Sha72] and Terr [Ter00]. It computes the fundamental unit of $\mathcal{O}$ in time $\mathrm{O}((\log \Delta + \sqrt{R})(\log \Delta)^2) = |\Delta|^{1/4+\mathrm{o}(1)}$. That algorithm can also be used to decide equivalence of $\mathcal{O}$-ideals and to calculate generators of principal $\mathcal{O}$-ideals. If the input ideals are reduced, then the running time of the equivalence algorithm admits the same running time bound as the regulator algorithm.

We let F be the field of fractions of $\mathcal{O}$. By $\mathcal{I}$ we denote the group of fractional invertible $\mathcal{O}$-ideals and by $\mathcal{P}$ we denote the group of principal fractional $\mathcal{O}$-ideals. Also, ε is the fundamental unit of $\mathcal{O}$.

10.1 Geometry of reduction

The algorithms described in this chapter are based on a geometric interpretation of reduction which we explain in this section. We let $\mathfrak{a}$ be a fractional $\mathcal{O}$-ideal.

10.1.1 Distance between ideals

We introduce a length of elements in F^*.

Definition 10.1.1. *For $\alpha \in F^*$ we set* $\operatorname{Log} \alpha = (1/2) \log |\sigma(\alpha)/\alpha|$.

For $\alpha \in F^*$ we can also write

$$\operatorname{Log} \alpha = \log |\sigma(\alpha)| - (1/2) \log |\mathrm{N}(\alpha)| \,. \tag{10.1}$$

Here are a few properties of this logarithm map.

Proposition 10.1.2.
1. *The map* $F^* \to \mathbb{R}$*,* $\alpha \mapsto \operatorname{Log}\alpha$ *is a homomorphism of the multiplicative group* F^* *into the additive group* $\mathbb{R}$*. The kernel of that homomorphism is* $\mathbb{Q}^* \cup \mathbb{Q}^*\sqrt{\Delta}$*.*
2. *If* η *is a unit in* $\mathcal{O}$*, then* $\operatorname{Log}\eta = -\log|\eta|$*. In particular, we have* $\operatorname{Log}\varepsilon = -R$*.*
3. *For any* $\alpha \in F^*$ *we have* $\operatorname{Log}\alpha = -\operatorname{Log}\sigma(\alpha)$*.*

Proof. Exercise 10.4.1. □

Next, we show how to construct a homomorphism of the group $\mathcal{P}$ of fractional principal $\mathcal{O}$-ideals into the circle group $\mathbb{R}/R\mathbb{Z}$.

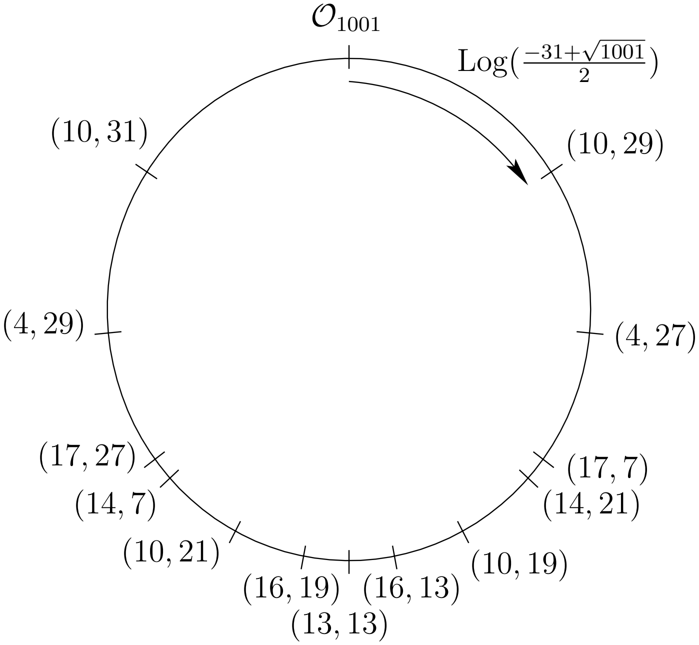

Fig. 10.1. Embedding the principle cycle of $\mathcal{O}_{1001}$ into $\mathbb{R}/R\mathbb{Z}$

Proposition 10.1.3. *The map*

$$d : \mathcal{P} \to \mathbb{R}/R\mathbb{Z}, \quad \alpha\mathcal{O} \mapsto \operatorname{Log}\alpha + R\mathbb{Z} \tag{10.2}$$

is a well defined homomorphism of the multiplicative group $\mathcal{P}$ *into the additive group* $\mathbb{R}/R\mathbb{Z}$*.*

Proof. We show that the map is well defined. Let $\alpha, \beta \in F^*$ with $\alpha\mathcal{O} = \beta\mathcal{O}$. Then α/β is a unit in $\mathcal{O}$ by Exercise 8.7.11. It follows from Theorem 8.3.5

that α/β is modulo sign a power of the fundamental unit ε of $\mathcal{O}$. Hence, $\operatorname{Log}\alpha/\beta = \operatorname{Log}\alpha - \operatorname{Log}\beta$ is an integer multiple of the regulator R.

By Lemma 10.1.2 the map is a homomorphism. □

We extend the map d from (10.2) to pairs of $\mathcal{O}$-ideals in the same ideal class. It follows from Proposition 10.1.3 that the map

$$d : F^*\mathfrak{a} \times F^*\mathfrak{a} \to \mathbb{R}/R\mathbb{Z}, \quad (\alpha\mathfrak{a}, \beta\mathfrak{a}) \mapsto \operatorname{Log}\beta - \operatorname{Log}\alpha + R\mathbb{Z} \tag{10.3}$$

is well defined (see Exercise 10.4.2).

We can now define the (oriented) distance between two $\mathcal{O}$-ideals in the same ideal class.

Definition 10.1.4. *Let $\mathfrak{a}$ and $\mathfrak{b}$ be equivalent $\mathcal{O}$-ideals. Then we call $d(\mathfrak{a}, \mathfrak{b})$ the* distance *from $\mathfrak{a}$ to $\mathfrak{b}$.*

Fixing one argument $\mathfrak{a}$, the distance function $d(\mathfrak{a}, \cdot)$ can also be viewed as a map of the $\mathcal{O}$-ideal class of $\mathfrak{a}$ to a circle of circumference R.

10.1.2 Cycles of reduced $\mathcal{O}$-ideals

Let $\mathfrak{a}$ be an $\mathcal{O}$-ideal and let $(a, b, c) = f = f_{\mathfrak{a}}$ be the corresponding normal form of discriminant Δ as defined in (8.9). Recall Lemma 9.1.9 which says

$$\rho(\mathfrak{a}) = \gamma(\mathfrak{a})\mathfrak{a} \tag{10.4}$$

with $\gamma(\mathfrak{a})$ from (9.2). Also, for reduced $\mathfrak{a}$ define

$$\gamma'(\mathfrak{a}) = \gamma(\mathfrak{a}^\sigma)^\sigma \tag{10.5}$$

and

$$\rho^{-1}(\mathfrak{a}) = \gamma'(\mathfrak{a})\mathfrak{a}\ . \tag{10.6}$$

This is proved in Exercise 10.4.7.

Set

$$\mathfrak{a}_i = \rho^i(\mathfrak{a})\ , \quad i \in \mathbb{Z}\ . \tag{10.7}$$

Also, let

$$f_i = f_{\mathfrak{a}_i} = (a_i, b_i, c_i), \quad i \in \mathbb{Z}\ . \tag{10.8}$$

Recall from Lemma 9.1.15 that $f_i = (\tau\rho)(f_{i-1})$.

Then the sequence $(\mathfrak{a}_i)_{i \in \mathbb{Z}_{\geq 0}}$ is eventually periodic of finite period length l. The period is the cycle of reduced ideals in the ideal class of $\mathfrak{a}$. If $\mathfrak{a}$ is principal, then the cycle is called the *principal cycle* of $\mathcal{O}$. We explain what the image of that cycle on the circle $\mathbb{R}/R\mathbb{Z}$ looks like. Figure 10.1 shows this image for $\Delta = 1001$. Set

$$\gamma_i = \gamma(\mathfrak{a}_i)\ , \quad \alpha_i = \prod_{j=0}^{i-1} \gamma_j\ , \quad i \in \mathbb{Z}_{\geq 0}\ . \tag{10.9}$$

Then

$$\mathfrak{a}_i = \rho(\mathfrak{a}_{i-1}) = \gamma_{i-1}\mathfrak{a}_{i-1} = \alpha_i \mathfrak{a}\ , \quad i \in \mathbb{Z}_{\geq 0}$$

by Lemma 9.1.9. Hence,

$$d(\mathfrak{a}_i, \mathfrak{a}_{i+1}) = \operatorname{Log}\gamma_i + R\mathbb{Z} \quad \text{and} \quad \operatorname{Log}\gamma_i = \frac{1}{2}\log\frac{\sqrt{\Delta}+b_i}{\sqrt{\Delta}-b_i} > 0, \quad i \in \mathbb{Z}\ . \tag{10.10}$$

Lemma 10.1.5. *If* $\mathfrak{a}$ *is reduced, then* $\operatorname{Log}\alpha_i < \operatorname{Log}\alpha_{i+1}$ *and* $\operatorname{Log}\alpha_l = R_\Delta$.

Proof. Note that $0 < b_i < \sqrt{\Delta}$ if f_i is reduced. In this case, $(\sqrt{\Delta}+b_i)/(\sqrt{\Delta}-b_i) > 1$. Hence $\operatorname{Log}\gamma_i > 0$ and $\operatorname{Log}\alpha_i < \operatorname{Log}\alpha_{i+1}$.

By (10.9) and Lemma 9.1.10

$$\gamma_i = \gamma(\mathfrak{a}_i) = -\theta_2(f_i) \quad \text{and} \quad \sigma(\gamma_i) = \sigma(-\theta_2(f_i)) = -\theta_1(f_i) = -\theta_1(f_i)$$

Proposition 8.3.7 says

$$\sigma(\prod_{i=0}^{l-1}\gamma_i) = (-1)^l\prod_{i=0}^{l-1}\theta_1(f_i) = (-1)^l\varepsilon_\Delta\ .$$

The second assertion follows now immediately from Proposition 10.1.2.

Let $\mathfrak{a}$ be again arbitrary and $\mathfrak{a}_k$ be the first reduced ideal in the sequence. If we map $\mathfrak{a}_i$ with $i \geq k$ to $\delta(\mathfrak{a}_k, \mathfrak{a}_i)$, then Lemma 10.1.5 shows that we traverse the circle $\mathbb{R}/R\mathbb{Z}$ in positive direction and encounter the reduced ideals in the order $\mathfrak{a}_k, \mathfrak{a}_{k+1}, \ldots, \mathfrak{a}_{k+l} = \mathfrak{a}_k$.

We prove upper and lower bounds on the distance from $\mathfrak{a}_i$ to $\mathfrak{a}_{i+1}$ and a lower bound for the distance from $\mathfrak{a}_i$ to $\mathfrak{a}_{i+2}$, $i \in \mathbb{Z}$.

Lemma 10.1.6.
1. $1/\sqrt{\Delta} < \operatorname{Log}\gamma_i < \frac{1}{2}\log\Delta$*,* $i \in \mathbb{Z}$*.*
2. $\operatorname{Log}\gamma_i + \operatorname{Log}\gamma_{i+1} > \log 2$*,* $i \in \mathbb{Z}$*.*

Proof. 1. Let $i \in \mathbb{Z}$. Then

$$\operatorname{Log}\gamma_i = \frac{1}{2}\log\frac{\sqrt{\Delta}+b_i}{\sqrt{\Delta}-b_i} = \frac{1}{2}\log\frac{(b_i+\sqrt{\Delta})^2}{4|a_ic_i|}\ .$$

Since f_i is reduced we have $1 \leq b_i < \sqrt{\Delta}$. Therefore, $|a_ic_i| \geq 1$ implies

$$\operatorname{Log}\gamma_i < \frac{1}{2}\log\sqrt{\Delta}\ .$$

Also, we have

$$\operatorname{Log}\gamma_i = \frac{1}{2}\log\frac{\sqrt{\Delta}+b_i}{\sqrt{\Delta}-b_i} \geq \frac{1}{2}\log\frac{\sqrt{\Delta}+1}{\sqrt{\Delta}-1} > \frac{1}{\sqrt{\Delta}}\ .$$

The last inequality follows from $\log(x+1) - \log(x-1) > 2/x$ for all $x > 1$.

2. To prove the second lower bound, we note that $|\gamma_i \gamma_{i+1}| = |\mu_i/\mu_{i+2}|$ with μ_i as in Section 6.8.3. It follows that

$$\operatorname{Log}\gamma_i + \operatorname{Log}\gamma_{i+1} = \frac{1}{2}(\log|\mu_{i+2}/\mu_i| + \log|\sigma(\mu_i/\mu_{i+2})|) \ .$$

Lemma 6.8.8 and Exercise 6.18.18 imply

$$\operatorname{Log}\gamma_i + \operatorname{Log}\gamma_{i+1} > \log 2 \ . \qquad \square$$

Corollary 10.1.7. *The length l of the cycle of any $\mathcal{O}$-ideal is bounded by $2R/\log 2 + 1$.*

Proof. It follows from Lemma 10.1.6 that $R = \sum_{i=1}^{l} \operatorname{Log}\gamma_i > \lceil l/2 \rceil \log 2$. This implies the assertion. □

We prove an upper bound for the distance traversed during reduction of a non-reduced ideal $\mathfrak{a}$. Recall that $\mathfrak{a}_k$ is the first reduced ideal in the sequence $\mathfrak{a}_i$, $i \geq 0$.

Lemma 10.1.8. *We have*

$$|\operatorname{Log}\alpha_k| < 1/2 \log a_0 \ .$$

Proof. We prove the upper bound for $\operatorname{Log}\alpha_k$. If $a_i > \sqrt{\Delta}$, then, by definition, $|b_i| \leq a_i$. Hence

$$\operatorname{Log}\gamma_i = \frac{1}{2}\log\frac{(b_i+\sqrt{\Delta})^2}{4|a_i c_i|} < \frac{1}{2}\log\frac{4a_i^2}{4|a_i c_i|} = \frac{1}{2}(\log|a_i| - \log|c_i|) \ .$$

Let m be the largest index for which $a_i > \sqrt{\Delta}$. Since $a_{i+1} = c_i$, we see

$$\sum_{i=0}^{m} \operatorname{Log}\gamma_i < \frac{1}{2}(\log|a_0| - \log|c_m|) \ .$$

If $k = m+1$, i.e. if $\mathfrak{a}_{m+1}$ is reduced, then we are done. If $\mathfrak{a}_{m+1}$ is not reduced, then $a_{m+1} < \sqrt{\Delta}$ implies that $\mathfrak{a}_{m+2}$ is indeed reduced, and $k = m+2$. Moreover, we know $b_{m+1} < 2a_{m+1} - \sqrt{\Delta}$ since $\mathfrak{a}_{m+1}$ is not reduced. Hence

$$\begin{aligned}\operatorname{Log}\gamma_{m+1} &= \log(b_{m+1} + \sqrt{\Delta}) - 1/2\log(4|a_{m+1}c_{m+1}|)\\ &< \log(2a_{m+1}) - 1/2\log(4a_{m+1}) = 1/2\ \log(a_{m+1}) = 1/2\ \log|c_m| \ .\end{aligned}$$

The lower bound for $\operatorname{Log}\alpha_k$ is proved in analogy using

$$\operatorname{Log}\gamma_i = \frac{1}{2}\log\frac{4|a_i c_i|}{(b_i - \sqrt{\Delta})^2} > \frac{1}{2}\log\frac{4|a_i c_i|}{4a_i^2} = \frac{1}{2}(\log|c_i| - \log|a_i|) \ . \qquad \square$$

10.2 A Terr algorithm

Embedding the set of reduced principal ideals into the circle $\mathbb{R}/R\mathbb{Z}$ of circumference R is similar to embedding a cyclic group of order n into the circle $\mathbb{R}/n\mathbb{Z}$ of circumference n. (See Figure 10.1.) While two neighboring group elements on the second circle have a fixed distance of 1, two reduced ideals have a distance of at least $1/\sqrt{\Delta}$ and at most $\frac{1}{2}\log\Delta$ (see Lemma 10.1.6). This indicates that computing the regulator is similar to computing the order of a cyclic group. Also, computing the logarithm of a generator of a reduced ideal is like computing the discrete logarithm of a group element.

In this section we will use those analogies to develop an algorithm for computing the regulator R and for solving the equivalence problem. That algorithm will be similar to the Terr algorithm described in Section 9.7.2.

The situation is the following. We are given reduced $\mathcal{O}$-ideals $\mathfrak{a}$ and $\mathfrak{b}$. The goal is to decide whether or not $\mathfrak{a}$ and $\mathfrak{b}$ are equivalent. Also, if $\mathfrak{a}$ and $\mathfrak{b}$ are equivalent, then we want to find $\lambda \in F$ with

$$\mathfrak{b} = \lambda\mathfrak{a}, \quad 0 \leq \operatorname{Log}\lambda < R .$$

10.2.1 Outline of the algorithm

The algorithm uses two sequences of reduced $\mathcal{O}$-ideals. The first sequence $\mathfrak{a}_0, \mathfrak{a}_1, \ldots$ starts at

$$\mathfrak{a}_0 = \mathfrak{a}$$

and is defined by

$$\mathfrak{a}_{i+1} = \rho(\mathfrak{a}_i) , \quad i \geq 0 .$$

It is called the *baby-step sequence* because the distance between two consecutive elements of the sequence is very small. Set

$$\alpha_i = \prod_{j=0}^{i-1} \gamma_j , \quad i \geq 0$$

with γ_j from Section 10.1.2. Then

$$\mathfrak{a}_i = \alpha_i\mathfrak{a} , \quad i \geq 0 . \tag{10.11}$$

Also, it follows from Lemma 10.1.6 that

$$\operatorname{Log}\alpha_{i+1} > \operatorname{Log}\alpha_i > 0 , \quad i \geq 0 , \tag{10.12}$$

$$\operatorname{Log}\alpha_{i+2} \geq \operatorname{Log}\alpha_i + \log 2 , \quad i \geq 0 , \tag{10.13}$$

and

$$\lim_{i\to\infty} \operatorname{Log}\alpha_i = \infty . \tag{10.14}$$

Initially, the baby-step sequence is calculated until $L \geq 0$ is found with

$$\operatorname{Log} \alpha_{2(L+1)} \geq \frac{1}{2} \log \Delta \ . \tag{10.15}$$

Define

$$s_i = \operatorname{Log} \alpha_{2(L+i)} \ , \quad i \geq 1 \ . \tag{10.16}$$

The second sequence $\mathfrak{b}_0, \mathfrak{b}_1, \dots$ starts at

$$\mathfrak{b}_0 = \mathfrak{b} \ .$$

It is called the *giant-step sequence* since the distances between the consecutive elements of that sequence become larger and larger. More precisely, the algorithm determines positive numbers $\delta_i \in F$ such that

$$\mathfrak{b}_i = (\delta_i / \alpha_{2(L+i)}) \mathfrak{b}_{i-1} \ , \quad i \geq 1 \ ,$$

is reduced. If

$$\beta_i = \delta_i / \alpha_{2(L+i)}, \quad i \geq 1 \ , \tag{10.17}$$

then we require

$$s_i - \frac{1}{2} \log \Delta < - \operatorname{Log} \beta_i \leq s_i \ , \quad i \geq 1 \ . \tag{10.18}$$

In section 10.2.3 we will show how to construct δ_i and $\mathfrak{b}$ with the desired properties.

The algorithm is based on the following proposition.

Proposition 10.2.1. *Assume that* $\mathfrak{a}$ *and* $\mathfrak{b}$ *are equivalent. Then the following are true.*

1. There are e, f *such that* $e \geq 0$ *and*

$$\mathfrak{b}_e = \mathfrak{a}_f \ , \quad f \in \{1, \dots, 2(e + L + 1)\} \ . \tag{10.19}$$

2. If (e, f) *is the lexicographically smallest pair that satisfies* (10.19)*, then for*

$$\lambda = \alpha_f / \prod_{i=1}^{e} \beta_i \ , \tag{10.20}$$

we have $\mathfrak{b} = \lambda \mathfrak{a}$ *and*

$$0 < \operatorname{Log} \lambda \leq R \ . \tag{10.21}$$

Also,

$$e < \lceil E_\Delta \rceil \tag{10.22}$$

where

$$E_\Delta = \log_2 \Delta + \sqrt{2R / \log 2} \ . \tag{10.23}$$

3. *If* $\mathfrak{a} = \mathfrak{b} = \mathcal{O}$ *and* e, f, *and* λ *are as in 2., then* $\operatorname{Log}\lambda = R$ *and the fundamental unit of* $\mathcal{O}$ *is* $1/\lambda$.

For $e = 1, 2, \dots$ the algorithm calculates the baby-step sequence $(\mathfrak{a}_i)_{1\le i\le 2(e+L+1)}$ and the giant-step $\mathfrak{b}_e$. If a *match* (10.19) is found, then the equivalence of $\mathfrak{a}$ and $\mathfrak{b}$ is proved and with λ from (10.20) we have $\mathfrak{b} = \lambda\mathfrak{a}$ and $0 < \operatorname{Log}\lambda \le R$.

Note that

$$\lambda = \prod_{i=0}^{f-1}\gamma_i \cdot \prod_{j=0}^{2L+1}\gamma_j^e \cdot \prod_{k=1}^{e-1}(\gamma_{2(L+k)}\gamma_{2(L+k)+1})^{e-k} \cdot \prod_{l=1}^{e}\delta_l^{-1} \; . \tag{10.24}$$

Since $\gamma_i = \gamma(\mathfrak{a}_i)$ can easily be determined from $\mathfrak{a}_i$, it suffices to store $\mathfrak{a}_i$, $1 \le i \le \max\{f, 2(e+L)\}$, and δ_i, $1 \le i \le e$, in order to compute the power product representation of λ in (10.24). This will actually be the representation of output λ of our algorithm.

If for no $e < \lceil E_\Delta \rceil$ a match (10.19) is found, then it is established that $\mathfrak{a}$ and $\mathfrak{b}$ are not equivalent. It follows from (10.22) that e and f are of the order of magnitude $\sqrt{2R} = \Delta^{1/4+o(1)}$. We will see that the running time of the algorithm is of the same order of magnitude.

To show Proposition 10.2.1 we prove the following auxiliary result.

Lemma 10.2.2. *There is some* $k \ge 0$ *with* $\operatorname{Log}\lambda_k \le s_{k+1}$. *Also, if* e *is the smallest such* k, *then* $\operatorname{Log}\lambda_e > 0$.

Proof. It follows from (10.18), the definition of s_i in (10.16), and (10.14) that

$$\lim_{i\to\infty}\operatorname{Log}\beta_i = -\infty \; . \tag{10.25}$$

This proves the existence of k. Let e be the smallest such k. Assume that $\operatorname{Log}\lambda_e \le 0$. Then $\operatorname{Log}\lambda_{e-1} = \operatorname{Log}\lambda_e - \operatorname{Log}\beta_e \le s_e$ by (10.18). This contradicts the minimality of e. □

Proof (of Proposition 10.2.1). Choose a positive $\lambda \in F$ with

$$\mathfrak{b} = \lambda\mathfrak{a} \tag{10.26}$$

and

$$0 < \operatorname{Log}\lambda \le R \; . \tag{10.27}$$

Define

$$\lambda_k = \lambda\prod_{i=1}^{k}\beta_i \; , \quad k \ge 0 \; . \tag{10.28}$$

Then

$$\mathfrak{b}_k = \lambda_k\mathfrak{a} \; , \quad k \ge 0 \; . \tag{10.29}$$

Proof of 1. Let e be as in Lemma 10.2.2. Then Lemma 10.2.2 and (10.27) imply $0 < \operatorname{Log} \lambda_e \leq R$.

Let f be the smallest positive index such that $\mathfrak{b}_e = \mathfrak{a}_f$ which exists since $\mathfrak{b}_e$ is reduced and sits on the cycle $\{\mathfrak{a}_i\}$ of reduced ideals in the class of $\mathfrak{a}$. Then $0 < \operatorname{Log} \alpha_f \leq R$. It follows that $\operatorname{Log} \alpha_f = \operatorname{Log} \lambda_e \leq \operatorname{Log} \alpha_{2(e+L+1)}$. Hence we also have $0 < f \leq 2(e + L + 1)$.

This concludes the proof of 1.

Proof of 2. Let (e, f) be the lexicographically smallest pair that satisfies (10.19). Let e' be the value of e in Lemma 10.2.2. Then the proof of 1. shows that

$$e \leq e' \tag{10.30}$$

and we have

$$0 < \operatorname{Log} \alpha_f = \operatorname{Log} \lambda_e + kR \tag{10.31}$$

for some integer k. Now $\operatorname{Log} \lambda_e = \operatorname{Log} \lambda + \sum_{i=1}^{e} \operatorname{Log} \beta_i \leq R$ since $\operatorname{Log} \lambda \leq R$ and $\operatorname{Log} \beta_i < 0$ by (10.18) and (10.15). Hence, (10.31) implies $k \geq 0$ and $\operatorname{Log} \alpha_f \geq \operatorname{Log} \lambda_e$. Since $f \leq 2(e + L + 1)$, it follows that $s_{e+1} = \operatorname{Log} \alpha_{2(e+L+1)} \geq \operatorname{Log} \lambda_e$. This implies $e \geq e'$. Together with (10.30) we have $e = e'$ and $\operatorname{Log} \lambda_e > 0$ by Lemma 10.2.2. The minimality of f implies $k = 0$. Hence $\operatorname{Log} \lambda = \operatorname{Log} \alpha_f - \sum_{i=1}^{e} \operatorname{Log} \beta_i$. Indeed, Exercise 10.4.5 implies $\lambda = \alpha_f / \prod_{i=1}^{e} \beta_i$ so that λ chosen at the beginning of the proof coincides with the one defined in the Proposition and (10.21) holds.

We prove the upper bound on e. Set $E = \lceil E_\Delta \rceil$. Then

$$\begin{aligned}
\log \lambda_E &= \operatorname{Log} \lambda + \sum_{i=1}^{E} \operatorname{Log} \beta_i \\
&\leq \operatorname{Log} \lambda - \sum_{i=1}^{E} (\operatorname{Log} \alpha_{2(L+i)} - \frac{1}{2} \log \Delta) \\
&\leq \operatorname{Log} \lambda + E/2 \log \Delta - \frac{E(E+1)}{2} \log 2 \ .
\end{aligned}$$

Now we have $E(E + 1)/2 \cdot \log 2 > E^2/2 \cdot \log 2 > E/2 \cdot \log \Delta + R$. Hence, $\lambda_E < 0$. Since $\operatorname{Log} \lambda_e > 0$ by Lemma 10.2.2 it follows that $e < E$.

Proof of 3. Let $\mathfrak{a} = \mathfrak{b} = \mathcal{O}$. Let λ be the number defined in 2. Then $\mathcal{O} = \lambda\mathcal{O}$. So λ is a unit in $\mathcal{O}$. Also $0 < \operatorname{Log} \lambda \leq R$. Finally, all factors of λ given in (10.24) are positive, hence so is λ. Proposition 10.1.2 then says that $1/\lambda$ is the fundamental unit, as desired. □

10.2.2 Auxiliary algorithms

In order to be able to construct the giant-step sequence we need a few auxiliary algorithms.

Algorithm `reduce`($\mathfrak{a}$) from Section 9.1.3 receives as input an $\mathcal{O}$-ideal $\mathfrak{a}$ and returns a pair $(\mathfrak{c}, \gamma)$ where $\mathfrak{c}$ is a reduced $\mathcal{O}$-ideal and $\gamma \in F$ with $\mathfrak{c} = \gamma\mathfrak{a}$.

Algorithm 10.1 `close`$(\mathfrak{a})$

Input: An $\mathcal{O}$-ideal $\mathfrak{a}$.
Output: A pair $(\mathfrak{c}, \gamma)$, where $\mathfrak{c}$ is a reduced $\mathcal{O}$-ideal, $\gamma \in F$, $\mathfrak{c} = \gamma\mathfrak{a}$, and $0 \leq \operatorname{Log}\gamma < \frac{1}{2}\log D$.

```
$(\mathfrak{c}, \gamma) \leftarrow \texttt{reduce}(\mathfrak{a})$
if $\operatorname{Log}\gamma < 0$ then
   while $\operatorname{Log}\gamma < 0$ do
      $(\mathfrak{c}, \gamma) \leftarrow (\rho(\mathfrak{c}), \gamma \cdot \gamma(\mathfrak{c}))$.
else
   while $\operatorname{Log}\gamma\gamma'(\mathfrak{c}) \geq 0$ do
      $(\mathfrak{c}, \gamma) \leftarrow (\rho^{-1}(\mathfrak{c}), \gamma \cdot \gamma'(\mathfrak{c}))$.
```

Next, we explain algorithm `close`$(\mathfrak{a})$. Input is the $\mathcal{O}$-ideal $\mathfrak{a}$. Output is a pair $(\mathfrak{c}, \gamma)$ such that $\mathfrak{c}$ is a reduced $\mathcal{O}$-ideal, $\gamma \in F$, $\mathfrak{c} = \gamma\mathfrak{a}$, and $0 \leq \operatorname{Log}\gamma < \frac{1}{2}\log\Delta$. This algorithm works as follows. First, the algorithm calculates

$$(\mathfrak{c}, \gamma) \leftarrow \texttt{reduce}(\mathfrak{a}) \ .$$

If $\operatorname{Log}\gamma < 0$, then $\mathfrak{c}$ is replaced by $\rho(\mathfrak{c})$ and γ by $\gamma \cdot \gamma(\mathfrak{c})$ with $\gamma(\mathfrak{a})$ from (9.2) until for the first time $\operatorname{Log}\gamma \geq 0$ is true. If $\operatorname{Log}\gamma \geq 0$, then $\mathfrak{c}$ is replaced by $\rho^{-1}(\mathfrak{c})$ and γ is replaced by $\gamma \cdot \gamma'(\mathfrak{c})$ with $\gamma'(\mathfrak{c})$ from (10.5) until for the last time $\operatorname{Log}\gamma \geq 0$. We prove the following bound on $\operatorname{Log}\gamma$.

Lemma 10.2.3. *If γ is constructed as described, then we have $0 \leq \operatorname{Log}\gamma < \frac{1}{2}\log\Delta$.*

Proof. Assume that the call `reduce`$(\mathfrak{a})$ returns $(\mathfrak{c}_0, \gamma_0)$.

First, let $\operatorname{Log}\gamma_0 < 0$. Then our algorithm calculates

$$\mathfrak{c}_i = \rho(\mathfrak{c}_{i-1}), \gamma_i = \gamma_{i-1}\gamma(\mathfrak{c}_{i-1}) \ , \quad i \geq 1$$

until the first γ_i is found with $\operatorname{Log}\gamma_i \geq 0$. That γ_i is γ. Now we have $0 \leq \operatorname{Log}\gamma_i = \operatorname{Log}\gamma_{i-1} + \operatorname{Log}\gamma(\mathfrak{c}_{i-1})$. Since $\operatorname{Log}\gamma_{i-1} < 0$ and $0 < \operatorname{Log}\gamma(\mathfrak{c}_{i-1}) < \frac{1}{2}\log\Delta$ by Lemma 10.1.6, it follows that $\operatorname{Log}\gamma < \frac{1}{2}\log\Delta$. The case $\operatorname{Log}\gamma_0 \geq 0$ is treated analogously. □

We explain how we can decide whether $\operatorname{Log}\gamma < 0$. We are given γ as

$$\gamma = \frac{x + y\sqrt{\Delta}}{2z}$$

with $x, y, z \in \mathbb{Z}$, $z > 0$. So

$$\operatorname{Log}\gamma = \frac{1}{2}\log\left|\frac{x - y\sqrt{\Delta}}{x + y\sqrt{\Delta}}\right| \ .$$

Hence, $\operatorname{Log}\gamma \geq 0$ if and only if

$$\left| \frac{x - y\sqrt{\Delta}}{x + y\sqrt{\Delta}} \right| \geq 1 \; . \tag{10.32}$$

A calculation with integers only decides whether or not (10.32) holds.

We analyze Algorithm `close` in a special situation that is used later. We want to bound the run-time of `close` and the size required to store its output. For the latter we define the *size* of a quadratic number α in standard representation $(x + y\sqrt{\Delta})/(2d)$ to be the sum of the sizes of x, y, d, and Δ, and denote it by $\operatorname{size}(\alpha)$.

Lemma 10.2.4. *If the input ideal* $\mathfrak{a}$ *in Algorithm* `close` *is the quotient of two reduced* $\mathcal{O}$*-ideals, then the running time of Algorithm* `close` *is* $\mathrm{O}((\log \Delta)^3)$ *and* $\operatorname{size}(\gamma) = \mathrm{O}((\log \Delta)^2)$.

Proof. It follows from Lemma 6.2.7 and Proposition 7.3.17 that $f_{\mathfrak{a}} = (a, b, c)$ has coefficient $a < \Delta$. Hence it follows from Theorem 6.6.4 and Lemma 6.6.3 that the reduction algorithm `reduce` takes time $\mathrm{O}((\log \Delta)^2)$.

Each single reduction in the while loop likewise takes time $\mathrm{O}((\log \Delta)^2)$. Lemmas 10.1.8 and 10.1.6 imply that the while loops in `close` are traversed no more than $2 \log_2 \Delta$ times. This proves the run-time bound.

Also, by (9.5) we can write the reducing number $\alpha(\mathfrak{a}) = (x + y\sqrt{\Delta})/z$ with integers x, y, z, $z > 0$ and $|x|, |y|, z = \mathrm{O}(\Delta^2)$. Each reducing number computed in the while loop likewise requires space in $\mathrm{O}(\log \Delta)$. Together with the previously given bound on the number of iterations in the loops, this yields the space bound. □

Note that Lenstra states in [Len82] that the number of iterations required in `close` is actually 0, 1 or 2. This implies quadratic run-time and linear space bounds for `close` applied to ideals of norm in $\mathrm{O}(\Delta)$.

10.2.3 Construction of the giant-steps

We explain how a giant-step is computed. We let $\mathfrak{b}_0 = \mathfrak{b}$. For some $e \geq 0$ we are given the giant-step ideal $\mathfrak{b}_e$ and the baby-step ideal $\mathfrak{a}_{2(L+e+1)}$. We compute

$$(\mathfrak{b}_{e+1}, \delta_{e+1}) \leftarrow \texttt{close}(\mathfrak{b}_e \mathfrak{a}_{2(L+e+1)}^{-1}) \; .$$

Then

$$\beta_{e+1} = \delta_{e+1} / \alpha_{2(L+e+1)} \; .$$

That number is not stored explicitly. This is too space consuming. It suffices to store δ_{e+1}.

Lemma 10.2.5. *We have* $s_{e+1} - \frac{1}{2} \log \delta < -\operatorname{Log} \beta_{e+1} \leq s_{e+1}$.

Proof. By construction we have $-\operatorname{Log} \beta_{e+1} = s_{e+1} - \operatorname{Log} \delta_{e+1}$. Since $-\frac{1}{2} \log \Delta < -\operatorname{Log} \delta_{e+1} \leq 0$ by Lemma 10.2.3, we have $s_{e+1} - \frac{1}{2} \log \Delta < \operatorname{Log} \beta_{e+1} \leq s_{e+1}$ as asserted. □

10.2.4 The complete algorithm

We present the algorithms for computing the fundamental unit of $\mathcal{O}$ and for deciding equivalence between two $\mathcal{O}$-ideals. In those algorithms the baby-step ideals $\mathfrak{a}_i$, $i \geq 1$, are used. They are stored in a hash table. Then deciding whether a given reduced $\mathcal{O}$-ideal is in that table takes time $\mathrm{O}(\log \Delta)$.

Each algorithm consists of a set-up stage in which the baby-step table is filled with a few ideals starting from $\rho(\mathcal{O})$, or $\rho(\mathfrak{a})$, respectively. This ensures that the following giant-steps have negative width. If in this set-up stage the target ideal ($\mathcal{O}$ or $\mathfrak{b}$) is found, then we can stop the algorithm immediately.

In the main loop, both algorithms compute one giant and two baby steps per iteration. After each giant-step, a table look-up in the baby-step table is performed. If this look-up was successful, then the algorithm terminates. The output allows the computation of the fundamental unit, or a relative generator using (10.24).

Note that `TerrEquivalent` presumes that an approximation to the regulator has been computed in advance. This is only used to obtain an integer close to and larger than E_Δ. (The rounding in the calculation of E need not be exact.) The pre-computation of R can be avoided by computing two giant step sequences, one beginning at $\mathfrak{b}$ and one at $\mathfrak{a}$, and terminating with result `nil` when there is a match of the second one with the baby step table.

Details of `TerrUnit` and `TerrEquivalent`, as well as a detailed description of the indicated variation can be found in [Vol03].

Example 10.2.6. Table 10.1 on page 230 lists the ideals computed in the course of the execution of `TerrUnit` for $\Delta = 2521$. Ideals in standard representation $a\mathbb{Z}+\mathbb{Z}(b+\sqrt{\Delta})/2$ are listed as (a, b). Distances given are distances to the unit ideal (1). We set $\omega = (1+\sqrt{\Delta})/2$. Finally, note that $1/2 \cdot \log \Delta \approx 3.916$.

The table shows that $\mathfrak{a}_{11} = \mathfrak{b}_6$ (connected by an arrow). Using (10.24) the data in the table yields the fundamental unit. We obtain $R \approx 85.768$.

10.2.5 Analysis of the algorithm

Proposition 10.2.7. *Algorithms* `TerrUnit` *and* `TerrEquivalent` *both require time* $\mathrm{O}((\log \Delta + \sqrt{R})(\log \Delta)^3)$ *and space* $\mathrm{O}((\log \Delta + \sqrt{R})(\log \Delta)^2)$.

Proof. It follows from (10.13) that $L = \mathrm{O}(\log \Delta)$. Also, it follows from Proposition 10.2.1 that Algorithms `TerrUnit` and `TerrEquivalent` both terminate with the correct output and $e = \mathrm{O}(\log \Delta + \sqrt{R})$.

In each iteration of the precomputation, the algorithms apply the reduction operator twice to reduced $\mathcal{O}$-ideals. That application has running time $\mathrm{O}((\log \Delta)^2)$. In each iteration of the main loop, both algorithms apply `close` once to the quotient of two reduced $\mathcal{O}$-ideals and the operator ρ at most twice

Algorithm 10.2 $\texttt{TerrUnit}(\mathcal{O})$

Input: The order $\mathcal{O}$

Output: $\mathfrak{a}_1, \dots, \mathfrak{a}_{2(e+L)}$, $\delta_1, \dots, \delta_e$, f such that λ from (10.24) is the fundamental unit of $\mathcal{O}$

```
$\mathfrak{a}_0 \leftarrow \mathcal{O}$.
$L \leftarrow -1$
repeat
   $L \leftarrow L + 1$
   $\mathfrak{a}_{2L+1} \leftarrow \rho(\mathfrak{a}_{2L})$, $\mathfrak{a}_{2L+2} \leftarrow \rho(\mathfrak{a}_{2L+1})$
until $\operatorname{Log} \alpha_{2L+2} > \frac{1}{2} \log \Delta$
if $\mathcal{O} = \mathfrak{a}_f$ for some $1 < f \leq 2(L+1)$ then
   return $\mathfrak{a}_1, \dots, \mathfrak{a}_f, f$
$e \leftarrow 0$
$\mathfrak{b}_0 \leftarrow \mathcal{O}$
loop
   $(\mathfrak{b}_{e+1}, \delta_{e+1}) \leftarrow \texttt{close}(\mathfrak{b}_e \mathfrak{a}_{2(L+e+1)}^{-1})$
   $e \leftarrow e + 1$
   if $\mathfrak{b}_e = \mathfrak{a}_f$ for some $f \leq 2(L+e)$ then
      return $\mathfrak{a}_1, \dots, \mathfrak{a}_{2(e+L)}, \delta_1, \dots, \delta_e, f$
   $\mathfrak{a}_{2L+2e+1} \leftarrow \rho(\mathfrak{a}_{2L+2e})$, $\mathfrak{a}_{2L+2e+2} \leftarrow \rho(\mathfrak{a}_{2L+2e+1})$
```

Algorithm 10.3 $\texttt{TerrEquivalent}(\mathfrak{a}, \mathfrak{b})$

Input: Reduced $\mathcal{O}$-ideals $\mathfrak{a}$ and $\mathfrak{b}$, regulator bound $R > R_\Delta$.

Output: If $\mathfrak{a}$ and $\mathfrak{b}$ are inequivalent, the algorithm returns `nil`. Else, the algorithm returns $\mathfrak{a}_1, \dots, \mathfrak{a}_{2(e+L)}$, $\delta_1, \dots, \delta_e$, f such that for λ as in (10.24) we have $\mathfrak{b} = \lambda \mathfrak{a}$ and $0 \leq \operatorname{Log} \lambda < R$.

```
$E \leftarrow \lceil \log_2 \Delta + \sqrt{2R/\log 2} \rceil$.
$\mathfrak{a}_0 \leftarrow \mathfrak{a}$.
$L \leftarrow -1$
repeat
   $L \leftarrow L + 1$
   $\mathfrak{a}_{2L+1} \leftarrow \rho(\mathfrak{a}_{2L})$, $\mathfrak{a}_{2L+2} \leftarrow \rho(\mathfrak{a}_{2L+1})$
until $\operatorname{Log} \alpha_{2L+2} > \frac{1}{2} \log \Delta$
if $\mathfrak{b} = \mathfrak{a}_f$ for some $0 \leq f \leq 2(L+1)$ then
   return $\mathfrak{a}_1, \dots, \mathfrak{a}_f, f$
$e \leftarrow 0$
$(\mathfrak{b}_0, \delta_0) \leftarrow (\mathfrak{b}, 1)$
while $e + 1 < E$ do
   $(\mathfrak{b}_{e+1}, \delta_{e+1}) \leftarrow \texttt{close}(\mathfrak{b}_e \mathfrak{a}_{2(L+e+1)}^{-1})$
   $e \leftarrow e + 1$
   if $\mathfrak{b}_e = \mathfrak{a}_f$ for some $f \leq 2(L+e)$ then
      return $\mathfrak{a}_1, \dots, \mathfrak{a}_{2(e+L)}, \delta_1, \dots, \delta_e, f$
   $\mathfrak{a}_{2L+2e+1} \leftarrow \rho(\mathfrak{a}_{2L+2e})$, $\mathfrak{a}_{2L+2e+2} \leftarrow \rho(\mathfrak{a}_{2L+2e+1})$
return nil
```

L	e	i	baby step	distance	giant step	distance	δ_e
			ideals $\mathfrak{a}_i$		ideals $\mathfrak{b}_e$		
0		1	(30,11)	2.203			
		2	(20,29)	2.426			
1		3	(21,13)	3.085			
		4	(28,43)	3.350			
2		5	(6,41)	4.630			
		6	(35,29)	5.776			
	1	7	(12,43)	6.435	(35,41)	-5.776	35
		8	(14,41)	7.715			
	2	9	(15,49)	8.861	(10,31)	-13.491	2
		10	(2,47)	11.065			
	3	11	(39,31)	12.770	(5,41)	-24.555	1
		12	(10,49)	13.491			
	4	13	(3,47)	15.694	(12,37)	-37.833	$1-\omega/5$
		14	(26, 5)	17.400			
	5	15	(24,43)	17.500	(9,35)	-54.379	$6-\omega/6$
		16	(7,41)	18.779			
	6				(39,31)	-72.998	$(-7+2\omega)/9$

Table 10.1. Baby and giant step ideals computed by `TerrUnit` for $\Delta = 2521$

each time with reduced $\mathcal{O}$-ideals as arguments. By Lemma 10.2.5 each application of `close` takes time $\mathrm{O}((\log\Delta)^3)$. Also, each application of ρ has running time $\mathrm{O}((\log\Delta)^2)$. This proves the running time estimate.

Both algorithms store $\mathrm{O}(\log\Delta + \sqrt{R})$ reduced $\mathcal{O}$-ideals and numbers δ_i. The size of a reduced $\mathcal{O}$-ideal is $\mathrm{O}(\log\Delta)$, and by Lemma 10.2.5 each δ_i requires space $\mathrm{O}((\log\Delta)^2)$. This implies the space estimate. □

10.3 Further applications

We may use computations in the infrastructure of a real quadratic field not only for the computation of its regulator, but also for the computation of its class group and for the factorization of its discriminant.

Class Group. In Section 9.7.1 we have explained how to compute the structure of a finite Abelian group from generators and relations. However, that algorithm and its improvements described in Section 9.7.3 cannot be immediately used in the context of real quadratic class groups since deciding equality of real quadratic ideal classes is more difficult. Nevertheless, the algorithms for deciding equivalence from this chapter can be used to construct a deterministic class group algorithm for real quadratic orders that runs in time $\Delta^{1/4+o(1)}$.

Factorization. Dan Shanks proposed an algorithm which utilizes the infrastructure to compute a factorization of a given number N. For reasons which become clear below he called it SQUFOF (*Squ*are *Fo*rm *F*actorization).

SQUFOF aims to compute an ambiguous ideal in a ring $\mathcal{O}$ whose discriminant Δ is a small multiple of N. For a primitive ambiguous ideal $\mathfrak{a}$ we have $\gcd(\mathrm{N}\mathfrak{a}, \Delta) > 0$ unless it is the unit ideal. This is clear if it has a nontrivial common divisor with the conductor of Δ. If it does not, then it is invertible and its factorization contains only prime ideals with norm dividing Δ, cf. Propositions 8.6.4 and 8.6.11. Note that the algorithm fails if the gcd computed equals Δ.

The idea is to walk along the principal cycle until an ideal $\mathfrak{c} = \rho^{2k}(\mathcal{O})$ is found which has square norm. If $\mathfrak{c}$ has norm coprime to the conductor of Δ, then there exists $\mathfrak{b}$ with $\mathfrak{b}^2 = \mathfrak{c}$, cf. Proposition 8.6.11. Let γ be a generator of $\mathfrak{c}$. Set $\beta = \gamma + \mathrm{N}\mathfrak{b}$, and $\mathfrak{a} = \beta^{-1}\mathfrak{b}$. Then

$$\mathfrak{a}^2 = \beta^{-2}\mathfrak{b}^2 = \frac{\gamma}{\beta^2}\mathcal{O} = \frac{\mathrm{N}\mathfrak{b}}{\mathrm{N}(\beta)}\mathcal{O} .$$

Hence $\mathfrak{a}$ is ambiguous. It differs from the unit ideal unless $\mathfrak{b}$ has been encountered on the cycle before $\mathfrak{c}$.

Unfortunately, it is possible that no square ideal is found on the principal cycle. This is especially likely if the regulator of the used order is small. Even if one assumes the regulator to be at the order of magnitude of $\sqrt{\Delta}$, there is no proven bound for the number of ideals on the principal cycle which need to be enumerated before a square ideal is found. It is, however, plausible and confirmed by experiments that $O(\sqrt[4]{\Delta})$ reductions suffice to find a square ideal.

Shanks used the language of forms to formulate his algorithm. He introduced for the first time a notion of distance between ideals on a cycle. The distance between ideals we introduced in section 10.1.1 is indeed only a slight modification of Shanks' distance function δ due to Lenstra [Len82]. On the principal cycle of forms with discriminant Δ, SQUFOF seeks to find a reduced form f which is a square of another, say g. It then computes a form h for which $2\delta(h, g) = \delta(1, f)$ where 1 denotes the unit form. The form h which corresponds to $\mathfrak{a}$ constructed above is ambiguous, and yields a (possibly trivial) factorization of Δ, cf. section 1.4.3.

10.4 Exercises

Exercise 10.4.1. Prove Proposition 10.1.2.

Exercise 10.4.2. Prove that the map in (10.3) is well defined.

Exercise 10.4.3. Prove Proposition 9.7.16 for the case $\Delta > 0$.

Exercise 10.4.4. Prove Theorem 9.2.3.

Exercise 10.4.5. Let $\mathcal{O}$ be a real quadratic order with field of fractions F and regulator R. Let $\mathfrak{a}$ and $\mathfrak{b}$ be two equivalent fractional $\mathcal{O}$-ideals. Prove that there is exactly one $\lambda \in F$ with $\mathfrak{b} = \lambda\mathfrak{a}$, $\lambda > 0$, and $0 < \operatorname{Log} \lambda \leq R$.

Exercise 10.4.6. Use the notation of Proposition 10.2.1. Let e be minimal such that (10.19) holds for some f and assume that $e > 1$. Prove that f is uniquely determined.

Exercise 10.4.7. Prove (10.6).

Exercise 10.4.8. Prove that the sequence $(\mathfrak{a}_i)_{i\in\mathbb{Z}}$ from Section 10.1.2 is cyclic of finite period length. Also prove that the period contains all reduced $\mathcal{O}$-ideals that are equivalent to $\mathfrak{a}$.

Exercise 10.4.9. Develop an algorithm that calculates the class group of a real quadratic order of discriminant Δ in time $\Delta^{1/4+o(1)}$.

Exercise 10.4.10. Prove that the bound for $\operatorname{Log} \gamma(\mathfrak{a})$ in Lemma 10.1.6 can be sharpened to $1/2 \log \Delta - \log 2$.

Chapter references and further reading

[Len82] Hendrik W. Lenstra, Jr., *On the calculation of regulators and class numbers of quadratic fields*, Journées Arithmétiques 1980 (Cambridge) (J. V. Armitage, ed.), London Mathematical Society Lecture Note Series, vol. 56, Cambridge University Press, 1982, pp. 123–150.

[Sch82] René Schoof, *Quadratic fields and factorization*, Computational methods in number theory (Hendrik W. Lenstra, Jr. and Robert Tijdeman, eds.), Mathematical Centre Tracts, vol. 154–155, Mathematisch Centrum, 1982, `http://cr.yp.to/bib/1982/schoof.html`, pp. 235–286.

[Schar] ———, *Computing Arakelov class groups*, Surveys in algorithmic number theory (to appear), `http://www.mat.uniroma2.it/~schoof/infranew2.pdf`.

[Sha72] Daniel Shanks, *The infrastructure of real quadratic fields and its applications*, Proceedings of the 1972 Number Theory Conference, Boulder, Colorado, 1972, pp. 217–224.

[Ter00] David C. Terr, *A modification of Shanks' baby-step giant-step algorithm*, Mathematics of Computation **69** (2000), no. 230, 767–773.

[Vol03] Ulrich Vollmer, *Rigorously analyzed algorithms for the discrete logarithm problem in quadratic number fields*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, 2003.

11

Subexponential Algorithms

In this chapter we describe probabilistic algorithms for computing class numbers and regulators. We also show how they can be extended to solve the equivalence problem. Those algorithms are much faster than the deterministic algorithms presented in Chapter 9. They use an approach dubbed *index calculus* which originated in work by Kraichik [Kra22] and seemingly independent work by Western and Miller [WM68]. The first proposals to apply this approach in the context of imaginary quadratic class groups came from Seysen [Sey87] and Hafner/McCurley [HM89]. In this chapter we let Δ be a fundamental discriminant. It is rather straightforward, however, to extend the algorithms of this chapter to the case of nonfundamental discriminants.

11.1 The function $L_x[a, b]$

For real numbers a, b, x with $x > e$ where e is Euler's constant, we define

$$L_x[a,b] = \exp(b(\log x)^a(\log\log x)^{(1-a)}) \ . \tag{11.1}$$

This function is used to describe the running time of algorithms. We explain its meaning. We have

$$L_x[0,b] = \exp(b(\log x)^0(\log\log x)^1) = (\log x)^b \tag{11.2}$$

and

$$L_x[1,b] = \exp(b(\log x)^1(\log\log x)^0) = x^b \ . \tag{11.3}$$

Consider an algorithm that computes the class number h_Δ. It receives as input Δ and returns h_Δ. The binary length of Δ is $\lfloor \log_2 |\Delta| \rfloor + 1$. If our algorithm has running time $L_{|\Delta|}[0, b]$, then it is a polynomial time algorithm; its complexity is bounded by a polynomial in the size of the input. The algorithm is considered efficient, although its real efficiency depends on the degree b of the polynomial. If the algorithm has running time $L_{|\Delta|}[1, b]$, then it is exponential; its complexity is bounded by an exponential function in the length of the input. The algorithm is considered inefficient. All class number algorithms, that we have seen so far, are exponential. If the algorithm has running time $L_{|\Delta|}[a, b]$ with $0 < a < 1$, then it is called *subexponential.* The algorithm is slower than polynomial but faster than exponential. In this section we will present probabilistic algorithms for the major computational problems such as the computation of regulators and class numbers that under the assumption of a certain Riemann hypothesis have expected running time bounded by $L_{|\Delta|}[1/2, b + o(1)]$ with some $b > 0$ and a function $o(1)$ that goes to zero as Δ goes to infinity. They are subexponential.

By $o(1)$ we always mean a function that can be evaluated in polynomial time, maps discriminants Δ to real numbers, and approaches 0 as $|\Delta|$ goes to infinity.

11.2 Preliminaries

The following two easy lemmas from elementary calculus will help us throughout in bounding the number of calls to probabilistic algorithms which might fail.

Lemma 11.2.1. *For all $x \geq 2$ we have*

$$(1 - 1/x)^x \geq 1/4 \, .$$

Proof. The function $f(x) = (1 - 1/x)^x$ increases monotonously with growing x (to the limit $1/e$). Indeed, consider the derivative $f'(x) = f(x)(\log(1 - 1/x) + 1/(x - 1))$. Set $g(x) = \log(1 - 1/x) + 1/(x - 1)$. This is positive at $x = 2$, tends to zero with growing x, and decreases monotonously itself since $g'(x) = -x^{-1}(x - 1)^{-2}$. It follows that $f'(x)$ is positive on $[2, \infty]$. □

Lemma 11.2.2. *Let $f \in \mathbb{Z}_{>1}$, $p \in \mathbb{R}$, $0 < p \leq 1$, and $l \in \mathbb{N}$ with*

$$pl > \log f \, . \tag{11.4}$$

Then

$$\left(1 - (1 - p)^l\right)^f \geq \frac{1}{4} \, . \tag{11.5}$$

Proof. The statement is trivial for $p = 1$. Assume $p < 1$. By Lemma 9.3.15 assumption (11.4) implies

$$2 \leq f < (1 - p)^{-l} \, .$$

Hence

$$\left(1 - (1 - p)^l\right)^f > \left(1 - (1 - p)^l\right)^{(1-p)^{-l}} \, .$$

An application of Lemma 11.2.1 yields the assertion. □

11.3 The factor base

Our algorithm seeks to compute the relations lattice of a large set generating the class group. These generators are given by representatives from a large set $\mathcal{F}$ of ideals. The algorithm obtains individual relations by finding two exponent vectors $\mathbf{v}$ and $\mathbf{w}$ such that $[\mathcal{F}]^{\mathbf{v}} = [\mathcal{F}]^{\mathbf{w}}$. The first vector, $\mathbf{v}$, is chosen randomly so that $[\mathcal{F}]^{\mathbf{v}}$ is a random element of the class group in a sense to be made precise later. The second vector $\mathbf{w}$ is obtained by finding a reduced representative of $[\mathcal{F}]^{\mathbf{v}}$ and factor it over $\mathcal{F}$ using its factorization into a power product of prime ideals. This is successful if all prime ideals occuring in the factorization actually are elements of $\mathcal{F}$. Thus, depending on the proportion of prime ideals in $\mathcal{F}$ within the set of all prime ideals occuring in the factorization of reduced ideals, the process may have to be repeated many times before it will yield a relation.

Since $\mathcal{F}$ needs to contain the prime ideals which the found reduced ideals are factored into, it is called a *factor base*. We will choose our factor base to contain the set $\mathcal{F}_z$ of all prime ideals $\mathfrak{p}(\Delta, p)$ and their conjugates for which p is a prime number with $\left(\frac{\Delta}{p}\right) = 1$ and

$$p \le L_{|\Delta|}[\frac{1}{2}, z]$$

for some positive real number

$$z \le 1$$

that is specified later. Depending on the problem we will want to solve, $\mathcal{F}$ will just contain the specified prime ideals, or one or two ideals in addition to them. Denote by f the cardinality of the factor base. By the prime number theorem there will be aproximately $L_{|\Delta|}[1/2, z]/(z\sqrt{\log|\Delta| \log\log|\Delta|})$ prime ideals in $\mathcal{F}$. Thus, in all cases f will be in $L_{|\Delta|}[1/2, z + o(1)]$.

We order $\mathcal{F}$ in some way and write

$$\mathcal{F} = (\mathfrak{p}_1, \ldots, \mathfrak{p}_f) . \tag{11.6}$$

All but at most two of the ideals $\mathfrak{p}_i$ are prime.

We need $[\mathcal{F}]$ to generate the class group. We will assume that the ERH holds and apply Proposition 9.5.3. This proposition says that $[\mathcal{F}]$ will generate the class group if $|\Delta|$ is sufficiently large. Comparing $c_3(\Delta)$ from (9.20) with $L_{|\Delta|}[1/2, z]$, we see that this is certainly the case if $\Delta < -157$ or $\Delta > 41$ which we will subsequently assume. We call an $\mathcal{O}$-ideal $\mathcal{F}$-smooth if all the factors in the prime ideal factorization of $\mathfrak{a}$ are in $\mathcal{F}$.

Algorithm `factorBase` shown on the following page computes the prime ideals the factor base is to contain.

Algorithm 11.1 `factorBase` (Δ, z)

Input: The discriminant Δ, the parameter z.
Output: The factor base $\mathcal{F}_z$ or `nil`.

```
Set $L \leftarrow \lceil L_{|\Delta|}[1/2, z] \rceil$, $k \leftarrow \lceil \log L \rceil$
Set $\mathcal{F} \leftarrow \emptyset$
for all primes $p < L$ do
  if kronecker$(\Delta, p) = 1$ then
    $i \leftarrow 0$
    repeat
      $i \leftarrow i + 1$
      $g \leftarrow$ primeForm$(\Delta, p)$
    until $i > 2k$ or $g \neq$ nil
    if $f \neq$ nil then
      $\mathcal{F} \leftarrow \mathcal{F} \cup \{L(g), L(g)^\sigma\}$
    else
      Return nil
Return $\mathcal{F}$
```

Lemma 11.3.1. *Algorithm* `factorBase` *succeeds with probability exceeding* $1/4$.

Proof. Corollary 3.4.26 implies that Algorithm `factorBase` succeeds with probability larger than

$$(1 - 2^{-k})^L \ .$$

This probability is by Lemma 11.2.2 larger than $1/4$ since $2k > \log L$. □

Lemma 11.3.2. *A call of* `factorBase` *takes time* $L_{|\Delta|}[1/2, z + o(1)]$.

Proof. Algorithm `factorBase` calls `primeForm` fewer than $2kL$ times. According to Corollary 3.4.26, each call to `primeForm` takes time $O(\log|\Delta| \cdot \log L + (\log L)^4)$. □

Finally, we prove a very loose bound for the cardinality of $\mathcal{F}$ which will be useful later on. It is exponential whereas we already know f to be subexponential.

Lemma 11.3.3. *If* $z \leq 1$, *then* $f < 8|\Delta|/h_\Delta + 1$.

Proof. We use the Chebyshev bound for the number $\pi(x)$ of prime numbers below x (see e.g. [HW79], Chapter 22)

$$\pi(x) < 9/8 \cdot x / \log(x) \ .$$

Thus we have $f < 2 + 9/4L/\log(L)$ where $L = L_{|\Delta|}[1/2, z]$ is the norm bound for the prime ideals in our factor base. Theorem 9.3.11 yields lower bounds for $8|\Delta|/h_\Delta$ which an easy calculation shows to be larger than $1 + 9/4L/\log(L)$ if $\Delta < -4$ or $\Delta > 5$. In this calculation we use again the lower bound $R_\Delta > \log((1 + \sqrt{\Delta})/2)$ for $\Delta > 8$. For the exceptional discriminants $\Delta \in \{-4, -3, 5, 8\}$ the statement of the Corollary can be directly verified. □

11.4 The imaginary quadratic case

Let $\Delta < 0$.

Our algorithm calculates random relations in the relation lattice $L([\mathcal{F}])$ until a generating system for that lattice is found. As described in Proposition 9.7.3, Hermite and Smith normal form computation can then be used to determine the structure of the class group Cl.

11.4.1 Random relations

The calculation of the random relations is done by applying Algorithm `randomRelation`. That algorithm selects an exponent vector $\mathbf{v} \in \mathbb{Z}^f_{0..|\Delta|-1}$ randomly with the uniform distribution. It calculates the reduced ideal $\mathfrak{a}$ in the ideal class $[\mathcal{F}^{\mathbf{v}+\mathbf{w}}]$ where $\mathbf{w}$ is some fixed offset vector that is also input for `randomRelation`. If the reduced ideal happens to be $\mathcal{F}$-smooth, that is,

$$\mathfrak{a} = \mathcal{F}^{\mathbf{a}} \tag{11.7}$$

for some $\mathbf{a} \in \mathbb{Z}^f$, then

$$[\mathcal{F}^{\mathbf{v}+\mathbf{w}}] = [\mathfrak{a}] = [\mathcal{F}^{\mathbf{a}}] \,. \tag{11.8}$$

Hence, the relation

$$\mathbf{z} = \mathbf{a} - \mathbf{w} - \mathbf{v} \in L([\mathcal{F}]) \tag{11.9}$$

is found.

Here is the algorithm.

Algorithm 11.2 `randomRelation` $(\Delta, \mathcal{F}, \mathbf{w})$

Input: The discriminant Δ, the factor base $\mathcal{F}$ of length f, the offset vector $\mathbf{w} \in \mathbb{Z}^f$.
Output: A relation $\mathbf{z}$ for $[\mathcal{F}]$ or `nil`.

```
Select $\mathbf{v} \in \mathbb{Z}^f_{0..|\Delta|-1}$ uniformly at random.
Calculate the reduced ideal $\mathfrak{a}$ in $[\mathcal{F}^{\mathbf{v}+\mathbf{w}}]$.
if $\mathfrak{a} = \mathcal{F}^{\mathbf{a}}$ with $\mathbf{a} \in \mathbb{Z}^f$ then
   Return $\mathbf{z} = \mathbf{v} + \mathbf{w} - \mathbf{a}$.
else
   Return nil
```

We explain how Algorithm `randomRelation` decides whether the reduced ideal $\mathfrak{a}$ factors over the factor base $\mathcal{F}$. Let $\mathfrak{a} = L(a, b, c)$ with a reduced form (a, b, c). We know from Proposition 8.6.11 that $\mathfrak{a}$ factors over $\mathcal{F}$ if and only if a factors into the norms of the prime ideals in $\mathcal{F}$. Proposition 8.6.11 also tells

Algorithm 11.3 `reducePowerProduct` $(\Delta, \mathcal{F}, \mathbf{u})$

Input: The discriminant Δ, the power product base $\mathcal{F} = \{\mathfrak{p}_1, \dots, \mathfrak{p}_f\}$, the exponent vector $\mathbf{u}$.
Output: The reduced ideal in $[\mathcal{F}^{\mathbf{u}}]$.

```
Initialize $\mathfrak{r} \leftarrow (1)$.
for ($i \leftarrow 1, i \leq f, i \leftarrow i + 1$) do
   if $u_i = 0$ then
      Initialize $\mathfrak{q} \leftarrow (1)$.
   else
      Initialize $\mathfrak{q} \leftarrow \mathfrak{p}_i$.
   for ($j \leftarrow \max(j \mid u_{ij}) - 1, j \geq 0, j \leftarrow j - 1$) do
      $\mathfrak{q} \leftarrow$ reduce($\mathfrak{q}^2$).
      if $u_{ij} = 1$ then
         $\mathfrak{q} \leftarrow$ reduce($\mathfrak{q} \cdot \mathfrak{p}_i$).
   $\mathfrak{r} \leftarrow$ reduce($\mathfrak{r} \cdot \mathfrak{q}$).
return $\mathfrak{r}$.
```

us how to determine the factorization of $\mathfrak{a}$ from the prime factorization of a. We use trial division to determine the prime factorization of a.[1]

For the computation of the reduced ideal in $[\mathcal{F}^{\mathbf{u}}]$, the algorithm `randomRelation` employs a fast exponentiation technique. Write the entries of $\mathbf{u}$ in binary form

$$\mathbf{u} = (u_1, \dots, u_f)\ , \qquad u_i = \sum_{j=0}^{k} u_{ij} 2^j\ , \quad \text{for } i = 1, \dots, f\ .$$

Algorithm `reducePowerProduct` shown on the top of this page applies the so-called left-right binary exponentiation method to the class group of an imaginary quadratic order.

Lemma 11.4.1. *On inputs discriminant Δ, power product base $\mathcal{F} = \{\mathfrak{p}_1, \dots, \mathfrak{p}_f\}$ with reduced ideals $\mathfrak{p}_i$, $i = 1, \dots, f$, and exponent vector $\mathbf{u}$ $\log_2 |\mathbf{u}|_\infty \leq k$, algorithm* `reducePowerProduct` *executes in time $O(f(\log|\mathbf{u}|_\infty)(\log|\Delta|)^2)$.*

Proof. Note that in algorithm `reducePowerProduct` the function `reduce` is applied at most $2(\lfloor \log_2 |\mathbf{u}|_\infty \rfloor) \cdot f$ times, each time to the product of two reduced ideals. By Proposition 9.1.8, these products have norm at most $|\Delta|$. Thus, Lemma 11.4.1 follows from Theorem 5.6.6. □

We estimate the running time of `randomRelation`.

[1] For large factor bases it is faster to use a factorization algorithm on the basis of elliptic curve arithmetic. If large numbers of ideals are to be tested for smoothness at the same time, then it will be advantageous to use Bernstein's algorithm [Ber]

Lemma 11.4.2. *If the entries of the offset vector* $\mathbf{w}$ *are bounded in absolute value by a polynomial in* $|\Delta|$*, then a call of* `randomRelation` *takes time* $L_{|\Delta|}[1/2, z + o(1)]$.

Proof. We need to bound the time needed for the computation of the reduced ideal in the second step, and the time needed for the trial division in the third.

For the second step we apply Lemma 11.4.1. Here $f = L_{|\Delta|}1/2, z + o(1)$ and $k = O(\log|\Delta|)$. Thus, this step can be performed in time $L_{|\Delta|}1/2, z + o(1)$.

Since the numbers to be factored in the third step are smaller than $\sqrt{|\Delta|}$, trial division can be performed in time $O((\log|\Delta|)^2)$ for each divisor tried. We need to test at most f factors. Thus, also the third step of `randomRelation` can be executed in time $L_{|\Delta|}1/2, z + o(1)$. □

We analyze the success probability of `randomRelation`. `randomRelation` can only be successful if there are reduced $\mathcal{O}$-ideals that factor completely into prime ideals from $\mathcal{F}$. We prove a lower bound for the number of such ideals. For $y \in \mathbb{R}_{>0}$ we call a positive integer *y-smooth* if it has only prime factors $\leq y$. Here is a lower bound for the number of smooth ideals of bounded norm.

Proposition 11.4.3. *For any* $\varepsilon > 0$ *there is a positive real number* $c(\varepsilon)$ *such that for any* $x, y \in \mathbb{R}_{>0}$ *and any fundamental discriminant* Δ *with*

$$\max\{(\log x)^{1+\varepsilon}, (\log|\Delta|)^{2+\varepsilon}\} \leq y \leq \exp((\log x)^{1-\varepsilon}) ,$$

the number of primitive integral $\mathcal{O}_\Delta$*-ideals with* y*-smooth norm* $\leq x$ *is at least*

$$x \exp(-u(\log u + \log\log u + c(\varepsilon)))$$

where $u = (\log x)/(\log y)$.

Proof. For $\Delta < 0$ this is proved in [Sey87]. That proof also works for the case $\Delta > 0$. For $\Delta > 0$ the statement also follows from [BH96]. □

From Proposition 11.4.3 we deduce the lower bound that we are interested in. We use the fact that by Proposition 9.1.8 an $\mathcal{O}$-ideal with norm $\leq \sqrt{|\Delta|}/2$ is reduced.

Lemma 11.4.4. *The number of reduced* $\mathcal{O}_\Delta$*-ideals that have* $L[\frac{1}{2}, z]$*-smooth norm is at least* $\sqrt{|\Delta|}L_{|\Delta|}[\frac{1}{2}, -1/(4z) + o(1)]$.

Proof. For any $z > 0$ there are $\Delta(z) > 0$ and ε with $0 < \varepsilon < 1/4$ such that

$$(\log|\Delta|)^{2+\varepsilon} \leq L_{|\Delta|}[\frac{1}{2}, z] \leq \exp(\log(\sqrt{|\Delta|/2})^{1-\varepsilon})$$

for any fundamental discriminant Δ with $|\Delta| > \Delta(z)$. We apply Proposition 11.4.3 with $x = \sqrt{|\Delta|}/2$, $y = L_{|\Delta|}[\frac{1}{2}, z]$ and this ε. Now

$$u(\log u + \log\log u + c(\varepsilon)) = (\frac{1}{4z} - o(1))\sqrt{\log|\Delta|\log\log|\Delta|} . \qquad (11.10)$$

Since by Proposition 9.1.8 an $\mathcal{O}$-ideal with norm $\leq \sqrt{|\Delta|}/2$ is reduced, this proves the assertion. □

.

We also use the following results.

Lemma 11.4.5. *Let $l \in \mathbb{Z}_{>0}$, and let L be a lattice of full rank in $\mathbb{Z}^l$ with determinant d. Then for any $\mathbf{w} \in \mathbb{Z}^l$ and any $m \in \mathbb{Z}_{>0}$ we have*

$$\frac{1}{d}(1 - \frac{d-1}{m}) \leq \frac{|\mathbb{Z}^l_{0..m-1} \cap (\mathbf{w} + L)|}{|\mathbb{Z}^l_{0..m-1}|} \leq \frac{1}{d}(1 + \frac{d-1}{m}) \ .$$

Proof. We prove the lemma by induction. The statement is clear for $l = 1$. Indeed, there are no less than

$$\left\lfloor \frac{m}{d} \right\rfloor \geq \frac{m-d+1}{d}$$

elements in $\mathbb{Z}_{0..m-1} \cap (\mathbf{w} + L)$, and no more than

$$\left\lceil \frac{m}{d} \right\rceil \leq \frac{m+d-1}{d} \ .$$

Assume $l > 1$. Let π be the projection

$$\pi : \mathbb{Z}^l \longrightarrow \mathbb{Z} : (a_1, \ldots, a_l) \longmapsto a_l \ .$$

Let $L' = \pi(L)$, and $L'' = L \cap \ker \pi$. If $d' = \det L'$, and $d'' = \det L''$, then $d'd'' = d$. We know already

$$\frac{1}{d'}(1 - \frac{d'-1}{m}) \leq \frac{|\mathbb{Z}_{0..m-1} \cap \pi(\mathbf{w} + L)|}{|\mathbb{Z}_{0..m-1}|} \leq \frac{1}{d'}(1 + \frac{d'+1}{m}) \ .$$

We consider the intersection of $(\mathbf{w}+L) \cap \mathbb{Z}^l_{0..m-1}$ with the hyperplanes $\pi^{-1}(x)$ with x running through $\mathbb{Z}_{0..m-1} \cap \pi(\mathbf{w} + L)$. For each such x choose some $\mathbf{v}_x \in L$ such that $x = \pi(\mathbf{w} + \mathbf{v}_x)$. Then

$$\pi^{-1}(x) \cap \mathbb{Z}^l_{0..m-1} \cap (\mathbf{w} + L) = \mathbb{Z}^l_{0..m-1} \cap (\mathbf{w} + \mathbf{v}_x + L'') \ .$$

We apply the induction hypothesis to each of these sets. Since $\mathbb{Z}^l_{0..m-1} \cap (\mathbf{w}+L)$ is their union we obtain

$$\frac{1}{d'd''}(1 - \frac{d'-1}{m})(1 - \frac{d''-1}{m}) \leq \frac{|\mathbb{Z}^l_{0..m-1} \cap (\mathbf{w} + L)|}{|\mathbb{Z}^l_{0..m-1}|}$$
$$\leq \frac{1}{d'd''}(1 + \frac{d'-1}{m})(1 + \frac{d''-1}{m}) \ .$$

This inequality implies the claim. □

Lemma 11.4.6. *The success probability of Algorithm* `randomRelation` *is bounded from below by* $L_{|\Delta|}[1/2, -1/(4z) - o(1)]$.

Proof. Let $\mathfrak{a}$ be an $\mathcal{F}$-smooth reduced $\mathcal{O}$-ideal. Let $\mathfrak{a} = \mathcal{F}^{\mathbf{a}}$ for some $\mathbf{a} \in \mathbb{Z}^f$. Then $\mathfrak{a}$ is found in `randomRelation` if and only if the selected exponent vector $\mathbf{v}$ is in the coset $\mathbf{a} - \mathbf{w} + L([\mathcal{F}])$. By Lemma 11.4.5 the probabiltity for `randomRelation` to find $\mathfrak{a}$ is at least

$$\frac{1}{h}(1 - \frac{h-1}{|\Delta|}) \; .$$

So by Lemma 11.4.4 the probability for `randomRelation` to find an $\mathcal{F}$-smooth $\mathfrak{a}$ is at least

$$\frac{\sqrt{|\Delta|}}{h} L[1/2, -1/4z - o(1)](1 - \frac{h-1}{|\Delta|}) \; .$$

From Theorem 9.3.11 we see that the the first factor is bounded from below by $1/\log|\Delta|$. Corollary 9.3.12 implies that the last factor is bounded from below by a constant. This yields the desired lower bound for the success probability of `randomRelation`. □

11.4.2 Computing a sublattice of full rank

In this section we give an algorithm which determines a sublattice of full rank in the relation lattice $L([\mathcal{F}])$ by calculating a strictly diagonally dominant $f \times f$ matrix with integer entries whose columns are relations in $L([\mathcal{F}])$.

The following auxiliary functions are needed. We let $g(\Delta, z)$ be a function which for fixed z goes to 0 as $-\Delta$ goes to infinity and such that the probability from Lemma 11.4.6 is at least

$$p(\Delta, z) = L_{|\Delta|}[1/2, -1/(4z) + g(\Delta, z)] \; . \tag{11.11}$$

Set

$$B_1(\Delta) = (f-1)|\Delta| + \log|\Delta| \; , \tag{11.12}$$

and let

$$\mathbf{e}_i = (\underbrace{0, \ldots, 0}_{i-1}, 1, \underbrace{0, \ldots, 0}_{f-i}), \quad 1 \le i \le f \; . \tag{11.13}$$

So $\mathbf{e}_i$ is the ith identity vector in $\mathbb{Z}^f$.

The listing of the algorithm which we will call `fullRank` is shown on page 242. For its analysis we need the following result.

Lemma 11.4.7. *Let* $\mathfrak{a}$ *be a reduced* $\mathcal{O}$*-ideal and let* $\mathbf{a} \in \mathbb{Z}^f$ *with* $\mathfrak{a} = \mathcal{F}^{\mathbf{a}}$. *Then the number of non-zero entries in* $\mathbf{a}$ *is at most* $\log|\Delta|$, *the entries are non-negative, and each entry is smaller than* $\log|\Delta|$.

Algorithm 11.4 `fullRank` $(\Delta, \mathcal{F}, z)$

Input: The discriminant Δ, the factor base $\mathcal{F}$, the parameter z.
Output: `nil` or relations $\mathbf{z}_1, \dots, \mathbf{z}_f \in L([\mathcal{F}])$ such that the matrix $(\mathbf{z}_1, \dots, \mathbf{z}_f)$ is strictly diagonally dominant.

```
$l = \lceil (\log f)/p(\Delta, z) \rceil$.
for $i = 1, 2, \dots, f$ do
  $j \leftarrow 0$
  repeat
    $j \leftarrow j + 1$
    $\mathbf{z}_i \leftarrow$ randomRelation$(\Delta, \mathcal{F}, B_1(\Delta)\mathbf{e}_i)$
  until $j = l$ or $\mathbf{z}_i \neq$ nil
  if $\mathbf{z}_i =$ nil then
    return nil
return $\mathbf{z}_1, \dots, \mathbf{z}_f$
```

Proof. We know from Lemma 5.4.1 that $\mathrm{N}(\mathfrak{a}) \leq \sqrt{|\Delta|/3}$. Let

$$\mathrm{N}(\mathfrak{a}) = \prod_{p \mid \mathrm{N}(\mathfrak{a})} p^{e(p)} \tag{11.14}$$

be the prime factorization of $\mathrm{N}(\mathfrak{a})$ and let l be the number of distinct prime factors in that factorization. By Proposition 8.6.11 the entries of $\mathbf{a}$ are 0 or $e(p)$ at prime ideals with norm p that divide $\mathrm{N}(\mathfrak{a})$.

We have

$$2^l \leq \mathrm{N}(\mathfrak{a}) \leq \sqrt{|\Delta|/3}\ .$$

It follows that

$$l < \frac{1}{2} \log_2(|\Delta|/3) < \log|\Delta|\ . \tag{11.15}$$

Also, for each prime p with $p \mid \mathrm{N}(\mathfrak{a})$ we have

$$2^{e(p)} \leq \mathrm{N}(\mathfrak{a}) \leq \sqrt{|\Delta|/3}\ . \tag{11.16}$$

As in (11.15) it follows that $e(p) < \log|\Delta|$. □

.

Proposition 11.4.8. *Algorithm* `fullRank` *outputs* `nil` *with probability smaller than* 3/4*. If* `fullRank` *does not output* `nil`*, then* `fullRank` *outputs a diagonally dominant relation matrix.*

Proof. By Lemma 11.4.6 the probability that l calls of `randomRelation` all yield `nil` is at most $(1 - p(\Delta, z))^l$. Hence, the probability that `fullRank` computes f relations is a least

$$(1 - (1 - p(\Delta, f))^l)^f\ .$$

Lemma 11.2.2 implies that `randomRelation` outputs `nil` with probability smaller than 3/4.

Suppose that `fullRank` outputs a relation matrix. For $1 \le i \le f$ the ith relation in that matrix is of the form

$$\mathbf{z} = \mathbf{v} + B_1(\Delta)\,\mathbf{e}_i - \mathbf{a} \tag{11.17}$$

with $\mathbf{v} = (v_1, \ldots, v_f) \in \mathbb{Z}^f_{0..|\Delta|-1}$, $\mathbf{a} = (a_1, \ldots, a_f) \in \mathbb{Z}^f_{\ge 0}$, and $\mathcal{F}^{\mathbf{a}} = \mathfrak{a}$ for some reduced $\mathcal{O}$-ideal $\mathfrak{a}$. From Lemma 11.4.7 we know that $|a_j| \le \log \Delta$. Thus our choice of $\mathbf{v}$ implies

$$\max(|a_j|, |v_j|) < |\Delta| \quad \text{for all } j \ne i\ . \tag{11.18}$$

If $\mathbf{z} = (z_1, \ldots, z_f)$ then

$$z_i \ge B_1(\Delta) - \log|\Delta| = (f-1)|\Delta| > \sum_{\substack{j=1,\ldots,f\\ j\ne i}} \max(|a_j|, |v_j|) \ge \sum_{\substack{j=1,\ldots,f\\ j\ne i}} |z_j|\ .$$

This shows that the relation matrix $(\mathbf{z}_1, \ldots, \mathbf{z}_f)$ is strictly diagonally dominant. □

We analyze Algorithm `fullRank`.

Proposition 11.4.9. *The running time of Algorithm* `fullRank` *is bounded from above by* $L_{|\Delta|}[1/2, 2z + 1/(4z) + o(1)]$.

Proof. In `fullRank` we have $l = L_{|\Delta|}[1/2, 1/(4z) + o(1)]$. Algorithm `fullRank` calls `randomRelation` at most $fl = L_{|\Delta|}[1/2, z + 1/4z + o(1)]$ times. By Lemma 11.4.2 each call of `randomRelation` takes time $L_{|\Delta|}[1/2, z + o(1)]$. Hence, the total running time of `fullRank` is $L_{|\Delta|}[1/2, 2z + 1/(4z) + o(1)]$. □

11.4.3 Computing $L([\mathcal{F}])$

Assume that we have applied `fullRank` successfully and have computed the relation matrix

$$Z = (\mathbf{z}_1, \ldots, \mathbf{z}_f)\ . \tag{11.19}$$

We explain the computation of a basis of the relation lattice $L([\mathcal{F}])$.

The idea is to compute sufficiently many additional relations

$$\mathbf{z}_{f+1}, \ldots, \mathbf{z}_{f+N}$$

such that the full sequence $S = (\mathbf{z}_1, \ldots, \mathbf{z}_{f+N})$ generates $L([\mathcal{F}])$ with high probability. Hermite and Smith normal form computation yield the structure of the class group as explained in Section 9.7.1.

We determine the number l of additional relations. For this purpose we estimate the determinant of the initial relation matrix Z. We set

$$B_2(\Delta) = (f+1)|\Delta| + \log|\Delta|\ . \tag{11.20}$$

Lemma 11.4.10. *We have* $|\det Z| \leq B_2(\Delta)^f$.

Proof. Fix $i \in \{1, \dots, f\}$. As in the proof of Proposition 11.4.8 we can write

$$\mathbf{z}_i = \mathbf{v}_i + B_1(\Delta)\,\mathbf{e}_i - \mathbf{a}\,, \quad 1 \leq i \leq n\,.$$

For each entry z_{ij} of $\mathbf{z}_i$ we have

$$-\log|\Delta| \leq z_{ij} < \begin{cases} |\Delta| & \text{if } i \neq j, \\ f|\Delta| + \log|\Delta| & \text{if } i = j, \end{cases}$$

and hence

$$|\mathbf{z}_i|^2 \leq (f^2+f-1)|\Delta|^2 + 2f|\Delta|\log|\Delta| + (\log|\Delta|)^2 < B_2(\Delta)^2 \quad \text{for all } 1 \leq i \leq f\,.$$

Proposition A.5.1 implies the assertion. □

It follows from Lemma 11.4.10 that we need to extend the relation lattice computed by `fullRank` at most $f\log_2(B_2(\Delta))$ times until we find the full relation lattice $L([\mathcal{F}])$.

We estimate the probability that a successful call to `randomRelation` extends a given sublattice of $L([\mathcal{F}])$.

Lemma 11.4.11. *Let L be a proper sublattice of $L([\mathcal{F}])$. Suppose that Algorithm* `randomRelation` *outputs a relation* $\mathbf{z}$. *Then the probability for* $\mathbf{z}$ *to be outside of L is at least* 2^{-17}.

Proof. Suppose that `randomRelation` is successful. Then the algorithm has found an $\mathcal{F}$-smooth reduced $\mathcal{O}$-ideal $\mathfrak{a}$. Fix one such $\mathfrak{a}$ and denote by S the set of vectors $\mathbf{v} \in \mathbb{Z}^f_{0..|\Delta|-1}$ that, when selected in `randomRelation`, yield $\mathfrak{a}$. `randomRelation` has selected one of those vectors and found a relation $\mathbf{z}(\mathbf{v})$. By N denote the number of vectors in S such that $\mathbf{z}(\mathbf{v})$ is outside of L. Set

$$S' = S \cap \mathbb{Z}^f_{0..|\Delta|-h_\Delta-1}\,. \tag{11.21}$$

If we can show that

$$N \geq |S'|/2\,, \tag{11.22}$$

then the conditional probability $N/|S|$ which we want to estimate is at least $|S'|/(2|S|)$. Lemma 11.4.5 says

$$\frac{|S'|}{|S|} \geq \frac{|\Delta| - 2h_\Delta + 1}{|\Delta| + h_\Delta - 1} \cdot \frac{(|\Delta| - h_\Delta)^{f-1}}{|\Delta|^{f-1}}\,.$$

We apply Corollary 9.3.12 and Lemmas 11.3.3 and 11.2.1 in order to see that this fraction is larger than 2^{-17}.

We prove (11.22). Let N' be the number of $\mathbf{v} \in S'$ such that the relation $\mathbf{z}(\mathbf{v})$ is inside of L. If $N' < |S'|/2$, then (11.22) is proved. Assume that $N' \geq$

$|S'|/2$. Since L is properly contained in $L([\mathcal{F}])$, there is a Hermite normal form basis vector $\mathbf{b}$ of $L([\mathcal{F}])$ that is not contained in L. Since the entries of $\mathbf{b}$ are non-negative and bounded by the class number h_Δ, it follows that for each $\mathbf{v} \in S'$ the vector $\mathbf{v} + \mathbf{b}$ is in S. Also, the relation $\mathbf{z}(\mathbf{v} + \mathbf{b}) = \mathbf{z}(\mathbf{v}) + \mathbf{b}$ is outside of L since $\mathbf{b}$ is not in L. This proves (11.22). □

This suggests the following algorithm.

Algorithm 11.5 `relationLattice` $(\Delta, \mathcal{F}, z)$

Input: The discriminant Δ, the factor base $\mathcal{F}$ with cardinality f, the parameter z.
Output: `nil` or relations $\mathbf{z}_1, \ldots, \mathbf{z}_N \in L[\mathcal{F}]$.

```
$N \leftarrow 1$, $i \leftarrow 0$, $k \leftarrow \lceil f \log_2 B_2(\Delta) \rceil$, $n \leftarrow \lceil 2^{17} \log k \rceil$, $l \leftarrow \lceil (\log kn)/p(\Delta, z) \rceil$
repeat
  $i \leftarrow i + 1$
  $\mathbf{z}_{N+1} \leftarrow$ randomRelation$(\Delta, \mathcal{F}, 0)$
  if $\mathbf{z}_{N+1} \neq$ nil then
    $N \leftarrow N + 1$
until $i \geq kln$ or $N \geq kn$
if $N < kn$ then
  return nil
else
  return $\mathbf{z}_1, \ldots, \mathbf{z}_N$
```

Proposition 11.4.12. *Algorithm* `relationLattice` *outputs* `nil` *with probability smaller than* $3/4$. *Assume that* `relationLattice` *outputs* $S = (\mathbf{z}_{f+1}, \ldots, \mathbf{z}_{f+N})$. *Then the probability that* $Z \cup S$ *generates* $L([\mathcal{F}])$ *is larger than* $1/4$.

Proof. The probability that `relationLattice` does not output `nil` is larger than

$$(1 - (1 - p(\Delta, z))^l)^{kn}$$

which in turn is larger than $1/4$ by Lemma 11.2.2 since $lp(\Delta, z) > \log kn$ and $p(\Delta, z)$ is a lower bound for the probability that a single call to `randomRelation` succeeds.

Assume that `relationLattice` has output $\mathbf{z}_1, \ldots, \mathbf{z}_{kn}$. Then the probability that $\mathbf{z}_i$ does not lie in the lattice L_{i-1} generated by Z and $\mathbf{z}_1, \ldots, \mathbf{z}_{i-1}$ (provided L_{i-1} is not yet equal to $L([\mathcal{F}])$) is larger than 2^{-17} by Lemma 11.4.11. Hence the probability that such an extension occurs k times or $L_i = L([\mathcal{F}])$ for some $i < kn$ is larger than

$$(1 - (1 - 2^{-17})^n)^k \ .$$

In either case, we have $L_n = L([\mathcal{F}])$. Since $n > 2^{17} \log k$, this probability is again larger than $1/4$. □

Proposition 11.4.13. *The running time of* `relationLattice` *is bounded by* $L[1/2, 2z + 1/(4z) + o(1)]$.

Proof. The proof is analogous to that of Proposition 11.4.9. □

11.4.4 Computing the structure of Cl_Δ

Taking the algorithms from the preceding two sections together we obtain an algorithm that computes with greater than constant probability the structure of Cl_Δ.

The correctness of the output is checked using Theorems 9.3.10 and 9.3.16. Set

$$\tilde{h} = 4/3 \cdot \sqrt{|\Delta|}/\pi \cdot \ell(\Delta, n) \tag{11.23}$$

where $\ell(\Delta, n)$ is defined in (9.12), and n is chosen larger than

$$\max(9(A \log \Delta + B)^2, n_0)$$

with n_0, A and B suitably selected from Table 9.1.

Then we have by Theorems 9.3.10 and 9.3.16

$$\tilde{h}/2 < h_\Delta < \tilde{h} \; . \tag{11.24}$$

Let h be the determinant of the lattice L_1 generated by the output of `fullRank` and `relationLattice`.

Lemma 11.4.14. *We have* $h = h_\Delta$ *if and only if* $h < \tilde{h}$.

Proof. Since L_1 is a full rank sublattice of L, we have

$$h = y h_\Delta \tag{11.25}$$

with a positive integer y.

Suppose that

$$h < \tilde{h} \; . \tag{11.26}$$

Then it follows from (11.24) and (11.26) that

$$h < \tilde{h} < 2h_\Delta \; . \tag{11.27}$$

This implies $y = 1$ and $h = h_\Delta$. Conversely, assume that $h = h_\Delta$. Then (11.24) implies (11.26). □

Algorithm 11.6 `classGroup` (Δ, ε)

Input: An imaginary quadratic discriminant Δ, parameter z.
Output: The structure of Cl_Δ or `nil`.

```
$\mathcal{F} \leftarrow$ factorBase$(\Delta, z)$, let $f \leftarrow |\mathcal{F}|$
$Z \leftarrow$ fullRank$(\Delta, \mathcal{F}, z)$
$S \leftarrow$ relationLattice$(\Delta, \mathcal{F}, z)$
if any of the preceding algorithms fail then return nil.
Determine the determinant $h$ of the lattice spanned by the columns of $Z \circ S$.
Set $\tilde{h} \leftarrow 4/3 \cdot \sqrt{|\Delta|}/\pi \cdot \ell(\Delta, n)$ with $n = \max(9(A \log \Delta + B)^2, n_0)$.
if $h < \tilde{h}$ then
    Determine the Smith normal form $A$ of $Z \circ S$ and transformation matrices $X$
    and $Y$ such that $A = X(\mathbb{Z} \circ S)Y$.
    Return $A$ and $X$.
else
    Return nil.
```

We analyze the performance of Algorithm `classGroup`.

Theorem 11.4.15. *Algorithm* `classGroup` *succeeds with probability larger than* $1/256$*. If it does succeed, then it returns the structure of* Cl_Δ*. Its running time is bounded by* $L_{|\Delta|}[1/2, \max(4z, 2z + 1/(4z)) + \mathrm{o}(1)]$.

Proof. The success probability of `classGroup` follows from Lemma 11.3.1 and Propositions 11.4.8 and 11.4.12.

The correctness of the output follows from Proposition 11.4.12 and Lemma 11.4.14. Proposition 9.7.3 shows that the structure of the class group can be read off the the data obtained.

The running time of `classGroup` is dominated by the relation generation in the second and third step and the computation of the Smith normal form of the relation matrix.

According to Propositions 11.4.9 and 11.4.13, the running time for the relation generation is bounded by $L_{|\Delta|}[1/2, 2z + 1/(4z) + o(1)]$.

Using the algorithm from Proposition A.5.19, the computation of the Smith normal form takes time $O(mf^3(\log f|\Delta|)^2)$ where f is the number of rows of the found relation lattice $Z \circ S$, i.e. the number of elements in the factor base, and m is the number of its columns, i.e. the number of generated relations. Here we have used that the entries in $Z \circ S$ are bounded in size by $B_1(\Delta) + |\Delta| + \log|\Delta|$, cf. Lemma 11.4.7. The numbers m and f are in $L_{|\Delta|}[1/2, z + o(1)]$. This yields a running time for this step bounded by $L_{|\Delta|}[1/2, 4z + o(1)]$.

Taking the two bounds together yields the desired running time bound for `classGroup`. □

The bound for the running time of `classGroup` in Theorem 11.4.15 is minimal if $z = 1/\sqrt{8}$. For this value of z it is $L_{|\Delta|}[1/2, \sqrt{2} + o(1)]$.

Corollary 11.4.16. *The class number of an imaginary quadratic order can be computed in time* $L_{|\Delta|}[1/2, \sqrt{2} + o(1)]$. □

11.5 The real quadratic case

Let $\Delta > 0$. Then $\mathcal{O}_\Delta$ is a real quadratic order. In addition to the structure of the class group we also wish to compute the fundamental unit ε_Δ.

11.5.1 The idea

The idea of the subexponential algorithm for computing the structure of Cl_Δ and the fundamental unit ε_Δ is to extend the relation lattice by a real component which records for each relation the length of a generator. The relations are again computed among the prime ideals of a factor base $\mathcal{F} = \mathcal{F}(\Delta, z)$ with cardinality $f = f(\Delta, z)$ as defined in section 11.3. The bound z will later be set such that $[\mathcal{F}]$ is a generating set for Cl_Δ.

Definition 11.5.1. *The* extended relation lattice $\tilde{L}(\mathcal{F})$ *for the set* $\mathcal{F} = \{\mathfrak{p}_1, \dots, \mathfrak{p}_f\}$ *of prime ideals is the lattice of all vectors* $(b_1, \dots, b_f, d) \in \mathbb{Z}^f \times \mathbb{R}$ *such that there exists an* $\alpha \in \mathcal{O}$ *such that*

$$\alpha = \prod_{i=1}^{f} \mathfrak{p}_i^{b_i} \quad \textit{and} \quad \operatorname{Log} \alpha = d \, . \tag{11.28}$$

The projection π that maps from $\mathbb{Z}^f \times \mathbb{R}$ to $\mathbb{Z}^f$ by forgetting the real component maps the extended relation lattice $\tilde{L}(\mathcal{F})$ naturally onto the relation lattice $L([\mathcal{F}])$.

Proposition 11.5.2. *The set* $\tilde{L}(\mathcal{F})$ *is an* $(f+1)$*-dimensional lattice. If the ideal classes of the elements of the sequence* $\mathcal{F}$ *generate the class group* Cl_Δ*, then its determinant is* $h_\Delta R_\Delta$.

Proof. The set $\tilde{L}(\mathcal{F})$ is an additive group. The lattice $L([\mathcal{F}])$ is an f-dimensional lattice with determinant h_Δ since $[\mathcal{F}]$ generates Cl_Δ.

Now let $\mathbf{b}_1, \dots, \mathbf{b}_f$ be f vectors in $\tilde{L}(\mathcal{F})$ such that $(\pi(\mathbf{b}_1), \dots, \pi(\mathbf{b}_l))$ is a basis of $L([\mathcal{F}])$. Also let $\mathbf{b}_{l+1} = (0, \dots, 0, R_\Delta) \in \mathbb{Z}^l \times \mathbb{R}$. Then $\mathbf{b}_{f+1} \in \tilde{L}(\mathcal{F})$. So we can write

$$(\mathbf{b}_1, \dots, \mathbf{b}_{f+1}) = \begin{pmatrix} b_{1,1} & \dots & b_{1,l} & 0 \\ \vdots & & \vdots & \vdots \\ b_{f,1} & \dots & b_{f,f} & 0 \\ \operatorname{Log} \alpha_1 & \dots & \operatorname{Log} \alpha_f & R_\Delta \end{pmatrix} \tag{11.29}$$

where $\alpha_j \in \mathcal{O}$ are such that $(\alpha_j) = \prod_{i=1}^{f} \mathfrak{p}_i^{b_{i,j}}$, $1 \le j \le f$. We claim that $\tilde{L}(\mathcal{F})$ is an $(f+1)$-dimensional lattice with basis $(\mathbf{b}_1, \dots, \mathbf{b}_{f+1})$. The vectors

$\mathbf{b}_1, \ldots, \mathbf{b}_{f+1}$ are linearly independent. Also, any linear combination of the $\mathbf{b}_j$ with integer coefficients is in $\tilde{L}(\mathcal{F})$. Conversely, let $\mathbf{b} \in \tilde{L}(\mathcal{F})$, and assume α satisfies (11.28). Then we can write $\pi(\mathbf{b}) = \sum_{k=1}^{f} x_i \pi(\mathbf{b}_i)$ with integer coefficients x_i, $1 \leq i \leq k$. Set

$$\beta = \prod_{j=1}^{f} \alpha_j^{x_j} \ .$$

Then $\beta\mathcal{O} = \mathcal{F}^{\pi(\mathbf{b})}$. Since both α and β generate the same ideal, it follows that β/α is a unit in $\mathcal{O}$. We can therefore write $\operatorname{Log} \beta/\alpha = x_{f+1} R_\Delta$ with an integer x_{f+1}. It follows that $\mathbf{b} = \sum_{j=1}^{f+1} x_j \mathbf{b}_j$. □

We generate random relations in $\tilde{L}(\mathcal{F})$ until a generating system

$$A = (\mathbf{a}_1, \ldots, \mathbf{a}_n) = \begin{pmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & \ldots & \vdots \\ a_{f,1} & \ldots & a_{f,n} \\ \operatorname{Log} \alpha_1 & \ldots & \operatorname{Log} \alpha_n \end{pmatrix} \tag{11.30}$$

is found where $\alpha_j \in F^*$ with

$$\alpha_j \mathcal{O} = \mathcal{F}^{\pi(\mathbf{a}_j)}, \quad 1 \leq j \leq n \ . \tag{11.31}$$

From that generating system a basis B of $\tilde{L}(\mathcal{F})$ of the form (11.29) is computed. The integer part $\pi(B)$ of the basis B – i.e., its first f rows – is used to compute the structure of Cl_Δ as in the imaginary quadratic case. The entry in the lower right corner of B is an approximation to the regulator R_Δ of $\mathcal{O}_\Delta$.

The fundamental unit ε_Δ can be computed from the regulator R_Δ. Since it takes $O(R_\Delta)$ bits to write down the standard representation of ε_Δ, we will require some compact representation.

11.5.2 Height

In this section we will introduce the height of a quadratic number. This is a more manageable measure for the size of a number than the size function introduced in section 10.2.2. Yet, it is still roughly proportional to the space required to write the number down in standard representation.

Definition 11.5.3. *For a quadratic number $\alpha \in \mathcal{O}$, the height $H(\alpha)$ is defined as the maximum of $|\alpha|$ and $|\sigma(\alpha)|$.*

Lemma 11.5.4. *If $\alpha \in \mathcal{O}$, then $H(\alpha) \geq 1$.*

Proof. If $\alpha \in \mathcal{O}$, then $1 \leq |\mathrm{N}(\mathfrak{a})| = |\alpha||\alpha^\sigma|$. Hence at least one of the numbers $|\alpha|$ or $|\alpha^\sigma|$ must be larger than or equal to 1. □

We relate the height of an integral number in $\mathcal{O}$ to its size as this was defined in section 10.2.2.

Lemma 11.5.5. *For non-zero* $\alpha \in \mathcal{O}$*, we have* $\log_2 H(\alpha) \leq \text{size}(\alpha) \leq 2\log_2 H(\alpha) + 1/2 \cdot \log_2 \Delta + 8$.

Proof. If $\alpha = (x + y\sqrt{\Delta})/2$, then $H(\alpha) = (|x| + |y|\sqrt{\Delta})/2$. It follows that at least one of the terms $|x|$ or $|y|\sqrt{\Delta}$ is larger than $H(\alpha)$. This implies the first inequality. Moreover, $|x|, |y|\sqrt{\Delta} \leq 2H(\alpha)$. Taking logarithms and adding the inequalities yields the second inequality. □

There is an easy relationship between the height and the length of a number introduced in section 10.1.1. On the one hand, numbers of small height have small length. On the other hand, numbers of small length and small norm have small height.

Lemma 11.5.6. *For any* $\alpha \in F^\star$ *we have* $|\operatorname{Log} \alpha| \leq \log H(\alpha)$.

Proof. By definition $\operatorname{Log}\alpha = (\log|\alpha| - \log|\sigma(\alpha)|)/2$. Each summand is smaller in absolute value than $1/2 \cdot \log H(\alpha)$. □

Lemma 11.5.7. *For any* $\alpha \in F^\star$ *with* $|\operatorname{Log}(\alpha)| < A$ *and* $|\mathrm{N}(\alpha)| < B$*, we have* $H(\alpha) < e^A\sqrt{B}$.

Proof. The assumptions on α imply both

$$e^{-2A} < |\alpha|/|\alpha^\sigma| < e^{2A} \quad \text{and} \quad |\alpha||\alpha^\sigma| < B\,.$$

Multiplying (or, respectively, dividing) the two inequalities and taking square roots yields the bound in the lemma. □

Lemma 11.5.8. *For any* $\alpha, \beta \in \mathcal{O}$*, we have* $H(\alpha\beta) \leq H(\alpha)H(\beta)$. □

Lemma 11.5.9. *For any* $\alpha \in F^\star$*, we have* $H(\alpha) = |\mathrm{N}(\alpha)|H(1/\alpha)$. □

We can bound the height of reducing numbers of primitive ideals which were defined in Section 9.1.3 on page 179.

Lemma 11.5.10. *Let* $\mathfrak{a}$ *be a primitive integral ideal, and* $\alpha = \alpha(\mathfrak{a})$ *be the reducing number of* $\mathfrak{a}$ *defined in* (9.4)*. Then* $d(\alpha)\,|\,\mathrm{N}(\mathfrak{a})$ *and* $1/\mathrm{N}(\mathfrak{a}) \leq H(\alpha) < 1$*. Moreover, if we write* $1/\alpha = \alpha'/d$ *with minimal positive* $d = d(1/\alpha)$ *and* $\alpha' \in \mathcal{O}$*, then* $d \leq \sqrt{\Delta}$ *and* $H(\alpha') < \mathrm{N}(\mathfrak{a})$.

Proof. Note that it is clear from (9.3) that $\mathrm{N}(\mathfrak{a})\alpha$ is integral. Hence $d(\alpha)\,|\,\mathrm{N}(\mathfrak{a})$. Since $H(d(\alpha)\alpha) = d(\alpha)H(\alpha)$, Lemma 11.5.4 implies the first inequality.

We prove the second inequality. Let $\mathfrak{a}_0 = \mathfrak{a}$, and $\mathfrak{a}_i = \rho(\mathfrak{a}_{i-1})$, $i = 1, \ldots, k$ be the sequence of ideals computed by `reduce`. Thus $\mathfrak{a}_k$ is the first reduced ideal in this sequence. Let $\mathfrak{a}_i = L(a_i, b_i, c_i)$. Then

$$\alpha(\mathfrak{a}) = \prod_{i=0}^{k-1} \gamma(\mathfrak{a}_i) = \prod_{i=0}^{k-1} \frac{-b_i + \sqrt{\Delta}}{2a_i} \ .$$

Thus, it suffices to show $2a_i > |b_i| + \sqrt{\Delta}$ for $i < k$. This inequality follows from the bound (6.1.1) for b_i if $a_i > \sqrt{\Delta}$ or $b_i < 0$. Assume $a_i < \sqrt{\Delta}$, and $b_i > 0$. Then $2a_i < b_i + \sqrt{\Delta}$ together with (6.1.1) would imply $|2a_i - \sqrt{\Delta}| < b_i$, and thus that $\mathfrak{a}_i$ is reduced. Since $i < k$ this is a contradiction.

We bound the denominator of $1/\alpha$. Let $\mathfrak{b} = \alpha\mathfrak{a}$. Then $\mathrm{N}(\mathfrak{b}) \leq \sqrt{\Delta}$. By Proposition 7.3.11 $\mathfrak{b} = \alpha\mathfrak{a}$ implies $(\mathrm{N}(\mathfrak{b})/\alpha) = \mathfrak{a}\cdot\sigma(\mathfrak{b})$. This shows that $\mathrm{N}(\mathfrak{b})/\alpha$ is integral, and $d(1/\alpha) \mid \mathrm{N}(\mathfrak{b})$.

Finally,

$$\begin{aligned} H(\alpha') = d(1/\alpha) \cdot H(1/\alpha) &\leq \mathrm{N}(\mathfrak{b}) \cdot H(1/\alpha) \\ &\leq \mathrm{N}(\mathfrak{b}) \cdot H(\alpha)/|\mathrm{N}(\alpha)| \\ &\leq H(\alpha) \cdot \mathrm{N}(\mathfrak{a}) < \mathrm{N}(\mathfrak{a}) \ . \end{aligned}$$ □

11.5.3 Compact representations

In real-quadratic number fields, numbers with small norm may have very large height. In this section we will introduce a representation of numbers in such a field which requires space that depends only logarithmically on its norm and quadratically on the logarithm of the discriminant. We will show how to find such a representation in polynomial time, and how to compute with it. Compact representations of quadratic integers were introduced by Johannes Buchmann, Christoph Thiel and Hugh Williams in [BTW95], and generalized to fields of arbitrary degree by Christoph Thiel in [Thi94], For a detailed exposition, see [Thi95].

Proposition 11.5.11. *Let $\Delta > 0$ and $\gamma \in \mathcal{O}$. Then there exist numbers $k \in \mathbb{N}$, and $\gamma_i \in F^\star$, $0 \leq i \leq k$, such that*

1.

$$\gamma = \gamma_0 \prod_{i=1}^{k} \gamma_i^{2^{k-i}} \ ; \tag{11.32}$$

2. for all j with $1 \leq j \leq k$, the ideal $\mathfrak{b}_j$ generated by

$$\beta_j = \prod_{i=1}^{j} \gamma_i^{2^{j-i}} \tag{11.33}$$

is reduced; and

3. the following size constraints are satisfied
a) $k \leq 2 + \log_2(1 + \log H(\gamma))$,
b) $H(\gamma_0) \leq |\mathrm{N}(\gamma)|$ and $d(\gamma_0) < \sqrt{\Delta}$,
c) $d(\gamma_i) \leq \Delta$ for $1 \leq i \leq k$, and
d) $H(\gamma_i) \leq \Delta$.

Definition 11.5.12. *Let $\gamma \in \mathcal{O}$. Any sequence $(k, \gamma_0, \ldots, \gamma_k)$ satisfying* (11.32) *and the size constraints of Proposition 11.5.11 is called a* compact representation *of γ.*

Proof. We will prove Proposition 11.5.11 constructively. If $\gamma \in \mathbb{Z}$, we may choose $k = 0$ and $\gamma_0 = \gamma$. Moreover, the case that γ is imprimitive reduces trivially to the primitive case. Thus we will assume that γ is primitive, and not in $\mathbb{Z}$.

Let α be the reducing number of (γ), and $\beta = \alpha\gamma$. Set $\gamma_0 = 1/\alpha$. Then $\gamma = \gamma_0\beta$. Due to Lemma 11.5.10, $H(\gamma_0) \leq |\mathrm{N}(\gamma)|$ and $d(\gamma_0) \leq \sqrt{\Delta}$. Thus constraint 3b is satisfied.

Denote $\operatorname{Log}\beta$ by δ. Taking conjugates if necessary we may assume that $\delta \geq 0$. If $\delta \leq 3/4 \cdot \log\Delta$, then $H(\beta) \leq \Delta$ by Lemma 11.5.7, since $|\mathrm{N}(\beta)| < \sqrt{\Delta}$. Moreover, $d(\beta) = 1$ since β generates an integral ideal. Thus if we set $k = 1$ and $\gamma_1 = \beta$ constraints 2, 3c and 3d are satisfied. Since γ is integral, we have $H(\gamma) \geq 1$. Hence $\log_2(1 + \log H(\gamma)) > 0$. This implies that k satisfies constraint 3a of the proposition. Thus, all assertions of the proposition are satisfied in this case.

Assume $\delta > 3/4 \cdot \log\Delta$. Let $k = \lceil \log_2 \delta \rceil + 1$. Then $k > 1$. Set $\delta_i = \delta/2^{k-i}$ for $1 \leq i < k$. Note that by Lemma 10.1.8

$$\delta = \operatorname{Log}\beta = \operatorname{Log}\gamma - \operatorname{Log}\gamma_0 \leq \operatorname{Log}\gamma + \frac{1}{2}\log|\mathrm{N}(\gamma)| = \log|\sigma(\gamma)| \leq \log H(\gamma) \ . \tag{11.34}$$

Hence $k \leq 2 + \log_2 \log H(\gamma)$. By Lemma 10.1.6 there exist $\beta_i \in \mathcal{O}$ such that (β_i) is reduced and

$$|\delta_i - \operatorname{Log}\beta_i| \leq 1/4 \cdot \log\Delta \ .$$

Set finally $\beta_0 = 1$ and $\beta_k = \beta$. We define $\gamma_i = \beta_i/\beta_{i-1}^2$ for $i = 1, \ldots, k$. This implies (11.33), and condition 2 is satisfied. We obtain

$$\gamma = \gamma_0\beta = \gamma_0\beta_k = \gamma_0 \prod_{i=1}^{k} \gamma_i^{2^{k-i}} \ .$$

We verify that constraints 3c and 3d are satisfied for $i > 1$. We have

$$\operatorname{Log}\gamma_i = 2(\delta_{i-1} - \operatorname{Log}\beta_{i-1}) - (\delta_i - \operatorname{Log}\beta_i) \ .$$

Hence $|\operatorname{Log}\gamma_i| \leq 3/4 \cdot \log\Delta$. Also, $\mathrm{N}(\beta_i) = \mathrm{N}(\gamma_i)\mathrm{N}(\beta_{i-1})^2$. Hence, $|\mathrm{N}(\gamma_i)| \leq \sqrt{\Delta}$. Thus Lemma 11.5.7 implies $H(\gamma_i) \leq \Delta$ which is constraint 3d for $i > 1$. Constraint 3c is likewise fulfilled since $\mathrm{N}(\beta_{i-1})^2\gamma_i = \beta_i(\beta_{i-1}^\sigma)^2 \in \mathcal{O}$ implies $d(\gamma_i) \mid \mathrm{N}(\beta_{i-1}^2) \leq \Delta$.

It remains to show $d(\gamma_1), H(\gamma_1) \leq \Delta$. By definition, $\gamma_1 = \beta_1$, and β_1 generates a reduced, hence integral ideal. We thus have $d(\gamma_1) = 1$.

Due to the choice of k, we have $\delta_1 \leq 1$. Hence $\operatorname{Log}\gamma_1 < 1 + 1/4 \cdot \log\Delta$. Since $2 < \log\Delta$ (otherwise $\Delta = 5$ and $\beta = 1$), this is smaller than $3/4 \cdot \log\Delta$. Also, $|\mathrm{N}(\gamma_1)| \leq \sqrt{\Delta}$ since $(\beta_1) = (\gamma_1)$ is reduced. Hence by Lemma 11.5.7 $H(\gamma_1) < \Delta$. Indeed, if $4 \geq \log\Delta$, then we may choose $\beta_1 = \gamma_1 = 1$. □

The construction in the proof of Proposition 11.5.11 can easily be turned into an algorithm for the computation of compact representations which we will call `compactRepresentation`. In order to compute β_i and γ_i for $i > 0$ we need to know $\operatorname{Log}\gamma$. For the computation of γ_0 we need to know the ideal generated by γ. Hence we will take these two quantities as input for our algorithm. From the proof it is also clear that we may restrict `compactRepresentation` to the case that (γ) is reduced, and $\operatorname{Log}\gamma$ is sufficiently large (in particular, positive).

First we will need an algorithm that computes the β_i and γ_i given δ_i and β_{i-1} for all i with $1 \leq i < k$. We use that we know $\delta_i - \operatorname{Log}(\beta_{i-1}^2)$ which by induction is smaller in absolute value than $1/2 \cdot \log\Delta$. Hence we need only reduce (β_{i-1}^2) and take a few corrective reduction steps.

Algorithm 11.7 `proxReduced` $(\mathfrak{b}, \delta)$

Input: Integral ideal $\mathfrak{b}$, distance $\delta \in \mathbb{Q}$.
Output: Reduced ideal $\mathfrak{c}$ and relative generator γ with $\mathfrak{c} = \gamma\mathfrak{b}$ and $|\operatorname{Log}\gamma - \delta| < 1/4 \cdot \log\Delta$.

```
$(\mathfrak{c}, \gamma) \leftarrow$ reduce($\mathfrak{b}$).
while $\operatorname{Log}\gamma \leq \delta - 1/4 \cdot \log\Delta$ do
    $\gamma \leftarrow \gamma(\mathfrak{c})\gamma$ and $\mathfrak{c} \leftarrow \rho(\mathfrak{c})$.
while $\operatorname{Log}\gamma \geq \delta + 1/4 \cdot \log\Delta$ do
    $\gamma \leftarrow \gamma'(\mathfrak{c})\gamma$ and $\mathfrak{c} \leftarrow \rho^{-1}(\mathfrak{c})$.
return $(\mathfrak{c}, \gamma)$.
```

Note that the comparisons in the controlling conditions of the while loops of `proxReduced` cannot be performed exactly. We determine the precision sufficient for the correctness of `proxReduced`.

Assume we compute approximations to the quantity $L = \operatorname{Log}\gamma - \delta + 1/4 \cdot \log\Delta$ with error smaller than ϵ and we execute the loop when $L \leq \epsilon$. Then it is guaranteed that $\operatorname{Log}\gamma - \delta + 1/4 \cdot \log\Delta > 0$ after the loop whether it was entered or not. If the loop was indeed executed at least once, then we also have $\operatorname{Log}\gamma - \delta + 1/4 \cdot \log\Delta < 2\varepsilon$ prior to the last run of the loop.

By Exercise 10.4.10 $\operatorname{Log}\gamma(\mathfrak{c}) < 1/2 \cdot \log\Delta - \log 2$ for any reduced $\mathfrak{c}$. Hence choosing $2\epsilon < \log 2$ guarantees $\operatorname{Log}\gamma < \delta + 1/4 \cdot \log\Delta$ after the first loop if this is entered at all.

We proceed analogously to ensure $\operatorname{Log}\gamma < \delta - 1/4 \cdot \log\Delta$ after the second loop. Thus independently of whether any loop was entered, none or both, the output $(\mathfrak{c}, \gamma)$ has the property $|\operatorname{Log}\gamma - \delta| < 1/4 \cdot \log\Delta$.

Next we require an algorithm for the computation of γ_k. This is done in analogy to `proxReduced`. Hence we will skip a listing, and lieave it to the reader to produce one in exercise 11.7.4. The algorithm will be called `reachReduced`. It takes as input two integral ideals $\mathfrak{b}$ and $\mathfrak{c}$ the second of which is reduced and an approximation to the length δ of a short relative generator. `reachReduced` computes γ with $\mathfrak{c} = \gamma\mathfrak{b}$ by reducing $\mathfrak{b}$ and then taking a few reduction steps, normal or inverse, until $\mathfrak{c}$ is reached. Note that by Lemma 10.1.6 the input $\mathfrak{c}$ is superfluous if we know the ideal generated by γ is reduced and the approximation to δ has error smaller than $1/(2\sqrt{\Delta})$.

The following lemma summarizes the properties of `proxReduced` and `reachReduced`. The run-time bounds are proved analogously to Lemma 10.2.4 on the basis of Lemmas 10.1.8 and 10.1.6.

Lemma 11.5.13. *Algorithms* `proxReduced` *and* `reachReduced` *are correct. If the input ideal $\mathfrak{b}$ has norm smaller Δ, then the running time of the algorithms is $O((\delta + \log\Delta)(\log\Delta)^2)$, or $O((|\delta - \operatorname{Log}\gamma| + \log\Delta)(\log\Delta)^2)$, respectively.* □

We are now able to state algorithm `compactRepresentation`.

Algorithm 11.8 `compactRepresentation` $(\mathfrak{c}, \delta)$

Input: Reduced principal ideal $\mathfrak{c}$ and $\delta \in \mathbb{Q}$ for which there exists a generator γ of $\mathfrak{c}$ such that δ approximates $\operatorname{Log}\gamma$.

Output: Numbers $k \in \mathbb{Z}$ and $\gamma_i \in F^\star$, $i = 1, \dots, k$, such that $\gamma = \prod_{i=1}^{k} \gamma_i^{2^{k-i}}$ generates $\mathfrak{c}$.

```
$k \leftarrow \lceil \log_2 \delta \rceil + 1$, $\delta_i \leftarrow \delta/2^{k-i}$ for $i = 1, \dots, k$. $\beta_0 \leftarrow 1$, $\mathfrak{b}_0 \leftarrow \mathcal{O}$, $\lambda_0 = 0$.
for $i = 1, \dots, k-1$ do
    $(\mathfrak{b}_i, \gamma_i) \leftarrow$ proxReduced($\mathfrak{b}_{i-1}^2, \delta_i - 2\lambda_{i-1}$). $\lambda_i \leftarrow 2\lambda_{i-1} + \operatorname{Log}\gamma_i$.
$\gamma_k \leftarrow$ reachReduced($\mathfrak{b}_{k-1}^2, \mathfrak{c}, \delta - 2\lambda_{k-1}$)
return $k, \gamma_1, \dots, \gamma_k$.
```

In order for `compactRepresentation` to be correct we need to ensure that `proxReduced` gets called with parameters of sufficient precision. This means that for each $i < k$ we need to compute $\operatorname{Log}\gamma_i$ with error smaller than $2^{i-k-1}/i \cdot \log 2$. The precision of the input to `reachReduced` only affects its run-time.

Lemma 11.5.14. *The output of* `compactRepresentation` *has the following properties:*
1. $k \leq \log_2(|\delta - \operatorname{Log}\gamma| + \log H(\gamma)) + 2$,
2. $H(\gamma_0) \leq |\mathrm{N}(\gamma)|$ and $d(\gamma_0) < \sqrt{\Delta}$,
3. $d(\gamma_i) \leq \Delta$ for $1 \leq i \leq k$,
4. $H(\gamma_i) \leq \Delta$ for $1 \leq i < k$ and $H(\gamma_k) \leq e^{|\delta - \operatorname{Log}\gamma|}\Delta^{3/4}$.
The run-time of `compactRepresentation` *is bounded by $O(k(\log\Delta + |\delta - \operatorname{Log}\gamma|)(\log\Delta)^2)$.*

Lemma 11.5.14 says that `compactRepresentation` indeed returns a compact representation of γ or $-\gamma$ if given (γ) and a sufficiently good approximation to $\operatorname{Log}\gamma$. Note that we may start with a poor approximation to $\operatorname{Log}\delta$, recompute δ using the output of a first call to `compactRepresentation`, and then call `compactRepresentation` again in order to obtain a true compact representation of γ.

Proof (of Lemma 11.5.14). The bound on k follows from (11.34). The bounds on the denominators of all γ_i, and the height of all γ_i except γ_k follow from the correctness of `proxReduced` and `reachReduced` and Lemma 11.5.7 as in the proof of Proposition 11.5.11.

We prove the bound on the height of γ_k. Define β_i by (11.33). Then $\mathfrak{b}_i = (\beta_i)$ and $\operatorname{Log}\beta_i = \lambda_i$ for $i < k$ by construction. From the correctness of `proxReduced` it follows that

$$|\delta_i - \lambda_i| = |\delta_i - (2\lambda_{i-1} + \operatorname{Log}\gamma_i)| = |(\delta_i - 2\lambda_{i-1}) - \operatorname{Log}\gamma_i| < 1/4 \cdot \log\Delta \quad (11.35)$$

for all $i < k$. Hence

$$\operatorname{Log}\gamma_k = \operatorname{Log}\gamma - \operatorname{Log}\beta_i^2 = (\operatorname{Log}\gamma - \delta) + 2(\delta_{k-1} - \lambda_{k-1}) < (1/2)\cdot\log\Delta + |\delta - \operatorname{Log}\gamma|\ .$$

If we now use Lemma 11.5.7 and apply, as before, $\mathrm{N}(\gamma_i) < \sqrt{\Delta}$, then we obtain the asserted bound on $H(\gamma_k)$.

The bound on the run-time, finally, follows from the bounds on the run-time of `proxReduced` and `reachReduced` in Lemma 11.5.13, if we take into account that their input is bound by $1/2 \cdot \log\Delta$ due to (11.35). □

Corollary 11.5.15. *Given two numbers $\alpha, \beta \in \mathcal{O}$ in compact representation, we can determine in time polynomial in $|\mathrm{N}(\alpha)|$, $|\mathrm{N}(\beta)|$ and $\log\Delta$:*

1. *the ideals (α) and $(1/\alpha)$,*
2. *the sign of α,*
3. *the norm of α,*
4. *a compact representation of $\alpha\beta$,*
5. *a compact representation of $d(1/\alpha)/\alpha$,*
6. *equality between α and β,*
7. *whether $|\alpha| > |\beta|$,*
8. *whether α divides β in $\mathcal{O}$, and, if the decision is positive, a compact representation of α/β.*

Proof. Assume we are given α in compact representation. For the components of this representation and the derived quantities we adopt the notation of Proposition 11.5.11 and its proof.

Note first that the computation of the ideals $\mathfrak{b}_j$ from $\mathfrak{b}_{j-1}$ involves only one squaring of reduced ideal $\mathfrak{b}_{j-1}$, the reduction of the result, and $O(\log\Delta)$ reductions. Hence it can be effected in time cubic in $\log\Delta$, cf. the explicit formulae of Proposition 8.4.8 and Corollary 7.3.18. We arrive at (α) by multiplying $\mathfrak{b}_k$ with (γ_0) which in view of the explicit formulae, the bound on the height of γ_0 and Lemma 11.5.5 is done in time $O((\log|\mathrm{N}(\alpha)| + \log\Delta)^2)$. Equation (7.10) yields $(1/\alpha)$ in quadratic time.

Equation (11.32) implies that the sign of α is the product of the signs of γ_0 and γ_k.

From (α) and the sign of α we immediately obtain $\mathrm{N}(\alpha)$.

The compact representation of $\alpha\beta$ is computed from $(\alpha)(\beta)$, the signs of α and β and $\operatorname{Log}\alpha + \operatorname{Log}\beta$. Likewise, $d(1/\alpha)$ and the compact representation of $d(1/\alpha)/\alpha$ is obtained from $1/(\alpha)$ and $-\operatorname{Log}\alpha$. Both computations can be performed using `compactRepresentation` within the quasi-cubic run-time bounds of Lemma 11.5.14.

Equality of α and β can be determined by comparing (α) and (β), $\operatorname{Log}\alpha$ and $\operatorname{Log}\beta$, and the signs of both numbers. This is done in cubic time.

We show how to effect a size comparison between $|\alpha|$ and $|\beta|$ if α and β are given by compact representations. Without loss of generality we may assume that $|\mathrm{N}(\alpha)| \leq |\mathrm{N}(\beta)|$. Note that this inequality can be checked exactly in cubic time. If $\beta \notin \mathbb{Z}$, we compare $|\alpha\sigma(\beta)|$ with $|\mathrm{N}(\beta)|$. Again, both numbers can be computed in quasi-cubic time. Assume $\beta \in \mathbb{Z}$ and call it b instead. If $|\alpha|$ is significantly smaller or larger than b, then the comparison is easily performed. Assume that this is not the case. Then $|\sigma(\alpha)|$ is not significantly larger than b either due to $|\mathrm{N}(\alpha)| < b^2$. Hence $H(\alpha)$ is not significantly larger than b, and the compact representation of α can be converted into the normal one in quasi-quadratic time. The size comparison between a real-quadratic number and an integer, finally, can be performed exactly in quadratic time.

The number α divides β in $\mathcal{O}$ if and only if (β/α) is an integral ideal. Hence it suffices to compute $(\beta)(1/\alpha)$ to establish divisibility. The quotient β/α is computed from (β/α), the signs of α and β and $\operatorname{Log}\beta - \operatorname{Log}\alpha$. All this can be done, as previously, in quasi-cubic time. □

11.5.4 Generating random relations

We explain the computation of random relations in the extended relation lattice $\tilde{L}(\mathcal{F})$. It is similar to the generation of the random relations in the imaginary quadratic case.

A random reduced ideal $\mathfrak{a}$ is determined with almost uniform distribution for which a decomposition

$$\mathfrak{a} = \gamma\mathcal{F}^{\mathbf{v}} \tag{11.36}$$

is known with $\gamma \in F^\star$ and $\mathbf{v} \in \mathbb{Z}^f$. This is done by first choosing $\mathbf{v}$ from some large cube, and then choosing randomly a reduced ideal in the class $[\mathcal{F}^{\mathbf{v}}]$. Suppose that $\mathfrak{a}$ admits another factorization

$$\mathfrak{a} = \mathcal{F}^{\mathbf{u}} \tag{11.37}$$

with $\mathbf{u} \in \mathbb{Z}^f$. Since $\mathfrak{a}$ is an almost uniformly distributed random reduced ideal, the probability for this to be true can be estimated as in Section 11.4.1. We obtain with

$$\mathbf{w} = (\mathbf{v} - \mathbf{u}, -\operatorname{Log}\gamma)$$

an extended relation in $\tilde{L}(\mathcal{F})$.

We describe the algorithm for computing the random reduced ideal $\mathfrak{a}$. In the imaginary quadratic case we have determined $\mathfrak{a}$ as the reduced ideal in a random ideal class. However, in the real quadratic case each ideal class contains a whole cycle of reduced ideals. We therefore proceed in two steps. First, we calculate a random ideal class C with almost uniform distribution. This is done as in the imaginary quadratic case. Then we choose a random reduced ideal from the cycle in C.

We explain Algorithm `randomReduced` which computes a random reduced ideal in a given ideal class C.

Algorithm 11.9 `randomReduced` $(\mathfrak{a}, B)$

Input: A reduced $\mathcal{O}$-ideal $\mathfrak{a}$, $B \in \mathbb{N}$ with $B > R_\Delta$.
Output: A random reduced ideal $\mathfrak{b}$ in the equivalence class of $\mathfrak{a}$, $\beta \in F$ with $\mathfrak{b} = \beta\mathfrak{a}$.

```
Choose $d \in \{0, \dots, B-1\}$ uniformly at random.
Determine $S_d(\mathfrak{a})$.
Choose $(\mathfrak{b}, \beta) \in S_d(\mathfrak{a})$ uniformly at random.
return $(\mathfrak{b}, \beta)$.
```

Let $\mathfrak{a}$ be some reduced $\mathcal{O}$-ideal in C. For any $d \in \mathbb{N}$ we define the cycle section

$$S_d = S_d(\mathfrak{a}) = \{(\mathfrak{b}, \beta) : \beta \in F^*, \mathfrak{b} = \beta\mathfrak{a}, \mathfrak{b} \text{ is reduced}, d \leq \operatorname{Log}\beta / \log\Delta < d+1\} . \tag{11.38}$$

Here is Algorithm `randomRelation` for the case $\Delta > 0$. The additional offset $\mathbf{w}$ is introduced in order to obtain a relation matrix in the first part of the computation of $\tilde{L}(\mathcal{F})$ whose integral part is diagonally dominant.

Algorithm 11.10 `randomRelation` $(\Delta, \mathcal{F}, \mathbf{w})$

Input: Discriminant Δ, factor base $\mathcal{F}$ of length f, offset vector $\mathbf{w}$.
Output: A relation for $\mathcal{F}$.

```
**loop**
    Choose $\mathbf{v} \in \mathbb{Z}^f_{0..\Delta-1}$ uniformly at random.
    Calculate $(\mathfrak{a}, \alpha) \leftarrow$ reduce$(\mathcal{F}^{\mathbf{v}+\mathbf{w}})$.
    $(\mathfrak{b}, \beta) \leftarrow$ randomReduced$(\mathfrak{a}, \Delta)$
    **if** $\mathfrak{b} = \mathcal{F}^{\mathbf{u}}$ with $\mathbf{u} \in \mathbb{Z}^f$ **then**
        Return $(\mathbf{w} = \mathbf{v} + \mathbf{w} - \mathbf{u}, -\operatorname{Log}\alpha\beta)$.
```

Deferring the question of how the cycle sections get enumerated for a moment, we prove a lower bound for the success probability of `randomRelation`. For this we need a lower bound for the probability that a particular $\mathfrak{b}$ is chosen by `randomReduced`.

Lemma 11.5.16. *Let* $\mathfrak{a}$ *and* $\mathfrak{b}$ *be two equivalent reduced ideals. Assume* $B\log\Delta > R_\Delta$ *for some* $B \in \mathbb{Z}$. *Then the probability that* `randomReduced` *returns* $\mathfrak{b}$ *on input of* $\mathfrak{a}$ *and* B *is in the interval* $((\log 2)/(2R_\Delta) - 1/(2B), (\log\Delta)/(2R_\Delta) + 1/(2B))$.

Proof. Let β be such that $\mathfrak{b} = \beta\mathfrak{a}$ and $0 \le \operatorname{Log}\beta < R$. Denote $\operatorname{Log}\beta$ by d_0. Then $\mathfrak{b}$ is chosen by `randomReduced` only if the selected index d fulfills

$$d\log\Delta \le d_0 + kR_\Delta < (d+1)\log\Delta$$

for some $k \in \mathbb{Z}_{\ge 0}$. If such d is selected then $\mathfrak{b}$ is chosen from $S_d(\mathfrak{a})$ with probability $s(d)^{-1}$ where $s(d) = |S_d(\mathfrak{a})|$. Hence the total probability that $\mathfrak{b}$ is output is

$$\frac{1}{B}\sum_{k\,:\,0\le d_0+kR<B\log\Delta} s(\lfloor (d_0+kR)/\log\Delta\rfloor)^{-1}\ .$$

Note that by Lemma 10.1.6, we have $2 \le s(d) < 2\log_2\Delta$. Moreover, $d_0 + kR_\Delta < B\log\Delta$ is equivalent to $k \le \lfloor (B\log\Delta - d_0)/R_\Delta\rfloor$. The assumption on B assures that there is at least one k satisfying this bound. Thus we obtain for the probability P that $\mathfrak{b}$ is chosen

$$\frac{1}{\log_2\Delta} \le \frac{2BP}{\lfloor (B\log\Delta + R_\Delta - d_0)/R_\Delta\rfloor} < 1\ .$$

This inequality implies the bounds given in the lemma. □

Lemma 11.5.17. *The success probability of algorithm* `randomRelation` *is bounded from below by* $L_\Delta[1/2, -1/(4z) - o(1)]$.

Proof. Let $\mathfrak{b}$ be an $\mathcal{F}$-smooth reduced ideal, say $\mathfrak{a} = \mathcal{F}^{\mathbf{b}}$ for some $\mathbf{b} \in \mathbb{Z}^f$. Then $\mathfrak{b}$ is found in `randomRelation` only if the selected exponent vector $\mathbf{v}$ is in the coset $\mathbf{b} - \mathbf{w} + L([\mathcal{F}])$. By Lemma 11.4.5 the probability for `randomRelation` to choose such $\mathbf{v}$ is at least

$$\frac{1}{h_\Delta}\left(1 - \frac{h_\Delta - 1}{\Delta}\right)\ .$$

For any such $\mathbf{v}$, Lemma 11.5.16 says that the probability that `randomReduced` finds $\mathfrak{b}$ if given a reduced ideal $\mathfrak{a}$ in the cycle of $\mathcal{F}^{\mathbf{v}+\mathbf{w}}$ is at least $(\log 2)/(2R_\Delta) - 1/(2\Delta)$. Note that $\Delta\log\Delta > R_\Delta$ due to Theorem 9.3.11.

So by Lemma 11.4.4, the probability for `randomRelation` to find an $\mathcal{F}$-smooth $\mathfrak{b}$ is at least

$$\left(\frac{\sqrt{\Delta}}{h_\Delta R_\Delta}\frac{\log 2}{2} - \frac{1}{2h_\Delta\sqrt{\Delta}}\right)\left(1 - \frac{h_\Delta - 1}{\Delta}\right) L_\Delta[1/2, -1/(4z) - o(1)]\ .$$

From Theorem 9.3.11 and Corollary 9.3.12 we see that

$$\sqrt{\Delta}/(h_\Delta R_\Delta) > 2/(\log\Delta + 2) \quad\text{and}\quad 1 - (h_\Delta - 1)/\Delta) > 1/2\ .$$

This yields the desired lower bound for the success probability of `randomRelation`. □

Let $\mathfrak{a}$ be a reduced $\mathcal{O}$-ideal and let $d \in \mathbb{N}$. We explain how the sets $S_d(\mathfrak{a})$ are enumerated in Algorithm `randomReduced`. The reduced $\mathcal{O}$-ideals that are the first components of the elements of $S_d(\mathfrak{a})$ are all reduced $\mathcal{O}$-ideals in a small interval of the cycle of all reduced $\mathcal{O}$-ideals in the class of $\mathfrak{a}$.

The enumeration proceeds in two steps. First, a reduced $\mathcal{O}$-ideal $\mathfrak{b}$ is computed that is contained in the interval that we need to enumerate. This is done by successive squaring in analogy to `compactRepresentation` from Section 11.5.3. Then all elements of $S_d(\mathfrak{a})$ are determined by successive application of the reduction operator ρ and its inverse. Here is a listing of the enumeration algorithm `cycleSection`

Algorithm 11.11 `cycleSection` $(\Delta, \mathfrak{a}, d)$

Input: Discriminant Δ, reduced ideal $\mathfrak{a}$, distance parameter $d \in \mathbb{N}$.
Output: Set S such that $S_d(\mathfrak{a}) \subset S$ and $|(S \setminus S_d(\mathfrak{a}))| \leq 2$.

```
$\delta \leftarrow (d + 1/2)\log\Delta$, $\underline{\delta} \leftarrow d \cdot \log\Delta$, $\overline{\delta} \leftarrow (d+1)\log\Delta$,
$k \leftarrow \lceil \log_2 \delta \rceil + 1$, $\delta_i \leftarrow \delta/2^{k-i}$ for $i = 1, \dots, k$. $\beta_0 \leftarrow 1$, $\mathfrak{b}_0 \leftarrow \mathcal{O}$, $\lambda_0 = 0$.
for $i = 1, \dots, k$ do
    $(\mathfrak{b}_i, \beta_i) \leftarrow$ proxReduced$(\mathfrak{b}_{i-1}^2, \delta_i - 2\lambda_{i-1})$. $\lambda_i \leftarrow 2\lambda_{i-1} + \operatorname{Log}\beta_i$.
$i \leftarrow 0$, $\mathfrak{c}_0 \leftarrow \mathfrak{b}_k$, $\gamma_0 \leftarrow \prod_{j=1}^{k} \beta_j^{2^{k-j}}$, $\lambda \leftarrow \lambda_k$.
while $\lambda < \overline{\delta}$ do
    $S \leftarrow S \cup \{(\mathfrak{c}_i, \gamma_i)\}$.
    $\gamma_{i+1} \leftarrow \gamma(\mathfrak{c}_i)\gamma_i$, $\lambda \leftarrow \lambda + \operatorname{Log}\gamma(\mathfrak{c}_i)$, $\mathfrak{c}_{i+1} \leftarrow \rho(\mathfrak{c}_i)$, $i \leftarrow i + 1$.
$i \leftarrow 0$, $\mathfrak{c}_0 \leftarrow \mathfrak{b}_k$, $\gamma_0 \leftarrow \prod_{j=1}^{k} \beta_j^{2^{k-j}}$, $\lambda \leftarrow \lambda_k$.
while $\lambda > \underline{\delta}$ do
    $S \leftarrow S \cup \{(\mathfrak{c}_i, \gamma_i)\}$.
    $\gamma_{i-1} \leftarrow \gamma'(\mathfrak{c}_i)\gamma_i$, $\lambda \leftarrow \lambda + \operatorname{Log}\gamma'(\mathfrak{c}_i)$, $\mathfrak{c}_{i-1} \leftarrow \rho^{-1}(\mathfrak{c}_i)$, $i \leftarrow i - 1$.
return $S$.
```

Note that the comparisons in the controlling conditions of the while loops should not be performed exactly since this would take time proportional to δ. Rather, they should be performed numerically in such a way that the loop is entered whenever δ_i is smaller than $\overline{\delta}$ (larger than $\underline{\delta}$), but possibly also slightly larger (smaller). This ensures that the output does contain $S_d(\mathfrak{a})$.

Lemma 11.5.18. *Algorithm* `cycleSection` *is correct. Its run-time is bounded by* $O((\log d + \log\log\Delta)(\log\Delta)^3)$.

Proof. The run-time bound follows from Lemmas 11.5.13 and 10.1.6. The proof of correctness is left as an exercise. □

Since it does not affect our analysis we will ignore that `cycleSection` may return a set which is slightly larger than desired. It is easy to introduce constant factors into the bounds of Lemma 11.5.16 in such a way that the divergence of `cycleSection` from its idealized behavior is taken into account.

In order to complete the running time analysis for `randomRelation`, we need to explain how the computation of the reduced ideal $\mathfrak{b}$ and its relative generator β is performed in the third step of the algorithm.

We use an adaptation of `reducePowerProduct` from Section 11.4.1 which we will call `reducePowerProductRQ`.

Algorithm 11.12 `reducePowerProductRQ` $(\Delta, \mathcal{F}, \mathbf{u})$

Input: The discriminant Δ, the power product base $\mathcal{F} = \{\mathfrak{p}_1, \dots, \mathfrak{p}_f\}$, the exponent vector $\mathbf{u}$ and bound k satisfying $\log_2|\mathbf{u}|_\infty < k$.
Output: A reduced ideal $\mathfrak{r}$ and $\alpha \in F$ such that $\mathfrak{r} = \alpha\mathcal{F}^{\mathbf{u}}$.

```
Initialize $\mathfrak{r} \leftarrow (1)$, $\alpha \leftarrow 1$.
for $(i \leftarrow 1, i \le f, i \leftarrow i+1)$ do
  if $u_i = 0$ then
    Initialize $\mathfrak{q} \leftarrow (1)$, $\beta \leftarrow 1$.
  else
    Initialize $\mathfrak{q} \leftarrow \mathfrak{p}_i$, $\beta \leftarrow 1$.
  for $(j \leftarrow \max(j \mid u_{ij}) - 1, j \ge 0, j \leftarrow j-1)$ do
    $(\mathfrak{q}, \gamma) \leftarrow$ reduce($\mathfrak{q}^2$).
    $\beta \leftarrow \beta^2\gamma$.
    if $u_{ij} = 1$ then
      $(\mathfrak{q}, \gamma) \leftarrow$ reduce($\mathfrak{q} \cdot \mathfrak{p}_i$).
      $\beta \leftarrow \beta\gamma$.
  $(\mathfrak{r}, \gamma) \leftarrow$ reduce($\mathfrak{r} \cdot \mathfrak{q}$).
  $\alpha \leftarrow \alpha\beta\gamma$.
return $(\mathfrak{r}, \alpha)$.
```

Note that `reducePowerProductRQ` should not compute either standard or compact representation of the relative generator α it outputs. The norm of $d(\alpha)\alpha$ is too large. Instead it keeps a list of all factors it encounters, and the exponents with which they enter into α. Alternatively, `reducePowerProductRQ` could be modified to output only (an approximation to) the length of the relative generator instead of the generator itself, given a maximal admissible error for this quantity.

Lemma 11.5.19. *The relative generator output of* `reducePowerProductRQ` *is a power product of no more than* $2f\lfloor\log_2|\mathbf{u}|_\infty\rfloor$ *factors* γ_i, $i = 1, \dots$ *with exponents bounded by* $|\mathbf{u}|_\infty$. *The denominator of each* γ_i *and the height of* $d(\gamma_i)\gamma_i$ *are bounded by* $|\Delta|$.

Proof. The number of factors and the bound for their exponents are obvious. The bounds for denominator and height follow from Lemma 11.5.10. □

Lemma 11.5.20. *The running time for* `reducePowerProductRQ` *is bounded from above by* $O(f(\log|\mathbf{u}|_\infty)(\log\Delta)^2)$.

Proof. The lemma is obtained by counting the number of calls to `reduce` and using Theorem 5.6.6.

Corollary 11.5.21. *The running time for* `randomRelation` *is bounded by* $L_\Delta[1/2, z + o(1)]$. □

11.5.5 Computing the Extended Relation Lattice

The computation of the extended relation lattice $\tilde{L}(\mathcal{F})$ proceeds in three steps. In the first step we compute $f = |\mathcal{F}|$ extended relations such that their integral parts form a diagonally dominant matrix, and thus generate a full rank sublattice M of $L([\mathcal{F}])$. In the second step we compute a sufficiently large set N of relations such that their integral parts together with M generate $L([\mathcal{F}])$ with high probability. These two steps are similar to the corresponding ones in the imaginary case. In a third step we compute two more relations $\mathbf{v}$ and $\mathbf{w}$. If we subtract from these relations a suitable linear combination of vectors in $M \cup N$, we obtain vectors whose integral parts are zero, and, hence, their real parts contain lengths of units. We will be able to show that these units generate the unit group with non-negligible probability. Thus $M \cup N \cup \{\mathbf{v}, \mathbf{w}\}$ generates $\tilde{L}(\mathcal{F})$.

We proceed through the steps described. First, we give the adaptation of `fullRank` from section 11.4.2.

We extend $g(\Delta, z)$ from section 11.4.2 to the domain with $\Delta > 0$ in such a way that for fixed z it goes to 0 as Δ goes to infinity and such that the probability from Lemma 11.5.17 is at least

$$p(\Delta, z) = L_\Delta[1/2, -1/(4z) + g(\Delta, z)]\ . \tag{11.39}$$

Set

$$B_1(\Delta) = (f-1)\Delta + \log \Delta\ . \tag{11.40}$$

The listing of `fullRank` can be found on page 262. Bounds for success probability and running time are given in the following two lemmas whose proof is mutatis mutandis the same as in the imaginary-quadratic case.

Proposition 11.5.22. *Algorithm* `fullRank` *outputs* `nil` *with probability smaller than* 3/4*. If* `fullRank` *does not output* `nil`*, then* `fullRank` *outputs a relation matrix whose integral part is diagonally dominant.* □

The running time of `fullRank` can also be bound as in the imaginary-quadratic case if we take into account Lemma 11.5.18.

Proposition 11.5.23. *The running time of Algorithm* `fullRank` *is bounded by* $L_\Delta[1/2, 2z + 1/(4z) + o(1)]$. □

Algorithm 11.13 `fullRank` $(\Delta, \mathcal{F}, z)$

Input: Discriminant Δ, factor base $\mathcal{F}$, parameter z.
Output: `nil` or relations $(\mathbf{z}_1, d_1), \ldots, (\mathbf{z}_f, d_f) \in \tilde{L}(\mathcal{F})$ such that the matrix $(\mathbf{z}_1, \ldots, \mathbf{z}_f)$ is strictly diagonally dominant.

```
$l = \lceil (\log f)/p(\Delta, z) \rceil$.
for $i = 1, 2, \ldots, f$ do
    $j \leftarrow 0$
    repeat
        $j \leftarrow j + 1$
        $(\mathbf{z}_i, d_i) \leftarrow$ randomRelation$(\Delta, \mathcal{F}, B_1(\Delta)\mathbf{e}_i)$
    until $j = l$ or $\mathbf{z}_i \neq$ nil
    if $\mathbf{z}_i =$ nil then
        return nil.
return $(\mathbf{z}_1, d_1), \ldots, (\mathbf{z}_f, d_f)$.
```

Assume we have applied `fullRank` and have computed the relation matrix

$$\tilde{Z} = \begin{pmatrix} \mathbf{z}_1 & \cdots & \mathbf{z}_f \\ d_1 & \cdots & d_f \end{pmatrix} \quad \text{with integral part } Z = (\mathbf{z}_1 \cdots \mathbf{z}_f) \; .$$

Again, we want to compute sufficiently many additional relations

$$\begin{pmatrix} \mathbf{z}_{f+1} & \cdots & \mathbf{z}_{f+N} \\ d_{f+1} & \cdots & d_{f+N} \end{pmatrix}$$

such that the full sequence $S = (\mathbf{z}_1, \ldots, \mathbf{z}_{f+N})$ of integral parts of all computed relations generates $L([\mathcal{F}])$.

We can bound the determinant of Z the same way as in the imaginary-quadratic case. With $B_2(\Delta)$ as defined in (11.20) we have $|\det Z| \leq B_2(\Delta)^f$. Hence we need to extend $\tilde{Z}$ at most $f \log_2(B_2(\Delta))$ times to get a relation lattice whose integral part equals $L([\mathcal{F}])$.

Lemma 11.5.24. *Let M be a proper sublattice of $L([\mathcal{F}])$. Suppose that Algorithm* `randomRelation` *outputs a relation $(\mathbf{z}, d)$. Then the probability for $\mathbf{z}$ to be outside of M is in $\Omega(1/(1 + \log \Delta))$.*

Proof. We need to adapt the proof of Lemma 11.4.11 only slightly taking into account that starting from a given exponent vector we arrive at a fixed ideal no longer deterministically, but only with a probability that we have bounded from above and below.

Suppose that `randomRelation` is successful. Then the algorithm has found an $\mathcal{F}$-smooth reduced $\mathcal{O}$-ideal $\mathfrak{b} = \mathcal{F}^{\mathbf{b}}$. Fix one such $\mathfrak{b}$ and denote by S the set of vectors $\mathbf{v} \in \mathbb{Z}^f_{0..\Delta-1}$ for which $\mathcal{F}^v \sim \mathfrak{b}$. `randomRelation` has selected one of those vectors and found a relation $(\mathbf{v} - \mathbf{b}, d)$. By N denote the number of vectors in S such that $\mathbf{v} - \mathbf{b}$ is outside of M. Set

$$S' = S \cap \mathbb{Z}^f_{0..\Delta - h_\Delta - 1} \, . \tag{11.41}$$

We prove

$$N \geq |S'|/2 \, . \tag{11.42}$$

as we did in (11.22).

The conditional probability which we want is $|N|/|S|$ multiplied with the minimum probability to reach $\mathfrak{b}$ from a vector in S divided by the maximum probability to reach $\mathfrak{b}$ from $S \setminus N$. By Lemma 11.5.16 the latter fraction is bounded from below by

$$\frac{\log 2 - R_\Delta/\Delta}{\log \Delta + R_\Delta/\Delta} \, . \tag{11.43}$$

The first factor $|N|/|S|$ is again at least $|S'|/(2|S|)$, and thus, by Lemma 11.4.5, larger than

$$\frac{|S'|}{|S|} \geq \frac{(1 - (h_\Delta - 1)/(\Delta - h_\Delta))}{(1 + (h_\Delta - 1)/\Delta)} \cdot \frac{(\Delta - h_\Delta)^{f-1}}{\Delta^{f-1}} \, . \tag{11.44}$$

We apply Corollary 9.3.12 and Lemmas 11.3.3 and Lemma 11.2.1 and see that the product of (11.43) and (11.44) is, for sufficiently large Δ, larger than $c/(1 + \log \Delta)$ for some suitable c. □

Let $p_{\text{new}}(\Delta, z)$ be a function which is for fixed z in $\Omega(1/(1 + \log \Delta))$ such that the probability in Lemma 11.5.24 is for any proper sub-lattice M of $L([\mathcal{F}])$ at least $p_{\text{new}}(\Delta, z)$. We are now ready to present the real-quadratic version of `relationLattice`.

Algorithm 11.14 `relationLattice` $(\Delta, \mathcal{F}, z)$

Input: The discriminant Δ, the factor base $\mathcal{F}$ with cardinality f, the parameter z.
Output: `nil` or relations $\mathbf{z}_1, \ldots, \mathbf{z}_N \in L[\mathcal{F}]$.

```
$N \leftarrow 1, i \leftarrow 0, k \leftarrow \lceil f \log_2 B_2(\Delta) \rceil, n \leftarrow \lceil p_{\text{new}}(\Delta, z) \log k \rceil, l \leftarrow \lceil (\log kn)/p(\Delta, z) \rceil$
repeat
  $i \leftarrow i + 1$
  $\mathbf{z}_{N+1} \leftarrow$ randomRelation$(\Delta, \mathcal{F}, 0)$
  if $(\mathbf{z}_{N+1}, d_{N+1}) \neq$ nil then
    $N \leftarrow N + 1$
until $i \geq kln$ or $N \geq kn$
if $N < kn$ then
  return nil
else
  return $(\mathbf{z}_1, d_1), \ldots, (\mathbf{z}_N, d_N)$
```

We can bound the success probability and running time of `relationLattice` as in the imaginary-quadratic case.

Proposition 11.5.25. *Algorithm* `relationLattice` *outputs* `nil` *with probability smaller than* 3/4. *Assume that* `relationLattice` *outputs* $(\mathbf{z}_{f+1}, d_{f+1}), \ldots, (\mathbf{z}_{f+N}, d_{f+N})$. *Then the probability that* $Z \cup \{\mathbf{z}_{f+1}, \ldots, \mathbf{z}_{f+N}\}$ *generates* $L([\mathcal{F}])$ *is larger than* 1/4. □

Proposition 11.5.26. *The running time of* `relationLattice` *is bounded by* $L[1/2, 2z + 1/(4z) + o(1)]$. □

Assume now that we have a set of extended relations $(\mathbf{z}_1, d_1), \ldots, (\mathbf{z}_N, d_N)$ whose integral parts generate $L([\mathcal{F}])$. Call the matrix containing $\mathbf{z}_1, \ldots, \mathbf{z}_N$ as column vectors A. Recall that $N = L_\Delta[1/2, z + o(1)]$. Assume, moreover, that two additional calls to `randomRelation` yielded two more extended relations $(\mathbf{z}_{N+i}, d_{N+i})$, $i = 1, 2$.

Then there exist $\mathbf{x}_i$ such that

$$A\mathbf{x}_i = \mathbf{v}_{N+i} \ , \quad i = 1, 2 \ . \tag{11.45}$$

Apply an algorithm satisfying the properties given in Proposition A.5.20 – call it `dSolv`– to find such $\mathbf{x}_i$. Since $|A|_\infty, |\mathbf{v}_i|_\infty < f\Delta + \log \Delta$ we get from Proposition A.5.20

$$\log|\mathbf{x}_i|_\infty = O(N \cdot \log \max(|A|_\infty, |\mathbf{v}_i|_\infty)) = L_\Delta[1/2, z + o(1)] \ . \tag{11.46}$$

Consider the real numbers

$$R_i = d_{N+i} - \sum_{j=1}^{N} x_{i,j} d_j \ , \quad i = 1, 2 \ . \tag{11.47}$$

Because of (11.45), $R_i = \operatorname{Log} \varepsilon_i$ for two units $\varepsilon_i = \varepsilon_\Delta^{a_i}$, $i = 1, 2$.

Let $\gcd(a_1, a_2) = y_1 a_1 + y_2 a_2 = a$. Then $y_1 l_1 + y_2 l_2 = a R_\Delta$. Assume that we have computed h_Δ from A as in section 11.4.4. Assume, moreover, that we have computed an approximation

$$\tilde{l} = \sqrt{\Delta}/2 \cdot l(\Delta, n) \tag{11.48}$$

to $\sqrt{\Delta}/2L(1, \chi_\Delta)$ which by Theorem 9.3.10 equals $h_\Delta R_\Delta$. Then integers y_i can be determined from R_i by computing the continued fraction expansion of l_1/l_2. The details of this procedure are explained in [Mau00], Chapter 12, where the algorithm is called `rgcd_cfrac`. For our purposes we require as output only the recombined regulator multiple aR_Δ.

Algorithm 11.15 `regulator` $(\Delta, \mathcal{F}, S, z)$

Input: Discriminant Δ, factor base $\mathcal{F}$ with cardinality f, set S of extended relations $(\mathbf{z}_1, d_1), \ldots, (\mathbf{z}_N, d_N)$, parameter z.
Output: `nil` or R, an approximation to the regulator R_Δ.

```
$l = \lceil (\log 2)/p(\Delta, z) \rceil$.
for $i = 1, 2$ do
  $j \leftarrow 0$
  repeat
    $j \leftarrow j + 1$
    $(\mathbf{z}_{N+i}, d_{N+i}) \leftarrow$ randomRelation$(\Delta, \mathcal{F}, 0)$
  until $j = l$ or $\mathbf{z}_i \neq$ nil
  if $\mathbf{z}_i =$ nil then
    return nil.
$\mathbf{x}_i \leftarrow$ dSolv$(\mathbf{z}_1, \ldots, \mathbf{z}_N; \mathbf{z}_{N+i})$, $i = 1, 2$.
$R_i \leftarrow d_{N+i} - \sum_{j=1}^{N} x_{i,j} d_j$, $i = 1, 2$.
return $R \leftarrow$ rgcd_cfrac$(R_1, R_2)$.
```

The following lemma helps us to bound the probability that $a = 1$.

Lemma 11.5.27. *Let $A, B, M \in \mathbb{Z}$ with $\log(|A - B| + 1) < M/100$. Consider the set $S = \{ (x, y) \in \mathbb{Z}^2 \mid A \leq x < A + M, B \leq y < B + M \}$. If $M \gg 0$ then there are more than $M^2/2$ pairs $(x, y) \in S$ with $\gcd(x, y) = 1$.*

Proof. We define the following subsets of S:

$$T = \{ (x, y) \in S \mid \gcd(x, y) \neq 1 \},$$
$$T_p = \{ (x, y) \in S \mid p \mid \gcd(x, y) \}$$

where p denotes some prime number. We need to show that $|T| < M^2/2$. We will show instead that

$$\sum_{p \leq M} |T_p| + |\bigcup_{p > M} T_p| < M^2/2$$

which is certainly sufficient.

Let $p \leq M$. Then a simple counting argument shows that $|T_p| < (1 + \lfloor M/p \rfloor)^2$. Thus

$$\begin{aligned} \sum_{p \leq M} |T_p| &< \sum_{p \leq M} (1 + M/p)^2 \\ &< M(\log\log M + O(1)) + M^2 P(2) \end{aligned}$$

where P is the prime zeta function, and $P(2) = 0.452....$

Let $p > M$. Then $|T_p| \leq 1$. For any $d \in \mathbb{Z}$ we define yet another set $U_d = \{ (x, y) \in S \mid x - y = d \}$. Assume $T_p \cap U_d \neq \emptyset$. Then $p \mid d$. Moreover,

$|d| < |A - B| + M$ by the definition of S, and d has no more than $\log(|A - B| + M)$ prime divisors. Hence

$$|U_d \cap \bigcup_{p>M} T_p| < \log(|A - B| + M) \ .$$

From this we deduce

$$\begin{aligned} |\bigcup_{p>M} T_p| &= \sum_{d=A-B-M}^{A-B+M} |U_d \cap \bigcup_{p>M} T_p| \\ &< 2M(\log(|A - B| + M)) < M^2/50 + M \log M \ . \end{aligned}$$

Adding the two estimates we obtain the desired result for sufficiently large M. □

Proposition 11.5.28. *Algorithm* `regulator` *outputs* `nil` *with probability smaller than* $3/4$. *For sufficiently large* Δ, *if* `regulator` *does not output* `nil`, *then it outputs an approximation to the regulator* R_Δ *with probability larger* $1/2$.

Proof. We omit the proof of the given bound for the success probability of `regulator` since it is the same as for `fullRank`. Assume that `regulator` has output some R.

We estimate the probability that $\gcd(a_1, a_2) = 1$. It suffices to do so conditional on $\mathbf{z}_{N+i}$, $\mathbf{x}_i$, $i = 1$, 2, being some fixed vectors. In this case, R_i depends on the distance parameters drawn in the i-th call to `cycleSection` that leads to a successful completion of `randomRelation`, according to (11.47). It follows that each a_i varies in an interval $[A_i, C_i]$ of length $\Delta \log \Delta / R_\Delta$. The size bound (11.46) implies that

$$\log |A_1 - A_2| = L_\Delta[1/2, z + o(1)] \ .$$

For sufficiently large Δ, this is smaller than $(\Delta \log \Delta)/(100 R_\Delta)$. Thus, we can apply Lemma 11.5.27 and obtain the desired probability bound. □

Proposition 11.5.29. *Algorithm* `regulator` *executes in time bounded by* $L_\Delta[1/2, \max(z + 1/(4z), 3z) + o(1)]$.

Proof. The algorithm `regulator` executes $l = L_\Delta[1/2, 1/(4z) + o(1)]$ calls to `randomRelation`, two to `dSolv`, and one to `rgcd_cfrac`.

The first sub-algorithm requires time $L_\Delta[1/2, z+o(1)]$ by Corollary 11.5.21, the second $L_\Delta[1/2, 3z + o(1)]$ by Proposition A.5.20. According to [Mau00], the running time of `rgcd_cfrac` is bounded by $O(k^2)$ with $k = \log R_1 R_2$. The R_i, $i = 1, 2$, were defined by (11.47). Applying Lemma 11.5.19 and 11.5.6, we get that $d_i = \Delta \cdot L_\Delta[1/2, 2z + o(1)]$. Combinining this bound with (11.46), we see that R_i, $i = 1, 2$, can be computed in time $L_\Delta[1/2, 2z + o(1)]$ (with error in $O(1)$ which is sufficient). We also get $k = L_\Delta[1/2, z + o(1)]$ so that the call to `rgcd_cfrac` costs time $L_\Delta[1/2, 2z + o(1)]$.

Taking all these running time bounds together we obtain the bound in the proposition. □

Theorem 11.5.30. *Class number and regulator of a real quadratic order can be computed in time* $L_\Delta[1/2, \sqrt{2} + o(1)]$.

Proof. This computation is effected by successive execution of `fullRank`, `relationLattice`, a HNF (or determinant) computation on the matrix of integral parts of the relations found, and `regulator`. According to Propositions 11.5.23, 11.5.26, and A.5.3, and Lemma 11.5.29 this takes time $L_\Delta[1/2, 2z + 1/(4z) + o(1)]$, $L_\Delta[1/2, 2z + 1/(4z) + o(1)]$, $L_\Delta[1/2, 4z + o(1)]$, and $L_\Delta[1/2, \min(z + 1/(4z), 3z) + o(1)]$, respectively. This is minimized for $z = 1/\sqrt{8}$ yielding the time bound given in the theorem. □

We close this section by sketching a solution to the equivalence problem. Given two ideals $\mathfrak{a}$ and $\mathfrak{b}$ we would like to decide whether they are equivalent, and if so, find a relative generator. By computing $(\mathfrak{c}, \alpha) =$ `reduce`$(\mathfrak{a}/\mathfrak{b})$, we reduce this problem to the principal ideal problem (PIP) for $\mathfrak{c}$: It suffices to determine whether $\mathfrak{c}$ is principal and, if so, find a generator γ for it. In the latter case, we have $\mathfrak{b} = (\alpha/\gamma)\mathfrak{a}$.

In order to solve the PIP for $\mathfrak{c}$ we include this ideal in the factor base $\mathcal{F}$ (at the beginning of the sequence), and call `fullRank` and `relationLattice`. Let $(\mathbf{z}_1, d_1), \ldots, (\mathbf{z}_N, d_N)$ be the relations obtained. Then, $\mathfrak{c}$ is principal if and only if there is $\mathbf{x} \in \mathbb{Z}^N$ such that

$$(\mathbf{x}_1, \ldots, \mathbf{z}_N)\mathbf{x} = \mathbf{e}_1$$

where $\mathbf{e}_1 = (1, 0, \ldots, 0)$. We apply `dSolv` to find $\mathbf{x}$. If it does find a solution, then $\mathfrak{c}$ has a generator of length

$$d = \sum_{i=1}^{N} x_i \cdot d_i \; .$$

Finally, we apply `compactRepresentation` to $d \bmod R_\Delta$ to obtain a generator of $\mathfrak{c}$ in compact representation.

11.6 Practice

In practice, the algorithms are not used as specified in the preceding sections. Many improvements can be applied.

The easiest measure is to include only one of each pair of conjugated prime ideals in the factor base.

In `randomRelation` only the first $c_3(\Delta)$ entries of $\mathbf{v}$ need to be chosen randomly, the rest can be set to zero. This still ensures that $[\mathcal{F}^{\mathbf{v}}]$ is nearly uniformly chosen from Cl_Δ. The reduction of non-zero entries in the relation matrix speeds up the linear algebra step considerably.

The number of relations computed by algorithm `relationLattice` can be chosen to be only slightly larger than f. A more careful analysis than the one we have given here for simplicity of argument will show that the relations we obtain from `fullRank` and the thus modified `relationLattice` will still generate $L([\mathcal{F}])$ with great likelihood (constant if $\Delta < 0$, or $\Omega(1/\log(\Delta))$ if $\Delta > 0$). In fact, practical computations seem to indicate that `relationLattice` need only compute a constant number of relations.

Moreover, we do not need that the lattice spanned by the relation vectors the algorithm computes contain all relations between elements of the factor base. It suffices if this lattice contains all relations between elements of a generating system $\mathcal{G}$ for the class group. This further reduces the number of relations we need to compute.

The linear algebra step can then be adapted to compute the Hermite Normal Form of the lattice of relations between the elements of $\mathcal{G}$. The resulting algorithm is cubic instead of quartic in the size of the matrix.

Finally, relation computation can be sped up by not trial dividing each number N(A) occuring in a call to `randomRelation`. Instead, one can collect all (or large batches of) these numbers and factor them simultaneously using Bernstein's algorithm [Ber] in average polynomial time. Remember, trial division costs time $O(|\mathcal{F}|)$.

Taking all these measures together one can prove a bound of $L_{|\Delta|}[1/2, \sqrt{9/8} + o(1)]$ for the expected running time of `classGroup` [Vol02].

We also note that there are heuristic approaches to speedier relation generation which use sieving, cf. [Jac99a, Jac99b]. These are used in most modern computer algebra packages which allow the computation of class numbers and regulators of number fields. Their running time behavior corresponds roughly to the one expected from the theoretical algorithms.

11.7 Exercises

Exercise 11.7.1. Prove the existence of ε and $\Delta(z)$ in the proof of Proposition 11.4.4.

Exercise 11.7.2. Prove (11.10).

Exercise 11.7.3. Use Lemma 9.3.15 and Lemma 11.4.5 to prove that with S and S' from the proof of Lemma 11.4.11 we have $|S|/(2|S|') \geq 1/c$ for some efficiently computable positive integer c.

Exercise 11.7.4. Produce a listing for algorithm `reachReduced`.

Exercise 11.7.5. Let $\mathcal{O}$ be an imaginary quadratic order with discriminant Δ, and $\mathfrak{a}$ and $\mathfrak{b}$ be two $\mathcal{O}$-ideals. The goal of this exercise is to find an algorithm that solves the following two tasks: 1. Determine whether there exists $n \geq 0$ such that $\mathfrak{a} \sim \mathfrak{b}^n$, 2. in the positive case, determine one such n. The smallest non-negative number n is called the *discrete logarithm* of $[\mathfrak{a}]$ to the basis $[\mathfrak{b}]$. Obviously, we may restrict ourselves to the case that $\mathfrak{a}$ and $\mathfrak{b}$ are reduced.

Let $\mathcal{F} = (\mathfrak{b}, \mathfrak{a}, \mathfrak{p}_1, \ldots, \mathfrak{p}_f)$ with $\mathfrak{p}_i$ from (11.6). The results from Sections 11.4.1 through 11.4.3 show that we may compute a set of generators for the lattice of relations on $\mathcal{G}$ with cardinality $L_{|\Delta|}[1/2, z + o(1)]$ in time $L_{|\Delta|}[1/2, 2z + 1/(4z) + o(1)]$.

Use Proposition A.5.20 to find the order of $\mathfrak{b}$ in Cl_Δ, to determine whether $[\mathfrak{a}] \in \langle [\mathfrak{b}] \rangle$, and, if it does, to find the discrete logarithm n. Find a bound for the running time of the resulting algorithm.

Exercise 11.7.6. In this exercise, we will extend the result of the previous exercise to the real quadratic case.

Let $\mathcal{O}$ be a real quadratic order with discriminant Δ, and $\mathfrak{a}$ and $\mathfrak{b}$ be two $\mathcal{O}$-ideals. We seek to solve the following tasks: 1. Determine whether there exists $n \geq 0$ such that $\mathfrak{a} \sim \mathfrak{b}^n$, 2. in the positive case, determine one such n, and 3. find α (in compact representation) such that

$$\mathfrak{a} = \alpha \cdot \mathfrak{b}^n .$$

The smallest non-negative number n is called again the *discrete logarithm* of $\mathfrak{a}$ with respect to the basis $\mathfrak{b}$, and α its relative generator. Obviously, we can again restrict ourselves to the case that $\mathfrak{a}$ and $\mathfrak{b}$ are reduced.

Proceed as in the previous exercise to arrive at an algorithm that computes an extended relation of the form $(n, 1, 0, \ldots, 0, d)^{\mathrm{T}}$. Give a size bound for the last entry d output by your algorithm, and use Lemma 11.5.14 to prove a run-time bound for the conversion of d into a compact representation of the sought relative generator α.

Chapter references and further reading

[Abe94] Christine Abel, *Ein Algorithmus zur Berechnung der Klassenzahl und des Regulators reellquadratischer Ordnungen*, Ph.D. thesis, Universität des Saarlandes, Saarbrücken, Germany, 1994, German.

[BD89] Johannes Buchmann and Stephan Düllmann, *A probabilistic class group and regulator algorithm and its implementation*, Proc. Colloquium on Computational Number Theory, Walter de Gruyter, 1989, pp. 5–72.

[BD91a] ______, *Distributed class group computation*, Informatik (Johannes Buchmann, Harald Ganzinger, and Wolfgang J. Paul, eds.), Teubner-Texte zur Informatik, vol. 1, B. G. Teubner, 1991, pp. 68–81.

[BD91b] ______, *On the computation of discrete logarithms in class groups*, Advances in Cryptology – CRYPTO '90 (Alfred J. Menezes and Scott A. Vanstone, eds.), Lecture Notes in Computer Science, vol. 537, Springer-Verlag, 1991, pp. 134–139.

[Ber] Daniel J. Bernstein, *How to find small factors of integers*, Mathematics of Computation (to appear), `http://cr.yp.to/papers/sf.ps`.

[BH96] Johannes Buchmann and Christine S. Hollinger, *On smooth ideals in number fields*, Journal of Number Theory **59** (1996), no. 1, 82–87.

[BH03] Mark L. Bauer and Safuat Hamdy, *On class group computations using the number field sieve*, Advances in Cryptology – ASIACRYPT 2003 (Chi-Sung Laih, ed.), Lecture Notes in Computer Science, vol. 2894, Springer-Verlag, 2003, pp. 311–325.

[BJN$^+$98] Johannes Buchmann, Michael J. Jacobson, Jr., Stefan Neis, Patrick Theobald, and Damian Weber, *Sieving methods for class group computation*, Algorithmic Algebra and Number Theory: Selected Papers from a Conference Held at the Univerity of Heidelberg in October 1997, Springer-Verlag, 1998, pp. 3–10.

[BTW95] Johannes Buchmann, Christoph Thiel, and Hugh C. Williams, *Short representation of quadratic integers*, Computational Algebra and Number Theory, Sydney 1992 (Wieb Bosma and Alf J. van der Poorten, eds.), Mathematics and its Applications, vol. 325, Kluwer Academic Publishers, 1995, pp. 159–185.

[Buc90a] Johannes Buchmann, *Complexity of algorithms in algebraic number theory*, Number Theory, Banff, Alberta 1988 (Richard A. Mollin, ed.), Walter de Gruyter Publishers, 1990, pp. 37–53 (English).

[Buc90b] ______, *A subexponential algorithm for the determination of class groups and regulators of algebraic number fields*, Séminaire de Théorie des Nombres, Paris 1988–1989 (Catherine Goldstein, ed.), Progress in Mathematics, vol. 91, Birkhäuser, 1990, pp. 27–41.

[BW89] Johannes Buchmann and Hugh C. Williams, *On the existence of a short proof for the value of the class number and regulator of a real quadratic field*, Number Theory and Applications, Calgary 1988 (Richard A. Mollin, ed.), NATO ASI Series, Series C, vol. 265, Kluwer Academic Publishers, 1989, pp. 327–345.

[BW91] ______, *Some remarks concerning the complexity of computing class groups of quadratic fields*, J. Complexity **7** (1991), no. 3, 311–315 (English).

[CDO93] Henri Cohen, Francisco Diaz y Diaz, and Michel Olivier, *Calculs de nombres de classes et de régulateurs de corps quadratiques en temps sous-exponentiel*, Séminaire de Théorie des Nombres, Paris 1990–1991 (Sinnou David, ed.), Progress in Mathematics, vol. 108, Birkhäuser, 1993, French, pp. 35–46.

[CDO97] ______, *Subexponential algorithm for class group and unit computation*, Journal of Symbolic Computing **24** (1997), no. 3/4, 433–441.

[HM89] James L. Hafner and Kevin S. McCurley, *A rigorous subexponential algorithm for computation of class groups*, Journal of the American Mathematical Society **2** (1989), no. 4, 837–850 (English).

[HW79] Godfrey H. Hardy and Edward M. Wright, *An introduction to the theory of numbers*, Clarendon Press, Oxford, 1979.

[Jac99a] Michael J. Jacobson, Jr., *Applying sieving to the computation of quadratic class groups*, Mathematics of Computation **68** (1999), no. 226, 859–867.

[Jac99b] ______, *Subexponential class group computation in quadratic orders*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, Darmstadt, Germany, 1999.

[Kra22] Maurice Kraitchik, *Theéorie des nombres*, vol. 1, Gauthier-Villars, 1922.

[Lag80a] Jeffrey C. Lagarias, *On the computational complexity of determining the solvability or unsolvability of the equation $X^2 - dY^2 = -1$*, Transactions of the American Mathematical Society **260** (1980), 485–508.

[Lag80b] ———, *Worst-case complexity bounds for algorithms in the theory of integral quadratic forms*, Journal of Algorithms **1** (1980), 142–186.

[LP92] Hendrik W. Lenstra, Jr. and Carl Pomerance, *A rigorous time bound for factoring integers*, J. Amer. Math. Soc. **5** (1992), 483–516.

[Mau00] Markus Maurer, *Regulator approximation and fundamental unit computation for real quadratic orders*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, Darmstadt, Germany, 2000.

[Nei02] Stefan Neis, *Berechnung von Klassengruppen*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, Darmstadt, Germany, 2002, German.

[Pau96] Sachar Paulus, *An algorithm of subexponential type computing the class group of quadratic orders over principal ideal domains*, Algorithmic Number Theory, ANTS-II (Henri Cohen, ed.), Lecture Notes in Computer Science, vol. 1122, Springer-Verlag, 1996, pp. 243–257 (English).

[Sey87] Martin Seysen, *A probablistic factorization algorithm with quadratic forms of negative discriminant*, Mathematics of Computation **48** (1987), 757–780.

[Tes98a] Edlyn Teske, *New algorithms for finite abelian groups*, Ph.D. thesis, Technische Universität Darmstadt, Germany, 1998, Shaker Verlag, Aachen.

[Tes98b] ———, *A space efficient algorithm for group structure computation*, Math. Comput. **67** (1998), no. 224, 1637–1663 (English).

[Thi94] Christoph Thiel, *Under the assumption of the generalized riemann hypothesis verifying the class number belongs to NP$\cap$ co-NP*, Algorithmic number theory, ANTS-I (Leonard M. Adleman and Ming-Deh Huang, eds.), Lecture Notes in Computer Science, vol. 877, Springer-Verlag, 1994, pp. 234–247 (English).

[Thi95] Christoph Thiel, *On the complexity of some problems in algorithmic algebraic number theory*, Ph.D. thesis, Universität des Saarlandes, Saarbrücken, Germany, 1995.

[Vol00] Ulrich Vollmer, *Asymptotically fast discrete logarithms in quadratic number fields*, Algorithmic Number Theory, ANTS-IV (Wieb Bosma, ed.), Lecture Notes in Computer Science, vol. 1838, Springer-Verlag, 2000, pp. 581–594.

[Vol02] ———, *An accelerated Buchmann algorithm for regulator computation in real quadratic fields*, Algorithmic Number Theory, ANTS-V (Claus Fieker and David R. Kohel, eds.), Lecture Notes in Computer Science, vol. 2369, Springer-Verlag, 2002, pp. 148–162.

[WM68] A. E. Western and J. C. P. Miller, *Tables of indices and primitive roots*, Royal Society Mathematical Tables, Vol. 9, Published for the Royal Society at the Cambridge University Press, London, 1968. MR 39 #7792.

12

Cryptographic Applications

In this chapter, we will discuss several ways in which the theory of binary quadratic forms can be employed for cryptographic applications. Goals of cryptography encompass the maintenance of confidentiality, authenticity, integrity and non-reputability of electronic documents.

We will discuss two types of cryptographic techniques: public-key encryption and digital signatures. Public-key encryption allows the confidential transfer of data without prior exchange of secret key material. Digital signatures allow the verification of integrity of data and the non-reputable attribution of these data to the authentic signer.

A public-key encryption scheme consists of message and ciphertext spaces $\mathcal{M}$ and $\mathcal{C}$, public and private key spaces $\mathcal{K}_{\text{pub}}$ and $\mathcal{K}_{\text{priv}}$, a map

$$\phi : \mathcal{K}_{\text{priv}} \longrightarrow \mathcal{K}_{\text{pub}} \ ,$$

and encryption and decryption maps

$$E : \mathcal{M} \times \mathcal{K}_{\text{pub}} \longrightarrow \mathcal{C} \qquad D : \mathcal{C} \times \mathcal{K}_{\text{priv}} \longrightarrow \mathcal{M}$$

with roughly the following properties

- E, D and ϕ are easy to compute.
- E and D are mutual inverses in the sense that

$$D(E(m, \phi(k)), k) = m.$$

- Given $c = E(m, \phi(k))$ and $\phi(k)$, it is hard to find m. In particular it is hard to find k, i.e. to invert ϕ, or to find a possibly different k' for which

$$D(c, k') = m \ .$$

If a sender—lets call her Alice—wants to send an encrypted message to a recipient—say Bob—she must have access to the public key $\phi(k)$ of Bob. Public keys may be stored in public databases, and their authenticity needs to be maintained in some manner.

Public-key encryption schemes contrast with secret-key encryption schemes which have trivial key maps ϕ so that public and private key spaces coincide, sender and recipient must share a key, and keep it secret.

A digital signature scheme consists of message and signature spaces $\mathcal{M}$ and $\mathcal{S}$, public and private key spaces $\mathcal{K}_{\mathrm{pub}}$ and $\mathcal{K}_{\mathrm{priv}}$, a map

$$\phi : \mathcal{K}_{\mathrm{priv}} \longrightarrow \mathcal{K}_{\mathrm{pub}} ,$$

and signature and verification maps

$$S : \mathcal{M} \times \mathcal{K}_{\mathrm{priv}} \longrightarrow \mathcal{S} \qquad V : \mathcal{M} \times \mathcal{S} \times \mathcal{K}_{\mathrm{pub}} \longrightarrow \{\mathrm{true}, \mathrm{false}\}$$

with roughly the following properties

- S, V and ϕ are easy to compute.
- Authentic signatures verify
$$V(m, S(m,k), \phi(k)) = \mathrm{true} .$$
- For any given public key $\phi(k)$, it is hard to *forge a signature*, i.e. find any pair (m,s) such that
$$V(m, s, \phi(k)) = \mathrm{true} .$$

Again, if a verifier—lets call him Bob—wants to verify a signature as truly pertaining to a signer—call her Alice—he needs to have access to the public key of Alice.

The hardness referred to in the above definitions is measured by the computational effort an attacker needs to expand in order to subvert the mechanism, i.e. decrypt an encrypted message, or forge a signature. Usually an attacker is modeled to have additional resources. Thus he may be allowed to ask the recipient of an encrypted message to encrypt cipher texts of his choosing (with the exception of the one he tries to decrypt). Respectively, he may be allowed to request signatures of documents of his choosing which, of course, he is then not allowed to present as forgeries.

12.1 Problems

In order to prove that the attacker's task is hard, one shows that a successful attacker is also capable of solving a computational problem which is known or believed to be hard. We say the security of the crypto-system is *reduced* to the difficulty of the computational problem. Sometimes this is also phrased as saying that the problem is the *basis* of the security of the system.

The crypto-systems presented in this chapter have one of the following problems as their basis.

Problem 12.1.1 (Discrete Logarithm). For given discriminant Δ, and $\mathcal{O}(\Delta)$-ideals $\mathfrak{a}$ and $\mathfrak{b}$ find $x \in \mathbb{N}$ such that $\mathfrak{a} \sim \mathfrak{b}^x$ if it exists. If $\Delta > 0$, find in addition $\alpha \in K$, the fractional field of $\mathcal{O}(\Delta)$, such that $\mathfrak{a} = \alpha \mathfrak{b}^x$.

If in the previous problem, the discriminant is negative, or, if $\Delta > 0$, the relative generator α is disregarded, then we speak of the *Class group Discrete Logarithm Problem* (ClDLP). If $\mathfrak{b}$ is chosen to equal $\mathcal{O}$ then one calls the resulting problem the *order problem.* If $\Delta > 0$ and $\mathfrak{a}$ is known to be principal, the resulting problem is called *principal ideal problem.*

The following problem has its root in the famous key exchange protocol proposed by Whitfield Diffie and Martin Hellman [DH76]. It is weaker than the DL problem in the sense that anybody able to solve the DL problem is also able to solve the Diffie-Hellman (DH) problem, with about the same effort.

Problem 12.1.2 (Class group Diffie-Hellman Problem). Let a discriminant Δ and a $\mathcal{O}(\Delta)$-ideal $\mathfrak{a}$ be given. For two powers $\mathfrak{a}^a$ and $\mathfrak{a}^b$ of $\mathfrak{a}$ with unknown exponents a and b coming from some large finite subset of $\mathbb{Z}$, find $\mathfrak{a}^{ab}$.

For the class group Discrete Logarithm and the Diffie-Hellman problem to be difficult, it is necessary that the cardinality of the class group be sufficiently large. Real quadratic orders with large discriminants have typically small class groups, and large regulators. Thus for these orders, we need to formulate an analogue to the Diffie-Hellman Problem which works in the infrastructure.

Problem 12.1.3 (Infrastructure Diffie-Hellman Problem). Let a discriminant $\Delta > 0$ and a principal reduced $\mathcal{O}(\Delta)$-ideal $\mathfrak{g}$ be given. Denote the fraction field of $\mathcal{O}(\Delta)$ by K. Let further $\delta > 1/2 \log \Delta$ be given. For two reduced ideals $\mathfrak{a}$ and $\mathfrak{b}$ for which such $\alpha, \beta \in K$ exist that

$$\mathfrak{a} = \alpha \mathfrak{g}^a \ , \quad \mathfrak{b} = \beta \mathfrak{g}^b \ , \qquad |\operatorname{Log} \alpha| < \delta \ , \quad \text{and } |\operatorname{Log} \beta| < \delta$$

with unknown exponents a and b coming from some large finite subset of $\mathbb{Z}$, find reduced $\mathfrak{c}$ for which such $\gamma \in K$ exists that $\mathfrak{c} = \gamma \mathfrak{g}^{ab}$, and $|\operatorname{Log} \gamma| < \delta$.

Another problem which is weaker than the DL problem is the Root Problem (RP).

Problem 12.1.4 (Class Group Root Problem). For given discriminant Δ, $\mathcal{O}(\Delta)$-ideal $\mathfrak{a}$ and exponent $r \in \mathbb{N}$ which is prime to the class number h_Δ, find the $\mathcal{O}(\Delta)$-ideal $\mathfrak{b}$ for which $\mathfrak{a} \sim \mathfrak{b}^r$.

An analogue to the Class Group Root Problem which operates in the infrastructure of a real quadratic order can be formulated, but has not found use as the basis of a cryptographic protocol. Hence we omit it here, and leave it as an Exercise 12.5.1.

The reduction from Class Group Discrete Logarithm to Root Problem proceeds as follows. Assume we are given an algorithm which solves instances of the DL problem. We want to solve an instance of the root problem given by Δ, $\mathfrak{a}$ and r. Apply the DL algorithm to Δ, $\mathfrak{a}^{-1}$ and $\mathfrak{a}$. We obtain some $q \in \mathbb{N}$ such that $\mathfrak{a}^{q+1}$ is principal. Write $q + 1 = r^k n$ with $\gcd(n, r) = 1$. Then

$\mathfrak{a}^n$ is also principal since $\gcd(r, h_\Delta) = 1$. Let s be the inverse of r modulo n, i.e., $rs = tn + 1$ with integral t. Set $\mathfrak{b} = \mathfrak{a}^s$. Then

$$\mathfrak{b}^r = (\mathfrak{a}^s)^r = \mathfrak{a}^{tn+1} \sim \mathfrak{a} \ .$$

One can adapt Algorithm 11.6 in Chapter 11 which computes the class group of an imaginary quadratic order to solve the discrete logarithm problem, cf. exercise 11.7.5, and hence also the Diffie-Hellman and root problems. Basically, one adds argument and base of the discrete logarithm to the factor base and computes the full relation lattice as before. This relation lattice contains a vector which corresponds to the discrete logarithm relation $\mathfrak{a}\mathfrak{b}^{-k} = (1)$. If one applies the improvements to these algorithms listed in section 11.6, then the run-time of these algorithms is asymptotically in $L_{|\Delta|}[1/2, \sqrt{9/8} + o(1)]$.

Note that the imaginary quadratic class group algorithm can be adapted to solve the Root Problem more efficiently. The adaptation works as follows. Compute the relation lattice for the factor base

$$\mathcal{F} = \{\mathfrak{a}\} \cup \mathcal{F}_z = \{\mathfrak{a}, \mathfrak{p}_1, \ldots, \mathfrak{p}_f\}$$

with cardinality $f + 1$ where $\mathcal{F}_z$ was defined in Section 11.3. This can be done with algorithm `relationLattice` in time $L_{|\Delta|}[1/2, 2z + 1/(4z) + o(1)]$, cf. Propositions 11.4.13 and 11.5.26. If the improvements mentioned in Section 11.6 are employed, then this time bound drops to $L_{|\Delta|}[1/2, z + 1/(4z) + o(1)]$. Let A be the $(f+1) \times m$-matrix containing all relations obtained. Compute $\mathbf{x} \in \mathbb{Z}^m$ such that

$$\mathsf{A}\mathbf{x} \equiv (1, 0, \ldots, 0)^{\mathrm{T}} \mod r \ . \tag{12.1}$$

If

$$\mathsf{A}\mathbf{x} = (1 + k_0 r, k_1 r \ldots, k_f r)^{\mathrm{T}} \ , \quad k_i \in \mathbb{Z} \ ,$$

then

$$\mathfrak{a} \sim \mathfrak{b}^r \qquad \text{with} \quad \mathfrak{b} = \sigma(\mathfrak{a})^{k_0} \prod_{i=1}^{f} \sigma(\mathfrak{p}_i)^{k_i} \ ,$$

If the matrix A is sparse which can be achieved by the methods mentioned in Section 11.6, then the solution of 12.1 can be obtained in quadratic time, using a method originally proposed by Douglas Wiedemann [Wie86]. Thus the run-time bound for the linear algebra part of the algorithm is $L_{|\Delta|}[1/2, 2z + o(1)]$. Hence the total run-time can be bounded by $L_{|\Delta|}[1/2, \max(2z, z + 1/(4z)) + o(1)]$. This is minimized for $z = 1/2$ leading to the bound $L_{|\Delta|}[1/2, 1 + o(1)]$.

While the algorithms given here and in Chapter 11 are asymptotically fast, in practice algorithms are used for which run-time bounds are not known. The most efficient of these algorithms is the Multiple Polynomial Quadratic Sieve by Michael Jacobson [Jac99].

For $\Delta < 0$, he reports running times of less than an hour for the computation of the structure of class groups of random 40 digit discriminants, and running (CPU) times of less than 10 days for special 80 digit discriminants on a 296 MHz SUN UltraSPARC-II platform. The running time of his algorithms can likely be improved using the optimized linear algebra techniques of [GJS01].

For $\Delta > 0$, Jacobson *et al.* report in [JSW01] that using Jacobson's MPQS, the regulator of an order with 101-digit discriminant was computed in 3.8 years of CPU time on 550 MHz Pentium III machines.

12.2 Cryptographic algorithms in imaginary-quadratic orders

The intractable problems described in the previous section can be used as the security basis for several cryptographic algorithms.

Since the Diffie-Hellman problem in the class group is intractable, the Diffie Hellman key exchange protocol [DH76] and the Diffie Hellman integrated encryption scheme DHIES [BR97], a simple extension of the ElGamal encryption scheme [ElG85], can be implemented in this group yielding variants called IQ-DH and IQ-IES.

We give a brief description of IQ-DH and IQ-IES. For a detailed description, see [Ham02].

Algorithm 12.1 IQ-DH $(\Delta, \mathfrak{g}, c)$

Parameters: Discriminant $\Delta < 0$, reduced ideal $\mathfrak{g}$, exponent bound c
Parties: Entities A and B
End state: Secret ideal $\mathfrak{c}$ known to A and B

```
  A chooses randomly $a \in 1..c$, and calculates the reduced ideal $\mathfrak{a}$ in $[\mathfrak{g}^a]$.
  A sends $\mathfrak{a}$ to B.
  B chooses randomly $b \in 1..c$, and calculates the reduced ideal $\mathfrak{b}$ in $[\mathfrak{g}^b]$.
  B sends $\mathfrak{b}$ to A.
  A calculates the reduced ideal $\mathfrak{c}$ in $[\mathfrak{b}^a]$.
  B calculates the reduced ideal $\mathfrak{c}$ in $[\mathfrak{a}^b]$.
```

Both parties participating in a round of IQ-DH may use a public key-derivation function to turn the common ideal $\mathfrak{c}$ into a key for the symmetric encryption function they intend to use.

Algorithm 12.2 $\texttt{IQ-IES}\ (\Delta, \mathfrak{g}, c, (E, D), \texttt{genKey}, \texttt{MAC})$

Parameters: Discriminant $\Delta < 0$, reduced ideal $\mathfrak{g}$, exponent bound c, secret-key encryption scheme (E, D), key derivation function `genKey`, message authentication code `MAC`

Parties: Sender A and recipient B

End state: Secret message m received by B

```
{Key generation}
B chooses randomly $b \in 1..c$, and calculates the reduced ideal $\mathfrak{b}$ in $[\mathfrak{g}^b]$.
B publishes $\mathfrak{b}$, his public key. He keeps $b$ as his private key.
{Encryption}
A obtains $\mathfrak{b}$ and verifies that it was authentically published by B.
A chooses randomly $a \in 1..c$ and computes the reduced ideals $\mathfrak{a}$ in $[\mathfrak{g}^a]$, and $\mathfrak{c}$ in $[\mathfrak{b}^a]$.
A employs genKey to calculate the secret key pair $(k_{Enc}, k_{MAC})$ starting from $\mathfrak{c}$.
A encrypts message $m$ yielding cipher text $c \leftarrow E(m, k_{Enc})$, and computes the message authentication code $h \leftarrow \texttt{MAC}(c, k_{MAC})$.
A transmits $(\mathfrak{a}, c, h)$ to B.
{Decryption}
B calculates the reduced ideal $\mathfrak{c}$ in $[\mathfrak{a}^b]$.
B employs genKey to calculate the secret key pair $(k_{Enc}, k_{MAC})$ starting from $\mathfrak{c}$.
B computes the message authentication code $h' \leftarrow \texttt{MAC}(c, k_{MAC})$.
If $h \neq h'$, then B rejects the message.
B decrypts $c$ yielding the original message $m \leftarrow D(c, k_{Enc})$.
```

The encryption algorithm IQ-IES uses a a message authentication code. Message authentication codes are algorithms which, given some secret key, assign bit strings to messages. If a sender wants to prevent the alteration of a message during transport, then she applies a message authentication code to the message using some secret key, and transmits the resulting bitstring together with the message itself. The recipient of the message shares knowledge of the secret key. He computes the message authentication code of the received message, and compares it to the one he obtained together with the message. If both coincide he is assured that the message was not altered on the way.

A listing of IQ-IES can be found above.

The security proofs of IQ-DH, and IQ-IES carry directly over from the known proofs for other variants of the Diffie Hellman key exchange, e.g., for the Diffie-Hellman key exchange on elliptic curves, and the analysis of IES in [ABR01].

We turn to digital signature schemes which work in class groups of imaginary quadratic orders. The DL-based signature scheme by Taher ElGamal [ElG85] and the Digital Signature Algorithm DSA [DSS00] cannot be implemented in imaginary quadratic class groups since they require the knowledge of the group order. DSA can be modified, however, so as to avoid reliance on the knowledge of the group order. Safuat Hamdy and Bodo Möller de-

Algorithm 12.3 `IQ-DSA` $(\Delta, \mathfrak{g}, c, d, h)$

```
Parameters: Discriminant $\Delta < 0$, reduced ideal $\mathfrak{g}$, exponents bounds $c$ and $d$, hash
    function $h$
Parties: Signer A and Verifier B
End state: Authenticity of message $m$ verified by B
  {Key generation}
  A chooses randomly $a \in 1..c$, and calculates the reduced ideal $\mathfrak{a}$ in $[\mathfrak{g}^a]$.
  A publishes $\mathfrak{a}$, her public key. She keeps $a$ as her private key.
  {Signature}
  A chooses randomly $k \in 1..d$, and calculates the reduced ideal $\mathfrak{k}$ in $[\mathfrak{g}^k]$.
  A calculates $s \leftarrow -ah(m, \mathfrak{k}) + k$.
  A transmits $m$ together with her signature $(\mathfrak{k}, s)$ to B.
  {Verification}
  B obtains $\mathfrak{a}$ and verifies that it was authentically published as A's public key.
  B calculates the reduced ideal $\mathfrak{k}'$ in $[\mathfrak{g}^s \mathfrak{a}^{h(m,\mathfrak{k})}]$.
  B accepts $m$ as authentically signed by A if $\mathfrak{k}' = \mathfrak{k}$.
```

scribed such a variant in [HM00a], and called it IQ-DSA. The disadvantage of IQ-DSA compared to DSA is that it requires larger exponents and produces larger signatures.

We give a brief description of IQ-DSA. Like most signature schemes, IQ-DSA employs a hash function. Hash functions assign to any message a short bit string called its hash value. It is supposed to be difficult both to find a message which hashes to a given value—this property is called one-wayness—, and to find two messages which have equal hash values. This last property is usually called collision-freeness, although there certainly exists a pair of messages both hashing to the same value provided the message space has larger cardinality than the space of hash values.

A listing of IQ-DSA can be found above. The security proof of generalized DSA by Guillaume Poupard and Jacques Stern [PS98] can be applied to IQ-DSA to show that the security of IQ-DSA can be reduced to the DL problem in the class group: Assume we consider any probability smaller than $1/N$ to be negligible. Assume moreover that there is no algorithm that solves the DL problem in the sub-group of $\mathcal{O}(\Delta)$ generated by $\mathfrak{g}$ with non-negligible probability. Choose c to be N^2, choose d to be N^5, and limit the number of all signatures made with a given key to N. Then the success probability of any forgery algorithm is negligible.

Another DSA variant called RDSA was suggested by Ingrid Biehl *et al.* in [BBHM02]. A version of RDSA that uses exponents that are as small as DSA exponents was broken in [FP03].

Since the root problem in imaginary quadratic class groups is intractable, the Guillou-Quisquater signature scheme [GQ88] can be implemented in those groups. This variant of the Guillou-Quisquater scheme is referred to as IQ-GQ.

Algorithm 12.4 `IQ-GQ` $(\Delta, \mathfrak{g}, c, h)$

```
Parameters: Discriminant $\Delta < 0$, reduced ideal $\mathfrak{g}$, exponent bound $c$
Parties: Signer A and Verifier B
End state: Authenticity of message $m$ verified by B
  {Key generation}
  A chooses randomly $a \in 1..c$, and calculates the reduced ideal $\mathfrak{c}$ in $[\mathfrak{g}^a]$.
  A chooses randomly $n \in 1..c$, and calculates the reduced ideal $\mathfrak{n}$ in $[\mathfrak{g}^{-n}]$.
  A publishes $(\mathfrak{n}, n)$, her public key. She keeps $\mathfrak{a}$ as her private key.
  {Signature for message $m$}
  A chooses randomly $k \in 1..c$, and calculates the reduced ideal $\mathfrak{k}$ in $[\mathfrak{g}^k]$.
  A calculates the reduced ideal $\mathfrak{r}$ in $[\mathfrak{k}^n]$.
  A calculates $s \leftarrow h(m, \mathfrak{r})$, and the reduced ideal $\mathfrak{s}$ in $[\mathfrak{k}\mathfrak{a}^s]$.
  A transmits $m$ together with her signature $(s, \mathfrak{s})$.
  {Verification}
  B obtains $(\mathfrak{n}, n)$ and verifies that it was authentically published as A's public key.
  B calculates the reduced ideal $\mathfrak{r}'$ in $[\mathfrak{s}^n \mathfrak{n}^s]$.
  B accepts $m$ as authentically signed by A if $h(m, \mathfrak{r}') = s$.
```

If the hash function used in IQ-GQ is modeled as a function randomly drawn from a large set of functions with the same domain and image[1], IQ-GQ can be reduced to the IQ root problem (see [HM00a] and [Ham02]).

In [Ham02] Hamdy reports about experimental results concerning IQ-DSA and IQ-GQ. For example, for a 671-bit discriminant which guarantees a security comparable to 1024-Bit RSA generating a signature with IQ-GQ takes 139.06 ms and verifying a signature takes 93.74 ms on a 500MHz SunBlade 100. Hamdy offers a C-library with improved performance [Ham].

12.3 Cryptographic algorithms in real-quadratic orders

In this section we will show how the Infrastructure Diffie-Hellman Problem and the Principal Ideal Problem can be used as the security basis for cryptographic algorithms.

Let $\mathcal{O}$ be a real-quadratic order and $\mathfrak{g}$ a principal reduced $\mathcal{O}$-ideal. The Diffie-Hellman problem in the group $\langle \mathfrak{g} \rangle$ is simple to solve by looking at the norms of the ideals involved. If, however, we replace each $\mathfrak{g}^k$ by a reduced ideal with small distance to it, then the problem transforms into the Infrastructure Diffie-Hellman Problem. This problem is intractable provided the regulator of $\mathcal{O}$ is sufficiently large.

This idea was used by Johannes Buchmann and Hugh Williams for their proposal of an infrastructure variant of the Diffie-Hellman key exchange protocol (RQ-DH). We give a brief description of RQ-DH. For a detailed description, see [BW90].

[1] This proof technique is usually referred to as the Random Oracle Model

Algorithm 12.5 RQ-DH (Δ, d, c, ϵ)

```
Parameters: Discriminant $\Delta > 0$, basis distance $d$, exponent bound $c$, error tolerance $\epsilon$
Parties: Entities A and B
End state: Secret ideal $\mathfrak{e}$ known to A and B
A chooses randomly $a \in 1..c$, and calculates a reduced ideal $\mathfrak{a}$ which has a generator
  $\alpha$ satisfying $\operatorname{Log}\alpha - ad = \delta_A$ with $|\delta_A| < (1/4)\log\Delta$.
A sends $\mathfrak{a}$ and rational $\tilde{\delta}_A$ with $|\tilde{\delta}_A - \delta_A| < \epsilon/c$ to B.
B chooses randomly $b \in 1..c$, and calculates a reduced ideal $\mathfrak{b}$ which has a generator
  $\beta$ satisfying $\operatorname{Log}\beta - bd = \delta_B$ with $|\delta_B| < (1/4)\log\Delta$.
B calculates the reduced ideals $\mathfrak{d}_j$ for which relative generators $\beta_j$ exist such that
  $\mathfrak{d}_j = \beta_j\mathfrak{a}^b$ with $|\operatorname{Log}\beta_j + b\tilde{\delta}_A| \leq (1/4)\log\Delta$.
if B finds $j$ such that $|\operatorname{Log}\beta_j + b\tilde{\delta}_A| < \epsilon$ then
    B sets $\mathfrak{e} \leftarrow \mathfrak{d}_i$ and $s_B \leftarrow 0$.
else
    B sets $s_B \leftarrow 1$
B sends $\mathfrak{b}$, rational $\tilde{\delta}_B$ with $|\tilde{\delta}_A - \delta_A| < \epsilon/c$, and $s_B$ to A.
A calculates the reduced ideals $\mathfrak{c}_i$ for which relative generators $\alpha_i$ exist such that
  $\mathfrak{c}_i = \alpha_i\mathfrak{b}^a$ with $|\operatorname{Log}\alpha_i + a\delta_B| < (1/4)\log\Delta$.
if $s_B = 0$ then
    A finds $i$ such that $|\operatorname{Log}\alpha_i + a\tilde{\delta}_B| < 2\epsilon$, and sets $\mathfrak{e} = \mathfrak{c}_i$.
else
    if A finds $i$ such that $|\operatorname{Log}\alpha_i + b\tilde{\delta}_B| < \epsilon$, then
        A sets $\mathfrak{e} \leftarrow \mathfrak{c}_i$ and $s_A \leftarrow 0$;
        A transmits $s_A$ to $B$;
        B finds $j$ such that $|\operatorname{Log}\beta_j + b\tilde{\delta}_A| < 2\epsilon$ and sets $\mathfrak{e} \leftarrow \mathfrak{d}_j$.
    else
        A finds $i$ such that $\operatorname{Log}\alpha_{i-1} + a\tilde{\delta}_B + \epsilon < 0 < \operatorname{Log}\alpha_i + a\tilde{\delta}_B - \epsilon$,
        A sets $\mathfrak{e} \leftarrow \mathfrak{c}_i$ and $s_A \leftarrow 1$;
        A transmits $s_A$ to $B$;
        B finds $j$ such that $\operatorname{Log}\beta_{j-1} + b\tilde{\delta}_A + \epsilon < 0 < \operatorname{Log}\beta_j + b\tilde{\delta}_A - \epsilon$;
        B sets $\mathfrak{e} \leftarrow \mathfrak{d}_j$.
```

As the starting point for the RQ-DH protocol we could either choose a commonly known ideal $\mathfrak{g}$, or the length of one of its generators. For simplicity one chooses to start from some arbitrary basis distance d instead, say $d = 1$.

The participants of the protocol choose secret exponents a and b, respectively. Then they compute reduced ideals $\mathfrak{a} = (\alpha)$ and $\mathfrak{b} = (\beta)$ which have generators α and β of length approximately ad and bd. These ideals are transmitted to the other side. Unfortunately, the ideals $\mathfrak{a}^b$ and $\mathfrak{b}^a$ need not have a short relative generator, since

$$b\operatorname{Log}\alpha - a\operatorname{Log}\beta = b(\operatorname{Log}\alpha - ad) - a(\operatorname{Log}\beta - bd)$$

and each of the factors $\operatorname{Log}\alpha - ad$ and $\operatorname{Log}\beta - bd$ may be as large as $(1/4)\log\Delta$. It is therefore necessary to transmit these quantities alongside with $\mathfrak{a}$ and $\mathfrak{b}$.

Using these distances and working in analogy to the computation of the cycle section in section 11.5.4, both participants end up with overlapping sets of reduced ideals, each having generators of length approximately abd. It remains to settle on one of these ideals as the common key.

One approach for this task is for one side to choose randomly one of the ideals obtained, apply a one-way key-generation function to obtain a key for some message authentication code and apply it to the ideal it transmits. The other side can test this code with all candidate ideals as input for the key-generation function in succession, and determine the correct one for which the message authentication code verifies.

It is, however, also possible to limit the transmitted data required for the determination of the common ideal from among the sets obtained to a single bit sent by each side. The common ideal has a generator with length either very close to abd, or as small as possible while being larger than abd. Algorithm 12.3 shows the details of this determination.

A DH key exchange in an order with 768-bit discriminant can be performed in 2 seconds CPU time on an 800 MHz Pentium III machine [JvdP02].

Real-quadratic ElGamal encryption can also be implemented based on real quadratic Diffie-Hellman key exchange. Moreover, there is a variant of the Fiat-Shamir signature protocol [FS87], called PIP-FS, which relies on the intractability of the Principal Ideal Problem, and was proposed and analyzed in [BMM00].

Choosing the discriminant of the underlying order for PIP-FS at the same order of magnitude as is secure for crypto-systems in imaginary quadratic class groups at a security level corresponding to 1024-bit RSA, PIP-FS can be executed in about 3 seconds CPU time on a 300 MHz Pentium II. Key generation takes less than a minute.

12.4 Open Problems

This overview shows that much has been done in quadratic field cryptography. However, there are still many open problems.

Imaginary quadratic field cryptography is ready for practical use. Offered are a Java implementation [NFP], and a C-library for IQ-cryptography [Ham]. It would be nice to see IQ-cryptography being used in real applications.

Real quadratic field cryptography is still too inefficient to be used in practice. Security of RQ schemes hinges on the right choice of the underlying order. To increase security, the authors of [JSW01] suggest the use of orders with discriminants of a particular form, since in these orders the application of the currently fastest algorithm for the computation of the regulator is considerably slower than in the average case. The record computations also given in the paper give valuable data points for estimates of lower bounds on the size of cryptographically secure discriminants. In general, however, a comprehensive analysis and, on its basis, a recommendation for the right choice of parameters in RQ cryptography is still outstanding.

There are generalizations of quadratic field cryptography to arbitrary number fields. First ideas are described in [BP97], [MNP01] and [BBHM02]. A very interesting aspect of general number field cryptography is the fact that no attack is known that is sub-exponential in the field degree. This is due to the fact that breaking crypto-systems in number fields of degree n requires the computation of short vectors in n-dimensional lattices.

One possible set-up of number field cryptography is to use number fields with small regulators and large discriminants. They have large class groups and deciding equality in such class groups is easy. Therefore, the IQ-crypto-algorithms can be implemented in such fields. However, when the field degree is large, arithmetic in the class group is very inefficient. There is a big need for optimization.

Also, the right choice of parameters is an open problem. In particular, it is an interesting question how to generate infinite families of number fields of very large degree with small regulators. It is in principle also possible to use number fields with small class numbers and large regulators. Very little is known about such applications.

Finally, it is an open question to what extent general number field cryptography is resistant to quantum computer attacks. Discrete logarithms in the imaginary quadratic class group can be computed in polynomial time using standard techniques since there is a unique representation for each group element, see [Sho97]. Sean Hallgren has sketched in [Hal02] a polynomial time quantum algorithm for the computation of the regulator of a real quadratic order and the solution of the principal ideal problem (PIP) in it. His algorithm was extended to higher degree fields by himself [Hal05] and Arthur Schmidt and Ulrich Vollmer [SV05]. The run-time of their algorithms depend very badly on the degree of the field. Ways to improve this dependence remain to be studied.

12.5 Exercises

Exercise 12.5.1. Give an analogue to the Class Group Root Problem which operates in the infrastructure of a real-quadratic order.

Exercise 12.5.2. Prove that both participants obtain the same ideal after one round of the protocol given in Algorithm 12.3.

Chapter references and further reading

[ABR01] Michel Abdalla, Mihir Bellare, and Philip Rogaway, *An encryption scheme based on the Diffie-Hellman problem*, Progress in Cryptology —

CT-RSA 2001 (David Naccache, ed.), Lecture Notes in Computer Science, vol. 2020, Springer-Verlag, 2001, pp. 143–158.

[BBHM02] Ingrid Biehl, Johannes Buchmann, Safuat Hamdy, and Andreas Meyer, *A signature scheme based on the intractability of computing roots*, Designs, Codes and Cryptography **25** (2002), no. 3, 223–236.

[BBT94] Ingrid Biehl, Johannes Buchmann, and Christoph Thiel, *Cryptographic protocols based on discrete logarithms in real-quadratic orders*, Advances in Cryptology – CRYPTO '94 (Yvo G. Desmedt, ed.), Lecture Notes in Computer Science, vol. 839, Springer-Verlag, 1994, pp. 56–60 (English).

[BDW90] Johannes Buchmann, Stephan Düllmann, and Hugh C. Williams, *On the complexity and efficiency of a new key exchange system*, Advances in Cryptology – EUROCRYPT '89 (Jean-Jacques Quisquater and Joos Vandewalle, eds.), Lecture Notes in Computer Science, vol. 434, Springer-Verlag, 1990, pp. 597–616.

[BMM00] Johannes Buchmann, Markus Maurer, and Bodo Möller, *Cryptography based on number fields with large regulator*, Journal de Théorie des Nombres de Bordeaux **12** (2000), 293–307.

[BMT96] Ingrid Biehl, Bernd Meyer, and Christoph Thiel, *Cryptographic protocols based on real-quadratic A-fields (extended abstract)*, Advances in Cryptology – ASIACRYPT '96 (Kwangjo Kim and Tsutomu Matsumoto, eds.), Lecture Notes in Computer Science, vol. 1163, Springer-Verlag, 1996, pp. 15–25.

[BP97] Johannes Buchmann and Sachar Paulus, *A one way function based on ideal arithmetic in number fields*, Advances in Cryptology – CRYPTO '97 (Burton S. Kaliski, ed.), Lecture Notes in Computer Science, vol. 1294, Springer-Verlag, 1997, pp. 385–394.

[BPT04] Ingird Biehl, Sachar Paulus, and Tsuyoshi Takagi, *Efficient undeniable signature schemes based on ideal arithmetic in quadratic orders*, Designs, Codes and Cryptography **31** (2004), no. 2, 99–123.

[BR97] Mihir Bellare and Philip Rogaway, *Minimizing the use of random oracles in authenticated encryption schemes*, Information and Communications Security, ICIS '97 (Y. Han, T. Okamoto, and S. Quing, eds.), Lecture Notes in Computer Science, vol. 1334, Springer-Verlag, 1997, pp. 1–16.

[Bra90] Gilles Brassard (ed.), *Advances in cryptology – CRYPTO '89*, Lecture Notes in Computer Science, vol. 435, Springer-Verlag, 1990.

[BST02] Johannes Buchmann, Kouichi Sakurai, and Tsuyoshi Takagi, *An IND-CCA2 public-key cryptosystem with fast decryption*, Information Security and Cryptology - ICISC 2001 (Kwangjo Kim, ed.), Lecture Notes in Computer Science, vol. 2288, Springer-Verlag, 2002, pp. 51–71.

[BSW90] Johannes Buchmann, Renate Scheidler, and Hugh C. Williams, *Implementation of a key exchange protocol using real quadratic fields*, Proc. of EUROCRYPT '90, Lecture Notes in Computer Science, vol. 473, Springer-Verlag, 1990, pp. 98–109.

[BSW94] ———, *A key-exchange protocol using real quadratic fields*, Journal of Cryptology **7** (1994), 171–199.

[BW88] Johannes Buchmann and Hugh C. Williams, *A key-exchange system based on imaginary quadratic fields*, Journal of Cryptology **1** (1988), no. 2, 107–118 (English).

[BW90] ———, *A key-exchange system based on real quadratic fields*, in Brassard [Bra90], pp. 335–343.

[DH76] Whitfield Diffie and Martin E. Hellman, *New directions in cryptography*, IEEE Transactions on Information Theory **IT-22** (1976), no. 6, 644–654.

[DSS00] *Digital signature standard*, Federal Information Processing Standards Publication FIPS 186-2, NIST, 2000.

[ElG85] Taher ElGamal, *A public key cryptosystem and a signature scheme based on discrete logarithms*, IEEE Transactions on Information Theory **31** (1985), no. 4, 469–472.

[FP03] Pierre-Alain Fouque and Guillaume Poupard, *On the security of RDSA*, Advances in Cryptology – EURCRYPT 2003 (Eli Biham, ed.), Lecture Notes in Computer Science, vol. 2656, Springer-Verlag, 2003.

[FS87] Amos Fiat and Adi Shamir, *How to prove yourself: Practical solutions to identification and signature problems*, Advances in Cryptology – CRYPTO '86 (Andrew M. Odlyzko, ed.), Lecture Notes in Computer Science, vol. 263, Springer-Verlag, 1987, pp. 186–194.

[GJS01] Mark Giesbrecht, Michael J. Jacobson, Jr., and Arne Storjohann, *Algorithms for large integer matrix problems*, Applied algebra, algebraic algorithms and error-correcting codes (Melbourne, 2001), Lecture Notes in Computer Science, vol. 2227, Springer, Berlin, 2001, pp. 297–307.

[GQ88] Louis C. Guillou and Jean-Jacques Quisqater, *A practical zero-knowledge protocol fitted to security microprocessors minimizing both transmission and memory*, Advances in Cryptology – EUROCRYPT '88 (Christoph G. Günther, ed.), Lecture Notes in Computer Science, vol. 330, Springer-Verlag, 1988, pp. 123–128.

[Hal02] Sean Hallgren, *Polynomial-time quantum algorithms for Pell's equation and the principal ideal problem*, Proceedings of the thiry-fourth annual ACM symposium on the theory of computing, ACM Press, 2002, pp. 653–658.

[Hal05] ———, *Fast quantum algorithms for computing the unit group and class group of a number field*, Proceedings of the 37th annual ACM Symposium on Theory of Computing (Harold N. Gabow and Ronald Fagin, eds.), ACM Press, 2005, pp. 468–474.

[Ham] Safuat Hamdy, `libiq` — *a library for arithmetic in class groups of imaginary quadratic orders*, `http://www.math.ucalgary.ca/~hamdy/libiq.html`.

[Ham02] ———, *Über die Sicherheit und Effizienz kryptografischer Verfahren mit Klassengruppen imaginär-quadratischer Zahlkörper*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, Darmstadt, Germany, 2002, `http://www.informatik.tu-darmstadt.de/ftp/pub/TI/reports/hamdy.diss.pdf`.

[HJPT98] Detlef Hühnlein, Michael J. Jacobson, Jr., Sachar Paulus, and Tsuyoshi Takagi, *A cryptosystem based on non-maximal imaginary quadratic orders with fast decryption*, Advances in Cryptology – EUROCRYPT '98 (Kaisa Nyberg, ed.), Lecture Notes in Computer Science, vol. 1403, Springer-Verlag, 1998, pp. 294–307 (English).

[HJW03] Detlef Hühnlein, Michael J. Jacobson, Jr., and Damian Weber, *Towards practical non-interactive public key cryptosystems using non-maximal imaginary quadratic orders*, Designs, Codes and Cryptography **30** (2003), no. 3, 281–299.

[HM00a] Safuat Hamdy and Bodo Möller, *Security of cryptosystems based on class groups of imaginary quadratic orders*, Advances in Cryptology – ASIACRYPT 2000 (Tatsuaki Okamoto, ed.), Lecture Notes in Computer Science, vol. 1976, Springer-Verlag, 2000, pp. 234–247.

[HM00b] Detlef Hühnlein and Johannes Merkle, *An efficient NICE-Schnorr-type cryptosystem*, Practice and Theory in Public Key Cryptography, PKC 2000 (Hideki Imai and Yuliang Zheng, eds.), Lecture Notes in Computer Science, vol. 1751, Springer-Verlag, 2000, pp. 14–27.

[HMT98] Detlef Hühnlein, Andreas Meyer, and Tsuyoshi Takagi, *Rabin and RSA analogues based on non-maximal imaginary quadratic orders*, in Rhee and Imai [RI98], pp. 221–240.

[HP01] Detlef Hühnlein and Sachar Paulus, *On the implementation of cryptosystems based on real quadratic fields*, Selected Areas in Cryptography, SAC 2001 (Serge Vaudenay and Amr M. Youssef, eds.), Lecture Notes in Computer Science, vol. 2259, 2001, pp. 288–302.

[Hüh00] Detlef Hühnlein, *Kryptosysteme auf Basis Quadratischer Ordnungen*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, 2000.

[Hüh01a] Detlef Hühnlein, *Efficient implementation of cryptosystems based on non-maximal imaginary quadratic orders*, Selected Areas in Cryptography, SAC 2000 (Douglas R. Stinson and Stafford Tavares, eds.), Lecture Notes in Computer Science, vol. 2012, 2001, pp. 150–167.

[Hüh01b] Detlef Hühnlein, *Faster generation of NICE-schnorr-type signatures*, Topics in Cryptology - CT-RSA 2001 (Berlin) (D. Naccache, ed.), Lecture Notes in Computer Science, vol. 2020, Springer-Verlag, 2001, pp. 1–12.

[Jac99] Michael J. Jacobson, Jr., *Applying sieving to the computation of quadratic class groups*, Mathematics of Computation **68** (1999), no. 226, 859–867.

[Jac00] ———, *Computing discrete logarithms in quadratic orders*, Journal of Cryptology **13** (2000), no. 4, 473–492.

[Jac04] ———, *The security of cryptosystems based on class semigroups of imaginary quadratic non-maximal orders*, Australasian Conference on Information Security and Privacy, ACISP 2004 (Vijay Varadharajan Huaxiong Wang, Josef Pieprzyk, ed.), Lecture Notes in Computer Science, vol. 3108, Springer-Verlag, 2004, pp. 149–156.

[JSW01] Michael J. Jacobson, Jr., Renate Scheidler, and Hugh C. Williams, *The efficiency and security of a real quadratic field based-key exchange protocol*, Public-Key Cryptography and Computational Number Theory (Warsaw, Poland), de Gruyter, 2001, pp. 89–112.

[JSW06] ———, *An improved real quadratic field based key exchange protocol*, Journal of Cryptology **19** (2006), no. 2, 211–239.

[JvdP02] Michael J. Jacobson, Jr. and Alfred J. van der Poorten, *Computational aspects of NUCOMP*, Algorithmic Number Theory, ANTS-V (Claus Fieker and David R. Kohel, eds.), Lecture Notes in Computer Science, vol. 2369, Springer-Verlag, 2002, pp. 120–133.

[MNP01] Andreas Meyer, Stefan Neis, and Thomas Pfahler, *First implementation of cryptographic protocols based on algebraic number fields*, Information Security and Privacy, ACISP 2001, Sydney (Vijay Varadharajan and Yi Mu, eds.), Lecture Notes in Computer Science, vol. 2119, Springer, 2001, pp. 84–103.

[NFP] *NFProvider — a toolkit for the Java Cryptography Architecture (JCA/JCE) for Number Field Cryptography*, `http://www.informatik.tu-darmstadt.de/TI/Forschung/FlexiProvider/overview.html#NFProvider`, Part of the FlexiProvider toolkit.

[PS98] Guillaume Poupard and Jacques Stern, *Security analysis of a practical "on the fly" authentication and signature generation*, Advances in Cryptology – EUROCRYPT '98 (Kaisa Nyberg, ed.), Lecture Notes in Computer Science, vol. 1403, Springer-Verlag, 1998, pp. 422–436.

[PT98] Sachar Paulus and Tsuyoshi Takagi, *A generalization of the Diffie-Hellman problem and related cryptosystems allowing fast decryption*, in Rhee and Imai [RI98], pp. 211–220.

[PT00] ———, *A new public-key cryptosystem over a quadratic order with quadratic decryption time*, Journal of Cryptology **13** (2000), no. 2, 263–272. MR 1 748 525.

[RI98] Man Young Rhee and Hideki Imai (eds.), *The 1st international conference on information security and cryptology, ICISC '98*, DongKwang Publishing Company, Korea, 1998.

[Sho97] Peter W. Shor, *Polynomial time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM Journal on Computing **26** (1997), no. 5, 1484–1509.

[SV05] Arthur Schmidt and Ulrich Vollmer, *Polynomial time quantum algorithm for the computation of the unit group of a number field*, Proceedings of the 37th annual ACM Symposium on Theory of Computing (Harold N. Gabow and Ronald Fagin, eds.), ACM Press, 2005, pp. 475–480.

[Wie86] Douglas H. Wiedemann, *Solving sparse linear equations over finite fields*, IEEE Trans. Inf. Theory IT-32 (1986), 54–62.

A

Appendix

A.1 Vectors and matrices

We introduce a few general notions concerning vectors and matrices. Let S be a set. By S^n we denote the set of all n-tuples of elements in M. An element $\mathbf{v}$ of S^n is written as $\mathbf{v} = (v_1, \dots, v_n)$. Also, by $S^{(m,n)}$ we denote the set of all matrices with entries from M which have m rows and n columns. If $A \in S^{(m,n)}$ then we write $A = (a_{i,j})_{1 \le i \le m, 1 \le j \le n}$ or simply $A = (a_{i,j})$ where $a_{i,j}$ is the entry in the ith row and the jth column of A. If A and B are matrices with the same number of rows and with n_A and n_B columns, respectively, then $A \circ B$ is the matrix whose first n_A columns are the columns of A and whose second n_B columns are the columns of B. The matrix I_n is the unit matrix with dimension n. Its diagonal entries are 1, all others 0.

Let R be a commutative ring with unit element. The product of matrices over R is defined in the usual way. Let $A \in R^{(m,n)}$, $A = (a_{i,j})$ and let $\mathbf{v} \in R^n$, $\mathbf{v} = (v_1, \dots, v_n)$. Then the product $A\mathbf{v}$ of the matrix A with the vector $\mathbf{v}$ is defined as

$$A \cdot \mathbf{v} = A\mathbf{v} = \left(\sum_{j=1}^{n} a_{1,j} v_j, \dots, \sum_{j=1}^{n} a_{m,j} v_j \right) .$$

The result is a vector in R^m. Also, if $\mathbf{v} \in R^m$ then the product $\mathbf{v}A$ of the vector $\mathbf{v}$ with the matrix A is defined as

$$\mathbf{v} \cdot A = \mathbf{v}A = \left(\sum_{i=1}^{n} a_{i,1} v_i, \dots, \sum_{i=1}^{n} a_{i,n} v_i \right) .$$

The result is a vector in R^n.

If $A \in R^{(n,n)}$, then the determinant of A is denoted by $\det A$, the trace of A is denoted by $\operatorname{Tr} A$ and the characteristic polynomial is denoted by $c_A(X)$. If $\mathbf{v}_1, \dots, \mathbf{v}_n \in R^n$ then $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ is identified with the matrix with columns $\mathbf{v}_1, \dots, \mathbf{v}_n$.

If $\mathbf{v} \in \mathbb{Z}^l$, $\mathbf{v} = (v_1, \dots, v_l)$, then we define the norms

$$|\mathbf{v}| = |\mathbf{v}|_2 = \Big(\sum_{i=1}^{l} v_i^2\Big)^{1/2} , \quad \text{and} \quad |\mathbf{v}|_\infty = \max\{\, |v_i| \mid i = 1, \dots, l\,\} . \tag{A.1}$$

The norm $|\ |_2$ is called the *Euclidean length* of $\mathbf{v}$. It is induced by the scalar product

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{l} v_i w_i$$

for vectors $\mathbf{v} = (v_1, \dots, v_l)$ and $\mathbf{w} = (w_1, \dots, w_l)$.

A.2 Action of groups on sets

Let S be a non-empty set and let G be a group. A *left action* of G on S is a map

$$G \times S \to S , \quad (g, s) \mapsto gs = g \cdot s$$

which has the following properties.

1. If 1_G is the neutral element in G then $1_G s = s$ for any $s \in S$.
2. For $g, h \in G$ and $s \in S$ we have $g(hs) = (gh)s$.

Suppose we have a left action of G on S as described above. Call two elements s and t in S equivalent if there is a group element g such that $t = gs$. It is easy to see that this is an equivalence relation on S. If $s \in S$, then the equivalence class $\{gs : g \in G\}$ is called the G-*orbit* of s. It is denoted by Gs. The set S is a disjoint union of its G-orbits. The set of all G-orbits of S is denoted by $G\backslash S$.

A *right action* of G on S is a map

$$S \times G \to S , \quad (s, g) \mapsto sg = s \cdot g$$

that satisfies the following properties.

1. If 1_G is the neutral element in G then $s1_G = s$ for any $s \in S$.
2. For g, h and $s \in S$ we have $(sg)h = s(gh)$.

Suppose we have a right action of G on S as described above. The G-orbits of S are defined in analogy to the G-orbits for a left action of G. The G-orbit of $s \in S$ is denoted by sG. The set of all G-orbits of S is denoted by G/S.

A.3 The lemma of Gauss

Definition A.3.1. *The* content *of a polynomial f with integer coefficients is the greatest common divisor of its coefficients. It is denoted by* $\mathrm{cont}(f)$. *The polynomial f is called* primitive *if* $\mathrm{cont}(f) = 1$.

The following statement is known as the *Lemma of Gauss.*

Lemma A.3.2. *Let $f, g \in \mathbb{Z}[X]$, $f, g \neq 0$. Then* $\mathrm{cont}(fg) = \mathrm{cont}(f)\,\mathrm{cont}(g)$.

Proof. Write $f = \mathrm{cont}(f)f'$, $g = \mathrm{cont}(g)g'$ with $f', g' \in \mathbb{Z}[X]$. Then f' and g' are primitive and $fg = \mathrm{cont}(f)\,\mathrm{cont}(g)f'g'$. It suffices to prove that the polynomial $f'g'$ is primitive. If $f' = \sum_{i=0}^{m} a_i X^i$, $g' = \sum_{j=0}^{n} b_j X^j$ then $f'g' = \sum_{k=0}^{m+n} c_k X^k$ where $c_k = \sum_{i+j=k} a_i b_j$.

If $f'g'$ is not primitive, then there exists a prime number q such that $q \mid c_k$ for $0 \leq k \leq m + n$. Let s be the least non negative integer with $q \nmid a_s$. Then $s \leq m$ because f' is primitive. Likewise, let t be the least non-negative integer such that $q \nmid b_t$. Then $t \leq n$ because g' is primitive. Now we have $c_{s+t} = a_0 b_{s+t} + \ldots + a_{s-1} b_{t+1} + a_s b_t + a_{s+1} b_{t-1} \ldots + a_{s+t} b_0$, where $a_i = 0$ for $i > m$ and $b_j = 0$ for $j > n$. Because of the definition of s and t, every term $a_i b_j$ in this sum is divisible by q except for $a_s b_t$. But also c_{s+t} is divisible by q. This is a contradiction. □

A.4 Lattices

Definition A.4.1. *A* lattice *in* $\mathbb{R}^n$ *is a subset* $L = \mathbb{Z}\mathbf{v}_1 + \mathbb{Z}\mathbf{v}_2 + \ldots + \mathbb{Z}\mathbf{v}_k$ *where* $k \in \mathbb{Z}_{\geq 0}$ *and* $\mathbf{v}_1, \ldots, \mathbf{v}_k$ *are elements of* $\mathbb{R}^n$ *which are linearly independent over* $\mathbb{R}$. *The sequence* $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ *is called a* basis *of* L. *The elements of a lattice are called* lattice points.

Let L be a lattice in $\mathbb{R}^n$. Since the elements of a basis of L are linearly independent, each element of $\mathbb{R}^n$ has a unique representation as a linear combination of the basis elements.

Proposition A.4.2. *All bases of L have the same length.*

Proof. If L has a basis of length k then k is the dimension of the subspace generated by the elements of L. Hence k is an invariant of L. □

The length k of the bases of L is called the *dimension* of L. We characterize the set of all bases of L.

Theorem A.4.3. *Let $V = (\mathbf{v}_1, \ldots, \mathbf{v}_k)$ be a basis of L. Then the set of all bases of L is* $\{VU : U \in \mathrm{GL}(k, \mathbb{Z})\}$.

Proof. Let $W = (\mathbf{w}_1, \ldots, \mathbf{w}_k)$ be a basis of L. Then there is a matrix $U \in \mathbb{Z}^{k,k}$ such that $W = VU$. Likewise there is a matrix $U' \in \mathbb{Z}^{k,k}$ such that $V = WU'$. This implies $V = WU' = VUU'$. Because of the uniqueness of the representation of the lattice elements as linear combinations of the basis elements in V we must have $UU' = I_k$. This shows that $U \in \mathrm{GL}(k, \mathbb{Z})$.

Conversely, let $U \in \mathrm{GL}(k, \mathbb{Z})$. Then the vectors in VU are linearly independent. We must show that VU generates L over $\mathbb{Z}$. So let $\mathbf{v} \in V$. Then $\mathbf{v} \in L$ if and only if there is $\mathbf{x} \in \mathbb{Z}^k$ such that $\mathbf{v} = V\mathbf{x}$. This, in turn is true if and only if $\mathbf{v} = VU(U^{-1}\mathbf{x})$ and this concludes the proof. □

Example A.4.4. Let $f = (a, b, c)$ be a positive definite form of discriminant Δ. In the following way a two dimensional lattice in the plane can be constructed such that the values of af are the lengths of the lattice vectors. We choose the basis $\mathbf{v} = (a, 0)$, $\mathbf{w} = (b/2, \sqrt{|\Delta|}/2)$. Then $|x\mathbf{v} + y\mathbf{w}|^2 = (ax + (b/2)y)^2 - y^2\Delta/4 = af(x, y)$.

We prove another characterization of lattices. For this purpose we need the following geometric notion. If r is a non-negative real number and $\mathbf{c} \in \mathbb{R}^n$, then the *sphere* with center $\mathbf{c}$ of radius r in $\mathbb{R}^n$ is the set

$$\left\{\mathbf{v} \in \mathbb{R}^n : |\mathbf{v} - \mathbf{c}| \leq r\right\} .$$

A subset S of $\mathbb{R}^n$ is called *discrete* if for every positive real number r the sphere with center 0 and radius r contains only finitely many elements of S.

Theorem A.4.5. *A subset L of $\mathbb{R}^n$ is a lattice in V if and only if L is a discrete set of points and a subgroup of $\mathbb{R}^n$.*

For the proof, we require the notion of the Gram-Schmidt orthogonalization of a lattice basis which is also important in its own right.

Definition A.4.6. *Let L be a lattice in $\mathbb{R}^n$ with basis $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ be a basis of L. The vectors $(\mathbf{v}_1^*, \ldots, \mathbf{v}_k^*)$ defined by*

$$\mathbf{v}_1^* = \mathbf{v}_1 \ , \quad \mathbf{v}_i^* = \mathbf{v}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{v}_j^*, \mathbf{v}_i \rangle}{\langle \mathbf{v}_j^*, \mathbf{v}_j^* \rangle} \mathbf{v}_j^* \quad \textit{for } 1 < i \leq k \ .$$

are called the Gram-Schmidt orthogonalization *of* $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$.

The vectors $\mathbf{v}_i^*$, $1 \leq i \leq k$ are mutually orthogonal, i.e. for any $1 \leq i < j \leq k$ we have $\langle \mathbf{v}_i^*, \mathbf{v}_j^* \rangle = 0$. Given a basis $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$, we can define the projection π_i on $\mathbb{R}^n$ orthogonal to the space W_i spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_i$ in terms of the Gram-Schmidt orthogonalized basis

$$\pi_i : \mathbb{R}^n \longrightarrow W_i : \mathbf{v} \longmapsto \mathbf{v} - \sum_{j=1}^{i-1} \frac{\langle \mathbf{v}_j^*, \mathbf{v} \rangle}{\langle \mathbf{v}_j^*, \mathbf{v}_j^* \rangle} \mathbf{v}_j^* \ .$$

Proof (of Theorem A.4.5). Let L be a lattice in $\mathbb{R}^n$. By definition, L is a subgroup of $\mathbb{R}^n$. We show that L is discrete. Let $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ be a basis of L and let $(\mathbf{v}_1^*, \ldots, \mathbf{v}_k^*)$ be its Gram-Schmidt-orthogonalization.

Let $r \in \mathbb{R}_{>0}$ and suppose that $\mathbf{v} \in L$ with $|\mathbf{v}| \leq r$. Write

$$\mathbf{v} = x_1\mathbf{v}_1 + \ldots + x_k\mathbf{v}_k, \quad x_i \in \mathbb{Z}, 1 \leq i \leq k \ .$$

From the Cauchy-Schwarz inequality and from the fact that $\mathbf{v}_i^*$ is orthogonal to $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}$ it follows that

$$\sum_{j=i}^{k} x_j \langle \mathbf{v}_j, \mathbf{v}_i^* \rangle = \langle \mathbf{v}, \mathbf{v}_i^* \rangle \leq r|\mathbf{v}_i^*|, \quad 1 \leq i \leq k \ .$$

Therefore, there are only finitely many choices for x_k and by induction we see that there are only finitely many choices for all other x_i. Hence, the number of lattice points in the sphere centered at 0 with radius r is finite.

Conversely, assume that L is discrete and a subgroup $\mathbb{R}^n$. We show that L is a lattice by constructing a basis of L. This is done by induction. Let W be the subspace of $\mathbb{R}^n$ generated by the elements of L over $\mathbb{R}$. Let k be the dimension of W.

For $1 \leq i \leq k$ we construct vectors $\mathbf{v}_i \in L$ such that $\mathbf{v}_1, \ldots, \mathbf{v}_i$ generates an i-dimensional subspace W_i of W and such that $L \cap W_i$ is a lattice in W with basis $(\mathbf{v}_1, \ldots, \mathbf{v}_i)$. Then $W_k = W$ and $L = W \cap L$ has the basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$.

For $\mathbf{v}_1$ we choose a shortest non-zero vector in L. Such a vector exists because L is discrete. It follows from the minimality of the length of $\mathbf{v}_1$ that any other vector $\mathbf{v}$ in L with $\mathbf{v} = r\mathbf{v}_1$, $r \in \mathbb{R}$, must be an integer multiple of $\mathbf{v}_1$. Hence $W_1 \cap L$ is a one-dimensional lattice with basis $\mathbf{v}_1$.

Assume that $i \leq k$ and that $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}$ have already been constructed. Then there are vectors in L whose projection orthogonal to W_{i-1} is non-zero. Among all those vectors choose one whose projection orthogonal to W_{i-1} has minimal length and call it $\mathbf{v}_i$. Let $(\mathbf{v}_1^*, \ldots, \mathbf{v}_i^*)$ be the Gram-Schmidt-Orthogonalization of $(\mathbf{v}_1, \ldots, \mathbf{v}_i)$. Let $\mathbf{v}$ be a vector in $W_i \cap L$. Then $\mathbf{v}$ can be written as $\mathbf{v} = x_1\mathbf{v}_1 + \ldots + x_i\mathbf{v}_i$ with real coefficients x_j, $1 \leq j \leq i$. To show that $W_i \cap L$ is a lattice with basis $(\mathbf{v}_1, \ldots, \mathbf{v}_i)$ it suffices to prove that the coefficients x_i are integers. Since $\mathbf{v} - \lfloor x_i \rfloor \mathbf{v}_i$ also belongs to L we may assume that $0 \leq x_i < 1$. Thus we have to show that $x_1, \ldots, x_{i-1}$ are integers and that $x_i = 0$. The projection of $\mathbf{v}$ orthogonal to W_{i-1} is of length $x_i|\mathbf{v}_i^*|$ while the projection of $\mathbf{v}_i$ orthogonal to W_{i-1} is of length $|\mathbf{v}_i^*|$. The choice of $\mathbf{v}_i$ implies that $x_i = 0$. Therefore $\mathbf{v} \in W_{i-1} \cap L$ which by the induction hypothesis is a lattice with basis $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}$. This means that $x_1, \ldots, x_{i-1}$ are integers. □

As an application, we obtain a very simple characterization of sublattices of L.

Definition A.4.7. *A* sublattice *of L is a lattice which is contained in L.*

Corollary A.4.8. *A subset of L is a sublattice of L if and only if it is a subgroup of L.*

Proof. Let M be a subgroup of L. Then M is a subgroup of $\mathbb{R}^n$. Also, by Theorem A.4.5 L is a discrete subset of $\mathbb{R}^n$. This implies that M is a discrete subset of $\mathbb{R}^n$. Therefore, M is a lattice by Theorem A.4.5. □

A.5 Linear algebra over $\mathbb{Z}$

We discuss a few fundamental algorithms for linear algebra problems over the integers. We begin by explaining several simplified algorithms for square matrices. The properties of state-of-the-art algorithms for arbitrary rectangular matrices are listed in Section A.5.5.

For Sections A.5.1 through A.5.4, we let n be a positive integer and B be matrix in $\mathbb{Z}^{(n,n)}$.

A.5.1 Computing determinants

We explain an algorithm for computing the determinant of B. It proceeds as follows: First a bound for the absolute value of the determinant is established. Then the determinant is computed modulo many small primes whose products exceeds the bound. And finally, the Chines Remainder Theorem is used to piece the local data together.

Let $\mathbf{b}_1, \dots, \mathbf{b}_n$ be the column vectors of B. The value

$$H(B) = \prod_{i=1}^{n} |\mathbf{b}_i| \tag{A.2}$$

is called the *Hadamard bound* of B.

The following proposition was first proved by Jacques Hadamard in 1893. A proof can be found e.g. in [HJ85].

Proposition A.5.1. *We have* $|\det B| \leq H(B)$.

Thus we determine first

$$c = \left\lfloor \min\left\{ H(B), H(B^T) \right\} \right\rfloor \ . \tag{A.3}$$

By Proposition A.5.1 we have

$$-c \leq \det B \leq c \ . \tag{A.4}$$

We then compute a set P of prime numbers such that

$$\prod_{p \in P} p \geq c \ . \tag{A.5}$$

This is done using the sieve of Erathostenes (see [BS96] p. 296). For each prime $p \in P$ we compute

$$d_p = \det B \bmod p \ . \tag{A.6}$$

This can, for example, be done using Gaussian elimination. Then we determine the absolute smallest solution of the simultaneous congruence

$$d \equiv d_p \pmod{p} \ , \quad p \in P \ . \tag{A.7}$$

The determinant of B is d.

Example A.5.2. Let

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & -5 & 6 \\ 7 & 8 & 9 \end{pmatrix} .$$

We compute the value c from (A.3). We have $H(B)^2 = (1+16+49)(4+25+64)(9+36+81) = 773388$. We also have $H(B^T)^2 = (1+4+9)(16+25+36)(49+64+81) = 209132$. Hence $c = \lfloor\sqrt{209132}\rfloor = 457$. We set $P = 2, 3, 5, 7, 11$. Then $\prod_{p\in P} p = 2310 > 2\cdot 457$. Next we compute the determinant of B modulo each prime in P. We find that $\det B \equiv 0 \pmod 2$, $\det B \equiv 0 \pmod 3$, $\det B \equiv 0 \pmod 5$, $\det B \equiv 1 \pmod 7$, $\det B \equiv 10 \pmod{11}$. Chinese remaindering tells us that $\det B = 120$.

We use the running time estimate from [vG99] Theorem 5.12.

Proposition A.5.3. *If the entries of B are bounded by C then the computation of $\det B$ requires time $\mathrm{O}\big(n^4 \log^2(nC)(\log^2 n + (\log\log C)^2)\big)$.*

A.5.2 Diagonally dominant matrices

Definition A.5.4. *Let $n \in \mathbb{N}$. A square matrix $M \in \mathbb{C}^{(n,n)}$, $M = (m_{i,j})_{1\le i,j\le n}$ is called* strictly diagonally dominant *if $|m_{j,j}| > \sum_{i=1,i\ne j}^{n} |m_{i,j}|$.*

Lemma A.5.5. *A strictly diagonally dominant complex matrix is non-singular.*

Proof. Let $M = (m_{i,j})$ be strictly diagonally dominant. Assume M is singular, and there exists $\mathbf{v} = (v_i)$ with $\mathbf{v}M = 0$. Let j be such that $|v_i| \le |v_j|$ for all $i \ne j$. Then

$$v_j m_{j,j} + \sum_{i\ne j} v_i m_{i,j} = 0$$

and

$$|v_j| \cdot |m_{j,j}| \le \sum_{i\ne j} |v_i| \cdot |m_{i,j}| \le \sum_{i\ne j} |v_j| \cdot |m_{i,j}| < |v_j| \cdot |m_{j,j}| .$$

This contradiction proves the lemma. □

A.5.3 Hermite normal form

We define the Hermite normal form.

Definition A.5.6. *A non-singular matrix $H \in \mathbb{Z}^{n,n}$ is in* Hermite normal form *if H is an upper triangular matrix whose diagonal elements are positive and whose off diagonal elements are non-negative and smaller than the diagonal element in the same row.*

Example A.5.7. The matrix

$$H = \begin{pmatrix} 3 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

is in Hermite normal form.

We show that B can be transformed into Hermite normal form by an unimodular transformation.

Proposition A.5.8. *There is a matrix $H \in \mathbb{Z}^{(n,n)}$ in Hermite normal form and a matrix $V \in \mathrm{GL}(n, \mathbb{Z})$ such that $BV = H$. The matrix H is uniquely determined by B.*

Proof. We first prove the uniqueness. We will demonstrate the existence by presenting an algorithm for computing the Hermite normal form.

Let $T \in \mathrm{GL}(n, \mathbb{Z})$ and let $H = BT$ be in Hermite normal form. Then the columns of H form a basis of the lattice $L(B)$ in $\mathbb{Z}^n$ which is generated by the columns of B. Let $H = (h_{i,j}$. We have

$$h_{j,j} = \min\left\{ v_j > 0 : \text{ there is } (v_1, \ldots, v_j, 0, \ldots, 0) \in L \right\} . \tag{A.8}$$

Hence, $h_{j,j}$ is uniquely determined. The uniqueness of the diagonal elements implies the uniqueness of the off-diagonal elements.

We prove the existence of the Hermite normal form by presenting an algorithm for transforming B into its Hermite normal H. The algorithm is very similar to Gaussian elimination. Initially, we set $H = B$. Denote the entries of H by $h_{i,j}$, $1 \le i, j \le n$ and denote the column vectors of H by $(\mathbf{h}_1, \ldots, \mathbf{h}_n)$. Starting in the lower right corner, we transform H into upper triangular form, that is, we eliminate $h_{i,j}$ for $i = n, n-1, \ldots, 1$ and $j = i-1, \ldots, 1$. We explain one elimination. Let $i \in \{1, \ldots, \}$ and $j \in \{1, \ldots, i-1\}$. Assume that $b_{i,j} \neq 0$. To eliminate $b_{i,j}$ we apply the extended euclidean algorithm `xgcd` to the pair $b_{i,j}$, $b_{i,i}$. Let (g, x, y) be the result, i.e. $g = \gcd(b_{i,j}, b_{i,i})$ and $xb_{i,j} + yb_{i,i} = g$.

Then the matrix

$$U = \begin{pmatrix} b_{i,i}/g & x \\ -b_{i,j}/g & y \end{pmatrix} \tag{A.9}$$

has determinant 1. We replace $(\mathbf{b}_j, \mathbf{b}_i)$ by $(\mathbf{b}_j, \mathbf{b}_i)U = \big((b_{i,i}/g)\mathbf{b}_j - (b_{i,j}/g)\mathbf{b}_i, x\mathbf{b}_j + y\mathbf{b}_i\big)$. Then the ith entry of $\mathbf{b}_j$ is zero. Also, the columns of B still form a basis of $L(B)$ because $\det U = 1$. After the triangularization is finished, we make the diagonal elements positive by multiplying the columns, in which the diagonal element is negative, with -1. We also replace the elements above the diagonal by the smallest non-negative residue modulo the diagonal element in the same row. □

From the proof of uniqueness we immediately see that for any unimodular T and any $B \in \mathbb{Z}^{(n,n)}$, the Hermite Normal Form of B and BT coincide. Hence, this HNF is an invariant of the lattice $L(B)$.

Corollary A.5.9. *Let L be an n-dimensional lattice in $\mathbb{Z}^n$. Then L has exactly one basis in Hermite normal form which is called the HNF-basis of L.*

Here is an example of a HNF computation.

Example A.5.10. We compute the HNF of the matrix

$$B = \begin{pmatrix} 385083 & 56286 & 26979 \\ 211248 & 30786 & 14712 \\ 24879 & 3624 & 1731 \end{pmatrix} .$$

We obtain the following elimination steps.

$$H = \begin{pmatrix} 385083 & -113610 & 40017 \\ 211248 & -8574 & 3042 \\ 24879 & 0 & 3 \end{pmatrix} .$$

$$H = \begin{pmatrix} -331475898 & -113610 & 40017 \\ -25016058 & -8574 & 3042 \\ 0 & 0 & 3 \end{pmatrix} .$$

$$H = \begin{pmatrix} 12 & 78 & 40017 \\ 0 & 6 & 3042 \\ 0 & 0 & 3 \end{pmatrix} .$$

Finally, the off diagonal elements are reduced modulo the diagonal elements. We obtain.

$$H = \begin{pmatrix} 12 & 6 & 3 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{pmatrix} .$$

In the above HNF-algorithm the intermediate entries can become very large. For another instructive example see [HM91]. Therefore, we use the modification from [DKJ87]. In this modification, all computations are done modulo

$$d = |\det B| . \tag{A.10}$$

This is possible because of the following result.

Lemma A.5.11. *The columns of dI_n belong to the lattice $L(B)$.*

Proof. This follows from Cramer's rule. □

Thus after each elimination step, we can subtract a suitable linear combination of column vectors of dI_n thereby ensuring that we never encounter numbers larger than d^2. Note that ths reduction modulo d corresponds to a unimodular transformation of $H \circ dI_n$ which preserves the lattice $L(H \circ dI_n) = L(B)$.

We modify our example.

Example A.5.12. We again compute the HNF of the matrix

$$B = \begin{pmatrix} 385083 & 56286 & 26979 \\ 211248 & 30786 & 14712 \\ 24879 & 3624 & 1731 \end{pmatrix} .$$

We have

$$d = \det(B) = 216 .$$

Reducing B mod d we obtain

$$H = \begin{pmatrix} 171 & 6 & 195 \\ 0 & 66 & 24 \\ 39 & 0 & 3 \end{pmatrix} .$$

Then we obtain the following elimination steps.

$$H = \begin{pmatrix} 12 & 6 & 195 \\ 120 & 66 & 24 \\ 0 & 0 & 3 \end{pmatrix} .$$

$$H = \begin{pmatrix} 12 & 6 & 195 \\ 0 & 6 & 24 \\ 0 & 0 & 3 \end{pmatrix} .$$

We reduce the off diagonal elements modulo the diagonal elements and obtain the HNF

$$H = \begin{pmatrix} 12 & 6 & 3 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{pmatrix} .$$

We present the modular HNF algorithm formally on page 299.

We leave the proof of the correctness of this algorithm to the reader. Note that `hnfModular` does not calculate a transformation matrix T with $H = BT$.

We analyze the complexity of `hnfModular`.

Proposition A.5.13. *After computing the determinant of B and reducing the entries of B modulo that determinant, algorithm* `hnfModular` *has running time* $\mathrm{O}\big(n^3(\log d)^2\big)$.

Algorithm A.1 `eliminateRowModular` $(B,\ d,\ i)$

Input: $B \in \mathbb{Z}^{(n,n)}$, $d = \det B$, $i \in \{1, \dots, n\}$
Output: $H \in \mathbb{Z}^{(n,n)}$, such that $L(H \circ dI_n) = L(B \circ dI_n)$ and the first $i - 1$ entries in the ith row of H are zero

```
for ($j \leftarrow i-1$, $j \geq 1$, $j \leftarrow j-1$) do
    if $h_{i,j} \neq 0$ then
        $\mathbf{h} \leftarrow \mathbf{h}_i$.
        $(g, x, y) \leftarrow$ xgcd$(h_{i,j}, h_{i,i})$.
        $\mathbf{h}_j \leftarrow (h_{i,i}/g)\mathbf{h}_j - (h_{i,j}/g)\mathbf{h} \bmod d$.
        $\mathbf{h}_i \leftarrow x\mathbf{h}_j + y\mathbf{h} \bmod d$.
return $H$.
```

Algorithm A.2 `reduceColumnsModDiagonalModular` $(H,\ d)$

Input: $B \in \mathbb{Z}^{(n,n)}$ non-singular in upper triangular form with non-negative entries, $d \in \mathbb{Z}_{>0}$
Output: $H \in \mathbb{Z}^{(n,n)}$, such that $L(H \circ dI_n)$ is unchanged and $0 \leq h_{i,j} < h_{i,i}$, $1 \leq i \leq n$, $i < j \leq n$

```
for ($i \leftarrow n-1$, $i \geq 1$ , $i \leftarrow i-1$) do
    for ($j = i+1$, $j < n$, $j \leftarrow j+1$) do
        $m = \lfloor h_{i,j}/h_{i,i} \rfloor$.
        $\mathbf{h}_j = \mathbf{h}_j - m \cdot \mathbf{h}_i \bmod d$.
return H.
```

Algorithm A.3 `hnfModular` (B)

Input: $B \in \mathbb{Z}^{(n,n)}$ non-singular
Output: The Hermite normal form H of B

```
$d \leftarrow \det B$.
$H \leftarrow B \bmod d$.
for ($i \leftarrow n$, $i \geq 1$, $i \leftarrow i-1$) do
    Permute the first $i$ columns of $H$ in such a way that $h_{i,i} \neq 0$.
    eliminateRowModular$(H, d, i)$.
reduceColumnsModDiagonalModular $(H,\ d)$.
return $H$.
```

Proof. In `eliminateRowModular` on page 299 and `reduceColumnsModDiagonalModular` on page 299 the iteration proceeds over all $n(n-1)$ off-diagonal entries of H. For each of these entries one or two linear combinations of columns are computed. This leads to n^3 operations with numbers of size $O(\log d)$.

We see that it is the computation of the determinant of B which is the expensive part in the modular computation of the HNF of B. Indeed, modern algorithms compute the HNF directly, and read the determinant off it.

A.5.4 Smith normal form

Next, we define the Smith normal form.

Definition A.5.14. *A matrix $S \in \mathbb{Z}^{(n,n)}$ is in* Smith normal form *if S is a diagonal matrix, $S = \operatorname{diag}(s_1, \dots, s_n)$, its diagonal elements are positive and s_i divides s_{i-1} for $1 < i \leq n$.*

Proposition A.5.15. *Let B be a non-singular $n \times n$ matrix. Then there is a matrix $S \in \mathbb{Z}^{(n,n)}$ in Smith normal form, a matrix $U \in \mathbb{Z}^{(n,n)}$ which is invertible mod $\det B$ and $V \in \mathrm{GL}(n, \mathbb{Z})$ such that $BV = US$. The matrix S is uniquely determined by B.*

The existence part of the statement we will prove constructively in the remainder of this section. Uniqueness follows from the third claim of Proposition 9.7.3 if we apply it to the finite group $G = \mathbb{Z}^n / L(B)$ where $L(B)$ is the subgroup of $\mathbb{Z}^n$ generated by the columns of B. Note that the proof of Proposition 9.7.3 uses only the existence part of Proposition A.5.15.

We now present an algorithm for computing the Smith normal form. In this algorithm we use `eliminateColumnModular`(T, S, d, i), shown below, which eliminates the off-diagonal entries in the j-th column of S and updates the transformation T. In this algorithm we denote the *rows* of matrix S by $\mathbf{s}_j$, and the *columns* of T by $\mathbf{t}_j$.

Algorithm A.4 `eliminateColumnModular` (T, S, d, j)

Input: $T, S \in \mathbb{Z}^{(n,n)}$, $d \in \mathbb{Z}_{>0}$, $j \in \{1, \dots, n\}$
Output: $T, S \in \mathbb{Z}^{(n,n)}$ such that $L(TS \circ dI_n)$ is not changed and the first $j - 1$ entries in the jth column of S are zero

```
for $(i \leftarrow j-1,\ i \geq 1,\ i \leftarrow i-1)$ do
    $(d, x, y) \leftarrow \texttt{xgcd}(s_{i,j}, s_{i,i})$.
    $\mathbf{s} \leftarrow \mathbf{s}_j$.
    $\mathbf{s}_j \leftarrow x\mathbf{s} + y\mathbf{s}_i \bmod d$.
    $\mathbf{s}_i \leftarrow (s_{i,j}/d)\mathbf{s} - (s_{j,j}/d)\mathbf{s}_i \bmod d$.
    $\mathbf{t} \leftarrow \mathbf{t}_j$.
    $\mathbf{t}_j \leftarrow (-s_{i,j}/d)\mathbf{t}_i + (-s_{j,j}/d)\mathbf{t} \bmod d$.
    $\mathbf{t}_i \leftarrow y\mathbf{t}_i - x\mathbf{t}_j \bmod d$.
return $T, S$
```

We also use `eliminateRowAndColumnModular`(T, P, d, i) which eliminates the ith row and the ith column of P and updates the transformation T.

Proposition A.5.16. *If the entries of the matrix S are in absolute value bounded by d, then* `eliminateRowAndColumnModular` *(S, d, i) has running time $\mathrm{O}\big(n^2 (\log d)^3\big)$.*

Algorithm A.5 `eliminateRowAndColumnModular` (T, S, d, j)

Input: $T, S \in \mathbb{Z}^{(n,n)}$, $d \in \mathbb{Z}_{>0}$, $j \in \{1, \dots, n\}$
Output: $T, S \in \mathbb{Z}^{(n,n)}$ such that $L(TS \circ dI_n)$ is not changed and the first $j-1$ entries in the jth row and column of S are zero

```
repeat
  eliminateColumnModular($T, S, d, j$)
  eliminateRowModular($T, S, d, j$)
until all of the first $j-1$ entries in the $j$th row and column of $S$ are zero
return $T, S$
```

Algorithm A.6 `snfModular` (B)

Input: $B \in \mathbb{Z}^{(n,n)}$ non-singular
Output: The Smith normal form S of B and $T \in \mathbb{Z}^{(n,n)}$ non-singular mod $\det B$ such that $TS = BV$ for some $V \in \mathrm{GL}(n, \mathbb{Z})$

```
$d \leftarrow |\det(B)|$.
$S \leftarrow B$ mod $d$.
$T \leftarrow I_n$.
for ($i \leftarrow n$, $i \geq 1$, $i \leftarrow i-1$) do
  Permute the first $i$ columns of $S$ in such a way that $s_{i,i} \neq 0$.
  eliminateRowAndColumnModular($T, S, d, i$).
  while $s_{i,i} \nmid s_{kl}$ for some $(k, l) \in \{1, \dots, i-1\}^2$ do
    $\mathbf{s}_i \leftarrow \mathbf{s}_i + \mathbf{s}_k$.
    eliminateRowAndColumnModular($T, S, d, i$).
return $S$.
```

Proof. The algorithm applies `eliminateColumnModular` after `eliminateRowModular`. If in `eliminateColumnModular` the entry $s_{j,j}$ divides all entries $s_{i,j}$ for $1 \leq i < j$ then the jth row of S is unchanged since the gcd representations are $s_{j,j} = \gcd(s_{i,j}, s_{j,j}) = 0 \cdot s_{i,j} + 1 \cdot s_{j,j}$. So in this case, `eliminateRowAndColumnModular` terminates after the application of `eliminateColumnModular`. If in `eliminateColumnModular` the jth row of S is modified, then $s_{j,j}$ is replaced by a proper divisor of $s_{j,j}$. This can happen only $\mathrm{O}(\log d)$ times. □

Proposition A.5.17. *After computing the determinant of B and reducing the entries of B modulo that determinant, algorithm* `snfModular`(B) *has running time* $\mathrm{O}\big(n^3(\log d)^3\big)$.

Proof. By the same argument as in the proof of Proposition A.5.16, we can establish that the sub-algorithms `eliminateColumnModular` and `eliminateRowModular` are called no more than $O(n \cdot \log d)$ times. As we have seen before, each such call can be executed in time $O(n^2(\log d)^2)$. □

Example A.5.18. We compute the SNF of the matrix

$$B = \begin{pmatrix} 385083 & 56286 & 26979 \\ 211248 & 30786 & 14712 \\ 24879 & 3624 & 1731 \end{pmatrix} .$$

Initially, we have $T = I_3$, $S = B$. After applying `eliminateRowAndColumnModular` we obtain the following matrices

$$T = \begin{pmatrix} 1 & 0 & 65 \\ 0 & 1 & 8 \\ 0 & 0 & 1 \end{pmatrix} , \quad S = \begin{pmatrix} 12 & 6 & 0 \\ 120 & 66 & 0 \\ 0 & 0 & 3 \end{pmatrix} ,$$

and

$$T = \begin{pmatrix} 1 & 1 & 65 \\ 0 & 1 & 8 \\ 0 & 0 & 1 \end{pmatrix} , \quad S = \begin{pmatrix} 12 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{pmatrix} .$$

A.5.5 Algorithms for rectangular matrices

We extend the notions of Hermite and Smith normal forms to rectangular matrices, and list the complexity of fast algorithms for their computation. For simplicity, we assume that all matrices have full rank.

We say that a rectangular matrix $H \in \mathbb{Z}^{(n,m)}$ with $n \leq m$ is in Hermite normal form if $H = H' \circ 0$ where $H' \in \mathbb{Z}^{(n,n)}$ is in Hermite normal form in the sense of Definition A.5.6. We say that H is the Hermite normal form of a given rectangular full-rank matrix $B \in \mathbb{Z}^{(n,m)}$ with $n \leq m$ if there exists unimodular $U \in \mathbb{Z}^{(m,m)}$ such that $H = BU$. The matrix S is the Smith normal form of B if S is the Smith normal form of H' where $H = H' \circ 0$.

Proposition A.5.19 ([Sto00]). *There are algorithms that compute given a full-rank matrix $B \in \mathbb{Z}^{(n,m)}$ with $n \leq m$ its Hermite and Smith normal forms in time $O(n^3 m b^2)$ where b is a bound on the size of the entries of B.*

The given complexity will be improved if faster than standard matrix multiplication algorithms are employed. Another improvement is possible if one knows in advance that the Hermite normal form of the given matrix has few entries on the main diagonal which differ from 1, cf. [Vol03]. This improvement uses fast algorithms for the solution of linear Diophantine systems which are interesting in their own right.

For a matrix $B \in \mathbb{Z}^{n,m}$ we define $\mathcal{L}(B)$ to be the lattice generated by the columns of B in $\mathbb{Z}^n$, and $\mathcal{L}_{\mathbb{Q}}(B) = \mathcal{L}(B) \otimes \mathbb{Q}$ to be the sub-vector space generated by $\mathcal{L}(B)$ in $\mathbb{Q}^n$.

Proposition A.5.20 ([MS99]). *There is an algorithm that given a matrix $B \in \mathbb{Z}^{n,m}$ and a column vector $\mathbf{b} \in \mathcal{L}_{\mathbb{Q}}(B)$ computes minimal $d \in \mathbb{N}$ such that $d\mathbf{b} \in \mathcal{L}(B)$ and a solution vector $\mathbf{x} \in \mathbb{Z}^m$ for the system*

$$B\mathbf{x} = d\mathbf{b}$$

such that the size of the entries of $\mathbf{x}$ *is in* $O(n(b + \log m))$. *The algorithm executes in time* $O\ (n^2mb^2)$. *Here* b *is an upper bound for the size of the entries of* B *and* $\mathbf{b}$.

A.6 Exercises

Exercise A.6.1. Suppose that L is a lattice in $\mathbb{R}^n$ and that $\mathbf{v} \in L$ is a nonzero lattice point which is not an integer multiple of another lattice point. Prove that there is a basis of L with first element $\mathbf{v}$.

Exercise A.6.2. Let G be a group. Prove that the map $G \times G \to G$, $(g, h) \mapsto gh$ defines an action of G on G. Determine the number of G-orbits with respect to this decomposition.

Exercise A.6.3. Prove the correctness of algorithm `hnfModular`.

Exercise A.6.4. Generalization of `hnfModular` to rectangular matrices.

Exercise A.6.5. Non-modular HNF algorithm with transformation.

Chapter references and further reading

[BS96] Eric Bach and Jeffrey Shallit, *Algorithmic number theory*, MIT Press, Cambridge, Massachusetts and London, England, 1996.

[DKJ87] Paul D. Domich, Ravi Kannan, and Leslie E. Trotter Jr., *Hermite normal form computation using modular determinant arithmetic*, Mathematics of Operations Research **12** (1987), 50–59.

[GJS01] Mark Giesbrecht, Michael J. Jacobson, Jr., and Arne Storjohann, *Algorithms for large integer matrix problems*, Applied algebra, algebraic algorithms and error-correcting codes (Melbourne, 2001), Lecture Notes in Computer Science, vol. 2227, Springer, Berlin, 2001, pp. 297–307.

[HJ85] Roger A. Horn and Charles R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1985.

[HM91] James L. Hafner and Kevin S. McCurley, *Asymptotically fast triangularization of matrices over rings*, SIAM Journal on Computing **20** (1991), 1068–1083.

[MS99] Thom Mulders and Arne Storjohann, *Diophantine linear system solving*, International Symposium on Symbolic and Algebraic Computation, ISSAC '99 (Sam Dooley, ed.), ACM Press, 1999.

[Sto00] Arne Storjohann, *Algorithms for matrix canonical forms*, Ph.D. thesis, ETH Zürich, 2000.

[The00] Patrick Theobald, *Berechnung von Hermite-Normalformen*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, Darmstadt, Germany, 2000, German.

[vG99] Joachim von zur Gathen and Jürgen Gerhard, *Modern computer algebra*, Cambridge University Press, Cambridge, UK, 1999.

[Vol03] Ulrich Vollmer, *A note on the Hermite basis computation of large integer matrices*, International Symposium on Symbolic and Algebraic Computation, ISSAC '03 (J. Rafael Sendra, ed.), ACM Press, 2003, pp. 255–257.

Bibliography

Books

The following works provide other view-points on the subject of this book and lead the reader further into the fields of algebraic and algorithmic number theory.

[BS66] Zenon I. Borevich and Igor R. Shafarevich. *Number theory.* Academic Press, New York, 1966.

[BS96] Eric Bach and Jeffrey Shallit. *Algorithmic Number Theory.* MIT Press, Cambridge, Massachusetts and London, England, 1996.

[Buc04] Johannes Buchmann. *Introduction to Cryptography.* Springer-Verlag, second edition, 2004. Undergradute Texts in Mathematics.

[Bue89] Duncan A. Buell. *Binary quadratic forms.* Springer-Verlag, New York, 1989.

[Coh00] Henri Cohen. *A course in computational algebraic number theory*, volume 138 of *Graduate Texts in Mathematics.* Springer-Verlag, fourth corrected edition, 2000.

[Gau86] Carl Friedrich Gauss. *Disquisitiones Arithmeticae.* Springer-Verlag, New York, English edition, 1986. Translated by A. Clark.

[Lan66] Edmund Landau. *Elementary Number Theory.* Chelsea Publishing Company, second edition, 1966.

[Mol96] Richard A. Mollin. *Quadratics.* CRC Press, 1996.

[PZ89] Michael Pohst and Hans Zassenhaus. *Algorithmic Algebraic Number Theory.* CUP, 1989.

[Zag81] Don B. Zagier. *Zetafunktionen und quadratische Zahlkörper.* Springer-Verlag, 1981.

Surveys

[BB93] Ingrid Biehl and Johannes Buchmann. Algorithms for quadratic orders. In *Proceedings of Symposium on Mathematics of Computation*, 1993.

[BH01] Johannes Buchmann and Safuat Hamdy. A survey on IQ-cryptography. In *Public Key Cryptography and Computational Number Theory*, pages 1–15. Walter de Gruyter, 2001.

[Buc91] Johannes Buchmann. Number theoretic algorithms and cryptology. In Lothar Budach, editor, *Fundamentals of Computation Theory, FCT '91*, volume 529 of *Lecture Notes in Computer Science*, pages 16–21. Springer-Verlag, 1991.

[BW90] Johannes Buchmann and Hugh C. Williams. Quadratic fields and cryptography. In John H. Loxton, editor, *Number Theory and Cryptography*, volume 154 of *London Mathematical Society Lecture Note Series*, pages 9–25. Cambridge University Press, 1990.

[Len92] Hendrik W. Lenstra. Algorithms in algebraic number theory. *Bulletin of the AMS (N.S.)*, 26:211–244, 1992.

[LL90] Arjen K. Lenstra and Hendrik W. Lenstra Jr. Algorithms in number theory. In J. van Leeuwen, editor, *Handbook of theoretical computer science. Volume A. Algorithms and Complexity*, chapter 12, pages 673–715. Elsevier, 1990.

[MW92] Richard A. Mollin and Hugh C. Williams. On real quadratic fields of class number two. *Mathematics of Computation*, 59(200):625–632, 1992.

[Wil85] Hugh C. Williams. Continued fractions and number-theoretic computations. *The Rocky Mountain Journal of Mathematics*, 15(2):621–655, 1985. Number theory (Winnipeg, Man., 1983).

Other Referenced Works

[Abe94] Christine Abel, *Ein Algorithmus zur Berechnung der Klassenzahl und des Regulators reellquadratischer Ordnungen*, Ph.D. thesis, Universität des Saarlandes, Saarbrücken, Germany, 1994, German.

[ABR01] Michel Abdalla, Mihir Bellare, and Philip Rogaway, *An encryption scheme based on the Diffie-Hellman problem*, Progress in Cryptology — CT-RSA 2001 (David Naccache, ed.), Lecture Notes in Computer Science, vol. 2020, Springer-Verlag, 2001, pp. 143–158.

[Ajt98] Miklos Ajtai, *The shortest vector problem in l_2 is NP-hard for randomized reductions (extended abstract)*, Proceedings of the 35th annual ACM Symposium on Theory of Computing (Jeffrey Vitter, ed.), ACM Press, 1998, pp. 10–19.

[Apo86] Tom M. Apostol, *Introduction to analytic number theory*, Springer-Verlag, 1986.

[Bac90] Eric Bach, *Explicit bounds for primality testing and related problems*, Mathematics of Computation **55** (1990), 355–380.

[Bac95] ———, *Improved approximations for Euler products*, Fourth Conference of the Canadian Number Theory Association (Karl Dilcher, ed.), CMS Proceedings, vol. 15, Canadian Mathematical Society, 1995, pp. 13–28.

[BB99] Ingrid Biehl and Johannes Buchmann, *An analysis of the reduction algorithms for binary quadratic forms*, Voronoi's Impact on Modern Science, Institute of Mathematics Kyiv (Peter Engel and Halyna M. Syta, eds.), National Academy of Sciences of Ukraine, 1999, pp. 71–98.

[BBHM02] Ingrid Biehl, Johannes Buchmann, Safuat Hamdy, and Andreas Meyer, *A signature scheme based on the intractability of computing roots*, Designs, Codes and Cryptography **25** (2002), no. 3, 223–236.

[BBT94] Ingrid Biehl, Johannes Buchmann, and Christoph Thiel, *Cryptographic protocols based on discrete logarithms in real-quadratic orders*, Advances in Cryptology – CRYPTO '94 (Yvo G. Desmedt, ed.), Lecture Notes in Computer Science, vol. 839, Springer-Verlag, 1994, pp. 56–60 (English).

[BD89] Johannes Buchmann and Stephan Düllmann, *A probabilistic class group and regulator algorithm and its implementation*, Proc. Colloquium on Computational Number Theory, Walter de Gruyter, 1989, pp. 5–72.

[BD91a] ———, *Distributed class group computation*, Informatik (Johannes Buchmann, Harald Ganzinger, and Wolfgang J. Paul, eds.), Teubner-Texte zur Informatik, vol. 1, B. G. Teubner, 1991, pp. 68–81.

[BD91b] ———, *On the computation of discrete logarithms in class groups*, Advances in Cryptology – CRYPTO '90 (Alfred J. Menezes and Scott A. Vanstone, eds.), Lecture Notes in Computer Science, vol. 537, Springer-Verlag, 1991, pp. 134–139.

[BDW90] Johannes Buchmann, Stephan Düllmann, and Hugh C. Williams, *On the complexity and efficiency of a new key exchange system*, Advances in Cryptology – EUROCRYPT '89 (Jean-Jacques Quisquater and Joos Vandewalle, eds.), Lecture Notes in Computer Science, vol. 434, Springer-Verlag, 1990, pp. 597–616.

[Ber] Daniel J. Bernstein, *How to find small factors of integers*, Mathematics of Computation (to appear), `http://cr.yp.to/papers/sf.ps`.

[BH96] Johannes Buchmann and Christine S. Hollinger, *On smooth ideals in number fields*, Journal of Number Theory **59** (1996), no. 1, 82–87.

[BH03] Mark L. Bauer and Safuat Hamdy, *On class group computations using the number field sieve*, Advances in Cryptology – ASIACRYPT 2003 (Chi-Sung Laih, ed.), Lecture Notes in Computer Science, vol. 2894, Springer-Verlag, 2003, pp. 311–325.

[BJN$^+$98] Johannes Buchmann, Michael J. Jacobson, Jr., Stefan Neis, Patrick Theobald, and Damian Weber, *Sieving methods for class group computation*, Algorithmic Algebra and Number Theory: Selected Papers from a Conference Held at the Univerity of Heidelberg in October 1997, Springer-Verlag, 1998, pp. 3–10.

[BJT97] Johannes Buchmann, Michael J. Jacobson, Jr., and Edlyn Teske, *On some computational problems in finite abelian groups*, Mathematics of Computation **66** (1997), no. 220, 1663–1687.

[BMM00] Johannes Buchmann, Markus Maurer, and Bodo Möller, *Cryptography based on number fields with large regulator*, Journal de Théorie des Nombres de Bordeaux **12** (2000), 293–307.

[BMT96] Ingrid Biehl, Bernd Meyer, and Christoph Thiel, *Cryptographic protocols based on real-quadratic A-fields (extended abstract)*, Advances in Cryptology – ASIACRYPT '96 (Kwangjo Kim and Tsutomu Matsumoto, eds.), Lecture Notes in Computer Science, vol. 1163, Springer-Verlag, 1996, pp. 15–25.

[BP95] Johannes Buchmann and Sachar Paulus, *Algorithms for finite abelian groups*, Proceedings of Number Theoretic and Algebraic Methods in Computer Science (NTAMCS) '93 (Singapor) (van der Poorten, Shparlinski, and Zimmer, eds.), World Scientific, 1995.

[BP97] ———, *A one way function based on ideal arithmetic in number fields*, Advances in Cryptology – CRYPTO '97 (Burton S. Kaliski, ed.), Lecture Notes in Computer Science, vol. 1294, Springer-Verlag, 1997, pp. 385–394.

[BPT04] Ingird Biehl, Sachar Paulus, and Tsuyoshi Takagi, *Efficient undeniable signature schemes based on ideal arithmetic in quadratic orders*, Designs, Codes and Cryptography **31** (2004), no. 2, 99–123.

[BR97] Mihir Bellare and Philip Rogaway, *Minimizing the use of random oracles in authenticated encryption schemes*, Information and Communications Security, ICIS '97 (Y. Han, T. Okamoto, and S. Quing, eds.), Lecture Notes in Computer Science, vol. 1334, Springer-Verlag, 1997, pp. 1–16.

[Bra90] Gilles Brassard (ed.), *Advances in cryptology – CRYPTO '89*, Lecture Notes in Computer Science, vol. 435, Springer-Verlag, 1990.

[BS05] Johannes Buchmann and Arthur Schmidt, *Computing the structure of a finite abelian group*, Mathematics of Computation **74** (2005), 2017–2026.

[BST02] Johannes Buchmann, Kouichi Sakurai, and Tsuyoshi Takagi, *An IND-CCA2 public-key cryptosystem with fast decryption*, Information Security and Cryptology - ICISC 2001 (Kwangjo Kim, ed.), Lecture Notes in Computer Science, vol. 2288, Springer-Verlag, 2002, pp. 51–71.

[BSW90] Johannes Buchmann, Renate Scheidler, and Hugh C. Williams, *Implementation of a key exchange protocol using real quadratic fields*, Proc. of EUROCRYPT '90, Lecture Notes in Computer Science, vol. 473, Springer-Verlag, 1990, pp. 98–109.

[BSW94] ———, *A key-exchange protocol using real quadratic fields*, Journal of Cryptology **7** (1994), 171–199.

[BTW95] Johannes Buchmann, Christoph Thiel, and Hugh C. Williams, *Short representation of quadratic integers*, Computational Algebra and Number Theory, Sydney 1992 (Wieb Bosma and Alf J. van der Poorten, eds.), Mathematics and its Applications, vol. 325, Kluwer Academic Publishers, 1995, pp. 159–185.

[Buc90a] Johannes Buchmann, *Complexity of algorithms in algebraic number theory*, Number Theory, Banff, Alberta 1988 (Richard A. Mollin, ed.), Walter de Gruyter Publishers, 1990, pp. 37–53 (English).

[Buc90b] ———, *A subexponential algorithm for the determination of class groups and regulators of algebraic number fields*, Séminaire de Théorie des Nombres, Paris 1988–1989 (Catherine Goldstein, ed.), Progress in Mathematics, vol. 91, Birkhäuser, 1990, pp. 27–41.

[Bue89] Duncan A. Buell, *Binary quadratic forms*, Springer, New York, 1989.

[BW88] Johannes Buchmann and Hugh C. Williams, *A key-exchange system based on imaginary quadratic fields*, Journal of Cryptology **1** (1988), no. 2, 107–118 (English).

[BW89] ———, *On the existence of a short proof for the value of the class number and regulator of a real quadratic field*, Number Theory and Applications, Calgary 1988 (Richard A. Mollin, ed.), NATO ASI Series, Series C, vol. 265, Kluwer Academic Publishers, 1989, pp. 327–345.

[BW90] ———, *A key-exchange system based on real quadratic fields*, in Brassard [Bra90], pp. 335–343.

[BW91] ———, *Some remarks concerning the complexity of computing class groups of quadratic fields*, J. Complexity **7** (1991), no. 3, 311–315 (English).

[CDO93] Henri Cohen, Francisco Diaz y Diaz, and Michel Olivier, *Calculs de nombres de classes et de régulateurs de corps quadratiques en temps sous-exponentiel*, Séminaire de Théorie des Nombres, Paris 1990–1991 (Sinnou David, ed.), Progress in Mathematics, vol. 108, Birkhäuser, 1993, French, pp. 35–46.

[CDO97] ———, *Subexponential algorithm for class group and unit computation*, Journal of Symbolic Computing **24** (1997), no. 3/4, 433–441.

[Cox89] David A. Cox, *Primes of the form $x^2 + ny^2$*, Wiley, New York, 1989.

[CyDO01] Henri Cohen, Franzisco Diaz y Diaz, and Michel Olivier, *Algorithmic methods for finitely generated abelian groups*, Journal of Symbolic Computation **31** (2001), no. 1-2, 133–147.

[DH76] Whitfield Diffie and Martin E. Hellman, *New directions in cryptography*, IEEE Transactions on Information Theory **IT-22** (1976), no. 6, 644–654.

[DKJ87] Paul D. Domich, Ravi Kannan, and Leslie E. Trotter Jr., *Hermite normal form computation using modular determinant arithmetic*, Mathematics of Operations Research **12** (1987), 50–59.

[DSS00] *Digital signature standard*, Federal Information Processing Standards Publication FIPS 186-2, NIST, 2000.

[ElG85] Taher ElGamal, *A public key cryptosystem and a signature scheme based on discrete logarithms*, IEEE Transactions on Information Theory **31** (1985), no. 4, 469–472.

[FP03] Pierre-Alain Fouque and Guillaume Poupard, *On the security of RDSA*, Advances in Cryptology – EURCRYPT 2003 (Eli Biham, ed.), Lecture Notes in Computer Science, vol. 2656, Springer-Verlag, 2003.

[FS87] Amos Fiat and Adi Shamir, *How to prove yourself: Practical solutions to identification and signature problems*, Advances in Cryptology – CRYPTO '86 (Andrew M. Odlyzko, ed.), Lecture Notes in Computer Science, vol. 263, Springer-Verlag, 1987, pp. 186–194.

[GJS01] Mark Giesbrecht, Michael J. Jacobson, Jr., and Arne Storjohann, *Algorithms for large integer matrix problems*, Applied algebra, algebraic algorithms and error-correcting codes (Melbourne, 2001), Lecture Notes in Computer Science, vol. 2227, Springer, Berlin, 2001, pp. 297–307.

[GQ88] Louis C. Guillou and Jean-Jacques Quisqater, *A practical zero-knowledge protocol fitted to security microprocessors minimizing both transmission and memory*, Advances in Cryptology – EUROCRYPT '88 (Christoph G. Günther, ed.), Lecture Notes in Computer Science, vol. 330, Springer-Verlag, 1988, pp. 123–128.

[Hal02] Sean Hallgren, *Polynomial-time quantum algorithms for Pell's equation and the principal ideal problem*, Proceedings of the thiry-fourth annual ACM symposium on the theory of computing, ACM Press, 2002, pp. 653–658.

[Hal05] ———, *Fast quantum algorithms for computing the unit group and class group of a number field*, Proceedings of the 37th annual ACM Symposium on Theory of Computing (Harold N. Gabow and Ronald Fagin, eds.), ACM Press, 2005, pp. 468–474.

[Ham] Safuat Hamdy, `libiq` — *a library for arithmetic in class groups of imaginary quadratic orders*, `http://www.math.ucalgary.ca/~hamdy/libiq.html`.

[Ham02] ———, *Über die Sicherheit und Effizienz kryptografischer Verfahren mit Klassengruppen imaginär-quadratischer Zahlkörper*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, Darmstadt, Germany, 2002, `http://www.informatik.tu-darmstadt.de/ftp/pub/TI/reports/hamdy.diss.pdf`.

[HJ85] Roger A. Horn and Charles R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1985.

[HJPT98] Detlef Hühnlein, Michael J. Jacobson, Jr., Sachar Paulus, and Tsuyoshi Takagi, *A cryptosystem based on non-maximal imaginary quadratic orders with fast decryption*, Advances in Cryptology – EUROCRYPT '98 (Kaisa Nyberg, ed.), Lecture Notes in Computer Science, vol. 1403, Springer-Verlag, 1998, pp. 294–307 (English).

[HJW03] Detlef Hühnlein, Michael J. Jacobson, Jr., and Damian Weber, *Towards practical non-interactive public key cryptosystems using non-maximal imaginary quadratic orders*, Designs, Codes and Cryptography **30** (2003), no. 3, 281–299.

[HM89] James L. Hafner and Kevin S. McCurley, *A rigorous subexponential algorithm for computation of class groups*, Journal of the American Mathematical Society **2** (1989), no. 4, 837–850 (English).

[HM91] ———, *Asymptotically fast triangularization of matrices over rings*, SIAM Journal on Computing **20** (1991), 1068–1083.

[HM00a] Safuat Hamdy and Bodo Möller, *Security of cryptosystems based on class groups of imaginary quadratic orders*, Advances in Cryptology – ASIACRYPT 2000 (Tatsuaki Okamoto, ed.), Lecture Notes in Computer Science, vol. 1976, Springer-Verlag, 2000, pp. 234–247.

[HM00b] Detlef Hühnlein and Johannes Merkle, *An efficient NICE-Schnorr-type cryptosystem*, Practice and Theory in Public Key Cryptography, PKC 2000 (Hideki Imai and Yuliang Zheng, eds.), Lecture Notes in Computer Science, vol. 1751, Springer-Verlag, 2000, pp. 14–27.

[HMT98] Detlef Hühnlein, Andreas Meyer, and Tsuyoshi Takagi, *Rabin and RSA analogues based on non-maximal imaginary quadratic orders*, in Rhee and Imai [RI98], pp. 221–240.

[HP01] Detlef Hühnlein and Sachar Paulus, *On the implementation of cryptosystems based on real quadratic fields*, Selected Areas in Cryptography, SAC 2001 (Serge Vaudenay and Amr M. Youssef, eds.), Lecture Notes in Computer Science, vol. 2259, 2001, pp. 288–302.

[HS06] Safuat Hamdy and Filip Saidak, *Arithmetic properties of class numbers of imaginary quadratic fields*, JP Journal of Algebra, Number Theory and Applications **6** (2006), no. 1, 129–148.

[Hua42] Loo-keng Hua, *On the least solution of Pell's equation*, Bulletin of the American Mathematical Society **48** (1942), 731–735.

[Hüh00] Detlef Hühnlein, *Kryptosysteme auf Basis Quadratischer Ordnungen*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, 2000.

[Hüh01a] Detlef Hühnlein, *Efficient implementation of cryptosystems based on non-maximal imaginary quadratic orders*, Selected Areas in Cryptography, SAC 2000 (Douglas R. Stinson and Stafford Tavares, eds.), Lecture Notes in Computer Science, vol. 2012, 2001, pp. 150–167.

[Hüh01b] Detlef Hühnlein, *Faster generation of NICE-schnorr-type signatures*, Topics in Cryptology - CT-RSA 2001 (Berlin) (D. Naccache, ed.), Lecture Notes in Computer Science, vol. 2020, Springer-Verlag, 2001, pp. 1–12.

[HW79] Godfrey H. Hardy and Edward M. Wright, *An introduction to the theory of numbers*, Clarendon Press, Oxford, 1979.

[IR82] Kenneth Ireland and Michael Rosen, *A classical introduction to modern number theory*, Springer-Verlag, New York, 1982.

[Jac98] Michael J. Jacobson, Jr., *Experimental results on class groups of real quadratic fields*, Algorithmic Number Theory, ANTS-III (Joe P. Buhler, ed.), Lecture Notes in Computer Science, vol. 1423, Springer-Verlag, 1998, pp. 463–474.

[Jac99a] ———, *Applying sieving to the computation of quadratic class groups*, Mathematics of Computation **68** (1999), no. 226, 859–867.

[Jac99b] ———, *Subexponential class group computation in quadratic orders*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, Darmstadt, Germany, 1999.

[Jac00] ———, *Computing discrete logarithms in quadratic orders*, Journal of Cryptology **13** (2000), no. 4, 473–492.

[Jac04] ———, *The security of cryptosystems based on class semigroups of imaginary quadratic non-maximal orders*, Australasian Conference on Information Security and Privacy, ACISP 2004 (Vijay Varadharajan Huaxiong Wang, Josef Pieprzyk, ed.), Lecture Notes in Computer Science, vol. 3108, Springer-Verlag, 2004, pp. 149–156.

[JRW06] Michael J. Jacobson, Jr., S. Ramachandran, and H.C. Williams, *Numerical results on class groups of imaginary quadratic fields*, Algorithmic Number Theory, ANTS-VII (Michael Pohst Florian Hess, Sebastian Pauli, ed.), Lecture Notes in Computer Science, vol. 4076, Springer-Verlag, 2006, pp. 87–101.

[JSW01] Michael J. Jacobson, Jr., Renate Scheidler, and Hugh C. Williams, *The efficiency and security of a real quadratic field based-key exchange protocol*, Public-Key Cryptography and Computational Number Theory (Warsaw, Poland), de Gruyter, 2001, pp. 89–112.

[JSW06a] Michael J. Jacobson, Jr., Reginald E. Sawilla, and Hugh C. Williams, *Efficient ideal reduction in quadratic fields.*, International Journal of Mathematics and Computer Science **1** (2006), no. 1, 83–116 (English).

[JSW06b] Michael J. Jacobson, Jr., Renate Scheidler, and Hugh C. Williams, *An improved real quadratic field based key exchange protocol*, Journal of Cryptology **19** (2006), no. 2, 211–239.

[JvdP02] Michael J. Jacobson, Jr. and Alfred J. van der Poorten, *Computational aspects of NUCOMP*, Algorithmic Number Theory, ANTS-V (Claus Fieker and David R. Kohel, eds.), Lecture Notes in Computer Science, vol. 2369, Springer-Verlag, 2002, pp. 120–133.

[Kra22] Maurice Kraitchik, *Theéorie des nombres*, vol. 1, Gauthier-Villars, 1922.

[Lag80a] Jeffrey C. Lagarias, *On the computational complexity of determining the solvability or unsolvability of the equation $X^2 - dY^2 = -1$*, Transactions of the American Mathematical Society **260** (1980), 485–508.

[Lag80b] ———, *Worst-case complexity bounds for algorithms in the theory of integral quadratic forms*, Journal of Algorithms **1** (1980), 142–186.

[Len82] Hendrik W. Lenstra, Jr., *On the calculation of regulators and class numbers of quadratic fields*, Journées Arithmétiques 1980 (Cambridge) (J. V. Armitage, ed.), London Mathematical Society Lecture Note Series, vol. 56, Cambridge University Press, 1982, pp. 123–150.

[LLL82] Arjen K. Lenstra, Hendrik W. Lenstra, Jr., and László Lóvasz, *Factoring polynomials with rational coefficients*, Mathematische Annalen **261** (1982), 515–534.

[LP92] Hendrik W. Lenstra, Jr. and Carl Pomerance, *A rigorous time bound for factoring integers*, J. Amer. Math. Soc. **5** (1992), 483–516.

[Mat61] George B. Mathews, *Theory of numbers*, 2 ed., Chelsea, New York, 1961.

[Mat70] Juri Matijasevič, *Enumerable sets are diophantine*, Soviet Mathematics. Doklady **11** (1970), no. 2, 354–358.

[Mau00] Markus Maurer, *Regulator approximation and fundamental unit computation for real quadratic orders*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, Darmstadt, Germany, 2000.

[MNP01] Andreas Meyer, Stefan Neis, and Thomas Pfahler, *First implementation of cryptographic protocols based on algebraic number fields*, Information Security and Privacy, ACISP 2001, Sydney (Vijay Varadharajan and Yi Mu, eds.), Lecture Notes in Computer Science, vol. 2119, Springer, 2001, pp. 84–103.

[MS99] Thom Mulders and Arne Storjohann, *Diophantine linear system solving*, International Symposium on Symbolic and Algebraic Computation, ISSAC '99 (Sam Dooley, ed.), ACM Press, 1999.

[Nei02] Stefan Neis, *Berechnung von Klassengruppen*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, Darmstadt, Germany, 2002, German.

[NFP] *NFProvider — a toolkit for the Java Cryptography Architecture (JCA/JCE) for Number Field Cryptography*, `http://www.informatik.tu-darmstadt.de/TI/Forschung/FlexiProvider/overview.html#NFProvider`, Part of the FlexiProvider toolkit.

[Pau96] Sachar Paulus, *An algorithm of subexponential type computing the class group of quadratic orders over principal ideal domains*, Algorithmic Number Theory, ANTS-II (Henri Cohen, ed.), Lecture Notes in Computer Science, vol. 1122, Springer-Verlag, 1996, pp. 243–257 (English).

[PS98] Guillaume Poupard and Jacques Stern, *Security analysis of a practical "on the fly" authentication and signature generation*, Advances in Cryptology – EUROCRYPT '98 (Kaisa Nyberg, ed.), Lecture Notes in Computer Science, vol. 1403, Springer-Verlag, 1998, pp. 422–436.

[PT98] Sachar Paulus and Tsuyoshi Takagi, *A generalization of the Diffie-Hellman problem and related cryptosystems allowing fast decryption*, in Rhee and Imai [RI98], pp. 211–220.

[PT00] ———, *A new public-key cryptosystem over a quadratic order with quadratic decryption time*, Journal of Cryptology **13** (2000), no. 2, 263–272. MR 1 748 525.

[PZ85] Michael Pohst and Hans Zassenhaus, *Über die Berechnung von Klassenzahlen und Klassengruppen algebraischer Zahlkörper*, Journal für die Reine und Angewandte Mathematik **361** (1985), 50–72.

[RI98] Man Young Rhee and Hideki Imai (eds.), *The 1st international conference on information security and cryptology, ICISC '98*, DongKwang Publishing Company, Korea, 1998.

[Sch82] René Schoof, *Quadratic fields and factorization*, Computational methods in number theory (Hendrik W. Lenstra, Jr. and Robert Tijdeman, eds.), Mathematical Centre Tracts, vol. 154–155, Mathematisch Centrum, 1982, `http://cr.yp.to/bib/1982/schoof.html`, pp. 235–286.

[Sch85] René Schoof, *Elliptic curves over finite fields and the computation of square roots mod p.*, Mathematics of Computation **44** (1985), 483–494 (English).

[Sch91] Arnold Schönhage, *Fast reduction and composition of binary quadratic forms*, International Symposium on Symbolic and Algebraic Computation, ISSAC '91 (Stephen M. Watt, ed.), ACM Press, 1991, pp. 128–133.

[Sch06] Arthur Schmidt, *Quantum algorithm for solving the discrete logarithm problem in the class group of an imaginary quadratic field and security comparison of current cryptosystems at the beginning of quantum computer age*, Emerging Trends in Information and Communication Security (Günter Müller, ed.), Lecture Notes in Computer Science, vol. 3995, Springer-Verlag, 2006, pp. 481–493.

[Schar] René Schoof, *Computing Arakelov class groups*, Surveys in algorithmic number theory (to appear), `http://www.mat.uniroma2.it/~schoof/infranew2.pdf`.

[Sey87] Martin Seysen, *A probablistic factorization algorithm with quadratic forms of negative discriminant*, Mathematics of Computation **48** (1987), 757–780.

[Sha71] Daniel Shanks, *Class number, a theory of factorization and genera*, 1969 Number Theory Institute (Providence, R.I.), Proceedings of Symposia in Pure Mathematics, vol. 20, AMS, 1971, pp. 415–440.

[Sha72] ———, *The infrastructure of real quadratic fields and its applications*, Proceedings of the 1972 Number Theory Conference, Boulder, Colorado, 1972, pp. 217–224.

[Sha89] ———, *On Gauss and composition I, II*, Number Theory and Applications, Calgary 1988 (Richard A. Mollin, ed.), NATO ASI Series, Series C, vol. 265, Kluwer Academic Publishers, 1989, pp. 163–178, 179–204.

[Sho97] Peter W. Shor, *Polynomial time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM Journal on Computing **26** (1997), no. 5, 1484–1509.

[Sto00] Arne Storjohann, *Algorithms for matrix canonical forms*, Ph.D. thesis, ETH Zürich, 2000.

[SV05] Arthur Schmidt and Ulrich Vollmer, *Polynomial time quantum algorithm for the computation of the unit group of a number field*, Proceedings of the 37th annual ACM Symposium on Theory of Computing (Harold N. Gabow and Ronald Fagin, eds.), ACM Press, 2005, pp. 475–480.

[Ter00] David C. Terr, *A modification of Shanks' baby-step giant-step algorithm*, Mathematics of Computation **69** (2000), no. 230, 767–773.

[Tes98a] Edlyn Teske, *New algorithms for finite abelian groups*, Ph.D. thesis, Technische Universität Darmstadt, Germany, 1998, Shaker Verlag, Aachen.

[Tes98b] ———, *A space efficient algorithm for group structure computation*, Math. Comput. **67** (1998), no. 224, 1637–1663 (English).

[Tes99] ———, *The Pohlig-Hellman method generalized for group structure computation*, J. Symbolic Comput. **27** (1999), no. 6, 521–534. MR 2000f:20090

[The00] Patrick Theobald, *Berechnung von Hermite-Normalformen*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, Darmstadt, Germany, 2000, German.

[Thi94] Christoph Thiel, *Under the assumption of the generalized riemann hypothesis verifying the class number belongs to NP ∩ co-NP*, Algorithmic number theory, ANTS-I (Leonard M. Adleman and Ming-Deh Huang, eds.), Lecture Notes in Computer Science, vol. 877, Springer-Verlag, 1994, pp. 234–247 (English).

[Thi95] Christoph Thiel, *On the complexity of some problems in algorithmic algebraic number theory*, Ph.D. thesis, Universität des Saarlandes, Saarbrücken, Germany, 1995.

[vdP03] Alfred van der Poorten, *A note on NUCOMP*, Mathematics of Computation **72** (2003), no. 244, 1935–1946 (electronic).

[vG99] Joachim von zur Gathen and Jürgen Gerhard, *Modern computer algebra*, Cambridge University Press, Cambridge, UK, 1999.

[Vol00] Ulrich Vollmer, *Asymptotically fast discrete logarithms in quadratic number fields*, Algorithmic Number Theory, ANTS-IV (Wieb Bosma, ed.), Lecture Notes in Computer Science, vol. 1838, Springer-Verlag, 2000, pp. 581–594.

[Vol02] ———, *An accelerated Buchmann algorithm for regulator computation in real quadratic fields*, Algorithmic Number Theory, ANTS-V (Claus Fieker and David R. Kohel, eds.), Lecture Notes in Computer Science, vol. 2369, Springer-Verlag, 2002, pp. 148–162.

[Vol03a] ———, *A note on the Hermite basis computation of large integer matrices*, International Symposium on Symbolic and Algebraic Computation, ISSAC ’03 (J. Rafael Sendra, ed.), ACM Press, 2003, pp. 255–257.

[Vol03b] ———, *Rigorously analyzed algorithms for the discrete logarithm problem in quadratic number fields*, Ph.D. thesis, Technische Universität Darmstadt, Fachbereich Informatik, 2003.

[WM68] A. E. Western and J. C. P. Miller, *Tables of indices and primitive roots*, Royal Society Mathematical Tables, Vol. 9, Published for the Royal Society at the Cambridge University Press, London, 1968. MR 39 #7792.

[Wie86] Douglas H. Wiedemann, *Solving sparse linear equations over finite fields*, IEEE Trans. Inf. Theory IT-32 (1986), 54–62.

Index