BOOK REVIEWS

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 35, Number 4, October 1998, Pages 333–338
S 0273-0979(98)00758-7

Quadratics, by Richard A. Mollin, CRC Press, Boca Raton, FL, 1996, xx+387 pp., $74.95, ISBN 0-8493-3983-9

A *quadratic order* is a ring $\mathcal{O} = \mathbf{Z}[\alpha]$ generated over the ring $\mathbf{Z}$ of rational integers by a root α of some irreducible quadratic polynomial $X^2 + bX + c \in \mathbf{Z}[X]$. We have $\mathcal{O} = \mathbf{Z}[(D + \sqrt{D})/2]$, with $D = b^2 - 4c$ the *discriminant* of $\mathcal{O}$. The order $\mathcal{O}$ is determined up to ring isomorphism by its discriminant, which may be any non-square integer congruent to 0 or 1 modulo 4. Orders of positive discriminant are called *real quadratic*, orders of negative discriminant *imaginary quadratic*.

In the 18th century, Euler used quadratic irrationalities in order to find integral solutions to various Diophantine equations. Using the order $\mathbf{Z}[(-1 + \sqrt{-3})/2]$ of discriminant -3 generated by the third roots of unity, he found that the only integral solutions to the Fermat equation $x^3 + y^3 = z^3$ are the trivial solutions having $xyz = 0$. Similarly, he interpreted a 17th-century English method for finding positive integral solutions to the equation $x^2 = dy^2 + 1$ for given non-square integer d, which he incorrectly attributed to John Pell, in terms of the *continued fraction expansion* of the quadratic irrationality $\sqrt{d}$. The first step in the argument consists of rewriting the equation as $(x + y\sqrt{d})(x - y\sqrt{d}) = 1$. In modern terms, finding the solutions to the equation essentially amounts to the determination of the unit group of the order $\mathbf{Z}[\sqrt{d}]$.

The first systematic investigation of what we have come to regard as the arithmetic properties of quadratic orders was undertaken by Gauss in the *Disquisitiones Arithmeticae* (1801). As Gauss held the opinion that quadratic irrationalities should not be used in a theory dealing primarily with integers, he avoided any explicit reference to orders and cast his theory exclusively in terms of *binary quadratic forms*. Such forms look like $aX^2 + bXY + cY^2$ for integers a, b and c; if we have $\gcd(a, b, c) = 1$, then the form is said to be *primitive*. The group $\mathrm{SL}_2(\mathbf{Z})$ of integral unimodular matrices has a natural right action on the set of forms given by

$$(aX^2 + bXY + cY^2)^M = a(pX + qY)^2 + b(pX + qY)(rX + sY) + c(rX + sY)^2$$

for $M = \left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$. Unimodular transformations leave the *discriminant* $D = b^2 - 4ac$ of the form $aX^2 + bXY + cY^2$ invariant, and forms in the same orbit assume the same values on $\mathbf{Z}^2$. By a process called *composition* of forms, Gauss showed how to obtain from two primitive forms of discriminant D a third primitive form of discriminant D. In modern terminology, he went on to prove that this composition makes the set $\mathcal{C}_D$ of $\mathrm{SL}_2(\mathbf{Z})$-orbits of primitive forms of given discriminant D into

1991 *Mathematics Subject Classification.* Primary 11R11.

a finite abelian group, the *class group* of discriminant D. Its order is the *class number* $h(D)$ of D. Gauss gave several applications of his theory, such as a proof of the quadratic reciprocity law and a proof of Legendre's theorem giving necessary and sufficient conditions for the ternary quadratic equation $ax^2 + by^2 + cz^2 = 0$ to have non-trivial solutions.

Despite the obvious strength of Gauss's theory, his definition of composition is somewhat involved, and the proof that it gives rise to a group structure on $\mathcal{C}_D$ involves lengthy calculations. During the 19th century, it became clear through the work of Dirichlet, Kummer and Dedekind that Gauss's theory fits into the more general framework of algebraic number theory. In this theory, one associates to every finite field extension K of $\mathbf{Q}$, usually called a *number field*, a ring of integers $\mathcal{O}_K$ of K. This ring is free as an abelian group of rank equal to the degree of K over $\mathbf{Q}$. It may fail to be a unique factorization domain like $\mathbf{Z}$ itself, but it does admit *unique ideal factorization*. The class group $\mathcal{C}(\mathcal{O}_K)$ of $\mathcal{O}_K$, which is the quotient of the group of invertible $\mathcal{O}_K$-ideals by the subgroup of principal ideals, is a finite abelian group measuring the extent to which unique element factorization fails in $\mathcal{O}_K$. Dirichlet showed that the unit group $\mathcal{O}_K^*$ of $\mathcal{O}_K$ is a finitely generated abelian group and derived *analytic class number formulas* for the order of $\mathcal{C}(\mathcal{O}_K)$, cf. [6]. For a subring $\mathcal{O} \subset \mathcal{O}_K$ of finite index, the class group $\mathcal{C}(\mathcal{O})$ is also finite, and $\mathcal{O}^*$ is a subgroup of finite index in $\mathcal{O}_K^*$.

For quadratic fields K, the ring $\mathcal{O}_K$ is the maximal quadratic order contained in K. Every quadratic order is a subring of finite index in the ring of integers of its field of fractions. If $\mathcal{O}$ is a quadratic order of discriminant $D < 0$, the class group $\mathcal{C}(\mathcal{O})$ defined in terms of ideals coincides with the class group $\mathcal{C}_D$ as defined by Gauss. More explicitly, the class of the quadratic form $aX^2 + bXY + cY^2$ of discriminant D in $\mathcal{C}_D$ corresponds to the class of the ideal with $\mathbf{Z}$-basis $[2a, -b+\sqrt{D}]$ in $\mathcal{C}(\mathcal{O})$. For real quadratic orders, the same is true if a certain sign-condition is taken into account. Each of the two interpretations of quadratic class groups has its merits: the ideal interpretation usually yields smooth conceptual proofs, whereas the explicit nature of the composition of forms makes them a convenient vehicle for computations. The simple form of the dictionary between the languages of ideals and forms shows that the translation of statements in either direction is entirely straightforward, and that there is no intrinsic mathematical distinction between them.

A third interpretation of quadratic class groups is of a more recent nature. It is a consequence of *class field theory*, which was discovered during the late 19th century by Kronecker, Weber and Hilbert and had become firmly established by 1925. This theory shows that the class group of an order $\mathcal{O}$ in a quadratic number field K can be viewed as the Galois group $\mathrm{Gal}(R/K)$ of the ring class field $R = R(\mathcal{O})$ of $\mathcal{O}$ over K. The ring class field R corresponding to $\mathcal{O}$ is the maximal abelian extension of K satisfying certain ramification restrictions related to $\mathcal{O}$. Again, this new interpretation is useful in both directions: one may derive results on the class group of $\mathcal{O}$ by constructing suitable field extensions of K inside $R(\mathcal{O})$, and one obtains field extensions of K with specific arithmetic properties from the knowledge of the class group of $\mathcal{O}$. For an attractive description of the interplay between the three descriptions, the interested reader should consult the book by Cox [2].

Quadratic class groups are the simplest examples of the class groups that occur in algebraic number theory, and they are very accessible from a computational point

of view. Nevertheless, many of the basic questions one can ask about them are still unanswered.

Imaginary quadratic orders are in many ways better understood than real quadratic orders. For negative D, every class of forms in $\mathcal{C}_D$ contains a unique *reduced* binary quadratic form, and there is an efficient reduction algorithm to transform a form into an $\mathrm{SL}_2(\mathbf{Z})$-equivalent reduced form. This feature makes it easy to perform computations in the class group. In addition, the ring class fields corresponding to imaginary quadratic orders can be generated explicitly by values of a modular function, the j-function.

The class number $h(D)$ behaves somewhat irregularly for negative D, but the general tendency for $h(D)$ to grow roughly as $\sqrt{-D}$, for D tending to $-\infty$, was already noticed by Gauss. It is not hard to find an upper bound $O(|D|^{1/2+\epsilon})$ for $h(D)$ for every $\epsilon > 0$. In 1935, Siegel showed by complex analytic methods that there is a lower bound $h(D) > C_\epsilon |D|^{1/2-\epsilon}$ for every $\epsilon > 0$. The constant $C_\epsilon > 0$ is however not *effective*: what the proof shows is that there cannot be *two* large negative values of D violating an inequality of this type. As a consequence, Siegel's theorem cannot be used to show that the list of 13 negative discriminants of class number 1 known to Gauss is complete. However, the method does imply that there is at most one discriminant of class number 1 that is not in the list. Negative discriminants of class number 1 possess remarkable properties. For instance, the value $h(-163) = 1$ implies that Euler's polynomial $x^2 - x + 41$ of discriminant -163 assumes only prime values on the integers in the interval $[-39, 40]$, and it explains Hermite's observation that $e^{\pi\sqrt{163}}$ is less than 10^{-12} away from an integer. The effectiveness problems encountered in Siegel's result and its generalizations are related to our lack of knowledge of the location of zeroes of zeta-functions, and usually there are stronger results if one is willing to assume certain generalized Riemann hypotheses. The completeness of the class number 1 list was not established unconditionally until 1967, when Baker and Stark independently proved this result, and Stark showed in addition how to fill the 'gap' in a 1952 proof of Heegner. Their methods, and the relation between the class number 1 problem and the determination of integral points on modular curves, are discussed in the appendix of [5]. These results have now been superseded by the work of Goldfeld, Gross and Zagier. As explained in Oesterlé's 1984 Bourbaki talk [4], one proceeds by constructing a modular elliptic curve E over $\mathbf{Q}$ for which the associated L-function has a zero of order ≥ 3 in $s = 1$. From such an L-function, one obtains an effective lower bound of the form $h(D) > C_E \log(-D)$. Such inequalities have been used in recent years to obtain complete lists of discriminants $D < 0$ of given small class numbers.

The behavior of class numbers of real quadratic orders turns out to be more complicated. The difficulties are related to the unit group $\mathcal{O}^*$ of the order. This group equals $\{\pm 1\}$ for all $D < -4$, but for positive D it is generated by $\{\pm 1\}$ and a *fundamental unit* ε_D of infinite order. It has been known for a long time that even for small D, the fundamental unit can be large. In a 1657 challenge problem to the English mathematicians proposed by Fermat, one is asked to find the smallest positive integer x for which $433x^2 + 1$ is a square. The solution $x =$ 5,025,068,784,834,899,736 makes clear that the order $\mathbf{Z}[\sqrt{433}]$ does not contain any 'obvious' units besides ± 1. Euler's formalization of the English method to produce such solutions is known as the *continued fraction algorithm*. Using the embedding of a real quadratic order $\mathcal{O}$ as a lattice in $\mathbf{R}^2$ via the two ring homomorphisms

$\mathcal{O} \to \mathbf{R}$, one can give a geometric description of the algorithm in terms of walks along lattice points in the plane.

For positive D, the object that exhibits a regular growth as a function of D is not the class group $\mathcal{C}_D$ itself, of size $h(D)$, but the *Arakelov class group* $\mathrm{Pic}(D)$, which is a compact topological group of 'size' $h(D)\cdot|\log|\varepsilon_D||$. Reduced quadratic forms may be viewed as elements of $\mathrm{Pic}(D)$. As the fibres of the natural surjection $\mathrm{Pic}(D) \to \mathcal{C}_D$ are the cosets of a circle group $\mathbf{R}/\log|\varepsilon_D|\mathbf{Z} \subset \mathrm{Pic}(D)$, the reduced forms can be grouped in *cycles*. For forms lying in the same cycle, their difference in $\mathrm{Pic}(D)$ lies in $\mathbf{R}/\log|\varepsilon_D|\mathbf{Z}$ and may be viewed as their relative *distance*. This distance function was discovered empirically in the seventies by the late Daniel Shanks, who showed that explicit computations with forms in what we now call $\mathrm{Pic}(D)$ can be performed efficiently on a computer. He coined the term *infrastructure* to describe the group structure of $\mathrm{Pic}(D)$, and used it to develop various practical algorithms to compute $h(D)$ and $\log|\varepsilon_D|$. A group-theoretical basis was provided by H. W. Lenstra [3]. In this setting, the continued fraction algorithm may be viewed as a method to 'walk down' the forms in a cycle in $\mathrm{Pic}(D)$ by repeatedly performing a 'reduction step'. Shanks's algorithms, which include a *baby-step giant-step* trick to speed up such walks, can be found in [1].

An analog of Siegel's theorem mentioned above shows that $h(D)\cdot|\log|\varepsilon_D||$ roughly grows as $\sqrt{D}$ for D tending to infinity, but the asymptotic behavior of the separate factors $h(D)$ and $\log|\varepsilon_D|$ is an open problem. It is easy to write down families of discriminants D for which $\log|\varepsilon_D|$ is small, and for such families Siegel's methods yield solutions to the class number 1 problem under assumption of the generalized Riemann hypothesis. On the other hand, it seems that ε_D is large in most cases, and this would imply that $h(D)$ is often small. There are exact predictions for the average behavior of $h(D)$ for $D \to \infty$ known as the *Cohen-Lenstra heuristics*. A proof of these heuristics seems to be beyond all current methods. Even though $h(D)$ should be equal to 1 for a proportion $P \approx .75446$ of the prime discriminants D, no one has been able to show that there are infinitely many D of class number 1. An additional complication in comparison with the imaginary case is the absence of a real analog of the theory of modular functions, which might be used to generate the ring class fields.

Any author writing a textbook on quadratic orders must make a selection from the numerous topics that might be included in such a book. Inclusion of the class field theoretic aspects, as in [2], necessitates knowledge of a fair amount of algebraic number theory on the part of the reader. Inclusion of modern methods to derive effective lower bounds on class numbers requires an exposition of material not directly related to quadratic orders, such as L-series of elliptic curves. Inclusion of the Arakelov point of view of the cycle structure of real quadratic class groups implies inclusion of some abstract algebra, and possibly some topology as well. Inclusion of a geometric description of the continued fraction algorithm means inclusion of pictures of plane lattices and arguments from the geometry of numbers.

Mollin's book includes none of the above. His motivation for writing the book, he states, is a 'search for truth and beauty.' Along the lines that we have sketched in this review, he distinguishes a 'form theory', an 'ideal theoretic approach' and a 'class field theory approach' to his subject. Forms and ideals provide to most people equivalent ways to view one and the same object, and for them the multiplication formulas for ideals on page 10 of the book are also composition formulas for

binary quadratic forms. Mollin holds a surprisingly different opinion. He adheres painstakingly to an exclusively ideal theoretic presentation, and finds the literature 'seriously deficient in terms of ideal-theoretic proofs of keystone concepts.' Quadratic forms, which occur implicitly everywhere throughout the book, are mentioned explicitly only in an Appendix E, a 'Gazetteer of Forms'. Here we learn that for the presentation so far, 'an end goal is to have subsumed the older material by updating concepts and introducing notation and meaning which are consistent with modern algebra and number theory, while, at the same time, eliminating confusing conflicts.' The Gazetteer tells us that the precise correspondence between forms and ideals is beyond the scope of the book since it 'involves certain identifications via ideals with norms prime to the conductor' and requires class field theory. Already for historical reasons, this cannot be true, and the simple formulas in [2] for moving back and forth between forms and ideals show that it isn't. If there is confusion, it is not in the mathematics. It is a fact that the identification with forms leads to the important distinction between ordinary and narrow ideal class groups in the real quadratic case, and the book chooses to work with the ordinary class group. As the narrow ideal class group often leads to the simpler theorems, there seems to be no reason to discuss this distinction only in an appendix.

The book focuses on the infrastructure and the continued fraction approach which 'have not received the respect and currency which they so richly deserve.' It may be true that the original presentation of the infrastructure did not immediately make clear whether this structure is part of the beauty of pure mathematics. However, the concept nowadays fits neatly in the Arakelov point of view, and its algorithmic qualities have secured its inclusion in the standard textbook [1] on computational number theory. Furthermore, the infrastructure yields an excellent opportunity to illustrate the use of modern algebra, which was notoriously lacking in the original presentation. Sadly enough, the discussion of the infrastructure in Chapter 2 of Mollin's book is actually a description of the continued fraction algorithm, and the fundamental distance function, which becomes a group homomorphism when it is defined appropriately, is hidden in a poorly formulated exercise. The distance function reemerges only in Chapter 8, in a description of Shanks's algorithm to find real quadratic class numbers. The reader is however referred to the literature for complete proofs with respect to this basic application. In addition, he is told about a classical theorem of Lévy (theorem 8.1.1), which should imply that reduction steps along a cycle, or equivalently continued fraction steps for a quadratic irrational number α, are of length close to $\pi^2/(12\log 2)$ for 'all but finitely many α.' Unfortunately, Lévy's theorem is a result about almost all real numbers α in the sense of the Lebesgue measure, so it implies nothing for quadratic irrationals.

Despite the book's aim to use modern algebra and number theory, all proofs proceed by manipulations of formulas and explicit calculations. The proof of the first structural result in the book, a result of Gauss on the number of ambiguous ideal classes in $\mathcal{C}_D$ on page 16, is a characteristic example. It uses the explicit composition formulas and invokes a couple of exercises to deal with various special cases. Moreover, the underlying assumption that an ambiguous ideal is a product of ambiguous prime ideals is completely unjustified. Given the author's aims, it is curious that structural concepts such as subgroups, generators and homomorphisms, which are characteristic of modern algebra and very useful in the context of quadratic orders, hardly occur in his book. Is it an excuse that Gauss, whose

death in 1855, we read on page 3, 'left a void in mathematics which remains to be filled,' could have proved most of the results in this book?

Each reader may form his own opinion on Mollin's idiosyncratic style. A special flavor is added by the numerous footnotes that point out minor errors to be found in the literature, comment on the history, provide explanations on the side, supply various anecdotes and contain biographical material on individuals ranging from Euler to Amerigo Vespucci.

The limited scope of the topics included in the book and the amount of synthesis achieved in their presentation make Mollin's book unsuitable as a general introduction to the subject of quadratic orders. It will however be a useful text for those who want to acquaint themselves with Mollin's research. There is a strong emphasis on class number results for discriminants of a special form. For instance, for an odd discriminant $D = b^2 - 4m^t < 0$ one can often show that the 'obvious' ideal of index m yields a class of order t in $\mathcal{C}_D$. Similarly, if one looks at the square-free odd discriminants $D = n^2 + 1$, the unit $n + \sqrt{D}$ in the corresponding order is so small that $h(D)$ tends to infinity with n by Siegel's theorem. In such cases, one can find the D with $h(D) = 1$ as in the original class number 1 problem for $D < 0$ if one either assumes suitable Riemann hypotheses or allows the existence of a large unknown exceptional value of D having $h(D) = 1$. Other topics in the book are generalizations of Euler's prime producing polynomial $x^2 - x + 41$, algorithms for computing real quadratic class numbers and applications in cryptography. The final section contains philosophical insights provided by Andrew Lazarus.

References

1. H. Cohen, *A Course in Computational Algebraic Number Theory*, Springer Graduate Texts in Mathematics 138, 1993. MR **94i:**11105
2. D. A. Cox, *Primes of the form $x^2 + ny^2$*, Wiley-Interscience, 1989. MR **90m:**11016
3. H. W. Lenstra, *On the calculation of regulators and class numbers of quadratic fields*, Journées Arithmétiques 1980 (J. Armitage, ed.), London Math. Soc. Lecture Note Ser., vol. 56, Cambridge University Press, 1982. MR **86g:**11080
4. J. Oesterlé, *Nombres de classes des corps quadratiques imaginaires*, Sém. Bourbaki 1983-84, exposé 631, Astérisque **121–122** (1985). MR **86k:**11064
5. J.-P. Serre, *Lectures on the Mordell-Weil theorem*, Vieweg, 1989. MR **90e:**11086
6. D. B. Zagier, *Zetafunktionen and quadratische Körper*, Springer, 1981. MR **82m:**10002

Peter Stevenhagen
Universiteit van Amsterdam
E-mail address: psh@wins.uva.nl