CHAPTER I

Transition to Modern Number Theory

Abstract. This chapter establishes Gauss's Law of Quadratic Reciprocity, the theory of binary quadratic forms, and Dirichlet's Theorem on primes in arithmetic progressions.

Section 1 outlines how the three topics of the chapter occurred in natural sequence and marked a transition as the subject of number theory developed a coherence and moved toward the kind of algebraic number theory that is studied today.

Section 2 establishes quadratic reciprocity, which is a reduction formula providing a rapid method for deciding solvability of congruences $x^2 \equiv m \bmod p$ for the unknown x when p is prime.

Sections 3–5 develop the theory of binary quadratic forms $ax^2 + bxy + cy^2$, where a, b, c are integers. The basic tool is that of proper equivalence of two such forms, which occurs when the two forms are related by an invertible linear substitution with integer coefficients and determinant 1. The theorems establish the finiteness of the number of proper equivalence classes for given discriminant, conditions for the representability of primes by forms of a given discriminant, canonical representatives of the finitely many proper equivalence classes of a given discriminant, a group law for proper equivalence classes of forms of the same discriminant that respects representability of integers by the classes, and a theory of genera that takes into account inequivalent forms whose values cannot be distinguished by linear congruences.

Sections 6–7 digress to leap forward historically and interpret the group law for proper equivalence classes of binary quadratic forms in terms of an equivalence relation on the nonzero ideals in the ring of integers of an associated quadratic number field.

Sections 8–10 concern Dirichlet's Theorem on primes in arithmetic progressions. Section 8 discusses Euler's product formula for $\sum_{n=1}^{\infty} n^{-s}$ and shows how Euler was able to modify it to prove that there are infinitely many primes $4k + 1$ and infinitely many primes $4k + 3$. Section 9 develops Dirichlet series as a tool to be used in the generalization, and Section 10 contains the proof of Dirichlet's Theorem. Section 8 uses some elementary real analysis, and Sections 9–10 use both elementary real analysis and elementary complex analysis.

1. Historical Background

The period 1800 to 1840 saw great advances in number theory as the subject developed a coherence and moved toward the kind of algebraic number theory that is studied today. The groundwork had been laid chiefly by Euclid, Diophantus, Fermat, Euler, Lagrange, and Legendre. Some of what those people did was remarkably insightful for its time, but what collectively had come out of their labors was more a collection of miscellaneous results than an organized theory. It was Gauss who first gave direction and depth to the subject, beginning with

his book *Disquisitiones Arithmeticae* in 1801. Dirichlet built on Gauss's work, clarifying the deeper parts and adding analytic techniques that pointed toward the integrated subject of the future. This chapter concentrates on three jewels of classical number theory—largely the work of Gauss and Dirichlet—that seem on the surface to be only peripherally related but are actually a natural succession of developments leading from earlier results toward modern algebraic number theory. To understand the context, it is necessary to back up for a moment.

Diophantine equations in two or more variables have always lain at the heart of number theory. Fundamental examples that have played an important role in the development of the subject are $ax^2+bxy+cy^2 = m$ for unknown integers x and y; $x_1^2+x_2^2+x_3^2+x_4^2 = m$ for unknown integers x_1, x_2, x_3, x_4; $y^2 = x(x-1)(x+1)$ for unknown integers x and y; and $x^n + y^n = z^n$ for unknown integers x, y, z.

In every case one can get an immediate necessary condition on a solution by writing the equation modulo some integer n. The necessary condition is that the corresponding congruence modulo n have a solution. For example take the equation $x^2 + y^2 = p$, where p is a prime, and let us allow ourselves to use the more elementary results of *Basic Algebra*. Writing the equation modulo p leads to $x^2 + y^2 \equiv 0 \bmod p$. Certainly x cannot be divisible by p, since otherwise y would be divisible by p, x^2 and y^2 would be divisible by p^2, and $x^2 + y^2 = p$ would be divisible by p^2, contradiction. Thus we can divide, obtaining $1 + (yx^{-1})^2 \equiv 0 \bmod p$. Hence $z^2 \equiv -1 \bmod p$ for $z \equiv xy^{-1}$. If p is an odd prime, then -1 has order 2, and the necessary condition is that there exist some z in $\mathbb{F}_p^\times$ whose order is exactly 4. Since $\mathbb{F}_p^\times$ is cyclic of order $p-1$, the necessary condition is that 4 divide $p-1$.

Using a slightly more complicated argument, we can establish conversely that the divisibility of $p-1$ by 4 implies that $x^2 + y^2 = p$ is solvable for integers x and y. In fact, we know from the solvability of $z^2 \equiv -1 \bmod p$ that there exists an integer r such that p divides $r^2 + 1$. Consider the possibilities in the integral domain $\mathbb{Z}[i]$ of Gaussian integers, where $i = \sqrt{-1}$. It was shown in Chapter VIII of *Basic Algebra* that $\mathbb{Z}[i]$ is Euclidean. Hence $\mathbb{Z}[i]$ is a principal ideal domain, and its elements have unique factorization. If p remains prime in $\mathbb{Z}[i]$, then the fact that p divides $(r+i)(r-i)$ implies that p divides $r+i$ or $r-i$ in $\mathbb{Z}[i]$. Then at least one of $\frac{r}{p} + i\frac{1}{p}$ and $\frac{r}{p} - i\frac{1}{p}$ would have to be in $\mathbb{Z}[i]$. Since $i\frac{1}{p}$ is not in $\mathbb{Z}[i]$, this divisibility does not hold, and we conclude that p does not remain prime in $\mathbb{Z}[i]$. If we write $p = (a+bi)(c+di)$ nontrivially, then $p^2 = |a+bi|^2|c+di|^2 = (a^2+b^2)(c^2+d^2)$ as an equality in $\mathbb{Z}$, and we readily conclude that $a^2 + b^2 = p$.

This much argument solves the Diophantine equation $x^2+y^2 = p$ for p prime. For p replaced by a general integer m, we use the identity

$$(x_1^2 + y_1^2)(x_2^2 + y_2^2) = (x_1x_2 - y_1y_2)^2 + (x_1y_2 + x_2y_1)^2,$$

which has been known since antiquity, and we see that $x^2 + y^2 = m$ is solvable if m is a product of odd primes of the form $4k + 1$. It is solvable also if $m = 2$ and if $m = p^2$ for any prime p. Thus $x^2 + y^2 = m$ is solvable whenever m is a positive integer such that each prime of the form $4k + 3$ dividing m divides m an even number of times. Using congruences modulo prime powers, we see that this condition is also necessary, and we arrive at the following result; historically it had already been asserted as a theorem by Fermat and was subsequently proved by Euler, albeit by more classical methods than we have used.

Proposition 1.1. The Diophantine equation $x^2+y^2 = m$ is solvable in integers x and y for a given positive integer m if and only if every prime number $p = 4k+3$ dividing m occurs an even number of times in the prime factorization of m.

The first step in the above argument used congruence information; we had to know the primes p for which $z^2 \equiv -1 \bmod p$ is solvable. The second step was in two parts—both rather special. First we used specific information about the nature of factorization in a particular ring of algebraic integers, namely $\mathbb{Z}[i]$. Second we used that the norm of a product is the product of the norms in that same ring of algebraic integers.

It is too much to hope that some recognizable generalization of these steps with $x^2 + y^2 = m$ can handle all or most Diophantine equations. At least the first step is available in complete generality, and indeed number theory—both classical and modern—deduces many helpful conclusions by passing to congruences. There is the matter of deducing something useful from a given congruence, but doing so is a finite problem for each prime. Like some others before him, Gauss set about studying congruences systematically. Linear congruences are easy and had been handled before. Quadratic congruences are logically the next step. The first jewel of classical number theory to be discussed in this chapter is the Law of Quadratic Reciprocity of Gauss, which appears below as Theorem 1.2 and which makes useful deductions possible in the case of quadratic congruences. In effect quadratic reciprocity allows one to decide easily which integers are squares modulo a prime p. Euler had earlier come close to finding the statement of this result, and Legendre had found the exact statement without finding a complete proof. Gauss was the one who gave the first complete proof.

Part of the utility of quadratic reciprocity is that it helps one to attack quadratic Diophantine equations more systematically. The second jewel of classical number theory to be discussed in this chapter is the body of results concerning representing integers by binary quadratic forms $ax^2+bxy+cy^2 = m$ that do not degenerate in some way. Lagrange and Legendre had already made advances in this theory, but Gauss's own discoveries were decisive. Dirichlet simplified the more advanced parts of the theory and investigated an aspect of it that Gauss had not addressed

and that would lead Dirichlet to his celebrated theorem on primes in arithmetic progressions.[1]

Lagrange had introduced the notion of the discriminant of a quadratic form and a notion of equivalence of such forms—two forms of the same discriminant being equivalent if one can be obtained from the other by a linear invertible substitution with integer entries. Equivalence is important because equivalent forms represent the same numbers. He established also a theory of reduced forms that specifies representatives of each equivalence class. For an odd prime p, $ax^2 + bxy + cy^2 = p$ is solvable only if the discriminant $b^2 - 4ac$ is a square modulo p, and Lagrange was hampered by not knowing quadratic reciprocity. But he did know some special cases, such as when 5 is a square modulo p, and he was able to deal completely with discriminant -20. For this discriminant, there are two equivalence classes, represented by $x^2 + 5y^2$ and $2x^2 + 2xy + 3y^2$, and Lagrange showed for primes p other than 2 and 5 that

$$\begin{aligned} x^2 + 5y^2 = p \qquad &\text{is solvable if and only if} \qquad p \equiv 1 \text{ or } 9 \bmod 20, \\ 2x^2 + 2xy + 3y^2 = p \qquad &\text{is solvable if and only if} \qquad p \equiv 3 \text{ or } 7 \bmod 20; \end{aligned}$$

the fact about $x^2 + 5y^2 = p$ had been conjectured earlier by Euler. Lagrange observed further that

$$\begin{aligned} &(2x_1^2 + 2x_1y_1 + 3y_1^2)(2x_2^2 + 2x_2y_2 + 3y_2^2) \\ &\qquad = (2x_1x_2 + x_1y_2 + y_1x_2 + 3y_1y_2)^2 + 5(x_1x_2 - y_1y_2)^2, \end{aligned}$$

from which it follows that the product of two primes congruent to 3 or 7 modulo 20 is representable as $x^2 + 5y^2$; this fact had been conjectured by Fermat.

Legendre added to this investigation the correct formula for quadratic reciprocity, which he incorrectly believed he had proved, and many of its consequences for representability of primes by binary quadratic forms. In addition, he tried to develop a theory of composition of forms that generalizes Lagrange's identity above, but he had only limited success.

In addition to establishing quadratic reciprocity, Gauss introduced the vital notion of "proper equivalence" for forms $ax^2+bxy+cy^2$ of the same discriminant—two forms of the same discriminant being properly equivalent if one can be obtained from the other by a linear invertible substitution with integer entries and determinant $+1$. In terms of this definition, he settled the representability of primes by binary quadratic forms, he showed that there are only finitely many proper equivalences classes for each discriminant, and he gave an algorithm for

[1] These matters are affirmed in Dirichlet's *Lectures on Number Theory*. The aspect that Gauss had not addressed and that provided motivation for Dirichlet is the value of the "Dirichlet class number" $h(D)$ defined below.

deciding whether two forms are properly equivalent. The main results of Gauss in this direction appear as Theorems 1.6 and 1.8 below. In addition, Gauss showed, without the benefit of having a definition of "group," in effect that the set of proper equivalence classes of forms with a given discriminant becomes a finite abelian group in a way that controls representability of nonprime integers; by contrast, Lagrange's definition of equivalence does not lead to a group structure. Gauss's main results in this direction, as recast by Dirichlet, appear as Theorem 1.12 below.

The story does not stop here, but let us pause for a moment to say what Lagrange's theory, as amended by Gauss, says for the above example, first rephrasing the context in more modern terminology. We saw earlier that unique factorization in the ring $\mathbb{Z}[i]$ of Gaussian integers is the key to the representation of integers by the quadratic form $x^2 + y^2$. For a general quadratic form $ax^2 + bxy + cy^2$ with discriminant $D = b^2 - 4ac$, properties of the ring R of algebraic integers in the field $\mathbb{Q}(\sqrt{D})$ are relevant for the questions that Gauss investigated. It turns out that R is a principal ideal domain if Gauss's finite abelian group of proper equivalence classes is trivial and that when D is "fundamental," there is a suitable converse.[2]

With the context rephrased we come back to the example. Consider the equation $x^2+5y^2 = p$ for primes p. The discriminant of x^2+5y^2 is -20, and the relevant ring of algebraic integers is $\mathbb{Z}[\sqrt{-5}]$, which is not a unique factorization domain. Thus the argument used with $x^2 + y^2 = p$ does not apply, and we have no reason to expect that solvability of $x^2 + 5y^2 \equiv 0 \bmod p$ is sufficient for solvability of $x^2+5y^2 = p$. Let us look more closely. The congruence condition is that -20 is a square modulo p. Thus -5 is to be a square modulo p. If we leave aside the primes $p = 2$ and $p = 5$ that divide 20, the Law of Quadratic Reciprocity will tell us that the necessary congruence resulting from solvability of $x^2 + 5y^2 = p$ is that p be congruent to 1, 3, 7, or 9 modulo 20. However, we can compute all residues n of $x^2 + 5y^2$ modulo 20 for n with $\mathrm{GCD}(n, 20) = 1$ to see that

$$x^2 + 5y^2 \equiv 1 \text{ or } 9 \bmod 20 \qquad \text{if } \mathrm{GCD}(x^2 + 5y^2,\ 20) = 1.$$

Meanwhile, the form $2x^2 + 2xy + 3y^2$ has discriminant -20, and we can check that solvability of $2x^2 + 2xy + 3y^2 = p$ leads to the conclusion that

$$2x^2 + 2xy + 3y^2 \equiv 3 \text{ or } 7 \bmod 20 \qquad \text{if } \mathrm{GCD}(2x^2 + 2xy + 3y^2,\ 20) = 1.$$

Lagrange's theory easily shows that representability of integers by a form depends only on the equivalence class of the form and that all primes congruent to 1, 3,

[2]In each of the situations (a) and (b) of Proposition 1.17 below, R is a principal ideal domain only if Gauss's group is trivial. In all other cases, Gauss's group is nontrivial, and R is a principal ideal domain only if the group has order 2.

7, or 9 modulo 20 are representable by some form. This example is special in that equivalence and proper equivalence come to the same thing. Gauss's multiplication rule for proper equivalence classes of forms with discriminant -20 produces a group of order 2, with $x^2 + 5y^2$ representing the identity class and $2x^2 + 2xy + 3y^2$ representing the other class. Consequently

$$\begin{aligned} p \equiv 1 \text{ or } 9 \bmod 20 \quad &\text{implies} \quad x^2 + 5y^2 = p \quad \text{solvable,} \\ p \equiv 3 \text{ or } 7 \bmod 20 \quad &\text{implies} \quad 2x^2 + 2xy + 3y^2 = p \quad \text{solvable.} \end{aligned}$$

In addition, the multiplication rule has the property that if m is representable by all forms in the class of $a_1x^2 + b_1xy + c_1y^2$ and n is representable by all forms in the class of $a_2x^2 + b_2xy + c_2y^2$, then mn is representable by all forms in the class of the product form. It is not necessary to have an explicit identity for the multiplication. Thus, for example, it follows without further argument that if p and q are primes congruent to 3 or 7 modulo 20, then $x^2 + 5y^2 = pq$ is solvable.

Let us elaborate a little about the rephrased context for Gauss's theory. We let D be the discriminant of the binary quadratic forms in question, and we assume that D is "fundamental." Let R be the ring of algebraic integers that lie in the field $\mathbb{Q}(\sqrt{D}\,)$. It turns out to be possible to define a notion of "strict equivalence" on the set of ideals of R in such a way that multiplication of ideals descends to a multiplication of strict equivalence classes. The strict equivalence classes of ideals then form a group, and this group is isomorphic to Gauss's group. In particular, one obtains the nonobvious conclusion that the set of strict equivalence classes of ideals is finite. The main result giving this isomorphism is Theorem 1.20. This rephrasing of the theory points to a generalization to algebraic number fields of degree higher than 2 and is a starting point for modern algebraic number theory.

Now we return to the work of Gauss. Even the example with $D = -20$ that was described above does not give an idea of how complicated matters can become. For discriminant -56, for example, the two forms x^2+14y^2 and $2x^2+7y^2$ take on the same residues modulo 56 that are prime to 56, but no prime can be represented by both forms. These two forms and the forms $3x^2 \pm 2xy + 5y^2$ represent the four proper equivalence classes. By contrast, there are only three equivalence classes in Lagrange's sense, and we thus get some insight into why Legendre encountered difficulties in defining a useful multiplication even for $D = -56$. Gauss's theory goes on to address the problem that $x^2 + 14y^2$ and $2x^2 + 7y^2$ take on one set of residues modulo 56 and prime to 56 while $3x^2 \pm 2xy + 5y^2$ take on a disjoint set of such residues. Gauss defined a "genus" (plural: "genera") to consist of proper equivalence classes like these that cannot be distinguished by linear congruences, and he obtained some results about this notion. Gauss's set of genera inherits a group structure from the group structure on the proper equivalence classes of forms, and the group structure for the genera enables one to work with genera easily.

The third jewel of classical number theory to be discussed in this chapter is Dirichlet's celebrated theorem on primes in arithmetic progressions, given below as Theorem 1.21. The statement is that if m and b are positive relatively prime integers, then there are infinitely many primes of the form $km+b$ with k a positive integer. The proof mixes algebra, a little real analysis, and some complex analysis.

What is not immediately apparent is how this theorem fits into a natural historical sequence with Gauss's theory of binary quadratic forms. In fact, the statement about primes in arithmetic progressions was thrust upon Dirichlet in at least two ways. Dirichlet thoroughly studied the work of those who came before him. One aspect of that work was Legendre's progress toward obtaining quadratic reciprocity; in fact, Legendre actually had a proof of quadratic reciprocity except that he assumed the unproved result about primes in arithmetic progressions for part of it and argued in circular fashion for another part of it. Another aspect of the work Dirichlet studied was Gauss's theory of multiplication of proper equivalence classes of forms, which Dirichlet saw a need to simplify and explain; indeed, a complete answer to the representability of composite numbers requires establishing theorems about genera beyond what Gauss obtained and has to make use of the theorem about primes in arithmetic progressions.

In addition, Dirichlet asked and settled a question about proper equivalence classes for which Gauss had published nothing and for which Jacobi had conjectured an answer: How many such classes are there for each discriminant D? Let us call this number the "Dirichlet class number," denoting it by $h(D)$. Dirichlet's answer has several cases to it. When D is fundamental, even, negative, and not equal to -4, the answer is

$$h(D) = \frac{2\sqrt{|D/4|}}{\pi} \sum_{\substack{n\ge 1,\\ \mathrm{GCD}(n,D)=1}} \left(\frac{D/4}{n}\right)\frac{1}{n},$$

with the sum taken over positive integers prime to D. Here when p is a prime not dividing D, $\left(\frac{D/4}{p}\right)$ is $+1$ if $D/4$ is a square modulo p and is -1 if not. For general $n = \prod p^k$ prime to D, $\left(\frac{D/4}{n}\right)$ is the product of the expressions $\left(\frac{D/4}{p}\right)^k$ corresponding to the factorization[3] of n. When $D = -4$, the quantity on the right side has to be doubled to give the correct result, and thus the formula becomes

$$h(-4) = \frac{4}{\pi} \sum_{n \text{ odd} \ge 1} \left(\frac{-1}{n}\right)\frac{1}{n} = \frac{4}{\pi} \sum_{n \text{ odd} \ge 1} \frac{(-1)^{(n-1)/2}}{n}.$$

The adjusted formula correctly gives $h(-4) = -1$, since Leibniz had shown more than a century earlier that $1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \frac{\pi}{4}$. Dirichlet was able to

[3]The expression $\left(\frac{D/4}{n}\right)$ is called a "Jacobi symbol." See Problems 9–11 at the end of the chapter.

evaluate the displayed infinite series for general D as a finite sum, but that further step does not concern us here. The important thing to observe is that the infinite series is always an instance of a series $\sum_{n=1}^{\infty} \chi(n)/n$ with χ a periodic function on the positive integers satisfying $\chi(m+n) = \chi(m)\chi(n)$. Dirichlet's derivation of a series expansion for his class numbers required care because the series is only conditionally convergent. To be able to work with absolutely convergent series, he initially replaced $\frac{1}{n}$ by $\frac{1}{n^s}$ for $s > 1$, thus initially treating series he denoted by $L(s, \chi) = \sum_{n=1}^{\infty} \chi(n)/n^s$.

As a consequence of this work, Dirichlet was familiar with series $L(s, \chi)$ and was aware of the importance of expressions $L(1, \chi)$, knowing that at least when $\chi(n) = \left(\frac{D}{n}\right)$, $L(1, \chi)$ is not 0 because it is essentially a class number. This nonvanishing turns out to be the core of the proof of the theorem on primes in arithmetic progressions. Dirichlet would have known about Euler's proof that the progressions $4n + 1$ and $4n + 3$ contain infinitely many primes, a proof that we give in Section 8, and he would have recognized Euler's expression $\sum_{n=1}^{\infty} (-1)^n/(2n+1)$ as something that occurs in his formula for $h(-4)$. Thus he was well equipped with tools and motivation for a proof of his theorem on primes in arithmetic progressions.

2. Quadratic Reciprocity

If p is an odd prime number and a is an integer with $a \not\equiv 0 \bmod p$, the **Legendre symbol** $\left(\frac{a}{p}\right)$ is defined by

$$\left(\frac{a}{p}\right) = \begin{cases} +1 & \text{if } a \text{ is a square modulo } p, \\ -1 & \text{if } a \text{ is not a square modulo } p. \end{cases}$$

Since $\mathbb{F}_p^{\times}$ is a cyclic group of even order, the squares form a subgroup of index 2. Therefore $a \mapsto \left(\frac{a}{p}\right)$ is a group homomorphism of $\mathbb{F}_p^{\times}$ into $\{\pm 1\}$, and we have $\left(\frac{a}{p}\right)\left(\frac{b}{p}\right) = \left(\frac{ab}{p}\right)$ whenever a and b are not divisible by p.

Theorem 1.2 (Law of Quadratic Reciprocity). If p and q are distinct odd prime numbers, then

(a) $\left(\dfrac{-1}{p}\right) = (-1)^{\frac{1}{2}(p-1)}$,

(b) $\left(\dfrac{2}{p}\right) = (-1)^{\frac{1}{8}(p^2-1)}$,

(c) $\left(\dfrac{p}{q}\right)\left(\dfrac{q}{p}\right) = (-1)^{[\frac{1}{2}(p-1)][\frac{1}{2}(q-1)]}$.

REMARKS. Conclusion (a) is due to Fermat and says that -1 is a square modulo p if and only if $p = 4n + 1$. We proved this result already in Section 1 and will not re-prove it here. Conclusion (b) is due to Euler and says that 2 is a square modulo p if and only if $p = 8n \pm 1$. Conclusion (c) is due to Gauss and says that if p or q is $4n + 1$, then $\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right)$ and otherwise $\left(\frac{p}{q}\right) = -\left(\frac{q}{p}\right)$. The proofs of (b) and (c) will occupy the remainder of this section.

EXAMPLES.

(1) This example illustrates how quickly iterated use of the theorem decides whether a given integer is a square. We compute $\left(\frac{17}{79}\right)$. We have

$$\left(\frac{17}{79}\right) = \left(\frac{79}{17}\right) = \left(\frac{11}{17}\right) = \left(\frac{17}{11}\right) = \left(\frac{6}{11}\right) = -\left(\frac{3}{11}\right) = +\left(\frac{11}{3}\right) = \left(\frac{2}{3}\right) = -1,$$

the successive equalities being justified by using (c), the formula $\left(\frac{a+kp}{p}\right) = \left(\frac{a}{p}\right)$, (c) again, $\left(\frac{a+kp}{p}\right) = \left(\frac{a}{p}\right)$ again, the formula $\left(\frac{a}{p}\right)\left(\frac{b}{p}\right) = \left(\frac{ab}{p}\right)$ and (b), (c) once more, $\left(\frac{a+kp}{p}\right) = \left(\frac{a}{p}\right)$ once more, and an explicit evaluation of $\left(\frac{2}{3}\right)$.

(2) Lemma 9.46 of *Basic Algebra* asserts that 3 is a generator of the cyclic group $\mathbb{F}_n^\times$ when n is prime of the form $2^{2^N} + 1$ with $N > 0$, and Theorem 1.2 enables us to give a proof. In fact, this n has $n \equiv 2 \bmod 3$ and $n \equiv 1 \bmod 4$. Thus $\left(\frac{3}{n}\right) = \left(\frac{n}{3}\right) = \left(\frac{2}{3}\right) = -1$. Since $\mathbb{F}_n^\times$ is a cyclic group whose order is a power of 2, every nonsquare is a generator. Thus 3 is a generator.

We prove two lemmas, give the proof of (b), prove a third lemma, and then give the proof of (c).

Lemma 1.3. If p is an odd prime and a is any integer such that p does not divide a, then $a^{\frac{1}{2}(p-1)} \equiv \left(\frac{a}{p}\right) \bmod p$.

PROOF. The multiplicative group $\mathbb{F}_p^\times$ being cyclic, let b be a generator. Write $a \equiv b^r \bmod p$ for some integer r. Since $\left(\frac{a}{p}\right) = (-1)^r$ and $a^{\frac{1}{2}(p-1)} \equiv (b^r)^{\frac{1}{2}(p-1)} = (b^{\frac{1}{2}(p-1)})^r \equiv (-1)^r \bmod p$, the lemma follows. □

Lemma 1.4 (Gauss). Let p be an odd prime, and let a be any integer such that p does not divide a. Among the least positive residues modulo p of the integers $a, 2a, 3a, \ldots, \frac{1}{2}(p-1)a$, let n denote the number of residues that exceed $p/2$. Then $\left(\frac{a}{p}\right) = (-1)^n$.

PROOF. Let $r_1, \ldots, r_n$ be the least positive residues exceeding $p/2$, and let $s_1, \ldots, s_k$ be those less than $p/2$, so that $n + k = \frac{1}{2}(p-1)$. The residues $r_1, \ldots, r_n, s_1, \ldots, s_k$ are distinct, since no two of $a, 2a, 3a, \ldots, \frac{1}{2}(p-1)a$ differ by a multiple of p. Each integer $p - r_i$ is strictly between 0 and $p/2$, and we cannot have any equality $p - r_i = s_j$, since $r_i + s_j = p$ would mean that $(u+v)a$ is divisible by p for some integers u and v with $1 \le u, v \le \frac{1}{2}(p-1)$. Hence

$$p - r_1, \ldots, p - r_n, s_1, \ldots, s_k$$

is a permutation of $1, \ldots, \frac{1}{2}(p-1)$. Modulo p, we therefore have

$$\begin{aligned} 1 \cdot 2 \cdots \tfrac{1}{2}(p-1) &\equiv (-1)^n r_1 \cdots r_n s_1 \cdots s_k \\ &\equiv (-1)^n a \cdot 2a \cdots \tfrac{1}{2}(p-1)a \\ &\equiv (-1)^n a^{\frac{1}{2}(p-1)} 1 \cdot 2 \cdots \tfrac{1}{2}(p-1), \end{aligned}$$

and cancellation yields $a^{\frac{1}{2}(p-1)} \equiv (-1)^n \bmod p$. The result follows by combining this congruence with the conclusion of Lemma 1.3. □

PROOF OF (b) IN THEOREM 1.2. We shall apply Lemma 1.4 with $a = 2$ after investigating the least positive residues of $2, 4, 6, \ldots, p-1$. We can list explicitly those residues that exceed $p/2$ for each odd value of p mod 8 as follows:

$$\begin{aligned} p &= 8k+1, && 4k+2, 4k+4, \ldots, 8k, \\ p &= 8k+3, && 4k+2, 4k+4, \ldots, 8k+2, \\ p &= 8k+5, && 4k+4, \ldots, 8k+2, 8k+4, \\ p &= 8k+7, && 4k+4, \ldots, 8k+4, 8k+6. \end{aligned}$$

If n denotes the number of such residues for a given p, a count of each line of the above table shows that

$$\begin{aligned} n &= 2k && \text{and} && (-1)^n = +1 && \text{for} && p = 8k+1, \\ n &= 2k+1 && \text{and} && (-1)^n = -1 && \text{for} && p = 8k+3, \\ n &= 2k+1 && \text{and} && (-1)^n = -1 && \text{for} && p = 8k+5, \\ n &= 2k+2 && \text{and} && (-1)^n = +1 && \text{for} && p = 8k+7. \end{aligned}$$

Thus Lemma 1.4 shows that $\left(\frac{2}{p}\right) = +1$ for $p = 8k \pm 1$ and $\left(\frac{2}{p}\right) = -1$ for $p = 8k \pm 3$. This completes the proof of (b). □

Lemma 1.5. If p is an odd prime and a is a positive odd integer such that p does not divide a, then $\left(\frac{a}{p}\right) = (-1)^t$, where $t = \sum_{u=1}^{\frac{1}{2}(p-1)} [ua/p]$. Here $[\,\cdot\,]$ denotes the greatest-integer function.

REMARKS. When $a = 2$, the equality $\left(\frac{a}{p}\right) = (-1)^t$ fails for $p = 3$, since $t = [2/3] = 0$.

PROOF. With notation as in Lemma 1.4 and its proof, we form each ua for $1 \le u \le \frac{1}{2}(p-1)$ and reduce modulo p, obtaining as least positive residue either some r_i for $i \le n$ or some s_j for $j \le k$. Then $ua/p = [ua/p] + p^{-1}(\text{some } r_i \text{ or } s_j)$. Hence

$$\sum_{u=1}^{\frac{1}{2}(p-1)} ua = \sum_{u=1}^{\frac{1}{2}(p-1)} p[ua/p] + \sum_{i=1}^{n} r_i + \sum_{j=1}^{k} s_j. \qquad (*)$$

The proof of Lemma 1.4 showed that $p-r_1, \ldots, p-r_n, s_1, \ldots, s_k$ is a permutation of $1, \ldots, \frac{1}{2}(p-1)$, and thus the sum is the same in the two cases:

$$\sum_{u=1}^{\frac{1}{2}(p-1)} u = \sum_{i=1}^{n} (p - r_i) + \sum_{j=1}^{k} s_j = np - \sum_{i=1}^{n} r_i + \sum_{j=1}^{k} s_j.$$

Subtracting this equation from $(*)$, we obtain

$$(a-1) \sum_{u=1}^{\frac{1}{2}(p-1)} u = p\Big(\sum_{u=1}^{\frac{1}{2}(p-1)} [ua/p] - n \Big) + 2 \sum_{i=1}^{n} r_i.$$

Replacing $\sum_{u=1}^{\frac{1}{2}(p-1)} u$ on the left side by its value $\frac{1}{8}(p^2-1)$ and taking into account that p is odd, we obtain the following congruence modulo 2:

$$(a-1)\tfrac{1}{8}(p^2-1) \equiv \sum_{u=1}^{\frac{1}{2}(p-1)} [ua/p] - n \mod 2.$$

Since a is odd, the left side is congruent to 0 modulo 2. Therefore $n \equiv \sum_{u=1}^{\frac{1}{2}(p-1)} [ua/p] \equiv t \mod 2$, and Lemma 1.4 allows us to conclude that $(-1)^t = (-1)^n = \left(\frac{a}{p}\right)$. □

PROOF OF (c) IN THEOREM 1.2. Let

$$S = \big\{(x, y) \in \mathbb{Z} \times \mathbb{Z} \,\big|\, 1 \le x \le \tfrac{1}{2}(p-1) \text{ and } 1 \le y \le \tfrac{1}{2}(q-1)\big\},$$

the number of elements in question being $|S| = \frac{1}{4}(p-1)(q-1)$. We can write $S = S_1 \cup S_2$ disjointly with

$$S_1 = \{(x, y) \mid qx > py\} \quad \text{and} \quad S_2 = \{(x, y) \mid qx < py\};$$

the exhaustion of S by S_1 and S_2 follows because $qx = py$ would imply that p divides qx and hence that p divides x, contradiction. We can describe S_1 alternatively as

$$S_1 = \left\{(x, y) \;\middle|\; 1 \le x \le \tfrac{1}{2}(p-1) \text{ and } 1 \le y < qx/p\right\},$$

and therefore $|S_1| = \sum_{x=1}^{\frac{1}{2}(p-1)} [qx/p]$, which is the integer t in Lemma 1.5 such that $(-1)^t = \left(\frac{q}{p}\right)$. Similarly we have $|S_2| = \sum_{y=1}^{\frac{1}{2}(q-1)} [py/q]$, which is the integer t in Lemma 1.5 such that $(-1)^t = \left(\frac{p}{q}\right)$. Therefore

$$(-1)^{\frac{1}{4}(p-1)(q-1)} = (-1)^{|S|} = (-1)^{|S_1|}(-1)^{|S_2|} = \left(\tfrac{q}{p}\right)\left(\tfrac{p}{q}\right),$$

and the proof is complete. □

3. Equivalence and Reduction of Quadratic Forms

A **binary quadratic form** over $\mathbb{Z}$ is a function $F(x, y) = ax^2 + bxy + cy^2$ from $\mathbb{Z} \times \mathbb{Z}$ to $\mathbb{Z}$ with a, b, c in $\mathbb{Z}$. Following Gauss,[4] we abbreviate this F as (a, b, c). We shall always assume, without explicitly saying so, that the **discriminant** $D = b^2 - 4ac$ is not the square of an integer and that F is **primitive** in the sense that $\mathrm{GCD}(a, b, c) = 1$. When there is no possible ambiguity, we may say "form" or "quadratic form" in place of "binary quadratic form."

Let $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ be a member of the group $\mathrm{GL}(2, \mathbb{Z})$ of integer matrices whose inverse is an integer matrix. The determinant of such a matrix is ± 1. We can use this matrix to change variables, writing

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \alpha x' + \beta y' \\ \gamma x' + \delta y' \end{pmatrix}.$$

Then $ax^2 + bxy + cy^2$ becomes

$$\begin{aligned} &a(\alpha x' + \beta y')^2 + b(\alpha x' + \beta y')(\gamma x' + \delta y') + c(\gamma x' + \delta y')^2 \\ &= (a\alpha^2 + b\alpha\gamma + c\gamma^2)x'^2 + (2a\alpha\beta + b\alpha\delta + b\beta\gamma + 2c\gamma\delta)x'y' + (a\beta^2 + b\beta\delta + c\delta^2)y'^2. \end{aligned}$$

[4] *Disquisitiones Arithmeticae*, Article 153. Actually, Gauss always assumed that the coefficient of xy is even and consequently wrote (a, b, c) for $ax^2 + 2bxy + cy^2$. To study $x^2 + xy + y^2$, for example, he took $a = 2, b = 1, c = 2$. The convention of working with $ax^2 + bxy + cy^2$ is due to Eisenstein.

If we associate the triple (a, b, c) of $F(x, y)$ to the matrix $\left(\begin{smallmatrix} 2a & b \\ b & 2c \end{smallmatrix}\right)$, then this formula shows that the triple (a', b', c') of the new form $F'(x', y')$ is associated to the matrix

$$\begin{pmatrix} 2a' & b' \\ b' & 2c' \end{pmatrix} = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.$$

From this equality of matrices, we see that

(i) the member $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ of $\mathrm{GL}(2, \mathbb{Z})$ has the effect of the identity transformation,

(ii) the member $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)\left(\begin{smallmatrix} \alpha' & \beta' \\ \gamma' & \delta' \end{smallmatrix}\right)$ of $\mathrm{GL}(2, \mathbb{Z})$ has the effect of applying first $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ and then $\left(\begin{smallmatrix} \alpha' & \beta' \\ \gamma' & \delta' \end{smallmatrix}\right)$.

These two facts say that we do not quite have the expected group action on forms on the left. Instead, we can say either that we have a group action on the right or that gF is obtained from F by operating by g^t. Anyway, there are orbits, and they are what we really need. The discriminant $D = b^2 - 4ac$ of the form F is evidently minus the determinant of the associated matrix $\left(\begin{smallmatrix} 2a & b \\ b & 2c \end{smallmatrix}\right)$, and the displayed equality of matrices thus implies that the discriminant of the form F' is $D(\alpha\delta - \beta\gamma)^2$. Since $(\alpha\delta - \beta\gamma)^2 = 1$ for matrices in $\mathrm{GL}(2, \mathbb{Z})$, we conclude that

(iii) each member of $\mathrm{GL}(2, \mathbb{Z})$ preserves the discriminant of the form.

Hence the group $\mathrm{GL}(2, \mathbb{Z})$ acts on the forms of discriminant D.

Forms in the same orbit under $\mathrm{GL}(2, \mathbb{Z})$ are said to be **equivalent**. Forms in the same orbit under the subgroup $\mathrm{SL}(2, \mathbb{Z})$ are said to be **properly equivalent**. A **proper equivalence class** of forms will refer to the latter relation. This notion is due to Gauss. Equivalence under $\mathrm{GL}(2, \mathbb{Z})$ is an earlier notion due to Lagrange, and we shall refer to its classes as **ordinary equivalence classes** on the infrequent occasions when the notion arises. Proper equivalence is necessary later in order to get a group operation on classes of forms. If one form can be carried to another form by a member of $\mathrm{GL}(2, \mathbb{Z})$ of determinant -1, we say that the two forms are **improperly equivalent**. Use of the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ shows that the form (a, b, c) is improperly equivalent to the form $(a, -b, c)$. In particular, $(a, 0, c)$ is improperly equivalent to itself.

The discriminant D is congruent to b^2 modulo 4 and hence is congruent to 0 or 1 modulo 4. All nonsquare integers D that are congruent to 0 or 1 modulo 4 arise as discriminants; in fact, we can always achieve such a D with $a = 1$ and with b equal either to 0 or to 1.

The discriminant is minus the determinant of the matrix $\left(\begin{smallmatrix} 2a & b \\ b & 2c \end{smallmatrix}\right)$ associated to

(a, b, c), and this matrix is real symmetric with trace $2(a+c)$. Since $D = b^2-4ac$ is assumed not to be the square of an integer, neither a nor c can be 0.

If $D > 0$, the symmetric matrix $\begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix}$ is indefinite, having eigenvalues of opposite sign. In this case the **Dirichlet class number** of D, denoted by $h(D)$, is defined to be the number[5] of all proper equivalence classes of forms of discriminant D.

If $D < 0$, then a and c have the same sign. The matrix $\begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix}$ is positive definite if a and c are positive, and it is negative definite if a and c are negative. Correspondingly we refer to the form (a, b, c) as **positive definite** or **negative definite** in the two cases. Since $g^t \begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix} g$ is positive definite whenever $\begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix}$ is positive definite, any form equivalent to a positive definite form is again positive definite. A similar remark applies to negative definite forms. Thus "positive definite" and "negative definite" are class properties. For any given discriminant $D < 0$, the **Dirichlet class number** of D, denoted by $h(D)$, is the number[6] of proper equivalence classes of *positive definite* forms of discriminant D.

The form (a, b, c) **represents** an integer m if $ax^2+bxy+cy^2 = m$ is solvable for some integers x and y. The form **primitively represents** m if the x and y with $ax^2 + bxy + cy^2 = m$ can be chosen to be relatively prime. In any event, GCD(x, y) divides m, and thus whenever a form represents a prime p, it primitively represents p.

Theorem 1.6. Fix a nonsquare discriminant D.

(a) The Dirichlet class number $h(D)$ is finite. In fact, any form of discriminant D is properly equivalent to a form (a, b, c) with $|b| \le |a| \le |c|$ and therefore has $3|ac| \le |D|$, and the number of forms of discriminant D satisfying all these inequalities is finite.

(b) An odd prime p with GCD$(D, p) = 1$ is primitively representable by some form (a, b, c) of discriminant D if and only if $\left(\frac{D}{p}\right) = +1$. In this case the number of proper equivalence classes of forms primitively representing p is either 1 or 2, and these classes are carried to one another by GL$(2, \mathbb{Z})$. In fact, if $\left(\frac{D}{p}\right) = +1$, then $b^2 \equiv D \bmod 4p$ for some integer b, and representatives of these classes may be taken to be $\left(p, \pm b, \frac{b^2-D}{4p}\right)$.

[5]This number was studied by Dirichlet. According to Theorem 1.20 below, it counts the "strict equivalence classes" of ideals in a sense that is introduced in Section 7. This number either equals or is twice the number of equivalence classes of ideals in the other sense that is introduced in Section 7. The latter is what is generalized in Chapter V in the subject of algebraic number theory, and the latter is how "class number" is usually defined in modern books in algebraic number theory. Consequently Dirichlet class numbers sometimes are twice what modern class numbers are. We use "Dirichlet class numbers" in this chapter and change to the modern "class numbers" in Chapter V.

[6]This number was studied by Dirichlet. See the previous footnote for further information.

We come to the proof after some preliminary remarks and examples. The argument for (a) is constructive, and thus the forms given explicitly in (b) can be transformed constructively into properly equivalent forms satisfying the conditions of (a). Hence we are led to explicit forms as in (a) representing p. A generalization of (b) concerning how a composite integer m can be represented if $\mathrm{GCD}(D, m) = 1$ appears in Problem 2 at the end of the chapter. What is missing in all this is a description of proper equivalences among the forms as in (a). We shall solve this question readily in Proposition 1.7 when $D < 0$. For $D > 0$, the answer is more complicated; we shall say what it is in Theorem 1.8, but we shall omit some of the proof of that theorem.

EXAMPLES.

(1) $D = -4$. Theorem 1.2a shows that the odd primes with $\left(\frac{D}{p}\right) = +1$ are those of the form $4k + 1$. Theorem 1.6a says that each proper equivalence class of forms of discriminant -4 has a representative (a, b, c) with $3|ac| \le 4$. Since $D < 0$, we are interested only in positive definite forms, which necessarily have a and c positive. Thus $a = c = 1$, and we must have $b = 0$. So there is only one class of (positive definite) forms of discriminant -4, namely $x^2 + y^2$, and Theorem 1.6b allows us to conclude that $x^2 + y^2 = p$ is solvable for each prime $p = 4k + 1$. In other words, we recover the conclusion of Proposition 1.1 as far as representability of primes is concerned.

(2) $D = -20$. To have $\left(\frac{D}{p}\right) = +1$ for an odd prime p, we must have either $\left(\frac{-1}{p}\right) = \left(\frac{5}{p}\right) = +1$ or $\left(\frac{-1}{p}\right) = \left(\frac{5}{p}\right) = -1$. Theorem 1.2 shows in the first case that $p \equiv 1 \bmod 4$ and $p \equiv \pm 1 \bmod 5$, while in the second case $p \equiv 3 \bmod 4$ and $p \equiv \pm 3 \bmod 5$. That is, p is congruent to one of 1 and 9 modulo 20 in the first case and to one of 3 and 7 modulo 20 in the second case. Let us consider the forms as in Theorem 1.6a. We know that $a > 0$ and $c > 0$. The inequality $3ac \le |D|$ forces $ac \le 6$. Since $|b| \le a \le c$, we obtain $a^2 \le 6$ and $a \le 2$. Since 4 divides D, b is even. Then $b = 0$ or $b = \pm 2$. So the only possibilities are $(1, 0, 5)$ and $(2, \pm 2, 3)$. Because of Theorem 1.6b, any prime congruent to one of 1, 3, 7, 9 modulo 20 is representable either by $(1, 0, 5)$ and not $(2, \pm 2, 3)$, or by $(2, \pm 2, 3)$ and not $(1, 0, 5)$. We can write down all residues modulo 20 for $x^2 + 5y^2$ and $2x^2 \pm 2xy + 3y^2$, and we find that the possible residues prime to 20 are 1 and 9 in the first case, and they are 3 and 7 in the second case. The conclusion for odd primes p with $\mathrm{GCD}(20, p) = 1$ is that

$$\begin{array}{lll} p \equiv 1 \text{ or } 9 \bmod 20 & \text{implies} & p \text{ is representable as } x^2 + 5y^2, \\ p \equiv 3 \text{ or } 7 \bmod 20 & \text{implies} & p \text{ is representable as } 2x^2 \pm 2xy + 3y^2. \end{array}$$

The residues modulo 20 have shown that $x^2 + 5y^2$ is not equivalent to either of $2x^2 \pm 2xy + 3y^2$, but they do not show whether $2x^2 \pm 2xy + 3y^2$ are properly

equivalent to one another. Hence the Dirichlet class number $h(-20)$ is either 2 or 3. It will turn out to be 2.

(3) $D = -56$. To have $\left(\frac{D}{p}\right) = +1$ for an odd prime p, we must have an odd number of the Legendre symbols $\left(\frac{-1}{p}\right)$, $\left(\frac{2}{p}\right)$, and $\left(\frac{7}{p}\right)$ equal to $+1$ and the rest equal to -1. We readily find from Theorem 1.2 that the possibilities with GCD(56, p) = 1 are

$$p \equiv 1, 3, 5, 9, 13, 15, 19, 23, 25, 27, 39, 45 \bmod 56.$$

Applying Theorem 1.6a as in the previous example, we find that $x^2 + 14y^2$, $2x^2 + 7y^2$, and $3x^2 \pm 2xy + 5y^2$ are representatives of all proper equivalence classes of forms of discriminant -56. Taking into account Theorem 1.6b and the residue classes of these forms modulo 56, we conclude for odd primes p that

$$\begin{aligned}
&\text{if } p \equiv \text{any of } 1, 9, 15, 23, 25, 39 \bmod 56, \text{ then}\\
&\qquad p \text{ is representable as } x^2 + 14y^2 \text{ or } 2x^2 + 7y^2,\\
&\text{if } p \equiv \text{any of } 3, 5, 13, 19, 27, 45 \bmod 56, \text{ then}\\
&\qquad p \text{ is representable as both of } 3x^2 \pm 2xy + 5y^2.
\end{aligned}$$

The question left unsettled by the argument so far is whether x^2+14y^2 is properly equivalent to $2x^2 + 7y^2$. Equivalent forms represent the same integers, and the integer 1 is representable by x^2+14y^2 but not by $2x^2+7y^2$. Hence the two forms are not equivalent and cannot be properly equivalent. According to Theorem 1.6b, the primes of the first line are therefore representable by either x^2+14y^2 or $2x^2 + 7y^2$ but *never* by both. Hence the Dirichlet class number $h(-56)$ is either 3 or 4. It will turn out to be 4.

(4) $D = 5$. The forms of discriminant 5 are indefinite. Applying Theorem 1.6a, we obtain $3|ac| \le 5$. Hence $|a| = |c| = 1$. Since D is odd, b is odd. The inequality $|b| \le |a|$ thus forces $|b| = 1$. Then $D = 1 - 4ac$ shows that $ac < 0$. The possibilities are therefore $(1, \pm1, -1)$ and $(-1, \pm1, 1)$. The Dirichlet class number $h(5)$ is at most 4. It will turn out to be 1. Let us take this fact as known. The odd primes p with $\left(\frac{D}{p}\right) = +1$ are $p = 5k \pm 1$. Under the assumption that the class number is 1, Theorem 1.6b shows that every such prime is representable as $x^2 + xy - y^2$.

PROOF OF THEOREM 1.6a. We consider the effect of two transformations in SL(2, $\mathbb{Z}$), one via $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and the other via $\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)$. Under these, the matrix associated to (a, b, c) becomes

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2c & -b \\ -b & 2a \end{pmatrix}$$

and $$\begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix}\begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix}\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2a & 2an+b \\ 2an+b & 2an^2+2bn+2c \end{pmatrix},$$

respectively. Thus the transformations are

$$(a, b, c) \longmapsto (c, -b, a), \tag{$*$}$$

$$(a, b, c) \longmapsto (a, 2an + b, c'). \tag{$**$}$$

Possibly applying $(*)$ allows us to make $|a| \le |c|$ while leaving $|b|$ alone. Since $a \neq 0$, we can apply $(**)$ with n the closest integer to $-\frac{b}{2a}$ to make $|b| \le |a|$. This step possibly changes c. Thus after this step, we again apply $(*)$ if necessary to make $|a| \le |c|$, and we apply $(**)$ again. In each pair of steps, we may assume that $|b|$ strictly decreases or else that $n = 0$. We cannot always be in the former case, since $|b|$ is bounded below by 0. Thus at some point we obtain $n = 0$. At this point, c does not change, and thus we have $|b| \le |a| \le |c|$, as required.

The inequalities $|b| \le |a| \le |c|$ imply that

$$4|ac| = |D - b^2| \le |D| + |b|^2 \le |D| + |ac|,$$

and hence $3|ac| \le |D|$. Since neither a nor c is 0, it follows that the inequalities $|b| \le |a| \le |c|$ imply that $|a|, |b|, |c|$ are all bounded by $|D|$. Therefore the Dirichlet class number $h(D)$ is finite. □

PROOF OF NECESSITY IN THEOREM 1.6b. Suppose x and y are integers with $\mathrm{GCD}(x, y) = 1$ and $ax^2 + bxy + cy^2 = p$. Then $ax^2 + bxy + cy^2 \equiv 0 \bmod p$. Choose u and v with $ux + vy = 1$. Routine computation shows that

$$\begin{aligned} 4(ax^2+bxy + cy^2)(av^2 - buv + cu^2) & \\ &= [u(xb + 2yc) - v(2xa + yb)]^2 - (b^2 - 4ac)(xu + yv)^2 \\ &= [u(xb + 2yc) - v(2xa + yb)]^2 - (b^2 - 4ac), \end{aligned}$$

and hence

$$0 \equiv [u(xb + 2yc) - v(2xa + yb)]^2 - (b^2 - 4ac) \bmod p.$$

Consequently $D \equiv [u(xb+2yc) - v(2xa + yb)]^2 \bmod p$, and D is exhibited as a square modulo p. □

PROOF OF SUFFICIENCY IN THEOREM 1.6b. Choose an integer solution b of $b^2 \equiv D \bmod p$. Since $b + p$ is another solution and has the opposite parity, we may assume that b and D have the same parity. Then $b^2 \equiv D \bmod p$ and $b^2 \equiv D \bmod 4$, so that $b^2 \equiv D \bmod 4p$. Since $\mathrm{GCD}(D, p) = 1$, p does not

divide b, and the forms $\left(p, \pm b, \frac{b^2-D}{4p}\right)$ are primitive. They have discriminant $b^2 - 4p\,\frac{b^2-D}{4p} = D$, they take the value p for $(x, y) = (1, 0)$, and they are improperly equivalent via $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Thus the forms in the statement of the theorem exist.

For the uniqueness suppose that a form (a, b, c) of discriminant D represents p, say with $ax_0^2 + bx_0y_0 + cy_0^2 = p$. Since this representation has to be primitive, we know that $\mathrm{GCD}(x_0, y_0) = 1$. Put $\left(\begin{smallmatrix} \alpha \\ \gamma \end{smallmatrix}\right) = \left(\begin{smallmatrix} x_0 \\ y_0 \end{smallmatrix}\right)$, and choose integers β and δ such that $\alpha\delta - \beta\gamma = 1$. Then $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ has determinant 1 and satisfies $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} x_0 \\ y_0 \end{smallmatrix}\right)$. The equality $ax_0^2 + bx_0y_0 + cy_0^2 = \frac{1}{2}(x_0 \;\; y_0)\left(\begin{smallmatrix} 2a & b \\ b & 2c \end{smallmatrix}\right)\left(\begin{smallmatrix} x_0 \\ y_0 \end{smallmatrix}\right)$ therefore yields

$$p = \tfrac{1}{2}\begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}\begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix}\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Consequently the form (a', b', c') associated to the matrix $\left(\begin{smallmatrix} \alpha & \gamma \\ \beta & \delta \end{smallmatrix}\right)\left(\begin{smallmatrix} 2a & b \\ b & 2c \end{smallmatrix}\right)\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ takes on the value p at $(x, y) = (1, 0)$ and is properly equivalent to (a, b, c). In particular, it is a form (p, b', c') for some b' and c' such that $b'^2 - 4pc' = D$.

Thus in the proof of uniqueness, we may assume that we have two forms (p, b', c') and (p, b'', c'') of discriminant D. Then $b''^2 \equiv D \equiv b'^2 \bmod 4p$. The conditions $b''^2 \equiv b'^2 \bmod p$ and $b''^2 \equiv b'^2 \bmod 4$ imply that $b'' \equiv \pm b' \bmod p$ and $b'' \equiv b' \bmod 2$ for one of the choices of sign. Thus $b'' \equiv \pm b' \bmod 2p$ for that choice of sign. Let us write $b'' = \pm b' + 2np$ for some integer n. The matrix equality

$$\begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix}\begin{pmatrix} 2p & \pm b' \\ \pm b' & 2c' \end{pmatrix}\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2p & 2pn \pm b' \\ 2pn \pm b' & 2(*) \end{pmatrix}$$

shows that $(p, \pm b', c')$ is properly equivalent to $(p, b'', *)$. Since the discriminant has to be D, we conclude that $* = c''$. That is, (p, b'', c'') is properly equivalent to $(p, \pm b', c')$ for that same choice of sign. Since (p, b', c') is improperly equivalent to $(p, -b', c')$, the proof of the theorem is complete. □

Our discussion of representability of primes p by binary quadratic forms of discriminant D when $\mathrm{GCD}(D, p) = 1$ will be complete once we have a set of representatives of proper equivalence classes with no redundancy. For discriminant $D < 0$, this step is not difficult and amounts, according to Theorem 1.6a, to sorting out proper equivalences among forms (a, b, c) with $b^2 - 4ac = D$ and $|b| \le |a| \le |c|$. Let us call a form with $D < 0$ **reduced** when it satisfies these conditions.

There are two redundancies that are easy to spot, namely

$$\begin{aligned}(a,b,a) &\text{ is properly equivalent to } (a,-b,a) \qquad \text{via} \quad \left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right),\\ (a,a,c) &\text{ is properly equivalent to } (a,-a,c) \qquad \text{via} \quad \left(\begin{smallmatrix}1&-1\\0&1\end{smallmatrix}\right).\end{aligned}$$

The result for $D < 0$ is that there are no other redundancies among reduced forms.

Proposition 1.7. Fix a negative discriminant D. With the exception of the proper equivalences of

$$\begin{aligned}&(a,b,a) \quad \text{to} \quad (a,-b,a)\\ \text{and} \qquad &(a,a,c) \quad \text{to} \quad (a,-a,c),\end{aligned}$$

no two distinct reduced positive definite forms of discriminant D are properly equivalent.

PROOF. Suppose that (a,b,c) is properly equivalent to (a',b',c'), that both are reduced, and that $a \geq a' > 0$. For some $\left(\begin{smallmatrix}\alpha&\beta\\\gamma&\delta\end{smallmatrix}\right)$ in $\mathrm{SL}(2,\mathbb{Z})$, we have $a' = a\alpha^2 + b\alpha\gamma + c\gamma^2$. Hence the inequalities $c \geq a$ and $|b| \geq -a$ imply that

$$a \geq a\alpha^2+b\alpha\gamma+c\gamma^2 \geq a(\alpha^2+\gamma^2)+b\alpha\gamma \geq a(\alpha^2+\gamma^2)-a|\alpha\gamma| \geq a|\alpha\gamma|, \quad (*)$$

and $\alpha\gamma$ equals 0 or ± 1. Thus the ordered pair (α,γ) is one of $(0,\pm1)$, $(\pm1,0)$, $(\pm1,1)$, $(\pm1,-1)$. Multiplying $\left(\begin{smallmatrix}\alpha&\beta\\\gamma&\delta\end{smallmatrix}\right)$ if necessary by $\left(\begin{smallmatrix}-1&0\\0&-1\end{smallmatrix}\right)$, which acts trivially on quadratic forms, we may assume that (α,γ) is one of $(0,1)$, $(1,0)$, $(1,\pm1)$. We treat these three cases separately.

Case 1. $(\alpha,\gamma) = (0,1)$. The condition $\alpha\delta-\beta\gamma = 1$ forces $\beta\gamma = -1$, and the formula $b' = 2a\alpha\beta + b\alpha\delta + b\beta\gamma + 2c\gamma\delta$ gives $(a',b',c') = (c,-b+2c\delta,*)$. Since $|b| \leq c$ and $|b-2c\delta| \leq c$, we must have $|\delta| \leq 1$. If $\delta = 0$, we are led to $(a',b',c') = (c,-b,a)$, which is reduced only if $c = a$, and this is the first of the two allowable exceptions. If $|\delta| = 1$, the triangle inequality gives $2c = |2c\delta| \leq |b| + |2c\delta - b| \leq c + c = 2c$, and therefore $|b| = c = |b-2c\delta|$. Then $b = -(b-2c\delta)$, and $b = c\delta = \pm c$. Since $|b| \leq a \leq c$, $b = \pm a$ also. Hence $(a',b',c') = (a,-b,a)$, and this is again the first of the two allowable exceptions.

Case 2. $(\alpha,\gamma) = (1,0)$. The condition $\alpha\delta - \beta\gamma = 1$ forces $\alpha\delta = 1$, and thus $(a',b',c') = (a,b+2a\beta,*)$. Since $|b| \leq a$ and $|b+2a\beta| \leq a$, we must have $|\beta| \leq 1$. If $\beta = 0$, then $(a',b',c') = (a,b,c)$, and there is nothing to prove. If $|\beta| = 1$, the triangle inequality gives $2a = |2a\beta| \leq |-b| + |2a\beta + b|$, and

therefore $|b| = a = |b + 2\beta a|$. Then $b = -(b + 2\beta a)$, and we conclude that $b = -a\beta = \pm a$ and $b + 2\beta a = \mp a$. Hence the proper equivalence in question is of (a, a, c) to $(a, -a, c)$, which is the second of the two allowable exceptions.

Case 3. $(\alpha, \gamma) = (1, \pm 1)$. From $(*)$ and the assumption that $a \geq a'$, we have $a \geq a' \geq a|\alpha\gamma| = a$. Thus $a = a'$, and the definition of a' shows that $a = a + b\gamma + c$. Hence $c = -b\gamma$, and $c = |b|$. Since $|b| \leq a \leq c$, we obtain $-b\gamma = a = c$. The formula $b' = 2a\alpha\beta + b\alpha\delta + b\beta\gamma + 2c\gamma\delta$ then simplifies to $b' = 2a\beta + b\delta + b\beta\gamma + 2a\gamma\delta = (2a + b\gamma)(\beta + \gamma\delta)$. From $\alpha\delta - \beta\gamma = 1$, we have $\delta - \beta\gamma = 1$ and thus also $\gamma\delta = \gamma + \beta$. Therefore $\beta + \gamma\delta = 2\beta + \gamma$, and this cannot be 0. So $|b'| \geq |2a + b\gamma| = |2a - a| = a = a'$. Since (a', b', c') is reduced, $|b'| = a' = a = c = |b|$, and the proper equivalence is of (a, a, a) to $(a, -a, a)$. This is an instance of both allowable exceptions, and the proof is complete. □

EXAMPLES, CONTINUED.

(2) $D = -20$. We saw earlier that the reduced positive definite forms with $D = -20$ are $x^2 + 5y^2$ and $2x^2 \pm 2xy + 3y^2$, i.e., $(1, 0, 5)$ and $(2, \pm 2, 3)$. The remarks preceding Proposition 1.7 show that $(2, 2, 3)$ is properly equivalent to $(2, -2, 3)$, and the proposition shows that $(1, 0, 5)$ is not properly equivalent to $(2, 2, 3)$. (We saw this latter conclusion for this example earlier by considering residues.) Consequently $h(-20) = 2$.

(3) $D = -56$. We saw earlier that the reduced positive definite forms with $D = -56$ are $x^2 + 14y^2$, $2x^2 + 7y^2$, and $3x^2 \pm 2xy + 5y^2$, i.e., $(1, 0, 14)$, $(2, 0, 7)$, $(3, 2, 5)$, and $(3, -2, 5)$. Proposition 1.7 shows that no two of these four forms are properly equivalent. Consequently $h(-56) = 4$.

Let us turn our attention to $D > 0$. We still have the proper equivalences of (a, b, a) to $(a, -b, a)$ and (a, a, c) to $(a, -a, c)$ as in the remarks before Proposition 1.7. But there can be others, and the question is subtle. Here are some simple examples.

EXAMPLES WITH POSITIVE DISCRIMINANT.

(1) $D = 5$. The forms with $D = 5$ satisfying the inequalities $|b| \leq |a| \leq |c|$ of Theorem 1.6a are $(1, \pm 1, -1)$ and $(-1, \pm 1, 1)$. The second standard equivalence allows us to discard one form from each pair, and we are left with $(1, 1, -1)$ and $(-1, -1, 1)$. The first of these two is equivalent to the second via $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Thus $h(5) = 1$, as was announced without proof in Example 4 earlier in this section.

(2) $D = 13$. The forms with $D = 13$ satisfying the inequalities $|b| \leq |a| \leq |c|$ of Theorem 1.6a are $(1, \pm 1, -3)$ and $(-1, \pm 1, 3)$. The second standard

equivalence allows us to discard one form from each pair, and we are left with $(1, 1, -3)$ and $(-1, -1, 3)$. The first of these two is equivalent to the second via $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) = \left(\begin{smallmatrix}1 & -2\\ 1 & -1\end{smallmatrix}\right)$. Thus $h(13) = 1$.

(3) $D = 21$. The forms with $D = 21$ satisfying the inequalities $|b| \le |a| \le |c|$ of Theorem 1.6a are $(1, \pm 1, -5)$ and $(-1, \pm 1, 5)$. The second standard equivalence allows us to discard one form from each pair, and we are left with $(1, 1, -5)$ and $(-1, -1, 5)$. These are not properly equivalent. In fact, the form $-x^2 - xy + 5y^2$ is -1 for $(x, y) = (1, 0)$, but $x^2 + xy - 5y^2 = -1$ is not even solvable modulo 3. Thus $h(21) = 2$.

Although the starting data for these three examples are similar, the outcomes are strikingly different. The idea for what to do involves starting afresh with the reduction question that was addressed in Theorem 1.6a. For discriminant $D > 0$, a different reduction is to be used. The reduction in question appears in Theorem 1.8a below, but some preliminary remarks are needed to explain the proof.

Two forms (a, b, c) and (a', b', c') of discriminant $D > 0$ will be said to be **neighbors** if $c = a'$ and $b + b' \equiv 0 \bmod 2c$. More precisely we say in this case that (a', b', c') is a **neighbor on the right** of (a, b, c) and that (a, b, c) is a **neighbor on the left** of (a', b', c'). A key observation is that neighbors are properly equivalent to one another. In fact, if (a', b', c') is a neighbor on the right of (a, b, c), define $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) = \left(\begin{smallmatrix}0 & -1\\ 1 & (b+b')/(2c)\end{smallmatrix}\right)$. Then computation gives

$$\begin{pmatrix}\alpha & \gamma\\ \beta & \delta\end{pmatrix}\begin{pmatrix}2a & b\\ b & 2c\end{pmatrix}\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix} = \begin{pmatrix}2c & b'\\ b' & (b-b')\frac{b+b'}{2c}\end{pmatrix}.$$

The lower right entry of this matrix is an even integer, since $b + b' \equiv 0 \bmod 2c$ and since, as a consequence, $b + b' \equiv 0 \bmod 2$. Hence (a, b, c) is transformed into (c, b', c'), where $c' = \frac{1}{2}(b - b')\,\frac{b+b'}{2c}$.

Let us call a primitive form (a, b, c) of discriminant $D > 0$ **reduced** when it satisfies the conditions

$$0 < b < \sqrt{D} \qquad \text{and} \qquad \sqrt{D} - b < 2|a| < \sqrt{D} + b.$$

The first inequality shows that b is bounded if D is fixed, and the equality $-4ac = D^2 - b^2$ shows that there are only finitely many possibilities for a and c. Consequently there are only finitely many reduced forms for given D.

From $|b| < \sqrt{D}$, we see that $b^2 < D = b^2 - 4ac$ and $ac < 0$; thus any reduced form has a and c of opposite sign. Then $D - b^2 = -4ac = (2|a|)(2|c|)$, and it follows that $2|a| > \sqrt{D} - b$ implies $2|c| < \sqrt{D} + b$ and that $2|a| < \sqrt{D} + b$ implies $2|c| < \sqrt{D} - b$. Consequently

$$\sqrt{D} - b < 2|c| < \sqrt{D} + b.$$

Theorem 1.8. Fix a positive nonsquare discriminant D.

(a) Each form of discriminant D is properly equivalent to some reduced form of discriminant D.

(b) Each reduced form of discriminant D is a neighbor on the left of one and only one reduced form of discriminant D and is a neighbor on the right of one and only one reduced form of discriminant D.

(c) The reduced forms of discriminant D occur in uniquely determined cycles, each one of even length, such that each member of a cycle is an iterated neighbor on the right to all members of the cycle and consequently is properly equivalent to all other members of the cycle.

(d) Two reduced forms of discriminant D are properly equivalent if and only if they lie in the same cycle in the sense of (c).

REMARKS. Conclusion (d) is the deepest part of the theorem, involving a subtle argument that in essence uses the periodic continued-fraction expansion of the roots z of the polynomial $az^2 + bz + c$ if (a, b, c) is a form under consideration. We shall prove (a) through (c), omitting the proof of (d), and then we shall return to the three examples $D = 5, 13, 29$ begun just above.

PROOF OF THEOREM 1.8a. If (a, b, c) is given and is not reduced, let m be the unique integer such that

$$\sqrt{D} - 2|c| < -b + 2cm < \sqrt{D}, \tag{$*$}$$

and define $(a', b', c') = (c, -b + 2cm, a - bm + cm^2)$. Then

$$\begin{aligned} b'^2 - 4a'c' &= (-b + 2cm)^2 - 4c(a - bm + cm^2) \\ &= b^2 - 4bcm + 4c^2m^2 - 4ac + 4bcm - 4c^2m^2 = b^2 - 4ac = D, \end{aligned}$$

and we observe that $a' = c$ and that $b + b' = 2cm \equiv 0 \bmod 2c$. Consequently (a', b', c') is a form of discriminant D and is a right neighbor to (a, b, c). By the remarks before the theorem, (a, b, c) is properly equivalent to (a', b', c').

We repeat this process at least once, obtaining (a'', b'', c''). If $|a''| < |a'|$, we repeat it again, obtaining (a''', b''', c'''), and we continue in this way. Eventually the strict decrease of the magnitude of the first entry must stop. To keep the notation simple, we may assume without loss of generality that $|a''| \geq |a'|$. The claim is that (a', b', c') is then reduced.

Put $u = \sqrt{D} - b'$ and $v = b' - (\sqrt{D} - 2|a'|)$. The inequalities $(*)$ show that $u > 0$ and $v > 0$. Therefore

$$\begin{aligned} 0 < v^2 + 2uv + 2u\sqrt{D} &= (u + v)^2 - u^2 + 2u\sqrt{D} \\ &= 4a'^2 - (D - 2b'\sqrt{D} + b'^2) + 2D - 2b'\sqrt{D} \\ &= 4a'^2 + D - b'^2 = 4a'^2 - 4a'c'. \end{aligned}$$

Since $|c'| = |a''| \geq |a'|$, this inequality shows that $a'c' < 0$. Therefore $b'^2 = D + 4a'c' < D$, and $|b'| < \sqrt{D}$.

From $a'c' < 0$ and $|a'| \leq |c'|$, we see that $4|a'|^2 \leq 4|a'c'| = -4a'c' = D - b'^2 \leq D$. Therefore $2|a'| < \sqrt{D}$. The inequality $\sqrt{D} - 2|c| < b'$ implies that $\sqrt{D} - b' < 2|c| = 2|a'|$. The right side has just been shown to be $< \sqrt{D}$, and therefore $b' > 0$. Hence $\sqrt{D} - b' < 2|a'| < \sqrt{D} < \sqrt{D} + b'$. □

PROOF OF THEOREM 1.8b. Suppose that (a, b, c) is reduced and that (a', b', c') is a reduced neighbor on the right of (a, b, c). Then we must have $a' = c$ and $b + b' \equiv 0 \bmod 2c$. Since $D - b' < 2|a'|$ and $b' < \sqrt{D}$, we have $\sqrt{D} - 2|a'| < b' < \sqrt{D}$. That is, $\sqrt{D} - 2|c| < b' < \sqrt{D}$. These inequalities in combination with the congruence $b + b' \equiv 0 \bmod 2c$ show that (a, b, c) uniquely determines b'. Since (a', b', c') is to have discriminant D, c' is uniquely determined also.

We turn this construction around to prove existence of a right neighbor. Define (a', b', c') in terms of (a, b, c) as in the proof of Theorem 1.8a. Then $a' = c$, and b' is the unique integer such that $b + b' \equiv 0 \bmod 2c$ and

$$\sqrt{D} - 2|c| < b' < \sqrt{D}.$$

The form (a', b', c') is a right neighbor of (a, b, c), and we are to show that (a', b', c') is reduced.

Since (a, b, c) is reduced, we have $\sqrt{D} - b < 2|c| < \sqrt{D} + b$ and $b < \sqrt{D}$. Let m be the integer such that $b + b' = 2m|c|$. Addition of the inequalities $b' - (\sqrt{D} - 2|c|) > 0$ and $\sqrt{D} + b - 2|c| > 0$ gives $2m|c| = b + b' > 0$, and thus $m > 0$. Hence $m - 1 \geq 0$. Addition of the inequalities $\sqrt{D} - b > 0$ and $b' - (\sqrt{D} - 2|c|) > 0$ gives $0 < b' - b + 2|c| = 2b' - (b + b') + 2|c| = 2b' - 2(m - 1)|c|$. Hence $2b' > 2(m - 1)|c| \geq 0$, and we see that $b' > 0$. Therefore $0 < b' < \sqrt{D}$.

The definition of b' gives $\sqrt{D} - b' < 2|c| = 2|a'|$. Addition of the inequalities $2(m - 1)|c| \geq 0$ and $\sqrt{D} - b > 0$ gives $b + b' - 2|c| + \sqrt{D} - b > 0$, which says that $2|a'| < \sqrt{D} + b'$. Therefore (a', b', c') is reduced.

Let R be the operation of passing from a reduced form (a, b, c) to its unique reduced right neighbor (a', b', c'). What we have just shown implies that R acts as a permutation of the finite set of reduced forms of discriminant D. This set being finite, let n be the order of R. Then the set $\{R^k \mid 0 \leq k \leq n - 1\}$ is a cyclic group of permutations of the set of reduced forms of discriminant D. The existence of a two-sided inverse of R as a permutation implies that each reduced form of discriminant D has exactly one left neighbor. Thus the existence and uniqueness of neighbors on one side for reduced forms, in the presence of the finiteness of the set, implies existence and uniqueness on the other side. □

PROOF OF THEOREM 1.8c. We continue with R as the operation of passing from a reduced form to its unique reduced right neighbor, letting $\{R^k \mid 0 \le k \le n-1\}$ be the finite cyclic group of powers of R. This group acts on the set of reduced forms of discriminant D, and the cycles in question are the orbits under this action. To see that each orbit has an even number of members, we recall that a reduced form (a, b, c) has a and c of opposite sign. Thus if, for example, a is positive, then $R^l(a, b, c) = (a', b', c')$ has $(-1)^l a'$ positive. If the orbit of (a, b, c) has k members, then $R^k(a, b, c) = (a, b, c)$. Consequently $(-1)^k a$ has to have the same sign as a, and k has to be even. Finally the members of each orbit are properly equivalent to one another because, as we observed before the statement of the theorem, a form is properly equivalent to each of its neighbors. □

EXAMPLES WITH POSITIVE DISCRIMINANT, CONTINUED.

(1) $D = 5$. The forms with $D = 5$ satisfying the inequalities of Theorem 1.8a are $(1, 1, -1)$ and $(-1, 1, 1)$, and these consequently represent all proper equivalence classes. They form a single cycle and are properly equivalent by Theorem 1.8c. Thus again we obtain the easy conclusion that $h(5) = 1$.

(2) $D = 13$. The forms with $D = 13$ satisfying the inequalities of Theorem 1.8a are $(1, 3, -1)$ and $(-1, 3, 1)$, which make up a single cycle. Thus $h(13) = 1$.

(3) $D = 21$. The forms with $D = 21$ satisfying the inequalities of Theorem 1.8a are $(1, 3, -2)$ and $(-2, 3, 1)$, which make up one cycle, and $(-1, 3, 2)$ and $(2, 3, -1)$, which make up another cycle. Thus $h(21) = 2$.

4. Composition of Forms, Class Group

The identity $(x_1^2 + y_1^2)(x_2^2 + y_2^2) = (x_1x_2 - y_1y_2)^2 + (x_1y_2 + x_2y_1)^2$, which can be derived by factoring the left side in $\mathbb{Q}(\sqrt{-1}\,)[x_1, y_1, x_2, y_2]$ and rearranging the factors, readily generalizes to an identity involving any form $x^2 + bxy + cy^2$ of nonsquare discriminant $D = b^2 - 4c$. We complete the square, writing the form as $(x - \frac{1}{2}by)^2 - \frac{1}{4}y^2D$ and factoring it as $\big(x - \frac{1}{2}by + \frac{1}{2}y\sqrt{D}\,\big)\big(x - \frac{1}{2}by - \frac{1}{2}y\sqrt{D}\,\big)$, and we obtain

$$\begin{aligned}(x_1^2 + bx_1y_1 + cy_1^2)(x_2^2 + bx_2y_2 + cy_2^2)&\\ = (x_1x_2 - cy_1y_2)^2 + b(x_1x_2 - cy_1y_2)(x_1y_2 + x_2y_1 + by_1y_2)&\\ + c(x_1y_2 + x_2y_1 + by_1y_2)^2.&\end{aligned}$$

Improving on an earlier attempt by Legendre, Gauss made a thorough investigation of how one might multiply two distinct forms of the same nonsquare discriminant, not necessarily with first coefficient 1, and Dirichlet reworked the theory and simplified it. Out of this work comes the following **composition formula**, of which the above formula is manifestly a special case.

Proposition 1.9. Let (a_1, b, c_1) and (a_2, b, c_2) be two primitive forms with the same middle coefficient b and with the same nonsquare discriminant D, hence with $a_1c_1 = a_2c_2 \neq 0$. Suppose that $j = c_1a_2^{-1} = c_2a_1^{-1}$ is an integer. Then the form (a_1a_2, b, j) is primitive of discriminant D, and it has the property that

$$\begin{aligned}(a_1x_1^2 + bx_1y_1 + cy_1^2)&(a_2x_2^2 + bx_2y_2 + cy_2^2)\\ &= a_1a_2(x_1x_2 - jy_1y_2)^2 + b(x_1x_2 - jy_1y_2)(a_1x_1y_2 + a_2x_2y_1 + by_1y_2)\\ &\quad + j(a_1x_1y_2 + a_2x_2y_1 + by_1y_2)^2.\end{aligned}$$

REMARKS. Consequently if an integer m is represented by the form (a_1, b, c_1) and an integer n is represented by the form (a_2, b, c_2), then mn is represented by the form (a_1a_2, b, j). For example we saw in an example with $D = -20$ immediately following the statement of Theorem 1.6 that any prime that is congruent to 3 or 7 modulo 20 is representable as $2x^2+2xy+3y^2$. If we have two such primes p and q, then p is representable by $(2, 2, 3)$ and q is representable by $(3, 2, 2)$. The proposition is applicable with $j = 1$ and shows that pq is representable by $(6, 2, 1)$. In turn, substitution using $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) = \left(\begin{smallmatrix}1 & 0\\ -1 & 1\end{smallmatrix}\right)$ changes this form to the properly equivalent form $(5, 0, 1)$. Thus pq is representable as $x^2 + 5y^2$.

PROOF. The form (a_1a_2, b, j) is primitive because any prime that divides $\mathrm{GCD}(a_1a_2, b, j)$ has to divide either $\mathrm{GCD}(a_1, b, j)$ or $\mathrm{GCD}(a_2, b, j)$ and then certainly has to divide $\mathrm{GCD}(a_1, b, c_1)$ or $\mathrm{GCD}(a_2, b, c_2)$. No such prime exists, and hence (a_1a_2, b, j) is primitive. The discriminant of (a_1a_2, b, j) is $b^2 - 4ja_1a_2 = D + 4a_1c_1 - 4ja_1a_2 = D + 4a_1c_1 - 4(c_1a_2^{-1})a_1a_2 = D$, as asserted, and the verification of the displayed identity is a routine computation. □

Let us say that two primitive forms (a_1, b_1, c_1) and (a_2, b_2, c_2) of the same nonsquare discriminant are **aligned** if $b_1 = b_2$ and if $j = c_1a_2^{-1} = c_2a_1^{-1}$ is an integer. In the presence of equal nonsquare discriminants D and the equal middle entries b, the rational number j is automatically an integer if $\mathrm{GCD}(a_1, a_2) = 1$. In fact, the equality $D - b^2 = -4a_1c_1 = -4a_2c_2$ shows that $D - b^2$ is divisible by $4a_1$ and by $4a_2$; since $\mathrm{GCD}(a_1, a_2) = 1$, $D - b^2$ is divisible by $4a_1a_2$, and the quotient $-j$ is an integer.

The idea is that each pair of classes of properly equivalent primitive forms of discriminant D has a pair of aligned representatives, and a multiplication of proper equivalence classes is well defined if the product is defined as the class of the composition of these aligned representatives in the sense of Proposition 1.9. This multiplication for proper equivalence classes will make the set of classes into a finite abelian group. This group will be defined as the "form class group" for the discriminant D, except that we use only the positive definite classes in the

case that $D < 0$. Before phrasing these statements as a theorem, we make some remarks and then state and prove two lemmas.

Let (a, b, c) be a form of nonsquare discriminant D, and let b' be an integer with $b' \equiv b \bmod 2a$. In this case the number $c' = (b'^2 - D)/(4a)$ is an integer; in fact, we certainly have the congruences $b'^2 \equiv b^2 \bmod 2a$ and $b'^2 \equiv b^2 \bmod 4$, and thus we obtain the automatic[7] consequence $b'^2 \equiv b^2 \bmod 4a$, the rewritten congruence $b'^2 \equiv D + 4ac \bmod 4a$, and the desired result $b'^2 - D \equiv 0 \bmod 4a$. Hence (a, b', c') is another form of discriminant D. We call (a, b', c') a **translate** of (a, b, c). The key observation about translates is that the translate (a, b', c') is properly equivalent to (a, b, c). This fact follows from the computation

$$\begin{pmatrix} 1 & 0 \\ l & 1 \end{pmatrix} \begin{pmatrix} 2a & b \\ b & 2c \end{pmatrix} \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2a & b + 2al \\ b + 2al & 2(al^2 + bl + c) \end{pmatrix} = \begin{pmatrix} 2a & b' \\ b' & 2c' \end{pmatrix},$$

valid for any integer l.

Lemma 1.10. If (a, b, c) is a primitive form of nonsquare discriminant and if $m \neq 0$ is an integer, then (a, b, c) primitively represents some integer relatively prime to m.

PROOF. Let

$$\begin{aligned} w_0 &= \text{product of all primes dividing } a, c, \text{ and } m, \\ x_0 &= \text{product of all primes dividing } a \text{ and } m \text{ but not } c, \\ y_0 &= \text{product of all primes dividing } m \text{ but not } a. \end{aligned}$$

Referring to the definitions, we see that any prime dividing m divides exactly one of w_0, x_0, and y_0. In particular, $\mathrm{GCD}(x_0, y_0) = 1$. We shall show that $\mathrm{GCD}(m, ax_0^2 + bx_0y_0 + cy_0^2) = 1$, and the proof will be complete. Arguing by contradiction, suppose that a prime p divides $\mathrm{GCD}(m, ax_0^2 + bx_0y_0 + cy_0^2)$. There are three cases for p, as follows.

Case 1. If p divides x_0, then the fact that p divides $ax_0^2 + bx_0y_0 + cy_0^2$ implies that p divides cy_0^2. Since p does not divide y_0, p divides c, in contradiction to the definition of x_0.

Case 2. If p divides y_0, then similarly p divides ax_0^2. Since p does not divide x_0, p divides a, in contradiction to the definition of y_0.

Case 3. If p divides w_0, then the fact that p divides a and c implies that p divides bx_0y_0. Since p divides neither x_0 nor y_0, p divides b, in contradiction to the fact that (a, b, c) is primitive. □

[7]The argument being used here—that a congruence modulo $2a$ implies the congruence of the squares modulo $4a$—will be used again later in this section without detailed comment.

Lemma 1.11. Suppose that (a_1, b, c_1) and (a_2, b, c_2) are properly equivalent forms of nonsquare discriminant. If l is an integer such that $\mathrm{GCD}(a_1, a_2, l) = 1$ and such that l divides $\mathrm{GCD}(c_1, c_2)$, then $(la_1, b, l^{-1}c_1)$ and $(la_2, b, l^{-1}c_2)$ are properly equivalent forms.

REMARK. Even if (a_1, b, c_1) and (a_2, b, c_2) are primitive, it does not follow that $(la_1, b, l^{-1}c_1)$ and $(la_2, b, l^{-1}c_2)$ are primitive. In fact, one need only take $l = 2$ and $(a_1, b, c_1) = (a_2, b, c_2) = (1, 2, 4)$.

PROOF. Since (a_1, b, c_1) and (a_2, b, c_2) are properly equivalent, there exists $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ with

$$\begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \begin{pmatrix} 2a_1 & b \\ b & 2c_1 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} 2a_2 & b \\ b & 2c_2 \end{pmatrix}.$$

We multiply both sides on the right by $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)^{-1}$, and the result is the system of four scalar equations

$$\begin{aligned} 2a_1\alpha + b\gamma &= 2a_2\delta - b\gamma, \\ 2a_1\beta + b\delta &= b\delta - 2c_2\gamma, \\ b\alpha + 2c_1\gamma &= -2a_2\beta + b\alpha, \\ b\beta + 2c_1\delta &= -b\beta + 2c_2\alpha. \end{aligned}$$

The second and third equations simplify to $a_1\beta + c_2\gamma = 0$ and $a_2\beta + c_1\gamma = 0$. Since l divides c_1 and c_2, these two simplified equations show that l divides $a_1\beta$ and $a_2\beta$. Since $\mathrm{GCD}(a_1, a_2, l) = 1$, it follows that l divides β.

Therefore the matrix $\left(\begin{smallmatrix} \alpha & l^{-1}\beta \\ l\gamma & \delta \end{smallmatrix}\right)$ of determinant 1 has integer entries. Direct computation shows that

$$\begin{pmatrix} \alpha & l\gamma \\ l^{-1}\beta & \delta \end{pmatrix} \begin{pmatrix} 2la_1 & b \\ b & 2l^{-1}c_1 \end{pmatrix} \begin{pmatrix} \alpha & l^{-1}\beta \\ l\gamma & \delta \end{pmatrix} = \begin{pmatrix} 2la_2 & b \\ b & 2l^{-1}c_2 \end{pmatrix}.$$

Consequently the forms $(la_1, b, l^{-1}c_1)$ and $(la_2, b, l^{-1}c_2)$ are properly equivalent. □

Theorem 1.12. Let D be a nonsquare discriminant, and let $\mathcal{C}_1$ and $\mathcal{C}_2$ be proper equivalence classes of primitive forms of discriminant D.

(a) There exist aligned forms $(a_1, b, c_1) \in \mathcal{C}_1$ and $(a_2, b, c_2) \in \mathcal{C}_2$, and these may be chosen in such a way that a_1 and a_2 are relatively prime to each other and to any integer $m \neq 0$ given in advance.

(b) If the product of $\mathcal{C}_1$ and $\mathcal{C}_2$ is defined to be the proper equivalence class of the composition of any aligned representatives of $\mathcal{C}_1$ and $\mathcal{C}_2$, as for example the ones in (a), then the resulting product operation is well defined on proper equivalence classes of primitive forms of discriminant D.

(c) Under the product operation in (b), the set of proper equivalence classes of primitive forms of discriminant D is a finite abelian group. The identity is the class of $(1, 0, -D/4)$ if $D \equiv 0 \bmod 4$ and is the class of $(1, 1, -(D-1)/4)$ if $D \equiv 1 \bmod 4$. The group inverse of the class of (a, b, c) is the class of $(a, -b, c)$.

REMARK. When $D < 0$, the proper equivalence classes of positive definite forms are a subgroup. In fact, if (a_1, b, c_1) and (a_2, b, c_2) are positive definite and are aligned, then a_1 and a_2 are positive, and therefore their composition (a_1a_2, b, j) has a_1a_2 positive and is positive definite. As was indicated in the discussion before Lemma 1.10, the **form class group** for discriminant D is defined to be the group in (c) if $D > 0$, and it is defined to be the subgroup of classes of positive definite forms if $D < 0$.

PROOF OF THEOREM 1.12a. By two applications of Lemma 1.10, $\mathcal{C}_1$ primitively represents some integer a_1 prime to m, and $\mathcal{C}_2$ primitively represents some integer a_2 prime to a_1m. Arguing as in the last part of the proof of Theorem 1.6b, we may assume without loss of generality that $(x, y) = (1, 0)$ yields these values in each case. Then $\mathcal{C}_1$ contains a form $(a_1, b_1, *)$ for some b_1, and $\mathcal{C}_2$ contains a form $(a_2, b_2, *)$ for some b_2. By the remarks before Lemma 1.10, $\mathcal{C}_1$ contains every translate $(a_1, b_1 + 2a_1l_1, *)$, and $\mathcal{C}_2$ contains every translate $(a_2, b_2 + 2a_2l_2, *)$.

Let us make specific choices of l_1 and l_2. We know that $b_1 \equiv D \equiv b_2 \bmod 2$, so that $b_2 - b_1$ is even. The construction of a_1 and a_2 was arranged to make $\mathrm{GCD}(a_1, a_2) = 1$, and therefore $\mathrm{GCD}(2a_1, 2a_2) = 2$. Since $b_2 - b_1$ is even, we can choose l_1 and l_2 such that $2a_1l_1 - 2a_2l_2 = b_2 - b_1$. Then $b_1 + 2a_1l_1 = b_2 + 2a_2l_2$, and we take the common value as b.

For this b, $\mathcal{C}_1$ contains the form $(a_1, b, *)$, and $\mathcal{C}_2$ contains the form $(a_2, b, *)$. Since we have arranged that $\mathrm{GCD}(a_1, a_2) = 1$, the remark immediately following the definition of "aligned" shows that these forms are aligned. □

PROOF OF THEOREM 1.12b. Suppose that

$$\begin{aligned} (a_1', b', *) &\quad \text{is properly equivalent to} \quad (a_1'', b'', *), \\ (a_2', b', *) &\quad \text{is properly equivalent to} \quad (a_2'', b'', *), \end{aligned}$$

with the vertical pairs aligned. We are to show that

$$(a_1'a_2', b', *) \quad \text{is properly equivalent to} \quad (a_1''a_2'', b'', *). \tag{$*$}$$

Theorem 1.12a applied to the integer $m = a_1'a_2'a_1''a_2''$ gives us an aligned pair of forms $(a_1, b, *)$ and $(a_2, b, *)$ in the respective proper equivalence classes such

that $\mathrm{GCD}(a_1, a_2) = 1$ and $\mathrm{GCD}(a_1a_2, m) = 1$. If we can show that

$$(a_1'a_2', b', *) \quad \text{is properly equivalent to} \quad (a_1a_2, b, *), \tag{$**$}$$

then we will have symmetrically that

$$(a_1''a_2'', b'', *) \quad \text{is properly equivalent to} \quad (a_1a_2, b, *),$$

and $(*)$ will follow from this fact and $(**)$ by transitivity of proper equivalence.

We can now argue as in the proof of Theorem 1.12a. We know that $b \equiv D \equiv b' \bmod 2$, so that $b' - b$ is even. The construction of a_1 and a_2 was arranged to make $\mathrm{GCD}(a_1a_2, a_1'a_2') = 1$, and therefore $\mathrm{GCD}(2a_1a_2, 2a_1'a_2') = 2$. Since $b_2 - b_1$ is even, we can choose l and l' such that $2a_1a_2l - 2a_1'a_2'l' = b' - b$. Then $b + 2a_1a_2l = b' + 2a_1'a_2'l'$, and we take the common value as B. This B has

$$B \equiv b \bmod 2a_1a_2 \qquad \text{and} \qquad B \equiv b' \bmod 2a_1'a_2'.$$

Thus

$$\begin{aligned} (a_1, b, *) &\quad \text{is properly equivalent to} \quad (a_1, B, *), \\ (a_2, b, *) &\quad \text{is properly equivalent to} \quad (a_2, B, *), \\ (a_1a_2, b, *) &\quad \text{is properly equivalent to} \quad (a_1a_2, B, *), \end{aligned} \tag{$\dagger$}$$

and similarly

$$\begin{aligned} (a_1', b', *) &\quad \text{is properly equivalent to} \quad (a_1', B, *), \\ (a_2', b', *) &\quad \text{is properly equivalent to} \quad (a_2', B, *), \\ (a_1'a_2', b', *) &\quad \text{is properly equivalent to} \quad (a_1'a_2', B, *). \end{aligned} \tag{$\dagger\dagger$}$$

By construction of b, $(a_1, b, *)$ is properly equivalent to $(a_1', b', *)$. This equivalence, in combination with the first line of $(\dagger)$ and the first line of $(\dagger\dagger)$, shows that

$$(a_1, B, *) \quad \text{is properly equivalent to} \quad (a_1', B, *). \tag{$\ddagger$}$$

Let us check that Lemma 1.11 is applicable to the two properly equivalent forms of $(\ddagger)$ and to the integer $l = a_2'$. In fact, $\mathrm{GCD}(a_1, a_2, l) = 1$ follows from $\mathrm{GCD}(a_1a_2, a_1'a_2') = 1$, and the problem is to show that $l = a_2'$ divides $(D - B^2)/(4a_1)$ and $(D - B^2)/(4a_1')$. To see this divisibility, we observe that $D - b'^2$ is divisible by $4a_1'a_2'$ because $(a_1', b', *)$ and $(a_2', b', *)$ are given as aligned; the congruence $b' \equiv B \bmod 2a_1'a_2'$ implies that $b'^2 \equiv B^2 \bmod 4a_1'a_2'$, and addition gives $D - B^2 \equiv 0 \bmod 4a_1'a_2'$. Meanwhile, $D - B^2$ is divisible by $4a_1$ because the third member of $(a_1, B, *)$ is an integer. Since $D - B^2$ is divisible also by $4a_1'a_2'$ and since $\mathrm{GCD}(a_1, a_1'a_2') = 1$, $D - B^2$ is divisible by

$4a_1a_1'a_2'$. Therefore $(D - B^2)/(4a_1)$ and $(D - B^2)/(4a_1')$ are divisible by a_2', and Lemma 1.11 is indeed applicable.

The application of Lemma 1.11 to (‡) with $l = a_2'$ shows that

$$(a_1a_2', B, *) \quad \text{is properly equivalent to} \quad (a_1'a_2', B, *).$$

Similarly $(a_2, B, *)$ is properly equivalent to $(a_2', B, *)$, and an application of Lemma 1.11 to this equivalence with $l = a_1$ shows that

$$(a_1a_2, B, *) \quad \text{is properly equivalent to} \quad (a_1a_2', B, *).$$

The two results together show that

$$(a_1a_2, B, *) \quad \text{is properly equivalent to} \quad (a_1'a_2', B, *).$$

Combining this equivalence with the third line of (†) and the third line of (††), we obtain (∗∗), and the proof of (b) is complete. □

PROOF OF THEOREM 1.12c. The set of proper equivalence classes is finite by Theorem 1.6a, and commutativity of multiplication is clear. Define δ to be 0 if $D \equiv 0 \bmod 4$ and to be 1 if $D \equiv 1 \bmod 4$. Let us see that the class of $(1, \delta, *)$ is the identity. If (a, b, c) has discriminant D, then $b \equiv \delta \bmod 2$, and hence $(1, b, *) = (1, \delta + 2 \cdot 1 \cdot \frac{1}{2}(b - \delta))$ is a translate of $(1, \delta, *)$. Consequently $(1, b, *)$ and $(1, \delta, *)$ are properly equivalent. Since Proposition 1.9 shows that the composition of (a, b, c) and $(1, b, *)$ is $(a, b, *)$, Theorem 1.12b allows us to conclude that the class of $(1, \delta, *)$ is the identity.

For inverses Theorem 1.12b shows that the product of the classes of (a, b, c) and $(a, -b, c)$ is the product of the classes of (a, b, c) and (c, b, a), which is the class of the composition $(a, b, c)(c, b, a)$. Proposition 1.9 shows that this composition is $(ac, b, 1)$. Since $(ac, b, 1)$ is properly equivalent to $(1, -b, ac)$ and since the latter is properly equivalent to $(1, \delta, *)$, the class of the composition $(a, b, c)(c, b, a)$ is the identity.

To complete the proof, we need to verify associativity. Let $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$ be three proper equivalence classes of primitive forms of discriminant D. Let (a_1, b_1, c_1) be a form in the class $\mathcal{C}_1$. Lemma 1.10 shows that $\mathcal{C}_2$ represents an integer a_2 prime to a_1, and then it follows that the form (a_2, b_2, c_2) is in $\mathcal{C}_2$ for some integers b_2 and c_2. A second application of Lemma 1.10 shows that $\mathcal{C}_3$ represents an integer a_3 prime to a_1a_2, and then it follows that the form (a_3, b_3, c_3) is in $\mathcal{C}_3$ for some integers b_3 and c_3. The middle components have $b_1 \equiv b_2 \equiv b_3 \equiv \delta \bmod 2$, and thus $\frac{1}{2}(b_j - \delta)$ is an integer for $j = 1, 2, 3$. Since a_1, a_2, a_3 are relatively prime in pairs, the Chinese Remainder Theorem shows that the congruences $x \equiv \frac{1}{2}(b_j - \delta) \bmod a_j$ have a common integer solution x for $j = 1, 2, 3$. Define

$b = 2x + \delta$. Then b is a solution of $b \equiv b_j \bmod 2a_j$ for $j = 1, 2, 3$. Write $b = b_j + 2a_j n_j$ for suitable integers n_j. Then $(a_j, b, *) = (a_j, b_j + 2a_j n_j, *)$ is a translate of (a_j, b_j, c_j) and consequently is properly equivalent to it. Thus $(a_j, b, *)$ lies in $\mathcal{C}_j$. Taking into account Theorem 1.12b and using Proposition 1.9, we see that $\mathcal{C}_1(\mathcal{C}_2\mathcal{C}_3)$ and $(\mathcal{C}_1\mathcal{C}_2)\mathcal{C}_3$ are both represented by the form $(a_1a_2a_3, b, *)$ and hence are equal. □

5. Genera

The theory of genera lumps proper equivalence classes of forms of a given discriminant according to their values in some way. There are at least two possible definitions of "genus," and it is a deep result that they lead to the same thing in all cases of interest. By way of background, we saw in Sections 2 and 3 for discriminant $D = -56$ that the number of proper equivalence classes of binary quadratic forms is exactly 4, representatives being $x^2 + 14y^2$, $2x^2 + 7y^2$, and $3x^2 \pm 2xy + 5y^2$. The last two are improperly equivalent and take the same values at integer points (x, y), and there are no other improper equivalences. Thus the first two take on a disjoint set of prime values from the values of $3x^2 \pm 2xy + 5y^2$ for integer points (x, y), and the sets of prime values taken on by $x^2 + 14y^2$ and $2x^2 + 7y^2$ at integer points are disjoint from one another.

Two possible lumpings of proper equivalence classes arise for this discriminant. One is to identify forms when their values modulo 56 include the same residues prime to 56. It is just a finite computation to see that

$x^2 + 14y^2$ and $2x^2 + 7y^2$	take on the residues	1, 9, 15, 23, 25, 39,
$3x^2 \pm 2xy + 5y^2$	take on the residues	3, 5, 13, 19, 27, 45.

Thus the first kind of lumping treats $x^2 + 14y^2$ and $2x^2 + 7y^2$ together because of the residues they take on, and it treats $3x^2 + 2xy + 5y^2$ and $3x^2 - 2xy + 5y^2$ together. Gauss proceeded by using this kind of lumping to define "genus."

The other lumping is to identify integer forms that take on the same *rational* values at rational points. Here $2x^2 + 7y^2 = 1$ for $(x, y) = (\frac{1}{3}, \frac{1}{3})$, and of course $x^2 + 14y^2 = 1$ for $(x, y) = (1, 0)$. Hence the sets of values of $x^2 + 14y^2$ and $2x^2 + 7y^2$ for x and y rational have a nonzero value in common. Lemma 1.13 below implies that the sets of rational values taken on by the two forms are identical. The second kind of lumping treats $x^2 + 14y^2$ and $2x^2 + 7y^2$ together because they take on the same rational values. We shall use this latter kind of lumping because, as Theorem 1.14 below shows, this is the definition that more quickly identifies the genus group once the form class group is known.

Problems 25–40 at the end of the chapter show that the two definitions of genus lead to the same thing for discriminants that are "fundamental" in a sense that we define in a moment.

We have defined two forms (a, b, c) and (a', b', c') with integer entries to be "properly equivalent" if there is a matrix $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ in $\mathrm{SL}(2, \mathbb{Z})$ with

$$\begin{pmatrix}\alpha & \gamma\\ \beta & \delta\end{pmatrix}\begin{pmatrix}2a & b\\ b & 2c\end{pmatrix}\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix} = \begin{pmatrix}2a' & b'\\ b' & 2c'\end{pmatrix}.$$

We say that two forms (a, b, c) and (a', b', c') with rational entries are **properly equivalent over** $\mathbb{Q}$ if there is a matrix $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ in $\mathrm{SL}(2, \mathbb{Q})$ such that the displayed equality holds. For emphasis we can refer to the original notion as "proper equivalence over $\mathbb{Z}$" when it is advisable to be more specific. It is evident that if two forms with rational entries are properly equivalent over $\mathbb{Q}$, then their sets of values at points (x, y) in $\mathbb{Q} \times \mathbb{Q}$ are the same.

Lemma 1.13. If (a, b, c) is a form with rational coefficients and with nonsquare discriminant D that takes on a nonzero value $q \in \mathbb{Q}$ for some (x_0, y_0) in $\mathbb{Q} \times \mathbb{Q}$, then (a, b, c) is properly equivalent over $\mathbb{Q}$ to $(q, 0, -D/(4q))$. Consequently two forms over $\mathbb{Q}$ of the same discriminant that take on a nonzero value in common over $\mathbb{Q}$ are properly equivalent over $\mathbb{Q}$.

PROOF. Suppose that $ax_0^2 + bx_0y_0 + cy_0^2 = q$. Put $\left(\begin{smallmatrix}\alpha\\ \gamma\end{smallmatrix}\right) = \left(\begin{smallmatrix}x_0\\ y_0\end{smallmatrix}\right)$. Since x_0 and y_0 cannot both be 0, we can choose rationals β and δ such that $\alpha\delta - \beta\gamma = 1$. Then $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ has determinant 1 and satisfies $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)\left(\begin{smallmatrix}1\\ 0\end{smallmatrix}\right) = \left(\begin{smallmatrix}x_0\\ y_0\end{smallmatrix}\right)$. The equality $ax_0^2 + bx_0y_0 + cy_0^2 = \frac{1}{2}(x_0 \;\; y_0)\left(\begin{smallmatrix}2a & b\\ b & 2c\end{smallmatrix}\right)\left(\begin{smallmatrix}x_0\\ y_0\end{smallmatrix}\right)$ therefore yields

$$q = \tfrac{1}{2}\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}\alpha & \gamma\\ \beta & \delta\end{pmatrix}\begin{pmatrix}2a & b\\ b & 2c\end{pmatrix}\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\begin{pmatrix}1\\ 0\end{pmatrix}.$$

It follows that (a, b, c) is properly equivalent over $\mathbb{Q}$ to some form (q, b', c') with b' and c' rational. Using a translation with a rational parameter, we see that (q, b', c') is properly equivalent over $\mathbb{Q}$ to a form $(q, 0, *)$. Inspection of the discriminant shows that this last form must be $(q, 0, -D/(4q))$. □

Two primitive integer forms having the same discriminant are said to be in the same **genus** (plural: *genera*) if they are properly equivalent over $\mathbb{Q}$. In view of Lemma 1.13 the condition is that they are primitive and take on a common nonzero value over $\mathbb{Q}$, or equivalently that they are primitive and take on the same set of values over $\mathbb{Q}$. Thus $x^2 + 14y^2$ and $2x^2 + 7y^2$ furnish an example of two

forms in distinct classes that are in the same genus. Two primitive integer forms that are in the same proper equivalence class over $\mathbb{Z}$ are in the same genus. The genus of the class $\mathcal{C}$ will be denoted by $[\mathcal{C}]$. The identity class will be denoted by $\mathcal{E}$, and $P = [\mathcal{E}]$ is called the **principal genus**. If (a, b, c) is an integer form representing a class $\mathcal{C}$, then Theorem 1.12c shows that $(a, -b, c)$ represents $\mathcal{C}^{-1}$. On the other hand, $\mathcal{C}$ and $\mathcal{C}^{-1}$ take on the same values over $\mathbb{Z}$, as we see by replacing (x, y) by $(x, -y)$, and it follows that $[\mathcal{C}] = [\mathcal{C}^{-1}]$.

For the main theorem about genera, we shall introduce an extra hypothesis on the discriminant D. A nonsquare integer D will be said to be a **fundamental discriminant** if D is not divisible by the square of any odd prime and if when D is even, $D/4$ is congruent to 2 or 3 modulo 4. It will be seen later that this condition is equivalent to the requirement that D be the "field discriminant" of some quadratic number field. Examples of discriminants that are not fundamental are $D = -12, -44, -108$.

With this condition imposed on D, any integer form (a, b, c) of discriminant D is automatically primitive. In fact, no odd prime p can divide $\text{GCD}(a, b, c)$, since then p^2 would divide D. If 2 were to divide $\text{GCD}(a, b, c)$, then $(a/2, b/2, c/2)$ would be an integer form, and $D/4 = (b/2)^2 - 4(a/2)(c/2)$ would be an integer congruent to 1 or 4 modulo 4.

Theorem 1.14. For a fundamental discriminant D, the principal genus P of primitive integer forms[8] is a subgroup of the form class group H, and the cosets of P are the various genera. Thus the set G of genera is exactly the set of cosets H/P and inherits a group structure from class multiplication. The subgroup P coincides with the subgroup of squares in H, and consequently every nontrivial element of G has order 2.

REMARKS. The group G is called the **genus group** of discriminant D. The hypothesis that D is fundamental is needed only for the conclusion that every member of P is a square in H. Since every nontrivial element of G has order 2 when D is fundamental, application of the Fundamental Theorem of Finitely Generated Abelian Groups or use of vector-space theory over a 2-element field shows that G is the direct sum of cyclic groups of order 2; in particular, the order of G is a power of 2. Problems 25–29 at the end of the chapter show that the order of G is 2^g, where $g + 1$ is the number of distinct prime factors of D.

PROOF. Let $V(\mathcal{C})$ denote the set of $\mathbb{Q}$ values assumed by forms in the class $\mathcal{C}$ at points (x, y) in $\mathbb{Q} \times \mathbb{Q}$. If S and S' are two genera and if $\mathcal{C}$ is a class in S and $\mathcal{C}'$ is a class in S', we define $S \cdot S' = [\mathcal{C}\mathcal{C}']$.

[8] As usual, we exclude the negative definite classes in the discussion.

To see that this product operation is well defined on the set G of genera, let $\mathcal{C}''$ be in S' also. Then $V(\mathcal{C}') = V(\mathcal{C}'')$. If q is in $V(\mathcal{C})$ and q' is in $V(\mathcal{C}') = V(\mathcal{C}'')$, then the prescription for multiplying classes shows that qq' is in $V(\mathcal{C}\mathcal{C}')$ and $V(\mathcal{C}\mathcal{C}'')$. Hence $V(\mathcal{C}\mathcal{C}') = V(\mathcal{C}\mathcal{C}'')$, and $[\mathcal{C}\mathcal{C}'] = [\mathcal{C}\mathcal{C}'']$. Therefore multiplication of genera is well defined. Define a function $\varphi : H \to G$ by $\varphi(\mathcal{C}) = [\mathcal{C}]$. Then the computation

$$\varphi(\mathcal{C}\mathcal{C}') = [\mathcal{C}\mathcal{C}'] = [\mathcal{C}][\mathcal{C}'] = \varphi(\mathcal{C})\varphi(\mathcal{C}')$$

shows that φ is a homomorphism of H onto G. The kernel of φ is $[\mathcal{C}] = P$, which is therefore a subgroup, and the image of φ, which is the set G of genera with its product operation, has to be a group.

For any class $\mathcal{C}$, the equality $[\mathcal{C}] = [\mathcal{C}^{-1}]$ implies that $[\mathcal{C}^2] = [\mathcal{C}][\mathcal{C}] = [\mathcal{C}][\mathcal{C}^{-1}] = [\mathcal{C}\mathcal{C}^{-1}] = [\mathcal{E}] = P$. Hence P contains all squares. Conversely let $\mathcal{C}$ be in P. Then $\mathcal{C}$ takes on the value 1 over $\mathbb{Q}$. If (a, b, c) is a form in the class $\mathcal{C}$, then there exist rationals r and s with $ar^2 + brs + cs^2 = 1$. Clearing fractions, we see that there exist integers x and y such that $ax^2 + bxy + cy^2 = n^2$ for some integer $n \neq 0$. Without loss of generality, we may assume that n is positive. Since (a, b, c) is primitive, a familiar argument allows us to make a substitution for which the value n^2 is taken on at $(x, y) = (1, 0)$. In other words, (a, b, c) is properly equivalent over $\mathbb{Z}$ to a form (n^2, b', c') for suitable integers b' and c'. The composition formula in Proposition 1.9 shows that the composition of $(n, b', c'n)$ with itself is (n^2, b', c'), and hence $\mathcal{C}$ is exhibited as the square of the class of $(n, b', c'n)$. Since $(n, b', c'n)$ has the same discriminant D as (n^2, b', c') and therefore as (a, b, c) and since D is fundamental, $(n, b', c'n)$ is primitive. Therefore $\mathcal{C}$ is the square of a class of primitive forms. If $\mathcal{C}$ is positive definite, then the above choice of the sign of n as positive makes $(n, b', c'n)$ positive definite. Hence the class of $(n, b', c'n)$ is in H. □

EXAMPLE. The discriminant $D = -56$ is fundamental, and we have seen that the form class group is of order 4 with representatives $x^2 + 14y^2$, $2x^2 + 7y^2$, and $3x^2 \pm 2xy + 5y^2$. We have seen also that $x^2 + 14y^2$ and $2x^2 + 7y^2$ both lie in the principal genus P. A group of order 4 must be isomorphic to the cyclic group C_4 or to $C_2 \times C_2$. In the first case the subgroup of squares has order 2, and in the second case the subgroup of squares has order 1. Since we have already found two elements in P, P has order exactly 2. By the theorem we must be in the first case. Hence H is of type C_4, and the genus group G is of type C_2. It is possible to check directly that $3x^2 + 2xy + 5y^2$ has order 4 by making computations similar to those for Problem 4d at the end of the chapter.

6. Quadratic Number Fields and Their Units

In this section we review material about quadratic number fields that appears in various places in *Basic Algebra*, and we determine the units in the ring of integers of such a number field.

Quadratic number fields are extension fields K of $\mathbb{Q}$ with $[K : \mathbb{Q}] = 2$. Such a field is necessarily of the form $K = \mathbb{Q}(\sqrt{m})$, where m is a uniquely determined square-free integer not equal to 0 or 1. The set $\{1, \sqrt{m}\}$ is a vector-space basis of K over $\mathbb{Q}$.

The extension $K/\mathbb{Q}$ is a Galois extension, and the Galois group $\mathrm{Gal}(K/\mathbb{Q})$ of automorphisms of K fixing $\mathbb{Q}$ has two elements. We denote the nontrivial element of the Galois group by σ; its values on the members of the vector-space basis are $\sigma(1) = 1$ and $\sigma(\sqrt{m}) = -\sqrt{m}$.

The norm $N = N_{K/\mathbb{Q}}$ and trace $\mathrm{Tr} = \mathrm{Tr}_{K/\mathbb{Q}}$ are given by $N(\alpha) = \alpha \cdot \sigma(\alpha)$ and $\mathrm{Tr}(\alpha) = \alpha + \sigma(\alpha)$. Thus $N(a + b\sqrt{m}) = a^2 - mb^2$ and $\mathrm{Tr}(a + b\sqrt{m}) = 2a$. These values are members of $\mathbb{Q}$. The norm is multiplicative in the sense that $N(\alpha\beta) = N(\alpha)N(\beta)$, and $N(1) = 1$.

The ring R of algebraic integers in K is the integral closure of $\mathbb{Z}$ in K. It works out to be

$$R = \begin{cases} \mathbb{Z}[\sqrt{m}\,] & \text{if } m \equiv 2 \text{ or } 3 \bmod 4, \\ \mathbb{Z}[\frac{1}{2}(\sqrt{m} - 1)] & \text{if } m \equiv 1 \bmod 4 \end{cases}$$

and is therefore a free abelian group of rank 2. The automorphism σ carries R to itself. The norm and trace of any member of R are in $\mathbb{Z}$; conversely any member of K whose norm and trace are in $\mathbb{Z}$ is in R. We define the algebraic integer δ to be given by

$$\delta = \begin{cases} -\sqrt{m} & \text{if } m \equiv 2 \text{ or } 3 \bmod 4, \\ \frac{1}{2}(1 - \sqrt{m}\,) & \text{if } m \equiv 1 \bmod 4. \end{cases}$$

Then $\{1, \delta\}$ is a $\mathbb{Z}$ basis of R. The norm and trace of δ are given by

$$N(\delta) = \delta \cdot \sigma(\delta) = \begin{cases} -m & \text{if } m \equiv 2 \text{ or } 3 \bmod 4, \\ \frac{1}{4}(1 - m) & \text{if } m \equiv 1 \bmod 4, \end{cases}$$

$$\mathrm{Tr}(\delta) = \delta + \sigma(\delta) = \begin{cases} 0 & \text{if } m \equiv 2 \text{ or } 3 \bmod 4, \\ 1 & \text{if } m \equiv 1 \bmod 4. \end{cases}$$

There is a general notion of **field discriminant** D, or **absolute discriminant**, for an algebraic number field, whose definition will be given in Chapter V. We shall not give that definition in general now but will be content to give the formula for D in the quadratic number field $\mathbb{Q}(\sqrt{m}\,)$, namely

$$D = \begin{cases} 4m & \text{if } m \equiv 2 \text{ or } 3 \bmod 4, \\ m & \text{if } m \equiv 1 \bmod 4. \end{cases}$$

The **units** of K are understood to be the members of the group $R^\times$ of units in the ring R. These are the members ε of R with $N(\varepsilon) = \pm 1$. In fact, if ε is a unit, then the equality $\varepsilon\varepsilon^{-1} = 1$ implies that $1 = N(1) = N(\varepsilon\varepsilon^{-1}) = N(\varepsilon)N(\varepsilon^{-1})$ and shows that $N(\varepsilon)$ is a unit in $\mathbb{Z}$. Thus $N(\varepsilon) = \pm 1$. Conversely if $N(\varepsilon) = \pm 1$, then $\pm\varepsilon\sigma(\varepsilon) = 1$ shows that $\sigma(\varepsilon) = \pm\varepsilon^{-1}$; since $\sigma(\varepsilon)$ is in R, ε is exhibited as in $R^\times$ and is therefore a unit.

For $m < 0$, the units of $\mathbb{Q}(\sqrt{m}\,)$ are easily determined. In fact, if $\varepsilon = a + b\delta$ with a and b in $\mathbb{Z}$, then $N(\varepsilon) = (a + b\delta)(a + b\sigma(\delta)) = a^2 + b\,\mathrm{Tr}\,\delta + b^2N(\delta)$ with each term equal to an integer and with the end terms ≥ 0. Sorting out the possibilities, we see that

$$R^\times = \begin{cases} \{\pm 1, \pm\sqrt{-1}\} & \text{if } m = -1, \\ \{\pm 1, \frac{1}{2}(\pm 1 \pm \sqrt{-3}\,)\} & \text{if } m = -3, \\ \{\pm 1\} & \text{for all other } m < 0. \end{cases}$$

The respective orders of $R^\times$ are 4, 6, and 2.

Determination of the units when $m > 0$ is more delicate. We require a lemma.

Lemma 1.15. If α is a real irrational number and if $N > 0$ is an integer, then there exist integers A and B with

$$|B\alpha - A| < \frac{1}{N} \qquad \text{and} \qquad 0 < B \leq N.$$

For this A and this B,

$$\left|\alpha - \frac{A}{B}\right| < \frac{1}{B^2}.$$

PROOF. Put $\alpha_n = n\alpha - [n\alpha]$, where $[\,\cdot\,]$ denotes the greatest-integer function. Then $0 \leq \alpha_n < 1$. We partition the half-open interval $[0, 1)$ into N subintervals $\left[\frac{t-1}{N}, \frac{t}{N}\right)$ with $1 \leq t \leq N$. For $0 \leq n \leq N$, the expression α_n takes on $N + 1$ distinct values because $\alpha_n = \alpha_m$ would imply that $(n - m)\alpha$ is in $\mathbb{Z}$. Hence there exist α_n and α_m with $n > m$ that lie in the same subinterval $\left[\frac{t-1}{N}, \frac{t}{N}\right)$. Then $|\alpha_n - \alpha_m| < \frac{1}{N}$. If we take $B = n - m$ and $A = [n\alpha] - [m\alpha]$, then $|B\alpha - A| = |\alpha_n - \alpha_m|$, and the inequality $|B\alpha - A| < \frac{1}{N}$ follows. Dividing this inequality by B gives $|\alpha - \frac{A}{B}| < \frac{1}{BN}$, and this is $\leq \frac{1}{B^2}$ because $N \geq B$. □

Proposition 1.16. For $K = \mathbb{Q}(\sqrt{m}\,)$ with $m > 0$, the units are the members of the infinite group

$$R^\times = \{(\pm 1)\varepsilon_1^n \mid n \in \mathbb{Z}\} \cong \mathbb{Z} \times C_2,$$

where ε_1 is the **fundamental unit**, defined as the least unit > 1.

REMARK. For example, when $m = 2$, the fundamental unit is $\varepsilon_1 = 1 + \sqrt{2}$.

PROOF. The units ω with $|\omega| = 1$ are ± 1, since the members of K are real numbers. We shall show shortly that there exists a unit ω with $|\omega| \neq 1$. Then ω or ω^{-1} has absolute value > 1. Let us say that $|\omega| > 1$. Then one of ω and $-\omega$ is > 1. Let us say that $\omega > 1$. Write $\omega = a + b\sqrt{m}$, so that $\sigma(\omega) = a - b\sqrt{m} = \pm\omega^{-1}$ has $|\sigma(\omega)| < 1$. Then

$$|2a| = |\omega + \sigma(\omega)| \leq |\omega| + |\sigma(\omega)| \leq |\omega| + 1$$

and

$$|2b\sqrt{m}| = |\omega - \sigma(\omega)| \leq |\omega| + |\sigma(\omega)| \leq |\omega| + 1$$

together show that there are only finitely many units ω' with $1 < |\omega'| < |\omega|$. Hence the existence of a unit ω with $|\omega| \neq 1$ implies the existence of a fundamental unit ε_1.

If ω' is any unit > 1, then we can choose a power ε_1^n of ε_1 with $\varepsilon_1^{n+1} > \omega' \geq \varepsilon_1^n$, by the archimedean property of $\mathbb{R}$. Then $\omega'\varepsilon_1^{-n}$ is a unit ≥ 1 with $|\omega'\varepsilon_1^{-n}| < \varepsilon_1$. Since ε_1 is fundamental, $\omega'\varepsilon_1^{-n}$ is 1, and thus $\omega' = \varepsilon_1^n$. Then it follows that the group of units has the asserted form.

Thus we need to exhibit some unit ω with $|\omega| \neq 1$. We apply Lemma 1.15 with $\alpha = \sqrt{m}$ and with N arbitrary. Then we obtain infinitely many pairs (A, B) of integers with $\left|\sqrt{m} - \frac{A}{B}\right| < \frac{1}{B^2} \leq 1$, hence with $|A/B| < 1 + \sqrt{m}$. For each such pair (A, B), the member $r = A - B\sqrt{m}$ of R has

$$\begin{aligned} |N(r)| &= \left|(A + B\sqrt{m})(A - B\sqrt{m})\right| = \left|\tfrac{A}{B} - \sqrt{m}\right| \left|B^2\right| \left|\tfrac{A}{B} + \sqrt{m}\right| \\ &\leq \tfrac{1}{B^2}\, B^2(1 + 2\sqrt{m}) = 1 + 2\sqrt{m}. \end{aligned}$$

Thus there are infinitely many r in R with $|N(r)| \leq 1 + 2\sqrt{m}$. Since the norm of an algebraic integer is in $\mathbb{Z}$, there is some integer n such that infinitely many $r \in R$ have $N(r) = n$. Among the elements $r \in R$ with $N(r) = n$, which we write as $r = A + B\sqrt{m}$ with A and B in $\frac{1}{2}\mathbb{Z}$, we consider the finitely many congruence classes of (A, B) modulo n, saying that two such (A, B) and (A', B') are congruent if $A - A'$ and $B - B'$ are integers divisible by n. Since infinitely many $r \in R$ have $N(r) = n$, there must be infinitely many of these in some particular congruence class. Take three such, say α_1, α_2, and α_3. Then

$$N(\alpha_1) = N(\alpha_2) = N(\alpha_3) = n$$

with

$$\frac{\alpha_1 - \alpha_2}{n} \text{ in } R \qquad \text{and} \qquad \frac{\alpha_1 - \alpha_3}{n} \text{ in } R.$$

Since $n = N(\alpha_2) = \alpha_2\sigma(\alpha_2)$, we see that

$$\frac{\alpha_1}{\alpha_2} = 1 + \Big(\frac{\alpha_1 - \alpha_2}{n}\Big)\sigma(\alpha_2).$$

Thus α_1/α_2 is exhibited as in R, and it has $N(\alpha_1/\alpha_2) = N(\alpha_1)/N(\alpha_2) = n/n = 1$. Hence α_1/α_2 is a unit different from $+1$. Arguing similarly with α_1/α_3, we see that α_1/α_3 is a unit different from $+1$ and not equal to α_1/α_2. Hence one of α_1/α_2 and α_1/α_3 is a unit whose absolute value is not 1. □

7. Relationship of Quadratic Forms to Ideals

We continue with K as the quadratic number field $\mathbb{Q}(\sqrt{m})$ and R as the ring of algebraic integers in K. Here $R = \mathbb{Z}[\delta]$, where $\delta = -\sqrt{m}$ if $m \equiv 2$ or $3 \bmod 4$ and $\delta = \frac{1}{2}(1 - \sqrt{m})$ if $m \equiv 1 \bmod 4$. Let D be the field discriminant of $\mathbb{Q}(\sqrt{m})$ as defined in Section 6.

The topic of this section is a relationship between nonzero ideals in R and binary quadratic forms with discriminant D. Binary quadratic forms with D as discriminant are automatically primitive.

The relationship is not a one-one correspondence of ideals to forms but a one-one correspondence of a certain kind of equivalence class of ideals to proper equivalence classes of forms. We saw in Theorem 1.12 that the latter collection has the structure of a finite abelian group, and we shall see in this section that the former collection has the natural structure of a finite abelian group as well. The correspondence is a group isomorphism, according to Theorem 1.20 below.

Consider nonzero ideals I in R. The first observation is that I is additively a free abelian group of rank 2. In fact, R itself is additively a free abelian group of rank 2, and the additive subgroup I has to be free abelian of rank ≤ 2. If r is a nonzero element in I, then $N(r) = r\sigma(r)$ is in I, and thus I contains a nonzero integer. If n is an integer in I, then $n\sqrt{m}$ is in I, and thus I contains a noninteger. Therefore I is a free abelian group of rank exactly 2, as asserted.

Certainly I can then be generated as an ideal by two elements, and our customary notation has been to write $I = (r_1, r_2)$ in this case. However, without an extra condition on them, the two ideal generators need not together be a $\mathbb{Z}$ basis for I because they need not generate all of I additively. It will be helpful to have separate notation when the generators are known to give a $\mathbb{Z}$ basis. Accordingly we shall write $I = \langle r_1, r_2 \rangle$ when r_1, r_2 give a $\mathbb{Z}$ basis of I. In this case it will be helpful also to regard the set $\{r_1, r_2\}$ as *ordered* with r_1 preceding r_2, and we shall often do so.

Now suppose that $I = \langle r_1, r_2 \rangle$ is a nonzero ideal, and consider the expression

$$r_1\sigma(r_2) - \sigma(r_1)r_2 = \det \begin{pmatrix} r_1 & \sigma(r_1) \\ r_2 & \sigma(r_2) \end{pmatrix}.$$

If I is written in terms of a second ordered $\mathbb{Z}$ basis as $I = \langle s_1, s_2 \rangle$, then the two

ordered bases are related by a matrix $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$ in $\mathrm{GL}(2,\mathbb{Z})$, the relationship being

$$\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}.$$

Hence

$$\begin{pmatrix} r_1 & \sigma(r_1) \\ r_2 & \sigma(r_2) \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} s_1 & \sigma(s_1) \\ s_2 & \sigma(s_2) \end{pmatrix},$$

and therefore

$$\det \begin{pmatrix} r_1 & \sigma(r_1) \\ r_2 & \sigma(r_2) \end{pmatrix} = \pm \det \begin{pmatrix} s_1 & \sigma(s_1) \\ s_2 & \sigma(s_2) \end{pmatrix},$$

where ± 1 is the determinant of $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)$. Consequently the expression

$$N(I) = \frac{|r_1\sigma(r_2) - \sigma(r_1)r_2|}{|\sqrt{D}\,|},$$

where D is the field discriminant of K, is independent of the choice of $\mathbb{Z}$ basis. It is called the **norm** of the ideal I. The factor of $\sqrt{D}$ in the denominator is a normalization factor that arranges for the norm of the ideal $I = R$ to be 1; in fact, we can write $R = \langle 1, \delta\rangle$ with δ as in the first paragraph of this section, and then

$$N(R) = \frac{|\sigma(\delta) - \delta|}{|\sqrt{D}\,|} = \left\{ \begin{array}{ll} \frac{|\sqrt{m}+\sqrt{m}\,|}{|\sqrt{4m}\,|} & \text{if } m \equiv 2 \text{ or } 3 \bmod 4 \\ \frac{|\frac{1}{2}(1+\sqrt{m})-\frac{1}{2}(1-\sqrt{m})|}{|\sqrt{m}\,|} & \text{if } m \equiv 1 \bmod 4 \end{array} \right\} = 1.$$

Since the norm of an element of R is given by $N(r) = r\sigma(r)$, it is immediate from the definition that

$$N(rI) = |N(r)|N(I) \qquad \text{for } r \in R.$$

Consequently the norm of the principal ideal (r) is given by

$$N((r)) = |N(r)|N(R) = |N(r)|1 = |N(r)| \qquad \text{for } r \in R.$$

Still with $I = \langle r_1, r_2\rangle$, let us observe that

$$\sigma\big(r_1\sigma(r_2) - \sigma(r_1)r_2\big) = -\big(r_1\sigma(r_2) - \sigma(r_1)r_2\big).$$

It follows that

$$r_1\sigma(r_2) - \sigma(r_1)r_2 \qquad \text{is} \qquad \left\{ \begin{array}{ll} \text{real} & \text{if } m > 0, \\ \text{imaginary} & \text{if } m < 0. \end{array} \right.$$

Since $r_1\sigma(r_2) - \sigma(r_1)r_2$ changes sign when r_1 and r_2 are interchanged, let us say that the expression $I = \langle r_1, r_2\rangle$ for I is **positively oriented** if $r_1\sigma(r_2) - \sigma(r_1)r_2$ is positive or positive imaginary,[9] **negatively oriented** if $r_1\sigma(r_2) - \sigma(r_1)r_2$ is negative or negative imaginary. If $I = \langle r_1, r_2\rangle$, then exactly one of the expressions $I = \langle r_1, r_2\rangle$ and $I = \langle r_2, r_1\rangle$ is positively oriented. The notion of orientation will be critical to setting up the correspondence between classes of ideals and classes of forms.

The set of nonzero ideals of R has a commutative associative multiplication that was introduced in *Basic Algebra*: if I and J are nonzero ideals, then IJ is defined to be the set of sums of products from the two ideals, the product IJ again being an ideal. Later in this section we shall recall some properties of this multiplication that were proved in *Basic Algebra*.

We define two equivalence relations on the set of nonzero ideals of I. We say that I and J are **equivalent** if there exist nonzero r and s in R with $(r)I = (s)J$. Here (r) and (s) are understood to be principal ideals. The ideals I and J are **strictly equivalent**, or **narrowly equivalent**, if equivalence occurs and if r and s can be chosen with $N(rs^{-1}) > 0$. Both relations are certainly reflexive and symmetric. To see transitivity, let $(r_1)I_1 = (r_2)I_2$ and $(s_2)I_2 = (s_3)I_3$. Then $(r_1s_2)I_1 = (r_2s_2)I_2 = (r_2s_3)I_3$, and I_1 is equivalent to I_3. If also $N(r_1r_2^{-1}) > 0$ and $N(s_2s_3^{-1}) > 0$, then the product $N((r_1s_2)(r_2s_3)^{-1})$ is positive, and I_1 is strictly equivalent to I_3. In other words, "equivalent" and "strictly equivalent" are equivalence relations.

The principal ideals form one full equivalence class under "equivalent." First of all, (r) is equivalent to (s) because $(s)(r) = (rs) = (r)(s)$. In the reverse direction, if I and (1) are equivalent, let $(r)I = (s)$. Then there exists $x \in I$ with $rx = s$. Hence sr^{-1} is in I, and $(sr^{-1}) \subseteq I$. In fact, equality holds: if y is in I, then the equality $ry = sz$ with z in R says that $y = (sr^{-1})z$, and y is in (sr^{-1}). In other words, $I = (sr^{-1})$.

In a sense, therefore, equivalence of ideals measures the extent to which nonprincipal ideals exist.

Multiplication is a class property of ideals relative to equivalence and to strict equivalence. In fact, if $(r)I = (r')I'$ and $(s)J = (s')J'$, then $(rs)IJ = (r's')I'J'$, and the assertion follows.

The theorem will be that multiplication of *strict* equivalence classes of ideals of R makes the set of such classes into an abelian group that is isomorphic to the finite abelian form class group of discriminant D. This result is not as beautiful as one might hope, since the identity class of ideals under strict equivalence need not match the set of all principal ideals. However, we can quantify the discrepancy. The relevant result is as follows.

[9] If $m < 0$, we adopt the convention that $\sqrt{m}$ is positive imaginary.

Proposition 1.17. Equivalence and strict equivalence are the same for ideals of R if and only if either

(a) $m > 0$ and the fundamental unit ε_1 has $N(\varepsilon_1) = -1$ or
(b) $m < 0$.

In the contrary case when $m > 0$ and the fundamental unit ε_1 has $N(\varepsilon) = +1$, a nonzero principal ideal (r) is strictly equivalent to (1) if and only if $N(r) > 0$; in particular, the principal ideal $(\sqrt{m})$ is not strictly equivalent to (1).

REMARKS. When $m > 0$, there are examples with $N(\varepsilon_1) = +1$ and examples with $N(\varepsilon_1) = -1$. Specifically when $m = 2$, $\varepsilon_1 = 1 + \sqrt{2}$, and this has $N(\varepsilon_1) = -1$. When $m > 0$ and m has any odd prime divisor p with $p \equiv 3 \bmod 4$, then $N(\varepsilon_1) = +1$; in fact, otherwise $\varepsilon_1 = x + y\sqrt{m}$ would imply that $-1 = N(\varepsilon_1) = x^2 - my^2$ and therefore that $-1 \equiv x^2 \bmod p$, but this congruence has no solutions by Theorem 1.2a.

PROOF. Suppose that $m > 0$ and $N(\varepsilon_1) = -1$. If $(r)I = (s)J$ with $N(rs^{-1}) < 0$, then $(\varepsilon_1 r)I = (s)J$ with $N(\varepsilon_1 rs^{-1}) > 0$. Thus equivalence implies strict equivalence in this case.

Suppose that $m < 0$. Then all norms of nonzero elements are > 0. Hence $N(rs^{-1}) > 0$ is an empty condition, and equivalence implies strict equivalence.

Conversely suppose that $m > 0$ and $N(\varepsilon_1) = +1$. Proposition 1.16 shows that the most general unit is $\varepsilon = \pm\varepsilon_1^n$, and consequently $N(\varepsilon) = N(\pm 1)N(\varepsilon_1)^n = +1$ for every unit. The element $\sqrt{m}$ is in R, and $N(\sqrt{m}) = -m < 0$. We know that the principal ideals (1) and $(\sqrt{m})$ are equivalent. Arguing by contradiction, suppose that they are strictly equivalent. Then $(r) = (r)(1) = (s)(\sqrt{m}) = (s\sqrt{m})$ for some r and s with $N(rs^{-1}) > 0$. Since the principal ideals generated by r and $s\sqrt{m}$ are the same, these elements must be related by $r = \varepsilon s\sqrt{m}$ for some unit ε. Then $N(rs^{-1}) = N(\varepsilon\sqrt{m}) = N(\varepsilon)N(\sqrt{m}) = -m < 0$, contradiction. The proposition follows. □

Once we have introduced group structures on the set of equivalence classes of ideals and the set of strict equivalence classes of ideals, it follows that the map that carries a strict equivalence class to the equivalence class containing it is a group homomorphism onto. If either of the conditions (a) and (b) in Proposition 1.17 is satisfied, then this homomorphism is one-one. Otherwise its kernel consists of the two strict equivalence classes of principal ideals—those whose generator has positive norm and those whose generator has negative norm.

At this point we could establish that the set of strict equivalence classes of ideals is a finite abelian group. The finiteness of the set of strict equivalence classes could be established directly by a geometric argument we give in Chapter V, and the group structure could be derived from the group structure on the set of "fractional ideals" of K that were introduced in Problems 48–53 at the end of Chapter VIII of *Basic Algebra*.

Although we could proceed with proofs along these lines, it is instructive to proceed in a different way. Rather than give a stand-alone proof of the finiteness of the number of strict equivalence classes of ideals, we prefer to derive this finiteness as part of the correspondence with proper equivalence classes of binary quadratic forms, since the number of such classes of binary quadratic forms has already been proved to be finite in Theorem 1.6a. The group structure then readily follows from this finiteness and the fact that R is a Dedekind domain.

Let us pause for a moment, therefore, to use results we already know in order to show how the group structure on the set of strict equivalence classes follows once it is known that there are only finitely many such classes. We know from Theorems 8.54 and 8.55 of *Basic Algebra* that R is a Dedekind domain and that R has unique factorization for its nonzero ideals. In other words, in terms of the already-defined multiplication of ideals, each nonzero ideal I in R is of the form $I = \prod_{j=1}^{k} P_j^{n_j}$, where the P_j are distinct nonzero prime ideals, the n_j are positive integers, and k is ≥ 0; moreover, this product expansion is unique up to the order of the factors.

Lemma 1.18. Let $\mathcal{H}$ be the set of strict equivalence classes of nonzero ideals in R, with its inherited commutative associative multiplication. If $\mathcal{H}$ is finite, then $\mathcal{H}$ is a group under this multiplication.

REMARKS. The group $\mathcal{H}$ will be seen in Theorem 1.20 to be isomorphic to the form class group of D. The set of ordinary equivalence classes is a quotient and is called the **ideal class group** of K. It will be generalized in Chapter V.

PROOF. The identity element of $\mathcal{H}$ is the strict equivalence class of the ideal $R = (1)$, and we are to prove the existence of inverses. Thus let I be given. For the sequence of ideals $I, I^2, I^3, \ldots$, the finiteness of $\mathcal{H}$ shows that two of these ideals must be strictly equivalent. Suppose that I^k is equivalent to I^{k+l} for some $k > 0$ and $l > 0$. Then there exist nonzero principal ideals (r) and (s) such that $(r)I^k = (s)I^{k+l}$. The uniqueness of factorization of ideals implies that we can cancel I^k from both sides of this equality, thereby obtaining $(r) = (s)I^l$. Let us define an element t in R. If $N(rs^{-1}) > 0$, we take t to be 1. Otherwise m must be positive, and we let $t = \sqrt{m}$, so that $N(t) < 0$. In both cases we then have $(rt)(1) = (s)(t)I^l$ with $N(rts^{-1}) > 0$, and the ideal $(t)I^l$ is strictly equivalent to (1). Hence the strict equivalence class of $(t)I^{l-1}$ is an inverse to the strict equivalence class of I, and $\mathcal{H}$ is a group. □

Now we define the mappings $\mathcal{F}$ and $\mathcal{I}$ that we shall use to establish the main result of this section. Let I be a nonzero ideal in R, and suppose that I is given by an expression $I = \langle r_1, r_2 \rangle$ that is positively oriented. We regard x and y as integer variables. To I, we associate the binary quadratic form

$$\mathcal{F}(I, r_1, r_2) = N(I)^{-1} N(r_1 x + r_2 y) = N(I)^{-1} (r_1 x + r_2 y)(\sigma(r_1)x + \sigma(r_2)y).$$

The associated 2-by-2 matrix for this form is

$$\frac{1}{N(I)}\begin{pmatrix} 2r_1\sigma(r_1) & r_1\sigma(r_2)+r_2\sigma(r_1) \\ r_1\sigma(r_2)+r_2\sigma(r_1) & 2r_2\sigma(r_2) \end{pmatrix} = \frac{1}{N(I)}\begin{pmatrix} r_1 & \sigma(r_1) \\ r_2 & \sigma(r_2) \end{pmatrix}\begin{pmatrix} \sigma(r_1) & \sigma(r_2) \\ r_1 & r_2 \end{pmatrix},$$

and the discriminant of the quadratic form is therefore

$$\begin{aligned} -\det\Big[\frac{1}{N(I)}\begin{pmatrix} r_1 & \sigma(r_1) \\ r_2 & \sigma(r_2) \end{pmatrix}\begin{pmatrix} \sigma(r_1) & \sigma(r_2) \\ r_1 & r_2 \end{pmatrix}\Big] &= N(I)^{-2}\big(r_1\sigma(r_2)-\sigma(r_1)r_2\big)^2 \\ &= |D|\,\frac{\big(r_1\sigma(r_2)-\sigma(r_1)r_2\big)^2}{\big|r_1\sigma(r_2)-\sigma(r_1)r_2\big|^2} \\ &= |D|(\operatorname{sgn} m) = D. \end{aligned}$$

Thus we have associated a quadratic form $\mathcal{F}(I, r_1, r_2)$ of discriminant D to an ideal I when I is given by a positively oriented expression $I = \langle r_1, r_2\rangle$. If $m < 0$, this quadratic form is positive definite because the coefficient of x^2, namely $N(I)^{-1}r_1\sigma(r_1) = N(I)^{-1}N(r_1)$, is positive when $m < 0$.

In the reverse direction we associate to an arbitrary form (a, b, c) of discriminant D an ideal $I = \mathcal{I}(a, b, c)$ given by a positively oriented expression $\langle r_1, r_2\rangle$. To begin with, if b is an integer with $b \equiv D \bmod 2$, let us define b' to be $\frac{1}{2}b$ if $D \equiv 0 \bmod 4$ and to be $\frac{1}{2}(b-1)$ if $D \equiv 1 \bmod 4$; in other words, $b' = \frac{1}{2}(b - \operatorname{Tr}(\delta))$ in both cases. The definition of $\mathcal{I}$ is to be

$$\mathcal{I}(a, b, c) = \begin{cases} \langle a, b'+\delta\rangle & \text{if } a > 0, \\ \langle \delta a, \delta(b'+\delta)\rangle & \text{if } a < 0. \end{cases}$$

The right sides in the above display make sense as ideals if the angular brackets are replaced by parentheses. To see that the definitions make sense, we thus need to check that $(a, b'+\delta) = \langle a, b'+\delta\rangle$ for all a and that the orientations are positive. Lemma 1.19a below shows that $(a, b'+\delta) = \langle a, b'+\delta\rangle$ if it is proved that a divides $N(b'+\delta)$, and the computation that verifies this equality is

$$\begin{aligned} N(b'+\delta) &= b'^2 + b'(\delta+\sigma(\delta)) + \delta\sigma(\delta) \\ &= \begin{cases} b'^2 + b' + \frac{1}{4}(1-m) & \text{if } D \equiv 1 \bmod 4, \\ b'^2 - m & \text{if } D \equiv 0 \bmod 4, \end{cases} \\ &= \begin{cases} \frac{1}{4}(b-1)^2 + \frac{1}{2}(b-1) + \frac{1}{4}(1-D) & \text{if } D \equiv 1 \bmod 4, \\ \frac{1}{4}b^2 - \frac{1}{4}D & \text{if } D \equiv 0 \bmod 4, \end{cases} \\ &= \tfrac{1}{4}(b^2 - D) \\ &= ac. \end{aligned}$$

From the definitions near the beginning of this section, the orientation of $\langle r_1, r_2\rangle$ is given by the sign of $(\sqrt{m}\,)^{-1}\big(r_1\sigma(r_2) - \sigma(r_1)r_2\big)$. Thus

$$\begin{aligned}\text{orientation}\langle a, b'+\delta\rangle &= \operatorname{sgn}\big((\sqrt{m}\,)^{-1}a(\sigma(\delta)-\delta)\big) = \operatorname{sgn} a,\\ \text{orientation}\langle \delta a, \delta(b'+\delta)\rangle &= \operatorname{sgn}\big((\sqrt{m}\,)^{-1}\big(\delta a\sigma(\delta b'+\delta^2) - \sigma(\delta)a\delta(b'+\delta)\big)\big)\\ &= \operatorname{sgn}\big((\sqrt{m}\,)^{-1}N(\delta)a(\sigma(\delta)-\delta)\big) = -\operatorname{sgn} a,\end{aligned}$$

and the orientations are positive in both cases.

Lemma 1.19.

(a) If $a \neq 0$ and b' are integers such that a divides $N(b'+\delta)$ in $\mathbb{Z}$, then $(a, b'+\delta) = \langle a, b'+\delta\rangle$ in the sense that the free abelian subgroup of R generated by a and $b'+\delta$ coincides with the ideal generated by a and $b'+\delta$.

(b) If I is any nonzero ideal in R, then I is of the form $I = \langle a, r\rangle$ for some integer $a > 0$ and some r in R.

PROOF. For (a), we are to show that $I' = \mathbb{Z}a + \mathbb{Z}(b'+\delta)$ is closed under multiplication by the generators 1 and δ of R. Closure of I' under multiplication by 1 is evident, and the formula $\delta a = -b'a + a(b'+\delta)$ shows that $\delta(\mathbb{Z}a) \subseteq I'$. Addition of $\delta b'$ to the sum of the two formulas $\delta^2 = \delta(\delta+\sigma(\delta)) - \delta\sigma(\delta) = \delta\,\mathrm{Tr}(\delta) - N(\delta)$ and $N(b'+\delta) = b'^2 + b'\,\mathrm{Tr}(\delta) + N(\delta)$ yields

$$\delta(b'+\delta) = -N(b'+\delta) + (b' + \mathrm{Tr}(\delta))(b'+\delta),$$

which shows that $\delta(b'+\delta) \subseteq I'$ because $N(b'+\delta)$ is by assumption an integer multiple of a.

For (b), we start from any $\mathbb{Z}$ basis $\{r_1, r_2\}$ of I, say with $r_1 = a_1 + b_1\delta$ and $r_2 = a_2 + b_2\delta$, and let $d = \mathrm{GCD}(b_1, b_2)$. Choose integers n_1 and n_2 with $n_1b_1 + n_2b_2 = d$. Then $\mathrm{GCD}(n_1, n_2) = 1$, and we can therefore find integers k_1 and k_2 with $\det\left(\begin{smallmatrix} k_1 & k_2\\ n_1 & n_2\end{smallmatrix}\right) = 1$. Consequently $\left(\begin{smallmatrix} s_1\\ s_2\end{smallmatrix}\right) = \left(\begin{smallmatrix} k_1 & k_2\\ n_1 & n_2\end{smallmatrix}\right)\left(\begin{smallmatrix} r_1\\ r_2\end{smallmatrix}\right)$ is a new $\mathbb{Z}$ basis of I of the form

$$\begin{aligned} s_1 &= c_1 + kd\delta,\\ s_2 &= c_2 + d\delta.\end{aligned}$$

If we put $a = s_1 - ks_2$ and possibly replace a by its negative, then $\{a, s_2\}$ is a $\mathbb{Z}$ basis of I of the required form. □

Theorem 1.20. The set $\mathcal{H}$ of strict equivalence classes of nonzero ideals relative to the field $K = \mathbb{Q}(\sqrt{m}\,)$ is a finite abelian group. Moreover, the mapping $\mathcal{F}$ that carries a positively oriented expression $I = \langle r_1, r_2\rangle$ for a nonzero ideal of R to a binary quadratic form depends only on I, not the ordered $\mathbb{Z}$ basis, and

descends to an isomorphism of the group $\mathcal{H}$ onto the form class group H for the discriminant D of the field K, i.e., the group of proper equivalence classes of binary quadratic forms of discriminant D, subject to the remark below. Moreover, the mapping $\mathcal{I}$ with domain all binary quadratic forms whose discriminant equals the field discriminant of K, sending such a form to a positively oriented expression for a nonzero ideal of R, descends to be defined from H to $\mathcal{H}$, and the descended map is the two-sided inverse of the isomorphism induced by $\mathcal{F}$.

REMARK. If $m < 0$, H is understood as usual to include only the classes of the positive definite forms.

PROOF. The proof proceeds in six steps.

Step 1. We show that the proper equivalence class of the quadratic form $\mathcal{F}(I, r_1, r_2)$ depends only on the ideal I, not the positively oriented expression $I = \langle r_1, r_2 \rangle$ for it. Thus the class of the form can be abbreviated as $\mathcal{F}(I)$.

Suppose that $I = \langle s_1, s_2 \rangle$ is another positively oriented expression for I. Then we can write $\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$ for a matrix $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ in $\mathrm{GL}(2, \mathbb{Z})$, and we have seen that

$$\begin{pmatrix} r_1 & \sigma(r_1) \\ r_2 & \sigma(r_2) \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} s_1 & \sigma(s_1) \\ s_2 & \sigma(s_2) \end{pmatrix}, \tag{$*$}$$

and that

$$\det \begin{pmatrix} r_1 & \sigma(r_1) \\ r_2 & \sigma(r_2) \end{pmatrix} = \pm \det \begin{pmatrix} s_1 & \sigma(s_1) \\ s_2 & \sigma(s_2) \end{pmatrix},$$

where ± 1 is the determinant of $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$. Since both expressions $I = \langle r_1, r_2 \rangle$ and $I = \langle s_1, s_2 \rangle$ are positively oriented, it follows that the sign in the determinant equation is plus, hence that $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ is in $\mathrm{SL}(2, \mathbb{Z})$. Substituting from $(*)$ into the formula for the matrix associated to the binary quadratic form $\mathcal{F}(I, r_1, r_2)$, we obtain the matrix

$$N(I)^{-1} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} s_1 & \sigma(s_1) \\ s_2 & \sigma(s_2) \end{pmatrix} \begin{pmatrix} \sigma(s_1) & \sigma(s_2) \\ s_1 & s_2 \end{pmatrix} \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}. \tag{$**$}$$

The product of the coefficient $N(I)^{-1}$ and the middle two matrices is the matrix associated to the quadratic form $\mathcal{F}(I, s_1, s_2)$, and $(**)$ therefore exhibits the two quadratic forms as properly equivalent.

Step 2. We show that the proper equivalence class $\mathcal{F}(I)$ does not change when we replace I by a strictly equivalent ideal.

Thus let $I = \langle r_1, r_2 \rangle$ and $J = \langle s_1, s_2 \rangle$ be expressions for I and J, and suppose that (r) and (s) are nonzero principal ideals such that $(r)I = (s)J$ and $N(s/r) > 0$. The formula

$$\det \begin{pmatrix} rr_1 & \sigma(rr_1) \\ rr_2 & \sigma(rr_2) \end{pmatrix} = r\sigma(r) \det \begin{pmatrix} r_1 & \sigma(r_1) \\ r_2 & \sigma(r_2) \end{pmatrix} = N(r) \det \begin{pmatrix} r_1 & \sigma(r_1) \\ r_2 & \sigma(r_2) \end{pmatrix}$$

shows that the expression $(r)I = \langle rr_1, rr_2\rangle$ is positively oriented if $N(r) > 0$ and is negatively oriented if $N(r) < 0$. Similarly $(s)J = \langle ss_1, ss_2\rangle$ is positively oriented if $N(s) > 0$ and is negatively oriented if $N(s) < 0$. Since $N(r/s) > 0$, $N(r)$ and $N(s)$ are both positive or both negative. Possibly replacing r and s by $r\sqrt{m}$ and $s\sqrt{m}$, we may assume that $N(r)$ and $N(s)$ are both positive. Then the matrix associated to the quadratic form $\mathcal{F}((r)I, rr_1, rr_2)$ is

$$\begin{aligned} N(rI)^{-1}&\begin{pmatrix} rr_1 & \sigma(rr_1)\\ rr_2 & \sigma(rr_2)\end{pmatrix}\begin{pmatrix} \sigma(rr_1) & \sigma(rr_2)\\ rr_1 & rr_2\end{pmatrix}\\ &= N(rI)^{-1}\begin{pmatrix} r_1 & \sigma(r_1)\\ r_2 & \sigma(r_2)\end{pmatrix}\begin{pmatrix} r & 0\\ 0 & \sigma(r)\end{pmatrix}\begin{pmatrix} \sigma(r) & 0\\ 0 & r\end{pmatrix}\begin{pmatrix} \sigma(r_1) & \sigma(r_2)\\ r_1 & r_2\end{pmatrix}\\ &= N(rI)^{-1}N(r)\begin{pmatrix} r_1 & \sigma(r_1)\\ r_2 & \sigma(r_2)\end{pmatrix}\begin{pmatrix} \sigma(r_1) & \sigma(r_2)\\ r_1 & r_2\end{pmatrix}\\ &= |N(r)|^{-1}N(I)^{-1}N(r)\begin{pmatrix} r_1 & \sigma(r_1)\\ r_2 & \sigma(r_2)\end{pmatrix}\begin{pmatrix} \sigma(r_1) & \sigma(r_2)\\ r_1 & r_2\end{pmatrix}\\ &= N(I)^{-1}\begin{pmatrix} r_1 & \sigma(r_1)\\ r_2 & \sigma(r_2)\end{pmatrix}\begin{pmatrix} \sigma(r_1) & \sigma(r_2)\\ r_1 & r_2\end{pmatrix}, \end{aligned}$$

while the matrix associated to $\mathcal{F}((s)J, ss_1, ss_2)$, by a similar computation, is

$$N(J)^{-1}\begin{pmatrix} s_1 & \sigma(s_1)\\ s_2 & \sigma(s_2)\end{pmatrix}\begin{pmatrix} \sigma(s_1) & \sigma(s_2)\\ s_1 & s_2\end{pmatrix}.$$

Since $(r)I = (s)J$, Step 1 shows that $\mathcal{F}((r)I, rr_1, rr_2)$ is properly equivalent to $\mathcal{F}((s)J, ss_1, ss_2)$.

Step 3. We show that $\mathcal{I}(a, b, c)$ depends only on the proper equivalence class of the binary quadratic form (a, b, c).

Problem 37 at the end of Chapter VII of *Basic Algebra* shows that $\mathrm{SL}(2, \mathbb{Z})$ is generated by $\alpha = \begin{pmatrix} 0 & -1\\ 1 & 0\end{pmatrix}$ and $\beta = \begin{pmatrix} 0 & 1\\ -1 & -1\end{pmatrix}$, hence by $\alpha\beta = \begin{pmatrix} 1 & 1\\ 0 & 1\end{pmatrix}$ and $\alpha^{-1} = \begin{pmatrix} 0 & 1\\ -1 & 0\end{pmatrix}$. Thus it is enough to handle $\alpha\beta$ and α^{-1}.

The operation of $\alpha\beta = \begin{pmatrix} 1 & 1\\ 0 & 1\end{pmatrix}$ on forms sends (a, b, c) into the translate $(a, b + 2a, *)$. Define $b' = \frac{1}{2}(b - \mathrm{Tr}(\delta))$ in the same way as when $\mathcal{I}$ was defined. If $a > 0$, then $\mathcal{I}(a, b, c) = (a, b' + \delta)$, and $\mathcal{I}(a, b + 2a, *) = (a, (b + 2a)' + \delta) = (a, b' + a + \delta)$; thus the two image ideals are the same. If $a < 0$, then the respective images are $(\delta)(a, b' + \delta)$ and $(\delta)(a, b' + a + \delta)$, and again the image ideals are the same.

To handle $\alpha^{-1} = \begin{pmatrix} 0 & 1\\ -1 & 0\end{pmatrix}$, we are to show that the ideals $\mathcal{I}(a, b, c)$ and $\mathcal{I}(c, -b, a)$ are strictly equivalent. We saw just after the definition of $\mathcal{I}$ that $N(b' + \delta) = ac$. There are four cases to the proof of the strict equivalence according to the signs of a and c. Let us use the symbol $\sim$ to denote "is strictly equivalent to."

Suppose that $a > 0$ and $c > 0$, so that $N(b' + \delta) > 0$. Then

$$\begin{aligned}\mathcal{I}(a, b, c) &= (a, b' + \delta) \sim (b' + \sigma(\delta))(a, b' + \delta) = (a(b' + \sigma(\delta)), N(b' + \delta)) \\ &= (a(b' + \sigma(\delta)), ac) = (a)(b' + \sigma(\delta), c) \\ &\sim (c, b' + \sigma(\delta)) = (c, -b' - \sigma(\delta)) = (c, (-b)' + \delta),\end{aligned}$$

the last equality holding because $b' + (-b)' = -\operatorname{Tr}\delta = -\delta - \sigma(\delta)$. The right side equals $\mathcal{I}(c, -b, a)$, and the strict equivalence is proved in this case.

Suppose that $a < 0$ and $c < 0$, so that $N(b' + \delta) > 0$. Then

$$\begin{aligned}\mathcal{I}(a, b, c) &= (\delta)(a, b' + \delta) \sim (b' + \sigma(\delta))(\delta)(a, b' + \delta) \\ &= (\delta)(a(b' + \sigma(\delta)), N(b' + \delta)) = (\delta)(a(b' + \sigma(\delta)), ac) \\ &= (a)(\delta)(b' + \sigma(\delta), c) \sim (\delta)(c, b' + \sigma(\delta)) \\ &= (\delta)(c, -b' - \sigma(\delta)) = (\delta)(c, (-b)' + \delta) = \mathcal{I}(c, -b, a),\end{aligned}$$

and the strict equivalence is proved in this case.

Suppose that $a > 0$ and $c < 0$, so that $N(b' + \delta) < 0$. Then $N(\delta)N(b' + \delta)$ is positive, and

$$\begin{aligned}\mathcal{I}(a, b, c) &= (a, b' + \delta) \sim (\delta)(b' + \sigma(\delta))(a, b' + \delta) \\ &= (\delta)(a(b' + \sigma(\delta)), N(b' + \delta)) = (\delta)(a(b' + \sigma(\delta)), ac) \\ &= (a)(\delta)(b' + \sigma(\delta), c) \sim (\delta)(c, b' + \sigma(\delta)) = (\delta)(c, -b' - \sigma(\delta)) \\ &= (\delta)(c, (-b)' + \delta) = \mathcal{I}(c, -b, a),\end{aligned}$$

and the strict equivalence is proved in this case.

Suppose that $a < 0$ and $c > 0$, so that $N(b' + \delta) < 0$. Then $N(\delta)^{-1}N(b' + \delta)$ is positive, and

$$\begin{aligned}\mathcal{I}(a, b, c) &= (\delta)(a, b' + \delta) \sim (b' + \sigma(\delta))(a, b' + \delta) \\ &= (a(b' + \sigma(\delta)), N(b' + \delta)) = (a(b' + \sigma(\delta)), ac) = (a)(b' + \sigma(\delta), c) \\ &\sim (c, -b' - \sigma(\delta)) = (c, (-b)' + \delta) = \mathcal{I}(c, -b, a),\end{aligned}$$

and the strict equivalence is proved in this case.

Step 4. We show that the mapping of the set H of proper equivalence classes of forms to itself induced by $\mathcal{F}\mathcal{I}$ is the identity.

Let the given form be (a, b, c). With b' defined to be $\frac{1}{2}(b - \operatorname{Tr}(\delta))$ as usual, we have seen that $N(b' + \delta) = ac$. Therefore a divides $N(b' + \delta)$, and Lemma 1.19a shows that $(a, b' + \delta) = \langle a, b' + \delta\rangle$ in the sense that the ideal generated by a and $b' + \delta$ matches the free abelian group generated by these two elements.

First suppose that $a > 0$. Then $\mathcal{I}(a, b, c) = (a, b' + \delta) = \langle a, b' + \delta \rangle$, and we know that this expression is positively oriented. Calculation gives

$$\begin{aligned} N(I) &= |\sqrt{D}\,|^{-1} \left| \det \left(\begin{smallmatrix} a & a \\ b'+\delta & b'+\sigma(\delta) \end{smallmatrix} \right) \right| \\ &= a|\sqrt{D}\,|^{-1}|\sigma(\delta) - \delta)| \\ &= a \times \begin{cases} |\sqrt{m}\,|/|\sqrt{m}\,| & \text{if } D \equiv 1 \bmod 4, \\ 2|\sqrt{m}\,|/|\sqrt{4m}\,| & \text{if } D \equiv 0 \bmod 4, \end{cases} \\ &= a. \end{aligned} \tag{$\dagger$}$$

Therefore the quadratic form $\mathcal{FI}(a, b, c)$ is

$$\begin{aligned} N(I)^{-1}&(ax + (b' + \delta)y)(ax + (b' + \sigma(\delta))y) \\ &= a^{-1}\big(a^2x^2 + a\big(2b' + (\delta + \sigma(\delta))\big)xy + N(b' + \delta)y^2\big) \\ &= ax^2 + \big(2b' + \mathrm{Tr}(\delta)\big)xy + cy^2 \\ &= ax^2 + bxy + cy^2, \end{aligned}$$

and we see that $\mathcal{FI}(a, b, c) = (a, b, c)$ when $a > 0$.

Next suppose that $a < 0$. Then $\mathcal{I}(a, b, c) = (\delta a, \delta(b' + \delta)) = \langle \delta a, \delta(b' + \delta) \rangle$, and we know that this expression is positively oriented. Since $a < 0$ cannot occur for $m < 0$, $N(\delta)$ is negative. Thus calculation gives

$$\begin{aligned} N(I) &= N((\delta)(a, b' + \delta)) = N((\delta)(-a, b' + \delta)) = |N(\delta)|N((-a, b' + \delta)) \\ &= |N(\delta)||a| = N(\delta)a, \end{aligned}$$

the next-to-last equality following from the calculation that gives ($\dagger$). Therefore the quadratic form $\mathcal{FI}(a, b, c)$ is

$$\begin{aligned} N(I)^{-1}&(a\delta x + (b' + \delta)\delta y)(a\sigma(\delta)x + (b' + \sigma(\delta))\sigma(\delta)y) \\ &= N(I)^{-1}N(\delta)(ax + (b' + \delta)y)(ax + (b' + \sigma(\delta))y) \\ &= a^{-1}\big(a^2x^2 + a\big(2b' + (\delta + \sigma(\delta))\big)xy + N(b' + \delta)y^2\big) \\ &= ax^2 + \big(2b' + \mathrm{Tr}(\delta)\big)xy + cy^2 \\ &= ax^2 + bxy + cy^2, \end{aligned}$$

and we see that $\mathcal{FI}(a, b, c) = (a, b, c)$ when $a < 0$.

Step 5. We show that the mapping of the set $\mathcal{H}$ of strict equivalence classes of ideals to itself induced by $\mathcal{I}\mathcal{F}$ is the identity. In view of Step 4, it follows that $\mathcal{F}$ and $\mathcal{I}$ are both one-one onto. Since Theorem 1.6a shows H to be finite, $\mathcal{H}$ has to be finite, and Lemma 1.18 shows that the multiplication on $\mathcal{H}$ makes $\mathcal{H}$ into an abelian group.

Let an ideal I be given, and apply Lemma 1.19b to write $I = \langle \widetilde{a}, r\rangle$ with $\widetilde{a} > 0$ an integer. The expression deciding orientation is $\widetilde{a}\sigma(r) - \sigma(\widetilde{a})r = \widetilde{a}(\sigma(r) - r)$, and this is multiplied by -1 if r is replaced by $-r$. Possibly changing r to $-r$ in the expression for I, we may therefore assume that the expression $I = \langle \widetilde{a}, r\rangle$ is positively oriented. Write $r = c + d\delta$. Then

$$\sigma(r) - r = d(\sigma(\delta) - \delta) = \left\{ \begin{array}{ll} 2d\sqrt{m} & \text{if } m \equiv 2 \text{ or } 3 \bmod 4 \\ d\sqrt{m} & \text{if } m \equiv 1 \bmod 4 \end{array} \right\} = d\sqrt{D}.$$

The orientation of I is given by $\widetilde{a}(\sigma(r) - r) = \widetilde{a}d\sqrt{D}$, and we deduce that $d > 0$ and that

$$N(I) = |\sqrt{D}\,|^{-1}\widetilde{a}|\sigma(r) - r| = \widetilde{a}d.$$

The definition of $\mathcal{F}$ gives $\mathcal{F}(I, \widetilde{a}, r) = N(I)^{-1}N(\widetilde{a}x + ry)$, which is a quadratic form whose x^2 coefficient is $a = N(I)^{-1}\widetilde{a}^{\,2} = d^{-1}\widetilde{a}$ and whose xy coefficient is

$$b = N(I)^{-1}\widetilde{a}\,\mathrm{Tr}(r) = d^{-1}\,\mathrm{Tr}(r) = d^{-1}(2c + d\,\mathrm{Tr}(\delta)) = 2d^{-1}c + \mathrm{Tr}(\delta).$$

With b' defined as usual to be $b' = \frac{1}{2}(b - \mathrm{Tr}(\delta))$, we see that $b' = d^{-1}c$. Consequently $\mathcal{I}\mathcal{F}(I, \widetilde{a}, r) = (a, b' + \delta) = (d^{-1}\widetilde{a}, d^{-1}c + \delta)$. The product of this ideal with (d) is $(\widetilde{a}, c + d\delta) = (\widetilde{a}, r) = I$, and thus $\mathcal{I}\mathcal{F}(I, \widetilde{a}, r)$ is strictly equivalent to I.

Step 6. We show that the mapping induced by $\mathcal{I}$ from the set H of proper equivalence classes of forms to the set $\mathcal{H}$ of strict equivalence classes of ideals respects the group operations in H and $\mathcal{H}$ and hence is an isomorphism.

Let two proper equivalence classes of forms with discriminant D be given, and use Theorem 1.12a to choose representatives (a, b, c) and $(\widetilde{a}, b, \widetilde{c}\,)$ with $\mathrm{GCD}(a, \widetilde{a}) = 1$. The composition of the forms is well defined and is $(a\widetilde{a}, b, *)$ for a suitable third entry in $\mathbb{Z}$. Let b' be $\frac{1}{2}(b - \mathrm{Tr}(\delta))$ as usual. We divide matters into cases according to the signs of a and $\widetilde{a}$.

Suppose that $a > 0$ and $\widetilde{a} > 0$. The definition of $\mathcal{I}$ shows that the ideals corresponding to the three quadratic forms in question are

$$(a, b' + \delta), \quad (\widetilde{a}, b' + \delta), \quad \text{and} \quad (a\widetilde{a}, b' + \delta).$$

The product of the first two ideals is $\big(a\widetilde{a}, a(b' + \delta), \widetilde{a}(b' + \delta), (b' + \delta)^2\big)$, and we are to show that this equals $(a\widetilde{a}, b' + \delta)$. In fact, the inclusion

$$\big(a\widetilde{a}, a(b' + \delta), \widetilde{a}(b' + \delta), (b' + \delta)^2\big) \subseteq (a\widetilde{a}, b' + \delta)$$

is clear. For the reverse inclusion we use the fact that $\mathrm{GCD}(a, \widetilde{a}) = 1$ to write $k_1 a + k_2 \widetilde{a} = 1$ for suitable integers k_1 and k_2. Then we see that $b' + \delta = k_1(a(b' + \delta)) + k_2(\widetilde{a}(b' + \delta))$, and the reverse inclusion follows.

Suppose that a and $\widetilde{a}$ are of opposite sign. By symmetry we may assume that $a > 0$ and $\widetilde{a} < 0$. The three ideals are then

$$(a, b' + \delta), \quad (\widetilde{a}\delta, (b' + \delta)\delta), \quad \text{and} \quad (a\widetilde{a}\delta, (b' + \delta)\delta),$$

while the product of the first two ideals is $\big(a\widetilde{a}\delta, a(b'+\delta)\delta, \widetilde{a}(b'+\delta)\delta, (b'+\delta)^2\delta\big) = (\delta)\big(a\widetilde{a}, a(b' + \delta), \widetilde{a}(b' + \delta), (b' + \delta)^2\big)$. From the previous paragraph this last ideal equals $(\delta)(a\widetilde{a}, b' + \delta) = (a\widetilde{a}\delta, (b' + \delta)\delta)$, and we have the required match.

Suppose that $a < 0$ and $\widetilde{a} < 0$. This time the product ideal is given by $(a\delta, (b' + \delta)\delta)(\widetilde{a}\delta, (b' + \delta)\delta) = (\delta^2)\big(a\widetilde{a}, a(b' + \delta), \widetilde{a}(b' + \delta), (b' + \delta)^2\big) = (\delta^2)(a\widetilde{a}, b' + \delta)$, the second equality following from the computation in the paragraph for a and $\widetilde{a}$ both positive. The ideal $(\delta^2)(a\widetilde{a}, b'+\delta)$ is strictly equivalent to $(a\widetilde{a}, b' + \delta)$ because $N(\delta^2) = N(\delta)^2$ is positive. Thus we have the required match on the level of strict equivalence classes. We conclude that the mapping of H to $\mathcal{H}$ is a group isomorphism. □

8. Primes in the Progressions $4n + 1$ and $4n + 3$

This section is the first of three sections about Dirichlet's Theorem on primes in arithmetic progressions, whose statement is as follows.

Theorem 1.21 (Dirichlet's Theorem). If m and b are relatively prime integers with $m > 0$, then there exist infinitely many primes of the form $km + b$ with k a positive integer.

We begin with the earlier treatment of the arithmetic progressions $4n + 1$ and $4n + 3$ by Euler. In 1737 Euler made the stunning discovery of the formula

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1 - p^{-s}},$$

valid for $s > 1$. Actually, the formula is valid for complex s with $\mathrm{Re}\, s > 1$, but Euler had not considered powers n^s with s complex by this time and did not need them for his purpose. Euler's formula is a consequence of unique factorization of integers. In fact, the product for $p \le N$ is

$$\prod_{p \le N} \frac{1}{1 - p^{-s}} = \prod_{p \le N} \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\right) = \sum_{\substack{n \text{ with} \\ \text{no prime} \\ \text{divisors} > N}} \frac{1}{n^s}.$$

Letting $N \to \infty$, we obtain the desired formula.

Built into the formula is the result of Euclid's that there are infinitely many primes, i.e., infinitely many primes in the arithmetic progression n. There are two ways to see this. In both cases one starts from the observation that the sum $\sum_{n=1}^{\infty} 1/n^s$ is $\geq \int_1^{\infty} (1/x^s)\,dx = 1/(s-1)$, from which it follows that the sum tends to infinity as s decreases to 1. In one case the argument continues with the observation that if there were only finitely many primes, then $\prod_{p \text{ prime}} \frac{1}{1-p^{-s}}$ would certainly have finite limit as s decreases to 1, and we arrive at a contradiction. In the other case the argument continues with the observation that the logarithm of $\frac{1}{1-p^{-s}}$ is comparable in size to $1/p^s$, hence that $\log \sum_{n=1}^{\infty} 1/n^s$ is comparable to $\sum_{p \text{ prime}} 1/p^s$. Since $\sum_{n=1}^{\infty} 1/n^s$ tends to infinity, $\sum_{p \text{ prime}} 1/p^s$ must tend to infinity, and we conclude that there are infinitely many primes. We shall return to this observation shortly in order to justify it more rigorously.[10]

Euclid's proof was much simpler: if there were only finitely many primes, then the sum of 1 and the product of all the primes would be divisible by none of the primes and would give a contradiction. The difficulty with Euclid's argument is that there is no apparent way to adapt it to treat primes of the form $4n+1$. Euler's argument, by contrast, does adapt to treat primes $4n+1$.

Before continuing, let us make rigorous the notion of comparing sizes of factors of an infinite product with terms of an infinite series. An infinite product $\prod_{n=1}^{\infty} c_n$ with $c_n \in \mathbb{C}$ and with no factor 0 is said to **converge** if the sequence of partial products converges to a finite limit and the limit is not 0. A necessary condition for convergence is that c_n tend to 1.

Proposition 1.22. If $|a_n| < 1$ for all n, then the following conditions are equivalent:

(a) $\prod_{n=1}^{\infty}(1+|a_n|)$ converges,

(b) $\sum_{n=1}^{\infty} |a_n|$ converges,

(c) $\prod_{n=1}^{\infty}(1-|a_n|)$ converges.

In this case, $\prod_{n=1}^{\infty}(1+a_n)$ converges.

PROOF. Condition (c) is equivalent to

(c′) $\prod_{n=1}^{\infty}(1-|a_n|)^{-1}$ converges.

For each of (a), (b), and (c′), convergence is equivalent to boundedness above. Since

$$1 + \sum_{n=1}^{N} |a_n| \leq \prod_{n=1}^{N} (1+|a_n|) \leq \prod_{n=1}^{N} \tfrac{1}{1-|a_n|},$$

[10] In fact, this argument is showing that $\sum 1/p$ diverges, which says something more than just that there are infinitely many primes.

we see that (c′) implies (a) and that (a) implies (b). To see that (b) implies (c′), we may assume, without loss of generality, that $|a_n| \le \frac{1}{2}$ for all n. Since $|x| \le \frac{1}{2}$ implies that

$$\log \tfrac{1}{1-x} \le |x| \sup_{|t|\le|x|\le\frac{1}{2}} \left|\tfrac{d}{dt} \log \tfrac{1}{1-t}\right| = |x| \sup_{|t|\le|x|\le\frac{1}{2}} \left(\tfrac{1}{1-t}\right) \le 2|x|,$$

we have

$$\log\Big(\prod_{n=1}^{N} \tfrac{1}{1-|a_n|}\Big) = \sum_{n=1}^{N} \log\Big(\tfrac{1}{1-|a_n|}\Big) \le 2\sum_{n=1}^{N} |a_n|.$$

Thus (b) implies (c′).

Now suppose that (a) holds. To prove that $\prod_{n=1}^{\infty}(1+a_n)$ converges, it is enough to show that $\prod_{n=M}^{N}(1+a_n)$ tends to 1 as M and N tend to ∞. In the expression

$$\left|\prod_{n=M}^{N} (1+a_n) - 1\right|,$$

we expand out the product, move the absolute values in for each term, and reassemble the product. The result is the inequality

$$\left|\prod_{n=M}^{N} (1+a_n) - 1\right| \le \prod_{n=M}^{N} (1+|a_n|) - 1.$$

By (a), the right side tends to 0 as M and N tend to ∞. Therefore so does the left side. This proves the proposition. □

Using this proposition and its proof, we can give a more rigorous justification for the comparison of $\log \sum_{n=1}^{\infty} n^{-s}$ and $\sum_{p \text{ prime}} p^{-s}$ in Euler's argument. Anticipating the notation that Riemann was to use for the function a century later, we introduce

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s},$$

at the moment just for real s with $s > 1$. (This function subsequently was named the **Riemann zeta function** and is defined and analytic for complex s with $\operatorname{Re} s > 1$. We postpone a more serious discussion of $\zeta(s)$ to Proposition 1.24 below.) We begin from the formula

$$\log \zeta(s) = \sum_{p \text{ prime}} \log \tfrac{1}{1-p^{-s}} = \sum_{p \text{ prime}} \Big(\tfrac{1}{p^s} + \tfrac{1}{2p^{2s}} + \tfrac{1}{3p^{3s}} + \cdots\Big).$$

Let us see that this expression equals

$$\sum_{p \text{ prime}} \tfrac{1}{p^s} + \text{bounded term} \qquad \text{as } s \downarrow 1.$$

Going over the second displayed line in the proof of Proposition 1.22, which applied when $|x| \le \frac{1}{2}$, we have

$$\begin{aligned}\left|\log\tfrac{1}{1-x} - x\right| &\le |x| \sup_{|t|\le|x|\le\frac{1}{2}} \left|\tfrac{d}{dt}\left(\log\tfrac{1}{1-t} - t\right)\right| \\ &= |x| \sup_{|t|\le|x|\le\frac{1}{2}} \left|\tfrac{1}{1-t} - 1\right| = |x| \sup_{|t|\le|x|\le\frac{1}{2}} \left|\tfrac{t}{1-t}\right| \le 2|x|^2.\end{aligned}$$

For $x = p^{-s}$ with $s > 1$, this inequality becomes

$$\left|\log\tfrac{1}{1-p^{-s}} - \tfrac{1}{p^s}\right| \le 2p^{-2s}.$$

Consequently

$$\begin{aligned}\left|\log\zeta(s) - \sum_{p\text{ prime}} \tfrac{1}{p^s}\right| &= \left|\sum_{p\text{ prime}} \left[\log\tfrac{1}{1-p^{-s}} - \tfrac{1}{p^s}\right]\right| \\ &\le \sum_{p\text{ prime}} \left|\log\tfrac{1}{1-p^{-s}} - \tfrac{1}{p^s}\right| \le 2\sum_{p\text{ prime}} p^{-2s}.\end{aligned}$$

The right side is $\le 2\sum_{n=1}^{\infty} n^{-2}$ for all $s > 1$, and we arrive at the desired formula

$$\log\zeta(s) = \sum_{p\text{ prime}} \frac{1}{p^s} + \text{bounded term} \qquad \text{as } s \downarrow 1.$$

Since we know that $\log\zeta(s)$ increases without bound as s decreases to 1, we can immediately conclude that there are infinitely many primes in the arithmetic progression n.

With this argument well understood as a prototype, let us modify it to treat primes $4k+1$ separately from primes $4k+3$. Euler needed one further key idea to succeed. It is tempting to replace the sum over all primes of p^{-s} in the above argument by

$$\sum_{\substack{p\text{ prime,}\\ p\equiv 1 \bmod 4}} \frac{1}{p^s} \qquad \text{or} \qquad \sum_{\substack{p\text{ prime,}\\ p\equiv 3 \bmod 4}} \frac{1}{p^s},$$

trace backward, and see what happens. What happens is that the expansion of the corresponding product of $(1-p^{-s})^{-1}$ as a sum does not yield anything very manageable. For example, with the first of the two sums, we are led to the logarithm of the series $\sum_{n=1}^{\infty} c(n)n^{-s}$, where $c(n)$ is 1 if n is a product of primes $4k+1$ and is 0 otherwise, and we have no direct way of deciding whether this diverges or converges as s decreases to 1.

Euler's key additional idea was to work with the sum and difference of the displayed series, rather than the two terms separately, and then to recover the two displayed series at the end. Let us see what this idea accomplishes. Tracing backward in the derivation of the formula $\log \zeta(s) = \sum_{p \text{ prime}} p^{-s} + \text{bounded term}$, we want to obtain a series $\sum_{p \text{ prime}} a_p p^{-s}$ from the logarithm of a product $\prod_p (1 - a_p p^{-s})^{-1}$ and be able to recognize this product as equal to a manageable series $\sum_{n=1}^{\infty} b_n n^{-s}$. Guided by what happens for $\zeta(s)$, we can hope that b_n will be readily computable from the a_p's and the unique factorization of n. The relevant identities, which we shall verify below, are as follows:

$$\sum_{n \text{ odd}} \frac{1}{n^s} = \prod_{\substack{p \text{ prime,} \\ p \text{ odd}}} \frac{1}{1 - p^{-s}},$$

$$\sum_{n \text{ odd}} \frac{(-1)^{\frac{1}{2}(n-1)}}{n^s} = \Big(\prod_{\substack{p \text{ prime,} \\ p=4k+1}} \frac{1}{1 - p^{-s}} \Big) \Big(\prod_{\substack{p \text{ prime,} \\ p=4k+3}} \frac{1}{1 + p^{-s}} \Big).$$

In more detail let us write

$$\chi_0(n) = \begin{cases} 0 & \text{if } n \equiv 0 \mod 2, \\ 1 & \text{if } n \equiv 1 \mod 2, \end{cases}$$

$$\chi_1(n) = \begin{cases} 0 & \text{if } n \equiv 0 \mod 2, \\ 1 & \text{if } n \equiv 1 \mod 4, \\ -1 & \text{if } n \equiv 3 \mod 4. \end{cases}$$

With χ equal to χ_0 or χ_1, we have $\chi(mn) = \chi(m)\chi(n)$ for all m and n. Consequently the two expressions $\sum_{n \text{ odd}} \frac{1}{n^s}$ and $\sum_{n \text{ odd}} \frac{(-1)^{\frac{1}{2}(n-1)}}{n^s}$ are both of the form

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s},$$

the function χ being χ_0 for the first series and being χ_1 for the second series. As we shall verify rigorously in the next section, the same argument via unique factorization that yields Euler's identity $\sum_{n=1}^{\infty} n^{-s} = \sum_{p \text{ prime}} \frac{1}{1-p^{-s}}$ gives a factorization

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod_{p \text{ prime}} \frac{1}{1 - \chi(p) p^{-s}}$$

because of the identity $\chi(mn) = \chi(m)\chi(n)$. Going over the argument that $\log \zeta(s)$ is the sum of $\sum_{p \text{ prime}} p^{-s}$ and a bounded term, we find that

$$\log L(s, \chi) = \sum_{p \text{ prime}} \frac{\chi(p)}{p^s} + g(s, \chi)$$

with $g(s, \chi)$ bounded as $s \downarrow 1$. The sum and difference for the two choices of $\chi(n)$ gives

$$\log(L(s, \chi_0)L(s, \chi_1)) = 2 \sum_{\substack{p \text{ prime} \\ p=4k+1}} \frac{1}{p^s} + (g(s, \chi_0) + g(s, \chi_1))$$

and

$$\log(L(s, \chi_0)L(s, \chi_1)^{-1}) = 2 \sum_{\substack{p \text{ prime} \\ p=4k+3}} \frac{1}{p^s} + (g(s, \chi_0) - g(s, \chi_1)) .$$

The function $L(s, \chi_0)$ is the product of $\zeta(s)$ and an elementary factor. In fact, a change of index of summation in the formula defining $\zeta(s)$ gives $2^{-s}\zeta(s) = \sum_{n \text{ even}} n^{-s}$. Subtracting this formula from the definition of $\zeta(s)$ gives

$$L(s, \chi_0) = \sum_{n \text{ odd}} \frac{1}{n^s} = (1 - 2^{-s})\zeta(s).$$

Therefore

$$\lim_{s \downarrow 1} L(s, \chi_0) = +\infty.$$

Meanwhile, the series $L(s, \chi_1) = \sum_{n \text{ odd}} \frac{(-1)^{\frac{1}{2}(n-1)}}{n^s}$ is alternating and converges for $s > 0$ by the Leibniz test. The convergence is uniform on compact sets, and the sum $L(s, \chi_1)$ is continuous for $s > 0$. Grouping the terms of this series in pairs, we see that $L(1, \chi_1)$ is positive.[11] Hence we have

$$0 < \lim_{s \downarrow 1} L(s, \chi_1) < +\infty.$$

Putting together the two limit relations for $L(s, \chi_0)$ and $L(s, \chi_1)$ as s decreases to 1, we see that

$$\log\big(L(s, \chi_0)L(s, \chi_1)\big) \qquad \text{and} \qquad \log\big(L(s, \chi_0)L(s, \chi_1)^{-1}\big)$$

both tend to $+\infty$ as $s \downarrow 1$. Referring to the values computed above for these expressions and taking into account that $\sum 1/p$ exceeds $\sum 1/p^s$ when $s > 1$, we see that

$$\sum_{\substack{p \text{ prime} \\ p=4k+1}} \frac{1}{p} \qquad \text{and} \qquad \sum_{\substack{p \text{ prime} \\ p=4k+3}} \frac{1}{p}$$

[11] We can even recognize the value of $L(1, \chi_1)$ as $\pi/4$ from the Taylor series of $\arctan x$, but the explicit value is not needed in the argument.

are both infinite. Hence there are infinitely many primes $4k + 1$, and there are infinitely many primes $4k + 3$.

The proof of the general case of Dirichlet's Theorem (Theorem 1.21) will proceed in similar fashion. We return to it in Section 10 after a brief but systematic investigation of the kinds of series and products that we have encountered in the present section.

9. Dirichlet Series and Euler Products

A series $\sum_{n=1}^{\infty} a_n n^{-s}$ with a_n and s complex is called a **Dirichlet series**. The first result below shows that the region of convergence and the region of absolute convergence for such a series are each right half-planes in $\mathbb{C}$ unless they are equal to the empty set or to all of $\mathbb{C}$. These half-planes may not be the same: for example, $\sum_{n=1}^{\infty}(-1)^n n^{-s}$ is convergent for $\operatorname{Re} s > 0$ and absolutely convergent for $\operatorname{Re} s > 1$.

Proposition 1.23. Let $\sum_{n=1}^{\infty} a_n n^{-s}$ be a Dirichlet series.

(a) If the series is convergent for $s = s_0$, then it is convergent uniformly on compact sets for $\operatorname{Re} s > \operatorname{Re} s_0$, and the sum of the series is analytic in this region.

(b) If the series is absolutely convergent for $s = s_0$, then it is uniformly absolutely convergent for $\operatorname{Re} s \geq \operatorname{Re} s_0$.

(c) If the series is convergent for $s = s_0$, then it is absolutely convergent for $\operatorname{Re} s > \operatorname{Re} s_0 + 1$.

(d) If the series is convergent at some s_0 and sums to 0 in a right half-plane, then all the coefficients are 0.

REMARK. The proof of (a) will use the **summation by parts** formula. Namely if $\{u_n\}$ and $\{v_n\}$ are sequences and if $U_n = \sum_{k=1}^{n} u_k$ for $n \geq 0$, then $1 \leq M \leq N$ implies

$$\sum_{n=M}^{N} u_n v_n = \sum_{n=M}^{N-1} U_n(v_n - v_{n+1}) + U_N v_N - U_{M-1} v_M. \tag{$*$}$$

PROOF. For (a), we write $a_n n^{-s} = a_n n^{-s_0} \cdot n^{-(s-s_0)} = u_n v_n$ and then apply the summation by parts formula $(*)$. The given convergence means that the sequence $\{U_n\}$ is convergent, and certainly v_n tends to 0 uniformly on any proper half-plane of $\operatorname{Re} s > \operatorname{Re} s_0$. Thus the second and third terms on the right side of $(*)$ tend to 0 with the required uniformity as M and N tend to ∞. For the first term, the sequence $\{U_n\}$ is bounded, and we shall show that

$$\sum_{n=1}^{\infty} |v_n - v_{n+1}| = \sum_{n=1}^{\infty} \left| \tfrac{1}{n^{s-s_0}} - \tfrac{1}{(n+1)^{s-s_0}} \right|$$

is convergent uniformly on compact sets for which $\operatorname{Re} s > \operatorname{Re} s_0$. Use of $(*)$ and the Cauchy criterion will complete the proof of convergence. For $n \le t \le n+1$, we have

$$\begin{aligned} |n^{-(s-s_0)} - t^{-(s-s_0)}| &\le \sup_{n\le t\le n+1} \left|\tfrac{d}{dt}(n^{-(s-s_0)} - t^{-(s-s_0)})\right| \\ &= \sup_{n\le t\le n+1} \left|\tfrac{s-s_0}{t^{s-s_0+1}}\right| \le \tfrac{|s-s_0|}{n^{1+\operatorname{Re}(s-s_0)}}. \end{aligned}$$

Thus

$$|v_n - v_{n+1}| = |n^{-(s-s_0)} - (n+1)^{-(s-s_0)}| \le \tfrac{|s-s_0|}{n^{1+\operatorname{Re}(s-s_0)}},$$

and $\sum_{n=1}^{\infty} |v_n - v_{n+1}|$ is uniformly convergent on compact sets with $\operatorname{Re} s > \operatorname{Re} s_0$, by the Weierstrass M-test. It follows that the given Dirichlet series is uniformly convergent on compact sets for which $\operatorname{Re} s > \operatorname{Re} s_0$. Since each term is analytic in this region, the sum is analytic.

For (b), we have

$$\left|\tfrac{a_n}{n^s}\right| = \left|\tfrac{a_n}{n^{s_0}}\right| \cdot \left|\tfrac{1}{n^{s-s_0}}\right| \le \left|\tfrac{a_n}{n^{s_0}}\right|.$$

Since the sum of the right side is convergent, the desired uniform convergence follows from the Weierstrass M-test.

For (c), let $\epsilon > 0$ be given. Then

$$\left|\tfrac{a_n}{n^{s_0+1+\epsilon}}\right| = \left|\tfrac{a_n}{n^{s_0}}\right| \, n^{-(1+\epsilon)}$$

with the first factor on the right bounded and the second factor contributing to a finite sum. Therefore we have absolute convergence at $s_0 + 1 + \epsilon$, and (c) follows from (b).

For (d), we may assume by (c) that there is absolute convergence at s_0. Suppose that $a_1 = \cdots = a_{N-1} = 0$. By (b), $\sum_{n=N}^{\infty} a_n n^{-s} = 0$ for $\operatorname{Re} s > \operatorname{Re} s_0$. The series

$$\sum_{n=N}^{\infty} a_n (n/N)^{-s} \tag{$**$}$$

is by assumption absolutely convergent at s_0, and $\operatorname{Re} s > \operatorname{Re} s_0$ implies

$$\left|a_n (n/N)^{-s}\right| \le \left|a_n (n/N)^{-s_0}\right|.$$

By dominated convergence we can take the limit of $(**)$ term by term as $s \to +\infty$. The only term that survives is a_N. Since $(**)$ has sum 0 for all s, we conclude that $a_N = 0$. This completes the proof. □

Proposition 1.24. The Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$, initially defined and analytic for $\operatorname{Re} s > 1$, extends to be meromorphic for $\operatorname{Re} s > 0$. Its only pole is at $s = 1$, and the pole is simple.

REMARK. Actually, $\zeta(s)$ extends to be meromorphic in $\mathbb{C}$ with no additional poles, but we do not need this additional information.

PROOF. For $\operatorname{Re} s > 1$, we have

$$\tfrac{1}{s-1} = \int_1^{\infty} t^{-s}\,dt = \sum_{n=1}^{\infty} \int_n^{n+1} t^{-s}\,dt.$$

Thus $\operatorname{Re} s > 1$ implies

$$\zeta(s) = \tfrac{1}{s-1} + \sum_{n=1}^{\infty} \left(\tfrac{1}{n^s} - \int_n^{n+1} t^{-s}\,dt \right) = \tfrac{1}{s-1} + \sum_{n=1}^{\infty} \int_n^{n+1} (n^{-s} - t^{-s})\,dt.$$

It is enough to show that the series on the right side converges uniformly on compact sets for $\operatorname{Re} s > 0$. Thus suppose that $\operatorname{Re} s \geq \sigma > 0$ and $|s| \leq C$. The proof of Proposition 1.23a showed that $|n^{-s} - t^{-s}| \leq |s|\, n^{-(1+\operatorname{Re} s)}$. Hence

$$\left| \int_n^{n+1} (n^{-s} - t^{-s})\,dt \right| \leq \int_n^{n+1} |n^{-s} - t^{-s}|\,dt \leq |s|\, n^{-(1+\operatorname{Re} s)} \leq C n^{-(1+\sigma)}.$$

Since $\sum_{n=1}^{\infty} n^{-(1+\sigma)} < \infty$, the desired uniform convergence follows from the Weierstrass M-test. □

Proposition 1.25. Let $Z(s) = \sum_{n=1}^{\infty} a_n n^{-s}$ be a Dirichlet series with all $a_n \geq 0$. Suppose that the series is convergent in some half-plane and that the sum extends to be analytic for $\operatorname{Re} s > 0$. Then the series converges for $\operatorname{Re} s > 0$.

PROOF. By assumption the series converges somewhere, and therefore $s_0 = \inf \left\{ s \geq 0 \;\middle|\; \sum_{n=1}^{\infty} a_n n^{-s} \text{ converges} \right\}$ is a well-defined real number ≥ 0. Arguing by contradiction, suppose that $s_0 > 0$. Since $\sum a_n n^{-s}$ converges uniformly on compact sets for $\operatorname{Re} s > s_0$ by Proposition 1.23a and since the terms of the series are analytic, we can compute the derivatives of the series term by term. Thus

$$Z^{(N)}(s_0 + 1) = \sum_{n=1}^{\infty} \tfrac{a_n (-\log n)^N}{n^{s_0+1}}. \tag{$*$}$$

The Taylor series of $Z(s)$ about $s_0 + 1$ is

$$Z(s) = \sum_{N=0}^{\infty} \tfrac{1}{N!} (s - s_0 - 1)^N Z^{(N)}(s_0 + 1)$$

and is convergent at $s = \frac{1}{2}s_0$, since $Z(s)$ is analytic in the open disk centered at $s_0 + 1$ and having radius $s_0 + 1$. Thus

$$Z(\tfrac{1}{2}s_0) = \sum_{N=0}^{\infty} \tfrac{1}{N!}(1 + \tfrac{1}{2}s_0)^N (-1)^N Z^{(N)}(s_0 + 1),$$

with the series convergent. Substituting from $(*)$, we have

$$Z(\tfrac{1}{2}s_0) = \sum_{N=0}^{\infty} \sum_{n=1}^{\infty} \tfrac{a_n (\log n)^N}{N!\, n^{s_0+1}} (1 + \tfrac{1}{2}s_0)^N.$$

This is a series with terms ≥ 0, and Fubini's Theorem allows us to interchange the order of summation and obtain

$$Z(\tfrac{1}{2}s_0) = \sum_{n=1}^{\infty} \sum_{N=0}^{\infty} \tfrac{a_n}{n^{s_0+1}} \tfrac{(\log n)^N (1+\frac{1}{2}s_0)^N}{N!} = \sum_{n=1}^{\infty} \tfrac{a_n}{n^{s_0+1}} e^{(\log n)(1+\frac{1}{2}s_0)} = \sum_{n=1}^{\infty} a_n n^{-\frac{1}{2}s_0}.$$

In other words, the assumption $s_0 > 0$ led to a point between 0 and s_0 (namely $\frac{1}{2}s_0$) for which there is convergence. This contradiction proves that $s_0 = 0$. Therefore $\sum_{n=1}^{\infty} a_n n^{-s}$ converges for $\operatorname{Re} s > 0$. □

We shall now examine special features of Dirichlet series that allow the series to have product expansions like the one for $\zeta(s)$, namely $\sum_{n=1}^{\infty} n^{-s} = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}}$. Consider a formal product

$$\prod_{p \text{ prime}} (1 + a_p p^{-s} + \cdots + a_{p^m} p^{-ms} + \cdots).$$

If this product is expanded without regard to convergence, the result is the Dirichlet series $\sum_{n=1}^{\infty} a_n n^{-s}$, where $a_1 = 1$ and a_n is given by

$$a_n = a_{p_1^{r_1}} \cdots a_{p_k^{r_k}} \qquad \text{if } n = p_1^{r_1} \cdots p_k^{r_k}.$$

Suppose that the Dirichlet series $\sum_{n=1}^{\infty} a_n n^{-s}$ is in fact absolutely convergent in some right half-plane. Then every rearrangement is absolutely convergent to the same sum, and the same conclusion is valid for subseries. If E is a finite set of primes and if $\mathbb{N}(E)$ denotes the set of positive integers requiring only members of E for their factorization, then we have

$$\prod_{p \in E} (1 + a_p p^{-s} + \cdots + a_{p^m} p^{-ms} + \cdots) = \sum_{n \in \mathbb{N}(E)} a_n n^{-s}.$$

Letting E swell to the whole set of positive integers, we see that the infinite product has a limit in the half-plane of absolute convergence of the Dirichlet

series, and the limit of the infinite product equals the sum of the series. The sum of the series is 0 only if one of the factors on the left side is 0. In particular, the sum of the series cannot be identically 0, by Proposition 1.23d. Thus the limit of the infinite product can can be given by only this one Dirichlet series.

Conversely if an absolutely convergent Dirichlet series $\sum_{n=1}^{\infty} a_n n^{-s}$ has the property that its coefficients are **multiplicative**, i.e.,

$$a_1 = 1 \quad \text{and} \quad a_{mn} = a_m a_n \quad \text{whenever GCD}(m, n) = 1,$$

then we can form the above infinite product and recover the given series by expanding the product and using the formula $a_n = a_{p_1^{r_1}} \cdots a_{p_k^{r_k}}$ when $n = p_1^{r_1} \cdots p_k^{r_k}$. In this case we say that the Dirichlet series $\sum_{n=1}^{\infty} a_n n^{-s}$ has the infinite product as an **Euler product**. Many functions in elementary number theory give rise to multiplicative sequences; an example is $a_n = \varphi(n)$, where φ is the Euler φ function.

If the coefficients are **strictly multiplicative**, i.e., if

$$a_1 = 1 \quad \text{and} \quad a_{mn} = a_m a_n \quad \text{for all } m \text{ and } n,$$

then the p^{th} factor of the infinite product simplifies to

$$1 + a_p p^{-s} + \cdots + (a_p p^{-s})^m + \cdots = \frac{1}{1 - a_p p^{-s}}.$$

As a consequence we obtain the following proposition.

Proposition 1.26. If the coefficients of the Dirichlet series $\sum_{n=1}^{\infty} a_n n^{-s}$ are strictly multiplicative, then the Dirichlet series has an Euler product of the form

$$\sum_{n=1}^{\infty} \frac{a_n}{n^s} = \prod_{p \text{ prime}} \frac{1}{1 - a_p p^{-s}},$$

valid in its region of absolute convergence.

REMARK. We refer to the kind of Euler product in this proposition as a **first-degree Euler product**.

This is what happens with $\zeta(s)$, for which all the coefficients are 1, and with $a_n = \chi_0(n)$ and $a_n = \chi_1(n)$ as in the previous section. Conversely an Euler product expansion of the form in the proposition forces the coefficients of the Dirichlet series to be strictly multiplicative.

A Dirichlet series $\sum_{n=1}^{\infty} a_n n^{-s}$ with $|a_n| \le n^c$ for some real c is absolutely convergent for $\operatorname{Re} s > c + 1$. This fact leads us to a convergence criterion for first-degree Euler products.

Proposition 1.27. A first-degree Euler product $\prod(1 - a_p p^{-s})^{-1}$ with $|a_p| \leq p^c$ for some real c and all primes p defines an absolutely convergent Dirichlet series for $\operatorname{Re} s > c + 1$ and hence a valid identity $\sum_{n=1}^{\infty} a_n n^{-s} = \prod_{p \text{ prime}} (1 - a_p p^{-s})^{-1}$ in that region.

PROOF. The coefficients a_n are strictly multiplicative, and thus $|a_n| \leq n^c$ for all n. The absolute convergence follows. □

10. Dirichlet's Theorem on Primes in Arithmetic Progressions

In this section we shall prove Dirichlet's Theorem as stated in Theorem 1.21. Recall from Section 8 that the proof of Dirichlet's Theorem for the progressions $4n + 1$ and $4n + 3$ required taking the sum and difference of two expressions, working with them, and then passing back to the original expressions. Generalizing this step involves recognizing this process as Fourier analysis on the 2-element group $(\mathbb{Z}/4\mathbb{Z})^\times$. This kind of Fourier analysis was discussed in Section VII.4 of *Basic Algebra*. Let us begin by reviewing what is needed from that section of *Basic Algebra* and then pinpoint the Fourier analysis that was the key to the argument in Section 8.

Let G be a finite abelian group, such as $(\mathbb{Z}/m\mathbb{Z})^\times$. A **multiplicative character** of G is a homomorphism of G into the circle group $S^1 \subseteq \mathbb{C}^\times$. The multiplicative characters of G form a finite abelian group $\widehat{G}$ under pointwise multiplication:

$$(\chi\chi')(g) = \chi(g)\chi'(g).$$

In this setting we recall the statement of the Fourier inversion formula.

THEOREM 7.17 OF *Basic Algebra* (Fourier inversion formula). Let G be a finite abelian group, and introduce an inner product on the complex vector space $C(G, \mathbb{C})$ of all functions from G to $\mathbb{C}$ by the formula

$$\langle F, F' \rangle = \sum_{g \in G} F(g)\overline{F'(g)},$$

the corresponding norm being $\|F\| = \langle F, F\rangle^{1/2}$. Then the members of $\widehat{G}$ form an orthogonal basis of $C(G, \mathbb{C})$, each χ in $\widehat{G}$ satisfying $\|\chi\|^2 = |G|$. Consequently $|\widehat{G}| = |G|$, and any function $F : G \to C$ is given by the "sum of its Fourier series":

$$F(g) = \frac{1}{|G|} \sum_{\chi \in \widehat{G}} \Big(\sum_{h \in G} F(h)\overline{\chi(h)} \Big) \chi(g).$$

EXAMPLE. With the two-element group $G = \{\pm 1\}$, there are two multiplicative characters, with $\chi_0(+1) = \chi_0(-1) = 1$, $\chi_1(+1) = 1$, and $\chi_1(-1) = -1$. We can think of the Fourier-coefficient mapping as carrying any complex-valued function F on G to the function $\widehat{F}$ on $\widehat{G}$ given by $\widehat{F}(\chi) = \sum_{h\in G} F(h)\overline{\chi(h)}$. The inversion formula says that F is recovered as $F = \frac{1}{2}\big(\widehat{F}(\chi_0)\chi_0 + \widehat{F}(\chi_1)\chi_1\big)$. A basis for the 2-dimensional space of complex-valued functions on G consists of the two functions F^+ and F^-, with F^+ equal to 1 at $+1$ and 0 at -1 and with F^- equal to 0 at $+1$ and 1 at -1. The multiplicative characters are given by $\chi_0 = F^+ + F^-$ and $\chi_1 = F^+ - F^-$. For these two functions the inversion formula reads $F^+ = \frac{1}{2}(\chi_0 + \chi_1)$ and $F^- = \frac{1}{2}(\chi_0 - \chi_1)$. In Section 8 the roles of F^+ and F^- are played by functions of s, not by scalars, with F^+ corresponding to $\sum_{p\equiv 1 \bmod 4} p^{-s}$ and F^- corresponding to $\sum_{p\equiv 3 \bmod 4} p^{-s}$. We are to consider the functions of s corresponding to their sum χ_0 and to their difference χ_1. The results of Section 9 show that these are the series that come from Euler products. The role of the Fourier inversion formula is to ensure that we can reconstruct $\sum_{p\equiv 1 \bmod 4} p^{-s}$ and $\sum_{p\equiv 3 \bmod 4} p^{-s}$ from the sum and difference. The general proof of Dirichlet's Theorem is a direct generalization of this argument for $m = 4$.

Fix an integer $m > 1$. A **Dirichlet character modulo** m is a function $\chi : \mathbb{Z} \to S^1 \cup \{0\}$ such that

(i) $\chi(j) = 0$ if and only if $\mathrm{GCD}(j, m) > 1$,
(ii) $\chi(j)$ depends only on the residue class j mod m,
(iii) when regarded as a function on the residue classes modulo m, χ is a multiplicative character of $(\mathbb{Z}/m\mathbb{Z})^\times$.

In particular, a Dirichlet character modulo m determines a multiplicative character of $(\mathbb{Z}/m\mathbb{Z})^\times$. Conversely each multiplicative character of $(\mathbb{Z}/m\mathbb{Z})^\times$ defines a unique Dirichlet character modulo m as the lift of the multiplicative character on the set $\{j \in \mathbb{Z} \mid \mathrm{GCD}(j, m) = 1\}$ and as 0 on the rest of $\mathbb{Z}$. For example the multiplicative character on $(\mathbb{Z}/4\mathbb{Z})^\times$ that is 1 at 1 mod 4 and is -1 at 3 mod 4 lifts to the Dirichlet character that is 1 at integers congruent to 1 modulo 4, is -1 at integers congruent to 3 modulo 4, and is 0 at even integers. It will often be notationally helpful to use the same symbol for the Dirichlet character and the multiplicative character of $(\mathbb{Z}/m\mathbb{Z})^\times$. Because of this correspondence, the number of Dirichlet characters modulo m matches the order of $\widehat{G}$ for $G = (\mathbb{Z}/m\mathbb{Z})^\times$, which matches the order of G and is $\varphi(m)$, where φ is the Euler φ function. The **principal** Dirichlet character modulo m, denoted by χ_0, is the one built from the trivial character of $(\mathbb{Z}/m\mathbb{Z})^\times$:

$$\chi_0(j) = \begin{cases} 1 & \text{if } \mathrm{GCD}(j, m) = 1, \\ 0 & \text{if } \mathrm{GCD}(j, m) > 1. \end{cases}$$

Each Dirichlet character modulo m is strictly multiplicative, in the sense of the previous section. We assemble each as the coefficients of a Dirichlet series, the associated **Dirichlet L function**, by the definition

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}.$$

Proposition 1.28. Fix m, and let χ be a Dirichlet character modulo m.

(a) The Dirichlet series $L(s, \chi)$ is absolutely convergent for $\operatorname{Re} s > 1$ and is given in that region by a first-degree Euler product

$$L(s, \chi) = \prod_{p \text{ prime}} \frac{1}{1 - \chi(p)p^{-s}}.$$

(b) If χ is not principal, then the series for $L(s, \chi)$ is convergent for $\operatorname{Re} s > 0$, and the sum is analytic for $\operatorname{Re} s > 0$.

(c) For the principal Dirichlet character χ_0 modulo m, $L(s, \chi_0)$ extends to be meromorphic for $\operatorname{Re} s > 0$. Its only pole for $\operatorname{Re} s > 0$ is at $s = 1$, and the pole is simple. It is given in terms of the Riemann zeta function by

$$L(s, \chi_0) = \zeta(s) \prod_{\substack{p \text{ prime,} \\ p \text{ dividing } m}} (1 - p^{-s}).$$

PROOF. For (a), the boundedness of χ implies that the series is absolutely convergent for $\operatorname{Re} s > 1$. Since χ is strictly multiplicative, $L(s, \chi)$ has a first-degree Euler product by Proposition 1.26, and the product is convergent in the same region.

For (b), let us notice that $\chi \neq \chi_0$ implies the equality

$$\sum_{n=1}^{m} \chi(n + b) = 0 \qquad \text{for any } b, \tag{$*$}$$

since the member of $(\mathbb{Z}/m\mathbb{Z})^\times$ that corresponds to χ is orthogonal to the trivial character, by the Fourier inversion formula as quoted above from *Basic Algebra*. For s real and positive, let us write

$$\tfrac{\chi(n)}{n^s} = \chi(n) \cdot \tfrac{1}{n^s} = u_n v_n$$

in the notation of the summation by parts formula that follows the statement of Proposition 1.23, and let us put $U_n = \sum_{k=1}^{n} u_k$. Equation $(*)$ implies that $\{U_n\}$ is bounded, say with $|U_n| \le C$. Summation by parts then gives

$$\left| \sum_{n=M}^{N} \tfrac{\chi(n)}{n^s} \right| \le \sum_{n=M}^{N-1} C \left(\tfrac{1}{n^s} - \tfrac{1}{(n+1)^s} \right) + \tfrac{C}{N^s} + \tfrac{C}{M^s} = \tfrac{2C}{M^s}.$$

This expression tends to 0 as M and N tend to ∞. Therefore the series $L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$ is convergent for s real and positive. By Proposition 1.23a the series is convergent for $\operatorname{Re} s > 0$, and the sum is analytic in this region.

For (c), let $\operatorname{Re} s > 1$. From the product formula in (a) with χ set equal to χ_0, we have

$$L(s, \chi_0) = \prod_{\substack{p \text{ prime,} \\ p \text{ not dividing } m}} \tfrac{1}{1-p^{-s}}.$$

Using the Euler product expansion of $\zeta(s)$, we obtain the displayed formula of (c). The remaining statements in (c) follow from Proposition 1.24, since the product over primes p not dividing m is a finite product. □

By Proposition 1.28b, $L(s, \chi)$ is well defined and finite at $s = 1$ if χ is not principal. The main step in the proof of Dirichlet's Theorem is the following lemma.

Lemma 1.29. $L(1, \chi) \neq 0$ if χ is not principal.

PROOF. Let $Z(s) = \prod_{\chi} L(s, \chi)$. Exactly one factor of $Z(s)$ has a pole at $s = 1$, according to Proposition 1.28. If any factor has a zero at $s = 1$, then $Z(s)$ is analytic for $\operatorname{Re} s > 0$. Assuming that $Z(s)$ is indeed analytic, we shall derive a contradiction.

Being the finite product of absolutely convergent Dirichlet series for $\operatorname{Re} s > 1$, $Z(s)$ is given by an absolutely convergent Dirichlet series. We shall prove that the coefficients of this series are ≥ 0. More precisely we shall prove for $\operatorname{Re} s > 1$ that

$$Z(s) = \prod_{p \text{ with } \mathrm{GCD}(p,m)=1} \frac{1}{\left(1 - p^{-f(p)s}\right)^{g(p)}}, \qquad (*)$$

where $f(p)$ is the order of p in $(\mathbb{Z}/m\mathbb{Z})^{\times}$ and where $g(p) = \varphi(m)/f(p)$, φ being Euler's φ function. The factor $(1 - p^{-f(p)s})^{-1}$ is given by a Dirichlet series with all coefficients ≥ 0. Hence so is the $g(p)^{\text{th}}$ power, and so is the product over p of the result. Thus $(*)$ will prove that all coefficients of $Z(s)$ are ≥ 0.

To prove $(*)$, we write, for $\operatorname{Re} s > 1$,

$$Z(s) = \prod_{\chi} L(s, \chi) = \prod_{p} \Big(\prod_{\chi} \frac{1}{1 - \chi(p)p^{-s}} \Big) = \prod_{\substack{p \text{ with} \\ \mathrm{GCD}(p,m)=1}} \Big(\prod_{\chi} \frac{1}{1 - \chi(p)p^{-s}} \Big).$$

Fix p not dividing m. We shall show that

$$\prod_{\chi} \left(1 - \chi(p)p^{-s}\right) = \left(1 - p^{-fs}\right)^{g}, \qquad (**)$$

where f is the order of p in $(\mathbb{Z}/m\mathbb{Z})^\times$ and where $g = \varphi(m)/f$; then $(*)$ will follow.

The function $\chi \to \chi(p)$ is a homomorphism of $(\mathbb{Z}/m\mathbb{Z})^\times$ into the subgroup $\{e^{2\pi ik/f}\}$ of S^1 and is onto some cyclic subgroup $\{e^{2\pi ik/f'}\}$ with f' dividing f. Let us see that $f' = f$. In fact, if $f' < f$, then $p^{f'} \not\equiv 1 \bmod m$, while $\chi(p^{f'}) = \chi(p)^{f'} = 1$ for all χ; since $\chi(p^{f'}) = \chi(1)$ for all χ, the χ's cannot span all functions on $(\mathbb{Z}/m\mathbb{Z})^\times$, in contradiction to the Fourier inversion formula (Theorem 7.17 of *Basic Algebra*).

Thus $\chi \to \chi(p)$ is onto $\{e^{2\pi ik/f}\}$. In other words, $\chi(p)$ takes on all f^{th} roots of unity as values, and the homomorphism property ensures that each is taken on the same number of times, namely $g = \varphi(m)/f$ times. If X is an indeterminate, we then have

$$\prod_\chi (1 - \chi(p)X) = \Big(\prod_{k=0}^{f-1} (1 - e^{2\pi ik/f}X)\Big)^g = (1 - X^f)^g.$$

Then $(**)$ follows and so does $(*)$. Hence all the coefficients of the Dirichlet series of $Z(s)$ are ≥ 0. We have already observed that this series, as the finite product of absolutely convergent series for $\operatorname{Re} s > 1$, is absolutely convergent for $\operatorname{Re} s > 1$. Thus Proposition 1.25 applies and shows that the Dirichlet series of $Z(s)$ converges for $\operatorname{Re} s > 0$.

Since the coefficients of the series are positive, the convergence is absolute for s real and positive. By Proposition 1.23b the convergence is absolute for $\operatorname{Re} s > 0$. Therefore the Euler product expansion $(*)$ is valid for $\operatorname{Re} s > 0$.

For primes p not dividing m and for real $s > 0$, we have

$$\begin{aligned}\frac{1}{\big(1 - p^{-fs}\big)^g} &= (1 + p^{-fs} + p^{-2fs} + \cdots)^g \geq 1 + p^{-fgs} + p^{-2fgs} + \cdots \\ &= 1 + p^{-\varphi(m)s} + p^{-2\varphi(m)s} + \cdots = \frac{1}{1 - p^{-\varphi(m)s}}.\end{aligned}$$

In combination with $(*)$, this inequality gives

$$\begin{aligned}&Z(s)\Big(\prod_{p \text{ dividing } m} \frac{1}{1 - p^{-\varphi(m)s}}\Big) \\ &\qquad = \Big(\prod_{p \text{ with GCD}(p,m)=1} \frac{1}{(1 - p^{-fs})^g}\Big)\Big(\prod_{p \text{ dividing } m} \frac{1}{1 - p^{-\varphi(m)s}}\Big) \\ &\qquad \geq \prod_{p \text{ prime}} \frac{1}{1 - p^{-\varphi(m)s}} = \sum_{n=1}^\infty \frac{1}{n^{\varphi(m)s}}.\end{aligned}$$

The sum on the right is $+\infty$ for $s = 1/\varphi(m)$, while the left side is finite for that s. This contradiction completes the proof of the lemma. □

PROOF OF THEOREM 1.21. First we show for each Dirichlet character χ modulo m that

$$\log L(s,\chi) = \sum_{p \text{ prime}} \tfrac{\chi(p)}{p^s} + g(s,\chi) \tag{$*$}$$

for *real* numbers $s > 1$, with $g(s,\chi)$ remaining bounded as $s \downarrow 1$. In this statement we have not yet specified a branch of the logarithm, and we shall choose it presently. Fix p and define, for $s \geq 1$, a value of the logarithm of the p^{th} factor of the Euler product of $L(s,\chi)$ in Proposition 1.28a by

$$\log\left(\tfrac{1}{1-\chi(p)p^{-s}}\right) = \tfrac{\chi(p)}{p^s} + \tfrac{1}{2}\tfrac{\chi(p^2)}{p^{2s}} + \tfrac{1}{3}\tfrac{\chi(p^3)}{p^{3s}} + \cdots = \tfrac{\chi(p)}{p^s} + g(s,p,\chi). \tag{$**$}$$

In Section 8 we obtained the inequality $|\log(1-x)^{-1} - x| \leq 2|x|^2$ for real x with $|x| \leq \frac{1}{2}$, but the proof remains valid for complex x with $|x| \leq \frac{1}{2}$. Since $x = \chi(p)p^{-s}$ is complex with $|\chi(p)p^{-s}| \leq \frac{1}{2}$, we obtain

$$|g(s,p,\chi)| = \left|\log\left(\tfrac{1}{1-\chi(p)p^{-s}}\right) - \chi(p)p^{-s}\right| \leq 2|\chi(p)p^{-s}|^2 \leq 2p^{-2}.$$

Since $\sum_{p \text{ prime}} p^{-2} \leq \sum_{n=1}^{\infty} n^{-2} < \infty$, the series $\sum_p g(s,p,\chi)$ is uniformly convergent for $s \geq 1$. Let $g(s,\chi)$ be the continuous function $\sum_p g(s,p,\chi)$. Summing $(**)$ over primes p, we obtain

$$\sum_p \log\left(\tfrac{1}{1-\chi(p)p^{-s}}\right) = \sum_p \tfrac{\chi(p)}{p^s} + g(s,\chi).$$

Because of the validity of the Euler product expansion of $L(s,\chi)$ in Proposition 1.28a, the left side represents a branch of $\log L(s,\chi)$. This proves $(*)$.

For each b prime to m, define a function F_b on the positive integers by

$$F_b(n) = \begin{cases} 1 & \text{if } n \equiv b \bmod m, \\ 0 & \text{otherwise.} \end{cases}$$

The Fourier inversion formula (Theorem 7.17 of *Basic Algebra*) gives

$$\sum_\chi \overline{\chi(b)}\chi(n) = \varphi(m)F_b(n). \tag{$\dagger$}$$

Multiplying $(*)$ by $\overline{\chi(b)}$, summing on χ, and using $(\dagger)$ to handle the term that is summed over p prime, we obtain

$$\varphi(m) \sum_{\substack{p \text{ prime,} \\ p = km+b}} p^{-s} = \sum_\chi \overline{\chi(b)} \log L(s,\chi) - \sum_\chi \overline{\chi(b)} g(s,\chi). \tag{$\dagger\dagger$}$$

The term $\sum_\chi \overline{\chi(b)}g(s,\chi)$ is bounded as $s \downarrow 1$, according to $(*)$. The term $\overline{\chi_0(b)} \log L(s,\chi_0)$ is unbounded as $s \downarrow 1$, by Proposition 1.28c. For χ nonprincipal, the term $\overline{\chi(b)} \log L(s,\chi)$ is bounded as $s \downarrow 1$, by Proposition 1.28b and Lemma 1.29. Therefore the left side of $(\dagger\dagger)$ is unbounded as $s \downarrow 1$. Hence the number of primes contributing to the sum is infinite. □

11. Problems

1. Fix an odd integer $m > 1$. Let P be the set of odd primes $p > 0$ such that $x^2 \equiv m \bmod p$ is solvable and such that p does not divide m. Show that P is nonempty and that there is a finite set S of arithmetic progressions such that the members of P are the odd primes > 0 that lie in at least one member of S.

2. Let D be a nonsquare integer, and let m be an odd integer with $\mathrm{GCD}(D, m) = 1$. By suitably adapting the proof of Theorem 1.6,
   (a) prove that if m is primitively representable by some binary quadratic form of discriminant D, then $x^2 \equiv D \bmod m$ is solvable,
   (b) prove that if $x^2 \equiv D \bmod m$ is solvable and m is odd, then m is primitively representable by some binary quadratic form of discriminant D.

3. For a fixed discriminant D, let H be the group of proper equivalence classes of binary quadratic forms of discriminant D, and let H' be the set of ordinary equivalence classes of discriminant D. Inclusion of a proper equivalence class into the ordinary equivalence class that contains it gives a map f of H onto H'. Give an example in which H' can admit no group structure for which f is a group homomorphism.

4. (a) Show that if (a, b, c) has order 3 in the form class group, then the product of any two integers of the form $ax^2 + bxy + cy^2$ is again of that form.
   (b) Show that $h(-23) = 3$.
   (c) Using the general theory, show that the class of $2x^2 + xy + 3y^2$ has order 3.
   (d) Find an explicit formula for (X, Y) in terms of (x_1, y_1) and (x_2, y_2) such that $(2x_1^2 + x_1y_1 + 3y_1^2)(2x_2^2 + x_2y_2 + 3y_2^2) = 2X^2 + XY + 3Y^2$.

5. If two integer forms are improperly equivalent over $\mathbb{Z}$, prove that they are properly equivalent over $\mathbb{Q}$.

6. Verify for the fundamental discriminant $D = -67$ that $h(D) = 1$. (Educational note: It is known that the only negative fundamental discriminants D with $h(D) = 1$ are $-3, -4, -7, -8, -11, -19, -43, -67, -163$. It is known also that the only other nonsquare $D < 0$ for which $h(D) = 1$ are $-12, -16, -28, -27$.)

7. This problem carries out the algorithm suggested by Theorem 1.8 to find representatives of all proper equivalence classes of binary quadratic forms (a, b, c) of discriminant $316 = 4 \cdot 79$. For each of these, b will be even.
   (a) For each even positive b with $b < \sqrt{4 \cdot 79}$, factor $(b^2 - 4 \cdot 79)/4$ as a product ac in all possible ways such that $a > 0$ and such that both $|a|$ and $|c|$ lie between $\sqrt{79} - b/2$ and $\sqrt{79} + b/2$, obtaining 16 forms (a, b, c). Expand the list by adjoining each form $(-a, b, -c)$, so that the expanded list has 32 members.

(b) Arrange the 32 members of the expanded list of (a) into 6 cycles, obtaining 2 cycles of length 4 and 4 cycles of length 6.
(c) Conclude that $h(4 \cdot 79) = 6$.

8. For discriminant $D = -47$, the class number is $h(-47) = 5$, and the reduced binary quadratic forms are $(1, 1, 12)$, $(2, 1, 6)$, $(2, -1, 6)$, $(3, 1, 4)$, $(3, -1, 4)$. Show what the multiplication table is for the proper equivalence classes of these forms.

Problems 9–11 concern the Jacobi symbol, which is a generalization of the Legendre symbol. Let m and n be integers with $n > 0$ odd, and let $n = p_1^{k_1} \cdots p_r^{k_r}$ be the prime factorization of n. The **Jacobi symbol** $\left(\frac{m}{n}\right)$ is defined to be 0 if $\mathrm{GCD}(m, n) > 1$ and is defined to be $\prod_{j=1}^{r} \left(\frac{m}{p_j}\right)^{k_j}$ if $\mathrm{GCD}(m, n) = 1$, where $\left(\frac{m}{p_j}\right)$ is a Legendre symbol. The Jacobi symbol therefore extends the domain of the Legendre symbol, and it depends only on the residue m mod n. Even when $\mathrm{GCD}(m, n) = 1$, the Jacobi symbol does *not* encode whether m is a square modulo n, however, since $\left(\frac{-1}{21}\right) = +1$ and since the residue -1 is not a square modulo 21.

9. Suppose that n and n' are odd positive integers and that m and m' are integers. Verify that
(a) $\left(\frac{mm'}{nn'}\right) = \left(\frac{m}{n}\right)\left(\frac{m'}{n'}\right)$,
(b) $\left(\frac{m^2}{n}\right) = \left(\frac{m}{n^2}\right) = 1$ if $\mathrm{GCD}(m, n) = 1$.

10. Prove for all odd positive integers n that
(a) $\left(\frac{-1}{n}\right) = (-1)^{\frac{1}{2}(n-1)}$,
(b) $\left(\frac{2}{n}\right) = (-1)^{\frac{1}{8}(n^2-1)}$.

11. **(Quadratic reciprocity)** Prove for all odd positive integers m and n satisfying $\mathrm{GCD}(m, n) = 1$ that $\left(\frac{m}{n}\right) = (-1)^{[\frac{1}{2}(m-1)][\frac{1}{2}(n-1)]}\left(\frac{n}{m}\right)$.

Problems 12–13 indicate, without spelling out what the group G is, two uses of Dirichlet's Theorem in the subject of "elliptic curves." No knowledge of the subject of elliptic curves is assumed, however.

12. Suppose that G is a finite abelian group whose order $|G|$ divides $p + 1$ for all sufficiently large primes p with $p \equiv 3 \bmod 4$. It is to be shown that $|G|$ divides 4 by means of multiple applications of Dirichlet's Theorem.
(a) Deduce that 8 does not divide $|G|$ by considering the arithmetic progression $8k + 3$.
(b) Deduce that 3 does not divide $|G|$ by considering the arithmetic progression $12k + 7$.
(c) Deduce that no odd prime $q > 3$ divides $|G|$ by considering the arithmetic progression $4qk + 3$.

13. Suppose that G is a finite abelian group whose order $|G|$ divides $p+1$ for all sufficiently large primes p with $p \equiv 2 \bmod 3$. It is to be shown that $|G|$ divides 6 by means of multiple applications of Dirichlet's Theorem.
    (a) Deduce that 4 does not divide $|G|$ by considering the arithmetic progression $12k+5$.
    (b) Deduce that 9 does not divide $|G|$ by considering the arithmetic progression $9k+2$.
    (c) Deduce that no odd prime $q > 3$ divides $|G|$ by considering the arithmetic progression $3qk+2$.

Problems 14–19 develop some elementary properties of ideals and their norms in quadratic number fields. Notation is as in Sections 6–7. In particular, the number field is $K = \mathbb{Q}(\sqrt{m})$, the ring R of algebraic integers in it has $\mathbb{Z}$ basis $\{1, \delta\}$, and σ is the nontrivial automorphism of K fixing $\mathbb{Q}$.

14. Prove that if $I = \langle a, r\rangle$ is a nonzero ideal in R with $a \in \mathbb{Z}$ and $r \in R$, then a divides $N(s)$ for every s in I.

15. Prove that any nonzero ideal I in R can be written as $I = \langle a, b + g\delta\rangle$ with a, b, and g in $\mathbb{Z}$ and with $a > 0$, $0 \le b < a$, and $0 < g \le a$. Prove also that the $\mathbb{Z}$ basis with these properties is unique, and it has the properties that g divides a and b and that ag divides $N(b + g\delta)$.

16. Let a, b, and g be integers satisfying $a > 0$, $0 \le b < a$, and $0 < g \le a$ with g dividing a and b and with ag dividing $N(b + g\delta)$. Prove that the ideal $I = (a, b + g\delta)$ in R has $\{a, b + g\delta\}$ as a $\mathbb{Z}$ basis.

17. Prove that if $I = \langle a, r\rangle$ is a nonzero ideal in R with $a \in \mathbb{Z}$, $r \in R$, and $r = c + d\delta$ for integers c and d, then $N(I) = |ad|$.

18. (a) Prove that if I is a nonzero ideal in R, then $N(I)$ is the number of elements in R/I.
    (b) Deduce that if $I \subseteq J$ are nonzero ideals in R, then $N(J)$ divides $N(I)$, and $I = J$ if and only if $N(J) = N(I)$.

19. (a) Using the Chinese Remainder Theorem, prove that if I and J are nonzero ideals in R with $I + J = R$, then $N(IJ) = N(I)N(J)$.
    (b) Let P be a nonzero prime ideal in R, and let $p > 0$ be the prime number such that $P \cap \mathbb{Z} = (p)\mathbb{Z}$. Then R/P is a vector space over $\mathbb{Z}/p\mathbb{Z}$, and its order is of the form p^f for some integer $f > 0$. Show by induction on the integer $e > 0$ that R/P^e has order p^{ef}.
    (c) Using unique factorization of ideals, deduce that if I and J are any two nonzero ideals in R, then $N(IJ) = N(I)N(J)$.
    (d) Prove that any nonzero ideal I of R has $I\sigma(I) = (N(I))$.

Problems 20–24 concern the splitting of prime ideals when extended to quadratic number fields. Fix a quadratic number field $\mathbb{Q}(\sqrt{m})$, and let R, D, δ, and σ be as

in Sections 6–7. Let $p > 0$ be a prime in $\mathbb{Z}$. According to Theorem 9.62 of *Basic Algebra*, the unique factorization of the ideal $(p)R$ in R is one of the following: $(p)R = (p)$ is already prime in R, $(p)R = P_1P_2$ is the product of two distinct prime ideals, or $(p)R = P^2$ is the square of a prime ideal.

20. Deduce from the formula $N((p)R) = p^2$ that if P is a nontrivial factor in the unique factorization of the ideal $(p)R$, then $N(P) = p$.

21. This problem concerns the prime $p = 2$.
    (a) Use Problem 15 to prove that if $D \equiv 5 \bmod 8$, then $(2)R$ is a prime ideal in R.
    (b) Prove that if $D \equiv 1 \bmod 8$, then $(2)R$ factors into the product of two distinct prime factors as $(2)R = \langle 2, \delta\rangle\langle 2, 1+\delta\rangle$.
    (c) Prove that if D is even and $D/4 \equiv 3 \bmod 4$, then $(2)R = (2, 1+\delta)^2$ exhibits $(2)R$ as the square of a prime ideal.
    (d) Prove that if D is even and $D/4 \equiv 2 \bmod 4$, then $(2)R = (2, \delta)^2$ exhibits $(2)R$ as the square of a prime ideal.

22. Let p be an odd prime.
    (a) Prove that if D is odd, then $(p)R$ has a nontrivial factorization into prime ideals if and only if $x^2 + x + \frac{1}{4}(1 - D) \equiv 0 \bmod p$ has a solution, and in this case a factorization of $(p)R$ is as $(p)R = (p, x+\delta)(p, x+\sigma(\delta))$.
    (b) Prove that if D is even, then $(p)R$ has a nontrivial factorization into prime ideals if and only if $x^2 \equiv 0 \bmod (D/4)$ has a solution, and in this case a factorization of $(p)R$ is as $(p)R = (p, x+\delta)(p, x+\sigma(\delta))$.
    (c) Deduce from (a) and (b) that $(p)R$ has a nontrivial factorization into prime ideals if and only if D is a square modulo p.

23. Let p be an odd prime such that D is a square modulo p, so that Problem 22c gives a nontrivial factorization of $(p)R$ into prime ideals of the form $(p)R = (p, x+\delta)(p, x+\sigma(\delta))$ for some integer x. Let $I = (p, x+\delta)$.
    (a) Prove that if D is odd, then $\sigma(I) = I$ if and only if the integer x is $\frac{1}{2}(p-1)$.
    (b) Prove that if D is even, then $\sigma(I) = I$ if and only if the integer x is 0.

24. Let p be an odd prime such that D is a square modulo p, so that Problem 22c gives a nontrivial factorization of $(p)R$ into prime ideals of the form $(p)R = (p, x+\delta)(p, x+\sigma(\delta))$ for some integer x. Using the previous problem, show that the two factors on the right are the same ideal if and only if p divides D.

Problems 25–29 seek to identify the genus group explicitly for fundamental discriminants D. Let $K = \mathbb{Q}(\sqrt{m})$ be the corresponding quadratic number field, let R be the ring of algebraic integers in K, and let σ be the nontrivial automorphism of K fixing $\mathbb{Q}$. Let $E = \{p_1, \dots, p_{g+1}\}$ with $g \geq 0$ be the set of distinct prime divisors of D. The goal of this set of problems is to prove that the order of the genus group is 2^g and to exhibit ideals in R representing each genus. Recall from Theorem 1.20

that strict equivalence classes of ideals correspond to proper equivalence classes of binary quadratic forms and therefore that each genus corresponds to a set of proper equivalence classes of binary quadratic forms.

25. Let the form class group H for discriminant D be isomorphic to a product of cyclic groups of orders $2^{k_1}, \ldots, 2^{k_r}, q_1^{l_1}, \ldots, q_s^{l_s}$, where $k_1, \ldots, k_r$ and $l_1, \ldots, l_s$ are positive integers and $q_1, \ldots, q_s$ are odd primes that are not necessarily distinct. Prove that the genus group has order 2^r and is abstractly isomorphic to the subgroup of H of elements whose order divides 2. (Educational note: Thus a goal of the present set of problems is to show that $r = g$.)

26. According to Problems 20–24, the nonzero prime ideals of R are of three kinds:
    (i) unique distinct ideals $I = (p, b+\delta)$ and $\sigma(I) = (p, b+\sigma(\delta))$ with product $(p)R$ if p is an odd prime not dividing D such that $x^2 \equiv D \bmod p$ is solvable, or if $p = 2$ and $D \equiv 1 \bmod 8$,
    (ii) the ideal $(p)R$ if p is an odd prime not dividing D such that $x^2 \equiv D \bmod p$ is not solvable, or if $p = 2$ and $D \equiv 5 \bmod 8$,
    (iii) a unique ideal I_p with $I_p^2 = (p)R$ if p divides D.

    For each subset $S \subseteq E$ of the $g + 1$ distinct prime divisors of D, define $J_s = \prod_{p \in S} I_p$.
    (a) Using unique factorization of ideals in R, show that any nonzero proper ideal I in R with $\sigma(I) = I$ is of the form $(a)J_S$ for some $a \in \mathbb{Z}$ and some subset $S \subseteq E$.
    (b) By considering norms of ideals, show that I uniquely determines S in (a).

27. (a) The element $x = -1$ of K has $N(x) = 1$ and factors as $x = \sigma(y)y^{-1}$ for the element $y = \sqrt{m}$ of K. For all other elements x of K with norm 1, verify the formula

$$\frac{1+x}{1+\sigma(x)} = \frac{(1+x)x}{(1+\sigma(x))x} = \frac{(1+x)x}{x+x\sigma(x)} = \frac{(1+x)x}{1+x} = x,$$

    and explain why it shows that x is of the form $\sigma(y)y^{-1}$ for some $y \neq 0$ in K. (Educational note: This result is a special case of **Hilbert's Theorem 90**, which is a theorem in the cohomology of groups and appears in Chapter III. The general theorem says for a finite Galois extension K/k with Galois group Γ that the cohomology H^1 of the group Γ with coefficients in the abelian group $K^\times$ is 0.)
    (b) Show that the element y in (a) can be taken to be in R and that all such y's in R are $\mathbb{Z}$ multiples of one of them y_0, which is unique up to a factor of -1.

28. Let I be a nonzero ideal in R whose class in the ideal class group $\mathcal{H}$ has order 2, i.e., an ideal I such that $I^2 = (x)$ for some element $x \in R$.
    (a) Show that the element $xN(I)^{-1}$ of K has norm 1.

(b) Show that the corresponding element y_0 of R from the previous problem has the property that $\sigma((y_0)I) = (y_0)I$.

(c) Using either y_0 or $y_0\sqrt{m}$ from (b), deduce that for any nonzero ideal I in R with I^2 principal, there is a *strictly* equivalent ideal J_S for some subset $S \subseteq E$ of the $g+1$ prime divisors of E. Consequently the order of the genus group is a power of 2 equal to at most 2^{g+1}.

29. This problem shows that the number of ideals J_S in the previous problem that are mutually strictly inequivalent is exactly 2^g. To get at this fact, the problem investigates properties of principal ideals $I = (x)$ in R with the properties that $\sigma(I) = I$ and $N(x) > 0$. Since $\sigma(I) = I$, it must be true that $\sigma(x) = \varepsilon x$ for some unit ε in R, and then $N(\sigma(x)) = N(x)$ implies that $N(\varepsilon) = +1$. Matters now split into cases along the lines of the hypotheses of Proposition 1.17.

(a) Under the assumption that $m < 0$ and that m is neither -1 nor -3, show that if a principal ideal $I = (x)$ in R has $\sigma(I) = I$, then x is in $\mathbb{Z}$ or in $\mathbb{Z}\sqrt{m}$.

(b) Under the assumption that $m < 0$, show that the only subsets S of E for which the ideal J_S is principal are $S = \varnothing$ and S equal to the set of all prime divisors of m, i.e., S equal to E for D odd and for D even with $D/4 \equiv 2 \bmod 4$ and S equal to $E - \{2\}$ for D even with $D/4 \equiv 2 \bmod 4$.

(c) Under the assumption that $m < 0$, Proposition 1.17 says that strict equivalence for ideals coincides with equivalence. Show how to conclude from this fact and the results of (a) and (b) that the order of the genus group is 2^g when $m < 0$.

(d) Under the assumption that $m > 0$ and that the fundamental unit ε_1 has norm -1, Proposition 1.17 says that strict equivalence for ideals coincides with equivalence. With I, x, and ε as in the statement of the problem, show that $\varepsilon = \pm\varepsilon_1^{2n}$ for some integer $n \geq 0$. Deduce that $\sigma(\varepsilon_1^n x) = s\varepsilon_1^n x$ for a suitable choice of sign s, and show as a consequence that J_S is principal for the same S's as in (b) and that the order of the genus group is 2^g.

(e) Under the assumption that $m > 0$ and that the fundamental unit ε_1 has norm $+1$, Proposition 1.17 says that strict equivalence for ideals is distinct from equivalence; in particular, there are two strict equivalence classes of principal ideals: those with a generator of positive norm and those with a generator of negative norm. Let y_0^+ and y_0^- be the elements produced by Problem 27 that satisfy $\varepsilon_1 = \sigma(y_0^+)(y_0^+)^{-1}$ and $-\varepsilon_1 = \sigma(y_0^-)(y_0^-)^{-1}$. Prove that exactly one of y_0^+ and y_0^- has positive norm, so that two of the principal ideals (1), (y_0^+), (y_0^-), $(\sqrt{m}\,)$ are strictly equivalent to (1), and two are not. Prove that all four of these principal ideals are of the form J_S and that they are distinct. By expressing elements arising from Problem 27 for the most general unit in R in terms of y_0 and ε_1, show that no other J_S is a principal ideal. Show as a consequence that the number of strict equivalence classes of ideals among the J_S's is 2^g.

Problems 30–34 show that proper equivalence over $\mathbb{Q}$ for two integer forms of fundamental discriminant D implies proper equivalence over $\mathbb{Z}/D\mathbb{Z}$. Consequently the order of the genus group is at most the number of classes of integer forms of discriminant D under proper equivalence over $\mathbb{Z}/D\mathbb{Z}$. It will follow from the next set of problems, concerning "genus characters," that the number of such classes is at least 2^g, where $g+1$ is the number of distinct prime divisors of D. In combination with Problem 29, this result shows that the number of genera equals 2^g. Throughout this set of problems, let D be a fundamental discriminant.

30. Let (a_1, b_1, c_1) be a binary quadratic form over $\mathbb{Z}$ of discriminant D. Using Lemma 1.10, prove that (a_1, b_1, c_1) is properly equivalent over $\mathbb{Z}$ to a form (a, b, c) of discriminant D such that $\mathrm{GCD}(a, D) = 1$.

31. Suppose that (a, b, c) is a binary quadratic form over $\mathbb{Z}$ of discriminant D such that $\mathrm{GCD}(a, D) = 1$.
    (a) Prove that if D is odd, then (a, b, c) is properly equivalent over $\mathbb{Z}$ to a form (a, kD, lD) for some integers k and l.
    (b) Prove that if D is even, then (a, b, c) is properly equivalent over $\mathbb{Z}$ to a form $(a, 2kD, -a(D/4) + lD)$ for some integers k and l.

32. Suppose that (a, kD, lD) is a form over $\mathbb{Z}$ having odd discriminant D, satisfying $\mathrm{GCD}(a, D) = 1$, and taking on an integer value r relatively prime to D for some rational (x, y). Write x and y as fractions with a positive common denominator as small as possible: $x = u/w$ and $y = v/w$.
    (a) Prove that $\mathrm{GCD}(w, D) = 1$, and conclude that $a \equiv d^2 r \bmod D$ for some integer d relatively prime to D.
    (b) Suppose that $(a', k'D, l'D)$ is a second form over $\mathbb{Z}$ having discriminant D, satisfying $\mathrm{GCD}(a', D) = 1$, and taking on the value r at some rational point. Prove that $a' \equiv as^2 \bmod D$ for some s relatively prime to D.
    (c) Suppose that (a, b, c) and (a', b', c') are forms over $\mathbb{Z}$ of the same odd discriminant with $\mathrm{GCD}(a, D) = \mathrm{GCD}(a', D) = 1$, and suppose that these forms are properly equivalent over $\mathbb{Q}$. Deduce that (a, b, c) and (a', b', c') are properly equivalent over $\mathbb{Z}/D\mathbb{Z}$ in the sense that there exists a matrix $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ in $\mathrm{SL}(2, \mathbb{Z}/D\mathbb{Z})$ such that substitution of $x = \alpha x' + \beta y'$ and $y = \gamma x' + \delta y'$ leads from $ax^2 + bxy + cy^2$ modulo D to $a'x'^2 + b'x'y' + c'y'^2$ modulo D.

33. Suppose that $(a, 2kD, -a(D/4) + lD)$ is a form over $\mathbb{Z}$ having even discriminant D, satisfying $\mathrm{GCD}(a, D) = 1$, and taking on an integer value r relatively prime to D for some rational (x, y). Write x and y as fractions with a positive common denominator as small as possible: $x = u/w$ and $y = v/w$.
    (a) Prove that $\mathrm{GCD}(w, D) = 1$, and obtain a congruence relating a and r modulo D.

(b) Suppose that $(a', 2k'D, -a'(D/4) + l'D)$ is a second form over $\mathbb{Z}$ having discriminant D, satisfying $\mathrm{GCD}(a', D) = 1$, and taking on the value r at some rational point. Prove that $\left(\frac{a}{p}\right) = \left(\frac{a'}{p}\right)$ for every odd prime p dividing D.

(c) In the setting of (b), suppose in addition that $D/4 \equiv 3 \bmod 4$. Prove that $a \equiv a' \bmod 4$.

(d) In the setting of (b), suppose in addition that $D/4 \equiv 2 \bmod 4$. Prove for $D/4 \equiv 2 \bmod 8$ that $a' \equiv \pm a \bmod 8$, and prove for $D/4 \equiv 6 \bmod 8$ that either $a' \equiv a \bmod 8$ or $a' \equiv 3a \bmod 8$.

(e) Suppose that (a, b, c) and (a', b', c') are forms over $\mathbb{Z}$ of the same even discriminant with $\mathrm{GCD}(a, D) = \mathrm{GCD}(a', D) = 1$, and suppose that these forms are properly equivalent over $\mathbb{Q}$. Deduce that (a, b, c) and (a', b', c') are properly equivalent over $\mathbb{Z}/D\mathbb{Z}$.

34. Why does it follow from Problems 30–33 that the order of the genus group for discriminant D is at least as large as the number of proper equivalence classes under $\mathrm{SL}(2, \mathbb{Z}/D\mathbb{Z})$ of integer forms of discriminant D?

Problems 35–40 introduce "genus characters." In fact, genus characters are already implicit in Problems 32 and 33. Throughout this set of problems, let D be a fundamental discriminant, and suppose that D has exactly $g + 1$ distinct prime factors. The content of these problems will be summarized in Problem 40. Call two binary quadratic forms over $\mathbb{Z}$ of discriminant D **similar modulo** D if they take on the same residues r modulo D that are relatively prime to D. Proper equivalence over $\mathbb{Z}$ via $\mathrm{SL}(2, \mathbb{Z})$ implies proper equivalence modulo D via $\mathrm{SL}(2, \mathbb{Z}/D\mathbb{Z})$, and this in turn implies similarity modulo D in the sense that was just defined. Problems 30–31 show that it is enough to study forms $ax^2 \bmod D$ for D odd, where $\mathrm{GCD}(a, D) = 1$, and to study forms $a(x^2 - (D/4)y^2)$ for D even, again where $\mathrm{GCD}(a, D) = 1$. Initially the **genus characters** are functions of pairs (similarity class, r), where r is a residue modulo D with $\mathrm{GCD}(r, D) = 1$ such that r is represented by the form modulo D. The values of these functions are $\left(\frac{r}{p}\right)$ for each odd prime $p > 0$ dividing D, as well as the indicated one of the following for $p = 2$ if D is even:

$$\begin{aligned}
\xi(r) = \left(\tfrac{-1}{r}\right) &= (-1)^{\frac{1}{2}(r-1)} && \text{if } D \text{ is even and } D/4 \equiv 3 \bmod 4,\\
\eta(r) = \left(\tfrac{2}{r}\right) &= (-1)^{\frac{1}{8}(r^2-1)} && \text{if } D \text{ is even and } D/4 \equiv 2 \bmod 8,\\
\xi(r)\eta(r) = \left(\tfrac{-2}{r}\right) &= (-1)^{\frac{1}{2}(r-1)+\frac{1}{8}(r^2-1)} && \text{if } D \text{ is even and } D/4 \equiv 6 \bmod 8.
\end{aligned}$$

Thus $g + 1$ expressions have been defined for each ordered pair (similarity class, r).

35. Using Problems 32 and 33, show that the genus characters are independent of the residue r modulo D with $\mathrm{GCD}(r, D) = 1$ such that r is represented by the form modulo D. Therefore the residue a in the quadratic form, either $ax^2 \bmod D$ for D odd or $a(x^2 - (D/4)y^2)$ for D even, can be used as r, and the genus characters are $g + 1$ functions defined on the set of similarity classes modulo D.

36. Prove that the genus characters respect the operation of multiplication of proper equivalence classes of forms over $\mathbb{Z}$.

37. The product of all $g+1$ genus characters is 1 in every case. A sketch of the argument for D odd is as follows: Since $D \equiv 1 \bmod 4$, D has an even number $2t$ of prime factors $4k+3$. Use of the Jacobi symbol with a odd and p varying over the (odd) prime divisors of D gives

$$\prod_p \left(\tfrac{a}{p}\right) = \prod_{p=4k+1} \left(\tfrac{a}{p}\right) \prod_{p=4k+3} \left(\tfrac{a}{p}\right) = \xi(a)^{2t} \prod_{p=4k+1} \left(\tfrac{p}{a}\right) \prod_{p=4k+3} \left(\tfrac{p}{a}\right) = \left(\tfrac{D}{a}\right),$$

and the right side is $+1$ by Problem 2a. Using this sketch as a guide, show that the product of all $g+1$ genus characters is 1 for the cases that D is even and
(a) $D/4 \equiv 3 \bmod 4$,
(b) $D/4 \equiv 2 \bmod 8$,
(c) $D/4 \equiv 6 \bmod 8$.

38. If D is even, let α be ξ if $D/4 \equiv 3 \bmod 4$, η if $D/4 \equiv 2 \bmod 8$, and $\xi\eta$ if $D/4 \equiv 6 \bmod 8$. Let $p \mapsto s_p$ be any function to $\{\pm 1\}$ from the set of distinct prime divisors of D. Using Dirichlet's Theorem on primes in arithmetic progressions, prove that there exists a prime q such that $\left(\frac{q}{p}\right) = s_p$ for each odd prime divisor p of D and $\alpha(q) = s_2$ in case D is even.

39. With α as in the previous problem, let $p \mapsto s_p$ be any function to $\{\pm 1\}$ from the set of distinct prime divisors of D such that $\prod_p s_p = +1$, and choose a prime q as in the previous problem. Prove that q is primitively representable by some integer binary quadratic form of discriminant D and that the values of the genus characters on this form are the numbers s_p. Conclude that the number of distinct similarity classes modulo D is at least 2^g.

40. For the quadratic number field $K = \mathbb{Q}(\sqrt{m})$ with discriminant D, suppose that D has $g+1$ distinct prime divisors. Conclude that the following equivalence classes of binary quadratic forms over $\mathbb{Z}$ of discriminant D coincide and that the number of such classes is 2^g:
    (i) classes relative to proper equivalence over $\mathbb{Q}$, i.e., genera,
    (ii) classes relative to proper equivalence over $\mathbb{Z}/D\mathbb{Z}$,
    (iii) classes relative to similarity modulo D.