Chapter 12
Quadratic forms and sums of squares

Our goal in this chapter is to develop the theory of quadratic forms so we can give another proof of Theorem 9.8, especially in the three square case. Our exposition follows [31, Part 3, Chap IV] closely. We start with the basic theory of quadratic forms and explain the notion of equivalence. We then discuss the concept of representability of an integer by a quadratic form. Since the goal of the chapter is to give a proof of the Three Square Theorem we set the stage by giving a proof of the Two Squares Theorem in §12.2. In this section we develop the theory of binary quadratic forms with integral coefficients, determine representatives for the equivalence classes of positive definite binary quadratic forms of a given discriminant, and use this knowledge to prove the Two Squares Theorem. In the next two sections we develop the analogous theory for ternary quadratic forms and prove the Three Squares Theorem. In the Notes to this chapter, we explain Gauss's beautiful composition law for binary quadratic forms.

12.1 Quadratic forms with integral coefficients

In Chapters 5 and 9 we determined what numbers can be represented as a sum of two, three, or four squares. One way to view these results is to think of them as theorems about the numbers that are represented by certain quadratic forms. For example, if we let

$$f(x, y) = x^2 + y^2,$$

then Theorem 5.2 tells us what $f(\mathbb{Z}^2)$ is. This is an example of a *quadratic form with integral coefficients*.

Definition 12.1. Let $A = (a_{ij})_{1 \le i,j \le n}$ be an $n \times n$ symmetric matrix with integer entries. We call a function $f : \mathbb{Z}^n \to \mathbb{Z}$ defined by

$$f(x_1, \ldots, x_n) = \sum_{\substack{1 \le i \le n \\ 1 \le j \le n}} a_{ij} x_i x_j$$

R. Takloo-Bighash, *A Pythagorean Introduction to Number Theory*,
Undergraduate Texts in Mathematics, https://doi.org/10.1007/978-3-030-02604-2_12

a *quadratic form with integral coefficients associated to the matrix* A. We define the *discriminant of the form* f, denoted disc f, to be the determinant of the matrix A. We call a quadratic form f with integral coefficients *primitive* if it is not an integral multiple of another quadratic form with integral coefficients.

For example, if $n = 2$ and

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

with $a, b, c \in \mathbb{Z}$, then the quadratic form associated to A is

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

It is easy to check that

$$f(x, y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = v^T A v$$

with $v = \begin{pmatrix} x \\ y \end{pmatrix}$. This is of course a completely general fact: If f is the quadratic form associated to the matrix A, then

$$f(x_1, \ldots, x_n) = v^T A v \tag{12.1}$$

with $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ the column vector with entries $x_1, \ldots, x_n$.

Lemma 12.2. *The quadratic form f uniquely determines the matrix A.*

Proof. Suppose f is associated to matrices $A = (a_{ij})_{1 \le i,j \le n}$ and $A' = (a'_{ij})_{1 \le i,j \le n}$. Then we have

$$v^T A v = v^T A' v \tag{12.2}$$

for all v. We will prove $A = A'$ by induction on n. If $n = 1$, then

$$a_{11} x_1^2 = a'_{11} x_1^2$$

for all $x_1 \in \mathbb{Z}$ immediately implies $a_{11} = a'_{11}$. Now suppose the lemma is true for $n - 1$. Let $w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{n-1} \end{pmatrix} \in \mathbb{Z}^{n-1}$ be a column vector, and for each $1 \le j \le n$ let $w(j) = \begin{pmatrix} w_1(j) \\ w_2(j) \\ \vdots \\ w_n(j) \end{pmatrix}$ be the vector in $\mathbb{Z}^n$ which is defined as follows:

$$w_i(j) = \begin{cases} w_i & i < j; \\ 0 & i = j; \\ w_{i-1} & i \geq j. \end{cases}$$

For example, if $n = 3$ and $w = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, then

$$w(3) = \begin{pmatrix} x \\ y \\ 0 \\ z \end{pmatrix}.$$

Next, for each $n \times n$ matrix $B = (b_{kl})_{1 \leq k,l \leq n}$ and each $1 \leq j \leq n$ define a matrix $B(j)$ to be the $(n-1) \times (n-1)$ matrix which is obtained from B by deleting the jth rows and jth column of B, i.e., if we write $B(j) = (b_{kl}(j))_{1 \leq k,l \leq n-1}$, then

$$b_{kl}(j) = \begin{cases} b_{kl} & k, l < j; \\ b_{k,l+1} & k < j, l > j; \\ b_{k+1,l} & k > j, l < j; \\ b_{k+1,l+1} & k > j, l > j. \end{cases}$$

For example, if

$$B = \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{pmatrix},$$

then

$$B(1) = \begin{pmatrix} f & g & h \\ j & k & l \\ n & o & p \end{pmatrix}, \quad B(3) = \begin{pmatrix} a & b & d \\ e & f & h \\ m & n & p \end{pmatrix}.$$

The importance of the matrix $B(j)$ lies in the fact that for each $w \in \mathbb{Z}^{n-1}$ and each $B \in \mathrm{M}_n(\mathbb{Z})$ we have

$$w(j)^T B w(j) = w^T B(j) w. \tag{12.3}$$

Now we go back to Equation (12.2), and apply it to column vectors of the form $w(j)$, $1 \leq j \leq n$. For each j we have

$$w^T A(j) w = w(j)^T A w(j) = w(j)^T A' w(j) = w^T A'(j) w.$$

Since we are assuming the lemma is true for $n-1$, this last equation implies that for each j,

$$A(j) = A'(j).$$

The assertion now follows from Exercise 12.2. □

Since the matrix A is symmetric and $x_i x_j = x_j x_i$ for all i, j, we have

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} a_{ii} x_i^2 + 2 \sum_{1 \le i < j \le n} a_{ij} x_i x_j,$$

This points to a caveat in our theory, namely that the quadratic forms that we consider have *even* coefficients for their "mixed" terms, i.e., the terms of the form $x_i x_j$ with $i \neq j$. This means that our theory does not include quadratic forms like

$$x^2 + xy + y^2, \quad x^2 + y^2 + z^2 + 3xy + 4xz.$$

One way to avoid this problem is to consider matrices that are not symmetric, or by allowing the off diagonal terms in A be half integers, but either of these ideas brings about complications that we do not want to deal with in this book. We refer the reader to Cassels [12] for a more thorough treatment of quadratic forms over the field of rational numbers.

Definition 12.3. For quadratic forms f and g with integral coefficients, we say f is *equivalent* to g, and write $f \sim g$, if there is a matrix $P = (p_{ij}) \in \mathrm{SL}_n(\mathbb{Z})$ such that

$$f(\sum_{j=1}^{n} p_{1j} x_j, \sum_{j=1}^{n} p_{2j} x_j, \ldots, \sum_{j=1}^{n} p_{nj} x_j) = g(x_1, x_2, \ldots, x_n).$$

For example if $f(x, y) = x^2 + y^2$ and $g(x, y) = x^2 + 2xy + 2y^2$, then $f \sim g$. The reason is that $f(x + y, y) = g(x, y)$, i.e., the definition holds with $P = \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$.

In the above notation, note that

$$\begin{pmatrix} \sum_{j=1}^{n} p_{1j} x_j \\ \sum_{j=1}^{n} p_{2j} x_j \\ \vdots \\ \sum_{j=1}^{n} p_{nj} x_j \end{pmatrix} = P. \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Now if we suppose f and g are associated to the matrices A and B, respectively, then $f \sim g$ means

$$(Pv)^T A (Pv) = v^T B v$$

for all $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{Z}^n$. Since transposition is order reversing, $(XY)^T = Y^T X^T$, this equation now implies

$$v^T (P^T A P) v = v^T B v.$$

Lemma 12.2 says

$$P^T A P = B. \tag{12.4}$$

It is clear that this process can be reversed, meaning if there is $P \in \mathrm{SL}_n(\mathbb{Z})$ such that Equation (12.4) holds, then $f \sim g$. We summarize this discussion as the following lemma:

Lemma 12.4. *Suppose f, g are quadratic forms associated to matrices A, B. Then $f \sim g$ if and only if there is $P \in \mathrm{SL}_n(\mathbb{Z})$ such that*

$$P^T A P = B.$$

This lemma has the following important consequence:

Proposition 12.5. *The relation $\sim$ on quadratic forms is an equivalence relation that preserves the discriminant.*

Proof. We need to show that $\sim$ is symmetric, reflexive, and transitive, and that if $f \sim g$, then $\det f = \det g$. We use Lemma 12.4 repeatedly.

Reflexive. We need: $f \sim f$. Clearly I_n, the $n \times n$ identity matrix, is in $\mathrm{SL}_n(\mathbb{Z})$, and $A = I_n^T A I_n$.

Symmetry. We need: $f \sim g$ implies $g \sim f$. Suppose f and g are associated to A, B, respectively. If there is a matrix $P \in \mathrm{SL}_n(\mathbb{Z})$ such that $P^T A P = B$, then since $(P^T)^{-1} = (P^{-1})^T$, $(P^{-1})^T B (P^{-1}) = A$, and $P^{-1} \in \mathrm{SL}_n(\mathbb{Z})$. This means $g \sim f$.

Transitive. We need: $f \sim g$ and $g \sim h$ implies $f \sim h$. Suppose f, g, h are associated to A, B, C, respectively, and that there are $P, Q \in \mathrm{SL}_n(\mathbb{Z})$ such that $P^T A P = B$ and $Q^T B Q = C$. Then

$$C = Q^T B Q = Q^T P^T A P Q = (PQ)^T A (PQ).$$

Determinant preservation. We need: $f \sim g$ implies $\det f = \det g$. Suppose f, g are associated to A, B, respectively, and that there is $P \in \mathrm{SL}_n(\mathbb{Z})$ such that $B = P^T A P$. We have

$$\operatorname{disc} g = \det B = \det(P^T A P) = \det(P^T) \det P \det A$$

by multiplicativity of determinant. Then we note that $\det P^T = \det P = 1$ as transposition does not change the value of determinant. This means that

$$\operatorname{disc} g = \det A = \operatorname{disc} f.$$

□

Definition 12.6. For a quadratic form f and an integer m, we say f *represents* m if there are integers $x_1, \ldots, x_n$ such that

$$f(x_1, \ldots, x_n) = m.$$

We call f *positive definite* if for all $x_1, \ldots, x_n \in \mathbb{Z}^n$, not all of which are zero, we have

$$f(x_1, \ldots, x_n) > 0.$$

The following proposition is central to our discussion:

Proposition 12.7. *Suppose f, g are quadratic forms, and $f \sim g$. Then*

1. *The quadratic forms f and g represent the exact same set of numbers.*
2. *The quadratic form f is positive definite if and only if the quadratic form g is.*

Proof. Following the notation of Equation (12.1) write

$$f(x_1, \ldots, x_n) = v^T A v, \quad g(x_1, \ldots, x_n) = v^T B v$$

with $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ the column vector with entries $x_1, \ldots, x_n$. For simplicity we write $f(v)$ and $g(v)$ instead of $f(x_1, \ldots, x_n)$ and $g(x_1, \ldots, x_n)$, respectively. The assumption on the f and g means there is a matrix $P \in \mathrm{SL}_n(\mathbb{Z})$ such that $B = P^T A P$. In terms of f and g this means that for all v, $g(v) = f(P \cdot v)$. As a result, $g(\mathbb{Z}^n) = f(P \cdot \mathbb{Z}^n)$. Once we show $P \cdot \mathbb{Z}^n = \mathbb{Z}^n$, the first assertion follows. Since P has integer entries, $P \cdot \mathbb{Z}^n \subset \mathbb{Z}^n$. Similarly, since $P \in \mathrm{SL}_n(\mathbb{Z})$, P^{-1}, too, has integer entries. Therefore, $P^{-1} \cdot \mathbb{Z}^n \subset \mathbb{Z}^n$. Multiplying by P gives $\mathbb{Z}^n \subset P \cdot \mathbb{Z}^n$. Putting the inclusions $P \cdot \mathbb{Z}^n \subset \mathbb{Z}^n$ and $\mathbb{Z}^n \subset P \cdot \mathbb{Z}^n$ together gives $P \cdot \mathbb{Z}^n = \mathbb{Z}^n$, and we are done with the first part. The second statement follows from the first statement, and the statement that for $v \in \mathbb{Z}^n$, $v = 0$ if and only if $Pv = 0$. □

Lemma 12.8. *If for a quadratic form f,* disc f *is square-free, then f is primitive. Equivalence preserves primitivity.*

Proof. If $f = mg$, then disc $f = m^n$ disc g. This observation implies the first assertion. The second statement is obvious. □

12.2 Binary forms

We now discuss the case where $n = 2$, the so-called binary forms, in detail. Here we do not address questions of representability of integers by binary forms. The wonderful book Cox [14], especially Chapter 1, provides an accessible introduction to this important topic.

Suppose we have a binary quadratic form f which is associated to the symmetric matrix

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Then $\operatorname{disc} f = \det A = ac - b^2$.

Lemma 12.9. *The form f is positive definite if and only if $a > 0$ and* $\operatorname{disc} f > 0$.

Proof. Suppose f is positive definite. Since $f(1, 0) = a$ we immediately see $a > 0$. Next,

$$0 < f(-b, a) = ab^2 - 2b^2a + ca^2 = -b^2a + ca^2 = a(ac - b^2) = a\operatorname{disc} f.$$

Since we have already established $a > 0$, $a\operatorname{disc} f > 0$ implies $\operatorname{disc} f > 0$.

Now suppose $a > 0$ and $\operatorname{disc} f > 0$. Then

$$af(x, y) = a^2x^2 + 2abxy + acy^2 = (a^2x^2 + 2abxy + b^2y^2) + (ac - b^2)y^2$$

$$= (ax + by)^2 + (\operatorname{disc} f)y^2.$$

Since $a > 0$ and $\operatorname{disc} f > 0$, the identity

$$af(x, y) = (ax + by)^2 + (\operatorname{disc} f)y^2 \tag{12.5}$$

shows that $f(x, y) \geq 0$, and $f(x, y) = 0$ only if $(\operatorname{disc} f)y^2 = 0$ and $(ax + by)^2 = 0$, which immediately implies $x = y = 0$. This means f is positive definite. □

Theorem 12.10. *Every equivalence class of positive definite binary quadratic forms contains a form f whose associated matrix $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ satisfies*

$$2|b| \leq a \leq c.$$

Proof. Suppose we have a positive definite form g associated to $A_g = \begin{pmatrix} a_0 & b_0 \\ b_0 & c_0 \end{pmatrix}$. We wish to show that there is a form f with $f \sim g$ for which the inequalities of the theorem hold. Let a be the smallest positive number represented by g. There are integers r, t such that $g(r, t) = a$. We claim $\gcd(r, t) = 1$. Otherwise, if $p \mid r$ and $p \mid t$, then $p^2 \mid a$, and we would have $g(r/p, t/p) = a/p^2$, and that contradicts the choice of a. Since $\gcd(r, t) = 1$, there are integers s, u such that $ru - st = 1$. By Theorem 2.23 if we fix one solution s_0, u_0 every other solution is of the form

$$s(h) = s_0 + rh, \quad u(h) = u_0 + ht, \quad h \in \mathbb{Z}.$$

Now consider the functions $a(h), b(h)$, and $c(h)$, for $h \in \mathbb{Z}$, defined by the following matrix identity

$$\begin{pmatrix} a(h) & b(h) \\ b(h) & c(h) \end{pmatrix} = \begin{pmatrix} r & s(h) \\ t & u(h) \end{pmatrix}^T \begin{pmatrix} a_0 & b_0 \\ b_0 & c_0 \end{pmatrix} \begin{pmatrix} r & s(h) \\ t & u(h) \end{pmatrix}.$$

Explicitly, we have

$$\begin{cases} a(h) = a_0 r^2 + 2b_0 rt + c_0 t^2 = a, \\ b(h) = s(h)(ra_0 + tb_0) + u(h)(rb_0 + tc_0), \\ c(h) = a_0 s(h)^2 + 2b_0 s(h)u(h) + c_0 u(h)^2. \end{cases}$$

Simplification gives

$$\begin{aligned} b(h) &= s_0(a_0 r + b_0 t) + u_0(b_0 r + c_0 t) + (a_0 r^2 + 2b_0 rt + c_0 t^2)h \\ &= s_0(a_0 r + b_0 t) + u_0(b_0 r + c_0 t) + ah. \end{aligned}$$

Since the coefficient of h is $a > 0$, and h is arbitrary, we may choose an h_0 so that $b(h_0)$ satisfies $|b(h_0)| \leq a/2$. The expression for $c(h)$ shows that

$$c(h_0) = g(s(h_0), u(h_0)),$$

and consequently $a \leq c(h_0)$. It is clear that the quadratic form associated to the matrix

$$\begin{pmatrix} a(h_0) & b(h_0) \\ b(h_0) & c(h_0) \end{pmatrix}$$

satisfies the requirements. □

Definition 12.11. A primitive binary form $f(x, y) = ax^2 + 2bxy + cy^2$ is called *reduced* if its coefficients satisfy the inequalities of Theorem 12.10.

For example, the forms $x^2 + y^2$ and $4x^2 + 2xy + 5y^2$ are reduced, and $5x^2 + 2xy + 4y^2$ is not.

Corollary 12.12. *Every positive definite binary quadratic form of discriminant* 1 *is equivalent to* $x^2 + y^2$.

Proof. By Theorem 12.10 and Proposition 12.5 every such quadratic form is equivalent to a quadratic form whose associated matrix $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ satisfies $2|b| \leq a \leq c$ and $ac - b^2 = 1$. Then we have

$$a^2 \leq ac = b^2 + 1 \leq \frac{a^2}{4} + 1.$$

Consequently, $a^2 \leq 4/3$. From this inequality it follows that $a = 1$. Since $2|b| \leq 1$, we see $b = 0$. Since $ac = b^2 + 1 = 1$, we see $c = 1$. □

Let us now use this last result to give another proof for Theorem 5.7, namely that every prime of the form $4k + 1$ is a sum of two squares.

One more proof of Theorem 5.7. Suppose p is of the form $4k + 1$. We wish to show that p is represented by the binary quadratic form $x^2 + y^2$. Since by Corollary 12.12 every positive definite binary form of discriminant 1 is equivalent to $x^2 + y^2$, and by

Proposition 12.7 equivalent forms represent the same set of numbers, it suffices to find some positive definite binary form

$$ax^2 + 2bx + cy^2$$

with discriminant 1 which represents p. We will show that we may even take $a = p$, see Exercise 12.7. Clearly, the form

$$g(x, y) = px^2 + 2bxy + cy^2$$

represents p, as $g(1, 0) = p$. We just need to choose b, c so that disc $g = 1$. We have

$$\operatorname{disc} g = pc - b^2.$$

As a result, the existence of b, c is equivalent to $b^2 \equiv -1 \bmod p$, or $(-1/p) = +1$. But for p of the form $4k + 1$ this is a consequence of Equation (6.3). □

12.3 Ternary forms

In this section we study quadratic forms in three variables. Our goal here is to prove the analogue of Corollary 12.12 in this setting. Namely, we will prove:

Theorem 12.13. *Every positive definite ternary quadratic form of discriminant* 1 *is equivalent to* $x^2 + y^2 + z^2$.

The proof of this theorem, though in principle similar to the proof of Corollary 12.12, is fairly complicated. The reader might want to skip the rest of this subsection in the first reading and go straight to §12.4 where the Three Square Theorem is proved.

Theorem 12.14. *Suppose* f *is a ternary quadratic form associated to the symmetric matrix*

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Then f *is positive definite if and only if*

- $a_{11} > 0$;
- $\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} > 0$;
- $\det A > 0$.

Before we prove the theorem, we need a lemma that is the analogue of Equation (12.5) for ternary forms:

Lemma 12.15. *With notations as above,*

$$a_{11} f(x, y, z) = (a_{11}x + a_{12}y + a_{13}z)^2 + K(y, z)$$

with $K(y,z)$ a binary quadratic form associated to the matrix

$$\begin{pmatrix} a_{11}a_{22} - a_{12}^2 & a_{11}a_{23} - a_{12}a_{13} \\ a_{11}a_{23} - a_{12}a_{13} & a_{11}a_{33} - a_{13}^2 \end{pmatrix}.$$

Furthermore, $\operatorname{disc} K = a_{11} \operatorname{disc} f$. *Finally, if f is positive definite, K will be positive definite.*

Proof. Every statement in the lemma, except for the last one, is a straightforward computation; see Exercise 12.10. The last statement follows from Lemma 12.9. □

We can now prove the theorem:

Proof of Theorem 12.14. Since $a_{11} = f(1,0,0)$, we see that $a_{11} > 0$ if f is positive definite. So we will assume $a_{11} > 0$.

If f is positive definite, Lemma 12.15 implies that K is positive definite. Lemma 12.9, applied to K, implies that $a_{11}a_{22} - a_{12}^2 > 0$ and $\operatorname{disc} K = a_{11} \operatorname{disc} f > 0$. These are the conditions required by the theorem.

Conversely, suppose the inequalities of the theorem are satisfied. Then, as above, it follows that K is positive definite. Suppose, to achieve a contradiction, that f is not positive definite. Then for some $(x,y,z) \neq (0,0,0)$, $f(x,y,z) \leq 0$. Then we have

$$(a_{11}x + a_{12}y + a_{13}z)^2 + K(y,z) \leq 0.$$

Since K is positive definite, this equation implies $K(y,z) = 0$ and $a_{11}x + a_{12}y + a_{13}z = 0$. The first of these implies $y = z = 0$, and then we conclude $x = 0$ as well. □

Our next theorem is the analogue of Exercise 12.3 for ternary forms.

Theorem 12.16. *Every positive definite ternary quadratic form f of discriminant d is equivalent to some quadratic form g whose associated matrix $A = (a_{ij})$ satisfies*

$$a_{11} \leq \frac{4}{3}\sqrt[3]{d}, \quad 2|a_{12}| \leq a_{11}, \quad 2|a_{13}| \leq a_{11}.$$

Proof. Suppose f is associated to the matrix B, and let a_{11} be the smallest natural number represented by f. Then there are integers c_{11}, c_{21}, c_{31} such that

$$a_{11} = f(c_{11}, c_{21}, c_{31}).$$

As in the proof of Theorem 12.10 we have

$$\gcd(c_{11}, c_{21}, c_{31}) = 1.$$

Exercise 12.11 shows that there is a 3×3 matrix $C = (c_{ij})$ whose first column is the numbers c_{11}, c_{21}, c_{31} and whose determinant is 1. Let g be the quadratic form whose associated matrix is

$$D = C^T B C.$$

Now

$$g(1,0,0) = f(c_{11}, c_{21}, c_{31}) = a_{11}.$$

Next, consider a form h whose associated matrix is

$$E = \begin{pmatrix} 1 & r & s \\ 0 & t & u \\ 0 & v & w \end{pmatrix}^T D \begin{pmatrix} 1 & r & s \\ 0 & t & u \\ 0 & v & w \end{pmatrix}.$$

Here we assume $r, s, t, u, v, w \in \mathbb{Z}$, and $tw - uv = 1$, so that for every r, s the determinant of the transformation matrix is 1.

We write $D = (b_{kl})$ and $E = (a_{kl})$. If

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & r & s \\ 0 & t & u \\ 0 & v & w \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

then one can check

$$b_{11}x_1 + b_{12}x_2 + b_{13}x_3 = a_{11}y_1 + a_{12}y_2 + a_{13}y_3.$$

Now we apply Lemma 12.15 to obtain positive definite binary forms K and L such that

$$a_{11}g(x_1, x_2, x_3) = (b_{11}x_1 + b_{12}x_2 + b_{13}x_3)^2 + K(x_2, x_3)$$

and

$$a_{11}h(y_1, y_2, y_3) = (a_{11}y_1 + a_{12}y_2 + a_{13}y_3)^2 + L(y_2, y_3).$$

The form K is transformed to L via $\begin{pmatrix} t & u \\ v & w \end{pmatrix}$. The form L has discriminant $a_{11}\operatorname{disc} f$, and the coefficient of y_2^2 is $a_{11}a_{22} - a_{12}^2$. Consequently, by Exercise 12.3 we can choose u, v, w, t such that

$$a_{11}a_{22} - a_{12}^2 \le \frac{2}{\sqrt{3}}\sqrt{a_{11}d}.$$

It is easy to see that

$$a_{12} = ra_{11} + tb_{12} + vb_{13}$$

and

$$a_{13} = sa_{11} + ub_{12} + wb_{13}.$$

Since r, s are arbitrary, we can choose them so that

$$|a_{12}| \le a_{11}/2, \quad |a_{13}| \le a_{11}/2.$$

Also, since $a_{22} = h(0, 1, 0)$, we must have $a_{22} \ge a_{11}$. Hence,

$$a_{11}^2 \le a_{11}a_{22} = (a_{11}a_{22} - a_{12}^2) + a_{12}^2 \le \frac{2}{\sqrt{3}}\sqrt{a_{11}d} + \frac{a_{11}^2}{4},$$

from which it immediately follows that

$$a_{11} \leq \frac{4}{3}\sqrt[3]{d}. \qquad \square$$

Now we proceed to prove the main theorem of this section:

Proof of Theorem 12.13. By Theorem 12.16 we know that our quadratic form is equivalent to a form whose associated matrix has the properties

$$a_{11} \leq 4/3, \quad 2|a_{12}| \leq a_{11}, \quad 2|a_{13}| \leq a_{11}.$$

Clearly, $a_{11} = 1$, $a_{12} = 0$, and $a_{13} = 0$. Consequently, our form is equivalent to a form

$$g = x_1^2 + K(x_2, x_3)$$

with K a positive definite binary quadratic form of discriminant 1. By Corollary 12.12 there is a transformation $\begin{pmatrix} t & u \\ v & w \end{pmatrix}$ that sends K to $x_2^2 + x_3^2$. Finally, $\begin{pmatrix} 1 & 0 & 0 \\ 0 & t & u \\ 0 & v & w \end{pmatrix}$ sends g to $x_1^2 + x_2^2 + x_3^2$, and we are done. $\square$

12.4 Three squares

In this section we give a proof of the most non-trivial part of Theorem 9.8. Namely, we will prove that if n is not of the form $4^a(8k + 7)$ then n is a sum of three squares. Clearly if $n = x^2 + y^2 + z^2$, then $4n = (2x)^2 + (2y)^2 + (2z)^2$, so we may factor out any factor 4^m from n and assume that either n is odd or it is twice an odd number. This means that we may assume

$$n \equiv 1, 2, 3, 5, 6 \mod 8.$$

Theorem 12.13 and Proposition 12.7 imply that it suffices to find a positive definite ternary form of discriminant 1 that represents n. This means we need to find a 3×3 matrix (a_{ij}) with integer entries and three integers x_1, x_2, x_3 such that

$$a_{11} > 0, \quad a_{11}a_{22} - a_{12}^2 > 0, \quad \det(a_{ij}) = 1,$$

and

$$n = \sum_{ij} a_{ij}x_i x_j.$$

We take

$$a_{13} = a_{31} = 1, \quad a_{23} = a_{32} = 0, \quad a_{33} = n, \quad x_1 = 0, \quad x_2 = 0, \quad x_3 = 1.$$

Then if we set $b = a_{11}a_{22} - a_{12}^2$, computing the determinant of (a_{ij}) using the bottom row gives

$$1 = \det(a_{ij}) = \det \begin{pmatrix} a_{11} & a_{12} & 1 \\ a_{21} & a_{22} & 0 \\ 1 & 0 & n \end{pmatrix} = -a_{22} + n \det \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$$
$$= -a_{22} + nb.$$

So we just need

- $a_{11} > 0$;
- $b = a_{11}a_{22} - a_{12}^2 > 0$;
- $a_{22} = bn - 1$.

If $n > 1$, then $a_{11} > 0$ is a consequence of the other statements. The reason for this is that

$$a_{22} = bn - 1 > b - 1 \geq 0,$$

and

$$a_{11}a_{22} = a_{12}^2 + b > 0.$$

The latter implies $a_{11} > 0$. So we need

- $b = a_{11}a_{22} - a_{12}^2 > 0$;
- $a_{22} = bn - 1$,

or, equivalently, we need to show that there is $b > 0$ such that the equation

$$X^2 \equiv -b \mod (bn - 1)$$

has a solution. We separate the cases where n is even or odd.

The even case: $n \equiv 2, 6 \mod 8$. Since $\gcd(4n, n - 1) = 1$, Dirichlet's Arithmetic Progression Theorem, Theorem 5.11, shows that there is a natural number v such that

$$p = 4nv + n - 1 = (4v + 1)n - 1$$

is prime. Note that $p \equiv 1 \mod 4$. Let $b = 4v + 1 > 0$. By Theorem 7.3 we have

$$\left(\frac{-b}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{b}{p}\right) = \left(\frac{b}{p}\right) = \left(\frac{p}{b}\right) = \left(\frac{bn - 1}{b}\right) = \left(\frac{-1}{b}\right) = +1.$$

The odd case: $n \equiv 1, 3, 5 \mod 8$. First let us assume $n \equiv 3 \mod 8$. Then $(n-1)/2$ is odd, and consequently, $\gcd(4n, (n - 1)/2) = 1$. By Dirichlet's Arithmetic Progression Theorem, Theorem 5.11, there is an integer v such that

$$p = 4nv + \frac{n - 1}{2} = \frac{(8v + 1)n - 1}{2}$$

is prime, and $p \equiv 1 \mod 4$. Set $b = 8v + 1$. Then $b > 0$ and $2p = bn - 1$. Since $b \equiv 1 \mod 8$, by Theorem 7.3, $(-2/b) = 1$. Then

$$\left(\frac{-b}{p}\right) = \left(\frac{b}{p}\right) = (-1)^{\frac{b-1}{2}\cdot\frac{p-1}{2}}\left(\frac{p}{b}\right) = \left(\frac{p}{b}\right) = \left(\frac{p}{b}\right)\left(\frac{-2}{b}\right)$$
$$= \left(\frac{-2p}{b}\right) = \left(\frac{1-nb}{b}\right) = \left(\frac{1}{b}\right) = 1.$$

If $n \equiv 1, 5 \bmod 8$, then we consider primes of the form $p = 4nv + \frac{3n-1}{2}$, and we let $b = 8v + 3$. The remainder of the argument is completely similar; see Exercise 12.13.

Exercises

12.1 Verify Equation (12.3).

12.2 This exercise uses the notations of the proof of Lemma 12.2. Suppose $A, A' \in M_n(R)$ for some ring R, and suppose for all j we have $A(j) = A'(j)$. Show that $A = A'$.

12.3 Show that every positive definite binary quadratic form of discriminant d is equivalent to a quadratic form whose associated matrix $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ satisfies

$$2|b| \leq a \leq \frac{2}{\sqrt{3}}\sqrt{d}.$$

12.4 Show that a reduced binary quadratic form cannot be equivalent to a different reduced binary quadratic form.

12.5 Show that for every natural number d there are only finitely many equivalence classes of positive definite binary quadratic forms of discriminant d.

12.6 Find representatives for equivalence classes of positive definite binary quadratic forms of discriminant d when

a. $d = 2$;
b. $d = 3$;
c. $d = 5$.

12.7 We say that a binary form f represents m *properly* if there are $a, b \in \mathbb{Z}$ with $\gcd(a, b) = 1$ such that $f(a, b) = m$. Show that a binary quadratic form represents an integer m properly if and only if it is equivalent to a binary form $mx^2 + bxy + cy^2$ for some $b, c \in \mathbb{Z}$.

12.8 Find reduced forms that are equivalent to the following forms:

a. $4x^2 + y^2$;
b. $9x^2 + 2xy + y^2$;
c. $126x^2 + 74xy + 13y^2$.

12.9 (✠) List all reduced primitive positive definite binary quadratic forms of discriminant bounded by 100. For each d, find the number of forms with that discriminant.

12.10 Prove Lemma 12.15.

12.11 Suppose $a, b, c \in \mathbb{Z}$ are such that $\gcd(a, b, c) = 1$. Then prove that there are integers d, e, f, g, h, i such that the matrix

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

has determinant 1.

12.12 Prove that the Three Square Theorem implies the Four Square Theorem.

12.13 Finish the proof of the Three Square Theorem for $n \equiv 1, 5 \bmod 8$.

12.14 Show that if $p > 17$ is a prime number $p \equiv 5 \bmod 12$ then p is a sum of three distinct positive squares. Hint: Use the identity,

$$9(a^2 + b^2) = (2a - b)^2 + (2a + 2b)^2 + (2b - a)^2.$$

Notes

Gauss Composition

The easy identity

$$(x^2 + y^2)(z^2 + w^2) = (xz + yw)^2 + (xw - zy)^2 \tag{12.6}$$

has been known for hundreds of years. As we noted in the Notes to Chapter 3, the master Indian mathematician Brahmagupta discovered the more general identity

$$(x^2 + dy^2)(z^2 + dw^2) = (xz + dyw)^2 + d(xw - yz)^2 \tag{12.7}$$

at some point in the seventh century CE. Over a thousand years later, Lagrange discovered the identities

$$(2x^2 + 2xy + 3y^2)(2z^2 + 2zw + 3w^2) = (2xz + xw + yz + 3yw)^2 + 5(xw - yz)^2, \tag{12.8}$$

and

$$(3x^2 + 2xy + 5y^2)(3z^2 + 2zw + 5w^2) = (3x^2 + xw + yz + 5yw)^2 + 14(xw - yz)^2. \tag{12.9}$$

All of these identities are of the form

$$f(x, y)f(z, w) = g(B_1(x, y, z, w), B_2(x, y, z, w)); \tag{12.10}$$

with f and g positive definite binary quadratic forms of the same discriminant, and B_1, B_2 homogeneous quadratic forms in the four variables x, y, z, w. The binary quadratic forms in Equation (12.6) have discriminant 1, in Equation (12.7) they have discriminant d, in Equation (12.8) they have discriminant 5, and in Equation (12.9) they have discriminant 14. Gauss proved a truly impressive theorem that generalizes all such identities. In fact, he showed the following theorem: Let f_1, f_2 be positive definite binary quadratic forms of discriminant d. Then there are homogeneous polynomials B_1, B_2 of degree 2 in the variables x, y, z, t such that

$$f_1(x, y) f_2(z, w) = g(B_1(x, y, z, w), B_2(x, y, z, w));$$

for some positive definite binary quadratic form g of discriminant d. Gauss called the quadratic form g the *composition of* f_1 *and* f_2, and for that reason the theorem is called the *composition law*. The binary quadratic forms we studied in this chapter all had an even middle coefficient, i.e., they were of the form $ax^2 + 2bxy + cy^2$ with b an integer. Gauss considered the more general quadratic forms $ax^2 + bxy + cy^2$ with b integral. For such forms the discriminant as we defined it is not necessarily an integer, so the discriminant is generally defined to be $4ac - b^2 \in \mathbb{Z}$. Gauss illustrated his theory with the following example:

$$\begin{aligned}&(4x^2 + 3xy + 5y^2)(3z^2 + zw + 6w^2)\\ &= (xz - 3xw - 2yz - 3yw)^2 + (xz - 3xw - 2yz - 3yw)(xz + xw + yz - yw)\\ &\quad +9(xz + xw + yz - yw)^2.\end{aligned}$$

Let us denote the composition of the forms f_1 and f_2 by $f_1 \circ f_2$. An important feature of Gauss's composition is that if f_1 is equivalent to a form f_1', then $f_1 \circ f_2 \sim f_1' \circ f_2$. This means that the composition provides a well-defined operation on the finite set of equivalence classes of binary quadratic forms of discriminant d, turning it into a finite abelian group, the *class group of binary forms*. It was Dirichlet who interpreted the composition of binary quadratic forms in terms of ideal multiplication, whereby connecting the class group of binary forms to the ideal class group of modern algebraic number theory. After about 200 years since the publication of [21], in a series of groundbreaking works, Manjul Bhargava generalized the Gauss composition laws and found numerous other composition laws. Gauss's proof of his composition law is *extremely* complicated; see [21, Ch. V]. Cox [14, §3] contains a motivated introduction to Gauss's theory of quadratic forms. We refer the reader to Andrew Granville's lecture at a summer school in 2014 for a review of Gauss's work and the works of other mathematicians that preceded it, as well as an introduction to Bhargava's works:

http://www.crm.umontreal.ca/sms/2014/pdf/granville1.pdf