Chapter I

# Foundations

## 1. Quadratic Forms and Quadratic Spaces

Throughout this book, a field always means a field of characteristic different from 2, unless it is stated otherwise.

An ($n$-ary) *quadratic form* over a field $F$ is a polynomial $f$ in $n$ variables over $F$ that is homogeneous of degree 2. It has the general form

$$f(X_1, \ldots, X_n) = \sum_{i,j=1}^{n} a_{ij} X_i X_j \in F[X_1, \ldots, X_n] = F[X].$$

To render the coefficients symmetric, it is customary to rewrite $f$ as

$$f(X) = \sum_{i,j} \tfrac{1}{2}(a_{ij} + a_{ji}) X_i X_j = \sum_{i,j} a'_{ij} X_i X_j,$$

where $a'_{ij} = \frac{1}{2}(a_{ij} + a_{ji})$. In this way, $f$ determines uniquely a *symmetric* matrix $(a'_{ij})$, which we shall denote by $M_f$. In terms of matrix notations, we have

$$f(X) = (X_1, \ldots, X_n) \cdot M_f \cdot \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = X^t \cdot M_f \cdot X \quad (t = \text{ transpose})$$

where $X$ is viewed as a column vector.

Let $f$ and $g$ be $n$-ary quadratic forms. We say that $f$ is *equivalent* to $g$ ($f \cong g$) if there exists an invertible matrix $C \in \mathrm{GL}_n(F)$ such that $f(X) = g(C \cdot X)$. This means that there exists a nonsingular, homogeneous linear substitution of the variables $X_1, \ldots, X_n$ that takes the form $g$ to the

form $f$. Since

$$g(C\cdot X) = (C\cdot X)^t\cdot M_g\cdot(C\cdot X) = X^t\cdot(C^t\cdot M_g\cdot C)\cdot X,$$

the equivalence condition $f(X) = g(C\cdot X)$ stipulated above amounts to a matrix equation[1]

$$M_f = C^t\cdot M_g\cdot C.$$

Thus, *equivalence* of forms amounts to *congruence* of the associated symmetric matrices. For example, if we perform the homogeneous linear substitution, $X_1 \mapsto X_1 + X_2$, $X_2 \mapsto X_1 - X_2$, the binary form $g(X_1, X_2) = X_1X_2$ goes to

$$\begin{aligned} g(C\cdot X) &= g(X_1 + X_2, X_1 - X_2) = (X_1 + X_2)(X_1 - X_2) \\ &= X_1^2 - X_2^2. \end{aligned} \tag{1.1}$$

Thus $g$ is equivalent to $f(X_1, X_2) = X_1^2 - X_2^2$. The corresponding matrix equation reads:

$$M_f = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = C^t\cdot M_g\cdot C.$$

As expected, the equivalence of forms is indeed an equivalence relation.

Let $F^n$ denote the space of $n$-tuples, given the usual $F$-vector space structure. Let $e_1, \ldots, e_n$ be the standard basis for $F^n$ given by the unit vectors. Any quadratic form, $f$, gives rise to a map $Q_f : F^n \to F$ defined by sending a column tuple

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum x_ie_i$$

to $Q_f(x) := x^t\cdot M_f\cdot x \in F$. We shall refer to $Q_f$ as the *quadratic map* defined by $f$. In terms of quadratic maps, the notion of equivalence of forms, $f \cong g$, amounts to the existence of a linear automorphism $C$ of $F^n$ such that $Q_f(x) = Q_g(C\cdot x)$, for every column tuple $x$.

Note that the quadratic map $Q_f$ determines uniquely the quadratic form $f$ (not only the equivalence class of $f$). In fact, suppose $Q_f = Q_g$ as maps from $F^n$ to $F$. For any $i$, we have

$$(M_f)_{ii} = Q_f(e_i) = Q_g(e_i) = (M_g)_{ii}.$$

For $i \neq j$, we have

$$Q_f(e_i + e_j) = (M_f)_{ii} + (M_f)_{jj} + 2(M_f)_{ij},$$

[1] To fully justify this equation, we should note, of course, that $C^t\,M_g\,C$ remains a symmetric matrix.

and a similar equation for $Q_g(e_i + e_j)$. It follows immediately (since $F$ has characteristic not 2) that $(M_f)_{ij} = (M_g)_{ij}$. Thus $M_f = M_g$, and $f = g \in F[X]$.

Let us examine the "quadratic map" $Q_f$ more closely. It has the following remarkable properties:

(1) $Q_f$ is "quadratic" in the sense that

$$Q_f(ax) = a^2 Q_f(x) \quad \text{(for all } x \in F^n, \text{ and } a \in F).$$

This follows since $Q_f(ax) = (ax)^t M_f (ax) = a^2 (x^t M_f x)$.

(2) If we "polarize" $Q_f$ by setting

$$B_f(x, y) = [Q_f(x + y) - Q_f(x) - Q_f(y)]/2,$$

then $B_f : F^n \times F^n \to F$ is a *symmetric bilinear* pairing. Here, symmetry is clear, and bilinearity follows easily from the observation that

$$\begin{aligned} B_f(x, y) &= [(x + y)^t M_f (x + y) - x^t M_f x - y^t M_f y]/2 \\ &= [x^t M_f y + y^t M_f x]/2 = x^t M_f y. \end{aligned}$$

Note that, in (2), the quadratic map $Q_f$ may be recaptured from the symmetric bilinear pairing $B_f$ by "depolarization"; that is,

$$Q_f(x) = B_f(x, x), \quad \text{for any } x \in F^n.$$

This remark motivates a slightly more "geometric" (coordinate-free) approach to the notion of a quadratic form. Let $V$ be any finite-dimensional $F$-vector space, and $B : V \times V \to F$ a symmetric bilinear pairing on $V$. We shall call the pair $(V, B)$ a "quadratic space," and associate with it a "quadratic map," $q = q_B : V \to F$, given by $q(x) = B(x, x)$ $(x \in V)$. As in (1) and (2) above, we have $q(ax) = B(ax, ax) = a^2 B(x, x) = a^2 q(x)$, and

$$\begin{aligned} q(x + y) - q(x) - q(y) &= B(x + y, x + y) - B(x, x) - B(y, y) \\ &= B(x, y) + B(y, x) \\ &= 2B(x, y). \end{aligned}$$

Since $q$ and $B$ determine each other, it is legitimate to write $(V, q)$ to represent the quadratic space $(V, B)$. If we coordinatize $V$, that is, choose a basis $e_1, \ldots, e_n$ for it, then the quadratic space $(V, B)$ gives rise to a quadratic form

$$f(X_1, \ldots, X_n) = \sum_{i,j} B(e_i, e_j) X_i X_j, \quad \text{with } M_f = (B(e_i, e_j)).$$

Note that if we identify $V$ with $F^n$ via the given coordinatization, then $q = q_B$ corresponds precisely to the quadratic map $q_f$ associated with the form $f$. If we coordinatize $V$ differently, by choosing another basis, $e'_1, \ldots, e'_n$, the

quadratic form $f'$ resulting from the new coordinatization will be equivalent to $f$. In fact, if $e'_i = \sum_k c_{ki} e_k$, then

$$\begin{aligned}(M_{f'})_{ij} &= B\left(\sum_k c_{ki}e_k, \sum_\ell c_{\ell j}e_\ell\right)\\ &= \sum_{k,\ell} c_{ki} \cdot B(e_k, e_\ell) \cdot c_{\ell j}\\ &= (C^t \cdot M_f \cdot C)_{ij},\end{aligned}$$

where $C = (c_{k\ell})$. Thus the quadratic space $(V, B)$ determines uniquely an *equivalence class* of quadratic forms, which we shall denote by $(f_B)$.

If $(V, B), (V', B')$ are quadratic spaces, we say that they are *isometric* ($\cong$) if there exists a linear isomorphism $\tau : V \to V'$ such that

$$B'(\tau(x), \tau(y)) = B(x, y) \quad \text{for all } x, y \in V;$$

that is, $\tau$ is "inner product preserving". It is clear that

$$(V, B) \cong (V', B') \iff (f_B) = (f_{B'}).$$

Thus, there is a one-one correspondence between the equivalence classes of $n$-ary quadratic forms and the isometry classes of $n$-dimensional quadratic spaces. In proving theorems, *we shall often prefer to argue geometrically with quadratic spaces*, and then pass back freely to quadratic forms, viewing the above one-one correspondence as an identification.

Let $(V, B)$ be a quadratic space, and let $M$ be a symmetric matrix associated to one of the forms in the equivalence class $(f_B)$. The following elementary fact from linear algebra is well-known.

**Proposition 1.2.** *The following statements are equivalent*:

(1) *$M$ is a nonsingular matrix.*

(2) *$x \mapsto B(\ , x)$ defines an isomorphism $V \to V^*$, where $V^*$ denotes the vector space dual of $V$.*

(3) *For $x \in V$, $B(x, y) = 0$ for all $y \in V$ implies that $x = 0$.*

If one (and hence all) of these statements holds, we shall say that $(V, B)$ is a *regular* (or nonsingular) quadratic space, or that $q_B$ is a *nonsingular* quadratic form. [Note: The zero quadratic space satisfies (2) and (3) above (forget about (1)!), hence should be called regular, also.]

Let $(V, B)$ be a quadratic space, and $S$ be a subspace of $V$. Then $(S,\ B|(S \times S))$ is a quadratic space in its own right. As usual, the *orthogonal complement* of $S$ is defined by

$$S^\perp = \{x \in V \mid B(x, S) = 0\}.$$

The orthogonal complement of $V$ itself is called the "radical" of $(V, B)$, denoted by $V^\perp = \operatorname{rad} V$. Observe that $(V, B)$ is regular iff $\operatorname{rad} V = 0$. However, if $(V, B)$ is regular, the subspace $S$ of $V$ need not be regular. (For instance, $B|S \times S$ may be zero.)

**Proposition 1.3.** *Let $(V, B)$ be a regular quadratic space, and $S$ be a subspace of $V$. Then:*

(1) (Dimension Formula) $\dim S + \dim S^\perp = \dim V$.

(2) $(S^\perp)^\perp = S$.

**Proof.** Let $\varphi : V \to V^*$ be the linear isomorphism defined in (2) of the preceding proposition. Then $S^\perp$ is precisely the subspace of $V$ annihilated by the functionals in $\varphi(S)$. By the usual duality theory in linear algebra, we have

$$\begin{aligned} \dim S^\perp &= \dim V^* - \dim \varphi(S) \\ &= \dim V - \dim S, \end{aligned}$$

since $\varphi$ is an isomorphism. This establishes (1). Applying (1) twice, we get

$$\dim (S^\perp)^\perp = \dim V - (\dim V - \dim S) = \dim S.$$

Since $(S^\perp)^\perp \supseteq S$, (2) follows immediately. □

Note that neither conclusion would hold in 1.3 if the quadratic space $(V, B)$ was not assumed to be *regular*: the case of the zero form $(B \equiv 0)$ provides, for instance, an extreme counterexample.

## 2. Diagonalization of Quadratic Forms

Throughout this book, we shall write $\dot{F}$ to denote the multiplicative group, $F \setminus \{0\}$, of the field $F$.

**Definition 2.1.** Let $f$ be a quadratic form over $F$, and $d \in \dot{F}$. We shall say that $f$ *represents* $d$ if there exist $x_1, \ldots, x_n$ in $F$ such that $f(x_1, \ldots, x_n) = d$. Note that $(x_1, \ldots, x_n)$ is automatically a nonzero vector. We shall write $D_F(f) = D(f)$ to denote the set of values in $\dot{F}$ represented by $f$. This set clearly depends only on the equivalence class of $f$.

If $(V, B)$ is any quadratic space corresponding to the equivalence class of $f$, then

$$D(f) = \{d \in \dot{F} \mid \text{ there exists } v \in V \text{ such that } q_B(v) = d\}.$$

We shall also denote this by $D(V)$ if the pairing $B$ is clear from the context.

If $a, d \in \dot{F}$, then, clearly, $d \in D(f)$ iff $a^2 d \in D(f)$. Thus, $D(f)$ consists of a union of cosets of $\dot{F}$ modulo $\dot{F}^2$. We shall call $\dot{F}/\dot{F}^2$ the *group of square*

*classes* of $F$. By abuse of notation, we may also regard $D(f)$ as a subset of $\dot{F}/\dot{F}^2$. In $\dot{F}$, the subset $D(f)$ is always closed under inverses, since $d \in D(f)$ implies that $d^{-1} = (d^{-1})^2 d \in D(f)$.

In general, $D(f)$ is *not* a subgroup of $\dot{F}$. For one thing, $D(f)$ may not contain 1. Even if $D(f)$ contains 1, it may fail to be closed under multiplication. For instance, consider the form $f = X^2 + Y^2 + Z^2$ over the field of rational numbers $\mathbb{Q}$. Here, $D(f)$ consists of (nonzero) rational numbers which are sums of three squares of rational numbers. We have $1, 2, 2^{-1}, 14 \in D(f)$, but $2^{-1} \cdot 14 = 7$ is known to be *not* in $D(f)$ in elementary number theory.

If the value set $D(f)$ for a form $f$ happens to be closed under multiplication, then (ignoring the case of the 0-dimensional form) $1 \in D(f)$, and $D(f)$ is a subgroup of $\dot{F}$. In this case, we say $f$ is a *group form* over $F$. For instance, by the classical 2-square, 4-square and 8-square identities, the quadratic forms $\sum_{i=1}^{n} X_i^2$ for $n = 1, 2, 4, 8$ *are group forms over any field* $F$. This result will be substantially generalized in a future chapter. As an interesting case in point, we can cite Lagrange's Theorem which says that any positive integer is a sum of four squares of integers. This beautiful theorem implies readily that $D_{\mathbb{Q}}(W^2 + X^2 + Y^2 + Z^2)$ is the group of positive rational numbers $\dot{\mathbb{Q}}^+$.

Next we want to introduce *orthogonal sums*. If $(V_1, B_1)$, $(V_2, B_2)$ are quadratic spaces, we may form $(V, B)$, where $V = V_1 \oplus V_2$, and $B$ is the pairing $V \times V \to F$ given by

$$B((x_1, x_2), (y_1, y_2)) = B_1(x_1, y_1) + B_2(x_2, y_2).$$

$B$ is clearly symmetric and bilinear, so $(V, B)$ is a new quadratic space. We have $B(V_1, V_2) = 0$, and $B \mid (V_i \times V_i) = B_i$ $(i = 1, 2)$. We shall henceforth denote $(V, B)$ by $V_1 \perp V_2$; this is called the *orthogonal sum* of $(V_1, B_1)$ and $(V_2, B_2)$. Note that

$$\begin{aligned} q_B(x_1, x_2) &= B((x_1, x_2), (x_1, x_2)) \\ &= B_1(x_1, x_1) + B_2(x_2, x_2) \\ &= q_{B_1}(x_1) + q_{B_2}(x_2). \end{aligned} \tag{2.2}$$

This dictates, incidentally, how orthogonal sums should be defined in the category of quadratic forms. For example, if $q_1$ is the binary form $X^2 + 2Y^2$, and $q_2$ is the ternary form $5XY - Z^2$, then their orthogonal sum is to be taken as

$$q_1 \perp q_2 = U^2 + 2V^2 + 5XY - Z^2,$$

in the five variables $U$, $V$, $X$, $Y$, $Z$.

It is easy to see that the orthogonal sum of two quadratic spaces is regular iff each of them is regular. This fact (especially the "if" part) will be used freely without mention throughout.

For $d \in F$, we shall write $\langle d \rangle$ to denote the isometry class of the 1-dimensional space corresponding to the quadratic form $dX^2$. Clearly, $\langle d \rangle$ is regular iff $d \in \dot{F}$.

**Representation Criterion 2.3.** *Let $(V, B)$ be a quadratic space, and $d \in \dot{F}$. Then $d \in D(V)$ iff there exists another quadratic space $(V', B')$ together with an isometry $V \cong \langle d \rangle \perp V'$.*

**Proof.** If we have $V \cong \langle d \rangle \perp V'$, then $d \in D(\langle d \rangle \perp V') = D(V)$. Conversely, suppose $d \in D(V)$, so there exists $v \in V$ with $q(v) = d$ (where $q = q_B$). We first make a reduction to the case where $V$ is regular. Take any subspace $W$ such that $V = (\operatorname{rad} V) \oplus W = (\operatorname{rad} V) \perp W$. By (2.2), we have $D(V) = D(W)$, and $W$ is clearly regular. We may thus assume that $V$ itself is regular. The quadratic subspace $F \cdot v$ is isometric to $\langle d \rangle$, and $(F \cdot v) \cap (F \cdot v)^{\perp} = 0$. Since

$$\dim (F \cdot v) + \dim (F \cdot v)^{\perp} = \dim V$$

by Proposition 1.3, we conclude that $V \cong \langle d \rangle \perp (F \cdot v)^{\perp}$. □

The first consequence of the above criterion is the existence of "orthogonal bases" in any quadratic space.

**Corollary 2.4.** *If $(V, B)$ is any quadratic space over $F$, then there exist scalars $d_1, \ldots, d_n \in F$ such that $V \cong \langle d_1 \rangle \perp \cdots \perp \langle d_n \rangle$. (In other words, any $n$-ary quadratic form is equivalent to some diagonal form, $d_1 X_1^2 + \cdots + d_n X_n^2$.)*

**Proof.** If $D(V)$ is empty, then $B \equiv 0$ and $V$ is isometric to an orthogonal sum of $\langle 0 \rangle$'s. If there exists some $d \in D(V)$, then $V \cong \langle d \rangle \perp V'$ for some $(V', B')$, and the proof proceeds by induction on $\dim V$. □

**Notation.** In the rest of the book, we shall abbreviate the diagonal form $\langle d_1 \rangle \perp \cdots \perp \langle d_n \rangle$ by $\langle d_1, \ldots, d_n \rangle$. The special $n$-ary diagonal form $\langle d, \ldots, d \rangle$ will be abbreviated as $n \langle d \rangle$. For instance, $3 \langle a \rangle \perp 2 \langle b \rangle$ means the quinary form $\langle a, a, a, b, b \rangle$.

**Corollary 2.5.** *If $(V, B)$ is a quadratic space (not necessarily regular) and $S$ is a regular subspace, then:*

(1) $V = S \perp S^{\perp}$.

(2) *If $T$ is a subspace of $V$ such that $V = S \perp T$, then $T = S^{\perp}$.*

**Proof.** (1) $\Rightarrow$ (2). If $V = S \perp T$, then, clearly, $T \subseteq S^\perp$. But

$$\dim T = \dim V - \dim S = \dim S^\perp$$

by (1) (not by Proposition 1.3!), so we must have $T = S^\perp$.

(1) Since $S \cap S^\perp = \operatorname{rad} S = 0$, it suffices to show that $V$ is spanned by $S$ and $S^\perp$. By 2.4, $S$ has an orthogonal basis $x_1, \ldots, x_p$, and the regularity of $S$ implies that $B(x_i, x_i) \neq 0$ for all $i$. Given any $z \in V$, consider the vector

$$y = z - \sum_{i=1}^{p} \frac{B(z, x_i)}{B(x_i, x_i)}\, x_i.$$

We have

$$\begin{aligned} B(y, x_j) &= B(z, x_j) - \sum_{i=1}^{p} \frac{B(z, x_i)}{B(x_i, x_i)}\, B(x_i, x_j) \\ &= B(z, x_j) - \frac{B(z, x_j)}{B(x_j, x_j)}\, B(x_j, x_j) = 0. \end{aligned}$$

Thus, $y \in S^\perp$, and

$$z = y + \sum_{i=1}^{p} \frac{B(z, x_i)}{B(x_i, x_i)}\, x_i \in S \perp S^\perp. \qquad \square$$

**Corollary 2.6.** *Let $(V, B)$ be a regular quadratic space. Then a subspace $S$ is regular iff there exists $T \subseteq V$ such that $V = S \perp T$.*

**Proof.** For the "only if" part, take $T = S^\perp$, and apply (1) of the above corollary. Conversely, if $V = S \perp T$, then $\operatorname{rad} S \subseteq \operatorname{rad} V = 0$, so $S$ is regular (and $T = S^\perp$). $\square$

We discuss now the *determinant* of a nonsingular quadratic form $f$. This is defined to be $d(f) = \det(M_f) \cdot \dot{F}^2$ (an element of $\dot{F}/\dot{F}^2$), where $M_f$ is the symmetric matrix associated with $f$. Note that if $f \cong g$, then $M_f = C^t M_g C$ for some nonsingular $C$, and hence

$$d(f) = \det(M_f) \cdot \dot{F}^2 = \det(M_g) \cdot (\det C)^2 \cdot \dot{F}^2 = d(g).$$

This shows that $d(f)$ is an *invariant* of the equivalence class of $f$. Also, we have clearly

$$d(f_1 \perp f_2) = d(f_1) d(f_2) \in \dot{F}/\dot{F}^2.$$

Let $(V, B)$ be a (regular) quadratic space that corresponds to the equivalence class of $f$. If $V \cong \langle d_1, \ldots, d_n \rangle$ is a "diagonalization" of $V$, then $d(f) = d_1 \cdots d_n \cdot \dot{F}^2$. It is sometimes convenient to call this quantity the determinant of $V$, written $d(V)$.

## 3. Hyperbolic Plane and Hyperbolic Spaces

In this section, we shall introduce the important notion of a hyperbolic quadratic space. We begin by defining "isotropy" and "anisotropy".

**Definition 3.1.** Let $v$ be a nonzero vector in a quadratic space $(V, B)$. We say that $v$ is an *isotropic vector* if $B(v, v) = 0$ (or equivalently, if $q(v) = 0$, where $q = q_B$), and say that $v$ is *anisotropic* otherwise. The quadratic space $(V, B)$ is said to be *isotropic* if it contains a (nonzero) isotropic vector, and is said to be *anisotropic* otherwise. (Note that anisotropic spaces are necessarily regular.) Finally, we shall say that $(V, B)$ is *totally isotropic* if all nonzero vectors in $V$ are isotropic (that is, $B \equiv 0$).

Whether or not the zero vector should be taken as anisotropic leads mostly to fruitless debate. It will be essentially harmless to side-step this issue altogether, which is exactly what we are going to do. The important thing to keep in mind is just that when we try to say something interesting about isotropic or anisotropic vectors $v$, these vectors are understood to be nonzero. Finally, we should note that, according to Definition 3.1, the 0-dimensional quadratic form is technically *anisotropic.*

**Theorem 3.2.** *Let $(V, q)$ be a 2-dimensional quadratic space. The following four statements are equivalent:*

(1) *$V$ is regular and isotropic.*
(2) *$V$ is regular, with $d(V) = -1 \cdot \dot{F}^2$.*
(3) *$V$ is isometric to $\langle 1, -1 \rangle$.*
(4) *$V$ corresponds to the equivalence class of the binary quadratic form $X_1X_2$.*

**Proof.** We have already seen that (3) $\Leftrightarrow$ (4) in 1.1.

(1) $\Rightarrow$ (2): Let $x_1, x_2$ be an orthogonal basis for $V$. Regularity of $V$ implies that $q(x_i) = d_i \neq 0$ $(i = 1, 2)$. Let $ax_1 + bx_2$ be an isotropic vector, with (say) $a \neq 0$. Then

$$0 = q(ax_1 + bx_2) = a^2 d_1 + b^2 d_2 \implies d_1 = -(ba^{-1})^2 \cdot d_2$$
$$\implies d(V) = d_1 d_2 \cdot \dot{F}^2 = -1 \cdot \dot{F}^2.$$

(2) $\Rightarrow$ (3): Under the hypothesis (2), we have clearly a diagonalization, $V \cong \langle a, -a \rangle$ for some $a \in \dot{F}$. By the argument in 1.1, we know that $aX_1^2 - aX_2^2$ is equivalent to $aX_1X_2$. The latter clearly represents all elements in $\dot{F}$. In particular, $(V, q)$ itself represents 1. By the Representation Criterion, we conclude that $V \cong \langle 1, -1 \rangle$.

(3) $\Rightarrow$ (1) is trivial. $\square$

The isometry class of a 2-dimensional quadratic space satisfying the conditions in Theorem 3.2 is called the "hyperbolic plane" (presumably because the graphs of the equations $X_1X_2 = d$ are called hyperbolas). The hyperbolic plane will be denoted by $\mathbb{H}$, and will play a very special role throughout the subsequent developments. An orthogonal sum of hyperbolic planes will be called a *hyperbolic space*. The corresponding quadratic form may be taken either as $(X_1^2 - X_2^2) + \cdots + (X_{2m-1}^2 - X_{2m}^2)$ or as $X_1X_2 + \cdots + X_{2m-1}X_{2m}$.

**Definition 3.3.** A quadratic form (or quadratic space) is called *universal* if it represents all the nonzero elements of $F$. (Of course, any such form is a group form over $F$.)

**Theorem 3.4.** *Let $(V, B)$ be a regular quadratic space. Then:*

(1) *Every totally isotropic subspace $U \subseteq V$ of positive dimension $r$ is contained in a hyperbolic subspace $T \subseteq V$ of dimension $2r$.*

(2) *$V$ is isotropic iff $V$ contains a hyperbolic plane (necessarily as an orthogonal summand, by Corollary 2.5(1)).*

(3) *$V$ is isotropic $\Rightarrow$ $V$ is universal.*

**Proof.** We show first that (1) $\Rightarrow$ (2) $\Rightarrow$ (3), and then come back to prove (1). (1) $\Rightarrow$ (2) is clear by putting $r = 1$ in (1). (2) $\Rightarrow$ (3) is also clear because the form $X_1X_2$ corresponding to $\mathbb{H}$ is obviously universal. We shall now prove (1) by induction on $r$. Take any basis $x_1, \ldots, x_r$ in $U$, and let $S$ be the span of $x_2, \ldots, x_r$. Of course, $U^\perp \subseteq S^\perp$. Since $V$ is regular, we may apply 1.3 to get

$$\dim S^\perp = \dim V - \dim S > \dim V - \dim U = \dim U^\perp.$$

This means that there exists a vector $y_1$ that is orthogonal to $x_2, \ldots, x_r$, but not orthogonal to $x_1$. In particular, $x_1, y_1$ are linearly independent vectors (since $x_1$ is isotropic). The subspace $H_1 = Fx_1 + Fy_1$ has determinant

$$d(H_1) = \begin{vmatrix} 0 & B(x_1, y_1) \\ B(x_1, y_1) & B(y_1, y_1) \end{vmatrix} \cdot \dot{F}^2 = -1 \cdot F^2,$$

so $H_1 \cong \mathbb{H}$ by Theorem 3.2. We have thus a splitting $V = H_1 \perp V'$, where $V' = H_1^\perp$ contains $x_2, \ldots, x_r$ (Corollary 2.5). Since $V'$ is regular (Corollary 2.6), the proof proceeds by induction. □

**Remark.** (A) Since the hyperbolic plane $\mathbb{H}$ has a diagonalization $\langle 1, -1 \rangle$, the fact that $\mathbb{H}$ is universal amounts to the fact that any (nonzero) element in $F$ is a difference of two squares. This fact can be checked directly by using the following equation:

$$a = \left(\frac{a+1}{2}\right)^2 - \left(\frac{a-1}{2}\right)^2 \quad (\forall a \in F),$$

which will be used many times over in the sequel.[2]

(B) If one prefers to give a direct argument for (3), one can proceed as follows. Fix an isotropic vector $x_1$, and take $y_1 \in V$ such that $B(x_1, y_1) \neq 0$. Then $B(tx_1 + y_1, tx_1 + y_1) = 2tB(x_1, y_1) + B(y_1, y_1)$ clearly assumes all values as $t$ ranges over $F$.

**Corollary 3.5 (First Representation Theorem).** *Let $q$ be a regular quadratic form, and $d \in \dot{F}$. Then, $d \in D(q)$ iff $q \perp \langle -d \rangle$ is isotropic.*

**Proof.** We may assume that $q(X) = a_1X_1^2 + \cdots + a_nX_n^2$. If there exists an equation $d = a_1x_1^2 + \cdots + a_nx_n^2$ ($x_i \in F$), then

$$a_1x_1^2 + \cdots + a_nx_n^2 + (-d) \cdot 1^2 = 0,$$

so the form $q \perp \langle -d \rangle$ is isotropic. Conversely, let $(x_1, \ldots, x_{n+1})$ be an isotropic vector for $q \perp \langle -d \rangle$, so $a_1x_1^2 + \cdots + a_nx_n^2 - dx_{n+1}^2 = 0$. If $x_{n+1} \neq 0$, then

$$d = a_1\left(\frac{x_1}{x_{n+1}}\right)^2 + \cdots + a_n\left(\frac{x_n}{x_{n+1}}\right)^2 \in D(q).$$

If, on the contrary, $x_{n+1} = 0$, then $(x_1, \ldots, x_n) \neq 0$ is an isotropic vector for $q$ itself. By (3) of the theorem, $D(q) = \dot{F}$, so, of course, $d \in D(q)$. □

**Corollary 3.6.** *Let $q_1, q_2$ be regular forms of positive dimensions. Then $q := q_1 \perp q_2$ is isotropic iff $D(q_1) \cap -D(q_2) \neq \varnothing$.*

**Proof.** For the "if" part, take an element $a \in D(q_1) \cap -D(q_2)$. If $q_1(x) = a$ and $q_2(y) = -a$, then $(x, y) \neq 0$ is an isotropic vector for $q$.

For the "only if" part, we may assume that $q_1, q_2$ are anisotropic. (If, say $q_2$ is isotropic, then $D(q_1) \cap -D(q_2) = D(q_1) \neq \varnothing$ by 3.4(3).) Suppose $q(x, y) = 0$, where $(x, y) \neq 0$. Say $x \neq 0$. Then $a := q_1(x) \neq 0$, and we have $a \in D(q_1) \cap -D(q_2)$, as desired. □

**Corollary 3.7.** *For a positive integer $r$, the following two statements are equivalent (over a given field $F$):*

(1) *Any regular quadratic form of dimension $r$ over $F$ is universal.*

(2) *Any quadratic form of dimension $r + 1$ over $F$ is isotropic.*

The easy proof of this is left to the reader. The nice fact expressed in this corollary will form the basis for the investigation of the "$u$-invariant" of a field in Ch. XI.

---

[2] Of course, it is essential here that $\mathrm{char}(F) \neq 2$. If $\mathrm{char}(F) = 2$, a difference of two squares is simply a square, but we may very well have $F \neq F^2$.

## 4. Decomposition Theorem and Cancellation Theorem

In this section, we come to some of the most important classical theorems in quadratic form theory, which first appeared in Witt's seminal paper [Wi], ca. 1937. Note that 4.1 and 4.2 below are proved for *arbitrary* quadratic spaces $(V, q)$, without any regularity assumptions on $q$.

**Witt's Decomposition Theorem 4.1.** *Any quadratic space $(V, q)$ splits into an orthogonal sum*

$$(V_t, q_t) \perp (V_h, q_h) \perp (V_a, q_a),$$

*where $V_t$ is totally isotropic, $V_h$ is hyperbolic (or zero), and $V_a$ is anisotropic ("Witt decomposition"). Furthermore, the isometry types of $V_t, V_h, V_a$ are all uniquely determined.*

**Proof.** For existence, take any subspace $V_0$ such that

$$V = (\operatorname{rad} V) \oplus V_0 = (\operatorname{rad} V) \perp V_0.$$

Then $V_t = \operatorname{rad} V$ is totally isotropic, and $V_0$ is regular. If $V_0$ is isotropic, we may write $V_0 = H_1 \perp V_1$ (by 3.4(2)), where $H_1 \cong \mathbb{H}$. If $V_1$ is again isotropic, we may further write $V_1 = H_2 \perp V_2$, where $H_2 \cong \mathbb{H}$. After a finite number of steps, we achieve a decomposition

$$V_0 = (H_1 \perp \cdots \perp H_r) \perp V_a \quad (r \geq 0),$$

where $H_1 \perp \cdots \perp H_r = V_h$ is hyperbolic (or zero), and $V_a$ is anisotropic. This proves the existence part. The uniqueness part will be deduced from the Cancellation Theorem, which reads as follows.

**Witt's Cancellation Theorem 4.2.** *If $q, q_1, q_2$ are arbitrary quadratic forms, then $q \perp q_1 \cong q \perp q_2 \Longrightarrow q_1 \cong q_2$.*

To establish the uniqueness part in 4.1, suppose $V$ has another "Witt decomposition," $V = V_t' \perp V_h' \perp V_a'$. Since $V_t'$ is totally isotropic and $V_h' \perp V_a'$ is regular, we have

$$\operatorname{rad} V = (\operatorname{rad} V_t') \perp \operatorname{rad}(V_h' \perp V_a') = V_t',$$

so $V_t' = V_t$. By the Cancellation Theorem, we have now $V_h \perp V_a \cong V_h' \perp V_a'$. Write $V_h \cong m \cdot \mathbb{H}$ (orthogonal sum of $m$ copies of $\mathbb{H}$) and $V_h' \cong m' \cdot \mathbb{H}$. By cancelling $\mathbb{H}$ one at a time, we conclude that $m = m'$ since $V_a, V_a'$ are both anisotropic. After all $m$ hyperbolic planes have been cancelled, we arrive at $V_a \cong V_a'$, completing the proof of 4.1.

**Definition 4.3.** The integer $m$ $(= \frac{1}{2} \dim V_h)$ uniquely determined in the Witt decomposition above is called the *Witt index* of the quadratic space $(V, q)$. The isometry class of $V_a$ is called the *anisotropic part* of $(V, q)$.

**Corollary 4.4.** *If $(V,q)$ is regular, the Witt index $m$ of $V$ equals the dimension of any maximal totally isotropic subspace of $V$.*

**Proof.** Let $U$ be any maximal totally isotropic subspace in $V$, with $\dim U = r$. By Theorem 3.4, there exists a hyperbolic space $T \supseteq U$ with dimension $2r$. Since $T$ is regular, 2.5 implies that $V = T \perp T^{\perp}$. Here, $T^{\perp}$ must be anisotropic. (If $0 \neq x \in T^{\perp}$ is an isotropic vector, then $U + F \cdot x$ will be a totally isotropic subspace of $V$ properly containing $U$, a contradiction.) The uniqueness part of 4.1 implies that $T \cong V_h$, so

$$m = \tfrac{1}{2} \dim V_h = \tfrac{1}{2}(2r) = r = \dim U. \qquad \square$$

Our next goal is to establish Witt's Cancellation Theorem 4.2. To present the proof, we need the notion of a "hyperplane reflection." Let $(V,B,q)$ be any quadratic space. We shall write $\mathrm{O}_q(V) = \mathrm{O}(V)$ to denote the group of isometries of $(V,q)$. This so-called *orthogonal group* is the symmetry group which underlies the geometry of our quadratic space. The following important construction associates an element $\tau_y \in \mathrm{O}(V)$ to every *anisotropic* vector $y \in V$. As a self-map from $V$ to itself, $\tau_y$ is defined by

$$\tau_y(x) = x - \frac{2B(x,y)}{q(y)}\, y \quad \text{for every } x \in V. \tag{4.5}$$

(1) $\tau_y$ is evidently a linear endomorphism.

(2) $\tau_y$ is the identity map on $(F \cdot y)^{\perp}$. In fact, in the above formula, if $B(x,y) = 0$, then $\tau_y(x) = x$. Furthermore, if we apply $\tau_y$ to $y$ itself, we get

$$\tau_y(y) = y - \frac{2B(y,y)}{q(y)} \cdot y = y - 2y = -y.$$

In particular, $\tau_y$ is an involution ($\tau_y^2 = 1$): it leaves the hyperplane $(F \cdot y)^{\perp}$ pointwise fixed, and "reflects" the vector $y$ across $(F \cdot y)^{\perp}$ to $-y$.

(3) $\tau_y \in \mathrm{O}(V)$ by the following straightforward calculation:

$$\begin{aligned} B(\tau_y(x), \tau_y(x')) &= B\left(x - \frac{2B(x,y)}{q(y)}\, y,\ x' - \frac{2B(x',y)}{q(y)}\, y\right) \\ &= B(x,x') + \frac{4B(x,y)\,B(x',y)}{q(y)^2}\, B(y,y) \\ &\quad - \frac{4B(x,y)\,B(x',y)}{q(y)} \\ &= B(x,x') \quad (\text{since } B(y,y) = q(y)). \end{aligned}$$

Alternatively, $\tau_y \in \mathrm{O}(V)$ can also be deduced from the fact that $\tau_y \,|\, F \cdot y$ and $\tau_y \,|\, (F \cdot y)^{\perp}$ are both isometries.

(4) As a linear automorphism, $\tau_y$ has determinant $-1$.

**Remark 4.6.** The set of hyperplane reflections $\{\tau_y \mid q(y) \neq 0\}$ is closed under conjugation in the orthogonal group $\mathrm{O}(V)$. In fact, if $\sigma \in \mathrm{O}(V)$, then one has $\sigma\tau_y\sigma^{-1} = \tau_{\sigma y}$. The verification is straightforward:

$$\begin{aligned}(\sigma\tau_y\sigma^{-1})(x) &= \sigma[\tau_y(\sigma^{-1}x)] \\ &= \sigma\left[\sigma^{-1}x - \frac{2B(\sigma^{-1}x, y)}{q(y)}\, y\right] \\ &= x - \frac{2B(x, \sigma y)}{q(\sigma y)}\,\sigma y \\ &= \tau_{\sigma y}(x) \quad \text{for every } x \in V.\end{aligned}$$

It follows, in particular, that the subgroup of $\mathrm{O}(V)$ generated by all $\tau_y$ (where $q(y) \neq 0$) is normal in $\mathrm{O}(V)$. In Section 7, we shall show that this normal subgroup actually coincides with $\mathrm{O}(V)$.

**Proof of 4.2.** Suppose it is given that $q \perp q_1 \cong q \perp q_2$.

**Step 1.** *Cancellation holds if $q$ is totally isotropic and $q_1$ is regular.* In fact, let $M_1, M_2$ be the symmetric matrices associated with $q_1$ and $q_2$. The hypothesis implies that $\begin{pmatrix} 0 & 0 \\ 0 & M_1 \end{pmatrix}$ is congruent to $\begin{pmatrix} 0 & 0 \\ 0 & M_2 \end{pmatrix}$, so there exists an invertible matrix $E = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ such that

$$\begin{pmatrix} 0 & 0 \\ 0 & M_1 \end{pmatrix} = E^t \begin{pmatrix} 0 & 0 \\ 0 & M_2 \end{pmatrix} E = \begin{pmatrix} * & * \\ * & D^t M_2 D \end{pmatrix}.$$

In particular, $M_1 = D^t M_2 D$. Since $M_1$ is nonsingular, so is $D$, and hence $M_1, M_2$ are congruent. This gives $q_1 \cong q_2$.

**Step 2.** *Cancellation holds if $q$ is totally isotropic.* To see this, diagonalize $q_1, q_2$ and assume that $q_1$ has exactly $r$ zero coefficients in the diagonalization, while $q_2$ has $r$ zeros or more. Rewriting the hypothesis, we have

$$q \perp r\,\langle 0\rangle \perp q_1' \cong q \perp r\,\langle 0\rangle \perp q_2'.$$

Since $q_1'$ is regular, Step 1 implies that $q_1' \cong q_2'$. By tagging on $r$ terms of $\langle 0\rangle$, we conclude that $q_1 \cong q_2$.

**Step 3** (*General case*). Let $\langle a_1, \ldots, a_n\rangle$ be a diagonalization of $q$. Inducting on $n$, we are reduced to the case $n = 1$. The case $a_1 = 0$ has been handled in Step 2, so we may assume that $q = \langle a_1\rangle$, $a_1 \neq 0$. The hypothesis now reads: $\langle a_1\rangle \perp q_1 \cong \langle a_1\rangle \perp q_2$. The Cancellation Theorem in this case is clearly equivalent to the following result.

**Proposition 4.7.** *Let $(V, q)$ be a quadratic space, and let $x, y$ be vectors in $V$ such that $q(x) = q(y) \neq 0$. Then there exists an element $\tau \in \mathrm{O}(V)$ such that $\tau(x) = y$.*

**Proof.** Geometrically, if we consider the reflection with respect to the hyperplane $(F\cdot(x-y))^{\perp}$, then $x$ should be taken to $y$. But, is $x-y$ necessarily anisotropic? We derive first the "law of parallelogram":

$$\begin{aligned} q(x+y)+q(x-y) &= B(x+y,\, x+y)+B(x-y,\, x-y) \\ &= 2B(x,x)+2B(y,y) \\ &= 4q(x)\neq 0. \end{aligned}$$

This implies that $q(x+y)$, $q(x-y)$ cannot be both zero. Replacing $y$ by $-y$ if necessary, we may assume that $q(x-y)\neq 0$. [If we can find $\tau\in \mathrm{O}(V)$ such that $\tau(x)=-y$, the isometry $-\tau$ will take $x$ to $y$.] Applying the reflection $\tau_{x-y}$ to $x$, we obtain

$$\tau_{x-y}(x) = x - \frac{2B(x,\, x-y)}{q(x-y)}\cdot(x-y).$$

But

$$\begin{aligned} q(x-y) &= B(x-y,\, x-y) \\ &= B(x,x)+B(y,y)-2B(x,y) \\ &= 2\,[B(x,x)-B(x,y)] \\ &= 2B(x,\, x-y). \end{aligned}$$

Thus, $\tau_{x-y}(x) = x-(x-y) = y$, as we claimed at the beginning of the proof. □

## 5. Witt's Chain Equivalence Theorem

The theorem in the section title (also originating from Witt's classical paper [Wi]) describes the equivalence of two diagonal quadratic forms in terms of the equivalence of *binary* diagonal forms. First, let us prove the following easy fact for binary forms.

**Proposition 5.1.** *Let $q=\langle a,b\rangle$, $q'=\langle c,d\rangle$ be regular binary forms. Then $q\cong q'$ iff $d(q)=d(q')$, and $q$, $q'$ represent a common element $e\in \dot{F}$.*

**Proof.** "Only if" is clear. Conversely, assume that $d(q)=d(q')\in \dot{F}/\dot{F}^2$ and $e\in D(q)\cap D(q')$. By the Representation Criterion (2.3), we know that $q\cong\langle e,e'\rangle$ for some $e'\in\dot{F}$. Taking determinants, we have $ab\dot{F}^2=ee'\dot{F}^2$, so $q\cong\langle e,\, abe\rangle$. Similarly, $q'\cong\langle e,\, cde\rangle$. But $ab\dot{F}^2=cd\dot{F}^2$, so $q\cong q'$. □

We shall now introduce the notion of *simple equivalence* for diagonal forms. Let $q=\langle a_1,\dots,a_n\rangle$ and $q'=\langle b_1,\dots,b_n\rangle$. We shall say that $q$ and $q'$ are *simply-equivalent*, if there exist two indices, $i$ and $j$, such that

(1) $\langle a_i,\, a_j\rangle\cong\langle b_i,\, b_j\rangle$, and

(2) $a_k=b_k$ whenever $k$ is different from $i$ and $j$.