NOTES ON THE ALGEBRAIC THEORY OF QUADRATIC FORMS

KYLE ORMSBY

Caveat emptor — These notes are in draft form and will evolve throughout the semester. Please contact the author at ormsbyk@reed.edu with any comments or corrections.

I learned most of this material form T.Y. Lam's classic [Lam05]. Portions of these notes are adapted from that text and [MH73, Szy97].

CONTENTS

1. *Amuse-gueule*

1.1. **The Hasse principle.** Every good Pythagorean — or Babylonian cuneiform scribe, for that matter — confronts the equation

$$a^2 + b^2 = c^2$$

and wonders how to solve it with integers. Since the equation is homogeneous, it suffices to find rational solutions and then clear denominators. This leads to the idea of solving

$$x^2 + y^2 - z^2 = 0$$

with $(x, y, z) \in \mathbb{Q}^3$. More generally, given a homogeneous quadratic polynomial $f(x_1, \ldots, x_n)$ with coefficients in a field k, we can ask whether there are solutions to

$$f(x_1, \ldots, x_n) = 0$$

in k^n. We always have the trivial solution $x_1 = \cdots = x_n = 0$, so we should in fact search for *nontrivial* solutions $(x_1, \ldots, x_n) \in \mathsf{k}^n \smallsetminus \{(0, \ldots, 0)\}$.

Such a polynomial f is called a *quadratic form*, and when $f = 0$ has nontrivial solutions (over k), we call f *isotropic* (over k). Which integral quadratic forms f are isotropic over $\mathbb{Q}$? A necessary condition is that f be isotropic over $\mathbb{R}$, and we will see later that there are simple and effective tests of isotropicity over $\mathbb{R}$. In order to discover stronger conditions, suppose that f is isotropic over $\mathbb{Q}$ with nontrivial solution $(x_1, \ldots, x_n) \in \mathbb{Z}^n$. Dividing out by the greatest common divisor of the x_i's, we may assume that $\gcd(x_1, \ldots, x_n) = 1$ — a *primitive* integral solution. Reducing the equation $f = 0$ modulo any integer m, we get

$$f(x_1, \ldots, x_n) \equiv 0 \pmod{m}.$$

We say that $(x_1, \ldots, x_n)$ is a *primitive solution modulo* m when $\gcd(x_1, \ldots, x_n, m) = 1$. This is now (nearly) enough terminology to state the famous *strong Hasse principle*:

Theorem 1.1 (Strong Hasse principle — first version)**.** *Let f be a regular integral quadratic form. Then f is isotropic over $\mathbb{Q}$ if and only if it is isotropic over $\mathbb{R}$ and has a primitive solution modulo m for all positive integers m.*

The ring $\mathbb{Z}/m\mathbb{Z}$ is unpleasant to work with when m is composite because of the presence of zero divisors. Theorem 1.1 is better approached via the p-adic numbers $\mathbb{Q}_p$, which we will develop in due course. It is common to write $\mathbb{R} = \mathbb{Q}_\infty$ and call $\{\text{positive prime integers}\} \cup \{\infty\}$ the set of *places* of $\mathbb{Q}$. This leads to the following reformulation of the strong Hasse principle:

Theorem 1.2 (Strong Hasse principle — second version)**.** *Let f be a regular*[1] *quadratic form over $\mathbb{Q}$. Then f is isotropic over $\mathbb{Q}$ if and only if it is isotropic over $\mathbb{Q}_p$ for all places p.*

1.2. **Fantastic forms and where to find them.** The algebraic theory of quadratic forms has deep roots in number theory, but quadratic — or bilinear — algebra is pervasive in modern mathematics.

Real quadratic forms and the second derivative test. Fix an open set $U \subseteq \mathbb{R}^n$ and suppose $f\colon U \to \mathbb{R}$ has continuous second order partial derivatives. The *Hessian* $H(x)$ of f at a point $x \in U$ is the $n \times n$ real matrix of second order partial derivatives of f evaluated at x. If $x_0 \in U$ is a critical point of f, then $f(x_0) + (x - x_0)^T H(x_0)(x - x_0)$ is a good approximation to f in a neighborhood of x.

Since $H(x_0)$ is a symmetric matrix, $q(x) = x^T H(x_0) x$ is a (real) quadratic form. If q is positive definite (so $q(x) \geq 0$ with equality only for $x = 0$), then f attains a local minimum at x_0; if q is negative definite, then f attains a local maximum at x_0; if q is regular and indefinite, then f has a

[1] We will define regularity properly quite soon, but not yet. Loosely speaking, a quadratic form is regular when you cannot eliminate variables via a change of basis.

saddle point at x_0; and finally if q is not regular then the second derivative test is inconclusive at x_0.

Later, we will see that regular real quadratic forms are determined (up to *isometry*) by dimension and *signature*. This is the content of *Sylvester's law of inertia*.

Pfister forms and the level of a field. Suppose that -1 is a sum of squares in a field k. The *level of* k is the smallest integer $s(\mathsf{k})$ such that -1 is a sum of s squares in k. Remarkably, $s(\mathsf{k})$ is always a power of 2. The proof of this result depends on Pfister's theory of multiplicative quadratic forms.

Intersection forms on even-dimensional manifolds. A closed oriented $4k$-manifold M has an *intersection (symmetric bilinear) form*

$$H_{2k}(M;\mathsf{k}) \times H_{2k}(M;\mathsf{k}) \to H_{2k}(M;\mathsf{k})$$

where $H_{2k}(M;\mathsf{k})$ is the "middle homology group" of M with coefficients in a field k. Shortly, we will see that symmetric bilinear forms are in bijective correspondence with quadratic forms when the characteristic of k is not 2. The intersection form is a powerful tool for studying the topology of closed manifolds.

Enumerative geometry. Loosely speaking, a (smooth) algebraic variety over a field k is a subset of k^n cut out by some number of polynomial equations (without singularities). Classical enumerative geometry counts the number of certain types of structures (*e.g.*, lines) on particular classes of complex smooth algebraic varieties (*e.g.*, cubic surfaces in $\mathbb{P}^3_{\mathbb{C}}$). Indeed, a famous result says that there are exactly 27 lines on every smooth cubic surface in $\mathbb{P}^3_{\mathbb{C}}$. The proof of this result proceeds via computation of a certain "self-intersection" number as an Euler characteristic of an associated vector bundle.

Recent work of Kass–Wickelgren and Levine shows how similar results can be obtained over any field when we "count" with quadratic forms instead of integers. Taking the dimension of the quadratic forms in question recovers the classical results over $\mathbb{C}$.

2. Some linear algebra

In the next two lectures, we will review some important concepts from linear algebra necessary for our study of quadratic forms. Henceforth, k will always denote a field. (At some point we will also require that the characteristic of k is not 2, but we do not need this assumption yet.)

Recall that a k-*vector space* V is an Abelian group $(V,+)$ along with a scalar multiplication $\mathsf{k} \times V \to V$ satisfying the usual distributivity and associativity axioms. For a *vector* $v \in V$ and *scalar* $\lambda \in \mathsf{k}$, the scalar product of v by λ is denoted $\lambda \cdot v$ or λv.

2.1. Linear transformations and Hom spaces.

Given k-vector spaces V and W, a *linear transformation* (or k-*linear map*) from V to W is a function $f\colon V \to W$ such that $f(\lambda v + w) = \lambda f(v) + f(w)$ for all $\lambda \in \mathsf{k}$ and $v, w \in V$.

Definition 2.1. The set of k-linear transformations from V to W is denoted $\mathrm{Hom}_{\mathsf{k}}(V,W)$ (or $\mathrm{Hom}(V,W)$ if the field is clear from context). When equipped with pointwise addition and scalar multiplication, $\mathrm{Hom}_{\mathsf{k}}(V,W)$ becomes a vector space called a *Hom space*.

By pointwise addition and scalar multiplication, we mean that for $f, g \in \mathrm{Hom}_{\mathsf{k}}(V,W)$, $v \in V$, and $\lambda \in \mathsf{k}$,

$$(f+g)(v) = f(v) + g(v),$$
$$(\lambda \cdot f)(v) = \lambda \cdot (f(v)).$$

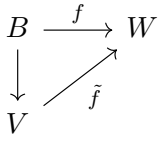
2.2. **Bases and matrices.** A *basis* of a k-vector space V is a subset $B \subseteq V$ of vectors which are linearly independent and span V. Every vector space has a basis, the cardinality of every basis is the same, and this cardinality is called the *dimension* of V. Linear transformations are determined by values on a basis in the following sense.

Proposition 2.2. Let V be a k-vector space with basis B and let W be another k-vector space. Then there is a bijection

$$W^B \longrightarrow \mathrm{Hom}_{\mathsf{k}}(V, W)$$

where W^B is the set of functions from B to W.

It is easy to see that restricting a linear transformation $f\colon V \to W$ to B gives a function in the opposite direction. The special part of the proposition is that any function $B \to W$ extends uniquely to a linear transformation $V \to W$. This is sometimes expressed diagramatically by saying that for every function $f\colon B \to W$ there is a unique linear tranformation $\tilde{f}\colon V \to W$ such that the diagram



commutes (where $B \to V$ is the inclusion function).[2]

If $B = \{v_i \mid i \in I\} \subseteq V$ is a basis and $v \in V$, then there is a unique way to express v as a linear combination of elements of B, $v = \sum_{i\in I} \lambda_i v_i$. Suppose now that $f\colon V \to W$ is a linear transformation between finite-dimensional vector spaces with ordered bases $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$, respectively. The *matrix* of f with respect to these ordered bases is the $m \times n$ array of scalars

$$A = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1} & \lambda_{m2} & \cdots & \lambda_{mn} \end{pmatrix}$$

where $f(v_i) = \sum_{j=1}^{m} \lambda_{ji} w_j$. In other words, the i-th column of A consists of the *coordinates* of $f(v_i)$ with respect to the ordered basis $\{w_1, \ldots, w_m\}$. Matrix multiplication is defined so that composition of linear transformations corresponds to multiplication of the corresponding matrices (after choosing ordered bases). We denote the set of $m \times n$ matrices over k by $M_{m\times n}(\mathsf{k})$ and give it the usual vector space structure.

2.3. **Direct sum.** We may combine two k-vector spaces via the operation of direct sum. More generally, we can form the direct sum of an infinite collection of k-vector spaces.

Definition 2.3. Given k-vector spaces V and W, their *direct sum*, $V \oplus W$, is the Cartesian product $V \times W$ equipped with termwise addition and scalar multiplication: $(v, w)+(v', w') = (v+v', w+w')$ and $\lambda(v, w) = (\lambda v, \lambda w)$.

If $\{V_i \mid i \in I\}$ is a set of k-vector spaces, then $\bigoplus_{i\in I} V_i$ is the subset of $\prod_{i\in I} V_i$ in which all but finitely many coordinates are 0; it is also equipped with termwise addition and scalar multiplication.

The set $\prod_{i\in I} V_i$ is also a vector space, but it has different formal properties when I is infinite. Direct sum has the following universal property.

[2]This discussion puts us ε away from talking about a vector space as being free on its basis in categorical language, but to do so properly we would also have to introduce the notion of naturality.

Proposition 2.4. If $\{V_i \mid i \in I\}$ is a set of k-vector spaces and W is another k-vector space, then there is an isomorphism

$$\mathrm{Hom}_{\mathsf{k}}\left(\bigoplus_{i\in I} V_i, W\right) \longrightarrow \prod_{i\in I} \mathrm{Hom}_{\mathsf{k}}(V_i, W).$$

You should be able to easily write down the appropriate map, and you will verify this proposition for 2-fold direct sums in your homework.

If I is a set and V is a vector space, we will write $V^{\oplus I}$ for the direct sum $\bigoplus_{i\in I} V$. If $I = \{1, \ldots, n\}$, we will write $V^{\oplus n}$ for $V^{\oplus I}$. Finally, if $V = \mathsf{k}$, then we may write k^n for $\mathsf{k}^{\oplus n}$, depending on mood. The *standard (ordered) basis* for k^n is $\{e_1, \ldots, e_n\}$ where e_i is the n-tuple with 1 in the i-th position and 0's elsewhere. If not explicitly mentioned, we will use the standard ordered bases for k^n and k^m when representing a linear transformation $\mathsf{k}^n \to \mathsf{k}^m$ as a matrix.

2.4. **Duals.** Dual vector spaces play an important role in linear algebra and an outsized one in the theory of symmetric bilinear and quadratic forms.

Definition 2.5. The k-linear *dual* of a k-vector space V is the Hom space

$$V^* = \mathrm{Hom}_{\mathsf{k}}(V, \mathsf{k}).$$

Elements of V^* are called *linear functionals* or *dual vectors*.

There is a canonical map

$$\begin{aligned} V &\longrightarrow (V^*)^* \\ v &\longmapsto (f \mapsto f(v)) \end{aligned}$$

which is always injective and is an isomorphism when V is finite-dimensional. (We call the map *canonical* because it does not depend on the choice of a basis or coordinates.) For V finite-dimensional, it is also the case that $V \cong V^*$, but this isomorphism in non-canonical. Indeed, after choosing an ordered basis $\{v_1, \ldots, v_n\}$ of V, we create a *dual basis* $\{v_1^*, \ldots, v_n^*\}$ where $v_i^*(v_j)$ is either 1 or 0 depending on whether $j = i$ or $j \neq i$. (This is a typical use of Proposition 2.2 — make sure you understand how it is being invoked.) It is straightforward to prove that $\{v_1^*, \ldots, v_n^*\}$ is a basis of V^* and the linear map taking v_i to v_i^* is an isomorphism.

Given a linear transformation $f\colon V \to W$, we can form the *dual transformation* $f^* : W^* \to V^*$ which takes $g : W \to \mathsf{k}$ to the composite linear functional $g \circ f$. In your homework, you will prove that this defines an injective linear transformation

$$\mathrm{Hom}_{\mathsf{k}}(V, W) \longrightarrow \mathrm{Hom}_{\mathsf{k}}(W^*, V^*)$$

which is an isomorphism when both vector spaces are finite-dimensional.

3. The quadratic square: quadratic forms, symmetric matrices, quadratic spaces, and symmetric bilinear forms

Henceforth, we adopt the convention that k is a field of characteristic different from 2, and that any other "arbitrary" field does not have characteristic 2. In many cases, this assumption will not be necessary, but it is crucial for our impending equivalence between symmetric bilinear and quadratic forms.

Our goal currently is to define four structures, all different but congruent vertices of the same square. We begin with the objects discussed in our *amuse-guele*, quadratic forms.

Definition 3.1. An *n-ary quadratic form over* k is a polynomial $f \in \mathsf{k}[x_1, \ldots, x_n]$ that is homogeneous of degree 2,

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

where not all of the $\lambda_{ij} \in \mathsf{k}$ are 0.

A presentation of the above form has some redundancy: $x^2+xy+2yx+y^2 = x^2+4xy-yx+y^2$, *&c.* We rectify this by symmetrizing our coefficients. Let $\lambda'_{ij} = \frac{1}{2}(\lambda_{ij} + \lambda_{ji})$. (This uses our $\operatorname{char} \mathsf{k} \neq 2$ assumption!) Then

$$f(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} \lambda'_{ij} x_i x_j$$

and $\lambda'_{ij} = \lambda'_{ji}$. The example form above has symmetric presentation $x^2 + \frac{3}{2}xy + \frac{3}{2}yx + y^2$.

In this fashion, every n-ary quadratic form over k determines a *unique* symmetric $n \times n$ matrix $A_f = (\lambda'_{ij})_{i,j=1}^n \in \operatorname{Sym}_{n\times n}(\mathsf{k}) \subseteq M_{n\times n}(\mathsf{k})$. Additionally, an n-ary quadratic form may be recovered from a nonzero symmetric matrix $A \in \operatorname{Sym}_{n\times n}(\mathsf{k})$. Indeed, let x denote the column vector with entries $x_1, x_2, \ldots, x_n$ and define

$$f_A(x) = x^\top A x.$$

It is easy to check that f_A is an n-ary quadratic form, and the assignments $f \mapsto A_f$ and $A \mapsto f_A$ are mutually inverse bijections between n-ary quadratic forms over k and $\operatorname{Sym}_{n\times n}(\mathsf{k}) \smallsetminus \{0\}$. These are the first two faces of our tetrahedron.

An n-ary quadratic form f also defines a *quadratic map* $q_f \colon \mathsf{k}^n \to \mathsf{k}$ given by evaluation of f. This function is quadratic in the sense that $q_f(\lambda x) = \lambda^2 q_f(x)$ for all $\lambda \in \mathsf{k}$ and $x \in \mathsf{k}^n$. The following definition abstracts this behavior from the vector space k^n to an arbitrary finite-dimensional k-vector space.

Definition 3.2. A *quadratic space over* k is a finite-dimensional k-vector space V equipped with a function $q \colon V \to \mathsf{k}$ satisfying $q(\lambda x) = \lambda^2 q(x)$ for all $\lambda \in \mathsf{k}$, $x \in V$, and such that the *polarization* of q,

$$\begin{aligned} B_q : V \times V &\longrightarrow \mathsf{k} \\ (x, y) &\longmapsto \frac{1}{2}(q(x+y) - q(x) - q(y)), \end{aligned}$$

is a symmetric bilinear form.

Definition 3.3. A *symmetric bilinear form* on a k-vector space V is a function $B \colon V \times V \to \mathsf{k}$ which is

(1) *symmetric*: $B(x, y) = B(y, x)$ for all $x, y \in V$, and
(2) *bilinear*: linear in each variable.

The *Gram matrix* of a symmetric bilinear form B relative to an ordered basis $v_1, \ldots, v_n$ of V is

$$A_B = (B(v_i, v_j))_{i,j=1}^n \in \operatorname{Sym}_{n\times n}(\mathsf{k}).$$

We may *depolarize* a symmetric bilinear form $B : V \times V \to \mathsf{k}$ to create a quadratic map

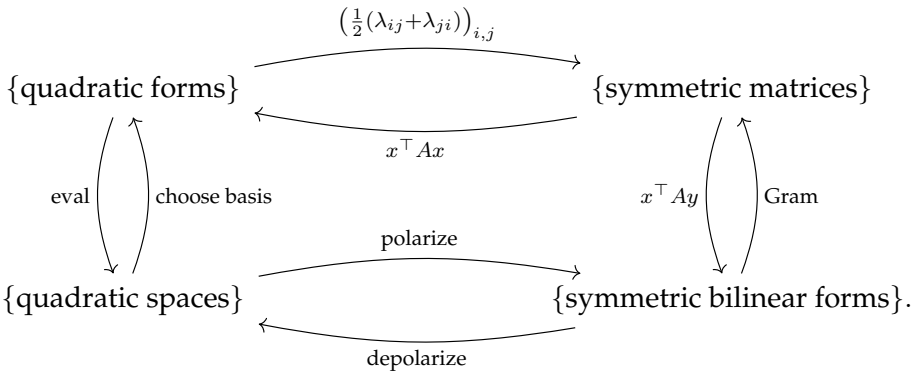
$$\begin{aligned} q_B : V &\longrightarrow \mathsf{k} \\ x &\longmapsto B(x, x). \end{aligned}$$

It is straightforward to check that polarization and depolarization are mutually inverse.

We may also produce a symmetric bilinear form from a symmetric matrix. Indeed, given $A \in Sym_{n\times n}(\mathsf{k})$, the map $\mathsf{k}^n \times \mathsf{k}^n \to \mathsf{k}$ taking $(x, y) \mapsto x^\top A y$ is symmetric bilinear.

We can summarize the above definitions and relations with the following diagram:

(□)



While the composable horizontal arrows are inverse bijections, the vertical arrows involve coordinatization and are only inverses if we restrict to vector spaces of the form k^n with standard ordered basis. We will see, though, that when we consider each structure up to an appropriate notion of isomorphism, all arrows become bijections.

Example 3.4. It will be instructive to chase a particular quadratic form, say $x^2-4xy+3y^2$, through these transformations. Expressed symmetrically, this is the same as $x^2 - 2xy - 2yx + 3y^2$ with symmetric matrix $\begin{pmatrix} 1 & -2 \\ -2 & 3 \end{pmatrix}$. We can then check that

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & -2 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^2 - 4xy + 3y^2,$$

as expected.

The associated quadratic space is (k^2, q) where q is the function $\mathsf{k}^2 \to \mathsf{k}$ given by $xe_1 + ye_2 \mapsto x^2 - 4xy + 3y^2$. Polarizing this quadratic map, we get

$$B((x,y),(x',y')) = \frac{1}{2}(q(x+x', y+y') - q(x,y) - q(x',y')) = xx' - 2xy' - 2x'y + 3yy'.$$

Evaluating at pairs of standard basis vectors, we get a Gram matrix $\begin{pmatrix} 1 & -2 \\ -2 & 3 \end{pmatrix}$ like before. Finally, depolarizing B gives

$$B((x,y),(x,y)) = x^2 - 2xy - 2xy + 3y^2 = x^2 - 4xy + 3y^2,$$

as anticipated.

4. Equivalence, congruence, and isometry

We now undertake the task of deciding when to consider two quadratic forms (or quadratic spaces, or symmetric matrices, or symmetric bilinear forms) to be the same. Let f and g be n-ary quadratic forms over k. We say that f is *equivalent* to g if there is an invertible matrix $B \in \mathrm{GL}_n(\mathsf{k})$ such that

$$f(x) = g(Bx).$$

This immediately translates into a condition on associated symmetric matrices. Indeed, we learn that

$$x^\top A_f x = (Bx)^\top A_g (Bx) = x^\top (B^\top A_g B) x,$$

which implies the matrix equation

$$A_f = B^\top A_g B.$$

We call two symmetric matrices $A, A' \in \mathrm{Sym}_{n\times n}(\mathsf{k})$ *congruent* if there exists $B \in \mathrm{GL}_n(\mathsf{k})$ such that $A = B^\top A' B$. We see then that equivalence of forms corresponds to congruence of symmetric matrices.

Example 4.1. Consider the change of coordinates effected by the matrix $B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. Applied to the form $g(x, y) = xy$ we get

$$g\left(B\begin{pmatrix} x \\ y \end{pmatrix}\right) = g(x+y, x-y) = (x+y)(x-y) = x^2 - y^2.$$

Thus g is equivalent to $h(x, y) = x^2 - y^2$. The corresponding matrix congruence is

$$A_h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = B^\top A_g B.$$

Suppose that (V, q) and (V', q') are quadratic spaces. We call them *equivalent* if there is a linear isomorphism $f\colon V \to V'$ such that $q = q' \circ f$. This clearly corresponds to equivalence of quadratic forms in the quadratic square (□) of the previous section.

If (V, B) and (V', B') are symmetric bilinear forms, we say they are *isometric* if there is a linear isomorphism $f\colon V \to V'$ such that

$$B'(f(x), f(y)) = B(x, y)$$

for all $x, y \in V$. Given an ordered basis $\{v_1, \ldots, v_n\}$ of V, we know that $\{f(v_1), \ldots, f(v_n)\}$ is an ordered basis of V' since f is an isomorphism. It follows that the Gram matrices of (V, B) and (V', B') agree when we choose f-compatible bases of V and V'.

We can now see that the four corners of the quadratic square (□) are the same when viewed up to equivalence (of quadratic forms or spaces), congruence (of symmetric matrices), or isometry (of symmetric bilinear forms). We will freely move between these perspectives as convenience dictates. The coordinate-free worlds of quadratic spaces and symmetric bilinear forms are often convenient in theoretical arguments, with symmetric bilinear forms of prime importance when we adopt a geometric perspective. (After all, a symmetric bilinear form is nearly an inner product space.) Quadratic forms and symmetric matrices conveniently package their data in a manner easily appreciated by both the human visual and computer algebra systems, and cannot be neglected.

5. Regular forms

In this section, we will use the geometric perspective granted by symmetric bilinear forms to understand the regularity condition advertised in the *amuse-guele*.

Theorem 5.1. *Let (V, B) be a symmetric bilinear form with Gram matrix A and quadratic form f (relative to some ordered basis). Then the following statements are equivalent:*

(a) $A \in \mathrm{GL}_n(\mathsf{k})$ *(*i.e.*, $\det A \neq 0$),*
(b) $x \mapsto B(\,, x)$ *defines an isomorphism $V \to V^*$,*
(c) for $x \in V$, if $B(x, y) = 0$ for all $y \in V$, then $x = 0$.

Definition 5.2. If any (and hence all) of the equivalent conditions in Theorem 5.1 hold, we call (V, B), A, f, and the associated quadratic space (V, q) *regular* or *nonsingular*. We also call the 0 vector space with trivial symmetric bilinear form, symmetric matrix, quadratic form, or quadratic space *regular* or *nonsingular*. Quadratic objects which are not regular are called *singular* or *degenerate*.

Proof of Theorem 5.1. We show that (a) $\iff$ (b) and (b) $\iff$ (c). First suppose that $A \in \mathrm{GL}_n(\mathsf{k})$ is the Gram matrix for B relative to an ordered basis $\alpha = \{v_1, \ldots, v_n\}$ of V. We may determine the matrix for $f\colon x \mapsto B(\ , x)$ by computing

$$f(v_j) = \sum_{k=1}^{n} A_{kj} v_k^*.$$

Indeed, $\left(\sum_{k=1}^n A_{kj} v_k^*\right)(v_i) = A_{ij} = B(v_i, v_j) = (f(v_j))(v_i)$. It follows that the matrix for f with respect to α and α^* is equal to A. Since $A \in \mathrm{GL}_n(\mathsf{k})$, we know that f has an inverse (the linear transformation induced by A^{-1} relative to bases α^* and α) and is an isomorphism.

Now assume that $f\colon x \mapsto B(\ , x)$ is an isomorphism. As above, the matrix for f relative to α and α^* is equal to A. Since f is invertible, so is A.

Continue to suppose that $f\colon x \mapsto B(\ , x)$ is an isomorphism, now with the goal of deducing (c). If for some $x \in V$, $B(x, y) = 0$ for all $y \in V$, then $B(y, x) = 0$ for all $y \in V$ as well by symmetry of B. It follows that $f(x) = 0$, the trivial linear functional. Since f is an isomorphism, we know that $x = 0$, which proves (c).

Finally, suppose that (c) is true. Invoking the symmetry of B, we immediately see that $f\colon x \mapsto B(\ , x)$ is injective. Since V and V^* have the same dimension, we conclude that f is an isomorphism. □

Example 5.3. The forms $x^2 + y^2$ and xy are regular. Indeed, they have matrices $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}$ with determinants 1 and $-1/4$, respectively.

Example 5.4. The form $x^2 + 2xy + y^2$ is singular. It has matrix $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ with determinant 0. Note that $x^2 + 2xy + y^2 = (x + y)^2$, and thus can be expressed as a regular quadratic form in the single variable $x + y$. We will investigate this idea further when we study totally isotropic spaces and Witt's decomposition theorem.

For the remainder of this section, we investigate some geometric constructions that are especially nice when working with regular symmetric bilinear forms.

Suppose that (V, B) is a symmetric bilinear form and that $U \leq V$ is a subspace of V. Restricting B to $U \times U$ results in another symmetric bilinear form $(U, B|_{U\times U})$.

Definition 5.5. In the above setting, the *orthogonal complement* of U is

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

Note that (V, B) is regular if and only if $V^\perp = 0$. The orthogonal complement of V is sometimes called the *radical* of (V, B).

Proposition 5.6. Suppose that (V, B) is a regular quadratic space and that U is a subspace of V. Then

(a) $\dim U + \dim U^\perp = \dim V$, and
(b) $(U^\perp)^\perp = U$.

Proof. Since (V, B) is regular, we know that $f\colon V \to V^*,\ x \mapsto B(\ , x)$ is an isomorphism. The reader may check that the dual inclusion $i\colon U \hookrightarrow V$ is a surjective map $i^*\colon V^* \to U^*$. Let $g = i^* \circ f\colon V \to U^*$, which is surjective since f is an isomorphism and i^* is surjective. By the rank-nullity theorem,

$$\dim V = \dim \ker g + \dim U^* = \dim \ker g + \dim U.$$

By definition,

$$\ker g = \{v \in V \mid f(v)(u) = 0 \text{ for all } u \in U\}.$$

We also have $f(v)(u) = B(u, v)$, whence $\ker g = U^\perp$. Substituting into our previous dimension equation gives

$$\dim V = \dim U^\perp + \dim U,$$

so we have proved (a).

For (b), first note that we always have $U \subseteq (U^\perp)^\perp$. An application of (a) to the subspace $U^\perp \leq V$ gives

$$\dim U^\perp + \dim(U^\perp)^\perp = \dim V,$$

so $\dim(U^\perp)^\perp = \dim V - \dim U^\perp$. Again by (a) (now applied to $U \leq V$), we see that the right-hand side also equals $\dim U$. □

Remark 5.7. Given that the dimensions of U and $U^\perp$ add to give $\dim V$, it might be tempting to guess that $V = U \oplus U^\perp$. This is false in general. Consider the bilinear form given by $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and let U denote the span of $(1, 1)$. In this case, $U^\perp = U$.

Remark 5.8. The regularity hypothesis in Proposition 5.6 is essential. If B is the 0 map (and $V \neq 0$), then $V^\perp = V$ and $\dim V + \dim V^\perp = 2 \dim V$.

6. Diagonalization of forms

First, fix the following notation and terminology: $\mathsf{k}^\times = \mathsf{k} \smallsetminus \{0\}$ is the multiplicative group of units in k and $\mathsf{k}^{\boxtimes} = \{\lambda^2 \mid \lambda \in \mathsf{k}^\times\}$ is the subgroup of squares in $\mathsf{k}^\times$. We call $\mathsf{k}^\times/\mathsf{k}^{\boxtimes}$ the *group of square classes* of k.

Definition 6.1. Let f be a quadratic form over k and take $\lambda \in \mathsf{k}^\times$. We say that f *represents* $\lambda \in \mathsf{k}^\times$ if there exist $x_1, \ldots, x_n \in \mathsf{k}$ such that $f(x_1, \ldots, x_n) = \lambda$. We write $D_{\mathsf{k}}(f) = D(f)$ for the set of elements of $\mathsf{k}^\times$ represented by f; in other words,

$$D_{\mathsf{k}}(f) = \{f(x) \mid x \in \mathsf{k}^n\} \smallsetminus \{0\}.$$

Proposition 6.2. The following three statements about $D(f)$ are true.

(a) The set $D(f)$ only depends on the equivalence class of f.
(b) The set $D(f)$ is a union of cosets in $\mathsf{k}^\times/\mathsf{k}^{\boxtimes}$.
(c) The set $D(f)$ is closed under taking inverses.

Proof. To prove (a), suppose that g is equivalent to f, with $f(x) = g(Bx)$ for some $B \in \mathrm{GL}_n(\mathsf{k})$. Since B induces a bijection from $\mathsf{k}^n \smallsetminus \{0\}$ to itself, we see that $D(f) = D(g)$.

To prove (b), we need to show that for all $a \in \mathsf{k}^\times$, $a^2\lambda \in D(f)$ whenever $\lambda \in D(f)$. If $\lambda \in D(f)$, then $\lambda = f(x)$ for some $x \in \mathsf{k}^n \smallsetminus \{0\}$, and hence $f(ax) = a^2 f(x) = a^2\lambda$.

Part (c) follows easily from (b) since $\lambda^{-1} = (\lambda^{-1})^2\lambda$. □

Remark 6.3. The set $D(f)$ is *not* a subgroup of $\mathsf{k}^\times$ in general. For instance, the form $-x^2$ does not represent 1 over $\mathbb{Q}$ or $\mathbb{R}$. The set $D(f)$ also need not be closed under multiplication, as can be seen by considering the form $f = x^2 + y^2 + z^2$ over $\mathbb{Q}$. Then $1 = 1^2 + 0^2 + 0^2$, $2 = 1^2 + 1^2 + 0^2$, and $14 = 3^2 + 2^2 + 1^2$ are in $D(f)$. Hence $2^{-1} \in D(f)$ as well. If $D(f)$ were closed under multiplication, the $7 = 2^{-1} \cdot 14$ would be in $D(f)$. A consequence of the strong Hasse principle Theorem 1.2 is that

$$D(f) = \{\lambda \in \mathbb{Q}^\times \mid \lambda > 0 \text{ and } -a \text{ is a square in } \mathbb{Q}_2\}.$$

After introducing the p-adic rationals and proving Theorem 1.2, we will show that $-7 \notin \mathbb{Q}_2^{\boxtimes}$, so $D(f)$ is not closed under multiplication.

Remark 6.4. When $D(f)$ is closed under multiplication, we automatically get that $1 \in D(f)$ and we call f a *group form* over k. For instance, the formula

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 - x_2y_2)^2 + (x_1y_2 + x_2y_1)^2$$

due to Brahmagupta (7th centure C.E.) and Fibonacci (1202) implies that $x_1^2 + x_2^2$ is a group form. In 1748, Euler discovered the identity

$$\begin{aligned}(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) = \\ (x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4)^2 + (x_1y_2 + x_2y_1 + x_3y_4 - x_4y_3)^2 + \\ (x_1y_3 - x_2y_4 + x_3y_1 + x_4y_2)^2 + (x_1y_4 + x_2y_3 - x_3y_2 + x_4y_1)^2\end{aligned}$$

so $x_1^2 + x_2^2 + x_3^2 + x_4^2$ is also a group form. Using the octonions (an 8-dimensional, non-associative number system), Graves (1843) and Cayley (1845) found the eight-square identity

$$\begin{aligned}(x_1^2 + \cdots + x_8^2)(y_1^2 + \cdots + y_8^2) = \\ (x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4 - x_5y_5 - x_6y_6 - x_7y_7 - x_8y_8)^2 + \\ (x_1y_2 + x_2y_1 + x_3y_4 - x_4y_3 + x_5y_6 - x_6y_5 - x_7y_8 + x_8y_7)^2 + \\ (x_1y_3 - x_2y_4 + x_3y_1 + x_4y_2 + x_5y_7 + x_6y_8 - x_7y_5 - x_8y_6)^2 + \\ (x_1y_4 + x_2y_3 - x_3y_2 + x_4y_1 + x_5y_8 - x_6y_7 + x_7y_6 - x_8y_5)^2 + \\ (x_1y_5 - x_2y_6 - x_3y_7 - x_4y_8 + x_5y_1 + x_6y_2 + x_7y_3 + x_8y_4)^2 + \\ (x_1y_6 + x_2y_5 - x_3y_8 + x_4y_7 - x_5y_2 + x_6y_1 - x_7y_4 + x_8y_3)^2 + \\ (x_1y_7 + x_2y_8 + x_3y_5 - x_4y_6 - x_5y_3 + x_6y_4 + x_7y_1 - x_8y_2)^2 + \\ (x_1y_8 - x_2y_7 + x_3y_6 + x_4y_5 - x_5y_4 - x_6y_3 + x_7y_2 + x_8y_1)^2\end{aligned}$$

which was originally discovered by Degen in 1818 via different methods.

A remarkable theorem of Pfister (1965) — proved below as Corollary 18.12 — states that $x_1^2 + \cdots + x_{2^n}^2$ is a group form for any $n \geq 1$. This theorem is all the more remarkable given Hurwitz's theorem (1898) which states that

$$(x_1^2 + \cdots + x_m^2)(y_1^2 + \cdots + y_m^2) = z_1^2 + \cdots + z_m^2$$

for some z_i homogeneous quadratic polynomials in $x_1, \ldots, x_m, y_1, \ldots, y_m$ if and only $m = 2$, 4, or 8; naturally, Pfister's z_i are *not* quadratic forms. We will study Pfister's theory later in the course.

Given symmetric bilinear forms (V, B) and (V', B') over k, we may form their *orthogonal sum*, $V \perp V'$, consisting of the vector space $V \oplus V'$ equipped with the map $B \perp B' \colon (V \oplus V') \times (V \oplus V') \to \mathsf{k}$ given by

$$(B \perp B')((x_1, y_1), (x_2, y_2)) = B(x_1, x_2) + B'(y_1, y_2),$$

which is clearly symmetric and bilinear. This has the effect of making $V \oplus 0$ and $0 \oplus V'$ orthogonal to each other, with $B \perp B'$ restricting on these subspaces to B and B'. Observe further that on associated quadratic spaces we get

$$\begin{aligned}q_{B \perp B'}(x, y) &= (B \perp B')((x, y), (x, y)) \\ &= B(x, x) + B'(y, y) \\ &= q_B(x) + q_{B'}(y),\end{aligned}$$

leading to a natural definition of orthogonal sum of quadratic spaces and forms. The definition on symmetric bilinear forms also mandates that orthogonal sum of symmetric matrices should be

block sum:

$$A \perp A' = \begin{pmatrix} A & 0 \\ 0 & A' \end{pmatrix}.$$

From this description, we easily see that the orthogonal sum of two forms is regular if and only if both of the constituent forms is regular.

Definition 6.5. A quadratic form f is *diagonal* when it is of the form $\lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2$. We denote this form by $\langle \lambda_1, \ldots, \lambda_n \rangle$.

Note that $\langle \lambda_1, \ldots, \lambda_n \rangle$ has a matrix of symmetrized coefficients which is diagonal with entries $\lambda_1, \ldots, \lambda_n$ on the diagonal. We also have

$$\langle \lambda_1, \ldots, \lambda_n \rangle = \langle \lambda_1 \rangle \perp \cdots \perp \langle \lambda_n \rangle .$$

We will use the following two lemmata on our way to proving that every quadratic form is equivalent to a diagonal one.

Lemma 6.6. Suppose U and U' are linear subspaces of V and that (V, B) is a symmetric bilinear form. If $U \cap U' = 0$, $\dim U + \dim U' = \dim V$, and $B(U, U') = 0$, then $V \cong U \oplus U'$ and $B \cong B|_U \perp B|_{U'}$.

Proof. Homework. (Do not invoke any of the following material in your proof.) □

Lemma 6.7 (Representation Criterion)**.** Let (V, B) be a symmetric bilinear form, and $\lambda \in \mathsf{k}^\times$. Then $\lambda \in D(B)$ if and only if $B \cong \langle \lambda \rangle \perp B'$ for some other symmetric bilinear form (V', B').

Proof. Suppose $B \cong \langle \lambda \rangle \perp B'$ and let $q' = q_{B'}$. Then $(\langle \lambda \rangle \perp q')(1, 0) = \lambda$. By Proposition 6.2(a), we conclude that B also represents λ.

For the converse, suppose that $\lambda \in D(B)$ with $\lambda = q(v)$, where $q = q_B$. Without loss of generality, we may assume that V is regular.[3] Now $U = \operatorname{span}\{v\} \cong \langle \lambda \rangle$, and $U \cap U^\perp = 0$. By Proposition 5.6, $\dim U + \dim U^\perp = \dim V$. By Lemma 6.6 we may conclude that

$$B \cong \langle \lambda \rangle \perp B|_{U^\perp}.$$

□

Theorem 6.8 (Diagonalizability of quadratic forms)**.** *If (V, B) is a symmetric bilinear form over* k*, then there exist $\lambda_1, \ldots, \lambda_n \in \mathsf{k}$ such that*

$$B \cong \langle \lambda_1, \ldots, \lambda_n \rangle .$$

Proof. If $D(B) = \varnothing$, then B is the 0 form which can be written as $\langle 0, \ldots, 0 \rangle$ where there are $\dim V$ many 0's. If $D(B) \neq \varnothing$, then there exists some $\lambda \in D(B)$, whence Lemma 6.7 implies that $B \cong \langle \lambda \rangle \perp B'$, and this proves the theorem by induction. □

In order to lighten our notational load, we shall now start abusing notation and referring to a symmetric bilinear form (V, B) as just V (rather than B). This has the advantage of allowing use to refer to the restriction of this form to a subspace U, namely $(U, B|_U)$, as just U, without reference to $B|_U$.

Corollary 6.9. If (V, B) is a symmetric bilinear form and U is a subspace of V, then

(a) if U is regular, then $V \cong U \perp U^\perp$,
(b) if U is regular and U' is a subspace of V such that $V \cong U \perp U'$, then $U' = U^\perp$, and
(c) if V is regular, then U is regular if and only if there exists $U' \leq V$ such that $V \cong U \perp U'$.

[3] The argument goes like this: There is a subspace W such that $V = V^\perp \oplus W = V^\perp \perp W$, and $D(V) = D(W)$. Clearly W is regular.

Proof. To prove (a), note that regularity of U implies that $U \cap U^\perp = 0$. Thus by Lemma 6.6, it suffices to show that U and $U^\perp$ span V. By Theorem 6.8, we may diagonalize U so that it has an orthogonal basis $u_1, \ldots, u_p$ with $B(u_i, u_i) \neq 0$ for all i (by regularity of U). Given $v \in V$, use the old Gram-Schmidt trick to construct

$$u' = v - \sum_{i=1}^{p} \frac{B(v, u_i)}{B(u_i, u_i)} u_i.$$

A straightforward computation show that $B(u', u_j) = 0$ for all j, whence $u' \in U^\perp$. We conclude that

$$v = u' + \sum_{i=1}^{p} \frac{B(v, u_i)}{B(u_i, u_i)} u_i$$

is in the span of U and $U^\perp$.

It is now easy to show that (a) implies (b). If $V = U \perp U'$, then $U' \subseteq U^\perp$ and $\dim V = \dim U + \dim U'$. By (a), $\dim V$ also equals $\dim U + \dim U^\perp$, so $\dim U' = \dim U^\perp$, and we conclude that $U' = U^\perp$.

The left-to-right direction of (c), follows by taking $U' = U^\perp$ and applying (a). For the right-to-left direction, suppose V is regular, $U \leq V$, and $V \cong U \perp U'$. By Theorem 6.8, we may diagonalize $U \cong \langle \lambda_1, \ldots, \lambda_m \rangle$ and $U' \cong \langle \lambda_{m+1}, \ldots, \lambda_n \rangle$, whence $V \cong \langle \lambda_1, \ldots, \lambda_n \rangle$. If any of the $\lambda_i = 0$, then V is not regular, contradicting a hypothesis. Thus U has a diagonalization without 0's and hence is regular. □

Remark 6.10. It is fruitful to compare part (c) of the corollary with Proposition 5.6(a). While (for (V, B) regular) we always have the dimension formula $\dim U + \dim U^\perp = \dim V$, we only have the decomposition $U \perp U^\perp \cong V$ when U is a regular subspace.

Having deduced these theoretical corollaries, let's return to diagonalization itself and think about how to explicitly determine the diagonalization of a quadratic form.

Example 6.11. Consider a binary quadratic form $ax^2 + bxy + cy^2$ for which $a \neq 0$. *Completing the square* produces the formula

$$ax^2 + bxy + cy^2 = \frac{1}{4a}\left((2ax + by)^2 - (b^2 - 4ac)y^2\right).$$

We conclude that

$$ax^2 + bxy + cy^2 \cong \left\langle \frac{1}{4a}, \frac{-(b^2 - 4ac)}{4a} \right\rangle.$$

Since $\langle \lambda^2 \rangle \cong \langle 1 \rangle$ and $\langle \lambda^{-1} \rangle \cong \langle \lambda \rangle$ for all $\lambda \in \mathsf{k}^\times$, we can rewrite this as

$$ax^2 + bxy + cy^2 \cong \langle a, -a(b^2 - 4ac) \rangle.$$

If $a = 0$ and $c \neq 0$, a symmetric argument gives that

$$bxy + cy^2 \cong \langle c, -cb^2 \rangle.$$

If $b \neq 0$, this is equivalent to $\langle c, -c \rangle \cong \langle 1, -1 \rangle = h$.

The final nontrivial case is $a = c = 0$ and $b \neq 0$. We have already seen that $xy \cong h$, and we leave it as an exercise to show that $bxy \cong h$ as well.

Example 6.12. While less well-known, we may actually complete the square in more than two variables. Here we exhibit this process by example and trust that the reader can generalize it. Consider the form

$$f = x^2 - \frac{2}{3}xy + 5xz + 3y^2 + 2yz - 8z^2.$$

In order to complete the square relative to x, we note that there are mixed terms $-\frac{2}{3}xy$ and $5xz$ involving x, and that

$$\left(x - \frac{1}{3}y + \frac{5}{2}z\right)^2 = x^2 - \frac{2}{3}xy + 5xz + \frac{1}{9}y^2 - \frac{5}{3}yz + \frac{25}{4}z^2.$$

As such, letting $\tilde{x} = x - \frac{1}{3}y + \frac{5}{2}z$, we may rewrite the original form f as

$$f = \tilde{x}^2 + \frac{26}{9}y^2 + \frac{11}{3}yz - \frac{57}{4}z^2.$$

Applying the previous example to the form $\frac{26}{9}y^2 + \frac{11}{3}yz - \frac{57}{4}z^2$ reveals that

$$f \cong \left\langle 1, \frac{26}{9}, -\frac{41678}{81} \right\rangle.$$

Since 9 and 81 are squares, this reduces to

$$f \cong \langle 1, 26, -41678 \rangle.$$

Remark 6.13. In the previous example, we always had access to a squared variable from which to start completing the square. If our form only consists of mixed terms, we would have to first perform a linear change of variables to produce a square and then continue with completing the square.

What if our form is given to us as a symmetic matrix? or if we simply prefer working with matrices? With a couple of extra observations, the proof of Theorem 6.8 may be turned into an algorithm that works well in this context.

Begin with an $n \times n$ symmetric matrix $A \in \mathrm{Sym}_{n\times n}(\mathsf{k})$ and suppose that $\lambda = A_{11}$ (the top left entry) is nonzero. (This is equivalent to $B_A(e_1, e_1) \neq 0$, *i.e.*, to the assumption that we may take the standard basis vector e_1 as part of the orthogonal basis in the proof of Theorem 6.8. We will handle the case $\lambda = 0$ shortly.) We may decompose A as

$$A = \left(\begin{array}{c|c} \lambda & v^\top \\ \hline v & A' \end{array}\right)$$

where $v \in \mathbb{R}^{n-1}$ is a column vector and $A' \in \mathrm{Sym}_{(n-1)\times(n-1)}(\mathsf{k})$. Define

$$A_2 = A' - \lambda^{-1} v v^\top$$

and let

$$B = \left(\begin{array}{c|c} 1 & \lambda^{-1}v^\top \\ \hline 0 & I_{n-1} \end{array}\right) \in \mathrm{GL}_n(\mathsf{k}).$$

A direct computation shows that

$$A = B^\top \left(\begin{array}{c|c} \lambda & 0^\top \\ \hline 0 & A_2 \end{array}\right) B.$$

(This should be compared with the "Gram-Schmidt" step in the proof of Corollary 6.9(a).)

If $\lambda = A_{11} = 0$, then we need to change variables to get something nonzero in the upper left corner. There are many ways to do so, and we will only briefly sketch one method: Pick i and j such that $a = A_{ij} \neq 0$. Assume for simplicity that neither i nor j is 0. Change variables so that $x_i \mapsto x_1 + x_2$, $x_j \mapsto x_1 - x_2$, $x_1 \mapsto x_i$, and $x_2 \mapsto x_j$ while all other variables are fixed. This results in a symmetric matrix with upper left-hand block $\begin{pmatrix} 2a & 0 \\ 0 & -2a \end{pmatrix}$. We may then proceed as above. (We leave it to the reader to work out the cases in which one or both of i and j are not distinct from 1 and 2.)

Applying the above process iteratively results in a diagonalization of A. We exhibit this for a specific 3×3 symmetric matrix presently.

Example 6.14. Consider the form $f = 2xy + 4xz + 2yz$ which has matrix

$$A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}.$$

The coordinate change $x \mapsto x + y$, $y \mapsto x - y$, $z \mapsto z$ transforms A into

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 3 \\ 0 & -2 & 1 \\ 3 & 1 & 0 \end{pmatrix}.$$

We now form

$$A_2 = \begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 0 \\ 3 \end{pmatrix} \begin{pmatrix} 0 & 3 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 1 & -\frac{9}{2} \end{pmatrix}$$

and conclude that

$$A \cong \begin{pmatrix} 2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -\frac{9}{2} \end{pmatrix}.$$

Either via another round of matrix manipulations or an application of Example 6.12, the right-hand matrix is in turn congruent to

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

We conclude that $f \cong \langle 2, -2, -1 \rangle \cong \langle 1, -1, -1 \rangle$, where the final equivalence uses the observation that $\langle \lambda, -\lambda \rangle \cong h$ for all $\lambda \in \mathsf{k}^\times$.

For some curious examples of diagonalizations of n-ary quadratic forms (n arbitrary), see T.Y. Lam's article [Lam99].

These algorithms are implemented in the computer algebra system SageMath. The relevant documentation is available at this link.[4] In order to diagonalize a quadratic form `q` (initialized via the command `QuadraticForm`), one calls `q.rational_diagonal_form()`. While this is convenient for checking one's work, remember that there are many diagonalizations of a given quadratic form.

Remark 6.15. Many multivariable calculus courses use the diagonalizability of real quadratic forms in order classify critical points via the Hessian matrix. Curiously, almost all of these texts invoke the spectral theorem which says that an $n \times n$ real matrix is orthogonally diagonalizable if and only if it is symmetric. Here *orthogonal diagonalizability* of A means that there is an invertible matrix P such that $P^{-1} = P^\top$ and $P^{-1}AP$ is diagonal. While this argument is perfectly accurate, it might leave the impression that diagonalization of real quadratic forms requires deep, analytical results (like the spectral theorem). Nothing could be further from the truth. While the above discussion might not be recognizable to an ancient Babylonian, the essential idea (completing the square) would be familiar.

[4] https://doc.sagemath.org/html/en/reference/quadratic_forms/sage/quadratic_forms/quadratic_form.html

7. Hyperbolic spaces

We have already seen the binary quadratic form $h = x^2 - y^2$ and exhibited its equivalence with xy in Example 4.1. The name h is for *hyperbolic*, and we call the associated quadratic space the *hyperbolic plane*. A quadratic space is a called a *hyperbolic space* if it is equivalent to an orthogonal sum of hyperbolic planes, $h \perp h \perp \cdots \perp h = n\langle 1, -1\rangle \cong n\langle 1\rangle \perp n\langle -1\rangle$. (Here we use the convention that $n \cdot f$ is the n-fold orthogonal sum of f with itself.) These simple spaces will play an outsized role in the theory of quadratic forms, largely because of their connection to the notion of isotropy.

Definition 7.1. Let (V, B) be a symmetric bilinear form and take $0 \neq v \in V$. We say that v is an *isotropic vector* if $B(v, v) = 0$, and otherwise we call v *anisotropic*. We call (V, B) an *isotropic form* if it contains an isotropic vector, and otherwise we call (V, B) *anisotropic*. When $B = 0$, we call (V, B) *totally isotropic*.

We need the notion of the *determinant* of a regular quadratic form in order to state the following theorem. Given a regular quadratic form f, let $A_f \in \mathrm{Sym}_{n\times n}(\mathsf{k})$ denote its associated symmetric matrix. We would like to create an invariant of the equivalence class of f out of the matrix determinant $\det(A_f)$. Recall that the forms equivalent to f correspond to symmetric matrices of the form $B^\top A_f B$ where $B \in \mathrm{GL}_n(\mathsf{k})$, then note that

$$\det(B^\top A_f B) = \det(B)^2 \cdot \det(A_f).$$

Thus we make the following definition.

Definition 7.2. The *determinant* of a regular quadratic form f over k with associated symmetric matrix A_f is

$$\det(f) = \det(A_f) \cdot \mathsf{k}^{\boxtimes} \in \mathsf{k}^\times/\mathsf{k}^{\boxtimes}.$$

The preceding discussion shows that $\det(f)$ is constant on the equivalence class of f. Given that orthogonal sum of quadratic forms corresponds to block sum of symmetric matrices, it is also easy to see that

$$\det(f \perp g) = \det(f)\det(g) \in \mathsf{k}^\times/\mathsf{k}^{\boxtimes}. \tag{7.3}$$

Proposition 7.4. The determinant of a diagonal quadratic form is given by

$$\det\langle \lambda_1, \ldots, \lambda_n\rangle = \lambda_1 \cdots \lambda_n \mathsf{k}^{\boxtimes}.$$

Proof. The associated matrix is diagonal with entries $\lambda_1, \ldots, \lambda_n$ along the diagonal. □

We are now ready to state a theorem linking the hyperbolic plane, isotropy, and determinants. Recall that we will freely translate our definitions through the (□), thus the determinant of a symmetric bilinear form will be the square class of the determinant of its Gram matrix.

Theorem 7.5. *Let (V, B) be a symmetric bilinear form with $\dim V = 2$. Then the following statements are equivalent:*

(a) B is regular and isotropic,
(b) B is regular with $\det B = -1 \cdot \mathsf{k}^{\boxtimes}$,
(c) $B \cong \langle 1, -1\rangle$,
(d) B corresponds to the equivalence class of the binary quadratic form xy.

Proof. We have already seen that (c) and (d) are equivalent in Example 4.1. We now check that (a) implies (b). Choose an orthogonal basis v_1, v_2 of V. By regularity, $\mu_i = B(v_i, v_i) \neq 0$ for $i = 1, 2$. For an isotropic vector $v \in V$, write $v = \lambda_1 v_1 + \lambda_2 v_2$ and, without loss of generality, suppose that

$\lambda_1 \neq 0$. Then $0 = \lambda_1^2\mu_1 + \lambda_2^2\mu_2$, whence $\mu_1 = -(\lambda_2\lambda_1^{-1})^2\mu_2$. It follows that $\det B = \mu_1\mu_2 \cdot \mathsf{k}^{\boxtimes} = -1 \cdot \mathsf{k}^{\boxtimes}$, as desired.

We now check that (b) implies (c). The form B has a diagonalization $\langle a, b\rangle$ for some $a, b \in \mathsf{k}^\times$, and we know that $\det B = ab \cdot \mathsf{k}^{\boxtimes} = -1 \cdot \mathsf{k}^{\boxtimes}$. As such, b has the same square class as $-a^{-1}$, which has the same square class of $-a$. Thus $B \cong \langle a, -a\rangle$, which is isometric to the symmetric bilinear form corresponding to axy. This binary quadratic form clearly represents all elements of $\mathsf{k}^\times$, hence so does B. Thus B represents 1 and the Representation Criterion Lemma 6.7 implies that $B \cong \langle 1, \lambda\rangle$ with $\lambda \equiv -1 \pmod{\mathsf{k}^{\boxtimes}}$. Thus $B \cong \langle 1, -1\rangle$, as desired.

Finally, the implication (c) $\implies$ (a) is trivial, so we have proven the theorem. □

Definition 7.6. If $D(B) = \mathsf{k}^\times$, then we call B *universal*.

Remark 7.7. Every universal form is a group form.

Theorem 7.8. *Let (V, B) be a regular symmetric bilinear form. Then*

(a) every totally isotropic subspace $U \leq V$ of dimension r is contained in a hyperbolic subspace $T \leq V$ of dimension $2r$,
(b) V is isotropic if and only if V contains a hyperbolic plane, and
(c) if V is isotropic, then V is universal.

Proof. We prove (a) by induction on r. If $r = 0$, then U is contained in the 0-fold orthogonal sum of hyperbolic planes. Let $v_1, \ldots, v_r$ be an orthogonal basis of U and let $S = \operatorname{span}\{v_2, \ldots, v_r\}$. Note that $U^\perp \subseteq S^\perp$. Since V is regular, we have

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

by Lemma 6.6. Thus there exists a vector w_1 orthogonal to $v_2, \ldots, v_r$, but not orthogonal to v_1. In particular, v_1, w_1 are linearly indpendent vectors since $B(v_1, v_1) = 0$. The subspace $H = \operatorname{span}\{v_1, w_1\}$ has determinant

$$\det H = \det \begin{pmatrix} 0 & B(v_1, w_1) \\ B(v_1, w_1) & B(w_1, w_1) \end{pmatrix} \cdot \mathsf{k}^{\boxtimes} = -1 \cdot \mathsf{k}^{\boxtimes},$$

so $H \cong h$ by Theorem 7.5. By Corollary 6.9, we get that $V = H \perp H^\perp$, and $H^\perp$ is regular and contains $v_2, \ldots, v_r$. The proof of (a) now follows by induction.

We leave it as an exercise for the reader that (b) and (c) now follow easily. □

Remark 7.9. Universality of h can be checked directly via the identity

$$\lambda = \left(\frac{\lambda + 1}{2}\right)^2 - \left(\frac{\lambda - 1}{2}\right)^2.$$

Corollary 7.10 (First Representation Theorem)**.** Let f be a regular quadratic form and let $\lambda \in \mathsf{k}^\times$. Then $\lambda \in D(f)$ if and only if $f \perp \langle -\lambda\rangle$ is isotropic.

Proof. Without loss of generality, we may assume throughout that $f = \langle \lambda_1, \ldots, \lambda_n\rangle$.

First suppose that $\lambda \in D(f)$. If $\lambda = \sum \lambda_i x_i^2$ for some $x_i \in \mathsf{k}$, then $\left(\sum \lambda_i x_i^2\right) + (-\lambda) \cdot 1^2 = 0$, so $f \perp \langle -\lambda\rangle$ is isotropic.

Conversely, suppose that $f \perp \langle -\lambda\rangle$ has isotropic vector $(x_1, \ldots, x_{n+1})$. If $x_{n+1} \neq 0$, then

$$\lambda = \sum \lambda_i \left(\frac{x_i}{x_{n+1}}\right)^2 \in D(f).$$

If $x_{n+1} = 0$, then $(x_1, \ldots, x_n) \neq 0$ is an isotropic vector for f, whence $D(f) = \mathsf{k}^\times$, so $\lambda \in D(f)$. □

Corollary 7.11. Let f and g be regular quadratic forms of positive dimensions. Then $f \perp g$ is isotropic if and only if $D(f) \cap -D(g) \neq \varnothing$.

Proof. Begin by assuming that $\lambda \in D(f) \cap -D(g)$ with $f(x) = \lambda$ and $g(y) = -\lambda$. Then $(x, y) \neq 0$ is an isotropic vector for $f \perp g$.

Now assume that $D(f) \cap -D(g) \neq \varnothing$. Without loss of generality, f and g are anisotropic.[5] Suppose $f(x) + g(y) = 0$ where $(x, y) \neq 0$. Say $x \neq 0$. Then $f(x) \neq 0$ and $f(x) \in D(f) \cap -D(g)$, as desired. □

Corollary 7.12. For $r > 0$, the following are equivalent:

(a) Any regular quadratic form of dimension r over k is universal.
(b) Any quadratic form of dimension $r + 1$ over k is isotropic.

Proof. First suppose that all regular quadratic forms of dimension r over k are universal. Any quadratic form f of dimension $r + 1$ over k has a diagonalization $\langle \lambda_1, \ldots, \lambda_{r+1} \rangle$. If all $\lambda_i = 0$, then we certainly have an isotropic form. If some λ_i, say λ_{r+1}, is nonzero, then $-\lambda_{r+1} \in D(f)$, and the previous corollary implies that f is isotropic, as desired.

Now suppose that all quadratic forms of dimension $r + 1$ over k are isotropic, and let f denote a regular quadratic form of dimension r over k. For $\lambda \in \mathsf{k}^\times$, we have that $f \perp \langle -\lambda \rangle$ is isotropic, whence $\lambda \in D(f)$ by the previous corollary. We conclude that f is universal, as desired. □

Remark 7.13. The *u-invariant* of a field k is

$$u(\mathsf{k}) = \max\{\dim f \mid f \text{ an anisotropic form over } \mathsf{k}\} \in \mathbb{N} \cup \{\infty\}.$$

By the above corollaries, we may also express $u(\mathsf{k})$ as the minimum n such that forms of dimension $> n$ over k are isotropic, and also as the minimum n such that forms of dimension $\geq n$ over k are universal. (Here we set $\min \varnothing = \infty$.)

8. Witt decomposition and cancellation

Ernst Witt's 1937 paper *Theorie der quadratischen Formen in beliebigen Körpern* (Theory of quadratic forms in arbitrary fields) initiated the modern algebraic theory of quadratic forms. Amongst its results are the decomposition and cancellation theorems which we introduce presently.

Theorem 8.1 (Witt decomposition theorem)**.** *Any quadratic space (V, q) is equivalent to 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, and V_a is anisotropic; furthermore, the isometry types of V_t, V_h, and V_a are all uniquely determined.

Theorem 8.2 (Witt cancellation theorem)**.** *If f, f', and f'' are arbitrary quadratic forms, then*

$$f \perp f' \cong f \perp f'' \quad \textit{implies} \quad f' \cong f''.$$

Note that neither theorem has a regularity hypothesis. The decomposition theorem permits a decomposition (as an orthogonal sum) of any quadratic space into pieces of proscribed forms, and the cancellation theorem allows us to "cancel" the f from the original equivalence to deduce $f' \cong f''$.

Proof of Theorem 8.1. We begin with the decomposition

$$V = V^\perp \perp V_0$$

where V_0 is any linear subspace such that $V = V^\perp \oplus V_0$. Then $V^\perp$ is totally isotropic and V_0 is regular. If V_0 is isotropic, then we may iteratively apply Theorem 7.8(b) to deduce that

$$V_0 = (H_1 \perp \cdots \perp H_r) \perp V_a$$

[5] If, say, g is isotropic, then g is universal and $D(f) \cap -D(g) = D(f) \neq \varnothing$.

where each $H_i \cong h$ and V_a is anisotropic (but possibly 0). This proves existence of a Witt decomposition of V.

To prove uniqueness, we invoke Witt cancellation Theorem 8.2. (We will prove Witt cancellation shortly; naturally, that proof will not depend on Witt decomposition.) Suppose that there is 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

$$V^\perp = (V_t')^\perp \perp (V_h' \perp V_a')^\perp = V_t',$$

so the totally isotropic pieces match. By Theorem 8.2, we now have $V_h \perp V_a \cong V_h' \perp V_a'$. Suppose $V_h \cong m \cdot h$ and $V_h' \cong m' \cdot h$. Cancelling one h at a time, we conclude that $m = m'$ since V_a and V_a' are both anisotropic. Cancelling V_h and V_h', we finally get that $V_a \cong V_a'$. □

Before we prove the Theorem 8.2, we give a name to $\frac{1}{2}\dim V_h$.

Definition 8.3. For a quadratic space (V, q) with Witt decomposition $V_t \perp V_h \perp V_a$, call the natural number $\frac{1}{2}\dim V_h$ the *Witt index* of q.

Proposition 8.4. If (V, q) is a regular quadratic space, then the Witt index of q equals the dimension of any maximal totally isotropic subspace of V.

Proof. Let $U \leq V$ be a maximal totally isotropic subspace of V, and set $r = \dim U$. By Theorem 7.8, there is a hyperbolic space $T \leq V$ such that $U \leq T$ and $\dim T = 2r$. Since T is regular, $V = T \perp T^\perp$. Note that $T^\perp$ is anisotropic: if $0 \neq x \in T^\perp$ is isotropic, then $U + \operatorname{span}\{x\}$ will be a larger totally isotropic subspace of V, contradicting maximality of U. By the uniqueness portion of Theorem 8.1, we know that the Witt index of q is half the dimension of T, which is $r = \dim U$, as desired. □

In order to prove the cancellation theorem, we need to develop some facts about the *orthogonal group*

$$\mathrm{O}(V) = \mathrm{O}_q(V) = \{\varphi : V \to V \mid \varphi \text{ an isometry of } (V, q)\}.$$

of isometries of V. In particular, we may associate the following *reflection across* v isometry to every anisotropic vector $v \in V$:

$$\begin{aligned} \rho_v : V &\longrightarrow V \\ x &\longmapsto x - \frac{2B(x,v)}{B(v,v)}v. \end{aligned}$$

It is clear that ρ_v is a linear transformation. Also observe that if $x \in \operatorname{span}\{v\}^\perp$, then $\rho_v(x) = x$, so ρ_v restricts to the identity transformation on $\operatorname{span}\{v\}^\perp$. Meanwhile,

$$\rho_v(v) = v - \frac{2B(v,v)}{B(v,v)}v = v - 2v = -v.$$

It follows that ρ_v is an involution which fixes $\operatorname{span}\{v\}^\perp$ pointwise and reflects v across $\operatorname{span}\{v\}^\perp$ to $-v$.

While the above analysis shows that ρ_v is a linear isomorphism, we need to also prove that ρ_v is an isometry, whence $\rho_v \in \mathrm{O}(V)$. We may compute

$$\begin{aligned} B(\rho_v(x), \rho_v(y)) &= B\left(x - \frac{2B(x,v)}{B(v,v)}v, y - \frac{2B(y,v)}{B(v,v)}v\right) \\ &= B(x,y) + \frac{4B(x,v)B(y,v)}{B(v,v)^2}B(v,v) - \frac{4B(x,v)B(y,v)}{B(v,v)} \\ &= B(x,y), \end{aligned}$$

so we indeed have $\rho_v \in \mathrm{O}(V)$.

Finally, we note that ρ_v has determinant -1. This is easy to see if we choose a basis for V compatible with the decomposition

$$V = \mathrm{span}\{v\} \perp \mathrm{span}\{v\}^\perp$$

and use the above observations.

Remark 8.5. It is actually the case that reflections through anisotropic vectors generate the group $\mathrm{O}(V)$. We will not use this result in the sequel and thus do not go through the proof here. The reader may easily check that $\langle \rho_v \mid v \in V \text{ anisotropic} \rangle$ is normal in $\mathrm{O}(V)$; it is more difficult to prove that this subgroup is all of $\mathrm{O}(V)$.

Remark 8.6. The reader may be familiar with the orthogonal group $\mathrm{O}(n)$ of real $n \times n$ matrices A with $A^\top = A^{-1}$. This is in fact an orthogonal group in the above sense, namely the one associated with $(\mathbb{R}^n, n\langle 1 \rangle)$. In order to see this, let's examine which matrices $A \in \mathrm{GL}_n(\mathsf{k})$ correspond to isometries of a regular symmmetric bilinear form (k^n, B). It is generally the case that we may form the B-adjoint operator $A^\dagger$ of any linear transformation A of k^n. This satisfies the universal property $B(v, Aw) = B(A^\dagger v, w)$ for all $v, w \in \mathsf{k}^n$. If A is an isometry, then $B(v,w) = B(Av, Aw) = B(A^\dagger A v, w)$ for all v, w. By regularity of B, we get that $A^\dagger A = I$, *i.e.*, $A^\dagger = A^{-1}$. The reader may check that for the form $n\langle 1 \rangle$, the adjoint operator simply takes the transpose of a matrix, *i.e.*, $A^\dagger = A^\top$, so $\mathrm{O}(n) = \mathrm{O}_{n\langle 1 \rangle}(\mathbb{R}^n)$.

Proposition 8.7. Let (V, B) be a symmetric bilinear form, and let v and w be anisotropic vectors in V with $B(v,v) = B(w,w)$.[6] Then there exists $\rho \in \mathrm{O}(V)$ such that $\rho(v) = w$.

Proof Sketch. Let $q = q_B$. Use the parallelogram law from your homework to deduce that we cannot have both $q(v+w)$ and $q(v-w)$ equal to 0. Without loss of generality, we may assume $q(v-w) \neq 0$.[7] We now check (via computation) that the reflection ρ_{v-w} takes v to w. Indeed,

$$\rho_{v-w}(v) = v - \frac{2B(v, v-w)}{B(v-w, v-w)} \cdot (v-w)$$

and

$$\begin{aligned} B(v-w, v-w) &= B(v,v) + B(w,w) - 2B(v,w) \\ &= 2(B(v,v) - B(v,w)) \\ &= 2B(v, v-w). \end{aligned}$$

Thus $\rho_{v-w}(v) = v - (v-w) = w$, and we may take $\rho = \rho_{v-w}$. □

Proof of Theorem 8.2. Suppose that $f \perp f' \cong f \perp f''$. We begin by showing that cancellation holds ($f' \cong f''$) when f is totally anisotropic and f' is regular. In this case, the associated matrices of $f \perp f'$ and $f \perp f''$ take the forms $\begin{pmatrix} 0 & 0 \\ 0 & A_{f'} \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ 0 & A_{f''} \end{pmatrix}$, respectively. Since $f \perp f' \cong f \perp f''$, there exists an invertible matrix $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ (where a, b, c, d are block matrices) such that

$$\begin{pmatrix} 0 & 0 \\ 0 & A_{f'} \end{pmatrix} = B^\top \begin{pmatrix} 0 & 0 \\ 0 & A_{f''} \end{pmatrix} B = \begin{pmatrix} * & * \\ * & d^\top A_{f''} d \end{pmatrix}.$$

Thus $A_{f'} = d^\top A_{f''} d$. The reader may check that d is nonsingular (because f' is regular), so we conclude that $f' \cong f''$ in this case.

[6] The draft of these notes from Friday, September 28 omitted the hypothesis $B(v,v) = B(w,w)$ — whoops! In a more rested or better caffeinated state, we could have realized that isometries preserve B, so the hypothesis is necessary.

[7] This proceeds by replacing w by $-w$. If $\rho \in \mathrm{O}(V)$ satisfies $\rho(v) = -w$, then $-\rho \in \mathrm{O}(V)$ takes v to w.

Now consider the case in which f is toally anisotropic and f' is not necessarily regular. We may diagonalize f' and f'' and assume (without loss of generality) that f' has exactly r zeroes in its diagonalization while f'' has r or more zeroes in its diagonalization. Then

$$f \perp r\langle 0\rangle \perp f'_0 \cong f \perp r\langle 0\rangle \perp f''_0.$$

Since f'_0 is regular, the previous paragraph implies that $f'_0 \cong f''_0$. Taking orthogonal sum with $r\langle 0\rangle$ implies that $f' \cong f''$.

We may now handle the general case. Let $\langle \lambda_1, \dots, \lambda_n\rangle$ be a diagonalization of f. Inducing[8] on n, we are reduced to the case $n = 1$. If $\lambda_1 = 0$, we are done by the previous case, so assume that $f = \langle \lambda_1\rangle$ and $\lambda_1 \neq 0$. We aim to show that $\varphi : \langle \lambda_1\rangle \perp f' \cong \langle \lambda_1\rangle \perp f''$ implies $f' \cong f''$ (where we are thinking of the equivalence of forms φ as a linear isomorphim $V \to W$ that preserves symmetric bilinear forms). Let $v \in V$ have first coordinate 1 and other coordinates 0, and similarly define $w \in W$. Then $B(v, v) = B'(w, w) = \lambda_1 \neq 0$. Let $\rho \in \mathrm{O}(V)$ denote the isometry of V taking v to $\varphi^{-1}(w)$ (guaranteed to exist by Proposition 8.7). Then $\varphi \circ \rho : V \to V'$ is an isometry taking v to w. Hence the restriction of $\varphi \circ \rho$ to $\mathrm{span}\{v\}^\perp$ corresponds to an equivalence $f' \cong f''$, as desired. □

9. Chain equivalence

We begin with a simple criterion for the equivalence of binary forms.

Proposition 9.1. Let $f = \langle a, b\rangle$ and let $g = \langle c, d\rangle$ be regular binary forms. Then $f \cong g$ if and only if $\det f = \det g$ and $D(f) \cap D(g) \neq \varnothing$.

Proof. The forwards implication is clear. For the converse, assume $\det f = \det g \in \mathsf{k}^\times/\mathsf{k}^{\boxtimes}$ and $\lambda \in D(f) \cap D(g)$. By the First Representation Criterion (Corollary 7.10), we have $f \cong \langle \lambda, \lambda'\rangle$ for some $\lambda' \in \mathsf{k}^\times$. The determinant condition implies that $ab\mathsf{k}^{\boxtimes} = \lambda\lambda'\mathsf{k}^{\boxtimes}$, whence $f \cong \langle \lambda, ab\lambda\rangle$. Similarly, $g \cong \langle \lambda, cd\lambda\rangle$. Since $ab\mathsf{k}^{\boxtimes} = cd\mathsf{k}^{\boxtimes}$, we get that $f \cong g$. □

It turns out that equivalence of forms is the same thing as iterative equivalence of binary subforms, a notion we make precise in the following definition.

Definition 9.2. Let $f = \langle \lambda_1, \dots, \lambda_n\rangle$ and $g = \langle \mu_1, \dots, \mu_n\rangle$ for $\lambda_i, \mu_i \in \mathsf{k}$. We call f and g *simply equivalent* if there are indices i and j such that $\langle \lambda_i, \lambda_j\rangle \cong \langle \mu_i, \mu_j\rangle$ and $\lambda_k = \mu_k$ for $k \neq i, j$. (We make the convetion that when $i = j$ we interpret $\langle \lambda_i, \lambda_j\rangle$ to be just $\langle \lambda_i\rangle$.)

Two diagonal forms f and g are *chain equivalent* if there exists a sequence of diagonal forms $f = f_0, f_1, \dots, f_m = g$ with each f_i simply equivalent to f_{i+1}, $0 \leq i \leq m - 1$. In this case, we write $f \approx g$.

The reader should take care to note that simple equivalence is *not* an equivalence relation; rather, it generates the equivalence relation of chain equivalence. By Proposition 9.1, we know that $f \approx g$ implies that $f \cong g$. The converse is true as well:

Theorem 9.3. *If $f = \langle \lambda_1, \dots, \lambda_n\rangle$ and $g = \langle \mu_1, \dots, \mu_n\rangle$ are arbitrary diagonal forms of the same dimension, then $f \cong g$ implies that $f \approx g$.*

Proof. We first reduce the problem to regular forms. Given a diagonal form $f = \langle \lambda_1, \dots, \lambda_n\rangle$ and σ a permutation of $\{1, \dots, n\}$, we define $f^\sigma = \langle \lambda_{\sigma(1)}, \dots, \lambda_{\sigma(n)}\rangle$, which gives a right action of the permutation group Σ_n on diagonal forms. Since Σ_n is generated by transpositions, we know that $f \approx f^\sigma$. We may thus arrange for all of the 0's in our diagonal forms to come first. By Witt deccomposition and cancellation, we may assume f and g are regular.

[8] Mel Hochster (a preeminent commutative algebraist based at the University of Michigan) insists that the word "induct" is a back-formation and that when we perform induction, we are inducing, not inducting. The author is too enamored with this argument to disagree.

We now proceed by induction on the dimension n. If $n = 1$ or 2, then there is nothing to prove. Fix $n \geq 3$. By the well-ordering principle,[9] we may choose a diagonal form $f' = \langle \zeta_1, \ldots, \zeta_n \rangle$ which is chain-equivalent to f and such that $\langle \zeta_1, \ldots, \zeta_p \rangle$ represents μ_1, and p is as small as possible.

We claim that $p = 1$. Suppose for contradiction that $p > 1$ and write $\mu_1 = \zeta_1\tau_1^2 + \cdots + \zeta_p\tau_p^2$. Since p is minimal, $\delta = \zeta_1\tau_1^2 + \zeta_2\tau_2^2 \neq 0$. By Proposition 9.1, $\langle \zeta_1, \zeta_2 \rangle \cong \langle \delta, \zeta_1\zeta_2\delta \rangle$. Thus f is chain equivalent to f' which is chain equivalent to $\langle \delta, \zeta_1\zeta_2\delta, \zeta_3, \ldots, \zeta_p, \ldots, \zeta_n \rangle$. Permuting terms, we learn that

$$f \approx \langle \delta, \zeta_3, \ldots, \zeta_p, \ldots, \zeta_n, \zeta_1\zeta_2\delta \rangle .$$

Since $\mu_1 = \delta + \zeta_3\tau_3^2 + \ldots + \zeta_p\tau_p^2$ is represented by $\langle \delta, \zeta_3, \ldots, \zeta_p \rangle$, we have reached a contradiction.

Since $p = 1$, we know that $\langle \zeta_1 \rangle \cong \langle \mu_1 \rangle$, whence $f \approx \langle \mu_1, \zeta_2, \ldots, \zeta_n \rangle \cong \langle \mu_1, \ldots, \mu_n \rangle$. By cancellation,

$$\langle \zeta_2, \ldots, \zeta_n \rangle \cong \langle \mu_2, \ldots, \mu_n \rangle .$$

By the induction hypothesis, these forms are also chain equivalent. We conclude that

$$f \approx \langle \mu_1, \zeta_2, \ldots, \zeta_n \rangle \approx \langle \mu_1, \mu_2, \ldots, \mu_n \rangle = g,$$

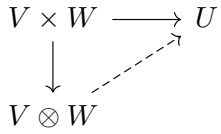
as desired. □

10. Tensor product of vector spaces and quadratic forms

Thus far, we have frequently used orthogonal sum, $\perp$, on quadratic and symmetric bilinear forms. This "addition operation" actually distributes over a "multiplication" given by the tensor (or Kronecker) product of forms. In order to develop this notion, we begin by investigating tensor products of vector spaces.

10.1. **Tensor products in linear algebra.** Given k-vector spaces V and W, we may form their *tensor product* $V \otimes W = V \otimes_{\mathsf{k}} W$. The universal property of $V \otimes W$ is embedded in the following proposition, and a construction of $V \otimes W$ is contained within its proof.

Proposition 10.1. For every pair of k-vector spaces V and W, there is a k-vector space $V \otimes W$ admitting a bilinear map $V \times W \to V \otimes W$ and such that every k-bilinear map $V \times W \to U$ factors uniquely as



where $V \otimes W \to U$ is a linear transformation.

Proof. Let I be the set underlying $V \times W$ and define $\widetilde{T}$ to be the k-vector space with basis I, *i.e.*, $\widetilde{T} = \mathsf{k}^{\oplus I}$.[10] For each $i = (v, w) \in I$, we will also let (v, w) denote the image of i in $\widetilde{T}$. (This element of the direct sum has entry 1 in the i-th coordinate and 0 in all other coordinates.) Define $R \leq \widetilde{T}$ to be the subspace spanned by vectors of the form

(i) $(v_1 + v_2, w) - (v_1, w) - (v_2, w)$ for $v_i \in V$, $w \in W$,
(ii) $(v, w_1 + w_2) - (v, w_1) - (v, w_2)$ for $v \in V$, $w_i \in W$,
(iii) $(\lambda v, w) - \lambda(v, w)$ for $\lambda \in \mathsf{k}$, $v \in V$, $w \in W$,
(iv) $(v, \lambda w) - \lambda(v, w)$ for $\lambda \in \mathsf{k}$, $v \in V$, $w \in W$,

[9] Recall that a set is well-ordered if it is totally ordered and every nonempty subset has a least element under the ordering. The well-ordering principle (or theorem) says that every set has a well-ordering; famously, it is equivalent to the axiom of choice. The author is unaware of a constructive proof of this theorem, but one would be welcome.

[10] Note that in typical cases, I is infinite, so we need to be careful here with the distinction between direct sum and product.

and then define

$$V \otimes W = \widetilde{T}/R.$$

We now verify that $V \otimes W$ satisfies the universal property. To give a linear transformation $\widetilde{B} : \widetilde{T} \to U$ is equivalent to giving a funciton $B : V \times W \to U$ (by the universal property of direct sums). Furthermore, this map factors through $V \otimes W$ if and only if it sends each element of R to 0. We may check this property on the spanning vectors of R. But these are exactly the conditions for bilinearity of the map B, so we are done. The reader may check that the universal properties invoked also give uniqueness of the map. □

Remark 10.2. For $(v, w) \in V \times W$, let $v \otimes w$ denote the corresponding vector in $V \otimes W$. We call such an element of $V \otimes W$ a *simple tensor*. The reader should beware that generic elements of $V \otimes W$ are linear combinations of simple tensors, not just simple tensors.

Remark 10.3. For $n \in \mathbb{N}$, we will let $V^{\otimes n}$ denote the n-fold tensor product of V with itself; if $n = 0$, we make the convention that $V^{\otimes 0} = \mathsf{k}$.

For finite-dimensional vector spaces, we can be a fair bit more concrete and write down a basis for $V \otimes W$ in terms of bases for V and W.

Proposition 10.4. Suppose $e_1, \ldots, e_n$ and $f_1, \ldots, f_m$ are bases of V and W, respectively. Then

$$\{e_i \otimes f_j \mid 1 \leq i \leq n,\ 1 \leq j \leq m\}$$

is a basis of $V \otimes W$. In particular, $\dim V \otimes W = \dim V \cdot \dim W$.

Proof. Given $v = \sum \lambda_i e_i$ and $w = \sum \mu_j f_j$, bilinearity of $V \times W \to V \otimes W$ implies that

$$v \otimes w = \sum_{i,j} \lambda_i \mu_j e_i \otimes f_j.$$

Since simple tensors span $V \otimes W$, we now know that $\{e_i \otimes f_j\}$ spans $V \otimes W$. In particular, $\dim V \otimes W \leq nm$.

To achieve the opposite inequality, we will temporarily denote the standard basis of k^{nm} by $g_1, \ldots, g_{nm}$. The reader may check that the assignment

$$V \times W \longrightarrow k^{nm}$$
$$\left(\sum_i \lambda_i e_i, \sum_j \mu_j f_j \right) \longmapsto \sum_{i,j} \lambda_i \mu_j g_{(i-1)m+j}$$

is a bilinear map. The induced map $V \otimes W \to k^{nm}$ sends $e_i \otimes f_j$ to $g_{(i-1)m+j}$. Since

$$\{(i, j) \mid 1 \leq i \leq n,\ 1 \leq j \leq m\} \longrightarrow \{1, \ldots, nm\}$$
$$(i, j) \longmapsto (i-1)m + j$$

is a bijection, the map $V \otimes W \to k^{nm}$ is surjective, implying that $\dim V \otimes W \geq nm$ as well. We conclude that $\dim V \otimes W = nm$ and that $\{e_i \otimes f_j\}$ is a basis. □

In your homework, you will verify the following crucial properties of tensor products.

Proposition 10.5 (Hom-tensor duality)**.** For all k-vectors spaces V and W, the map

$$\mathrm{Hom}(V \otimes W, U) \longrightarrow \mathrm{Hom}(V, \mathrm{Hom}(W, U))$$
$$f \longmapsto (v \mapsto (w \mapsto f(v \otimes w)))$$

is an isomorphism.

Proposition 10.6. For k-vector spaces V and W where V is finite-dimensional, the map

$$\begin{aligned} V^* \otimes W &\longrightarrow \mathrm{Hom}(V, W) \\ f \otimes w &\longmapsto (v \mapsto f(v)w) \end{aligned}$$

is an isomorphism.

Proposition 10.7. For finite-dimensional k-vector spaces V and W, the map

$$\begin{aligned} V^* \otimes W^* &\longrightarrow (V \otimes W)^* \\ f \otimes g &\longmapsto (v \otimes w \mapsto f(v)g(w)) \end{aligned}$$

is an isomorphism.

10.2. **Tensor products of quadratic forms.** Suppose that (V, B) is a symmetric bilinear form. Then B is a bilinear map $V \times V \to \mathsf{k}$ and thus induces a unique linear transformation $V \otimes V \to \mathsf{k}$ compatible with the canonical map $V \times V \to V \otimes V$. By Proposition 10.5, such a map is "the same" as a linear transformation $V \to V^*$. Abusing notation, we will also refer to $V \to V^*$ as B.

Now suppose that (V, B) and (W, B') are symmetric bilinear forms. We may then form a linear transformation

$$V \otimes W \to V^* \otimes W^* \cong (V \otimes W)^*$$

by taking the taking the tensor product of $B : V \to V^*$ and $B' : W \to W^*$ composed with the isomorphism provided by Proposition 10.7. Further abusing notation, we will refer to this linear transformation (and its associated symmetric bilinear form) as $B \otimes B'$. This makes $(V \otimes W, B \otimes B')$ a symmetric bilinear form called the *tensor* or *Kronecker product* of (V, B) and (W, B').

In the following discussion, we make this construction explicit in all corners of the quadratic square (□). First note that on simple tensors, we have

$$(B \otimes B')(v_1 \otimes w_1, v_2 \otimes w_2) = B(v_1, v2)B'(w_1, w_2),$$

and this rule may be extended by bilinearity to all elements of $(V \otimes W) \times (V \otimes W)$.

If $q = q_B$ and $q' = q_{B'}$, then the quadratic map $q \otimes q' = q_{B \otimes B'}$ satisfies

$$\begin{aligned} (q \otimes q')(v \otimes w) &= (B \otimes B')(v \otimes w, v \otimes w) \\ &= B(v, v)B'(w, w) \\ &= q(v)q'(w). \end{aligned}$$

Now suppose that V has ordered basis $v_1, \ldots, v_m$ and W has ordered basis $w_1, \ldots, w_n$. Let $\lambda_{ij} = B(v_i, v_j)$ and $\mu_{k\ell} = B'(w_k, w_\ell)$, so that $A_B = (\lambda_{ij})$ and $A_{B'} = (\mu_{k\ell})$. Endow $V \otimes W$ with the ordered basis

$$v_1 \otimes w_1, \ldots, v_1 \otimes w_n, v_2 \otimes w_1, \ldots, v_2 \otimes w_n, \ldots, v_n \otimes w_1, \ldots, v_n \otimes w_m.$$

Then $A_{B \otimes B'}$ is the block matrix

$$\begin{pmatrix} \lambda_{11}A_{B'} & \lambda_{12}A_{B'} & \cdots & \lambda_{1m}A_{B'} \\ \lambda_{21}A_{B'} & \lambda_{22}A_{B'} & \cdots & \lambda_{2m}A_{B'} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1}A_{B'} & \lambda_{m2}A_{B'} & \cdots & \lambda_{mm}A_{B'} \end{pmatrix},$$

frequently referred to as the *Kronecker product* of A_B and $A_{B'}$. In particular, for diagonal forms $q = \langle \lambda_1, \ldots, \lambda_m \rangle$ and $q' = \langle \mu_1, \ldots, \mu_n \rangle$ we have

$$q \otimes q' \cong \langle \lambda_1\mu_1, \ldots, \lambda_1\mu_n, \ldots, \lambda_m\mu_1, \ldots, \lambda_m\mu_n \rangle .$$

To determine the tensor product of two quadratic forms, one takes the Kronecker product of their symmetric matrices then turns that back into a quadratic form; the general formula is unilluminating.

Example 10.8. Consider the quadratic forms $f = x^2 + 4xy + 3y^2$ and $g = 2x^2 + 16xy + y^2$ which have symmetric matrices

$$\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 & 8 \\ 8 & 1 \end{pmatrix}.$$

The Kronecker product of these matrices is

$$\begin{pmatrix} 2 & 8 & 4 & 16 \\ 8 & 1 & 16 & 2 \\ 4 & 16 & 6 & 24 \\ 16 & 2 & 24 & 3 \end{pmatrix}.$$

It follows that

$$f \otimes g = 2x^2 + 16xy + 8xz + 32xw + y^2 + 32yz + 4yw + 6z^2 + 48zw + 3w^2.$$

Proposition 10.9. Up to equivalence, the tensor product of quadratic forms satisfies the commutative and associative laws, and distributes over orthogonal sum:

(a) $q \otimes q' \cong q' \otimes q$,
(b) $q \otimes (q' \otimes q'') \cong (q \otimes q') \otimes q''$,
(c) $q \otimes (q' \perp q'') \cong (q \otimes q') \perp (q \otimes q'')$.

Proof. These results are immediate, especially if one works with diagonal forms. □

The following easy lemma will have significant ramifications in our subsequent studies. Recall that for $n \in \mathbb{N}$ and a form q, we write nq for the n-fold orthogonal sum of q with itself.

Lemma 10.10. If q is a regular quadratic form and $h = \langle 1, -1 \rangle$, then

$$q \otimes h \cong (\dim q)h.$$

Proof. Without loss of generality, $q = \langle \lambda_1, \ldots, \lambda_n \rangle$ for some $\lambda_i \in \mathsf{k}^\times$. By distribution of $\otimes$ over $\perp$, we get

$$q \otimes h \cong \langle \lambda_1, -\lambda_1 \rangle \perp \cdots \perp \langle \lambda_n, -\lambda_n \rangle.$$

By a homework problem, each summand is equivalent to h, giving the desired equivalence. □

This is remarkable! Tensor multiplication with the hyperbolic plane is insensitive to everything except the dimension of the form being multiplied.

11. Group completion

The operations $\perp$ and $\otimes$ make the set of isometry classes of regular quadratic forms look an awful lot like a commutative ring. Indeed, we can associatively and commutatively add and multiply, the 0 form is a unit for $\perp$, $\langle 1 \rangle$ is a unit for $\otimes$, and $\otimes$ distributes over $\perp$ (Proposition 10.9). But we are missing a crucial operation: subtraction. This type of structure is actually common in mathematics; it is called a (commutative) semi-ring. Underlying every semi-ring is an Abelian monoid, a set with an associative, commutative addition operation and 0 element. Our present goal is to make a natural construction which turns every Abelian monoid into an Abelian group; when applied to a semi-ring, the multiplication will come along for the ride and we will get a ring. In the next section we will apply this construction to quadratic forms.

Definition 11.1. A *semigroup* $(M, \bullet)$ is a set M together with a binary operation $\bullet$ which is associative. If there exists $e \in M$ that acts as an identity element ($e \bullet m = m = m \bullet e$ for all $m \in M$), then we call M a *monoid*. If $\bullet$ is commutative ($m \bullet n = n \bullet m$ for all $m, n \in M$), we call M an *Abelian monoid*. In a generic monoid, we will typically drop $\bullet$ from our notation and write $mn = m \bullet n$; in an Abelian monoid, we will typically write $+$ for $\bullet$.

Example 11.2. You have probably encountered a great many semigroups and monoids in your mathematical career already.

(1) $(\mathbb{Z}_{>0}, +)$ is an Abelian semigroup.
(2) $(\mathbb{N}, +)$ is an Abelian monoid.
(3) $(\mathbb{Z}_{>0}, \cdot)$ is an Abelian monoid.
(4) If S is a totally ordered set and $a \bullet b = \min\{a, b\}$, then $(S, \bullet)$ is an Abelian semigroup. If $s = \sup S$ exists, then $(S, \bullet)$ is an Abelian monoid with identity element s.
(5) Given a set S, $(2^S, \cap)$ and $(2^S, \cup)$ are Abelian monoids. (Here 2^S denotes the set of subsets of S.)
(6) Every group has an underlying monoid which is Abelian whenever the monoid is. Additionally, every subset of a group which is closed under the group operation is a semigroup; if the subset contains the identity element, it is also a monoid.
(7) Given a set S, $(S^S, \circ)$ is a monoid called the *transformation monoid* of S. In any category, the endomorphisms of a given object form a monoid under composition.
(8) Given a (unital) ring $(R, +, \cdot)$, $(R, \cdot)$ is a monoid.
(9) The set of all finite strings over a fixed alphabet Σ with the concatenation operation is the *free monoid* over Σ. (Here we are including the empty string, which serves as the identity. If the empty string is excluded, this is the *free semigroup* over Σ.)
(10) Given a field k, write $M(\mathsf{k})$ for the set of isometry classes of regular quadratic forms over k. Then $(M(\mathsf{k}), \perp)$ is an Abelian monoid with identity 0 and $(M(\mathsf{k}), \otimes)$ is an Abelian monoid with identity $\langle 1 \rangle$.

Definition 11.3. Given monoids M and N, a *monoid homomorphism* $f\colon M \to N$ is a function between underlying sets satisfying $f(e) = e$ and $f(mn) = f(m)f(n)$ for all $m, n \in M$.

Notably, the condition $f(e) = e$ does not come along for free in this setting. The reader may check that in a cancellative monoid (see the Definition 11.9), this condition follows from preservation of the monoid operation.

Given a monoid M, we would like to form the "smallest" group admitting a homomorphism from M. Similar to our handling of the tensor product of vector spaces, the following proposition states the universal property of such an object, and the proof contains a construction/definition.

Proposition 11.4. Given a monoid M, there exists a group M^{gp} (called the *group completion* or *universal enveloping group* of M) and a homomorphism $M \to M^{\mathrm{gp}}$ such that if G is a group and $M \to G$ is a homomorphism, there is a unique homomorphism $M^{\mathrm{gp}} \to G$ such that the diagram

$$\begin{array}{ccc} M & \longrightarrow & G \\ \downarrow & \nearrow & \\ M^{\mathrm{gp}} & & \end{array}$$

commutes.

Proof Sketch. Let $\underline{M} = \{\underline{m} \mid m \in M\}$ and set

$$M^{\mathrm{gp}} = \langle \underline{M} \mid \underline{e} = e, \underline{mn} = \underline{m} \cdot \underline{n} \text{ for all } m, n \in M \rangle .$$

(This is standard notation for presenting a group via generators and relations.) The map $M \to M^{\mathrm{gp}}$ is given by $m \mapsto \underline{m}$. Given a group G and homomorphism $f\colon M \to G$, the function $\underline{m} \mapsto f(m)$

is well-defined and a homomorphism. Moreover, in order for the diagram to commute, the map $M^{\mathrm{gp}} \to G$ must take these values. □

Remark 11.5. The reader should compare this construction and its universal property with Abelianization, $G \mapsto G^{\mathrm{Ab}}$.

Remark 11.6. Given a monoid homomorphism $f\colon M \to N$, composition with $N \to N^{\mathrm{gp}}$ results in a homomorphism $M \to N^{\mathrm{gp}}$. By Proposition 11.4, we get a unique compatible homomorphism $f^{\mathrm{gp}}\colon M^{\mathrm{gp}} \to N^{\mathrm{gp}}$. The reader may check that $\mathrm{id}_M^{\mathrm{gp}} = \mathrm{id}_{M^{\mathrm{gp}}}$ and $(f \circ g)^{\mathrm{gp}} = f^{\mathrm{gp}} \circ g^{\mathrm{gp}}$, so $(\ )^{\mathrm{gp}}$ is a functor from monoids to groups. In fact, this functor is left adjoint to the "forgetful" functor that takes a group to its underlying monoid.

Word problems in groups make the group completion of an arbitrary monoid unwieldy to work with in general. In the case of Abelian monoids, we can be much more concrete.

Proposition 11.7. If M is an Abelian monoid, then the following statements are true:

(a) Every element of M^{gp} can be expressed as $\underline{m} - \underline{n}$ for some $m, n \in M$.
(b) If $m, n \in M$, then $\underline{m} = \underline{n} \in M^{\mathrm{gp}}$ if and only if $m + \ell = n + \ell$ for some $\ell \in M$.
(c) The monoid map $M \times M \to M^{\mathrm{gp}}$ taking $(m, n) \mapsto \underline{m} - \underline{n}$ is surjective.
(d) The set underlying the group completion M^{gp} is the set-theoretic quotient of M by the equivalence relation generated by $(m, n) \sim (m', n')$ whenever there exists $\ell \in M$ such that $m+n'+\ell = m' + n + \ell$.

Proof. First observe that since M is Abelian, we can rewrite

$$M^{\mathrm{gp}} = \mathbb{Z}\underline{M}/R$$

where $\mathbb{Z}\underline{M} = \mathbb{Z}^{\oplus \underline{M}}$ is the free Abelian group on $\underline{M}$ and R is the subgroup (necessarily normal) generated by $\underline{m+n} - \underline{m} - \underline{n}$. In a free Abelian group (such as $\mathbb{Z}\underline{M}$), every element is a difference of sums of (potentially repeated) generators. Since $\underline{m_1} + \cdots + \underline{m_k} = \underline{m_1 + \cdots + m_k}$ in M^{gp}, we can "group positive and negative terms" to achieve the description given in (a). Part (c) follows immediately.

For (b), suppose $\underline{m} - \underline{n} = 0$ in M^{gp}. This means that $\underline{m} - \underline{n}$ is an element of R, and hence

$$\underline{m} - \underline{n} = \left(\sum \underline{a_i + b_i} - \underline{a_i} - \underline{b_i}\right) - \left(\sum \underline{c_j + d_j} - \underline{c_j} - \underline{d_j}\right)$$

in $\mathbb{Z}\underline{M}$ for some $a_i, b_i, c_j, d_j \in M$. (Here we are using that elements of R can be written as a difference of a sum of generators.) By doing some arithmetic in $\mathbb{Z}\underline{M}$, we get that

$$\underline{m} + \left(\sum \underline{a_i} + \underline{b_i}\right) + \left(\sum \underline{c_j + d_j}\right) = \underline{n} + \left(\sum \underline{a_i + b_i}\right) + \left(\sum \underline{c_j} + \underline{d_j}\right).$$

In a free Abelian group, two sums of generators are equal if and only if they have the same number of terms, and those terms differ by a permutation. Hence the terms on the left and right of the above display differ only by a permutation. It follows that we may "remove the underlines" and translate the above identity into one that holds in M:

$$m + \left(\sum a_i + b_i\right) + \left(\sum c_j + d_j\right) = n + \left(\sum a_i + b_i\right) + \left(\sum c_j + d_j\right).$$

This proves (b), and (d) follows immediately from (a) and (b). □

Remark 11.8. Given the above proposition, some authors choose to define M^{gp} as the quotient of $M \times M$ described in part (d). In fact, you may have seen exactly this construction in Math 112 when building $(\mathbb{Z}, +)$ from $(\mathbb{N}, +)$. This is also closely related to the construction of fields of fractions of integral domains and, more generally, localizations of commutative rings at multiplicative subsets.

For special classes of Abelian monoids, group completion is even better behaved.

Definition 11.9. A monoid M is *cancellative* (or has the *cancellation property*) if for all $m, n, \ell \in M$, the equality $mn = m\ell$ implies that $n = \ell$.

Corollary 11.10. For an Abelian monoid M, the natural map $M \to M^{\mathrm{gp}}$ is injective if and only if M is cancellative.

Proof. First suppose that M is a cancellative Abelian monoid and suppose that $\underline{m} = \underline{n} \in M^{\mathrm{gp}}$. By Proposition 11.7(b), we know there is some $\ell \in M$ such that $m + \ell = n + \ell$. Since M is cancellative, $m = n$, whence $M \to M^{\mathrm{gp}}$ is injective.

Now suppose that M is Abelian and $M \to M^{\mathrm{gp}}$ is injective. Again by Proposition 11.7(b), we know that $\underline{m} = \underline{n} \in M^{\mathrm{gp}}$ if and only if $m + \ell = n + \ell$ for some $\ell \in M$. But injectivity of $M \to M^{\mathrm{gp}}$ says that $\underline{m} = \underline{n}$ if and only if $m = n$. Enchaining equivalences, we learn that $m + \ell = n + \ell$ if and only if $m = n$, which is the cancellation property. □

We now study the situation in which an Abelian monoid $(M, +)$ also supports a "multiplication" that distributes over $+$.

Definition 11.11. A (unital) *semiring* $(M, +, \cdot)$ consists of an Abelian monoid $(M, +)$ and a monoid $(M, \cdot)$ such that $m \cdot (n + \ell) = (m \cdot n) + (m \cdot \ell)$ and $(n + \ell) \cdot m = (n \cdot m) + (\ell \cdot m)$ for all $m, n, \ell \in M$.

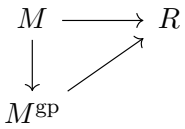
Remark 11.12. Some authors refer to semirings as *rigs* because they are rings without negatives. While exceptionally cute, this author finds it too difficult to distinguish rig from ring when reading quickly.

Example 11.13. (a) The natural numbers with standard addition and multiplication form a semiring.
(b) For a fixed group G, isomorphism classes of finite G-sets with disjoint union and Cartesian product form a semiring.
(c) The set of isometry classes $M(\mathsf{k})$ of regular quadratic forms over k with orthogonal sum and tensor product form a semiring.

Proposition 11.14. Let $(M, +, \cdot)$ be a semiring and let M^{gp} denote the group completion of $(M, +)$. Then $\cdot$ extends to M^{gp} and makes M^{gp} a ring.

Proof. Without loss of generality, we may replace M^{gp} with the quotient $(M \times M)/\sim$ of Proposition 11.7(d). Inspired by the formula $(m - n)(m' - n') = (mm' + nn') - (mn' + nm')$ which holds in rings, we define $(m, n) \cdot (m', n')$ to be $(mm' + nn', mn' + nm')$. It is now rote to check that this makes M^{gp} a ring. □

Remark 11.15. More is true. With the induced ring structure, $(\ )^{\mathrm{gp}}$ is a functor from semirings to rings which is left adjoint to the forgetful functor. In particular, if M is a semiring, R is a ring, and $f : M \to R$ respects addition, multiplication, and units, then there is a unique ring homomorphism $f^{\mathrm{gp}} : M^{\mathrm{gp}} \to R$ such that



commutes.

Example 11.16. (a) $(\mathbb{N}, +, \cdot)^{\mathrm{gp}} = (\mathbb{Z}, +, \cdot)$.
(b) For a group G, let $\mathrm{Iso}(G\,\mathrm{Fin})$ denote isomorphism classes of finite G-sets. Then $(\mathrm{Iso}(G\,\mathrm{Fin}), \amalg, \times)^{\mathrm{gp}}$ is called the *Burnside ring* of G, and is typically denoted $A(G)$ (or $B(G)$, or $\Omega(G)$).
(c) In the next section, we will thoroughly explore $\mathrm{GW}(\mathsf{k}) = (M(\mathsf{k}), \perp, \otimes)^{\mathrm{gp}}$, the *Grothendieck-Witt ring* of k.

12. The Witt and Grothendieck-Witt rings

Definition 12.1. The *Grothendieck-Witt ring* of k, denoted $\mathrm{GW}(\mathsf{k})$, is the group completion of $M(\mathsf{k})$, the semiring of isometry classes of regular quadratic forms over k with operations $\perp$ and $\otimes$. We write $+$ and $\cdot$ for $\perp$ and $\otimes$ in $\mathrm{GW}(\mathsf{k})$.

By Theorem 8.2, $M(\mathsf{k})$ is a cancellative monoid, and thus (by Corollary 11.10) the natural map $M(\mathsf{k}) \to \mathrm{GW}(\mathsf{k})$ is an injective homomorphism. In fact, we will make the usual abuse of notation and consider $M(\mathsf{k})$ to be a subset of $\mathrm{GW}(\mathsf{k})$. Additionally, rather than referring to the isometry class of q as $\underline{q}$ or $[q]$ and belaboring the distinction between $q \cong q'$ and $\underline{q} = \underline{q}'$, we will simply write $q = q'$ in $\mathrm{GW}(\mathsf{k})$, as long as it is clear from context that we are working in the Grothendieck-Witt ring.

By Proposition 11.7, we have that every element of $\mathrm{GW}(\mathsf{k})$ may be written as $q - q'$ for $q, q' \in M(\mathsf{k})$. If $q, q' \in M(\mathsf{k})$ and $q = q'$ in $\mathrm{GW}(\mathsf{k})$, then Corollary 11.10 implies that $q = q'$ in $M(\mathsf{k})$ as well, further justifying the abuses of the previous paragraph.

We have already seen that dimension gives a homomorphism $\dim\colon M(\mathsf{k}) \to \mathbb{Z}$. As such, we get a dimension homomorphism $\dim\colon \mathrm{GW}(\mathsf{k}) \to \mathbb{Z}$, and we may compute it on "formal differences" as $\dim(q - q') = \dim q - \dim q'$.

Definition 12.2. The *fundamental ideal* of $\mathrm{GW}(\mathsf{k})$ is

$$\mathrm{GI}(\mathsf{k}) = \ker(\dim\colon \mathrm{GW}(\mathsf{k}) \to \mathbb{Z}).$$

Since the dimension homomorphism is manifestly surjective, we get the isomorphism

$$\mathrm{GW}(\mathsf{k})/\mathrm{GI}(\mathsf{k}) \cong \mathbb{Z}$$

for all fields k.

Proposition 12.3. The fundamental ideal is additively generated (as a subgroup) by the expressions $\langle\lambda\rangle - \langle 1\rangle$, $\lambda \in \mathsf{k}^\times$.

Proof. Elements of the form $\langle\lambda\rangle - \langle 1\rangle$ are clearly in $\mathrm{GI}(\mathsf{k})$, giving us one inclusion. If $z \in \mathrm{GI}(\mathsf{k})$, then $z = q - q'$ for $q, q' \in M(\mathsf{k})$ of the same dimension. Diagonalizing, we may write $q = \langle\lambda_1, \ldots, \lambda_n\rangle$ and $q' = \langle\mu_1, \ldots, \mu_n\rangle$. Then

$$z = \sum(\langle\lambda_i\rangle - \langle\mu_i\rangle) = \left(\sum \langle\lambda_i\rangle - \langle 1\rangle\right) - \left(\sum \langle\mu_i\rangle - \langle 1\rangle\right).$$

This shows that z is in the subgroup generated by "virtual forms" of the form $\langle\lambda\rangle - \langle 1\rangle$, proving the opposite inclusion. □

Now consider the ideal $(h) = h\,\mathrm{GW}(\mathsf{k})$ generated by the hyperbolic form $h = \langle 1, -1\rangle$. Recall that Lemma 10.10 tells us that $hq = \dim(q)h$. As such,

$$(h) = \mathbb{Z}h,$$

the set of integer multiples of h. To remind us that this ideal has such a simple form, we will typically write it as $\mathbb{Z}h$ rather than (h).

Definition 12.4. The *Witt ring* of k is the quotient ring

$$\mathrm{W}(\mathsf{k}) = \mathrm{GW}(\mathsf{k})/\mathbb{Z}h.$$

Somewhat surprisingly, the Witt ring actually predates the Grothendieck-Witt ring, having been introduced (by another name) in Witt's 1937 paper. The following proposition exhibits how this was possible.

Proposition 12.5. (a) The elements of $\mathrm{W}(\mathsf{k})$ are in bijective correspondence with isometry classes of anisotropic forms over k.

(b) Two regular forms q and q' represent the same class in $\mathrm{W}(\mathsf{k})$ if and only if their anisotropic parts (in the sense of the Witt Decomposition Theorem 8.1) are isometric.
(c) If $\dim q = \dim q'$, then q and q' represent the same class in $\mathrm{W}(\mathsf{k})$ if and only if $q \cong q'$.

Proof. First note that (a) implies (b) and (c). Indeed, (b) is immediate given that $q_h + q_a \equiv q_a \pmod{\mathbb{Z}h}$. For (c), suppose $\dim q = \dim q'$ and $q = q'$ in $\mathrm{W}(\mathsf{k})$. By (b), $q_a \equiv q'_a$, and the dimension condition guarantees that $q_h \cong q'_h$ as well, whence $q \cong q'$. The other direction of (c) is trivial.

To prove (a), first note that $h = 0$ in $\mathrm{W}(\mathsf{k})$ implies that $\langle -1\rangle = -\langle 1\rangle$ in $\mathrm{W}(\mathsf{k})$ and, more generally, $\langle -\lambda\rangle = -\langle\lambda\rangle$ in $\mathrm{W}(\mathsf{k})$. It follows that every element of $\mathrm{W}(\mathsf{k})$ is represented by a (non-virtual) quadratic form. By Witt decomposition and the fact that hyperbolic spaces are trivial in $\mathrm{W}(\mathsf{k})$, we know that $q = q_a$ in $\mathrm{W}(\mathsf{k})$, where q_a is the anisotropic part of q. Thus every element of $\mathrm{W}(\mathsf{k})$ is represented by an anisotropic form. It remains to show that if q and q' are anisotropic and $q = q' \in \mathrm{W}(\mathsf{k})$, then $q \cong q'$. We have $q = q' + mh \in \mathrm{GW}(\mathsf{k})$ for some integer m. Swapping q and q' if necessary, we may assume that $m \geq 0$. Thus $q \cong q' \perp mh$. Since q is anisotropic, we may conclude that $m = 0$, so $q \cong q'$ as desired. □

Now consider the image of $\mathrm{GI}(\mathsf{k})$ under the quotient homomorphism $\mathrm{GW}(\mathsf{k}) \to \mathrm{W}(\mathsf{k})$. We denote this ideal $\mathrm{I}(\mathsf{k})$, and call it the *fundamental ideal* of $\mathrm{W}(\mathsf{k})$. We will interpret this ideal through the lens of the following commutative diagram

(ℵ)

$$\begin{array}{ccccccccc}
 & & & & 0 & & 0 & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & 0 & \longrightarrow & \mathbb{Z}h & \xrightarrow{\cong} & 2\mathbb{Z} & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathrm{GI}(\mathsf{k}) & \longrightarrow & \mathrm{GW}(\mathsf{k}) & \longrightarrow & \mathbb{Z} & \longrightarrow & 0 \\
 & & \downarrow\cong & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathrm{I}(\mathsf{k}) & \longrightarrow & \mathrm{W}(\mathsf{k}) & \xrightarrow[\dim_0]{} & \mathbb{Z}/2\mathbb{Z} & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

in which the rows and columns are exact.[11] There are two observations hiding here. First, when restricted to $\mathrm{GI}(\mathsf{k})$, the quotient homomorphism is an isomorphism onto $\mathrm{GI}(\mathsf{k}) \cong \mathrm{I}(\mathsf{k})$. This follows because $\mathrm{GI}(\mathsf{k}) \cap \mathbb{Z}h = 0$. Second, an anisotropic form q is in $\mathrm{I}(\mathsf{k})$ if and only if its dimension is 0; the map $\mathrm{W}(\mathsf{k}) \to \mathbb{Z}/2\mathbb{Z}$ takes $q + \mathbb{Z}h \mapsto \dim q + 2\mathbb{Z}$. We state and prove this second observation in the following proposition.

Proposition 12.6. A regular quadratic form q represents an element of $\mathrm{I}(\mathsf{k}) \subseteq \mathrm{W}(\mathsf{k})$ if and only if $\dim q$ is even.

Proof. We begin with the right-to-left implication. It suffices to consider a binary quadratic form $q = \langle\lambda, \mu\rangle$. Note that $\langle\lambda\rangle - \langle -\mu\rangle \mapsto \langle\lambda, \mu\rangle$ under the quotient map $\mathrm{GW}(\mathsf{k}) \to \mathrm{W}(\mathsf{k})$. By the definition of $\mathrm{I}(\mathsf{k})$, we conclude that $q \in \mathrm{I}(\mathsf{k})$.

For the left-to-right implication, suppose that q is regular and in $\mathrm{I}(\mathsf{k})$. By definition, $q = q_1 - q_2 + mh \in \mathrm{GW}(\mathsf{k})$ where $m \in \mathbb{Z}$ and $\dim q_1 = \dim q_2$. It follows that $\dim q = \dim q_1 - \dim q_2 + 2m = 2m$, as desired. □

[11] A sequence $A \xrightarrow{f} B \xrightarrow{g} C$ of homomorphisms is *exact* at B when $\ker g = \operatorname{im} f$. We are claiming that every composable pair of horizontal or vertical maps is exact at its shared term. Note that exactness of $0 \to A \to B \to C \to 0$ (a *short exact sequence*) implies that $C \cong B/A$, where A is identified with its isomorphic image in B.

13. The I-adic filtration of the Witt ring and the "dimterminant" homomorphism

Given a commutative ring R and an ideal $J \subseteq R$, we may consider the *J-adic filtration* of R,

$$R \supseteq J \supseteq J^2 \supseteq J^3 \supseteq J^4 \supseteq \cdots,$$

where J^k is the k-fold product of J with itself. Such a filtration has *filtration quotients*

$$R/J, \quad J/J^2, \quad J^2/J^3, \quad J^3/J^4, \quad \ldots.$$

Loosely speaking, one might try to understand properties of R by first understanding all the filtration quotients and then trying to glue these pieces together.

In the cases of the Grothendieck-Witt and Witt rings, we will concern ourselves with the $\mathrm{GI}(\mathsf{k})$-adic and $\mathrm{I}(\mathsf{k})$-adic filtrations. Since $\mathrm{GI}(\mathsf{k}) \cong \mathrm{I}(\mathsf{k})$, the associated filtration quotients will only differ at the first stage. This stage is easy for us to identify via (ℵ): $\mathrm{GW}(\mathsf{k})/\mathrm{GI}(\mathsf{k}) \cong \mathbb{Z}$, and $\mathrm{W}(\mathsf{k})/\mathrm{I}(\mathsf{k}) \cong \mathbb{Z}/2\mathbb{Z}$.

We now consider the second filtration quotient, $\mathrm{GI}(\mathsf{k})/\mathrm{GI}(\mathsf{k})^2 \cong \mathrm{I}(\mathsf{k})/\mathrm{I}(\mathsf{k})^2$. In order to determine this quotient's structure, we will achieve the more ambitious goal of identifying $\mathrm{W}(\mathsf{k})/\mathrm{I}(\mathsf{k})^2$ (again with k implicit).

Recall that there is a monoid homomorphism $\det\colon M(\mathsf{k}) \to \mathsf{k}^\times/\mathsf{k}^{\boxtimes}$ which takes the square class of the determinant of a symmetric matrix representing a regular quadratic form. This extends to a group homomorphism $\det\colon \mathrm{GW}(\mathsf{k}) \to \mathsf{k}^\times/\mathsf{k}^{\boxtimes}$ via the formula

$$\det(q - q') = \det(q)/\det(q')$$

for $q' \neq 0$. Since $\det h = -1 \cdot \mathsf{k}^{\boxtimes}$, and this class is nontrivial in fields lacking a square root of -1, the map $\det$ does not extend to $\mathrm{W}(\mathsf{k})$. We remedy this via a clever construction.

If q is a regular quadratic form with $\dim q = n$, define the *signed determinant* of q by

$$\det_{\pm} q = (-1)^{n(n-1)/2} \det q \in \mathsf{k}^\times/\mathsf{k}^{\boxtimes}.$$

This map satisfies $\det_\pm h = \mathsf{k}^{\boxtimes}$, but it is no longer a homomorphism: $\det_\pm(q \perp q') \neq \det_\pm(q) \cdot \det_\pm(q')$ in general. We need to change the codomain and add some information to $\det_\pm$ to get a homomorphism out of $\mathrm{W}(\mathsf{k})$.

Define $Q(\mathsf{k})$ to be the set $\mathbb{Z}/2\mathbb{Z} \times \mathsf{k}^\times/\mathsf{k}^{\boxtimes}$ and equip it with the binary operation

$$(m + 2\mathbb{Z}, \lambda\mathsf{k}^{\boxtimes}) \cdot (n + 2\mathbb{Z}, \mu\mathsf{k}^{\boxtimes}) = ((m+n) + 2\mathbb{Z}, (-1)^{mn}\lambda\mu\mathsf{k}^{\boxtimes}).$$

The reader may check that this is a commutative and associative product with identity element $(2\mathbb{Z}, \mathsf{k}^{\boxtimes})$. The inverse of $(n + 2\mathbb{Z}, \lambda\mathsf{k}^{\boxtimes})$ with respect to this operation is $(n + 2\mathbb{Z}, (-1)^n\lambda\mathsf{k}^{\boxtimes})$ since

$$(n + 2\mathbb{Z}, \lambda\mathsf{k}^{\boxtimes}) \cdot (n + 2\mathbb{Z}, (-1)^n\lambda\mathsf{k}^{\boxtimes}) = (2n + 2\mathbb{Z}, (-1)^{n^2}(-1)^n\lambda^2\mathsf{k}^{\boxtimes}) = (2\mathbb{Z}, \mathsf{k}^{\boxtimes}).$$

Thus $(Q(\mathsf{k}), \cdot)$ is a group. Furthermore, the map $\mathsf{k}^\times/\mathsf{k}^{\boxtimes} \to Q(\mathsf{k})$ given by $\lambda\mathsf{k}^{\boxtimes} \mapsto (2\mathbb{Z}, \lambda\mathsf{k}^{\boxtimes})$ is an injective homomorphism which identifies $\mathsf{k}^\times/\mathsf{k}^{\boxtimes}$ as an index two subgroup of $Q(\mathsf{k})$. In other words, we get a short exact sequence of Abelian groups

$$1 \to \mathsf{k}^\times/\mathsf{k}^{\boxtimes} \to Q(\mathsf{k}) \to \mathbb{Z}/2\mathbb{Z} \to 0$$

which is an extension of $\mathsf{k}^\times/\mathsf{k}^{\boxtimes}$ by $\mathbb{Z}/2\mathbb{Z}$. This extension is split (*i.e.*, $Q(\mathsf{k})$ is the product of $\mathsf{k}^\times/\mathsf{k}^{\boxtimes}$ and $\mathbb{Z}/2\mathbb{Z}$) if and only if $-1 \in \mathsf{k}^{\boxtimes}$.

Proposition 13.1. The assignment

$$\begin{aligned}(\dim_0, \det_\pm)\colon M(\mathsf{k}) &\longrightarrow Q(\mathsf{k}) \\ q &\longmapsto (\dim_0(q), \det_\pm q)\end{aligned}$$

(which we will glibly refer to as the *dimterminant*) is a surjective monoid homomorphism. Its extension to $\mathrm{GW}(\mathsf{k}) \to Q(\mathsf{k})$ factors through $\mathrm{W}(\mathsf{k})$ and induces an isomorphism $\mathrm{W}(\mathsf{k})/\mathrm{I}(\mathsf{k})^2 \cong Q(\mathsf{k})$.

Proof. We first check that the map is a monoid homomorphism. Suppose q and q' are (isometry classes of) regular quadratic forms with $\dim q = n$ and $\dim q' = n'$. Then

$$\begin{aligned}(\dim_0, \det_\pm)(q) \cdot (\dim_0, \det_\pm)(q') &= (n + 2\mathbb{Z}, (-1)^{n(n-1)/2} \det(q)) \cdot (n', (-1)^{n'(n'-1)/2} \det(q')) \\ &= ((n+n') + 2\mathbb{Z}, (-1)^{nn'}(-1)^{(n(n-1)+n'(n'-1))/2} \det(q)\det(q')) \\ &= ((n+n') + 2\mathbb{Z}, (-1)^{(n+n')(n+n'-1)/2} \det(q \perp q')) \\ &= (\dim_0, \det_\pm)(q \perp q'),\end{aligned}$$

as desired. Additionally, we have $(\dim_0, \det_\pm)(\langle\lambda\rangle) = (1 + 2\mathbb{Z}, \lambda\mathsf{k}^{\boxtimes})$ and $(\dim_0, \det_\pm)(\langle 1, -\lambda\rangle) = (2\mathbb{Z}, \lambda\mathsf{k}^{\boxtimes})$. It follows that $(\dim_0, \det_\pm)$ is surjective.

By the universal property of group completion, we get a unique surjective group homomorphism $(\dim_0, \det_\pm) : \mathrm{GW}(\mathsf{k}) \to Q(\mathsf{k})$. Moreover,

$$(\dim_0, \det_\pm)(h) = (2\mathbb{Z}, (-1)\det(h)) = (2\mathbb{Z}, \mathsf{k}^{\boxtimes}),$$

the identity element of $Q(\mathsf{k})$, so this map factors through $\mathrm{W}(\mathsf{k}) = \mathrm{GW}(\mathsf{k})/\mathbb{Z}h$.

The next step is to check that this map is trivial on $\mathrm{I}(\mathsf{k})^2$. We know that $\mathrm{I}(\mathsf{k})$ is additively generated by binary forms $\langle 1, \lambda\rangle$, so $\mathrm{I}(\mathsf{k})^2$ is additively generated by the four-dimensional forms $\langle 1, \lambda\rangle \otimes \langle 1, \mu\rangle = \langle 1, \lambda, \mu, \lambda\mu\rangle$. We have

$$(\dim_0, \det_\pm)(\langle 1, \lambda, \mu, \lambda\mu\rangle) = (2\mathbb{Z}, (-1)^4\lambda \cdot \mu \cdot \lambda\mu\mathsf{k}^{\boxtimes}) = (2\mathbb{Z}, \mathsf{k}^{\boxtimes}),$$

so the map factors as a surjective homomorphism $f\colon \mathrm{W}(\mathsf{k})/\mathrm{I}(\mathsf{k})^2 \to Q(\mathsf{k})$.

Finally, we construct an inverse $g\colon Q(\mathsf{k}) \to \mathrm{W}(\mathsf{k})/\mathrm{I}(\mathsf{k})^2$ in order to prove that f is an isomorphism. This map is given by

$$g(2\mathbb{Z}, \lambda\mathsf{k}^{\boxtimes}) = \langle 1, -\lambda\rangle + \mathrm{I}(\mathsf{k})^2, \qquad g(1 + 2\mathbb{Z}, \lambda) = \langle\lambda\rangle + \mathrm{I}(\mathsf{k})^2.$$

In a homework problem, you will check that g is a homomorphism. (All that is required is rote computation.) Clearly $f \circ g = \mathrm{id}_{Q(\mathsf{k})}$, so g splits the surjection f. Since $g(1 + 2\mathbb{Z}, \lambda\mathsf{k}^{\boxtimes}) = \langle\lambda\rangle + \mathrm{I}(\mathsf{k})^2$, we learn that g is surjective, whence f and g are inverse isomorphisms. □

We immediately get the following corollary.

Corollary 13.2. If $\mathrm{I}(\mathsf{k})^2 = 0$, then $\mathrm{W}(\mathsf{k}) \cong Q(\mathsf{k})$ as an Abelian group.

Since $f\colon \mathrm{W}(\mathsf{k})/\mathrm{I}(\mathsf{k})^2 \to Q(\mathsf{k})$ is injective, we also have the following result.

Corollary 13.3. The ideal $\mathrm{I}(\mathsf{k})^2$ consists of classes of even-dimensional forms q for which $\det(q) = (-1)^{n(n-1)/2}\mathsf{k}^{\boxtimes}$, where $n = \dim q$.

Finally, if we restrict f to $\mathrm{I}(\mathsf{k})/\mathrm{I}(\mathsf{k})^2$, then the "$\dim_0$-coordinate" always takes the value $2\mathbb{Z}$. This implies the following corollary.

Corollary 13.4. The restriction of $f\colon \mathrm{W}(\mathsf{k})/\mathrm{I}(\mathsf{k})^2 \to Q(\mathsf{k})$ to $\mathrm{I}(\mathsf{k})/\mathrm{I}(\mathsf{k})^2$ induces an isomorphism

$$\mathrm{I}(\mathsf{k})/\mathrm{I}(\mathsf{k})^2 \cong \mathsf{k}^\times/\mathsf{k}^{\boxtimes}.$$

This is our desired determination of the second filtration quotient of the $\mathrm{I}(\mathsf{k})$-adic filtration of $\mathrm{W}(\mathsf{k})$. Later, we will return to this filtration and discuss the *Milnor conjecture* (proved by Voevodsky), which asserts an isomorphism between $\mathrm{I}(\mathsf{k})^n/\mathrm{I}(\mathsf{k})^{n+1}$ and $k_n^M(\mathsf{k})$, the n-th mod 2 Milnor K-theory group of k.

14. First computations of Witt and Grothendieck–Witt rings

For certain fields we can completely determine the structure of $\mathrm{GW}(\mathsf{k})$ and $\mathrm{W}(\mathsf{k})$. Here we expose several of these examples, focusing on quadratically closed fields, the real numbers, and finite fields.

Definition 14.1. A field k is *quadratically closed* if every element of k is a square, *i.e.*, $\mathsf{k}^{\boxtimes} = \mathsf{k}^{\times}$.

Proposition 14.2. A field k is quadratically closed if and only if $\dim\colon \mathrm{GW}(\mathsf{k}) \to \mathbb{Z}$ is an isomorphism. In this case, $\dim_0\colon \mathrm{W}(\mathsf{k}) \cong \mathbb{Z}/2\mathbb{Z}$.

Proof. If k is quadratically closed, then $\langle\lambda\rangle \cong \langle 1\rangle$ for all $\lambda \in \mathsf{k}^{\times} = \mathsf{k}^{\boxtimes}$, and thus $q \cong (\dim q)\langle 1\rangle$ for all regular quadratic forms q. This implies that $\dim$ is an isomorphism.

Conversely, if $\dim$ is an isomorphism, then $\langle\lambda\rangle \cong \langle 1\rangle$ for all $\lambda \in \mathsf{k}^{\times}$, which is only possible if $\mathsf{k}^{\times} = \mathsf{k}^{\boxtimes}$. □

To handle $\mathsf{k} = \mathbb{R}$, we recall some facts you proved in your homework.

Lemma 14.3. If $\langle\lambda_1, \dots, \lambda_n\rangle \cong \langle\mu_1, \dots, \mu_n\rangle$ over $\mathbb{R}$, then there are the same number of positive λ_i's as there are positive μ_i's.

Proof. Since $\mathbb{R}^{\times}/\mathbb{R}^{\boxtimes} = \{\pm 1\}\mathbb{R}_{>0}$, both forms may be rewritten as $r\langle 1\rangle \perp (n-r)\langle -1\rangle$ and $s\langle 1\rangle \perp (n-s)\langle -1\rangle$. The Witt Decomposition Theorem 8.1 and Cacellation Theorem 8.2 allows us to group together and cancel hyperbolic terms. The remaining equivalent anisotropic forms must consist of the same number of all 1's or all -1's. This allows us to conclude that $r = s$. □

The above lemma guarantees that the following definition is well-posed.

Definition 14.4. The *signature* of a real quadratic form q is

$$\mathrm{sgn}(q) = n_+ - n_-$$

where n_+ is the number of positive terms and n_- is the number of negative terms in any diagonalization of q.

Note that $\mathrm{sgn}\colon \mathrm{GW}(\mathbb{R}) \to \mathbb{Z}$ is in fact a ring homomorphism, and since $\mathrm{sgn}\langle 1, -1\rangle = 0$, it descends to a homomomorphism (with the same name) $\mathrm{sgn}\colon \mathrm{W}(\mathbb{R}) \to \mathbb{Z}$.

Proposition 14.5. The signature homomorphism $\mathrm{sgn}\colon \mathrm{W}(\mathbb{R}) \to \mathbb{Z}$ is an isomorphism.

Proof. By Proposition 12.5, the elements of $\mathrm{W}(\mathbb{R})$ are in bijective correspondence with isometry classes of anisotropic forms, which are all of the form $n\langle 1\rangle$ or $n\langle -1\rangle$ over $\mathbb{R}$. It immediately follows that sgn is an isomorphism. □

We can make a similarly nice statement for the Grothendieck–Witt ring.

Proposition 14.6. Over $\mathsf{k} = \mathbb{R}$,

(a) the isometry class of every regular quadratic form is determined by its dimension and signature, and
(b) $\mathrm{GW}(\mathsf{k}) \cong \mathbb{Z}[h]/(h^2 - 2h)$.

Proof. Part (a) is a direct consequence of Lemma 14.3. For part (b), note that (a) implies that Lemma 14.3 implies that every element of $\mathrm{GW}(\mathbb{R})$ takes the form $m + nh$ for unique integers m, n. We define a function $f\colon \mathrm{GW}(\mathsf{k}) \to \mathbb{Z}[h]/(h^2 - 2h)$ by the formula $f(m+nh) = m+nh$, which is a bijection since the codomain has additive basis $\{1, h\}$. By Lemma 10.10, we have $h^2 = 2h \in \mathrm{GW}(\mathsf{k})$, and this makes it easy to check that f is a homomorphism, and hence an isomorphism. □

Remark 14.7. Alternately, we may express $\mathrm{GW}(\mathbb{R})$ as the integral group ring on C_2, $\mathbb{Z}[C_2]$, by considering $\{\langle 1\rangle, \langle -1\rangle\}$ as a free $\mathbb{Z}$-basis of $\mathrm{GW}(\mathbb{R})$. Here, for a group G, $\mathbb{Z}[G] = \bigoplus_{g\in G} \mathbb{Z}\{g\}$ with multiplication extended bilinearly from $(1g)\cdot(1h) = 1(gh)$ (with the final product gh occuring in G). We leave the details of the isomorphisms $\mathrm{GW}(\mathbb{R}) \cong \mathbb{Z}[C_2] \cong \mathbb{Z}[x]/(x^2-1)$ to the reader.

It is also interesting to note that the above results hold for all *Euclidean* fields k, *i.e.*, those which are Pythagorean with $[\mathsf{k}^\times : \mathsf{k}^{\boxtimes}] = 2$.

We now consider finite fields $\mathsf{k} = \mathbb{F}_q$, where $q = p^n$, $p > 2$ prime.[12] Since $\mathsf{k}^\times$ is cyclic of order $q-1$, we have $\mathsf{k}^\times/\mathsf{k}^{\boxtimes} = \{\mathsf{k}^{\boxtimes}, s\mathsf{k}^{\boxtimes}\}$ for some nonsquare $s \in \mathsf{k}^\times \smallsetminus \mathsf{k}^{\boxtimes}$.

Lemma 14.8. For $\mathsf{k} = \mathbb{F}_q$, any class $s \in \mathsf{k}^\times \smallsetminus \mathsf{k}^{\boxtimes}$ is a sum of two squares in $\mathsf{k}^\times$.

Proof. First suppose that $-1 \in \mathsf{k}^{\boxtimes}$. Then $\langle 1,1\rangle \cong \langle 1,-1\rangle = h$ is universal, so s is a sum of two squares.

If $-1 \notin \mathsf{k}^{\boxtimes}$, consider the sets $\mathsf{k}^{\boxtimes}$ and $1+\mathsf{k}^{\boxtimes}$ which are subsets of k of the same cardinality, $(q-1)/2$. These sets are not equal since $1 \in \mathsf{k}^{\boxtimes}$ but $1 \notin 1+\mathsf{k}^{\boxtimes}$. It follows that some $1+\lambda^2$ is not in $\mathsf{k}^{\boxtimes}$. Since $1+\lambda^2 \neq 0$ (lest $-1 \in \mathsf{k}^{\boxtimes}$), we know that $1+\lambda^2 \in \mathsf{k}^\times \smallsetminus \mathsf{k}^{\boxtimes}$. It follows that

$$s\mathsf{k}^{\boxtimes} = (1+\lambda^2)\mathsf{k}^{\boxtimes},$$

whence s is also a sum of two squares. □

Lemma 14.8 implies that every regular binary form over $\mathbb{F}_q$ is universal.

Proposition 14.9. Over $\mathsf{k} = \mathbb{F}_q$, every regular binary form is universal.

Proof. Since 1 and s are the only square classes, there at most three nonequivalent regular binary forms,

$$f_1 = \langle 1,1\rangle, \quad f_2 = \langle 1,s\rangle, \quad \text{and} \quad f_3 = \langle s,s\rangle.$$

Lemma 14.8 implies that each of these is universal. □

Theorem 14.10. *Let* k *be a field in which every regular binary form is universal (such as* $\mathbb{F}_q$*). Then*

(a) two regular quadratic forms are isometric if and only if they have the same dimension and the same determinant,
(b) $\mathrm{GI}(\mathsf{k})^2 \cong \mathrm{I}(\mathsf{k})^2 = 0$ *and* $\mathrm{GI}(\mathsf{k}) \cong \mathrm{I}(\mathsf{k}) \cong \mathsf{k}^\times/\mathsf{k}^{\boxtimes}$*, and*
(c) $\mathrm{W}(\mathsf{k}) \cong Q(\mathsf{k})$*, and* $\mathrm{GW}(\mathsf{k}) \cong \mathbb{Z} \oplus \mathrm{GI}(\mathsf{k})$ *with trivial multiplication on* $\mathrm{GI}(\mathsf{k})$*.*

Proof. First note that, since any regular binary form $\langle \lambda_1, \lambda_2\rangle$ represents 1, we have $\langle \lambda_1,\lambda_2\rangle \cong \langle 1, \lambda_1\lambda_2\rangle$. It follows by induction that

$$q = \langle \lambda_1, \ldots, \lambda_n\rangle \cong \langle 1, \ldots, 1, \det(q)\rangle,$$

which proves (a).

By Proposition 12.3, $\mathrm{GI}(\mathsf{k})^2$ is additively generated by

$$(\langle\lambda_1\rangle - \langle 1\rangle)(\langle\lambda_2\rangle - \langle 1\rangle) = \langle\lambda_1\lambda_2\rangle + \langle 1\rangle - \langle\lambda_1\rangle - \langle\lambda_2\rangle = 0,$$

so $\mathrm{GI}(\mathsf{k})^2 = 0$. The rest of (b) follows from Corollary 13.4.

The first part of (c) has already been stated in Corollary 13.2. For the second part of (c), note that

$$0 \to \mathrm{GI}(\mathsf{k}) \to \mathrm{GW}(\mathsf{k}) \xrightarrow{\dim} \mathbb{Z} \to 0$$

[12] The reader may safely assume $q = p$ in the following discussion if they have not studied general finite fields previously.

is a *split short exact sequence*.[13] The ring structure follows since $\mathrm{GI}(\mathsf{k})^2 = 0$. □

Corollary 14.11. Let $\mathsf{k} = \mathbb{F}_q$. If $q \equiv 1 \pmod 4$, then $\mathrm{W}(\mathsf{k})$ is isomorphic to the group ring $\mathbb{F}_2[\mathsf{k}^\times/\mathsf{k}^{\boxtimes}]$; if $q \equiv 3 \pmod 4$, then $\mathrm{W}(\mathsf{k}) \cong \mathbb{Z}/4\mathbb{Z}$. In both cases, $\mathrm{GW}(\mathsf{k}) \cong \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$ with trivial multiplication on the second summand.

Proof. As a group, we know that $\mathrm{W}(\mathsf{k}) \cong Q(\mathsf{k})$ and that $Q(\mathsf{k})$ sits in a short exact sequence,

$$0 \to \mathsf{k}^\times/\mathsf{k}^{\boxtimes} \to Q(\mathsf{k}) \to \mathbb{Z}/2\mathbb{Z} \to 0.$$

When $q \equiv 1 \pmod 4$, this sequence is split by $s(n + 2\mathbb{Z}) = n\langle 1\rangle$, but when $q \equiv 3 \pmod 4$ the sequence does not split (essentially because $h \not\cong 2\langle 1\rangle$ in this case). We leave the remaining details to the reader. □

Remark 14.12. Later, we will see that if $\mathsf{k} = F(t)$ with F algebraically closed, then every regular binary form over k is universal, at which point Theorem 14.10 will once again provide a computation of $\mathrm{W}(\mathsf{k})$ and $\mathrm{GW}(\mathsf{k})$. In this case, an $\mathbb{F}_2$-basis of $\mathsf{k}^\times/\mathsf{k}^{\boxtimes}$ is given by $\{(t-\lambda)\mathsf{k}^{\boxtimes} \mid \lambda \in F\}$, so the Witt and Grothendieck–Witt rings have the same cardinality as F, but we still have precise control over their structure.

15. Presentations of the Witt and Grothendieck–Witt rings

Our present aim is to present the Grothendieck–Witt and Witt rings (of a given field k) in terms of generators and relations in the category of commutative rings. A *presentation* of a commutative ring A consists of a set of generators S and relations $R \subseteq \mathbb{Z}[S]$ such that $A \cong \mathbb{Z}[S]/(R)$. Here $\mathbb{Z}[S]$ is the free commutative ring on S, and (R) is the ideal in $\mathbb{Z}[S]$ generated by R. The following proposition specifies the universal property of $\mathbb{Z}[S]$, and a construction of $\mathbb{Z}[S]$ appears in the proof. A formal definition of a presentation then follows.

Proposition 15.1. Given a set S, there exists a commutative ring $\mathbb{Z}[S]$ and function $S \to \mathbb{Z}[S]$ such that for any commutative ring A and function $f\colon S \to A$, there exists a unique ring homomorphism $\tilde{f}\colon \mathbb{Z}[S] \to A$ such that

$$\begin{array}{ccc} S & \xrightarrow{f} & A \\ \downarrow & \nearrow_{\tilde{f}} & \\ \mathbb{Z}[S] & & \end{array}$$

commutes.

Proof. As the notation suggests, we set $\mathbb{Z}[S]$ to be the polynomial ring over $\mathbb{Z}$ with variables the elements of S. The function $S \to \mathbb{Z}[S]$ takes $s \in S$ to the variable of the same name. Given a function $f\colon S \to A$ to a commutative ring A, we define $\tilde{f}$ to be the function which evaluates a polynomial in S at the values given by f. The reader may check that this defines a ring homomorphism, and it is the only such map that will make the diagram commute. □

Definition 15.2. A *presentation* of a commutative ring A consists of a set S, a set $R \subseteq \mathbb{Z}[S]$, and a function $f\colon S \to A$ such that the homomorphism $\tilde{f}\colon \mathbb{Z}[S] \to A$ induces an isomorphism $\mathbb{Z}[S]/(R) \cong A$. In this case, we write $A \cong \langle S \mid R\rangle$.

Theorem 15.3. *Let* $[\mathsf{k}^\times] = \{[\lambda] \mid \lambda \in \mathsf{k}^\times\}$ *and let* R *be the set consisting of expressions of one of the following three forms:*

[13] We have already discussed what it means to be a short exact sequence. The word *split* means that there is a homomorphism $s\colon \mathbb{Z} \to \mathrm{GW}(\mathsf{k})$ such that $\dim \circ s = \mathrm{id}_{\mathbb{Z}}$. Indeed, we may take $s(n) = n\langle 1\rangle$. When a short exact sequence is split, the middle term is automatically the direct sum of the first and last terms. (*Warning*: Many short exact sequences are not split, *e.g.*, $0 \to \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/4\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0$.)

(a) $[1] - 1$,
(b) $[\lambda\mu] - [\lambda][\mu]$ *for* $\lambda, \mu \in \mathsf{k}^\times$,
(c) $[\lambda] + [\mu] - [\lambda + \mu](1 + [\lambda\mu])$ *for* $\lambda, \mu, \lambda + \mu \in \mathsf{k}^\times$.

Then

$$\mathrm{GW}(\mathsf{k}) \cong \left\langle [\mathsf{k}^\times] \mid R \right\rangle.$$

Before proving this result, we establish the following prepatory lemma.

Lemma 15.4. For every $\lambda \in \mathsf{k}^\times$, $[\lambda^2] = 1$ in $\langle [\mathsf{k}^\times] \mid R \rangle$.

Proof. By relation (c),

$$[\lambda] + [\lambda] = [2\lambda](1 + [\lambda^2])$$

in $\langle [\mathsf{k}^\times] \mid R \rangle$. We may also perform the following computation of $[\lambda] + [\lambda]$ in $\langle [\mathsf{k}^\times] \mid R \rangle$:

$$\begin{aligned} [\lambda] + [\lambda] &= [\lambda]([1] + [1]) && \text{(by (b) and the distributive law)} \\ &= [\lambda][2](1 + 1) && \text{(by (c) and (a))} \\ &= [2\lambda](1 + 1) && \text{(by (b)).} \end{aligned}$$

It follows that $[2\lambda](1 + [\lambda^2]) = [2\lambda](1 + 1)$. By (b), we know that $[2\lambda]$ is a unit in $\langle [\mathsf{k}^\times] \mid R \rangle$, so $1 + [\lambda^2] = 1 + 1$. Subtracting 1 from both sides gives $[\lambda^2] = 1$, as desired. □

Proof of Theorem 15.3. We take $f \colon [\lambda] \mapsto \langle \lambda \rangle$ to be the map $S \to \mathrm{GW}(\mathsf{k})$, which induces a homomorphism $\mathbb{Z}[[\mathsf{k}^\times]] \to \mathrm{GW}(\mathsf{k})$ by Proposition 15.1. Since $\mathrm{GW}(\mathsf{k})$ is additively generated by $\{\langle \lambda \rangle \mid \lambda \in \mathsf{k}^\times\}$, we know that $\tilde{f}$ is surjective.

We know that $(R) \subseteq \ker \tilde{f}$ because $\langle 1 \rangle = 1$, $\langle \lambda \rangle \langle \mu \rangle = \langle \lambda\mu \rangle$, and $\langle \lambda \rangle + \langle \mu \rangle = \langle \lambda + \mu \rangle (1 + \langle \lambda\mu \rangle)$.[14] Thus $\tilde{f}$ factors through $\langle [\mathsf{k}^\times] \mid R \rangle$ to give a surjective ring homomorphism $\tilde{\tilde{f}} \colon \langle [\mathsf{k}^\times] \mid R \rangle \to \mathrm{GW}(\mathsf{k})$.

We now produce an inverse to $\tilde{\tilde{f}}$, $\mathrm{GW}(\mathsf{k}) \to \langle [\mathsf{k}^\times] \mid R \rangle$.[15] By the univeral property of group completion, we can construct a monoid homomorphism $\varphi \colon M(\mathsf{k}) \to \langle [\mathsf{k}^\times] \mid R \rangle$. Given a quadratic form q equivalent to $\langle \lambda_1, \ldots, \lambda_n \rangle$, we would like to define $\varphi(q) = [\lambda_1] + \cdots + [\lambda_n]$. We must show that this assignment is well-defined. Once we have done so, it follows immediately that $\tilde{\varphi} \colon \mathrm{GW}(\mathsf{k}) \to \langle [\mathsf{k}^\times] \mid R \rangle$ is our desired inverse.

To check well-definition of φ amounts to showing that $\langle \mu_1, \ldots, \mu_n \rangle$ is another diagonalization of q, then $\sum [\lambda_i] = \sum [\mu_i]$ in $\langle [\mathsf{k}^\times] \mid R \rangle$. By Theorem 9.3, we may assume that $\langle \lambda_1, \ldots, \lambda_n \rangle$ is simply equivalent to $\langle \mu_1, \ldots, \mu_n \rangle$. Without loss of generality, further assume that $\lambda_i = \mu_i$ for $i \geq 3$ and that $\langle \lambda_1, \lambda_2 \rangle \cong \langle \mu_1, \mu_2 \rangle$. We must show that $[\lambda_1] + [\lambda_2] = [\mu_1] + [\mu_2]$ in $\langle [\mathsf{k}^\times] \mid R \rangle$ whenever $\langle \lambda_1, \lambda_2 \rangle \cong \langle \mu_1, \mu_2 \rangle$.

Since $\langle \lambda_1, \lambda_2 \rangle \cong \langle \mu_1, \mu_2 \rangle$ there exist $x, y \in \mathsf{k}$ and $z \in \mathsf{k}^\times$ such that $\mu_1 = \lambda_1 x^2 + \lambda_2 y^2$ and $\lambda_1 \lambda_2 = \mu_1 \mu_2 z^2$. First suppose that $x = 0$ or $y = 0$, in which case we may assume that $x = 0$ without loss of generality. Then $\mu_1 = \lambda_2 y^2$, so $[\mu_1] = [\lambda_2 y^2] = [\lambda_2]$ in $\langle [\mathsf{k}^\times] \mid R \rangle$ by (b) and Lemma 15.4. We also have

$$[\lambda_1] = \left[\mu_2 \cdot \frac{\mu_1}{\lambda_2} \cdot z^2 \right] = [\mu_2 (yz)^2] = [\mu_2] \in \left\langle [\mathsf{k}^\times] \mid R \right\rangle.$$

Thus $[\lambda_1] + [\lambda_2] = [\mu_1] + [\mu_2]$, as desired.

[14] The final equality follows easily from Proposition 9.1, but you also checked it by producing an explicit equivalence between $\langle \lambda, \mu \rangle$ and $\langle \lambda + \mu, (\lambda + \mu)\lambda\mu \rangle$ on a homework.

[15] It suffices to exhibit an inverse function. We will construct the map as a group homomorphism, and the reader may check that it is actually a ring isomorphism.

Now suppose that $x, y \neq 0$. Then in $\langle [\mathsf{k}^\times] \mid R \rangle$,

$$\begin{aligned}
[\lambda_1] + [\lambda_2] &= [\lambda_1 x^2] + [\lambda_2 y^2] && \text{(by Lemma 15.4)} \\
&= [\lambda_1 x^2 + \lambda_2 y^2](1 + [\lambda_1 \lambda_2 (xy)^2]) && \text{(by (c))} \\
&= [\mu_1](1 + [\mu_1 \mu_2]) \\
&= [\mu_1] + [\mu_2] && \text{(by (c)).}
\end{aligned}$$

We conclude that $\varphi \colon M(\mathsf{k}) \to \langle [\mathsf{k}^\times] \mid R \rangle$ is well-defined, and it is clearly a monoid homomorphism. As indicated previously, the induced map $\tilde{\varphi} \colon \mathrm{GW}(\mathsf{k}) \to \langle [\mathsf{k}^\times] \mid R \rangle$ is then clearly an inverse to $\tilde{\tilde{f}}$, concluding our proof. □

Corollary 15.5. Let $WR \subseteq \mathbb{Z}[[\mathsf{k}^\times]]$ be the set consisting of expressions of one of the following three forms:

(a) $[1] - 1$,
(b) $[\lambda\mu] - [\lambda][\mu]$ for $\lambda, \mu \in \mathsf{k}^\times$,
(c) $[\lambda] + [\mu] - [\lambda + \mu](1 + [\lambda\mu])$ for $\lambda, \mu, \lambda + \mu \in \mathsf{k}^\times$,
(d) $[1] + [-1]$.

Then

$$\mathrm{W}(\mathsf{k}) \cong \left\langle [\mathsf{k}^\times] \mid WR \right\rangle.$$

Proof. This is immediate since we have added in the relation (d) which mandates $h = 0$. □

16. Orderings and signatures

We have already seen that the signature of a quadratic form is a powerful invariant when working over the field of real numbers. It turns out that there is a signature invariant associated with every ordering of a field. (The signature of Section 14 is the signature relative to the unique ordering of $\mathbb{R}$.) Presently, we will develop the theory of field orderings to the extent necessary to understand signatures in generality. We begin with a seemingly unrelated definition.

Definition 16.1. Given a field k, let $\sigma(\mathsf{k})$ denote the set of sums of squares in k, and let $\sigma^\times(\mathsf{k}) = \sigma(\mathsf{k}) \smallsetminus \{0\}$. We call k *nonreal* if $-1 \in \sigma(\mathsf{k})$. If $-1 \notin \sigma(\mathsf{k})$, we call k *formally real*.

An ordering on a field is a fairly different animal.

Definition 16.2. An *ordering* of a field k is a proper subset $P \subsetneq \mathsf{k}^\times$ which is closed under addition and multiplication, and satisfies $P \cup (-P) = \mathsf{k}^\times$.

The elements of P are called *positive* (relative to P). When $\lambda - \mu \in P$, we write $\mu < \lambda$ or $\mu <_P \lambda$. In your homework, you will check that $<_P$ is a total (linear) ordering of k, and also that every total order of k satisfying $a < b \implies a + c < b + c$ and $0 < a, 0 < b \implies 0 < ab$ arises in this way.

Proposition 16.3. If P is an ordering on k, then $\sigma^\times(\mathsf{k}) \subseteq P$. In particular, k has characteristic 0.

Proof. If $\lambda \in \mathsf{k}^\times$, then λ or $-\lambda$ is in P (since $P \cup (-P) = \mathsf{k}^\times$). Either way, $(\pm\lambda)^2 = \lambda^2$ is in P since P is closed under multiplication. Since P is also closed under addition, $\sigma^\times(\mathsf{k}) \subseteq P$.

To deduce that k has characteristic 0, note that $1 = 1^2 \in P$, hence every sum $1 + 1 + \cdots + 1 \in P$. Since $P \subseteq \mathsf{k}^\times$, no such sum is 0. □

The following theorem links ordered for formally real fields.

Theorem 16.4 (Artin–Schreier)**.** *A field has an ordering if and only if it is formally real.*

Proof. If a field k has an ordering P, then $1 \in P$, hence $-1 \notin P$ and, in particular, $-1 \notin \sigma^\times(\mathsf{k})$, *i.e.*, k is formally real.

Now suppose that k is formally real. Define a *quasi-ordering* of k to be a subset of $\mathsf{k}^\times$ which is closed under addition and multiplication. Let P_0 be the set of sums of nonzero squares in k. (Careful: *a priori*, P_0 might not equal $\sigma^\times(\mathsf{k})$.) It is easy to check that P_0 is closed under addition and multiplication. Furthermore, if $0 \in P_0$, then

$$0 = \lambda_1^2 + \cdots + \lambda_n^2$$

with $\lambda_1 \neq 0$, so

$$-1 = (\lambda_2/\lambda_1)^2 + \cdots + (\lambda_n/\lambda_1)^2.$$

This contradicts the formally real hypothesis on k, so $P_0 = \sigma^\times(\mathsf{k})$ is a quasi-ordering of k.

By Zorn's lemma, P_0 is contained in a quasi-ordering P of k which is maximal with respect to inclusion.[16] We claim that for $\lambda \in \mathsf{k}^\times$, either $\lambda \in P$ or $-\lambda \in P$ (and hence P is an ordering). Define

$$Q = P \cup \lambda P \cup (P + \lambda P),$$

which is closed under addition and multiplication. If Q contains 0, then $0 = \pi' + \lambda\pi$ with $\pi, \pi' \in P$. Thus

$$-\lambda = \pi'/\pi = \pi'\pi(\pi^{-1})^2 \in P$$

since P is closed under multiplication and $k^{\boxtimes} \subseteq \sigma^\times(\mathsf{k}) \subseteq P$. On the other hand, if Q does not contain 0, then Q is a quasi-ordering of k. But $Q \supseteq P$ and P is maximal, so $Q = P$, and it follows that $\lambda \in P$. We have thus shown that P is an ordering, which completes the proof. □

For future reference, we highlight a lemma proved *en passant* in the above argument.

Lemma 16.5. A field k is nonreal if and only if 0 is a sum of nonzero squares in k.

Proof. The second paragraph of the proof of Theorem 16.4 gives the reverse implication. If k is nonreal, then $-1 = \sum \lambda_i^2$ implies that $0 = 1^2 + \sum \lambda_i^2$. □

Definition 16.6. The set X_k of all orderings of k is called the *Harrison space* of k.[17] The set $\bigcap_{P \in X_\mathsf{k}} P \subseteq \mathsf{k}^\times$ is called the set of *totally positive* elements of $\mathsf{k}^\times$.

Remark 16.7. Note that when $X_\mathsf{k} = \varnothing$ (*i.e.*, k is nonreal), $\bigcap_{P \in \varnothing} P = \mathsf{k}^\times$.

Lemma 16.8. Let k be formally real and set $F = \mathsf{k}(\sqrt{\lambda})$ be a quadratic extension of k, $\lambda \in \mathsf{k}^\times \smallsetminus \mathsf{k}^{\boxtimes}$. Then F is nonreal if and only if $-\lambda \in \sigma^\times(\mathsf{k})$.

Proof. If $-\lambda \in \sigma^\times(\mathsf{k})$, then $(\sqrt{\lambda})^2 + (-\lambda) = 0$ implies that F is nonreal by Lemma 16.5. Conversely, if F is nonreal, then we may write

$$-1 = \sum (a_i + b_i\sqrt{\lambda})^2$$

for some $a_i, b_i \in \mathsf{k}$. Since $-1 = -1 + 0 \cdot \sqrt{\lambda}$, this tells us that

$$-1 = \sum a_i^2 + \lambda \sum b_i^2.$$

[16] Recall that Zorn's lemma says that in a partially ordered set S in which every chain has an upper bound in S, the set S contains at least one maximal element. Here a *chain* is a totally ordered subset of S. In the above paragraph's argument, we are taking S to be the set of quasi-orderings on k with partial order given by the subset relation. Check that chains of quasi-orderings have upper bounds.

[17] For the topologically inclined: There is a topology on X_k generated by the sets $U_\lambda = \{P \in X_\mathsf{k} \mid \lambda \in P\}$, $\lambda \in \mathsf{k}^\times$. This makes X_k compact and totally disconnected.

Since k is formally real, we know that $\sum b_i^2 \neq 0$. It follows that

$$-\lambda = \left(1 + \sum a_i^2\right)\left(\sum b_i^2\right)^{-1}.$$

Since $\sigma^\times(\mathsf{k})$ is a group, we conclude that $-\lambda \in \sigma^\times(\mathsf{k})$, as desired. □

Theorem 16.9 (Artin). *An element $\lambda \in \mathsf{k}^\times$ is totally positive if and only if it is a nonzero sum of squares,* i.e.,

$$\bigcap_{P \in X_\mathsf{k}} P = \sigma^\times(\mathsf{k}).$$

Proof. If k is nonreal, this statement is a tautology. Suppose that k is formally real. We have already seen in Proposition 16.3 that $\sigma^\times(\mathsf{k}) \subseteq \bigcap_{P \in X_\mathsf{k}} P$. Suppose $\lambda \in k^\times \smallsetminus \sigma^\times(\mathsf{k})$. By Lemma 16.8, $F = \mathsf{k}(\sqrt{-\lambda})$ is formally real. If this is the case, we may pick an ordering P on F, and the reader may check that $P \cap \mathsf{k}^\times$ is an ordering on k. Since $\lambda = -(\sqrt{-\lambda})^2$, we know that $\lambda \in -P \cap \mathsf{k}^\times$, so λ is not totally positive in k. □

We are now a definition and a few results away from defining the signature relative to an ordering.

Definition 16.10. Let P be an ordering of a field k. A symmetric bilinear form (V, B) over k is *postive definite* (relative to P) if $B(v, v) \in P$ for all $v \in V \smallsetminus \{0\}$, and *negative definite* if $B(v, v) \in -P$ for all $v \in V \smallsetminus \{0\}$.

We need the following analogue of Sylvester's law of intertia to define the signature of a regular quadratic form over k relative to an ordering $P \in X_\mathsf{k}$.

Theorem 16.11 (Sylvester's law of intertia for an ordered field). *Any regular symmetric bilinear form (V, B) over an ordered is isometric to $V^+ \perp V^-$ with V^+ postive definite and V^- negative definite. The dimensions of V^+ and V^- are isometry invariants.*

Proof. Choose an orthogonal basis $v_1, \ldots, v_n$ for V, let V^+ be spanned by the v_i such that $B(v_i, v_i) \in P$, and let V^- be spanned by the v_j such that $B(v_j, v_j) \in -P$. Then $V \cong V^+ \perp V^-$ with V^+ postive definite and V^- negative definite.

Now let W be a positive definite subspace of V. Then $W \cap V^- = 0$, so

$$\dim W \leq \dim V - \dim V^- = \dim V^+.$$

It follows that $\dim V^+$ is the maximum possible dimension of a positive definite subspace of V, and this quantity is clearly invariant under isometry. □

Definition 16.12. If (k, P) is an ordered field and q is a regular quadratic form over k, define the *P-signature* of q to be

$$\operatorname{sgn}_P(q) = \dim V^+ - \dim V^-$$

where (V, B) is the symmetric bilinear form associated with q.

Theorem 16.11 implies that P-signature is an isometry invariant, and it thus defines a function $\operatorname{sgn}_P \colon M(\mathsf{k}) \to \mathbb{Z}$. Working with diagonalizations, it is easy to check that $\operatorname{sgn}_P(f \perp g) = \operatorname{sgn}_P(f) + \operatorname{sgn}_P(g)$, that $\operatorname{sgn}_P(f \otimes g) = \operatorname{sgn}_P(f)\operatorname{sgn}_P(g)$, and that $\operatorname{sgn}_P(\langle 1 \rangle) = 1$. (In fact, $\operatorname{sgn}_P(\langle \lambda \rangle) = \pm 1$ according to whether $\lambda \in P$ or $\lambda \in -P$.) Thus sgn_P is a semiring homomorphism that extends uniquely to a ring homomorphism $\operatorname{sgn}_P \colon \mathrm{GW}(\mathsf{k}) \to \mathbb{Z}$. We also have $\operatorname{sgn}_P(h) = 1 - 1 = 0$, so sgn_P further extends to $\mathrm{W}(\mathsf{k})$.

Remark 16.13. Some texts produce P-signatures via an alternate method involving taking the real closure of (k, P). A field is called *real closed* if it is formally real, but no proper algebraic extension of it is formally real. A field extension $F \supseteq \mathsf{k}$ is called a *real closure* of (k, P) if (a) F is real closed, (b) F is algebraic over k, and (c) $P = \mathsf{k}^\times \cap F^{\boxtimes}$. As it turns out, real closures exist and are unique up to order-preserving isomorphism. We let k_P denote the real closure of (k, P). Then $\mathrm{W}(\mathsf{k}_P) \cong \mathbb{Z}$, and sgn_P is isomorphic to the extension of scalars homomorphism $\mathrm{W}(\mathsf{k}) \to \mathrm{W}(\mathsf{k}_P)$.

17. Total signature and Pfister's local-global principle

Many fields have multiple (even infinitely many) orderings, so it behooves us to consider all of the P-signatures ($P \in X_\mathsf{k}$) at once. The *total signature* is the homomorphism

$$\begin{aligned} \mathrm{sgn}\colon \mathrm{W}(\mathsf{k}) &\longrightarrow \prod_{P \in X_\mathsf{k}} \mathbb{Z} \\ q &\longmapsto (\mathrm{sgn}_P(q))_P. \end{aligned}$$

Pfister's local-global principle computes the kernel of this map.

Theorem 17.1 (Pfister's local-global principle)**.** *For any field* k, $\ker(\mathrm{sgn}) = \mathrm{W}(\mathsf{k})_{\mathrm{tors}}$, *the torsion subgroup of* $\mathrm{W}(\mathsf{k})$. *Moreover, every element of* $\mathrm{W}(\mathsf{k})_{\mathrm{tors}}$ *has order a power of* 2.

A full proof of this theorem can be found in [Lam05, §VIII.3] or [MH73, §III.3]. Both proofs depend on certain information about prime ideals in $\mathrm{W}(\mathsf{k})$ and a computation of the kernel of the homomorphism $\mathrm{W}(\mathsf{k}) \to \mathrm{W}(\mathsf{k}(\sqrt{\lambda})$ induced by extension of scalars. In the interest of time, we will not give all the details here. Note, though, that the inclusion $\ker(\mathrm{sgn}) \subseteq \mathrm{W}(\mathsf{k})_{\mathrm{tors}}$ is relatively easy to prove. Indeed, if $q \in \mathrm{W}(\mathsf{k})$ is torsion, then $\mathrm{sgn}_P(q) = 0$ since $\mathbb{Z}$ is torsion-free. Since $\ker(\mathrm{sgn}) = \bigcap_{P \in X_\mathsf{k}} \ker(\mathrm{sgn}_P)$, this shows that $\ker(\mathrm{sgn}) \subseteq \mathrm{W}(\mathsf{k})_{\mathrm{tors}}$. For the converse, the crucial thing to prove is that if $q \in \mathrm{W}(\mathsf{k})$ is not 2-primary torsion, then there exists an ordering $P \in X_\mathsf{k}$ such that $\mathrm{sgn}_P(q) \neq 0$.

Remark 17.2. The "local-global" nature of Theorem 17.1 is best understood in the context of Remark 16.13. In this form, the total signature looks like

$$\mathrm{W}(\mathsf{k}) \longrightarrow \prod_{P \in X_\mathsf{k}} \mathrm{W}(\mathsf{k}_P).$$

We think of each real closure k_P as "local" information, and of k itself as "global." Pfister's theorem says that taken together (over all $P \in X_\mathsf{k}$), the local information detects global torsion.

There are many nice interpretations of Theorem 17.1. Suppose, for instance, that quadratic forms f and g over k have the same P-signature for all $P \in X_\mathsf{k}$. We know, then, that there exists a postive integer n such that $nf \perp \langle -1 \rangle ng$ is hyperbolic; moreover, the smallest such n is a power of 2.

It is also nice to see that Theorem 17.1 describes the torsion in $\mathrm{GW}(\mathsf{k})$. Indeed, $\mathrm{GW}(\mathsf{k})_{\mathrm{tors}} \subseteq \mathrm{GI}(\mathsf{k})$ since classes with nonzero dimension cannot be torsion. Since $\mathrm{GI}(\mathrm{tors}$

We would also like to understand the cokernel of sgn, but first we must cut its codomain down to a more reasonable size. To do so requires the topology on X_k introduced in footnote 17. Let $\mathbb{Z}^{X_\mathsf{k}}$ denote the ring of *continuous* functions from X_k (with the Harrison topology) to $\mathbb{Z}$ (with the discrete topology). The function $\mathrm{sgn}(q) : X_\mathsf{k} \to \mathbb{Z}$ taking $P \mapsto \mathrm{sgn}_P(q)$ is continuous, so the total signature factors through $\mathbb{Z}^{X_\mathsf{k}}$.

Theorem 17.3. *The cokernel* $\mathbb{Z}^{X_\mathsf{k}} / \mathrm{sgn}(\mathrm{W}(\mathsf{k}))$ *is a* 2*-primary torsion group.*

Corollary 17.4. Suppose that $|X_\mathsf{k}| = r < \infty$. Then, as an Abelian group, $\mathrm{W}(\mathsf{k})$ is isomorphic to $T \oplus \mathbb{Z}^r$ where T is a 2-primary torsion group.

18. Pfister forms

Given $\lambda_1, \dots, \lambda_n \in \mathsf{k}^\times$, the *$n$-fold Pfister form* determined by these scalars is

$$\langle\langle \lambda_1, \dots, \lambda_n \rangle\rangle = \bigotimes_{i=1}^{n} \langle 1, \lambda_i \rangle .$$

By convention, the 0-fold Pfister form is $\langle\langle\rangle\rangle = \langle 1 \rangle$, and we may compute

$$\begin{aligned} \langle\langle \lambda \rangle\rangle &= \langle 1, \lambda \rangle , \\ \langle\langle \lambda, \mu \rangle\rangle &= \langle 1, \lambda, \mu, \lambda\mu \rangle , \\ \langle\langle \lambda, \mu, \nu \rangle\rangle &= \langle 1, \lambda, \mu, \nu, \lambda\mu, \lambda\nu, \mu\nu, \lambda\mu\nu \rangle . \end{aligned}$$

Clearly $\dim \langle\langle \lambda_1, \dots, \lambda_n \rangle\rangle = 2^n$.

Pfister forms arise as norm forms of quaternion and Cayley-Dickson algebras, and as trace forms of multi-quadratic field extensions. They are also important in the study of the $\mathrm{I}(\mathsf{k})$-adic filtration of $\mathrm{W}(\mathsf{k})$.

Proposition 18.1. The n-fold Pfister form $\langle\langle \lambda_1, \dots, \lambda_n \rangle\rangle$ is hyperbolic if any of the λ_i has the square class of -1. If $\lambda_1 \in \mathsf{k}^{\boxtimes}$, then

$$\langle\langle \lambda_1, \lambda_2, \dots, \lambda_n \rangle\rangle \cong 2 \langle\langle \lambda_2, \dots, \lambda_n \rangle\rangle .$$

Proof. If $\lambda_i \in -\mathsf{k}^{\boxtimes}$, then there is a factor of h in $\prod_{j=1}^{n} \langle 1, \lambda_j \rangle$, and Lemma 10.10 implies that

$$\langle\langle \lambda_1, \dots, \lambda_n \rangle\rangle = 2^{n-1} h.$$

If $\lambda_1 \in \mathsf{k}^{\boxtimes}$, then $\langle 1, \lambda_1 \rangle = 2 \langle 1 \rangle$, given the desired equivalence. □

Proposition 18.2. The ideal $\mathrm{I}(\mathsf{k})^n \subseteq \mathrm{W}(\mathsf{k})$ is additively generated by the n-fold Pfister forms.

Proof. We saw in Proposition 12.3 that $\mathrm{GI}(\mathsf{k}) \subseteq \mathrm{GW}(\mathsf{k})$ is additively generated by $\langle 1 \rangle - \langle \lambda \rangle$, $\lambda \in \mathsf{k}^\times$. Since $-\langle \lambda \rangle = \langle -\lambda \rangle$ in $\mathrm{W}(\mathsf{k})$, we know that $\langle 1, -\lambda \rangle = \langle\langle -\lambda \rangle\rangle$, $\lambda \in \mathsf{k}^\times$ generate $\mathrm{I}(\mathsf{k})$. For the statement regarding $\mathrm{I}(\mathsf{k})^n$, note that

$$\langle\langle \lambda_1, \dots, \lambda_n \rangle\rangle = \langle\langle \lambda_1 \rangle\rangle \cdots \langle\langle \lambda_n \rangle\rangle .$$

□

Presently, we will concern ourselves with how the properties of n-fold Pfister forms are built up from those of 1- and 2-fold Pfister forms. Recall that $D(q) = D_{\mathsf{k}}(q)$ is the set of nonzero values in k represented by a quadratic form q.

Proposition 18.3. (a) For any $\mu \in D(\langle\langle \lambda_1 \rangle\rangle)$, $\langle\langle \lambda_1, \lambda_2 \rangle\rangle \cong \langle\langle \lambda_1, \lambda_2 \mu \rangle\rangle$.
(b) For any $\mu \in D(\langle \lambda_1, \lambda_2 \rangle)$, $\langle\langle \lambda_1, \lambda_2 \rangle\rangle \cong \langle\langle \mu, \lambda_1 \lambda_2 \rangle\rangle$.

Proof. For (a), note that $\langle\langle \lambda_1, \lambda_2 \rangle\rangle = \langle 1, \lambda_1 \rangle \perp \langle \lambda_2, \lambda_1 \lambda_2 \rangle$. The second term is equivalent to $\langle \lambda_2 \rangle \langle 1, \lambda_1 \rangle \cong \langle \lambda_2 \rangle \langle \mu, \mu\lambda_1 \rangle$, where the last equivalence uses Proposition 9.1. We thus have the chain of equivalences

$$\langle\langle \lambda_1, \lambda_2 \rangle\rangle \cong \langle 1, \lambda_1 \rangle \perp \langle \mu\lambda_2, \mu\lambda_1\lambda_2 \rangle \cong \langle\langle \lambda_1, \lambda_2 \mu \rangle\rangle .$$

For (b), compute

$$\langle\langle \lambda_1, \lambda_2 \rangle\rangle \cong \langle 1, \lambda_1\lambda_2, \lambda_1, \lambda_2 \rangle \cong \langle 1, \lambda_1\lambda_2, \mu, \lambda_1\lambda_2\mu \rangle \cong \langle\langle \mu, \lambda_1\lambda_2 \rangle\rangle$$

□

The following definition is in analogy with that of chain equivalence.

Definition 18.4. Let $\langle\langle \lambda_1, \dots, \lambda_n \rangle\rangle$ and $\langle\langle \mu_1, \dots, \mu_n \rangle\rangle$ be two n-fold Pfister forms over k. We say that these forms are *simply P-equivalent* if there exist indices i, j such that

(a) $\langle\langle \lambda_1, \lambda_2 \rangle\rangle \cong \langle\langle \mu_1, \mu_2 \rangle\rangle$, and
(b) $\lambda_k = \mu_k$ for any $k \neq i, j$.

(Note that when $i = j$, we take (a) to mean that $\langle\langle \lambda_i \rangle\rangle \cong \langle\langle \mu_i \rangle\rangle$.) We say that two n-fold Pfister forms f and g are *chain P-equivalent* if there exists a sequence of n-fold Pfister forms $f = \varphi_0, \varphi_1, \ldots, \varphi_m = g$ such that φ_i is simply P-equivalent to φ_{i+1} for $0 \leq i \leq m-1$. When two Pfister forms f, g are chain P-equivalent, we write $f \approx g$.

Clearly $f \approx g$ implies $f \cong g$. The converse is true as well (see [Lam05, Theorem X.1.12]), but we will not present the proof. Since transpositions generate the symmetric group, we have

$$\langle\langle \lambda_1, \ldots, \lambda_n \rangle\rangle \approx \left\langle\left\langle \lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(n)} \right\rangle\right\rangle$$

for any $\sigma \in \Sigma_n$.

Since any Pfister form f represents 1, we may write $f \cong \langle 1 \rangle \perp f'$. We call f' the *pure subform* of f. By the Witt Cancellation Theorem 8.2, the isometry type of f' is well-defined.

Theorem 18.5. *[Pure suform theorem] Let $f = \langle\langle \lambda_1, \ldots, \lambda_n \rangle\rangle$, $n \geq 1$, and let $\mu \in D(f')$. Then there exist $\mu_2, \ldots, \mu_n \in \mathsf{k}^\times$ such that*

$$f \approx \langle\langle \mu, \mu_2, \ldots, \mu_n \rangle\rangle .$$

Proof. We proceed by induction on n. If $n = 1$, then $f = \langle 1, \lambda_1 \rangle$ and $f' = \langle \lambda_1 \rangle$. We have $\langle \mu \rangle \cong \langle \lambda_1 \rangle$, so the result follows.

Now assume the theorem holds for $(n-1)$-fold Pfister forms. Let

$$g = \langle\langle \lambda_1, \ldots, \lambda_{n-1} \rangle\rangle \cong \langle 1 \rangle \perp g'.$$

Then $f = g \otimes \langle 1, \lambda_n \rangle \cong g \perp \langle \lambda_n \rangle \otimes g$, so $f' \cong g' \perp \langle \lambda_n \rangle \otimes g$. We have $\mu \in D(f')$, so there exist $x \in D(g') \cup \{0\}$ and $y \in D(g) \cup \{0\}$ such that $\mu = x + \lambda_n y$. Furthermore, $y = t^2 + y_0$ for some $t \in \mathsf{k}$ and $y_0 \in D(g') \cup \{0\}$.

Proceeding by cases, suppose that $y = 0$. Then $\mu = x \in D(g')$. By the inductive hypothesis, there exist $\nu_2, \ldots, \nu_{n-1} \in \mathsf{k}^\times$ such that $g \approx \langle\langle x, \nu_2, \ldots, \nu_{n-1} \rangle\rangle$. Thus

$$f \approx \langle\langle x, \nu_2, \ldots, \nu_{n-1}, \lambda_n \rangle\rangle = \langle\langle \mu, \nu_2, \ldots, \nu_{n-1}, \lambda_n \rangle\rangle ,$$

and we are done.

Now suppose that $y \neq 0$. We will first show that $f \approx \langle\langle \lambda_1, \ldots, \lambda_{n-1}, \lambda_n y \rangle\rangle$. If $y_0 = 0$, then $y = t^2 \in \mathsf{k}^{\boxtimes}$, and this is obvious. So we may assume $y_0 \in D(g')$. By the inductive hypothesis, $g \approx \langle\langle y_0, \mu_2, \ldots, \mu_{n-1} \rangle\rangle$ for some $\mu_i \in \mathsf{k}^\times$. Thus

$$\begin{aligned} f &\approx \langle\langle y_0, \mu_2, \ldots, \mu_{n-1}, \lambda_n \rangle\rangle \\ &\approx \left\langle\left\langle y_0, \mu_2, \ldots, \mu_{n-1}, \lambda_n (t^2 + y_0) \right\rangle\right\rangle \quad \text{by Proposition 18.3(a)} \\ &\approx \langle\langle \lambda_1, \ldots, \lambda_{n-1}, \lambda_n y \rangle\rangle , \end{aligned}$$

as claimed.

If $x = 0$, then $\lambda_n y = \mu$, and we are done. Thus we may assume $x \in D(g')$. By our inductive hypothesis, $g \approx \langle x, \nu_2, \ldots, \nu_{n-1} \rangle$ for some $\nu_i \in \mathsf{k}^\times$. Thus

$$\begin{aligned} f &\approx \langle\langle x, \nu_2, \ldots, \nu_{n-1}, \lambda_n y \rangle\rangle \\ &\approx \langle\langle x + \lambda_n y, \nu_2, \ldots, \nu_{n-1}, \lambda_n xy \rangle\rangle \quad \text{by Proposition 18.3(b)} \\ &\approx \langle\langle \mu, \nu_2, \ldots, \nu_{n-1}, \lambda_n xy \rangle\rangle . \end{aligned}$$

This completes the proof. □

In the course of the above proof, we encountered the following phenomenon.

Proposition 18.6. Let $g = \langle\langle \lambda_1, \ldots, \lambda_{n-1} \rangle\rangle$ and let $y \in D(g)$. Then for any $\lambda_n \in \mathsf{k}^\times$,

$$\langle\langle \lambda_1, \ldots, \lambda_{n-1}, \lambda_n \rangle\rangle \approx \langle\langle \lambda_1, \ldots, \lambda_{n-1}, \lambda_n y \rangle\rangle .$$

In particular, $\langle\langle \lambda_1, \ldots, \lambda_{n-1}, y \rangle\rangle \approx 2g$ and $\langle\langle \lambda_1, \ldots, \lambda_{n-1}, -y \rangle\rangle \cong 2^{n-1}h$.

Proof. The first portion is proved in the $y \neq 0$ case above. By setting $\lambda_n = \pm 1$, we get the second statement. □

Theorem 18.7. *If a Pfister form is isotropic, then it is hyperbolic.*

Proof. Suppose that f is an isotropic Pfister form. Then f contains a hyperbolic plane, hence $-1 \in D(f')$. By Theorem 18.5, $f \approx \langle\langle -1, \ldots \rangle\rangle$, which is hyperbolic. □

We now introduce similarity factors, a concept crucial to the development of Pfister's theory.

Definition 18.8. For any quadratic form q,

$$G(q) = G_{\mathsf{k}}(q) = \{\lambda \in \mathsf{k}^\times \mid \langle\lambda\rangle \otimes q \cong q\}$$

is called the group of *similarity factors of* q.

Remark 18.9. We clearly have $1 \in G(q)$, and $\langle\lambda\rangle \cong \langle\lambda^{-1}\rangle$, so $G(q)$ is closed under inverses. If $\lambda, \mu \in G(q)$, then $\langle\lambda\mu\rangle \otimes q \cong \langle\lambda\rangle \otimes (\langle\mu\rangle \otimes q) \cong \langle\lambda\rangle \otimes q \cong q$, so $G(q)$ is in fact a subgroup of $\mathsf{k}^\times$.

The fact that $G(q)$ is a group will be essential in proving that Pfister forms are group forms. We will also need the following lemma.

Lemma 18.10. If f and g are regular quadratic forms with $\dim f = \dim g$ and $f \perp g$ is hyperbolic, then $f \cong \langle -1 \rangle \otimes g$.

Proof. Assume that $\dim f = \dim g$ and $f \perp g \cong nh$ for some $n \geq 0$. Then $f + g = 0$ in $\mathrm{W}(\mathsf{k})$, so $f = \langle -1 \rangle \otimes g$ in $\mathrm{W}(\mathsf{k})$. Since f and $\langle -1 \rangle \otimes g$ have the same dimension, it follows that $f \cong \langle -1 \rangle \otimes g$ by Proposition 12.5(c). □

We are now ready to prove one of the most remarkable properties of Pfister forms.

Theorem 18.11. *For any Pfister form f, $D(f) = G(f)$. In particular, f is a group form.*

Proof. Suppose that $\lambda \in G(f)$ so that $\langle\lambda\rangle \otimes f \cong f$. Then f has a diagonalization with λ as one of its coefficients, hence $\lambda \in D(f)$. This proves that $G(f) \subseteq D(f)$.

Now suppose that $\lambda \in D(f)$. The Pfister form $f \otimes \langle\langle -\lambda \rangle\rangle \cong f \perp (\langle -\lambda \rangle \otimes f)$ contains a subform $\langle \lambda, -\lambda \rangle \cong h$. By Proposition 18.6, $f \perp (\langle -\lambda \rangle \otimes f)$ is hyperbolic, so Lemma 18.10 implies that $f \cong \langle\lambda\rangle \otimes f$. It follows that $D(f) \subseteq G(f)$ as well, so $D(f) = G(f)$. □

Corollary 18.12. Over any field k, $2^n \langle 1 \rangle$ is a group form. In particular, the product of two sums of 2^n squares is itself a sum of 2^n squares.

Proof. This follows immediately from Theorem 18.11 once we observe that $2^n \langle 1 \rangle$ is the n-fold Pfister form $\langle\langle 1, \ldots, 1 \rangle\rangle$. □

See Remark 6.4 for explicit formulas when $n = 1, 2, 3$ and further discussion of this problem.

19. Multiplicative forms

We have just seen that Pfister forms are group forms. Interpreted in the context of function fields, this has an important consequence. Let $\mathbf{x} = (x_1, \ldots, x_{2^m})$ and let $\mathbf{y} = (y_1, \ldots, y_{2^m})$ be commuting variables over k and let $L = \mathsf{k}(\mathbf{x}, \mathbf{y})$ be the field of rational functions in these variables. If f is an m-fold Pfister form over k, then f_L (which is just f with its coefficients considered as elements of L) is an m-fold Pfister form over L. In particular, $f(\mathbf{x}) \cdot f(\mathbf{y}) = f(z_1, \ldots, z_{2^m})$ for some rational functions $z_i \in \mathsf{k}(\mathbf{x}, \mathbf{y})$. The following definition formalizes this property.

Definition 19.1. Let f be an n-dimensional quadratic form over k and let $L = \mathsf{k}(\mathbf{x}, \mathbf{y})$ for $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$ commuting variables over k. Call f *multiplicative* if $f(\mathbf{x})f(\mathbf{y}) \in D_L(f)$.

Remark 19.2. By the first paragraph of this section, every Pfister form is multiplicative. Also, if f is isotropic, then $D_L(f) = L^\times$, so isotropic forms are multiplicative as well.

Proposition 19.3. A quadratic form f over k is multiplicative if and only if $D_F(f)$ is a group for all field extensions $F \supseteq \mathsf{k}$.

Proof. If $D_F(f)$ is a group for all field extensions $F \supseteq \mathsf{k}$, then, in particular, $D_L(f)$ is a group for $L = \mathsf{k}(\mathbf{x}, \mathbf{y})$. It follows that $f(\mathbf{x})f(\mathbf{y}) \in D_L(f)$, as desired.

Now suppose that f is multiplicative in the sense of Definition 19.1, with

$$f(\mathbf{x})f(\mathbf{y}) = f(z_1(\mathbf{x}, \mathbf{y}), \ldots, z_n(\mathbf{x}, \mathbf{y})).$$

For any field extension $F \supseteq \mathsf{k}$ and $\lambda, \mu \in F^n$, we see that

$$f(\lambda)f(\mu) = f(z_1(\lambda, \mu), \ldots, z_n(\lambda, \mu)),$$

for $D_F(f)$ is closed under multiplication. We have previously observed that this condition implies that $D_F(f)$ is a group.[18] □

We need one more notion of multiplicativity before stating our main theorem.

Definition 19.4. An n-dimensional quadratic form f over k is *strongly multiplicative* if $f(\mathbf{x}) \in G_{\mathsf{k}(x)}(f)$ for $\mathbf{x} = (x_1, \ldots, x_n)$.

Remark 19.5. (a) The above condition says that $\langle f(\mathbf{x})\rangle \otimes f \cong f$ over $\mathsf{k}(\mathbf{x})$.
(b) If f is a Pfister form over k, then $f(\mathbf{x})$ is clearly in $D_{\mathsf{k}(\mathbf{x})}(f)$, and $D_{\mathsf{k}(\mathbf{x})} = G_{\mathsf{k}(\mathbf{x})}$ by Theorem 18.11, so f is strongly multiplicative.
(c) If f is hyperbolic over k, then $G_{\mathsf{k}(\mathbf{x})}(f) = \mathsf{k}(\mathbf{x})^\times$, so f is strongly multiplicative.

The following theorem characterizes anisotropic Pfister forms. (Recall from Theorem 18.7 that isotropic Pfister forms are hyperbolic.)

Theorem 19.6. *For any anisotropic quadratic form over* k*, the following are equivalent:*
(a) f is a Pfister form,
(b) f is multiplicative,
(c) f is strongly multiplicative.

Proof sketch. We have already observed that Pfister form are strongly multiplicative, so (a) $\implies$ (c).

To prove (c) $\implies$ (b), suppose that $\langle f(\mathbf{x})\rangle \otimes f \cong f$ over $\mathsf{k}(\mathbf{x})$. Then f represents $f(\mathbf{x})f(\mathbf{y})$ over $\mathsf{k}(\mathbf{x}, \mathbf{y})$, so f is multiplicative.

Finally, we prove (b) $\implies$ (a). Assume that f is multiplicative, in which case it is a group form over k (Proposition 19.3), so $1 \in D_{\mathsf{k}}(f)$. It follows that f contains the 0-fold Pfister form $\langle 1\rangle$. Choose r maximal such that f contains an r-fold Pfister form φ over k, say $f \cong \varphi \perp f_0$. If $\dim f_0 = 0$, we are done, so assume $f_0 \cong \langle \lambda, \ldots\rangle$.

We claim that $\varphi \perp \langle\lambda\rangle\varphi$ is an $(r+1)$-fold Pfister form which is a subform of f. If so, we have reached a contradiction since φ was maximal. The proof of this claim depends on material we have not developed, namely the Third Representation Theorem [Lam05, Theorem IX.2.8]. □

Remark 19.7. Note that (a) $\implies$ (c) $\implies$ (b) does not depend on the anisotropicity hypothesis above.

[18] The trick was to notice that $D_F(f)$ is closed under inverses: if $f(x) = \lambda$, then $f(\lambda^{-1}x) = \lambda^{-2}\lambda = \lambda^{-1}$.

We now now that multiplicative forms are either isotropic or they are anisotropic Pfister forms. It turns out that isotropic strongly multiplicative forms are precisely the hyperbolic forms, see [Lam05, Theorem X.2.9].

By definition, multiplicative quadratic forms satisfy the identity

$$f(\mathbf{x})f(\mathbf{y}) = f(z_1, \dots, z_n)$$

for some $z_i \in \mathsf{k}(\mathbf{x}, \mathbf{y})$. The following theorem tells us something about the form which the z_i may take.

Theorem 19.8. *An n-dimensional quadratic form f over k is strongly multiplicative if and only if there exist $z_1, \dots, z_n \in \mathsf{k}(\mathbf{x})[\mathbf{y}]$ that are homogeneous degree 1 in $\mathbf{y} = (y_1, \dots, y_n)$.*

Proof. First suppose that f is strongly multiplicative, so $\langle f(\mathbf{x})\rangle \otimes f \cong f$ over $\mathsf{k}(\mathbf{x})$. This means that there is a matrix $B \in \mathrm{GL}_n(\mathsf{k}(\mathbf{x}))$ such that

$$f(\mathbf{x})f(\mathbf{y}) = f(\mathbf{y}B).$$

Thus we may take $(z_1, \dots, z_n) = \mathbf{y}B$, which clearly has the desired form. The converse follows by reversing these steps. □

20. A glimpse into function fields and the Hauptsatz

For R a commutative ring and $J \subseteq R$ an ideal, the J-adic filtration

$$R \supseteq J \supseteq J^2 \supseteq J^3 \supseteq \cdots$$

of R is called *Hausdorff* if $\bigcap_n J^n = 0$. The Arason–Pfister Hauptsatz tells us that the $\mathrm{I}(\mathsf{k})$-adic filtration of $\mathrm{W}(\mathsf{k})$ is Hausdorff:

Theorem 20.1 (Arason–Pfister Hauptsatz)**.** *For all fields k,*

$$\bigcap_{n\geq 0} \mathrm{I}(\mathsf{k})^n = 0.$$

This is an immediate consequence of the following theorem.

Theorem 20.2. *If q is a positive-dimensional anisotropic form in $\mathrm{I}(\mathsf{k})^n$, then $\dim q \geq 2^n$.*

For nontorsion elements $q \in \mathrm{W}(\mathsf{k})$, Pfister's local-global principle (Theorem 17.1) tells us that there is an ordering $P \in X_{\mathsf{k}}$ such that $\mathrm{sgn}_P(q) \neq 0$. If $q \in \mathrm{I}(\mathsf{k})^n$, then $\mathrm{sgn}_P(q) = 2^n k$ for some $k \in \mathbb{Z} \smallsetminus \{0\}$ (because $\mathrm{sgn}_P \mathrm{I}(\mathsf{k}) \subseteq 2\mathbb{Z}$). It follows that $\dim q \geq 2^n|k| \geq 2^n$, so the conclusion of Theorem 20.2 follows for nontorsion forms.

Despite the ease of this argument in the nontorsion case, Theorem 20.2 for torsion elements is quite deep and one of the high points of the algebraic theory of quadratic forms. A full proof would take significant space and time, but we will sketch the main ideas here. See [Lam05, §X.5] for the details.

The main tool in Arason–Pfister's proof is the function field of a quadratic form. The definition of this object depends on the following lemma.

Lemma 20.3. Fix $n \geq 1$ and let $f(x_0, \dots, x_n)$ be a regular $(n+1)$-dimensional quadratic form over k. Then f is reducible in $\mathsf{k}[x_0, \dots, x_n]$ if and only if $n = 1$ and $f \cong h$.

Proof. Suppose that f is reducible. Then it factors as a product of two linear forms (linear polynomials with constant term 0). This happens if and only if f is isometric to $x_0x_1 \cong h$. □

Definition 20.4. For $f \ncong h$ a regular quadratic form over k of dimension at least 2, the *function field* of f is

$$\mathsf{k}[f] = \mathrm{Frac}\, \mathsf{k}[\mathbf{x}]/(f(\mathbf{x})),$$

the field of fractions of $\mathsf{k}[\mathbf{x}]/(f(\mathbf{x}))$, where $\mathbf{x} = (x_0, \dots, x_n)$ is the set of variables for f.

Remark 20.5. By Lemma 20.3, we know that f is irreducible as a polynomial in the above definition. Since $\mathsf{k}[\mathbf{x}]$ is a unique factorization domain, it follows that the ideal $(f(\mathbf{x}))$ is prime, so $\mathsf{k}[\mathbf{x}]/(f(\mathbf{x}))$ is an integral domain, so it has a field of fractions.

Remark 20.6. If we diagonalize f as $\langle \lambda_0, \dots, \lambda_n \rangle$, then

$$\mathsf{k}[f] \cong \operatorname{Frac} \mathsf{k}[\mathbf{x}]/(\lambda_0 x_0^2 + \cdots + \lambda_n x_n^2).$$

Thus in $\mathsf{k}[f]$, we have $x_0 = \sqrt{-(\lambda_1 x_1^2 + \cdots + \lambda_n x_n^2)/\lambda_0}$. It follows that $\mathsf{k}[f]$ is a quadratic extension of a rational function field in n variables, *i.e.*,

$$\mathsf{k}[f] = \mathsf{k}(x_1, \dots, x_n)\left(\sqrt{-(\lambda_1 x_1^2 + \cdots + \lambda_n x_n^2)/\lambda_0}\right).$$

Remark 20.7. For $\lambda \in \mathsf{k}^\times$, we have $(\lambda f(\mathbf{x})) = (f(\mathbf{x}))$ as ideals in $\mathsf{k}[\mathbf{x}]$. It follows that

$$\mathsf{k}[f] = \mathsf{k}[\langle \lambda \rangle \otimes f],$$

so function fields do not distinguish similar quadratic forms. Note, though, that $\mathsf{k}[f] = \mathsf{k}[g]$ does not necessarily imply that $f = \langle \lambda \rangle \otimes g$ for some $\lambda \in \mathsf{k}^\times$.

Clearly $\mathsf{k}[f]$ is a field extension of k, so extension of scalars induces a homomorphism $\operatorname{res}_{\mathsf{k}}^{\mathsf{k}[f]} \colon \mathrm{W}(\mathsf{k}) \to \mathrm{W}(\mathsf{k}[f])$. Since $f(\mathbf{x}) = 0$ in $\mathsf{k}[f]$, we know that f is isotropic over f. It turns out to be quite interesting to study what other quadratic forms become isotropic, or even hyperbolic, over $\mathsf{k}[f]$. This latter class is precisely the kernel of $\operatorname{res}_{\mathsf{k}}^{\mathsf{k}[f]}$, which we give a special name, $\mathrm{W}(\mathsf{k}[f]/\mathsf{k})$.

Theorem 20.8. *If $q \in \mathrm{W}(\mathsf{k}[f]/\mathsf{k})$ and $1 \in D_{\mathsf{k}}(f)$, then $f(\mathbf{x}) \in G_{\mathsf{k}(\mathbf{x})}(q)$ where $\dim f = n+1$ and $\mathbf{x} = (x_0, \dots, x_n)$. If q is anisotropic and $\lambda \in D_{\mathsf{k}}(q)$, then $\langle \lambda \rangle \otimes f$ is a subform of q; in particular, $\dim q \geq \dim f$ as long as $\dim q \neq 0$.*

We defer the proof of Theorem 20.8 for the moment and instead see how it is used to prove the Hauptsatz.

Proof of Theorem 20.2. Suppose that f is a positive-dimensional anisotropic form in $\mathrm{I}(\mathsf{k})^n$. Since n-fold Pfister forms generate $\mathrm{I}(\mathsf{k})^n$, we know that

$$f = \varepsilon_1 \varphi_1 + \cdots + \varepsilon_r \varphi_r$$

where each $\varepsilon_i \in \{\pm 1\}$ and the φ_i are anisotropic n-fold Pfister forms. To show that $\dim f \geq 2^n$, we proceed by induction on r. If $r = 1$, then $f = \langle \pm 1 \rangle \varphi_1$, so $\dim q = 2^n$, as desired.

For the induction step, consider the image of f in $L = \mathsf{k}[\varphi_1]$. Writing $g_L = \operatorname{res}_{\mathsf{k}}^L g$ and noting that $(\varphi_1)_L = 0$ (since isotropic Pfister forms are hyperbolic by Proposition 18.6), we get

$$f_L = \varepsilon_2 (\varphi_2)_L + \cdots + \varepsilon_r (\varphi_r)_L \in \mathrm{I}(L)^n.$$

If f_L is hyperbolic, then Theorem 20.8 implies that $\dim f \geq \dim \varphi_1 = 2^n$. Thus we may assume that $(f_L)_{\mathrm{an}}$, the anisotropic part of f_L, is a positive-dimensional form in $\mathrm{I}(L)^n$. Invoking the inductive hypothesis over L implies that $\dim_L (f_L)_{\mathrm{an}} \geq 2^n$. But then

$$\dim_{\mathsf{k}} f = \dim_L f_L \geq \dim_L (f_L)_{\mathrm{an}} \geq 2^n,$$

as desired. □

We conclude this section by sketching the proof of Theorem 20.8, which was crucial to the above argument. See [Lam05, Theorem 4.5] for details.

Proof sketch for Theorem 20.8. Since $1 \in D_{\mathsf{k}}(f)$, we have $f \cong \langle 1 \rangle \perp f'$, whence

$$L = \mathsf{k}[f] = F(\sqrt{-f'(\mathbf{x}')})$$

where $\mathbf{x}' = (x_1, \ldots, x_n)$ and $F = \mathsf{k}(\mathbf{x}')$. Without loss of generality, we may assume that q is anisotropic over k. One can then check that $f_F = \operatorname{res}^F_{\mathsf{k}} f$ is anisotropic (using, for instance, [Lam05, Lemma IX.1.1]). By hypothesis, $f_L = \operatorname{res}^L_{\mathsf{k}} f = \operatorname{res}^L_F f_F$ is hyperbolic. One then checks that this implies

$$f_F \cong g \otimes \langle 1, f'(\mathbf{x}') \rangle \tag{20.9}$$

over F (using [Lam05, Theorem VII.3.2]). Over the rational function $F(x_0) = \mathsf{k}(\mathbf{x})$, we know that $\langle 1, f'(\mathbf{x}') \rangle$ represents $x_0^2 + f'(\mathbf{x}') = f(\mathbf{x})$. Since $\langle 1, f'(\mathbf{x}') \rangle$ is a Pfister form over $\mathsf{k}(\mathbf{x})$, Theorem 18.11 implies that $f(\mathbf{x})$ is a similarity factor of $\langle 1, f'(\mathbf{x}') \rangle$. By (20.9), it follows that $f(\mathbf{x}) \in G_{\mathsf{k}(\mathbf{x})}(q)$ as well. The final claim is another consequence of the Third Representation Theorem [Lam05, Theorem IX.2.8]. □

The missing components of our sketch amount to information about the functorial properties of W under quadratic and transcendental field extensions, important aspects of quadratic forms which the reader is invited to study in [Lam05, Chapters VII & IX].

Since we referenced it twice above, we will also state the Third Representation Theorem.

Theorem 20.10 (Third Representation Theorem)**.** *For any form f and any anisotropic form g over k, the following are equivalent:*

(a) $f(\mathbf{x}) \in D_{\mathsf{k}(\mathbf{x})}(g)$,
(b) f *is a subform of* g *over* k,
(c) $D_F(f) \subseteq D_F(g)$ *for any field extension* $F \supseteq \mathsf{k}$.

In particular, if $f(\mathbf{x}) \in D_{\mathsf{k}(\mathbf{x})}(g)$, *then* $\dim g \geq \dim f$.

As a matter of terminology, we note that when $f(\mathbf{x}) \in D_{\mathsf{k}(\mathbf{x})}(g)$, we say that g *dominates* f.

21. Quaternion algebras

21.1. **Construction and first properties.** We now undertake the study of quaternion algebras over a field k, which are certain 4-dimensional associative, noncommutative algebras. These algebras are equipped with a norm which is a 2-fold Pfister form, and we will freely use our knowledge about such objects to deduce theorems.

Some comments on k-algebras are in order before we get started in earnest. A k-algebra A is a k-vector space equipped with a product $A \times A \to A$ which is k-bilinear. This makes A into a ring, but A need not be commutative. Given a set of generators S, we are entitled to talk about the free k-algebra $\mathsf{k}\langle S \rangle$ generated by S. It has basis consisting of words in S with product extended bilinearly from concatenation of words. (These are sometimes referred to as non-commuting polynomial algebras.) We can then specify an algebra by generators and relations (a presentation) by setting a generating set S and relations $R \subseteq \mathsf{k}\langle S \rangle$. If (R) denotes the *two-sided* ideal generated by R, then $\mathsf{k}\langle S \rangle / (R)$ is the k-algebra on generators S with relations R.

A homomorphism of k-algebras $f \colon A \to B$ is a k-linear map which respects products. The free k-algebra $\mathsf{k}\langle S \rangle$ satisfies the obvious universal property with respect to such maps, and we have a k-algebra homomorphism $\tilde{g} \colon \mathsf{k}\langle S \rangle / (R) \to B$ induced by a function $g \colon S \to B$ if and only if $g(r) = 0$ for all $r \in R$.

Let $a, b \in \mathsf{k}^\times$. Then the *quaternion algebra* $A = \left(\frac{a,b}{\mathsf{k}}\right)$ is the k-algebra with generators i and j defined by the relations

$$i^2 = a, \quad j^2 = b, \quad ij = -ji.$$

We define $k = ij \in A$. Then $k^2 = (ij)(ij) = -i^2j^2 = -ab$, and

$$ik = -ki = aj, \quad kj = -jk = bi.$$

It follows that any two of the elements i, j, k anticommute.

Remark 21.1. The reader may be most familiar with the case $a = b = -1$, $\mathsf{k} = \mathbb{R}$. In this case, $\left(\frac{-1,-1}{\mathbb{R}}\right)$ is a division algebra — it satisfies all the field properties except for commutativity of multiplication. It is not the case, though, that all quaternion algebras are division algebras.

Proposition 21.2. The set $\{1, i, j, k\}$ form a k-basis of $\left(\frac{a,b}{\mathsf{k}}\right)$.

Proof. Let $L = \mathsf{k}(\alpha, \beta)$ where $\alpha^2 = -a$ and $\beta^2 = b$, and let

$$i_0 = \begin{pmatrix} 0 & \alpha \\ -\alpha & 0 \end{pmatrix}, \quad j_0 = \begin{pmatrix} 0 & \beta \\ \beta & 0 \end{pmatrix}$$

in $M_{2\times 2}(L)$. We may compute

$$i_0^2 = aI, \quad j_0^2 = bI, \quad i_0 j_0 = \begin{pmatrix} \alpha\beta & 0 \\ 0 & -\alpha\beta \end{pmatrix} = -j_0 i_0.$$

As such, there is a k-algebra homomorphism $\varphi\colon \left(\frac{a,b}{\mathsf{k}}\right) \to M_{2\times 2}(L)$ with $\varphi(i) = i_0$ and $\varphi(j) = j_0$. Since $\{I, i_0, j_0, i_0 j_0\}$ are linearly independent over L, it follows that $\{1, i, j, ij\}$ are linearly independent over k. These elements clearly span $\left(\frac{a,b}{\mathsf{k}}\right)$, so we are done. □

In the following proposition, $\cong$ denotes isomorphism of k-algebras; two k-algebras A and B are isomorphic when there is a bijective k-algebra homomorphism $A \to B$.

Proposition 21.3. (a) $\left(\frac{a,b}{\mathsf{k}}\right) \cong \left(\frac{ax^2,by^2}{\mathsf{k}}\right)$ for all $a, b, x, y \in \mathsf{k}^\times$.

(b) $\left(\frac{-1,1}{\mathsf{k}}\right) \cong M_{2\times 2}(\mathsf{k})$.

Proof. (a) Let $A = \left(\frac{a,b}{\mathsf{k}}\right)$ with the usual basis $\{1, i, j, k\}$, and let $A' = \left(\frac{ax^2,by^2}{\mathsf{k}}\right)$ with basis $\{1, i', j', k'\}$ such that $i'^2 = ax^2$, $j'^2 = by^2$, and $i'j' = -j'i'$. By direct computation in A, we have

$$(xi)^2 = x^2 i^2 = ax^2, \quad (yj)^2 = y^2 j^2 = by^2, \quad (xi)(yj) = xy(ij) = -xy(ji) = -(yj)(xi).$$

It follows that there is a k-algebra isomorphism $\varphi\colon A' \to A$ such that $\varphi(i') = i$ and $\varphi(j') = j$.

(b) With $a = -1$ and $b = 1$, we may take $\alpha = \beta = 1 \in \mathsf{k}$ in the proof of Proposition 21.2. This induces a k-algebra isomorphism $\varphi\colon \left(\frac{-1,1}{\mathsf{k}}\right) \to M_{2\times 2}(\mathsf{k})$ with

$$\varphi(i) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad \varphi(j) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

□

By Proposition 21.3(b), we see that 2×2 matrix algebras play a special role in the theory of quaternion algebras. Combined with part (a), we see that when $-1//a$ and $1/b$ have square roots in k, $\left(\frac{a,b}{\mathsf{k}}\right)$ is a matrix algebra. (Of course, this is equivalent to $-a$ and b having square roots.) It is also the case that for a field extension $L \supseteq \mathsf{k}$, we have $L \otimes_{\mathsf{k}} \left(\frac{a,b}{\mathsf{k}}\right) \cong \left(\frac{a,b}{L}\right)$. Thus by extending scalars to $L = \mathsf{k}(\sqrt{-a}, \sqrt{b})$, we get $\left(\frac{a,b}{L}\right) \cong M_{2\times 2}(L)$.

The *center* of a k-algebra A is $Z(A) = \{z \in A \mid xz = zx \text{ for all } x \in A\}$. We have $Z(M_{2\times 2}(\mathsf{k})) = \mathsf{k}I \cong \mathsf{k}$, the scalar matrices. This observation is crucial to the following proof.

Proposition 21.4. The center of $\left(\frac{a,b}{\mathsf{k}}\right)$ is $\mathsf{k} \cdot 1 \cong \mathsf{k}$.

Proof. Let $L = \mathsf{k}(\alpha, \beta)$ with $\alpha^2 = -a$, $\beta^2 = b$. Then

$$L \otimes_{\mathsf{k}} \left(\frac{a,b}{\mathsf{k}}\right) \cong \left(\frac{a,b}{L}\right) \cong M_{2\times 2}(L)$$

which has center L. It follows that the center of $\left(\frac{a,b}{\mathsf{k}}\right)$ is k. □

A matrix algebra $M_{2\times 2}(\mathsf{k})$ is also *simple*. This means that it has no nontrivial ideals. Indeed, suppose that $I \subseteq M_{2\times 2}(\mathsf{k})$ is an ideal containing a matrix M with a nonzero entry a_{ij} in the (i,j) position. Let $E_{k\ell}$ denote the matrix with 1 in the (k,ℓ) position and 0's elsewhere, and note that

$$a_{ij} E_{k\ell} = E_{ki} M E_{j\ell} \in I.$$

Thus

$$\begin{pmatrix} \lambda/a_{ij} & 0 \\ 0 & \lambda/a_{ij} \end{pmatrix} a_{ij} E_{k\ell} = \lambda E_{k\ell} \in I$$

for all $\lambda \in \mathsf{k}$ and $1 \leq k, \ell \leq 2$. Taking sums of such matrices, we see that $I = M_{2\times 2}(\mathsf{k})$. This entire argument generalizes to show that $M_{n\times n}(R)$ is simple whenever R is a division ring.[19]

Proposition 21.5. Every quaternion algebrea is simple.

Proof. Let $L = \mathsf{k}(\alpha, \beta)$ for $\alpha^2 = -a$, $\beta^2 = b$. If $I \subseteq \left(\frac{a,b}{\mathsf{k}}\right)$ is a nontrivial ideal, then $L \otimes_{\mathsf{k}} I$ is a nontrivial ideal of $\left(\frac{a,b}{L}\right) \cong M_{2\times 2}(L)$ which is simple. Since $M_{2\times 2}(L)$ is simple, I must be trivial. □

The quaternions with "no real part" play a special role in the theory.

Definition 21.6. A quaternion $\alpha + \beta i + \gamma j + \delta k \in A = \left(\frac{a,b}{\mathsf{k}}\right)$ is called *pure* if $\alpha = 0$. The k-vector space of pure quaternions in A is denoted A_0.

Remark 21.7. Any endomorphism of a finite-dimensional simple algebra necessarily has trivial kernel and is thus an automorphism.

Proposition 21.8. Let $0 \neq v \in A = \left(\frac{a,b}{\mathsf{k}}\right)$. Then $v \in A_0$ if and only if $v \notin \mathsf{k}$ and $v^2 \in \mathsf{k}$.

Proof. If $v = \alpha + \beta i + \gamma j + \delta k$, then

$$v^2 = (\alpha^2 + a\beta^2 + b\gamma^2 - ab\delta^2) + 2\alpha(\beta i + \gamma j + \delta k)$$

by direct computation. Thus if $v \in A_0$, then $v^2 = \alpha^2 + a\beta^2 + b\gamma^2 - ab\delta^2 \in \mathsf{k}$. If $v \notin \mathsf{k}$ and $v^2 \in \mathsf{k}$, then the above equation implies $\alpha = 0$, so $v \in A_0$. □

Corollary 21.9. If A and A' are quaternion algebras over k and $\varphi\colon A \to A'$ is a k-algebra isomorphism, then $\varphi(A_0) = A'_0$.

We conclude this section by making some statements about the original quaternion algebra, $\mathbb{H} = \left(\frac{-1,-1}{\mathbb{R}}\right)$.

[19] Wedderburn's theorem says that all finite-dimensional simple k-algebras are of this form.

Proposition 21.10. The quaternion algebra $\mathbb{H}$ is isomorphic to the $\mathbb{R}$-subalgebra of $M_{2\times 2}(\mathbb{C})$ consisting of matrices of the form $\begin{pmatrix} \alpha & -\overline{\beta} \\ \beta & \overline{\alpha} \end{pmatrix}$ where $\alpha, \beta \in \mathbb{C}$. The group of unit quaternions

$$U = \{x + yi + zj + wk \mid x^2 + y^2 + z^2 + w^2 = 1\} \leq \mathbb{H}^\times$$

is isomorphic to the special unitary group $\mathrm{SU}(2)$.

The first identification is derived by considering how expressions of the form $\alpha + j\beta$ multiply when $\alpha, \beta \in \mathbb{C}$. The group $\mathrm{SU}(2)$ consists of 2×2 unitary matrices ($AA^* = A^*A = I$ where $(\ )^*$ is conjugate transpose) with determinant 1. This identifies the group structure on the 3-sphere $S^3 = U$.

21.2. **Norm forms of quaternion algebras.** The reader will recall that the complex numbers have an automorphism given by complex conjugation: $z \mapsto \overline{z}$, where $\overline{a + bi} = a - bi$ for $a, b \in \mathbb{R}$. The quaternion algebra $A = \left(\frac{a,b}{\mathsf{k}}\right)$ supports a similar operation which we call the *bar involution*. If $z = \alpha + \beta i + \gamma j + \delta k$ for $\alpha, \beta, \gamma, \delta \in \mathsf{k}$, we define

$$\overline{z} = \alpha - (\beta i + \gamma j + \delta k).$$

We may then compute

$$\overline{z + w} = \overline{z} + \overline{w}, \quad \overline{zw} = \overline{w}\,\overline{z}, \quad \overline{\overline{z}} = z, \quad \text{and} \quad \overline{\lambda z} = \lambda \overline{z}$$

for $z, w \in A$ and $\lambda \in \mathsf{k}$. This means that $\overline{(\ )} \colon A \to A$ is an *anti-automorphism* of order 2.

Definition 21.11. The *norm* and *trace* maps on $A = \left(\frac{a,b}{\mathsf{k}}\right)$ are defined as

$$\begin{aligned} \mathrm{N} \colon A &\longrightarrow \mathsf{k} \quad \text{and} \quad \mathrm{Tr} \colon A \longrightarrow \mathsf{k} \\ z &\longmapsto z\overline{z} \qquad\qquad\quad z \longmapsto z + \overline{z}. \end{aligned}$$

Remark 21.12. In order to check that the codomains are accurate, one may check that $\overline{\mathrm{N}\, z} = \mathrm{N}\, z$ and $\overline{\mathrm{Tr}\, z} = \mathrm{Tr}\, z$.

Now consider the symmetric bilinear form

$$\begin{aligned} B \colon A \times A &\longrightarrow \mathsf{k} \\ (z, w) &\longmapsto \frac{1}{2}(z\overline{w} + w\overline{z}) = \frac{1}{2}\,\mathrm{Tr}(z\overline{w}). \end{aligned}$$

The associated quadratic map takes $z \mapsto B(z, z) = \frac{1}{2}\,\mathrm{Tr}(z\overline{z}) = z\overline{z} = \mathrm{N}\, z$, so we conclude that N is a quadratic map on A. We may thus refer to N as the *norm form* of A.

We leave it to the reader to produce a proof of the following proposition (presumably by direct computation and basic properties of Tr).

Proposition 21.13. The quadratic space (A, B) has orthogonal basis $\{1, i, j, k\}$ and associated diagonalization

$$\langle 1, -a, -b, ab\rangle = \langle 1, -a\rangle \otimes \langle 1, -b\rangle = \langle\langle -a, -b\rangle\rangle\,.$$

Remark 21.14. It follows that if $z = \alpha + \beta i + \gamma j + \delta k$ with $\alpha, \beta, \gamma, \delta \in \mathsf{k}$, then

$$\mathrm{N}\, z = \alpha^2 - \beta^2 a - \gamma^2 b + \delta^2 ab.$$

Thus $\mathrm{N}\,\overline{z} = \mathrm{N}\, z$ for all $z \in A$, whence the bar involution is an isometry of the quadratic space A.

Also note that $\mathbb{H} = \left(\frac{-1,-1}{\mathbb{R}}\right)$, we get

$$\mathrm{N}(\alpha + \beta i + \gamma j + \delta k) = \alpha^2 + \beta^2 + \gamma^2 + \delta^2$$

the (squared) Euclidean norm on $\mathbb{R}^4$.

Proposition 21.15. (a) For $z, w \in A = \left(\frac{a,b}{\mathsf{k}}\right)$, $\mathrm{N}(zw) = (\mathrm{N}\, z)(\mathrm{N}\, w)$.

(b) A quaternion $z \in A$ is a unit if and only if $\mathrm{N}\, z \neq 0$ (*i.e.*, z is anisotropic in the quadratic space A).

Proof. (a) We compute

$$\mathrm{N}(zw) = zw\overline{zw} = z(w\overline{w})\overline{z} = (\mathrm{N}\, w)z\overline{z} = (\mathrm{N}\, z)(\mathrm{N}\, w).$$

(b) If z^{-1} exists, then $1 = \mathrm{N}(1) = \mathrm{N}(zz^{-1}) = (\mathrm{N}\, z)(\mathrm{N}\, z^{-1})$, so $\mathrm{N}\, z \neq 0$. Conversely, if $\mathrm{N}\, z \neq 0$, then the equation $z\overline{z} = \mathrm{N}(z) \cdot 1$ implies that $z^{-1} = (1/\mathrm{N}\, z) \cdot \overline{z}$.

□

Remark 21.16. From Proposition 21.15(a), we directly see that N is a group form over k. Of course, given Theorem 18.11 we already knew this since N is a Pfister form. Since N is strongly multiplicative by Theorem 19.6, we also have that $\lambda \in D_{\mathsf{k}}(\mathrm{N})$ if and only if $\langle\lambda\rangle \otimes \mathrm{N} \cong \mathrm{N}$.

We can now state and prove an important theorem that classifies quaternion algebras via their norm forms.

Theorem 21.17. *For $A = \left(\frac{a,b}{\mathsf{k}}\right)$ and $A' = \left(\frac{c,d}{\mathsf{k}}\right)$ with norm forms N and N′, respectively, the following statements are equivalent:*

(a) A is isomorphic to A' as k-algebras,
(b) N is isometric to N′, and
(c) A_0 is isometric to A_0' as quadratic spaces.

Proof. The equivalence (b) $\iff$ (c) is clear from Witt's Cancellation Theorem 8.2. We now show (a) $\implies$ (b). Suppose that $\varphi\colon A \cong A'$ as k-algebras. We must show that φ is an isometry $\mathrm{N} \cong \mathrm{N}'$. By Corollary 21.9, $\varphi(A_0) = A_0'$. Write $z = \alpha + z_0$ for $\alpha \in \mathsf{k}$ and $z_0 \in A_0$. Then $\overline{z} = \alpha - z_0$, $\varphi(z) = \alpha + \varphi(z_0)$, and $\varphi(\overline{z}) = \alpha - \varphi(z_0)$. Since $\varphi(z_0) \in A_0'$, we have $\overline{\varphi(z)} = \varphi(\overline{z})$. Thus

$$\mathrm{N}'\,\varphi(z) = \varphi(z) \cdot \overline{\varphi(z)} = \varphi(z) \cdot \varphi(\overline{z}) = \varphi(z\overline{z}) = \varphi(\mathrm{N}\, z) = \mathrm{N}\, z$$

as desired.

Finally, we show that (c) $\implies$ (a). Assume that σ is an isometry $\sigma\colon (A_0, \mathrm{N}\,|_{A_0}) \cong (A_0', \mathrm{N}'\,|_{A_0'})$. Then

$$\mathrm{N}'\,\sigma(i) = \mathrm{N}\, i = -a \quad \text{and} \quad \mathrm{N}'\,\sigma(i) = \sigma(i)\overline{\sigma(i)} = -\sigma(i)^2$$

because $\overline{z_0} = -z_0$ for $z_0 \in A_0'$. Therefore $\sigma(i)^2 = a$, and similarly $\sigma(j)^2 = b$. Lastly, since $B(i,j) = 0$, we know that $B(\sigma(i), \sigma(j)) = 0$, whence $\sigma(i)\sigma(j) = -\sigma(j)\sigma(i) \in A'$. Taken together, these imply that $A' \cong \left(\frac{a,b}{\mathsf{k}}\right) = A$. □

Corollary 21.18. The quaternion algebras $\left(\frac{a,a}{\mathsf{k}}\right)$ and $\left(\frac{a,-1}{\mathsf{k}}\right)$ are isomorphic.

Proof. The associated norm forms are

$$\langle 1, -a, -a, a^2\rangle \quad \text{and} \quad \langle 1, -a, 1, -a\rangle$$

which are isometric. □

We have already seen that in special circumstances, $\left(\frac{a,b}{\mathsf{k}}\right)$ is isomorphic to the 2×2 matrix algebra $M_{2\times 2}(\mathsf{k})$. When this happens, we call $\left(\frac{a,b}{\mathsf{k}}\right)$ *split* (or *split over* k, more verbosely, a *split quaternion algebra over* k). The following theorem characterizes split quaternion algebras in a number of useful ways.

Theorem 21.19. *For $A = \left(\frac{a,b}{\mathsf{k}}\right)$ with norm form N, the following statements are equivalent:*

(a) A *is split,*
(b) A *is not a division algebra,*
(c) N *is isotropic,*
(d) N *is hyperbolic,*
(e) $\mathrm{N}|_{A_0}$ *is isotropic,*
(f) $(\langle a\rangle - 1)(\langle b\rangle - 1) = 0 \in \mathrm{GW}(\mathsf{k})$ *or* $\mathrm{W}(\mathsf{k})$,
(g) the binary form $\langle a, b\rangle$ *represents* 1,
(h) $a \in \mathrm{N}_{L/\mathsf{k}}(L)$ *where* $L = \mathsf{k}(\sqrt{b})$ *and* $\mathrm{N}_{L/\mathsf{k}} \colon L \to \mathsf{k}$ *is the field norm for the extension* $\mathsf{k} \subseteq L$.

Of course, we can use the above conditions to determine when $A = \left(\frac{a,b}{\mathsf{k}}\right)$ is *not* split, in which case A is a division algebra!

Proof. Since $\left(\frac{1,-1}{\mathsf{k}}\right)$ is split and has norm form $2h$, Theorem 21.17 implies that (a), (d), (f), and (g) are equivalent. By Theorem 18.7, isotropic Pfister forms are hyperbolic, so these are all equivalent to (c) as well.

Next, we show that (d) $\implies$ (e) $\implies$ (c). If (d) holds, then $\mathrm{N}|_{A_0} \cong \langle -1, 1-1\rangle$ is isotropic, so (e) holds. Also then N is isotropic, so (c) holds.

Given Proposition 21.15(b), we also have (a) $\implies$ (b) $\implies$ (c). This proves the equivalence of (a)–(g).

Finally, we show (g) $\iff$ (h). If $b \in \mathsf{k}^{\boxtimes}$, then $L = \mathsf{k}$, $\mathrm{N}_{L/\mathsf{k}} = \mathrm{id}$, and (g) and (h) are clearly equivalent. Assume then that $b \in \mathsf{k}^{\times} \smallsetminus \mathsf{k}^{\boxtimes}$. In this case, $\mathrm{N}_{L/\mathsf{k}} \colon x + y\sqrt{b} \mapsto x^2 - by^2$ for $x, y \in \mathsf{k}$. Thus (h) holds if and only if $a \in D_{\mathsf{k}}(\langle 1, -b\rangle)$. If $a = x^2 - by^2$ and $x \neq 0$, then $1 = a(1/x)^2 + b(y/x)^2$; if $x = 0$, then $\langle a, b\rangle \cong \langle -b, b\rangle \cong h$ which is universal. The converse follows from similar computations. □

The most classical of the above conditions is (f), which says that $1 \in D_{\mathsf{k}}(\langle a, b\rangle)$, *i.e.*, that there exists a solution $(x, y) \in \mathsf{k}^2$ to the equation $ax^2 + by^2 = 1$. This is called the *Hilbert equation*, and is closely related to the notion of a Hilbert symbol, which we will study later.

Corollary 21.20. (a) For all $a \in \mathsf{k}^{\times}$, $\left(\frac{1,a}{\mathsf{k}}\right)$ and $\left(\frac{a,-a}{\mathsf{k}}\right)$ are both split.

(b) For $1 \neq a \in \mathsf{k}^{\times}$, $\left(\frac{a,1-a}{\mathsf{k}}\right)$ is split; this is known as the *Steinberg relation*.

(c) The quaternion algebra $\left(\frac{-1,a}{\mathsf{k}}\right)$ splits if and only if a is a sum of two squares in k.

Proof. For (a) and (b), just note that $\langle 1, a\rangle$, $\langle a, -a\rangle \cong h$, and $\langle a, 1-a\rangle$ all represent 1. For (c), a computation shows that $\langle -1, a\rangle$ represents 1 if and only if $\langle 1, 1\rangle$ represents a. □

Corollary 21.21. If k is a field over which every binary form is universal (*e.g.*, $\mathsf{k} = \mathbb{F}_q$, k quadratically closed, or $\mathsf{k} = F(t)$ for F algebraically closed), then $\left(\frac{a,b}{\mathsf{k}}\right)$ is split for all $a, b \in \mathsf{k}^{\times}$. □

We will now explore some examples that utilize our new theorems.

Example 21.22. If $a, b \in \mathbb{Q}_{<0}$, then $D_{\mathbb{Q}}(\langle a, b\rangle) \subseteq \mathbb{Q}_{<0}$, and thus $\langle a, b\rangle$ does not represent 1. As such, $\left(\frac{a,b}{\mathbb{Q}}\right)$ is nonsplit in this case. This can also be seen by extending scalars to $\mathbb{R}$:

$$\mathbb{R} \otimes_{\mathbb{Q}} \left(\frac{a, b}{\mathbb{Q}}\right) \cong \left(\frac{a, b}{\mathbb{R}}\right) \cong \left(\frac{-1, -1}{\mathbb{R}}\right) = \mathbb{H},$$

which is a division algebra.

Example 21.23. Consider the quaternion algebras $A = \left(\frac{-1,-1}{\mathbb{Q}}\right)$ and $A' = \left(\frac{-2,-3}{\mathbb{Q}}\right)$. The reader may check the isometries

$$\langle 1,1,1,1\rangle \cong \langle 1,1,2,2\rangle \cong \langle 1,3,6,2\rangle \cong \langle 1,2,3,6\rangle .$$

Of course, the first and last forms are the norm forms of A and A', respectively, so $A \cong A'$.

Example 21.24. We can show that the quaternion algebra $A = \left(\frac{5,-3}{\mathbb{Q}}\right)$ is a division algebra, but $L \otimes_{\mathbb{Q}} A$ is split for $L = \mathbb{Q}(\sqrt{17})$. Indeed, A is split if and only there are relatively prime integers x, y, z such that $5x^2 - 3y^2 = z^2$. The reader may check that this is impossible (perhaps by working mod 3). Meanwhile, the equation $5 \cdot 2^2 - 3 = 17$ implies that $\langle 5, -3\rangle$ represents 1 over $L = \mathbb{Q}(\sqrt{17})$, so $L \otimes_{\mathbb{Q}} A$ is split.

22. Local fields

We will now take a brief detour in order to introduce an important class of fields, namely local fields. These include the p-adic rationals $\mathbb{Q}_p$ and function fields over finite fields $\mathbb{F}_q(t)$. This material will link back with quaternion algebras and quadratic forms when we study Hilbert reciprocity in the next section.

Definition 22.1. A *discretely valued field* is a field k equipped with a surjective homomorphism $v\colon \mathsf{k}^\times \to (\mathbb{Z}, +)$ such that $v(a+b) \geq \min\{v(a), v(b)\}$ for $a, b, a+b \in \mathsf{k}^\times$. By convention, we define $v(0) = \infty$. Then

$$\mathscr{O}_v = \{x \in \mathsf{k} \mid v(x) \geq 0\}$$

is a subring of k called the *valuation ring* of (k, v).

Note that the homomorphism condition on v amounts to $v(ab) = v(a) + v(b)$. This, combined with $v(a+b) \geq \min\{v(a), v(b)\}$, makes it clear that $\mathscr{O}_v$ really is a subring.

Example 22.2. (a) The most famous discretely valued field is $\mathbb{Q}$ with the p-adic valuation v_p, p a prime. For $a \in \mathbb{Z}$, we let $v_p(a) = n$ when $a = p^n m$, $p \nmid m$. If $a/b \in \mathbb{Q}$, then $v_p(a/b) = v_p(a) - v_p(b)$. The reader may check that v_p satisfies the properties of a valuation.

(b) Similarly, let $p \in \mathsf{k}[t]$ be an irreducible polynomial. Measuring the divisibility of the numerator and denominator of a rational function over k induces a valuation on $\mathsf{k}(t)$ also denoted v_p.

(c) The field of Laurent series $\mathsf{k}(\!(t)\!)$ also supports a discrete valuation. This consists of series of the form

$$f = \sum_{i=m}^{\infty} \lambda_i t^i$$

where $m \in \mathbb{Z}$ and $\lambda_i \in \mathsf{k}$ under formal addition and multiplication. If $\lambda_m \neq 0$ in the above expression, then we define $v(f) = m$.

Definition 22.3. A commutative ring R is called *local* if it has a unique maximal ideal.

Proposition 22.4. The valuation ring $\mathscr{O}_v$ of a discretely valued field (k, v) is a local ring with unique maximal ideal

$$\mathfrak{m}_v = \{x \in \mathsf{k} \mid v(x) \geq 1\}.$$

Furthermore,

$$\operatorname{Frac} \mathscr{O}_v = \mathsf{k}$$

and

$$\mathscr{O}_v^\times = \{x \in \mathsf{k} \mid v(x) = 0\}.$$

Proof. First note that if $x \in \mathsf{k} \smallsetminus \mathscr{O}_v$, then $v(x) < 0$, whence $0 = v(1) = v(xx^{-1}) = v(x) + v(x^{-1})$ implies that $v(x^{-1}) > 0$, *i.e.*, $x^{-1} \in \mathscr{O}_v$. It follows that $x = 1/x^{-1} \in \operatorname{Frac} \mathscr{O}_v$, and this is enough to show that $\operatorname{Frac} \mathscr{O}_v = \mathsf{k}$.

Next, we show that $\mathscr{O}_v^\times = \{x \in \mathsf{k} \mid v(x) = 0\}$. Indeed, if $x, y \in \mathscr{O}_v$ with $xy = 1$, then applying v we find that $0 = v(x) + v(y)$. Since $v(x), v(y) \geq 0$, it follows that $v(x) = 0$. This implies that $\mathscr{O}_v^\times \subseteq \{x \in \mathsf{k} \mid v(x) = 0\}$. Now suppose that $x \in \mathsf{k}$ has $v(x) = 0$. Then $x \neq 0$ since $v(0) = \infty$, so x has an inverse $x^{-1} \in \mathsf{k}$. The equation $1 = xx^{-1}$ implies that $0 = v(x) + v(x^{-1}) = v(x^{-1})$, so x^{-1} is in fact in $\mathscr{O}_v$. This proves that $\mathscr{O}_v^\times = \{x \in \mathsf{k} \mid v(x) = 0\}$.

We see now that $\mathfrak{m}_v = \mathscr{O}_v \smallsetminus \mathscr{O}_v^\times$ is the set of nonunits in $\mathscr{O}_v$. The reader may check that properties of v imply that $\mathfrak{m}_v$ is an ideal. If $I \subseteq \mathscr{O}_v$ is an ideal properly containing $\mathfrak{m}_v$, then I contains a unit of $\mathscr{O}_v$ and thus $I = \mathscr{O}_v$. This shows that $\mathfrak{m}_v$ is maximal. If $J \subseteq \mathscr{O}_v$ is another maximal ideal, then J contains no units of $\mathscr{O}_v$ and thus $J \subseteq \mathfrak{m}_v$. By maximality of J, $J = \mathfrak{m}_v$, so $\mathfrak{m}_v$ is the unique maximal ideal of the local ring $\mathscr{O}_v$. □

We can say yet more about the structure of $\mathfrak{m}_v$ and the rest of the ideals in $\mathscr{O}_v$.

Definition 22.5. Any element $\pi \in \mathscr{O}_v$ with $v(\pi) = 1$ is called a *uniformizer* of $\mathscr{O}_v$.

Proposition 22.6. If π is a uniformizer of $\mathscr{O}_v$, then

$$\mathfrak{m}_v = (\pi).$$

If π' is another uniformizer of $\mathscr{O}_v$, then $\pi' = u\pi$ for some unit $u \in \mathscr{O}_v^\times$.

Proof. Moral exercise. □

The following proposition specifies all ideals in $\mathscr{O}_v$.

Proposition 22.7. The ring $\mathscr{O}_v$ is a principal ideal domain; in fact, every ideal in $\mathscr{O}_v$ is of the form $\mathfrak{m}_v^n = (\pi^n)$ for some $n \in \mathbb{N}$ and uniformizer π. The full lattice of ideals in $\mathscr{O}_v$ is

$$\mathscr{O}_v \supsetneq \mathfrak{m}_v \supsetneq \mathfrak{m}_v^2 \supsetneq \cdots \supsetneq 0$$

with $\bigcap_i \mathfrak{m}_v^i = 0$.

Proof. Moral exercise. □

Since $\mathfrak{m}_v$ is maximal, we know that $\mathscr{O}_v/\mathfrak{m}_v$ is a field, and it gets a special name.

Definition 22.8. The field $\kappa(v) = \mathscr{O}_v/\mathfrak{m}_v$ is called the *residue field* of (k, v).

Remark 22.9. The study of discrete valuation fields (DVFs) is essentially the same as that of *discrete valuation rings* (DVRs). A DVR is, by definition, a local principal ideal domain. Thus every DVF (k, v) has an associated DVR $\mathscr{O}_v$. Given a DVR A with maximal ideal $\mathfrak{m}$ and fraction field k, we can define a valuation

$$v_{\mathfrak{m}} \colon \mathsf{k} \to \mathbb{Z}$$

which takes $a \in \mathfrak{m}^n$ to $v(a) = n$ for $n \in \mathbb{N}$, and takes $a/b \in \mathsf{k}$ to $v(a) - v(b)$. It is straightforward to check that $\mathscr{O}_{v_{\mathfrak{m}}} = A$.

We may attach the metric

$$\begin{aligned} \mathrm{d}_v \colon \mathsf{k} \times \mathsf{k} &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \exp(-v(x - y)) \end{aligned}$$

to a discretely valued field (k, v).[20] Under this metric, x and y are close together when the valuation of $x - y$ is large, which is equivalent to $x - y \in \mathfrak{m}_v^n$ for n large.

[20] There is nothing special about using the base e exponential — we could replace it with $a^{-v(x-y)}$ for any real number $a > 1$ and we would get an isometric metric space. When working with the p-adic valuation, it is traditional to take $a = p$.

Recall that a metric space (X, d) is called *complete* if every Cauchy sequence in X converges (to an element of X). We call (k, v) a *complete discretely valued field* (or CDVF) if k is complete with respect to d_v. Those familiar with inverse limits may verify the following proposition.

Proposition 22.10. A discrete valuation field (k, v) is complete if and only if the natural map $\mathscr{O}_v \to \lim_n \mathscr{O}_v/\mathfrak{m}_v^n$ is surjective (and hence an isomorphism of rings).

This essentially says that if $x_1, x_2, \ldots$ is a sequence in $\mathscr{O}_v$ with $x_{i+1} \equiv x_i \pmod{\mathfrak{m}_v^i}$ for all i, then $\mathscr{O}_v$ contains an element x such that $x \equiv x_i \pmod{\mathfrak{m}_v^i}$ for all i.

With any luck, you have seen in an analysis class that every metric space has a *completion*. If (X, d) is a metric space, its completion $(\overline{X}, \overline{\mathrm{d}})$ is a metric space containing X as a dense subspace and such that if $f\colon X \to Y$ is a uniformly continuous function, then there is a unique uniformly continuous function $\overline{f}\colon \overline{X} \to Y$ extending f. The completion may be constructed as the set of equivalence classes of Cauchy sequences in X, where if $x = (x_i)$ and $y = (y_i)$ are Cauchy sequences, then $\overline{\mathrm{d}}(x, y) = \lim_i d(x_i, y_i)$; we declare two Cauchy sequences equivalent when $\overline{\mathrm{d}}(x, y) = 0$.

The completion k_v of any discretely valued field (k, v) is in fact a complete discretely valued field. If $x = (x_i)$ and $y = (y_i)$ are equivalence classes of Cauchy sequences in $(\mathsf{k}, \mathrm{d}_v)$, we define their sum and product through coordinatwise addition and multiplication. (The reader may check that this is well-defined.) We recover a valuation by taking $-\log \overline{\mathrm{d}}(x, 0)$, which is automatically complete.

Example 22.11. The completion of $\mathbb{Q}$ with respect to the p-adic valuation v_p (p a prime number) is called $\mathbb{Q}_p$, the p-adic rationals. The valuation ring of $\mathbb{Q}_p$ is denoted $\mathbb{Z}_p$ and is called the p-adic integers. We have $\mathbb{Z}_p = \lim_i \mathbb{Z}/p^i\mathbb{Z}$. We may take p as a uniformizer of $\mathbb{Z}_p$, and $\kappa(v_p) = \mathbb{Z}_p/p\mathbb{Z}_p \cong \mathbb{F}_p$.

Example 22.12. The completion of $\mathsf{k}(t)$ with respect to the t-adic valuation is the field of formal Laurent series $\mathsf{k}(\!(t)\!)$. The valuation ring of $\mathsf{k}(\!(t)\!)$ is $\mathsf{k}[\![t]\!]$, the ring of formal power series, t is a uniformizer, and the residue field is k.

We will now investigate the connections between arithmetic in a complete discretely valued field (k, v) and its residue field $\kappa(v)$. For $x \in \mathscr{O}_v$, let $\overline{x} = x + \mathfrak{m}_v \in \kappa(v) = \mathscr{O}_v/\mathfrak{m}_v$.

Lemma 22.13. Let (k, v) be a CDVF with characteristic not 2. For any $u \in \mathscr{O}_v^\times$, u is a square in k (or $\mathscr{O}_v^\times$) if and only if $\overline{u}$ is a squre in $\kappa(v)$.

Proof. The forwards implication is trivial. For the backwards implication, suppose $\mathfrak{m} = \mathfrak{m}_v$ and suppose that $\overline{u} \in \kappa(v)^{\boxtimes}$. We will construct a sequence (x_i) in $\mathscr{O}_v^\times$ such that

$$x_i^2 \equiv u \pmod{\mathfrak{m}^i} \quad \text{and} \quad x_{i+1} \equiv x_i \pmod{\mathfrak{m}^i}$$

for all $i \geq 1$. If we have such a sequence, then its limit x will satisfy $x^2 - u = \lim_i (x_i^2 - u) = 0$, as desired.

Since $\overline{u}$ is a square in $\kappa(v)$, x_1 exists. For induction, suppose that we have constructed the element x_i as required. Let $x_{i+1} = x_i + \pi^i z$ where π is a uniformizer and z is to be determined. We have $x_i^2 - u = \pi^i y$ for some $y \in \mathscr{O}_v$, so

$$x_{i+1}^2 - u \equiv (x_i + \pi^i z)^2 - u \equiv \pi^i (y + 2x_i z) \pmod{\mathfrak{m}^{i+1}}.$$

Since the characteristic of $\kappa(v)$ is not 2, we have $2x_i \in \mathscr{O}_v^\times$. Set $z = (\pi - y)/(2x_i)$ so that $y + 2x_i z = \pi$. Then $x_{i+1} \in \mathscr{O}_v^\times$, and

$$x_{i+1}^2 \equiv u \pmod{\mathfrak{m}^{i+1}} \quad \text{and} \quad x_{i+1} \equiv x_i \pmod{\mathfrak{m}^i}$$

as needed. □

Corollary 22.14. Under the hypotheses of Lemma 22.13, a nonzero element $u\pi^n$ ($u \in \mathscr{O}_v^\times$, $n \in \mathbb{Z}$) is a square in k if and only if n is even and $\overline{u} \in \kappa(v)^{\boxtimes}$.

Lemma 22.13 enables us to define a group homomorphism $i\colon \kappa(v)^\times/\kappa(v)^{\boxtimes} \to \mathsf{k}^\times/\mathsf{k}^{\boxtimes}$ in the following fashion. Given $\overline{u} \in \kappa(v)^{\boxtimes}$, let u be any lifting of $\overline{u}$ to $\mathscr{O}_v^\times$. If u' is another lifting, then $\overline{u/u'} = 1$, so Corollary 22.14 implies that $u \in u'\mathsf{k}^{\boxtimes}$. The rule $i(\overline{u}) = u\mathsf{k}^{\boxtimes}$ is thus a well-defined homomorphism.

Corollary 22.15. Under the hypotheses of Lemma 22.13, the sequence

$$1 \longrightarrow \kappa(v)^\times/\kappa(v)^{\boxtimes} \xrightarrow{i} \mathsf{k}^\times/\mathsf{k}^{\boxtimes} \xrightarrow{\overline{v}} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0$$

is split exact. (Here $\overline{v}$ is the mod 2 reduction of v.)

Proof. Exactness is clear given Corollary 22.14. The splitting is given by $1 + 2\mathbb{Z} \mapsto \pi\mathsf{k}^{\boxtimes}$. □

We are now ready to compute the Grothendieck–Witt ring of a CDVF (k, v) in terms of the Grothendieck–Witt ring of $\kappa(v)$, at least when $\operatorname{char}\kappa(v) \neq 2$. Since $\mathrm{GW}(\mathbb{F}_q)$ is already known by Corollary 14.11, this will give a complete computation of $\mathrm{GW}(\mathbb{Q}_p)$, $p > 2$, and $\mathrm{GW}(\mathbb{F}_q(\!(t)\!))$, $2 \nmid q$.

Consider the rule $\langle\overline{u}\rangle \mapsto \langle u\rangle$, $u \in \mathscr{O}_v^\times$. We claim that this extends linearly to a well-defined ring homomorphism $\mathrm{GW}(\kappa(v)) \to \mathrm{GW}(\mathsf{k})$. Indeed, we may simply check that the relations Theorem 15.3 are satisfied. Abusing notation, we will also call this map $i\colon \mathrm{GW}(\kappa(v)) \to \mathrm{GW}(\mathsf{k})$.

We now define a second homomorphism $j\colon \mathrm{GW}(\kappa(v)) \to \mathrm{GW}(\mathsf{k})$ as the composite

$$\mathrm{GW}(\kappa(v)) \xrightarrow{i} \mathrm{GW}(\mathsf{k}) \xrightarrow{\langle\pi\rangle\cdot} \mathrm{GW}(\mathsf{k})$$

which is a homomorphism of Abelian groups. We see that $j(h_{\kappa(v)}) = \langle\pi\rangle\, h_{\mathsf{k}} = h_{\mathsf{k}}$, so the pair (i, j) induce a group homomorphism

$$f\colon \mathrm{GW}(\kappa(v)) \oplus \mathrm{GW}(\kappa(v))/\mathbb{Z}\cdot(h, -h) \longrightarrow \mathrm{GW}(\mathsf{k}).$$

Theorem 22.16 (Springer's Theorem)**.** *For any CDVF (k, v) with $\operatorname{char}\kappa(v) \neq 2$, the map f is a group isomorphism.*

This result appears as [Lam05, Theorem VI.1.4] where a full proof is given. The idea is to define an inverse g to f on generators $\langle x\rangle$, $x \in \mathsf{k}^\times$, of $\mathrm{GW}(\mathsf{k})$. The rule is that if $x = u\pi^m$ ($u \in \mathscr{O}_v^\times$ and $m \in \mathbb{Z}$), then

$$g(x) = \begin{cases} (\langle\overline{u}\rangle, 0) & \text{if } m \text{ is even,} \\ (0, \langle\overline{u}\rangle) & \text{if } m \text{ is odd.} \end{cases}$$

With some effort, one then shows that this assignment respects the relations in the presentation of $\mathrm{GW}(\mathsf{k})$ as an Abelian group. This establishes g as a well-defined group homomorphism, and it is clearly a two-sided inverse to f.

We also note the following immediate corollary of Theorem 22.16.

Corollary 22.17. For any CDVF (k, v) with $\operatorname{char}\kappa(v) \neq 2$, the pair (i, j) induces a group isomorphism

$$\mathrm{W}(\kappa(v)) \oplus \mathrm{W}(\kappa(v)) \cong \mathrm{W}(\mathsf{k}).$$

We now specialize our discussion from complete discretely valued fields to local fields.

Definition 22.18. A CDVF (k, v) is called a *local field* if the residue field $\kappa(v)$ is finite. In this case, the cardinality $q = |\kappa(v)|$ is called the *residue order* of (k, v). If $2 \mid q$ we call k a *dyadic* local field, and if $2 \nmid q$, we call k a *nondyadic* local field.

Remark 22.19. Some authors call this class of fields the *non-Archimedean local fields*. The *Archimedean local fields* are $\mathbb{R}$ and $\mathbb{C}$. In these notes, we will always use the term *local field* to refer to the non-Archimedean ones.

Remark 22.20. Local fields have been classified and fall into two camps: equicharacteristic and mixed characteristic. In the equicharacteristic case, $\operatorname{char} \mathsf{k} = \operatorname{char} \kappa(v)$, and it turns out that $\mathsf{k} \cong \mathbb{F}_q(\!(t)\!)$ with the t-adic valuation.

In the mixed characteristic case, $\operatorname{char} \mathsf{k} = 0$ and $\operatorname{char} \kappa(v) = p$ for p the unique prime divisor of the residue field. It turns out that such a k is a finite algebraic extension of $\mathbb{Q}_p$, carrying the unique extension of v_p. Such fields are often called *p-adic fields*.

We have the following omnibus theorem on square classes, quaternion algebras, and (Grothendieck–)Witt rings of local fields.

Theorem 22.21. *Let k be a nondyadic local field with uniformizer π, and let $u \in \mathscr{O}_v^\times$ be such that $\overline{u} \notin \kappa(v)^{\boxtimes}$. Then*

(a) $\mathsf{k}^\times/\mathsf{k}^{\boxtimes}$ *has order* 4 *with square classes represented by* $1, u, \pi, u\pi$,
(b) every quadratic form over k *of dimension* ≥ 5 *is isotropic,*
(c) there is a unique 4*-dimensional anisotropic quadratic form over* k*, namely*

$$\varphi_{\mathsf{k}} = \langle 1, -u, -\pi, u\pi \rangle = \langle\!\langle -u, -\pi \rangle\!\rangle \,,$$

(d) there is a unique nonsplit quaternion algebra over k*, namely* $\left(\frac{\pi,u}{\mathsf{k}}\right)$,
(e) if the residue order $q \equiv 1 \pmod 4$*, then*

$$\mathrm{GW}(\mathsf{k}) \cong \mathbb{Z} \oplus (\mathbb{Z}/2\mathbb{Z})^3$$

additively and $\mathrm{W}(\mathsf{k}) \cong \mathbb{F}_2[C_2 \times C_2]$ *as a ring,*
(f) if $q \equiv 3 \pmod 4$*, then*

$$\mathrm{GW}(\mathsf{k}) \cong \mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$$

additively and $\mathrm{W}(\mathsf{k}) \cong (\mathbb{Z}/4\mathbb{Z})[C_2]$ *as a ring.*

Proof. Part (a) follows from Corollary 22.15. For (b), note that Corollary 22.17 implies that for any CDVF (k, v), if every quadratic form of dimension $n+1$ over $\kappa(v)$ is isotropic, then every quadratic form of dimension $2n+1$ over k is isotropic. Since every 3-dimensional quadratic form is isotropic over $\mathbb{F}_q$, (b) follows. Part (c) can be similarly deduced from Springer's Theorem 22.16, using the fact that $\langle 1, -\overline{u} \rangle$ is the unique anisotropic form over $\mathbb{F}_q$. With (c) established, (d) immediately follows by Theorem 21.19.

The proofs of (e) and (f) follow from Theorem 22.16 and Corollary 22.17 and our determination of $\mathrm{GW}(\mathbb{F}_q)$ in Corollary 14.11. The details are all worked out in [Lam05, Theorem VI.2.2]. □

Remark 22.22. The reader is encourage to imagine several of the many immediate consequences of this theorem. For instance, there are exactly 16 isometry classes of anisotropic forms over a local field!

Notably, the theorem above did not cover the case of dyadic local fields. This is the moment at which we must confront our sins and recognize the importance of characteristic 2 fields: the structure of the Grothendieck–Witt ring of a 2-adic field depends on quadratic theory over is characteristic 2 residue field. These notes will not rehearse the intricate steps needed to recover our balance. The reader is encouraged to look near Theorem VI.2.10 in [Lam05] if they wish to see why the following statement is true.

Theorem 22.23. *For an arbitrary local field* k*, there exists* $u \in \mathscr{O}_v^\times$ *such that* $F = \mathsf{k}(\sqrt{u})$ *is unramified*[21] *(and the square class of* u *in* k *is uniquely determined); furthermore, there is a unique nonsplit quaternion algebra over* k*, namely* $\left(\frac{\pi,u}{\mathsf{k}}\right)$.

The crucial takeaway is that in both the dyadic and nondyadic cases, there is only one quaternion algebra over a local field which is a division algebra.

23. Hilbert reciprocity

In the last section, we emphasized that there are only two isomorphism classes of quaternion algebras over a local field: split and nonsplit. The same is true over the real numbers. Indeed, for $a, b \in \mathbb{R}^\times$, $\langle a, b\rangle$ represents 1 if and only if one or both of a and b are positive. By Theorem 21.19, this implies that $\left(\frac{a,b}{\mathbb{R}}\right)$ is split if and only if at least one of a, b is positive. If $a, b < 0$, then $\langle\langle -a, -b\rangle\rangle \cong 4\langle 1\rangle$, and it follows that $\left(\frac{a,b}{\mathbb{R}}\right) \cong \left(\frac{-1,-1}{\mathbb{R}}\right) = \mathbb{H}$. These observations allow us to make the following definition.

Definition 23.1. For k a (non-Archimedean) local field or $\mathbb{R}$, the *Hilbert symbol* of k is the function

$$(\ ,\ )_\mathsf{k} \colon \mathsf{k}^\times/\mathsf{k}^{\boxtimes} \times \mathsf{k}^\times/\mathsf{k}^{\boxtimes} \to \{\pm 1\}$$

taking values

$$(a,b)_\mathsf{k} = \begin{cases} 1 & \text{if } \left(\frac{a,b}{\mathsf{k}}\right) \text{ is split,} \\ -1 & \text{if } \left(\frac{a,b}{\mathsf{k}}\right) \text{ is nonsplit.} \end{cases}$$

If $\mathsf{k} = \mathbb{Q}_p$, we write $(\ ,\ )_p = (\ ,\ )_{\mathbb{Q}_p}$, and if $\mathsf{k} = \mathbb{R}$, we write $(\ ,\ )_\infty = (\ ,\ )_\mathbb{R}$.

Note that the Hilbert symbol determines the isomorphism class of $\left(\frac{a,b}{\mathsf{k}}\right)$ over such fields. When convenient, we will also consider the Steinberg symbol as a map $\mathsf{k}^\times \times \mathsf{k}^\times \to \{\pm 1\}$.

Proposition 23.2. The Hilbert symbol is symmetric, bimultiplicative, and satisfies the Steinberg relation. In equations, these properties say that

- $(a,b)_\mathsf{k} = (b,a)_\mathsf{k}$,
- $(ab,c)_\mathsf{k} = (a,c)_\mathsf{k}(b,c)_\mathsf{k}$ and $(a,bc)_\mathsf{k} = (a,b)_\mathsf{k}(a,c)_\mathsf{k}$, and
- $(a, 1-a)_\mathsf{k} = 1$ for $1 \neq a \in \mathsf{k}^\times$.

Proof. Symmetry is obvious and we have already proven the Steinberg relation in Corollary 21.20(b). We leave bimultiplicativity as a (difficult?) exercise for the reader. □

Remark 23.3. More is true: the Hilbert symbol is nondegenerate as well, meaning that if $(a,b)_\mathsf{k} = 1$ for all $a \in \mathsf{k}^\times$, then $b \in \mathsf{k}^{\boxtimes}$.

Our primary goal in this section is to sketch a proof of Hilbert reciprocity, which relates the Hilbert symbols for $\mathbb{Q}_p$ and $\mathbb{R}$.

Theorem 23.4 (Hilbert reciprocity)**.** *Let* $P = \{\mathbb{R}, \mathbb{Q}_2, \mathbb{Q}_3, \mathbb{Q}_5, \ldots\}$*. For all* $a, b \in \mathbb{Q}^\times$*,*

$$\prod_{\mathsf{k}\in P} (a,b)_\mathsf{k} = 1.$$

In other words, the set $\Lambda = \{\mathsf{k} \in P \mid (a,b)_\mathsf{k} = -1\}$ *is finite with even cardinality.*

[21] Ramification is an important topic in local field theory that we will not touch on. There are several equivalent conditions for an extension $\mathsf{k} \subseteq L$ to be unramified. Perhaps most convenient in this setting would be that a uniformizer π of k remains a uniformizer in L; alternatively (and equivalently), $[L : \mathsf{k}]$ is equal to the index of the residue fields.

Our proof will pass through the Witt ring of $\mathbb{Q}$. While we do not have time to present a proof of the following theorem, we have developed enough material to make its statement quite precise. Let $\partial_\infty\colon \mathrm{W}(\mathbb{Q}) \to \mathbb{Z}$ be the signature homomorphism. Let $\partial_2\colon \mathrm{W}(\mathbb{Q}) \to \mathbb{Z}/2\mathbb{Z}$ be the map $q \mapsto v_2(\det q) + 2\mathbb{Z}$ where v_2 is the 2-adic valuation. For $p > 2$ prime, let $\partial_p\colon \mathrm{W}(\mathbb{Q}) \to \mathrm{W}(\mathbb{F}_p)$ be the composite $\mathrm{W}(\mathbb{Q}) \to \mathrm{W}(\mathbb{Q}_p) \to \mathrm{W}(\mathbb{F}_p)$ where the map $\mathrm{W}(\mathbb{Q}_p) \to \mathrm{W}(\mathbb{F}_p)$ is the *second residue homomorphism*: the unique homomorphism taking $\langle u\rangle \mapsto 0$ and $\langle pu\rangle \mapsto \langle \overline{u}\rangle$ for $u \in \mathbb{Z}_p^\times$. (This is essentially projection onto the second factor in Springer's Theorem 22.16.) Finally, let $\Omega = \{\infty, 2, 3, 5, 7, 11, \ldots\}$ and let

$$\partial = \bigoplus_{p\in\Omega} \partial_p\colon\ \mathrm{W}(\mathbb{Q}) \to \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \bigoplus_{p\neq 2,\infty} \mathrm{W}(\mathbb{Q}_p).$$

Theorem 23.5 ([Lam05, VI.4])**.** *The map ∂ is an isomorphism of Abelian groups.*

Since $\mathrm{W}(\mathbb{Q})$ splits as a direct sum, we know that every homomorphism $\chi\colon\ \mathrm{W}(\mathbb{Q}) \to A$ to an Abelian group is uniquely determined by homomorphisms

$$\chi_\infty\colon \mathbb{Z} \to A, \quad \chi_2\colon \mathbb{Z}/2\mathbb{Z} \to A, \quad \chi_p\colon\ \mathrm{W}(\mathbb{F}_p) \to A$$

such that $\chi = \sum_{p\in\Omega} \chi_p \circ \partial_p$. The determination of the maps χ_p, $p \in \Omega$, for a given homomomorphism χ is what W. Scharlau [Sch72] calls the *reciprocity law for* χ. We will see that Hilbert reciprocity follows from Scharlau reciprocity for a particular χ applied to the norm forms $\langle\langle -a, -b\rangle\rangle$. In order to proceed, we need to know a littlbe bit about the Witt ring of $\mathbb{Q}_2$.

Proposition 23.6. As a ring,

$$\mathrm{W}(\mathbb{Q}_2) \cong \mathbb{Z}/8\mathbb{Z}[s,t]/(s^2, t^2, 2s, 2t, st - 4).$$

Additive generators for $\mathrm{W}(\mathbb{Q}_2)$ may be taken as $\langle 1\rangle$, $\langle 1, -2\rangle$, and $\langle 1, -5\rangle$, and these map to 1, s, and t, respectively, in the above isomorphism.

Proof. See [Lam05, Theorem VI.2.29 and Remark VI.2.31]. □

In order to deduce Hilbert reciprocity, we henceforth fix χ to be the composite homomorphism

$$\mathrm{W}(\mathbb{Q}) \xrightarrow{\mathrm{res}_{\mathbb{Q}}^{\mathbb{Q}_2}} \mathrm{W}(\mathbb{Q}_2) \xrightarrow{\eta} \mathbb{Z}/8\mathbb{Z},$$

where η is determined by the rules

$$\eta(\langle 1\rangle) = 1 + 8\mathbb{Z}, \quad \eta(\langle 1, -2\rangle) = 0 + 8\mathbb{Z}, \quad \eta(\langle 1, -5\rangle) = 4 + 8\mathbb{Z}.$$

The following theorem expresses Scharlau reciprocity for this particular χ (*i.e.*, it determines the corresponding χ_p for each $p \in \Omega$).

Theorem 23.7. *For $\chi = \eta \circ \mathrm{res}_{\mathbb{Q}}^{\mathbb{Q}_2}$, we have*

$$\begin{aligned}
\chi_\infty(n) &= n + 8\mathbb{Z},\\
\chi_2 &= 0,\\
\chi_p(\langle 1\rangle) &= p - 1 + 8\mathbb{Z},\\
\chi_p(\varphi_p) &= 4 + 8\mathbb{Z}
\end{aligned}$$

for $p \neq 2, \infty$ and φ_p the unique binary anisotropic form over $\mathbb{F}_p$.

Before proving this theorem, we will see how it implies Hilbert reciprocity.

Proof of Theorem 23.4. Fix $a, b \in \mathbb{Q}^\times$ and let $\Lambda = \{\mathsf{k} \in P \mid (a,b)_\mathsf{k} = -1\}$ where $P = \{\mathbb{R}, \mathbb{Q}_2, \mathbb{Q}_3, \ldots\}$. We aim to show that $|\Lambda|$ is even and finite.

Let $q = \langle\!\langle -a, -b \rangle\!\rangle$ denote the norm form of $\left(\frac{a,b}{\mathsf{k}}\right)$, and let $p \in \Omega \smallsetminus \{2, \infty\}$. Then

$$\begin{aligned} p \in \Lambda &\iff q \text{ is anisotropic over } \mathbb{Q}_p \\ &\iff \partial_p(q) = \varphi_p \\ &\iff \chi_p(\partial_p(q)) = 4 + 8\mathbb{Z}. \end{aligned}$$

Similarly,

$$\begin{aligned} \infty \in \Lambda &\iff q \cong \langle\!\langle -1, 1 \rangle\!\rangle \text{ over } \mathbb{R} \\ &\iff \chi_\infty(\partial_\infty(q)) = 4 + 8\mathbb{Z}. \end{aligned}$$

Since $\chi_2 = 0$ and $\chi = \sum_{p \in \Omega} \chi_p \circ \partial_p$, it follows that

$$\chi(q) = 4|\Lambda \smallsetminus \{2\}| + 8\mathbb{Z}. \tag{23.8}$$

The final ingredient is to figure out what happens at $p = 2$. If q is isotropic over $\mathbb{Q}_2$, then $q_{\mathbb{Q}_2} = 0 \in \mathrm{W}(\mathbb{Q}_2)$ and $\chi(q) = 0 + 8\mathbb{Z}$. Thus (23.8) implies that $|\Lambda| = |\Lambda \smallsetminus \{2\}|$ is even. If q is anisotropic over $\mathbb{Q}_2$, then $\chi(q) = \eta(4\langle 1\rangle) = 4 + 8\mathbb{Z}$ and $2 \in \Lambda$. Again (23.8) implies that $|\Lambda|$ must be even, and this completes the proof. □

To truly complete the proof of Hilbert reciprocity, we must verify the Scharlau reciprocity formulæ for χ.

Proof sketch for Theorem 23.7. The formula for χ_∞ is clear by its definition. In order to determine χ_2, consider that $\chi(\langle -1, 2\rangle) = 8\mathbb{Z}$, while $\partial_p(\langle -1, 2\rangle) = 0$ for $2 \neq p \in \Omega$ and $\partial_2(\langle -1, 2\rangle) = v_2(-2) = 1 + 2\mathbb{Z}$. It follows that $\chi_2 = 0$.

Now consider $p \in \Omega \smallsetminus \{\infty, 2\}$. Note that $\partial_r(\langle -1, p\rangle) = 0$ for $r \neq p$ and $\partial_p(\langle -1, p\rangle) = \langle 1\rangle$. Thus $\chi_p(\langle 1\rangle) = \chi(\langle -1, p\rangle) = \chi(\langle p\rangle) - 1 = p - 1 + 8\mathbb{Z}$, where the last identity follows from the structure of $\mathrm{W}(\mathbb{Q}_2)$.

It remains to calculate $\chi_p(\varphi_p)$ for $p > 2$. If $p \equiv 3, 7 \pmod 8$, then $\varphi_p = \langle 1, 1\rangle$ and thus

$$\chi_p(\varphi_p) = 2\chi_p(\langle 1\rangle) = 2p - 2 + 8\mathbb{Z} = 4 + 8\mathbb{Z}.$$

The $p \equiv 5 \pmod 8$ case follows from a similar computation in $\mathrm{W}(\mathbb{Q}_2)$. The $p \equiv 1 \pmod 8$ case is the hardest. It depends on more calculations in $\mathrm{W}(\mathbb{Q}_2)$ and *Gauss's Lemma*, a result from number theory which states the following:

> If $p \equiv 1 \pmod 8$ is prime, then there exists an odd prime $q < \sqrt{p}$ such that p is not a square modulo q.

One uses Gauss's Lemma to select such a q and then sets $\varphi = \langle\!\langle -p, -q \rangle\!\rangle$. The prime p is a square in $\mathbb{Q}_2$, so $\chi(\varphi) = 0$. A corollary of Theorem 22.21 implies that $\left(\frac{p,q}{\mathbb{Q}_r}\right)$ splits for $r \in \Omega \smallsetminus \{p, q\}$, so $\partial_r(\varphi) = 0$ for $r \in \Omega \smallsetminus \{p, q\}$. Thus

$$0 = \chi(\varphi) = \chi_p\partial_p(\varphi) + \chi_q\partial_q(\varphi) = \chi_p(\langle -1, q\rangle) + \chi_q(\langle -1, p\rangle).$$

Since p is not a square modulo q, we have $\varphi_q \cong \langle -1, p\rangle$. We may make the inductive hypothesis[22] that $\chi_q(\varphi_q) = 4 + 8\mathbb{Z}$, which then forces $\chi_p(\langle -1, q\rangle) = 4 + 8\mathbb{Z}$. It follows that $\langle -1, q\rangle \cong \varphi_p$ over $\mathbb{F}_p$. We conclude that $\chi_p(\varphi_p) = 4 + 8\mathbb{Z}$, as desired. □

[22] The careful reader will note at this point that we did not need the full $q < \sqrt{p}$ power of Gauss's Lemma, but rather only that such a $q < p$ existed.

Modulo several lacunæ — namely Gauss's Lemma and the computations of $\mathrm{W}(\mathbb{Q})$ and $\mathrm{W}(\mathbb{Q}_2)$ — the reader should now have an appreciation of how Hilbert reciprocity is a special case of Scharlau reciprocity (for a particular χ) applied to norm forms of quaternion algebras. The content in this section was based on Scharlau's original paper [Sch72] (and Lam's recapitulation thereof in [Lam05, VI.5]). The paper [Sch72] is a font of wisdom on this subject, and much of it is approachable by the reader at this point. Of particular interest might be Scharlau reciprocity for rational function fields, which is also covered in [Lam05, IX.4].

References

[Lam99] T.Y. Lam. On the diagonalization of quadratic forms. *Math. Mag.*, 72(3):231–235, 1999.

[Lam05] T.Y. Lam. *Introduction to quadratic forms over fields*, volume 67 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2005.

[MH73] J. Milnor and D. Husemoller. *Symmetric bilinear forms*. Springer-Verlag, New York-Heidelberg, 1973. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 73.

[Sch72] Winfried Scharlau. Quadratic reciprocity laws. *J. Number Theory*, 4:78–97, 1972.

[Szy97] K. Szymiczek. *Bilinear algebra*, volume 7 of *Algebra, Logic and Applications*. Gordon and Breach Science Publishers, Amsterdam, 1997. An introduction to the algebraic theory of quadratic forms.