Witt's theorem of quadratic forms

math center

The following suspicious problem appears in a random centest problem set for reasons beyond my comprehension. I attempted it and successfully failed.

Exercise. Find the maximal dimension of a subspace $W \subset \mathrm{Mat}_{n\times n}(\mathbb{R})$ such that for any $A, B \in W$, $\mathrm{tr}(AB) = 0$.

A natural candidate of W is the space of all strictly upper triangular matrices. This clearly satisfies the condition. Moreover, this space is maximal, in the sense that no other matrix may be included into W. To see this, expand

$$\mathrm{tr}(AB) = \sum_{i,j} a_{ij} b_{ji}. \tag{1}$$

If $C \notin W$ and C has zero diagonal, then there is some i, j such that $i < j$ and $\mathrm{tr}(CE_{ij}) \neq 0$. Otherwise suppose that C has some non-zero diagonal entries. We take the 'strictly upper triangular' part $A \in W$ of C, so that $C - A$ is lower triangular. It follows that $\mathrm{tr}((C - A)^2) > 0$, again a contradiction.

However, showing that this W has the maximal dimension requires some theory.

1 Quadratic forms

Throughout this note, let k be a field and assume $\mathrm{char}(k) \neq 2$. All vector spaces are assumed finite dimensional.

Definition 1.1. A *quadratic form* on a vector space V is a map $Q : V \to k$ satisfying

1. $Q(\lambda x) = \lambda^2 Q(x)$ for all $\lambda \in k$, $x \in V$.
2. $(x, y) \mapsto Q(x + y) - Q(x) - Q(y)$ is a bilinear form on V.

The pair (V, Q) is called a *quadratic space*.

Since $\mathrm{char}(k) \neq 2$, we have a bijective correspondence between quadratic forms and symmetric bilinear forms, given by

$$x \cdot y = \frac{1}{2}\left(Q(x + y) - Q(x) - Q(y)\right),$$

$$Q(x) = x \cdot x.$$

Given a basis $\{e_i\}_{i=1}^n$ of V, the quadratic form is associated to a matrix $A = (a_{ij})$ given by $a_{ij} = e_i \cdot e_j$. Let $\mathbf{x} = \sum_{i=1}^n x_i e_i$. We have

$$Q(\mathbf{x}) = \sum_{i,j=1}^n a_{ij} x_i x_j = \mathbf{x}^\top A \mathbf{x}.$$

If we transform the basis with a matrix S^{-1}, the new quadratic form is given by $A' = S^\top AS$. It follows that the determinant of A is invariant up to a square. This is called the *discriminant* of the quadratic form, denoted as $\Delta(V) \in k/k^{\times 2}$.

The morphisms of quadratic spaces are the *metric morphisms* that are k-linear and preserve this bilinear form.

The quadratic space is said non-degenerate if the bilinear form is non-degenerate, i.e. the evident map $\theta : V \to V^\vee$ is an isomorphism.

We also have the notion of orthogonality from the bilinear form. Given a subspace $U \subset V$, we can define its orthogonal complement $U^\perp$. If V is non-degenerate, this fits into the exact sequence

$$0 \longrightarrow U^\perp \longrightarrow V \longrightarrow U^\vee \longrightarrow 0. \tag{2}$$

In particular, the *radical* of V is defined as $\mathrm{rad}(V) = V^\perp$. We now have three notions of non-degeneracy:

- $\Delta(V) \neq 0$;
- θ is an isomorphism;
- $\mathrm{rad}(V) = 0$.

Under our assumptions, these all coincide.

Now we prove Witt's theorem. Let (V, Q) and (V', Q') be isomorphic quadratic spaces. Let $U \subset V'$ be a subspace and let

$$s : U \hookrightarrow V'$$

be an injective metric morphism. We wish to extend s.

Lemma 1.2. If U is degenerate, then there exists a $U_1 \subset V$ which contains U as a proper subspace and an injective metric morphism $s_1 : U_1 \hookrightarrow V$ extending s.

Proof. Pick a nonzero $x \in \mathrm{rad}(U)$ and a linear functional $f \in U^\vee$ such that $f(x) = 1$. Then there exists $y \in V$ such that $u \cdot y = f(u)$ for all $u \in U$. Moreover, we assume that $y \cdot y = 0$ by replacing y with $y - \frac{1}{2}Q(y)x$. Put $U_1 = U \oplus ky$.

With the same construction as above, we can find an element $y' \in V' \setminus s(U)$ corresponding to the functional fs^{-1}, and by adding a multiple of $s(x) \in \mathrm{rad}(s(U))$, we may assume that $y' \cdot y' = 0$. Then $y \mapsto y'$ gives the desired metric morphism $U_1 \hookrightarrow s(U) \oplus V'$. □

Theorem 1.3 (Witt)**.** Let $V \simeq V'$ be isomorphic *non-degenerate* quadratic spaces. Then any such injective metric morphism $s : U \hookrightarrow V'$ can be extended to an isomorphism of V and V'.

Proof. For simplicity, we assume that $V = V'$. And by lemma 1.2, we may assume that U is non-degenerate. We proceed by induction on $\dim(U)$.

If $\dim(U) = 1$, let $x \in U$ and put $y := s(x)$ so that $x \cdot x = y \cdot y \neq 0$. It follows that $x + y$ is orthogonal to $x - y$. Then there must exist a $\epsilon \in \{\pm 1\}$ such that $z := x + \epsilon y$ satisfies $z^2 \neq 0$. Otherwise,

$$0 = Q(x + y) + Q(x - y) = Q(2x),$$

a contradiction. Let σ be the reflection along z. This is a metric morphism as it fixes the orthogonal complement of kz. In particular, we have

$$2\sigma(x) = \sigma(x + \epsilon y) + \sigma(x - \epsilon y) = -x - \epsilon y + x - \epsilon y = -2\epsilon y.$$

Clearly, $-\epsilon\sigma$ is the desired automorphism.

Now suppose $\dim(U) > 1$. Pick a nontrivial orthogonal decomposition $U = U_1 \oplus U_2$. By the induction hypothesis, $s|_{U_1}$ can be extended to some automorphism of V. With (the inverse of) this map we may assume that s is the identity on U_1. Then $s(U_2) \subset U_1^{\perp}$. By the induction hypothesis, $s|_{U_2}$ can be extended to an automorphism of $U_1^{\perp}$, which can be further extended to V setting it to be identiy on U_1. □

This elegant proof is from *A Course in Arithmetic* by Serre.

Now, the problem can be easily solved. In view of (1), we have a quadratic form Q on a real vector space V, and we are looking for the largest subspace W such that $Q|_W = 0$. This is called an *isotropic* subspace.

We've found a candidate W_0 that is at least a local maximum. Suppose that there is a W with larger dimension. Any embedding $W_0 \hookrightarrow W$ would be a metric morphism. By Witt's theorem, we can pull W back to an isotropic subspace that contains W_0 as a proper subspace. Contradiction.

2 Some more theory

Proposition 2.1. Any quadratic space (V, Q) has an orthogonal basis.

Proof. We use induction on $\dim(V)$. If V is isotropic, any basis is orthogonal. Otherwise suppose $Q(v) \neq 0$. It is clear that the orthogonal complement of kv is a hyperplane in V, so the induction hypothesis applies and the proof is complete. □

Lemma 2.2. Let (V, Q) be a quadratic space. If U is a non-degenerate subspace of V, then $V = U \oplus U^{\perp}$.

Proof. Clearly $U \cap U^{\perp} = 0$. To see that they span V, we take an orthogonal basis $\{e_i\}_{i=1}^{n}$ of U. We have $e_i \cdot e_i \neq 0$ for all i. Given $x \in V$, it's clear that

$$x - \sum_{i=1}^{n} \frac{e_i \cdot x}{e_i \cdot e_i} e_i \quad \in x + U$$

is orthogonal to U. This completes the proof. □

Proposition 2.3. Let U be a subspace of a quadratic space V such that $V = U + U^{\perp}$. (By the above lemma, this condition is always satisfied if U is non-degenerate.) Then any orthogonal basis B of U can be extended to an orthogonal basis of V.

Proof. We use induction on $\dim(V)$. The base case $\dim(V) = 0$ is trivial. Consider

- If $U^{\perp} = V$, let W be a direct sum complement of U and pick an orthogonal basis C of W. Then $B \cup C$ is an orthogonal basis of V.
- Otherwise, pick any basis A of the isotropic subspace $\mathrm{rad}(U) = U \cap U^{\perp}$ and extend it to an orthogonal basis C of $U^{\perp}$ with the induction hypothesis. Then $B \cup (C \setminus A)$ is an orthogonal basis of V.

This completes the proof. □

Let $\{e_i\}_{i=1}^n$ be an orthogonal basis. Then there exists $\lambda_1, \lambda_2, \cdots, \lambda_n$ in k such that

$$\mathbf{x} = \sum_{i=1}^n x_i e_i \Longrightarrow Q(\mathbf{x}) = \sum_{i=1}^n \lambda_i x_i^2.$$

The matrix of Q is thus diagonal, and we see that $\Delta(V) = \lambda_1 \lambda_2 \cdots \lambda_n$. In the rest of this section, *we assume that the spaces are non-degenerate*, so $\lambda_i \neq 0$. Hence we may view the λ_i as lying in $k^\times/k^{\times 2}$.

Conversely, for $\lambda \in k^\times/k^{\times 2}$, introduce the notation $\langle\lambda\rangle$ for the one-dimensional quadratic space. Denote the orthogonal direct sum of two quadratic spaces (V, Q) and (V', Q') as $V + V'$.

Corollary 2.4. Let (V, Q) be a non-degenerate quadratic space over k. Then

1. $V \simeq \langle\lambda_1\rangle + \langle\lambda_2\rangle + \cdots + \langle\lambda_n\rangle$, where $\lambda_i \in k^\times/k^{\times 2}$.
2. If $\lambda \in Q(V)\backslash\{0\}$, then there exists a non-degenerate quadratic space H such that $V \simeq \langle\lambda\rangle + H$.

Proof. 1 is proposition 2.1. 2 follows from extending $\{v\}$ with proposition 2.3, where v is a (non-isotropic) vector satisfying $Q(v) = \lambda$. □

Let $M(k)$ be the set of isomorphism classes of non-degenerate quadratic spaces over k. Then $M(k)$ clearly forms a commutative monoid under $+$, with identity being the zero space. Moreover, $M(k)$ has the left and right cancellation property by Witt's theorem.

In fact, we also have a natural mulplication on $M(k)$, hence making it a commutative *semiring*. This is the tensor product. Let's quickly summarize these operations.

Definition 2.5. Let (V, Q) and (V', Q') be elements of $M(k)$. Let $v \in V$, $v' \in V'$. We have

- Their orthogonal sum is $(V \oplus V', Q_+)$, where $Q_+(v + v') := Q(v) + Q'(v')$.
- Their tensor product is $(V \otimes_k V', Q_\otimes)$, where $(v \otimes v') \cdot (w \otimes w') := (v \cdot w)(v' \cdot w')$.

Definition 2.6. The *Grothendieck–Witt ring* $\widehat{W}(k)$ is obtained by using the Grothendieck construction on the semiring $M(k)$.

The elements of $\widehat{W}(k)$ can be (non-uniquely) represented as

$$n_1\langle\lambda_1\rangle + n_2\langle\lambda_2\rangle + \cdots + n_m\langle\lambda_m\rangle,$$

where $n_i \in \mathbb{Z}\backslash\{0\}$ and $\lambda_i \in k^\times/k^{\times 2}$ are distinct. We also have the dimension map $\dim : \widehat{W}(k) \longrightarrow \mathbb{Z}$ (via the universal property of Grothendieck group).

Definition 2.7. The *hyperbolic plane* is the quadratic space $\mathbb{H} = \langle 1\rangle + \langle -1\rangle$.

Definition 2.8. A nonzero vector $v \in V$ is *isotropic* if $Q(v) = 0$. A quadratic space V is *isotropic* if $Q|_V = 0$, and is *anisotropic* if $0 \notin Q(V \setminus \{0\})$. In particular, the zero space is anisotropic.

Lemma 2.9. Equivalent definitions of hyperbolic planes. TFAE:

1. V is a hyperbolic plane.
2. $V \simeq \langle\lambda\rangle + \langle -\lambda\rangle$ for some $\lambda \in k^\times/k^{\times 2}$.

3. V is spanned by isotropic vectors x, y such that $x \cdot y \neq 0$.

Proof. Trivial. $\square$

Proposition 2.10. Hyperbolic planes as subspaces. Let V be a non-degenerate quadratic space.

1. If $x \in V$ is an isotropic vector, then there is a subspace $U \subset V$ containing x such that $U \simeq \mathbb{H}$.

2. More generally, if $W \subset V$ is an isotropic subspace of dimension d, then there exists a $2d$-dimensional subspace U containing W such that $U \simeq d\mathbb{H}$.

3. V can be decomposed as $V \simeq V_h + V_a$, where V_a is anisotropic and $V_h \simeq d\mathbb{H}$ for some $d \in \mathbb{N}$. This decomposition is unique up to isomorphisms of V_h and V_a.

Proof. Let y be a vector such that $x \cdot y \neq 0$. Then $z = y - \dfrac{y \cdot y}{2x \cdot y} x$ is isotropic and $x \cdot z \neq 0$. By lemma 2.9, this proves 1.

For 2, we use induction on d. Let $w \in W$ be a nonzero vector and pick a subspace L such that $W = kw \oplus L$. By (2), we have $\dim(L^\perp) = \dim(W^\perp) + 1$, so we may pick $v \in L^\perp$ such that $v \cdot w \neq 0$. With the exact same argument as above, we see that $H := \operatorname{span}(w, v)$ is a hyperbolic plane. By lemma 2.2, we have $V = H \oplus H^\perp$. Now we may pass to the pair $L \subset H^\perp$ for induction.

Now we prove 3. The existence of such factorization is easy. We take a maximal isotropic subspace, say of degree d, and extend it to a sum of d hyperbolic spaces with clause 2. This is clearly a non-degenerate subspace. By lemma 2.2, it remains to see that its orthogonal complement is anisotropic. Indeed, if it contains an isotropic vector, it must contain a hyperbolic plane by clause 1. This contradicts the maximality of the isotropic subspace.

In the end of section 1, we've shown the uniqueness of V_h. The uniqueness of its complement, V_a, immediately follows by virtue of Witt's theorem. $\square$

The idea is that an isotropic subspaces carries no information beyond their dimension. To remove them, in view of proposition 2.2, we should find some minimal non-degenerate subspace containing the isotropic subspace, which turns out to be simply a sum of hyperbolic planes. This also leads to our definition of Witt rings.

Definition 2.11. The *Witt ring* $W(k)$ of a field k is $\widehat{W}(k)$ modulo the ideal $(\mathbb{H})$ generated by the hyperbolic plane $\mathbb{H}$.

In fact, it is easy to see that $(\mathbb{H}) = \mathbb{Z}\mathbb{H}$, i.e. its elements are multiples of $\mathbb{H}$ or its formal additive inverse. Since the tensor product is given by $\langle\lambda\rangle \oplus \langle\lambda'\rangle = \langle\lambda\lambda'\rangle$, tensoring with a hyperbolic plane always results in an orthogonal sum of hyperbolic planes by lemma 2.9.

By the above proposition, the elements of the Witt ring is in one-to-one correspondence with the isomorphism classes of anisotropic quadratic spaces.

Example 2.12. k is quadratically closed iff $\dim : \widehat{W}(k) \longrightarrow \mathbb{Z}$ is a ring isomorphism. In this case, $\mathbb{H} = 2\langle 1\rangle$, and thus $W(k) = \mathbb{Z}/2\mathbb{Z}$.

3 The Reals

In this section, let $k = \mathbb{R}$. We first find the rings $\widehat{W}(k)$ and $W(k)$.

We have $k/k^2 = \{\pm 1\}$, so after diagonalizing, any non-degenerate quadratic space V over k has the form $n_+\langle 1\rangle + n_-\langle -1\rangle$ where $n_+, n_- \in \mathbb{N}$. We'll show that these are in fact unique invariants by our next proposition.

Definition 3.1. The *signature* of V is defined as $n_+ - n_-$. This induces a ring homomorphism $\mathrm{sgn} : \widehat{W}(k) \longrightarrow \mathbb{Z}$.

Proposition 3.2. The map sgn induces a ring isomorphism $W(k) \simeq \mathbb{Z}$. We also have that $\widehat{W}(k) \simeq \mathbb{Z}[G]$, the integral group ring of the group $G = \{\pm 1\}$.

Proof. The first part is trivial. For the second part, it suffices to show that $\{\langle 1\rangle, \langle -1\rangle\}$ is a basis of (the group) $\widehat{W}(k)$. We've shown that they are spanning. To see their linear independence, suppose that $n_+\langle 1\rangle + n_-\langle -1\rangle = 0$ in $\widehat{W}(k)$. Passing to $W(k)$, we have $s\langle 1\rangle = 0$ where $s := n_+ - n_-$. Now we use the proporty of $\mathbb{R}$ to note that $|s|\langle 1\rangle$ is anisotropic, hence $s = 0$ by the one-to-one correspondence. Therefore $n_+ = n_-$ and by checking the dimension we have $n_+ = n_- = 0$. □

Corollary 3.3 (Sylvester's Law of Inertia)**.** Two non-degenerate quadratic spaces over $\mathbb{R}$ are isomorphic iff they have the same dimension and the same signature.

Proof. Specifying such a space is equivalent to specifying the pair (n_+, n_-), which is equivalent to giving the pair $(n_+ + n_-, n_+ - n_-)$. □

In summary, a real quadratic space V have a unique factorization as $\mathrm{rad}(V) + n_+\langle 1\rangle + n_-\langle -1\rangle$. Let's return to the problem at the beginning as an example of finding these invariants.

All we need is (1). Observe that $\mathrm{rad}(V) = 0$, i.e. V is non-degenerate. One checks that the following is an orthogonal basis of V:

$$\{E_{ii} : i\} \cup \{E_{ij} + E_{ji} : i < j\} \cup \{E_{ij} - E_{ji} : i < j\}\,.$$

The invariants may then be read off as $n_+ = n + (n^2 - n)/2$ and $n_- = (n^2 - n)/2$. It follows that the dimension of the maximal isotropic subspace is $\min(n_+, n_-) = (n^2 - n)/2$. Apparently, such maximal subspace is far from unique.

Remark 3.4. The result from this section holds if we replace $\mathbb{R}$ with any ordered field where every positive element is a square. Such fields are called *Euclidean*. Real closed fields are, by definition, Euclidean.

Checkpoint: The original problem has been solved and understood. We see that the theory of quadratic forms over $\mathbb{C}$ and $\mathbb{R}$ is very mild. In the following sections we'll provide a similar classification of quadratic forms over finite fields and $\mathbb{Q}_p$ as a mere introduction to the topic.

4 Finite fields

Let $k = \mathbb{F}_q$ be a finite field of characteristic $p \neq 2$.

Theorem 4.1 (Chevalley-Warning)**.** Let $\{f_i\}_{i=1}^m$ be polynomials in $k[X_1, \ldots, X_n]$ and denote the set of their common zeros as V. If $\sum \deg(f_i) < n$, then p divides $|V|$.

Proof. Put $P = \prod_{i=1}^{m} \left(1 - f_i^{q-1}\right)$. We notice that P is the characteristic function of V. It follows that $|V| \equiv \sum_{x \in k^n} P(x) \pmod p$.

To see that the right hand side is 0, we note that $\deg(P) < n(q-1)$, so for each monomial in P, there exists at least one X_j such that its exponent is less than $q-1$. The result follows from the elementary fact that

$$0 \leq l < q - 1 \Longrightarrow \sum_{a \in k} a^l = 0,$$

a direct consequence of the cyclic nature of the multiplicative group $k^\times$. □

Corollary 4.2. Any quadratic form of at least 3 variables over a finite field has a nontrivial zero.

Theorem 4.3. Let's write $k^{\times 2}/k^\times = \{1, u\}$ since it has order 2. Then there exists exactly two (up to isomorphism) non-degenerate quadratic spaces of each dimension n:

- $n\langle 1 \rangle$, with discriminant $\Delta = 1$;
- $(n-1)\langle 1 \rangle + \langle u \rangle$, with discriminant $\Delta = u$.

Hence the discriminant is a complete invariant of such spaces.

Proof. Induct on n. The base case $n = 1$ is trivial. By proposition 2.3, it suffices to show that any non-degenerate quadratic space of dimension ≥ 2 contains a vector x such that $Q(x) = 1$. Indeed we prove a stronger statement: every 2-dimensional non-degenerate quadratic space V is *universal*, meaning that $Q(V) = k$.

- If V contains an isotropic vector, then it is a hyperbolic plane by proposition 2.10. The hyperbolic plane is universal.
- Otherwise, V is anisotropic. Apply corollary 4.2 to $V + \langle -\lambda \rangle$ for any $\lambda \in k^\times$, and we see that $\lambda \in Q(V)$.

The proof is complete. □

From the above proof, we extract the following definition and an equivalent formulation.

Definition 4.4. We say that a quadratic space V represents a scalar $\lambda \in k$ if $\lambda \in Q(V \setminus \{0\})$.

Lemma 4.5. If V is non-degenerate, it represents $\lambda \in k^\times$ if and only if $V + \langle -\lambda \rangle$ represents 0 (i.e. it contains an isotropic vector).

5 Quaternion algebras

Fix a field k of characteristic $p \neq 2$.

Definition 5.1. Given two nonzero scalars $a, b \in k^\times$, the *quaternion algebra* $(a, b)_k$ is the k-algebra generated by two elements x_1, x_2, subject to the relations

$$\begin{aligned} x_1^2 &= a \cdot 1, \\ x_2^2 &= b \cdot 1, \\ x_1 x_2 &= -x_2 x_1. \end{aligned}$$

The uniqueness of this definition is a part of the following proposition.

Proposition 5.2. Up to algebra isomorphism, the quaternion algebra $\mathcal{A} := (a,b)_k$ is unique, and has $\{1, x_1, x_2, x_1x_2\}$ as a basis. Moreover, $\mathcal{A}$ is a central simple k-algebra, hence represents an element in the Brauer group $\mathrm{Br}(k)$.

Proof. Put $x_3 = x_1x_2$ and $B = \{1, x_1, x_2, x_3\}$. Clearly, B spans $\mathcal{A}$ as a vector space and multiplication between its elements is specified by the relations. To prove uniqueness, it remains to show that B is linearly independent. In fact, we contend that an element $y = \alpha_0 + \alpha_1x_1 + \alpha_2x_2 + \alpha_3x_3$ lies in the center of $\mathcal{A}$ if and only if $\alpha_1 = \alpha_2 = \alpha_3 = 0$. This follows from the observation that for $i, j \in \{1, 2, 3\}$,

$$i \neq j \Longrightarrow x_ix_j = -x_jx_i. \tag{3}$$

Hence for $i \in \{1, 2, 3\}$, $[x_i, y]$ is a linear combination of x_j and x_k, where $\{i, j, k\} = \{1, 2, 3\}$. If $y \in Z(\mathcal{A}) \setminus k$, this implies that some x_j and x_k are collinear, contradicting (3).

Finally, we need that $\mathcal{A}$ is simple. This is quite simple. Given y as above and supposing that $\alpha_i \neq 0$ for some i, pick $j \neq i$. By taking the commutator of y with x_j and then x_i, we are guaranteed to obtain an invertible element of $\mathcal{A}$. In other words, the two-sided ideal generated by y is $\mathcal{A}$, which means that $\mathcal{A}$ is simple. □

Remark 5.3. The quaternion algebra is a special case of the *Clifford algebra* of a quadratic space V which can be explicitly constructed as follows. Define the two-sided ideal $I := \langle x \otimes x - Q(x) \cdot 1 : x \in V \rangle$ of the tensor algebra $T(V)$, and the Clifford algebra is the quotient $\mathrm{Cl}(V) := T(V)/I$. If $V = \langle a \rangle + \langle b \rangle$, this of course coincides with $(a,b)_k$. With some modification, the Clifford algebra gives rise to an invariant of quadratic spaces also taking value in the Brauer group.

Lemma 5.4. A quaternion algebra $(a,b)_k$ is either a division ring or isomorphic to $\mathrm{Mat}_{2\times 2}(k)$.

Proof. This is Wedderburn's theorem (Theorem 3.4 here), applied to a simple algebra which is finite-dimensional over a field. □

Example 5.5. The prototypical quaternions $\mathbb{H} = (-1,-1)_{\mathbb{R}}$ is a division ring.

Definition 5.6. Define the conjugation map $\overline{\cdot}$ as

$$\overline{\alpha_0 + \alpha_1x_1 + \alpha_2x_2 + \alpha_3x_3} = \alpha_0 - \alpha_1x_1 - \alpha_2x_2 - \alpha_3x_3.$$

The *(reduced) norm form* on the quaternion algebra is then defined as $Q(x) = x\overline{x} \in k$.

We have $Q(\alpha_0 + \alpha_1x_1 + \alpha_2x_2 + \alpha_3x_3) = \alpha_0^2 - \alpha_1^2a - \alpha_2^2b + \alpha_3^2ab$. Note that the reduced norm is related to the usual algebra norm by

$$Q(x)^2 = \mathrm{N}_{\mathcal{A}|k}(x). \tag{4}$$

We can check this by direct computation, or more conveniently, by reducing to the split case $\mathrm{Mat}_{2\times 2}(k)$. Put $K = k(\sqrt{a})$. By the criterion given in 5.10, we have $\mathcal{A} \otimes K \simeq (a,b)_K \simeq \mathrm{Mat}_{2\times 2}(K)$. This extension of scalar preserves conjugation and the norm form. Also define the reduced trace as $T(x) = x + \overline{x}$. For all $x \in (a,b)_k \subset (a,b)_K$, we have

$$x^2 - T(x)x + Q(x) = 0.$$

View x as a 2×2 matrix over K. Assuming that $x \notin k$, this implies $Q(x) = \det(x)$; if $x \in k$, the same result holds. The rest is easy. As a left module over itself, $\mathcal{A} \otimes K \simeq K^2 \oplus K^2$. We have

$$\mathrm{N}_{\mathcal{A}|k}(x) = \mathrm{N}_{\mathcal{A}\otimes K|K}(x) = \det(x)^2 = Q(x)^2.$$

Lemma 5.7. We equip the quaternion algebra $(a,b)_k$ with the norm form Q and view it as a non-degenerate quadratic space. Then we have

$$\begin{aligned}(a,b)_k &\simeq \langle 1\rangle + \langle -a\rangle + \langle -b\rangle + \langle ab\rangle \\ &= (1-\langle a\rangle)(1-\langle b\rangle)\,, \text{ if viewed in } W(k).\end{aligned} \tag{5}$$

Proof. Proven in above discussion. □

Lemma 5.8. Two quaternion algebras are isomorphic if and only if they are isomorphic as quadratic spaces when equipped with the norm form.

Proof. Let $\mathcal{A} = (a,b)_k$ and $\mathcal{A}' = (a',b')_k$ be two quaternion algebras. Suppose we have an isomorphism $\varphi : \mathcal{A} \to \mathcal{A}'$ of k-algebras. φ sends scalars (elements of k) to scalars. Less trivially, it also preserves the *pure quaternions*, i.e., elements of $\operatorname{span}\{x_1, x_2, x_3\}$. Indeed, an element y is either a scalar or a pure quaternion iff $y^2 \in k$. Therefore, if $y = y_0 + y_{\text{pure}}$, we have

$$\varphi(\overline{y}) = \varphi(y_0 - y_{\text{pure}}) = \varphi(y_0) - \varphi(y_{\text{pure}}) = \overline{\varphi(y)},$$

so φ commutes with conjugation, and hence with the norm form. This proves the 'only if' direction.

Conversely, suppose that $\mathcal{A} \simeq \mathcal{A}'$ as quadratic spaces. By Witt's theorem and the formula in (5), there is a copy of $\langle -a\rangle + \langle -b\rangle$ contained in $\mathcal{A}'_{\text{pure}}$ the subspace of pure quaternions in $\mathcal{A}'$. Since $y \in \mathcal{A}'_{\text{pure}} \implies Q'(y) = y\overline{y} = -y^2$, this subspace encodes the precise relations in definition 5.1. By uniqueness (proposition 5.2), $\mathcal{A}$ embeds into $\mathcal{A}'$ as algebras, and the result follows. □

As a result, the quaternion algebra $(a,b)_k$ may equivalently be defined for $a, b \in k^\times / k^{\times 2}$, as a square factor does not affect the quadratic space in (5).

Lemma 5.9. Let $\mathcal{A} = (a,b)_k$. Then the conjugation map $\overline{\cdot}$ is an algebra isomorphism $\mathcal{A}^{\text{op}} \to \mathcal{A}$. In particular, the norm form Q satisfies $Q(xy) = Q(x)Q(y)$ for all $x, y \in \mathcal{A}$.

Proof. The conjugation map is a bijective k-linear map, so it suffices to check that $\overline{xy} = \overline{y}\,\overline{x}$ for the basis elements $1, x_1, x_2, x_3$. We've already proved this in (3). Now, we have

$$Q(xy) = xy\overline{xy} = x(y\overline{y})\overline{x} = Q(x)Q(y),$$

which proves the second part. □

Proposition 5.10. Let $\mathcal{A} = (a,b)_k$. TFAE:

1. $\mathcal{A}$ *splits* over k, i.e., it represents the identity element in the Brauer group $\operatorname{Br}(k)$;
2. $\mathcal{A}$ is isomorphic to $\operatorname{Mat}_{2\times 2}(k)$;
3. $\mathcal{A}$ contains an isotropic vector;
4. $\mathcal{A}$ contains an isotropic pure quaternion;
5. the quadratic space $\langle a\rangle + \langle b\rangle$ represents 1, i.e., $1 = as^2 + bt^2$ for $s, t \in k$.

Proof. The equivalence $1 \iff 2$ is lemma 5.4. $4 \implies 3$ is trivial.

$2 \iff 3$: I claim that an element $x \in \mathcal{A}$ is invertible iff it's not isotropic. If x is isotropic, then $Q(x) = x\overline{x} = 0$, so x is not invertible. Conversely, suppose that $Q(x) = x\overline{x} \in k^\times$. Note that x is algebraic over k, so this implies that x is invertible. Since x commutes with $\overline{x}$, it is clear that $Q(x)^{-1}\overline{x}$ is the inverse of x.

$2, 3 \implies 4$: By lemma 5.8, it suffices to prove 4 for one pair of a, b. Let's pick $a = 1$, $b = -1$. (5) states that

$$(1, -1)_k \simeq \langle 1 \rangle + (\langle -1 \rangle + \langle 1 \rangle + \langle -1 \rangle).$$

The second summand corresponds to the pure quaternions, and clearly contains an isotropic vector.

$4 \iff 5$: By (5), we have

$$\begin{aligned} \mathcal{A}_{\text{pure}} &\simeq \langle -a \rangle + \langle -b \rangle + \langle ab \rangle \\ &\simeq \langle b \rangle + \langle a \rangle + \langle -1 \rangle, \quad \left(\text{multiply by} - ab \in k^\times\right). \end{aligned}$$

Since $\langle a \rangle + \langle b \rangle$ is non-degenerate, it is either a hyperbolic plane or anisotropic. The result follows. ☐

Corollary 5.11. For all $a \in k^\times$, we have

$$(1, a)_k \simeq (a, -a)_k \simeq (a, 1-a)_k \simeq \mathrm{Mat}_{2\times 2}(k).$$

Proof. Use criterion 3 and 5. ☐

Example 5.12. In section 4, we proved that every 2-dimensional quadratic space over a finite field is universal. Hence by criterion 5, every quaternion algebra over a finite field splits.

Proposition 5.13. For $a, b, c \in k^\times / k^{\times 2}$, we have

$$(a, b)_k \otimes (a, c)_k \simeq (a, bc)_k \otimes \mathrm{Mat}_{2\times 2}(k).$$

Computing in $\mathrm{Br}(k)$, this is

$$(a, b)_k \, (a, c)_k = (a, bc)_k \, .$$

Proof. Let $\{1, x_1, x_2, x_3\}$ and $\{1, x_1', x_2', x_3'\}$ be the bases of $(a, b)_k$ and $(a, c)_k$, respectively. Put

$$\begin{aligned} &\xi_1 = x_1 \otimes 1, \quad \xi_2 = x_2 \otimes x_2' \\ \implies &\xi_1^2 = a, \quad \xi_2^2 = bc, \quad \xi_1\xi_2 = -\xi_2\xi_1. \end{aligned}$$

and

$$\begin{aligned} &\zeta_1 = 1 \otimes x_2', \quad \zeta_2 = x_1 \otimes x_3' \\ \implies &\zeta_1^2 = c, \quad \zeta_2^2 = -a^2 c, \quad \zeta_1\zeta_2 = -\zeta_2\zeta_1. \end{aligned}$$

Denote the subalgebras generated by $\{\xi_1, \xi_2\}$ and $\{\zeta_1, \zeta_2\}$ as $\mathcal{A} \simeq (a, bc)_k$ and $\mathcal{B} \simeq \left(c, -a^2 c\right)_k$, respectively. We have $\mathcal{B} \simeq \mathrm{Mat}_{2\times 2}(k)$ by the previous corollary.

It remains to show that $\mathcal{A} \otimes \mathcal{B}$ is isomorphic to $(a, b)_k \otimes (a, c)_k$. From inspecting the basic elements, we know that the elements of $\mathcal{A}$ commutes with those of $\mathcal{B}$. So we have an algebra homomorphism $\mathcal{A} \otimes \mathcal{B} \to (a, b)_k \otimes (a, c)_k$ given by multiplication. It is injective because $\mathcal{A} \otimes \mathcal{B}$ is central simple. As both sides have the same dimension, it must be an isomorphism. ☐

In particular, it follows that any quaternion algebra $(a, b)_k$ lies in 2-torsion of the Brauer group $\mathrm{Br}(k)$. In fact, we have a deep result providing a partial converse to this.

Theorem 5.14 (Merkurje)**.** The quaternion algebras generate the 2-torsion of the Brauer group $\mathrm{Br}(k)$.

6 Hasse invariant

Definition 6.1. Let V be a non-degenerate quadratic space diagonalized as $V \simeq \langle\lambda_1\rangle + \langle\lambda_2\rangle + \cdots + \langle\lambda_n\rangle$. The *Hasse invariant* of V is defined as the k-algebra

$$S(V) := \bigotimes_{i<j} (\lambda_i, \lambda_j)_k \, .$$

The End

Compiled on 2025/07/15.
Home page