Chapter 3

Quaternion algebras and quadratic forms

By Wedderburn's theorem (Theorem 2.2.6) and the fact that a central division algebra over F must have square dimension, we can explicitly list the types of simple F-algebras in small dimensions. In 1-dimension, there is just F. In 2- and 3-dimensions, there are just quadratic and cubic fields (exercise below). In 4-dimensions, we can have quartic fields, 4-dimensional division algebras and $M_2(F)$. So 4-dimensions is the smallest case where we get something besides fields. These lowest dimensional noncommutative simple algebras are what we will call quaternion algebras. From this point of view, they are the simplest (only 2% pun intended) noncommutative algebras, and thus a natural object of study.

> **Exercise 3.0.1.** Let A be a simple F-algebra of dimension < 4. Show A is a field.

3.1 Construction

Definition 3.1.1. *A* **quaternion algebra** *over F is a four-dimensional central simple F-algebra.*

We will often denote quaternion algebras by B, and use the letter A for a non-necessarily quaternion algebra. (In number theory it's common to use B, or sometimes D, for quaternion algebras. I believe this is because A often was used for an abelian ring, so B was used for something nonabelian (belian?), but my memory is not entirely trustworthy. Caution: if you see D to denote a quaternion algebra somewhere else (including my papers), it does not necessarily mean a quaternion *division* algebra—in these notes I will try to restrict the use of D solely for division algebras.)

Note the central condition rules out fields, so any quaternion algebra over F is either a noncommutative division algebra or the split matrix algebra $M_2(F)$.

Now you might wonder why we allow matrix algebras to be called quaternion algebras if our original motivation was to generalize $\mathbb{H}$. Why not require that all quaternion algebras are division algebras? One reason is that the theories are closely related and sometimes it is useful to consider both matrix algebras and division algebras together. Another reason is that some (in fact, almost all) local components of global quaternion division algebras will be local matrix algebras, as previewed in Section 2.7. Of course, in the end it is just terminology and what has become standard practice. (Nevertheless, some people seem to use quaternion algebra to mean quaternion division algebra, either out of laziness or ignorance. But that does not mean you should. It just means I will doubt whether you know anything about quaternion algebras.)

Hilbert symbols

There is a well-known way to construct quaternion algebras.

Definition 3.1.2. *Let F be a field of characteristic not 2, and $a, b \in F^\times$. The* **(algebra) Hilbert symbol** $\left(\frac{a,b}{F}\right)$ *is the quaternion algebra with F-basis $1, i, j, k$ and multiplication satisfying*

$$i^2 = a, \quad j^2 = b, \quad ij = -ji = k. \tag{3.1.1}$$

That is to say, to construct $\left(\frac{a,b}{F}\right)$ you adjoin formal square roots[1] $i = \sqrt{a}$ and $j = \sqrt{b}$ to F (formally meaning *not* inside the algebraic closure $\overline{F}$, or else they will commute) with the relation $k := ij = -ji$ so

$$k^2 = (ij)(-ji) = -ij^2i = -ibi = -bi^2 = -ab.$$

(If you do not like my liberal use of radicals, just adjoin formal non-commuting variables i, j, k to F modulo the above relations.) Then we note,

$$ik = i^2j = aj, \quad ki = -ji^2 = -aj, \quad jk = -j^2i = -bi, \quad kj = ij^2 = bi.$$

Here we are extending multiplication $\left(\frac{a,b}{F}\right) \times \left(\frac{a,b}{F}\right) \to \left(\frac{a,b}{F}\right)$ so it is F-bilinear and we get a well-defined multiplication map

$$(x + yi + zj + wk)(x' + y'i + z'j + w'k) = x'' + y''i + z''j + w''k,$$

since the product of any two of $1, i, j, k$ lies in one of the following sets: F, Fi, Fj or Fk. Here x, y, z, w, etc denote elements of F. Using linearity, associativity of this multiplication follows from associativity of the elements i, j, k. (Note if $a, b \in \{\pm 1\}$, the set $\{\pm 1, \pm i, \pm j, \pm k\}$ forms an abelian group of order 8, which is the quaternion group Q_8 if $a = b = -1$.) Thus $\left(\frac{a,b}{F}\right)$ is a 4-dimensional associative algebra.

> **Exercise 3.1.1.** Check that $\left(\frac{a,b}{F}\right)$ is a CSA.

[1] I use notation like $i = \sqrt{a}$ because I find it suggestive. However it is ambiguous, and formally requires suitable interpretation. For instance, if there is already a square root of a in F, I do not mean that i is one of these—it is just a formal new symbol such that $i^2 = a$. Similarly, if $a = b$, the notation $i = \sqrt{a}$ and $j = \sqrt{b}$ does not mean $i = j$—they are both just formal symbols such that $i^2 = j^2 = a$. Thus the $i = \sqrt{a}$ notation should be viewed along the same lines as $\int e^x \, dx = e^x + C$ in calculus or $\frac{1}{3}x^2 - 5\sqrt{x} = O(x^2)$ with big O notation—it doesn't really mean equality of two objects, but just membership in an equivalence class of objects satisfying a certain property.

Therefore $\left(\frac{a,b}{F}\right)$ is indeed a quaternion algebra, and the definition is justified.

We had to require $\operatorname{char} F \neq 2$ here to ensure we get a noncommutative algebra. In characteristic 2, this construction gives a non-central algebra (note $ij = ji$), and thus not a quaternion algebra, so one should modify the construction. However, we are not interested in fields of characteristic 2 for this class. Hence, for simplicity:

From now on, we will assume $\operatorname{char} F \neq 2$.

Note that $\left(\frac{a,b}{F}\right)$ contains the 3 distinct (though possibly isomorphic) quadratic subalgebras generated by i, j and k as depicted in this diagram:

$$\begin{tikzcd}
 & \left(\frac{a,b}{F}\right) \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
F(i) \simeq F \oplus F\sqrt{a} \arrow[dr, dash] & F(j) \simeq F \oplus F\sqrt{b} \arrow[d, dash] & F(k) \simeq F \oplus F\sqrt{-ab} \arrow[dl, dash] \\
 & F &
\end{tikzcd}$$

The direct sums here denote direct sums as vector spaces, not ring direct sums. For instance, $F \oplus F\sqrt{a}$ will be ring isomorphic to either the $F \oplus F$ if $\sqrt{a} \in F$ and $F \oplus F\sqrt{a}$ is ring isomorphic to the quadratic field $F(\sqrt{a})$ if $\sqrt{a} \notin F$.

Exercise 3.1.2. Show $F(i) \simeq F \oplus F$ if a is a square in F and $F(i)/F$ is a quadratic field extension otherwise.

We have only seen a couple of explicit quaternion algebras so far (though Theorem 2.7.5 tells us there should be many). We can easily realize the two we have seen via Hilbert symbols.

Example 3.1.1. $\left(\frac{-1,-1}{\mathbb{R}}\right) = \mathbb{H}$.

Example 3.1.2. For any field F (with characteristic not 2), $\left(\frac{1,1}{F}\right) \simeq M_2(F)$. We can explicate the isomorphism by sending $i \mapsto \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$, $j \mapsto \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}$, and $k \mapsto \begin{pmatrix} & 1 \\ -1 & \end{pmatrix}$.

In fact any quaternion algebra (remember the characteristic is not 2!) is given by this Hilbert symbol construction.

Theorem 3.1.3. *Let B be a quaternion algebra over F. Then there exist $a, b \in F^\times$ such that $B \simeq \left(\frac{a,b}{F}\right)$.*

Proof. By Example 3.1.2, it suffices to assume B is a quaternion division algebra.

Let $K = F(\alpha)$ be a maximal (whence quadratic) subfield of B. Then as in (2.5.1), we have an algebra embedding $B \hookrightarrow M_2(K)$. Consider some $\beta \in B - K$. Then $F(\beta)$ is a subfield of B as B is division, and $F(\beta)$ must be quadratic as $\beta \notin F$. The subalgebra of B generated by α, β must be a simple algebra of dimension 3 or 4, but cannot be a cubic field as the maximal fields are quadratic, so it must be all of B by Exercise 3.0.1.

We may choose α so that $\alpha^2 = a \in F^\times$. By Skolem–Noether, the nontrivial Galois automorphism $\alpha \mapsto -\alpha$ of K/F is given by conjugation by some $\beta \in B^\times$:

$$\beta\alpha\beta^{-1} = -\alpha.$$

Since this β does not commute with α, $\beta \notin K$, $L = F(\beta)$ is a distinct quadratic subfield of B, and $B = KL$ (i.e., B is generated by α and β) by the above discussion. One the other hand, β^2 commutes with α:

$$\beta^2\alpha\beta^{-2} = \beta(\beta\alpha\beta^{-1})\beta^{-1} = \beta(-\alpha)\beta^{-1} = -\beta\alpha\beta^{-1} = \alpha.$$

So β^2 commuting with generators α, β of B implies $\beta^2 \in Z(B) = F$. Hence $b = \beta^2 \in F^\times$.

Then it is easy to check that $\alpha \mapsto i$, $\beta \mapsto j$ gives an isomorphism $B \simeq \left(\frac{a,b}{F}\right)$. □

We remark a couple of general facts that fall out of the above proof.

Corollary 3.1.4. *Let B be a quaternion division algebra. Then any $\alpha \in B - F$ generates a quadratic subfield of B.*

We also implicitly proved the following when B is division, but it is true without the division hypothesis.

Exercise 3.1.3. Let B be a quaternion algebra and K a quadratic subfield. Then the centralizer $C_B(K) = K$.

Given a quaternion algebra B, the Hilbert symbol realization for B is not unique—i.e., we may have $\left(\frac{a,b}{F}\right) \simeq \left(\frac{a',b'}{F}\right)$ for $(a,b) \neq (a',b')$. This is evident if $F = \mathbb{R}$, as there are infinitely many choices for (a,b) in the Hilbert symbol, but by Frobenius's theorem the only quaternion algebras are $M_2(\mathbb{R})$ and $\mathbb{H}$ up to isomorphism. It is also evident from the construction that

$$\left(\frac{a,b}{F}\right) = \left(\frac{b,a}{F}\right) \tag{3.1.2}$$

as this just switches i and j. Here is another simple case of coincidences of Hilbert symbols.

Exercise 3.1.4. If c and d are squares in $F^\times$, show that, for $a, b \in F^\times$,

$$\left(\frac{ac,bd}{F}\right) = \left(\frac{a,b}{F}\right)$$

Note the above two cases of Hilbert symbol isomorphisms are not enough to explain all such isomorphisms. For instance, when $F = \mathbb{R}$, by varying the Hilbert symbol parameters by squares, we reduce to 4 possibilities: $\left(\frac{1,1}{\mathbb{R}}\right)$, $\left(\frac{1,-1}{\mathbb{R}}\right)$, $\left(\frac{-1,1}{\mathbb{R}}\right)$ and $\left(\frac{-1,-1}{\mathbb{R}}\right)$. The middle two are the same by (3.1.2), but this still leaves 3 cases of Hilbert symbols. On the other hand, by Frobenius's theorem we know there are only two real quaternion algebras up to isomorphism, $M_2(\mathbb{R}) \simeq \left(\frac{1,1}{\mathbb{R}}\right)$ and $\mathbb{H} = \left(\frac{-1,-1}{\mathbb{R}}\right)$. Corollary 3.1.5 below will resolve the situation over $\mathbb{R}$ by telling us $\left(\frac{1,-1}{\mathbb{R}}\right) \simeq M_2(\mathbb{R})$.

Later we will use quadratic forms to give general criteria for when two Hilbert symbols are isomorphic.

Next, let us consider matrix presentations for quaternion algebras. By Theorem 3.1.3, we may as well assume $B = \left(\frac{a,b}{F}\right)$. Let $K = F(\sqrt{a})$, which will just be F if a is a square. Then

$$i \mapsto \begin{pmatrix} \sqrt{a} & \\ & -\sqrt{a} \end{pmatrix}, \quad j \mapsto \begin{pmatrix} & b \\ 1 & \end{pmatrix}, \quad k \mapsto \begin{pmatrix} & b\sqrt{a} \\ -\sqrt{a} & \end{pmatrix} \tag{3.1.3}$$

induces an algebra homomorphism $\phi : B \hookrightarrow M_2(K)$. To see this, one just needs to check compatibility with (3.1.1), i.e.,

$$\phi(i)^2 = a, \quad \phi(j)^2 = b, \quad \phi(k) = \phi(i)\phi(j) = -\phi(j)\phi(i),$$

which is elementary. Note this matrix embedding generalizes Example 3.1.2.

> **Exercise 3.1.5.** Fix $a, b \in F^\times$ and let $K = F(\sqrt{a})$. Show that (3.1.3) induces an F-algebra isomorphism
>
> $$\left(\frac{a,b}{F}\right) \simeq \left\{ \begin{pmatrix} \alpha & b\beta \\ \overline{\beta} & \overline{\alpha} \end{pmatrix} : \alpha, \beta \in K \right\} \subset M_2(K)$$
>
> where $\alpha \mapsto \overline{\alpha}$ denotes the nontrivial element of $\mathrm{Gal}(K/F)$ if K/F is quadratic, or an isomorphism
>
> $$\left(\frac{a,b}{F}\right) \simeq M_2(F)$$
>
> if $K = F$.

Note for $\mathbb{H} = \left(\frac{-1,-1}{\mathbb{R}}\right)$, this gives the matrix representation $\left\{ \begin{pmatrix} \alpha & -\beta \\ \overline{\beta} & \overline{\alpha} \end{pmatrix} : \alpha, \beta \in \mathbb{C} \right\}$ from Exercise 2.1.10.

Corollary 3.1.5. *We have $\left(\frac{a,b}{F}\right) \simeq M_2(F)$ if a is a square or b is a square in $F^\times$.*

Proof. This is immediate from (3.1.2) and the previous exercise. □

Personally, I generally prefer to do quaternion calculations using matrix representations, though many people often work with the $1, i, j, k$ basis. As with anything, you can do what you like. For addition it doesn't matter, but for multiplication I think the matrix form is more convenient.

> **Exercise 3.1.6.** For this exercise only, suppose F has characteristic 2. Let $a, b \in F^\times$. Show $F \oplus Fi \oplus Fj \oplus Fk$ can be made a quaternion algebra where i, j, k are symbols

satisfying the multiplication rules:

$$i^2 + i = a, \quad j^2 = b, \quad k = ij = j(1+i).$$

The canonical involution

An **involution** ι of B is an anti-automorphism of order 2, i.e., an F-linear map $\iota : B \to B$ such that $\iota(\alpha\beta) = \iota(\beta)\iota(\alpha)$ and $\iota(\iota(\alpha)) = \alpha$. In other words, ι is an F-algebra homomorphism $B \to B^{\mathrm{opp}}$ which is its own inverse. Since B is simple, such a ι must be an algebra isomorphism, whence $B \simeq B^{\mathrm{opp}}$ if an involution exists.[2] Note $\iota(\alpha\beta) = \iota(\beta)\iota(\alpha)$ means ι cannot be the identity, as B is not commutative.

The **canonical involution** is given by $\overline{\alpha} = x - yi - zj - wk$ where $\alpha = x + yi + zj + wk \in B$. In other words, the canonical involution just interchanges the square roots $\pm i$ of a, $\pm j$ of b and $\pm k$ of $-ab$ in the Hilbert symbol constructions. This is analogous to Galois conjugation for quadratic fields, and we will see below that the canonical involution is compatible with Galois conjugation. Certainly it restricts to Galois conjugation on $F(i)$, $F(j)$ and $F(k)$ when these are quadratic fields. Furthermore, the fixed points of the involution are F:

$$\{\alpha \in B : \overline{\alpha} = \alpha\} = F.$$

Exercise 3.1.7. Check the canonical involution is indeed an involution on B. In particular $\overline{\overline{\alpha}} = \alpha$ and $\overline{(\alpha\beta)} = \overline{\beta}\,\overline{\alpha}$.

Exercise 3.1.8. Show $B \otimes B \simeq M_4(F)$. (Cf. Exercise 2.7.2.)

Example 3.1.3. Let's compute the canonical involution on $M_2(F) = \left(\frac{1,1}{F}\right)$. Using the isomorphism from Example 3.1.2, we have

$$x + yi + zj + wk = \begin{pmatrix} x+y & z+w \\ z-w & x-y \end{pmatrix},$$

so

$$\overline{x + yi + zj + wk} = \begin{pmatrix} x-y & -z-w \\ -z+w & x+y \end{pmatrix}.$$

Thus, in terms of matrix coefficients, the canonical involution is given by

$$\overline{\begin{pmatrix} x & y \\ z & w \end{pmatrix}} = \begin{pmatrix} w & -y \\ -z & x \end{pmatrix}.$$

In particular, if $g \in \mathrm{GL}_2(F)$, then

$$\overline{g} = \det(g)g^{-1}.$$

[2] Since not all CSAs are isomorphic to their opposite algebras, not all CSAs have involution. In fact, most don't. This is one of the features that makes quaternion algebras especially nice.

Example 3.1.4. Now suppose $B = \left(\frac{a,b}{F}\right)$ is a quaternion division algebra, which we can identify with $\left\{ \begin{pmatrix} \alpha & b\beta \\ \overline{\beta} & \overline{\alpha} \end{pmatrix} : \alpha, \beta \in K \right\}$ as in Exercise 3.1.5. Then calculating as in the previous example gives

$$\overline{\begin{pmatrix} \alpha & b\beta \\ \overline{\beta} & \overline{\alpha} \end{pmatrix}} = \begin{pmatrix} \overline{\alpha} & -b\beta \\ -\overline{\beta} & \alpha \end{pmatrix}.$$

If $g \in B^\times$, then

$$\overline{g} = N(g)g^{-1}$$

where $N\begin{pmatrix} \alpha & b\beta \\ \overline{\beta} & \overline{\alpha} \end{pmatrix} = \alpha\overline{\alpha} - b\beta\overline{\beta}$ is the reduced norm.

Recall from Section 2.5, we have (reduced) norm and trace maps $N : B \to F$ and $\mathrm{tr} : B \to F$ which are multiplicative and additive group homomorphisms. Further, since $\deg B = 2$, the characteristic polynomial p_α of $\alpha \in B$ is a quadratic polynomial and

$$p_\alpha(x) = x^2 - \mathrm{tr}(\alpha)x + N(\alpha).$$

Proposition 3.1.6. *Let $B = \left(\frac{a,b}{F}\right)$. Then the reduced norm and trace and of $\alpha \in B$ are given by*

$$N(\alpha) = \alpha\overline{\alpha},$$
$$\mathrm{tr}(\alpha) = \alpha + \overline{\alpha}.$$

Explicitly, for $x, y, z, w \in F$, we have

$$N(x + yi + zj + wk) = x^2 + ay^2 + bz^2 + abw^2,$$
$$\mathrm{tr}(x + yi + zj + wk) = 2x.$$

This can be proved by simple calculation using Examples 3.1.3 and 3.1.4.

Exercise 3.1.9. Prove the above proposition.

Exercise 3.1.10. Show $\mathrm{tr}(\overline{\alpha}) = \mathrm{tr}(\alpha)$ and $N(\overline{\alpha}) = N(\alpha)$.

Thus the relationship between norm, trace and canonical involution for quaternion algebras is analogous to the relationship between norm, trace and Galois conjugation for quadratic fields. In fact, often reduced norm and trace are defined by their expressions in terms of the canonical involution. (This only works for quaternion algebras though, not general CSAs.) Next we show canonical involution and Galois conjugation of quadratic subfields are compatible on B.

Lemma 3.1.7. *Let K be a quadratic subfield of B. Then the canonical involution $\alpha \mapsto \overline{\alpha}$ restricted to K acts as the nontrivial Galois automorphism of K/F.*

Proof. Say $K = F(\alpha)$ where $\alpha^2 = a \in F^\times$. Note $\alpha\overline{\alpha} = N(\alpha) = N(\overline{\alpha}) = \overline{\alpha}\alpha$, i.e. α and $\overline{\alpha}$ commute. Since $C_B(K) = K$, this means $\overline{\alpha} \in K$. By the fact that $\alpha \mapsto \overline{\alpha}$ is a canonical involution, it restricts to an involution on K, which must be an automorphism as K is commutative. This automorphism is nontrivial, but fixes F. □

We now to justify the use of the word "canonical."

Proposition 3.1.8. *Suppose there is an isomorphism of quaternion algebras $\phi : \left(\frac{a,b}{F}\right) \xrightarrow{\sim} \left(\frac{a',b'}{F}\right)$. Then it respects the canonical involution, i.e., $\phi(\overline{\alpha}) = \overline{\phi(\alpha)}$.*

Proof. Note $\overline{\alpha} = \mathrm{tr}\alpha - \alpha$. Since ϕ is an isomorphism, α and $\phi(\alpha)$ have the same reduced characteristic polynomials. (Either $\alpha \in F$ and $p_\alpha = (x-\alpha)^2$, or $\alpha \notin F$ and the characteristic polynomial is the same as the minimal polynomial.) In particular $\mathrm{tr}\phi(\alpha) = \mathrm{tr}\alpha$. Thus

$$\phi(\overline{\alpha}) = \phi(\mathrm{tr}\alpha - \alpha) = \mathrm{tr}\phi(\alpha) - \phi(\alpha) = \overline{\phi(\alpha)}.$$

□

This means that for any quaternion algebra B (not given a priori as $\left(\frac{a,b}{F}\right)$), we can define the canonical involution on B by fixing an isomorphism $B \xrightarrow{\sim} \left(\frac{a,b}{F}\right)$ for some (a, b) and pulling back the involution on $\left(\frac{a,b}{F}\right)$, and this definition does not depend upon the choice of (a, b) or on the choice of the isomorphism. More directly, we can just define the canonical involution on any quaternion algebra B by

$$\overline{\alpha} = \mathrm{tr}\alpha - \alpha.$$

3.2 Quadratic forms

One way to classify quaternion algebras is by using quadratic forms. This approach is taken in [Vig80] and [MR03]. This is more efficient than classifying CSAs over number fields as described in Section 2.7 and specializing to quaternion algebras. In this section we will review some basic theory of quadratic forms and explain how quadratic forms are related to quaternion algebras. We won't include much in the way of proofs, but what we will need can be found in any standard reference for quadratic forms. However, nothing we do in this section is difficult—it is mostly just learning terminology. You should be able to fill in all details yourself. (Though later we will quote local-global results without proof, which are nontrivial.)

Some general references for quadratic forms are Serre [Ser73], Gerstein [Ger08], Cassels [Cas78], O'Meara [O'M00], and Shimura [Shi10] (ordered roughly by my level of familiarity with them). Another standard reference is Lam [Lam05], though this focuses on rationality questions rather than integrality questions. In fact, much of what we need can be found in a good linear algebra book like Hoffman–Kunze [HK71]. There's also a short chapter on general quadratic forms in my Number Theory II notes [Marb] if you just want a quick introduction, with a bit of a different perspective than the one here (a large part of that course focused on binary quadratic forms).

Quadratic and bilinear forms

Let R be a Dedekind domain, e.g., the ring of integers of a number field or p-adic field.[3] Let F be the field of fractions of R. As before, we assume char $F \neq 2$.

Let M be a free R-module of finite rank. A **(R-)bilinear form** on M is a map

$$\phi : M \times M \to F$$

which is R-linear in each variable, i.e., $\phi(rx+x', y) = r\phi(x,y) + \phi(x', y)$ and $\phi(x, ry+y') = r\phi(x,y) + \phi(x,y')$ for all $r \in R$, $x, x', y, y' \in M$.[4] We say ϕ is **symmetric** if $\phi(x,y) = \phi(y,x)$ for all $x, y \in M$.

If $\{e_1, \dots, e_n\}$ is a basis for M, we can associate to a bilinear form ϕ the matrix $A = (a_{ij}) \in M_n(F)$ where $a_{ij} = \phi(e_i, e_j)$. Then if $x = \sum x_i e_i$ and $y = \sum y_j e_j$, we can express ϕ in terms of matrix multiplication by

$$\phi(x,y) = xAy := \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} A \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

It is clear that A is symmetric if and only if ϕ is. If a basis is understood, sometimes we abuse terminology and call A the bilinear form.

Let ϕ be a symmetric bilinear form on M. The map $Q : M \to F$ given by

$$Q(x) = \phi(x,x)$$

is the **quadratic form** (over R) associated to ϕ. We call the pair (M, Q) (or (M, ϕ)) a **quadratic R-module**. If $R = F$ is a field so $V = M$ is a vector space, we call (V, Q) (or (V, ϕ)) a **quadratic space**.

> **Example 3.2.1.** If $R = F = \mathbb{R}$ and $V = M = \mathbb{R}^n$, then a symmetric bilinear form on V is the same as an inner product. For instance, $\phi(x,y) = x \cdot y$ (the standard dot product) a symmetric bilinear form and $Q(x) = x \cdot x = \|x\|^2 = \sum x_i^2$ is just the square of the Euclidean norm.

It is clear that $Q(rx) = r^2 Q(x)$ for $r \in R$, $x \in M$. Explicitly with respect to the basis $e_1, \dots, e_n$,

$$Q(x) =: Q(x_1, \dots, x_n) = xAx = \sum_{i=1}^{n} a_{ii} x_i^2 + 2 \sum_{1 \le i < j \le n} a_{ij} x_i x_j.$$

In other words, after fixing a basis, we can simply view Q as a homogenous quadratic polynomial on $R^n \simeq M$ over F. If one prefers, one could take this as the definition of quadratic form.

We call A a matrix for Q. Any matrix for Q with respect to another basis will be similar to A.

[3] See the beginning of Chapter 4 for a brief recollection of what Dedekind domain is.

[4] One often denotes bilinear forms by B, but we are using that symbol for quaternion algebras. I'm sure I will find some conflict with ϕ later, at which point I may switch to $\langle \cdot, \cdot \rangle$ for my bilinear form.

Exercise 3.2.1. Check $\phi(x,y) = \frac{Q(x+y)-Q(x)-Q(y)}{2}$ for all $x, y \in M$.

This says that a bilinear form determines a quadratic form and conversely (This is not true in characteristic 2.)

The **dimension** of a quadratic form Q is $\dim Q = \dim_R M$. In light of the polynomial point of view, if $\dim Q = n$, we also say Q is an n**-ary** quadratic form, or a quadratic form in n **variables**. In small dimensions, I will sometimes use x, y, etc. for elements of R or F rather than M and write quadratic forms as $Q(x,y) = x^2 + y^2$ say, rather than the more cumbersome $Q(x) = Q(x_1, x_2) = x_1^2 + x_2^2$. When $n = 2$, 3, or 4, we call Q **binary**, **ternary**, or **quaternary**, respectively.

The **discriminant** (or **determinant**) of $\operatorname{disc} Q$ is $\det A$, where A is a symmetric matrix representing Q as above.

Let ϕ be a symmetric bilinear form on M given by a matrix A in a basis $\{e_1, \dots, e_n\}$, and Q the associated quadratic form. We say ϕ or Q is **non-degenerate** if $\phi(x_0, y) = 0$ for all y implies $x_0 = 0$.[5] This is equivalent to A being invertible, i.e., $A \in \mathrm{GL}_n(F)$. One can typically restrict to working with non-degenerate quadratic forms just by restricting to a submodule on which your form is non-degenerate.

Example 3.2.2. Let $R = \mathbb{Z}$ so $F = \mathbb{Q}$. If $M = \mathbb{Z}^n$ and ϕ is the (non-degenerate) bilinear form associated to $A = I$, the identity matrix, then the quadratic form $Q : \mathbb{Z}^n \to \mathbb{Q}$ is just the sum of squares:

$$Q(x) = Q(x_1, \dots, x_n) = xIx = x_1^2 + \cdots + x_n^2.$$

(Essentially the same quadratic form in Example 3.2.1, but now over $\mathbb{Z}$.) More generally, if $A = \operatorname{diag}(a_1, \dots, a_n)$, then $Q(x_1, \dots, x_n) = a_1 x_1^2 + \cdots + a_n x_n^2$. A quadratic form associated to a diagonal matrix A is called a **diagonal form**. This diagonal form has discriminant $\operatorname{disc} Q = a_1 a_2 \cdots a_n$.

A classical problem in number theory (say, when $R = \mathbb{Z}$ or $R = \mathbb{Q}$) is to determine what numbers are **represented** by a quadratic form Q, i.e., for which $a \in F$, $Q(x) = a$ has a solution. In the above example this just means which numbers are expressible as the sum of n integer squares.[6]

A special case of representation problems is understanding the solutions to $Q(x) = 0$. We say Q is **isotropic** if there exists $x \neq 0$ in M such that $Q(x) = 0$; otherwise Q is **anisotropic**. Just by positivity, the example of a sum of n squares $x_1^2 + \cdots + x_n^2$ is anisotropic—however on $\mathbb{Z}[i]^n$ or $\mathbb{C}^n$ it is isotropic (for $n > 1$).[7]

[5] Some authors call degenerate singular, and non-degenerate non-singular or regular.

[6] More generally, one can look at *representation numbers* $r_Q(n)$, which is the number of solutions to $Q(x) = n$. When $R = \mathbb{Z}$, the numbers $r_Q(n)$ are Fourier coefficients of a modular form, and one can use modular forms to study these numbers. See, for instance, my Modular Forms notes [Mara].

[7] Over $\mathbb{C}$, we can consider forms related to inner products, i.e., *skew-symmetric* linear forms rather than symmetric linear forms. This leads to the notion of **Hermitian forms**. An example on $\mathbb{C}^n$ is $|x_1|^2 + \cdots + |x_n|^2$.

More generally, we have the following notions of positivity and negativity for quadratic forms. Suppose $F \subset \mathbb{C}$. We say Q is **positive definite** (resp. **negative definite**) if $Q(x) > 0$ (resp. $Q(x) < 0$) for all $x \in M - \{0\}$. If Q is positive or negative definite, we say Q is **definite**, and **indefinite** otherwise. It is easy to see that for Q to be definite, we need both $F \subset \mathbb{R}$ and Q to be anisotropic.

If Q is a non-degenerate diagonal form $a_1x_1^2 + \cdots + a_nx_n^2$ with $F \subset \mathbb{R}$, then positive (resp. negative) definite just means $Q(x) \geq 0$ (resp. $Q(x) \leq 0$) for all x because positivity (resp. negativity) will imply anisotropy. Of course Q being positive definite just means each $a_i > 0$ and negative definite means each $a_i < 0$. Thus if Q is indefinite, it takes on both positive and negative values. On the other hand, if Q is degenerate, e.g., $Q(x, y) = x^2 + 0 \cdot y^2$ it can be indefinite but need not take on negative values.[8]

Here are a couple of examples.

> **Example 3.2.3.** Let $R = \mathbb{Z}$, $M = R^2$ and consider the non-degenerate bilinear form ϕ given by the matrix $A = \begin{pmatrix} 1 & -\frac{3}{2} \\ -\frac{3}{2} & 1 \end{pmatrix}$ with respect to the basis $e_1 = (1, 0)$, $e_2 = (0, 1)$. Then
>
> $$Q(x, y) = x^2 - 3xy + y^2.$$
>
> Note there is a difference between the coefficients of $Q(x, y)$ (as a quadratic polynomial) being integral and the coefficients of the matrix A being integral. This is indefinite, but anisotropic as can be seen by the quadratic formula. However this form becomes isotropic over $\mathbb{Z}[\sqrt{5}]$.

The next example is fundamental.

> **Example 3.2.4.** Let $R = F$, $V = M = F^2$. Then $Q(x, y) = x^2 - y^2$ is isotropic, e.g., $Q(1, 1) = 0$. The quadratic space (V, Q) is called the **hyperbolic plane**[9], and is a fundamental quadratic space.

Note the quadratic form for the hyperbolic plane *splits* as a product $Q(x, y) = (x - y)(x + y)$. More generally one can construct isotropic forms by taking the product of two linear forms with distinct zeroes.

> **Exercise 3.2.2.** Let $M = M_2(R)$ and $Q(x) = \det x$. Determine explicitly the bilinear form associated to Q.

[8] Usually one would call forms like $x^2 + 0 \cdot y^2$ *positive semidefinite*. Often one reserves "indefinite" for forms which actually take on both positive and negative values, but we will just be concerned with non-degenerate forms, in which case these two notions of indefiniteness agree.

[9] The terminology hyperbolic plane for a quadratic space is not, as far as I know, directly related to the hyperbolic plane which arises in hyperbolic geometry (i.e., the upper half-plane with the hyperbolic metric). Rather, the equation $Q(x, y) = x^2 - y^2 = c$, so the level sets of Q on the hyperbolic plane are just hyperbolas. In hyperbolic geometry, $x_1^2 + \cdots + x_{n-1}^2 - x_n^2 = 1$ defines an $(n-1)$-dimensional hyperboloid, and in particular $x^2 + y^2 - z^2$ can be used to give a model for the hyperbolic hyperbolic plane.

Exercise 3.2.3. Let $R = F$ and $M = \mathcal{A}$ be a CSA over F. Define the **trace form** on $\mathcal{A}$ by $T(x, y) = \mathrm{tr}(xy)$.

(i) Show T is a symmetric F-bilinear form on $\mathcal{A}$.

(ii) Show that T is nondegenerate.

(iii) If $\mathcal{A} = \left(\frac{a,b}{F}\right)$, compute the trace form and its associated quadratic form with respect to the basis $\{1, i, j, k\}$.

Exercise 3.2.4. Let $R = F$ and $M = \mathcal{A}$ be a CSA over F of degree n. Show that the reduced norm $N : \mathcal{A} \to F$ is a quadratic form if and only if $n = 2$.

Isometry groups and equivalence

In light of (3.2.1), we often think of $Q(x-y)$ as providing a measure of the "distance" between x and y on a quadratic space (V, Q). For isotropic forms, this clearly cannot be used to define a metric because multiple points would have distance 0 from 0. Still, we sometimes think of Q as providing some sort of geometry on V. This leads to the terminology that isomorphisms of quadratic spaces or modules preserving a quadratic form are called *isometries*.

Let M and M' be free R-modules of finite rank and ϕ and ϕ' be symmetric bilinear forms on M and M' with associated quadratic forms Q and Q'. A linear map $L : M \to M'$ is an **isometry** of (M, ϕ) with (M', ϕ') if it is an R-module isomorphism such that $\phi'(Lx, Ly) = \phi(x, y)$ for all $x, y \in M$, or equivalently (cf. Exercise 3.2.1) such that $Q'(Lx) = Q(x)$. In this case, we call the quadratic forms **equivalent** and write $Q \simeq Q'$.

Exercise 3.2.5. With notation above, let A and A' be matrices for ϕ and ϕ' with respect to some bases. Show $Q \simeq Q'$ if and only if $A' = {}^t g A g$ for some $g \in \mathrm{GL}_n(R)$.

This says that two quadratic forms are equivalent one can be obtained from the other by an invertible linear change of variables.

Example 3.2.5. Suppose $R = \mathbb{Z}$ and $Q(x, y) = ax^2+bxy+cy^2$, $Q'(x, y) = a'x^2+b'xy+c'y^2$ are binary quadratic forms. Then Q and Q' are given by matrices $A = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$ and $A' = \begin{pmatrix} a' & b'/2 \\ b'/2 & c \end{pmatrix}$. The above exercise say $Q \equiv Q'$ if and only if

$$\begin{pmatrix} a' & b'/2 \\ b'/2 & c \end{pmatrix} = {}^t g \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} g$$

for some $g \in \mathrm{GL}_2(\mathbb{Z})$. For instance, take $Q(x, y) = 2x^2 + y^2$ and $Q'(x, y) = 2x^2+4xy+3y^2$. Then $Q \simeq Q'$ because

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

The next exercise deals with some classical theory of binary quadratic forms, in case you have not seen it before (you can also see my Number Theory II notes [Marb], for instance).

Exercise 3.2.6. Let $Q(x,y) = ax^2 + bxy + cy^2$ be a binary quadratic form over $R = \mathbb{Z}$.

(i) Show $\Delta := b^2 - 4ac = -4 \operatorname{disc} Q$ and that $\Delta > 0$ if and only if Q is definite.

(ii) Suppose Q is positive definite. We say Q is *reduced* if $|b| \le a \le c$ and $b \ge 0$ if $a = |b|$ or $a = c$. Show any Q is equivalent to a reduced form—in fact it is *properly equivalent* to a reduced form, i.e., equivalent to one by a change of variables matrix $g \in \mathrm{SL}_2(\mathbb{Z})$ not just $g \in \mathrm{GL}_2(\mathbb{Z})$.

(iii) Conclude that the set of proper equivalence classes $\mathrm{Cl}(\Delta)$ of binary quadratic forms with discriminant $-\frac{\Delta}{4}$ is finite for $\Delta > 0$. (In fact, Gauss showed $\mathrm{Cl}(\Delta)$ as the structure of a finite abelian group, called the form class group. If Δ is a fundamental discriminant, i.e., the discriminant of an imaginary quadratic field $K/\mathbb{Q}$, then $\mathrm{Cl}(\Delta) \simeq \mathrm{Cl}(\mathcal{O}_K)$.

For representation problems, i.e., which numbers are represented by Q, it suffices to consider forms up to equivalence as equivalent forms $Q : M \to F$, $Q' : M' \to F$ must have the same image.

Consequently, many of the properties of quadratic forms we discussed are invariant under isometry: dimension, isotropy, definiteness. The exercise also implies being non-degenerate is an invariant. However, the discriminant is not an invariant of equivalences—but it is up to squares. Namely, if $A' = {}^t g A g$, then $\det A' = (\det g)^2 \det A$, so $\operatorname{disc} Q / \operatorname{disc} Q' \in R^{\times 2}$ if $Q \simeq Q'$. (Recall $R^{\times 2}$ denotes the squares in $R^\times$.) Conversely, given Q and any $\lambda \in R^\times$, there exists a $Q' \equiv Q$ such that $\operatorname{disc} Q' = \lambda^2 \operatorname{disc} Q$.

For the rest of this section, we just work with quadratic forms over fields, i.e., $R = F$.

Let (V, Q) be a quadratic space of dimension n over a field F with bilinear form ϕ. If Q has some property such as being non-degenerate or isotropic, we say the same for (V, Q). However, if (V, Q) and (V', Q') with $Q \simeq Q'$, we say (V, Q) and (V', Q') are **isometric**, rather than equivalent, but also write $(V, Q) \simeq (V', Q')$.

Let W be an F-linear subspace of V. Then, by restriction (W, Q) is also a quadratic space, which we call a quadratic subspace. For $v, w \in V$, say v and w are **orthogonal** if $\phi(v, w) = 0$. Define the **orthogonal complement** of W by

$$W^\perp = \{v \in V : \phi(v, w) = 0 \text{ for all } w \in W\}.$$

Since ϕ is bilinear, $W^\perp$ is also a subspace. Note that $V^\perp = 0$ if and only if (V, Q) is non-degenerate.

Exercise 3.2.7. Let (V, Q) be a quadratic space and (W, Q) a quadratic subspace. Suppose (W, Q) is non-degenerate. Show $V = W \oplus W^\perp$.

Exercise 3.2.8. Let (V, Q) be a isotropic non-degenerate quadratic space of dimension ≥ 2. Show (V, Q) contains a hyperbolic plane, i.e., a subspace isometric to the hyperbolic plane from Example 3.2.4.

> **Exercise 3.2.9.** Show the hyperbolic plane is **universal**, i.e., the quadratic form represents every element of F. Conclude any non-degenerate isotropic form in dimension ≥ 2 is universal. (Note: anisotropic forms and degenerate isotropic forms may also be universal—e.g., $Q(x) = x^2$ on $\mathbb{C}$.)

Over fields (characteristic not 2), we can reduce our study to diagonal forms.

Theorem 3.2.1. *Let (V, Q) be a quadratic space over F. Then Q is equivalent to a diagonal form.*

Proof. With the above concept of orthogonality, one just does the Gram–Schmidt process to find an orthogonal basis to make Q diagonal. You can do this as an exercise for yourself or [HK71, Thm 10.3]. □

The classification of quadratic forms over $\mathbb{R}$ and $\mathbb{C}$ is well known and was proven by the mid 19th century. We will not need this, but it is something every mathematician should know.

Theorem 3.2.2 (Sylvester's law of inertia)**.** *Any non-degenerate quadratic form in n variables over $\mathbb{C}$ is equivalent to $x_1^2 + \cdots + x_n^2$. The equivalence classes of non-degenerate quadratic forms over $\mathbb{R}$ are uniquely represented by the forms $x_1^2 + \cdots + x_m^2 - x_{m+1}^2 - \cdots - x_n^2$ $(0 \leq m \leq n)$.*

Proof. The statement for $\mathbb{C}$ is a simple consequence of Theorem 3.2.1, because we can transform $\sum a_i x_i^2$ into $\sum x_i^2$ by the transformation $\sqrt{a_i} x_i \mapsto x_i$. One half of the statement for $\mathbb{R}$ is similar, except now one needs to account for ±'s as not all elements of $\mathbb{R}$ have real square roots. Show the above forms are inequivalent is the main point—you can try it as an exercise yourself or see [HK71, Sec 10.2]. The case of $n = 3$, which is most relevant to quaternions, is below in Exercise 3.3.4. □

It is called a law of inertia because it says that, over $\mathbb{R}$, the number of $+1$'s and -1's in a diagonal representation of the form are invariant under a change of basis. Really this is what is called Sylvester's law of inertia. I included the classification over $\mathbb{C}$ in the same theorem for convenience. In the law of inertia, the number of $+1$'s on the diagonal minus the number of -1's, i.e., $m - (n - m) = 2m - n$ is called the **signature** of Q, and this theorem says the signature determines Q up to equivalence.

Definition 3.2.3. *Let (V, Q) be a quadratic space over a field F. The* **orthogonal group** $\mathrm{O}(V)$ *of V is the group of isometries of (V, Q) with itself. The* **special orthogonal group** $\mathrm{SO}(V)$ *is the subgroup of* $\mathrm{O}(V)$ *consisting of transformations with determinant* 1.

(One might write $\mathrm{O}(V, Q)$ or $\mathrm{SO}(V, Q)$ to be precise, but it's customary to suppress the form Q in the orthogonal group notation.)

If Q is associated to a matrix A with respect to a basis $\{e_1, \ldots, e_n\}$, then we have

$$\mathrm{O}(V) \simeq \left\{ g \in \mathrm{GL}_n(F) : {}^t g A g = A \right\}$$

and

$$\mathrm{SO}(V) \simeq \left\{ g \in \mathrm{GL}_n(F) : {}^t g A g = A, \, \det g = 1 \right\}.$$

Example 3.2.6. Let $V = F$ and $Q(x) = x^2$. Then $\mathrm{O}(V) = \{\pm 1\}$ and $\mathrm{SO}(V) = 1$. That is, there are two isometries of F with itself—the identity and scaling by -1 (an "orientation-reversing" reflection), since the quadratic form does not see multiplication by -1. In fairly general settings, the special orthogonal group can be viewed as the group of "orientation-preserving" isometries.

Example 3.2.7. When $F = \mathbb{R}$ and $A = I$ is the identity matrix, $\mathrm{SO}(V) \simeq \mathrm{SO}(n)$ as defined in Remark 2.4.2. More generally, if $F = \mathbb{R}$ and Q is the non-degenerate quadratic form $x_1^2 + \cdots + x_m^2 - x_{m+1}^2 - \cdots - x_n^2$, then $\mathrm{SO}(V)$ is isomorphic to the classical Lie group

$$\mathrm{SO}(m, n-m) = \left\{ g \in \mathrm{GL}_n(F) : {}^t g \begin{pmatrix} I_m & \\ & -I_{n-m} \end{pmatrix} g = \begin{pmatrix} I_m & \\ & -I_{n-m} \end{pmatrix}, \det g = 1 \right\}.$$

(So $\mathrm{SO}(n,0) = \mathrm{SO}(n)$.) It is well known that $\mathrm{SO}(m, n-m) = \mathrm{SO}(m', n-m')$ if and only if $m' = m$ or $m' = n - m$.

Example 3.2.8. Let $V = F^n$ and $Q(x) = 0$ be the (degenerate) quadratic form which is identically 0. Then $\mathrm{O}(V) \simeq \mathrm{GL}_n(F)$ and $\mathrm{SO}(V) = \mathrm{SL}_n(F)$.

However, normally when one talks about orthogonal groups of quadratic forms, one restricts to non-degenerate forms. Ooh, look at me, I'm all high and mighty, I don't mix with degenerates.

Exercise 3.2.10. Show $\mathrm{SO}(2) = \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R}/2\pi\mathbb{Z} \right\}$.

Exercise 3.2.11. Compute explicitly $\mathrm{SO}(V)$ where (V, Q) is the hyperbolic plane over F.

Exercise 3.2.12. Suppose (V, Q) is non-degenerate. Show $g \in \mathrm{O}(V)$ implies $\det g = \pm 1$. Conclude $\mathrm{SO}(V)$ has index at most 2 in $\mathrm{O}(V)$.

3.3 Norm forms

Consider a quaternion algebra B over F. It is clear (B, N) is a quadratic space over F, where N is the reduced norm.[10] In other words, N defines a quaternary quadratic form over F. To emphasize that N is a quadratic form, we often call it the *norm form*.

[10] You might see the phrase (B, N)-pair in algebraic groups or representation theory. It has nothing to do with what we are talking about. There B and N are something like a Borel subgroup and a normalizer of the diagonal, e.g., for GL(2) one can take $B = \left\{ \begin{pmatrix} * & * \\ & * \end{pmatrix} \right\}$ and $N = \left\{ \begin{pmatrix} * & \\ & * \end{pmatrix}, \begin{pmatrix} & * \\ * & \end{pmatrix} \right\}$.

We'll go back to using ϕ for things which are not a bilinear form. Let's denote the bilinear form for N by

$$\langle \alpha, \beta \rangle = \frac{N(\alpha+\beta) - N(\alpha) - N(\beta)}{2} = \frac{\alpha\overline{\beta} + \beta\overline{\alpha}}{2} = \frac{1}{2}\mathrm{tr}(\alpha\overline{\beta}).$$

Explicitly, if $B = \left(\frac{a,b}{F}\right)$, Proposition 3.1.6 implies

$$\langle x + yi + zj + wk, x' + y'i + z'j + w'k \rangle = xx' - ayy' - bzz' + abww'.$$

In particular, $1, i, j, k$ is an orthogonal basis for B, i.e., the vector space decomposition $B \simeq F \oplus Fi \oplus Fj \oplus Fk$ is really a decomposition into lines which are orthogonal with respect to the norm form.

Proposition 3.3.1. *Two quaternion algebras B and B' over F are isomorphic if and only if the quadratic spaces (B, N) and (B', N) are isometric, i.e., if and only if the norm forms $N_{B/F}$ and $N_{B'/F}$ are equivalent.*

Proof. If $\phi : B \to B'$ is an isomorphism, then it induces an isometry as α and $\phi(\alpha)$ must have the same characteristic polynomial for any $\alpha \in B$ by Skolem–Noether.

The other direction will follow from Proposition 3.3.2. □

> **Example 3.3.1.** Suppose $B \simeq M_2(F)$ is the split quaternion algebra. Then the norm form is equivalent to the isotropic form $Q(x, y, z, w) = x^2 - y^2 - z^2 + w^2$ by Example 3.1.2.

Suppose $F \subset \mathbb{C}$. We call B **definite** (resp. **indefinite**) its norm form is definite (resp. indefinite). Writing $B = \left(\frac{a,b}{F}\right)$, we see the norm form $x^2 - ay^2 - bz^2 + abw^2$ is definite if and only if $F \subset \mathbb{R}$ and $a < 0$ and $b < 0$. If B is definite, then the norm form must be positive definite and thus anisotropic.

> **Example 3.3.2.** Over $F = \mathbb{R}$, there are two quaternion algebras up to isomorphism: $M_2(\mathbb{R})$ and $\mathbb{H}$, with norm forms given by $x^2 - y^2 - z^2 + w^2$ and $x^2 + y^2 + z^2 + w^2$, i.e., the quaternary real quadratic forms of signature 0 and 4. The former is indefinite and isotropic, whereas the latter is positive definite and anisotropic. In particular $M_2(\mathbb{R})$ is indefinite and $\mathbb{H}$ is definite.

By Sylvester's law of inertia, not all equivalence classes of real quaternary quadratic forms are norm forms of quaternion algebras, e.g., the "hyperbolic" form $x^2 + y^2 + z^2 - w^2$ of signature 2 is not. On the other hand, we will see below that there is one-to-one correspondence between classes of ternary quadratic forms and classes of quaternion algebras.

Pure quaternions

Let B_0 denote the set of trace 0 elements in B, i.e., those of the form $xi + yj + zk$. The elements in B_0 are called **pure quaternions**. Equivalently, $B_0 = \{\alpha \in B : \overline{\alpha} = -\alpha\}$. This is a clear analogue of the purely imaginary numbers $iy \in \mathbb{C}$. We make B_0 a quadratic space (B_0, N_0) by restricting the norm map. We call N_0 the **restricted norm form**.

Explicitly, if $B = \left(\frac{a,b}{F}\right)$, then writing $B_0 = \{xi + yj + zk : x, y, z \in F\}$ and

$$N_0(xi + yj + zk) = -ax^2 - by^2 + abz^2$$

is a ternary quadratic form. If B is definite, this ternary form is positive definite and anisotropic; if B is indefinite, the ternary form is indefinite. The associated bilinear form is

$$\langle xi + yj + zk, x'i + y'j + z'k\rangle = -axx' - byy' + abzz'.$$

Note that we can also describe B_0 as the orthogonal complement $F^\perp$ of F in the quadratic space (B, N). Further B_0 breaks up as the direct sum of 3 pairwise orthogonal lines $Fi \oplus Fj \oplus Fk$. Orthogonality can be expressed in terms of a nice multiplicative criterion:

> **Exercise 3.3.1.** Let B be a quaternion algebra over F, and $\alpha, \beta \in B_0$. Show $\alpha\beta = -\beta\alpha$ if and only if α and β are orthogonal in (B_0, N_0). Deduce that if α, β are orthogonal, then $\alpha\beta$ is orthogonal to α and β.

Proposition 3.3.2. *Two quaternion algebras B and B' over F are isomorphic if and only if the 3-dimensional quadratic spaces (B_0, N_0) and (B'_0, N_0) of pure quaternions are isometric.*

Proof. Suppose first that $\phi : B \to B'$ is an isomorphism. By Proposition 3.3.1, then the quadratic spaces (B, N) and (B', N) are isometric. Since ϕ takes F to F, being an isometry implies it take $F^\perp = B_0$ to $F^\perp = B'_0$, and thus gives an isometry of (B_0, N_0) with (B'_0, N_0).

Now suppose $\phi : (B_0, N_0) \to (B'_0, N_0)$ is an isometry. We can extend ϕ to a linear map $B = F \oplus B_0 \to B' = F \oplus B'_0$ by sending 1 to 1. This is an isometry $(B, N) \to (B', N)$ by orthogonality.

Identify $B = \left(\frac{a,b}{F}\right)$. Let $i' = \phi(i)$, $j' = \phi(j)$ and $k' = \phi(k)$. As ϕ is an isometry, i', j' and k' are pairwise orthogonal and $B'_0 = Fi' \oplus Fj' \oplus Fk'$. Note for any $\alpha \in B_0$ or B'_0, $N_0(\alpha) = \alpha\overline{\alpha} = -\alpha^2$. Hence

$$(i')^2 = \phi(i)^2 = -N_0(\phi(i)) = -N_0(i) = -i^2 = a.$$

Similarly $(j')^2 = b$ and $(k')^2 = -ab$. Further, by Exercise 3.3.1, $i'j' = -j'i'$ and $i'j'$ is orthogonal to i' and j', thus $i'j' = ck'$ for some $c \in F$. Taking norms gives

$$N(ij) = N(i'j') = c^2N(k') = c^2N(k),$$

so $c = \pm 1$. As described in the construction of the algebra Hilbert symbol, these relations on i', j' and k' determine B', and $1 \mapsto 1$, $i \mapsto i'$, $j \mapsto j'$, $k \mapsto ck'$ defines an isomorphism of B with B'. □

Corollary 3.3.3. *For $a, b \in F^\times$, $\left(\frac{a,b}{F}\right) \simeq \left(\frac{a,-ab}{F}\right)$.*

This can be proved directly from the definition of $\left(\frac{a,b}{F}\right)$, and basically tells us that we can replace j with k. However, the argument is even easier using restricted norms.

Proof. By the proposition, it suffices to show $-ax^2 - by^2 + abz^2$ is equivalent to $-ax^2 + aby^2 - a^2bz^2$. Simply change variables for the latter form by sending z to $a^{-1}z$ and interchange y and z. □

Exercise 3.3.2. Show that if $(B, N) \simeq (B', N)$, then $(B_0, N) \simeq (B'_0, N)$. Use this to finish the proof of Proposition 3.3.1.

Example 3.3.3. Let $F = \mathbb{R}$ and $B = M_2(F)$ or $B = \mathbb{H}$. Then the restricted norm form N_0 is equivalent to $Q_1 : -x^2 - y^2 + z^2$ or $Q_2 : x^2 + y^2 + z^2$. By Sylvester's law of inertia, any non-degenerate real ternary quadratic form is equivalent to $\pm Q_1$ or $\pm Q_2$.

Another way to state the above example is that the quaternion algebras over $\mathbb{R}$ correspond to non-degenerate real ternary quadratic forms with positive discriminant. This generalizes to the following classification in terms of ternary forms.

Theorem 3.3.4. *There is a 1-1 correspondence between isomorphism classes of quaternion algebras over F and equivalence classes of non-degenerate ternary quadratic forms over F with square discriminant. This correspondence is given by $B \mapsto N_0$.*

Proof. Recall that though the discriminant is not invariant under isometry, the property of the discriminant being square is.

Given any quaternion algebra B, it is isomorphic to some $\left(\frac{a,b}{F}\right)$, which has restricted norm $-ax^2 - by^2 + abz^2$. This is non-degenerate and has discriminant $(ab)^2$. Thus by Proposition 3.3.2 it suffices to show the correspondence $B \mapsto N_0$ is surjective on classes.

Let Q be a non-degenerate ternary quadratic form with square discriminant. By Theorem 3.2.1, Q is equivalent to a diagonal form, say $-ax^2 - by^2 + cz^2$, which has discriminant $abc = \lambda^2$. Hence we can write the form as $-ax^2 - by^2 + \frac{\lambda^2}{ab}z^2$. Now making the change of variable $z' = \frac{\lambda}{ab}z$, we see Q is equivalent to $-ax^2 - by^2 + abz^2$, the norm form of $\left(\frac{a,b}{F}\right)$. □

We can rephrase this correspondence without the discriminant condition by using a weaker notion of equivalence of quadratic forms. Let us say quadratic forms Q_1 and Q_2 over F are **similar** if $Q_2 \simeq \lambda Q_1$ for some $\lambda \in F^\times$. We remark that if Q_1 and Q_2 are similar, they may not represent the same numbers, but it is easy to determine the numbers represented by Q_2 in terms of the numbers represented by Q_1, as they just differ by some scalar λ. In particular Q_1 is isotropic if and only if Q_2 is.

Corollary 3.3.5. *There is a 1-1 correspondence between isomorphism classes of quaternion algebras over F and similarity classes of non-degenerate ternary quadratic forms over F induced by the map $B \mapsto N_0$.*

Exercise 3.3.3. Prove this corollary.

Exercise 3.3.4. Use Frobenius' theorem to deduce the $n = 3$ case of Sylvester's law of inertia.

The correspondence with ternary quadratic forms also provide us a way to understand the multiplicative group $B^\times$ of a quaternion algebra. Consider the **projective group** $PB^\times = B^\times/Z(B^\times) = B^\times/F^\times$. (For a vector space V, the projectivization PV is the space of lines in V, i.e., the space of non-zero vectors modulo scalars.) In particular, if $B = M_2(F)$, then $PB^\times = \mathrm{PGL}_2(F) = \mathrm{GL}_2(F)/F^\times$, the **projective linear group**.

Theorem 3.3.6. *Let B/F be a quaternion algebra. The action of $B^\times$ on B_0 by conjugation induces an isomorphism $PB^\times \simeq \mathrm{SO}(B_0)$.*

Proof. We don't actually need this result, so I'll refer to [Vig80, Thm I.3.3] or [MR03, Thm 2.4.1] for details. The proof in these references uses a theorem of Cartan which implies that every element in $\mathrm{O}(B_0)$ is a product of at most 3 reflections. This means every nontrivial element of $\mathrm{SO}(B_0)$ is a product of exactly 2 reflections, and one can construct these isometries with the action of $B^\times$.

However, I'll give you a different type of moral argument. First note for $\alpha \in B^\times$, the map $\beta \mapsto \alpha\beta\alpha^{-1}$ is an (inner) automorphism of B, and thus an isometry of B_0 with itself. These isometries arising from conjugation must have determinant 1. (If one wants, one can check this by a tedious calculation—for this it might help to recall that $\alpha^{-1} = N(\alpha)\overline{\alpha}$.) It is easy to see this map from $B^\times \to \mathrm{SO}(B_0)$ has kernel $Z(B^\times) = F^\times$. The determinant 1 isometries from B_0 to itself actually extend to algebra automorphisms $B \to B$ (see the end of the proof of Proposition 3.3.2, where $c = \pm 1$ is the determinant of the isometry). Such automorphisms must be inner by Skolem–Noether, thus the map $B^\times \to \mathrm{SO}(B_0)$ is surjective. □

> **Example 3.3.4.** We have the exceptional isomorphism $\mathrm{PGL}_2(F) \simeq \mathrm{SO}(1,2;F)$, the orthogonal group of the ternary diagonal form $-x^2 - y^2 + z^2$ over F. One often says PGL_2 is a split form of $\mathrm{SO}(3)$, meaning it is isomorphic to the special orthogonal group of *some* non-degenerate 3-dimensional quadratic space. The term split for an orthogonal group agrees with the terminology in Remark 2.4.2.

> **Example 3.3.5.** If $F = \mathbb{R}$ and $B = \mathbb{H}$, then $\mathbb{H}^\times/\mathbb{R}^\times \simeq \mathrm{SO}(3)$.

Geometrically, $\mathrm{SO}(3)$ is the group of rotations of $\mathbb{R}^3$ about the origin. Hamilton's quaternions provide a convenient and efficient way to study rotations in 3-space. The next exercises explore this a little.

> **Exercise 3.3.5.** Let $\mathbb{H}^\times$ act on $\mathbb{H}_0 = \mathbb{R}i \oplus \mathbb{R}j \oplus \mathbb{R}k$ via $\alpha\beta\alpha^{-1}$ for $\beta \in \mathbb{H}_0$ and $\alpha \in \mathbb{H}$. Show $\alpha = i$ and $\alpha = 1 + i$ correspond to a rotations about the i-axis in $\mathbb{H}_0$. Determine the angles of rotation. Interpret geometrically the identity $(1 + i)(1 + i) = 2i$.

> **Exercise 3.3.6.** Use multiplication in $\mathbb{H}$ to geometrically describe the actions of $\alpha_1 = 1+i$, $\alpha_2 = 1 + j$, and $\alpha = \alpha_1\alpha_2$ on $\mathbb{H}_0$ (action as in previous exercise).

Splitting criteria

A basic problem is to determine when two quaternion algebras $\left(\frac{a,b}{F}\right)$ and $\left(\frac{a',b'}{F}\right)$ are isomorphic. By Corollary 3.3.5, this reduces to checking if the quadratic forms $-ax^2 - by^2 + abz^2$ and $-a'x^2 - b'y^2 + a'b'z^2$ are similar (see Corollary 3.3.3 for one example). Scaling the first form by $\frac{1}{ab}$, making the substitutions $x \mapsto bx$, $y \mapsto ay$, and interchanging x and y shows the first form is similar to $-ax^2 - by^2 + z^2$. Thus it suffices to check if $-ax^2 - by^2 + z^2$ is similar to $-a'x^2 - b'y^2 + z^2$.

Let us now consider the special case that $\left(\frac{a',b'}{F}\right) \simeq M_2(F)$, so our question becomes: when is $\left(\frac{a,b}{F}\right)$ split?

Proposition 3.3.7. *Let $B = \left(\frac{a,b}{F}\right)$. The following are equivalent:*

(1) B is split, i.e., $B \simeq M_2(F)$;

(2) the norm form N is isotropic;

(3) the restricted norm form N_0 is isotropic;

(4) $ax^2 + by^2 = 1$ has a solution with $x, y \in F$; and

(5) either a is a square or $K = F(\sqrt{a})$ is a quadratic extension field and $b \in N_{K/F}(K^\times)$.

Proof. We have already observed the implication (1) $\implies$ (2) in Example 3.3.1 and (1) $\implies$ (3) is similarly obvious. In fact (1) and (2) are clearly equivalent because $N(\alpha) = 0$ for $\alpha \in B$ if and only if $\alpha \notin B^\times$. (3) $\implies$ (2) is trivial. Thus we have (1) $\iff$ (2) $\iff$ (3).

For (3) $\implies$ (4), note that (3) is equivalent to $ax^2 + by^2 = z^2$ having a nontrivial solution by the similarity of forms described above. If there is a nontrivial solution with $z \neq 0$, we can divide by z and change variables to get (4). If there is a nontrivial solution with $z = 0$, then we have $-\frac{a}{b}$ is a square, so change of variables shows $ax^2 + by^2 - z^2$ is equivalent to $a(x^2 - y^2) - z^2$, which has a (nontrivial) zero with $z \neq 0$ because the hyperbolic plane $x^2 - y^2$ is universal by Exercise 3.2.9. (One could also use Exercise 3.2.8.) Thus we still get (4).

For (4) $\implies$ (5), suppose a is not a square and (x, y) is a nontrivial solution as in (4). Then $b = (1/y)^2 - a(x/y)^2 = N_{K/F}(\frac{1}{y} + \frac{x}{y}\sqrt{a})$.

For (5) $\implies$ (1), we already observed this is the case if a is a square (Corollary 3.1.5), so assume a is not a square and $b = N_{K/F}(z_0 + x_0\sqrt{a}) = z_0^2 - ax_0^2$ for some $x_0, z_0 \in F$. Then $-ax_0^2 - b \cdot 1^2 + z_0^2 = 0$, i.e., the form $-ax^2 - by^2 + z^2$ is isotropic, so the restricted norm form N_0 being similar is also isotropic. Thus (5) $\implies$ (3) $\iff$ (1). □

Exercise 3.3.7. Let $a \in F^\times$. Show $\left(\frac{a,-a}{F}\right)$ and $\left(\frac{a,1-a}{F}\right)$ are split, assuming $a \neq 1$ in the latter case. Conclude that $\left(\frac{a,b}{F}\right)$ is split if $-\frac{a}{b} \in F^{\times 2}$. (The latter statement generalizes the fact that $\left(\frac{a,b}{\mathbb{R}}\right)$ is split if $ab < 0$.)

Condition (3) is often expressed in terms of the classical Hilbert symbol. Define the **(quadratic) Hilbert symbol** $(a, b)_F \in \{\pm 1\}$ to be $+1$ if $ax^2 + by^2 = z^2$ has a nontrivial solution over F, and -1 otherwise. Note if F is p-adic or a number field, this is $+1$ if and only if $ax^2 + by^2 = z^2$ has a nontrivial solution over $\mathcal{O}_F$, by clearing denominators.

We have the following connection between the quadratic Hilbert symbol and the classical quadratic residue (Legendre) symbol:

> **Exercise 3.3.8.** Let p be an odd prime and $a \in \mathbb{Z} - p\mathbb{Z}$. Show
>
> $$(a, p)_{\mathbb{Q}_p} = \left(\frac{a}{p}\right) = \begin{cases} +1 & a \text{ is a square mod } p, \\ -1 & \text{else.} \end{cases}$$

When $p = 2$, the Hilbert symbol also yields the Kronecker symbol:

> **Exercise 3.3.9.** Let $a \in \mathbb{Z}$ be odd. Show
>
> $$(a, 2)_{\mathbb{Q}_2} = \left(\frac{a}{2}\right) = \begin{cases} +1 & a \equiv 1, 7 \bmod 8, \\ -1 & a \equiv 3, 5 \bmod 8. \end{cases}$$
>
> Moreover, if a, b odd, show
>
> $$(a, b)_{\mathbb{Q}_2} = (-1)^{\frac{a-1}{2}\frac{b-1}{2}}.$$

In general, $(a, b)_F = 1$ is equivalent to N_0 being isotropic as described above, so (1) $\iff$ (3) can be rephrased as: $\left(\frac{a,b}{F}\right)$ is split if and only if $(a, b)_F = +1$.

Sometimes this is phrased as an equality of symbols. We define the **Hasse invariant** $\epsilon(a, b) \in \{\pm 1\}$ to be $+1$ if $\left(\frac{a,b}{F}\right)$ is split and -1 otherwise.[11] Thus we can recast the equivalence (1) $\iff$ (3) as

Corollary 3.3.8. *For $a, b \in F^\times$ we have $\epsilon(a, b) = (a, b)_F$.*

> **Exercise 3.3.10.** Show the quadratic Hilbert symbol satisfies the following properties for $a, b, c \in F^\times$:
>
> (1) $(a, b)_F = (b, a)_F$;
>
> (2) $(a, b)_F = (ac, bd)_F$ if $c, d \in F^{\times 2}$;
>
> (3) $(a, -a)_F = (a, 1 - a)_F = 1$; and
>
> (4) $(a, b)_F (a, c)_F = (a, bc)_F$ if F is p-adic.

Now we can apply these results to some explicit splitting questions.

[11] Recall we also define a Hasse invariant $\operatorname{inv} A_v$ for CSAs over local fields in Section 2.7. When A_v is a quaternion algebra over F_v, then $\operatorname{inv} A_v \in \frac{1}{2}\mathbb{Z}/\mathbb{Z} = \left\{0, \frac{1}{2}\right\}$ with $\operatorname{inv} A_v = 0$ if and only if $A_v \simeq M_2(F_v)$. This space is an additive group isomorphic to $\mathbb{Z}/2\mathbb{Z}$ whereas $\{\pm 1\}$ is a multiplicative representation of the same group. Consequently, translating additive notation to multiplicative notation shows these two notions of Hasse invariants agree.

Exercise 3.3.11. Determine which of the following quaternion algebras are split: $\left(\frac{2,3}{\mathbb{Q}}\right)$, $\left(\frac{-2,3}{\mathbb{Q}}\right)$, $\left(\frac{-2,-3}{\mathbb{Q}}\right)$, $\left(\frac{-2,-3}{\mathbb{Q}_3}\right)$, $\left(\frac{-2,-3}{\mathbb{Q}_5}\right)$, $\left(\frac{-2,-3}{\mathbb{Q}_7}\right)$.

Recall that every quadratic field embeds in the split quaternion algebra $M_2(\mathbb{Q})$ (Proposition 2.4.1). A weaker statement is true for quaternion division algebras.

Exercise 3.3.12. Given any quadratic number field $K = \mathbb{Q}(\sqrt{d})$, show there exists a quaternion *division* algebra $B/\mathbb{Q}$ such that K embeds as a subfield of B.

3.4 Local-global principle

In this section, F is a number field. Suppose (V, Q) is a quadratic space over F with a bilinear form $\langle \cdot, \cdot \rangle$. Then $\langle \cdot, \cdot \rangle$ extends to a bilinear form on $V \otimes F_v$, for any completion F_v. Let Q_v denote the associated quadratic form on $V \otimes F_v$.

We have the following well known local-global principle for quadratic forms.

Theorem 3.4.1 (Hasse–Minkowski)**.** *Let Q and Q' be quadratic forms over a number field F. Then*

(1) Q is isotropic if and only if Q_v is for all places v of F; and

(2) $Q \simeq Q'$ if and only $Q_v \simeq Q'_v$ for all places v of F.

Usually just the first statement is called the Hasse–Minkowski theorem, but the second follows from the first. These statements are also called the strong Hasse principle and the weak Hasse principle, respectively. See, e.g. [Ser73] or [Cas78] for proofs when $F = \mathbb{Q}$ (what Minkowski originally proved), or e.g. [O'M00] for the general case (completed by Hasse).

Recall that for an algebra A over a number field F, A_v denotes the completion $A \otimes F_v$. In particular, if B is a quaternion algebra over F, then B_v will be a quaternion algebra over F_v for all v.

Theorem 3.4.2. *Let B, B' be quaternion algebras over a number field F. Then $B \simeq B'$ if and only if $B_v \simeq B'_v$.*

Proof. This is an immediate consequence of the Hasse–Minkowski theorem and either Proposition 3.3.1 or Proposition 3.3.2. □

This is the Albert–Brauer–Hasse–Noether theorem (Theorem 2.7.1) in the case of quaternion algebras. Recall the general Albert–Brauer–Hasse–Noether theorem reduces to Hasse's norm theorem, whereas the quaternionic case reduces to the Hasse–Minkowski theorem. We stress that the Hasse–Minkowski theorem really is simpler than Hasse's norm theorem, though we do not have time to prove it. Now is a good time to read [Ser73] if you have not already, and convince yourself it is not too hard.

To finish the classification of quaternion algebras over number fields (i.e., the quaternionic case of Theorem 2.7.5), we need to prove (i) the local classification, i.e. that there is a unique quaternion division algebra over a p-adic field (we already know the classifications over $\mathbb{R}$ and

$\mathbb{C}$), i.e. that the Hasse invariant $\epsilon(a,b)$ determines B_v over a p-adic field; and (ii) determine when we can patch local quaternion algebras together to get a global quaternion algebra.

For this classification—specifically to determine quaternion algebras over p-adic fields—it will be useful to consider orders in quaternion algebras (which is our eventual goal of study anyway). In the next chapter we will define orders in CSAs and consider some of their basic properties. Then we will finish our classification in the following chapter, and subsequently move on to the study of arithmetic of quaternion algebras. Then we can use the arithmetic of quaternion algebras to repay some of our debt to the theory of quadratic forms with applications such as the determining numbers represented by certain ternary or quaternary quadratic forms.