Quaternion Algebras and Quadratic Forms

by

Zi Yang Sham

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Mathematics

in

Pure Mathematics

Waterloo, Ontario, Canada, 2008

Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be electronically available to the public.

Abstract

The main goal of this Masters' thesis is to explore isomorphism types of quaternion algebras using the theory of quadratic forms, number theory and algebra. I would also present ways to characterize quaternion algebras, and talk about how quaternion algebras are important in Brauer groups by describing a theorem proved by Merkurjev in 1981.

Acknowledgements

I need to thank my supervisor Professor David McKinnon for his kind guidance throughout the duration of my Masters' degree, and Professor Rahim Moosa and Professor Ken Davidson for being my readers. I would like to thank every instructor in the University of Waterloo who taught me for the past five years, without all of you I could not possibly be finishing this degree in pure mathematics. I am also grateful to Lalit Jain and Collin Roberts for helping me with typesetting. Finally my thanks go to my parents and Yunzhi for their loving support and concern.

Dedication

To everyone who likes pure mathematics

Contents

Chapter 1

Quadratic Forms

(From Chapter **I** of [8])

1.1 Quadratic Forms and Quadratic Spaces

An n-ary quadratic form (i.e. a 2-form) over a field F is a polynomial f in n variables over F that is homogeneous of degree 2. (Please note that throughout this article, the characteristic of F is assumed not to be 2.) It has the general form

$$f(X_1, \cdots, X_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} X_i X_j$$

where $a_i j \in F$, for all i, j. Since F is a field, $X_i X_j = X_j X_i$ for any i, j, we can make the coefficients symmetric by rewriting f as

$$f(X_1, \cdots, X_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \tfrac{1}{2}(a_{ij} + a_{ji}) X_i X_j$$

Now f determines uniquely a symmetric matrix M_f where $(M_f)_{ij} = \frac{1}{2}(a_{ij} + a_{ji})$.

For convenience we write $f(X_1, \cdots, X_n)$ as $f(X)$ and view X as a column vector, so then in terms of matrix notation, $f(X)$ satisfies

$$f(X) = (X_1, \cdots, X_n) M_f \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = X^t M_f X$$

Let f and g be quadratic forms, we say that f is equivalent to g (or $f \cong g$) if there exists an invertible matrix $C \in GL_n(F)$ such that $f(X) = g(CX)$. Notice that

$$f(X) = g(CX) = (CX)^t M_g (CX) = X^t (C^t M_g C) X$$

this implies that $M_f = C^t M_g C$. Thus, equivalence of forms can be regarded as congruence of the associated symmetric matrices and it is an equivalence relation. We can define the *quadratic map* Q_f defined by f to be $Q_f : F^n \to F$ such that $Q_f(x) = x^t M_f x$, for any $x \in F^n$ viewed as a column vector. In relation with the equivalence of forms, $f \cong g$ amounts to the existence of a linear automorphism C of F^n such that $Q_f(x) = Q_g(Cx)$ for every column tuple x. It's easy to that the quadratic map Q_f determines uniquely the quadratic form f. We also have the property that $Q_f(ax) = a^2 Q_f(x)$ for any $a \in F$.

In addition to the quadratic map, we can "polarize" Q_f by defining

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

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

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

We can get back Q_f from B_f by "depolarization", that is

$$Q_f(x) = B_f(x,x)$$

Now we are in the position to define quadratic spaces.

Let V be a finite dimensional F-vector space, and $B : V \times V \to F$ be a symmetric bilinear pairing on V. We call the pair (V, B) a quadratic space, and associate it with a quadratic map denoted by q_B or q when the context is clear. It is defined by $q(x) = B(x, x)$. As described above, we have

$$q(ax) = B(ax, ax) = a^2 B(x,x) = a^2 q(x)$$

and

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

Since q and B determines each other, we can use (V, q) to represent (V, B). While (V, B) determines a unique quadratic map, it also determines a unique *equivalence class* of quadratic forms in the following way. If we choose a basis $e_1, e_2, \cdots, e_n$ for V, then the quadratic space (V, B) gives rise to a quadratic form over F

$$f(X_1, \cdots, X_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} B(e_i, e_j) X_i X_j$$

with the associated matrix n by n M_f such that $(M_f)_{ij} = B(e_i, e_j)$. Note that if we identify V with F^n using the given coordination, then q_B corresponds precisely to the quadratic map q_f associated with the form f.

If we choose another basis $e'_1, e'_2, \cdots, e'_n$ for V, and write $e'_i = \sum_{k=1}^{n} c_{ki} e_k$ for some $c_{ki} \in F$ and for each i, we have

$$\begin{aligned}(M'_f)_{ij} &= B(\sum_{k=1}^{n} c_{ki} e_k, \sum_{l=1}^{n} c_{lj} e_l) \\ &= \sum_{k=1}^{n} \sum_{l=1}^{n} c_{ki} B(e_k, e_l) c_{lj} \\ &= (C^t M_f C)_{ij}\end{aligned}$$

where C is the matrix with $(C)_{kl} = c_{kl}$. Therefore we have that M'_f and M_f are congruent, so that the forms f' and f are equivalent. This unique equivalence class of forms determined by (V, B) is denoted by (f_B)

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

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

Notice that if f is the quadratic form corresponding to a basis $e_1, e_2, \cdots, e_n$ for V, and f' is a quadratic form corresponding to the basis $\tau(e_1), \tau(e_2), \cdots, \tau(e_n)$

for V', then

$$\begin{aligned}(M_f)_{ij} &= B(e_i, e_j) \\ &= B'(\tau(e_i), \tau(e_j)) \\ &= (M'_f)_{ij}\end{aligned}$$

From this, it is clear that $(V, B) \cong (V', B') \Leftrightarrow (f_B) = (f'_B)$.

Here is a summary of the above results.

- An n-ary quadratic form f determines uniquely the following,

    1. a symmetric n by n matrix M_f
    2. a quadratic map $Q_f : F^n \to F$ defined by $Q_f(x) = x^t M_f x$
    3. a symmetric bilinear pairing $B_f : F^n \times F^n \to F$,

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

- We can "depolarize" B_f to get back Q_f by $Q_f(x) = B_f(x, x)$.

- If $B : V \times V \to F$ is a symmetric bilinear pairing on a finite dimensional F-vector space V, the quadratic space (V, B) determines uniquely the following,

    1. a quadratic map q_B (or q) such that $q_B(x) = B(x, x)$. Since q_B and B determines each other, we can write (V, B) as (V, q).
    2. an equivalence class of quadratic forms. Equivalent quadratic forms correspond to equivalent bases of V.

- There is a one-to-one correspondence between the equivalence classes of n-ary quadratic forms and the isometry classes of n-dimensional quadratic spaces, so we can freely identify them.

Definition 1.1.1 *Let (V, B) be a quadratic space, and M a symmetric matrix associated to one of the quadratic forms in the equivalence class (f_B). We say (V, B) is a regular (or non-singular) quadratic space if one of the following equivalent conditions holds*

1. *M is a non-singular matrix.*

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

3. *If $x \in V$ such that $B(x, y) = 0$ for all $y \in V$, then $x = 0$.*

Even though the zero quadratic space (in which $B \equiv 0$) does not satisfy condition (1), we call it a regular quadratic space too.

Definition 1.1.2 *Let (V, B) be a quadratic space, and S be a subspace of V. Then $(S, B|_{S\times S})$ is also a quadratic space. The the orthogonal complement of S is defined by*

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

The orthogonal complement of V itself is called the *radical* of (V, B) and it's denoted by $V^{\perp} = \text{rad } V$. Observe that (V, B) is regular iff rad $V = 0$. However, if (V, B) is regular, the subspace S of V need not be regular. For instance, consider $(\mathbb{R}^2, B)$ where $B((a, b), (c, d)) = bc + ad$ and let $S = \text{span}\{(0, 1)\}$, $B|_{S\times S} \equiv 0$.

Analogous to vector spaces, quadratic spaces satisfy the following dimension theorem.

Proposition 1.1.3 *Let (V, B) be a* ***regular*** *quadratic space, and S be a subspace of V. Then,*

1. $\dim S + \dim S^{\perp} = \dim V$

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

Proof: Consider the isomorphism $\varphi : V \to V^*$ defined in **Definition 1.1.1**. Then $S^{\perp}$ is the subspace of V annihilated by the functionals in $\varphi(S)$. By the usual duality theory in linear algebra, we have

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

This proves (1). And by applying (1) twice,

$$\dim (S^{\perp})^{\perp} = \dim V - (\dim V - \dim S) = \dim S$$

and since $(S^{\perp})^{\perp} \supseteq S$, result (2) follows. $\square$

1.2 Diagonalization of Quadratic Forms

Definition 1.2.1 *Let f be a (n-ary) quadratic form over F, and $d \in \dot{F}$, where $\dot{F}$ is the multiplicative group of non-zero elements in F.*
We say f represents d if there exist $x_1, x_2, \cdots, x_n \in F$ such that $f(x_1, \cdots, x_n) = d$. The set of elements in $\dot{F}$ represented by f is denoted by $D(f)$ or sometimes $D_F(f)$. This set clearly depends only on the equivalence class of f. And if (V, B) is any quadratic space corresponding to the equivalence class of f, then $D(f)$ (or in this case $D(V)$) is exactly the set of values represented by q_B.

Group structure of $D(f)$

Since f is a quadratic form, if $a, d \in \dot{F}$, then clearly we have $d \in D(f)$ iff $a^2 d \in D(f)$. Thus $D(f)$ consists of a union of cosets of $\dot{F}$ modulo $\dot{F}^2$. We call $\dot{F}/\dot{F}^2$ the *group of square classes of* F.

The set $D(f)$ is always closed under inverses, since $d \in D(f)$ iff $d^{-1} = (d^{-1})^2 d \in D(f)$. However, f might not represent 1, so $D(f)$ might not contain the identity and is thus not a group. Even if it contains 1, it may not be closed under multiplication. Consider the form $f = X^2 + Y^2 + Z^2$ over $\mathbb{Q}$, then $D(f)$ contains $1, 2, 2^{-1}, 14$. However $2^{-1} \cdot 14 = 7$, and 7 is not a sum of three squares in $\mathbb{Q}$. Note that if $D(f)$ happens to be closed under multiplication, then for any $d \in D(f)$, $D(f)$ will contain $d \cdot d^{-1} = 1$, which makes it a subgroup of $\dot{F}$. In this case we call f a *group form* over F.

Definition 1.2.2 *If (V_1, B_1), (V_2, B_2) are quadratic spaces, the orthogonal sum $V_1 \perp V_2 = (V, B)$ is defined with $V = V_1 \oplus V_2$, and $B : V \times V -> F$ is given by*

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

for any $x_1, y_1 \in V_1$ and $x_2, y_2 \in V_2$.

Clearly, this B is symmetric and bilinear, which makes (V, B) a quadratic space. If we identify V_1 with the set $\{(x, 0) : x \in V_1\}$, and V_2 with the set $\{(0, x) : x \in V_2\}$, we have $B(V_1, V_2) = 0$. Also $B|_{V_1 \times V_1} = B_1$ since $B_2(0, 0) = 0$, similarly, $B|_{V_2 \times V_2} = B_2$. This justifies why we call it an orthogonal sum. As for the associated quadratic form, for any $x_1 \in V_1$ and $x_2 \in V_2$

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

Proposition 1.2.3 *The quadratic space $(V, B) = V_1 \perp V_2$ is regular if and only if (V_1, B_1) and (V_2, B_2) are regular.*

Proof: Say $\beta_1 = \{e_1, \cdots, e_n\}$ and $\beta_2 = \{e_{n+1}, \cdots, e_{n+m}\}$ are bases of V_1 and V_2 respectively. Let M be the matrix associated with (V, B), then $M_{ij} = B(e_i, e_j)$. We already saw that $B(V_1, V_2) = 0$, so if the matrices corresponding to B_1 and B_2 are M_1 and M_2 respectively, it's clear that M has the form,

$$\begin{pmatrix} M_1 & O \\ O & M_2 \end{pmatrix}$$

where O represents a block of zeroes. From this, it's clear that M is invertible if and only if M_1 and M_2 are invertible. The result then follows from the definition of regular spaces immediately.$\square$

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

Theorem 1.2.4 ***Representation Criterion***
Let (V, B) be a quadratic space, and $d \in \dot{F}$. Then $d \in D(V)$ iff there exists another quadratic space (V', B') together with an isometry $V \cong \langle d \rangle \perp V'$.

Proof: First assume $V \cong \langle d \rangle \perp V'$. Then since $q_B \equiv dX^2 + q'_B$, d is represented by $(1, 0)$, here $1 \in \langle d \rangle$ and $0 \in V'$. In other words, $d \in D(\langle d \rangle \perp V') = D(V)$. Conversely, suppose $d \in D(V)$, there exists $v \in V$ such that $q_B(v) = d$.
The radical of V, rad V, is a subspace of V, and the associated quadratic map of rad V is identically zero. Let W be a subspace of V such that $V = (\mathrm{rad}\, V) \oplus W$, since $\mathrm{rad}\, V$ is orthogonal to every vector, we have $V = (\mathrm{rad}\, V) \perp W$. Since $q_V = q_{\mathrm{rad}V} + q_W$, and we have $q_{\mathrm{rad}V} = 0$, $q_V = q_W$ and thus they represents the same set of values, i.e. $D(V) = D(W)$. Here W is a regular space, so we may assume without loss of generality that V is regular. Consider the linear

isomorphism $\tau : F \cdot v \to \langle d \rangle$ defined by $\tau(v) = 1$. For any $a, b \in F$

$$B(av, bv) = abB(v, v) = abd = d\tau(av)\tau(bv)$$

which shows that the quadratic subspace $F \cdot v$ is isometric to $\langle d \rangle$, and

$$(F \cdot v) \cap (F \cdot v)^{\perp} = 0$$

Since V is regular, by **Proposition 1.1.3**, we have

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

Therefore we conclude that $V \cong \langle d \rangle \perp (F \cdot v)^{\perp}$. (Note that when V is not regular, rad V is also contained in $(F \cdot v)^{\perp}$) $\square$

By repeatedly applying the Representation Criterion, we have proved the existence of an *orthogonal basis*. This is stated as the following corollary.

Corollary 1.2.5 *If (V, B) is a quadratic space over F, then there exist scalars $d_1, d_2, \cdots, d_n \in F$ such that $V \cong \langle d_1 \rangle \perp \cdots \perp \langle d_n \rangle$. (In other words, any n-ary quadratic form is equivalent to some diagonal form, $d_1 X_1^2 + \cdots + d_n X_n^2$, also denoted by $\langle d_1, \cdots, d_n \rangle$)*

Proof: If $D(V)$ is empty, then B is identically zero. In this case, every pair of vectors in V is orthogonal, so that V is isometric to an orthogonal sum of $\langle 0 \rangle$'s, so we can take any basis of V. If there exists some $d \in D(V)$, then by the Representation Criterion, we have $V \cong \langle d \rangle \perp V'$ for some (V', B'), and the result follows, by induction on $\dim V$.$\square$

Note that the special n-ary quadratic form $\langle d, \cdots, d \rangle$ is denoted by $n\langle d \rangle$. For example, $3\langle a \rangle \perp 4\langle b \rangle$ means $\langle a, a, a, b, b, b, b \rangle$

Corollary 1.2.6 *If (V, B) is a quadratic space and S is a regular subspace, then:*

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

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

Proof: (1) Since S is regular, $S \cap S^{\perp} = 0$. Since we already have the dimension theorem for regular subspaces, it suffices to show that V is spanned by S and $S^{\perp}$. By **Corollary 1.2.5**, S has an orthogonal basis $x_1, \cdots, x_p$. The regularity of S implies that $B(x_i, x_i) \neq 0$ for all i, since if $B(x_i, x_i) = 0$ for some i, the matrix associated with B will not be invertible. Given any $z \in V$, consider

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

Then for any j

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

which says that $y \in S^{\perp}$, and so

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

This finishes the proof of (1).

(2) If $V = S \perp T = S \oplus T$, then $T \subseteq S^{\perp}$. Therefore by (1) we have,

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

and it follows that $T = S^{\perp}$. $\square$

Corollary 1.2.7 *Let (V, B) be a regular quadratic space. A subspace S is regular iff there exists $T \subseteq V$ such that $V = S \perp T$.*

Proof: If S is regular, take $T = S^{\perp}$.
Conversely if $V = S \perp T$, then rad $S \subseteq$ rad $V = 0$, thus S is regular. $\square$

Definition 1.2.8 *The determinant of a nonsingular quadratic form f is defined to be $d(f) = \det(M_f) \cdot \dot{F}^2$ which is an element of $\dot{F}/\dot{F}^2$. Note that if $f \cong g$, then there are some nonsingular C such that $M_f = C^t M_g C$. We have*

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

That is, $d(f)$ is an invariant of the equivalence class of f. By considering block diagonal matrices, we see that

$$d(f_1 \perp f_2) = d(f_1) d(f_2)$$

So if $V \cong \langle d_1, \cdots, d_n \rangle$ and V corresponds to f, then we have $d(f) = d_1 \cdots d_n \cdot \dot{F}^2$. In this case, $d(f)$ is called the determinant of V and can be denoted by $d(V)$.

Proposition 1.2.9 *Let $q = \langle a, b \rangle$, $q' = \langle c, d \rangle$ be regular binary quadratic forms. (So that a, b, c, d are all nonzero.) Then $q \cong q'$ iff $d(q) = d(q')$, and q, q' represent a common element $e \in \dot{F}$.*

Proof: The only if part is clear. Conversely, assume that $d(q) = d(q') \in \dot{F}/\dot{F}^2$ and $e \in D(q) \cap D(q')$. By the Representation Criterion, we know that $q \cong \langle e, e' \rangle$ for some $e' \in \dot{F}$, since q has dimension 2. Taking their determinants, we have $ab\dot{F}^2 = ee'\dot{F}^2$, so $e' = abe$. Therefore we have $q \cong \langle e, abe \rangle = eX^2 + abeY^2$, and similarly $q' \cong \langle e, cde \rangle = eX^2 + cdeY^2$. But from $d(q) = d(q')$ we have that $ab\dot{F}^2 = cd\dot{F}^2$, so $abeY^2$ and $cdeY^2$ are isometric and thus $q \cong q'$. $\square$

1.3 Hyperbolic Plane and Hyperbolic Spaces

Definition 1.3.1 *Let v be a nonzero vector in a quadratic space (V, B). We say that v is an isotropic if $B(v, v) = q_B(v) = 0$, and anisotropic otherwise. The quadratic space (V, B) is said to be isotropic if it contains an isotropic vector, and it is anisotropic otherwise. And (V, B) is totally isotropic if every non-zero vector in V is isotropic, i.e. $B \equiv 0$.*

Theorem 1.3.2 *Let (V, q) be a 2-dimensional quadratic space. The following are equivalent.*

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

*Note: A 2-dimensional quadratic space satisfying any of the above statements is called a **hyperbolic plane**, and it can be denoted by $\mathbb{H}$.*
*An orthogonal sum of hyperbolic planes is called **hyperbolic space**, with its corresponding quadratic form in the form*

$$X_1X_2 + \cdots + X_{2m-1}X_{2m} \quad \text{or} \quad (X_1^2 - X_2^2) + \cdots + (X_{2m-1}^2 - X_{2m}^2)$$

Proof:
(3) $\Leftrightarrow$ (4) Let $g(X_1, X_2) = X_1X_2$, and C be the invertible linear transformation

$$(X_1, X_2) \mapsto (X_1 + X_2, X_1 - X_2)$$

then $g(C(X_1, X_2)) = (X_1 + X_2)(X_1 - X_2) = X_1^2 - X_2^2 = \langle 1, -1 \rangle$

(1) $\Rightarrow$ (2) Let x_1, x_2 be an orthogonal bases for V, so $B(x_1, x_2) = 0$. The quadratic space V is regular implies that $q(x_i) = d_i \neq 0$, for $i = 1, 2$. If $ax_1 + bx_2$ is an isotropic vector, then $a, b \neq 0$, and,

$$0 = q(ax_1 + bx_2) = a^2 d_1 + b^2 d_2$$

which implies that $d_1 = -(ba^{-1})^2 d_2$ and we have

$$d(V) = d_1 d_2 \cdot \dot{F}^2 = -(ba^{-1}d_2)^2 \dot{F}^2 = -1 \cdot \dot{F}^2$$

(2) $\Rightarrow$ (3) Assuming (2), and say $q = aX^2 + bY^2$ for some $a, b \in \dot{F}$, such that $d(V) = ab \cdot \dot{F}^2 = -1 \cdot \dot{F}^2$. Therefore $ab \in -1 \cdot \dot{F}^2$ and equivalently $a/b \in -1 \cdot \dot{F}^2$, and there exists $k \in \dot{F}$ such that $a/b = -k^2$. By applying the linear transformation $Y \mapsto kY$, we see that

$$q \cong aX^2 + bk^2Y^2 = aX^2 + b(-a/b)Y^2 = aX^2 - aY^2$$

Therefore the associated quadratic form is equivalent to aXY by using similar argument in proving (3) $\Leftrightarrow$ (4). Now the map $aX \mapsto X$ gives

$$q \cong aXY \cong XY \cong X^2 - Y^2 = \langle 1, 1 \rangle$$

(3) $\Rightarrow$ (1) For the quadratic form $\langle 1, -1 \rangle = X_1^2 - X_2^2$, $(1, 1)$ is an isotropic vector of the quadratic space. $\square$

Next, we are going to see how to find a decomposition of a quadratic space by considering its hyperbolic "parts".

Theorem 1.3.3 *Let (V, B) be a regular quadratic space. Then:*

1. *Every totally isotropic subspace of $U \subseteq V$ of positive dimension r is con-*

tained in a hyperbolic subspace $T \subseteq V$ of dimension $2r$.

2. *V is isotropic iff V contains a hyperbolic plane.*

Proof: We can prove (1) by using induction on r. First, let $\{x_1, x_2, \cdots, x_r\}$ be a basis of U, and let $S = \text{span}\{x_2, \cdots, x_r\}$, so that $\text{dim}S = r - 1$. We have that $U^\perp \subseteq S^\perp$. Since V is regular, the dimension formula applies,

$$\text{dim}S^\perp = \text{dim}V - \text{dim}S > \text{dim}V - \text{dim}U = \text{dim}U^\perp$$

Thus there exists $y \in S^\perp$ such that y is not in $U^\perp$. In other words, y is orthogonal to all of $x_2, \cdots, x_r$, but not orthogonal to x_1. Assume for contradiction that $y = ax_1$ for some $a \in F$. Since U is a totally isotropic space, x_1 is isotropic and $B(x_1, x_1) = 0$. Then

$$B(y, x_1) = B(ax_1, x_1) = aB(x_1, x_1) = 0$$

which means y is orthogonal to x_1 which contradicts the property of y. Therefore we have that y and x_1 are linearly independent. Consider the subspace $H = Fx_1 + Fy$ which has determinant

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

We want to show that H is regular. Assume that $ax_1 + by \in H$ is such that $B(ax_1 + by, cx_1 + dy) = 0$ for any $c, d \in F$. Then since $B(x_1, x_1) = 0$ we have

$$(ad + bc)B(y, x) + bdB(y, y) = 0$$

By looking at the coefficients of c, d, and the fact that $B(y, x) \neq 0$, it's easy to see that $a = b = 0$. Hence H is regular. So by **Theorem 1.3.2**, this shows that

H is a hyperbolic plane. Since H is regular, we can write $V = H \perp V'$, where V' contains $x_2, \cdots, x_r$. By **Corollary 1.2.7**, V' is regular, and the result follows by induction.

(1)$\Rightarrow$(2) If V is an isotropic space, it contains an isotropic vector v. However for any $a \in F$

$$B(av, av) = a^2 B(v, v) = 0$$

that is, V contains at least a 1-dimensional totally isotropic subspace U spanned by v. This subspace U satisfies the condition in the statement of (1) with $r = 1$. Therefore, V contains a hyperbolic plane. The converse is clear, since a hyperbolic plane is represented by $X_1^2 - X_2^2$, the space spanned by $(1, 1)$ is a totally isotropic subspace. $\square$

1.4 Witt's Decomposition and Cancellation

These two classical theorems in quadratic form theory first appeared in Witt's seminal paper in 1937. Please note that both theorems are proved for arbitrary quadratic spaces (V, q), without any regularity assumptions on (V, q). First, let us look at the statements of the theorems.

Witt's Decomposition Theorem 1.4.1 *Let (V, q) be a quadratic space. Then*

$$(V, q) = (V_t, q_t) \perp (V_h, q_h) \perp (V_\alpha, q_\alpha)$$

where V_t is totally isotropic, V_h is hyperbolic (or zero), and V_α is anisotropic, and V_t, V_h, V_α are uniquely determined up to isometries.

Witt's Cancellation Theorem 1.4.2 *If q, q_1, q_2 are arbitrary quadratic forms, then $q \perp q_1 \cong q \perp q_2 \Rightarrow q_1 \cong q_2$.*

We will need to apply the Cancellation theorem to prove the Decomposition theorem.

Proof of Witt's Decomposition Theorem: To show existence, let V_0 be such that

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

since V_0 is obviously orthogonal to rad$V$. If V_0 were not regular, there would be an element $r \in V_0$ such that, $B(r, v) = 0$ for any $v \in V_0$. However we also have $B(r, w) = 0$ for any $w \in \mathrm{rad}V$, that means r is orthogonal to every vector in V, i.e. $r \in \mathrm{rad}V$, which is a contradiction. So V_0 is regular, and rad$V$ is obviously totally isotropic. If V_0 is isotropic, by **Theorem 1.3.3**, it contains a hyperbolic plane H_1, and we can write $V_0 = H_1 \perp V_1$. If V_1 is again isotropic, we may further write $V_1 = H_2 \perp V_2$, where H_2 is a hyperbolic plane. After a finite number of steps, we achieve a decomposition

$$V_0 = (H_1 \perp \cdots \perp H_m) \perp V_\alpha$$

so now $V_h = H_1 \perp \cdots \perp H_m$ is hyperbolic (or zero, if V_0 is not isotropic), and V_α is anisotropic. For uniqueness, assume that V has another decomposition $V_t' \perp V_h' \perp V_\alpha'$. Since V_t' is totally isotropic and $V_h' \perp V_\alpha'$ is regular, we have

$$\mathrm{rad}V = \mathrm{rad}(V') \perp \mathrm{rad}(V_h' \perp V_\alpha') = V_t'$$

So by the Cancellation Theorem, $V_h \perp V_\alpha \cong V_h' \perp V_\alpha'$. Write $V_h \cong m \cdot \mathbb{H}$, that is, m orthogonal copies of hyperplanes, and $V_h' \cong m' \cdot \mathbb{H}$. By using the Cancellation theorem to cancel one $\mathbb{H}$ at a time, we conclude that $m = m'$ and $V_\alpha \cong V_\alpha'$, since V_α, V_α' are anisotropic. This finishes the proof of Witt's Decomposition Theorem. $\square$

Definition 1.4.3 *The integer $m = \frac{1}{2}\text{dim}V_h$ uniquely determined in the proof of Witt's Decomposition Theorem is called the Witt index of the quadratic space (V, q). The isometry class of V_α is called the anisotropic part of (V, q).*

To establish Witt's Cancellation Theorem, we need to introduce the notion of a *hyperplane reflection*. Say (V, q) is any quadratic space, we will write $O_q(V) = O(V)$ to denote the group of of isometries of (V, q), it is sometimes called *orthogonal group*. Next we are going to associate an element $\tau_y \in O(V)$ to every *anisotropic* vector $y \in V$. As a map from V to itself, τ_y is defined by

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

for any $x \in V$. Since $B(x, y) = (q(x + y) - q(x) - q(y))/2$, $B(x, y)$ is linear in x. This shows that

1. τ_y is a linear endomorphism.

2. τ_y is the identity map on $(F \cdot y)^\perp$. To see this, consider when $B(x, y) = 0$, then $\tau_y(x) = x$. Also, we have

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

   Therefore $(\tau_y)^2$ is the identity map and we say that τ_y is an *involution*. In other words, it fixes the hyperplane $(F \cdot y)^\perp$, and reflects the vector y across $(F \cdot y)^\perp$ to $-y$.

3. $\tau_y \in O(V)$, that is, τ_y is an isometry. This is proved as follows

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

4. As a linear automorphism, τ_y has determinant -1.

Proposition 1.4.4 *Let (V, q) be a quadratic space, and x, y be vectors such that $q(x) = q(y) \neq 0$. Then there exists an isometry $\tau \in O(V)$ such that $\tau(x) = y$.*

Proof: First, we claim that for such a pair of x, y, we have that $q(x-y)$ and $q(x+y)$ cannot be both zero. Consider

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

which proves the claim. Assume $q(x-y) \neq 0$. First of all

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

$$\text{Thus,} \quad \tau_{x-y}(x) = x - \frac{2B(x, x-y)}{q(x-y)}(x-y) = x - (x-y) = y$$

If instead $q(x+y) \neq 0$, then $\tau_{x+y}(x) = -y$ and $-\tau_{x+y}$ is the function that we are looking for. □

Proof of Witt's Cancellation Theorem:

Suppose that $q \perp q_1 \cong q \perp q_2$.

Case 1: Assume that q is totally isotropic and q_1 is regular. Then the symmetric bilinear form B and the matrix associated with q are identically zero. Let M_1 and M_2 be the matrix corresponding to q_1 and q_2 respectively. The hypothesis $q \perp q_1 \cong q \perp q_2$ implies that $\begin{pmatrix} 0 & 0 \\ 0 & M_1 \end{pmatrix}$ is congruent to $\begin{pmatrix} 0 & 0 \\ 0 & M_2 \end{pmatrix}$, so there exists an invertible matrix $E = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ such that

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

In particular, $M_1 = D^t M_2 D$. Since M_1 and D are invertible, M_1, M_2 are congruent and thus $q_1 \cong q_2$.

Case 2: Assume that q is totally isotropic. Without loss of generality, assume that there are exactly r zeros in the diagonalization of q_1, and that that of q_2 has at least r zeros. Then we can rewrite $q \perp q_1 \cong q \perp q_2$ as

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

here $q \perp r\langle 0 \rangle$ is totally isotropic and q_1' is regular. Using the result in *Case 1*, we have that $q_1' \cong q_2'$ and $q_1 \cong r\langle 0 \rangle \perp q_1' \cong r\langle 0 \rangle \perp q_2' \cong q_2$. Therefore the cancellation also holds for *Case 2*.

Case 3: No assumptions on q, q_1 and q_2. Let $\langle a_1, \cdots, a_n \rangle$ be a diagonalization of q. By inducing on n, we are reduced to the case $n = 1$, as we can add one a_i at a time. If $a_1 = 0$, this is reduced to *Case 2*, in which we proved that the cancellation holds. Thus we may assume that $a_1 \neq 0$. Then the hypothesis $q \perp q_1 \cong q \perp q_2$ becomes $(V, B) \cong \langle a_1 \rangle \perp q_1 \cong \langle a_1 \rangle \perp q_2$. Let $\phi_i : (V, B) \to \langle a_1 \rangle \perp q_i$ be such isometries, then we can pick $x, y \in V$ such that $\phi_1(x) = 1 \perp \vec{0}$ and $\phi_2(y) = 1 \perp \vec{0}$, then

$$(F \cdot x) \perp q_1 \cong \langle a_1 \rangle \perp q_1 \cong \langle a_1 \rangle \perp q_2 \cong (F \cdot y) \perp q_2$$

By **Proposition 1.4.4**, and assuming that $x - y \neq 0$ we know that τ_{x-y} is an isometry such that $\tau_{x-y}(x) = y$. Moreover, for any vector z in the quadratic space corresponding to q_1,

$$\begin{aligned} B(y, \tau_{x-y}(z)) &= B\left(y, z - \frac{2B(z, x-y)}{q(x-y)}(x-y)\right) \\ &= B(z, y) + \frac{2B(z, x-y)}{q(x-y)}[B(x, y) - B(y, y)] \\ &= B(z, y) + B(z, x) - B(z, y) = 0 \end{aligned}$$

since $B(z, x) = 0$, $B(x, x) = B(y, y)$ and $q(x - y) = 2[B(x, y) - B(x, x)]$. This shows that the image of z under τ_{x-y} is orthogonal to y, in other words, τ_{x-y} is an isometry that takes the orthogonal complement of $(F \cdot x)$ to the orthogonal complement $(F \cdot y)$. That is, $q_1 \cong q_2$ and Witt's Cancellation Theorem is proved.

□

Chapter 2

Quaternion Algebras

(From [4], Chapter **I**II of [8], [10], Chapter 22 of [11])

2.1 Basic Properties of Quaternion Algebras

In 1843, William Rowan Hamilton discovered the real quaternions $\mathbb{H}$. It is a non-commmutative algebra of dimension 4 over the real numbers $\mathbb{R}$. We write

$$\mathbb{H} = \{\alpha + \beta i + \gamma j + \delta k \mid \alpha, \beta, \gamma, \delta \in \mathbb{R}\}$$

with addition defined by

$$(\alpha+\beta i+\gamma j+\delta k)+(\alpha'+\beta' i+\gamma' j+\delta' k) = (\alpha+\alpha')+(\beta+\beta')i+(\gamma+\gamma')j+(\delta+\delta')k$$

scalar multiplication defined by $\lambda(\alpha + \beta i + \gamma j + \delta k) = \lambda\alpha + \lambda\beta i + \lambda\gamma j + \lambda\delta k$ for any $\lambda, \alpha, \beta, \gamma, \delta, \alpha', \beta', \gamma', \delta' \in \mathbb{R}$.
A natural basis for this vector space over $\mathbb{R}$ is $\{1, i, j, k\}$

The set $\mathbb{H}$ is also a non-commutative ring with multiplication defined by

$$i^2 = j^2 = k^2 = -1,\ ij = -ji = k$$

with the usual distributivity. We can see that

$$\frac{1}{\alpha + \beta i + \gamma j + \delta k} = \frac{\alpha - \beta i - \gamma j - \delta k}{\alpha^2 + \beta^2 + \gamma^2 + \delta^2}$$

if at least one of $\alpha, \beta, \gamma, \delta$, is non-zero. This shows that $\mathbb{H}$ is a division algebra. In the case where the real part of a quaternion is zero, that is, $\alpha = 0$, it is called a *pure quaternion*.

The real numbers $\mathbb{R}$ is a subring of $\mathbb{H}$ identified with the set of quaternions with $\beta = \gamma = \delta = 0$. The complex numbers $\mathbb{C}$ is also a subring of $\mathbb{H}$ identified with $\mathbb{R} + \mathbb{R}i$.

In general, we can have quaternions over an arbitrary field F, and i^2 and j^2 need not equal -1. (Assume that the characteristic of F is not 2.) For non-zero $a, b \in F$, we define the quaternion algebra $A = (\frac{a,b}{F})$ to be the 4 dimensional F-algebra on two generators i, j with the defining relations

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

Like in the real quaternions, we define $k = ij$ and now $k^2 = -ab$, and

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

We say that any two of the elements i, j, k *anticommute*. And here, $\{1, i, j, k\}$ is a basis for A over F so that A has dimension 4 over F.
Therefore in the case where $a = b = -1$ and $F = \mathbb{R}$, $(\frac{-1,-1}{\mathbb{R}})$ is the real quaternions.

2.2 Determining the Isomorphism Type

For a general quaternion algebra A over a field F, we are interested in its isomorphism type. While A can be a division algebra (e.g.$\mathbb{H}$), it is also possible that A is isomorphic to M_2F, the algebra of all 2×2 matrices with entries from F. In fact, these are the only possibilities! To see this, let us first show that A is a central simple algebra. Since i, j, k do not commute pairwise, and A is a F-vector space, A has center F. Also, A has no non-trivial two-sided ideal, in other words, it is simple. (See P.232 Lemma 3 of [4]) Thus A is a central simple algebra. Consider the following theorem.

Theorem 2.2.1 ***Artin-Wedderburn Theorem***
A semisimple ring R is isomorphic to a product of n_k by n_k matrix rings over division rings D_k, for some integers n_k, both of which are uniquely determined up to permutation of the index k.

And as an immediate corollary,

Corollary 2.2.2 *Any central simple algebra which is finite dimensional over its center F is isomorphic to an algebra M_nD, where n is a positive integer and D is a division algebra over F.*

Say the quaternion algebra is $(\frac{a,b}{F})$, $\dim_F(\frac{a,b}{F}) = 4$ and $\dim_F M_nD = n^2$ $\dim_F D$, so the only possibilities are $n = 1$ and $D = (\frac{a,b}{F})$, or $n = 2$ and $D = F$. Notice that when $n = 1$, $D = (\frac{a,b}{F})$ is a division algebra. Whereas in the case where $n = 2$, $(\frac{a,b}{F}) \cong M_2D$, in other words, it is *split*. (If an F-algebra is isomorphic to a full matrix algebra over F we say that the algebra is split.)

The question is, how do we determine when it is a division algebra, and when it is split? The answer is to look at its norm form.

Definition 2.2.3 *Let $q \in (\frac{a,b}{F})$, say $q = \alpha + \beta i + \gamma j + \delta k$ with $\alpha, \beta, \gamma, \delta \in F$. Denote the conjugate of q, by $\bar{q} = \alpha - \beta i - \gamma j - \delta k$*

Define the norm form $N : (\frac{a,b}{F}) \to F$ *by* $N(q) = q\bar{q}$ *for every* $q \in (\frac{a,b}{F})$ *and the trace by* $T(x) = x + \bar{x}$.

Note that if $q = \alpha + \beta i + \gamma j + \delta k$,

$$N(q) = q\bar{q} = \bar{q}q = \alpha^2 - a\beta^2 - b\gamma^2 + ab\delta^2$$

which is a quadratic form in four variables $\alpha, \beta, \gamma, \delta$ and it is denoted by $< 1, -a, -b, ab >$. This $< 1, -a, -b, ab >$ corresponds to the diagonal entry of the matrix representation of the quadratic form. In fact, the quaternion algebra $(\frac{a,b}{F})$ can be considered as a quadratic space, with the associated quadratic form being the norm form N. The symmetric bilinear pairing B is then given by $B(x, y) = (x\bar{y} + y\bar{x})/2 = T(x\bar{y})/2$.

The conjugation function is called an *involution*. In general, an *F-involution* (or an *involution of the first kind*) of an algebra A is a map $\sigma : A \to A$ which is F-linear and satisfies

1. $\sigma(x + y) = \sigma(x) + \sigma(y)$ for all $x, y \in A$
2. $\sigma(xy) = \sigma(y)\sigma(x)$ for all $x, y \in A$
3. $\sigma(\sigma(x)) = x$ for all $x \in A$

In the case of the real quaternions $\mathbb{H}$, we have seen that the inverse of an element q is $\frac{\bar{q}}{N(q)}$, and in $\mathbb{H}$, we have

$$N(q) = \alpha^2 + \beta^2 + \gamma^2 + \delta^2$$

which is represented by $< 1, 1, 1, 1 >$. This is a sum of 4 squares in $\mathbb{R}$, and since $\mathbb{R}$ is a real closed field, $N(q)$ is zero if and only if $q = 0$. This is saying that the only element that has no inverse in $\mathbb{H}$ is zero, implying that $\mathbb{H}$ is a division algebra. However, when we are dealing with general quaternion algebras over arbitrary fields, it's possible that the norm of a non-zero element q is zero, which further implies that q (and $\bar{q}$) are zero divisors. As a matter of fact, we have the following theorem.

Theorem 2.2.4 *The quaternion algebra $(\frac{a,b}{F})$ is a division algebra if and only if its norm form $N : (\frac{a,b}{F}) \to F$ satisfies $N(q) = 0 \Rightarrow q = 0$, i.e. the norm form is anisotropic.*

Proof: If $(\frac{a,b}{F})$ is a division algebra, for $q \in (\frac{a,b}{F})$ if $N(q) = q\bar{q} = 0$, either $q = 0$ or $\bar{q} = 0$, both of which implies $q = 0$, so that the norm form is anisotropic. For the other direction, we see that $q^{-1} = \frac{\bar{q}}{N(q)}$ if $N(q) \neq 0$. Now $N(q) = 0 \Rightarrow q = 0$, that says that every non-zero element of $(\frac{a,b}{F})$ is invertible, i.e. $(\frac{a,b}{F})$ is a division algebra. $\square$

This theorem gives us a criterion to determine whether a quaternion algebra is a division algebra or not. And as an immediate consequence, we have that the quaternion algebra is split (i.e. $\cong M_2(F)$) if and only if the norm form is isotropic, i.e. the norm of a non-zero element is zero.

Theorem 2.2.5 ***Identification Theorem for Quaternion Algebras***
Let B be a 4-dimensional algebra over a field F ($\mathrm{char}F \neq 2$), and let $c, d \in \dot{F}$ and $u, v \in B$ be such that,

$$u^2 = c,\ v^2 = d,\ \textit{and}\ uv = -vu$$

then $B \cong (\frac{c,d}{F})$. (From Page 351 of [11])

Proof: Let $A = (\frac{c,d}{F})$, and $h : A \to B$ be F-linear such that,

$$h(1) = 1, h(i) = u, h(j) = v, h(k) = uv$$

It is clear that h preserves addition, multiplication, and the anti-commutativity of i, j, k. The kernel of a homomorphism is an ideal of the domain. Here A is a central simple algebra, therefore h cannot have a non-zero kernel. And h is obviously surjective, thus h is an isomorphism. $\square$

Definition 2.2.6 *Let A be a quaternion algebra, an element $v = \alpha+\beta i+\gamma j+\delta k$ is said to be a pure quaternion if $\alpha = 0$. The F-vector space of pure quaternions of A is denoted by A_0.*

Proposition 2.2.7 *Let $A = (\frac{a,b}{F})$, and v be a non-zero element of A. Then $v \in A_0$ iff $v \notin F$ and $v^2 \in F$.*

Proof: If $v = \alpha + \beta i + \gamma j + \delta k$, we have

$$v^2 = (\alpha^2 + a\beta^2 + b\gamma^2 - ab\delta^2) + 2\alpha(\beta i + \gamma j + \delta k)$$

Therefore when $\alpha = 0$, $v^2 \in F$. Conversely, if $v \notin F$, then one of β, γ, δ must be non-zero. For $v^2 \in F$ to be true, the above equation implies that $\alpha = 0$, and hence v is a pure quaternion.$\square$

Corollary 2.2.8 *If $A = (\frac{a,b}{F})$, $A' = (\frac{a',b'}{F'})$, and $\varphi : A \to A'$ is an F-algebra isomorphism, then $\varphi(A_0) = A'_0$. In particular, A_0 is stable under any F-algebra endomorphism of A.*

Proof: Since φ is an F-algebra isomorphism, by **Proposition 2.2.7** we have

$$\begin{aligned} v \in A_0 &\Leftrightarrow v \notin F, v^2 \in F \\ &\Leftrightarrow \varphi(v) \notin F, \varphi(v)^2 \in F \\ &\Leftrightarrow \varphi(v) \in A'_0 \end{aligned}$$

The second conclusion is clear since A is a central simple algebra and every F-algebra endomorphism of A is an automorphism. $\square$

We have come to one of our major theorems linking quaternion algebras and quadratic forms.

Theorem 2.2.9 *For $A = (\frac{a,b}{F})$, $A' = (\frac{a',b'}{F})$, the following statements are equivalent:*

1. *A and A' are isomorphic as F-algebras.*

2. *A and A' are isometric as quadratic spaces.*

3. *A_0 and A'_0 are isometric as quadratic spaces.*

In other words, to determine whether two quaternion algebras are isomorphic, we only have to check if their norm forms are isometric. This will be important in finding the isomorphism class of a quaternion algebra.

Proof: (1) $\Rightarrow$ (2) Suppose $\varphi : A \to A'$ is an F-algebra homomorphism, then by **Corollary 2.2.8** we have that $\varphi(A_0) = A'_0$. If $x = \alpha + x_0$ where $\alpha \in F$ and $x_0 \in A_0$, then $\bar{x} = \alpha - x_0$, and hence $\varphi(x) = \alpha + \varphi(x_0)$ and $\varphi(\bar{x}) = \alpha - \varphi(x_0)$. Since $\varphi(x_0) \in A'_0$, we have $\overline{\varphi(x)} = \varphi(\bar{x})$. Therefore,

$$N(\varphi(x)) = \varphi(x)\overline{\varphi(x)} = \varphi(x)\varphi(\bar{x}) = \varphi(N(x)) = N(x)$$

so φ is an isometry from A to A'.

(2) $\Rightarrow$ (3) If $A = \langle 1 \rangle \perp A_0$ and $A' = \langle 1 \rangle \perp A'_0$ are isometric, then by Witt's Cancellation Theorem, A_0 and A'_0 are isometric.

(3) $\Rightarrow$ (1) Let $\sigma : A_0 \to A'_0$ be an isometry (which is a linear isomorphism). Then,

$$N(\sigma(i)) = N(i) = -a$$

and

$$N(\sigma(i)) = \sigma(i)\overline{\sigma(i)} = \sigma(i)\sigma(\bar{i}) = -\sigma(i)^2$$

clearly $\sigma(i)^2 = a$, and similarly $\sigma(j)^2 = b$. Finally,

$$0 = B(i,j) = B(\sigma(i), \sigma(j)) = (-\sigma(i)\sigma(j) - \sigma(j)\sigma(i))/2$$

implies that $\sigma(i)\sigma(j) = -\sigma(j)\sigma(i)$ and hence $A' \cong (\frac{a,b}{F}) = A$ by **Theorem 2.2.5**. $\square$

Since isomorphic quaternion algebras are isometric as quadratic spaces and vice versa, from now on, we will freely interchange between $A = (\frac{a,b}{F})$ and $\langle 1, -a, -b, ab\rangle$ or $X_1^2 - aX_2^2 - bX_3^2 + abX_4^2$ which is its norm form. The elements a, b are always non-zero, so that $\langle 1, -a, -b, ab\rangle$ is always a regular form. Let us look at some examples of isomorphic quaternion algebras.

Examples 1

1. The quaternion algebra $A = (\frac{a,b}{F})$ is isomorphic to $B = (\frac{b,a}{F})$ because their norm forms $\langle 1, -a, -b, ab\rangle$ and $\langle 1, -b, -a, ab\rangle$ are isometric. In fact, the isometry will be sending $X_2 \mapsto X_3$ and $X_3 \mapsto X_2$.

2. For any $x, y \in \dot{F}$, $A = (\frac{a,b}{F})$ is isomorphic to $B = (\frac{ax^2, by^2}{F})$. The elements $u = xi$ and $v = yj$ in A satisfy $u^2 = ax^2$, $v^2 = by^2$ and $uv = -vu$. So by the Identification Theorem for Quaternion Algebras, A and B are isomorphic.

3. In $A = (\frac{a,b}{F})$, since the elements $u = i$ and $v = k$ satisfy $u^2 = a$, $v^2 = -ab$ and $uv = -vu$, by the Identification Theorem, $A \cong (\frac{a,-ab}{F})$.

4. Since a^2X^2 are isometric to X^2 by the linear isomorphism $X \mapsto aX$, we have

$$\langle 1, -a, 1, -a\rangle \cong \langle 1, -a, a^2, -a\rangle \cong \langle 1, -a, -a, a^2\rangle$$

Therefore $(\frac{a,a}{F}) \cong (\frac{a,-1}{F})$.

5. Let $A = M_2(F)$, $u = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $v = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then $u^2 = -I$ and $v^2 = I$ and $uv = -vu$, where I is the 2 by 2 identity matrix. Therefore by the Identification Theorem, we have $A \cong (\frac{1,-1}{F})$.

Theorem 2.2.10 *For $A = (\frac{a,b}{F})$, the following statements are equivalent:*

1. $A \cong (\frac{1,-1}{F})$
2. *A is isotropic as a quadratic space. (So by* **Theorem 2.2.4**, *$A \cong M_2(F)$)*
3. *A is hyperbolic as a quadratic space.*
4. *The binary form $\langle a, b\rangle$ represents* 1.

Proof:

(1)⇒(2) $(\frac{1,-1}{F}) \cong \langle 1, -1, 1, -1\rangle$ is isotropic because $N(1 + i) = 0$.

(2)⇒(1) If A is isotropic, $A \cong M_2(F)$, and so $A \cong (\frac{1,-1}{F})$ by *Example* 1(5).

(1)⇔(3) is the definition of a 4-dimensional hyperbolic space, which has the associated form $\langle 1, -1, 1, -1\rangle$.

(1)⇒(4) Assume that $A \cong (\frac{1,-1}{F})$, then we also have $\langle 1, -a, -b, ab\rangle \cong \langle 1, -1, 1, -1\rangle$. Consider $q = \langle 1, -1\rangle$, q represents 1, so by **Proposition 1.2.9**

$$q \cong \langle a, 1 \cdot -1 \cdot a\rangle \cong \langle a, -a\rangle$$

similarly $q \cong \langle b, -b\rangle$. Therefore

$$\langle 1, -a, -b, ab\rangle \cong \langle 1, -1, 1, -1\rangle \cong \langle a, -a, b, -b\rangle$$

By Witt's Cancellation Theorem, we can cancel the $-a$ and $-b$ and get

$$q' = \langle 1, -ab\rangle \cong \langle a, b\rangle = q''$$

Since $(1,0)$ is a vector such that $q'(1,0) = 1$. Also q' and q'' are isometric, so there exists $(x,y) \in F \times F$ such that $q''(x,y) = ax^2 + by^2 = 1$.

(4)$\Rightarrow$(1) Now if $\langle a,b\rangle$ represents 1, then $\langle a,b\rangle \cong \langle 1,ab\rangle$ by **Proposition 1.2.9**. Now

$$\langle 1,-a,-b,ab\rangle \cong \langle 1,-1,-ab,ab\rangle \cong \langle 1,-1,1,-1\rangle$$

which implies $(\frac{a,b}{F}) \cong (\frac{1,-1}{F})$. $\square$

Note: In the above theorem, the equivalence (1)$\Leftrightarrow$(4) is also called *Hilbert's Criterion* for the splitting of the quaternion algebra A. Whether the form $\langle a,b\rangle$ represents 1, can also be written as whether the *Hilbert equation* $ax^2 + by^2 = 1$ has a solution over a field F. In elementary number theory, this equation is used to define the *Hilbert symbol* over a local field K as follows,

$$(a,b) = \begin{cases} 1 & \text{if } z^2 = ax^2 + by^2 \text{ has a non-zero solution } (x,y,z) \in K^3 \\ -1 & \text{otherwise} \end{cases}$$

From the theorems we had above, we have that $(\frac{a,b}{F})$ splits if $(a,b) = 1$, and it is a division algebra if $(a,b) = -1$. Here are some examples of quaternion algebras that split.

Examples 2

1. For any $a \in \dot{F}$, $(\frac{a,-a}{F})$ is split because of Hilbert's Criterion. The binary form $\langle a,-a\rangle$ represents 1, since

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

2. If again $a \in \dot{F}$, $\langle 1,a\rangle$ represents 1 obviously so $(\frac{1,a}{F})$ is split.

3. If $a \neq 0, 1$, then let $u = \begin{pmatrix} 0 & a \\ 1 & 0 \end{pmatrix}$, $v = \begin{pmatrix} 1 & -a \\ 1 & -1 \end{pmatrix}$.

   Then $u^2 = aI$ and $v^2 = (1-a)I$ and $uv = -vu$, so by the Identification Theorem, $M_2(F) \cong (\frac{a,(1-a)}{F})$.

Corollary 2.2.11 *The algebra $A = (\frac{-1,a}{F})$ splits iff a is a sum of two squares in F (not necessarily non-zero).*

Proof:

If the imaginary number $i \in F$, then

$$\left(\frac{1+a}{2}\right)^2 + \left(\frac{1-a}{2}i\right)^2 = a$$

so that a is always a sum of two squares. Therefore if $a = X^2 + Y^2$ for some $X, Y \in F$, $X^2 + Y^2 - a(1)^2 - a(0)^2 = 0$ which implies that the norm form $\langle 1, 1, -a, -a \rangle$ is isotropic. The algebra A is always split.

Instead if $i \notin F$

$$\left(\frac{-1,a}{F}\right) \text{ splits}$$

$\Leftrightarrow$ $\langle -1, a \rangle$ represents 1

$\Leftrightarrow$ there are $X, Y \in F$ such that $-X^2 + aY^2 = 1$ where Y cannot be zero

$\Leftrightarrow$ there are $X, Y \in F$ such that $a = Y^{-2} + X^2Y^{-2}$

That is a is a sum of two squares in F. □

Corollary 2.2.12 *For any prime $p \equiv 1 \bmod 4$, $(\frac{-1,-p}{\mathbb{Q}}) \cong (\frac{-1,-1}{\mathbb{Q}})$ is a division algebra, and $(\frac{-1,p}{\mathbb{Q}}) \cong M_2(\mathbb{Q})$*

Proof: The norm form of $(\frac{-1,-p}{\mathbb{Q}})$ is $X_1^2 + X_2^2 + pX_3^2 + pX_4^2$ which is positive over the non-zero rationals and thus anisotropic, so by **Theorem 2.2.4** it is a division algebra. By Fermat's Theorem, p is a sum of two squares, say $p = c^2 + d^2$. Let $u = i$, and $v = (cj + dk)/p$, then $u^2 = -1$, $v^2 = (-pc^2 - pd^2)/p^2 = -1$, and $uv = -vu$. We can then apply the Identification Theorem to get

$$\left(\frac{-1,-p}{\mathbb{Q}}\right) \cong \left(\frac{-1,-1}{\mathbb{Q}}\right)$$

We already have $p = c^2 + d^2$, and we can apply **Corollary 2.2.11** to get $(\frac{-1,p}{\mathbb{Q}}) \cong M_2(\mathbb{Q})$. $\square$

2.3 Quaternion Algebras over Different Fields

In general there is no procedure to decide if two quadratic forms are isometric, or if two quaternion algebras are isomorphic. This question is specific to a field. Two forms isometric over a field need not be isometric over another field. It is exactly because of this that a theory of quadratic forms becomes necessary. To illustrate this, let us look at quaternion algebras over different fields.

- The complex numbers $\mathbb{C}$

  $(\frac{a,b}{\mathbb{C}})$ is isomorphic to $M_2(\mathbb{C})$ for any non-zero $a, b \in \mathbb{C}$ because the norm form $\langle 1, -a, -b, ab\rangle$ is always isotropic. ($\because (\sqrt{a})^2 - a(1)^2 = 0$)

- The real numbers $\mathbb{R}$

  Whenever $a, b \in \mathbb{R}$ are negative, the norm form is $\langle 1, -a, -b, ab\rangle$ and the norm of a non-zero real number is always a sum of positive numbers. Therefore it is anisotropic and $(\frac{a,b}{\mathbb{R}})$ is a division algebra which is isomorphic to the real quaternions $\mathbb{H}$, by Frobenius's Theorem on Real Division Algebra.

Otherwise, if at least one of a, b is positive, $\langle 1, -a, -b, ab \rangle$ is isotropic, since either $(\sqrt{a})^2 - a \cdot 1^2 = 0$ or $(\sqrt{b})^2 - b \cdot 1^2 = 0$. Thus the form is isotropic and $(\frac{a,b}{\mathbb{R}})$ is split.

- The p-adic fields $\mathbb{Q}_p$ where p is a prime

  This is the completion of the field $\mathbb{Q}$ with respect to the p-adic absolute value on $\mathbb{Q}$. For each prime p there is a unique quaternion division algebra over $\mathbb{Q}_p$. This follows from the fact that, up to isometry, there is a unique anisotropic quadratic form of dimension 4 and it is the norm form of a quaternion algebra. However the proof of this is beyond the scope of this article; please see T.Y.Lam [8] Chapter VI for details.

- The finite fields $\mathbb{F}_n$ with n elements

  By the 1905 theorem of Wedderburn, any finite division ring is commutative. However, $(\frac{a,b}{\mathbb{F}_n})$ is not commutative, therefore it is not a division ring and it splits.

- The rational numbers $\mathbb{Q}$

  There are infinitely many non-isomorphic quaternion algebras over $\mathbb{Q}$, to see this, consider the following lemmas.

Lemma 1 *A positive integer n is a sum of two squares of integers if and only if n can be factored as ab^2 such that a is not divisible by any prime that is $3 \bmod 4$, $a, b \in \mathbb{Z}$. (See [2])*

Proof: First assume that n can be factored as ab^2 such that a is not divisible by any prime that is $3 \bmod 4$. Then a is a product of 2 and primes which are $1 \bmod 4$. By Fermat's Theorem, 2 and every prime that is $1 \bmod 4$ is a sum of two squares. Also, product of sums of squares is also a sum of squares, since

$$(\alpha^2 + \beta^2)(\gamma^2 + \delta^2) = (\alpha\gamma - \beta\delta)^2 + (\alpha\delta + \beta\gamma)^2$$

Therefore $a = a_1^2 + a_2^2$ for some $a_1, a_2 \in \mathbb{Z}$, and $n = a_1^2 b^2 + a_2^2 b^2$ which is a sum of two squares.

Now say n is a sum of two squares of integers $x^2 + y^2$. We will proceed by induction on n. Assume that every sum of squares that is smaller than n can be factored as the form ab^2 with a not divisible by primes congruent to $3 \bmod 4$. If n does not have prime factors that are $3 \bmod 4$, then n is of course in the form ab^2 with that property. So say there exists $q \equiv 3 \bmod 4$ such that $q \mid n = x^2 + y^2$. Since q is $3 \bmod 4$, q is irreducible in $\mathbb{Z}[i]$, and thus $q \mid (x + yi)$ and $q \mid (x - yi)$. This implies that $q \mid x$ and $q \mid y$, let $x = qx'$ and $y = qy'$. Now $n = q^2(x'^2 + y'^2)$ and $x'^2 + y'^2$ is a smaller sum of squares. By induction hypothesis $x'^2 + y'^2$ is of the form ab^2 such that a is not divisible by any prime $3 \bmod 4$, it follows that $n = aq^2b^2$ is also of the desired form. □

Lemma 2 *For any prime $q \equiv 3 \bmod 4$, $(\frac{-1,-q}{\mathbb{Q}})$ and $(\frac{-1,q}{\mathbb{Q}})$ are non-isomorphic division algebras. (See page 362 of [11])*

Proof: If they were isomorphic, their norm forms would be isometric and we have

$$\langle 1, 1, q, q \rangle \cong \langle 1, 1, -q, -q \rangle$$

By Witt's Cancellation Theorem, we cancel the two 1's and get $\langle q, q \rangle \cong \langle -q, -q \rangle$ which is obviously wrong. The form $\langle q, q \rangle$ is always non-negative and $\langle -q, -q \rangle$ is never positive. Therefore these two algebras are non-isomorphic.

Now, consider the form $\langle -1, -q \rangle = -X^2 - qY^2$, it is negative for any non-zero $X, Y \in \mathbb{Q}$. Hence the form does not represent 1, and by Hilbert's Criterion, $(\frac{-1,-q}{\mathbb{Q}})$ is a division algebra.

Next, we want to show that $(\frac{-1,q}{\mathbb{Q}})$ is a division algebra. Using Hilbert's Criterion once again, we aim to prove that the binary form $\langle -1, q \rangle$ does not represent 1. Assume the contrary, there exist $x, y \in \mathbb{Q}$ such that $-x^2 + qy^2 = 1$. Rearranging the terms, this is saying there exist $a, b, c, d \in \mathbb{Z}$ such that $gcd(a, b) = gcd(c, d) = 1$ and $q = (a/b)^2 + (c/d)^2$, we have

$$q(bd)^2 = (ad)^2 + (bc)^2$$

In the prime factorization of $q(bd)^2$, the exponent of q must be odd, and $q \equiv 3 \bmod 4$. However $q(bd)^2$ is a sum of squares here, this contradicts **Lemma 1**. We conclude that $\langle -1, q\rangle$ does not represent 1 and by Hilbert's Criterion $(\frac{-1,q}{\mathbb{Q}})$ is a division algebra.$\square$

Proposition 2.3.1 *Let $p \equiv q \equiv 3 \bmod 4$ be distinct prime numbers. We have*

$$\left(\frac{-1,p}{F}\right) \ncong \left(\frac{-1,q}{F}\right), \left(\frac{-1,p}{F}\right) \ncong \left(\frac{-1,-q}{F}\right), \left(\frac{-1,-p}{F}\right) \ncong \left(\frac{-1,-q}{F}\right)$$

Proof: Assume for contradiction that $(\frac{-1,-p}{F}) \cong (\frac{-1,-q}{F})$, then $\langle 1, 1, p, p\rangle \cong \langle 1, 1, q, q\rangle$. By Witt's Cancellation Theorem, $\langle p, p\rangle \cong \langle q, q\rangle$. However consider

$$S_p = \left\{ p\left(\frac{X^2}{Z^2} + \frac{Y^2}{Z^2}\right) \,\middle|\, X, Y, Z \in \mathbb{Z}\right\}$$

and also define S_q in a similar manner. So S_p is the set of values $\langle p, p\rangle$ represents. In the prime factorizations $X^2 + Y^2$ and Z^2, p and q both appears even number of times by **Lemma 1**, therefore S_p and S_q cannot be the same set of rationals, $\langle p, p\rangle$ and $\langle q, q\rangle$ are thus non-isometric. This is a contradiction and hence $(\frac{-1,-p}{F}) \ncong (\frac{-1,-q}{F})$. Similarly, we also have that $(\frac{-1,p}{F}) \ncong (\frac{-1,q}{F})$. The proof of $(\frac{-1,p}{F}) \ncong (\frac{-1,-q}{F})$ is easy. The set S_{-p} is not positive whereas S_q is not negative, therefore $\langle 1, 1, -p, -p\rangle$ and $\langle 1, 1, q, q\rangle$ cannot be isometric. $\square$

From **Lemma 2** and **Proposition 2.3.1**, we have that *there are infinitely many non-isomorphic non-split quaternion algebras over* $\mathbb{Q}$, namely $(\frac{-1,\pm q}{\mathbb{Q}})$ where $q \equiv 3 \bmod 4$.

Chapter 3

The Brauer Group and the Theorem of Merkurjev

(From Section 4.6 and 4.7 of [4], [10])

3.1 Properties of the Brauer Group

Closely related to quaternion algebras is the Brauer group. This group consists of similarity classes of central simple algebras over a specific field, with the group operation being tensor product over that field. Quaternion algebras and tensor products of them have order 1 or 2 in this group, we'll justify this by using tools from algebra and quadratic form theory. A.A. Albert conjectured that the subgroup generated by all the quaternion algebras over a field actually contains all the elements of order 2 in the Brauer group. This theorem was finally proved by Merkurjev in 1981 using tools from Milnor K-theory. Let us start by looking at some properties of central simple algebras.

Proposition 3.1.1 *If B is an algebra over F, then $M_n(B) \cong M_n(F) \otimes_F B$.*

Proposition 3.1.2 $M_m(F) \otimes_F M_n(F) \cong M_{mn}(F)$

These are basic results of tensor products of matrix algebras and proofs can be easily found in many algebra books, so the proofs will be left to the reader. Please see Jacobson, [4] page 216 for more details.

Theorem 3.1.3 *If A is a finite dimensional central simple algebra over a field F, then the enveloping algebra $A^e = A \otimes_F A^{op}$ is isomorphic to $M_n(F)$, where $n = \dim A$ and A^{op} is the opposite algebra of A, that is, A with multiplication in reverse order.*

Proof: (Sketch) A can be regarded as an A^e-module. Then A is irreducible and $End_{A^e} A = F$. Also, A is finite dimensional over F. Hence by the density theorem A^e maps onto $End_F A$. Since both A^e and $End_F A$ has dimension n^2, we therefore have an isomorphism of A^e onto $End_F A$. Since $End_F A \cong M_n(F)$, the result follows. $\square$

Now we are in a position to define the Brauer group over a field.

Definition 3.1.4 *In the Brauer Group $B(F)$ over a field F, the elements are similarity classes of central simple algebra. Let A and B be central simple algebras over F. We say that A and B are similar, denoted by $A \sim B$, if for some positive integers m, n such that $M_m(A) \cong M_n(B)$ as $F-algebras$, or equivalently $M_m(F) \otimes A \cong M_n(F) \otimes B$. If $[A]$ denotes the similarity class of A, the group operation is defined by $[A][B] = [A \otimes B]$*

The similarity condition is clearly reflexive and symmetric. Now if we have

$$M_m(F) \otimes A \cong M_n(F) \otimes B, \; and \; M_r(F) \otimes B \cong M_s(F) \otimes C$$

then consider

$$M_{mr}(F) \otimes A \cong M_r(F) \otimes M_m(F) \otimes A \cong M_r(F) \otimes M_n(F) \otimes B$$
$$\cong M_n(F) \otimes M_r(F) \otimes B \cong M_n(F) \otimes M_s(F) \otimes C \cong M_{ns}(F) \otimes C$$

Therefore the similarity relation is a equivalence relation.

Suppose we have $A \sim A'$ and $B \sim B'$. Then there exist positive integers m, m', n, n' such that $M_m(F) \otimes A \cong M'_m(F) \otimes A'$ and $M_n(F) \otimes B \cong M'_n(F) \otimes B'$. This implies that $M_{mn} \otimes A \otimes B \cong M_{m'n'} \otimes A' \otimes B'$. Hence $A \otimes B \sim A' \otimes B'$ and the binary group operation is well defined. Obviously the group operation is also associative and commutative. The identity element is $[F]$, that is, if $A \cong M_n(F)$ for some n, A belongs to the identity class. Finally, **Theorem 3.1.3** implies that $[A^{op}]$ is the inverse of $[A]$. Therefore we have that the Brauer group over a field F is an abelian group.

If A is finite dimensional central simple over F, it is also Artinian. By Artin-Wedderburn Theorem we can write $A \cong M_n(F) \otimes \Delta$, where Δ is a finite dimensional central division algebra. Conversely, if Δ is such an algebra, $M_n(F) \otimes \Delta$ is finite dimensional central simple over F. Also, since $M_n(\Delta)$ is a simple Artinian Ring, if $M_n(\Delta) =\cong M'_n(\Delta')$ for division algebras Δ, Δ', then $n = n'$ and $\Delta \cong \Delta'$. So the division algebra Δ in $A \cong M_n(F) \otimes \Delta$ is determined up to isomorphism. Thus a similarity class $[A]$ contains a single isomorphism class of finite dimensional central division algebras and distinct similarity classes are associated with non-isomorphic division algebras.

3.2 The Role of Quaternion Algebras in the Brauer Group

The tensor product $A = A_1 \otimes_F A_2$ of two quaternion algebras A_1, A_2 is called a biquaternion algebra. As a consequence of Wedderburn's theorem on central simple algebras, this 16-dimensional algebra is isomorphic to one of the following:

1. A is a division algebra.

2. A is split, i.e. A is isomorphic to $M_4(F)$.

3. A is isomorphic to $M_2(D)$ for some quaternion division algebra D.

These three cases correspond to different similarity classes in the Brauer group over F, since their division algebras are different. Being analogous to the norm form of a quaternion algebra, we can define the *Albert quadratic form* of the biquaternion algebra

$$A = \left(\frac{a_1, b_1}{F}\right) \otimes \left(\frac{a_2, b_2}{F}\right)$$

as the 6-dimensional quadratic form $\phi_A = \langle a_1, b_1, -a_1b_1, -a_2, -b_2, a_2b_2 \rangle$. A theorem of Albert says the following.

Theorem 3.2.1 *Let A be a biquaternion algebra, then*

1. *A is a division algebra if and only if ϕ_A is anisotropic.*

2. *A is split if and only if ϕ_A is hyperbolic, i.e. $\phi_A \cong \langle 1, -1, 1, -1, 1, -1 \rangle$.*

Otherwise, A is isomorphic to $M_2(D)$ for some quaternion division algebra D.

For a proof of this theorem, see T.Y.Lam [8] page 70.

Consider the special case when $A_1 = A_2 = (\frac{a,b}{F})$ and $B = A_1 \otimes A_2$, for some non-zero $a, b \in F$. Then

$$\phi_B = \langle a, b, -ab, -a, -b, ab \rangle$$

From Example 2(1) in **Section 2.2**, we know that $\langle a, -a \rangle$ represents 1. So by **Proposition 1.2.9**,

$$\langle a, -a \rangle \cong \langle 1, -a^2 \rangle \cong \langle 1, -1 \rangle$$

and we have $\phi_B = \langle 1, -1, 1, -1, 1, -1 \rangle$. That is, B is split, by the theorem of Albert. The above implies that if A is a quaternion algebra over F, we have $[A][A] = [A \otimes A] = 1$. Therefore non-split quaternion algebras have order 2 in the Brauer group.

Another way to view this is the following. If $A = (\frac{a,b}{F})$, then consider $u = i$ and $v = j$ in the *opposite algebra* A^{op}. Now

$$(u^2)^{op} = a,\ (v^2)^{op} = b,\ (uv)^{op} = (ij)^{op} = ji = -ij = -(ji)^{op} = -(vu)^{op}$$

So by the Identification Theorem for Quaternion Algebras, $A^{op} \cong (\frac{a,b}{F}) = A$. We have already seen that $[A^{op}]$ is the inverse of $[A]$, therefore we have

$$[A][A] = [A^{op}][A] = 1$$

and $[A]$ has at most order 2 in $B(F)$.

Brauer Groups Over Different Fields

1. By Frobenius's Theorem on Real Division Algebras in 1877, $B(R) \cong \{\pm 1\}$ for any real closed field R, with the only non-trivial element being $(\frac{-1,-1}{R})$.

2. The fact that there are infinitely non-isomorphic quaternion division algebras over the rationals $\mathbb{Q}$ was proven earlier. Therefore $B(\mathbb{Q})$ is infinite.

3. If F is the completion of a number field at a finite place, then there exists an isomorphism inv : $B(F) \cong \mathbb{Q}/\mathbb{Z}$. This is one of the central facts in local class field theory.

3.3 The Theorem of Merkurjev

In 1981, Merkurjev proved the conjecture suggested by Albert which is an implication of the following theorem.

Theorem 3.3.1 *(Merkurjev) Let k_2F denote the reduced Milnor K-theory group of the field F generated by the symbols $[a,b]$ and Br_2F be the subgroup of $B(F)$ generated by all the elements of order ≤ 2. The map $\alpha : k_2F \rightarrow Br_2F$ such that*

$$\alpha([a,b]) = \left[\left(\frac{a,b}{F}\right)\right]$$

for any $a, b \in \dot{F}$ is an isomorphism.

The proof of this theorem is nowhere close to trivial, and is way beyond the scope of this article. However, we can check intuitively why this is right.

The reduced Milnor K-theory group k_2F of the field F, is a multiplicative group generated by the bimultiplicative symbols $[a,b]$ with $a, b \in F$ satisfying the set of relations

$$\begin{aligned} [a, 1-a] &= 1 & (a \in \dot{F}, a \neq 1) \\ [a,b] &= [b,a] & (a,b \in \dot{F}) \\ [a,a] &= [a,-1] & (a \in \dot{F}) \end{aligned}$$

We have already seen that quaternion algebras over F satisfy the same kind of relations.

$$\left(\frac{a, 1-a}{F}\right) \cong \left(\frac{1, -1}{F}\right), \; \left(\frac{a, b}{F}\right) \cong \left(\frac{b, a}{F}\right), \; \left(\frac{a, a}{F}\right) \cong \left(\frac{a, -1}{F}\right)$$

Therefore the map α is well-defined.

Albert proved that a central simple F-algebra A has order ≤ 2 in $B(F)$ if and only if A has an F-involution, but was unable to show that such an algebra is a tensor product of quaternion algebras. Merkurjev's result provided an affirmative answer to Albert's question. The surjectivity of the map α amounts to the fact that any element of order 2 in $B(F)$ is expressible as a product of quaternion algebras.

Chapter 4

Characterization of Quaternion Algebras

4.1 Three Similar Theorems

As a consequence of Albert's work and Merkurjev's Theorem, we know that if A is an algebra which admits an F-involution, then it is a tensor product of quaternion algebras. And if A is of dimension 4 over F, then of course we have a quaternion algebra. One of the oldest and most important results is the theorem of Frobenius on real division algebras.

Theorem 4.1.1 ***Frobenius's Theorem on Real Division Algebras***
If $\mathcal{D}$ is a finite dimensional division algebra over $\mathbb{R}$, then $\mathcal{D} = \mathbb{R}$, $\mathcal{D} = \mathbb{R}(i) = \mathbb{C}$ or $\mathcal{D} = (\mathbb{H})$, the division algebra of real quaternions.

Note: The theorem is actually true for finite dimensional division algebras over any real closed field, here we will present a proof of it with the real closed field being $\mathbb{R}$. The proof is from Chapter 13 of [7]

Proof: If $\mathcal{D} = \mathbb{R}$, we are done. So we may assume that $dim_{\mathbb{R}}\mathcal{D} \geq 2$. Take an element $\alpha \in \mathcal{D}\backslash\mathbb{R}$, then $\mathbb{R}[\alpha]$ is a proper algebraic extension of $\mathbb{R}$, so $\mathbb{R}[\alpha] \cong \mathbb{C}$. Fix a copy of $\mathbb{C}$ in $\mathcal{D}$, and view $\mathcal{D}$ as a left vector space over $\mathbb{C}$. Let i denote the complex number $\sqrt{-1} \in \mathbb{C}$.

Let

$$\mathcal{D}^+ = \{d \in \mathcal{D} : \ di = id\} \supseteq \mathbb{C}$$

$$\mathcal{D}^- = \{d \in \mathcal{D} : \ di = -id\}$$

These are $\mathbb{C}$-subspaces of ${}_{\mathbb{C}}\mathcal{D}$ (left vector space) with $\mathcal{D}^+ \cap \mathcal{D}^- = 0$. Also, for any $d \in \mathcal{D}$, let $d^+ = id + di$, and $d^- = id - di$, then we have

$$id^+ = i^2 d + idi = -d + idi = di^2 + idi = d^+ i$$

so that $d^+ \in \mathcal{D}^+$, and similarly $d^- \in \mathcal{D}^-$. Since $d = (2i)^{-1}(d^+ + d^-) \in \mathcal{D}^+ + \mathcal{D}^-$, $\mathcal{D}$ is a direct sum of $\mathcal{D}^+$ and $\mathcal{D}^-$, i.e. $\mathcal{D} = \mathcal{D}^+ \oplus \mathcal{D}^-$.

For any $d^+ \in \mathcal{D}^+$, $\mathbb{C}[d^+] = \mathbb{C}$ since $\mathbb{C}$ is algebraically closed, thus $\mathcal{D}^+ = \mathbb{C}$. If $\mathcal{D}^- = 0$, we are done; $\mathcal{D} = \mathcal{D}^+ = \mathbb{C}$. Assume $\mathcal{D}^- \neq 0$. Fix an element $z \in \mathcal{D}^- \backslash \{0\}$ (so that $z \notin \mathbb{C}$). Consider the injective $\mathbb{C}$-linear map $\mu : \mathcal{D}^- \to \mathcal{D}^+$ sending $x \mapsto xz$. Since $dim_{\mathbb{C}}\mathcal{D}^+ = 1$, it follows that $dim_{\mathbb{C}}\mathcal{D}^- = 1$, and so

$$dim_{\mathbb{R}}\mathcal{D} = 2 \dim_{\mathbb{C}} \mathcal{D} = 4$$

Therefore the element z is algebraic over $\mathbb{R}$, but $\mathbb{C}$ is already the algebraic closure of $\mathbb{R}$, so $z^2 \in \mathbb{R} + \mathbb{R}z$. On the other hand, $z^2 = \mu(z) \in \mathcal{D}^+ = \mathbb{C}$ but $z \notin \mathbb{C}$, then

$$z^2 \in \mathbb{C} \cap (\mathbb{R} + \mathbb{R}z) = \mathbb{R}$$

If $z^2 > 0$ in $\mathbb{R}$, then $\pm z \in \mathbb{R}$, which contradicts $z \notin \mathbb{C}$. Thus $z^2 < 0$ in $\mathbb{R}$. Since every positive real number is a square, there exists $r \in \mathbb{R}$ such that $z^2 = -r^2$. Letting $j = z/r$, we have $i^2 = j^2 = -1$, and $ji = -ij$, which shows that

$$\mathcal{D} = \mathbb{C} \oplus \mathbb{C}j = \mathbb{R} \oplus \mathbb{R}i \oplus \mathbb{R}j \oplus \mathbb{R}ij$$

and $\mathcal{D}$ is a copy of the real quaternions. □

In case the center of the division algebra is not $\mathbb{R}$, but a general field F, we can identify a quaternion algebra using the following theorem. The proof of this theorem is exactly similar to the above theorem.

Theorem 4.1.2 *Let $A \neq F$ be a simple F-algebra of dimension ≤ 4 with center F. Then A is isomorphic to a quaternion algebra F. (Section 3.5 of [8])*

Proof: By Wedderburn's Theorem, $A \cong M_n(D)$ for some positive integer n and some division algebra D. Since $dim_F A \leq 4$, we have $A \cong M_2(F) \cong (\frac{1,1}{F})$ (in which case we are done), or $n = 1$ and A is a division algebra of dimension 4 over F. Say A is a division algebra, fix a non-central element $i \in A$, and let K be the field $F(i)$. Since K is a field, the center of it is K, and thus K cannot be equal to A. Also, $F \subsetneq K \subsetneq A$, implies that $dim_F K = 2$ and $dim_F A = 4$. Therefore K is a quadratic field extension of F, and we may assume that $i \in K$ to have been chosen such that $i^2 = a \in F \backslash \{0\}$ (if $\operatorname{char} F \neq 2$). Let $f : A \to A$ be the inner automorphism $f(x) = i^{-1}xi$. Then $f^2 = Id$, and we have, as in the previous proof, an eigenspace decomposition $A = A^+ \oplus A^-$, where

$$A^+ = \{a \in A : \ f(a) = a\} = \{a \in A : \ ai = ia\}$$

$$A^- = \{a \in A : \ f(a) = -a\} = \{a \in A : \ ai = -ia\}$$

Now fix an element $j \in A^-$. Since $K \subseteq A^+$ and $K \cdot j \subseteq A^-$, we must have $K = A^+$ and $K \cdot j = A^-$, by considering their dimensions as an F-module and the fact that $dim_F A = 4$. Since $j \in A^-$, $ij = -ji$ and $ij^2 = j^2 i$; that is, $j^2 \in A^+ = K$. Also, $F(j)$ is a quadratic field extension of F, so j satisfies a quadratic equation $j^2 + cj - b = 0$, for some $b, c \in F$. Now $cj = b - j^2 \in K$ implies that $c = 0$ and $j^2 = b \in F \backslash 0$. Therefore

$$A = K \oplus K \cdot j = F \oplus Fi \oplus Fj \oplus Fij \cong \left(\frac{a, b}{F} \right)$$

and A is isomorphic to a quaternion algebra.$\square$

As another way to characterize quaternion algebras, we will also present a proof of the following theorem. Gerstenhaber and Yang [3] gave a proof of a modified form of the theorem of Frobenius stated above, by weakening the assumption and the conclusion. The exact statement is as follows.

Theorem 4.1.3 ***Modified Frobenius's Theorem***
If $\mathcal{D}$ is a division ring containing a real closed field $\mathcal{R}$, such that $\mathcal{D}$ is a finite dimensional left vector space over $\mathcal{R}$, then either $\mathcal{D} = \mathcal{R}$, $\mathcal{D} = \mathcal{R}(i)$ or $\mathcal{D}$ is the quaternion algebra over a real closed field $\mathcal{F}$ such that $\mathcal{F}(i) \cong \mathcal{R}(i)$

The proof of this theorem requires quite a lot of tools from the theory of real closed fields, and here are some of them that we will see in the proof.

4.2 Properties of Real Closed Fields

Recall the definitions of formally real fields and real closed fields. (From [5] and [9])

Definition 4.2.1 *A field is called formally real if $\sum_{r=1}^{n} a_r^2 = 0$ iff $a_r = 0$ for any r.*

Definition 4.2.2 *A field Φ is called real closed if Φ is formally real and no proper algebraic extension of Φ is formally real.*

Theorem 4.2.3 *If Φ is real closed, then any element of Φ is either a square or negative of a square.*

Proof: Say $a \in \Phi$ is not a square. Then $\Omega = \Phi(\sqrt{a})$ is a proper algebraic extension of Φ. And since no algebraic extension of a real closed field is formally real, Ω is not formally real. Therefore, there exist $b_i, c_i \in \Phi$, c_i not all zero, *s.t.*

$$\sum (b_i + c_i\sqrt{a})^2 = 0$$

Expanding, we have

$$\sum (b_i^2 + c_i^2 a) + \sum 2b_i c_i \sqrt{a} = 0$$

Since $\sqrt{a} \notin \Phi$, we have $\sum (b_i^2 + c_i^2 a) = 0 = \sum 2b_i c_i$.
Here, $\sum c_i^2 \neq 0$ since Φ is formally real. Moreover, $\Sigma(\Phi)$, the set of sums of squares, is closed under addition, multiplication and inverse.(To see why $\Sigma(\Phi)$ is closed under inverse, if α is a sum of squares, then $\alpha\alpha^{-2}$ is a sum of squares.) Thus

$$-a = (\sum b_i^2)(\sum c_i^2)^{-1} \in \Sigma(\Phi)$$

But $-1 \notin \Sigma(\Phi)$ since Φ is formally real. This implies that $a \notin \Sigma(\Phi)$. This shows that if an element is not a square, then it is not a sum of squares. Taking the contrapositive, if an element is a sum of squares, then it is actually a square. But we have already shown that if a is not a square, then $-a \in \Sigma(\Phi)$ which implies that $-a$ is a square.

Therefore, either a is a square, or $-a$ is a square.$\square$

Theorem 4.2.4 *Φ is a real closed field if and only if Φ is a field, $i \notin \Phi$ and $\Phi(i)$ is algebraically closed.*

Proof: First assume that Φ is a real closed field. Clearly, $\sqrt{-1} \notin \Phi$, let's consider the algebraic extension $\Phi(\sqrt{-1})$ of Φ.

Step 1: Show that every element in $\Phi(\sqrt{-1})$ has a square root in $\Phi(\sqrt{-1})$.

Proof of 1: First of all, if $\alpha \in \Phi$, then we proved that α is either a square or the negative of a square in Φ. But since $\sqrt{-1} \in \Phi(\sqrt{-1})$, the negative of a square in Φ is a square in $\Phi(\sqrt{-1})$.

Check that for any $x \in \Phi$, $x + \sqrt{x^2+1} \geq 0$ or $x - \sqrt{x^2+1} \geq 0$

otherwise, $x + \sqrt{x^2+1} < 0$ and $x - \sqrt{x^2+1} < 0$

$\Rightarrow (x + \sqrt{x^2+1})(x - \sqrt{x^2+1}) > 0$

$\Rightarrow -1 = x^2 - x^2 - 1 > 0$ which is a contradiction.

Let $\sqrt{-1} = i$. Now consider a general element $\alpha + \beta i$ where $\beta \neq 0$ and $\alpha, \beta \in \Phi$.

Let $\sigma = \sqrt{\frac{\alpha}{\beta} + \sqrt{\frac{\alpha^2}{\beta^2} + 1}}$ if $\left(\frac{\alpha}{\beta} + \sqrt{\frac{\alpha^2}{\beta^2} + 1}\right) > 0$

If not, then let $\sigma = \sqrt{\frac{\alpha}{\beta} - \sqrt{\frac{\alpha^2}{\beta^2} + 1}}$ with $\left(\frac{\alpha}{\beta} - \sqrt{\frac{\alpha^2}{\beta^2} + 1}\right) > 0$

Therefore, $\sigma \in \Phi$. Then one can check that

$$\left(\sqrt{\frac{\beta}{2}}\left(\sigma + i\sigma^{-1}\right)\right)^2 = \alpha + \beta i$$

where $\sqrt{\frac{\beta}{2}}\sigma \in \Phi$ and $\sqrt{\frac{\beta}{2}}\sigma^{-1} \in \Phi$. This finishes Step 1.

Step 2: Show that for any $f(x) \in \Phi[x]$, $f(x)$ has a root in $\Phi(\sqrt{-1})$.

Proof of 2: Let $f(x) \in \Phi[x]$. Let E be the splitting field of $(x^2+1)f(x)$ over Φ. That means $\sqrt{-1} \in E$, so we may assume that $E \supseteq \Phi(\sqrt{-1})$. We have seen that ordered fields, and thus real closed fields, have characteristic zero. That means E is also of characteristic zero, which in turn implies that the extension E/Φ is separable. Therefore, E is Galois over Φ. Let G be the Galois group, say $G = 2^n m$ where m is odd. By Sylow's First Theorem, Sylow 2-subgroups exist. Let H be a subgroup of G of order 2^n. Let F be the corresponding subfield of E, *i.e.* the subfield fixed by automorphisms in H. We have $[E : F] = 2^n$ and $[F : \Phi] = m$. However, since Φ is a real closed field, by **Theorem ??**, every polynomial of odd degree is reducible in Φ. This implies that Φ does not have an extension of odd degree. So, $m = 1$, $F = \Phi$ and $G = H$. Since G has order 2^n, G is solvable.

If $n = 1$, $E = \Phi(\sqrt{-1})$ and this means $\Phi(\sqrt{-1})$ is the splitting field of $(x^2 + 1)f(x)$. Therefore, $f(x)$ has a root in $\Phi(\sqrt{-1})$ and we are done.

If $n > 1$ and $E \neq \Phi(\sqrt{-1})$, by the Galois Correspondence, there is a subfield K of E such that $[K : \Phi(\sqrt{-1})] = 2$. However, by the result of Step 1, we have that every polynomial of degree 2 over $\Phi(\sqrt{-1})$ is reducible. *Therefore, there does NOT exist an algebraic extension of* $\Phi(\sqrt{-1})$ *of degree 2.* And this gives a contradiction. Done for Step 2.

Finally, for any $g(x) \in \Phi(\sqrt{-1})[x]$, $g(x)\overline{g(x)} \in \Phi[x]$, where the bar on top denotes the conjugate. If a is a root of $g(x)\overline{g(x)}$, then $\bar{a}$ is also a root, since $x - a$ and $x - \bar{a}$ both divide the polynomial. This implies that either a or $\bar{a}$ is a root of $g(x)$. But we have already shown that every polynomial in $\Phi[x]$ has a root in $\Phi(\sqrt{-1})$, so $g(x)$ must have a root in $\Phi(\sqrt{-1})$. This shows that $\Phi(\sqrt{-1})$ is algebraically closed, and we are done for forwards.

Conversely, assume that Φ is a field, $i \notin \Phi$ and $\Phi(i)$ is algebraically closed.

Let $f(x) \in \Phi[x]$ be an irreducible polynomial and let θ be a root of $f(x)$ in $\Phi(i)$. Then $[\Phi(\theta) : \Phi] \leq [\Phi(i) : \Phi] = 2$. So the irreducible polynomials in $\Phi[x]$ has degree 1 or 2.

Next, let's show that Φ is formally real. Consider the polynomial $g(x) \in \Phi[x]$, where

$$g(x) = (x^2 - a)^2 + b^2$$

with $a, b \in \Phi$, $a \neq 0 \neq b$, then

$$g(x) = (x - \sqrt{a + bi})(x + \sqrt{a + bi})(x - \sqrt{a - bi})(x + \sqrt{a - bi})$$

Therefore the linear factors are not in $\Phi[x]$, which means that $g(x)$ factors as two irreducible polynomials in $\Phi[x]$.

However $(x - \sqrt{a - bi})(x + \sqrt{a - bi}) = x^2 - (a - bi)$ is not in $\Phi[x]$. Same for the two linear factors with $\sqrt{a + bi}$. This implies that the only possible irreducible factors of $g(x)$ are

$$(x + \sqrt{a + bi})(x + \sqrt{a - bi}) \quad and \quad (x - \sqrt{a + bi})(x - \sqrt{a - bi})$$

or

$$(x - \sqrt{a + bi})(x + \sqrt{a - bi}) \quad and \quad (x + \sqrt{a + bi})(x - \sqrt{a - bi})$$

In either case, we have $\sqrt{a^2 + b^2} \in \Phi$. In other words, a sum of two non-zero squares is a square in Φ. Inductively, any sum of squares is a square. Since i is not in Φ, -1 is not a sum of squares, implying that Φ is formally real.

Since the degree of irreducible polynomials in $\Phi[x]$ is 1 or 2, any proper algebraic extension of Φ is isomorphic to $\Phi(i)$, so the extension is not formally real. Φ is real closed.$\square$

Theorem 4.2.5

A proper algebraic extension of a real closed field is algebraically closed.

Proof: Let $\mathcal{R}$ be a real closed field. Let α be algebraic over $\mathcal{R}$.

If $i \in \mathcal{R}(\alpha)$, then $\mathcal{R}(\alpha)$ contains $\mathcal{R}(i)$ which is algebraically closed. That means

$\mathcal{R}(i)$ contains the element α since α is algebraic over $\mathcal{R}$. Thus $\mathcal{R}(i) = \mathcal{R}(\alpha)$ by double inclusion. In this case, $\mathcal{R}(\alpha)$ is a proper algebraic extension of $\mathcal{R}$ and is algebraically closed.

Instead if $i \notin \mathcal{R}(\alpha)$, consider $\mathcal{R}(\alpha, i)$. Since α is algebraic over $\mathcal{R}$, and $\mathcal{R}(i)$ is algebraically closed, $\mathcal{R}(\alpha, i) = \mathcal{R}(i)$. We thus have $[\mathcal{R}(\alpha, i) : \mathcal{R}] = 2$. Also, $i \notin \mathcal{R}(\alpha)$ implies that $[\mathcal{R}(\alpha, i) : \mathcal{R}(\alpha)] = 2$. Moreover, we also have $\mathcal{R}(\alpha) \supseteq \mathcal{R}$, that means $\mathcal{R}(\alpha) = \mathcal{R}$. Therefore $\mathcal{R}(\alpha)$ is not a proper extension of $\mathcal{R}$. $\square$

Theorem 4.2.6 ***(Artin)***

If $\mathcal{C}$ is any algebraically closed field of characteristic zero and $\mathcal{R}$ is a proper subfield of $\mathcal{C}$ such that $[\mathcal{C} : \mathcal{R}] < \infty$, then $[\mathcal{C} : \mathcal{R}] = 2$ and $\mathcal{C} = \mathcal{R}(i)$.

The proof of this theorem can be found in Jacobson, Basic Algebra II, Second Edition, p.674. One can show that $\mathcal{R}$ can be ordered by defining a non-zero element of $\mathcal{R}$ to be positive if it is the norm of an element in $\mathcal{C}$. It follows that $\mathcal{R}$ is a real closed field.

Remark: When $\mathcal{C}$ is a larger field than that of all algebraic numbers then the real closed field $\mathcal{R}$ is not determined up to isomorphism.

4.3 $\mathbb{R}$ need not be in the center of $\mathcal{D}$

Referring back to the Modified Frobenius Theorem (**Theorem 4.1.3**), in this section, we are going to see that the division ring D does not necessarily have $\mathbb{R}$ in the center.

(From a paper of A. BIAŁYNICKI-BIRULA [1].)

Proposition 4.3.1 *Let $\mathcal{R}$ be a real closed field of power continuum, then the field $\mathcal{R}(\sqrt{-1})$ is isomorphic to the field of complex numbers.*

Proof: Since $\mathcal{R}$ is real closed, $\mathcal{R}(i)$ is algebraically closed, by **Theorem 4.2.4**. Also, algebraically closed fields of power continuum and of characteristic zero

are isomorphic to the field of complex numbers.
Therefore if $\mathcal{R}$ is a real closed field of power continuum, then $\mathcal{R}(i)$ is isomorphic to the field of complex numbers.$\square$

There exist non-isomorphic real closed fields of power continuum, for example the field of all real numbers $\mathbb{R}$ and the real closure of the ordered field $\mathbb{R}(t)$, where $0 < t < a$ for any positive $a \in \mathbb{R}$. See [1] and page 655 and 656 of [4].

Also, by the **Remark** after **Theorem 4.2.6**, an immediate consequence of the proposition is as follows:

Corollary 4.3.2 *There exist non-isomorphic real closed subfields $\mathcal{R}$ and $\mathcal{R}'$ of the field of complex numbers $\mathbb{C}$ such that $\mathcal{R}(i)$ and $\mathcal{R}'(i)$ are isomorphic to $\mathbb{C}$.*

Since $\mathbb{C}$ contains $\mathbb{R}$ (which is real closed), there exists a subfield $\mathcal{R}'$ of $\mathbb{C}$ such that $\mathbb{R} \ncong \mathcal{R}'$ and $\mathbb{C} = \mathbb{R}(i) \cong \mathcal{R}'(i)$.
Now let a quaternion algebra over $\mathcal{R}'$ be $\mathcal{Q}$. Since $\mathcal{Q}$ contains $\mathcal{R}'(i) \cong \mathbb{C}$, $\mathcal{Q}$ contains a copy $\mathcal{R}$ of $\mathbb{R}$. Also, $\mathcal{Q}$ has dimension two over $\mathbb{C} \cong \mathcal{R}(i) \cong \mathcal{R}'(i)$, therefore $\mathcal{Q}$ is a four dimensional left vector space over $\mathcal{R}$. Since $\mathcal{R}$ is not isomorphic to $\mathcal{R}'$, $\mathcal{R}$ is not in the center of $\mathcal{Q}$.
This example shows that a division ring $\mathcal{D}$ may contain $\mathbb{R}$ and be a finite dimensional left vector space over $\mathbb{R}$, but that $\mathbb{R}$ need not be contained in the center of $\mathcal{D}$.

4.4 A Few Lemmas and the Proof

Assuming $\mathcal{D}$ is a division ring containing a real closed field $\mathcal{R}$ such that $\mathcal{D}$ is a finite dimensional left vector space over $\mathcal{R}$. Let $i \in \mathcal{D}$ satisfy $i^2 = -1$. We distinguish two cases:

Case 1: $\mathcal{R}$ is a maximal subfield of $\mathcal{D}$.

In other words, there is no (commutative) field $\mathcal{F}$ contained in $\mathcal{D}$ and properly containing $\mathcal{R}$. Let $\mathcal{Z}$ = center of $\mathcal{D}$. We can say that $i \notin \mathcal{Z}$. (If i were in $\mathcal{Z}$, $\mathcal{R}(i)$ would be commutative, and is a proper extension of $\mathcal{R}$ in $\mathcal{D}$)

Let $\mathcal{A} = \mathcal{D} \otimes \mathcal{Z}(i)$, tensor product taken over $\mathcal{Z}$. This is the ring obtained by extending $\mathcal{D}$ to have i in it's center.

By identifying $\mathcal{D}$ and the subring $\{d \otimes 1 : d \in \mathcal{D}\}$, we may say that $\mathcal{D}$ is contained in $\mathcal{A}$, and every element in $\mathcal{A}$ may be written in the form $a + bi$ with a,$b \in \mathcal{D}$. (Since $d \otimes (z_1 + z_2 i) = d \otimes z_1 + d \otimes z_2 i = dz_1 \otimes 1 + dz_2 \otimes i$) Hence the algebra $\mathcal{A}$ contains a copy of $\mathcal{R}(i)$, which we will denote by $\mathcal{C}$ and which is algebraically closed. Any basis of $\mathcal{D}$ over $\mathcal{R}$ is also a basis of $\mathcal{A}$ over $\mathcal{C}$.

Case 2: $\mathcal{R}$ is not a maximal subfield of $\mathcal{D}$

Since $\mathcal{D}$ is a finite dimensional left vector space over $\mathcal{R}$, $\mathcal{D}$ contains a proper algebraic extension $\mathcal{C}$ of $\mathcal{R}$. However, by **Theorem 4.2.5**, $\mathcal{C}$ can only be the algebraic closure of $\mathcal{R}$. We can write $\mathcal{D} \supseteq \mathcal{C} = \mathcal{R}(i)$. Note that here, i commutes with every element in $\mathcal{R}$. Now, let $\mathcal{A} = \mathcal{D}$, so that in both cases, $\mathcal{A}$ contains an algebraically closed field $\mathcal{R}(i)$.

In this part, a few lemmas will be proven, in order to prove **Theorem 4.1.3**. In the following proofs, we will often refer to *Case 1* and *Case 2* described above. When a statement about $\mathcal{A}$ does not specify which case we are dealing with, it will be meant to hold for both.

Lemma 3 *Let x, y be non-zero elements of $\mathcal{A}$. Then there exists $\lambda \in \mathcal{R}$ such that $x\lambda y \neq 0$.*

Proof: In *Case 1*, if $xy \neq 0$ we can take $\lambda = 1$ so that $x\lambda y \neq 0$. Thus we may assume that $xy = 0$. By writing $x = a + bi$ and $y = c + di$, where $a, b, c, d \in \mathcal{D}$, we have $xy = 0$ if and only if $a^{-1}b + dc^{-1} = 0$ and $(dc^{-1})^2 = -1$ by the following observations.

Observation 1: In *Case 1* $\mathcal{A}$ is obtained by extending $\mathcal{D}$ so that i is in the center, therefore i commutes with a, b, c, and d.

Assume $xy = 0$, then $0 = xy = (a + bi)(c + di) = ac - bd + (ad + bc)i$.

Keep in mind that these products are tensor products, only written like usual products. And the tensor product is taken over $\mathcal{Z}$ which does not contain i. It should actually be written like $0 = (ac - bd) \otimes 1 + (ad + bc) \otimes i$.

From here it's obvious that $(ad + bc) = 0$ which means $(dc^{-1} + a^{-1}b) = 0$ since $\mathcal{D}$ is a division ring.

Observation 2: Assume $xy = 0$, then

$$0 = (a + bi)(c + di) = (a + bi)(a - bi)(c + di)(c - di) = (a^2 + b^2)(c^2 + d^2)$$

and since $\mathcal{D}$ is a division ring, we have $c^2 + d^2 = 0$ and so $(dc^{-1})^2 = -1$.

Observation 3: If $a^{-1}b + dc^{-1} = 0$ and $(dc^{-1})^2 = -1$, therefore

$$0 = c^2 + d^2 = (a^2 + b^2)(c^2 + d^2) = (a + bi)(c + di)(a - bi)(c - di)$$

so that $((ac - bd) + (ad + bc)i)((ac - bd) - (ad + bc)i) = 0$. By the assumption, $ad + bc = 0$, therefore we have $(ac - bd)^2 = 0$ implying that $(ac - bd) = 0$. With $ad + bc = 0$ and $(ac - bd) = 0$, it's easy to see that $xy = (a + bi)(c + di) = 0$.

Now $(dc^{-1})^2 = -1$ implies that $dc^{-1} \notin \mathcal{R}$, since $\mathcal{R}$ is formally real. If it's not in $\mathcal{R}$ but it commutes with everything in $\mathcal{R}$, this would give a proper commutative extension of $\mathcal{R}$ in $\mathcal{D}$. However, $\mathcal{R}$ is maximal in $\mathcal{D}$ in *Case 1*, so this can't happen. Therefore, there exists a $\lambda \in \mathcal{R}$ such that $\lambda dc^{-1}\lambda^{-1} \neq dc^{-1}$. Then $a^{-1}b + (\lambda d)(\lambda c)^{-1} \neq 0$, hence $x\lambda y \neq 0$.

And for *Case 2*, $\mathcal{A}$=$\mathcal{D}$ which is a division ring and we can take $\lambda = 1$. $\square$

Right multiplication by $a \in \mathcal{A}$ to elements in $\mathcal{A}$ is a linear transformation on $\mathcal{A}$ (as a vector space over $\mathcal{C}$). Let this transformation be $R_a : \mathcal{A} \to \mathcal{A}$ such that $R_a(x) = xa$ for any $x \in \mathcal{A}$. In particular R_λ is defined for all $\lambda \in \mathcal{C}$.
Left multiplication L_a by an element $a \in \mathcal{A}$ is not always a linear transformation over $\mathcal{C}$. L_a is a $\mathcal{C}$-linear transformation if and only if a commutes with every element of $\mathcal{C}$. In particular, L_λ is a linear transformation for every $\lambda \in \mathcal{C}$.
For $\lambda \in \mathcal{C}$, let $C_\lambda = R_\lambda - L_\lambda$, so that $C_\lambda(x) = x\lambda - \lambda x$ for all $x \in \mathcal{A}$. Note that if $\lambda \in \mathcal{R}$, then C_λ is a linear transformation of $\mathcal{D}$ into itself. Also C_λ is a derivation of $\mathcal{D}$. (A derivation is a function $d : \mathcal{D} \to \mathcal{D}$ satisfying $d(ab) = ad(b) + d(a)b$)

Lemma 4 *If $\lambda \in \mathcal{R}$ and C_λ is nilpotent, then C_λ is the zero transformation. i.e. λ is in the center of $\mathcal{D}$.*

Proof: Let a be an element of $\mathcal{D}$. First let's assume ${C_\lambda}^2(a) = 0$. C_λ is a derivation of $\mathcal{D}$, so we have $C_\lambda(a^2) = aC_\lambda(a) + C_\lambda(a)a$ and so

$$ {C_\lambda}^2(a^2) = C_\lambda(a)C_\lambda(a) + a{C_\lambda}^2(a) + {C_\lambda}^2(a)a + C_\lambda(a)C_\lambda(a) = 2C_\lambda(a)^2 $$

and ${C_\lambda}^3(a^2) = 2C_\lambda(a){C_\lambda}^2(a) + 2{C_\lambda}^2(a)C_\lambda(a) = 0$ With these base cases, applying induction, one can prove that ${C_\lambda}^m(a^n) = 0$ if $m > n$, and

$$ {C_\lambda}^n(a^n) = n!C_\lambda(a)^n \quad \text{for} \quad n > 1 $$

Since $\mathcal{D}$ is finite dimensional over $\mathcal{R}$, a satisfies some equation of the form

$$ a^n + r_1 a^{n-1} + \cdots + r_n = 0 $$

Applying C_λ on the above equation n times, we have $n!C_\lambda(a)^n = 0$, which means $C_\lambda(a)$ is nilpotent. However $C_\lambda(a)$ belongs to $\mathcal{D}$ which is a division algebra, that means $C_\lambda(a) = 0$. Here we proved that if $a \in \mathcal{D}$ and ${C_\lambda}^2(a) = 0$ then $C_\lambda(a) = 0$.

As in the assumption in the lemma, C_λ is nilpotent. For $b \in \mathcal{D}$, $\exists$ a least non-negative integer m such that ${C_\lambda}^m(b) = 0$. But if $m > 1$,

$${C_\lambda}^m(b) = {C_\lambda}^2({C_\lambda}^{m-2}(b)) = 0$$

and from above, we know that $C_\lambda({C_\lambda}^{m-2}(b)) = 0$, that is ${C_\lambda}^{m-1}(b) = 0$ which contradicts the minimality of m. Therefore $m = 1$ and C_λ is the zero transformation. This finishes the proof.$\square$

The transformations R_λ commute for all $\lambda \in \mathcal{C}$. Therefore they have the same generalized eigenspaces. Thus, $\mathcal{A}$ can be written as a direct sum of subspaces $\mathcal{A} = A_1 + \cdots + A_k$ such that,

1) $R_\lambda(A_n) \subseteq A_n$ for $\lambda \in \mathcal{C}$ and $n = 1, 2, \cdots, k$,

2) for each λ and n, R_λ has only one generalized eigenvalue λ_n in A_n, i.e. $R_\lambda - L_{\lambda_n}$ restricted to A_n is nilpotent.

3) Each A_n is irreducible. In other words, it cannot be expressed as a direct sum of proper subspaces with properties 1) and 2).

In each A_n, there exists a non-zero d_n, unique up to left multiplication by elements in $\mathcal{C}$, which is an eigenvector simultaneously for all R_λ for $\lambda \in \mathcal{C}$. That is, given $\lambda \in \mathcal{C}$ and n, there exists $\lambda_n \in \mathcal{C}$ such that $d_n\lambda = \lambda_n d_n$. Define $\sigma_n : \mathcal{C} \to \mathcal{C}$ by $\sigma_n(\lambda) = \lambda_n$. If $\lambda \neq 0$, then $\lambda_n \neq 0$. And since d_n is an eigenvector for all R_λ simultaneously, σ_n will preserve addition and multiplication, so it is an isomorphism of $\mathcal{C}$ into itself.

Note that if for some $\lambda \in \mathcal{C}$ we have $\sigma_n(\lambda) = \lambda$, then $R_\lambda - L_\lambda$ is nilpotent. (This is explained in the second property of the decomposition of $\mathcal{A}$.)

Lemma 5 *Let $a_1, \cdots, a_m$ be non-zero elements of $\mathcal{A}$ and suppose that for each n there exists an isomorphism τ_n from $\mathcal{C}$ to $\mathcal{C}$ such that $a_n\lambda = \tau_n(\lambda)a_n$ for all $\lambda \in \mathcal{C}$. If the τ_n are distinct, for $n = 1, \cdots, m$, then $\{a_1, \cdots a_m\}$ is linearly independent over $\mathcal{C}$.*

Proof: Assuming the contrary, there is a relation of the form $\sum \mu_n a_n = 0$ with $\mu_n \in \mathcal{C}$ not all zero. A is a division ring, so the relation has more than one term. Suppose that this relation has the minimum number of non-zero terms. If for any $\lambda \in \mathcal{C}$, $\tau_j(\lambda) = \tau_i(\lambda)$, for all i, j such that $\mu_i \neq 0$, $\mu_j \neq 0$, these τ_j will not be distinct, which is a contradiction. Let λ be an element of $\mathcal{C}$ and fix a j such that $\tau_j(\lambda) \neq \tau_i(\lambda)$ for some $i \neq j$ such that $\mu_i, \mu_j \neq 0$. Then, $\tau_j(\lambda)^{-1} \sum \mu_n a_n \lambda = 0$ and using $a_n\lambda = \tau_n(\lambda)a_n$ given in the lemma, we have $\sum \mu_n \tau_j(\lambda)^{-1}\tau_n(\lambda)a_n = 0$. Subtracting this from $\sum \mu_n a_n = 0$ gives a relation with fewer non-zero terms, which is a contradiction. The lemma is proven. □

Lemma 6 *Let σ and τ be isomorphisms from $\mathcal{C}$ to $\mathcal{C}$ and suppose there exist non-zero elements a and b of $\mathcal{A}$ such that $a\lambda = \sigma(\lambda)a$ and $b\lambda = \tau(\lambda)b$ for all $\lambda \in \mathcal{C}$. Then there is a non-zero $c \in \mathcal{A}$ such that $c\lambda = \sigma(\tau(\lambda))c$ for all $\lambda \in \mathcal{C}$.*

Proof: By **Lemma 3**, there exists a $\mu \in \mathcal{C}$ such that $a\mu b \neq 0$. Let $c = a\mu b$, then

$$c\lambda = (a\mu b)\lambda = a\tau(\lambda)b = a\tau(\lambda)\mu b = \sigma(\tau(\lambda)(a\mu b) = \sigma(\tau(\lambda))c$$

Lemma 7 *The isomorphisms $\sigma_1, \sigma_2, \cdots, \sigma_k$ of $\mathcal{C}$ generate a finite group G of order 1 or 2.*
Let $\mathcal{F} \subseteq \mathcal{C}$ denote the fixed field of G. i.e. $\lambda \in \mathcal{F}$ if and only if $\sigma(\lambda) = \lambda$ for all $\sigma \in G$. We have either $[\mathcal{C} : \mathcal{F}] = 2$, $\mathcal{F}$ is a real closed field, and $\mathcal{C} = \mathcal{F}(i)$, or else $\mathcal{C} = \mathcal{F}$.

Proof: Let G denote the semigroup generated by $\sigma_1, \sigma_2, \cdots, \sigma_k$. By **Lemma 6**, if $\tau \in G$, then there exists an $a_\tau \in \mathcal{A}$ such that $a_\tau\lambda = \tau(\lambda)a_\tau$ for all $\lambda \in \mathcal{C}$.

(The a and b in **Lemma 6** exist because of the existence of eigenvectors d_n in each A_n) By **Lemma 5** the a_τ corresponding to distinct elements τ of G are linearly independent over $\mathcal{C}$, hence finite in number since $\mathcal{A}$ is finite dimensional over $\mathcal{C}$. Therefore G is finite, from which it follows that G is a group. Since $[\mathcal{C} : \mathcal{F}]$ equals the order of the group G, which is finite, by Artin's Theorem, either $\mathcal{C} = \mathcal{F}$ (in which case G has order 1), or $[\mathcal{C} : \mathcal{F}]$, $\mathcal{F}$ is real closed, $\mathcal{C} = \mathcal{F}(i)$ and G has order 2. □

Note that any element of $\mathcal{C}$ which is in the center of $\mathcal{A}$ must commute with a_τ, as defined above, for any $\tau \in G$. Therefore it must be in the fixed field $\mathcal{F}$ of G.

Proposition 4.4.1

Let $\mathcal{D}$ be a division ring over a real closed field $\mathcal{R}$ over which $\mathcal{D}$ has finite dimension as a left vector space. If $\mathcal{R}$ is a maximal subfield of $\mathcal{D}$ (i.e. Case 1), then $\mathcal{D} = \mathcal{R}$.

Proof: By construction in *Case 1*, i is in the center of $\mathcal{A}$ and therefore, as explained above, in the fixed field $\mathcal{F}$ of G. Therefore $\mathcal{F}$ is not a real field and by **Lemma 7**, $\mathcal{C} = \mathcal{F}$. It follows that the generalized eigenvalues of R_λ are equal to λ for any $\lambda \in \mathcal{C}$ and so $R_\lambda - L_\lambda$ is nilpotent. In particular this is true for any $\lambda \in \mathcal{R}$, by **Lemma 4**, λ is in the center of $\mathcal{D}$. So $\mathcal{R}$ is in the center of $\mathcal{D}$. By the general form of Frobenius Theorem (**Theorem 4.1.1**), $\mathcal{D} = \mathcal{R}$, $\mathcal{D} = \mathcal{R}(i)$ or $\mathcal{D}$ is the quaternion algebra over $\mathcal{R}$. However for the last two cases $\mathcal{R}$ is not a maximal subfield of $\mathcal{D}$, so we conclude that $\mathcal{D} = \mathcal{R}$.

Proposition 4.4.2

In Case 2, $\mathcal{F}$ is the center of $\mathcal{D}$.

Proof: The center $\mathcal{Z}$ of $\mathcal{D}$ must be contained in $\mathcal{C}$, since $\mathcal{C} \subseteq \mathcal{D}$ is algebraically closed. In *Case 2*, $\mathcal{A} = \mathcal{D}$, so $\mathcal{C} \cap \mathcal{Z}$ is in $\mathcal{F}$. For the other containment, if

$\lambda \in \mathcal{F}$, then λ is fixed by the group G, and so R_λ has generalized eigenvalue λ. Therefore $R_\lambda - L_\lambda$ is nilpotent, and by **Lemma 4**, λ is in $\mathcal{Z}$. So the proposition is proved by double inclusion.$\square$

$\mathcal{C}$ has dimension 2 over $\mathcal{F}$ and $\mathcal{D}$ is finite dimensional over $\mathcal{C}$, so $\mathcal{D}$ is a finite dimensional division algebra over the real closed field $\mathcal{F}$ and we can then apply the Frobenius Theorem. Although we have $\mathcal{C} = \mathcal{F}(i)$, and $\mathcal{C} = \mathcal{R}(i)$, we cannot say that $\mathcal{R} \cong \mathcal{F}$, as explained before using the power continuum argument. (Note that if $\mathcal{R}$ is the field of real algebraic numbers, then $\mathcal{R} \cong \mathcal{F}$)
The results we have now is a proof of **Theorem 4.1.3**, which we'll restate below.

Theorem 4.1.3 ***Modified Frobenius's Theorem***

If $\mathcal{D}$ is a division ring containing a real closed field $\mathcal{R}$, such that $\mathcal{D}$ is a finite dimensional left vector space over $\mathcal{R}$, then either $\mathcal{D} = \mathcal{R}$, $\mathcal{D} = \mathcal{R}(i)$ or $\mathcal{D}$ is the quaternion algebra over a real closed field $\mathcal{F}$ such that $\mathcal{F}(i) \cong \mathcal{R}(i)$

References

[1] A. Białynicki-Birula, On Subfields of Countable Codimension, Proceedings of AMS, Volume 35, Number 2, October 1972.

[2] Pete L. Clark, Sums of Two Squares, University of Georgia http://math.uga.edu/∼pete/4400twosquares.pdf

[3] Murray Gerstenhaber and C.T. Yang, Division Rings Containing a Real Closed Field, Duke Math. J. Volume 27, Number 4 (1960), 461-465. MR0114830

[4] Nathan Jacobson, Basic Algebra II, Second Edition, W.H. Freeman and Company, New York (1989).

[5] Nathan Jacobson, Lectures in Abstract Algebra Vol. 3, Theory of Fields and Galois Theory, Van Nostrand, Princeton, NJ (1964).

[6] Irving Kaplansky, Linear Algebra and Geometry, A Second Course, Oston, Allyn and Bacon (1969).

[7] Tsit-Yuen Lam, A First Course in Noncommutative Rings, Graduate Texts in Mathematics , Vol. 131, 2nd ed., 2001, XIX, 385 p.

[8] Tsit-Yuen Lam, Introduction to Quadratic Forms over Fields, Graduate Studies in Math. **67**, AMS, Providence, Rhode Island (2004).

[9] Serge Lang, Algebra, Reading, Mass. : Addison-Wesley, c1965.

[10] David W. Lewis, Quaternion Algebras and the Algebraic Legacy of Hamilton's Quaternions, Irish Math. Soc. Bulletin **57** (2006), 41-64.

[11] Kazimierz Szymiczek, Bilinear Algebra: An Introduction to the Algebraic Theory of Quadratic Forms, Algebra, Logic and Applications Series Volume 7, Gordon and Breach Science Publishers, 1997.