QUATERNION ALGEBRAS

KEITH CONRAD

1. Introduction

A complex number is a sum $a+bi$ with $a, b \in \mathbf{R}$ and $i^2 = -1$. Addition and multiplication are given by the rules

$$(a+bi)+(c+di) = (a+c)+(b+d)i, \quad (a+bi)(c+di) = (ac-bd)+(ad+bc)i. \tag{1.1}$$

This definition doesn't explain what i "is". In 1833, Hamilton [3, p. 81] proposed bypassing the mystery about the meaning of i by declaring $a+bi$ to be an ordered pair (a, b). That is, he defined $\mathbf{C}$ to be $\mathbf{R}^2$ with addition and multiplication rules inspired by (1.1):

$$(a,b)+(c,d) = (a+c, b+d), \quad (a,b)(c,d) = (ac-bd, ad+bc).$$

The additive identity is $(0,0)$, the multiplicative identity is $(1,0)$, and from addition and scalar multiplication of real vectors we have $(a,b) = (a,0)+(0,b) = a(1,0)+b(0,1)$, which looks like $a+bi$ if we define i to be $(0,1)$. Real numbers occur as the pairs $(a,0)$.

Hamilton asked himself if it was possible to multiply triples (a,b,c) in a nice way that extends multiplication of complex numbers (a,b) when they are thought of as triples $(a,b,0)$. In 1843 he discovered a way to multiply in *four* dimensions, not three, at the cost of abandoning commutativity of multiplication. His construction is called the quaternions.

After meeting the quaternions in Section 2, we will see in Section 3 how they can be generalized to a construction called a quaternion algebra. Sections 4 and 5 explore quaternion algebras over fields not of characteristic 2.

2. Hamilton's Quaternions

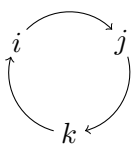
Definition 2.1. The *quaternions* are

$$\mathbf{H} = \{a+bi+cj+dk : a, b, c, d \in \mathbf{R}\},$$

where the following multiplication conditions are imposed:

- $i^2 = j^2 = k^2 = -1$,
- $ij = k$, $ji = -k$, $jk = i$, $kj = -i$, $ki = j$, $ki = j$, $ik = -j$,
- every $a \in \mathbf{R}$ commutes with i, j, and k.

To remember the rules for multiplying i, j, and k by each other, put them in alphabetical order around a circle as below. Products following this order get a plus sign, and products going against the order get a minus sign, *e.g.*, $jk = i$ and $ik = -j$.



The rules for multiplication among i, j, and k are enough, with the distributive law, to multiply all quaternions.

Example 2.2. $(i+j)(i-j) = i^2 - ij + ji - j^2 = -1 - k - k - (-1) = -2k$, while $i^2 - j^2 = -1 - (-1) = 0$.

Example 2.3. A quaternion with $a = 0$ is called a *pure quaternion*, and the square of a pure quaternion is a negative sum of three squares:

$$(bi + cj + dk)^2 = -b^2 - c^2 - d^2. \tag{2.1}$$

The multiplicative rules involving i, j, and k can be derived from $i^2 = j^2 = -1$ and $ij = k = -ji$ using associativity, *e.g.*,

$$k^2 = (ij)(ij) = (ij)(-ji) = i(-j^2)i = i(-(-1))i = i^2 = -1$$

and

$$jk = j(ij) = (ji)j = (-ij)j = -i(jj) = -i(-1) = i, \quad ik = i(ij) = (ii)j = -j.$$

While multiplication in $\mathbf{H}$ is typically noncommutative, multiplication in $\mathbf{H}$ by real numbers is commutative: $aq = qa$ when $a \in \mathbf{R}$ and $q \in \mathbf{H}$. This commuting property singles out $\mathbf{R}$ inside $\mathbf{H}$: the only quaternions that commute with all quaternions are real numbers (Exercise 2.2). In the terminology of ring theory, the set of elements of a ring that commute with every element of the ring is called the *center* of the ring, so the center of $\mathbf{H}$ is $\mathbf{R}$. The ring $\mathrm{M}_2(\mathbf{R})$ also has center $\mathbf{R}$: the matrices in $\mathrm{M}_2(\mathbf{R})$ that commute with all matrices in $\mathrm{M}_2(\mathbf{R})$ are the scalar diagonal matrices $\left(\begin{smallmatrix} a & 0 \\ 0 & a \end{smallmatrix}\right) = aI_2$, which is a natural copy of $\mathbf{R}$ in $\mathrm{M}_2(\mathbf{R})$.

For a quaternion $q = a + bi + cj + dk$, its *conjugate* $\overline{q}$ is defined to be

$$\overline{q} = a - bi - cj - dk.$$

This is analogous to complex conjugation on $\mathbf{C}$, where $\overline{a+bi} = a - bi$. Complex conjugation interacts well with addition and multiplication in $\mathbf{C}$:

$$\overline{z+w} = \overline{z} + \overline{w}, \quad \overline{zw} = \overline{z}\,\overline{w}.$$

For $z = a + bi$ in $\mathbf{C}$, $z\overline{z} = a^2 + b^2$. The absolute value $|a+bi|$ is defined to be $\sqrt{a^2+b^2}$, so $|z|^2 = z\overline{z}$. If $z \neq 0$ in $\mathbf{C}$ then $|z| > 0$ and $1/z = \overline{z}/|z|^2$.

Conjugation on $\mathbf{H}$ has properties similar to conjugation on $\mathbf{C}$: in $\mathbf{H}$,

$$\overline{q_1 + q_2} = \overline{q_1} + \overline{q_2}, \quad \overline{q_1 q_2} = \overline{q_2}\,\overline{q_1}, \quad \overline{\overline{q}} = q. \tag{2.2}$$

Note that conjugation switches the order of multiplication. The *norm* of q is

$$\mathrm{N}(q) = q\overline{q} = a^2 + b^2 + c^2 + d^2.$$

Check that q commutes with its conjugate: $\overline{q}q = q\overline{q}$. Since $\overline{q_1 q_2} = \overline{q_2}\,\overline{q_1}$, the norm is multiplicative:

$$\mathrm{N}(q_1 q_2) = q_1 q_2 \overline{q_1 q_2} = q_1 q_2 \overline{q_2}\,\overline{q_1} = q_1\,\mathrm{N}(q_2)\overline{q_1} = q_1\overline{q_1}\,\mathrm{N}(q_2) = \mathrm{N}(q_1)\,\mathrm{N}(q_2). \tag{2.3}$$

If $q \neq 0$ then $\mathrm{N}(q) > 0$, so $\overline{q}/\mathrm{N}(q)$ is an inverse for q on both the left and right:

$$q\frac{\overline{q}}{\mathrm{N}(q)} = \frac{\overline{q}}{\mathrm{N}(q)}q = \frac{\mathrm{N}(q)}{\mathrm{N}(q)} = 1.$$

Example 2.4. The quaternion $i + j$ has conjugate $-i - j$ and norm 2, so the inverse of $i + j$ is $\frac{1}{2}(-i - j)$.

A ring in which every nonzero element has a two-sided multiplicative inverse is called a *division ring*, so $\mathbf{H}$ is a division ring. We set $\mathbf{H}^\times = \mathbf{H} - \{0\}$, just like with fields. A commutative division ring is a field, and the center of a division ring is a field (Exercise 2.3). The quaternions were the first example of a noncommutative division ring, and the following theorem provides a conceptual role for them in algebra among all division rings.

Theorem 2.5 (Frobenius, 1878)**.** *Each division ring with center* $\mathbf{R}$ *that is finite-dimensional as a vector space over* $\mathbf{R}$ *is isomorphic to* $\mathbf{R}$ *or* $\mathbf{H}$*.*

Proof. See [4, pp. 219–220]. □

Theorem 2.5 does not include $\mathbf{C}$, even though it is a division ring and is finite-dimensional over $\mathbf{R}$, since the center of $\mathbf{C}$ is $\mathbf{C}$, not $\mathbf{R}$.

The term "conjugation" has two meanings: the operation $q \mapsto \overline{q}$ (generalizing complex conjugation) and the operation $\mathbf{x} \mapsto q\mathbf{x}q^{-1}$ for $q \neq 0$ (conjugation in the sense of group theory). The four-dimensional space $\mathbf{H}$ can be used to describe rotations in $\mathbf{R}^3$ by using conjugation in the second sense on the three-dimensional subspace of pure quaternions in $\mathbf{H}$, which is described in Exercise 2.8 (a), (c). See also [2, Chap. 7] or [6, Sect. 5]. In computer code, the composition of rotations when described in terms of multiplication of quaternions has some advantages over other approaches to rotations (no "gimbal lock" and less data to store: 4 coordinates of a quaternion vs. 9 components of a matrix). This makes quaternions a practical tool in computer graphics (search on the internet for "slerp").

The complex numbers are a 2-dimensional vector space over $\mathbf{R}$, so the set $\mathrm{End}_{\mathbf{R}}(\mathbf{C})$ of all $\mathbf{R}$-linear maps $\mathbf{C} \to \mathbf{C}$ is a noncommutative ring that is isomorphic to $\mathrm{M}_2(\mathbf{R})$ by using a basis of $\mathbf{C}$ over $\mathbf{R}$ to turn linear maps $\mathbf{C} \to \mathbf{C}$ into 2×2 real matrices. To each complex number $z = a + bi$, associate the $\mathbf{R}$-linear map $m_z \colon \mathbf{C} \to \mathbf{C}$, where $m_z(w) = zw$. Not only is each m_z an $\mathbf{R}$-linear map $\mathbf{C} \to \mathbf{C}$, but m_z is additive and multiplicative in z: $m_{z+z'} = m_z + m_{z'}$ and $m_{zz'} = m_z \circ m_{z'}$ as mappings $\mathbf{C} \to \mathbf{C}$. Therefore $z \mapsto m_z$ is a ring homomorphism $\mathbf{C} \to \mathrm{End}_{\mathbf{R}}(\mathbf{C})$ and it is injective since $m_z(1) = z$.

Using the $\mathbf{R}$-basis $\{1, i\}$ for $\mathbf{C}$, m_z has matrix representation $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ since $m_z(1) = a + bi$ (1st column) and $m_z(i) = zi = -b + ai$ (2nd column). Therefore $a + bi \mapsto \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ is an embedding of rings $\mathbf{C} \to \mathrm{M}_2(\mathbf{R})$. That means these 2×2 matrices add and multiply in the same way that complex numbers add and multiply in (1.1):

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} + \begin{pmatrix} c & -d \\ d & c \end{pmatrix} = \begin{pmatrix} a+c & -(b+d) \\ b+d & a+c \end{pmatrix},$$
$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} c & -d \\ d & c \end{pmatrix} = \begin{pmatrix} ac-bd & -(ad+bc) \\ ad+bc & ac-bd \end{pmatrix}.$$

Complex conjugation and the squared absolute value of a complex number z can be described in terms of matrix operations on $[m_z]$: $[m_{\overline{z}}] = [m_z]^\top$ and $|z|^2 = \det[m_z]$.

To express $\mathbf{H}$ in terms of 2×2 complex matrices, we want to convert each quaternion into a $\mathbf{C}$-linear map $\mathbf{H} \to \mathbf{H}$, since $\mathrm{M}_2(\mathbf{C})$ consists of $\mathbf{C}$-linear maps $\mathbf{C}^2 \to \mathbf{C}^2$ and $\mathbf{H}$ is 2-dimensional as a complex vector space, so $\mathbf{H} \cong \mathbf{C}^2$ once we pick a $\mathbf{C}$-basis of $\mathbf{H}$. This requires care because $\mathbf{C}$ does not commute with $\mathbf{H}$, *e.g.*, $ij = -ji$ and $ik = -ki$. We can view $\mathbf{H}$ as a left $\mathbf{C}$-vector space ($z \cdot q = zq$) or as a right $\mathbf{C}$-vector space ($z \cdot q = qz$), and the choice affects whether a particular mapping $\mathbf{H} \to \mathbf{H}$ is $\mathbf{C}$-linear.

Example 2.6. For each $q \in \mathbf{H}$ the mapping $\ell_q \colon \mathbf{H} \to \mathbf{H}$ where $\ell_q(\mathbf{x}) = q\mathbf{x}$ for all $\mathbf{x} \in \mathbf{H}$, is $\mathbf{C}$-linear when $\mathbf{H}$ is a right $\mathbf{C}$-vector space but not when $\mathbf{H}$ is a left $\mathbf{C}$-vector space: for

$z \in \mathbf{C}$, $q(\mathbf{x}z) = (q\mathbf{x})z$ always, but typically $q(z\mathbf{x}) \neq z(q\mathbf{x})$ (if $\mathbf{x} \neq 0$, $z \in \mathbf{C} - \mathbf{R}$, and $q \in \mathbf{H} - \mathbf{C}$).

As both a left and right vector space over $\mathbf{C}$, $\mathbf{H}$ has basis $\{1, j\}$: for $a, b, c, d \in \mathbf{R}$,

$$a + bi + cj + dk = (a + bi) + (c + di)j = (a + bi) + j(c - di).$$

Therefore each $q \in \mathbf{H}$ can be written uniquely as $z + wj$ for some $z, w \in \mathbf{C}$ or as $z + jw$ for some $z, w \in \mathbf{C}$. Passing between $\{1, j\}$ as a left $\mathbf{C}$-basis and a right $\mathbf{C}$-basis involves complex conjugation of the coefficient of j: since $ij = -ji$,

$$z + wj = z + j\overline{w}, \quad z + jw = z + \overline{w}j.$$

$\mathbf{H}$ as a right $\mathbf{C}$-vector space. For $q \in \mathbf{H}$, let $\ell_q \colon \mathbf{H} \to \mathbf{H}$ by $\ell_q(\mathbf{x}) = q\mathbf{x}$. Each ℓ_q is $\mathbf{C}$-linear, *e.g.*, $\ell_q(\mathbf{x}z) = q(\mathbf{x}z) = (q\mathbf{x})z = \ell_q(\mathbf{x})z$, and ℓ_q is additive and multiplicative in q (*e.g.*, $\ell_{q_1} \circ \ell_{q_2} = \ell_{q_1 q_2}$). We can recover q from ℓ_q since $\ell_q(1) = q$. Therefore $q \mapsto \ell_q$ is an embedding $\mathbf{H} \to \mathrm{End}_{\mathbf{C}}(\mathbf{H})$ that is additive and multiplicative. Using the basis $\{1, j\}$ for $\mathbf{H}$ as a right $\mathbf{C}$-vector space, if we write $q = z + jw$ for $z, w \in \mathbf{C}$ then $\ell_q(1) = q = z + jw$ and $\ell_q(j) = qj = zj + jwj = j\overline{z} + j(j\overline{w}) = -\overline{w} + j\overline{z}$. Therefore $\mathbf{H}$ embeds additively and multiplicatively into $\mathrm{M}_2(\mathbf{C})$ by

$$q = z + jw \mapsto [\ell_q] = \begin{pmatrix} z & -\overline{w} \\ w & \overline{z} \end{pmatrix}. \tag{2.4}$$

For example,

$$[\ell_i] = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad [\ell_j] = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad [\ell_k] = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}.$$

This embedding could be used to prove multiplication in $\mathbf{H}$ is associative since multiplication is associative in $\mathrm{M}_2(\mathbf{C})$.

$\mathbf{H}$ as a left $\mathbf{C}$-vector space. For $q \in \mathbf{H}$, let $r_q \colon \mathbf{H} \to \mathbf{H}$ by $r_q(\mathbf{x}) = \mathbf{x}\overline{q}$. This is $\mathbf{C}$-linear, *e.g.*, $r_q(z\mathbf{x}) = (z\mathbf{x})\overline{q} = z(\mathbf{x}\overline{q}) = zr_q(\mathbf{x})$. We use $\overline{q}$ in the definition of r_q to make r_q multiplicative in q: for all $\mathbf{x} \in \mathbf{H}$, $(r_{q_1} \circ r_{q_2})(\mathbf{x}) = \mathbf{x}\overline{q_2}\,\overline{q_1} = \mathbf{x}\overline{q_1 q_2} = r_{q_1 q_2}(\mathbf{x})$, so $r_{q_1} \circ r_{q_2} = r_{q_1 q_2}$. Since $r_q(1) = \overline{q}$, we can recover q from r_q, so $q \mapsto r_q$ is an embedding $\mathbf{H} \to \mathrm{End}_{\mathbf{C}}(\mathbf{H})$ that is additive and multiplicative. Using the basis $\{1, j\}$ for $\mathbf{H}$ as a left $\mathbf{C}$-vector space, if we write $q = z + wj$ for $z, w \in \mathbf{C}$ then

$$r_q(1) = \overline{q} = \overline{z} + \overline{wj} = \overline{z} + \overline{j}\,\overline{w} = \overline{z} - j\overline{w} = \overline{z} - wj$$

and

$$r_q(j) = j\overline{q} = j(\overline{z} - wj) = zj - \overline{w}jj = \overline{w} + zj.$$

Therefore $\mathbf{H}$ embeds additively and multiplicatively into $\mathrm{M}_2(\mathbf{C})$ by

$$q = z + wj \mapsto [r_q] = \begin{pmatrix} \overline{z} & \overline{w} \\ -w & z \end{pmatrix}. \tag{2.5}$$

For example,

$$[r_i] = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}, \quad [r_j] = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad [r_k] = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}.$$

While the two embeddings $\mathbf{H} \to \mathrm{M}_2(\mathbf{C})$ in (2.4) and (2.5) are different functions (they have different effects on i and on j), their images are the same: in both cases, the image is the set of complex matrices whose diagonal entries are complex conjugates and whose off-diagonal entries are negative complex conjugates.

Conjugation on $\mathbf{H}$ and the norm on $\mathbf{H}$ can be described in terms of matrix operations using either of the above 2×2 matrix representations of $\mathbf{H}$: by writing

$$[\ell_{\overline{q}}] = \overline{[\ell_q]}^\top \quad \text{and} \quad \mathrm{N}(q) = \det[\ell_q], \qquad [r_{\overline{q}}] = \overline{[r_q]}^\top \quad \text{and} \quad \mathrm{N}(q) = \det[r_q]. \tag{2.6}$$

For example, the first equation says the matrix representation of left multiplication by q on $\mathbf{H}$ as a right $\mathbf{C}$-vector space using the basis $\{1, j\}$ is the transpose of the complex conjugate of the matrix $[\ell_q]$.

Exercises.

1. Verify properties of quaternionic conjugation: $\overline{q_1 + q_2} = \overline{q}_1 + \overline{q}_2$, $\overline{q_1 q_2} = \overline{q}_2 \overline{q}_1$, $\overline{\overline{q}} = q$, $\overline{cq} = c\overline{q}$ for $c \in \mathbf{R}$, and $\overline{q} = q \Leftrightarrow q \in \mathbf{R}$.
2. Show the center of $\mathbf{H}$ is $\mathbf{R}$: $\{q \in \mathbf{H} : qq' = q'q \text{ for all } q' \in \mathbf{H}\} = \mathbf{R}$.
3. Show the center of a division ring is a field. (The main point is to show the inverse of a nonzero element in the center is also in the center.)
4. Verify the equations in (2.6).
5. Let $f\colon \mathbf{H} \to \mathrm{M}_2(\mathbf{C})$ by $f(z+wj) = \left(\begin{smallmatrix} z & -w \\ \overline{w} & \overline{z} \end{smallmatrix}\right)$ where $z, w \in \mathbf{C}$. Show (i) f is an injective ring homomorphism with the same image as (2.4) and (2.5), (ii) $\det([f_q]) = \mathrm{N}(q)$, and (iii) f is *not* $\mathbf{C}$-linear when $\mathbf{H}$ is viewed as either a left $\mathbf{C}$-vector space or as a right $\mathbf{C}$-vector space.
6. Verify that the image of $\mathbf{C}$ in $\mathrm{M}_2(\mathbf{R})$ under the embedding $a + bi \mapsto \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ is $\{A \in \mathrm{M}_2(\mathbf{R}) : \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)A = A\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)\}$ and the common image of $\mathbf{H}$ in $\mathrm{M}_2(\mathbf{C})$ from the embeddings (2.4) and (2.5) is $\{A \in \mathrm{M}_2(\mathbf{C}) : \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)A = \overline{A}\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)\}$, where $\overline{A}$ is the matrix with entries that are complex conjugates of the entries of A.
7. For $q \in \mathbf{H}^\times$, let $R_q\colon \mathbf{H} \to \mathbf{H}$ by $R_q(r) = qrq^{-1}$.
   a) Show R_q is a ring automorphism of $\mathbf{H}$.
   b) Show $R_{q_1} \circ R_{q_2} = R_{q_1 q_2}$. Does $R_{q_1+q_2} = R_{q_1} + R_{q_2}$?
   c) For $q, q' \in \mathbf{H}^\times$, show $R_q(r) = R_{q'}(r)$ for all $r \in \mathbf{H}$ if and only if $q' = cq$ for some $c \in \mathbf{R}^\times$.
8. Let $\mathbf{H}^0 = \mathbf{R}i + \mathbf{R}j + \mathbf{R}k$. These are the pure quaternions. Define $\mathrm{Tr}\colon \mathbf{H} \to \mathbf{R}$ by $\mathrm{Tr}(q) = q + \overline{q} = 2a$, where a is the real component of q. The number $\mathrm{Tr}(q)$ is called the trace of q. Then $\mathbf{H}^0 = \{q \in \mathbf{H} : \mathrm{Tr}(q) = 0\}$.
   a) Show $\mathrm{Tr}(qq') = \mathrm{Tr}(q'q)$ for all q and q' in $\mathbf{H}$. Use this to show $R_q(\mathbf{H}^0) = \mathbf{H}^0$ for all $q \in \mathbf{H}^\times$, where R_q is defined in the previous exercise.
   b) If $R_q(r) = R_{q'}(r)$ for all $r \in \mathbf{H}^0$, is $q' = cq$ for some $c \in \mathbf{R}^\times$?
   c) Show $\mathbf{H}^0 = \{q \in \mathbf{H} : q^2 \leq 0\}$, and use this to prove in another way that $R_q(\mathbf{H}^0) = \mathbf{H}^0$ for all $q \in \mathbf{H}^\times$.
   d) For $q \in \mathbf{H}$, show $q^2 = -1$ if and only if $q = bi + cj + dk$ with $b^2 + c^2 + d^2 = 1$. That is, the solutions to $q^2 + 1 = 0$ in $\mathbf{H}$ form a sphere of pure quaternions.
9. Identify $\mathbf{H}^0$ with $\mathbf{R}^3$ by $bi + cj + dk \leftrightarrow (b, c, d)$. If $q = bi + cj + dk$, write $\mathbf{q}$ for (b, c, d) as a vector in $\mathbf{R}^3$.
   a) Show multiplication of pure quaternions can be described in terms of the dot product and cross product on $\mathbf{R}^3$: $q_1, q_2 \in \mathbf{H}^0 \Longrightarrow q_1 q_2 = -(\mathbf{q}_1 \cdot \mathbf{q}_2) + \mathbf{q}_1 \times \mathbf{q}_2$, where the cross product $\mathbf{q}_1 \times \mathbf{q}_2$ is computed in $\mathbf{R}^3$ and then viewed as a pure quaternion. In particular, $\mathbf{q}_1$ and $\mathbf{q}_2$ are perpendicular in $\mathbf{R}^3$ if and only if q_1 and q_2 anti-commute (that is, $q_1 q_2 = -q_2 q_1$).

b) What are the constraints on the coordinates of $x_1i + x_2j + x_3k$ in order for it to anti-commute with $i + j$?

c) For $q_1, q_2, q_3 \in \mathbf{H}^0$, show

$$\mathbf{q}_1 \times (\mathbf{q}_2 \times \mathbf{q}_3) = \frac{1}{2}(q_1q_2q_3 - q_2q_3q_1).$$

3. Quaternion Algebras: Introduction

Let F be a field. Hamilton's quaternions $\mathbf{H}$ can be generalized to allow coefficients in F:

$$\mathbf{H}(F) = \{a + bi + cj + dk : a, b, c, d, \in F\}$$

where i, j, and k have the same multiplicative rules as in $\mathbf{H} = \mathbf{H}(\mathbf{R})$. Conjugation and the norm on $\mathbf{H}(F)$ are defined in the same way as in $\mathbf{H}$, and their properties in (2.2) and (2.3) continue to be valid. If F does not have characteristic 2, so $1 \neq -1$ in F, then the center of $\mathbf{H}(F)$ is F. If F has characteristic 2 then $\mathbf{H}(F)$ is commutative. *From now on, F is assumed to have characteristic not 2.*

Example 3.1. The ring $\mathbf{H}(\mathbf{Q})$ is a division ring since it is a subring of the division ring $\mathbf{H}(\mathbf{R})$ and the inverse of a nonzero element q of $\mathbf{H}(\mathbf{Q})$ in $\mathbf{H}(\mathbf{R})$ is $\overline{q}/\operatorname{N}(q)$, which is in $\mathbf{H}(\mathbf{Q})$.

Example 3.2. The ring $\mathbf{H}(\mathbf{F}_7)$ is not a division ring: $2^2+3^2+1^2 = 0$ in $\mathbf{F}_7$, so $(2i+3j+k)^2 = -2^2 - 3^2 - 1^2 = 0$ using (2.1) in $\mathbf{H}(\mathbf{F}_7)$. A quaternion that squares to 0 can't have a multiplicative inverse, so $2i+3j+k$ is a nonzero element of $\mathbf{H}(\mathbf{F}_7)$ without a multiplicative inverse in $\mathbf{H}(\mathbf{F}_7)$.

A broader generalization of $\mathbf{H}$ than $\mathbf{H}(F)$ was introduced by Dickson in 1906 [1].

Definition 3.3. A *quaternion algebra* over F is a ring that is a 4-dimensional vector space over F with a basis $1, u, v, w$ with the following multiplicative relations: $u^2 \in F^\times$, $v^2 \in F^\times$, $w = uv = -vu$, and every $c \in F$ commutes with u and v. When $a = u^2$ and $b = v^2$, this ring is denoted $(a, b)_F$.

More explicitly, for a and b in $F^\times$ the ring $(a, b)_F$ looks as follows. As a vector space over F it can be written as

$$(a, b)_F = F + Fu + Fv + Fw,$$

and the multiplicative relations among u, v, w, and elements of F are

- $u^2 = a$ and $v^2 = b$,
- $w := uv = -vu$,
- every $c \in F$ commutes with u and v.

Example 3.4. In this notation $\mathbf{H}(F) = (-1, -1)_F$, so this is a quaternion algebra where $a = b = -1$.

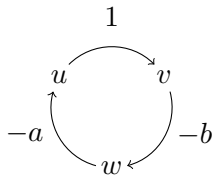
In Table 1 are products among u, v, and w, where each entry is the product of the row label times the column label (in that order: multiplication is noncommutative). For example, $vw = v(uv) = (vu)v = -uv^2 = -ub = -bu$. Note u, v, and w each square to a nonzero element of F and they anti-commute: $uv = -vu$, $uw = -wu$, and $vw = -wv$.

We can make a circular diagram for products of u, v, and w that is similar to the one for i, j, and k. In the picture below we write u, v, and w in alphabetical order, with 1 on the arc from u to v, $-b$ on the arc from v to w, and $-a$ on the arc from w to u. The product of two of u, v, and w is the third one times the number on the arc between the

	u	v	w
u	a	w	av
v	$-w$	b	$-bu$
w	$-av$	bu	$-ab$

TABLE 1. Products of u, v, and w in $(a,b)_F$.

two factors, with an additional sign if the multiplication is going against the arrows, *e.g.*, $vw = (-b)u = -bu$ and $uw = -(-a)v = av$.



Example 3.5. We have $(2,3)_{\mathbf{Q}} = \mathbf{Q} + \mathbf{Q}u + \mathbf{Q}v + \mathbf{Q}w$ where $u^2 = 2$, $v^2 = 3$, and $w = uv = -vu$ with $w^2 = -6$.

The multiplicative rules on u, v, and w are consistent with the axioms of a ring because we can realize the operations in $(a,b)_F$ as addition and multiplication of certain 2×2 matrices (Exercise 3.8). Since F doesn't have characteristic 2, $(a,b)_F$ is noncommutative because u and v don't commute. The center of $(a,b)_F$ is F (Exercise 3.2).

For $q = x_0 + x_1u + x_2v + x_3w$, define the *conjugate* and *norm* of q to be

$$\overline{q} = x_0 - x_1u - x_2v - x_3w, \quad \mathrm{N}(q) = q\overline{q} = x_0^2 - ax_1^2 - bx_2^2 + abx_3^2. \tag{3.1}$$

As with $\mathbf{H}$, $\overline{q}q = q\overline{q}$ in $(a,b)_F$ and the calculations in (2.2) and (2.3) remain valid, so the norm is a multiplicative function $(a,b)_F \to F$.

Example 3.6. For $q = x_0 + x_1u + x_2v + x_3w$ in $(2,3)_{\mathbf{Q}}$,

$$\mathrm{N}(q) = x_0^2 - 2x_1^2 - 3x_2^2 + 6x_3^2.$$

Example 3.7. Generalizing Example 2.3, an element of $(a,b)_F$ with $x_0 = 0$ is called a *pure quaternion*. Its square is a scalar: for $x, y, z \in F$,

$$(xu + yv + zw)^2 = ax^2 + by^2 - abz^2 \in F. \tag{3.2}$$

This property essentially characterizes the pure quaternions (Exercise 3.6 ii).

Theorem 3.8. *An element q of $(a,b)_F$ has a two-sided multiplicative inverse in $(a,b)_F$ if and only if $\mathrm{N}(q) \neq 0$.*

Proof. Suppose $qq' = 1$ for some q'. Then $\mathrm{N}(q)\,\mathrm{N}(q') = \mathrm{N}(1) = 1$ in F, so $\mathrm{N}(q) \in F^\times$.

Conversely, suppose $\mathrm{N}(q) \in F^\times$. Since $\mathrm{N}(q)$ commutes with all elements of $(a,b)_F$, the equation $\mathrm{N}(q) = q\overline{q} = \overline{q}q$ can be rewritten as

$$q \cdot \frac{1}{\mathrm{N}(q)}\overline{q} = \frac{1}{\mathrm{N}(q)}\overline{q} \cdot q = 1,$$

so $\overline{q}/\mathrm{N}(q)$ is a 2-sided inverse of q. □

Here are quaternion algebras over $\mathbf{Q}$ besides $\mathbf{H}(\mathbf{Q})$ that are division rings.

Theorem 3.9. *Let a be an integer and p be an odd prime such that $a \not\equiv \square \bmod p$.*[1] *Then $(a,p)_{\mathbf{Q}}$ is a division ring.*

Proof. By Theorem 3.8, to show $(a,p)_{\mathbf{Q}}$ is a division ring it suffices to show every nonzero element of $(a,p)_{\mathbf{Q}}$ has a nonzero norm. We will prove the contrapositive: an element of $(a,p)_{\mathbf{Q}}$ with norm 0 must be 0.

Let $q = x_0 + x_1u + x_2v + x_3w$ in $(a,p)_{\mathbf{Q}}$. Using the formula for $\mathrm{N}(q)$ in (3.1),

$$\mathrm{N}(q) = x_0^2 - ax_1^2 - px_2^2 + apx_3^2,$$

so we can't show $\mathrm{N}(q) = 0 \Rightarrow q = 0$ using positivity as we can for $\mathbf{H}$: $\mathrm{N}(q)$ can be either positive or negative. To show $\mathrm{N}(q) = 0 \Rightarrow q = 0$, the property $a \not\equiv \square \bmod p$ will be crucial.

If $\mathrm{N}(q) = 0$ then

$$x_0^2 - ax_1^2 - px_2^2 + apx_3^2 = 0 \Longrightarrow x_0^2 - ax_1^2 = p(x_2^2 - ax_3^2). \tag{3.3}$$

Multiplying through the last equation by a common denominator of x_0, x_1, x_2, and x_3, we can assume the x_i's are all in $\mathbf{Z}$. Then if we reduce mod p,

$$x_0^2 - ax_1^2 \equiv 0 \bmod p \Longrightarrow x_0^2 \equiv ax_1^2 \bmod p.$$

If $x_1 \not\equiv 0 \bmod p$ then we can solve for $a \bmod p$ in the congruence to see that $a \equiv \square \bmod p$, which is false. Therefore $x_1 \equiv 0 \bmod p$, so $x_0^2 \equiv 0 \bmod p$, and thus $x_0 \equiv 0 \bmod p$. Then in $x_0^2 - ax_1^2 = p(x_2^2 - ax_3^2)$ the left side is divisible by p^2, so $x_2^2 - ax_3^2 \equiv 0 \bmod p$, and an argument similar to the one above shows x_2 and x_3 are divisible by p.

Since every x_i is divisible by p, write $x_i = px_i'$ with $x_i' \in \mathbf{Z}$. Then

$$x_0^2 - ax_1^2 = p(x_2^2 - ax_3^2) \Longrightarrow p^2(x_0'^2 - ax_1'^2) = p(p^2)(x_2'^2 - ax_3'^2) \Longrightarrow x_0'^2 - ax_1'^2 = p(x_2'^2 - ax_3'^2).$$

This last equation is the same as the right side of (3.3), with x_i' in place of x_i. Then as before, each x_i' is divisible by p, so each x_i is divisible by p^2. Repeating this argument shows each x_i is divisible by arbitrarily high powers of p, so each x_i must be 0. □

Example 3.10. The rings $(2,3)_{\mathbf{Q}}$ and $(2,5)_{\mathbf{Q}}$ are division rings since 2 mod 3 and 2 mod 5 are not squares.

Example 3.11. For prime p with $p \equiv 3 \bmod 4$, $(-1,p)_{\mathbf{Q}}$ is a division ring since $-1 \not\equiv \square \bmod p$. We will look at $(-1,p)_{\mathbf{Q}}$ for $p \not\equiv 3 \bmod 4$ in Example 4.19.

Remark 3.12. The converse of Theorem 3.9 is false: $(a,p)_{\mathbf{Q}}$ can be a division ring when $a \equiv \square \bmod p$. For example, $3 \equiv \square \bmod 11$ and $(3,11)_{\mathbf{Q}}$ is a division ring (Example 4.2).

Quaternion algebras are related to hyperbolic manifolds [7], number theory [8, Chap. 5], [9, §III.9], [10], and quadratic forms [5, Chap. III].

Exercises.

1. Verify the multiplication table for u, v, w in Table 1.
2. Show the center of $(a,b)_F$ is F.
3. Show the set of elements of $(a,b)_F$ that anti-commute with u is $Fv + Fw$, and the elements of $(a,b)_F$ that anti-commute with u and square to b are those $xv + yw$ $(x, y \in F)$ such that $x^2 - ay^2 = 1$.

[1] Here and later, $\square$ means a square: something of the form x^2. Here it means a is *not* a square mod p.

4. (Conjugation and norm)
   a) Check properties of conjugation on $(a,b)_F$: $\overline{q_1+q_2} = \overline{q}_1 + \overline{q}_2$, $\overline{q_1q_2} = \overline{q}_2\overline{q}_1$, $\overline{\overline{q}} = q$, $\overline{cq} = c\overline{q}$ for $c \in F$, and $\overline{q} = q \Leftrightarrow q \in F$.
   b) For $q = x_0 + x_1u + x_2v + x_3w$, show $q\overline{q} = \overline{q}q = x_0^2 - ax_1^2 - bx_2^2 + abx_3^2$.
5. For $a \in \mathbf{Z}$, show that if $a \equiv 3$ or $5 \bmod 8$ then $(a,2)_{\mathbf{Q}}$ is a division ring. This should be considered an analogue of Theorem 3.9 when $p = 2$.
6. Let $(a,b)_F^0 = Fu + Fv + Fw$ be the pure quaternions in $(a,b)_F$.
   (i) If r is pure and q is invertible in $(a,b)_F$, show qrq^{-1} is pure. (Hint: Set $\mathrm{Tr}(q) = q + \overline{q}$, show Tr has properties similar to the trace on $\mathbf{H}$, and show $(a,b)_F^0 = \{q \in (a,b)_F : \mathrm{Tr}(q) = 0\}$.)
   (ii) For $q \in (a,b)_F$, show $q^2 \in F \iff q \in F$ or q is pure. Therefore the pure quaternions in $(a,b)_F$ are precisely the q satisfying $q^2 \in F$ with $q \notin F$, along with 0. (Hint: write $q = x_0 + q_0$ where q_0 is pure. Use the right side to compute q^2, noting x_0 and q_0 commute. By (3.2), $q_0^2 \in F$.)
7. Suppose $a, b \in \mathbf{R}^\times$ with $a > 0$. Show $(a,b)_{\mathbf{Q}}$ becomes a subring of $\mathrm{M}_2(\mathbf{R})$ by mapping
$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad u \mapsto \begin{pmatrix} \sqrt{a} & 0 \\ 0 & -\sqrt{a} \end{pmatrix}, \quad v \mapsto \begin{pmatrix} 0 & -1 \\ -b & 0 \end{pmatrix}, \quad w \mapsto \begin{pmatrix} 0 & -\sqrt{a} \\ \sqrt{a}b & 0 \end{pmatrix}$$
for the basis of $(a,b)_{\mathbf{Q}}$ and extending this to all of $(a,b)_{\mathbf{Q}}$ by $\mathbf{Q}$-linearity.
8. Let's generalize the embedding of $\mathbf{H}$ into $\mathrm{M}_2(\mathbf{C})$ in (2.4) to an embedding of $(a,b)_F$ into a 2×2 matrix ring.
   For $a \in F^\times$, the ring $F[t]/(t^2 - a)$ is a field if $a \neq \square$ in F, while $F[t]/(t^2 - a) \cong F \times F$ if $a = \square$ in F. Verify that the map $(a,b)_F \to \mathrm{M}_2(F[t]/(t^2-a))$ given by
$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad u \mapsto \begin{pmatrix} t & 0 \\ 0 & -t \end{pmatrix}, \quad v \mapsto \begin{pmatrix} 0 & -1 \\ -b & 0 \end{pmatrix}, \quad w \mapsto \begin{pmatrix} 0 & -t \\ bt & 0 \end{pmatrix},$$
and extended to all of $(a,b)_F$ by F-linearity, is an injective ring homomorphism.

4. Isomorphisms Between Quaternion Algebras

An *isomorphism* between two quaternion algebras A and A' over a field F is a ring isomorphism $f\colon A \to A'$ that fixes the elements of F (that is, $f(c) = c$ for all $c \in F$). To show two quaternion algebras are isomorphic we will take a low-brow approach by working with well-chosen bases of them.

Definition 4.1. A basis of $(a,b)_F$ having the form $\{1, e_1, e_2, e_1e_2\}$ where $e_1^2 \in F^\times$, $e_2^2 \in F^\times$, and $e_1e_2 = -e_2e_1$ is called a *quaternionic basis* of $(a,b)_F$.

For instance, the defining basis $\{1,u,v,w\}$ of $(a,b)_F$ is a quaternionic basis. In a quaternionic basis $(e_1e_2)^2 = -e_1^2e_2^2$ and the three elements e_1, e_2, e_1e_2 anti-commute.

There are quaternionic bases of $(a,b)_F$ other than $\{1,u,v,uv\}$, and different choices of a quaternionic basis reveal isomorphisms between different quaternion algebras on account of the multiplicative relations among the basis elements:

(1) $\{1,v,u,vu\}$ is a quaternionic basis of $(a,b)_F$, so $(a,b)_F \cong (b,a)_F$,
(2) $\{1,u,w,uw\}$ is a quaternionic basis of $(a,b)_F$, so $(a,b)_F \cong (a,-ab)_F$,
(3) $\{1,v,w,vw\}$ is a quaternionic basis of $(a,b)_F$, so $(a,b)_F \cong (b,-ab)_F$,
(4) $\{1,cu,dv,(cu)(dv)\}$ is a quaternionic basis of $(a,b)_F$ for all $c,d \in F^\times$, so $(a,b)_F \cong (ac^2,bd^2)_F$ for all nonzero c and d in F.

Example 4.2. The quaternion algebra $(3,11)_{\mathbf{Q}}$ is a division ring since $(3,11)_{\mathbf{Q}} \cong (11,3)_{\mathbf{Q}}$ and $11 \not\equiv \square \bmod 3$. Therefore we can use Theorem 3.9 with $a = 11$ and $p = 3$.

Using the second quaternionic basis with $b = 1$, $(a,-a) \cong (a,1)_F$, and with $b = -1$ we get $(a,a)_F \cong (a,-1)_F$. Using the fourth quaternionic basis, up to isomorphism $(a,b)_F$ only depends on a and b up to multiplication by nonzero squares in $F^\times$. For instance, $(a,c^2)_F \cong (a,1)_F$ and $(c^2,b)_F \cong (1,b)_F \cong (b,1)_F$. The quaternion algebra $(a,1)_F$ turns out to be a familiar ring.

Theorem 4.3. *For all $a \in F^\times$, $(a,1)_F \cong \mathrm{M}_2(F)$.*

This shows the ring $\mathrm{M}_2(F)$ is a quaternion algebra over F and that

$$(a,c^2)_F \cong (a,-a)_F \cong \mathrm{M}_2(F). \tag{4.1}$$

Proof. Send the basis $1, u, v, w$ of $(a,1)_F$ to $\mathrm{M}_2(F)$ as follows:

$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad u \mapsto \begin{pmatrix} 0 & 1 \\ a & 0 \end{pmatrix}, \quad v \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad w \mapsto \begin{pmatrix} 0 & -1 \\ a & 0 \end{pmatrix}.$$

Since $1 \neq -1$ in F, 1 and v are not sent to the same matrix. Extend this mapping by F-linearity to a function $(a,b)_F \to \mathrm{M}_2(F)$:

$$x_0 + x_1 u + x_2 v + x_3 w \mapsto \begin{pmatrix} x_0 + x_2 & x_1 - x_3 \\ a(x_1 + x_3) & x_0 - x_2 \end{pmatrix}. \tag{4.2}$$

The image of $1, u, v, w$ in $\mathrm{M}_2(F)$ is a linearly independent set, so by a dimension count this F-linear mapping $(a,b)_F \to \mathrm{M}_2(F)$ is a bijection. It fixes F, in the sense that $c \in F$ in $(a,b)_F$ goes to cI_2 in $\mathrm{M}_2(F)$. It is left to the reader to check (4.2) is multiplicative (Exercise 4.1). □

Definition 4.4. We call $\mathrm{M}_2(F)$, or a quaternion algebra isomorphic to $\mathrm{M}_2(F)$, a *trivial* or *split* quaternion algebra over F. If $(a,b)_F \ncong \mathrm{M}_2(F)$ we say $(a,b)_F$ is a *non-split* quaternion algebra.

Example 4.5. Let $F = \mathbf{R}$. Then

$$(a,b)_{\mathbf{R}} \cong \begin{cases} \mathbf{H}, & \text{if } a < 0 \text{ and } b < 0, \\ \mathrm{M}_2(\mathbf{R}), & \text{if } a > 0 \text{ or } b > 0. \end{cases}$$

Example 4.6. Let $F = \mathbf{C}$. All elements of $\mathbf{C}^\times$ are squares in $\mathbf{C}$, so $(a,b)_{\mathbf{C}} \cong \mathrm{M}_2(\mathbf{C})$ for all a and b in $\mathbf{C}^\times$: all quaternion algebras over $\mathbf{C}$ are split.

These examples tell us that, up to isomorphism, there are two quaternion algebras over $\mathbf{R}$ and one quaternion algebra over $\mathbf{C}$. Over $\mathbf{Q}$ the situation is completely different: there are *infinitely many non-isomorphic quaternion algebras* over $\mathbf{Q}$. We'll see this in Section 5.

Example 4.7. If p is prime and $p \equiv 1 \bmod 4$ then $\mathbf{H}(\mathbf{F}_p) \cong \mathrm{M}_2(\mathbf{F}_p)$ since -1 is a square in $\mathbf{F}_p$. We'll see in Corollary 4.24 that every quaternion algebra over $\mathbf{F}_p$ is isomorphic to $\mathrm{M}_2(\mathbf{F}_p)$.

Example 4.8. The quaternion algebras $(2,3)_{\mathbf{Q}}$ and $(2,5)_{\mathbf{Q}}$ are both division rings (Example 3.10), but the quaternion algebras $(2,3)_{\mathbf{R}}$ and $(2,5)_{\mathbf{R}}$ are not division rings: both are isomorphic to $\mathrm{M}_2(\mathbf{R})$.

Definition 4.9. For a and b in $\mathbf{Q}^\times$ we say $(a,b)_{\mathbf{Q}}$ *splits over* $\mathbf{R}$ if $(a,b)_{\mathbf{R}} \cong \mathrm{M}_2(\mathbf{R})$ and we say $(a,b)_{\mathbf{Q}}$ is *non-split over* $\mathbf{R}$ if $(a,b)_{\mathbf{R}} \ncong \mathrm{M}_2(\mathbf{R})$ (that is, $(a,b)_{\mathbf{R}} \cong \mathbf{H}$).

For example, $(2,3)_{\mathbf{Q}}$ and $(2,5)_{\mathbf{Q}}$ both split over $\mathbf{R}$, while $\mathbf{H}(\mathbf{Q}) = (-1,-1)_{\mathbf{Q}}$ is non-split over $\mathbf{R}$. More generally, for a field extension $F \subset K$ and $a, b \in F^\times$, we say $(a,b)_F$ *splits over* K when $(a,b)_K \cong \mathrm{M}_2(K)$.

Since $(a,b)_{\mathbf{Q}}$ splits over $\mathbf{R}$ when a or b is positive, while $(a,b)_{\mathbf{Q}}$ is non-split over $\mathbf{R}$ when a and b are both negative, the formula for $\mathrm{N}(q)$ in (3.1) shows the norm on $(a,b)_{\mathbf{Q}}$ has positive and negative values when $(a,b)_{\mathbf{Q}}$ splits over $\mathbf{R}$, while the norm on $(a,b)_{\mathbf{Q}}$ has only positive values (and 0) when $(a,b)_{\mathbf{Q}}$ is non-split over $\mathbf{R}$. It is a hard theorem that the norm mapping $\mathrm{N}\colon (a,b)_{\mathbf{Q}} \to \mathbf{Q}$ is surjective when $(a,b)_{\mathbf{Q}}$ splits over $\mathbf{R}$ and has image $\mathbf{Q}_{\geq 0}$ if $(a,b)_{\mathbf{Q}}$ is non-split over $\mathbf{R}$.

Example 4.10. Since $(2,3)_{\mathbf{Q}}$ is split over $\mathbf{R}$, the equation $x_0^2 - 2x_1^2 - 3x_2^2 + 6x_3^2 = r$ has a rational solution for every $r \in \mathbf{Q}$.

There are analogies between quadratic fields $\mathbf{Q}(\sqrt{d})$ and quaternion algebras $(a,b)_{\mathbf{Q}}$. Real quadratic fields (when $x^2 - d$ splits into linear factors over $\mathbf{R}$) are analogous to the $(a,b)_{\mathbf{Q}}$ that split over $\mathbf{R}$ and imaginary quadratic fields (when $x^2 - d$ does not split into linear factors over $\mathbf{R}$) are analogous to the $(a,b)_{\mathbf{Q}}$ that are non-split over $\mathbf{R}$. See Table 2.

Quadratic Field $\mathbf{Q}(\sqrt{d})$	Quaternion Algebra $(a,b)_{\mathbf{Q}}$
Real Quadratic ($d > 0$)	Split over $\mathbf{R}$
Imaginary Quadratic ($d < 0$)	Non-split over $\mathbf{R}$

TABLE 2. Analogous Quadratic Fields and Quaternion Algebras over $\mathbf{Q}$.

Here are two examples of the analogies.

(1) Sign of norm values: There are norm functions $\mathrm{N}\colon \mathbf{Q}(\sqrt{d}) \to \mathbf{Q}$ and $\mathrm{N}\colon (a,b)_{\mathbf{Q}} \to \mathbf{Q}$, where $\mathrm{N}(x + y\sqrt{d}) = x^2 - dy^2$ for quadratic fields. If $d > 0$ the norm on $\mathbf{Q}(\sqrt{d})$ has both positive and negative values, while if $d < 0$ the norm on $\mathbf{Q}(\sqrt{d})$ has only positive values (and 0). If $(a,b)_{\mathbf{Q}}$ splits over $\mathbf{R}$ then the norm on $(a,b)_{\mathbf{Q}}$ has both positive and negative values, while if $(a,b)_{\mathbf{Q}}$ is non-split over $\mathbf{R}$ then the norm on $(a,b)_{\mathbf{Q}}$ has only positive values (and 0).

(2) Integral units: The ring $\mathbf{Z}[\sqrt{d}]$ is analogous to the ring $(a,b)_{\mathbf{Z}} = \mathbf{Z} + \mathbf{Z}u + \mathbf{Z}v + \mathbf{Z}w$. Units in $\mathbf{Z}[\sqrt{d}]$ are those $x + y\sqrt{d}$ with norm ± 1, and units (invertible elements) of $(a,b)_{\mathbf{Z}}$ are the elements with norm ± 1. If $d > 0$ there are infinitely many units in $\mathbf{Z}[\sqrt{d}]$ (theory of Pell's equation), but if $d < 0$ the unit group of $\mathbf{Z}[\sqrt{d}]$ is finite. If $(a,b)_{\mathbf{Q}}$ splits over $\mathbf{R}$ then $(a,b)_{\mathbf{Z}}$ has infinitely many units, while if $(a,b)_{\mathbf{Q}}$ is non-split over $\mathbf{R}$ then $(a,b)_{\mathbf{Z}}$ has finitely many units.

A more subtle analogy is that when $d > 0$ and $(a,b)_{\mathbf{Q}}$ is split over $\mathbf{R}$, the infinitely many units in $\mathbf{Z}[\sqrt{d}]$ and $(a,b)_{\mathbf{Z}}$ both form finitely-generated groups.

Example 4.11. The quaternion algebra $(2,3)_{\mathbf{Q}}$ is split over $\mathbf{R}$, so infinitely many elements of $(2,3)_{\mathbf{Z}}$ have norm ± 1. Some examples are $1 + u$ (similar to $1 + \sqrt{2}$ in $\mathbf{Z}[\sqrt{2}]$), $2 + v$ (similar to $2 + \sqrt{3}$ in $\mathbf{Z}[\sqrt{3}]$), and $(1+u)^2(2+v) = 6 + 4u + 3v + 2u$.

The easiest way to know $(a,b)_F \cong \mathrm{M}_2(F)$ is when $b = \square$ in $F^\times$ (see (4.1)). There is a weaker condition on b than being a square that still implies $(a,b)_F \cong \mathrm{M}_2(F)$, and to describe this condition we need to know something about numbers of the form $x^2 - ay^2$.

Lemma 4.12. *For $a \in F^\times$, the set of nonzero $x^2 - ay^2$ with $x, y \in F$ is a subgroup of $F^\times$.*

Proof. The number 1 has this form ($x = 1, y = 0$). Number of this form are closed under multiplication since

$$(x_1^2 - ay_1^2)(x_2^2 - ay_2^2) = (x_1x_2 + ay_1y_2)^2 - a(x_1y_2 + x_2y_1)^2.$$

Nonzero numbers of this form are closed under inversion using the trivial identity $1/t = t/t^2$:

$$\frac{1}{x^2 - ay^2} = \frac{x^2 - ay^2}{(x^2 - ay^2)^2} = \left(\frac{x}{x^2 - ay^2}\right)^2 - a\left(\frac{y}{x^2 - ay^2}\right)^2.$$

□

Definition 4.13. For $a \in F^\times$, let $N_a = N_a(F)$ be the set of all nonzero $x^2 - ay^2$ where $x, y \in F$.

By Lemma 4.12 N_a is a subgroup of $F^\times$, and $(F^\times)^2 \subset N_a$ using $y = 0$.

Theorem 4.14. *If a is a square in F then $N_a = F^\times$.*

Proof. Write $a = c^2$ for $c \in F^\times$. Then $x^2 - ay^2 = x^2 - c^2y^2 = x^2 - (cy)^2 = (x - cy)(x + cy)$. The change of variables $x' = x - cy$ and $y' = x + cy$ is invertible[2] ($x = (x' + y')/2$ and $y = (y' - x')/(2c)$), so $N_a = \{x'y' : x', y' \in F^\times\}$, which assumes all values in $F^\times$ by choosing $y' = 1$. □

Remark 4.15. The converse of Theorem 4.14 is generally false: N_a could be $F^\times$ without a being a square in F. For instance, if $F = \mathbf{F}_p$ for odd primes p then we'll see in Corollary 4.24 that $N_a = \mathbf{F}_p^\times$ for all $a \in \mathbf{F}_p^\times$, but only half the elements of $\mathbf{F}_p^\times$ are squares. There are important cases where the converse of Theorem 4.14 is true, such as $F = \mathbf{Q}$ and $F = \mathbf{R}$ (and $F = \mathbf{Q}_p$ for a prime p, if you know what $\mathbf{Q}_p$ is).

We call N_a the *norm subgroup* of $F^\times$ associated to a, since $x^2 - ay^2 = (x + y\sqrt{a})(x - y\sqrt{a})$ is usually called a norm.

Theorem 4.16. *If $b \in N_a$ then $(a, b)_F \cong \mathrm{M}_2(F)$.*

As a special case this includes $(a, b)_F \cong \mathrm{M}_2(F)$ if $b = \square$ in F.

Proof. Write $b = x_0^2 - ay_0^2$ with x_0 and y_0 in F. The ring $(a, b)_F$ has a quaternionic basis (Definition 4.1):

$$1, \quad u, \quad x_0v + y_0w, \quad u(x_0v + y_0w).$$

The fourth element is $x_0w + y_0av$ by the formulas for uv and uw. Why is this a basis? It is linearly independent since the 4th term is $ay_0v + x_0w$ and the change of basis matrix from v, w to $x_0v + y_0w, ay_0v + x_0w$ has determinant $\left|\begin{smallmatrix} x_0 & y_0 \\ ay_0 & x_0 \end{smallmatrix}\right| = b \neq 0$ (Exercise 4.10). Therefore the above set of four elements of $(a, b)_F$ is linearly independent over F, so it is a basis of $(a, b)_F$. Why is this basis quaternionic? We have $(x_0v + y_0w)^2 = bx_0^2 - aby_0^2 = b^2$, and u and $x_0v + y_0w$ anti-commute. Therefore $b \in N_a \Rightarrow (a, b)_F \cong (a, b^2)_F \cong (a, 1)_F \cong \mathrm{M}_2(F)$. □

Example 4.17. We have $(a, 1 - a)_F \cong \mathrm{M}_2(F)$ if $a \neq 0, 1$ since $1 - a = x^2 - ay^2$ with $x = y = 1$.

Example 4.18. The quaternion algebra $(3, 11)_{\mathbf{Q}}$ is a division ring since $11 \not\equiv \square \mod 3$ (Example 4.2), while $(3, -11)_{\mathbf{Q}} \cong \mathrm{M}_2(\mathbf{Q})$ since $-11 = x^2 - 3y^2$ with $x = 1$ and $y = 2$.

Example 4.19. When p is a prime and $p = 2$ or $p \equiv 1 \mod 4$, Fermat's two-square theorem says p is a sum of two squares in $\mathbf{Z}$. Therefore $p \in N_{-1}(\mathbf{Q})$, so $(-1, p)_{\mathbf{Q}} \cong \mathrm{M}_2(\mathbf{Q})$.

[2]Here it is crucial that F does not have characteristic 2.

Theorem 4.20. *A quaternion algebra $(a,b)_F$ that is not a division ring is isomorphic to $\mathrm{M}_2(F)$.*

Proof. Since we already know that $(c^2,b)_F \cong \mathrm{M}_2(F)$, we can assume a is not a square in F.

By Theorem 3.8, if $(a,b)_F$ is not a division ring it contains a nonzero element q with $\mathrm{N}(q) = 0$. Let $q = x_0 + x_1u + x_2v + x_3w$ with its coefficients not all equal to 0. Then

$$\mathrm{N}(q) = 0 \Longrightarrow x_0^2 - ax_1^2 - bx_2^2 + abx_3^2 = 0 \Longrightarrow x_0^2 - ax_1^2 = b(x_2^2 - ax_3^2).$$

Since a is not a square in F, we must have $x_2^2 - ax_3^2 \neq 0$ by contradiction: if $x_2^2 - ax_3^2 = 0$ then $x_3 = 0$ (if $x_3 \neq 0$ we could solve for a to see it is a square in F), so also $x_2 = 0$, and that implies $x_0^2 - ax_1^2 = 0$, so also $x_1 = 0$ and $x_0 = 0$, but then $q = 0$.

Solving for b,

$$b = \frac{x_0^2 - ax_2^2}{x_2^2 - ax_3^2} \in N_a,$$

so $(a,b)_F \cong \mathrm{M}_2(F)$ by Theorem 4.16. □

By this theorem, all $(a,b)_F$ that are not isomorphic to $\mathrm{M}_2(F)$ are division rings, so

Non-split quaternion algebras = quaternion algebras that are division rings.

Theorem 4.21. *For a and b in $F^\times$, $(a,b)_F \cong \mathrm{M}_2(F)$ if and only if $b \in N_a$.*

Proof. If $b \in N_a$ then $(a,b)_F \cong \mathrm{M}_2(F)$ by Theorem 4.16. Conversely, suppose $(a,b)_F \cong \mathrm{M}_2(F)$. To show $b \in N_a$, we can assume a is not a square in $F^\times$, since if a were a square then $N_a = F^\times$ by Theorem 4.14, so obviously $b \in N_a$. When $(a,b)_F$ is not a division ring and a is not a square, the proof of Theorem 4.20 shows $b \in N_a$. □

We have seen several sufficient conditions that imply $(a,b)_F \cong \mathrm{M}_2(F)$:

- $b = c^2$: (4.1).
- $b = -a$: (4.1).
- $b = 1 - a$: Example 4.17.
- $b = x^2 - ay^2$ for some $x, y \in F$: Theorem 4.21.

The last condition includes the previous ones as special cases, and Theorem 4.21 tells us the last condition is not only sufficient but necessary as well.

Remark 4.22. We consider $\mathrm{M}_2(F)$ to be the "trivial" quaternion algebra, so the following quaternion algebras $(a,b)_F$ are considered trivial: $(a,c^2)_F$, $(a,-a)_F$, $(a,1-a)_F$, and more generally $(a,x^2-ay^2)_F$. In other areas of math there are objects depending on two variables that turn out to be "trivial" in situations that resemble the conditions above. Probably the first instance of this historically was the Hilbert symbol $(a,b)_p$ where p is a prime number and a and b are nonzero rational numbers (or even nonzero p-adic numbers). The Steinberg symbol $\{a,b\}_F$ for a field F (of characteristic not 2) is a universal construction subject to rules typified by $\{a,1-a\}_F$ being considered "trivial."

Corollary 4.23. *For a field F not of characteristic 2, $\mathbf{H}(F)$ is a division ring if and only if -1 is not a sum of two squares in F.*

Proof. We will prove the negations of both conditions are equivalent: to say $\mathbf{H}(F)$ is not a division ring means $\mathbf{H}(F) \cong \mathrm{M}_2(F)$, and by Theorem 4.21, $(-1,-1)_F \cong \mathrm{M}_2(F)$ if and only if $-1 = x^2 - (-1)y^2 = x^2 + y^2$ for some x and y in F. □

Corollary 4.24. *For every odd prime p, all quaternion algebras over $\mathbf{F}_p$ are isomorphic to $\mathrm{M}_2(\mathbf{F}_p)$.*

This includes Example 4.7 as a special case.

Proof. We will show for all nonzero a and b in $\mathbf{F}_p$ that $b = x^2 - ay^2$ for some x and y in $\mathbf{F}_p$. Write the equation $b = x^2 - ay^2$ as $b + ay^2 = x^2$. Let's count how many values each side of the equation has as x and y run over $\mathbf{F}_p$. The total number of squares in $\mathbf{F}_p$ is $(p+1)/2$ (*not* $(p-1)/2$, because we include 0 as a square), so $|\{x^2 : x \in \mathbf{F}_p\}| = (p+1)/2$, and since $a \neq 0$ we also have $|\{b + ay^2 : y \in \mathbf{F}_p\}| = (p+1)/2$. If the equation $b + ay^2 = x^2$ had no solution in $\mathbf{F}_p$ then $\{x^2\}$ and $\{b + ay^2\}$ would be disjoint subsets of $\mathbf{F}_p$, but their sizes add up to $(p+1)/2 + (p+1)/2 = p + 1 > \mathbf{F}_p$, so we'd have a contradiction. Therefore there exist some x and y in $\mathbf{F}_p$ such that $b + ay^2 = x^2$. □

In Theorem 4.21, the condition $(a, b)_F \cong \mathrm{M}_2(F)$ is symmetric in a and b since $(a, b)_F \cong (b, a)_F$. Therefore the condition $b \in N_a$ has to be symmetric in a and b even though it doesn't look symmetric in a and b at first glance. (That is, being able to solve $b = x^2 - ay^2$ in F may not look obviously equivalent to being able to solve $a = x^2 - by^2$ in F, particularly if one of the squares in a solution is 0.) The following alternate form of Theorem 4.21 replaces "$b \in N_a$" with conditions on an equation involving a and b that are visibly symmetric.

Theorem 4.25. *For a and b in $F^\times$, the following conditions are equivalent:*

1) $(a, b)_F \cong \mathrm{M}_2(F)$,
2) *the equation* $ax^2 + by^2 = 1$ *has a solution* (x, y) *in* F,
3) *the equation* $ax^2 + by^2 = z^2$ *has a solution* (x, y, z) *in* F *other than* $(0, 0, 0)$.

Proof. Exercise 4.12. □

Corollary 4.26. *For a and b in $F^\times$, the following conditions are equivalent:*

1) $(a, b)_F$ *is a division ring,*
2) *the equation* $ax^2 + by^2 = 1$ *has no solution* (x, y) *in* F,
3) *the only solution to* $ax^2 + by^2 = z^2$ *in* F *is* $(0, 0, 0)$.

Proof. Negate each part of Theorem 4.25. □

Exercises.

1. Show (4.2) is multiplicative.
2. If $e_1, e_2 \in (a, b)_F$ satisfy the conditions $e_1^2 \in F^\times$, $e_2^2 \in F^\times$, $e_1e_2 = -e_2e_1$, and e_1^2 and e_2^2 are not in $F^{\times 2}$, show $\{1, e_1, e_2, e_1e_2\}$ is a linearly independent set. Remember, F has characteristic $\neq 2$. (Hint: If $x_0 + x_1e_1 + x_2e_2 + x_3e_1e_2 = 0$ where $x_i \in F$, write this as $(x_0 + x_1e_1) + (x_2 + x_3e_1)e_2 = 0$ and multiply on the left by $x_2 - x_3e_1$. Note $x_2^2 - x_3^2e_1^2 \neq 0$ unless $x_2 = x_3 = 0$, since $e_1^2 \notin F^{\times 2}$.)
3. Under the isomorphism $(a, 1)_F \cong \mathrm{M}_2(F)$ determined by
$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad u \mapsto \begin{pmatrix} 0 & 1 \\ a & 0 \end{pmatrix}, \quad v \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad w \mapsto \begin{pmatrix} 0 & -1 \\ a & 0 \end{pmatrix}$$
what element of $(a, 1)_F$ corresponds to the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$?
4. Under the isomorphism $(1, 1)_F \cong \mathrm{M}_2(F)$ as in the previous exercise (with $a = 1$), let $x_0 + x_1u + x_2v + x_3w$ in $(1, 1)_F$ correspond to $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ in $\mathrm{M}_2(F)$. Write $\alpha, \beta, \gamma, \delta$ in terms of x_0, x_1, x_2, x_3 and vice versa. Check that, under this isomorphism, the norm on $(1, 1)_F$ corresponds to the determinant on $\mathrm{M}_2(F)$, but conjugation on $(1, 1)_F$ does

not correspond to the transpose on $\mathrm{M}_2(F)$. What operation on $\mathrm{M}_2(F)$ corresponds to conjugation on $(1,1)_F$?

5. When $a \neq -b$ in $F^\times$, check $\{1, u+v, w, (u+v)w\}$ is a quaternionic basis of $(a,b)_F$. Therefore $(a,b)_F \cong (a+b, -ab)_F$. For example, $(2,3)_{\mathbf{Q}} \cong (5,-6)_{\mathbf{Q}}$.
6. Show -1 is not a sum of two squares in the field $\mathbf{Q}(\sqrt{-7})$, so $\mathbf{H}(\mathbf{Q}(\sqrt{-7}))$ is a division ring. Is $\mathbf{H}(\mathbf{Q}(\sqrt{-2}))$ a division ring?
7. If $a, b \in F^\times$ satisfy $a + b = c^2$ for some $c \in F$, show $(a,b) \cong \mathrm{M}_2(F)$.
8. By Theorem 4.16,
$$b = x_0^2 - ay_0^2 \text{ for some } x_0, y_0 \in F \Longrightarrow (a,b)_F \cong \mathrm{M}_2(F)$$
by using the quaternionic basis $\{1, u, x_0v + y_0w, u(x_0v + y_0w)\}$ of $(a,b)_F$. What is wrong with the following alternate proof of the above implication?

   If $x_0 = 0$, then $y_0 \neq 0$ and $(a,b)_F = (a, -ay_0^2)_F \cong (a,-a)_F \cong \mathrm{M}_2(F)$.

   If $x_0 \neq 0$, then $(y_0u + v)^2 = x_0^2$ and the set $\{1, u, y_0u + v, u(y_0u + v)\}$ is a basis of $(a,b)_F$, so $(a,b)_F \cong (a, x_0^2)_F \cong \mathrm{M}_2(F)$.
9. A quaternion algebra over F is isomorphic to $\mathrm{M}_2(F)$ precisely when it has a nonzero element with norm 0. Prove $(a,-a)_F \cong \mathrm{M}_2(F)$ if $a \neq 0$, and $(a, 1-a)_F \cong \mathrm{M}_2(F)$ if $a \neq 0, 1$ by finding specific nonzero elements in them with norm 0.
10. In F^n, let $\mathbf{v}_1$ and $\mathbf{v}_2$ be linearly independent. For $a, b, c, d \in F$, show $a\mathbf{v}_1 + b\mathbf{v}_2$ and $c\mathbf{v}_1 + d\mathbf{v}_2$ are linearly independent in F^n if and only if the determinant $\left|\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right| = ad - bc$ is nonzero.
11. Decide if the following are division rings: $(2,-5)_{\mathbf{Q}}$, $(6,10)_{\mathbf{Q}}$, $(6,-10)_{\mathbf{Q}}$, $(5,11)_{\mathbf{Q}}$, $(5,-11)_{\mathbf{Q}}$.
12. Prove Theorem 4.25.

5. Isomorphism and Norms

Theorem 4.21 can be expressed as: $(a,b)_F \cong (a,1)_F$ if and only if $b \in N_a$. The following theorem generalizes this.

Theorem 5.1. *For $a, b, b' \in F^\times$, $(a,b)_F \cong (a,b')_F$ if and only if $b/b' \in N_a$.*

Proof. The direction ($\Leftarrow$) is much simpler, so we do that first. Suppose $b/b' = x_0^2 - ay_0^2$ for some x_0 and y_0 in F. Let $\{1, u, v, uv\}$ be the usual quaternionic basis of $(a,b')_F$. Check that

$$1, \quad u, \quad x_0v + y_0w, \quad u(x_0v + y_0w), \tag{5.1}$$

is also a quaternionic basis of $(a,b')_F$. Here $u^2 = a$ and $(x_0v + y_0w)^2 = b'x_0^2 - ab'y_0^2 = b'(b/b') = b$, so $(a,b')_F \cong (a,b)_F$.

To prove the reverse direction, that $(a,b)_F \cong (a,b')_F \Rightarrow b/b' \in N_a$, the isomorphic quaternion algebras $(a,b)_F$ and $(a,b')_F$ are either both division rings or both not division rings.

First suppose $(a,b)_F$ and $(a,b')_F$ are not division rings, so they are isomorphic to $\mathrm{M}_2(F)$. Then $b \in N_a$ and $b' \in N_a$ by Theorem 4.21. Since N_a is a subgroup of $F^\times$, $b/b' \in N_a$.

Next suppose $(a,b)_F$ and $(a,b')_F$ are division rings, so in particular a is not a square in F. Let $\{1, u, v, uv\}$ be the standard quaternionic basis of $(a,b')_F$, so

$$u^2 = a, \quad v^2 = b', \quad uv = -vu.$$

Since $(a,b')_F \cong (a,b)_F$, $(a,b')_F$ contains a quaternionic basis $\{1, u_0, v_0, u_0v_0\}$ where

$$u_0^2 = a, \quad v_0^2 = b, \quad u_0v_0 = -v_0u_0.$$

The polynomial $T^2 - a$ is irreducible over F since a is not a square in F, and both u and u_0 are roots of this polynomial in $(a,b')_F$, so Theorem B.2 implies $u = qu_0q^{-1}$ for some nonzero $q \in (a,b')_F$. Set $\widetilde{v} = qv_0q^{-1}$, so $\widetilde{v}^2 = (qv_0q^{-1})(qv_0q^{-1}) = qv_0^2q^{-1} = qbq^{-1} = b$. Then

$$u_0v_0 = -v_0u_0 \Longrightarrow (qu_0q^{-1})(qv_0q^{-1}) = -(qv_0q^{-1})(qu_0q^{-1}) \Longrightarrow u\widetilde{v} = -\widetilde{v}u.$$

The elements of $(a,b')_F$ that anti-commute with u are $Fv+Fw$ (Exercise 3.3), so $\widetilde{v} = xv+yw$ for some x and y in F. Then

$$b = \widetilde{v}^2 = (xv+yw)^2 = b'x^2 - b'ay^2 = b'(x^2 - ay^2) \Longrightarrow \frac{b}{b'} \in N_a.$$

□

Remark 5.2. Theorem 5.1 gives us a new proof that $(a,-ab)_F \cong (a,b)_F$: $-a = x^2 - ay^2$ with $x = 0$ and $y = 1$. It might appear that this theorem reproves $(a,bc^2)_F \cong (a,b)_F$ ($c^2 = x^2 - ay^2$ with $x = c$ and $y = 0$), but this is circular because we will be using that isomorphism in the proof of Theorem 5.1.

Example 5.3. $(2,3)_{\mathbf{Q}} \cong (2,21)_{\mathbf{Q}}$ since $21/3 = 7 = x^2 - 2y^2$ using $x = 3$ and $y = 1$.

Example 5.4. To decide if the division rings $(2,3)_{\mathbf{Q}}$ and $(2,5)_{\mathbf{Q}}$ are isomorphic is equivalent to deciding if $5/3 = x^2 - 2y^2$ for some rational numbers x and y. This equation has no rational solution (Exercise 5.3a), so $(2,3)_{\mathbf{Q}} \ncong (2,5)_{\mathbf{Q}}$.

Corollary 5.5. *For distinct primes p and q that are $3 \bmod 4$, $(-1,p)_{\mathbf{Q}}$ is not isomorphic to $(-1,q)_{\mathbf{Q}}$.*

Proof. If $(-1,p)_{\mathbf{Q}} \cong (-1,q)_{\mathbf{Q}}$ then $q/p \in N_{-1}(\mathbf{Q})$, so $q/p = x^2 + y^2$ for some rational numbers x and y. Write x and y with a common denominator: $x = m/d$ and $y = n/d$ with integers m, n, and d where $d \neq 0$. Then

$$qd^2 = p(m^2 + n^2).$$

Since $p \neq q$, $m^2 + n^2 \equiv 0 \bmod q$. That implies m and n are divisible by q (because $-1 \bmod q$ is not a square), so qd^2 is divisible by q^2, and thus $q|d$. In the equation $qd^2 = p(m^2+n^2)$ the numbers m, n, and d are all divisible by q, so we can divide through by q^2 and get a similar equation where m, n, and d are replaced by m/q, n/q, and d/q. Repeating this *ad infinitum* d is divisible by arbitrarily high powers of q, a contradiction. □

There are infinitely many primes congruent to $3 \bmod 4$, so Corollary 5.5 shows there are infinitely many non-isomorphic quaternion algebras over $\mathbf{Q}$!

Theorem 5.6. *If $f\colon A_1 \to A_2$ is an isomorphism of quaternion algebras over F then we have $\overline{f(q)} = f(\overline{q})$ for all $q \in A_1$. In particular, $\mathrm{N}(f(q)) = \mathrm{N}(q)$.*

Proof. Once we show $\overline{f(q)} = f(\overline{q})$ for all $q \in A_1$ we get

$$\mathrm{N}(f(q)) = f(q)\overline{f(q)} = f(q)f(\overline{q}) = f(q\overline{q}) = f(\mathrm{N}(q)) = \mathrm{N}(q)$$

since $\mathrm{N}(q) \in F$ and f fixes all elements of F.

Write $q = x_0 + q_0$ where $x_0 \in F$ and q_0 is a pure quaternion in A_1 (Example 3.7). Then

$$f(\overline{q}) = f(x_0 - q_0) = x_0 - f(q_0). \tag{5.2}$$

We will show $f(q_0)$ is a pure quaternion in A_2. This is obvious if $q_0 = 0$, so assume $q_0 \neq 0$. Since q_0 is pure in A_1 we have $q_0^2 \in F$ by (3.2), so also $f(q_0)^2 \in F$. By Exercise 3.6 ii, either $f(q_0)$ is pure or $f(q_0) \in F$. We can't have $f(q_0) \in F$ since $f(F) = F$, f is injective, and $q_0 \notin F$ (the only pure quaternion in F is 0). Thus $f(q_0)$ is pure, so $\overline{f(q_0)} = -f(q_0)$, which turns (5.2) into

$$f(\overline{q}) = x_0 + \overline{f(q_0)} = \overline{x_0 + f(q_0)} = \overline{f(x_0 + q_0)} = \overline{f(q)}.$$

□

Corollary 5.7. *If A_1 and A_2 are isomorphic quaternion algebras over F then the norm maps $A_1 \to F$ and $A_2 \to F$ have the same image.*

Proof. This is immediate from $\mathrm{N}(q) = \mathrm{N}(f(q))$ for an isomorphism $f\colon A_1 \to A_2$. □

Example 5.8. The quaternion algebras $\mathbf{H}(\mathbf{Q})$ and $(2,3)_{\mathbf{Q}}$ are division rings (for $(2,3)_{\mathbf{Q}}$ see Example 3.10), but they are not isomorphic quaternion algebras over $\mathbf{Q}$ since the norm on $\mathbf{H}(\mathbf{Q})$ is nonnegative and the norm on $(2,3)_{\mathbf{Q}}$ has negative values by Example 3.6.

Exercises.

1. In the proof of Theorem 5.1, show (5.1) is a quaternionic basis of $(a,b)_F$.
2. Let $a, b, b' \in F^\times$.
   a) If $(a,b)_F \cong \mathrm{M}_2(F)$, prove $(a,b')_F \cong (a,bb')_F$. (For example, if $p = 2$ or $p \equiv 1 \bmod 4$, then we already know $(p,-1)_{\mathbf{Q}} \cong \mathrm{M}_2(\mathbf{Q})$, so $(p,r)_{\mathbf{Q}} \cong (p,-r)_{\mathbf{Q}}$ for all $r \in \mathbf{Q}^\times$.) Is the converse true?
   b) By part a, $(2,3)_{\mathbf{Q}} \cong (2,-3)_{\mathbf{Q}}$. Show $(-2,3)_{\mathbf{Q}} \cong \mathrm{M}_2(\mathbf{Q})$ and $(-2,-3)_{\mathbf{Q}} \cong \mathbf{H}(\mathbf{Q})$.
3. If p is a prime number such that $p \equiv 3$ or $5 \bmod 8$, then $(p,2)_{\mathbf{Q}}$ is a division ring by Exercise 3.5.
   a) Show the equation $5/3 = x^2 - 2y^2$ has no rational solution, so $(3,2)_{\mathbf{Q}} \ncong (5,2)_{\mathbf{Q}}$. (Hint: Mimic the proof of Corollary 5.5.)
   b) For distinct primes p and q that are 3 or 5 mod 8 (this means $p \equiv 3$ or $5 \bmod 8$ and $q \equiv 3$ or $5 \bmod 8$), show $(p,2)_{\mathbf{Q}}$ is not isomorphic to $(q,2)_{\mathbf{Q}}$.
   c) Prove the converse of Exercise 3.5 is false by showing $(15,2)_{\mathbf{Q}}$ and $(33,2)_{\mathbf{Q}}$ are division rings. Neither 15 nor 33 is 3 or 5 mod8.

Appendix A. Sum of square identities using $\mathbf{C}$ and $\mathbf{H}$

In $\mathbf{C}$, the fact that $\overline{zw} = \overline{z}\,\overline{w}$ implies a formula for the product of sums of two squares as a sum of two squares:

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2. \tag{A.1}$$

Writing $z = a + bi$ and $w = c + di$, the left side is $z\overline{z}w\overline{w}$ and the right side is $zw\overline{zw} = zw\overline{z}\,\overline{w} = z\overline{z}w\overline{w}$.

Quaternionic multiplication in $\mathbf{H}$ leads to a four-square identity generalizing (A.1):

$$\begin{aligned}(a_1^2 + b_1^2 + c_1^2 + d_1^2)(a_2^2 + b_2^2 + c_2^2 + d_2^2) \;=\; & (a_1a_2 - b_1b_2 - c_2c_2 - d_1d_2)^2 + \\ & (a_1b_2 + b_1a_2 + c_1d_2 - d_1c_2)^2 + \\ & (a_1c_2 - b_1d_2 + c_1a_2 + d_1b_2)^2 + \\ & (a_1d_2 + b_1c_2 - c_1b_2 + d_1a_2)^2.\end{aligned}$$

To prove this, a sum of four squares is the norm of a quaternion: $a^2 + b^2 + c^2 + d^2 = \mathrm{N}(a + bi + cj + dk)$. Letting $q_1 = a_1 + b_1 i + c_1 j + d_1 k$ and $q_2 = a_2 + b_2 i + c_2 j + d_2 k$ in $\mathbf{H}$, feeding these into the formula $\mathrm{N}(q_1)\,\mathrm{N}(q_2) = \mathrm{N}(q_1 q_2)$ from (2.3) implies the four-square identity above.

It's hard to imagine the four-square identity could be found without quaternions, but it was! Unbeknownst to Hamilton, the four-square identity was written down by Euler in 1748, a hundred years before the discovery of quaternions.

Appendix B. Conjugates in a division ring

The label "conjugates" in the heading of this appendix doesn't refer to the conjugation operation on a quaternion algebra, but rather to conjugation in the sense of group theory: elements x and y in a group G are called conjugate when $y = gxg^{-1}$ for some $g \in G$. For a division ring D, its nonzero elements are a group under multiplication and we call x and y in D conjugate if $y = dxd^{-1}$ for some $d \in D^\times = D - \{0\}$.

When F is a field, a polynomial of degree n in $F[t]$ has at most n roots in F. Surprisingly, over a division ring a polynomial of degree n can have more than n roots. For example, the polynomial $t^2 + 1$ has *infinitely many* roots in $\mathbf{H}$: by (2.1), a pure quaternion $bi + cj + dk$ with $b^2 + c^2 + d^2 = 1$ is a root of $t^2 + 1$. As if to compensate for there being more roots than the degree, all these roots turn out to be conjugate to each other in the sense of group theory: if $x^2 + 1 = 0$ and $y^2 + 1 = 0$ in $\mathbf{H}$ then $y = qxq^{-1}$ for some $q \in \mathbf{H}^\times$. This is a special case of the following general theorem of Dickson. Recall (Exercise 2.3) that the center of a division ring is a field.

Theorem B.1 (Dickson)**.** *Let D be a division ring with center F and $f(t)$ be an irreducible polynomial in $F[t]$. All roots of $f(t)$ in D are conjugate to each other: if $f(x) = 0$ and $f(y) = 0$ for x and y in D then $y = dxd^{-1}$ for some $d \in D^\times$.*

A proof of Theorem B.1 can be found in [4, Theorem 16.8]. We use this result (in Section 5) only when $\dim_F(D) < \infty$ and $\deg f = 2$, so we'll prove Theorem B.1 in this special case:

Theorem B.2. *Let D be a division ring with center equal to a field F, and assume $\dim_F(D) < \infty$. If $f(t) = t^2 + c_1 t + c_0 \in F[t]$ is irreducible over F then x and y in D that satisfy $f(x) = 0$ and $f(y) = 0$ are conjugate: $y = dxd^{-1}$ for some $d \in D^\times$.*

Proof. We have $x^2 + c_1 x + c_0 = 0$ and $y^2 + c_1 y + c_0 = 0$. Therefore $y^2 + c_1 y = x^2 + c_1 x$, and by adding yx to both sides we can write this equation in the clever way

$$y(y + x + c_1) = (y + x + c_1)x$$

If $y + x + c_1 \neq 0$ then set $d = y + x + c_1$, so $yd = dx$ and $d \neq 0$ in D. Thus $y = dxd^{-1}$.

What if $y + x + c_1 = 0$? In that case, to find a nonzero d that makes $yd = dx$ requires a bit of linear algebra. Define $L\colon D \to D$ by $L(d) = dx - yd$ for all $d \in D$. Since F is the center, L is F-linear:

$$L(d_1 + d_2) = (d_1 + d_2)x - y(d_1 + d_2) = d_1 x - y d_1 + d_2 x - y d_2 = L(d_1) + L(d_2),$$

$$L(cd) = (cd)x - y(cd) = cdx - cyd = c(dx - yd) = cL(d),$$

where $c \in F$. Since $y + x + c_1 = 0$, $L(d) = dx + (x + c_1)d$. This formula implies $xL(d) = L(d)x$:

$$xL(d) = x(dx + (x + c_1)d) = xdx + (x^2 + c_1 x)d = xdx - c_0 d,$$

$$L(d)x = dx^2 + (x + c_1)dx = d(-c_1 x - c_0) + xdx + c_1 dx = xdx - c_0 d,$$

and these two values are equal. Thus $L(d)$ commutes with x for all $d \in D$.

Since $f(x) = 0$ and $f(t)$ has no roots in F, $x \notin F$. All of $L(D)$ commutes with x, and not all of D commutes with x (otherwise x would be in the center of D, which is F), so $L(D)$ is a proper subspace of D: L is not surjective. A basic theorem from linear algebra says a linear map $V \to V$ where $\dim_F(V) < \infty$ is one-to-one if and only if it is onto. Since $L\colon D \to D$ is not onto and $\dim_F(D) < \infty$, L is is not one-to-one: $\ker L \neq \{0\}$. Therefore some nonzero $d \in D$ satisfies $L(d) = 0$, which means $dx = yd$. Thus $y = dxd^{-1}$. □

References

[1] L. E. Dickson, "Linear algebras in which division is always uniquely possible," Bull. Amer. Math. Soc. **12** (1905-6), 441–442.

[2] H.-D. Ebbinghaus *et al.*, *Numbers*, Springer-Verlag, New York, 1990.

[3] W. Hamilton, *Mathematical Papers, Vol. III, Algebra*, Cambridge Univ. Press, 1967.

[4] T. Y. Lam, *A First Course in Noncommutative Rings*, Springer-Verlag, New York, 1991.

[5] T. Y. Lam, *Introduction to Quadratic Forms over Fields*, Amer. Math. Soc., Providence, 2005.

[6] T. Y. Lam, "Hamilton's quaternions," pp. 429–454 of *Handbook of algebra, Vol. 3*, North-Holland, Amsterdam, 2003.

[7] C. Maclachlan and A. Reid, *The arithmetic of hyperbolic 3-manifolds*, Springer-Verlag, New York, 2003.

[8] T. Miyake, *Modular forms*, Springer-Verlag, Berlin, 1989.

[9] J. H. Silverman, *The Arithmetic of Elliptic Curves*, Springer-Verlag, New York, 1986.

[10] M.-F. Vignéras, *Arithmétique des algèbres de quaternions*, Springer-Verlag, Berlin, 1980.