Why quaternion algebras have rank 4

Darij Grinberg

March 29, 2019

Contents

1. The statement

This brief note is devoted to a simple (and well-known) result in noncommutative algebra, which is not deep but nevertheless subtler than it appears. It concerns the so-called *quaternion algebras:*

Definition 1.1. Let $\mathbf{k}$ be a commutative ring[1]. Let $a \in \mathbf{k}$ and $b \in \mathbf{k}$. The *quaternion algebra* $H_{a,b}$ is defined to be the $\mathbf{k}$-algebra with generators i and j and relations

$$i^2 = a, \qquad j^2 = b, \qquad ij = -ji. \tag{1}$$

We notice that the well-known *algebra of (Hamilton) quaternions* $\mathbb{H}$ is a particular case of this construction: namely, it is the quaternion algebra $H_{-1,-1}$ when $\mathbf{k} = \mathbb{R}$.

Now, the result that we will be discussing is the following:

Theorem 1.2. Let $\mathbf{k}$ be a commutative ring. Let $a \in \mathbf{k}$ and $b \in \mathbf{k}$. Then, $(1, i, j, ij)$ is a basis of the $\mathbf{k}$-module $H_{a,b}$.

The purpose of this note is to show two things:

1. that Theorem 1.2 is not obvious, and should not be taken lightly[2];
2. that Theorem 1.2 is nevertheless easy to prove, and there are various ways to do so.

Remark 1.3. Definition 1.1 is not the only possible way to define the quaternion algebra $H_{a,b}$. Some authors (e.g., Keith Conrad in [Conrad15, §3]) instead **define** $H_{a,b}$ to be a certain rank-4 free $\mathbf{k}$-module with basis e, i, j, k and with multiplication rules

$$\begin{array}{llll} e \cdot e = e, & e \cdot i = i, & e \cdot j = j, & e \cdot k = k, \\ i \cdot e = i, & i \cdot i = ae, & i \cdot j = k, & i \cdot k = aj, \\ j \cdot e = j, & j \cdot i = -k, & j \cdot j = be, & j \cdot k = -bi, \\ k \cdot e = k, & k \cdot i = -aj, & k \cdot j = bi, & k \cdot k = -abe. \end{array}$$

This definition gives a $\mathbf{k}$-algebra that is isomorphic to the $H_{a,b}$ from our Definition 1.1[3]; however, the proof of this isomorphism is not immediately obvious (it is, in fact, tantamount to proving Theorem 1.2). Moreover, this definition requires laborious verifications in order to convince oneself that it is well-defined; in fact, the associativity of its multiplication must be checked. We shall be using Definition 1.1 as the definition of $H_{a,b}$.

2. Isn't it obvious?

I shall first try to explain why Theorem 1.2 is not self-evident (even in "good weather" – e.g., when $\mathbf{k}$ is a field, and a and b are nonzero[4]).

[1] Some conventions: The word "ring" always means "associative ring with 1" in this note. When $\mathbf{k}$ is a commutative ring, then a $\mathbf{k}$-algebra is supposed to be associative and unital, and to satisfy the axiom $\lambda (ab) = (\lambda a) b = a (\lambda b)$ for every $\lambda \in \mathbf{k}$ and every two elements a and b of the $\mathbf{k}$-algebra.

[2] This note has been written in response to numerous incorrect proofs found in students' homework.

[4] I am calling this "good weather" because one might intuitively expect that claims such as Theorem 1.2 become easier to prove when $\mathbf{k}$ is assumed to be a field, and a and b to be

The $\mathbf{k}$-algebra $H_{a,b}$ in Definition 1.1 is an example of a (noncommutative) $\mathbf{k}$-algebra with two generators. You probably are used to $\mathbf{k}$-algebras with one generator. These are noticeably simpler: they are all commutative, and they can be written as polynomial rings modulo ideals. For example, here is a "one-variable" analogue of $H_{a,b}$:

Example 2.1. Let $\mathbf{k}$ be a commutative ring, and let $a \in \mathbf{k}$. Let C_a be the $\mathbf{k}$-algebra with generator i and relation $i^2 = a$. I claim that $(1, i)$ is a basis of the $\mathbf{k}$-module C_a.

Indeed, the definition of C_a shows that $C_a \cong \mathbf{k}[X] / \left(X^2 - a\right)$ as $\mathbf{k}$-algebras (where $\mathbf{k}[X]$ is the polynomial ring over $\mathbf{k}$ in one variable X); the isomorphism sends i to $\overline{X}$ (where $\overline{p}$ denotes the remainder class of any $p \in \mathbf{k}[X]$ in the quotient ring $\mathbf{k}[X] / \left(X^2 - a\right)$). But the polynomial $X^2 - a$ is monic, and thus every polynomial in $\mathbf{k}[X]$ can be divided by $X^2 - a$ with remainder. As a consequence of this, the $\mathbf{k}$-module $\mathbf{k}[X] / \left(X^2 - a\right)$ has basis $\left(\overline{1}, \overline{X}\right)$. Due to our isomorphism $C_a \cong \mathbf{k}[X] / \left(X^2 - a\right)$, this shows that the $\mathbf{k}$-module C_a has basis $(1, i)$. Thus, our claim about C_a is proven.

Can we generalize this argument to the $H_{a,b}$ in Theorem 1.2? It is not immediately clear how. The $\mathbf{k}$-algebra $H_{a,b}$ is not commutative (at least we do not see why it should be[5]), so we cannot identify it with a quotient of a polynomial ring, and even if we could, it would be a polynomial ring in two variables, and how would division with remainder work in that ring? (There **is** an analogue of division with remainder for multivariate polynomials, and there **is** such a thing as noncommutative polynomials; with some luck you might be able to make these things work together nicely and possibly get an analogous proof, but that will require some creativity to say the least.)

So it is not that simple to prove Theorem 1.2.

You might think that at least the $\mathbf{k}$-linear independency of $1, i, j$ in Theorem 1.2 is obvious, because all of the relations in (1) are "of degree 2" (whatever this means). But this is not a valid proof, and in fact such reasoning can completely fail:

Example 2.2. Let S be the $\mathbb{R}$-algebra with generators i and j and relations

$$i^2 = 1, \qquad j^2 = 1, \qquad ij = 2ji. \tag{2}$$

(Looks similar to (1), doesn't it?) I claim that the elements $1, i, j$ of S are not $\mathbb{R}$-linearly independent. Actually, I claim that S is a trivial ring (i.e., all elements of S are equal to 0). Why is this so?

nonzero. However, this does not actually happen here; the general case of Theorem 1.2 is no harder, and nothing is gained by making any assumptions on $\mathbf{k}$, a and b. (That said, such assumptions can be helpful in other, similar situations.)

[5] It is commutative when $2 = 0$ in $\mathbf{k}$. But this is probably not the most interesting use case of quaternion algebras...

Well, by repeated application of (2), we find that

$$1 = \underbrace{1}_{=i^2=ii} \cdot \underbrace{1}_{=j^2=jj} = i \underbrace{ij}_{=2ji} j = 2ijij = 2\left(\underbrace{ij}_{=2ji}\right)^2 = 2\left(2ji\right)^2$$
$$= 8\left(ji\right)^2 = 8j \underbrace{ij}_{=2ji} i = 16 \underbrace{jj}_{=j^2=1} \underbrace{ii}_{=i^2=1} = 16. \tag{3}$$

Of course, this doesn't mean that we have magically proven that the real numbers 1 and 16 are equal; in fact, (3) is merely an equality inside S. But this equality shows that S is a trivial ring: In fact, subtracting 1 from (3), we obtain $0 = 15$, and multiplying this by the scalar $\frac{1}{15} \in \mathbb{R}$, we obtain $0 = 1$ (in S). But every ring which satisfies $0 = 1$ is a trivial ring.

Thus, the fact that the relations (2) are "of degree 2" did not prevent S from collapsing to a trivial ring. In general, if an algebra is given by generators and relations, it is not easy to tell how "large" it is (e.g., what dimension it has), and even whether it is trivial.

I hope you now have an idea of where the difficulty in Theorem 1.2 lies.

Example 2.2 also hints at the importance of Theorem 1.2: You might be tempted to prove some an equality between two elements of $\mathbf{k}$ by comparing their images under the canonical homomorphism $\mathbf{k} \to H_{a,b}$ [6]; but without knowing that this canonical homomorphism is injective, such a proof would not work. Theorem 1.2 implies that the canonical homomorphism $\mathbf{k} \to H_{a,b}$ is injective (because this homomorphism sends 1 to 1, and Theorem 1.2 shows that 1 is an entry of a basis of $H_{a,b}$).

Results like Theorem 1.2 (that is, results which state that an algebra given by generators and relations has a particular basis) are often called *PBW-like theorems*, in honor of the PBW (Poincaré-Birkhoff-Witt) theorem which makes such a claim about the universal enveloping algebra of a Lie algebra. Two surveys about such theorems are Bergman's [Bergma78] and Shepler's and Witherspoon's [SheWit14]; they will give you enough artillery to destroy Theorem 1.2 many times over. In this note, we will fight with bare hands instead; Theorem 1.2 is not trivial, but it is not **that** tough either.

[6] For example, the famous Euler four-square identity

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

can be proven in this way.

3. Spanning

Clearly, Theorem 1.2 will follow if we can prove the following two lemmas:

Lemma 3.1. Let $\mathbf{k}$ be a commutative ring. Let $a \in \mathbf{k}$ and $b \in \mathbf{k}$. Then, the sequence $(1, i, j, ij)$ spans the $\mathbf{k}$-module $H_{a,b}$.

Lemma 3.2. Let $\mathbf{k}$ be a commutative ring. Let $a \in \mathbf{k}$ and $b \in \mathbf{k}$. Then, the sequence $(1, i, j, ij)$ of vectors in $H_{a,b}$ is $\mathbf{k}$-linearly independent.

Lemma 3.1 is easy. Let me sketch two proofs – an "uncombed" algorithmic one, and a slick algebraic one. In truth, the two proofs are essentially equivalent, and the second is merely a way of rewriting the first in a more streamlined way. The first proof is more intuitive, to make up for it.

First proof of Lemma 3.1 (sketched). The $\mathbf{k}$-algebra $H_{a,b}$ is generated by i and j. Thus, the $\mathbf{k}$-module $H_{a,b}$ is spanned by all finite products of i's and j's (in arbitrary order, and including the empty product). It therefore suffices to show that each such product is a $\mathbf{k}$-linear combination of $1, i, j, ij$.

But this is easy: We can first reduce any product of i's and j's to a product of the form $i^n j^m$ (by repeatedly replacing ji factors by $-ij$, because the relations (1) imply $ji = -ij$), and then repeatedly apply $i^2 = a$ and $j^2 = b$ until neither i nor j appears in any power greater than 1. For example, the product $ijiiijjijj$ is thus simplified as follows:

$$\begin{aligned} i \underbrace{ji}_{=-ij} iij \underbrace{ji}_{=-ij} j &= ii \underbrace{ji}_{=-ij} i \underbrace{ji}_{=-ij} jj = iii \underbrace{ji}_{=-ij} ijjj = -iiii \underbrace{ji}_{=-ij} jjj \\ &= - \underbrace{ii}_{=i^2=a} \underbrace{ii}_{=i^2=a} i \underbrace{jj}_{=j^2=b} \underbrace{jj}_{=j^2=b} = -a^2 b^2 i. \end{aligned}$$

□

Second proof of Lemma 3.1. Let U be the $\mathbf{k}$-submodule of $H_{a,b}$ spanned by $1, i, j, ij$. Our goal is then to prove that $U = H_{a,b}$.

Notice that $ji = -ij$ in $H_{a,b}$ (since $ij = -ji$). Hence, $\underbrace{ji}_{=-ij} j = -i \underbrace{jj}_{=j^2=b} = -bi$.

We shall use the following notation: If $p_1, p_2, \ldots, p_k$ are some vectors in a $\mathbf{k}$-module V, then $\langle p_1, p_2, \ldots, p_k \rangle$ shall denote the $\mathbf{k}$-submodule of V spanned by $p_1, p_2, \ldots, p_k$. Using this notation, the definition of U rewrites as follows: $U = \langle 1, i, j, ij \rangle$. Thus,

$$iU = i \langle 1, i, j, ij \rangle = \left\langle \underbrace{i1}_{=i}, \underbrace{ii}_{=i^2=a}, ij, \underbrace{ii}_{=i^2=a} j \right\rangle = \langle i, a, ij, aj \rangle \subseteq U$$

(since all of i, a, ij, aj belong to U). Also, from $U = \langle 1, i, j, ij \rangle$, we obtain

$$jU = j\langle 1, i, j, ij \rangle = \left\langle \underbrace{j1}_{=j}, \underbrace{ji}_{=-ij}, \underbrace{jj}_{=j^2=b}, \underbrace{jij}_{=-bi} \right\rangle = \langle j, -ij, b, -bi \rangle \subseteq U$$

(since all of $j, -ij, b, -bi$ belong to U). Now, recall that the $\mathbf{k}$-algebra $H_{a,b}$ is generated by i and j. Therefore, from $iU \subseteq U$ and $jU \subseteq U$, we obtain $H_{a,b}U \subseteq U$ [7].

But $1 \in U$, and thus $H_{a,b} = H_{a,b}\underbrace{1}_{\in U} \subseteq H_{a,b}U \subseteq U$. Combining this with $U \subseteq H_{a,b}$ (which is obvious), we obtain $U = H_{a,b}$. As we have explained, this completes the proof of Lemma 3.1. □

4. Linear independency: a not-quite-proof

Now that Lemma 3.1 is proven, it remains to verify Lemma 3.2. This is harder. Here is an "almost-proof" of a particular case, which is insufficient, but which I shall show because it demonstrates a rather enticing trap to fall into:

Incomplete proof of a particular case of Lemma 3.2. Let us try to prove Lemma 3.2 in the particular case when $\mathbf{k} = \mathbb{R}$ and $a = -1$ and $b = -1$.

So let us assume that $\mathbf{k} = \mathbb{R}$, $a = -1$ and $b = -1$. (Thus, of course, $H_{a,b} = H_{-1,-1}$ is the ring $\mathbb{H}$ of quaternions over $\mathbb{R}$.) We need to show that the sequence $(1, i, j, ij)$ of vectors in $H_{-1,-1}$ is $\mathbf{k}$-linearly independent. Thus, assume that $\alpha, \beta, \gamma, \delta$ are four elements of $\mathbf{k}$ satisfying $\alpha 1 + \beta i + \gamma j + \delta ij = 0$ in $H_{-1,-1}$. We must prove that $\alpha = \beta = \gamma = \delta = 0$.

We have $ii = i^2 = a = -1$, $jj = j^2 = b = -1$ and $ji = -ij$ (since $ij = -ji$).

Multiplying the equation $\alpha 1 + \beta i + \gamma j + \delta ij = 0$ by i from the right, we obtain $\alpha i + \beta ii + \gamma ji + \delta iji = 0$. Since $ii = -1$, $ji = -ij$ and $i\underbrace{ji}_{=-ij} = -\underbrace{ii}_{=-1}j = j$, this rewrites as $\alpha i - \beta 1 - \gamma ij + \delta j = 0$. In other words, $-\beta 1 + \alpha i + \delta j - \gamma ij = 0$.

Multiplying the equation $\alpha 1 + \beta i + \gamma j + \delta ij = 0$ by j from the right, we obtain $\alpha j + \beta ij + \gamma jj + \delta ijj = 0$. Since $jj = -1$ and $i\underbrace{jj}_{=-1} = -i$, this rewrites as $\alpha j + \beta ij - \gamma 1 - \delta i = 0$. In other words, $-\gamma 1 - \delta i + \alpha j + \beta ij = 0$.

[7] The same argument, in a bit more detail: Let G be the subset $\{x \in H_{a,b} \mid xU \subseteq U\}$ of $H_{a,b}$. It is straightforward to see that G is closed under addition, multiplication and scaling (by elements of $\mathbf{k}$), and that G contains 0 and 1. Thus, G is a $\mathbf{k}$-subalgebra of $H_{a,b}$. Moreover, this $\mathbf{k}$-subalgebra G contains i (since $iU \subseteq U$) and j (since $jU \subseteq U$). Thus, this $\mathbf{k}$-subalgebra G must be the whole $H_{a,b}$ (because the $\mathbf{k}$-algebra $H_{a,b}$ is generated by i and j, and therefore any $\mathbf{k}$-subalgebra of $H_{a,b}$ that contains i and j must be the whole $H_{a,b}$). In other words, every $x \in H_{a,b}$ belongs to G. In other words, every $x \in H_{a,b}$ satisfies $xU \subseteq U$ (since this is what it means for x to belong to G). In other words, $H_{a,b}U \subseteq U$, qed.

Multiplying the equation $\alpha 1 + \beta i + \gamma j + \delta ij = 0$ by ij from the right, we obtain $\alpha ij + \beta iij + \gamma jij + \delta ijij = 0$. Since $\underbrace{ii}_{=-1} j = -j$, $\underbrace{ji}_{=-ij} j = -i \underbrace{jj}_{=-1} = i$ and $i \underbrace{ji}_{=-ij} j = -\underbrace{ii}_{=-1} \underbrace{jj}_{=-1} = -1$, this rewrites as $\alpha ij - \beta j + \gamma i - \delta 1 = 0$. In other words, $-\delta 1 + \gamma i - \beta j + \alpha ij = 0$.

Thus, we have found the four equations

$$\begin{cases} \alpha 1 + \beta i + \gamma j + \delta ij = 0; \\ -\beta 1 + \alpha i + \delta j - \gamma ij = 0; \\ -\gamma 1 - \delta i + \alpha j + \beta ij = 0; \\ -\delta 1 + \gamma i - \beta j + \alpha ij = 0 \end{cases}.$$

These equations are linear in $\alpha, \beta, \gamma, \delta$, and so we can solve them by a sort of Gaussian elimination[8]. For example, if we add together the first equation, i times the second equation, j times the third equation, and ji times the fourth equation, then we obtain $4\delta ij = 0$ (after some simplifications); multiplying this with ij from the right again, we obtain $-4\delta = 0$. Since $\mathbf{k} = \mathbb{R}$, we can divide this by -4, and obtain $\delta = 0$. This suggests that we have achieved at least part of our goal (of proving that $\alpha = \beta = \gamma = \delta = 0$). However, we have not! The equality $\delta = 0$ we have proven is an equality inside $H_{a,b}$, whereas the equality $\alpha = \beta = \gamma = \delta = 0$ that we want to prove is an equality in $\mathbf{k}$. We cannot derive the equality $\delta = 0$ in $\mathbf{k}$ from the equality $\delta = 0$ inside $H_{a,b}$, because (at the current stage) we cannot even be sure that $H_{a,b}$ is not the trivial ring. So this approach is not particularly hopeful. □

5. Linear independency: proof by construction

Now, we need to prove Lemma 3.2. The first proof that I will show is based upon the following idea: Theorem 1.2 shows that $H_{a,b}$ is a free $\mathbf{k}$-module of rank 4; thus, the $\mathbf{k}$-algebra structure of $H_{a,b}$ can be regarded as a $\mathbf{k}$-algebra structure on $\mathbf{k}^4$, once Theorem 1.2 is proven. We are going to reverse this cart, and first construct a $\mathbf{k}$-algebra structure on $\mathbf{k}^4$, and then prove that it is isomorphic to $H_{a,b}$, and use this to prove Theorem 1.2. Before we give this proof, we state two obvious facts:

Lemma 5.1. Let $\mathbf{k}$ be a commutative ring. Let P and Q be two $\mathbf{k}$-modules. Let $f : P \to Q$ be a $\mathbf{k}$-linear map. Let $v_1, v_2, \ldots, v_n$ be some vectors in a $\mathbf{k}$-module P. If the vectors $f(v_1), f(v_2), \ldots, f(v_n)$ are $\mathbf{k}$-linearly independent, then the vectors $v_1, v_2, \ldots, v_n$ are also $\mathbf{k}$-linearly independent.

[8] Be careful, since the coefficients are elements of the noncommutative algebra $H_{-1,-1}$ (and you do not know $H_{-1,-1}$ well yet; you cannot tell whether some element of $H_{-1,-1}$ is zero or not); but tricks such as adding a multiple of one equation to another still work, of course.

Lemma 5.1 is a simple linear-algebraic fact[9].

The next lemma is a universal property of $H_{a,b}$. Indeed, recall that $H_{a,b}$ is the $\mathbf{k}$-algebra with generators i and j and relations (1). In other words, $H_{a,b}$ is the quotient of the free $\mathbf{k}$-algebra with generators I and J (here I am using uppercase letters in order to tell them apart from the i and j in $H_{a,b}$) modulo the ideal generated by $I^2 - a$, $J^2 - b$ and $IJ + JI$. Combining the universal property of a free $\mathbf{k}$-algebra and that of a quotient algebra, we thus obtain the following universal property of $H_{a,b}$:

Lemma 5.2. Let $\mathbf{k}$ be a commutative ring. Let $a \in \mathbf{k}$ and $b \in \mathbf{k}$. Let A be any $\mathbf{k}$-algebra. Let $\mathfrak{i}$ and $\mathfrak{j}$ be any two elements of A satisfying

$$\mathfrak{i}^2 = a, \qquad \mathfrak{j}^2 = b, \qquad \mathfrak{ij} = -\mathfrak{ji}. \tag{4}$$

Then, there exists a unique $\mathbf{k}$-algebra homomorphism $\Phi : H_{a,b} \to A$ satisfying $\Phi(i) = \mathfrak{i}$ and $\Phi(j) = \mathfrak{j}$.

Now, we can prove Lemma 3.2:

First proof of Lemma 3.2. Consider the free $\mathbf{k}$-module $\mathbf{k}^4$ of all vectors $(x_1, x_2, x_3, x_4)^T$ (with $x_1, x_2, x_3, x_4 \in \mathbf{k}$). Define a multiplication on $\mathbf{k}^4$ as follows:

$$\begin{aligned}
&(x_1, x_2, x_3, x_4)^T \cdot (y_1, y_2, y_3, y_4)^T \\
&= (x_1y_1 + ax_2y_2 + bx_3y_3 - abx_4y_4, x_1y_2 + x_2y_1 - bx_3y_4 + bx_4y_3, \\
&\qquad x_1y_3 + ax_2y_4 + x_3y_1 - ax_4y_2, x_1y_4 + x_2y_3 - x_3y_2 + x_4y_1)^T
\end{aligned} \tag{5}$$

for every $(x_1, x_2, x_3, x_4)^T, (y_1, y_2, y_3, y_4)^T \in \mathbf{k}^4$. Straightforward computations reveal that this multiplication is $\mathbf{k}$-bilinear[10] and associative, and that $(1, 0, 0, 0)^T \in \mathbf{k}^4$ is a neutral element of this multiplication. Thus, $\mathbf{k}^4$ becomes a $\mathbf{k}$-algebra with unity $(1, 0, 0, 0)^T$.

Set $\mathfrak{i} = (0, 1, 0, 0)^T$ and $\mathfrak{j} = (0, 0, 1, 0)^T$. Thus, $\mathfrak{i}$ and $\mathfrak{j}$ are two elements of the $\mathbf{k}$-algebra $\mathbf{k}^4$. They satisfy the relations (4) (this is checked by straightforward computation). Hence, Lemma 5.2 (applied to $A = \mathbf{k}^4$) shows that there exists

[9] It follows by observing that if $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = 0$, then $\lambda_1 f(v_1) + \lambda_2 f(v_2) + \cdots + \lambda_n f(v_n) = f\left(\underbrace{\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n}_{=0}\right) = f(0) = 0$.

[10] I.e., we have

$$\begin{aligned}
(u + v) w &= uw + vw && \text{for all } u, v, w \in \mathbf{k}^4; \\
(\lambda u) w &= \lambda (uw) && \text{for all } u, w \in \mathbf{k}^4 \text{ and } \lambda \in \mathbf{k}; \\
u (v + w) &= uv + uw && \text{for all } u, v, w \in \mathbf{k}^4; \\
u (\lambda w) &= \lambda (uw) && \text{for all } u, w \in \mathbf{k}^4 \text{ and } \lambda \in \mathbf{k}.
\end{aligned}$$

a unique $\mathbf{k}$-algebra homomorphism $\Phi : H_{a,b} \to \mathbf{k}^4$ satisfying $\Phi(i) = \mathfrak{i}$ and $\Phi(j) = \mathfrak{j}$. Consider this Φ.

Now, Φ is a $\mathbf{k}$-algebra homomorphism. Hence, it sends 1 to the unity of $\mathbf{k}^4$, which is $(1,0,0,0)^T$. In other words, $\Phi(1) = (1,0,0,0)^T$. Also, $\Phi(i) = \mathfrak{i} = (0,1,0,0)^T$ and $\Phi(j) = \mathfrak{j} = (0,0,1,0)^T$. Finally, since Φ is a $\mathbf{k}$-algebra homomorphism, we have

$$\Phi(ij) = \underbrace{\Phi(i)}_{=(0,1,0,0)^T} \underbrace{\Phi(j)}_{=(0,0,1,0)^T} = (0,1,0,0)^T (0,0,1,0)^T = (0,0,0,1)^T$$

(this results from another straightforward computation inside $\mathbf{k}^4$). Hence, the map Φ sends the elements $1, i, j, ij$ of $H_{a,b}$ to the elements $(1,0,0,0)^T, (0,1,0,0)^T, (0,0,1,0)^T, (0,0,0,1)^T$ of $\mathbf{k}^4$. The latter four elements are $\mathbf{k}$-linearly independent (since they form the standard basis of $\mathbf{k}^4$); thus, the former four elements are also $\mathbf{k}$-linearly independent (according to Lemma 5.1, applied to $P = H_{a,b}$, $Q = \mathbf{k}^4$, $f = \Phi$ and $(v_1, v_2, \ldots, v_n) = (1, i, j, ij)$). This proves Lemma 3.2. □

Let me make a few remarks about the proof just given.

1. Our definition of the multiplication on $\mathbf{k}^4$ was no stroke of genius; it was tailored to our goal in a straightforward way. In fact, our goal (to prove Lemma 3.2 and thus Theorem 1.2) was to prove that the $\mathbf{k}$-module $H_{a,b}$ has basis $(1, i, j, ij)$. If we assume (for a moment) that this holds, then it becomes possible to identify the elements of $H_{a,b}$ with vectors in $\mathbf{k}^4$ (namely, by identifying every element $x_1 + x_2 i + x_3 j + x_4 ij \in H_{a,b}$ with the vector $(x_1, x_2, x_3, x_4)^T \in \mathbf{k}^4$). This identification makes $\mathbf{k}^4$ into a $\mathbf{k}$-algebra, and the multiplication on this $\mathbf{k}$-algebra is given by (5) (because the multiplication on $H_{a,b}$ is given by

$$\begin{aligned}
&(x_1 + x_2 i + x_3 j + x_4 ij)(y_1 + y_2 i + y_3 j + y_4 ij) \\
&= (x_1 y_1 + a x_2 y_2 + b x_3 y_3 - ab x_4 y_4) + (x_1 y_2 + x_2 y_1 - b x_3 y_4 + b x_4 y_3) i \\
&\qquad + (x_1 y_3 + a x_2 y_4 + x_3 y_1 - a x_4 y_2) j + (x_1 y_4 + x_2 y_3 - x_3 y_2 + x_4 y_1) ij,
\end{aligned}$$

as a straightforward computation shows). Now, of course, this is "wishful thinking", because when we are proving the claim that the $\mathbf{k}$-module $H_{a,b}$ has basis $(1, i, j, ij)$, we cannot assume that very same claim to hold; thus, we were not able to identify the elements of $H_{a,b}$ with vectors in $\mathbf{k}^4$ in our proof. But we nevertheless were able to define the $\mathbf{k}$-algebra structure on $\mathbf{k}^4$ that would result from such an identification (because this structure is simply given by the explicit equation (5), which makes no reference to the alleged identification), and then we have used Lemma 5.2 to define a homomorphism Φ from $H_{a,b}$ to this $\mathbf{k}$-algebra $\mathbf{k}^4$. It is easy to see that this

Φ is a $\mathbf{k}$-algebra isomorphism; that said, all that we have actually used is the existence of Φ.

From this point of view, the above proof of Lemma 3.2 looks like a magic trick: In a sense, we have created the $\mathbf{k}$-algebra $\mathbf{k}^4$ out of thin air by requiring the multiplication to be the one we wanted $H_{a,b}$ to have. But it is a valid proof, as long as one actually does make all the straightforward verifications that I have left to the reader. The most important of these verifications is that of associativity; this is the point at which a failure of Lemma 3.2 (if this lemma were to fail) would have probably revealed itself. For example, if we tried to similarly prove the (false) statement that the $\mathbb{R}$-algebra S from Example 2.2 has basis $(1, i, j, ij)$, then we would have to prove that the multiplication on $\mathbb{R}^4$ defined by

$$\begin{aligned}
&\left(x_1, x_2, x_3, x_4\right)^T \cdot \left(y_1, y_2, y_3, y_4\right)^T \\
&= \Bigg(x_1 y_1 + x_2 y_2 + x_3 y_3 + \frac{1}{2} x_4 y_4, x_1 y_2 + x_2 y_1 + \frac{1}{2} x_3 y_4 + x_4 y_3, \\
&\qquad x_1 y_3 + x_2 y_4 + x_3 y_1 + \frac{1}{2} x_4 y_2, x_1 y_4 + x_2 y_3 + \frac{1}{2} x_3 y_2 + x_4 y_1 \Bigg)^T
\end{aligned}$$

for every $\left(x_1, x_2, x_3, x_4\right)^T, \left(y_1, y_2, y_3, y_4\right)^T \in \mathbb{R}^4$ is associative. But it is not (for example, we have $(uv)\, w \neq u\,(vw)$ for $u = (0,0,0,1)^T$, $v = (0,0,1,0)^T$ and $w = (0,0,0,1)^T$). This lack of associativity shows that S cannot have basis $(1, i, j, ij)$ (although it does not show that S is the trivial ring; that is a stronger statement).

2. In the proof of Lemma 3.2, we left to the reader the annoying chore of checking that the multiplication on $\mathbf{k}^4$ is associative.[11] This is a somewhat lengthy computation, and one might wonder whether it can be simplified. Indeed, it can; here is a way to reduce the amount of work necessary:

   Let (e_1, e_2, e_3, e_4) be the standard basis of the $\mathbf{k}$-module $\mathbf{k}^4$ (so each e_i is a vector whose i-th coordinate is 1 and whose all other coordinates are 0). We want to check the equality $(uv)\, w = u\,(vw)$ for any $u, v, w \in \mathbf{k}^4$. This equality is $\mathbf{k}$-linear in each of u, v, w (because the multiplication on $\mathbf{k}^4$ is $\mathbf{k}$-bilinear[12]). Thus, we can WLOG assume that each of u, v, w belongs to the basis (e_1, e_2, e_3, e_4) of $\mathbf{k}^4$. This leaves us 4 possibilities for each of u, v, w, and therefore $4^3 = 64$ possibilities for the triple (u, v, w); we have to check the equality $(uv)\, w = u\,(vw)$ for each of these 64 possible triples. To simplify our life, we can create a multiplication table for the basis (e_1, e_2, e_3, e_4). To simplify our life yet further, we can notice that the equality $(uv)\, w = u\,(vw)$ is obviously true when at least one of u, v, w equals e_1 (because e_1 is the

[11] We also left some other chores to the reader, but those are far easier.

[12] This is clear from a look at its definition.

neutral element of the multiplication on $\mathbf{k}^4$ [13]). Thus, we can WLOG assume that none of u, v, w equals e_1; this leaves only 3 possibilities for each of u, v, w, and therefore only $3^3 = 27$ possibilities for the triple (u, v, w). Still, this argument requires some manual labor. We will see a nicer proof in the next section.

6. Linear independency: proof by representation

Now, we shall see another proof of Lemma 3.2, which is less straightforward but also requires less computation. The general philosophy behind this proof is that algebras are often best understood by studying their representations; more specifically, if we want to prove that some elements of a $\mathbf{k}$-algebra A are linearly independent, it helps to show that their actions on some representation of A are linearly independent (as endomorphisms of this representation). We shall thus prove Lemma 3.2 by constructing a representation of $H_{a,b}$ such that the actions of $1, i, j, ij$ on this representation are linearly independent. For the sake of simplicity, we shall use a matrix representation, i.e., a $\mathbf{k}$-algebra homomorphism from $H_{a,b}$ into a matrix ring; so the actions of $1, i, j, ij$ will be matrices.

Second proof of Lemma 3.2. Consider the $\mathbf{k}$-algebra $\mathbf{k}^{4\times 4}$ of all 4×4-matrices over $\mathbf{k}$. Let I_4 denote the identity matrix in $\mathbf{k}^{4\times 4}$; this is the unity of the $\mathbf{k}$-algebra $\mathbf{k}^{4\times 4}$. Define two elements $\mathfrak{i}$ and $\mathfrak{j}$ of $\mathbf{k}^{4\times 4}$ by

$$\mathfrak{i} = \begin{pmatrix} 0 & a & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & a \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad \text{and} \qquad \mathfrak{j} = \begin{pmatrix} 0 & 0 & b & 0 \\ 0 & 0 & 0 & -b \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}. \tag{6}$$

Then, straightforward computations show that $\mathfrak{i}^2 = aI_4$, $\mathfrak{j}^2 = bI_4$ and $\mathfrak{ij} = -\mathfrak{ji}$. Hence, Lemma 5.2 (applied to $A = \mathbf{k}^{4\times 4}$) shows that there exists a unique $\mathbf{k}$-algebra homomorphism $\Phi : H_{a,b} \to \mathbf{k}^{4\times 4}$ satisfying $\Phi(i) = \mathfrak{i}$ and $\Phi(j) = \mathfrak{j}$. Consider this Φ.

Now, Φ is a $\mathbf{k}$-algebra homomorphism. Hence, it sends 1 to the unity of $\mathbf{k}^{4\times 4}$, which is I_4. In other words, $\Phi(1) = I_4$. Also, $\Phi(i) = \mathfrak{i}$ and $\Phi(j) = \mathfrak{j}$. Finally, since Φ is a $\mathbf{k}$-algebra homomorphism, we have $\Phi(ij) = \underbrace{\Phi(i)}_{=\mathfrak{i}}\underbrace{\Phi(j)}_{=\mathfrak{j}} = \mathfrak{ij}$. Hence, the map Φ sends the elements $1, i, j, ij$ of $H_{a,b}$ to the elements $I_4, \mathfrak{i}, \mathfrak{j}, \mathfrak{ij}$ of $\mathbf{k}^{4\times 4}$.

[13] This, of course, has to be checked, but we need to check this anyway, and it is quite easy to check.

The latter four elements are $\mathbf{k}$-linearly independent[14]. Hence, the former four elements are also $\mathbf{k}$-linearly independent (according to Lemma 5.1, applied to $P = H_{a,b}$, $Q = \mathbf{k}^{4\times 4}$, $f = \Phi$ and $(v_1, v_2, \ldots, v_n) = (1, i, j, ij)$). This proves Lemma 3.2. □

Let me discuss this proof a little bit:

1. It is similar to the first proof of Lemma 3.2, but it differs in one important fact: Instead of constructing a $\mathbf{k}$-algebra structure on $\mathbf{k}^4$ (as we did in the first proof), we have now used an already existing $\mathbf{k}$-algebra structure on $\mathbf{k}^{4\times 4}$ (namely, the one given by matrix multiplication). This has the advantage that we did not have to prove associativity. That said, we still had to make some computations (in order to prove $\mathfrak{i}^2 = aI_4$, $\mathfrak{j}^2 = bI_4$ and $\mathfrak{ij} = -\mathfrak{ji}$); but these are much shorter than what was needed in the first proof.

2. You might wonder how I have guessed the correct matrices $\mathfrak{i}$ and $\mathfrak{j}$ in (6) which make the proof work. The answer is that I have used a similar kind of "wishful thinking" that helped me guess the multiplication rule (5) in the first proof. Namely, if we assume for a moment that Theorem 1.2 is proven, then it becomes possible to identify the elements of $H_{a,b}$ with vectors in $\mathbf{k}^4$ (namely, by identifying every element $x_1 + x_2 i + x_3 j + x_4 ij \in H_{a,b}$ with the vector $(x_1, x_2, x_3, x_4)^T \in \mathbf{k}^4$), and consequently the $\mathbf{k}$-algebra $H_{a,b}$ acts on the $\mathbf{k}$-module $\mathbf{k}^4$. The matrices $\mathfrak{i}$ and $\mathfrak{j}$ defined in (6) are precisely the actions of i and j on this module (written as matrices). Now, of course, we can define these two matrices explicitly (as in (6)) even if we do **not** assume that Theorem 1.2 is proven; so the proof does not

[14]This can be checked as follows: We have

$$I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \mathfrak{i} = \begin{pmatrix} 0 & a & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & a \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$\mathfrak{j} = \begin{pmatrix} 0 & 0 & b & 0 \\ 0 & 0 & 0 & -b \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \qquad \mathfrak{ij} = \begin{pmatrix} 0 & 0 & 0 & -ab \\ 0 & 0 & b & 0 \\ 0 & -a & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Hence, if $\alpha, \beta, \gamma, \delta \in \mathbf{k}$ are scalars, then the first column of the matrix $\alpha I_4 + \beta\mathfrak{i} + \gamma\mathfrak{j} + \delta\mathfrak{ij}$ is $\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$. Consequently, if $\alpha, \beta, \gamma, \delta \in \mathbf{k}$ are scalars satisfying $\alpha I_4 + \beta\mathfrak{i} + \gamma\mathfrak{j} + \delta\mathfrak{ij} = 0$, then $\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = 0$. In other words, the matrices $I_4, \mathfrak{i}, \mathfrak{j}, \mathfrak{ij}$ are linearly independent, qed.

actually rely on our wishful thinking, but the wishful thinking was crucial in finding it.

7. to be continued

[To be continued, eventually:

- proof by base change to an extension where $ax^2 + by^2 = 1$ is solvable (works only under certain circumstances).
- proof using the Clifford-PBW theorem (basis theorem for Clifford algebra of quadratic form).
- proof of Clifford-PBW theorem.
- reference to diamond lemma.
- conclusion, and $\mathbf{k} \to H_{a,b}$ is an injection.
- proof using Cayley-Dickson process.

]

References

[Conrad15] Keith Conrad, *Quaternion algebras*, version 10 May 2015.

[Bergma78] George M. Bergman, *The diamond lemma for ring theory*, Advances in Mathematics, Volume 29, Issue 2, February 1978, pp. 178–218, doi:10.1016/0001-8708(78)90010-5.
See also the list of errata on Bergman's website.

[SheWit14] Anne V. Shepler, Sarah Witherspoon, Poincaré-Birkhoff-Witt Theorems, arXiv:1404.6497v1.