THE HURWITZ THEOREM ON SUMS OF SQUARES BY LINEAR ALGEBRA

KEITH CONRAD

1. Introduction

From commutativity of multiplication (for numbers), a product of two squares is a square: $x^2y^2 = (xy)^2$. A more interesting identity is the following one, which expresses a sum of two squares times a sum of two squares as another sum of two squares:

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 - x_2y_2)^2 + (x_1y_2 + x_2y_1)^2. \tag{1.1}$$

There is also an identity like this for a sum of four squares:

$$\begin{aligned}(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) \;=\;& (x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4)^2 + \\ & (x_1y_2 + x_2y_1 + x_3y_4 - x_4y_3)^2 + \\ & (x_1y_3 + x_3y_1 - x_2y_4 + x_4y_2)^2 + \\ & (x_1y_4 + x_4y_1 + x_2y_3 - x_3y_2)^2.\end{aligned} \tag{1.2}$$

This was discovered by Euler in the 18th century, forgotten, and then rediscovered in the 19th century by Hamilton in his work on quaternions. Shortly after Hamilton's rediscovery of (1.2) Cayley discovered a similar 8-square identity.

In all of these sum-of-squares identities, the terms being squared on the right side are all bilinear expressions in the x's and y's: each such expression, like $x_1y_2+x_2y_1$ for sums of two squares, is a linear combination of the x's when the y's are fixed and a linear combination of the y's when the x's are fixed.

It was natural for mathematicians to search for a similar 16-square identity next, but they were unsuccessful. At the end of the 19th century Hurwitz [3] proved his famous "1,2,4,8 theorem," which says that further identities of this kind are *impossible.*

Theorem 1.1 (Hurwitz, 1898)**.** *If there is an identity*

$$(x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) = z_1^2 + \cdots + z_n^2 \tag{1.3}$$

for $x_1, \dots, x_n, y_1, \dots, y_n$ in $\mathbf{C}$, where each z_k is a $\mathbf{C}$-bilinear function of the x's and the y's, then $n = 1, 2, 4$ or 8.

To prove Theorem 1.1, we first show in Section 2 that the existence of a bilinear formula like (1.3) leads to a set of equations in $n \times n$ matrices over $\mathbf{C}$. Then we show that the matrix equations can be solved only when $n = 1, 2, 4$, or 8. The method we use, based on [5], depends on a linear independence property of certain matrix products we will see in Section 3 and is a simplified version of Hurwitz's original argument. As an application of Hurwitz's theorem, in Section 4 we see which Euclidean spaces $\mathbf{R}^n$ can admit a multiplication resembling the usual vector cross product on $\mathbf{R}^3$.

2. The Hurwitz Matrix Equations

Lemma 2.1. *Let V be a finite-dimensional vector space over $\mathbf{C}$. If there is a pair of invertible anti-commuting linear operators on V, then $\dim V$ is even.*

Proof. Suppose $L, L' \colon V \to V$ are linear, invertible, and $LL' = -L'L$. Taking the determinant of both sides, $(\det L)(\det L') = (-1)^{\dim V}(\det L')(\det L)$. Since L and L' have nonzero determinants, $1 = (-1)^{\dim V}$, so $\dim V$ is even. □

We return to (1.3). That z_k is a bilinear functions of the x's and y's means

$$z_k = \sum_{i,j=1}^{n} a_{ijk} x_i y_j \tag{2.1}$$

for some $a_{ijk} \in \mathbf{C}$. For example, in the case $n = 2$ we see by (1.1) that we can use

$$z_1 = x_1y_1 - x_2y_2, \quad z_2 = x_1y_2 + x_2y_1. \tag{2.2}$$

We can collect the two equations in (2.2) as components of the equation

$$\begin{aligned} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} &= \begin{pmatrix} x_1y_1 - x_2y_2 \\ x_1y_2 + x_2y_1 \end{pmatrix} \\ &= \begin{pmatrix} x_1 & -x_2 \\ x_2 & x_1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \left(x_1 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + x_2 \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}. \end{aligned}$$

From (1.2), in the $n = 4$ case we can use

$$\begin{aligned} z_1 &= x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4, \\ z_2 &= x_1y_2 + x_2y_1 + x_3y_4 - x_4y_3, \\ z_3 &= x_1y_3 + x_3y_1 - x_2y_4 + x_4y_2, \\ z_4 &= x_1y_4 + x_4y_1 + x_2y_3 - x_3y_2, \end{aligned}$$

so

$$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} = (x_1A_1 + x_2A_2 + x_3A_3 + x_4A_4) \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix},$$

where A_1, A_2, A_3, and A_4 are 4×4 matrices with entries 0, 1, and -1. For example,

$$A_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The reader can work out A_3 and A_4.

Such matrix equations can be developed in the $n \times n$ case too. The scalar equation (2.1) for $k = 1, \dots, n$ is the same as the single equation

$$\begin{aligned}
\begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} &= \begin{pmatrix} \sum_{i,j} a_{ij1} x_i y_j \\ \vdots \\ \sum_{i,j} a_{ijn} x_i y_j \end{pmatrix} \\
&= \begin{pmatrix} \sum_j \left(\sum_i a_{ij1} x_i\right) y_j \\ \vdots \\ \sum_j \left(\sum_i a_{ijn} x_i\right) y_j \end{pmatrix} \\
&= \begin{pmatrix} \sum_i a_{i11} x_i & \dots & \sum_i a_{in1} x_i \\ \vdots & \ddots & \vdots \\ \sum_i a_{i1n} x_i & \dots & \sum_i a_{inn} x_i \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.
\end{aligned} \tag{2.3}$$

The $n \times n$ matrix in the last expression can be expressed as a sum of n matrices, each one containing only one x_i which can then be pulled out as a coefficient:

$$x_1 \begin{pmatrix} a_{111} & \dots & a_{1n1} \\ \vdots & \ddots & \vdots \\ a_{11n} & \dots & a_{1nn} \end{pmatrix} + \dots + x_n \begin{pmatrix} a_{n11} & \dots & a_{nn1} \\ \vdots & \ddots & \vdots \\ a_{n1n} & \dots & a_{nnn} \end{pmatrix}.$$

This sum can be written as $x_1 A_1 + \dots + x_n A_n$, where A_i is an $n \times n$ matrix with (j,k)-entry a_{ikj}. (Why the index reversal on the subscripts? That is in the nature of how matrix-vector multiplication works: look at the $n = 2$ case to convince yourself in a concrete case that this index reversal is not an error.) Now (2.3) reads as

$$\mathbf{z} = (x_1 A_1 + \dots + x_n A_n)\mathbf{y} = A_{\mathbf{x}}\mathbf{y},$$

where we set $A_{\mathbf{x}} = x_1 A_1 + \dots + x_n A_n$.

The right side of (1.3) is

$$z_1^2 + \dots + z_n^2 = \mathbf{z} \cdot \mathbf{z},$$

where $\cdot$ is the dot product $(a_1, \dots, a_n) \cdot (b_1, \dots, b_n) = \sum_{k=1}^n a_k b_k$ on $\mathbf{C}^n$ (not the Hermitian inner product!). Thus

$$\begin{aligned}
z_1^2 + \dots + z_n^2 &= \mathbf{z} \cdot \mathbf{z} \\
&= A_{\mathbf{x}}\mathbf{y} \cdot A_{\mathbf{x}}\mathbf{y} \\
&= (A_{\mathbf{x}}^\top A_{\mathbf{x}}\mathbf{y}) \cdot \mathbf{y}
\end{aligned}$$

The left side of (1.3) is

$$\left(\sum x_i^2\right) \mathbf{y} \cdot \mathbf{y} = \left(\left(\sum x_i^2\right) \mathbf{y}\right) \cdot \mathbf{y}.$$

Therefore

$$(A_{\mathbf{x}}^\top A_{\mathbf{x}}\mathbf{y}) \cdot \mathbf{y} = \left(\left(\sum x_i^2\right) \mathbf{y}\right) \cdot \mathbf{y}.$$

Comparing the two sides as $\mathbf{y}$ varies shows

$$A_{\mathbf{x}}^\top A_{\mathbf{x}} = \left(\sum x_i^2\right) I_n. \tag{2.4}$$

Expanding the left side of (2.4) using $A_{\mathbf{x}} = x_1A_1 + \cdots + x_nA_n$, we have

$$A_{\mathbf{x}}^\top A_{\mathbf{x}} = \sum_{i=1}^{n} \left(A_i^\top A_i\right) x_i^2 + \sum_{\substack{1\leq i,j\leq n\\ i\neq j}} \left(A_i^\top A_j + A_j^\top A_i\right) x_ix_j,$$

so (2.4) is equivalent to the system of matrix equations

$$A_i^\top A_i = I_n, \quad A_i^\top A_j + A_j^\top A_i = O \text{ for } i \neq j. \tag{2.5}$$

These are the *Hurwitz matrix equations*. The rest of the proof of Theorem 1.1 is now devoted to showing these equations in $n \times n$ matrices can exist only if n is 1, 2, 4, or 8. Without loss of generality we take $n > 2$.

We normalize the matrices A_i to make one of them the identity, as follows. By (2.5), A_i is an invertible matrix whose inverse is $A_i^\top$. Set

$$B_i = A_iA_n^\top.$$

Now (2.5) is easily seen to be equivalent to

$$B_n = I_n, \quad B_i^\top B_i = I_n, \quad B_i^\top B_j + B_j^\top B_i = O \text{ for } i \neq j. \tag{2.6}$$

Taking $j = n$ in the third equation shows $B_i^\top = -B_i$ for $i \neq n$. Therefore the $n-1$ matrices $B_1, \ldots, B_{n-1}$ satisfy

$$B_i^\top = -B_i, \quad B_i^2 = -I_n, \quad B_iB_j = -B_jB_i \text{ for } i \neq j. \tag{2.7}$$

We see immediately from (2.7) and Lemma 2.1 that n is *even*. Next we will prove that (2.7) for even $n > 2$ forces $n = 4$ or 8.

Up to this point, although our scalars have been taken to be complex it is worth noting that we have not used anything special about complex numbers. For example, if we had a sum of squares identity (1.3) for real x's and y's and the z's are real bilinear functions of the x's and y's then the exact same argument as above would yield a set of Hurwitz matrix equations with real matrices. Nothing has to be changed in the derivation at all. (For example, Lemma 2.1 is valid for with real scalars in place of complex scalars using the same proof.)

3. Conclusion via Linear Algebra

We will use a lemma about linear independence of certain matrix products. Let m be a positive *even* integer and $C_1, \ldots, C_m$ be matrices in some $\mathrm{M}_d(\mathbf{C})$ which are pairwise anticommuting and each C_i^2 is a nonzero scalar diagonal matrix. (For instance, in the notation of (2.7), we can use $B_1, B_2, \ldots, B_{n-2}$ in $\mathrm{M}_n(\mathbf{C})$. We take out $B_n = I_n$ since it is not anti-commuting with the other B_i's, and we then take out B_{n-1} because we need an even number of anti-commuting matrices and $n-1$ is odd.) While $B_i^2 = -I_n$ for all i, for the purpose of what we are going to do for now with these C's, we don't need to assume C_i^2 is the *same* scalar for all i.) From the m matrices $C_1, \ldots, C_m$, we get 2^m products of different terms. Specifically, for an m-tuple $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m) \in \{0,1\}^m$, set

$$C^{\boldsymbol{\delta}} = C_1^{\delta_1} \cdots C_m^{\delta_m}.$$

Note C_i is $C^{\boldsymbol{\delta}}$ where $\delta_i = 1$ and other δ_j's are 0. The number of different $\boldsymbol{\delta}$'s is 2^m.

Lemma 3.1. *With notation as in the previous paragraph, the* 2^m *matrices* $C^{\boldsymbol{\delta}}$ *are linearly independent in* $\mathrm{M}_d(\mathbf{C})$*. In particular,* $2^m \leq d^2$ *when* m *is even.*

In the course of the proof, the condition that m is even will only be needed at the very end.

Proof. This proof is a bit tedious, so the reader may want to skip the proof on a first reading to see how it gets used and then return here later. Suppose there is a non-trivial linear relation

$$\sum_{\boldsymbol{\delta}} b_{\boldsymbol{\delta}} C^{\boldsymbol{\delta}} = O, \tag{3.1}$$

where the $b_{\boldsymbol{\delta}}$'s are in $\mathbf{C}$ and are not all 0. Take such a relation with as few nonzero coefficients as possible.

First we show that we can assume $b_{\mathbf{0}} \neq 0$. Since the C_i's anti-commute and square to a nonzero scalar matrix, $C^{\boldsymbol{\delta}'}C^{\boldsymbol{\delta}'}$ is a nonzero scalar matrix for any $\boldsymbol{\delta}'$. Moreover, as $\boldsymbol{\delta}$ varies and $\boldsymbol{\delta}'$ is fixed,

$$\{C^{\boldsymbol{\delta}}C^{\boldsymbol{\delta}'} : \boldsymbol{\delta} \in \{0,1\}^m\} = \{(\text{nonzero scalar})C^{\boldsymbol{\delta}} : \boldsymbol{\delta} \in \{0,1\}^m\}.$$

Therefore, picking $\boldsymbol{\delta}'$ such that $b_{\boldsymbol{\delta}'} \neq 0$, multiplying (3.1) on the right by $C^{\boldsymbol{\delta}'}$ gives a linear relation with the same number of nonzero coefficients as in (3.1) but now the coefficient of $C^{\mathbf{0}} = I_d$ is nonzero. We may henceforth impose this condition on the minimal relation (3.1).

Now we use conjugations to show most terms in (3.1) are zero. By anti-commutativity,

$$C_i C_j C_i^{-1} = \begin{cases} C_j, & \text{if } i = j, \\ -C_j, & \text{if } i \neq j. \end{cases}$$

Therefore

$$C_i C^{\boldsymbol{\delta}} C_i^{-1} = \pm C^{\boldsymbol{\delta}}. \tag{3.2}$$

What is the exact recipe for the ± sign? It depends on how many coordinates in $\boldsymbol{\delta}$ equal 1. For $\boldsymbol{\delta} \in \{0,1\}^m$, let its *weight* $w(\boldsymbol{\delta})$ be the number of i's with $\delta_i = 1$. For instance, $w(\mathbf{0}) = 0$. We get the more precise version of (3.2):

$$C_i C^{\boldsymbol{\delta}} C_i^{-1} = \varepsilon_{\boldsymbol{\delta},i} C^{\boldsymbol{\delta}}, \tag{3.3}$$

where

$$\varepsilon_{\boldsymbol{\delta},i} = \begin{cases} (-1)^{w(\boldsymbol{\delta})}, & \text{if } \delta_i = 0, \\ (-1)^{w(\boldsymbol{\delta})-1}, & \text{if } \delta_i = 1. \end{cases} \tag{3.4}$$

For instance, $\varepsilon_{\mathbf{0},i} = 1$ for all i.

Pick i from 1 to n and conjugate (3.1) by C_i. By (3.3), we get

$$\sum_{\boldsymbol{\delta}} \varepsilon_{\boldsymbol{\delta},i} b_{\boldsymbol{\delta}} C^{\boldsymbol{\delta}} = O. \tag{3.5}$$

Since $\varepsilon_{\mathbf{0},i} = 1$, subtract (3.5) from (3.1) to get the linear relation

$$\sum_{\boldsymbol{\delta}} (1 - \varepsilon_{\boldsymbol{\delta},i}) b_{\boldsymbol{\delta}} C^{\boldsymbol{\delta}} = O. \tag{3.6}$$

Here the coefficient of the term for $\boldsymbol{\delta} = \mathbf{0}$ is 0, while we arranged for it to be nonzero in (3.1). Therefore (3.6) is a linear relation with fewer nonzero terms than the nonzero relation of minimal length. Hence all terms in (3.6) vanish. That is,

$$\boldsymbol{\delta} \neq \mathbf{0}, b_{\boldsymbol{\delta}} \neq 0 \Longrightarrow \varepsilon_{\boldsymbol{\delta},i} = 1.$$

This holds for every i from 1 to n, so each $\boldsymbol{\delta} \neq \mathbf{0}$ with a nonzero coefficient in (3.1) has $\varepsilon_{\boldsymbol{\delta},i}$ independent of i. Then δ_i is independent of i by (3.4), so $\boldsymbol{\delta} = (1, 1, \ldots, 1)$. Then $w(\boldsymbol{\delta}) = m$, so $\varepsilon_{\boldsymbol{\delta},i} = (-1)^{m-1} = -1$, since m is *even*. This is a contradiction since $-1 \neq 1$. We have shown $b_{\boldsymbol{\delta}} = 0$ for $\boldsymbol{\delta} \neq \mathbf{0}$, but then the linear relation (3.1) has just one nonzero term, which is impossible. □

Remark 3.2. Nothing is special about complex scalars for the matrices in this lemma. The proof of Lemma 3.1 works for matrices in $\mathrm{M}_d(F)$ for any field F in which $-1 \neq 1$.

Returning now to the setting of the proof of Theorem 1.1, apply Lemma 3.1 to the matrices $B_1, \ldots, B_{n-2}$. (Recall n is even.) We conclude $2^{n-2} \leq n^2$. It is easy to see this inequality, for even $n > 2$, holds only for $n = 4$, 6, and 8. The possibility $n = 6$ in Theorem 1.1 will be eliminated by studying eigenspaces for B_1. We will see that when $n > 4$, $n/2$ is even, so $n \neq 6$.

Consider the B_j's as linear operators on $\mathbf{C}^n$. Since $B_j^2 = -I_n$, the only eigenvalues of B_j are $\pm i$. Let U and W be the two eigenspaces of B_1:

$$U = \{v : B_1 v = iv\}, \quad W = \{v : B_1 v = -iv\}.$$

For any $v \in \mathbf{C}^n$, the decomposition

$$v = \frac{v - iB_1 v}{2} + \frac{v + iB_1 v}{2}$$

has the first term in U and the second term in W. (More generally, if B is an $n \times n$ matrix and $B^2 = c^2 I_n$ then $(1/2)(v + (1/c)Bv)$ is in the c-eigenspace of B for any $v \in \mathbf{C}^n$. Take $c = i$ to get the displayed terms above.) Thus $\mathbf{C}^n = U + W$. Since U and W are eigenspaces of B_1 for different eigenvalues, $U \cap W = \{0\}$. Therefore $\mathbf{C}^n = U \oplus W$, so $n = \dim U + \dim W$.

Our goal is to show the eigenspaces U and W have the same dimension. Since B_j is invertible for every j, $\dim U = \dim B_j(U)$ and $\dim W = \dim B_j(W)$. Easily $B_1(U) \subset U$ and $B_1(W) \subset W$. Of greater interest is that, for $j = 2, 3, \ldots, n-1$, $B_j(U) \subset W$ and $B_j(W) \subset U$. To show this, pick $u \in U$. Then by anticommutativity,

$$B_1(B_j u) = -B_j(B_1 u) = -B_j(iu) = -iB_j u,$$

so $B_j u \in W$. Thus, $B_j(U) \subset W$. That $B_j(W) \subset U$ is analogous. Then

$$\dim U = \dim B_j(U) \leq \dim W \text{ and } \dim W = \dim B_j(W) \leq \dim U,$$

so $\dim U = \dim W$. Therefore from the decomposition $\mathbf{C}^n = U \oplus W$ we get $\dim U = \dim W = n/2$.

Although the maps B_j $(j > 1)$ send U to W and *vice versa*, we can get self-maps on one of these subspaces by composition of each B_j with, say, B_2. For $j = 2, 3, \ldots, n-1$, the composite $L_j = B_2 \circ B_j$ is an invertible linear operator on $\mathbf{C}^n$ and $L_j(U) = B_2(B_j(U)) \subset B_2(W) \subset U$, so $L_j(U) \subset U$. When $n > 4$, a calculation using (2.7) shows that L_3 and L_4 are anti-commuting on $\mathbf{C}^n$:

$$L_3L_4 = (B_2B_3)(B_2B_4) = -B_2^2B_3B_4 = B_3B_4$$

and

$$L_4L_3 = (B_2B_4)(B_2B_3) = -B_2^2B_4B_3 = B_4B_3 = -B_3B_4 = -L_3L_4.$$

(More generally, L_j and L_k are anti-commuting for any different $j, k > 2$.) Viewing L_3 and L_4 as linear operators not on $\mathbf{C}^n$ but on the vector space U [1], their anticommutativity on U forces $\dim U = n/2$ to be even by Lemma 2.1, so n must be a multiple of 4. This eliminates the choice $n = 6$ and concludes the proof of Hurwitz's theorem!

We stated and proved Hurwitz's theorem over $\mathbf{C}$, but the theorem goes through with variables in $\mathbf{R}$ too. After all, if we have an n-term bilinear sum of squares identity over $\mathbf{R}$ then it leads to $n \times n$ real matrices satisfying the Hurwitz matrix equations by the same reasoning as above. By viewing these matrices as acting on $\mathbf{C}^n$ and running through the above eigenspace argument we obtain $n = 1$, 2, 4, or 8. (By similar reasoning, the proof goes through with $\mathbf{C}$ replaced by any field F in which $-1 \neq 1$; if necessary we may have to enlarge the field as we do when passing from $\mathbf{R}$ to $\mathbf{C}$ to contain square roots of -1 if they are not already in F, in order for the eigenspace argument to make sense.)

While Hurwitz proved only the dimension constraints $n = 1$, 2, 4, and 8 for an n-dimensional bilinear sum of squares identity, something stronger is true: up to a linear change of variables there is only one bilinear sum of squares identity in each of these dimensions. For a proof, see [1], [4, §7.6], or [5, Appendix, Chap. 1]. (Such identities are associated to multiplication in the real numbers, complex numbers, quaternions, and octonions. Readers unfamiliar with the quaternions and octonions can look in [2].)

4. Vector products

We will use Hurwitz's theorem to explore the following question: does the cross product on $\mathbf{R}^3$ have an analogue on $\mathbf{R}^n$ for any $n > 3$? After we specify what properties we want such a product to satisfy, we will see the choices are quite limited.

The multiplication on $\mathbf{R}^n$ should assign to any v and w in $\mathbf{R}^n$ a third vector in $\mathbf{R}^n$, to be denoted $v \times w$. It is natural to insist that this product be $\mathbf{R}$-bilinear in v and w:

$$(4.1) \qquad (v_1 + v_2) \times w = v_1 \times w + v_2 \times w, \quad v \times (w_1 + w_2) = v \times w_1 + v \times w_2,$$

and

$$(4.2) \qquad (cv) \times w = c(v \times w), \quad v \times (cw) = c(v \times w),$$

where $c \in \mathbf{R}$. One consequence of bilinearity is that multiplication by $\mathbf{0}$ is $\mathbf{0}$:

$$(4.3) \qquad v \times \mathbf{0} = \mathbf{0}, \quad \mathbf{0} \times w = \mathbf{0}.$$

[1] At this step we need a theory of linear operators on general vector spaces and not just on the "column spaces" $\mathbf{C}^d$

Let us also ask that, as with the cross product on $\mathbf{R}^3$, the product be orthogonal to both factors: for all v and w in $\mathbf{R}^n$,

$$v \cdot (v \times w) = 0, \quad w \cdot (v \times w) = 0. \tag{4.4}$$

(This property is not satisfied by other kinds of products in linear algebra, such as matrix multiplication on $\mathrm{M}_d(\mathbf{R}) = \mathbf{R}^{d^2}$ with $\cdot$ being the dot product on $\mathrm{M}_d(\mathbf{R})$ given by $(a_{ij})\cdot(b_{ij}) = \sum_{i,j} a_{ij}b_{ij}$.)

Lastly, we ask that the magnitude $||v \times w||$ be determined by the same formula which works for the cross product in three dimensions:

$$||v \times w||^2 = ||v||^2||w||^2 - (v \cdot w)^2. \tag{4.5}$$

When $n = 1$, a product on $\mathbf{R}^n = \mathbf{R}$ satisfying (4.5) must be identically zero. Indeed, the dot product on $\mathbf{R}$ is the ordinary product, so (4.5) becomes $|x \times y|^2 = x^2y^2 - (xy)^2 = 0$, so $x \times y = 0$. So we only care about the case $n > 1$.

The assumption (4.5) looks more complicated than the earlier assumptions. The following result expresses (4.5) in simpler terms, but it is in the form (4.5) that we will actually use the assumption.

Theorem 4.1. *Let $\times$ be a product on $\mathbf{R}^n$ which satisfies (4.1), (4.2), and (4.4). Then (4.5) is equivalent to the following two conditions together:*

1. *for all $v \in \mathbf{R}^n$, $v \times v = \mathbf{0}$,*
2. *if $||v|| = 1, ||w|| = 1$, and $v \perp w$, then $||v \times w|| = 1$.*

Proof. It is easy to see that (4.5) implies the two conditions in the theorem. Now we assume the two conditions and derive (4.5).

First suppose v and w are linearly dependent, say $w = cv$ for some $c \in \mathbf{R}$. Then

$$||v \times w||^2 = ||v \times (cv)||^2 = c^2||v \times v||^2 = 0$$

and

$$||v||^2||w||^2 - (v \cdot w)^2 = c^2||v||^4 - c^2(v \cdot v)^2 = c^2||v||^4 - c^2||v||^4 = 0,$$

so the two sides of (4.5) both equal 0.

Now suppose v and w are linearly independent. Let $u = v - \frac{v\cdot w}{w\cdot w}w$, so $u \cdot w = 0$. Then $u/||u||$ and $w/||w||$ are perpendicular unit vectors, so by assumption the product $(u/||u||) \times (w/||w||)$ is a unit vector. By bilinearity, the unit length of this product implies

$$||u \times w|| = ||u||||w||. \tag{4.6}$$

Since $w \times w = \mathbf{0}$, $u \times w = v \times w$ by bilinearity and (4.3). Squaring both sides of (4.6),

$$||v \times w||^2 = ||u||^2||w||^2 \tag{4.7}$$

From the definition of u,

$$\begin{aligned}
||u||^2 &= u \cdot u \\
&= \left(v - \frac{v \cdot w}{w \cdot w} w\right) \cdot \left(v - \frac{v \cdot w}{w \cdot w} w\right) \\
&= v \cdot v - 2\frac{(v \cdot w)^2}{w \cdot w} + \frac{(v \cdot w)^2}{w \cdot w} \\
&= v \cdot v - \frac{(v \cdot w)^2}{w \cdot w} \\
&= ||v||^2 - \frac{(v \cdot w)^2}{||w||^2}.
\end{aligned}$$

Substituting this into (4.7) gives

$$||v \times w||^2 = ||v||^2 ||w||^2 - (v \cdot w)^2.$$

□

Theorem 4.2. *For $n \geq 1$, assume there is a multiplication $\times \colon \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ satisfying (4.1), (4.2), (4.4), and (4.5). Then $n = 1$, 3, or 7.*

We have seen the $n = 1$ case is quite dull, so the only interesting cases in Theorem 4.2 are 3 and 7.

Proof. We use the multiplication $\times$ on $\mathbf{R}^n$ to define a product, say $\odot$, on $\mathbf{R}^{n+1}$. Write vectors in $\mathbf{R}^{n+1}$ in the form (x, v), where $x \in \mathbf{R}$ and $v \in \mathbf{R}^n$. Note that the dot product of such vectors can be expressed in terms of dot products of the components:

$$(x, v) \cdot (y, w) = xy + v \cdot w.$$

For (x, v) and (y, w) in $\mathbf{R}^{n+1}$, define

$$(x, v) \odot (y, w) = (xy - v \cdot w, xw + yv + v \times w). \tag{4.8}$$

This formula makes sense (even if it seems a bit mysterious) since $xy - v \cdot w \in \mathbf{R}$ and $xw + yv + v \times w \in \mathbf{R}^n$. While $(1, \mathbf{0})$ is a 2-sided identity for $\odot$, we won't be using this explicitly.

This product $\odot$ on $\mathbf{R}^{n+1}$ has two key properties. The first is that it is a bilinear function of (x, v) and (y, w). That is, fixing one of these vector pairs in $\mathbf{R}^{n+1}$, the right side of (4.8) is a linear function of the other pair.

The second key property of $\odot$ is that it is multiplicative for lengths:

$$||(x, v) \odot (y, w)||^2 = ||(x, v)||^2 ||(y, w)||^2. \tag{4.9}$$

We verify this by writing the left side as a dot product and expanding:

$$\begin{aligned}
||(x, v) \odot (y, w)||^2 &= (xy - v \cdot w, xw + yv + v \times w) \cdot (xy - v \cdot w, xw + yv + v \times w) \\
&= (xy - v \cdot w)^2 + (xw + yv + v \times w) \cdot (xw + yv + v \times w)
\end{aligned}$$

By (4.4), $v \times w$ is orthogonal to $xw + yv$. Therefore $(xw + yv + v \times w) \cdot (xw + yv + v \times w)$ equals

$$(xw + yv) \cdot (xw + yv) + (v \times w) \cdot (v \times w) = x^2 ||w||^2 + 2xy(v \cdot w) + y^2 ||v||^2 + ||v \times w||^2.$$

Adding this to $(xy - v \cdot w)^2 = x^2y^2 - 2xy(v \cdot w) + (v \cdot w)^2$ gives

$$||(x,v) \odot (y,w)||^2 = x^2y^2 + (v \cdot w)^2 + x^2||w||^2 + y^2||v||^2 + ||v \times w||^2.$$

By (4.5), this simplifies to

$$\begin{aligned} ||(x,v) \odot (y,w)||^2 &= x^2y^2 + x^2||w||^2 + y^2||v||^2 + ||v||^2||w||^2 \\ &= (x^2 + ||v||^2)(y^2 + ||w||^2) \\ &= ||(x,v)||^2||(y,w)||^2, \end{aligned}$$

so we have established (4.9).

Now we show the connection between $\odot$ and Hurwitz's theorem. Pick two vectors $(x_1, \dots, x_{n+1})$ and $(y_1, \dots, y_{n+1})$ in $\mathbf{R}^{n+1}$. Their $\odot$-product is a third vector $(z_1, \dots, z_{n+1})$, where the components are computed according to (4.8). Writing (4.9) with the terms moved to opposite sides,

$$(x_1^2 + \cdots + x_{n+1}^2)(y_1^2 + \cdots + y_{n+1}^2) = z_1^2 + \cdots + z_{n+1}^2. \tag{4.10}$$

This identity holds for all real values of the variables. From the first key property of $\odot$, the z_k's are bilinear functions of the x_i's and y_j's. Thus, (4.10) and Hurwitz's theorem over $\mathbf{R}$ tell us $n+1$ is 1, 2, 4, or 8, so n is 0, 1, 3, or 7. Discard the case $n = 0$ and we are done. □

Up to a linear change of variables, it can be shown that the only product on $\mathbf{R}^3$ satisfying the conditions of Theorem 4.2 is the usual cross product. To see the construction of a product on $\mathbf{R}^7$ satisfying the conditions of Theorem 4.2, consult [2, pp. 278–279].

Appendix A. Lemma 3.1 revisited

The linear independence conclusion of Lemma 3.1 continues to hold under a weaker assumption than the C_i's having scalar squares: invertibility of the C_i's is sufficient. However, the proof becomes more involved, since we can't reduce immediately to the case when $b_{\mathbf{0}} \neq 0$. Here is the general result along these lines for readers who have heard of rings.

Theorem A.1. *Let F be a field and A be an associative ring with identity containing F. Suppose $a_1, \dots, a_m$ are m pairwise anticommuting units in A, where m is even. For $\boldsymbol{\delta} \in \{0,1\}^m$, set*

$$a^{\boldsymbol{\delta}} = a^{\delta_1} \cdots a^{\delta_m}.$$

The 2^m products $a^{\boldsymbol{\delta}}$ are linearly independent over F.

This has Lemma 3.1 as a special case taking $A = \mathrm{M}_d(\mathbf{C})$ and $a_i = C_i$.

Proof. Let $w(\boldsymbol{\delta})$ be the number of i's with $\delta_i = 1$. Then

$$a_i a_j a_i^{-1} = \begin{cases} a_j, & \text{if } i = j, \\ -a_j, & \text{if } i \neq j, \end{cases}$$

so

$$a_i a^{\boldsymbol{\delta}} a_i^{-1} = \varepsilon_{\boldsymbol{\delta},i} a^{\boldsymbol{\delta}}, \tag{A.1}$$

where

$$\varepsilon_{\boldsymbol{\delta},i} = \begin{cases} (-1)^{w(\boldsymbol{\delta})}, & \text{if } \delta_i = 0, \\ (-1)^{w(\boldsymbol{\delta})-1}, & \text{if } \delta_i = 1. \end{cases}$$

Since $w(\boldsymbol{\delta})$, by definition, is the number of i's such that $\delta_i = 1$, we get a global constraint linking the signs $\varepsilon_{\boldsymbol{\delta},1}, \dots, \varepsilon_{\boldsymbol{\delta},m}$:

$$\prod_{i=1}^{m} \varepsilon_{\boldsymbol{\delta},i} = (-1)^{mw(\boldsymbol{\delta})}(-1)^{w(\boldsymbol{\delta})} = (-1)^{w(\boldsymbol{\delta})}. \tag{A.2}$$

The last equality uses the evenness of m.

Suppose there is a nontrivial linear dependence relation among the $a^{\boldsymbol{\delta}}$'s, say

$$\sum_{\boldsymbol{\delta}} b_{\boldsymbol{\delta}} a^{\boldsymbol{\delta}} = 0, \tag{A.3}$$

for some coefficients $b_{\boldsymbol{\delta}} \in F$ not all zero. Pick such a nontrivial relation with a minimal number of nonzero coefficients. Fixing i between 1 and n, conjugate (A.3) by a_i. By (A.1), we get

$$\sum_{\boldsymbol{\delta}} \varepsilon_{\boldsymbol{\delta},i} b_{\boldsymbol{\delta}} a^{\boldsymbol{\delta}} = 0.$$

Adding and subtracting this from (A.3) gives

$$\sum_{\boldsymbol{\delta}} (1 - \varepsilon_{\boldsymbol{\delta},i}) b_{\boldsymbol{\delta}} a^{\boldsymbol{\delta}} = 0, \quad \sum_{\boldsymbol{\delta}} (1 + \varepsilon_{\boldsymbol{\delta},i}) b_{\boldsymbol{\delta}} a^{\boldsymbol{\delta}} = 0. \tag{A.4}$$

Suppose $\boldsymbol{\delta}'$ has $b_{\boldsymbol{\delta}'} \neq 0$. Then one of the linear relations in (A.4) has no $\boldsymbol{\delta}'$-term, so it is shorter than the minimal nontrivial relation. Thus *all* terms in the shorter relation have coefficient 0. That is, any $\boldsymbol{\delta}$ where $b_{\boldsymbol{\delta}} \neq 0$ has $1 \pm \varepsilon_{\boldsymbol{\delta},i} = 0$, taking $-$ if $\varepsilon_{\boldsymbol{\delta}',i} = 1$ and $+$ if $\varepsilon_{\boldsymbol{\delta}',i} = -1$. In other words,

$$b_{\boldsymbol{\delta}} \neq 0 \Longrightarrow \varepsilon_{\boldsymbol{\delta},i} = \varepsilon_{\boldsymbol{\delta}',i}$$

for all i. Therefore, multiplying over all i and using (A.2) tells us $(-1)^{w(\boldsymbol{\delta})} = (-1)^{w(\boldsymbol{\delta}')}$ for all $\boldsymbol{\delta}$ where $b_{\boldsymbol{\delta}} \neq 0$.

This implies, when $b_{\boldsymbol{\delta}} \neq 0$, that

$$\delta_i = 0 \Longrightarrow \varepsilon_{\boldsymbol{\delta},i} = (-1)^{w(\boldsymbol{\delta}')}. \tag{A.5}$$

Since $\varepsilon_{\boldsymbol{\delta},i} = \varepsilon_{\boldsymbol{\delta}',i}$ when $b_{\boldsymbol{\delta}} \neq 0$, we can rewrite (A.5) as

$$\delta_i = 0 \Longrightarrow \varepsilon_{\boldsymbol{\delta}',i} = (-1)^{w(\boldsymbol{\delta}')}$$

when $b_{\boldsymbol{\delta}} \neq 0$. Thus, when $b_{\boldsymbol{\delta}} \neq 0$,

$$\delta_i = 0 \Longrightarrow \delta_i' = 0.$$

Similarly,

$$\delta_i = 1 \Longrightarrow \delta_i' = 1,$$

so in fact $\boldsymbol{\delta} = \boldsymbol{\delta}'$. That is, the minimal nontrivial linear relation among the $a^{\boldsymbol{\delta}}$'s has just one nonzero term. But then it reads $b_{\boldsymbol{\delta}'} a^{\boldsymbol{\delta}'} = 0$, which is impossible. □

For a further discussion of results of this kind, see [5, p. 37].

References

[1] M. L. Curtis, *Abstract Linear Algebra*, Springer-Verlag, New York, 1990.

[2] H.-D. Ebbinghaus *et al.*, *Numbers*, Springer-Verlag, New York, 1991.

[3] A. Hurwitz, "Über die Composition der quadratichen Formen von beliebig vielen Variabeln," Werke, Band II, Basel 1932, 565-571.

[4] N. Jacobson, *Basic Algebra I*, 2nd ed., W.H. Freeman and Co., New York, 1985.

[5] D. Shapiro, *Compositions of Quadratic Forms*, de Gruyter, New York, 2000.