THE HURWITZ THEOREM ON SUMS OF SQUARES

KEITH CONRAD

1. Introduction

When multiplication is commutative, a product of two squares is a square: $x^2y^2 = (xy)^2$. A more profound identity is the 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:

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

These are polynomial identities, so they are valid when we substitute for the variables elements of any field (or, for that matter, elements of any commutative ring).

In the 19th century, after the 4-square identity (1.2) was popularized by Hamilton in his work on quaternions (it had been found by Euler in the 18th century but then forgotten), Cayley discovered a similar 8-square identity. In all of these sum-of-squares identities, the terms being squared in the product 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 [4] proved his famous "1,2,4,8 theorem," which says that further identities of this kind are *impossible*.

Theorem 1.1 (Hurwitz, 1898)**.** *Let F be a field of characteristic not equal to 2. If*

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

for all $x_1, \ldots, x_n, y_1, \ldots, y_n$ in F, where each z_k is an F-bilinear function of the x's and the y's, then $n = 1, 2, 4$ or 8.

Hurwitz's original proof was stated for $F = \mathbf{C}$, but the field of scalars only needs to be of characteristic not equal to 2 for his proof to work. Nothing would be lost if you take $F = \mathbf{C}$ in the rest of this discussion. (What if the field F has characteristic 2? Then there *is* an identity as in (1.3) for all n because a sum of squares in characteristic 2 is again a square.)

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 F. Then we show by two different methods that the matrix equations can be solved only when $n = 1, 2, 4$, or 8. The

first method, in Section 3, will involve a linear independence property of certain matrix products, and is a simplified version of Hurwitz's original argument. Our treatement is based on [6]. The second method, in Section 4, will use representation theory. This method is due to Eckmann [2] (see also [3, pp. 141-144]). Sections 3 and 4 can be read independently of each other. As an application of Hurwitz's theorem, we show in Section 5 that the only Euclidean spaces which can admit a multiplication resembling the usual vector cross product are $\mathbf{R}$ (a degenerate case, it turns out), $\mathbf{R}^3$, and $\mathbf{R}^7$.

While Hurwitz proved only the dimension constraints $n = 1, 2, 4$, and 8, it is also the case that, up to a linear change of variables, the only sum of squares identities in these dimensions are the ones associated to multiplication in the four classical real division algebras of dimensions 1, 2, 4, and 8: the real numbers, complex numbers, quaternions, and octonions. For a proof of this more precise result, see [5, §7.6] or [6, Appendix, Chap. 1]. Readers unfamiliar with algebra in the quaternions and octonions can look in [1].

2. The Hurwitz Matrix Equations

Lemma 2.1. *Let V be a finite-dimensional vector space over F, where F does not have characteristic 2. 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 non-zero determinants, $1 = (-1)^{\dim V}$ in F, so $\dim V$ is even since the characteristic of F is not 2. □

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 $a_{ijk} \in F$. 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 $k = 1, \ldots, n$ is the same as the single equation

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

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:

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

This sum can be written as $x_1A_1 + \cdots + x_nA_n$, 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

$$\mathbf{z} = (x_1A_1 + \cdots + x_nA_n)\mathbf{y} = A_{\mathbf{x}}\mathbf{y},$$

where we set $A_{\mathbf{x}} = x_1A_1 + \cdots + x_nA_n$.

With this notation, the right side of (1.3) is

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

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 (since F has more than 2 elements)

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

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

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

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

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

These are the *Hurwitz matrix equations*. (The actual entries in the A_i's won't matter anymore.) 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

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

(We write $i\neq j$ rather than $i<j$ to make things more symmetric; it doesn't change anything.) 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.

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 $\mathrm{M}_d(F)$ which are pairwise anticommuting and each C_i^2 is a non-zero scalar diagonal matrix. (For instance, in the notation of (2.7), we can use $B_1,B_2,\ldots,B_{n-2}$ in $\mathrm{M}_n(F)$. 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 $\mathrm{M}_d(F)$. 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. 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 F and are not all 0. Take such a relation with as few non-zero 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 non-zero scalar matrix, $C^{\boldsymbol{\delta}'} C^{\boldsymbol{\delta}'}$ is a non-zero scalar matrix for any $\boldsymbol{\delta}'$. Moreover, as $\boldsymbol{\delta}$ varies and $\boldsymbol{\delta}'$ is fixed,

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

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 non-zero coefficients as in (3.1) but now the coefficient of $C^{\mathbf{0}} = I_d$ is non-zero. 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 $\pm$ 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 non-zero in (3.1). Therefore (3.6) is a linear relation with fewer non-zero terms than the non-zero 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 non-zero 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 $-1 \neq 1$ in F. We have shown $b_{\boldsymbol{\delta}} = 0$ for $\boldsymbol{\delta} \neq \mathbf{0}$, but then the linear relation (3.1) has just one non-zero term, which is impossible. □

Returning now to the setting of the proof of Theorem 1.1, apply Lemma 3.1 to the matrices $B_1, \ldots, B_{n-2}$ in $\mathrm{M}_n(F)$. (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 $n > 4$, $\frac{n}{2}$ is even, so $n \neq 6$.

Consider the B_j's as linear operators on $V := \overline{F}^n$, where $\overline{F}$ is an algebraic closure of F. Since $B_j^2 = -I_n$, the eigenvalues of B_j are $\pm i = \pm\sqrt{-1}$.

Let $\langle \cdot, \cdot \rangle$ be the standard inner product on V: $\langle (a_1, \ldots, a_n), (b_1, \ldots, b_n) \rangle = \sum_{k=1}^n a_k b_k$. Since $B_j^\top = -B_j$, we have

$$\langle B_j v, w \rangle = -\langle v, B_j w \rangle$$

for any v and w in V. There is an eigenspace decomposition $V = U \oplus W$ for the action of B_1, where

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

Since B_1 is injective and U and W are eigenspaces, $B_1(U) = U$ and $B_1(W) = W$. Of greater interest is that, for $j = 2, 3, \ldots, n-1$, $B_j(U) = W$ and $B_j(W) = U$. To see this, it will suffice to show $B_j(U) \subset W$ and $B_j(W) \subset U$. (Then, by injectivity of B_j, we'd get $\dim U \leq \dim W$ and $\dim W \leq \dim U$, so these inequalities are equalities and B_j maps U onto W and W onto U by the rank-nullity theorem.)

For $v \in U$,

$$B_1(B_j v) = -B_j(B_1 v) = -B_j(iv) = -iB_j v,$$

so $B_j v \in W$. Thus, $B_j(U) \subset W$. That $B_j(W) \subset U$ is analogous. It follows, as noted already, that $\dim U = \dim W = \frac{n}{2}$.

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 $C_j = B_2 \circ B_j$ is an invertible linear operator on U. For $n > 4$, a direct calculation

shows that C_3 and C_4 are anti-commuting on U (as are C_j and C_k for any different $j, k > 2$). This forces $\dim U$ to be even by Lemma 2.1. Thus $2|\frac{n}{2}$, so $4|n$. This eliminates the choice $n = 6$ and concludes the first proof of Hurwitz's theorem.

4. Conclusion via Representation Theory

Assuming $n > 2$ is even, we will derive $n = 4$ or 8 from (2.7) using representation theory.

Consider the group of matrices generated by the B_i's. These consist of the matrix products

$$\pm B_1^{a_1} \cdots B_{n-1}^{a_{n-1}},$$

where $a_i = 0$ or 1. Note $-I_n \neq I_n$ since F doesn't have characteristic 2.

Let G be a group generated by elements $g_1, \ldots, g_{n-1}$ such that

$$g_i^2 = \varepsilon \neq 1,\ \varepsilon^2 = 1,\ g_i g_j = \varepsilon g_j g_i \text{ for } i \neq j. \tag{4.1}$$

(The Hurwitz matrix equations, or rather their consequence in (2.6), led us to an example of such a group.) Every element of G has the form

$$\varepsilon^{a_0} g_1^{a_1} \cdot \cdots \cdot g_{n-1}^{a_{n-1}},$$

where $a_i = 0$ or 1. Also, if G exists then the subgroup $\{g_1, \ldots, g_m\}$ for $2 \leq m \leq n-2$ has the same formal properties (4.1) as G, but with fewer generators. Note ε commutes with all the g_i's, so $\varepsilon \in Z(G)$.

We now show the following four facts:

(a) $\#G = 2^n$,
(b) $[G, G] = \{1, \varepsilon\}$,
(c) If $g \notin Z(G)$, then the conjuagcy class of g is $\{g, \varepsilon g\}$,
(d) The evenness of n implies

$$Z(G) = \{1, \varepsilon, g_1 \cdots g_{n-1}, \varepsilon g_1 \cdots g_{n-1}\}.$$

(a) Certainly $\#G \leq 2^n$. We need to show that if

$$\varepsilon^{a_0} g_1^{a_1} \cdots g_{n-1}^{a_{n-1}} = 1, \tag{4.2}$$

then all a_i are even (or simply equal 0 if we assume, as we can, that $a_i = 0$ or 1).

Well, if $n - 1 = 2$ and (4.2) holds with $a_2 = 1$ then g_2 is in the group generated by ε and g_1, hence in the group generated by g_1 since $\varepsilon^2 = g_1$. That implies g_2 commutes with g_1, which is not the case. So $a_2 = 0$ and $\varepsilon^{a_0} g_1^{a_1} = 1$. Since g_1 has order 4, it doesn't lie in the 2 element group generated by ε, so $a_1 = 0$. Therefore $a_0 = 0$.

Now assume $n - 1 > 2$ and $a_{n-1} = 1$. Multiplying each side of (4.2) by g_{n-1} on the right, we move $g_{n-1}^2 = \varepsilon$ over to the ε term (since $\varepsilon \in Z(G)$) and get

$$\varepsilon^{a_0'} g_1^{a_1} \cdots g_{n-2}^{a_{n-2}} = 1,$$

where $a_0' = 0$ or 1 since ε has order 2. Since the group generated by $\varepsilon, g_1, \ldots, g_{n-2}$ has the same formal properties as G, we see by induction that

$$a_1 = \cdots = a_{n-2} = 0.$$

Thus $\varepsilon^{a_0'} g_{n-1} = 1$, so $g_{n-1} \in \{1, \varepsilon\}$, a contradiction.

(b) Since $n - 1 \geq 2$, (4.1) gives $g_1 g_2 g_1^{-1} g_2^{-1} = \varepsilon$, so ε lies in $[G, G]$. Since $\varepsilon \in Z(G)$, the group $G/\{1, \varepsilon\}$ is abelian by the defining properties of G, so $[G, G] = \{1, \varepsilon\}$.

(c) This is obvious from b.

(d) An element g of G lies in the center if and only if $gg_i = g_ig$ for all i. Write

$$g = \varepsilon^{a_0} g_1^{a_1} \cdots g_{n-1}^{a_{n-1}},$$

where $a_i = 0$ or 1. Then (using $g_ig_jg_i^{-1} = \varepsilon g_j$)

$$\begin{aligned} gg_i = g_ig &\iff \varepsilon^{a_0} g_1^{a_1} \cdots g_{n-1}^{a_{n-1}} = g_i \varepsilon^{a_0} g_1^{a_1} \cdots g_{n-1}^{a_{n-1}} g_i^{-1} \\ &\iff \varepsilon^{a_0} g_1^{a_1} \cdots g_{n-1}^{a_{n-1}} = \varepsilon^{a_0 + \sum_{\substack{j=1 \\ j\neq i}}^{n-1} a_j} g_1^{a_1} \cdots g_{n-1}^{a_{n-1}}. \end{aligned}$$

Since ε has order 2, we see

$$g \in Z(G) \Longleftrightarrow \sum_{\substack{j=1 \\ j\neq i}}^{n-1} a_j \equiv 0 \bmod 2 \text{ for all } i.$$

For $i \neq k$, we get

$$\sum_{\substack{j=1 \\ j\neq i}}^{n-1} a_j \equiv \sum_{\substack{j=1 \\ j\neq k}}^{n-1} a_j \bmod 2,$$

so $a_i \equiv a_k \bmod 2$. Thus $a_1 = \cdots = a_{n-1}$, so $g = \varepsilon^{a_0}$ or $\varepsilon^{a_0} g_1 \cdots g_{n-1}$. Hence

$$g \in Z(G) \Longleftrightarrow (n-2)a_1 \equiv 0 \bmod 2,$$

so $Z(G)$ has the elements as indicated for n even. (That n is even was only used in the last line. If instead n were odd, then $Z(G) = \{1, \varepsilon\}$.)

We now bring in representation theory. The original Hurwitz problem gave an n-dimensional (faithful) representation of G over F, which we view as a representation over the algebraic closure $\overline{F}$. Which irreducible representations of G over $\overline{F}$ can occur in this n-dimensional representation? Since $\overline{F}$ doesn't have characteristic 2, the characteristic doesn't divide the order of G, so classical representation theory applies.

Since $G/[G,G]$ has size 2^{n-1}, G has 2^{n-1} representations of degree 1. The number of representations equals the number of conjugacy classes. We already computed the conjugacy classes of G, so we can read off the number of conjugacy classes. Since n is even, G has

$$4 + \frac{1}{2}(2^n - 4) = 2^{n-1} + 2$$

conjugacy classes. (If n were odd, there would be $2 + \frac{1}{2}(2^n - 2) = 2^{n-1} + 1$ conjugacy classes.) Thus, for even n, G has two irreducible representations of degree greater than 1. Let f_i be the degrees of the irreducible representations of G over $\overline{F}$. Since $\#G = \sum f_i^2$ and all f_i divide $\#G$ (hence all f_i are powers of 2), we see (since $n - 1 > 1$) that G has two irreducible representations of degree $2^{\frac{n}{2}-1} > 1$ if n is even. (If n were odd, G would have just one irreducible representation of degree $2^{\frac{n-1}{2}} > 1$.)

Our problem gave us an n-dimensional representation of G where ε is represented by $-I_n$, hence by the negative of the identity map on any subspace. Since $\varepsilon \in [G,G]$, it is sent to 1 under all 1-dimensional representations. Therefore our n-dimensional representation of G has no irreducible subrepresentations of degree 1. Thus, for even $n > 2$ we must have

$$2^{\frac{n}{2}-1} | n.$$

Letting $n = 2^r s$ for $r \geq 1$ and s odd, we have $\frac{n}{2} - 1 \leq r$, so

$$2^r \leq n \leq 2r + 2.$$

This implies $n = 4$ or 8.

5. Vector products

We use the Hurwitz 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:

$$(v_1 + v_2) \times w = v_1 \times w + v_2 \times w, \quad v \times (w_1 + w_2) = v \times w_1 + v \times w_2, \tag{5.1}$$

and

$$(cv) \times w = c(v \times w), \quad v \times (cw) = c(v \times w), \tag{5.2}$$

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

$$v \times \mathbf{0} = \mathbf{0}, \quad \mathbf{0} \times w = \mathbf{0}. \tag{5.3}$$

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

$$v \cdot (v \times w) = 0, \quad w \cdot (v \times w) = 0. \tag{5.4}$$

This property is satisfied by the cross product on $\mathbf{R}^3$, thus motivating this condition. However, it is not satisfied by other kinds of products in linear algebra. For instance, matrix multiplication on $\mathrm{M}_d(\mathbf{R}) = \mathbf{R}^{d^2}$ is an $\mathbf{R}$-bilinear product but (5.4) isn't satisfied when v and w are matrices, $\times$ means matrix multiplication, and $\cdot$ is 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:

$$||v \times w||^2 = ||v||^2 ||w||^2 - (v \cdot w)^2. \tag{5.5}$$

When $n = 1$, a product on $\mathbf{R}^n = \mathbf{R}$ satisfying (5.5) must be identically zero. Indeed, the dot product on $\mathbf{R}$ is the ordinary product, so (5.5) becomes $|x \times y|^2 = x^2 y^2 - (xy)^2 = 0$, so $x \times y = 0$. So we only care about the case $n > 1$.

The assumption (5.5) looks more complicated than the earlier assumptions. The following result expresses (5.5) in simpler terms, but it is in the form (5.5) that we will actually use the assumption.

Theorem 5.1. *Let $\times$ be a product on $\mathbf{R}^n$ which satisfies* (5.1), (5.2), *and* (5.4). *Then* (5.5) *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 (5.5) implies the two conditions in the theorem. Now we assume the two conditions and derive (5.5).

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 (5.5) 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

$$||u \times w|| = ||u||||w||. \tag{5.6}$$

Since $w \times w = \mathbf{0}$, $u \times w = v \times w$ by bilinearity and (5.3). Squaring both sides of (5.6),

$$||v \times w||^2 = ||u||^2||w||^2 \tag{5.7}$$

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 (5.7) gives

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

□

Theorem 5.2. *For $n \geq 1$, assume there is a multiplication $\times \colon \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ satisfying* (5.1), (5.2), (5.4), *and* (5.5). *Then $n = 1$, 3, or 7.*

We have seen the $n = 1$ case is quite dull, so the only interesting cases in Theorem 5.2 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

$$(x, v) \odot (y, w) = (xy - v \cdot w, xw + yv + v \times w). \tag{5.8}$$

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 (5.8) is a linear function of the other pair.

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

$$||(x,v)\odot(y,w)||^2 = ||(x,v)||^2||(y,w)||^2. \tag{5.9}$$

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 (5.4), $v\times w$ 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 (5.5), 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 (5.9).

Now we show the connection between $\odot$ and Hurwitz's theorem. Pick two vectors $(x_1,\ldots,x_{n+1})$ and $(y_1,\ldots,y_{n+1})$ in $\mathbf{R}^{n+1}$. Their $\odot$ product is a third vector $(z_1,\ldots,z_{n+1})$, where the components are computed according to (5.8). Writing (5.9) with the terms moved to opposite sides,

$$(x_1^2 + \cdots + x_{n+1}^2)(y_1^2 + \cdots + y_{n+1}^2) = z_1^2 + \cdots + z_{n+1}^2. \tag{5.10}$$

This identity holds for all values of the variables, so it is a formal algebraic identity as well. From the first key property of $\odot$, the z_k's are bilinear functions of the x_i's and y_j's. Thus, (5.10) and Hurwitz's theorem tell us $n+1$ is 1, 2, 4, or 8, so n is 0, 1, 3, or 7. The case $n=0$ is discarded. □

Up to a linear change of variables, it can be shown that the only product on $\mathbf{R}^3$ satisfying the conditions of Theorem 5.2 is the usual cross product. A product on $\mathbf{R}^7$ satisfying the conditions of Theorem 5.2 can be constructed, but the details are somewhat tedious. See [1, 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 is sufficient. However, the proof becomes a little 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.

Theorem A.1. *Let F be a field not of characteristic 2 and A be an associative F-algebra. Suppose $a_1,\ldots,a_m$ are m pairwise anticommuting units in A, where m is even. For $\boldsymbol{\delta} \in \{0,1\}^m$, set*

$$a^{\boldsymbol{\delta}} = a_1^{\delta_1}\cdots a_m^{\delta_m}.$$

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

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 $\varepsilon_{\boldsymbol{\delta},1}, \ldots, \varepsilon_{\boldsymbol{\delta},m}$:

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

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

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

Pick a $\boldsymbol{\delta}'$ such that $b_{\boldsymbol{\delta}'} \neq 0$. Since $\varepsilon_{\boldsymbol{\delta}',i}$ is ± 1, one of the linear relations in (A.4) has no $\boldsymbol{\delta}'$-term, so it has fewer nonzero terms than the minimal nontrivial relation (A.3). Thus *all* terms in the shorter relation have coefficient 0. That is, any $\boldsymbol{\delta}$ where $b_{\boldsymbol{\delta}} \neq 0$ has $1 + \varepsilon_{\boldsymbol{\delta},i} = 0$ if $\varepsilon_{\boldsymbol{\delta}',i} = -1$ and $1 - \varepsilon_{\boldsymbol{\delta},i} = 0$ 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. Multiplying these equation 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 non-zero 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 [6, p. 37].

References

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

[2] B. Eckmann, "Gruppentheoretischer Beweis des Satzes von Hurwitz-Radon über die Komposition quadratichen Formen," *Comment. Math. Helv.* **15** (1942), 358-366.

[3] I. Herstein, *Noncommutative Rings*, Mathematical Association of America, 1968.

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

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

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