THE HURWITZ THEOREM ON SUMS OF SQUARES

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:

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

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 [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 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 C, 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 representation theory in Section 3 that the matrix equations can be solved only when $n = 1, 2, 4$, or 8. This method is due to Eckmann [2] (see also [3, pp. 141-144]).

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 [1], [5, §7.6], or [6, Appendix, Chap. 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 $k = 1, \dots, n$ is the same as the single equation

$$\begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} \sum_{i,j} a_{ij1} x_i y_j \\ \vdots \\ \sum_{i,j} a_{ijn} x_i y_j \end{pmatrix} \tag{2.3}$$

$$= \begin{pmatrix} \sum_j \left(\sum_i a_{ij1} x_i\right) y_j \\ \vdots \\ \sum_j \left(\sum_i a_{ijn} x_i\right) y_j \end{pmatrix}$$

$$= \begin{pmatrix} \sum_i a_{i11} x_i & \dots & \sum_i a_{in1} x_i \\ \vdots & \ddots & \vdots \\ \sum_i a_{i1n} x_i & \dots & \sum_i a_{inn} x_i \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

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:

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

This sum can be written as $x_1 A_1 + \cdots + x_n A_n$, 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

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

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

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

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

(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. Using Representation Theory

Consider the group of matrices generated by the B_i's. These are 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. Since $n > 2$ is even, we have $n \geq 4$.

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

$$g_i^2 = \varepsilon \neq 1, \ \varepsilon^2 = 1, \ g_ig_j = \varepsilon g_jg_i \text{ for } i \neq j. \tag{3.1}$$

(Does such a group exist? Well, for our purposes we only have to note that the Hurwitz matrix equations, or rather their consequence in (2.6), led us to an example of such a group, so if the Hurwitz matrix equations hold for some $n \times n$ matrices then there is 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 $\langle g_1, \ldots, g_m\rangle$ for $2 \leq m \leq n-2$ has the same formal properties (3.1) 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 $Z(G)$ has four elements:

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

We will not need representation theory to do this.

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

$$\varepsilon^{a_0} g_1^{a_1} \cdots g_{n-1}^{a_{n-1}} = 1, \tag{3.2}$$

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 (3.2) 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 (3.2) 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 $n-1 \geq 2$, (3.1) 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 $gg_i = g_i g$ 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 $g_i g_j g_i^{-1} = \varepsilon g_j$)

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

Since ε has order 2, we see

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

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.

References

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

[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.