Cayley–Dickson Construction for Beginners: An Accessible Proof of Hurwitz's Sums Of Squares Theorem

Ezra Brown and Adrian Rice

June 25, 2021

1 Introduction

The purpose of this paper is to give a simple proof, intelligible to undergraduates, that a particular multiplicative formula for sums of n squares can only occur when $n = 1, 2, 4$, and 8, a result originally proved by Hurwitz in 1898. We begin with a brief survey of the history of sums of squares, leading to a discussion of the related topic of normed division algebras over the real numbers.[1] This story culminates with a crucial paper by Dickson in 1919 that not only contained an exposition of Hurwitz's 1898 proof, but which also outlined a new process for producing division algebras over the reals. That process, now called Cayley-Dickson construction, is intimately connected with the product formula for sums of squares and the dimensions necessary for its existence. For this reason, we present an introduction to Cayley-Dickson construction for beginners, together with a proof of Hurwitz's theorem accessible to anyone with a basic knowledge of undergraduate algebra.

2 Historical Background

The question in which we are interested is: for which values of n does it hold that

$$\left(x_1^2 + x_2^2 + \cdots + x_n^2\right)\left(y_1^2 + y_2^2 + \cdots + y_n^2\right) = z_1^2 + z_2^2 + \cdots + z_n^2 \tag{1}$$

where $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n, z_1, z_2, \ldots, z_n \in \mathbb{Z}$ and $n \in \mathbb{N}$? Now, since

$$x_1^2 y_1^2 = (x_1 y_1)^2$$

[1] For a more detailed discussion, see [12].

it is trivial that formula (1) holds when $n = 1$, and as early as the 3rd century, Diophantus was aware that it is also true when $n = 2$. He observed that the number 65 could be written as two different sums of integer squares, namely $16 + 49$ and $64 + 1$, since it is itself the product of two sums of two squares, namely 13×5 or $(3^2 + 2^2)(2^2 + 1^2)$. The formula that he implicitly used would today be written as

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 \mp x_2y_2)^2 + (x_2y_1 \pm x_1y_2)^2$$

which is, of course, equation (1) when $n = 2$.

By the 17th century, it was realized that no extension of identity (1) for the $n = 3$ case would be possible. For example, as the French mathematician Albert Girard noticed in 1625, $3 = 1^2 + 1^2 + 1^2$ and $13 = 3^2 + 2^2 + 0^2$ both have three-square sum representations, but their product 39 does not.

The next major step came with Euler who in 1748 announced that he had derived an expression for formula (1) when $n = 4$, namely:

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

By now it was at least intuitively clear that the dimensions of any further extensions were likely to be of the form $n = 2^m$. Corroboration came seventy years after Euler's result, when a relatively unknown Danish mathematician, Carl Ferdinand Degen, managed to extend it still further, proving the $n = 8$ case:

$$\begin{aligned}
&\left(x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2 + x_8^2\right)\left(y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2 + y_6^2 + y_7^2 + y_8^2\right) \\
&\quad = (x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4 - x_5y_5 - x_6y_6 - x_7y_7 - x_8y_8)^2 \\
&\quad + (x_1y_2 + x_2y_1 + x_3y_4 - x_4y_3 + x_5y_6 - x_6y_5 - x_7y_8 + x_8y_7)^2 \\
&\quad + (x_1y_3 - x_2y_4 + x_3y_1 + x_4y_2 + x_5y_7 + x_6y_8 - x_7y_5 - x_8y_6)^2 \\
&\quad + (x_1y_4 + x_2y_3 - x_3y_2 + x_4y_1 + x_5y_8 - x_6y_7 + x_7y_6 - x_8y_5)^2 \\
&\quad + (x_1y_5 - x_2y_6 - x_3y_7 - x_4y_8 + x_5y_1 + x_6y_2 + x_7y_3 + x_8y_4)^2 \\
&\quad + (x_1y_6 + x_2y_5 - x_3y_8 + x_4y_7 - x_5y_2 + x_6y_1 - x_7y_4 + x_8y_3)^2 \\
&\quad + (x_1y_7 + x_2y_8 + x_3y_5 - x_4y_6 - x_5y_3 + x_6y_4 + x_7y_1 - x_8y_2)^2 \\
&\quad + (x_1y_8 - x_2y_7 - x_3y_6 + x_4y_5 - x_5y_4 - x_6y_3 + x_7y_2 + x_8y_1)^2
\end{aligned} \tag{3}$$

This 8-squares formula was rediscovered independently a quarter of a century later—in a completely different mathematical context—when in 1843 an Irish mathematician by the name of John Thomas Graves created a new system of hypercomplex numbers. He had been inspired by the recent work of his friend, William Rowan Hamilton, who earlier that year had created the 4-dimensional extension of complex numbers [10, pp. 106–110], known as the *quaternions*:[2]

$$\mathbb{H} = \{x_1 + x_2 i + x_3 j + x_4 k : x_1, x_2, x_3, x_4 \in \mathbb{R}, i^2 = j^2 = k^2 = ijk = -1\}.$$

Because of the fundamental equation connecting the three imaginary quantities i, j, k, this new algebra turned out to be noncommutative with respect to multiplication since, for example, $ij = k$ but $ji = -k$.

Furthermore, by analogy with the algebra of complex numbers, letting the conjugate of a quaternion $z_1 = x_1 + x_2 i + x_3 j + x_4 k$ be simply $\overline{z_1} = x_1 - x_2 i - x_3 j - x_4 k$, and defining the *norm* function $N(z) = z \cdot \overline{z}$, Hamilton found that $N(z_1) = x_1^2 + x_2^2 + x_3^2 + x_4^2$ and that, for all $z_1, z_2 \in \mathbb{H}$,

$$N(z_1)N(z_2) = N(z_1 z_2). \tag{4}$$

In other words, he had discovered that $\mathbb{H}$ was a new kind of structure called a *normed division algebra* over the real numbers, adding to the two that were previously known, namely $\mathbb{R}$ and $\mathbb{C}$. Now, you may be familiar with real normed algebras, but it doesn't hurt to review some definitions. So here we go.

An *n-dimensional real algebra* $\mathcal{A}$ is an n-dimensional vector space over the real numbers $\mathbb{R}$ equipped with a multiplication that is left- and right-distributive over vector addition and satisfies $(av)(bw) = (ab)(vw)$ for all real numbers a and b and all vectors $v, w \in \mathcal{A}$. We call $\mathcal{A}$ a *division algebra* if every nonzero $v \in \mathcal{A}$ has both a left- and a right-multiplicative inverse. Finally, $\mathcal{A}$ is a *real normed algebra* if there exists a mapping N from $\mathcal{A}$ to the nonnegative real numbers such that $N(v) = 0$ if and only if v is the zero vector and, most importantly, for all $v, w \in \mathcal{A}$,

$$N(v)N(w) = N(vw).$$

These criteria were, of course, satisfied by Hamilton's quaternions.

Spurred on by Hamilton's work, Graves came up with an 8-dimensional extension of $\mathbb{H}$, known today as the *octonions* [1]. This new real algebra $\mathbb{O}$ contained numbers of the form

$$z = x_1 + x_2 i_1 + x_3 i_2 + x_4 i_3 + x_5 i_4 + x_6 i_5 + x_7 i_6 + x_8 i_7$$

[2]To discover why Hamilton was unable to come up with a 3-dimensional extension of $\mathbb{C}$, see our paper [2].

where $x_1, x_2, \ldots x_8 \in \mathbb{R}$ and the seven basic imaginary components $i_1, i_2, \ldots, i_7$ were governed by the following rules:

$$\begin{aligned} &i_1^2 = \cdots = i_7^2 = -1, \\ &i_\alpha i_\beta = -i_\beta i_\alpha, \\ &i_\alpha i_\beta = i_\gamma \implies i_{\alpha+1} i_{\beta+1} = i_{\gamma+1}, \\ &i_\alpha i_\beta = i_\gamma \implies i_{2\alpha} i_{2\beta} = i_{2\gamma}, \end{aligned}$$

with all subscripts belonging to $\{1, 2, 3, 4, 5, 6, 7\}$.

Again, due obviously to the second of the above equations, multiplication in this new algebra is noncommutative. But it also turned out to be nonassociative, since in general $i_\alpha(i_\beta i_\gamma) \neq (i_\alpha i_\beta) i_\gamma$. Furthermore, letting the conjugate of an octonion z_1 be $\overline{z_1} = x_1 - x_2 i_1 - x_3 i_2 - x_4 i_3 - x_5 i_4 - x_6 i_5 - x_7 i_6 - x_8 i_7$ resulted in the norm function $N(z_1) = z_1 \cdot \overline{z_1} = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2 + x_8^2$. Now, for $\mathbb{O}$ to be a normed division algebra, equation (4) needed to hold for all $z_i \in \mathbb{O}$. In fact, not only was this the case, but if $z_1 = x_1 + x_2 i_1 + x_3 i_2 + x_4 i_3 + x_5 i_4 + x_6 i_5 + x_7 i_6 + x_8 i_7$ and $z_2 = y_1 + y_2 i_1 + y_3 i_2 + y_4 i_3 + y_5 i_4 + y_6 i_5 + y_7 i_6 + y_8 i_7$, then equation (4) produced Degen's 8-squares formula (3), which is exactly how Graves was led to this result in the first place.

But Graves was not the only mathematician inspired by Hamilton's discovery of quaternion algebra. Around the same time, and completely independently, the English mathematician Arthur Cayley created an identical system of 8-dimensional algebra which he published in 1845 [3]. Graves had also written up his work for publication. The problem was that he had entrusted his manuscript to the care of his friend, Hamilton, whose memory and organization were not perhaps as good as they could have been. The consequence was that, although Graves' work was eventually published [10, pp. 648–656], Cayley's octonions appeared first, winning him much of the credit for their discovery. Indeed, for many years octonions were better known as *Cayley numbers*.

It was quickly realized that the existence of the Cayley-Graves 8-dimensional normed algebra $\mathbb{O}$ is a necessary and sufficient condition for Degen's 8-squares formula, that Hamilton's 4-dimensional quaternions $\mathbb{H}$ had the same relationship to Euler's 4-squares formula, and that the complex numbers $\mathbb{C}$ and the 2-squares formula were similarly co-dependent. Therefore, the product formula (1) for sums of n squares would hold if and only if a corresponding n-dimensional normed algebra over the real numbers could be found. But did any more of these algebras exist? As early as the 1840s, Cayley and others began to believe that the answer was no. But it was not until 1898 that the German mathematician Adolf Hurwitz managed to prove it [11]. Our explanation of his proof will come later, but first let's convince ourselves that the only possible normed algebras over the reals are indeed $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, and $\mathbb{O}$.

3 The Cayley-Dickson construction

In 1919, the American mathematician Leonard Eugene Dickson published a noteworthy paper entitled "On Quaternions and Their Generalization and the History of the Eight Square Theorem" [8]. In it, he explained an ingenious method he had recently devised[3] that could not only construct the normed algebra $\mathbb{C}$ from $\mathbb{R}$, but could also build $\mathbb{H}$ from $\mathbb{C}$, and $\mathbb{O}$ from $\mathbb{H}$, all using exactly the same procedure.

That procedure was based on an idea first published by Hamilton in 1835 [9]. Consider two real numbers, x and y. Let the ordered pair (x, y) denote the complex number $z = x + yi$, and define its conjugate $\overline{z}$ to be $x - yi$, or $(x, -y)$. Hamilton defined multiplication in $\mathbb{C}$ to be equivalent to

$$(x_1, y_1)(x_2, y_2) = (x_1x_2 - y_2y_1, y_2x_1 + y_1x_2).$$

Since addition was defined merely as the addition of corresponding components, and the multiplicative inverse (or division) operation was easily proved to be

$$z^{-1} = \frac{\overline{z}}{N(z)},$$

it was clear that Hamilton had devised a new method of obtaining the two-dimensional algebra of $\mathbb{C}$ from the one-dimensional algebra of $\mathbb{R}$.

What Dickson now did was to modify Hamilton's definition of multiplication slightly, so that it became:

$$(x_1, y_1)(x_2, y_2) = (x_1x_2 - \overline{y_2}y_1, y_2x_1 + y_1\overline{x_2}). \tag{5}$$

This clearly had no effect on the construction of $\mathbb{C}$ from $\mathbb{R}$, since for all $x \in \mathbb{R}$, $\overline{x} = x$. However, it did result in a remarkable generalization of Hamilton's method which, given an n-dimensional algebra $\mathcal{A}$, enabled the immediate construction of an algebraic extension of $\mathcal{A}$ with dimension $2n$. In effect this amounted to defining the set $\mathbb{C}$ to be equal to $\mathbb{R} + \mathbb{R}i$, where $i^2 = -1$. Similarly, for quaternions, $\mathbb{H}$ could be defined as $\mathbb{C} + \mathbb{C}j$, where $i^2 = j^2 = k^2 = ijk = -1$, since it can be easily shown that, if $x_1, x_2, y_1, y_2 \in \mathbb{R}$,

$$(x_1 + x_2i) + (y_1 + y_2i)j = x_1 + x_2i + y_1j + y_2k \in \mathbb{H}$$

and if $x_1, x_2, y_1, y_2 \in \mathbb{C}$, then

$$(x_1 + y_1j)(x_2 + y_2j) = (x_1x_2 - \overline{y_2}y_1) + (y_2x_1 + y_1\overline{x_2})j \in \mathbb{H}.$$

[3] Dickson's idea had first appeared in [6, pp. 72–73] and was explained in more detail in [7, pp. 15–16].

Today, this process of successively building algebras of dimension 2^n is known as *Cayley-Dickson construction*.[4] But, given that it is a generalization of a method due to Hamilton, why is the name of *Cayley* attached to it?

One possible reason lies in the groundbreaking paper of 1858 that introduced the algebra of matrices to the world [4]. In this paper, Cayley put forward the idea of representing quaternions in terms of linear combinations of 2×2 matrices

$$\mathbf{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \mathbf{I} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \qquad \mathbf{J} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad \mathbf{K} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}.$$

In addition to satisfying the fundamental equations for the base elements in $\mathbb{H}$, such as

$$\mathbf{I^2} = \mathbf{J^2} = \mathbf{K^2} = \mathbf{IJK} = -\mathbf{1},$$

these matrices could be used to represent any quaternion $x_1 + x_2 i + x_3 j + x_4 k$ as

$$\begin{aligned} x_1\mathbf{1} + x_2\mathbf{I} + x_3\mathbf{J} + x_4\mathbf{K} &= x_1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + x_2 \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} + x_3 \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} \\ &= \begin{bmatrix} x_1 + ix_2 & x_3 + ix_4 \\ -x_3 + ix_4 & x_1 - ix_2 \end{bmatrix} \\ &= \begin{bmatrix} z_1 & w_1 \\ -\overline{w_1} & \overline{z_1} \end{bmatrix} \end{aligned}$$

where $z_1 = x_1 + ix_2$ and $w_1 = x_3 + ix_4 \in \mathbb{C}$. Neatly, but not at all coincidentally,

$$\det \begin{bmatrix} x_1 + ix_2 & x_3 + ix_4 \\ -x_3 + ix_4 & x_1 - ix_2 \end{bmatrix} = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

showing that the determinant of this matrix representation of a quaternion is simply equal to its norm.

Standard results from linear algebra could then be used to derive crucial properties of $\mathbb{H}$. For example, the associative, distributive and noncommutative properties of quaternion multiplication followed immediately from the corresponding attributes of matrices. Also, the fact that for any 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$,

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

thus meant that any quaternion z with a non-zero norm would have the inverse

[4] For more detail, see [1], [5].

$$z^{-1} = \frac{\overline{z}}{N(z)}.$$

And importantly, since for any $n \times n$ matrices A and B, $\det AB = \det A \det B$, the fundamental equation (4) immediately followed.

Most crucially, as Dickson would have realized, if $x_1, x_2, y_1, y_2 \in \mathbb{C}$, then multiplication in $\mathbb{H}$ is defined by the matrix product

$$\begin{bmatrix} x_1 & y_1 \\ -\overline{y_1} & \overline{x_1} \end{bmatrix} \begin{bmatrix} x_2 & y_2 \\ -\overline{y_2} & \overline{x_2} \end{bmatrix} = \begin{bmatrix} x_1x_2 - \overline{y_2}y_1 & y_2x_1 + y_1\overline{x_2} \\ -y_2x_1 + y_1\overline{x_2} & x_1x_2 + \overline{y_2}y_1 \end{bmatrix}$$

which is exactly his formula (5). So perhaps this applicability of matrices to the generation of $\mathbb{H}$ from $\mathbb{C}$ is one reason for the association of Cayley's name with this procedure.

On the other hand, as Dickson himself noted, the Cayley-Dickson process arose in direct response to Cayley's foundational work on octonions from 1845. Dickson realized that by taking $\mathbb{H}$, re-labeling $i = i_1, j = i_2, k = i_3$, and defining i_4 to be a new square-root of -1 such that $i_1i_4 = i_5, i_2i_4 = i_6$, and $i_3i_4 = i_7$, he could define the octonions $\mathbb{O}$ as $\mathbb{H} + \mathbb{H}i_4$. Thus if $x_1, x_2, y_1, y_2 \in \mathbb{H}$, then the product

$$(x_1 + y_1i_4)(x_2 + y_2i_4) = (x_1x_2 - \overline{y_2}y_1) + (y_2x_1 + y_1\overline{x_2})i_4$$

would define all multiplication in $\mathbb{O}$.

Of course, there is no need to stop there, and the Cayley-Dickson construction can be used to build further composition algebras of increasing dimension: 16, 32, 64, and so on *ad infinitum*. But there is a problem. As the dimension doubles, key algebraic properties are successively lost, the first of which is trivial conjugation, which clearly holds in $\mathbb{R}$ but not in $\mathbb{C}$. This has a knock-on effect, resulting in the loss of commutativity in multiplication when moving from $\mathbb{C}$ to $\mathbb{H}$, which likewise results in the nonassociativity of $\mathbb{O}$. But things get worse, because at dimension 16 the next property to be lost is the non-existence of zero divisors. In other words, from this point for any Cayley-Dickson algebra $\mathcal{A}$, any element $x \in \mathcal{A}$ could have a nonzero partner $y \in \mathcal{A}$ such that $xy = 0$. This naturally has the consequence that the crucial equation (4) no longer holds in general, meaning that no composition algebras over the reals with dimension 2^n will be normed division algebras if $n > 3$.

This gives us an intuitive idea why the only possible normed algebras over the reals are indeed $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, and $\mathbb{O}$. But it doesn't prove it. Nor does it say anything about the existence (or non-existence) of normed algebras with dimensions other than powers of 2. To resolve the matter conclusively, we need to return to sums of squares and look at Hurwitz's 1898 proof. Our exposition will be based on an

expanded version given by Dickson in his paper of 1919, where he explained the nature of, and the rationale for, his presentation of Hurwitz's Theorem [8, p. 159]:

> Since experience shows that graduate students fail to follow various steps merely outlined by Hurwitz, we shall here give the proof in detailed, amplified form.

But Dickson's exposition is also lacking in certain respects: it gives quite lengthy explanations of some relatively easy portions, while skipping over the more sophisticated details of others. And it would certainly not be fully intelligible to today's undergraduates. We therefore give our own step-by-step, and hopefully intelligible, elucidation of Dickson's expanded proof of Hurwitz's Theorem.

4 Our expanded proof of Dickson's expanded proof of Hurwitz's Theorem

Theorem 4.1 (Hurwitz's Sums of Squares Theorem)**.** *Let n be a positive integer for which there exists an identity of the form* (1)

$$(x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) = z_1^2 + \cdots + z_n^2$$

where $z_k = \sum_{i,j=1}^{n} A_{ijk} x_i y_j$ and the A_{ijk} are constants independent of the values of the x_i and the y_j. Then $n = 1, 2, 4,$ or 8 and no other values.

Proof. Our proof is in six steps.

(1) Quadratic forms and their matrices.

An *n-ary quadratic form* is a homogeneous polynomial of degree 2 in n variables. Quadratic forms with 2, 3, and 4 variables are called *binary, ternary* and *quaternary* quadratic forms, respectively. Thus, $4z_1^2 + z_1 z_2 + 6z_2^2$ is a binary quadratic form and $z_1^2 + z_2^2 + z_3^2 + z_4^2$ is a quaternary quadratic form.

Consider the quadratic form

$$\begin{aligned} F(z) = \sum_{i,j=1}^{n} b_{ij} z_i z_j = {} & b_{11} z_1^2 + b_{12} z_1 z_2 + \ldots + b_{1n} z_1 z_n \\ & + b_{21} z_2 z_1 + b_{22} z_2^2 + \ldots + b_{2n} z_2 z_n \\ & + \ldots \\ & + b_{n1} z_n z_1 + b_{n2} z_n z_2 + \ldots + b_{nn} z_n^2. \end{aligned} \tag{6}$$

The matrix of this quadratic form is $B = [b_{ij}]$, and in the special case when $B = I_n$, $F(z) = z_1^2 + z_2^2 + \ldots + z_n^2$.

If we now substitute $z_i = a_{i1}y_1 + a_{i2}y_2 + \ldots + a_{in}y_n,$, where the a_{ij} are all scalars, then equation (6) becomes

$$\begin{aligned} G(y) = \sum_{i,j=1}^{n} m_{ij}y_iy_j = m_{11}y_1^2 + m_{12}y_1y_2 + \ldots + m_{1n}y_1y_n \\ + m_{21}y_2y_1 + m_{22}y_2^2 + \ldots + m_{2n}y_2y_n \\ + \ldots \\ + m_{n1}y_ny_1 + m_{n2}y_ny_2 + \ldots + m_{nn}y_n^2, \end{aligned} \tag{7}$$

where the m_{ij} are various linear combinations of the a_{ij} and b_{ij}. The matrix of this quadratic form $G(y)$ is A^TBA, where $A = [a_{ij}]$. Hence, if $B = I_n$, then the matrix of the resulting quadratic form is equal to $A^TI_nA = A^TA$.

Now let $M = [m_{ij}] = A^TA$. Once again, if $M = I_n$, then equation (7) becomes $G(y) = y_1^2 + y_2^2 + \ldots + y_n^2$. Thus if we want $\sum_{i=1}^n y_i^2 = \sum_{i=1}^n z_i^2$, this will hold provided $m_{ij} = 1$ if $i = j$ and 0 otherwise, i.e. if $M = I_n$. Now suppose we want $x_1^2\sum_{i=1}^n y_i^2 = \sum_{i=1}^n z_i^2$, where x_1^2 is a scalar. This will hold provided $m_{ij} = x_1^2$ if $i = j$ and 0 otherwise, i.e. if $M = x_1^2 \cdot I_n$.

In a similar way, let x_k^2 be a scalar for $k = 1, 2, \ldots, n$. Then $x_1^2 + x_2^2 + \ldots + x_n^2$ is also a scalar and it follows that

$$(x_1^2 + x_2^2 + \ldots + x_n^2)\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} z_i^2$$

provided $m_{ij} = x_1^2 + x_2^2 + \ldots + x_n^2$ if $i = j$ and 0 otherwise, i.e. if $M = A^TA = (x_1^2+x_2^2+\ldots+x_n^2)I_n$. Hence, the existence of the identity $\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n y_i^2 = \sum_{i=1}^n z_i^2$, i.e. formula (1), is equivalent to the existence of the equality

$$A^TA = (x_1^2 + x_2^2 + \ldots + x_n^2)I_n. \tag{8}$$

For the case $n = 4, (x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) = z_1^2 + z_2^2 + z_3^2 + z_4^2,$ where

$$\begin{aligned} z_1 &= x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4, \\ z_2 &= x_2y_1 + x_1y_2 - x_4y_3 + x_3y_4, \\ z_3 &= x_3y_1 + x_4y_2 + x_1y_3 - x_2y_4, \\ z_4 &= x_4y_1 - x_3y_2 + x_2y_3 + x_1y_4, \end{aligned} \tag{9}$$

and the relevant matrix A is given by[5]

$$A = \begin{bmatrix} x_1 & -x_2 & -x_3 & -x_4 \\ x_2 & x_1 & -x_4 & x_3 \\ x_3 & x_4 & x_1 & -x_2 \\ x_4 & -x_3 & x_2 & x_1 \end{bmatrix}.$$

(2) The matrices A_i and relations among them.

Notice that the matrix A can be written as the linear combination

$$A = x_1A_1 + \ldots + x_nA_n, \tag{10}$$

where the entries of the A_i are $0, 1$, or -1. In our example above, for instance, we may write A as

$$x_1\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} + x_2\begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} + x_3\begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} + x_4\begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

where the A_i's are the matrices with real coefficients x_i.

Substituting the linear combinations of type (10) for A and A^T into equation (8) leads to

$$(x_1A_1^T + \ldots + x_nA_n^T)(x_1A_1 + \ldots + x_nA_n) = (x_1^2 + x_2^2 + \ldots + x_n^2)I_n \tag{11}$$

Expanding the left side of the above equation gives us

$$\sum_{i,j=1,\ldots,n} x_ix_jA_i^TA_j = \left(\sum_{i=1}^{n} x_i^2\right)I_n.$$

The coefficient of $A_i^TA_j$ is x_ix_j; since $x_ix_j = x_jx_i$, it follows that

$$A_i^TA_i = I_n \text{ for } i = 1, \ldots, n, \text{ and } A_i^TA_j + A_j^TA_i = 0.$$

But what about about the individual A_i's? Hurwitz used the equalities $A_i^TA_i = I_n = A_iA_i^T$ to make a change of variables that replaces the terms x_nA_n and $x_nA_n^T$ by the term x_nI_n. He did this by defining new matrices $B_i = A_n^TA_i$ for $i = 1, \ldots, n-1$.

[5]The equivalence of formulae (2) and (9) is not coincidental!

Then for $i < n$, $A_i A_i^T = I_n$ implies $A_i = A_n A_n^T A_i = A_n B_i$, and so $A_i^T = B_i^T A_n^T$. When we make these substitutions in equation (11) and un-distribute the factors A_n^T and A_n, the result is as follows:

$$\begin{aligned}(x_1^2 + \ldots + x_n^2) I_n &= \left(\sum_{i=1}^{n-1} x_i A_i^T + x_n A_n^T\right)\left(\sum_{i=1}^{n-1} x_i A_i + x_n A_n\right) \\ &= \left(\sum_{i=1}^{n-1} x_i B_i^T + x_n I_n\right) A_n^T A_n \left(\sum_{i=1}^{n-1} x_i B_i + x_n I_n\right) \qquad (12) \\ &= \left(\sum_{i=1}^{n-1} x_i B_i^T + x_n I_n\right)\left(\sum_{i=1}^{n-1} x_i B_i + x_n I_n\right).\end{aligned}$$

This time, the coefficients of $x_i x_j$ are I_n if $i = j$ and 0 otherwise, and the coefficients of $x_i x_n$ are 0 for $i < n$. Hence, for $i, j = 1, \ldots, n-1$,

(a) $B_i^T B_i = I_n$,
(b) $B_i^T B_j + B_j^T B_i = 0$ for $i \neq j$,
(c) $B_i + B_i^T = 0$.

From (a) and (c) we deduce that $B_i^2 = -B_i^T B_i = -I_n$. Using (b) and (c), we deduce that $B_i^T B_j + B_j^T B_i = -B_i B_j - B_j B_i = 0$, and hence $B_i B_j = -B_j B_i$.[6]

(3) Symmetric and skew-symmetric matrices

We now need two definitions. A square matrix M is *symmetric* provided that $M^T = M$, and *skew-symmetric* provided $M^T = -M$. Thus, $B_i^T = -B_i$ and $(B_i B_j)^T = B_j^T B_i^T = (-1)^2 B_j B_i = B_j B_i = -B_i B_j$. Hence the B_i and the $B_i B_j$ are skew-symmetric. As for a triple product,

$$(B_i B_j B_k)^T = B_k^T B_j^T B_i^T = (-1)^3 (B_k B_j B_i) = (-1)^3 (-1^3) B_i B_j B_k = B_i B_j B_k,$$

because the three substitutions $B_r^T = -B_r$ contribute a factor of $(-1)^3$ to the product. In addition, it takes three pairwise adjacent swaps to reverse $B_k B_j B_i$, each such swap contributes a factor of -1, and so the "swapping" contributes $(-1)^3$ to the product. Hence, $(B_i B_j B_k)^T = (-1)^{3+3} B_i B_j B_k = B_i B_j B_k$, and so a product of three distinct B_i's is symmetric.

More generally, it takes $\binom{r}{2}$ such swaps to reverse a product of r distinct factors, and there is a straightforward inductive proof of this fact. Thus, if $1 \leq i_1 < i_2 < \ldots < i_r \leq n-1$, then

$$(B_{i_1} B_{i_2} \cdots B_{i_r})^T = B_{i_r}^T \cdots B_2^T B_1^T = (-1)^r B_{i_r} \cdots B_2 B_1 = (-1)^{r+\binom{r}{2}} B_{i_1} B_{i_2} \cdots B_{i_r}.$$

[6] Given any distinct B_i, B_j, B_k, the fact that $B_i^2 = B_j^2 = B_k^2 = B_i B_j B_k = -I_n$ is not a coincidence either!

Now, $r + \binom{r}{2} = \binom{r+1}{2} = \frac{(r+1)r}{2}$, and this number is even if $r \equiv 0$ or $3 \mod 4$, and odd if $r \equiv 1$ or $2 \mod 4$. Thus, a product of $1, 2, 5, 6, \ldots$ B_i's is skew-symmetric, and a product of $3, 4, 7, 8, \ldots$ B_i's is symmetric – and so is the identity matrix I_n.

The skew-symmetry of the individual B_i's also reveals a highly significant fact. Since the B_i's are $n \times n$ matrices such that $B_i^T = -B_i$, this means that, for $n > 1$, $\det B_i = (-1)^n \det B_i$, meaning that if n is odd, then $\det B_i = 0$. But since $B_i^2 = -I_n$, $\det B_i \neq 0$. Therefore, apart from the trivial case when $n = 1$, n cannot be odd. And in what follows, that fact is going to be crucial.

Now for an important result.

(4) Linear independence of a set of 2^{n-2} $n \times n$ matrices.

Proposition 1. *At least half of the 2^{n-1} matrices in*

$$S = \{I, B_{i_1}, B_{i_1}B_{i_2}, \cdots, B_{i_1}B_{i_2}\cdots B_{i_r} | 1 \leq i_1 < i_2 < \ldots < i_r \leq n-1\}$$

form a linearly independent set.

Proof. First note that $|S| = 2^{n-1}$ since any element of S either does or does not contain $B_1, B_2, \ldots, B_{n-1}$. Let R be any linear combination of the matrices in S involving constants $a_1, a_2, \ldots$ — not all zero — such that $R = 0$. Clearly, this means that the matrices in R are linearly dependent. We call this linear dependecy *irreducible* if R cannot be written as $R_1 + R_2$, where $R_1 = 0, R_2 = 0$, and no element of S belongs to both R_1 and R_2.

In particular, an irreducible linear combination R cannot contain both symmetric and skew-symmetric matrices—and this is the key to proving this proposition. To see this, suppose the contrary. This would mean that $R = 0$ could be rewritten as $M = K$, where M and K are linear combinations of only symmetric and skew-symmetric matrices, respectively. Thus,

$$M = M^T = K^T = -K = -M, \text{ implying } M = -M = 0 \text{ and so } K = 0.$$

Next, we show that multiplication by any number of matrices B_i permutes the 2^{n-1} members of

$$S = \{I, B_{i_1}, B_{i_1}B_{i_2}, \cdots, B_{i_1}B_{i_2}\cdots B_{i_r} | i_1 < i_2 < \cdots < i_r \leq n-1\}$$

if we ignore sign changes and orderings. The reason is that $B_i^2 = -I_n$ so that B_i is an involution (up to sign) on the set S. Multiplication by B_i permutes the products in S by mapping those products in S that contain the factor B_i to the products in S that do not have the factor B_i and vice-versa. By induction (and ignoring signs

and orderings), multiplication by a product of matrices in S is a permutation of the elements of S.

We have already proved that for $n > 1$, n cannot be odd, so we now assume that n is even. An irredicible linear dependency $R = \sum_{i=1}^{r} a_i S_i = 0$, where S_i is a product of the B_j's and the a_i are nonzero, can be rewritten as $I_n = \sum c_i T_i$ by multiplying $R = 0$ through by the inverse of one term and re-arranging the equation.[7] I_n is symmetric, and so the T_i must also be symmetric. Thus $T_i = B_{i_1} \cdots B_{i_j}$ is a product of j factors, with $j \equiv 0$ or $3 \mod 4$.

If T_i has $4k$ factors, multiplying $I_n = \sum c_i T_i$ through by B_{i_1} yields the skew-symmetric B_{i_1} on the left-hand side and a symmetric term on the right. Thus, no T_i can be a product of $4k$ terms.

If T_i has $4k + 3$ factors and this number is less than $n - 1$, then multiplying $I_n = \sum c_i T_i$ through by a non-factor B_j yields the skew-symmetric B_j on the left-hand side and a symmetric term on the right. Thus, no T_i can be a product of $4k+3$ terms with $4k + 3 < n - 1$.

So the only possible linear dependency must be of the form $I_n = cB_1 \cdots B_{n-1}$ for some constant c, where $n - 1 \equiv 3 \mod 4$. Hence, $n \equiv 0 \mod 4$. As for c, we know that $I_n^2 = c^2(B_1 \cdots B_{n-1})^2$. As we have seen, it takes $\binom{r}{2}$ adjacent swaps to reverse a product of r distinct factors. Hence, since $n \equiv 0 \mod 4$, we see that

$$\begin{aligned}(B_1 \cdots B_{n-1})^2 &= (-1)^{\binom{n-1}{2}}(B_1 \cdots B_{n-1})(B_{n-1} \cdots B_1) \\ &= (-1)^{\binom{n-1}{2}+n-1} I_n^{n-1} \\ &= (-1)^{\binom{n}{2}} I_n \\ &= I_n.\end{aligned}$$

Therefore, $I_n = c^2 \cdot I_n$, and so $c = \pm 1$. We thus see that the only possible linear dependencies are of the form $I_n = (\pm 1)(B_1 \cdots B_{n-1})$—where $n \equiv 0 \mod 4$—and those obtained by multiplying this equation by matrices from the set S. These dependencies take the form of equalities between two products, one of which contains more than half of the B_i's and the other, fewer than half. This means that if $n \equiv 0 \mod 4$, then half of the 2^{n-1} matrices in the set S will be linearly independent.

Finally, we note that if $n \equiv 2 \mod 4$, then there can be no linear dependencies, and so all 2^{n-1} of the products in S will be linearly independent.

Thus, the set S contains at least 2^{n-2} linearly independent $n \times n$ matrices. □

We are almost finished . . .

[7] Notice that this new relation $I_n = \sum c_i T_i$ is also irreducible, since we can multiply it through by a suitable constant c_i and matrix B_j to return us to our original irreducible $R = 0$.

(5) The $(1, 2, 4, 6, 8)$ Restriction.

Proposition 2. *If n satisfies the conditions of Proposition 1, then $n = 1, 2, 4, 6,$ or 8.*

Proof. The n^2 distinct $n \times n$ matrices whose entries are all zeros except for a single 1 are clearly linearly independent. The 2^{n-2} linearly independent matrices from the set S are a subset of these n^2 matrices. Hence $2^{n-2} \leq n^2$. For $n \leq 8$, this inequality is true by inspection, and we show by induction that the inequality is false if $n > 8$.

For $n = 9$, $2^{n-2} = 2^7 = 128 > 81 = 9^2 = n^2$, so the inequality is false for $n = 9$.

Next, assume that $n \geq 9$ and that $2^{n-2} > n^2$. Then $2^{n-1} = 2 \cdot 2^{n-2} > 2 \cdot n^2 = n^2 + n^2 > n^2 + 2n + 1 = (n+1)^2$, because $n^2 > 2n + 1$ whenever $n \geq 3$. It follows that $2^{n-2} > n^2$ for all $n \geq 9$.

Since we know that for $n > 1$, n cannot be odd, this leaves only the possibilities $1, 2, 4, 6$ and 8. □

We have now proved that the existence of the sums-of-n-squares identity (1) is equivalent to the existence of the identity (8), which itself is equivalent to (12). This last equation holds if and only if there exist skew-symmetric matrices B_i such that 2^{n-2} of their 2^{n-1} possible products are linearly independent. This is equivalent to saying that $2^{n-2} \leq n^2$, which is only true if $n = 1, 2, 4, 6$, or 8. There is thus only one thing now left to do.

(6) Elimination of the $n = 6$ case.

We finally eliminate $n = 6$. To begin with, $6 \equiv 2 \bmod 4$, so that if a sums-of-six-squares identity exists, then all $32 = 2^5 = 2^{6-1}$ of the relevant matrices form a linearly independent set. Sixteen of these matrices — namely, the five B_i, the ten B_iB_j, and the product $B_1B_2B_3B_4B_5$ — are skew-symmetric. Call these matrices $M_1, M_2, \ldots, M_{16}$, and let a_{ij}^k be the $(i,j)^{\text{th}}$ entry of M_k.

Suppose there exist constants $c_1, c_2, \ldots, c_{16}$ — not all zero — such that

$$c_1M_1 + c_2M_2 + \ldots + c_{16}M_{16} = 0.$$

Then for each of the 36 pairs (i, j) and each k,

$$c_1a_{i,j}^1 + c_2a_{i,j}^2 + \ldots + c_{16}a_{i,j}^{16} = 0.$$

However, the M_k are skew-symmetric. This has two consequences:

(a) For all i and k, $a_{i,i}^k = -a_{i,i}^k = 0$, so the six linear combinations when $i = j$ are identically zero.

(b) For all i, j, k with $i \neq j$, $a^k_{i,j} = -a^k_{j,i}$, so the 15 linear combinations for $i < j$ are the negatives of the 15 linear combinations for $i > j$, and so they contribute nothing new.

Hence, there are only 15 distinct linear equations relating these 16 matrices, and a system of 15 homogeneous linear equations in 16 unknowns — namely, $c_1, c_2, \ldots, c_{16}$ — has more unknowns than equations ... and so it has nontrivial solutions. In short, the 16 skew-symmetric matrices are linearly dependent, contrary to the assumption that the 32 relevant matrices are linearly independent. Hence, there is no sums-of-six-squares identity.

Thus, the sums-of-n-squares identity (1) exists for $n = 1, 2, 4,$ and 8 and for no other positive integers n. This completes the proof of Hurwitz's Theorem. □

References

[1] Baez, J. C. (2002). The octonions. *Bull. Amer. Math. Soc.* 39: 145–205.

[2] Brown, E., Rice, A. (2020). Why Hamilton couldn't multiply triples, to appear.

[3] Cayley, A. (1845). On Jacobi's elliptic functions ... and on quaternions. *Phil. Mag.* 26: 208–211; *Collected Math. Papers* 1: 127.

[4] Cayley, A. (1858). A memoir on the theory of matrices. *Phil. Trans. Roy. Soc. London* 148: 17–37; *Collected Math. Papers* 2: 475–496.

[5] Conway, J. H., Smith, D. A. (2003). *On quaternions and octonions: their geometry, arithmetic, and symmetry.* A K Peters/CRC Press.

[6] Dickson, L. E. (1912). Linear algebras, *Trans. Amer. Math. Soc.* 13: 59–73.

[7] Dickson, L. E. (1914). *Linear algebras.* Cambridge: Cambridge University Press.

[8] Dickson, L. E. (1919). On quaternions and their generalization and the history of the eight square theorem, *Ann. Math.* (2) 20: 155–171.

[9] Hamilton, W. R. (1835). Theory of conjugate functions, or algebraic couples; with a preliminary and elementary essay on algebra as the science of pure time. *Trans. Roy. Irish Acad.* 17: 293–422; *Mathematical Papers* 3: 76–96.

[10] Hamilton, W. R. (1967). *The Mathematical Papers of Sir William Rowan Hamilton*, vol. 3, ed. H. Halberstam and R. E. Ingram. Cambridge: Cambridge University Press.

[11] Hurwitz, A. (1898). Über die Composition der quadratischen Formen von beliebig vielen Variablen. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 309–316; *Mathematische Werke* 2: 565–571.

[12] Rice, A., Brown, E. (2016). Commutativity and collinearity: a historical case study of the interconnection of mathematical ideas. Part I. *BSHM Bulletin: Jour. Brit. Soc. Hist. Math.* 31: 1–14.