Products of Sums of Squares

Lecture notes for a mini-course at the Universidad de Talca, December 1999.

Daniel Shapiro [1]

Lecture 1: Introduction and History

The set of squares is closed under multiplication because: $x^2y^2 = z^2$, where $z = xy$. Similarly the sums of two squares also form a multiplicatively closed set, because:

$$(x_1^2 + x_2^2) \cdot (y_1^2 + y_2^2) = z_1^2 + z_2^2$$

where $z_1 = x_1y_1 - x_2y_2$ and $z_2 = x_2y_1 + x_1y_2$.

In this example z_1 and z_2 are bilinear forms in $X = (x_1, x_2)$ and $Y = (y_1, y_2)$. This 2-square identity can be interpreted as the "law of moduli" for complex numbers: $|\alpha\beta| = |\alpha| \cdot |\beta|$, where $\alpha = x_1 + ix_2$ and $\beta = y_1 + iy_2$.

In 1748 Euler used a 4-square identity in his attempt to prove that every positive integer is a sum of four squares. Here is aversion of Euler's identity: $(x_1^2 + x_2^2 + x_3^2 + x_4^2) \cdot (y_1^2 + y_2^2 + y_3^2 + y_4^2) = z_1^2 + z_2^2 + z_3^2 + z_4^2$ where

$$z_1 = x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4$$

$$z_2 = x_2y_1 + x_1y_2 - x_4y_3 + x_3y_4$$

$$z_3 = x_3y_1 + x_4y_2 + x_1y_3 - x_2y_4$$

$$z_4 = x_4y_1 - x_3y_2 + x_2y_3 + x_1y_4.$$

After Hamilton's discovery of the quaternion algebra in 1843, this formula was interpreted as the law of moduli for quaternions. In 1848 the octonion algebra was discovered by Graves and Cayley, providing a similar 8-square identity.

For the next 50 years various mathematicians tried to construct a 16-square identity, and began to suspect that none can exist. Finally in 1898 Hurwitz

[1]Partially supported by a grant from the Fundación Andes. It is also a pleasure to thank the Universidad de Talca and my colleagues there for their generous hospitality.

proved that result using ideas from linear algebra. To discuss Hurwitz's ideas we make a basic definition.

Definition 1. *A* **composition formula** *of size* $[r, s, n]$ *is a formula of the type*

$$(x_1^2 + x_2^2 + \cdots + x_r^2) \cdot (y_1^2 + y_2^2 + \cdots + y_s^2) = z_1^2 + z_2^2 + \cdots + z_n^2$$

where $X = (x_1, \ldots, x_r)$ *and* $Y = (y_1, \ldots, y_s)$ *are systems of indeterminates and each* $z_k = z_k(X, Y)$ *is a bilinear form in* X *and* Y*, with coefficients in a field* F*.*

Let's view X, Y and Z as column vectors. Then $z_1^2 + z_2^2 + \cdots + z_n^2 = Z^\mathsf{T} Z$, where "$\mathsf{T}$" denotes the transpose. Since Z is linear in Y we have $Z = AY$ where A is an $n \times s$ matrix with entries in $F(X)$. The composition formula then becomes: $(\sum x_i^2) Y^\mathsf{T} Y = Z^\mathsf{T} Z = Y^\mathsf{T} A^\mathsf{T} A Y$. Since Y consists of indeterminates, this equation is equivalent to

$$A^\mathsf{T} A = (x_1^2 + x_2^2 + \cdots + x_r^2) I_s.$$

But Z is also linear in X, so the entries of A are linear forms in X. Therefore $A = x_1 A_1 + \cdots + x_r A_r$, where each A_i is an $n \times s$ matrix with constant entries. Substitute this into the equation above and cancel like terms to find:

There are $n \times s$ matrices $A_1, \ldots, A_r$ over F satisfying

$$A_i^\mathsf{T} A_i = I_s,$$

$$A_i^\mathsf{T} A_j + A_j^\mathsf{T} A_i = 0 \quad \text{whenever} \quad i \neq j.$$

This system of "Hurwitz Matrix Equations" exists if and only if there is a composition formula over F of size $[r, s, n]$. Since rectangular matrices are hard to deal with, let's restrict to square matrices: assume $s = n$. In that special case the system is greatly simplified by defining the $n \times n$ matrices $B_i = A_1^{-1} A_i$. Then $B_1, \ldots, B_r$ satisfy the equations above and $B_1 = I_n$. Therefore, for i and j ranging from 2 to r:

$$B_i^\mathsf{T} = -B_i$$

$$B_i^2 = -I_n,$$

$$B_iB_j + B_jB_i = 0 \quad \text{whenever} \quad i \neq j.$$

Such a system of $r-1$ $n \times n$ matrices exists if and only if there is a composition formula of size $[r, n, n]$. The existence of such matrices puts some restrictions on n. For example,

Lemma 2. *If there exists a composition of size $[r, n, n]$, then:*

$$r \geq 2 \quad \textit{implies} \quad 2 \mid n.$$

$$r \geq 3 \quad \textit{implies} \quad 4 \mid n.$$

$$2^{r-2} \leq n^2.$$

Proof. See Exercise 2 below for an outline of the ideas. □

This information suffices to prove the Hurwitz Theorem of 1898.

Corollary 3. *A composition of size $[n, n, n]$ implies n = 1, 2, 4, or 8.*

These ideas were pushed further by Hurwitz (1923)[2] over the complexes, and by Radon (1922) over the reals. They determined all the possible sizes for compositions of type $[r, n, n]$. The answer involves the so-called Hurwitz-Radon function $\rho(n)$ determined by the following rules:

If $n = 1$, 2, 4 or 8 then $\rho(n) = n$.

If k is odd, then $\rho(2^m k) = \rho(2^m)$.

$\rho(16n) = 8 + \rho(n)$.

Theorem 4. *(Hurwitz-Radon) There exists a composition of size $[r, n, n]$ if and only if $r \leq \rho(n)$.*

This theorem generalizes the older result: $\rho(n) = n \iff n = 1$, 2, 4, or 8. The proof involves a careful analysis of those anti-commuting square matrices B_j. Hurwitz considered those matrices to have complex entries, but his ideas work just as well using any field of coefficients, provided that $2 \neq 0$ in F.

[2]Hurwitz died in 1919.

After the work of Hurwitz and Radon in the 1920s several mathematicians found new proofs of their theorem. For example Eckmann (1943) used the representation theory of certain finite groups to prove the theorem over $\mathbb{R}$, and Dubisch (1946) and Lee (1948) used the representation theory of Clifford algebras. However the first results about compositions of size $[r, s, n]$ when $r, s < n$ were obtained by topologists.

Some further history might help explain the connection with topology. We have seen that an n-dimensional algebra for which the norm is multiplicative ($|xy| = |x| \cdot |y|$) provides a composition formula of size $[n, n, n]$. The classical examples, $\mathbb{R}$ (reals), $\mathbb{C}$ (complexes), $\mathbb{H}$ (quaternions), and $\mathbb{O}$ (octonions), are real division algebras of dimensions 1, 2, 4, and 8. Can there be any other dimensions for division algebras over $\mathbb{R}$? We consider only finite dimensional real division algebras.

Frobenius (1877) proved that if A is an associative real division algebra with 1, then $A \cong \mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. Zorn (1933) investigated alternative rings (these satisfy a weak form of associativity: $x \cdot xy = xx \cdot y$ and $y \cdot xx = yx \cdot x$) and proved that the alternative real division algebras with 1 are exactly the four standard examples. Still weaker algebraic properties have been analyzed, like the "flexible property" ($x \cdot yx = xy \cdot x$). But what if all assumptions of associativity are dropped? Define a *real division algebra* to be simply a vector space $A \cong \mathbb{R}^n$ with a bilinear multiplication which admits no nontrivial zero divisors. Is its dimension restricted to 1, 2, 4 or 8?

We make a more general definition.

Definition 5. *Let f be a real bilinear map of size $[r, s, n]$. That is, $f : \mathbb{R}^r \times \mathbb{R}^s \to \mathbb{R}^n$.*

f is **normed** *if $|f(x, y)| = |x| \cdot |y|$ for every $x \in \mathbb{R}^r$ and $y \in \mathbb{R}^s$.*

f is **nonsingular** *if $f(x, y) = 0$ implies that either $x = 0$ or $y = 0$.*

Then a composition formula is exactly a normed bilinear map of that size. Since $|x| = 0$ implies $x = 0$, every normed map is nonsingular. A nonsingular bilinear map of size $[n, n, n]$ is exactly an n-dimensional real division algebra. (We can always arrange an identity element. See Exercise 4.)

In the late 1930s the topologists Stiefel and Hopf attacked the problem of real division algebras. They noted that a nonsingular bilinear map $\mathbb{R}^r \times \mathbb{R}^s \to \mathbb{R}^n$

induces a map of real projective spaces $\mathbb{P}^{r-1} \times \mathbb{P}^{s-1} \to \mathbb{P}^{n-1}$. Hopf applied his newly constructed theory of the ring structure of cohomology, while Stiefel (a former student of Hopf) applied his characteristic classes of vector bundles. Both methods led to the same result, published in 1940.

Hopf-Stiefel Theorem 6. *If there exists a nonsingular bilinear map of size* $[r, s, n]$ *over* $\mathbb{R}$, *then*

$$(x+y)^n = 0 \quad \textit{in the ring} \quad \mathbb{F}_2[x,y]/(x^r, y^s).$$

Here $\mathbb{F}_2 \cong \mathbb{Z}/2\mathbb{Z}$ is the field of two elements. By the Binomial Theorem this criterion can also be stated as: $\binom{n}{k}$ is even whenever $n - s < k < r$. This Hopf-Stiefel criterion is fairly easy to apply. For example, a nonsingular bilinear $[3, 5, 6]$ cannot exist over $\mathbb{R}$ because $\binom{6}{3}$ is odd.

Corollary 7. *(Hopf-Stiefel) If there is an n-dimensional real division algebra then n is a power of 2.*

Proof. Check that $(x+y)^n = x^n + y^n$ in $\mathbb{F}_2[x,y]$ if and only if n is a power of 2. See Exercise 4(a). □

Of course the expectation was that the only possible division algebra dimensions here are 1, 2, 4 and 8. That was finally proved in 1958 as a consequence of the Bott Periodicity Theorem, and those ideas were later incorporated into topological K-theory. Topologists studied nonsingular bilinear maps for other reasons as well. For example a map of size $[r, n, n]$ provides a system of $r-1$ linearly independent vector fields on the sphere S^{n-1}. A nonsingular $[r, r, n]$ provides an immersion of $\mathbb{P}^{r-1}$ into euclidean space $\mathbb{R}^{n-1}$. Nonsingular bilinear maps are also involved in the analysis of the homotopy groups of spheres.

Let's summarize what has been discussed so far by introducing some new notations.

Definition 8. *Let* r, s *be positive integers.*

$$r \circ s = \min\{n : (x+y)^n = 0 \textit{ in } \mathbb{F}_2[x,y]/(x^r, y^s)\}.$$

$$r \# s = \min\{n : \textit{there is a nonsingular bilinear } [r,s,n] \textit{ over } \mathbb{R}\}.$$

$$r * s = \min\{n : \textit{there is a normed bilinear } [r,s,n] \textit{ over } \mathbb{R}\}.$$

We already know some basic facts about these functions:

$$r \circ s \quad \leq \quad r \# s \quad \leq \quad r * s.$$

$$r * n = n \quad \text{if and only if} \quad r \leq \rho(n).$$

The first inequality here is a restatement of the Hopf-Stiefel Theorem. The second statement is the Hurwitz-Radon Theorem.

The values of $r \# s$ and $r * s$ are known for only a few values of r, s. For example, if there exists a composition formula of size $[r, s, r \circ s]$ then all three values are equal. Such constructions can be done whenever r is small:

Lemma 9. *If* $r \leq 9$ *then* $r * s = r \# s = r \circ s$.

Similarly $10 \circ 10 = 16$ and there is a normed $[10, 10, 16]$, as we will see in Lecture 2. Therefore $10 * 10 = 10 \# 10 = 10 \circ 10 = 16$. On the other hand, for larger values the lower bound $r \circ s$ is not sharp. For example, $16 \circ 16 = 16$, but Hurwitz-Radon implies $16 * 16 > 16$. In fact, K. Y. Lam (1972), (1985) proved that $16 \# 16 = 23$ while $16 * 16 > 23$. It is easy to construct a composition of size $[16, 16, 32]$ and it is conjectured that $16 * 16 = 32$. The best result known in this direction is due to Lam and Yiu (1989): $29 \leq 16 * 16 \leq 32$.

To clarify the Hopf-Stiefel result, and for future use, let's investigate those values $r \circ s$ in more detail. For small values of r and s it is not hard to check the binomial coefficients to determine the value of $r \circ s$. A table of some of the small values is given below.

When I look at this table of the values of $r \circ s$, some patterns seem to jump out. For example,

- The rows and columns are non-decreasing.
- The entries 2, 4, 8, 16 occur in triangular patterns.
- The upper left 2×2 square, with various constants added, is repeated in several places. Similar patterns hold for the upper left 4×4 square.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	2	2	4	4	6	6	8	8	10	10	12	12	14	14	16	16	18
3	3	4	4	4	7	8	8	8	11	12	12	12	15	16	16	16	19
4	4	4	4	4	8	8	8	8	12	12	12	12	16	16	16	16	20
5	5	6	7	8	8	8	8	8	13	14	15	16	16	16	16	16	21
6	6	6	8	8	8	8	8	8	14	14	16	16	16	16	16	16	22
7	7	8	8	8	8	8	8	8	15	16	16	16	16	16	16	16	23
8	8	8	8	8	8	8	8	8	16	16	16	16	16	16	16	16	24
9	9	10	11	12	13	14	15	16	16	16	16	16	16	16	16	16	25
10	10	10	12	12	14	14	16	16	16	16	16	16	16	16	16	16	26
11	11	12	12	12	15	16	16	16	16	16	16	16	16	16	16	16	27
12	12	12	12	12	16	16	16	16	16	16	16	16	16	16	16	16	28
13	13	14	15	16	16	16	16	16	16	16	16	16	16	16	16	16	29
14	14	14	16	16	16	16	16	16	16	16	16	16	16	16	16	16	30
15	15	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	31
16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	32
17	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	32

Table of values of $r \circ s$ when $r, s \leq 17$.

These observations can be formulated algebraically as follows.

Proposition 10. *$r \circ s$ is a commutative binary operation on $\mathbb{Z}^+$.*

(1) If $r \leq r'$ then $r \circ s \leq r' \circ s$.

(2) $r \circ s = 2^m$ if and only if $r, s \leq 2^m$ and $r + s > 2^m$.

(3) If $r \leq 2^m$ then $r \circ (s + 2^m) = (r \circ s) + 2^m$.

These properties are not hard to prove directly from the definition of $r \circ s$. Moreover the $r \circ s$ function is uniquely determined by these rules.

In the remaining lectures we will mention how this same function $r \circ s$ arises in quite different ways in the study of composition formulas.

Exercises.

EXERCISE 1. (a) Check that for the original $[2,2,2]$ formula the matrix A is $\begin{bmatrix} x_1 & -x_2 \\ x_2 & x_1 \end{bmatrix}$. Then $A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $A_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

(b) Write the 4×4 matrix A corresponding to Euler's $[4,4,4]$ identity.

(c) Construct an 8-square identity. This can be viewed as an 8×8 matrix A with orthogonal rows, where each row is a signed permutation of $(x_1, \dots, x_8)$. (Another method is described in Lecture 2.)

EXERCISE 2. **Proof of the 1, 2, 4, 8 Theorem**.

Suppose F is a field in which $2 \neq 0$, and let $V = F^n$. (a) View an $n \times n$ matrix A as a mapping $V \to V$. If $A^{\mathsf{T}}A = I_n$ then A is an isometry of V, preserving lengths and angles (using the standard scalar product). If also $A^{\mathsf{T}} = -A$ show that Av is always perpendicular to v. Prove that there is an orthogonal basis of V of the type $\{v_1, Av_1, v_2, Av_2, \dots\}$. Consequently n must be even.

(b) Suppose $A, B \in GL_n(F)$ are skew-symmetric isometries, as in part (a) and suppose $AB = -BA$. Show that there is an orthogonal basis of V of the type $\{v_1, Av_1, Bv_1, ABv_1, v_2, Av_2, Bv_2, ABv_2, \dots\}$. Consequently n must be a multiple of 4.

(c) Suppose $C_1, \dots, C_k$ are nonsingular $n \times n$ matrices which anticommute in pairs: $C_iC_j = -C_jC_i$ whenever $i \neq j$. Ignoring scalar multiples, there are 2^k matrices obtained as products of distinct C_j's. If k is even, prove that these 2^k matrices are linearly independent.

(d) Complete the proof of the Hurwitz 1, 2, 4, 8 Theorem.

(Hints. (a) Recall that the scalar product $\langle x, y\rangle$ satisfies $\langle Ax, y\rangle = \langle x, A^{\mathsf{T}}y\rangle$. Choose $v_1 \in V$ with nonzero length: $\langle v_1, v_1\rangle \neq 0$. Choose v_2 with nonzero length in $\{v_1, Av_1\}^{\perp}$, the orthogonal complement. Etc.

(b) Choose $v_1 \in V$ with nonzero length. Then v_1, Av_1, Bv_1, ABv_1 are mutually perpendicular. Choose v_2 with nonzero length in $\{v_1, Av_1, Bv_1, ABv_1\}^{\perp}$. Etc.

(c) If false, choose a dependence relation with the minimal number of nonzero terms. We may assume that I_n is one of the terms. Conjugate by C_j to get another relation which must be identical to the original (otherwise their difference is a shorter relation). Then each nonzero term must commute with every C_j.)

EXERCISE 3. **Inverses, but not a division algebra.**

Let $A = \mathbb{R}^3$ with basis $\{1, i, j\}$. Fix an element $\omega \in A$ and define a bilinear multiplication on A by requiring 1 to be an identity element, $i^2 = j^2 = -1$ and $ij = -ji = \omega$. If $\alpha = a + bi + cj$ define $\overline{\alpha} = a - bi - cj$.

(a) Show that $\alpha\overline{\alpha} = \overline{\alpha}\alpha = a^2 + b^2 + c^2$.

(b) If $\omega \neq 0$ show that $\alpha\beta = \beta\alpha = 0$ implies $\alpha = 0$ or $\beta = 0$.

(c) Deduce that every nonzero element of A has a (2-sided) inverse, and that this inverse is unique. However A is not a division algebra.

EXERCISE 4. **The operation r ∘ s**

(a) Express $n = 2^m n_0$ where n_0 is odd. Then: $r \circ n = n \iff r \leq 2^m$.
In particular, $n \circ n = n \iff n$ is a power of 2.

(b) $r \circ (s + s') \leq r \circ s \ + \ r \circ s'$, and similarly for $\#$ and $*$.

(c) If $r \leq \max\{\rho(s), \rho(s-1), \ldots, \rho(s-9)\}$ then $r \circ s = r \# s = r * s$.
For example, $10 * s = 10 \circ s$ for any $s \equiv 0, 1, \ldots, 9 \pmod{32}$.

(Hints. (a) $(x + y)^n = (x^{2^m} + y^{2^m})^{n_0}$ in $\mathbb{F}_2[x, y]$.

(c) Use Lemma 9. Given $r * (s - k) = s - k$ where $0 \leq k \leq 9$. Then $r * s \ \leq \ s - k \ + \ r * k \ = s - k \ + \ r \circ k \ = r \circ s$.)

EXERCISE 5. **Nonsingular maps.**

(a) If there is a nonsingular bilinear $[r, s, n]$ then $r \# s \leq n$. Moreover if $r \# s = n$ that map must be surjective.

(b) Let $\mathcal{P}_n = \{\text{real polynomials of degree } < n\}$. Then $\dim \mathcal{P}_n = n$ and multiplication provides a nonsingular bilinear map $c_{r,s} : \mathcal{P}_r \times \mathcal{P}_s \to \mathcal{P}_{r+s-1}$. Hence $r \circ s \leq r \# s \leq r + s - 1$.

(c) For which r, s does $r \circ s = r + s - 1$? In those cases we know that $r \circ s = r \# s$.

(d) For which r, s is that map $c_{r,s}$ surjective?

(Hints. (a) If the given map f is not surjective, choose a line $L \subseteq \mathbb{R}^n$ with $L \cap \text{image}(f) = 0$ and consider $\overline{f} : \mathbb{R}^r \times \mathbb{R}^s \to \mathbb{R}^n/L$.

(c) It depends on the dyadic (base 2) expansions of $r - 1$ and $s - 1$.)

References

[1] L. E. Dickson, On quaternions and their generalizations and the history of the eight square theorem, *Ann. Math.* **20** (1919) 155-171.

[2] A. Hurwitz, Über die Komposition der quadratischen Formen von beliebig vielen Variabeln, *Nach. Ges. der Wiss. Göttingen*, (1898) 309-316.

[3] A. Hurwitz, Über die Komposition der quadratischen Formen, *Math. Ann.* **88** (1923) 1-25.

[4] H. Hopf, Ein topologischer Beitrag zur reellen Algebra, *Comment. Math. Helv.* **13** (1940-41) 219-239.

[5] K. Y. Lam and P. Yiu, Geometry of normed bilinear maps and the 16-square problem, *Math. Ann.* **284** (1989) 437-447.

[6] J. Radon, Lineare Scharen orthogonaler Matrizen, *Abh. Math. Sem. Univ. Hamburg* **1** (1922) 1-14.

[7] E. Stiefel, Über Richtungsfelder in den projektiven Räumen und einen Satz aus der reellen Algebra, *Comment. Math. Helv.* **13** (1940-41) 201-218.

All the topics mentioned in these lectures will appear in greater detail in:

D. Shapiro, *Compositions of Quadratic Forms*, W. deGruyter, Berlin, 2000.

Department of Mathematics
The Ohio State University
Columbus, OH 43210, U.S.A.
email: shapiro@math.ohio-state.edu