Products of Sums of Squares

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

Daniel Shapiro

Lecture 3: Arbitrary Fields

Today we investigate composition formulas where coefficients are allowed to be in some field F (always assuming $2 \neq 0$ in F). In the first lecture we mentioned that the Hurwitz-Radon Theorem remains valid over any field. This settles the question about compositions when $s = n$, that is for sizes $[r, n, n]$.

However the topological methods applied over $\mathbb{R}$ by Hopf, Stiefel, and others do not generalize to other fields in any obvious way. For example, we know from the intercalate matrices that there is a formula of size $[3, 5, 7]$ over $\mathbb{Z}$, and hence over any field. That formula can be described more directly as follows.

$$(x_1^2 + x_3^2 + x_3^2) \cdot (y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2) = \\ (x_1^2 + x_2^2 + x_3^2) \cdot (y_1^2 + y_2^2 + y_3^2 + y_4^2) + (x_1 y_5)^2 + (x_2 y_5)^2 + (x_3 y_5)^2.$$

Since the first term on the right is expressible as a sum of 4 squares (by the 4-square identity), the entire quantity is a sum of 7 squares, as claimed.

Can there exist a composition of size $[3, 5, 6]$ over some field? What about $[3, 6, 7]$ and $[4, 5, 7]$? These are known to be impossible over $\mathbb{R}$ by the Hopf-Stiefel result, because

$$3 \circ 5 = 7, \qquad 3 \circ 6 = 8, \qquad 4 \circ 5 = 8.$$

The Hopf-Stiefel theorem was generalized by Behrend (1939) to allow coefficients in any real closed field. Behrend (a student of Hopf) gave an independent proof of the result using methods of real algebraic geometry. His proof does not extend to any fields where -1 is a sum of squares.

For complex coefficients there is a nice trick due to K.-Y. Lam and T.Y. Lam. They discovered this during an ELAM conference in Mexico in 1982.

The Lam-Lam Lemma 1. *If there exists a composition formula of size* $[r, s, n]$ *over* $\mathbb{C}$*, then there is a nonsingular bilinear* $[r, s, n]$ *over* $\mathbb{R}$*. Consequently,* $r \# s \leq n$*.*

Proof. Suppose there is a formula $(x_1^2 + x_2^2 + \cdots + x_r^2) \cdot (y_1^2 + y_2^2 + \cdots + y_s^2) = z_1^2 + z_2^2 + \cdots + z_n^2$ where each z_k is is a bilinear form in X, Y with coefficients in $\mathbb{C}$. Express $z_k = u_k + iv_k$, where u_k and v_k are bilinear in X, Y with coefficients in $\mathbb{R}$. Compare the real parts in the given formula to find:

$$(x_1^2 + x_2^2 + \cdots + x_r^2) \cdot (y_1^2 + y_2^2 + \cdots + y_s^2) = u_1^2 - v_1^2 + \cdots + u_n^2 - v_n^2.$$

Now consider the map $f : \mathbb{R}^r \times \mathbb{R}^s \to \mathbb{R}^n$ defined by:

$$f(a, b) = (u_1(a, b), \cdots, u_n(a, b)).$$

Then f is bilinear and the formula above shows that it is nonsingular. □

This result immediately extends to any field isomorphic to a subfield of $\mathbb{C}$:

CoroLLary 2. *If there is a composition of size* $[r, s, n]$ *over a field* F *of characteristic zero, then* $r \# s \leq n$*.*

Imitating notations introduced in the first lecture, we define

$$r *_F s = \min\{n : \text{there exists an } [r, s, n] \text{ formula over } F\}.$$

Consequently,

$$r \circ s \;\leq\; r \# s \;\leq\; r *_F s, \quad \text{whenever } F \text{ has characteristic zero.}$$

Here the first inequality is the Hopf-Stiefel Theorem, and the second is the CoroLLary above. In particular we know that formulas of sizes $[3, 5, 6]$, $[3, 6, 7]$, $[4, 5, 7]$ are impossible over any field of characteristic zero.

Those algebraic statements about fields of characteristic zero were proved using topology over $\mathbb{R}$. Are there elementary algebraic methods to attack this problem? One interesting approach arises from the work of Pfister.

Let F be any field (in which $2 \neq 0$) and define

$$D_F(n) = \{a \in F^* : a \text{ is a sum of } n \text{ squares in } F\}.$$

Then $D_F(n)$ is a subset of the multiplicative group F^*, and it is closed under multiplication by squares in F^*.

The standard 2-square identity shows that $D_F(2)$ is closed under multiplication. In fact it is a group. (To get inverses note that $a^{-1} = a \cdot (a^{-1})^2$.) Similarly $D_F(4)$ and $D_F(8)$ are groups. For those people knowing Hurwitz's 1, 2, 4, 8 Theorem, the following result of Pfister (1965) came as a surprise.

Theorem 3. *$D_F(2^m)$ is a group, for every field F and every integer $m \geq 0$.*

An elementary proof is described in Exercise 1. As one example we may apply this when $m = 4$ and $F = \mathbb{R}(X, Y)$ to get a 16-square identity:

$$(x_1^2 + x_2^2 + \cdots + x_{16}^2) \cdot (y_1^2 + y_2^2 + \cdots + y_{16}^2) = z_1^2 + z_2^2 + \cdots + z_{16}^2.$$

Of course these expressions $z_k \in \mathbb{R}(X, Y)$ **must** involve denominators, by the Hurwitz Theorem.

What are the multiplicative properties of the other sets $D_F(n)$? For example, the $[3, 5, 7]$-formula implies that $D_F(3)D_F(5) \subseteq D_F(7)$. Conversely, suppose $c \in D_F(7)$. Then $c = c_1^2 + \cdots + c_7^2 = (c_1^2 + c_2^2 + c_3^2)\left(1 + \frac{c_4^2+c_5^2+c_6^2+c_7^2}{c_1^2+c_2^2+c_3^2}\right)$. That last fraction is a sum of 4 squares (since $D_F(4)$ is a group), and hence $c \in D_F(3)D_F(5)$. This proves that $D_F(3)D_F(5) = D_F(7)$ for every F.

Pfister analyzed the product $D_F(r) \cdot D_F(s)$ and discovered that it always equals another one of the sets $D_F(n)$. The value n there that works for every field F is exactly the Hopf-Stiefel function!

Proposition 4. *$D_F(r) \cdot D_F(s) = D_F(r \circ s)$, for every field F and every r, s.*

Pfister proved this result in 1965 without knowing the connection with the Hopf-Stiefel result. That relationship was pointed out by Köhnen, one of Pfister's students, in 1978. As one consequence of this calculation, we see that the binary operation $r \circ s$ is associative, a property not obvious from the original definition.

Let's use Pfister's calculation to prove that compositions of size $[3, 5, 6]$ are impossible over some fields. Suppose a $[3, 5, 6]$ exists over F. By substituting field elements for the indeterminates in the formula we obtain:

$D_F(3)D_F(5) \subseteq D_F(6)$. Then, by Pfister, $D_F(7) = D_F(6)$, and the same argument shows that $D_K(7) = D_K(6)$, for every field K containing F. We want to derive a contradiction here, but over some fields (like $\mathbb{C}$) those sets *are* equal. However, Cassels (1963) settled the question for rational functions over "formally real" fields. (F is formally real if -1 is not expressible as a sum of squares in F.)

Lemma 5. *Suppose $K = F(x_1, \ldots, x_n)$ where F is a formally real field. Then $1 + x_1^2 + \cdots + x_n^2$ cannot be expressed as a sum of n squares in K. Consequently, $D_K(n) \neq D_K(n+1)$.*

Therefore a formula of size $[3, 5, 6]$ is impossible over any formally real field. Generalizing this argument we get a parallel of the Hopf-Stiefel theorem:

Corollary 6. *Suppose F is formally real. If there is a composition of size $[r, s, n]$ over F then $r \circ s \leq n$.*

The proof sketched above is fairly easy (compared to the topology used earlier). But unfortunately every formally real field has characteristic zero, so this corollary is weaker than the earlier result, which was proved using the topology.None of the methods above is of any use when the field F has positive characteristic. Could there be a $[3, 5, 6]$-formula over some finite field? The first analysis of this problem valid over arbitrary fields was done by Adem (a topologist), who examined rectangular matrices directly. During the original Hurwitz proof we saw that an $[r, s, n]$-formula over F is equivalent to an $n \times s$ matrix A, whose entries are linear forms in X, and which satisfies

$$A^{\mathsf{T}} A = \alpha I_s,$$

where $\alpha = x_1^2 + \cdots + x_r^2 \in F(X)$. Using the standard dot product for column vectors in the vector space $F(X)^n$, this equation says that the s columns of A are orthogonal vectors all of "length" α. Hurwitz and Radon dealt with the case of square matrices: when $s = n$. Adem moved one step away, considering the case $s = n - 1$.

For concreteness let's work again with $[3, 5, 6]$. Starting from the given 6×5 matrix A we want to find one new column vector v so that the 6×6 matrix $\hat{A} = (A\ v)$ will have linear form entries and satisfy $\hat{A}^{\mathsf{T}} \hat{A} = \alpha I_6$. If we could do that, then the Hurwitz proof would provide a $[3, 6, 6]$-formula over F,

and that is impossible by Hurwitz-Radon: $\rho(6) = 2$. That vector v needs to have linear form entries, to be orthogonal to the 5 columns of A, and to have "length" α. The 5 given columns in 6-dimensional space have a line as their orthogonal complement, so the vector v is determined up to a scalar (if it exists). A formula for v can be found using determinants, generalizing the "cross product" of two vectors in 3-space. Further analysis of that formula shows that we *can* remove common factors from the entries of v to obtain linear forms and get the correct length. Done.

It turns out that this method works whenever n is even. When n is odd that complementary vector v can be taken with constant entries, and the whole problem restricts to the $(n-1)$-dimensional space $(v)^{\perp}$. Summarizing all this, we obtain:

Adem's Theorem 7. *Suppose F is any field with characteristic $\neq 2$. If there is a composition of size $[r, n-1, n]$ over F, then:*

If n is even, there is an $[r, n, n]$ over F. Then $r \leq \rho(n)$.

If n is odd, there is an $[r, n-1, n-1]$ over F. Then $r \leq \rho(n-1)$.

This result eliminates the small cases $[3, 5, 6]$ and $[3, 6, 7]$ over *any* field. The size $[4, 5, 7]$ has codimension 2 and is not covered by Adem's Theorem above. However Adem did eliminate that case, and his result was generalized by Yuzvinsky, Adem, Gauchman and Toth, and Shapiro. The arguments involve more complicated linear algebra, but are still fairly elementary.

Theorem 8. *Suppose F is a field with characteristic not 2, and there is a composition of size $[r, n-2, n]$ over F.*

If n is odd then: either $r \leq \rho(n-1)$, or $r = 3$ and $n \equiv 3 \pmod 4$.

If n is even then: either $r \leq \rho(n)$ or $r \leq \rho(n-2)$.

It seems difficult analyze the question in codimension 3 using similar ideas from linear algebra. Further information can be obtained if we are willing to apply more powerful machinery. The original proof of the Hopf theorem can be modified, with the cohomology groups of projective spaces replaced by certain Chow groups, and these groups can be calculated using some sophisticated algebraic K-theory. This leads to the following result proved with Szyjewski.

Theorem 9. *Suppose F is a field with characteristic $\neq 2$ and there is a composition of size $[r, s, n]$ over F.*

If n is even then $r \circ s \leq n$.

If n is odd then $r \circ s \leq n + 1$.

For example $5 \circ 9 = 13$ and $5 \circ 10 = 14$. The theorem above implies that $[5, 9, 12]$-formulas cannot exist over any field, but no information is known about compositions of size $[5, 10, 13]$ over fields ofpositive characteristic. They are impossible in characteristic zero by CoroLLary 2.

It is tempting to conjecture that the composition formula sizes are independent of the ring of coefficients. That is: If there is a composition of size $[r, s, n]$ over some field F then there must exist a composition of that size over $\mathbb{Z}$. There are no known counterexamples to this conjecture, but on the other hand, very few composition formulas are known so far.

Exercises.

EXERCISE 1. $D_F(2^m)$ **is a group.**

(a) Let $n = 2^m$. Prove the following lemma.

Lemma. *Suppose $c = c_1^2 + c_2^2 + \cdots + c_n^2$. Then there exists an $n \times n$ matrix C having first row $(c_1, c_2, \ldots, c_n)$ and satisfying $C \cdot C^{\mathsf{T}} = C^{\mathsf{T}} \cdot C = cI_n$.*

(b) Prove $D_F(n)$ is a group, as follows: If $c, d \in D_F(n)$ have corresponding matrices C, D as in the lemma, then $A = CD^{\mathsf{T}}$ satisfies $AA^{\mathsf{T}} = cdI_n$.

(c) Prove that there is an n-square identity $(x_1^2 + \cdots + x_n^2) \cdot (y_1^2 + \cdots + y_n^2) = z_1^2 + z_2^2 + \cdots + z_n^2$, where each z_k is linear in Y with coefficients in the field $F(X)$. Moreover, we can arrange $z_1 = x_1y_1 + \cdots + x_ny_n$.

(Hints. (a) Express $c = a + b$ where a, b are sums of 2^{m-1} of the c_i^2. By induction, let A, B be the matrices corresponding to a, b. If $a \neq 0$, define $C = \begin{bmatrix} A & B \\ \diamondsuit & A^{\mathsf{T}} \end{bmatrix}$, where the entry $\diamondsuit$ is to be filled in. What if $a = 0$?

Parts (b) and (c) follow quickly from the Lemma.)

EXERCISE 2. **Pfister's products.**

Let F be a field and write $D(n)$ for $D_F(n)$.

(a) Prove $D(r)D(s) \subseteq D(r \circ s)$ by induction on $r+s$ using the following steps. Suppose $r \leq s$ and $2^m < s \leq 2^{m+1}$.

If $r \geq 2^m$ then $r \circ s = 2^{m+1}$.

If $r < 2^m < s$ express $s = s' + 2^m$ so that $r \circ s = (r \circ s') + 2^m$. Apply the hypothesis on $D(r)D(s')$.

(b) For the converse, first prove $D(2^m)D(2^m+1) = D(2^{m+1})$. Complete the proof of Proposition 4.

EXERCISE 3. **Full maps.**

A bilinear map $f : F^r \times F^s \to F^n$ is ***full*** if image(f) spans F^n.

(a) There is a full bilinear $[r,s,n]$ if and only if $n \leq rs$.

(b) Suppose f is a direct sum of bilinear maps g_1 and g_2. If g_1 and g_2 are full prove that f must be full. Does the converse hold?

(c) Prove that every composition of size $[r, s, r *_F s]$ over F must be full.

(d) If $r > 2$ and there is a full composition of size $[r, n-1, n]$, explain why n must be even.

(Hints. (a) Begin with the tensor product. (d) Check Adem's Theorem.)

EXERCISE 4. **Nonsingular maps.**

Define $r \#_F s = \min\{n : \text{there is a nonsingular bilinear } [r,s,n] \text{ over } F\}$. Note: the inequality $r \# s \leq r * s$ over $\mathbb{R}$ does not generalize to fields where a sum of nonzero squares can equal zero.

(a) Show that

$$\max\{r,s\} \leq r \#_F s \leq r + s - 1.$$

(b) If F admits field extensions of every degree $\leq n$, prove that $r \#_F s = \max\{r,s\}$ whenever $r, s \leq n$.

(c) Prove: $r \#_F s = r + s - 1$ for every r, s if and only if F is algebraically closed.

(d) Define $Deg(F)$ to be the set of degrees of finite field extensions of F. For example $Deg(\mathbb{C}) = \{1\}$ and $Deg(\mathbb{R}) = \{1, 2\}$. Suppose $Deg(F) = \{1, 2, 4, \ldots, 2^m, \ldots\}$. (Such fields can be constructed using Galois theory.) Prove: $r \circ s \leq r \#_F s$.

(Hints. (a) Recall the "Cauchy product" map $c_{r,s}$ from Lecture 1, Exer. 5.

(c) Suppose $f : U \times V \to W$ is a nonsingular bilinear $[r, s, n]$, where n is minimal. Then $f^{\otimes} : U \otimes V \to W$ is linear and surjective. Let $\mathcal{K} = \ker(f^{\otimes})$ and $\mathcal{D} = \{u \otimes v : u \in U, v \in V\}$. Explain why $\mathcal{D}$ is an algebraic set of dimension $r + s - 1$ and $\mathcal{K} \cap \mathcal{D} = \{0\}$. If F is algebraically closed, count dimensions to prove $r + s - 1 \leq n$.

Conversely, $2 \#_F s = s + 1$ if and only if every polynomial of degree s in $F[x]$ has a root in F.)

(d) Given nonsingular $[2^m, 2^m, 2^m]$ for every m, use direct sums to construct a nonsingular $[r, s, r \circ s]$.)

References

[1] J. Adem, On the Hurwitz problem over an arbitrary field, I, II, *Bol. Soc. Mat. Mexicana* **25** (1980) 29-51, **26** (1981) 29-41.

[2] A. Pfister, Darstellung von -1 als Summe von Quadraten in einem Körper, *J. London Math. Soc.* **40** (1965) 159-165.

[3] D. B. Shapiro, On the Hurwitz problem over an arbitrary field, *Bol. Soc. Mat. Mexicana* **29** (1984) 1-4.

[4] D. B. Shapiro and M. Szyjewski, Product formulas for quadratic forms, *Bol. Soc. Mat. Mexicana* **37** (1992) 463-474.

[5] S. Yuzvinsky, On the Hopf condition over an arbitrary field, *Bol. Soc. Mat. Mexicana* **28** (1983) 1-7.

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

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

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