Pfister's Work on Sums of Squares

A.R. Rajwade

Historically the theory of quadratic forms was regarded as a topic in number theory. However, Witt's paper "Theorie der quadratischen Formen in beliebigen Körpern" of 1937[15] opened up a new chapter in the theory: that of combining the number theoretic aspect with the algebraic development, by the creation of the famous Witt ring.

Then triggered off by Cassel's paper "On the representation of rational functions as sums of squares" of 1964[4], Albrecht Pfister, about 1966, come up with his celebrated structure theorems, giving birth to a purely algebraic theory of quadratic forms. Special cases of arithmetical aspect of Pfister's theory are his beautiful results about sums of squares and Pfister forms.

Our object here is to give a brief exposition of Pfister's work on sums of squares and related topics, one of the most beautiful and self contained set of results in any field (pun intended).

So let K be a field. We make the following

Definition 1. The smallest integer $s = s(K)$ for which the equation $-1 = a_1^2 + a_2^2 + \cdots + a_s^2$ $(a_j \in K)$ is solvable, is called the *Stufe* (often referred to as *level*) of K. If the equation has no solution, we put $s = \infty$ and call K *formally real*.

In 1932, Van der Waerden had posed the problem of enquiring which numbers can occur as Stufe. For example 3 can never occur as Stufe. Indeed if $-1 = x^2+y^2+z^2 (x, y, z \in K)$, then $0 = 1 + x^2 + y^2 + z^2$. Multiplying by $1 + x^2$ gives

$$\begin{aligned} 0 &= (1+x^2)^2 + (1+x^2)(y^2+z^2) && \text{(a)} \\ &= (1+x^2)^2 + (y+xz)^2 + (z-xy)^2 && \text{(b)} \end{aligned}$$

Now $1 + x^2 \neq 0$, otherwise Stufe $K = 1$. Hence (b) gives

$$-1 = \left(\frac{y+xz}{1+x^2}\right)^2 + \left(\frac{z-xy}{1+x^2}\right)^2$$

showing $s(K) \neq 3$.

It may similarly be shown that no odd number can be the Stufe of any field, but can 6 or 10 or 12 be the Stufe of a suitable field? We shall have to experiment with various fields and then make a conjecture. The rationals Q and the reals R, being formally real, are of no use, the complexes C provide a trivial example: $-1 = \iota^2$ giving $s(C) = 1$. So let us look at the imaginary quadratic fields. Indeed we have the following

Theorem 1. *Let $D > 0$ be a square free integer; then the Stufe $s(K)$ of $K = Q\left(\sqrt{-D}\right)$ is*

$$\begin{cases} 1 & \text{if } D = 1, \\ 2 & \text{if } D \neq 8b+7, \\ 4 & \text{if } D = 8b+7. \end{cases}$$

If $D < 0$, then K is of course formally real.

Proof: [12]. Writing $D = a^2+b^2+c^2+d^2,\ a, b, c, d, \in Z$, we see that $0 = (\sqrt{-D})^2 + a^2 + b^2 + c^2 + d^2$, giving $s(K) \leq 4$. Now $s(K) = 1$ if and only if $\sqrt{-1} \in K$ and this happens only in the case $D = 1$. If $D \not\equiv 7 \pmod 8$, then D is a sum of three squares and so $0 = (\sqrt{-D})^2 + a^2 + b^2 + c^2$, giving $s(K) \leq 3$. But we have already seen that s cannot be 3, hence $s(K) = 2$, the case $s = 1$ being fully cleared.

Finally let $D \equiv 7 \pmod 8$. If $s(K)$ were < 4, then it would be equal to 2, i.e.

$$-1 = (a_1 + b_1\sqrt{-D})^2 + (a_2 + b_2\sqrt{-D})^2, a_1, b_1, a_2, b_2, \in Q.$$

Here, without loss of generality, we may suppose that $b_1 \neq 0$. Equating reals and imaginaries, we get the following two equations:

$$a_1^2 + a_2^2 - D(b_1^2 + b_2^2) = -1$$
$$a_1 b_1 + a_2 b_2 = 0$$

These imply $D = (\frac{a_2}{b_1})^2 + (\frac{b_1}{b_1^2+b_2^2})^2 + b(\frac{b_2}{b_1^2+b_2^2})^2$. Thus D is a sum of 3 rational squares which is a contradiction since $D \equiv 7 \pmod 8$. Thus $s(K)$ not be less than 4 as required. □

Let us next look at all the finite fields. We have the following easy

Theorem 2. *Let F_q be the finite field of $q = p^\alpha$ elements; then*

$$s(F_q) = \left\{ \begin{array}{l} 1 \ \textit{if either } p = 2 \textit{ or } p \equiv 1(4), \text{ or } p \equiv 3(4), 2|\alpha, \\ 2 \ \textit{otherwise i.e. if } p \equiv 3(4), 2 \not| \alpha. \end{array} \right\}$$

Proof: First let $p = 2$. Then $-1 = 1 = 1^2$ giving $s(F_2^\alpha) = 1$

Next if $p \equiv 1(4)$, then $(\frac{-1}{p}) = 1$ i.e. $-1 = x^2$ is solvable in $F_p \subset F_{p^\alpha}$ (for all α). So $s(F_{p^\alpha}) = 1$.

Let now $p \equiv 3(4)$. Let $A = \left\{ -1 - X^2 \middle| X = 1, 2, \ldots, \frac{p-1}{2}, 0 \right\}$ and B $= \left\{ Y^2 \middle| Y = 1, 2, \ldots, \frac{p-1}{2}, 0 \right\}$, A, B, $\subset F_p$. Then $|A| = |B| = \frac{p+1}{2}$. Hence by the pigeon hole principle, there exist $X_0, Y_0 \in F_p$ such that $Y_0^2 = -1 - X_0^2$, i.e. $-1 = X_0^2 + Y_0^2$ in F_p. But -1 is not a square in F_p since $p \equiv 3(4)$. It follows that $s(F_p) = 2$ if $p \equiv 3(4)$.

Now $F_p^2 = F_p(\sqrt{-1})$ and here $-1 = (\sqrt{-1})^2$; so $s(F_{p^2}) = 1$. But $F_{p^\alpha} \supset F_{p^2}$ if $2|\alpha$, so F_{p^α} has Stufe 1 if $2|\alpha$.

If $s(F_{p^\alpha}) = 1$ even for $2 \not| \alpha$, then $-1 = X^2$ is solvable in $F_{p^\alpha}(2 \not| \alpha)$, so $F_{p^\alpha} \supset F_p(X) = F_p(\sqrt{-1}) \cong F_{p^2}$ which is false since $2 \not| \alpha$. Hence $s(F_{p^\alpha}) = 2$ if $2 \not| \alpha$. □

That more or less exhausts all the easy fields using elementary methods. Even allowing the Hasse-Minkowski theorem all those algebraic number fields K for which $s(K)$ exists finite can be dealt with in a single go. The exact result is the following:

Theorem A. *Let $K = Q(\alpha)$ be an algebraic number field with $[K:Q] = n$ finite. Then $s(K)$ exists finite iff K is totally complex (i.e. all the zeros of irr (α, Q) are non-real) and then $s(K) \leq 4$ i.e. equals 1, 2 or 4, because we have seen that Stufe can not be 3 and in fact*

$$s(K) = \begin{cases} 1 & \text{iff } i \in K (i = \sqrt{-1}) \\ 4 & \text{iff } i \notin K \text{ and for all primes } y|2, \text{ the local degree } [K_y : Q_2] \text{ is odd.} \end{cases}$$

For a proof see [13], p 261.

We see that experimentation is not easy and so it was all the more surprising, when Pfister proved the following beautiful

Theorem 3. *For any field K, the Stufe $s(K)$, if finite, is always a power of 2. Conversely, every power of 2 is the Stufe of some field K.*

We shall give a proof of this result in the sequel. In showing that 3 cannot occur as Stufe, the transition from equation (a) to (b) (see before theorem 1) is the crucial step in the process. We have, more generally, the curious looking identity.

$$(X_1^2 + X_2^2)(Y_1^2 + Y_2^2) = (X_1Y_1 - X_2Y_2)^2 + (X_1Y_2 + X_2Y_1)^2 \tag{1}$$

which tells its that *a product of two sums of two squares is itself a sum of two squares.* Known to the Greeks, (1) is equivalent to the statement, The norm of the product of two complex numbers Z_1, Z_2, is the product of their norms:

$$|Z_1 Z_2|^2 = |Z_1|^2 |Z_2|^2 \tag{1'}$$

for writing $Z_1 = X_1 + iX_2$, $Z_2 = Y_1 + iY_2$, we see that $Z_1.Z_2 = (X_1Y_1 - X_2Y_2) + i(X_1Y_2 + Y_2X_2)$ and so (1) and (1′) are the same.

This identity (1) enables us to prove another curious result:

For any field K, the set $G_2(K) = \{a \in K^* | a = x^2 + y^2, x, y, \in K\}$ is a multiplicative group.

For, the closure property is the identity (1) while if

$$a = x^2 + y^2 \in G_2(K), \text{ then } \frac{1}{a} = \frac{a}{a^2} = \frac{x^2 + y^2}{a^2} = \left(\frac{x}{a}\right)^2 + \left(\frac{y}{a}\right)^2 \in G_2(K).$$

The following striking identity was already known to Euler in 1770 and he used it to prove Lagrange's theorem that every positive integer is a sum of four squares.

$$(X_1^2 + X_2^2 + X_3^2 + X_4^2)(Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2) = (Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2) \tag{2}$$

where

$$\begin{aligned} Z_1 &= X_1Y_1 - X_2Y_2 - X_3Y_3 - X_4Y_4 \\ Z_2 &= X_1Y_2 + X_2Y_1 + X_3Y_4 - X_4Y_3 \\ Z_3 &= X_1Y_3 + X_3Y_1 - X_2Y_4 + X_4Y_2 \\ Z_4 &= X_1Y_4 + X_4Y_1 + X_2Y_3 - X_3Y_2 \end{aligned}$$

The discovery of quaternions by William Hamilton, in 1843, brought out the real significance of the identity (2) in as much as (2) is simply the fact that the norm of a product of two quaternions is equal to the product of their norms.

Almost immediately after Hamilton's discovery of the quaternions, Arthur Cayley, in 1845 discovered the octonions (the Cayley numbers) which give rise to the incredible looking identity

$$(X_1^2 + \cdots + X_8^2)(Y_1^2 + \cdots + Y_8^2) = Z_1^2 + \cdots + Z_8^2 \tag{3}$$

where

$$\begin{aligned}
Z_1 &= X_1Y_1 - X_2Y_2 - X_3Y_3 - X_4Y_4 - X_5Y_5 - X_6Y_6 - X_7Y_7 - X_8Y_8,\\
Z_2 &= X_1Y_2 + X_2Y_1 + X_3Y_4 - X_4Y_3 + X_5Y_6 - X_6Y_5 - X_7Y_8 + X_8Y_7,\\
Z_3 &= X_1Y_3 + X_3Y_1 - X_2Y_4 + X_4Y_2 + X_5Y_7 - X_7Y_5 + X_6Y_8 - X_8Y_6,\\
Z_4 &= X_1Y_4 + X_4Y_1 + X_2Y_3 - X_3Y_2 + X_5Y_8 - X_8Y_5 - X_6Y_7 + X_7Y_6,\\
Z_5 &= X_1Y_5 + X_5Y_1 - X_2Y_6 + X_6Y_2 - X_3Y_7 + X_7Y_3 - X_4Y_8 + X_8Y_4,\\
Z_6 &= X_1Y_6 + X_6Y_1 + X_2Y_5 - X_5Y_2 - X_3Y_8 + X_8Y_3 + X_4Y_7 - X_7Y_4,\\
Z_7 &= X_1Y_7 + X_7Y_1 + X_2Y_8 - X_8Y_2 + X_3Y_5 - X_5Y_3 - X_4Y_6 + X_6Y_4,\\
Z_8 &= X_1Y_8 + X_8Y_1 - X_2Y_7 + X_7Y_2 + X_3Y_6 - X_6Y_3 + X_4Y_5 - X_5Y_4.
\end{aligned}$$

Although the identity emerges most naturally from Cayley numbers, it was discovered nearly a quarter of a century earlier by C.F. Degan (1822) with minor sign differences.

Degan stated (erroneously of course) that there is a like formula for 2^n squares. For the case of 16 squares, he gave the literal parts of the 16 bilinear functions $Z_1, Z_2, \ldots, Z_{16}$ but left most of the signs undetermined, saying that the only difficulty is the prolixity of the ambiguities of signs.

Degan was also aware of the 2– and 4– variable Pfister forms $X_1^2 - aX_2^2$ and $X_1^2 + aX_2^2 + b(X_3^2 + aX_4^2)$ both of which satisfy identities similar to (1) and (2).

As before, if we define

$$G_4 = \{a \in K^* \mid a = x_1^2 + \cdots + x_4^2, x_i \in K\}$$

and G_8 similarly, then it follows from (2) and (3) respectively that G_4 and G_8 are groups under multiplication, so that we have the chain of inclusions

$$K^{*^2} \subseteq G_2 \subseteq G_4 \subseteq G_8 \subseteq K^*.$$

A great many unsuccessful attempts followed Degan's discovery of (3), to extend formulae (1), (2) and (3) to a similar 16 term identity, and many workers, realizing the impossibility of such an extension, tried giving convincing arguments to prove the impossibility. Hamilton's and Cayley's discoveries had reduced the problem to the determination of the so-called normed algebras over the real numbers **R**; the four known ones being **R** (of dimension 1),

the complex numbers **C** (of dimension 2), the quaternions **H** (of dimension 4) and the octonions **O** (of dimension 8). It is an astonishing observation how the axioms of the ordered field **R** gradually drop off as we move up these higher dimensional hypercomplex systems: **C** is, no doubt a field, commutative and associative (under multiplication) and a division ring, but the order property is lost. **H** is only an associative division ring; thus commutativity, and order are both lost. Finally **O** is not even associative-it is merely a division ring; thus commutativity, associativity and order are all lost.

The half century following the discovery of these quaternions and octonions then saw many attempts to find a 16-dimensional hypercomplex system over the reals and several erroneous affirmations were given. Finally in 1898, Hurwitz [6] gave a decisive solution to the problem about the dimensionality of all possible normed algebras over **R** and so also about the possible values of n for which there is an identity of the type (3) with n terms. More precisely we have the following.

Theorem 4 (Hurwitz-1898). *Let K be a field with char $K \neq 2$. The only values of n for which there is an identity of the type*

$$(X_1^2 + \cdots + X_n^2)(Y_1^2 + \cdots + Y_n^2) = Z_1^2 + \cdots + Z_n^2 \tag{4}$$

where the Z_k are bilinear functions of the X_i and the Y_j, coefficients in K are $n = 1, 2, 4, 8$.

Actually Hurwitz proved this only over **C** but his proof generalizes to any field K with char $K \neq 2$. We give here a proof given by Dickson in his beautiful expository paper [5] of 1919. A proof using normed algebras can be found in A.A. Albert's *Studies in Modern Algebra* [2].

Proof: (Dickson). The idea is to convert (4) into a system of matrix equations. The bilinearity condition on the Z_k can be written as

$$\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = AY,$$

where the a_{ij} are linear functions of $X_1, X_2, \ldots, X_n$. Then (4) becomes

$$(X_1^2 + \cdots + X_n^2)(Y_1, \ldots, Y_n) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} I_n = (Z_1, \ldots, Z_n) \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = Y'A'AY,$$

i.e.

$$(Y_1, Y_2, \ldots, Y_n)[(X_1^2 + X_2^2 + \cdots + X_n^2)I_n - A'A] \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = 0,$$

and since this is true for all $Y_1, Y_2, \ldots, Y_n$, it follows that

$$A'A = (X_1^2 + \cdots + X_n^2)I_n \tag{a}$$

Now

$$\begin{aligned} A &= \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} \\ &= \begin{pmatrix} b_{11}^{(11)}X_1 + b_{11}^{(12)}X_2 + \cdots, b_{12}^{(11)}X_1 + b_{12}^{(12)}X_2 + \cdots \\ b_{21}^{(11)}X_1 + b_{21}^{(12)}X_2 + \cdots, b_{22}^{(11)}X_1 + b_{22}^{(12)}X_2 + \cdots \\ \cdots \end{pmatrix} \\ &= A_1X_1 + A_2X_2 + \cdots + A_nX_n \quad \text{say} \end{aligned}$$

By (a),

$$\begin{aligned} &(A_1'X_1 + A_1'X_2 + \cdots + A_n'X_n)(A_1X_1 + A_2X_2 + \cdots + A_nX_n) \\ &\quad = (X_1^2 + X_2^2 + \cdots + X_n^2)I_n. \end{aligned}$$

Since this is true for all X_j, we have

(1) $A_j'A_j = I_n (j = 1, 2, \ldots, n)$, hence also $A_jA_j' = I_n$.
(2) $A_j'A_k + A_k'A_j = 0, 1 \leq j, k \leq n, j \neq k$.

Conversely, the existence of such a system implies that (4) holds with Z_k bilinear in the X_i and the Y_j. Note also that if $n = 1$, (2) is vacuous, and (1) can be trivially satisfied so we may suppose $n > 1$.

Now let $B_i = A_n'A_i (i = 1, 2, \ldots, n-1)$. The B's are easily seen to satisfy

(1) $B_i'B_i = I_n$
(2) $B_i' + B_i = 0 \quad (i, j = 1, 2, \ldots, n-1)$
(3) $B_i'B_j + B_j'B_i = 0 \quad (i \neq j)$

Hence we have

$$\left.\begin{array}{ll} \text{(i)} & B_i' = -B_i (i = 1, 2, \ldots, n-1) \text{ i.e. the } B_i \\ & \quad \text{are skew} - \text{symmetric matrices} \\ \text{(ii)} & B_i^2 = -I (i = 1, 2, \ldots, n-1) \\ \text{(iii)} & B_iB_j = -B_jB_i, \ i, j = 1, 2, \ldots, n-1 \quad i \neq j. \end{array}\right\} \tag{b}$$

It follows that $|B_i| = |B_i'| = |-B_i| = (-1)^n|B_i|$, and since $|B_i| \neq 0$ we must have n even. Hence

Proposition 1. *There is no identity of the type* (4) *if* $n(>1)$ *is odd.*

In future, therefore, we suppose n to be even. Now consider the following set $\mathcal{G}$ of $n \times n$ matrices:

$$\{I, B_{i_1}, B_{i_1}B_{i_2}, B_{i_1}B_{i_2}B_{i_3}, \ldots, B_{i_1}B_{i_2}, \ldots, B_{i_{n-2}} \text{ and } B_1B_2, \ldots, B_{n-1} (i_1 < n, i_1 < i_2 < n, \ldots)\}.$$

Here B_{i_1} takes $n-1$ values viz. $B_1, B_2, \ldots, B_{n-1}$, while $B_{i_1}B_{i_2}$ takes $\binom{n-1}{2}$ values viz. $B_1B_2, B_1B_3, \ldots$ etc. So altogether there are $1 + \binom{n-1}{1} + \cdots + \binom{n-1}{n-1} = 2^{n-1}$ elements in the set $\mathcal{G}$. Let $G = B_{i_1}B_{i_2}, \ldots, B_{i_r} \in \mathcal{G}$. Then we have

Lemma 1. *G is symmetric if $r \equiv 0$ or 3 modulo 4, and skew-symmetric if $r \equiv 1$ or 2 modulo 4.*

Proof:

$$\begin{aligned} G' &= B'_{i_r}, \ldots, B'_{i_1} = (-1)^r B_{i_r}, \ldots, B_{i_1} \\ &= (-1)^r(-1)^{r-1} B_{i_1}(B_{i_r}, \ldots, B_{i_2}) \end{aligned}$$

by (iii) of (b) to commute B_{i_1} successively with $B_{i_2}, \ldots, B_{i_r}$,

$$\begin{aligned} &= (-1)^r(-1)^{r-1}(-1)^{r-2} B_{i_1}B_{i_2}(B_{i_r}, \ldots, B_{i_3}) \\ &\qquad \text{and so on} \\ &= (-1)^r(-1)^{r-1}\ldots(-1)^2(-1)B_{i_1}B_{i_2}, \ldots, B_{i_r} \\ &= (-1)^{1+2+\cdots+r}G \\ &= (-1)^{r(r+1)/2}G \\ &= \begin{cases} G & \text{if } r \equiv 0, 3, (4) \\ -G & \text{if } r \equiv 1, 2, (4) \end{cases} \end{aligned}$$

□

Lemma 2. *Let $M \in \mathcal{G}$. Then the set $M\mathcal{G} = \{MG | G \in \mathcal{G}\}$ is simply a permutation of $\mathcal{G}$ with each term prefixed with either $+1$ or -1.*

Proof: The result is clear if the multiplier M is B_1, since then the product will contain or lack B_1 according as the multiplicand of $\mathcal{G}$ lacks or contains B_1 (use again (b)).

If the multiplier is B_2, we first replace $B_1B_2, \ldots$, wherever it appears, by $B_2B_1, \ldots$, and see that the former argument applies.

After thus proving our statement when the multiplier is any B_i, we see that it holds when the multiplier is any product of the B's. □

An Example: $n = 4$.

$$\mathcal{G} = \{I, B_1, B_2, B_3, B_1B_2, B_2B_3, B_1B_3, B_1B_2B_3\}.$$

Then

$$\begin{aligned} B_3\mathcal{G} &= \{B_3, B_3B_1, B_3B_2, B_3^2, B_3B_1B_2, B_3B_2B_3, B_3B_1B_3, B_3B_1B_2B_3\}. \\ &= \{B_3, -B_1B_3, -B_2B_3, -I, B_1B_2B_3, B_2, B_1, -B_1B_2\} \\ &= \{-I, B_1, B_2, B_3, -B_1B_2, -B_2B_3, -B_1B_3, B_1B_2B_3\}. \end{aligned}$$

Our aim is now the following.

Proposition 2. *At least half of the elements of $\mathcal{G}$ are linearly independent.*

With this in view, we look for any linear relations that can exist amongst the elements of $\mathcal{G}$.

Definition: A relation $\lambda_1 G_1 + \lambda_2 G_2 + \cdots + \lambda_s G_s = 0, G_j \in \mathcal{G}, \lambda_j \in \mathbf{R}$, or $R = 0$ for short, is called *irreducible* if it is not possible to express R as $R_1 + R_2$, where $R_1 = 0$, $R_2 = 0$ represent two linear relations that hold between the subsets R_1 and R_2 of R with $R_1 \cap R_2 = \emptyset$, i.e. there are no matrices common to R_1 and R_2.

We have the following.

Lemma 3. *An irreducible relation $R = 0$ cannot involve both symmetric and skew-symmetric matrices.*

Proof: Let M_1 be the subset of all symmetric matrices in R and M_2 the set of all skew-symmetric matrices in R. Then $M_1 + M_2 = 0$, i.e. $M_1 = -M_2$. Hence $M_1 = M_1' = -M_2' = M_2$, i.e. $M_1 = M_2$. It follows that $M_1 = 0$, $M_2 = 0$ which contradicts the irreducibility of $R = 0$. □

Now let $R = 0$ be any irreducible relation between the matrices of $\mathcal{G}$. By multiplying R by a suitable λG ($\lambda \in \mathbf{R}, G \in \mathcal{G}$) we get a new relation $T = 0$, one term of which is I and all the remaining terms are products of matrices of $\mathcal{G}$ by real constants. For suppose μG ($\mu \in \mathbf{R}, G \in \mathcal{G}$) is a term in R which we wish should become I in the relation $T = 0$. One just multiplies $R = 0$ by $\pm \mu^{-1} G^{-1}$ and notes that one of $\pm G^{-1} \in \mathcal{G}$.

For example if $4B_2B_3$ is one term of R, then on multiplying $R = 0$ by

$$-\frac{1}{4}(B_2B_3)^{-1} = -\frac{1}{4}B_3^{-1}B_2^{-1} = -\frac{1}{4}(-B_3)(-B_2) = -\frac{1}{4}B_3B_2 = \frac{1}{4}B_2B_3,$$

we get what is required.

This new relation $T = 0$ is also irreducible, for if $T = 0$ were to split as $T_1 = 0$, $T_2 = 0$, then since $T = \lambda G R$ we have $\lambda^{-1}G^{-1}T = R$ and so $R = 0$ splits as $\lambda^{-1}G^{-1}T_1 = 0$, $\lambda^{-1}G^{-1}T_2 = 0$, which gives a contradiction.

Hence we may suppose that $T = 0$ looks like

$$I = \sum c_{i_1 i_2 i_3} B_{i_1} B_{i_2} B_{i_3} + \sum d_{i_1 i_2 i_3 i_4} B_{i_1} B_{i_2} B_{i_3} B_{i_4} + \cdots \qquad (*)$$

where by Lemma 3, each of the matrices $B_{i_1}B_{i_2}B_{i_3}$, $B_{i_1}B_{i_2}B_{i_3}B_{i_4}$, etc. is symmetric since I is symmetric. That is why no singleton B_i nor any of the products B_iB_j of two B's can be involved in $(*)$ since B_i and B_iB_j are skew-symmetric by Lemma 1.

Now multiply $(*)$ throughout on the right by B_i to obtain an irreducible relation which then involves only skew-symmetric matrices since one term (on the left side) is the skew-symmetric matrix B_i. But by Lemma 1, $B_{i_1}B_{i_2}B_{i_3}B_{i_4}$ is symmetric. So all the c_j are 0 if only i is distinct from $i_1 i_2 i_3$. Since i may have any value $\leq n - 1$, we see that each c is 0 unless $n - 1 = 3$ for then i cannot be chosen different from i_1, i_2, i_3.

Next we show that all the d's are 0 too; for multiply $(*)$ by B_{i_4} and it becomes

$$\begin{aligned} B_{i_4} &= \sum d_{i_1 i_2 i_3 i_4} B_{i4} B_{i_1} B_{i_2} B_{i_3} B_{i_4} + \cdots \\ \text{But } B_{i_4} B_{i_1} B_{i_2} B_{i_3} B_{i_4} &= (-1)^3 B_{i_1} B_{i_2} B_{i_3} B_{i_4}^2 = B_{i_1} B_{i_2} B_{i_3}. \text{ So } (*) \text{ becomes} \\ B_{i_4} &= \sum d_{i_1 i_2 i_3 i_4} B_{i_1} B_{i_2} B_{i_3} + \cdots. \end{aligned}$$

Here $B_{i_1}B_{i_2}B_{i_3}$ is symmetric, while B_{i_4} is skew-symmetric (by Lemma 1). It follows that all the d's are 0 too.

The method used in proving $c = 0$ applies when the number r of factors in $B_{i_1}B_{i_2}, \ldots, B_{i_r}$ is $\equiv 3(4)$ and $r < n - 1$. Similarly the method used in proving $d = 0$ also applies when $r \equiv 0(4)$.

Hence if our relation exists, it has the form

$$I = kB_1B_2, \ldots, B_{n-1}$$

the right hand term being the only survivor. Now I is symmetric so $B_1B_2, \ldots, B_{n-1}$ is symmetric i.e. $n - 1 \equiv 0$ or $3(4)$, but n is even so $n - 1 \equiv 3(4)$ i.e. $n \equiv 0(4)$

We have thus proved the following.

If an irreducible relation between the elements of $\mathcal{G}$ does exist, then $n \equiv 0(4)$.

Now square this relation to get

$$\begin{aligned} I &= k^2B_1B_2, \ldots, B_{n-1}B_1B_2, \ldots, B_{n-1} \\ &= k^2(-1)^{n-2}B_2 \ldots B_{n-1}B_2 \ldots B_{n-1} \\ &= \ldots\ldots\ldots\ldots \\ &= k^2(-1)^{\frac{1}{2}(n-1)n}I. \end{aligned}$$

Since $n \equiv 0(4)$, we see that $k^2 = 1$ i.e. $k = \pm 1$. Hence we have the following.

Lemma 4. *If $n \equiv 2(4)$ then the 2^{n-1} matrices of $\mathcal{G}$ are linearly independent, while for $n \equiv 0(4)$, they are either linearly independent or are connected by the relations which arise from the relation $I = \pm B_1B_2, \ldots, B_{n-1}$ through multiplication by the various elements of $\mathcal{G}$, but are connected by no further irreducible linear relations.*

Example: let $n = 4$. Then

$$\mathcal{G} = \{I, B_1, B_2, B_3, B_1B_2, B_2B_3, B_1B_3, B_1B_2B_3\}$$

and these eight matrices are either linearly independent or are connected by the following four irreducible linear relations and no others:

$$I = \pm B_1B_2B_3,\ B_1 = \mp B_2B_3,\ B_2 = \pm B_1B_3,\ B_3 = \mp B_1B_2.$$

These express $B_1B_2B_3$, B_2B_3, B_1B_3, B_1B_2 linearly in terms of I, B_1, B_2, B_3; so that these latter matrices are, in any case, linearly independent.

Now consider all the irreducible linear relations that exist between the element of $\mathcal{G}$. As we have seen, they are all of the type

$$G \cdot I = \pm G \cdot B_1B_2, \ldots, B_{n-1} (G \in \mathcal{G})$$

and no others. Now reduce the right side of this using (b). Then one of G or the reduced right side obviously contains fewer than half of the B's while the other contains more than half of the B's.

Thus these irreducible linear relations merely serve to express the products containing more than half of the B's in terms of those with less than half of the B's.

So in every case (i.e. irrespective of whether $n \equiv 0$ or $2 \pmod 4$) the 2^{n-2} matrices of $\mathcal{G}$, which are products of less than $\frac{n-1}{2}$ B's, are linearly independent. Hence for all values of n (necessarily even) if there is to be an identity of the type (4), the 2^{n-2} matrices of $\mathcal{G}$ consisting of the product of at most $\frac{n-2}{2}$ B's are linearly independent.

This completes the proof of Proposition 2. □

We can now give a proof of our main result.

The elements of $\mathcal{G}$ are all $n \times n$ matrices and the maximum number of linearly independent $n \times n$ matrices is n^2 since they form, over the reals, a vector space of dimension n^2. Hence by the proposition we get

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

This is satisfied if $n \leq 8$ but fails if $n = 10$. Now if it fails for $n = m$, then it fails for $n = m + 1$ for we have

$$\begin{aligned} 2^{m+1-2} &= 2 \cdot 2^{m-2} > 2 \cdot m^2 \text{ (since the relation fails for } m) \\ &> (m+1)^2 \text{ if } m \geq 3. \end{aligned}$$

It follows that if an identity of the type (4) exists, then $n \leq 8$ (and n is even). For $n = 2, 4, 8$ we already have the required type of identities. It remains to dispose off the case $n = 6$.

Suppose an identity exists for $n = 6$. Then since $6 \equiv 2(4)$, we see that

the 2^5 matrices of $\mathcal{G}$ are linearly independent. (i)

Of these 32 matrices, 16 are skew-symmetric by Lemma 1, viz. the ones that are products of 1, 2 or 5 B's. But

between any 16 *skew-symmetric* 6×6 *matrices there exists a linear relation.* (ii)

This is because the 15 matrices

$$\begin{pmatrix} 0 & 1 & 0 & \dots \\ -1 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 & \dots \\ 0 & 0 & 0 & \dots \\ -1 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}, \dots$$

with a 1 in the one place above the main diagonal, -1 in the corresponding place below, and 0's elsewhere, form a basis for the subspace of all 6×6 skew-symmetric matrices and so this subspace has dimension 15. This proves (ii).

(i) and (ii) above are contradictory. Hence no identity of type (4) can exist for $n = 6$.

That at last completes the proof of Hurwitz's theorem. □

Remark: The proof works for any field K of characteristic $\neq 2$.

Although the impossibility of the identity (4) for n $\neq 1, 2, 4, 8$ has been proved, it was under the stringent restriction that the Z_k are bilinear polynomials in the X_i and the Y_j. One

could look into the possibility of the existance of other values of n for which (4) holds, if we relax this bilinear condition and allow the Z_k to be more general polynomials in the X_i and the Y_j. However, in 1966, Frank Adams [1] showed that when n is not 1, 2, 4, 8, there are no identities of the type (4) even if the Z_k are allowed to be any bi-skew, continuous functions of the X_i and the Y_j (where a mapping $f : K^r \times K^s \to K^n$ is called bi-skew if

$$f(-\underline{x}, \underline{y}) = f(\underline{x}, -\underline{y}) = -f(x, y) \text{ for all } \underline{x} \in K^r, \underline{y} \in K^s).$$

It was thus totally unexpected when in 1965, Albrecht Pfister [10] proved the following remarkable.

Theorem 5. *Let K be a field and let $n = 2^m$ be a power of 2. Then there are identities.*

$$(X_1^2 + \cdots\cdots + X_n^2)(Y_1^2 + \cdots\cdots + Y_n^2) = Z_1^2 + \cdots\cdots + Z_n^2 \tag{5}$$

where the Z_k are linear functions of the Y_j with coefficients in $K(X_1, \ldots, X_n)$: $Z_k = \sum_j T_{kj} \cdot Y_j$ with $T_{kj} \in K(X_1, \ldots, X_n)$.

Conversely suppose n is not a power of 2. Then there is a field K such that there is no identity (5) with the $Z_k \in K(X_1, \ldots, X_n, \quad Y_1, \ldots, Y_n)$. Here the Z_k are not even demanded to be linear in the Y_j.

We shall now give proofs of theorems 3 and 5. In the process we shall get other results which are interesting in their own right.

The proof of the first part of theorem 5 requires no elaborate algebraic machinery and is, indeed, remarkably simple. We dispose of it first.

Proof of the first part of theorem 5: We use induction on m. We know that (5) holds for $m = 1, 2, 3$ (see (1), (2), (3)). Suppose it holds for m. Write $T = (T_{ij})$ so that

$$\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix} = T \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \tag{i}$$

Then (5) can be written as

$$(X_1^2 + \cdots + X_n^2)(Y_1^2 + \cdots + Y_n^2) = (X_1^2 + \cdots + X_n^2)(Y_1, \ldots, Y_n) \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

$$= (Z_1^2 + \cdots + Z_n^2) = (Z_1, \ldots, Z_n) \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$$

$$= (Y_1, \ldots, Y_n) T'T \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \text{ by (i), i.e.}$$

$$(X_1^2 + \cdots + X_n^2)(Y_1, \ldots, Y_n) I_n \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} - (Y_1, \ldots, Y_n) T'T \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = 0$$

or

$$(Y_1, \ldots, Y_n)\{(X_1^2 + \cdots + X_n^2)I_n - T'T\}\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = 0.$$

Since this is true for all $Y_1, \ldots, Y_n$, we must have

$$TT' = (X_1^2 + \cdots + X_n^2)I_n$$

and so also $T'T = (X_1^2 + \cdots + X_n^2)I_n$, T being orthogonal.

We now prove (5) for $2^{m+1} = 2n$. Write

$$(X_1, \ldots, X_{2n}) = (\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$$

where $\mathbf{X}^{(1)} = (X_1, \ldots, X_n)$ and $\mathbf{X}^{(2)} = (X_{n+1}, \ldots, X_{2n})$. By the induction hypothesis there exist two matrices $T^{(1)}, T^{(2)}$ say, corresponding to $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ respectively such that

$$\left.\begin{aligned} &(X_1^2 + \cdots + X_n^2)I_n = \mathbf{X}^{(1)}\mathbf{X}^{(1)'}I_n = T^{(1)}T^{(1)'} = T^{(1)'}T^{(1)} \\ &\text{and} \\ &(X_{n+1}^2 + \cdots + X_{2n}^2)I_n = \mathbf{X}^{(2)}\mathbf{X}^{(2)'}I_n = T^{(2)}T^{(2)'} = T^{(2)'}T^{(2)} \end{aligned}\right\} \quad \text{(ii)}$$

and we wish to show that there exists a matrix T say, such that

$$T'T = (X_1^2 + \cdots + X_n^2 + X_{n+1}^2 + \cdots + X_{2n}^2)I_{2n}. \quad \text{(iii)}$$

Try $T = \begin{pmatrix} T^{(1)} & T^{(2)} \\ T^{(2)} & X \end{pmatrix}$-a partitional matrix, where X will be determined by (iii). We have

$$T'T = \begin{pmatrix} T^{(1)'} & T^{(2)'} \\ T^{(2)'} & X' \end{pmatrix}\begin{pmatrix} T^{(1)} & T^{(2)} \\ T^{(2)} & X \end{pmatrix},$$

and using block multiplication of matrices this equals

$$\begin{pmatrix} T^{(1)'}T^{(1)} + T^{(2)'}T^{(2)} & T^{(1)'}T^{(2)} + T^{(2)'}X \\ T^{(2)'}T^{(1)} + X'T^{(2)} & T^{(2)'}T^{(2)} + X'X \end{pmatrix}$$

$$= \begin{pmatrix} (X_1^2 + \cdots + X_n^2 + X_{n+1}^2 + \cdots + X_{2n}^2)I_n & A \\ B & C \end{pmatrix}$$

say; we want to choose X so that $A = B = 0$ and

$$C = (X_1^2 + \cdots + X_n^2 + X_{n+1}^2 + \cdots + X_{2n}^2)I_n.$$

To make $A = 0$ we have to have $X = -T^{(2)'^{-1}} T^{(1)'} T^{(2)}$. This automatically makes $B = 0$ (just check). Now it seems too much to expect C to be what we want. But we have

$$\begin{aligned} C &= T^{(2)'} T^{(2)} + T^{(2)'} T^{(1)} T^{(2)'^{-1}{}'} T^{(2)'^{-1}} T^{(1)'} T^{(2)} \\ &= (X_{n+1}^2 + \cdots + X_{2n}^2) I_n + (X_{n+1}^2 + \cdots + X_{2n}^2)^{-1} T^{(2)'} T^{(1)} T^{(1)'} T^{(2)} \\ &= (X_{n+1}^2 + \cdots + X_{2n}^2) I_n + (X_{n+1}^2 + \cdots + X_{2n}^2)^{-1} \\ &\quad (X_1^2 + \cdots + X_n^2) T^{(2)'} T^{(2)} \\ &= (X_{n+1}^2 + \cdots + X_{2n}^2) I_n + (X_1^2 + \cdots + X_n^2) I_n \\ &= (X_1^2 + \cdots + X_n^2 + X_{n+1}^2 + \cdots + X_{2n}^2) I_n. \end{aligned}$$

This completes the proof of the first part of Theorem 5. □

We now come to a result of Cassels [4], which, in a way, was the starting point of this whole business and which is an indispensible tool in our further developments.

Cassels' Lemma (1964). *Let $f(X) \in K(X)$ be a polynomial with coefficients in K. If $f(X)$ is a sum of n squares of elements of the field $K(X)$, then it is a sum of n squares of elements of the ring $K[X]$.*

Note: What is new in this enunciation is that the same number n of squares suffice in $K[X]$; without this condition, the result had been proved by Artin [3].

Proof: There are three trivial cases of the lemma which we dispose of first.

(i) $n = 1$. Then $f(X) = (p(X)|q(X))^2$, *so* $q(X)\big|p(X)$.

(ii) char $K = 2$. Then $a^2 + b^2 = (a + b)^2$ and so if

$$f(X) = \gamma_1^2(X) + \cdots + \gamma_n^2(X)$$

then combining two squares at a time into one, $f(X)$ reduces to a single square, i.e. we land up in case (i).

(iii) -1 is a sum of $n - 1$ squares of elements in K.

Say $-1 = b_1^2 + \cdots + b_{n-1}^2$. Then for any $f(X)$, we have

$$\begin{aligned} f(X) &= \left(\frac{f+1}{2}\right) - \left(\frac{f-1}{2}\right)^2 \\ &= \left(\frac{f+1}{2}\right)^2 + \left(b_1 \frac{(f-1)}{2}\right)^2 + \cdots + \left(b_{n-1} \frac{(f-1)}{2}\right)^2 \end{aligned}$$

a sum of n squares of elements of $K[X]$.

So now let us suppose none of these three cases holds and let

$$f(X) = (p_1(X)/q_1(X))^2 + \cdots + (p_n(X)/q_n(X))^2.$$

Dropping the X from now on and clearing the denominators, this gives

$$fZ^2 = Y_1^2 + \cdots + Y_n^2, \quad Z, Y_1, \ldots, Y_n \in K[X], Z \neq 0.$$

Thus the equation

$$fZ^2 = Y_1^2 + \cdots + \cdots + Y_n^2 \tag{a}$$

has a solution $(Z, Y_1, \ldots, Y_n)$ with $Z \neq 0$ and we have to show that there exists a solution of (a) with $Z \in K (Z \neq 0)$, i.e. with degree of Z (in X) $= 0$. Now since (a) has a solution with $Z \neq 0$, so there is a solution, call it $(\zeta, \eta_1, \ldots, \eta_n)$, with $\zeta \neq 0$ for which deg ζ is as small as possible:

$$f\zeta^2 = \eta_1^2 + \cdots + \eta_n^2. \tag{b}$$

We shall show that this degree is 0 i.e. that $\zeta \in K$, by showing that if not, then there exists a solution, say $(\zeta^*, \eta_1^*, \ldots, \eta_n^*)$ with $\zeta^* \neq 0$ and deg $\zeta^* <$ deg ζ.

So suppose deg $\zeta > 0$. By the division al gorithm in $K[X]$, we can write, for $j = 1, 2, \ldots, n$,

$$\eta_j = \lambda_j \zeta + \gamma_j$$

where either $\gamma_j = 0$ or deg $\gamma_j <$ deg ζ. i.e.

$$\eta_j/\zeta = \lambda_j + \gamma_j/\zeta = \lambda_j + \Lambda_j, \tag{c}$$

say. Note that not all the γ_j can be zero, otherwise ζ divides all the η_j so that (b) becomes $f = \lambda_1^2 + \cdots + \lambda_n^2$ -a contradiction, since the degree of ζ was least possible.

Now let

$$\zeta^* = \zeta \left\{ \sum_i \lambda_i^2 - f \right\} - 2 \left\{ \sum_i \lambda_i \eta_i - f\zeta \right\}$$

and

$$\eta_j^* = \eta_j \left\{ \sum_i \lambda_i^2 - f \right\} - 2\lambda_j \left\{ \sum \lambda_i \eta_i - f\zeta \right\}.$$

Then visibly, all of $\zeta^*, \eta^*, \ldots, \eta_n^* \in K[X]$. We now claim

(i) that $(\zeta^*, \eta_1^*, \ldots, \eta_n^*)$ is a solution of (a),
(ii) $\zeta^* \neq 0$ and
(iii) deg $\zeta^* <$ deg ζ.

This would then contradict the definition of ζ and so would prove the lemma.

We prove (i) by brute force: we must show that $\sum_j \eta_j^{*2} - f\zeta^{*2} = 0$ i.e. that

$$\sum_j \left[\eta_1^2 \left\{ \sum_i \lambda_i^2 - f \right\}^2 + 4\lambda_j^2 \left\{ \sum_i \lambda_i \eta_i - f\zeta \right\}^2 \right.$$
$$\left. -4\lambda_j \eta_j \left\{ \sum_i \lambda_i^2 - f \right\} \left\{ \sum_i \lambda_i \eta_i - f\zeta \right\} \right]$$
$$= f \left[\zeta^2 \left\{ \sum_i \lambda_i^2 - f \right\}^2 + 4 \left\{ \sum_i \lambda_i \eta_i - f\zeta \right\}^2 \right.$$
$$\left. -4\zeta \left\{ \sum_i \lambda_i^2 - f \right\} \left\{ \sum_i \lambda_i \eta_i - f\zeta \right\} \right]$$

Here the first terms from both sides cancel since $\sum \eta_j^2 = f\zeta^2$ and it remains to prove that

$$4\left\{\sum_i \lambda_i\eta_i - f\zeta\right\}\left[\left(\sum \lambda_i\eta_i - f\zeta\right)\sum \lambda_j^2 - \left(\sum_i \lambda_i^2 - f\right)\sum_j \lambda_j\eta_j - \left(\sum_i \lambda_i\eta_i - f\zeta\right)f + \left(\sum \lambda_i^2 - f\right) + f\zeta\right] = 0$$

Here the expression in square brackets just cancels out.

To prove (ii) and (iii) we substitute for λ_j from (c) in ζ^*. Then

$$\begin{aligned}\zeta^* &= \zeta\left(\sum_i\left(\frac{\eta_i^2}{\zeta^2} + \Lambda_i^2 - \frac{2\eta_i\Lambda_i}{\zeta}\right) - f\right) - 2\left(\sum_i\left(\frac{\eta_i}{\zeta} - \Lambda_i\right)\eta_i - \eta\zeta\right)\\ &= \zeta(\Lambda_i^2 + \cdots + \Lambda_n^2) \text{ (using } f\zeta^2 = \eta_i^2 + \cdots + \eta_n^2)\\ &= \zeta\sum_i \gamma_i^2/\zeta^2 = 1/\zeta\sum_i \gamma_i^2.\end{aligned}$$

Here not all the γ_i are zero (as already noted) and so $\sum \gamma_i^2$ is non-zero since otherwise by equating the coefficient of the highest power in X to 0, we find that 0 is a sum of at most n squares of elements of K, which is the third trivial case of the lemma. Thus $\zeta^* \neq 0$, which proves (ii).

Finally $\zeta^* = 1/\zeta\sum_i \gamma_i^2$ giving $\zeta\zeta^* = \sum_i \gamma_i^2$. Equating degrees, we get deg ζ + deg $\zeta^* = 2\max_i$ (deg γ_i) $<$ 2 deg ζ since deg $\gamma_i <$ deg ζ (for all i). Thus deg $\zeta^* <$ deg ζ, which proves (iii).

This completes the proof of Cassels' lemma. □

Remark: The solution $(\zeta^*, \eta_1^*, \ldots, \eta_n^*)$ does not just come out of the blue. It is the second point of intersection Q of the quadric (a) with the line joining the points $P = (\zeta, \eta_1, \ldots, \eta_n)$ (on the quadric) and $P' = (1, \lambda_1, \ldots, \lambda_n)$ (in space) in the n-dimensional projective space over the field $K(X)$. The simplest way to get this point Q is as follows: a general point of the line PP' is

$$(\theta\zeta + \varphi, \theta\eta_1 + \varphi\lambda_1, \ldots, \theta\eta_n + \varphi\lambda_n)$$

θ/φ being a parameter for various points, $\varphi = 0$ giving the point P. To get Q we substitute this general point in the quadric (a):

$$f(\theta^2\zeta^2 + \varphi^2 + 2\theta\varphi\zeta) = \sum_{j=1}^{n}(\theta^2\eta_j^2 + \varphi^2\lambda_j^2 + 2\theta\varphi\eta_j\lambda_j)$$

i.e. $\theta^2(\zeta^2 f - \sum\eta_j^2) + 2\theta\varphi(f\zeta - \sum\lambda_j\eta_j) + \varphi^2(f - \sum\lambda_j^2) = 0$. But $\zeta^2 f = \sum\eta_j^2$ so this becomes $2\theta\varphi(f\zeta - \sum\lambda_j\eta_j) + \varphi^2(f - \sum\lambda_j^2) = 0$. This has a root $\varphi = 0$ as expected giving the point P. The other root is $\theta/\varphi = -(\sum\lambda_j^2 - f)/2(\sum\lambda_j\eta_j - \zeta f)$,

and substituting this in the general point and multiplying by a suitable factor (allowed in a projective space) we get our point Q as required.

We now deduce a few corollaries from this lemma.

Corollary 1. *Let* char $K \neq 2$ *let* $f(X_1, \ldots, X_m) \in K(X_1, \ldots, X_m)$ *be a sum of* n *squares of elements of* $K(X_1, \ldots, X_m)$. *Let* $a_1, a_2, \ldots, a_m \in K$ *be such that* $f(a_1, \ldots, a_m)$ *is defined (i.e. the dominator is not* 0). *Then* $f(a_1, \ldots, a_m)$ *is a sum of* n *squares in* K.

Remark: The point is that although $f(X_1, \ldots, X_m)$ is defined at $(a_1, \ldots, a_m)$, it may well happen that the summands $f_j^2(X_1, \ldots, X_m)$ of the right hand side of $f(X_1, \ldots, X_m) = f_1^2 + \cdots + f_n^2$ may not be defined at $(a_1, \ldots, a_m)$, but still according to the corollary, $f(a_1, \ldots, a_m)$ is a sum of n squares in K.

Proof: We use induction on m. For $m = 1$, we have

$$f(X) = g(X)/h(X) = \gamma_1^2(X) + \cdots + \gamma_n^2(X).$$

Then $gh = (\gamma_1 h)^2 + \cdots + (\gamma_n h)^2$. Thus gh, which is in $K[X]$, is a sum of n squares in $K(X)$ and so by Cassels' lemma, it is a sum of n squares in $K[X]$:

$$gh = f_i^2 + \cdots + f_n^2 \ (f_j \in K[X]).$$

Hence $g(X)/h(X) = (\frac{f_1(X)}{h(X)})^2 + \cdots + (\frac{f_n(X)}{h(X)})^2$. Now by hypothesis, $f(a) = g(a)/h(a)$ is defined; i.e. $h(a) \neq 0$, so each $f_j(a)/h(a)$ is defined.

Let now $m > 1$. Let $L = K(X_1, \ldots, X_{m-1})$. Assume the result for $m - 1$ variables and let $g(X_1, \ldots, X_m)/h(X_1, \ldots, X_m)$ be a rational function which is a sum of n squares in $K(X_1, \ldots, X_m)$. Regard g/h as a rational function of X_m belonging to $L(X_m)$. So by the case $m = 1$, we see that $g(X_1, \ldots, X_{m-1}, a_m)/h(X_1, \ldots, X_{m-1}, a_m)$ is a sum of n squares in $L = K(X_1, \ldots, X_{m-1})$. So by the induction hypothesis $g(a_1, \ldots, a_m)/h(a_1, \ldots, a_m)$ is a sum of n squares in K. This completes the proof of the corollary. □

Corollary 2. *Suppose* $n = 2^m$. *Let* G_n *be the set of all non-zero elements of* K *which are sums of* n *squares in* K. *Then* G_n *is a group under multiplication.*

Proof: Let $\alpha\beta \in G_n$ say, $\alpha = \alpha_1^2 + \cdots + \alpha_n^2$, $\beta = \beta_1^2 + \cdots + \beta_n^2$. Then $\alpha^{-1} = \alpha/\alpha^2 = (\alpha_1/\alpha)^2 + \cdots + (\alpha_n/\alpha)^2 \in G_n$ and it remains to prove that $\alpha\beta \in G_n$. Consider the identity

$$(X_1^2 + \cdots + X_n^2)(Y_1^2 + \cdots + Y_n^2) = Z_1^2 + \cdots + Z_n^2$$

which exists since $n = 2^m$. In this let $X_1 \to \alpha_1, \ldots, X_n \to \alpha_n, Y_1 \to \beta_1, \ldots, Y_n \to \beta_n$. Then the left side is well defined and equal to $\alpha\beta$ and so by Corollary 1, the right side is a sum of n squares of elements of K, i.e. $\alpha\beta \in G_n$ as required. □

We see that it is the identity (5) that does the trick.

We can now prove the first part of Theorem 3: $s(K)$ *is always a power of* 2.

Proof of the first part of Theorem 3: Let

$$n = 2^m \leq s(K) < 2^{m+1} \qquad (*)$$

Then $a_1^2 + \cdots + a_n^2 + a_{n+1}^2 + \cdots + a_s^2 + 1 = 0$ $(a_j \in K)$. Let $A = a_1^2 + \cdots + a_n^2$, $B = a_{n+1}^2 + \cdots + a_s^2 + 1$. Here A, B are both non-zero, otherwise $s(K) < s$. Also A, B, both $\in G_n$ (by adding a suitable number of 0^2's to B if necessary). Then $A + B = 0$ so $A = -B$ i.e. $-1 = B/A \in G_n$ since G_n is a group i.e. $-1 = c_1^2 + \cdots + c_n^2$ giving $s(K) \leq n$. Comparing with $(*)$, we get $s(K) = n = 2^m$.

To prove the remaining parts of Theorems 3 and 5 we need to deduce some more corollaries from Cassels' lemma; see [4].

Corollary 3. *Let* char $K \neq 2$. *A necessary and sufficient condition for* $X^2 + d \in K[X]$ *to be a sum of* n *squares in* $K(X)$ *(and so in* $K[X]$ *by Cassels' lemma) is that either*

(i) -1 *is a sum of* $n-1$ *squares in* K *or*
(ii) d *is a sum of* $n-1$ *squares in* K.

Proof: If $-1 = b_1^2 + \cdots + b_{n-1}^2$, then for any polynomial $f(X) \in K[X]$, we have

$$\begin{aligned} f &= \left(\frac{f+1}{2}\right)^2 - \left(\frac{f-1}{2}\right)^2 \\ &= \left(\frac{f+1}{2}\right)^2 + \left(\frac{b_1(f-1)}{2}\right)^2 + \cdots + \left(\frac{b_{n-1}(f-1)}{2}\right)^2. \end{aligned}$$

In particular $X^2 + d$ is a sum of n squares.

If d is a sum of $n-1$ squares then visibly $X^2 + d$ is a sum of n squares in $K[X]$.

For the converse, suppose $X^2 + d$ is a sum of n squares in $K[X]$. If (i) holds, well and good; otherwise let $X^2 + d = p_1^2(X) + \cdots + p_n^2(X)$ say. Here we may suppose the $p_j(X)$ to be linear poynomials in X for if not, then equating to 0 the coefficient of the highest power of X gives (i). Then

$$X^2 + d = (a_1X + b_1)^2 + \cdots + (a_nX + b_n)^2 \qquad (*)$$

Now one of the equations $C = \pm(a_nC + b_n)$ is always solvable in K. For if $a_n \neq 1$ then $C = +(a_nC + b_n)$ is solvable, while if $a_n = 1$ then $C = -(a_nC + b_n)$ is solvable since char $K \neq 2$. Now put $X = C$ in $(*)$:

$$C^2 + d = (a_1C + b_1)^2 + \cdots + (a_{n-1}C + b_{n-1})^2 + (a_nC + b_n)^2.$$

Cancelling C^2 with $(a_nC + b_n)^2$ we see that d is a sum of $n-1$ squares in K. This completes the proof. □

Corollary 4. *Let* $\mathbf{R}$ *be the field of real numbers. Then* $X_1^2 + \cdots + X_n^2$ *is not a sum of* $n-1$ *squares of elements in* $\mathbf{R}(X_1, \ldots, X_n)$.

Proof: We use induction on n. For $n = 1$, the result is trivial. So suppose the result is true for $n - 1$. Let $K = \mathbf{R}(X_1, \ldots, X_{n-1})$, $X_n = X$ and $d = X_1^2 + \cdots + X_{n-1}^2$. If $X_1^2 + \cdots + X_n^2$ is a sum of $n - 1$ squares in $K(X) = \mathbf{R}(X_1, \ldots, X_n)$, then by Corollary 3, $d = X_1^2 + \cdots + X_{n-1}^2$ is a sum of $n - 2$ squares in K, since -1 is clearly not a sum of $n - 2$ squares in K-indeed not a sum of squares at all in K, which is formally real. This contradicts the induction hypothesis and completes the proof of Corollary 4. □

We are now in a position to complete the proofs of the remaining parts of Theorems 3 and 5.

Every power of 2 *is the Stufe of some field* K.

Proof: Let $n = 2^m$ and let $K = \mathbf{R}(X_1, \ldots, X_{n+1}, Y)$ where $X_1, \ldots, X_{n+1}$ are independent transcendentals over $\mathbf{R}$ and Y satisfies the equation

$$Y^2 + X_1^2 + \cdots + X_{n+1}^2 = 0 \tag{i}$$

We claim that $s(K) = n = 2^m$. In any case by (i), $s(K) \leq n + 1$ and so is at most n since $n + 1$ cannot be a power of 2 whereas $s(K)$ *is* (except in the trivial case $n = 1$ i.e. $m = 0$).

If $s(K) < n$ then there exist $t_1, \ldots, t_n \in K$, not all zero such that

$$t_1^2 + \cdots + t_n^2 = 0 \tag{ii}$$

Let $L = \mathbf{R}(X_1, \ldots, X_{n+1})$ so that $K = L(Y)$. By (i), Y is algebraic over L of degree 2 and so each element of K is a linear polynomial in Y with coefficients from L. Write $t_j = a_j + Yb_j, a_j, b_j \in L$. Then by (ii) we see that

$$\sum a_j^2 + Y^2 \sum b_j^2 = 0$$

and

$$\sum a_j b_j = 0.$$

Here not all the a_j are zero, otherwise $\sum b_j^2 = 0$ and so each $b_j = 0$ since the $b_j \in L = \mathbf{R}(X_1, \ldots, X_{n+1})$ is formally real. Then each t_j would be zero which is not true. Similarly not all the b_j are zero. Hence

$$\begin{aligned} -Y^2 &= \sum_{j=1}^{n} a_j^2 / \sum_{j=1}^{n} b_j^2 \in G_n \text{ (by the group property of } G_n) \\ &= c_1^2 + \cdots + c_n^2, \text{ say } c_j \in L, \end{aligned}$$

i.e. $X_1^2 + \cdots + X_{n+1}^2$ is a sum of n squares in L which contradicts Corollary 4. Thus $s(K)$ is *not* less than n and so equals n. This completes the proof. □

Remark: The proof also works for $\mathbf{Q}(X_1, \ldots, X_{n+1}, Y)$.

Finally we prove the remaining part of Theorem 5.

Suppose n *is not a power of* 2. *Then there is a field* K *such that there is no identity*

$$(X_1^2 + \cdots + X_n^2)(Y_1^2 + \cdots + Y_n^2) = z_1^2 + \cdots + Z_n^2$$

with $Z_j \in K(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$.

Proof: Let $2^{m-1} < n < 2^m$. Let K be a field having Stufe $2^m = \nu$, say. Then $a_1^2+\cdots+a_n^2+a_{n+1}^2+\cdots+a_\nu^2+1=0$. Let $A=a_1^2+\cdots+a_n^2$, $B=a_{n+1}^2+\cdots+a_\nu^2+1$; hence $A, B \in G_n$ and if an identity of the above type exists, then G_n is a group (see the proof of Corollary 2). So $-1 = B/A \in G_n$, i.e. $-1 = C_1^2+\cdots+C_n^2$ ($C_j \in K$) hence $s(K) \leq n < \nu$. But K was chosen to have Stufe ν. This gives a contradiction and so completes the proof. □

Remark 1. In our examples of fields with high Stufe both the fields $\mathbf{R}(X_1, \ldots, X_{n+1}, Y)$ and $\mathbf{Q}(X_1, \ldots, X_{n+1}, Y)$ are of high transcendence degree over $\mathbf{R}$ or $\mathbf{Q}$ as the case may be. We have the following

Problem: *Does high Stufe always imply high degree of transcendence? (over* $\mathbf{R}$ *or* $\mathbf{Q}$*).*

Let us now go back to the identity (4):

$$(X_1^2+\cdots+X_n^2)(Y_1^2+\cdots+Y_n^2)=(Z_1^2+\cdots+Z_n^2)$$

where the Z_k are bilinear functions of the X_i and the Y_j with coefficients in the field K.

There are three obvious ways of generalizing this identity (one of which we have already looked at in theorem 5). They are

a) Allow the Z_k to be the rational functions of the X_i and the Y_j : $Z_k \in K(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$. Then, as we have seen in theorem 5, such identities can be found for each power $n = 2^m (m = 0, 1, \ldots)$ of 2 and for no other value of n.
b) Consider the (r, s, n) identity

$$(X_1^2+\cdots+X_r^2)(Y_1^2+\cdots+Y_s^2)=Z_1^2+\cdots+Z_n^2 \tag{6}$$

with Z_k bilinear in the X_i and the Y_j, and determine, for given r, s the least value of n for which (6) holds. We could, alternatively look for the maximum value of r, for given s and n, for which (6) holds.

For general values of r, s, n, little is known about (6). However, for $s = n$, Hurwitz and Radon gave a solution of (6) in 1922–3, for the field R of real numbers. Before giving the exact statement of the Hurwitz-Radon theorem we make the following

Definition 2. We say the triple (r, s, n) is *admissible* over K if (6) holds.

Thus (r, s, rs) is trivially admissible over any field K; so that what we want is the most economical n for which (r, s, n) is admissible for a given pair r, s of integers. In view of this we have the

Definition 3. We denote by r^*s (or rather r_k^*s) the least n for which (r, s, n) is admissible/K.

We have the trivial bounds.

$$\max(r, s) \leq r^*s \leq r.s$$

It is not easy to determine r^*s, even for small values or r, s.

Alternatively, as already said above, we could ask, for given s, n the maximum value of r for which (r, s, n) is admissible/K. This is the approach adopted by Hurtwitz and Radon in their treatment of (6) for the field R of real numbers. Simultaneously, Hurwitz solved (6), in this special case, for the field C of complex numbers, published posthumously in 1923. Various authors have since dealt with other fields.

As illustrations of definitions 2 and 3, we have the following:

Examples:

(i) (n, n, n) is admissible over **R,** indeed over any field K, char $K \neq 2$, iff $n = 1, 2, 4, 8$. Thus is Hurwitz's theorem (Theorem 4).

(ii) $(1, n, n)$, and indeed (r, s, rs), is admissible for all n, r, s over any field K.

(iii) If char$K = 2$, then $r *_K s = 1$ for all r, s for then $a^2 + b^2 = (a + b)^2$.

(iv) $8 * 8 = 8$ for $\max(8, 8) \leq 8 * 8 \leq 8$. Similarly $4 * 4 = 4$ and $2 * 2 = 2$.

(v) *The* 16*-square problem*: Before Hurwitz, studies about the (r, s, n)-identites (6) were exclusively restricted to the polynomial ring $\mathbf{Z}[X_1, \ldots, X_r, Y_1, \ldots, Y_s]$ over **Z.** One then speaks of the $(r, s, n)_{\mathbf{Z}}$-identities. It has recently been confirmed that $16 *_{\mathbf{Z}} 16 = 32$, thereby completing the solution of the so-called 16-square problem in the integer coefficient (case see [16]). However, the integer $\nu = 16 *_{\mathbf{R}} 16$ is not known to date. Various methods developed by K.Y. Lam and J. Adem narrow down the range of ν to $23 \leq \nu \leq 32$. The values 23, 24 were subsequently ruled out by Lam and Yuzvinski. By going more deeply into the geometry of sums of square formulae and using sophisticated algebraic topology, it has now been established by Lam and Yiu that $29 \leq \nu \leq 32$.

It is trivial to see that $\nu \leq 32$; indeed

$$\left(\sum_1^{16} X_j^2\right)\left(\sum_1^{16} Y_j^2\right) = \left(\sum_1^{8} X_j^2 + \sum_9^{16} X_j^2\right)\left(\sum_1^{8} Y_j^2 + \sum_9^{16} Y_j^2\right)$$

which is a sum of 32 squares, using the 8-square identity four times.

(vi) Amongst small values of r, s, n, even (10, 11, 25) is not known to be admissible or otherwise.

Definition 4. For any positive integer n, define the so-called *Radon function* $\rho(n)$ as follows:

Write $n = 2^m \cdot u(u \text{ odd})$; then

$$\rho(n) = \begin{cases} 2m + 1 \\ 2m \\ 2m \\ 2m + 2 \end{cases} \text{ according as } m \equiv \begin{cases} 0 \\ 1 \\ 2 \\ 3 \end{cases} \text{ (modulo 4).}$$

Equivalently write $m = 4a + b, 0 \leq b \leq 3$; then

$$\rho(n) = 8a + 2^b.$$

We now have the following

Theorem B (Radon, Hurwitz-1922, 1923). *The triple (r, n, n) is admissible over the field of real numbers (indeed over any field K, char $K \neq 2$) iff $r \leq \rho(n)$.*

For a proof see [13], p 127 or [14].

(c) Instead of a "product formula" (4) for the form $X_1^2 + \cdots + X_n^2$, look for such a formula for more general quadratic forms $q(X_1, \ldots, X_n)$; i.e. determine other quadratic forms $q(X_1, \ldots, X_n)$, if any, for which

$$q(X_1, \ldots, X_n) \cdot q(Y_1, \ldots, Y_n) = q(Z_1, \ldots, Z_n) \tag{7}$$

where $Z_k \in K(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$

Examples: (i) If $q(X_1, X_2) = X_1^2 + aX_2^2 (a \in K)$, then we have the curious identity (cf. identity (1))

$$(X_1^2 + aX_2^2)(Y_1^2 + aY_2^2) = (X_1Y_1 + aX_2Y_2)^2 + a(X_1Y_2 - aX_2Y_1)^2.$$

(ii) In 4 variables, we have the striking identity (cf. identity(2))

$$\begin{aligned}(X_1^2 + aX_2^2 + bX_3^2 + baX_4^2)(Y_1^2 + aY_2^2 + bY_3^2 + baY_4^2) \\ = (X_1Y_1 + aX_2Y_2 + bX_3Y_3 + abX_4Y_4)^2 \\ + a(-X_1Y_2 + X_2Y_1 - bX_3Y_4 + bX_4Y_3)^2 \\ + b(-X_1Y_3 + X_3Y_1 + aX_2Y_4 - aX_4Y_2)^2 \\ + ab(-X_1Y_4 + X_4Y_1 - bX_2Y_3 + bX_3Y_2)^2\end{aligned}$$

Pfister has given a complete solution of (c). He shows that for every power $n = 2^m$ of 2, there is this form in n variables generalizing the forms $X_1^2 + aX_2^2$ and $X_1^2 + aX_2^2 + bX_3^2 + abX_4^2$ by an obvious induction, which satisfies a product identity. These are the so called Pfister forms. Further there are no other forms that satisfy a product formula (7). For a proof of these results see [10], [11], [13].

In the identity (5), we have proved that the Z_k may be chosen linear functions of the Y_j with coefficients in $K(X_1, \ldots, X_n)$:

$$Z_k = \sum_{j=1}^{n} T_{kj} Y_j \text{ with } T_{kj} \in K(X_1, \ldots, X_n).$$

For $n = 2, 4$, and 8, these T_{kj} are linear in the X_i as well. It is natural to enquire how simple we can take the T_{kj} as functions of the X_i. By theorem 4, they can not all be taken linear in the X_i nor indeed polynomials, by Frank Adam's theorem so that some of them at least have to have a denominator; but can we make at least some of them linear forms in the X_i?

Let us first see what we can do with the first term Z_1 and prove the following

Theorem 6. *Let $n = 2^m$ and let $X_1, \ldots, X_n, Y_1, \ldots, Y_n \in K$. Then*

$$(X_1^2 + \cdots + X_n^2)(Y_1^2 + \cdots + Y_n^2) = (X_1Y_1 + \cdots + X_nY_n)^2 + Z_2^2 + \cdots + Z_n^2$$

for some $Z_2, \ldots, Z_n \in K$.

We first prove: Let $n = 2^m$ and let $c = c_1^2 + \cdots + c_n^2$ ($c_j \in K$). Then there exists an $n \times n$ matrix S with first row $(c_1, \ldots, c_n)$ such that $SS' = S'S = cI_n$.

Proof: First let $c = 0$. If all the $c_j = 0$, we take S to be the zero matrix.

So suppose, say, $c_1 \neq 0$. Let R be the row vector $(c_1, \ldots, c_n)$ and take $S = c_1^{-1}R'R$, which has first row R as required. Further

$$\begin{aligned} SS' &= c_1^{-1}R'Rc_1^{-1}R'R \\ &= c_1^{-2}R'(RR')R \\ &= 0, \end{aligned}$$

since $RR' = c_1^2 + \cdots + c_n^2 = c = 0$. Similarly $S'S = 0$ and the proof is complete. So we may now suppose that $c \neq 0$ and we proceed by induction on m.

Write

$$\begin{aligned} \underline{R} = (c_1, \ldots, c_{2^m}) &= (c_1, \ldots, c_{2^{m-1}}, c_{2^{m-1}+1}, \ldots, c_{2^m}) \\ &= (\underline{R}_1, \underline{R}_2). \end{aligned}$$

Let $a = c_1^2 + \cdots + c_{2^{m-1}}^2$, $b = c_{2^{m-1}+1}^2 + \cdots + c_{2^m}^2$, so that $c = a + b$. Here since $c \neq 0$, so a, b cannot be both zero; say, without loss of generality, that $a \neq 0$. By the induction hypothesis, there exist square matrices S_1, S_2 of size 2^{m-1} such that

$$\begin{aligned} S_1S_1' &= S_1'S_1 = aI_{2^{m-1}} \\ S_2S_2' &= S_2'S_2 = bI_{2^{m-1}}. \end{aligned}$$

Furthermore the first row of S_1 is $(c_1, \ldots, c_{2^{m-1}})$ and that of S_2 is $(c_{2^{m-1}+1}, \ldots, c_2^m)$. Now let

$$S = \begin{pmatrix} S_1 & S_2 \\ -a^{-1}S_1'S_2'S_1 & S_1' \end{pmatrix}.$$

This has first row equal to R as required and an easy matrix computation gives $SS' = S'S = cI_n$, e.g.

$$\begin{aligned} SS' &= \begin{pmatrix} S_1 & S_2 \\ -a^{-1}S_1'S_2'S_1 & S_1' \end{pmatrix} \begin{pmatrix} S_1' & -a^{-1}S_1'S_2S_1 \\ S_2' & S_1 \end{pmatrix} \\ &= \begin{pmatrix} aI_{2^{m-1}} + bI_{2^{m-1}} & -a^{-1}aS_2S_1 + S_2S_1 \\ -a^{-1}S_1'S_2'aI_{2^{m-1}} + S_1'S_2' & a^{-2}S_1'S_2'S_1S_1'S_2S_1 + S_1'S_1 \end{pmatrix} \\ &= \begin{pmatrix} cI_{2^{m-1}} & 0 \\ 0 & bI_{2^{m-1}} + aI_{2^{m-1}} \end{pmatrix} = cI_{2^m}. \end{aligned}$$

□

Proof of Theorem 6: Write

$$\begin{aligned} X &= X_1^2 + \cdots + X_n^2 \\ Y &= Y_1^2 + \cdots + Y_n^2. \end{aligned}$$

Then there exist $n \times n$ matrices U, V such that $UU' = U'U = XI_n$, $VV' = V'V = YI_n$ and

$$U \text{ has 1st row} = (X_1, \ldots, X_n),$$
$$V \text{ has 1st row} = (Y_1, \ldots, Y_n).$$

Then

$$XYI_n = XVV' = V(U'U)V' = (VU')(VU')' = V(U'U)V' = WW',$$
$$\text{where } W = VU'.$$

This equation says that if $(Z_1, \ldots, Z_n)$ is the first row of W, then $XY = Z_1^2 + \cdots + Z_n^2$. But since $W = VU'$, we have $Z_1 = X_1Y_1 + \cdots + X_nY_n$. □

Theorem 6 enables us to give another proof of the important group property of the set. $G_n = \{a \in K^* | a = \alpha_1^2 + \cdots + \alpha_n^2, \alpha_j \in K\}$, when $n = 2^m$, a power of 2 (see corollary 2 after cassels' lemma). For, theorem 6 implies closure of G_n under multiplication while as before $a^{-1} \in G_n$ whenever $a \in G_n$.

Going back to our enquiry about how simple we can take the Z_k as functions of the X_i, we now state the following striking result of Shapiro:

Theorem 7 (Shapiro-1978). *Let $n = 2^m$ and let K be any field. In the n-square identity* (5)

$$(X_1^2 + \cdots + X_n^2)(Y_1^2 + \cdots + Y_n^2) = Z_1^2 + \cdots + Z_n^2 \qquad (*)$$

with Z^k linear in the Y_j with coefficient in $K(X_1, \ldots, X_n)$, the first r terms $Z_1, \ldots, Z_r$ of the right side of (5) *can also be taken linear in the X_i iff $r \leq \rho(n)$, the Radon function!*

For a proof of this result, see [13], page 183.

Incidentally, we note that in $(*)$ above we can easily arrange a formula where 8 of the Z_k are bilinear (when $n \geq 8$). To do this start with the known (8, 8, 8) bilinear identity and apply the "doubling" process given in Pfister's theorem 5. Indeed write the 8-square identity twice over, once for the variables $X_1, \ldots, X_8; Y_1, \ldots, Y_8; Z_1, \ldots, Z_8$ and once for $X_9, \ldots, X_{16}; Y_9, \ldots, Y_{16}; Z_9, \ldots, Z_{16}$. Thus

$$\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_8 \end{pmatrix} = \begin{pmatrix} X_1 & -X_2 & -X_3 & -X_4 & -X_5 & -X_6 & -X_7 & -X_8 \\ X_2 & X_1 & -X_4 & X_3 & -X_6 & X_5 & X_8 & -X_7 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ X_8 & X_7 & -X_6 & -X_5 & X_4 & X_3 & -X_2 & -X_1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_8 \end{pmatrix}$$

and

$$\begin{pmatrix} Z_9 \\ Z_{10} \\ \vdots \\ Z_{16} \end{pmatrix} = \begin{pmatrix} X_9 & -X_{10} & -X_{11} & -X_{12} & -X_{13} & -X_{14} & -X_{15} & -X_{16} \\ X_{10} & X_9 & -X_{12} & X_{11} & -X_{14} & X_{13} & X_{16} & -X_{15} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ X_{16} & X_{15} & -X_{14} & -X_{13} & X_{12} & X_{11} & -X_{10} & X_9 \end{pmatrix} \begin{pmatrix} Y_9 \\ Y_{10} \\ \vdots \\ Y_{16} \end{pmatrix},$$

(simply read off the identity (3)); say $\underline{Z}_1 = S_1 \underline{Y}_1$ and $\underline{Z}_2 = S_2 \underline{Y}_2$ for short. Then

$$\begin{pmatrix} \underline{Z}_1 \\ \underline{Z}_2 \end{pmatrix} = \begin{pmatrix} S_1 & S_2 \\ S_2 & -S_2'^{-1} S_1' S_2 \end{pmatrix} \begin{pmatrix} \underline{Y}_1 \\ \underline{Y}_2 \end{pmatrix}$$

by Pfister's Theorem 5. We see that $Z_1, Z_2, \ldots, Z_8$ are bilinear in the X_i and the Y_j as claimed. The process can be repeated for $32, 64, \ldots,$ variables; the 8 bilinear terms will persist.

But of course even for $n = 16$, Theorem 7 is stronger than the above method as it gives us nine fully bilinear terms.

This problem was posed by Baeza and solved by Shapiro in a letter to Baeza in 1976.

Pfister has other very interesting results about Hilbert's 17th problem in the function fields $\mathbf{R}(X, Y)$ and more generally in $\mathbf{R}(X_1, X_2, \ldots, X_n)$. We refer our readers again to [13], [9].

References

[1] J.F. Adams, Vector fields on Sphere, *Annals of Maths* **75** (1962), 603–632.

[2] A.A. Albert (editor), Studies in Modern Algebra, vol. 2, MAA Studies in Maths (1963).

[3] E. Artin, Uber die Zerlegung definiter Funktionen in Quadrate, *Hamb. Abh.* **5** (1927), 100–115.

[4] J.W.S. Cassels, On the representation of rational functions as sums of squares, *Acta Arith.* **9** (1964), 79–82.

[5] L.E. Dickson, On quaternions and their generalizations and the history of the 8-square theorem, *Annals of Maths.* **20** (1919), 155–171.

[6] Adolf Hurwitz, Uber der Komposition der Quadratischen Formen von beliebig vielen Variabeln, Nachrichten von der Königlichen Gessellschaft der Wissenschaften in Göttingen (1898), 309–316; = Math. werke, II 565–571.

[7] A. Pfister, Zur Darstellung von-1 also Summe von Quadraten in einem Körper, JLMS, **40** (1965), 159–165.

[8] A. Pfister, Zur Darstellung definiter Funktionen als Summe Von Quadraten, *Inventiones Math.* **4** (1967), 224–236.

[9] A. Pfister, Hilbert's 17th problem and related problem on definite forms, *Proc. of Symposia in Pure Maths.* **28** (1976), 483–489.

[10] A. Pfister, Multiplicative quadratische Formen, *Arch. Math.* **16** (1965), 363–370.

[11] A. Pfister, Quadratrische Fromen in beliebigen Körpern, *Inventiones Maths.* **1** (1966), 116–132.

[12] A.R. Rajwade, A note on the Stufe of quadratic fields. *Indian J. Pure and App. Maths.* **6** (1975), 725–6.
[13] A.R. Rajwade, Squares, *London Math Soc.* Lecture note series no. 171, (1993).
[14] J. Radon, Lineare Scharen orthogonaler Matrizen, *Abh. Math. Sem Univ. Hamburg* **1** (1922), 1–14.
[15] E. Witt, Theorie der quadratischen Formen in beliebigen Körpern, *J. reine angew. Math.* 17 **6** (1937), 31–34.
[16] Paul Yiu, On the product of 2 sums of 16 squares as a sum of squares on integral bilinear forms, *Quart. J. Maths.* (2) **41** (1990), 463–500.

Centre for Advanced Study in Mathematics
Panjab University
Sector 14
Chandigarh 160 014