PACIFIC JOURNAL OF MATHEMATICS
Vol. 118, No. 2, 1985

SUMS OF SQUARES OF MATRICES

MORRIS NEWMAN

In memoriam Ernst G. Straus

The principal results of this paper are the following: Every integral 2×2 matrix is the sum of at most 3 integral squares, and this is best possible. Every integral $n \times n$ matrix with $n > 2$ is the sum of at most k integral squares, where $k = 7$ if n is even, and $k = 9$ if n is odd. Every $n \times n$ matrix over $GF(2)$ is the sum of at most 2 matrix squares, and this is best possible.

1. Introduction. The purpose of this paper is to show that every integral $n \times n$ matrix is the sum of a fixed number of integral squares $k = k(n)$, where

$$k = \begin{cases} 3, & n = 2, \\ 7, & n \text{ even and } > 2, \\ 9, & n \text{ odd and } > 1. \end{cases}$$

The result for $n = 2$ is best possible. The proof depends on the fact that over $GF(2)$, any matrix is the sum of at most 2 matrix squares, and this is best possible.

The possibility of representing a matrix as a sum of squares was already considered by L. Carlitz in [**1**], R. Ciampi in [**2**], D. Gondard and P. Ribenboim in [**3**], and M. Griffin and M. Krusemeyer in [**4**]; and some overlap naturally occurs. In particular, the case $n = 2$ (Theorem 5 of this paper) already occurs in [**1**] and [**4**], and Theorem 3 of this paper appears in [**4**]. The proofs are quite different, however and the theorems leading up to the proof of Theorem 3 are new and of independent interest. Also, the proof of Theorem 5 is quite direct, and somewhat simpler than the previous proofs.

We conclude the paper by listing some open problems.

The following non-standard notation will be employed in the paper: The $n \times n$ companion matrix

$$C = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \\ a_1 & a_2 & a_3 & \cdots & a_n \end{bmatrix}$$

will be denoted by

$$C = CM(a_1, a_2, \ldots, a_n);$$

and the $n \times n$ matrix A whose last row is $(a_1, a_2, \ldots, a_n)$ and which is 0 elsewhere will be denoted by

$$A = RM(a_1, a_2, \ldots, a_n).$$

We believe that this is just the sort of problem that would have appealed to Ernst Straus. In fact, he had turned to classical matrix theory just before his death, and had written several highly interesting papers in collaboration with Moshe Goldberg. The combination of matrix theory and number theory required by this problem would undoubtedly have interested him.

2. Matrices over $GF(2)$. In this section the underlying field is $GF(2)$, unless otherwise stated. In the theorem that follows, A is an $n \times n$ matrix and $p(x)$ its minimal polynomial. $p(x)$ may be written as

$$p(x) = p_0(x^2) + xp_1(x^2) = p_0(x)^2 + xp_1(x)^2,$$

where $p_0(x^2)$ is the even part of $p(x)$ and $xp_1(x^2)$ is the odd part of $p(x)$. Then we have

THEOREM 1. *A polynomial* $f(x)$ *exists such that* $A = f(A)^2$, *if and only if* $(p_0(x), p_1(x)) = 1$.

Proof. Suppose first that such a polynomial exists. Then $A = f(A)^2 = f(A^2)$. Since $p(x)$ is the minimal polynomial of A, this implies that $f(x^2) - x \equiv 0 \bmod p(x)$, so that for some polynomial $q(x)$, $f(x^2) - x = p(x)q(x)$. Decompose $q(x)$ into its even and odd parts:

$$q(x) = q_0(x^2) + xq_1(x^2).$$

Then

$$f(x^2) = x + \left(p_0(x^2) + xp_1(x^2)\right)\left(q_0(x^2) + xq_1(x^2)\right),$$

which implies that

$$p_0(x^2)q_1(x^2) + p_1(x^2)q_0(x^2) = 1,$$
$$p_0(x^2)q_0(x^2) + x^2p_1(x^2)q_1(x^2) = f(x^2),$$

so that

$$p_0(x)q_1(x) + p_1(x)q_0(x) = 1, \tag{1}$$

$$p_0(x)q_0(x) + xp_1(x)q_1(x) = f(x). \tag{2}$$

Thus certainly $(p_0(x), p_1(x)) = 1$.

Now suppose that $(p_0(x), p_1(x)) = 1$. Then $q_0(x)$, $q_1(x)$ may be determined so that (1) holds. Having determined $q_0(x)$, $q_1(x)$, define $f(x)$ by (2). Then all the steps in the first part of the proof are reversible, and we may conclude that $f(x^2) - x \equiv 0 \bmod p(x)$. Since $p(x)$ is the minimal polynomial of A, this implies that $f(A^2) - A = 0$, so that

$$A = f(A^2) = f(A)^2.$$

This concludes the proof.

An easy consequence of Theorem 1 is

THEOREM 2. *Suppose that A is non-derogatory, with characteristic polynomial $p(x) = p_0(x^2) + xp_1(x^2)$. Then a matrix B exists such that $A = B^2$ if and only if $(p_0(x), p_1(x)) = 1$.*

Proof. Since A is non-derogatory, its characteristic polynomial and minimal polynomial coincide. Hence if $(p_0(x), p_1(x)) = 1$, Theorem 1 implies the existence of the matrix B. Conversely, if the matrix B exists, then A and B must commute, and so B must be a polynomial in A, since A is non-derogatory. But then Theorem 1 implies that $(p_0(x), p_1(x)) = 1$. This concludes the proof.

Since any companion matrix is non-derogatory, an immediate corollary is

COROLLARY 1. *Suppose that A is a companion matrix, with characteristic polynomial $p(x) = p_0(x^2) + xp_1(x^2)$. Then A is a square if and only if $(p_0(x), p_1(x)) = 1$.*

These results were derived in order to prove the following theorem, which is the main result of this section.

THEOREM 3. *Every $n \times n$ matrix over $GF(2)$ is the sum of 2 squares of $n \times n$ matrices over $GF(2)$, and this is in general best possible.*

Proof. The main tool necessary to prove this result is the theorem of the Frobenius canonical form (also known as the rational canonical form), which states that every matrix over a field F is similar to a direct sum of companion matrices over F (see [5] or [6] for a convenient reference). Thus it is only necessary to prove the theorem for a companion matrix, since the direct sum of matrices which are sums of 2 squares is itself a sum of 2 squares, and a matrix similar to a sum of 2 squares is itself a sum of 2 squares. Accordingly, let C be an $n \times n$ companion matrix over $GF(2)$. If $n = 1$ then $C = [0]$ or $[1]$, both of which are squares. If $n = 2$ then C has the form

$$C = \begin{bmatrix} 0 & 1 \\ a & b \end{bmatrix}.$$

If $b = 0$, then

$$C = \begin{bmatrix} 0 & 1 \\ a & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ a & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ a & 1 \end{bmatrix}^2 + \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}^2.$$

If $b = 1$ and $a = 0$, then

$$C = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}^2.$$

If $b = 1$ and $a = 1$, then

$$C = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^2.$$

Hence we may assume that $n > 2$, and that

$$C = CM(a_1, a_2, \ldots, a_n).$$

We write $C = D + E$, where

$$D = RM(a_1 + k_1, a_2 + k_2, \ldots, a_{n-1} + k_{n-1}, 1)$$

and

$$E = CM(k_1, k_2, \ldots, k_{n-1}, 1 + a_n).$$

Now $D = D^2$, so that in order to complete the proof, it is only necessary to show that $k_1, k_2, \ldots, k_{n-1}$ may always be determined so that each of the matrices

$$E_0 = CM(k_1, k_2, \ldots, k_{n-1}, 0),$$
$$E_1 = CM(k_1, k_2, \ldots, k_{n-1}, 1)$$

is a square. Now the characteristic polynomials of E_0 and E_1 are respectively

$$f(x) = x^n + 0 \cdot x^{n-1} + k_{n-1} \cdot x^{n-2} + \cdots + k_1,$$
$$g(x) = x^n + 1 \cdot x^{n-1} + k_{n-1} \cdot x^{n-2} + \cdots + k_1.$$

First suppose that n is even, so that $n = 2m$, where $m > 1$. Make the choice

$$f(x) = x^{2m} + x + 1,$$
$$g(x) = x^{2m} + x^{2m-1} + 1.$$

Then

$$f_0(x) = 1 + x^m, \qquad f_1(x) = 1,$$
$$g_0(x) = 1 + x^m, \qquad g_1(x) = x^{m-1}.$$

Thus $(f_0(x), f_1(x)) = 1$, $(g_0(x), g_1(x)) = 1$, and Corollary 1 implies that both E_0 and E_1 are squares. Next suppose that n is odd, so that $n = 2m + 1$, where $m \geq 1$. Make the choice

$$f(x) = x^{2m+1} + x + 1,$$
$$g(x) = x^{2m+1} + x^{2m} + 1.$$

Then

$$f_0(x) = 1, \qquad f_1(x) = 1 + x^m,$$
$$g_0(x) = 1 + x^m, \qquad g_1(x) = x^m,$$

so that once again

$$(f_0(x), f_1(x)) = 1, \qquad (g_0(x), g_1(x)) = 1,$$

and Corollary 1 implies that both E_0 and E_1 are squares.

Thus it has been shown that in all cases C is the sum of 2 squares; and if $n > 1$, this result is best possible, since Corollary 1 implies that companion matrices exist which are certainly not squares. This concludes the proof.

Theorem 3 will be used in the next section in the following form, which we state as a corollary:

COROLLARY 2. *Let A be an integral $n \times n$ matrix. Then integral $n \times n$ matrices B, C exist such that $A \equiv B^2 + C^2 \bmod 2$.*

3. Matrices over the integers. In this section it will be shown that every integral $n \times n$ matrix with $n > 2$ is the sum of a fixed number s of integral squares. The proof gives $s = 9$, but this is not best possible. In the next section it will be shown that $s = 3$ is best possible for the case of integral 2×2 matrices. The fact that s does not depend on n is the significant fact here. It is interesting to note that this global result is an immediate consequence of the local results modulo 2 proved in the previous section.

The theorem to be proved is

THEOREM 4. *Suppose that $n > 1$. Then if n is even, every integral $n \times n$ matrix is the sum of at most 7 integral squares, and if n is odd, every integral $n \times n$ matrix is the sum of at most 9 integral squares.*

Proof. Suppose first that n is even. Write $n = 2m$, $m \geq 1$. Let M be any integral $n \times n$ matrix, and write

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

where A, B, C, D are all $m \times m$ matrices. Set

$$M_0 = \begin{bmatrix} A - I & B \\ C & D \end{bmatrix}.$$

We now note the following identities:

$$M_1 = \begin{bmatrix} 0 & B \\ C & I \end{bmatrix}^2 = \begin{bmatrix} BC & B \\ C & CB + I \end{bmatrix}, \qquad \begin{bmatrix} 0 & I \\ U & 0 \end{bmatrix}^2 = \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix},$$

where U is any integral $m \times m$ matrix. Then

$$M_0 - M_1 = \begin{bmatrix} A - BC - I & 0 \\ 0 & D - CB - I \end{bmatrix} = \begin{bmatrix} A_1 & 0 \\ 0 & D_1 \end{bmatrix}.$$

Set

$$M_2 = \begin{bmatrix} 0 & I \\ D_1 & 0 \end{bmatrix}^2 = \begin{bmatrix} D_1 & 0 \\ 0 & D_1 \end{bmatrix}.$$

Then

$$M_0 - M_1 - M_2 = \begin{bmatrix} A_1 - D_1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} A_2 & 0 \\ 0 & 0 \end{bmatrix}.$$

By Corollary 2, integral matrices A_3, A_4, A_5 exist such that $A_2 = A_3^2 + A_4^2 + 2A_5$. Set

$$M_3 = \begin{bmatrix} A_4 & 0 \\ 0 & 0 \end{bmatrix}^2, \quad M_4 = \begin{bmatrix} A_4 & 0 \\ 0 & 0 \end{bmatrix}^2.$$

Then

$$M_0 - M_1 - M_2 - M_3 - M_4 = \begin{bmatrix} 2A_5 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} (A_5 + I)^2 - A_5^2 - I & 0 \\ 0 & 0 \end{bmatrix}.$$

Thus

$$M_0 = S_5 + \begin{bmatrix} -A_5^2 - I & 0 \\ 0 & 0 \end{bmatrix}, \qquad M = S_5 + \begin{bmatrix} -A_5^2 & 0 \\ 0 & 0 \end{bmatrix},$$

where S_5 is the sum of 5 integral squares. But

$$\begin{bmatrix} -A_5^2 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -A_5^2 & 0 \\ 0 & -A_5^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & A_5^2 \end{bmatrix},$$

a sum of 2 integral squares. It follows that M is the sum of 7 integral squares.

Suppose now that n is odd, so that $n \geq 3$. Let M be any integral $n \times n$ matrix. Write M as

$$M = \begin{bmatrix} a & x \\ y & N \end{bmatrix},$$

where a is an integer and N is an integral $(n-1) \times (n-1)$ matrix. Put

$$M_1 = \begin{bmatrix} 0 & x \\ y & I \end{bmatrix}^2 = \begin{bmatrix} xy & x \\ y & yx + I \end{bmatrix}.$$

Then

$$M - M_1 = \begin{bmatrix} a - xy & 0 \\ 0 & N - yx - I \end{bmatrix} = \begin{bmatrix} a_1 & 0 \\ 0 & N_1 \end{bmatrix}.$$

Now put

$$M_2 = \begin{bmatrix} 0 & 1 \\ a_1 & 0 \end{bmatrix}^2 + 0_{n-2} = \begin{bmatrix} a_1 & 0 \\ 0 & a_1 \end{bmatrix} + 0_{n-2},$$

where 0_{n-2} is the $(n-2) \times (n-2)$ zero matrix. Then

$$M - M_1 - M_2 = \begin{bmatrix} 0 & 0 \\ 0 & N_3 \end{bmatrix},$$

where N_3 is an integral $(n-1) \times (n-1)$ matrix. By the previous result, N_3 is the sum of at most 7 integral squares. Hence M is the sum of at most 9 integral squares, and the proof of the theorem is concluded.

4. Matrices of order 2.

In this section the integral 2 × 2 matrices are examined in more detail. The first theorem to be proved is

THEOREM 5. *Every integral* 2 × 2 *matrix is the sum of at most* 3 *integral squares, and this is best possible.*

Proof. Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

be any integral 2×2 matrix. Put

$$B = \begin{bmatrix} x & b \\ c & 1 - x \end{bmatrix},$$

where x is an integer to be determined later. Then

$$B^2 = B - (x - x^2 - bc)I,$$

so that

$$A - B^2 = A - B + (x - x^2 - bc)I,$$

$$A - B^2 = \begin{bmatrix} a - bc - x^2 & 0 \\ 0 & d - bc - (x - 1)^2 \end{bmatrix} = \begin{bmatrix} a' & 0 \\ 0 & d' \end{bmatrix}.$$

Then

$$d' - a' = d - a - 1 + 2x.$$

Now if $d - a - 1$ is odd, choose $x = 0$; and if $d - a - 1$ is even, choose x so that $d - a - 1 + 2x = 0$. In the first case,

$$A - B^2 - \begin{bmatrix} 0 & 1 \\ y & 0 \end{bmatrix}^2 = \begin{bmatrix} a' - y & 0 \\ 0 & d' - y \end{bmatrix};$$

setting

$$r = \tfrac{1}{2}(a' + d' + 1), \quad s = \tfrac{1}{2}(a' + d' - 1), \quad y = a' - r^2 = d' - s^2,$$

r, s, y are integers and

$$A - B^2 - \begin{bmatrix} 0 & 1 \\ y & 0 \end{bmatrix}^2 = \begin{bmatrix} r^2 & 0 \\ 0 & s^2 \end{bmatrix},$$

so that A is the sum of 3 integral squares. In the second case,

$$A - B^2 = \begin{bmatrix} a' & 0 \\ 0 & a' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ a' & 0 \end{bmatrix}^2,$$

so that A is the sum of 2 integral squares. Hence it has been shown that A is always the sum of at most 3 integral squares.

To complete the proof, it is necessary to exhibit a matrix which is not the sum of 2 integral squares. Such a matrix, for example, is $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 3 \end{smallmatrix}\right]$. The proof is as follows: Suppose the contrary, and let A, B be integral 2×2 matrices for which

$$A^2 + B^2 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}. \tag{3}$$

Put $t = \operatorname{tr}(A)$, $d = \det(A)$, $\tau = \operatorname{tr}(B)$, $\delta = \det(B)$, so that $A^2 = tA - dI$, $B^2 = \tau B - \delta I$. Then

$$(4) \qquad tA + \tau B = \begin{bmatrix} 1 + d + \delta & 0 \\ 0 & 3 + d + \delta \end{bmatrix},$$

$$(5) \qquad t^2 + \tau^2 = 4 + 2(d + \delta).$$

It follows from (5) that t and τ are either both even or both odd. If t and τ are both odd, then (5) implies that $d + \delta$ is odd. But then (4) implies that $A \equiv B \bmod 2$, which contradicts (3). If t and τ are both even, then (5) implies that $d + \delta$ is even. But this contradicts (4), and the proof is concluded.

If the representing matrices are restricted to be unimodular, then a universal result no longer holds. For example, the following theorem is true:

THEOREM 6. *Let* $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ *be an integral* 2×2 *matrix. Then in order for the equation*

$$A = \sum_{i=1}^{t} A_i^2, \qquad A_i \in \mathrm{SL}(2, Z)$$

to have a solution, it is necessary and sufficient that

$$a + d \equiv b \equiv c \bmod 2.$$

The proof is a straightforward calculation modulo 2 and will be omitted. Thus for example the matrices $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$ cannot be written as sums of squares of matrices of SL(2, Z). A related result in this direction may be found in [7].

5. Some open questions. The following questions appear worthy of consideration:

(a) If it were true that over GF(2) the number of distinct $n \times n$ matrix squares exceeded half the total number of $n \times n$ matrices (which of course is 2^{n^2}) then there would be a very simple proof of Theorem 3. Unfortunately this is not always so. However, it does bring up the interesting question of determining this number.

(b) Determine the least positive integer $k = k(n)$ such that every integral $n \times n$ matrix with $n > 1$ is the sum of at most k integral squares. The results of this paper show that $k(2) = 3$, $k(2n) \leq 7$, $k(2n + 1) \leq 9$, $k(2n + 1) \leq 2 + k(2n)$. It is proved in [4] that $k(3) = 3$. Quite possibly $k(n) = 3$ for all $n > 1$.

(c) Generalize the problem to the case of the ring of integers of an algebraic number field.

(d) Generalize the problem to the case of arbitrary powers.

(e) Investigate the structure of the sets

$$S_{k,n} = \left\{ \sum_{i=1}^{k} A_i^2, A_i \in \mathrm{SL}(n, Z) \right\}.$$

References

[1] L. Carlitz, *Solution to problem* 140 (proposed by I. Connell), Canad. Math. Bull., **11** (1968), 615–619.

[2] R. Ciampi, *Characterization of a class of matrices as sums of squares*, Linear Algebra Appl., **3** (1970), 45†0.

[3] D. Gondard and P. Ribenboim, *Le* 17^e *probème de Hilbert pour les matrices*, Report of the Algebra Group, Queen's Papers in Pure and Applied Math., **36** (1973), 325–335.

[4] M. Griffin and M. Krusemeyer, *Matrices as sums of squares*, Linear and Multilinear Algebra, **5** (1977), 33–44.

[5] C. C. MacDuffee, *The Theory of Matrices*, Chelsea, New York (1946).

[6] M. Newman, *Integral Matrices*, Academic Press, New York (1972).

[7] ______, *Symmetric completions and products of symmetric matrices*, Trans. Amer. Math. Soc., **186** (1973), 191–201.

[8] O. Taussky, *History of sums of squares in algebra*, American Mathematics Heritage, Algebra and Applied Mathematics, **13** (1981), 73–90.

Received February 13, 1984. The preparation of this paper supported by a grant from the National Science Foundation.

University of California
Santa Barbara, CA 93106