PACIFIC JOURNAL OF MATHEMATICS
Vol. 32, No. 1, 1970

MATRIC POLYNOMIALS WHICH ARE HIGHER COMMUTATORS

EDMOND D. DIXON

Let A be an $n \times n$ matrix defined over a field F of characteristic greater than n. For each $n \times n$ matrix X we define

$$X_1 = [A, X]_0 = X \tag{1}$$
$$X_{h+1} = [A, X]_h = [A, X_h] = AX_h - X_hA$$

for each positive integer h. Then X is defined to be k-commutative with A if and only if

$$[A, X]_k = 0 \ , \qquad [A, X]_{k-1} \neq 0 \ . \tag{2}$$

Let $P(x)$ be a polynomial such that $P(A) \neq 0$. Specifically, assume that

$$P(A) = \sum_{i=p}^{n-1} \lambda_i A^i \neq 0 \tag{3}$$

where p is a positive integer, each λ_i is a scalar from F, and $\lambda_p \neq 0$. In this paper we study, for each positive integer k, the matrices X such that

$$[A, X]_k = P(A) \ . \tag{4}$$

We specify a polynomial $P(A)$ in the form (3) and show how the maximal value of k for which (4) has a solution depends on the polynomial $P(A)$. In Theorem 3 it is assumed that A is nonderogatory. Since the only matrices which commute with A in this case are polynomials in A, we are, in effect, establishing a more precise bound for k in (2) by predetermining X_k.

In the derogatory case, a matrix which is not a polynomial in A may commute with A. However, Theorem 4 shows that if we choose a polynomial $P(A)$ as X_k, then the maximal value of k depends on the polynomial P.

The problem of determining the maximal value of k for which (2) has a solution has been studied by Roth [8] and others. Roth's results are stated in terms of the maximal degrees of the elementary divisors of the matrix A. In particular, he showed that there exists a matrix X satisfying (2) for some A if $k \leqq 2n - 1$.

Nilpotent case. Throughout the paper we assume that A is in Jordan canonical form, since $[a, X]_k = P(A)$ if and only if

$$[BAB^{-1}, BXB^{-1}]_k = BP(A)B^{-1} \ .$$

The following notation introduced by W. V. Parker is used to simplify the proofs of the theorems.

DEFINITION. Let M_s for any integer s such that $-n+1 \leqq s \leqq m-1$ be the set of all $n \times m$ matrices in which all elements are zero except those for which $j-i=s$ (i denotes the row and j denotes the column in which the element appears). If $s > m-1$, M_s is defined to be the set consisting of only the zero matrix. A particular member of M_s will be denoted by D_s and will be called an s-stripe matrix. Note that if X is any $n \times m$ matrix then X can be written uniquely as $X = \sum_{s=-n+1}^{m-1} D_s$ where D_s is an element of M_s.

If A_1 and A_2 are $n \times n$ and $m \times m$ nilpotent nonderogatory matrices in Jordan canonical form and if $D_s = (d_{ij})$ is an $n \times m$ element of M_s where s is any integer such that $-n+1 \leqq s \leqq m-1$, let $f(D_s) = A_1D_s - D_sA_2$ and $f^k(D_s) = A_1f^{k-1}(D_s) - f^{k-1}(D_s)A_2$. It is easily seen that $f^k(D_s)$ is an element of M_{s+k}. Notice that the element in the ij position of $f(D_s)$, where $j-i=s+1$, is $d_{i+1,j}-d_{i,j-1}$ for $i \neq 1$. The element in the nj position is $-d_{n,j-1}$ if $j \neq 1$; the element in the $i1$ position is $d_{i+1,1}$ if $i \neq n$; and the element in the $n1$ position is zero.

LEMMA 1. *If A is an $n \times n$ nilpotent nonderogatory matrix in Jordan canonical form, if X is an $n \times n$ matrix, and if*

$$M = [A, X] = AX - XA \ ,$$

then the trace of M is zero and the trace of every subdiagonal stripe of M is zero.

Proof. Any $n \times n$ matrix X may be written as $\sum_{s=-n+1}^{n-1} D_s$ where D_s is an element of M_s. Thus

$$[A, X] = \left[A, \sum_{s=-n+1}^{n-1} D_s\right] = \sum_{s=-n+1}^{n-1} [A, D_s] \ .$$

If $s < 0$, then $[A, D_s]$ is a matrix such that the sum of the nonzero elements is zero. The matrix $[A, D_s]$ forms the $(s+1)$-stripe of M. This completes the proof of the lemma.

If A is an $n \times n$ nilpotent nonderogatory matrix in Jordan canonical form then for any positive integer $s < n$, $(A^T)^sA^s$ plays the part of a "lower identity" which we denote by L_s. That is,

$$(A^T)^sA^s = \begin{pmatrix} 0 & 0 \\ 0 & I_{n-s} \end{pmatrix} = L_s \ . \tag{5}$$

Similarly,

$$A^s(A^T)^s = \begin{pmatrix} I_{n-s} & 0 \\ 0 & 0 \end{pmatrix} = U_s \tag{6}$$

which we call an "upper identity".

Using the above, we prove the following lemma.

LEMMA 2. *Let A be an $n \times n$ nilpotent nonderogatory matrix in Jordan canonical form. Let L_s and U_s be as defined above. Then*

$$L_s(I - A)L_{s+k} = (I - A)L_{s+k} \tag{7}$$

and

$$U_{s+k}(I - A)U_s = U_{s+k}(I - A) , \tag{8}$$

where k is any positive integer less than $n - s$.

Proof. If we partition $I - A$ as follows:

$$(I - A) = \begin{pmatrix} M & 0 \\ * & N \end{pmatrix}$$

where M is $s \times (s + k)$, then

$$L_s(I - A)L_{s+k} = \begin{pmatrix} 0 & 0 \\ * & N \end{pmatrix} L_{s+k} = \begin{pmatrix} 0 & 0 \\ 0 & N \end{pmatrix} = (I - A)L_{s+k} .$$

The proof of (8) is similar.

Let $V = (1, 1, \cdots, 1)$, a $1 \times n$ vector, and let $V_s = VD_s$. That is, V_s is the vector in which each element represents a column sum in D_s, and since the columns in D_s have at most one nonzero element, V_s simply displays these elements in the form of a row vector. To simplify the notation we will let $V_{s+k} = VD_{s+k}$ where $D_{s+k} = [A, D_s]_k$ for some matrix D_s. In other words, the added subscript, k, implies that V_{s+k} is the result of k commutations. From now on, s will denote a nonnegative integer, $0 \leqq s \leqq n - 1$, and subdiagonal stripes of X will be denoted by D_{-s}. Also, the nontrivial subvector in V_s will be denoted by w_{n-s}, and the nontrivial subvector in V_s will be denoted by $\hat{w}_{n-s}$. Thus

$$V_s = (0, 0, \cdots, 0, d_{1,s+1}, d_{2,s+2}, \cdots, d_{n-s,n}) = (0_s, w_{n-s}) . \tag{9}$$

Similarly,

$$V_{-s} = (d_{s+1,1}, d_{s+2,2}, \cdots, d_{n,n-s}, 0, \cdots, 0) = (\hat{w}_{n-s}, 0_s) . \tag{10}$$

The following lemma is a vital part of the proof of Theorem 1.

LEMMA 3. *If k is a positive integer and if V_s, A, U_s, and L_s are as defined above, then*
(i) $V_{s+k} = V_s(I - A)^k L_k$,
(ii) $V_{-s+k} = V_{-s} U_s (I - A)^k$ *if* $k \leqq s$,
(iii) $V_{-s+k} = V_{-s} U_s (I - A)^k L_{k-s}$ *if* $k > s$.

Proof. Case (i). If $k = 1$, from (7) and (9)

$$V_s(I - A)L_{s+1} = (0_s, w_{n-s})\begin{pmatrix} 0 & 0 \\ 0 & N \end{pmatrix} .$$

In this case N has dimensions $(n - s)\times(n - s - 1)$, so N has (-1)'s on the diagonal and 1's on the first subdiagonal. But

$$(0_s, w_{n-s})\begin{pmatrix} 0 & 0 \\ 0 & N \end{pmatrix} = (0_s, w_{n-s})N = (0_{s+1}, w_{n-s-1})$$

where w_{n-s-1} has only $n-s-1$ elements of the form $(d_{i+1,s+i+1} - d_{i,s+i})$, and this is V_{s+1}. Therefore

$$V_{s+1} = V_s(I - A)L_{s+1} .$$

Similarly,

$$V_{s+2} = V_{s+1}(I - A)L_{s+2} = V_s(I - A)L_{s+1}(I - A)L_{s+2} .$$

But by Lemma 2,

$$L_{s+1}(I - A)L_{s+2} = (I - A)L_{s+2} .$$

Thus $V_{s+2} = V_s(I - A)^2 L_{s+2}$, and by induction it follows that

$$V_{s+k} = V_s(I - A)^k L_{s+k} . \tag{11}$$

In particular,

$$V_{0+k} = V_0(I - A)^k L_k . \tag{12}$$

Case (ii). From (10),

$$V_{-s}U_s(I - A) = V_{-s}\begin{pmatrix} I_{n-s} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} M & 0 \\ * & N \end{pmatrix} = (\hat{w}_{n-s}, 0_s)\begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix}$$

where M has dimensions $(n - s) \times (n - s + 1)$ and so has 1's on the diagonal and (-1)'s on the first superdiagonal. But

$$(\hat{w}_{n-s+1}, 0_s)\begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix} = (\hat{w}_{n-s+1}, 0_{s-1})$$

where $\hat{w}_{n-s+1}$ has $n - s + 1$ elements

$$d_{s+i+1,i+1} - d_{s+i,i}, \ (i = 0, 1, \cdots, n - s + 1) ,$$

and $d_{s,0} = d_{n+1,n-s+1} = 0$. This is $V[A, D_{-s}] = V_{-s+1}$. Similarly,

$$V_{-s+2} = V_{-s+1} U_{s-1}(I - A) = V_{-s} U_s(I - A) U_{s-1}(I - A) .$$

But by Lemma 2, $U_s(I - A)U_{s-1} = U_s(I - A)$. Thus

$$V_{-s+2} = V_{-s} U_s(I - A)^2 ,$$

and by induction it follows that if $k \leqq s$,

$$V_{-s+k} = V_{-s} U_s(I - A)^k . \tag{13}$$

In particular,

$$V_{-s+s} = V_{-s} U_s(I - A)^s . \tag{14}$$

Case (iii). When $k > s$, we divide the problem into two parts. Using case (i) we have

$$V_{-s+k} = V_{-s+s}(I - A)^{k-s} L_{k-s} . \tag{15}$$

But by case (ii), $V_{-s+s} = V_{-s} U_s(I - A)^s$. Thus

$$\begin{aligned} V_{-s+k} &= V_{-s} U_s(I - A)^s (I - A)^{k-s} L_{k-s} \\ &= V_{-s} U_s(I - A)^k L_{k-s} . \end{aligned}$$

This completes the proof of the lemma.

Using the above lemmas we prove Theorem 1, which establishes a precise upper bound for k in the case where A is nilpotent and $[A, X]_k = P(A) \neq 0$.

THEOREM 1. *Let A be an $n \times n$ nilpotent nonderogatory matrix. Let p be a positive integer such that $p < n$. Let*

$$\lambda_i (i = p, p + 1, \cdots, n - 1)$$

be scalars from F such that $\lambda_p \neq 0$. Then there exists a matrix X such that

$$[A, X]_k = \sum_{i=p}^{n-1} \lambda_i A^i \neq 0 \tag{16}$$

if and only if $k \leqq 2p$.

Proof. We first prove the case where $\lambda_i = 0$ for all $i > p$. We may assume without loss of generality that $\lambda_p = 1$ since $[A, X]_k = A^p$ if and only if $[A, \lambda_p X]_k = \lambda_p A^p$.

If there exists a matrix X satisfying (16) where A is nilpotent, then $[A, X]_k = [A, \sum_{s=-n+1}^{n-1} D_s]_k = A^p$. Thus we must have

$$[A, D_{s-k}]_k = \begin{cases} 0 & \text{if } s \neq p \\ A^p & \text{if } s = p \end{cases} . \tag{17}$$

Therefore, for $s = p$,

$$\begin{aligned} V[A, D_{p-k}]_k &= V_{(p-k)+k} = VD_p = VA^p \\ &= (0, 0, \cdots, 0, 1, 1, \cdots, 1) , \end{aligned}$$

which we will call $(0_p, E_{n-p})$. If $k \leqq p$, from (11),

$$V_{(p-k)+k} = V_{p-k}(I - A)^k L_p .$$

Using an argument similar to that used in proving lemma 2, we find that $(I - A)^k L_p$ can be written as $\begin{pmatrix} 0 & 0 \\ 0 & N_k \end{pmatrix}$ where N_k has dimensions $(n - p + k) \times (n - p)$. Since this matrix has a square submatrix of order $n - p$ with 1's on the diagonal, zeros below, it has rank $n - p$.

Now rewriting (12) as

$$(0_p, E_{n-p}) = (0_{p-k}, w_{n-p+k}) \begin{pmatrix} 0 & 0 \\ 0 & N_k \end{pmatrix}$$

we see that solving this equation is equivalent to solving $E_{n-p} = (w_{n-p+k}) N_k$. The augmented matrix for this equation is $\begin{pmatrix} N_k \\ E_{n-p} \end{pmatrix}$, and since N_k has rank $n - p$, the augmented matrix also has rank $n - p$. Thus the system has a solution with $(n - p + k) - (n - p) = k$ parameters.

Now if $k > p$ we refer to equation (15) and set

$$V_{(p-k)+k} = V_{p-k} U_{k-p} (I - A)^k L_p . \tag{18}$$

But the product on the right may be written as $\begin{pmatrix} 0 & H_k \\ 0 & 0 \end{pmatrix}$.

If $k = 2p$ then H_k is square of order $n - p$. Since it has minus signs in a checkerboard pattern, we may transform it into a matrix with nonnegative elements or nonpositive elements (depending on whether p is even or odd) by multiplying on the left and right by the matrix $D = \text{diag.}\ (-1, 1, -1, \cdots, (-1)^{n-p})$. Thus the determinant of H_k will be unchanged and the resulting matrix has determinant

$$(-1)^p \prod_{i=0}^{n-p-1} \frac{\binom{2p+i}{p}}{\binom{p+i}{p}} \neq 0$$

(see Muir, Vol. 3, p. 451). Hence H_k is nonsingular. Furthermore,

$(-1)^p H_k$ is positive definite since the principal subdeterminants are all positive by the same argument.

Thus if $k = 2p$ we may rewrite the equation (18) as

$$(0_p, E_{n-p}) = (\hat{w}_{n-p}, 0_p)\begin{pmatrix} 0 & H_k \\ 0 & 0 \end{pmatrix} .$$

But solving this system is equivalent to solving

$$E_{n-p} = \hat{w}_{n-p} H_k \ , \tag{19}$$

and since H_k is nonsingular, this system has a unique solution. A solution for $k = 2p$ implies the existence of matrices X satisfying $[A, X]_k = A^p$ for all $k < 2p$.

Next we show that there is no solution for $k = 2p + 1$, and thus for any $k > 2p$, by the following argument. Since H_k is nonsingular, equation (19) is equivalent to $E_{n-p} H_k^{-1} = \hat{w}_{n-p}$. Multiplying both sides of this equation by the $(n - p) \times 1$ column vector E_{n-p}^T gives

$$E_{n-p} H_k^{-1} E_{n-p}^T = \hat{w}_{n-p} E_{n-p}^T = \sum_{i=1}^{n-p} d_{p+i,i} \ . \tag{20}$$

This is the sum of the nonzero elements in D_{-p}. By Lemma 1, if $[A, X] = D_{-p}$, then $\sum_{i=1}^{n-p} d_{p+i,i} = 0$. But since $(-1)^p H_k$ is positive definite, $(-1)^p H_k^{-1}$ is also. Thus the product on the left in (20) is not zero and there does not exist a solution for $k > 2p$.

This completes the proof in the case where $[A, X]_k = \lambda A^p$. In the case where $[A, X]_k = \lambda_p A^p + \lambda_{p+1} A^{p+1} + \cdots + \lambda_{n-1} A^{n-1}$, we see that X may be written as $\sum_{i=p}^{n-1} X_{1i}$ where $[A, X_{1i}]_k = \lambda_i A^i$.

If A is derogatory then the Jordan canonical form for A is diag. $(A_1, A_2, \cdots, A_s)$ where $s > 1$. Theorem 1 can also be extended to the derogatory case. The method of proof is similar to that used in Theorem 1.

THEOREM 2. *Let A be an $n \times n$ nilpotent matrix. Let p be a positive integer such that $p < n_i$ where n_i is the dimension of the largest block in the Jordan canonical form for A. Let λ_i $(i = p, p + 1, \cdots, n - 1)$ be scalars from F such that $\lambda_p \neq 0$. Then there exists a matrix X such that*

$$[A, X]_k = \sum_{i=p}^{n_i - 1} \lambda_i A^i \neq 0 \tag{23}$$

if and only if $k \leqq 2p$.

Some remarks about the integer p are in order here. If the Jordan canonical form for A is diag. $(A_1, A_2, \cdots, A_s)$ we may assume without loss of generality that the dimension n_i of A_i is greater than or equal to the dimension n_{i+1} of A_{i+1} for $i = 1, 2, \cdots, s-1$. Since $A^p = \text{diag.}$ $(A_1^p, A_2^p, \cdots, A_s^p)$, p must be less than n_1 if A^p is to be different from zero. However, A_i^p may be zero for some $i > 1$.

Notice that since the Jordan canonical form for a nilpotent matrix is the same as the rational canonical form for that matrix, the constructions for the matrices X in Theorems 1 and 2 may be done with rational operations.

The general case. Here it is not assumed that A is nilpotent. We assume that A is in Jordan canonical form. Again we choose a polynomial $P(A)$ which we desire to write as a higher commutator of A. Theorems 3 and 4 establish the maximal value for k in equation (4).

THEOREM 3. *Let A be an $n \times n$ nonderogatory matrix in Jordan canonial form $\alpha I + N$ where N is the nilpotent matrix with 1's on the first superdiagonal and zeros elsewhere. Let $P(A)$ be a polynomial in A such that $P(A) \neq 0$. Let t be the multiplicity of α as a root of $P(x)$. Then there exists an $n \times n$ matrix X such that*

$$[A, X]_k = P(A) \tag{24}$$

if and only if $k \leqq 2t$.

Proof. If $A = (\alpha I + N)$ then

$$[A, X]_k = [(\alpha I + N), X]_k = [\alpha I, X]_k + [N, X]_k = [N, X]_k \,.$$

Thus condition (24) becomes $[N, X]_k = P(\alpha I + N) = \sum_{i=1}^{n-1} \lambda_i N^i$ where $\lambda_i = p^{(i)}(\alpha)/i!$. Now by Theorem 1, (24) has a solution if and only if $k \leqq 2t$.

THEOREM 4. *Let $A = \text{diag.}\ (A_1, A_2, \cdots, A_s)$ where $A_i = (\alpha_i I + N_i)$ $(i = 1, 2, \cdots, s)$ where each N_i is as in Theorem 3. Let P be a polynomial such that $P(A) \neq 0$. Let $A_{i_1}, A_{i_2}, \cdots, A_{i_t}$ be the blocks of A such that $P(A_{i_j}) \neq 0$. Let m_{i_j} be the multiplicity of $(x-\alpha_{i_j})$ in $P(x)$. Let $m = \text{min.}\ \{m_{i_j}\}$. Then there exists an $n \times n$ matrix X such that*

$$[A, X]_k = P(A) \tag{25}$$

if and only if $k \leqq 2m$.

Proof. If $A = \text{diag.}\ (A_1, A_2, \cdots, A_s)$ then

$$P(A) = \text{diag.}\ (P(A_1), P(A_2), \cdots, P(A_s)) \,.$$

If $P(A_t) = 0$ for some A_t, then there exists a matrix $X_t \neq 0$ such that

$[A_t, X_t]_k = P(A_t) = 0$ for any positive integer k. Thus we need only consider those A_i for which $P(A_i) \neq 0$. Assume that $P(A_i) \neq 0$ for all $i = 1, 2, \cdots, s$. Then if we let

$$X = \text{diag.}\ (X_1, X_2, \cdots, X_s)$$

where $[A_i, X_i]_k = P(A_i)$, the matrix X will satisfy (25). Assume without loss of generality that the degree of $(x - \alpha_1)$ in $P(x)$ is $m = \min. \{m_i\}$. Then $[A_1, X_1] = P(A_1)$ if and only if $k \leqq 2m$. Thus $[A, X]_k = P(A)$ if and only if $k \leqq 2m$.

Bibliography

1. M.A. Drazin, J.W. Dungey, and K.W. Gruenberg, *Some theorems on commutative matrices,* J. London Math. Soc. **26** (1951) 221-228.
2. N. Jacobson, *Lie Algebras,* Interscience Publishers, New York, 1962.
3. N. H. McCoy, *On quasi-commutative matrices,* Trans. Amer. Math. Soc. **36** (1934), 327-340.
4. N. T. Muir, *Theory of determinants,* Dover, New York, 1920.
5. W. V. Parker, *Matrices and polynomials,* Math. Monthly **61** (1954), 182-183.
6. ———, *The matrix equation $AX = XB$,* Duke Math. J. **17** (1950), 43-51.
7. W. V. Parker and J. C. Eaves, *Matrices,* Ronald Press, New York, 1960.
8. W.E. Roth, *On K-commutative matrices,* Trans. Amer. Math. Soc. **39** (1936), 483-495.
9. O. Taussky, and H. Wielandt, *Linear relations between higher additive commutators,* Proc. Amer. Math. Soc. **13** (1962), 732-735.

Received December 6, 1968.

TENNESSEE TECHNOLOGICAL UNIVERSITY