ON MATRIX COMMUTATORS

MARVIN MARCUS AND NISAR A. KHAN

1. Introduction. Let A, B, and X be n-square matrices over an algebraically closed field F of characteristic 0. Let $[A, B] = AB - BA$ and set $(A, B) = [A, [A, B]]$. Recently several proofs **(1; 3; 5)** of the following result have appeared: *if* $\det (AB) \neq 0$ *and* $(A, B) = 0$ *then* $A^{-1}B^{-1}AB - I$ *is nilpotent.* In **(2)** McCoy determined the general form of any X satisfying

$$(1.1) \qquad (A, X) = 0$$

in the case that A has a single elementary divisor corresponding to each eigenvalue, that is, A is *non-derogatory.* In Theorem 1 we determine the structure of any matrix X satisfying (1.1) and also give a formula for the dimension of the linear space of all such X in terms of the degrees of the elementary divisors of A. Moreover, we apply our results to obtain a condition that B be a polynomial in A. It is a classical result **(6**, p. 150**)** that *if* $[X, B] = 0$ *whenever* $[A, X] = 0$ *then* B *is a scalar polynomial in* A. We prove in Theorem 2 that *if* $(X, B) = 0$ *whenever* $(A, X) = 0$ *then* B *is a scalar polynomial in* A.

We also obtain the result that *the dimension of the linear space* $K(A)$ *of all such matrices* B *is precisely the number of distinct eigenvalues of* A.

2. Solutions of $(A,X) = 0$. Let A have the distinct eigenvalues $\lambda_1, \ldots, \lambda_q$ and let

$$(x - \lambda_i)^{e_{ij}},$$

$j = 1, \ldots, n_i$, $i = 1, \ldots, q$ be the elementary divisors of A with the notation chosen as follows:

For each $i = 1, \ldots, q$

$$e_{i1} > e_{i2} > \ldots > e_{in_i}$$

and

$$(x - \lambda_i)^{e_{ij}}$$

appears with multiplicity r_{ij}, $j = 1, \ldots, n_i$.

THEOREM 1. *The number of linearly independent solutions of*

$$(1.1) \qquad (A, X) = 0$$

is

Received February 9, 1959; in revised form July 22, 1959. The work of the first author was supported by U.S. National Science Foundation Grant N.S.F. G-5416. The second author is a Post-doctorate Fellow of the National Research Council of Canada.

https://doi.org/10.4153/CJM-1960-022-2 Published online by Cambridge University Press

$$(2.1) \qquad \sum_{i=1}^{q} \left\{ \sum_{j=1}^{n_i} (2e_{ij} - 1)r_{ij}^2 + 4 \sum_{j<k}^{n_i} r_{ij}r_{ik}e_{ik} \right\}.$$

It is clear that we may assume A to be in Jordan canonical form J. Set

$$p_i = \sum_{j=1}^{n_i} r_{ij}e_{ij},$$

the algebraic multiplicity of λ_i.

We write

$$(2.2) \qquad A = \sum_{i=1}^{q}{}^{\cdot} (\lambda_i I_{p_i} + V_i)$$

where I_{p_i} is a p_i-square identity matrix, V_i is a p_i-square matrix with only 1 and 0 in the superdiagonal, all other elements 0, and $\sum^{\cdot}$ indicates direct sum.

We also write

$$(2.3) \qquad V_i = \sum_{j=1}^{n_i}{}^{\cdot} U_{ij},\ J_i = \lambda_i I_{p_i} + \sum_{j=1}^{n_i}{}^{\cdot} U_{ij}$$

and hence

$$(2.4) \qquad A = \sum_{i=1}^{q}{}^{\cdot} \sum_{j=1}^{n_i}{}^{\cdot} (\lambda_i I e_{ij} + U_{ij}).$$

where U_{ij} is the direct sum of the e_{ij}-square auxiliary unit matrix repeated r_{ij} times, $j = 1, \ldots, n_i$. We partition X conformally with the partitioning of A indicated in (2.4). Now consider a block of X, call it C, that corresponds to λ_i and λ_j for $i \neq j$. A result obtained in both **(2)** and **(4)** is

LEMMA 1. $C = 0$.

From Lemma 1 we conclude that

$$X = \sum_{i=1}^{q}{}^{\cdot} X_i$$

and X_i is a p_i-square matrix. To determine the structure of X_i we may obviously confine our attention to the case in which A has a single eigenvalue with several elementary divisors.

LEMMA 2. *Let A be an n-square matrix with the single eigenvalue λ and let $(x - \lambda)^{\nu_i}$ be an elementary divisor of A of multiplicity r_i, $i = 1, \ldots, t$, $\nu_1 > \ldots > \nu_t$, $\sum_{i=1}^{t} r_i \nu_i = n$. Then the most general matrix X satisfying* (1.1) *contains*

$$\sum_{i=1}^{t} (2\nu_i - 1)r_i^2 + 4 \sum_{i<j}^{t} r_i r_j \nu_j$$

arbitrary parameters.

https://doi.org/10.4153/CJM-1960-022-2 Published online by Cambridge University Press

Proof. Since (1.1) holds for $A - \lambda I$ if and only if it holds for A we may assume $\lambda = 0$ without loss of generality. Thus we assume

$$A = \sum_{i=1}^{t}{}^{\cdot} \sum_{j=1}^{r_i}{}^{\cdot} U_i \tag{2.5}$$

where U_i is a ν_i-square matrix with 1 along the superdiagonal and 0 elsewhere. If $\nu_i = 1$ then U_i is the 1-square 0 matrix. We partition X conformally with A in (2.5):

$$X = (X_{ij}),$$

and observe from (1.1) that

$$U_i^2 X_{ij} + X_{ij} U_j^2 - 2U_i X_{ij} U_j = 0. \tag{2.6}$$

For the sake of simplicity of notation, we take $U_i = U$ as m-square, $U_j = V$ as n-square, and $X_{ij} = C = (c_{ij})$ as an $m \times n$ matrix. There are three essentially distinct cases to consider:

$$\text{(i) } m = n, \qquad \text{(ii) } m > n, \qquad \text{(iii) } m < n.$$

The case (i), $m = n$, is done in **(2)** and **(4)** and in this case $c_{ij} = 0$ $i > j$, and the elements of each diagonal parallel (or equal) to the main diagonal are in arithmetic progression. Hence the number of arbitrary parameters in C in case (i) is $2(n-1) + 1 = 2n - 1$. The case $n \leqslant 2$ is not considered in **(2)** but it is trivial to see that the number of parameters there is also $2n - 1$.

Case (ii): $m > n$. We have from (2.6)

$$U^2C + CV^2 - 2UCV = 0. \tag{2.7}$$

We assume in what follows that $m \geqslant 3$. The case $m = 2$, $n = 1$ will be disposed of later. Let ϵ_i be the unit column n-vector with 1 in position i, $i = 1, \ldots, n$. We evaluate the transform of ϵ_i by the left side of (2.7) to obtain

$$U^2 c_i + c_{i-2} = 2Uc_{i-1}, \tag{2.8}$$

where c_i denotes the ith column of C, $i = 1, \ldots, n$. Co-ordinatewise (2.8) becomes for $i = 1, \ldots, n$

$$(2.9)_i \quad [c_{3i}, c_{4i}, \ldots, c_{mi}, 0, 0] + [c_{1,i-2}, c_{2,i-2}, \ldots, c_{m-2,i-2}, c_{m-1,i-2}, c_{m,i-2}]$$
$$= 2[c_{2,i-1}, c_{3,i-1}, \ldots, c_{m-1,i-1}, c_{m,i-1}, 0].$$

We show first that $c_{ij} = 0$ for $i > j$, $i \geqslant 3$. We compute the $(i+s)$ co-ordinate of $(2.9)_i$ where s is one of the integers $0, \ldots, m - i$;

$$c_{i+s+2,i} + c_{i+s,i-2} - 2c_{i+s+1,i-1} = 0, \quad i = 1, \ldots, m - (s+2). \tag{2.10}$$

We first note that from (2.8) for $i = 1$ we have

$$U^2 c_i = 0$$

and hence

$$c_{31} = c_{41} = \ldots = c_{m1} = 0. \tag{2.11}$$

https://doi.org/10.4153/CJM-1960-022-2 Published online by Cambridge University Press

Consider the system (2.10) for $i = 2, 3, \ldots, m - (s+2)$ in succession and obtain

$$\begin{aligned} c_{s+4,2} &= 2c_{s+3,1} = 0 \qquad \text{(by 2.11)} \\ c_{s+5,3} &= -c_{s+3,1} + 2c_{s+4,2} = 0 \\ &\cdots \\ c_{m,m-(s+2)} &= -c_{m-2,m-s-4} + 2c_{m-1,m-s-3} = 0. \end{aligned}$$

Thus $c_{ij} = 0$ for $i > j$, $i \geqslant 3$. We now consider in succession the $(i-1)$ co-ordinate of $(2.9)_i$ for $i = 2, \ldots, n$ (since $n + 1 \leqslant m$), to obtain

$$c_{i+1,i} + c_{i-1,i-2} - 2c_{i,i-1} = 0.$$

Setting i successively equal to $2, \ldots, n$ we have

$$c_{32} = 2c_{21},\ c_{43} = 3c_{21}, \ldots, c_{n+1,n} = nc_{21}.$$

Hence there is only one arbitrary parameter c_{21} in this diagonal of C. We next consider the elements c_{ij} for $i \leqslant j$. We compute the rth co-ordinate of $(2.9)_{r+s}$, $r = 1, \ldots, n - s$, where s is one of the integers $2, \ldots, n-1$, to obtain

$$\begin{aligned} c_{3,s+1} + c_{1,s-1} &= 2c_{2,s} \\ c_{4,s+2} + c_{2,s} &= 2c_{2,s+1} \\ &\cdots \\ c_{n-s+2,n} + c_{n-s,n-2} &= 2c_{n-s+1,n-1}. \end{aligned}$$

Hence

$$c_{1,s-1}, \quad c_{2,s}, \quad c_{3,s+1}, \ldots, c_{n-s+2,n}$$

are in arithmetic progression. Thus in each diagonal

$$d_{s-1} = c_{1,s-1}, \ldots, c_{n-s+2,n} \qquad s = 2, \ldots, n-1$$

there are two arbitrary parameters. In the diagonals d_{n-1} and d_n we accumulate three more arbitrary parameters in C, $c_{1,n-1}$, c_{1n}, c_{2n}. Hence the total number of arbitrary parameters in C for $m > n$ is

$$3 + 2(n-2) + 1 = 2n.$$

We compute easily that for $m = 2$, $n = 1$, C involves $2n = 2$ arbitrary parameters as well.

Case (iii): $n > m$. We reduce this to case (ii) as follows:

Let P_k denote the k-square permutation matrix with 1 in each of the positions $(k - j, j + 1)$, $j = 0, \ldots, k - 1$. Taking the transpose of (2.7) we have

$$C'(U')^2 + (V')^2C' - 2V'C'U' = 0. \tag{2.12}$$

Now observe that

$$\begin{aligned} U' &= P_m U P_m, \\ V' &= P_n V P_n, \end{aligned}$$

and substituting in (2.12) using $P_m^2 = I_m$, $P_n^2 = I_n$ we have

$$C'P_mU^2P_m + P_nV^2P_nC' - 2P_nVP_nC'P_mUP_m = 0.$$

Now pre-multiplying by P_n and post-multiplying by P_m we have

$$(P_nC'P_m)U^2 + V^2(P_nC'P_m) - 2V(P_nC'P_m)U = 0.$$

Now $P_nC'P_m$ is $n \times m$, the situation of case (ii). Hence $P_nC'P_m$ has $2m$ arbitrary parameters and C has $2m$ arbitrary parameters.

Returning to the statement of Lemma 2, we conclude from case (i) that any block in the partitioning of X corresponding to equal U_i's contains $2\nu_i - 1$ arbitrary parameters and there are r_i^2 such blocks for each i. Also from (ii) and (iii) any block in X corresponding to U_i and U_j, $i < j$, contains $2\nu_j$ arbitrary parameters (since for $i < j$, $\nu_i > \nu_j$). Hence the total number of arbitrary parameters in X is

$$\sum_{i=1}^{t} (2\nu_i - 1)r_i^2 + 4\sum_{i<j}^{t} r_i r_j \nu_j.$$

We return now to the proof of Theorem 1. By Lemma 1 we need only add the total number of parameters of the q main diagonal blocks of X corresponding to each λ_i. By Lemma 2, this number for a fixed λ_i is

$$\sum_{j=1}^{ni} (2e_{ij} - 1)r_{ij}^2 + 4\sum_{j<k}^{ni} r_{ij}r_{ik}e_{ik}.$$

Summing this for $i = 1, \ldots, q$ we obtain the formula (2.1) and the proof is complete.

3. The space $K(A)$. Let P be a non-singular matrix satisfying

$$P^{-1}AP = J \tag{3.1}$$

and let $Y = P^{-1}XP$ and $C = P^{-1}BP$. Then $(P^{-1}AP, P^{-1}BP) = P^{-1}(A, B)P$ implies that $B \in K(A)$ if and only if $C \in K(J)$. As indicated earlier, if $(J, Y) = 0$, then

$$Y = \sum_{s=1}^{q}{}^{\cdot}\, Y_s \tag{3.2}$$

and Y_s is p_s-square, $s = 1, \ldots, q$. If

$$Y_s = (Y_{ij}), i, j = 1, \ldots, m_s, m_s = \sum_{j=1}^{ns} r_{sj} \tag{3.3}$$

indicates a partitioning of Y_s conformally with the partitioning of J_s in (2.3) then we have seen in the proof of Theorem 1 that a block in (3.3) is an $e_{si} \times e_{sj}$ rectangular matrix with the following structure:

1. $e_{si} > e_{sj}$.

https://doi.org/10.4153/CJM-1960-022-2 Published online by Cambridge University Press

(i) each diagonal, except the element in the upper right corner, parallel or equal to the diagonal starting from the upper left corner involves two arbitrary parameters, and the elements in each of these diagonals are in arithmetical progression;

(ii) there is one non-zero diagonal immediately below the diagonal starting from the upper left corner, containing one parameter only. The elements are of the form $a, 2a, 3a, \ldots, e_{sj}a$ for an arbitrary $a \in F$;

(iii) all other elements are zero.

2. $e_{si} < e_{sj}$.

(i) the diagonal d ending in the lower right corner and those above it each involve two arbitrary parameters and the elements are in arithmetical progression, with the exception of the upper right corner element which is arbitrary;

(ii) the diagonal immediately below d contains one parameter and the elements are of the form $e_{si}a, \ldots, 3a, 2a, a$ for an arbitrary $a \in F$;

(iii) all other elements are zero.

3. $e_{si} = e_{sj}$.

The block is upper triangular. Each diagonal involves two arbitrary parameters and the elements are in arithmetical progression with the exception of the upper right corner element which is arbitrary.

Let $L(J)$ be the linear space of all Y satisfying $(J, Y) = 0$.

LEMMA 3. *$(Y, C) = 0$ for each $Y \in L(J)$ if and only if*

$$C = \sum_{j=1}^{q}{}^{\cdot}\, c_j I_{p_j} \tag{3.4}$$

where $c_j \in F$, $j = 1, 2, \ldots, q$.

Proof. The sufficiency of (3.4) is clear. By the above description of $L(J)$,

$$\sum_{j=1}^{q}{}^{\cdot}\, x_j I_{p_j} \in L(J)$$

for any $x_j \in F$.

Hence $(Y, C) = 0$ implies

$$C = \sum_{s=1}^{q}{}^{\cdot}\, C_s,$$

C_s is p_s-square. Now $(Y, C) = 0$ implies that $(Y_s, C_s) = 0$, $s = 1, \ldots, q$. We may choose $Y \in L(J)$ with

$$Y_s = \sum_{j=1}^{n_s}{}^{\cdot} \sum_{\alpha=1}^{r_{sj}}{}^{\cdot}\, x_{j\alpha} I_{e_{sj}}$$

for arbitrary $x_{j\alpha} \in F$ and conclude that

$$C_s = \sum_{j=1}^{n_s}{}^{\cdot}\, C_{sj},$$

https://doi.org/10.4153/CJM-1960-022-2 Published online by Cambridge University Press

where C_{sj} is a direct sum of r_{sj} e_{sj}-square matrices $j = 1, \ldots, n_s$. Let M_{sj} be any one of the e_{sj}-square blocks whose direct sum comprises C_{sj}.

We next show that C_s is a scalar multiple of the identity by noting first that $(Y_s, C_s) = 0$ implies that $(D_{sj}, M_{sj}) = 0$ where D_{sj} is an e_{sj}-square diagonal matrix with diagonal elements (in arithmetical progression) along the main diagonal. Hence M_{sj} is diagonal. Let

$$M_{sj} = \text{diag}\,(\alpha_1, \alpha_2, \ldots, \alpha_{e_{sj}}).$$

Now we may choose Y_s such that the equation $(E_{sj}, M_{sj}) = 0$ holds, where E_{sj} is e_{sj}-square and

$$E_{sj} = \begin{bmatrix} 0 & x & 0 & \ldots & 0 \\ 0 & 0 & x+y & \ldots & 0 \\ . & . & . & & . \\ 0 & 0 & 0 & \ldots & x + (e_{sj} - 2)y \\ 0 & 0 & 0 & \ldots & 0 \end{bmatrix}.$$

Now $(E_{sj}, M_{sj}) = 0$ is equivalent (for the case $e_{sj} \geqslant 3$, the case $e_{sj} \leqslant 2$ is trivial) to

$$E_{sj}^2 M_{sj} + M_{sj} E_{sj}^2 = 2E_{sj} M_{sj} E_{sj}.$$

Elementwise we have

$$(x + (t-3)y)\,(x + (t-2)y)\,(\alpha_p + \alpha_{p-2} - 2\alpha_{p-1}) = 0$$

for $t = 3, \ldots, e_{sj}$ and for arbitrary x, y. Hence we conclude that

$$\alpha_1, \alpha_2, \ldots, \alpha_{e_{sj}}$$

are in arithmetical progression and thus we may write

$$M_{sj} = \text{diag}\,(\alpha, \alpha + \beta, \ldots, \alpha + (e_{sj} - 1)\beta). \tag{3.5}$$

We next show that $\beta = 0$. To this end we choose Y_s such that $(Y_s, C_s) = 0$ implies the following:

$$(F_{sj}, M_{sj}) = 0, \tag{3.6}$$

where

$$F_{sj} = \begin{bmatrix} 1 & 1 & \ldots & 0 \\ 0 & 2 & \ldots & 0 \\ . & . & & . \\ 0 & 0 & \ldots & 1 \\ 0 & 0 & \ldots & e_{sj} \end{bmatrix}.$$

From (3.5) we obtain by computing the (1, 2) element of (3.6) that $\beta = 0$. Hence

$$C_s = \sum_{j=1}^{n_s}{}^{\cdot}\, C_{sj}.$$

where

$$C_{sj} = \sum_{\alpha=1}^{r_{sj}}{}^{\cdot}\, x_{j\alpha} I_{e_{sj}},$$

We next show that all $x_{j\alpha}$ are equal. Let

$$y_1 \equiv x_{11}, y_2 = x_{12}, \ldots, y_{r_{s1}} = x_{1r_{s1}}, y_{r_{s1}+1} = x_{21}, \ldots, y_{m_s} = x_{n_s r_{sn_s}}.$$

Then from $(Y_s, C_s) = 0$, it follows that

$$\sum_{j=1}^{m_s} y_v Y_{uj} Y_{jv} + \sum_{j=1}^{m_s} y_u Y_{uj} Y_{jv} = 2 \sum_{j=1}^{m_s} y_j Y_{uj} Y_{jv}, \quad u, v = 1, 2, \ldots, m_s.$$

We have, by computing the block in the upper left corner (that is, the one conformal with Y_{11}),

$$y_1 \sum_{j=1}^{m_s} Y_{1j} Y_{j1} = \sum_{j=1}^{m_s} y_j Y_{1j} Y_{j1}.$$

$$\text{(3.7)} \qquad \sum_{j=2}^{m_s} (y_1 - y_j) Y_{1j} Y_{j1} = 0.$$

For a fixed t, $1 \leqslant t \leqslant m_s$, choose the $e_{s1} \times e_{st}$ matrix (recalling that $e_{s1} \geqslant e_{st}$)

$$Y_{1t} = \sum_{j=1}^{e_{st}} G_{jj},$$

where G_{jj} is an $e_{s1} \times e_{st}$ matrix with 1 in the (j, j) position and zeros elsewhere. Also choose

$$Y_{t1} = \sum_{j=1}^{e_{st}} H_{jj},$$

where H_{jj} is $e_{st} \times e_{s1}$ with 1 in the (j, j) position and zeros elsewhere, and let $Y_{1j} = 0$ for $j \neq t$. Then (3.7) becomes

$$(y_1 - y_t) I_{e_{s1}} = 0.$$

Hence $y_1 = y_t$, $t = 2, 3, \ldots, m_s$ and C has the form indicated in (3.4).

THEOREM 2. *If* $(X, B) = 0$ *for any* X *satisfying* $(A, X) = 0$, *then* B *is a scalar polynominal in* A.

Moreover the dimension of the linear space $K(A)$ *of all such* B *is given by*

$$\dim K(A) = q,$$

where q *is the number of distinct eigenvalues of* A.

Proof. We have seen that $B \in K(A)$ if and only if $C = P^{-1}BP \in K(J)$ where $P^{-1}AP = J$ is the Jordan canonical form of A.

Moreover, by Lemma 3

$$C = \sum_{j=1}^{q}{}^{\cdot}\, c_j I_{p_j},$$

where $c_j \in F$ are arbitrary. This implies immediately that

$$\dim K(A) = q.$$

Now

$$B = PCP^{-1} = P\Big(\sum_{j=1}^{q}{}^{\cdot}\, c_j I_{p_j}\Big)P^{-1}.$$

Let 0_p be the p-square matrix of zeros and let

$$E_j = P(0_{u_j} \dotplus I_{p_j} \dotplus 0_{v_j})P^{-1},$$

where

$$u_j = \sum_{i=1}^{j-1} p_i,\ v_j = n - \sum_{i=1}^{j} p_i,\ j = 1, 2, \ldots, q.$$

Then the E_j are the principal idempotents of A corresponding to the λ_j respectively and each E_j is a scalar polynomial $f_j(A)$ in A (**7**, p. 29).

Hence

$$B = \sum_{j=1}^{q} c_j E_j = \sum_{j=1}^{q} c_j f_j(A)$$
$$= f(A)$$

where

$$f(x) = \sum_{j=1}^{q} c_j f_j(x).$$

References

1. I. N. Herstein, *On a theorem of Putnam and Wintner*, Proc. Amer. Math. Soc., *9* (1958), 363–364.
2. N. H. McCoy, *On quasi-commutative matrices*, Trans. Amer. Math. Soc., *36* (1934), 327–340.
3. C. R. Putnam and A. Wintner, *On the spectra of group commutators*, Proc. Amer. Math. Soc., *9* (1958), 360–362.
4. W. E. Roth, *On k-commutative matrices*, Trans. Amer. Math. Soc., *39* (1936), 483–495.
5. O. Taussky, *A note on the group commutator of A and A**, J. Wash. Acad. Sci., *48* (1958), 305.
6. H. W. Turnbull and A. C. Aitken, *An introduction to the theory of canonical matrices* (London, 1932).
7. J. H. M. Wedderburn, *Lectures on matrices*, Amer. Math. Soc. Colloq. Publ. *17* (New York, 1934).

The University of British Columbia
and
Muslim University, Aligarh

https://doi.org/10.4153/CJM-1960-022-2 Published online by Cambridge University Press