ON MATRIX COMMUTATORS OF HIGHER ORDER

D. W. ROBINSON

Introduction. Let F_n be the collection of n-by-n matrices over a field F. For Y in F_n let Δ_Y be the mapping on F_n given by $X\Delta_Y = XY - YX$. In this paper we study the following

PROPOSITION. *Let A and B be in F_n and let m be a positive integer. If $B\Delta_X{}^m = 0$ whenever $X\Delta_A{}^m = 0$, then B is a polynomial in A with coefficients in F.*

The case $m = 1$ is the classical result that if B commutes with every matrix that commutes with A, then B is a scalar polynomial in A (cf. **10**, p. 106; **5**, p. 48; or **2**, p. 536). The case $m = 2$ has been investigated by M. Marcus and N. A. Khan (**4**) when F is algebraically closed and of characteristic zero, and the proposition itself has been established by M. F. Smiley (**8**) provided F is algebraically closed and of characteristic zero or prime $p \geqslant n$.

Recently, O. Taussky (**9**) has asked if this proposition is valid in case F is not algebraically closed. In this paper we provide an affirmative answer to this question. Indeed, with F an arbitrary field of elements, the proposition is an immediate corollary of the classical theorem and the two theorems given below.

1. Preliminary considerations. In this section we give the basic facts that are needed for the proofs of the theorems.

First, we call $X \in F_n$ *semi-simple* if the minimum polynomial of X is relatively prime to its derivative, and prove (**8**, p. 353 and **6**, p. 776)

LEMMA. *If F is a field and $X \in F_n$ is semi-simple, then $B\Delta_X{}^m = 0$ for some positive integer m implies $B\Delta_X = 0$.*

Proof. Let $M(x)$ be the minimum polynomial of X. If $B\Delta_X{}^2 = 0$, then we have identically (cf. **7**, p. 487)

$$0 = B\Delta_{M(X)} = M'(X)(B\Delta_X).$$

If, moreover, X is semi-simple, then $M'(X)$ is non-singular and $B\Delta_X = 0$. That is, X semi-simple and $B\Delta_X{}^2 = 0$ implies $B\Delta_X = 0$. The lemma now follows by induction on m.

Second, let A and B be in F_n. We say that B has property P in F relative to A if $X\Delta_A{}^2 = 0$ and X semi-simple implies $B\Delta_X = 0$. Also, for m a positive integer, we say that B has property P_m in F relative to A if $X\Delta_A{}^m = 0$ implies

Received February 17, 1964.

$B\Delta_X{}^m = 0$. We observe that the proposition above may now be restated to read that B has property P_m in F relative to A only if B is a polynomial in A with coefficients in F. Furthermore, we note from the lemma that property P_m for $m > 1$ implies property P.

Third, we shall have occasion to use the well-known fact that if $A \in F_n$ is cyclic and commutes with $B \in F_n$, then B is a polynomial in A with coefficients in F (**5**, p. 45).

Finally, we note that if F is of characteristic prime p, for A and B in F_n and f a positive integer,

$$B\Delta_A{}^{p^f} = \sum_{k=0}^{p^f} (-1)^k \binom{p^f}{k} A^k B A^{p^f-k} = BA^{p^f} - A^{p^f}B = B\Delta_{A^{p^f}}.$$

2. The separable case. In this section we consider the case in which the minimum polynomial $M(x)$ of $A \in F_n$ is separable (i.e., none of the irreducible factors of $M(x)$ has a zero derivative (**11**, p. 65), and prove

THEOREM 1. *Let F be a field and let $A \in F_n$. Assume that the minimum polynomial of A is separable and let $B \in F_n$ have property P in F relative to A. Then B is a polynomial in A with coefficients in F.*

Proof. We first show that it is sufficient to consider the case in which the minimum polynomial $M(x)$ of A has only one irreducible factor. Indeed, let $E_1, \ldots, E_k$ be the principal idempotents of A associated with the respective irreducible factors $\pi_1(x), \ldots, \pi_k(x)$ of $M(x)$ (**1**, pp. 130–132). Since A commutes with each E_i, by property P, B also commutes with each E_i. It follows that $E_i B$ has property P relative to $E_i A$ in the algebra of matrices of the form $E_i X$ where $E_i X = XE_i$ and $X \in F_n$. Thus, by assumption,

$$E_i B = Q_i(E_i A) = E_i Q_i(A) \quad \text{for some } Q_i(x) \in F[x] \ (i = 1, \ldots, k).$$

Consequently, since $E_i = R_i(A)$ for some $R_i(x) \in F[x]$ $(i = 1, \ldots, k)$,

$$B = E_1 Q_1(A) + \ldots + E_k Q_k(A)$$

is a polynomial in A.

Second, we show that it will suffice to consider the case in which A is cyclic. Indeed, let $\pi(x)$ of degree r be the sole irreducible factor of $M(x)$. By a similarity transformation we may assume that $A = \operatorname{diag}(A_1, \ldots, A_t)$, where each A_j is cyclic with minimum polynomial $(\pi(x))^{s_j}$, $s_1 \geqslant s_2 \geqslant \ldots \geqslant s_t$, and each A_j has the canonical form

$$A_j = \begin{bmatrix} C & U & 0 & \ldots & 0 & 0 \\ 0 & C & U & \ldots & 0 & 0 \\ & & & \ldots & & \\ 0 & 0 & 0 & \ldots & C & U \\ 0 & 0 & 0 & \ldots & 0 & C \end{bmatrix},$$

where C is the companion matrix of $\pi(x)$ and U is the r-by-r matrix with 1 in the $(r, 1)$-position and zero everywhere else. If $X_j = \operatorname{diag}(0, \ldots, I_{rs_j}, \ldots, 0)$ is conformal with A, then property P implies that B commutes with X_j $(j = 1, \ldots, t)$. Thus, $B = \operatorname{diag}(B_1, \ldots, B_t)$ and is conformal with A. Now for $j \neq 1$, let G_j be the block matrix conformal with A such that the $(1, 1)$ block is I_{rs_1}, the $(1, j)$ block is

$$\begin{bmatrix} I_{rs_j} \\ 0 \end{bmatrix},$$

and the remaining blocks are all zero. Since

$$A_1 = \begin{bmatrix} A_j & * \\ 0 & * \end{bmatrix},$$

it is clear that $G_j \Delta_A = 0$. Also, since G_j is idempotent, property P implies that $B\Delta_{G_j} = 0$. Thus,

$$B_1 = \begin{bmatrix} B_j & * \\ 0 & * \end{bmatrix}.$$

Next, it is evident that property P of B relative to A implies that B_1 has property P relative to A_1. Since A_1 is cyclic, we have by assumption

$$B_1 = Q(A_1) = \begin{bmatrix} Q(A_j) & * \\ 0 & * \end{bmatrix}$$

for some $Q(x) \in F[x]$. Thus, $B_j = Q(A_j)$ $(j = 1, \ldots, t)$ and $B = Q(A)$.

Finally, to complete the proof of the theorem, let A be cyclic with minimum polynomial $(\pi(x))^s$, where $\pi(x)$ of degree r is irreducible in $F[x]$. Since $\pi(x)$ is separable, by a similarity transformation we may assume that

$$A = \begin{bmatrix} C & I_r & 0 & \ldots & 0 & 0 \\ 0 & C & I_r & \ldots & 0 & 0 \\ & & & \ldots & & \\ 0 & 0 & 0 & \ldots & C & I_r \\ 0 & 0 & 0 & \ldots & 0 & C \end{bmatrix},$$

where again C is the companion matrix of $\pi(x)$ (**3**, p. 115). Also, as in (**8**), let $D = \operatorname{diag}(I_r, 2I_r, \ldots, sI_r)$ and $N = A - \operatorname{diag}(C, C, \ldots, C)$ and find by direct calculation that $D\Delta_A = -N$ and $N\Delta_A = 0$.

If the field F is of characteristic zero, then D and $D + N$ are semi-simple and property P implies that $B\Delta_D = B\Delta_{D+N} = 0$. If F is of characteristic prime p, then $D + N$ is not necessarily semi-simple. However, by introducing $G = DN$, we observe that $D + G$ and $D + G + N$ as well as D are semi-simple. This is because each has minimum polynomial $(x - 1), (x - 2) \ldots (x - s)$ when $p \geqslant s$ and $(x - 1)(x - 2) \ldots (x - p)$ when $p < s$. Thus again property P implies that $B\Delta_D = B\Delta_{D+N} = 0$.

Next, partition B into r-by-r blocks b_{ij} $(i, j = 1, \ldots, s)$. If F is of characteristic zero, then the results of the preceding paragraph easily require that

$B = \operatorname{diag}(b, \ldots, b)$, where $b = b_{11}$. Also, if F is of characteristic p, then it follows that $b_{ij} = b_{1,kp+1}$ if $j - i = kp$ $(k = 0, 1, 2, \ldots)$ and $b_{ij} = 0$ otherwise; that is, if $B_{kp} = \operatorname{diag}(b_{kp}, \ldots, b_{kp})$, where $b_{kp} = b_{1,kp+1}$, then

$$B = B_0 I + B_p N^p + B_{2p} N^{2p} + \ldots .$$

Now, since $\pi(x)$ is separable, $S = \operatorname{diag}(C, C, \ldots, C)$ is semi-simple. Moreover, it is clear that $S\Delta_A = 0$. Thus, by property P, $B\Delta_S = 0$; that is, $b\Delta_C = 0$ and $b_{kp}\,\Delta_C = 0$ in the respective cases above. Consequently, because C is cyclic, b and b_{kp} are polynomials in C. It follows that $B\Delta_A = 0$, and because A is cyclic, B is a polynomial in A. This completes the proof of Theorem 1.

3. The inseparable case. In this section we drop the separability condition, but we are required now to replace property P by the stronger property P_m.

THEOREM 2. *Let F be a field of characteristic prime p. Suppose that $B \in F_n$ has property P_m in F relative to $A \in F_n$ and let the non-negative integer f be determined by $p^{f-1} < m \leqslant p^f$. Then B is a polynomial in A^{p^f} with coefficients in F.*

Proof. As in the proof of Theorem 1, we note first that it is sufficient to consider again the case in which the minimum polynomial of A has only one irreducible factor. This follows from the proof above together with the fact that with F of characteristic p, the idempotent $E_i = R_i(A)$ is expressible as a polynomial in A^{p^f}. Indeed,

$$E_i = E_i^{p^f} = (R_i(A))^{p^f} = R_i^{(p^f)}(A^{p^f}),$$

where the coefficients of the polynomial $R_i^{(p^f)}(x)$ are simply the p^f powers of the respective coefficients of $R_i(x)$.

Furthermore, the second part of the proof of Theorem 1 may be applied directly to show that it is again sufficient to consider A to be cyclic. Thus, we now assume that A has minimum polynomial $(\pi(x))^s$, where $\pi(x)$ of degree r is irreducible over F and $n = rs$. Furthermore, we suppose $\pi(x)$ is inseparable of degree p^e (**11**, p. 67). Without loss of generality we take

$$A = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ & & & \ldots & \\ a_0 & a_1 & a_2 & \ldots & a_{n-1} \end{bmatrix}$$

in natural normal form, where

$$(\pi(x))^s = x^n - a_{n-1}x^{n-1} - \ldots - a_1 x - a_0.$$

Because $\pi(x)$ is inseparable of degree p^e, it is expressible as a polynomial in x^{p^e}; that is, $a_i = 0$ unless $p^e \mid i$. Therefore, for any integer f such that $1 \leqslant p^f \leqslant p^e$, it follows that

$$A^{p^f} = \begin{bmatrix} 0 & I_{p^f} & 0 & \cdots & 0 \\ 0 & 0 & I_{p^f} & \cdots & 0 \\ & & & \cdots & \\ a_0 I_{p^f} & a_{p^f} I_{p^f} & a_{2p^f} I_{p^f} & \cdots & a_{(n_f-1)p^f} I_{p^f} \end{bmatrix}$$

where $n = p^f n_f$. Therefore, if $F^{(p^f)}$ is defined to be the field of elements of the form aI_{p^f} with $a \in F$, then A^{p^f} may be considered as a matrix over $F^{(p^f)}$ $(1 \leqslant p^f \leqslant p^e)$.

Furthermore, as we now show by finite induction on f, if $p^{f-1} < p^e$ and $m > p^{f-1}$, then property P_m implies that B may also be considered as a matrix over $F^{(p^f)}$. Specifically, since it is obvious that B may be considered as a matrix over $F^{(p^0)}$, suppose that B is a matrix over $F^{(p^f)}$ for $1 \leqslant p^f < p^e$ and let $m > p^f$. Define

$$D = \operatorname{diag}(I_{p^f}, 2I_{p^f}, \ldots, n_f I_{p^f}).$$

Since $p \mid n_f$, and $a_i = 0$ unless $p^e \mid i$, it follows by direct calculation that

$$D\Delta_A{}^{p^f} = D\Delta_{A^{p^f}} = -A^{p^f}.$$

Therefore, since $m > p^f$, $D\Delta_A{}^m = 0$. Also, by choosing $G = DA^{p^f}$, we have $(D + G)\Delta_A{}^m = 0$. Because $p \mid n_f$ it is easily shown that both D and $D + G$ are semi-simple. Therefore, property P_m and the lemma above imply that B commutes with both D and $D + G$. This result, however, implies that B, which is considered as a matrix over $F^{(p^f)}$, is partitioned into p-by-p scalar blocks over $F^{(p^f)}$. That is, B may be considered as a matrix over $F^{(p^{f+1})}$.

We now complete the proof of the theorem in case $m \leqslant p^e$. Thus, suppose that $p^{f-1} < m \leqslant p^f \leqslant p^e$. By property P_m, $B\Delta_A{}^m = 0$. Therefore,

$$B\Delta_{A^{p^f}} = B\Delta_A{}^{p^f} = 0.$$

That is, A^{p^f} commutes with B. Consequently, since both B and A^{p^f} may be considered as matrices over $F^{(p^f)}$ and A^{p^f} is clearly cyclic over $F^{(p^f)}$, B is necessarily a polynomial in A^{p^f}.

Finally, we consider the case $p^e < m$. Let $\pi(x) = \pi_0(x^{p^e})$, $n = p^e n_0$, and $r = p^e r_0$. Also, let $\mathfrak{A} = A^{p^e}$ and $\mathfrak{F} = F^{(p^e)}$. Since $\mathfrak{A}$ is a cyclic matrix over $\mathfrak{F}$ of order $n_0 = r_0 s$, and $(\pi_0(x))^s$ is the minimum polynomial of $\mathfrak{A}$ with $\pi_0(x)$ separable, by a similarity transformation we may consider

$$\mathfrak{A} = \begin{bmatrix} \mathfrak{C} & \mathfrak{I}_{r_0} & 0 & \cdots & 0 & 0 \\ 0 & \mathfrak{C} & \mathfrak{I}_{r_0} & \cdots & 0 & 0 \\ & & & \cdots & & \\ 0 & 0 & 0 & \cdots & \mathfrak{C} & \mathfrak{I}_{r_0} \\ 0 & 0 & 0 & \cdots & 0 & \mathfrak{C} \end{bmatrix},$$

where $\mathfrak{C}$ is the companion matrix of $\pi_0(x)$ over $\mathfrak{F}$, $\mathfrak{I}_{r_0}$ is the r_0-by-r_0 identity matrix over $\mathfrak{F}$, and there are s blocks of $\mathfrak{C}$ on the main diagonal **(3**, p. 115**)**.

https://doi.org/10.4153/CJM-1965-052-9 Published online by Cambridge University Press

As a consequence of the proof of Theorem 1 above it follows that B is a polynomial in $\mathfrak{A}$; specifically,

$$B = \mathfrak{B}_0 \mathfrak{I} + \mathfrak{B}_p \mathfrak{N}^p + \mathfrak{B}_{2p} \mathfrak{N}^{2p} + \ldots,$$

where $\mathfrak{N} = \mathfrak{A} - \text{diag}(\mathfrak{C}, \mathfrak{C}, \ldots, \mathfrak{C})$ and $\mathfrak{B}_{kp} = \text{diag}(\mathfrak{b}_{kp}, \mathfrak{b}_{kp}, \ldots, \mathfrak{b}_{kp})$ with $\mathfrak{b}_{kp}$ some polynomial in $\mathfrak{C}$. We now show in fact that $m > p^{e+j-1}$ implies that

$$B = \mathfrak{B}_0 \mathfrak{I} + \mathfrak{B}_{pj} \mathfrak{N}^{pj} + \mathfrak{B}_{2pj} \mathfrak{N}^{2pj} + \ldots .$$

Indeed, since the case $j = 1$ is given above, we proceed by induction on j and assume that $m > p^{e+(j+1)-1}$. Let $s = p^j q + t$, $0 \leqslant t < p^j$, and choose

$$\mathfrak{D} = \text{diag}(\overbrace{\mathfrak{I}_{r0}, \ldots, \mathfrak{I}_{r0}}^{p^j}, \ldots, \overbrace{q\mathfrak{I}_{r0}, \ldots, q\mathfrak{I}_{r0}}^{p^j}, \overbrace{(q+1)\mathfrak{I}_{r0}, \ldots, (q+1)\mathfrak{I}_{r0}}^{t}).$$

By direct calculation,

$$\mathfrak{D}\Delta_{\mathfrak{A}^{pj}} = -\mathfrak{N}^{pj},$$

which is a polynomial in $\mathfrak{A}$. Thus, since $m > p^{e+j}$, it follows by property P_m that $B\Delta_{\mathfrak{D}}{}^m = 0$. Since $\mathfrak{D}$ is semi-simple, B commutes with $\mathfrak{D}$. Consequently, $\mathfrak{B}_i = 0$ unless $p^{j+1}|i$, which provides the desired conclusion.

Now, with $p^e \leqslant p^{e+j-1} < m \leqslant p^{e+j}$, we show that B is a polynomial in $A^{p^{e+j}} = \mathfrak{A}^{pj}$. Let $\mathfrak{S} = \text{diag}(\mathfrak{C}, \mathfrak{C}, \ldots, \mathfrak{C})$. Then $\mathfrak{A}^{pj} = \mathfrak{S}^{pj} + \mathfrak{N}^{pj}$ and $\mathfrak{A}^{ph} = \mathfrak{S}^{ph}$ for some sufficiently large $h \geqslant j$. Suppose X commutes with $\mathfrak{A}^{pj}$. Then X commutes with $\mathfrak{A}^{ph} = \mathfrak{S}^{ph}$. That is $X\Delta_{\mathfrak{S}}{}^{ph} = 0$. But, since the minimum polynomial $\pi_0(x)$ of $\mathfrak{S}$ is irreducible and separable, $\mathfrak{S}$ is semi-simple and $X\Delta_{\mathfrak{S}} = 0$. Therefore, X commutes with $\mathfrak{N}^{pj} = \mathfrak{A}^{pj} - \mathfrak{S}^{pj}$. Also, since $\mathfrak{B}_{kpj}$ is a polynomial in $\mathfrak{S}$, X commutes with $\mathfrak{B}_{kpj}$. Therefore, from the form of B obtained above, X commutes with B. That is, B commutes with every matrix that commutes with $A^{p^{e+j}}$. Consequently, by the classical theorem, B is a polynomial in $A^{p^{e+j}}$. This completes the proof of Theorem 2.

The author expresses his appreciation to the referee for his very helpful suggestions.

References

1. N. Jacobson, *Lectures in abstract algebra*, vol. II (Princeton, 1953).
2. P. Lagerstrom, *A proof of a theorem on commutative matrices*, Bull. Amer. Math. Soc., *51* (1945), 535–536.
3. A. I. Mal'cev, *Foundations of linear algebra* (San Francisco, 1963).
4. M. Marcus and N. A. Khan, *On matrix commutators*, Can. J. Math., *12* (1960), 269–277.
5. W. V. Parker, *The matrix equation AX = XB*, Duke Math. J., *17* (1950), 43–51.
6. D. W. Robinson, *A note on k-commutative matrices*, J. Mathematical Phys. *2* (1961), 776–777.
7. W. E. Roth, *On k-commutative matrices*, Trans. Amer. Math. Soc., *39* (1936), 483–495.
8. M. F. Smiley, *Matrix commutators*, Can. J. Math., *13* (1961), 353–355.
9. O. Taussky, *Matrix commutators of higher order*, Bull. Amer. Math. Soc., *69* (1963), 738.
10. J. H. M. Wedderburn, *Lectures on matrices*, Amer. Math. Soc. Colloq. Publ., vol. *17* (1934).
11. O. Zariski and P. Samuel, *Commutative algebra*, vol. I (Princeton, 1958).

Brigham Young University