Pacific Journal of Mathematics

ON A CLASS OF MATRIX EQUATIONS

ROBERT CHARLES THOMPSON

Vol. 20, No. 2 October 1967

PACIFIC JOURNAL OF MATHEMATICS
Vol. 20, No. 2, 1967

ON A CLASS OF MATRIX EQUATIONS

R. C. THOMPSON

Let K be a field of characteristic p, p zero or prime, and let L be the algebraic closure of K. Let $M_n(K)$ denote the matrix algebra of n-square matrices with elements in K. The commutator of A, $B \in M_n(K)$ is defined by $(A, B) = AB - BA$. It is the object of this paper to examine the following two questions.

I. Given exactly one of the three matrices $A, B, C \in M_n(K)$, to determine necessary and sufficient conditions in order that the other two matrices will exist in $M_n(K)$ such that

$$(1) \qquad C = (A, B),\ (C, A) = 0,\ C \neq 0\ .$$

II. Given exactly one of the three matrices $A, B, C \in M_n(K)$, to determine necessary and sufficient conditions in order that the other two matrices will exist in $M_n(K)$ such that

$$(2) \qquad C = (A, B),\ (C, A) = 0,\ (C, B) = 0,\ C \neq 0\ .$$

We shall obtain complete solutions to all these problems, except that, in Question I when C is the given matrix and $0 < p \leqq n$, we obtain only a partial solution. As a consequence of our results, we are able to find conditions that are sufficient, and sometimes necessary and sufficient, in order that solutions exist in $M_n(K)$ for certain complicated families of commutator equations related to (1) or (2).

Our proofs use only the theory of matrix similarity. Certain of our results (Theorems 1, 3, 4) are close to known results in the theory of Lie algebras. Nevertheless, we present full proofs.

Question II has already been examined by McCoy [4] when K is the complex number field. In our discussion of Question II, we obtain results that partially duplicate and partially complement McCoy's results.

2. Additional notation. Let $0_{\alpha\beta}$ denote an $\alpha \times \beta$ matrix of zeros, let I_α denote the α-square identity matrix. Frequently we simply write 0 and I. We let λ be an indeterminate and if $m(\lambda) = m_0 + m_1\lambda + \cdots + m_{r-1}\lambda^{r-1} + \lambda^r$, we let $C(m(\lambda))$ be the companion matrix of $m(\lambda)$. It is the matrix 2.2 on p. 252 of [5] when $r > 1$, and $C(m(\lambda)) = (-m_0)$ when $r = 1$. For the informed reader we take our companion matrices to have the stripe of one's on the diagonal just above the main diagonal. Let T_α be a linear combination of powers of $C(\lambda^\alpha)$. Then T_α is constant on each diagonal parallel to (or equal to) the main diagonal, and T_α has only zeros strictly below

the main diagonal. We use T_α to define $\alpha \times \beta$ *stripe matrix* M. For $\alpha < \beta$, $M = (0_{\alpha,\beta-\alpha}, T_\alpha)$; for $\alpha = \beta$, $M = T_\alpha$; for $\alpha > \beta$,

$$M = \begin{bmatrix} T_\beta \\ 0_{\alpha-\beta,\beta} \end{bmatrix}.$$

The direct sum of matrices A, B is denoted by $A + B$ or by diag (A, B).

Given matrix $A \in M_n(K)$, the nonconstant polynomials on the main diagonal of the Smith canonical form of the polynomial matrix $\lambda I - A$ are called the invariant factors of A. Each of the invariant factors of A can be factored into a product

$$p_1(\lambda)^{e_1} p_2(\lambda)^{e_2} \cdots$$

of powers of distinct polynomials $p_1(\lambda)$, $p_2(\lambda)$, $\cdots$, which are irreducible over K. We call these powers of irreducible polynomials

$$p_1(\lambda)^{e_1},\ p_2(\lambda)^{e_2},\ \cdots,$$

the elementary divisors of A over K. Within $M_n(K)$, A is similar to the direct sum of the companion matrices of its elementary divisors over K.

3. Question *I*.

THEOREM 1. *Let $p = 0$ or $p > n$. Let C, U_i, $V_i \in M_n(K)$ satisfy*

$$C = \sum_{i=1}^{s} (U_i, V_i),\ (C, U_i) = 0,\ 1 \leqq i \leqq s\ . \tag{3}$$

Then C is nilpotent. Conversely, if C is nilpotent, $A, B \in M_n(K)$ exist such that

$$C = (A, B),\ (C, A) = 0\ . \tag{4}$$

THEOREM 2. *Let $0 < p \leqq n$. Let $C \in M_n(L)$. Then $A, B \in M_n(L)$ exist satisfying* (4) *if and only if C is similar within $M_n(L)$ to some matrix $C_1 + \cdots + C_t$ in which*: (i) *for $1 \leqq i \leqq t$, C_i has a single eigenvalue γ_i*; (ii) *whenever $\gamma_i \neq 0$, C_i partitions as*

$$C_i = (C_{i\alpha\beta})_{1 \leqq \alpha,\beta \leqq k(i)}\ ,$$

such that block $C_{i\alpha\beta}$ is an $e_{i\alpha} \times e_{i\beta}$ stripe matrix, and $e_{i\alpha} \equiv e_{i\beta} \equiv 0$ (mod p), $1 \leqq \alpha, \beta \leqq k(i)$.

THEOREM 3. *Let $A, B, C \in M_n(L)$ satisfy* (4). *Then $A^+, B^+ \in M_n(L)$ exist such that $C = (A^+, B^+)$, $(C, A^+) = 0$, A^+ is nilpotent, and for*

$p \neq 2$, B^+ has zero trace.

THEOREM 4. *If $C \in M_n(K)$ is nilpotent then $B \in M_n(K)$ exists such that*

$$C = (C, B) . \tag{5}$$

Moreover, if $p \neq 2$, B may be chosen to have zero trace.

Proofs. We first establish part of Theorem 1. Suppose (3) holds. After a similarity transformation by an element of $M_n(L)$, we may suppose C is in Jordan canonical form: $C = C_1 + C_2 + \cdots + C_t$ where C_i has a single eigenvalue γ_i and $\gamma_i \neq \gamma_j$ for $i \neq j$. Partition $U_i = (U_{i\alpha\beta})_{1\leqq\alpha,\beta\leqq t}$, $V_i = (V_{i\alpha\beta})_{1\leqq\alpha,\beta\leqq t}$, conformally with the partitioning of C. Then $(C, U_i) = 0$ forces $C_\alpha U_{i\alpha\beta} = U_{i\alpha\beta} C_\beta$, $1 \leqq \alpha, \beta \leqq t$, $1 \leqq i \leqq s$. It is well known that, because C_α and C_β do not have a common eigenvalue when $\alpha \neq \beta$, this equation implies $U_{i\alpha\beta} = 0$ for $\alpha \neq \beta$. Then (3) immediately yields

$$C_\alpha = \sum_{i=1}^{s} (U_{i\alpha\alpha}, V_{i\alpha\alpha}) , \quad 1 \leqq \alpha \leqq s ,$$

which forces trace $C_\alpha = 0$, hence $\gamma_\alpha = 0$, for all α. Hence C is nilpotent. This proves half of Theorem 1. This half of Theorem 1 is well known in the theory of Lie Algebras. The proof usually given there uses Newton's identities on symmetric polynomials. Our proof avoids use of this device from outside linear algebra.

We next establish part of Theorem 3. After a similarity transformation by an element of $M_n(L)$, we may suppose A in Jordan canonical form, hence let $A = A_1 + \cdots + A_t$ where A_i has α_i as its only eigenvalue and $\alpha_i \neq \alpha_j$ if $i \neq j$. Then $(C, A) = 0$ forces $C = C_1 + \cdots + C_t$. Partition $B = (B_{ij})_{1\leqq i,j\leqq t}$. Then $C = (A, B)$ yields $C_i = (A_i, B_{ii})$ and $(C, A) = 0$ yields $(C_i, A_i) = 0$; $1 \leqq i \leqq t$. If we put $A^+ = \mathrm{diag}(A_1 - \alpha_1 I, A_2 - \alpha_2 I, \cdots, A_t - \alpha_t I)$ and $B^+ = \mathrm{diag}(B_{11}, B_{22}, \cdots, B_{tt})$, then easily $C = (A^+, B^+)$, $(C, A^+) = 0$, and A^+ is nilpotent. This proves part of Theorem 3.

We now prove Theorem 4. It suffices to establish the result when C is in rational canonical form. Since the rational canonical form of a nilpotent matrix is the direct sum of matrices of the form $C(\lambda^n)$. it suffices to assume that $C = C(\lambda^n)$. Let

$$B = \beta I_n - \mathrm{diag}(n-1, n-2, \cdots, 1, 0) ,$$

where $\beta \in K$. Then

$$C(\lambda^n) = (C(\lambda^n), B) .$$

Moreover, trace $B = n\beta - n(n-1)/2$. If $n \not\equiv 0 \pmod{p}$ then $\beta \in K$ can be found to make trace $B = 0$. If $n \equiv 0 \pmod{p}$ and p is odd, then trace $B = 0$ for any $\beta \in K$. This proves Theorem 4. An appeal to Theorem 4 completes the proof of Theorem 1.

We now prove Theorem 2. We may, as in the proof of Theorem 1, assume $C = C_1 + \cdots + C_t$, $A = A_1 + \cdots + A_t$, $B = (B_{ij})_{1 \leqq i,j \leqq t}$, where C_i has γ_i as its only eigenvalue. Then $C_i = (A_i, B_{ii})$, $(C_i, A_i) = 0$, $1 \leqq i \leqq t$. Let us simplify the notation and now denote C_i, A_i, B_{ii} by C, A, B, respectively. Then (4) holds, C has γ as its only eigenvalue, and by Theorem 3 we may take A to be nilpotent. After a similarity transformation by an element of $M_n(L)$, we may take

$$(6) \qquad A = C(\lambda^{e_1}) + C(\lambda^{e_2}) + \cdots + C(\lambda^{e_k}),\ e_1 \leqq e_2 \leqq \cdots \leqq e_k\ .$$

Let $B = (B_{ij})_{1 \leqq i,j \leqq k}$, where B_{ij} is $e_i \times e_j$. It is well known that $(C, A) = 0$ forces $C = (C_{ij})_{1 \leqq i,j \leqq k}$ where C_{ij} is an $e_i \times e_j$ stripe matrix. Moreover $C = (A, B)$ yields

$$(7) \qquad C_{ij} = C(\lambda^{e_i})B_{ij} - B_{ij}C(\lambda^{e_j})\ ,$$

for $1 \leqq i, j \leqq k$. From (7) follows: whenever $e_i = e_j$, trace $C_{ij} = 0$. Thus all square blocks in C have trace zero.

We now have to show that if $\gamma \neq 0$ then $e_i \equiv 0 \pmod{p}$ for $1 \leqq i \leqq k$. Suppose, for some fixed i with $1 \leqq i \leqq k$, that $e_i \not\equiv 0 \pmod{p}$. Choose integers u, v such that all the e_i-square blocks in C are $C_{\alpha\beta}$ with $u < \alpha, \beta \leqq v$. Since $e_i \not\equiv 0 \pmod{p}$ any e_i-square $C_{\alpha\beta}$ must have zero main diagonal since it has trace zero. Hence the first column of any $C_{\alpha\beta}$ with $u < \alpha, \beta \leqq v$ is a zero column. If $\alpha \leqq u$ and $\beta > u$ or if $u < \alpha \leqq v$ and $\beta > v$, then the first column of $C_{\alpha\beta}$ is a zero column since any such $C_{\alpha\beta}$ is an $e_\alpha \times e_\beta$ stripe matrix with $e_\alpha < e_\beta$. If $\alpha > v$ and $\beta > u$ then the first column of $C_{\alpha\beta}$ is all zeros, except (perhaps) for the top element. Thus the columns of C passing through the first columns of $C_{1,u+1}, C_{1,u+2}, \cdots, C_{1k}$ are entirely zero except for a fixed set of $k - v$ positions. Since $k - u > k - v$, these columns must be dependent. Thus C is singular. But γ is the only eigenvalue of C and $\gamma \neq 0$. Hence $e_i \equiv 0 \pmod{p}$.

To complete the proof of Theorem 2, we show that if $C = (C_{ij})_{1 \leqq i,j \leqq k}$ with each C_{ij} an $e_i \times e_j$ stripe matrix such that $e_i \equiv e_j \equiv 0 \pmod{p}$, then with A given by (6), we can find $B = (B_{ij})_{1 \leqq i,j \leqq k}$ such that B_{ij} is $e_i \times e_j$ and (7) holds for $1 \leqq i, j \leqq k$. First let $e_i > e_j$. Put row α of B_{ij} equal to $(\alpha - 1)$ (row $\alpha - 1$ of C_{ij}) for $2 \leqq \alpha \leqq e_j + 1$, and put all other rows of B_{ij} equal to zero. Then (7) holds. Next suppose that $e_i = e_j$. Then set row α of B_{ij} equal to $(\alpha - 1)$ (row $\alpha - 1$ of C_{ij}) for $2 \leqq \alpha \leqq e_i$, and set the first row of B_{ij} equal to zero. Then, because $e_i \equiv 0 \pmod{p}$, (7) holds. Finally suppose $e_i < e_j$.

This time put column $e_j - \alpha$ of B_{ij} equal to $(-\alpha)$ (column $e_j - \alpha + 1$ of C_{ij}) for $1 \leqq \alpha \leqq e_i$, and put all other columns of B_{ij} equal to zero. Then again (7) holds. This finishes the proof of Theorem 2. Note that, when $i = j$, trace $B_{ii} = ce_i(e_i - 1)/2$, where c is the (1, 2) element of C_{ii}. Since $e_i \equiv 0 \pmod{p}$, we find trace $B_{ii} = 0$ whenever p is odd. Combining this fact with Theorem 4 yields the remaining part of Theorem 3.

Theorem 2 is somewhat unsatisfactory in that whether $C \in M_n(L)$ can be represented in the form (4) depends only on the similarity class of C, and hence the necessary and sufficient condition should be a condition on the elementary divisors of C. In one particular case it is possible to obtain a condition involving the elementary divisors of C.

THEOREM 5. *Let $C \in M_n(L)$ and let $0 < p \leqq n$.* (i) *Suppose $A, B \in M_n(L)$ exist to satisfy* (4). *If C is not nilpotent let $\gamma \neq 0$ be an eigenvalue of C of multiplicity $< 2p$. Then the full set of elementary divisors (over L) of C belonging to γ is given by $(\lambda - \gamma)^a$ ($t - b$ times) and $(\lambda - \gamma)^{a+1}$ (b times), where a, b, t are integers such that $0 \leqq b < t$, $a > 0$, $p = at + b$.* (ii) *Suppose that each nonzero eigenvalue γ of C satisfies the following condition: the elementary divisors belonging to γ come in disjoint sets of the form $(\lambda - \gamma)^a$ ($t - b$ times), $(\lambda - \gamma)^{a+1}$ (b times), where integers a, b, t satisfy $0 \leqq b < t$, $a > 0$, $at + b \equiv 0 \pmod{p}$. Then $A, B \in M_n(L)$ exist to satisfy* (4).

Proof. Suppose (4) holds. Then, by Theorem 2, we may take $C = C_1 + \cdots + C_t$ where, say, C_1 has eigenvalue γ. From the fact that the multiplicity of γ is $< 2p$ and the form of C_1, it follows that C_1 is a p-square stripe matrix. We have merely to determine the elementary divisors of the stripe matrix C_1. This was accomplished by Gantmacher in [2]. Here is an alternative proof which establishes a result slightly more general than Gantmacher's.

LEMMA 1. *Let $W = (w_{\alpha\beta})_{1 \leqq \alpha, \beta \leqq n}$ be n-square. Let t be a fixed integer, $1 \leqq t < n$. Let $w_{\alpha\beta} = 0$ whenever $\beta - \alpha < t$ and $w_{\alpha\beta} \neq 0$ whenever $\beta - \alpha = t$. Then the elementary divisors of W are λ^a with multiplicity $t - b$ and λ^{a+1} with multiplicity b, where $n = at + b$, $0 \leqq b < t$.*

Proof of Lemma. Suppose, for a fixed r with $1 \leqq r < n - t$, that $w_{ij} = 0$ for all i, j for which $1 \leqq i < r$, $j - i > t$. Fix integer s so that $s \leqq n$ and $s - r < t$. Perform the following similarity transformation of W: add $-(w_{r,t+r})^{-1}w_{rs}$ times column $t + r$ to column

s, then in the resulting matrix add $(w_{r,t+r})^{-1}w_{rs}$ times row s to row $t+r$. The matrix obtained this way satisfies all the hypotheses imposed on W, has the same elements on diagonal (i, j) with $j-i=t$ as W, has the same row r as W except that position (r, s) is now zero. Thus, in W, we may, one by one, replace the elements to the right of $w_{r,t+r}$ with zeros, and we may do this for $r=1, 2, \cdots, n-t-1$. So in W we may assume $w_{ij}=0$ whenever $j-i \neq t$.

If $S=\operatorname{diag}(\delta_1, \delta_2, \cdots, \delta_n)$ with $\delta_1=\cdots=\delta_t=1$, $\delta_{t+i}=(w_{i,t+i}\delta_i)^{-1}$ for $1 \leqq i \leqq n-t$, then in SWS^{-1}, all positions (i, j) with $j-i \neq t$ are zero, and all positions (i, j) with $j-i=t$ are one. So in W assume $w_{ij}=0$ if $j-i \neq t$, $w_{ij}=1$ if $j-i=t$.

Now consider $\lambda I_n - W$. Given i, $1 \leqq i \leqq n-t$, find integers α, β such that $i=\alpha t+\beta$, $1 \leqq \beta \leqq t$, $\alpha \geqq 0$. In $\lambda I_n - W$ add $\lambda^{\alpha+1}$ times column $t+i$ to column β for $1 \leqq i \leqq n-t$. Call the resulting matrix W_1. As columns $1, 2, \cdots, t$ of W_1 are entirely zero down to row $n-t$, we may add appropriate multiples of row α to rows $\alpha+1, \alpha+2, \cdots, n$ for $1 \leqq \alpha \leqq n-t$, so that the resulting matrix, call it W_2, becomes a generalized permutation matrix (has exactly one nonzero element in each row and in each column). The nonconstant entries of W_2 occur in the lower left $t \times t$ block of W_2, and they give the elementary divisors of W.

To apply Lemma 1 to Theorem 5, let C_1 be an n-square stripe matrix with diagonal element γ, then let $W=C_1-\gamma I_n$. This proves Theorem 5 (i).

For Theorem 5 (ii) let $m=at+b$. We now know that m-square matrix Γ with elementary divisors $(\lambda-\gamma)^a$ ($t-b$ times) and $(\lambda-\gamma)^{a+1}$ (b times) is similar to the m-square stripe matrix Γ_1 whose first row is $(\gamma, 0, 0, \cdots, 0, 1, 0, \cdots, 0)$, where the 1 occurs at position $t+1$. Then if B has zero first row and row i of B is $(i-1)$ (row $i-1$ of Γ_1) for $2 \leqq i \leqq m$, we find $\Gamma_1=(C(\lambda^m), B)$, $(\Gamma_1, C(\lambda^m))=0$. By taking direct sums in an obvious manner we obtain Theorem 5 (ii).

THEOREM 6. *Let $C \in M_n(K)$ and let $0<p \leqq n$. Suppose for each nonzero eigenvalue γ of C, each elementary divisor $(\lambda-\gamma)^e$ of C over L belonging to γ occurs with multiplicity $k \equiv 0 \pmod p$ whenever $e \not\equiv 0 \pmod p$. Then $A, B \in M_n(K)$ exist such that (4) holds.*

Proof. Form the elementary divisors of C over K. The elementary divisors of C over L are obtained from the elementary divisors of C over K by decomposing the elementary divisors over K into products of powers of distinct linear factors. Let $p(\lambda)^e$ be an elementary divisor of C over K, where $p(\lambda)$ is irreducible over K. Either

$p(\lambda)$ has simple roots or else $p(\lambda)$ has multiple roots and in the latter event $p(\lambda)$ is a polynomial in λ^p. First suppose that $p(\lambda)$ has multiple roots or that $p(\lambda)$ has simple roots but $e \equiv 0 \pmod{p}$. Let $p(\lambda)^e = a_0 + a_1\lambda + \cdots + a_{m-1}\lambda^{m-1} + \lambda^m$. Then $a_j = 0$ if $j \not\equiv 0 \pmod{p}$, and $m \equiv 0 \pmod{p}$. Put $B = \operatorname{diag}(1, 2, 3, \cdots, m)$. Then $C(p(\lambda)^e) = (C(p(\lambda)^e), B)$ since $ja_{j-1} = 0$ if $j \not\equiv 1 \pmod{p}$ and $ja_{j-1} = a_{j-1}$ if $j \equiv 1 \pmod{p}$.

Now let $p(\lambda)^e$ be an elementary divisor of C over K such that $e \not\equiv 0 \pmod{p}$ and $p(\lambda)$ has simple roots. Let λ_0 be a zero of $p(\lambda)$. Then $(\lambda - \lambda_0)^e$ is an elementary divisor of C over L, hence appears as an elementary divisor of C over L with multiplicity $k \equiv 0 \pmod{p}$. Consequently $p(\lambda)^e$ is an elementary divisor of C over K with multiplicity $k \equiv 0 \pmod{p}$. Let $\Gamma = \operatorname{diag}(C(p(\lambda)^e), \cdots, C(p(\lambda)^e))$, ($k$ direct summands). Then it will be shown in the proof of Theorem 9 that A, B exist over K such that $\Gamma = (A, B)$, $(\Gamma, A) = (\Gamma, B) = 0$. This completes the proof of Theorem 6.

THEOREM 7. *Let $A \in M_n(K)$. Then $B, C \in M_n(K)$ exist satisfying* (1) *if and only if at least one elementary divisor of A over L is nonlinear.*

Proof. Suppose each elementary divisor of A over L is linear. After a similarity transformation of (1) by an element of $M_n(L)$, we may suppose $A = A_1 + \cdots + A_k$, where A_i is scalar with diagonal element α_i, and $\alpha_i \neq \alpha_j$ if $i \neq j$. Then $(C, A) = 0$ forces $C = C_1 + \cdots + C_k$. Partition $B = (B_{ij})_{1 \leqq i, j \leqq k}$. Then $C_i = (A_i, B_{ii}) = 0$ as A_i is scalar. Hence $C = 0$.

Suppose now A possesses an elementary divisor over L which is nonlinear. This means that A possesses an elementary divisor $p(\lambda)^e$ over K which has multiple roots. If $p(\lambda)$ has multiple roots then $p(\lambda)$ (and hence $p(\lambda)^e$) is a polynomial in λ^p:

$$p(\lambda)^e = a_0 + a_1\lambda + \cdots + a_{m-1}\lambda^{m-1} + \lambda^m ,$$

where $a_j = 0$ if $j \not\equiv 0 \pmod{p}$, and $m \equiv 0 \pmod{p}$. Let $B = (b_{ij})$ be m-square, with $b_{i+1,i} = i$ for $1 \leqq i \leqq m - 1$, and all other b_{ij} zero. Then $I_m = (C(p(\lambda)^e), B)$, because $ja_j = 0$ since $a_j = 0$ if $j \not\equiv 0 \pmod{p}$ and $j = 0$ (in K) if $j \equiv 0 \pmod{p}$. Clearly $(I_m, C(p(\lambda)^e)) = 0$.

To complete the proof we may suppose that, for each elementary divisor $p(\lambda)^e$ of A over K, the polynomial $p(\lambda)$ has simple roots. Hence for at least one elementary divisor $p(\lambda)^e$ over K, we have $e > 1$. Now when $p(\lambda)$ has simple roots, $C(p(\lambda)^e)$ is similar over K to

$$R = \begin{pmatrix} C(p(\lambda)) & I_f & 0 & 0 \cdots & 0 & 0 \\ 0 & C(p(\lambda)) & I_f & 0 \cdots & 0 & 0 \\ 0 & 0 & C(p(\lambda)) & I_f \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \cdots & C(p(\lambda)) & I_f \\ 0 & 0 & 0 & 0 \cdots & 0 & C(p(\lambda)) \end{pmatrix}$$

In R there are e blocks $C(p(\lambda))$ on the main block diagonal, and f is the degree of $p(\lambda)$. To prove this, one need only show that R and $C(p(\lambda)^e)$ are similar in $M_m(L)$, and this is not difficult. If we set $B = \text{diag}(I_f, 2I_f, \cdots, eI_f)$, then $C = (R, B) \neq 0$ and $(C, R) = 0$.

THEOREM 8. *Let $B \in M_n(K)$. Then $A, C \in M_n(K)$ exist such that* (1) *holds*: (i) *when $n > 2$ or when $n = p = 2$, if and only if B is not scalar*; (ii) *when $n = 2 \neq p$, if and only if the eigenvalues of B are distinct and in K.*

Proof. We may assume B is in rational canonical form: $B = C(p_1(\lambda)) + C(p_2(\lambda)) + \cdots + C(p_r(\lambda))$, where $p_1(\lambda), p_2(\lambda), \cdots, p_r(\lambda)$ are the invariant factors of B; $p_i(\lambda) \mid p_{i+1}(\lambda)$ for $1 \leqq i < r$. First suppose $r > 1$ and $p_r(\lambda)$ is not linear. Let X be a (degree $p_1(\lambda)$) $\times$ (degree $p_r(\lambda)$) matrix whose only nonzero element is a one at the top left corner. Let $A \in M_n(K)$ have X occupying the extreme top right corner; all other positions in A are zero. Let $C = (A, B)$. Then C is all zeros except that the upper right corner is

$$-C(p_1(\lambda))X + XC(p_r(\lambda)) \,. \tag{8}$$

Since $p_r(\lambda)$ is not linear, the top element of the second column of (8) is 1. Hence $C \neq 0$. But $CA = AC = 0$, hence $(C, A) = 0$.

This calculation fails if $r = 1$ or if $p_r(\lambda)$ is linear. When $p_r(\lambda)$ is linear, B is scalar. So let $r = 1$. Let $n \geqq 3$ and let B be a companion matrix. Take matrix A to be entirely zero except for a single one in extreme lower left corner. Then $C = (A, B) \neq 0$ and $CA = 0 = AC$.

Now let $n=2=p$. Let $B=C(\lambda^2+\beta_2\lambda+\beta_1)$. Put $A=C(\lambda^2+(1+\beta_1))$. Then $C = (A, B) \neq 0$ but $(C, A) = 0$.

Finally let $n = 2 \neq p$. If $C = (A, B)$ and $(C, A) = 0$, then also $C = (A, B - \beta I)$ and $(C, A) = 0$. Therefore in searching for C, A belonging to a given B, we may change B to $B - \beta I$, and hence assume we are working with a B which has zero trace. Cast B into rational canonical form. Since B clearly cannot be scalar if we are to have $C \neq 0$, we may suppose $B = C(\lambda^2 - \beta)$. Now note that if $C = (A, B)$ with $(C, A) = 0$ then also $C = (A - \alpha I, B)$ with

$(C, A - \alpha I) = 0$, and also $\alpha C = (\alpha A, B)$ with $(\alpha C, \alpha A) = 0$. By combining these two observations we may assume that the (1, 1) element of A is zero and that one of the nonzero elements of A is a one. Thus we may assume A is in one of the following three forms:

$$A = \begin{bmatrix} 0 & 1 \\ \alpha_1 & \alpha_2 \end{bmatrix}; \quad A = \begin{bmatrix} 0 & 0 \\ \alpha_1 & 1 \end{bmatrix}; \quad A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

In each of the last two possibilities $((A, B), A) \neq 0$. Thus we may confine our search for A to the first possibility. Then

$$C = (A, B) = \begin{bmatrix} \beta - \alpha_1 & -\alpha_2 \\ \alpha_2\beta & \alpha_1 - \beta \end{bmatrix},$$

$$(C, A) = \begin{bmatrix} -\alpha_2(\alpha_1 + \beta) & 2(\beta - \alpha_1) - \alpha_2^2 \\ 2\alpha_1(\alpha_1 - \beta) - \alpha_2^2\beta & \alpha_2(\beta + \alpha_1) \end{bmatrix}.$$

If $\alpha_2 = 0$ then $(C, A) = 0$ forces $\beta - \alpha_1 = 0$, which in turn forces $C = 0$. Since we want $C \neq 0$, we must have $\alpha_2 \neq 0$. Then $(C, A) = 0$ forces $\alpha_1 = -\beta$ and then $\alpha_2^2 = 4\beta$. Thus β must be a square in K, and then

$$C = \begin{bmatrix} 2\beta & -2\beta^{1/2} \\ 2\beta^{3/2} & -2\beta \end{bmatrix}.$$

Thus for $C \neq 0$, β must be a nonzero square in K. Let $\rho, -\rho$ be the eigenvalues of B. The $\beta = \rho^2$, so that β is a nonzero square in K if and only if $\rho, -\rho$ are distinct and in K. Moreover when β is a nonzero square in K and $\alpha_1 = -\beta$, $\alpha_2^2 = 4\beta$, we find $C = (A, B) \neq 0$ but $(C, A) = 0$. In terms of the original B with arbitrary trace, we find that B must have its eigenvalues distinct and in K. This completes the proof.

4. **Question II.** Let $C, A_\varphi, B_\varphi \in M_n(L)$ satisfy

$$(9) \qquad C = \sum_{\varphi=1}^{\theta} (A_\varphi, B_\varphi), \quad (C, A_\varphi) = (C, B_\varphi) = 0, \ 1 \leqq \varphi \leqq \theta,$$

where θ is a fixed positive integer. After a similarity transformation of (9) by an element of $M_n(L)$, we may assume $C = C_1 + \cdots + C_k$, where C_i has γ_i as its only eigenvalue and $\gamma_i \neq \gamma_j$ if $i \neq j$. Then $(C, A_\varphi) = (C, B_\varphi) = 0$ forces $A_\varphi = A_{\varphi 1} + \cdots + A_{\varphi k}$, $B_\varphi = B_{\varphi 1} + \cdots + B_{\varphi k}$, and then

$$C_i = \sum_{\varphi=1}^{\theta} (A_{\varphi i}, B_{\varphi i}), \ (C_i, A_{\varphi i}) = (C_i, B_{\varphi i}) = 0, \ 1 \leqq i \leqq k.$$

So let us change notation and consider (9) when C has a single

eigenvalue γ and is in Jordan canonical form. So now we let $C = C_1 + \cdots + C_k$ where C_i is e_i-square, with $e_1 \leqq e_2 \leqq \cdots \leqq e_k$, such that

$$C_i = \begin{pmatrix} \gamma & 1 & 0 & 0 \cdots 0 & 0 \\ 0 & \gamma & 1 & 0 \cdots 0 & 0 \\ 0 & 0 & \gamma & 1 \cdots 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \quad \vdots & \vdots \\ 0 & 0 & 0 & 0 \cdots \gamma & 1 \\ 0 & 0 & 0 & 0 \cdots 0 & \gamma \end{pmatrix} \quad \text{if } e_i > 1 ,$$

$C_i = (\gamma)$ if $e_i = 1$. Partition $A_\varphi = (A_{\varphi ij})_{1 \leqq i, j \leqq k}$ and $B_\varphi = (B_{\varphi ij})_{1 \leqq i, j \leqq k}$, where $A_{\varphi ij}$ and $B_{\varphi ij}$ are $e_i \times e_j$. Then $(C, A_\varphi) = 0$ and $(C, B_\varphi) = 0$ yield $C_i A_{\varphi ij} = A_{\varphi ij} C_j$, $C_i B_{\varphi ij} = B_{\varphi ij} C_j$, and it is well known that these equations force $A_{\varphi ij}$, $B_{\varphi ij}$ to be stripe matrices; $1 \leqq \varphi \leqq \theta$; $1 \leqq i, j \leqq k$.

We now require additional notation. Let the first row of $e_i \times e_j$ stripe matrix $A_{\varphi ij}$ be

$$(\underbrace{0, 0, \cdots, 0}_{e_j - e_i}, a_{\varphi ij_1}, a_{\varphi ij_2}, \cdots, a_{\varphi ij_{e_i}})$$

when $e_j \geqq e_i$, and

$$(a_{\varphi ij_1}, a_{\varphi ii_2}, \cdots, a_{\varphi ij_{e_j}})$$

when $e_i \geqq e_j$. Similarly we let $b_{\varphi ij_1}, b_{\varphi ij_2}, \cdots,$ be the not necessarily zero elements of the first row of the stripe matrix $B_{\varphi ij}$.

From (9) we get

$$C_i = \sum_{\varphi=1}^{\theta} \sum_{t=1}^{k} (A_{\varphi it} B_{\varphi ti} - B_{\varphi it} A_{\varphi ti}), \quad 1 \leqq i \leqq k . \tag{10}$$

A routine computation shows that the first row of the stripe matrix $A_{\varphi it} B_{\varphi ti}$ is e_i-tuple

$$(\underbrace{0, 0, \cdots, 0}_{e_i - e_t}, a_{\varphi it_1} b_{\varphi ti_1}, a_{\varphi it_1} b_{\varphi ti_2} + a_{\varphi it_2} b_{\varphi ti_1}, \cdots) \tag{11}$$

if $i \geqq t$; and

$$(\underbrace{0, 0, \cdots, 0}_{e_t - e_i}, a_{\varphi it_1} b_{\varphi ti_1}, a_{\varphi it_1} b_{\varphi ti_2} + a_{\varphi it_2} b_{\varphi ti_1}, \cdots) \tag{12}$$

if $t \geqq i$. In (11), the second nonzero coordinate is absent $e_t = 1$. In (12), all coordinates are zero if $e_t - e_i \geqq e_i$, and there is only one nonzero coordinate if $e_t - e_i = e_i - 1$.

From (10), (11), (12) we find that the (1, 1) element of C_i is

$$\gamma = \sum_{\varphi=1}^{\theta} \sum_{\substack{t: \\ t \neq i \\ e_t = e_i}} (a_{\varphi it_1} b_{\varphi ti_1} - b_{\varphi it_1} a_{\varphi ti_1}) , \tag{13}$$

and if $e_i > 1$, the (1, 2) element of C_i is

$$
\begin{aligned}
1 = &\sum_{\varphi=1}^{\theta} \sum_{\substack{t: \\ |e_t - e_i| = 1}} (a_{\varphi i t 1} b_{\varphi t i 1} - b_{\varphi i t 1} a_{\varphi t i 1}) \\
&+ \sum_{\varphi=1}^{\theta} \sum_{\substack{t: \\ t \neq i \\ e_t = e_i}} (a_{\varphi i t 1} b_{\varphi t i 2} + a_{\varphi i t 2} b_{\varphi t i 1} - b_{\varphi i t 1} a_{\varphi t i 2} - b_{\varphi i t 2} a_{\varphi t i 1}) .
\end{aligned} \tag{14}
$$

Note that, in (13) and (14), the terms with $t = i$ collapse to zero, hence it is necessary to sum only over values of t different from i.

Suppose $\gamma \neq 0$. Then it follows from (13) that at least one $t \neq i$ exists such that $e_t = e_i$. Fix i, and let $q + 1, q + 2, \cdots, q + r$ be all α with $e_\alpha = e_i$. Let j, s be two fixed integers such that $q + 1 \leqq j < s \leqq q + r$. Then in sum (13) for C_j, the terms

$$
\sum_{\varphi=1}^{\theta} (a_{\varphi j s 1} b_{\varphi s j 1} - b_{\varphi j s 1} a_{\varphi s j 1}) \tag{15}
$$

are contributed by C_s, whereas in the sum (13) for C_s, the terms

$$
\sum_{\varphi=1}^{\theta} (a_{\varphi s j 1} b_{\varphi j s 1} - b_{\varphi s j 1} a_{\varphi j s 1}) \tag{16}
$$

are contributed by C_j. Note that (15) and (16) add to zero. This means:

$$
\sum_{j=q+1}^{q+r} \gamma = 0 .
$$

Consequently the number r of integers α with $e_\alpha = e_i$ satisfies $r \equiv 0 \pmod{p}$. This means: if $\gamma \neq 0$ and $(\lambda - \gamma)^e$ is an elementary divisor of C over L with multiplicity r, then $r \equiv 0 \pmod{p}$.

Now let $\gamma = 0$. Select an integer i such that $e_i > 1$. Recall that $e_1 \leqq e_2 \leqq \cdots \leqq e_k$. Find the set S of consecutive integers of maximal length containing i such that whenever α and $\alpha + 1 \in S$ then $e_{\alpha+1} - e_\alpha = \delta_\alpha$ with $\delta_\alpha = 0$ or 1. Let $S = \{q + 1, q + 2, \cdots, q + r\}$. Then either $q = 0$ or $q > 0$ and $e_q < e_{q+1} - 1$; and either $q + r = k$ or $q + r < k$ and $e_{q+r+1} > e_{q+r} + 1$.

Because of (14), there exists at least one integer $t \neq i$ such that $e_t - e_i = 0$ or ± 1. Thus the length of S is at least two. We wish to examine the circumstances under which $e_{q+1} > 1$. So assume $e_{q+1} > 1$.

Let $j < s$ be two members of S. If $e_s - e_j = 0$ then in the sum (14) for C_j, C_s contributes the terms

$$
\sum_{\varphi=1}^{\theta} (a_{\varphi j s 1} b_{\varphi s j 2} + a_{\varphi j s 2} b_{\varphi s j 1} - b_{\varphi j s 1} a_{\varphi s j 2} - b_{\varphi j s 2} a_{\varphi s j 1}) \tag{17}
$$

whereas in the sum (14) for C_s, C_j contributes the terms

$$\sum_{\varphi=1}^{\theta} (a_{\varphi s j_1} b_{\varphi j s_2} + a_{\varphi s j_2} b_{\varphi j s_1} - b_{\varphi s j_1} a_{\varphi j s_2} - b_{\varphi s j_2} a_{\varphi j s_1}) . \tag{18}$$

Note that (17) and (18) add to zero. If $e_s - e_j = 1$ then in the sum (14) for C_j, C_s contributes (15) whereas in the sum (14) for C_s, C_j contributes the term (16), and (15) and (16) sum to zero. If $e_s - e_j > 1$, then C_s contributes no terms to the sum (14) for C_j and C_j contributes no terms to the sum (14) for C_s. Because of the maximality of S, no integer outside S contributes any terms to the sum (14) for C_i, where i is any integer in S. By virtue of these remarks we find

$$\sum_{q+1}^{q+r} 1 = 0 .$$

Hence $r \equiv 0 \pmod{p}$. This argument does not work when $e_{q+1} = 1$ since then equation (14) does not exist when $i = q + 1$. This completes the proof of the first part of Theorem 9.

THEOREM 9. *Suppose* C, A_φ, $B_\varphi \in M_n(K)$ *and satisfy* (9). *Then the elementary divisors of* C *over* L *can be classified into disjoint sets of the types* (20), (21), (22) *below. Conversely, if* $C \in M_n(K)$ *is such that its elementary divisors over* L *classify into disjoint sets* (20), (21), (22), *then* A, $B \in M_n(K)$ *exist such that*

$$C = (A, B), \ (C, A) = (C, B) = 0 . \tag{19}$$

(20) $\lambda^{e_1}, \lambda^{e_2}, \cdots, \lambda^{e_r}$ *with* $e_1 = 1$, $e_{i+1} - e_i = \delta_i = 0$ *or* 1, $1 \leqq i < r$;

(21) $\lambda^{e_1}, \lambda^{e_2}, \cdots, \lambda^{e_r}$ *with* $e_1 > 1$, $e_{i+1} - e_i = \delta_i = 0$ *or* 1, $1 \leqq i < r$, *and* $r \equiv 0 \pmod{p}$;

(22) $(\lambda - \gamma)^e$, $\gamma \neq 0$, *with multiplicity* $r \equiv 0 \pmod{p}$.

Note if $p = 0$ or $p > n$, C cannot have any elementary divisors in classifications (21) or (22).

We now have to establish the converse part of Theorem 9. Consider the elementary divisors of C over K. If $p(\lambda)^e$ is such an elementary divisor then either $p(\lambda) = \lambda$ and then $p(\lambda)^e$ falls into classification (20) or (21); or else $p(\lambda) \neq \lambda$. In the latter event let

$$p(\lambda)^{e_1}, \ p(\lambda)^{e_2}, \ \cdots, \ p(\lambda)^{e_t} \tag{23}$$

be all the elementary divisors of C over K involving $p(\lambda)$. Let $\gamma \in L$ be a zero of $p(\lambda)$ with multiplicity ν. Then all the elementary divisors of C over L involving γ are

$$(\lambda - \gamma)^{\nu e_1}, (\lambda - \gamma)^{\nu e_2}, \cdots, (\lambda - \gamma)^{\nu e_t} . \tag{24}$$

By the hypotheses of the theorem, the multiplicity of

$$(\lambda - \gamma)^{\nu e_1}$$

is $r \equiv 0 \pmod{p}$. This implies that the number of integers $e_1, e_2, \cdots, e_t$ equal to e_1 is r. Hence the multiplicity of

$$p(\lambda)^{e_1}$$

in (23) is $r \equiv 0 \pmod{p}$.

To complete the proof it will suffice to assume that

$$C = C(p(\lambda)^{e_1}) \dotplus \cdots \dotplus C(p(\lambda)^{e_r}) \tag{25}$$

and that either $p(\lambda) = \lambda$ and the integers $r, e_1, \cdots, e_r$ satisfy (20) or (21) above; or else that $p(\lambda) \neq \lambda$ and

$$e_1 = e_2 = \cdots = e_r = e;\ r \equiv 0 \pmod{p}\ . \tag{26}$$

If in each of these three cases we find $A, B \in M_n(K)$ such that (19) holds, then the general case will follow by taking direct sums.

Suppose in (25) that $p(\lambda) = \lambda$ and that the integers $e_1, \cdots, e_r, r$ satisfy (20). We may assume $e_1 = 1$, $e_2 = 2$, since in the more general situation we need only consider $0 \dotplus C$, $0 \dotplus A$, $0 \dotplus B$. So assume $e_1 = 1$, $e_2 = 2$. We present $A = (A_{\alpha\beta})_{1 \leqq \alpha, \beta \leqq r}$ in partitioned form, where block $A_{\alpha\beta}$ is $e_\alpha \times e_\beta$ and is zero whenever $\beta \neq \alpha - 1$. We set

$$A_{\alpha,\alpha-1} = (r + 1 - \alpha) I_{e_\alpha} \text{ if } e_\alpha = e_{\alpha-1}; \tag{27}$$

$$A_{\alpha,\alpha-1} = \begin{bmatrix} (r + 1 - \alpha) I_{e_{\alpha-1}} \\ 0_{1, e_{\alpha-1}} \end{bmatrix} \text{ if } e_\alpha - 1 = e_{\alpha-1}\ ; \tag{28}$$

$2 \leqq \alpha \leqq r$. We present $B = (B_{\alpha\beta})_{1 \leqq \alpha, \beta \leqq r}$ in partitioned form also, where block $B_{\alpha\beta}$ is $e_\alpha \times e_\beta$ and is zero whenever $\beta \neq \alpha + 1$. We set

$$B_{\alpha,\alpha+1} = C(\lambda^{e_\alpha}) \text{ if } e_\alpha = e_{\alpha+1}\ ; \tag{29}$$

$$B_{\alpha,\alpha+1} = (0_{e_\alpha,1}, I_{e_\alpha}) \text{ if } e_{\alpha+1} = e_\alpha + 1\ ; \tag{30}$$

$1 \leqq \alpha \leqq r - 1$. Then with C given by (25), (19) holds.

Next let $p(\lambda) = \lambda$ and let the integers $e_1, \cdots, e_r, r$ satisfy (21) above. We use the A, B just constructed. Because $r = 0$ (in K), (19) still holds, where C is given by (25).

Now suppose $e_1, \cdots, e_r$ satisfy (26). We now set $A_{\alpha,\alpha-1} = (r + 1 - \alpha)\, C(p(\lambda)^e)$ for $2 \leqq \alpha \leqq r$; $B_{\alpha,\alpha+1} = I_f$ for $1 \leqq \alpha \leqq r - 1$; here $f =$ degree $p(\lambda)^e$. Otherwise we keep A, B exactly as constructed above. Then with C given by (25), (19) still holds, since $r = 0$ (in K). This completes the proof of Theorem 9 (and also of Theorem 6).

THEOREM 10. *Let $A \in M_n(K)$. Then $B, C \in M_n(K)$ exist satisfying (2) if and only if*: (i) *when $p = 2$, at least one eigenvalue of A has a nonlinear elementary divisor over L*; (ii) *when $p \neq 2$, at least one eigenvalue of A has multiplicity not less than 3 and a nonlinear elementary divisor over L.*

Proof. Suppose C, A, B satisfy (2). By Theorem 7, A must have a nonlinear elementary divisor over L. Suppose $p \neq 2$ and that each eigenvalue possessing a nonlinear elementary divisor over L has multiplicity < 3, hence multiplicity 2. After a similarity transformation by an element of $M_n(L)$ we may suppose $A = A_1 + \cdots + A_k$ where A_i has α_i as its only eigenvalue and $\alpha_i \neq \alpha_j$ for $i \neq j$. Our assumptions imply that A_i is either scalar or 2-square. From $(C, A) = 0$ we get $C = C_1 + \cdots + C_k$. Let $B = (B_{ij})_{1 \leq i,j \leq k}$. Then $(C, B) = 0$ implies $(C_i, B_{ii}) = 0$; $(C, A) = 0$ implies $(C_i, A_i) = 0$; and $C = (A, B)$ implies $C_i = (A_i, B_{ii})$; $1 \leq i \leq k$. If A_i is scalar, $(A_i, B_{ii}) = 0$. Thus $C_i \neq 0$ is possible only if C_i is 2-square. Any 2-square matrix is either scalar or nonderogatory. If C_i is nonderogatory then $(C_i, A_i) = 0$ and $(C_i, B_{ii}) = 0$ force A_i, B_{ii} to be polynomials in C_i, hence commutative, hence $C_i = (A_i, B_{ii}) = 0$. If C_i is scalar, $C_i = (A_i, B_{ii})$ forces trace $C_i = 0$, hence $C_i = 0$. Thus $C = 0$. Therefore the conditions of the theorem are necessary.

Suppose the conditions of the theorem are satisfied. Consider the elementary divisors of A over K. These are of the form $p(\lambda)^e$ where $p(\lambda)$ is irreducible over K. Either all $p(\lambda)$ have simple roots or else there exists an elementary divisor $p(\lambda)^e$ in which $p(\lambda)$ has multiple roots. In the latter event let $m =$ degree $p(\lambda)^e$. In the proof of Theorem 7, $B \in M_m(K)$ was found such that $I_m = (C(p(\lambda)^e), B)$. This completes the proof of the theorem if some $p(\lambda)$ has multiple roots.

We may therefore assume that for each elementary divisor $p(\lambda)^e$ of A over K, $p(\lambda)$ has simple roots. Suppose $e \geq 3$ for some elementary divisor $p(\lambda)^e$ over K. Then $C(p(\lambda)^e)$ is similar over K to the matrix R in the proof of Theorem 7. Let

$$B = \begin{pmatrix} 0 & 0 \cdots 0 & I_f & 0 \\ 0 & 0 \cdots 0 & 0 & 2I_f \\ 0 & 0 \cdots 0 & 0 & 0 \\ \vdots & \vdots \quad \vdots & \vdots & \vdots \\ 0 & 0 \cdots 0 & 0 & 0 \end{pmatrix}.$$

Then if $C = (R, B)$, C is entirely zero except for a block I_f in the upper right corner. We find that $(C, R) = 0$ and $CB = BC = 0$.

To finish the proof we may now assume that $e \leq 2$ for each

elementary divisor $p(\lambda)^e$ of A over K. For $p \neq 2$, the conditions of the theorem imply that A has either a couple of elementary divisors over K of the form $p(\lambda)^2$, $p(\lambda)^2$; or else a couple of elementary divisors of the form $p(\lambda)$, $p(\lambda)^2$. In the first event, if we take

$$A = \begin{bmatrix} C(p(\lambda)) & I & 0 & 0 \\ 0 & C(p(\lambda)) & 0 & 0 \\ 0 & 0 & C(p(\lambda)) & I \\ 0 & 0 & 0 & C(p(\lambda)) \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

then (2) holds. In the second event, if we take

$$A = \begin{bmatrix} C(p(\lambda)) & I & 0 \\ 0 & C(p(\lambda)) & 0 \\ 0 & 0 & C(p(\lambda)) \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & I \\ 0 & 0 & 0 \end{bmatrix},$$

then again (2) holds.

The calculations of the above paragraphs hold if $p = 2$. To finish the proof we need only consider the case when $p = 2$ and $A = C(p(\lambda)^2)$. But $p(\lambda^2)$ is a polynomial in λ^2, hence as in the proof of Theorem 7, $I = (C(p(\lambda^2)), B)$ for a certain B over K. This completes the proof of Theorem 10.

It should be pointed out that much of the argument of § 4 is very close to arguments first found by McCoy [4].

5. The number of similarity classes of commutators. In § 5 only, we let $p(n)$ denote the number of partitions of n and $p^d(n)$ denote the number of partitions of n into distinct summands.

THEOREM 11. *Let $p = 0$ or $p > n$. Then*: (i) *the number of similarity classes of matrices $C \in M_n(K)$ for which a representation* (3) *or* (4) *exists in $M_n(K)$ is $p(n)$*; (ii) *the number of similarity classes of matrices $C \in M_n(K)$ for which a representation* (9) *or* (19) *exists in $M_n(K)$ is $p^d(n)$.*

Proof. In view of Theorem 1, (i) is just a count on the number of similarity classes of nilpotent matrices; the result is well known and trivial. To prove (ii) we use the fact that (20) of Theorem 9 is the necessary and sufficient condition for (9) or (19) to hold, when $p = 0$ or $p > n$. Then from $e_{i+1} - e_i = \delta_i$ for $1 \leqq i < r$, and from $e_1 + \cdots + e_r = n$, we get

$$r + \sum_{j=1}^{r-1} (r - j)\delta_j = n \,. \tag{31}$$

Since each δ_j is 0 or 1, the result immediately follows from (31).

6. Some lemmas. Suppose $C \in M_n(K)$ has elementary divisors

$$\lambda^{e_1}, \lambda^{e_2}, \cdots, \lambda^{e_r} \tag{32}$$

with

$$e_{i+1} - e_i = \delta_i = 0 \text{ or } 1, \ 1 \leqq i < r \,, \tag{33}$$

and with $e_2 > 1$. Let $B = (B_{\alpha\beta})_{1 \leqq \alpha, \beta \leqq r}$ where $e_\alpha \times e_\beta$ matrix $B_{\alpha\beta}$ is zero if $\beta \neq \alpha + 1$, and $B_{\alpha,\alpha+1}$ is given by (29) or (30), for $1 \leqq \alpha < r$.

Let $\sigma(1) = 0$ and for $i > 1$ define $\sigma(i)$ to be the number of integers α for which $e_\alpha = e_{\alpha+1}$, $1 \leqq \alpha < i$.

LEMMA 2. (i) *If $\sigma(r) < e_1$ then the elementary divisors of B are: λ^α with multiplicity one if $e_\alpha < e_{\alpha+1}$, and with multiplicity two if $e_\alpha = e_{\alpha+1}$, $1 \leqq \alpha < r$; together with λ^r with multiplicity $e_1 - \sigma(r)$.* (ii) *If $\sigma(r) \geqq e_1$, let u be the largest integer for which $\sigma(u) < e_1$. Then the elementary divisors of B are: λ^α with multiplicity one if $e_\alpha < e_{\alpha+1}$, and with multiplicity two if $e_\alpha = e_{\alpha+1}$, for $1 \leqq \alpha < u$; λ^u with multiplicity equal to the number of integers $\beta \geqq u$ for which $e_\beta = u$ (always at least once); if $u < e_r$, λ^α with multiplicity equal to the number of integers β for which $e_\beta = \alpha$ (always at least once), for all α such that $u < \alpha \leqq e_r$.*

Proof. Note that

$$e_i = i - \sigma(i) + e_1 - 1 \,, \tag{34}$$

(verify by induction on i) and if α is the largest integer $< i$ for which $e_\alpha = e_{\alpha+1}$, then

$$\sigma(i) = \sigma(\alpha) + 1 \,. \tag{35}$$

Given i, $1 \leqq i < r$, define $e_i \times e_j$ matrix M_{ij} for $j \leqq i$, as follows. If $e_i < e_{j+1}$, then $M_{ij} = 0$ for all j, $1 \leqq j \leqq i$. If $e_i = e_{i+1}$ and $\sigma(i) < e_1$ then M_{i1} is entirely zero except for λ^i occupying position

$(e_i, e_1 - \sigma(i))$, and $M_{i2}, M_{i3}, \cdots, M_{ii}$ are all zero matrices. If $e_i = e_{i+1}$ and $\sigma(i) \geqq e_1$ then all M_{ij} are zero matrices except for

$$M_{i,i-e_i+1}$$

which has

$$\lambda^{e_i}$$

in position $(e_i, 1)$ as the only nonzero element.

At a later point it will be necessary to observe that a certain matrix Γ_r is a generalized permutation matrix. For this purpose observe the following facts. If β is the smallest integer for which $e_\beta = e_{\beta+1}$ then $\sigma(\beta) = 0$ and $M_{\beta 1}$ has its nonzero element in the extreme lower right corner. If $i > \beta$ is such that $\sigma(i) < e_i$ and $e_i = e_{i+1}$, let α be the largest integer $< i$ with $e_\alpha = e_{\alpha+1}$. Then (35) implies that the nonzero element of M_{i1} occurs one column to the left of the nonzero element of $M_{\alpha 1}$. If $\sigma(r) \geqq e_1$, take u to be the largest integer for which $\sigma(u) < e_1$. Then $\sigma(u+1) = e_1 = \sigma(u) + 1$ and $e_u = e_{u+1}$. If $u < r$, it follows that M_{u1} has a nonzero element in the extreme lower left corner. Next observe that, because of (34), $i - e_i + 1 = \sigma(i) - e_1 + 2$, so that if $\sigma(i) \geqq e_1$ and $e_i = e_{i+1}$, then $M_{i1} = 0$ and, in addition, if α is the smallest integer for which $\alpha > i$ and $e_\alpha = e_{\alpha+1}$, then (by (35)), $\alpha - e_\alpha + 1 = (i - e_i + 1) + 1$.

Next, given integer t, $1 \leqq t \leqq r$, define $e_t \times e_j$ matrix N_{tj} for $j \leqq t$ as follows. For nonnegative $i \leqq \min(t-2, e_t - 2)$, let $N_{t,t-i}$ be all zeros except for λ^{i+1} occupying position $(i+1, 1)$. For $i = \min(t-1, e_t - 1)$ let $N_{t,t-i}$ be all zeros except for λ^{i+1} occupying all positions (α, β) with $\alpha - \beta = i$. For $i > \min(t-1, e_t - 1)$, set $N_{t,t-i} = 0$.

Note that, when $t \leqq e_t$, then for $i = \min(t-1, e_t - 1)$, we have $t - i = 1$ and $N_{t,t-i}$ has λ^t occupying the diagonal of positions (α, β) with $\alpha - \beta = t - 1$, and that the number of zeros to the right of λ^t in the last row of $N_{t,t-i}$ is $\sigma(t)$ (by 34)). Moreover, if $t \leqq e_t$, then by (34), $\sigma(t) < e_1$, so that if q is the largest integer $< t$ with $e_q = e_{q+1}$, the nonzero element in the last row of M_{q1} is one column to the right of the nonzero element in the last row of N_{t1}. When $t > e_t$, then with $i = \min(t-1, e_t - 1)$, we have $t - i = t - e_t + 1 \geqq 2$; and for this i, $N_{t,t-i}$ has

$$\lambda^{e_t}$$

in the extreme lower left as the only nonzero element. Moreover, if α is the largest integer $< t$ with $e_\alpha = e_{\alpha+1}$, then

$$t - e_t + 1 = (\alpha - e_\alpha + 1) + 1 .$$

Finally note that no two of $N_{t1}, N_{t2}, \cdots, N_{tt}$ have a nonzero element in the same row. These remarks will also be used below to see that Γ_r is a generalized permutation matrix.

For fixed t, $1 \leqq t \leqq r$, let Γ_t be a polynomial matrix presented in partitioned form as follows. For $i < t$, block row i of Γ_t is $(M_{i1}, M_{i2}, \cdots, M_{ii}, -B_{i,i+1}, 0, \cdots, 0)$. (Observe that each row of block row i has exactly one nonzero element.) Block row t of Γ_t is $(N_{t1}, N_{t2}, \cdots, N_{tt}, -B_{t,t+1}, 0, \cdots, 0)$ when $t < r$, and block row t of Γ_t is $(N_{r1}, N_{r2}, \cdots, N_{rr})$ when $t = r$. For all i, $t < i < r$, block row i of Γ_t is

$$(0, \cdots, 0, \lambda I_{e_i}, -B_{i,i+1}, 0, \cdots, 0)$$

where the block

$$\lambda I_{e_i}$$

appears in the main block diagonal position. For $t < r$, block row r of Γ_t is

$$(0, 0, \cdots, 0, \lambda I_{e_r}) .$$

Observe that each row in block row r of Γ_r contains exactly one nonzero element.

Note that when $t = 1$, $\Gamma_1 = \lambda I - B$. This is so since for $t = 1$, $\min(t - 1, e_t - 1) = 0$, and then $N_{11} = \lambda I_{e_1}$. We now supply elementary row and column operations that successively convert Γ_1 into $\Gamma_2, \Gamma_3, \cdots, \Gamma_r$.

Let $t < r$. To convert Γ_t into Γ_{t+1}, first add λ times the row of Γ_t passing through row i of the block $-B_{t,t+1}$, to the row of Γ_t passing through row $i + 1$ of the block

$$\lambda I_{e_{t+1}} ,$$

for $1 \leqq i \leqq e_{t+1} - 1$. Call the resulting matrix Γ'_t. Block row $t + 1$ of Γ'_t is now $(N_{t+1,1}, \cdots, N_{t+1,t+1}, -B_{t+1,t+2}, 0, \cdots, 0)$ ($-B_{t+1,t+2}$ is absent if $t + 1 = r$). In Γ'_t the columns of Γ'_t passing through columns $2, 3, \cdots, e_{t+1}$ of block $-B_{t,t+1}$ are entirely zero except for -1's appearing in $-B_{t,t+1}$. So if we add appropriate multiples of these columns of Γ'_t to the columns of Γ'_t to the left of block $-B_{t,t+1}$, we can replace all elements in blocks $N_{t1}, N_{t2}, \cdots, N_{tt}$ with zeros, except that, when $e_t = e_{t+1}$, the nonzero element in the last row of the block $N_{t,t-i}$, with $i = \min(t - 1, e_t - 1)$, cannot get replaced with zero. If $\sigma(t) < e_1$, (34) implies $t \leqq e_t$, hence $t - 1 = \min(t - 1, e_t - 1)$, hence the element that does not get replaced with zero is λ^t at position $(e_t, e_1 - \sigma(t))$ in N_{t1}. Thus when $e_t = e_{t+1}$ and $\sigma(t) < e_1$, our column operations convert N_{t1} into M_{t1}. If $\sigma(t) \geqq e_1$, then (34) yields $t > e_t$,

so that $\min(t-1, e_t - 1) = e_t - 1$, hence the element that does not get replaced with zero is

$$\lambda^{e_t}$$

at position $(e_t, 1)$ of

$$N_{t,t-e_t+1}\,.$$

Thus, whether $\sigma(t) < e_1$ or $\sigma(t) \geqq e_1$, the blocks $N_{t1}, \cdots, N_{tt}$ in Γ'_t are converted into $M_{t1}, \cdots, M_{tt}$. Hence Γ'_t becomes Γ_{t+1}.

The remarks above show that Γ_r is a generalized permutation matrix. As each nonconstant element of Γ_r is a power of λ, to obtain the invariant factors we need only locate all of the nonconstant polynomials in Γ_r.

Let $\sigma(r) < e_1$. Then from the block $M_{\alpha 1}$ in Γ_r we obtain λ^α precisely when $e_\alpha = e_{\alpha+1}$, $1 \leqq \alpha < r$. From the block $N_{r,r-\alpha}$ we get $\lambda^{\alpha+1}$, $0 \leqq \alpha < r-1$. From the block N_{r1} we obtain λ^r exactly $e_1 - \sigma(r)$ times. This completes the proof of (i).

Let $\sigma(r) \geqq e_1$. Since $\sigma(1) = 0 < e_1$, $\sigma(r) \geqq e_1$, there exists a largest positive integer u with $\sigma(u) < e_1$. Since $\sigma(2) \leqq 1 < e_2$, clearly $u \geqq 2$. We must have $\sigma(u+1) = e_1$ and also that $e_u = e_{u+1}$. From (34) with $i = u+1$, we get $e_{u+1} = u$. Hence $u = e_u = e_{u+1}$. Let $\alpha < u$. Then $\sigma(\alpha) \leqq \sigma(u) < e_1$. Thus the only one of blocks $M_{\alpha 1}, M_{\alpha 2}, \cdots, M_{\alpha\alpha}$ which can be nonzero is $M_{\alpha 1}$, and in $M_{\alpha 1}$ we find a nonzero element, λ^α, if and only if $e_\alpha = e_{\alpha+1}$. Let $\alpha > u$, but $\alpha < r$. (Such an α need not exist in all cases.) Then $\sigma(\alpha) \geqq e_1$, so the only block among $M_{\alpha 1}, M_{\alpha 2}, \cdots, M_{\alpha\alpha}$ that can be nonzero is $M_{\alpha,\alpha-e_\alpha+1}$, and in this block we find a nonzero element, λ^{e_α}, if and only if $e_\alpha = e_{\alpha+1}$. Thus if $j > u$, we pick up λ^j exactly as many times as there are integers α with $e_\alpha = e_{\alpha+1}(= j)$. And we pick up λ^u as many times as there are integers $\alpha > u$ with $u = e_\alpha = e_{\alpha+1}$. Since $\sigma(u) < e_1$ and $e_u = e_{u+1}$, we also pick up λ^u once in block M_{u1}. Finally, in the blocks $N_{rr}, N_{r,r-1}, \cdots, N_{r,r-e_r+1}$ we pick up $\lambda, \lambda^2, \cdots, \lambda^{e_r}$, each exactly once. This completes the proof of Lemma 1.

COROLLARY 1. *Let* $e_1 = 1$. *Then the elementary divisors of* B *are*

$$\lambda^{e_i},\ 1 \leqq i \leqq r\,.$$

Proof. If $\sigma(r) = 0$, then $e_1 < e_2 < \cdots < e_r$, so $e_i = i$ for $1 \leqq i \leqq r$, hence Lemma 2(i) gives the result. If $\sigma(r) \geqq 1$ then $\sigma(r) \geqq e_1$. Because $e_1 = 1$, (34) gives $e_i \leqq i$, so $e_{u-1} < u$. Hence λ^u occurs as many times as there are integers β with $u = e_\beta$. Because $\sigma(u) < e_1 = 1$, $\sigma(u) = 0$. So $1 = e_1 < e_2 < \cdots < e_{u-1} < e_u = u$, hence for $\alpha < u$,

λ^α occurs exactly once, hence exactly as many times as there are β with $e_\beta = \alpha$. This completes the proof.

For the next Lemma, let $e_1, \cdots, e_r$ be positive integers with $e_1 \leqq e_2 \leqq \cdots \leqq e_r$. Set $e_0 = 0$. Let $A = (A_{\alpha\beta})_{1\leqq\alpha,\beta\leqq r}$ where $A_{\alpha\beta}$ is $e_\alpha \times e_\beta$ and $A_{\alpha\beta} = 0$ if $\alpha \neq \beta + 1$, and where

$$A_{\alpha,\alpha-1} = \begin{bmatrix} a_\alpha I_{e_{\alpha-1}} \\ 0_{e_\alpha - e_{\alpha-1}, e_{\alpha-1}} \end{bmatrix},$$

for $1 < \alpha \leqq r$, with each $a_\alpha \neq 0$.

LEMMA 3. *The elementary divisors of A are λ^{r-i+1} with multiplicity $e_i - e_{i-1}$, for $1 \leqq i \leqq r$.*

Proof. For $i \leqq \alpha$ let $M_{i\alpha}$ be an $e_i \times e_\alpha$ matrix constructed according to the following rule: the element of $M_{i\alpha}$ at position (ξ, ξ) is $\lambda^{\alpha-i+1}$, for all ξ such that $e_{i-1} < \xi \leqq e_i$. (If $e_{i-1} = e_i$, there are no such ξ.) All other elements of $M_{i\alpha}$ are zero. Now let $\Delta_t = (\Delta_{tij})_{1\leqq i,j\leqq r}$ be presented in partitioned form, as follows. Block Δ_{tij} is $e_i \times e_j$; $\Delta_{t,i,i-1} = -A_{i,i-1}$ for $2 \leqq i \leqq r$; $\Delta_{tit} = M_{it}$ for $1 \leqq i \leqq t$; $\Delta_{tii} = \lambda I_{e_i}$ for all $i > t$; and all other blocks Δ_{tij} in Δ_t are zero.

Note that $\Delta_1 = \lambda I - A$. We now supply elementary row and column operations to convert Δ_t into Δ_{t+1}. In Δ_t add $(a_{t+1})^{-1}\lambda$ times the column passing through column i of block $-A_{t+1,t}$ to the column of Δ_t passing through column i of block $\lambda I_{e_{t+1}}$, for $1 \leqq i \leqq e_t$. Call the resulting matrix Δ_t'. In Δ_t', we find $M_{1,t+1}, M_{2,t+1}, \cdots, M_{t+1,t+1}$ occupying the first $t+1$ positions in the block column $t+1$. Moreover, in Δ_t', the rows passing through the first e_t rows of block $-A_{t+1,t}$ are entirely zero, except for the nonzero elements in the main diagonal of $-A_{t+1,t}$. By adding appropriate multiples of the rows of Δ_t' passing through the first e_t rows of $-A_{t+1,t}$ to the rows of Δ_t' above block $-A_{t+1,t}$, we can replace blocks $M_{1t}, M_{2t}, \cdots, M_{tt}$ in Δ_t' with zero blocks. This converts Δ_t into Δ_{t+1}.

It is not difficult to verify that Δ_r is a generalized permutation matrix. The nonconstant entries appear in the blocks $M_{1r}, \cdots, M_{rr}$; from block M_{ir} we pick up λ^{r-i+1} exactly $e_i - e_{i-1}$ times. This completes the proof.

For the next Lemma let $A = (A_{\alpha\beta})_{1\leqq\alpha,\beta\leqq r}$ where each $A_{\alpha\beta}$ is m-square, $A_{\alpha\beta} = 0$ if $\alpha \neq \beta + 1$, and $A_{\alpha,\alpha-1}$ is nonsingular for $1 < \alpha \leqq r$.

LEMMA 4. *The elementary divisors of A are λ^r with multiplicity m.*

Proof. Let $S = \operatorname{diag}(I_m, A_{21}, A_{32}A_{21}, \cdots, A_{r,r-1}A_{r-1,r-2}\cdots A_{21})$. Then

S is nonsingular and $S^{-1}AS$ has the same structure as A, except now the block in position $(\alpha, \alpha - 1)$ is I_m, for $2 \leqq \alpha \leqq r$. To complete the proof apply Lemma 1 to the matrix $S^{-1}AS$.

7. **Chains of commutators.** We say matrix $C \in M_n(K)$ supports a chain of commutators (type I) of length t if matrices X_i, $Y_i \in M_n(K)$ exist such that

$$Y_{i-1} = (X_i, Y_i), \ (Y_{i-1}, Y_i) = 0, \ 1 \leqq i \leqq t , \tag{36}$$

with $C = Y_0$. We say $C \in M_n(K)$ supports a chain of commutators (type II) of length t if matrices X_i, $Y_i \in M_n(K)$ exist such that

$$Y_{i-1} = (X_i, Y_i), \ (Y_{i-1}, X_i) = (Y_{i-1}, Y_i) = 0, \ 1 \leqq i \leqq t , \tag{37}$$

with $C = Y_0$.

THEOREM 12. *Let $0 < p \leqq n$. Then $C \in M_n(L)$ supports a chain of commutators* (36) *of type I and length t if and only if C satisfies the condition of Theorem* 2. *Moreover, if C supports some commutator chain of type I, C also supports a commutator chain* (36) *of type I and arbitrary length t in which $Y_1 = Y_2 = \cdots = Y_t$, $X_2 = X_3 = \cdots = X_t$, Y_1 is nilpotent, and when $p \neq 2$, X_1 and X_2 have trace zero (hence are commutators).*

Proof. If C supports a commutator chain of type I, Theorem 2 applies to C and hence C satisfies the condition of Theorem 2. Conversely, if C satisfies the condition of Theorem 2, then by Theorem 3, $C = (A^+, B^+)$ with $(C, A^+) = 0$, with A^+ nilpotent, and (for $p \neq 2$) with trace $B^+ = 0$. If $X_1 = -B^+$, $Y_1 = A^+$, then $C = (X_1, Y_1)$, and $(C, Y_1) = 0$. Since Y_1 is nilpotent, Theorem 4 shows $Y_1 = (X_2, Y_1)$ for some X_2 which has trace zero when $p \neq 2$. Iterating this last identity produces the desired commutator chain for C. This also proves Theorem 13.

THEOREM 13. *If $C \in M_n(K)$ is nilpotent then C supports a commutator chain* (36) *of type I of arbitrary length t within $M_n(K)$ in which $C = Y_1 = \cdots = Y_t$, $X_1 = \cdots = X_t$, and (for $p \neq 2$) trace $X_1 = 0$.*

Suppose now $C \in M_n(K)$ supports a commutator chain (37) of type II. Then Theorem 9 applies and so the elementary divisors of C over L classify into the types (20), (21), (22) of Theorem 9. In the next theorems we provide sufficient conditions on the elementary divisors of C over K in order that C support a commutator chain of type II of arbitrary length t.

THEOREM 14. *Let $C \in M_n(K)$. Suppose the elementary divisors of C are given by* (20). *Then C supports a chain of commutators* (37) *of type II of arbitrary length t within $M_n(K)$ in which $C, Y_1, Y_2, \cdots, Y_t$ are similar elements of $M_n(K)$ and $X_1, X_2, \cdots, X_t$ are similar elements of $M_n(K)$. In addition, if $p = 0$ or $p \geqq r$ and if the multiplicity of integer i in $e_1, e_2, \cdots, e_r$ is monotone nonincreasing as i increases, we may also have X_1 similar to C.*

Proof. In the proof of Theorem 9 we found certain $A, B \in M_n(K)$ such that (19) holds. Corollary 1 may be used to show that B is similar to C. So set $X_1 = A$, $Y_1 = B$. Then the elementary divisors of Y_1 satisfy (20) of Theorem 9, so $Y_1 = (X_2, Y_2)$ with $(Y_1, X_2) = (Y_1, Y_2) = 0$, and with Y_2 similar to Y_1. Moreover X_2 depends on the elementary divisors of Y_1 in the same way as X_1 depends on the elementary divisors of C. Hence X_2 is similar to X_1. An iteration now produces the desired chain.

Let $m(i)$ be the multiplicity of integer i in $e_1, e_2, \cdots, e_r$. In the second assertion of the theorem, we know $m(1) \geqq m(2) \geqq \cdots \geqq m(e_r)$. Choose integers $k_1 < k_2 < \cdots < k_d$ so that

$$\begin{aligned} m(1) = m(2) \cdots = m(k_1) > m(k_1 + 1) = m(k_1 + 2) = \cdots \\ = m(k_2) > m(k_2 + 1) = \cdots > \cdots > m(k_{d-1} + 1) = \cdots = m(k_d)\,, \end{aligned}$$

where $k_d = e_r$. Let C_i be a matrix with elementary divisors

$$\lambda, \lambda^2, \lambda^3, \cdots, \lambda^{k_i} \tag{38}$$

for $1 \leqq i \leqq d$. Then C is similar to

$$\operatorname{diag}(C_1, C_1, \cdots, C_1, C_2, C_2, \cdots, C_2, \cdots, C_d, C_d, \cdots, C_d)$$

where C_d appears $m(k_d)$ times and C_i appears $m(k_i) - m(k_{i+1})$ times, $1 \leqq i < d$. If $p \geqq r$ then from $r \geqq e_r = k_d \geqq k_i$, we have $p \geqq k_i$ in (38) for $1 \leqq i \leqq d$. If we can show that C_i supports a commutator chain (37) with $Y_0 = C_i$ and with all X_i and all Y_i similar to C_i, then by taking direct sums we will also get C supporting a commutator chain (37) with $Y_0 = C$ and all X_i and all Y_i similar to C.

Hence to complete the proof of Theorem 14, we may suppose that the elementary divisors of C are (20) with $e_i = i$ for $1 \leqq i \leqq r$, and $p = 0$ or $p \geqq r$. In this case the proof above gives $C = (A, B)$ where the blocks $A_{\alpha,\alpha-1}$ of A, given by (28), have $r + 1 - \alpha \neq 0$, $2 \leqq \alpha \leqq r$. We may use Lemma 3 to compute that the elementary divisors of A are $\lambda, \lambda^2, \cdots, \lambda^r$. Hence A is similar to C. This completes the proof.

COROLLARY 2. *In Theorem* 14, *if the additional conditions are*

satisfied, each matrix X_i, Y_i *appearing in the chain* (37) *of type II supported by* C *also supports a commutator chain of type II, and each matrix appearing in any of these chains also supports a commutator chain of type II, etc. We may construct these chains so that any matrix appearing in any of these chains is similar to* C *in* $M_n(K)$.

THEOREM 15. *Let* $p \neq 0$. *Let* $C \in M_n(K)$ *have elementary divisors* (21). *Then* C *supports a commutator chain* (37) *of type II and arbitrary length* t *within* $M_n(K)$ *in which* $Y_1, Y_2, Y_3, \cdots, Y_t$ *are similar elements of* $M_n(K)$ *and* $X_2, X_3, \cdots, X_t$ *are similar elements of* $M_n(K)$.

Proof. In the proof of Theorem 9, we expressed C in the form (19) for a certain $A, B \in M_n(K)$. The elementary divisors of B may be calculated from Lemma 2, and they turn out to satisfy the conditions (20) of Theorem 9. Hence if we set $X_1 = A$, $Y_1 = B$, then Theorem 14 applied to Y_1 produces the rest of the chain.

THEOREM 16. *Let* $p \neq 0$. *Let* $g(\lambda)$ *be a polynomial over* K *of degree* m *with* $m \equiv 0 \pmod{p}$. *Let* $C = C(g(\lambda)) + \cdots + C(g(\lambda))$ *where* $C(g(\lambda))$ *appears* r *times with* $r \equiv 0 \pmod{p}$. *Then* C *supports a commutator chain* (37) *of type II and arbitrary length* t *within* $M_{mr}(K)$ *such that* X_2 *and* Y_1 *are similar elements of* $M_{mr}(K)$ *and such that* $Y_2, Y_3, \cdots, Y_t, X_3, X_4, \cdots, X_t$ *are all similar in* $M_{mr}(K)$. *If* $g(\lambda)$ *has nonzero constant term then we also have* X_1 *similar to* X_2.

Proof. By taking direct sums it suffices to consider the case $r = p$. Let $A_{\alpha,\alpha-1} = (p + 1 - \alpha)\, C(g(\lambda))$ for $2 \leqq \alpha \leqq p$ and let $B_{\alpha,\alpha+1} = I_m$ for $1 \leqq \alpha < p$. Let $A = (A_{\alpha\beta})_{1\leqq\alpha,\beta\leqq p}$ and $B = (B_{\alpha\beta})_{1\leqq\alpha,\beta\leqq p}$ where $A_{\alpha\beta}$ and $B_{\alpha\beta}$ are all m-square, where $A_{\alpha\beta} = 0$ if $\beta \neq \alpha - 1$, and where $B_{\alpha\beta} = 0$ if $\beta \neq \alpha + 1$. Then $C = (A, B)$, $(C, A) = (C, B) = 0$. The elementary divisors of B can be found from Lemma 4; they are λ^p with multiplicity m. If $g(\lambda)$ has nonzero constant term, the elementary divisors of A can also be found from Lemma 4; they are also λ^p with multiplicity m. Let $X_1 = A$, $Y_1 = B$. Let C_1 be p^2-square with elementary divisors λ^p with multiplicity p. Matrix B is similar to the direct sum of m/p copies of C_1. It will therefore suffice to construct the rest of the chain for C_1. The elementary divisors of C_1 fall into classification (21) of Theorem 9, with $r = p$ and $e_1 = e_2 = \cdots = e_p = p$. So, as in the proof of Theorem 9, we express $C_1 = (A_1, B_1)$ with $(C_1, A_1) = (C_1, B_1) = 0$. The elementary divisors of B_1 can be computed from Lemma 2; they turn out to be $\lambda, \lambda^2, \lambda^3, \cdots, \lambda^{p-1}$,

each with multiplicity two, and λ^p with multiplicity one. The elementary divisors of A_1 can be computed from Lemma 3; they turn out to be λ^p with multiplicity p. Since B_1 satisfies all the conditions in Theorem 14, the existence of the rest of the commutator chain now follows from Theorem 14. This finishes the proof.

We remark that by varying this general theme, many other commutator chains may be constructed involving matrices from a small number of similarity classes. Let the reader experiment for himself!

We now present two corollaries which summarize some of our results and which show interesting parallels between commutator chains of types I and II.

COROLLARY 3. *Let $p = 0$ or $p > n$. Let $C \in M_n(K)$. Then each of the following statements* (a), (b), (c), (d), (e) *implies all of the others.*

(a) *C supports a commutator chain* (36) *of type* I *within $M_n(K)$ of some length t;*

(b) *C supports a commutator chain* (36) *of type* I *within $M_n(K)$ of arbitrary length t;*

(c) *C supports a commutator chain* (36) *of type* I *within $M_n(K)$ of arbitrary length t in which $C = Y_1 = \cdots = Y_t$ and $X_1 = \cdots = X_t$;*

(d) *C is nilpotent;*

(e) *C is a sum of commutators* (3) *within $M_n(K)$.*

COROLLARY 4. *Let $p = 0$ or $p > n$. Let $C \in M_n(K)$. Then each of the following statements* (a), (b), (c), (d), (e) *implies all of the others.*

(a) *C supports a commutator chain* (37) *of type* II *within $M_n(K)$ of some length t;*

(b) *C supports a commutator chain* (37) *of type* II *within $M_n(K)$ of arbitrary length t;*

(c) *C supports a commutator chain* (37) *of type* II *within $M_n(K)$ of arbitrary length t in which $Y_1, Y_2, \cdots, Y_t$ are all similar to C, and $X_1, X_2, \cdots, X_t$ are all similar;*

(d) *C is nilpotent with elementary divisors in the form* (20);

(e) *C is a sum of commutators* (9) *within $M_n(K)$.*

8. Construction of block stripe matrices with prescribed eigenvalue. Let C be presented in partitioned form as $C = (C_{ij})_{1 \leqq i,j \leqq k}$, where C_{ij} is an $e_i \times e_j$ stripe matrix, $1 \leqq i, j \leqq k$. Because of Theorem 2, it is of interest to determine when such a matrix can have a single nonzero eigenvalue. We show that matrices C with these prpoerties are extremely plentiful, and we construct them all. In the process

of this, we use a theorem (Theorem 17) due to Williamson, and in Theorem 18 we develop what amounts to an extension of Williamson's result. Let

$$e_1 \leqq e_2 \leqq \cdots \leqq e_k \tag{39}$$

be any nondecreasing sequence of positive integers and let

$$f_1 < f_2 < \cdots < f_s \tag{40}$$

be the distinct integers in (39), where f_i appears with multiplicity m_i in (39), $1 \leqq i \leqq s$.

Theorems 17 and 18 deal with matrices more general than the matrix C above in which we are interested.

THEOREM 17. *For $1 \leqq i, j \leqq r$, let V_{ij} be an e-square triangular matrix (zeros below the main diagonal) and let* $\operatorname{diag}(v_{ij1}, v_{ij2}, \cdots, v_{ije})$ *be the main diagonal of V_{ij}. Let $V_\alpha = (v_{ij\alpha})_{1\leqq i,j\leqq r}$, and let $V = (V_{ij})_{1\leqq i,j\leqq r}$. Then V and* $\operatorname{diag}(V_1, V_2, \cdots, V_e)$ *have the same eigenvalues.*

Proof. In [6, Theorem 1], a similarity transformation of V is exhibited which converts V into a block triangular form with blocks $V_1, V_2, \cdots, V_e$ on the main block diagonal.

Note that, because of Theorem 17, when evaluating the eigenvalues of V, one may replace all elements strictly above the main diagonal of V_{ij} with zeros, $1 \leqq i, j \leqq r$.

THEOREM 18. *Let matrix M_{ij} be $e_i \times e_j$ such that*: (i) *if $e_i \geqq e_j$, the (α, β) entry of M_{ij} is zero whenever $\alpha > \beta$*; (ii) *if $e_i < e_j$, the (α, β) entry of M_{ij} is zero whenever $\alpha \geqq \beta$; $1 \leqq i, j \leqq k$. Let $M = (M_{ij})_{1\leqq i,j\leqq k}$. Let $N_{ij} = M_{ij}$ if $e_i = e_j$, and let N_{ij} be the $e_i \times e_j$ zero matrix if $e_i \neq e_j$; $1 \leqq i, j \leqq k$. Let $N = (N_{ij})_{1\leqq i,j\leqq k}$. Then M and N have the same eigenvalues.*

Proof. The theorem says that in evaluating the eigenvalues of M, the nonsquare blocks don't matter. We prove the theorem by induction on the integer s in (40). If $s = 1$, all blocks in M are square and so $M = N$. Now let $s > 1$. Find integers a, $b(a < b)$ such that $e_\alpha < f_{s-1}$ if $\alpha \leqq a$; $e_\alpha = f_{s-1}$ if $a < \alpha \leqq a + b$; and $e_\alpha = f_s$ if $a + b < \alpha \leqq k$. If $s = 2$, then $a = 0$ and there are no $e_\alpha < f_{s-1}$. We now partition certain of the blocks in M. We let

$$M_{\alpha\beta} = \begin{bmatrix} M'_{\alpha\beta} \\ 0_{f_s - f_{s-1}, e_\beta} \end{bmatrix}$$

for $\alpha > a + b$ and $1 \leqq \beta \leqq a + b$;

$$M_{\alpha\beta} = (M'_{\alpha\beta}, M''_{\alpha\beta})$$

where $M'_{\alpha\beta}$ is $e_\alpha \times f_{s-1}$ and $M''_{\alpha\beta}$ is $e_\alpha \times (f_s - f_{s-1})$, for $1 \leqq \alpha \leqq a + b$ and $\beta > a + b$;

$$M_{\alpha\beta} = \begin{bmatrix} M'_{\alpha\beta} & M''_{\alpha\beta} \\ 0_{f_s - f_{s-1}, f_{s-1}} & M'''_{\alpha\beta} \end{bmatrix}$$

for $\alpha > a + b$ and $\beta > a + b$.

After an interchange of rows in M and then the same interchange of columns in the result (a similarity transformation), we transform M into $\tilde{M}$ where (let $\nu = a + b$):

$$\tilde{M} = \begin{bmatrix} M_{11} & \cdots & M_{1\nu} & M'_{1,\nu+1} & \cdots & M'_{1k} & M''_{1,\nu+1} & \cdots & M''_{1k} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ M_{\nu 1} & \cdots & M_{\nu\nu} & M'_{\nu,\nu+1} & \cdots & M'_{\nu k} & M''_{\nu,\nu+1} & \cdots & M''_{\nu k} \\ M'_{\nu+1,1} & \cdots & M'_{\nu+1,\nu} & M'_{\nu+1,\nu+1} & \cdots & M'_{\nu+1,k} & M''_{\nu+1,\nu+1} & \cdots & M''_{\nu+1,k} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ M'_{k1} & \cdots & M'_{k\nu} & M'_{k,\nu+1} & \cdots & M'_{kk} & M''_{k,\nu+1} & \cdots & M''_{kk} \\ 0 & \cdots & 0 & 0 & & 0 & M'''_{\nu+1,\nu+1} & \cdots & M'''_{\nu+1,k} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & & 0 & M'''_{k,\nu+1} & \cdots & M'''_{kk} \end{bmatrix}$$

Thus the eigenvalues of M are the eigenvalues of

$$\begin{bmatrix} M_{11} & \cdots & M'_{1k} \\ \vdots & & \vdots \\ M'_{k1} & \cdots & M'_{kk} \end{bmatrix}, \tag{41}$$

together with the eigenvalues of

$$\begin{bmatrix} M'''_{\nu+1,\nu+1} & \cdots & M'''_{\nu+1,k} \\ \vdots & & \vdots \\ M'''_{k,\nu+1} & \cdots & M'''_{kk} \end{bmatrix}. \tag{42}$$

However, the induction hypothesis applies to (41), since the blocks in (41) have dimensions $f_1 \times f_1, \cdots, f_{s-1} \times f_{s-1}$. Therefore in evaluating the eigenvalues of (41) we may replace all rectangular blocks in (41) with zero blocks. Using the induction hypothesis for $s - 2$, we get that the eigenvalues of M are the eigenvalues of (43), together with the eigenvalues of (44) and (42):

$$\begin{bmatrix} M_{11} & \cdots & M_{1a} \\ \vdots & & \vdots \\ M_{a1} & \cdots & M_{aa} \end{bmatrix} \tag{43}$$

$$(44)\qquad \begin{bmatrix} M_{a+1,a+1} & \cdots & M_{a+1,\nu} & M'_{a+1,\nu+1} & \cdots & M'_{a+1,k} \\ \vdots & & \vdots & \vdots & & \vdots \\ M_{\nu,a+1} & \cdots & M_{\nu\nu} & M'_{\nu,\nu+1} & \cdots & M'_{\nu k} \\ M'_{\nu+1,a+1} & \cdots & M'_{\nu+1,\nu} & M'_{\nu+1,\nu+1} & \cdots & M'_{\nu+1,k} \\ \vdots & & \vdots & \vdots & & \vdots \\ M'_{k,a+1} & \cdots & M'_{k\nu} & M'_{k,\nu+1} & \cdots & M'_{kk} \end{bmatrix}.$$

This is because the blocks in (44) are all the f_{s-1}-square blocks in (41). By Theorem 17, we can replace all elements above the main diagonal in each block of (44) with zeros. Because the main diagonal of each block M'_{ij}, $a < i \leqq \nu$, $\nu < j \leqq k$, is already zero, it follows that the eigenvalues of (44) are the eigenvalues of (45) together with the eigenvalues of (46).

$$(45)\qquad \begin{bmatrix} M_{a+1,a+1} & \cdots & M_{a+1,\nu} \\ \vdots & & \vdots \\ M_{\nu,a+1} & \cdots & M_{\nu\nu} \end{bmatrix},$$

$$(46)\qquad \begin{bmatrix} M'_{\nu+1,\nu+1} & \cdots & M'_{\nu+1,k} \\ \vdots & & \vdots \\ M'_{k,\nu+1} & \cdots & M'_{kk} \end{bmatrix}.$$

Now, by Theorem 18 for $s-2$, the eigenvalues of N are the eigenvalues of the direct sum of (43), (45), and (47).

$$(47)\qquad \begin{bmatrix} M_{\nu+1,\nu+1} & \cdots & M_{\nu+1,k} \\ \vdots & & \vdots \\ M_{k,\nu+1} & \cdots & M_{kk} \end{bmatrix}.$$

But the same row and column interchanges that converted M into $\tilde{M}$ converts (47) into a matrix of the form

$$\begin{bmatrix} X & Y \\ 0 & Z \end{bmatrix}$$

where X is (46) and Z is (42). Therefore the eigenvalues of (42) and (46) together are the eigenvalues of (47). This completes the proof.

It follows from Theorems 17 and 18 that all $C = (C_{ij})_{1\leqq i,j\leqq k}$ (where each C_{ij} is an $e_i \times e_j$ stripe matrix) for which C has a single eigenvalue γ can be constructed as follows. To construct the main diagonals of the f_i-square blocks in C, select any m_i-square nilpotent matrix E_i and let $F_i = \gamma I_{m_i} + E_i$. Let $F_i = (f_{\alpha\beta})_{1\leqq\alpha,\beta\leqq m_i}$. Let $a = f_1m_1 + \cdots + f_{i-1}m_{i-1}$. Then the blocks in C which are to be f_i-square

are $C_{\alpha+\alpha,\alpha+\beta}$ for $1 \leqq \alpha, \beta \leqq m_i$. Let the diagonal element of $C_{\alpha+\alpha,\alpha+\beta}$ be $f_{\alpha\beta}$, for $1 \leqq \alpha, \beta \leqq m_i$. Do this for $1 \leqq i \leqq s$. Choose arbitrary values for the remaining stripes in the blocks of C. The C has γ as its only eigenvalue.

Added in proof. Theorem 18 appears to overlap the result contained in the corollary on p. 240 of the paper *On the characteristic values of the matrix* $f(A, B)$, by W. E. Roth, Trans. Amer. Math. Soc., vol. **39** (1936), 234–243.

References

1. A. A. Albert and B. Muckenhaupt, *On matrices of trace zero*, Michigan Math. J. **4** (1957), 1-3.
2. F. R. Gantmacher, *The theory of matrices*, Vol. 1, Chelsea, New York, 1959, p. 157.
3. C. C. MacDuffee, *The theory of matrices*, Chelsea, New York, 1956, p. 90.
4. N. H. McCoy, *On quasi-commutative matrices*, Trans. Amer. Math. Soc. **36** (1934), 327-340.
5. R. M. Thrall and L. Tornheim, *Vector spaces and matrices*, Wiley, New York, 1958, p. 252.
6. J. Williamson, *The latent roots of a matrix of special type*, Bull. Amer. Math. Soc. **37** (1931), 585-590.

Received December 2, 1965. The preparation of this paper was supported in part by the U. S. Air Force.

University of California, Santa Barbara

PACIFIC JOURNAL OF MATHEMATICS

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should be typewritten (double spaced). The first paragraph or two must be capable of being used separately as a synopsis of the entire paper. It should not contain references to the bibliography. Manuscripts may be sent to any one of the four editors. All other communications to the editors should be addressed to the managing editor, Richard Arens at the University of California, Los Angeles, California 90024.

50 reprints per author of each article are furnished free of charge; additional copies may be obtained at cost in multiples of 50.

The *Pacific Journal of Mathematics* is published monthly. Effective with Volume 16 the price per volume (3 numbers) is $8.00; single issues, $3.00. Special price for current issues to individual faculty members of supporting institutions and to individual members of the American Mathematical Society: $4.00 per volume; single issues $1.50. Back numbers are available.

Subscriptions, orders for back numbers, and changes of address should be sent to Pacific Journal of Mathematics, 103 Highland Boulevard, Berkeley 8, California.

Printed at Kokusai Bunken Insatsusha (International Academic Printing Co., Ltd.), No. 6, 2-chome, Fujimi-cho, Chiyoda-ku, Tokyo, Japan.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS, A NON-PROFIT CORPORATION

The Supporting Institutions listed above contribute to the cost of publication of this Journal, but they are not owners or publishers and have no responsibility for its content or policies.

Pacific Journal of Mathematics
Vol. 20, No. 2 October, 1967