Roth's Removal Rule Revisited

Robert E. Hartwig
Mathematics Department
North Carolina State University
Raleigh, North Carolina 27650

Submitted by Richard A. Brualdi

ABSTRACT

The theory of companion matrices is used to give explicit representations for the matrices needed in Roth's removal rule. These are then used to give simple proofs for the cyclic decomposition theorem, as well as for Roth's similarity theorem for matrices over a field.

1. INTRODUCTION

In his fundamental paper [12] Roth introduced his removal rule, which states that one may remove the matrix C in the block matrix $\begin{bmatrix} A & C \\ 0 & D \end{bmatrix}$ by a *similarity* transformation, provided that the matrix equation $AX - XD = C$ has a solution X. In this case it is easily seen that the matrix $\begin{bmatrix} I & X \\ 0 & I \end{bmatrix}$ will do. He went on to prove his similarity theorem, which states that for matrices over a field the converse is also true.

Perhaps the most important interpretation of this removal rule is in terms of invariant subspaces. Indeed, the rule says that given a linear map T on a vector space, V, and an invariant subspace W_1 with complementary subspace W_2, then one may find an *invariant* complementary subspace W_3, provided the matrix of T restricted to W_1 and the matrix of the induced map on V/W_1 satisfy the consistency condition $AX - XD = C$. What has not been brought out before is the fact that the removal rule actually tells you where and how to find the subspace W_3. Consequently it is of considerable importance in the study of canonical forms and matrix theory in general [8]. In this paper we shall develop the removal rule for companion matrices, and apply this to give an easy algorithmic proof for the cyclic decomposition theorem (CDT; also called the rational canonical form theorem). It in essence lays bare what makes the CDT tick. We shall then examine the semisimple case, and

52 Vanderbilt Ave., New York, NY 10017 0024-3795/83/010091-25$3.00

conclude by using the removal rule to give a constructive proof of Roth's similarity theorem. Roth's similarity theorem has been studied by a number of authors [1–4, 12], and has been proven under quite general conditions, such as for matrices over a commutative ring and beyond. For example see Theorem 4.2 in [4]. Most proofs however are existence proofs, using dimensional analysis, module theory, or λ-matrix theory, and do not furnish a tractable solution to the equation $AX - XD = C$. The removal rule for companion matrices generalizes a result due to Feinberg [1], and its application simplifies considerably the proof of Roth's theorem given in this paper. In particular, we shall not need to use determinantal divisors, but merely use some simple properties of rank, of annihilating polynomials, and of companion matrices. In addition, the removal rule gives a much more tractable solution to the desired matrix equation. Throughout this paper, all our matrices will be over a field $\mathfrak{F}$. We shall use $\mathfrak{F}_{m\times n}$ to denote the set of $m\times n$ matrices over $\mathfrak{F}$, and use $\mathfrak{F}^n$ for $\mathfrak{F}_{n\times 1}$. For a matrix $A\in\mathfrak{F}_{n\times n}$, we shall denote the range, the nulspace, the rank, the characteristic polynomial, and the minimal polynomial by $R(A)$, $N(A)$, $\rho(A)$, $\Delta_A(\lambda)$, and $\psi_A(\lambda)$ respectively. For a vector $\mathbf{x}\in\mathfrak{F}^n$, we denote its minimal polynomial with respect to A by $\psi_{\mathbf{x}}(\lambda)$. We say that $\mathbf{x}$ *belongs to* ψ_A if $\psi_{\mathbf{x}}(\lambda)=\psi_A(\lambda)$. As always, we shall denote similarity by $\approx$ and use $\otimes$ to denote the right direct product.

For a matrix $X=[\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n]$ we shall write $\mathrm{col}(X)=[\mathbf{x}_1^T,\dots,\mathbf{x}_n^T]^T$ and shall freely use the column lemma,

$$\mathrm{col}(AXB)=(B^T\otimes A)\,\mathrm{col}(X).$$

For a monic polynomial $f(\lambda)=f_0+f_1\lambda+\cdots+\lambda^k$, let

$$L[f(\lambda)] = \begin{bmatrix} 0 & & & -f_0 \\ 1 & & & -f_1 \\ & \ddots & & \vdots \\ 0 & & 1 & -f_{k-1} \end{bmatrix} \quad S[f(\lambda)] = \begin{bmatrix} f_1 & f_2 & \cdot & \cdot & \cdot & f_k \\ f_2 & & & & \cdot & \\ \vdots & & & \cdot & & \\ \cdot & & \cdot & & & 0 \\ \cdot & \cdot & & & & \\ f_k & & & & & \end{bmatrix}$$

and define

$$\mathfrak{f}_i(\lambda)=f_{i+1}+f_{i+2}\lambda+\cdots+f_k\lambda^{k-i-1}, \qquad i=0,1,\dots,k-1. \quad (1.1)$$

It is well known and easily verified that

$$LS = SL^T \tag{1.2}$$

and that for any vector $\mathbf{x} = [x_0, x_1, \dots, x_{k-1}]^T$,

$$[\mathbf{x}, L\mathbf{x}, \dots, L^{k-1}\mathbf{x}] = x(L), \tag{1.3}$$

where $x(\lambda) = x_0 + x_1\lambda + \cdots + x_{k-1}\lambda^{k-1}$.

This identity gives rise to the following surprising result:

$$\mathrm{col}[I, L, L^2, \dots, L^{k-1}] = \mathrm{col}\begin{bmatrix} I \\ L \\ L^2 \\ \vdots \\ L^{k-1} \end{bmatrix}. \tag{1.4}$$

We shall also need the fact that

$$(S \otimes I)\,\mathrm{col}(I) = (I \otimes S)\,\mathrm{col}(I) \tag{1.5}$$

Now if $f(D) = 0$, then by the division algorithm,

$$(\lambda I - D)\left(\sum_{i=0}^{k-1} D_i \lambda^i\right) = f(\lambda) I,$$

where

$$D_i = \mathfrak{f}_i(D) = \sum_{j=0}^{k-i-1} f_{i+j+1} D^j, \qquad i = 0, 1, \dots, k-1. \tag{1.6}$$

In addition we have the chain conditions [5, p. 105]

$$\begin{aligned} D[I, D, \dots, D^{k-1}] &= [I, D, \dots, D^{k-1}](L[f(\lambda)] \otimes I) \\ &= [I, D, \dots, D^{k-1}](I \otimes D), \end{aligned} \tag{1.7}$$

as well as

$$[D_0, D_1, \dots, D_{k-1}] = [I, D, \dots, D^{k-1}](S[f(\lambda)] \otimes I). \tag{1.8}$$

Lastly, we shall also make use of the fact that both the consistency of $AX - XD = C$ and the similarity of

$$\begin{bmatrix} A & C \\ 0 & D \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}$$

remain invariant under a similarity transformation applied to A and/or D, with an appropriate modification of C.

2. THE MINIMAL POLYNOMIAL

In this section we shall derive several elementary results dealing with the minimal polynomial of a block triangular matrix. These will then be used in Section 7.

LEMMA 1. *Let*

$$M = \begin{bmatrix} A_1 & & & \\ & A_2 & ? & \\ & & \ddots & \\ & 0 & & \\ & & & A_k \end{bmatrix}, \qquad N = \begin{bmatrix} A_1 & & & \\ & A_2 & 0 & \\ & 0 & \ddots & \\ & & & A_k \end{bmatrix},$$

and

$$S_{ij} = \begin{bmatrix} A_i & & & \\ & A_{i+1} & ? & \\ & 0 & \ddots & \\ & & & A_j \end{bmatrix}, \qquad T_{ij} = \begin{bmatrix} A_i & & & \\ & A_{i+1} & 0 & \\ & 0 & \ddots & \\ & & & A_i \end{bmatrix}$$

be sections of M and N respectively. Then

$$\operatorname{lcm}(\psi_{A_1}, \dots, \psi_{A_k}) \mid \psi_M(\lambda) \mid \prod_{i=1}^{k} \psi_{A_i}(\lambda), \tag{2.1a}$$

$$\psi_N(\lambda) = \operatorname{lcm}(\psi_{A_1}, \dots, \psi_{A_k}), \tag{2.1b}$$

If $M \approx N$ *then* $\psi_{S_{ij}}(\lambda) = \psi_{T_{ij}}(\lambda) = \operatorname{lcm}(\psi_r : i \leqslant r \leqslant j)$. (2.1c)

Proof. (a), (b): These are left as an exercise, and are proven in [8].

(c) First note that $\rho(M) \geqslant \rho(N) = \sum_{i=1}^{k} \rho(A_i)$. Now let $p(\lambda) = \psi_{T_{ij}}(\lambda)$. Then by (a), $p \mid \psi_{S_{ij}}$. On the other hand,

$$p(M) = \begin{bmatrix} p(A_1) & & & ? & \\ & \ddots & & & \\ & & p(S_{ij}) & & \\ & 0 & & \ddots & \\ & & & & p(A_k) \end{bmatrix}$$

$$\approx p(N) = \begin{bmatrix} p(A_1) & & & 0 & \\ & \ddots & & & \\ & & p(N_{ij}) & & \\ & 0 & & \ddots & \\ & & & & p(A_k) \end{bmatrix},$$

with $p(N_{ij}) = 0$. Thus

$$\rho(p(N)) = \sum_{\substack{t<i\\t>j}} \rho[p(A_t)] = \rho(p(M)) \geqslant \sum_{\substack{t<i\\t>j}} \rho[p(A_k)] + \rho[p(S_{ij})]$$

and so $p(S_{ij}) = 0$ and $\psi_{S_{ij}} \mid p$ as desired. ■

An immediate consequence is the following, which we shall use in constructing the solution to Roth's similarity problem.

Lemma 2. *Let*

$$M = \left[\begin{array}{ccc|ccc} A_1 & & 0 & & & \\ & \ddots & & & C & \\ 0 & & A_m & & & \\ \hline & & & D_1 & & 0 \\ & 0 & & & \ddots & \\ & & & 0 & & D_n \end{array}\right],$$

$$N=\begin{bmatrix} A_1 & & & & & 0 \\ & \ddots & & & & \\ & & A_m & & & \\ & & & D_1 & & \\ & 0 & & & \ddots & \\ & & & & & D_n \end{bmatrix}$$

with $C=[C_{ij}]$ *partitioned conformally. Also let*

$$S_{ij}=\left[\begin{array}{ccc|ccc} A_i & & 0 & C_{i1} & \cdots & C_{ij} \\ & \ddots & & \vdots & & \vdots \\ 0 & & A_m & C_{m1} & \cdots & C_{mj} \\ \hline & & & D_1 & & 0 \\ & 0 & & & \ddots & \\ & & & 0 & & D_j \end{array}\right] \quad \text{and} \quad W_{ij}=\begin{bmatrix} A_i & C_{ij} \\ 0 & D_j \end{bmatrix}.$$

If $M \approx N$, *then*

$$\operatorname{lcm}\left(\psi_{A_i},\psi_{D_j}\right) \mid \psi_{W_{ij}} \mid \operatorname{lcm}\left(\psi_{A_t}: t \geqslant i,\ \psi_{D_u}: u \leqslant j\right).$$

Proof. By (2.1a), the first part is clear. Also by (2.1c), $\psi_{S_{ij}} = \operatorname{lcm}(\psi_{A_t}: t \geqslant i, \psi_{D_u}: u \leqslant j)$. Next, for any polynomial $q(\lambda)$,

$$q\left(S_{ij}\right)=\begin{bmatrix} q(A_i) & \hat{C}_{ij} \\ 0 & q(D_j) \end{bmatrix},$$

where $\hat{C}_{ij}$ is exactly the (1,2) block in $q(W_{ij})$. Hence $q(S_{ij})=0 \Rightarrow q(W_{ij})=0$ and $\psi_{W_{ij}} \mid \psi_{S_{ij}}$, completing the proof. ■

LEMMA 3. *If*

$$\begin{bmatrix} A & C \\ 0 & D \end{bmatrix} \approx \begin{bmatrix} A & 0 \\ 0 & E \end{bmatrix},$$

then $\psi_D \mid \psi_E$.

Proof.

$$\rho\begin{bmatrix}\psi_E(A) & ? \\ 0 & \psi_E(D)\end{bmatrix} = \rho\begin{bmatrix}\psi_E(A) & 0 \\ 0 & 0\end{bmatrix} \quad \Rightarrow \quad \psi_E(D)=0 \quad \Rightarrow \quad \psi_D \mid \psi_E.$$

If in addition, $C=0$, then by symmetry $\psi_D = \psi_E$. ■

3. SOME PRELIMINARY RESULTS

Let

$$M=\begin{bmatrix}A & C \\ 0 & D\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}A & 0 \\ 0 & D\end{bmatrix},$$

and suppose that $f(\lambda)=f_0+f_1\lambda+\cdots+\lambda^k$ is a monic polynomial over $\mathcal{F}$. Then it is easily seen that

$$f(M)=\begin{bmatrix}f(A) & \tilde{C} \\ 0 & f(D)\end{bmatrix},$$

where

$$\begin{aligned}\tilde{C}=\mathcal{K}(C)&=f_1C+f_2(AC+CD)+\cdots\\&\quad+\left(A^{k-1}C+A^{k-2}CD+\cdots+CD^{k-1}\right)\\&=\left[I,A,\dots,A^{k-1}\right](S\otimes C)\begin{bmatrix}I\\D\\ \vdots \\ D^{k-1}\end{bmatrix}.\end{aligned} \tag{3.1}$$

Now let $\mathcal{G}(X)=AX-XD$. Clearly both $\mathcal{G}$ and $\mathcal{K}$ are linear maps on $\mathcal{F}_{n\times m}$. Our first observation is that when f annihilates A and D, then the range of $\mathcal{G}$ must be contained in the nulspace of $\mathcal{K}$.

LEMMA 4. *If $f(A)=0$, $f(D)=0$, then $R(\mathcal{G}) \subseteq N(\mathcal{H})$.*

Proof. Clearly $f(M)=0 \Leftrightarrow f(A)=0$, $f(D)=0$ and $\tilde{C}=0$. Now if $C = AX - XD$ and $f(A)=0$, $f(D)=0$, then by telescoping induction, $\tilde{C}= \mathcal{H}(C)=(f_1 A + f_2 A^2 + \cdots + A^k)-(f_1 D + \cdots + D^k) = -f_0 X + f_0 X = 0$, as desired. Alternatively, one could use (3.1) combined with (1.2) and (1.7). ■

In general, one cannot expect equality to hold, since every matrix M has annihilating polynomials. Thus the fundamental question is, when does equality hold? Before we examine this question, let us observe that the linear map $\mathcal{H}$ is very closely related to the block blinear map [6]

$$\tau(Y)=\left[I, A,\dots,A^{k-1}\right](Y\otimes C)\begin{bmatrix} I \\ D \\ \vdots \\ D^{k-1}\end{bmatrix}. \tag{3.2}$$

Using the shift condition (1.6) and applying (1.2), we see that

$$A\tau(Y)-\tau(Y)D=\mathcal{Q}\left[(LY-YL^T)\otimes C\right]\mathcal{D}, \tag{3.3}$$

where $\mathcal{Q}=[I, A,\dots,A^{k-1}]$, $\mathcal{D}=[I, D^T,\dots,(D^T)^{k-1}]^T$, and $L=L[f(\lambda)]$. Hence it would be natural to look for a solution to

$$LY-YL^T=E_{11}, \tag{3.4}$$

since this would give $A\tau(Y)-\tau(Y)D=C$. Unfortunately (3.4) will be inconsistent is general. If we now use the fact that $f(M)=0$, and that $C=\mathcal{Q}(S\otimes C)\mathcal{D}=0$, then by (1.7) we have

$$p(A)\tilde{C}=\mathcal{Q}\left[p(L)S\otimes C\right]\mathcal{D}=0$$

for any polynomial $p(\lambda)$. It would thus be sufficient to find a polynomial $p(\lambda)$ for which

$$LY-YL^T=E_{11}+p(L)S \tag{3.5}$$

is consistent. Using (1.2) we may rewrite this as

$$LZ-ZL=E_{1,n}+p(L), \tag{3.6}$$

where $Z = YS^{-1}$. Again this will not be consistent in general, but stands a much better chance than (3.4). We shall take up this question in Section 6. Let us now show that if A is nonderogatory (i.e., $\psi_A = \Delta_A$) and $\psi_M = \psi_A$, then C may be removed by similarity.

4. THE REMOVAL RULE FOR COMPANION MATRICES

If $A = L[f(\lambda)]$ is a companion matrix, then unlike the invariant subspace notation, the direct sum notation is simplest when applied to *lower* block triangular matrices. We will obtain the desired upper triangular results by transposition. This is easier than using a permutation matrix to interchange matrices in the direct product notation.

THEOREM 1. *Let*

$$M = \begin{bmatrix} A & 0 \\ B & D \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix},$$

where $A = L[f(\lambda)]$, $f(\lambda) = f_0 + f_1\lambda + \cdots + \lambda^n$. *If* $f(M) = 0$, *then* $DY - YA = B = [\mathbf{b}_1, \ldots, \mathbf{b}_n]$ *has a solution*

$$Y = -\left[\mathbf{0}, \mathbf{b}_1, D\mathbf{b}_1 + \mathbf{b}_2, \ldots, D^{n-2}\mathbf{b}_1 + D^{n-3}\mathbf{b}_2 + \cdots + \mathbf{b}_{n-1}\right].$$

Proof. First of all, it is clear that $\psi_M = f(\lambda)$, since $g(M) = 0$ implies $g(A) = 0$ and hence $f \mid g$. Next, from (3.1) we have that $f(M) = 0$ if and only if $f(A) = 0$, $f(D) = 0$ and

$$\tilde{B} = \left[I, D, \ldots, D^{n-1}\right](S \otimes B)\begin{bmatrix} I \\ A \\ \vdots \\ A^{n-1} \end{bmatrix} = 0,$$

where $S = S[f(\lambda)]$.

Now the consistency condition $\tilde{B} = 0$ can be rewritten in column form $\tilde{\mathbf{b}} = H\mathbf{b} = \mathbf{0}$, where $\mathbf{b} = \operatorname{col}(B)$ and

$$H = \sum_{i,j=0}^{n-1} (A^T)^j f_{i+j+1} \otimes D^i = \sum_{i=0}^{n-1} A_i^T \otimes D^i, \tag{4.1}$$

in which

$$A_i = \mathfrak{f}_i(A) = \sum_{j=0}^{n-j-1} f_{i+k+1}A^j, \qquad i = 0,1,\dots,n-1. \tag{4.2}$$

To identify the matrix H, we use the following trick [5, p. 112]. In the identity

$$(\lambda I - A^T)\operatorname{adj}(\lambda I - A^T) = f(\lambda)I, \tag{4.3}$$

we replace λ by D. This gives

$$(I\otimes D - A^T\otimes I)\left(\sum_0^{n-1} A_i^T\otimes D^i\right) = GH = 0. \tag{4.4}$$

Likewise $HG = 0$, from which it is clear that $R(G) \subseteq N(H)$. We shall now show that in fact $R(G) = N(H)$. This will crucially depend on the specific structure of the companion matrix. It is easily verified that if

$$R(\lambda) = \left[\begin{array}{c|c} f_0(\lambda), f_1(\lambda), \dots, f_{n-2}(\lambda) & 1 \\ \hline -I & 0 \end{array}\right]$$

and

$$K(\lambda) = \begin{bmatrix} 1 & & & \\ \lambda & 1 & 0 & \\ \vdots & & \ddots & \\ \lambda^{n-1} & \cdots & \lambda & 1 \end{bmatrix},$$

then [7, p. 164]

$$R(\lambda)(\lambda I - A^T)K(\lambda) = \begin{bmatrix} f(\lambda) & & 0 & \\ & 1 & & \\ 0 & & \ddots & \\ & & & 1 \end{bmatrix} \tag{4.5}$$

and

$$\operatorname{adj}(\lambda I - A^T) = K(\lambda)\begin{bmatrix} 1 & & & & \\ & f(\lambda) & & 0 & \\ & & f(\lambda) & & \\ & 0 & & \ddots & \\ & & & & f(\lambda) \end{bmatrix} R(\lambda).$$

Replacing λ by D, and recalling that $f(D) = 0$, now gives

$$R(D)GK(D) = \begin{bmatrix} 0 & & & \\ & I & 0 & \\ & 0 & \ddots & \\ & & & I \end{bmatrix} = I - E, \tag{4.6}$$

and

$$H = K(D)\begin{bmatrix} I & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} R(D).$$

It is now easily seen that $H\mathbf{b} = \mathbf{0} \Leftrightarrow ER(D)\mathbf{b} = \mathbf{0} \Leftrightarrow \mathbf{b} = R^{-1}(D)(I - E)R(D)\mathbf{b} \Leftrightarrow \mathbf{b} = GK(D)R(D)\mathbf{b} \Leftrightarrow \mathbf{b} = G\mathbf{y}$ for some $\mathbf{y}$. Consequently $R(G) = N(H)$, and $DY - YA = B$ has a solution $Y = \operatorname{col}^{-1}\{[K(D)R(D)]\mathbf{b}\}$. ■

REMARKS.

(1) It follows also that $R(H) = N(G)$.

(2) The matrix $T = K(D)R(D)$ is an inner inverse of G (i.e. $GTG = G$), of a somewhat simpler form than the one used in [7, p. 165]. It is further clear that $I - GT = H^- H$, where $H^- = R^{-1}(D)EK^{-1}(D)$.

(3) An analogous result holds in the case where $D = L[g(\lambda)]$ is a companion matrix and $g(M) = 0$. In this case it is easier to use the matrix $\begin{bmatrix} A & C \\ 0 & D \end{bmatrix}$.

(4) The results of this section, as well as the previous, remain valid over any commutative ring with 1.

If $B=[\mathbf{b}_1,\mathbf{b}_2,\dots,\mathbf{b}_n]$, then the matrix solution has the form

$$Y=\operatorname{col}^{-1}\left(\begin{bmatrix} I & & & \\ D & I & & 0 \\ \vdots & & \ddots & \\ D^{n-1} & \cdots & D & I \end{bmatrix}\begin{bmatrix} \mathbf{v} \\ -\mathbf{b}_1 \\ \vdots \\ -\mathbf{b}_{n-1} \end{bmatrix}\right)$$

$$=\left[\mathbf{v},D\mathbf{v},\dots,D^{n-1}\mathbf{v}\right]-\left[\mathbf{0},\mathbf{b}_1,D\mathbf{b}_1+\mathbf{b}_2,\dots,D^{n-2}\mathbf{b}_1+\cdots+\mathbf{b}_{n-1}\right], \tag{4.7}$$

where

$$\mathbf{v}=f_0(D)\mathbf{b}_1+f_1(D)\mathbf{b}_2+\cdots+f_{n-1}(D)\mathbf{b}_n$$
$$=D_0\mathbf{b}_1+D_1\mathbf{b}_2+\cdots+D_{n-1}\mathbf{b}_n$$

and $D_i=f_i(D)$, $i=0,1,\dots,n-1$. Now by (1.8) $\mathbf{v}=[I,D,\dots,D^{n-1}](S\otimes I)\mathbf{b}$, in which $(S\otimes I)\mathbf{b}$ collapses, with the aid of (1.4), to

$$(S\otimes B)\operatorname{col}(I)=(S\otimes B)\begin{bmatrix} I \\ A \\ \vdots \\ A^{n-1}\end{bmatrix}\mathbf{e}_1.$$

Hence $\mathbf{v}=\tilde{B}\mathbf{e}_1=\mathbf{0}$. Consequently the solution takes the simple form

$$Y=-\left[\mathbf{0},\mathbf{b}_1,D\mathbf{b}_1+\mathbf{b}_2,\dots,D^{n-2}\mathbf{b}_1+D^{n-3}\mathbf{b}_2+\cdots+\mathbf{b}_{n-1}\right]$$

$$=-\left[I,D,\dots,D^{n-1}\right](I\otimes B)\begin{bmatrix} J^T \\ J^{T^2} \\ \vdots \\ J^{T^n}\end{bmatrix}$$

$$=-\sum_{i=0}^{n-1}D^iB(J^T)^{i+1}, \tag{4.8}$$

where

$$J=\begin{bmatrix} 0 & & & 0 \\ 1 & & & \\ & \ddots & & \\ & 0 & 1 & 0 \end{bmatrix}.$$

The form of this solution may of course be verified directly.

It should be noted that on the basis of (1.3), J^T will *never* be a polynomial in A, so that we cannot expect to be able to write this in the block-blinear form of (3.2). Let us now examine some closely related results.

COROLLARY 1. *Let*

$$M=\begin{bmatrix} A & 0 \\ B & D \end{bmatrix} \quad \text{and} \quad N=\begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix},$$

with $A=L[f(\lambda)]$. *Let* $h(M)=0$, *where* $h=f(\lambda)r(\lambda)$ *is monic and* $(\psi_D(\lambda), r(\lambda))=1$. *Then* $DY-YA=B$ *is consistent.*

Proof. The proof is similar to that of Theorem 1, except that now $h(A^T)=0$ implies that $(\lambda I-A^T)H(\lambda)=h(\lambda)I$, where $H(\lambda)=\Sigma_0^{n-1}H_i^T\lambda^i$ and $H_i=h_i(A)$ as in (1.5). Then (4.5) still holds, but

$$H(\lambda)=K(\lambda)\begin{bmatrix} r(\lambda) & & & 0 \\ & h(\lambda) & & \\ & & \ddots & \\ 0 & & & h(\lambda) \end{bmatrix}R(\lambda).$$

This gives the same expression for G, but now

$$H=K(D)\begin{bmatrix} r(D) & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix}R(D).$$

Since $(\psi_D, r(\lambda))=1$, it follows that $r(D)$ is invertible. It is now again clear that $H\mathbf{b}=\mathbf{0} \Leftrightarrow \mathbf{b}=GK(D)R(D)\mathbf{b}$ and that the solution Y has the same form as above. ■

COROLLARY 2. *Let* $M=\begin{bmatrix} A & C \\ 0 & D \end{bmatrix}$ *and* $N=\begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}$, *with* $A=L[f(\lambda)]$ *and* $f(M)=0$. *Then* $AX-XD=C$ *has a solution*

$$X=\mathcal{P}(I\otimes C)\mathcal{D}=\sum_{i=0}^{n-1}\sum_{i=0}^{n-1}P^iCD^j \tag{4.9}$$

where

$$\mathcal{P}=\left[P,P^2,\dots,P^n\right],\qquad \mathcal{D}=\begin{bmatrix} I \\ D \\ \vdots \\ D^{n-1} \end{bmatrix},$$

and $P=SJS^{-1}$, *with*

$$J=\begin{bmatrix} 0 & & & & 0 \\ 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ 0 & & & 1 & 0 \end{bmatrix}\quad and\quad S=\begin{bmatrix} f_1 & f_2 & \cdot & \cdot & \cdot & f_n \\ f_2 & & & & \cdot & \\ \cdot & & & \cdot & & \\ \cdot & & \cdot & & & \\ \cdot & \cdot & & & & \\ f_n & & & & & 0 \end{bmatrix}.$$

Proof. If we set $\mathcal{J}=[J,J^2,\dots,J^n]$, then by (1.2) and (1.7)

$$\begin{aligned} AX-XD &= \left[A\mathcal{P}-\mathcal{P}(A^T\otimes I)\right](I\otimes C)\mathcal{D} \\ &= S\{A^T\mathcal{J}-\mathcal{J}(A^T\otimes I)\}(I\otimes S^{-1})(I\otimes C)\mathcal{D}. \end{aligned}$$

Now $A^T\mathcal{J}-\mathcal{J}(A^T\otimes I)=[I,0,\dots,0]-\mathbf{e}_n[\mathbf{e}_1^TS,\mathbf{e}_2^TS,\dots,\mathbf{e}_n^TS]$, in which the second term reduces with the aid of (1.5) to $-\mathbf{e}_n\,\mathrm{col}(I)^T(S\otimes I)=-\mathbf{e}_n\mathbf{e}_1^T[I,A^T,\dots,(A^T)^{n-1}](S\otimes I)=-E_{n1}S^{-1}\mathcal{A}(S\otimes I)(I\otimes S)$, and where $\mathcal{A}=[I,A,\dots,A^{n-1}]$. Hence $AX-XD=S[I,0,\dots,0](I\otimes S^{-1})(I\otimes C)\mathcal{D}-SE_{n1}S^{-1}\mathcal{A}(S\otimes I)(I\otimes C)\mathcal{D}=[I,0,\dots,0](I\otimes C)\mathcal{D}-SE_{n1}S^{-1}\mathcal{A}(S\otimes C)\mathcal{D}=C$, since the second term vanishes by (3.1).

REMARKS.

(1) The solution in (4.9) can be obtained by applying Theorem 1 to the matrices

$$\begin{bmatrix} A & 0 \\ C^TS^{-1} & D^T \end{bmatrix}\quad\text{and}\quad\begin{bmatrix} A & 0 \\ 0 & D^T \end{bmatrix}.$$

These are obtained from M^T and N^T by a similarity transformation on A^T. In fact $X = -SY^T$, where Y is the solution to $D^TY - YA = C^TS^{-1}$, given by (4.8).

(2) The matrix $P = SJS^{-1}$ can be simplified if we use the fact that

$$S^{-1} = \begin{bmatrix} \mathbf{e}_n^T \\ \mathbf{e}_n^T A \\ \vdots \\ \mathbf{e}_n^T A^{n-1} \end{bmatrix}.$$

Then

$$P = S \begin{bmatrix} \mathbf{0}^T \\ \mathbf{e}_n^T \\ \mathbf{e}_n^T A \\ \vdots \\ \mathbf{e}_n^T A^{n-2} \end{bmatrix}.$$

Alternatively, if we set $J = L + \mathbf{f}\mathbf{e}_n^T$, with $\mathbf{f}^T = [f_0, f_1, \dots, f_{n-1}]^T$, then $P = I + S\mathbf{f}\mathbf{e}_n^T S^{-1} = I + S\mathbf{f}\mathbf{e}_n^T A^{n-2}$. Likewise $SE_{n1}S^{-1}$ reduces to E_{1n}.

Corollary 3. *If*

$$M = \begin{bmatrix} A & 0 \\ B & D \end{bmatrix}, \qquad N = \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix},$$

with $A = L(p^k)$, $D = L(q^l)$, *and* p, q *prime polynomials, then the following are equivalent*:

(i) $DY - YA = B$ is consistent,
(ii) $M \approx N$,
(iii) $\psi_M = \psi_N$.

Proof. (iii)$\Rightarrow$(i): If $p \neq q$ then $\psi_M = p^k q^l$, and it is well-known that the equation in (i) has a unique solution. Alternatively, we could use Corollary 1, which actually gives a simpler solution. If $p = q$, then $\psi_N = p^t$, $t = \max(k, l)$, and so $\psi_M = \psi_A$ or $\psi_M = \psi_D$ and either Theorem 1 or Corollary 2 applies. ■

We should remark here that the results of Theorem 1 and Corollary 3 essentially solve the problem of giving a tractable consistency condition and a simple particular solution to the matrix equation

$$H_k X - X H_l = C, \qquad k \leqslant l, \tag{4.10}$$

where $H_k = H(p^k)$ is the hypercompanion matrix of p^k. Indeed, if $L_l = L(p^l)$ and $Q_l = [\mathbf{z}, L_l\mathbf{z}, \ldots, L_l^{r-1}\mathbf{z}; p(L_l)\mathbf{z}, \ldots, p(L_l)L_l^{r-1}\mathbf{z}; \cdots p^{l-1}(L_l)L_l^{r-1}\mathbf{z}]$, with $\psi_{\mathbf{z}} = p^l$, then $Q_l^{-1}L_l Q_l = H_l$. Hence we obtain

$$L_k Y - Y L_l = B, \tag{4.11}$$

where $Y = Q_k X Q_l^{-1}$ and $B = Q_k C Q_l^{-1}$. It then follows from (3.1) that (4.10) is consistent precisely when the block-bilinear form $\tilde{B}$ vanishes, that is, when

$$\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} H_k^i f_{i+j+1} C H_l^j = 0, \tag{4.12}$$

where $f(\lambda) = p^l(\lambda)$ and $n = rl = \partial f$. Moreover, a particular solution is given by the block-bilinear form

$$X = -\sum_{i=0}^{n-1} H_k^i C \left(Q_l^{-1} J^T Q_l\right)^{i+1}. \tag{4.13}$$

The expression in parentheses is independent of the choice of $\mathbf{z}$, as long as $\psi_{\mathbf{z}} = p^l$. This may be seen as follows. If $\mathbf{z}^T = [z_0, z_1, \ldots, z_{n-1}]$ and $z(\lambda) = \sum_{i=0}^{n-1} z_i \lambda^i$, then Q_l can be written as $z(L_l)M$, where $z(L_l)$ is invertible and

$$M = \left[\begin{bmatrix} I_r \\ 0 \end{bmatrix}, p(L_l) \begin{bmatrix} 0 \\ I_r \\ 0 \end{bmatrix}, \ldots, p^{l-1}(L_l) \begin{bmatrix} 0 \\ \vdots \\ I_r \end{bmatrix} \right].$$

Now $z(L_l)^{-1} J^T z(L_l)$ is independent of the choice of $\mathbf{z}$, which follows from the following fact:

Lemma. *If $L = L(f)$ and $g(\lambda) = g_0 + \cdots + g_{n-1}\lambda^{n-1}$, then $g(L)J^T = J^T g(L)$ if and only if $g(\lambda) = g_0$.*

5. THE CYCLIC DECOMPOSITION THEOREM

Let us now use the removal rule for companion matrices to give a quick constructive proof of the cyclic decomposition theorem. It should be pointed out that there is a *weak* version of this theorem, which states that

$$A \approx A_I = \begin{bmatrix} L(\psi_1) & & & 0 \\ & L(\psi_2) & & \\ & & \ddots & \\ 0 & & & L(\psi_r) \end{bmatrix}$$

for some polynomials $\psi_1, \psi_2, \ldots, \psi_r$. On the other side there is the *strong* version, which in addition states that we may select the ψ_i so that $\psi_{i+1} \mid \psi_i$, and $\psi_1 = \psi_A$. In this case the ψ_i are unique and are called the invariant factors of A. The weak version is usually proven using the cyclic decomposition theorem for abelian groups [9, p. 162], but uses induction and gives no real information as to the actual transformation that has been used. The strong version of the CDT is traditionally proven either by induction or by using quotient spaces, and is also not transparent as to what goes on.

THEOREM 2. *Let* $A \in \mathfrak{F}_{n \times n}$. *Then*

$$A \approx A_I = \begin{bmatrix} L(\psi_1) & & 0 \\ & \ddots & \\ 0 & & L(\psi_r) \end{bmatrix},$$

where $\psi_1 = \psi_A = f_0 + f_1\lambda + \cdots + \lambda^k$ *and* $\psi_{i+1} \mid \psi_i$, $i = 1, 2, \ldots, r = 1$.

Proof. Let $A = A_1$ and $\psi_A = \psi_1$. From the primary decomposition theorem we know that there exists a vector $\mathbf{x}$ that belongs to ψ_1, i.e., $\psi_{\mathbf{x}} = \psi_1$. Selecting the vectors $[\mathbf{x}, A\mathbf{x}, \ldots, A^{k-1}\mathbf{x}]$ as first part of our basic matrix Q, we have

$$AQ = Q\left[\begin{array}{c|c} L(\psi_1) & C_1 \\ \hline 0 & A_2 \end{array}\right].$$

Since $\psi_1 = \psi_A$, it is clear that $\psi_1(A_2) = 0$ and $\psi_{A_2} \mid \psi_1$. It is now clear that

Corollary 2 is applicable, and

$$A \approx \begin{bmatrix} L(\psi_1) & 0 \\ 0 & A_2 \end{bmatrix}.$$

Repeating with A_2 yields the desired canonical form. The uniqueness follows from Lemma 3. For if

$$\begin{bmatrix} L(\psi_1) & & & 0 \\ & L(\psi_2) & & \\ & & \ddots & \\ 0 & & & L(\psi_r) \end{bmatrix} \approx \begin{bmatrix} L(\phi_1) & & & 0 \\ & L(\phi_2) & & \\ & & \ddots & \\ 0 & & & L(\phi_s) \end{bmatrix}$$

with $\psi_1 = \phi_1 = \psi$ and $\psi_{i+1} \mid \psi_i, \phi_{i+1} \mid \phi_i$, then by Lemma 3,

$$\begin{bmatrix} L(\psi_2) & & 0 \\ & \ddots & \\ 0 & & L(\psi_r) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} L(\phi_2) & & \\ & \ddots & \\ & & L(\phi_s) \end{bmatrix}$$

have the same minimal polynomial. That is, $\psi_2 = \phi_2$. Now repeat. ∎

6. THE SEMISIMPLE CASE

Recall that a matrix $M \in F_{n \times n}$ is *semisimple* [10, p. 263] if $\psi_M(\lambda) = p_1(\lambda)p_2(\lambda) \cdots p_s(\lambda)$, where the $p_i(\lambda)$ are distinct primes. If the p_i are linear, we usually call M *simple*. In either case the removal rule takes on a particularly simple form.

THEOREM 3. *Let* $\psi_M(\lambda) = f(\lambda) = f_0 + f_1\lambda + \cdots + \lambda^k$, *and let* $L = L[f(\lambda)]$. *Then the following are equivalent*:

(i) *M is semisimple*,
(ii) $(f, f') = 1$,
(iii) $\mathbf{e}_1 \in R[f'(L)]$,
(iv) $LZ - ZL = E_{1n} + p(L)$ *is consistent for some polynomial* $p(\lambda)$,
(v) *every invariant subspace has an invariant complement.*

Proof. (i)$\Rightarrow$(ii): Since $f(\lambda)$ has no repeated factors and the $p_i(\lambda)$ are distinct primes, this is clear.

(ii)$\Rightarrow$(iii): Since $(f, f')=1$ and $f=\psi_M$, it follows that $f'(L)$ is invertible.

(iii)$\Rightarrow$(iv): Consider $LZ-ZL=E_{1n}+p(L)$, with $p(\lambda)$ a polynomial. This is exactly Equation (3.6). To check consistency, recall that

$$(\lambda I-L^T)K(\lambda)=\left[\begin{array}{c|c} 0 & \\ \vdots & -I \\ \hline f(\lambda) & f_0(\lambda)f_1(\lambda)\cdots f_k(\lambda) \end{array}\right]. \tag{6.1}$$

On replacing λ by D we have

$$(I\otimes D-L^T\otimes I)K(D)=\left[\begin{array}{c|c} 0 & \\ \vdots & -I \\ \hline f(D) & f_0(D), f_1(D),\ldots,f_{k-2}(D) \end{array}\right], \tag{6.2}$$

where $G=I\otimes D-L^T\otimes I$ is given by (4.6) with $n=k$. Now consider the special case where $D=L$, and $\mathfrak{f}_i(D)=L_i$. Then on taking columns through (3.6) and using (1.3) with $p(L)=[\mathbf{p}, L\mathbf{p},\ldots,L^{k-1}\mathbf{p}]$, we obtain

$$G\mathbf{z}=\begin{bmatrix} \mathbf{p} \\ L\mathbf{p} \\ \vdots \\ L^{k-1}\mathbf{p} \\ \mathbf{e}_1+L^{k-1}\mathbf{p} \end{bmatrix}=\mathbf{w}. \tag{6.3}$$

This will be consistent precisely when the system

$$\left[\begin{array}{c|c} 0 & -I \\ \hline 0 & L_0L_1\cdots L_{k-2} \end{array}\right]\mathbf{u}=\mathbf{w} \tag{6.4}$$

is consistent, where $\mathbf{u}=K(L)^{-1}\mathbf{z}$. Using elementary row operations, this reduces to

$$\begin{bmatrix} 0 & -I \\ 0 & 0 \end{bmatrix} \mathbf{u} = \begin{bmatrix} \mathbf{p} \\ L\mathbf{p} \\ \vdots \\ L^{k-2}\mathbf{p} \\ \mathbf{q} \end{bmatrix}, \tag{6.5}$$

where $\mathbf{q} = (L_0 + LL_1 + L^{k-1}L_{k-1})\mathbf{p} + \mathbf{e}_n$. Clearly (6.5) is consistent exactly when $\mathbf{q} = \mathbf{0}$, that is, when $\mathbf{e}_1 \in R(T)$, where $T = L_0 + LL_1 + \cdots + L^{k-1}L_{k-1}$. Using (1.6) and adding terms, we see that

$$T = f'(L), \qquad \text{where} \quad f'(\lambda) = \sum_{i=0}^{k-1} if_i\lambda^{i-1}.$$

Consequently, (3.6) is consistent if and only if $\mathbf{e}_1 \in R[f'(L)]$.

(iv) $\Rightarrow$ (v): Let

$$M \approx \begin{bmatrix} A & C \\ 0 & D \end{bmatrix} = M_1.$$

It suffices to show that C can be removed by a similarity transformation. Using the results of Section 3, we recall that the consistency of (3.6) suffices for $AX - XD = C$ to have a solution, and so we are done.

(v) $\Rightarrow$ (i): This was shown in [10, p. 264]. ∎

REMARKS.

(1) If $\mathbf{e}_1 \in R[f'(L)]$, then one can directly show that $(f, f') = 1$. Indeed, since $f'(L)$ commutes with L, $R[f'(L)]$ is an L-invariant subspace. Thus, span$\{\mathbf{e}_1, L\mathbf{e}_1, \ldots, L^{k-1}\mathbf{e}_1\} \subseteq R[f'(L)]$, which ensures that $f'(L)$ is invertible and $(f, f') = 1$.

(2) The construction of part (iv) in Theorem 3 actually also shows that *any* matrix X can be written in the form $LZ - ZL - p(L)$, for suitable Z and $p(\lambda)$. In fact, on taking columns, we obtain (6.5) modified to

$$\begin{bmatrix} 0 & -I \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{k-1} \end{bmatrix} = \begin{bmatrix} \mathbf{p} \\ L\mathbf{p} \\ \vdots \\ L^{k-2}\mathbf{p} \\ \mathbf{r} \end{bmatrix} + \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \vdots \\ \mathbf{x}^{k-2} \\ \mathbf{s} \end{bmatrix},$$

where $\mathbf{r} = f'(L)\mathbf{p}$ and $\mathbf{s} = L_0\mathbf{x}_0 + L_1\mathbf{x}_1 + \cdots + L_{k-1}\mathbf{x}_{k-1}$. This gives $\mathbf{p} = -[f'(L)]^{-1}\mathbf{s}$ and $\mathbf{u}_{i+1} = -(\mathbf{x}_i + L^i\mathbf{p})$, $i = 0,1,\ldots,k-2$, with $\mathbf{u}_0$ arbitrary. From this we may recover $\mathbf{z} = K(D)\mathbf{u}$ and hence Z.

(3) The equivalence of parts (i) and (v) in Theorem 3 may also be seen from the fact that M is semisimple precisely when the ring $R = \mathfrak{F}[M]$ is semisimple. Since both chain conditions obviously hold in R, every left R-module is completely reducible. In particular every R-submodule of ${}_R\mathfrak{F}^n$ is a direct summand, which translates back to (v). The converse follows similarly.

(4) The equivalence of (i) and (v) may also be proven using the method of Section 4. Indeed, if

$$M = \begin{bmatrix} A & C \\ 0 & D \end{bmatrix}$$

and if $f(M) = 0$, then $f(A) = 0$, $f(D) = 0$, and $\tilde{C} = 0$. Again $\tilde{C} = 0$ if and only if $H\mathbf{c} = 0$, where

$$H = \sum_{i=0}^{k-1} D_i^T \otimes A^i \quad \text{and} \quad D_i = \mathfrak{f}_i(D).$$

On the other hand $f(D^T) = 0$ implies that

$$(\lambda I - D^T)H(\lambda) = f(\lambda)I,$$

where

$$H(\lambda) = \sum_{0}^{k-1} D_i^T \lambda^i.$$

Formally differentiating this with respect to λ shows that

$$(\lambda I - D^T)H'(\lambda) + H(\lambda) = f'(\lambda)I, \tag{6.6}$$

Replacing λ by A now gives

$$GH'(A) + H = I \otimes f'(A) \tag{6.7}$$

where

$$G = (I \otimes A - D^T \otimes I)$$

and

$$H'(A) = \sum_{i=0}^{k-1} iD_i^T \otimes A^{i-1}.$$

Now $(f, f') = 1$ implies that $f'(A)^{-1}$ exists and is a polynomial in A. Consequently, $I \otimes f'(A)$ and G commute. Hence if $H\mathbf{c} = \mathbf{0}$, then we obtain

$$G\left[I \otimes f'(A)^{-1}\right] H'(A)\mathbf{c} = \mathbf{c} \tag{6.8}$$

ensuring that $\mathbf{c} \in R(G)$, and $AX - XD = C$ is consistent. The solution we found is given by

$$X = \operatorname{col}^{-1}\left\{ \sum_{i=0}^{k-1} iD_i^T \otimes A^{i-1}(f'(A))^{-1}\mathbf{c} \right\}, \tag{6.9}$$

which may be written in block-bilinear form

$$X = \left[I, A, \dots, A^{k-1}\right](Y \otimes C) \begin{bmatrix} I \\ D \\ \vdots \\ D^{k-1} \end{bmatrix}$$

for some matrix Y depending on the coefficients in $f'(A)^{-1} = \sum_{i=0}^{k-1} \gamma_i A^i$.

(5) The case where

$$M = \begin{bmatrix} A & C \\ 0 & D \end{bmatrix}$$

is semisimple is not the only case where a block-bilinear solution of the form (3.2) solves $AX - XD = C$. The other well-known case is where $(\Delta_A, \Delta_D) = 1$ [6]. In fact, using (6.2) with D replaced by L_A and setting $L = L_D$, $f = \Delta_D$, we see that $I \otimes L_A - L_D^T \otimes I$ is invertible, as $f(L_A)$ is invertible. Hence $L_A Y - YL_D = E_{11}S$ is consistent, as well as $L_A Y - YL_D^T = E_{11}$. The latter ensures that

$$X = \left[I, A, \dots, A^{m-1}\right](Y \otimes C) \begin{bmatrix} I \\ D \\ \vdots \\ D^{m-1} \end{bmatrix}$$

will solve $AX - XD = C$ uniquely.

7. ROTH'S SIMILARITY THEOREM

We conclude this paper by applying the removal rule for companion matrices, in the form of Corollary 2, to give a constructive algorithm for finding the solution to $AX - XD = C$. Again let

$$M = \begin{bmatrix} A & C \\ 0 & D \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}.$$

Without loss of generality, using the CDT, let

$$A = \begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix}, \qquad D = \begin{bmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_n \end{bmatrix},$$

where $A_i = L(p_i^{k_i})$ and $D_j = L(q_j^{l_j})$. Furthermore, let us assume that the elementary divisors $p_i^{k_i}$ have been ordered in *decreasing* order, so that $k_i \geqslant k_{i+1}$ whenever $p_{i+1} = p_i$, and that the elementary divisors of D have been ordered in *increasing* order. That is, $l_j \leqslant l_{j+1}$ whenever $q_{j+1} = q_j$. Also let $C = [C_{ij}]$ be partitioned conformally. It is well known that if

$$A_i X_{ij} - X_{ij} D_j = C_{ij} \tag{7.1}$$

has a solution for every value of i and j, then the block matrix $X = [X_{ij}]$ solves $AX - XD = C$. There are two cases to be considered.

Case (i): $p_i \neq q_j$. In this case $(p_i, q_j) = 1$ and (7.1) has a unique solution [7, p. 171], which can be expressed as a block-bilinear solution. Alternatively, Corollary 1 can be used.

Case (ii): $p_i = q_j$. Let

$$W_{ij} = \begin{bmatrix} A_i & C_{ij} \\ 0 & D_j \end{bmatrix} \quad \text{and} \quad S_{ij} = \left[\begin{array}{cccc|ccc} A_i & & & 0 & C_{i1} & \cdots & C_{ij} \\ & A_{i+1} & & & & & \\ & & \ddots & & \vdots & & \vdots \\ 0 & & & A_m & C_{m1} & \cdots & C_{mj} \\ \hline & & & & D_1 & & 0 \\ & & 0 & & & \ddots & \\ & & & & 0 & & D_j \end{array}\right].$$

Then by Lemma 1(a), $\psi_{W_{ij}} = p_i^t$, where $t = \max(k_i, l_i)$. On the other hand, by Lemma 2, $\psi_{W_{ij}} \mid \psi_{S_{ij}}$, where $\psi_{S_{ij}} = \mathrm{lcm}(\psi_{A_u} : u \geqslant i,\ \psi_{D_v} : v \leqslant j)$. From the way the elementary divisors have been ordered, it follows that $\psi_{S_{ij}}(\lambda) = \mathrm{lcm}(\psi_{A_i}, \psi_{D_j}) \cdot r(\lambda)$, where $r(\lambda)$ is coprime to the first factor. Hence $\psi_{W_{ij}} = \mathrm{lcm}(\psi_{A_i}, \psi_{D_j})$. This means that Theorem 1, or Corollary 2, may be applied to each W_{ij} independently, *without* changing the matrices C_{ij} in the process. Thus we may remove the C_{ij} one by one without any further calculations, and (7.1) has a solution for each i and j, as desired.

8. CONCLUSIONS

Several of the results, such as those in Sections 3 and 4, remain valid over any commutative ring with 1. It is not known whether the method of annihilating polynomials can be used to prove Roth's theorem for commutative rings. On the other hand, Theorem 1, will no longer hold if $\mathcal{F}$ is replaced by $\mathbb{Z}$, and the conditions $A = L[f(\lambda)]$ and $f(M) = 0$ are replaced by the assumptions that A is indecomposable under similarity and $\Delta_A(M) = 0$. For example, let

$$A = D = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Then $AX - XD = C$ has no solution over $\mathbb{Z}$. We close with a question related to the above content: when exactly does $AX - XD = C$ have a block-bilinear solution $\sum_{ij} A^i y_{ij} C B^j$? Partial answers were given in [6] and [11].

The author wishes to thank the referee for making several valuable suggestions, including remarks (2) and (3) and the proof of remark (1) following Theorem 3.

REFERENCES

1 R. B. Feinberg, Similarity of partitioned matrices, *J. Res. Nat. Bur. Standards Sect. B* 79*B*:117–125 (1975).

2 H. Flanders and H. K. Wimmer, On the matrix equations $AX - XB = C$ and $AX - YB = C$, *SIAM J. Appl. Math.* 32:707–710 (1977).

3 R. M. Guralnick, Roth's theorem and decomposition of modules, *Linear Algebra Appl.* 39:155–165 (1980).

4 ______, Similarity of matrices over local rings, *Linear Algebra Appl.* 41:161–174 (1981).

5 W. Gustafson, Roth's theorems over commutative rings, *Linear Algebra Appl.* 23:245–251 (1979).
6 R. E. Hartwig, Resultants and the solution of $AX - XB = -C$, *SIAM J. Appl. Math.* 23:104–117 (1972).
7 ______, $AX - XB = C$, resultants and generalized inverses, *SIAM J. Appl. Math* 28:154–183 (1975).
8 ______, From Jacobi to Jordan, to appear.
9 I. N. Herstein, *Topics in Algebra*, Ginn, Waltham, Mass., 1964, p. 162.
10 K. Hoffman and R. Kunze, *Linear Algebra*, Prentice-Hall, Englewood Cliffs, N.J., 1961.
11 E. de Souza and S. P. Bhattacharyya, Controllability, observability, and the solution of $AX - XB = C$, *Linear Algebra Appl.* 39:167–188 (1981).
12 W. E. Roth, The equations $AX - YB = C$ and $AX - XB = C$ in matrices, *Proc. Amer. Math. Soc.* 3:392–396 (1952); MR 13,900.

Received 27 January 1982; revised 17 May 1982