Factored Forms for Solutions of
$AX - XB = C$ and $X - AXB = C$ in Companion Matrices

P. Lancaster
Department of Mathematics and Statistics
University of Calgary
Calgary, Alberta, Canada

L. Lerer*
Department of Mathematics
Technion — Israel Institute of Technology
Haifa, Israel

and

M. Tismenetsky
Department of Mathematics and Statistics
University of Calgary
Calgary, Alberta, Canada

Submitted by Hans Schneider

ABSTRACT

The main concern of this paper is linear matrix equations with block-companion matrix coefficients. It is shown that general matrix equations $AX - XB = C$ and $X - AXB = C$ can be transformed to equations whose coefficients are block companion matrices: $\hat{C}_L X - XC_M = \text{diag}(I \ 0 \ \cdots \ 0)$ and $X - \hat{C}_L X C_M = \text{diag}(I \ 0 \ \cdots \ 0)$, respectively, where $C_L$ and $C_M$ stand for the first and second block-companion matrices of some monic $r \times r$ matrix polynomials $L(\lambda) = \lambda^s I + \sum_{j=0}^{s-1} \lambda^j L_j$ and $M(\lambda) = \lambda^t I + \sum_{j=0}^{t-1} \lambda^j M_j$. The solution of the equations with block-companion coefficients is reduced to solving vector equations $Sx = \rho$, where the matrix $S$ is $r^2 I \times r^2 I \ [I - \max(s, t)]$ and enjoys some symmetry properties.

*Partially supported by the fund for promotion of research at the Technion—Israel Institute of Technology.

1. INTRODUCTION

If A, B, and C are matrices of sizes $m \times m$, $n \times n$, and $m \times n$ respectively, the linear matrix equations

\[ AX - XB = C \]  
(1.1)

and

\[ X - AXB = C \]  
(1.2)

arise frequently in various problems of pure and applied mathematics. In particular, the cases in which A and B are block-companion matrices are of interest. Equations of such type appear, for instance, when studying vector differential or finite-difference equations of the form

\[
\frac{d^s x(t)}{dt^s} + L_{s-1} \frac{d^{s-1} x(t)}{dt^{s-1}} + \cdots + L_0 x(t) = 0, \\
\]

\[
x_{i+s} + L_{s-1} x_{i+s-1} + \cdots + L_0 x_i = 0 \quad (i = 0, 1, \ldots),
\]

in which the coefficients $L_j (j = 0, 1, \ldots, s - 1)$ are matrices of size $r \times r$ and $x(t) \ (0 < t < \infty)$ or $x_0, x_1, \ldots$ are $r$-dimensional vectors. Using the idea of linearization, together with the Lyapunov stability theory, one sees that the stability properties of the solutions of Equation (1.3) [respectively, (1.4)] are closely related to the solutions of the matrix equation $C_L X + X C_L^* = -W$ [respectively, $X + C_L X C_L^* = -W$], where $W = W^* > 0$ and $C_L$ is the (first) companion matrix of the monic $r \times r$ matrix polynomial $L(\lambda) = \lambda^s I + \sum_{j=0}^{s-1} \lambda^j L_j$:

\[
C_L = \begin{bmatrix}
0 & I & & 0 \\
& \ddots & \ddots & \vdots \\
0 & 0 & \ddots & I \\
- L_0 & - L_1 & \cdots & - L_{s-1}
\end{bmatrix}.
\]

An additional example of an equation (1.2) with block-companion coefficients can be found in [11].

The main objects of study of this paper are equations

\[
\hat{C}_L X - X C_M = \text{diag}\{R, 0, \ldots, 0\}
\]
(1.6)

and

\[
X - \hat{C}_L X C_M = \text{diag}\{R, 0, \ldots, 0\},
\]
(1.7)
where \( L(\lambda) = \lambda^r I + \sum_{j=1}^{r-1} \lambda^j L_j \) and \( M(\lambda) = \lambda^s I + \sum_{j=1}^{s-1} \lambda^j M_j \) are monic \( r \times r \) matrix polynomials and \( \hat{C}_L \) stands for the second companion matrix of \( L: \hat{C}_L = [C_{L}]^T \). We are primarily concerned with the following two independent problems: (a) Under what conditions can the general equations (1.1) and (1.2) be transformed to equations in companion matrices (1.6) and (1.7) respectively? (b) What numerical improvements in solving equations (1.6) and (1.7) may follow from the special structure of the coefficients in these equations? For the case of scalar polynomials \( L(\lambda) \) and \( M(\lambda) \) problem (a) has been investigated in [19], while problem (b) appears in [12] and [22].

Firstly we show that under practically interesting hypotheses the solutions of the general equations (1.1) and (1.2) can be expressed via those of equations (1.6) and (1.7) respectively. To be more precise, assume that the pairs \((A, C)\) and \((C, B)\) are controllable and observable respectively (these conditions arise naturally in systems theory), and let \( C = U^* V \) be a rank decomposition of \( C \). Let \( r = \text{rank} \, C \), and let \( s \) and \( t \) be two integers such that the matrices \( \text{row}(A^i U^*)_{i=0}^{s-1} \) and \( \text{col}(V B^i)_{i=0}^{t-1} \) are of full rank. Following the lead given in [18] and taking advantage of the modern theory of matrix polynomials, we can associate with the pairs \((A, U^*)\) and \((V, B)\) two manic \( r \times r \) matrix polynomials \( L_A(\lambda) \) and \( L_B(\lambda) \) of degrees \( s \) and \( t \) respectively such that Equation (1.1) is solvable if and only if the equation

\[
\hat{C}_L A X_0 - X_0 C L_B = \text{diag}[I, 0, \ldots, 0]
\]

is solvable and any solution \( X \) of (1.1) can be represented in the form

\[
X = \text{row}(A^i U^*)_{i=0}^{s-1} X_0 \text{col}(V B^i)_{i=0}^{t-1},
\]

where \( X_0 \) is a solution of (1.8).

Remark that although Equation (1.8) may have matrix coefficients of larger size than the original equation (1.1), the new coefficients are sparse and the information implicit in the equation (1.1) is organized in a more explicit way in the new equation (1.8). In particular, this transformation sharply illuminates the role of the rank of the right-hand term \( C \) in the structure of the solutions of (1.1).

As to problem (b), we show that the sparse structure of the coefficients in equations (1.6)–(1.7) allows us to reduce their solution to solving a vector equation

\[
S x = \rho
\]

whose coefficient matrix \( S \) has size \( pr^2 \times pr^2 \) \( [p = \max(s, t)] \). Note that a
straightforward rewriting of (1.6) or (1.7) in a vector-equation form leads to a coefficient matrix (the nivellateur) of size $s^2 r^2 \times t^2 r^2$. Moreover, the matrix $S$ in (1.10) has a certain symmetric structure. In fact $S$ can be chosen to be a Hankel or Bezout matrix associated with the polynomials $L(\lambda)$ and $M(\lambda)$. This structure of $S$ may yield computational advantages in some problems.

Making a direct comparison of Equation (1.1) with the transformed equation (1.8), it is found that for sufficiently small values of the integer $r = \text{rank } C$ the matrix $S$ in Equation (1.10) [corresponding to (1.8)] may actually have size less than $n^2 \times m^2$, which is the size of the nivellateur of (1.1).

The paper is organized as follows. The results concerning problem (a) are presented in Sections 3–4. The formula (1.9) is proved in Theorem 3. In Theorem 4 (which includes Theorem 1 of [20]) it is shown how the controllability and observability hypotheses which appear in Theorem 3 can be removed. In Theorem 8 the implications of these results for the symmetric equation $AX + XA^* = C$ are worked out.

Solution of problem (b) is the content of Sections 6–7. Combining the results of Sections 3–4 and 6–7, it is seen in Section 8 that under the hypotheses of Theorem 4 (or the stronger ones of Theorem 3) Equation (1.1) can be transformed to an equivalent equation (1.10) in which $S$ is a Bezout or a block-Hankel matrix.

For convenience the analysis of this paper concentrates on Equation (1.1), but there are parallel results for Equation (1.2), and they are indicated in the final section of the paper. Here, we also include some discussion of the interesting case when $\text{rank } C = 1$ (cf. [13]).

It will be necessary to establish some ideas from the theory of matrix polynomials as well as connections with the more familiar ideas of controllability and observability of matrix pairs. This is done in Section 2.

2. PRELIMINARIES

Consider first the theory of matrix polynomials. Let $L(\lambda) = \lambda^l I + \sum_{i=0}^{l} \lambda^i L_i$ denote a monic $n \times n$ matrix polynomial. Thus $L_0, L_1, \ldots , L_{l-1}, I \in \mathbb{C}^{n \times n}$, the space of $n \times n$ complex matrices. Then let

$$
C_L = \begin{bmatrix}
0 & I & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & I \\
-L_0 & -L_1 & \cdots & -L_{l-1}
\end{bmatrix}, \quad \hat{C}_L = \begin{bmatrix}
0 & \cdots & 0 & -L_0 \\
I & \cdots & \cdots & \cdots \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & I & -L_{l-1}
\end{bmatrix}
$$

(2.1)
be the first and second companion matrices (respectively) associated with the given matrix polynomial. Denote by \( \sigma(L) \) the spectrum of \( L(\lambda) \):

\[
\sigma(L) = \{ \lambda \in \mathbb{C} : \det L(\lambda) = 0 \},
\]

and note that \( \sigma(L) = \sigma(\lambda I - C_L) = \sigma(\lambda I - \hat{C}_L) \).

Now we introduce some notational devices. By \( \text{row}(A_i) \) we denote the one-row block matrix \([A, A_2, \ldots, A_i] \). Similarly, \( \text{col}(A_i) \) is written for the corresponding one-column block matrix.

A pair of matrices \((X, T)\) where \( X \) is \( n \times n \) and \( T \) is \( n \times n \) is called a **standard pair** for \( L(\lambda) \) if these two conditions are satisfied:

1. \( \text{col}(XT_0) \) is nonsingular,
2. \( XT_0 + \sum_{j=0}^{l-1} L_j XT_j = 0 \).

It is known (see [7]) that a standard pair determines a monic matrix polynomial completely, i.e. it can be used to determine the coefficients \( L_0, L_1, \ldots, L_{l-1} \) explicitly. In fact, if \((X, T)\) is a standard pair for \( L(\lambda) \), then

\[
\begin{bmatrix}
L_0 & L_1 & \cdots & L_{l-1}
\end{bmatrix} = XT^0 \left( \text{col}(XT^0) \right)^{-1}.
\]

There is a dual notion: a **left standard pair** for \( L(\lambda) \) is a pair of matrices \((S, Y)\) such that \((Y^T, ST)\) is a standard pair of \( L^T(\lambda) = \lambda I + \sum_{i=0}^{l-1} \lambda^i L_i^T \).

Turn now to a more general examination of matrix pairs. A pair of matrices \((Q, T)\) is said to be (right) admissible of order \( p \) if \( Q \) is \( r \times p \), \( T \) is \( p \times p \), and \( \ker Q \neq \{0\} \), where \( \ker \) stands for kernel, or nullspace. Define \( K_i = \ker(\text{col}(QT^j)) \) for \( i = 1, 2, \ldots \); it is easily seen that \( K_1 \supsetneq K_2 \supsetneq \cdots \) with strict inclusion for increasing \( i \) until equality first holds, and equality holds thereafter. Thus, there is a least positive integer \( l (l \leq p) \), called the **index** of the admissible pair \((Q, T)\), for which \( K_l = K_{l+1} \). The **kernel** of the pair \((Q, T)\) is the subspace of \( \mathbb{C}^p \) defined by

\[
\ker(Q, T) = \bigcap_{j=1}^{\infty} \ker(QT^j) = \ker(\text{col}(QT^{j-1})^\infty_{j=1}),
\]

or simply \( \ker(Q, T) = K_l \), where \( l \) is the index of \((Q, T)\).

It follows easily from the definition that

\[
l \leq p + 1 - \text{rank } Q.
\]
An admissible pair is said to be observable, or (right) independent, if \( \text{Ker}(Q, T) = \{0\} \). Otherwise, \( \text{Ker}(Q, T) \) is sometimes known as the unobservable subspace, and we note that it is necessarily \( T \)-invariant.

As an important example, note that a standard pair \((X, T)\) as introduced above is admissible of order \( nl \). It is also independent with index \( l \).

A pair of matrices \((T, R)\) is said to be (left) admissible of order \( p \) if \( T \) is \( p \times p \) and \( R \) is \( p \times r \). Such a pair is said to be (left) independent or controllable if the right admissible pair \((R^*, T^*)\) is right independent; and more generally, the index of \((T, R)\) is defined to be that defined above for \((R^*, T^*)\).

Now suppose that \((Q_1, T_1)\) and \((Q, T)\) are right admissible pairs of orders \( p_1 \) and \( p \), respectively, with \( p_1 \leq p \). Then \((Q, T)\) is said to be an extension of \((Q_1, T_1)\) if there is a full rank (i.e. left invertible) matrix \( S \) such that
\[
QS = Q_1, \quad TS = ST_1.
\]
Under the same conditions \((Q_1, T_1)\) is called a restriction of \((Q, T)\). For a left admissible pair \((T, R)\) extensions and restrictions are defined by applying the above definitions to the right admissible pair \((R^*, T^*)\).

Note that right admissible pairs \((Q, T)\) and \((R, S)\) are said to be similar if there is a nonsingular \( M \) such that
\[
Q = RM \quad \text{and} \quad T = M^{-1}SM.
\]
As an important illustration, note that all standard pairs for a fixed monic matrix polynomial are similar to one another in this sense.

The following lemmas will play important parts in our analysis. The first is well known and easily established.

**Lemma 1.** An admissible pair
\[
\begin{bmatrix}
Q_1 & Q_2 \\
T_1 & T_2
\end{bmatrix}
\]
is an extension of admissible pair \((Q_1, T_1)\). Conversely, every extension of \((Q_1, T_1)\) is similar to a pair of this form.

**Lemma 2** [9]. Let \( Q \) be \( r \times p \) and \( T \) be \( p \times p \). If \((Q, T)\) is independent with index \( l \) (so that \( nl \geq p \)), then for any integer \( l \geq 1 \) there is an \( r \times r \) monic
matrix polynomial $L_T(\lambda)$ of degree 1 such that $(Q, T)$ is a restriction of each standard pair for $L_T(\lambda)$.

Under the conditions of Lemma 2, it follows from Lemma 1 that the matrix polynomial $L_T(\lambda)$ has a standard pair of the form

$$\hat{Q} = \begin{bmatrix} Q & Q_1 \end{bmatrix}, \quad \hat{T} = \begin{bmatrix} T & T_2 \\ 0 & T_1 \end{bmatrix}.$$ 

The spectrum of $T_1$ is determined by the choice of $L_T(\lambda)$. Hence, we call it the *supplementary spectrum of $T$ with respect to $L_T(\lambda)$* and write $\sigma(T_1) = \sigma_s(T)$. Subsequently, we omit reference to the polynomial $L_T(\lambda)$ when this polynomial is clear from the context. Note, in particular, that extensions of $(Q, T)$ always exist, even with $\sigma_s(T)$ a singleton outside $\sigma(T)$—a situation of special interest in [9].

3. TRANSFORMATION OF THE EQUATION $AX - XB = C$

Consider the equation in complex matrices

$$AX - XB = C \quad (3.1)$$

where $A$ is $m \times m$, $B$ is $n \times n$, and $C$ is $m \times n$. We are interested in $m \times n$ solution matrices $X$. Our first result is formulated under the hypotheses that $(C^*, A^*)$ and $(C, B)$ are independent pairs. In the second theorem the independence hypotheses will be relaxed.

In these results $r$ will always denote the rank of $C$ and $C = U^*V$ is a full-rank decomposition of $C$, i.e., $U$ and $V$ are full rank (or right invertible) matrices of sizes $r \times m$ and $r \times n$, respectively. It is easily seen that the independence of $(C^*, A^*)$ and $(C, B)$ implies that of $(U, A^*)$ and $(V, B)$, respectively. Note that Lemma 2 plays an important part in the formulation of Theorem 3, and the companion matrices appearing in (3.2) are as defined in (2.1).

**Theorem 3.** Let $(C^*, A^*)$ and $(C, B)$ be independent pairs and rank $C = r$. Let $C = U^*V$ be a full-rank decomposition of $C$, and let $L_A(\lambda), L_B(\lambda)$ be $r \times r$ monic matrix polynomials of degree $l_A, l_B$ respectively such that $(V, B)$ is a restriction of any standard pair of $L_B(\lambda)$, and $(A, U^*)$ is a restriction of
any left standard pair of \( L_A(\lambda) \). Consider the equation

\[
\hat{C}_{L_A} X_0 - X_0 C_{L_B} = \begin{bmatrix} I_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.
\] (3.2)

(a) If (3.2) has a solution \( X_0 \), then (3.1) has a solution

\[
X = \text{row} \left[ A^i U^* \right]_{i=0}^{n-1} X_0 \text{col} \left[ V B^j \right]_{j=0}^{n-1}.
\] (3.3)

(b) If the supplementary spectra of \( A \) and \( B \) are chosen so that

\[
\emptyset = \sigma_s(B) \cap \sigma_s(A) = \sigma_s(A) \cap \sigma_s(B),
\] (3.4)

then Equations (3.1) and (3.2) are equivalent and all solutions are related by (3.3).

Before proving the theorem, a number of remarks can usefully be made. Note first that the matrix on the right of (3.2) has blocks of size \( r \times r \), and is itself of size \( l_A r \times l_B r \). Also, \( l_A \) and \( l_B \) can take any integer values not less than the indices of the pairs \((C^*, A^*)\) and \((C, B)\), respectively.

Observe also that consideration of the homogeneous equation \([C = 0\] in (3.1)] is excluded by the independence hypotheses. Another extreme case, in which \( C \) is nonsingular, is also of no interest, for one could then take \( U = I \) and \( V = C \), and it is easily seen that \( L_{BC}(\lambda) = I_\lambda - BC^{-1}, \ L_A(\lambda) = I_\lambda - A \), and (3.3) becomes \( X = X_0 C \). Thus, it is the cases in which \( C \) is singular but nonzero which are relevant in this analysis.

Now observe that a solution of the equation (3.2) can be found by solving a vector equation \( S x = \rho \) with a \( l r^2 \times l r^2 \) coefficient matrix \( S \), where \( l = \max(l_A, l_B) \); this result will be proved in Section 6. Using the bound (2.2), we see that the size of \( S \) can be estimated from above by the integer \( N = (n-r+1)r^2 \) (we assume, for simplicity, \( n = m \)). It is interesting to note that for small values of \( r \) the integer \( N \) is less than \( n^2 \), which is the size of the nivellateur corresponding to (3.1).

Proof. Since \((V, B)\) is a restriction of a right standard pair of \( L_B(\lambda) \), it follows from Lemma 1 that \( L_B(\lambda) \) has a standard pair of the form

\[
V_B = \begin{bmatrix} V & V_1 \end{bmatrix}, \quad T_B = \begin{bmatrix} B & B_2 \\ 0 & B_1 \end{bmatrix}.
\] (3.5)
Similarly, \( L_A(\lambda) \) has a left standard pair of the form

\[
T_A = \begin{bmatrix} A & 0 \\ A_2 & A_1 \end{bmatrix}, \quad U_A = \begin{bmatrix} U^{*} \\ U_1^{*} \end{bmatrix}.
\] (3.6)

Denote \( F_A = \text{row}(T_A^{i=1} U_A^{i=1}) \), \( F_B = \text{col}(V_B T_B^{i=1} V_B^{i=1}) \), and note that \( F_A, F_B \) are nonsingular and satisfy

\[
F_A T_{L_A} = T_A F_A, \quad U_A = F_A \begin{bmatrix} I & 0 & \cdots & 0 \end{bmatrix}^T,
\] (3.7)

\[
C_{L_B} F_B = F_B T_B, \quad V_B = \begin{bmatrix} I & 0 & \cdots & 0 \end{bmatrix} F_B.
\]

Multiplying (3.2) by \( F_A \) and \( F_B \) from the left and right respectively, and using (3.7), we rewrite (3.2) in an equivalent form

\[
T_A \tilde{X} - \tilde{X} T_B = U_A V_B
\] (3.8)

with

\[
\tilde{X} = F_A X_0 F_B.
\] (3.9)

Substituting from (3.5) and (3.6) in (3.8) and decomposing \( \tilde{X} \) in blocks of suitable sizes, we rewrite (3.8) as

\[
\begin{bmatrix} A & 0 \\ A_2 & A_1 \end{bmatrix} \begin{bmatrix} X & 0 & X_1 \\ X_2 & X_3 \end{bmatrix} - \begin{bmatrix} X & 0 & X_1 \\ X_2 & X_3 \end{bmatrix} \begin{bmatrix} B & B_2 \\ 0 & B_1 \end{bmatrix} = \begin{bmatrix} U^{*} V & U^{*} V_1 \\ U_1^{*} V & U_1^{*} V_1 \end{bmatrix}.
\] (3.10)

Comparing the entries in the left upper corner on both sides of this equation, we infer that the matrix \( X \) is a solution of (3.1).

Representing

\[
F_A = \begin{bmatrix} \text{row}(A^{i=1} U^*) \\ \hat{F}_A \end{bmatrix},
\]

\[
F_B = \begin{bmatrix} \text{col}(V B^{i=1}) \\ \hat{F}_B \end{bmatrix}
\]

with some \( \hat{F}_A \) and \( \hat{F}_B \), we obtain (3.3) from (3.9).
Now, conversely, let $X$ be a solution of (3.1), and assume that the hypotheses of part (b) are met. Then $\sigma(A) = \sigma(A_1)$ and $\sigma(B) = \sigma(B_1)$ so that (3.4) implies $\sigma(A_1) \cap \sigma(B) = \emptyset$ and $\sigma(B_1) \cap \sigma(A) = \emptyset$. This ensures the existence of matrices $X_1$ and $X_2$ (of appropriate sizes) such that

\[
AX_1 - X_1B_1 = U^*V_1 + XB_2, \quad (3.11)
\]
\[
A_1X_2 - X_2B = U_1^*V - A_2X. \quad (3.12)
\]

In addition, (3.4) also implies $\sigma(A_1) \cap \sigma(B_1) = \emptyset$ and therefore a matrix $X_3$ exists such that

\[
A_1X_3 - X_3B_1 = U_1^*V_1 - A_2X_1 + X_2B_2. \quad (3.13)
\]

Now Equations (3.11)–(3.13) together with (3.1) can be rewritten in a concise way as (3.10). Reversing the argument of the first part of the proof, we find that (3.10) is equivalent to (3.2).

Now we relax the assumptions concerning the independence of the pairs $(A, C)$ and $(C, B)$ but retain the assumption that $C \neq 0$ throughout. Let us agree to denote by $A_0$ the restriction of $A$ to the controllable subspace $\text{Im} \text{row}(A^t\text{inv}^{-1}U^*)_{i=1}^\infty$ (which is $A$-invariant) and let $B_0^*_0$ denote the restriction of $B^*$ to the $B^*$-invariant orthogonal complement of the unobservable subspace $\text{Ker} \text{col}(VB^t)^{\infty}_{i=1}$. It is not difficult to see that there exist nonsingular matrices $S$ and $K$ such that

\[
S^{-1}BS = \begin{bmatrix} B_0 & 0 \\ B_1 & \hat{B} \end{bmatrix}, \quad VS = \begin{bmatrix} V_0 & 0 \end{bmatrix}, \quad (3.14)
\]
\[
K^*AK^{-1} = \begin{bmatrix} A_0 & A_1 \\ 0 & \hat{A} \end{bmatrix}, \quad K^*U^* = \begin{bmatrix} U_0^* \\ 0 \end{bmatrix}, \quad (3.15)
\]

and the pairs $(V_0, B_0)$ and $(A_0, U_0^*)$ are independent. Such representations were extensively used in [19] for analysis of the equation $AX + XA^* = C$.

If $l_A, l_B$ now denote the indices of $(A_0, U_0^*)$ and $(B_0, V_0)$ respectively, then using Lemma 2 again, we construct $r \times r$ monic matrix polynomials $L_{A_0}(\lambda)$ and $L_{B_0}(\lambda)$ of degrees $l_A, l_B$ respectively, such that $(V_0, B_0)$ is a restriction of any standard pair of $L_{B_0}(\lambda)$ and $(A_0, U_0^*)$ is a restriction of any left standard pair of $L_{A_0}(\lambda)$. 
THEOREM 4. Given matrices $A$, $B$, and $C$ of sizes $m \times m$, $n \times n$, and $m \times n$, respectively, complete the above construction and assume that

\[ \varnothing = \sigma(\hat{A}) \cap \sigma(B) = \sigma(\hat{B}) \cap \sigma(A). \]  

(3.16)

If an $l_A r \times l_B r$ solution $X_0$ of

\[ \hat{C}_{L_{A_0}} X_0 - X_0 C_{L_{B_0}} = \begin{bmatrix} I_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \]  

(3.17)

exists, then (3.1) is solvable and a solution $X$ of (3.1) is given by (3.3).

Furthermore, if the supplementary spectra of $A_0$ and $B_0$ are chosen so that

\[ \varnothing = \sigma_*(B_0) \cap \sigma(A_0) = \sigma_*(A_0) \cap \sigma(B_0) = \sigma_*(A_0) \cap \sigma_*(B_0), \]

then (3.1) is solvable if and only if (3.17) is solvable, and any solution of (3.1) has the representation (3.3).

Proof. Eq. (3.1) can be rewritten as

\[ \begin{bmatrix} A_0 & A_1 \\ 0 & \hat{A} \end{bmatrix} \hat{X} - \hat{B} \begin{bmatrix} B_0 & 0 \\ B_1 & \hat{B} \end{bmatrix} = \begin{bmatrix} U_0^* V_0 & 0 \\ 0 & 0 \end{bmatrix} \]

with

\[ \hat{X} = K^* XS. \]  

(3.18)

The conditions (3.16) imply that the matrix $\hat{X}$ must have the form

\[ \hat{X} = \begin{bmatrix} X_1 & 0 \\ 0 & 0 \end{bmatrix} \]

with $X_1$ satisfying

\[ A_0 X_1 - X_1 B_0 = U_0^* V_0. \]  

(3.19)
So the solvability of Equation (3.1) is equivalent to the solvability of (3.19). Now to complete the proof, we apply Theorem 3 and obtain

\[ X_1 = \text{row} \left( A_i \left( A_i^{-1} U_i \right) \right) X_0 \text{col} \left( V_j \left( V_j^{-1} \right) \right), \tag{3.20} \]

where \( X_0 \) is the solution of (3.17).

But from (3.18) and (3.20) it follows that

\[ X = K^{*-1} X_1 = K^{*-1} \begin{bmatrix} X_1 & 0 \\ 0 & 0 \end{bmatrix} S^{-1} \]

\[ = K^{*-1} \begin{bmatrix} \tilde{U}^* \\ 0 \end{bmatrix} X_0 \begin{bmatrix} \tilde{V} \\ 0 \end{bmatrix} S^{-1}, \tag{3.21} \]

where

\[ \tilde{U}^* = \text{row} \left( A_i \left( A_i^{-1} U_i \right) \right), \quad \tilde{V} = \text{col} \left( V_j \right). \]

Furthermore, (3.14) and (3.15) yield

\[ \text{row} \left( A_i \left( A_i^{-1} U_i \right) \right) = K^{*-1} \begin{bmatrix} \tilde{U}^* \\ 0 \end{bmatrix}, \]

\[ \text{col} \left( V_j \left( V_j^{-1} \right) \right) = \begin{bmatrix} \tilde{V} \\ 0 \end{bmatrix} S^{-1}, \]

and the required form of the solution follows from (3.21).

\[ \text{Corollary 5. If } \sigma(A) \cap \sigma(B) = \emptyset \text{ and } C \neq 0, \text{ then with a suitable choice of } L_A \text{ and } L_B \text{ the unique solution of (3.1) can be represented in the form (3.3), where } X_0 \text{ is the (unique) solution of (3.17) and all other notation is preserved.} \]

**Proof.** Indeed, the conditions (3.16) are obviously satisfied and, consequently, Theorem 4 can be applied.

Theorem 3 and Corollary 5 show that independence of the pairs \((A, C)\) and \((C, B)\) and uniqueness of the solution of (3.1) [i.e. \( \sigma(A) \cap \sigma(B) = \emptyset \)] are in fact two extremal cases of the general conditions (3.16).
The rank of solution matrices $X$ of (3.1) is a question of some interest (see [13] and [20]), and in the case of a unique solution the representation (3.3) immediately gives:

**Corollary 6.** If $\sigma(A) \cap \sigma(B) = \emptyset$ and $l_A = \text{ind}(C^*, A^*)$, $l_B = \text{ind}(C, B)$, then

$$\text{rank } X \leq \min(r_1, r_2)$$

where

$$r_1 = \text{rank} \text{col}(CB^{-1})^{l_B}_{j=1},$$

$$r_2 = \text{rank} \text{row}(A^{-1}C)^{l_A}_{j=1}.$$

In particular, the nonsingularity of the matrix $X$ implies the controllability and the observability of the pairs $(A, C)$ and $(C, B)$, respectively.

We conclude this section by showing that Theorem 1 from [20] can be deduced as a corollary of Theorem 4.

**Corollary 7.** Let (3.1) have a unique solution, and let $\alpha(\lambda)$ and $\beta(\lambda)$, with degrees $\mu$ and $v$ respectively, be coprime (scalar) monic polynomials such that

$$\alpha(A)C = 0,$$

$$C\beta(B) = 0.$$  \hspace{1cm} (3.22)

Then the unique solution of (3.1) has the representation

$$X = \sum_{j=1}^{v} \sum_{i=1}^{\mu} \gamma_{ij} A^{i-1}C B^{j-1}$$

$$= [U^*, AU^*, \ldots, A^{\mu-1}U^*] [\Gamma \otimes I_r] \begin{bmatrix} V \\ V B \\ \vdots \\ V B^{v-1} \end{bmatrix},$$  \hspace{1cm} (3.23)
where $\Gamma = (\gamma_{ij})_{i,j=1}^{n}$ is the unique solution of

\[
\hat{C}_\alpha \Gamma - \Gamma C_\beta = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{bmatrix}.
\]

**Proof.** First note that Eqs. (3.22) can be rewritten as

\[
\alpha(A)U^* = 0, \quad (3.25)
\]
\[
V\beta(B) = 0. \quad (3.26)
\]

Using Lemma 3 from [20], we observe that (3.25) implies divisibility of $\alpha(\lambda)$ by the minimal polynomial of $A_0$, i.e. $\alpha(A_0) = 0$. Now define a matrix polynomial $L_{A_0}(\lambda) = \alpha(\lambda)I_r$, and observe that, since $U_0^*(\alpha_0I) + A_0U_0^*(\alpha_1I) + \cdots = \alpha(A)U^* = 0$, $(A_0, U_0^*)$ is a restriction of a left standard pair of this polynomial. Analogously $(V_0, B_0)$ is a restriction of a right standard pair of this polynomial. Since $\alpha(\lambda)$ and $\beta(\lambda)$ are coprime, conditions of the type (3.4) are satisfied and we can apply the second assertion of Theorem 4. Note that in our case one can write

\[
C_{L_{A_0}} = C_\beta \otimes I_r, \quad C_{V_0} = C_\alpha \otimes I_r,
\]

and therefore the equation (3.17) can be written as

\[
\begin{bmatrix}
C_\alpha \otimes I_r \\
C_\beta \otimes I_r
\end{bmatrix} X_0 - X_0 \begin{bmatrix}
C_\beta \otimes I_r \\
C_\alpha \otimes I_r
\end{bmatrix} = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{bmatrix} \otimes I_r.
\]

It is obvious that the (unique) solution of this equation is given by $X_0 = \Gamma \otimes I_r$, where $\Gamma$ is the solution of (3.24). Now the formula (3.23) is just a consequence of (3.3).

## 4. TRANSFORMATION OF THE EQUATION $AX + XA^* = C$

The special case of Equation (3.1) when $B = -A^*$, so that

\[
AX + XA^* = C,
\]

(4.1)
and \( C^* = C \) is of special interest. In this case the rank factorization of \( C \) can be symmetric:

\[
C = V^*DV, \quad (4.2)
\]

where \( \text{rank } C = \text{rank } V = r \), \( V \) is \( r \times n \), \( D = \text{diag}[I_{r_1} - I_{r_2}] \) with \( r_1 + r_2 = r \), and \( D = I \) if and only if \( C \) is positive semidefinite. Theorems 3 and 4 can now be applied with \( U^* \) replaced by \( V^*D \). We shall simply make a statement of the special case obtained from the final clause of the more general Theorem 4.

First, we need the construction. Let \( A_0 \) be the restriction of \( A \) to the controllable subspace of \( (A, C) \) [i.e. of \( (A, V^*D) \)]. Then there is a nonsingular \( K \) such that

\[
K^*AK^{-1} = \begin{bmatrix}
A_0 & A_1 \\
0 & \hat{A}
\end{bmatrix}, \quad KV^*D = \begin{bmatrix}
U_0^* \\
0
\end{bmatrix},
\]

where \( (A_0, U_0^*) \) are independent. Let \( l \) be the index of \( (A_0, U_0^*) \) and \( L_{A_0}(\lambda) \) be a monic \( r \times r \) matrix polynomial of degree \( l \) such that \( (A_0, U_0^*) \) is a restriction of any left standard pair of \( L_{A_0}(\lambda) \).

**Theorem 8.** Given an admissible pair \( (A, C) \) of order \( n \) with \( C^* = C \neq 0 \), complete the above construction and assume that

\[
\lambda \in \sigma(\hat{A}) \text{ implies } -\bar{\lambda} \notin \sigma(A). \quad (4.3)
\]

If also the supplementary spectrum of \( A_0 \) is chosen so that \( \lambda \in \sigma(A_0) \) implies \( -\bar{\lambda} \notin \sigma(L_{A_0}(\lambda)) \), then the equations (4.1) and

\[
\hat{C}_{L_{a_0}}X_0 + X_0\hat{C}_{L_{a_0}} = \begin{bmatrix}
D & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{bmatrix} \quad (4.4)
\]

are equivalent, and \( X = Q_A X_0 Q_A^* \), where \( Q_A = \text{row}[A^jV^*D]^{-1}_{j=0} \).

Note that the hypothesis (4.3) is automatically satisfied if \( (A, C) \) is independent, i.e. controllable, and in this case \( A_0 = A \).

**Proof.** First, it is easily verified that the condition (4.3) corresponds to (3.16) in the case that \( B = -A^* \), and so, using the proof of Theorem 4, the problem is reduced to that in which the pair \( (A, C) \) is independent, i.e. controllable, and in this case we may take \( A_0 = A, U_0^* = V^*D \).
We now have the pair \((A, V^*D)\) as a restriction of any left standard pair of \(L_A(\lambda) = \sum_{j=0}^{l} \lambda^j L_j\), and there is a left standard pair of the form

\[
T_A = \begin{bmatrix} A & 0 \\ A_1 & A \end{bmatrix}, \quad U_A^* = \begin{bmatrix} V^*D \\ U_1^* \end{bmatrix}.
\]

Then it is easily verified that \((V, A^*)\) is a restriction of any right standard pair for \(DL_A(\lambda)D\) [and we define \(L_A^*(\lambda) = \sum_{j=0}^{l} \lambda^j L_j^*\)], and hence that \((V, -A^*)\) is a restriction of any right standard pair for the polynomial

\[
M(\lambda) = D\left( \sum_{j=0}^{l} (-1)^l - j \lambda^j L_j^* \right) D. \quad (4.5)
\]

Thus, by Theorem 3, \(AX + XA^* = C\) corresponds to the equation

\[
\dot{C}_{L_A} \tilde{x}_0 - \dot{x}_0 M = \text{diag}[I, 0, \ldots, 0]. \quad (4.6)
\]

Define \(\dot{D} = \text{diag}[D, D, \ldots, D]\) and \(P = \text{diag}[I, -I, \ldots, (-1)^l I]\) and it is easily verified that \(C_M = -DP C_{L_A} D\); consequently,

\[
\dot{C}_{L_A} \tilde{x}_0 + \dot{x}_0 D P C_{L_A} D \dot{D} = \text{diag}[I, 0, \ldots, 0].
\]

Then observe that \(C_{L_A} = \dot{C}_{L_A}\), and introduce a new unknown matrix \(X_0 = \tilde{x}_0 \dot{D} P\) to obtain (4.4).

The relationship between \(X\) and \(\tilde{x}_0\) of Equation (4.6) is, noting (4.5),

\[
X = \text{row} \left[ A^{-1} V^* D \right]_{j=1}^{l} \dot{\tilde{x}}_0 \text{col} \left[ (-1)^{l-j-1} V A^* D \right]_{j=1}^{l}.
\]

Since \(\dot{\tilde{x}}_0 = X_0 \dot{D}\), the conclusion \(X = Q_A X_0 Q_A^*\) is obtained.

It is interesting that for Equation (4.4) the controllability condition is automatically satisfied. Thus, when \(D > 0\) the inertia theorem of Wimmer becomes applicable (see [21]).

We illustrate the technique with an example.

**Example.** Consider the equation (4.1) where

\[
A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}
\]
and write

\[ V = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \]

so that \( C = V^*V \). Note that \( \sigma(A) = \{-1, 1, 2\} \) and that the pair

\[ A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}, \quad V^* = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \]

is independent so that the hypothesis (4.3) need not concern us, and we take \( A_0 = A \) in the theorem. Extend \((A, V^*)\) to the pair

\[ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & \alpha \end{bmatrix}, \quad \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 2 \\ 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \]

where \( \alpha \in \mathbb{R}, \alpha \notin \{-1, 0, -1, -2\} \). This is found to be a left standard pair for

\[ L_A(\lambda) = \begin{bmatrix} \lambda^2 - (\alpha + 1)\lambda + \alpha & 0 \\ 0 & \lambda^2 - \lambda - 2 \end{bmatrix} \]

and \( \sigma(L_A(\lambda)) = \{-1, 1, 2, \alpha\} \), \( \sigma_0(A) = \{\alpha\} \).

The equivalent equation (4.4) is therefore

\[ \begin{bmatrix} 0 & 0 & -\alpha & 0 \\ 0 & 0 & 0 & 2 \\ 1 & 0 & \alpha + 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} X_0 + X_0 \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \alpha & 0 & \alpha + 1 & 0 \\ 0 & 2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \]

and its solutions are related to those of \( AX + XA^* = C \) by \( X = Q_A X_0 Q_A^* \), where

\[ Q_A = \begin{bmatrix} \sqrt{2} & 0 & \sqrt{2} & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}. \]

In fact this can be verified directly, given that the general solutions are (in
terms of parameters $a, b$ and $y, z$, respectively)

$$
X = \begin{bmatrix}
1 & 2a & -a \\
2b & 1 & 1 \\
-b & 1 & 1
\end{bmatrix}, 
X_0 = \begin{bmatrix}
\frac{\alpha^2 + 3\alpha + 1}{2\alpha(\alpha + 1)} & 2ay & -\frac{1}{2\alpha} & -ay \\
2az & \frac{1}{4} & -2z & \frac{1}{4} \\
-\frac{1}{2\alpha} & -2y & \frac{1}{2\alpha(\alpha + 1)} & y \\
-az & \frac{1}{4} & z & -\frac{1}{4}
\end{bmatrix}.
$$

5. FURTHER BACKGROUND INFORMATION

The linear equations $AX - XB = C$ and $X - AXB = C$ are now to be examined in the case of block-companion coefficient matrices $A$ and $B$. It will be shown how they can be transformed to matrix-vector equations with coefficient matrices of block-Hankel or Bezout type. Some more preparations are necessary before the main result can be stated.

Let $A(\lambda), B(\lambda)$ be $r \times r$ monic matrix polynomials of degrees $s$ and $t$ respectively, and assume, without loss of generality, that $t \geq s$. Consider the associated equation

$$
C_A X - X \hat{C}_B = \begin{bmatrix}
0 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & 0 \\
0 & \cdots & 0 & R
\end{bmatrix}, \quad \text{(5.1)}
$$

where $C_A, \hat{C}_B$ are the first and second companion matrices for $A(\lambda)$ and $B(\lambda)$, respectively. Note that the blocks on the right are $r \times r$ and the whole matrix (as well as $X$) has size $sr \times tr$. The dual equation is

$$
X - C_A X \hat{C}_B = \begin{bmatrix}
0 & \cdots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \cdots & 0 & 0 \\
0 & \cdots & 0 & R
\end{bmatrix}, \quad \text{(5.2)}
$$

We shall return to equations of the form (3.2) in Theorem 13.

It is well known that solutions of (5.1) and (5.2) are necessarily of block-Hankel and block-Toeplitz form, respectively. For clarity, we establish some preliminary lemmas related to these remarks.
Let
\[ E_m = \text{diag}[I, I, \ldots, I, 0], \]
where there are \( m \) diagonal blocks of size \( r \times r \), and let \( K_m \) be the block matrix with \( m \times m \) blocks, each of size \( r \times r \), with blocks \( I \) on the superdiagonal and zero blocks elsewhere. A little manipulation serves to establish the first result.

**Lemma 9.** Let \( M \) be an \( mr \times nr \) matrix with \( m \) (block) rows and \( n \) (block) columns. Then:

(a) \( M \) is a block-Hankel matrix if and only if
\[ K_m ME_n = E_m MK_n^T. \]

(b) \( M \) is a block-Toeplitz matrix if and only if
\[ E_m ME_n = K_m MK_n^T. \]

Then to describe the form of solutions of (5.1) and (5.2) we have:

**Lemma 10.** Let \( C_A, \hat{C}_B \) be defined as above, and let \( M \) be any \( sr \times tr \) matrix for which \( E_s ME_t = 0 \). Then:

(a) any solution \( X \) of \( C_A X - X \hat{C}_B = M \) is a block-Hankel matrix.

(b) any solution \( X \) of \( X - C_A X \hat{C}_B = M \) is a block-Toeplitz matrix.

**Proof.** (a): Write \( C_A = K_s + A_0, \hat{C}_B = K_t^T + B_t \), and note that \( E_s K_m = K_m, E_s A_0 = 0, B_t E_t = 0 \). Multiply the relation \( C_A X - X \hat{C}_B = M \) on the left and right by \( E_s, E_t \) respectively to obtain \( K_s X E_t - E_s X K_t^T = 0 \), and use part (a) of Lemma 9.

(b): This is obtained on decomposing \( C_A, \hat{C}_B \) as in part (a) and multiplying the relation \( X - C_A X \hat{C}_B = M \) on left and right by \( E_s, E_t \), respectively.

We remark that the block-Hankel structure of \( X \) in part (a), for example, is no longer valid in the case of the equation \( C_A X + X \hat{C}_B = M \). Before going on to the main results concerning the solutions of (5.1) and (5.2), we introduce the concept of bezoutian matrices. The appropriate form of the bezoutian for this analysis is defined as follows for matrix polynomial \( A(\lambda), B(\lambda) \) as above. With the above convention that \( \deg B(\lambda) = t \geq s = \deg A(\lambda) \) it will be convenient to write \( A(\lambda) = \sum_{j=0}^{t} \lambda^j A_j \) with \( A_{s+1} = \cdots = \)
\( A_t = 0 \) when \( s < t \). Define matrices \( \Gamma_{ij} \) of size \( r^2 \times r^2 \) by the relation
\[
\sum_{i, j=0}^{t-1} \Gamma_{ij} \lambda^i \mu^j = \frac{1}{\lambda - \mu} \left[ B(\lambda) \otimes A(\mu) - B(\mu) \otimes A(\lambda) \right]. \tag{5.3}
\]

Then the bezoutian matrix \( \text{Bez}(B, A) \) defined by \( A(\lambda) \) and \( B(\lambda) \) is the matrix \( [\Gamma_{ij}]_{i,j=0}^{t-1} \). This bezoutian is considered in [2], [14], and [15], and in a slightly different context in [8]. It is, in fact, a special case of a more general definition introduced in [1], [3] and studied in [17] in connection with the general question of least common multiples for matrix polynomials (see [6]).

To see why the bezoutian should play a role in this analysis, consider the well-known fact that the solution of \( AX - XB = C \) is equivalent to that of a matrix vector equation \( \mathcal{B}x = c \) with the nivellateur \( \mathcal{B} = I_n \otimes A - B^T \otimes I_m \), and \( x \) and \( c \) are the stacked vectors of \( X \) and \( C \) respectively. (Recall that, given an \( m \times k \) matrix \( M = [M_1 \ M_2 \ \cdots \ M_k] \), the vector \( [M_1^T \ M_2^T \ \cdots \ M_k^T]^T \) (where \( T \) denotes the transpose) is said to be the vector obtained by stacking the columns of \( M \), or merely the stacked vector of (columns of) \( M \).

Now choose, in (5.3), \( A(\lambda) = \lambda I - A \) and \( B(\lambda) = \lambda I - B^T \), and it is easily verified that the matrix \( \text{Bez}(B, A) \) is just the nivellateur \( \mathcal{B} \) introduced above. This simple remark and the known properties of the bezoutian readily imply some interesting properties (see, for example, [16]) of the map \( X \rightarrow AX - XB \).

More generally, the kernel of the bezoutian of general matrix polynomials \( A(\lambda) \) and \( B(\lambda) \) is known to characterize their common spectral data (see [2] and [17]), and when the linear equation \( AX - XB = C \) has the form (5.1), it is the common spectral data of the underlying polynomials \( A(\lambda) \) and \( B(\lambda) \) which determine general solutions of the equation. Thus, one may expect the bezoutian of \( A(\lambda) \) and \( B(\lambda) \) to play a role in reformulations of the equation. In fact, the recasting of Equation (5.1) in our main result relies heavily on known properties of the bezoutian and its symmetries. We describe the two major results of this kind here.

Let \( S_A \) be the \( sr \times sr \) matrix of block-Hankel form:
\[
S_A = \begin{bmatrix}
A_1 & A_2 & \cdots & \cdots & A_{s-1} & I \\
A_2 & I & & & & 0 \\
\vdots & & & & & \vdots \\
\vdots & & & & & \vdots \\
A_{s-1} & I & & & & \vdots \\
I & & & & & 0
\end{bmatrix},
\]

and \( x \) and \( c \) are the stacked vectors of \( X \) and \( C \) respectively. (Recall that, given an \( m \times k \) matrix \( M = [M_1 \ M_2 \ \cdots \ M_k] \), the vector \( [M_1^T \ M_2^T \ \cdots \ M_k^T]^T \) (where \( T \) denotes the transpose) is said to be the vector obtained by stacking the columns of \( M \), or merely the stacked vector of (columns of) \( M \).)
which we refer to as the symmetrizer for \( A(\lambda) \), since \( S_A C_A \) also has block symmetry about the main diagonal. The symmetrizer \( S_B \) for \( C_B \) is defined similarly.

The roles of \( A(\lambda) \) and \( B(\lambda) \) in the structure of \( \text{Bez}(A, B) \) is demonstrated more clearly in the representation obtained in [2]. For future convenience, we express the result in terms of the pair \( B^T(\lambda) \) and \( A(\lambda) \):

\[
\text{Bez}(B^T, A) = (S_{B^T} \otimes I_r) \sum_{j=0}^{s} C_{R^T} \otimes A_j = \sum_{j=0}^{s} (S_{B^T}C_{R^T}) \otimes A_j. \tag{5.4}
\]

Define the block-Hankel matrix \( H = [H_{i+j-1}]_{i,j=1} \) using the first coefficients of the expansion

\[
B^T(\lambda)^{-1} \otimes A(\lambda) = \sum_{j=0}^{\infty} \lambda^{-j} H_j, \tag{5.5}
\]

valid for all sufficiently large \( |\lambda| \) [recall that \( \deg B(\lambda) - t \geq s = \deg A(\lambda) \)]. Then let \( S = S_{B^T} \otimes I_r \), and it can be proved that

\[
\text{Bez}(B^T(\lambda), A(\lambda)) = ShS. \tag{5.6}
\]

6. EQUATIONS IN COMPANION MATRICES

With the notation and results of the preceding section the main theorem of this part of the paper can now be stated and readily proved. The solution of (5.1) with the assumption \( t \geq s \) is to be transformed to a matrix-vector equation \( Sx = y \) in which \( S \) is a bezoutian matrix or is of block-Hankel form.

**Theorem 11.** Any solution of (5.1) is a block-Hankel matrix of the form \( X = \text{col}(X_1, C_B^{-1}_0)_{t-1} \). If \( x \) denotes the vector obtained by stacking the columns of \( X_1 \), then

\[
Bx = \begin{bmatrix} \rho & 0 & \cdots & 0 \end{bmatrix}^T \quad (t - 1 \text{ zero blocks}), \tag{6.1}
\]

where \( B = \text{Bez}(B^T, A) \) and \( \rho \) denotes the stacked vector of columns of \( -R \).

**Proof.** The block-Hankel form of \( X \) is established in Lemma 10. Taking advantage of this, the next idea is to focus on the first block row of \( X \) and show that the remaining \( s - 1 \) block rows are generated by a simple recursion.
Thus, write $X = \text{col}[X_j]_{j=1}^s$, and it is easily verified directly from (5.1) that for $j = 1, 2, \ldots, s$,

$$X_j = X_j \hat{C}_B^{-1},$$

(6.2)

i.e. $X = \text{col}[X_j \hat{C}_B^{-1}]_{j=1}^s$, and (since $A_s = I_r$)

$$A_0 X_1 + A_1 X_1 \hat{C}_B + \cdots + X_1 \hat{C}_B^s = -[0 \ 0 \ \cdots \ 0 \ R].$$

(6.3)

Using Kronecker products, the equation (6.3) has the equivalent formulation

$$\left( \sum_{j=0}^s C_B^j \otimes A_j \right) x = \rho_1,$$

where $x$ and $\rho_1$ are the vectors obtained by stacking the columns of $X_1$ and $[0 \ \cdots \ 0 \ R]$ respectively. Multiply the last equation from the left by $S = S_B \otimes I_r$, and using (5.4) it is found that

$$\text{Bez}(B^T, A) x = S \rho_1,$$

(6.4)

and (6.1) follows.

**Theorem 12.**

(a) If $t \geq s$, then any solution $X$ of (5.1) is a block-Hankel matrix of the form

$$X = \text{col}[\tilde{X}_i \tilde{C}_B^{-1}]_{i=1}^s S_B^{-1},$$

(6.5)

where the stacked vector $\tilde{x}$ of the matrix $\tilde{X}_1$ is found from the equation

$$H \tilde{x} = [0, \ldots, 0, \rho^T]^T \quad (t - 1 \text{ zero blocks}),$$

(6.6)

$\rho$ is the stacked vector of $-R$, and $H = [H_{i+j-1}]_{i,j-1}$ is a block-Hankel matrix generated by the Laurent expansion of the rational matrix function $W(\lambda) = B^T(\lambda)^{-1} \otimes A(\lambda)$ for sufficiently large $|\lambda|$, 

$$W(\lambda) = \sum_{i=0}^{\infty} \lambda^{-i} H_i.$$ 

(6.7)
(b) If \( s \geq t \), then any solution \( X \) of (5.1) is a block-Hankel matrix of the form

\[
X = S_A^{-1} \text{row} \left[ \hat{C}_A^{j-1} \hat{x}^T \right]_{j=1}^t,
\]

where the stacked vector \( \hat{x} \) of \( \hat{X}_1 \) is found from the equation

\[
H\hat{x} = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}^T \rho^T \quad (s - 1 \text{ zero blocks}),
\]

\( \rho \) is the stacked vector of \( R^T \), and \( H = \left[ H_{i+j-1} \right]_{i,j=1}^s \) is a block-Hankel matrix generated by the Laurent expansion (6.7) for the rational matrix function \( W(\lambda) = A^{-1}(\lambda) \otimes B^T(\lambda) \).

**Proof.** Use the representation (5.6) to get from (6.4) the equivalent equation

\[
HSx = \rho_A = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}^T \rho^T,
\]

where \( H \) is defined using (5.5). Setting \( \tilde{x} = Sx \), we have an equation of the form (6.6). It is easily verified that \( \tilde{x} \) is the stacked vector of columns of the matrix \( \tilde{X}_1 = X_1S_B \). Since \( C_B = S_B^{-1} \hat{C}_B S_B \), it follows that \( \tilde{X}_1C_B^j = X_1\hat{C}_B^j S_B, \) \( j = 0,1,2,\ldots, \) and, in view of (6.2), (6.1) is obtained, and for the case \( t \geq s \) the proof is complete.

The assumption \( t \geq s \) is required by our use of the representation (5.6) and hence of the expansion (5.5). Applying the already proved result to the transpose of Equation (5.1), the second statement of the theorem in the case \( t \leq s \) is easily derived.

The solution of equations of the form

\[
\hat{C}_A X_0 - X_0 C_B = \begin{bmatrix} R & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},
\]

which include those of the first topic of this paper, can also be transformed into simple matrix-vector problems. Consider the relation \( X_0 = S_A X S_B \). Recalling that \( C_B = S_B^{-1} \hat{C}_B S_B \) and \( C_A = S_A^{-1} \hat{C}_A S_A \), it is easily verified that \( X_0 \) satisfies (6.10) if and only if \( X \) is a solution of (5.1). Using this fact we obtain from the
last two theorems:

**Theorem 13.**

(a) If \( t \geq s \), then any solution \( X_0 \) of (6.10) is given by

\[
X_0 = S_A \operatorname{col} \left[ \tilde{X}_1 C_B^{(1)} \right]_{j=1}^t,
\]

and \( \tilde{X}_1 \) is defined by (6.6).

(b) If \( s \geq t \), then any solution of (6.10) is given by

\[
X_0 = \operatorname{row} \left[ \hat{C}_A^{(1)} \right]_{j=1}^t S_B,
\]

where \( \tilde{X}_1 \) is defined by (6.9).

Finally, consider equations in companion matrices with the symmetry of Equation (4.4). Thus, for a second companion matrix \( \hat{C}_A \) associated with monic matrix polynomial \( A(\lambda) = \sum_{j=0}^l \lambda A_j \) we examine

\[
\hat{C}_A X + X \hat{C}_A^* = \operatorname{diag}[R, 0, \ldots, 0].
\]

Taking advantage of the device used in Equations (4.5), (4.6), let

\[
B(\lambda) = (-1)^j A^*(-\lambda) = (-1)^j \sum_{j=0}^l (-\lambda)^j A_j^*,
\]

and observe that \( \hat{C}_A^* = -P C_B P \), where \( P = \operatorname{diag}[I, -I, \ldots, (-1)^l I] \). Thus, (6.13) is equivalent to

\[
\hat{C}_A X_0 - X_0 C_B = \operatorname{diag}[R, 0, \ldots, 0],
\]

where \( X_0 = XP \). This equation is of the form (6.10) so that Theorem 13 can be applied. Observing that \( S_B = (-1)^l PS_A P \), part (b) of Theorem 13 gives:

**Theorem 14.** Let \( \hat{C}_A \) be the second companion matrix of an \( n \times n \) monic matrix polynomial \( A(\lambda) = \sum_{j=0}^l \lambda A_j \), and \( R \) be any \( n \times n \) hermitian matrix. Then any solution of equation (6.13) is given by

\[
X = (-1)^{l-1} \operatorname{row} \left[ \hat{C}_A^{(1)} \right]_{j=1}^l PS_A^*,
\]

(6.14)
FACTORED FORMS FOR SOLUTIONS

where $P = \text{diag}[I, -I, \ldots, (-1)^{t-1}I]$ and $\tilde{X}_1$ is given by (6.9), with

$$
\sum_{j=0}^{\infty} \lambda^{-j} H_j = A^{-1}(\lambda) \otimes (-1)^t \tilde{A}(-\lambda)
$$

for all sufficiently large $|\lambda|$ (and $\tilde{A}(\lambda) = \sum_{j=0}^{t} \lambda^j \tilde{A}_j$).

(Note that $\tilde{A}$ is the matrix whose elements are the conjugates of those of $A$.) It should also be mentioned that results similar to those of Theorems 13 and 14, but expressed in terms of bezoutian matrices, can be easily established.

7. SYMMETRIC FORM OF THE SOLUTIONS

Substituting (6.11) or (6.12) in (3.3) and assuming that the conditions of Theorem 3 are satisfied, the solution of the equation $AX - XB = C$ can be reduced to solving an appropriate equation of the form $Hx = y$ with a block-Hankel matrix $H$. Using this idea, we present in this section the symmetric representation of the solution of $AX - XB = C$. This form of the solution is easily derived by combining Theorems 3 and 12, Corollary 5, and the formula $X_0 = S_A X S_B$, which provides a linkage between the solutions $X$ and $X_0$ of the equations (5.1) and (6.10), respectively.

THEOREM 15. Let the equation $AX - XB = C$ have a unique solution, or let the pairs $(A, C)$ and $(C, B)$ be controllable and observable, respectively. If the polynomials $L_A(\lambda)$, $L_B(\lambda)$ of degrees $s$ and $t$, respectively, are defined as in Theorem 3, and if $X$ denotes a solution of (3.1), then

$$
X = \text{row}(A_i^{-1}U^*)S_{L_A}X S_{L_B}\text{col}(V B_i^{-1})^t_{-1},
$$

where $U^*V = C$ and the block-Hankel matrix $\hat{X}$ is determined by Theorem 12 with $R = I_r$.

Consider the Lyapunov equation (4.1). Preserving the notation of Section 4 and putting $B = -A^*$, we choose $L_B(\lambda) = M(\lambda)$ and hence

$$
S_{L_B} = (-1)^{t-1} \hat{D} P S_{L_A} \hat{P} \hat{D}.
$$
Furthermore, \( U^* = V^* D \),
\[
P^D \text{col}(VB^{i-1})_{i=1}^l = (\text{row}(A^{i-1}V^* D)_{i=1}^l)^*,
\]
and then by Theorem 15
\[
X = (-1)^{l-1} Q \hat{X}DPQ^*,
\]
where
\[
Q = \text{row}(A^{i-1}V^* D)_{i=1}^s S_{LA}
\]
and \( \hat{X} \) is a solution of (4.6). Recalling that \( C_M \) in (4.6) is the matrix
\[
\hat{D}P \hat{C}_{LA}^* P \hat{D},
\]
the following consequence of Theorem 15 is derived.

**Theorem 16.** If the equation \( AX + XA^* = C \) has a unique solution, or if
the pair \((A, C)\) is controllable, then with a suitable choice of \( L_A(\lambda) \) any
solution \( X \) is given by
\[
X = (-1)^{l-1} Q \hat{X}Q^*
\]
where \( \hat{X} \) is a solution of the equation
\[
C_{LA} \hat{X} + \hat{X}C_{LA}^* = \begin{bmatrix}
D & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{bmatrix}.
\]

8. **THE EQUATION** \( X - AXB = C \)

Results analogous to those in Sections 3, 4, 6, and 7 apply to equations of
the form
\[
X - AXB = C \quad (8.1)
\]
and, in particular, the equation (5.2). Since there are only technical differ-
ces in proofs, we restrict ourselves to an explicit presentation of ana-
logues of Theorems 3, 12, and 16.

Recall that Equation (8.1) is solvable for any right-hand side \( C \) if and only
if \( 1 \not\in \sigma(A) \circ \sigma(B) \), where
\[
\sigma(A) \circ \sigma(B) = \{ \gamma = \lambda \mu | \lambda \in \sigma(A), \mu \in \sigma(B) \}.
\]
FACTORED FORMS FOR SOLUTIONS

This explains the conditions on the spectra in the next theorem which replace the conditions (3.4) in Theorem 3.

**Theorem 17.** Let the notation of Theorem 3 be preserved, and let the pairs \((A, C)\) and \((C, B)\) be independent. Consider the equation

\[
X_0 - C_L A X_0 C_L B = \text{diag}[I_r, 0, \ldots, 0].
\]  

(8.2)

(a) If (8.2) has a solution \(X_0\), then (8.1) has a solution in the form (3.3).

(b) If the supplementary spectra are chosen so that

\[1 \notin \sigma_s(B) \circ \sigma(A), \quad 1 \notin \sigma_s(A) \circ \sigma(B), \quad 1 \notin \sigma_s(A) \circ \sigma_s(B),\]

then Equations (8.1) and (8.2) are equivalent and all solutions are related by (3.3).

Results similar to those of Section 6 can also be formulated for the solutions of (5.2) or of (8.2). We present one result of this kind.

**Theorem 18.** If \(t \geq s\), then any solution \(X\) of (5.2) is a block-Toeplitz matrix of the form

\[
X = \text{col}\left[ \tilde{X} \tilde{C}_B^{-i} \right]_{i=1}^{s} S_B^{-1}
\]  

(8.3)

where the stacked vector \(\tilde{x}\) of the matrix \(\tilde{X}_s\) is found from the equation

\[
H\tilde{x} = \begin{bmatrix} 0 & \cdots & 0 & \rho^T \end{bmatrix}^T,
\]  

(8.4)

\(\rho\) is the stacked vector of \(R\), \(H = [H_{i+j-1}]_{i,j=1}^s\) is a block-Hankel matrix generated by the Laurent expansion of the rational matrix function \(W(\lambda) = B^T(\lambda)^{-1} \otimes A_\infty(\lambda)\), and \(A_\infty(\lambda) = \lambda^t A(\lambda^{-1})\).

Note that the solutions \(X\) of (5.2) and \(X_0\) of (8.2) are related by \(X_0 = S_{L_A} X S_{L_B}\).

Putting \(B = A^*\) in (5.2), it is found that any solution \(X\) of the equation

\[
X - C_\Lambda X C_A^* = \text{diag}[0, 0, \ldots, 0, I]
\]  

(8.5)

(which plays an important role in the theory of autoregressive processes) is of the form

\[
X = \text{col}\left[ \tilde{X}_i C_A^{-i} \right]_{i=1}^{s} S_A^{-1},
\]  

where \(\tilde{X}_i\) is evaluated on solving (8.4).
Note that solving Equation (5.2) is in a certain sense equivalent to the inverse problem for Toeplitz matrices, which consists, briefly, in constructing a block-Toeplitz matrix via its (and its transposed) images on a relatively small number of vectors. More precisely, the following connection holds.

**Proposition 19.** Let $\tilde{X} = [X_{i-j}]_{i,j=0}^{s,t}$ be a block-Toeplitz matrix such that

$$
\begin{bmatrix}
A_0 & A_1 & \cdots & A_{s-1} & I
\end{bmatrix} \tilde{X} = \begin{bmatrix}
0 & \cdots & 0 & R
\end{bmatrix}
$$

and

$$
\tilde{X} = \begin{bmatrix}
B_0 \\
B_1 \\
\vdots \\
B_{t-1} \\
I
\end{bmatrix} = \begin{bmatrix}
0 \\
0 \\
\vdots \\
0 \\
R
\end{bmatrix}.
$$

Then the truncated matrix $X = [X_{i-j}]_{i,j=0}^{s,t-1}$ is a solution of (5.2).

Conversely, if $X = [X_{i-j}]_{i,j=0}^{s,t-1}$ is a solution of (5.2), then the matrix $\tilde{X} = [X_{i-j}]_{i,j=0}^{s,t-1}$ with

$$
X_{-t} = -X_{-t+1}B_{t-1} - \cdots - X_{-1}B_1 - X_0B_0,
$$

$$
X_s = -A_{s-1}X_{s-1} - \cdots - A_1X_1 - A_0X_0
$$

satisfies equations (8.6)–(8.7).

Note that the inverse problem for Toeplitz matrices appears in some cases in system theory as the inverse problem of an autoregressive process. In fact, Proposition 19 in the special case

$$
s = t, \quad \tilde{X} = \tilde{X}^*, \quad R = R^* > 0, \quad B(\lambda) = A^*(\lambda)
$$

is pointed out in [11], where it is applied to finding the ladder canonical realization of the autoregressive model given by $I, A_{t-1}, \ldots, A_0$. Different algorithms for computing the unique solution in the case (8.8) are also presented in [11]. We note that Theorem 15 shows the possibility of using the generalized Levinson-Szegö algorithm for solving problems of this kind.

Now note that in the scalar case, when the size of blocks $r = 1$ (and $s = t$), the inverse problem for Toeplitz matrices defined by (8.6)–(8.7) is actually solved in [10] (see also [5, Chapter 3, Section 6]), where an explicit formula for $\tilde{X}$ (and $X$) is given via the parameters $A_i$ and $B_i$. Using these results and Proposition 19, we obtain the following explicit formula for the solution of
FACTORED FORMS FOR SOLUTIONS

In this statement $B_{\infty}(\lambda)$ is the scalar polynomial whose zeros are the reciprocals of those of $\bar{B}(\lambda)$. Thus, if $B(\lambda) = \sum_{j=0}^{r} \lambda^j b_j$, then $B_{\infty}(\lambda) = \sum_{j=0}^{r} \lambda^{-j} b_{-j}$.

PROPOSITION 20. Equation (5.2) (with $r = 1$ and $s = t$) is solvable if and only if the (scalar) polynomials $A(\lambda)$ and $B_{\infty}(\lambda)$ have no common zeros. In this case the solution $\hat{X}$ of (5.2) is invertible and is given by

$$\hat{X} = \left[ \begin{array}{cccc} 1 & & & \\ \lambda^{s-1} & \ddots & & \\ & \ddots & \ddots & \\ & & \lambda & \lambda^{s-1} \end{array} \right]$$

$$\times \left[ \begin{array}{cccc} 1 & B_{s-1} & \cdots & B_1 \\ \cdots & \ddots & \ddots & \vdots \\ \cdots & & \ddots & B_{s-1} \\ \cdots & \cdots & \cdots & 1 \end{array} \right]$$

$$- \left[ \begin{array}{ccc} B_0 & & \\ B_1 & \ddots & \\ \vdots & \ddots & \ddots \\ B_{s-1} & \cdots & B_1 \end{array} \right]$$

$$\times \left[ \begin{array}{cccc} A_0 & A_1 & \cdots & A_{s-1} \\ \cdots & \ddots & \ddots & \vdots \\ \cdots & \cdots & \ddots & A_1 \\ \cdots & \cdots & \cdots & A_0 \end{array} \right]^{-1} R^2. \quad (8.9)$$

Note that the matrix in brackets can be inverted by fast algorithms, since its displacement rank is equal to two (see e.g. [4]).

We sketch the proof of Proposition 20. If $X$ is a solution of (5.2) then, by Proposition 19, we obtain a Toeplitz matrix $\hat{X}$ satisfying (8.6) and (8.7). Then use Theorem III.6.2 of [5] to deduce that $X$ is invertible and is given by (8.9). But it is known (see [17]) that (8.9) is the bezoutian matrix associated with
A(λ) and B∞(λ), and its invertibility implies that A(λ) and B∞(λ) have no common zeros. Use the same ideas to prove the converse statement.

Using Theorem 17 and the relations \( C_B = S_B^{-1} \hat{C}_B S_B \), \( C_A = S_A^{-1} \hat{C}_A S_A \), one can deduce from Proposition 20 the following result concerning Equation (8.1) in case \( n = m \), i.e., when \( C \) is a square matrix.

**Corollary 21.** Let the pairs \((A, C)\) and \((C, B)\) be independent and rank \( C = 1 \). Then Equation (8.1) is solvable if and only if \( 1 \notin \sigma(A) \cap \sigma(B) \). In this case the (unique) solution \( X \) of (8.1) is invertible and is found as

\[
X = \text{row}(A^i U^*)_{i=0}^{n-1} S_A \hat{X} S_B \text{col}(V B^i)_{i=0}^{n-1},
\]

where \( C = U^* V \) is the rank factorization of \( C \), \( \hat{X} \) is defined by (8.9), and the (scalar) coefficients \( A_j \) and \( B_j \) in (8.9) are found as follows:

\[
\begin{bmatrix}
A_0 \\
A_1 \\
\vdots \\
A_{n-1}
\end{bmatrix} = \left[ \text{row}(A^i U^*)_{i=0}^{n-1} \right]^{-1} A^n U^*
\]

\[
\begin{bmatrix}
B_0 & B_1 & \cdots & B_{n-1}
\end{bmatrix} = VB^n \left[ \text{col}(V B^i)_{i=0}^{n-1} \right]^{-1}.
\]

The general case of \( r > 1 \) and \( s \neq t \) is much more complicated and will be discussed in future publications. In conclusion, notice that results similar to Propositions 19 and 20 and Corollary 21 hold true for Equation (5.1) and the corresponding inverse problem for block-Hankel matrices. Note also that the problem of invertibility of solutions of (1.1) has been investigated in [13]. It is shown there, in particular, that when rank \( C = 1 \) and the pairs \((A, C)\) and \((C, B)\) are independent, the solutions of a consistent equation are invertible. A result parallel to Corollary 21 (for a Lyapunov-type equation) shows that, in this case, (1.1) is consistent if and only if \( \sigma(A) \cap \sigma(B) = \emptyset \).

**References**


Received 28 June 1982; revised 18 July 1983