Symmetric, Positive Semidefinite, and Positive Definite Real Solutions of $AX = XA^T$ and $AX = YB$

Antony Jameson
School of Engineering
Princeton University
Princeton, New Jersey 08541

Eliezer Kreindler
Department of Electrical Engineering
Technion— Israel Institute of Technology
Haifa 32000, Israel

and

Peter Lancaster*
Department of Mathematics and Statistics
The University of Calgary
Calgary, Alberta, Canada T2N 1N4

Submitted by Richard A. Brualdi

ABSTRACT

Given the equations $AX = XA^T$ and $AX = YB$ with arbitrary nonzero real matrices A and B of the same size, we seek all real solutions X and Y which are: (1) symmetric, (2) symmetric and positive semidefinite, and (3) symmetric and positive definite. Necessary and sufficient conditions for the existence of such solutions and their general forms are derived.

1. INTRODUCTION

An inverse problem of linear optimal control requires the solution of equations of the form

$$AX = YB \tag{1.1}$$

*Research supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

655 Avenue of the Americas, New York, NY 10010 0024-3795/92/$3.50

where A, B are real $n \times m$ matrices, and the unknowns X and Y are required to be real and symmetric, and possibly positive definite or semidefinite. No assumption is made about the relative sizes of m and n; however, it will be convenient to assume throughout that $A \neq 0$ and $B \neq 0$. Throughout the paper *definite and semidefinite matrices are assumed to be real and symmetric*. See [9] for partial solutions to problems of this kind and for some history of the problem in the control-theoretic context. Here, the results of [9] will be extended and presented in algebraic form.

The strategy adopted here is to first solve for the $m \times m$ matrix X, and then for the $n \times n$ matrix Y in terms of X. Multiplication of (1.1) on the left by B^T shows at once that for a symmetric Y, X must be such that B^TAX is symmetric. Hence, we first solve for the symmetric solutions X of the equation

$$MX = XM^T, \tag{1.2}$$

where, in this case, M is the $m \times m$ real matrix B^TA. Nonsingular and nonsingular symmetric solutions of Equation (1.2) are considered by Taussky and Zassenhaus [16].

The equations (1.1) and (1.2) are, respectively, special cases of the well-known equations

$$AX - YB = C \tag{1.3}$$

and

$$AX - XB = C, \tag{1.4}$$

treated in books [7, 12, 13] and in many papers [1, 3, 6, 8, 10, 11, 15], to mention a few. For early references—to Frobenius (1878), Sylvester (1884), and Cayley (1885), among others—see [13]. More recent results do not easily apply to the problem of finding *symmetric* solutions X and Y of (1.1), and we therefore give an independent self-contained analysis. However, the interested reader may wish to consult recent papers by Don [5] and Chu [4] treating (among other topics) symmetric solutions of $AX = B$, which, in particular, contain general solutions consistent with results obtained here for our problem. [Indeed, Equation (1.1) may be seen as a special case of $AX = B$ in which B has a special form.] Also, the interesting and wide-ranging paper of Magnus [14] (Theorem 4.1, in particular) includes analysis of the symmetric solution matrices X and Y of Equation (1.3). However, problems with definiteness conditions on the solution matrices are not considered in any of these papers. We also remark that general solutions of $AX = B$ derived in [9] seem to have been overlooked in many subsequent papers.

In Section 2 we consider solutions of (1.2) that are symmetric, positive semidefinite, or positive definite. In Section 3 we introduce the notion of matrices X that are admissible for the solution of (1.1), i.e., that satisfy further conditions guaranteeing the existence of symmetric matrices Y for which (1.1) is fulfilled with $X^T = X$, $X \geqslant 0$, and $X > 0$. Section 4 is then devoted to the description of the solution pairs ($X^T = X$, $Y^T = Y$), ($X \geqslant 0$, $Y^T = Y$), and ($X > 0$, $Y^T = Y$). Section 5 contains a key lemma, and in Sections 6 and 7, pairs with $Y \geqslant 0$ and $Y > 0$, respectively, are discussed.

2. SOLUTIONS OF $MX = XM^T$

We first obtain the general solution X of Equation (1.2), where M is a real $m \times m$ matrix, and then specialize to solutions which are real and symmetric, positive semidefinite, and positive definite. Let J be a Jordan normal form for M, and

$$M = VJV^{-1}. \tag{2.1}$$

With each Jordan block, say of size three, associate the permutation matrix

$$P_0 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \tag{2.2}$$

and similarly for other sizes. Then define a block-diagonal matrix P of permutation matrices P_0 each with the size of the corresponding Jordan block of J. Then $PJ = J^TP$, $P^2 = I$, and $J^T = PJP$.

Now the equation $MX = XM^T$ is equivalent to

$$VJV^{-1}X = XV^{-T}PJPV^T,$$

where $V^{-T} = (V^T)^{-1}$, or

$$J(V^{-1}XV^{-T}P) = (V^{-1}XV^{-T}P)J.$$

Defining $W = V^{-1}XV^{-T}P$, we have

$$X = V(WP)V^T. \tag{2.3}$$

Then, using (2.1), (2.3), and $PJ^T = JP$, the equation $MX = XM^T$ is equivalent to

$$JW = WJ.$$

We thus have

LEMMA 2.1. *All solutions X of $MX = XM^T$ are characterized by Equation* (2.3) *where W varies over all matrices commuting with J.*

We next specialize to real solutions $X^T = X$, $X \geqslant 0$, and $X > 0$. To that end we may assume that J has the block-diagonal form

$$J = \operatorname{diag}[J_r, J_c, \bar{J}_c] \tag{2.4}$$

where J_r is a real Jordan matrix, J_c has all its eigenvalues in the open upper half of the complex plane, and $\bar{J}_c$ is its complex conjugate; the corresponding matrix P is

$$P = \operatorname{diag}[P_r, P_c, P_c], \tag{2.5}$$

where again P_r and P_c are block-diagonal matrices of permutation matrices of the type (2.2), each with the size of the corresponding Jordan block in J_r, J_c, and $\bar{J}_c$. Also, with a consistent partitioning,

$$V = [V_r, V_c, \bar{V}_c], \tag{2.6}$$

where V_r is real and the columns of V are made up of eigenvectors and generalized eigenvectors of M.

The detailed structure of the matrices W will be important in this analysis, and the first observation is that they have block-diagonal form

$$W = \operatorname{diag}[W_r, W_1, W_2] \tag{2.7}$$

consistent with that of J in (2.4). For the more detailed structure of W_r, W_1, and W_2 we refer to Section 12.4 of [12]. They are necessarily block-diagonal matrices if the corresponding blocks J_r, J_c, or $\bar{J}_c$ (respectively) are either diagonal or nonderogatory. The nonzero blocks of W_r, W_1, and W_2 corresponding to eigenvalues with nonlinear elementary divisors have upper triangular Toeplitz structure, so that the corresponding blocks in WP have

Hankel form. It is important to note that such a matrix (of size two or more) can never be positive definite, but may be positive semidefinite.

Observe also that we may write

$$V^T = \begin{bmatrix} V_r^T \\ V_c^T \\ \overline{V}_c^T \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & I \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ V_c^* \\ \overline{V}_c^* \end{bmatrix} = KV^*, \tag{2.8}$$

where ( )* denotes the conjugate transpose and

$$K = \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & I \\ 0 & I & 0 \end{bmatrix}.$$

Then Equation (2.3) takes the form of a congruence:

$$X = V(WPK)V^*, \tag{2.9}$$

where

$$WPK = \begin{bmatrix} W_r P_r & 0 & 0 \\ 0 & 0 & W_1 P_c \\ 0 & W_2 P_c & 0 \end{bmatrix}. \tag{2.10}$$

Thus, $X \geqslant 0$, or $X > 0$, if and only if $WPK \geqslant 0$, or $WPK > 0$, respectively. We immediately see that $W_1 = W_2 = 0$ is a necessary condition for $X \geqslant 0$, and that $X > 0$ requires that all the spectrum of M be real and that all elementary divisors be linear. In this case we have $P_r = I$.

These considerations lead to the following result (note that a matrix is said to be *simple* if all elementary divisors are linear, i.e. if it is similar to a diagonal matrix):

THEOREM 2.1

(a) *The equation* $MX = XM^T$ *with* M *real has real nonzero symmetric solutions* X. *Furthermore, writing* $M = VJV^{-1}$, *where* J *is a Jordan matrix of the form* (2.4), *all such solutions have the form*

$$X = V(WP)V^T,$$

where $W = \text{diag}[W_r, W_1, W_2]$ *commutes with* J, P *is the permutation defined by* (2.5), $W_r P_r$ *is real symmetric,* $W_2 = \overline{W}_1$, *and* $W_1 P_c$ *is symmetric.*

(b) *All real solution* $X \geqslant 0$ *of* $MX = XM^T$ *are obtained as in part* (a) *with, in addition,* $W_1 = W_2 = 0$ *and by choosing* W_r *so that* $W_r P_r \geqslant 0$. *Nonzero positive semidefinite solutions exist if and only if* M *has at least one real eigenvalue.*

(c) *Real solutions* $X > 0$ *exist for the equation* $MX = XM^T$ *if and only if* M *is simple with all eigenvalues real. All such solutions have the form* (2.3) *with* $P = P_r = I$, *and are obtained by choosing* $W = W_r$ *so that* $W > 0$.

To illustrate the form of the matrices W appearing in part (a) of the theorem, suppose that M has just one real eigenvalue with elementary divisors of degree two and three. Then $W_r P_r$ is a real matrix of the form

$$W_r P_r = \left[\begin{array}{cc|ccc} a_0 & a_1 & 0 & b_0 & b_1 \\ 0 & a_0 & 0 & 0 & b_0 \\ \hline b_0 & b_1 & c_0 & c_1 & c_2 \\ 0 & b_0 & 0 & c_0 & c_1 \\ 0 & 0 & 0 & 0 & c_0 \end{array}\right] \left[\begin{array}{cc|ccc} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{array}\right]$$

$$= \left[\begin{array}{cc|ccc} a_1 & a_0 & b_1 & b_0 & 0 \\ a_0 & 0 & b_0 & 0 & 0 \\ \hline b_1 & b_0 & c_2 & c_1 & c_0 \\ b_0 & 0 & c_1 & c_0 & 0 \\ 0 & 0 & c_0 & 0 & 0 \end{array}\right].$$

Proof of Theorem 3.1. (a): It is clear that the three conditions, $W_r P_r$ real symmetric, $W_2 = \overline{W}_1$, and $W_1 P_c$ symmetric, when combined with (2.3) and (2.7), determine real symmetric solutions X.

Conversely, if X is real then (2.3) implies

$$X = V(WP)V^T = \overline{W}(\overline{WP})\overline{V}^T,$$

and, since $\overline{V} = VK$ [Equation (2.8)],

$$V(WP)V^T = V(K\overline{W}PK)V^T.$$

As V is nonsingular, it is easily seen that this implies $W_r = \overline{W}_r$ and $W_2 = \overline{W}_1$.

If, in addition, X is symmetric, then also $W_r P_r$ and $W_1 P_c$ must be symmetric. This proves part (a).

Parts (b) and (c) follow from (a), taking into account Equations (2.9) and (2.10). ∎

3. ADMISSIBLE MATRICES

Now let us consider the equation $AX = YB$. Since $(B^T A)X$ must be symmetric, the symmetric solutions X must be among those described by Theorem 2.1, if we take $M = B^T A$. In addition, any such X must be such that the equation $AX = YB$ (for Y) is consistent. Now $AX = YB$ is equivalent to $B^T Y^T = (AX)^T$, and this is consistent if and only if $\operatorname{Im}(AX)^T \subset \operatorname{Im} B^T$, where the symbol $\subset$ is used to denote either strict inclusion or equality. This in turn is equivalent to $\operatorname{Ker} B \subset \operatorname{Ker} AX$, and this is the form of the consistency condition that we will use. Obviously, this is trivially satisfied when B has full rank and $m \leqslant n$, for then $\operatorname{Ker} B = \{0\}$.

Now, it will be convenient to have a formal definition.

Definition 3.1. A real $m \times m$ matrix X will be called *admissible* if

$$(B^T A)X = X(B^T A)^T, \tag{3.1}$$

$X^T = X$, and $\operatorname{Ker} B \subset \operatorname{Ker} AX$.

Thus, admissible matrices are those generated by Theorem 2.1 (with $M = B^T A$) for which solution pairs (X, Y) of $AX = YB$ also exist; it will be seen in Section 4 that when X is admissible there are also (without further conditions) solution pairs (X, Y) with $Y^T = Y$.

Before proving the main theorem on existence of admissible matrices, it is convenient to establish a lemma.

Lemma 3.1. *If* $(B^T A)X = X(B^T A)^T$, $X^T = X \neq 0$, *and* $\operatorname{Im} X \subset \operatorname{Im} B^T A$, *then X is admissible.*

Proof of Theorem 3.1. We have only to show that $\operatorname{Ker} B \subset \operatorname{Ker} AX$. Let S be the orthogonal projection onto $\operatorname{Im} B^T$ along $\operatorname{Ker} B$. Then we have

$$\operatorname{Im} X \subset \operatorname{Im} B^T A \subset \operatorname{Im} B^T,$$

so that $SX = X$, and consequently $SXA^T = XA^T$. Since $X^T = X$ and $S^T = S$, we obtain $AXS = AX$. Now if $u \in \operatorname{Ker} B$ then $Su = 0$, so that also $AXu = 0$, i.e. $\operatorname{Ker} B \subset \operatorname{Ker} AX$. ∎

THEOREM 3.1. *Let real $n \times m$ matrices A and B be given. Then:*

(a) *There exist nonzero admissible matrices X.*

(b) *There exist admissible matrices $X \geqslant 0$ if and only if B^TA has at least one real eigenvalue. If B^TA has a nonzero real eigenvalue or a zero eigenvalue with a nonlinear elementary divisor, then there exist nonzero admissible matrices $X \geqslant 0$.*

(c) *There exist admissible matrices $X > 0$ if and only if B^TA is a simple matrix with all eigenvalues real and*

$$\operatorname{rank} AB^T = \operatorname{rank} A. \tag{3.2}$$

The general form of these matrices X is given in Theorem 2.1, parts (a), (b), and (c), respectively.

REMARK 3.1. The ungainly statement of part (b) is required because if zero is the only real eigenvalue of B^TA and it has only linear elementary divisors, then a *nonzero* admissible $X \geqslant 0$ may or may not exist (see examples below).

Proof of Theorem 3.1. (a): With $M = B^TA$ Equation (2.1) implies that

$$\operatorname{Im} B^TA = \operatorname{Im} VJ = \operatorname{Im} V_rJ_r + \operatorname{Im} V_cJ_c + \operatorname{Im} \bar{V}_c\bar{J}_c.$$

Let us reduce J_r further in the form $J_r = \operatorname{diag}[J_0, J_\rho]$ where J_0 is nilpotent (has all eigenvalues zero) and J_ρ is nonsingular. Partition V_r accordingly: $V_r = [V_0, V_\rho]$, and we have

$$\operatorname{Im} B^TA = \operatorname{Im} V_0J_0 + \operatorname{Im} V_\rho J_\rho + \operatorname{Im} V_cJ_c + \operatorname{Im} \bar{V}_c\bar{J}_c. \tag{3.3}$$

Now X must have the form given in Equation (2.3), where J commutes with W, and hence

$$\operatorname{Im} X = \operatorname{Im} V_0W_0 + \operatorname{Im} V_\rho W_\rho + \operatorname{Im} V_cW_1 + \operatorname{Im} \bar{V}_cW_2, \tag{3.4}$$

where W_r in (2.7) is $W_r = \operatorname{diag}[W_0, W_\rho]$, corresponding to the above form of

J_r. Comparing (3.3) and (3.4), it is clear that W can be chosen in a manner consistent with Theorem 2.1(a) and, at the same time, satisfy $\operatorname{Im} X \subset \operatorname{Im} B^T A$. Because of the nullity of J_0, care must only be taken to ensure that $\operatorname{Im} V_0 W_0 \subset \operatorname{Im} V_0 J_0$. This can generally be achieved by setting $W_0 = 0$ and then using Lemma 3.1.

If $B^T A$ has *only* the zero eigenvalue, assume that it is already reduced to Jordan canonical form J. We show how to find a block-diagonal W satisfying $JW = WJ$ [and hence an X from (2.3)] so that $\operatorname{Ker} B \subset \operatorname{Ker} AX$. If the Jordan block in question has size one, then we may suppose $B^T A = [0]$, the zero matrix of size one. Since $B \neq 0$, B is a nonzero column vector and $\operatorname{Ker} B = \{0\}$. So we can take any scalar $W \neq 0$ to satisfy $JW = WJ$ and generate a symmetric $X \neq 0$. Clearly, $\operatorname{Ker} B \subset \operatorname{Ker} AX$ and hence X is admissible. When there is a Jordan block of size greater than one, we again consider size three as typical. Thus, consider

$$B^T A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad X = \begin{bmatrix} x_1 & x_2 & x_3 \\ x_2 & x_3 & 0 \\ x_3 & 0 & 0 \end{bmatrix}. \tag{3.5}$$

It is easily seen that either $\operatorname{Ker} B = \{0\}$ or $\operatorname{Ker} B$ is the span of the third unit coordinate vector. In the first case $\operatorname{Ker} B \subset \operatorname{Ker} AX$ trivially for any choice of x_1, x_2, x_3 not all zero. In the second case choose $x_3 = 0$ and x_1, x_2 not both zero to obtain a nonzero admissible matrix.

(b): Referring to Theorem 2.1(b), we may use the same argument as in part (a). The second statement of part (b) is required because the case in which $B^T A$ has zero as the only real eigenvalue and its elementary divisors are linear does not ensure the existence of an admissible $X \geqslant 0$ with $X \neq 0$. Consider two examples. First, if

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

then $B^T A = 0$. Thus, zero is the only eigenvalue of $B^T A$, and trivially the elementary divisors are linear. In this case we may take

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

which is nonzero, positive semidefinite, and admissible. In contrast, if $A = I_3$

and

$$B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix},$$

then $B^TA = B^T$ and has zero as the only real eigenvalue, and it is simple. However, a little calculation verifies that there is *no* nonzero, positive semidefinite, admissible X.

(c): First, if $X > 0$ and is admissible, then $AX = YB$ for some Y so that $AXA^T = YBA^T$. But now A and AXA^T have the same rank, so

$$\operatorname{rank} A = \operatorname{rank} AXA^T = \operatorname{rank} YBA^T \leqslant \operatorname{rank} BA^T = \operatorname{rank} AB^T.$$

But $\operatorname{rank} AB^T \leqslant \operatorname{rank} A$, so equality obtains throughout and $\operatorname{rank} AB^T = \operatorname{rank} A$, as required. It follows from Theorem 2.1(c) that B^TA is a simple matrix with all eigenvalues real.

Conversely, given that B^TA is simple with all real eigenvalues, we can construct an $X > 0$ such that $(B^TA)X = X(B^TA)^T$ as indicated in Theorem 2.1(c). We have only to show that given the rank condition (3.2), there is an X of this form for which the consistency condition $\operatorname{Ker} B \subset \operatorname{Ker} AX$ is satisfied.

It has already been remarked that if B has full rank then the consistency condition is satisfied. So let $\operatorname{rank} B = r_B < m$, and first reduce B to canonical form by a real equivalence transformation,

$$\overline{B} = EBF = \begin{array}{c} \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} \begin{array}{l} n - r_B \\ r_B \end{array} \\ \begin{array}{cc} m - r_B & \quad r_B \end{array} \end{array} \tag{3.6}$$

where E and F are suitable nonsingular matrices. If we define

$$\overline{A} = E^{-T}AF^{-T}, \qquad \overline{X} = F^TXF, \qquad \overline{Y} = E^{-T}YE^{-1},$$

then it is easily verified that $\overline{A}\,\overline{X} = \overline{Y}\overline{B}$, $\overline{X} > 0$, $(\overline{B}^T\overline{A})\overline{X} = \overline{X}(\overline{B}^T\overline{A})^T$, and $\overline{B}^T\overline{A}$ is similar to B^TA. Further, the conditions

$$\operatorname{rank} AB^T = \operatorname{rank} A \quad \text{and} \quad \operatorname{rank} \overline{A}\,\overline{B}^T = \operatorname{rank} \overline{A}$$

are equivalent. It follows that, without loss of generality, we may assume that

B is given in the canonical form (3.6). Then with a partition of A consistent with (3.6) we have

$$AB^T = \begin{bmatrix} 0 & A_{12} \\ 0 & A_{22} \end{bmatrix},$$

and $\operatorname{rank} AB^T = \operatorname{rank} A$ implies that the first $m - r_B$ columns of A are linearly dependent on the last r_B columns. In other words, there is a matrix K such that

$$A = \begin{bmatrix} A_{12}K & A_{12} \\ A_{22}K & A_{22} \end{bmatrix}.$$

But also

$$B^TA = \begin{bmatrix} 0 & 0 \\ A_{22}K & A_{22} \end{bmatrix},$$

and it obviously has $m - r_B$ linearly independent eigenvectors corresponding to the zero eigenvalue of the form

$$v = \begin{bmatrix} v_1 \\ -Kv_1 \end{bmatrix}.$$

Thus, we may write

$$VA = \begin{bmatrix} V_{11} & 0 \\ -KV_{11} & V_{22} \end{bmatrix},$$

where V_{11} and V_{22} are nonsingular. As B^TA is simple, J is diagonal and has the form $J = \operatorname{diag}[0, J_2]$, where the zero matrix has size $m - r_B$ (and J_2 may be singular). Note also that $P = P_r = I$.

As in Theorem 2.1(c) [see also Equation (2.8)], we have $X = V(W_r P_r)V^T$. We may now choose W_r to be diagonal (although more general choices are also admissible). Thus, let $W = \operatorname{diag}[W_1, W_2]$, where W_1 has size $m - r_B$ and W_1, W_2 are both positive definite and diagonal. Then certainly $X > 0$.

The subspace $\operatorname{Ker} B$ is clearly spanned by vectors of the form $\begin{bmatrix} u \\ 0 \end{bmatrix}$, where u is an arbitrary $(m - r_B)$-vector. Thus, if

$$x = \begin{bmatrix} u \\ 0 \end{bmatrix} \in \operatorname{Ker} B,$$

then

$$AXx = \begin{bmatrix} A_{12}K & A_{12} \\ A_{22}K & A_{22} \end{bmatrix} \begin{bmatrix} V_{11} & 0 \\ -KV_{11} & V_{22} \end{bmatrix} \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} V_{11}^T & -V_{11}^T K^T \\ 0 & V_{22}^T \end{bmatrix} \begin{bmatrix} u \\ 0 \end{bmatrix}. \tag{3.7}$$

Multiplying out the right-hand side, it is found that $AXx = 0$, i.e. $\operatorname{Ker} B \subset \operatorname{Ker} AX$, as required. ■

4. SOLUTIONS WITH $Y^T = Y$

We first consider the problem of finding solution pairs (X, Y) for $AX = YB$ where both X and Y are real symmetric and no definiteness conditions are imposed. The solution pairs $(X \geqslant 0,\ Y^T = Y)$ and $(X > 0,\ Y^T = Y)$ then follow at once from Theorem 3.1. It will be convenient to use the language of generalized inverses here (see Reference [2], for example). For any (possibly rectangular) matrix M a matrix X is a (1)-inverse of M if $MXM = M$. If, in addition, X satisfies $XMX = X$, then we call X a (2)-inverse of M. It is well known that such inverses always exist but are not unique. The first two theorems of this section include only cosmetic improvements of the results in [9].

THEOREM 4.1. *If X is a nonzero admissible matrix, then there is a nonzero Y such that $Y^T = Y$ and $AX = YB$. Also, there exists a real $m \times n$ matrix U such that*

$$AX(UB) = AX, \tag{4.1}$$

and the general real symmetric solution Y of $AX = YB$ is given by

$$Y = AXU + U^T XA^T - U^T B^T AXU + Y_0, \tag{4.2}$$

where Y_0 is any real symmetric matrix for which $Y_0 B = 0$.

Proof. We first verify that there is a U for which (4.1) is satisfied. Let B^I be any (1)-inverse of B. Then it is easily seen that $I - B^I B$ is a projection onto $\operatorname{Ker} B$. Since X is admissible, we have $\operatorname{Ker} B \subset \operatorname{Ker} AX$, and it follows that $AX(I - B^I B) = 0$. Thus, we can take $U = B^I$ in Equations (4.1) and (4.2).

We now proceed by verification. Multiply (4.2) on the right by B, and using (4.1) and the symmetry of B^TAX, it is found that $AX = YB$ and $Y^T = Y$. The general solution has the form (4.2) because $AX = Y_1B$ and $AX = Y_2B$ imply $(Y_1 - Y_2)B = 0$. ■

At the expense of the further rank condition [shown in Theorem 6.1 to be necessary for the solution pair $(X > 0,\ Y \geqslant 0)$]

$$\operatorname{rank} B^TA = \operatorname{rank} A, \tag{4.3}$$

a more compact form of the general solution (4.2) can be obtained.

THEOREM 4.2. *If X is an admissible matrix and the rank condition* (4.3) *holds, then the general real symmetric solution Y of $AX = YB$ is given by*

$$Y = AX(B^TAX)^{\#}XA^T + Y_0, \tag{4.4}$$

where $(B^TAX)^{\#}$ denotes any (2)*-inverse of B^TAX, and Y_0 is any real symmetric matrix for which $Y_0B = 0$.*

Proof. Let $U = (B^TAX)^{\#}XA^T$. As $(B^TAX)^{\#}$ is a (1)-inverse and B^TAX is symmetric,

$$B^TAX(B^TAX)^{\#}XA^TB = B^TAX,$$

or

$$B^TA(XUB - X) = 0.$$

Then $\operatorname{rank} B^TA = \operatorname{rank} A$ implies that also

$$A(XUB - X) = 0.$$

Thus, U satisfies Equation (4.1)

Now put $U = (B^TAX)^{\#}XA^T$ in (4.2), and use the fact that $(B^TAX)^{\#}$ is a (2)-inverse and the symmetry of B^TAX to obtain (4.4). ■

Combining the existence statement of Theorem 4.1 with Theorem 3.1, we obtain the first major result on the existence of solution pairs of $AX = YB$. Here, and in the sequel, the phrase "a nonzero pair (X, Y)" means that both $X \neq 0$ and $Y \neq 0$.

THEOREM 4.3. *Let real $n\times m$ matrices A and B be given. Then:*

(a) *There exist nonzero solution pairs $(X^T=X,\ Y^T=Y)$ of $AX=YB$.*

(b) *There exist solution pairs $(X\geqslant 0,\ Y^T=Y)$ of $AX=YB$ if and only if B^TA has at least one real eigenvalue. If B^TA has a nonzero eigenvalue or zero eigenvalue with a nonlinear elementary divisor, then there exist solution pairs $(X\geqslant 0,\ Y^T=Y)$ with $X\neq 0$ (see Remark 3.1).*

(c) *There exist solution pairs $(X>0,\ Y^T=Y)$ of $AX=YB$ if and only if B^TA is a simple matrix with all eigenvalues real and* $\operatorname{rank} AB^T=\operatorname{rank} A$.

These solutions X and Y all have the forms given in Theorems 2.1 and 4.1, respectively.

5. PRELIMINARY RESULTS WITH $Y\geqslant 0$ AND $Y>0$

In this section a technical lemma is established which will play an important part in the examination of solution pairs (X,Y) with $Y\geqslant 0$ and $Y>0$. We first need some geometric ideas. Let $b_1,\dots,b_r$ be an orthogonal basis for $\operatorname{Im} B$, and write

$$\mathscr{B}_1=[b_1,b_2,\dots,b_r],$$

an $n\times r$ matrix, where r is the rank of B. Then define $Q=\mathscr{B}_1\mathscr{B}_1^T$, the orthogonal projection onto $\operatorname{Im} B$.

Since $(\operatorname{Im} B)^\perp=\operatorname{Ker} B^T$ [where $(\ )^\perp$ denotes the orthogonal complement], there is an orthonormal basis $b_{r+1},\dots,b_n$ for $\operatorname{Ker} B^T$ such that $b_1,b_2,\dots,b_n$ is an orthonormal basis for $\mathbb{R}^n$. Then, if we define

$$\mathscr{B}_2=[b_{r+1},\dots,b_n], \tag{5.1}$$

we have $\mathscr{B}_2\mathscr{B}_2^T=I-Q$, the orthogonal projection onto $\operatorname{Ker} B^T$.

For an $n\times n$ symmetric matrix Y we define the *compression* $Y_{11}=\mathscr{B}_1^TY\mathscr{B}_1$. Then construct an orthonormal basis of eigenvectors for Y_{11}, say $u_1,\dots,u_r$ such that

$$Y_{11}u_j=0 \qquad \text{for} \quad j=1,2,\dots,d$$

and

$$Y_{11}u_k=\lambda_k u_k\neq 0 \qquad \text{for} \quad k=d+1,\dots,r.$$

Thus d is the dimension of $\operatorname{Ker} Y_{11}$. Now let

$$U_1 = [u_1, \dots, u_d], \qquad U_2 = [u_{d+1}, \dots, u_r], \tag{5.2}$$

and define the projection

$$R = \mathscr{B}_1 U_1 U_1^T \mathscr{B}_1^T. \tag{5.3}$$

It is easily seen now that

$$[\mathscr{B}_1 U_1 \quad \mathscr{B}_1 U_2 \quad \mathscr{B}_2] \tag{5.4}$$

is an orthogonal matrix, and

$$\begin{bmatrix} U_1^T \mathscr{B}_1^T \\ U_2^T \mathscr{B}_1^T \\ \mathscr{B}_2^T \end{bmatrix} Y [\mathscr{B}_1 U_1 \quad \mathscr{B}_1 U_2 \quad \mathscr{B}_2] = \begin{bmatrix} 0 & 0 & Y_{13} \\ 0 & Y_{22} & Y_{23} \\ Y_{13}^T & Y_{23}^T & Y_{33} \end{bmatrix}, \tag{5.5}$$

where Y_{22} is nonsingular and, in general, Y_{13}, Y_{23}, Y_{33} are nonzero. The block matrix on the right is just the representation of Y_1 with respect to the orthonormal basis for $\mathbb{R}^n$ defined by the columns of (5.4) (see Chapter 4 of [12], for example).

The relevance of the following lemma for investigation of the solutions of $AX = YB$ with $Y \geqslant 0$ can be seen from the fact that $B^T AX = B^T YB \geqslant 0$ follows from $Y \geqslant 0$.

LEMMA 5.1. *Let X be a nonzero admissible matrix for which $B^T AX \geqslant 0$. Let $B^\dagger$ be the Moore-Penrose inverse of B (see* [2, 12], *for example), and use Theorem* 4.1 *to define a corresponding solution of $AX = YB$ by*

$$Y_1 = AXB^\dagger + (B^\dagger)^T XA^T - (B^\dagger)^T (B^T AX) B^\dagger. \tag{5.6}$$

Let R be the orthogonal projection defined by Y_1 as in the above construction (see equation (5.3)), *and let $Y_0 = \alpha(I - BB^\dagger)$ and $Y = Y_1 + Y_0$. Then:*

(a) *There exist numbers $\alpha > 0$ such that (X, Y) is a nonzero solution pair with $Y \geqslant 0$ if and only if $RY_1 = 0$.*

(b) *There exist numbers $\alpha > 0$ such that (X, Y) is a nonzero solution pair with $Y > 0$ if and only if $R = 0$.*

Note that the orthogonal projection onto Im B may now be written in the forms $Q = \mathscr{B}_1\mathscr{B}_1^T = BB^\dagger = (B^\dagger)^T B^T$.

Proof of Lemma 5.1. Since $AX = Y_1 B$ and $B^T AX \geqslant 0$, we have $B^T Y_1 B \geqslant 0$. In other words, Y_1 is positive semidefinite on the image of B. In the representation (5.5) of Y_1 this implies that $Y_{22} > 0$ (when $B \neq 0$).

It follows from (5.6) that $Y_1 Q = AXB^\dagger$, and substituting this back into (5.6) gives

$$Y_1 = Y_1 Q + QY_1 - QY_1 Q,$$

or

$$(I-Q)Y_1(I-Q) = 0.$$

In other words, Y_1 is zero on $\mathrm{Im}(I-Q) = \mathrm{Ker}\, B^T$. Since $I - Q = \mathscr{B}_2\mathscr{B}_2^T$ [see (5.1)], this implies that in the representation (5.5) of Y_1 we have $Y_{33} = 0$. Thus, there is an orthonormal basis for $\mathbb{R}^n$ in which the representation of Y_1 has the form

$$\begin{bmatrix} 0 & 0 & Y_{13} \\ 0 & Y_{22} & Y_{23} \\ Y_{13}^T & Y_{23}^T & 0 \end{bmatrix}, \tag{5.7}$$

where $Y_{22} > 0$.

Now further solutions of $AX = YB$ can be generated by adding to Y_1 symmetric matrices Y_0 with the property that $Y_0 B = 0$. Clearly, for any real α, matrices

$$Y_0 = \alpha(I-Q) = \alpha\left(I - \mathscr{B}_1\mathscr{B}_1^T\right)$$

have these properties. Since $\mathscr{B}_1^T\mathscr{B}_2 = 0$ and $\mathscr{B}_2^T\mathscr{B}_2 = I$

$$\mathscr{B}_2^T Y_0 \mathscr{B}_2 = \alpha I.$$

Thus, in the representation (5.5) for Y_0 we have $Y_{33} = \alpha I$ and all other blocks are zero. Combining this with the representation (5.7) for Y_1, it is found that a matrix $Y = Y_1 + Y_0$ satisfies $AX = YB$ and has the representation

$$\begin{bmatrix} 0 & 0 & Y_{13} \\ 0 & Y_{22} & Y_{23} \\ Y_{13}^T & Y_{23}^T & \alpha I \end{bmatrix}. \tag{5.8}$$

Now we set about the proof of statements (a) and (b). (a): Suppose first that $RY_1 = 0$. From (5.5) we have

$$Y_{13} = U_1^T \mathscr{B}_1^T Y_1 \mathscr{B}_2.$$

But, using (5.3), $RY_1 = 0$ implies $U_1^T \mathscr{B}_1^T Y_1 = 0$ and hence $Y_{13} = 0$. Now the four bottom right blocks of (5.8) are congruent to

$$\begin{bmatrix} Y_{22} & 0 \\ 0 & \alpha I - Y_{23}^T Y_{22}^{-1} Y_{23} \end{bmatrix},$$

and so, by choosing α large enough, we can make

$$\begin{bmatrix} Y_{22} & Y_{23} \\ Y_{23}^T & \alpha I \end{bmatrix}$$

positive definite and hence $Y \geqslant 0$.

Conversely, if $Y \geqslant 0$, it obviously follows from (5.8) that we must have $Y_{13} = 0$, and from (5.5) we obtain

$$U_1^T \mathscr{B}_1^T Y_1 [\mathscr{B}_1 U_1 \quad \mathscr{B}_1 U_2 \quad \mathscr{B}_2] = 0.$$

Since the matrix on the right is unitary, this implies $U_1^T \mathscr{B}_1^T Y_1 = 0$ and hence $RY_1 = 0$.

(b): If $R = 0$, then the kernel of $\mathscr{B}_1^T Y_1 \mathscr{B}_1$ is trivial and the first row and column of blocks in (5.5) and (5.8) simply do not appear. Then, as above, α can be chosen so that $Y > 0$. For the converse, we can only have $Y > 0$ in (5.8) if $R = 0$. ∎

EXAMPLE. Let

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then

$$B^TA = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad B^{\dagger} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

It is easily verified that all admissible matrices have the form

$$X = \begin{bmatrix} 0 & 0 & x_{13} \\ 0 & \xi & \xi \\ x_{13} & \xi & x_{33} \end{bmatrix}.$$

for an arbitrary real parameters x_{13}, ξ, x_{33}. From (5.6) is found that

$$Y_1 = \begin{bmatrix} 0 & 0 & -x_{13} \\ 0 & \xi & 0 \\ -x_{13} & 0 & 0 \end{bmatrix}$$

and $Y_0 = \operatorname{diag}[\alpha, 0, 0]$. Thus, X determines the class of "partners"

$$Y = \begin{bmatrix} \alpha & 0 & -x_{13} \\ 0 & \xi & 0 \\ -x_{13} & 0 & 0 \end{bmatrix}. \tag{5.9}$$

The representation (5.8) for Y is found to be

$$\begin{bmatrix} 0 & 0 & -x_{13} \\ 0 & \xi & 0 \\ -x_{13} & 0 & \alpha \end{bmatrix},$$

and also

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

For case (a) the hypothesis $RY_1 = 0$ is equivalent to $x_{13} = 0$, and in this case we can choose α so that $Y \geqslant 0$. Since $R \neq 0$, it follows from case (b), and is obvious from (5.9), that there are no solutions $Y = Y_1 + Y_0$ with $Y > 0$.

6. SOLUTIONS WITH $Y \geqslant 0$

In this section we establish results concerning the three cases $X^T = X$, $X \geqslant 0$, $X > 0$ together with $Y \geqslant 0$. The first of these is obtained immediately from Theorem 4.3(b) by transposition: since there is no hypothesis on the relative size of m and n, the roles of X and Y are reversed if, in the equation $AX = YB$, A is replaced by B^T and B is replaced by A^T. The other two cases are covered in the next theorem. We note again a certain complication with the case of nonnegative solution pairs which (as in Theorems 3.1 and 4.3) arises when B^TA is nilpotent (i.e. has only zero eigenvalues).

THEOREM 6.1

(a) *If B^TA has a positive eigenvalue, then $AX = YB$ has a nonzero solution pair $(X \geqslant 0, Y \geqslant 0)$. Conversely, if there exists a nonzero solution pair $(X \geqslant 0, Y \geqslant 0)$, then AB has a nonnegative eigenvalue.*

(b) *The equation $AX = YB$ has solution pairs $(X > 0, Y \geqslant 0)$ if and only if B^TA is simple with all eigenvalues nonnegative, and*

$$\operatorname{rank} AB^T = \operatorname{rank} A = \operatorname{rank} B^TA. \tag{6.1}$$

Proof. (a): Let $\lambda > 0$ be an eigenvalue of B^TA. If λ has a linear elementary divisor, choose the corresponding entry w in W_ρ [of Equation (3.4) to be positive. If λ has an elementary divisor of degree two, there are blocks in WP and JWP, respectively, of the form

$$\begin{bmatrix} w & w_1 \\ w_1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \lambda w + w_1 & \lambda w_1 \\ \lambda w_1 & 0 \end{bmatrix}. \tag{6.2}$$

In this case, choose $w_1 = 0$ and $w > 0$. Thus, when X has an elementary divisor of degree one or two, there is a semidefinite block of JWP. Clearly, the construction can be extended to elementary divisors of any degree, and on setting all other entries of W equal to zero we obtain

$$X = wvv^T, \qquad B^TAX = \lambda wvv^T, \tag{6.3}$$

where $B^TAv = \lambda v$, $v \neq 0$, and $\lambda > 0$. Clearly, $X \neq 0$ and $X \geqslant 0$. Also $\operatorname{Im} X \subset \operatorname{Im} B^TAX$ because $\lambda \neq 0$, so the admissibility of X follows from Lemma 3.1. Note also that the admissible X in (6.3) is such that $B^TAX \geqslant 0$, are required by Lemma 5.1.

We are now to apply Lemma 5.1(a) to prove the existence of $Y \geqslant 0$ such that (X, Y) is a nonzero solution pair. With X given by (6.3), Equation (5.6) gives

$$Y_1 = w\left[Avv^TB^\dagger + (B^\dagger)^T vv^TA^T - (B^\dagger)^T(B^TAX)B^\dagger\right].$$

Writing $v^T = \lambda^{-1}v^TA^TB$ and $BB^\dagger = Q$, this takes the form

$$Y_1 = \frac{w}{\lambda}(Avv^TA^TQ + QAvv^TA^T - QAvv^TA^TQ). \tag{6.4}$$

To construct the projector R determined by X (and hence Y_1), first form

$$Y_{11} = \mathscr{B}_1^T Y_1 \mathscr{B}_1 = \frac{w}{\lambda}\mathscr{B}_1^TAvv^TA^T\mathscr{B}_1,$$

since $Q\mathscr{B}_1 = \mathscr{B}_1$. Let $u = \mathscr{B}_1^TAv$, and we have

$$Y_{11} = \frac{w}{\lambda}uu^T.$$

The eigenvalues of Y_{11} are $(w/\lambda)\|u\|^2$ and zeros. We now find an orthogonal U such that

$$UY_{11}U^T = \operatorname{diag}\left[0,\dots,0,\frac{w}{\lambda}\|u\|^2\right].$$

In fact, if $a = (1/\|u\|)u$ then

$$U = [u_1,\dots,u_{r-1},a],$$

where $u_1,\dots,u_{r-1}$, a form an orthonormal system. We write $U_1 = [u_1,u_2,\dots,u_{r-1}]$, $U = [U_1, a]$. Then (5.3) gives

$$\begin{aligned} R &= \mathscr{B}_1U_1U_1^T\mathscr{B}_1^T = \mathscr{B}_1(UU^T - aa^T)\mathscr{B}_1^T \\ &= \mathscr{B}_1(I - aa^T)\mathscr{B}_1^T, \\ &= Q\left(I - \frac{1}{\|u\|^2}Avv^TA^T\right)Q, \end{aligned}$$

after some simplification [using $u = \mathscr{B}_1^TAv$ and $a = (1/\|u\|)u$].

For brevity, write $b = Av$; then

$$R = Q\left(I - \frac{1}{\|u\|^2}bb^T\right)Q,$$

and (6.4) gives

$$QY_1 = \frac{w}{\lambda}Qbb^T.$$

Hence

$$\begin{aligned}RY_1 &= \frac{w}{\lambda}Q\left(I - \frac{1}{\|u\|^2}bb^T\right)Qbb^T,\\ &= \frac{w}{\lambda}\left(1 - \frac{b^TQb}{\|u\|^2}\right)Qbb^T.\end{aligned}$$

But $b^TQb = v^TA^T\mathscr{B}_1\mathscr{B}_1^TAv = u^Tu = \|u\|^2$ and so $RY_1 = 0$. It follows from Lemma 5.1(a) that there are nonzero matrices $Y = Y_1 + Y_0 \geqslant 0$ for which $AX = YB$.

Conversely, if $AX = YB$, $X \geqslant 0$, and $Y \geqslant 0$, then

$$0 \leqslant B^TYB = B^TAX = V(JWPK)V^*.$$

This implies $JWPK \geqslant 0$ and hence $W_1 = W_2 = 0$ [in Equation (2.10)] and $J_rW_rP_r \geqslant 0$. Since we also have $W_rP_r \geqslant 0$, and $W_r \neq 0$ (since $X \neq 0$), it is seen from the earlier discussion that this demands the presence of a nonnegative eigenvalue for B^TA. For example, in Equation (6.2) we must have $w_1 = 0$, $w > 0$, and $\lambda \geqslant 0$.

(b): To prove part (b) assume first that B^TA is simple with all eigenvalues nonnegative and that (6.1) holds. Since $\operatorname{rank} AB^T = \operatorname{rank} A$ we obtain the existence of an admissible $X > 0$ from Theorem 3.1(c). With such an X, and as $\operatorname{rank} B^TA = \operatorname{rank} A$, we may employ a Moore-Penrose inverse in Theorem 4.2, and define $Y_0 = \alpha(I - BB^\dagger) \geqslant 0$ to obtain a solution pair (X, Y) with $X > 0$, $Y \geqslant 0$.

For the converse, let $X = LL^T$, where L is nonsingular. Then $L^{-1}(B^TYB)L^{-T} = L^{-1}(B^TAX)L^{-T} = L^{-1}(B^TA)L \geqslant 0$. This implies that B^TA is simple with nonnegative values. Furthermore, $AX = YB$ and $X > 0$ imply

$$\operatorname{rank} A = \operatorname{rank} AX = \operatorname{rank} YB,$$

and also

$$\operatorname{rank} B^T A = \operatorname{rank} B^T AX = \operatorname{rank} B^T YB.$$

But $Y \geqslant 0$ implies that $\operatorname{rank} B^T YB = \operatorname{rank} YB$, and consequently $\operatorname{rank} B^T A = \operatorname{rank} A$. The necessity of $\operatorname{rank} AB^T = \operatorname{rank} A$ has been shown in Theorem 3.1(c). ∎

Notice that since, according to Theorem 6.1(b), $\operatorname{rank} B^T A = \operatorname{rank} A$ is necessary for the solution pair $(X > 0,\ Y \geqslant 0)$, one may use Theorem 4.2 to obtain a general form for $Y \geqslant 0$ when $X > 0$. We replace in Equation (4.4) the 2-inverse $(\ )^{\#}$ by the Moore-Penrose inverse $(\ )^{\dagger}$ and require, in addition to $Y_0 B = 0$, that $Y_0 \geqslant 0$. As in the proof of Theorem 6.1(b), $X > 0$ implies $B^T AX \geqslant 0$, whence $(B^T AX)^{\dagger} \geqslant 0$. This gives $Y \geqslant 0$ in Equation (4.4).

7. SOLUTIONS WITH $Y > 0$

Of the three cases $X^T = X$, $X \geqslant 0$, and $X > 0$, together with $Y > 0$, the first two cases are already covered by transposition of the equation (1.1) and interchange of X and Y. Thus, solution pairs $(X^T = X,\ Y > 0)$ are characterized using Theorem 4.3(c), and solution pairs $(X \geqslant 0,\ Y > 0)$ are characterized similarly using Theorem 6.1(b). The remaining case is covered by:

THEOREM 7.1. *The equation $AX = YB$ has a solution pair $(X > 0,\ Y > 0)$ if and only if $B^T A$ is simple with all eigenvalues nonnegative, and*

$$\operatorname{rank} AB^T = \operatorname{rank} A = \operatorname{rank} B = \operatorname{rank} B^T A. \tag{7.1}$$

Proof. Let $B^T A = VJV^{-1}$, where $J = \operatorname{diag}[0, J_1]$ and $J_1 > 0$ with size $\rho = \operatorname{rank} B^T A$. Let W be an $m \times m$ diagonal matrix, with $W > 0$, and define $X = VWV^T$. Then $X > 0$ and $B^T AX = V(JW)V^T = V_1(J_1 W_1)V_1^T$, where $V = [V_0\ V_1]$ and V_1 is $m \times \rho$. Furthermore, since $\operatorname{rank}(AB^T) = \operatorname{rank} A$, it follows, as in the proof of Theorem 3.1(c), that X is admissible.

Now we have a solution pair (X, Y) with Y_1 given by (5.6). Let $r = \operatorname{rank} B$ and $\mathscr{B}_1 = BS$, where S is $m \times r$ and has full rank. Then

$$Y_1 \mathscr{B}_1 = Y_1 BS = AXS$$

and

$$\begin{aligned} Y_{11} &= \mathscr{B}_1^T Y_1 \mathscr{B}_1 = S^T(B^TAX)S \\ &= (V_1^TS)^T(J_1W_1)(V_1^TS), \end{aligned}$$

where $J_1W_1 > 0$. Now Y_{11} is $r \times r$ and the matrix on the right has rank ρ, and hence $\rho \leqslant r$. But by hypothesis $\rho = r$, and hence Y_{11} is nonsingular. It follows that in (5.2) $V_1 = 0$ and hence, by (5.3), $R = 0$. Lemma 5.1(b) now applies to show that there is a solution pair (X, Y) with $Y > 0$.

Conversely, if $AX = YB$ with $X > 0$ and $Y > 0$, then rank A = rank B and the remaining conditions follow from Theorem 6.1(b). ∎

Finally, let us consider how Theorem 4.2 may be used in the determination of pairs $(X > 0, Y > 0)$. We first claim a simple lemma whose proof is left to the reader.

LEMMA 7.1. *Let B be a real $n \times m$ matrix. Then Y_0 is a real symmetric matrix for which $Y_0B = 0$ if and only if*

$$Y_0 = \mathscr{B}_2 D \mathscr{B}_2^T, \tag{7.2}$$

where $\mathscr{B}_2$ is defined as in Equation (5.1) *and D is some real symmetric matrix of size n-rank B.*

Furthermore, all matrices $Y_0 \geqslant 0$ for which $Y_0B = 0$ are determined by choosing matrices $D \geqslant 0$ in (7.2).

THEOREM 7.2. *Let the rank conditions* (7.1) *hold. Then for each positive definite admissible matrix X, the matrix*

$$Y = AX(B^TAX)^{\dagger}XA^T + Y_0,$$

where $Y_0B = 0$, determines a solution pair $(X > 0, Y > 0)$ if and only if Y_0 has the form (7.2) *with $D > 0$.*

Proof. Given an admissible $X > 0$, it follows from Theorem 4.2 and Lemma 7.1 that the set of all solution pairs $(X, Y^T = Y)$ is described by

$$Y = AX(B^TAX)^{\dagger}XA^T + Y_0, \tag{7.3}$$

where Y_0 is given by (7.2). If we assume $D > 0$, then it follows that $Y_0 \geqslant 0$. From Theorem 7.1 we know there is at least one solution Y with $Y > 0$, and hence $B^TAX = B^TYB \geqslant 0$, and hence $(B^TAX)^\dagger \geqslant 0$. Thus, for every Y given by (7.3) with $D > 0$ we have $Y \geqslant 0$.

We show that, in fact, $Y > 0$. For any nonzero $x \in \mathbb{R}^n$ let $x = x_1 + x_2$, where $x_1 = Qx \in \operatorname{Im} B$ and $x_2 = (I - Q)x \in (\operatorname{Im} B)^\perp$. Note that $Y_0 x_1 = 0$. Thus, with Y given by (7.3),

$$x^TYx = x^TAX(B^TAX)^\dagger XA^Tx + x_2^TY_0x_2,$$

and each term on the right must be nonnegative. Furthermore, if $x_2 \neq 0$ then, as $x_2 = \mathscr{B}_2 y$ for some $y \neq 0$,

$$x_2^TY_0x_2 = y^T\mathscr{B}_2^T(\mathscr{B}_2 D\mathscr{B}_2^T)\mathscr{B}_2 y = y^TDy > 0.$$

Thus, $x_2 \neq 0$ implies $x^TYx > 0$.

On the other hand, if $x_2 = 0$, then $x = x_1 = Bz$ for some $z \neq 0$ and, from (7.3),

$$x^TYx = z^T(B^TAX)(B^TAX)^\dagger(XA^TB)z = z^T(B^TAX)z. \tag{7.4}$$

Now $z^TB^T = x^T \neq 0$ and $\operatorname{rank} B^T = \operatorname{rank} B^TA$ implies that $z^TB^TA \neq 0$. Let $B^TAX = Z^TZ$, where Z is real and has full rank. Then $z^TB^TA = z^TZ^TZX^{-1} \neq 0$ and hence $Zz \neq 0$. From (7.4) we obtain

$$x^TYx = z^TB^TAXz = z^TZ^TZz > 0.$$

If follows that $Y > 0$, as required.

Conversely, let $(X > 0,\ Y > 0)$ be a solution pair and Y have the form (7.3). For any vector y of size n define

$$x = \left[I - B(B^TAX)^\dagger XA^T\right]y. \tag{7.5}$$

Then calculate x^TYx using the formula (7.3) and the fact that $Y_0B = 0$. It is easily found that

$$x^TYx = y^TY_0y.$$

Since $Y > 0$, it follows that $y^T Y_0 y \geqslant 0$ for all y and hence that $Y_0 \geqslant 0$. It now follows from the lemma that $D \geqslant 0$.

If $Dz = 0$ for some $z \neq 0$, then $z = \mathscr{B}_2^T y_0$ for some $y_0 \neq 0$, and $y_0^T Y_0 y_0 = y_0^T \mathscr{B}_2 D \mathscr{B}_2^T y_0 = 0$. Then define x_0 as in (7.5) with y replaced by y_0. As $x_0 = 0$ implies $B_2^T y_0 = 0$, we deduce that $x_0 \neq 0$. Also, as in the preceding paragraph,

$$x_0^T Y x_0 = y_0^T Y_0 y_0 = 0,$$

which contradicts $Y > 0$. Hence $D > 0$. ∎

8. CONCLUSIONS

The first step of this work has been the formulation of existence theorems for solutions X of $MX = XM^T$, where M is real and X is required to be symmetric, positive semidefinite, or positive definite. This is contained in Theorem 2.1, which also describes the three solution sets.

Given arbitrary nonzero $n \times m$ real matrices A and B, conditions for the existence of symmetric matrices X and Y for which $AX = YB$ have also been examined. Nine types of solution pairs occur with X and Y either symmetric, positive semidefinite, or positive definite. For simplicity Table 1 summarizes only sufficient conditions for the existence of nonzero solution pairs, i.e. with both $X \neq 0$ and $Y \neq 0$. More details are found in Theorems 4.3, 6.1, and 7.1, in particular. In the cases ($X \geqslant 0$, $Y \geqslant 0$) and ($X > 0$, $Y > 0$) the product B^TA can be replaced wherever it appears by AB^T.

General solution forms for the solution pairs in Table 1 are also generated by our results. The existence of admissible matrices X is characterized in Theorem 3.1. They are the matrices X constructed as described in Theorem 2.1 (with $M = B^TA$) for which $\operatorname{Ker} B \subset \operatorname{Ker} AX$. Then for each admissible X a particular solution pair (X, Y) is generated by Theorem 4.1 [or Theorem 4.2 when the condition (4.3) applies]. Since $AX = YB$ is a linear equation in Y, all solutions for a given X are obtained simply by adding to the particular Y symmetric matrices $\hat{Y}$ for which $\hat{Y}B = 0$ as described in Lemma 7.1. However, it is not so obvious how to guarantee that $Y + \hat{Y} \geqslant 0$ when $Y \geqslant 0$, for example. Taking $\hat{Y} \geqslant 0$ is sufficient, but not necessary.

We leave for further study the problems of finding "minimal" solutions. These may be interpreted in different ways. Minimal in the sense of matrix norms is one possibility, which has already been studied for the equation $AX = B$ (see [15], for example). Here, the problem of finding solution pairs

TABLE 1

SUFFICIENT CONDITIONS FOR EXISTENCE OF NONZERO, REAL SOLUTION PAIRS (X,Y) OF $AX = BY$[a]

	$X^T = X$	$X \geqslant 0$	$X > 0$
$Y^T = Y$	Theorem 4.3(a): Nil.	Theorem 4.3(b): B^TA has a nonzero real eigenvalue.	Theorem 4.3(c): B^TA simple. $\sigma(B^TA) \subset \mathbb{R}$. $r(AB^T) = r(A)$.
$Y \geqslant 0$	Theorem 4.3(b): A^TB has a nonzero real eigenvalue.	Theorem 6.1(a): B^TA has a positive eigenvalue.	Theorem 6.1(b): B^TA simple. $\sigma(B^TA) \subset \mathbb{R}^+$. $r(AB^T) = r(A) = r(B^TA)$.
$Y > 0$	Theorem 4.3(c): AB^T simple. $\sigma(AB^T) \subset \mathbb{R}$. $r(B^TA) = r(B)$.	Theorem 6.1(b): AB^T simple. $\sigma(AB^T) \subset \mathbb{R}^+$. $r(B^TA) = r(B) = r(AB^T)$.	Theorem 7.1: B^TA simple. $\sigma(B^TA) \subset \mathbb{R}^+$. $r(AB^T) = r(A) = r(B) = r(B^TA)$.

[a]Here, $r(M)$ is the rank of M, $\sigma(M)$ is the spectrum of M, $\mathbb{R}$ is the real numbers, and $\mathbb{R}^+$ is the nonnegative real numbers. The applicable theorems are indicated. Items below the table's diagonal have been obtained by symmetry.

(X, Y_0) with the property that $Y_0 \leqslant Y$ for all other pairs (X,Y) also suggests itself.

The authors are grateful to K.-w. E. Chu for helpful discussions and comments.

REFERENCES

1 J. K. Baksalary and R. Kala, The matrix equation $AX - YB = C$, *Linear Algebra Appl.* 25:41–43 (1979).
2 A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, Wiley, New York, 1974.
3 K. E. Chu, Singular value and generalized singular value decompositions and the solution of linear matrix equations, *Linear Algebra Appl.* 88/89:83–98 (1987).
4 K. E. Chu, Symmetric solutions of linear matrix equations by matrix decompositions, *Linear Algebra Appl.* 119:35–50 (1989).
5 F. J. H. Don, On the symmetric solutions of a linear matrix equation, *Linear Algebra Appl.* 93:1–7 (1987).
6 H. Flanders and H. K. Wimmer, On the matrix equation $AX - XB = C$ and $AX - YB = C$, *SIAM J. Appl. Math.* 32:707–710 1977.
7 F. R. Gantmacher, *The Theory of Matrices*, vol. 1, Chelsea, New York, 1959.

8 A. Jameson, Solution of the equation $AX + XB = C$ by inversion of an $M \times M$ or $N \times N$ Matrix, *SIAM J. Appl. Math.* 16:1020–1023 (1968).

9 A. Jameson and E. Kreindler, Inverse problem of linear optimal control, *SIAM J. Control* 11:1–19 (1973).

10 V. Kučera, The matrix equation $AX + XB = C$, *SIAM J. Appl. Math.* 26:15–25 (1974).

11 P. Lancaster, Explicit solutions of linear matrix equations, *SIAM Rev.* 12:544–566 (1970).

12 P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, 2nd ed., Academic, Orlando, Fla., 1985.

13 C. C. MacDuffee, *The Theory of Matrices*, Chelsea, New York, 1946.

14 J. R. Magnus, L-structured matrices and linear matrix equations, *Linear and Multilinear Algebra* 14:67–88 (1983).

15 W. E. Roth, The equations $AX - YB = C$ and $AX - XB = C$ in matrices, *Proc. Amer. Math. Soc.* 3:392–396 (1952).

16 O. Taussky and H. Zassenhaus, On the similarity transformation between a matrix and its transpose, *Pacific J. Math.* 9:893–896 (1959).

Received 4 February 1988; final manuscript accepted 30 October 1990