On the existence of a positive definite solution of the matrix equation X = ATX-1A = I

Engwerda, J.C.

Published in:
Linear Algebra and its Applications

Publication date:
1993

On the Existence of a Positive Definite Solution of the Matrix Equation $X + A^{T}X^{-1}A = I$

Jacob C. Engwerda
Department of Econometrics
Tilburg University
P.O. Box 90153
5000 LE Tilburg, the Netherlands

Submitted by Richard A. Brualdi

ABSTRACT

The question is raised under which conditions on the real (square) matrix A the matrix equation $X + A^{T}X^{-1}A = I$ has a real symmetric positive definite solution X. Both necessary and sufficient solvability conditions on A are derived. Moreover, we give an algorithm to calculate the solution. For a number of special cases we also present an analytic solution.

I. INTRODUCTION

The central issue in this paper is to find solvability conditions for the existence of a positive definite solution X of the matrix equation $X + A^{T}X^{-1}A = I$. This problem can be viewed as a natural extension of giving solvability conditions for the scalar problem $x + a^{2}/x = 1$. From calculus we know that the existence of the real square root $\sqrt{1 - 4a^{2}}$ plays here an important role. We will see that this condition generalizes straightforwardly to the matrix case, if A has the additional property that it is normal (i.e. $A^{T}A = AA^{T}$). However, if A has not this additional property, things become more complicated.

We show that the general problem has a solution if and only if a related recursive algorithm converges to a positive definite solution. Moreover we use this algorithm to prove that, provided the matrix A satisfies a certain condition, the matrix equation is solvable, and to calculate a solution numerically.

655 Avenue of the Americas, New York, NY 10010 0024-3795/93/$6.00

Separately, we derive a number of necessary conditions and show by means of a counterexample that these are in general not sufficient.

The paper is organized as follows. First, in Section 2 we introduce some notation and study the general problem together with the recursive algorithm. Then, we derive a number of necessary conditions. Section 4 contains a number of special cases in which a solution exists. Before we discuss the main results in Section 6, we give in Section 5 an example of this equation in the field of optimal control theory.

II. THE GENERAL PROBLEM

Mathematically, the problem analyzed in this paper is to find conditions under which

$$\exists X > 0: \qquad X + A^{\mathsf{T}} X^{-1} A = I, \tag{1}$$

where X and I are real square $\bar{n} \times \bar{n}$ matrices. Here $X > 0$ means that X is symmetric positive definite, A^{T} denotes the transpose of A, and I is the identity matrix. In the sequel also the notation $\geqslant$ is used to indicate that a matrix is symmetric positive semidefinite, and $A > B$ is used as a different notation for $A - B > 0$. Moreover, Ker A denotes the kernel of A, and Im A its image.

Further on we show that this problem has a solution if and only if the following recursion problem is solvable: $\forall n \in \mathbb{N}$, is

$$X_n > AA^{\mathsf{T}} \tag{2a}$$

if

$$X_0 = I, \qquad X_{n+1} = I - A^{\mathsf{T}} X_n^{-1} A. \tag{2b}$$

To prove this result we start with some intermediate results which are interesting in themselves. The first thing we prove is that in fact it suffices to solve the problem (1) for invertible matrices. We show that in case A is not invertible, (1) can be reduced to a similar problem with an invertible matrix A. How this can be accomplished is the content of Theorem 1. Its proof contains an algorithm which will be used again later on.

GENERAL OBSERVATION 1. *If we can solve the problem* (1) *whenever the matrix* A *is invertible, then we can solve it without this invertibility restriction too.*

Proof. We prove this theorem by reducing the problem to a similar problem of lower dimension. The reduction is achieved via the following algorithm:

(i) If A is invertible, then the algorithm is finished.
(ii) Else, there exists an orthogonal transformation T such that

$$A = T^{\mathsf{T}}\begin{pmatrix} A_{11} & 0 \\ A_{21} & 0 \end{pmatrix}T.$$

Consequently, (1) has a solution if and only if (iff) the following problem is solvable:

$$\exists Y > 0: \qquad Y + \begin{pmatrix} A_{11}^{\mathsf{T}} & A_{21}^{\mathsf{T}} \\ 0 & 0 \end{pmatrix} Y^{-1} \begin{pmatrix} A_{11} & 0 \\ A_{21} & 0 \end{pmatrix} = I,$$

which is the case if and only if

$$\exists Z > 0: \qquad Z + A_{11}^{\mathsf{T}} Z^{-1} A_{11} = I - A_{21}^{\mathsf{T}} A_{21}, \quad \text{where } G := I - A_{21}^{\mathsf{T}} A_{21} > 0.$$

Now define $A := G^{-1/2} A_{11} G^{-1/2}$. Then this problem can be rewritten in the original form (1), unless $A_{11} = 0$. If $A_{11} \neq 0$, we return to (i); otherwise, $Z := I - A_{21}^{\mathsf{T}} A_{21}$ and the algorithm stops. ■

So, to solve the problem (1) we can restrict ourselves to invertible matrices. But from the algorithm it is clear that then the solvability conditions for noninvertible matrices become rather involved. For that reason we will not make this invertibility assumption with respect to A from the outset.

The following preparatory lemma gives a lower bound for any solution to (1).

LEMMA 2. *If the problem* (1) *has a solution* X, *then* $X > AA^{\mathsf{T}}$.

Proof. Rewriting (1) yields that

$$X = I - A^{\mathsf{T}} X^{-1} A. \tag{i}$$

Since X is positive definite, application of Schur's lemma to (i) [see e.g. Kailath (1980, p. 656)] yields that $AA^{\mathsf{T}} - X$ is invertible and in particular that

$$X^{-1} = I - A^{\mathsf{T}}(AA^{\mathsf{T}} - X)^{-1}A. \tag{ii}$$

Now consider $(X - AA^{\mathsf{T}})^{-1}$. Applying Schur's lemma once again yields that

$$\begin{aligned}(X - AA^{\mathsf{T}})^{-1} &= X^{-1} - X^{-1}A(A^{\mathsf{T}}X^{-1}A - I)^{-1}A^{\mathsf{T}}X^{-1}\\ &= X^{-1} + X^{-1}AX^{-1}A^{\mathsf{T}}X^{-1},\end{aligned}$$

which is clearly positive definite. ■

COROLLARY 3. *If the problem* (1) *has a solution* X, *then* $I - AA^{\mathsf{T}} - A^{\mathsf{T}}A > 0$.

Proof. Since $X + A^{\mathsf{T}}X^{-1}A = I$, we obtain by substitution of (ii) from Lemma 2 that

$$X + A^{\mathsf{T}}A - A^{\mathsf{T}\,2}(AA^{\mathsf{T}} - X)^{-1}A^{2} = I.$$

So $X - AA^{\mathsf{T}} = I - AA^{\mathsf{T}} - A^{\mathsf{T}}A - A^{\mathsf{T}\,2}(X - AA^{\mathsf{T}})^{-1}A^{2}$. Application of Lemma 2 yields that

$$I - AA^{\mathsf{T}} - A^{\mathsf{T}}A = (X - AA^{\mathsf{T}}) + A^{\mathsf{T}\,2}(X - AA^{\mathsf{T}})^{-1}A^{2} > 0. \quad ■$$

A similar result to Lemma 2 holds with respect to the problem (2).

LEMMA 4. *If the problem* (2) *has a solution, then there exists a positive constant* α *such that* $X_n > \alpha I \ \forall n \in \mathbb{N}$.

Proof. The proof is similar to that of our General Observation 1.

(i) In case the matrix A is invertible, the above statement is trivial.
(ii) In case A is not invertible, we decompose A again into

$$A =: T^{\mathsf{T}}\begin{pmatrix} A_{11} & 0 \\ A_{21} & 0 \end{pmatrix}T$$

and note that the algorithm of Theorem 1 implies that the next algorithm has a solution $X' > 0$:

$$X'_0 = I,$$

$$X'_{n+1} = I - A_{21}^{\mathsf{T}} A_{21} - A_{11}^{\mathsf{T}} X_n'^{-1} A_{11}, \qquad \text{where} \quad X_n = T^{\mathsf{T}} \begin{pmatrix} X'_n & 0 \\ 0 & I \end{pmatrix} T.$$

Rewriting this equation yields

$$\exists Y > 0: \qquad Y_0 = I,$$

$$Y_{n+1} = I - A''^{\mathsf{T}} Y_n^{-1} A'',$$

where

$$A'' := \left(I - A_{21}^{\mathsf{T}} A_{21}\right)^{-1/2} A_{11} \left(I - A_{21}^{\mathsf{T}} A_{21}\right)^{-1/2}.$$

If A_{11} is nonzero, then we return to (i). If $A_{11} = 0$, then it is clear that $X'_n = I - A_{21}^{\mathsf{T}} A_{21} > 0 \ \forall n$, and thus $X_n > \alpha I \ \forall n$ for some α too. Finally notice that this algorithm stops after at most $\bar{n} - 1$ iterations, and that the nested solution

$$X = T_1^{\mathsf{T}} \begin{pmatrix} T_2^{\mathsf{T}} & \cdots & T_2 & 0 \\ & 0 & & I \end{pmatrix} T_1$$

can always be estimated by αI for some α, which completes the proof. ■

Using these two lemmas, we can prove that the problem (1) has a solution whenever (2) has one, and vice versa.

THEOREM 5. *The problem* (1) *has a solution if and only if the problem* (2) *has one.*

Proof. $\Leftarrow$: First we prove that X_n is a monotonically nonincreasing sequence. This is proved by induction. Note that the initialization step is trivially satisfied, for

$$X_0 - X_1 = I - (I - A^{\mathsf{T}}A) = A^{\mathsf{T}}A \geqslant 0.$$

Now, let $X_n \leqslant X_{n-1}$. Then, since $X_n > 0$, we have that $X_n^{-1} \geqslant X_{n-1}^{-1}$. So $X_n - X_{n+1} = A^{\mathsf{T}}(X_{n-1}^{-1} - X_n^{-1})A \geqslant 0$, which completes the induction step. Therefore X_n is a monotonically nonincreasing sequence which is, according to Lemma 4, bounded from below by some positive definite matrix. Consequently X_n converges to a positive definite limit which satisfies Equation (1).

$\Rightarrow$: We prove this part also by induction.

According to Corollary 3, we have that whenever (1) has a solution, $I - A^{\mathsf{T}}A - AA^{\mathsf{T}}$ is positive definite. Since $X_1 - AA^{\mathsf{T}} = I - A^{\mathsf{T}}A - AA^{\mathsf{T}}$, this completes the first part of the proof.

Now assume that $X_i - AA^{\mathsf{T}} > 0 \;\; \forall i \leqslant n$. Then it is easily seen by induction that $X_i - X = A^{\mathsf{T}}(X^{-1} - X_{i-1}^{-1})A \geqslant 0 \;\; \forall i \leqslant n$. So in particular $X - AA^{\mathsf{T}} \leqslant X_n - AA^{\mathsf{T}}$. Application of this inequality yields that

$$\begin{aligned}
X_{n+1} - AA^{\mathsf{T}} &= I - AA^{\mathsf{T}} - A^{\mathsf{T}}X_n^{-1}A \\
&= I - A^{\mathsf{T}}\left(X_n - AA^{\mathsf{T}}\right)^{-1}A - AA^{\mathsf{T}} \\
&= \left[I + A^{\mathsf{T}}\left(X_n - AA^{\mathsf{T}}\right)^{-1}A\right]^{-1} - AA^{\mathsf{T}} \\
&\geqslant \left[I + A^{\mathsf{T}}\left(X - AA^{\mathsf{T}}\right)^{-1}A\right]^{-1} - AA^{\mathsf{T}} \\
&= I - A^{\mathsf{T}}X^{-1}A - AA^{\mathsf{T}} \\
&= X - AA^{\mathsf{T}} \\
&> 0,
\end{aligned}$$

which completes the proof. ∎

III. NECESSARY CONDITIONS

In this section we discuss a number of conditions on A that must be satisfied in order to solve the matrix equation. Moreover we show by means of a counterexample that these conditions are in general not sufficient to solve the problem.

We start this section again with a preliminary lemma. In this lemma, as well as in the rest of the paper, we use $r(A)$ to denote the spectral radius of the matrix A (i.e. $\max_{\lambda_i}|\lambda_i|$, where λ_i are the eigenvalues of A).

LEMMA 6. *Let P and Q be two arbitrary compatible matrices. Then* $r(P^{\mathsf T}Q - Q^{\mathsf T}P) \leqslant r(P^{\mathsf T}P + Q^{\mathsf T}Q)$.

Proof. By elementary calculus we have that

$$r(P^{\mathsf T}Q - Q^{\mathsf T}P) = r\left(\begin{pmatrix} P^{\mathsf T} & Q^{\mathsf T} \end{pmatrix}\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}\begin{pmatrix} P \\ Q \end{pmatrix}\right).$$

Since $r(AB) = r(BA)$ for any two compatible matrices, we have that

$$r\left(\begin{pmatrix} P^{\mathsf T} & Q^{\mathsf T} \end{pmatrix}\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}\begin{pmatrix} P \\ Q \end{pmatrix}\right) = r\left(\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}\begin{pmatrix} P \\ Q \end{pmatrix}\begin{pmatrix} P^{\mathsf T} & Q^{\mathsf T} \end{pmatrix}\right).$$

Now, $r(A) \leqslant \|A\|_2$, where $\|\cdot\|_2$ denotes the operator norm (i.e. the largest singular value of A). So

$$\begin{aligned} r\left(\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}\begin{pmatrix} P \\ Q \end{pmatrix}\begin{pmatrix} P^{\mathsf T} & Q^{\mathsf T} \end{pmatrix}\right) &\leqslant \left\|\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}\begin{pmatrix} P \\ Q \end{pmatrix}\begin{pmatrix} P^{\mathsf T} & Q^{\mathsf T} \end{pmatrix}\right\|_2 \\ &\leqslant \left\|\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}\right\|_2 \left\|\begin{pmatrix} P \\ Q \end{pmatrix}\begin{pmatrix} P^{\mathsf T} & Q^{\mathsf T} \end{pmatrix}\right\|_2. \end{aligned}$$

As

$$\begin{pmatrix} P \\ Q \end{pmatrix}\begin{pmatrix} P^{\mathsf T} & Q^{\mathsf T} \end{pmatrix}$$

is a normal matrix, and $\|A\|_2 = r(A)$ for any matrix A of this type, we can rewrite the above expression as follows:

$$\begin{aligned} r\left(P^{\mathsf T}Q - Q^{\mathsf T}P\right) &\leqslant \left\|\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}\right\|_2 \left\|\begin{pmatrix} P \\ Q \end{pmatrix}\begin{pmatrix} P^{\mathsf T} & Q^{\mathsf T} \end{pmatrix}\right\|_2 \\ &= 1 r\left(\begin{pmatrix} P \\ Q \end{pmatrix}\begin{pmatrix} P^{\mathsf T} & Q^{\mathsf T} \end{pmatrix}\right) \\ &= r\left(\begin{pmatrix} P^{\mathsf T} & Q^{\mathsf T} \end{pmatrix}\begin{pmatrix} P \\ Q \end{pmatrix}\right) \\ &= r(P^{\mathsf T}P + Q^{\mathsf T}Q), \end{aligned}$$

which completes the proof. ∎

THEOREM 7. *Assume that the problem* (1) *is solvable. Then the matrix A satisfies the following inequalities*:

(i) $r(A) \leqslant \frac{1}{2}$,
(ii) $r(A + A^{\mathsf{T}}) \leqslant 1$,
(iii) $r(A - A^{\mathsf{T}}) \leqslant 1$.

Proof. (i): Let x be an eigenvector corresponding to an eigenvalue λ of A. Then rewriting the equality

$$x^{\mathsf{T}}Xx + x^{\mathsf{T}}A^{\mathsf{T}}X^{-1}Ax = x^{\mathsf{T}}x$$

yields

$$x^{\mathsf{T}}Xx + |\lambda|^2 x^{\mathsf{T}}X^{-1}x = x^{\mathsf{T}}x,$$

from which we deduce that

$$|\lambda|^2 = \frac{x^{\mathsf{T}}(I - X)x}{x^{\mathsf{T}}X^{-1}x}. \qquad (*)$$

Since X is a symmetric positive definite matrix, we can make a singular value decomposition of X into $U^{\mathsf{T}}\Sigma U$, where U is an orthogonal matrix and $\Sigma = \operatorname{diag}(\sigma_i^2)$ [see e.g. Kailath (1980, p. 667)].

Now, introduce the variable $y = Ux$. Then we have from $(*)$ that

$$|\lambda|^2 = \frac{y^{\mathsf{T}}(I - \Sigma)y}{y^{\mathsf{T}}\Sigma^{-1}y}.$$

So it suffices to prove that $y^{\mathsf{T}}(I - \Sigma)y/y^{\mathsf{T}}\Sigma^{-1}y \leqslant \frac{1}{4}$, or equivalently, that $y^{\mathsf{T}}(I - \Sigma - \frac{1}{4}\Sigma^{-1})y \leqslant 0$. As

$$y^{\mathsf{T}}\left(I - \Sigma - \tfrac{1}{4}\Sigma^{-1}\right)y = \sum_{i=1}^{n} y_i^2\left(1 - \sigma_i^2 - \frac{1}{4\sigma_i^2}\right)$$

$$= \sum_{i=1}^{n} -y_i^2\left(\sigma_i^2 - \tfrac{1}{2}\right)^2\frac{1}{\sigma_i^2},$$

which is clearly smaller than zero, this proves the first claim.

(ii): To prove the other two claims we introduce the following notation:

$$P := X^{1/2} - X^{-1/2}A,$$

$$Q := X^{1/2} + X^{-1/2}A.$$

With this notation Equation (1) can be rewritten as either $P^{\mathsf{T}}P = I - A - A^{\mathsf{T}}$ or $Q^{\mathsf{T}}Q = I + A + A^{\mathsf{T}}$. Since both $P^{\mathsf{T}}P$ and $Q^{\mathsf{T}}Q$ are positive semidefinite, this proves claim (ii).

(iii): Using the above notation we have, moreover, that

$$A - A^{\mathsf{T}} = \tfrac{1}{2}(P^{\mathsf{T}}Q - Q^{\mathsf{T}}P).$$

Application of Lemma 6 yields then that

$$\begin{aligned} r(A - A^{\mathsf{T}}) &= \tfrac{1}{2}r(P^{\mathsf{T}}Q - Q^{\mathsf{T}}P) \leqslant \tfrac{1}{2}r(P^{\mathsf{T}}P - Q^{\mathsf{T}}Q) \\ &= \tfrac{1}{2}r(I - A - A^{\mathsf{T}} + I + A + A^{\mathsf{T}}) \\ &= \tfrac{1}{2}r(2I) \\ &= 1. \end{aligned}$$

The proof of this part is completed by noting that

$$\begin{aligned} r\left((A - A^{\mathsf{T}})(A - A^{\mathsf{T}})^{\mathsf{T}}\right) = r\left(-(A - A^{\mathsf{T}})^2\right) &= r\left((A - A^{\mathsf{T}})^2\right) \\ &= \left[r(A - A^{\mathsf{T}})\right]^2. \end{aligned}$$

■

Other necessary conditions can be formulated too, e.g. $r(AA^{\mathsf{T}} + A^{\mathsf{T}}A) < 1$ (see Corollary 3) or $r(A^2 + A^{\mathsf{T}\,2}) \leqslant \frac{1}{2}$. These additional conditions, however, do not give much extra information about A. Moreover, even together with the conditions posed in Theorem 6, they are not yet sufficient to conclude solvability of the matrix equation, as will be shown in Example 8. For that reason we will not go into any further details on this subject here.

EXAMPLE 8. Let

$$A = \begin{pmatrix} 0.5 & -0.45 \\ 0.45 & 0 \end{pmatrix}.$$

Then all necessary conditions mentioned before are satisfied. However, from the simulation results performed with the algorithm (2) (see Appendix 1) it is clear that X is not positive definite. So according to Theorem 5, the problem (1) does not have a solution.

We conclude this section with two examples of the 2×2 matrix case in which the above-mentioned conditions are sufficient. They might be useful in future research on obtaining a general analytic expression for a solution of the equation. That the stated solutions indeed satisfy the equation can be verified by elementary calculation.

EXAMPLE 9. Let

$$A = \begin{pmatrix} 0 & a_{12} \\ 0 & a_{22} \end{pmatrix}.$$

Then, with

$$x_{22} = \frac{1 - a_{12}^2 + \sqrt{(1 - a_{12}^2)^2 - 4a_{22}^2}}{2},$$

the matrix

$$X = \begin{pmatrix} 1 & 0 \\ 0 & x_{22} \end{pmatrix}$$

satisfies Equation (1). Moreover, X can be rewritten as follows:

$$X = \tfrac{1}{2}\left(I - G + \sqrt{(I + G)^2 - 4A^{\mathsf{T}}A}\right), \qquad (*)$$

where $G = (A - A^{\mathsf{T}})(A - A^{\mathsf{T}})^{\mathsf{T}}$.

EXAMPLE 10. Let

$$A = \begin{pmatrix} 0 & a_{12} \\ a_{21} & 0 \end{pmatrix}.$$

Then, with

$$x_{11} = 1 + a_{12}^2 - a_{21}^2 + \sqrt{(1 - a_{21} + a_{12}^2) - 4a_{12}^2} \quad \text{and}$$

$$x_{22} = 1 + a_{21}^2 - a_{12}^2 + \sqrt{(1 - a_{12}^2 + a_{21}^2)^2 - 4a_{21}^2},$$

the matrix

$$X = \frac{1}{2}\begin{pmatrix} x_{11} & 0 \\ 0 & x_{22} \end{pmatrix}$$

satisfies Equation (1). Moreover, X can be rewritten like $(*)$ in Example 9 with G replaced by $G = AA^{\mathsf{T}} - A^{\mathsf{T}}A$.

Note that the two analytic solutions stated here do not coincide. Now, one might hope that this is due to the fact that in Example 10 the matrix A is generically invertible, whereas in Example 9 it is not. Stated differently, one might guess that the analytic solution presented in Example 10 solves the problem whenever A is invertible. But unfortunately this is not the case. Take e.g.

$$A = \begin{pmatrix} 0.2 & 0.4 \\ 0.05 & 0.25 \end{pmatrix};$$

then

$$X := \begin{pmatrix} 0.95 & -0.117 \\ -0.117 & 0.701 \end{pmatrix},$$

and simple calculations show that this is a counterexample for the conjecture. However, there is a class of matrices for which this formula does make sense. These are the normal matrices. In the next section we will see that if matrix A is normal, condition (i) of Theorem 7 is sufficient in itself to conclude solvability of Equation (1), and that a solution is given by $(*)$ where G is as in Example 10.

IV. SOME SPECIAL CASES

Using the theory developed in the previous section, we derive in the present section a sufficient condition for the existence of a solution. The claim is that whenever the operator norm of A is smaller than $\frac{1}{2}$, then there exists a solution. In particular, if A is normal, this implies that the equation has a solution iff the spectral radius of A is smaller than $\frac{1}{2}$. We first prove this last-mentioned result, since in that case a geometric approach is possible which facilitates a constructive proof.

THEOREM 11. *Let A be normal. Then the problem* (1) *has a solution iff* $r(A) \leqslant \frac{1}{2}$.

Proof. That the spectral condition is necessary was already proved in Theorem 7. To prove its sufficiency we recall from elementary matrix theory (see e.g. Horn and Johnson, p. 105) the result that matrix A is normal iff there is a real orthogonal matrix U such that

$$U^{\mathsf{T}}AU = \operatorname{diag}(D_i),$$

where each D_i is either a real 1×1 matrix or a real 2×2 matrix of the form

$$D_i = \begin{pmatrix} \lambda_i & \mu_i \\ -\mu_i & \lambda_i \end{pmatrix}.$$

An immediate consequence of this result is that the problem (1) is solvable iff $\exists Z > 0$: $Z + D^{\mathsf{T}}Z^{-1}D = I$, where $D := \operatorname{diag}(D_i)$ and $r(D_i) \leqslant \frac{1}{2}$. By construction we show now that this problem always has a diagonal solution $Z := \operatorname{diag}(z_i)$.

To that end we first consider the case that D_i is a 1×1 matrix. Then we have to solve the equation $z_i + D_i/z_i = 1$. Since $D_i \leqslant \frac{1}{2}$, it is easily seen that this quadratic equation always has a positive solution.

In case D is a 2×2 matrix, we note the assumption $r(D_i) \leqslant \frac{1}{2}$ in particular implies that $1 - 4(\lambda_i^2 + \mu_i^2)$ is nonnegative. Now take

$$Z = \operatorname{diag}\left(1 + \sqrt{1 - 4(\lambda_i^2 + \mu_i^2)}\right).$$

Straightforward calculation shows then that Z indeed satisfies the equation $Z + D^{\mathsf{T}}Z^{-1}D = I$ and that, moreover, Z is positive definite. This completes the proof. ∎

REMARK 12. By some matrix manipulation it can be shown that $X_1 = \frac{1}{2}[I + (I - 4A^{\mathsf{T}}A)^{1/2}]$ and $X_2 = \frac{1}{2}[I - (I - 4A^{\mathsf{T}}A)^{1/2}]$ always satisfy the equation. These expressions clearly generalize the scalar case.

A question which now immediately arises in this context is whether the set of all solutions satisfying Equation (1) has a smallest (X') and largest (X'') element, respectively, in the sense that any other solution X satisfies the inequality $X' \leqslant X \leqslant X''$. In the particular case of Theorem 11 a natural guess at X' and X'' would then be X_2 and X_1 respectively. This remains, however, a topic for future research.

In the next theorem we show that in general the condition that the largest singular value of matrix A is smaller than $\frac{1}{2}$ is sufficient to conclude that the problem (1) has a solution. The proof is given by showing that (2) has a solution under this assumption. The disadvantage of this approach is that the connection with the analytic solution is lost.

THEOREM 13. *Let σ^2 denote the largest singular value of A. Then the problem* (1) *has a solution if $\sigma^2 \leqslant \frac{1}{2}$.*

Proof. Consider the "equivalent" problem (2). We show by induction that under the above-mentioned assumption, $X_n > AA^{\mathsf{T}} + \frac{1}{4}I$.

The first step is rather trivial. For, since $\sigma^2 \leqslant \frac{1}{2}$, $AA^{\mathsf{T}} \leqslant \frac{1}{4}I$. Consequently, $X_0 = I > \frac{1}{2}I \geqslant AA^{\mathsf{T}} + \frac{1}{4}I$.

Now assume that $X_n \geqslant AA^{\mathsf{T}} + \frac{1}{4}I$. Then

$$\begin{aligned} X_{n+1} &= I - A^{\mathsf{T}}\left(X_n - AA^{\mathsf{T}} + AA^{\mathsf{T}}\right)^{-1} A \\ &= \left[I + A^{\mathsf{T}}\left(X_n - AA^{\mathsf{T}}\right)^{-1} A\right]^{-1} \\ &\geqslant \left(I + 4A^{\mathsf{T}}A\right)^{-1} \\ &\geqslant \tfrac{1}{2}I \\ &\geqslant AA^{\mathsf{T}} + \tfrac{1}{4}I. \end{aligned}$$

So $X_n > AA^{\mathsf{T}}$ $\forall n \in \mathbb{N}$. Therefore, according to Theorem 5, the problem (1) has a solution. ■

V. AN EXAMPLE FROM CONTROL THEORY

In this section we give an example in the field of control theory, where the solvability of Equation (1) plays an important role.

Consider the following optimal control problem:

$$\min_{u[0,\cdot]} \lim_{N\to\infty} J_N \qquad \text{w.r.t.} \quad x(k+1) = A'x(k) + Bu(k),$$

$$x(\cdot) \in \mathbb{R}^n,\ u(\cdot) \in \mathbb{R}^n, \quad x(0) = x. \quad (3)$$

with the additional constraint that $\lim_{N\to\infty} x(N) = 0$, where

$$J_N = \sum_{k=0}^{N-1} \{x^{\mathsf{T}}(k)Qx(k) + u^{\mathsf{T}}(k)Ru(k)\}$$

and both Q and R are symmetric. It is well known that it is difficult to find explicit solvability conditions for this so-called indefinite linear quadratic (LQ) control problem [see e.g. Jonckheere and Silverman (1978) and Molinari (1973)]. We will show that in case B is invertible the solvability of an appropriate equation of the type (1) plays a crucial role. But first we state sufficient general solvability conditions for the problem (3).

THEOREM 14. *The problem* (3) *has a solution if there exists a real solution* K' *of the algebraic Riccati equation*

$$K = A'^{\mathsf{T}}\left\{K - KB(R + B^{\mathsf{T}}KB)^{-1}B^{\mathsf{T}}K\right\}A' + Q \qquad \text{(ARE)}$$

which additionally satisfies the requirements

(i) $R + B^{\mathsf{T}}K'B > 0$,
(ii) $r(A' + BF) < 1$, where $F = -(R + B^{\mathsf{T}}K'B)^{-1}B^{\mathsf{T}}K'A'$.

Proof. It is well known that by introducing the variable $v(k) = u(k) - Fx(k)$ the cost functional can be rewritten as

$$\min_{u[0,\cdot]} \lim_{N\to\infty} J_N = \min_{u[0,\cdot]} \lim_{N\to\infty} \left[J_N + x^{\mathsf{T}}(N)K'x(N) - x^{\mathsf{T}}(N)K'x(N)\right]$$

$$= \min_{u[0,\cdot]} \lim_{N\to\infty} \left(\sum_{k=0}^{N-1} v^{\mathsf{T}}(k)(R + B^{\mathsf{T}}K'B)v(k)\right.$$

$$\left. + x^{\mathsf{T}}(0)K'x(0) - x^{\mathsf{T}}(N)K'x(N)\right), \qquad (*)$$

where $x(k+1) = (A' + BF)x(k) + Bv(k)$.

Now take $v(\cdot) = 0$. Then, due to our assumption on $r(A' + BF)$, $x(N)$ converges to zero. Consequently, the minimum value of the problem (3) is always equal or smaller than $x^{\mathsf{T}}K'x$.

Moreover, since the control sequence must be such that $x(N)$ converges to zero and $R + B^{\mathsf{T}}K'B > 0$, we have from $(*)$ that always $\min \lim J_N \geqslant x^{\mathsf{T}}K'x$. So $v(k) = 0$ solves the problem, which completes the proof. ■

Thus the problem left to be solved is to give conditions under which there exists a real symmetric matrix K' to (ARE) which additionally satisfies Theorem 14(i) and (ii).

THEOREM 15. *Let* $M := RB^{-1}A'B$ *and* $N := B^{\mathsf{T}}A'^{\mathsf{T}}B^{-\mathsf{T}}RB^{-1}A'B + R + B^{\mathsf{T}}QB$. *There exists a real symmetric solution* K' *to* (ARE) *satisfying Theorem* 14(i) *if and only if*

(1) $N > 0$,
(2) *the problem* (1) *has a solution with* $A := N^{-1/2}MN^{-1/2}$.

Proof. Consider (ARE). Some elementary matrix manipulation shows that (ARE) has a real symmetric solution satisfying (i) if and only if the following equation has such a solution:

$$R + B^{\mathsf{T}}KB = -B^{\mathsf{T}}A'^{\mathsf{T}}KB(R + B^{\mathsf{T}}KB)^{-1}B^{\mathsf{T}}KA'B + B^{\mathsf{T}}A'^{\mathsf{T}}KA'B + R + B^{\mathsf{T}}QB.$$

This equation can be rewritten as

$$R + B^{\mathsf{T}}KB = -B^{\mathsf{T}}A'^{\mathsf{T}}B^{-\mathsf{T}}(-R + R + B^{\mathsf{T}}KB) \times (R + B^{\mathsf{T}}KB)^{-1}(R + B^{\mathsf{T}}KB - R)B^{-1}A'B + B^{\mathsf{T}}A'^{\mathsf{T}}KA'B + R + B^{\mathsf{T}}QB,$$

which yields

$$R + B^{\mathsf{T}}KB = -B^{\mathsf{T}}A'^{\mathsf{T}}B^{-\mathsf{T}}R(R + B^{\mathsf{T}}KB)^{-1}RB^{-1}A'B + B^{\mathsf{T}}A'^{\mathsf{T}}B^{-\mathsf{T}}RB^{-1}A'B + R + B^{\mathsf{T}}QB.$$

So, introducing $Y := R + B^{\mathsf{T}}KB$, we see that there is a solution satisfying Theorem 14(i) if and only if there exists a real positive definite solution Y to

$$Y = -M^{\mathsf{T}}Y^{-1}M + N.$$

The stated conditions (1) and (2) now immediately result from this equation. ■

Combining the main results of this section and the previous one, we have the following corollary.

COROLLARY 16. *With the notation of Theorem* 15, *the indefinite LQ problem* (3) *has a solution if the following conditions are satisfied*:

(1) $N > 0$,
(2) $\|N^{-1/2}MN^{-1/2}\|_2 \leqslant \frac{1}{2}$,
(3) $r((R + B^{\mathsf{T}}KB)^{-1}M) < 1$.

Note that in the definite LQ problems condition (3) is always satisfied.

VI. CONCLUDING REMARKS

In this paper we have introduced a nonlinear equation which directly extends the well-known scalar quadratic equation. It turned out that it is rather difficult to find necessary and sufficient conditions for the existence of a real symmetric positive definite solution. For that reason we formulated a recursive algorithm from which a solution can always be calculated numerically whenever the equation is solvable.

Of course, the equation has in general more than one solution. Therefore the question arises whether all solutions can be ordered in some way, and in particular, whether there exist a smallest and a largest element. Drawing a parallel with the properties of the solutions satisfying the algebraic Riccati equation [see e.g. Willems (1971), Trentelman (1987), and Ran and Vreugdenhil (1988)], we believe that this minimal and maximal element exist, and that our recursive algorithm converges to the maximal one. But this remains a topic for future research.

Here, we have concentrated on finding solvability conditions which can be easily verified, and the derivation of an analytic solution. We showed that whenever the operator norm of the matrix A is smaller then $\frac{1}{2}$, the equation is always solvable. In case A is normal, this condition is both necessary and sufficient, and we gave an analytic solution. To find an explicit solution in other cases was rather difficult. Only in the 2×2 case for some particular situations were general formulas derived, which unfortunately do not solve the equation in general.

Since we were not able to solve the general problem, we also derived a number of simple nontrivial necessary conditions that are expressed in terms of spectral radii.

In this context it is interesting to note that in a recent paper Lerer (1989) also studied quadratic matrix equations. He treats these problems from a factorization point of view. It may be that this different approach will give rise to additional explicit solvability conditions. But this is again a matter for future research.

We concluded the discussion with an example from optimal control theory. It concerns the indefinite linear quadratic optimization problem. We showed that in case the input matrix B is invertible, the optimal control problem can in essence be reduced to the question whether a special quadratic matrix equation of the type we studied is solvable. Using the developed theory, we gave sufficient solvability conditions. In fact, the analysis can be straightforwardly extended to right invertible B matrices. The general case is however more difficult, and it is yet unclear whether our analysis can be used to solve the Riccati equation in that case too.

Finally, we note that the results we obtained in Section 5 with respect to (right) invertible matrices B are more general than those obtained by Ran and Vreugdenhil (1988) in that the matrix R is here not assumed *a priori* to be positive definite.

APPENDIX 1

The first four values for X_n in $X_{n+1} = I - A^{\mathsf{T}} X_n^{-1} A$, $X_0 = I$, with

$$A = \begin{pmatrix} 0.500 & -0.450 \\ 0.450 & 0.000 \end{pmatrix},$$

are

$$X_0 = \begin{pmatrix} 1.000 & 0.000 \\ 0.000 & 1.000 \end{pmatrix},$$

$$X_1 = \begin{pmatrix} 0.548 & 0.225 \\ 0.225 & 0.798 \end{pmatrix},$$

$$X_2 = \begin{pmatrix} 0.459 & 0.347 \\ 0.347 & 0.582 \end{pmatrix},$$

$$X_3 = \begin{pmatrix} 0.439 & 0.414 \\ 0.414 & 0.196 \end{pmatrix},$$

$$X_4 = \begin{pmatrix} 0.433 & 0.465 \\ 0.465 & 1.463 \end{pmatrix}.$$

REFERENCES

Horn, R. A. and Johnson, C. A. 1985. *Matrix Analysis*, Cambridge, U.P., London.

Jonckheere, E. A. and Silvermann, L. M. 1978. Spectral theory of the linear quadratic optimal control problem; discrete-time single-input case, *IEEE Trans. Circuits and Systems* CAS-25.

Kailath, T. 1980. *Linear Systems*, Prentice-Hall, Englewood Cliffs, N.J.

Lerer, L. 1989. The algebraic Riccati equation and inverse spectral problems for rational matrix functions, presented at MTNS-89, Amsterdam.

Molinari, B. P. 1973. The stabilizing solution of the algebraic Riccati equation, *SIAM J. Control Optim.* 11:262–271.

Ran, A. C. M. and Vreugdenhil, R. 1988. Existence and comparison theorems for algebraic Riccati equations for continuous- and discrete-time systems, *Linear Algebra Appl.* 99:63–83.

Trentelman, H. L. 1987. Families of linear-quadratic problems, *IEEE Trans. Automat. Control* AC-32:323–329.

Willems, J. C. 1971. Least squares stationary optimal control and the algebraic Riccati equation, *IEEE Trans. Automat. Control* AC-16:621–634.

Received 11 August 1989; final manuscript accepted 21 May 1990