A Review of the Matrix Riccati Equation

VLADIMÍR KUČERA

This paper reviews some basic results regarding the matrix Riccati equation of the optimal control and filtering theory. The theoretical exposition is divided into three parts dealing respectively with the steady-state algebraic equation, the differential equation, and the asymptotic properties of the solution. At the end a survey of existing computational techniques is given.

INTRODUCTION

As usual, $\mathbb{R}$ denotes the field of real numbers, $\mathbb{R}^n$ stands for the $n$-dimensional vector space over $\mathbb{R}$, a prime denotes the transpose of a matrix, an asterisk denotes the complex conjugate transpose of a matrix, and $P \succeq Q$ means that $P - Q$ is hermitian or real symmetric nonnegative matrix. Square brackets represent matrices composed of the symbols inside.

In order to get a better motivation for the problems to be discussed we first pose the underlying physical problem.

Given the linear, continuous-time, constant system

\begin{align*}
\frac{dx}{dt} &= Ax(t) + Bu(t), \quad x(t_0) = x_0, \\
y(t) &= Hx(t),
\end{align*}

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^r$, and $y \in \mathbb{R}^p$ are the state, the input, and the output of the system respectively and $A$, $B$, $H$ are constant matrices over $\mathbb{R}$ of appropriate dimensions, find a control $u(t)$ over $t_0 \leq t \leq t_f$ which for any $x_0 \in \mathbb{R}^n$ minimizes the cost functional

\begin{align*}
\mathcal{J} &= \frac{1}{2}x'(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f}(x'Qx + u'u) \, dt.
\end{align*}

with $S \succeq 0$, $Q \succeq 0$. 


This problem is referred to as the least squares optimal control problem and it can be solved by the minimum principle of Pontryagin [1], [19], [29], [32], by the dynamic programming of Bellman [1], [2], [7], [15], [19], [32] or by the second method of Lyapunov [33].

The minimum value $f_o$ of (3) is given as

\[ f_o = \frac{1}{2} x'(t_0) P(t_0) x(t_0) \]

and it is attained if and only if the control

\[ u(t) = -B'P(t)x(t) \]

is used. Here $P$ is an $n \times n$ matrix solution of the Riccati differential equation

\[ \frac{dP}{dt} = -P(t)BB'P(t) + P(t)A + A'P(t) + Q, \]

\[ P(t_0) = S. \]

Note that this equation must be solved backward from $t_f$ to $t_0$ in order to obtain the optimal control.

One special case is frequent in applications, namely $t_f \to \infty$, the so called regulator problem. In this particular case it may happen that $P(t)$ approaches a finite constant, $P_\infty$, as $t \to \infty$, or, equivalently, as $t \to -\infty$ in (6). Then

\[ f_o = \frac{1}{2} x'(t_0) P_\infty x(t_0); \]

the control law

\[ u(t) = -B'P_\infty x(t) \]

is independent of time and $P_\infty$ satisfies the quadratic algebraic equation

\[ -PBB'P + PA + A'P + Q = 0. \]

In the sections to follow we first investigate the algebraic equation (9), then the differential equation (6) and the asymptotic behaviour of the solution of (6) as $t \to -\infty$. Finally some computing techniques for both equation (6) and (9) are surveyed.

**THE QUADRATIC EQUATION**

The matrix equation (9) has been extensively studied [6], [16], [22], [23], [26], [30], [36]. It is well-known that it can possess a variety of solutions. First of all (9) may have no solution at all. If it does have one, there can be both real and complex solutions, some of them being hermitian or symmetric. There can be even infinitely many solutions. Due to the underlying physical problem, however, only nonnegative
solutions are of interest to us. Therefore, we are mainly concerned with the existence and uniqueness of such a solution.

In this section we summarize some long-standing as well as recent results [22], [23], [26], [30] on (9) which will prove useful later. First of all, write

\[ Q = C'C, \quad S = D'D. \]

Then \( \lambda \) is said to be an uncontrollable eigenvalue [13], [22] of the pair \((A, B)\) if there exists a row vector \( w \neq 0 \) such that \( wA = \lambda w \) and \( wB = 0 \). Similarly, \( \lambda \) is an unobservable eigenvalue of the pair \((C, A)\) if there exists a vector \( z \neq 0 \) such that \( Az = \lambda z \) and \( Cz = 0 \).

The pair \((A, B)\) is said to be stabilizable [35] if a matrix \( L \) over \( \mathbb{R} \) exists such that \( A + BL \) is stable (i.e., all its eigenvalues have negative real parts), or, equivalently, if the unstable eigenvalues of \((A, B)\) are controllable [13], [35].

Analogically, the pair \((C, A)\) is defined to be detectable [35] if a matrix \( F \) over \( \mathbb{R} \) exists such that \( FC + A \) is stable, or, if the unstable eigenvalues of \((C, A)\) are observable [13], [35].

A nonnegative solution of (9) is said to be an optimizing solution [28] if it yields the optimal control (8); it is called a stabilizing solution [28] if the control (8) is stable. We shall denote these solutions \( P_n \) and \( P_s \), respectively.

Further we introduce the \( 2n \times 2n \) matrix

\[ M = \begin{bmatrix} A & -BB^T \\ -C'C & -A' \end{bmatrix}. \]

Unless otherwise stated we shall henceforth assume that the \( M \) matrix is diagonalizable, that is, it has \( 2n \) eigenvectors. This assumption is made for the sake of simplicity and is by no means essential.

Let \( Ma_i = \lambda_i a_i, \quad r_iM = \lambda'_i r_i, \quad i = 1, 2, \ldots, 2n \),

and write

\[ a_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \quad r'_i = \begin{bmatrix} u'_i \\ v'_i \end{bmatrix}, \]

where \( x_i \in \mathbb{R}^n, \ y_i \in \mathbb{R}^n, \ u_i \in \mathbb{R}^n \) and \( v_i \in \mathbb{R}^n \).

Thus the \( a_i \) is a column vector whereas the \( r_i \) is a row vector. They are sometimes called the right and the left eigenvectors of \( M \), respectively.

It is well-known that the eigenvectors can be chosen so that

\[ r_i a_j = 0, \quad i \neq j, \quad \pm 0, \quad i = j. \]

The following seems to have been proved first in [10], [26] and [30].
Theorem 1. Each solution \( P \) of (9) has the form

\[
P = YX^{-1},
\]

where

\[
X = [x_1, x_2, \ldots, x_n],
\]
\[
Y = [y_1, y_2, \ldots, y_n]
\]
correspond to such a choice of eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_n \) of \( M \) that \( X^{-1} \) exists.

Conversely, all solutions are generated in this way.

Proof. Let \( P \) satisfies (9) and set

\[
K = A - BB'P,
\]

the closed-loop system matrix. Then we infer from (9) that

\[
PK = -Q - A'P
\]

and hence

\[
M \begin{bmatrix} I \\ P \end{bmatrix} = \begin{bmatrix} I \\ P \end{bmatrix} K.
\]

Let

\[
J = X^{-1} K X = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)
\]

be the Jordan canonical form of \( K \) and set \( PX = Y \). Then (13) yields

\[
M \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix} J.
\]

Since \( J \) is diagonal, the columns of \( \begin{bmatrix} X \\ Y \end{bmatrix} \) constitute the eigenvectors of \( M \) associated with \( \lambda_1, \lambda_2, \ldots, \lambda_n \) and \( P = YX^{-1} \).

The converse can be proved by reversing the arguments.

Corollary. The matrix \( K = A - BB'P \) given by the solution (12) has the eigenvalues \( \lambda_i \) associated with the eigenvectors \( x_i \), \( i = 1, 2, \ldots, n. \)

Proof. The \( J \) matrix is the Jordan form of \( K \) and \( X \) is the associated transformation matrix.

Theorem 2. Let \( \lambda_i \) be an eigenvalue of \( M \) and \( \begin{bmatrix} x_i \\ y_i \end{bmatrix} \) the corresponding right eigenvector. Then \( -\lambda_i \) is an eigenvalue of \( M \) and \( \begin{bmatrix} y_i \\ x_i \end{bmatrix} \) the corresponding left eigenvector.
Proof. By direct verification, making use of the identity

$$M' = \begin{bmatrix} 0, & -I \\ I, & 0 \end{bmatrix} M \begin{bmatrix} 0, & -I \\ I, & 0 \end{bmatrix}$$

and the fact that a left eigenvector of $M$ associated with $\lambda$ is a transposed right eigenvector of $M'$ associated with $\lambda$.

Note that $M$ being a real matrix, its eigenvalues occur in quadruples $(\lambda, \lambda^*, -\lambda, -\lambda^*)$.

The procedure described in Theorem 1 is the time-domain counterpart of the spectral factorization in the frequency domain.

We also point out that there can exist at most one stabilizing solution due to the symmetry of the eigenvalues of $M$. If $P = YX^{-1}$ is such a solution, then the matrix $X^*Y$ is hermitian.

The following fundamental theorems originated in [22], [23], [24].

**Theorem 3.** The stabilizing solution of (9) exists if and only if $(A, B)$ is stabilizable and $\text{Re} \ \lambda > 0$ for all eigenvalues $\lambda$ of $M$.

Proof. Necessity: Suppose the stabilizing solution $P_s$ exists. It is associated with $n$ stable eigenvalues of $M$ and hence no eigenvalue can have zero real part due to their symmetrical distribution about the imaginary axis.

Moreover, the matrix $L = -B'P_s$ stabilizes $A + BL$, i.e., the pair $(A, B)$ is stabilizable.

Sufficiency: Suppose the hypothesis holds and let no stabilizing solution exist. Then either (i) we cannot choose $n$ stable eigenvalues of $M$, a contradiction, or (ii) we can do so but the $X$ matrix in (14) is singular.

If (ii) is the case, write $z$ for any nonzero vector of $\mathcal{N}(X)$, the null space of $X$.

Since $X^*Y = Y^*X$, we have

$$0 = YXz = X^*Yz .$$

By virtue of (10) and (14),

$$AX - BB'Y = XJ ,$$

$$-C'CX - A'Y = YJ .$$

The first equation yields

$$z^*Y^*AXz - z^*Y^*BB'Yz = z^*Y^*XJz$$

and hence by (15) and by the definition of $z$

$$BYz = 0 .$$

But

$$0 = AXz - BB'Yz = XJz ,$$
which means that $\mathcal{A}(X)$ is a $J$-invariant subspace of $R^p$. Hence there exists at least one nonzero vector $\mathbf{z} \in \mathcal{A}(X)$ such that

$$J\mathbf{z} = \mu \mathbf{z}$$

where $\mu$ coincides with one of the stable eigenvalues of $M$.

The second equation (16) postmultiplied by $\mathbf{z}$ yields

$$-A'Y\mathbf{z} = YJ\mathbf{z}.$$  

Collecting (17) through (19) gives us

$$\begin{bmatrix} (Y\mathbf{z})' A &= -\mu (Y\mathbf{z})' , & Re (-\mu) > 0 , \\ (Y\mathbf{z})' B &= 0 . \end{bmatrix}$$

Thus $(A, B)$ is not stabilizable, contradicting our hypothesis.

**Theorem 4.** The stabilizing solution is the only nonnegative solution of (9) if and only if $(C, A)$ is detectable.

**Proof.** Sufficiency: Assuming $(C, A)$ detectable we shall demonstrate that any solution $P$ of (9) is the stabilizing solution. Suppose to the contrary that a $X$ exists such that

$$K\mathbf{z} = X\mathbf{z} , \quad Re X > 0 , \quad K = A - BB'P.$$  

On rearranging equation (9) reads

$$PBB'P + PK + K'P + C'C = 0$$

and hence

$$(\lambda + \lambda^*) z^* Pz = -z^* PBB'Pz - z^* C'Cz .$$

Since $\lambda + \lambda^* = 2 \text{Re} \lambda \geq 0$, the left hand side of this equation is nonnegative, while the right hand side is nonpositive. Therefore either is zero and

$$B'Pz = 0 ,$$

$$Cz = 0 ,$$

which in turn implies

$$Az = \lambda \mathbf{z} , \quad Re \lambda \geq 0 ,$$

Thus $(C, A)$ is not detectable contradicting our hypothesis. Hence $K$ is stable.

However, there is at most one stabilizing solution, i.e., the solution $P$ is unique. For another proof refer to [36].

**Necessity:** By contradiction, suppose an undetectable eigenvalue $\lambda_1$ of $(C, A)$
exists. We shall show the existence of at least two nonnegative solutions of equation (9). One of them is the stabilizing solution \( P_s = YY^{-1} \) by hypothesis.

To form another solution \( P_1 = YX^{-1} \) we substitute the eigenvector \( \begin{bmatrix} z_1 \\ 0 \end{bmatrix} \) of \( M \) associated with \( \lambda_1 \) for the eigenvector \( \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \) associated with \( -\lambda_1 \), thus obtaining

\[
X_1 = [z_1, x_2, \ldots, x_n], \quad Y_1 = [0, y_2, \ldots, y_n].
\]

Also set

\[
\hat{X} = [x_2, \ldots, x_n], \quad \hat{Y} = [y_2, \ldots, y_n].
\]

Now Theorem 2 along with (11) implies that

\[
\begin{bmatrix} z_{1}^*, 0^* \end{bmatrix} \begin{bmatrix} -y_1 \\ x_1 \end{bmatrix} = 0, \quad i = 2, 3, \ldots, n.
\]

Hence \( z_{1}^* \hat{Y} = 0 \) and

\[
(20) 
X_1^*Y_1 = \begin{bmatrix} 0, & z_{1}^* \hat{Y} \\ 0, & \hat{X}^*\hat{Y} \end{bmatrix} = \begin{bmatrix} 0, & 0 \\ 0, & \hat{X}^*\hat{Y} \end{bmatrix} \succeq 0
\]
because \( \hat{X}^*\hat{Y} \succeq 0 \).

To prove that \( X_1 \) is nonsingular, suppose the contrary is true. Then a vector \( v \neq 0 \) exists such that \( z_1 = \hat{X} v \) and, consequently,

\[
0 = z_{1}^* \hat{Y} = v^* \hat{X}^*\hat{Y};
\]

that is, \( \det \hat{X}^*\hat{Y} = 0 \). Observing that \( \det \hat{X}^*\hat{Y} \) is a principal minor of the nonnegative matrix \( \hat{X}^*\hat{Y} \), it is easy to see that

\[
[X_1, \hat{X}]^*\hat{Y} v = X^*(\hat{Y}v) = 0.
\]

This is a contradiction, however, as \( X \) is nonsingular and \( \hat{Y} v \neq 0 \).

Thus \( P_1 \) does exist and is different from \( P_s \) because it corresponds to a different \( n \)—tuple of eigenvalues of \( M \). \( \square \)

**Theorem 5.** Stabilizability of \((A, B)\) and detectability of \((C, A)\) is necessary and sufficient for equation (9) to have a unique nonnegative solution which stabilizes the closed-loop system.

**Proof.** Sufficiency part is a well-established result in [36]; it can also be inferred from Theorems 3 and 4.

Necessity part is a simple consequence of Theorems 3 and 4. \( \square \)

We note that optimality does not necessarily imply stability. Indeed, the optimizing solution minimizes the cost functional whereas the stabilizing solution makes the closed-loop stable, and this is quite a different property. However, in control applications an optimal system which is stable as well is desirable. Conditions for this case to hold have been stated in Theorem 5, first published in [22].
Further we turn our attention to the situation when these conditions are not in force. If \((C, A)\) is detectable and \((A, B)\) is not stabilizable, no nonnegative solution of (9) exists. Indeed, \((C, A)\) detectable implies by Theorem 4 that all solutions are the stabilizing solutions. But such a solution does not exist by Theorem 3.

Further let \((A, B)\) be stabilizable and \((C, A)\) be not necessarily detectable. In this case we characterize all nonnegative solutions of (9), see [23], [24].

Let \(\lambda_1, \lambda_2, \ldots, \lambda_p, p \geq 0\), be those eigenvalues of \(M\) that are also the undetectable eigenvalues of \((C, A)\), i.e.,
\[
\begin{align*}
A \xi_i &= \lambda_i \xi_i, \\
C \xi_i &= 0, \\
\Re \lambda_i &\geq 0, \quad i = 1, 2, \ldots, p.
\end{align*}
\]

An eigenvalue \(\lambda\) of \(A\) is called cyclic if any two eigenvectors of \(A\) associated with \(\lambda\) are linearly dependent. We restrict ourselves to the cyclic \(\lambda_i, i = 1, 2, \ldots, p\). Under this condition equation (9) has only a finite number of nonnegative solutions [23], [24]. Otherwise nondenumerable many nonnegative solutions exist as the example \(A = B = I, C = 0, n > 1\) shows, see [23], [24] for details.

Define
\[
\mathcal{E} = \{\lambda_1, \lambda_2, \ldots, \lambda_p\}
\]
and write \(\mathcal{E}_k, k = 1, 2, \ldots\) for the subsets of \(\mathcal{E}\). Note that all \(-\lambda_i, i = 1, 2, \ldots, p\) must be used to form the stabilizing solution if it exists, since \(\Re (-\lambda_i) \leq 0\).

Write \(P_k\) for the solution of (9) which is generated from the stabilizing solution by replacing all \(-\lambda_i\) with \(\lambda_i, \lambda_i \in \mathcal{E}_k\).

The theorems below have originally been proven in [23], [24].

**Theorem 6.** Suppose the stabilizing solution of equation (9) exists. Then the set of all \(P_k\) form the class of all nonnegative solutions of (9), and there are exactly \(2^p\) such solutions.

**Proof.** Existence of all \(P_k\) can be proved in an identical manner as the existence of \(P_1\) in the "if" part of the proof of Theorem 4.

To see that no other nonnegative solution of (9) exists, firstly substitute an eigenvalue \(\lambda \in \mathcal{E}\) for the eigenvalue \(-\lambda\). Arguing as in [23] we conclude that either \(X^*Y \not\geq 0\) or \(X\) is singular and hence the corresponding solution cannot be nonnegative.

Secondly make the substitution of an eigenvalue \(\lambda\), \(\Re \lambda \geq 0\), for an eigenvalue \(\mu\), \(\Re \mu < 0\), \(-\mu \not\geq \lambda\). Applying (11) and Theorem 2 we conclude that \(X^*Y \not\geq Y^*X\) and such a solution cannot be nonnegative, either.

The number of solutions follows from the fact that there are \(2^p\) subsets of a set consisting of \(p\) elements.

**Remark.** If the \(M\) matrix is not diagonalizable, a more refined analysis in [23]
shows that, in general, there exist nonnegative solutions, where
\[ Cz_{t+1} = 0, \quad j = 0, 1, \ldots, s_t - 1, \]
\[ \pm 0, \quad j = s_t \]
and
\[ Az_{t+1} = \lambda_t z_t, \quad j = 0, \]
\[ = \lambda_t z_{t+1} + z_{t+1-j}, \quad j = 1, 2, \ldots, s_t - 1. \]

Theorem 7. Any two nonnegative solutions \( P_k \), \( P_t \) of (9) satisfy
\[ P_k \preceq P_t \text{ if and only if } \mathcal{U}_k \subseteq \mathcal{U}_t. \]

Otherwise speaking, the set of all nonnegative solutions constitute a distributive lattice with respect to the partial ordering \( \preceq \). \( P_0 \) is the smallest nonnegative solution (the zero element of the lattice) and \( P_s \) is the largest nonnegative solution (the identity element of the lattice).

Proof. The claim hinges on the observation that the family of subsets \( \mathcal{U}_t \) constitute a distributive lattice with respect to the partial ordering by inclusion and that (22) is an isomorphism.

In particular, \( \mathcal{U}_k = \emptyset \) implies \( P_k = P_0 \) whereas \( \mathcal{U}_k = \mathcal{U} \) implies \( P_k = P_s \).

The relation (22) itself follows from (20).

Note that if \((C, A)\) is detectable then \( \mathcal{U} = \emptyset \) and hence \( P_0 = P_0 \). However, \((C, A)\) undetectable implies that \( \mathcal{U} \neq \emptyset \) and hence \( P_0 \neq P_s \); the two solutions never coincide:

The physical interpretation of different real nonnegative solutions is as follows [26]. Each nonnegative solution is a conditionally optimizing solution of (9), the condition being a certain degree of stability. Specifically, \( P_k \) stabilizes the undetectable eigenvalues of \((C, A)\) included in \( \mathcal{U}_k \) and no others. Equation (9) thus contains the optimal solutions for all degrees of stability [23], [24], [26]. The idea that the more undetectable eigenvalues is stabilized the higher is the cost \( f \) is made rigorous via the concept of lattice.

THE DIFFERENTIAL EQUATION

In this section we summarize some fundamental results on the Riccati equation (6) which are scattered in the literature [1], [6], [8], [9], [15], [17], [19], [20], [32], [36].

The well-known theorem on differential equations guarantees only local existence and uniqueness for the solution of (6); without further analysis we cannot conclude existence over \( t_0 \leq t \leq t_F \) because of the phenomenon of finite escape time.

Nonetheless, the following is proved in [6], [8], [9], [15], [36].
Theorem 8. Let $P(t, S, t_0)$ be the solution of (6) which passes through $S$ at $t = t_0$. Then given any $t_0$, $P(t, S, t_0)$ exists and is unique on $t_0 \leq t \leq t_f$ regardless of $S$.

Proof. In view of the local existence the $P(t, S, t_0)$ exists for some $t \leq t_f$. Let $\mathcal{V}(t, t_0)$ be the transition matrix of $A$. Then

$$P(t, S, t_0) = \mathcal{V}(t, t_0) S \mathcal{V}(t, t_0) + \int_t^{t_f} \mathcal{V}(t, \tau) \left[ Q - A(t, S, t_0) BB' A(t, S, t_0) \right] \mathcal{V}(t, \tau) d\tau \leq \mathcal{V}(t, t_0) S \mathcal{V}(t, t_0) + \int_t^{t_f} \mathcal{V}(t, \tau) Q \mathcal{V}(t, \tau) d\tau,$$

which can be verified by elementary differentiation. The a priori upper bound above provides a Lipschitz constant on any finite interval $t_0 \leq t \leq t_f$ no matter how large. Hence $P(t, S, t_0)$ exists and is unique globally for any $S$.

Theorem 9. For any $t_0 \leq t \leq t_f$ and any $S \geq 0$,

$$P(t, S, t_s) \succeq 0 .$$

Proof. If $P(t, S, t_s)$ satisfies (6), then $P^*(t, S, t_s)$ does so. Hence $P(t, S, t_s)$ is a hermitian matrix. Moreover, using (4), we obtain (23) since

$$\mathcal{F} = \min_{x^*} \mathcal{F} \geq 0 .$$

Theorem 10. The solution $P$ of (6) enjoys the following properties:

(24) $P(t_0, 0, t_1) \leq P(t_0, 0, t_2)$

for any $t_0 \leq t_1 \leq t_2$,

(25) $P(t_2, 0, t_2) \leq P(t_1, 0, t_1)$

for any $t_1 \leq t_2 \leq t_f$,

(26) $P(t, S_1, t_0) \leq P(t, S_2, t_0)$

for any nonnegative $S_1 \leq S_2$ and $t \leq t_f$.

Proof. Regarding (24), we have

$$x_0^{'} P(t_0, 0, t_1) x_0 = \min_{u} \int_{t_0}^{t_1} (x^{'Q} x + u'u) dt \leq \min_{u} \int_{t_0}^{t_2} (x^{'Q} x + u'u) dt = x_0^{'} P(t_0, 0, t_2) x_0$$

for any $x_0 \in \mathbb{R}^n$. 

Regarding (25), the substitution \( t_1 - t = \tau \) in (3), gives us

\[ P(t_0, 0, t_t) = P(0, 0, t_t - t_0) \]

and hence (25) follows from (24) on identifying \( t_0 \) with \( t_t \) and \( t_t \) in turn.

Finally, (26) is obvious since increasing the terminal penalty results in increased cost functional (3).

We conclude this section by emphasizing that we are not concerned with stability of the optimal closed-loop system since the control problem is defined over a finite interval.

**ASYMPTOTIC BEHAVIOUR OF THE SOLUTION**

Now we let \( t_1 \to \infty \) and investigate the properties of \( P(t) \) over the half-line \( t_0 \leq t < \infty \). Since the Riccati equation coefficients are independent of time, \( P(t_0, S, t_t) = P(0, S, t_1 - t_0) \) and it makes no difference to consider \( t \to -\infty \) and the half-line \(-\infty < t \leq t_t \) instead. The first arrangement reflects the time evolution of the system, whereas the other is convenient from the computational point of view.

It is easy to see that additional assumptions will be required to keep \( P(t) \) from passing off to infinity.

**Theorem 11.** If \((A, B)\) is stabilizable, then the solution \( P(t, S, t_t) \) of (6) is bounded on \(-\infty < t \leq t_t \) regardless of \( S \) and \( t_t \).

**Proof.** Consider a control

\[ \hat{u}(t) = L \hat{x}(t), \quad t_0 \leq t, \]

such that the closed-loop system matrix \( A + BL \) is stable. Then

\[
x_0 P(t_0, S, t_t) x_0 = \min_{u} \left[ x'(t_t) S x(t_t) + \int_{t_0}^{t_t} (x'Qx + u'u) \, dt \right] \leq \int_{t_0}^{t_t} (\hat{x}'Q \hat{x} + \hat{u}'\hat{u}) \, dt = x_0 W(t_0, t) x_0
\]

since \( \hat{x}(t_t) \to 0 \) as \( t_1 \to \infty \), and \( x_0 = \hat{x}_0 \).

Here \( W(t_0, t) \) is a finite matrix over \( R \) for any \( t \leq t_t \). Hence \( P(t_0, S, t) \) is bounded from above for any \( t_0 \), any \( t_0 \leq t < \infty \), and any \( S \).

Equivalently, \( P(t, S, t_t) \) is bounded for any \( t_t \), any \( -\infty < t \leq t_t \), and any \( S \).

Theorem 11 is proved in [36] without any appeal to variational ideas. Our proof, however, is considerably simpler and more intuitive.

First the asymptotic properties of \( P(t, 0, t_t) \) will be discussed and existence of an equilibrium solution of (6) deduced, see [15], [36].
Theorem 12. If \((A, B)\) is stabilizable, then
\[
\lim_{t \to -\infty} P(t, 0, t) = P_\infty,
\]
a finite nonnegative matrix.
Moreover, \(P_\infty\) is a fixed point solution of (6),
\[
P_\infty = P(t, P_\infty, t).
\]

Proof. By Theorem 10, (25), \(P(t, 0, t)\) is monotone nondecreasing in \(t\) in the ordering of nonnegative matrices. Further Theorem 11 implies that \(P(t, 0, t)\) is bounded from above and consequently there exists a finite matrix \(P_\infty \geq 0\) so that \(P(t, 0, t) \to P_\infty\) as \(t \to -\infty\).

Since \(P_\infty\) is independent of \(t\), it satisfies both (6) and (9), that is, \(P_\infty = P(t, P_\infty, t)\).

We remark that even if the stabilizing solution \(P_s\) of (9) exists, \(P_\infty = P_s\) need not be true. Naturally, there arises the question: To which nonnegative solution of (9) does \(P(t, 0, t)\) converge? The answer is contained in the two theorems below [26], [36].

Theorem 13. If \((A, B)\) is stabilizable, then
\[
P_\infty = P_0.
\]

Proof. Since \(S \geq 0\), we infer from Theorem 10, (26) that \(P(t, S, t) \geq P(t, 0, t)\) for any \(S\). Thus, as \(t \to -\infty\), \(P(t, 0, t) \to P_0\), the smallest nonnegative solution of (9).

Theorem 14. If \((A, B)\) is stabilizable and \((C, A)\) is detectable, then
\[
P_\infty = P_\infty = P_s.
\]

Proof. By hypothesis Theorem 5 implies that equation (9) possesses the unique nonnegative solution \(P_s = P_\infty\). Thus the claim follows from Theorem 13.

Now we generalize and consider the asymptotic behaviour of \(P(t, S, t)\) rather than of \(P(t, 0, t)\). Theorem 12 is not known to hold for arbitrary nonzero \(S \geq 0\) as yet, even though some conjectures so that effect has been made [26]. The difficulty stems from the fact that \(P(t, S, t)\) is not monotone nondecreasing in \(t\).

We avoid the impasse by assuming that not only \((A, B)\) is stabilizable but also \(\Re \lambda \neq 0\) for all eigenvalues \(\lambda\) of \(M\). This is equivalent to the assumption that \(P_s\) exists.

Thus we are ready to prove the original results of this paper.
Theorem 15. If $P_s$ exists then

$$\lim_{t \to \infty} P(t, S, t_f) = P_\infty,$$

a finite nonnegative matrix.

Moreover, $P_\infty$ is a fixed point solution of (6),

$$P_\infty = P(t, P_\infty, t_f).$$

Proof. For $S = 0$ the theorem reduces to Theorem 12. Thus it will suffice to show that the term $x'(t_f) S x(t_f)$ approaches zero as $t_f \to \infty$.

By the assumption of $\Re \lambda \neq 0$ this term goes either to zero or to infinity. The latter case is not possible, however, since by Theorem 11 the cost is bounded from above.

Observe again that, in general, $P_\infty$ need not coincide with either $P_s$ or $P_0$. And again we ask: To which nonnegative solution of (9) does $P(t, S, t_f)$ converge?

We have to take $S = D'D$ into account. Let $\mu_1, \mu_2, \ldots, \mu_q$, $q \geq 0$, be those eigenvalues of $M$ that are also the undetectable eigenvalues of $(D, A)$, i.e.,

$$A z_i = \mu_i z_i,$$

$$D z_i = 0,$$

$$\Re \mu_i \geq 0, \quad i = 1, 2, \ldots, q.$$

Define

$$\mathcal{D} = \{\mu_1, \mu_2, \ldots, \mu_q\}$$

in accordance with (21).

Theorem 16. If $P_\infty$ exists then

$$P_\infty = P_k$$

if and only if

$$\mathcal{C}_k = \mathcal{C} \cap \mathcal{D}.$$  

Thus $P(t)$ can be made to approach any nonnegative solution $P_k$ of (9) by taking an appropriate $S$.

Proof. To demonstrate the interaction of $S$ and $Q$ assume first that all undetectable eigenvalues $\lambda_i$ of $(C, A)$ belong to $\mathcal{D}$, where $\mathcal{D}$ corresponds to an arbitrary $S \geq 0$. Then $D z_i = 0$ and hence

$$\lim_{t \to \infty} P(t, S, t_f) = \lim_{t \to \infty} P(t, 0, t_f) = P_\infty$$

by Theorem 13. Note that $P_s$ is generated by the set $\mathcal{C} = \mathcal{C} \cap \mathcal{D}$.

Second, let one undetectable eigenvalue $\lambda_1$ of $(C, A)$ does not belong to $\mathcal{D}$. Then $D z_1 \neq 0$ and

$$\lim_{t \to \infty} P(t, S, t_f) = P_1$$
where $P_t$ is generated by the set $\mathcal{V}_t = \mathcal{V} - \{\lambda_i\} = \mathcal{V} \cap \mathcal{D}$. It is so since otherwise the cost would be infinite due to the term $x(t_t) D' D x(t_t)$. Consequently, $P_{\infty} = P_s$ if and only if $\mathcal{V}_k = \mathcal{V} \cap \mathcal{D}$.

Observe that $P_{\infty}$ is the smallest nonnegative solution of the lattice generated by the set $\mathcal{V} \cap \mathcal{D}$ rather than by the $\mathcal{V}$ itself.

**Theorem 17.** If $(A, B)$ is stabilizable and $(C, A)$ is detectable, then

$$P_{\infty} = P_s = P_t$$

for any $S \geq 0$.

**Proof.** By Theorem 5, equation (9) has the unique nonnegative solution $P_s = P_t$. Thus our claim is proved by referring to Theorem 16.

From the computational viewpoint Theorem 16 renders it possible to find an $S$ such that $P(t, S, t_t)$ will approach a desired equilibrium solution $P_{\infty}$. In particular,

$$P_{\infty} = P_s \quad \text{if and only if} \quad \mathcal{V} \cap \mathcal{D} = \emptyset$$

and

$$P_{\infty} = P_s \quad \text{if and only if} \quad \mathcal{V} \cap \mathcal{D} = \mathcal{V}.$$

The relation

$$\lim_{t \to \infty} P(t, S, t_t) = P_x \iff \mathcal{V} \cap \mathcal{D} = \mathcal{V}_k$$

is an equivalence and, therefore, it induces a decomposition of the cone of non-negative matrices $S$ into equivalence classes. Each class contains those $S$ that make $P(t, S, t_t)$ converge towards a particular $P_x$. Thus we see that $P_{\infty}$ is not a continuous function of $S$ [26]. This is an observation of crucial importance in computations.

**COMPUTATIONAL TECHNIQUES**

This section is concerned with the computational aspects. Various techniques of finding a solution to equations (6) and (9) are reviewed. It is very difficult to label a particular method as being superior to the others. Each method may prove better in one application but it may fail in another. Our aim is to bring the most important methods to the reader's attention and indicate their applicability, advantages and objections.

**Computing the solutions of the algebraic equation**

(i) **Eigenvector Solution**

This method has been reported in [2], [10], [12], [23], [26], [27], [30] and its essentials are stated in Theorem 1. Apart from its theoretical importance, it is computationally promising as efficient and accurate algorithms to compute eigenvalues of a matrix are now available.
This is the only method for computing any solution of (9), no matter whether it is nonnegative or not.

(ii) Iterative Solution

This technique has been first reported in [4], [20], [21] and is believed to be one of the best methods to find the stabilizing solution of (9). It is assumed that (9) possesses a unique nonnegative (stabilizing) solution $P_0$, which is being found by successive linear approximations.

Let $P_j$, $j = 0, 1, \ldots$ be the unique nonnegative solution of the linear algebraic equation

$$PJKJ + K'JPJ + C'C + L'JLJ = 0$$

where, recursively,

$$L_j = -B'P_{j-1}, \quad j = 1, 2, \ldots,$$

$$K_j = A + BL_j$$

and where $L_0$ is chosen such that the matrix $K_0 = A + BL_0$ is stable. Then

$$P_s \leq P_{j+1} \leq P_j \leq \ldots, \quad j = 0, 1, \ldots,$$

and

$$\lim_{j \to \infty} P_j = P_s.$$

Indeed, $K_0$ being stable, (27) has a unique nonnegative solution $P_0$. This yields a $K_1$ and, in turn, a $P_1$. A little manipulation with the associated cost functionals reveals that $P_s \leq P_1 \leq P_0$. $P_s$ being bounded, $K_1$ is stable etc. A theorem on monotonic convergence of nonnegative matrices guarantees the existence of a limit. Then (27) is identical to (9) when $j \to \infty$ and (28) holds by uniqueness of $P_s$.

The method provides monotonic and quadratic convergence. In fact, it is an ingenious modification of Newton’s method reported in [5].

(iii) Solution Using the Sign Function

This method is described in [31] and can be viewed as a simplification of (i) to find the stabilizing solution.

The matrix function sign $Z$ is defined in [31] by

$$\text{sign} \, Z = \lim_{k \to \infty} Z_{k+1},$$

where

$$Z_{k+1} = \frac{1}{4}(Z_k + Z_k^{-1}), \quad k = 0, 1, \ldots,$$

$$Z_0 = Z.$$

Define also

$$\text{sign}^+ \, Z = \frac{1}{4}(I + \text{sign} \, Z).$$
Then assuming $K = A - BB'P$ is stable and defining a matrix $V$ by

$$KV + VK' + BB' = 0,$$

it is easy to check that

$$M = \begin{bmatrix} I, -V \\ P, I - PV \end{bmatrix} \begin{bmatrix} K, 0 \\ 0, -K' \end{bmatrix} \begin{bmatrix} I - VP, V \\ -P, I \end{bmatrix}.$$

Hence

$$\text{sign}^+ M = \begin{bmatrix} I, -V \\ P, I - PV \end{bmatrix} \begin{bmatrix} 0, 0 \\ 0, I \end{bmatrix} \begin{bmatrix} I - VP, V \\ -P, I \end{bmatrix} = \begin{bmatrix} VP, -V \\ -(I - PV) P, I - PV \end{bmatrix}$$

and $P$ follows immediately if $V^{-1}$ exists.

Note that $M$ is not taken through a complete spectral factorization, but into two parts only.

**Computing the solution of the differential equation**

**iv) Transition Matrix Solution**

This old method was rediscovered in [8], [9], [15], [25]. It is important from the theoretical standpoint since it enables us to solve analytically equation (6) even in the time-varying case. However, analytic solutions can be found only in exceptional cases.

Denote

**(29)**

$$p(t) = P(t) x(t).$$

Then

**(30)**

$$\frac{d}{dt} \begin{bmatrix} x(t) \\ p(t) \end{bmatrix} = M \begin{bmatrix} x(t) \\ p(t) \end{bmatrix}$$

by virtue of (9), and the transversality condition becomes

**(31)**

$$p(t_i) = S x(t_i).$$

Let $\Phi(t, t_i)$ be the transition matrix for (30), i.e.,

**(32)**

$$\begin{bmatrix} x(t) \\ p(t) \end{bmatrix} = \begin{bmatrix} \Phi_{11}(t, t_i), \Phi_{12}(t, t_i) \\ \Phi_{21}(t, t_i), \Phi_{22}(t, t_i) \end{bmatrix} \begin{bmatrix} x(t_i) \\ p(t_i) \end{bmatrix}$$

where the $\Phi_{ij}(t, t_i)$ are $n \times n$ submatrices formed by partitioning $\Phi$. Then (29), (31) and (32) gives

$$P(t, S, t_i) = \left[ \Phi_{21}(t, t_i) + \Phi_{22}(t, t_i) S \right] \left[ \Phi_{11}(t, t_i) + \Phi_{12}(t, t_i) S \right]^{-1}.$$
(v) Negative Exponential Solution

This technique, reported first in [34], bypasses one difficulty arising in (iv), namely, that $\Phi(t_1, t_1)$ contains stable as well as unstable modes. As time proceeds, the unstable modes tend to dominate and accuracy is at stake.

The basic trick is to recast the equations so that only negative exponentials occur in computations.

Write

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}$$

for a $2n \times 2n$ matrix that diagonalizes $M,$

$$M = W \begin{bmatrix} J & 0 \\ 0 & -J \end{bmatrix} W^{-1},$$

and set

$$\begin{bmatrix} \xi(t) \\ \beta(t) \end{bmatrix} = \begin{bmatrix} x(t) \\ p(t) \end{bmatrix}. $$

Then

$$\begin{bmatrix} \dot{\xi}(t) \\ \dot{\beta}(t) \end{bmatrix} = \begin{bmatrix} e^{-j\omega t}, & 0 \\ 0, & e^{-j\omega t} \end{bmatrix} \begin{bmatrix} \xi(t) \\ \beta(t) \end{bmatrix}$$

and by (33),

$$\dot{\beta}(t) = R \dot{\xi}(t)$$

where

$$R = -[W_{12} - SW_{12}]^{-1} [W_{21} - SW_{11}].$$

Denoting

$$G(t, t_1) = e^{-j\omega t_1} Re^{-j\omega t},$$

we finally get by (29)

$$P(t, S, t_1) = \left[ W_{21} + W_{22} G(t, t_1) \right] \left[ W_{11} + W_{12} G(t, t_1) \right]^{-1}. $$

The steady-state solution $P_\infty$ is obtained as

$$P_\infty = \lim_{t \to \infty} P(t, S, t) = W_{21}W_{11}^{-1},$$

which is identical to equation (12).

(vi) Solution Via Numerical Integration

This is perhaps the most natural numerical technique to obtain the transient solution of (6). Usually the Runge-Kutta method is applied [18]. The method can also be applied to compute a desired steady-state solution as a function of $P(t_i) = S,$ see Theorem 16. However, for this purpose the numerical integration is time consuming and not very accurate.
### Table 1.

#### Nonnegative Solutions of Algebraic Equation (9)

<table>
<thead>
<tr>
<th>Condition</th>
<th>( \text{Re} \lambda \neq 0 ) for all ( \lambda )</th>
<th>( \text{Re} \lambda = 0 ) for some ( \lambda )</th>
</tr>
</thead>
<tbody>
<tr>
<td>((A, B)) stabilizable ( (C, A)) detectable</td>
<td>( P_s ) exists and it is the unique nonnegative solution, ( P_s = P_{s0} )</td>
<td></td>
</tr>
<tr>
<td>((A, B)) stabilizable ( (C, A)) not detectable</td>
<td>( P_s ) exists, there are other nonnegative solutions, ( P_s = P_a )</td>
<td>( P_s ) does not exist, there are other nonnegative solutions</td>
</tr>
<tr>
<td>((A, B)) not stabilizable ( (C, A)) detectable</td>
<td>No nonnegative solution exists including ( P_s ) and ( P_a )</td>
<td></td>
</tr>
<tr>
<td>((A, B)) not stabilizable ( (C, A)) not detectable</td>
<td>( P_s ) does not exist, other nonnegative solutions may or may not exist</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2.

#### Asymptotic Properties of Solutions to Riccati Equation (6)

<table>
<thead>
<tr>
<th>Condition</th>
<th>( t_T \to \infty )</th>
<th>( S = 0 )</th>
<th>( S \neq 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S = 0 )</td>
<td>( \text{Re} \lambda \neq 0 ) for all ( \lambda )</td>
<td>( \text{Re} \lambda = 0 ) for some ( \lambda )</td>
<td>( t_T ) finite</td>
</tr>
<tr>
<td>((A, B)) stabilizable ( (C, A)) detectable</td>
<td>( P_\infty ) exists, ( P_\infty = P_s )</td>
<td>( P_\infty = P_0 = P_s )</td>
<td>There is always a unique solution ( P(t) ) and it is nonnegative</td>
</tr>
<tr>
<td>((A, B)) stabilizable ( (C, A)) not detectable</td>
<td>( P_\infty ) exists, ( P_\infty = P_0 )</td>
<td>( P_\infty = P_s )</td>
<td>?</td>
</tr>
<tr>
<td>((A, B)) not stabilizable ( (C, A)) detectable</td>
<td>( P_\infty ) does not exist</td>
<td>( P_\infty ) does not exist</td>
<td></td>
</tr>
<tr>
<td>((A, B)) not stabilizable ( (C, A)) not detectable</td>
<td>( P_\infty ) may or may not exist</td>
<td>( P_\infty ) may or may not exist</td>
<td></td>
</tr>
</tbody>
</table>
To conclude this section we mention yet another group of methods to solve (9). The common feature is that equation (9) is transformed into a recurrent algebraic equation [14] whose solution coincides with the unique nonnegative solution of (9). The resulting matrix recurrent equation is then straightforward to solve.

CONCLUSIONS

The reader will have noticed that the interrelations of different nonnegative solutions of (9) are quite complicated. Letting $t \to \infty$ and studying the asymptotic behaviour of (6) makes the matters even worse. Therefore the following Tables 1 and 2 are provided to summarize the results for equations (9) and (6), respectively. We also stress that the aim of the paper was to study the matrix Riccati equation in general and not to restrict ourselves to common engineering applications. This approach in turn provides a deep insight into the problem of least squares control and the engineer is well prepared to cope with ill-behaved solutions.

(Received June 5, 1972.)

REFERENCES
