J. Austral. Math. Soc. Ser. B 26 (1985), 355–361

A CANONICAL FORM AND SOLUTION FOR THE MATRIX RICCATI DIFFERENTIAL EQUATION

R. B. LEIPNIK[1]

(Received 27 September 1982; revised 25 July 1983)

Abstract

A canonical form of the self-adjoint Matrix Riccati Differential Equation with constant coefficients is obtained in terms of extremal solutions of the self-adjoint Matrix Riccati Algebraic (steady-state) Equations. This form is exploited in order to obtain a convenient explicit solution of the transient problem. Estimates of the convergence rate to the steady state are derived.

Introduction

There is a long history of interest in quadratic differential equations in mechanics (Bolza [3]), control theory (Kalman [6]) and systems theory (Herman [5]). Reid [10] considered the special system

$$\dot{P} = -A - PB - DP - PCP \tag{1}$$

where $\dot{P} = dP/dt$, $P = P(t)$ is an $m \times n$ matrix with complex entries, and $A = A(t)$, $B = B(t)$, $D = D(t)$, $C = C(t)$ are respectively $m \times n$, $n \times n$, $m \times m$, $n \times m$ matrices with complex entries, which are locally Lebesgue integrable. Reid, (following Bernoulli) reduced (1) to the consideration of

$$\begin{bmatrix} 0 & -I_m \\ I_n & 0 \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix}^{\cdot} = \begin{bmatrix} A & D \\ B & C \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} \tag{2}$$

where $U = U(t)$ and $V = V(t)$ are respectively $n \times m$ and $m \times m$ matrices, obtaining many interesting results. If $m = n$, then the correspondence between (1)

[1] The Australian National University, Statistics Department, I.A.S., P. O. Box 4, Canberra, ACT 2600. Permanent address: Department of Applied Mathematics, University of California, Santa Barbara, California 93106, U.S.A.

and (2) is effected by the Bernoulli formula

$$P(t) = V(t)U^{-1}(t). \tag{3}$$

The self-adjoint case $D = B^*$, $A = A^*$, $C = C^*$ has many applications in the calculus of variations and hence in control theory. Also, the non-self-adjoint case appears in scattering theory (Reid [10], Redheffer [9]). In most engineering problems, self-adjointness (either real or Hermitian) is assumed, the coefficient matrices $A, B, C, D = B^*$ are taken as constant, and a self-adjoint solution $P = P^*$ is desired, starting from a self-adjoint initial matrix $P(0) = P^*(0)$. These problems can lead to numerical difficulties which we can try to circumvent. We shall obtain a canonical form (in terms of steady-state solutions) and an explicit solution convenient for high-speed computing and error estimation in medium to large systems ($n \geqslant 8$, say). Previously, Bellman [2] found an elegant solution for special A, B, C, and B. D. O. Anderson [1] found a general solution requiring somewhat more computing. Our approach exploits these ideas, as well as algebraic results of Coppel [4] and Kucera [7]. These results relate to the self-adjoint algebraic (steady-state) matrix Riccati equation

$$PB + B^*P + PCP = -A. \tag{4}$$

In Section 2, the canonical form for (1) is derived and expressed in terms of a pair of solutions of (4) and of bisymmetric matrices. In Section 3, functions of bisymmetric matrices are shown to be bisymmetric, permitting the (transient) solution of (1) to be expressed explicitly (in Section 4). Convergence rates of the transient solution to the steady-state are discussed in Section 5.

2. Canonical form

Consider now the equation

$$\dot{P} = -A - PB - B^*P - PCP \tag{5}$$

and make the transformation

$$P = Z^*P_1Z + H \quad \text{where } H = H^* = \text{constant}, Z = \text{constant}, \tag{6}$$

and note that P_1 satisfies

$$\dot{P}_1 = -B_1^*P_1 - P_1B_1 - P_1C_1P_1 - A_1 \tag{7}$$

where

$$\begin{aligned} B_1^* &= Z^{*-1}(B^* + HC)Z^*, \qquad C_1 = ZCZ^*, \\ A_1 &= Z^{*-1}(A + B^*H + HB + HCH)Z^{-1}. \end{aligned} \tag{8}$$

We wish to choose H, Z so that

$$B_1^* = -B_1, \qquad C_1 = -A_1 \tag{9}$$

https://doi.org/10.1017/S0334270000004550 Published online by Cambridge University Press

which yields our canonical form for (7). Condition (9) is clearly implied by the coupled equations

$$(Z^*Z)(B + CH) = -(B^* + HC)(Z^*Z), \tag{10}$$

$$(Z^*Z)C(Z^*Z) = -A - B^*H - HB - HCH, \tag{11}$$

The coupled system (10), (11) is uncoupled by the substitution

$$Z^*Z = (P^+ - P^-)/2, \qquad H = (P^+ + P^-)/2 \tag{12}$$

if P^+, P^- satisfy the equations

$$B^*P^+ + P^+B + P^+CP^+ = -A,$$
$$B^*P^- + P^-B + P^-CP^- = -A$$

and $(P^+)^* = P^+$, $(P^-)^* = P^-$ and if, in addition, the chosen P^+, P^- satisfy $P^+ > P^-$.

Motivation for the steps (6)–(12), the key novelty in the paper, can be supplied by noting that if P^+ is normalized to I and P^- to $-I$, then $Z^*Z = I$ and $H = 0$, so $P = Z^{-1}P_1Z$, $B_1 = ZBZ^{-1}$, $C_1 = ZCZ^{-1}$, $A_1 = ZAZ^{-1}$. The whole procedure is now merely a unitary transformation of the original problem valid when B is skew-Hermitian and C is already equal to $-A$. This useful comment was provided by a referee.

To see that (12) uncouples (10), (11), simply form

$$2(P^+CP^+ + P^-CP^-) = (P^+ + P^-)C(P^+ + P^-) + (P^+ - P^-)C(P^+ - P^-), \tag{13a}$$

$$2(P^+CP^+ - P^-CP^-) = (P^+ + P^-)C(P^+ - P^-) + (P^+ - P^-)C(P^+ + P^-). \tag{13b}$$

Addition and subtraction of (13a) and (13b) yields (10), (11). (In the special case, (13) becomes $2(2C) = 2C(2)$ and $0 = 0$).

Assume, from now on, that $C \leqslant 0$, (B, C) is controllable and the matrix

$$\begin{bmatrix} B & C \\ -A & -B^* \end{bmatrix}$$

has no pure imaginary eigenvalues. Then (4) has a unique, symmetric solution P^+ such that all eigenvalues of $B + CP^+$ have negative real part and a unique, symmetric solution P^- such that all eigenvalues of $B + CP^-$ have positive real part. Moreover P^+ and P^- are the maximal and minimal symmetric solutions of (27), and $P^+ > P^-$. (These results follow from Coppel [4, Corollaries of Theorems 6 and 2].)

https://doi.org/10.1017/S0334270000004550 Published online by Cambridge University Press

3. Bisymmetric matrices

We begin with the reduced equation

$$\dot{P}_1 = -B_1^* P_1 - P_1 B_1 - P_1 C_1 P_1 - A_1 \tag{14}$$

under the condition

$$B_1^* = -B_1, \qquad C_1 = -A_1 \tag{15}$$

obtained in Section 2.

If $V(t)$ is a nonsingular solution of the linear differential equation

$$\dot{V} = [B_1 + C_1 P_1(t)]V \tag{16}$$

then $V(t)$ and $U(t) = P_1(t)V(t)$ satisfy the extended equation

$$\begin{bmatrix} U \\ V \end{bmatrix}^{\cdot} = \begin{bmatrix} B_1 & C_1 \\ C_1 & B_1 \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} \tag{17}$$

Conversely, if $U(t)$, $V(t)$ satisfy (17) and $V(t)$ is nonsingular then $P_1(t) = U(t)V^{-1}(t)$ is a solution of (14).

The block matrix above is said to be in the bisymmetric form. The solution to (17) is

$$\begin{bmatrix} U \\ V \end{bmatrix} = \left(\exp \begin{bmatrix} B_1 & C_1 \\ C_1 & B_1 \end{bmatrix} t \right) \begin{bmatrix} U(0) \\ V(0) \end{bmatrix}. \tag{18}$$

To obtain an explicit formula for P, the partition of matrices of the form

$$\exp \begin{bmatrix} K & L \\ L & K \end{bmatrix}$$

is necessary. For any matrix power series

$$\theta(S) = \sum_n a_n S^n, \tag{19}$$

it is easy to show that if

$$S = \begin{bmatrix} K & L \\ L & K \end{bmatrix},$$

then

$$\theta(S) = \begin{bmatrix} X & Y \\ Y & X \end{bmatrix},$$

where

$$X = \tfrac{1}{2}(\theta(K + L) + \theta(K - L)), \; Y = \tfrac{1}{2}(\theta(K + L) - \theta(K - L)) \tag{20}$$

whether K, L commute or not.

https://doi.org/10.1017/S0334270000004550 Published online by Cambridge University Press

Let

$$K_n = \tfrac{1}{2}\big((K+L)^n + (K-L)^n\big) \text{ and } L_n = \tfrac{1}{2}\big((K+L)^n - (K-L)^n\big), \quad (21)$$

and note that (by an easy induction)

$$\begin{bmatrix} K & L \\ L & K \end{bmatrix}^n = \begin{bmatrix} K_n & L_n \\ L_n & K_n \end{bmatrix} \quad (22)$$

for all n, from which (20) follows by summation.

In particular, we conclude with the seemingly novel

$$\exp\begin{bmatrix} K & L \\ L & K \end{bmatrix} = \begin{bmatrix} X & Y \\ Y & X \end{bmatrix} \quad (23)$$

where

$$\left.\begin{aligned} X &= 1/2(\exp(K+L) + \exp(K-L)), \\ Y &= 1/2(\exp(K+L) - \exp(K-L)). \end{aligned}\right\} \quad (24)$$

4. Solution of the differential equation

Consider a solution $P(t)$ of (5) which is defined either for all $t \geqslant 0$, or all $t \leqslant 0$, or both.

Let $F = P(0)$, $P_1(0) = F_1 = Z^{*-1}(F-H)Z^{-1}$ and take $U(0) = F_1$, $V(0) = I$, where F_1 may be singular and

$$S = U(0) + V(0), \quad D = U(0) - V(0). \quad (25)$$

Then by Sections 2 and 3 we have $P_1 = UV^{-1}$, where

$$\left.\begin{aligned} U(t) &= \tfrac{1}{2}\big([\exp(B_1+C_1)t]S + [\exp(B_1-C_1)t]D\big), \\ V(t) &= \tfrac{1}{2}\big([\exp(B_1+C_1)t]S - [\exp(B_1-C_1)t]D\big). \end{aligned}\right\} \quad (26)$$

Also, since $B_1^* = B_1$, $C_1^* = C_1$, and $S = S^*$, $D = D^*$, we have

$$\left.\begin{aligned} U^* &= \tfrac{1}{2}\big(S[\exp(-B_1+C_1)t] + D[\exp(-B_1-C_1)t]\big), \\ V^* &= \tfrac{1}{2}\big(S[\exp(-B_1+C_1)t] - D[\exp(-B_1-C_1)t]\big). \end{aligned}\right\} \quad (27)$$

Now $B_1 + C_1$ commutes with $-B_1 - C_1$ and $B_1 - C_1$ commutes with $-B_1 + C_1$, trivially, whether B_1 and C_1 commute or not. Hence,

$$[\exp(B_1+C_1)t][\exp(-B_1-C_1)t] = I = [\exp(B_1-C_1)t][\exp(-B_1+C_1)t]$$

and thus by direct calculation from (26) and (27), $U^*V - V^*U = 0$. It follows that

$$P_1^* = (UV^{-1})^* = V^{-1*}U^* = V^{-1*}U^*VV^{-1} = V^{-1*}(V^*U)V^{-1} = P_1, \quad (28)$$

as desired. If $P(t_0)$ is defined, and $P(0)$ is not, time-shifts can be made.

Clearly $P(t) = Z^*P_1(t)Z + H$. But it is $Z^*Z = (P^+ - P^-)/2$ which is defined, not Z^* or Z separately. Since $(P^+ - P^-)/2$ is Hermitian and positive definite, it

https://doi.org/10.1017/S0334270000004550 Published online by Cambridge University Press

has a unique Hermitian, positive square root $Z_0 = Z_0^*$, obtained by direct or iterative procedures, which may be preferred to other values of Z.

Having Z and H permits the calculation of F_1, S, and D from (25). Also, B_1 and C_1 are obtainable from Z, H, and (8). This permits the determination of $U(t)$ and $V(t)$ from (26) and so $P_1(t) = UV^{-1}$ and finally $P(t)$ from (6). Clearly $P_1(0) = F_1 I^{-1} = F_1$ and $P(0) = F$ as desired.

A word on the calculation of the exponentials $\exp[(B_1 \pm C_1)t]$ which appear in (26). First, $B_1 + C_1 = Z(B + CP^+)Z^{-1}$ is stable (all its eigenvalues have negative real part) since $B + CP^+$ is stable, as mentioned following (13). Likewise, $C_1 - B_1 = (B_1 + C_1)^*$ is stable. It follows that as $t \to \pm\infty$ the exponentials tend to 0, ∞ and so $U(t)$ and $V(t)$ tend separately to ∞. Taking out a common factor from $U(t)$ and $V(t)$ shows that, nevertheless, $P_1(t)$ tends to a limit as $t \to \pm\infty$.

The numerical stability of the exponentials is an extremely important and difficult practical issue. A review paper devoted to e^{Mt} where M is of order < 200 (typical of control applications but not of large structure applications) is Moler and Van Loan [15]. Nineteen methods are considered, falling into five classes, in this long paper. Generally, when M is symmetric (or normal), the difficulties are very much reduced. Unfortunately, control system matrices are typically non-normal.

Squaring Method 3, in which $e^{At} = (e^{A/m})^{mt}$ is used for $m = 2^\nu$, where $\|A\| \leqslant m$, is well recommended. Of course, e^{Mt} can be obtained by integration of $\dot{x} = Mx$. This partially defeats the object of the present paper, but apart from that, the integration methods appear 10 to 20 times less efficient than squaring. Method 18, based on triangular block diagonalization, appears comparable to Method 3 in speed and accuracy. Trotter's formula

$$e^{Mt} = \lim_{m \to \infty} \left(e^{M_1/m} e^{M_2/m}\right)^{mt}$$

is the basis for the splitting method 19, where $M = M_1 + M_2$, and $m = 2^\nu$. The special choices $M_1 = (M + M^*)/2$ and $M_2 = (M - M^*)/2$ are particularly convenient, where $e^{M_1/m}$ and $e^{M_2/m}$ may themselves be computed by Method 3. A more satisfactory basis for the splitting method may be the Hausdorff expansion formula for $e^{(M_1+M_2)t}$ or as recently suggested by Kenney (unpublished), the use of Richardson extrapolation. However, since Methods 18 and 19 are conceptually more complicated and generally no faster than the squaring method, the latter appears most useful.

5. Convergence rate

Suppose the eigenvalues of $B + CP^+$ (or of $B_1 + C_1$) have real parts whose maximum $< r_1 < 0$ (discussed by Coppel [4]). Then as $t \to \infty$, $e^{(B_1+C_1)t} = O(e^{r_1 t})$.

https://doi.org/10.1017/S0334270000004550 Published online by Cambridge University Press

If $D = F_1 - I$ is nonsingular, then by (26),

$$P_1(t) = U(t)V^{-1}(t) = \big(D + O(e^{2r_1 t})\big)\big(-D + O(e^{2r_1 t})\big)^{-1} = -I + O(e^{2r_1 t}) \quad \text{as } t \to \infty.$$

Similarly, if $S = F_1 + I$ is nonsingular, then

$$P_1(t) = I + O(e^{-2r_1 t}) \quad \text{as } t \to -\infty.$$

Returning to the original matrix variable $P(t)$, it follows that

$$P(t) = P^- + O(e^{2r_1 t}),\ t \to \infty, \qquad \text{if } D^{-1} \text{ exists},$$

$$P(t) = P^+ + O(e^{-2r_1 t}),\ t \to -\infty, \quad \text{if } S^{-1} \text{ exists},$$

where $r_1 < 0$, and $P^+ > P^-$ are the extremal solutions of the algebraic matrix Riccati equation (4). Any less satisfactory result is due to numerical integration or numerical matrix-exponentiation errors in truncation, round-off, or algorithmic inadequacies.

The preceding results are reminiscent of Coppel's reduction of the system matrix to the form $L\tilde{M}L^{-1}$, where L is symplectic, and

$$\tilde{M} = \begin{bmatrix} \tilde{M}_{11} & 0 \\ 0 & -\tilde{M}_{11} \end{bmatrix}.$$

They differ in being derived through quadratic identities (numerically manageable) rather than through the characteristic function of the system matrix (numerically highly volatile) and so being more adaptable to conventional computing, the original motivation for the present work.

References

[1] B. D. O. Anderson and J. B. Moore, *Linear optimal control* (Prentice-Hall, Englewood Cliffs, New Jersey, 1971).

[2] R. Bellman, "Upper and lower bounds for the solutions of the matrix Riccati equation", *J. Math. Anal. Appl.* 17 (1967), 373–79.

[3] O. Bolza, *Vorlesungen über Variations-rechnung* (Teubner, Leipzig, 1909).

[4] W. A. Coppel, "Matrix quadratic equations", *Bull. Austral. Math. Soc.* 10 (1974), 377–401.

[5] R. Hermann, *Cartanian geometry, non linear waves, and control theory. Parts A and B, Interdisciplinary Math.* 20, 21 (Math. Sci. Press, Brookline, Mass., 1979).

[6] R. Kalman, "Contributions to the theory of optimal control", *Bol. Soc. Mat. Mexicana* 5 (1960), 102–119.

[7] V. Kucera, "A contribution to matrix quadratic equations", *IEEE Trans. Automat. Control* 17 (1972), 344–347.

[8] C. Moler and C. Van Loan, "Nineteen dubious ways to compute the exponential of a matrix", *SIAM Rev.* 20 (1978), 801–836.

[9] R. M. Redheffer, "Inequalities for a matrix Riccati equation", *J. Math. Mech.* 8 (1959), 349–377.

[10] W. T. Reid, *Riccati differential equations, Mathematics in Science and Engineering* 86 (Academic Press, New York, 1972).