Journal of Computational and Applied Mathematics 41 (1992) 377-383
North-Holland

CAM 1270

Letter Section

Analytical approximate solutions and error bounds for nonsymmetric Riccati matrix differential equations *

Lucas Jódar
Departamento de Matemática Aplicada, Universidad Politécnica de Valencia, Spain

Received 5 December 1991

Abstract

Jódar, L., Analytical approximate solutions and error bounds for nonsymmetric Riccati matrix differential equations, Journal of Computational and Applied Mathematics 41 (1992) 377-383.

In this paper we construct analytical approximate solutions for the nonsymmetric Riccati matrix differential equation. Given an admissible error $\epsilon > 0$, we determine an interval where we construct an approximate solution whose error is smaller than ϵ for all the points of the interval. The approximate solution is constructed in terms of matrices related to the data.

Keywords: Nonsymmetric Riccati equation; approximate solution; error bounds.

1. Introduction

In this paper we consider the nonsymmetric Riccati matrix differential equation

$$W'(t) = C - DW(t) - W(t)A - W(t)BW(t), \quad W(0) = W_0, \tag{1}$$

where $W(t)$, C, D, B and W_0 are $n \times n$ complex matrices, elements of $\mathbb{C}_{n\times n}$. Such equations appear in the invariant imbedding context [9], and for the symmetric case, where A is the adjoint matrix of D, equation (1) is important in the optimal control theory [1]. Most general

Correspondence to: Prof. L. Jódar, Departamento de Matemática Aplicada, Universidad Politécnica de Valencia, P.O. Box 22.012, Valencia, Spain.
* This work has been partially supported by the D.G.I.C.Y.T., grant PS90-0140.

methods to solve (1) are based on the transformation of (1) into the extended first-order linear system

$$\begin{bmatrix} U'(t) \\ V'(t) \end{bmatrix} = S \begin{bmatrix} U(t) \\ V(t) \end{bmatrix}, \quad \begin{bmatrix} U(0) \\ W(0) \end{bmatrix} = \begin{bmatrix} I \\ W_0 \end{bmatrix}, \quad S = \begin{bmatrix} A & B \\ C & -D \end{bmatrix}; \tag{2}$$

then solve the reduced equation (2) by numerical integration techniques, see [6] for details. For the case where B is invertible, a closed-form solution of (1) has been recently given in [5], however, for the general case such an expression is unknown in the literature. This motivates the search of analytical approximate solutions of the problem whose error is smaller than a prefixed admissible error ϵ.

If P is a matrix in $\mathbb{C}_{n \times m}$, we denote by $\| P \|$ its 2-norm, defined in [7, p.21]. We denote by I_m the identity matrix in $\mathbb{C}_{m \times m}$.

2. Analytical approximate solutions and error bounds

We begin this section with a lemma that guarantees an interval where the solution of (1) is defined.

Lemma 1. *Let us consider* (1) *and let* δ *be the positive constant defined by*

$$\delta = \| S \|^{-1} \ln\left(1 + \left\| \begin{bmatrix} I \\ W_0 \end{bmatrix} \right\|^{-1}\right), \tag{3}$$

where S *is given by* (2). *Then the solution of* (1) *in the interval* $|t| < \delta$ *is defined by*

$$W(t) = \left\{ [0, I] \exp(tS) \begin{bmatrix} I \\ W_0 \end{bmatrix} \right\} \left\{ [I, 0] \exp(tS) \begin{bmatrix} I \\ W_0 \end{bmatrix} \right\}^{-1}. \tag{4}$$

Proof. From [8, p.28], the solution of (1) is defined by $W(t) = V(t)(U(t))^{-1}$, where $U(t)$, $V(t)$ are the block components of system (2), and this solution is defined in the interval where $U(t) = [I, 0] \exp(tS)[\begin{smallmatrix} I \\ W_0 \end{smallmatrix}]$ is invertible. Note that for $t = 0$, $U(0) = I_n$. From the Perturbation Lemma [7, p.32], $U(t)$ is invertible at t if $\| U(t) - I_n \| < 1$. Note that

$$U(t) - U(0) = [I, 0] \exp(tS) \begin{bmatrix} I \\ W_0 \end{bmatrix} - I = [I, 0] \left\{ \sum_{k \geqslant 1} \frac{(St)^k}{k!} \right\} \begin{bmatrix} I \\ W_0 \end{bmatrix}$$

$$= [I, 0] (\exp(tS) - I_{2n}) \begin{bmatrix} I \\ W_0 \end{bmatrix}. \tag{5}$$

Taking norms in (5), it follows that

$$\| U(t) - U(0) \| \leqslant \left\| \begin{bmatrix} I \\ W_0 \end{bmatrix} \right\| \{ \exp(\| tS \|) - 1 \}.$$

Thus $\| U(t) - U(0) \| < 1$ if $(\exp(\| tS \|) - 1) \| [\begin{smallmatrix} I \\ W_0 \end{smallmatrix}] \| < 1$. Taking logarithms, this last inequality is equivalent to the condition

$$|t|\,\|S\| < \ln\left(1 + \left\|\begin{bmatrix} I \\ W_0 \end{bmatrix}\right\|^{-1}\right).$$

Hence the result is established. □

For the sake of clarity in the presentation of the next results we recall some properties related to the differentiation in Banach spaces whose proofs may be found in [3,4]. Let E and F be Banach spaces, let A be an open set in E and let f be a function from A into F. If $x \in A$ and f is Fréchet differentiable at x, we denote by $\mathrm{D}f(x)$ the Fréchet differential of f at x. It is well known that if f is a linear function, then, for any point $x \in A$, its Fréchet differential at x coincides with f, i.e., $\mathrm{D}f(x) = f$. If E, F and G are Banach spaces and g is a bilinear function from $E \times F$ into G, then for any point $(x, y) \in E \times F$, g is Fréchet differentiable at (x, y) and $\mathrm{D}g(x, y)$ is the linear function $(s, t) \to g(x, t) + g(s, y)$, see [3, Theorem 8.1.4].

If $f: E \to F$ is Fréchet differentiable at x and $\mathrm{D}f(x)$ is the linear mapping from E into F, we denote by $\|\mathrm{D}f(x)\| = \sup\{\|\mathrm{D}f(x)(z)\|;\ z \in E,\ \|z\| \leqslant 1\}$.

Lemma 2. *Let A, B, C and D be matrices in $\mathbb{C}_{n \times n}$, let F be the Banach space of all matrices in $\mathbb{C}_{n \times n}$ endowed with the 2-norm, and let $f: \mathbb{R} \times F \times \mathbb{R} \to F$ be the function defined by*

$$f(t, X, \lambda) = C - DX - XA - X(B + \lambda I)X. \tag{6}$$

Then f is differentiable at any point (t, X, λ), admits partial Fréchet differentials $\mathrm{D}_2 f(t, X, \lambda)$ and $\mathrm{D}_3 f(t, X, \lambda)$ and

$$\begin{aligned} &\|\mathrm{D}_2 f(t, X, \lambda)\| \leqslant \|A\| + \|D\| + 2\|BX\| + 2|\lambda|\,\|X\|, \\ &\|\mathrm{D}_3 f(t, X, \lambda)\| = \|X^2\|. \end{aligned} \tag{7}$$

Proof. Note that $f(t, X, \lambda)$ defined by (6) may be written in the form

$$f(t, X, \lambda) = C - DX - XA - XBX - \lambda X^2. \tag{8}$$

Now the function $g: F \to F$, defined by $g(X) = XBX$, is Fréchet differentiable at $X \in F$, and $\mathrm{D}g(X)$ is the linear mapping from F into F, defined by $\mathrm{D}g(X)T = XBT + TBX$, for any matrix T in F. In fact, note that

$$\frac{g(X+T) - g(X) - (XBT + TBX)}{\|T\|} = \frac{TBT}{\|T\|}. \tag{9}$$

Taking norms in (9), it is clear that the right-hand side of (9) tends to zero as $\|T\|$ tends to zero.

On the other hand, note that from (8) it is clear that $\mathrm{D}_3 f(t, X, \lambda)$ is the linear mapping which takes the value $-\lambda X^2$ at λ. Hence from the previous comments the proof of (7) is established. □

Now let us consider (1) with B singular, which in terms of the function f defined by (6) may be written in the form

$$W'(t) = f(t, W(t), 0), \quad W(0) = W_0. \tag{10}$$

From [8, p.28], if we consider the linear system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} U(t, \lambda) \\ V(t, \lambda) \end{bmatrix} = S(\lambda)\begin{bmatrix} U(t, \lambda) \\ V(t, \lambda) \end{bmatrix}, \quad \begin{bmatrix} U(0, \lambda) \\ V(0, \lambda) \end{bmatrix} = \begin{bmatrix} I \\ W_0 \end{bmatrix}, \; S(\lambda) = \begin{bmatrix} A & B+\lambda I \\ C & -D \end{bmatrix}, \tag{11}$$

then the solution of the perturbed problem (P_λ),

$$W'(t) = f(t, W(t), \lambda), \quad W(0) = W_0, \tag{12}$$

is given by $W(t) = V(t, \lambda)[U(t, \lambda)]^{-1}$, in the interval $(-c_0, c_0)$ where $U(t, \lambda)$ is invertible. Let h be a positive number such that

$$h \leqslant \inf\{\| S(\lambda) \|;\ \lambda \in (0, r(B))\}, \tag{13}$$

where

$$r(B) = \begin{cases} 1, & \text{if } \sigma(B) = \{0\}, \\ \min\{|z|;\ z \in \sigma(B),\ z \neq 0\}, & \text{if } \sigma(B) \neq \{0\}, \end{cases} \tag{14}$$

and $\sigma(B)$ denotes the set of all eigenvalues of B. From Lemma 1, the matrix $U(t, \lambda)$ is invertible for

$$|t| < h^{-1} \ln\left(1 + \left\| \begin{bmatrix} I \\ W_0 \end{bmatrix} \right\|^{-1}\right) = c, \quad 0 < \lambda < r(B), \tag{15}$$

and

$$B(\lambda) = B + \lambda I \quad \text{is invertible for } 0 < \lambda < r(B).$$

Let $\rho > 0$ and let H be the open ball in $\mathbb{C}_{n\times n}$ with centre at W_0 and radius ρ, $H = \{X \in \mathbb{C}_{n\times n},\ \| X - W_0 \| < \rho\}$. Let $\Omega = (0, r(B))$, $J = (-c, c)$, and note that from Lemma 2 it follows that

$$\begin{aligned} &\sup\{\|D_2 f(t, X, \lambda)\|;\ (t, X, \lambda) \in J \times H \times \Omega\} \\ &\quad \leqslant \| D \| + \| A \| + 2(\| B \| + r(B))(\| W_0 \| + \rho), \end{aligned} \tag{16}$$

$$\sup\{\|D_3 f(t, X, \lambda)\|;\ (t, X, \lambda) \in J \times H \times \Omega\} \leqslant (\| W_0 \| + \rho)^2.$$

Now if we denote by $W(t, \lambda)$ the solution of the problem (P_λ) defined by (12), and $W(t)$ is the solution of (1), from [3, Theorem 10.7.3] it follows that

$$\| W(t, \lambda) - W(t) \| \leqslant \gamma\lambda, \quad \text{for } |t| < c,\ 0 < \lambda < r(B), \tag{17}$$

where

$$\gamma = \frac{\beta(\exp(c\alpha) - 1)}{\alpha}, \qquad \alpha = \| D \| + \| A \| + 2(\| B \| + r(B))(\| W_0 \| + \rho),$$
$$\beta = (\| W_0 \| + \rho)^2. \tag{18}$$

Hence given an admissible error $\epsilon > 0$, if we take any value of λ such that

$$0 < \lambda < \min\left(r(B), \frac{\epsilon}{\gamma}\right), \tag{19}$$

where γ is determined by (18), then $W(t, \lambda)$, for $t \in (-c, c)$, is an approximate solution of (1), whose error is smaller than ϵ uniformly for $t \in (-c, c)$. Hence from [5, Corollary 1], the following result has been established.

Theorem 3. *Let us consider* (1) *where* B *is a singular matrix in* $\mathbb{C}_{n\times n}$ *and let* $\epsilon > 0$. *If* $\rho > 0$, c *is defined by* (15), $r(B)$ *is given by* (14) *and* α, β *and* γ *are defined by* (18), *taking* $B(\lambda) = B + \lambda I$, $B(\lambda)^{1/2}$ *a square root of* $B(\lambda)$, $0 < \lambda < r(B)$, *let us consider the matrix*

$$Z = \begin{bmatrix} 0 & I \\ B(\lambda)^{1/2}\left(C + DB(\lambda)^{-1}A\right)B(\lambda)^{1/2} & B(\lambda)^{1/2}\left(DB(\lambda)^{-1} - B(\lambda)^{-1}A\right)B(\lambda)^{-1/2} \end{bmatrix}, \tag{20}$$

let $J = \mathrm{Diag}(J_1, \ldots, J_k)$ *be the Jordan canonical form of* Z, *with* $J_j \in \mathbb{C}_{m_j\times m_j}$, *for* $1 \leqslant j \leqslant k$, *and let* $M = (M_{ij})$ *be an invertible matrix in* $\mathbb{C}_{2n\times 2n}$, *with* $M_{ij} \in \mathbb{C}_{n\times m_j}$, *for* $1 \leqslant i \leqslant 2$, $1 \leqslant j \leqslant k$, *such that* $MJ = ZM$, *then*

$$W(t, \lambda) = B(\lambda)^{-1/2}\left\{\sum_{s=1}^{k} M_{2s}\exp(tJ_s)D_s\right\} \times\left\{\sum_{s=1}^{k} M_{1s}\exp(tJ_s)D_s\right\}^{-1} B(\lambda)^{-1/2} - B(\lambda)^{-1}A, \tag{21}$$

where $D_s \in \mathbb{C}_{m_s\times n}$, *for* $1 \leqslant s \leqslant k$, *are determined by*

$$\begin{bmatrix} D_1 \\ \vdots \\ D_k \end{bmatrix} = M^{-1}\begin{bmatrix} B(\lambda)^{-1/2} \\ B(\lambda)^{1/2}W_0 + B(\lambda)^{-1/2}A \end{bmatrix}, \tag{22}$$

is an approximate solution of (1), *whose error is uniformly upper bounded by* ϵ *for all* $t \in (-c, c)$.

Now we illustrate the availability of the previous result with an example in $\mathbb{C}_{2\times 2}$.

Example 4. Let us consider the nonsymmetric Riccati equation of type (1) where

$$A = D = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad W_0 = 0 \in \mathbb{C}_{2\times 2}.$$

In this case B is singular and $\sigma(B) = \{0\}$. From (14) we have $r(B) = 1$ and straightforward computations show that $\| A \| = \| D \| = (\frac{1}{2})^{1/2}(3 + 5^{1/2})^{1/2}$; and if we denote by $S(\lambda)$ the matrix defined in (11), then in this case

$$\| S(\lambda) \| = \left(\tfrac{1}{2}(3 + 5^{1/2})\right)^{1/2} + 2 + \left(1 + (1 + \lambda^2)^{1/2} + \lambda^2\right)^{1/2}$$
$$\geqslant h = \| S(0) \| = 3.455\,346\,69.$$

Taking into account that in this case $W_0 = 0$, the constant c defined by (15) takes the value

$$c = h^{-1}\ln(2) = 0.200\,601\,34.$$

Taking $\rho = 1$, from (18) we have $\alpha = 9.236\,067\,977$, $\beta = 1$, $\gamma = 0.582\,222\,373$. Given $\epsilon > 0$, from (19), taking a positive value of λ such that $\lambda < \min(1, 1.717\,556\,807\,\epsilon)$, the solution of problem

(P_λ) is an approximate solution in $(-0.200\,601\,34, 0.200\,601\,34)$ whose error is smaller than ϵ uniformly in this interval. In this example we have

$$Z=\begin{bmatrix} 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 1-\lambda & -4 & 0 & 0\\ 0 & 1-\lambda & 0 & 0\end{bmatrix},\qquad J_1=\begin{bmatrix}(1-\lambda)^{1/2} & 1\\ 0 & (1-\lambda)^{1/2}\end{bmatrix},$$

$$J_2=\begin{bmatrix}-(1-\lambda)^{1/2} & 1\\ 0 & -(1-\lambda)^{1/2}\end{bmatrix},$$

$$M=\begin{bmatrix}(1-\lambda)^{1/2} & -2 & -(1-\lambda)^{1/2} & -2\\ (1-\lambda)^{1/2} & 0 & -(1-\lambda)^{1/2} & 0\\ -(1+\lambda) & -2(1-\lambda)^{1/2} & -(1+\lambda) & 2(1-\lambda)^{1/2}\\ 1-\lambda & 0 & 1-\lambda & 0\end{bmatrix},$$

$$M^{-1}=\begin{bmatrix}0 & \frac{1}{2}(1-\lambda)^{-1/2} & 0 & \frac{1}{2}(1-\lambda)^{-1/2}\\ -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4}(1-\lambda)^{-1/2} & -\frac{1}{4}(1+\lambda)(1-\lambda)^{-1/2}\\ 0 & -\frac{1}{2}(1-\lambda)^{-1/2} & 0 & \frac{1}{2}(1-\lambda)^{-1/2}\\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4}(1-\lambda)^{-1/2} & \frac{1}{4}(1+\lambda)(1-\lambda)^{-3/2}\end{bmatrix},$$

$$\exp(tJ_1)=\exp\left(t(1-\lambda)^{1/2}\right)\begin{bmatrix}1 & t\\ 0 & 1\end{bmatrix},\qquad \exp(tJ_2)=\exp\left(-t(1-\lambda)^{1/2}\right)\begin{bmatrix}1 & t\\ 0 & 1\end{bmatrix},$$

$$B(\lambda)^{-1}=\begin{bmatrix}\lambda^{-1} & -2\lambda^{-2}\\ 0 & \lambda^{-1}\end{bmatrix},\qquad B(\lambda)^{1/2}=\begin{bmatrix}\lambda^{1/2} & \lambda^{-1/2}\\ 0 & \lambda^{1/2}\end{bmatrix},$$

$$B(\lambda)^{-1/2}=\begin{bmatrix}\lambda^{-1/2} & -\lambda^{-3/2}\\ 0 & \lambda^{-1/2}\end{bmatrix},$$

$$M_{11}=\begin{bmatrix}(1-\lambda)^{1/2} & -2\\ (1-\lambda)^{1/2} & 0\end{bmatrix},\qquad M_{12}=\begin{bmatrix}-(1-\lambda)^{1/2} & -2\\ -(1-\lambda)^{1/2} & 0\end{bmatrix},$$

$$M_{21}=\begin{bmatrix}-(1+\lambda) & -2(1-\lambda)^{1/2}\\ 1-\lambda & 0\end{bmatrix},\qquad M_{22}=\begin{bmatrix}-(1+\lambda) & 2(1-\lambda)^{1/2}\\ 1-\lambda & 0\end{bmatrix},$$

The matrices D_1 and D_2 defined by (22) take the form

$$D_1=\begin{bmatrix}0 & \dfrac{1+(1-\lambda)^{1/2}}{2(1-\lambda)\lambda^{1/2}}\\ -\dfrac{1+(1-\lambda)^{1/2}}{4(\lambda-\lambda^2)^{1/2}} & \dfrac{(1+\lambda)\left((1-\lambda)^{3/2}+1-2\lambda\right)}{4(\lambda-\lambda^2)^{3/2}}\end{bmatrix},$$

$$D_2=\begin{bmatrix}0 & \dfrac{1-(1-\lambda)}{2(1-\lambda)\lambda^{1/2}}\\ \dfrac{1-(1-\lambda)^{1/2}}{4(\lambda-\lambda^2)^{1/2}} & \dfrac{(1+\lambda)\left((1-\lambda)^{3/2}-1+2\lambda\right)}{4(\lambda-\lambda^2)^{3/2}}\end{bmatrix}.$$

Let us take $\epsilon = 10^{-2}$; then from the previous computations, taking $\lambda = 0.017\,175\,568\,06$ in the above expressions and putting the resulting expressions into the corresponding expression (21), one gets the required approximate solution.

References

[1] M. Athans and P. Falb, *Optimal Control* (McGraw-Hill, New York, 1966).

[2] R.S. Bucy and P.D. Joseph, *Filtering for Stochastic Processes with Applications to Guidance* (Interscience, New York, 1968).

[3] J. Dieudonné, *Foundations of Modern Analysis* (Academic Press, New York, 1960).

[4] T.M. Flett, *Differential Analysis* (Cambridge Univ. Press, Cambridge, 1980).

[5] L. Jódar and E. Navarro, A closed form solution for nonsymmetric Riccati differential equations with invertible quadratic coefficient, *Appl. Math. Lett.* **4** (2) (1991) 17–20.

[6] C.R. Kenney and R.B. Leipnik, Numerical integration of the differential matrix Riccati equation, *IEEE Trans. Automat. Control* **AC-30** (1985) 962–970.

[7] J.M. Ortega, *Numerical Analysis, A Second Course* (Academic Press, New York, 1972).

[8] W.T. Reid, *Riccati Differential Equations* (Academic Press, New York, 1972).

[9] M.R. Scott, *Invariant Imbedding and its Applications to Ordinary Differential Equations* (Addison-Wesley, Reading, MA, 1973).