UNIVERSITY OF CAMBRIDGE

Numerical Analysis Reports

Approximate solutions with a priori error bounds for continuous-coefficient matrix Riccati equations

S. Blanes, L. Jódar and E. Ponsoda

DAMTP 1999/NA10
August, 1999

Department of Applied Mathematics and Theoretical Physics
Silver Street
Cambridge
England CB3 9EW

Approximate solutions with a priori error bounds for continuous-coefficient matrix Riccati equations

S. Blanes[1,2], L. Jódar[3] and E. Ponsoda[3]

[1] *Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, England* (S.Blanes@damtp.cam.ac.uk).

[2] *Departament de Matemàtiques. Universitat Jaume I, 12.071 Castellón, Spain.*

[3] *Departamento de Matemática Aplicada. Universidad Politécnica de Valencia, P.O. Box 22.012 Valencia, Spain.*

In this paper the exact solution of the non-symmetric matrix Riccati equation with continuous coefficients is approximated using Fer's approximations of the associated underlying linear system. Given an admissible error $\epsilon > 0$, the order n of Fer's truncation is determined so that in the previously guaranteed existence interval, the error of the approximated solution is less than ϵ. Some qualitative properties of the Fer's approximations are given and illustratives examples are included.

1. Introduction

In this paper we consider rectangular non-symmetric Riccati matrix differential equations of the type

$$W'(t) = C(t) - D(t)W(t) - W(t)A(t) - W(t)B(t)W(t)\,; \quad W(0) = W_0\,, \qquad (1)$$

where the unknown $W(t)$ lies in $C^{p\times q}$, and coefficients $C(t) \in C^{p\times q}$, $D(t) \in C^{p\times p}$, $A(t) \in C^{q\times q}$, $B(t) \in C^{q\times p}$ are continuous matrix valued functions. If $A^T(t)$ denotes the transpose matrix of $A(t)$, and $p = q$, then equation (1) is said to be symmetric when the coefficients in (1) are real matrices and $D(t) = A(t)^T$.

Initial value problems for Riccati differential systems arise frequently in important applications to classical control theory Casti (1977), Kailath (1980), Kratz (1995), Reid (1972), in decoupling techniques for both the analysis and numerical study of boundary value problems Ascher *et al* (1988), Keller & Lentini (1982), and in non-cooperative control theory appearing in economic or military problems Abou-Kandil (1986), Baghi (1984), Clemhout & Wan Jr. (1979), Mariton (1990) and the references therein. The case $p \neq q$ appears, for instance, in differential games when Nash strategies in non-cooperative control problems are tackled Cruz & Chen (1971), Starr & Ho (1969).

The Riccati equation (1) has been extensively studied in the autonomous case, but many real systems are non-autonomous, see Abou-Kandil (1986), Baghi (1984), Clemhout & Wan Jr. (1979), Mariton (1990), and this case has received little numerical treatment in the literature as far as the study of accuracy and a priori error bounds in terms of the data for the proposed numerical solutions are concerned. Some exceptions can be found in Dieci (1992), Jódar & Abou-Kandil (1988), Jódar & Cortés (1998), Jódar & Ponsoda (1995), Kenney & Leipnik (1985), Kunkel & Mehrmann (1990), Oshman & Bar-Itzhack (1984) and Schiff & Schnider (1996).

In Kenney & Leipnik (1985) the direct integration of (1) using well known routines and a method based on the modified Davison-Maki algorithm is proposed, but the numerical solutions are tested on a set of concrete examples without error bound information in terms of the data for the general case. In Oshman & Bar-Itzhack (1984) the solution of a symmetric Riccati equation is reconstructed itself starting fronm the previous determination of the eigenvalues and eigenvectors of the solution matrix. In Kunkel & Mehrmann (1990) an implicit Riccati matrix differential equation is considered, and a numerical method based on continuous singular value decompositions of matrix functions and Kronecker canonical forms of matrix pencils is proposed.

The study of the Riccati equation (1) is closely related to the underlying linear system

$$\left.\begin{array}{l} Y'(t) \;=\; S(t)\,Y(t)\,, \quad Y(0) \;=\; \begin{bmatrix} I_q \\ W_0 \end{bmatrix}, \\[2ex] Y(t) = \begin{bmatrix} U(t) \\ V(t) \end{bmatrix}, \quad S(t) = \begin{bmatrix} A(t) & B(t) \\ C(t) & -D(t) \end{bmatrix}, \end{array}\right\} \tag{2}$$

where I_q denotes the identity matrix of $C^{q\times q}$. Specifying $S(t)$ uniquely determines the related Riccati equation, but the converse is not true, see Schiff & Schnider (1996). In fact, equation (1) is unaffected by the substitution

$$\left.\begin{array}{l} A(t) \;\longrightarrow\; A(t) \,+\, f(t)\,I_q\,, \\[2ex] D(t) \;\longrightarrow\; D(t) \,+\, f(t)\,I_p\,, \end{array}\right\} \tag{3}$$

where $f(t)$ is a scalar function. This can be written in the equivalent form

$$S(t) \;\longrightarrow\; S(t) \,+\, f(t)\,I_{q+p}\,. \tag{4}$$

The method proposed in Jódar & Abou-Kandil (1988) is based on the vectorization of problem (2) and uses the successive approximation method. In Dieci (1992) an interesting numerical integration method is proposed and error bounds are given, but these are expressed in terms of the geometrical properties and involve dichotomy constants which are not known a priori in practice. In Schiff & Schnider (1996) a class of explicit one-step matrix methods for the numerical solution of problem (1) are given. This method exploits the matrix group structure underlying the equation, is simple and is computationally cheap to implement. In Choi & Laub (1990), Jódar & Ponsoda (1995) approximate solutions are constructed using matrix difference schemes. Finally, in Jódar & Cortés (1998), the exact solution of (1) is approximated by a rational matrix function with a pre-determined accuracy, for the case when the coefficients in (1)) are analytic functions.

Note that if $S(t)$ given by (2) is continuous in $[0,b]$, the equation (2) has a unique solution that, in general, is not exactly computable in a closed form, Flett (1980, p 99), Feynman (1951). This motivates the search for analytic approximate solutions with pre-determined error with respect to the exact theoretical solution. In

Blanes *et al* (1999), Iserles & Norsett (1999) efficient numerical methods preserving qualitative properties of the solution of problem

$$\Phi'(t) \;=\; S(t)\,\Phi(t)\,, \quad \Phi(0) \,=\, I\,, \tag{5}$$

where $\Phi(t) \in C^{(p+q)\times(p+q)}$ are developed, and in Blanes & Jódar (1998) error bounds for the approximations using Fer's factorization, see Fer (1958), are given. For the case where $S(t)$ is an analytic matrix function, the terms of the error in each order for Fer's approximations of problem (2) are studied in Iserles (1984) using the so called method of iterated conmutators.

The aim of this paper is the construction of analytic-numerical approximations of the solution of problem (1) with a priori error bounds using Fer's factorization, under the hypothesis of continuity for the coefficients. In this sense the approach proposed here may be regarded a continuation of Jódar & Cortés (1998) where coefficients of (1) are analytic matrix functions. Apart from this advantage, we will show that for the symmetric case, the approximated solutions constructed using Fer's factorization preserve qualitative properties of the exact solution. It is important to point out that although Fer's method is often regarded as computationally expensive, very efficient methods for its implementation have been recently proposed in Blanes (1998), Celledoni & Iserles (1998) or Sidje (1998). Another advantage of Fer's factorization for approximating solutions of Riccati equations is that Fer's approximations are invariant under transformation (4) and this fact permits the choice of appropriate scalar function $f(t)$ so that the guaranteed existence interval for the solution and the approximations may be bigger than in Jódar & Cortés (1998).

This paper is organized as follows. Section 2 expresses the exact solution of problem (1) with continuous coefficients using Fer's factorization of the solution of the underlying linear system (2). Appropriate formal expressions for the approximate solutions of (1) after truncating Fer's factorization of the solution of (2) are obtained. Some qualitative properties of these approximations are explored. Convergence and error analysis are treated in section 3. Some illustrative examples are included in section 4.

We recall that if $P(t)$ is an invertible and differentiable matrix function then

$$(P^{-1}(t))' \;=\; -\,P^{-1}(t)\,P'(t)\,P^{-1}(t)\,. \tag{6}$$

If $M(t)$ is a differentiable $C^{r\times r}$ valued matrix function, by Wilcox (1967) it follows that

$$\left(e^{M(t)}\right)' \;=\; \int_0^1 e^{sM(t)}\,M'(t)\,e^{-sM(t)}\,ds\,e^{M(t)}\,. \tag{7}$$

By (6) and (7) one gets

$$\left(e^{-M(t)}\right)' \;=\; \left(\left[e^{M(t)}\right]^{-1}\right)' \;=\; -\int_0^1 e^{-(1-s)M(t)}\,M'(t)\,e^{-sM(t)}\,ds\,. \tag{8}$$

Finally, if P and Q are two matrices in $C^{r\times r}$, let us denote the conmutator of P and Q by $[P,Q] \,=\, PQ - QP$.

If M is a matrix in $C^{p\times q}$ we denote by $\|\,M\,\|$ the square root of the maximum of the set $\{|z|\,;\quad z\ \textit{eigenvalue of}\ M^H M\,\}$, where M^H denotes the conjugate transpose of M. If P, Q are matrices in $C^{r\times r}$ and P is invertible, from the Banach lemma Golub & Van Loan (1993, p 28), it follows that any matrix Q such that $\|Q-P\| \,<\, \|P^{-1}\|^{-1}$ is also invertible and

$$\|\,P^{-1} \,-\, Q^{-1}\,\| \;<\; \|P^{-1}\|\,\|Q^{-1}\|\,\|P-Q\|\,. \tag{9}$$

Hence

$$\| P^{-1} - Q^{-1} \| \leq \|P^{-1}\| \left(\|Q^{-1} - P^{-1}\| + \|P^{-1}\| \right) \|P - Q\| ,$$

and

$$\| P^{-1} - Q^{-1} \| \leq \frac{\|P^{-1}\|^2 \|P - Q\|}{1 - \|P^{-1}\| \|P - Q\|} . \tag{10}$$

If $M(t)$ is an invertible differentiable matrix function defined on the interval $[0, t_0]$, and

$$t_0 \|M'(t)\| \leq 1 - \rho , \quad 0 < \rho < 1 ,$$

then by Freedman (1996, p 225) one gets

$$\| M^{-1}(t) \| < \frac{1}{\rho} .$$

2. The Riccati equation and Fer's factorization

Let $S(t)$ be the continuous function defined by (2) and let $k_0(t)$, $K_0(t)$ be such that

$$\| S(t) \| \leq k_0(t) , \quad K_0(t) = \int_0^t k_0(u) \, du . \tag{11}$$

From Flett (1980, p 114) one gets that $\Phi(t)$ defined by (4) verifies

$$\| \Phi(t) \| \leq e^{\int_0^b \|S(t)\| dt} \leq e^{K_0(b)} , \quad 0 \leq t \leq b . \tag{12}$$

Let w_0, q_0 and $q_0(t)$ be defined by

$$\| [A(t) \quad B(t)] \| \leq q_0(t) , \quad \max\{q_0(t) ; \; 0 \leq t \leq b\} \leq q_0 , \tag{13}$$

and

$$w_0 = \left\| \begin{bmatrix} I \\ W_0 \end{bmatrix} \right\| = \left(1 + \|W_0\|^2 \right)^{1/2} . \tag{14}$$

Note that the solution $Y(t)$ of (2) can be written in terms of $\Phi(t)$ in the form

$$Y(t) = \Phi(t) \begin{bmatrix} I \\ W_0 \end{bmatrix} .$$

Thus

$$U(t) = [I \quad 0] \, Y(t) = [I \quad 0] \, \Phi(t) \begin{bmatrix} I \\ W_0 \end{bmatrix} , \tag{15}$$

$$V(t) = [0 \quad I] \, Y(t) = [0 \quad I] \, \Phi(t) \begin{bmatrix} I \\ W_0 \end{bmatrix} . \tag{16}$$

By Reid (1972, p 28), the solution $W(t)$ of (1) is defined in the interval where $U(t)$ is invertible, and in such domain

$$W(t) = V(t) \, [U(t)]^{-1} . \tag{17}$$

The next lemma is related to Lemma 2.10 of Keller & Lentini (1982) and its proof is a consequence of Freedman (1996, p 225) and lemma 1 of Jódar & Ponsoda (1995).

LEMMA 1 Let $k_0(t)$, q_0 and w_0 be defined by (11), (13) and (14) respectively. Let $t_1 \le b$, satisfying

$$t_1\, q_0\, w_0\, e^{K_0(t_1)} \;<\; 1\,, \tag{18}$$

and let

$$N \;=\; q_0\, w_0\, e^{K_0(t_1)}.$$

Then the local solution $W(t)$ of (1) is given by (17) in $[0, t_1]$. Furthermore, if $Y(t)$ is the solution of (2), then $U(t)$, given by (15) is invertible and

$$\| \,(U(t))^{-1}\, \| \;<\; \frac{1}{1 - N t_1}\,, \quad 0 \le t \le t_1\,.$$

Let $\Phi_1(t)$ and $F_1(t)$ be defined by

$$\left.\begin{aligned} \Phi(t) \;&=\; e^{F_1(t)}\, \Phi_1(t)\,; \quad \Phi_1(0) = I\,, \\ F_1(t) \;&=\; \int_0^t S(u)\, du\,, \end{aligned}\right\} \tag{19}$$

where $\Phi(t)$ is the solution of (5). By (5), (8) and (19) one gets

$$\Phi_1(t) \;=\; e^{-F_1(t)}\, \Phi(t)\,,$$

$$(\Phi_1(t))' \;=\; (e^{-F_1(t)})'\, \Phi(t) \,+\, e^{-F_1(t)}\, \Phi'(t)$$

$$= \left\{ e^{-F_1(t)} S(t) e^{F_1(t)} \;-\; \int_0^1 e^{-(1-s)F_1(t)} F_1'(t) e^{(1-s)F_1(t)}\, ds \right\} \Phi_1(t)$$

$$= \left[e^{-F_1(t)} \left\{ S(t) \;-\; \int_0^1 e^{sF_1(t)} F_1'(t) e^{-sF_1(t)}\, ds \right\} e^{F_1(t)} \right] \Phi_1(t)\,,$$

or

$$\Phi_1'(t) \;=\; S_1(t)\, \Phi_1(t)\,, \quad \Phi_1(0) \;=\; I\,, \tag{20}$$

where

$$S_1(t) \;=\; e^{-F_1(t)} \left\{ S(t) \;-\; \int_0^1 e^{sF_1(t)} F_1'(t) e^{-sF_1(t)}\, ds \right\} e^{F_1(t)}\,.$$

Note that equation (20) is of the type (5). Inductively we obtain $\Phi_2(t), \ldots, \Phi_n(t)$, where

$$\left.\begin{aligned} \Phi(t) \;&=\; \widetilde{\Phi_n(t)}\, \Phi_n(t)\,, \\ \Phi_n'(t) \;&=\; S_n(t)\, \Phi_n(t)\,, \quad \Phi_n(0) \;=\; I\,, \\ \widetilde{\Phi_n(t)} \;&=\; e^{F_1(t)}\, e^{F_2(t)} \cdots e^{F_n(t)}\,, \\ F_n(t) \;&=\; \int_0^t S_{n-1}(u)\, du\,, \quad n \ge 1\,, \quad S_0(t) = S(t)\,, \\ S_n(t) \;&=\; e^{-F_n(t)} \left\{ S_{n-1}(t) \;-\; \int_0^1 e^{sF_n(t)} F_n'(t) e^{-sF_n(t)}\, ds \right\} e^{F_n(t)}\,, \quad n \ge 1\,. \end{aligned}\right\} \tag{21}$$

From (17), if we denote

$$\left.\begin{aligned} U_n(t) &= [0 \quad I]\, e^{F_1(t)} \cdots e^{F_n(t)}\, \Phi_n(t) \begin{bmatrix} I \\ W_0 \end{bmatrix}, \\ V_n(t) &= [I \quad 0]\, e^{F_1(t)} \cdots e^{F_n(t)}\, \Phi_n(t) \begin{bmatrix} I \\ W_0 \end{bmatrix}, \end{aligned}\right\} \tag{22}$$

the exact solution $W(t)$ of (1) can be written in the form

$$W(t) = V_n(t)\, [U_n(t)]^{-1},$$

in the domain where $U_n(t)$ is invertible. If we denote the matrix functions $\widetilde{U_n(t)}$ and $\widetilde{V_n(t)}$, by

$$\left.\begin{aligned} \widetilde{U_n(t)} &= [0 \quad I]\, e^{F_1(t)} \cdots e^{F_n(t)} \begin{bmatrix} I \\ W_0 \end{bmatrix}, \\ \widetilde{V_n(t)} &= [I \quad 0]\, e^{F_1(t)} \cdots e^{F_n(t)} \begin{bmatrix} I \\ W_0 \end{bmatrix}, \end{aligned}\right\} \tag{23}$$

in the domain where $\widetilde{U_n(t)}$ is invertible, one gets Fer's approximation to the solution of the problem (1), defined by

$$\widetilde{W_n(t)} = \widetilde{V_n(t)} \left[\widetilde{U_n(t)}\right]^{-1}. \tag{24}$$

Remark 1 Note that $S_n(t)$ defined by (21) can be written in the form

$$S_n(t) = e^{-F_n(t)} \left\{ \int_0^1 \left(S_{n-1}(t) - e^{sF_n(t)} S_{n-1}(t) e^{-sF_n(t)} \right) ds \right\} e^{F_n(t)},$$

because $F_n'(t) = S_{n-1}(t)$. Hence

$$\begin{aligned} S_n(t) &= e^{-F_n(t)} \left\{ -\int_0^1 \left(\int_0^v \frac{\partial}{\partial u} \left[e^{uF_n(t)} S_{n-1}(t) e^{-uF_n(t)} \right] du \right) dv \right\} e^{F_n(t)} \\ &= e^{-F_n(t)} \left\{ \int_0^1 \left(\int_0^v e^{uF_n(t)} \left[S_{n-1}(t), F_n(t) \right] e^{-uF_n(t)} du \right) dv \right\} e^{F_n(t)} \\ &= \int_0^1 \left\{ \int_0^v e^{-(1-u)F_n(t)} \left[S_{n-1}(t), F_n(t) \right] e^{(1-u)F_n(t)} du \right\} dv . \end{aligned}$$

Thus (21)-(22) is equivalent to the expression given in Blanes & Jódar (1998, p 192).

Remark 2 The computation of Fer's approximations (23)-(24) involves matrix exponentials. Very recently new and interesting methods for computing matrix exponentials have appeared in Celledoni & Iserles (1998) and Sidje (1998).

Remark 3 If $f(t)$ is a scalar function the exact solution of (1) is invariant under transformation (4). In fact, if

$$S(f,t) = S(t) + f(t)\, I_{p+q},$$

and $Y(t)$ is the solution of (2), then

$$Y(f,t) = e^{f(t) I_{p+q}}\, Y(t) = \begin{bmatrix} e^{f(t)}\, U \\ e^{f(t)}\, V \end{bmatrix},$$

is the solution of

$$\frac{d}{dt} Y(f,t) \;=\; S(f,t)\, Y(f,t)\,,$$

and the solution of the transformed Riccati equation is given by

$$W(f,t) \;=\; \left[e^{f(t)} V\right] \left[e^{f(t)} U\right]^{-1} \;=\; V\, U^{-1} \;=\; W(t)\,.$$

In general, see Schiff & Schnider (1996), numerical methods for solving (1) depend of $f(t)$, however Fer's method changes the approximation in the same way as the exact solution because $e^{F_1(t)}$ is transformed into

$$e^{f(t) I_{p+q} + F_1(t)} \;=\; e^{f(t)}\, e^{F_1(t)}\,,$$

and for $j \geq 2$ matrices $F_j(t)$ defined by (21) remain unaltered.

Remark 4 For the symmetric Riccati equation $D(t) = A(t)^T$ with $A(t)$ real and $B(t)$, $C(t)$ real and symmetric, solving (2) is equivalent to solving the Hamiltonian system with

$$H \;=\; \frac{1}{2} p^T\, B(t)\, p \;+\; p^T\, A(t)\, q \;-\; \frac{1}{2} q^T\, C(t)\, q\,,$$

$$q' \;=\; \frac{\partial H}{\partial p}\,, \quad p' \;=\; -\frac{\partial H}{\partial q}\,.$$

Using Fer's approach, which is a simplectic integrator, we are solving exactly a perturbed Hamiltonian system, or solving exactly a symmetric perturbed Riccati equation, and some qualitative properties are preserved. This also happens using other simplectic integrators, see Stuart & Humphries (1996, ch 8) for details.

3. Convergence and error analysis

In this section we address three questions. The first is the determination of the interval where Fer's approximation (24) is well defined. Secondly, since the convergence of Fer's factorization is local, its extension to a wider interval where the Riccati equation is solvable is treated. Finally error bounds of $\widetilde{W_n(t)}$ defined by (24) with respect to the exact solution $W(t)$ are proposed.

By remark 2 of Blanes & Jódar (1998, p 192) Fer's algorithm (21)-(22) provides the inequality

$$\| S_n(t) \| \;\leq\; k_n(t) \quad for \quad 0 \leq t \leq b\,.$$

Under the definition

$$K_n(t) \;=\; \int_0^t k_n(s)\, ds\,,$$

it follows that

$$K_n(t) \;\geq\; \int_0^t \|S_n(u)\|\, du \;\geq\; \|F_{n+1}(t)\|\,. \tag{25}$$

If after n steps we impose $\Phi_n(t) = I$, we are left with an approximation $\widetilde{\Phi(t)}$ to the exact solution $\Phi(t)$ of (5) such that if

$$\int_0^t \|S(u)\|\, du \;\leq\; K_0(t) \;\leq\; \psi \;=\; 0.8604065\,, \tag{26}$$

then, by Blanes & Jódar (1998, p 192), if $K_0(t) < \psi$, considering

$$K_n(t) = \psi \left(\frac{K_0(t)}{\psi} \right)^{2^n}\,, \tag{27}$$

one gets

$$K_n\,(t) \;<\; K_{n-1}(t) \;<\; \cdots \;<\; K_0(t)\,, \quad \lim_{n\longrightarrow\infty} K_n(t) \;=\; 0\,, \tag{28}$$

see Blanes (1998, p 43) for details. Let t_2, with $0 \leq t_2 \leq b$, be a real number verifying (26). Then Fer's factorization for $\Phi(t)$ converges in $[0, t_2]$. Let t_1 be given by lemma 1 and let $\delta = \min\{\, t_1\,,\, t_2, \}$. In the interval $0 \leq t \leq \delta$ one gets

$$\Phi(t) \;-\; \widetilde{\Phi(t)} \;=\; \Phi(t)\, D_n(t)\,, \quad D_n(t) \;=\; I - \Phi_n^{-1}(t)\,. \tag{29}$$

By (6), it follows that

$$\left(\Phi_n^{-1}(t)\right)' \;=\; -\,\Phi_n^{-1}(t)\, S_n(t)\,; \quad \Phi_n^{-1}(0) \;=\; I\,, \tag{30}$$

$$D_n(t) \;=\; -\int_0^t \Phi_n^{-1}(u)\, S_n(u)\, du\,. \tag{31}$$

Taking norms in (29) it follows that

$$\|\, \Phi(t) \;-\; \widetilde{\Phi(t)}\,\| \;\leq\; \|\Phi_n(t)\|\, \|D_n(t)\|\,. \tag{32}$$

By (30) and Flett (1980, p 144) one obtains

$$\|\Phi_n^{-1}(t)\| \;\leq\; e^{K_n(b)}\,, \tag{33}$$

and by (31)

$$\|D_n(t)\| \;\leq\; e^{K_n(b)}\, K_n(b)\,. \tag{34}$$

By (12), (29) and (31)-(34), one gets

$$\|\Phi(t) - \widetilde{\Phi(t)}\| \;\leq\; e^{K_0(\delta)}\, e^{K_n(\delta)}\, K_n(\delta)\,. \tag{35}$$

Note that by lemma 1,

$$\|\,(U(t))^{-1}\,\| \;\leq\; \frac{1}{1 - \delta q_0 w_0 e^{K_0(\delta)}} \;=\; M_1\,, \tag{36}$$

and by (12) and (16),

$$\|V(t)\| \;\leq\; w_0\, e^{K_0(\delta)} \;=\; M_2\,. \tag{37}$$

By (16) and (35) it follows that

$$\|V(t) - \widetilde{V(t)}\| \;\leq\; \|\,[\,0 \quad I\,]\,\|\, \|\Phi(t) - \widetilde{\Phi(t)}\| \left\| \begin{bmatrix} I \\ W_0 \end{bmatrix} \right\| \;\leq\; M_2\, \alpha_n\,, \tag{38}$$

where M_2 is given by (37) and

$$\alpha_n \;=\; e^{K_n(\delta)}\, K_n(\delta)\,. \tag{39}$$

Note that by (28) it follows that

$$\alpha_1 \;>\; \alpha_2 \;>\; \cdots \;>\; \alpha_n \;>\; \cdots\,, \quad \lim_{n\longrightarrow\infty} \alpha_n(t) = 0\,, \quad 0 \leq t \leq \delta\,.$$

Let us write

$$\left.\begin{aligned} W(t) - \widetilde{W(t)} \;&=\; V(t)(U(t))^{-1} \;-\; \widetilde{V(t)}(\widetilde{U(t)})^{-1} \\ = V(t)\left[(U(t))^{-1} - (\widetilde{U(t)})^{-1}\right] &+ \left[V(t) - \widetilde{V(t)}\right](U(t))^{-1} \\ + \left[V(t) - \widetilde{V(t)}\right]&\left[(U(t))^{-1} - (\widetilde{U(t)})^{-1}\right]\,, \end{aligned}\right\} \tag{40}$$

and note that

$$U(t) \, - \, \widetilde{U(t)} \;=\; [\, I \quad 0\,] \left(\Phi(t) - \widetilde{\Phi(t)} \right) \left[\begin{array}{c} I \\ W_0 \end{array} \right] ,$$

and by (35) one gets

$$\left. \begin{array}{rl} \| U(t) \, - \, \widetilde{U(t)} \| \;=\; \| \, [\, I \quad 0\,] \, \| \, \left\| \Phi(t) - \widetilde{\Phi(t)} \right\| \, \left\| \left[\begin{array}{c} I \\ W_0 \end{array} \right] \right\| \\ \\ \leq \, w_0 \, e^{K_0(\delta)} \, e^{K_n(\delta)} \, K_n(\delta) \;=\; M_2 \, \alpha_n \, . \end{array} \right\} \tag{41}$$

Let n be the first natural number such that

$$(1 + \xi) \, M_2 \, \alpha_n \;<\; 1 \, - \, \delta \, M_2 \, q_0 \, , \quad 0 < \xi \, ,$$

or in an equivalent form

$$\alpha_n \, < \, \frac{1}{1+\xi} \left(\frac{1}{M_2} \, - \, \delta \, q_0 \right) . \tag{42}$$

By (18), (37), (41) and (42) it follows that

$$\| U(t) - \widetilde{U(t)} \| \, \leq \, M_2 \alpha_n \, < \, \frac{1}{1+\xi} \, M_2 \left(\frac{1}{M_2} \, - \, \delta \, q_0 \right)$$

$$= \, \frac{1}{1+\xi} \, (1 \, - \, M_2 \, \delta \, q_0) \, \leq \, \frac{1}{1+\xi} \, \| \, (U(t))^{-1} \, \|^{-1} \, < \, \| \, (U(t))^{-1} \, \|^{-1} \, . \tag{43}$$

Note that by (43) one gets

$$\left. \begin{array}{l} M_1 \, M_2 \, \alpha_n \, = \, \dfrac{M_2 \, \alpha_n}{1 - \delta q_0 M_2} \, < \, \dfrac{1}{1+\xi} \, < \, 1 \, , \\ \\ 1 \, - \, M_1 \, M_2 \, \alpha_n \, > \, 1 \, - \, \dfrac{1}{1+\xi} \, = \, \dfrac{\xi}{1+\xi} \, . \end{array} \right\} \tag{44}$$

By (9), (10) and (41) one gets

$$\left\| (U(t))^{-1} \, - \, \left(\widetilde{U(t)} \right)^{-1} \right\| \, \leq \, \frac{\| \, (U(t))^{-1} \, \|^2 \, \| \, U(t) \, - \, \widetilde{U(t)} \, \|}{1 \, - \, \| \, (U(t))^{-1} \, \| \, \| \, U(t) \, - \, \widetilde{U(t)} \, \|} \, \leq \, \frac{M_1^2 \, M_2 \, \alpha_n}{1 \, - \, M_1 \, M_2 \, \alpha_n} \, , \tag{45}$$

where M_1, M_2 and α_n are defined by (36), (37) and (39) respectively. By (44)-(45) one gets

$$\left\| (U(t))^{-1} \, - \, \left(\widetilde{U(t)} \right)^{-1} \right\| \, \leq \, \left(1 + \frac{1}{\xi} \right) \, M_1^2 \, M_2 \, \alpha_n \, . \tag{46}$$

Taking norms in (40) and using (36), (37), (38) and (46) it follows that

$$\| W(t) - \widetilde{W_n(t)} \|$$

$$\leq \, \left(1 + \frac{1}{\xi} \right) M_1^2 M_2^2 \alpha_n + M_1 M_2 \alpha_n + \left(1 + \frac{1}{\xi} \right) M_1^2 M_2^2 \alpha_n^2$$

$$= \, H_1 \, \alpha_n \, + \, H_2 \, \alpha_n^2 \, , \quad 0 \leq t \leq \delta \, , \tag{47}$$

where

$$H_1 \, = \, M_1 M_2 \, + \, \left(1 + \frac{1}{\xi} \right) \, M_1^2 \, M_2^2 \, , \tag{48}$$

$$H_2 \;=\; \left(1+\frac{1}{\xi}\right)\, M_1^2\, M_2^2 \;=\; H_1 \;-\; M_1\, M_2\,. \tag{49}$$

From (47), given an admissible error $\epsilon > 0$, taking the first positive integer n satisfying (42) and

$$H_1\, \alpha_n \;+\; H_2\, \alpha_n^2 \;<\; \epsilon\,, \tag{50}$$

one gets

$$\|\, W(t) \;-\; \widetilde{W_n(t)}\,\| \;\leq\; \epsilon\,, \quad 0 \leq t \leq \delta\,. \tag{51}$$

Since the algebraic equation $z^2 + \frac{H_1}{H_2} z - \frac{\epsilon}{H_2} \;=\; 0$ has a unique positive root defined by

$$z_1 \;=\; -\frac{1}{2}\,\frac{H_1}{H_2} \;+\; \frac{1}{2}\sqrt{\left(\frac{H_1}{H_2}\right)^2 \;+\; \frac{4\epsilon}{H_2}}\,,$$

condition (50) can be expressed in the more explicit form

$$\alpha_n \;<\; -\frac{1}{2}\,\frac{H_1}{H_2} \;+\; \frac{1}{2}\sqrt{\left(\frac{H_1}{H_2}\right)^2 \;+\; \frac{4\epsilon}{H_2}} \;=\; z_1\,,$$

and taking into account (44) then

$$\alpha_n \;<\; \min\left\{\, [(1+\xi)M_1M_2]^{-1} \;,\; -\frac{1}{2}\,\frac{H_1}{H_2} \;+\; \frac{1}{2}\sqrt{\left(\frac{H_1}{H_2}\right)^2 \;+\; \frac{4\epsilon}{H_2}}\,\right\}. \tag{52}$$

Observe that α_n has an explicit dependence on ξ of the form $\alpha_n \;<\; \min\{A(\xi), B(\xi)\}$, $A(\xi)$ and $B(\xi)$ being decreasing and increasing functions respectively. The highest allowed value for α_n is given for $A(\xi) = B(\xi)$ which has the solution

$$\xi \;=\; \frac{M_1\, M_2 \;+\; 1}{\epsilon}\,.$$

Substituting into (52) gives the simple condition

$$\alpha_n \;<\; \frac{\epsilon}{M_1\, M_2\; (M_1\, M_2 + 1 + \epsilon)}\,. \tag{53}$$

Summarizing, the following result has been established.

Theorem 1 Let $\epsilon > 0$ and let $\delta = \min\{t_1, t_2\}$, where t_1 is defined by (18) and $K_0(t_2) < \psi = 0.8604065$. Let n_0 be the first positive integer satisfying (53) where M_1 and M_2 are defined by (36) and (37) respectively. Let $\widetilde{W_n(t)}$ be the Fer approximation to the exact solution $W(t)$ of (1), defined by (23)-(24) for $n \geq n_0$. Then $\widetilde{W_n(t)}$ is well defined and (51) holds uniformly for $0 \leq t \leq \delta\,,\; n \geq n_0$.

Remark 5 The invariance of the exact solution $W(t)$ of problem (1) under transformation (3) has an important advantage for finding a bigger existence domain and a better error bound. So, for instance, instead of considering (26) we may select $f_1(t)$ so that

$$\int_0^t \|\, S(u) \;-\; f_1(u)\, I_{p+q}\,\|\; du \;<\; \int_0^t \|S(u)\|\; du\,, \tag{54}$$

and for a bigger value of t one gets

$$\int_0^t \|\, S(u) \;-\; f_1(u)\, I_{p+q}\,\|\; du \;<\; \psi\,. \tag{55}$$

This fact also improves the error bounds using Fer's factorization. Regarding the invertibility of $U(t)$ in lemma 1, as a consequence of the existence domain of $W(t)$ remember that $U(t)$ is invertible in the domain $[0, t_1]$ where (see lemma 1)

$$t_1 \;<\; \left[w_0\, q_0\, e^{K_0} \right]^{-1} ,$$

and

$$q_0 \;\geq\; \| [\; A(t) - f_2(t) \qquad B(t)\;] \| \;, \quad K_0 \;\geq\; \int_0^t \| \, S(u) \;-\; f_2(u)\, I_{p+q} \, \| \; du \, .$$

Thus we could choose the function $f_2(t)$ which minimizes the product $q_0 e^{K_0}$. This function $f_2(t)$ may be different from $f_1(t)$ verifying (54)-(55).

Remark 6 The conclusion of theorem 1 is only valid in the interval $[0, \delta]$ where $\delta = \min\{t_1, t_2\}$. However, the exact solution $W(t)$ can be defined in some interval $[0, t_1]$ with $t_1 > t_2$. Let us suppose that $W(t)$ is defined in $[0, t_1]$ and split the interval into subintervals of length $h = t_1/N$ where N is a positive integer large enough to ensure that Fer's approximation is convergent in each subinterval. Let $h_j = jh$ with $0 \leq j \leq N$ and note that if $h_j \leq t \leq h_{j+1}$, then it follows that

$$\frac{d}{dt}\Phi(t, h_j) \;=\; S(t)\, \Phi(t, h_j) \,, \quad \Phi(h_j, h_j) \;=\; I \,, \quad h_j \leq t \leq h_{j+1} \,,$$

$$\Phi(t, 0) \;=\; \Phi(t, h_i)\, \Phi(h_i, h_{i-1}) \cdots \Phi(h, 0) \,, \quad h_i \leq t \leq h_{i+1} \,.$$

Thus we can approximate $\Phi(t) \;=\; \Phi(t, 0)$ by

$$\widetilde{\Phi(t,0)} \;=\; \widetilde{\Phi(t, h_i)}\, \widetilde{\Phi(h_i, h_{i-1})} \cdots \widetilde{\Phi(h, 0)} \,,$$

$$\widetilde{\Phi(h_{j+1}, h_j)} \;=\; e^{F_1(h_{j+1}, h_j)} \cdots e^{F_n(h_{j+1}, h_j)} \,.$$

Let us denote by $K_0(h_{j+1}, h_j)$ and $K_n(h_{j+1}, h_j)$ the quantities

$$\left. \begin{aligned} K_0(h_{j+1}, h_j) \;&=\; \int_{jh}^{(j+1)h} \|S(u)\| \, du \,, \\ K_n(h_{j+1}, h_j) \;&=\; \psi \left(\frac{K_0(h_{j+1}, h_j)}{\psi} \right)^{2^n} \end{aligned} \right\}$$

Then it is easy to prove (see Blanes & Jódar for details)

$$\left. \begin{gathered} \| \Phi(t, 0) - \widetilde{\Phi(t, 0)} \| \\ \leq e^{K_0(t,0)} \sum_{j=0}^{i} K_n(h_{i-j+1}, h_{i-j})\, e^{K_n(t, h_i) + K_n(h_i, h_{i-1}) + \cdots + K_n(h_{i-j+1}, h_{i-j})} \,, \\ h_i \;\leq t \;\leq\; h_{i+1} \,, \quad 1 \;\leq\; i \;\leq\; N-1 \,. \end{gathered} \right\}$$

Taking $h > 0$ small enough to ensure that

$$\delta_1 \;=\; \frac{h \max\{ \, \|S(t)\| \, ; \;\; 0 \leq t \leq t_1 \, \}}{\psi} \;<\; 1 \,,$$

one gets

$$K_n(h_{j+1}, h_j) \;\leq\; \psi \, \delta_1^{2^n} \;=\; \delta_{1,n}\, \psi \,, \quad \delta_{1,n} \;=\; \delta_1^{2^n} \,.$$

Hence

$$\|\Phi(t,0) - \widetilde{\Phi(t,0)}\| \leq N\, e^{K_0(t,0)}\, \Lambda_n\,, \quad \Lambda_n = \delta_{1,n}\, e^{(N+1)\,\delta_{1,n}}\,. \tag{56}$$

Since the condition $\delta = \min\{t_1, t_2\}$ is changed to $\delta = t_1$, then in (38) and (41) we can replace the coefficients α_n by $N\,\Lambda_n$. Following the same steps as previously, we obtain the condition

$$N\,\Lambda_n < \frac{\epsilon}{M_1\, M_2\, (M_1 M_2 + 1 + \epsilon)}\,. \tag{57}$$

Observe that, if desired, we can fix the order of the Fer expansion n, and look for N such that (57) is satisfied.

Hence, for a given value of $\epsilon > 0$, in an analogous way to theorem 1 one may select n and N large enough so that

$$\left\| W(t) - \widetilde{W_n(t)} \right\| < \epsilon\,, \quad 0 \leq t \leq t_1\,.$$

Remark 7 Condition $K_0(t) < \psi = 0.8604065$ for the convergence of Fer's approximation can be improved if

$$K_1(t) < \psi \left(\frac{K_0(t)}{\psi} \right)^2\,.$$

Note that by (25) one gets

$$\begin{aligned}
&\int_0^t \|\, S_1(z_1)\,\|\, dz_1 \\
&= \int_0^t \left\| \int_0^1 \left\{ \int_0^v e^{-(1-u)F_1(z_1)}\, [S(z_1), F_1(z_1)]\, e^{(1-u)F_1(z_1)}\, du \right\} dv \right\| dz_1 \\
&\leq \frac{1 - e^{2\,K_0(t)}\,(1 - 2K_0(t))}{(2\,K_0(t))^2} \int_0^t \left\| \int_0^{z_1} [S(z_1), S(z_2)]\, dz_2 \right\| dz_1 \leq K_1(t)\,.
\end{aligned}$$

We try to take a function $K_1(t)$ as close as possible to the left hand side of last inequality. Note that if $[S(t_1), S(t_2)] = 0$, then we can take $K_1 = 0$ because the first order of Fer's expansion gives the exact solution. If this is not the case, there are several situations where $[S(t_1), S(t_2)]$ is small enough to obtain better error bounds than in theorem 1:

(i) If $S(t)$ is a constant matrix with small time dependent perturbation.

(ii) $S(t)$ varies very slowly with time (adiabatically).

(iii) If $t_1,\, t_2 \in [t_0, t_0 + h]$ then $[S(t_1), S(t_2)] = O(h)$.

In these cases better bounds may be obtained than by considering (27) for $n = 1$. In an analogous way, a better convergence radius for Fer's algorithm can be obtained by taking

$$K_n(t) = \psi \left(\frac{K_1(t)}{\psi} \right)^{2^{n-1}}\,,$$

that tends to zero faster than (27).

3. Examples

In this section some illustrative examples are studied on Fer's approximations to the solution of the Riccati equation (1).

EXAMPLE 1 Consider the scalar Riccati problem presented in Schiff & Schnider (1996)

$$\dot{y} \,=\, \rho + \frac{y\,(y-t)}{\rho}\,; \quad -1 < t < 1\,, \quad y(-1) = -1 - m\,,$$

with $\rho \ll 1$. With the notation of the previous sections one gets

$$S(t) \,=\, \begin{bmatrix} -\dfrac{t}{2\,\rho} & \rho \\ & \\ -\dfrac{1}{\rho} & \dfrac{t}{2\rho} \end{bmatrix}, \quad w_0 \,=\, \sqrt{1+(1+m)^2}\,,$$

$$\|S(t)\| \,\leq\, \frac{\sqrt{5}}{2\rho}\,, \quad q_0 \,=\, \left\|\left(-\frac{t}{2\rho}, \rho\right)\right\|.$$

Taking into account remark 5 we may select an appropriate function $f_2(t)$ so that q_0 is smaller. In fact, since q_0 must satisfy

$$q_0 \,\geq\, \|\,[\,A(t) - f_2(t) \qquad B(t)\,]\,\|\,,$$

we can take $f_2(t) = A(t) = -t/2\rho$ and thus

$$q_0 \,=\, \|\,[\,0 \quad \rho\,]\,\| \,=\, \rho\,.$$

With this choice of f_2 we have $K_0(t) \leq \frac{t\sqrt{2}}{\rho}$. With the notation of previous sections the length of the interval $[-1, -1+t_1]$ where $U(t)$ is invertible is defined by

$$t_1\, q_0\, w_0\, e^{K_0(t_1)} \,=\, \eta$$

with $\eta < 1$. Hence, t_1 is the solution of

$$t_1\, e^{t_1\,\frac{\sqrt{2}}{\rho}} \,=\, \frac{\eta}{\rho\,\sqrt{1+(1+m)^2}}\,.$$

Taking into account lemma 1, and the fact that now $K_0(t) \,=\, t\frac{\sqrt{5}}{2\,\rho}$, then t_2 can be chosen by the inequality

$$t_2 \,<\, \frac{2}{\sqrt{5}}\,\rho\,\psi\,. \tag{58}$$

However we can improve this result if, considering remark 7 and its notation, we decompose $S(t)$ as

$$S(t) = S_1 + tS_2\,, \quad S_1 = \begin{bmatrix} 0 & \rho \\ -\dfrac{1}{\rho} & 0 \end{bmatrix}, \quad S_2 = \begin{bmatrix} -\dfrac{1}{2\rho} & 0 \\ 0 & \dfrac{1}{2\rho} \end{bmatrix}.$$

Then

$$\left\|\int_{t_0}^{t_0+z_1} [S(t_0+z_1), S(z_2)]dz_2\right\| = \left\|\int_{t_0}^{t_0+z_1} (t_0+z_1-z_2)[S_1, S_2]dz_2\right\| \leq z_1^2\,\|S_1\|\,\|S_2\| \leq \frac{z_1^2}{2\rho^2}\,.$$

So we can consider

$$K_1(t_0+t_2, t_0) \,=\, \frac{t_2^3}{6\rho^2}\,\frac{1-e^x(1-x)}{x^2} \,<\, \psi, \tag{59}$$

with $x = t_2 \dfrac{\sqrt{5}}{\rho}$. Then, $\delta = \min\{ t_1 \, , \, t_2 \}$ and

$$M_1 \; = \; \frac{1}{1-\eta} \, , \quad M_2 \; = \; \frac{\eta}{\delta \, \rho} \, .$$

In accordance with (57) one gets

$$\alpha_n \; < \; \frac{\epsilon (1-\eta)^2 \rho^2 \delta^2}{\eta^2 + (1+\epsilon)(1-\eta)\eta\rho\delta}$$

If $t_2 < t_1$ or n takes a large value, we may split the interval $[-1, -1+t_1]$ into N parts so that in each interval the method is convergent in accordance with remark 6. In this case $\delta = t_1$.

Let us consider $\rho = \frac{1}{25}$ and $m = \frac{1}{10}$. In order to avoid very small values of α_n it is interesting to take $\eta < \frac{1}{2}$. In particular, taking $\eta = \frac{1}{4}$ we obtain $t_1 = 0.1045$. Considering (58) we obtain $t_2 < 0.0385$ and from (59) $t_2 < 0.0978$ improving the convergence domain as well as the error bounds. Given an admisible error $\epsilon = 10^{-3}$ and taking $\delta = t_1$ gives $\alpha_n < 3.23 \cdot 10^{-7}$ and it is necessary to split the interval. If we decide to take $n = 2$ then $K_1 < 5.27 \cdot 10^{-4}$, which can be obtained with $N = 5$.

Example Consider the Riccati problem

$$W' = DW - W^2 + C \, , \quad W(0) = \frac{1}{1+e} \begin{bmatrix} 0 & a_2 + e - \frac{ke^2}{4} \\ -(a_1 + e + \frac{ke^2}{4}) & 0 \end{bmatrix} .$$

$$D = \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix} , \quad C = \frac{1}{1 + e\cos(t)} t \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix} ,$$

with $k = 1/\mu$, $c = \sqrt{9\mu - 8 + 2e^2 + k^2 e^4}$, $c_1 = \frac{3}{2}(1-\mu)$, $c_2 = 3 - c_1$, $a_1 = \frac{1}{4}(2c_1 + 1 - c)$, $a_2 = \frac{1}{4}(2c_2 + 1 - c)$, which appears in the study of the stability of the triangular points in the elliptic restricted three body problem. Here e is the eccentricity of the main bodies, so $0 < e < 1$, and μ is a parameter depending on the ratio of the masses of the main bodies where the condition $0 \ll \mu < 1$ is necessary, see Meire (1982).

In this case we have

$$w_0 = \frac{1 + a_2 + 2e - \frac{ke^2}{4}}{1+e} \, , \quad q_0 = 1 \, ,$$

and

$$\|S(t)\| \; = \; \sqrt{4 \; + \; \frac{c_2^2}{(1 \; + \; e \, \cos t)^2}} \, .$$

Note that $diag \, S(t) \; = \; 0$; the values of q_0 and K_0 cannot be improved taking into account remark 5.

The length t_1 of the existence interval for the solution is defined by

$$t_1 \, e^{\int_0^{t_1} \|S(u)\| du} \; = \; \eta \, \frac{1+e}{1 + a_2 + 2e - \frac{ke^2}{4}} \, , \tag{60}$$

with $\eta < 1$. We will take $\eta = \frac{4}{5}$.

Considering that $\|S(t)\| \leq \sqrt{4 + c_2^2/(1-e)^2}$ one gets

$$K_0(t_2, 0) \; = \; t_2 \sqrt{4 \; + \; \frac{c_2^2}{(1-e)^2}} \; < \; \psi \, . \tag{61}$$

This result may be improved considering that

$$\| [S(z_1), S(z_2)] \| \le \frac{2\sqrt{5}\, c_2}{(1-e)^2} |z_1^2 - z_2^2| ;$$

from remark 7 one gets

$$K_1(t_2, 0) = \frac{2\sqrt{5}\, c_2}{(1-e)^2} \frac{t_2^4}{12} \left(\frac{1 - e^x(1-x)}{x^2} \right) < \psi, \tag{62}$$

where $x = 2K_0(t_2, 0)$. Taking $\delta = \min\{ t_1 , t_2 \}$ one gets $M_1 = 5$ and $M_2 = \frac{4}{5\delta}$. Given an admissible error ϵ the order n of Fer's approximation is defined by

$$\alpha_n < \frac{\epsilon \delta^2}{16 + 4\delta(1+\epsilon)}. \tag{63}$$

Taking $e = 1/2$, $\mu = 99/100$ then, $c_1 = 3/200$, $c_2 = 199 c_1$, $k = 100/99$, $c = 1.7672$, $a_1 = -0.03668$, $a_2 = 1.44832$ and $w_0 = 2.6568$, then condition (60) gives $t_1 = 0.2$. Considering (61) we obtain $t_2 < 0.137$ and from (62) $t_2 < 0.377$ where now $t_2 > t_1$. Then $\delta = t_1$ and $M_1 M_2 = 20$. If $\epsilon = 10^{-3}$ condition for the required order n of Fer's approximation is defined by $\alpha_n < 2.5 \cdot 10^{-7}$. This gives $n = 3$ or, alternatively, $n = 2$ taking $N = 2$.

Acknowledgements

This work has been supported by the D.G.I.C.Y.T. grant PB96-1321-C02-02. The first author would like to express his thanks to the Department of Applied Mathematics and Theoretical Physics of the University of Cambridge for their hospitality and the Generalitat Valenciana for its financial support.

References

ABOU-KANDIL, H. 1986 Elaboration de structures de commande hiérarchisées: approches monocritře et multicritère. *These de Doctorat d'Etat.* Université Pierre et Marie Curie. Paris 6.

ASCHER U. M., MATTHEIJ, R. M., & RUSSELL, R. D. 1988 *Numerical Solutions of Boundary Value Problems for Ordinary Differential Equations.* Englewood Cliffs, NJ. Prentice-Hall.

BAGHI, A. 1984 *Stackelberg Differential Games in Economic Models. (Springer Lectures Notes in Control and Information Sciences).* Berlin. Springer.

BLANES, S. 1998 Estudio de sistemas dinámicos clásicos y cuánticos utilizando métodos algebraicos, *Ph. D. Thesis*, Univ. of Valencia.

BLANES, S., & JÓDAR, L. 1998 Continuous numerical solutions of coupled mixed partial differential systems using Fer's factorization, *J. Comput. Appl. Math.*, **101**, 189-202.

BLANES, S., CASAS, F., & ROS, J. 1999 *Improved higher order integrators based on Magnus expansion*, DAMTP, Univ. Cambridge, 1999/NA08, Cambridge.

CASTI, J. L. 1977 *Dynamical Systems and Their Applications: Linear Theory.* New York. Academic.

CELLEDONI, E., & ISERLES, A. 1998 *Approximate the Exponential from a Lie Algebra to a Lie Group*, DAMTP, Univ. Cambridge, 1998/NA3, Cambridge.

CHOI, C. H., & LAUB, A. J. 1990 Efficient matrix-valued algorithms for solving stiff Riccati differential equations. *IEEE Trans. Automatic Control*, Vol. **35, No. 7**, 770-776.

CLEMHOUT, S., & WAN JR., H. Y. 1979 Interactive economic dynamics and differential games. *J. Optimiz. Theory Appl.* **27**, 7-30.

CRUZ, J. B., & CHEN, C. I. 1971 Series Nash solution of two person nonzero sum linear quadratic games. *J. Optimiz. Theory Appl.*, **7**, 240-257.

DIECI, L. 1992 Numerical integration of the differential Riccati equation and some related issues. *SIAM J. Numer. Anal.*, **29**, 781-815.

FER, F. 1958 Résolution de l'équation matricielle $U' = pU$ par produit infini d'exponentielles matricielles, *Bull. Classe Sci. Acad. Roy. Bel.*, **44**, 818-829.

FEYNMAN, R. P. 1951 An operator calculus having applications in quantum electrodynamics. *Phys. Rev* **84** 108-128.

FLETT, T. M. 1980 *Differential Analysis*, Cambridge Univ. Press, New York.

FREEDMAN, H. I. 1996 An explicit estimate of the norm of the inverse of a matrix. *SIAM Rev.* **11**, 254-256.

GOLUB, G. H., & VAN LOAN, C. F. 1993 *Matrix Computations*, John Hopkins Univ. Press, Baltimore, Maryland.

ISERLES, A. 1984 Solving ordinary differential equations by exponentials of iterated conmutators. *Numer. Math.* **45**, 183-199.

ISERLES, A., & NORSETT, S. P. 1999 On the solution of linear differential equations in Lie groups. *Philosophical Trans. Royal Soc.* **A 357**, 983-1019.

JÓDAR, L., & ABOU-KANDIL, H. 1988 A resolution method for Riccati differential systems coupled in their quadratic therms. *SIAM J. Math. Anal.*, **19**, 1225-1230.

JÓDAR, L., & CORTÉS, J. C. 1998 Rational matrix approximation with a priori error bounds for non-symmetric matrix Riccati equations with analytic coefficients. *IMA J. Numer. Anal.*, **18**, 545-561.

JÓDAR, L., & PONSODA, E. 1995 Non-autonomous Riccati-type matrix differential equations: existence interval, construction of continuous numerical solutions and error bounds. *IMA J. Num. Anal.*, **15**, 61-74.

KAILATH, T. 1980 *Linear Systems.* Englewood Cliffs, NJ. Prentice-Hall.

KELLER, H. B., & LENTINI, M. 1982 Invariant imbedding, the box scheme and an equivalence between them. *SIAM J. Numer. Anal.* **19**, 942-962.

KENNEY, S. S., & LEIPNIK, R. B. 1985 Numerical integration of the differential matrix Riccati equation. *IEEE Trans. Aut. Control.* **AC-30**, 962-970.

KRATZ, W. 1995 *Quadratic Functionals in Variational Analysis and Control Theory.* Academie Verlag. Munchen.

KUNKEL, P., & MEHRMANN, M. 1990 Numerical solution of differential-algebraic Riccati equations. *linear Algebra Appl.* **137-138**, 39-66.

MARITON, M. 1990 *Jump Linear Systems in Automatic Control.* New York. Marcel Dekker.

MEIRE, R. 1982 On the stability of the triangular points in the elliptic restricted problem. *Astron. Astophys.* bf 110, 152-155.

OSHMAN, Y., & BAR-ITZHACK, I. Y. 1984 Eigenfactor solution of the matrix Riccati equation. A continuous square root algorithm. *Proc. 23rd Conf. on Decision and Control (Las Vegas, 1984)*, 503-508.

REID, W. T. 1972 *Riccati Differential Equations.* New York. Academic.

SCHIFF, J., & SCHNIDER, S. 1996 A natural approach to the numerical integration of Riccati differential equations. *Technical Report.* Bar Ilan Univ. Israel.

SIDJE, R. B. 1998 Expokit: a software package for computing matrix exponentials. *ACM Trans. Math. Software* **24**, 130-156.

STARR, A. W., & HO, Y. C. 1969 Non-zero sum differential games. *J. Optimiz. Theory Appl.*, **3**, 179-197.

STUART, A. M., & HUMPHRIES, A. R. 1996 *Dynamical Systems and Numerical Analysis.* Cambridge Univ. Press, New York.

WILCOX, R. M. 1967 Exponential operators and parameter diferentiation in quantum physics, *J. Math. Phys.* **8**, 962-982.