Pergamon

Computers Math. Applic. Vol. 32, No. 8, pp. 41–50, 1996

0898-1221/96 $15.00 + 0.00

PII: S0898-1221(96)00165-4

# Construction and Computation of Variable Coefficient Sylvester Differential Problems

L. JÓDAR, J. L. CORTÉS AND J. L. MORERA
Departamento de Matemática Aplicada, Universidad Politécnica de Valencia
P.O. Box 22.012, Valencia, Spain
ljodar@upvnet.upv.es

*(Received June 1996; accepted July 1996)*

**Abstract**—In this paper, initial value problems for Sylvester differential equations $X'(t) = A(t)X(t)+X(t)B(t)+F(t)$, with analytic matrix coefficients are considered. First, an exact series solution of the problem is obtained. Given a bounded domain $\Omega$ and an admissible error $\epsilon$, a finite analytic-numerical series solution is constructed, so that the error with respect to the exact series solution is uniformly upper bounded by $\epsilon$ in $\Omega$. An iterative procedure for the construction of the approximate solutions is included.

**Keywords**—Sylvester differential equation, Initial value problem, Frobenius method, Accuracy, Gronwall's inequality, Error bound.

## 1. INTRODUCTION

Sylvester matrix differential equations of the form

$$X'(t) = A(t)X(t) + X(t)B(t) + F(t), \qquad X(0) = C, \tag{1}$$

where the coefficient $A(t)$, $B(t)$, and $F(t)$, as well as the unknown $X(t)$ are matrices in $\mathbb{C}^{r\times r}$ are frequent in large-space flexible structures [1], jump linear systems [2], control of linear systems with non-Markovian modal changes [3], or when one uses semidiscretization techniques to solve scalar partial differential equations [4]. For the particular case where the coefficients are real matrices and $B(t)$ is the transposed matrix of $A(t)$, equation (1) becomes the Lyapunov differential equation. An account of examples, properties, and applications of the Lyapunov differential equation may be found in [5].

Several numerical integration methods of solving problem (1) for the case where $A(t)$, $B(t)$, and $F(t)$ are constant matrices, have been given in [6–10]. A modification of the Runge-Kutta method for problem (1) has been proposed in [11]. A method for constructing continuous numerical solutions of (1) has recently been given in [12] using linear unilateral associated problems and one-step matrix methods. However, the method proposed in [9] is expensive from a computational point of view.

Here we consider problem (1), where the coefficients $A(t)$, $B(t)$, and $F(t)$ are $\mathbb{C}^{r\times r}$ valued analytic functions in $|t| < c$, say

$$A(t) = \sum_{n\geq 0} A_n t^n, \qquad B(t) = \sum_{n\geq 0} B_n t^n, \qquad F(t) = \sum_{n\geq 0} F_n t^n, \qquad |t| < c. \tag{2}$$

This work has been partially supported by the Generalitat Valenciana Grant GV1118/93 and the D.G.I.C.Y.T. Grant PB93-0381.

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX

The aim of this paper is to construct analytic-numerical solutions of problem (1) with a prefixed accuracy in a domain $|t| \le A < c$, and its organization is as follows. Section 2 deals with the convergence proof of a series solution of problem (1) in $|t| < c$, under the hypothesis (2). Some important technical lemmas that will be used in the further error analysis are proved in Section 3. In Section 4, we address the following question. How to construct an analytic-numerical finite series solution in $|t| \le A$, whose error with respect to the exact infinite series solution is uniformly upper bounded by a prefixed admissible error $\varepsilon > 0$. An iterative procedure for the construction of such an approximate solution is also included.

Throughout this paper, the norm $\|D\|$ of a matrix $D$ in $\mathbb{C}^{r\times r}$, is the 2-norm of $D$, defined by [13, p. 53]

$$\|D\| = \sup x \neq 0 \frac{\|Dx\|}{\|x\|},$$

where for a vector $y$ in $\mathbb{C}^r$, $\|y\|$ denotes the usual Euclidean norm of $y$. If $x$ is a real number, we denote by $[x]$ its entire part.

## 2. CONVERGENCE OF THE SERIES SOLUTION

In this section, we seek an analytic series solution $X(t)$ of problem (1) of the form

$$X(t) = \sum_{n\ge 0} X_n t^n, \qquad |t| < c, \tag{3}$$

where $X_n$ are matrices in $\mathbb{C}^{r\times r}$ to be determined. Taking formal derivatives in (3), one gets

$$X'(t) = \sum_{n\ge 0} (n+1) X_n t^n. \tag{4}$$

Assuming for the moment that a solution of the form (3) exists, by Merten's theorem for the product of matrix series, by (1) and (3), it follows that

$$A(t)X(t) = \left(\sum_{n\ge 0} A_n t^n\right)\left(\sum_{n\ge 0} X_n t^n\right) = \sum_{n\ge 0}\left(\sum_{k=0}^{n} A_{n-k}X_k\right) t^n, \tag{5}$$

$$X(t)B(t) = \left(\sum_{n\ge 0} X_n t^n\right)\left(\sum_{n\ge 0} B_n t^n\right) = \sum_{n\ge 0}\left(\sum_{k=0}^{n} X_{n-k}B_k\right) t^n. \tag{6}$$

By imposing that $X(t)$ given by (3) satisfies (1), and taking into account (4)–(6), one gets

$$\begin{aligned}\sum_{n\ge 0}(n+1)X_n t^n &= \sum_{n\ge 0}\left(\sum_{k=0}^{n} A_{n-k}X_k\right)t^n + \sum_{n\ge 0}\left(\sum_{k=0}^{n} X_{n-k}B_k\right)t^n + \sum_{n\ge 0} F_n t^n \\ &= \sum_{n\ge 0}\left\{F_n \sum_{k=0}^{n}\left(A_{n-k}X_k + X_{n-k}B_k\right)\right\} t^n.\end{aligned} \tag{7}$$

Equating the coefficients of $t^n$ in (7), it follows that

$$(n+1)X_{n+1} = F_n \sum_{k=0}^{n}\left(A_{n-k}X_k + X_{n-k}B_k\right), \qquad n \ge 0, \quad X_0 = C. \tag{8}$$

Taking norms in (8), one gets

$$(n+1)\|X_{n+1}\| \le \|F_n\| \sum_{k=0}^{n}\left(\|A_{n-k}\|\,\|X_k\| + \|X_{n-k}\|\,\|B_k\|\right). \tag{9}$$

By Cauchy inequalities [14, p. 222], there exists a positive constant $M > 0$ such that

$$\|A_n\| \le \frac{M}{\rho^n}, \qquad \|B_n\| \le \frac{M}{\rho^n}, \qquad \|F_n\| \le \frac{M}{\rho^n}, \qquad 0 < \rho < c, \quad n \ge 0. \tag{10}$$

From (9) and (10), it follows that

$$(n+1)\|X_{n+1}\| \le \frac{M}{\rho^n}\left(1 + 2\sum_{k=0}^{n} \|X_k\| \rho^k\right), \qquad n \ge 0, \quad 0 < \rho < c. \tag{11}$$

Hence,

$$\|X_{n+1}\| \le \frac{M}{(n+1)\rho^n}\left(1 + 2\sum_{k=0}^{n} \|X_k\| \rho^k\right), \qquad n \ge 0, \quad 0 < \rho < c. \tag{12}$$

Let us introduce the sequence of positive numbers $\{\delta_n\}_{n\ge 0}$ defined by $\delta_0 = \|X_0\| = \|C\|$, and $\delta_n$ for $n \ge 0$ is the solution of the equation

$$\delta_{n+1} = \frac{M}{(n+1)\rho^n}\left(1 + 2\sum_{k=0}^{n} \delta_k \rho^k\right), \qquad n \ge 0. \tag{13}$$

By the definition of $\{\delta_n\}_{n\ge 0}$ and (12), using the induction principle, it is easy to prove that

$$\|X_n\| \le \delta_n, \qquad n \ge 0. \tag{14}$$

By (14), in order to prove the convergence of the series (3) where $X_n$ is given by (8), it is sufficient to guarantee the convergence of the numerical series

$$\sum_{n\ge 0} \delta_n t^n, \qquad 0 < |t| < c. \tag{15}$$

By the definition of $\delta_n$, see (13), one gets

$$(n+1)\delta_{n+1} - \rho^{-1} n \delta_n = 2M\delta_n, \qquad n > 0. \tag{16}$$

Hence,

$$\frac{\delta_{n+1}}{\delta_n} = \frac{2M\rho + n}{(n+1)\rho},$$

$$\lim n \to \infty \frac{\delta_{n+1}|t|^{n+1}}{\delta_n |t|^n} = \frac{|t|}{\rho} < 1, \qquad \text{if } |t| < \rho.$$

Thus, (15) converges in $|t| < \rho$, where $\rho$ is any positive number with $0 < \rho < c$, i.e., the series (5) converges in $|t| < c$. This means that the series (3),(8) is not only a formal solution, but the rigorous solution of problem (1).

REMARK 1. Given a point $t_0$ with $0 < |t_0| < c$, by the properties of the analytic functions, the functions $A(t)$, $B(t)$, and $F(t)$ admit a power series development of the form

$$A(t) = \sum_{n\ge 0} A_n(t_0)(t-t_0)^n, \qquad B(t) = \sum_{n\ge 0} B_n(t_0)(t-t_0)^n,$$
$$F(t) = \sum_{n\ge 0} F_n(t_0)(t-t_0)^n, \qquad |t - t_0| < c - |t_0|.$$

If we consider the initial value problem

$$X'(t) = A(t)X(t) + X(t)B(t) + F(t), \qquad X(t_0) = C(t_0), \qquad t_0 \le t < c, \tag{17}$$

by the previous arguments, it is easy to prove that the exact series solution of problem (17) is given by

$$X(t) = \sum_{n \geq 0} X_n(t_0)(t - t_0)^n, \qquad t_0 \leq t < c,$$

$$X_0(t_0) = C(t_0),$$

$$X_{n+1} = \frac{1}{n+1} \left\{ F_n(t_0) + \sum_{k=0}^{n} \left( A_{n-k}(t_0) X_k(t_0) + X_{n-k}(t_0) B_k(t_0) \right) \right\}, \qquad n \geq 0.$$

Also, by Cauchy's inequalities [14, p. 222] applied in the disk $|z - t_0| < c - |t_0|$, one gets

$$\|A_n(t_0)\| \leq \frac{M}{(c - |t_0|)^n}, \qquad \|B_n(t_0)\| \leq \frac{M}{(c - |t_0|)^n},$$
$$\|F_n(t_0)\| \leq \frac{M}{(c - |t_0|)^n}, \qquad n \geq 0,$$

where $M \geq \sup\{\|A(t)\|, \|B(t)\ \|F(t)\ |t - t_0| \leq c - |t_0|\}$. For the sake of clarity in the notation, in the following the coefficients of the power series expansions of $A(t)$, $B(t)$, and $F(t)$ about $t = (j-1)b_1$, $j > 1$, will be denoted by $A_n(j-1)$, $B_n(j-1)$, and $F_n(j-1)$, respectively.

## 3. TECHNICAL LEMMAS

We begin this section with a result that provides an *a priori* error bound of the theoretical solution of problem (1).

LEMMA 1. *Let $A(t)$, $B(t)$, and $F(t)$ be continuous $\mathbb{C}^{r \times r}$ valued functions in $[0, A]$, and let $X(t)$ be the solution of problem (1) in $[0, A]$. Then*

$$\|X(t)\| \leq \left( \|C\| + \int_0^A \|F(s)\| \, ds \right) \exp \left( \int_0^A (\|A(s)\| + \|B(s)\|) \, ds \right), \qquad 0 \leq t \leq A. \tag{18}$$

PROOF. By integrating in (1), one gets that the solution $X(t)$ verifies

$$X(t) - C = \int_0^t \{A(s)X(s) + X(s)B(s) + F(s)\} \, ds. \tag{19}$$

Let $f(t) = \|X(t)\|$ and $g(t) = \|A(t)\| + \|B(t)\|$. Taking norms in (18), it follows that

$$f(t) \leq f(0) + \int_0^t g(s) f(s) \, ds, \qquad 0 \leq t \leq A. \tag{20}$$

By application of Gronwall's inequality [15, p. 95] to (20), one gets (18).

LEMMA 2. *Let $A(t)$, $B(t)$, and $F(t)$ be continuous $\mathbb{C}^{r \times r}$ valued functions, and let $X_1(t)$ be the solution of*

$$X_1'(t) = A(t)X_1(t) + X_1(t)B(t) + F(t), \qquad X_1(\alpha) = P, \quad \alpha \leq t \leq \beta, \tag{21}$$

*and let $X_2(t)$ be the solution of*

$$X_2'(t) = A(t)X_2(t) + X_2(t)B(t) + F(t), \qquad X_2(\alpha) = Q, \quad \alpha \leq t \leq \beta. \tag{22}$$

*Then*

$$\|X_1(t) - X_2(t)\| \leq \|P - Q\| \exp\left((\beta - \alpha)(\|A(t)\| + \|B(t)\|)\right), \qquad \alpha \leq t \leq \beta. \tag{23}$$

PROOF. Let $G(t) = X_1(t) - X_2(t)$. By integrating (21) and (22), one gets

$$\begin{aligned} X_1(t) &= P + \int_\alpha^t \{A(s)X_1(s) + X_1(s)B(s) + F(s)\}\, ds, \qquad \alpha \le t \le \beta, \\ X_2(t) &= Q + \int_\alpha^t \{A(s)X_2(s) + X_2(s)B(s) + F(s)\}\, ds, \qquad \alpha \le t \le \beta, \\ G(t) &= P - Q + \int_\alpha^t \{A(s)\left(X_1(s) - X_2(s)\right) + \left(X_1(s) - X_2(s)\right)B(s)\}\, ds. \end{aligned} \tag{24}$$

Taking norms in (24) and denoting $g(t) = \|G(t)\|$, it follows that

$$g(t) \le \|P - Q\| + \int_\alpha^t (\|A(s)\| + \|B(s)\|)\, g(s)\, ds, \qquad \alpha \le t \le \beta. \tag{25}$$

By application of Gronwall's inequality to (25), see [15, p. 95], one gets (23). For the sake of clarity in the presentation, we state the following result about the summation of double series, whose proof may be found in [4, p. 173].

LEMMA 3. *Given a double sequence $\{a_{ij}\}$, $i \ge 1$, $j \ge 1$, let us suppose that*

$$\sum_{j\ge1} |a_{ij}| = b_i, \qquad i \ge 1, \tag{26}$$

*and that $\sum_{i\ge1} b_i$ converges. Then*

$$\sum_{i\ge1}\sum_{j\ge1} a_{ij} = \sum_{j\ge1}\sum_{i\ge1} a_{ij}. \tag{27}$$

## 4. CONSTRUCTION OF ACCURATE APPROXIMATIONS

In this section, we address the following question under the hypothesis (2). Given a bounded domain $[0, A]$, with $A < c$, and an admissible error $\varepsilon > 0$, how do we construct a finite approximate solution $X(t, \varepsilon)$ defined in $[0, A]$ so that the error with respect to the infinite series solution given in Section 2, be uniformly upper bounded by $\varepsilon$ in $[0, A]$.

Given $\varepsilon > 0$ and $A > 1$, let $h = [A] + 1$, and note that

$$b_1 = \frac{A}{[A] + 1} < 1, \qquad b_1 h = A, \quad h = [A] + 1. \tag{28}$$

Let $b$ and $a$ be positive numbers such that

$$0 < b_1 < b < 1, \qquad A < a < c, \qquad b < b_1 + (a - A), \tag{29}$$

where $b$ is defined by (28) and $c$ by (2). Note that, in this way, the interval $[0, A]$ has been divided in $h$ subintervals $[0, b_1], [b_1, 2b_1], \dots, [(h-1)b_1, A]$.

By the development of Section 2, we know that the exact solution of problem (1) in $[0, b_1]$ is given by

$$\begin{gathered} X_1(t) = \sum_{n\ge0} X_n t^n, \qquad 0 \le t \le b_1, \\ X_0 = C, \qquad X_{n+1} = \frac{1}{n+1}\left\{F_n + \sum_{k=0}^{n} (A_{n-k}X_k + X_{n-k}B_k)\right\}, \qquad n \ge 0. \end{gathered} \tag{30}$$

Now let us consider the truncated series of order $m$ of $X_1(t)$:

$$Y_1(t,m) = \sum_{n=0}^{m} X_n t^n, \qquad 0 \le t \le b_1. \tag{31}$$

For $|t| \le b_1$, it follows that

$$\|X_1(t) - Y_1(t,m)\| = \left\| \sum_{n\ge m+1} X_n t^n \right\| \le \sum_{n\le m+1} \|X_n\| \, b_1^n. \tag{32}$$

Let $\|X_n\| = \varphi_n$, and let $M > 0$ such that

$$\sup 0 \le t \le a \left\{\|A(t)\|, \|B(t)\|, \|F(t)\|, \|X(t)\|\right\} \le M, \tag{33}$$

and recall that by Lemma 1, such a value of $M$ is easy to obtain in terms of the data. By Cauchy's inequalities and Section 2, one gets (see (12))

$$\begin{aligned} \varphi_{n+1} &\le \frac{M}{(n+1)b^n}\left(1 + 2\sum_{k=0}^{n} \varphi_k b^k\right), \qquad n \ge 0, \\ \varphi_n &\le \frac{M}{nb^{n-1}}\left(1 + 2\sum_{k=0}^{n-1} \varphi_k b^k\right), \qquad n \ge 1. \end{aligned} \tag{34}$$

From (32) and (34), one gets

$$\begin{aligned} \|X_1(t) - Y_1(t,m)\| &\le \sum_{n\ge m+1} \left\{ \frac{M}{nb^{n-1}} \left(1 + 2\sum_{k=0}^{n-1} \varphi_k b^k\right)\right\} b_1^n \\ &= Mb \sum_{n\ge m+1} \frac{1}{n}\left(\frac{b_1}{b}\right)^n + \sum_{n\ge m+1} \left(\frac{2M}{nb^{n-1}} \sum_{k=0}^{n-1} \varphi_k b^k\right) b_1^n \\ &\le Mb \sum_{n\ge m+1} \left(\frac{b_1}{b}\right)^n + \sum_{n\ge m+1} \left(\frac{2M}{nb^{n-1}} \sum_{k=0}^{n-1} \varphi_k b^k\right) b_1^n. \end{aligned} \tag{35}$$

By the convergence of the series $\sum_{n\ge m+1} \varphi_n b_1^n$ and Lemma 3, we can write

$$\begin{aligned} \sum_{n\ge m+1} \left(\frac{2M}{nb^{n-1}} \sum_{k=0}^{n-1} \varphi_k b^k\right) b_1^n = {}& 2Mb\varphi_0 \sum_{j\ge 1} \frac{1}{m+j}\left(\frac{b_1}{b}\right)^{m+j} \\ & + 2Mb^2\varphi_1 \sum_{j\ge 1} \frac{1}{m+j}\left(\frac{b_1}{b}\right)^{m+j} + 2Mb^3\varphi_2 \sum_{j\ge 1} \frac{1}{m+j}\left(\frac{b_1}{b}\right)^{m+j} + \cdots \\ & + 2Mb^{m+1}\varphi_m \sum_{j\ge 1} \frac{1}{m+j}\left(\frac{b_1}{b}\right)^{m+j} \\ & + 2Mb^{m+2}\varphi_{m+1} \sum_{j\ge 2} \frac{1}{m+j}\left(\frac{b_1}{b}\right)^{m+j} \\ & + 2Mb^{m+3}\varphi_{m+2} \sum_{j\ge 3} \frac{1}{m+j}\left(\frac{b_1}{b}\right)^{m+j} + \cdots \\ & + 2Mb^{m+l}\varphi_{m+l-1} \sum_{j\ge l} \frac{1}{m+j}\left(\frac{b_1}{b}\right)^{m+j} + \cdots . \end{aligned}$$

Hence,

$$
\begin{aligned}
\sum_{n\geq m+1}\left(\frac{2M}{nb^{n-1}}\sum_{k=0}^{n-1}\varphi_k b^k\right)b_1^n &= \left\{2Mb\sum_{j\geq 1}\frac{1}{m+j}\left(\frac{b_1}{b}\right)^{m+j}\right\}(\varphi_0+b\varphi_1+\cdots+b^m\varphi_m)\\
&\quad + 2Mb\left\{b^{m+1}\varphi_{m+1}\sum_{j\geq 2}\frac{1}{m+j}\left(\frac{b_1}{b}\right)^{m+j}\right.\\
&\quad + b^{m+2}\varphi_{m+2}\sum_{j\geq 3}\frac{1}{m+j}\left(\frac{b_1}{b}\right)^{m+j}+\cdots\\
&\quad \left. + b^{m+l}\varphi_{m+l}\sum_{j\geq l}\frac{1}{m+j}\left(\frac{b_1}{b}\right)^{m+j}+\cdots\right\}.
\end{aligned}
\tag{36}
$$

As for $l \geq 1$, one gets

$$
\begin{aligned}
\sum_{j\geq l}\left(\frac{b_1}{b}\right)^{m+j} &= \left(\frac{b_1}{b}\right)^{m}\sum_{j\geq l}\left(\frac{b_1}{b}\right)^{j} = \left(\frac{b_1}{b}\right)^{m}\frac{(b_1/b)^l}{1-b_1/b}\\
&\leq \left(\frac{b_1}{b}\right)^{m}\frac{b_1/b}{1-b_1/b} = \frac{b_1}{b-b_1}\left(\frac{b_1}{b}\right)^{m},\\
\sum_{j\geq l}\frac{1}{m+j}\left(\frac{b_1}{b}\right)^{m+j} &\leq \sum_{j\geq l}\left(\frac{b_1}{b}\right)^{m+j},
\end{aligned}
$$

from (36), it follows that

$$
\begin{aligned}
\sum_{n\geq m+1}\left(\frac{2M}{nb^{n-1}}\sum_{k=0}^{n-1}\varphi_k b^k\right)b_1^n &\leq \frac{2Mb\,(b_1/b)^m}{1-b_1/b}\left[\sum_{n=0}^{m}b^n\varphi_n + \sum_{n\geq m+1}\varphi_n b^n\left(\frac{b_1}{b}\right)^{n-m+1}\right]\\
&\leq \frac{2Mb\,(b_1/b)^m}{1-b_1/b}\sum_{n\geq 0}b^n\varphi_n.
\end{aligned}
\tag{37}
$$

By Cauchy's inequality [14, p. 222] and (33), it follows that

$$
\|\varphi_n\| \leq \frac{M}{a^n}, \qquad n \geq 0, \tag{38}
$$

where $M$ is given by (33). Since

$$
\sum_{n\geq m+1}\left(\frac{b_1}{b}\right)^{n} = \frac{(b_1/b)^{m+1}}{1-b_1/b}, \tag{39}
$$

by (35),(37)–(39), it follows that

$$
\begin{aligned}
&\|X_1(t)-Y_1(t,m)\|\\
&\quad\leq \frac{Mb}{1-b_1/b}\left\{(b_1/b)^{m+1}+\frac{2M(b_1/b)^m}{1-b/a}\right\} \leq \frac{Mb}{1-b_1/b}\left(1+\frac{2M}{1-b/a}\right)\left(\frac{b_1}{b}\right)^{m}.
\end{aligned}
\tag{40}
$$

Let us suppose for a moment that we choose the first positive integer $m_1$ such that

$$
\left(\frac{b_1}{b}\right)^{m_1} < \frac{\varepsilon\,(1-b_1/b)}{3Mb\,(1+(2M/1-(b/a)))\,[h+(h-1)e^{Lb_1}]}, \tag{41}
$$

where

$$L = \max\{\|A(s)\| + \|B(s)\|;\ 0 \le s \le A\}. \tag{42}$$

Then from (40),(41), it follows that

$$\|X_1(t) - Y_1(t, m_1)\| \le \frac{\epsilon}{3\left[h + (h-1)e^{Lb_1}\right]}, \qquad |t| \le b_1. \tag{43}$$

Note that $m_1$ can be determined taking the first positive integer $m_1$ verifying

$$m_1 > \frac{\ln\left\{\dfrac{\epsilon\left(1 - \frac{b_1}{b}\right)}{3Mb\left(1 + \frac{2M}{1-\frac{b}{a}}\right)\left[h + (h-1)e^{Lb_1}\right]}\right\}}{\ln \frac{b_1}{b}}. \tag{44}$$

Now let us consider the initial value problem in $[b_1, 2b_1]$:

$$X'(t) = A(t)X(t) + X(t)B(t) + F(t), \qquad X(b_1) = Y_1(b_1, m_1). \tag{45}$$

By application of the method developed in Section 2, and taking into account Remark 1, the solution of (48) can be written in the form

$$\begin{aligned} X_2(t) &= \sum_{n\ge 0} X_n(1)\,(t-b_1)^n, \qquad b_1 \le t \le 2b_1\ , \\ X_0(1) &= Y_1(b_1, m_1), \\ X_{n+1}(1) &= \frac{1}{n+1}\left\{F_n(1) + \sum_{k=0}^{n}\left(A_{n-k}(1)X_k(1) + X_{n-k}(1)B_k(1)\right)\right\}, \end{aligned} \tag{46}$$

where $F_n(1)$, $B_n(1)$, and $A_n(1)$ are the Taylor coefficients of the power series expansions of $F(t)$, $B(t)$, and $A(t)$, respectively, about $t = b_1$. Note that from (33), and Cauchy's inequalities applied to $F(t)$, $B(t)$, and $A(t)$ in the disk $|z - b_1| < a - b_1 = a_1$, it follows that

$$\|A_n(1)\| \le \frac{M}{a_1^n}, \qquad \|F_n(1)\| \le \frac{M}{a_1^n}, \qquad \|B_n(1)\| \le \frac{M}{a_1^n}, \qquad n \ge 0. \tag{47}$$

If we truncate the series (46) by its $m_1^{\text{th}}$ partial sum

$$Y_2(t, m_1) = \sum_{n=0}^{m_1} X_n(1)\,(t-b_1)^n, \qquad b_1 \le t \le 2b_1, \tag{48}$$

then by Remark 1, (47) and (44) replacing $a$ by $a_1 = a - b$, if $m_2$ is the first positive integer verifying

$$m_2 > \frac{\ln\left\{\dfrac{\epsilon\left(1 - \frac{b_1}{b}\right)}{3Mb\left(1 + \frac{2M}{1-\frac{b}{a_1}}\right)\left[h + (h-1)e^{Lb_1}\right]}\right\}}{\ln \frac{b_1}{b}}, \tag{49}$$

it follows that

$$\|X_2(t) - Y_2(t, m_2)\| \le \frac{\epsilon}{h + (h-1)e^{Lb_1}}, \qquad b_1 \le t \le 2t_1, \tag{50}$$

where $L$ is given by (42). Inductively, going on from $[b_1, 2b_1]$ to $[2b_1, 3b_1]$, and so on, if we denote by $Y_{j-1}(t, m_{j-1})$ the approximation of

$$\begin{aligned} X_j(t) &= \sum_{n\ge 0} X_n(j-1)\,\left(t-(j-1)b_1\right)^n, \qquad (j-1)b_1 \le t \le jb_1, \\ X_0(j-1) &= Y_{j-1}\left(jb_1, m_{j-1}\right), \\ X_{n+1}(j-1) &= \frac{1}{n+1}\left\{F_n(j-1) + \sum_{k=0}^{m_j}\left(A_{n-k}(j-1)X_k(j-1) + X_{n-k}(j-1)B_k(j-1)\right)\right\}, \end{aligned} \tag{51}$$

where $m_j$ is the first positive integer verifying

$$m_2 > \frac{\ln\left\{\dfrac{\epsilon\left(1-\frac{b_1}{b}\right)}{3Mb\left(1+\frac{2M}{1-\frac{b}{a_{j-1}}}\right)\left[h+(h-1)e^{Lb_1}\right]}\right\}}{\ln\frac{b_1}{b}}, \tag{52}$$

and

$$a_{j-1} = a-(j-1)b_1, \qquad 1 \le j \le h, \tag{53}$$

by the previous arguments, the truncation series of order $m_j$ of $X_j(t)$, defined by

$$Y_j(t,m_j) = \sum_{n=0}^{m_j} X_n(j-1)\,(t-(j-1)b_1)^n \tag{54}$$

satisfies

$$\|X_j(t) - Y_j(t,m_j)\| \le \frac{\epsilon}{h+(h-1)e^{Lb_1}}, \qquad (j-1)b_1 \le t \le jb_1. \tag{55}$$

Note that in order to select $m_j$, we have used that the matrix coefficients $A_n(j-1)$, $B_n(j-1)$, $F_n(j-1)$ of the power series expansions of $A(t)$, $B(t)$, and $F(t)$, respectively, verify

$$\|A_n(j-1)\| \le \frac{M}{a_{j-1}^n}, \qquad \|B_n(j-1)\| \le \frac{M}{a_{j-1}^n}, \qquad \|F_n(j-1)\| \le \frac{M}{a_{j-1}^n}, \qquad n \ge 0. \tag{56}$$

Thus, the approximate solution $X(t,\epsilon)$ defined by

$$X(t,\epsilon) = Y_j(t,m_j), \qquad (j-1)b_1 \le t \le jb_1, \qquad 1 \le j \le h, \tag{57}$$

where $Y_1(t,m_1)$ by (30),(31), with $m = m_1$, and for $1 \le j \le h$, $Y_j(t,m_j)$ is defined by (51), being $m_j$ the first positive integer verifying (52).

Note that in the interval $[0,b_1]$, the approximation error between the exact series solution $X(t)$ given by (30) and $X(t,\epsilon)$ defined by (31), is the truncation error bounded by (43). However, in each subinterval $[(j-1)b_1, jb_1]$ for $2 \le j \le h$, we have two contributions to the error; the one coming from the consideration of an approximate initial condition at the $(j-1)b_1$, and the truncation error when one considers the $m_j^{\text{th}}$ partial sum instead the infinite series. Hence, for any $t \in [0,hb_1] = [0,A]$, by the previous comments and Lemma 2, one gets

$$\begin{aligned}\|X(t)-X(t,\epsilon)\| &\le \frac{\epsilon}{h+(h-1)e^{Lb_1}} + \sum_{j=1}^{h-1}\left[\frac{\epsilon}{h+(h-1)e^{Lb_1}} + \frac{\epsilon e^{Lb_1}}{h+(h-1)e^{Lb_1}}\right] \\ &= \frac{h\epsilon}{h+(h-1)e^{Lb_1}} + \frac{(h-1)\epsilon}{h+(h-1)e^{Lb_1}} \\ &= \frac{\epsilon}{h+(h-1)e^{Lb_1}}\left[h+(h-1)e^{Lb_1}\right] = \epsilon.\end{aligned} \tag{58}$$

Note that if $A < 1$, then taking $b_1 = A$, the approximate solution $X(t,\epsilon) = Y_1(t,m)$ defined by (31), satisfies also

$$\|X(t) - X(t,\epsilon)\| < \epsilon, \qquad 0 \le t \le b_1 = A. \tag{59}$$

Note that by the selection of $b$ given by (29), even in the last subinterval of the construction procedure $[(h-1)b_1, hb_1 = A]$, the distance $a-(h-1)b_1 = a_{h-1} = b_1 + (a-A) > b$. Thus, the series

$$\sum_{n\ge 0}\left(\frac{b}{a_{h-1}}\right)^n < +\infty,$$

and (40) holds, replacing $a$ by $a_{j-1}$, for $2 \le j \le h$.

Summarizing, the following result has been established.

THEOREM 1. *Let us consider the initial value problem (1) in the interval* $[0, A]$, *where* $0 < A < c$ *and the matrix coefficients* $A(t)$, $B(t)$, *and* $F(t)$, *are* $\mathbb{C}^{r\times r}$ *valued analytic functions in* $|t| < c$. *Let* $X(t)$ *be the exact series solution given by (3),(8). Given an admissible error* $\epsilon > 0$, *the following procedure provides the construction of an approximate solution* $X(t, \epsilon)$, *whose error with respect to* $X(t)$ *is uniformly bounded by* $\epsilon$ *in* $[0, A]$:

$$\|X(t) - X(t, \epsilon)\| < \epsilon. \tag{60}$$

CASE 1. $A < 1$, $\epsilon > 0$. Let $b_1 = A$, $h = 1$, $b = 1$, $a > 1$. Compute $M$ satisfying (33) using (18). Let $m_1$ be the first positive integer verifying (44). Let $X_0$ and $X_{n+1}$ for $0 \le n \le m_1 - 1$ be defined by (30). Then $X(t, \epsilon) = Y_1(t, m)$ defined by (31) is the approximate solution of problem (1) in $[0, A]$ verifying (54).

CASE 2. $A > 1$, $\epsilon > 0$. Let $h = [A] + 1$, $b_1 = A/([A] + 1)$, $A = b_1 h$, $A < a < c$, $b < b_1 + a - A$. Compute $M$ satisfying (33) using (18). Let $m_1$ be the first positive integer verifying (44). Let us compute $A_n = A_n(0)$, $B_n = B_n(0)$, and $F_n = F_n(0)$ for $0 \le n \le m_1$, given by (2). Let $Y_1(t, m_1)$ be defined by (30),(31). Let $j = 2$ and $a_1 = a - b_1$. Let $m_2$ be the first positive integer verifying (49) and $Y_2(t, m_2)$ be defined by (54) with $j = 2$ and $X_n(1)$ given by (51). Inductively, for each $j > 2$, given $Y_{j-1}(t, m_{j-1})$ defined by (51)–(54), for $j - 1$ instead of $j$, let $m_j$ be the first positive integer verifying (52), where $a_{j-1}$ is defined by (53). Let $Y_j(t, m_j)$ be defined by (51)–(54) in $(j - 1)b_1 \le t \le jb_1$. For $j = h$, construct $Y_h(t, m_h)$ by (51)–(54), where $m_h$ is the first positive integer verifying (52) with $a_{h-1}$ defined by (53). Then $X(t, \epsilon)$ defined by (57) is the required approximate solution of problem (1) in $[0, A]$ verifying (59).

## REFERENCES


1. M. Balas, Trends in large space structure control theory: Fondest hopes, wildest dreams, *IEEE Trans. Automatic Control* **27**, 522–535 (1982).
2. M. Mariton, *Jump Linear Systems in Automatic Control*, Marcel Dekker, New York 1990.
3. D.D. Sworder, Control of a linear system with non-Markovian modal changes, *J. Economic Dynamics and Control* **2**, 233–240 (1980).
4. K. Rektorys, *The Method of Discretization in Time and Partial Differential Equations*, Reidel, Dordrecht 1982.
5. Z. Gajic, M. Tair and J. Qureshi, *Lyapunov Matrix Equation in System Stability and Control*, Academic Press, New York, (1995).
6. S.M. Serbin and C.A. Serbin, A time-stepping procedure for $X' = A_1X + XA_2 + D$; $X(0) = C$, *IEEE Trans. Autom. Control* **25**, 1138–1141 (1980).
7. W. Hoskins, D. Meek and D. Walton, The numerical solution of $\dot{X} = A_1X + XA_2 + D$, $X(0) = C$, *IEEE Trans. Automatic Control* **22**, 881–882 (1977).
8. A. Barraud, A new numerical solution of $\dot{X} = A_1X + XA_2 + D$, $X(0) = C$, *IEEE Trans. Automatic Control* **22**, 976–977 (1977).
9. E. Davison, The numerical solution of $\dot{X} = A_1X + XA_2 + D$, $X(0) = C$, *IEEE Trans. Aut. Control* **20**, 566–567 (1975).
10. L. Jódar, Solving a class of generalized Lyapunov operator differential equation without the exponential operator function, *Publicacions Matemàtiques* **34**, 25–35 (1990).
11. M. Subrahmanyam, On a numerical method of solving the Lyapunov and Sylvester equations, *Int. J. Control* **43**, 433–439 (1986).
12. L. Jódar and E. Ponsoda, Computing continuous numerical solutions of matrix differential equations, *Computers Math. Applic.* **29** (4), 73–84 (1995).
13. G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD 1985.
14. J. Dieudonné, *Foundations of Modern Analysis*, Academic Press, New York, (1960).
15. T.M. Flett, *Differential Analysis*, Cambridge University Press, New York, (1980).
16. W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill New York 1964.