Approximating and computing nonlinear matrix differential models*

E. Defez⋆, M.M. Tung⋆, J.J. Ibáñez♮, J. Sastre†

⋆ Instituto de Matemática Multidisciplinar

♮ Instituto de Instrumentación para Imagen Molecular

† Instituto de Telecomunicaciones y Aplicaciones Multimedia

Universitat Politècnica de València, Spain

{edefez, mtung}@imm.upv.es, jjibanez@dsic.upv.es, jorsasma@iteam.upv.es

Abstract

Differential matrix models are an essential ingredient of many important scientific and engineering applications. In this work, we propose a procedure to represent the solutions of first-order matrix differential equations $Y'(x) = f(x, Y(x))$ with approximate matrix splines. For illustration of the method, we choose one scalar example, a simple vector model, and finally a Sylvester matrix differential equation as test.

Keywords and phrases. Higher-order matrix splines, first-order matrix differential equations.

1 Introduction

In this paper we propose a novel algorithm to tackle matrix differential equations of the first order. Matrix differential models are relevant for the description of many phenomena in physics and engineering, ranging from such diverse applications as control theory to game theory [1]. In particular, we will develop in this work a method for the numerical integration of first-order matrix differential

*Acknowledgments. This work has been supported by grant PAID-06-07/3283 from the Universitat Politècnica de València, Spain.

equations with initial conditions. For different examples of this class of problems, we also refer to Ref. [2].

In their seminal work, Loscalzo and Talbot introduce spline function approximations for solutions of scalar differential equations [3]. These spline solutions $S(x)$ are of degree $m = 2, 3$ and continuity class $\mathcal{C}^{m-1}$. Recently, this method has been used in the resolution of other scalar problems as discussed in Ref. [4]. The corresponding generalizations to the matrix framework have been carried out in Refs. [5, 6].

Unfortunately, as detected by Loscalzo and Talbot, their scalar procedure is divergent when higher-order spline functions are used [3, p. 444–445]. They have explicitly shown by numerical computations that the equation $y' = y, y(0) = 1$ contains noticeable divergences for splines of order $m > 3$. However, our new method avoids these problems with divergences for splines $S(x)$ of order m but only require them to be of differentiability class $\mathcal{C}^1$.

Throughout this work, we will adopt the notation for norms and matrix cubic splines as in the previous work [5] and common in matrix calculus. Following this nomenclature, we recall that the 2-norm of a rectangular $r \times s$ matrix $A \in \mathbb{C}^{r\times s}$ is

$$\|A\| = \sup_{z\neq 0} \frac{\|Az\|}{\|z\|},$$

where, as usual, for a vector $z \in \mathbb{C}^s$ the Euclidean norm is $\|z\| = (z^t z)^{\frac{1}{2}}$. Similarly, the 1-norm is given by $\|z\|_1 = \sum_{i=1}^{s} |z_i|$.

The Kronecker product $A \otimes B$ of $A = (a_{ij}) \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{r\times s}$ is defined by the following block matrix

$$A \otimes B \; = \; \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix}.$$

The column-vector operator on a matrix $A \in \mathbb{C}^{m\times n}$ is denoted by

$$\mathbf{vec}(A) = \begin{pmatrix} \mathbf{A}_{\bullet 1} \\ \vdots \\ \mathbf{A}_{\bullet n} \end{pmatrix}, \text{ where } \mathbf{A}_{\bullet k} = \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix}.$$

Here and in the following, we denote vectors and vector-valued functions by boldface characters.

If $Y=(y_{ij})\in\mathbb{C}^{p\times q}$ and $X=(x_{ij})\in\mathbb{C}^{m\times n}$, then the derivative of a matrix with respect to a matrix is defined by [11, p.62 and 81]:

$$\frac{\partial Y}{\partial X}=\begin{pmatrix} \dfrac{\partial Y}{\partial x_{11}} & \cdots & \dfrac{\partial Y}{\partial x_{1n}} \\ \vdots & & \vdots \\ \dfrac{\partial Y}{\partial x_{m1}} & \cdots & \dfrac{\partial Y}{\partial x_{mn}} \end{pmatrix}, \text{ where } \frac{\partial Y}{\partial x_{rs}}=\begin{pmatrix} \dfrac{\partial y_{11}}{\partial x_{rs}} & \cdots & \dfrac{\partial y_{1q}}{\partial x_{rs}} \\ \vdots & & \vdots \\ \dfrac{\partial y_{p1}}{\partial x_{rs}} & \cdots & \dfrac{\partial y_{pq}}{\partial x_{rs}} \end{pmatrix}.$$

If $X\in\mathbb{C}^{m\times n}, Y\in\mathbb{C}^{n\times v}, Z\in\mathbb{C}^{p\times q}$, then the following rule for the derivative of a matrix product with respect to another matrix applies [11, p.84]:

$$\frac{\partial XY}{\partial Z} \;=\; \frac{\partial X}{\partial Z}\left[I_q\otimes Y\right] \;+\; \left[I_p\otimes X\right]\frac{\partial Y}{\partial Z}\,, \tag{1.1}$$

where I_q and I_p denote the identity matrices of dimensions q and p, respectively. If $X\in\mathbb{C}^{m\times n}, Y\in\mathbb{C}^{u\times v}, Z\in\mathbb{C}^{p\times q}$, the following chain rule [11, p.88] is valid :

$$\frac{\partial Z}{\partial X} \;=\; \left[\frac{\partial\left[\mathbf{vec}(Y)\right]^t}{\partial X}\otimes I_p\right]\left[I_n\otimes\frac{\partial Z}{\partial\left[\mathbf{vec}(Y)\right]}\right]. \tag{1.2}$$

This paper is organized as follows. In Section 2, we give a description of the proposed method and give details of the corresponding procedure. Section 3 concludes the discussion with some numerical examples for the scalar, vector and matrix cases, respectively.

2 Description of the method

As usual, let us consider the following first-order matrix problem

$$\left.\begin{array}{rcl} Y'(x) &=& f(x,Y(x)) \\ \\ Y(a) &=& Y_a \end{array}\right\}\;,\;\; a\le x\le b, \tag{2.1}$$

where the unknown matrix is $Y(x)\in\mathbb{R}^{r\times q}$ with initial condition $Y_a\in\mathbb{R}^{r\times q}$. The matrix-valued function $f:[a,b]\times\mathbb{R}^{r\times q}\to\mathbb{R}^{r\times q}$ is of differentiability class $f\in\mathcal{C}^s\left(T\right)$, $s\ge 1$, with

$$T=\left\{(x,Y);\; a\le x\le b,\; Y\in\mathbb{R}^{r\times q}\right\}, \tag{2.2}$$

and f fulfills the global Lipschitz's condition

$$\|f(x, Y_1) - f(x, Y_2)\| \leq L\|Y_1 - Y_2\|,\ a \leq x \leq b, Y_1, Y_2 \in \mathbb{R}^{r\times q} \tag{2.3}$$

to guarantee the existence and uniqueness of the continuously differentiable solution $Y(x)$ of problem (2.1), see Ref. [7, p.99].

The partition of the interval $[a, b]$ shall be given by

$$\Delta_{[a,b]} = \{a = x_0 < x_1 < \ldots < x_n = b\},\ x_k = a + kh,\ k = 0, 1, \ldots, n, \tag{2.4}$$

where n is a positive integer with the corresponding step size $h = (b - a)/n$. We will construct in each subinterval $[a + kh, a + (k + 1)h]$ a matrix spline $S(x)$ of order $m \in \mathbb{N}$ with $1 \leq m \leq s$, where s is the order of the differentiability class of f. This will approximate the solution of problem (2.1) so that $S(x) \in C^1([a, b])$.

In the first interval $[a, a + h]$, we define the matrix spline as

$$\begin{aligned} S_{\big|_{[a,a+h]}}(x) &= Y(a) + Y'(a)(x-a) + \frac{1}{2!}Y''(a)(x-a)^2 + \frac{1}{3!}Y^{(3)}(a)(x-a)^3 \\ &+ \cdots + \frac{1}{(m-1)!}Y^{(m-1)}(a)(x-a)^{m-1} + \frac{1}{m!}A_0(x-a)^m, \end{aligned} \tag{2.5}$$

where $A_0 \in \mathbb{R}^{r\times q}$ is a matrix parameter to be determined. It is straightforward to check

$$S_{\big|_{[a,a+h]}}(a) = Y(a), \quad S'_{\big|_{[a,a+h]}}(a) = Y'(a) = f(a, Y(a)),$$

and therefore the spline satisfies the differential equation Eq. (2.1) at $x = a$.

We must obtain the values $Y''(a), Y^{(3)}(a), \ldots, Y^{(m-1)}(a)$, and A_0 in order to determine the matrix spline (2.5). To compute the second-order derivative $Y''(x)$, we follow the procedure given in Ref. [6] and use the nomenclature as already outlined in the introduction. We then obtain

$$\begin{aligned} Y''(x) &= \frac{\partial f(x, Y(x))}{\partial x} + \left[[\mathbf{vec}\ f(x, Y(x))]^T \otimes I_r\right] \frac{\partial f(x, Y(x))}{\partial\, \mathbf{vec}\ Y(x)} \\ &= g_1(x, Y(x)), \end{aligned} \tag{2.6}$$

where $g_1 \in \mathcal{C}^{s-1}(T)$. We are now in the position to evaluate $Y''(a) = g_1(a, Y(a))$ using (2.6). Similarly, we can assume that $f \in \mathcal{C}^s(T)$ for $s \geq 2$. Then, the second partial derivatives of f exist and are continuous. This yields the third derivative:

$$Y^{(3)}(x) = \frac{\partial^2 f(x, Y(x))}{\partial x^2} + \left([\mathbf{vec}\ f(x, Y(x))]^T \otimes I_r\right) \frac{\partial}{\partial x}\left(\frac{\partial f(x, Y(x))}{\partial\, \mathbf{vec}\ Y(x)}\right)$$

$$
\begin{aligned}
&+ \left(\frac{\partial \left[\mathbf{vec}\ f(x,Y(x))\right]^T}{\partial x} \otimes I_r\right) \frac{\partial f(x,Y(x))}{\partial\, \mathbf{vec}\ Y(x)} \\
&+ \left(\left[\mathbf{vec}\ f(x,Y(x))\right]^T \otimes I_r\right) \frac{\partial}{\partial\, \mathbf{vec}\ Y(x)} \left(\frac{\partial f(x,Y(x))}{\partial x}\right) \\
&+ \left(\left[\mathbf{vec}\ f(x,Y(x))\right]^T \otimes I_r\right) \left(\frac{\partial \left[\mathbf{vec}\ f(x,Y(x))\right]^T}{\partial\, \mathbf{vec}\ Y(x)} \otimes I_r\right) \frac{\partial f(x,Y(x))}{\partial\, \mathbf{vec}\ Y(x)} \\
&+ \left(\left[\mathbf{vec}\ f(x,Y(x))\right]^T \otimes I_r\right) \left(\left[\mathbf{vec}\ f(x,Y(x))\right]^T \otimes I_{r^2q}\right) \frac{\partial^2 f(x,Y(x))}{\left(\partial\, \mathbf{vec}\ Y(x)\right)^2} \\
&= g_2\left(x,Y(x)\right) \in \mathcal{C}^{s-2}\left(T\right). \qquad (2.7)
\end{aligned}
$$

Now we can evaluate $Y^{(3)}(a) = g_2\left(a,Y(a)\right)$ using (2.7). For all higher-order derivatives $Y^{(4)}(x),\dots,Y^{(m-1)}(x)$ we proceed in like manner and calculate

$$
\left.
\begin{array}{rcl}
Y^{(4)}(x) & = & g_3\left(x,Y(x)\right) \in \mathcal{C}^{s-3}\left(T\right) \\
 & \vdots & \\
Y^{(m-1)}(x) & = & g_{m-2}\left(x,Y(x)\right) \in \mathcal{C}^{s-(m-2)}\left(T\right)
\end{array}
\right\}. \qquad (2.8)
$$

A list of all these derivatives can be easily established by employing standard computer algebra systems. Substituting $x = a$ in (2.8), one gets $Y^{(4)}(a)$, …, $Y^{(m-1)}(a)$. In summary, all matrix parameters of the spline which were to be determined are known, except for A_0. To determine A_0, we suppose that (2.5) is a solution of problem (2.1) at $x = a+h$, which gives

$$
S'_{\big|_{[a,a+h]}}(a+h) = f\left(a+h, S_{\big|_{[a,a+h]}}(a+h)\right). \qquad (2.9)
$$

Next, we obtain from (2.9) the matrix equation with only one unknown A_0:

$$
\begin{aligned}
A_0 = \tfrac{(m-1)!}{h^{m-1}} \Big[& f\left(a+h, Y(a) + Y'(a)h + \dots + \tfrac{h^{m-1}}{(m-1)!}Y^{(m-1)}(a) + \tfrac{h^m}{m!}A_0\right) \\
& - Y'(a) - Y''(a)h - \tfrac{1}{2}Y^{(3)}(a)h^2 + \dots + \tfrac{1}{(m-2)!}Y^{(m-1)}(a)h^{m-2}\Big]. \qquad (2.10)
\end{aligned}
$$

Assuming that the implicit matrix equation (2.10) has only one solution A_0, the matrix spline (2.5) is totally determined in the interval $[a, a+h]$.

In the following interval $[a+h, a+2h]$, the matrix spline takes the form

$$
S_{\big|_{[a+h,a+2h]}}(x) = S_{\big|_{[a,a+h]}}(a+h) + \overline{Y'(a+h)}(x-(a+h)) +
$$

$$\frac{1}{2!}\overline{Y''(a+h)}(x-(a+h))^2+\cdots+\frac{1}{(m-1)!}\overline{Y^{(m-1)}(a+h)}(x-(a+h))^{m-1}$$
$$+\frac{1}{m!}A_1(x-(a+h))^m, \tag{2.11}$$

where

$$\overline{Y'(a+h)} = f\left(a+h, S_{\big|_{[a,a+h]}}(a+h)\right), \tag{2.12}$$

and $\overline{Y''(a+h)}, \ldots, \overline{Y^{(m-1)}(a+h)}$ are the similar results obtained after evaluating the respective derivatives of $Y(x)$ using $S_{\big|_{[a,a+h]}}(a+h)$ in (2.6)–(2.8). In more compact form, we may write

$$\begin{aligned} \overline{Y''(a+h)} &= g_1\left(a+h, S_{\big|_{[a,a+h]}}(a+h)\right), \\ &\vdots \\ \overline{Y^{(m-1)}(a+h)} &= g_{m-2}\left(a+h, S_{\big|_{[a,a+h]}}(a+h)\right). \end{aligned} \tag{2.13}$$

Note that matrix spline $S(x)$ defined by (2.5) and (2.11) is of differentiability class $\mathcal{C}^1([a,a+h]\cup[a+h,a+2h])$, contrary to the splines introduced by Loscalzo and Talbot [3], which were of class $\mathcal{C}^{m-1}([a,a+h]\cup[a+h,a+2h])$. By construction, spline (2.11) satisfies the differential equation (2.1) at $x=a+h$. and all of its coefficients are determined with the exception of $A_1\in\mathbb{R}^{r\times q}$.

The value of A_1 can be found by taking the spline (2.11) as a solution of (2.1) at point $x=a+2h$:

$$S'_{\big|_{[a+h,a+2h]}}(a+2h) = f\left(a+2h, S_{\big|_{[a+h,a+2h]}}(a+2h)\right).$$

An expansion yields the matrix equation with the only unknown A_1:

$$\begin{aligned} A_1 &= \frac{(m-1)!}{h^{m-1}}\left[f\left(a+2h, S_{\big|_{[a,a+h]}}(a+h)+\overline{Y'(a+h)}h+\frac{h^2}{2!}\overline{Y''(a+h)}+\right.\right. \\ &+ \left.\cdots+\frac{h^{m-1}}{(m-1)!}\overline{Y^{(m-1)}(a+h)}+\frac{h^m}{m!}A_1\right)-\overline{Y'(a+h)}-\overline{Y''(a+h)}h \\ &- \left.\cdots-\frac{1}{(m-2)!}\overline{Y^{(m-1)}(a+h)}h^{m-2}\right]. \end{aligned} \tag{2.14}$$

Let us assume that the matrix equation (2.14) has only one solution A_1. This way the spline is totally determined in the interval $[a+h,a+2h]$.

Iterating this process, we can construct the matrix spline approximation taking $[a+(k-1)h, a+kh]$ as the last subinterval. For the succeeding subinterval $[a+kh, a+(k+1)h]$, we define the corresponding matrix spline as

$$\begin{aligned}
&S_{\big|_{[a+kh,a+(k+1)h]}}(x) = S_{\big|_{[a+(k-1)h,a+kh]}}(a+kh) + \overline{Y'(a+kh)}(x-(a+kh)) \\
&+\tfrac{1}{2!}\overline{Y''(a+kh)}(x-(a+kh))^2 + \cdots + \\
&\tfrac{1}{(m-1)!}\overline{Y^{(m-1)}(a+kh)}(x-(a+kh))^{m-1} + \tfrac{1}{m!}A_k(x-(a+kh))^m,
\end{aligned} \tag{2.15}$$

where

$$\overline{Y'(a+kh)} = f\left(a+kh, S_{\big|_{[a+(k-1)h,a+kh]}}(a+kh)\right), \tag{2.16}$$

and in a similar manner one abbreviates

$$\begin{aligned}
\overline{Y''(a+kh)} &= g_1\left(a+kh, S_{\big|_{[a+(k-1)h,a+kh]}}(a+kh)\right), \\
&\vdots \\
\overline{Y^{(m-1)}(a+kh)} &= g_{m-2}\left(a+kh, S_{\big|_{[a+(k-1)h,a+kh]}}(a+kh)\right).
\end{aligned} \tag{2.17}$$

With this definition, the matrix spline $S(x) \in \mathcal{C}^1\left(\bigcup_{j=0}^{k}[a+jh, a+(j+1)h]\right)$ fulfills the differential equation (2.1) at point $x = a+kh$. As an additional requirement, we assume that $S_{\big|_{[a+kh,a+(k+1)h]}}(x)$ satisfies (2.1) at point $x = a+(k+1)h$:

$$S'_{\big|_{[a+kh,a+(k+1)h]}}(a+(k+1)h) = f\left(a+(k+1)h, S_{\big|_{[a+kh,a+(k+1)h]}}(a+(k+1)h)\right),$$

and expanding this expression gives

$$\begin{aligned}
A_k \;=\; &\tfrac{(m-1)!}{h^{m-1}}\Big[f\Big(a+(k+1)h, S_{\big|_{[a+kh,a+(k+1)h]}}(a+(k+1)h) + \overline{Y'(a+kh)}h \\
&+ \cdots + \tfrac{h^{m-1}}{(m-1)!}\overline{Y^{(m-1)}(a+kh)} + \tfrac{h^m}{m!}A_1\Big) - \overline{Y'(a+kh)} - \overline{Y''(a+kh)}h \\
&- \cdots - \tfrac{h^{m-2}}{(m-2)!}\overline{Y^{(m-1)}(a+kh)}\Big].
\end{aligned} \tag{2.18}$$

Observe that the final result (2.18) relates directly to equations (2.10) and (2.14), when setting $k = 0$ and $k = 1$. We will demonstrate that these equations have a unique solution using a fixed-point argument.

For a fixed h and k, we consider the matrix function $g : \mathbb{R}^{r\times q} \to \mathbb{R}^{r\times q}$ defined by

$$g(T) = \tfrac{(m-1)!}{h^{m-1}} \Big[f \Big(a+(k+1)h, S_{\big|_{[a+kh,a+(k+1)h]}} (a+(k+1)h) + \overline{Y'(a+kh)}h$$
$$+ \cdots + \tfrac{h^{m-1}}{(m-1)!}\overline{Y^{(m-1)}(a+kh)} + \tfrac{h^m}{m!}T \Big) - \overline{Y'(a+kh)} - \overline{Y''(a+kh)}h$$
$$- \cdots - \tfrac{h^{m-2}}{(m-2)!}\overline{Y^{(m-1)}(a+kh)} \Big] . \tag{2.19}$$

Relation (2.18) holds if and only if $A_k = g(A_k)$, that is, if A_k is a fixed point for function $g(T)$. By using the definition (2.19) of g and applying the global Lipschitz's condition (2.3) for f, it immediately follows that

$$\|g(T_1) - g(T_2)\| \leq \frac{Lh}{m} \|T_1 - T_2\| .$$

Taking $h < m/L$, the matrix function g is contractive. Therefore equation (2.18) has unique solutions A_k for $k = 0, 1, \ldots, n-1$, and the matrix spline is completely determined. In summary, we have proved the following theorem:

Theorem 2.1 *For the first-order matrix differential equation (2.1), let L be the corresponding Lipschitz constant defined by (2.3). We also consider the partition (2.4) with step size $h < m/L$. Then, the matrix spline $S(x)$ of order $m \in \mathbb{N}$ exists in each subinterval $[a+kh, a+(k+1)h]$, $k = 0, 1, \ldots, n-1$, as defined in the previous construction and is of class $\mathcal{C}^1[a,b]$.*

Observe that the so constructed splines have a global error of $O(h^{m-1})$, which follows from an analysis similar to Loscalzo and Talbot's work [3].

The approximate solution of (2.1) can be computed by means of matrix splines of order m in the interval $[a,b]$ with an error of the order $O(h^{m-1})$ under the conditions of Theorem 2.1. The procedure is as follows:

- Using any convenient computer-algebra system, obtain the matrix functions $g_1(x, Y(x)), \ldots, g_{m-2}(x, Y(x))$ given by (2.6)–(2.8) and determine the constants $Y''(a), \ldots, Y^{(m-1)}(a)$. Choose $n > L(b-a)/m$ so that $h = (b-a)/n$ with the partition $\Delta_{[a,b]}$ defined by Eq. (2.4).

- Solve equation (2.10) to find A_0, and determine $S_{\big|_{[a,a+h]}}(x)$ of Eq. (2.5).

- Iteratively, for $k = 1, \ldots, n-1$, solve equations (2.18) to find all A_k. Next, compute the splines $S_{\big|_{[a+kh,a+(k+1)h]}}(x)$ according to Eq. (2.15).

In order to find A_k for $k = 0, 1, \ldots, n-1$, one may solve equations (2.10) and (2.18) either explicitly [8], or by employing an iterative method [9]. For example, we can consider the recursion relation $T^s_{l+1} = g(T^s_l)$. Here, T^s_0 is an arbitrary matrix in $\mathbb{R}^{r\times q}$ for $s = 0, 1, \ldots, n-1$, and $g(T)$ is given by (2.19).

3 Numerical Examples

3.1 A scalar test problem

This simple test problem is motivated by Loscalzo and Talbot's seminal work on scalar spline function approximation for ordinary differential equations [3]. Unfortunately, their otherwise very efficient method had the drawback to be divergent for higher degree spline functions ($m > 3$). Here, we will compare our procedure with their test case for the spline solution of $y' = y$ with initial condition $y(0) = 1$.

Figure 1 depicts the error of fourth-order spline solutions for the Loscalzo-Talbot problem which were constructed by our proposed method. Observe that for $h = 0.01$ the results already reach the accuracy of 10^{-14}, compared to the serious error of the conventional Loscalzo-Talbot method [3]. It also becomes clear that a further reduction in step size h does not necessarily improve the approximation.

It may be interesting to study the increasing quality of the approximation with higher-order splines. Figure 2 shows how the solutions improve by taking $m = 4, 5, 6$, respectively, with a constant step size $h = 0.1$.

3.2 A non-linear vector system

As a second example of our method, we choose the following vector differential system for the interval $x \in [0, 1]$, which is clearly non-linear:

$$\left.\begin{aligned} y_1'(x) &= -1 + e^x - \sin x + \sin\left(y_2(x)\right) \\ y_2'(x) &= \frac{1}{4 + y_1^2(x)} - \frac{1}{5 + e^{2x} + 2\,e^x\,\cos x - \sin^2 x} \end{aligned}\right\} \qquad (3.1)$$

with the initial values

$$\left.\begin{aligned} y_1(0) &= 2 \\ y_2(0) &= \pi/2 \end{aligned}\right\}.$$

We can then rewrite the problem using vector notation $\mathbf{y}(x) = \begin{pmatrix} y_1(x) \\ y_2(x) \end{pmatrix}$ with $\mathbf{y}(0) = \begin{pmatrix} 2 \\ \pi/2 \end{pmatrix}$ to obtain the nonlinear vector problem $\mathbf{y}'(x) = \mathbf{f}(x, \mathbf{y}(x))$, where

$$\mathbf{f}\,(x, \mathbf{y}(x)) = \begin{pmatrix} -1 + e^x - \sin x + \sin\left(y_2(x)\right) \\ \dfrac{1}{4 + y_1^2(x)} - \dfrac{1}{5 + e^{2\,x} + 2\,e^x\ \cos x - \sin^2 x} \end{pmatrix}. \quad (3.2)$$

According to Ref. [6] this problem has the exact solution $y_1(x) = e^x + \cos x$ and $y_2(x) = \pi/2$, and hence for this test case we will be able to assess the exact error of our numerical estimates. Our proposed method serves to construct the splines of fifth order for the problem given in Eq. (3.1). For this we require to calculate $\mathbf{y}''(x)$, $\mathbf{y}^{(3)}(x)$ and $\mathbf{y}^{(4)}(x)$, which in general is straightforward. We may derive $\mathbf{y}''(x) = \begin{pmatrix} y_1''(x) \\ y_2''(x) \end{pmatrix}$ using a computer algebra system such as *Mathematica*, which readily produces:

$$\left.\begin{aligned} y_1''(x) &= e^x - \cos(x) + \cos\left(y_2(x)\right)y_2'(x) \\ y_2''(x) &= \frac{2e^{2x} + 2e^x\cos(x) - 2e^x\sin(x) - 2\cos(x)\sin(x)}{\left(5 + e^{2x} + 2e^x\cos(x) - \sin(x)^2\right)^2} - \frac{2y_1(x)y_1'(x)}{\left(4 + y_1(x)^2\right)^2} \end{aligned}\right\}. \quad (3.3)$$

Taking into account that $y_1(0) = 2, y_1'(0) = 1, y_2(0) = \pi/2$, and $y_2'(0) = 0$, it follows by Eq. (3.3) that $\mathbf{y}''(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. We similarly calculate the third-order derivative $\mathbf{y}^{(3)}(x) = \begin{pmatrix} y_1^{(3)}(x) \\ y_2^{(3)}(x) \end{pmatrix}$ with components:

$$\left.\begin{aligned} y_1^{(3)}(x) &= e^x + \sin(x) - \sin(y_2(x))\left(y_2'(x)\right)^2 + \cos(y_2(x)) y_2''(x) \\ y_2^{(3)}(x) &= -\frac{8\left(\cos(2x) - 2e^x\left(e^x - \sin(x)\right)\right)}{\left(9 + 2e^{2x} + 4e^x\cos(x) + \cos(2x)\right)^2} \\ &- \frac{64\left(e^x + \cos(x)\right)^2\left(e^x - \sin(x)\right)^2}{\left(9 + 2e^{2x} + 4e^x\cos(x) + \cos(2x)\right)^3} \\ &+ \frac{8\left(y_1(x)\right)^2\left(y_1'(x)\right)^2}{\left(4 + \left(y_1(x)\right)^2\right)^3} - \frac{2\left(y_1'(x)\right)^2}{\left(4 + \left(y_1(x)\right)^2\right)^2} - \frac{2y_1(x)y_1''(x)}{\left(4 + \left(y_1(x)\right)^2\right)^2} \end{aligned}\right\} \tag{3.4}$$

In like manner as before, we consider $y_1(0) = 2, y_1'(0) = 1, y_1''(0) = 0, y_2(0) = \pi/2, y_2'(0) = 0$, and $y_2''(0) = 0$ with (3.4) to deduce $\mathbf{y}^{(3)}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Similarly, we may then derive the explicit results for the components of $\mathbf{y}^{(4)}(x) = \begin{pmatrix} y_1^{(4)}(x) \\ y_2^{(4)}(x) \end{pmatrix}$. In the final step, it remains to substitute the known values $y_1(0) = 2, y_1'(0) = 1, y_1''(0) = 0, y_1'''(0) = 1, y_2(0) = \pi/2, y_2'(0) = 0, y_2''(0) = 0, y_2'''(0) = 0$, into the last expression to obtain $\mathbf{y}^{(4)} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$.

Also, it is not difficult to see that f, defined by (3.2), fulfills the global Lipschitz's condition

$$\left\|\mathbf{f}(x, \mathbf{y}) - \mathbf{f}(x, \mathbf{z})\right\|_1 \leq \left\|\mathbf{y} - \mathbf{z}\right\|_1, \quad 0 \leq x \leq 1, \quad \mathbf{y}, \mathbf{z} \in \mathbb{R}^2. \tag{3.5}$$

Comparing with the general form (2.3), we note that $L = 1$. Therefore, by Theorem 2.1 we need to take $h < 5$. In the following, for example we choose $h = 0.1$ and summarize the numerical results in Table 2. In each interval, we evaluated the difference between the estimates of our numerical approach and the exact solution, and then take the Fröbenius norm of this difference, following the procedure explained in Ref. [6]. Table 1 lists the maximum of these errors for each subinterval.

For the solution of the vector differential system (3.1), Figure 3 illustrates the approximation behavior of various splines of the fourth order ($m = 4$) with the different step sizes $h = 0.1, 0.01$, and $h = 0.001$. All vector splines lie well in the predicted range of Theorem 2.1 and provide excellent approximations for the problem at hand with the benefit of very low computational cost. Observe that at step size $h = 0.001$ the limit of machine precision is practically reached and explains the random fluctuations around 10^{-15}. Hence, it obviously is of lesser interest to obtain more accurate approximations for $m = 4$ and $h = 0.001$.

3.3 Sylvester matrix differential equation

In many areas of science and engineering linear matrix differential equations appear of the type

$$\left.\begin{array}{rcl} Y'(x) &=& A(x)Y(x) + Y(x)B(x) + C(x) \\ Y(a) &=& Y_a \end{array}\right\} \quad a \le x \le b, \tag{3.6}$$

where $Y(x), A(x), B(x), C(x) \in \mathbb{R}^{r\times r}$. The case of constant coefficients has been studied by several authors [10], whereas the variable-coefficient case has so far received little numerical treatment in the literature.

Following Ref. [6], we choose the following Sylvester problem (3.6) as a final example:

$$\begin{array}{rcl} A(x) &=& \begin{pmatrix} 0 & xe^{-x} \\ x & 0 \end{pmatrix}, \quad B(x) = \begin{pmatrix} 0 & x \\ 0 & 0 \end{pmatrix}, \\ C(x) &=& \begin{pmatrix} -e^{-x}(1+x^2) & -2e^{-x}x \\ 1 - e^{-x}x & -x^2 \end{pmatrix} \\ Y(0) &=& \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad Y(x) \in \mathbb{R}^{2\times 2},\ 0 \le x \le 1. \end{array} \tag{3.7}$$

According to [6] we know that this problem has the exact solution

$$Y(x) = \begin{pmatrix} e^{-x} & 0 \\ x & 1 \end{pmatrix}$$

with the Lipschitz constant $L = 2$. The higher-order derivatives of $Y(x)$ are required for the construction of the spline approximation and can be readily obtained.

For splines of the fifth order ($m = 5$), we take $n = 10$ partitions and $h = 0.1$. The results are summarized in Table 3, where the numerical estimates have been rounded to the sixth relevant digit. In Table 4, we evaluated the difference between the estimates of our numerical approach and the exact solution, and then take the Fröbenius norm of this difference. The maximum of these errors are indicated for each subinterval.

For the solution of the Sylvester matrix problem (3.6), Figure 4 depicts the approximation behavior of various splines of the fifth order ($m = 5$) with the different step sizes $h = 0.1, 0.01$, and $h = 0.001$. As before, all matrix splines lie well in the predicted range of Theorem 2.1. It becomes evident that the splines for step sizes $h = 0.01$ and $h = 0.001$ are almost indistinguishable and reach the same precision of almost 10^{-14}.

We also carried out the computations for the sixth order matrix splines ($m = 6$) with the step sizes $h = 0.1, 0.01$, and $h = 0.001$, and as expected, we could observe that $h = 0.01$ yields an accuracy close to machine precision. Interestingly, higher step sizes do not improve these approximations—the quality of approximation indeed deteriorates due to the accumulation of rounding errors.

3.4 The Hénon-Heiles system

The Hénon-Heiles equation [12] is a nonlinear nonintegrable Hamiltonian system defined by

$$\left.\begin{array}{rcl} x'' & = & -\dfrac{\partial V(x,y)}{\partial x} \\[2ex] y'' & = & -\dfrac{\partial V(x,y)}{\partial y} \end{array}\right\}, \quad a \le t \le b, \tag{3.8}$$

where the potential-energy function is conserved during motion and given by the following expression

$$V(x,y) = \frac{1}{2}\left(x^2 + y^2 + 2x^2y - \frac{2}{3}y^3\right).$$

The differential system (3.8) can be recast in vectorial form $\mathbf{u}'(t) = \mathbf{f}\,(t, \mathbf{u}(t))$, where $\mathbf{u}(t) = (u_1(t)\ u_2(t)\ u_3(t)\ u_4(t))^T \in \mathbb{R}^4$ and

$$\mathbf{f}\,(t, \mathbf{u}) = \begin{pmatrix} u_2 \\ -u_1 - 2u_1u_3 \\ u_4 \\ -u_3 - u_1^2 + u_3^2 \end{pmatrix},$$

and f satisfies

$$\|\mathbf{f}\,(t,\mathbf{y})-\mathbf{f}\,(t,\mathbf{z})\|_1 \le 5\,\|\mathbf{y}-\mathbf{z}\|_1\,, \quad 0 \le t \le 1, \quad \mathbf{y},\mathbf{z} \in \mathbb{R}^4.$$

Then, one gets that $L = 5$ and by Theorem 2.1 we need to take $h < m/5$. For these benchmark tests, we have taken $t \in [0,1]$ with the initial conditions $x(0) = 1, \dot{x}(0) = 0.5, y(0) = 1$, and $\dot{y}(0) = 0.5$. Since the solution is unknown, we have considered as reference values the results generated by MATLAB ODE solver *ode45*. The parameters ***RelTol*** and ***AbsTol*** were chosen to obtain the maximum precision ($RelTol = 2.22045 \cdot 10^{-14}$, $AbsTol = 1.0 \cdot 10^{-14}$). The ODE solver *ode45* allows to solve non-stiff differential equations and is based on the Runge-Kutta method.

The numerical estimates are shown in Figures 5 and 6. Figure 5 depicts the errors for splines of order $m = 4$ with variable step size $h = 0.1, 0.01, 0.001$, whereas in Figure 6 the step size $h = 0.1$ is fixed and the spline order varies $m = 4, 5, 6$.

As can be seen in Figure 5, the error is situated well within the expected margins improving with each lower value of h. On the other hand, in Figure 6 with $h = 0.1$ the error is not exceeding the predicted maximum estimate $O(h^{m-1})$ for $m = 4, 5, 6$.

4 Conclusions

This work focuses on the presentation of a new method for the numerical integration of first-order matrix differential equations of the type $Y'(x) = f(x, Y(x))$ in the interval $[a, b]$ using higher-order matrix splines ($m > 3$). Contrary to existing spline methods in the literature, this new method only requires first-order derivatives for the construction of the splines to provide a continuous approximation of order $O(h^{m-1})$. Additionally, our method is well-suited for implementation on numerical and/or symbolical computer systems.

For an explicit demonstration of our proposed method and its advantages over existing conventional methods, we discussed three numerical test cases with excellent results. It is hoped that this new approach to approximating matrix differential models will motivate and open up alternative avenues to tackle different related problems in science and engineering.

Acknowledgments

We wish to thank the unnamed referee for his valuable suggestions to improve the original manuscript.

References

[1] G. Freiling and A. Hochhaus, On a class of rational matrix differential equations arising in stochastic control, Linear Algebra Appl. 379 (2004) 43–68.

[2] U.M. Ascher, R.M.M. Mattheij, R.D. Russell, Numerical solutions of boundary value problems for ordinary differential equations, Prentice Hall, New Jersey, 1988.

[3] F.R. Loscalzo, T.D. Talbot, Spline function approximations for solutions of ordinary differential equations, SIAM J. Numer. Anal. 4(3) (1967) 433–445.

[4] E.A. Al-Said, M.A. Noor, Cubic splines method for a system of third-order boundary value problems, Appl. Math. Comput. 142 (2003) 195–204.

[5] E. Defez, L. Soler, A. Hervás, C. Santamaría, Numerical solutions of matrix differential models using cubic matrix splines, Comput. Math. Appl. 50 (2005) 693–699.

[6] E. Defez, A. Hervás, L. Soler, M.M. Tung, Numerical solutions of matrix differential models using cubic matrix splines II, Math. Comp. Modelling 46 (2007) 657–669.

[7] T.M. Flett, Differential Analysis, Cambridge University Press, 1980.

[8] P. Lancaster, Explicit solutions of linear matrix equations, SIAM Review 12 (1970) 544–566.

[9] J.M. Ortega, W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, 1972.

[10] A.Y. Barraud, Noveaux développements sur la résolution numérique de $X' = AX + XB + C; X(0) = C$, R.A.I.R.O. 16(4) (1982) 341–356.

[11] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, John Wiley, New York. (1981).

[12] M. Hénon and C. Heiles, The Applicability of the Third Integral of Motion: Some Numerical Experiments, Astron. J. 69 (1964) 73–79.

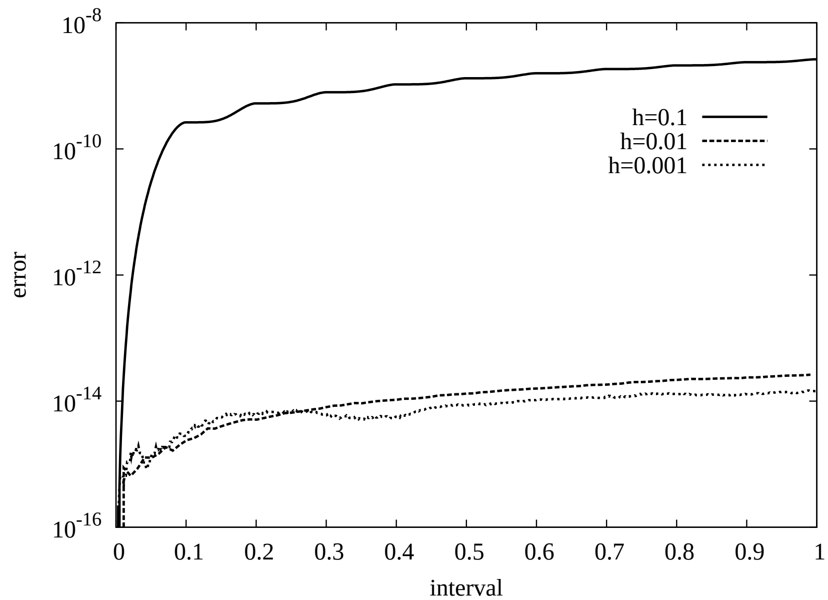

Figure 1: Error for the Loscalzo-Talbot problem with splines of fourth order ($m = 4$) using our proposed method for various step sizes.

Interval	$[0, 0.1]$	$[0.1, 0.2]$	$[0.2, 0.3]$	$[0.3, 0.4]$	$[0.4, 0.5]$
Max. error	8.2362×10^{-12}	4.8717×10^{-11}	1.27357×10^{-10}	2.50353×10^{-10}	4.24194×10^{-10}
Interval	$[0.5, 0.6]$	$[0.6, 0.7]$	$[0.7, 0.8]$	$[0.8, 0.9]$	$[0.9, 1.0]$
Max. error	6.55672×10^{-10}	9.51896×10^{-10}	1.32033×10^{-9}	1.7688×10^{-9}	2.30555×10^{-9}

Table 1: Approximation error for vector problem (3.1).

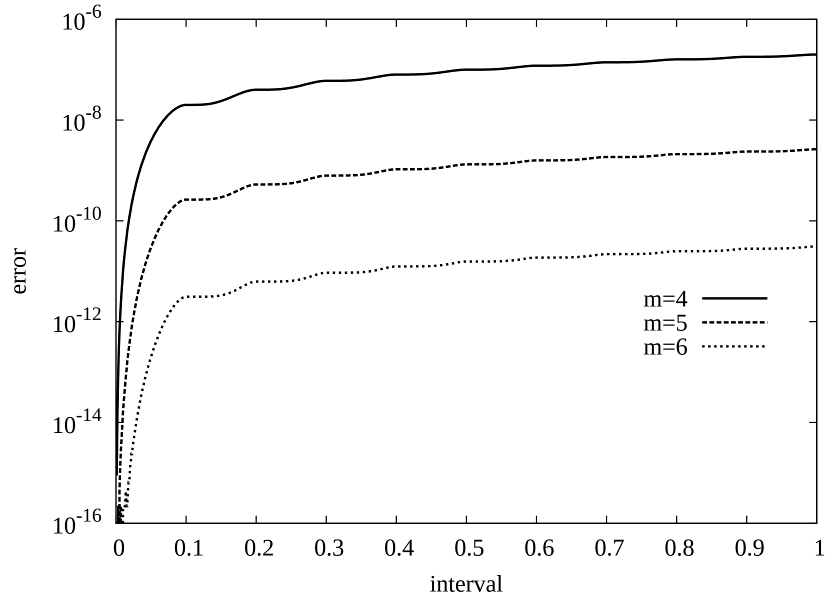

Figure 2: Errors for increasing spline orders ($m = 4, 5, 6$) solving the Loscalzo-Talbot problem. The step size is constant ($h = 0.1$).

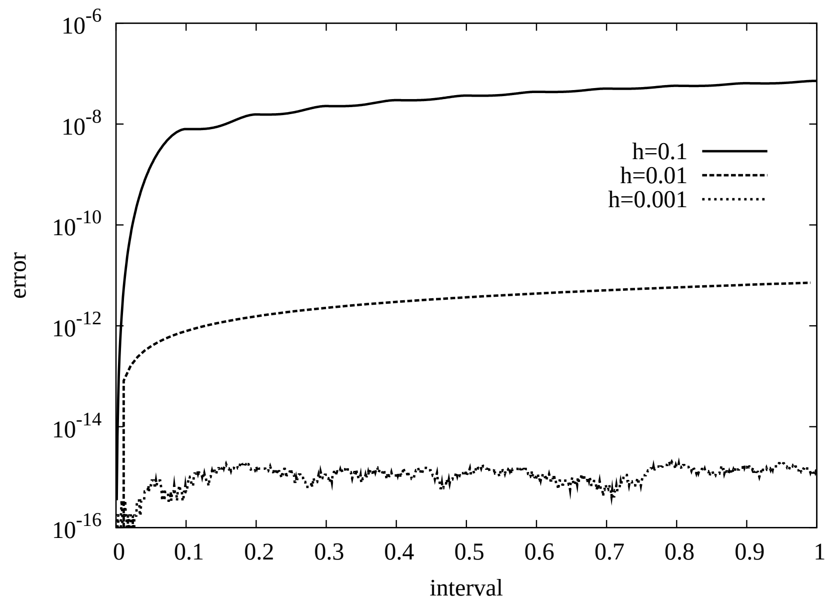

Figure 3: Representing the 2-norm error for the vector differential system (3.1) using splines of fourth order ($m = 4$).

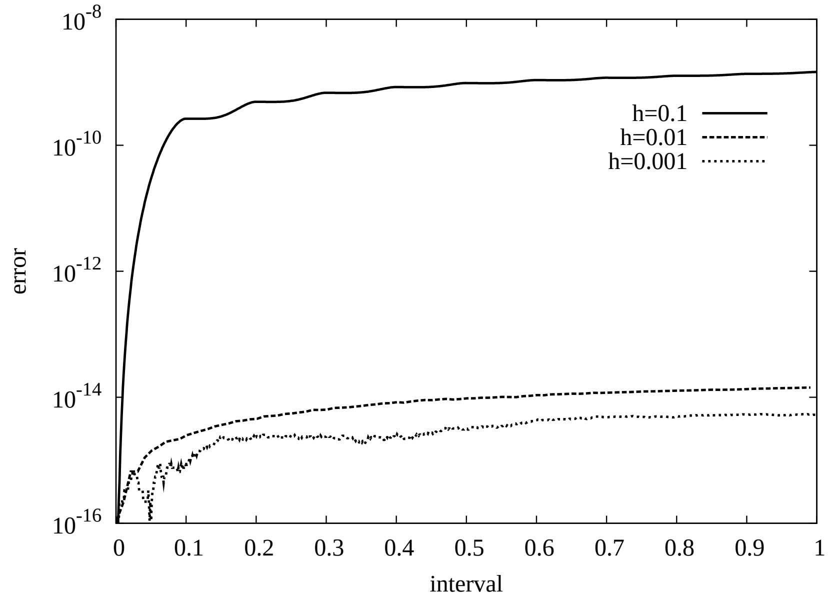

Figure 4: Representing the 2-norm error for the Sylvester matrix differential equation (3.6) using splines of fourth order ($m = 4$).

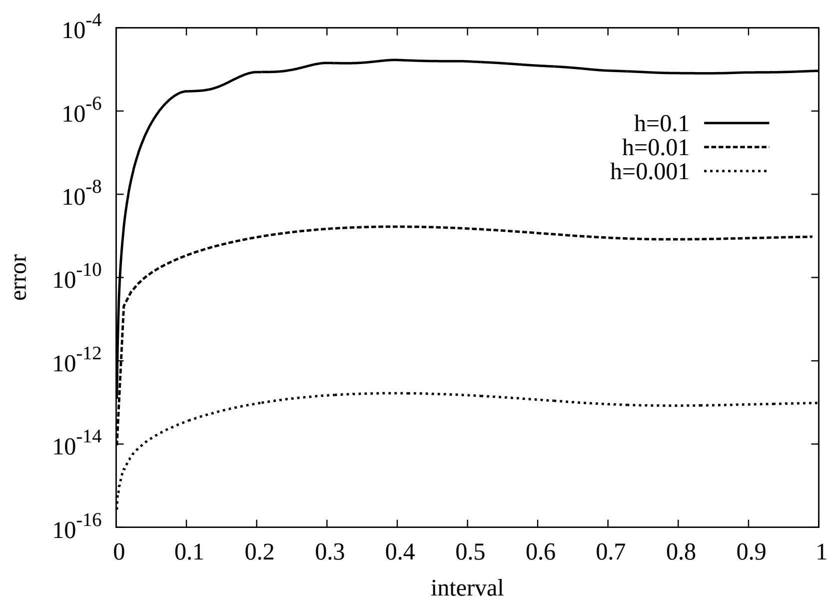

Figure 5: Error for the Hénon-Heiles problem with splines of fourth order ($m = 4$) using our proposed method for various step sizes.

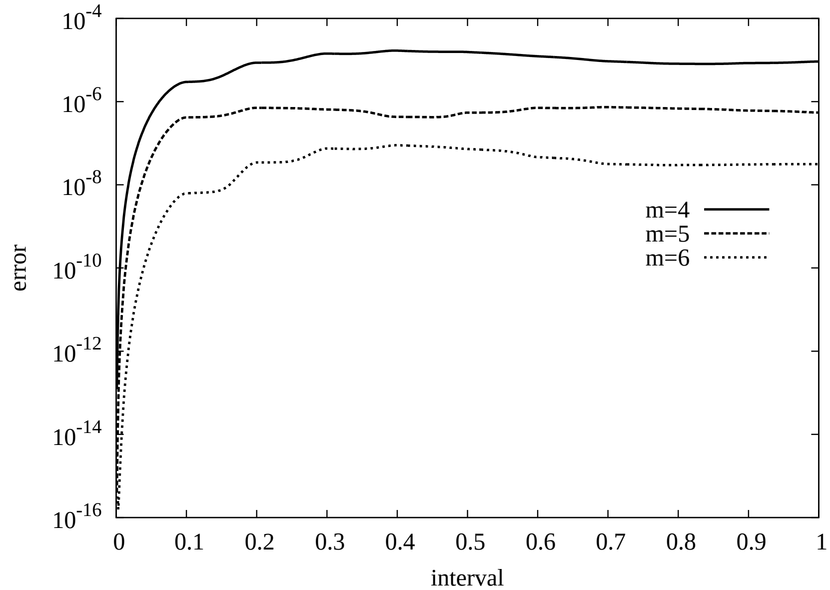

Figure 6: Errors for increasing spline orders ($m = 4, 5, 6$) solving the Hénon-Heiles problem. The step size is constant ($h = 0.1$).

Interval	Approximation
$[0, 0.1]$	$\begin{pmatrix} 2. + x + 0.166667x^3 + 0.0833333x^4 + 0.00833619x^5 \\ 1.5708 \end{pmatrix}$
$[0.1, 0.2]$	$\begin{pmatrix} 2. + 1.x - 3.98676 \times 10^{-7}x^2 + 0.166671x^3 + 0.0833075x^4 + 0.0083996x^5 \\ 1.5708 + 1.02341 \times 10^{-9}x^2 - 9.53254 \times 10^{-9}x^3 + 4.27272 \times 10^{-8}x^4 - 7.27159 \times 10^{-8}x^5 \end{pmatrix}$
$[0.2, 0.3]$	$\begin{pmatrix} 2. + 1.x - 9.78808 \times 10^{-6}x^2 + 0.166723x^3 + 0.0831609x^4 + 0.00856703x^5 \\ 1.5708 - 2.80891 \times 10^{-9}x + 2.68447 \times 10^{-8}x^2 - 1.2696 \times 10^{-7}x^3 + 2.96285 \times 10^{-7}x^4 - 2.72203 \times 10^{-7}x^5 \end{pmatrix}$
$[0.3, 0.4]$	$\begin{pmatrix} 2. + 1.00001x - 0.000070073x^2 + 0.166941x^3 + 0.0827649x^4 + 0.00885657x^5 \\ 1.5708 - 2.3641 \times 10^{-8}x + 1.51773 \times 10^{-7}x^2 - 4.84484 \times 10^{-7}x^3 + 7.68203 \times 10^{-7}x^4 - 4.83576 \times 10^{-7}x^5 \end{pmatrix}$
$[0.4, 0.5]$	$\begin{pmatrix} 2. + 1.00005x - 0.000295117x^2 + 0.167541x^3 + 0.0819626x^4 + 0.00928717x^5 \\ 1.5708 - 1.04291 \times 10^{-7}x + 5.05234 \times 10^{-7}x^2 - 1.21958 \times 10^{-6}x^3 + 1.46618 \times 10^{-6}x^4 - 7.01984 \times 10^{-7}x^5 \end{pmatrix}$
$[0.5, 0.6]$	$\begin{pmatrix} 1.99998 + 1.0002x - 0.000921692x^2 + 0.168862x^3 + 0.080566x^4 + 0.00987867x^5 \\ 1.5708 - 3.25859 \times 10^{-7}x + 1.26869 \times 10^{-6}x^2 - 2.46395 \times 10^{-6}x^3 + 2.3864 \times 10^{-6}x^4 - 9.21882 \times 10^{-7}x^5 \end{pmatrix}$
$[0.6, 0.7]$	$\begin{pmatrix} 1.99993 + 1.00062x - 0.00237386x^2 + 0.171395x^3 + 0.0783551x^4 + 0.0106518x^5 \\ 1.5708 - 8.18293 \times 10^{-7}x + 2.66421 \times 10^{-6}x^2 - 4.32971 \times 10^{-6}x^3 + 3.51164 \times 10^{-6}x^4 - 1.13697 \times 10^{-6}x^5 \end{pmatrix}$
$[0.7, 0.8]$	$\begin{pmatrix} 1.9998 + 1.00162x - 0.00534205x^2 + 0.175805x^3 + 0.0750753x^4 + 0.0116284x^5 \\ 1.5708 - 1.76297 \times 10^{-6}x + 4.93332 \times 10^{-6}x^2 - 6.89355 \times 10^{-6}x^3 + 4.80962 \times 10^{-6}x^4 - 1.34027 \times 10^{-6}x^5 \end{pmatrix}$
$[0.8, 0.9]$	$\begin{pmatrix} 1.99947 + 1.00376x - 0.0108784x^2 + 0.18297x^3 + 0.0704351x^4 + 0.0128313x^5 \\ 1.5708 - 3.38486 \times 10^{-6}x + 8.30591 \times 10^{-6}x^2 - 0.0000101804x^3 + 6.23218 \times 10^{-6}x^4 - 1.52432 \times 10^{-6}x^5 \end{pmatrix}$
$[0.9, 1.0]$	$\begin{pmatrix} 1.99873 + 1.00796x - 0.0205098x^2 + 0.19401x^3 + 0.0641039x^4 + 0.0142844x^5 \\ 1.5708 - 5.93162 \times 10^{-6}x + 0.000012961x^2 - 0.0000141487x^3 + 7.71598 \times 10^{-6}x^4 - 1.68162 \times 10^{-6}x^5 \end{pmatrix}$

Table 2: Vector approximation for system (3.1) in the interval $[0, 1]$.

Interval	Approximation
$[0, 0.1]$	$\begin{pmatrix} 1. - 1.x + 0.5x^2 - 0.166667x^3 + 0.0416667x^4 - 0.00816941x^5 & 0. \\ x. & 1. \end{pmatrix}$
$[0.1, 0.2]$	$\begin{pmatrix} 1. - 1.x + 0.499997x^2 - 0.166626x^3 + 0.0413976x^4 - 0.00739198x^5 & 0. \\ 1.x & 1. \end{pmatrix}$
$[0.2, 0.3]$	$\begin{pmatrix} 1. - 0.999997x + 0.499961x^2 - 0.166422x^3 + 0.0408023x^4 - 0.00668854x^5 & 0. \\ 1.x & 1. \end{pmatrix}$
$[0.3, 0.4]$	$\begin{pmatrix} 0.999999 - 0.999979x + 0.499834x^2 - 0.165957x^3 + 0.0399455x^4 - 0.00605204x^5 & 0. \\ 1.x & 1. \end{pmatrix}$
$[0.4, 0.5]$	$\begin{pmatrix} 0.999995 - 0.999925x + 0.499542x^2 - 0.16517x^3 + 0.0388822x^4 - 0.00547612x^5 & 0. \\ 1.x & 1. \end{pmatrix}$
$[0.5, 0.6]$	$\begin{pmatrix} 0.999983 - 0.999797x + 0.499x^2 - 0.16402x^3 + 0.0376596x^4 - 0.00495499x^5 & 0 \\ 1.x & 1. \end{pmatrix}$
$[0.6, 0.7]$	$\begin{pmatrix} 0.999954 - 0.999547x + 0.498127x^2 - 0.16249x^3 + 0.0363175x^4 - 0.00448346x^5 & 0 \\ 1.x & 1. \end{pmatrix}$
$[0.7, 0.8]$	$\begin{pmatrix} 0.999896 - 0.999117x + 0.496844x^2 - 0.160578x^3 + 0.0348899x^4 - 0.00405681x^5 & 0 \\ 1.x & 1. \end{pmatrix}$
$[0.8, 0.9]$	$\begin{pmatrix} 0.999792 - 0.998438x + 0.495083x^2 - 0.158291x^3 + 0.033405x^4 - 0.00367075x^5 & 0 \\ 1.x & 1. \end{pmatrix}$
$[0.9, 1.0]$	$\begin{pmatrix} 0.999617 - 0.997437x + 0.492785x^2 - 0.155651x^3 + 0.0318868x^4 - 0.00332143x^5 & 0 \\ 1.x & 1. \end{pmatrix}$

Table 3: Approximation for the Sylvester matrix problem (3.6).

Interval	$[0, 0.1]$	$[0.1, 0.2]$	$[0.2, 0.3]$	$[0.3, 0.4]$	$[0.4, 0.5]$
Max. error	2.6999×10^{-10}	5.1438×10^{-10}	7.36134×10^{-10}	9.38797×10^{-10}	1.1268×10^{-9}
Interval	$[0.5, 0.6]$	$[0.6, 0.7]$	$[0.7, 0.8]$	$[0.8, 0.9]$	$[0.9, 1.0]$
Max. error	1.30572×10^{-9}	1.48252×10^{-9}	1.66579×10^{-9}	1.86603×10^{-9}	2.09601×10^{-9}

Table 4: Approximation error for the Sylvester matrix problem (3.6).