Journal of Mathematics Research; Vol. 7, No. 3; 2015
ISSN 1916-9795 E-ISSN 1916-9809
Published by Canadian Center of Science and Education

Multivariate Canonical Polynomials in the Tau Method with Applications to Optimal Control Problems

Mohamed K. El-Daou[1], Khaled M. Al-Hamad[1] & Ahmed S. Zadeh[1]

[1] Applied Sciences Department, College of Technological Studies, P.O.Box 64287 Shuwaikh/B, 70453, Kuwait

Correspondence: Mohamed K. El-Daou, Applied Sciences Department, College of Technological Studies, P.O.Box 64287 Shuwaikh/B, 70453, Kuwait. E-mail: mk.eldaou@paaet.edu.kw

Received: May 2, 2015 Accepted: June 18, 2015 Online Published: July 11, 2015

doi:10.5539/jmr.v7n3p26 URL: http://dx.doi.org/10.5539/jmr.v7n3p26

Abstract

The Tau method is a highly accurate technique that approximates differential equations efficiently. This paper discusses two approaches of the Tau Method: recursive and spectral. In the recursive Tau, the approximate solution of the differential equation is obtained in terms of a special polynomial basis called *canonical polynomials*. The present paper extends this concept to the *multivariate canonical polynomial vectors* and proposes a self starting algorithm to generate those vectors. In the spectral Tau, the approximate solution is obtained as a truncated series expansions in terms of a set of orthogonal polynomials where the coefficients of the expansions are obtained by forcing the defect of the differential equation to vanish at the some selected points. In this paper we use the spectral Tau to solve a class of optimal control problems associated with a nonlinear system of differential equations. Some numerical examples that confirm our method are given.

Keywords: Canonical polynomials, Tau method, optimal control problems

AMS Subject Classification: 65L05 - 65L70

1. Introduction

The Tau method is a highly accurate technique that approximates differential equations without requiring the discretization of the given differential operator. Its basic idea is to perturb the right hand side of the differential equation in a way that an exact polynomial solution of the new equation can be found analytically. This method was devised in (Lanczos, 1956) to find polynomial approximations for simple linear ordinary differential equations (ODE) and it was extended later on to treat differential equations with different level of complexities, (see (Ortiz, 1969), (Ortiz & Samara, 1984), (El-Daou & Ortiz, 1992-1994), and (Liu & Pan, 1999)).

The Tau method has three equivalent approaches: Recursive, operational and spectral. The *recursive approach*, proposed in (Ortiz, 1969), permits to obtain an approximate polynomial solution expressed in terms of a special polynomials basis called *canonical polynomials*. This technique has been thoroughly investigated in a series of papers (see for example (Crisci & Russo, 1983), (Freilich & Ortiz, 1982) and (El-Daou & Ortiz, 1994-1998)). In the *operational Tau*, (see (Ortiz & Samara, 1981)), the ODE is transformed to a system of linear algebraic equations using some simple elementary matrices. The operational procedure was extended in (Liu & Pan, 1999) to solve mixed-order systems of linear ODEs with polynomial coefficients. In the same reference an automation of the operational approach has been reported. In (Ortiz & Samara, 1984) the operational Tau was shown effective in solving partial differential equations. In (Canuto, Hussaini, Quarteroni, & Zang, 2006) and (Gottlieb & Orszag, 1977), a *spectral approach* to the Tau method has been studied. This technique seeks an approximate solution in the form of a truncated series expansions of Chebyshev or Legendre polynomials. The coefficients of the series are computed by forcing the ODE to be exact at some selected points (called collocation points) and the supplementary conditions to be satisfied exactly. The spectral approach of the Tau Method guarantees spectral accuracy because the approximate solution is obtained in terms of orthogonal polynomials basis.

Although the three approaches of the Tau Method explained above appear to be different, it was shown in (El-Daou & Ortiz, 1992-1994) that they are equivalent in the sense that they yield the same approximate solution. However, the suitability of those approaches is judged by the problem under consideration. While the recursive and the operational Tau are more suitable from the computation point of view for ODE with polynomial coefficients, the spectral Tau enjoys a superiority when the ODE is nonlinear or its coefficients are not polynomials.

We point out that an important feature that distinguishes the Tau Method from the classical finite difference methods

is that the Tau solution is obtained in a closed form on the whole domain of integration without discretization, while in the finite difference methods the domain is divided into small elements of stepsize h on which depends the accuracy of the method.

The present paper is to extend the recursive approach of the Tau Method to the numerical solution of systems of linear ODEs and to give a practical procedure that permits to construct approximate solutions. Further, we will show that the spectral Tau method is highly effective in tackling a class of optimal control problems (see (Flores Tlacuahuac, Terrazas Moreno, & Biegler, 2008) and (Jaddu & Majdalawi, 2014)). To this end, we generalize in Section 2 the concept of canonical polynomials to be adapted for system of ODEs. An algorithm to compute the Tau solution in terms of the canonical vectors will be given in Section 3. Section 4 is concerned with applying the spectral Tau to solve an optimal control problem whose the constraints are given as a system of nonlinear ODEs. Numerical examples illustrating the efficiency of our method are provided throughout the paper.

2. Canonical Polynomial Vectors

Let us consider a system of linear ODEs of dimension $\nu \geq 1$ written in the matrix form as

$$\mathbf{Dy}(x) := \left[\mathbf{I}_\nu \frac{d}{dx} + \mathbf{A}(x)\right]\mathbf{y}(x) = \mathbf{f}(x)\ ; \quad x \in [0,1], \tag{1}$$

where

$$\begin{aligned} \mathbf{f}(x) &:= [f_1(x), f_2(x), \ldots, f_\nu(x)]^T, \\ \mathbf{y}(x) &:= [y_1(x), y_2(x), \ldots, y_\nu(x)]^T, \end{aligned} \tag{2}$$

$\mathbf{I}_\nu$ is the ν identity matrix and $\mathbf{A} := (A_{ij}(x))_{i,j=1}^{\nu}$ is a $\nu \times \nu$ matrix with $A_{ij}(x)$ being functions of x. The superscript T means "Transpose". We shall assume for simplicity that all the $A_{ij}(x)$'s are polynomials having the same degree d,

$$A_{ij}(x) = \sum_{m=0}^{d} a_{ij}^{(m)} x^m; \quad a_{ij}^{(m)} \in \mathbf{R},\ i, j = 1, 2, \ldots, \nu,$$

and more compactly we can write

$$\mathbf{A}(x) = \sum_{m=0}^{d} \mathbf{A}_m x^m, \tag{3}$$

where $\{\mathbf{A}_m = \left(a_{ij}^{(m)}\right)_{i,j=1}^{\nu}\ ; m = 0, 1, 2, \ldots, \mathrm{d}\}$ are constant matrices.

Let us associate to (1) the initial conditions

$$\mathbf{y}(0) = \mathbf{y}_0 := [y_1 \quad y_2 \quad \ldots \quad y_\nu]^T; \quad y_k \in \mathbf{R}. \tag{4}$$

When $\nu = 1$, system (1) reduces to a single equation $(Dy)(x) = f(x)$. This case was fully discussed in Ortiz (1969) wherein a Tau solution $\tilde{y}(x)$ that approximates $y(x)$ is obtained in the form

$$\tilde{y}(x) = \sum_{k=0}^{N} c_k q_k^*(x), \quad c_k \in \mathbf{R},$$

with $\{q_k^*(x); k \geq 0\}$ being a sequence of functions, called canonical polynomials associated with D each one of which is an exact solution of the differential equation

$$(Dq_k^*)(x) = x^k,\ k \geq 0.$$

In (Ortiz, 1969, Theorem 3.3) a self starting recursive formula that generates the $\{q_k^*; k \geq 0\}$ associated with a single ODE was developed. Next we extend the concept of canonical polynomial to systems of ODEs:

Definitions and Notation.

1. We call a vector $\mathbf{Q}_i^{(k)}(x)$ an ith canonical vector of order k associated with the operator $\mathbf{D}$ if

$$\mathbf{DQ}_k^{(i)} = x^k \mathbf{e}_i, \quad i = 1, 2, ..., \nu,$$

where $\mathbf{e}_i$ is the ith column of $\mathbf{I}_\nu$. Note that $\mathbf{Q}_i^{(k)}(x)$ is a $\nu \times 1$ matrix.

2. $Q_k^* := \begin{bmatrix} \mathbf{Q}_k^{(1)} & \mathbf{Q}_k^{(2)} & \dots & \mathbf{Q}_k^{(\nu)} \end{bmatrix}^T$ will be called a kth *canonical block*. Note that Q_k^* is a $\nu \times \nu$ matrix.

3. $\mathcal{E}_\nu := \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_\nu \end{bmatrix}^T$.

The next theorem is a generalization of Theorem 3.3 given in (Ortiz, 1969):

Theorem 1 *Suppose that the matrix $\mathbf{A}_d$ defined in (3) is non-singular. Then the canonical blocks $\{Q_k^*; k \geq 0\}$ satisfy the recursion:*

$$Q_{k+d}^* = \mathbf{A}_d^{-1}\left[x^k\mathcal{E}_\nu - kQ_{k-1}^* - \sum_{m=0}^{d-1}\mathbf{A}_m Q_{k+m}^*\right]; \quad k \geq 0. \tag{5}$$

In particular, if $d = 0$, then (5) becomes

$$Q_k^* = \mathbf{A}_0^{-1}\left[x^k\mathcal{E}_\nu - kQ_{k-1}^*\right]; \quad k \geq 0.$$

Proof. Let $k \geq 0$ and $i \in \{1, 2, \dots, \nu\}$. Let us apply the operator $\mathbf{D}$, defined by (1), to the vector $x^k\mathbf{e}_i$:

$$\begin{aligned}
\mathbf{D}[x^k\mathbf{e}_i] &= \mathbf{I}_\nu\frac{d}{dx}(x^k\mathbf{e}_i) + \mathbf{A}(x)x^k\mathbf{e}_i \\
&= kx^{k-1}\mathbf{e}_i + [A_{1i}(x) \quad A_{2i}(x) \quad \dots \quad A_{\nu i}(x)]^T x^k \\
&= kx^{k-1}\mathbf{e}_i + \sum_{j=1}^{\nu} A_{ji}(x)x^k\mathbf{e}_j \\
&= kx^{k-1}\mathbf{e}_i + \sum_{\substack{j=1\\ j\neq i}}^{\nu} A_{ji}(x)x^k\mathbf{e}_j + A_{ii}(x)x^k\mathbf{e}_i \\
&= kx^{k-1}\mathbf{e}_i + \sum_{\substack{j=1\\ j\neq i}}^{\nu}[\sum_{m=0}^{d} a_{ji}^{(m)}x^{k+m}]\mathbf{e}_j + \sum_{m=0}^{d} a_{ii}^{(m)}x^{k+m}\mathbf{e}_i \\
&= \mathbf{D}\left[k\mathbf{Q}_{k-1}^{(i)} + \sum_{\substack{j=1\\ j\neq i}}^{\nu}[\sum_{m=0}^{d} a_{ji}^{(m)}\mathbf{Q}_{k+m}^{(j)}] + \sum_{m=0}^{d-1} a_{ii}^{(m)}\mathbf{Q}_{k+m}^{(i)}\right] + a_{ii}^{(d)}x^{k+d}\mathbf{e}_i.
\end{aligned}$$

The last identity is due to the linearity of $\mathbf{D}$. Rearranging terms, and using the definition $\mathbf{D}\mathbf{Q}_k^{(i)} = x^k\mathbf{e}_i$, we obtain:

$$\mathbf{D}\left[x^k\mathbf{e}_i - k\mathbf{Q}_{k-1}^{(i)} - \sum_{\substack{j=1\\ j\neq i}}^{\nu}[\sum_{m=0}^{d} a_{ji}^{(m)}\mathbf{Q}_{k+m}^{(j)}] - \sum_{m=0}^{d-1} a_{ii}^{(m)}\mathbf{Q}_{k+m}^{(i)}\right] = a_{ii}^{(d)}x^{k+d}\mathbf{e}_i = a_{ii}^{(d)}\mathbf{D}\mathbf{Q}_{k+d}^{(i)}.$$

This implies (without loosing generality) that

$$\begin{aligned}
a_{ii}^{(d)}\mathbf{Q}_{k+d}^{(i)} &= x^k\mathbf{e}_i - k\mathbf{Q}_{k-1}^{(i)} - \sum_{\substack{j=1\\ j\neq i}}^{\nu}[\sum_{m=0}^{d} a_{ji}^{(m)}\mathbf{Q}_{k+m}^{(j)}] - \sum_{m=0}^{d-1} a_{ii}^{(m)}\mathbf{Q}_{k+m}^{(i)} \\
&= x^k\mathbf{e}_i - k\mathbf{Q}_{k-1}^{i} - \sum_{\substack{j=1\\ j\neq i}}^{\nu}[\sum_{m=0}^{d-1} a_{ji}^{(m)}\mathbf{Q}_{k+m}^{(j)} + a_{ji}^{(d)}\mathbf{Q}_{k+d}^{(j)}] - \sum_{m=0}^{d-1} a_{ii}^{(m)}\mathbf{Q}_{k+m}^{(i)} \\
&= x^k\mathbf{e}_i - k\mathbf{Q}_{k-1}^{(i)} - \sum_{\substack{j=1\\ j\neq i}}^{\nu}[\sum_{m=0}^{d-1} a_{ji}^{(m)}\mathbf{Q}_{k+m}^{(j)}] - \sum_{\substack{j=1\\ j\neq i}}^{\nu} a_{ji}^{(d)}\mathbf{Q}_{k+d}^{(j)} - \sum_{m=0}^{d-1} a_{ii}^{(m)}\mathbf{Q}_{k+m}^{(i)} \\
&= x^k\mathbf{e}_i - k\mathbf{Q}_{k-1}^{(i)} - \sum_{j=1}^{\nu}[\sum_{m=0}^{d-1} a_{ji}^{(m)}\mathbf{Q}_{k+m}^{(j)}] - \sum_{\substack{j=1\\ j\neq i}}^{\nu} a_{ji}^{(d)}\mathbf{Q}_{k+d}^{(j)},
\end{aligned}$$

which gives

$$\begin{aligned}\underbrace{a_{ii}^{(d)}\mathbf{Q}_{k+d}^{(i)}+\sum_{\substack{j=1\\ j\neq i}}^{\nu}a_{ji}^{(d)}\mathbf{Q}_{k+d}^{(j)}} &= x^k\mathbf{e}_i-k\mathbf{Q}_{k-1}^{(i)}-\sum_{j=1}^{\nu}[\sum_{m=0}^{d-1}a_{ji}^{(m)}\mathbf{Q}_{k+m}^{(j)}]\\ \sum_{j=1}^{\nu}a_{ji}^{(d)}\mathbf{Q}_{k+d}^{(j)} &= x^k\mathbf{e}_i-k\mathbf{Q}_{k-1}^{(i)}-\sum_{j=1}^{\nu}[\sum_{m=0}^{d-1}a_{ji}^{(m)}\mathbf{Q}_{k+m}^{(j)}]\\ &= x^k\mathbf{e}_i-k\mathbf{Q}_{k-1}^{(i)}-\sum_{m=0}^{d-1}[\sum_{j=1}^{\nu}a_{ji}^{(m)}\mathbf{Q}_{k+m}^{(j)}].\end{aligned}$$

Explicitly this means that for $i=1,2,...,\nu$

$$[a_{1i}^{(d)}\ \ a_{2i}^{(d)}\ \ ...\ \ a_{\nu i}^{(d)}]\begin{bmatrix}\mathbf{Q}_{k+d}^{(1)}\\ \mathbf{Q}_{k+d}^{(2)}\\ \vdots\\ \mathbf{Q}_{k+d}^{(\nu)}\end{bmatrix} = x^k\mathbf{e}_i-k\mathbf{Q}_{k-1}^{(i)}-\sum_{m=0}^{d-1}[a_{1i}^{(m)}\ \ a_{2i}^{(m)}\ \ ...\ \ a_{\nu i}^{(m)}]\begin{bmatrix}\mathbf{Q}_{k+m}^{(1)}\\ \mathbf{Q}_{k+m}^{(2)}\\ \vdots\\ \mathbf{Q}_{k+m}^{(\nu)}\end{bmatrix},$$

and therefore

$$\begin{bmatrix}a_{11}^{(d)} & a_{21}^{(d)} & \dots & a_{\nu 1}^{(d)}\\ a_{12}^{(d)} & a_{22}^{(d)} & \dots & a_{\nu 2}^{(d)}\\ \vdots & \vdots & \ddots & \vdots\\ a_{1\nu}^{(d)} & a_{2\nu}^{(d)} & \dots & a_{\nu\nu}^{(d)}\end{bmatrix}\begin{bmatrix}\mathbf{Q}_{k+d}^{(1)}\\ \mathbf{Q}_{k+d}^{(2)}\\ \vdots\\ \mathbf{Q}_{k+d}^{(\nu)}\end{bmatrix}=x^k\begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\\ \vdots\\ \mathbf{e}_\nu\end{bmatrix}-k\begin{bmatrix}\mathbf{Q}_{k-1}^{(1)}\\ \mathbf{Q}_{k-1}^{(2)}\\ \vdots\\ \mathbf{Q}_{k-1}^{(\nu)}\end{bmatrix}-\sum_{m=0}^{d-1}\begin{bmatrix}a_{11}^{(m)} & a_{21}^{(m)} & \dots & a_{\nu 1}^{(m)}\\ a_{12}^{(m)} & a_{22}^{(m)} & \dots & a_{\nu 2}^{(m)}\\ \vdots & \vdots & \ddots & \vdots\\ a_{1\nu}^{(m)} & a_{2\nu}^{(m)} & \dots & a_{\nu\nu}^{(m)}\end{bmatrix}\begin{bmatrix}\mathbf{Q}_{k+m}^{(1)}\\ \mathbf{Q}_{k+m}^{(2)}\\ \vdots\\ \mathbf{Q}_{k+m}^{(\nu)}\end{bmatrix}.$$

In matricial form we have

$$\mathbf{A}_dQ_{k+d}^*=x^k\mathcal{E}_\nu-kQ_{k-1}^*-\sum_{m=0}^{d-1}\mathbf{A}_mQ_{k+m}^*.$$

Since $\mathbf{A}_d$ is non-singular, we obtain the desired recursion

$$Q_{k+d}^*=\mathbf{A}_d^{-1}\left[x^k\mathcal{E}_\nu-kQ_{k-1}^*-\sum_{m=0}^{d-1}\mathbf{A}_mQ_{k+m}^*\right];\ \ k\geq 0,$$

and this completes the proof of (5).

Although all the Q_k^*'s satisfy the self starting recursive formula (5), in practice only $\{Q_k^*;k\geq d\}$ can be generated by this recursion while the remaining ones $\{Q_0^*,Q_1^*,...,Q_{d-1}^*\}$ are not computable by the same formula. These are then called *undefined*. This point will be clarified further in the following example which illustrates Algorithm (5):

Example 1. Consider the differential operator **D** defined by (1) with

$$\mathbf{A}(x)=\begin{pmatrix}\frac{16x}{5}-\frac{1}{5} & \frac{8}{5}-\frac{28x}{5} & 1-2x\\ \frac{2x}{5}+\frac{1}{10} & \frac{34x}{5}-\frac{4}{5} & 6x-\frac{1}{2}\\ -\frac{6x}{5} & -\frac{12x}{5} & -4x\end{pmatrix}.$$

Since all the entries of **A** are of order 1, $d=1$ and therefore (5) becomes

$$Q_{k+1}^*=\mathbf{A}_1^{-1}\left[x^k\mathcal{E}_3-kQ_{k-1}^*-\mathbf{A}_0Q_k^*\right];\ \ k\geq 0, \tag{6}$$

where

$$\mathbf{A}_1 = \begin{pmatrix} \frac{16}{5} & -\frac{28}{5} & -2 \\ \frac{2}{5} & \frac{34}{5} & 6 \\ -\frac{6}{5} & -\frac{12}{5} & -4 \end{pmatrix} \text{ and } \mathbf{A}_0 = \begin{pmatrix} -\frac{1}{5} & \frac{8}{5} & 1 \\ \frac{1}{10} & -\frac{4}{5} & -\frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}.$$

It is clearly seen that $Q_0^* := \{\mathbf{Q}_0^{(1)}, \mathbf{Q}_0^{(2)}, \mathbf{Q}_0^{(3)}\}$ is undefined because it can not be obtained from (6). However the execution of (6) for $k \geq 1$ produces $\{Q_1^*, Q_2^*, Q_3^*, \dots\}$ some of which are:

1) $Q_1^* := \{\mathbf{Q}_1^{(1)}, \mathbf{Q}_1^{(2)}, \mathbf{Q}_1^{(3)}\}$:

$$\begin{aligned} \mathbf{Q}_1^{(1)} &= \frac{1}{30}\mathbf{Q}_0^{(1)} - \frac{1}{60}\mathbf{Q}_0^{(2)} + \frac{8\mathbf{e}_1}{15} + \frac{7\mathbf{e}_2}{30} - \frac{3\mathbf{e}_3}{10} \\ \mathbf{Q}_1^{(2)} &= -\frac{4}{15}\mathbf{Q}_0^{(1)} + \frac{2}{15}\mathbf{Q}_0^{(2)} + \frac{11\mathbf{e}_1}{15} + \frac{19\mathbf{e}_2}{30} - \frac{3\mathbf{e}_3}{5} \\ \mathbf{Q}_1^{(3)} &= -\frac{1}{6}\mathbf{Q}_0^{(1)} + \frac{1}{12}\mathbf{Q}_0^{(2)} + \frac{5\mathbf{e}_1}{6} + \frac{5\mathbf{e}_2}{6} - \mathbf{e}_3. \end{aligned}$$

2) $Q_2^* := \{\mathbf{Q}_2^{(1)}, \mathbf{Q}_2^{(2)}, \mathbf{Q}_2^{(3)}\}$:

$$\begin{aligned} \mathbf{Q}_2^{(1)} &= -\frac{19}{36}\mathbf{Q}_0^{(1)} - \frac{17}{72}\mathbf{Q}_0^{(2)} + \frac{3}{10}\mathbf{Q}_0^{(3)} + \frac{8}{15}\mathbf{e}_1 x + \frac{7}{30}\mathbf{e}_2 x - \frac{3}{10}\mathbf{e}_3 x + \frac{\mathbf{e}_1}{180} - \frac{\mathbf{e}_2}{360} \\ \mathbf{Q}_2^{(2)} &= -\frac{7}{9}\mathbf{Q}_0^{(1)} - \frac{11}{18}\mathbf{Q}_0^{(2)} + \frac{3}{5}\mathbf{Q}_0^{(3)} + \frac{11}{15}\mathbf{e}_1 x + \frac{19}{30}\mathbf{e}_2 x - \frac{3}{5}\mathbf{e}_3 x - \frac{2\mathbf{e}_1}{45} + \frac{\mathbf{e}_2}{45} \\ \mathbf{Q}_2^{(3)} &= -\frac{31}{36}\mathbf{Q}_0^{(1)} - \frac{59}{72}\mathbf{Q}_0^{(2)} + \mathbf{Q}_0^{(3)} + \frac{5}{6}\mathbf{e}_1 x + \frac{5}{6}\mathbf{e}_2 x - \mathbf{e}_3 x - \frac{\mathbf{e}_1}{36} + \frac{\mathbf{e}_2}{72}. \end{aligned}$$

3) $Q_3^* := \{\mathbf{Q}_3^{(1)}, \mathbf{Q}_3^{(2)}, \mathbf{Q}_3^{(3)}\}$:

$$\begin{aligned} \mathbf{Q}_3^{(1)} &= -\frac{17\mathbf{Q}_0^{(1)}}{1080} + \frac{17\mathbf{Q}_0^{(2)}}{2160} + \frac{8}{15}\mathbf{e}_1 x^2 + \frac{7}{30}\mathbf{e}_2 x^2 - \frac{3}{10}\mathbf{e}_3 x^2 + \frac{1}{180}\mathbf{e}_1 x - \frac{1}{360}\mathbf{e}_2 x - \frac{443\mathbf{e}_1}{1080} - \frac{97\mathbf{e}_2}{2160} \\ \mathbf{Q}_3^{(2)} &= \frac{17}{135}\mathbf{Q}_0^{(1)} - \frac{17}{270}\mathbf{Q}_0^{(2)} + \frac{11}{15}\mathbf{e}_1 x^2 + \frac{19}{30}\mathbf{e}_2 x^2 - \frac{3}{5}\mathbf{e}_3 x^2 - \frac{2}{45}\mathbf{e}_1 x + \frac{1}{45}\mathbf{e}_2 x - \frac{97\mathbf{e}_1}{135} - \frac{19\mathbf{e}_2}{135} \\ \mathbf{Q}_3^{(3)} &= \frac{17}{216}\mathbf{Q}_0^{(1)} - \frac{17}{432}\mathbf{Q}_0^{(2)} + \frac{5}{6}\mathbf{e}_1 x^2 + \frac{5}{6}\mathbf{e}_2 x^2 - \mathbf{e}_3 x^2 - \frac{1}{36}\mathbf{e}_1 x + \frac{1}{72}\mathbf{e}_2 x - \frac{97\mathbf{e}_1}{216} + \frac{97\mathbf{e}_2}{432} - \frac{\mathbf{e}_3}{2}. \end{aligned}$$

One observes that every Q_k^*, $k \geq 1$, contains a linear combination of the undefined block $Q_0^* := \{\mathbf{Q}_0^{(1)}, \mathbf{Q}_0^{(2)}, \mathbf{Q}_0^{(3)}\}$. This characteristic is not confined to the differential operator discussed in Example 1. In fact, for any operator of the form (1), every canonical block Q_k^*, $k \geq d$ must involve a component formed of some undefined canonical blocks $\{Q_0^*, Q_1^*, ..., Q_{d-1}^*\}$. To justify this claim:

Execute (5) when $k = 0$:

$$\mathbf{A}_d Q_d^* = \mathcal{E}_\nu - \mathbf{A}_0 Q_0^* - \mathbf{A}_1 Q_1^* - \cdots - \mathbf{A}_{d-1} Q_{d-1}^*.$$

Similarly $k = 1$ gives:

$$\begin{aligned} \mathbf{A}_d Q_{d+1}^* &= x\mathcal{E}_\nu - Q_0^* - \sum_{m=0}^{d-1} \mathbf{A}_m Q_{m+1}^* \\ &= x\mathcal{E}_\nu - Q_0^* - \left(\mathbf{A}_0 Q_1^* + \mathbf{A}_1 Q_2^* + \cdots + \mathbf{A}_{d-1} Q_d^*\right) \\ &= x\mathcal{E}_\nu - Q_0^* - \mathbf{A}_0 Q_1^* - \mathbf{A}_1 Q_2^* - \cdots - \mathbf{A}_{d-1}\mathbf{A}_d^{-1}\left[\mathcal{E}_\nu - \mathbf{A}_0 Q_0^* - \mathbf{A}_1 Q_1^* - \cdots - \mathbf{A}_{d-1} Q_{d-1}^*\right] \\ &= x\mathcal{E}_\nu - \mathbf{A}_{d-1}\mathbf{A}_d^{-1}\mathcal{E}_\nu - Q_0^* - \mathbf{A}_0 Q_1^* - \mathbf{A}_1 Q_2^* - \cdots + \mathbf{A}_{d-1}\mathbf{A}_d^{-1}\left[\mathbf{A}_0 Q_0^* + \mathbf{A}_1 Q_1^* + \cdots + \mathbf{A}_{d-1} Q_{d-1}^*\right] \end{aligned}$$

$$= \quad x\mathcal{E}_\nu - \mathbf{A}_{d-1}\mathbf{A}_d^{-1}\mathcal{E}_\nu + \mathbf{R}_0\mathcal{Q}_0^* + \mathbf{R}_1\mathcal{Q}_1^* + ... + \mathbf{R}_{d-1}\mathcal{Q}_{d-1}^*,$$

for some constant square matrices $\{\mathbf{R}_j\}$ depending on $\{\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{d-1}\}$. Proceeding this way we obtain by induction the following corollary:

Corollary 2 *Let* $\{\mathbf{R}_k^{(j)}, k \geq 0,\ s = 0, 1, ..., d-1\}$ *be sequences of* $\nu \times \nu$ *matrices defined by the recursion:*

$$\begin{aligned} \mathbf{R}_s^{(j)} &= \mathbf{I}_\nu;\ s, j = 0, 1, 2, ..., d-1, \\ \mathbf{R}_s^{(k+d)} &= \mathbf{A}_d^{-1}\left[-k\mathbf{R}_s^{(k-1)} - \sum_{m=0}^{d-1}\mathbf{A}_m\mathbf{R}_s^{(k+m)}\right];\ \ k \geq 0 \text{ and } s = 0, 1, 2, ..., d-1. \end{aligned}$$

Then each $\mathcal{Q}_k^*$, $k \geq 1$, *can be written as*

$$\mathcal{Q}_k^* = \mathcal{Q}_k + \sum_{s=0}^{d-1}\mathbf{R}_s^{(k)}\mathcal{Q}_s^*, \tag{7}$$

where $\{\mathcal{Q}_k, k \geq 0,\ j = 1, 2, ..., \nu\}$ *are defined by the recursion*

$$\begin{aligned} \mathcal{Q}_k &= [\mathbf{0}\ \ \mathbf{0}\ \ \ldots\ \ \mathbf{0}]^T;\ \ k = 0, 1, 2, ..., d-1, \\ \mathcal{Q}_{k+d} &= \mathbf{A}_d^{-1}\left[x^k\mathcal{E}_\nu - k\mathcal{Q}_{k-1} - \sum_{m=0}^{d-1}\mathbf{A}_m\mathcal{Q}_{k+m}\right];\ \ k \geq 0. \end{aligned}$$

The proof this corollary follows from the fact that sequence $\{\mathcal{Q}_k^*, k \geq 0\}$ is unique by construction and from comparing both sides of the identity:

$$\mathcal{Q}_{k+d} + \sum_{s=0}^{d-1}\mathbf{R}_s^{(k+d)}\mathcal{Q}_s^* = \mathbf{A}_d^{-1}\left[x^k\mathcal{E}_\nu - k\left(\mathcal{Q}_{k-1} + \sum_{s=0}^{d-1}\mathbf{R}_s^{(k-1)}\mathcal{Q}_s^*\right) - \sum_{m=0}^{d-1}\mathbf{A}_m\left(\mathcal{Q}_{k+m} + \sum_{s=0}^{d-1}\mathbf{R}_s^{(k+m)}\mathcal{Q}_s^*\right)\right].$$

3. Construction of the Tau Method Approximation

In the Tau method we associate to (1)-(4) a perturbed problem of the form

$$\mathbf{D}\mathbf{Y}_N(x) := \left[\mathbf{I}_\nu\frac{d}{dx} + \mathbf{A}(x)\right]\mathbf{Y}_N(x) = \mathbf{f}(x) + \mathbf{H}_N(x), \tag{8}$$

$$\mathbf{Y}_N(0) = \mathbf{y}_0, \tag{9}$$

where $\mathbf{H}_N$ is a free perturbation term adjusted in a way that $\mathbf{Y}_N$ is a vector of polynomials that can be obtained analytically. Usually $\mathbf{H}_N$ is chosen either of the form

$$\mathbf{H}_N(x) = \boldsymbol{\tau}_0 V_N + \boldsymbol{\tau}_1 V_{N+1} + \cdots + \boldsymbol{\tau}_{r-1} V_{N+r-1}, \tag{10}$$

or

$$\mathbf{H}_N(x) = (\boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 x + \cdots + \boldsymbol{\tau}_{r-1} x^{r-1})V_N, \tag{11}$$

with N being a prescribed positive integer, and

$$\boldsymbol{\tau}_j := [\tau_j^{(1)}\ \ \tau_j^{(2)}\ \ \ldots\ \ \tau_j^{(\nu)}]^T,\quad j = 0, 1, 2, \ldots, r-1,$$

are r vectors each one of which is formed of ν free parameters where the positive integer r will be fixed later. $V_m(x)$ designates a polynomial of degree m that is usually chosen as the Chebyshev polynomial $T_m(x)$ or Legendre polynomial $P_m(x)$ shifted to the appropriate interval:

$$V_m(x) = \sum_{j=0}^{m} c_j^{(m)} x^j.$$

The unknown vectors $\{\boldsymbol{\tau}_j;\ j = 1, 2, \ldots, r-1\}$ are determined (i) by imposing the initial condition (9) on the desired approximate solution $\mathbf{Y}_N$, and (ii) by forcing $\mathbf{Y}_N$ to be independent of the undefined terms $\{\mathbf{Q}_j^*;\ j = 0, 1, \ldots, d-1\}$. The latter can be realized by setting the coefficient of each $\mathbf{Q}_j^*;\ j = 0, 1, \ldots, d-1$ equal to zero.

Since (9) is the only initial condition to be satisfied, and since there are d undefined canonical blocks $\{Q_0^*, Q_1^*, ..., Q_{d-1}^*\}$, we need $d+1$ vectors $\boldsymbol{\tau}_j$'s only i.e. we choose $r = d+1$.

Having decided the value of r, it becomes possible now to construct $\mathbf{Y}_N$ simultaneously with $\{\boldsymbol{\tau}_j;\, j = 1, 2, \ldots, r-1\}$:

Theorem 3 *Assume that $\mathbf{H}_N$ is of the from (10). Then an exact polynomial solution for the perturbed problem (8)-(9) is given as*

$$\mathbf{Y}_N(x) = \sum_{i=0}^{\alpha} \mathbf{f}_i^T Q_i(x) + \sum_{k=0}^{d} \sum_{i=0}^{N+k} c_i^{(N+k)} \boldsymbol{\tau}_k^T Q_i(x), \tag{12}$$

where each $\boldsymbol{\tau}_k^T$, $k = 0, 1, \ldots, d$, is a ν dimensional vector fixed by the linear system of algebraic equations:

$$\sum_{k=0}^{d} \Phi_{k,s} \boldsymbol{\tau}_k = \rho_s; \quad s = 0, 1, 2, ..., d-1, \tag{13}$$

$$\sum_{k=0}^{d} \Psi_k \boldsymbol{\tau}_k = \rho_d, \tag{14}$$

with

$$\Phi_{k,s} := \left[\sum_{i=0}^{N+k} c_i^{(N+k)} \mathbf{R}_s^{(i)}\right]^T, \qquad \rho_s := -\left[\sum_{i=0}^{\alpha} \mathbf{f}_i^T \mathbf{R}_s^{(i)}\right]^T,$$
$$\Psi_k := \left[\sum_{i=0}^{N+k} c_i^{(N+k)} Q_i(0)\right]^T, \qquad \rho_d := \mathbf{y}_0 - \left[\sum_{i=0}^{\alpha} \mathbf{f}_i^T Q_i(0)\right]^T,$$

for $k = 0, 1, \ldots, d$ and $s = 0, 1, \ldots, d-1$.

Proof. This follows once the right hand side of (8), $\mathbf{f}(x) + \mathbf{H}_N(x)$, is expressed in terms of $\{Q_i^*\}$. Let us consider first $\mathbf{H}_N(x)$:

Noting that $\boldsymbol{\tau}_j := [\tau_j^{(1)} \;\; \tau_j^{(2)} \;\; \ldots \;\; \tau_j^{(\nu)}]^T = \sum_{i=1}^{\nu} \tau_j^{(i)} \mathbf{e}_i$ and $V_m = \sum_{i=0}^{m} c_i^{(m)} x^i$, then $\mathbf{H}_N(x)$ introduced in (10) can be written as

$$\mathbf{H}_N(x) := \sum_{k=0}^{d} \boldsymbol{\tau}_k V_{N+k} = \sum_{k=0}^{d} \sum_{j=1}^{\nu} \tau_k^{(j)} V_{N+k} \mathbf{e}_j$$

$$= \sum_{k=0}^{d} \sum_{j=1}^{\nu} \tau_k^{(j)} \left(\sum_{i=0}^{N+k} c_i^{(N+k)} x^i\right) \mathbf{e}_j = \sum_{k=0}^{d} \sum_{j=1}^{\nu} \tau_k^{(j)} \left(\sum_{i=0}^{N+k} c_i^{(N+k)} x^i \mathbf{e}_j\right).$$

Since $\mathbf{D}\mathbf{Q}_i^{(j)} = x^i \mathbf{e}_j$ and $\mathbf{D}$ is linear, $\mathbf{H}_N(x)$ becomes

$$\mathbf{H}_N(x) = \sum_{k=0}^{d} \sum_{j=1}^{\nu} \tau_k^{(j)} \left(\sum_{i=0}^{N+k} c_i^{(N+k)} \mathbf{D}\mathbf{Q}_i^{(j)}\right) = \mathbf{D}\left[\sum_{k=0}^{d} \sum_{j=1}^{\nu} \tau_k^{(j)} \left(\sum_{i=0}^{N+k} c_i^{(N+k)} \mathbf{Q}_i^{(j)}\right)\right]$$
$$= \mathbf{D}\left[\sum_{k=0}^{d} \sum_{i=0}^{N+k} c_i^{(N+k)} \left(\sum_{j=1}^{\nu} \tau_k^{(j)} \mathbf{Q}_i^{(j)}\right)\right] = \mathbf{D}\left[\sum_{k=0}^{d} \sum_{i=0}^{N+k} c_i^{(N+k)} \boldsymbol{\tau}_k^T Q_i^*\right].$$

Hence,

$$\mathbf{H}_N(x) = \mathbf{D}\left[\sum_{k=0}^{d} \sum_{i=0}^{N+k} c_i^{(N+k)} \boldsymbol{\tau}_k^T Q_i^*\right]. \tag{15}$$

The same arguments apply to the function vector $\mathbf{f}(x)$ given by (2). Suppose that each entry in $\mathbf{f}(x)$ is a polynomial

of degree α written as $f_j(x) = \sum_{i=0}^{\alpha} f_{ji}x^i$. Then

$$\mathbf{f}(x) := \begin{bmatrix} \sum_{i=0}^{\alpha} f_{1i}x^i \\ \sum_{i=0}^{\alpha} f_{2i}x^i \\ \vdots \\ \sum_{i=0}^{\alpha} f_{\nu i}x^i \end{bmatrix} = \sum_{j=1}^{\nu}\sum_{i=0}^{\alpha} f_{ji}x^i\mathbf{e}_j = \sum_{j=1}^{\nu}\sum_{i=0}^{\alpha} f_{ji}\mathbf{D}\mathbf{Q}_i^{(j)} = \mathbf{D}\left[\sum_{j=1}^{\nu}\sum_{i=0}^{\alpha} f_{ji}\mathbf{Q}_i^{(j)}\right].$$

But

$$\sum_{j=1}^{\nu}\sum_{i=0}^{\alpha} f_{ji}\mathbf{Q}_i^{(j)} = \sum_{i=0}^{\alpha}\sum_{j=1}^{\nu} f_{ji}\mathbf{Q}_i^{(j)} = \sum_{i=0}^{\alpha}[f_{1i} \quad f_{2i} \quad \dots \quad f_{\nu i}] \begin{bmatrix} \mathbf{Q}_i^{(1)} \\ \mathbf{Q}_i^{(2)} \\ \vdots \\ \mathbf{Q}_i^{(\nu)} \end{bmatrix} = \sum_{i=0}^{\alpha} \mathbf{f}_i^T Q_i^*.$$

Therefore

$$\mathbf{f}(x) = \mathbf{D}\left[\sum_{i=0}^{\alpha} \mathbf{f}_i^T Q_i^*\right]. \tag{16}$$

Adding (15) and (16) we get

$$\mathbf{f}(x) + \mathbf{H}_N(x) = \mathbf{D}\left[\sum_{i=0}^{\alpha} \mathbf{f}_i^T Q_i^* + \sum_{k=0}^{d}\sum_{i=0}^{N+k} c_i^{(N+k)}\boldsymbol{\tau}_k^T Q_i^*\right]. \tag{17}$$

Thus the Tau problem (8) is written now in the form

$$\mathbf{D}\mathbf{Y}_N(x) = \mathbf{f}(x) + \mathbf{H}_N(x) = \mathbf{D}\left[\sum_{i=0}^{\alpha} \mathbf{f}_i^T Q_i^* + \sum_{k=0}^{d}\sum_{i=0}^{N+k} c_i^{(N+k)}\boldsymbol{\tau}_k^T Q_i^*\right]$$

which implies that $\mathbf{Y}_N$ is formally given by:

$$\mathbf{Y}_N(x) = \sum_{i=0}^{\alpha} \mathbf{f}_i^T Q_i^* + \sum_{k=0}^{d}\sum_{i=0}^{N+k} c_i^{(N+k)}\boldsymbol{\tau}_k^T Q_i^*.$$

Further, using (7) we can write $\mathbf{Y}_N$ in terms of $\{Q_i\}$:

$$\begin{aligned} \mathbf{Y}_N(x) &= \sum_{i=0}^{\alpha} \mathbf{f}_i^T \left(Q_i + \sum_{s=0}^{d-1} \mathbf{R}_s^{(i)} Q_s^*\right) + \sum_{k=0}^{d}\sum_{i=0}^{N+k} c_i^{(N+k)}\boldsymbol{\tau}_k^T \left(Q_i + \sum_{s=0}^{d-1} \mathbf{R}_s^{(i)} Q_s^*\right) \\ &= \sum_{i=0}^{\alpha} \mathbf{f}_i^T Q_i + \sum_{k=0}^{d}\sum_{i=0}^{N+k} c_i^{(N+k)}\boldsymbol{\tau}_k^T Q_i \\ &+ \sum_{i=0}^{\alpha} \mathbf{f}_i^T \left(\sum_{s=0}^{d-1} \mathbf{R}_s^{(i)} Q_s^*\right) + \sum_{k=0}^{d}\sum_{i=0}^{N+k} c_i^{(N+k)}\boldsymbol{\tau}_k^T \left(\sum_{s=0}^{d-1} \mathbf{R}_s^{(i)} Q_s^*\right) \\ &= \sum_{i=0}^{\alpha} \mathbf{f}_i^T Q_i + \sum_{k=0}^{d}\sum_{i=0}^{N+k} c_i^{(N+k)}\boldsymbol{\tau}_k^T Q_i \qquad (18) \\ &+ \sum_{s=0}^{d-1} \left\{\sum_{i=0}^{\alpha} \mathbf{f}_i^T \mathbf{R}_s^{(i)} + \sum_{k=0}^{d}\sum_{i=0}^{N+k} c_i^{(N+k)}\boldsymbol{\tau}_k^T R_s^{(i)}\right\} Q_s^*. \qquad (19) \end{aligned}$$

This expression holds for any choice of $\{\boldsymbol{\tau}_k\}$. Since (19) contains undefined canonical blocks only we call it residual. In order to eliminate this residual we set its coefficients equal to zero. That is for all $s = 0, 1, \dots, d-1$,

$$\sum_{k=0}^{d}\sum_{i=0}^{N+k} c_i^{(N+k)}\boldsymbol{\tau}_k^T \mathbf{R}_s^{(i)} = -\sum_{i=0}^{\alpha} \mathbf{f}_i^T \mathbf{R}_s^{(i)}.$$

Equivalently,

$$\sum_{k=0}^{d} \boldsymbol{\tau}_k^T \left[\sum_{i=0}^{N+k} c_i^{(N+k)} \mathbf{R}_s^{(i)} \right] = -\sum_{i=0}^{\alpha} \mathbf{f}_i^T \mathbf{R}_s^{(i)}.$$

Setting $\Phi_{k,s} := \left[\sum_{i=0}^{N+k} c_i^{(N+k)} \mathbf{R}_s^{(i)}\right]^T$ and $\rho_s := -\left[\sum_{i=0}^{\alpha} \mathbf{f}_i^T \mathbf{R}_s^{(i)}\right]^T$, we obtain (13)

$$\sum_{k=0}^{d} \Phi_{k,s} \boldsymbol{\tau}_k = \rho_s, \quad s = 0, 1, 2, ..., d-1.$$

With this choice of the $\boldsymbol{\tau}_j$'s, (18) reduces to (12):

$$\mathbf{Y}_N(x) = \sum_{i=0}^{\alpha} \mathbf{f}_i^T \mathcal{Q}_i + \sum_{k=0}^{d} \sum_{i=0}^{N+k} c_i^{(N+k)} \boldsymbol{\tau}_k^T \mathcal{Q}_i.$$

Imposing the initial condition (9) we get:

$$\mathbf{Y}_N(0) = \sum_{i=0}^{\alpha} \mathbf{f}_i^T \mathcal{Q}_i(0) + \sum_{k=0}^{d} \sum_{i=0}^{N+k} c_i^{(N+k)} \boldsymbol{\tau}_k^T \mathcal{Q}_i(0) = \mathbf{y}_0$$

which, in turn, leads to

$$\sum_{k=0}^{d} \boldsymbol{\tau}_k^T \left[\sum_{i=0}^{N+k} c_i^{(N+k)} \mathcal{Q}_i(0) \right] = \mathbf{y}_0 - \sum_{i=0}^{\alpha} \mathbf{f}_i^T \mathcal{Q}_i(0).$$

Setting $\Psi_k := \left[\sum_{i=0}^{N+k} c_i^{(N+k)} \mathcal{Q}_i(0)\right]^T$ and $\rho_d := \mathbf{y}_0 - \left[\sum_{i=0}^{\alpha} \mathbf{f}_i^T \mathcal{Q}_i(0)\right]^T$, we obtain (14)

$$\sum_{k=0}^{d} \Psi_k \boldsymbol{\tau}_k = \rho_d.$$

This completes the proof of the theorem.

The following corollary is a particular case of the previous theorem:

Corollary 4 *The assumptions of the previous theorem hold. If further $d = 0$ then*

$$\mathbf{Y}_N(x) = \sum_{i=0}^{\alpha} \mathbf{f}_i^T \mathcal{Q}_i(x) + \boldsymbol{\tau}_0^T \sum_{i=0}^{N} c_i^{(N)} \mathcal{Q}_i(x)$$

where

$$\boldsymbol{\tau}_0^T = \left[\mathbf{y}_0^T - \left(\sum_{i=0}^{\alpha} \mathbf{f}_i^T \mathcal{Q}_i(x_0) \right) \right] \left[\sum_{i=0}^{N} c_i^{(N)} \mathcal{Q}_i(x_0) \right]^{-1}.$$

Proof. Since $d = 0$, all the canonical elements are defined and therefore the result is obtained from Theorem 1.

Example 2. Let us solve the initial value problem

$$\begin{aligned} \mathbf{D}\mathbf{y}(x) &:= \left[\mathbf{I}_\nu \frac{d}{dx} + \mathbf{A}(x) \right] \mathbf{y}(x) = \mathbf{f}(x); \quad x \in [0, 1] \\ & \mathbf{y}(0) = \mathbf{y}_0 := [7 + e, \quad e - 1, \quad -1 - 3e]^T, \end{aligned}$$

where $\mathbf{A}(x)$ is given in Example 1 and

$$\mathbf{f}(x) = \begin{pmatrix} -\frac{18x^2}{5} + \frac{63x}{5} - 2 \\ \frac{4x^2}{5} - \frac{63x}{10} + 2 \\ \frac{8x^2}{5} - 1 \end{pmatrix}.$$

The exact solution is:

$$\begin{aligned} y_1(x) &= 3e^{-x^2} + 2e^{x-3x^2} + e^{x^2+1} + 2, \\ y_2(x) &= e^{-x^2} - e^{x-3x^2} + 2e^{x^2+1} + x - 1, \\ y_3(x) &= -e^{-x^2} - 3e^{x^2+1} - x. \end{aligned}$$

Here $d = 1$ and therefore $Q_0^* := \{\mathbf{Q}_0^{(1)}, \mathbf{Q}_0^{(2)}, \mathbf{Q}_0^{(3)}\}$ is the only undefined block. So we choose a perturbation term of the form:

$$\mathbf{H}_N(x) = \mathbf{f}(x) + \boldsymbol{\tau}_0 V_N(x) + \boldsymbol{\tau}_1 V_{N+1} = \mathbf{f}(x) + \begin{bmatrix} \tau_0^{(1)} \\ \tau_0^{(2)} \\ \tau_0^{(3)} \end{bmatrix} V_N(x) + \begin{bmatrix} \tau_1^{(1)} \\ \tau_1^{(2)} \\ \tau_1^{(3)} \end{bmatrix} V_{N+1}(x).$$

We applied the algorithm presented in Theorem 3 with $N = 10$ and $V_N = T_N(x)$, the Chebyshev polynomial. We found that

$$\begin{bmatrix} \tau_0^{(1)} \\ \tau_0^{(2)} \\ \tau_0^{(3)} \end{bmatrix} = \begin{bmatrix} 6.614722\text{E-}6 \\ -3.741327\text{E-}6 \\ 5.303844\text{E-}7 \end{bmatrix} \text{ and } \begin{bmatrix} \tau_1^{(1)} \\ \tau_1^{(2)} \\ \tau_1^{(3)} \end{bmatrix} = \begin{bmatrix} 5.878458\text{E-}7 \\ -3.341887\text{E-}7 \\ 4.921872\text{E-}8 \end{bmatrix}.$$

Figure 1 displays the exact errors in y_1, y_2 and y_3 while Figure 2 shows the three components of the perturbation term $H_N(x)$.

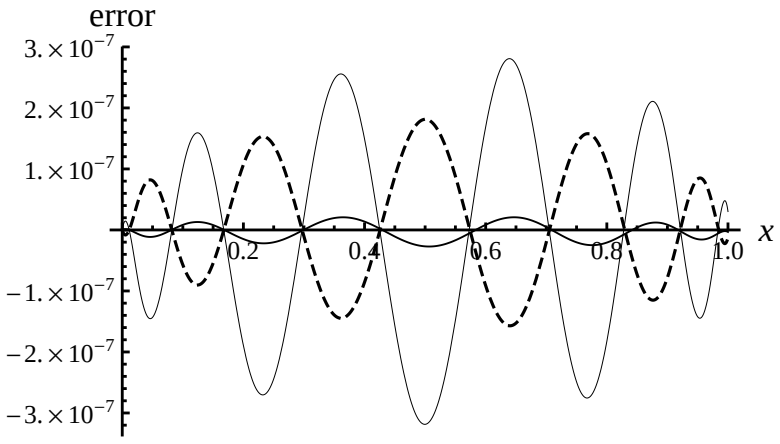

Figure 1.

Description: (Example 2): Plot of the exact errors in $y_1(x)$ (light), $y_2(x)$ (dashed), $y_3(x)$ (thick). Here N=10.

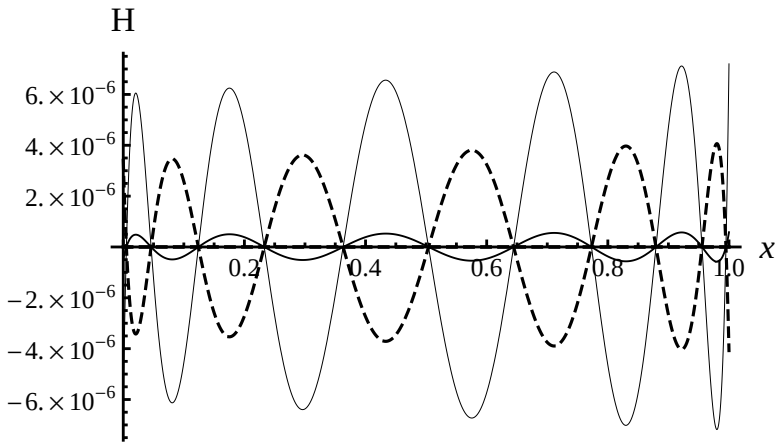

Figure 2.

Description: (Example 2): Plot of the Tau perturbations: $H_{10}^1(x)$ (light), $H_{10}^2(x)$ (dashed), $H_{10}^3(x)$ (thick). Here N=10.

Throughout this section we considered $\mathbf{H}_N$ as in (10). Following the same arguments we can derive analogous results when $\mathbf{H}_N$ is of the form (11). It is worth to note that the Tau method with a perturbation of the form (11) is equivalent to the spectral collocation method at the zeros of $V_N(x)$, see (El-Daou & Ortiz, 1994). This equivalence permits to solve nonlinear differential equations using the spectral approach more effectively than the recursive Tau. In the next section we recall the main features of the spectral Tau method and we illustrate it by solving an optimal control problem with constraints given as a system of nonlinear differential equations.

4. The Spectral Tau Method and Applications to Optimal Control Problem

Let us reconsider problem (1)-(4). In the spectral Tau method, the solution $\mathbf{y}(x) := [y_1(x), y_2(x), \ldots, y_\nu(x)]$ of (1)-(4) is approximated by a truncated series expansion of the form

$$\tilde{y}_i(x) = \sum_{j=0}^{N} c_{ij} V_j(x), \quad i = 1, 2, \ldots, \nu \tag{20}$$

where $\{c_{ij} \; ; i = 1, 2, \ldots, \nu, \;\; j = 0, 1, \ldots, N\}$ are the expansions coefficients. If the approximation $\tilde{\mathbf{y}}(x) := [\tilde{y}_1(x), \tilde{y}_2(x), \ldots, \tilde{y}_\nu(x)]$ is substituted in the differential equation (1), $(\mathbf{D}\tilde{\mathbf{y}} - \mathbf{f})(x)$ will not be identically zero unless $\tilde{\mathbf{y}}$ is the exact solution. Otherwise $(\mathbf{D}\tilde{\mathbf{y}} - \mathbf{f})(x)$ is called the residual.

The $(N + 1)\nu$ unknown coefficients $\{c_{ij} \; ; i = 1, 2, \ldots, \nu, \;\; j = 0, 1, \ldots, N\}$ are obtained by requiring the residual $\mathbf{D}\tilde{\mathbf{y}} - \mathbf{f}$ to be zero at the N zeros of $V_N(x)$, $\{\xi_k; k = 1, 2, \ldots, N\}$, that is

$$(\mathbf{D}\tilde{\mathbf{y}} - \mathbf{f})(\xi_k) = 0, \quad k = 1, 2, \ldots, N \qquad (N\nu \text{ equations}) \tag{21}$$

and by forcing $\tilde{\mathbf{y}}$ to satisfy the initial conditions (4), that is

$$\tilde{\mathbf{y}}(0) = \mathbf{y}_0 \qquad (\nu \text{ equations}). \tag{22}$$

Combining (21) and (22) one obtains a system of $\nu(N+1)$ equations with $\nu(N+1)$ unknowns $\{c_{ij} \; ; i = 1, 2, \ldots, \nu, \;\; j = 0, 1, \ldots, N\}$. Once the latter are obtained by solving (21)-(22), the spectral approximate solution can be computed from (20).

To illustrate the spectral procedure we shall consider the Hicks-Ray reactor problem where it is desired to minimize the quadrature cost functional

$$\int_0^{10} [a_1(X(t) - \bar{X})^2 + a_2(Y(t) - \bar{Y})^2 + a_3(U(t) - \bar{U})^2]dt \tag{23}$$

subject to the nonlinear constraints

$$X^{'}(t) = \theta(1-X) - \gamma X(t)e^{-r/Y}, \tag{24}$$
$$Y^{'}(t) = \theta(y_f - Y) + \gamma X(t)e^{-r/Y} - \alpha(Y(t) - y_c)U(t) \tag{25}$$

$$X(0) = x_0, \quad Y(0) = y_0. \tag{26}$$

The two states are denoted by $X(t)$ and $Y(t)$, the control is denoted by $U(t)$ and all other parameters are constants.

Our procedure starts with replacing the nonlinear constraints (24)-(25) by an infinite sequence of linear systems of ODEs of the form (1). This is accomplished by employing an iterative procedure described as follows:

Let $\phi(X,Y) := Xe^{-r/Y}$ and $W(t) := e^{-r/Y}$. The Taylor's series expansions of the bivariate function $\phi(X,Y)$ near a given point (X_0, Y_0) allows to write

$$\begin{aligned}
\phi(X,Y) &= \phi(X_0,Y_0) + \frac{\partial\phi}{\partial X}(X_0,Y_0)(X-X0) + \frac{\partial\phi}{\partial Y}(X_0,Y_0)(Y-Y_0) + O_2 \\
&= X_0e^{-r/Y_0} + e^{-r/Y_0}(X-X_0) + \frac{rX_0e^{-r/Y_0}}{Y_0^2}(Y-Y_0) + O_2 \\
&= X_0W_0 + W_0X - W_0X_0 + \frac{rX_0W_0}{Y_0^2}Y - \frac{rX_0W_0}{Y_0} + O_2, \qquad (W_0 = e^{-r/Y_0}) \\
&= W_0X + \frac{rX_0W_0}{Y_0^2}Y - \frac{rX_0W_0}{Y_0} + O_2,
\end{aligned} \tag{27}$$

where $O_2 = O(\|X-X_0\|^2) + O(\|Y-Y_0\|^2)$. Dropping O_2 and using (27) in (24) yields:

$$\begin{aligned}
\frac{dX}{dt} &\approx \theta(1-X) - \gamma[W_0X + \frac{rX_0W_0}{Y_0^2}Y - \frac{rX_0W_0}{Y_0}] \\
&\approx \theta - \theta X - \gamma W_0X - \frac{\gamma rX_0W_0}{Y_0^2}Y + \frac{\gamma rX_0W_0}{Y_0} \\
&\approx -(\theta + \gamma W_0)X - (\frac{\gamma rX_0W_0}{Y_0^2})Y + \theta + \frac{\gamma rX_0W_0}{T_0}.
\end{aligned} \tag{28}$$

Again using (27)) and the approximation

$$UY \approx U_0Y_0 + U_0(Y-Y_0) + Y_0(U-U_0) = U_0Y + Y_0U - Y_0U_0$$

in (25), we get

$$\frac{dY}{dt} \approx \gamma W_0X + (\theta + \frac{\gamma rX_0W_0}{Y_0^2} + \alpha U_0)]Y + \theta y_f - \frac{\gamma rX_0W_0}{Y_0} + \alpha(y_c - Y_0)U + \alpha Y_0U_0. \tag{29}$$

Setting $(X_k, Y_k, U_k) = (X, Y, U)$ and $(X_{k-1}, Y_{k-1}, U_{k-1}) = (X_0, Y_0, U_0)$ in equations (28)-(29) we obtain the sequence of linear systems

$$\frac{dX_k}{dt} + (\theta + \gamma W_{k-1})X_k + (\frac{\gamma rX_{k-1}W_{k-1}}{Y_{k-1}^2})Y_k = \theta + \frac{\gamma rX_{k-1}W_{k-1}}{Y_{k-1}}, k = 1,2,3,\ldots \tag{30}$$

$$\frac{dY_k}{dt} - \gamma W_{k-1}X_k - (\theta + \frac{\gamma rX_{k-1}W_{k-1}}{Y_{k-1}} + \alpha U_{k-1})Y_k = \theta y_f - \frac{\gamma rX_{k-1}W_{k-1}}{Y_{k-1}} + \alpha(y_c - Y_{k-1})U_k + \alpha Y_{k-1}U_{k-1}, \tag{31}$$

which is of the form (1) with

$$\mathbf{A}_{k-1}(x) = \begin{pmatrix} (\theta + \gamma W_{k-1}) & -(\frac{\gamma rX_{k-1}W_{k-1}}{Y_{k-1}}) \\ -\gamma W_{k-1} & -(\theta + \frac{\gamma rX_{k-1}W_{k-1}}{Y_{k-1}} + \alpha U_{k-1}) \end{pmatrix}$$

and

$$\mathbf{f}_{k-1}(x) = \begin{pmatrix} \theta + \frac{\gamma rX_{k-1}W_{k-1}}{Y_{k-1}} \\ \theta y_f - \frac{\gamma rX_{k-1}W_{k-1}}{Y_{k-1}} + \alpha(y_c - Y_{k-1})U_k + \alpha Y_{k-1}U_{k-1} \end{pmatrix}.$$

In a more compact form we have the sequence of linear systems

$$\begin{aligned} \mathbf{D}_k\mathbf{y}_k(t) \quad &:= \quad \left[\mathbf{I}_\nu \frac{d}{dt} + \mathbf{A}_{k-1}(t)\right]\mathbf{y}_k(t) = \mathbf{f}_{k-1}(t),; \quad t \in [0,10] \\ &\mathbf{y}_k(0) = \mathbf{y}_0, \quad k = 1,2,3,\ldots. \end{aligned}$$

where $\mathbf{y}_k = [X_k(t) \;\; Y_k(t)]^T$.

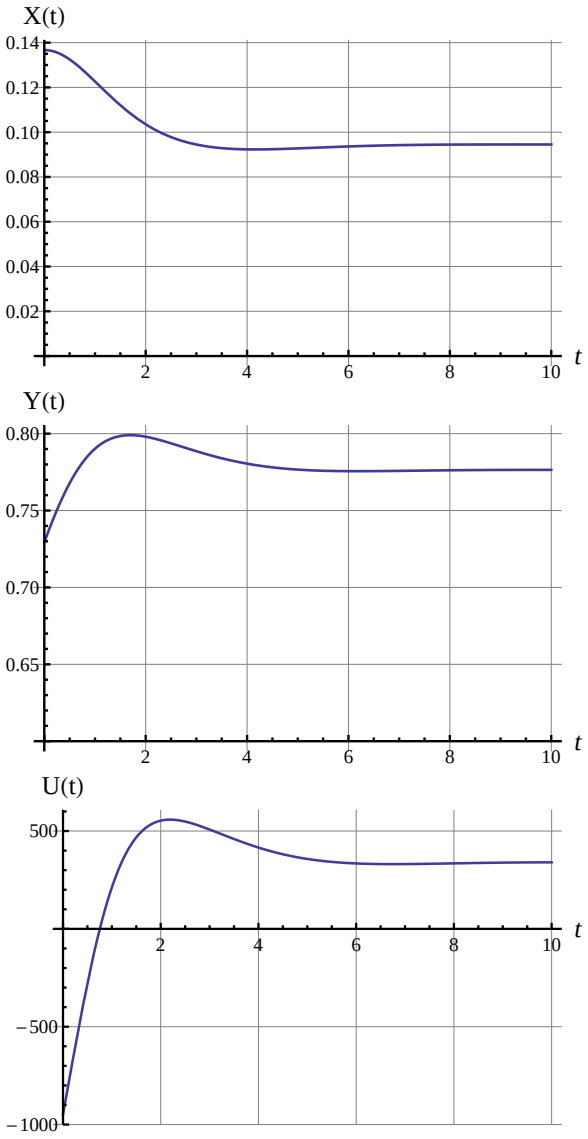

Figure 3.
Description. Hicks-ray Problem (23)-(27): (top) $\tilde{X}_{20}(t)$, (middle) $\tilde{Y}_{20}(t)$, (bottom) $\tilde{U}_{20}(t)$

This system is solved iteratively by the spectral tau method algorithm formed by (20)-(21)-(22). In order to start the iteration $\tilde{X}_0$, $\tilde{Y}_0$ and $\tilde{U}_0$ should be supplied by the user. Once these are provided $\tilde{X}_1$, $\tilde{Y}_1$ and $\tilde{U}_1$ are computed. Thus all the functions with subscript $k-1$ such as $\tilde{X}_{k-1}$, $\tilde{Y}_{k-1}$, $\tilde{U}_{k-1}$ are assumed to be known

We approximate $X_k(t)$, $Y_k(t)$ and U_k by three polynomials $\tilde{X}_k(t)$, $\tilde{Y}_k(t)$ and $\tilde{U}_k(t)$ respectively where

$$\tilde{X}_k(t) := \sum_{i=0}^{N} a_{ki}V_i(t), \qquad \tilde{Y}_k(t) := \sum_{i=0}^{N} b_{ki}V_i(t), \qquad \tilde{U}_k(t) := \sum_{i=0}^{N} u_{ki}V_i(t)$$

with $\{a_{ki},\ b_{ki},\ u_{ki},\ i = 0,1,\ldots,N\}$ being unknown coefficients.

Table 1. Coefficients of spectral tau approximate solution of problem (23)-(27).

Legendre	$\tilde{X}_{20}(x)$	$\tilde{Y}_{20}(x)$	$\tilde{U}_{20}$
P_0	0.099814135111705	0.779838514745205	319.95865312844278
P_1	−0.014447716903033	−0.004517449723467	115.95821518780970
P_2	0.017243639847273	−0.005516546127986	−296.39604842504507
P_3	−0.011124324495591	0.017784093547327	412.09559281837257
P_4	0.002005198828965	−0.019427432915259	−335.33644132195188
P_5	0.002938556082931	0.011518103104032	159.72323036315646
P_6	−0.003151934016184	−0.003500488896576	−28.99969101442239
P_7	0.001709145185311	−0.000334621114142	−23.63514313563081
P_8	−0.000542974832060	0.001150886895462	29.59528722423919
P_9	−0.000015936705712	−0.000840970019439	−19.69521820926024
P_{10}	0.000171488334340	0.000389567146567	8.77261343700043
P_{11}	−0.000145155600082	−0.000095199806246	−1.76810582580361
P_{12}	0.000075051535310	−0.000027982688861	−1.14605985892688
P_{13}	−0.000021916545586	0.000049087658321	1.56440363715893
P_{14}	−3.1233724658601E-6	−0.000032774535344	−1.03938505118715
P_{15}	9.0726214806500E-6	0.000013666636462	0.44790206015741
P_{16}	−6.8813183628923E-6	−2.4066743266386E-6	−0.08342508746112
P_{17}	3.2393901421164E-6	−1.6468007215665E-6	−0.06120983314422
P_{18}	−7.0768963477172E-7	1.9284582308677E-6	0.08092585154985
P_{19}	−5.7590158083240E-7	−1.2889474903878E-6	−0.05569526025167
P_{20}	4.9469939551189E-7	5.0912752638160E-7	0

Clearly the coefficients $\{a_{ki},\ b_{ki},,\ i = 0, 1, \ldots, N\}$ of $\tilde{X}_k(t)$ and $\tilde{Y}_k(t)$ will depend on $\{u_{ki}; i = 0, 1, \ldots, N\}$. To determine the values of these u_{ki}'s, we substitute $\tilde{X}_k(t)$, $\tilde{Y}_k(t)$ and $\tilde{U}_k(t)$ in the objective function (23), and integrate it exactly to end up with the problem of minimizing a multivariate function of the form:

$$\min \Psi_k(u_{k0}, u_{k1}, \ldots, u_{kN})$$

that can be achieved using the direct approach. That is we solve the algebraic linear system formed by the gradient,

$$\frac{\partial \Psi}{\partial u_{kj}} = 0, \quad j = 0, 1, 2, \ldots, N$$

to obtain the unknowns $\{u_{k0}, u_{k1}, \ldots, u_{kN}\}$, and then we test the Hessian to verify the optimality.

For each k-cycle, we construct $\{\tilde{X}_k, \tilde{Y}_k, \tilde{U}_k\}$ by the spectral Tau Algorithm, and repeat this process until a prescribed convergence tolerance ϵ is satisfied; that is until the iteration counter k reaches a certain k^* such that

$$\max\{\|\tilde{X}_{k^*} - \tilde{X}_{k^*-1}\|_\infty, \|\tilde{Y}_{k^*} - \tilde{Y}_{k^*-1}\|_\infty, \|\tilde{U}_{k^*} - \tilde{U}_{k^*-1}\|_\infty\} \leq \epsilon.$$

We consider then $\{\tilde{X}_{k^*}, \tilde{Y}_{k^*}, \tilde{U}_{k^*}\}$ as the Tau approximation for $\{X, Y, U\}$. It is worth noting that the convergence of $\{(\tilde{X}_k, \tilde{Y}_k, \tilde{U}_k),\ k \geq 1\}$ to the exact solution (X, Y, U) is guaranteed by Kantorovich Theorem that imposes conditions on the starting values $\{\tilde{X}_0, \tilde{Y}_0, \tilde{U}_0\}$ and on the entries of the matrix $\mathbf{A}$.

Figure 3 shows the profiles of the approximated states functions $X(t)$ and $Y(t)$ and the control $U(t)$ computed by the spectral Tau method with $N = 20$. These were obtained when $k^* = 18$ with tolerance $\epsilon = 10^{-12}$. This problem does not have an exact solution to compare but the same results can be obtained if this the problem is solved using the concept of Hamiltonian. Table lists the coefficients of $\tilde{X}_{20}$, $\tilde{Y}_{20}$ and $\tilde{U}_{20}$. The minimum value of the objective function = 2402.02746.

References

Canuto, C., Hussaini, M. Y., Quarteroni, A. & Zang, Th. A. (2006). *Spectral Methods: Fundamentals in Single Domains*, Springer Berlin.

Crisci, M.R., & Russo E. (1983). An extension of Ortiz' recursive formulation of the Tau Method to certain linear systems of ordinary differential equations. *Maths. Comput., 41*, 27-42.

El-Daou, M.K., & Al Hamad, K. M. (2012). Computation of the canonical polynomials and applications to some optimal control problems, *Num. Algo., 61*(4), 545-566. http://dx.doi.org/10.1007/s11075-012-9550-5

El-Daou, M.K., & Ortiz E. L. (1994). A recursive formulation of collocation in terms of canonical polynomials. *Computing, 52*, 177-202. http://dx.doi.org/10.1007/BF02238075.

El-Daou, M.K., & Ortiz E. L. (1998). A recursive formulation of Galerkins method in terms of the tau method: Bounded and Unbounded domains. *Comput. Math. Appl., 35*(12), 83-94.

El-Daou, M.K., Ortiz, E. L., & Samara, H. (1992). A unified approach to the tau methods and Chebyshev series expansions techniques. *Comput. Math. Appl., 25*, 73-82.

Flores Tlacuahuac, A., Terrazas Moreno, S., & Biegler, L. T. (2008). On Global Optimization of Highly Nonlinear Dynamic Systems. *Industrial and Engineering Chemistry Research, 47*, 2643-2655.

Freilich, J.H., & Ortiz, E.L. (1982). Numerical solution of systems of ordinary differential equations with the Tau Method: An error analysis. *Maths. Comput., 39*, 467-479.

Gottlieb, D., & Orszag, S. (1977). *Numerical Analysis of Spectral Methods: Theory and Applications*, SIAM, Pennsylvania.

Jaddu, H. & Majdalawi, A. (2014). Legendre Polynomials Iterative Technique for Solving a Class of Nonlinear Optimal Control Problems. *Int. Jr. of Cont. and Autom., 7*(3), 17-28. http://dx.doi.org/10.14257/ijca.2014.7.3.03

Lanczos, C. (1956). *Applied Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey.

Liu, K,M., & Ortiz, E.L. (1989). Numerical solution of ordinary and partial functional-differential eigenvalue problems with the Tau Method. *Computing, 41*, 205-217.

Liu, K.M., & Pan, C.K. (1999). The automatic solution to systems of ordinary differential equations by the tau method. *Comput. Math. Appls, 38*(910), 197-210. http://dx.doi.org/10.1016/S0898-1221(99)00275-8

Ortiz, E.L. (1969). The Tau Method. *SIAM J. Numer. Anal., 6*, 480-492. http://dx.doi.org/10.1137/0706044

Oritz, E.L., & Samara, H. (1981). An operational approach to the Tau method for the numerical solution of non-linear differential equations. *Computing, 27*(1), 15-25. http://dx.doi.org/10.1007/BF02243435

Oritz, E.L., & Samara, H. (1983). Numerical solution of differential eigenvalue problems with an operational approach to the Tau Method. *Computing, 31*, 95-103.

Oritz, E.L., & Samara, H. (1984). Numerical solution of partial differential equations with variable coefficients with an operational approach to the Tau Method *Computers Math. Applic., 10*(1), 5-13.

Ortiz, E.L., & Pham Ngoc Dinh, A. (1987). Linear Recursive Schemes Associated with Some Nonlinear Partial Differential Equations in One Dimension and the Tau Method. *SIAM Journal on Mathematical Analysis, 18*(2). http://dx.doi.org/10.1137/0518035