An International Journal
computers & mathematics
with applications

PERGAMON Computers and Mathematics with Applications 37 (1999) 41–48

# The Errors in Calculating the Pseudospectral Differentiation Matrices for Čebyšev-Gauss-Lobatto Points

R. BALTENSPERGER AND J.-P. BERRUT
Institut de Mathématiques, Université de Fribourg
CH-1700 Fribourg/Pérolles, Switzerland
⟨richard.baltensperger⟩⟨jean-paul.berrut⟩@unifr.ch

(Received and accepted May 1998)

**Abstract**—We discuss here the errors incurred using the standard formula for calculating the pseudospectral differentiation matrices for Čebyšev-Gauss-Lobatto points. We propose explanations for these errors and suggest more precise methods for calculating the derivatives and their matrices.


**Keywords**—Spectral methods, Differentiation matrix, Čebyšev points, Roundoff error, Barycentric formula.

## 1. INTRODUCTION

Recent years have seen widespread use of spectral and pseudospectral methods for the solution of partial differential equations [1–3]. The main reason is that, due to their infinite order, very good approximations of the solution are in general obtained with relatively few values of the solution when the latter is sufficiently differentiable.

In the matrix version of pseudospectral methods, one typically expands the approximate solution $u_n$ as an interpolating polynomial in its Lagrangian form

$$u_n(x) := \sum_{k=0}^{n} u_n(x_k)\, L_k(x), \tag{1}$$

so that the unknown coefficients are directly the values of the corresponding function at the interpolation points. The $L_k(x)$ in (1) are the Lagrange polynomials. The Čebyšev pseudospectral method on $[-1,1]$ with boundary values uses the Čebyšev-Gauss-Lobatto points

$$x_k := \cos\frac{k\pi}{n}, \qquad k = 0(1)n, \tag{2}$$

as interpolation points. For nonperiodic problems, it yields much better results than the Fourier pseudospectral method, at least as long as merely low-order derivatives of $u_n$ are involved. These derivatives can be estimated at the collocation points by differentiating (1) and evaluating the resulting expression. This yields

$$u_n^{(p)}(x_j) = \sum_{k=0}^{n} u_n(x_k)\, L_k^{(p)}(x_j), \qquad p = 1, 2, \ldots, \tag{3}$$

0898-1221/98/$ - see front matter  Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX
PII: S0898-1221(98)00240-5

or in matrix notation

$$\mathbf{u}^{(p)} = \mathbf{D}^{(p)}\mathbf{u}, \tag{4}$$

where

$$\mathbf{u} := \left[u_n(x_0), \dots, u_n(x_n)\right]^{\mathsf{T}}, \qquad \mathbf{u}^{(p)} := \left[u_n^{(p)}(x_0), \dots, u_n^{(p)}(x_n)\right]^{\mathsf{T}}, \tag{5}$$

and $\mathbf{D}^{(p)}$ is the $(n+1)\times(n+1)$ matrix whose entries are given by

$$D_{jk}^{(p)} := L_k^{(p)}(x_j), \qquad j,k = 0(1)n. \tag{6}$$

Such derivatives are calculated a great many times when solving time independent problems with iterative methods or time evolution problems with the method of lines. The differentiation matrices have been the subject of a significant number of articles, among them [4–7]. For example, calculating the differentiation matrices can be useful for studying the stability of pseudospectral methods applied to partial differential equations [6,8].

As noted in [9], errors much larger than machine precision arise when calculating the matrix $\mathbf{D}^{(p)}$ for Čebyšev points. For example, the errors in the first derivative computed as in (4) grow like $n^4$. We present here an explanation for these errors. For the Čebyšev-Gauss-Lobatto points (2), we then suggest ways of computing the differentiation matrices with smaller error growth.

## 2. COMPUTATION OF DERIVATIVES WITH THE MATRIX METHOD

For the Čebyšev collocation points (2), the first and second collocation derivative matrices $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$ can be computed analytically [1,2]. The entries of $\mathbf{D}^{(1)}$ are given by

$$D_{jk}^{(1)} = \begin{cases} \dfrac{2n^2+1}{6}, & \text{if } j=k=0, \\ -\dfrac{2n^2+1}{6}, & \text{if } j=k=n, \\ -\dfrac{x_j}{2(1-x_j^2)}, & \text{if } j=k\neq 0,n, \\ \dfrac{c_j}{c_k}\dfrac{1}{x_j-x_k}, & \text{if } j\neq k, \end{cases} \tag{7}$$

with $c_0 = 2$, $c_n = 2(-1)^n$, $c_k = (-1)^k$, $k = 1(1)n-1$. The second derivative matrix can be given explicitly also (see [10,11]), and here the relation $\mathbf{D}^{(p)} = (\mathbf{D}^{(1)})^p$ holds, which is not true for all collocation methods [11].

The matrix (7) is not skew symmetric, as opposed to the Fourier differentiation matrix. If the collocation derivative is computed by matrix-vector multiplication, then the total number of operations is $2n^2$. For small problems, matrix techniques are often faster than transform methods (which, for Čebyšev-Gauss-Lobatto points, can be applied by means of the FFT) and, unlike the latter, matrix multiplication is easily amenable to vectorization. Moreover, as noted by Fornberg [3, p. 8], "the pseudospectral method becomes particularly easy to apply to differential equations with variable coefficients and nonlinearity, since these give rise only to products of numbers (rather than to problems of determining the expansions coefficients for product of expansions)".

We now estimate the errors incurred when calculating the derivatives of the Čebyšev polynomial approximation by comparing numerically calculated derivatives with the exact derivatives of an example function. We measure the error in the numerical approximation $u_n$ with the maximum- or $L_\infty$-error

$$E_\infty := \max_{0\le k\le n} \left|u_n(x_k) - u(x_k)\right|,$$

here for the example function

$$u(x) := \sin(x), \qquad -1 \le x \le 1. \tag{8}$$

The justification for this (simple) choice is that, as noted in [9], the results are essentially independent of the complexity of the function (see Section 5 for other examples). And, since it is very simple, $u$ and its derivatives are approximated very accurately with small $n$ and the growth of the numerical error becomes visible. The calculation of the $E_\infty$-errors for the derivatives was performed at the same Čebyšev-Gauss-Lobatto points $x_k$. The computations were done on an AlphaServer 2100A 5/300.

Table 1. Errors in approximating the derivatives of $\sin(x)$ and maximum value of the sum of the elements of the rows using (7) (top) and (12) (bottom), both times with the relation $\mathbf{D}^{(2)} = (\mathbf{D}^{(1)})^2$.

| $n$ | 16 | 32 | 64 | 128 | 256 | 512 | 1024 |
|---|---|---|---|---|---|---|---|
| $E_\infty^{(1)}$ | $8.69 \cdot 10^{-14}$ | $4.93 \cdot 10^{-12}$ | $2.00 \cdot 10^{-11}$ | $5.40 \cdot 10^{-10}$ | $1.17 \cdot 10^{-8}$ | $1.50 \cdot 10^{-7}$ | $1.70 \cdot 10^{-6}$ |
| $\max_j \left( \left\| \sum_{k=0}^{n} D_{jk}^{(1)} \right\| \right)$ | $9.17 \cdot 10^{-14}$ | $5.85 \cdot 10^{-12}$ | $2.36 \cdot 10^{-11}$ | $6.42 \cdot 10^{-10}$ | $1.39 \cdot 10^{-8}$ | $1.78 \cdot 10^{-7}$ | $2.02 \cdot 10^{-6}$ |
| $E_\infty^{(2)}$ | $1.00 \cdot 10^{-11}$ | $2.10 \cdot 10^{-9}$ | $4.25 \cdot 10^{-8}$ | $3.96 \cdot 10^{-6}$ | $3.41 \cdot 10^{-4}$ | $1.50 \cdot 10^{-2}$ | $7.05 \cdot 10^{-1}$ |
| $\max_j \left( \left\| \sum_{k=0}^{n} D_{jk}^{(2)} \right\| \right)$ | $1.12 \cdot 10^{-11}$ | $2.49 \cdot 10^{-9}$ | $5.05 \cdot 10^{-8}$ | $4.71 \cdot 10^{-6}$ | $4.06 \cdot 10^{-4}$ | $1.78 \cdot 10^{-2}$ | $8.37 \cdot 10^{-1}$ |
| $E_\infty^{(1)}$ | $2.12 \cdot 10^{-13}$ | $4.13 \cdot 10^{-13}$ | $2.77 \cdot 10^{-12}$ | $3.44 \cdot 10^{-11}$ | $1.19 \cdot 10^{-9}$ | $1.43 \cdot 10^{-9}$ | $6.61 \cdot 10^{-8}$ |
| $\max_j \left( \left\| \sum_{k=0}^{n} D_{jk}^{(1)} \right\| \right)$ | $2.56 \cdot 10^{-13}$ | $5.68 \cdot 10^{-13}$ | $3.64 \cdot 10^{-12}$ | $4.00 \cdot 10^{-11}$ | $1.41 \cdot 10^{-9}$ | $1.70 \cdot 10^{-9}$ | $7.86 \cdot 10^{-8}$ |
| $E_\infty^{(2)}$ | $2.58 \cdot 10^{-11}$ | $2.91 \cdot 10^{-10}$ | $3.91 \cdot 10^{-9}$ | $2.33 \cdot 10^{-7}$ | $3.09 \cdot 10^{-5}$ | $2.43 \cdot 10^{-4}$ | $2.97 \cdot 10^{-2}$ |
| $\max_j \left( \left\| \sum_{k=0}^{n} D_{jk}^{(2)} \right\| \right)$ | $3.09 \cdot 10^{-11}$ | $3.49 \cdot 10^{-10}$ | $4.66 \cdot 10^{-9}$ | $2.76 \cdot 10^{-7}$ | $3.67 \cdot 10^{-5}$ | $2.89 \cdot 10^{-4}$ | $3.52 \cdot 10^{-2}$ |

The first and third row of Table 2 show the growth of the maximum error $E_\infty$ for both the first and the second derivatives as a function of the number $n+1$ of collocation points. Although not necessary, we chose $n = 2^\ell$ in order to allow the comparison with the transform method. We went up to $n = 1024$, about the largest size giving rise to systems of equations for the approximate differential equation that can be directly solved on workstations. As pointed out by Breuer and Everson [9], the $E_\infty$-error in the first derivative grows like $n^4$. The second derivative was computed using the relation $\mathbf{D}^{(2)} = \left(\mathbf{D}^{(1)}\right)^2$; its $E_\infty$-error is observed to increase as $n^6$.

## 3. SOURCES OF ERRORS

In [9], Breuer and Everson explain that the error spoiling the first derivative is due to roundoff, (on our computer, the machine precision $\nu$ is about $10^{-16}$) affecting the approximation of $x_1 = \cos 1/n$, so that, for large $n$,

$$\widehat{x}_1 = 1 - \frac{1}{2n^2} + \nu + \mathcal{O}\left(n^{-4}, \nu^2\right).$$

Using this expression, we may expand $\widehat{D}_{01}^{(1)}$ as a series in $\nu$,

$$\widehat{D}_{01}^{(1)} = \frac{-2}{x_0 - \widehat{x}_1} \approx -\frac{2}{1/2n^2 - \nu} = -4n^2 + \mathcal{O}\left(n^4 \nu\right), \tag{9}$$

which shows that the error in $D_{01}^{(1)}$ grows like $n^4\nu$, dominating any $n^2$-inner-product accumulation error in $\mathbf{D}^{(1)}\mathbf{u}$.

Another way of understanding the error is to look at a certain relation among the entries of the differentiation matrices. If the interpolated function $u$ takes the value one everywhere ($u(x) \equiv 1$), then the polynomial (1) interpolates $u$ exactly, so that $\sum_{k=0}^{n} L_k(x) \equiv 1$. Then we get for the derivatives of this sum, $\sum_{k=0}^{n} L_k^{(p)}(x) \equiv 0$ ($p = 1, 2, \dots$) and in particular $\sum_{k=0}^{n} L_k^{(p)}(x_j) = 0$ ($p = 1, 2, \dots$), so that

$$L_j^{(p)}(x_j) = - \sum_{k=0,\ k\neq j}^{n} L_k^{(p)}(x_j). \tag{10}$$

That is, every diagonal element of the differentiation matrix should equal the negative sum of all other elements on its row. This reflects the fact that all the derivatives of constant functions vanish.

On the other hand, we see in the formulas (7) that the elements $D_{00}^{(1)}$ and $D_{nn}^{(1)}$ are calculated precisely even when the other elements on their row are subject to large errors, so that the relations $D_{00}^{(1)} = -\sum_{k=1}^{n} D_{0k}^{(1)}$ and $D_{nn}^{(1)} = -\sum_{k=0}^{n-1} D_{nk}^{(1)}$ are less and less satisfied as $n$ increases. The error is $\mathcal{O}(n^4\nu)$ for $\mathbf{D}^{(1)}$ and $\mathcal{O}(n^6\nu)$ for $\mathbf{D}^{(2)}$. We can experience that the maximum error incurred in calculating the $p^{\text{th}}$ ($p = 1, 2$) derivative of $u$ grows at the same rate as the maximum value of the sum of the elements of the row (top half of Table 1).

If we replace $u$ by a modified example function that takes the value 0 at the extremities, say

$$u(x) := \sin\left(\frac{\pi}{2}(x+1)\right), \tag{11}$$

then we see that the results improve sharply (see Table 2). This demonstrates that the main cause of error is the very accurate calculation of the first and the last elements of the matrix $\mathbf{D}^{(p)}$, for which the error is only of order $\nu$. We can therefore hope for better results by modifying the way the diagonal elements of the differentiation matrix are calculated.

Table 2. $E_\infty^{(1)}$ and $E_\infty^{(2)}$-error for the function (11) using (7) and the relation $\mathbf{D}^{(2)} = (\mathbf{D}^{(1)})^2$.

| $n$ | 16 | 32 | 64 | 128 | 256 | 512 | 1024 |
|---|---|---|---|---|---|---|---|
| $E_\infty^{(1)}$ | $7.33 \cdot 10^{-15}$ | $2.75 \cdot 10^{-14}$ | $4.32 \cdot 10^{-13}$ | $1.28 \cdot 10^{-12}$ | $2.37 \cdot 10^{-12}$ | $1.24 \cdot 10^{-11}$ | $4.28 \cdot 10^{-11}$ |
| $E_\infty^{(2)}$ | $8.52 \cdot 10^{-13}$ | $1.19 \cdot 10^{-11}$ | $5.92 \cdot 10^{-10}$ | $5.00 \cdot 10^{-9}$ | $5.11 \cdot 10^{-8}$ | $5.29 \cdot 10^{-7}$ | $1.16 \cdot 10^{-5}$ |

## 4. ALLEVIATION OF THE ERRORS

We now present three different ways of alleviating the errors.

(i) As noted by Tang and Trummer [7], the evaluation of the elements of $\mathbf{D}^{(1)}$ by formula (7) is prone to cancellation. The first way of diminishing the errors is therefore to use trigonometric identities and replace (7) by

$$D_{jk}^{(1)} = \begin{cases} -\dfrac{x_j}{2\sin^2(j\pi/n)}, & \text{if } k = j \neq 0, n, \\ -\dfrac{c_j}{2c_k}\dfrac{(-1)^{j+k}}{\sin((j+k)\pi/2n)\sin((j-k)\pi/2n)}, & \text{if } j \neq k, \end{cases} \tag{12}$$

with $c_k$ as before (without modifying the formulas in (7) for $j = k = 0$ and $j = k = n$). We could also use trigonometric identities for the matrix $\mathbf{D}^{(2)}$ given explicitly (see [10,11]), but here we use the relation $\mathbf{D}^{(2)} = (\mathbf{D}^{(1)})^2$, as before.

One can see the improvement of the results in the bottom half of Table 1. The error is again close to that affecting the sum of the row. We gain one power of 10 for the approximation of the first derivative and two for the second.

We have also experimented with the algorithm proposed by Welfert [12], but this did not improve the results obtained with (12).

(ii) The second and best way we tried makes use of the barycentric representation of the polynomial of degree $\leq n$ interpolating between the Čebyšev-Gauss-Lobatto points [13,14]:

$$u_n(x) = \frac{\sum_{k=0}^{n}((-1)^k\delta_k)/(x-x_k)u_n(x_k)}{\sum_{k=0}^{n}((-1)^k\delta_k)/(x-x_k)}, \qquad \delta_k := \begin{cases} \frac{1}{2}, & k = 0 \text{ or } n, \\ 1, & \text{otherwise.} \end{cases} \tag{13}$$

In [15], Schneider and Werner present a formula for differentiating rational functions written in their barycentric form. They also suggest an algorithm for calculating all derivatives of such functions. Their algorithm for calculating the derivative of a function with values $\mathbf{f} := [f_0, \dots, f_n]^\top$ involves first-order divided differences of the interpolated values

$$\frac{f_k - f_j}{x_k - x_j}, \qquad k = 0(1)j-1,\ j+1(1)n.$$

Let us expand the calculated quotient of the term involving $x_0 - x_1$ into its Taylor series

$$\begin{aligned} \frac{f_0 - f_1}{x_0 - \widehat{x}_1} &= f_0' + \frac{f_0''}{2!}(x_0 - \widehat{x}_1) + \frac{f_0'''}{3!}(x_0 - \widehat{x}_1)^2 + \cdots \\ &\approx f_0' + \frac{f_0''}{2!}\left(\frac{1}{2n^2} - \nu\right) + \frac{f_0'''}{3!}\left(\frac{1}{2n^2} - \nu\right)^2 + \cdots \\ &= f_0' + \mathcal{O}\left(n^{-2}, \nu\right). \end{aligned}$$

It follows that, as soon as $n$ is sufficiently large, the error is of order $\nu$, the term involving $x_0 - x_1$ is less sensitive than (9) to rounding errors and we can therefore hope for better results, which is confirmed in the top half of Table 3. The error in the derivative of our test function $u(x) = \sin(x)$ does not grow as fast as before. The results are also better than those obtained with trigonometric identities, compare with the bottom half of Table 1. This way of calculating the differentiation matrix is even better than the transform method via the FFT, whose error are given in the bottom half of Table 3. Breuer and Everson have found that in the case of the transform method the error is proportional to $\mathcal{O}(n^2\nu)$ for the first derivative and $\mathcal{O}(n^4\nu)$ for the second. It should be noted, however, that neither the transform method nor the above algorithm produce the differentiation matrices.

Table 3. $E_\infty^{(1)}$ and $E_\infty^{(2)}$-error using the algorithm of Schneider and Werner (top) and the transform method (bottom).

| $n$ | 16 | 32 | 64 | 128 | 256 | 512 | 1024 |
|---|---|---|---|---|---|---|---|
| $E_\infty^{(1)}$ | $6.00 \cdot 10^{-15}$ | $2.18 \cdot 10^{-14}$ | $3.95 \cdot 10^{-14}$ | $5.71 \cdot 10^{-14}$ | $3.15 \cdot 10^{-13}$ | $2.58 \cdot 10^{-12}$ | $1.67 \cdot 10^{-11}$ |
| $E_\infty^{(2)}$ | $5.29 \cdot 10^{-13}$ | $7.60 \cdot 10^{-12}$ | $3.55 \cdot 10^{-11}$ | $3.55 \cdot 10^{-10}$ | $1.03 \cdot 10^{-8}$ | $1.82 \cdot 10^{-7}$ | $5.18 \cdot 10^{-6}$ |
| $E_\infty^{(1)}$ | $1.22 \cdot 10^{-14}$ | $9.08 \cdot 10^{-14}$ | $2.14 \cdot 10^{-13}$ | $1.18 \cdot 10^{-12}$ | $1.22 \cdot 10^{-11}$ | $6.13 \cdot 10^{-11}$ | $1.42 \cdot 10^{-9}$ |
| $E_\infty^{(2)}$ | $1.05 \cdot 10^{-12}$ | $2.54 \cdot 10^{-11}$ | $6.49 \cdot 10^{-11}$ | $6.28 \cdot 10^{-9}$ | $4.56 \cdot 10^{-8}$ | $1.05 \cdot 10^{-6}$ | $1.80 \cdot 10^{-4}$ |

(iii) Another way of shedding light on this issue is to calculate the differentiation matrices $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$ from (13) with the formula proposed by Schneider and Werner [15]. This yields the first derivative matrix as

$$D_{jk}^{(1)} = \begin{cases} \dfrac{\delta_k}{\delta_j}\dfrac{1}{x_j - x_k}, & \text{if } j \neq k, \\ -\displaystyle\sum_{i=0, i\neq j}^{n} \dfrac{\delta_i}{\delta_j}\dfrac{1}{x_j - x_i}, & \text{if } j = k, \end{cases} \tag{14}$$

and the second derivative matrix as

$$D_{jk}^{(2)} = \begin{cases} D_{jk}^{(1)}\left(D_{jj}^{(1)} - \dfrac{1}{x_j - x_k}\right), & \text{if } j \neq k, \\ 2\left(D_{jj}^{(1)}\right)^2 + 2 \displaystyle\sum_{i=0, i\neq j}^{n} D_{ji}^{(1)} \dfrac{1}{x_j - x_i}, & \text{if } j = k. \end{cases} \tag{15}$$

In Table 4, we display the errors in the first and second derivatives and we can see that they are again close to the errors incurred in summing the rows.

This method is again better than the transform method via the FFT (compare the top half of Table 4 with the bottom half of Table 3) .

The errors appearing in Table 4 could seem surprising, especially for the first derivative, where we explicitly have the relation (10). They are due to smearing [14]. If we rearrange the summation of the elements in every row from the smallest to the largest (in absolute value), the error diminishes, see the bottom half of Table 4.

We have also experimentally used the relation (10) and the proposed rearrangement for calculating the second differentiation matrix; we observe a larger improvement than for the first derivative, compare top and bottom half of Table 4 for the second derivative matrix. After rearrangement, the results are similar to those obtained using the algorithm of Schneider and Werner.

We have also experimented to combine the relation (12) with (14) and (15), but did not see any improvement.

Table 4. Errors in approximating derivatives of $\sin(x)$ and maximum value of the sum of the elements of the rows using (14) and (15), without (top) and with (bottom) rearrangement and the relation (10).

| $n$ | 16 | 32 | 64 | 128 | 256 | 512 | 1024 |
|---|---|---|---|---|---|---|---|
| $E_\infty^{(1)}$ | $1.29 \cdot 10^{-14}$ | $7.41 \cdot 10^{-14}$ | $2.92 \cdot 10^{-13}$ | $1.11 \cdot 10^{-12}$ | $1.67 \cdot 10^{-11}$ | $1.55 \cdot 10^{-11}$ | $4.99 \cdot 10^{-11}$ |
| $\max_j \left(\left\lvert \sum_{k=0}^{n} D_{jk}^{(1)} \right\rvert\right)$ | $7.77 \cdot 10^{-15}$ | $1.53 \cdot 10^{-14}$ | $2.70 \cdot 10^{-13}$ | $1.80 \cdot 10^{-12}$ | $1.72 \cdot 10^{-11}$ | $2.96 \cdot 10^{-11}$ | $2.57 \cdot 10^{-11}$ |
| $E_\infty^{(2)}$ | $2.18 \cdot 10^{-12}$ | $3.04 \cdot 10^{-11}$ | $8.93 \cdot 10^{-10}$ | $5.65 \cdot 10^{-9}$ | $7.51 \cdot 10^{-7}$ | $2.82 \cdot 10^{-6}$ | $5.93 \cdot 10^{-5}$ |
| $\max_j \left(\left\lvert \sum_{k=0}^{n} D_{jk}^{(2)} \right\rvert\right)$ | $1.82 \cdot 10^{-12}$ | $2.91 \cdot 10^{-11}$ | $9.31 \cdot 10^{-10}$ | $4.61 \cdot 10^{-9}$ | $9.24 \cdot 10^{-7}$ | $3.10 \cdot 10^{-6}$ | $7.92 \cdot 10^{-5}$ |
| $E_\infty^{(1)}$ | $1.59 \cdot 10^{-14}$ | $7.41 \cdot 10^{-14}$ | $1.86 \cdot 10^{-13}$ | $7.08 \cdot 10^{-13}$ | $3.82 \cdot 10^{-12}$ | $7.09 \cdot 10^{-12}$ | $3.66 \cdot 10^{-11}$ |
| $\max_j \left(\left\lvert \sum_{k=0}^{n} D_{jk}^{(1)} \right\rvert\right)$ | $6.44 \cdot 10^{-15}$ | $1.53 \cdot 10^{-14}$ | $4.24 \cdot 10^{-14}$ | $1.14 \cdot 10^{-13}$ | $9.71 \cdot 10^{-13}$ | $1.75 \cdot 10^{-12}$ | $3.25 \cdot 10^{-11}$ |
| $E_\infty^{(2)}$ | $8.79 \cdot 10^{-13}$ | $1.49 \cdot 10^{-11}$ | $5.06 \cdot 10^{-11}$ | $1.93 \cdot 10^{-9}$ | $5.78 \cdot 10^{-8}$ | $8.12 \cdot 10^{-7}$ | $5.46 \cdot 10^{-6}$ |
| $\max_j \left(\left\lvert \sum_{k=0}^{n} D_{jk}^{(2)} \right\rvert\right)$ | $5.40 \cdot 10^{-13}$ | $1.42 \cdot 10^{-12}$ | $5.82 \cdot 10^{-11}$ | $1.12 \cdot 10^{-9}$ | $1.97 \cdot 10^{-8}$ | $5.91 \cdot 10^{-7}$ | $2.04 \cdot 10^{-6}$ |

Our experiments demonstrate that maintaining the relation (10) is more important than calculating certain elements of the differentiation matrix precisely. It looks as if, as $n$ increases, preventing $D_{00}^{(p)}$ and $D_{nn}^{(p)}$ from following the continuous deterioration of the other elements results in an "unbalanced" differentiation operator.

## 5. FURTHER COMPUTATIONAL EXAMPLES

In order to demonstrate that the above results are independent of the function whose derivatives are approximated, we conclude with results for the derivatives of two other functions using the standard formula (7) (denoted (1) in Tables 5 and 6), the trigonometric identities (12) (denoted (2)), the transform method (denoted (3)), the algorithm of Schneider and Werner (denoted (4)) and finally the explicit formulas (14) and (15) with rearrangement (denoted (5)).

The first example we used,

$$u(x) := \frac{\sin(8x)}{(x+1.1)^{3/2}}$$

is that given in [9], slightly modified in order to avoid the value 0 at one of the extremities. The errors in approximating the first two derivatives are presented in Table 5.

Table 5. Errors in approximating the first (top) and second (bottom) derivatives of $\sin(8x)/(x+1.1)^{3/2}$ using five different techniques.

| $n$ | 16 | 32 | 64 | 128 | 256 | 512 | 1024 |
|---|---|---|---|---|---|---|---|
| (1) | $5.96 \cdot 10^{0}$ | $1.23 \cdot 10^{-2}$ | $2.29 \cdot 10^{-8}$ | $2.01 \cdot 10^{-8}$ | $4.34 \cdot 10^{-7}$ | $5.58 \cdot 10^{-6}$ | $6.32 \cdot 10^{-5}$ |
| (2) | $5.96 \cdot 10^{0}$ | $1.23 \cdot 10^{-2}$ | $2.21 \cdot 10^{-8}$ | $1.17 \cdot 10^{-9}$ | $4.43 \cdot 10^{-8}$ | $5.36 \cdot 10^{-8}$ | $2.45 \cdot 10^{-6}$ |
| (3) | $5.96 \cdot 10^{0}$ | $1.23 \cdot 10^{-2}$ | $2.22 \cdot 10^{-8}$ | $1.09 \cdot 10^{-10}$ | $3.81 \cdot 10^{-10}$ | $3.57 \cdot 10^{-9}$ | $4.01 \cdot 10^{-8}$ |
| (4) | $5.96 \cdot 10^{0}$ | $1.23 \cdot 10^{-2}$ | $2.22 \cdot 10^{-8}$ | $8.87 \cdot 10^{-12}$ | $9.09 \cdot 10^{-11}$ | $9.44 \cdot 10^{-11}$ | $2.75 \cdot 10^{-10}$ |
| (5) | $5.96 \cdot 10^{0}$ | $1.23 \cdot 10^{-2}$ | $2.22 \cdot 10^{-8}$ | $2.07 \cdot 10^{-11}$ | $4.98 \cdot 10^{-11}$ | $1.83 \cdot 10^{-10}$ | $1.83 \cdot 10^{-10}$ |
| (1) | $1.15 \cdot 10^{3}$ | $8.68 \cdot 10^{0}$ | $6.27 \cdot 10^{-5}$ | $1.48 \cdot 10^{-4}$ | $1.27 \cdot 10^{-2}$ | $5.58 \cdot 10^{-1}$ | $2.62 \cdot 10^{1}$ |
| (2) | $1.15 \cdot 10^{3}$ | $8.68 \cdot 10^{0}$ | $6.10 \cdot 10^{-5}$ | $8.39 \cdot 10^{-6}$ | $1.16 \cdot 10^{-3}$ | $8.94 \cdot 10^{-3}$ | $1.10 \cdot 10^{0}$ |
| (3) | $1.15 \cdot 10^{3}$ | $8.68 \cdot 10^{0}$ | $6.11 \cdot 10^{-5}$ | $2.29 \cdot 10^{-7}$ | $8.75 \cdot 10^{-6}$ | $1.01 \cdot 10^{-4}$ | $2.96 \cdot 10^{-3}$ |
| (4) | $1.15 \cdot 10^{3}$ | $8.68 \cdot 10^{0}$ | $6.11 \cdot 10^{-5}$ | $3.04 \cdot 10^{-8}$ | $1.80 \cdot 10^{-6}$ | $1.67 \cdot 10^{-6}$ | $7.07 \cdot 10^{-5}$ |
| (5) | $1.15 \cdot 10^{3}$ | $8.68 \cdot 10^{0}$ | $6.11 \cdot 10^{-5}$ | $5.01 \cdot 10^{-8}$ | $1.63 \cdot 10^{-6}$ | $2.89 \cdot 10^{-5}$ | $1.81 \cdot 10^{-4}$ |

For $n = 16$, 32, and 64, the results are almost the same with all techniques: the discretization error dominates the calculation error. For larger $n$, the formulas of Schneider and Werner give the best results for the approximations of the first two derivatives.

Our second example is

$$u(x) := \frac{1}{1+x^2}.$$

We have again calculated the approximation error for the first two derivatives. The results are presented in Table 6.

Table 6. Errors in approximating the first (top) and second (bottom) derivatives of $1/(1+x^2)$ using the five different techniques.

| $n$ | 16 | 32 | 64 | 128 | 256 | 512 | 1024 |
|---|---|---|---|---|---|---|---|
| (1) | $1.70 \cdot 10^{-5}$ | $2.85 \cdot 10^{-11}$ | $1.17 \cdot 10^{-11}$ | $3.21 \cdot 10^{-10}$ | $6.94 \cdot 10^{-9}$ | $8.92 \cdot 10^{-8}$ | $1.01 \cdot 10^{-6}$ |
| (2) | $1.70 \cdot 10^{-5}$ | $2.55 \cdot 10^{-11}$ | $1.93 \cdot 10^{-12}$ | $2.05 \cdot 10^{-11}$ | $7.06 \cdot 10^{-10}$ | $8.53 \cdot 10^{-10}$ | $3.93 \cdot 10^{-8}$ |
| (3) | $1.70 \cdot 10^{-5}$ | $2.55 \cdot 10^{-11}$ | $1.65 \cdot 10^{-13}$ | $1.21 \cdot 10^{-12}$ | $3.98 \cdot 10^{-12}$ | $1.19 \cdot 10^{-11}$ | $3.08 \cdot 10^{-10}$ |
| (4) | $1.70 \cdot 10^{-5}$ | $2.55 \cdot 10^{-11}$ | $1.24 \cdot 10^{-13}$ | $2.95 \cdot 10^{-13}$ | $5.96 \cdot 10^{-13}$ | $1.74 \cdot 10^{-12}$ | $1.55 \cdot 10^{-11}$ |
| (5) | $1.70 \cdot 10^{-5}$ | $2.55 \cdot 10^{-11}$ | $1.14 \cdot 10^{-13}$ | $2.12 \cdot 10^{-13}$ | $1.58 \cdot 10^{-12}$ | $7.28 \cdot 10^{-12}$ | $3.41 \cdot 10^{-11}$ |
| (1) | $2.91 \cdot 10^{-3}$ | $1.87 \cdot 10^{-8}$ | $2.51 \cdot 10^{-8}$ | $2.35 \cdot 10^{-6}$ | $2.03 \cdot 10^{-4}$ | $8.92 \cdot 10^{-3}$ | $4.19 \cdot 10^{-1}$ |
| (2) | $2.91 \cdot 10^{-3}$ | $1.75 \cdot 10^{-8}$ | $2.44 \cdot 10^{-9}$ | $1.38 \cdot 10^{-7}$ | $1.84 \cdot 10^{-5}$ | $1.45 \cdot 10^{-4}$ | $1.76 \cdot 10^{-2}$ |
| (3) | $2.91 \cdot 10^{-3}$ | $1.74 \cdot 10^{-8}$ | $8.68 \cdot 10^{-11}$ | $2.21 \cdot 10^{-9}$ | $1.18 \cdot 10^{-7}$ | $5.96 \cdot 10^{-7}$ | $3.31 \cdot 10^{-5}$ |
| (4) | $2.91 \cdot 10^{-3}$ | $1.74 \cdot 10^{-8}$ | $1.32 \cdot 10^{-10}$ | $4.77 \cdot 10^{-10}$ | $3.00 \cdot 10^{-9}$ | $1.34 \cdot 10^{-7}$ | $3.55 \cdot 10^{-6}$ |
| (5) | $2.91 \cdot 10^{-3}$ | $1.74 \cdot 10^{-8}$ | $1.16 \cdot 10^{-10}$ | $9.20 \cdot 10^{-10}$ | $1.06 \cdot 10^{-8}$ | $3.51 \cdot 10^{-7}$ | $7.63 \cdot 10^{-6}$ |

For the same reason as in Table 5, the first two columns are almost the same and, again, the formulas of Schneider and Werner yield the best results.

## 6. CONCLUSION

We have discussed some errors incurred when calculating the pseudospectral differentiation matrices for Čebyšev-Gauss-Lobatto points and we have suggested different methods for alleviating these errors. The results are better than those obtained through the transform method.

To compute the derivative of a function in cases where the transform method is not adequate, one should use method (ii). On the other hand, method (iii) is to be preferred for computing the differentiation matrices for solving time evolution problems with the method of lines or time independent problems with iterative methods.

We were not aware of the article [16] at the time we performed the above computations. Its authors use the trigonometric identities (12) in calculating the top half of the matrix $D^{(1)}$. Then they notice that the relation $D^{(1)}_{n-j,n-k} = -D^{(1)}_{jk}$ gives the bottom half of the matrix with smaller cancellation error than (7). The results for the first derivative matrix are almost the same as those displayed in the top half of Table 3. For the second derivative matrix (calculated as $\mathbf{D}^{(2)} = \left(\mathbf{D}^{(1)}\right)^2$), the improvement with respect to the bottom half of Table 1, where we also used trigonometric identities, is a mere power of 10.

## REFERENCES


1. J.P. Boyd, *Chebyshev and Fourier Spectral Methods*, Springer Verlag, Berlin, (1989).
2. P. Canuto, M.Y. Hussaini, A. Quarteroni and T. Zang, *Spectral Methods in Fluid Dynamics*, Springer, New York, (1988).
3. B. Fornberg, *A Practical Guide to Pseudospectral Methods*, Cambridge University Press, Cambridge, (1996).
4. D. Funaro, A preconditioning matrix for the Chebyshev differencing operator, *SIAM J. Numer. Anal.* **24**, 1024–1031 (1987).
5. D. Funaro, Computing the inverse of the Chebyshev collocation derivative, *SIAM J. Sci. Stat. Comput.* **9**, 1050–1057 (1988).
6. L.N. Trefethen and M.R. Trummer, An instability phenomenon in spectral methods, *SIAM J. Numer. Anal.* **24**, 1008–1023 (1987).
7. T. Tang and M.R. Trummer, Boundary layer resolving pseudospectral methods for singular perturbation problems, *SIAM J. Sci. Comput.* **17**, 430–438 (1996).
8. S.C. Reddy and L.N. Trefethen, Stability of the method of lines, *Numer. Math.* **62**, 235–267 (1992).
9. K.S. Breuer and R.M. Everson, On the errors incurred calculating derivatives using Chebyshev polynomials, *J. Comput. Phys.* **99**, 56–67 (1992).
10. R. Peyret, *Introduction to Spectral Methods*, von Karman Institute Lecture Series, Rhode-St-Genese, (1986).
11. B.D. Welfert, A remark on pseudospectral differentiation matrices, Department of Mathematics, Arizona State University, Tempe, AZ, (1992).
12. B.D. Welfert, Generation of pseudospectral differentiation matrices I, *SIAM J. Numer. Anal.* **34**, 1640–1657 (1997).