IJMMS 25:12 (2001) 801–810
PII. S0161171201003490
http://ijmms.hindawi.com

THE RESIDUAL AND THE ERROR FOR COLLOCATION METHODS

A. H. AHMED

(Received 25 May 1999)

ABSTRACT. We use earlier results on collocation matrices related to the solution and its derivatives in describing asymptotic relation between the error on the highest derivative and the residuals. That justifies the use of the form of the residual rather than its norm in measuring the error.

2000 Mathematics Subject Classification. Primary 65-XX.

1. Introduction. This paper extends the analysis of collocation methods for solution of ordinary differential equations using operator methods. It is based on the results described in Wright [11], Gerrard and Wright [7], Ahmed and Wright [3], and Ahmed [1], where references to related work (e.g., Karpilovskaja [9], Vainikko [10], Kantorovich and Akilov [8], and Anselone [5]) may be found. In the first three references it was shown that the norms of certain collocation matrices which relate to the solution and its derivatives and the norms of the corresponding inverse operators are asymptotically the same. These results are used here to study the behaviour of similar matrices related to the residual. Some useful results are stated and discussed.

Before investigating the results, we need to introduce the following assumptions and notations.

We consider an mth order differential equation of the form

$$x^m(t) + \sum_{j=0}^{n-1} p_j(t)x^j(t) = y(t), \tag{1.1}$$

with n associated homogeneous boundary conditions. This may be written in the operator form

$$(D^m - T)x = y, \tag{1.2}$$

where $(D^m x)(t) = (d^m x/dt^m)(t)$. In (1.2) we suppose x in X and y in Y where Y are suitable Banach spaces. The operator $D^m - T$ denoted by D^*, with the associated conditions is assumed to be invertible.

Let X_n and Y_n be subspaces of X and Y, respectively, and a projection $\varphi_n : Y \to Y_n$. An approximate solution $x_n \in X_n$ is found where possible by applying the projection φ_n to equation (1.2) with x_n substituted for x, that is,

$$\varphi_n(D^m x_n - Tx_n - y) = 0. \tag{1.3}$$

The subscript n will be related to the dimension of the subspaces X_n and Y_n which are assumed equal. It is also assumed that $\varphi_n D^m x_n = D^m x_n$, that is, D^m restricted

to X_n establishes a bijection between X_n and Y_n. It follows that x_n satisfies

$$(D^m - \varphi_n T)x_n = \varphi_n y. \tag{1.4}$$

Without loss of generality, we assume that the equation is satisfied in the range $[-1,1]$. The coefficients $p_j(t)$ in equation (1.1) are assumed to be continuous in the range. The space Y is taken as $R[-1,1]$, the space of Riemann integrable functions on $[-1,1]$, as approximate solutions with finite discontinuities in their mth derivatives are considered. X is taken as $D^{-1}Y$. The space Y_n will be taken to be polynomial or piecewise polynomial subspace defined below. The infinity norm is used throughout.

For global collocation Y_n is taken as the space of polynomial of degree $n-1$ on $[-1,1]$ and φ_n is the polynomial projection based on collocation points $\{\xi_j\}$, $j = 1,\dots,n$. For piecewise polynomial collocation the range is divided into n subintervals by the break points $-1 = t_0 < t_1 < \cdots < t_n = 1$. In each subinterval q points are used, chosen as

$$\xi_{jk} = \frac{((t_k - t_{k-1})\xi_j^* + (t_k + t_{k-1}))}{2}, \quad j = 1,\dots,q,\ k = 1,\dots,n, \tag{1.5}$$

where $\{\xi_j^*\}$, $j = 1,\dots,q$ are given reference points in $[-1,1]$. The subspace Y_n consists of functions which are polynomials of degree $q-1$ in each subinterval $J_i = [t_{i-1}+, t_{i-1}]$, $i = 1,\dots,n$. The projection ϕ_n is the corresponding interpolation projection based on these collocation points, which is equivalent to polynomial interpolation in each subinterval.

As in [1, 7] the global polynomial collocation points are assumed to be zeros of some polynomial chosen from a set of polynomials orthogonal with respect to a weight function $p(t)$ continuous in $(-1,1)$ and for which

$$\Omega^2 = \int_{-1}^{1} \rho(t)\,dt, \qquad \Omega^{*2} = \int_{-1}^{1} \frac{1}{\rho(t)}\,dt \tag{1.6}$$

are finite.

For piecewise polynomial collocation it is assumed that the points $\{\xi_j^*\}$ are chosen so that the corresponding interpolatory quadrature weights are positive as in [1, 3].

We introduce as in [9] Y_n^* to be the space generated by $\{D_{z_k}^*\}$ where $\{z_k\}$ is a basis of X_n, $k = 1,\dots,n$ for the global case and $k = 1,\dots,n_q$ for the piecewise case. Now if D^{*-1} is well defined, then D^* establishes a bijection between X_n and $Y_n^* = \varphi_n^* Y$. Thus $\varphi_n^* D^* x_n = D^* x_n$, for all x_n in X_n and φ_n^* defines a linear projection from Y to Y_n^*. It follows, then, that x_n satisfies

$$D^* x_n = \varphi_n^* y. \tag{1.7}$$

Some further definitions and notation are needed before stating the results precisely. First, the evaluation operator $\Phi_n : Y \to \mathbb{R}^n$ is needed to give a vector consisting of the values of a function at the collocation points. Second, an extension operator $\Psi_n : \mathbb{R}^n \to Y$ is needed to give a function whose values at the collocation points agree with the components of the vector. We define an additional evaluation operator $\Phi_p : Y \to \mathbb{R}^p$ relating to a set of evaluation points $\{t_i\}_{i=1,2,\dots,p}$ such that

$$\triangle_p = \max\{s_1 + 1, s_{i+1} - s_i, 1 - s_p\} \longrightarrow 0 \quad \text{as } p \longrightarrow \infty. \tag{1.8}$$

We note that the properties $\|\Phi_n\| = \|\Psi_n\| = 1$ and $\varphi_n \Psi \Phi = \varphi_n$ hold.

2. Definition of a collocation matrix relate to the residual. From (1.7), we get

$$x_n - D^{*-1}\varphi_n^* y, \tag{2.1}$$

and from (1.4), we get

$$x_n = (D^m - \varphi_n T)^{-1}\varphi_n y, \tag{2.2}$$

therefore

$$D^{*-1}\varphi_n^* y = (D^m - \varphi_n T)^{-1}\varphi_n y. \tag{2.3}$$

If we multiply from the left by $D^* = (D^m - T)$, we get

$$\varphi_n^* y = (D^m - T)(D^m - \varphi_n T)^{-1}\varphi_n y. \tag{2.4}$$

Applying Φ_p and using the result $\varphi_n \Psi_n \Phi_n = \varphi_n$, we get

$$\Phi_p \varphi_n^* y = \Phi_p (D^m - T)(D^m - \varphi_n T)^{-1}\varphi_n \Psi_n \Phi_n y. \tag{2.5}$$

Defining the vectors $\underline{y}_p^* = \Phi_p \varphi_n^* y$ and $\underline{y}_n = \Phi_n y$ gives

$$\underline{y}_p^* = W_p^* \underline{y}_n, \tag{2.6}$$

where W_n^* is the matrix defined by

$$W_p^* = \Phi_p (D^m - T)(D^m - \varphi_n T)^{-1}\varphi_n \Psi_n. \tag{2.7}$$

If we define the residual r_n by $(1 - \varphi_n^*)y$, then $\Phi_p r_n$ will be

$$\Phi_p r_n = \Phi_p y - \underline{y}_n^* = (\Phi_p \Psi_n - W_p^*)\underline{y}_n. \tag{2.8}$$

If we denote the matrix $(\Phi_p \Psi_n - W_p^*)$ by R_p then R_p relates the values of the right-hand side at the collocation points to the values of the residual at the evaluation points.

3. The residual and the error. If we define the vector $\underline{x}_n^{(j)}$ by $\Phi_p x_n^{(j)}$, $j = 0, 1, \ldots, n$ and apply the same arguments, applied to (2.5), to (2.2), we get

$$\underline{x}_n^{(j)} = \Phi_p D^{(j)} (D^m - \varphi_n T)^{-1}\varphi_n \Psi_n \underline{y} = Q_p^{(j)} \underline{y}, \quad j = 0, 1, \ldots, m. \tag{3.1}$$

Then it is shown in [7, 11] if the evaluation points are restricted to be at the collocation points that

$$\begin{aligned} \left\| Q_p^{(m)} \right\| &\longrightarrow \left\| D^{(m)} (D^{(m)} - T)^{-1} \right\| \quad \text{as } n \longrightarrow \infty, \\ \left\| Q_p^{(m)} - \Phi_p D^{(m)} (D^{(m)} - T)^{-1} \Psi_n \right\| &\longrightarrow 0 \quad \text{as } n \longrightarrow \infty \end{aligned} \tag{3.2}$$

for the global and piecewise method, respectively. It is also shown in [3], without restriction on evaluation points, that

$$\left\| Q_p^{(j)} \right\| \longrightarrow \left\| D^{(j)} (D^{(m)} - T)^{-1} \right\| \quad \text{as } n \longrightarrow \infty, \tag{3.3}$$

$$\left\| Q_p^{(j)} - \Phi_p D^{(j)} (D - T)^{-1} \Psi_n \right\| \longrightarrow 0 \quad \text{as } n \longrightarrow \infty, \text{ for } j = 0, 1, \ldots, m-1, \tag{3.4}$$

for both methods.

If we express the matrix W_p^* in terms of these Q matrices, we get from (2.7)

$$W_p^* = Q_p^{(m)} + \sum_{j=0}^{m-1} \Phi_p p_j Q_p^{(j)}. \tag{3.5}$$

Therefore

$$\begin{aligned} R_p &= W_p^* - \Phi_p (D^{(m)} - T)(D^m - T)^{-1}\Psi_n \\ &= Q_p^{(m)} - \Phi_p D^m (D^m - T)^{-1}\Psi_n + \sum_{j=0}^{m-1} \Phi_p p_j \left[Q_p^{(j)} - \Phi_p D^{(j)} (D^{(m)} - T)^{-1}\Psi_n \right]. \end{aligned} \tag{3.6}$$

Now if we denote the matrix $Q_p^{(m)} - \Phi_p D^{(m)} (D^{(m)} - T)^{-1}\Psi_n$ by $E_p^{(m)}$, then $E_p^{(m)}$ relates the values of the right-hand side at the collocation points to the values of the error on the highest derivative of the solution at the evaluation points and has the following relation with R_p.

By (3.6)

$$R_p - E_p^{(m)} = \sum_{j=0}^{m-1} \Phi_p p_j \left[Q_p^{(j)} - \Phi_p D^{(j)} (D^m - T)^{-1}\Psi_n \right], \tag{3.7}$$

and hence by (3.4)

$$\left\| R_p - E_p^{(m)} \right\| \le \sum_{j=0}^{m-1} \|\Phi_p\| |P_j| \left\| Q_p^{(j)} - \Phi_p^{(j)} D (D^m - T)\Psi_n \right\| \longrightarrow 0 \quad \text{as } n \longrightarrow \infty. \tag{3.8}$$

This result shows that the residual and the error on the highest derivative of the solution are asymptotically the same. It justifies the use of the form of the residual rather than its norm in estimating the error and which has significant advantages as stated in Ahmed and Wright [4]. It has also practical importance in mesh selection algorithms as described in Carey and Humphrey [6], Wright, Ahmed, and Seleman [12], and Ahmed [2].

4. Illustrative examples. For illustration only two problems are presented briefly here. More problems with detailed discussions are found on [4, 12]. These problems are:

Problem (1)

$$\begin{aligned} &x''(t) + 2\gamma t x'(t) + 2\gamma x(t) = 0, \\ &x(0) = 0, \qquad x(1) = \bar{e}^{\gamma}, \quad \gamma = 1. \end{aligned} \tag{4.1}$$

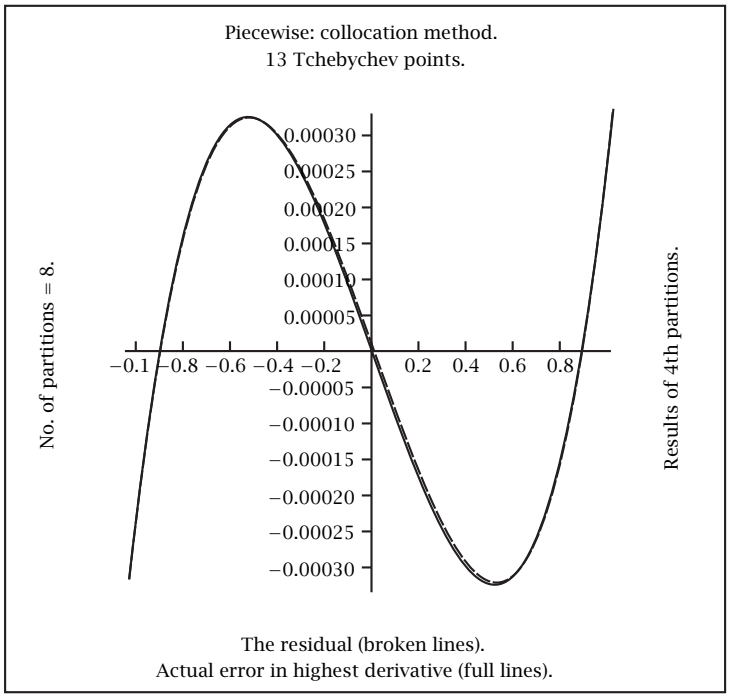
The solution of this problem is $\bar{e}^{\gamma t}$.

Problem (2)

$$\begin{aligned} &x''(t) + (3\cot(t) + 2\tan(t))x'(t) + 2x(t) = 0, \\ &x(30^\circ) = 4, \qquad x(60^\circ) = \frac{4}{3}. \end{aligned} \tag{4.2}$$

The solution of this problem is $1/\sin^2(t)$.

The two problems are solved using piecewise collocation method with 8 partitions and 3 collocation points. The collocation points are chosen as Tchebychev and Gauss points, respectively. In Figures 4.1, 4.2, 4.3, and 4.4, the residuals and the errors on the highest derivative of the solution are plotted on broken lines and full lines,


Piecewise: collocation method.
13 Tchebychev points.
No. of partitions = 8.
Results of 4th partitions.
0.00030
0.00025
0.00020
0.00015
0.00010
0.00005
−0.1 −0.8 −0.6 −0.4 −0.2
0.2 0.4 0.6 0.8
−0.00005
−0.00010
−0.00015
−0.00020
−0.00025
−0.00030
The residual (broken lines).
Actual error in highest derivative (full lines).

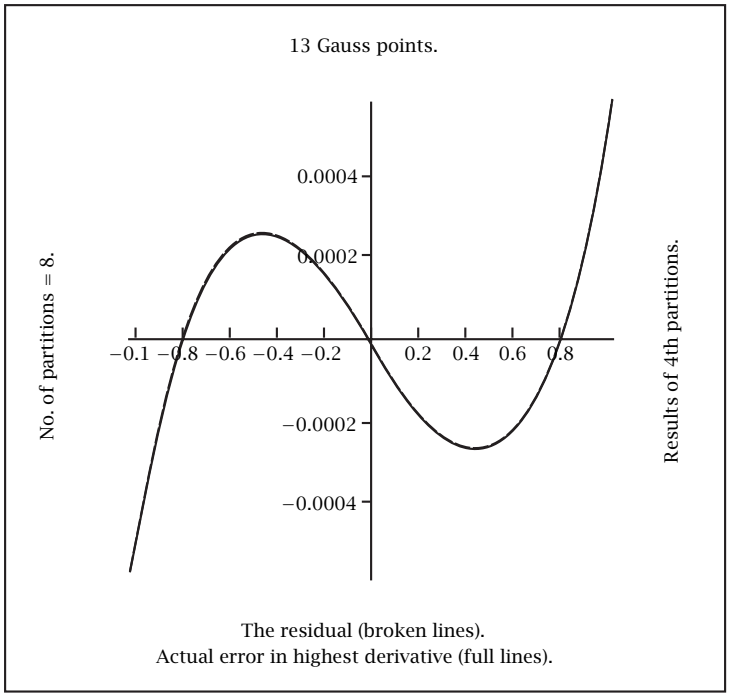

13 Gauss points.
No. of partitions = 8.
Results of 4th partitions.
0.0004
0.0002
−0.1 −0.8 −0.6 −0.4 −0.2
0.2 0.4 0.6 0.8
−0.0002
−0.0004
The residual (broken lines).
Actual error in highest derivative (full lines).

FIGURE 4.1.

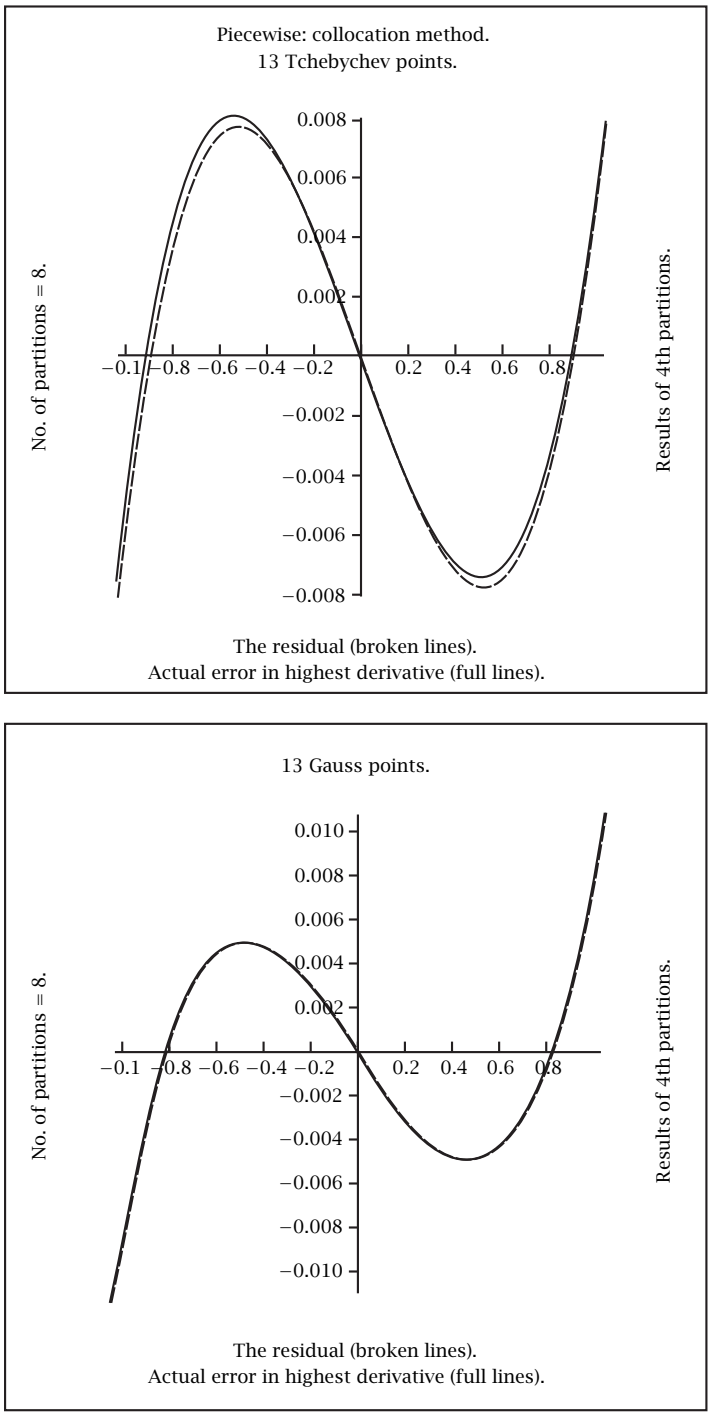
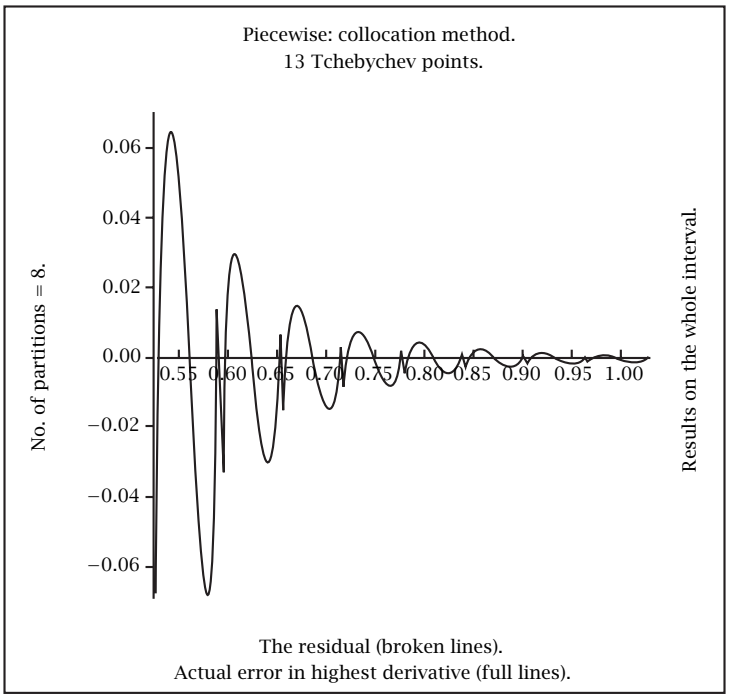

Piecewise: collocation method.
13 Tchebychev points.
No. of partitions = 8.
Results of 4th partitions.
0.008
0.006
0.004
0.002
−0.1 −0.8 −0.6 −0.4 −0.2
0.2 0.4 0.6 0.8
−0.002
−0.004
−0.006
−0.008
The residual (broken lines).
Actual error in highest derivative (full lines).
13 Gauss points.
No. of partitions = 8.
Results of 4th partitions.
0.010
0.008
0.006
0.004
0.002
−0.1 −0.8 −0.6 −0.4 −0.2
0.2 0.4 0.6 0.8
−0.002
−0.004
−0.006
−0.008
−0.010
The residual (broken lines).
Actual error in highest derivative (full lines).

FIGURE 4.2.


Piecewise: collocation method.
13 Tchebychev points.
No. of partitions = 8.
Results on the whole interval.
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00
The residual (broken lines).
Actual error in highest derivative (full lines).

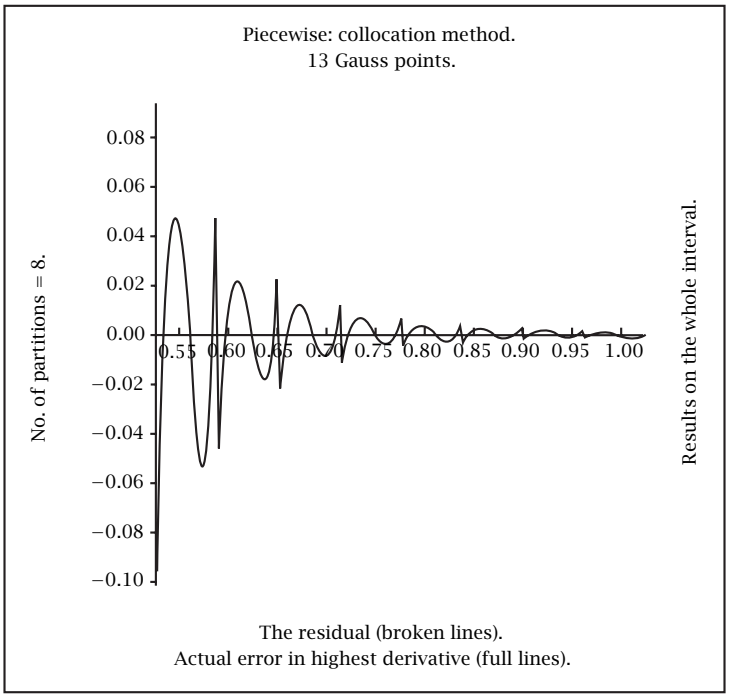
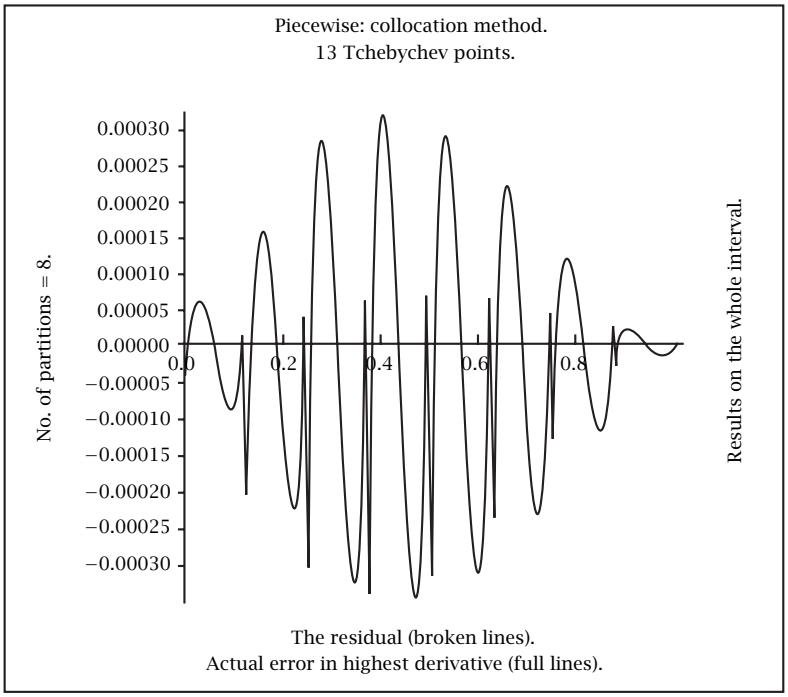

Piecewise: collocation method.
13 Gauss points.
No. of partitions = 8.
Results on the whole interval.
0.08
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.08
−0.10
0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00
The residual (broken lines).
Actual error in highest derivative (full lines).

FIGURE 4.3.


Piecewise: collocation method.
13 Tchebychev points.
0.00030
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
−0.00005
−0.00010
−0.00015
−0.00020
−0.00025
−0.00030
0.0
0.2
0.4
0.6
0.8
No. of partitions = 8.
Results on the whole interval.
The residual (broken lines).
Actual error in highest derivative (full lines).

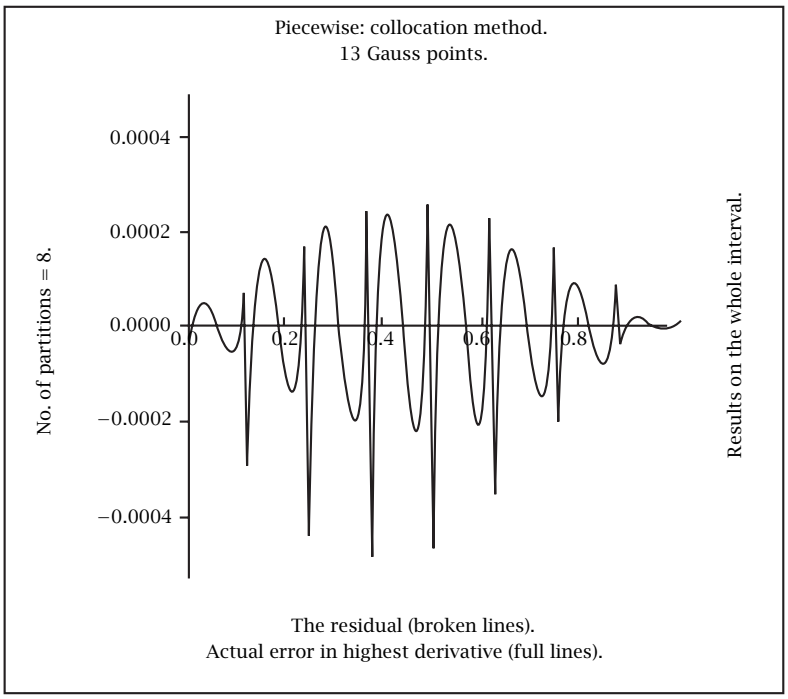

Piecewise: collocation method.
13 Gauss points.
0.0004
0.0002
0.0000
−0.0002
−0.0004
0.0
0.2
0.4
0.6
0.8
No. of partitions = 8.
Results on the whole interval.
The residual (broken lines).
Actual error in highest derivative (full lines).

FIGURE 4.4.

respectively. Figures 4.1 and 4.2 describe the comparison on one partition (the fourth partition) with Tchebychev and with Gauss points for problems (1) and (2), respectively. Figures 4.3 and 4.4 do the same comparison on the whole interval. We observe that in all figures the broken lines coincide with the full lines as is expected by the theory. However, with ill-conditioned problems as shown in [4, 12] large values of n may be needed before the residual gives close estimate of the error. It may be also useful to observe the minimax and the minimum over square properties with Tchebychev and Gauss points, respectively. This is also consistent with the theory and with similar results given in [1] between the interpolation operator and the collocation operator. One more point to observe is the discontinuity at the joint points. That is expected since continuity is not assumed on the highest derivative of the solution.

5. Conclusion. The result presented in this paper extends the asymptotic properties of the collocation matrices which relate to the solution and it is derivatives described in [3, 7, 11] to a matrix related to the residual. It gives stronger justification for the use of the residual in error estimates and mesh selection criteria described in [2, 4, 6, 12].

References

[1] A. H. Ahmed, *Asymptotic properties of collocation projection norms*, Comput. Math. Appl. **19** (1990), no. 4, 45–50. MR 90i:65141. Zbl 716.65067.

[2] ———, *Improved implementations of adaptive algorithms in collocation for boundary value problems*, Int. J. Comput. Math. **66** (1998), no. 3-4, 267–275. MR 98m:65203. Zbl 892.65052.

[3] A. H. Ahmed and K. Wright, *Further asymptotic properties of collocation matrix norms*, IMA J. Numer. Anal. **5** (1985), no. 2, 235–246. MR 86i:65042. Zbl 571.65077.

[4] ———, *Error estimation for collocation solution of linear ordinary differential equations*, Comput. Math. Appl. Part B **12** (1986), no. 5-6, 1053–1059. MR 88c:65062. Zbl 627.65092.

[5] P. M. Anselone, *Collectively Compact Operator Approximation Theory and Applications to Integral Equations*, Prentice-Hall, New Jersey, 1971. MR 56#1753. Zbl 228.47001.

[6] G. F. Carey and D. L. Humphrey, *Finite element mesh refinement algorithm using element residuals*, Codes for Boundary Value Problems in Ordinary Differential Equations (B. Childs, E. Denman, M. Scott, P. Nelson, and J.-W. Daniel, eds.), Lecture Notes Comput. Sci., vol. 76, Springer-Verlag, Berlin, New York, 1979, Proceedings of a Working Conference held at the University of Houston, Houston, Texas, May 14–17, 1978., pp. 243–249. MR 82b:65074. Zbl 434.65055.

[7] C. Gerrard and K. Wright, *Asymptotic properties of collocation matrix norms. II. Piecewise polynomial approximation*, IMA J. Numer. Anal. **4** (1984), no. 3, 363–373. MR 85i:65095. Zbl 542.65044.

[8] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces*, The Macmillan Co., New York, 1964. MR 35#4699. Zbl 127.06104.

[9] E. B. Karpilovskaja, *Convergence of the collocation method*, Soviet Math. Dokl. **4** (1963), 1070–1073, translated from Dokl. Akad. Nauk SSSR 151, 766–769 (1963). Zbl 192.21704.

[10] G. M. Vainikko, *On the stability and convergence of the collocation method*, Differential Equations **1** (1965), 186–194, translated from Differ. Uravn. 1, 244–254 (1965). Zbl 171.36503.

[11] K. Wright, *Asymptotic properties of collocation matrix norms. 1. Global polynomial approximation*, IMA J. Numer. Anal. **4** (1984), no. 2, 185–202. MR 85m:65072. Zbl 542.65043.

[12] K. Wright, A. H. Ahmed, and A. H. Seleman, *Mesh selection in collocation for boundary value problems*, IMA J. Numer. Anal. **11** (1991), no. 1, 7–20. MR 91m:65219. Zbl 716.65068.

A. H. Ahmed: Computer Centre, University of Khartoum, Sudan

Current address: Instituto de Matemática Pura e Aplicada, IMPA, Estrada Dona Castorina 110-Jardim Botânico, CEP 22460-320-Rio de Janeiro-RJ, Brazil