Journal of Engineering Mathematics, Vol. 9, No. 2, April 1975
Noordhoff International Publishing – Leyden
Printed in The Netherlands

Bounds for the solution of a second order differential equation with mixed boundary conditions

R. A. USMANI*

Department of Computer Science, University of Manitoba, Winnipeg, Canada

(Received August 20, 1974)

SUMMARY
A finite difference scheme is given for the numerical approximation of the real solution of the second order linear differential equation, lacking the first derivative, with mixed boundary conditions. The matrix associated with the resulting system of linear equations is tridiagonal and the overall discretization error is $O(h^4)$. The derived error bound is at most four times larger than the observed maximum error in absolute value for the numerical problem considered.

1. Introduction

It is well-known that the solution of the boundary value problem

$$y''(x) = f(x)\,y(x) + g(x)\,, \qquad a \leqq x \leqq b\,, \tag{1.1}$$

subject to the boundary conditions

$$\begin{aligned} y'(a) - c\,y(a) &= A\,, \\ y'(b) + d\,y(b) &= B\,, \end{aligned} \tag{1.2}$$

is unique provided

$$\begin{aligned} &\text{(i)}\ \ c \geqq 0\,, \quad d \geqq 0\,, \quad c+d > 0\,, \\ &\text{(ii)}\ f(x) \geqq 0 \ \text{ for } \ x \in [a, b]\,, \end{aligned} \tag{1.3}$$

see Henrici [4, p. 385].

The numerical approximation of the solution of the real linear boundary value problem (1.1)–(1.2) by finite difference methods has been considered by many authors. The reader is referred to Fox [2] for many elaborate finite difference schemes for obtaining an approximate solution of the preceding boundary value problem. A second order finite difference scheme for the numerical solution of a more general two point boundary value problem is developed and analysed by Aziz and Hubbard [1]. In defining any finite difference scheme for the numerical solution of Eqns. (1.1)–(1.2), we first introduce a finite set of grid points

$$x_m = a + (m-1)h\,, \qquad m = 1, 2, \ldots, N\,, \tag{1.4}$$

where $x_1 = a$, $x_N = b$, and $h = (b-a)/(N-1)$. We also require that the discretization error

$$e_i = y(x_i) - y_i \tag{1.5}$$

that is the difference between the exact solution $y(x_i)$ of the problem, Eqns. (1.1)–(1.2), at the grid point x_i and its approximation y_i obtained by solving the finite difference equations can be made arbitrarily small as the step size h tends to zero. We also need a bound on the discretization error e_i. Such error bounds are useful in proving convergence of the finite difference method, in the comparison of different numerical methods, for Richardson extrapolation, and in assisting the computer in the selection of h. However, this choice of the step-size h based on these error bounds may not be very realistic.

* This research was carried out during author's sabbatical leave at the Computer Centre Physics Department, The A.M.U. Aligarh, U. P., India.

The author [6] considered the numerical solution of Eqns. (1.1)–(1.2) satisfying the conditions (1.3) and proved that the resulting error is $O(h^4)$ based on the finite difference equations

$$\text{(i)}\quad -y(x_1)+y(x_2)=hy'(x_1)+\frac{h^2}{12}(5y''(x_1)+y''(x_2))+\frac{h^3}{12}y'''(x_1)-\tfrac{1}{180}h^5y^{(5)}(a_1),\quad x_1<a_1<x_2\ ;$$

$$\text{(ii)}\quad y(x_{m-1})-2y(x_m)+y(x_{m+1})=\frac{h^2}{12}(y''(x_{m-1})+10y''(x_m)+y''(x_{m+1}))-\frac{h^6}{240}y^{(6)}(a_m),\quad x_{m-1}<a_m<x_{m+1},\quad m=2,3,\ldots,N-1\ ;$$

and

$$\text{(iii)}\quad -y(x_{N-1})+y(x_N)=hy'(x_N)-\frac{h^2}{12}(y''(x_{N-1})+5y''(x_N))+\frac{h^3}{12}y'''(x_N)-\tfrac{1}{180}h^5y^{(5)}(a_N),\quad x_{N-1}<a_N<x_N\,. \tag{1.6}$$

The details of development of the formulas (1.5) are given in [6]. The bound on $\|e\|$ derived by Usmani is

$$\|e\|=\max_i|e_i|<\left(\frac{\overline{M}_5}{180}+\frac{(b-a)M_6}{480}\right)\left(\frac{2}{c}+b-a\right)h^4=E\,, \tag{1.7}$$

where $M_6=\max|y^{(6)}(x)|$ for $a\leqq x\leqq b$,

$$\overline{M}_5=\max.\left(\max_{a\leqq x\leqq x_1}|y^{(5)}(x)|\,,\quad \max_{x_{N-1}\leqq x\leqq b}|y^{(5)}(x)|\right),$$

and $e=(e_m)$ is the N-dimensional error vector. Note that for a vector $v=(v_i)$, $\|v\|=\max_i|v_i|$, and for a matrix $M=(m_{ij})$, $\|M\|=\max_i\Sigma_j|m_{ij}|$. The actual experiments show that the quantity E, in general, is 18 times larger than E_0, the observed maximum error in absolute value. We might remark that in deriving Eqn. (1.7) the author assumed that

$$c>0\,, \tag{1.8}$$

and d introduced in Eqns. (1.3) may be zero. In boundary value problems where $c>0$, $d>0$, the inequality (1.7) turns out to be too crude. In the sections that follow we will examine a technique to sharpen the inequality (1.7).

2. Purpose

Assume that the exact solution of the problem (1.1)–(1.2) is $y(x)\in C^6$. The replacement of the problem (1.1)–(1.2) by the difference equations (1.6) leads to the system of linear equations (2.1) in the unknowns y_m $(m=1,2,\ldots,N)$ where y_m is the numerical approximation to $y(x_m)$.

$$My=b\,. \tag{2.1}$$

Here $M=(m_{ij})$ is an $(N\times N)$ matrix, $y=(y_i)$. $b=(b_i)$ are N-dimensional column vectors, and

$$\begin{aligned}
m_{1,1}&=1+u\,, & m_{N,N}&=1+v\,,\\
m_{ij}&=2+\tfrac{10}{12}h^2f_i\,, & i&=j=2,3,\ldots,N-1\,,\\
&=-1+\tfrac{1}{12}h^2f_{i+1}\,, & j-i&=1\,,\\
&=-1+\tfrac{1}{12}h^2f_{i-1}\,, & i-j&=1\,,\\
&=0\,,\quad\text{otherwise}\ ;
\end{aligned} \tag{2.2}$$

$$\begin{aligned}
u&=hc+\tfrac{5}{12}h^2f_1+\tfrac{1}{12}h^3(cf_1+f_1')\,,\\
v&=hd+\tfrac{5}{12}h^2f_N+\tfrac{1}{12}h^3(df_N-f_N')\,,
\end{aligned} \tag{2.3}$$

and $f_i\equiv f(x_i)$, $f_i'\equiv f'(x_i)$ etc. Note that the quantities u and v are positive for sufficiently small values of $h>0$, provided we assume that

$$c>0\,, \quad d>0\,. \tag{2.4}$$

The purpose of this paper is to sharpen the inequality given by Eqn. (1.7) under the assumption (2.4), instead of Eqn. (1.8).

The error equation

$$Me = T \tag{2.5}$$

is obtained in the usual manner, as in [6]. The column vector $T=(t_i)$ is given below.

$$\begin{aligned} t_1 &= \tfrac{1}{180} h^5 y^{(5)}(a_1)\,, \quad t_N = \tfrac{1}{180} h^5 y^{(5)}(a_N)\,, \\ t_m &= \tfrac{1}{240} h^6 y^{(6)}(a_m)\,, \quad (m=2,3,\ldots,N-1)\,. \end{aligned} \tag{2.6}$$

The matrix $M>P$ provided $f(x)$ satisfies Eqns. (1.3), where $P=(\bar{P}_{ij})$ is an $N\times N$ matrix such that

$$\begin{aligned} \bar{P}_{1,1} &= 1+u\,, \quad \bar{P}_{N,N} = 1+v\,, \\ \bar{P}_{i,j} &= 2\,, \qquad i,j=2,3,\ldots,N-1\,, \\ &= -1\,, \quad |i-j|=1\,, \\ &= 0\,, \quad \text{otherwise}\,. \end{aligned} \tag{2.7}$$

The matrix M is irreducible provided

$$h<(12/f_M)^{\frac{1}{2}}\,, \quad f_M = \max_{a\leqq x\leqq b}.\, f(x)\,, \tag{2.8}$$

see [4, Corollary of Theorem 7.2]. It also follows from Theorem 7.4 [4] that M is a monotone matrix. Similarly P given by Eqns. (2.7) is also a monotone matrix, therefore it follows that

$$0 \leqq M^{-1} \leqq P^{-1}\,. \tag{2.9}$$

A further analysis now depends on the properties of P^{-1}. We now attempt to determine $P^{-1}=(P_{ij})$ explicitly. The author relies entirely on the theory of linear difference equations for generating the elements of the matrix P^{-1}, see [4, 5].

3. Inversion of the matrix P, and error bounds

Concerning $P^{-1}=(P_{ij})$, we will prove the following theorem.

Theorem 3.1: *The matrix $P^{-1}=(P_{ij})$ is symmetric and*

$$\begin{aligned} P_{ij} &= (1+(i-1)u)\cdot(1+(N-j)v)/D>0\,, \qquad i\leqq j\,, \\ &= (1+(j-1)u)\cdot(1+(N-i)v)/D>0\,, \qquad i\geqq j\,, \end{aligned}$$

where $D=u+v+(N-1)uv$, and u, v are given by Eqns. (2.3).

Remark. It can be easily seen that D is the determinant of the matrix P.

Proof: On multiplying the rows of P by the jth column of P^{-1}, we obtain the following difference equations.

$$\begin{aligned} &\text{(i)} \quad (1+u)P_{1,j}-P_{2,j}=0\,, \\ &\text{(ii)} \quad -P_{i-1,j}+2P_{i,j}-P_{i+1,j}=0\,, \qquad i=2,3,\ldots,j-1\,, \\ &\text{(iii)} \quad -P_{j-1,j}+2P_{j,j}-P_{j+1,j}=1\,, \\ &\text{(iv)} \quad -P_{i-1,j}+2P_{i,j}-P_{i+1,j}=0\,, \qquad i=j+1,j+2,\ldots,N-1\,, \\ &\text{(v)} \quad -P_{N-1,j}+(1+v)P_{N,j}=0\,, \end{aligned} \tag{3.1}$$

The solution of Eqn. (3.1) (ii) with initial condition Eqn. (3.1) (i) is easily seen to be

$$P_{ij} = C_1(1+iu/(1-u)), \qquad i \leqq j, \tag{3.2}$$

where C_1 is independent of i, but may depend on j. Similarly the solution of difference Eqn. (3.1) (iv) with associated condition 3.1 (v) is

$$P_{ij} = C_2(1-iv/(1+Nv)), \qquad i \geqq j, \tag{3.3}$$

and the arbitrary constant C_2 depends only on j. The element P_{jj} can be obtained either from Eqns. (3.2) or (3.3). On equating the expressions for P_{jj} obtained from Eqns. (3.2) and (3.3) respectively, we obtain an equation in the unknowns C_1 and C_2 in the form

$$C_1(1+ju/(1-u)) = C_2(1-jv/(1+Nv)). \tag{3.4}$$

Also, on substituting in Eqn. (3.1) (iii), the values of P_{ij} $(i=j-1, j, j+1)$ derived from Eqns. (3.2) and (3.3) respectively we obtain

$$C_1(1+(j+1)u/(1-u)) - C_2(1-(j+1)v/(1+Nv)) = 1. \tag{3.5}$$

On solving Eqns. (3.4) and (3.5) for C_1 and C_2, we obtain

$$C_1 = (1-u)(1+(N-j)v)/D, \tag{3.6}$$

$$C_2 = (1+Nv)(1+(j-1)u)/D. \tag{3.7}$$

The theorem follows on substituting the values of C_1 and C_2 in Eqns. (3.2) and (3.3) respectively.

Now define

$$R_i = (P_{i,1}+P_{i,N}), \quad S_i = \sum_{j=2}^{N-1} P_{ij}. \tag{3.8}$$

Concerning R_i and S_i we prove the following lemmas.

Lemma 3.2:

$$R_i = (2+(i-1)u+(N-i)v)/D,$$

and if $w = \max(u, v)$, *then*

$$R_i \leqq (2+(N-1)w)/D \textit{ for all } i.$$

Lemma 3.3:

$$S_i = (H+iG-i^2D)/2D,$$

and hence

$$S_i \leqq (G^2+4HD)/8D^2 \textit{ for all } i,$$

where

$$G = (2N-1)u+3v+(N^2-1)uv,$$

$$H = (2N-4)-2(N-1)u+N(N-3)v-N(N-1)uv.$$

Proof of Lemma 3.2

Using Theorem 3.1, we get

$$R_i = P_{i,1}+P_{i,N} = \frac{(1+(N-i)v)+(1+(i-1)u)}{D}$$

$$= (2+(i-1)u+(N-i)v)/D,$$

and clearly R_i is independent of i if $u=v$ and equals $(2+(N-1)u)/D$. Further, if $u \neq v$, and $w = \max.(u, v)$, then

$$R_i \leqq (2+(N-1)w)/D \text{ for all } i.$$

Proof of Lemma 3.3

Consider

$$S_i = \sum_{j=2}^{i} p_{ij} + \sum_{j=i+1}^{N-1} p_{ij}$$
$$= \frac{1}{D}\left[(1+(N-i)v)\sum_{j=2}^{i}(1+(j-1)u)+(1+(i-1)u)\sum_{j=i+1}^{N-1}(1+(N-j)v)\right],$$

using Theorem 3.1. On simplifying the expression on the right of the equality sign, we get the desired expression for S_i. Treat S_i, as given by Lemma 3.3, as a function of the real variable i. Then it is easily verified that S_i attains its maximum for $i=G/2D$. Furthermore, it is proved that

$$\max_i \; S_i = (G^2+4HD)/8D^2 .$$

We now turn back to the error equation (2.5) and write it in the form

$$|e| = M^{-1}|T| \leqq P^{-1}|T| ,$$

using Eqn. (2.9), and thus

$$|e_i| \leqq \sum_{j=1}^{N} P_{ij}|t_j| = (P_{i1}|t_1|+P_{iN}|t_N|) + \sum_{j=2}^{N-1} P_{ij}|t_j| \leqq \frac{h^5}{180}\overline{M}_5 R_i + \frac{h^6}{240} M_6 S_i , \tag{3.9}$$

using Eqns. (2.6), (3.8), and definitions of $\overline{M}_5$, M_6 as in Eqn. (1.7).

We finally obtain

$$\begin{aligned} \|e\| &= \max_i |e_i| \\ &\leq \frac{h^5}{180}\overline{M}_5 \cdot \frac{2+(N-1)w}{D} + \frac{h^6}{1920} M_6 \cdot \frac{G^2+4HD}{D^2} \\ &= O(h^4), \quad \text{(see Lemmas 3.2 and 3.3)}, \end{aligned} \tag{3.10}$$

on substituting the values of D, G, and H etc.

4. A numerical example

In this section we will compare the observed maximum error in absolute value (henceforth to be designated by E_0) with its bound $\|e\|$ based on Eqn. (3.10) for the following boundary value problem

$$\begin{aligned} &y''(x) = y(x) - 4x\,e^x , \qquad 0 \leqq x \leqq 1 , \\ &y'(0) - y(0) = 1 , \; y'(1) + y(1) = -e . \end{aligned} \tag{4.1}$$

We can easily verify that the analytical solution of Eqns. (4.1) is

$$y(x) = x(1-x)e^x ,$$

and $\overline{M}_5 = 25e$, $M_6 = 36e$.

In order to reduce the rounding errors to a minimum, the bounds $\|e\|$ based on Eqn. (3.10) were calculated using "extended precision arithmetic", on an IBM 1130 Computer at the Aligarh Muslim University, Aligarh, India. The problem under consideration is solved with $h=2^{-m}$, $m=1, 2, \ldots, 8$. The numerical results are summarized in the accompanying table. Clearly the error bound $\|e\|$ based on Eqn. (3.10) is at most 4 times larger than E_0.

TABLE 1

No. of unknowns y_i	h	$\|e\|$ *based on Eqn.* (3.10)	E_0	$\|e\|/E_0$
3	$\frac{1}{2}$	0.140×10^{-1}	0.540×10^{-2}	2.6
5	$\frac{1}{4}$	0.115×10^{-2}	0.364×10^{-3}	3.1
9	$\frac{1}{8}$	0.816×10^{-4}	0.232×10^{-4}	3.5
17	$\frac{1}{16}$	0.542×10^{-5}	0.146×10^{-5}	3.7
33	$\frac{1}{32}$	0.349×10^{-6}	0.913×10^{-7}	3.8
65	$\frac{1}{64}$	0.222×10^{-7}	0.571×10^{-8}	3.9
129	$\frac{1}{128}$	0.140×10^{-8}	0.357×10^{-9}	3.9
257	$\frac{1}{256}$	0.876×10^{-10}	0.212×10^{-10}	4.0

REFERENCES

[1] A. K. Aziz and B. E. Hubbard, Bounds for the solution of the Sturm–Liouville problem with application to finite difference methods, *J. Soc. Indust. Math.*, 12, 1, March (1964) 163–178.
[2] L. Fox, *The numerical solution of two-point boundary value problems in ordinary differential equations*, Oxford University Press (1957).
[3] C. F. Fischer and R. A. Usmani, Properties of some tridiagonal matrices and their application to boundary value problems, *SIAM J. Numer. Anal.*, 6 (1969) 127–142.
[4] P. Henrici, *Discrete variable methods in ordinary differential equations*, John Wiley, New York (1962).
[5] D. E. Rutherford, Some continuant determinants arising in physics and chemistry II, *Proc. Roy. Soc. Edinburgh* A63 (1952) 232–241.
[6] R. A. Usmani, Integration of second order linear differential equation with mixed boundary conditions, *Intern. J. Computer Math.*, (1973) 389–397.