TR/19/85 August 1985

Multiderivative methods for nonlinear second order boundary value problems.

E.H. Twizell

and

S.I.A . Tirmizi

Abstract

Second, fourth and sixth order methods are developed and analysed for the numerical solution of non-linear second order boundary value problems.

The methods arise from a two-step recurrence relation involving exponential terms, these being replaced by Padé approximants .
The methods are tested on two problems from the literature.

w925929x

1. INTRODUCTION AND DEVELOPMENT

Consider the general second order boundary value problem given by

$$y''(x) = f(x,y) \quad , \quad a_D < x < a_1 \tag{1}$$

$$y(a_0) = A_0 \quad , \quad y(a_1) = A_1 \tag{2}$$

where a_0, a_1, A_0, A_1 are finite constants. It will be assumed that a unique solution, $y(x)$, to (1) and (2) exists for $a_0 \leq x \leq a_1$; for a general discussion of existence and uniqueness to (1) and (2) the reader is referred to the text by Henrici [1],for instance. It will further be assumed that $y(x)$ and $f(x,y)$ are sufficiently often differentiable with respect to x in the interval $a_0 \leq x \leq a_1$.

Numerous numerical methods for solving (1) with (2) have appeared in the literature. Commonly used finite difference methods are discussed by the authors of many texts (see, for instance [1,2,3]); the problem is treated using variational techniques in [3]; and shooting methods have received coverage in [4], for example. The possibility of using spline functions to obtain a smooth solution of (1) with (2) was first discussed in [5], and in [6] cubic, quartic, quintic and sextic splines are used to solve the boundary value problem. Attention was drawn in [7] to the connection between a cubic spline solution and the solution obtained using the well known Numerov method. The replacement of (2) by mixed boundary conditions was considered in [8] where the classical second order finite difference method was used to compute the solution.

Suppose the independent variable x is incremented using a constant step size $h = (a_1 - a_0) / (N+1)$ where N is a positive integer. The solution will be computed at the points $x_m = a_0 + mh$ $(m = 1,2,...,N)$ and the notation y_m will be used to denote the solution of a numerical method at x_m; clearly $y_0 = A_0$ and $y_{N+1} = A_1$.

The multiderivative methods to be discussed in the paper are based on the recurrence relation

$$- y(x-h) + \{\exp(hD) + \exp(-hD)\}\, y(x) - y(x+h) = 0 \quad , \tag{3}$$

where $D = d/dx$. Using this relation, each numerical method will determine the solution vector $\underline{Y} = (y_1, y_2, ..., y_N)^T$, T denoting transpose, implicitly.

The methods are obtained by approximating the exponential terms in (3) by (M,K) Padé replacements of the form

$$\exp(hD) = [Q_M(hD)]^{-1} P_K(hD) + 0(h^{M+K+1}) \tag{4}$$

where $P_K(hD)$ and $Q_M(hD)$ are polynomials in hD of degrees K and M, respectively. Using such an approximation in (3), and clearing all inverses, leaves only even powers of D. It then follows that f(x,y) and its second, fourth, etc., derivatives with respect to x may occur in the resulting multiderivative method. Clearly, the need to use such derivatives of f implies that $y'(x)$ becomes involved. This derivative is estimated to the required accuracy by using enough terms of the forward, central or backward differentiation formulas, as appropriate, given, for example, in the text by Gerald and Wheatley [9].

The local truncation error associated with the numerical method based on the (M,K) Pade approximant, at the point $x=x_m$, takes the form

$$t_m = t(x_m) = c_{p+2}\, h^{p+2}\, y^{(p+2)}(x_m) + C_{p+4}\, h^{p+4}\, y^{(p+4)}(x_m) \ldots\ldots \tag{5}$$

where $p = 2[\frac{1}{2}(M+K)]$. Here p is the order of the method and the C_{p+q} (q-2,4,...) are constants; C_{p+2} is the error constant of the method. For consistency, $p \geq 1$ and so the methods based on the use of the (0,1) or (1,0) Padé approximants in (3) are inconsistent. The error constants

for eleven of the multiderivative methods arising from a sample of values of M and K are given in Table 1.

Table 1 here

2. SECOND ORDER METHODS

All numerical methods to be discussed in the paper have the form

$$-\delta^2 y_m + h^2 \sum_{r=-1}^{1} a_r \tilde{f}_{m+r} + h^4 \sum_{r=-1}^{1} a_r^* \tilde{f}''_{m+r} + h^6 \sum_{r=-1}^{1} a_r^{**} \tilde{f}^{(iv)}_{m+r} = 0 \ , \quad m = 1,2,....,N \ , \tag{6}$$

where $\tilde{f}_m = f(x_m, y_m)$, $\tilde{f}''_m = d^2 f(x_m, y_m)/dx^2$, $\tilde{f}^{(iv)}_m = d^4 f(x_m, y_m)/dx^4$, δ^2

is the central difference operator defined by

$$\delta^2 y_m = y_{m-1} - 2y_m + y_{m+1} \ , \tag{7}$$

and the a_r, a_r^*, a_r^{**} (r = -1,0,1) are parameters which depend on the chosen Padé approximant.

Visual examination of the error constants of each of the five second order methods of the family (see Table 1) shows that the numerical methods based on the (1,2) and (2,1) Pade approximants are the most accurate. The parameters a_r, a_r^*, a_r^{**} for the method based on the (1,2) Padé. approximant are given by

$$a_{-1} = a_1 = \frac{1}{9} \ , \ a_0 = \frac{7}{9} \ , \quad a_r^* = a_r^{**} = 0 \qquad (r = -1,0,1) \tag{8}$$

and those for the method based on the (2,1) Pade approximant are given in the Appendix. It is obvious from (8) that no derivatives of f(x,y) are involved in the method based on the (1,2) Pade approximant . It is equally obvious from the Appendix that d^2f/dx^2 is required for the (2,1) method and it may be concluded, therefore, that the (1,2) method is, overall, the best second order method of the family arising from (3).

The method based on the (0,2) Pade approximant is, of course, the well known second order linear multistep method; its parameters, and those of the other second order methods of the family, are given in the Appendix.

The solution vector $\underline{Y}$ for each method is obtained by solving a non-linear algebraic system of order N of the form

$$\underline{F}(\underline{Y}) = \underline{0} \tag{9}$$

using the Newton-Raphson method. In the case of the method {(6) ,(8)} based on the (1,2) Padé approximant, the Jacobian of $\underline{F}(\underline{Y})$ in (9) is tridiagonal.

3. FOURTH ORDER METHODS

Nine entries of the Padé Table lead to fourth order multiderivative methods when used in (3). The error constants of three of these methods are given in Table 1, the error constants of the other six methods being no smaller in modulus than these three. From Table 1, it is evident that the use of the (2,3) Padé approximant in (3) leads to the fourth order multiderivative method with smallest modulus error constant.

The parameters a_r, a_r^*, a_r^{**} for this method are given by

$$a_{-1} = a_1 = \frac{3}{50}\ ,\ a_0 = \frac{22}{25}\ ,\ a_{-1}^* = a_1^* = -\frac{1}{400}\ ,\ a_0^* = \frac{17}{600}\ ,\ a_r^{**} = 0 \qquad (r = -1, 0, 1) \tag{10}$$

while parameters for the other fourth order methods appearing in Table 1 are given in the Appendix. After making approximations for y', as noted in § 1,the solution vector $\underline{Y}$ is obtained by solving a non-linear system of the form (9).

4. SIXTH ORDER METHODS

It was observed in §§2,3 that the (1,2) and (.2,3) Pade approximants give the smallest moduli error constants for the second and fourth order numerical methods arising from (3). A check of the local truncation errors of the sixth order multiderivative methods of the family reveals that the (3,4) and (4,3) padé approximants give the smallest moduli error constants. The (4,3) method, however, requires the sixth order derivative of f(x,y) with respect to x while the (3,4) method requires only the sixth derivative of y(x) and is therefore more economical to implement. It is reasonable to assume that, for any (even) order p, the method based on the $(\frac{1}{2}p, \frac{1}{2}p+1)$ Padé approximant yields the multiderivative method with smallest modulus error constant; furthermore, this multiderivative method is more economical to implement than any other such method with the same modulus error constant.

The parameters for the (3,4) method are given by

$$a_{-1} = a_1 = \frac{2}{49}, \quad a^*_{-1} = a^*_1 = -\frac{1}{980}, \quad a^*_0 = \frac{131}{2940},$$
$$a^{**}_{-1} = a^{**}_1 = \frac{1}{44100}, \quad a^{**}_0 = \frac{31}{88200}, \qquad (11)$$

and after making approximations for y', as required, the solution vector $\underline{Y}$ is obtained by solving a non-linear system of the form (9).

Parameters for the method based on the (3,.3.) Pade approximant are given in the Appendix.

5. CONVERGENCE

Let $\underline{y} = (y(x_1), y(x_2), \ldots, y(x_N))^T$, let

$$g_m(\underline{y}) = h^2 \sum_{r=-1}^{1} a_r f_{m+r} + h^4 \sum_{r=-1}^{1} a^*_r f''_{m+r} + h^6 \sum_{r=-1}^{1} a^{**}_r f^{(iv)}_{m+r}, \quad m = 1, \ldots, N, \qquad (12)$$

where $f_m = f(x_m, y(x_m)), f''_m = d^2 f(x_m, y(x_m))$ and $f^{(iv)}_m = d^4 f(x_m, y(x_m))$,

so that $\underline{G}(\underline{y}) = (g_1, g_2, \ldots, g_N)^T$, and let $\underline{t} = (t_1, t_2, \ldots, t_N)^T$. Then every member of the family of multiderivative methods mentioned in §§2,3,4 can be expressed in system from as

$$A\underline{y} + \underline{G}(\underline{y}) = \underline{t} \qquad (13)$$

where A is the tridiagonal matrix

$$A = \begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix} \qquad (14)$$

It is easy to see that the nonlinear system (9) takes the form

$$A\underline{Y} + \underline{G}(\underline{Y}) = \underline{0} \qquad (15)$$

for each multiderivative method, where $\underline{O}$ is the zero vector of order N.

Now let $\underline{E} = (e_1, e_2, \ldots, e_N)^T = \underline{y} - \underline{Y}$ and define q_m, q_m^*, q_m^{**}, $m = 1, 2, \ldots, N$, by

$$q_m = \left(f_m - \tilde{f}_m\right)/e_m, \quad q_m^* = \left(f_m'' - \tilde{f}_m''\right)/e_m, \quad q_m^{**} = \left(f_m^{(iv)} - \tilde{f}_m^{(iv)}\right)/e_m. \qquad (16)$$

Subtracting (15) from (14) gives

$$(A + M)\underline{E} = \underline{t} \qquad (17)$$

where $M=[M_{i,j}]$ $(i,j = 1, 2, \ldots, N)$ is the tridiagonal matrix with $M_{ij} = 0$ for $|i - j| > 1$ and

$$M_{1,1} = h^2 a_0 q_1 + h^4 a_0^* q_1^* + h^6 a_0^{**} q_1^{**},$$

$$M_{1,2} = h^2 a_1 q_2 + h^4 a_1^* q_2^* + h^6 a_1^{**} q_2^{**},$$

$$M_{i,i\pm1} = h^2 a_{\pm1} q_{i\pm1} + h^4 a_{\pm1}^* q_{i\pm1}^* + h^6 a_{\pm1}^{**} q_{i\pm1}^{**}, \quad i = 2, \ldots, N - 1.$$

$$M_{i,i} = h^2 a_0 q_i + h^4 a_0^* q_i^* + h^6 a_0^{**} q_i^{**}, \quad i = 2, \ldots, N - 1,$$

$$M_{N,N-1} = h^2 a_{-1} q_{N-1} + h^4 a_{-1}^* q_{N-1}^* + h^6 a_{-1}^{**} q_{N\pm1}^{**},$$

$$M_{N,N} = h^2 a_0 q_N + h^4 a_{-1}^* q_N^* + h^6 a_0^{**} q_N^{**}. \qquad (18)$$

Defining, next, U by

$$U = \max_{a_0 \le x \le a_1} \left\{ \left| \frac{\partial f}{\partial y} \right|, \left| \frac{\partial f''}{\partial y} \right|, \left| \frac{\partial f^{(iv)}}{\partial y} \right| \right\} \qquad (19)$$

it is seen that all of q_m, q_m^* and q_m^{**} $(m = 1,2,..., N)$ are bounded in modulus by U. It then follows that

$$\|M\| \le h^2 \left\{ \sum_{r=-1}^{1} |a_r| + h^2 \sum_{r=-1}^{1} |a_r^*| + h^4 \sum_{r=-1}^{1} |a_r^{**}| \right\} U, \qquad (20)$$

so that

$$\|M\| \le \begin{cases} h^2 U, & \text{for the } (1,2) \text{ Pade method,} \\ h^2 \left(1 + \frac{1}{30} h^2 \right) U, & \text{for the } (2,3) \text{ Pade method,} \\ h^2 \left(1 + \frac{137}{2940} h^2 + \frac{1}{2520} h^4 \right) U, & \text{for the } (3,4) \text{ Pade method;} \end{cases} \qquad (21)$$

in (20) and (21) the norm referred to is the maximum norm.

It is well known that

$$\| A^{-1} \| \le (a_1 - a_0)^2/(8h^2) \qquad (22)$$

and from (5) it is seen that

$$\| \underline{t} \| = 0(h^{P+2}) . \qquad (23)$$

Then, provided $U<8/(a_1 - a_o)^2$, it can be shown that

$$\| A^{-1} \| . \| M \| < 1 \qquad (24)$$

for all methods of the family.

It follows from (17) that

$$\| \underline{E} \| \le \| A^{-1} \| . \| \underline{t} \| / (1 - \| A^{-1} \| . \| M \|) \qquad (25)$$

and thus from (22), (23) and (24) that

$$\| E \| = 0(h^P) .$$

The methods based on the use of the (1,2), (2,3) and (3,4) Padé approximants, the most accurate and economical of their respective orders, are thus second, fourth and sixth order convergent,, respectively.

6. NUMERICAL RESULTS

To test the effectiveness of the second, fourth and sixth order methods developed in the paper, each was tested on the following problems which were also used by Lal and Moffatt [10], Usmani [6,8,11] and Jain [12].

Problem 1

$$y''(x) = \frac{3}{2} y^2, \qquad 0 < x < 1$$

with boundary conditions

$$y(0) = 4 , y(1) = 1 .$$

The exact solution of Problem 1 is

$$y(x) = 4/(1+x)^2 .$$

Following Jain [12], the interval [0,1] was divided in equal sub-intervals of width $h = 2^{-m}$ ($m = 3, \ldots, 6$); the corresponding values of N are then given by $N = 2^m - 1$ .The numerical results for the three novel methods are recorded in Table 2, together with the equivalent results of the sixth order method of Jain [12] which was based on Lobatto quadrature. It was found that no more than four iterations were required to obtain $\| \underline{E} \|$ to three significant figures.

Table 2 here

It is noted from Table 2 that, for each of the three methods, $\| \underline{E} \|$ is reduced by the factor 2^p (approximately), where p is the order of the method, as h is successively halved. It is also noted that the novel sixth order method gives smaller error moduli than the corresponding methods outlined in Jain [12] and Usmani [6] where, for h =0.1, $\| \underline{E} \|$ =0.3E-04.

Problem 2

$$y''(x) = \frac{1}{2}(1 \pm x + y)^3 \quad , \qquad 0 < x < 1 \quad ,$$

with boundary conditions

$$y(0) = y(1) = 0 \ ,$$

for which the exact solution is

$$y(x) = 2/(2-x) - x - 1 \ .$$

For this problem, too, the interval [0,1] was divided into sub-intervals of width $h = 2^{-m}$ $(m = 3, \ldots, 6)$ so that, again, $N = 2^m - 1$. The values of $\| \underline{E} \|$ for each of the three methods are reported in Tables 3,4 5.

Tables 3,4,5 here

It is noted from these tables that, as for Problem 1, $\| \underline{E} \|$ is diminished by the approximate factor 2^p, where p is the order of the method, as h is successively halved.

Table 5 also includes values of $\|\underline{E}\|$ relating to the sixth order method based on Lobatto quadrature reported in [12] and, for m=3 and 4, on the sixth order method of Usmani [6] which was based on a sextic spline formulaiton. It is noted for comparison purposes that, for m = 3, an earlier sixth order method of Usmani [131 gives $\|\underline{E}\| = 0.8E\text{-}05$. Table 3 also contains, for m = 3 and 4, the values of $\|\underline{E}\|$ obtained by Usmani [6] using his second order cubic spline method, while Table 4 contains the values of $\|\underline{E}\|$ obtained using the fourth order methods of Usmani [6] based on quartic and quintic splines.

Overall, the results obtained using the novel methods to solve Problem 2 are superior to those obtained using competitive methods. It was found that, as for Problem †, no more than four iterations were needed to obtain $\| \underline{E} \|$ to three significant figures.

7. CONCLUSION

Second, fourth and sixth order multiderivative methods have been developed and analysed for the numerical solution of nonlinear second order boundary value problems.

The methods were seen to arise from the replacements of the exponential terms by Padé approximants in a three-point recurrence relation.

The principal parts of the local truncation errors of the methods were seen to be small in modulus; this was reflected in the numerical results obtained for two problems from the literature.

REFERENCES

1. HENRICI, P.:*Discrete variable methods in ordinary differential equations,* John Wiley and Sons, New York (1962).

2. LAMBERT, J.D.: *Computational methods in ordinary differential equations,* John Wiley and Sons, Chichester (1973).

3. BURDEN, R.L., FAIRES, J.D. and REYNOLDS, A.C.: *Numerical analysis,* 3rd edition, Prindle, Weber and Schmidt, Boston (1981).

4. ROBERTS, S.M. and SHIPMAN, J.S.: *Two-point boundary value problems, shooting methods,* American Elsevier, New York (1972).

5. AHLBERG, J.H., NILSON, E.N. and WALSH, J.L.: *The theory of splines and their applications,* Academic Press, New York (1967).

6. USMANI, R.A. : "Spline solutions for nonlinear two point boundary value problems. "Internat. J. Math. & Math. Sci., V.3 No.1 (1980), pp.151-167.

7. ALBASINEY, E.L. and HOSKINS, W.D.: "Cubic spline solutions to two point boundary value problems. "Computer Journal, V. 12 (1969), pp.151-153.

8. USMANI, R.A.: "Integration of nonlinear equations with mixed boundary conditions." Proc. Fourth Manitoba Conf. Num. Math. Winnipeg (1974), pp.361-373.

9. GERALD, C.F. and WHEATLEY, P.0. : *Applied numerical analysis,* 3rd edition, Addison-Wesley, Reading, Massachusetts (1984).

10. LAL, M. and MOFFATT, D.: "Picard's successive approximation for nonlinear two point boundary value problems." J. Comp. Appl. Math., V.8 No.4 (1982) pp.233-236.

11. USMANI, R.A.: "A note on the numerical integration of nonlinear problems." The Aligarh Bull, of Math., V.3 (1973), pp.15-25.

12. JAIN, M.K. : *Numerical solution of differential equations, 2nd edition,* Wiley Eastern Limited, New Dehli (1984).

13. USMANI, R.A.: "Integration of nonlinear boundary value problems of class M." Proc. Manitoba Conf. Num. Math. Winnipeg (1971), pp.589-606.

Table 1: Error constants.

Order	Padé approximant	error constant
2	(1,1)	$C_4 = \frac{1}{6}$
	(0,2)	$C_4 = -\frac{1}{12}$
	(1,2)	$C_4 = \frac{1}{36}$
	(2,1)	$C_4 = -\frac{1}{36}$
	(2,0)	$C_4 = -\frac{7}{12}$
4	(2,2)	$C_6 = -\frac{1}{360}$
	(2,3)	$C_6 = -\frac{1}{3600}$
	(3,2)	$C_6 = \frac{1}{3600}$
6	(3,3)	$C_8 = \frac{1}{50400}$
	(3,4)	$C_8 = \frac{1}{705600}$
	(4,3)	$C_8 = -\frac{1}{705600}$

Table 2. Values of $\| E \|$ for Problem 1 with $h=2^{-m}$ $(m = 3,4,5,6)$

M	N	Method			
		(1,2) Pade order 2	(2,3) Pade order 4	(3,4) Pade order 6	Jain [12] order 6
3	7	0.26E-2	0.13E-4	0.45E-6	0.49E-5
4	15	0.63E-3	0.71E-6	0.61E-8	0.80E-7
5	31	0.16E-3	0.43E-7	0.89E-10	0.13E-8
6	63	0.39E-3	0.26E-8	0.13E-11	0.20E-10

Table 3. Values of $\| \underline{E} \|$ for Problem 2 with $h=2^{-m}$ (m = 3,4,5,6) using second order methods

		Method	
m	N	Usmani [6] cubic spline	(1,2) Padé
3	7	0.12E-2	0.40E-3
4	15	0.00	0.98E-4
5	31	-	0.24E-4
6	63	-	0.61E-5

Table 4. Values of $\| \underline{E} \|$ for Problem 2 with $h=2^{-m}$ (m = 3,4,5,6) using four order methods

		Method		
M	N	Usmani [6] quartic spline	Usman i [6] qu in t i c spline	(2,3) Padé
3	7	0.16E-4	0.65E-5	0.13E-5
4	15	0.11E-5	0.22E-6	0.73E-7
5	31	-	-	0.45E-8
6	63	-	-	0.28E-9

Table 5. Values of $\| \underline{E} \|$ for Problem 2 with $h = 2^{-m}$ (m = 3,4,5,6) using sixth order methods

		Method		
m	N	Usmani [6] sextic spline	Jain [12]	(3,4) Padé
3	7	0.78E-5	0.27E-6	0.43E-8
4	15	0.20E-6	0.44E-8	0.57E-10
5	31	-	0.7 2E-10	0.84E-12
6	63	-	0.4 3E-11	0.13E-13

APPENDIX. The parameters $a_r, a_r^*, a_r^{**} (r = -1,0,1)$ for the methods not detailed in §§ 2,3,4.

$(0,2)$ Pade: $a_{-1} = a_1 = a_{-1}^* = a_0^* = a_1^* = a_{-1}^{**} = a_0^{**} = a_1^{**} = 0; \quad a_0 = 1.$

$(1,1)$ Pade: $a_{-1} = a_1 = \frac{1}{4}, a_0 = \frac{1}{2}; \quad a_{-1}^* = a_0^* = a_1^* = a_{-1}^{**} = a_0^{**} = a_1^{**} = 0.$

$(2,1)$ Pade: $a_{-1} = a_1 = \frac{1}{9}, a_0 = \frac{7}{9}; \quad a_{-1}^* = a_1^* = -\frac{1}{16}, a_0^* = a_{-1}^{**} = a_0^{**} = a_1^{**} = 0.$

$(2,0)$ Pade: $a_0 = 1 \quad a_{-1}^* = a_1^* = -\frac{1}{2}, \; a_{-1} = a_1 = a_0^* = a_{-1}^{**} = a_0^{**} = a_1^{**} = 0.$

$(2,2)$ Pade: $a_{-1} = a_1 = \frac{1}{12}, a_0 = \frac{5}{6}; \quad a_{-1}^* = a_1^* = \frac{-1}{144} = a_0^* = \frac{1}{72}, a_{-1}^{**} = a_0^{**} = a_1^{**} = 0$

$(3,2)$ Pade: $a_{-1} = a_1 = \frac{3}{50}, a_0 = \frac{22}{25}, a_{-1}^* = a_1^* = -\frac{1}{400}, a_0^* = \frac{17}{600},$

$a_{-1}^{**} = a_1^{**} = \frac{1}{3600}, a_0^{**} = 0.$

$(3,3)$ Pade: $a_{-1} = a_1 = \frac{1}{20}, a_0 = \frac{9}{10}, a_{-1}^* = a_1^* = -\frac{1}{600}, a_0^* = \frac{11}{300},$

$a_{-1}^{**} = a_1^{**} = \frac{1}{14400}, a_0^{**} = \frac{1}{7200}.$