MATHEMATICS OF COMPUTATION, VOLUME 30, NUMBER 133
JANUARY 1976, PAGES 92–103

Tridiagonal Fourth Order Approximations to General Two-Point Nonlinear Boundary Value Problems with Mixed Boundary Conditions

By Robert S. Stepleman

Abstract. This paper develops fourth order discretizations to the two-point boundary value problem

$$y^{(2)}(t) = f(t, y(t), y^{(1)}(t)),$$

$$\alpha_0 y(0) - \beta_0 y^{(1)}(0) = \delta_0, \quad \alpha_1 y(1) + \beta_1 y^{(1)}(1) = \delta_1.$$

These discretizations have the desirable properties that they are tridiagonal and of "positive type".

1. Introduction. In this paper we consider discretization techniques for the nonlinear two-point boundary value problem

$$\text{(1.1)} \quad \begin{aligned} &\text{(a)} \quad y^{(2)}(t) = f(t, y(t), y^{(1)}(t)), \\ &\text{(b)} \quad \alpha_0 y(0) - \beta_0 y^{(1)}(0) = \delta_0, \quad \alpha_1 y(1) + \beta_1 y^{(1)}(1) = \delta_1, \end{aligned}$$

where $f: I \times R^2 \longrightarrow R^2$ and $I = [-\epsilon, 1 + \epsilon]$, for some $\epsilon > 0$. Here $y^{(k)}(t)$ represents the kth derivative. We also assume

$$\text{(1.2)} \quad \alpha_0 + \alpha_1 > 0, \quad \alpha_0, \beta_0, \alpha_1, \beta_1 \geqslant 0, \quad \alpha_0 + \beta_0 > 0, \quad \alpha_1 + \beta_1 > 0.$$

In particular, we shall derive for the first time, direct finite-difference analogues of (1.1) which have the following two key properties. First, they will have solutions which approximate the solution to (1.1) with order h^4 globally over the mesh points of a uniform mesh of width h; and second, when applied to the linear problem

$$\text{(1.3)} \quad y^{(2)}(t) + p(t) y^{(1)}(t) + q(t) y(t) = r(t)$$

with condition (1.1b), the linear system of equations resulting will be both tridiagonal and of "positive type" (as will the Jacobian matrix of (1.1)). Thus, we will have a method that will yield a high order solution and be easy to analyze. The method will have the same number of matrix operations as solving (1.1) to order h^2; however, there will be more functional evaluations.

Recent work on algorithms for two-point boundary value problems for first order nonlinear systems (see e.g. Keller [7]) has produced methods that can be applied in much more general circumstances than ours. This does not mean the techniques in this paper are only of historical interest; however, in that context they do close a gap

Received August 22, 1974; revised February 19, 1975.
AMS (MOS) subject classifications (1970). Primary 65L10.
Key words and phrases. Boundary value problems, mixed boundary conditions, fourth order discretization.

in the theory and application of positive type finite-difference methods to two-point boundary value problems. We shall show, for problem (1.1), that the methods in this paper are competitive and useful.

The classical finite-difference algorithms for (1.1) are tridiagonal and of positive type, but converge only of order h^2 (see e.g. Keller [6], or Aziz and Hubbard [2] for the linear case). These algorithms combined with Richardson extrapolation (when applicable) do give an order h^4 method, however, an indirect one. A comparison between this method and our direct order h^4 algorithm will be given in Section 4. Collocation techniques have also been applied to (1.1) to give fourth order methods (see e.g. Daniel and Swartz [4] or Russell and Shampine [10]). Depending on the basis used for the splines in this technique, the resulting matrix problem may be anywhere from an effective bandwidth of five, to a matrix problem requiring $O(h^{-2})$ operations to solve the linear system. A more serious difficulty with these methods is that, at least for some basis that give "small" bandwidth, terms of $O(h^{-1})$ can appear on the subdiagonal with terms of $O(1)$ on the diagonal; this can give stability problems due to roundoff error and partial pivoting may be necessary (see [4, pp. 18–22]). Since our algorithm is of positive type, it is diagonally dominant and this difficulty cannot occur. In Pereyra [9], it is suggested that difference corrections could be applied to the $O(h^2)$ analogue of (1.1) to obtain an $O(h^4)$ solution. This has not yet been done or rigorously justified, so it is not clear how the two methods compare. However, based on some other results in [9], it appears that Pereyra's method might well be the best way to obtain high order solutions of (1.1). Shoosmith [11] suggests replacing all derivatives in (1.1) by their fourth order finite-difference analogues. This yields a five diagonal matrix, which is not of positive type (but possibly is monotone). The local truncation error of this method is order h^4, but no global truncation error estimates or stability results are given when the boundary conditions contain the first derivative.

We note that we make no attempt to derive conditions for the existence and uniqueness of solutions to (1.1), being concerned here only with the numerical analysts problem of determining the conditions under which a given numerical method will converge and its order of convergence. Hence we assume that at least one solution to (1.1) exists, and let y denote any such solution.

In Section 2 we will consider the local truncation error of the method. Since all the proofs in that section are simple in idea, based on Taylor series and algebraic manipulation, but somewhat lengthy because of the details involved, we shall omit most of them. In Section 3 we obtain our global error estimate and stability result, while in Section 4 we consider some numerical experience and computational details about the method.

Extensions to partial differential equations of the idea of obtaining high order finite-difference approximations to a complicated operator that have the same matrix structure as the approximation to some simpler operator are possible. For one application of this idea see Stepleman [12].

2. The Method and Local Truncation Error. We shall consider two slightly different discretizations. The one with the fewer function evaluations will require stronger

hypothesis on f to be of fourth order. Thus, which should be used in a given situation will depend on f.

Let $h = 1/N$, N some integer, $t_{hn} = (n-1)h$, $n = 0, \ldots, N+2$ and $n = 3/2, N + \frac{1}{2}$. Define $y_{hn} = y(t_{hn})$ and $y^{(k)}_{hn} = y^{(k)}(t_{hn})$. Consider first the approximation at the boundary. If the global discretization error is to be fourth order, we will need an approximation to $y^{(1)}(0)$ and $y^{(1)}(1)$ of that order. Set

$$y'^{0}_{mh,i} = (y_{h,i+m} - y_{h,i-m})/2mh. \tag{2.1}$$

A standard technique is to use $y'^{0}_{h,1}$ to approximate $y^{(1)}(0)$. However,

$$y'^{0}_{h,1} = y^{(1)}(0) + \frac{h^2}{6} y^{(3)}(0) + O(h^4) \tag{2.2}$$

so that this discretization is only $O(h^2)$. If we had an approximation to $y^{(3)}(0)$ that was also $O(h^2)$, we could then use (2.2) in the obvious way to get an $O(h^4)$ discretization of $y^{(1)}(0)$. This is what we want to do, recalling, however, that we have an added constraint that the resulting discretization gives rise to a tridiagonal matrix problem.

We now introduce some notation. Set

$$\begin{cases} y'^{+}_{mh,i} = (3y_{hi} - 4y_{h,i-m} + y_{h,i-2m})/2mh, \\ y'^{-}_{mh,i} = (4y_{h,i+m} - 3y_{hi} - y_{h,i+2m})/2mh, \end{cases} \tag{2.3}$$

and

$$f^{0}_{mh,i} = f(t_{hi}, y_{hi}, y'^{0}_{mh,i}) \tag{2.4}$$

with similar definitions for $f^{+}_{mh,i}$ and $f^{-}_{mh,i}$. Also, set

$$y^{a}_{mh,i} = (y_{h,i+m} + y_{h,i-m})/2. \tag{2.5}$$

A simple Taylor series argument shows:

LEMMA 2.1. *Let $y \in C^5[I]$. Then*

$$y^{a}_{mh,i} = y_{hi} + \frac{(mh)^2}{2} y^{(2)}_{hi} + O(h^4) \tag{2.6}$$

and

$$\begin{cases} y'^{+}_{mh,i} = y^{(1)}_{hi} - \dfrac{(mh)^2}{3} y^{(3)}_{h,i-m} - \dfrac{(mh)^3}{12} y^{(4)}_{h,i-m} + O(h^4), \\ y'^{-}_{mh,i} = y^{(1)}_{hi} - \dfrac{(mh)^2}{3} y^{(3)}_{h,i+m} + \dfrac{(mh)^3}{12} y^{(4)}_{h,i+m} + O(h^4). \end{cases} \tag{2.7}$$

Since these discretizations are only order h^2, we will need to improve them before we use them. We introduce some new discretizations to this end. Set:

$$\hat{y}_{hi} = y^{a}_{h/2,i} - \frac{h^2}{8} f(t_{hi}, y^{a}_{h/2,i}, y'^{0}_{h/2,i}), \qquad i = \frac{3}{2}, \quad N + \frac{1}{2}, \tag{2.8}$$

$$\begin{cases} y^{p0}_{mh,i} = y'^{0}_{mh,i} - \dfrac{mh}{12}(f^{+}_{mh,i+m} - f^{-}_{mh,i-m}), \\ y^{p+}_{mh,i} = y'^{+}_{mh,i} + \dfrac{mh}{6}(f^{+}_{mh,i} - f^{-}_{mh,i-2m}), \\ y^{p-}_{mh,i} = y'^{-}_{mh,i} + \dfrac{mh}{6}(f^{+}_{mh,i+2m} - f^{-}_{mh,i}), \end{cases} \tag{2.9}$$

and

$$f^{p0}_{mh,i} = f(t_{hi}, y_{hi}, y^{p0}_{mh,i}) \tag{2.10}$$

with similar definitions for $f^{p-}_{mh,i}$ and $f^{p+}_{mh,i}$.

The next lemma, whose proof is based on the Mean Value Theorem and some manipulation is a key to our local truncation error results.

LEMMA 2.2. *Let $y \in C^5[I]$ with $f_p(r, p, q)$ and $f_q(r, p, q)$ bounded and $f_q(r, p, q)$ Lipschitz continuous in q on $I \times R^2$. Then*

$$y^{(3)}_{hi} = (f^{+}_{mh,i+m} - f^{-}_{mh,i-m})/2mh + O(h^2). \tag{2.11}$$

Also,

$$\begin{cases} y^{p0}_{mh,i} = y^{(1)}_{hi} + O(h^4), \\ y^{p+}_{mh,i} = y^{(1)}_{hi} + O(h^3), \\ y^{p-}_{mh,i} = y^{(1)}_{hi} + O(h^3), \end{cases} \tag{2.12}$$

and $\hat{y}_{hi} = y_{hi} + O(h^4)$.

Proof. Calculate that

$$y^{(3)}_{hi} - \frac{f^{+}_{mh,i+m} - f^{-}_{mh,i-m}}{2mh} = y^{(3)}_{hi} - \frac{y^{(2)}_{h,i+m} - y^{(2)}_{h,i-m}}{2mh}$$

$$+ \frac{y^{(2)}_{h,i+m} - f^{+}_{mh,i+m}}{2mh} + \frac{f^{-}_{mh,i-m} - y^{(2)}_{h,i-m}}{2mh} = T_h.$$

Then using the Mean Value Theorem on the second and third term gives

$$T_h = [y^{(3)}_{hi} - (y^{(2)}_{h,i+m} - y^{(2)}_{h,i-m})/2mh]$$

$$+ f_q(\theta_1)(y^{(1)}_{h,i+m} - y^{(1)}_{h,i-m} - y'^{+}_{mh,i+m} + y'^{-}_{mh,i-m})/2mh$$

$$+ (f_q(\theta_2) - f_q(\theta_1))(y'^{-}_{mh,i-m} - y^{(1)}_{h,i-m})/2mh.$$

Here $\theta_1, \theta_2 \in I \times R^2$ and $\theta_{1i} = \theta_{2i} + O(h)$, $i = 1, 2, 3$. That the first bracket is order h^2 follows from the differentiability of y, that the second is follows from (2.7) and that the third is follows from (2.7) and the Lipschitz continuity of f_q.

The result (2.12) follows from (2.11) and (2.7), while the last conclusion follows from (2.6) and a very similar argument.

We would like to approximate $y^{(3)}(0) \doteq (f^{+}_{h,2} - f^{-}_{h,0})/2h$. However, since this directly involves y_{h0}, y_{h1}, and y_{h2}, it will not give a triangular matrix approximation. We would like an order h^2 approximation that does not involve y_{h0}. What we will do is to use (2.3) and interpolate a point midway between t_{h1} and t_{h2}. Thus, to approximate $y^{(3)}(0)$ we use the expression

$$(4f^{p0}_{h/2,3/2} - 3f^{p-}_{h/2,1} - f^{p+}_{h/2,2})/h.$$

However, this contains the nonmeshpoint $y_{h,3/2}$. We substitute $\hat{y}_{h,3/2}$ for this point whenever it appears. Thus, we approximate

$$y^{(3)}(0) \doteq (4\hat{f}^{p0}_{h/2,3/2} - 3\hat{f}^{p-}_{h/2,1} - \hat{f}^{p+}_{h/2,2})/h \equiv y^{(3)-}_{h,1},$$

where the $\hat{}$ denotes the substitution of $\hat{y}_{h,3/2}$ for $y_{h,3/2}$. In a similar manner, we approximate

$$y^{(3)}(1) = (3\hat{f}^{p+}_{h/2,N+1} - 4\hat{f}^{p0}_{h/2,N+\frac{1}{2}} + \hat{f}^{p-}_{h/2,N})/h = y^{(3)+}_{h,N+1}.$$

Here the $\hat{}$ denotes the substitution of $\hat{y}_{h,N+\frac{1}{2}}$ for $y_{h,N+\frac{1}{2}}$.

Then in a manner very similar to Lemma 2.2 we prove:

LEMMA 2.3. *Let the hypothesis of Lemma 2.2 hold. Then*

$$y^{(3)}(0) = y^{(3)-}_{h,1} + O(h^2), \qquad y^{(3)}(1) = y^{(3)+}_{h,N+1} + O(h^2),$$

and

$$\begin{cases} y^{(1)}(0) = y'^{0}_{h,1} - \dfrac{h^2}{6} y^{(3)-}_{h,1} + O(h^4), \\ y^{(1)}(1) = y'^{0}_{h,N+1} - \dfrac{h^2}{6} y^{(3)+}_{h,N+1} + O(h^4). \end{cases} \tag{2.13}$$

Because of the first term on the right in (2.13) these discretizations still contain the point y_{h0} or $y_{h,N+2}$, respectively. However, these will disappear when we combine them with the interior discretization, which we now consider.

At the points t_{hn}, $N = 1, \ldots, N+1$, we would like to approximate the differential equation to order h^4. The standard discretization for the second derivative satisfies

$$y^{(2)}_{hi} = (y_{h,1+i} - 2y_{hi} + y_{h,i-1})/h^2 - \frac{h^2}{12} y^{(4)}(t_{hi}) + O(h^4). \tag{2.14}$$

Thus, if we can approximate $y^{(4)}(t_{hi})$ to order h^2 using only $y_{h,1+i}$, y_{hi}, and $y_{h,i-1}$ we will have what we want. Since

$$(y^{(2)}(t_{h,i+1}) - 2y^{(2)}(t_{hi}) + y^{(2)}(t_{h,i-1}))/h^2 = y^{(4)}(t_{hi}) + O(h^2), \tag{2.15}$$

we need only approximate the expression on the left in (2.15) to $O(h^2)$. Thus, we approximate this expression by $(f^{p-}_{h,i-1} - 2f^{p0}_{h,i} + f^{p+}_{h,i+1})/h^2$.

Then in a manner analogous to Lemma 2.2, we have:

LEMMA 2.4. *Suppose the hypothesis of Lemma 2.2 holds. Then*

$$y^{(4)}(t_{hi}) = (f^{p-}_{h,i-1} - 2f^{p0}_{h,i} + f^{p+}_{h,i+1})/h^2 + O(h^2), \qquad 1 \leqslant i \leqslant N+1. \tag{2.16}$$

Then combining (2.14), (2.15), (2.16) we approximate problem (1.1) by

$$y_{h,i+1} - 2y_{hi} + y_{h,i-1} = \frac{h^2}{12}(f^{p-}_{h,i-1} + 10f^{p0}_{hi} + f^{p+}_{h,i+1}) + h^2\tau_{hi}, \qquad i = 1, \ldots, N+1, \tag{2.17}$$

$$\begin{cases} y_{h0} = \dfrac{2h}{\beta_0}[\delta_0 - \alpha_0 y_{h1}] + y_{h2} - \dfrac{h^3}{3} y^{(3)-}_{h,1} + h\tau_{h0}, \\ y_{h,N+2} = \dfrac{2h}{\beta_1}[\delta_1 - \alpha_1 y_{h,N+1}] + y_{hN} + \dfrac{h^3}{3} y^{(3)+}_{h,N+1} + h\tau_{h,N+2}. \end{cases} \tag{2.18}$$

Substituting (2.18) into (2.17) where required for $i = 1$ and $N + 1$ gives rise to the required tridiagonal matrix problem. Here the τ_{hi}, $i = 0, \ldots, N + 2$, represent the local truncation errors caused by replacing all derivatives by difference quotients.

Combining Lemma 2.1–Lemma 2.4 yields:

THEOREM 2.5. *Suppose the hypothesis of Lemma* 2.2 *holds. If, in addition,* $y \in C^6(I)$, *then*

$$\max_{0 \leqslant i \leqslant N+2} |\tau_{hi}| = O(h^4). \tag{2.19}$$

Discretizations, somewhat like (2.17) and (2.18), can be found in the literature for problems that do not involve $y^{(1)}$ in the boundary condition or in the function (see e.g. Collatz [3, Chapter III] or Allen [1]). However, none appear for the general problem (1.1) which are order h^4.

If higher partials of f exist and are bounded, we can save a function evaluation at the interior grid points by using the discretization:

$$y_{h,i+1} - 2y_{hi} + y_{h,i-1} = \frac{h^2}{12}\left[f^{-}_{h,i-1} + 4f^{0}_{hi} + 6f^{p0}_{hi} + f^{+}_{h,i+1}\right] + h^2\tau^{*}_{hi}. \tag{2.20}$$

In this case we have the following result:

THEOREM 2.6. *Let* $y \in C^6(I)$ *and* $f_q(r, p, q)$ *be twice continuously differentiable in* r, p, q *with bounded partials through order three on* $I \times R^2$. *Then*

$$\max_{0 \leqslant i \leqslant N+2} |\tau^{*}_{hi}| = O(h^4). \tag{2.21}$$

It should be noted that if, say, $\beta_0 = 0$ then t_{h0} is ignored and t_{h1} is considered a boundary point. In this case (2.17) starts with $i = 2$ instead of $i = 1$.

3. Global Discretization Error. In this section we show that if y is the solution to (1.1) and u_{hn} the solution to either (2.17) or (2.20) and (2.18) with the local truncation error set to zero, then

$$\max_{1 \leqslant n \leqslant N+1} |u_{hn} - y_{hn}| = O(h^4). \tag{3.1}$$

LEMMA 3.1. *Set for* $N = 1/h$, A_N *to be the* $(N + 1) \times (N + 1)$ *tridiagonal matrix*

$$A_N = \begin{bmatrix} \alpha_0^N + \gamma_0 h & -\alpha_0^N & & \\ -1 - \beta_1^N h & 2 + (\beta_1^N + \alpha_1^N)h + \gamma_1^N h^2 & -1 - \alpha_1^N h & \\ & \cdot & \cdot & \cdot \\ & & -1 - \beta_{N-1}^N h & 2 + (\beta_{N-1}^N + \alpha_{N-1}^N)h + \gamma_{N-1}^N h^2 \\ & & -\beta_N^N & \beta_N^N + \gamma_N h \end{bmatrix}.$$

Suppose $\gamma_i \geqslant 0$, $i = 0, \ldots, N$, $\beta_N^N, \alpha_0^N \geqslant 0$, $\beta_N^N + \gamma_N \geqslant m > 0$, $\alpha_0^N + \gamma_0 \geqslant m$, $N = 1, 2, \ldots$ *and* $\gamma_0 + \gamma_N > 0$. *Also assume* $|\alpha_i^N| \leqslant K_1$, $|\beta_i^N| \leqslant K_2$, $i = 0, \ldots, N$, *where* K_1 *and* K_2 *are constants independent of* N. *Then for* N *sufficiently large* (h *sufficiently small*) A_N *is nonsingular,* $A_N^{-1} = (r_{ij})$ *satisfies* $r_{ij} \geqslant 0$, $i, j = 1, \ldots, N + 1$ *and*

$$\text{(3.2)}\qquad \begin{cases} \max\limits_{1\leqslant j\leqslant N+1} \max(r_{j,1}, r_{j,N+1}) = O(N), \\ \|A_N^{-1}\|_\infty = O(N^2). \end{cases}$$

Proof. That A_N^{-1} exists and $A_N^{-1} \geqslant 0$ for N sufficiently large follows since A_N is irreducibly diagonally dominant with positive diagonal and nonpositive off diagonal. Since $\gamma_0 + \gamma_N > 0$, we assume without loss of generality $\gamma_N > 0$. We now show (3.2). Set $V_i = \exp(2s) - \exp((i-1)sh)$, $i = 1, \ldots, N+1$, $s > 0$. Calculate

$$\begin{aligned}(A_N V)_i &= (-1 - \beta_i^N h)(\exp(2s) - \exp((i-1)sh)) \\ &\quad + [(2 + \beta_i^N + \alpha_i^N)h + \gamma_i^N h^2]\,[\exp(2s) - \exp(ish)] \\ &\quad + (-1 - \alpha_i^N h)\,(\exp(2s) - \exp((i+1)sh)), \qquad i = 2, \ldots, N, \\ &\geqslant (1 + \beta_i^N h)\exp((i-1)sh) - [2 + (\beta_i^N + \alpha_i^N)h]\exp(ish) \\ &\quad + (1 + \alpha_i^N h)\exp((i+1)sh) \\ &= \exp((i-1)sh)[(1 + \beta_i^N h) - [2 + (\beta_i^N + \alpha_i^N)h]\exp(sh) \\ &\qquad\qquad + (1 + \alpha_i^N h)\exp(2sh)] \\ &= \exp((i-1)sh)h^2\left[\frac{(\exp(sh) - 1)}{h}\left(\frac{\exp(sh) - 1}{h} + (\alpha_i^N - \beta_i^N)\right)\right].\end{aligned}$$

Then since $\lim_{h\to 0}((\exp(sh) - 1)/h) = s$, it follows for $s > K_1 + K_2$ and h sufficiently small that $(A_N V)_i \geqslant Kh^2$ where K is a constant independent of h or i.

Calculate

$$\begin{aligned}(A_N V)_1 &= (\exp(2s) - 1)\,(\alpha_0^N + \gamma_0 h) - \alpha_0^N(\exp(2s) - \exp(sh)) \\ &= \alpha_0^N(\exp(sh) - 1) + (\exp(2s) - 1)\gamma_0 h \\ &\geqslant h\left[\alpha_0^N \frac{(\exp(sh) - 1)}{h} + (\exp(2s) - 1)\gamma_0\right] \geqslant Kh\end{aligned}$$

for $s > 0$ since $\alpha_0^N \geqslant 0$, $\gamma_0 \geqslant 0$, $\alpha_0^N + \gamma_0 \geqslant m > 0$.

$$\begin{aligned}(A_N V)_{N+1} &= (\exp(2s) - \exp(s))((\beta_N^N + \gamma_N h) - \beta_N^N(\exp(2s) - \exp((1-h)s))) \\ &= (\exp(2s) - \exp(s))(\gamma_N h + \beta_N^N(\exp((1-h)s) - \exp(s))) \\ &\geqslant \gamma_N h\left[\exp(2s) - \exp(s) - \frac{K_2}{\gamma_N}\frac{\exp(s) - \exp((1-h)s)}{h}\right] \geqslant Kh\end{aligned}$$

for s so large that $e^s > 1 + K_2 s/\gamma_N$ and h sufficiently small. Since $(A^{-1}AV) = V$, and $(AV)_K \geqslant 0$, $K = 1, \ldots, N+1$, we have for h sufficiently small,

$$\sum_{j=1}^{N+1} r_{ij} \min_{1\leqslant K\leqslant N+1} (AV)_K \leqslant V_i,$$

so that for some constant K

$$\max_{1 \leqslant i \leqslant N+1} \sum_{j=1}^{N+1} r_{ij} \leqslant \frac{K}{h^2}.$$

Similarly, for $i = 1$ or $N + 1$ and h sufficiently small there is a constant K such that

$$r_{ji} < K/h, \qquad j = 1, \ldots, N + 1.$$

Thus, (3.2) follows.

THEOREM 3.2. *Suppose $w \in R^{N+1}$ satisfies for $h = 1/N$*

$$A_N w = h^2 f_N + h^2 \tau_N, \tag{3.3}$$

where

$$\max_{2 \leqslant i \leqslant N} |\tau_i| \leqslant Ch^4, \qquad |\tau_i| \leqslant Ch^3, \quad i = 1, N + 1, \tag{3.4}$$

and

$$|f_{N,i}| \leqslant Kh\|w\|_\infty, \quad i = 2, \ldots, N, \qquad |f_{N,i}| \leqslant K\|w\|_\infty, \quad i = 1, N + 1. \tag{3.5}$$

Then for N sufficiently large

$$\|w\|_\infty = O(h^4). \tag{3.6}$$

Proof. From (3.3)

$$w_i = h^2 \sum_{j=1}^{N+1} r_{ij} f_{N,j} + h^2 \sum_{j=1}^{N+1} r_{ij} \tau_{N,j}.$$

Then from Lemma 3.1 and (3.4) and (3.5) we obtain

$$\begin{aligned} |w_i| &\leqslant h^2(r_{i1} + r_{i,N+1})(|f_{N,1}| + |f_{N,N+1}| + |\tau_{N,1}| + |\tau_{N,N+1}|) \\ &\quad + h^2 \sum_{j=2}^{N} r_{ij}(|\tau_{N,j}| + |f_{N,j}|) \\ &\leqslant h(2K\|w\|_\infty + 2Ch^3) + h^2\|A_N^{-1}\|_\infty(Ch^4 + Kh\|w\|_\infty); \end{aligned}$$

and thus,

$$\|w\|_\infty \leqslant 3kh\|w\|_\infty + (2C + h^2\|A_N^{-1}\|_\infty)h^4.$$

The conclusion now follows.

The following corollary provides both a global error estimate and stability result.

COROLLARY 3.3. *Let $y \in C^6[I]$ and the hypothesis of Theorem 2.2 hold. Suppose in addition*

$$f_p(r, p, q) \geqslant 0 \quad \text{on } I \times R^2. \tag{3.7}$$

Then with u_{hn} the solution to (2.17) and (2.18) with $\tau_{hn} = 0$, $n = 0, \ldots, N + 2$

$$\max_{1 \leqslant n \leqslant N+1} |u_{hn} - y_{hn}| = O(h^4). \tag{3.8}$$

Proof. Let $e_{hn} = y_{hn} - u_{hn}$, $n = 1, \ldots, N + 1$. Then e_{hn} satisfies

$$e_{h,n-1} - 2e_{hn} + e_{h,n+1}$$

$$= \frac{h^2}{12}\left[f^{p-}_{h,n-1}(y) - f^{p-}_{h,n-1}(u) + 10(f^{p0}_{h,n}(y) - f^{p0}_{h,n}(u)) + f^{p+}_{h,n+1}(y) - f^{p+}_{h,n+1}(u)\right]$$

$$+ h^2\tau_{hn}, \qquad 1 \leqslant n \leqslant N+1.$$

Here, for example, $f^{p0}_{h,n}(u) = f(t_{hn}, u_{hn}, u^{p0}_{hn})$. Then using the Mean Value Theorem repeatedly we obtain

$$\left\{12 - h^2 f_p(\theta_1) + \left[\frac{3}{2} f_q(\theta_1) + 5 f_q(\theta_2) - \frac{1}{2} f_q(\theta_3)\right] h + \frac{h^2}{24}(f_q(\theta_4) + 3f_q(\theta_5))(f_q(\theta_1) + 10 f_q(\theta_2) + f_q(\theta_3))\right\} e_{h,n-1}$$

$$+ \left\{-24 - 10 f_p(\theta_2)h^2 - 2h(f_q(\theta_1) - f_q(\theta_3)) - \frac{h^2}{6}(f_q(\theta_4) + f_q(\theta_5))(f_q(\theta_1) + 10 f_q(\theta_2) + f_q(\theta_3))\right\} e_{hn}$$

$$+ \left\{12 - h^2 f_p(\theta_3) + \left[\frac{1}{2} f_q(\theta_1) - 5 f_q(\theta_2) - \frac{3}{2} f_q(\theta_3)\right] h + \frac{h^2}{24}(3f_q(\theta_4) + f_q(\theta_5))(f_q(\theta_1) + 10 f_q(\theta_2) + f_q(\theta_3))\right\} e_{h,n+1}$$

$$= \frac{h^3}{12}(f_p(\theta_5)e_{h,n-1} - f_p(\theta_4)e_{h,n+1})(f_q(\theta_1) - 20 f_q(\theta_2) + f_q(\theta_3)) + 12h^2\tau_{hn}.$$

Here $\theta_i \in [0, 1] \times R^2$, $i = 1, \ldots, 5$, and θ_i depends on n. Similarly, the boundary conditions (2.18) give:

$$\frac{\beta_0}{2} e_{h0} + \left\{\alpha_0 + \frac{\beta_0}{6}[3 f_q(\theta_6) - 4 f_q(\theta_7) + f_q(\theta_8)]\right\} h e_{h1} + \left\{-\frac{\beta_0}{2} + \frac{\beta_0 h}{6}[4 f_q(\theta_7) - 3 f_q(\theta_6) - f_q(\theta_8)]\right\} e_{h2} = h^2 F_0 + \tau_{h0},$$

$$\frac{\beta_1}{2} e_{h,N+2} + \left\{\alpha_1 + \frac{\beta_1}{6}[3 f_q(\theta_9) - 4 f_q(\theta_{10}) + f_q(\theta_{11})]\right\} h e_{h,N+1} + \left\{-\frac{\beta_1}{2} + \frac{\beta_1 h}{6}[-3 f_q(\theta_9) + 4 f_q(\theta_{10}) - f_q(\theta_{11})]\right\} e_{hN} = h^2 F_1 + \tau_{h,N+2}.$$

Here F_0 and F_1 satisfy

$$|F_0| \leqslant K_0 \max(|e_{h2}|, |e_{h1}|), \qquad |F_1| \leqslant K_1 \max(|e_{hN}|, |e_{h,N+1}|)$$

with K_0, K_1 constant independent of h.

After eliminating (when necessary) e_{h0} and $e_{h,N+2}$, we can now apply the last theorem and obtain the desired result using (1.2) and Theorem 2.5.

In an exactly analogous manner the next corollary follows.

COROLLARY 3.4. *Let the hypothesis of Theorem* 2.6 *hold. Suppose in addition* (3.7) *holds. Then with* u_{hn} *the solution to* (2.20) *and* (2.18) *with the local truncation error set to zero* (3.8) *holds.*

The next results show that the discrete problems have a solution.

THEOREM 3.5. *Consider the system of equations.*

$$A_N w = h^2 f_N + h^2 b, \tag{3.9}$$

where b *is a constant vector and* f_N *is as in Theorem* 3.2. *Then* (3.9) *has a solution for* h *sufficiently small.*

Proof. The system (3.9) is equivalent to the system

$$w = h^2 A_N^{-1} f_N + h^2 A_N^{-1} b.$$

Then, proceeding exactly as in Theorem 3.2 using Lemma 3.1, we can conclude that as long as h is sufficiently small

$$\| w\|_\infty \leqslant \frac{h^2 \| A_N^{-1}\|_\infty \| b\|_\infty}{1 - K^* h} = T$$

where K^* is independent of h. Thus, the sphere $S = \{ w | \ \| w\|_\infty \leqslant T\}$ is mapped into itself; and the conclusion follows by the Brouwer fixed point theorem.

Since by the Mean Value Theorem, our discrete problems are equivalent to $A_N u = h^2 F(0) + h^2 F + h^2 \delta$, where $[F(0)]_i = f(t_{hi}, 0, 0)$, $\delta = (\delta_0, 0, \ldots, 0, \delta_1)^T$ and F is given analogously to that in Corollary 3.3, the last theorem applies.

Using the results we have obtained here, it is not difficult to follow standard techniques to show that under slightly strengthened hypothesis Newton's Method converges to the solution of our discrete systems. For more on this see Henrici [5], Keller [6] or Lees [8].

4. Numerical consideration. In this section we will consider the method applied to problem (1.1) when the first derivative appears in both the differential equation and the boundary condition. (If, for example, no first derivative appears at all, then the algorithm reduces to the well-known Numerov method; see Lees [8].)

Table 4.1 gives a comparison of the amount of work needed to solve problem (1.1) for a fixed N, three fourth order methods and a linear f. In the table, function evaluations refer to an evaluation of $f(t, y(t), y^{(1)}(t))$. Method (1) is (2.17)–(2.18) and (2.18)–(2.20); Method (2) is the classical $O(h^2)$ algorithm followed by Richardson extrapolation; Method (3) refers to collocation type procedures. We need not consider Keller's Method here as it cannot compete with (2) for this simple problem. We have chosen a linear equation to remove the number of iterations for solving the nonlinear equations as a variable. However, if one considers, say, the number of operations per Newton step, there is very little difference in the conclusions that can be drawn.

From Table 4.1 it appears that the superiority of (1) or (2) on the basis of work done depends on the relative cost of multiplication versus the particular functional evaluation. The comparison between (1) and (3) is even more complex. At least two things must be considered, both dependent on the basis chosen for the space of splines.

Measured	(1)	(2)	(3)
Function Evaluations	$4N - 5N$	$3N$	N
Multiplications to Solve Linear Equations	$5N$	$15N$	$11N - O(N^2)$

TABLE 4.1

We must be aware not only of the bandwidth that the choice gives, but of its stability properties. As was pointed out in Section 1, some methods that give small bandwidth allow large subdiagonal elements relative to the diagonal elements, leaving the possibility of instability due to roundoff error. This, of course, can be corrected by pivoting in the linear system. It remains an open question whether this is often needed; however, it is clear that the small bandwidth methods, which appear to need less work than our algorithm, must be used with some care.

Since we could not find any problems in the literature for collocation when the boundary conditions contain derivatives, we will not consider these any further. However, based on some comparisons given in [10], for the no derivative case, it seems reasonable to expect collocation to give errors of about the same magnitude as our method.

In the following table, which describes some of the numerical experiments performed, Methods (1) and (2) remain the same while Method (3) is now the classical $O(h^2)$ algorithm. All experiments were performed on the Spectra 7 in double precision, and all errors are measured in the maximum norm. The boundary value problems solved were:

(A)
$$\begin{cases} y^{(2)} = [(y^{(1)})^2 + y^2]/2e^x, \\ y(0) - y^{(1)}(0) = 0, \quad y(1) + y^{(1)}(1) = 2e, \end{cases}$$

(B)
$$\begin{cases} y^{(2)} = [e^{2y} + (y^{(1)})^2]/2, \\ y(0) - y^{(1)}(0) = 0, \quad y(1) + y^{(1)}(1) = -\ln 2 - \tfrac{1}{2}, \end{cases}$$

(C)
$$\begin{cases} y^{(2)} = (y + xy^{(1)})/(1 + x), \\ y(0) - 2y^{(1)}(0) = -1, \quad y(1) + 2y^{(1)}(1) = 3e. \end{cases}$$

The solutions of (A) and (C) are $y(x) = e^x$, while (B) has the solution $y(x) = \log(1/(1 + x))$.

Table 4.2 shows several things. First, that the error in our method is order h^4. Second, that for the same number of points (but more work) you get much better answers than the classical order h^2 algorithm. Finally, that the answers in these particular cases seem to be somewhat less accurate than the $O(h^2)$ algorithm plus Richardson extrapolation. In Table 4.2 the notation, say, .13(−4) means $.13 \times 10^{-4}$.

Differential Equation	# Points	Error (1)	Error (2)	Error (3)
(A)	4	.13(−4)	.63(−5)	.11(−1)
	8	.78(−6)	.40(−6)	.28(−2)
	16	.32(−7)		.70(−3)
(B)	4	.34(−3)	.17(−4)	.73(−2)
	8	.56(−4)	.12(−5)	.18(−2)
	16	.37(−5)		.46(−3)
(C)	4	.58(−4)	.93(−5)	.10(−1)
	8	.41(−5)	.60(−6)	.26(−2)
	16	.26(−6)		.66(−3)

TABLE 4.2

RCA
David Sarnoff Research Center
Princeton, New Jersey 08540

1. B. T. ALLEN, "A new method of solving second-order differential equations when the first derivative is present," *Comput. J.*, v. 8, 1965/66, pp. 392–394. MR 32 #6675.

2. A. K. AZIZ & B. E. HUBBARD, "Bounds for the solution of the Sturm-Liouville problem with application to finite difference methods," *J. Soc. Indust. Appl. Math.*, v. 12, 1964, pp. 163–178. MR 29 #2981.

3. L. COLLATZ, *The Numerical Treatment of Differential Equations*, Springer, Berlin, 1966.

4. J. DANIEL & B. SWARTZ, *Extrapolated Collocation for Two-Point Boundary-Value Problems Using Cubic Splines*, Technical Report LA-DC-72-1520, Los Alamos Scientific Laboratory, Los Alamos, 1972.

5. P. HENRICI, *Discrete Variable Methods in Ordinary Differential Equations*, Wiley, New York, 1962. MR 24 #B1772.

6. H. B. KELLER, *Numerical Methods for Two-Point Boundary-Value Problems*, Blaisdell, Waltham, Mass., 1968. MR 37 #6038.

7. H. B. KELLER, "Accurate difference methods for nonlinear two-point boundary value problems," *SIAM J. Numer. Anal.*, v. 11, 1974, pp. 305–320.

8. M. LEES, "Discrete methods for nonlinear two-point boundary value problems," in *Numerical Solution of Partial Differential Equations* (Proc. Sympos. Univ. Maryland, 1965), Academic Press, New York, 1966, pp. 59–72. MR 34 #2196.

9. V. PEREYRA, *High Order Finite Difference Solution of Differential Equations*, Technical Report STAN-CS-73-348, Computer Science Dept., Stanford University, 1973.

10. M. RUSSELL & L. SHAMPINE, "A collocation method for boundary value problems," *Numer. Math.*, v. 19, 1972, pp. 1–28.

11. J. SHOOSMITH, *A Study of Monotone Matrices With an Application to the High-Order, Finite-Difference Solution of a Linear Two-Point Boundary-Value Problem*, Dissertation, Department of Applied Mathematics and Computer Science, University of Virginia, Charlottesville, 1973.

12. R. STEPLEMAN, "High order solution of mildly nonlinear elliptic boundary value problems," *Proceedings of the AICA International Symposium on Computer Methods for Partial Differential Equations*, Lehigh University, 1975.