MATHEMATICS OF COMPUTATION, VOLUME 28, NUMBER 127, JULY 1974, PAGES 743-755

A Comparison of Some Numerical Methods for Two-Point Boundary Value Problems*

By James M. Varah

Dedicated to the memory of G. Immerzeel

Abstract. In this paper we discuss and compare two useful variable mesh schemes for linear second-order two-point boundary value problems: the midpoint rule and collocation with cubic Hermite functions. We analyze the stability of the block-tridiagonal factorization for solving the linear systems, compare the amount of computer time required, and test the methods on some particular numerical problems.

1. Introduction. Recently, there has been a great deal of interest in "global" methods for the numerical solution of two-point boundary value problems. By this, we mean methods which find a solution of the form

$$u^{(N)}(x) = \sum_{j=1}^{N} a_j \phi_j(x), \tag{1.1}$$

where the $\{\phi_j(x)\}$ are piecewise polynomials. Most of the work reported on these methods deals with the theoretical aspects, and it is our purpose here to report on the computational aspects, in particular, to compare the most economical computational versions of these methods with appropriate finite-difference methods. This is not meant to be an all-encompassing survey; rather, a comparison between some computationally feasible schemes on the same problems.

We will fix our attention on the linear second-order equation

$$\begin{aligned} Ly &\equiv y'' + p(x)y' + q(x)y = f(x), \\ \alpha y(a) + \beta y'(a) &= g_0, \qquad \alpha' y(b) + \beta' y'(b) = g_1. \end{aligned} \tag{1.2}$$

Define the mesh partition $a = x_0 < x_1 < \cdots < x_m = b$, and let $h \equiv \max|x_{i+1} - x_i|$. Of the global methods, the most attractive computationally is the *collocation method* since no quadrature sums are required (see [7] for a detailed comparison). In particular, deBoor and Swartz [2] have shown that collocation at the two Gaussian points within each interval (x_i, x_{i+1}) with piecewise Hermite polynomials of degree 3 gives a method of order 4. This defines the coefficients $\{a_j\}$ in (1.1) by the linear system

$$Lu^{(N)}(\xi_{ij}) = f(\xi_{ij}), \qquad i = 1, \ldots, m, j = 1, 2,$$

where ξ_{i1}, ξ_{i2} denote the two Gaussian points in (x_{i-1}, x_i).

Received September 4, 1973

AMS (MOS) subject classifications (1970). Primary 65L10; Secondary 65F05, 65B05.

Key words and phrases. Two-point boundary value problems, collocation, Galerkin method, finite-differences, operation counts, numerical stability.

*Supported by NRC(Canada) grant #A8240.

The corresponding finite-difference method used should have the same versatility and accuracy; thus, we will use the well-known *midpoint rule* popularized by Keller [4]. This is second-order accurate for any mesh partition, and can be extended to a method of order 4 by one extrapolation. This method is normally derived from finite-difference equations, but, in Section 2, we will show that it is equivalent to a particular discrete Galerkin method.

Both these methods are $O(h^4)$ and both solve explicitly for approximations to $\{u(x_i), u'(x_i)\}$, $i = 0, \ldots, m$, if we use a natural basis for the cubic Hermite polynomials. In Section 3, we discuss the form of the linear systems generated by these methods, and their solution via block-tridiagonal factorization. In particular, we show that the factorization is stable *without pivoting*. Then, in Section 4, we compare the amount of work required to set up and solve these systems, and also compare collocation with higher-order piecewise polynomials and more extrapolations of the midpoint rule. Finally, in Section 5, we give some results for the $O(h^4)$ methods on some particular numerical examples.

Both of these methods (collocation and midpoint rule) can be used on more complicated problems, e.g. first-order systems of two-point boundary value problems, and nonlinear problems, but we feel that this would not change the comparison greatly.

2. The Midpoint Rule as a Global Method. The midpoint rule is usually defined for a first-order system of equations (see Keller [4]). In this form, it is easily seen as a collocation method for the system using piecewise linear functions (see Russell [5]). However, for our Eq. (1.2), it can also be viewed as a discrete Galerkin method. Consider (1.2) in selfadjoint form with homogeneous boundary conditions:

$$(r(x)u')' - s(x)u = f,$$

$$u(a) = u(b) = 0.$$

Then the midpoint rule is defined by first forming the equivalent first-order system

$$\begin{aligned} u' &= v/r, \\ v' &= su + f, \end{aligned} \tag{2.1}$$

and then differencing across each mesh interval and averaging, so the approximation is second-order accurate. This gives

$$\begin{aligned} u_j &= u_{j-1} + \frac{h_j}{2r_{j-1/2}}(v_j + v_{j-1}), \\ v_j &= v_{j-1} + \frac{h_j}{2}s_{j-1/2}(u_j + u_{j-1}) + h_j f_{j-1/2} \end{aligned} \tag{2.2}$$

for $j = 1, \ldots, m$, and, of course, $u_0 = u_m = 0$. Here $u_j = u(x_j)$, $h_j = x_j - x_{j-1}$, and $r_{j-1/2} = r(x_j - h_j/2)$. Although this is a finite-difference method, we can consider it a global method if (for example) we let the $\{u_j\}$ be the coefficients of a piecewise linear or "roof function" expansion. (Indeed, for the original formulation (1.2), the $\{u_j\}$ and $\{v_j\}$ together can be considered as the coefficients of a piecewise cubic Hermite polynomial basis; we will use this fact later.) Now, consider the Ritz-

Galerkin method for (2.1). This finds as a solution, $u^{(N)} = \sum c_j \phi_j(x)$, where the coefficients $\{c_j\}$ satisfy

$$A\mathbf{c} = \mathbf{b},$$
$$a_{ij} = -\int_a^b (r\phi_i'\phi_j' + s\phi_i\phi_j)\,dx, \qquad b_i = \int_a^b f\phi_i\,dx. \tag{2.3}$$

Of course, these matrix elements cannot be evaluated exactly; instead, some quadrature rule is applied, giving a discrete Galerkin method.

THEOREM 2.1. *The values $\{u_j\}$ obtained from the midpoint rule* (2.2) *are precisely the coefficients $\{c_j\}$ obtained from the Galerkin method* (2.3) *using the midpoint quadrature rule on piecewise linear functions.*

Proof. First, consider the discrete Galerkin method. Since $\phi_i(x)$ is the normalized piecewise linear function with support in $[x_{i-1}, x_{i+1}]$, it is easy to see that the midpoint quadrature rule gives

$$-a_{jj} = \frac{r_{j+1/2}}{h_{j+1}} + \frac{r_{j-1/2}}{h_j} + \frac{h_{j+1}s_{j+1/2}}{4} + \frac{h_j s_{j-1/2}}{4},$$
$$a_{j,j+1} = \frac{r_{j+1/2}}{h_{j+1}} - \frac{h_{j+1}}{4}s_{j+1/2}, \qquad a_{j-1,j} = \frac{r_{j-1/2}}{h_j} - \frac{h_j}{4}s_{j-1/2}, \tag{2.4}$$
$$b_j = \frac{h_j}{2}f_{j-1/2} + \frac{h_{j+1}}{2}f_{j+1/2}.$$

Note that A is symmetric and tridiagonal.

Now consider the midpoint rule (2.2). To get an expression involving only $\{u_j\}$, use the first of these equations to define v_{j+1}:

$$v_{j+1} = -v_j + \frac{2r_{j+1/2}}{h_{j+1}}(u_{j+1} - u_j)$$
$$= v_{j-1} - \frac{2r_{j-1/2}}{h_j}(u_j - u_{j-1}) + \frac{2r_{j+1/2}}{h_{j+1}}(u_{j+1} - u_j).$$

Now, if we add the second equation of (2.2) with indices $j, j+1$, we obtain

$$\frac{-h_j}{2}s_{j-1/2}u_{j-1} - \left(\frac{h_j}{2}s_{j-1/2} + \frac{h_{j+1}}{2}s_{j+1/2}\right)u_j - \frac{h_{j+1}}{2}s_{j+1/2}u_{j+1} + (v_{j+1} - v_{j-1})$$
$$= h_j f_{j-1/2} + h_{j+1}f_{j+1/2}.$$

Now, substitute for $(v_{j+1} - v_{j-1})$ from above, and we have

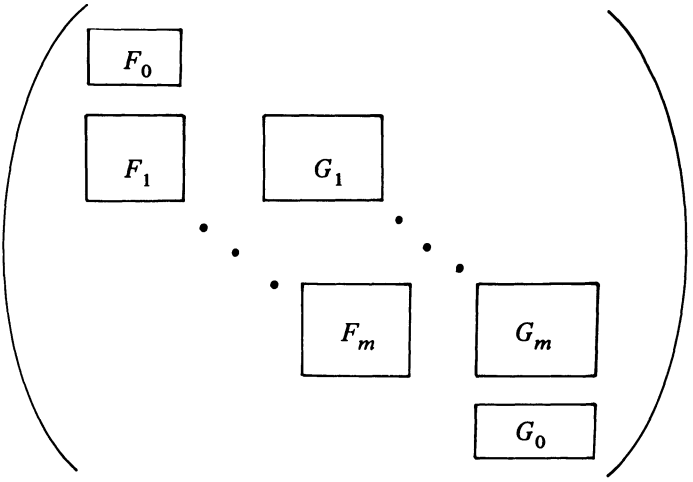
$$\left(\frac{2r_{j-1/2}}{h_j} - \frac{h_j}{2}s_{j-1/2}\right)u_{j-1} - \left(\frac{2r_{j-1/2}}{h_j} + \frac{2r_{j+1/2}}{h_{j+1}} + \frac{h_j}{2}s_{j-1/2} + \frac{h_{j+1}}{2}s_{j+1/2}\right)u_j$$
$$+ \left(\frac{2r_{j+1/2}}{h_{j+1}} - \frac{h_{j+1}}{2}s_{j+1/2}\right)u_{j+1} = h_j f_{j-1/2} + h_{j+1}f_{j+1/2}.$$

But this is precisely (2.4) multiplied by 2. Moreover, the boundary conditions are $u_0 = u_m = 0$ for both schemes, so the $\{u_j\}$ and $\{c_j\}$ are identical. Q.E.D.

Notice that, although the form (2.4) gives a symmetric tridiagonal matrix, it is much more work to set up this linear system than the original midpoint rule (2.2). Perhaps other Galerkin schemes could be similarly economized by rearrangement of the linear system.

3. Solution of the Linear Systems. The collocation method, using piecewise cubic Hermite polynomials with natural basis and collocating at the two Gaussian points in each subinterval, gives a linear system of the form

(3.1)



where all blocks are 2×2 except F_0 and G_0 (1×2). Of course, the midpoint rule gives a system of exactly the same form. An effective way of solving this system is by block-tridiagonal factorization; i.e., we put the matrix into the form

$$\begin{bmatrix} B_0 & C_0 & & \\ A_1 & B_1 & \ddots & \\ & \ddots & \ddots & C_{m-1} \\ & & A_m & B_m \end{bmatrix}$$

where each block is 2×2, and use the iteration

$$\left.\begin{aligned} U_0 &= B_0, \\ L_i &= A_i U_{i-1}^{-1}, \\ U_i &= B_i - L_i C_{i-1}, \end{aligned}\right\} \quad i = 1, \cdots, m, \tag{3.2}$$

to form a block-LU factorization. Then, we solve the linear system by a forward and backward substitution with 2×2 blocks.

Just as with the normal LU decomposition, we must ensure that this block-LU factorization is *stable*, i.e. $\|L_i\|, \|U_i\| \leqq K$. Since we can rewrite (3.2) as

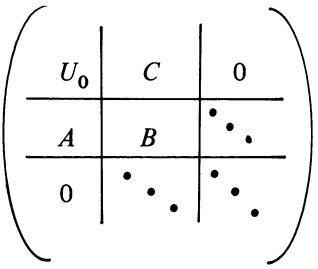
$$\begin{aligned} U_0 &= B_0, \\ U_i &= B_i - A_i U_{i-1}^{-1} C_{i-1}, \qquad i = 1, \ldots, m, \end{aligned}$$

this is equivalent to $\kappa(u_i) = \|U_i\| \|U_i^{-1}\| \leqq K$.

Conditions for stability were given in [9, Theorem 2.2]; unfortunately, these conditions do not hold here, so we must examine the iteration more closely.

Assume now that the problem (1.2) has constant coefficients (i.e., $p(x) = q(x) = 0$) and uniform mesh h. Then both systems have the block-tridiagonal form

(3.3)



with $A = \left(\begin{smallmatrix} a_1 a_2 \\ 0\ 0 \end{smallmatrix}\right)$, $C = \left(\begin{smallmatrix} 0\ 0 \\ c_1 c_2 \end{smallmatrix}\right)$, $B = \left(\begin{smallmatrix} b_1 b_2 \\ b_3 b_4 \end{smallmatrix}\right)$.

LEMMA 3.1. *The block-tridiagonal factorization* (3.2) *with a matrix of form* (3.3) *gives explicitly* $U_i = \left(\begin{smallmatrix} \alpha_i \beta_i \\ b_3 b_4 \end{smallmatrix}\right)$ *with*

$$\alpha_{i+1} = b_1 + c_1(e - 1/r_i), \qquad \beta_{i+1} = b_2 + c_2(e - 1/r_i),$$

where

$$r_i = K^i(r_0 + L/(a-1)) - L/(a-1) \qquad (\text{if } K \neq 1)$$
$$= r_0 + iL \qquad (\text{if } K = 1)$$

and

$$K = \frac{d_1/d_2 - e}{d_4/d_2 + e}, \qquad L = \frac{1}{d_4/d_2 + e}, \qquad r_0 = \frac{1}{e + (a_2\alpha_0 - a_1\beta_0)/(b_4\alpha_0 - b_4\beta_0)},$$

e *is either root of* $e^2 d_2 + e(d_4 - d_1) + d_3 = 0$, *and*

$$d_1 = \begin{vmatrix} b_1 & b_2 \\ b_3 & b_4 \end{vmatrix}, \qquad d_2 = \begin{vmatrix} b_3 & b_4 \\ c_1 & c_2 \end{vmatrix}, \qquad d_3 = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}, \qquad d_4 = \begin{vmatrix} a_1 & a_2 \\ c_1 & c_2 \end{vmatrix}.$$

Proof. The ith step of iteration (3.2) applied to (3.3) is $U_{i+1} = B - AU_i^{-1}C$ giving $U_{i+1} = \left(\begin{smallmatrix} \alpha_{i+1} \beta_{i+1} \\ b_3 b_4 \end{smallmatrix}\right)$, with $\alpha_{i+1} = b_1 - c_1\delta_i/\gamma_i$, $\beta_{i+1} = b_2 - c_2\delta_i/\gamma_i$, where $\delta_i = a_2\alpha_i - a_1\beta_i$, $\gamma_i = b_4\alpha_i - b_3\beta_i$. Of course, this is a nonlinear iteration for α_i, β_i, but we can express the iteration in terms of δ_i, γ_i:

$$\delta_{i+1} = -d_3 + d_4\delta_i/\gamma_i, \quad : \quad \gamma_{i+1} = d_1 + d_2\delta_i/\gamma_i.$$

This is still nonlinear, but if we define the new variable $r_i = 1/(e + \delta_i/\gamma_i)$, for e a constant, we obtain

$$r_{i+1} = \frac{d_1 + d_2\delta_i/\gamma_i}{ed_1 - d_3 + (ed_2 + d_4)\delta_i/\gamma_i}.$$

Now, if e is defined as in the statement of the lemma,

$$r_{i+1} = \frac{d_1 + d_2\delta_i/\gamma_i}{(ed_2 + d_4)(e + \delta_i/\gamma_i)} = \frac{d_1 + d_2(e + \delta_i/\gamma_i) - d_2 e}{(ed_2 + d_4)(e + \delta_i/\gamma_i)} = Kr_i + L,$$

with K, L as given above. This is linear, and hence

$$\begin{aligned} r_i &= K^i(r_0 + L/(a-1)) - L/(a-1), && \text{if } K \neq 1, \\ &= r_0 + iL, && \text{if } K = 1. \end{aligned}$$

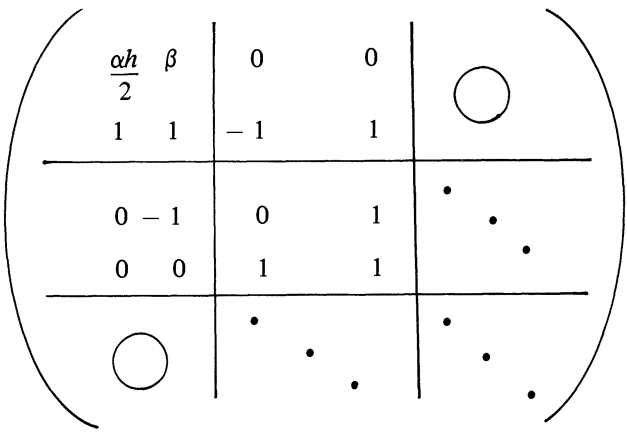
The expressions for α_{i+1}, β_{i+1} then follow easily. Q.E.D.

Now, to show stability of the block-tridiagonal factorization, we need only show that the $\{\alpha_i\}, \{\beta_i\}$ remain bounded and $\alpha_i/\beta_i \neq b_3/b_4$ (i.e., det $(U_i) \neq 0$). First, consider the midpoint rule. For our Eq. (1.2), the midpoint rule is

$$\begin{aligned} u_j - u_{j-1} &= h_j(v_j + v_{j-1})/2, \\ v_j - v_{j-1} &= -h_j p(x_{j-1/2})(v_j + v_{j-1})/2 - h_j q(x_{j-1/2})(u_j + u_{j-1})/2 \\ &\quad + h_j f(x_{j-1/2}) \end{aligned} \tag{3.4}$$

which for $p = q = 0$ and constant h gives the linear system

(3.5)



after multiplying the even rows by $-2/h$ and the odd columns by $h/2$.

COROLLARY 3.2. *Consider the midpoint rule applied to the constant coefficient problem*

$$\begin{aligned} y'' &= f(x), \\ \alpha y(a) + \beta y'(a) &= g_0, \\ \alpha' y(b) + \beta' y'(b) &= g_1, \end{aligned}$$

with constant mesh size h.

(i) *If* $\beta/\alpha \leqq 0$, *the block-tridiagonal factorization* (3.2) *is stable for all h; in fact,* $-1 \leqq \alpha_i \leqq 0$, $1 \leqq \beta_i \leqq 2$.

(ii) *If* $\beta/\alpha > 0$, *the factorization is unstable for certain values of h.*

Proof. Lemma 3.1 applied to (3.5) gives

$$r_i = -1 + 2\beta/h\alpha - 2i,$$

$$\alpha_{i+1} = 1/r_i, \qquad \beta_{i+1} = 1 - 1/r_i.$$

If $\beta/\alpha \leqq 0$, r_i stays bounded away from zero for all h, so the $\{\alpha_i\}$ and $\{\beta_i\}$ remain bounded as above. In this case, this ensures stability because $-\det(U_i) = \beta_i - \alpha_i \geqq 1$, so $\|U_i^{-1}\|_\infty \leqq \|U_i\|_\infty \leqq 3$. Note also that if $\alpha = 0$, then $\alpha_i = 0$, $\beta_i = 1$ for all i. However, if $\beta/\alpha > 0$, we have $r_i = 0$ if $h = \beta/\alpha(i + 1/2)$ for some i, in which case the factorization is unstable and, in fact, breaks down. Q.E.D.

Now consider the collocation scheme on the same constant coefficient problem. Recall the mesh $a = x_0 < x_1 < \cdots < x_m = b$, and recall that we must collocate at two points ξ_{i1}, ξ_{i2} in each interval (x_{i-1}, x_i). Let the points be placed symmetrically in the interval at $((x_{i-1} + x_i)/2) \pm \rho((x_i - x_{i-1})/2)$, with $0 < \rho < 1$ (for Gaussian points, $\rho = 1/\sqrt{3}$). Then the linear system to solve is

(3.6)

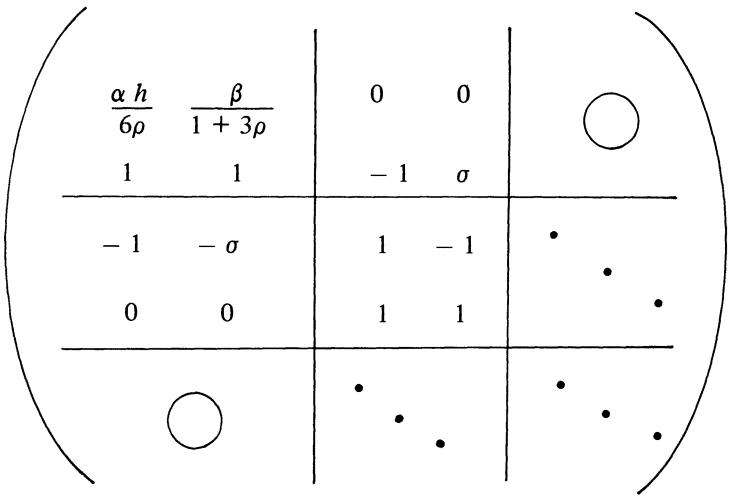
$$\left(\begin{array}{cc|cc|cc}
\alpha & \beta & 0 & 0 & & \\
\phi_0''(\xi_{11}) & \phi_1''(\xi_{11}) & \phi_2''(\xi_{11}) & \phi_3''(\xi_{11}) & & \\
\hline
\phi_0''(\xi_{12}) & \phi_1''(\xi_{12}) & \phi_2''(\xi_{12}) & \phi_3''(\xi_{12}) & 0 & 0 \\
0 & 0 & \phi_2''(\xi_{21}) & \phi_3''(\xi_{21}) & \phi_4''(\xi_{21}) & \phi_5''(\xi_{21}) \\
\hline
 & & \cdot & & \cdot & \\
 & & & \cdot & & \cdot \\
 & & & & & \cdot
\end{array}\right)$$

The $\{\phi_i(x)\}$ are the usual natural basis for cubic Hermite polynomials (see Schultz [8, p. 27]). Thus, ϕ_{2i}, ϕ_{2i+1} have support (x_{i-1}, x_{i+1}) and $\phi_{2i}(x_i) = 1$, $\phi_{2i+1}'(x_i) = 1$. If we scale successive columns by multiplying by h^2, h respectively, the matrix (3.6) becomes

$$\left(\begin{array}{cc|cc|cc}
\alpha h^2 & \beta h & 0 & 0 & & \\
-6\rho & -1-3\rho & 6\rho & 1-3\rho & & \\
\hline
6\rho & 3\rho - 1 & -6\rho & 1+3\rho & & \\
0 & 0 & -6\rho & -1-3\rho & & \\
\hline
 & & \cdot & & \cdot & \\
 & & & \cdot & & \cdot \\
 & & & & & \cdot
\end{array}\right)$$

Moreover, it simplifies the stability analysis if we divide successive columns by (-6ρ), $-1 - 3\rho$; then the collocation matrix becomes

(3.7)



where $\sigma = (3\rho - 1)/(3\rho + 1)$. Since $0 < \rho < 1$, $-1 < \sigma < \frac{1}{2}$ and for the Gaussian points, $\sigma = 2 - \sqrt{3}$.

COROLLARY 3.3. *Consider the cubic Hermite collocation scheme applied to the constant coefficient problem*

$$y'' = f(x),$$
$$\alpha y(a) + \beta y'(a) = g_0,$$
$$\alpha' y(b) + \beta' y'(b) = g_1,$$

with constant mesh size h, with collocation at symmetric points in each subinterval.

(i) *If* $\beta/\alpha \leqq 0$, *the block-tridiagonal factorization is stable for all h; in fact,*

$$0 \leqq \alpha_i \leqq 1 - \sigma,$$
$$\sigma - 1 \leqq \beta_i \leqq \sigma - 1 - \sigma/(1 - \sigma), \qquad \text{if } \sigma \leqq 0,$$
$$\sigma - 1 - \sigma/(1 - \sigma) \leqq \beta_i \leqq \sigma - 1, \qquad \text{if } \sigma \geqq 0.$$

(ii) *If* $\beta/\alpha > 0$, *the factorization is unstable for certain values of h.*

Proof. Applying Lemma 3.1 to (3.7), we obtain

$$r_i = \frac{1 - 6\rho\beta/h\alpha(1 + 3\rho) + i(1 + \sigma)}{1 - \sigma},$$
$$\alpha_{i+1} = 1/r_i, \qquad \beta_{i+1} = \sigma - 1 - \sigma/r_i.$$

Recall $0 < \rho < 1$, $-1 < \sigma < \frac{1}{2}$. If $\beta/\alpha \leqq 0$, then $1/(1 - \sigma) \leqq r_i < \infty$, giving the bounds on α_i, β_i as above. Again, we need only ensure the boundedness of $\{\alpha_i\}$, $\{\beta_i\}$. Also, if $\alpha = 0$, $\alpha_i = 0$, $\beta_i = \sigma - 1$ for all i. However, if $\beta/\alpha > 0$, $r_i = 0$ possibly for certain values of h and i in which case the factorization is unstable. Q.E.D.

Of course, the above analysis only shows that the block-tridiagonal factorization is stable for constant coefficients and constant h; however, a similar analysis could probably be done for variable h, and the bounds may involve the mesh ratios. Also,

for variable coefficients, the matrix elements only change by $O(h)$. In any case, the stability can be monitored during the decomposition and if the intermediate $\|L_i\|$, $\|U_i\|$ become too large, one can shift to a partial pivoting routine.

4. Work Estimates. Now let us compare the work required (we measure this in units of M, the average time for a multiplication or division) for the two $O(h^4)$ schemes. For our problem, (1.2) and the given mesh $a = x_0 < x_1 < \cdots < x_m = b$, both schemes give a matrix of the form (3.1). The collocation scheme has the general matrix element

$$L\phi_j(\xi_i) = \phi_j''(\xi_i) + p(\xi_i)\,\phi_j'(\xi_i) + q(\xi_i)\phi_j(\xi_i) \tag{4.1}$$

where the $\{\phi_j\}$ are the natural basis for cubic Hermite polynomials. Recall from Section 3 that, for stability of the matrix factorization, we must scale by dividing successive columns of the matrix by h_j^2, h_j. Now, we can assume that the constants used in the evaluation of $\phi_j(\xi_i)$, $\phi_j'(\xi_i)$, $\phi_j''(\xi_i)$ are done beforehand, so each element (4.1) takes $4M$ (i.e., 4 multiplications/divisions). Also let E denote the time required for evaluating $p(x)$, $q(x)$, and $f(x)$ at some point. Then the total setup time for the collocation matrix is $(4M)(8m) + (2m)E$. Moreover, the matrix elements have no fixed value; this will make a difference later when we discuss the work required to solve the system.

For the midpoint rule, the matrix elements are much easier to evaluate; if we scale as in Section 3 (i.e., multiply row $2j$ by $-2/h_j$, column $2j-1$ by $h_j/2$), we have, using the notation of (3.1),

$$F_j = \begin{pmatrix} 1 & 1 \\ \dfrac{h_j^2}{4} q(x_{j-1/2}) & -1 + \dfrac{h_j}{2} p(x_{j-1/2}) \end{pmatrix},$$

$$G_j = \begin{pmatrix} -h_{j+1}/h_j & 1 \\ \dfrac{h_j h_{j+1}}{4} q(x_{j-1/2}) & 1 + \dfrac{h_j}{2} p(x_{j-1/2}) \end{pmatrix}, \tag{4.2}$$

$$F_0 = \begin{pmatrix} \alpha h_1 \\ 2 \end{pmatrix} \beta, \qquad G_0 = \begin{pmatrix} \alpha' h_m \\ 2 \end{pmatrix} \beta'.$$

If we store the terms involving only the mesh sizes, the total setup time is $(3M + E)m$.

Now, consider solving these systems by block-tridiagonal factorization (3.2). The relevant operations are as follows, where we assume arbitrary elements for the collocation matrix but use the ones appearing in the midpoint matrix (4.2):

	collocation	midpoint rule
solve $L_i(U_{i-1}) = A_i$	$6M$	$2M$
form $U_i = B_i - L_i C_{i-1}$	$2M$	$1M$
solve $L_i y_{i-1} + y_i = f_i$	$2M$	$2M$
solve $U_i x_i + C_i x_{i+1} = y_i$	$6M$	$3M$
total for m blocks	$16Mm$	$8Mm$

Of course, with the midpoint rule, we must extrapolate once to get an $O(h^4)$ method; this involves placing new points at the midpoints of each subinterval, solving on the expanded mesh by the midpoint rule, and then forming the right linear combination $(-\frac{1}{3}, \frac{4}{3})$ of these two solutions on the original mesh. This, of course, involves setting up and solving a second system of exactly double the size, so the time for the full $O(h^4)$ method is $(11M + E)3m + 2m$ (for the extrapolation), or $(35M + 3E)m$. For collocation, the total is $(48M + 2E)m$. This easily gives

THEOREM 4.1. *The midpoint rule with extrapolation is faster than collocation with cubic Hermite polynomials on the same mesh for problem* (1.2), *provided* $E/M < 13$, *where* E/M *is the number of equivalent multiplications required to evaluate* $p(x)$, $q(x)$, *and* $f(x)$.

Notice that because three evaluations of the coefficient functions in each subinterval are required for the midpoint rule, and only two for collocation, the collocation scheme is cheaper except for quite simple functions.

Note. Recently, G. Immerzeel has observed that if the trapezoidal method of averaging is used instead of the midpoint rule (see [4, p. 15]), then one extrapolation again gives a fourth-order method, but the number of evaluations is reduced to two per interval, the same as collocation, and the number of multiplications is not increased. So this method (which has almost identical error properties as the midpoint rule) is *always* cheaper than collocation.

Another appropriate comparison of higher-order methods involves collocation with higher degree piecewise polynomials and more extrapolations of the midpoint rule. DeBoor and Swartz [2] show that using piecewise polynomials of degree $2n - 1$, which are only $C^{(1)}$ at the mesh points, and collocating at the $2n - 2$ Gauss points in each subinterval gives a method of order $4n - 4$ *at the mesh points*. Even though such a method is only practical for small values of n, we can compare the work required with the correspondingly accurate method obtained by extrapolating the midpoint rule.

First, consider the midpoint rule: for a method of order $4n - 4$ at the basic mesh points, we need to extrapolate $2n - 3$ times. Normally (see Keller [4]), this is done by subdividing each h_i into $h_i/2$, $h_i/4$, $h_i/8$, ... However, this quickly involves too many points. Here, we propose the sequence $h_i/2$, $h_i/3$, $h_i/4$, ... This sequence has not been used even for extrapolation with initial value problems because of the possible unlimited growth of the roundoff error (see Gragg [3]). However, experiments of G. Immerzeel have shown that this roundoff error growth is not large in the range of practical computation (10 or 12 extrapolations). For this sequence, we must set up and solve midpoint rule systems like (4.2) with $m, 2m, 3m, \ldots, pm$ blocks, where we define $p = 2n - 2$. Proceeding as with the $O(h^4)$ method, we see the total time required is

$$\left[11\frac{p(p+1)}{2} + 2\frac{p(p-1)}{2}\right]mM + \frac{p(p+1)}{2}mE. \tag{4.3}$$

The collocation method, on the other hand, requires only one solution of a larger system; we have $(2n - 2)$ basis functions at each of the $(m - 1)$ interior nodes and n at each endpoint, giving # unknowns $= 2n + (m - 1)(2n - 2)$. Similarly, there are $(2n - 2)$ collocation equations for each of the m intervals, which together with the boundary conditions give $m(2n - 2) + 2$ equations. Because the

basis functions have support over only two intervals, the linear system is again of the form (3.1) with F_i and G_i $p \times p$, except F_0 and G_0 which are $1 \times n$ and F_1 and G_m ($p \times n$). Now, the block-tridiagonal factorization has the first and last diagonal blocks $n \times n$, and the rest $p \times p$.

We again assume the basis functions have been evaluated beforehand, so the total setup time for each matrix element is $4M$. Also, we assume the block-tridiagonal factorization is stable; then we find (as in [9, p. 867] with $q = p/2$) the solution time is $(13p^3/12 + 2p^2 - p/3)mM$. Here, unlike [9], we have included the lower-order terms since we are interested in small values of p. Thus, the total time for collocation is

$$\left(\frac{13}{12}p^3 + 10p^2 - \frac{p}{3}\right)mM + pmE. \tag{4.4}$$

Comparing (4.3) and (4.4), we obtain easily

THEOREM 4.2. *Extrapolation with the midpoint rule is faster than the above described collocation procedure of order* $4n - 4$ *when* $E/M < 13n/3 + 29/6$.

Notice that again the outcome depends on how complex the functions $p(x)$, $q(x)$, $f(x)$ are. The more accuracy desired however, the more attractive extrapolation appears. Also, the extrapolation procedure is *much* easier to program.

5. Numerical Examples. Here, we will report the results of the two $O(h^4)$ schemes on some test problems. It is not our purpose to obtain extremely high accuracy even though this is certainly possible; rather, we are interested in reasonable accuracy (3 or 4 significant figures) with a small number of mesh points. Each example is of the form (1.2) on the mesh $a = x_0 < x_1 < \cdots < x_m = b$. For each example, we give the interior mesh points used $(x_1, \ldots, x_{m-1})$. Since both collocation with cubic Hermite polynomials (denoted $CH3$) and the midpoint rule with one extrapolation (denoted $M + E$) provide approximations to the solution and its first derivative at the mesh points, it is natural to use the uniquely defined piecewise cubic Hermite polynomial as the global solution and then measure the maximum error between this function and the true solution over the whole interval. We give this error for $CH3$ and $M + E$, and also for the cubic Hermite interpolate of the exact solution at the same mesh points (denoted INT). We give this last error for reference: we can hardly expect the approximate solution to give results better than interpolating the exact solution! We could also give first derivative errors, but, in all cases, they were no larger than a factor of 10.

Example 1.

$$y'' + (2\gamma x)y' + 2\gamma y = 0, \qquad \text{exact solution.}$$

$$y(0) = 1, \quad y(1) = e^{-\gamma}, \quad y = e^{-\gamma x^2}.$$

γ	m	interior mesh points	INT	$CH3$	$M + E$
10	5	.2, .4, .6, .8	.0029	.0036	.0025
10	5	.137, .302, .457, .703	.0008	.0028	.0027
20	5	.2, .4, .6, .8	.0063	.0102	.0054
20	5	.107, .234, .327, .561	.0014	.0049	.0051

Example 2.

$$y'' + (3 \cot x + 2 \tan x)y' + \gamma y = 0,$$

$$y(a) = g_0, \quad y(b) = g_1 \qquad (0 < a < b < \pi/2).$$

This has the Fourier series solution $y(x) = \sum_0^\infty a_k \cos^k x$, with

$$a_{k+2} = ((k(k+1) - \gamma)/(k-1)(k+2))a_k,$$

$a_1 = 0$, and a_0, a_3 determined by the boundary conditions. (For $\gamma = 2$ and the proper boundary conditions, $y(x) = \csc^2 x$.) This is almost the same as Problem 2 of Russell and Shampine [6] except they measure x in degrees not radians. With $a = 30°$, $b = 60°$, $g_0 = 0, g_1 = 5$, $\gamma = 0.7$, their solution has a sharp rise near $x = a$ (this can be predicted if their problem is converted to radians; then there is a small constant $(\pi/180)^2$ in the y'' term).

With x in radians and the same boundary conditions, the solution rises slowly from 0 at $x = \pi/6$ to 5 at $x = \pi/3$. The errors are as follows:

γ	m	interior mesh points	*INT*	*CH3*	*M + E*
0.7	5	$6\pi/30, 7\pi/30, 8\pi/30, 9\pi/30$	.0026	.0024	.0023

A harder example is the same problem from $a = \pi/18$ (10°) to $b = 8\pi/18$ (80°). This has a sharp rise near $x = a$.

γ	m	interior mesh points	*INT*	*CH3*	*M + E*
0.7	5	13°, 17°, 27°, 50°	.021	.016	.015

Example 3.

$$\varepsilon y'' - (2 - x^2)y = -1, \qquad y'(0) = y(1) = 0.$$

This is the right half of the singular perturbation problem discussed by Carrier [1] and also used in [6].

ε	m	interior mesh points	*INT*	*CH3*	*M + E*
.01	5	.3, .6, .8, .9	.002	.0024	.0023
.0001	5	.4, .85, .96, .99	.020	.015	.21
.0001	7	.3, 6, .85, .95, .97, .99	.006	.008	.042

Notice that, for a small ε, the midpoint rule gives inferior results; this is because the first derivative is poorly approximated, particularly with only a few mesh points.

Department of Computer Science
University of British Columbia
Vancouver, British Columbia, Canada

1. G. F. CARRIER, "Singular perturbation theory and geophysics," *SIAM Rev.*, v. 12, 1970, pp. 175-193.

2. G. DE BOOR & B. SWARTZ, "Collocation at Gaussian points," *SIAM J. Numer. Anal.*, v.10, 1973, pp. 582-606.

W. B. GRAGG, "On extrapolation algorithms for ordinary initial value problems," *J. Soc. Indust. Appl. Math. Ser. B Numer. Anal.*, v. 2, 1965, pp. 384-403. MR **34** #2191.

4. H. B. KELLER, "Accurate difference methods for linear ordinary differential systems subject to linear constraints," *SIAM J. Numer. Anal.*, v. 6, 1969, pp. 8-30. MR **40** #6776.

5. R. D. RUSSELL, "Collocation for systems of boundary value problems," *SIAM J. Numer. Anal.* (Submitted.)

6. R. D. RUSSELL & L. F. SHAMPINE, "A collocation method for boundary value problems," *Numer. Math.*,v. 19, 1972, pp. 1-28. MR **46** #4737.

7. R. D. RUSSELL & J. M. VARAH, "Equivalences in global methods for two-point boundary value problems," (In preparation.)

8. M. H. SCHULTZ, *Spline Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1973.

9. J. M. VARAH, "On the solution of block-tridiagonal systems arising from certain finite-difference equations," *Math. Comp.*, v. 26, 1972, pp. 859-868.