Minimising Truncation Error in Finite Difference Approximations to Ordinary Differential Equations

By M. R. Osborne

Abstract. It is shown that the error in setting up a class of finite difference approximations is of two kinds: a quadrature error and an interpolation error. In many applications the quadrature error is dominant, and it is possible to take steps to reduce it. In the concluding section an attempt is made to answer the question of how to find a finite difference formula which is best in the sense of minimising the work which has to be done to obtain an answer to within a specified tolerance.

1. Introduction. This paper has two main aims:

(i) to provide general schemes for generating difference approximations which make best use of available information in the sense of minimising truncation error, and

(ii) to provide a criterion for comparing the utility of particular difference approximations.

Consideration is restricted to finite difference approximations to ordinary linear differential equations, and to difference approximations which require only values of the coefficients in the differential equation for their construction. Difference approximations are called *classical* if they are satisfied exactly whenever the solution to the differential equation is a polynomial of sufficiently low degree.

The first aim was motivated by the recent appearance of several papers in which Gaussian-type quadrature formulae were used to reduce the truncation error in finite difference approximations to special differential equations (see for example [1]). The author has proposed [2] a scheme for generating classical finite difference approximations, and the question whether Gaussian-type quadrature formulae could be used naturally suggested itself. The answer is developed in Sections 2, 3 and 4. First, a slight generalisation of the author's scheme and a brief resume of the error analysis are given. It is shown that the error falls into two parts called the quadrature error and the interpolation error, and that the quadrature error is dominant. In Section 3 the term quadrature error is justified by deriving an explicit form for the appropriate quadrature. This turns out to be an integral containing a positive weight function. This suggests Gaussian quadrature, and its use is exemplified in Section 4.

An interesting feature of the author's scheme is that it has a natural generalisation which permits the construction of a range of nonclassical approximations. Particular examples of these have been produced before by several authors—for example, by Hersch [4] and Rose [5] who effectively rediscovers Hersch's work. This generalisation is discussed in Section 5.

In the final section a basis for comparing the utility of particular difference schemes is suggested. This is applied to discuss several of the difference equations

Received April 7, 1966. Revised September 9, 1966.

constructed in previous sections. The conclusion to be drawn would seem to be that the law of diminishing returns applies to the search for difference approximations of high accuracy, and that comparatively simple formulae are most useful.

A characteristic feature of the references quoted above is that they restrict attention to difference approximations having the same order as the differential equation to which they approximate. Such approximations have proved popular in particular for the numerical solution of boundary-value problems. Here finite difference approximations of this type only are considered, but this should not be thought of as implying any restriction on the methods used.

2. The Scheme for Difference Approximation. In this section an outline is given of a technique for constructing finite-difference approximations to the differential equation

$$L(y) = \frac{d^n y}{dx^n} + \sum_{i=0}^{n-1} a_i(x) \frac{d^i y}{dx^i} = f(x). \tag{2.1}$$

A more detailed account can be found in [2]. The approximation is classical as it is found by first fitting an interpolation polynomial to y, and then finding a difference equation satisfied by the interpolation polynomial.

Let S_1 be the set of points $x_1, x_2, \cdots, x_{n+1}$ where $x_p < x_q$ if $p < q$, and $x_{n+1} - x_1 = nh$. The quantity h defines the scale of the difference mesh. Also let S_2 be the set of points $\xi_1, \xi_2, \cdots, \xi_m$ where S_1 and S_2 need not be disjoint. Let z be the interpolation polynomial to y which satisfies the conditions

(i) $\Delta(1, 2, \cdots, r+1)z = \Delta(1, 2, \cdots, r+1)y$, $r = 0, 1, \cdots, n-1$.

(Here $\Delta(1, 2, \cdots, p)$ is the divided difference operator defined on the points of S_1 whose suffices are indicated. When $p = 1$ the corresponding operator is the identity.)

(ii) $L(z)(\xi_i) = f(\xi_i)$, $i = 1, 2, \cdots, ns + 1$.

Provided only that h is small enough, z can be found by

(a) fitting a polynomial to $z^{(n)}(\xi_i)$, $i = 1, 2, \cdots, m$ (regarding them as formal parameters) and integrating n times,

(b) finding the constants of integration using the conditions (i), and

(c) using the conditions (ii) to determine actual values for the formal parameters $z^{(n)}(\xi_i)$.

To carry out stage (c) note that for every $p = 0, 1, \cdots, n-1$, and $i = 1, 2, \cdots, m$, $z^{(p)}(\xi_i)$ is expressible as a linear combination of the values of z on S_1 and $z^{(n)}$ on S_2. Let $\mathbf{w}(z)$ be the vector whose components are the values of z on S_1, then the vector $\mathbf{z}^{(p)}$ whose components are the values of $z^{(p)}$ on S_2 permits a representation having the form

$$\mathbf{z}^{(p)} = B_p \mathbf{w}(z) + C_p \mathbf{z}^{(n)} \tag{2.2}$$

where B_p has (m) rows and $(n+1)$ columns, and C_p has (m) rows and (m) columns, $p = 0, 1, \cdots, n-1$. Note that the components of C_p are obtained by integrating the interpolation polynomial for $z^{(n)}$ so that they are $O(h^{n-p})$ as $h \to 0$. The conditions (ii) can be written in matrix form

$$\mathbf{z}^{(n)} = -\sum_{i=0}^{n-1} A_i \mathbf{z}^{(i)} + \mathbf{f} \tag{2.3}$$

where the A_i are diagonal matrices. Combining Eqs. (2.2) and (2.3) gives

$$\left(I + \sum_{i=0}^{n-1} A_i\, C_i\right)\mathbf{z}^{(n)} = -\left(\sum_{i=0}^{n-1} A_i\, B_i\right)\mathbf{w}(z) + \mathbf{f} \tag{2.4}$$

and this equation determines $\mathbf{z}^{(n)}$ provided h is small enough as the components of the C_i tend to zero as $h \to 0$ (noted above).

The constants of integration appear in z only in terms of degree $\leqq n - 1$. Therefore

$$\Delta(1, 2, \cdots, n + 1)z = \sum_{i=1}^{m} v_i\, z^{(n)}(\xi_i) = \mathbf{v}^T\, \mathbf{z}^{(n)} \tag{2.5}$$

where the v_i depend only on the points of S_1 and S_2 and satisfy $\sum_{i=1}^{m} v_i = 1/n!$, whence

$$\Delta(1, 2, \cdots, n + 1)z = -\mathbf{v}^T\left(I + \sum_{i=0}^{n-1} A_i\, C_i\right)^{-1}\left\{\left(\sum_{i=0}^{n-1} A_i\, B_i\right)\mathbf{w}(z) - \mathbf{f}\right\}. \tag{2.6}$$

Eq. (2.6) is a difference equation which is satisfied exactly by the interpolation polynomial z. It is the desired finite-difference approximation to Eq. (2.1).

To examine the error in Eq. (2.6) write $y = t + R$ where t consists of the first $ns + n + 1$ terms of the Taylor series for y, and R is the remainder. Making use of the fact that t satisfies Eq. (2.5) it follows after some manipulation that

$$\begin{aligned}\Delta(1, 2, \cdots, n + 1)y + \mathbf{v}^T\left(I + \sum_{i=0}^{n-1} A_i\, C_i\right)^{-1}&\left\{\left(\sum_{i=0}^{n-1} A_i\, B_i\right)\mathbf{w}(y) - \mathbf{f}\right\} \\ = \Delta(1, 2, \cdots, n + 1)R - \mathbf{v}^T\mathbf{R}^{(n)} &- \mathbf{v}^T\left(I + \sum_{i=0}^{n-1} A_i\, C_i\right)^{-1} \\ &\cdot\left\{\sum_{i=0}^{n-1} A_i(\mathbf{R}^{(i)} - B_i\,\mathbf{w}(R) - C_i\,\mathbf{R}^{(n)})\right\}.\end{aligned} \tag{2.7}$$

This equation shows that the error in using Eq. (2.6) as an approximate difference equation is a linear combination of the errors in the Eqs. (2.2) and (2.5). The errors introduced by Eq. (2.2) are called the interpolation errors. In general they will be $O(h^{m+n-j})$, $j = 0, 1, \cdots, n - 1$. The error introduced by Eq. (2.5) is called the quadrature error. The significance of this term will be made clear in the next section. It can be expected to be $O(h^m)$ so that usually it will dominate the interpolation errors.

It would appear that little can be done about reducing the interpolation errors, but the actual contribution of these terms in any actual case depends on the nonzero coefficients in Eq. (2.1). For example if a_{n-1} is nonzero then the interpolation error contains terms $O(h^{m+1})$, but if only a_0 is nonzero then the interpolation errors are $O(h^{m+n})$. A difference equation in which the error has the same order of magnitude as the interpolation error will be called *optimum*.

There is little scope for optimisation in the general case, and specifications of S_1 and S_2, such that the quadrature error is $O(h^{m+1})$ provided ns is even, are given in [2]. In this case the quadrature and interpolation errors are of the same order of magnitude so that these formulae are already optimum. *To obtain difference*

formulae of substantially higher accuracy it is necessary to specialise the differential equation.

3. The Quadrature Formula. In this section it is shown that the quadrature error is identical with the error in the numerical evaluation of an integral representation of the divided difference defined on the points of S_1. This integral representation has the form

$$\Delta(1, 2, \cdots, n+1)y = \int_{-\infty}^{\infty} Ty^{(n)}\, dx \tag{3.1}$$

where T is defined by

(i) $T = 0$, $x < x_1$ and $x > x_{n+1}$,

(ii) $T, T^{(1)}, \cdots, T^{(n-2)}$ continuous on S_1,

(iii) $T^{(n)} = 0$ except on S_1, and

(iv) an appropriate scaling condition.

Eq. (3.1) is readily verified. First the right-hand side obviously vanishes when y is a polynomial of degree $< n$. Second, by Green's theorem,

$$\int_{-\infty}^{\infty} Ty^{(n)}\, dt = \int_{-\infty}^{\infty} yT^{(n)}\, dx = \sum_{i=1}^{n+1} \lambda_i\, y(x_i)$$

as $T^{(n)}$ vanishes except at the points x_i where T has (possibly) discontinuous $(n-1)$st derivative so that $T^{(n)}$ is expressible as a linear combination of δ functions with peaks on S_1.

From the conditions (i)–(iii) specifying T it follows that (for certain constants K_i to be determined)

$$\begin{aligned} T &= K_1(x - x_1)^{n-1}, \qquad x_1 \leqq x < x_2, \\ &= K_1(x - x_1)^{n-1} + K_2(x - x_2)^{n-1}, \qquad x_2 \leqq x < x_3, \\ &= \sum_{i=1}^{n+1} K_i(x - x_i)^{n-1} \equiv 0, \qquad x \geqq x_{n+1}. \end{aligned}$$

If coefficients of powers of x are equated to zero in this last equation there results

$$\sum_{i=1}^{n+1} K_i x_i^r = 0, \qquad r = 0, 1, \cdots, n-1, \tag{3.2}$$

which shows that the K_i are proportional to the coefficients in the divided-difference operator defined on the points of S_1. Therefore

$$K_i = \gamma / \prod_{s=1;\, s \neq i}^{n+1} (x_i - x_s) \tag{3.3}$$

where γ is a scale factor to be determined.

To calculate γ put $y = x^n$ in Eq. (3.1). Then

$$\Delta(1, \cdots, n+1)x^n = n! \int_{x_1}^{x_{n+1}} T\, dx$$

$$\therefore \int_{x_1}^{x_{n+1}} T\, dx = \frac{\Delta(1, \cdots, n+1)x^n}{n!}.$$

But, by direct calculation,

$$\int_{x_1}^{x_{n+1}} T\,dt = \sum_{i=1}^{n+1} K_i \int_{x_i}^{x_{n+1}} (x - x_i)^{n-1}\,dx$$

$$= \sum_{i=1}^{n+1} \frac{K_i}{n} (x_{n+1} - x_i)^n$$

$$= \frac{\gamma}{n} \sum_{i=1}^{n+1} \left((x_{n+1} - x_i)^n / \prod_{s \neq i} (x_i - x_s) \right)$$

$$= \frac{\gamma}{n} \Delta(1, 2, \cdots, n+1)(x_{n+1} - x)^n$$

$$= \frac{\gamma}{n} (-1)^n \Delta(1, 2, \cdots, n+1) x^n$$

whence

$$\gamma = \frac{(-1)^n}{(n-1)!}. \tag{3.4}$$

Example. In the case $n = 2$, $x_1 = -h$, $x_2 = 0$, $x_3 = h$,

$$\Delta(1, 2, 3)y = \frac{h^{-2}}{2} \delta^2 y(0) = \int_{-h}^{h} T y^{(2)}\,dx \tag{3.5}$$

where

$$T = (h + x)/2h^2, \qquad -h \leqq x < 0,$$
$$= (h - x)/2h^2, \qquad 0 \leqq x < h.$$

If the interpolation polynomial z is inserted in Eq. (3.1) and the integrations carried out, there is obtained the result

$$\Delta(1, 2, \cdots, n+1)z = \int_{-\infty}^{\infty} T z^{(n)}\,dt = \sum_{j=1}^{m} v_j\, z^{(n)}(\xi_j). \tag{3.6}$$

The numbers v_j are identical with those in Eq. (2.5). This follows at once because the $z^{(n)}(\xi_j)$ can be chosen arbitrarily. Thus Eq. (2.5) can be interpreted as a quadrature formula for the integral in Eq. (3.1).

The use of Gaussian quadrature with T as weight function to improve the accuracy of Eq. (2.5) now suggests itself. For this it is sufficient that T be positive, and this will now be demonstrated. (I am indebted to the referee for this proof.)

Assume $T < 0$ for $t_1 \leqq x \leqq t_2$.

Let $K = \max |\, T \,|$, and $y^{(n)} = \epsilon/K$, $x_1 \leqq x < t_1$, $t_2 < x \leqq x_{n+1}$, $= \epsilon/K + \eta(x - t_1)(t_2 - x)$, $t_1 \leqq x \leqq t_2$, where ϵ and η are > 0.

Note that $y^{(n)} > 0$ and continuous in $[x_1, x_{n+1}]$. It can obviously be modified to be arbitrarily many times differentiable as well. For this $y^{(n)}$ we have by the standard properties of divided differences

$$\Delta(1, 2, \cdots, n+1)y = y^{(n)}(\xi)/n!$$

where ξ is a mean value in $[x_1, x_{n+1}]$.

$$\therefore\ 0 < y^{(n)}(\xi)/n! = \int_{x_1}^{x_{n+1}} Ty^{(n)}\,dt = I_1 + I_2 \tag{3.7}$$

where $I_1 = \{\int_{x_1}^{t_1} + \int_{t_2}^{x_{n+1}}\} Ty^{(n)}\,dt$ and $I_2 = \int_{t_1}^{t_2} Ty^{(n)}\,dt$.

Now $|I_1| < \epsilon$, and $I_2 < 0$. Further $|I_2|$ can be made as large as desired by choosing η large enough. Therefore the right-hand side of (3.7) can be made negative by suitable choice of ϵ and η. This is a contradiction so that $T \geqq 0$ in $[x_1, x_{n+1}]$.

The decision to use Gaussian quadrature fixes the points of S_2 as the zeros of the orthogonal polynomial of degree m with respect to T as weight function. The corresponding quadrature formula will be exact for polynomials of degree $2m - 1$ (i.e., whenever y is a polynomial of degree $2m + n - 1$) so that the error in the optimised form of (2.5) will be $O(h^{2m})$ as $T = O(h^{-1})$ and the range of integration is $O(h)$. The error in the optimised quadrature formula is (for $m > 1$) smaller than the interpolation error.

4. Some Examples. In all but very special cases the construction of difference approximations rapidly becomes extremely tedious as the order of the differential equations increases, and for this reason the examples considered in this section refer to the equation

$$d^2y/dx^2 + f(x)y = g(x). \tag{4.1}$$

Let S_1 consist of the points $x_1 = -h_1$, $x_2 = 0$, $x_3 = h_2$, then

$$\begin{aligned} T &= (x + h_1)/h_1(h_1 + h_2), \qquad -h_1 \leqq x < 0, \\ &= (h_2 - x)/h_2(h_1 + h_2), \qquad 0 \leqq x < h_2, \end{aligned} \tag{4.2}$$

and

$$\int_{-h_1}^{h_2} Tx^r\,dx = \frac{1}{(r+1)(r+2)} \frac{h_2^{r+1} + (-1)^r h_1^{r+1}}{h_2 + h_1}. \tag{4.3}$$

Even in the case $s = 1$, the problem of computing the quadrature points for the weight function T requires the solution of three nonlinear equations in three unknowns. This presents little difficulty on a computer, but does not make for ease of presentation. However the most important special case (where $h_1 = h_2 = h$) is readily soluble. The quadrature points will be the zeros of a cubic polynomial $P = x^3 + Ax^2 + Bx + C$ where P must satisfy the orthogonality conditions

$$\int_{-h}^{h} TP\,dx = 0 = Ah^2/6 + C,$$

$$\int_{-h}^{h} TPx\,dx = 0 = h^2/15 + B/6,$$

$$\int_{-h}^{h} TPx^2\,dx = 0 = Ah^2/15 + C/6,$$

so that

$$A = C = 0, \qquad B = -2h^2/5 \quad \text{and} \quad P = x(x^2 - 2h^2/5). \tag{4.4}$$

Thus the points of S_2 are $\xi_1 = -h(\frac{2}{5})^{1/2}$, $\xi_2 = 0$, $\xi_3 = h(\frac{2}{5})^{1/2}$. The quadrature weights are $v_1 = v_3 = 5/48$, $v_2 = 7/24$.

The difference equation can now be derived using the method of Section 2. However, in this case, it is easy to write down an interpolation polynomial which has an error $O(h^5)$ as the values of $y^{(2)}$ are given on S_1 by the differential equation. Writing $y(x_i) = y_i$ this interpolation polynomial is

$$Q = y_2 + xh^{-1}\mu\delta\left(y_2 - \frac{h^2}{6}y_2^{(2)}\right) + \frac{x^2}{2}y_2^{(2)} + \frac{x^3}{6}h^{-1}\mu\delta y_2^{(2)} + \frac{x^4}{24}h^{-2}\delta^2 y_2^{(2)} \tag{4.5}$$

giving the difference equation

$$\begin{aligned}\Delta(1, 2, 3)y &= (h^{-2}/2)\delta^2 y_2 \\ &= -1/48\{5f(-(\tfrac{2}{5})^{1/2}h)Q(-(\tfrac{2}{5})^{1/2}h) + 14f(0)Q(0) \\ &\quad + 5f((\tfrac{2}{5})^{1/2}h)Q((\tfrac{2}{5})^{1/2}h) - 5g(-(\tfrac{2}{5})^{1/2}h) - 14g(0) - 5g((\tfrac{2}{5})^{1/2}h)\}\end{aligned} \tag{4.6}$$

where the second derivatives have been evaluated from

$$y^{(2)}(x) = -f(x)Q(x) + g(x) + O(h^5). \tag{4.7}$$

However, if Lobatto quadrature is used to reduce the quadrature error, then the general case $s = 1$ is quite tractable. The resulting difference equation has an error of $O(h^4)$ which is the same as that of the Numerov equation, and it may be useful for problems in which graded meshes are necessary.

The use of Lobatto quadrature fixes $\xi_1 = -h_1$ and $\xi_3 = h_2$, and leaves ξ_2 free to be adjusted to give maximum accuracy. By the usual argument, ξ_2 is given by the equation

$$\int_{-h_1}^{h_2} T(x + h_1)(h_2 - x)(x - \xi_2)\,dx = 0 \tag{4.8}$$

which has the solution

$$\xi_2 = \frac{h_2 - h_1}{5} \cdot \frac{2h_1^2 + 5h_1h_2 + 2h_2^2}{h_1^2 + 3h_1h_2 + h_2^2}. \tag{4.9}$$

The corresponding quadrature weights are

$$\begin{aligned}v_1 &= \frac{1}{12} \cdot \frac{3h_2^4 + 6h_2^3h_1 + 9h_2^2h_1^2 + 6h_2h_1^3 + h_1^4}{(h_1 + h_2)(2h_2^3 + 8h_2^2h_1 + 12h_2h_1^2 + 3h_1^3)}, \\ v_2 &= \frac{1}{12} \cdot \frac{(h_2^2 + 3h_1h_2 + h_1^2)^3}{(3h_2^3 + 12h_2^2h_1 + 8h_2h_1^2 + 2h_1^3)(2h_2^3 + 8h_2^2h_1 + 12h_2h_1^2 + 3h_1^3)}, \\ v_3 &= \frac{1}{12} \cdot \frac{h_2^4 + 6h_2^3h_1 + 9h_2^2h_1^2 + 6h_2h_1^3 + 3h_1^4}{(h_1 + h_2)(3h_2^3 + 12h_2^2h_1 + 8h_2h_1^2 + 2h_1^3)}.\end{aligned} \tag{4.10}$$

When $h_1 = h_2$ then $\xi_2 = 0$, and the quadrature weights reduce to those appropriate to the Numerov formula.

When $m = 5$ and the points of S_1 are equispaced, then Lobatto quadrature is again tractable. In this case the error in the resulting quadrature formula is $O(h^8)$, while the interpolation error is $O(h^7)$, so that this formula is optimal.

Formulae for approximating to boundary conditions can be derived using

similar techniques to those discussed above. Assume, for example, that S_1 consists of the points $-h_1$, η, and h_2. Then (3.1) takes the form

$$\frac{1}{(h_1+\eta)(h_1+h_2)}y_1+\frac{1}{(\eta+h_1)(\eta-h_2)}y(\eta)+\frac{1}{(h_2-\eta)(h_2+h_1)}y_3 = \int_{-h_1}^{h_2} Ty^{(2)}\,dx \tag{4.11}$$

where

$$\begin{aligned} T &= (x+h_1)/(h_1+\eta)(h_1+h_2), && -h_1 \leqq x < \eta,\\ &= (h_2-x)/(h_2-\eta)(h_1+h_2), && \eta \leqq x < h_2. \end{aligned}$$

If this equation is differentiated with respect to η and then η set $= 0$, there results

$$\frac{1}{h_1+h_2}\left\{-\frac{1}{h_1^2}y_1+\left(\frac{1}{h_1^2}-\frac{1}{h_2^2}\right)y_2+\frac{1}{h_2^2}y_3\right\}-\frac{1}{h_1h_2}y_{(0)}^{(1)} = \int_{-h_1}^{h_2} Vy^{(2)}\,dx \tag{4.12}$$

where

$$\begin{aligned} V &= \frac{\partial T}{\partial \eta}\bigg|'_{\eta=0} = \frac{-1}{h_1+h_2}\frac{x+h_1}{h_1^2}, && -h_1 \leqq x < 0,\\ &= \frac{1}{h_1+h_2}\frac{h_2-x}{h_2^2}, && 0 \leqq x < h_2. \end{aligned}$$

If $h_1 = h_2 = h$, then fitting a quadratic to the values of $y^{(2)}$ on S_1 and integrating leads to the familiar formula

$$y_{(0)}^{(1)} = h^{-1}\mu\delta y_2 - \tfrac{1}{6}h\mu\delta y_2^{(2)} \tag{4.13}$$

which is exact whenever y is a polynomial of degree $\leqq 4$. Again Gaussian quadrature can be used to increase accuracy. Here V changes sign, but xV is positive and can be used as a weight function provided $x = 0$ is a quadrature point. The remaining quadrature points in the case $h_1 = h_2$ have the form $\pm\alpha$ where

$$\int_{-h}^{h} Vx(x^2-\alpha^2)\,dx = 0 \tag{4.14}$$

giving $\pm\alpha = \pm(3/10)^{1/2}h$. The corresponding quadrature weights are $-1/(12\alpha)$, 0, $1/(12\alpha)$. The formula that results when this quadrature is used to evaluate the integral in Eq. (4.12) is exact whenever y is a polynomial of degree $\leqq 6$.

5. Derivation of Some Nonclassical Formulae. The techniques described in Sections 2 and 3 are based on a partitioning of the operator L of the form (writing $d/dx = D$)

$$\begin{aligned} L &= L_1 + L_2,\\ L_1 &= D^n,\\ L_2 &= \sum_{i=0}^{n-1} a_i(x)D^i. \end{aligned}$$

The significant characteristics of the partitioning are

(i) the orders of L and L_1 are the same,

(ii) the equation $L_1(y) = 0$ is readily soluble, and

(iii) a difference equation satisfied by all solutions of $L_1(y) = 0$ is readily determined.

Any other partitioning of L which has these three properties provides a possible basis for generating finite difference approximations to Eq. (2.1). Actually, condition (iii) is a consequence of condition (ii) for let $v_1, v_2, \cdots, v_n$ be a fundamental set of solutions to the equation $L_1(y) = 0$, then the linear dependence of any other solution of them over the points of S_1 gives

$$\begin{vmatrix} y(x_1) & \cdots & y(x_{n+1}) \\ v_1(x_1) & \cdots & v_1(x_{n+1}) \\ \vdots & & \\ v_n(x_1) & \cdots & v_n(x_{n+1}) \end{vmatrix} = 0$$

which is written here as

$$\chi(1, 2, \cdots, n+1)y = 0. \tag{5.1}$$

Note that there is no scaling associated with the operator χ in contrast to Δ where the scale is fixed by convention.

The program of Section 2 can be followed through in this case also. However, some technique such as variation of parameters is needed to generate the interpolation to y from that to $L_1(y)$ so that it is perhaps best to go straight to the formula which corresponds to (3.1). This has the form

$$\chi(1, 2, \cdots, n+1)y = \int_{x_1}^{x_{n+1}} TL_1(y)\, dx \tag{5.2}$$

where now T is characterised by the conditions

(i) $T = 0, x < x_1, x \geqq x_{n+1}$,

(ii) $T, T^{(1)}, \cdots, T^{(n-2)}$ continuous on the points of S_1,

(iii) $\tilde{L}_1(T) = 0$ except at the points of S_1.

Here $\tilde{L}_1$ is the differential operator adjoint to L_1.

Example 1. Consider the self-adjoint differential equation

$$(d/dx)(p\, dy/dx) + qy = f. \tag{5.3}$$

Let v_1 and v_2 form a fundamental set of solutions, and assume that they satisfy the conditions

$$v_1(x_1) = v_2(x_3) = 0, \qquad v_1(x_2) = v_2(x_2).$$

Then a possible choice for T is

$$\begin{aligned} T &= v_1(x), \qquad x_1 \leqq x < x_2, \\ &= v_2(x), \qquad x_2 \leqq x < x_3. \end{aligned}$$

Differentiating T in the first integration by parts gives

$$dT/dx = H(x - x_1)H(x_2 - x)\, dv_1/dx + H(x - x_2)H(x_3 - x)\, dv_2/dx$$

where $H(x)$ is the Heaviside unit function. The second integration by parts brings in the δ functions which give the difference equation

$$p(x_1)\frac{dv_1(x_1)}{dx}y(x_1) - p(x_2)\left(\frac{dv_1(x_2)}{dx} - \frac{dv_2(x_2)}{dx}\right)y(x_2) - p(x_3)\frac{dv_2(x_3)}{dx}y(x_3) = \int_{x_1}^{x_3} Tf\,dx. \tag{5.4}$$

Example 2. Consider the special case $p = 1$, $q > 0$, and define $L_1 = D^2 + q_2$ where $q_2 = q(x_2)$, then

$$v_1(x) = \frac{\sin (q_2)^{1/2}(x - x_1)}{\sin (q_2)^{1/2}(x_2 - x_1)},$$

$$v_2(x) = \frac{\sin (q_2)^{1/2}(x_3 - x)}{\sin (q_2)^{1/2}(x_3 - x_2)}.$$

If x_1 is specialised to $x_2 - h$ and x_3 to $x_2 + h$ then (5.4) becomes

$$y_1 - 2\cos (h(q_2)^{1/2})y_2 + y_3 = \frac{\sin (h(q_2)^{1/2})}{(q_2)^{1/2}}\int_{x_1}^{x_3} T(q_2 - q)y\,dx. \tag{5.5}$$

This formula is given by Hersch in [4] and his derivation has been followed closely here. An application of this equation to an eigenvalue problem has been given in [3].

A range of difference equations can be obtained by substituting different interpolations for y on the right-hand side of Eq. (5.5). If, for example, the interpolation polynomial given by Eq. (4.5) is used, and if the resulting integral can be evaluated exactly, then the interpolation error in Eq. (5.5) will be $O(h^8)$. However the left-hand side of this equation tends to $h^2L_1(y)$ as $h \to 0$ so that the error is only $O(h^6)$ on a scale comparable with that used in Eq. (4.6). Gaussian quadrature with respect to T as weight function can also be used. If a three-point Gaussian formula is used then the quadrature error will be $O(h^8)$, and the error on a scale comparable with that used in Eq. (4.6) is again $O(h^6)$.

Example 3. Let Q be the quadratic interpolation polynomial fitted to q on the points of S_1. In this case define $L_1 = D^2 + Q$. Explicit formulae for v_1 and v_2 do not exist in general, but they can be generated to any degree of accuracy by Taylor series methods. Assuming that T is positive on S_1, Gaussian quadrature can be used to estimate $\int_{x_1}^{x_3} T(Q - q)\,ydx$. It is again most convenient to compute y from (4.5), and in this case the error (again using the scale appropriate to (4.6)) is $O(h^8)$ as $Q - q$ is $O(h^3)$.

6. Assessing the Difference Equations. In the two previous sections several formulae have been suggested which offer different compromises between accuracy and ease of construction. In this section an attempt is made to provide a criterion for selecting between them. The following assumptions are made.

A. That a realistic bound of the form Kh^r can be found for the error in the solution to the difference equation. It is assumed that $K = O(1)$ as $h \to 0$, and that r is the order of the error in the difference approximation measured in the scale appropriate to Eq. (4.6).

B. That the number of evaluations of the coefficients in the differential equation is an adequate measure of the work done in obtaining an approximate solution to the differential equation.

This last is really two assumptions: (i) that the work done in setting up the difference equation dominates the work done in solving it, and (ii) that the work done in setting up the difference equation is effectively the work done in evaluating the coefficients in the differential equation at the appropriate points.

Note that while B is a realistic assumption for our purposes it does not generalise. For example, in solving a Fredholm integral equation of the second kind by finite differences $O(n^2)$ function evaluations are required in setting up the linear equations. The matrix of this set of equations is full, and its solution requires $O(n^3)$ multiplications. In this case it is likely that the work of solution would be dominant. Thus assumption B takes account of the band structure of the matrices produced by finite difference approximations to ordinary differential equations.

If E is the permitted tolerance for the error in the solution, then h must satisfy

$$h \leqq (E/K)^{1/r}. \tag{6.1}$$

Also let J be the average number of new evaluations of coefficients required in computing the difference equation at each mesh point (assuming that values at the $(i+1)$st point are computed after those at the ith, and that common values are reused). Then the work necessary to integrate the differential equation from $x = a$ to $x = b$ is approximately

$$W = J(b-a)(K/E)^{1/r}. \tag{6.2}$$

To compare two methods (referred to by suffices 1 and 2) the ratio W_1/W_2 is appropriate. This contains the terms K_1^{1/r_1} and K_2^{-1/r_2} which are difficult to specify precisely as they are dependent on the error constants, on fairly high derivatives of the solution, and on the conditioning of the original problem and that of the difference approximations. However, these terms tend to cancel one another out, and the exponents $1/r_1$ and $1/r_2$ tend to reduce their influence strongly. Accepting this as an argument for ignoring the terms in K_1 and K_2 largely on the basis of expediency, we are led to define a *relative efficiency index*

$$R_{12} = \frac{J_1}{J_2} E^{(1/r_2 - 1/r_1)}. \tag{6.3}$$

Example 1. Consider Eq. (5.3) with $p = 1$. Two possible finite difference approximations are

(i) $\delta^2 y_i + h^2 q_i y_i = h^2 f_i$ (standard), and

(ii) $\delta^2 y_i + h^2(1 + \frac{1}{12}\delta^2)(q_i y_i - f_i) = 0$ (Numerov).

In (i) the truncation error is

$$-\frac{h^4}{12}\left(\frac{d^4y}{dx^4}\right)_i + O(h^6)$$

and in (ii) it is

$$\frac{h^6}{240}\left(\frac{d^6y}{dx^6}\right)_i + O(h^8).$$

Thus $r_1 = 2$ and $r_2 = 4$. Clearly $J_1 = J_2 = 1$ so that if $E = 10^{-6}$ then $R_{12} = 10^{1.5}$. This indicates that method (i) would require about 30 times as many mesh points as method (ii) to give six correct decimal places.

It is interesting that in this case at least the error constants contribute little to the ratio W_1/W_2 for $(1/12)^{1/2}/(1/240)^{1/4} \approx 1.1$.

Example 2. Compare now the Numerov formula with the formula (4.6) and the Gaussian type formulae suggested in Examples 2 and 3 of Section 5. Again we take $E = 10^{-6}$.

(i) Numerov compared with (4.6). Here $J_2 = 3$, $r_2 = 5$, $R_{12} = \frac{1}{3}\, 10^{3/10} \approx .7$.

(ii) Numerov compared with the Gaussian formula of Section 5, Example 2. Here $J_2 = 3$, $r_2 = 6$. However, the quadrature points and weights must also be evaluated (consider Eqs. (4.9) and (4.10)). Depending on the complexity of the coefficients, an effective J_2 may be expected to range between 3 and (say) 7.

For these extremes

$$J_2 = 3, \qquad R_{12} = \tfrac{1}{3}(10)^{1/2} \approx 1,$$

$$J_2 = 7, \qquad R_{12} = \tfrac{1}{7}(10)^{1/2} \approx .4.$$

(iii) Numerov compared with the Gaussian formula of Section 5, Example 3. Here $r_2 = 8$ giving $R_{12} = 10^{.75}/J_2$.

In this case the number of coefficient evaluations (3) cannot be expected to be a reasonable measure of the work involved in setting up the difference equation as there are no closed formulae for the quadrature points and weights. However, R_{12} cannot be greater than the value obtained by taking $J_2 = 3$. This value $\cong 2$.

From these figures it is clear that the Numerov formula is very attractive even when compared with the very accurate formulae based on Gaussian quadrature. An additional feature in its favour when solving eigenvalue problems is that the eigenvalue parameter would appear linearly in it if it entered the original differential equation linearly. This is not true for any of the more accurate formulae considered.

However, note that R_{12} depends only on the two difference approximations and not at all on the differential equation to be solved. Its use must therefore be tempered by discretion. What it can do is provide a prior guide to a suitable difference approximation by considering those features which always contribute to the work of solution.

Of course, if an estimate is known for the magnitudes of the appropriate derivatives of the solution of the differential equation then their contribution to the term $K_1^{1/r_1}/K_2^{1/r_2}$ can be estimated. Note also that these terms depend on the choice of scales for the independent and dependent variables, and that the use of R_{12} can only be appropriate if "sensible" scales are adopted.

7. Acknowledgments. The author wishes to acknowledge the influence of Mr. P. M. Keeping and Mr. D. Kershaw of Edinburgh University on the material presented in this paper. The author and Mr. Keeping have worked on the use of Radau quadrature to minimise truncation error in finite-difference formulae suitable for the integration of stiff systems of differential equations. It is hoped to publish this

work (which antedates the present paper) shortly. The material shown in Sections 3 and 4 was produced in close collaboration with Mr. Kershaw.

Computer Centre
Australian National University
Canberra, Australia

1. J. T. DAY, "A one-step method for the numerical solution of second order linear ordinary differential equations," *Math. Comp.*, v. 18, 1964, pp. 664–668. MR **29** #5385.
2. M. R. OSBORNE, "A method for finite-difference approximation to ordinary differential equations," *Comput. J.*, v. 7, 1964, pp. 58–65. MR **31**, #5338.
3. M. R. OSBORNE & S. MICHAELSON, "The numerical solution of eignvalue problems in which the eigenvalue parameter appears nonlinearly, with an application to differential equations," *Comput. J.*, v. 7, 1964, pp. 66–71. MR **31** #4167.
4. J. HERSCH, "Contribution a la méthode des équations aux différences," *Z. Angew. Math. Phys.*, v. 9a, 1958, pp. 129–180. MR **21** #1708.
5. M. E. ROSE, "Finite difference schemes for differential equations," *Math. Comp.*, v. 18, 1964, pp. 179–195. MR **32** #605.