The use of higher derivatives in quadrature formulae

By J. D. Lambert and A. R. Mitchell

Some new high accuracy quadrature formulae involving derivatives of the integrand are derived and compared with existing formulae.

1. Introduction

Interpolation formulae using function values and derivatives at several points have been known for some time. An extensive list of references relating to this topic can be found in Hildebrand (1956).

Quadrature formulae involving derivatives of the integrand have been proposed by several authors. For example, Squire (1961) produced a class of quadrature formulae in which the value of an integral over a finite range is expressed in terms of the integrand and its derivatives at the end points of the range. Also, Hammer and Wicke (1960) derived formulae of a special type which make use of the values of the integrand and its derivatives at special points unequally spaced within the range of integration. The evaluation of the coefficients and the special points was carried out by Struble (1960) for a variety of cases.

It is the object of the present paper to generate the general family of quadrature formulae which involve derivatives of the integrand at equally spaced intervals, and to examine the usefulness of such formulae.

2. Optimum formulae

In order to evaluate the integral $\int_{x_0}^{x_k} f(x)dx$, consider the class of formulae

$$y_k = y_0 + \sum_{s=1}^{l} \sum_{t=0}^{k} a_{st} h^s y_t^{(s)}, \; (k \geqslant 1, \; l \geqslant 1) \qquad (1)$$

where
$$\frac{dy}{dx} = f(x);$$

y_0, y_k are the values of y at the points x_0, x_k; $y_t^{(s)}(s = 1, 2, \ldots l; \; t = 0, 1, \ldots k)$ are the l successive derivatives of y at the $(k + 1)$ neighbouring points $x_0 + th$, h being the distance between consecutive points along the x-axis, and a_{st} are coefficients to be determined.

If y_0, y_k and the derivatives of y up to order l at the $(k + 1)$ points are expanded as Taylor series about a convenient origin, the expansions substituted into (1) and the coefficients of the various powers of h equated to zero, then the following equations are obtained

$$\sum_{t=0}^{k} a_{1,t} = k \qquad (j = 1)$$

$$\sum_{s=1}^{j-1} \sum_{t=1}^{k} t^s a_{j-s,t} + \sum_{t=0}^{k} a_{j,t} = \frac{k^j}{j!} \qquad (j = 2, 3, \ldots l)$$

$$\sum_{s=1}^{l} \frac{1}{(j-s)!} \sum_{t=1}^{k} t^{j-s} a_{s,t} = \frac{k^j}{j!} \qquad (j = l+1, l+2, \ldots).$$

The first $(k + 1)l$ of these equations can be solved to give the values of $a_{s,t} (s = 1, 2, \ldots l; \; t = 0, 1, \ldots k)$. With these values (1) now determines a class of formulae with minimum truncation error.

The values of the coefficients and the form of the principal part of the truncation error are independent of the choice of origin. If, however, the mid-point of the range is taken as origin, the labour of calculating the coefficients is considerably reduced. The form of the terms after the first in the truncation error series depends on the choice of origin, and if the mid-point is chosen, the second term vanishes. Accordingly the values of $a_{s,t}$ and the first two non-vanishing terms of the truncation error are given in Table 1 for $l = 2$ and in Table 2 for $l = 3$. Coefficients have been quoted for $k \leqslant 5$ in the case $l = 2$, and for $k \leqslant 4$ in the case $l = 3$. For larger values of k, the coefficients become excessively unwieldy, and in any case, the coefficients of the function values are no longer all positive, a situation which can lead to an adverse accumulation of rounding error. (A negative coefficient for a function value has in fact already appeared when $k = 4$, $l = 3$.) Henceforth the formula of class (1) with $k = \alpha$, $l = \beta$ will be referred to as $[\alpha; \beta]$. For example, Simpson's rule is the formula [2; 1], the general Newton–Cotes formula is $[n\,; 1]$, and the general formula of the class derived by Squire is $[1\,; n]$.

3. Sub-optimum formulae

Formulae involving derivatives of the integrand with an increased truncation error can also be obtained from (1) by putting certain of the coefficients a_{st} equal to zero. If p of the coefficients are taken to be zero, then only the first $(k + 1)l - p$ equations of (2) need be solved to give the remaining coefficients. The formulae obtained are referred to as *sub-optimum formulae*.

A class of sub-optimum formulae is now derived for the numerical evaluation of $\int_{\alpha}^{\beta} \eta(x)dx$, where η is a solution of the equation $\eta'' = g(x, \eta)$ and is tabulated at regular intervals of x. Examples of such functions η are the Airy integral, the Whittaker function, and the Mathieu function. If in (1) all the coefficients except a_{1t} and $a_{3t} (t = 0, \ldots k)$ are equated to zero, the class of formulae is obtained which involves the values of η and its second derivative, the latter being readily obtained in terms of η from the differential equation. These formulae are quoted for $k = 1, 2, 3, 4$ in Table 3.

Table 1
$l = 2$

Coefficients

k	1	2	3	4	5
a_{10}	$\frac{1}{2}$	$\frac{7}{15}$	$\frac{93}{224}$	$\frac{3,202}{8,505}$	$\frac{319,085}{912,384}$
a_{11}	$\frac{1}{2}$	$\frac{16}{15}$	$\frac{243}{224}$	$\frac{8,192}{8,505}$	$\frac{691,875}{912,384}$
a_{12}		$\frac{7}{15}$	$\frac{243}{224}$	$\frac{11,232}{8,505}$	$\frac{1,270,000}{912,384}$
a_{13}			$\frac{93}{224}$	$\frac{8,192}{8,505}$	$\frac{1,270,000}{912,384}$
a_{14}				$\frac{3,202}{8,505}$	$\frac{691,875}{912,384}$
a_{15}					$\frac{319,085}{912,384}$
a_{20}	$\frac{1}{12}$	$\frac{1}{15}$	$\frac{57}{1,120}$	$\frac{116}{2,835}$	$\frac{36,975}{1,064,448}$
a_{21}	$-\frac{1}{12}$	0	$-\frac{81}{1,120}$	$\frac{512}{2,835}$	$-\frac{314,375}{1,064,448}$
a_{22}		$-\frac{1}{15}$	$\frac{81}{1,120}$	0	$-\frac{272,500}{1,064,448}$
a_{23}			$-\frac{57}{1,120}$	$\frac{512}{2,835}$	$\frac{272,500}{1,064,448}$
a_{24}				$-\frac{116}{2,835}$	$\frac{314,375}{1,064,448}$
a_{25}					$-\frac{36,975}{1,064,448}$

Truncation Errors

$h^5y^{(5)}$	$-\frac{1}{720}$				
$h^7y^{(7)}$	$-\frac{1}{40,320}$	$-\frac{1}{4,725}$			
$h^9y^{(9)}$		$-\frac{1}{113,400}$	$-\frac{9}{313,600}$		
$h^{11}y^{(11)}$			$-\frac{53}{27,596,800}$	$-\frac{8}{1,964,655}$	
$h^{13}y^{(13)}$				$-\frac{724}{1,915,538,625}$	$-\frac{144,425}{230,150,688,768}$
$h^{15}y^{(15)}$					$-\frac{2,875}{38,626,689,024}$

Table 2 $l = 3$

Coefficients

k	1	2	3	4
a_{10}	$\frac{1}{2}$	$\frac{41}{105}$	$\frac{3,849}{9,856}$	$\frac{1,257,482}{3,648,645}$
a_{11}	$\frac{1}{2}$	$\frac{128}{105}$	$\frac{10,935}{9,856}$	$\frac{6,848,512}{3,648,645}$
a_{12}		$\frac{41}{105}$	$\frac{10,935}{9,856}$	$-\frac{1,617,408}{3,648,645}$
a_{13}			$\frac{3,849}{9,856}$	$\frac{6,848,512}{3,648,645}$
a_{14}				$\frac{1,257,482}{3,648,645}$
a_{20}	$\frac{1}{10}$	$\frac{2}{35}$	$\frac{2,799}{49,280}$	$\frac{52,552}{1,216,215}$
a_{21}	$-\frac{1}{10}$	0	$-\frac{2,187}{49,280}$	$\frac{290,816}{1,216,215}$
a_{22}		$-\frac{2}{35}$	$\frac{2,187}{49,280}$	0
a_{23}			$-\frac{2,799}{49,280}$	$-\frac{290,816}{1,216,215}$
a_{24}				$-\frac{52,552}{1,216,215}$
a_{30}	$\frac{1}{120}$	$\frac{1}{315}$	$\frac{153}{49,280}$	$\frac{2,408}{1,216,215}$
a_{31}	$\frac{1}{120}$	$\frac{16}{315}$	$\frac{2,187}{49,280}$	$\frac{126,976}{1,216,215}$
a_{32}		$\frac{1}{315}$	$\frac{2,187}{49,280}$	$-\frac{184,788}{1,216,215}$
a_{33}			$\frac{153}{49,280}$	$\frac{126,976}{1,216,215}$
a_{34}				$\frac{2,408}{1,216,215}$

Truncation Errors

$h^7 y^{(7)}$	$\frac{1}{100,800}$			
$h^9 y^{(9)}$	$\frac{1}{7,257,600}$			
$h^{11} y^{(11)}$		$\frac{1}{130,977,000}$		
$h^{13} y^{(13)}$		$\frac{1}{5,108,103,000}$	$\frac{9}{5,637,632,000}$	
$h^{15} y^{(15)}$			$\frac{171}{2,209,951,744,000}$	
$h^{17} y^{(17)}$				$\frac{478}{162,983,603,908,125}$
$h^{19} y^{(19)}$				$\frac{4,727}{27,870,196,268,289,375}$
$h^{21} y^{(21)}$				

Table 3

Coefficients

k	1	2	3	4
a_{10}	$\frac{1}{2}$	$\frac{5}{21}$	$\frac{3}{56}$	$\frac{8,674}{39,105}$
a_{11}	$\frac{1}{2}$	$\frac{32}{21}$	$\frac{81}{56}$	$\frac{57,344}{39,105}$
a_{12}		$\frac{5}{21}$	$\frac{81}{56}$	$\frac{24,384}{39,105}$
a_{13}			$\frac{3}{56}$	$\frac{57,344}{39,105}$
a_{14}				$\frac{8,674}{39,105}$
a_{30}	$-\frac{1}{24}$	$-\frac{1}{315}$	$\frac{9}{1,120}$	$-\frac{1,912}{821,205}$
a_{31}	$-\frac{1}{24}$	$\frac{32}{315}$	$\frac{351}{1,120}$	$\frac{102,400}{821,205}$
a_{32}		$-\frac{1}{315}$	$\frac{351}{1,120}$	$\frac{56,064}{821,205}$
a_{33}			$\frac{9}{1,120}$	$\frac{102,400}{821,205}$
a_{34}				$-\frac{1,912}{821,205}$

Truncation Errors

$h^5y^{(5)}$	$-\frac{1}{120}$			
$h^7y^{(7)}$	$-\frac{1}{5,040}$			
$h^9y^{(9)}$		$-\frac{1}{396,900}$	$\frac{39}{313,600}$	
$h^{11}y^{(11)}$		$-\frac{1}{13,097,700}$	$\frac{211}{27,596,800}$	
$h^{13}y^{(13)}$				$-\frac{42,608}{554,867,688,375}$
$h^{15}y^{(15)}$				$-\frac{80}{13,316,824,521}$

Table 4

Numerical Evaluation of $\int_{-1}^{+1} \frac{dx}{x+2}$ (Theoretical value 1·098,612,288,668)

	METHOD	NUMBER OF VALUES	MESH SIZE	ORDER OF TRUNCATION ERROR	RESULT	ERROR
1	Formula [2; 3] once	8	1	11	1·098,667,854	+0·000,035,565
2	Formula [2; 3] twice	12	$\frac{1}{2}$	11	1·098,612,522	+0·000,000,233
3	Sub-optimum formula	25	$\frac{1}{10}$	9	1·098,612,288,785	+0·000,000,000,117
4	Struble ($l = 3$)	3	Uneven mesh (3 points)	7	1·096,652,562	−0·001,959,727
5	Struble ($l = 3$)	5	Uneven mesh (5 points)	11	1·098,590,615	−0·000,021,674
6	9-point Newton–Cotes once	9	$\frac{1}{4}$	10	1·098,616,867	+0·000,004,578
7	9-point Newton–Cotes twice	17	$\frac{1}{8}$	10	1·098,612,304	+0·000,000,035
8	9-point Newton–Cotes three times	25	$\frac{1}{12}$	10	1·098,612,289,926	+0·000,000,001,258
9	Weddle's Rule four times	25	$\frac{1}{12}$	7	1·098,612,332	+0·000,000,044

Sub-optimum formulae can also be chosen in such a way that the coefficients are simpler than those of the corresponding optimum formulae, and the increase in truncation error may be compensated for by the greater simplicity of calculation. In many cases, a solution of desired accuracy is most readily obtained by employing such a sub-optimum formula with a suitably small mesh length. For example, a sub-optimum formula with $k = 2$, $l = 4$, which is particularly suitable for repeated application is

$$y_2 - y_0 = \frac{h}{63}(31f_0 + 64f_1 + 31f_2) + \frac{5h^2}{63}(f_0^{(1)} - f_2^{(1)}) - \frac{h^4}{945}(f_0^{(3)} - f_2^{(3)}),$$

with a principal truncation error of $\frac{1}{198,450}h^9f^{(8)}$. If this formula is used n times to cover the range of integration $2nh$, the result

$$y_{2n} - y_0 = \frac{h}{63}[31f_0 + 64f_1 + 62f_2 + 64f_3 + \ldots + 62f_{2n-2} + 64f_{2n-1} + 31f_{2n}] + \frac{5h^2}{63}[f_0^{(1)} - f_{2n}^{(1)}] - \frac{h^4}{945}[f_0^{(3)} - f_{2n}^{(3)}]$$

is obtained.

4. Numerical results

It is very difficult to get a fair means of comparing various classes of quadrature formulae. The numerical results obtained for the evaluation of $\int_{-1}^{+1} \frac{1}{x+2}dx$ by a selection of formulae are set out in Table 4. The figure entered in the column headed "number of values" denotes the total number of values of the function and of its derivatives employed in the formula.

It is of particular interest to compare methods 3 and 8, which both involve the same number of function values. Method 8 (9-point Newton–Cotes used three times) has a smaller mesh size and a higher-order truncation error than method 3, which uses the sub-optimum formula of the last Section. Nevertheless, method 3 gives the more accurate result. This is due to the very small numerical factor in the principal part of the truncation error. The smallness of these factors is a general feature of the methods proposed in the present paper, as will be seen from Tables 1, 2, and 3. It is thus possible to obtain accuracy comparable with the Newton–Cotes formulae with a very much larger mesh size. Provided the derivatives of the integrand can be evaluated at the end points of the range without too much difficulty, method 3 again compares favourably with method 8 in ease of computation on a *desk* machine, since the coefficients are much simpler.

5. Concluding remarks

In general, it is impossible to single out one particular formula of class (1) as giving the best balance between accuracy and labour of computation, since the latter varies greatly with the form of the function and its derivatives. Nevertheless, the formulae using higher derivatives developed in the present paper can have a distinct advantage over existing formulae when the integrand, although given analytically as a continuous function of x, is most readily evaluated by reference to a tabulation at discrete values of x. Provided such a function can be differentiated, sufficient accuracy may be achieved by using formulae involving higher derivatives with a mesh size large enough to avoid the need for interpolation of the tabulated function.

Finally, although quadrature formulae are frequently employed in the solution of first-order ordinary differential equations, it is not intended that the formulae derived in the present paper be used for that purpose. The formulae given by (1) are designed to perform quadrature with the maximum accuracy, and the question of stability, which arises if they are used in a step-by-step calculation of an initial-value problem, has not been considered. In any case, a wider class of finite-difference formulae involving higher derivatives which can be used in step-by-step solutions of the equation $y' = f(x, y)$ has already been discussed by the present authors (1962).

6. Acknowledgement

The authors are indebted to Dr. J. C. P. Miller, Mathematical Laboratory, Cambridge University, for some helpful remarks during the preparation of the present paper.

References

HAMMER, P. C., and WICKE, H. W. (1960). "Quadrature formulae involving derivatives of the integrand," *Math. Comp.*, Vol. 14, pp. 3–7.

HILDEBRAND, F. B. (1956). *Introduction to Numerical Analysis*, Chapter 8, McGraw-Hill, New York.

LAMBERT, J. D., and MITCHELL, A. R. (1962). "On the solution of $y' = f(x, y)$ by a class of high accuracy difference formulae of low order," *Z.A.M.P.*, Vol. 13, pp. 223–232.

SQUIRE, W. (1961). "Some applications of quadrature by differentiation," *Journ. Soc. Indust. Appl. Math.*, Vol. 9, pp. 94–108.

STRUBLE, G. (1960). "Tables for use in quadrature formulae involving derivatives of the integrand," *Math. Comp.*, Vol. 14, pp. 8–12.

Note on the solution of certain tri-diagonal systems of linear equations

By D. J. Evans and C. V. D. Forrington

This note describes a technique for solving the tri-diagonal systems of linear equations encountered in the numerical solution of certain types of partial differential equations.

In some methods of numerical solution of second-order elliptic difference equations with constant coefficients, it is necessary to solve repeatedly tri-diagonal systems of a certain restricted form. It is usual to solve these systems by an algorithm based on Gaussian elimination, which leads to a simple process involving $9N$ arithmetic operations, including N divisions, where N is the order of the system.

We have considered a method, which when applicable, reduces the number of operations and requires only a small number of divisions independent of N. Thus the advantage in speed may be considerable on machines where the division operation is slow compared with the other operations.

Consider a restricted system of the form

$$\begin{bmatrix} a & b & & & & \\ b & a & b & & & \\ & b & a & b & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & b & a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \\ x_N \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \cdot \\ \cdot \\ \cdot \\ d_N \end{bmatrix} \quad (1)$$

or in matrix notation $Ax = d$.

We attempt to factorize A into the product $\boldsymbol{LL}^T$, where $\boldsymbol{L}$ is a lower triangular matrix and $\boldsymbol{L}^T$ its transpose.

Now consider the product $\boldsymbol{LL}^T$ when

$$\boldsymbol{L} = \begin{bmatrix} r & & & & \\ q & r & & & \\ & q & r & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & \cdot & \\ & & & q & r \end{bmatrix} \quad (2)$$

then

$$\boldsymbol{LL}^T = \begin{bmatrix} r^2 & qr & & & \\ qr & q^2 + r^2 & qr & & \\ & qr & q^2 + r^2 & qr & \\ & & \cdot & & \\ & & & \cdot & \\ & & & \cdot & \\ & & & qr & q^2 + r^2 \end{bmatrix}. \quad (3)$$

By comparing $\boldsymbol{LL}^T$ and $\boldsymbol{A}$ we see that values of q and r can be chosen so that $\boldsymbol{LL}^T \equiv \boldsymbol{A}$ for every element except the first element in the first row. This is so when $q^2 + r^2 = a$ and $qr = b$ which gives

$$r^2 = \tfrac{1}{2}[a \pm \sqrt{(a^2 - 4b^2)}], \; q = b/r. \quad (4)$$