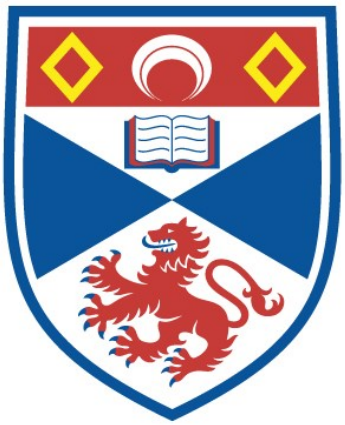
FINITE DIFFERENCE TECHNIQUES OF IMPROVED ACCURACY

John Denholm Lambert

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1963

FINITE DIFFERENCE TECHNIQUES

OF

IMPROVED ACCURACY.

A thesis presented by

John D. Lambert, B.Sc.,

to

the University of St. Andrews,

in application for

the degree of

Doctor of Philosophy.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

DECLARATION.

I declare that the following thesis is a record of research work carried out by me, that the thesis is my own composition, and that it has not previously been presented in application for a higher degree.

[redacted]

PREFACE.

In 1954, I graduated from the University of St. Andrews, with the degree of B.Sc., with Honours in Mathematics, and from 1954 to 1957 I was a member of a Fluid Dynamics Research Group in the Royal Naval Scientific Service. In September 1957, I was appointed as Assistant Professor in Mathematics at the Memorial University of Newfoundland. I returned to this country in the summer of 1958, and, for the academic year 1958-9, held the post of Lecturer in Mathematics at Magee University College, Londonderry. During that year, I carried out research work under the direction of Dr S.C.R. Dennis, then of Queen's University, Belfast. In October 1959, I was appointed to my present post as Lecturer in Applied Mathematics in St. Salvator's College, University of St. Andrews, and from that date, was admitted, under Ordinance 61, as a part-time Research Student, under the supervision of Dr A. R. Mitchell.

CERTIFICATE.

I certify that J. D. Lambert has spent the equivalent of nine terms in part-time research work under my direction, and is thus qualified to submit the accompanying thesis in application for the degree of Doctor of Philosophy.

[redacted].

Research Supervisor.

CONTENTS.

continued overleaf.

Page.

CHAPTER IV. HIGH ACCURACY QUADRATURE FORMULAE.

CHAPTER I.

INTRODUCTION.

I.1. Introductory Remarks.

If a physical problem can be formulated mathematically, it is more than likely that its solution will eventually depend on the solution of a differential equation, either partial or ordinary. Even when we restrict our attentions to the latter, the class of such equations which can be solved exactly by some analytical process is sadly restricted. Such a seemingly innocent equation as

$$y' = x^2 + y^2 ,$$

admits of no solution in terms of elementary functions. A similar example, of considerable physical significance, arises in the problem of the constant velocity flow of a viscous incompressible fluid over a semi-infinite flat plate. Using Prandtl's boundary-layer approximations, this problem can be reduced to that of solving the ordinary differential equation

$$2y''' + yy'' = 0 ,$$

an equation which has resisted all analytical onslought for over half a century. Indeed, this example gives weight to a remark recently made by Todd [1] that " in discussing properties of matter, (we) regard the medium as continuous, set up differential equations, look at them for a while, give up, and replace them by difference equations."

Even when an analytical solution for a differential equation is available in closed form, the requirements of the physical problem may be such that a tabulation of the solution for a range of values of the argument is ultimately required. The labour of such a

tabulation frequently exceeds that of computing an accurate numerical solution directly from the differential equation. This is especially true when a high speed digital computer is available. Such machines are not at their best when making a tabulation from a given closed formula, while they delight in the repetitive nature of a step-by-step calculation. Indeed it can happen that the numerical solution is more accurate than the tabulation based on the exact solution, when the latter involves special functions for which tables (or sub-routines) are available only to a certain degree of accuracy.

In recent years, increased interest in all branches of numerical analysis has been aroused by the advent of the high speed digital computer. However, there are those who argue, with some justification, that the rapid growth of the study of the computer as a tool has created rather too much interest in the new art of efficient programming, while too little thought has been given to effecting a comparable increase in the efficiency of the basic numerical processes involved. One fears that many users of expensive computing machinery seldom look past the method of Adams and Bashforth [2] (1883) for the solution of an ordinary differential equation, or Simpson's Rule (c. 1750) for the evaluation of a definite integral, although at the time when these methods were devised, there was small demand for a numerical solution of high accuracy, and little means of obtaining it, except at the cost of excessively laborious desk calculation.

It is the major purpose of this thesis to propose finite difference techniques of improved accuracy for the numerical solution

of ordinary differential equations, and for the numerical evaluation of definite integrals, the former problem being discussed in Chapter II, and the latter in Chapter IV. In Chapter III the stability of the formulae evolved in Chapter II is studied.

I.2 Elementary Finite Difference Replacements for the Equation $y' = f(x, y)$.

Consider the following initial value problem involving a first order ordinary differential equation.

$$y' = f(x, y) \quad ; \quad y(x_0) = y_0 \tag{1.1}$$

Consider also equally spaced points on the x-axis,

$$x_r = x_0 + rh, \qquad r = 0, 1, 2, \ldots ,$$

where the constant h is the mesh length. We denote by $y_r^{(s)}$ the derivative $\frac{d^s y}{dx^s}$ evaluated at $x = x_r$, and it is understood that $y_r^{(0)} \equiv y_r$.

The Taylor expansion for y_{r+1} about x_r is

$$y_{r+1} = y_r^{(0)} + hy_r^{(1)} + \frac{h^2}{2!} y_r^{(2)} + \frac{h^3}{3!} y_r^{(3)} + \ldots \tag{1.2}$$

Truncating this series after two terms leads to the following elementary finite difference replacement for (1.1)

$$y_{r+1} = y_r + hf_r , \tag{1.3}$$

where $f_r = f(x_r, y_r)$. Since y_0 is known, this formula can be used with $r = 0$ to yield a value for y_1. Using this value, and putting $r = 1$, a value for y_2 is then obtained, and hence, in turn, $y_3, y_4, \ldots y_n \ldots$ are found by a step-by-step process.

The truncation error of the formula (1.3) at the step r is

defined to be the difference between the value of y_{r+1} predicted by (1.3) and the value of y_{r+1} given by the differential equation (1.1), provided that no round-off error is made. That is

$$\text{Truncation Error} = (y_{r+1})_{\text{difference eqn.}} - (y_{r+1})_{\text{differential eqn.}} \qquad (1.4)$$

Throughout this thesis truncation error will always be defined in this sense. On comparing (1.2) with (1.3), it is evident that the truncation error for formula (1.3) is the infinite series

$$-\frac{h^2}{2!}y_r^{(2)} - \frac{h^3}{3!}y_r^{(3)} - \cdots$$

The leading term of this series is called the principal part of the truncation error, or the principal truncation error, and will be denoted by $(TE)_1$. Thus for formula (1.3),

$$(TE)_1 = -\tfrac{1}{2}h^2y^{(2)}.$$

For reasons which will be discussed in Chapter II, section 3, we prefer not to attach the suffix r to the derivative in the expression for $(TE)_1$. We shall have occasion to refer also to the second term in the truncation error series, and this will be denoted by $(TE)_2$. It should be noted that no term in the truncation error can be fully evaluated without knowledge of the solution y of the differential equation.

Formula (1.3) is not the only finite difference replacement for (1.1) which involves the points x_r, x_{r+1} only. For, consider the

Taylor expansion of y_{r-1} about x_r.

$$y_{r-1} = y_r^{(0)} - hy_r^{(1)} + \frac{h^2}{2!} y_r^{(2)} - \frac{h^3}{3!} y_r^{(3)} + \cdots$$

Truncating the series after the second term leads to the formula

$$y_r = y_{r-1} + hf_r \quad ,$$

or, equally, $$y_{r+1} = y_r + hf_{r+1} \tag{1.5}$$

which constitutes an alternative finite difference replacement for (1.1), with a principal truncation error

$$(TE)_1 = +\tfrac{1}{2} h^2 y^{(2)} \quad .$$

Although formulae (1.3) and (1.5) are of comparable accuracy, the latter is, in practice, more difficult to apply, since at each step the unknown y_{r+1} appears on both sides of the equation, necessitating the solution, possibly by an iterative method, of an algebraic or transcendental equation. Formulae with this property will be termed <u>implicit</u>, while those which, like (1.3), yield the new value of y directly will be termed <u>explicit</u>. For the pair of formulae at present under discussion there is, of course, no reason for not preferring the easier explicit formula. In general, however, the most accurate available implicit formula involving a specified number of points turns out to be more accurate than the best available explicit formula. In addition, stability requirements (see Chapter III) are less stringent for implicit formulae than for explicit.

I.3 Finite Difference Replacements of Higher Order.

Formulae (1.3) and (1.5) are, of course, of very low accuracy, and yield acceptable results only if the mesh length is excessively small. (Even then, accumulation of round-off error can set a limit to the accuracy attainable.) An obvious way to increase the accuracy is to base the finite difference replacement for (1.1) on more than two consecutive mesh points. A formula which involves values of the unknown y at the k+1 neighbouring points $x_r, x_{r+1}, \ldots x_{r+k}$ is said to be a formula of order k. Clearly, a step-by-step numerical solution of (1.1) based on a k-th finite difference formula cannot proceed unless an additional k-1 values $y_1, y_2, \ldots y_{k-1}$ are known. The question of obtaining these additional starting values is discussed in Chapter II, section 7.

Let us consider possible ways of forming a finite difference replacement of second order for (1.1). Clearly, a possible method is to proceed exactly as in the derivation of (1.3), but with the finite difference replacement for the derivative $y^{(1)}$ based on three points rather than two. This yields the explicit formula

$$y_{r+2} = y_r + 2hf_{r+1} \qquad (1.6)$$

with $$(TE)_1 = -\frac{2h^3}{3!} y^{(3)}$$

The truncation error is an order higher than that of (1.3) or (1.5).

An alternative second order formula can be formed by considering

the equation

$$y_{r+2} = y_{r+1} + \int_{x_{r+1}}^{x_{r+2}} y' dx \qquad (1.7)$$

The definite integral is evaluated as accurately as possible, bearing in mind that y' $(= f(x, y))$ is available only at the three given mesh points. Accordingly we calculate the unique polynomial of degree two passing through the three points (x_r, f_r), (x_{r+1}, f_{r+1}), (x_{r+2}, f_{r+2}) in the x-f plane. Integrating this polynomial over the range $[x_{r+1}, x_{r+2}]$ and substituting in (1.7) leads to the formula

$$y_{r+2} = y_{r+1} + \frac{h}{12}(5f_{r+2} + 8f_{r+1} - f_r) \qquad (1.8)$$

with $$(TE)_1 = +\frac{h^4}{24} y^{(4)}$$

This is the well known second order implicit Adam's formula, and its principal truncation error is an order higher than that of (1.6).

This result can also be improved upon by considering the equation

$$y_{r+2} = y_r + \int_{x_r}^{x_{r+2}} y' \, dx,$$

and proceeding as before to obtain the formula

$$y_{r+2} = y_r + \frac{h}{3}(f_{r+2} + 4f_{r+1} + f_r) \qquad (1.9)$$

with $$(TE)_1 = +\frac{h^5}{90} y^{(5)} .$$

This, the celebrated Simpson's Formula, turns out to be the most accurate second order formula obtainable.

This rather haphazard derivation of different formulae of the same order is unsatisfactory. It is not obvious at the outset that formula (1.9) will be more accurate than (1.8). It is even less obvious that formula (1.9) is the best second order formula available. If we go on to consider higher order formulae involving four or more mesh points, the situation is more confused, and the most accurate formulae are not always known. One of the objects of this thesis will be to systematically derive the formulae, of given order, which have the greatest accuracy (henceforth called optimum formulae).

Among the devices commonly used to produce finite difference replacements for (1.1) are calculations involving the difference operators Δ , ∇ and δ . Throughout this thesis we shall steadfastly avoid the use of these operators, since it is felt that they hinder the process of systematisation by disguising familiar formulae in a variety of garbs.

The most general finite difference formula of order k will be a relationship of the type

$$\sum_{t=0}^{k} a_{0t}\, y_{r+t} \;+\; h \sum_{t=0}^{k} a_{1t}\, y^{(1)}_{r+t} \;=\; 0 \qquad (1.10)$$

This expression will be our starting point for the project of deriving the optimum k-th order formula. The method for calculating the optimum values of the undetermined multipliers a_{0t} , a_{1t}, ($t = 0,1,\ldots k$) will be discussed in Chapter II.

I.4 Finite Difference Formulae Involving Higher Derivatives.

In practice, formulae of the type (1.10) operate by virtue of the fact that $y^{(1)}$ is known as a function of x and y from the differential equation (1.1) ; but so also is $y^{(2)} = f_x + f_y y^{(1)}$ known as a function of x and y , as are $y^{(3)}$, $y^{(4)}$ etc. Thus a novel method of incorporating in the difference formula more information from the differential equation suggests itself, namely that we replace (1.10) by the wider calss of finite difference formulae,

$$\sum_{t=0}^{k} a_{0t}\, y_{r+t} + h \sum_{t=0}^{k} a_{1t}\, y^{(1)}_{r+t} + h^2 \sum_{t=0}^{k} a_{2t}\, y^{(2)}_{r+t} + \cdots$$

$$\cdots + h^{\ell} \sum_{t=0}^{k} a_{\ell t}\, y^{(\ell)}_{r+t} = 0 \qquad (1.11)$$

Chapter II will be devoted to the determination and study of formulae of the class (1.11). Our original plan of deriving systematically the optimum formulae, which involve first derivatives only, will thus be included in the wider project, as the special case $\ell = 1$.

The inclusion of the higher derivatives in the difference formulae, not surprisingly, effects considerable increases in accuracy. (Since the additional information is basically analytical information concerning the original differential equation (1.1), the new formulae of improved accuracy turn out not merely to be tantamount to higher order formulae of class (1.10).) Moreover, when we consider, in Chapter III, the problem of stability, it transpires that the new formulae have much superior characteristics.

The same device of including higher derivatives can produce new formulae of improved accuracy applicable to the problem of quadrature ; this is the subject of Chapter IV.

The only branch of numerical analysis in which the idea of using the higher derivatives of a known function appears to have been studied in any detail is the theory of interpolation. The labour involved in the high accuracy interpolation necessary for the construction of a tabulation of some of the transcendental functions is much diminished when a formula is used which involves the values of the derivative as well as of the function. Such formulae have been considered by Fort [3] and Salzer [4] .

The possibility of using higher derivatives in quadrature formulae has received scant attention. Squires [5] has produced a class of quadrature formulae involving derivatives of the integrand, but restricts his investigation to first order formulae only. Hammer and Wicke [6] have produced quadrature formulae of the Gaussian type, in which derivatives of the integrand are involved, but the use of special irregularly spaced abscissae makes the methods extremely cumbersome when high accuracy is sought.

With the exception of the formula known as Milne's starting procedure,[7] the numerical solution of ordinary differential equations by finite difference formulae involving higher derivatives has, to the author's knowledge, never been investigated.

CHAPTER II.

HIGH ACCURACY FORMULAE FOR THE SOLUTION OF ORDINARY DIFFERENTIAL EQUATIONS.

II.1 Introduction.

Consider the differential equation

$$y' = f(x, y) \qquad (2.1)$$

subject to the boundary condition $y = y_0$ at $x = x_0$. Along the x-axis, the mesh points x_r are given by

$$x_r = x_0 + rh, \qquad r = 0, 1, 2, \ldots ,$$

h being the distance between consecutive points. In order to obtain a numerical solution of (2.1), we shall consider the class of difference equations

$$\sum_{s=0}^{\ell} \sum_{t=0}^{k} a_{st} h^s y_{n+t}^{(s)} = 0, \quad k \geqslant 1, \quad \ell \geqslant 1 \qquad (2.2)$$

An equation of this class is a relationship involving the values of y_r ($\equiv y_r^{(0)}$) and its successive derivatives $y_r^{(s)}$ ($\equiv (\frac{d^s y}{dx^s})_{x=x_r}$) at the $k+1$ neighbouring points $x_n, x_{n+1}, \ldots x_{n+k}$.

Provided that the function $f(x, y)$ can be differentiated partially as often as may be required, the derivatives $y^{(s)}$ can be obtained in terms of x and y from the differential equation (2.1). Thus

$$\begin{aligned} y^{(1)} &= f(x, y) = F_1(x, y) \\ y^{(2)} &= (F_1)_x + (F_1)_y F_1 = F_2(x, y) \\ y^{(3)} &= (F_2)_x + (F_2)_y F_2 = F_3(x, y) \end{aligned} \qquad (2.3)$$

$$y^{(m)} = (F_{m-1})_x + (F_{m-1})_y F_1 = F_m(x, y)$$

where $(F_j)_x \equiv \frac{\partial F_j}{\partial x}$, $(F_j)_y \equiv \frac{\partial F_j}{\partial y}$

Let us assume that the coefficients a_{st} have been determined, and that the values of y are known at the k neighbouring points $x_0, x_1, x_2, \ldots x_{k-1}$. (See Chapter II section 7, for methods for obtaining additional starting values.) Then an equation of class (2.2) with $n=0$ can be used to find a value for y_k. Applied again with $n=1$, it yields a value for y_{k+1}, and, continuing in this manner, a complete step-by-step numerical solution of the difference equation (2.2) is obtained. In general, it will be necessary to solve, at each step, an algebraic or transcendental equation for y_{n+k}; that is, the formula is <u>implicit</u>, in the sense defined in Chapter I. If, however, a set of coefficients a_{st} can be found with

$$a_{sk} = 0, \quad s = 1, 2, 3, \ldots . l .,$$

then y_{n+k} will be given directly at each step, and the formula is <u>explicit</u>.

II.2 **An Illustration : Second Order Formulae Involving No Higher Derivatives.**

Before discussing the general problem of determining possible values for the coefficients a_{st}, we shall illustrate the method to be employed by considering the specific example given by putting $k = 2$ and $\ell = 1$ in (2.2). This gives a class of second order finite difference formulae for the solution of (2.1), in which no higher derivatives are employed. It can therefore be expected that some of the well known second order formulae, mentioned in Chapter I, will emerge. With $k = 2$ and $\ell = 1$, (2.2) becomes

$$a_{00}y_n + a_{01}y_{n+1} + a_{02}y_{n+2} + h(a_{10}y_n^{(1)} + a_{11}y_{n+1}^{(1)} + a_{12}y_{n+2}^{(1)}) = 0 \qquad (2.4.)$$

Clearly, the a_{st} are arbitrary to the extent of a constant factor, and we can therefore, without loss of generality, put $a_{02} = -1$. We now expand $y_{n+t}^{(s)}$, $(s = 0,1;\ t = 1,2)$ as Taylor series about the point x_n.

$$y_{n+1}^{(0)} = y_n^{(0)} + hy_n^{(1)} + \frac{h^2}{2!}y_n^{(2)} + \ldots + \frac{h^m}{m!}y_n^{(m)} + \ldots$$

$$y_{n+2}^{(0)} = y_n^{(0)} + 2hy_n^{(1)} + \frac{(2h)^2}{2!}y_n^{(2)} + \ldots + \frac{(2h)^m}{m!}y_n^{(m)} + \ldots$$

$$y_{n+1}^{(1)} = y_n^{(1)} + hy_n^{(2)} + \frac{h^2}{2!}y_n^{(3)} + \ldots + \frac{h^m}{m!}y_n^{(m+1)} + \ldots$$

$$y_{n+2}^{(1)} = y_n^{(1)} + 2hy_n^{(2)} + \frac{(2h)^2}{2!}y_n^{(3)} + \ldots + \frac{(2h)^m}{m!}y_n^{(m+1)} + \ldots$$

When these expressions are substituted into (2.4), the following equation is obtained.

$$A_0 y_n^{(0)} + A_1 h y_n^{(1)} + A_2 h^2 y_n^{(2)} + \ldots + A_m h^m y_n^{(m)} + \ldots = 0 \qquad (2.5)$$

The first few coefficients A_m are as follows

$$A_0 = a_{00} + a_{01} - 1 \qquad (2.5(i))$$

$$A_1 = a_{01} - 2 + a_{10} + a_{11} + a_{12} \qquad (2.5(ii))$$

$$A_2 = a_{01}/2! - 2^2/2! + a_{11} + 2a_{12} \qquad (2.5(iii))$$

$$A_3 = a_{01}/3! - 2^3/3! + a_{11}/2! + 2^2 a_{12}/2! \qquad (2.5(iv))$$

$$A_4 = a_{01}/4! - 2^4/4! + a_{11}/3! + 2^3 a_{12}/3! \qquad (2.5(v))$$

$$A_5 = a_{01}/5! - 2^5/5! + a_{11}/4! + 2^4 a_{12}/4! \qquad (2.5(vi))$$

If the values of the coefficients a_{st} could be chosen in such a way that each of the coefficients A_m in (2.5) vanished identically, then (2.4), with $y^{(1)}$ replaced by $f(x, y)$, would represent the differential equation (2.1) exactly. With a fixed number of undetermined multipliers a_{st} available, this, of course, cannot in general be done. The best that can be achieved is to make as many as possible of the leading coefficients $A_0, A_1, A_2 \ldots$ identically zero. The remaining non-zero terms on the left hand side of (2.5) now constitute the truncation error, as defined in (1.4). The first of these non-zero terms is then $(TE)_1$, the principal truncation error of the formula (2.4).

It is one of the advantages of this type of derivation that the

leading terms of the truncation error series are readily obtained as an integral part of the calculation. Indeed, the general term of the truncation error series for a particular formula can be obtained with little difficulty, allowing a detailed examination, for any given differential equation, of the convergence of the solution of the finite difference equation to the solution of the differential equation.

Since, for the formula at present under consideration, there are five undetermined multipliers a_{st}, it should be possible to put

$$A_0 = A_1 = A_2 = A_3 = A_4 = 0,$$

and solve the resulting equations (2.5(i)) - (2.5(v)) to determine the coefficients a_{st} uniquely. With these values for a_{st}, (2.4) now determines the second order finite difference replacement of maximum accuracy for (2.1). Such a formula will be called an <u>optimum formula</u>. Its principal truncation error is

$$(TE)_1 = A_5 h^5 y^{(5)},$$

where A_5 can now be evaluated from (2.5(vi)).

In all step-by-step numerical solutions for initial value problems, the question of possible instability must be taken into account. A detailed examination of this phenomenon is made in the next Chapter, where it is shown that the stability of any particular formula is largely governed by the numerical values of the coefficients a_{st} appearing in it. Thus, it may happen that the optimum formula obtained

above turns out to be unstable, and must be rejected. Thus, instead of evaluating the optimum coefficients it would be wiser to introduce some degree of flexibility into the formula by retaining one of the coefficients a_{st} as a variable parameter, the eventual value of which would be dictated by stability requirements. Since the coefficients a_{1t} in (2.4) are all premultiplied by h, and therefore make a second order contribution to the stability characteristics of the formula, one of the coefficients a_{0t}, namely a_{00}, is retained as a parameter. Hence we put $a_{00} = a$, and solve for the four remaining undetermined coefficients a_{st} in terms of a, by setting

$$A_0 = A_1 = A_2 = A_3 = 0 ,$$

and solving the four resulting equations (2.5(i))-(2.5(iv)). The two leading terms in the truncation error series for the formula (2.4), now defined by the new values of a_{st} are

$$(TE)_1 = A_4 h^4 y^{(4)}$$

and

$$(TE)_2 = A_5 h^5 y^{(5)} ,$$

where A_4 and A_5 can be evaluated in terms of a. It should be observed that the optimum formula can quickly be recovered by putting $(TE)_1 = 0$, and solving for a, which now has its optimum value.

If this procedure is followed out for the formula (2.4), the following results are obtained

$$a_{00} = a \qquad a_{01} = 1-a \qquad a_{02} = -1$$
$$a_{10} = \frac{5a-1}{12} \qquad a_{11} = \frac{2}{3}(1+a) \qquad a_{12} = \frac{5-a}{12} \tag{2.6}$$
$$A_4 = \frac{1-a}{24} \qquad A_5 = \frac{17-13a}{360}$$

Equation (2.4) now becomes

$$y_{n+2} = ay_n + (1-a)y_{n+1} + \frac{h}{12}\left((5a-1)y_n^{(1)} + 8(1+a)y_{n+1}^{(1)} + (5-a)y_{n+2}^{(1)}\right), \tag{2.7}$$

with $$(TE)_1 = \frac{1-a}{24} h^4 y^{(4)} ,$$

and $$(TE)_2 = \frac{17-13a}{360} h^5 y^{(5)} \tag{2.8}$$

The optimum formula is obtained by putting $(TE)_1 = 0$, that is $a = 1$. With this value for the parameter a, (2.7) reduces to

$$y_{n+2} = y_n + \frac{h}{3}\left(y_n^{(1)} + 4y_{n+1}^{(1)} + y_{n+2}^{(1)}\right) \tag{2.9}$$

On putting $y^{(1)} = f(x, y)$, we see that this formula is identical with (1.9), Simpson's formula. Its principal truncation error, obtained by putting $a=1$ in (2.8) is $\frac{1}{90} h^5 y^{(5)}$.

If we choose $a = 0$, then (2.7) reduces to

$$y_{n+2} = y_{n+1} + \frac{h}{12}\left(-y_n^{(1)} + 8y_{n+1}^{(1)} + 5y_{n+2}^{(1)} \right) \tag{2.10}$$

with $$(TE)_1 = \frac{1}{24} h^4 y^{(4)} .$$

This is identically (1.8), the second order implicit Adam's Formula

Another simple choice for a is -1. With this value, (2.7) becomes

$$y_{n+2} = -y_n + 2y_{n+1} + \tfrac{1}{3}h(-y_n^{(1)} + y_{n+2}^{(1)}) \qquad (2.11)$$

with $$(TE)_1 = \frac{1}{12} h^4 y^{(4)} .$$

This formula does not appear to be particularly well known, although its stability characteristics are, in some respects, superior to those of either Simpson's or Adams' formula.

There would be an obvious advantage in choosing $a = 5$, since (2.7) would then become an explicit formula. Unfortunately, as will be demonstrated in Chapter III, it is essential from considerations of stability that $-1 \leq a \leq +1$.

Explicit formulae can be obtained, at the cost of an increased truncation error, by stipulating in advance that $a_{12} = 0$. If $a_{00} \equiv a$ is retained as a parameter, then the number of undetermined coefficients a_{st} is reduced from four to three. Consequently, we can demand only that A_0, A_1 and A_2 vanish. Thereupon solving (2.5(i)) - (2.5(iii)) for the three undetermined coefficients a_{st} leads to the following formula

$$y_{n+2} = ay_n + (1-a)y_{n+1} + \tfrac{1}{3}h \left[(a-1)y_n^{(1)} + (a+3)y_{n+1}^{(1)} \right] \qquad (2.12)$$

with $$(TE)_1 = \frac{a-5}{12} h^3 y^{(3)}$$

and $$(TE)_2 = \frac{a-9}{24} h^4 y^{(4)}$$

The optimum explicit formula corresponds to a = 5 (and is identically formula (2.7) with a = 5) but, as has already been stated, this is unacceptable from the standpoint of stability.

It is of interest to observe that if a = +1, (2.12) reduces to

$$y_{n+2} = y_n + 2hy'_{n+1}$$

with $$(TE)_1 = -\tfrac{1}{3} h^3 y^{(3)}.$$

This is identical with (1.6), one of the elementary finite difference formulae discussed in Chapter I.

II.3 The Choice of Origin for the Taylor Expansions.

Consider equation (2.4) of the preceding section,

$$a_{00}y_n + a_{01}y_{n+1} + a_{02}y_{n+2} + h(a_{10}y_n^{(1)} + a_{11}y_{n+1}^{(1)} + a_{12}y_{n+2}^{(1)}) = 0 \quad (2.4)$$

and choose a_{02} to be -1. We now expand $y_{n+t}^{(s)}$, $(s = 0, 1 ; t = 0, 1, 2)$ as Taylor Series not, as previously, about x_n, but about an arbitrary mesh point x_q, where $q = n - p$, p being an integer, positive or negative.

$$y_n^{(0)} = y_q^{(0)} + phy_q^{(1)} + \frac{(ph)^2}{2!}y_q^{(2)} + \ldots + \frac{(ph)^m}{m!}y_q^{(m)} + \ldots$$

$$y_{n+1}^{(0)} = y_q^{(0)} + (p+1)hy_q^{(1)} + \frac{(p+1)^2h^2}{2!}y_q^{(2)} + \ldots + \frac{(p+1)^mh^m}{m!}y_q^{(m)} + \ldots$$

$$y_{n+2}^{(0)} = y_q^{(0)} + (p+2)hy_q^{(1)} + \frac{(p+2)^2h^2}{2!}y_q^{(2)} + \ldots + \frac{(p+2)^mh^m}{m!}y_q^{(m)} + \ldots$$

$$y_n^{(1)} = y_q^{(1)} + phy_q^{(2)} + \frac{(ph)^2}{2!}y_q^{(3)} + \ldots + \frac{(ph)^m}{m!}y_q^{(m+1)} + \ldots$$

$$y_{n+1}^{(1)} = y_q^{(1)} + (p+1)hy_q^{(2)} + \frac{(p+1)^2h^2}{2!}y_q^{(3)} + \ldots + \frac{(p+1)^mh^m}{m!}y_q^{(m+1)} + \ldots$$

$$y_{n+2}^{(1)} = y_q^{(1)} + (p+2)hy_q^{(2)} + \frac{(p+2)^2h^2}{2!}y_q^{(3)} + \ldots + \frac{(p+2)^mh^m}{m!}y_q^{(m+1)} + \ldots$$

If these expansions are substituted into (2.4), the following equation is obtained.

$$B_0 y_q^{(0)} + B_1 hy_q^{(1)} + B_2 h^2y_q^{(2)} + \ldots + B_m h^m y_q^{(m)} + \ldots = 0 \quad (2.13)$$

The first few coefficients B_m are as follows.

$$B_0 = a_{00} + a_{01} - 1 \qquad (2.13(i))$$

$$B_1 = pa_{00} + (p+1)a_{01} - (p+2) + a_{10} + a_{11} + a_{12} \qquad (2.13(ii))$$

$$B_2 = \frac{p^2}{2!} a_{00} + \frac{(p+1)^2}{2!} a_{01} - \frac{(p+2)^2}{2!} + pa_{10} + (p+1)a_{11} + (p+2)a_{12} \qquad (2.13(iii))$$

$$B_3 = \frac{p^3}{3!} a_{00} + \frac{(p+1)^3}{3!} a_{01} - \frac{(p+2)^3}{3!} + \frac{p^2}{2!} a_{10} + \frac{(p+1)^2}{2!} a_{11} + \frac{(p+2)^2}{2!} a_{12} \qquad (2.13(iv))$$

$$B_4 = \frac{p^4}{4!} a_{00} + \frac{(p+1)^4}{4!} a_{01} - \frac{(p+2)^4}{4!} + \frac{p^3}{3!} a_{10} + \frac{(p+1)^3}{3!} a_{11} + \frac{(p+2)^3}{3!} a_{12} \qquad (2.13(v))$$

$$B_5 = \frac{p^5}{5!} a_{00} + \frac{(p+1)^5}{5!} a_{01} - \frac{(p+2)^5}{5!} + \frac{p^4}{4!} a_{10} + \frac{(p+1)^4}{4!} a_{11} + \frac{(p+2)^4}{4!} a_{12} \qquad (2.13(vi))$$

Equations (2.13(i)) - (2.13(vi)) are compared with (2.5(i)) - (2.5(vi)), and, after some manipulation, the following relationship between the A_m and the B_m are established.

$$\begin{aligned} B_0 &= A_0 \\ B_1 &= pA_0 + A_1 \\ B_2 &= \frac{p^2}{2!} A_0 + pA_1 + A_2 \\ B_3 &= \frac{p^3}{3!} A_0 + \frac{p^2}{2!} A_1 + pA_2 + A_3 \\ B_4 &= \frac{p4}{4!} A_0 + \frac{p^3}{3!} A_1 + \frac{p^2}{2!} A_2 + pA_3 + A_4 \\ B_5 &= \frac{p^5}{5!} A_0 + \frac{p^4}{4!} A_1 + \frac{p^3}{3!} A_2 + \frac{p^2}{2!} A_3 + pA_4 + A_5 \end{aligned} \qquad (2.14)$$

If we proceed, as in the preceding section, and choose $a_{00} = a$ to be a variable parameter, the remaining four undetermined coefficients a_{st} are found by putting $B_0 = B_1 = B_2 = B_3 = 0$. From (2.14), this implies that $A_0 = A_1 = A_2 = A_3 = 0$, and the values of a_{st} which satisfy this set of equations are those given by (2.6), and the formula (2.7) is again produced.

Thus, the finite difference formula is independent of the choice of origin for the Taylor expansions.

In practice, the origin is therefore chosen in such a way that the labour of obtaining the coefficients is minimised.

Further, it follows from (2.14) that once the a_{st} have been determined, then

$$B_4 = A_4$$

and

$$B_4 = pA_4 + A_5 .$$

Hence, the coefficient in $(TE)_1$ is also independent of the choice of origin, but it is important to notice that the value of $(TE)_1$ does depend on the choice of origin. For, if the expansions are about x_n, then

$$(TE)_1 = A_4 h^4 y_n^{(4)},$$

and if about x_q, then

$$(TE)_1 = A_4 h^4 y_q^{(4)}$$

However, since the formula (2.7) is independent of the choice of origin for the Taylor expansions, it follows that the <u>total</u> truncation error is likewise independent of this choice. Thus, there is no advantage to be gained by choosing x_q in such a way that the derivative $y_q^{(4)}$ is small, or even zero. In such a case, the principal truncation error would simply give a very misleading indication of the magnitude of the total truncation error.

Since B_5 is not identically equal to A_5, the coefficient in $(TE)_2$ is, in general, dependent on the choice of origin. If, however, the formula is optimised by choosing the value of the parameter a in such a way that $(TE)_1$ vanishes (i.e. $A_4 = 0$), then for this optimum formula $A_5 = B_5$, and the coefficient of $(TE)_2$ becomes independent of the choice of origin. For this reason, the truncation error of the formula (2.7) is better quoted as

$$(TE)_1 = \frac{1-a}{24} h^4 y^{(4)}$$

$$(TE)_2 = \frac{1}{90} h^5 y^{(5)} \quad \text{when} \quad a = 1.,$$

since, in this form, both the coefficients quoted are invariant under change of origin.

The proof of the result given in this section can readily be extended to cover higher order formulae, involving higher dreivatives. Thus, the conclusions reached above are pertinent to all formluae of class (2.2), and, consequently, truncation errors will always be quoted in the manner indicated.

II.4 Finite Difference Formulae Involving Higher Derivatives : the [k ; ℓ] Formulae. (Lambert and Mitchell [8])

We now return to the problem of determining the coefficients a_{st} in the general formula of class (2.2). We expand y and its derivatives up to order ℓ at the $k+1$ neighbouring points $x_n, x_{n+1}, \ldots x_{n+k}$ as Taylor series about x_n. That is

$$y_{n+t}^{(s)} = y_n^{(s)} + thy_n^{(s+1)} + \frac{(th)^2}{2!}y_n^{(s+2)} + \ldots + \frac{(th)^m}{m!}y_n^{(s+m)} + \ldots$$

where $t = 0, 1, 2, \ldots k$

$s = 0, 1, 2, \ldots \ell$.

If these expansions are substituted into (2.2), the following equation is obtained

$$C_0y_n + C_1hy_n^{(1)} + C_2h^2y_n^{(2)} + \ldots + C_mh^my_n^{(m)} + \ldots = 0 \quad (2.15)$$

The coefficients C_m are defined as follows.

$$C_m = \sum_{s=0}^{m} \frac{1}{(m-s)!} \sum_{t=0}^{k} t^{m-s} a_{st} \quad \text{for } m = 0, 1, 2, \ldots \ell .$$

$$C_m = \sum_{s=0}^{\ell} \frac{1}{(m-s)!} \sum_{t=0}^{k} t^{m-s} a_{st} \quad \text{for } m = (\ell+1), (\ell+2), \ldots \quad (2.16)$$

It is clear from (2.2) that there is no loss of generality in arbitrarily assigning a value to one of the coefficients a_{st}, and we accordingly choose $a_{0k} = -1$. There therefore remain $(k+1)(\ell+1)$ undetermined coefficients a_{st} . However, as pointed out in section 2 of this

Chapter, it is desirable that certain of these coefficients be retained as variable parameters, for the purpose of controlling the stability characteristics of the final formula. Moreover, it is well known that as the order of the formula is increased, so the chances of instability are enhanced, and, indeed, for $k > 2$, a single variable parameter proves to be inadequate. Hence, in general, $k - 1$ parameters are retained, and these are chosen to be a_{0t}, $t=0, 1, \ldots k-2$. Thus, there finally remain $\ell(k + 1) + 1$ undetermined coefficients a_{st}, and these are found in terms of the variable parameters by putting

$$C_m = 0 , \quad m = 0, 1, 2, \ldots (k+1) \qquad (2.17)$$

When these values of the coefficients a_{st} are substituted into (2.2), the required finite difference formulae are defined. This has been done for the following range of values for k and ℓ.

$k = 1$; $\ell = 1, 2, 3, 4.$

$k = 2$; $\ell = 1, 2, 3.$

$k = 3$; $\ell = 1, 2.$

$k = 4$; $\ell = 1.$

The values of the coefficients a_{st} for this range of formulae are displayed Table 1., where the parameters a_{00}, a_{01} and a_{02} have been replaced by a, b and c respectively. The values of the coefficients have, of course, been calculated exactly, no round-off errors having been introduced.

The leading terms in the truncation error series are

TABLE 1. COEFFICIENTS, IMPLICIT CASE.

k	1				2			3		4
l	1	2	3	4	1	2	3	1	2	1
a_{00}	1	1	1	1	a	a	a	a	a	a
a_{01}	-1	-1	-1	-1	$1-a$	$1-a$	$1-a$	b	b	b
a_{02}					-1	-1	-1	$1-a-b$	$1-a-b$	c
a_{03}								-1	-1	$1-a-b-c$
a_{04}										-1
a_{10}	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{12}(1-5a)$	$\frac{1}{240}(11+101a)$	$-\frac{(421-5669a)}{13,440}$	$\frac{1+8a-b}{24}$	$\frac{397+7136a+2436}{18,144}$	$\frac{-19+243a-8b+11c}{720}$
a_{11}	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}(1+a)$	$\frac{8}{15}(1+a)$	$\frac{64}{105}(1+a)$	$\frac{-5+32a+13b}{24}$	$\frac{89+640a+327b}{672}$	$\frac{53+459a+136b-37c}{360}$
a_{12}					$\frac{1}{12}(5-a)$	$\frac{1}{240}(101+11a)$	$\frac{5669-421a}{13,440}$	$\frac{19+8a+13b}{24}$	$\frac{313+416a+327b}{672}$	$\frac{-11+27a+38b+19c}{30}$
a_{13}								$\frac{9-b}{24}$	$\frac{6893+1640a+2436b}{18,144}$	$\frac{323+189a+136b+173c}{360}$
a_{14}										$\frac{251-27a-8b-19c}{720}$
a_{20}		$\frac{1}{12}$	$\frac{1}{10}$	$\frac{3}{28}$		$\frac{1}{240}(3+13a)$	$-\frac{(47-303a)}{4,440}$		$\frac{163+1376a+93b}{30,240}$	
a_{21}		$-\frac{1}{12}$	$-\frac{1}{10}$	$-\frac{3}{28}$		$\frac{1}{6}(1-a)$	$\frac{1}{8}(1-a)$		$\frac{269-512a+339b}{3,360}$	
a_{22}						$\frac{-1}{240}(13+3a)$	$-\frac{(303-47a)}{4,440}$		$\frac{851-608a-339b}{3,360}$	
a_{23}									$\frac{-1283-256a-93b}{30,240}$	
a_{30}			$\frac{1}{120}$	$\frac{1}{84}$			$-\frac{(41-169a)}{40,320}$			
a_{31}			$\frac{1}{120}$	$\frac{1}{84}$			$\frac{8}{315}(1+a)$			
a_{32}							$\frac{169-41a}{40,320}$			
a_{40}				$\frac{1}{1680}$						
a_{41}				$\frac{-1}{1680}$						

TABLE 2 TRUNCATION ERRORS, IMPLICIT CASE

k	1				2			3		4
ℓ	1	2	3	4	1	2	3	1	2	1
$h^3 y^{(3)}$	$\frac{1}{12}$									
$h^4 y^{(4)}$					$\frac{1}{24}(1-a)$					
$h^5 y^{(5)}$		$\frac{-1}{720}$			$\frac{1}{90}(a=1)$			$\frac{19+8a+11b}{720}$		
$h^6 y^{(6)}$								$\frac{65+16a-33b}{1440}$		$\frac{27-27a-11c}{1440}$
$h^7 y^{(7)}$			$\frac{1}{100,800}$			$\frac{-1-a}{9,450}$		$\frac{3}{1540}\left(a=1, b=\frac{27}{11}\right)$		$\frac{641-513a-206-241c}{15,120}$
$h^8 y^{(8)}$						$\frac{-1}{1,344}(a=-1)$				$\frac{6125-4077a-3206-2125c}{120,960}$
$h^9 y^{(9)}$				$\frac{-1}{25,401,600}$					$\frac{313+416a+1036}{25,401,600}$	$\frac{1}{2625}\left(a=1, b=\frac{32}{5}, c=0\right)$
$h^{10} y^{(10)}$							$\frac{1-a}{14,515,200}$		$\frac{3041+3520a+927b}{152,409,600}$	
$h^{11} y^{(11)}$							$\frac{1}{130,977,000}(b=1)$		$\frac{1}{3,172,400}\left(a=1, b=\frac{-729}{103}\right)$	

$$(TE)_1 = C_{\ell(k+1)+1} \left(h \frac{d}{dx} \right)^{\ell(k+1)+1} y_n$$

$$(TE)_2 = C_{\ell(k+1)+2} \left(h \frac{d}{dx} \right)^{\ell(k+1)+2} y_n \qquad (2.18)$$

$$\cdot \qquad \cdot \qquad \cdot$$

$$(TE)_k = C_{k\ell+k+\ell} \left(h \frac{d}{dx} \right)^{k\ell+k+\ell} y_n .$$

The coefficients, C_m, in $(TE)_1$, $(TE)_2$, . . $(TE)_{k-1}$, (which do not depend on the choice of origin for the Taylor expansions) are quoted in Table 2.

The optimum formula is obtained by finding the values of the variable parameters for which

$$(TE)_1 = (TE)_2 = \cdot \cdot \cdot = (TE)_{k-1} = 0 ,$$

and its principal truncation error is then found by substituting these values into the expression for $(TE)_k$. The coefficient of the resulting principal truncation error (which is independent of the choice of origin for the Taylor expansions) is also quoted in Table 2.

The class of <u>explicit</u> formulae is derived in a similar manner, first stipulating that $a_{sk} = 0$ for $s = 1, 2, \ldots \ell$. There are now fewer coefficients a_{st} to be determined, and if the same parameters are retained, then $(TE)_1$ is of order $\ell k + 1$, and the principal part of the truncation error for the optimum formula is of order $\ell k + k$.

The coefficients a_{st} have been calculated for the class of

k	1				2			3		4
l	1	2	3	4	1	2	3	1	2	1
a_{00}	1	1	1	1	a	a	a	a	a	a
a_{01}	-1	-1	-1	-1	1-a	1-a	1-a	b	b	b
a_{02}					-1	-1	-1	1-a-b	1-a-b	c
a_{03}								-1	-1	1-a-b-c
a_{04}										-1
a_{10}	1	1	1	1	$-\frac{1}{2}(1-a)$	$\frac{1}{2}(3+a)$	$-\frac{1}{2}(13-a)$	$\frac{1}{12}(5+4a-b)$	$\frac{1}{240}(581+112a+11b)$	$\frac{1}{24}(-9+9a+c)$
a_{11}					$\frac{1}{2}(3+a)$	$-\frac{1}{5}(1-a)$	$\frac{1}{2}(15+a)$	$\frac{1}{3}(-4+4a+2b)$	$\frac{1}{15}(38+16a+8b)$	$\frac{1}{24}(37+27a+8b-5c)$
a_{12}								$\frac{1}{12}(23+4a+5b)$	$\frac{1}{240}(-949+112a+101b)$	$\frac{1}{24}(-59+27a+32b+19c)$
a_{13}										$\frac{1}{24}(55+9a+8b+9c)$
a_{20}		$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$		$\frac{1}{12}(7+a)$	$-\frac{1}{10}(29-a)$		$\frac{1}{240}(173+16a+3b)$	
a_{21}						$-\frac{1}{12}(a-17)$	$-\frac{1}{10}(31+a)$		$\frac{1}{6}(27+b)$	
a_{22}									$\frac{1}{240}(637-16a-13b)$	
a_{30}			$\frac{1}{6}$	$\frac{1}{6}$			$-\frac{1}{120}(49-a)$			
a_{31}							$+\frac{1}{120}(111+a)$			
a_{40}				$\frac{1}{24}$						

TABLE 3 COEFFICIENTS, EXPLICIT CASE.

TABLE 4. TRUNCATION ERRORS, EXPLICIT CASE

k	1				2			3		4
i	1	2	3	4	1	2	3	1	2	1
$h^2 y^{(2)}$	$-1/2$									
$h^3 y^{(3)}$		$-1/6$			$\frac{-5+a}{12}$					
$h^4 y^{(4)}$			$-1/24$		$-\frac{1}{6}$ $(a=5)$			$\frac{-9+b}{24}$		
$h^5 y^{(5)}$				$-1/120$		$\frac{-31-a}{720}$		$\frac{-193+4a+17b}{360}$		$\frac{-251+27a+8b+19c}{720}$
$h^6 y^{(6)}$						$-\frac{1}{40}$ $(a=-3)$		$\frac{-1}{20}$ $a=10, b=9$		$\frac{-977+81a+32b+65c}{1440}$
$h^7 y^{(7)}$							$\frac{-209+a}{100,800}$		$\frac{-106-2a-b}{9,450}$	$\frac{-21,559+1431a+568b+124c}{30,240}$
$h^8 y^{(8)}$							$\frac{-1}{2520}$ $(a=209)$		$\frac{-(539+128a+16b)}{604,800}$	$\frac{-1}{70}$ $\left(a=\frac{47}{3}, b=64, c=-36\right)$
$h^9 y^{(9)}$									$\frac{-1}{1680}$ $\left(a=\frac{245}{2}, b=-351\right)$	

explicit formulae, for the same ranges of k and ℓ as in the case of the implicit formulae. Those values are given in Table 3, and the truncation errors in Table 4.

Throughout this thesis, the implicit formula of class (2.2) of order k, involving derivatives up to order ℓ, will be denoted by

$$\left[k\,;\,\ell\right]_{imp}\,,$$

and the corresponding explicit formula by

$$\left[k\,;\,\ell\right]_{exp}\,.$$

II.5 Assessment of the $[k ; \ell]$ formulae.

Not all of the formulae derived in the preceding section are, of course, new. In particular, many of the classical formulae are contained in the sub-class $[k ; 1]$. It has already been pointed out that the optimum form of $[2 ; 1]_{imp}$ coincides with Simpson's formula. The formula $[4 ; 1]_{imp}$, with $a = 1$, $b = -8/19$ and $c = 0$ yields

$$y_{n+4} = y_n - \frac{8}{19}(y_{n+1} - y_{n+3}) + \frac{6h}{19}(y_n^{(1)} + 4y_{n+1}^{(1)} + 4y_{n+3}^{(1)} + y_{n+4}^{(1)}),$$

which is Gambo's Formula. (See Collatz [9] , p.536). Putting each of the variable parameters equal to zero in either an implicit or an explicit $[k ; 1]$ formula gives the general implicit or explicit k-th order Adam's Formula.

Within the sub-class $[k ; 1]$, formulae of improved accuracy can be produced only by increasing the order. In the wider class $[k ; \ell]$, however, higher accuracy can be obtained either by increasing the order (increasing k), or by increasing the number of derivatives of y involved (increasing ℓ). It is one of our purposes to point out that in many cases the latter alternative is preferable.

From the standpoint of the amount of labour involved, much depends on the form of the function $f(x, y)$ appearing in the differential equation (2.1). If this is a complicated function, then it can be argued that the evaluation of the higher derivatives at the mesh points might prove laborious. This is in some degree offset by the fact that

the use of the higher derivatives makes it possible to keep the order of the formula low, thus obviating the need for calculating many additional starting values.

The main argument fo preferring the high derivative formula is, however, concerned with stability. It is well known that if the order of the finite difference formula greatly exceeds that of the differential equation, then the chance of the numerical solution exhibiting instability is considerably increased. It will be shown in Chapter III that, by increasing ℓ rather than k , formulae of high accuracy can be produced, which in no way increase the danger of instability.

Finally, as far as implicit methods are concerned, there is yet another reason for preferring the high derivative formulae. The order of the principal truncation error of the optimum $[k\ ;\ \ell]_{imp}$ formula has been shown to be $k\ell + k + \ell$, and, by the symmetry of this expression it would appear that there is nothing to be gained, in terms of truncation error, by increasing ℓ rather than k. However, it will be seen from Table 2 that the numerical factor in the optimum truncation error diminishes much more rapidly with increasing ℓ than with increasing k . For example, we have

for optimum $[1\ ;\ 4]_{imp}$, $(TE)_1 = -\frac{1}{25,401,600}\ h^9 y^{(9)}$, while

for optimum $[4\ ;\ 1]_{imp}$, $(TE)_1 = +\frac{1}{2,625}\ h^9 y^{(9)}$,

the principal truncation error of the high derivative formula being ten thousand times smaller than that of the high order formula. Indeed

if h is given the not unusual value of 0.1, the principal truncation error for $[1\ ;\ 4]_{imp}$ assumes the rather fantastic order of $10^{-16}\, y^{(9)}$. For explicit formulae, the order of the optimum truncation error is $k\ell + k$, and there is an apparent advantage in increasing k rather than ℓ. However, stability requirements - always more difficult to satisfy in explicit formulae - exclude the formulae with smallest truncation errors, when $k \geqslant 2$.

When implicit $[k\ ;\ \ell]$ formulae are used, the solution for y_{n+k} at each step may be obtained either directly or iteratively. As a refinement of the latter technique, a $[k\ ;\ \ell]_{exp}$ formula can be used in conjunction with a $[k\ ;\ \ell]_{imp}$ to form a predictor-corrector pair. Since some of the derivatives evaluated in the corrector are also used in the predictor, this scheme shows no great increase in computational labour over methods involving a single implicit $[k\ ;\ \ell]$ formula. In this context, it is of interest to consider the well known predictor-corrector pair which constitute Milne's Method. [10]. These are

$$y_{n+4} = y_n + \frac{4h}{3}(2y'_{n+1} - y'_{n+2} + 2y'_{n+3}) \quad ; \quad (TE)_1 = -\frac{28}{90} h^5 y^{(5)} \qquad (2.19(i))$$

$$y_{n+2} = y_n + \frac{h}{3}(y'_n + 4y'_{n+1} + y'_{n+2}) \quad ; \quad (TE)_1 = +\frac{1}{90} h^5 y^{(5)} \qquad (2.19(ii))$$

Referring to Tables 1 and 3, it will be seen that the predictor, (2.19(i)), is a $[4\ ;\ 1]^*_{exp}$, while the corrector, (2.19(ii)), is the optimum $[2\ ;\ 1]_{imp}$. We can readily devise analogous pairs of predictor-corrector formulae, which involve higher derivatives. For example, $[3\ ;\ 2]_{exp}$ with $a = -1$, $b = +1$, provides the predictor formula,

---

* Put $a = 1$, $b = c = 0$.

$$y_{n+3} = -y_n + y_{n+1} + y_{n+2} + 2h(y_n^{(1)} + y_{n+1}^{(1)} - 2y_{n+2}^{(1)}) + \frac{2h^2}{3}(y_n^{(2)} + 7y_{n+1}^{(2)} + 4y_{n+2}^{(2)})$$

with $$(TE)_1 = -\frac{1}{90} h^7 y^{(7)} \qquad (2.20(i))$$

A suitable corrector is afforded by the optimum $\left[2\ ;\ 2\right]_{imp}$ formula,

$$y_{n+2} = -y_n + 2y_{n+1} - \frac{3h}{5}(y_n^{(1)} - y_{n+2}^{(1)}) - \frac{h^2}{24}(y_n^{(2)} - 8y_{n+1}^{(2)} + y_{n+2}^{(2)}),$$

with $$(TE) = -\frac{1}{1344} h^8 y^{(8)} \qquad (2.20(ii))$$

The $\left[k\ ;\ l\right]$ formulae can also be used to find a numerical solution for a system of simultaneous first order ordinary differential equations. As an illustration, consider the following set of three simultaneous differential equations in the three dependent variables y, z and w, with x as independent variable.

$$\begin{aligned} y^{(1)} &= \frac{dy}{dx} = f_1(x, y, z, w) \\ z^{(1)} &= \frac{dz}{dx} = f_2(x, y, z, w) \\ w^{(1)} &= \frac{dw}{dx} = f_3(x, y, z, w) \end{aligned} \qquad (2.21)$$

We first observe that the higher derivatives of the dependent variables with respect to x can still be obtained analytically in terms of x, y, z and w, after the manner of equation (2.3). Thus,

$$\begin{aligned} y^{(2)} &= \frac{d^2y}{dx^2} = (f_1)_x + (f_1)_y f_1 + (f_1)_z f_2 + (f_1)_w f_3 = F_1 \\ y^{(3)} &= \frac{d^3y}{dx^3} = (F_1)_x + (F_1)_y f_1 + (F_1)_z f_2 + (F_1)_w f_3 = F_2 \end{aligned} \qquad (2.22)$$

and so on.

Suppose that the values of y, z and w are known at the k neighbouring points $x_n, x_{n+1}, \ldots x_{n+k-1}$. If one of the $[k;l]_{imp}$ formulae is applied to each equation in (2.21), and the higher derivatives evaluated by means of relations of the type (2.22), then a set of three simultaneous equations in the three unknowns y_{n+k}, z_{n+k} and w_{n+k} is obtained, and can in general be solved. The labour is considerably reduced if $[k;l]_{exp}$ formulae are employed, in which case y_{n+k}, z_{n+k} and w_{n+k} are, in turn, given directly.

It follows that the $[k;l]$ formulae can be used to solve ordinary differential equations of second or higher order. For, consider the general m-th order ordinary differential equation

$$y^{(m)} = f(x, y, y^{(1)}, y^{(2)}, \ldots y^{(m-1)}) \qquad (2.23)$$

Writing $y = Y_1$, (2.23) is equivalent to the following set of m simultaneous first order ordinary differential equations in the m dependent variables $Y_1, Y_2, \ldots Y_m$, x being the independent variable.

$$Y_1^{(1)} = Y_2$$

$$Y_2^{(1)} = Y_3$$

$$\vdots \qquad \vdots$$

$$Y_{m-1}^{(1)} = Y_m$$

$$Y_m^{(1)} = f(x, Y_1, Y_2, \ldots Y_{m-1})$$

When a differential equation of the type (2.23) arises from a physical problem, it is, of course, far more usual for the necessary boundary conditions to be given in the form of the initial values of y, $y^{(1)}$, . . . $y^{(m-1)}$, rather than in the form of the initial values y_0, y_1, . . y_{m-1} In such a case, the method proposed here is particularly convenient, if the set of first order equations is solved by a formula of class $[k\,;\ell]$ of low order, since the initial values of Y_1, Y_2, ... Y_m are at hand, and the labour and inaccuracies in computing additional starting values are minimized. With existing formulae of the type

$$\sum_{t=0}^{k} a_{0t}\, y_{n+t} + \sum_{t=0}^{k} a_{mt}\, h^m\, y^{(m)}_{n+t} = 0$$

for the solution of (2.23), it is necessary to use the given initial values of y and its derivatives to estimate the values of y at the m points x_0, x_1, ... x_{m-1} .

A case of particular importance is the initial value problem

$$y^{(2)} = f(x, y, y^{(1)}) \quad ; \quad y(x_0) = y_0 \ , \ y^{(1)}(x_0) = y_0^{(1)}, \tag{2.24}$$

which can readily be solved with high accuracy using a formula of class $[k\,;\ell]$ on the pair of equations

$$y^{(1)} = \eta \quad ; \quad \eta^{(1)} = f(x, y, \eta)$$

Unfortunately, it is not possible to incorporate higher derivatives in formulae which would give a single step-by-step algorithm for solving problems of the type (2.24). For, on attempting to evaluate the higher derivatives analytically, we find

$$y^{(3)} = f_x + f_y\, y^{(1)} + f_{y^{(1)}}\, f \quad ,$$

and $y^{(1)}$ is not available in an analytical form.

II.6 Numerical Examples.

An example is now given to demonstrate that accurate stable formulae of low order can be used to obtain numerical solutions of high accuracy for ordinary differential equations of first order. The example considered is

$$y^{(1)} = e^{-x} - y , \tag{2.25}$$

subject to the boundary condition $y = 1$ at $x = 0$. The theoretical solution is

$$y = (1 + x)e^{-x} , \tag{2.26}$$

and this is used to check the accuracy of the numerical solution of (2.25). Three separate numerical solutions are computed. In each case, the mesh length h is taken to be 0.1 .

I. Formula $\left[1 ; 3\right]_{imp}$.

From Table 1, it is seen that this formula is

$$y_{n+1} = y_n + \tfrac{1}{2}h(y_n^{(1)} + y_{n+1}^{(1)}) + \frac{h^2}{10}(y_n^{(2)} - y_{n+1}^{(2)}) + \frac{h^3}{120}(y_n^{(3)} + y_{n+1}^{(3)})$$

From Table 2, its principal truncation error is seen to be $\frac{h^7 y^{(7)}}{100,800}$.

No additional starting values are required.

II. Fourth Order Runge-Kutta Method. ([11] p 72.)

A solution by this method is included, since the Runge-Kutta formulae are well established as means of obtaining solutions of high accuracy.

The formula used is

$$y_{n+1} = y_n + \tfrac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

where

$$k_1 = h\, f(x_n, y_n)$$

$$k_2 = h\, f(x_n + \tfrac{1}{2}h,\ y_n + \tfrac{1}{2}k_1)$$

$$k_3 = h\, f(x_n + \tfrac{1}{2}h,\ y_n + \tfrac{1}{2}k_2)$$

$$k_4 = h\, f(x_n + h,\ y_n + k_3)$$

There exist no precise expressions for the principal truncation errors of formulae of this type ; in this case the error is of order h^5. Perhaps a fairer comparison with I would have been made if we had chosen a sixth order Runge-Kutta formula, in which case the truncation error would have been of order h^7. However, due to the fact that an m-th order Runge-Kutta formula requires more than m substitutions into the differential equation when $m > 4$, the higher order formulae become excessively complicated, and their application demands an exorbitant amount of labour. (See, for example, Milne [11]) The formula quoted above is the most accurate Runge-Kutta formula in common usage.

No additional starting values are required.

III Formula $[4 ; 1]_{Imp}$.

It has been pointed out that class $[k ; 1]$ has already been studied in some detail. Urabe [12] has shown that the most satisfactory

formulae of this class are the Adams type formulae, and we therefore choose as our last example the most accurate Adams formula available from Table 1, which is $[4\ ;\ 1]_{imp}$ with $a = b = c = 0$. This gives

$$y_{n+4} = y_{n+3} + \frac{h}{720}(-19y_n^{(1)} + 106y_{n+1}^{(1)} - 264y_{n+2}^{(1)} + 646y_{n+3}^{(1)} + 251y_{n+4}^{(1)}) ,$$

with $$(TE)_1 = -\frac{3}{160} h^6 y^{(6)} .$$

Three additional starting values are required for this method, and, for the purposes of the present comparison, these are taken directly from the theoretical solution of the differential equation.

Table 5 shows the numerical solution of the equation (2.25) for the range $x = 0(0.1)1.0$, by each of the three methods described, and compares the results with the theoretical solution. It is seen that, by method I, the error is always less than 2 in the eleventh decimal place ; by method II, it is less than 2 in the seventh decimal place ; and, by method III, it is less than 3 in the seventh decimal place. The labour of computing by method I is no more than by method III. The labour by method II is considerably greater.

A second example is given to demonstrate that accurate solutions of higher order differential equations are also attainable, using the $[k\ ;\ l]$ formulae. The problem considered is

$$y^{(2)} = -y \quad ; \quad y(0) = 0,\ y^{(1)}(0) = 1 , \qquad (2.27)$$

the theoretical solution being $y = \sin x$. The second order differential equation is replaced by the pair of first order equations,

TABLE 5. SOLUTIONS OF EQUATION (2.25).

	Theoretical.	Formula $[1;3]_{imp}$		4th order Runge-Kutta.		Formula $[4;1]_{imp}$	
x	y	y	Error in 12^{th} place	y	Error in 9^{th} place	y	Error in 9^{th} place.
0	<u>1·000,000,000,000</u>	<u>1·000,000,000,000</u>		<u>1·000,000,000</u>		<u>1·000,000,000</u>	
0·1	0·995,321,159,840	0·995,321,159,845	+5	0·995,321,090	-70	<u>0·995,321,160</u>	
0·2	0·982,476,903,694	0·982,476,903,703	+9	0·982,476,784	-120	<u>0·982,476,904</u>	
0·3	0·963,063,686,886	0·963,063,686,899	+13	0·963,063,535	-152	<u>0·963,063,687</u>	
0·4	0·938,448,064,450	0·938,448,066,465	+15	0·938,447,893	-171	0·938,447,995	-69
0·5	0·909,795,989,569	0·909,795,989,586	+17	0·909,795,810	-180	0·909,795,867	-122
0·6	0·878,098,617,750	0·878,098,617,769	+19	0·878,098,437	-181	0·878,098,453	-165
0·7	0·844,195,016,445	0·844,195,016,465	+20	0·844,196,842	-174	0·844,196,820	-196
0·8	0·808,792,135,411	0·808,792,135,431	+20	0·808,791,971	-164	0·808,791,916	-219
0·9	0·772,482,353,507	0·772,482,353,525	+18	0·772,482,203	-151	0·772,482,118	-235
1·0	0·735,758,882,343	0·735,758,882,361	+18	0·735,758,747	-135	0·735,758,636	-246

STARTING VALUES ARE UNDERLINED.

$$y^{(1)} = \eta \quad ; \quad y(0) = 0$$
$$\eta^{(1)} = -y \quad ; \quad \eta(0) = 1 \qquad (2.28)$$

The formula used is the implicit $[1 ; 4]$, which is

$$y_{n+1} = y_n + \tfrac{1}{2}h(y_n^{(1)} + y_{n+1}^{(1)}) + \tfrac{3h^2}{28}(y_n^{(2)} - y_{n+1}^{(2)}) + \frac{h^3}{84}(y_n^{(3)} + y_{n+1}^{(3)}) + \frac{h^4}{1680}(y_n^{(4)} - y_{n+1}^{(4)}) , \qquad (2.29)$$

with a similar equation for η . From (2.28), we obtain the following expressions for the successive derivatives of y and η .

$$y^{(1)} = +\eta \; ; \; y^{(2)} = -y \; ; \; y^{(3)} = -\eta \; ; \; y^{(4)} = +y$$
$$\eta^{(1)} = -y \; ; \; \eta^{(2)} = -\eta \; ; \; \eta^{(3)} = +y \; ; \; \eta^{(4)} = +\eta$$

Substituting these expressions into the pair of finite difference equations and simplifying, we obtain the algorithms

$$y_{n+1} = \frac{A^2 - B^2}{A^2 + B^2} y_n + \frac{2AB}{A^2 + B^2} \eta_n$$
$$\eta_{n+1} = \frac{A^2 - B^2}{A^2 + B^2} \eta_n - \frac{2AB}{A^2 + B^2} y_n \qquad (2.30)$$

where $A = 1 - \frac{3h^2}{28} + \frac{h^4}{1680} \; ; \; B = \frac{h}{2} - \frac{h^3}{84}$

The values of y_0 and η_0 are known, and no additional starting values are required. The principal truncation error of the finite difference formula employed is

	Theoretical Solution	Numerical Solution		Theoretical Solution	Numerical Solution	
x	$y = \sin x$	y	Error in 8th place	$\eta = \cos x$	η	Error in 8th place
0	0·000,000,00	0·000,000,00		1·000,000 00	1·000,000 00	
1·0	0·841,470 99	0·841,470,96	−3	0·540,302,31	0·540,302,34	+3
2·0	0·909,297,43	0·909,297,46	+3	−0·416,146,84	−0·416,146,77	+7
3·0	0·141,120,01	0·141,120,12	+11	−0·989,992,50	−0·989,992,48	+2
4·0	−0·756,802,50	−0·756,802,40	+10	−0·653,643,62	−0·653,643,74	−12
5·0	−0·958,924,28	−0·958,924,33	−5	+0·283,662,19	+0·283,662,00	−19
6·0	−0·279,415,50	−0·279,415,72	−22	+0·960,170,29	+0·960,170,22	−7
7·0	+0·656,986,60	+0·656,986,40	−20	+0·753,902,25	+0·753,902,43	+18
8·0	+0·989,358,25	+0·989,358,29	+4	−0·145,500,03	−0·145,499,73	+30
9·0	+0·412,118,49	+0·412,118,50	+31	−0·911,130,26	−0·911,130,12	+14
10·0	−0·544,021,11	−0·544,020,79	+32	−0·839,071,53	−0·839,071,73	−20

TABLE 6. SOLUTIONS OF EQUATION (2.27).

$$(TE)_1 = - \frac{h^9 y^{(9)}}{25,401,600} \quad .$$

The numerical factor in this error is so small, that we are able to obtain an accurate solution even if h is given the unusually large value of 1.0. The solution of (2.27) calculated using this value of h is set out in Table 6, for the range $x = 0(1.0)10.0$.

Despite the large mesh length, and the fact that the solution oscillates strongly within the range under consideration, the error is never more than 32 in the eighth decimal place.

It is of interest to note that a calculation based on formula (2.29), with $h = 0.1$, would be capable of evaluating the sine and cosine functions correct to approximately seventeen decimal places.

II.7 Starting Values.

If a finite difference formula of order k is used to obtain a numerical solution of a first order differential equation, then, in addition to the single boundary condition presumed given with the differential equation, $k - 1$ new starting values will, in general, be required. It is important that these additional starting values be obtained with sufficient precision, since the accuracy of the whole computation depends on them.

One method commonly used is to apply a fourth-order Runge-Kutta method ([11] p.72), preferably with a mesh length smaller than that to be used in the main calculation. Formula of this type have the advantage that they, themselves, never require additional starting values.

Another common procedure is to find the additional starting values directly from a Taylor expansion about the point at which the original boundary condition is given. This is, of course, equivalent to using formulae of the class $[1\ ;\ \ell]_{exp}$. However, since accuracy, rather than ease of application is desirable when finding starting values, it would be better to use one of the $[1\ ;\ \ell]_{imp}$ formulae. Indeed, as mentioned in Chapter I, W. E. Milne [7] has derived the following formula, which involves higher derivatives, specifically for the evaluation of starting values.

$$y_1 = y_0 + \tfrac{1}{2}h(y_0^{(1)} + y_1^{(1)}) + \frac{h^2}{10}(y_0^{(2)} - y_1^{(2)}) + \frac{h^3}{120}(y_0^{(3)} + y_1^{(3)}) \quad ,$$

with $$(TE)_1 = \frac{h^7 y^{(7)}}{100,800} .$$

This formula, known as Milne's Starting Procedure, is identically $\left[1 ; 3\right]_{imp}$. Even better accuracy can be obtained by using $\left[1 ; 4\right]_{imp}$, for which

$$(TE)_1 = -\frac{h^9 y^{(9)}}{25,401,600} .$$

Once the first additional starting value has been found, the next is most readily obtained by using a formula such as the optimum $\left[2 ; 4\right]_{imp}$, for which

$$(TE)_1 = \frac{h^{11} y^{(11)}}{130,977,000} .$$

The increase in accuracy over $\left[1 ; 4\right]_{imp}$ is welcome, in that it offsets any tendency for the round-off errors to accumulate, since the formula now uses the first derived starting value, as well as the original boundary value.

An illustration of the use of $\left[k ; l\right]$ formulae for obtaining additional starting values is afforded by the example considered in Table 5. The three additional starting values necessary for the application of the Adams' method, $\left[4 ; 1\right]_{imp}$, which were, in fact, taken from the theoretical solution of the differential equation, might equally well have been taken from the solution by the $\left[1 ; 3\right]_{imp}$ method, since this agrees with the theoretical solution to more decimal places than are required for the Adams' method.

CHAPTER III.

THE STABILITY OF THE [k ; l] FORMULAE.

III. Introduction.

In previous chapters of this thesis, much reference has been made to truncation error. There is, of course, another source of error, which is unavoidable in any numerical process, namely round-off error. Thus, in discussing the numerical solution of an initial value problem, there are, in fact, three distinct solutions to be considered. These are (a) the theoretical solution of the differential system,

(b) the theoretical solution of the difference system, and

(c) the numerical solution of the difference system.

The difference between (a) and (b) is due to truncation error, while that between (b) and (c) is the result of round-off error. In this context, the inaccuracies in the additional starting values, which must be computed when the order of the difference system exceeds that of the differential, can conveniently be regarded as round-off error, since they are propagated in the same manner as round-off error.

It is, of course, normal practice in any numerical calculation to control the influence of round-off error on the eventual result, by retaining an adequate number of guarding figures throughout the computation. However, when we are dealing with an initial value problem, and the numerical solution is obtained from a step-by-step process, it can happen that, despite all precautions, the growth of round-off error is so rapid ~~is so rapid~~ that it swamps the whole computation, rendering the final numerical result totally unreliable. When this happens, we say that the method is exhibiting <u>instability</u> . Whether or not this phenomenon will arise in a given problem depends on the

particular finite difference replacement used, and, to a lesser extent, on the size of the mesh length, and on the form of the original differential equation. Thus, even when we are convinced that the use of a particular finite difference replacement will ensure that (b) is acceptably close to (a), we cannot assert, without further investigation, that the same will be true for (c).

The possibility of instability does not arise whenever a finite difference formula is employed, but only when it is used in a step-by-step process. Indeed, one way to avoid instability when dealing with an initial value problem, is to first convert it to a boundary-value problem, a suggestion originally made by Allen and Severn [13]. (See also the papers by Mitchell and Rutherford [14], and Fox and Mitchell [15].) However, for the purposes of this chapter, it will be assumed that the numerical solution of the differential equation will always be obtained from a step-by-step process.

Interest in the problem of instability was first aroused in 1950 by a paper of Todd's [16]. The differential equation considered by Todd is of second order, whereas the $[k\,;\,\ell]$ formulae developed in Chapter II are designed primarily for use with first order differential equations. Nevertheless, the example is still relevant, since instability can arise in an initial value problem involving any order of differential equation. The equation considered is

$$y^{(2)} = -y, \tag{3.1}$$

with $$y(0) = 0, \quad \text{and} \quad y^{(1)}(0) = 1 \quad ,$$

the theoretical solution being $y = \sin x$. Two finite difference formulae are considered. These are

$$y_{n+2} = 2y_{n+1} - y_n + h^2 y^{(2)}_{n+1} \quad ; \quad (TE)_1 = \frac{1}{12} h^4 y^{(4)} \tag{3.2}$$

$$y_{n+4} = 16y_{n+3} - 30y_{n+2} + 16y_{n+1} - y_n - 12h^2 y^{(2)}_{n+2} \quad ; \quad (TE)_1 = \frac{2}{15} h^6 y^{(6)} \tag{3.3}$$

The mesh length, h, is chosen to be 0.1, and the starting values are taken directly from the theoretical solution of the differential equation. Working to five decimal places, the following results are obtained for y when $x = 0.8$.

Theoretical solution of differential equation	$y = 0.71736$
Numerical solution by (3.2)	$y = 0.71728$
Numerical solution by (3.3)	$y = -2.67357$

Thus, the formula with the better truncation error gives a hopelessly inaccurate answer. The difficulty is not overcome by retaining more guarding figures, for if the numerical solution by (3.3) is repeated, working now to <u>ten</u> decimal figures, the result

$$y(0.8) = 0.71864,22373$$

is obtained. It is of interest to note that this result is still poorer than that obtained from (3.2), working to five places. In any case, if the solution by (3.3) is continued, still working to ten places, the trouble quickly shows itself again, and we get the impossible result $y(1.1) = 4.37411,56871$.

The formula (3.3) is here exhibiting instability, and the advantages of its superior truncation error are totally nullified.

Various aspects of the problem of instability have been studied by a number of different authors, and, indeed, there is not even general agreement on the meaning of the term "stability". Authors such as Dahlquist [17], [18] and Richtmyer [19] define a method to be stable if the solution (c) defined above is sufficiently close to (a) . Others, including Rutishauser [20] , Mitchell and Craggs [21] and Todd [16] have taken the view that stability is concerned only with the growth of round-off error, and these authors require that (c) be sufficiently close to (b) in order that the method be deemed stable. Whether (b) is close to (a) is then regarded as the separate problem of convergence.

As may be inferred from previous remarks, we shall in this thesis adopt the latter attitude. Our reasons for this choice are as follows. It is the major purpose of this thesis to propose certain finite difference techniques of improved accuracy. The $[k\,;\,l]$ formulae, which constitute one class of such techniques, are intended to be used in step-by-step computations, and it is therefore incumbent on us to determine whether such procedure will, in practice, be feasible, or whether the growth of round-off error will prove uncontrollable. We do not, and cannot assert that the $[k\,;\,l]$ formulae will afford, for any differential equation, a numerical solution arbitrarily close to the theoretical solution of the differential equation.

III.2 Stability Analysis.

When a finite difference formula of the class (2.2) is applied to the numerical solution of the initial value problem

$$y^{(1)} = f(x, y) \quad ; \quad y(x_0) = y_0 \tag{2.1}$$

the resulting algorithm is

$$\sum_{t=0}^{k} a_{0t} y_{n+t} + \sum_{s=1}^{l} \sum_{t=0}^{k} a_{st} h^s F_s(x_{n+t}, y_{n+t}) = 0 \quad , \tag{3.4}$$

where the functions $F_m(x, y)$ are defined by equation (2.3). We assume that any additional starting values required have been obtained, so that the values of $y_0, y_1, \ldots y_{k-1}$ are known. With these initial values, equation (3.4) possesses a unique theoretical solution at every subsequent mesh point. We indicate by y_n^* the value of this theoretical solution at the mesh point x_n. The actual numerical value, y_n, which we obtain at x_n as a result of the step-by-step calculation, will not coincide exactly with y_n^*, the difference being attributable to the round-off error in the calculation up to that point. We define this difference to be ϵ_n, where

$$y_n = y_n^* + \epsilon_n \tag{3.5}$$

If we have procured numerical values for y at all points up to and including x_n, then we obtain a numerical value for y_{n+1} by applying the algorithm (3.4) with appropriate values for the suffices. This gives the equation

$$\sum_{t=0}^{k} a_{0t} y_{n-k+l+t} + \sum_{s=1}^{\ell} \sum_{t=0}^{k} a_{st} h^s F_s(x_{n-k+l+t}, y_{n-k+l+t}) = 0 \qquad (3.6)$$

which can be solved to give y_{n+1}. Substituting from (3.5) into (3.6), and setting $n - k + 1 = p$ for convenience, we obtain

$$\sum_{t=0}^{k} a_{0t}(y^*_{p+t} + \epsilon_{p+t}) + \sum_{s=1}^{\ell} \sum_{t=0}^{k} a_{st} h^s F_s(x_{p+t}, y^*_{p+t} + \epsilon_{p+t}) = 0 \qquad (3.7)$$

We now expand $F_s(x_{p+t}, y^*_{p+t} + \epsilon_{p+t})$ as a Taylor series about y^*_{p+t}, and ignore powers of ϵ_{p+t} higher than the first, to get

$$\sum_{t=0}^{k} a_{0t}(y^*_{p+t} + \epsilon_{p+t}) + \sum_{s=1}^{\ell} \sum_{t=0}^{k} a_{st} h^s \left\{ F_s(x_{p+t}, y^*_{p+t}) + \epsilon_{p+t} \frac{\partial F_s}{\partial y}(x_{p+t}, y^*_{p+t}) \right\} \qquad (3.8)$$

$$= 0$$

Since y^*_{p+t} satisfies the difference equation (3.4) exactly, (3.8) reduces to

$$\sum_{t=0}^{k} a_{0t} \epsilon_{p+t} + \sum_{s=1}^{\ell} \sum_{t=0}^{k} a_{st} h^s \frac{\partial F_s}{\partial y}(x_{p+t}, y^*_{p+t}) \epsilon_{p+t} = 0 \qquad (3.9)$$

This is <u>the linearised error equation for the general $[k, \ell]$ formula</u>, and the form of its solution determines how the round-off error will be propagated as the computation continues. Equation (3.9) is a homogeneous linear difference equation of order k, but since its coefficients are variable, it is not, in general, possible to find an analytical solution in closed form. In any event, in a practical problem, we would have no knowledge of the theoretical solution y^*_{p+t} which appears in (3.9), and could only replace it by y_{p+t} which would be available as a <u>numerical</u> value at each mesh point. Thus, the

variable coefficients in (3.9) could not be obtained as continuous analytical functions of x, and this would preclude any possibility of finding an analytical solution in closed form for the full linearized error equation.

Following Mitchell and Craggs [21] and Todd [16], we can, as a first approximation, consider the situation in the limit as h tends to zero. Equation (3.9) now reduces to

$$\sum_{t=0}^{k} a_{0t} \epsilon_{p+t} = 0 , \tag{3.10}$$

a linear difference equation with constant coefficients. A standard solution in closed form is now available. Putting $\epsilon_j = \lambda^j$ in (3.10) leads to the equation

$$\rho(\lambda) = a_{0k}\lambda^k + a_{0k-1}\lambda^{k-1} + \cdots + a_{01}\lambda + a_{00} = 0 \tag{3.11}$$

This equation - the auxiliary equation of (3.10) - will be referred to as the <u>basic characteristic equation</u> of the finite difference formula (3.4). Its roots, which may be real or complex, distinct or multiple, will be denoted by $\lambda_1, \lambda_2 \cdots \lambda_r$. The polynomial $\rho(\lambda)$ can therefore be written in the form

$$\rho(\lambda) = a_{0k}(\lambda-\lambda_1)^{\alpha_1}(\lambda-\lambda_2)^{\alpha_2}\cdots(\lambda-\lambda_r)^{\alpha_r}, \tag{3.12}$$

where $\alpha_1, \alpha_2, \cdots \alpha_r$ are positive integers, and

$$\alpha_1 + \alpha_2 + \cdots + \alpha_r = k .$$

The general solution for (3.10) is then

$$\epsilon_j = \left(\sum_{m=1}^{\alpha_1} A_{1m} j^{m-1}\right)\lambda_1^j + \left(\sum_{m=1}^{\alpha_2} A_{2m} j^{m-1}\right)\lambda_2^j + \cdots + \left(\sum_{m=1}^{\alpha_r} A_{rm} j^{m-1}\right)\lambda_r^j \qquad (3.13$$

where the $A_{1m}, A_{2m}, \cdots A_{rm}$ are constants.

It follows that the conditions for ϵ_j to remain bounded as $j \to \infty$ are that

(i) all distinct roots of (3.11) lie on or within the unit circle,

and (ii) all roots of (3.11) of multiplicity greater than one lie within the unit circle.

The second condition follows, since, for any integer q,

$$\lim_{j\to\infty} j^q |\lambda|^j = 0 \quad , \quad \text{if} \quad |\lambda| < 1.$$

If, however, a root of multiplicity $\alpha > 1$ lies <u>on</u> the unit circle, then

$$\epsilon_j \sim \sum_{m=1}^{\alpha} A_m j^{m-1} \qquad \text{as} \quad j \to \infty \quad ,$$

and ϵ_j will consequently be unbounded.

Despite this last remark, we make the following definition.

<u>A finite difference formula of the type (3.4) exhibits strong instability if any root of the basic characteristic equation (3.11) lies outside the unit circle.</u>

It follows that whether or not a particular finite difference formula of the class (2.2) exhibits strong instability depends only on the coefficients a_{0t}, $t = 0, 1, 2, \cdots k$. It does not

depend on the particular form of the function $f(x, y)$ appearing in the differential equation, or on the size of the mesh length h.

Strongly unstable finite difference formulae will be rejected as being unsuitable for the step-by-step solution of initial value problems. Our reasons for not rejecting a formula whose basic characteristic equation has a multiple root on the unit circle will be given later.

Equation (3.10) was obtained from the linearized error equation (3.9) by putting $h = 0$. As a refinement, we now ignore only h^2 and higher powers of h in (3.9), yielding in place of (3.10) the equation

$$\sum_{t=0}^{k} \left\{ a_{0t} + h a_{1t} \frac{\partial F_i}{\partial y} (x_{p+t}, y^*_{p+t}) \right\} \epsilon_{p+t} = 0 \qquad (3.14)$$

We are, of course, no more able to give a general solution in closed form for this equation than for the full linearized error equation (3.9). Rutishauser [20] has suggested, however, that (3.14) be converted into a linear difference equation with constant coefficients by assigning a constant value, β, to the variable term $\frac{\partial F_i}{\partial y} (x_{p+t}, y^*_{p+t})$. A solution in closed form can then be obtained, and is of the form (3.13) with λ_i replaced by μ_i, where μ_i are the roots of the equation

$$\tau(\mu) = (a_{0k} + h\beta a_{1k})\mu^k + (a_{0k-1} + h\beta a_{1k-1})\mu^{k-1} + \cdots$$

$$\cdots + (a_{01} + h\beta a_{11})\mu + (a_{00} + h\beta a_{10}) = 0 \qquad (3.15)$$

We shall call this equation the <u>secondary characteristic equation</u> of the finite difference formula (3.4).

The only differential equation for which $\frac{\partial F_i}{\partial y}$ is in fact constant, is

$$y^{(1)} = \beta y + g(x) , \quad \text{where } \beta = \text{constant.}$$

Nevertheless, we can give the following heuristic argument in defence of the practice of applying Rutishauser's technique to a general problem. It is assumed that we are dealing only with those finite difference formulae which have been shown to have no strong instability, that is, formulae for which the roots, λ_i, of the basic characteristic equation lie on or within the unit circle. We are particularly interested in those roots, λ_i, which lie <u>on</u> the unit circle. The values of these roots were calculated on the assumption that $h = 0$, yet we must, in practice, compute with a small, but non-zero value of h. We ask the question, "As the value of h moves away from zero to a small positive value, do the roots λ_i which used to lie on the unit circle, move inside or outside the circle?" If the latter is the case, we expect a mildly increasing error ϵ_j; if the former, we expect ϵ_j to remain bounded. Thus it is reasonable to regard the secondary characteristic equation as a perturbed form of the basic characteristic equation, and its roots, μ_i, as perturbed values of the roots λ_i. We do not intend to interpret the solution of (3.14) obtained by setting $\frac{\partial F_i}{\partial y} = \beta$ as an accurate estimate of the error ϵ_j at a general mesh point. We mean to use the secondary characteristic equation only to determine whether the roots μ_i, which are the

perturbed values of those roots λ_i for which $|\lambda_i| = 1$, lie within or without the unit circle. We make the following definition applicable only when the finite difference formula employed has already been shown not to be strongly unstable.

The numerical solution of an algorithm of the type (3.4) exhibits weak instability if any root of the secondary characteristic equation (3.15) lies outside the unit circle.

Clearly, weak instability depends not only on the coefficients in the finite difference formula, but also on the size of the mesh length h, and on the choice of the value assigned to the constant β, a choice influenced by the particular differential equation under consideration. Thus, unlike strong instability, it is not a property simply of the $[k\,;\,l]_\alpha$ formula employed, but of the particular calculation being performed. Calculations which exhibit weak instability need not be rejected out of hand. Provided that the step-by-step calculation is not pursued for an excessive number of steps, the elementary precaution of retaining an adequate number of guarding figures beyond the decimal place in which the truncation error is expected to appear, is an adequate safeguard.

We can now give reasons why the occurrence, on the unit circle, of a multiple root of the basic characteristic equation, was not regarded as an indication of strong instability. When such a root, of multiplicity α, is perturbed, it is usual for it to be replaced

by α distinct roots μ_i of the secondary characteristic equation. If any of these falls outside the unit circle, then weak instability will ensue. It is only as a result of freak algebraic coincidences that any of the perturbed roots μ_i will be multiple, and, in any case, such an occurrence would come to light when a test for weak instability is made. (The multiple root μ_i can then, usually, be separated into distinct roots by choosing another value for h.) It is therefore felt that when formulae whose basic characteristic equations have multiple roots on the unit circle, are applied to practical problems, they will exhibit, at worst, weak instability.

In practice, the method of choosing the value assigned to the constant β is determined by the particular differential equation being tackled. First, the function $\frac{\partial F_1}{\partial y} \equiv \frac{\partial f}{\partial y}$ is constructed from the function $f(x, y)$ appearing in the differential equation (2.1). If $\frac{\partial F_1}{\partial y}$ is a function of x only (in other words, if the differential equation (2.1) is linear) then the value for β can be chosen, and the process investigated for weak instability, before the numerical solution of the problem is instigated. In the majority of such cases, it is sufficient to give β a single value, which is "typical" of the values taken by $\frac{\partial F_1}{\partial y}$ in the range of the problem. Indeed, it is frequently enough to know merely the sign of β. If, however, $\frac{\partial F_1}{\partial y}$ varies rapidly with x, we can regard β as being a parameter in the secondary characteristic equation, and calculate how the roots μ_i vary, as β runs through the range of values assumed by $\frac{\partial F_1}{\partial y}$ in the range of the problem. If $\frac{\partial F_1}{\partial y}$ is a function of y as well as of x,

(non-linear differential equation) then we must calculate the numerical solution for y for a few steps, using the algorithm (3.4), in order to estimate the magnitude of $\frac{\partial f}{\partial y}$. (In practice, the given initial value for y can often be sufficient for this purpose.)

It is of interest to compare the process outlined above with a new method for obtaining a numerical solution for a linear differential equation, put forward in a recent paper by Allen [22]. In this paper, Allen recommends that over a small local range of integration, the variable coefficients should be replaced by constants.

In certain cases, study of the criterion for weak instability can give an indication of the maximum value that can be given to the mesh length h, if weak instability is to be avoided. Finally, it is emphasised once more that the procedures recommended above, in connection with weak instability, should be applied only to problems employing finite difference formulae which have been shown not to be strongly unstable.

Some examples are now given to illustrate a few of the remarks made above.

As an illustration of the mechanism of strong instability, consider the example (3.1) of Todd's, mentioned in the preceding section. Although the differential equation is here of second order, a theory of strong instability can readily be constructed for such an equation, on exactly the lines indicated in the present section.

This leads to the basic characteristic equation

$$\lambda^4 - 16\lambda^3 + 30\lambda^2 - 16\lambda + 1 = 0 \qquad (3.16)$$

for the finite difference formula (3.3). The roots of this equation are $1, 1, 7 - \sqrt{48}, 7 + \sqrt{48}$, the last root being responsible for the strong instability. The observation might be made that, in setting out the numerical results of Todd's example, we have not quoted a theoretical solution for the difference system, and that divergence between the numerical solution and the theoretical solution of the differential equation does not constitute instability, as we have defined it. However, when we observe that the numerical value for y at $x = 0.8$ is -2.67357 or $+0.71864,22373$ according as the rounding-off is done at the fifth or the tenth decimal place, it is clear that the numerical solution is incapable of approximating satisfactorily to the theoretical solution of the difference system, whatever it may be; this is the essence of the phenomenon of strong instability.

The basic characteristic equation of the other finite difference formula (3.2), used in Todd's example is

$$\lambda^2 - 2\lambda + 1 = 0, \qquad (3.17)$$

with roots $\lambda_i = +1, +1$. Despite the double root, no strong instability is observed.

To illustrate the relative innocuity of weak instability, and to vindicate our reasons for not regarding a double root, on the unit circle, of the basic characteristic equation, as being indicative of

anything worse than weak instability, we consider the $[2\ ;\ 1]_{imp}$ formula. Quoting from Table 1 , we see that

$$a_{00} = a \quad ; \quad a_{01} = 1 - a \quad ; \quad a_{02} = -1$$

$$a_{10} = \tfrac{1}{12}(5a - 1) \quad ; \quad a_{11} = \tfrac{2}{3}(1 + a) \quad ; \quad a_{12} = \tfrac{1}{12}(5 - a)$$

We consider two cases (i) $a = +1$, and (ii) $a = -1$.

Testing first for strong instability, we find that the basic characteristic equation is

$$-\lambda^2 + (1-a)\lambda + a = 0 \ , \qquad (3.18)$$

which has roots $+1$ and $-a$. Thus, in neither of the cases considered will the formula exhibit strong instability, but we observe that in case (ii) there is a double root on the unit circle. Examining for weak instability, the secondary characteristic equation is found to be

$$\left[-1 + h\beta\tfrac{5-a}{12}\right]\mu^2 + \left[1-a + h\beta\tfrac{2(1+a)}{3}\right]\mu + \left[a + h\beta\tfrac{5a-1}{12}\right] = 0 \qquad (3.19)$$

<u>Case (i)</u> $a = +1$, Equation (3.19) reduces to

$$\left[-1 + \tfrac{h\beta}{3}\right]\mu^2 + 4\tfrac{h\beta}{3}\mu + \left[1 + \tfrac{h\beta}{3}\right] = 0$$

Hence,

$$\mu = \frac{\tfrac{2}{3}h\beta \pm \sqrt{1 + \tfrac{1}{3}h^2\beta^2}}{1 - \tfrac{1}{3}h\beta}$$

$$\doteq \left[\tfrac{2}{3}h\beta \pm (1 + \tfrac{1}{6}h^2\beta^2)\right] \Big/ (1 - \tfrac{1}{3}h\beta)$$

The roots μ_1 and μ_2 are therefore

$$\mu_1 \doteq \frac{1 + \frac{1}{2}h\beta}{1 - \frac{1}{2}h\beta} \quad ; \quad \mu_2 \quad \frac{-1 + \frac{1}{2}h\beta}{1 - \frac{1}{2}h\beta}$$

If $\beta > 0$, then $\mu_1 > 1$

If $\beta < 0$, then $\mu_2 < -1$

Thus, excluding the trivial case of $\beta = 0$, weak instability is unavoidable with this formula.

Case (ii) $a = -1$. Equation (3.19) reduces to

$$\left(-1 + \frac{h\beta}{2}\right)\mu^2 + 2\mu + \left(-1 - \frac{h\beta}{2}\right) = 0$$

Hence $$\mu = \frac{2 \pm h\beta}{2 - h\beta} \quad .$$

The roots for μ_1 and μ_2 are

$$\mu_1 = 1 \quad ; \quad \mu_2 = \frac{2 + h\beta}{2 - h\beta}$$

The double root, at $+1$, of the basic characteristic equation has been replaced by two distinct roots, one of which remains on the unit circle. From the expression for the other root, it follows that if $\beta > 0$, there is a weak instability, while if $\beta < 0$, there is not.

We now consider some numerical results given by these two formulae. Writing the $[2 ; 1]_{imp}$ formula in the manner of the algorithm (3.4), we obtain

$$-y_{n+2} + (1-a)y_{n+1} + ay_n + \frac{h}{12}\Big[(5-a)F_1(x_{n+2},y_{n+2}) + 8(1+a)F_1(x_{n+1},y_{n+1}) + (5a-1)F_1(x_n,y_n)\Big] = 0 \quad (3.20)$$

We choose $h = 0.1$, and $F_1(x, y) = -y$. This choice of the function F_1 is forced upon us, since, for the purposes of the present demonstration, it is necessary that (3.20) be restricted to a form for which a theoretical solution in closed form can be computed. In Case (i), (3.20) reduces to

$$3.1\, y_{n+2} + 0.4\, y_{n+1} - 2.9\, y_n = 0 , \qquad (3.21)$$

and in Case (ii), to

$$2.1\, y_{n+2} - 4.0\, y_{n+1} + 1.9\, y_n = 0 . \qquad (3.22)$$

The value of $\beta = \partial F/\partial y$ is -1, so we regard (3.21) as having a weak instability, but not so (3.22).

Two initial conditions are required, and these are chosen, quite arbitrarily, to be $y_0 = 0$ and $y_1 = 1$. In particular, no attempt has been made to choose initial conditions which approximate to those of the differential equation $y^{(1)} = -y$. The theoretical solutions of (3.21) and (3.22) do not represent the solution of any differential equation, and in fact are quite distinct from each other. The theoretical and numerical solutions of (3.21) and (3.22) are given in Table 7, the numerical solutions being rounded-off in the third decimal place.

In both cases, the rounded numerical solution is satisfactory. It clearly would have been wasteful to have rejected (3.22) on the grounds that the basic characteristic equation had a double root on

TABLE 7.

	Equation (3.21)		Equation (3.22)	
x	Theoretical Solution	Numerical Solution	Theoretical Solution	Numerical Solution
0	0	<u>0</u>	0	<u>0</u>
0·1	1·0	<u>1·0</u>	1·0	<u>1·0</u>
0·2	−0·129,03	−0·129	1·904,76	1·905
0·3	+0·952,13	+0·952	2·723,36	2·724
0·4	−0·243,57	−0·244	3·463,99	3·465
0·5	+0·922,13	+0·922	4·134,08	4·135
0·6	−0·346,84	−0·347	4·740,36	4·741
0·7	+0·907,39	+0·907	5·288,90	5·289
0·8	−0·441,54	−0·442	5·785,19	5·785
0·9	+0·905,83	+0·906	6·234,22	6·234
1·0	−0·529,94	−0·530	6·640,49	6·640

STARTING VALUES UNDERLINED.

the unit circle. Also, the weak instability in (3.21) has failed to show itself in the numerical results. This is not surprising, since the root outside the unit circle has the value 1.003,870, and this number raised to the power ten is only of the order of 1.4 . Hence, no meteoric increase in round-off error could be expected in ten steps. Such a calculation would need to be continued for a very large number of steps indeed, before the weak instability would constitute a serious threat to the method.

III.3 The Strong Stability of the $[k\ ;\ l]$ Formulae .

It is clear from the considerations of the preceding section that the only stability test, which can be meaningfully applied to a general class of finite difference formulae, is the test for strong instability. Weak instability is a function of the particular calculation being considered, and, moreover, since its presence by no means calls for the rejection of the finite difference formula involved, there is little point in attempting to examine the general class of $[k\ ;\ l]$ formulae in this respect.

It will have been noticed that the higher derivatives of y have played no part in the analysis of stability. Indeed, the various authors mentioned in the two previous sections have dealt exclusively with formulae which belong to the class $[k\ ;\ 1]$. The criterion for strong instability, depending as it does only on the values of the coefficients a_{0t} , is quite unaffected by the values of the coefficients of the higher derivatives. It is true that in the analysis of weak instability, equation (3.14) can be replaced by equations such as

$$\sum_{t=0}^{k}\left[a_{0t} + ha_{1t}\left(\frac{\partial F_1}{\partial y}\right)_{p+t} + h^2a_{2t}\left(\frac{\partial F_2}{\partial y}\right)_{p+t}\right]\epsilon_{p+t} = 0 ,$$

which would be subsequently solved by setting $\frac{\partial F_1}{\partial y}=\beta$ and $\frac{\partial F_2}{\partial y}=\gamma$, β and γ being constants. In effect, this is achieving nothing more than allowing the coefficients of the higher derivatives to influence one's choice of the value for the constant in Rutishauser's

procedure. In practice, this is seldom necessary.

For every formula of the class $[k\,;\,l]$, the basic characteristic equation will have a root $\lambda = +1$. This is a consequence of the fact that

$$\sum_{t=0}^{k} a_{0t} = 0 ,$$

which is the first of the equations (2.17) which define the coefficients a_{st}. This root can be regarded as the "genuine" root, corresponding to the differential equation. For formulae of order one - that is, of the same order as the differential equation - it is the only root. When the order of the difference formula exceeds that of the differential equation, other roots, termed by Rutishauser [20], "eingeschleppten", or "smuggled roots", will arise. If our choice of formulae is restricted to those of class $[k\,;\,1]$, which do not involve higher derivatives, greater accuracy can be achieved only by increasing k, and hence increasing the number of smuggled roots. It becomes increasingly difficult to avoid strong instability without sacrificing accuracy, and the optimum formulae soon become unusable. Thus, the formula $[4\,;\,1]_{imp}$ takes its optimum form when $a = 1$, $b = 32/5$ and $c = 0$. (See Tables 1 and 2.) The roots of the basic characteristic equation are then $+1$, -1, $(-32 \pm 30.4)/10$, and the formula clearly has a strong instability. The difficulty is even more severe with explicit formulae, and, indeed, all the explicit $[k\,;\,l]$ formulae quoted in Table 3 exhibit strong instability if the optimum

values of the parameters are employed.

It is felt to be one of the major advantages possessed by formulae of low order, which involve higher derivatives of y, that they permit high accuracy to be achieved without ~~attend~~ attendant instability difficulties. Thus, the optimum $[2\,;\,3]_{imp}$ formula exhibits no strong instability, yet has a truncation error of

$$\frac{1}{130,977,000}\, h^{11} y^{(11)}.$$

For any formula of the $[k\,;\,\ell]$ class, the number of variable parameters incorporated in the coefficients a_{st}, as displayed in Tables 1 and 3, has been chosen so that it is always possible to assign their values in such a way that the resulting formula does not exhibit strong instability. The conditions that must be satisfied by the coefficients of a polynomial, in order that its zeros should lie on, or within, the unit circle, are not easily laid down in a convenient form, and, hence, no attempt is made to define the ranges in which the parameters must lie in order that the formula avoid strong instability. Instead, following Hull and Luxembourg [23], it is recommended that the first step in choosing an appropriate $[k\,;\,\ell]$ formula for a particular calculation be to specify the positions of the roots of the basic characteristic equation. (Considerations of weak instability, for the problem in hand may help determine this choice.) A simple calculation then yields the values to be taken by the parameters.

CHAPTER IV.

HIGH ACCURACY QUADRATURE FORMULAE.

IV.1 Introduction.

If the equation (2.1) is replaced by the simpler equation

$$y^{(1)} = f(x) \tag{4.1}$$

with $y = y_0$ at $x = x_0$, then the theoretical solution for y at $x = x_p$ is

$$y_p = y_0 + \int_{x_0}^{x_p} f(x)\,dx \tag{4.2}$$

The numerical evaluation of the integral on the right hand side of (4.2) constitutes the problem of quadrature, or, properly, the problem of single quadrature. There are three distinct categories of problem.

(i) f(x) given numerically as a tabulation.

In this case, formulae involving derivatives of y of order higher than the first (that is, involving derivatives of the integrand) cannot be applied, since there is no means of obtaining analytical expressions for such derivatives.

(ii) f(x) given analytically, but a primitive for f(x) is not known.

A well known example is $f(x) = e^{-x^2}$. In such a case, derivatives of the integrand are available as analytical expressions, and quadrature formulae of improved accuracy, which involve derivatives of the integrand can be used.

(iii) f(x) given analytically, and a primitive for f(x) is known.

There is then no call to resort to a numerical evaluation.

Except where specific reservations to the contrary are made, we shall assume that the function $f(x)$ appearing in (4.2) falls into category (ii).

The $[k\,;\,l]$ formulae developed in Chapter II can, of course, be used to afford a step-by-step solution of (4.1) for the independent variable y, and, by (4.2), a numerical estimate for the integral $\int_{x_0}^{x_p} f(x)dx$ is obtained. Such a procedure is, however, not to be recommended, since it needlessly introduces the danger of instability, and the general inaccuracy due to the accumulation of round-off error made at each step, into a problem in which these difficulties can be completely avoided. For, putting $a_{00} = 1$, $a_{0k} = -1$ and $a_{0t} = 0$ $(t = 1, 2, \ldots k-1)$ in equation (2.2), we obtain the following class of difference equations

$$y_k - y_0 = \sum_{s=1}^{l} \sum_{t=0}^{k} a_{st}\, h^s\, y_t^{(s)} \qquad k \geqslant 1,\ l \geqslant 1 \qquad (4.3)$$

(The suffix n in (2.2) has been replaced by 0, since the formula will no longer be used in a step-by-step process.)

The derivatives $y_t^{(s)}$, $(s = 1, 2, \ldots l\ ;\ t = 0, 1, 2, \ldots k)$ can be found as functions of x only, from the differential equation (4.1), and thus if the a_{st} are determined, (4.3) gives, directly, an

estimate for

$$\int_{x_0}^{x_k} f(x)\,dx = y_k - y_0 \qquad (4.4)$$

If $\int_{x_0}^{x_p} f(x)\,dx$ is required, where $p = qk$, (p and q integers), then

$$\int_{x_0}^{x_p} f(x)\,dx = \int_{x_0}^{x_k} f(x)\,dx + \int_{x_k}^{x_{2k}} f(x)\,dx + \ldots + \int_{x_{(q-1)k}}^{x_{qk}} f(x)\,dx \qquad (4.5)$$

and each of the integrals on the right hand side of (4.5) can be evaluated individually by a formula of class (4.3). The process is no longer a step-by-step one, and the round-off error in the evaluation of any one of the integrals on the right hand side of (4.5) does not influence the evaluation of the next. The total error - both truncation and round-off - cannot exceed the sum of the errors in the individual quadratures.

Conversely, it is not recommended that the quadrature formulae developed in this Chapter be applied to a step-by-step solution of the differential equation

$$y^{(1)} = f(x, y)\,,$$

although this is technically possible. There is no advantage in using quadrature formulae in preference to the more accurate $[k\,;\ell]$ formulae of Chapter II. For this reason, we give no discussion of the stability characteristics of the quadrature formulae, when

used in a step-by-step process.

IV.2 Quadrature Formulae Involving Higher Derivatives : the $[k ; \ell]$ Quadrature Formulae. (Lambert and Mitchell [24])

The determination of the coefficients a_{st} appearing in formula (4.3) follows the pattern of the derivation of the coefficients in the $[k ; \ell]$ formulae of Chapter II. Taylor expansions for $y_t^{(s)}$ ($t = 0, 1, \ldots k$; $s = 1, 2, \ldots \ell$) and for y_0 and y_k are substituted into (4.3), yielding the following equation.

$$D_0 y_0 + D_1 h y_0^{(1)} + D_2 h^2 y_0^{(2)} + \ldots + D_m h^m y_0^{(m)} + \ldots = 0 \qquad (4.6)$$

where

$$\begin{aligned}
D_0 &= 0 \\
D_1 &= -k + \sum_{t=0}^{k} a_{1t} \\
D_m &= -\frac{k^m}{m!} + \sum_{s=1}^{m-1} \frac{1}{(m-s)!} \sum_{t=1}^{k} t^{m-s} a_{st} + \sum_{t=0}^{k} a_{mt} \quad ; \; m = 2, 3, \ldots \ell . \\
D_m &= -\frac{k^m}{m!} + \sum_{s=1}^{\ell} \frac{1}{(m-s)!} \sum_{t=1}^{k} t^{m-s} a_{st} \quad ; \; m = \ell+1, \ell+2, \ldots
\end{aligned} \qquad (4.7)$$

Since there is no question of stability to be considered, there is little point in retaining parameters among the a_{st}, and so we put

$$D_m = 0 \quad ; \; m = 1, 2, \ldots (k+1)\ell \quad ,$$

and solve fortthe $(k+1)\ell$ undetermined coefficients a_{st} ($t=0,1, \ldots k$; $s=1,2, \ldots \ell$) With these values for a_{st}, (3.3) now determines the class of optimum quadrature formulae.

The principal truncation error is then

$$(TE)_1 = D_{k+l+1} \left(h \frac{d}{dx}\right)^{k+l+1} y \qquad (4.8)$$

Once again, the values of the coefficients a_{st} and the coefficient in the principal truncation error can be shown to be independent of the choice of origin for the Taylor expansions. In practice, the algebraic labour in solving for the a_{st} is reduced to a minimum, if the expansions are made about the mid-point of the range.

The values of the coefficients a_{st} and the principal truncation errors are quoted in Tables 8, 9 and 10 as follows :

Table 8 :- $l = 1$; $k = 1,2,3,4,5,6,7,8.$

Table 9 :- $l = 2$; $k = 1,2,3,4,5.$

Table 10 :- $l = 3$; $k = 1,2,3,4.$

For values of k greater than those considered, the formulae become excessively unwieldy. Moreover, in certain cases, the coefficients, a_{lt}, of the function values are no longer all positive, a situation which can adversely affect the accumulation of round-off error. (A negative coefficient for a function value has already appeared when $k = 4$, $l = 3$.)

The quadrature formula of the class derived above, which involves derivatives of y up to order l, (that is, derivatives of the integrand up to order $l-1$) at $k+1$ neighbouring points, will be referred to as a [k ; l] quadrature formula. All of the formulae

Coefficients

TABLE 8 Quadrature Formulae : $\ell = 1$

K	1	2	3	4	5	6	7	8
a_{10}	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{3}{8}$	$\frac{14}{45}$	$\frac{95}{288}$	$\frac{41}{140}$	$\frac{5,257}{17,280}$	$\frac{3,956}{14,175}$
a_{11}	$\frac{1}{2}$	$\frac{4}{3}$	$\frac{9}{8}$	$\frac{64}{45}$	$\frac{375}{288}$	$\frac{216}{140}$	$\frac{25,039}{17,280}$	$\frac{23,552}{14,175}$
a_{12}		$\frac{1}{3}$	$\frac{9}{8}$	$\frac{24}{45}$	$\frac{250}{288}$	$\frac{27}{140}$	$\frac{9261}{17,280}$	$-\frac{3,712}{14,175}$
a_{13}			$\frac{3}{8}$	$\frac{64}{45}$	$\frac{250}{288}$	$\frac{272}{140}$	$\frac{20,923}{17,280}$	$\frac{41,984}{14,175}$
a_{14}				$\frac{14}{45}$	$\frac{375}{288}$	$\frac{27}{140}$	$\frac{20,923}{17,280}$	$-\frac{18,160}{14,175}$
a_{15}					$\frac{95}{288}$	$\frac{216}{140}$	$\frac{9,261}{17,280}$	$\frac{41,984}{14,175}$
a_{16}						$\frac{41}{140}$	$\frac{25,039}{17,280}$	$-\frac{3,712}{14,175}$
a_{17}							$\frac{5,257}{17,280}$	$\frac{23,552}{14,175}$
a_{18}								$\frac{3,956}{14,175}$

Truncation Errors

$h^3 y^{(3)}$	$\frac{1}{12}$							
$h^5 y^{(5)}$		$\frac{1}{90}$	$\frac{3}{80}$					
$h^7 y^{(7)}$				$\frac{8}{945}$	$\frac{275}{12,096}$			
$h^9 y^{(9)}$						$\frac{9}{1,400}$	$\frac{8,183}{518,400}$	
$h^{11} y^{(11)}$								$\frac{2,368}{467,775}$

Coefficients

TABLE 9. Quadrature Formulae : $l = 2$.

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}		$\frac{1}{2}$	$\frac{243}{224}$	$\frac{11,232}{8,505}$	$\frac{1,270,000}{912,384}$
a_{13}			$\frac{93}{224}$	$\frac{8,192}{8,505}$	$\frac{1,270,000}{912,384}$
a_{14}				$\frac{3202}{8,505}$	$\frac{691,875}{912,384}$
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

$h^5 y^{(5)}$	$-\frac{1}{720}$				
$h^7 y^{(7)}$		$-\frac{1}{4,725}$			
$h^9 y^{(9)}$			$-\frac{9}{313,600}$		
$h^{11} y^{(11)}$				$-\frac{8}{1,964,655}$	
$h^{13} y^{(13)}$					$-\frac{144,425}{230,150,684,768}$

Coefficients

TABLE 10. QUADRATURE FORMULAE : $l=3$

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}		$\frac{41}{105}$	$\frac{10,935}{9,856}$	$-\frac{1,017,408}{3,648,645}$
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}	$\frac{1}{120}$	$\frac{16}{315}$	$\frac{2,187}{49,280}$	$\frac{120,976}{1,216,215}$
a_{32}		$\frac{1}{315}$	$\frac{2,187}{49,280}$	$-\frac{184,788}{1,216,215}$
a_{33}			$\frac{153}{49,280}$	$\frac{120,976}{1,216,215}$
a_{34}				$\frac{2,408}{1,216,215}$

Truncation Errors

$h^7 y^{(7)}$	$\frac{1}{100,800}$			
$h^{11} y^{(11)}$		$\frac{1}{130,977,000}$		
$h^{15} y^{(15)}$			$\frac{9}{5,637,032,000}$	
$h^{17} y^{(17)}$				$\frac{478}{162,983,603,908,125}$

appearing in Tables 8,9 and 10 are optimum, in the sense that, for the given k and ℓ, the order of the principal truncation error is, in each case, as high as possible.

Reference to Tables 8, 9 and 10 will show that certain of the $[k\,;\,\ell]$ formulae have principal truncation errors of an order higher, by unity, than that predicted by (4.8). This is due to a coincidental algebraic symmetry, which causes the coefficient $D_{k\ell+\ell+1}$ to vanish identically, when the coefficients a_{st} take their optimum values. In such cases

$$D_{k\ell+\ell+2}\ (h\frac{d}{dx})^{k\ell+\ell+2}\ y$$

is considered to be the principal truncation error, since it can be shown, that in these cases only, the coefficient $D_{k\ell+\ell+2}$ is invariant under change of origin of the Taylor expansions. Formulae with the property just described will be said to have <u>extra-predicted</u> truncation errors. From Tables 8, 9 and 10, it will be seen that the following $[k\,;\,\ell]$ quadrature formulae have extra-predicted truncation errors.

$[2\,;\,1]$, $[4\,;\,1]$, $[6\,;\,1]$, $[8\,;\,1]$, $[2\,;\,3]$, $[4\,;\,3]$

Not all of the formulae quoted are, of course, new. For example, the $[2\,;\,1]$ quadrature formula is seen from Table 8 to be

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{3}(f_0 + 4f_1 + f_2)\,,$$

the celebrated <u>Simpson's Rule</u>, which owes at least a part of its

popularity to the fact that it has an extra-predicted truncation error. In fact, the class of $[k\ ;\ 1]$ quadrature formulae given in Table 8 is the well known class of Newton-Cotes formulae ([25] pp71-9) , and is included here for completeness and comparison. The class of $[1\ ;\ l]$ quadrature formulae has already been derived by Squires [5], in 1961, but it is believed that the remaining formulae of Tables 9 and 10 are new.

A method for deriving quadrature formulae of the Gaussian type, in which higher derivatives of the integrand are incorporated, has recently been devised by Hammer and Wicke [6] . As in all formulae of the Gaussian type, the integrand - and its derivatives - must be evaluated at special points unequally spaced within the range of integration. Unfortunately, when high accuracy is sought, the abscissae of these special points must be calculated to an accuracy comparable with that of the formula itself, and this makes the eventual numerical evaluation of the integrand and its derivatives tedious in the extreme, high accuracy interpolation from tables being frequently unavoidable. The evaluation of the coefficients and the abscissae of the special points for the Hammer-Wicke formulae has been carried out by Struble, [26].

IV.3 Numerical Examples.

That substantial improvements in accuracy can be obtained by making use of the higher derivatives of the integrand is illustrated by the following group of numerical estimates for the integral $\int_0^{\pi/2} \sin x \, dx = 1$. In each case, the mesh length is $h = \pi/4$.

Table 11. $\int_0^{\pi/2} \sin x \, dx$.

$[k ; \ell]$ quadrature formula	Result
$[2 ; 1]$ (Simpson's Rule)	1.002,28
$[2 ; 2]$	1.000,026
$[2 ; 3]$	0.999,999,999,7

A result of high accuracy is thus obtained, using values of the integrand and its derivatives at the points 0, $\pi/4$, $\pi/2$ only. This illustrates one of the features of the $[k ; \ell]$ quadrature formulae, namely, that the accuracy obtainable is not limited by the (usually) fixed interval at which values of the integrand and its derivatives are available.

A further example is given, because the sine function, since its derivatives all lie between -1 and $+1$, is rather flattering to all quadrature formulae. This second example is $\int_{-1}^{+1} \frac{dx}{x+2}$, which has a theoretical value 1.098,612,288,668. The numerical estimates for this integral given by various quadrature formulae are given below.

Table 12. $\int_{-1}^{+1} \frac{dx}{x+2}$ (Theoretical value 1.098,612,288,668)

Method	Number of values	h	Order of TE	Result	Error
[2 ; 3] once	8	1	11	1.098,667,854	$+3.6\ 10^{-5}$
[2 ; 3] twice	12	$\frac{1}{2}$	11	1.098,612,522	$+\ 2.3\ 10^{-7}$
Struble, $\ell=3$	3	-	7	1.096,652,562	$-2.0\ 10^{-3}$
Struble, $\ell=3$	5	-	11	1.098,590,615	$-2.2\ 10^{-5}$
[8 ; 1] once	9	$\frac{1}{4}$	10	1.098,616,867	$+4.6\ 10^{-6}$
[8 ; 1] twice	17	1/8	10	1.098,612,304	$+3.5\ 10^{-8}$
[8 ; 1] thrice	25	1/12	10	1.098,612,289,926	$+1.3\ 10^{-9}$

No value for the mesh length h is quoted for Struble's method, since the mesh points are unevenly spaced.

The figure entered in the column headed "Number of values" denotes the total number of values of the function and its derivatives that must be evaluated, and is meant to give some indication of the amount of labour involved. However, in the case of Struble's formulae, the figure is rather misleading, and it is fair to say that the labour in computing the results by these formulae far exceeded that of any other method listed in Table 12. The fact that the mesh length took particularly simple values in the first two methods listed ,(and, with $[k\ ;\ \ell]$ formulae involving higher derivatives, it is nearly always possible to arrange for this to be the case) made the numerical work somewhat easier than in the three Newton-Cotes formulae tested.

IV.4 Sub-Optimum [k ; ℓ] Quadrature Formulae.

At the cost of increasing the truncation error, it is possible, by putting certain of the coefficients a_{st} equal to zero, to derive from (4.3) additional quadrature formulae involving derivatives of the integrand, which are specially suitable for certain types of problem. If p of the coefficients a_{st} are set equal to zero, then only the first $(k+1)\ell - p$ of the coefficients D_m appearing in (4.6) can be equated to zero, and the resulting equations solved to determine the remaining a_{st}. Formulae so obtained will be referred to as sub-optimum [k ; ℓ] quadrature formulae , or simply sub-optimum formulae, since they have no counterpart among formulae for solving differential equations. Three types of sub-optimum formulae will be discussed here.

(a) Sub-optimum formulae of high efficiency.

One aspect in which the problem of quadrature differs form that of solving numerically a first order differential equation, has not yet been mentioned. In the latter problem, it is not in general desirable - and, indeed, not often possible - to obtain a solution of high accuracy by using a formula with a poor truncation error, and compensating by taking the mesh length, h, excessively small. The vast increase in computational labour precludes such a process, and, in any event, the increased accumulation of round-off error by the time a given abscissa is reached can fix a limit to the accuracy that can, in practice, be obtained. This is not so in the problem of quadrature,

where highly accurate results can often be obtained by the repeated application of low accuracy formulae, using very small mesh lengths - hence the popularity of Simpson's Rule. There is no great increase of computational labour, provided that the function values are readily available from tables, and, since the process is not step-by-step, accumulation of round-off error is never severe. Hence, in those problems for which the integrand is tabulated for a very small interval, the case for using formulae involving higher derivatives must rest, not only on high accuracy, but on the ease with which such accuracy is achieved. A new sub-optimum formula, derived with this end in view, is the three-point formula

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{63}\left(31f_0 + 64f_1 + 32f_2\right) + \frac{5h^2}{63}\left(f_0^{(1)} - f_2^{(1)}\right) - \frac{h^4}{945}\left(f_0^{(3)} - f_2^{(3)}\right) \tag{4.9}$$

with $$(TE)_1 = \frac{1}{198,450}\,h^9\,f^{(8)}\,.$$

If this formula is used n times, to cover the range of integration 2nh , the result

$$\int_{x_0}^{x_{2n}} f(x)\,dx = \frac{h}{63}\left(31f_0 + 64f_1 + 62f_2 + 64f_3 + \cdots + 64f_{2n-1} + 31f_{2n}\right) + \frac{5h^2}{63}\left(f_0^{(1)} - f_{2n}^{(1)}\right) - \frac{h^4}{945}\left(f_0^{(3)} - f_{2n}^{(3)}\right) \tag{4.10}$$

is obtained. Compare this with Simpson's Rule applied n times to the same range.

$$\int_{x_0}^{x_{2n}} f(x)\,dx = \frac{h}{3}\left(f_0 + 4f_1 + 2f_2 + 4f_3 + \cdots + 4f_{2n-1} + f_{2n}\right)$$

with $$(TE)_1 = \frac{1}{90} h^5 f^{(4)} \quad .$$

It is seen that, in (4.10), the evaluation of the derivatives at the beginning and end of the range is virtually the only increase in labour compared with Simpson's Rule, for a very considerable increase in accuracy.

If (4.10) is used with a mesh length of 0.1 to evaluate the integral $\int_{-1}^{+1} \frac{dx}{x+2}$ (the second example of the last section) the result 1.098,612,288,785 is obtained. The error is +0.000,000,000,117 , and the method is more accurate than any considered in Table 12. It is of interest to compare this result with that previously obtained using the nine-point Newton-Cotes formula (i.e. $[8\ ;\ 1]$) three times. Both use twenty-five values, and although the truncation error of the Newton-Cotes method is of an order higher than that of the sub-optimum formula, the latter is, nevertheless, the more accurate. This is in part due to the very small numerical factor appearing in the principal truncation error of (4.9). The sub-optimum formula is also much easier to use. (See coefficients of $[8\ ;\ 1]$ in Table 8.)

Another sub-optimum formula with the same property is the five-point formula

$$\int_{x_0}^{x_4} f(x)\,dx = \frac{h}{945}(434f_0 + 1024f_1 + 864f_2 + 1024f_3 + 434f_4) + \frac{4h^2}{63}(f_0^{(1)} - f_4^{(1)}) \qquad (4.11)$$

with $$(TE)_1 = -\frac{16}{99,225} h^9 f^{(8)}$$

The optimum $[1\,;\,2]$ and $[2\,;\,2]$ quadrature formulae also have the property that, when they are applied repeatedly, the derivatives of the integrand need be evaluated only at the extreme ends of the range.

(b) Sub-optimum formulae for use near singularities.

It is, in general, impossible to find any meaningful numerical estimate for $\int_\alpha^\beta f(x)\,dx$, when $f(x)$ has a singularity within the range $[\alpha,\beta]$. The difficulty is usually overcome by isolating a small range which includes the singularity, and using some analytical approximation, valid near the singularity, to afford an estimate of the integral over that range. A numerical estimate of the integral over the remaining adjoining ranges is then obtained, in the usual way, from quadrature formulae. As an example, consider the problem of evaluating

$$\int_0^A \frac{\log(1+x)}{x^{3/2}}\,dx\ ,$$

the integrand having a singularity at the origin. We split the range of integration into the two ranges $[0\,,\,a]$ and $[a\,,\,A]$. For small x,

$$\frac{\log(1+x)}{x} = x^{-\frac{1}{2}} - \tfrac{1}{2}x^{\frac{1}{2}} + \tfrac{1}{3}x^{3/2} - \ldots$$

$$\int_0^a \frac{\log(1+x)}{x}\,dx = 2a^{\frac{1}{2}} - \tfrac{1}{3}a^{3/2} + \frac{2}{15}a^{5/2} - \ldots \qquad (4.12)$$

The choice for a is now made in such a way that the error, in truncating the series (4.12) at a convenient point, is comparable

with the truncation error in evaluating $\int_a^A \frac{\log(1+x)}{x}\,dx$.

Thus, we frequently have to apply a quadrature formula to the problem of evaluating $\int_a^b f(x)\,dx$, where $f(x)$ has a singularity close to, but not within, the range $[a, b]$, a situation in which it is notoriously difficult to achieve high accuracy.

The difficulty is basically one of polynomial representability. Provided that the function $f(x)$ is continuous and single-valued, it is always possible to find a polynomial of degree p which will pass through the $p+1$ points (x_0, f_0) , (x_1, f_1) , . . (x_p, f_p) . A Newton-Cotes quadrature formula simply evaluates the area under this interpolating polynomial, as an estimate for $\int_{x_0}^{x_p} f(x)\,dx$. There is, of course, no guarantee that the polynomial approximation to the function $f(x)$ is reasonable at points other than the mesh points, but, provided that the mesh length is small enough, the representation by a polynomial is usually adequate, in the sense that the polynomial has no turning points in the range $[x_0, x_p]$ other than those which correspond to turning points of the function $f(x)$. If, however, the values of the function are altering rapidly - and such is the case near a singularity - then the polynomial may be able to pass through the assigned points only at the cost of assuming extraneous maxima and minima within the range. That is, the interpolating polynomial looses "smoothness". The effect is illustrated by Figure 1, in which four curves are plotted for the range $[0.5 , 2.5]$ of x . Curve I is the graph of the

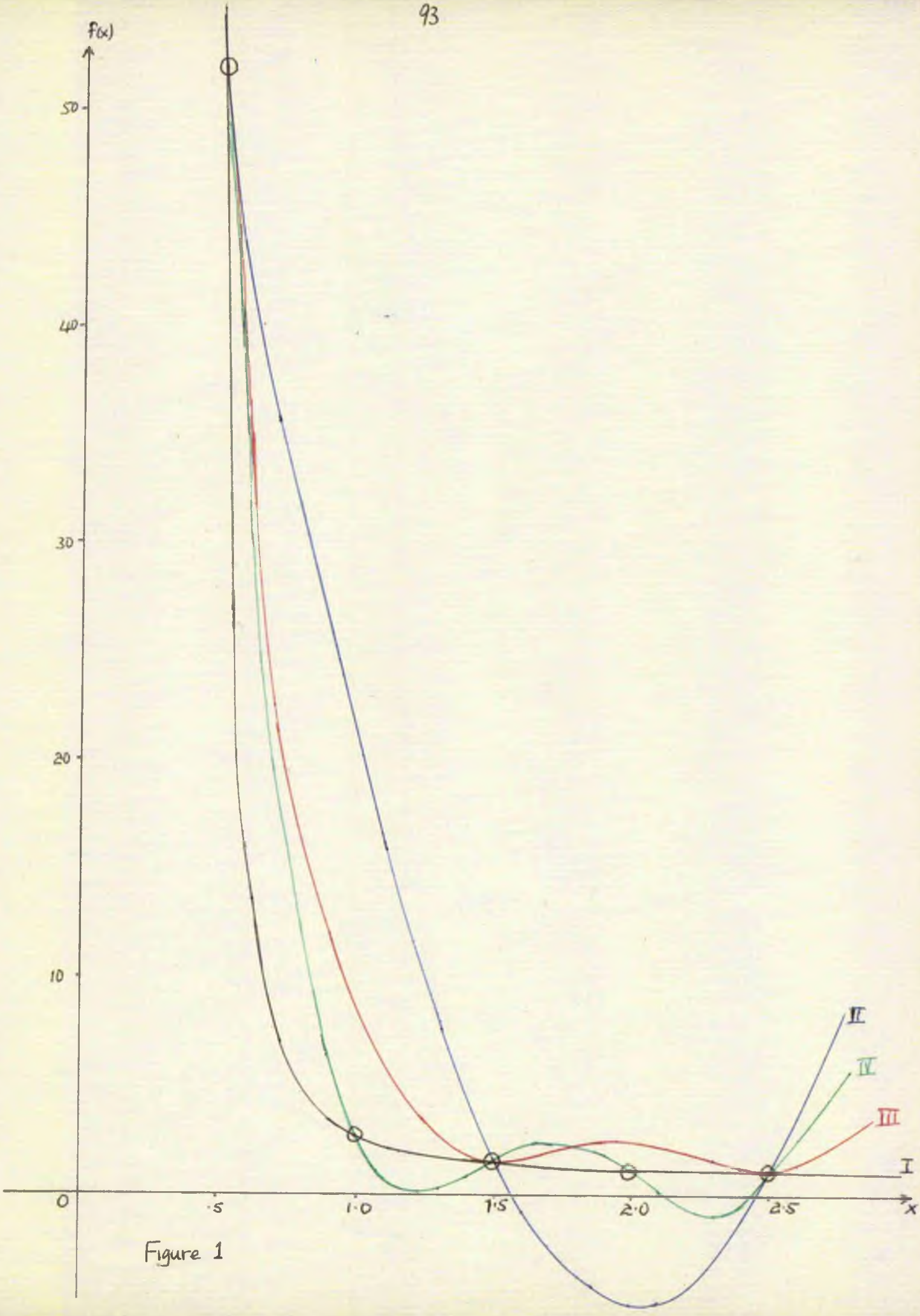


Figure 1

function e^{1/x^2}, which has a singularity at $x = 0$. Curve II is the graph of the unique interpolating polynomial of degree two, which assumes the values of $f(x)$ at the points $x = 0.5, 1.5, 2.5$. Its representation of the function $f(x)$ elsewhere is clearly far from satisfactory. Curve III is the graph of the unique polynomial of degree four, which assumes the values of $f(x)$ at the points $x = 0.5, 1.5, 2.5,$ and, moreover, has the same first derivative as $f(x)$ at the points $x = 1.5, 2.5$. Although this curve also oscillates, its representation of $f(x)$ is more satisfactory. That the oscillation is not removed, by requiring the interpolating polynomial to pass through a larger number of assigned points, is illustrated by Curve IV, which is the graph of the unique polynomial of degree four, which takes the values of $f(x)$ at the points $x = 0.5, 1.0, 1.5, 2.0, 2.5$.

To evaluate the integral $\int_{0.5}^{2.5} f(x)\,dx$ by calculating the area under the Curve II, is tantamount to using Simpson's Rule (optimum $[2\ ;\ 1]$ quadrature formula.) Calculating the area under Curve III is equivalent to using a new sub-optimum quadrature formula

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{15}\left(4f_0 + 16f_1 + 10f_2\right) - \frac{h^2}{15}\left(4f_1^{(1)} + 2f_2^{(1)}\right) \qquad (4.13)$$

with $$(TE)_1 = -\frac{1}{450}\,h^6 f^{(5)}$$

It can be argued, that in this example, the best representation of $f(x)$ would be got by choosing a large number of points on the

graph of $f(x)$, and linking them by a polygonal arc. The area under such a curve is then the estimate given by the repeated application of the Trapezoidal Rule (optimum $[1 ; 1]$ quadrature formula). This would indeed give a superior result, but it presupposes that the values of the integrand are available in a finer tabulation. If it were the case that e^{1/x^2} were tabulated only at $x = 0.5(1.0)n$, then the result by the trapezoidal rule would be poor (see Fig. 1), and we would have to choose between Simpson's Rule and formulae of the type (4.13). Nor would it be any improvement to find additional values of the function by interpolating from the tabulation, since, near a singularity, high accuracy interpolation suffers precisely the same difficulties as high accuracy quadrature.

Formula (4.13) is an example of a weighted sub-optimum formula, "weighted" in the sense that the higher derivatives are introduced at special points. At first sight, one might be inclined to introduce the higher derivatives at the points nearest to the singularity, on the grounds that this is the end of the range nearest the troublesome region. However, a glance at Figure 1 shows that there would be little to be gained by requiring the polynomial to have the same slope as $f(x)$ at the point $x = 0.5$, since all the polynomials considered have approximately the same slope there. In contrast, it is at the end of the range furthest from the singularity that the differences in the slopes are greatest. Thus, if we permit ourselves a quadrature formula of degree five, rather than four, we would choose, for this problem, not the optimum $[2 ; 2]$ quadrature

formula, but the new sub-optimum formula

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{30}(7f_0 + 16f_1 + 37f_2) - \frac{h^2}{15}(8f_1^{(1)} + 7f_2^{(1)}) + \frac{h^3}{15^2}f_2^{(2)} \qquad (4.14)$$

with $$(TE)_1 = \frac{+1}{1890}\,h^7 f^{(6)}$$

Many problems will require formulae of higher accuracy than (4.13) and (4.14), if the contribution from the range nearest the singularity is to match, in accuracy, contributions from neighbouring, less difficult, ranges. A weighted formula of higher accuracy, for use when there is a singularity of the integrand close to the left hand end of the range, will specify more high derivatives at the right hand end of the range. Such a formula is

$$\begin{aligned} 197{,}120\int_{x_0}^{x_3} f(x)\,dx = -h(&32{,}016f_0 - 2{,}009{,}853f_1 + 68{,}199{,}408f_2 \\ &- 65{,}630{,}211f_3) \\ &+ h^2(614{,}547f_1^{(1)} - 23{,}164{,}704f_2^{(1)} - 40{,}838{,}553f_3^{(1)}) \\ &- h^3(3{,}989{,}088f_2^{(2)} - 11{,}390{,}506f_3^{(2)}) \\ &- h^4(1{,}801{,}386f_3^{(3)}) + h^5(163{,}782f_3^{(4)}) - h^6(6{,}966f_3^{(5)}) \end{aligned} \qquad (4.15)$$

with $$(TE)_1 = -\frac{2{,}853}{9{,}865{,}856{,}000}\,h^{13} f^{(12)}$$

If this formula seems more unpleasant than most quadrature formulae, it should be remembered that it is designed to do an unpleasant job.

Formulae (4.13), (4.14) and (4.15) are weighted to cope with a singularity close to the left hand end of the range $[x_0, x_n]$. If the singularity is close to the right hand end of the same interval, a suitable set of formulae is readily obtained from (4.13) - (4.15) by making the transformation

$$h \longrightarrow -h$$

$$f_j^{(s)} \longrightarrow f_{n-j}^{(s)} \quad , \quad s = 0, 1, 2, \ldots$$

An example is now given to illustrate the advantages of these weighted formulae. The integral evaluated is $\int_{0.1}^{0.4} x^{-\frac{1}{2}}\,dx$. The following values were obtained.

Table 13. $\int_{0.1}^{0.4} x^{-\frac{1}{2}}dx$ (Theoretical value 0.632,455,532,034)

	Method	h	Result	Error
1	Formula (4.15)	0.1	0.632,475,724	+0.000,02
2	Optimum 3 ; 3	0.1	0.637,270,236	+0.005
3	Formula (4.15) adapted for right hand singularity	0.1	0.640,556,330	+0.008

Method 2 , the optimum $[3 ; 3]$ quadrature formula, has the same order of truncation error as Method 1 , and, in fact, its numerical coefficient is about 180 times smaller. It produces a much poorer result, because its derivatives are not distributed to best advantage. Method 3 is Method 1 adapted to deal with a right hand singularity. Applied to the present problem, it thus displays a distribution of derivatives even more unfavourable than that of Method 2 , and the result is correspondingly poorer still.

(c) Sub-optimum formulae for the integration of some special functions.

A class of sub-optimum formulae is now derived for the numerical evaluation of $\int_{\alpha}^{\beta} \eta(x)\,dx$, where η is a function defined by a differential equation of the type

$$\frac{d^2\eta}{dx^2} = g(x, \eta) \quad , \tag{4.16}$$

and is available as a tabulation at regular intervals of x. Examples of such functions, η, are the Airy integral, the Whittaker function, and the Mathieu function.

The suggestion that sub-optimum formulae of this type should be developed was put forward, during correspondence, by Dr J.C.P.Miller of the Mathematical Laboratory, Cambridge.

If in (4.3), all the coefficients a_{st}, other than a_{1t} and a_{3t} $(t = 0, 1, \ldots k)$ are equated to zero, then the following class of quadrature formulae is obtained.

$$\int_{x_0}^{x_k} f(x)\,dx = h \sum_{t=0}^{k} a_{1t}\, y_t^{(1)} + h^3 \sum_{t=0}^{k} a_{3t} y_t^{(3)}$$

Putting $y^{(1)} = \eta(x)$, we obtain

$$\int_{x_0}^{x_k} \eta(x)\,dx = h \sum_{t=0}^{k} a_{1t}\, \eta_t + h^3 \sum_{t=0}^{k} a_{3t} \eta_t^{(2)} \quad , \tag{4.17}$$

Coefficients

TABLE 14. Sub-optimum quadrature formulae.

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}				$-\frac{1,912}{821,205}$

$h^5 y^{(5)}$	$-\frac{1}{120}$			
$h^9 y^{(9)}$		$-\frac{1}{396,900}$	$\frac{39}{313,600}$	
$h^{13} y^{(13)}$				$-\frac{42,608}{554,867,688,375}$

where $\eta^{(2)}$ is readily obtained in terms of η and x from the differential equation (4.16) which defines the special function. The coefficients a_{1t} , a_{3t} (t = 0, 1, . . k) , and the principal truncation errors for formulae of class (4.17) are quoted in Table 14, for k = 1, 2, 3, 4.

IV.5 $[k\,;\,l]$ Formulae for Repeated Quadrature.

(Lambert and Mitchell, [27])

Quadrature formulae of the type derived in section 2 of this chapter can also be devised for the evaluation of the repeated integral

$$I_N = \underbrace{\int_a^b \int_a^x \cdots \int_a^x}_{N\text{-fold}} f(v)\,(dv)^N \quad ; \quad N \gg 2 \qquad (4.18)$$

It is, of course, possible to reduce such an integral to a single integral in the well known convolution form,

$$\underbrace{\int_a^b \int_a^x \cdots \int_a^x}_{N\text{-fold}} f(v)\,(dv)^N = \frac{1}{(N-1)!} \int_a^b (b-v)^{N-1} f(v)\,dv\,, \qquad (4.19)$$

and the right hand side can be evaluated by using one of the $k\,;\,$ quadrature formulae already obtained. It will be shown, however, that such a procedure is, in general, less accurate than the direct use of a formula designed for repeated quadrature. The latter type of formula turns out to have a more favourable truncation error, and, moreover, quadrature of the convolution form has the added disadvantage that the value of the function at the right hand end of the interval, being perforce multiplied by zero, cannot enter into the calculation. Thus, difficulty over polynomial representability may well arise, if the integrand varies rapidly at that end of the interval.

Suppose the function $y(x)$ is such, that $f(x) = y^{(N)}(x)$.

Then, by repeated integration, we obtain the result.

$$I_N = y(b) - \sum_{m=1}^{N} \frac{1}{(m-1)!} (b-a)^{m-1} y^{(m-1)}(a) \qquad (4.20)$$

If we introduce the $k+1$ equally spaced mesh points x_r, where

$$x_r = x_0 + rh \quad ; \quad r = 0, 1, 2, \ldots k$$

and $$x_0 = a \, , \quad x_k = b \, ,$$

then, in the notation employed throughout this thesis, (4.20) can be written

$$I_N = y_k - \sum_{m=1}^{N} \frac{1}{(m-1)!} (kh)^{m-1} y_0^{(m-1)} \qquad (4.21)$$

The repeated integral I_N can thus be evaluated if we can find a class of difference equations

$$I_N = y_k - \sum_{m=1}^{N} \frac{1}{(m-1)!} (kh)^{m-1} y_0^{(m-1)} = \sum_{s=N}^{l} \sum_{t=0}^{k} a_{st} h^s y^{(s)}, \quad l \geq N \qquad (4.22)$$

This class of difference equations is identically the class (2.2), in which the suffix n has been replaced by 0, and where the following values have been assigned to certain of the coefficients a_{st}.

$$\begin{aligned} &a_{00} = 1 \; ; \; a_{0k} = -1 \; ; \; a_{0t} = 0 \, , \; t = 1, 2, \ldots k-1 \\ &a_{st} = 0 \, , \; s = 1, 2, \ldots N-1 \; ; \; t = 1, 2, \ldots k-1 \\ &a_{s0} = k^s/s! \; ; \; s = 1, 2, \ldots N-1 \, . \end{aligned} \qquad (4.23)$$

There remain $(k+1)(\ell - N + 1)$ undetermined coefficients a_{st}, $(s = N, N+1, \ldots \ell;\ t = 0, 1, \ldots k)$ which can be evaluated by the technique already used to derive the $[k\ ;\ \ell]$ quadrature formulae.

Taylor expansions for y and its derivatives about x_0 are substituted into (4.22), yielding the result

$$E_0 y_0 + E_1 h y_0^{(1)} + E_2 h^2 y_0^{(2)} + \ldots + E_m h^m y_0^{(m)} + \ldots = 0 \qquad (4.24)$$

where

$$\begin{aligned}
E_m &= 0\ ;\ m = 0, 1, 2, \ldots N-1 \\
E_N &= -\frac{k^N}{N!} + \sum_{t=0}^{k} a_{Nt} \\
E_m &= -\frac{k^m}{m!} + \sum_{s=N}^{m-1} \frac{1}{(m-s)!} \sum_{t=1}^{k} t^{m-s} a_{st} + \sum_{t=0}^{k} a_{mt}\ ;\ m = N+1, N+2, \ldots \\
E_m &= -\frac{k^m}{m!} + \sum_{s=N}^{\ell} \frac{1}{(m-s)!} \sum_{t=1}^{k} t^{m-s} a_{st}\ ;\ m = \ell+1, \ell+2, \ldots
\end{aligned} \qquad (4.25)$$

We now put $E_m = 0$, $m = N, N+1, \ldots \{(k+1)(\ell+1) - Nk - 1\}$, and solve for the remaining $(k+1)(\ell - N+1)$ undetermined a_{st}. With these values, (4.22) is now the class of optimum quadrature formulae for an N-fold repeated integral of the type (4.18).

The principal truncation error is

$$(TE)_1 = E_{(k+1)(\ell+1)-Nk} \left(h \frac{d}{dx}\right)^{(k+1)(\ell+1)-Nk} y\ . \qquad (4.26)$$

As before, the optimum coefficients a_{st} and the numerical coefficient

in the principal truncation error are independent of the choice of origin for the Taylor expansions. For this set of formulae, it happens that the derivation of the coefficients is least laborious when x_0 is chosen as the origin.

The optimum coefficients a_{st}, and the principal truncation errors are set out as follows.

Table 15 ; N = 2 : k = 1 ; ℓ = 2, 3, 4, 5.
k = 2 ; ℓ = 2, 3, 4.
k = 3 ; ℓ = 2, 3.
k = 4 ; ℓ = 2.
k = 5 ; ℓ = 2.

Table 16 ; N = 3 : k = 1 ; ℓ = 3, 4, 5.
k = 2 ; ℓ = 3, 4.
k = 3 ; ℓ = 3, 4.
k = 4 ; ℓ = 3.
k = 5 ; ℓ = 3.

Table 17 ; N = 4 : k = 1 ; ℓ = 4, 5.
k = 2 ; ℓ = 4, 5.
k = 3 ; ℓ = 4.
k = 4 ; ℓ = 4.

A study of these tables substantiates the conclusion reached in other parts of this thesis, that it is better to use finite difference formulae of low order, involving higher derivatives, than to attempt

Coefficients TABLE 15. Formulae for Repeated Quadrature: $N = 2$

K	1				2			3		4	5
l	2	3	4	5	2	3	4	2	3	2	2
a_{20}	$\frac{1}{3}$	$\frac{7}{20}$	$\frac{5}{14}$	$\frac{13}{36}$	$\frac{2}{3}$	$\frac{79}{105}$	$\frac{228}{315}$	$\frac{39}{40}$	$\frac{1206}{1120}$	$\frac{56}{45}$	$\frac{2,050}{2,016}$
a_{21}	$\frac{1}{6}$	$\frac{3}{20}$	$\frac{2}{14}$	$\frac{5}{36}$	$\frac{4}{3}$	$\frac{118}{105}$	$\frac{384}{315}$	$\frac{108}{40}$	$\frac{2,187}{1,120}$	$\frac{192}{45}$	$\frac{16,875}{2,016}$
a_{22}						$\frac{19}{105}$	$\frac{18}{315}$	$\frac{27}{40}$	$\frac{1,458}{1,120}$	$\frac{48}{45}$	$\frac{2,500}{2,016}$
a_{23}								$\frac{6}{40}$	$\frac{189}{1,120}$	$\frac{64}{45}$	$\frac{6,250}{2,016}$
a_{24}										0	$\frac{4,250}{2,016}$
a_{25}											$\frac{275}{2,016}$
a_{30}		$\frac{3}{60}$	$\frac{13}{210}$	$\frac{17}{252}$		$\frac{10}{105}$	$\frac{34}{315}$		$\frac{135}{1,120}$		
a_{31}		$\frac{-2}{60}$	$\frac{-8}{210}$	$\frac{-10}{252}$		$\frac{-16}{105}$	$\frac{-32}{315}$		$\frac{-486}{1,120}$		
a_{32}						$\frac{-4}{105}$	$\frac{-2}{315}$		$\frac{-243}{1,120}$		
a_{33}									$\frac{-36}{1,120}$		
a_{40}			$\frac{4}{840}$	$\frac{7}{1008}$			$\frac{2}{315}$				
a_{41}			$\frac{3}{840}$	$\frac{5}{1,008}$			$\frac{16}{315}$				
a_{42}							0				
a_{50}				$\frac{6}{15,120}$							
a_{51}				$\frac{-4}{15,120}$							

Truncation Errors

$h^4 y^{(4)}$	$\frac{1}{24}$										
$h^5 y^{(5)}$					$\frac{-2}{45}$						
$h^6 y^{(6)}$		$\frac{-1}{1440}$									
$h^7 y^{(7)}$								$\frac{9}{160}$		$\frac{-16}{315}$	
$h^8 y^{(8)}$			$\frac{1}{201,600}$			$\frac{-1}{4725}$					$\frac{1,375}{24,192}$
$h^9 y^{(9)}$											
$h^{10} y^{(10)}$				$\frac{-1}{50,803,200}$					$\frac{-27}{627,200}$		
$h^{11} y^{(11)}$							$\frac{-1}{13,017,700}$				

Coefficients TABLE 16. Formulae for Repeated Quadrature : N=3

k	1			2		3		4	5
ℓ	3	4	5	3	4	3	4	3	3
$a_{\ell 0}$	$\frac{3}{24}$	$\frac{4}{30}$	$\frac{23}{168}$	$\frac{9}{15}$	$\frac{68}{105}$	$\frac{105}{80}$	$\frac{3285}{2240}$	$\frac{744}{315}$	$\frac{29,125}{8,064}$
$a_{\ell 1}$	$\frac{1}{24}$	$\frac{1}{30}$	$\frac{5}{168}$	$\frac{12}{15}$	$\frac{64}{105}$	$\frac{243}{80}$	$\frac{4,374}{2240}$	$\frac{2,176}{315}$	$\frac{101,375}{8,064}$
$a_{\ell 2}$				$\frac{-1}{15}$	$\frac{8}{105}$	0	$\frac{2,187}{2,240}$	$\frac{96}{315}$	$\frac{1,250}{8,064}$
$a_{\ell 3}$						$\frac{9}{80}$	$\frac{334}{2,240}$	$\frac{384}{315}$	$\frac{36,750}{8,064}$
$a_{\ell 4}$								$\frac{-40}{315}$	$\frac{-4,375}{8,064}$
$a_{\ell 5}$									$\frac{1,375}{8,064}$
$a_{4 0}$		$\frac{2}{120}$	$\frac{7}{336}$		$\frac{8}{105}$		$\frac{351}{2,240}$		
$a_{4 1}$		$\frac{-1}{120}$	$\frac{-3}{336}$		$\frac{-16}{105}$		$\frac{-1,458}{2,240}$		
$a_{4 2}$					$\frac{-2}{105}$		$\frac{-729}{2,240}$		
$a_{4 3}$							$\frac{-54}{2,240}$		
$a_{5 0}$			$\frac{5}{3,360}$						
$a_{5 1}$			$\frac{3}{3,360}$						

Truncation Errors.

$h^5 y^{(5)}$	$\frac{1}{80}$								
$h^6 y^{(6)}$				$\frac{-2}{45}$					
$h^7 y^{(7)}$		$\frac{-1}{5,040}$				$\frac{81}{1120}$			
$h^8 y^{(8)}$								$\frac{-32}{315}$	
$h^9 y^{(9)}$			$\frac{1}{725,760}$		$\frac{-2}{14,175}$				$\frac{12,625}{145,152}$
$h^{10} y^{(10)}$									
$h^{11} y^{(11)}$							$\frac{-81}{1,724,800}$		

Coefficients TABLE 17: Formulae for Repeated Quadrature: N=4.

k	1		2		3	4
l	4	5	4	5	4	4
a_{40}	$\frac{4}{120}$	$\frac{6}{168}$	$\frac{16}{45}$	$\frac{24}{63}$	$\frac{351}{280}$	$\frac{2816}{945}$
a_{41}	$\frac{1}{120}$	$\frac{1}{168}$	$\frac{16}{45}$	$\frac{16}{63}$	$\frac{648}{280}$	$\frac{7168}{945}$
a_{42}			$\frac{-2}{45}$	$\frac{2}{63}$	$\frac{-81}{280}$	$\frac{-768}{945}$
a_{43}					$\frac{127}{280}$	$\frac{1024}{945}$
a_{44}						$\frac{-160}{945}$
a_{50}		$\frac{5}{1,260}$		$\frac{40}{945}$		
a_{51}		$\frac{-2}{1,260}$		$\frac{-80}{945}$		
a_{52}				$\frac{-8}{945}$		

Truncation Errors.

$h^6 y^{(6)}$	$\frac{+1}{360}$					
$h^7 y^{(7)}$			$\frac{-8}{315}$			
$h^8 y^{(8)}$		$\frac{-1}{24,192}$			$\frac{+297}{4,480}$	
$h^9 y^{(9)}$						$\frac{-256}{2025}$
$h^{10} y^{(10)}$				$\frac{-1}{14,175}$		

to gain comparable accuracy by increasing the number of points, and ignoring the higher derivatives. A good illustration is afforded by the pair of formulae

$$N = 3,\ k = 1,\ \ell = 5$$

and

$$N = 3,\ k = 5,\ \ell = 3.$$

Both have principal truncation errors of order nine, but the numerical coefficient in the principal truncation error of the latter is $\frac{18,625}{145,152}$, against $\frac{1}{725,760}$ of the former. Moreover, the coefficients of the integrand and its derivatives are rather easier in the low order formula.

The formulae for which $\ell = N$ do not, of course, involve derivatives of the integrand, and can be used even if $f(v)$ is given only as a tabulation at equal interval. They are the generalization, to the N-fold integral, of the Newton-Cotes formulae, although it is of interest to note that they do not appear to be anywhere in the literature.

Formulae for the numerical evaluation of repeated integrals of the type (4.18) have been derived by Salzer [28], [29], and also in an H.M.S.O. pamphlet, [30]. These papers however are restricted to the case where the range of integration $[a, b]$ is only one mesh length (i.e. $b = a+h$), and they employ values of the integrand outside this range. (The derivatives of the integrand are not involved.) Since the value of I_N over the range $[a, c]$ cannot, for $N \geqslant 2$, be deduced from the values of I over the ranges

$[a , b]$ and $[b , c]$, without knowledge of the derivatives $y^{(s)}$ $(s = 1, 2, \ldots N-1)$ at $x = b$, the formulae proposed in these papers cannot be adequately compared with the formulae derived in this section.

It remains only to compare the process of evaluating I_N by a formula of class (4.22) with that of evaluating the convolution form (4.19) by one of the $[k ; l]$ quadrature formulae derived in section 2 of this chapter. A fair comparison can be made only if we stipulate the number of points involved within the range of integration, and the number of derivatives of the integrand, $f(v)$, employed. Let the former be $K+1$, and the latter be L . The appropriate single quadrature formula for use on the convolution form is then the optimum $[K ; L+1]$ quadrature formula, whose principal truncation error is, by (4.8), of order

$$K(L + 1) + (L + 1) + L = (K+1)(L+1) + 1 .$$

The appropriate formula of class (4.22) is, however, given by $k = K$, $l = L + N$, and its principal truncation error is, by (4.26), of order

$$(K+1)(L+N+1) - NK = (K+1)(L+1) + N .$$

Hence the truncation error for the formula of class (4.22) is of order $N-1$ higher than that for a comparable single quadrature of the convolution form. The numerical coefficients of the principal truncation errors of the two methods are roughly comparable, but

tend to favour the method employing a direct repeated quadrature formula.

In the case $N = 2$, however, the gain of $N-1$ in the order of the truncation error will be nullified if the single quadrature formula, used on the convolution form, has an extra-predicted truncation error. (See IV - 2). In such cases the order of the principal truncation error is the same for both methods.

The labour by the convolution method is perhaps greater, owing to the presence of the factor $(b - v)^{N-1}$ in the integrand., Moreover, the already mentioned danger concerning polynomial representability, when the convolution method is adopted, is a further drawback.

Finally, an example is given. The integral considered is

$$\int_0^{0.1}\int_0^{x}\int_0^{x}\int_0^{x} \frac{24}{(1+v)^5}\,(dv)^4 \ .$$

The following results were obtained

Table 18.

	Theoretical Value	0.000,090,909,09
1	$N = 4,\ k = 1,\ \ell = 5$	0.000,090,897,85
2	$N = 4,\ k = 2,\ \ell = 5$	0.000,090,909,07
3	Convolution, by $k = 1,\ \ell = 2$	0.000,083,333,33
4	Convolution, by $k = 2,\ \ell = 2$	0.000,090,894,00

Methods 1 and 3 are directly comparable in terms of number

points involved, and number of derivatives of the integrand employed, as are methods 2 and 4 .

References.

[1] TODD, J. (Editor) Survey of Numerical Analysis. McGraw-Hill, New York, (1962).

[2] BASHFORTH, F. & ADAMS, J.C. An attempt to test the theories of capillary action . . . with an explanation of the method of integration employed. Cambridge University Press, (1883).

[3] FORT, T. Finite Differences and Difference Equations in the Real Domain. Oxford University Press, (1948).

[4] SALZER, H.E. New Formulae for facilitating osculatory interpolation. Journ. Res. Natl. Bur. Standards, 52 pp 211-216, (1954)

[5] SQUIRES, W. Some applications of quadrature by differentiation. Journ. Soc. Indust. Appld. Math., 9, pp 94-108, (1961).

[6] HAMMER, P.C. & WICKE, H.W. Quadrature formulae involving derivatives of the integrand. Math. Comp., 14, pp 3-7, (1960).

[7] MILNE, W.E. A note on the numerical integration of differential equations. Journ. Res. Natl. Bur. Standards, 43, pp 537-542, (1949).

[8] LAMBERT, J.D. & MITCHELL, A.R. On the solution of $y'=f(x, y)$ by a class of high accuracy difference formulae of low order. Zeit. angew. Math. und Phys., 13, pp 223-232, (1962).

[9] COLLATZ, L. The Numerical Treatment of Differential Equations. 3rd Ed., Springer-Verlag, Berlin, (1960).

[10] MILNE, W.E. Numerical integration of ordinary differential equations. Amer. Math. Month., 33, pp455-460, (1926).

[11] MILNE, W.E. Numerical Solution of Differential Equations. Wiley, New York, (1953).

[12] URABE, M. Theory of errors in numerical integration of ordinary differential equations. Tech. Summary Report, No. 183, M.R.C., Madison, (1960).

[13] ALLEN, D.N. deG. & SEVERN, R.T. The application of relaxation methods to the solution of non-elliptic partial differential equations. Quart. Journ. Mech. App. Math., 4 ,p 209, (1951).

[14] MITCHELL, A.R. & RUTHERFORD, D.E. On the theory of relaxation. Proc. Glasgow Math. Assoc., 1, p 101, (1953).

[15] FOX, L. & MITCHELL, A.R. Boundary-value techniques for the numerical solution of initial value problems in ordinary differential equations. Quart. Journ. Mech. App. Math., 10, pp 232-243, (1957).

[16] TODD, J. Solution of differential equations by recurrence relations. Math. Tab & other aids Computation, 4, p 39, (1950)

[17] DAHLQUIST, G. Convergence and stability in the numerical integration of ordinary differential equations. Math. Scand. 4, pp 33-53, (1956).

[18] DAHLQUIST, G. Stability and error bounds in the numerical integration of ordinary differential equations. Kungl. Tekn. Högsk. Handl. Stockholm, no. 130, pp 1 - 87, (1959).

[19] RICHTMYER, R.D. Difference Methods for Initial-Value Problems. Interscience, New York, (1957).

[20] RUTISHAUSER, H. Ueber die Instabilität von Methoden zur Integration gewöhnlicher Differentialgleichungen. Zeit. angew. Math. und Phys., 3, pp65-74, 1952.

[21] MITCHELL, A.R. & CRAGGS, J. W. Stability of Difference relations in the solution of ordinary differential equations. Math. Tab.& other aids Computation, 7, p 127, (1953).

[22] ALLEN, D.N. de G. A suggested approach to finite-difference representation of differential equations, with an application to determine temperature distributions near a sliding contact. Quart. Journ. Mech. Appl. Math., 15, p 11,(1962).

[23] HULL, T.E. & LUXEMBURG, W.A.J. Numerical methods and existence theorems for ordinary differential equations. Num. Math., 2, pp 30-41, (1960).

[24] LAMBERT, J.D. & MITCHELL, A.R. The use of higher derivatives in quadrature formulae. Computer Journ., 5, p 322, (1963).

[25] HILDEBRAND, F. B. Introduction to numerical analysis. McGraw-Hill, New York, (1956)

[26] STRUBLE, G. Tables for use in quadrature formulae involving derivatives of the integrand. Math. Comp., 14, pp 8-12, (1960).

[27] LAMBERT, J. D. & MITCHELL, A.R. Repeated quadrature using derivatives of the integrand. Submitted to Zeit. Angew Math, und Phys.

[28] SALZER, H.E. Table of coefficients for double quadrature without differences, for integrating second order differential equations. Journ. Math. Phys., 24, (1945).

[29] SALZER, H.E. Coefficients for repeated integration with central differences. Journ. Math. Phys., 28, (1949-50).

[30] Interpolation and allied tables. H.M.S.O., (1956).