THE NUMERICAL SOLUTION OF VOLTERRA INTEGRAL EQUATIONS

by

Janet P. Rae, B.Sc.

A dissertation presented for the degree of Master of Science

of the University of St. Andrews

1974

ABSTRACT

The aim of this dissertation is to present a survey of numerical methods for solving Volterra equations of the first and second kinds. At present this subject matter is to be found only in research papers; this dissertation collects all such work and from it attempts to present a concise and logical account of the subject. Due to the nature of the dissertation, however, only the more important and practical methods can be included. In several places new methods are suggested. Numerical examples are included where it has seemed relevant and interesting. All calculations were performed on the University of St. Andrews' computer, an IBM 360, using the language FORTRAN IV with double-precision arithmetic. Discussions on the methods are included within each chapter.

In Chapter one we give a brief introduction to the equations together with important definitions, theorems and a lemma.

In Chapter two we examine numerical methods for the equation of the second kind. We find that the equation is readily adaptable to numerical treatment, but that some problems in numerical stability can result. (Section 1-4). A new variant on the block-by-block methods is suggested in Section 2-3, together with some error analysis.

In Chapter three we examine the linear equation of the first kind. The work presented here is very new and indeed the subject is not yet exhausted. We find that the equation does not lend itself readily to numerical treatment and many difficulties are incurred. In Sections 3-3 and 3-4, the author considers some applications of the convergence theory of 3-1 and 3-2, and the possibility of the use of expansion methods is also discussed in Section 5.

ACKNOWLEDGEMENT

I would like to thank Dr. G.M. Phillips for his help and encouragement in the preparation of this dissertation.

I declare that this dissertation is of my own composition and that the work of which it is a record has been carried out by myself. It has not been submitted in any previous application for a degree or professional qualification.

25th April 1975.

CONTENTS

CHAPTER 1

INTRODUCTION

1 - 1 Volterra Integral equations

The non-linear Volterra integral equation of the second kind is

$$y(x) = g(x) + \int_a^x K(x,t,y(t))\,dt \,, \qquad a \le x \le b \,, \qquad \ldots.(1\text{-}1\text{-}1)$$

where the function y(t) is to be determined. The equation is linear if $K(x,t,y(t)) = k(x,t)y(t)$ and in this case $k(x,t)$ is called the Kernel.

The Volterra equation of the first kind is

$$g(x) = \int_a^x K(x,t,y(t))\,dt \,, \qquad a \le x \le b \,, \qquad \ldots.(1\text{-}1\text{-}2)$$

However, in this Thesis we shall only consider the linear equation:

$$g(x) = \int_a^x K(x,t)\,y(t)\,dt \,, \qquad a \le x \le b \,. \qquad \ldots.(1\text{-}1\text{-}3)$$

In contrast to the Volterra case, Fredholm equations possess a fixed upper limit to the integral. That is, the Fredholm equations of the second and first kinds are, respectively;

$$y(x) = g(x) + \int_a^b K(x,t,y(t))\,dt \,, \qquad a \le x \le b \,,$$

$$g(x) = \int_a^b K(x,t,y(t))\,dt \,, \qquad a \le x \le b \,.$$

We notice that we can recast (1-1-1) and (1-1-2) as Fredholm equations with integrand $K_F(x,t,y(t))$, where

$$K_F(x,t,y(t)) = K(x,t,y(t)) \qquad a \le t \le x$$
$$= 0 \qquad x < t \le b$$

In this thesis we restrict ourselves to equations possessing non-singular integrands. That is, $K(x,t,y(t))$ has no singularities in the range of integration. If the equation is linear, and $K(x,t) = K(x-t)$, then the kernel is said to be a convolution kernel.

If $K(x,t)$ is of the form

$$K(x,t) = \sum_{K=1}^{N} X_K(x)\, T_K(t)$$

then the kernel is said to be of finite rank N. This type of kernel is also termed degenerate. It is assumed that $X_K(x), T_K(t)$, $K=1,\ldots,N$ are linearly independent sets of functions.

It is known (see Tricomi [23]) that the eigenvalue problem for the Volterra equation

$$\int_a^x K(x,t,y(t))\,dt = \lambda y(x),$$

where λ is a parameter, possesses only the trivial solution.

Volterra integral equations are related to initial value problems. For example, consider the linear differential equation

$$u^{(n)}(x) = a_0(x)u^{(n-1)}(x) + \ldots + a_{n-1}(x)u(x) + g(x) \qquad \ldots.(1\text{-}1\text{-}4)$$

with initial values

$$u(0) = u'(0) = \ldots = u^{(n-1)}(0) = 0 \quad , \qquad 0 \le x \le b.$$

Suppose we define $y(x) = u^{(n)}(x)$, then we have

$$u^{(v)}(x) = \frac{1}{(n-v-1)!} \int_0^x (x-t)^{n-v-1} y(t)\, dt \qquad 0 \le v \le n-1$$

[See Cochran [6]]

It then follows that (1-1-4) can be written as the linear Volterra equation

$$y(x) = g(x) + \int_0^x K(x,t)\, y(t)\, dt \qquad 0 \le x \le b \qquad \ldots (1\text{-}1\text{-}5)$$

where

$$K(x,t) = \sum_{v=0}^{n-1} \frac{a_v(x)}{v!} (x-t)^v \qquad \ldots (1\text{-}1\text{-}6)$$

Similarly, the integral equation (1-1-5) with kernel (1-1-6) can be written in the form (1-1-4). In Chapter 2 we shall find that many of the 'standard' finite difference methods for initial value problems can be modified to suit the Volterra equation of the second kind.

1 - 2 Existence and uniqueness theorems

The following existence and uniqueness theorem for non-linear Volterra equations of the second kind follows from the well-known Picard iteration. The theorems below are taken from Linz [15]; proofs may be found in Tricomi [23], for example.

Theorem 1-1 Consider the non-linear Volterra integral equation of the second kind

$$y(x) = g(x) + \int_a^x K(x,t, y(t))\, dt , \qquad a \le x \le b ,$$

with (a) $g(x)$ continuous and bounded in $[a,b]$.

(b) $K(x,t,y)$ uniformly continuous in x and t for all finite y.

(c) $K(x,t,y)$ satisfying a uniform Lipschitz condition

$$|K(x,t,y_1) - K(x,t,y_2)| \leq L\,|y_1 - y_2|,$$

where L is a constant independent of x and t.

If conditions (a), (b) and (c) are satisfied, then the solution $y(x)$ is unique and continuous.

The linear Volterra equation of the first kind

$$g(x) = \int_a^x K(x,t)\,y(t)\,dt\,, \qquad a \leq x \leq b \qquad .... (1\text{-}2\text{-}1)$$

can be converted into an integral equation of the second kind provided $K(x,t)$ is differentiable with respect to its first argument, and that $g(x)$ is differentiable. To show this, differentiate (1-2-1) with respect to x

$$g'(x) = K(x,x)\,y(x) + \int_a^x \frac{\partial K}{\partial x}(x,t)\,y(t)\,dt\,.$$

That is

$$y(x) = \frac{g'(x)}{K(x,x)} - \int_a^x \frac{\partial K/\partial x\,(x,t)}{K(x,x)}\,y(t)\,dt\,. \qquad .... (1\text{-}2\text{-}2)$$

Equation (1-2-2) is a linear Volterra equation of the second kind with kernel

$$-\partial K/\partial x\,(x,t)\;/\;K(x,x).$$

Theorem 1-2 now follows.

Theorem 1-2 Consider the linear Volterra equation of the first kind

$$g(x) = \int_a^x K(x,t)y(t)\,dt \qquad a \le t \le x \le b$$

with (a) $g^{(1)}(x)$ continuous and bounded in $[a,b]$

(b) $\frac{\partial K(x,t)}{\partial x}$ uniformly continuous in x and t and $K(x,x) \neq 0$

(c) $g(a) = 0$.

Then if conditions (a), (b) and (c) are satisfied, the solution y(x) is unique and continuous.

Condition (c) is clearly necessary, or (1-2-1) is not consistent. From now on we will assume that the conditions stated in Theorems (1-1) and (1-2) hold.

Finally, we state an important lemma, due to Linz[15], which will be used several times in Chapters 2 and 3.

Lemma 1 If we have a sequence $\{\varepsilon_i\}$, $i = 0, 1, \ldots$ satisfying

$$|\varepsilon_n| \le A \sum_{i=0}^{n-1} |\varepsilon_i| + B \qquad n = s, s+1, \ldots$$

with finite $A, B > 0$

and

$$\sum_{i=0}^{s-1} |\varepsilon_i| \le \eta$$

with finite $\eta > 0$, then $|\varepsilon_n| \le (B + A\eta)(1 + A)^{n-s}$.

CHAPTER 2

THE VOLTERRA EQUATION OF THE SECOND KIND

Section 1: MULTISTEP METHODS

1 - 1 Introduction and basic definitions

The non-linear Volterra equation of the second kind is

$$y(x) = g(x) + \int_a^x K(x,t,y(t))\,dt \qquad a \le x \le b \qquad \ldots.(1\text{-}1\text{-}1)$$

Consider the grid of equally-spaced points $x_n = x_o + nh$, with $x_o = a, x_N = b, n = 0,1,\ldots, N$. Writing $x = x_n$ in (1-1-1) we obtain

$$y(x_n) = g(x_n) + \int_a^{x_n} K(x_n,t,y(t))\,dt \qquad n = 0,1,\ldots,N. \quad \ldots.(1\text{-}1\text{-}2)$$

If we now replace the integral by a quadrature rule we have

$$y(x_n) = g(x_n) + h \sum_{i=0}^{n} \omega_{ni}\, K(x_n, x_i, y(x_i)) + R_{n,h},$$

where w_{ni}, $i = 0,\ldots,n$, are the weights for the rule and $R_{n,h}$ is the truncation error of the rule.

We can obtain the scheme

$$Y_n = g(x_n) + h \sum_{i=0}^{n} \omega_{ni}\, K(x_n, x_i, Y_i) \qquad \ldots.(1\text{-}1\text{-}3)$$

If $Y_o, \ldots, Y_{n-1}$, are known, then (1-1-3) gives a scheme for determining Y_n, an approximation to $y(x_n)$. If $K(x_n, x_n, Y_n) \neq 0$ and $w_{nn} \neq 0$, the scheme (1-1-3) is implicit; that is we have a non-zero term involving Y_n on the right-hand side. Y_n can be obtained from the

iterative procedure

$$Y_n^{(r)} = g(x_n) + h\sum_{i=0}^{n-1} \omega_{ni}\, K(x_n, x_i, Y_i) + h\,\omega_{nn}\, K(x_n, x_n, Y_n^{(r-1)}) \quad r=1,2,\ldots \quad \ldots.(1\text{-}1\text{-}4)$$

We can take

$$Y_n^{(0)} = g(x_n)$$

or

$$Y_n^{(0)} = g(x_n) + h\sum_{i=0}^{n-1} \omega'_{ni}\, K(x_n, x_i, Y_i)$$

where the weights w'_{ni} are those belonging to an open Newton-Cotes rule.

Equation (1-1-3) is of the form

$$Y_n = \phi(Y_n),$$

and hence the iteration (1-1-4) will converge to the unique solution Y_n, provided that

$$hL|\omega_{nn}| < 1 \qquad \ldots.(1\text{-}1\text{-}5)$$

where L is the Lipschitz constant introduced in Chapter 1.

This can be guaranteed by taking h small enough. The result follows from the contraction mapping theorem [10].

In the case of the linear Volterra equation

$$K(x,t,y(t)) = K(x,t)\,y(t),$$

and (1-1-3) reduces to the following scheme for Y_n

$$Y_n = \{g(x_n) + h\sum_{i=0}^{n-1} \omega_{ni}\, K(x_n, x_i) Y_i\} \,/\, \{1 - h\,\omega_{nn}\, K(x_n, x_n)\}.$$

We notice from (1-1-1) that Y_o can be obtained exactly from

$$Y_o = g(x_o)$$

In the following sections we will assume that this value of Y_o is used. We now state two definitions which will prove themselves to be important in the following sections.

Definition 1-1 Consistency. (Linz [15])

The approximation method (1-1-3) is said to be consistent if it is based on a convergent quadrature.

Definition 1-2 Repetition factor.

The approximation method (1-1-3) is said to possess a repetition factor ρ, if ρ is the smallest integer such that

$$\omega_{n+\rho,i} = \omega_{n,i} \qquad i = \alpha, \alpha+1, \ldots, n-\beta$$

for integers α and β independent of n.

This definition, introduced by Linz [15], is found to be very important when we consider the problem of the stability of (1-1-3), and also when we consider the application of multistep methods to the Volterra equation of the first kind (see Chapter 3). The restriction of the values of i allows for the fact that different quadratures may need to be introduced over the first few values of x_i, $i=o,\ldots,n$. for certain values of n. This idea is made clearer in Section (1-3).

1 - 2 Convergence and stability

For convenience, we rewrite (1-1-3) below:

$$Y_n = g(x_n) + h \sum_{i=0}^{n}{}' \omega_{ni} K(x_n, x_i, Y_i). \qquad \ldots.(1\text{-}2\text{-}1)$$

Now any quadrature, (bar the simple Euler and trapezoidal rules used in each interval $[x_i, x_{i+1}]$), used in (1-2-1) will necessitate the use of starting values $Y_1, \ldots, Y_s$ say, before we can use (1-2-1) with n=s,s+1,...,N. Let us suppose that the errors incurred through using (1-2-1) are given by ε_i, where

$$\varepsilon_i = y(x_i) - Y_i \qquad i = 1, 2, \ldots, N$$

and that the starting errors satisfy

$$\sum_{i=1}^{s}{}' |\varepsilon_i| \leq \eta$$

for some finite $\eta > 0$.

Now the equation corresponding to (1-2-1) involving exact values of y(x) is

$$y(x_n) = g(x_n) + h \sum_{i=0}^{n}{}' \omega_{ni} K(x_n, x_i, y(x_i)) + R_{n,h}. \qquad \ldots.(1\text{-}2\text{-}2)$$

Subtracting (1-2-1) from (1-2-2) gives

$$\varepsilon_n = h \sum_{i=0}^{n}{}' \omega_{ni} \{K(x_n, x_i, y(x_i)) - K(x_n, x_i, Y_i)\} + R_{n,h} \qquad \ldots.(1\text{-}2\text{-}3)$$

Taking moduli in (1-2-3) and using the Lipschitz condition gives

$$|\varepsilon_n| \leq hLW \sum_{i=1}^{n}{}' |\varepsilon_i| + R(h),$$

where

$$|\omega_{ni}| \leq W, \qquad 0 \leq i \leq n \leq N$$

$$|R_{n,h}| \leq R(h), \qquad n \leq N.$$

Thus

$$|\varepsilon_n| \leq \frac{hLW}{\{1-hLW\}} \sum_{i=1}^{n-1} |\varepsilon_i| + \frac{R(h)}{\{1-hLW\}} \qquad n = S, S+1, \ldots, N$$

Then provided $hLW < 1$, we obtain, from Lemma 1, Chapter 1,

$$|\varepsilon_n| \leq \left\{ \frac{R(h)}{1-hLW} + hLW\eta \right\} e^{\{LW/1-hLW\}x_n}.$$

Thus we have $\lim_{\substack{h\to 0 \\ n\to\infty}} |\varepsilon_n| = 0$ with nh fixed, provided the scheme is consistent.

Notice here that if R(h) is $O(h^p)$, then we need only have $\eta = O(h^{p-1})$ to obtain an overall convergence of $O(h^p)$. In other words we need only obtain the starting values to an accuracy of $O(h^{p-1})$ to obtain solutions to the integral equation which are accurate to $O(h^p)$. The results so far in this section are summed up in the following theorem.

Theorem 1. The scheme (1-2-1) is convergent of order $O(h^p)$ provided the quadrature rule has a truncation error $O(h^p)$ and the starting values have been obtained to an accuracy $O(h^{p-1})$.

Stability considerations

When we speak of the numerical stability of a difference scheme such as (1-2-1) we are concerned with the behaviour of the errors inherent in the difference scheme itself. We must be careful not to confuse this with any properties (such as ill-conditioning) of the integral equation itself. Due to the liaison between Volterra

equations of the second kind and initial value problems it might be thought that the numerical stability of the scheme (1-2-1) could be dealt with in the same fashion as that for difference methods for solving initial value problems. However, although the theory for initial value problems has been studied in depth, the theory is fairly complex. See, for example, Henrici [9]. If we attempt to study the numerical stability of the scheme (1-2-1) in the same way we find the problems are compounded. For example, consider the simple initial value problem

$$y'(x) = y(x) \qquad \text{with} \quad y(0) = 1 ,$$

which has integral equation equivalent

$$y(x) = 1 + \int_0^x y(t)\, dt$$

and solution $y(x)=e^x$. For the initial problem, solution by a multistep method leads to the scheme

$$\sum_{j=0}^{k} \alpha_j Y_{n+j} = h \sum_{j=0}^{k} \beta_j Y_{n+j} , \qquad \ldots.(1\text{-}2\text{-}3)$$

where Y_j is the approximation to $y(x_j)$ by the multistep method; α_j and β_j are constants with $\alpha_k = 1$. The errors ε_j satisfy

$$\sum_{j=0}^{k} (\alpha_j - h\beta_j)\, \varepsilon_{n+j} = \phi_{h,n}$$

where $\phi_{h,n}$ is the truncation error of the method.

The homogeneous part of this equation has characteristic polynomial

$$p(z) = \sum_{j=0}^{k_1} (\alpha_j - h\beta_j) z^j .$$

If any of the zeros of p(z) lie outside the unit circle we find that the spurious components in the solution of the difference equation (1-2-3) (introduced since the order of the difference equation is higher than the original differential equation) may dominate the true solution. If any of the zeros of p(z) be outside the unit circle the method (1-2-3) is said to be unstable.[14]

For the Volterra equation, the difference scheme is

$$Y_n = 1 + h \sum_{i=0}^{n} \omega_{ni} Y_i$$

and the errors satisfy

$$\varepsilon_n = 1 + h \sum_{i=0}^{n} \omega_{ni} \varepsilon_i + R_{n,h} .$$

The homogeneous equation has characteristic polynomial

$$q(z) = z^n - 1 - h \sum_{i=0}^{n} \omega_{ni} z^i .$$

However, the weights w_{ni} vary with n, and the determination of the zeros of q(z) will not be straightforward even for this simple equation. The problem of numerical stability of the scheme (1-2-1) has therefore been reformulated by Linz [15] in a more tractable way.

Consider a small change g(x), say $\delta g(x)$, producing a small change in the solution from y(x) to $y(x) + \delta y(x)$.

Then

$$y(x) + \delta y(x) = g(x) + \delta g(x) + \int_a^x K(x,t, y(t) + \delta y(t)) \, dt$$

Subtracting this equation from (1-1-1) gives

$$\delta y(x) = \delta g(x) + \int_a^x \frac{\partial K}{\partial y}(x,t, y(t)) \, \delta y(t) \, dt \qquad \ldots\ldots(1\text{-}2\text{-}4)$$

if we neglect second order terms in $\delta y(t)$.

The equation (1-2-4) can be thought of as representing the condition of the problem as it reflects the sensitivity of the solution to small changes in g(x). Now discretization errors and starting errors can be considered as causing a perturbation in (1-1-1). For a numerically stable scheme we would desire that the errors be characterized by the same growth rate as $\delta y(x)$ in (1-2-4). If the errors grow at a faster rate than $\delta y(x)$ in (1-2-4), the scheme could not be considered stable.

This argument prompted Linz [15] into the following definition of numerical stability of the scheme (1-2-1).

Definition 1-3 Numerical stability.

If the error in the finite difference scheme (1-2-1) has as its dominant component $\varepsilon(x_n)$, say, where $\varepsilon(x)$ satisfies an equation of the form

$$\varepsilon(x) = G(x) + \int_a^x \frac{\partial K}{\partial y}(x,t, y(t)) \, \varepsilon(t) \, dt \qquad \ldots\ldots(1\text{-}2\text{-}5)$$

then the scheme (1-2-1) is said to be numerically stable.

Clearly, the errors could propagate at a rate slower than (1-2-5) and the scheme would still be stable. Also if (1-2-5) is satisfied then $\varepsilon(x_n)$ may still not be small. However, in this case the original equation

is badly conditioned and we cannot expect anything better. We conclude that if we obtain an asymptotic expansion for $\varepsilon(x_n)$, then for a numerically stable scheme we require that the leading term satisfies an equation of the form (1-2-5).

If the leading term satisfies an equation of the form

$$\varepsilon(x) = G(x) + \lambda \int_a^x \frac{\partial K}{\partial y}(x,t,y(t))\,\varepsilon(t)\,dt, \qquad \ldots(1\text{-}2\text{-}6)$$

where λ is a parameter introduced by the finite difference scheme, we may find that, for $\lambda \neq 1$, the scheme is stable for certain kernels $K(x,t,y(t))$. This situation is known as <u>weak instability</u>. We shall meet equations of the form (1-2-6) in Section (1-4).

1 - 3 Choice of quadrature and the repetition factor.

(i) Newton-Cotes rules

The simplest way in which we can approximate to the integral in (1-1-2) is to use a repeated Euler rule in each interval $[x_i, x_{i+1}]$, $i=0,\ldots,n$. We then obtain the explicit scheme for y_n

$$Y_n = g(x_n) + h \sum_{i=0}^{n-1} K(x_n, x_i, Y_i), \qquad n=1,2,\ldots N.$$

Similarly by using a repeated trapezoidal rule we obtain the implicit scheme

$$Y_n = g(x_n) + \frac{h}{2} \sum_{i=0}^{n-1} \{ K(x_n, x_i, Y_i) + K(x_n, x_{i+1}, Y_{i+1}) \}.$$

Open Newton-Cotes rules could be used to obtain explicit schemes; however, these rules appear to be unpopular, possibly because they are suspected of causing high rounding errors in computation.

For higher order closed Newton-Cotes rules the exact form of (1-2-1) will depend on the value of n. For example, consider the use of Simpson's rule. For n even, n=2r, say, we can apply Simpson's rule in a straightforward manner as follows

$$Y_{2r} = g(x_{2r}) + h/3 \sum_{i=0}^{r-1} \{ K(x_{2r}, x_{2i}, Y_{2i}) + 4K(x_{2r}, x_{2i+1}, Y_{2i+1}) + K(x_{2r}, x_{2i+2}, Y_{2i+2}) \}, \quad r=1,2,\ldots \qquad \ldots(1\text{-}3\text{-}1)$$

where we need to supply a starting value Y_1.

However, for nodd, n=2r+1, say, we need to supply an ancillary rule either over $\{x_0, x_1, x_2, x_3\}$ or over $\{x_{2r-2}, x_{2r-1}, x_{2r}, x_{2r+1}\}$. The three-eighths rule seems a natural choice. It has a local truncation error $O(h^5)$ (Simpson's rule in composite form has a truncation error $O(h^4)$). Placing the three-eighths rule at the lower end gives us

$$Y_{2r+1} = g(x_{2r+1}) + \frac{3h}{8} \{ K(x_{2r+1}, x_0, Y_0) + 3K(x_{2r+1}, x_1, Y_1) + 3K(x_{2r+1}, x_2, Y_2) + K(x_{2r+1}, x_3, Y_3) \} + h/3 \sum_{i=2}^{r} \{ K(x_{2r+1}, x_{2i-1}, Y_{2i-1}) + 4K(x_{2r+1}, x_{2i}, Y_{2i}) + K(x_{2r+1}, x_{2i+1}, Y_{2i+1}) \}. \qquad \ldots(1\text{-}3\text{-}2)$$

$r=1,2,\ldots$

We shall call the scheme given by (1-3-1) and (1-3-2) SIMPSON (i) [After Noble[16]]. Placing the three-eighths rule at the upper end gives us

$$Y_{2r+1} = g(x_{2r+1}) + h/3 \sum_{i=0}^{r-2} \{ K(x_{2r+1}, x_{2i}, Y_{2i}) + 4K(x_{2r+1}, x_{2i+1}, Y_{2i+1}) + K(x_{2r+1}, x_{2i+2}, Y_{2i+2}) \} + 3/8\, h \{ K(x_{2r+1}, x_{2r-2}, Y_{2r-2}) + 3K(x_{2r+1}, x_{2r-1}, Y_{2r-1}) + 3K(x_{2r+1}, x_{2r}, Y_{2r}) + K(x_{2r+1}, x_{2r+1}, Y_{2r+1}). \qquad \ldots(1\text{-}3\text{-}3)$$

$r=1,2,\ldots$

We shall call the scheme given by (1-3-1) and (1-3-3), SIMPSON (ii). We now show the relevance of the idea of a repetition factor introduced in Section (1-1). Compare the weights attached to Y_j, $j=2i,2i+1,\ldots$ in SIMPSON (i) and SIMPSON (ii) in the following tables.

TABLE 1: SIMPSON (i)

Weight for x	Y_{2i}	Y_{2i+1}	Y_{2i+2}	Y_{2i+3}	
$x=x_{2r}$	2/3	4/3	2/3	4/3	$i\geq 2$
$x=x_{2r+1}$	4/3	2/3	4/3	2/3	
$x=x_{2r+2}$	2/3	4/3	2/3	4/3	

TABLE 2: SIMPSON (ii)

Weight for x	Y_{2i}	Y_{2i+1}	Y_{2i+2}	Y_{2i+3}	
$x=x_{2r}$	2/3	4/3	2/3	4/3	$i\leq r-2$
$x=x_{2r+1}$	2/3	4/3	2/3	4/3	$i\leq r-3$
$x=x_{2r+2}$	2/3	4/3	2/3	4/3	$i\leq r-1$

For SIMPSON (i) we notice that a pattern in the weights is formed, and the pattern is repeated every alternate row. For SIMPSON (ii) the pattern is repeated every row. In accordance with the definition, then, the repetition factor for SIMPSON (i) is 2 and the repetition

factor for SIMPSON (ii) is 1.

Other Newton-Cotes schemes. Clearly, other Newton-Cotes rules could be combined in a similar way, but more than one ancillary quadrature would need to be introduced in most cases. The combinations considered above are probably the best for most practical purposes.

(ii) Gregory quadrature.

The Gregory rule for a function f(x) is

$$\int_{x_0}^{x_n} f(x)\,dx \doteq h\{\tfrac{1}{2}f(x_0) + f(x_1) + \cdots + f(x_{n-1}) + \tfrac{1}{2}f(x_n)\}$$
$$- h/12\{\nabla f(x_n) - \Delta f(x_0)\} - h/24\{\nabla^2 f(x_n) + \Delta^2 f(x_0)\}$$
$$- 19/720 h\{\nabla^3 f(x_n) - \Delta^3 f(x_0)\} - \ldots$$

(See Phillips and Taylor [21].)

The rule can be rewritten as

$$\int_{x_0}^{x_n} f(x)\,dx \doteq \sum_{r=0}^{n} f(x_r) + h\sum_{j=0}^{k} A_j^k\{f(x_j) + f(x_{n-j})\}.$$

The truncation error, $E_{n,k}(f)$, is given by

$$E_{n,k}(f) = h^{k+3}(b_k - (n-k)a_{k+1})f^{(k+2)}(\xi_1), \quad x_0 \le \xi_1 \le x_n, \quad \ldots(1\text{-}3\text{-}4(i))$$

when k is even, and

$$E_{n,k}(f) = h^{k+2}c_k f^{(k+1)}(\xi_2) + h^{k+3}(n-k)a_{k+1}f^{(k+2)}(\xi_3), \quad \ldots(1\text{-}3\text{-}4(ii))$$
$$x_0 \le \xi_2, \xi_3 \le x_n$$

when k is odd, where

$$A_j^k = \sum_{i=j}^{k} a_i \binom{i}{i-j}(-1)^{i-j}, \quad a_i = -\int_0^1 \binom{s}{i+1}ds$$

$$b_k = \frac{1}{(k+2)!}\int_0^k (s - k/2)s(s-1)\ldots(s-k)\,ds,$$

$$c_k = \int_0^1 \binom{s}{k+1}ds.$$

The advantage of using Gregory quadrature in (1-2-1) is that we can apply the rule directly for any value of n. We notice that the rule requires k-1 starting values, with $f(x_o)$ known.

We obtain the scheme

$$Y_n = g(x_n) + h\sum_{r=0}^{n} K(x_n, x_r, Y_r) + h\sum_{j=0}^{k} A_j^k \{K(x_n, x_j, Y_j) + K(x_n, x_{n-j}, Y_{n-j})\},$$

$$n = K, K+1, \ldots, N. \qquad \ldots.(1\text{-}3\text{-}5)$$

The scheme is convergent of order (k+2), and we notice that in view of (1-1-5), the iteration (1-1-4) will converge provided

$$h(1 + A_0^k)L < 1$$

We find the scheme has repetition factor 1, with the repetition extending over the first (n-k-1) values of Y_i. Table 3 shows the form of the repetition for $n \geq 2k+1$.

TABLE 3: THE GREGORY RULE AND THE REPETITION FACTOR

Weight for x	Y_o	Y_1		Y_k	Y_{k+1}	Y_{k+2}	Y_{k+3}
$x=x_{2k+1}$	$1+A_o^k$	$1+A_1^k$	....	$1+A_k^k$	$1+A_{k-1}^k$	$1+A_{k-2}^k$	$1+A_{k-3}^k$
$x=x_{2k+2}$	$1+A_o^k$	$1+A_1^k$	....	$1+A_k^k$	1	$1+A_k^k$	$1+A_{k-1}^k$
$x=x_{2k+3}$	$1+A_o^k$	$1+A_1^1$	....	$1+A_k^k$	1	1	$1+A_k^k$

If we subtract (1-3-5) from the analogous equation involving exact values of y(x) we obtain

$$\varepsilon_n = h \sum_{r=0}^{n} \{ K(x_n, x_r, y(x_r)) - K(x_n, x_r, Y_r) \} + h \sum_{j=0}^{K} A_j^K \{ K(x_n, x_j, y(x_j)) - K(x_n, x_j, Y_j) \} + h \sum_{j=0}^{K} A_j^K \{ K(x_n, x_{n-j}, y(x_{n-j}) - K(x_n, x_{n-j}, Y_{n-j}) \} + E_{n,K}(K)$$

Phillips [20] obtains the following error bound for ε_n

$$|\varepsilon_n| \le \delta e^{2(x_n - x_0)L} + h^{K+2} M_{K+2} |a_{K+1}| \{ e^{2(x_n - x_0)L} - 1 \} / L \quad \ldots . (1\text{-}3\text{-}6)$$

where

$$|\varepsilon_j| \le \delta \quad , \quad 1 \le j \le K-1$$

and

$$\max_{x_n} \max_{a \le t \le b} \left| \frac{\partial^{K+2}}{\partial t^{K+2}} K(x_n, t, y(t)) \right| = M_{K+2}$$

The bound (1-3-6) is only of practical use if we know an a priori value of M_{k+2}. For some applications, an estimate may be obtainable.

We applied the Gregory rule to the equation

$$y(x) = \cos x - \sin x + 2 \int_0^x \sin(x-t)\, y(t)\, dt$$

in $o \le x \le 1$, with k=4, h=0.05, and <u>exact</u> starting values Y_1, Y_2, Y_3.

Some numerical results are listed in Table 4 together with values of the error bound (1-3-6).

Notice that for this equation $K(t,t)=0$, and so the scheme is explicit.

TABLE 4: NUMERICAL RESULTS FOR THE GREGORY RULE

Value of x	Error in Computed Solution	Value of Error Bound
0.20	0.75×10^{-10}	0.85×10^{-9}
0.30	0.36×10^{-9}	0.14×10^{-8}
0.40	0.63×10^{-9}	0.20×10^{-8}
0.50	0.87×10^{-9}	0.28×10^{-8}
0.60	0.11×10^{-9}	0.37×10^{-8}
0.70	0.13×10^{-9}	0.48×10^{-8}
0.80	0.15×10^{-9}	0.60×10^{-8}
0.90	0.17×10^{-9}	0.75×10^{-8}
1.00	0.19×10^{-9}	0.93×10^{-8}

(iii) Gaussian quadrature

If values of y(x) are required at the equally spaced points $x_n=x_o+nh, n=0,\ldots,N$, the only Gaussian rule applicable is the Lobatto rule. All Gaussian rules have the favourable property that the truncation error is $O(h^{2m-1})$, where m is the number of points employed in the rule.

If we apply the four-point Lobatto rule to the integral in (1-1-2), that is, in each interval $[x_i, x_{i+1}], i=0,\ldots,N$, we obtain

$$Y_n = g(x_n) + h/12 \sum_{i=0}^{n-1} \{K(x_n, x_i, Y_i) + 5K(x_n, x_{i+r}, Y_{i+r}) + 5K(x_n, x_{i+s}, Y_{i+s}) + K(x_n, x_{i+1}, Y_{i+1})\}, \quad n=1,2,\ldots \qquad \ldots(1\text{-}3\text{-}7)$$

where $x_{i+r} = x_i + (5-\sqrt{5})/10\, h$, $x_{i+s} = x_i + (5+\sqrt{5})/10\, h$.

We immediately observe that we do not know the values Y_{i+r}, Y_{i+s}, $i=0,\ldots,n-1$ and that we require starting values Y_r, Y_s. Campbell and Day [4] suggest a method of determining Y_r, Y_s to $O(h^6)$, and Jaim and Sharma [12] have proposed a method of finding Y_{i+r}, Y_{i+s} in terms of Y_i and certain kernel values. We do not examine these methods here, but merely comment that the amount of computation required to obtain these values is large and the use of the rule is not justified. Equally accurate solutions can be obtained with far less effort by using Gregory's rule, for example.

1 - 4 Numerical Stability and the repetition factor

(i) The simple schemes

In this subsection we obtain an asymptotic error estimate for the scheme (1-2-1) when we use a repeated trapezoidal rule. The scheme is

$$Y_n = g(x_n) + h/2 \sum_{i=0}^{n-1} \{ K(x_n, x_i, Y_i) + K(x_n, x_{i+1}, Y_{i+1}) \}$$

Thus we obtain

$$\varepsilon_n = h/2 \sum_{i=0}^{n-1} \{ (\partial K/\partial y)_i \, \varepsilon_i + (\partial K/\partial y)_{i+1} \, \varepsilon_{i+1} \} + R_{n,h},$$

where

$$(\partial K/\partial y)_i = \partial K/\partial y \, (x_n, x_i, y(x))_{x=x_i}$$

and

$$R_{n,h} = -h^2/12 \{ dK/dt \, (x_n, t, y(t))_{t=x_n} - dK/dt \, (x_n, t, y(t))_{t=x_0} \} + O(h^3).$$

This result follows from the Euler-Maclaurin sum formula and the fact that each $\varepsilon_i = O(h^2)$.

On introducing the scaled error $e_n = \varepsilon_n / h^2$ we obtain

$$e_n = h/2 \sum_{i=0}^{n-1} \{ (\partial K/\partial y)_i e_i + (\partial K/\partial y)_{i+1} e_{i+1} \}$$
$$- 1/12 \{ \frac{dK}{dt} (x_n, t, y(t))_{t=x_n} - \frac{dK}{dt} (x_n, t, y(t))_{t=x_0} + O(h).$$

If we now consider the integral equation

$$e(x) = \int_a^x \partial K/\partial y \ (x, t, y(t)) e(t) \, dt + q(x)$$

where $q(x) = -1/12 \{ dK/dt \ (x, t, y(t))_{t=x} - dK/dt \ (x, t, y(t))_{t=x_0}$

Then e_n differs from $e(x_n)$ by terms of $O(h)$ only. Thus

$$e_n = e(x_n) + O(h)$$

so that

$$\varepsilon_n = h^2 e(x_n) + O(h^3).$$

Now $e(x_n)$ satisfies an integral equation of the form (1-2-5). Thus, ignoring the effects of rounding errors in computations, we see that the trapezoidal rule is <u>stable</u>. The same result holds for the Euler method.

(ii) <u>Higher-order methods</u>

Unfortunately this argument will not hold for higher order methods. The presence of starting errors (and in practise rounding errors) may cause instability. We can no longer identify ε_n with $e(x_n)$ as in (i). A different approach to the problem due to Noble [16],

which includes the effect of starting errors has been tried and has produced some interesting results. Noble presents his theory applied to the schemes SIMPSON (i) and SIMPSON (ii) and, attaining results, he then generalises the theory so that it is applicable to all multistep methods. We now give an outline of the theory in the same way, for, as will be seen, it is easier to deal with a specific example first and then generalise.

We can obtain equations for the errors ε_i, $i=1,\ldots,N$ in SIMPSON (i) as follows:

$$\varepsilon_{2r} = u_{2r} + 1/3\, h K_{2r,2}\, \varepsilon_2 + 1/3\, h \sum_{i=1}^{r-1} \{ K_{2r,2i}\, \varepsilon_{2i} + 4 K_{2r,2i+1}\, \varepsilon_{2i+1} + K_{2r,2i+2}\, \varepsilon_{2i+2} \} \quad \ldots(1\text{-}4\text{-}1)$$

$$\varepsilon_{2r+1} = V_{2r+1} + 9/8\, h K_{2r+1,2}\, \varepsilon_2 + 3/8\, h K_{2r+1,3}\, \varepsilon_3 + 1/3\, h \sum_{i=2}^{r-1} \{ K_{2r+1,2i-1}\, \varepsilon_{2i-1} + 4 K_{2r+1,2i}\, \varepsilon_{2i} + K_{2r+1,2i+1}\, \varepsilon_{2i+1} \} \quad \ldots(1\text{-}4\text{-}2)$$

with

$$u_{2r} = 4/3\, h K_{2r,1}\, \varepsilon_1 + R_{2r}; \quad V_{2r+1} = 9/8\, K_{2r+1,1}\, \varepsilon_1 + R_{2r+1}.$$

For Simpson (ii) we have ε_{2r} as above and

$$\varepsilon_{2r+1} = W_{2r+1} + h/3 \sum_{i=1}^{r-2} \{ K_{2r+1,2i}\, \varepsilon_{2i} + 4 K_{2r+1,2i+1}\, \varepsilon_{2i+1} + K_{2r+1,2i+2}\, \varepsilon_{2i+2} + 3/8\, h \{ K_{2r+1,2r-2}\, \varepsilon_{2r-2} + 3 K_{2r+1,2r-1}\, \varepsilon_{2r-1} + 3 K_{2r+1,2r}\, \varepsilon_{2r} + K_{2r,2r+1,}\, \varepsilon_{2r+1} \} \quad \ldots(1\text{-}4\text{-}3)$$

with $W_{2r+1} = 4/3\, h K_{2r+1,1}\, \varepsilon_1 + R_{2r+1}$

We have used the notation

$$K_{ij} = \partial K / \partial y\, (x_i, x_j, y(x))_{x=x_j}$$

and R_{2r}, R_{2r+1} refer to terms involving truncation errors and terms involving ε_i^2. The terms U_{2r}, V_{2r+1}, W_{2r+1} have been isolated since they contain the starting errors. Now suppose that U_{2r}, V_{2r+1}, W_{2r+1} vary

smoothly with r. Then ε_{2r} varies "smoothly" with r and ε_{2r+1} varies "smoothly" with r, but in general ε_r may not vary "smoothly" with r. This argument led Noble to treat ε_{2r}, ε_{2r+1} as arising from different functions, as follows

Let

$$\varepsilon_{2r} = x_1(t_{2r}) + O(h) , \qquad \ldots.(1\text{-}4\text{-}4)$$

$$\varepsilon_{2r+1} = y_1(t_{2r+1}) + O(h) , \qquad \ldots.(1\text{-}4\text{-}5)$$

where $x_1(t)$ and $y_1(t)$ are continuously differentiable functions. Also, suppose there exist continuously differentiable functions U(t), V(t), W(t) such that

$$u_{2r} = u(t_{2r}) + O(h),$$

$$v_{2r+1} = v(t_{2r+1}) + O(h),$$

$$w_{2r+1} = w(t_{2r+1}) + O(h).$$

Now let us consider SIMPSON(i). From (1-4-1) and (1-4-2) we can write

$$x_1(t_{2r}) = u(t_{2r}) + \tfrac{1}{3}(S_1 + 2S_2) + O(h), \qquad \ldots(1\text{-}4\text{-}6)$$

$$y_1(t_{2r+1}) = V(t_{2r+1}) + \tfrac{1}{3}(2S_3 + S_4) + O(h) \qquad \ldots(1\text{-}4\text{-}7)$$

where

$$S_1 = 2h\{K_{2r,2}\,x_1(t_2) + K_{2r,4}\,x_1(t_4) + \ldots + K_{2r,2r-2}\,x_1(t_{2r-2}) + \tfrac{1}{2}K_{2r,2r}\,x_1(t_{2r})\}$$

$$= \int_a^{t_{2r}} \partial K/\partial y\,(x_{2r}, t, y(t))\,x_1(t)\,dt + O(h)$$

$$= I(t_{2r}) + O(h).$$

Here we have considered all the terms in S_1, except the last, as arising from an Euler approximation to the integral $I(t_{2r})$ with step-size 2h. The discrepancy from the true Euler approximation is O(h). Similarly

$$S_2 = J(t_{2r}) + O(h),$$

where

$$J(t_{2r}) = \int_0^{t_{2r}} \partial K/\partial y\,(x_{2r}, t, y(t))\, y_1(t)\, dt.$$

We find that

$$S_3 = I(t_{2r}) + O(h),$$

$$S_4 = J(t_{2r+1}) + O(h).$$

From (1-4-4) and (1-4-5) it follows that

$$x_1(s) = u(s) + 1/3 \int_a^s \partial K/\partial y\,(s,t,y(t))\, x_1(t)\, dt + 2/3 \int_a^s \partial K/\partial y\,(s,t,y(t))\, y_1(t)\, dt \quad \ldots(1\text{-}4\text{-}8)$$

$$y_1(s) = v(s) + 2/3 \int_a^s \partial K/\partial y\,(s,t,y(t))\, x_1(t)\, dt + 1/3 \int_a^s \partial K/\partial y\,(s,t,y(t))\, y_1(t)\, dt$$

If we introduce

$$z_1(s) = \alpha\, x_1(s) + \beta\, y_1(s)$$

Then

$$z_1(s) = q_1(s) + \lambda\alpha \int_a^s \partial K/\partial y\,(s,t,y(t))\, x_1(t)\, dt + \lambda\beta \int_a^s \partial K/\partial y\,(s,t,y(t))\, y_1(t)\, dt$$

$$= q_1(s) + \lambda \int_a^s \partial K/\partial y\,(s,t,y(t))\, z_1(t)\, dt.$$

That is, we have uncoupled the equations (1-4-8), provided the constant λ can be found such that

$$1/3\,\alpha + 2/3\,\beta = \lambda\alpha,$$

$$2/3\,\alpha + 1/3\,\beta = \lambda\beta, \quad \ldots.(1\text{-}4\text{-}9)$$

with

$$q_1(s) = \alpha\, u(s) + \beta\, v(s).$$

Equations (1-4-9) can be written as the eigenvalue problem

$$A\,\underset{\sim}{\alpha} = \lambda\,\underset{\sim}{\alpha}, \quad \ldots(1\text{-}4\text{-}10)$$

with

$$A = \begin{bmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}, \quad \underset{\sim}{\alpha} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

The eigenvalues and corresponding eigenvectors are found to be

$$\lambda_1 = 1, \quad \underset{\sim}{a}_1 = [1 \; 1]^T,$$

$$\lambda_2 = -1/3, \quad \underset{\sim}{a}_2 = [1 \; -1]^T.$$

The integral equations satisfied by

$$z_1(s) = x_1(s) + y_1(s), \quad z_2(s) = x(s) - y(s)$$

are

$$z_1(s) = U(s) + V(s) + \int_a^s \frac{\partial K}{\partial y}(s, t, y(t))\, z_1(t)\, dt, \qquad \ldots.(1\text{-}4\text{-}11)$$

$$z_2(s) = U(s) - V(s) - 1/3 \int_a^s \frac{\partial K}{\partial y}(s, t, y(t))\, z_2(t)\, dt. \qquad \ldots.(1\text{-}4\text{-}12)$$

Examination of U_{2r} and V_{2r+1} in (1-4-2) shows that U(s)-V(s) is non-zero even for small h. Thus with (1-4-4) and (1-4-5) we find that

$$\varepsilon_n = 1/2 \{ z_1(x_n) + (-1)^n z_2(x_n) \} + O(h).$$

We see that the dominant part of the error contains a component $z_1(x_n)$ which satisfies an integral equation (1-4-11) of the desired form, but that it also contains a spurious oscillating component $(-1)^n z_2(x_n)$ where z_2 satisfies (1-4-12). For certain kernels the behaviour of $z_2(x)$ could be different from $z_1(x)$ and instability would result. We therefore term SIMPSON (i) <u>weakly unstable</u>.

Let us now examine SIMPSON (ii)

The equations analogous to (1-4-8) are

$$x_1(s) = U(s) + 1/3 \int_a^s \partial K/\partial y\,(s, t, y(t))\, x_1(t)\, dt + 2/3 \int_a^s \partial K/\partial y\,(s, t, y(t))\, y_1(t)\, dt$$

$$y_1(s) = V(s) + 1/3 \int_a^s \partial K/\partial y\,(s, t, y(t))\, x_1(t)\, dt + 2/3 \int_a^s \partial K/\partial y\,(s, t, y(t))\, y_1(t)\, dt \qquad \ldots.(1\text{-}4\text{-}13)$$

(We have used the same notation for convenience).

The matrix A is now $A = \begin{bmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{bmatrix}$

with eigenvalues and corresponding eigenvectors

$\lambda_1 = 1 \qquad \underset{\sim}{\alpha}_1 = [1 \quad 2]^T$

$\lambda_2 = 0 \qquad \underset{\sim}{\alpha}_2 = [1 \quad -1]^T.$

The equations analogous to (1-4-11), (1-4-12) are

$$z_1(s) = u(s) + 2\,w(s) + \int_a^s \partial k/\partial y\,(s,t,y(t))\,z_1(t)\,dt, \quad \ldots.(1\text{-}4\text{-}14)$$

$$z_2(s) = u(s) - w(s). \quad \ldots.(1\text{-}4\text{-}15)$$

Inspection of (1-4-1) and (1-4-3) shows that for, sufficiently small h, $U_{2r} \simeq W_{2r+1}$, so that $U(S) \simeq W(S)$ and hence $x_1(S) \simeq y_1(s)$.
We find that

$$\varepsilon_r = 1/3\; z_1(x_r) + O(h).$$

Now $z_1(x)$ satisfies (1-4-14) and we see, in contrast to SIMPSON (i) that there is no spurious component in the error term which could cause instability. We conclude that SIMPSON (ii) is numerically stable.

The generalised theory.

Suppose we are using a repeated k-step quadrature formula as the main rule, with weights $w_0, w_1, \ldots, w_k$. The generalisation of (1-4-3) and (1-4-13) is

$$\underset{\sim}{x}(s) = \underset{\sim}{u}(s) + \int_a^s \partial k/\partial y\,(s,t,y(t))\;C\;\underset{\sim}{x}(t)\,dt, \quad \ldots.(1\text{-}4\text{-}16)$$

where the vectors $\underset{\sim}{x}$ and $\underset{\sim}{u}$ contain k elements, and C is a kxk matrix whose first row is

$$1/k \{ (w_0 + w_k), w_1, \ldots, w_{k-1} \}$$

If the method has a repetition factor 1, then all the rows of C are the same. If the method has a repetition factor greater than one, then the remaining rows of C will contain the same elements as the first row but each element will be shifted in position an appropriate number of places. Normally, as in SIMPSON(i), the shift is one place to the right per row.

From (1-4-16) we find that the matrix analogous to A in (1-4-10) is C^T. (1-4-10) can then be generalised to give

$$C^T P^T = P^T \Lambda$$

where the columns of P^T are the eigenvectors $P_1, \ldots, P_k$ of C^T and the matrix Λ is diagonal having as elements the eigenvalues $\lambda_1, \ldots, \lambda_k$ of C^T. Premultiplying (1-4-16) by P and denoting

$$P\underset{\sim}{x}(s) = \underset{\sim}{z}(s)$$

$$P\underset{\sim}{u}(s) = \underset{\sim}{\mu}(s).$$

we find

$$\underset{\sim}{z}(s) = \underset{\sim}{\mu}(s) + \int_a^s \partial k/\partial y \,(s,t,y(t)) \Lambda \underset{\sim}{z}(t)\, dt,$$

which is analogous to (1-4-11) and (1-4-12), and (1-4-14) and (1-4-15). Explicitly, we have

$$z_i(s) = \mu_i(s) + \lambda_i \int_a^s \partial k/\partial y \,(s,t,y(t)) z_i(t)\, dt \qquad \ldots.(1\text{-}4\text{-}17)$$

$$i = 1, \ldots, k.$$

Now, since $\sum_{j=0}^{K} \omega_j/K = 1$, the sum of the elements in each column of C^T is unity, and so at least one eigenvalue of C^T is unity. At least one equation in (1-4-17) therefore has the same form as the perturbation equation (1-2-5), as we should expect. The stability of the scheme will therefore be governed by the size of the eigenvalues in the remaining k-1 equations of (1-4-17). If the repetition factor is equal to 1, we find that all the columns of C^T are the same. The k-1 remaining eigenvalues of C^T are therefore zero.

Thus (1-4-17) reduces to

$$z_1(s) = \mu_1(s) + \int_a^s \partial K/\partial y\,(s,t,y(t))\, z_1(t)\, dt,$$

$$z_i(s) = \mu_i(s) \quad , \qquad i = 2, \ldots, K.$$

Now $\mu_i(s)$, $i = 2, \ldots, K$ depend only on the starting errors and truncatio errors, that is, they will be small. This in turn makes $z_i(S)$, i=2,...,k, small. We see that the dominant part of the errors will behave in the same manner as $z_1(S)$, and we conclude that methods with repetition factor one are stable. Note that the Gregory rule will therefore always produce a stable scheme.

An Example: The midpoint method

The midpoint rule applied in intervals of length $2h$ can produce a scheme with a repetition factor of two as follows:

$$Y_{2r} = g(x_{2r}) + 2h \sum_{i=0}^{r-1} K(x_{2r}, x_{2i+1}, Y_{2i+1}) \quad , \qquad r=1,2,\ldots,$$

$$Y_{2r+1} = g(x_{2r+1}) + h/2 \{ K(x_{2r+1}, x_0, Y_0) + K(x_{2r+1}, x_1, Y_1)$$

$$+ 2h \sum_{i=0}^{r} K(x_{2r+1}, x_{2i}, Y_{2i}) \quad , \qquad r=1,2,\ldots ,$$

We note that this scheme has the advantage of being explicit.

The matrix $C^T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

with eigenvalues and eigenvectors

$$\lambda_1 = 1, \quad \underset{\sim}{\alpha}_1 = [1, 1]^T,$$

$$\lambda_2 = -1, \quad \underset{\sim}{\alpha}_2 = [1 \; -1]^T,$$

thus

$$z_1(s) = u(s) + v(s) + \int_a^s \partial k/\partial y \; (s, t, y(t))\, z_1(t)\, dt,$$

$$z_2(s) = u(s) - v(s) - \int_a^s \partial k/\partial y \; (s, t, y(t))\, z_2(t)\, dt,$$

and

$$\varepsilon_r = z_1(x_r) + (-1)^r z_2(x_r) + O(h).$$

A numerical example

We consider the simple equation

$$y(x) = 1 - \int_0^x y(t)\, dt \qquad \ldots(1\text{-}4\text{-}18)$$

with solution $y(x) = e^{-x}$.

If we first examine SIMPSON(i), then equation (1-4-12) becomes

$$z_2(x) = u(x) - v(x) + 1/3 \int_0^x z_2(t)\, dt.$$

with solution

$$z_2(x) = \{u(0) - v(0)\} e^{1/3 x} + e^{-1/3 x} \int_0^x e^{1/3 t} \, d/dt \, (u(t) - v(t))\, dt$$

$$= q(x) + \{u(0) - v(0)\} e^{1/3 x}$$

The weak positive exponential factor in $z_2(x)$ will cause a certain amount of instability. The degree of instability will depend on whether starting errors are present.

For the midpoint method we find $z_2(x) = \tau(x) + \{u(0) - v(0)\} e^x$

Here, $z_2(x)$ is growing in completely the opposite way to $z_1(x)$ and could eventually dominate. Severe instability will result.

Numerical results for equation (1-4-18) were computed using SIMPSON(i) and SIMPSON(ii). The starting value Y_1 was computed using a special starting procedure, which will be discussed in Section 1-5. The error in this starting value is $O(h^3)$. A step-size h=0.1 was taken and the results are displayed in Table 5.

TABLE 5: NUMERICAL INSTABILITY WITH SIMPSON'S RULE

Value of x	Error in SIMPSON(i)	Error in SIMPSON(ii)
2.0	0.66×10^{-5}	0.37×10^{-5}
2.1	-0.35×10^{-6}	0.34×10^{-5}
2.2	0.62×10^{-5}	0.30×10^{-5}
2.3	-0.11×10^{-5}	0.28×10^{-5}
2.4	0.59×10^{-5}	0.25×10^{-5}
2.5	-0.18×10^{-5}	0.23×10^{-5}
2.6	0.58×10^{-5}	0.20×10^{-5}
2.7	-0.24×10^{-5}	0.19×10^{-5}
2.8	0.58×10^{-5}	0.17×10^{-5}
2.9	-0.30×10^{-5}	0.15×10^{-5}
3.0	0.58×10^{-5}	0.14×10^{-5}

The instability in SIMPSON(i) was not apparent until x=2.0, thus errors in the two schemes are listed above from x=2.0 to x=3.0. It is interesting to note that at this stage the moduli of the errors in SIMPSON(i) are not significantly worse than those of SIMPSON(ii).

For larger values of x, however, the errors in SIMPSON(i) were found to increase more rapidly, whilst the errors in SIMPSON (ii) continued to decrease.

Comments

The arguments used by Noble in his theory are heuristic - for example there is no rigorous justification for considering ε_{2r} and ε_{2r+1} as being derived from separate functions [equations (1-4-6) and (1-4-7)]. However, his results agree with previous work, where the behaviour of the error has been studied by difference methods for the equation (1-4-18)[see Linz [15]]. Numerical work verifies the theory too.

Accepting the theory, we obtain the interesting result that the numerical stability of a scheme is related to its repetition factor. It is curious that the position of an ancillary quadrature can be vitally important.

1 - 5 Some methods for finding starting values

We noted in Section (1-1) that if the quadrature used in (1-2-1) is of accuracy $O(h^p)$, then we only require the starting values to be accurate to $O(h^{p-1})$ to obtain an overall accuracy in the solution of $O(h^p)$. If we are using an n-step repeated Newton-Cotes rule we will require n-1 starting values, and if we are using Gregory quadrature we will require k-1 starting values.

Suppose we are using the Gregory rule with k=2. The scheme is of accuracy $O(h^5)$ if we obtain the starting value Y_1 to $o(h^4)$. A possible scheme is obtaining Y_1 is the following:

$$Y_1^{(0)} = g(x_1) + hK(x_1, x_0, Y_0) \quad , \qquad [\, y(x_1) - Y_1^{(0)} = o(h^2) \,]$$

$$Y_{1/2}^{(0)} = g(x_{1/2}) + \tfrac{1}{2} hK(x_{1/2}, x_0, Y_0) \quad , \qquad [\, y(x_{1/2}) - Y_{1/2}^{(0)} = o(h^2) \,]$$

$$Y_1^{(1)} = g(x_1) + \tfrac{1}{2} h \{ K(x_1, x_0, Y_0) + K(x_1, x_1, Y_1^{(0)}) \}, \quad [\, y(x_1) - Y_1^{(1)} = o(h^3) \,]$$

$$Y_{1/2}^{(1)} = g(x_{1/2}) + \tfrac{1}{4} h \{ K(x_{1/2}, x_0, Y_0) + K(x_{1/2}, x_{1/2}, Y_{1/2}^{(0)}) \}, \quad [\, y(x_{1/2}) - Y_{1/2}^{(1)} = o(h^3) \,]$$

Finally we obtain

$$Y_1 = g(x_1) + \tfrac{1}{6} h \{ K(x_1, x_0, Y_0) + 4K(x_1, x_{1/2}, Y_{1/2}^{(1)}) + K(x_1, x_1, Y_1^{(1)}) \},$$

$$[\, y(x_1) - Y_1 = o(h^4) \,].$$

where $x_{\frac{1}{2}} = x_0 + h/2$ and $Y_{\frac{1}{2}}$ is the approximation to $y(x^{\frac{1}{2}})$.

If we are using the Gregory rule with k=3 we will need an additional starting value Y_2 to an accuracy $O(h^4)$. This value could be obtained in the following way:

$$Y_2^{(0)} = g(x_2) + 2h K(x_2, x_1, Y_1^{(1)}) \qquad [\, y(x_2) - Y_2^{(0)} = o(h^3) \,]$$

$$Y_2 = g(x_2) + h/3 \{ K(x_1, x_0, Y_0) + 4K(x_2, x_1, Y_1) + K(x_2, x_2, Y_2^{(0)}) \}$$

$$[\, y(x_2) - Y_2 = o(h^4) \,]$$

The value $Y_1^{(1)}$ could be used as the starting value required by a repeated Simpson rule, and indeed this value was used in the computation of the numerical example in Section 1 - 4. The above methods were adapted from Noble's starting scheme [16]. In a series of papers, Day [7] gives some higher order procedures developed in the same way as above.

The first paper in the series is [7], and the reader is referred to this for more details.

Another method which can be employed under certain circumstances makes use of a Taylor series expansion. Suppose y(x) can be expanded in a Taylor series about the point x=a:

$$y(x) = y(a) + (x-a)y'(a) + \ldots + \frac{(x-a)^r}{r!} y^{(r)}(\xi) \qquad \ldots(1\text{-}5\text{-}1)$$

where

$$a \leq \xi \leq x.$$

Differentiating the integral equation (1-1-1) with respect to x we obtain

$$y^{(n)}(x) = g^{(n)}(x) + \sum_{s=0}^{n-1} \left\{ \frac{d^{n-1-s}}{dx^{n-1-s}} K^{(s)}(x,x,y(x)) \right\} + \int_a^x K^{(n)}(x,t,y(t)) \, dt \qquad \ldots.(1\text{-}5\text{-}2)$$

where

$$K^{(s)}(x,t,y(t)) = \partial^s K / \partial x^s \, (x,t,y(t))$$

Here we have assumed that g(x) and K(x,t,y(t)) are sufficiently differentiable. Putting x=a in (1-5-2) we obtain

$$y^{(n)}(a) = g^{(n)}(a) + \sum_{s=0}^{n-1} \left\{ \frac{d^{n-1-s}}{dx^{n-1-s}} K^{(s)}(x,x,y(x)) \right\}_{x=a} \qquad \ldots(1\text{-}5\text{-}3)$$

We can therefore obtain the values of $y^{(n)}(a)$, n=1,2,..., in succession from (1-5-3) and values of y(x) at x=h,2h,... can then be found from

substitution into (1-5-1) to as many terms as required for a given accuracy. This method will obviously only be practical if $K(x,t,y(t))$ and $g(x)$ are known analytically, and then only if the derivatives in (1-5-3) exist.

Other methods

In the following two sections we discuss two finite difference methods which are self-starting. These methods provide a simpler way of determining starting values for the scheme (1-2-1), and, by using a block-by-block method of section 2, we can obtain starting values to an arbitrarily high degree of accuracy.

Section 2: BLOCK-BY-BLOCK METHODS

2 - 1 Introduction

This method was first suggested by Young [27] and has been extended by several authors including Linz [15], O'Neill and Byrne [17] and Campbell and Day [5]. Weiss [25] gives a complete analysis of the method but we shall not follow his theory here. Instead we present an easier and more natural account of the method. By choice of suitable quadrature formulae the block-by-block method enables us to obtain a 'block' of values at each stage of the process and the method has the advantage of requiring no starting values.

We first need to develop a set of quadrature formulae. Consider the interpolating polynomial for $F(t)$ where,

$$F(t) = F(x_0) + \binom{s}{1} \Delta F(x_0) + \cdots + \binom{s}{p} \Delta^p F(x_0) + h^{p+1} \binom{s}{p+1} F^{(p+1)}(\eta_s),$$

where $t = x_0 + sh$ and $x_0 \le \eta_s \le x_0 + sh$.

Integrating over $[x_0, x_i]$, where $x_i = x_0 + ih$, we obtain

$$\int_{x_0}^{x_i} F(t)\,dt = h \sum_{j=0}^{p} d_{ij} \Delta^j F(x_0) + h^{p+2} \int_0^i \binom{s}{p+1} F^{(p+1)}(\eta_s)\,ds, \quad \ldots\ldots (2\text{-}1\text{-}1)$$

where

$$d_{ij} = \int_0^i \binom{s}{j} ds.$$

(2-1-1) can be written as

$$\int_{x_0}^{x_i} F(t)\,dt = h\sum_{j=0}^{p} c_{ij}\,F(x_j) + h^{p+2}\int_0^i \binom{s}{p+1} F^{(p+1)}(\eta_s)\,ds$$

where

$$c_{ij} = \sum_{r=j}^{p} (-1)^{r+j}\binom{r}{j} dir \qquad \ldots.(2\text{-}1\text{-}2)$$

For i=p (2-1-2) becomes the closed Newton Cotes rules. We shall refer to the quadrature (2-1-2) as the interpolatory rule.

2 - 2: The Block Method

In the integral equation (1-1-1) we choose a positive integer p and write $x=x_{mp+i}$. (1-1-1) can then be written as:

$$y(x_{mp+i}) = g(x_{mp+i}) + \int_{x_0}^{x_{mp}} K(x_{mp+i}, t, y(t))\,dt + \int_{x_{mp}}^{x_{mp+i}} K(x_{mp+i}, t, y(t))\,dt \qquad \ldots.(2\text{-}2\text{-}1)$$

where m is a non-negative integer and $1\le i\le p$.

The intention is to determine a block of values $Y_{mp+1},\ldots, Y_{mp+p}$ as approximations in the solution of the integral equation at x_{mp+1}, $\ldots$, x_{mp+p}. Suppose we replace the first integral in (2-2-1) by a convergent quadrature with weights w_j, j=o,...,mp. and the second integral by the interpolatory rule.

We then obtain

$$Y_{mp+i} = g(x_{mp+i}) + h\sum_{j=0}^{mp} \omega_j\,K(x_{mp+i}, x_j, Y_j) + h\sum_{j=0}^{p} c_{ij}\,K(x_{mp+i}, x_{mp+j}, Y_{mp+j}), \qquad \ldots(2\text{-}2\text{-}2)$$

for i=1,...,p and m=o,1,...

Now if $\{Y_1,\ldots, Y_p\},\{Y_{p+1},\ldots,Y_{2p}\},\ldots,\{Y_{mp-p+1},\ldots,Y_{mp}\}$ are known, equation (2-2-2) gives up a system of non-linear equations to determine the block of values $\{Y_{mp+1},\ldots,Y_{mp+p}\}$. With m=0, (2-2-2) gives us a system of equations to determine $\{Y_1,\ldots,Y_p\}$ using the interpolatory rule only. If we let $\epsilon_{mp+i}=y(x_{mp+i}) - Y_{mp+i}$, and let R_{mp+i} represent the truncation errors due to <u>both</u> quadratures, then we obtain from (2-2-2):

$$|\epsilon_{mp+i}| \le hL\sum_{j=0}^{mp}|\omega_j||\epsilon_j| + hL\sum_{s=0}^{p}|\epsilon_{mp+s}|\cdot\left|\sum_{j=s}^{p}(-1)^j\binom{j}{s}d_{ij}\right| + |R_{mp+i}|$$

$$\le hLW\sum_{j=0}^{mp}|\epsilon_j| + hLD_i\,\epsilon_m + |R_{mp+i}|$$

$$i=1,\ldots,p.$$

where

$$\epsilon_m = \max_{1\le i\le p}|\epsilon_{mp+i}|\ ,\quad D_i=\sum_{s=0}^{p}\left|\sum_{j=s}^{p}(-1)^j\binom{j}{s}d_{ij}\right|,$$

$$W=\max_{0\le j\le mp}|\omega_j|$$

Thus

$$\epsilon_m \le hLWp\sum_{j=0}^{m-1}\epsilon_j + hLD\,\epsilon_m + R_m \qquad \ldots(2\text{-}2\text{-}3)$$

where

$$R_m=\max_{1\le i\le p}|R_{mp+i}|\ ,\quad D=\max_i D_i.$$

From (2-2-3) we obtain

$$\epsilon_m \le \frac{hLWp}{\{1-hLD\}}\sum_{j=0}^{m-1}\epsilon_j + \frac{R_m}{\{1-hLD\}}$$

provided

$$hLD<1 \qquad \ldots(2\text{-}2\text{-}4)$$

By Lemma 1, Chapter 1, we obtain convergence of the scheme (2-2-2) and the order of convergence is that of the least accurate quadrature in (2-2-2). To solve the system of non-linear equations in (2-2-2), for each value of m, by iteration, we write

$$\underset{\sim}{Y}_m = \underset{\sim}{\phi}(\underset{\sim}{Y}_m) \qquad \ldots(2\text{-}2\text{-}5)$$

where

$$\underset{\sim}{Y}_m^T = \{ Y_{mp+1}, \ldots, Y_{mp+p} \}.$$

Then by the contraction mapping theorem we see that the sequence

$$\{ \underset{\sim}{Y}_m^{(r)} \}_{r=0}^{\infty}$$

generated from

$$\underset{\sim}{Y}_m^{(r+1)} = \underset{\sim}{\phi}(\underset{\sim}{Y}_m^{(r)}) \qquad \ldots(2\text{-}2\text{-}6)$$

with $\underset{\sim}{Y}_m^{(0)}$ arbitrary, will converge to $\underset{\sim}{Y}_m$, the unique solution of (2-2-5) provided that (2-2-4) holds. A convenient choice of $\underset{\sim}{Y}_m^{(0)}$ might be

$$\underset{\sim}{Y}_m^{(0)} = \underset{\sim}{g}_m$$

where

$$\underset{\sim}{g}_m^T = \{ g(x_{mp+1}), \ldots, g(x_{mp+p}) \}$$

Another iterative scheme that could be used is a "Gauss-Seidel" type, that is

$$\begin{aligned} Y_{mp+i}^{(r+1)} = g(x_{mp+i}) &+ h \sum_{j=0}^{p} \omega_j K(x_{mp+i}, x_j, Y_j) \\ &+ h \sum_{j=0}^{i-1} c_{ij} K(x_{mp+i}, x_{mp+j}, Y_{mp+j}^{(r+1)}) \\ &+ h \sum_{j=i}^{p} c_{ij} K(x_{mp+i}, x_{mp+j}, Y_{mp+j}^{(r)}), \end{aligned}$$

$$r = 0, 1, \ldots$$

where $Y_{mp}^{(r)} = Y_{mp}$, for all r.

We would expect, intuitively, that this iterative scheme will converge faster than the scheme (2-2-6) this may not necessarily be always true.

2 - 3 Choice of quadrature

The choice of repeated Newton Cotes quadrature as the main quadrature in (2-2-2) has been considered by several authors. (See, for example, Linz [15]). In this case, the scheme becomes

$$Y_{mp+i} = g(x_{mp+i}) + h \sum_{r=0}^{m-1} \sum_{j=0}^{p} c_j \, K(x_{mp+i}, x_{rp+j}, Y_{rp+j}) + h \sum_{j=0}^{p} c_{ij} \, K(x_{mp+i}, x_{mp+j}, Y_{mp+j}), \quad 1 \le i \le p, \qquad \ldots.(2\text{-}3\text{-}1)$$

where the $c_j = c_{pj}$, $j=0,\ldots,p$ are the weights for the Newton-Cotes rule. Another convenient choice for the main quadrature is the Gregory rule, and this will now be pursued by the present writer.

Now the error incurred by use of the interpolatory rule is $0(h^{p+1})$ and the accuracy of the Gregory rule is $0(h^{k+2})$. (See (1-3-4(i)) and (1-3-4(ii)).) Let us consider the case of **p** odd, and apply the Gregory rule in $[x_0, x_{mp}]$ with $k=p-1$. The analysis that follows can be modified in an obvious manner for the case of **p** even.

The scheme (2-2-2) becomes

$$Y_{mp+i} = g(x_{mp+i}) + h\sum_{j=0}^{mp} K(x_{mp+i}, x_j, Y_j)$$
$$+h\sum_{j=0}^{p-1} A_j^{p-1}\{K(x_{mp+i}, x_j, Y_j) + K(x_{mp+i}, x_{mp-j}, Y_{mp-j})$$
$$+h\sum_{j=0}^{p} C_{ij}\, K(x_{mp+i}, x_{mp+j}, Y_{mp+j}), \qquad 1\le i\le p. \qquad \ldots(2\text{-}3\text{-}2)$$

We now derive an estimate for the error in each block of computed values.

The equation (2-3-2) involving exact values $y(x_{mp+i})$ is

$$y(x_{mp+i}) = g(x_{mp+i}) + h\sum_{r=0}^{mp} K(x_{mp+i}, x_r, y(x_r))$$
$$+h\sum_{j=0}^{p-1} A_j^{p-1}\{K(x_{mp+i}, x_j, y(x_j)) + K(x_{mp+i}, x_{mp-j}, y(x_{mp-j})\}$$
$$+h\sum_{j=0}^{p} C_{ij}\, K(x_{mp+i}, x_{mp+j}, y(x_{mp+j}))$$
$$+h^{p+2}(b_{p-1} + (mp-p+1)a_p)\,K^{(p+1)}(\eta_1) + \int_0^i \binom{s}{p+1} K^{(p+1)}(\eta_s)\,ds, \qquad \ldots(2\text{-}3\text{-}3)$$

where $K^{(p+1)}(\eta)$ is an abbreviation for

$$\left[\frac{d^{p+1}}{dt^{p+1}}\, K(x_{mp+i}, \tau, y(t))\right]_{t=\eta}$$

and in (2-3-3) we have $x_0 \le \eta_1 \le x_{mp}$, $x_{mp} \le \eta_s \le x_{mp} + sh$

We write M as an upper bound for this (p+1)th derivative.

From (2-3-2) and (2-3-3) we have

$$|\varepsilon_{mp+i}| \le hL\sum_{j=0}^{mp} |\varepsilon_j| + hL\sum_{j=0}^{p-1} |A_j^{p-1}|\{|\varepsilon_j| + |\varepsilon_{mp-j}|\}$$
$$+hL\sum_{s=0}^{p} |\varepsilon_{mp+s}|\left|\sum_{j=s}^{p} (-1)^j \binom{s}{s} d_{ij}\right|$$
$$+h^{p+2} M(d + |b_{p-1}| + (mp-p+1)|a_p|),$$

where

$$d = \max_i \int_0^i \left| \binom{s}{p+1} \right| ds = \int_0^p \left| \binom{s}{p+1} \right| ds .$$

Now $\varepsilon_0 = 0$, so we can write

$$\epsilon_m \le h L_p \sum_{r=0}^{m-1} \epsilon_r + hLB(\epsilon_0 + \epsilon_{m-1}) + hLD\,\epsilon_m + h^{p+2} K_m \qquad \ldots(2\text{-}3\text{-}4)$$

where

$$B = \sum_{j=0}^{p-1} |A_j^{p-1}| = |a_0| + 2|a_1| + \ldots + 2^{p-1}|a_{p-1}|, \qquad \ldots(2\text{-}3\text{-}5)$$

as in Phillips [20], and

$$K_m = M(d + |b_{p-1}| + (mp - p + 1)|a_p|) \qquad \ldots(2\text{-}3\text{-}6)$$

Now for m=0 (2-3-4) simplifies to

$$\epsilon_0 \le hLD\,\epsilon_0 + h^{p+2} M d \qquad \ldots(2\text{-}3\text{-}7)$$

Consider now the difference equation

$$(1 - hLD)u_m = hLp \sum_{r=0}^{m-1} u_r + hLB(u_0 + u_{m-1}) + h^{p+2} K_m \qquad \ldots(2\text{-}3\text{-}8)$$

Replacing m in (2-3-8) by m+1, subtracting from (2-3-8) gives us, on using (2-3-6)

$$(1 - hLD)(u_{m+1} - u_m) = hLp\,u_m + hLB(u_m - u_{m-1}) + h^{p+2} M_p |a_p| \qquad \ldots(2\text{-}3\text{-}9)$$

This equation has a particular solution u given by

$$u = -h^{p+1} M |a_p| / L .$$

The characteristic equation associated with the difference equation (2-3-9) is

$$q(z) = (1 - \theta D)(z^2 - z) - \theta p z - \theta B(z - 1).$$

where $\theta = hL$. As $h \to 0$ we see that the equation $q(z)=0$ has roots $z=0$ and $z=1$. For $h \neq 0$, we see that

$$q(1) = -\theta p < 0$$

and

$$q(1+\theta t) = (1-\theta D)\theta t(1+\theta t) - \theta p(1+\theta t) - \theta^2 B t$$

$$= (1+\theta t)(\theta t - \theta p - \theta^2 D t - \theta B) + \theta B.$$

Now $q(1+\theta t) > 0$, if $t = p + B/(1-\theta D)$ and we note that t is positive from (2-2-4).

If we now restrict h further, so that $\theta D = hLD < 1/2$ say, then $p+B < t < 2(p+B)$. There exists a zero z_0 of $q(z)$ such that $z_0 < 1 + 2hL(p+B)$ and thus

$$z_0^m \leq e^{2mhL(p+B)}.$$

From (2-3-7) we note that

$$\epsilon_0 \leq \frac{h^{p+2} Md}{1-hLD} = u_0 \qquad \ldots.(2\text{-}3\text{-}10)$$

A family of solutions of (2-3-8) is given by

$$u_m = C z_0^m - h^{p+1} M|ap|/L \qquad \ldots.(2\text{-}3\text{-}11)$$

Now if we compare the values of u_0 in (2-3-10) and (2-3-11) we see that we need to choose

$$C = h^{p+1} M|ap|/L + h^{p+2} Md/(1-hLD).$$

By induction $\epsilon_m \leq u_m$ for all $m \geq 0$ and thus we obtain the error bound

$$\epsilon_m \leq \frac{h^{p+1}}{L} M|ap| [e^{2mhL(p+B)} - 1] + \frac{h^{p+2} Md}{1-hLD} e^{2mhL(p+B)} \qquad \ldots.(2\text{-}3\text{-}12)$$

This error bound can only really be regarded as being of qualitative interest, because of the difficulty in evaluation M and L in practise. In general we will only be able to obtain a posteriori estimates of these quantities.

The two values of p which will probably be required in practise most often are p=3 and p=5, (Giving schemes $O(h^5)$ and $O(h^7)$ and in order to use the scheme (2-3-2) we will require the p values of A_j^{p-1} and the p(p+1) values of d_{ij} to give us the weights c_{ij}. We list values of d_{ij} in Table 6.

TABLE 6 VALUES OF dij, $1 \le i \le 5$, $o \le j \le 5$

1	1/2	-1/12	1/24	-19/720	3/60
2	2	1/3	0	-1/90	1/90
3	9/2	9/4	3/8	-3/80	3/160
4	8	20/8	8/3	14/45	0
5	25/2	175/12	75/8	425/144	95/288

d_{ij} is found in the ith row and (j+1)th column of the array.

The weights c_{ij} may then be found by using (2-1-2).
The coefficients A_j^{p-1}, $o \le j < p-1$ are given below.

A_j^2 = 5/8, 1/6, -1/24

A_j^4 = -193/288, 77/240, -7/30, 73/720, -3/160

Finally to use the error bound (2-3-11) we require the following coefficients.

p=3 : $|a_3|$ = 19/720, B = 5/6, d = 49/720, D = 3

p=5 : $|a_5|$ = 863/60480, B = 121/90, d = 2459/60480, D = 5 .

Noting that D_i is the sum of the moduli of the weights in each interpolatory rule $1 \le i \le p$, we have $D_i \ge i$, so that $D \ge p$.

If L is large, h may need to be chosen small in order to satisfy (2-2-4).

Numerical Example

We took the equation $y(x) = \cos x - \sin x + 2\int_0^x \sin(x-t)\, y(t)\, dt$ with solution $y(x) = e^{-x}$ and found numerical solutions in [0,1] by the block method studied in this section. With p=5 and L=2 sin 1, we took h=0.05. Thus (2-2-4) was satisfied and we could also apply the error bound (2-3-11) [Note $hLD < \frac{1}{2}$ with our choice of h]. We found we could take M=16.

Some numerical results are shown in TABLE 7.

TABLE 7: SOME NUMERICAL RESULTS FOR THE BLOCK METHOD

Range of values of x in each block	Maximum modulus of error in block	Value of theoretical error bound
0.05 - 0.25	0.14×10^{-8}	0.88×10^{-8}
0.30 - 0.50	0.14×10^{-8}	0.66×10^{-8}
0.55 - 0.75	0.11×10^{-8}	0.23×10^{-7}
0.80 - 1.00	0.17×10^{-8}	0.72×10^{-7}

We performed the standard iteration (2-2-6) until the maximum modulus of the difference between successive iterates was less than $(0.05)^{10}$. For the first block five iterations were needed, for the rest of the blocks six iterations were needed. We can compare the results in Table 7 with those in Table 4, where the Gregory rule was used in a straightforward manner.

2 - 4 A Modified Method

A disadvantage of all the block methods considered so far can be that values of the kernel are required outside the region $a \le t \le x \le b$. Now if the kernel is not defined when $t>x$ or if it is badly behaved in this region, we are in difficulties. If we are faced with such a kernel we can avoid this difficulty by using a different approximation to the second integral in (2-2-1). Let us approximate to the integral by a Newton-Cotes rule with step-size ih/p, $1 \le i \le p-1$. That is

$$\int_{x_{mp}}^{x_{mp+i}} K(x_{mp+i}, t, y(t))\,dt = \frac{ih}{p}\int_0^p K\left(x_{mp+i}, x_{mp}+\frac{ihs}{p}, y\left(x_{mp}+\frac{ihs}{p}\right)\right) ds$$

$$\simeq \frac{ih}{p}\sum_{r=0}^{p} c_r\, K\left(x_{mp+i}, x_{mp}+\frac{irh}{p}, y\left(x_{mp}+\frac{irh}{p}\right)\right) \qquad \ldots(2\text{-}4\text{-}1)$$

We do not know, or require, values of $y(x_{mp}+irh/p)$, $r=1,\ldots,p-1$. However, these values can be found by Lagrangian enterpolation on $y(x_{mp})$, $y(x_{mp+1}),\ldots, y(x_{mp+p})$. (The error incurred from this is $O(h^{p+1})$. From (2-4-1) we see that values of the kernel are no longer required outside $a \le t \le x \le b$ and the difficulty is avoided. This is provided that values of the kernel are known at the intermediate mesh points $x=x_{mp}+irh/p$, $r=1,\ldots,p-1$.

Example 1. Suppose we are using a repeated Simpson's rule as the main quadrature. Then the above approximation to the integral in $[x_{2m}, x_{2m+1}]$, using Lagrangian interpolation on the unknown $Y_{2m+\frac{1}{2}}$, yields the expressio for Y_{2m+1};

$$Y_{2m+1} = g(x_{2m+1}) + h/3 \sum_{r=0}^{m-1} \{ K(x_{2m+1}, x_{2r}, Y_{2r}) + 4K(x_{2m+1}, x_{2r+1}, Y_{2r+1}) + K(x_{2m+1}, x_{2r+2}, Y_{2r+2}) \} + h/6 \{ K(x_{2m+1}, x_{2m}, Y_{2m}) + 4K(x_{2m+1}, x_{2m+1/2}, 3/8\, Y_{2m} + 3/4\, Y_{2m+1} - 1/8\, Y_{2m+2}) + K(x_{2m+1}, x_{2m+1}, Y_{2m+1}) \}$$

Example 2. Suppose we are using Gregory quadrature as the main rule with p=3. We then approximate to the integral in (x_{2m}, x_{3m+i}) by the three-eighths rule at the point x_{3m}, $x_{3m+i/3}$, $x_{3m+2i/3}$, x_{3m+i}. Values of Y at these intermediate mesh points are then found from cubic Lagrangian interpolation on $Y_{3m}, \ldots, Y_{3m+3}$.

Example 3. So far we have only considered using quadratures which make use of equally spaced points. Lobatto four-point quadrature can be used in $[x_{mp}, x_{mp+i}]$ with p=4, and a modified scheme can be obtained in an analogous way to the above. A repeated five-point Newton-Cotes rule can be used as the main rule, or a Gregory rule with k=4, to give an accuracy of $O(h^7)$, which is the same order of accuracy as the Lobatto rule.

The following approximation to the integral in $[x_{4m}, x_{4m+i}]$ is taken from Campbell and Day [5].

$$\int_{x_{4m}}^{x_{4m+i}} K(x_{4m+i}, t, y(t))\, dt \simeq h/12 \sum_{j=0}^{i-1} \{ K(x_{4m+i}, x_{4m+j}, y(x_{4m+j})) + 5K(x_{4m+i}, x_{4m+j+r}, y(x_{4m+j+r})) + 5K(x_{4m+i}, x_{4m+j+s}, y(x_{4m+j+s})) + K(x_{4m+i}, x_{4m+j+1}, y(x_{4m+j+1})) \},$$

$i = 1, 2, 3.$

where $r=\left(\frac{5-\delta s}{10}\right)h$, $s=\left(\frac{5+\delta s}{10}\right)h$.

The unknown $y(x_{4m+j+r})$, $y(x_{4m+j+s})$, $j=0,1,2$ can be obtained from Lagrangian interpolation or $y(x_{4m+j})$, $j=0,1,\ldots,4$.

Section 3: RUNGE-KUTTA METHODS

3 - 1 Introduction

Let us recall the Runge-Kutta method for the initial value problem

$$y'(x) = F(x, y(x)) \quad , \quad y(x_0) = y_0 \, . \qquad \ldots(3\text{-}1\text{-}1)$$

It consists of replacing the result of truncating a Taylor series expansion for $y(x_n)$,

$$y(x_n) = y(x_{n-1}) + h\,y'(x_{n-1}) + \frac{h^2}{2!}\,y''(x_{n-1}) + \cdots \qquad \ldots(3\text{-}1\text{-}2)$$

by an approximation to $y(x_n)$,

$$y_n = y_{n-1} + \alpha_0 K_0 + \cdots + \alpha_p K_p \quad , \quad p \geq 0,$$

where

$$\begin{aligned} K_0 &= hF(x_{n-1}, y_{n-1}) \\ K_1 &= hF(x_{n-1} + \mu_1 h, \; y_{n-1} + \lambda_{10} K_0) \\ &\vdots \\ K_p &= hF(x_{n-1} + \mu_n h, \; y_{n-1} + \lambda_{p0} K_0 + \cdots + \lambda_{p,p-1} K_{p-1}). \end{aligned} \qquad \ldots(3\text{-}1\text{-}3)$$

We determine the parameters α_i, μ_i, λ_{ij} so that when y_{n-1} is replaced by $y(x_{n-1})$ and expanded in a Taylor series about $(x_{n-1}, y(x_{n-1}))$ in powers of h, the coefficients of a certain number of the leading terms agree with those in (3-1-2). [The values of the derivatives in (3-1-2) are expressed in terms of derivatives of $F(x,y)$ at $(x_{n-1}, y(x_{n-1}))$ on using (3-1-1)]. With these values of the parameters (3-1-3) then yields an approximation y_n to $y(x_n)$.

For example, the case p=1 can lead to the scheme

$$y_n = y_{n-1} + \tfrac{1}{2}(K_0 + K_1) \qquad n=0,1,\ldots$$

where $K_0 = hF(x_{n-1}, y_{n-1})$

$$K_1 = hf(x_n, y_{n-1} + K_0)$$

and the scheme is $O(h^3)$.

Now, returning to the integral equation

$$y(x) = g(x) + \int_a^x K(x,t,y(t))\,dt,$$

we find that we cannot apply a Runge-Kutta method in such a straightforward manner. However, it is possible to apply an analogous method although its application is not so easy. The analysis presented in the next section is due to Bel'tyukov [3]. Other Runge-Kutta methods (e.g. Pouzet [22]) have been proposed, but Bel'tyukov's is simpler and computationally more efficient.

3 - 2 The Runge-Kutta Method

We write the integral equation in the form

$$y(x) = g(x) + \int_a^{x_{n-1}} K(x,t,y(t))\,dt + \int_{x_{n-1}}^x K(x,t,y(t))\,dt$$

$$= \mu_{n-1}(x) + \gamma_n(x), \qquad \ldots.(3\text{-}2\text{-}1)$$

say, where

$$\mu_{n-1}(x) = g(x) + \int_a^{x_{n-1}} K(x,t,y(t))\,dt \qquad \ldots.(3\text{-}2\text{-}2)$$

$$\gamma_n(x) = \int_{x_{n-1}}^x K(x,t,y(t))\,dt. \qquad \ldots.(3\text{-}2\text{-}3)$$

(3-2-3) is now written as

$$\gamma_n(x) = \int_{x_{n-1}}^x K^*(x,t,\gamma_n(t))\,dt \qquad \ldots.(3\text{-}2\text{-}4)$$

where

$$K^*(x,t,\gamma_n(t)) = K(x,t,\gamma_n(t)+\mu_{n-1}(t)) \qquad \ldots.(3\text{-}2\text{-}5)$$

We set $x_n=x_o+nh$, with $x_o=a$, (as before), and suppose that approximations $y_1,\ldots,y_{n-1}$ to $y(x_1),\ldots,y(x_{n-1})$ are known. The idea now is to use a Runge-Kutta process on (3-2-4) to determine an approximation to $\gamma_n(x)$. Then y_n, the approximation to $y(x_n)$, is to be found from (3-2-1). (We note that $\mu_{n-1}(x_n)$ could be determined by means of quadrature). We now show how $\gamma_n(x)$ is to be determined. First, consider a Taylor series expansion for $\gamma_n(x_n)$,

$$\gamma_n(x_n) = \gamma_n(x_{n-1}) + h\gamma_n'(x_{n-1}) + \frac{h^2}{2!}\gamma_n''(x_{n-1}) + \ldots \qquad \ldots.(3\text{-}2\text{-}6)$$

We can obtain expressions for the derivatives in (3-2-6) from differentiating (3-2-4) with respect to x, and setting $x=x_{n-1}$. We note that all terms under the integral sign will then disappear. Writing $K^*=K^*(x,t,z)$, we have,

$$\gamma_n(x_{n-1})=0$$

$$\gamma_n'(x_{n-1})=K^*$$

$$\gamma_n''(x_{n-1}) = 2K^* + K_t^* + K^*K_z^*$$

$$\gamma_n'''(x_{n-1}) = 3K_{xx}^* + 3K_{xt}^* + K_{tt}^* + 3K_{xz}^* + 2K^*K_{tz}^* + K^{*2}K_{zz}^* + 2K_x^*K_z^* + K_t^*K_z^* + K^*K_z^{*2} \qquad \ldots(3\text{-}2\text{-}7)$$

$$\vdots$$

In (3-2-7), all the partial derivatives of $K^*(x,t,z)$ are evaluated at $(x_{n-1}, x_{n-1}, Y_n(x_{n-1}))$. That is $(x_{n-1}, x_{n-1}, 0)$. Now consider an approximation $\tilde{Y}_n$ to $Y_n(x_n)$ of the form

$$\tilde{Y}_n = \sum_{s=1}^{r} A_{rs} K_s , \qquad \ldots(3\text{-}2\text{-}8)$$

where

$$K_1 = hK^*(x_{n-1} + d_1 h, x_{n-1} + \beta_1 h, 0)$$

$$K_2 = hK^*(x_{n-1} + d_2 h, x_{n-1} + \beta_2 h, \gamma_{21} K_1)$$

$$K_r = hK^*(x_{n-1} + d_r h, x_{n-1} + \beta_r h, \gamma_{r1} K_1 + \cdots + \gamma_{rr} K_{r-1})$$

The parameters d_i, β_i, γ_{ij} are to be chosen so that the terms in a Taylor series expansion of (3-2-8) coincide with the terms in (3-2-6), using (3-2-7) to the largest possible power of h, for a given value of r.

Expanding K_1 we obtain

$$\frac{K_1}{h} = K^* + d_1 h K_x^* + \tfrac{1}{2}\{ d_1^2 h^2 K_{xx}^* + 2\beta_1 d_1 h^2 K_{xt}^* + \beta_1^2 h^2 K_{tt}^* \} + O(h^3),$$

where again the partial derivatives are evaluated at $(x_{n-1}, x_{n-1}, 0)$.

Similarly expanding K_2, K_3, ..., we obtain for (3-2-8)

$$\tilde{Y}_n = hK^* \sum_{s=1}^{r} A_{rs} + h^2 K_x^* \sum_{s=1}^{r} d_s A_{rs} + h^2 K_t^* \sum_{s=1}^{r} \beta_s A_{rs} + h^2 K_z^* K^* \sum_{s=2}^{r} A_{rs} \sum_{v=1}^{s-1} \gamma_{sv} + O(h^3). \qquad \ldots(3\text{-}2\text{-}9)$$

Using (3-2-7), (3-2-6) becomes

$$Y_n(x_n) = hK^* + \frac{h^2}{2!}\{ 2K_x^* + K_t^* + K^* K_z^* \} + O(h^3). \qquad \ldots(3\text{-}2\text{-}10)$$

We are now in a position to match coefficients of h in (3-2-9) and (3-2-10) for specific values of r.

Case 1 : r=1

Here we obtain from (3-2-9)

$$\tilde{Y}_n = hK^* A_{11} + h^2 K_x^* A_{11} + h^2 K_y^* \beta_1 A_{11} + O(h^3).$$

Comparing coefficients with (3-2-10) we obtain $A_{11} = 1$.

The terms in h^2 cannot be matched completely, but we can make the discrepancy as small as possible by choosing $\alpha_1 = 1, \beta_1 = 1/2$.

We therefore obtain

$$\tilde{Y}_n = hK^*(x_n, x_{n-1} + 1/2h, 0)$$

$$= hK(x_n, x_{n-1} + 1/2h, \mu_{n-1}(x_{n-1} + 1/2h))$$

Thus

$$y_n = \mu_{n-1}(x_n) + hK(x_n, x_{n-1} + 1/2h, \mu_{n-1}(x_{n-1} + 1/2h)) \qquad n=1,2,\ldots$$

is a Runge-Kutta approximation to $y(x_n)$ of order h^2.

Case 2 : r=2

Here

$$\tilde{Y}_n = hK^*(A_{21} + A_{22}) + h^2 K_x^* (\alpha_1 A_{21} + \alpha_2 A_{22}) + h^2 K_y^* (\beta_1 A_{21} + \beta_2 A_{22})$$
$$+ h^2 K_z^* K^* A_{22} \gamma_{21} + O(h^3).$$

We can match all the terms in h^2 with (3-2-10) to give

$$A_{21} + A_{22} = 1,$$
$$\alpha_1 A_{21} + \alpha_2 A_{22} = 1,$$
$$2\beta_1 A_{21} + \beta_2 A_{22} = 1,$$
$$2 A_{22} \gamma_{21} = 1.$$

That is, we have four equations in five unknowns. If we regard the first three equations as a linear system in A_{21} and A_{22}, we obtain the consistency condition

$$(2\beta_1 - 1)(\alpha_1 - 1) = (2\beta_1 - 1)(\alpha_2 - 1).$$

Provided this condition is satisfied, the parameters can be chosen arbitrarily. We desire to make the terms in h^3 in (3-2-9) and (3-2-10) match as closely as possible. It can be verified (see Bel'tyukov) that the choice

$$d_1=\alpha_2=1,\ \beta_1=0,\ \beta_2=\gamma_{21}=2/3,\ A_{21}=1/4,\ A_{22}=3/4$$

makes six of these terms match. With this choice we obtain

$$y_n=\mu_{n-1}(x_n)+1/4\,K_1+3/4\,K_2, \qquad n=1,2,\ldots$$

with

$$K_1=hK(x_n, x_{n-1}, y_n)$$

$$K_2=hK(x_n, x_{n-1}+2/3\,h, \mu_{n-1}(x_{n-1}+2/3\,h)+2/3\,K_1)$$

Case 3 : r=3

In a similar way we can match terms up to and including the terms in h^3 to give us a method which is $O(h^4)$. We obtain thirteen equations in twelve unknowns that are consistent.

One set of values of the parameters gives us the scheme

$$y_n=\mu_{n-1}(x_n)+1/4\,(3K_2+K_3), \qquad n=1,2,\ldots$$

where

$$K_1=hK(x_{n-1}+h/3, x_{n-1}, y_{n-1}),$$

$$K_2=hK(x_{n-1}+2/3\,h, x_{n-1}+2/3\,h, \mu(x_{n-1}+2/3\,h)+2/3\,K_1)$$

$$K_3=hK(x_n, x_{n-1}, y_{n-1}+K_2-K_1)$$

Now in all three cases we require the value of $\mu_{n-1}(x_n)$. Also for r=1, we require $\mu(x_{n-1}+1/2\,h)$ and for r=2 and r=3 we require $\mu(x_{n-1}+2/3\,h)$ Now these values are unknown, but by using (3-2-2), and the formerly calculated values $y_1,\ldots,y_{n-1}$, approximations to these values could be obtained by a quadrature method to the required accuracy.

3 - 3 Discussion

Clearly, the need to evaluate the quantities $\mu_{n-1}(x_n)$, $\mu_{n-1}(x_{n-1}+{}^1/_2h)$ and $\mu_{n-1}(x_{n-1}+{}^2/_3h)$ is a drawback. Coupled with the need to evaluate many kernel values for the computation of K_1, K_2 ,... it is doubtful whether there is any advantage in this approach over the straightforward quadrature methods discussed in Section 1. Although we have considered only Runge-Kutta methods up to $O(h^4)$ in accuracy, higher order methods are clearly possible, although they would be far more complex. The errors in the schemes could be determined, but these would contain many values of the various derivatives of $K(x,t,z)$ and would be of little use in practise.

The best application of Runge-Kutta methods is probably in finding starting values for a quadrature method. For example the $O(h^3)$ method could produce a starting value Y_1 as follows

$$Y_1 = g(x_1) + \frac{K_1}{4} + \frac{3K_2}{4},$$

with

$$K_1 = hK(x_1, x_0, g(x_0)),$$

$$K_2 = hK(x_1, x_0 + {}^2/_3 h, g(x_0 + {}^2/_3 h) + {}^2/_3 K_1).$$

This could be used as the starting value required by Simpson's rule, used in composite form.

Similarly, the $O(h^4)$ method would give

$$Y_1 = g(x_1) + {}^1/_4 (3K_2 + K_3),$$

with

$$K_1 = hK(x_0 + h/3, x_0, g(x_0)),$$

$$K_2 = hK(x_1, x_0 + {}^2/_3 h, g(x_0 + {}^2/_3 h) + {}^2/_3 K_1),$$

$$K_3 = hK(x_1, x_0, g(x_0) + K_2 - K_1).$$

This could be used as the starting value required by the Gregory rule with k=2.

Section 4: THE FINITE RANK METHOD

4 - 1 Introduction

This method differs from the methods considered so far in this chapter in that is is applicable only to linear Volterra equations of the second kind. The theory associated with the method has been studied in depth by Anselone and Gonzalez-Fernandez [1] and has been applied to both Volterra and Fredholm equations of the second kind by Espinosa-Maldonado and Byrne [8]. The theory was developed principally for Fredholm equations, but it is equally applicable to Volterra equations. As we shall see, the method possesses the advantage that it provides a global approximation to the solution of the equation, as distinct from the finite difference methods considered so far.

4 - 2 Development of the method

The linear Volterra equation of the second kind, with a continuous kernel $K_v(x,t)$,

$$y(x) = g(x) + \int_a^x K_v(x,t)\, y(t)\, dt, \qquad a \le t \le x \le b, \qquad \ldots.(4\text{-}2\text{-}1)$$

can be written as the Fredholm equation

$$y(x) = g(x) + \int_a^b K(x,t)\, y(t)\, dt, \qquad a \le x \le b, \qquad \ldots(4\text{-}2\text{-}2)$$

where $$K(x,t) = \begin{cases} K_v(x,t) & a \le t \le x, \\ 0 & a \le x < t \le b. \end{cases} \qquad \ldots.(4\text{-}2\text{-}3)$$

The Fredholm equation (4-2-2) is said to possess a mildly discontinuous kernel K(x,t). (K(x,t) may be discontinuous when x=t). The method will now be applied to (4-2-2) and the results will be eventually put in terms of the Volterra kernel by means of (4-2-3). The method is more easily developed if we rewrite (4-2-2) in operator notation.

If we let K denote the integral operator,

$$(Kf)(x) = \int_a^b K(x,t) f(t)\, dt \quad ,$$

then K possesses a mildly discontinuous kernel K(x,t).

The equation (4-2-2) can be rewritten as

$$(I-K)y = g \quad , \qquad \ldots.(4\text{-}2\text{-}4)$$

where I is the identity operator.

We now suppose it is possible to rewrite (4-2-4) in the form

$$(I-KL)y = u, \qquad \ldots.(4\text{-}2\text{-}5)$$

where L is an integral operator with kernel L(x,t). Now replace K in (4-2-5) by K_N, say, where K_N is the finite dimensional operator

$$(K_N f)(x) = h \sum_{k=0}^{N}{}' \omega_{Nk} K(x,x_k) f(x_k)$$

where $x_k = x_o + kh$, and $x_o = a, x_N = b$; $h = b-a/N$.

In other words, we are approximately to the integral in (4-2-5) by a quadrature with weights $w_{N,k}$, k=o,...,N. We now have the equation

$$(I - K_N L) y_N = u. \qquad \ldots.(4\text{-}2\text{-}6)$$

Notice now that KL in (4-2-5) is an integral operator with kernel

$$KL(x,t) = \int_a^b K(x,s) L(s,t)\, ds, \qquad \ldots.(4\text{-}2\text{-}7)$$

but that $K_N L$ in (4-2-6) is an integral operator with kernel

$$K_N L(x,t) = \sum_{K=0}^{N} \omega_{NK} K(x,x_K) L(x_K,t) \quad \ldots.(4\text{-}2\text{-}8)$$

We see that (4-2-5) is an integral equation possessing a kernel of finite rank N, and standard methods exist for solving such an equation. (See Tricomi [23]). But before we continue in this vein we must examine whether it is possible to write (4-2-4) in the form (4-2-5). Suppose $(I+K)^{-1}$ exists. Then we can write (4-2-4) as

$$(I-K^2)y = (I+K)g,$$

which is in the form (4-2-5) with L=K, U=(I+K)g. Now $(I+K)^{-1}$ exists if and only if the eigenvalue problem

$$Kx = \lambda x$$

does not have a solution $\lambda = -1$. However, for the Volterra kernel there is no eigenvalue problem and so the existence of $(I+K)^{-1}$ is established. Returning now to (4-2-6) we can write (4-2-6) explicitly as

$$y_N(x) - h\int_a^b \sum_{K=0}^{N} \omega_{NK} K(x,x_K) L(x_K,t) y_N(t)\, dt = u(x) \quad \ldots.(4\text{-}2\text{-}9)$$

Let

$$\xi_K = \int_a^b L(x_K,t) y_N(t)\, dt$$

$$= \int_{x_0}^{x_K} K_\nu(x_K,t)\, y_N(t)\, dt, \qquad K=1,\ldots,N.$$

with the above choice of L. So (4-2-9) becomes

$$y_N(x) - h\sum_{K=0}^{N} \omega_{NK} K(x,x_K)\xi_K = u(x).$$

Now multiply this equation by $L(x_i,x)$, i=1,...,N and integrate with respect to x from a to b to obtain

$$\xi_i - \sum_{K=1}^{N} A_{iK}\,\xi_K = B_i \qquad i = 1,\ldots,N, \qquad \ldots(4\text{-}2\text{-}10)$$

where

$$A_{iK} = h\,\omega_{NK}\int_a^b K(x,x_K)\,L(x_i,x)\,dx$$

$$= h\,\omega_{NK}\int_{x_K}^{x_i} K_V(x,x_K)\,K_V(x_i,x)\,dx\,,$$

$$i,K = 1,\ldots,N, \qquad \ldots(4\text{-}2\text{-}11)$$

and

$$B_i = \int_a^b L(x_i,x)\,u(x)\,dx$$

$$= \int_{x_0}^{x_i} K_V(x_i,x)\,u(x)\,dx\,, \qquad \ldots(4\text{-}2\text{-}12)$$

$$i = 1,\ldots,N,$$

where $$u(x) = g(x) + \int_a^x K_V(x,t)\,g(t)\,dt. \qquad \ldots(4\text{-}2\text{-}13)$$

Now (4-2-10) is a system of N equations in the N unknowns ξ_i, $i = 1,\ldots,N$. If we solve these equations for the ξ_i, we can obtain $y_N(x)$ from (4-2-9). That is

$$y_N(x) = g(x) + \int_a^x K_V(x,t)\,g(t)\,dt + h\sum_{K=1}^{N} \omega_{N,K}\,K_V(x,x_K)\,\xi_K. \qquad \ldots(4\text{-}2\text{-}14)$$

We notice that (4-2-14) gives us a global approximation to y(x), if the integral

$$\int_a^x K_V(x,t)\,g(t)\,dt$$

can be evaluated analytically.

Anselone and Gonzalez-Fernandez have examined convergence properties of the operators K_N and K and the convergence of $y_N(x)$ to y(x). The

most important results are contained in the following theorem.

Theorem 2 (Anselone and Gonzalez-Fernandez)

Let X be the space of functions f which are proper Riemann integrable on [a,b]with norm

$$\|f\| = \sup_{a\le x\le b} |f(x)|.$$

Then if the quadrature used in (4-2-14) is convergent and $W_{Nk} \ge 0$, k=o,...,N, we have

$$\|y_N - y\| \to 0 \quad \text{as } N \to \infty$$

with the error bound

$$\|y_N - y\| \le \|(I - K_N L)^{-1}\| \cdot \left\{ \frac{\|K_N L - K L\| \cdot \|y_N\|}{1 - \|(I - K_N L)^{-1}\| \cdot \|K_N L - K L\|} \right\}$$

4 - 3 Discussion

The method described will produce uniformly convergent solutions to the integral equation provided the conditions in the theorem are satisfied. The only restrictive condition is that all the weights in the quadrature must be positive. This prevents us from using high order Newton-Cotes rules, however, normally, Simpson's rule or the three-eighth's rule is adequate.

The setting up of the system of equations (4-2-10) requires considerable computation. We need first to evaluate u(x) and then B_i, i=1,...,N and A_{ik}, i,k=1,...,N. In general these quantities will not be available analytically and approximations will need to be found by quadrature. It would seem desirable to use a quadrature with the same order of accuracy as the quadrature used in (4-2-8). The method possesses the advantages over the multistep methods of Section 1 in that it requires no starting values and that if (4-2-13) is available exactly, a global approximation is obtained for y(x).

SECTION 5: EXPANSION METHODS

5 - 1 Introduction

The purpose of this section is to give a brief introduction to the application of expansion methods to the Volterra equation of the second kind. These methods have not been widely applied to the Volterra equation, but there appears to be no reason for this. We attempt to solve

$$y(x) = g(x) + \int_a^x K(x,t,y(t))\,dt \quad , \quad a \le t \le x \le b, \qquad \ldots.(5\text{-}1\text{-}1)$$

by means of an approximate representation of the form

$$y(x) \simeq Y(x) = \sum_{j=1}^{n} a_j \psi_j(x), \qquad \ldots.(5\text{-}1\text{-}2)$$

where the $\psi_j(x)$, j=o,...,n are suitably chosen functions and the a_j, j=1,...,n are to be determined. The method is particularly applicable when we have some prior knowledge of the behaviour of y(x). For example, we may know that y(x) is even, or odd. Substitution of (5-1-2) into (5-1-1) gives us the <u>residual</u> R(x) where

$$R(x) = Y(x) - g(x) - \int_a^x K(x,t,Y(t))\,dt . \qquad \ldots.(5\text{-}1\text{-}3)$$

There are two methods of approach now. We can either minimise the residual in someway to obtain values of a_j, j=1,...,n, or we can minimise the error e(x), where

$$e(x) = y(x) - Y(x), \qquad \ldots.(5\text{-}1\text{-}4)$$

and obtain values of a_j, j=1,...,n.

From (5-1-3) and (5-1-4) we obtain

$$e(x) = -R(x) - \int_a^x K(x,t,y(t))\,dt + \int_a^x K(x,t,Y(t))\,dt,$$

which yields $\|e\| \le \|R\| + L(b-a)\|e\|$

where $\|e\| = \sup_{a \le x \le b} |e(x)|$.

If $L(b-a) < 1$ we have

$$\|e\| \leq \|R\| / (1 - L(b-a)). \qquad \ldots(5\text{-}1\text{-}5)$$

If $Y(x)$ is derived from some residual minimising technique we will be able to evaluate $R(x)$ from (5-1-3) (or, at least, calculate $R(x)$ at a discrete number of points). An error bound can then be obtained from (5-1-5). We now examine briefly some methods of minimising $R(x)$ and $e(x)$.

5 - 2 Residual minimising techniques

(i) Collocation

Here we simply set $R(x)=0$ at certain points x_i, $i=1,\ldots,n$. These points are called collocation points. We obtain the non-linear system of equations

$$\sum_{j=1}^{n} a_j \psi_j(x_k) = g(x_k) + \int_a^{x_k} K(x_k, t, \sum_{j=1}^{n} a_j \psi_j(t)) \, dt, \quad k=1,\ldots,n.$$

If the integral equation is linear, we obtain the linear system

$$\sum_{j=1}^{n} a_j \{ \psi_j(x_k) - \int_a^{x_k} K(x_k, t) \psi_j(t) \, dt \} = g(x_k), \quad k=1,\ldots,n.$$

(ii) Least squares

Here we minimise $I = \int_a^b R^2(x) \, dx$. This has the advantage that $R(x)$ is minimised in a particular way over the whole region of interest $[a,b]$. The coefficients a_r $r=1,\ldots,n$ are then found by setting $\partial I / \partial a_r = 0$, $r=1,\ldots,n$ and solving the resulting system of equations. For the linear Volterra equation we obtain the following linear system:

$$\sum_{j=1}^{n} a_j \{ \int_a^b \Phi_k(x) \Phi_j(x) \, dx \} = \int_a^b \Phi_k(x) g(x) \, dx, \qquad \ldots(5\text{-}2\text{-}1)$$

$$k=1,\ldots,n,$$

where

$$\underline{\Phi}_j(x) = \psi_j(x) - \int_a^x K(x,t)\,\psi_j(t)\,dt.$$

Normally we will need to approximate to the integrals in (5-2-1) by quadrature. For the non-linear equation we obtain a system of non-linear equations which will need to be solved by some iteration method.

(iii) Galerkin's method

Here we let

$$I_r = \int_a^b R(x)\,\underline{\Psi}_r(x)\,dx \qquad r=1,\ldots,n,$$

where the $\underline{\Psi}_r(x)$ are suitably chosen functions, and then set each $I_r = 0$. For the linear Volterra equation we obtain the system

$$\sum_{j=1}^{n} a_j \left\{ \int_a^b \underline{\Psi}_k(x)\,\underline{\Phi}_j(x)\,dx \right\} = \int_a^b \underline{\Psi}_k(x)\,g(x)\,dx,$$

$$k=1,\ldots,n, \qquad \ldots(5\text{-}2\text{-}2)$$

If the $\psi_r(x)$ are orthogonal polynomials with respect to a weight function $w(x)$, that is they satisfy

$$\int_a^b w(x)\,\psi_r(x)\,\psi_j(x)\,dx = 0, \qquad r \neq j,$$

then it is convenient to choose $\underline{\Psi}_r(x) = \psi_r(x)\,w(x)$.

(iv) Chebyshev approximation

Here we attempt to find

$$\min_{a_r} \max_{a \le x \le b} |R(x)|.$$

In practise this problem is replaced by

$$\min_{a_r} \max_{i} |R(x_i)|$$

for some choice of x_i, i=o,...,m, lying in [a,b] and where m>>n. This is a discrete non-linear minimax problem for the a_r, r=1,...,n. Algorithms have been developed by Osborne and Watson [18] and the linear problem is discussed by Barrodale and Young in [2].

5 - 3 Error minimising techniques

This is a relatively new technique, due to Watson [24].

The error (5-1-4) satisfies

$$e(x) + Y(x) = g(x) + \int_a^x K(x,t, Y(t))\,dt + \int_a^x \partial K/\partial Y\,(x,t, Y(t))\,e(t)\,dt$$

if we retain only first order terms in e(t). We therefore have

$$e(x) = p(x) + \int_a^x \frac{\partial K}{\partial Y}(x,t,Y(t))\,e(t)\,dt, \qquad \ldots.(5\text{-}3\text{-}1)$$

where

$$p(x) = g(x) - Y(x) + \int_a^x K(x,t, Y(t))\,dt.$$

If we discretize (5-3-1) on the equispaced points x_i, i=1,..., m (m>>n) with x_i=a+ih; x_m=b, and approximate to the integral in (5-3-1) by quadrature, we obtain

$$\varepsilon_i = p(x_i) + h \sum_{j=0}^{i} \omega_{ij} \frac{\partial K}{\partial Y}(x_i, x_j, Y(x_j))\,\varepsilon_j \quad , i=1,\ldots,m, \qquad \ldots.(5\text{-}3\text{-}2)$$

where, for any convergent quadrature,

$$\lim_{\substack{i\to\infty \\ h\to 0 \\ (ih\ \text{fixed})}} \varepsilon_i = e(x_i).$$

This follows since (5-3-1) is a Volterra equation of the second kind in e(x). Then (5-3-2) gives us the system of equations

$$\underset{\sim}{\varepsilon} = K^{-1}\underset{\sim}{p} \qquad \ldots.(5\text{-}3\text{-}3)$$

where

$$\underset{\sim}{\varepsilon} = \{\varepsilon_1, \ldots, \varepsilon_m\}^T, \qquad \underset{\sim}{a} = \{a_1, \ldots, a_n\}^T,$$

and K is an mxn matrix. We have assumed that K^{-1} exists. In (5-3-3) the matrix K and the vector p depend non-linearly on a. Osborne and Watson now continue by solving the non-linear discrete minimax problem

$$\min_{a_r} \max_{1 \le x_j \le m} |\varepsilon(x_j)|.$$

However, if the integral equation is linear, then K^{-1} is independent of a_r, r=1,...,n, and we could also minimise ε_i, $i = 1, \ldots, m$ in a discrete least-squares sense. That is, we find

$$\min_{a_r} \sum_{j=1}^{m} \varepsilon_j^2 \qquad \ldots.(5\text{-}3\text{-}4)$$

and obtain a system of linear equations for the a_r, r=1,...,n

5 - 4 Discussion

We have briefly examined several ways of obtaining an approximate solution to the integral equation using a series expansion. For the non-linear Volterra equation, the methods reduce to a system of non-linear equations in the unknown coefficients; in most cases. The solution of this problem has not been discussed here, and, depending on the form of the kernel, it may be difficult to find.. The methods are clearly more easily applied to the linear equation.

The error minimisation techniques possess the advantage that they enable us to obtain actual values for the error at a discrete number of points [using 5-3-3]; however they require more computation in setting up the system of equations, [in the case of (5-3-4)]. We shall return to the least-squares problem (5-3-4) in Chapter 3 for the equation of the first kind.

CHAPTER 3

THE VOLTERRA EQUATION OF THE FIRST KIND

Section 1: INTRODUCTION

1 - 1 The Volterra equation of the first kind

In this chapter we consider the linear, non-singular Volterra equation of the first kind

$$\int_a^x K(x,t)\, y(t)\,dt = g(x) \qquad a \le x \le b. \qquad \ldots.(1\text{-}1\text{-}1)$$

Conditions for this equation to have a unique and continuous solution were given in Chapter 1, and these conditions are assumed to hold. In Chapter 1 we saw that (1-1-1) could be converted into an equation of the second kind by differentiation; that is

$$y(x) = \frac{g'(x)}{K(x,x)} - \int_a^x \frac{\partial K/\partial x\,(x,t)}{K(x,x)}\, y(t)\,dt \qquad \ldots.(1\text{-}1\text{-}2)$$

Now, we could find numerical solutions of (1-1-1) by applying the methods of Chapter 2 to (1-1-2). However, it is desirable to have direct methods for solving (1-1-1) especially if K(x,t) or g(x) are given only in tabulated form. It is known that numerical differentiation can be a sensitive process.

1 - 2 Some problems

Equations of the first kind have always been viewed with caution. They are usually suspected of being ill-conditioned or ill-posed. By ill-conditioning we mean that a small perturbation in g(x) (for example, from experimental data) can produce a large perturbation in the solution y(x). By being ill-posed we mean that even though a solution y(x) exists, (in (1-1-1)) and is unique, it may not depend continuously on the data g(x). Equation (1-1-2) informs us that y(x) depends continuously on $g^{(1)}(x)$. A simple example demonstrates that the Volterra equation can be ill-posed. Consider

$$\int_0^x y(t)\,dt = g(x)$$

If we perturb g(x) to form g(x) + e(x), then the perturbed equation

$$\int_0^x y(t)\,dt = g(x) + e(x)$$

has no solution unless e(0) = 0, and, in view of (1-1-2), unless $e^{(1)}(x)$ exists.

This suggests that any numerical procedure for solving (1-1-1) must be examined carefully, as the introduction of discretization errors could have a large effect on the solution.

Difficulties are soon apparent when Fredholm equations of the first kind are solved by standard numerical methods. For example an obvious method of solving

$$g(x) = \int_a^b K(x,t)\,y(t)\,dt \qquad a \le x \le b.$$

is to replace the integral by a quadrature with weights $w_{Ni}, i=0,\ldots,N$

and solve the resulting system of equations. That is

$$g(x_j) = h \sum_{i=0}^{N} \omega_{Ni}\, K(x_j, x_i) Y_i \,, \qquad j = 0, \ldots, N \,,$$

where $x_j = a + jh$, in the usual way.

In general the solutions Y_i are found to oscillate wildly about the true solution $y(x_i)$ and the numerical results are worthless. This result led to more complicated smoothing techniques (see Phillips [19]). We might therefore be justified in being cautious in attempting to solve (1-1-1) by a quadrature method. In sections 2 and 3 quadrature methods are investigated.

Section 2 MULTISTEP METHODS I: AN EXAMINATION OF SOME SIMPLE RULES

2 - 1: Convergence and stability considerations

Consider the grid of equally spaced points $x_i=x_o+ih$ with $x_o=a$, $x_N=b$. If we write $x=x_n$ $(n\le N)$ in (1-1-1) and replace the integral by a quadrature with weight $w_{n,i}$ $(i=o,\dots,n)$ we obtain

$$h\sum_{i=0}^{n} \omega_{ni}\, K(x_n,x_i)\,Y_i = g(x_n), \qquad \dots.(2\text{-}1\text{-}1)$$

where Y_i is the numerical approximation to $y(x_i)$. This can be interpreted as a scheme for Y_n,

$$Y_n = \frac{g(x_n)}{h\,\omega_{nn}K(x_n,x_n)} - \sum_{i=0}^{n-1} \frac{\omega_{ni}\,K(x_n,x_i)}{\omega_{nn}\,K(x_n,x_n)}\,Y_i, \qquad \dots.(2\text{-}1\text{-}2)$$

If Y_i, $i=o,\dots,n-1$ are known and $w_{nn}\neq 0$. From (2-1-2) we see that we cannot obtain a general convergence result as we did in Chapter 2 for the equation of the second kind. This problem is discussed in Section 3.

The concept of numerical stability introduced in Chapter 2 can be carried over to this chapter. If we perturb $g(x)$ to $g(x)+\delta g(x)$, this produces a perturbation $\delta y(x)$ in $y(x)$ which satisfies

$$\delta g(x) = \int_a^x K(x,t)\,\delta y(t)\,dt.$$

We thus have the following definition.

Definition 1 A multistep method is numerically stable if the leading term $e(x)$ in an asymptotic expansion of the error due to discretization satisfies an equation of the form

$$h(x) = \int_a^x K(x,t)\,e(t)\,dt.$$

In other words, e(x) grows in essentially the same way as the solution y(x). As before, this definition does not imply that the errors will always be small, since the equation (1-1-1) may be ill-conditioned.

Another, more intuitive, concept of stability has been suggested by Noble [16]. Again, consider the equation

$$\int_0^x y(t)\,dt = g(x) \qquad \ldots\ldots(2\text{-}1\text{-}3)$$

with solution $y(x)=g'(x)$. Clearly a suitable multistep method is one which produces approximations $Y(x_i,h)$ which satisfy

$$\lim_{h\to 0} Y(x_i,h) = g'(x_i)$$

(i fixed)

In the following sections we shall find it instructive to examine whether a given scheme has this property. Indeed we shall find it easier to examine this property than to determine whether a scheme is stable in the sense of definition 1.

2 - 2: The Euler and midpoint schemes

If we approximate to the integral in each interval $[x_i,x_{i+1}]$ by the rectangular rule we obtain the Euler scheme

$$h\sum_{i=0}^{n-1} K(x_n,x_i)Y_i = g(x_n)$$

which we may express as

$$Y_{n-1} = \frac{g(x_n)}{hK(x_n,x_{n-1})} - \sum_{i=0}^{n-2} \frac{K(x_n,x_i)}{K(x_n,x_{n-1})}\,Y_i \qquad , \quad n=1,2,\ldots$$

Similarly, by approximating to the integral over $[x_i,x_{i+1}]$ by its midpoint value $h\,K(x_n,x_{i+\frac12},Y_{i+\frac12})$ we find the midpoint scheme

$$Y_{n-1/2} = \frac{g(x_n)}{hK(x_n,x_{n-1/2})} - \sum_{i=0}^{n-2} \frac{K(x_n,x_{i+1/2})}{K(x_n,x_{n-1/2})}\,Y_{i+1/2} \qquad n=1,2,\ldots$$

Here $x_{i+\frac{1}{2}}$ denotes the value $x_i+h/2$.

Let us now apply the above methods to

$$\int_0^x y(t)\,dt = g(x) \qquad \ldots(2\text{-}2\text{-}1)$$

to obtain an intuitive idea of the stability of the scheme.

For the Euler scheme, if we subtract the expressions for Y_n and Y_{n-1} we obtain

$$\begin{aligned} \gamma_n &= g(x_{n+1}) - g(x_n)/h \\ &= g'(x_n) + O(h) \end{aligned} \qquad \ldots.(2\text{-}2\text{-}2)$$

In the same way we obtain the following for the midpoint method.

$$\begin{aligned} \gamma_{n+\frac{1}{2}} &= g(x_{n+1}) - g(x_n)/h \\ &= g'(x_{n+1/2}) + O(h^2). \end{aligned}$$

From these results we would expect, intuitively, that the Euler and midpoint methods produce converging, stable results for the general Volterra equation, and that the order of convergence is respectively one and two. Linz [15] confirms these results by rigorous proofs. The results are summarised in the following theorem:

Theorem 1 The Euler and midpoint methods are convergent of orders 1 and 2 respectively and the errors $\varepsilon_n = y(x_n) - Y_n$, $n=0,1,\ldots$ satisfy

$$\varepsilon_n = h\,e(x_n) + O(h^2) \qquad \text{[Euler method]}$$

$$\varepsilon_{n+1/2} = h^2\,e(x_{n+1/2}) + O(h^3) \qquad \text{[midpoint method]}$$

where $e(x)$ satisfies a Volterra equation of the form

$$e(x) = -\int_a^x \frac{\partial K/\partial x\,(x,t)}{K(x,x)}\,e(t)\,dt + \ell(x)$$

We conclude that both methods are numerically stable.

2 - 3: The trapezoidal method

Approximating to the integral in $[x_i, x_{i+1}]$ by the trapezoidal rule, we obtain

$$g(x_n) = \frac{h}{2} \sum_{i=0}^{n-1} \{ K(x_n, x_i) Y_i + K(x_n, x_{i+1}) Y_{i+1} \}.$$

We notice that in order to produce a scheme for Y_i, $i=1,2,\ldots$ we require a starting value Y_0. Applying the method to equation (2-2-2) we find

$$g(x_{n+1}) - g(x_n) = h/2 \, (Y_n + Y_{n+1}).$$

This difference equation has solution

$$Y_{n+1} = \frac{2}{h} g(x_{n+1}) + \frac{4}{h} \sum_{i=1}^{n} (-1)^i g(x_{n+1-i}) + (-1)^{n+1} Y_0 \qquad \ldots(2\text{-}3\text{-}1)$$

Clearly the right-hand side cannot be interpreted as a suitable approximation to $g'(x_{n+1})$. We thus except the method to possess undesired stability properties. This result has been confirmed by Kobayasi [13]. Both Linz [15] and Kobayasi establish convergence of the scheme, and the results are summarised in the following theorem.

Theorem 2 (a) The trapezoidal method is convergent of order 2

(b) Suppose that $K(x,t)$ is differentiable in both its arguments and that $y(x)$ is differentiable. Now let $e(x)$ be the solution of the Volterra equation

$$\int_a^x K(x,t)\, e(t)\, dt = -1/12 \left[\partial/\partial t \, (K(x,t)\, y(t)) \right]_{t=a}^{t=x},$$

and let $d(x)$ be the solution of the initial value problem

$$K(x,x)\, d'(x) + \frac{\partial K}{\partial t}(x,t)_{t=x}\, d(x) = 0 \quad , \quad d(a) = \delta_0,$$

where $\delta_0 = \begin{cases} -e(a) & q \geq 3 \\ \varepsilon_0 - e(a) & q = 2 \\ \varepsilon_0 & q = 1 \end{cases}$

Then if the starting error ε_0 in the trapezoidal method is $O(h^q)$, the error ε_n satisfies

$$\varepsilon_n = h^2 e(x_n) + (-1)^n h^r d(x_n) + O(h^{r+1}).$$

where r is the minimum value of (2,q).

We see that the asymptotic error expansion contains a term which oscillates with the value of n. For a particular kernel this term could dominate the term in $e(x_n)$, and we deduce the scheme is weakly unstable. We now examine the effect of this term in two examples.

Example 1 Consider the equation

$$\int_0^x (1+x-t) y(t)\, dt = -1 + x + e^{-x}$$

with solution

$$y(x) = xe^{-x}$$

We find

$$e(x) = 1/12\,(x-1)(x-4)e^{-x}, \quad d(x) = \delta_0 e^x$$

For an exact starting value we have $r=2$, $\delta_0 = -1/3$.

Then

$$\varepsilon_n = h^2 e(x_n) + (-1)^n h^2 d(x_n) + O(h^3)$$

$$= 1/12\,(x_n-1)(x_n-4)h^2 e^{-x_n} - (-1)^n h^2/3\, e^{x_n} + O(h^3).$$

We notice that e(x) is growing in the same way as the solution of the equation, but that d(x) is behaving in the opposite way. As n increases the second term in the expansion of ε_n will dominate the term in $e^{-x}n$ and eventually the numerical solution will bear no resemblance to the true solution. The effect of the oscillating positive exponiential factor is shown clearly in Figure 1.

Example 2 Consider the equation

$$\int_0^x \cos(x-t)\,y(t)\,dt = \sin x,$$

with solution

$$y(x) = 1$$

We find $e(x) = 1/12$, $d(x) = \delta_0$.

For an exact starting value $r = 2$, $\delta_0 = -1/12$

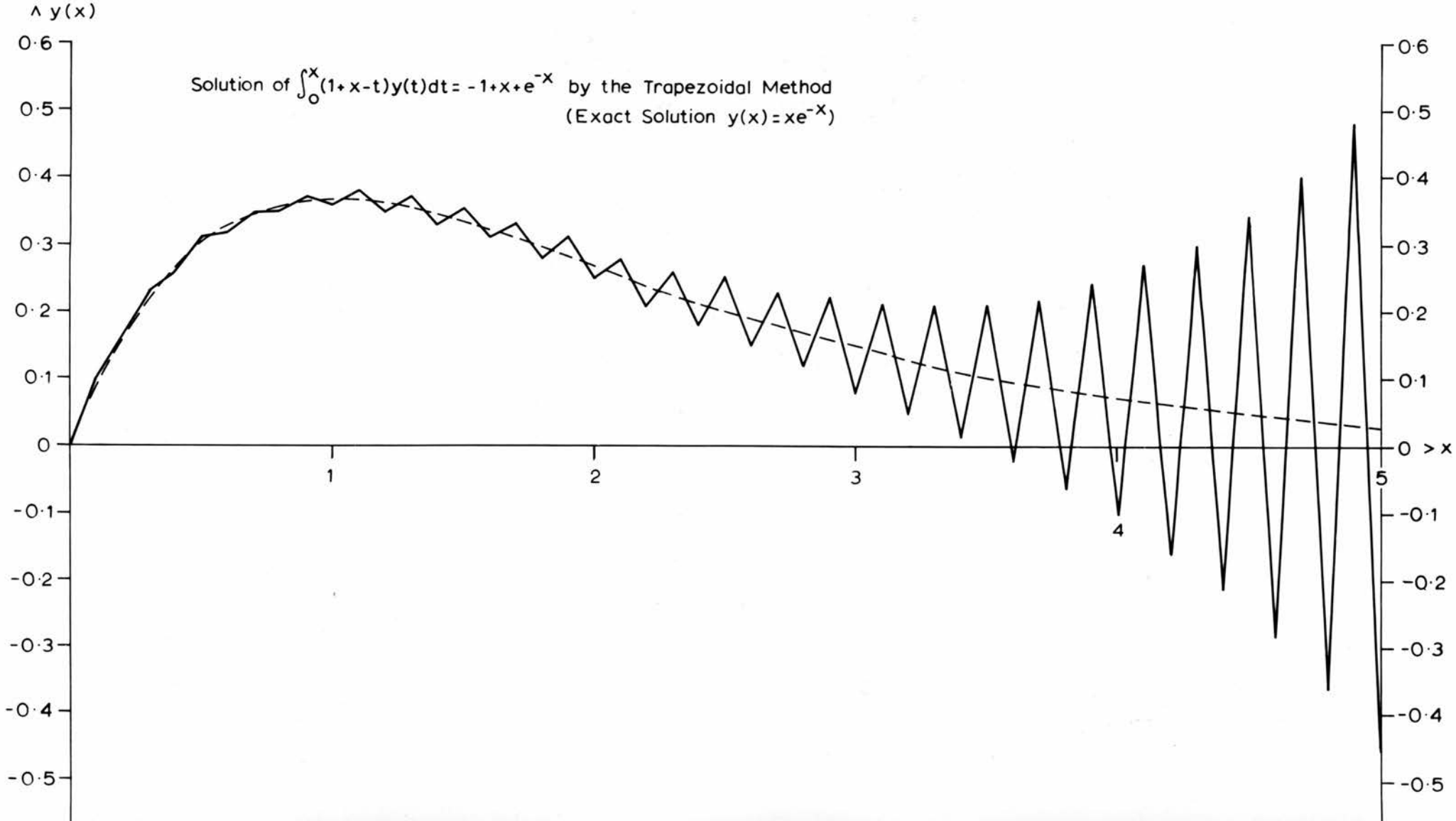
Then

$$\varepsilon_n = h^2/12 - (-1)^n h^2/12 + O(h^3).$$

Here the effect of the oscillating term is minor. For n even the first two terms in the expansion cancel and we can expect errors of $O(h^3)$. For n odd, the terms sum to give an error $h^2/6+O(h^3)$. Numerical results confirm these statements and are given in Table 1.

TABLE 1 : THE TRAPEZOIDAL METHOD

VALUE OF X	ERROR IN NUMERICAL SOLUTION
0.1	0.1664×10^{-2}
0.2	0.0
0.3	0.1664×10^{-2}
0.4	0.0
0.5	0.1664×10^{-2}
0.6	0.0


∧ y(x)
Solution of $\int_0^x (1+x-t)y(t)dt = -1+x+e^{-x}$ by the Trapezoidal Method
(Exact Solution $y(x) = xe^{-x}$)
0·6
0·5
0·4
0·3
0·2
0·1
0
-0·1
-0·2
-0·3
-0·4
-0·5
1
2
3
4
5
> x

The step-size h was chosen as 0.1. The pattern of values of the errors in Table 1 continued throughout all the values calculated (up to x=5). The errors were calculated up to ten decimal places, and for this level of accuracy, every alternate error was found to be zero, as listed above.

2 - 4: Higher-order methods

For higher order quadrature, approximate solutions of (2-2-1) in the form (2-2-2) or (2-3-1) cannot be obtained. Also, it is not possible to obtain any convergence results by the methods used in the proofs of Theorems 1 and 2. A different approach to the convergence problem is needed and we shall see in the next chapter how a theory is developed. However, it is interesting to first examine the growth of errors produced by some standard quadrature when applied to the simple equation (2-2-1). Let us examine the Gregory rule with k=1. (See Chapter 2, Section 1 - 3)

Gregory's rule with k=1

In this case we find, analogous to (2-4-1),

$$-\tfrac{1}{12}\varepsilon_{n-1} + \tfrac{8}{12}\varepsilon_n + \tfrac{5}{12}\varepsilon_{n+1} = O(h^2).$$

The characteristic polynomial equation is

$$5z^2 + 8z - 1 = 0,$$

which has a root in $[-1,-2]$. The method therefore produces diverging results for the equation.

Clearly we see that higher order methods need to be examined carefully for their convergence properties. In the next section a convergence criterion is established, and then applied to standard quadrature methods.

Thus

$$^{7}/_{24}\,\varepsilon_{2r-2} + {}^{5}/_{24}\,\varepsilon_{2r-1} - {}^{19}/_{24}\,\varepsilon_{2r} - {}^{3}/_{8}\,\varepsilon_{2r+1} = O(h^3) \qquad \ldots(2\text{-}4\text{-}1)$$

The homogeneous equation has characteristic polynomial equation

$$9z^3 + 19z^2 - 5z - 7 = 0.$$

This equation has a root in $[-2,-3]$ and consequently $\varepsilon_{2r} \to \infty$ as $r \to \infty$

We see that the method is divergent for the equation (2-2-1), and convergence for a more general Volterra equation seems unlikely.

(ii) Gregory's rule with k=1

In this case we find, analogous to (2-4-1),

$$-{}^{1}/_{12}\,\varepsilon_{n-1} + {}^{8}/_{12}\,\varepsilon_{n} + {}^{5}/_{12}\,\varepsilon_{n+1} = O(h^2).$$

The characteristic polynomial equation is

$$5z^2 + 8z - 1 = 0,$$

which has a root in $[-1,-2]$. The method therefore produces diverging results for the equation.

Clearly we see that higher order methods need to be examined carefully for their convergence properties. In the next section a convergence criterion is established, and then applied to standard quadrature methods.

Section 3: MULTISTEP METHODS II: A CONVERGENCE CRITERION

3 - 1 The determination of a convergence criterion

Holyhead [11] has recently produced a theory which gives a necessary and sufficient condition for the convergence of the multistep method (2-1-2). In the analysis which follows we shall be making use of the concept of a repetition factor, previously introduced in Chapter 2.

Suppose the multistep method (2-1-2) requires s starting values $\tilde{Y}_i$, i=o,...,s-1 where

$$\tilde{\varepsilon}_i = y(x_i) - \tilde{Y}_i \quad , \qquad i=0,\dots,s-1,$$

and that the quadrature in (2-1-2) is convergent of order h^P. From (2-1-1) we have

$$h \sum_{i=0}^{n} \omega_{ni}\, K(x_n, x_i)\, \varepsilon_i = R_{n,h} \qquad n=s,\dots,N$$

with

$$\varepsilon_i = \tilde{\varepsilon}_i \qquad i=0,\dots,s-1 \quad , \qquad \dots.(3\text{-}1\text{-}1)$$

where $R_{n,h}$ is the truncation error the quadrature.

Equation (3-1-1) has the matrix representation

$$\bar{\underline{\Psi}}_h \underset{\sim}{\varepsilon} = \underset{\sim}{r} \qquad \dots.(3\text{-}1\text{-}2)$$

where

$$\underset{\sim}{\varepsilon} = \{\varepsilon_0, \dots, \varepsilon_N\}^T$$

$$\underset{\sim}{r} = \{h\tilde{\varepsilon}_0, \dots, h\tilde{\varepsilon}_{s-1}, R_{s,h}, \dots, R_{N,h}\}^T.$$

The matrix $\bar{\underline{\Psi}}_h$ is NxN and is lower triangular. Its first r rows are those of the unit matrix, the remainder of its rows contain the

elements $\{\psi_{h\,i,j}\}$ where

$$\psi_{h\,i,j} = h\,\omega_{ij}\,K(x_i, x_j) \qquad i = s, \ldots, N$$

$$j = 0, \ldots, i.$$

As an example let us consider a method using Simpson's rule as the 'main' rule, together with an ancillary rule placed at the end of every alternate row. Suppose the weights in this ancillary rule are d_0, d_1, d_2, d_3. The matrix $\underline{\Psi}_h$ is shown in figure 2. [We have set $K(x,t)=1$ here for clarity.]

FIGURE 2 : THE MATRIX $\underline{\Psi}_h$ FOR A SIMPSON RULE METHOD

$$\underline{\Psi}_h = h \begin{bmatrix}
1 & & & & & & & & & & \\
0 & 1 & & & & & & & & & \\
1/3 & 4/3 & 1/3 & & & & & & & & \\
d_0 & d_1 & d_2 & d_3 & & & & & & & \\
1/3 & 4/3 & 2/3 & 4/3 & 1/3 & & & & & & \\
\updownarrow & & & & & & & & & & \\
1/3 & 4/3 & 1/3+d_0 & d_1 & d_2 & d_3 & & & & & \\
\updownarrow & \updownarrow & & & & & & & & & \\
1/3 & 4/3 & 2/3 & 4/3 & 2/3 & 4/3 & 1/3 & & & & \\
\updownarrow & \updownarrow & \updownarrow & & & & & & & & \\
1/3 & 4/3 & 2/3 & 4/3 & 1/3+d_0 & d_1 & d_2 & d_3 & & & \\
\updownarrow & \updownarrow & \updownarrow & \updownarrow & & & & & & & \\
1/3 & 4/3 & 2/3 & 4/3 & 2/3 & 4/3 & 2/3 & 4/3 & 1/3 & & \\
\updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & & & & & & \\
1/3 & 4/3 & 2/3 & 4/3 & 2/3 & 4/3 & 1/3+d_0 & d_1 & d_2 & d_3 & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}$$

The method has repetition factor $\rho = 1$, and this repetition is indicated by the arrows in figure 2. We also notice that a 'block' of elements is repeated in a diagonal fashion from the sixth row of

$\underline{\Psi}_h$ onwards, and that each block extends over two rows, and over five elements in each row. For a general multistep method, the blocks will each extend over μ rows, and over λ elements in each row. The repetition will begin at row$(v+1)$, say. It will prove convenient to denote $\rho = \lambda - 1$. The type of repetition indicated in figure 2 is common to all multistep methods. For methods with $\rho > 1$, it is found that in every case we have $\mu = 1$.

Let us now return to (3-1-2). We can write

$$\| \tilde{\varepsilon} \| \leq \| \bar{\underline{\Psi}}_h^{-1} \| . \| \tilde{\tau} \|$$

where

$$\| \tilde{\varepsilon} \| = \max_{0 \leq i \leq N} | \varepsilon_i | \quad , \quad \| \bar{\underline{\Psi}}_h^{-1} \| = \max_i \sum_{j=0}^{N} | \Psi_h^{-1}{}_{i,j} |$$

where $\{ \Psi_h^{-1}{}_{i,j} \}$ are the elements of $\bar{\underline{\Psi}}_h^{-1}$.

Now if we suppose that each $\tilde{\varepsilon}_i = O(h^{p-1})$, then we can write

$$\| \tilde{\varepsilon} \| \leq \| \bar{\underline{\Psi}}_h^{-1} \| \, C h^p \qquad \text{for some } C > 0.$$

What we now desire is a bound on $\| \bar{\underline{\Psi}}_h^{-1} \|$ which is of the form

$$\| \bar{\underline{\Psi}}_h^{-1} \| \leq M h^{-r} \qquad \ldots . (3\text{-}1\text{-}3)$$

where $r < p$, for some integer r and some $M > 0$, which is independent of h. This would give us a convergent method of order $p - r$.

Holyhead [11] has obtained a criterion for a multistep method to satisfy in order to produce a bound (3-1-3). We first premultiply $\bar{\underline{\Psi}}_h$ by $K_h^{-1} D_h$, where

$$K_h = \text{diag}\,(1, \ldots, 1, \omega_{ss} K_{ss}, \ldots, \omega_{NN} K_{NN}).$$

with K_{ss} denoting $K(x_s, x_s)$. The matrix D_h is a "differentiation matrix" and is shown in figure 3. Now, from row $(v+1)$ onwards we find that an element a_{ij} of $K_h^{-1} D_h \underline{\Psi}_h$ is given by

$$a_{ij} = \begin{cases} \omega_{ij} K_{ij} - \omega_{i-\rho,j} K_{i-\rho,j} & j = 0, \ldots, i-\rho \\ \omega_{ii} K_{ii} & \\ \dfrac{\omega_{ij} K_{ij}}{\omega_{ii} K_{ii}} & j = i-\rho+1, \ldots, i-1 \\ 1 & j = i \end{cases}$$

However, for $j=0,\ldots,i-\lambda$, we have $\omega_{ij} = \omega_{i-\rho,j}$. Hence, using Taylor's theorem gives

$$a_{ij} = \begin{cases} O(h) & j = 0, \ldots, i-\lambda \\ \dfrac{\omega_{ij} - \omega_{i-\rho,j}}{\omega_{ii}} + O(h) & j = i-\lambda+1, \ldots, i-\rho \\ \dfrac{\omega_{ij}}{\omega_{ii}} + O(h) & j = i-\rho+1, \ldots, i-1 \\ 1 & j = i \end{cases}$$

We can thus write

$$K_h^{-1} D_h \underline{\Psi}_h = A_N + h L_h.$$

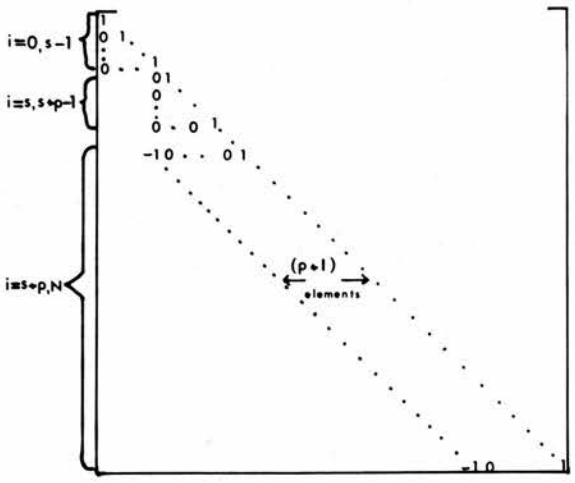
where A_N and L_h are shown in figure 3. For $i = v, \ldots, N$ the $(i+1)^{th}$ row of A_N is

$$\left(0, \ldots, 0, \frac{\omega_{i,i-\lambda+1} - \omega_{i-\rho,i-\lambda+1}}{\omega_{ii}}, \ldots, \frac{\omega_{i,i-\rho} - \omega_{i-\rho,i-\rho}}{\omega_{ii}}, \right.$$

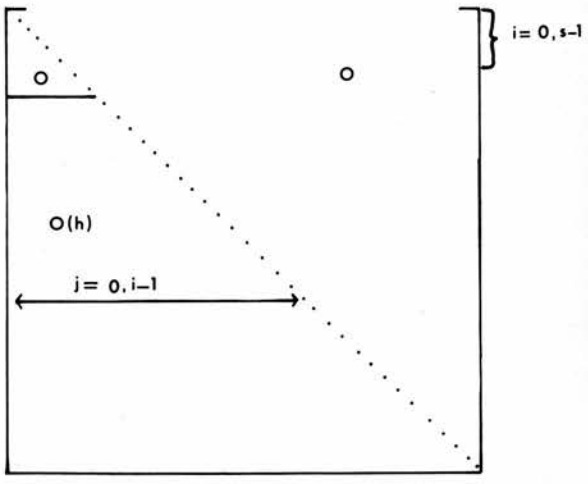
$$\left. \frac{\omega_{i,i-\rho+1}}{\omega_{ii}}, \ldots, \frac{\omega_{i,i-1}}{\omega_{ii}}, 1, 0, \ldots, 0 \right) \qquad \ldots.(3\text{-}1\text{-}4)$$

That is, $(0, \ldots, \alpha_{i1}, \ldots, \alpha_{ip}, 1, 0, \ldots, 0)$ $\ldots.(3\text{-}1\text{-}5)$

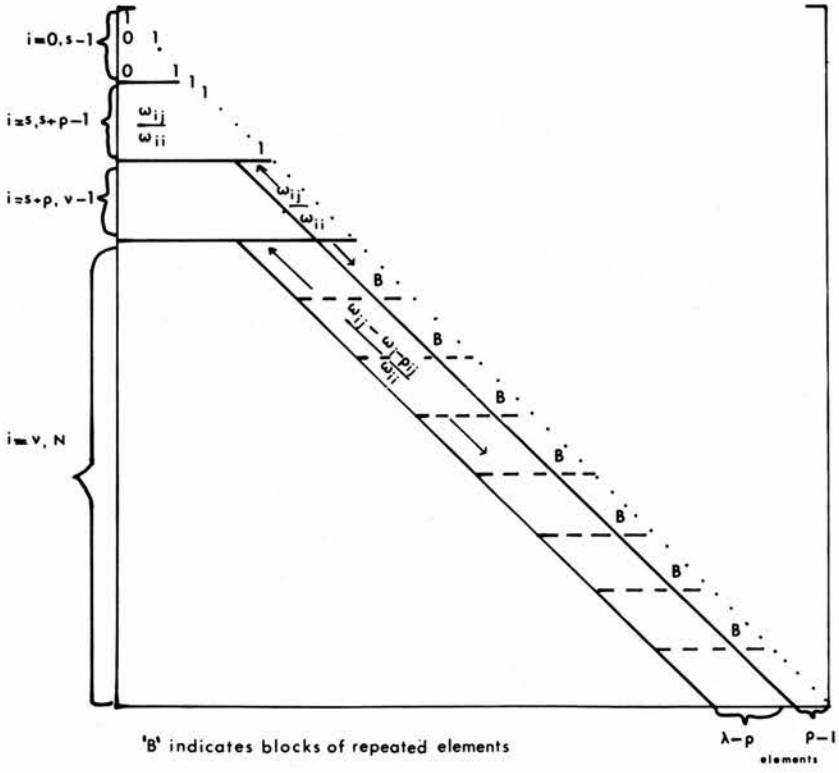
where there are $(i+1-\lambda)$ zeros at the beginning of the row and



THE MATRIX D_h



THE MATRIX L_h



THE MATRIX A_N

Figure 3

$p = \lambda - 1$, as before. As shown in figure 2, it is found that the matrix A_N possesses the same repetition quality as $\underline{\Psi}_h$. That is, blocks of size $\mu \times p$ are repeated down its length. It was this property that enabled Holyhead to prove the following important theorem.

Theorem 3 The matrix A_N is uniformly bounded with respect to N if and only if $G = G_\mu \; G_{\mu-1} \; \ldots \; G_1$ has all its eigenvalues lying strictly within the unit circle, where

$$G_{i-v+1} = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \\ -\alpha_{i1} & \cdots & \cdots & \cdots & \cdots & \cdots & -\alpha_{ip} \end{bmatrix} \quad , i = v, v+1, \ldots, v+\mu-1$$

and the α_{ij} $1 \le j \le p$ are given in equations (3-1-4) and (3-1-5).

From this theorem, Holyhead shows that

$$\| (A_N + h L_h)^{-1} \| \le M_3 \qquad \text{for some } M_3 > 0 \text{ independent of h.}$$

Now, since $K_h^{-1} D_h \underline{\Psi}_h = A_N + h L_h$,

we have $$\underline{\Psi}_h^{-1} = (A_N + h L_h)^{-1} K_h^{-1} D_h$$

Now $\| K_h^{-1} \|$ is uniformly bounded with respect to N and $\| D_h \| = 2/h$; thus we obtain the result

$$\| \underline{\Psi}_h^{-1} \| \le M/h \qquad , \text{ for some M independent of h.}$$

A sufficient condition for convergence has therefore been established. Holyhead states that for the necessary condition we must have the additional property that all eigenvalues of G which lie on the unit

circle are roots of $\lambda^{\rho}=1$. The results of this section are summed up in Theorem 4 below.

Theorem 4 If the quadrature of (2-1-3) is convergent of order h^{p} and the starting values are convergent of order h^{p-1}, then a necessary and sufficient condition for the multistep method (2-1-2) to be convergent of order $p-1$ that all the eigenvalues of

$$G = G_{\mu} \ldots . G_{1} \qquad \text{(with } G_{i} \text{ as in Theorem 3)}$$

be inside or on the unit circle, and that those on the unit circle are roots of

$$\lambda^{\rho}=1$$

where ρ is the repetition factor of the method.

3 - 2 A stability criterion

The results of section 3-1 can be extended to include a stability criterion. Noble's stability argument (Section 2-1) requires that when $K(x,t)=1$, the multistep method should reduce to a "differentiation rule" for $g(x)$ for each value of $x=x_i$, $i>s$. Now when $K(x,t)=1$, the multistep method (2-1-2) reduces to

$$h\sum_{i=0}^{n} \omega_{ni}\,\gamma_i = g(x_n) \qquad i=s,s+1,\ldots,N.$$

That is

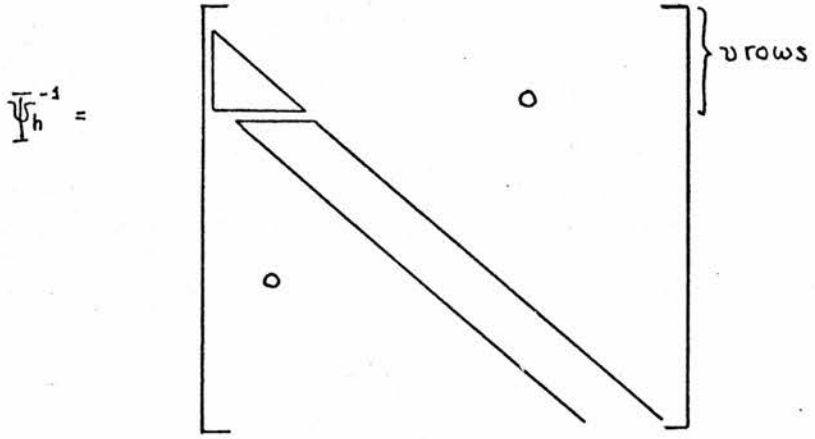
$$\underset{\sim}{\gamma} = \underline{\Psi}_h^{-1}\,\underset{\sim}{g}$$

with

$$\underset{\sim}{\gamma} = \{\tilde{\gamma}_0,\ldots,\tilde{\gamma}_{s-1},\gamma_s,\ldots,\gamma_N\}^{T}, \quad \underset{\sim}{g} = \{g(x_0),\ldots,g(x_N)\}^{T}$$

We therefore require that $\underline{\Psi}_h^{-1}$ be <u>banded</u>. (See figure 4)

FIGURE 4 : REQUIRED FORM OF $\bar{\Psi}_h^{-1}$ FOR STABILITY



Now when $K(x,t)=1$, all elements of L_h are zero. Thus

$$A_N = K_h^{-1} D_h \bar{\Psi}_h,$$

that is

$$\bar{\Psi}_h^{-1} = A_N^{-1} K_h^{-1} D_h.$$

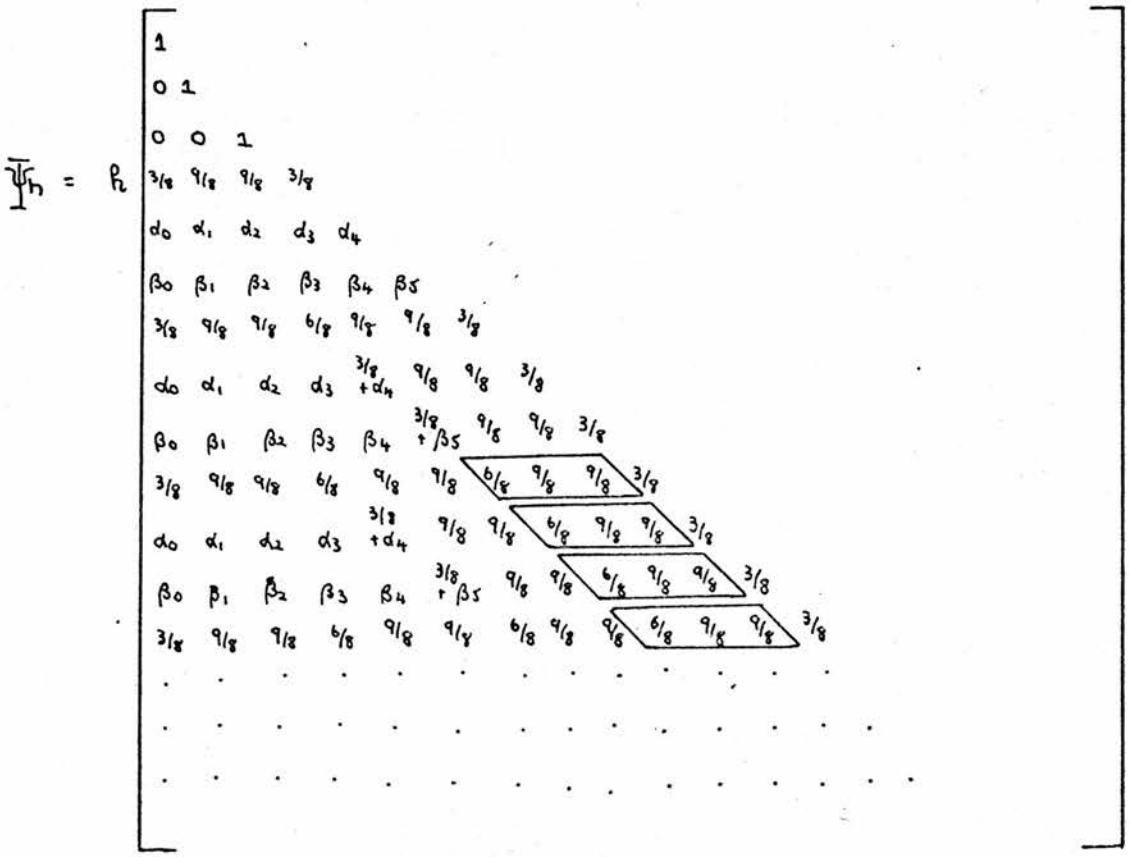
We see we now require A_N^{-1} to be banded, and this in turn requires that all the eigenvalues of G are zero. [See Holyhead for details.] This result is stated in Theorem 5.

Theorem 5 A convergent multistep method is stable, in the sense of Noble, if and only if all eigenvalues of G are zero.

3 - 3 Applications - methods with a repetition factor greater than one

Specific examples of multistep methods are now investigated by the author for their convergence and stability properties. Newton-Cotes rules with ancillary quadratures placed at the beginning of the rows of $\bar{\Psi}_h$ will always produce methods with $\rho>1$. For example, suppose we are using the three-eights rule as the main rule. We will then require two ancillary rules. The matrix $\bar{\Psi}_h$ (with $K(x,t)=1$) is shown in figure 5.

FIGURE 5 : THE THREE-EIGHTH'S RULE



Here $v=9$, $\rho=3$, $\mu=1$ and $p=3$. The α_i, $i=0,\ldots,4$; β_i, $i=0,\ldots,5$ are the weights in the ancillary quadratures.

In general, suppose that the weights in the repeated main quadrature are $\omega_0, \omega_1, \ldots, \omega_n$. Then, for the class of methods considered here we have

$$\mu=1,\ \rho=n=p\ , \qquad \text{and so} \quad \lambda-\rho=1$$

The first element in the last row of $G=G_1$ is $-\{(\omega_0+\omega_n)-\omega_n\}$

$$= -\omega_0/\omega_n.$$

The i^{th} element $(2\le i\le n)$ in the last row of G is $-\omega_{i-1}/\omega_n$.

We find

$$G-\lambda I = \begin{bmatrix} -\lambda & 1 & & & & 0 \\ 0 & -\lambda & 1 & & & \\ \vdots & & \ddots & \ddots & & \\ & & & -\lambda & 1 & \\ -\omega_0/\omega_n & & \cdots & & \left(-\frac{\omega_{n-1}}{\omega_n}-\lambda\right) & \end{bmatrix}$$

On expanding

$$|G-\lambda I|=0$$

We obtain

$$\lambda^n(-1)^n + \lambda^{n-1}\omega_{n-1}/\omega_n(-1)^n + \ldots + (-1)^n\omega_0/\omega_n = 0$$

That is

$$\omega_n\lambda^n + \omega_{n-1}\lambda^{n-1} + \ldots + \omega_1\lambda + \omega_0 = 0$$

Now if the weights are those of a Newton-Cotes rule we have the symmetry property

$$\omega_i = \omega_{n-i} \qquad i = 0,\ldots,n$$

If n is odd, we can immediately observe that a zero of

$$p(\lambda) = \omega_n\lambda^n + \ldots + \omega_1\lambda + \omega_0$$

occurs at $\lambda=-1$

For convergence we require that any root on the unit circle must be roots of

$$\lambda^p = 1$$

That is

$$\lambda^n = 1$$

This condition is not satisfied and we conclude that Newton-Cotes methods with an even number of weights and with a repetition factor greater than one yield diverging methods.

Let us now examine the case when n is even. For the Newton-Cotes weights we have $\omega_n = \omega_0 > 0$. Thus as $\lambda \to -\infty, p(\lambda) \to \infty$. Now let $n=2r$, say.

Then

$$p(-1) = w_{2r} - w_{2r-1} + w_{2r+2} + \cdots + w_2 - w_1 + w_0$$

$$= 2\{w_0 - w_1 + w_2 + \cdots + (-1)^r w_{r-2} - (-1)^r w_{r-1}\} + (-1)^r w_r$$

Now if $p(-1) < 0$, we have a root in $(-\infty, -1)$. It can easily be verified that all the Newton-Coles rules up to the case n=8 (at least) satisfy this condition.

We conclude therefore that <u>any</u> practical Newton-Cotes method (with $\rho > 1$) will yield diverging methods.

3 - 4 Applications - Methods with a repetition factor of one

In these methods the main rule is applied at the beginning of each row with ancillary quadratures needed at the ends of rows. If we first consider Newton-Cotes rules, we find we need to consider each case separately. We consider a Simpson rule method as follows.

(i) <u>Simpson's Rule</u> Suppose we use a repeated Simpson rule together with an ancillary quadrature with weights $\alpha_0, \alpha_1, \alpha_2, \alpha_3$. The matrix $\underline{\bar{\psi}}_h$ is shown in figure 2. For this method $v=5$, $\mu=2$ and $\rho=4$

Thus

$$G = G_2 G_1$$

where

$$G_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & b_1 & b_2 & b_3 \end{bmatrix}, \qquad G_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ a_1 & a_2 & a_3 & a_4 \end{bmatrix},$$

with

$$a_1 = 3\alpha_0 - 1$$

$$a_2 = 3\alpha_1 - 4$$

$$a_3 = 3\alpha_2 - 3, \quad a_4 = 3\alpha_3 - 4$$

$$b_1 = (1 - 3\alpha_0)/3\alpha_3$$

$$b_2 = (4 - 3\alpha_1)/3\alpha_3$$

$$b_3 = (1 - 3\alpha_2)/3\alpha_3$$

Therefore

$$G = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & b_1 & b_2 & b_3 \\ 0 & a_1 + a_4 b_1 & a_2 + a_4 b_2 & a_3 + a_4 b_3 \end{bmatrix}$$

The eigenvalues of G are solutions of $\lambda(\lambda^3 + C_1\lambda^2 + C_2\lambda + C_3)=0$ where the C_i are combinations of the a_i and b_i.

For any choice of α_i, $i=0,\ldots,3$, we find $C_3=0$. [Using $C_3=|G|$.]

Evaluating $|G-\lambda I|=0$,

we find that

$$\lambda^2 - \{a_4b_3 + b_2 + a_3\}\lambda + \{b_2a_3 - b_3a_2 - a_1 - a_4b_1\}=0$$

Substituting, we find

$$d_3\lambda^2 - \{d_1 + d_3 - 4d_2\}\lambda + \{d_1 - 4d_0\}=0 \quad \ldots.(3\text{-}4\text{-}1)$$

is the equation satisfied by the remaining eigenvalues.

The most obvious choice of the α_i is $d_0=3/8$, $d_1=d_2=9/8$, $d_3=3/8$, or $d_0=1/2$, $d_1=d_2=1$, $d_3=1/2$. It is easily seen however that both sets of weights give a root of

$$\rho(\lambda)= d_3\lambda^2 - \{d_1 + d_3 - 4d_2\}\lambda + \{d_1 - 4d_0\}$$

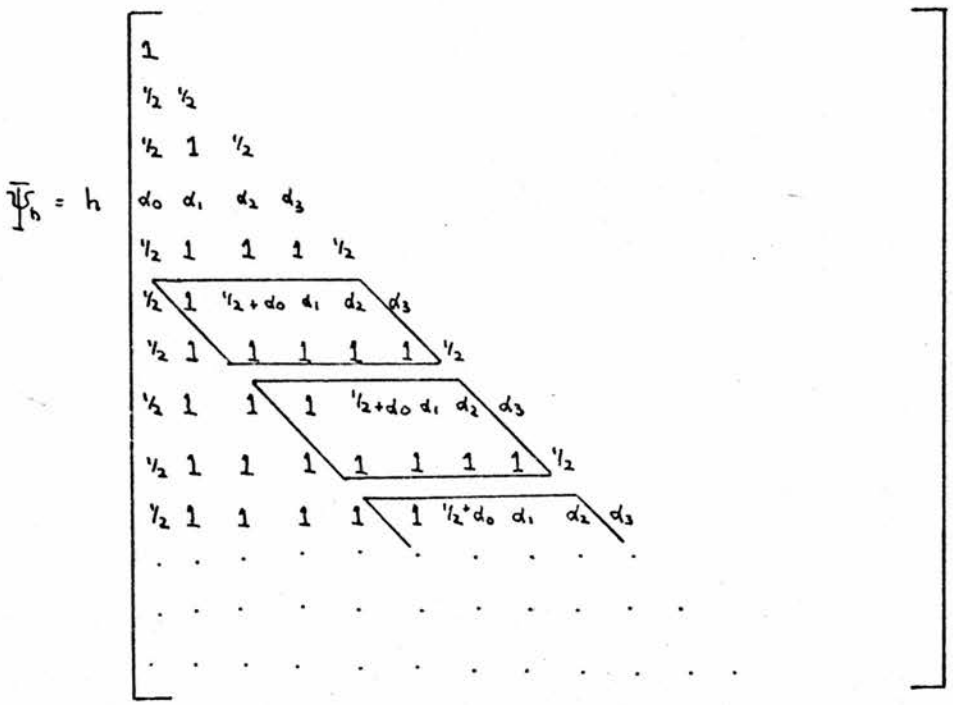
outside the unit circle.

Clearly, there is no simple way in which Simpson's rule can be used successfully. Now the desired properties of any scheme are convergence and stability. For the method being considered we could ensure this by showing the ancillary rule to be $O(h^2)$ in accuracy. We would then obtain two free parameters in the weights which could be chosen so that both eigenvalues in (3-4-1) become zero. This is not really satisfactory, however, as we would only obtain a method which is O(h) in accuracy. In the next sub-section, however, the author puts this idea, (due to Holyhead), into better use.

(ii) A stable trapezoidal rule We have already seen that the trapezoidal rule, applied in a direct way, gives a convergent method for the integral equation, but that the solutions obtained are subject to instability. (Section 2-3). However by combining the trapezoidal rule with an ancillary rule, the author now shows that stability can be enforced.

Suppose we place the ancillary quadrature, with weights d_0, d_1, d_2, d_3 at the end of every alternate row. The matrix $\bar{\Psi}_h$ is then as shown in figure 6.

FIGURE 6 : THE TRAPEZOIDAL METHOD



For this method $v=5$, $\mu=2$, $p=4$.

Let us choose the ancillary quadrature so that it has a truncation error $O(h^2)$. We then have a system analogous to that of (i). That

is, we have two zero eigenvalues of the matrix G, and, from (3-4-1), the remaining two satisfy

$$\lambda^2 - \{a_4 b_3 + b_2 + a_3\}\lambda + \{b_2 a_3 - b_3 a_2 - a_1 - a_4 b_1\} = 0$$

where, in this case

$$b_1 = 1 - 2d_0/2d_3 \quad , b_2 = 1 - d_1/d_3 \quad , b_3 = 1 - 2d_3/2d_3$$

$$a_1 = 2d_0 - 1 \, , \; a_2 = 2(d_1 - 1) \, , \; a_3 = 2(d_2 - 1) \, , \; a_4 = 2(d_3 - 1)$$

We find that

$$\lambda^2 + \left(\frac{d_1 + d_3 - 2d_2}{d_3}\right)\lambda + \left(\frac{d_1 - 2d_0}{d_3}\right) = 0$$

For both eigenvalues to be zero, we require that

$$d_1 = 2d_0 \qquad \ldots.(3\text{-}4\text{-}2)$$

$$d_2 = \frac{d_1 + d_3}{2} \, . \qquad \ldots.(3\text{-}4\text{-}3)$$

For the rule to have a truncation error $O(h^2)$ we require

$$d_0 = 3x$$

$$d_1 = 3y$$

$$d_2 = 9/2 - 9x - 6y$$

$$d_3 = -3/2 + 6x + 3y \, ,$$

where x and y are free parameters.

Under the conditions (3-4-2) and (3-4-3) we find that

$$x = 7/40, \quad y = 17/20.$$

Thus the required weights are

$$a_0 = 21/40$$

$$a_1 = 21/20$$

$$a_2 = 33/40$$

$$a_3 = 3/5 .$$

We have thus produced a stable method. We again decided to take the equation

$$\int_0^x (1+x-t)\, y(t)\, dt = -1 + x + e^{-x},$$

with solution $y(x) = xe^{-x}$, and apply the method of this section to it. An exact starting value was used, and the errors obtained are shown in table 2.

TABLE 2 : THE MODIFIED TRAPEZOIDAL METHOD

VALUE OF X	ERROR
0.1	0.63×10^{-2}
0.2	-0.20×10^{-2}
0.3	0.10×10^{-1}
0.4	-0.12×10^{-1}
0.5	0.13×10^{-1}
0.6	-0.91×10^{-2}
0.7	0.86×10^{-2}
0.8	-0.61×10^{-2}
0.9	-0.53×10^{-2}
1.0	-0.39×10^{-2}

The results continued in a stable fashion for all the values calculated; (up to x=5). We can compare these results with those shown in Figure 1, where the trapezoidal method was used on the same equation in a direct manner.

(iii) A convergent, stable, third order method. Holyhead combines the three-eighth's rule with two ancillary rules, and by employing the technique used in (ii), obtains a convergent and stable third order method. The matrix $\underline{\Psi}_h$ is shown in figure 7.

FIGURE 7 : THE THREE-EIGHTH'S RULE

$$\underline{\Psi}_h = h \begin{bmatrix}
1 & & & & & & & & & \\
0 & 1 & & & & & & & & \\
0 & 0 & 1 & & & & & & & \\
3/8 & 9/8 & 9/8 & 3/8 & & & & & & \\
\alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & & & & & \\
\beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & & & & \\
3/8 & 9/8 & 9/8 & 6/8 & 9/8 & 9/8 & 3/8 & & & \\
3/8 & 9/8 & 9/8 & 3/8+\alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & & \\
3/8 & 9/8 & 9/8 & 3/8+\beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & \\
3/8 & 9/8 & 9/8 & 6/8 & 9/8 & 9/8 & 6/8 & 9/8 & 9/8 & 3/8 \\
. & . & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . \quad .
\end{bmatrix}$$

Here $v=7$, $\mu=3$, $\rho=6$.

In this case, the eigenvalues of G satisfy

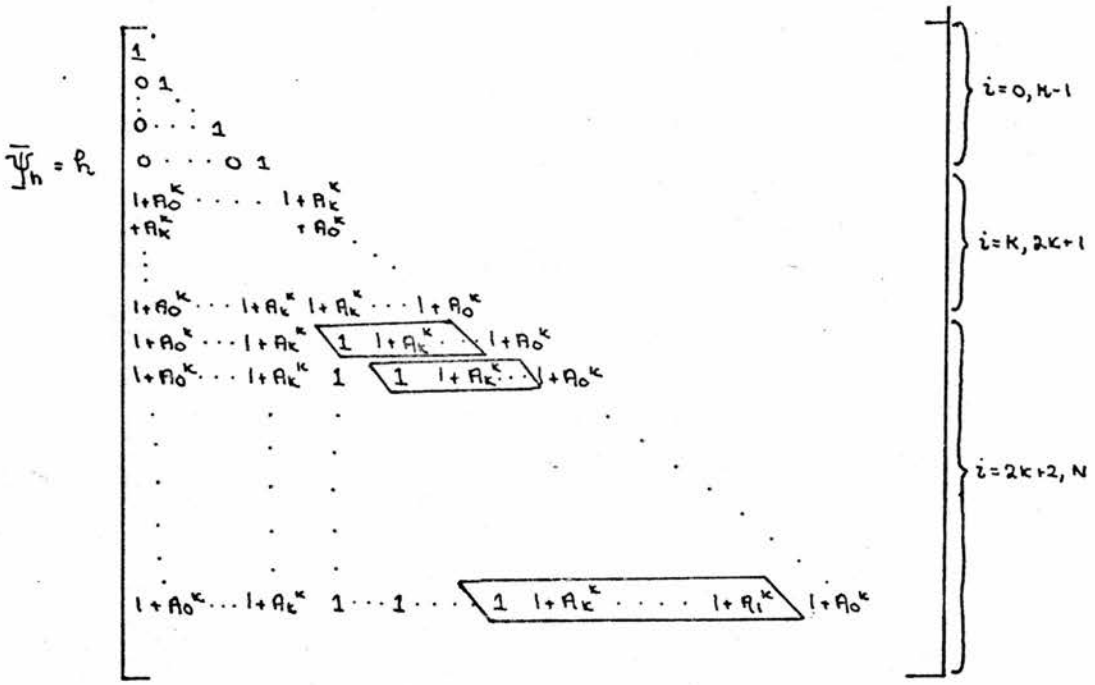
$$\lambda^2(\lambda^4 + a_1\lambda^3 + a_2\lambda^2 + a_3\lambda + a_4) = 0$$

where the a_i, $i=1,\ldots,4$ are dependent on α_i, β_i the weights in the ancillary rules. Again we find $a_4=0$ for any choice of α_i, β_i. Choosing the two rules to be of accuracy $O(h^4)$, (the same as the three-eighth's rule in composite form), we have three free parameters among

the weights α_i, β_i . These can be chosen so that the remaining three eigenvalues are zero. Three non-linear equations in the three parameters are then obtained, which Holyhead solves and obtains explicit values for the weights α_i, β_i . For full details and a numerical example, consult Holyhead.

(iv) Gregory's Rule. The use of Gregory's rule for the integral equation gives a method with a repetition factor of one. The matrix $\bar{\Psi}_h$ is shown in figure 8, for the cases of k>o.

FIGURE 8 : GREGORY'S RULE



For this method we find $\nu = 2k+2$, $\mu = 1$, $\rho = k+1$.

The matrix $G = G_1$ is thus $(k+1) \times (k+1)$. [Notice that when k=o, (trapezoidal method), the matrix G collapses.] The first element in the last row of G is $A_K^K / 1 + A_0^K$ and the i^{th} element in the last row of G is $-(A_{K-i+2}^K - A_{K-i+1}^K) / 1 + A_0^K$, $2 \le i \le K+1$.

The eigenvalues of G are solutions of

$$\lambda^{k+1} + \lambda^k \frac{(A_1^k - A_0^k)}{1+A_0^k} + \cdots + \lambda \frac{(A_k^k - A_{k-1}^k)}{1+A_0^k} - \frac{A_k^k}{1+A_0^k} = 0$$

That is,

$$\lambda^{k+1}(1+A_0^k) + \lambda^k (A_1^k - A_0^k) + \cdots + \lambda (A_k^k - A_{k-1}^k) - A_k^k = 0$$

The writer now studies the question of the roots of

$$p(\lambda) = \lambda^{k+1}(1+A_0^k) + \lambda^k (A_1^k - A_0^k) + \cdots + \lambda (A_k^k - A_{k-1}^k) - A_k^k$$

for the cases of k even and k odd separately.

Case 1 : k even

Since $|A_0^k| < 1$, we find $p(\lambda) \to -\infty$ as $\lambda \to -\infty$

Now $p(-1) = -1 + 2\{ -A_0^k + A_1^k + \cdots - A_k^k \}$

$= -1 + 2\{ |A_0^k| + |A_1^k| + \cdots + |A_k^k| \}$,

since the A_i^k oscillate in sign with $A_0^k < 0$

Thus $p(-1) = -1 + 2B_k$

where $$B_k = \sum_{j=0}^{k} |A_j^k|$$

$$= |a_0| + 2|a_1| + \cdots + 2^k |a_k| \qquad \text{[See Phillips][20]}$$

$$> 1/2,$$

since $|a_0| = 1/2$

Thus, $p(-1) > 0$, and we deduce that $p(-1)$ has a root in $(-\infty, -1)$.

Case 2 : k odd

Here $p(\lambda) \to \infty$ as $\lambda \to -\infty$

We find $p(-1) = 1 + 2\{ A_0^k - A_1^k + \cdots + A_k^k \}$

$= 1 - 2\{ |A_0^k| + |A_1^k| + \cdots + |A_k^k| \}$

$= 1 - 2B_k$

$< 0.$

We deduce that $\rho(\lambda)$ has a root in $(-\infty, -1)$.

We can conclude, therefore, that the Gregory rule, with k>o, will produce divergent methods.

3 - 5 Concluding Remarks

We have seen that the convergence of the multistep methods depends on the size of the eigenvalues of a certain matrix G. We have shown that when we apply Newton-Cotes formulae to give methods with a repetition factor of one, we obtain diverging results. However, we have seen that when we apply Newton-Cotes formulae to give methods with a repetition factor greater than one, we can, by skillful choice of weights in the ancillary rules, obtain methods which are convergent and stable. We have shown the disappointing result that the Gregory rule, with k>o, gives divergent methods.

In section (3-4) we examined means of obtaining a third order convergent method. Clearly, higher order methods are possible, although the analysis required to determine the values of the unknown weights in the ancillary rules would become increasingly complex.

SECTION 4: A CLASS OF BLOCK-BY-BLOCK METHODS

4 - 1 Introduction

In Section 3 we saw that many 'standard' quadrature methods, when applied to the Volterra equation of the first kind, were found to be non-convergent. Special methods had to be devised in order to produce convergence. We might, therefore, be justified in being cautious in applying block-by-block methods to the equation. Indeed some numerical experiments with block methods at first indicated that the methods were non-convergent (see Linz [15]).

However, recent work, due to Weiss [25] , shows that some block methods are convergent, provided we take a sufficiently small step-size. Weiss also shows that we may need to be careful about the stability of particular schemes. The methods outlined in this section can be considered as generalisations of the block methods discussed in Chapter 2, Section 2.

4 - 2 The Block Methods

We consider the grid of points x_i, $i=0,1,2,\ldots,N$ such that

$$x_i = x_0 + ih$$

where $x_0 = a$, $x_N = b$, $h = (b-a)/N$ together with the sub-grid

$$x_{ij} = x_i + u_j h \quad , \quad j = 1,\ldots,n$$

where the u_j satisfy $0 \le u_1 < u_2 < \ldots < u_n = 1$.

We notice that if $u_1 = 0$, then $x_{i1} = x_i$, and that $x_{in} = x_{i+1}$. The aim now is to obtain block methods over each set of points $\{x_{ir},\ldots,x_{in}\}$, $i=0,\ldots,n$, obtaining the block of approximate values $\{Y_{ir},\ldots, Y_{ir}\}$, where

$$r = \begin{cases} 1 & , u_1 \neq 0 \\ 2 & , u_1 = 0 \end{cases}$$

This avoids the blocks overlapping.

Thus, in the integral equation

$$\int_a^x K(x,t)\,y(t)\,dt = g(x) \quad , \qquad a \le t \le x \le b,$$

we set $x = x_{ij}$ and split the interval of integration (x_o, x_{ij}) into (x_o, x_i) and (x_i, x_{ij}). We can then obtain

$$g(x_{ij}) = \sum_{\ell=0}^{i-1} \int_{x_\ell}^{x_{\ell+1}} K(x_{ij},t)\,y(t)\,dt + \int_{x_i}^{x_{ij}} K(x_{ij},t)\,y(t)\,dt \qquad \ldots(4\text{-}2\text{-}1)$$

Now consider the Lagrange coefficient function

$$h_k(x) = \frac{\Pi(x)}{(x-u_k)\,\Pi'(u_k)}$$

where

$$\Pi(x) = (x-u_1)\ldots(x-u_n)$$

If, in (4-2-1), we replace $K(x,t)y(t)$ by its interpolating polynomial

$$\sum_{k=1}^{n} h_k\left(\frac{t-x_\ell}{h}\right) K(x, x_{\ell k})\, y(x_{\ell k}) \qquad \ell = 0, \ldots, i-1,$$

and perform the subsequent integration, we obtain the scheme

$$g(x_{ij}) = h \sum_{\ell=0}^{i-1} \sum_{k=1}^{n} a_k\, K(x_{ij}, x_{\ell k})\, Y_{\ell k} + h \sum_{k=1}^{n} a_{jk}\, K(x_{ij}, x_{ik})\, Y_{ik} \qquad \ldots(4\text{-}2\text{-}2)$$

for $j = r, \ldots, n$, $i = 0, \ldots, N$, and where

$$a_k = \int_0^1 h_k(s)\,ds \quad , \quad a_{jk} = \int_0^{u_j} h_k(s)\,ds \quad , \quad j = 1, \ldots, n.$$

For each value of i (4-2-2) gives us a system of linear equations to be solved for the values $\{Y_{ir}, \ldots, Y_{in}\}$.

Choice of $u_1, \ldots, u_{n-1}$

If we choose equally spaced points, that is

$$u_i = \frac{i-1}{n-1} \qquad , \qquad i = 1, \ldots, n-1$$

then we obtain Newton-Cotes rules and interpolatory rules in the scheme (4-2-2). (See Chapter 2, Section 2) With this choice of u_i, the block method (4-2-2) is equivalent to the first block method considered in Section 2 - 3, Chapter 2, except that here the blocks are size h, instead of size (n-1)h.

The choice of symmetrically spaced points can lead to Lobatto quadrature rules in (4-2-2). For example the choice

$$u_1 = 0, \; u_2 = 5 - \sqrt{5}/10, \; u_3 = 5 + \sqrt{5}/10, \; u_4 = 1,$$

gives us a four point Lobatto quadrature rule as the main rule in (4-2-2). We notice that the restriction $u_n = 1$ prevents us from obtaining Radau and Chebyshev rules.

In scheme (4-2-2) we find that values of the kernel K(x,t) are required outside $a \le t \le x \le b$. For certain kernels this may produce difficulties. A different approximation to the second integral in (4-2-1) avoids this. (See Section 2 - 4, Chapter 2)

Suppose that in the integral

$$\int_{x_i}^{x_{ij}} K(x_{ij}, t)\, y(t)\, dt \qquad \ldots (4\text{-}2\text{-}3)$$

we replace y(t) by its interpolating polynomial

$$\sum_{k=1}^{n} \frac{L_k(t - x_i)}{h} \, y(x_{ik}) \qquad \text{where } L_k(t) \text{ is as before.}$$

If we first write $t = x_i + u_j h s$ in (4-2-3) we find

$$\int_{x_i}^{x_{ij}} K(x_{ij}, t)\, y(t)\, dt = \int_0^1 h u_j \, K(x_{ij}, x_i + u_j h s)\, y(x_i + u_j h s)\, ds$$

$$\int_{x_i}^{x_{ij}} K(x_{ij},z)\,y(t)\,dt \;\doteq\; \sum_{r=1}^{n}\int_0^1 h u_j\, K(x_{ij}, x_i + u_j h s)\, l_r(u_j s)\, y(x_{ir})\,ds$$

$$\doteq \sum_{r=1}^{n}\sum_{k=1}^{n} h u_j a_k\, K(x_{ij}, x_i + u_j u_k h)\, l_r(u_j u_k)\, y(x_{ir})$$

on applying the quadrature rule used to form the first summation in (4-2-2).

We thus obtain the modified scheme

$$g(x_{ij}) = h\sum_{i=0}^{i-1}\sum_{k=1}^{n} a_k\, K(x_{ij},x_{ik})\,\gamma_{ik} + h\sum_{r=1}^{n}\sum_{k=1}^{n} u_j a_k\, K(x_{ij},x_i+u_j u_k h)\cdot l_r(u_j u_k)\,\gamma_{ir} \qquad \ldots(4\text{-}2\text{-}4)$$

$j=1,\ldots,n;\ i=0,\ldots,N$

This scheme is a generalisation of the schemes considered in Chapter 2, section 2 - 4.

4 - 3 Convergence theorems

In this section we find that we need to consider the two cases $u_1 \neq 0$, $u_1 = 0$ separately. The reason for this is soon apparent.

Case 1 : $u_1 \neq 0$

Let us first examine scheme (4-2-2). We find that the errors

$\varepsilon_{ij} = y(x_{ij}) - \gamma_{ij}$ satisfy

$$h\sum_{k=1}^{n} a_{jk}\, K(x_{ij}, x_{ik})\,\varepsilon_{ik} + h\sum_{i=0}^{i-1}\sum_{k=1}^{n} a_k\, K(x_{ij},x_{ik})\,\varepsilon_{ik} + R_{ij,h} = 0, \qquad \ldots(4\text{-}3\text{-}1)$$

$j=1,\ldots,n,$

where $R_{ij,h}$ denotes the truncation error of the quadrature rules. If we subtract equation (4-3-1), with i replaced by i-1 and j=n, from equation (4-3-1), and divide the resulting equation by $hK(x_i,x_i)$, we obtain

$$\sum_{k=1}^{n} a_{jk}\,\frac{K(x_{ij},x_{ik})}{K(x_i,x_i)}\,\varepsilon_{ik} = -\sum_{i=0}^{i-1}\sum_{k=1}^{n} a_k\left(\frac{K(x_{ij},x_{ik}) - K(x_i,x_{ik})}{K(x_i,x_i)}\right)\varepsilon_{ik}$$

$$-\frac{(R_{ij,h} - R_{i-1,n,h})}{hK(x_i,x_i)}, \qquad \ldots(4\text{-}3\text{-}2)$$

$j=1,\ldots,n.$

Now, for sufficiently differentiable K(x,t) and y(t) we can find a constant C>o such that

$$\frac{R_{ij,h} - R_{i-1,n,h}}{hK(x_i,x_i)} \leq Ch^n$$

Taking moduli in equation (4-3-2) we can define constants M, K, a, >o, such that

$$\left| \sum_{k=1}^{n} a_{jk} \frac{K(x_{ij}, x_{ik})}{K(x_i,x_i)} \varepsilon_{ik} \right| \leq \frac{hKa}{M} \sum_{i=0}^{i-1} \sum_{k=1}^{n} |\varepsilon_{ik}| + Ch^n \qquad \ldots.(4\text{-}3\text{-}3)$$

If we write $e_i = \|\underset{\sim}{\varepsilon_i}\|$ where $\underset{\sim}{\varepsilon_i} = \{\varepsilon_{i1}, \ldots, \varepsilon_{in}\}^T$

and denote by $\bar{A}^{(i)}$ the matrix with elements $\left\{ a_{jk} \frac{K(x_{ij},x_{ik})}{K(x_i,x_i)} \right\}$, $j,k=1,\ldots,n$.

then if $L = \max_{1 \leq i \leq N} \| (\bar{A}^{(i)} - I) \|$

we find, after some manipulation with (4-3-3) that

$$e_i \leq \frac{hKan}{M} \sum_{i=0}^{i-1} e_i + Ch^n.$$

With $e_0 \leq C_2 h^n$, for some $C_2 > o$, we find, using Lemma 1, Chapter 1, that

$e_i \to 0$ as $h \to 0$ and $i \to \infty$ with ih fixed.

We deduce that the scheme (4-2-2) with $u_1 \neq o$ is convergent of order n. Turning to the modified scheme (4-2-4), we find that we can obtain an equation analogous to (4-3-2) if we make use of the identity

(See Weiss [25], page 14)

$$u_j \sum_{k=1}^{n} a_k L_r(u_j, u_k) = a_{jr}, \qquad \ldots.(4\text{-}3\text{-}4)$$

on setting j=n.

Convergence of the scheme (4-2-4) then follows.

Case 2 : $u_1 = o$

Again let us examine scheme (4-2-2) first. We find that equation (4-3-1) becomes

$$h \sum_{k=2}^{n} a_{jk} K(x_{ij}, x_{ik}) \varepsilon_{ik} + h \sum_{i=0}^{i-1} \sum_{k=1}^{n} a_k K(x_{ij}, x_{ik}) \varepsilon_{ik}$$

$$+ h a_{j1} K(x_{ij}, x_{i1}) \varepsilon_{i1} + R_{ij,h}. \qquad \ldots(4\text{-}3\text{-}5)$$

$j=2,\ldots,n.$

We see we now have an 'extra' term in ε_{i1} on the left-hand side, and the previous analysis is no longer valid. Weiss shows that convergence can be obtained, but that it is subject to certain restrictions on the choice of u_i $(i=2,\ldots,n-1)$. The proof of the convergence result is complex and lengthy. Therefore we merely state the result in the following theorem.

Theorem 6

The schemes (4-2-2) and (4-2-4) with $u_1 = o$ are convergent if and only if $-1 \le \eta \le 1$, where

$$\eta = \prod_{k=1}^{n-1} (1-u_k) \Big/ \prod_{k=2}^{n} (-u_k),$$

and the order of convergence is n-1 if $\eta = 1$, and n otherwise.

4 - 4 Stability results

We now examine the schemes (4-2-2) and (4-2-4) with $K(x,t)=1$. The two schemes are, in fact, identical. This is a consequence of the identity (4-3-4). Thus setting $K(x,t)=1$ in (4-2-2) or (4-2-4), replacing i by i-1 and setting j=n, subtracting this equation from

the original (4-2-2) or (4-2-4) gives us

$$g(x_{ij}) - g(x_{i1}) = \sum_{k=1}^{n} h a_{jk} Y_{ik} \qquad j = 1, \ldots, n \qquad \ldots (4\text{-}4\text{-}1)$$

We again need to consider the two cases $u_1 \neq 0$, $u_1 = 0$ separately.

Case 1 : $u_1 \neq 0$ $(r=1)$

Here (4-4-1) gives us the system of linear equations

$$A \underset{\sim}{Y} = \underset{\sim}{g},$$

where

$$\underset{\sim}{Y} = \{ Y_{i1}, \ldots, Y_{in} \}^T$$

$$\underset{\sim}{g} = 1/h \{ g(x_{i1}) - g(x_{i1}), \ldots, g(x_{in}) - g(x_{i1}) \}^T$$

and A is the matrix with elements $\{a_{ij}\}$ $i,j=1,\ldots,n$.

We see that each Y_{ij} $(j=1,\ldots,n)$ can be obtained in terms of $g(x_i)$, $g(x_{i1}),\ldots, g(x_{in})$ and we can interpret this as a 'differentiation rule', for Y_{ij}. We deduce that the scheme is stable in the sense of Noble.

Case 2 : $u_1 = 0$ $(r=2)$

The situation here is quite different. Y_{i1} is now known from the previous block of values, and (4-4-1) gives a system of linear equations for $\{Y_{i2}, \ldots, Y_{in}\}^T$. That is

$$g(x_{ij}) - g(x_{i1}) - h a_{j1} Y_{i1} = \sum_{k=2}^{n} h a_{jk} Y_{ik} \qquad , \quad j = 2, \ldots, n.$$

We now have the term in Y_{i1} on the left hand side, and the scheme cannot be interpreted in the same way as above. We therefore do not expect the scheme to be stable.

In some lengthy analysis, Weiss obtains asymptotic expansions for the errors in the schemes (4-2-2) and (4-2-4). He finds that for the cases $u_1>0$ and $u_1=0$, $|\eta|<1$, the leading terms in the asymptotic expansions are characterised by stable growth (in accordance with Definition 1). However, for the cases $u_1=0$, $|\eta|\geq 1$, he finds that the leading terms are now of the desired form, and the methods are weakly unstable. As Weiss's expansions appear to be of limited practical use, they are not quoted here.

4 - 5 Applications

Here we consider schemes with symmetrically positioned u_i, $i=1,\ldots,n$. This class of schemes includes methods based on Newton-Cotes rules and Lobatto rules. We find

$$\eta = \frac{\prod_{k=2}^{n-1}(1-u_k)}{(-1)\prod_{k=2}^{n-1}(-u_k)}$$

$$= \frac{\prod_{k=2}^{n-1}(u_k)}{(-1)^{n-1}\prod_{k=2}^{n-1}(u_k)}$$

Thus

$$\eta = (-1)^{1-n}$$

We deduce that the schemes are convergent, but are weakly unstable. From the stability results of Section 4 - 4, we see that the most satisfactory schemes are those with $u_1 \neq 0$. This class of schemes will not yield familiar quadrature rules; however, schemes can be easily constructed.

A numerical example

We consider the equation

$$1+x-\sin x-\cos x=\int_0^x (1+x-t)\,y(t)\,dt \qquad 0\le x\le 1.2$$

with solution $y(x)=\sin x$. If we set $u_1=0$, $u_2=\frac{1}{2}$, $u_3=1$, we find Simpson's rule as the first quadrature in scheme (4-2-2). Choosing a step-size $h=0.3$, we obtain approximations to $y(x)$ at $x=0.15, 0.30,\ldots, 1.2$, and choosing $h=0.15$ we obtain approximations at $x=0.075, 0.15,\ldots, 1.2$. Weiss computes these values, using schemes (4-2-2), and some results, showing convergence, are given below.

x	Error with h=0.3	Errors with h=0.15
0.3	-3.6×10^{-4}	-9.2×10^{-5}
0.6	-2.3×10^{-4}	-6.1×10^{-5}
0.9	5.5×10^{-4}	1.3×10^{-4}
1.2	2.1×10^{-3}	5.2×10^{-4}

Table 3 : BLOCK-BY-BLOCK METHOD [TABLE TAKEN FROM WEISS [25]]

Section 5 : EXPANSION METHODS

5 - 1 Introduction

The purpose of this section is to investigate the use of expansion methods for solving Volterra equations of the first kind. The methods have been tried rather unsuccessfully on Fredholm equations of the first kind; the lack of success being due to the extreme ill-conditioning of the system of equations resulting from the procedure, especially when this system is large.

We would expect that the system of equations resulting from the application of an expansion method to a Volterra equation of the first kind is likely to be ill-conditioned. Before examining the methods, therefore, we need to have available an algorithm specifically devised for solving ill-conditioned systems. Wilkinson [26] has devised such an algorithm; the method being essentially Crout's with partial pivoting and inner-product accumulation. for full details, consult Wilkinson [26]. With this algorithm in mind, we move on to examine some expansion methods.

5 - 2 A collocation method

Given the equation

$$g(x) = \int_a^x K(x,t)\, y(t)\, dt \quad , \qquad a \le x \le b,$$

we consider approximating to y(t) by $\tilde{y}(t)$ where

$$\tilde{y}(t) = \sum_{j=1}^{N} a_{j-1}\, \psi_{j-1}(t) \qquad \ldots.(5\text{-}2\text{-}1)$$

In (5-2-1) the functions $\psi_{j-1}(t), j=1,\ldots,N$ must be prescribed, and the constants $a_{j-1}, j=1,\ldots,N$ are to be determined. We therefore obtain

$$g(x)=\sum_{j=1}^{N} a_{j-1}\int_a^x K(x,t)\,\psi_{j-1}(t)\,dt. \qquad \ldots.(5\text{-}2\text{-}2)$$

If we now set $x=x_r$, $r=1,\ldots,N$, with $a\leq x_r\leq b$, we obtain from (5-2-2) a system of equations for the N unknowns a_{j-1}, $j=1,\ldots,N$. That is

$$\underset{\sim}{g} = K\underset{\sim}{a} \quad , \qquad \ldots.(5\text{-}2\text{-}3)$$

where

$$\underset{\sim}{g} = \{g(x_1),\ldots,g(x_N)\}^T \quad ,$$

$$\underset{\sim}{a} = \{a_0,\ldots,a_{N-1}\}^T \quad ,$$

and K is the NxN matrix with elements

$$\int_a^{x_r} K(x_r,t)\,\psi_{j-1}(t)\,dt, \qquad r,j=1,\ldots,N. \qquad \ldots.(5\text{-}2\text{-}4)$$

The points x_r, $r=1,\ldots,N$ are known as collocation points. In general we will not be able to determine the integrals (5-2-4) exactly. However, if we choose the collocation points to be equally spaced, such that

$$x_j = a+jh \quad , \quad j=1,\ldots,N \quad ,$$

with $x_N = b$, we can approximate to the integrals by any standard quadrature. Now the introducture of quadrature means that the elements of the matrix K will be slightly perturbed from their true values. If the matrix K is ill-conditioned we clearly need to make these perturbations as small as possible to avoid the possibility of large changes in $\underset{\sim}{a}$ arising. The use of a high-order quadrature is thus desired.

Choice of approximating functions This type of method is obviously most suitable if we have some prior knowledge of the behaviour of y(x). If a polynomial approximation is appropriate, then a suitable choice of $\tilde{y}(x)$ is

$$\tilde{y}(x) = \sum_{r=1}^{N} a_{r-1} T_{r-1}(x),$$

where the $T_{r-1}(x)$ are the Chebyshev polynomials $T_{r-1}(x) = \cos((r-1)\cos^{-1}x)$ for this choice, we would need to transform the range of y(x) from $[a,b]$ to $[-1,1]$. This is easily done by means of a linear transformation. The choice

$$\hat{y}(x) = \sum_{r=1}^{N} a_{r-1} x^{r-1}$$

is not suitable as it is known to lead to ill-conditioning problems.

A numerical example We took the equation

$$\int_0^x (1+x-t)\, y(t)\, dt = -1 + x + e^{-x}$$

with solution $y(x) = xe^{-x}$

on the range $0 \leq x \leq 1$. For this equation it was convenient to use modified Chebyshev polynomials $T_r^*(x)$ defined on the range $[0,1]$. That is

$$T_r^*(x) = \cos(r \cos^{-1}(2x-1)), \qquad r=0,\ldots,N-1.$$

The collocation method was performed taking N=5. The method was first performed with (5-2-4) being evaluated exactly, and again using Gregory quadrature to evaluate (5-2-4). Collocation points were chosen as $x_r = (0.2)Xr$, $r=1,\ldots,5$, and the Gregory rule was used with step-size h=0.1 and with $k=2$. (See Chapter 1, Section 1-3). The accuracy of the Gregory rule is then $O(h^4)$. Errors are compared in Table 4 below. (The system (5-2-3) was solved using Wilkinson's algorithm[25]).

TABLE 4 : COLLOCATION METHOD

VALUE OF X	ERRORS IN 'EXACT' METHOD	ERRORS IN 'QUADRATURE' METHOD
0.2	$0.29X10^{-4}$	$0.47X10^{-5}$
0.4	$-0.15X10^{-4}$	$-0.38X10^{-4}$
0.6	$0.14X10^{-4}$	$0.90X10^{-6}$
0.8	$-0.29X10^{-4}$	$-0.40X10^{-4}$
1.0	$0.15X10^{-3}$	$0.15X10^{-3}$
	Condition number of matrix = 53.4	Condition number of matrix = 53.5

'EXACT' METHOD 'QUADRATURE' METHOD

Approximate expansion: $\tilde{y}(x)=0.15\times10^{-3}+x-0.98x^2$, $\tilde{y}(x)=0.13\times10^{-3}+x-0.98x^2$
$+0.44x^3-0.09x^4$ $+0.44x^3-0.09x^4$

True expansion: $y(x)=x-x^2+0.5x^3-0.16x^4$.

It can be seen that both methods produced good results with the quadrature method in fact producing the better results. This result is, of course, due to the fact that the errors in the Gregory rule have resulted in a 'perturbed' expansion, which has turned out to be closer to the true expansion. This will not always be the case.

In general, of course, we will not be able to determine the magnitude of the errors, and the optimum size of N is not known. If we increase N we obtain a larger system of equations to solve and the risk of ill-conditioning grows. This is a drawback as we would like to choose N large so that the collocation points cover the region $[a,b]$ as densely as possible. In the next section we suggest a much improved method which also enables us to obtain actual estimates of the errors incurred.

5 - 3 Error minimising techniques

A well-known method for solving integral equations of the second kind is the residual minimising technique. In this method we first replace $y(x)$ by a series of expansion (as in 5-2), form the residual, $R(x)$ say, and then minimise this residual in a particular way (for example, in a least squares sense). The residual may be minimised over a discrete number of points spanning the region of interest, with the number of these points exceeding the number of terms in the series expansion. If the equation is linear we then obtain a linear system of equations to solve for the unknown coefficients in the series expansion.

For the Volterra equation of the first kind, we obtain the residual

$$R(x) = g(x) - \int_a^x K(x,t)\,\tilde{y}(t)\,dt \qquad , \quad a \le x \le b,$$

$$= g(x) - \sum_{j=1}^{N} a_{j-1} \int_a^x K(x,t)\,\psi_{j-1}(t)\,dt.$$

If we write

$$e(x) = y(x) - \tilde{y}(x),$$

then

$$R(x) = \int_a^x K(x,t)\,e(t)\,dt.$$

However, if we minimise R(x), we will not be able to determine how e(x) is being minimised; if at all. We cannot obtain any bounds for the magnitude of e(x) in terms of R(x), unlike the situation for the equation of the second kind (See Chapter 2, Section 5-1). We therefore look at ways of minimising the error directly. This idea is due to Watson[24], who applied it to integral equations of the second kind. From (5-2-2) we can obtain

$$g(x) - \sum_{i=1}^{N} a_{i-1} \int_a^x K(x,t)\,\psi_{i-1}(t)\,dt = \int_a^x K(x,t)\,e(t)\,dt.$$

We now choose equally spaced points x_j , j=1,...,M with

$$x_j = x_0 + jh \quad , \; x_0 = a, \; x_M = b,$$

and where M≫N. Then we have

$$\int_a^{x_j} K(x_j,t)\,e(t)\,dt = g(x_j) - \sum_{i=1}^{N} a_{i-1} \int_a^{x_j} K(x_j,t)\,\psi_{i-1}(t)\,dt, \qquad j=1,\ldots,M. \qquad \ldots.(5\text{-}3\text{-}1)$$

Now (5-3-1) is an integral equation of the first kind for e(t). Replacing the integral on the left-hand side of (5-3-1) by a convergent quadrature for the equation gives us approximations ε_i to $e(t_1)$, i=1,...,N with

$$\lim_{h\to 0} \varepsilon_i = e(t_i).$$

However, we are limited in the choice of quadrature available. From Section 2 we know that the Euler, midpoint and trapezoidal methods produce convergent methods. In Section 3-4 we managed to obtain a higher order convergent method, but this was complex in nature. We therefore proceed to replace the integral containing e(t) in (5-3-1) by the midpoint rule. We observe that the error of the rule is $O(h^2)$ but this low order of accuracy could be off-set by increasing M; thus making h smaller. This will also mean that the x_j, j=1,...,M cover [a,b] more densely. Equation (5-3-1) thus becomes

$$h\sum_{i=0}^{j-1} K(x_j, x_{i+1/2})\,\varepsilon_{i+1/2} = g(x_j) - \sum_{i=1}^{N} a_{i-1}\,\overline{\psi}_i(x_j), \qquad j=1,\ldots,M, \qquad \ldots.(5\text{-}3\text{-}2)$$

where $$\overline{\psi}_i(x_j) = \int_a^{x_j} K(x_j,t)\,\psi_{i-1}(t)\,dt, \qquad \ldots.(5\text{-}3\text{-}3)$$

and we have written $\varepsilon_{i+1/2}$ to denote $\varepsilon_i + h/2$.

From (5-3-2) we obtain the system of equations

$$\underset{\sim}{\varepsilon} = K^{-1}\underset{\sim}{g} - K^{-1}\Phi\,\underset{\sim}{a} \qquad \ldots.(5\text{-}3\text{-}4)$$

where

$$\underset{\sim}{\varepsilon} = \{\varepsilon_{1/2}, \ldots, \varepsilon_{N-1/2}\}^T,$$

$$\underset{\sim}{a} = \{a_0, \ldots, a_{N-1}\}^T,$$

$$\underset{\sim}{g} = \{g(x_1), \ldots, g(x_N)\}^T,$$

and where K is an MxM lower triangular matrix with elements $\{K_{jl}\}$ such that

$$K_{jl} = K(x_j, x_{l+1/2}).$$

Φ is an MxN matrix with elements $\phi_{ij} = \Psi_j(x_i)$.

One method of procedure now is to find, from (5-3-4)

$$\min_{\underset{\sim}{a}} \|\underset{\sim}{\varepsilon}\|,$$

where

$$\|\underset{\sim}{\varepsilon}\| = \max_{1\le i\le M} |\varepsilon_{i-1/2}|,$$

which is a linear programming problem. However, a discrete least squares method can be applied more easily as follows.

We rewrite the system (5-3-4) explicitly as

$$\varepsilon_i^* = \sum_{j=1}^{M} K_{ij}^* g_j - \sum_{j=1}^{N} a_{j-1}\,\phi_{ij}^* \quad , i = 1, \ldots, M \qquad \ldots.(5\text{-}3\text{-}5)$$

where $\varepsilon_i^* = \varepsilon_{i-1/2}$, $i = 1, \ldots, M$,

$g_j = g(x_j)$. $j = 1, \ldots, M$,

and K_{ij}^* are the elements of the matrix K^{-1} , $i,j = 1,\ldots,M$

ϕ_{ij}^* are the elements of the matrix $K^{-1}\Phi$, $i=1,\ldots,M$, $j=1,\ldots,N$

The discrete least squares method requires that we set

$$\partial/\partial a_{j-1} \sum_{i=1}^{M} \varepsilon_i^{*2} = 0 \qquad j=1,\ldots,N.$$

From (5-3-5) we therefore obtain

$$\sum_{i=1}^{M} \phi_{ik}^* \left\{ \sum_{j=1}^{N} K_{ij}^* q_j - \sum_{j=1}^{N} \phi_{ij}^* a_{j-1} \right\} = 0 \qquad k=1,\ldots,N$$

That is,

$$\sum_{i=1}^{M} \phi_{ik}^* \sum_{j=1}^{N} K_{ij}^* q_j = \sum_{j=1}^{N} a_{j-1} \sum_{i=1}^{M} \phi_{ik}^* \phi_{ij}^* \qquad \ldots.(5\text{-}3\text{-}6)$$

$k=1,\ldots,N$

We therefore have a system of linear equations to solve for the a_{j-1}, $j=1,\ldots,N$. The system (5-3-6) may at first look complex to set up, but we are reminded that the matrix K is lower triangular, and thus the elements K_{ij}^*, $i,j=1,\ldots,N$, are easy to obtain. Once the a_{j-1}, $j=1,\ldots,N$ are obtained we can substitute the values in (5-3-5) to obtain the $\varepsilon_{i-\frac{1}{2}}$, $i=1,\ldots,N$ the midpoint rule approximations to the errors

$$e(x_{i-\frac{1}{2}}) = y(x_{i-\frac{1}{2}}) - \tilde{y}(x_{i-\frac{1}{2}}) \quad , \qquad i=1,\ldots,N.$$

Thus, in addition to being able to utilise M N points for the minimisation of errors, we see that we are able to obtain direct estimates of the errors, within the bounds of the truncation error of the midpoint rule. We notice that the integrals in (5-3-3) may not always be available exactly. However, they may be determined to an arbitrarily high degree of accuracy by using quadrature.

A Numerical example

Again we chose the equation

$$\int_0^x (1+x-t)\,y(t)\,dt = 1+x+e^{-x},$$

to be solved for $y(x)$ in $0 \le x \le 1$.

We took the modified Chebyshev polynomials $T_{r-1}^{*}(t)$, $r=1,\ldots,N$, with $N=5$, as the approximating functions. We also took $M=50$ with the step-size $h=0.02$. The results at the first few x_i, $i=1,\ldots,10$ are shown in Table 5.

VALUE OF X	ERROR IN COMPUTED RESULT	PREDICTED ERROR (FROM (5-3-5))	RESIDUAL
0.01	-0.72×10^{-4}	-0.72×10^{-4}	-0.85×10^{-6}
0.03	-0.27×10^{-4}	-0.28×10^{-4}	-0.18×10^{-5}
0.05	0.34×10^{-5}	0.29×10^{-5}	-0.21×10^{-5}
0.07	0.23×10^{-4}	0.23×10^{-4}	-0.19×10^{-5}
0.09	0.35×10^{-4}	0.34×10^{-4}	-0.13×10^{-5}
0.11	0.39×10^{-4}	0.39×10^{-4}	-0.56×10^{-6}
0.13	0.38×10^{-4}	0.37×10^{-4}	0.21×10^{-6}
0.15	0.33×10^{-4}	0.33×10^{-4}	0.94×10^{-6}
0.17	0.26×10^{-4}	0.26×10^{-4}	0.16×10^{-5}
0.19	0.17×10^{-4}	0.17×10^{-4}	0.20×10^{-5}
0.21	0.75×10^{-5}	0.67×10^{-5}	0.23×10^{-5}
0.23	-0.19×10^{-5}	-0.20×10^{-5}	0.24×10^{-5}

Condition number of matrix of coefficients = 8.2

True expansion $y(x) = x-x^2+0.5x^3-0.16x^4+\ldots$

Approximate expansion $\tilde{y}(x)=0.99\times10^{-4}+x-0.98x^2+0.44x^3-0.09x^4$

TABLE 5 : THE LEAST SQUARES METHOD

The good results shown in Table 5 continued throughout all the values of $\tilde{y}(x)$ calculated. For interest, values of the residual were also computed, and the magnitude of the residual was normally less than the magnitude of the error. The approximate expansion is seen to fit fairly well with the first few terms of the true expansion. To the degree of accuracy given here, however, it is the same as that obtained by the collocation method of Section 5-2.

The condition number of the matrix of coefficients is surprisingly low. This may not, of course, always be the case.

5 - 4 Discussion

We have briefly examined the application of some expansion methods to the Volterra equation of the first kind. In Section 5-2 we examined a collocation method and noted its limitations. In Section 5-3 we examined a least-squares method which enabled us to obtain actual estimates for the errors incurred at a discrete number of points.

In both of these methods we need to be careful of the validity of the expansion obtained for $y(x)$. For example, in Section 5-3 we obtained numerical solutions to a particular integral equation in the interval $[0,1]$. Now this integral equation has solution

$$y(x) = xe^{-x},$$

for all x, with its maximum value occuring at x=1. We cannot expect the expansion we obtained for $y(x)$ to yield very accurate values outside this interval. The situation here will be common to most Volterra integral equations except, for example, when the solution is periodic or linear. The method is perhaps better used only for obtaining discrete values of $y(x)$ within $[0,1]$.

Another difficulty common to all methods of this type is the

choice of N, the degree of the approximating polynomial. Increasing N may not necessarily guarantee better results, as the system of linear equations becomes larger and more ill-conditioned. We may find that the errors at discrete points remain small, but that the exapansion itself is totally invalid.

The methods should therefore be used with care; their best application being when some prior knowledge of $y(x)$ is available (to determine the appropriate expansion) and then taking only a small number of terms in this expansion.

REFERENCES

1. Anselone P.H. and Gonzalez-Fernandez J.M. Uniformly convergent solutions of Fredholm Integral equations, J. Math Anal. and Appl. 10(1965), 519-536.

2. Barrodale I and Young A., Computational experience in solving linear operator equations using the Chebyshev norm, Numerical approximations to functions and data, edited by J.G. Hayes, The Athlone Press (1970), 115-142.

3. Bel'tyukov B.A., An analog of the Runge-Kutta method for solution of non-linear Volterra-type integral equations, Differential Equations, Vol. 1, (1966), 417-426.

4. Campbell G.M. and Day J.T.,The Numerical solution of non-linear Volterra integral equations, BIT 10 (1970), 10-19.

5. Campbell G.M. and Day J.T., A block-by-block method for the numerical solution of Volterra integral equations, BIT 11 (1971), 120-124.

6. Cochran J.A., Analysis of linear integral equations, McGraw-Hill (1972).

7. Day J.T., A starting method for solving non-linear Volterra integral equations, Math. Comp. 21 (1967), 179-187.

8. Espinosa-Maldonado R. and Byrne G.D., Solution of linear integral equations by Gregory's method, BIT 10 (1970), 457-464.

9. Henrici P., Discrete variable methods in ordinary differential equations, Wiley (1962).

10. Henrici P., E'ements of Numerical analysis, Wiley (1964).

11. Holyhead, P.A.W. and McKee, S., and Taylor, P.J., Multistep methods for solving linear Volterra integral equations of the first kind, to be published in SIAM J. Numer. Anal.

12. Jaim M.K. and Sharma K.D., Numerical solutions of linear differential equations and Volterra integral equations using Lobatto quadrature formula, Comp. J. 10 (1967), 101-107.

13. Kobayasi M., On numerical solution of the Volterra integral equations of the first kind by trapezoidal rule, Rep. Stat. Appl. Res. JUSE 14(2) (1967), 65-78.

14. Lambert J.D., Computational methods in ordinary differential equations, Wiley (1973).

15. Linz P., The Numerical solution of Volterra integral equations by finite difference methods, MRC Technical Summary Report no. 825, (1967).

16. Noble B., The numerical solution of non-linear integral equations and related topics, Non-linear integral equations edited by P.H. Anselone, University of Wisconsin Press (1964), 215-318.

17. O'Neill J.C. and Byrne G.D., A starting method for the numerical solution of Volterra's integral equation of the second kind, BIT 8 (1968), 43-47.

18. Osborne M.R. and Watson G.A., An algorithm for minimax approximation in the non-linear case, Comp. J. 12 (1969), 63-68.

19. Phillips D.L., A technique for the numerical solution of certain integral equations of the first kind, J. Assoc. Comp. Mach.9 (1962), 84-96.

20. Phillips G.M., An error estimate for Volterra integral equations, BIT 11 (1971), 181-186.

21. Phillips G.M. and Taylor, P.J., Theory and Applications of Numerical Analysis, Academic Press (1973).

22. Pouzet P., Methode d'integration numerique des equations integrales et integro-differentielles du type Volterra de second espece, formules de Runge-Kutta, in Symposium on the numerical treatment of Integral and Integro-differential equations, Birkhauser (1960), 362-368.

23. Tricomi F.G., Integral equations, Interscience (1957).

24. Watson G.A., On estimating best approximations of functions defined by integral equations, Comp. J. 16 (1973), 77-79.

25. Weiss R., Numerical procedures for Volterra integral equations, Ph.d. Thesis, Canberra (1972).

26. Wilkinson J., The solution of ill-conditioned equations, Mathematical methods for Digital Computers Vol.2, edited by A. Ralston and H.S. Wilf, Wiley (1965).

27. Young A., The application of product integration to the numerical solution of integral equations, Proc. Roy. Soc. A224 (1954), 561-573.