THE UNIVERSITY OF BRITISH COLUMBIA

FACULTY OF GRADUATE STUDIES

PROGRAMME OF THE

FINAL ORAL EXAMINATION

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

of

RIAZ AHMAD USMANI

B.Sc. The Aligarh Muslim University, India, 1956

M.Sc. The Aligarh Muslim University, India, 1957

THURSDAY, April 13, 1967 at 3:30 p.m.

In Room 102 Mathematics Building

COMMITTEE IN CHARGE

Chairman: I. McT. Cowan

C. Clark
A. H. Cayford
E. Leimanis

C. Froese
J. Brenner
M. MacMillan

External Examiner: Dr. Ralph G. Stanton
Department of Mathematics
University of Manitoba.

Research Supervisor: C. Froese.
NUMERICAL SOLUTION OF BOUNDARY VALUE PROBLEMS
FOR ORDINARY DIFFERENTIAL EQUATIONS

Abstract

In the numerical solution of the two point boundary value problem,
\[ y'' = f(x,y), \quad y(a) = y_a, y(b) = y_b, -\infty < a < b < \infty, \]
the usual method is to approximate the problem by a finite difference analogue of the form
\[ \sum_{i=0}^{k} a_i y_{n+i} = h^2 \sum_{i=0}^{k} \beta_i y''_{n+i}, \quad y_0 = y_a, \quad y_{N+1} = y_b \]
\[ n = 0,1, \ldots, N-1 \quad \ldots (2) \]
with \( k = 2 \), and the truncation error \( T.E. = O(h^4) \) or \( O(h^6) \), where \( h \) is the step-size. Varga (1962) has obtained error bounds for the former when the problem \((1)\) is linear and of class \( M \).

In this thesis, more accurate finite difference methods are considered. These can be obtained in essentially two different ways, either by increasing the value \( k \) in difference equations \((2)\), or by introducing higher order derivatives. Several methods of both types have been derived. Also, it is shown how the initial value problem \( y' = \phi(x,y) \) can be formulated as a two point boundary value problem and solved using the latter approach.
Error bounds have been derived for all of these methods for linear problems of class M. In particular, more accurate bounds have been derived than those obtained by Varga (1962) and Aziz and Hubbard (1964). Some error estimates are suggested for the case where $\frac{df}{dy} < 0$, but these are not accurate bounds, especially when $\frac{df}{dy}$ is not a constant.

In the case of non-linear differential equations, sufficient conditions are derived for the convergence of the solution of the system of equations (2) by a generalized Newton's method.

Some numerical results are included and the observed errors compared with theoretical error bounds.

GRADUATE STUDIES

Real Variable Theory                  S. Nash
Point Set Topology                    S. Cleveland
Theory of Functions                   R. Cleveland
Functional Analysis                   C. W. Clark
Celestial Mechanics                   E. Leimanis
Numerical Analysis                   C. Froese.
PUBLICATIONS


NUMERICAL SOLUTION OF BOUNDARY VALUE PROBLEMS IN ORDINARY DIFFERENTIAL EQUATIONS

by

Riaz Ahmad Usmani

B.Sc.(Hons.), 1956; M.Sc., 1957,

The Aligarh Muslim University, ALIGARH, India.

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Mathematics

We accept this thesis as conforming to the required standard:

The University of British Columbia

March, 1967
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Mathematics
The University of British Columbia
Vancouver 8, Canada
Date April 13, 1967
ABSTRACT

In the numerical solution of the two point boundary value problem,

\[ y'' = f(x, y), \quad y(a) = y_a, \quad y(b) = y_b, \quad -\infty < a < b < \infty, \quad \ldots \] (1)

the usual method is to approximate the problem by a finite difference analogue of the form

\[ \sum_{i=0}^{k} a_i y_{n+i} = h^2 \sum_{i=0}^{k} \beta_i y''_{n+i}, \quad y_0 = y_a, \quad y_{N+1} = y_b, \quad n = 0, 1, \ldots, N-1 \quad \ldots \] (2)

with \( k = 2 \), and the truncation error \( T.E. = O(h^4) \) or \( O(h^6) \), where \( h \) is the step-size. Varga (1962) has obtained error bounds for the former when the problem (1) is linear and of class M, i.e. \( f(x, y) \) is continuous, bounded and \( f_y > 0 \), but his method can be extended immediately to include the latter case as well.

In this thesis, more accurate finite difference methods are considered. These can be obtained in essentially two different ways, either by increasing the value of \( k \) in difference equations (\( \Delta E \)) (2), or by introducing higher order derivatives. Methods with \( k = 4 \) are considered as a one parameter family. When \( k > 2 \), additional \( \Delta E \)'s are required for points near the boundaries. Such \( \Delta E \)'s for \( k = 4 \) have been derived so that the overall resulting error in these methods turns out to be \( O(h^6) \). These methods are shown to be more accurate than a similar one proposed by Bramble and Hubbard (1964). Difference methods involving higher order derivatives of the type

\[ \sum_{i=0}^{k} a_i y_{n+i} = h^2 \sum_{i=0}^{k} \beta_i y''_{n+i} + h^3 \sum_{i=0}^{k} \gamma_i y'''_{n+i} + \ldots + h^t \sum_{i=0}^{k} \theta_i y^{(t)}_{n+i} \quad \ldots \] (3)
with \( k = 2, t = 3 \) and 4 are considered when (1) is linear. Now additional equations are required to eliminate \( y' \) from the expressions for the higher derivatives obtained from (1) through over differentiation. Such equations have been determined so that the overall method has T.E. \( = 0(h^8) \) and \( 0(h^{10}) \) for \( t = 3 \) and 4, respectively. Also, it is shown how the initial value problem \( y' = \phi(x,y) \) can be formulated as a two point boundary value problem and solved using these equations.

Error bounds have been derived for all of these methods for linear problems of class \( M \). In particular, more accurate bounds have been derived than those obtained by Varga (1962) and Aziz and Hubbard (1964). Some error estimates are suggested for the case where \( \frac{\partial f}{\partial y} < 0 \), but these are not accurate bounds, especially when \( \frac{\partial f}{\partial y} \) is not a constant.

In the case of non-linear differential equations, sufficient conditions are derived for the convergence of the solution of the system of equations (2) by a generalized Newton's method.

Some numerical results are included and the observed errors compared with theoretical error bounds. The ratio of the maximum absolute error and the theoretical bound was usually greater than 0.1, for the methods based on (3). Various methods are compared by treating the error as a function of cost, assuming the major cost is either (i) in the evaluations of the functions involved or (ii) in the solution of the system of equations. Under both assumptions, the method based on (3) with \( t = 4 \) gave the lowest error for a given cost over a wide range. It also compared favourably with well-known predictor-corrector methods for initial value problems.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II BOUNDARY VALUE PROBLEM OF CLASS M</td>
<td>9</td>
</tr>
<tr>
<td>III DIFFERENCE EQUATIONS OF HIGHER ORDERS</td>
<td>29</td>
</tr>
<tr>
<td>IV A CLASS OF HIGH ACCURACY DIFFERENCE FORMULAS OF LOW ORDERS</td>
<td>49</td>
</tr>
<tr>
<td>V LINEAR BOUNDARY VALUE PROBLEMS WHERE $\frac{\partial f(x,y)}{\partial y} &lt; 0$</td>
<td>61</td>
</tr>
<tr>
<td>VI NON-LINEAR DIFFERENTIAL EQUATIONS</td>
<td>69</td>
</tr>
<tr>
<td>VII BOUNDARY VALUE TECHNIQUES AS APPLIED TO INITIAL VALUE PROBLEMS</td>
<td>79</td>
</tr>
<tr>
<td>VIII EXPERIMENTAL RESULTS AND CONCLUDING REMARKS</td>
<td>90</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>117</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>27</td>
</tr>
<tr>
<td>II</td>
<td>34</td>
</tr>
<tr>
<td>III</td>
<td>35</td>
</tr>
<tr>
<td>IV</td>
<td>68</td>
</tr>
<tr>
<td>V</td>
<td>82</td>
</tr>
<tr>
<td>VI</td>
<td>91</td>
</tr>
<tr>
<td>VII</td>
<td>92</td>
</tr>
<tr>
<td>VIII</td>
<td>103</td>
</tr>
<tr>
<td>IX</td>
<td>104</td>
</tr>
<tr>
<td>X</td>
<td>105</td>
</tr>
<tr>
<td>XI</td>
<td>106</td>
</tr>
<tr>
<td>XII</td>
<td>107</td>
</tr>
<tr>
<td>XIII</td>
<td>110</td>
</tr>
<tr>
<td>XIV</td>
<td>111</td>
</tr>
<tr>
<td>XV</td>
<td>113</td>
</tr>
<tr>
<td>XVI</td>
<td>116</td>
</tr>
</tbody>
</table>

- **TABLE I**: EXPERIMENTS WITH PROBLEM (2.18)
- **TABLE II**: DISTRIBUTION OF ROOTS OF $G(s) = 0$ for $2 < s < 3$
- **TABLE III**: TABULATION OF THE FUNCTION $-10^3 G(s)$
- **TABLE IV**: EXPERIMENTS WITH PROBLEM (5.10)
- **TABLE V**: COEFFICIENTS OF THE Δ.E. (7.6)
- **TABLE VI**: OPTIMUM CHOICE OF $c$ IN $M-3(c)$
- **TABLE VII**: EXPERIMENTS WITH PROBLEM (A) WITH $h = 1/8$
- **TABLE VIII**: EXPERIMENTS WITH PROBLEM (D) WITH $h = 1/8$
- **TABLE IX**: EXPERIMENTS WITH PROBLEM (D) WITH $h = 1/2$ USING $M-5$ AND $M-6$
- **TABLE X**: EXPERIMENTS WITH PROBLEM (B)
- **TABLE XI**: EXPERIMENTS WITH PROBLEM (E)
- **TABLE XII**: EXPERIMENTS WITH PROBLEM (8.2)
- **TABLE XIII**: EXPERIMENTS WITH NON-LINEAR PROBLEM (8.3), $h = 0.1$, $\mu = 3/2$
- **TABLE XIV**: EXPERIMENTS WITH PROBLEM (8.4) USING $\hat{M}-1$, $\hat{M}-4$ and $\hat{M}-5$
- **TABLE XV**: EXPERIMENTS WITH PROBLEM (8.4) USING $M^*-1$ AND $M^*-5$
- **TABLE XVI**: COMPARISON OF $\hat{M}-1$, $\hat{M}-4$ AND $\hat{M}-5$ WITH STANDARD INITIAL VALUE TECHNIQUES
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>DESCRIPTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\log_{10}</td>
<td>\text{Error}</td>
</tr>
<tr>
<td>2</td>
<td>$\log_{10}</td>
<td>\text{Error}</td>
</tr>
<tr>
<td>3</td>
<td>$\log_{10}</td>
<td>\text{Error}</td>
</tr>
<tr>
<td>4</td>
<td>$\log_{10}</td>
<td>\text{Error}</td>
</tr>
<tr>
<td>5</td>
<td>$\log_{10}</td>
<td>\text{Error}</td>
</tr>
<tr>
<td>6</td>
<td>$\log_{10}</td>
<td>\text{Error}</td>
</tr>
<tr>
<td>7</td>
<td>$\log_{10}</td>
<td>\text{Error}</td>
</tr>
<tr>
<td>8</td>
<td>$\log_{10}</td>
<td>\text{Error}</td>
</tr>
<tr>
<td>9</td>
<td>$\log_{10}</td>
<td>\text{Error}</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENT

The author wishes to thank his advisor, Professor Charlotte Froese for guidance and helpful suggestions given during the preparation of this thesis and Professor A.H. Cayford for reading and commenting on the manuscript in the final form.

The author also expresses his appreciation to the National Research Council of Canada for the financial assistance through its summer grants and the University of British Columbia Computing Centre for making available to him the facilities for the numerical experiments on their IBM 7040 computer.
To my wife
CHAPTER I
INTRODUCTION

1.1 GENERAL CONSIDERATIONS

The general solution of an nth order differential equation (D.E.)
\[ y^{(n)}(x) = F(x, y(x), y'(x), \ldots, y^{(n-1)}(x)), \]
for a real function \( y(x) \), normally depends on \( n \) parameters. These parameters are determined in an initial value problem by prescribing the values
\[ y^{(m)}(x) = A_m \quad (m = 0, 1, \ldots, n-1) \]
at a fixed point \( x = a \). If the conditions are based on more than one point, then the problem is called a boundary value problem. The boundary conditions usually have the form
\[ V_i [y(x), y'(x), \ldots, y^{(n-1)}(x)] = 0 \quad (i = 0, 1, \ldots, n-1) \]
at the boundary points \( x = a \) and \( x = b \) which may include \( \pm \infty \). Here \( y^{(m)}(x) \) denotes the value of the mth derivative of \( y(x) \) at the point \( x \). The functions \( F \) and \( V_i \) are in general nonlinear. In the present thesis, we shall deal with the special class of boundary value problems for which the D.E. is of second order (linear or nonlinear).

1.2 BOUNDARY VALUE PROBLEM OF CLASS M

A boundary value problem is said to be of class \( M \), see [12], Chapter 7, if it is of the form
\[ y''(x) = f(x, y), \quad y(a) = y_a, \quad y(b) = y_b \quad \ldots(1.1) \]
where
(i) \( -\infty < a < b < \infty \)
(ii) \( y_a \) and \( y_b \) are arbitrary constants
(iii) \( f(x, y) \) is a continuous function of two variables with \( \frac{\partial f(x, y)}{\partial y} \)
continuous, bounded and non-negative in the strip $S$ defined by

$$a \leq x \leq b \quad \text{and} \quad -\infty < y < \infty.$$  

Although in what follows, we will also consider boundary value problems in which the above condition viz. $\frac{\partial f(x,y)}{\partial y} \geq 0$, will be violated in $S$, we will concentrate mostly on problems of class $M$. The existence of a unique solution of the boundary value problem of class $M$ is assured (Henrici [12, pp. 347]).

1.3 NUMERICAL SOLUTION OF BOUNDARY VALUE PROBLEMS

Numerically there are two methods for solving a boundary value problem. The "shooting" technique is that of solving the initial value problem

$$y'' = f(x,y), \quad y(a) = y_a, \quad y'(a) = c$$

using some appropriate method. A correction procedure is then used to choose new values of $y'(a)$ such that $y(b)$ converges to the desired value $y_b$. This technique will not be considered in this thesis, even though there are problems where it is a feasible procedure.

The second method is the numerical approximation of the D.E. (1.1) at a finite set of grid points,

$$x_n = a + nh \quad (n = 0,1,..., N+1)$$

where

$$x_0 = a, \quad x_{N+1} = b \quad \text{and} \quad h = (b-a)/(N+1).$$

Here $N$ is an appropriate positive integer. A scheme is then designed for the determination of the numbers $y_n$, which it is hoped, will closely approximate the values $y(x_n)$ of the true solution of (1.1). A convenient way to obtain such a scheme is to demand that the $y_n$ satisfy a difference equation ($\Delta \text{E.}$) of the form
\[
\sum_{i=0}^{k} \alpha_i y_{n+i} = h^2 \sum_{i=0}^{k} \beta_i y''_{n+i} \quad (n = 0, 1, \ldots, N-k+1) . \quad \ldots (1.2)
\]

Here \( y'' = f(x, y) \), \( \alpha_k \neq 0 \) and \( |\alpha_0| + |\beta_0| \neq 0 \).

We normalize (1.2) by choosing \( \alpha_k = 1 \). The positive integer \( k \) is called the order of the A.E. This method always leads to a system of \( N-k+2 \) simultaneous equations involving \( y_1, y_2, \ldots, y_N \), possibly in a nonlinear fashion, \( y_0 \) and \( y_{N+1} \) being determined by the boundary conditions. For \( k > 2 \), additional equations are required.

The purpose of this thesis is to formulate more accurate procedures than the known numerical techniques for the solution of the boundary value problems and to obtain accurate error bounds to assist one in the selection of the step size \( h \). It is aimed to show how such error bounds may be obtained for certain finite difference analogues approximating a class of problems of the form (1.1).

In the case of linear boundary value problems, this purpose has been achieved in two ways:

(a) By increasing the order \( k \) of the A.E. from \( k = 2 \) to \( k = 4 \). In Chapter II we approximate the linear D.E. by a well known A.E. of order \( k = 2 \) of the form (1.2) for which the resulting error is \( O(h^4) \). But in Chapter III, a general one parameter family of A.E.'s of order \( k = 4 \) is developed (most of the known A.E.'s of order \( k = 4 \) become special cases of this A.E.) and it is proved that the resulting error is \( O(h^6) \).

(b) By using A.E.'s involving higher derivatives. The second important aspect is to modify A.E. (1.2) to the following form

\[
\sum_{i=0}^{k} \alpha_i y_{n+i} = h^2 \sum_{i=0}^{k} \beta_i y''_{n+i} + h^3 \sum_{i=0}^{k} \gamma_i y'''_{n+i} + \ldots + h^t \sum_{i=0}^{k} \theta_i y^{(t)}_{n+i}, \quad t > 3 ,
\]

\ldots (1.3)
and to use it in approximating the linear D.E. (1.1). These ideas are expanded in Chapter IV in connection with linear two point boundary value problems of class M.

The linear boundary problems not of class M are dealt with in Chapter V using the same ideas. In Chapter VI, we consider nonlinear boundary value problems of class M. The iterative procedure used is merely a generalization of Newton's method for solving a system of nonlinear algebraic equations.

Finally in Chapter VII, we show that the boundary value techniques studied in Chapters II and IV can be suitably applied to solve a class of initial value problems of the form
\[ y' = f(x) y + g(x), \quad y(a) = y_a . \]
The last chapter is devoted to numerical results and conclusions.

1.4 Before we start a systematic discussion of linear boundary value problems in the next chapter, we would like to introduce some terms employed frequently in the thesis.

We assume that the true solution \( y(x) \) of the problem (1.1) satisfies a Δ.E. of the form
\[
\sum_{i=0}^{k} \alpha_i y(x_{n+i}) = h^2 \sum_{i=0}^{k} \beta_i y''(x_{n+i}) + T_{n+k}, \quad \ldots (1.4)
\]
where \( T_{n+k} \) is the truncation error (T.E.). Here \( y''(x_i) = f(x_i, y(x_i)) \), and if we introduce the discretization error
\[ e_n = y(x_n) - y_n , \]
we obtain the error equation by subtracting (1.2) from (1.4) namely
\[
\sum_{i=0}^{k} \alpha_i e_{n+i} = h^2 \sum_{i=0}^{k} \beta_i \sigma_{n+i} e_{n+i} + T_{n+k}, \quad \ldots (1.5)
\]
where $\sigma_i e_i = f(x_i, y(x_i)) - f(x_i, v_i)$, so that $\sigma_i$ will be some value of $\frac{\partial f(x,y)}{\partial y}$ by "mean value theorem" in two variables.

In actual practice, in any machine calculations, equation (1.2) will not be solvable exactly because of round-off and other errors which may be introduced in the evaluation of $f(x,y)$. The round-off error depends primarily on the number of digits being carried out during computation but partly on the way in which the calculation is organized. However in this thesis, we want to concentrate on the accuracy of the difference approximation itself and therefore all such errors will be neglected.

**Difference Operator**

We associate with the A.E. (1.2) the operator

$$L[y(x); h] = 1 + \sum_{i=0}^{k} \frac{\alpha_i}{4} y(\alpha_i + ih) - h^2 \sum_{i=0}^{k} \frac{\beta_i}{2} y''(\alpha_i + ih)$$  \hspace{1cm} (1.6)

This operator acts on any function $y(x)$ which has a second order derivative, but we assume that $y(x)$ has continuous derivatives of sufficiently high orders. We may then expand (1.6) using Taylor's formula in powers of $h$ and obtain, on collecting terms containing like powers of $h$,

$$L[y(x); h] = C_0 y(x) + C_1 h y'(x) + C_2 h^2 y''(x) + \ldots,$$  \hspace{1cm} (1.7)

where

$$C_0 = \sum_{i=0}^{k} \alpha_i$$

$$C_1 = \sum_{i=0}^{k} i \alpha_i$$

$$C_2 = \frac{1}{2} \sum_{i=0}^{k} i^2 \alpha_i - \sum_{i=0}^{k} \frac{\beta_i}{2}$$  \hspace{1cm} (1.8)

and

$$C_q = \frac{1}{q} \sum_{i=0}^{k} i^q \alpha_i - \frac{1}{(q - 2)!} \sum_{i=0}^{k} i^{q-2} \beta_i \quad (q=3,4,\ldots).$$
Now we define the degree of the Δ.E. as the unique integer \( p \) such that
\[
C_q = 0 \ (q=0,1, \ldots, p+1); \quad C_{p+2} \neq 0.
\]
Then by definition we have
\[
L[y(x); h] = C_{p+2} h^{p+2} y^{(p+2)}(x) + O(h^{p+3}). \quad \ldots\ldots(1.9)
\]
We also note [using (1.3)] that the T.E. is given by
\[
T_{n+k} = C_{p+2} h^{p+2} y^{(p+2)}(x_n) + O(h^{p+3}). \quad \ldots\ldots(1.10)
\]
However, for a number of difference operators, it turns out that we can write
\[
L[y(x); h] = C_{p+2} h^{p+2} y^{(p+2)}(\xi) \quad \ldots\ldots(1.11)
\]
where \( \xi \) is a suitable number in the interval \((x, x+kh)\). We shall refer to (1.11) as the generalized mean value theorem. This theorem does not hold for all operators, but every operator of degree \( p \) can be represented in the form
\[
L[y(x); h] = h^{p+2} \int_0^k G(s) y^{(p+2)}(x + hs) \, ds, \quad \ldots\ldots(1.12)
\]
by using Taylor's formula with the integral representation of the remainder in the expansion of (1.6). The kernel \( G(s) \) is given as follows:
\[
G(s) = \alpha_k \frac{(k-s)^{p+1}}{(p+1)!} - \beta_k \frac{(k-s)^{p-1}}{(p-1)!}, \quad s \in [k-1, k]
\]
and for \( s \in [\mu, \mu + 1] \) (\( \mu = 0, 1, \ldots, k-2 \))
\[
G(s) = \frac{1}{(p+1)!} \left[ \alpha_k (k-s)^{p+1} + \alpha_{k-1} (k-1-s)^{p+1} + \ldots + \alpha_{\mu+1} (\mu+1-s)^{p+1} \right]

- \frac{1}{(p-1)!} \left[ \beta_k (k-s)^{p-1} + \beta_{k-1} (k-1-s)^{p-1} + \ldots + \beta_{\mu+1} (\mu+1-s)^{p-1} \right].
\]
An operator \( L[y(x); h] \) is called definite if the kernel \( G(s) \) does not change sign in \([0,k]\). For a definite operator with \( y^{(p+2)}(x) \) continuous, the generalized mean value theorem holds since by the "second law of mean" (1.12) can be written as (1.11). However, if \( G(s) \) does change sign, we still have
\[ |L [y(x); h]| < h^{p+2} \sum_{k} G_{p+2} \]

where
\[ G = \int_{0}^{k} |G(s)| \, ds , \]

and
\[ M_{p+2} = \max_{x<\xi<x+kh} |y^{(p+2)}(\xi)|. \]

### Associated Polynomials

Sometimes it is convenient to associate with the Δ.E. (1.2), the polynomials

\[ a(x) = \sum_{i=0}^{k} a_i x^i \quad \text{with} \quad a_k = 1. \]

\[ \beta(x) = \sum_{i=0}^{k} \beta_i x^i \.
\]

It should be noted that if the Δ.E. (1.2) has a positive degree \( p \geq 1 \), then we have

\[ C_0 = C_1 = C_2 = 0 . \]

In terms of the polynomials \( a(x) \) and \( \beta(x) \), the above conditions may be expressed in the form

\[
\begin{align*}
\alpha(1) & = 0 \\
\alpha'(1) & = 0 \\
\alpha''(1) & = 2 \beta(1)
\end{align*}
\]

The first two conditions of (1.16) imply that \( x = 1 \) is a root of \( a(x) = 0 \) of at least multiplicity 2. Thus we can write \( a(x) \) in the form

\[ a(x) = (x - 1)^2 \gamma(x) \]

where \( \gamma(x) \) is a monic polynomial of degree \( k-2 \).

### Vector Norm

Let \( v \) be a vector with components \( v_1, v_2, \ldots, v_n \). The vector norm \(||v||| \) employed is defined by
\[ |v| = \max_i |v_i| \] \hspace{1cm} \ldots (1.18)

This also induces a norm on the matrices as follows:

\[ |A| = \max \quad |Av| \quad \text{with} \quad |v| = 1 , \]

where \( A = (a_{ij}) \). It turns out that the matrix norm \(|A|\) subordinate to the vector norm defined by (1.18) is given by

\[ |A| = \max \sum_{i}^{n} |a_{ij}| . \] \hspace{1cm} \ldots (1.19)

For a detailed discussion of the norms on vectors and matrices, the reader is referred to Faddeeva [7].
CHAPTER II

BOUNDARY VALUE PROBLEMS OF CLASS M

2.1

We consider the boundary value problem of class M viz.

\[ y'' = f(x,y), \quad y(a) = y_a, \quad y(b) = y_b \]  

...(2.1)

with \( \frac{\partial f(x,y)}{\partial y} \geq 0 \) and replace it by a Δ.E. of the form (1.2). The coefficients \( a \)'s and \( \beta \)'s are usually chosen in such a way that the Δ.E. (1.2) turns out to be of as high degree as possible. The truncation error of the Δ.E. is thus

\[ \text{T.E.} = C_{p+2} h^{p+2} y^{(p+2)} + O(h^{p+3}), \quad C_{p+2} \neq 0. \]  

...(2.2)

The resulting system of algebraic equations for \( y_n \) is implicit in any case; however another difficulty may arise. As in any system of equations, we need as many equations for the determination of the unknowns as there are unknowns. Since \( y_0 \) and \( y_{N+1} \) are determined by the boundary conditions, the unknowns in our case will be \( y_1, y_2, \ldots, y_N \). If the order \( k \) of the Δ.E. is greater than 2, the new unknowns such as \( y_{-1} \) or \( y_{N+2} \) are introduced for which there are no equations. This difficulty is overcome by suitably modifying the Δ.E. near the boundaries; however it does not arise at all if \( k=2 \), the smallest possible value of \( k \). In view of relations (1.16) if \( k=2 \) and the Δ.E. has a positive degree \( p \), then it is necessarily of the form

\[ y_n - 2 y_{n+1} + y_{n+2} = h^2 (\beta_0 y_n'' + \beta_1 y_{n+1}'' + \beta_2 y_{n+2}'') \]  

...(2.3)

with \( \beta_0 + \beta_1 + \beta_2 = 1 \).

Now we shall apply representation (1.12) in order to prove the generalized mean value theorem (1.11) for the operator associated with the
Δ.E. (2.3). Expanding around \( x \), using Taylor's formula with integral remainder, we get

\[
L[y(x); h] = y(x) - 2y(x+h) + y(x+2h) - h^2 \left[ \beta_0 y''(x) + \beta_1 y''(x+h) + \beta_2 y''(x+2h) \right]
\]

\[
= h^{p+2} \int_0 G(s) y^{(p+2)}(x+hs) \, ds
\]

where

\[
G(s) = \begin{cases} 
\frac{(2-s)^{p+1}}{(p+1)!} - \beta_2 \frac{(2-s)^{p-1}}{(p-1)!}, & 1 \leq s \leq 2 \\
\frac{(2-s)^{p+1}}{(p+1)!} - 2 \frac{(1-s)^{p+1}}{(p+1)!} - \beta_2 \frac{(2-s)^{p-1}}{(p-1)!} - \beta_1 \frac{(1-s)^{p-1}}{(p-1)!}, & 0 \leq s \leq 1.
\end{cases}
\]

For \( p=1 \), \( \beta_0 + \beta_1 + \beta_2 = 1 \),

\[
G(s) = \begin{cases} 
\frac{1}{2} ((2-s)^2 - 2 \beta_2), & 1 \leq s \leq 2 \\
\frac{1}{2} (2 \beta_0 - s^2), & 0 \leq s \leq 1.
\end{cases}
\]

It is easily shown that \( G(s) \) changes sign if

(i) \( 0 < \beta_2 < \frac{1}{2} \) or if \( 0 < \beta_0 < \frac{1}{2} \)

(ii) both \( \beta_0, \beta_2 \geq \frac{1}{2} \)

(iii) \( \beta_0, \beta_2 \leq 0 \),

and \( G(s) \) is of constant sign if \( \beta_2 \geq \frac{1}{2}, \beta_0 \leq 0 \) or \( \beta_0 \geq \frac{1}{2}, \beta_2 \leq 0 \).

For \( p=2 \), we have by (1.8)

\[
\beta_0 + \beta_1 + \beta_2 = 1, \beta_1 + 2\beta_2 = 1.
\]

The above conditions together imply that \( \beta_0 = \beta_2 \) so that

\[
G(s) = \begin{cases} 
\frac{1}{3} (2-s)^3 - \beta_2 (2-s), & 1 \leq s \leq 2 \\
G(2-s), & 0 \leq s \leq 1
\end{cases}
\]

i.e. the function \( G(s) \) is symmetric about the line \( s=1 \).
Also
\[
G(s) = \begin{cases} 
0, & \beta_2 = 0 \\
< 0, & \beta_2 \geq \frac{1}{6} 
\end{cases}
\]

For p=3, by (1.8), we have
\[
\beta_0 + \beta_1 + \beta_2 = 1, \quad \beta_1 + 2\beta_2 = 1, \quad \beta_1 + 4\beta_2 = \frac{7}{6}.
\]
The solution of this system is given by
\[
\beta_0 = \beta_2 = \frac{1}{12}, \quad \beta_1 = \frac{10}{12},
\]
and the corresponding coefficient \(C_5\) vanishes. Thus there is no operator of degree \(p=3\) associated with the A.E. (2.3). However \(C_6 = -1/240\), hence the operator associated with the A.E. (2.3) with \(\beta_0=\beta_2=1/12, \beta_1=10/12\) turns out to be of degree \(p=4\) and the corresponding kernel \(G(s)<0\) in \([0,2]\). The question of \(p\) being greater than 4 does not arise at all, since all the \(\beta's\) are determined already with \(p=4\).

Thus the operator associated with the A.E. (2.3) is definite, provided \(p=1, 2\) or 4. In case \(p=1\), either \(\beta_2 \geq 1/2, \beta_0 \leq 0\) or \(\beta_0 \geq 1/2, \beta_2 \leq 0\) and if \(p=2\), then either \(\beta_2 = 0\) or if \(\beta_2 
eq 0\), then \(\beta_2 \geq 1/6\).

2.2 A.E. OF ORDER 2

If we make use of A.E. (2.3) in approximating the D.E. (2.1) setting \(n=0,1,\ldots, N-1\) in (2.3), we get the following system of equations
\[
Jy + h^2Df(y) = d,
\]
where
\[
J = \begin{bmatrix}
2 & -1 \\
-1 & 2 & -1 \\
\vdots & \ddots & \ddots \\
& \ddots & \ddots & \ddots \\
& & -1 & 2 & -1 \\
& & & -1 & 2
\end{bmatrix},
\]
and
\[
D = \begin{bmatrix}
\beta_1 & \beta_2 \\
\beta_0 & \beta_1 & \beta_2 \\
\beta_0 & \beta_1 & \beta_2
\end{bmatrix},
\]

\[
\begin{bmatrix}
\beta_1 & \beta_2 \\
\beta_0 & \beta_1 & \beta_2 \\
\beta_0 & \beta_1
\end{bmatrix}
\]

Here we are concerned with the case where the given D.E. (2.1) is linear i.e. $f(x,y)$ is of the form

$$f(x,y) = g(x) y + s(x), \quad g(x) \geq 0 \text{ in } S.$$  

Defining a new diagonal matrix $G$

$$G = \begin{bmatrix}
    g_1 & & \\
    & g_2 & \\
    & & g_3 \\
    & & & \ddots \\
    & & & & g_N
\end{bmatrix}$$

and the vector $s = \begin{bmatrix}
    s_1 \\
    s_2 \\
    s_3 \\
    \vdots \\
    s_N
\end{bmatrix}$

where $g_i = g(x_i)$ and $s_i = s(x_i)$, we may write the vector $f(y)$ introduced in (2.4) as

$$f(y) = G y + s.$$  

The algebraic system (2.4) now reduces to the system of linear equations

$$A y = b \quad \ldots(2.5)$$

where $A = J + h^2 D G$ and $b = d - h^2 D s$.

The numerical solution of the system (2.5) is relatively easy owing to the fact that the matrix $A$ shares with the matrices $J$ and $D$ the property of having non-zero elements only on the main diagonal and the two diagonals adjacent to it. Matrices of this kind are called tridiagonal. In fact, if
\[ A = (a_{ij}), \text{ we have by virtue of (2.5)} \]
\[ a_{n,n} = 2 + h^2 \beta_2 g_n \quad (n = 1, 2, \ldots, N) \]
\[ a_{n,n-1} = -1 + h^2 \beta_0 g_{n-1} \quad (n = 2, 3, \ldots, N) \]
\[ a_{n,n+1} = -1 + h^2 \beta_2 g_{n+1} \quad (n = 1, 2, \ldots, N-1) \]
and \[ a_{n,m} = 0 \text{ for } |n - m| > 1. \]

Linear systems involving non-singular tridiagonal matrices are most easily solved by a modification of the Gaussian elimination algorithm. We leave the details of the algorithm for brevity [12, pp. 375]. We might note that the whole process of solving the system (2.5) requires approximately 3N additions, 3N multiplications and 2N divisions.

2.3 THE DISCRETIZATION ERROR IN LINEAR PROBLEM

Now we shall establish bounds for the discretization error. We note that the true solution \( y(x) \) of the linear problem satisfies the system
\[ A \ y(x) = b + \Gamma(y) \quad \ldots(2.6) \]
where
\[ \Gamma(y) = \begin{bmatrix} y^{(p+2)}(x_1 + \theta_1 h) \\ y^{(p+2)}(x_2 + \theta_2 h) \\ \vdots \\ y^{(p+2)}(x_N + \theta_N h) \end{bmatrix} \]

Here \( |\theta_1| < 1 \), it being assumed that the operator associated with the A.E. (2.3) is definite so that the generalized mean value theorem holds. Now subtracting (2.5) from (2.6), we get the error equation
\[ Ae = \Gamma(y) \quad \ldots(2.7) \]

where \( e \) denotes the error vector.

The matrix \( A \) has the following properties provided \( \beta_1 \geq 0 \) and \( h^2 \beta_1 g_M < 1 \)
(i = 0, 1, 2), where \( g_M = \max_{a \leq x \leq b} g(x) \).

\[
\begin{align*}
(i) & \quad A \text{ is real, irreducible matrix} \\
(ii) & \quad A \text{ is monotone matrix} \\
& \quad \text{i.e. } A^{-1} \geq 0 \\
& \quad \text{and } 0 \leq A^{-1} \leq J^{-1}.
\end{align*}
\]

\( J^{-1} \) is known, in fact if \( J^{-1} = (j_{mn}) \), then

\[
\begin{cases}
\frac{n(N+1-m)}{N+1}, & n \leq m \\
\frac{m(N+1-n)}{N+1}, & n \geq m,
\end{cases}
\]

See Rutherford [23].

The truth of the above properties of the matrix \( A \) follows from the following well known theorems:

**Theorem 2.1**

A tridiagonal matrix \( A = (a_{ij}) \) is irreducible if

\[
(i) \quad a_{i,i-1} \neq 0 \quad (i = 2,3,\ldots, n) \\
(ii) \quad a_{i,i+1} \neq 0 \quad (i = 1,2,\ldots, n-1)
\]

**Theorem 2.2**

Let the matrix \( A = (a_{ij}) \) be irreducible and satisfy the conditions

\[
(i) \quad a_{ij} \leq 0 \quad i \neq j, \quad (i, j = 1,2,\ldots, n) \\
(ii) \quad \sum_{j=1}^{n} a_{ij} \geq 0, \quad i = 1,2,\ldots,n
\]

Then \( A \) is monotone.

**Theorem 2.3**

Let the matrices \( A \) and \( B \) be monotone, and assume that \( A-B \geq 0 \), then

\[ B^{-1} - A^{-1} \geq 0. \]
For the proof of the above theorems, see Henrici [12, pp. 360].

Now we consider the error equation (2.7) and rewrite it in the form
\[ e = A^{-1} \Gamma(y) . \]

From the above equation, we deduce
\[ |e_m| \leq |C_{p+2}| h^{p+2} M_{p+2} \sum_{n=1}^{N} j_{mn} , \]
using (2.8 (ii)),
\[ \leq |C_{p+2}| h^{p+2} M_{p+2} x \frac{m(N+1-m)}{2} \]
\[ \leq |C_{p+2}| h^{p} M_{p+2} \times \frac{(x_m-a) (b-x_m)}{2} \]
\[ \leq \frac{M}{2} h^{p} (x_m-a) (b-x_m) , \]
\[ (m = 1, 2, \ldots, N) \]
where \( M = |C_{p+2}| M_{p+2} \)

Also
\[ ||e|| = \max_{1 \leq m \leq N} |e_m| = M h^{p} (b-a)^{2}/8 \]

We summarize these results in the following theorem.

**Theorem 2.4**

Let \( y(x) \) be the true solution of the linear boundary value problem of class \( M \) and \( y_n \) be its discrete approximation with an error \( e_n \). Let the error equation be
\[ Ae = \Gamma(y) , \quad (\text{see (2.7)}). \]

where \( A \) is a monotone matrix such that \( A - J \geq 0 \)

and \[ ||\Gamma(y)|| \leq |C_{p+2}| h^{p+2} M_{p+2} , \]
then
\[ |e_i| \leq \frac{M h^p (x_i-a) (b-x_i)}{2}, \quad (i = 1, 2, \ldots, N) \]

and

\[ ||e|| \leq \frac{M h^p (b-a)^2}{8}, \]

where

\[ M = \frac{|C_{p+2}|}{M_{p+2}}, \]

or equivalently

\[ ||e|| = O(h^p). \]

A special case of Theorem 2.4 was proved by Varga [25], see Theorem 6.2, pp. 165, in which the approximation of the linear D.E. is based on the A.E.

\[ -y_n + 2y_{n+1} - y_{n+2} + h^2 y''_{n+1} = 0 \quad \ldots (2.10a) \]

with \( p = 2 \) and \( M = \frac{1}{12} M_4 \). Thus

\[ ||e|| \leq h^2 M_4 (b-a)^2/96. \]

A more accurate approximation is based on the A.E.

\[ -y_n + 2y_{n+1} - y_{n+2} + (h^2/12) (y''_n + 10 y''_{n+1} + y''_{n+2}) = 0 \quad \ldots (2.10b) \]

with \( p = 4 \) and \( M = \frac{1}{240} M_6 \) and thus

\[ ||e|| \leq h^4 M_6 (b-a)^2/1,920. \quad \ldots (2.11) \]

The difference operators associated with the above A.E.'s have been proved to be definite and hence the generalized mean value theorem (1.11) holds, (Section 2.1).

### 2.4 MORE PRECISE ERROR BOUNDS

The error bounds obtained in Theorem 2.4 depends on the fact that

\[ 0 \leq A^{-1} \leq J^{-1} \]

...and thus on

\[ ||A^{-1}|| \leq ||J^{-1}|| = \begin{cases} \sqrt{N(N+2)/8}, & N \text{ even} \\ (N+1)^2/8, & N \text{ odd} \end{cases} \]

\[ \leq (N+1)^2/8 = (b-a)^2/8 h^2. \]
However, in view of Theorem 2.3 and properties of the matrix $A$, given by (2.8), $||A^{-1}||$ monotonically increases as $g(x)$ decreases and attains its maximum value $(b-a)^2/8 \ h^2$ when $g(x) = 0$. Setting $A_M = $ matrix $A$ in which each element $g_i$ has been replaced by $g_M = \max_{a \leq x \leq b} g(x)$ and $A_m = $ matrix $A$ in which each $g_i$ has been replaced by $g_m = \min_{a \leq x \leq b} g(x)$, then from the theory of monotone matrices, it follows that

$$||A_M^{-1}|| \leq ||A_m^{-1}|| \leq ||A_m^{-1}|| \leq ||J^{-1}||,$$ ...

(2.12)

equality holds iff $g(x) = 0$. In order to obtain $||A_m^{-1}||$ which would enable us to get a better error bound than already obtained, we consider

$$A_m = \begin{bmatrix}
2 + h^2 \beta_1 g_m & -1 + h^2 \beta_2 g_m \\
-1 + h^2 \beta_0 g_m & 2 + h^2 \beta_1 g_m & -1 + h^2 \beta_2 g_m \\
& \cdots & \cdots & \cdots & \cdots \\
& & \cdots & \cdots & \cdots \\
-1 + h^2 \beta_0 g_m & 2 + h^2 \beta_1 g_m & -1 + h^2 \beta_2 g_m \\
& & & \cdots & \cdots \\
-1 + h^2 \beta_0 g_m & 2 + h^2 \beta_1 g_m & -1 + h^2 \beta_2 g_m
\end{bmatrix},$$

with $\beta_0 = \beta_2$ and $\beta_0 + \beta_1 + \beta_2 = 1$ and this will be the case for $p = 2$ or 4.

Then

$$A_m = \begin{bmatrix}
\mu & -1 \\
-1 & \mu & -1 \\
& \cdots & \cdots & \cdots \\
& \cdots & \cdots & \cdots \\
& & \cdots & \cdots & \cdots \\
-1 & \mu & -1 \\
& & & \cdots & \cdots \\
-1 & \mu & -1
\end{bmatrix} \begin{bmatrix}
\lambda \\
\lambda \\
\cdots \\
\cdots \\
\cdots \\
\lambda
\end{bmatrix} = P_N^{-1}(\mu) \ Q \ , \ \text{where} \ \lambda = 1 - h^2 \beta_0 g_m$$

and $\mu = (2 + h^2 \beta_1 g_m)/(1 - h^2 \beta_0 g_m)$.

Therefore

$$A_m^{-1} = Q^{-1} P_N^{-1}(\mu) ,$$
We now proceed to determine exact expressions for $P_n^{-1}(\mu)$ and $||P_n^{-1}(\mu)||$.

Let $D_n$ denote the determinant of the matrix $P_n(\mu)$, then it is easy to show using Laplace's method of expanding determinants that $D_n$ satisfies a Δ.E. with constant coefficients viz. (2.13a).

\[
D_n = \mu D_{n-1} - D_{n-2} \quad (n = 1, 2, \ldots) \quad \text{...(2.13a)}
\]

\[D_{-1} = 0, \quad D_0 = 1\]

Please note that for a given problem and for a fixed $h$, $\mu$ is constant. On solving the above Δ.E. with initial conditions $D_{-1} = 0$, $D_0 = 1$, we obtain

\[
D_n = \begin{cases} 
\frac{\sinh (n+1)\theta}{\sinh \theta}, & \text{if } \mu = 2 \cosh \theta > 2 \\
\frac{\sinh (n+1)\theta}{\sin \theta}, & \text{if } \mu = 2 \\
\frac{\sin (n+1)\theta}{\sin \theta}, & \text{if } \mu = 2 \cos \theta < 2
\end{cases}
\]

...(2.13b)

We first determine the inverse of the matrix $P_n(\mu)$ for smaller value of say $n = 1, 2, 3$ and $4$. This can readily be done using the Gauss elimination method or the adjoint method for finding the inverse of a matrix if $n$ is not very large. The $P_n^{-1}(\mu)$ for $n = 1, 2, 3$ and $4$ is given below:

\[
P_1^{-1}(\mu) = \frac{1}{D_1} [1] = \frac{1}{D_1} [D_0^0 D_0^0]
\]

\[
P_2^{-1}(\mu) = \frac{1}{D_2} \begin{bmatrix}
D_0^1 & D_0^0 \\
D_0^0 & D_1^0
\end{bmatrix}
\]
The form of the above inverse matrices suggests that in general
\[ P_n^{-1}(\mu) = (p_{ij}) \]
\[ \ldots (2.13c) \]
where
\[ p_{ij} = p_{ji} = D_{j-1} D_{n-1-i} / D_n, \ j \leq i. \]

We will prove the above result (2.13c) using induction argument below.

We assume that the above result is true for the matrix \( P_{n-1}(\mu) \), i.e. if \( P_{n-1}^{-1}(\mu) = (p_{ij}) \), then \( p_{ij} = p_{ji} = D_{j-1} D_{n-1-i} / D_{n-1} \), \ j \leq i. \ Let the sequence of the matrices \( P_1(\mu), P_2(\mu), \ldots P_n(\mu) \) be written as follows:

\[ P_1(\mu) = [\mu], P_2(\mu) = \begin{bmatrix} P_1(\mu) & -1 \\ -1 & \mu \end{bmatrix}, P_3(\mu) = \begin{bmatrix} P_2(\mu) & 0 \\ 0 & 1 \end{bmatrix}, \ldots \]

\[ P_n(\mu) = \begin{bmatrix} P_{n-1}(\mu) & 0 \\ 0 & 1 \\ \vdots \\ 0 & 0 \end{bmatrix} \]

In general, we partition the matrix \( P_n(\mu) \) into 4 sub-matrices which we write as
\[ P_n(\mu) = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \]
Here $\alpha_{11} = P_{n-1}(\mu)$, $\alpha_{21} = [0 \ 0 \ldots 0 -1]$, (1×n-1) row matrix, $\alpha_{12}$, a (n-1×1) column matrix and equals the transpose of $\alpha_{21}$ and finally $\alpha_{22} = \mu$.

If we write

$$P_n^{-1}(\mu) = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}$$

partitoned exactly in the same form as $P_n(\mu)$ is, then we have

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} = I_n,$$

where $I_n$ is the unit matrix of order $n$. On multiplying the 2 matrices on the left, we get four matrix equations namely

$$\begin{align*}
\alpha_{11} \beta_{12} + \alpha_{12} \beta_{22} &= 0 \\
\alpha_{21} \beta_{12} + \alpha_{22} \beta_{22} &= 1 \\
\alpha_{11} \beta_{11} + \alpha_{12} \beta_{21} &= I_{n-1} \\
\alpha_{21} \beta_{11} + \alpha_{22} \beta_{21} &= 0 .
\end{align*}$$

Solving these matrix equations for $\beta_{11}, \beta_{12}, \beta_{21}$ and $\beta_{22}$, we have

$$\begin{align*}
\beta_{22} &= D \\
\beta_{21} &= -D (\alpha_{21} \alpha_{11}^{-1}) \\
\beta_{12} &= -D (\alpha_{11}^{-1} \alpha_{12}) \\
\beta_{11} &= \alpha_{11}^{-1} + (\alpha_{11}^{-1} \alpha_{12}) D (\alpha_{21} \alpha_{11}^{-1}) \\
&= \alpha_{11}^{-1} + (\alpha_{12} \alpha_{21}) / D
\end{align*}$$

where

$$D = (\alpha_{22} - \alpha_{21} \alpha_{11}^{-1} \alpha_{12})^{-1}.$$ 

On substituting $\alpha_{11}^{-1} = P_n^{-1}(\mu)$ and the values of the matrices $\alpha_{12}, \alpha_{21}$ and $\alpha_{22}$ in the above relations, we get,

$$\begin{align*}
\beta_{22} &= D_{n-1} / D_n \\
\beta_{21} &= (D_0, D_1, \ldots, D_{n-2}) / D_n \\
\beta_{12} \text{ being equal to the transpose of } \beta_{21}
\end{align*}$$
and

\[ \beta_{11} = \beta_{n-1} + \beta_{21}/D \]

\[ = (D_{j-1} D_{n-1-i}/D_{n-1}) + \frac{1}{D_n D_{n-1}} (D_{j-1} D_{i-1}) \]

\[ = \frac{1}{D_n} (D_{j-1} D_{n-1-i} + D_{i-1}), \ j \leq i \]

\[ = (D_{j-1} D_{n-i}/D_n), \ j \leq i, \]

using the identity

\[ (D_n D_{n-1-i} + D_{i-1})/D_{n-1} = D_{n-i} \quad (i = 0, 1, 2, ..., n) \quad \ldots (2.13d) \]

**Proof of the Identity**

We will prove the identity for \( u \leq 2 \) using (2.13a).

Consider the \( \Delta.E. \) (2.13a) in the form

\[ D_r = \mu D_{r-1} - D_{r-2} \quad (r = 1, 2, 3, ...) \]

\[ D_{-1} = 0, \ D_0 = 1 \]

We now consider the above \( \Delta.E. \) for \( r = n \) and \( r = n-i \), and eliminate \( \mu \) between these two relations and write the result in the form

\[ (*) \ D_n D_{n-i-1} - D_{n-1} D_{n-i} = D_{n-1} D_{n-i-2} - D_{n-2} D_{n-i-1}. \]

We then consider the above \( \Delta.E. \) for \( r = n-1 \) and \( r = n-i-1 \) and prove

\[ (**) D_{n-1} D_{n-i-2} - D_{n-2} D_{n-i-1} = D_{n-2} D_{n-3} - D_{n-3} D_{n-i-2}. \]

Next we consider the \( \Delta.E. \) for \( r = n-2 \) and \( r = n-i-2; \ldots; \) and finally \( r = i+1 \) and \( r = 1 \).

Combining the results of the form \( (*) \) and \( (**) \), we get

\[ (***) D_r D_{r-i-1} - D_{r-1} D_{r-i} = \text{Constant} = \lambda \ \text{(say)}, \ (r = n, n-1, \ldots, i). \]

In order to evaluate \( \lambda \), we put \( r = i \) in \( (***) \) and thus \( \lambda = -D_{i-1} \).

Hence

\[ D_n D_{n-i-1} - D_{n-1} D_{n-i} = \lambda = -D_{i-1} \quad (i = 0, 1, 2, \ldots, n), \]
or \( \frac{(D_n D_{n-i-1} - D_{i-1})}{D_{n-1}} = D_{n-i} \),

and this completes the proof of identity (2.13d).

Now substituting \( \beta_{11}, \beta_{12}, \beta_{21} \) and \( \beta_{22} \) we finally obtain

\[
P_n^{-1}(\mu) = (p_{ij})
\]

where \( p_{ij} = p_{ji} = \left( \frac{D_{j-1} D_{n-i}}{D_n} \right), \ j \leq i \)

and this completes the proof of (2.13c).

For the determination of \( ||P_n^{-1}(\mu)|| \), we will require two more identities which we give below:

\[
\sum_{j=0}^{n-1} D_j = \frac{1}{\mu-2} \left[ D_1 - D_{i-1} - 1 \right], \ (\mu \neq 2) \quad \ldots(2.13e)
\]

**Proof of Identity**

We write (2.13a) for \( n = 1, 2, \ldots, i \) and then summing up all the \( i \) equations columnwise, we get

\[
D_1 = \mu D_0 - D_{-1}
\]

\[
D_2 = \mu D_1 - D_0
\]

\[
D_3 = \mu D_2 - D_1
\]

\[
\cdots
\]

\[
D_i = \mu D_{i-1} - D_{i-2}
\]

\[
-D_0 + D_1 + (\sum_{j=0}^{i-1} D_j) = \mu \sum_{j=0}^{i-1} D_j - (\sum_{j=0}^{i-1} D_j - D_{i-1})
\]

\[
(\mu - 2) \sum_{j=0}^{i-1} D_j = (D_1 - D_{i-1} - 1)
\]

or \( \sum_{j=0}^{i-1} D_j = \frac{1}{\mu-2} [D_1 - D_{i-1} - 1], \ (\mu \neq 2) \).
\[ D_{n-1} D_i - D_{i-1} D_{n-i-1} = D_n \]  
...(2.13f)

The above identity can also be proved using the A.E. (2.13a) and following the technique employed for proving (2.13d).

We now state the theorem giving \( ||P_n^{-1}(\mu)|| \) below:

**Theorem 2.5**

\[
||P_n^{-1}(\mu)|| \leq \begin{cases} 
[1 - \text{sech} \left( \frac{\mu+1}{2} \theta \right)]/(\mu-2), & \text{if } \mu > 2 \\
(n+1)^2/8, & \text{if } \mu = 2 \\
\frac{n}{\sin \theta \sin (n+1) \theta}, & \text{if } \mu < 2 
\end{cases}
\]

equality holds only if \( n \) is odd positive integer and \( \mu \geq 2 \).

**Proof**

For \( \mu > 2 \)

\[
\sum_{j=1}^{n} |P_{ij}| = \frac{1}{D_n} \left[ \sum_{j=1}^{i} D_{j-1} D_{n-i} + \sum_{j=i+1}^{n} D_{n-j} D_{i-1} \right]
\]

\[
= \frac{1}{D_n} [D_{n-1} \sum_{j=1}^{i} D_{j-1} + D_{i-1} \sum_{j=i+1}^{n} D_{n-j}]
\]

\[
= \frac{D_{n-1}}{\mu-2} D_n [D_i - D_{i-1} - 1] + \frac{D_{i-1}}{\mu-2} D_n [D_{n-i} - D_{n-i-1} - 1] \quad \text{using (2.13e)},
\]

\[
= \frac{1}{\mu-2} \left[ \frac{D_i}{D_n} D_{n-i} - \frac{D_{i-1}}{D_n} D_{n-i-1} - \frac{D_{n-i}}{D_n} + \frac{D_{i-1}}{D_n} \right]
\]

\[
= \frac{1}{\mu-2} [1 - \frac{D_{n-i} + D_{i-1}}{D_n}] \quad \text{using (2.13f)}.
\]

Since \( P_n^{-1}(\mu) \) is symmetric and also is such that

\[
\sum_{j=1}^{n} P_{ij} = \sum_{j=1}^{n} P_{n+1-i,j}.
\]

therefore, row sums increase with \( i \), hence \( \sum_{j=1}^{n} P_{ij} \) is maximum for
\[ i = \begin{cases} \frac{n}{2}, & \text{n even} \\ \frac{n+1}{2}, & \text{n odd} \end{cases} \]

thus,

\[ ||P_n^{-1}(\mu)|| = \begin{cases} \frac{1}{\mu-2} \left(1 - \frac{D_{n/2} + D_{n/2} - 1}{D_n}\right), & \text{n even} \\ \frac{1}{\mu-2} \left(1 - \frac{2D_{n-1/2}}{D_n}\right), & \text{n odd} \end{cases} \]

\[ \leq \frac{1}{\mu-2} \left(1 - \text{sech} \left(\frac{n+1}{2} \theta\right)\right), \mu > 2, \]

(equality holds for n odd positive integer).

The case for \(\mu = 2\) has been given in the beginning of the Section 2.4.

For \(\mu < 2\), \(D_n = \sin(n+1)\theta / \sin \theta\),

and

\[ \sum_{j=1}^{n} |p_{ij}| = \sum_{j=1}^{i} \left|\frac{D_{j-1}D_{n-i}}{D_{n}}\right| + \sum_{j=i+1}^{n} \left|\frac{D_{n-j}D_{i-1}}{D_{n}}\right| \]

\[ = \sum_{j=1}^{i} \left|\frac{\sin(j\theta) \sin(n-i+1)\theta}{\sin \theta \sin(n+1)\theta}\right| + \sum_{j=i+1}^{n} \left|\frac{\sin(n-j+1)\theta \sin(i\theta)}{\sin \theta \sin(n+1)\theta}\right| \]

\[ < \frac{1}{|\sin \theta \sin(n+1)\theta|} \left|i + (n-i)\right| \]

\[ < n/|\sin \theta \sin(n+1)\theta| \]

(an expression independent of \(i\)).

Thus \(||P_n^{-1}(\mu)|| < n/|\sin \theta \sin(n+1)\theta|, \mu < 2\).

This completes the proof of the theorem 2.5.

The above results enable us to write down the norm of another matrix \(\tilde{P}_n(\mu)\), closely related to the matrix \(P_n(\mu)\). The norm of the matrix \(\tilde{P}_n(\mu)\) defined below will be required later in Section 3.3.
\[
\beta_n(y) = \begin{bmatrix}
\mu & 1 \\
1 & \mu & 1 \\
& & \ddots \\
& & & 1 & \mu & 1 \\
& & & & & 1 & \mu
\end{bmatrix}
= [-\beta_n(-\mu)]
\]...
(2.14)

Let \( D_n \) denote the determinant of the matrix \( \beta_n(y) \). Expansion of the determinant then leads to the following A.E.

\[
\tilde{D}_n = \mu \tilde{D}_{n-1} - \tilde{D}_{n-2} , \tilde{D}_{-1} = 0 , \tilde{D}_0 = 1
\]
from which immediately follows (see (2.13a)) that

\[
\tilde{D}_n(\mu) = D_n(|\mu|)
\]
The inverse of the matrix \( \beta_n(y) \) is known explicitly. If \( \beta_n^{-1}(\mu) = (\beta_{ij}) \), then

\[
\beta_{ij} = \beta_{ji} = (-1)^{i+j} \tilde{D}_{j-1} \tilde{D}_{n-1}/\tilde{D}_n , j \leq i,
\]
see Rutherford (1952), [23].

The elements of the matrix \( \beta_n^{-1}(\mu) \) can also be deduced using (2.13c). Now since \( \tilde{D}_n(\mu) = D_n(|\mu|) \), and \( D_n \) is an even or odd function of \( \mu \) according as \( n \) is an even or odd positive integer, we notice that the elements of the inverse matrices \( \beta_n^{-1}(\mu) \) and \( \beta_n^{-1}(\mu) \) are the same in magnitude, but differ in sign only. Hence

\[
||\beta_n^{-1}(\mu)|| = ||\beta_n^{-1}(\mu)||
\]...
(2.15)

In our present discussion

\[
\mu = (2 + h^2 \beta_1 g_m)/(1 - h^2 \beta_0 g_m) > 2
\]
This follows from the fact that \( \beta_1 > 0 \), \( g(x) > 0 \) and \( h^2 \beta_0 g_m < 1 \). The latter condition viz. \( h^2 \beta_0 g_m < 1 \) is an immediate consequence of the assumption

\[
h^2 \beta_i g_m < 1 \quad (i = 0, 1, 2)\]
which we made in proving the irreducibility of the matrix \( A \) (see 2.8(i)).

Thus finally, for \( \mu > 2 \)

\[
||A_m^{-1}|| \leq H_N(\mu)/h^2 g_m
\]

...(2.16)

where

\[
H_N(\mu) = [1 - \text{sech} \left( \frac{N+1}{2} \theta \right), 2 \cosh \theta = \mu
\]

(N refers to the order of the associated matrix),

\[
= [1 - \text{sech} \left( \frac{N+1}{2} \log \frac{\mu + \sqrt{\mu^2 - \xi}}{2} \right)].
\]

Thus again using the error equation (2.7) we get for the linear boundary value problem of class M the more accurate error bound

\[
||e|| \leq M h^p H_N(\mu)/g_m
\]

...(2.17)

Now in view of the above, we restate the Theorem 2.4 as follows:

**Theorem 2.6**

Let \( y(x) \in C^{p+2} \) be the true solution of the linear boundary value problem of class M (given by (2.1) for \( f(x,y) = g(x)y + s(x) \)) and let \( y_n \) be its discrete approximation based on \( \Delta E. \) (2.3). Furthermore, let \( \min. g(x) = g_m > 0 \) and \( h \) be such that \( h^2 \beta_i g_m < 1 \) (\( i = 0,1,2 \)), then

\[
||e|| \leq M h^p H_N(\mu)/g_m
\]

or

\[
||e|| = O(h^p).
\]

**Particular Cases of Theorem 2.6**

(i) Using \( p = 2, \beta_0 = \beta_2 = 0, \beta_1 = 1, M = M_4/12 \), we get

\[
||e|| \leq h^2 M_4 H_N(\mu)/12 g_m, \mu = 2 + h^2 g_m
\]

...(2.18)

(ii) Using \( p = 4, \beta_0 = \beta_2 = 1/12, \beta_1 = 10/12 \) and \( M = M_6/240 \), we get

\[
||e|| \leq h^4 M_6 H_N(\mu)/240 g_m, \mu = \frac{2+(10/12)h^2 g_m}{1-(1/12)h^2 g_m}
\]

...(2.19)

The error bound given by (2.19) is better than the one given by (2.11) as is displayed by the following Table I. For experimentation, we chose the
problem

\[ y'' - y = x^2 - 2, \quad y(0) = 0, \quad y(b) = (2 \sinh b/\sinh 1) - b^2 \quad \ldots (2.20) \]

with \( y(x) = (2 \sinh x/\sinh 1) - x^2 \) for a series of values of \( b \) and \( h \) respectively, (Table I).

** Table I

<table>
<thead>
<tr>
<th>( b )</th>
<th>( h )</th>
<th>Error bound ( E_1 ) based on (2.11)</th>
<th>Error bound ( E_2 ) based on (2.19)</th>
<th>( E_1/E_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/4</td>
<td>0.4069 E-5**</td>
<td>0.3684 E-5</td>
<td>1.10</td>
</tr>
<tr>
<td>1/8</td>
<td>0.2543 E-6</td>
<td>0.2303 E-6</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>1/16</td>
<td>0.1589 E-7</td>
<td>0.1439 E-7</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>1/32</td>
<td>0.9934 E-9</td>
<td>0.8995 E-9</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1/4</td>
<td>0.6510 E-4</td>
<td>0.2390 E-4</td>
<td>2.72</td>
</tr>
<tr>
<td>1/8</td>
<td>0.4069 E-5</td>
<td>0.1494 E-5</td>
<td>2.72</td>
<td></td>
</tr>
<tr>
<td>1/16</td>
<td>0.2543 E-6</td>
<td>0.9336 E-7</td>
<td>2.72</td>
<td></td>
</tr>
<tr>
<td>1/32</td>
<td>0.1589 E-7</td>
<td>0.5835 E-8</td>
<td>2.72</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1/4</td>
<td>0.1042 E-2</td>
<td>0.3253 E-4</td>
<td>32.02</td>
</tr>
<tr>
<td>1/8</td>
<td>0.6510 E-4</td>
<td>0.2033 E-5</td>
<td>32.02</td>
<td></td>
</tr>
<tr>
<td>1/16</td>
<td>0.4069 E-5</td>
<td>0.1271 E-6</td>
<td>32.02</td>
<td></td>
</tr>
<tr>
<td>1/32</td>
<td>0.2543 E-6</td>
<td>0.7942 E-8</td>
<td>32.02</td>
<td></td>
</tr>
</tbody>
</table>

** We write 0.4069 E-5 for 0.4069 \times 10^{-5}. We will adopt this convention throughout this thesis.
An $O(h^4)$ three point finite difference analogue was also formulated by Aziz and Hubbard (1964), [2]. In fact the problem considered by them was

$$\begin{align*}
L y &= -y'' + g(x) y = s(x) \\
y(0) &= y(1) = 0
\end{align*} \quad \text{...(2.21)}$$

where $g(x) \geq 0$, $g(x)$, $s(x)$ are of class $C^4$ on $[0,1]$ and $y(x) \in C^6$. The main result of their paper is that the discretization error satisfies the inequality

$$|e| \leq \frac{1}{4} \left[ 1 - \frac{5h^2}{48} M \right]^{-1} \left[ \frac{7 h^4 M}{720} + \frac{h^4}{6} \left| s^{(4)}(x) \right| M \right] \quad \text{...(2.22)}$$

The error bound given by (2.19) is better than the one given by either (2.11) or (2.22). For instance in the problem $y'' - y = -x e^x$, $y(0) = y(1) = 0$, the error bounds given by (2.11), (2.19) and (2.22) are $0.51 \ E^{-5}$, $0.46 \ E^{-5}$ and $26.0 \ E^{-5}$ respectively with $h = 0.1$.

In the last chapter, we will compare the accuracy of the method proposed by Aziz and Hubbard with other known techniques and with those developed in the present thesis. The numerical technique in which the two point linear boundary value problem is replaced by $\Delta E.$ (2.10) with the error bound as given by (2.19) will henceforth be designated for reference purposes as $M^{-1}$. Similarly, the method proposed by Aziz and Hubbard will be referred to as $M^{-AH}$. 
3.1

In the last chapter, we considered the use of A.E. (1.2) with \( k = 2 \) for the numerical solution of linear boundary value problem of class M. We proved that the resulting error in one case is \( O(h^2) \), while in the other case it is \( O(h^4) \). The two different error bounds each of \( O(h^4) \) are given by (2.11) and (2.19) respectively. Now if we want to consider finite difference analogues in which the resulting error is \( O(h^t) \), \( t>4 \), then one way to do this is to consider (A.E.) (1.2) with \( k>2 \). Other methods for obtaining accurate A.E.'s without increasing the order \( k \) will be considered in the next chapter.

We have already seen in Chapter I (1.17) that a A.E. of positive degree has an associated polynomial

\[
a(x) = \sum_{i=0}^{k} a_i x^i = (x-1)^2 \gamma(x) ,
\]

...(3.1)

where \( \gamma(x) \) is a monic polynomial of degree \( k-2 \). In fact any choice of the polynomial \( \gamma(x) \) readily determines \( a \)'s. We then determine the coefficients \( \beta \)'s from a set of linear equations

\[
C_i = 0 \ (i = 2,3,...,k+2) ,
\]

...(3.2)

see (1.8). The above system of linear equations satisfied by \( \beta \)'s can be written in matrix form as follows:
where all summations in (3.3) extend from \( i=0 \) to \( i=k \). The matrix associated with the above system is a Vandermonde matrix, hence it is non-singular and the coefficients \( \beta \)'s are therefore uniquely determined.

The determination of \( \beta_i \)'s for a fixed value of \( k \) can in general be effected in two ways. We can either invert this Vandermonde matrix, Parker [21], or we can use Gauss elimination method. In next section, we will consider A.E. (1.2) for \( k = 4 \) and develop a one parameter family of A.E.'s to approximate the linear boundary value problem of class M.

3.2 A.E. OF ORDER \( k = 4 \)

In view of (3.1), \( \alpha(x) = (x-1)^2 \gamma(x) \) where \( \gamma(x) \) is now a monic polynomial of degree 2. Hence

\[
\alpha(x) = (x-1)^2 (x^2 + cx + d)
\]

\[
= x^4 + (c-2) x^3 + (d-2c+1) x^2 + (c-2d) x + d,
\]

where \( c \) and \( d \) are real arbitrary constants.

Thus

\[
\begin{bmatrix}
\alpha_0 \\
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\alpha_4 \\
\end{bmatrix} =
\begin{bmatrix}
d \\
c - 2d \\
-2c + d + 1 \\
c - 2 \\
1 \\
\end{bmatrix}
\]  

...(3.4)
Any choice of $c$ and $d$ determines $\alpha$'s and then we determine $\beta$'s by solving the system (3.3) for $k = 4$, and T.E. using (1.10). An easy calculation shows that

$$
\begin{align*}
\beta_0 &= (-c + 19d - 1)/240 \\
\beta_1 &= (12c + 102d + 2)/120 \\
\beta_2 &= (97c + 7d + 7)/120 \\
\beta_3 &= (12c + 2d + 102)/120 \\
\beta_4 &= (-c - d + 19)/240
\end{align*}
\quad \text{...(3.5)}
$$

and

$$
\text{T.E.} = \begin{cases} 
\frac{d-1}{240} h^7 y(7), d \neq 1 \\
\frac{31c-190}{60,400} h^8 y(8), d=1, c \neq \frac{190}{31} \\
-\frac{79}{585,900} h^{10} y(10), d=1, c = \frac{190}{31}
\end{cases}
\quad \text{...(3.6)}
$$

However, for obtaining symmetric $\Delta E$'s, we will require that $d=1$. The T.E. of the corresponding $\Delta E$ will be $O(h^8)$ and we might expect that the resulting error in the D.E. will be $O(h^6)$ at most. Thus for $d=1$, the coefficients $\alpha$'s and $\beta$'s are given as follows:

$$
\begin{align*}
\alpha_0 &= \alpha_4 = 1 \\
\alpha_1 &= \alpha_3 = c-2 \\
\alpha_2 &= -2c + 2 \\
\beta_0 &= \beta_4 = (-c + 18)/240 \\
\beta_1 &= \beta_3 = (12c + 104)/120 \\
\beta_2 &= (97c + 14)/120
\end{align*}
\quad \text{...(3.7)}
$$

We notice that $\beta$'s are non-negative for any value of $c$ satisfying $-14/97 \leq c \leq 18$. Thus we arrive at a one parameter family of symmetric
\[ -y_n + y_{n+4} + (2-c) (y_{n+1} + y_{n+3}) + (2c-2) y_{n+2} + h^2 \sum_{i=0}^{4} \beta_i y''_{n+i} = 0 \quad \ldots (3.8) \]

where \( c \) is an arbitrary parameter and \( \beta \)'s are given by \((3.7)\). The T.E. associated with \((3.8)\) is given by \((3.6)\).

Each equation \((3.8)\) may be considered as being centred on \( x_{n+2} \) and represents an equation for \( y_{n+2} \). But because two additional points are involved on either side, it cannot be used as an equation for either \( y_1 \) or \( y_N \) and therefore two additional equations must be introduced. The \( \Delta.E.'s \) used near the boundaries are given by \((3.9)\) below. The reason for this particular choice of \( \Delta.E.'s \) \((3.9)\) is that the matrix \( J(c) \) defined below by \((3.11)\) becomes real, symmetric, irreducible and monotone for any \( c > 2 \).

\[
\begin{cases}
(i) & -c y_0 + (2c-1) y_1 + (2-c) y_2 - y_3 + h^2 \sum_{i=0}^{3} \bar{\beta}_i y''_i = 0 \\
(ii) & -y_{N-2} + (2-c) y_{N-1} + (2c-1) y_N - c y_{N+1} + h^2 \sum_{i=0}^{3} \bar{\beta}_{3-i} y''_{N-2+i} = 0 \quad \ldots (3.9)
\end{cases}
\]

where \( \bar{\beta}_0 = c/12, \bar{\beta}_1 = (10c+1)/12, \bar{\beta}_2 = (c+10)/12 \) and \( \bar{\beta}_3 = 1/12 \). The T.E. associated with \((3.9)\) is given by

\[
\text{T.E.} = \begin{cases}
\frac{-(c+1)}{240} h^6 y^{(6)}, \quad c \neq 1 \\
\frac{-1}{240} h^7 y^{(7)}, \quad c = -1 \text{ for } \Delta.E. (3.9i) \\
\frac{1}{240} h^7 y^{(7)}, \quad c = -1 \text{ for } \Delta.E. (3.9ii)
\end{cases} \ldots (3.10)
\]

The usual details of the derivation of equation \((3.9)\) have been omitted for brevity. We notice that the coefficients \( \bar{\beta} \)'s are positive for any value of \( c > 0 \). Now we would like to find those values of the parameter \( c \)
for which the Δ.E.'s (3.8) and (3.9) are such that the difference operators associated with them become definite and the generalized mean value theorem (1.11) holds.

In case of Δ.E. (3.8), we have

\[
L[y(x); h] = [y(x) + y(x+4h)] + (c-2) [y(x+h) + y(x+3h)] \\
+ (2-2c) y(x+2h) - h^2 \left[ \frac{-c+18}{240} [y''(x) + y''(x+4h)] \\
+ \frac{3c+26}{30} [y''(x+h) + y''(x+3h)] + \frac{97c+14}{120} y''(x+2h) \right] \\
= h^8 \int_0^4 G(s) y^{(8)}(x+hs) \, ds
\]

where

\[
G(s) = \begin{cases} 
\frac{(4-s)^7}{7!} - \frac{c+18}{240} \cdot \frac{(4-s)^5}{5!} , & 3 < s \leq 4 \\
\frac{(4-s)^7}{7!} + (c-2) \frac{(3-s)^7}{7!} - \frac{c+18}{240} \cdot \frac{(4-s)^5}{5!} - \frac{3c+26}{30} \cdot \frac{(3-s)^5}{5!} , & 2 \leq s \leq 3
\end{cases}
\]

and \( G(s) = G(4-s), 0 \leq s \leq 2 \).

i.e. the function \( G(s) \) is symmetric about the line \( s=2 \).

Now for \( 3 < s \leq 4 \)

\[
G(s) = \frac{(4-s)^5}{6!} \left[ \frac{(4-s)^2}{7} - \frac{18-c}{40} \right] \leq 0 ,
\]

provided \( c \leq 86/7 \).

Furthermore, \( G(2) = (96c - 448)/(5 \times 8!) \leq 0 \),

provided \( c \leq 14/3 \),

and

\[
G(3) = \frac{(3c - 86)/(5 \times 8!)}{c \leq 86/7}.
\]

Numerical evidence indicates that \( G(s) \) has no zeros in the interval \((2,3)\).

The distribution of the zeros for a series of values of \( c \) in \([2, 14/3]\) is shown in Table II.
TABLE II

DISTRIBUTION OF ZEROS OF G(s) FOR 2<s<3

<table>
<thead>
<tr>
<th>c</th>
<th>The interval in which all the real zeros of G(s) lie</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>(1, 1.5), (3, 3.5), (12.5, 13)</td>
</tr>
<tr>
<td>2.5</td>
<td>(1, 1.5), (3, 3.5), (10, 10.5)</td>
</tr>
<tr>
<td>3</td>
<td>(1, 1.5), (3, 3.5), (9, 9.5)</td>
</tr>
<tr>
<td>3.5</td>
<td>(1, 1.5), (3, 3.5), (8.5, 9)</td>
</tr>
<tr>
<td>4</td>
<td>(1, 1.5), (3, 3.5), (8, 8.5)</td>
</tr>
<tr>
<td>4.5</td>
<td>(1, 1.5), (3, 3.5), (7.5, 8)</td>
</tr>
<tr>
<td>4.67</td>
<td>(1, 1.5), (3, 3.5), (7, 7.5)</td>
</tr>
</tbody>
</table>

The function $-10^3 G(s)$, $2<s<3$, is tabulated below in Table III for the same sequence of values of c. This table also shows that the function $G(s)$ is of constant sign in [2,3].
Thus it follows that \( G(s) \leq 0 \) for \( 2 \leq s \leq 3 \). Combining the results proved above we get,

\[
G(s) \leq 0 \text{ for } s \in [0,4],
\]
provided \( 2 \leq c \leq 14/3 \), so that the generalized mean value theorem holds for these values of \( c \).

In a similar way, it can be proved that the operator associated with the \( \Delta \text{E.} \) (3.9) is also definite and satisfies \( G(s) \leq 0 \) for \( c > 2 \).

Now we make use of the \( \Delta \text{E.} \) (3.8) at the interior points and the \( \Delta \text{E.}'s \) (3.9) at the boundaries and consequently the linear boundary value problem of class \( M \) is replaced by

\[
A(c) y = \bar{b} \quad \ldots (3.11)
\]

where \( A(c) = J(c) + h^2 D(c) G \) and \( \bar{b} = d - h^2 D(c) G \).

Here
\[ J(c) = \begin{bmatrix}
2c-1 & 2c & -1 \\
2c & 2c-2 & 2c & -1 \\
-1 & 2c & 2c-2 & 2c & -1 \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
-1 & 2c & 2c-2 & 2c & -1 \\
\end{bmatrix}, \quad \ldots (3.12) \]

\[ D(c) = \begin{bmatrix}
\beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 \\
\beta_1 & \beta_2 & \beta_3 & \beta_4 \\
\beta_2 & \beta_3 & \beta_4 & \vdots \\
\vdots & \vdots & \vdots & \ddots \\
\beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 \\
\end{bmatrix}. \]

\[ d = \begin{bmatrix}
\frac{c}{h^2} y_a - h^2 \beta_0 f(a, y_a) \\
y_a - h^2 \beta_0 f(a, y_a) \\
0 \\
0 \\
\vdots \\
0 \\
y_b - h^2 \beta_4 f(b, y_b) \\
c y_b - h^2 \beta_0 f(b, y_b) \\
\end{bmatrix}, \]

the matrix $G$ and vector $s$ have been given before, Section 2.2. Finally, the error equation is obtained in the form
\[ A(c) \mathbf{e} = \Gamma(y) , \quad \ldots (3.13) \]

where now,
\[
\Gamma(y) = \begin{bmatrix}
\frac{(c+1)}{240} h^6 y^{(6)} (\xi_1) \\
-\frac{31c-190}{60,480} h^8 y^{(8)} (\xi_2) \\
\quad \ddots \\
-\frac{31c-190}{60,480} h^8 y^{(8)} (\xi_{N-1}) \\
\frac{c+1}{240} h^6 y^{(6)} (\xi_N)
\end{bmatrix}
\]

for any \( c \) satisfying \( 2 < c \leq 14/3 \),

and
\[
||\Gamma(y)|| \leq h^2 K . \quad \ldots (3.14)
\]

Here \( K = \max \left[ h^4 \frac{(c+1)}{M_6/240} , h^6 |31c-190| M_8/60, 480 \right] \).

3.3 PROPERTIES OF MATRICES \( J(c) \) AND \( A(c) \):

We shall prove that the matrices \( J(c) \) and \( A(c) \) are irreducible and monotone.

The geometrical concept of irreducibility by means of graph theory is quite useful, see Varga [25], Chapter I. We will state here the main theorem which will find application in our analysis.

Let \( A = (a_{ij}) \) be any \((n \times n)\) complex matrix. Consider any \( n \) distinct points \( P_1, P_2, \ldots, P_n \) in the plane, which we shall call nodes. For every non-zero entry \( a_{ij} \) of the matrix, we connect node \( P_i \) to \( P_j \) by means of a path \( P_i P_j \) directed from \( P_i \) to \( P_j \), as shown in the adjoining figure.
When $i = j$ and $a_{ij} \neq 0$, we get what is called a loop.

A directed graph is strongly connected if, for any ordered pair of nodes $P_i$ and $P_j$, there exists a path

$$P_i \rightarrow P_{m_1} \rightarrow P_{m_2} \rightarrow \ldots \rightarrow P_{m_{r-1}} \rightarrow P_{m_r} = P_j$$

connecting $P_i$ to $P_j$.

**Theorem 3.1**

An $(n \times n)$ complex matrix $A = (a_{ij})$ is irreducible iff its directed graph is strongly connected.

For proof see Varga [25], Theorem 1.6. From the above theorem we immediately deduce the following corollary.

**Corollary 3.2**

If $A = (a_{ij})$ is a five band matrix such that

- $a_{i,i-1} \neq 0$ for $i = 2, 3, \ldots, n$,
- $a_{i,i+1} \neq 0$ for $i = 1, 2, \ldots, n-1$,

then $A$ is irreducible.

**Proof:** In the proof of the theorem, we only require the directed subgraph $G(A)$ of the matrix $A = (a_{ij})$ namely

$$G(A): P_1 \rightarrow P_2 \rightarrow P_3 \rightarrow \ldots \rightarrow P_{n-2} \rightarrow P_{n-1} \rightarrow P_n$$

Clearly the subgraph $G(A)$ of the matrix $A$ is strongly connected, hence $G(A)$, its directed graph, is strongly connected as well. Therefore $A$ is an irreducible matrix.

As a consequence of this corollary we deduce that the matrix
A(c) defined by (3.11) is irreducible if
\[
\begin{align*}
(i) & \quad (2-c) + h^2 \beta_2 g_i \neq 0 \quad (i = 2, N-1) \\
(ii) & \quad (2-c) + h^2 \beta_j g_i \neq 0 \quad (j = 1,3), (i = j, j+1, \ldots, N+j-3).
\end{align*}
\] ...(3.15)
These conditions will be satisfied for $c > 2$ whenever
\[
\begin{align*}
& h^2 \beta_2 g_M < c-2 \\
& h^2 \beta_j g_M < c-2, \quad (j = 1,3).
\end{align*}
\] ...(3.16)
This is the case when either $\beta_2 = \beta_1 = \beta_3 = 0$ or else for sufficiently small values of $h$.

Having proved $A(c)$ to be irreducible, it can be proved using Theorem 2.2 to be monotone as well for sufficiently small values of $h$ satisfying (3.16), provided the parameter $c$ satisfies the inequality.
\[
2 < c \leq 18
\] ...(3.17)

In an analogous manner we can also prove that the matrix $J(c)$ defined by (3.12) is irreducible and monotone. The fact that the matrices $A(c)$ and $J(c)$ are both irreducible and monotone will be crucial for the error estimates in our later development. We further have
\[
A(c) \geq J(c)
\]
if $c$ satisfies the inequality (3.17), hence from Theorem 2.3 it follows that
\[
0 \leq [A(c)]^{-1} \leq [J(c)]^{-1}
\] ...(3.18)
provided $h$ and $c$ satisfy (3.16) and (3.17) respectively. Now we proceed to determine $||J(c)^{-1}||$. The matrix $J(c)$ can be factored as a product of two tridiagonal matrices $J$ and $\bar{P}_N(c)$ and hence
\[
||J(c)^{-1}|| = ||[J \bar{P}_N(c)]^{-1}||
\] ...(3.19)
\[
\leq ||\bar{P}_N^{-1}(c)|| \cdot ||J^{-1}||
\]
\[
\leq ||\bar{P}_N^{-1}(c)|| \cdot (N+1)^2 / 8
\]
where $J$ is defined by (2.4). The matrix $\bar{P}_N(c)$ was introduced in Section 2.4 and shown to have an inverse such that

$$||\bar{P}_N^{-1}(c)|| \leq H_N(c)/(c-2).$$

We finally obtain

$$||J^{-1}(c)|| \leq H_N(c) (b-a)^2/[8(c-2)h^2].$$

Now going back to the error equation (3.13), we get

$$||e|| \leq ||A^{-1}(c)|| ||\Gamma(y)||$$

$$\leq H_N(c) (b-a)^2 K/8(c-2),$$

where $K$ is introduced in (3.14). We summarize these results in the following theorem.

**Theorem 3.3.**

Let $y(x) \in C^8$ be the true solution of the linear boundary value problem of class $M$ and let $y_n$ be its discrete approximation with an error $e_n$ defined by the equation (3.13). Let $A(c)$ be a monotone matrix such that $A(c) \geq J(c)$, and let $||\Gamma(y)|| \leq h^2 K$, then $||e||$ is given by (3.21).

The numerical method in which we use $\Delta.E.$ (3.8) with $T.E. = 0(h^6)$ at the interior points and $\Delta.E.'s$ (3.9) with $T.E. 0(h^6)$ near the boundaries will henceforth be referred to as $M-2(c)$. The error estimate will be based on (3.21).

3.4 AN $O(h^6)$ FINITE-DIFFERENCE ANALOGUE

Since the approximations near the boundaries in the last section were not as accurate as at the interior points, the overall accuracy was $O(h^4)$ rather than $O(h^6)$ as desired, we therefore introduce more accurate $\Delta.E.'s$ instead of (3.9) near the boundaries. We explain briefly how $\Delta.E.$ (3.9) is modified to obtain new $\Delta.E.'s$. The $T.E.$ associated with (3.9) is
0(h^6) and we want to develop new Δ.E. whose T.E. will be O(h^8). This can be done by considering a Δ.E. of order k = 3 connecting y_0, y_1, y_2, y_3 and then y_{N-2}, y_{N-1}, y_N and y_{N+1} respectively through the usual analysis in the form

\[
\begin{align*}
(i) & \quad -c y_0 + (2c-1) y_1 + (2-c) y_2 - y_3 + h^2 \sum_{i=0}^{3} \bar{e}_i y_i'' + h^3 (y_0 y_0'' + y_1 y_1'') = 0 \\
(ii) & \quad -y_{N-2} + (2-c) y_{N-1} + (2c-1) y_N - c y_{N+1} + h^2 \sum_{i=0}^{3} \bar{e}_{3-i} y_{N-2+i}'' + h^3 (y_N y_N'' + y_{N+1} y_{N+1}'') = 0,
\end{align*}
\] ...(3.22)

where

\[
\begin{align*}
\bar{e}_0 &= (58c - 57)/360 \\
\bar{e}_1 &= (91c + 16)/120 \\
\bar{e}_2 &= (10c + 115)/120 \\
\bar{e}_3 &= (-c + 24)/360 \\
y_0 &= y_3 = (2c - 3)/60 \\
y_1 &= y_2 = (c - 4)/20
\end{align*}
\]

and the T.E. associated with (3.22) is given by

\[
\text{T.E.} = \begin{cases} \\
\frac{71c - 349}{302,400} h^8 y^{(8)} , & c \neq \frac{349}{71} \\
\frac{-1.051}{4,294,080} h^9 y^{(9)} , & c = \frac{349}{71} \text{ for (3.22i)} \\
\frac{1.051}{4,294,080} h^9 y^{(9)} , & c = \frac{349}{71} \text{ for (3.22ii)} \\
\end{cases}
\] ...(3.23)

Using the D.E. viz. \( y'' = g(x) y + s(x) \), we get

\[
y''' = g'(x) y + g(x) y' + s'(x).
\] ...(3.24)

We now substitute for \( y''_0, y''_1, y''_N \) and \( y''_{N+1} \) in the Δ.E.'s (3.22) and thus these resulting equations contain terms involving \( y'_0, y'_1, y'_N \) and \( y'_{N+1} \).
In order to obtain expression for the first derivative of \( y(x) \) at \( x = x_0 \), \( x_1 \), \( x_N \) and \( x_{N+1} \) we proceed as follows:

We first consider \( \text{A.E.'s} \) of the form

\[
\begin{align*}
(i) & \quad \sum_{i=0}^{3} a_i y_i' = h y_0' + h^2 \sum_{i=0}^{3} b_i y_i'' + h^3 c_0 y_0''', \\
(ii) & \quad \sum_{i=0}^{3} a_i y_i' = h y_1' + h^2 \sum_{i=0}^{3} d_i y_i'' + h^3 c_1 y_1'''.
\end{align*}
\]

The reason for including terms like \( y''_1 \) in the above \( \text{A.E.'s} \) is that the resulting \( \text{T.E.} \) associated with them can be made at least \( O(h^7) \). The coefficients \( b \)'s, \( c \)'s, and \( d \)'s are given as follows:

\[
\begin{align*}
a_0 &= \lambda + 2\mu - 1, \\
a_1 &= -2\lambda - 3\mu + 1, \\
a_2 &= \lambda, \\
a_3 &= \mu.
\end{align*}
\]

where \( \lambda \) and \( \mu \) are arbitrary real constants. Please note that the coefficients \( a \)'s satisfy the system

\[
\begin{align*}
\sum_{i=0}^{3} a_i + a_1 + a_2 + a_3 &= 0, \\
a_1 + 2a_2 + 3a_3 &= 1.
\end{align*}
\]

\[
\begin{align*}
b_0 &= (82\lambda + 186\mu + 271)/720, \\
b_1 &= (94\lambda + 192\mu + 17)/120, \\
b_2 &= (26\lambda + 258\mu - 5)/240, \\
b_3 &= (-2\lambda + 24\mu + 1)/360.
\end{align*}
\]
\[
\begin{align*}
\begin{bmatrix}
d_0 = (8\lambda + 14\mu - 7)/120 \\
d_1 = (194\lambda + 402\mu - 97)/240 \\
d_2 = (16\lambda + 138\mu - 5)/120 \\
d_3 = (-2\lambda + 14\mu + 1)/240 \\
c_0 = (2\lambda + 6\mu + 7)/120 \\
c_1 = (-6\lambda - 18\mu + 17)/120 \\
\end{bmatrix}
\]

and the T.E. associated with (3.25) is given by

\[
T.E. = \frac{42(\lambda-2\mu) - 53}{50,400} h^7 y^{(7)}, \text{ for (3.24i)}
\]

\[
T.E. = \frac{42(2\lambda+\mu) - 53}{50,400} h^7 y^{(7)}, \text{ for (3.24ii)}
\]

Again substituting \(y'''\) and \(y''''\) (using (3.24)) in \(\Delta.E.'s\) (3.25) and solving for \(y_0'\) and \(y_1'\), we get respectively

\[
\begin{align*}
&\text{(i) } y_0' = h^{-1} (1+h^2 c_0 g_0)^{-1} \left[ \sum_{i=0}^{3} a_i y_i - h^2 \sum_{i=0}^{3} b_i y_i'' - h^3 c_0 (g_0 y_0 + s_0') \right], \\
&\text{(ii) } y_1' = h^{-1} (1+h^2 c_1 g_1)^{-1} \left[ \sum_{i=0}^{3} a_i y_i - h^2 \sum_{i=0}^{3} d_i y_i'' - h^3 c_1 (g_1 y_1 + s_1') \right].
\end{align*}
\]

In order to obtain expressions for \(y_N'\) and \(y_{N+1}'\), we consider \(\Delta.E.'s\)

\[
\begin{align*}
&\text{(i) } \sum_{i=0}^{3} a_{3-i} y_{N-2+i} = -h y_N' + h^2 \sum_{i=0}^{3} d_{3-i} y_{N-2+i} - h^3 c_1 y_N''', \\
&\text{(ii) } \sum_{i=0}^{3} a_{3-i} y_{N-2+i} = -h y_{N+1}' + h^2 \sum_{i=0}^{3} b_{3-i} y_{N-2+i} - h^3 c_0 y_{N+1}'''.
\end{align*}
\]

These \(\Delta.E.'s\) (3.27) involving \(y_N'''\) and \(y_{N+1}'''\) are deduced from the \(\Delta.E.'s\) (3.25) on replacing \(x\) by \(-x\) and then shifting the origin from \(x_{-3}\) to \(x_{N-2}\):

Again substituting \(y_N'''\) and \(y_{N+1}'''\) (using (3.24)) in \(\Delta.E.'s\) (3.27) and then solving for \(y_N'\) and \(y_{N+1}'\), we get respectively
\[
\begin{align*}
(i) & \quad y'_N = h^{-1} \left( 1 + h^2 c_1 g_N \right)^{-1} \left[ - \sum_{i=0}^{3} a_{3-i} y_{N-2+i} + h^2 \sum_{i=0}^{3} d_{3-i} y''_{N-2+i} 
\quad - h^3 c_1 \left( \varepsilon_N y'_{N-2} + s_N \right) \right], \\
(ii) & \quad y'_{N+1} = h^{-1} \left( 1 + h^2 c_0 g_{N+1} \right)^{-1} \left[ - \sum_{i=0}^{3} a_{3-i} y_{N+2+i} + h^2 \sum_{i=0}^{3} b_{3-i} y''_{N+2+i} 
\quad - h^3 c_0 \left( \varepsilon'_{N+1} y''_{N+2} + s'_{N+1} \right) \right].
\end{align*}
\]...

Thus the A.E.'s (3.22i) and (3.22ii) are used at the two boundaries respectively, after \( y''_0, y''_1, y''_N \) and \( y'''_{N+1} \) have been replaced using (3.24) and then the corresponding first derivative terms viz. \( y'_0, y'_1, y'_N \) and \( y'_{N+1} \) have been substituted using (3.26) and (3.28) respectively. The optimum values of \( \lambda \) and \( \mu \) can be obtained by solving the system

\[
\begin{align*}
42 \left( \lambda - 2\mu \right) &= 53, \\
42 \left( 2\lambda + \mu \right) &= 53, 
\end{align*}
\]...

because the T.E. associated with the A.E.'s (3.25) and (3.27) will be least in that case. Solving (3.29), we get

\[ \lambda = 53/70, \quad \mu = -53/210, \]

and with these choices of \( \lambda \) and \( \mu \) the T.E. associated with the A.E.'s (3.25) and (3.27) becomes \( O(h^8) \).

If we now make use of these new modified A.E.'s at the boundaries which have been obtained by overdifferentiation and A.E. (3.8) at the interior points, then the linear boundary value problem of class M is replaced by

\[
\bar{A}(c) y = \bar{b}, \]

where \( \bar{A}(c) \) differs from \( A(c) \) given by (3.11) in the first and the last row only. The elements of \( \bar{A}(c) \) that are different from the corresponding
elements of \( A(c) \) are given below, neglecting terms \( O(h^3) \),

\[
\begin{align*}
\bar{a}_{11} &= 2c-1 + h^2 \bar{\beta}_1 g_1 + a_1 (T_{0,0} + T_{1,1}) \\
\bar{a}_{12} &= 2c-1 + h^2 \bar{\beta}_2 g_2 + a_2 (T_{0,0} + T_{1,1}) \\
\bar{a}_{13} &= -1 + h^2 \bar{\beta}_3 g_3 + a_3 (T_{0,0} + T_{1,1}) \\
\bar{a}_{N,N-2} &= -1 + h^2 \bar{\beta}_3 g_{N-2} + a_3 (T_{1,N} + T_{0,N+1}) \\
\bar{a}_{N,N-1} &= 2c-1 + h^2 \bar{\beta}_2 g_{N-1} + a_2 (T_{1,N} + T_{0,N+1}) \\
\bar{a}_{N,N} &= 2c-1 + h^2 \bar{\beta}_1 g_N + a_1 (T_{1,N} + T_{0,N+1}),
\end{align*}
\]

where \( T_{i,j} = h^2 \gamma_i g_j (1 + h^2 c_i g_j)^{-1} \).

We would like to establish criteria for \( \bar{A}(c) \) to be irreducible and monotone such that \( \bar{A}(c) > J(c) \). We find that the matrix \( \bar{A}(c) \) defined by (3.30) is irreducible if \( \bar{a}_{12} \neq 0 \), \( \bar{a}_{N,N-1} \neq 0 \) and \( (2-c) + h^2 \beta_j g_M \neq 0 \) \((j = 1, 3)\). These conditions will be satisfied for \( c > 2 \) whenever

\[
\begin{cases}
0 < h^2 \bar{\beta}_2 g_2 + a_2 (T_{0,0} + T_{1,1}) < c-2 \\
0 < h^2 \bar{\beta}_2 g_{N-1} + a_2 (T_{1,N} + T_{0,N+1}) < c-2 \\
h^2 \beta_j g_M < c-2 \quad (j=1,3)
\end{cases} \tag{3.31}
\]

Since \( \beta_j > 0 \) for all values of \( c \) satisfying (3.17), the last condition is satisfied for sufficiently small values of \( h \). We now proceed to examine the first two conditions of (3.31). We have seen above that \( \lambda = 53/70 \) and \( \mu = -53/210 \). With this choice of \( \lambda \) and \( \mu \) we find

\[
\begin{align*}
a_1 &= 17/70, \quad a_2 = 53/70, \quad a_3 = -53/210; \\
c_0 &= 7/120, \quad c_1 = 17/120.
\end{align*}
\]

The coefficients \( \bar{\beta} \)'s and \( \gamma_0 \) are positive for any value of \( c \) satisfying (3.17). However \( \gamma_1 = (c-4)/20 < 0 \), for \( c \) satisfying \( 2 < c < 4 \). We now consider
\[ h^2 \tilde{g}_2 g_2 + a_2 (T_{0,0} + T_{1,1}) = h^2 \tilde{g}_2 g_2 + a_2 h^2 \left[ \gamma_0 g_0 \left( 1 + \frac{7h^2}{120} g_0 \right)^{-1} + \gamma_1 g_1 \left( 1 + \frac{17h^2}{120} g_1 \right)^{-1} \right] \\
= h^2 \tilde{g}_2 (g_1 + h g_1 + \ldots) + a_2 h^2 [\gamma_0 (g_1 h g_1 + \ldots) + \gamma_1 g_1] \\
= h^2 g_1 \left[ \frac{10c+115}{120} + \frac{53}{70} \left( \frac{2c-3}{60} + \frac{c-4}{20} \right) \right] \\
= \frac{123c+646}{840} h^2 g_1 ,
\]

Similarly
\[ h^2 \tilde{g}_2 g_{N-1} + a_2 (T_{1,N} + T_{0,N+1}) = h^2 \tilde{g}_2 g_{N-1} + a_2 h^2 \left[ \gamma_1 g_N \left( 1 + \frac{7h^2}{120} g_N \right)^{-1} + \gamma_0 g_{N+1} \left( 1 + \frac{17h^2}{120} g_{N+1} \right)^{-1} \right] \\
= \frac{123c+646}{840} h^2 g_N ,
\]

neglecting all terms \( O(h^3) \).

Now it is evident that the first two conditions (3.31) will be satisfied if
\[
0 \leq \frac{123c + 646}{840} h^2 g_m < c-2 ,
\]
and this can be forced to be satisfied for sufficiently small values of \( h \).

Now we want the matrix \( A(c) > J(c) \) such that \( A^{-1}(c) \leq J^{-1}(c) \). This will be the case if the following six conditions are satisfied, namely
\[
\begin{align*}
  h^2 \tilde{g}_1 g_1 + a_1 (T_{0,0} + T_{1,1}) & \geq 0 , \\
  h^2 \tilde{g}_{4-i} g_{N-3+i} + a_{4-i} (T_{1,N} + T_{0,N+1}) & \geq 0 , \\
(i = 1,2,3).
\end{align*}
\]

Again through the usual analysis employed above in establishing the first two conditions (3.31) we prove that the conditions (3.32) will be satisfied if
\[
\begin{align*}
(i) & \quad \frac{654c + 61}{840} h^2 g_m > 0 \\
(ii) & \quad \frac{123c + 646}{840} h^2 g_m > 0 \quad \cdots (3.33) \\
(iii) & \quad \frac{-60c + 327}{2520} h^2 g_m > 0 ,
\end{align*}
\]
at least for the values of \( c \) satisfying \( 2 < c < 14/3 \). But the above three conditions are obviously satisfied for those values of \( c \). Thus the conditions (3.32) are satisfied at least for \( c \) satisfying \( 2 < c < 14/3 \), and for sufficiently small values of \( h \).

Having proved that \( \bar{A}(c) \) is irreducible and monotone such that \( \bar{A}(c) > J(c) \) and \( \bar{A}^{-1}(c) < J^{-1}(c) \), we obtain the error equation in the following form:

\[
\tilde{A}(c) e = \bar{\Gamma}(y),
\]

where

\[
\bar{\Gamma}(y) = h^8 \begin{bmatrix}
\frac{71c-349}{302,400} y(8) (\xi_1) \\
- \frac{31c-190}{60,480} y(8) (\xi_2) \\
\vdots \\
- \frac{31c-190}{60,480} y(8) (\xi_{N-1}) \\
\frac{71c-349}{302,400} y(8) (\xi_N)
\end{bmatrix}
\]

for any value of \( c \) satisfying \( 2 < c < 14/3 \), and \( ||\bar{\Gamma}(y)|| \leq h^2 \bar{K} \).

Here \( \bar{K} = \max. \ h^6 M_8 \begin{bmatrix} 71c-349 \\ 302,400 \\
31c-190 \\
60,480
\end{bmatrix} , c \neq \frac{349}{71} \text{ or } \frac{190}{31} \). One can easily show

\[
\bar{K} = \begin{cases} 
\frac{71c-349}{302,400} h^6 M_8 , & 5.75 = \frac{1299}{266} \leq c \leq \frac{601}{84} = 7.15 \\
\frac{31c-190}{60,480} h^6 M_8 , & \text{otherwise.}
\end{cases}
\]

Thus for \( c \) satisfying \( 2 < c < 14/3 \), \( \bar{K} \) turns out to be equal to \( \frac{190-31c}{60,480} h^6 M_8 \). Now substituting \( \bar{K} \) instead of \( K \) in (3.21), we get

\[
||e|| \leq \frac{h^6 (190-31c)(b-a)^2 M_8 H_N(c)}{483,840 (c-2)} , \quad 2 < c < 14/3 , \quad \ldots (3.35)
\]
or equivalently

\[ ||e|| = O(h^6). \]

The above error bound is analogous to the one obtained earlier (Theorem 3.3). The method described above will henceforth be referred to as M-3(c). The resulting error is based on (3.35).

Recently an \( O(h^4) \) finite difference analogue was formulated by Bramble and Hubbard [3]. The coefficient matrix was based on a five point A.E.

(i) \( h^2 y''_n = \frac{1}{12} (-y_{n-2} + 16 y_{n-1} - 30 y_n + 16 y_{n+1} - y_{n+2}) \),

at the interior points and by a three point A.E.

(ii) \( h^2 y''_n = y_{n-1} - 2 y_n + y_{n+1} \) \hspace{1cm} ...(3.36)

at the boundary points of the range of integration.

In fact the problem considered by Bramble and Hubbard was the one given by (2.19). The main result of their paper was that the discretization error satisfies the inequality

\[ ||e|| \leq h^4 \left[ \frac{M_4}{2} + \frac{M_6}{720} \right]. \]  

...(3.37)

In our case \( ||e|| \) based on five point A.E. (3.8) is given by (3.21), which turns out to be a better bound than the one given by (3.37). We postpone comparison of the method proposed by Bramble and Hubbard (henceforth referred to as M-BH) with other methods till the last chapter.
CHAPTER IV
A CLASS OF HIGH ACCURACY DIFFERENCE FORMULAS OF
LOW ORDER

4.1

In the present chapter we consider ΔE.'s of order k = 2 which are of the form

\[ \Sigma a_i y_{n+i} = h^2 \Sigma \beta_i y''_{n+i} + h^3 \Sigma \gamma_i y'''_{n+i} + \ldots + h^t \Sigma \theta_i y^{(t)}_{n+i}, \quad \ldots(4.1) \]

where \( t \) is a positive integer \( \geq 3 \). In this chapter all summations extend from \( i = 0 \) to \( i = 2 \). Since \( y'' = g(x) y + s(x) \), in general \( y^{(n)}(x) \), \( n \geq 3 \), cannot be expressed in terms of \( y(x) \) and higher derivatives of \( g(x) \) and \( s(x) \) alone. The derivatives of \( y(x) \) higher than 2 will contain \( y'(x) \).

Therefore we need expressions for \( y'_{n+i}, i = 0, 1, 2 \). These expressions can be obtained by considering auxiliary ΔE.'s of the form

\[ \Sigma a_i y_{n+i} = h y'_{n+m} + h^2 \Sigma b_i y''_{n+i} + \ldots + h^t \Sigma \phi_i y^{(t)}_{n+i}, \quad \ldots(4.2) \]

\((m = 0, 1, 2)\). In general the coefficients \( b \)'s, \( c \)'s, \ldots etc. depend on \( m \).

We obtain by successive differentiation of the D.E. viz. \( y'' = g(x) y + s(x) \) the expressions for higher derivatives \( y''' \), \( y^{(4)} \), \ldots etc.

Assume

\[ y^{(t)}(x) = \theta^t(x) y + \phi^t(x) y' + \psi^t(x), \quad t \geq 3 \]

\((4.3)\)

where the functions \( \theta^t(x) \), \( \phi^t(x) \) and \( \psi^t(x) \) depend on \( g(x) \), \( s(x) \) and on their higher derivatives. On substituting the expressions for higher derivatives given by (4.3) in ΔE. (4.2), we get

\[ h y'_{n} (1 + h^2 c_0 \phi^3_n + \ldots) + h^3 y'_{n+1} (c_1 \phi^3_{n+1} + \ldots) + h^3 y'_{n+2} (c_2 \phi^3_{n+2} + \ldots) + \ldots(4.4a) \]

\[ h^3 y'_{n} (c_0 \phi^3_n + \ldots) + h y'_{n+1} (1 + h^2 c_1 \phi^3_{n+1} + \ldots) + h^3 y'_{n+2} (c_2 \phi^3_{n+2} + \ldots) + \ldots(4.4b) \]
\( h^3 y'_n (c_0 \phi_n^3 + ... ) + h^3 y'_{n+1} (c_1 \phi_{n+1}^3 + ... ) + h y'_{n+2} (1+ h^2 c_2 \phi_{n+2}^3 + ... ) + \eta(m)=0 \) \hspace{1cm} (4.4c)

where \( \eta(m) = -\Sigma a_i y_{n+i} + h^2 \Sigma [ (b_i g_{n+i} + h c_i \phi_{n+i}^3 + ... ) y_{n+i} \\
+ (b_i s_{n+i} + h c_i \psi_{n+i}^3 + ... ) ] \).

We solve the above system (4.4) for the unknowns \( y'_n, y'_{n+1} \) and \( y'_{n+2} \).

However we can simplify the above system (4.4) to a very great extent if we consider the auxiliary A.E. (4.2) in the form

\[ \Sigma a_i y_{n+i} = h y'_{n+m} + h^2 \Sigma b_i y''_{n+i} + h^3 c_i \frac{d}{(4)} \]

Again on substituting in (4.5) the expressions for higher derivatives of \( y(x) \) as given by (4.3) and solving it for \( y'_{n+m} \), we get

\[ h y'_{n+m} = -(1+h^2 \frac{d}{(3)} + h^3 \frac{d}{(4)} + ... )^{-1} \eta_{n+m} \]

where \( \eta_{n+m} = -\Sigma a_i y_{n+i} + h^2 \Sigma (b_i g_{n+i} + s_{n+i}) \\
+ h^3 (c_i \theta_{n+m}^3 + h d_m \phi_{n+m}^4 + ... ) y_{n+m} \\
+ h^3 (c_i \psi_{n+m}^3 + h d_m \psi_{n+m}^4 + ... ) \), \( m = 0,1,2 \).

We now obtain expressions for higher derivatives \( y''_{n+i}, y''_{n+i} \) etc. in the form, (using (4.3) and (4.6)),

\[ y^{(t)}_{n+m} = \theta_{n+m}^{(t)} y_{n+m} - \phi_{n+m}^{(t)} h^{-1} (1+h^2 c_m \phi_{n+m}^3 + h^3 d_m \phi_{n+m}^4 + ... )^{-1} \eta_{n+m} + \psi_{n+m}^{(t)} \hspace{1cm} (m = 0,1,2), \ t \geq 3 \]

We substitute the expressions for higher derivatives \( y^{(t)}(x), \ t \geq 3 \) as given by (4.7) into the A.E. (4.1) and get the required A.E. to be used in the numerical solution of the linear boundary value problem of class M.

In what follows we will consider the A.E. (4.1) for \( t = 3 \) and 4 in detail and the auxiliary A.E.'s (4.5) for \( \bar{t} = 2 \) and 3 respectively.
It is proved that the resulting error based on these methods is $O(h^q)$, $q = 5, 6, 7$ and 8.

4.2 **THE CASE FOR $t = 3$**

It can be shown (as in (2.3)) that the Δ.E. (4.1) for any positive integer $m$ is necessarily of the form

$$y_n - 2y_{n+1} + y_{n+2} = h^2 \sum \beta_i y''_{n+i} + \ldots + h^t \sum \theta_i y^{(t)}_{n+i}, \quad \ldots (4.8)$$

A special case of (4.8) for $t = 3$ is

$$y_n - 2y_{n+1} + y_{n+2} = h^2 \sum \beta_i y''_{n+i} + h^3 \sum \beta_i y'''_{n+i}, \quad \ldots (4.9)$$

where

$$\beta_0 = \beta_2 = 2/15$$
$$\beta_1 = 11/15$$
$$\gamma_0 = -\gamma_2 = 1/40$$
$$\gamma_1 = 0$$

and T.E. of (4.9) is given by

$$\text{T.E.} = \frac{29}{302,400} \cdot h^8 y^{(8)}.$$  

Since $\gamma_1 = 0$, therefore we only require auxiliary Δ.E.'s of the form (4.5) for $m = 0$ and 2. We first consider (4.5) for $\bar{t} = 2$, $m = 0$ namely

$$\sum a_i y_{n+i} = h y'_{n} + h^2 \sum b_i y''_{n+i}. \quad \ldots (4.10)$$

Using Taylor's formula for expanding $y_{n+i}$ and $y''_{n+i}$ around $x_n$, one can show that $a$'s used in the Δ.E. (4.10) satisfy the following system:

$$\begin{bmatrix}
    a_0 + a_1 + a_2 = 0 \\
    a_1 + 2a_2 = 1 
\end{bmatrix} \quad \ldots (4.11)$$

The general solution of the system (4.11) is given by

$$a_0 = \lambda - 1, \quad a_1 = -2\lambda + 1, \quad a_2 = \lambda, \quad \ldots (4.12)$$

where $\lambda$ is an arbitrary real number. Using the above values of $a_0$, $a_1$ and $a_2$ we find through the usual analysis
\[ b_0 = \frac{(2\lambda+7)}{24},\ b_1 = \frac{(10\lambda+3)}{12}\ and\ b_2 = \frac{(2\lambda-1)}{24}, \]

and T.E. associated with (4.10) is given by
\[ \text{T.E.} = \frac{1}{45} h^5 y^{(5)}. \]

We might remark that the D.E.'s (4.9) and (4.10) are both such that the operators associated with them are definite and the generalized mean value theorem holds.

The D.E. (4.5) for \( t = 2 \) and \( m = 2 \) is symmetric to the D.E. (4.10) and can be obtained by replacing \( x \) by \(-x\) or by flipping the interval \([x_n, x_n + 2h]\) in the form
\[ \sum a_{2-1} y_{n+i} = -h y'_{n+2} + h^2 \sum b_{2-1} y''_{n+i}. \hspace{1cm} \ldots(4.13) \]

The T.E. of (4.13) is given by
\[ \text{T.E.} = -\frac{1}{45} h^5 y^{(5)}. \]

On substituting the value of \( y''_{n+1} \) using the D.E. in (4.10) and (4.13) and solving for \( y'_{n} \), \( y'_{n+2} \) respectively, we get
\[ \begin{cases} (i) \ h y'_{n} = \sum a_{i} y_{n+i} - h^2 \sum b_{i} (g_{n+i} y_{n+i} + s_{n+i}) \\ (ii) \ h y'_{n+2} = \sum a_{2-1} y_{n+i} + h^2 \sum b_{2-1} (g_{n+i} y_{n+i} + s_{n+i}) \end{cases} \hspace{1cm} \ldots(4.14) \]

Similarly on substituting \( y''_{n+1} \), \( y''_{n+2} \) (using D.E.) in the D.E. (4.9) and then substituting for \( h y'_{n} \), \( h y'_{n+2} \) using (4.14), we get the required D.E. which is used to approximate the boundary value problem in the form
\[ (-1 + A_n) y_n + (2 + B_{n+1}) y_{n+1} + (-1 + C_{n+2}) y_{n+2} + d_{n+1} = 0, \hspace{1cm} \ldots(4.15) \]

\[(n = 0, 1, 2, \ldots, N-1), \]

where
\[ A_i = h^2 (\beta_{0} g_i + \gamma_{0} a_{0} g_i - \gamma_{2} a_{2} g_{i+2}) \ (i = 0, 1, \ldots, N-1), \]
\[ B_i = h^2 (\beta_{1} g_i + \gamma_{0} a_{i-1} - \gamma_{2} a_{2} g_i) \ (i = 1, 2, \ldots, N), \]
\[ C_i = h^2 (\beta_{2} g_i + \gamma_{0} a_{i-2} - \gamma_{2} a_{2} g_i) \ (i = 2, 3, \ldots, N+1), \]

plus terms \( O(h^3) \).
Here $\beta_0 = 2/15$, $\beta_1 = 11/15$, $\beta_2 = 2/15$, $\beta_0 = -\gamma_2 = 1/40$, and $a_0 = \lambda - 1$, $a_1 = -2\lambda + 1$, $a_2 = \lambda$.

Thus
\[ A_i = h^2(\beta_0 g_i + \gamma_0 a_0 g_i - \gamma_2 a_2 g_i + 2h g_i + \ldots)) \]
\[ = h^2 g_i (\beta_0 + \gamma_0 a_0 - \gamma_2 a_2) \]
\[ = h^2 g_i \left( \frac{2}{15} + \frac{\lambda - 1 + \lambda}{40} \right) \]
\[ = \frac{6\lambda + 13}{120} h^2 g_i > 0 \text{ for } \lambda > -13/6. \]

Similarly
\[ B_i = \frac{47 - 6\lambda}{60} h^2 g_i > 0 \text{ for } \lambda < 47/6 \]
and
\[ C_i = \frac{6\lambda + 13}{120} h^2 g_i > 0 \text{ for } \lambda > -13/6, \]
neglecting the terms $O(h^3)$.

Thus for sufficiently small values of $h$, $A_i$'s, $B_i$'s and $C_i$'s are positive and also $A_i$'s, $C_i$'s are each less than unity provided
\[
\begin{align*}
(i) & \quad -\frac{13}{6} < \lambda < \frac{47}{6} \\
(ii) & \quad 0 < \frac{6\lambda + 13}{120} h^2 g_i < 1. 
\end{align*}
\]

Now the associated matrix with the system of linear equations (4.15) has all the nice properties given by (2.8) provided the conditions (4.16) are satisfied. We can now obtain an error bound of the form (2.9). In fact we get
\[
|e_i| \leq \frac{h^5 M_5 g_M (x_i-a)(b-x_i)}{1,800} \\
||e|| \leq h^5 M_5 g_M (b-a)^2/7,200 ,
\]
or equivalently $||e|| = O(h^5)$.

In this case we will not be able to obtain the more accurate error bound of the form (2.15) for the simple reason that we will not be
able to associate with the matrix $A$, the matrix $A_m$ (defined in Section 2.4) with the desired properties viz. $A^2 = A_m$ and $A_m$ is symmetric and can be factored as a product of $P_N(u)$ and $Q$.

4.3 **MORE ACCURATE EXPRESSIONS FOR $y'_{n+m}$** ($m = 0, 1, 2$)

We have seen that although the $\Delta.E.$ (4.5) is such that its T.E. $= O(h^8)$, yet the error resulting from it is $O(h^5)$. The reason is that the T.E. of the auxiliary $\Delta.E.'s$ is $O(h^5)$, and when we substitute $y'_n$ and $y'_{n+2}$ using (4.14) in (4.9), the resulting $\Delta.E.$ turns out to have a T.E. $O(h^7)$. Thus the accuracy of the numerical procedure discussed above mainly depends on the accuracy of the auxiliary $\Delta.E.'s$. Therefore we would like to obtain more accurate auxiliary $\Delta.E.'s$. Hence we now consider (4.5) with $t = 3$ and $m = 0$ namely

$$\Sigma a_i y_{n+i} = h y'_n + h^2 \Sigma b_i y''_{n+i} + h^3 c_0 y'''_n,$$

... (4.18)

where $a's$ are given by (4.12) and $b's$ are given as follows:

$$b_0 = (10\lambda + 47)/120$$
$$b_1 = (50\lambda + 7)/60$$
$$b_2 = (10\lambda - 1)/120$$
$$c_0 = 1/15$$

and T.E.

$$\text{T.E.} = \begin{cases} 
\frac{-2\lambda}{480} h^6 y^{(6)}, & \lambda \neq 1/2 \\
- \frac{1}{1,575} h^7 y^{(7)}, & \lambda = 1/2 
\end{cases}$$

We similarly develop a $\Delta.E.$ containing $h y'_{n+1}$ in the form

$$\Sigma a_i y_{n+i} = h y'_{n+1} + h^2 \Sigma d_i y''_{n+i} + h^3 c_1 y'''_{n+1},$$

... (4.19)

where

$$d_0 = (5\lambda - 4)/60$$
$$d_1 = 5(2\lambda - 1)/12$$
\[ d_2 = (5\lambda - 1)/60 \]
\[ c_1 = 7/60 \]

and
\[ T.E. = \begin{cases} 
\frac{1-2\lambda}{480} h^6 y^{(6)}, & \lambda \neq 1/2 \\
-\frac{11}{50,400} h^7 y^{(7)}, & \lambda = 1/2.
\end{cases} \]

However \( \Delta.E. \) (4.5) for \( t = 3 \) and \( m = 2 \) can be obtained using \( \Delta.E. \) (4.18).

We replace \( x \) by \(-x\), \( y' \) by \(-y'\), \( y'' \) by \(-y''\) in (4.18), rearrange the suffixes and get
\[ \Delta\alpha \ y_{n+i} = -h y'_{n+2} + h^2 \Sigma b_{2-i} y''_{n+i} - h^3 c_0 y'''_{n+2}, \quad \ldots \text{(4.20)} \]

where the T.E. of (4.20) is the same as that of (4.18) when \( \lambda \neq 1/2 \), and is of opposite sign for \( \lambda = 1/2 \). Now following the analysis used in the previous section, we can prove that the resulting error in the boundary value problem is \( O(h^6) \), provided the conditions (4.16) are satisfied.

This result will also be a special case of the Theorem 2.4 for \( p = 6 \) and \( M = 29 M_e/302,400 \). The inequality (4.21) gives the error bound based on this method.

\[ ||e|| \leq \frac{29 h^6 M_e (b-a)^2}{2,419,200} \quad \ldots \text{(4.21)} \]

or \( e = O(h^6) \). The more accurate error bound of the form (2.15) does not apply here for the same reasons mentioned in the previous section. The numerical method in which we use \( \Delta.E. \) (4.9) and auxiliary \( \Delta.E.'s \) (4.18) and (4.20) will henceforth be referred to as M-4. The error estimate is given by the inequality (4.21).

4.4 THE CASE FOR \( t = 4 \)

We now consider (4.8) for \( t = 4 \) namely
\[ y_n - 2y_{n+1} + y_{n+2} = h^2 \Sigma b_{i} y''_{n+i} + h^3 \Sigma y_{i} y''_{n+i} + h^4 \Sigma \delta_{i} y^{(4)}_{n+i}, \quad \ldots \text{(4.22)} \]
where \( \beta_2 = 912Q \),
\( \beta_2 = 912Q \),
\( \gamma_2 = 177Q \),
\( \gamma_2 = 177Q \),
\( \delta_2 = 11Q \),
\( \delta_2 = 11Q \),
and \( Q = 1/10,080 \).

The T.E. of (4.22) is given by
\[
T.E. = \frac{-17}{5,588,352,000} h^{12} y(12).
\]

If we use D.E. to substitute expressions for \( y'' \), \( y''' \) and \( y^{(4)} \) in (4.22) and then substitute expressions for \( h y'_{n+m} \) \( (m = 0,1,2) \) obtained from auxiliary A.E.'s (4.18), (4.19) and (4.20), we get the required A.E. The T.E. of this A.E. is \( O(h^9) \), again limited by the accuracy of the expressions approximating \( y'_{n+m} \) \( (m = 0,1,2) \). The matrix associated with the system of linear equations will be monotone and \( >J \) provided \( \lambda \) and \( h \) are chosen so as to satisfy the following conditions.

\[
\begin{align*}
(i) \quad -245/118 &< \lambda < 1435/118 \\
(ii) \quad 0 &< \frac{118\lambda + 245}{3,360} h^2 s_M < 1
\end{align*}
\]

We omit the details and state the error bound obtained for a linear boundary value problem of class M below in (4.23).

\[
||e|| \leq \frac{177 h^7 M \delta_M (b-a)^2}{63,504,000}
\]

or \( e = O(h^7) \). Please note that the error estimate given by (4.23) is a special case of the theorem 2.4 for \( p = 7 \) and \( M = 177 M \delta_M / 7,938,000 \). The more accurate error bound of the form (2.15) does not apply here as well.
4.5 \textit{AN O}\(h^8\) \textit{FINITE DIFFERENCE ANALOGUE}

Another special case of A.E. (4.8) will be
\[ y_n - 2y_{n+1} + y_{n+2} = h^2 \sum \beta_i y_{n+i} + h^4 \sum \gamma_i y_{n+i}^{(4)}, \quad \ldots \tag{4.24} \]
where
\[ \beta_0 = \beta_2 = 660Q \]
\[ \beta_1 = 13,800Q \]
\[ \gamma_0 = \gamma_2 = -13Q \]
\[ \gamma_1 = 626Q \]
\[ 15,120Q = 1 \]
and the T.E. associated with the above A.E. is given by
\[ \text{T.E.} = \frac{-59 h^{10} y^{(10)}}{76,204,800}. \]

The operator associated with the A.E. (4.24) can be easily shown to be definite. Again using the D.E. and the auxiliary A.E.'s (4.18), (4.19) and (4.20), we can obtain expressions for \(y_{n+m}^{(m = 0,1,2)}\). Replacing \(y''\) and \(y^{(4)}\) in (4.24) by using D.E. and then substituting explicit expressions for \(y_{n+m}^{(m)}\), the required A.E. is obtained with a T.E. \(O(h^{10})\).

Here the D.E. is approximated by a system of linear equations of the form (4.15) where now
\[ A_i = h^2 \beta_i + O(h^3), \]
\[ B_i = h^2 \beta_i + O(h^3), \]
and \[ C_i = h^2 \beta_i + O(h^3). \]
Since \(\beta_i > 0\), therefore \(A_i, B_i, C_i > 0\) and if \(h\) be chosen so that
\[ 0 < h^2 \beta_i g_i < 1 \quad (i = 0,2), \]
then the associated matrix \(A\) will be monotone and \(> J\).

As before, for sufficiently small values of \(h\) we state the error bound obtained in the form
\[ ||e|| \leq \frac{h^8 (590 M_{10} + 15,436 M_7 g_M^1) (b-a)^2}{6,096,384,000} \]  

...(4.25)

The method given above will henceforth be referred to as M-5, where the error estimate is given by (4.25).

We have seen that the accuracy of the numerical method mainly depends on the accuracy of the auxiliary Δ.E.'s of the form (4.5). We have considered (4.5) for \( \bar{t} = 2, \) and 3 in the previous discussion. By choosing Δ.E.'s (4.5) for \( \bar{t} > 3, \) and combining them with a suitable Δ.E. of the form (4.8), we can produce numerical methods in which \( e = O(h^r), \) \( r \geq 9. \) The cases in which \( r = 5, 6, 7 \) and 8 have been discussed in this chapter in detail.

4.6 LINEAR BOUNDARY VALUE PROBLEMS OF CLASS M, WHERE \( g(x) = \text{Constant} \)

We now consider a special class of two point boundary value problem of class M of the form (1.1), where

\[
f(x,y) = a \cdot y + s(x).
\]

Here \( a \) is a real positive arbitrary constant. For this class of problems extremely accurate finite difference analogues can be developed by following the technique, discussed in this chapter, without using any auxiliary Δ.E.'s of the form (4.5). The reason for this is the following:

From

\[
y(2v) = a^v y + \sum_{m=1}^{v} a^{v-m} s(2m-2)(x),
\]

and

\[
y(2v+1) = a^v y' + \sum_{m=1}^{v} a^{v-m} s(2m-1)(x),
\]

we notice that the higher derivatives of \( y \) of even order do not contain \( y' \), hence Δ.E. (4.8) can be modified in the form

\[
y_n - 2y_{n+1} + y_{n+2} = h^2 \sum \beta_i y_{n+i} + h^4 \sum \gamma_i y_n^{(4)} + \ldots + h^t \sum \theta_i y_n^{(2t)}, \ldots (4.26)
\]

where \( t \) is a positive integer \( >1. \) We will consider two cases of (4.26)
in detail namely for \( t = 2 \) and \( t = 3 \). The A.E. (4.24) already considered is a special case of (4.26) for \( t = 2 \). We will not repeat it here again. However, if we approximate the two point boundary value problem of class M with constant coefficient by A.E. (4.24), then the resulting error is \( O(h^8) \). In fact the error bound obtained is as follows:

\[ ||e|| \leq \frac{59 h^8 M_{10} (b-a)^2}{609,638,400} \]  

The error bound for linear problem with constant coefficient obtained here is a special case of (2.9) for \( p = 8 \) and \( M = 59 M_{10}/76,204,800 \). In this case however, we can obtain a more precise error bound using the theory developed in Section 2.4. Without giving the computational details, we state a more precise bound of the form (2.15) below:

\[ ||e|| \leq \frac{59 h^8 M_{10} H_N(\mu)}{76,204,800 \alpha} \]  

where \( \mu = (2+h^2 \alpha \beta_1 + h^4 \gamma_0)/(1-h^2 \alpha \beta_0 - h^4 \alpha^2 \gamma_0) > 2 \) for all sufficiently small values of \( h > 0 \).

We now consider (4.26) for \( t = 3 \) viz.

\[ y_n - 2y_{n+1} + y_{n+2} = h^2 \sum \beta_i y_{n+i} + h^4 \sum \gamma_i y_{n+i} + h^6 \sum \delta_i y_{n+i}^6, \]  

where the coefficients \( \beta 's, \gamma 's, \) and \( \delta 's \) are obtained through the usual analysis and are given by

\[
\begin{align*}
\beta_0 &= \beta_2 = 1,154,160Q \\
\beta_1 &= 36,943,200Q \\
\gamma_0 &= \gamma_2 = -16,632Q \\
\gamma_1 &= 2,150,064Q \\
\delta_0 &= \delta_2 = 127Q \\
\delta_1 &= 29,230Q
\end{align*}
\]

and \( 39,251,520Q = 1 \).
Also the T.E. associated with (4.28) is given by

\[ T.E. = \frac{-45,469}{13! \times 272,580} \cdot h^{14} y(14). \]

Then since the \( \beta \)'s are positive, it follows from (2.9) that

\[ \|e\| \leq \frac{45,469 h^{12} M_{14} (b-a)^{2}}{13! \times 2,180,640}, \]

...(4.30)

for all sufficiently small values of \( h \).

In this case as well we derive a more accurate error bound.

It follows from (2.15) that

\[ \|e\| \leq \frac{45,469 h^{12} M_{14} H_{N}(u)}{13! \times 272,580} \alpha, \]

...(4.31)

where

\[ \mu = (2+h^{2} a \beta_{1} + h^{4} a^{2} \gamma_{1} + h^{6} a^{3} \delta_{1})/(1-h^{2} a \beta_{0} - h^{4} a^{2} \gamma_{0} - h^{6} a^{3} \delta_{0}). \]

The numerical method based on A.E. (4.29) will henceforth be referred to as M-6, where the error estimate is based on (4.31).
CHAPTER V

LINEAR BOUNDARY PROBLEMS WHERE \( \frac{\partial f(x,y)}{\partial y} < 0 \)

5.1

We know that the solution of the boundary value problem of class M namely

\[ y'' = f(x,y), \ y(a) = y_a, \ y(b) = y_b \]

is unique, provided the function \( f(x,y) \), in addition to satisfying the usual conditions of the existence theorems is such that \( \frac{\partial f(x,y)}{\partial y} > 0 \) in \( S \).

However if \( \frac{\partial f(x,y)}{\partial y} < 0 \), then the solution of the above boundary value problem is not necessarily unique. For instance, the boundary value

\[ y'' + \pi^2 y = 0, \ y(0) = y(1) = 0 \]

has \( y(x) = C \sin x \) as a solution for arbitrary values of \( C \). Thus this problem has no unique solution. On the other hand, the solution of

\[ y'' + \pi^2 y = x + \pi^2 \frac{x^3}{6}, \ y(0) = 0, \ y(1/2) = 1/48 \]

is unique. In fact

\[ y'' + \lambda^2 y = g(x), \ y(a) = y(b) = 0 \]

has a unique solution whenever \( \lambda \neq n\pi/(b-a), \ n = \pm 1, \pm 2, \ldots \). The purpose of this chapter is to obtain error bounds for such problem where the solution is unique.

5.2

Let us consider the problem in the form

\[ y'' = -g(x) y + s(x), \ g(x) > 0 \ in \ S, \quad \ldots (5.1) \]

\[ y(a) = y_a, \ y(b) = y_b. \]

We assume that the above boundary value problem (5.1) has a unique solution. We then replace the D.E. (5.1) by the Δ.E. (2.3) namely

\[ y_n - 2y_{n+1} + y_{n+2} = \frac{h^2}{12} \left( \beta_0 y''_n + \beta_1 y''_{n+1} + \beta_2 y''_{n+2} \right) \]
which has a truncation error determined through
\[ T.E. = C_{p+2} h^{p+2} y^{(p+2)}(\xi), \quad \xi \in [x_n, x_{n+2}], \]
where \( \beta_0 + \beta_1 + \beta_2 = 1, \beta_0 = \beta_2, \beta_j \geq 0 \) \((j = 0, 1, 2)\). We have already shown that the operator associated with the above \( \Delta E \) is definite, see Section 2.1. The error equation is obtained in a usual manner and can be put in the form
\[ \Delta e = \Gamma(y) \quad \ldots (5.2) \]
where now
\[
A = \begin{bmatrix}
A_1 & B_2 \\
B_1 & A_2 & B_3 \\
& & \ddots & \ddots \\
& & & \ddots & \ddots \\
& & & & & B_{N-2} & A_{N-1} & B_N \\
& & & & & B_{N-1} & A_N
\end{bmatrix},
\]
so that
\[
A_n = 2 - h^2 \beta_1 g_n,
\]
\[
B_n = -1 - h^2 \beta_0 g_n, \quad (n = 1, 2, \ldots, N)
\]
and
\[ ||\Gamma(y)|| = |C_{p+2}| h^{p+2} M_{p+2}. \]

We shall at first consider a special case of D.E. (5.1) in which \( g(x) = \text{constant} = g \), say. Using the results of the section 2.4, we get
\[ ||A^{-1}|| < \frac{N}{(1 + h^2 \beta_0 g) \sin \theta \sin(N+1)\theta} \quad \ldots (5.3) \]
where
\[ 2 \cos \theta = \mu = (2 - h^2 \beta_1 g)/(1 + h^2 \beta_0 g) < 2. \]

Setting \( \sin \theta = h \sqrt{g + h^2 \frac{g}{2} (\beta_0^2 - \beta_1^2/4)} / (1 + h^2 \beta_0 g) \),

we finally get
\[ ||e|| < \frac{|C_{p+2}| h^p M_{p+2} (b-a)}{\sqrt{g + h^2 \frac{g}{2} (\beta_0^2 - \beta_1^2/4)} |\sin(N+1)\theta|} \quad \ldots (5.4) \]
We state the above result in the form of a theorem as follows:

**Theorem 5.1**

Let \( y(x) \in C^{p+2} \) be the true solution of the boundary value problem \( y'' = -g \, y + s(x) \), \( y(a) = y_a \), \( y(b) = y_b \), where \( g \) is a real positive constant; also let \( y_n \) be its discrete approximation based on the \( \Delta.E. \) (2.3), then

\[
||e|| < \frac{|C_{p+2}| \, h^p \, M_{p+2} (b-a)}{\sqrt{g + h^2 g^2 (\beta_0^2 - \beta_1^2/4)} |\sin(b-a)\theta|},
\]

or

\[
||e|| = O(h^p), \text{ where } 2 \cos \theta = \frac{(2 - h^2 \beta_1 g)}{(1 + h^2 \beta_0 g)} < 2.
\]

**Particular Cases of Theorem 5.1:**

**Case (a)** - A special form of \( \Delta.E. \) (2.3) is of degree \( p = 2 \) with \( \beta_0 = \beta_2 = 0 \), \( \beta_1 = 1 \) and \( T.E. = \frac{h^4}{12} y^{(4)}(\xi) \), hence

\[
||e|| < \frac{h^2 M_4 (b-a)}{12 \sqrt{g - h^2 g^2/4} |\sin(b-a)\theta|}, \quad \ldots \quad (5.5)
\]

where \( 2 \cos \theta = 2 - h^2 g \).

**Case (b)** - Another and more accurate form of \( \Delta.E. \) (2.3) is of degree \( p = 4 \) with \( \beta_0 = \beta_2 = 1/12 \), \( \beta_1 = 10/12 \) and \( T.E. = -h^6 y^{(6)}(\xi)/240 \), hence

\[
||e|| < \frac{h^6 M_6 (b-a)}{240 \sqrt{g - h^2 g^2/6} |\sin(b-a)\theta|}, \quad \ldots \quad (5.6)
\]

where \( 2 \cos \theta = \frac{(24 - 10 h^2 g)}{(12 + h^2 g)} \).

We know that the D.E. under consideration viz.

\[
y'' = -g \, y + s(x), \quad y(a) = y_a, \quad y(b) = y_b
\]

does not have a unique solution whenever

\[0 < \sqrt{g} \, (b-a) = n \pi, \quad n = 1, 2, 3, \ldots\]

We also have

\[
2 \cos \theta = \frac{(2 - h^2 \beta_1 g)}{(1 + h^2 \beta_0 g)} < 2.
\]
Expanding cosθ in powers of θ and the expression on the right hand side in powers of h, we get

\[ 2(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \ldots) = (2 - h^2 \beta_0 g)(1 - h^2 \beta_0 g + h^4 \beta_0^2 g^2 - \ldots), \]

\[ = 2 - (2\beta_0 + \beta_1) h^2 g + O(h^4), \]

\[ = 2 - h^2 g + O(h^4), \]

or

\[ \theta \approx \sqrt{g} h. \]

Consequently

\[ \sin((N+1)g) \approx \sin[(N+1)\sqrt{g} h] \approx \sin((b-a)\sqrt{g}) \approx 0 \]

if \( 0 < (b-a) \sqrt{g} = n \pi, \) (n = 1,2,\ldots).

Therefore the expression on the right hand side of (5.4) which has \( \sin(N+1)\theta \) in its denominator could be very large (if not infinite) provided the range of integration \( (b-a) \) is such that

\[ 0 < (b-a) \sqrt{g} \approx n \pi, \] \( n = 1,2,3,\ldots \), \( \ldots(5.7) \)

and in that case mathematically the system of linear equations arising from the D.E. may have a unique solution but numerically the problem is ill-conditioned.

We consider the boundary value problem

\[ y'' + E^2 y = \sin x, \] \( E^2 \neq 0, \) \( 1 \) and \( E \neq \frac{n\pi}{b} \)

where n is an integer).

\[ y(0) = 1 \]

\[ y(b) = \sin( Eb ) + \cos( Eb ) + \sin b / (E^2 - 1), \] \( \) see (8.2).

We choose those values of \( E \) and \( b \) whose product approximately equals \( n\pi, \) for some positive integer \( n \). For example, for \( E = 2 \) and \( b = 11, \)

\( Eb = 22 \approx 7\pi. \) Note that \( 7\pi \) is approximately equal to 21.99. The "max. absolute error" in the numerical solution with \( h = 1/8 \) is found to be
0.20 $E^{-1}$ and the theoretical estimate based on the inequality (5.6) is equal to 0.56 $E^{-1}$. The ratio of "max. absolute error" to its estimate is 0.35. On choosing $E = 11$, and $b = 2$, the "max. absolute error" and the error estimate turn out to be 0.10 $E^{-1}$ and 0.28 $E^{-1}$ respectively, and their ratio being 0.36. On taking $E = b = 5$ ($E_b = 25 \approx 8\pi$), the "max. absolute error" is equal to 0.68 $E^{-1}$, while the error estimate based on (5.6) is 0.18. The ratio of "max. absolute error" to its estimate is 0.38. We tried several other combinations of $E$ and $b$. The results are similar to those discussed above.

The above experimental results confirm that the expression for error bound (5.4) is predominantly range dependent and whenever range $(b-a)$ satisfies (5.7), the corresponding "max. absolute error" and its estimates both become large.

5.3 AN $O(h^8)$ ERROR BOUND

We can obtain more accurate results of the form (5.4) for the boundary value problem (5.1) with $g(x) = \text{constant}$, by using A.E.'s of the form (4.26). We first consider (4.26) for $t = 2$ and show that under certain conditions the resulting error is $O(h^8)$. Here the resulting system of linear equations will be of the form (5.2)

where

$$A_n = 2 - h^2 \beta_1 g + h^4 \gamma_1 g^2,$$

$$B_n = -1 - h^2 \beta_0 g + h^4 \gamma_0 g^2,$$

$$59 M_{10} h^{10},$$

and $||\Gamma(y)|| \leq \frac{76,204,800}{76,204,800},$

for the coefficients $\beta_1$'s and $\gamma_1$'s, see A.E. (4.24). The constant $\mu$ introduced in Section 2.4 will now be given by
\[ \mu = \frac{(2-h^2 \beta_1 g + h^4 \gamma_1 g^2)/(1+h^2 \beta_0 g - h^4 \gamma_0 g^2)}{1+h^2 \beta_0 g - h^4 \gamma_0 g^2} \].

The denominator of \( \mu \) is clearly greater than 1 because \( \beta_0 > 0 \) and \( \gamma_0 < 0 \).

The numerator will be less than 2 provided
\[ \beta_1 > h^2 \gamma_1 g \]

or on substituting the values of \( \beta_1 \) and \( \gamma_1 \), we have
\[ h^2 g < 6900/313 \].

Thus \( \mu < 2 \) if \( h \) satisfies the above inequality.

Finally using the theory developed in Section 2.4, we get
\[ \| A^{-1} \| < \frac{N}{(1+h^2 \beta_0 g - h^4 \gamma_0 g^2) \sin \theta |\sin(N+1)\theta|} \]
where \( 2 \cos \theta = \mu < 2 \),
and
\[ \| e \| < \frac{N \| \Gamma(y) \|}{(1+h^2 \beta_0 g - h^4 \gamma_0 g^2) \sin \theta |\sin(N+1)\theta|} \]

On substituting
\[ (1+h^2 \beta_0 g - h^4 \gamma_0 g^2) \sin \theta = h \sqrt{g+h^2 g^2 (\beta_0-\beta_1/4)-h^2 g^2 (2\gamma_0+\gamma_1)-2h^4 g^3 (\beta_0 \gamma_0-\beta_1 \gamma_1/4)} \]
\[ + h^6 g^4 (\gamma_0-\gamma_1/4), \]
\[ = h \sqrt{g + 0(h^2)}, \]
we get
\[ \| e \| < \frac{59 M_{10} h^8}{76,204,800|\sin(N+1)\theta| \sqrt{g + 0(h^2)}} \]
\[ \ldots (5.8) \]

or
\[ \| e \| = 0(h^8). \]

In a similar manner we can use A.E. (4.26) for \( t = 3 \) and show that the resulting error in the corresponding case is \( 0(h^{12}) \). We omit the details here for brevity.
5.4 THE GENERAL CASE WHEN \( g(x) \neq \text{constant} \)

In determining the error bounds (5.4) and (5.8) respectively, we have so far assumed that \( g(x) \) is constant. The general case when \( g(x) \) is not a constant cannot be dealt with by the above method. We suggest that error bound should be obtained numerically.

We consider the error equation (5.2) namely

\[ A e = \Gamma(y) . \]

Thus

\[ ||e|| \leq |c_p + 2| h^{p+2} M_p + 2 ||A^{-1}|| . \]  

...(5.9)

In order to obtain the error bound given above, we will compute \( A^{-1} \) and hence the norm \( ||A^{-1}|| \) numerically.

In actual practice, the error estimate obtained by the above technique is not very good, especially when the range of integration is fairly large. By the error estimate not being very good, we mean that the ratio of maximum absolute error to its estimate is very small. We tried the error estimate given by (5.9) in various situations. The results for the problem (5.10) are displayed in Table IV on the next page.

\[ y'' + x^2 y = x + x^3 + 2/x^3 , \]  

...(5.10)

\[ y(1) = 2, \ y(b) = b + \frac{1}{b} \text{ with } y(x) = x + \frac{1}{x} , \]

and \( M_n \leq n ! \). The Table IV shows that for the range of integration \([1,2], [1,4] \) and \([1,8] \), the error estimate is too crude so that the ratio of maximum absolute error to its estimate becomes <<1. Also this ratio goes on decreasing as the range of integration increases and therefore the error estimate based on (5.9) does not reflect well the variation in the actual error. Besides the D.E. (5.10) we also tried the following
TABLE IV
EXPERIMENTS WITH PROBLEM (5.10)

<table>
<thead>
<tr>
<th>b</th>
<th>h</th>
<th>Max. Abs. Error</th>
<th>Error Estimate based on (5.9)</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1/4</td>
<td>0.18 E-3</td>
<td>0.36 E-2</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>1/8</td>
<td>0.13 E-4</td>
<td>0.29 E-3</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>1/16</td>
<td>0.82 E-6</td>
<td>0.19 E-4</td>
<td>0.043</td>
</tr>
<tr>
<td>4</td>
<td>1/4</td>
<td>0.21 E-3</td>
<td>0.56 E-2</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>1/8</td>
<td>0.14 E-4</td>
<td>0.46 E-2</td>
<td>0.030</td>
</tr>
<tr>
<td>8</td>
<td>1/4</td>
<td>0.38 E-3</td>
<td>0.15 E-1</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Two problems:

\[ y'' + \left(1 - \frac{2}{x^2}\right) y = -\frac{2 \cos x}{x^2} \]  
\text{...(5.11a)}

with boundary conditions such that the true solution of the D.E. turns out to be \( y(x) = \sin x / x \).

\[ y'' + \left(\frac{3 + 4x}{16x^2}\right) y = 0, \ y(1) = 1, \ y(b) \]  
\text{...(5.11b)}

is so chosen that the true solution of the problem turns out to be

\[ y(x) = x^{1/4} \left[ \cos (\sqrt{x} - 1) - 0.5 \sin (\sqrt{x} - 1) \right] . \]

The conclusions are identical to those obtained for the problem (5.10). The experimental results based on theorem 5.1 will be given in the last chapter.
6.1

If the function \( f(x,y) \) introduced in D.E. (1.1) is not linear in \( y \), one cannot hope to solve the system of equations (2.4) namely

\[
J y + h^2 D f(y) = d,
\]

by algebraic methods. Some iterative procedure must be adopted. The following discussion is devoted to proving the existence and uniqueness of the solution of the system (2.4). At the same time, this solution will be shown to be an approximation to the solution of the corresponding boundary value problem. The basis for the theorem which we shall state is a method which is a generalization of Newton's method. Kantorovich [16] proved a very general result which guarantees the convergence of Newton's method without even assuming the existence of a solution. We shall state Kantorovich's result in a form applicable to our discussion.

**Theorem 6.1**

Let the \( n \) equations

\[
\phi_i (y_1, y_2, \ldots, y_n) = 0, \quad (i = 1, 2, \ldots, n)
\]

for \( n \) unknowns \( y_1, y_2, \ldots, y_n \) be written in the vector form as

\[
\phi(y) = 0.
\]

Let \( A(y) = (a_{ij}) \) denote the matrix with elements

\[
a_{ij} = \frac{\partial \phi_i}{\partial \phi_j},
\]

then the Newton's method is written in the form

\[
y^{(v+1)} = y^{(v)} - [A^{-1}(y^{(v)})] \phi (y^{(v)}), \quad (v = 0, 1, \ldots) \quad ...(6.1)
\]
Now if the following conditions are satisfied:

(i) For \( y = y^{(0)} \), the initial approximation, the matrix \( A (y^{(0)}) \) has an inverse \( \Gamma_0 \) and an upper bound for its norm is known i.e.

\[
||\Gamma_0|| \leq B_0.
\] ...(6.2)

(ii) The vector \( y^{(0)} \) approximately satisfies the system of equations in the sense that the first correction vector \( \Gamma_0 \phi (y^{(0)}) \) is such that

\[
||\Gamma_0 \phi (y^{(0)})|| \leq \eta_0
\] ...(6.3)

(iii) In the region defined by inequality (6.6) below, the components of the vector \( \phi(y) \) are twice continuously differentiable with respect to the components of \( y \) and satisfy

\[
\sum_{j,k=1}^{n} \left| \frac{\partial^2 \phi_i}{\partial y_j \partial y_k} \right| \leq K_0 \quad (i = 1,2, \ldots, n)
\] ...(6.4)

(iv) The constants \( B_0, \eta_0 \) and \( K_0 \) introduced above satisfy the inequality

\[
0 < h_0 = B_0 \eta_0 K_0 < 1/2,
\] ...(6.5)

then the system of equations \( \phi(y) \) has a solution \( y^* \) which is located in the hypercube

\[
||y - y^{(0)}|| \leq \frac{1 - \sqrt{1 - 2h_0}}{h_0} \eta_0.
\] ...(6.6)

Moreover, the successive approximation \( y^{(v)} \) defined above exist and converge to \( y^* \), and a bound of \( ||y^{(v)} - y^*|| \) is given by the inequality

\[
||y^{(v)} - y^*|| \leq \frac{1}{2^{v-1}} (2h_0)^{2^v-1} \eta_0
\] ...(6.7)

For the proof of the above theorem, see [12], Theorem 7.6.

We would like to remind ourselves that at present we are attempting to solve the system of nonlinear algebraic equations (2.4) by Newton's method. Now for a given \( y \), the residual vector \( r(y) \) is defined as
\[ r(y) = J y + h^2 D f(y) - d. \quad \ldots (6.8) \]

For our problem, the matrix \( A(y) \) introduced in the Theorem 6.1 is given by

\[ A(y) = J + h^2 D F(y), \quad \ldots (6.9) \]

(the matrices \( J \) and \( D \) have been introduced in the Section 2.2).

\[
\begin{bmatrix}
  f_y(x_1, y_1) \\
  \vdots \\
  f_y(x_N, y_N)
\end{bmatrix}
\]

where

\[
F(y) =
\begin{bmatrix}
  f_y(x_1, y_1) \\
  \vdots \\
  f_y(x_N, y_N)
\end{bmatrix},
\]

and we assume that \( f_{yy}(x, y) \) exists and is bounded in \( S \). Now \( y^{(0)} \) is a vector believed to be close to the actual solution of (6.1) so that the residual vector

\[ r(y^{(0)}) = J y^{(0)} + h^2 D f(y^{(0)}) - d \]

is small. We now solve the linearized system

\[ r(y^{(0)}) + (J + h^2 D F(y^{(0)})) \Delta y^{(0)} = 0, \quad \ldots (6.10) \]

and its solution is given by

\[ \Delta y^{(0)} = -[J + h^2 D F(y^{(0)})]^{-1} r(y^{(0)}) = -[A(y^{(0)})]^{-1} r(y^{(0)}) \]

provided that the inverse of the matrix \( A(y) \) defined by (6.9) exists for \( y = y^{(0)} \). Then the vector \( y^{(1)} = y^{(0)} + \Delta y^{(0)} \) will be a better approximation to the exact solution, the residual vector \( r(y^{(1)}) \) will be smaller, and the process can be repeated with \( y^{(1)} \) taking the place of \( y^{(0)} \), etc., until the convergence is achieved.

Note that the solution of the linearized system (6.10) is
greatly facilitated by the fact that the matrix $A(y^{(0)})$ is again tri-diagonal. In fact, if $A(y^{(0)}) = (a_{mn})$, we have

$$
\begin{align*}
    a_{n,n-1} &= -1 + h^2 \beta_0 f_y(x_{n-1}, y_{n-1}) \quad (n = 2, 3, \ldots, N) \\
    a_{n,n} &= 2 + h^2 \beta_1 f_y(x_n, y_n) \quad (n = 1, 2, \ldots, N) \quad \ldots (6.11) \\
    a_{n,n+1} &= -1 + h^2 \beta_2 f_y(x_{n+1}, y_{n+1}) \quad (n = 1, 2, \ldots, N-1)
\end{align*}
$$

and all the other elements are zero. Now the method described earlier for the linear boundary value problem in Chapter II is immediately applicable. The only extra work that is to be required is to evaluate the residual vector $r(y^{(v)})$ and the partial derivatives $f_y(x_n, y_n)$.

6.2 **CONVERGENCE OF NEWTON'S METHOD FOR THE SYSTEM (2.4)**

We shall verify that the conditions of the Theorem 6.1 can be satisfied for the system of the nonlinear equations (2.4) based on the A.E. (2.10b). Since $\beta_i \geq 0$, $f_y \geq 0$, the matrix $A(y)$ given by the equation (6.9) has all the right kind of properties stated in (2.8). Thus we have

$$
0 \leq [A(y)]^{-1} \leq J^{-1}.
$$

But $||J^{-1}|| \leq (N+1)^2/8$, it follows that (6.2) is satisfied for

$$
B_0 = (N+1)^2/8 = (b-a)^2/8h^2. \quad \ldots (6.12)
$$

Since

$$
\phi_i(y) = -y_{i-1} + 2y_i - y_{i+1} + (h^2/12) [f(x_{i-1}, y_{i-1}) + 10 f(x_i, y_i)
$$

$$
+ f(x_{i+1}, y_{i+1})],
$$

hence

$$
\sum_{j,k=1}^{N} \frac{\partial^2 \phi_i(y)}{\partial y_j \partial y_k} = (h^2/12) [f_{yy}(x_{i-1}, y_{i-1}) + 10 f_{yy}(x_i, y_i)
$$

$$
+ f_{yy}(x_{i+1}, y_{i+1})], \quad (i = 1, 2, \ldots, N)
$$
or
\[
\frac{1}{N} \sum_{j,k=1}^{N} \left| \frac{\partial^2 \phi_{j,k}(y)}{\partial y_j \partial y_k} \right| \leq h^2 \text{ max. } \left| f_{yy}(x,y) \right|
\]

Thus the condition (6.4) of the theorem 6.1 is satisfied for
\[
K_0 = h^2 L_2 ,
\]
where
\[
L_2 = \text{ maximum } \left| f_{yy}(x,y) \right|_{(x,y) \in S}.
\]

Now let the initial approximations \( y^{(0)} \) be defined by
\[
y_n^{(0)} = Q(x_n) \quad (n = 1, 2, \ldots, N) ,
\]
where \( Q(x) \in C^6 \) satisfying \( Q(a) = y_a \), \( Q(b) = y_b \), and let \( Q_n = \max_x \left| Q^{(n)}(x) \right| \),
\[
R = \max_x \left| Q''(x) - f(x,Q(x)) \right| , \ \forall x \in [a,b] .
\]

It then follows (see (2.10b)) that
\[
\left| -Q(x) + 2Q(x+h) - Q(x+2h) + (h^2/12) \left[ Q''(x) + 10Q''(x+h) + Q''(x+2h) \right] \right|
\]
\[
\leq h^6 Q_6 / 240 .
\]

Hence
\[
|r_i(y^{(0)})| = \left| -Q(x_{i-1}) + 2Q(x_i) - Q(x_{i+1}) + (h^2/12) \left[ f(x_{i-1},Q(x_{i-1})) + 10 f(x_i,Q(x_i)) + f(x_{i+1},Q(x_{i+1})) \right] \right|
\]
\[
= \left| -Q(x_{i-1}) + 2Q(x_i) - Q(x_{i+1}) + (h^2/12) \left[ Q''(x_{i-1}) + 10Q''(x_i) + Q''(x_{i+1}) \right] 
- (h^2/12) \left[ (Q''(x_{i-1}) - f(x_{i-1},Q(x_{i-1})) + 10(Q''(x_i) - f(x_i,Q(x_i))) 
+ (Q''(x_{i+1}) - f(x_{i+1},Q(x_{i+1}))) \right] \right|
\]
\[
< h^6 Q_6 / 240 + h^2 (R + 10R + R) 
< h^6 Q_6 / 240 + h^2 R , \quad (i = 1, 2, \ldots, N) .
\]

Thus (6.3) is satisfied for
\[
\eta_0 = \frac{(b-a)^2}{8} (R + h^4 Q_6 / 240) .
\]
The main condition (6.5) of the Theorem 6.1, which guarantees convergence of Newton's process, thus turns out to be satisfied if

\[ \left( \frac{1}{64} \right) \frac{1}{(b-a)^4} L_2 (R + h^4 Q_6 / 240) \leq \frac{1}{2}. \]  

(6.16a)

If as an initial approximation \( Q(x) \) a polynomial of degree \( <6 \) is chosen, then \( Q_6 = 0 \), and the method will converge if the quantity \( R \) defined by the equation (6.14) satisfies the inequality

\[ R \leq \frac{32}{(b-a)^4 L_2}. \]  

(6.16b)

Having thus demonstrated the existence of a solution of the system (2.4) uniqueness can now be proved.

If \( y^* \) and \( y^{**} \) are any two solutions of the system (2.4), viz.

\[ J y + h^2 D f(y) = d, \]

then we may put

\[ f(x_n, y^*_n) - f(x_n, y^{**}_n) = (y^*_n - y^{**}_n) \eta_n, \]  

(6.17)

where \( \eta_n \) is a value of \( \frac{\partial f}{\partial y} \). The vector \( y^* - y^{**} \) is easily seen to satisfy the system

\[ (J + h^2 D B) (y^* - y^{**}) = 0, \]  

(6.18)

where \( J \) and \( D \) are defined by equation (2.4), and \( B \) is the diagonal matrix with positive elements \( \eta_n \). The matrix \( J + h^2 D B \) is known to be non-singular and monotone. It follows that the vector \( y^* - y^{**} = 0 \).

6.3 The conditions of convergence of Newton's method namely (6.16a) and (6.16b) can be improved by using a better estimate \( B_0 \) for the norm of the matrix \( \Gamma_0 \) using techniques of the section 2.4. In fact, one proves that (6.2) is satisfied for

\[ B_0 = \frac{H_N(\mu)}{h^2 L}. \]
where \( u = \frac{(24 + 10 h^2 L)}{(12 - h^2 L)} > 2 \),

\[ L = \min. \sqrt{\frac{\partial^2 f(x,y)}{\partial y^2}}, (x,y) \in S \] and \( H_N(u) \) has been introduced by (2.14).

Thus (6.16a) can be rewritten in the form

\[ \frac{L^2}{L^2} H_N^2(u) (R + \frac{h^4}{240} \frac{Q_0}{L^2}) \leq \frac{1}{2}, \]

and (6.16b) in the form

\[ R \leq \frac{L^2}{2 L^2 H_N^2(u)}. \]

### 6.4

Now we will consider the solution of the nonlinear D.E. (1.1) by D.E.'s of higher order. In particular we consider the use of D.E. (3.8) at the interior points and D.E.'s (3.9) near the boundary points. When the two point nonlinear boundary value problem of class M is replaced by D.E.'s (3.8) and (3.9), we get the system

\[ J(c)y + h^2 D(c) f(y) = d \quad \cdots (6.19) \]

where \( J(c) \) and \( D(c) \) have been introduced by (3.12). Now we will develop criteria for the convergence of Newton's method, as applied to the system (6.19). We will estimate \( B_0, n_0 \) and \( K_0 \) as introduced in the Theorem 6.1.

For our purpose

\[ A(y) = J(c) + h^2 D(c) F(y), \quad \cdots (6.20) \]

where \( F(y) \) is defined in (6.9). We also have

\[ [A(y)]^{-1} \leq J^{-1}(c), \]

provided \( c, h \) satisfy the inequalities (3.16) and (3.17). Thus the condition (6.2) is now satisfied for

\[ B_0 = \frac{(b-a)^2}{8h^2(c-2)} H_N(c), \quad \cdots (6.21) \]

see inequality (3.20). Again using notations of the Theorem 6.1, we have
\[ \sum_{j,k=1}^{N} \frac{\partial^2 \phi_{j,k}}{\partial y_j \partial y_k} = \begin{cases} h^2 L_2 (c+1), & (i = 1,N) \\ h^2 L_2 (c+2), & (i = 2,3,\ldots,N-1) \end{cases} \]

Thus we have
\[ k_0 = h^2 L_2 (c+2). \] ...(6.22)

Also following a usual analysis, as in the previous section
\[ \left| r_i(y^{(0)}) \right| \leq \begin{cases} h^2 R (c+1) + h^6 G_1 Q_6, & (i = 1,N) \\ h^2 R (c+2) + h^8 G_2 Q_8, & (i = 2,3,\ldots,N-1) \end{cases} \]

for some real constants \( G_1 \) and \( G_2 \). In fact \( G_1 = \frac{c+1}{240} \) and \( G_2 = \frac{|31c-190|}{60,480} \), \( c \neq \frac{190}{31} \), the other symbols having been defined in the previous section.

Thus (6.3) would be satisfied for
\[ h_0 = \frac{(b-a)^2}{8(c-2)} H_N(c) \left[ R(c+2) + \sigma \right], \] ...(6.23)

where \( \sigma = \max. [h^4 G_1 Q_6, h^6 G_2 Q_8] \). Hence the condition for the convergence of Newton's method in the present case becomes
\[ h_0 = \frac{(b-a)^4 (c+2) L_2 \left[ R(c+2) + \sigma \right] H_N(c)}{64 (c-2)^2} \leq \frac{1}{2}. \] ...(6.24)

Also as before, if the initial approximation \( Q(x) \) is a polynomial of degree \(<6\), then \( \sigma = 0 \) and the method will converge if the quantity \( R \) satisfies the inequality
\[ R \leq \frac{32 (c-2)^2}{(b-a)^4 (c+2) L_2}. \] ...(6.25)

6.5 We also used a \( \Delta.E. \) of order \( k = 2 \) obtained by overdifferentiation for solving nonlinear boundary value problem of class \( M \). In particular we used the \( \Delta.E. \) (4.9). We used auxiliary \( \Delta.E.'s \) (4.18) and (4.20) to obtain \( y_n' \) and \( y_{n+2}' \) respectively. We first substituted
\[ y'' = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} y'. \]
using the D.E., and then the expressions for $y'_{n+m} (m = 0,2)$ in the A.E. (4.9).

We will present numerical results using the above technique as well as the techniques discussed in Sections 6.2 and 6.4, respectively. We might remark that we will not try to use A.E.'s of the form (4.1) for $t \geq 4$ to solve a nonlinear boundary value problem of class M. The main reason is that

$$y^{(4)} = f_{xx} + f f_y + 2 f_{xy} y' + (y')^2 f_{yy},$$

therefore when we substitute $y''$, $y'''$, ... etc. using the D.E. and then the expressions for $y'_{n+m} (m = 0,1,2)$ obtained from auxiliary A.E.'s (4.18), (4.19) and (4.20) in A.E. (4.1), the resulting A.E. becomes quite complicated and at the same time it retains a T.E. = $O(h^8)$, which is not at all an improvement on the previous situation.
7.1 GENERAL

In the present chapter it is shown that the boundary value techniques developed to solve the two point boundary problem can be used to solve a certain class of initial value problems which are usually solved by step-by-step method. Allen and Severn [1] suggested a method whereby a first order system, normally solved by initial value techniques, might be replaced by a second order boundary value problem. No indication was given of the merit of this procedure, except when extended to a certain partial differential equations. Later Fox [10] proposed to treat a derived second order equation; but remarked that either initial value or boundary value techniques might be used, the choice depending on the nature of the solution. No great details were given regarding this choice. Fox and Mitchell [11] in 1957 investigated these two methods by which a first order D.E. can be transformed into a second order D.E., solvable either by initial value or boundary value techniques. Their illustrative examples were based on linear D.E.'s with constant coefficients.

In what follows, we will consider a certain family of first order linear D.E.'s and solve them by the boundary value techniques, developed in Chapters II and IV. It will be proved that the resulting error is $O(h^t)$, $t = 4, 6$ or 8.
7.2 **LINEAR D.E. OF THE FIRST ORDER**

Consider a D.E. of the first order in the form

\[ y' = \phi(x,y) , \quad y(a) = y_a . \]  

...(7.1)

Fox [10] produces a second order D.E. by direct differentiation of the original problem (7.1) and subsequent elimination of the first derivative. For instance (7.1) is replaced by

\[ y'' = \frac{3\phi}{3x} + \frac{3\phi}{3y} \phi(x,y) . \]  

...(7.2)

For boundary conditions, he takes the given condition \( y(a) = y_a \) and the first order equation (7.1) applied at \( x = b \) viz.

\[ y(a) = y_a , \quad y'(b) = \phi(b,y_b) . \]  

...(7.3)

On approximating (7.2) - (7.3) by a Δ.E. of the form (1.2) or (1.3), we get a system of algebraic equations, linear if \( \phi(x,y) \) is linear. If \( \phi(x,y) \) is nonlinear, then the resulting system of algebraic equations must be solved by some iterative method. However in this chapter, we want to show how the techniques developed in Chapters II and IV for solving two point linear boundary value problem of class M (M-1, M-4 and M-5) can be modified to solve a linear D.E. of the first order (7.1) and error bounds can be conveniently obtained provided certain conditions (to be stated later) are satisfied. Thus we assume

\[ \phi(x,y) = f(x) y + g(x) , \]

hence the derived second order boundary value problem takes the form

\[
\begin{align*}
    y'' &= [f^2(x) + f'(x)] y + [f(x) g(x) + g'(x)] \\
      &= F(x) y + G(x) \\
    y(a) &= y_a , \quad y'(b) = f(b) y(b) + g(b),
\end{align*}
\]  

...(7.4)
which we want to solve by modifying the direct methods M-1, M-4 and M-5, respectively.

This time the points \( x_n \) are given by

\[
x_n = a + nh \quad (n = 0, 1, 2, \ldots, N)
\]

with \( h = (b-a)/N \).

We first use the A.E. (2.10b) to replace the boundary value problem (7.4). This A.E. is such that the operator associated with it is definite and the generalized mean value theorem (1.11) holds, see Section 2.1. We then obtain the following system of (N-1) linear algebraic equations in N unknowns \( y_1, y_2, \ldots, y_N \),

\[
(-1 + \frac{h^2}{12} P_n) y_n + (2 + \frac{10h^2}{12} P_{n+1}) y_{n+1} + (-1 + \frac{h^2}{12} P_{n+2}) y_{n+2} + \frac{h^2}{12} (G_n + 10 G_{n+1} + G_{n+2}) = 0, \quad (n = 0, 1, \ldots, N-2). \tag{7.5}
\]

Thus we have (N-1) equations in N unknowns. We develop another equation by considering a A.E. of the form

\[
\sum_{i=0}^{k} a_i y_{n+i} = h \sum_{i=0}^{k} b_i y'_{n+i} + \ldots + h^j \sum_{i=0}^{k} \theta_{i} y^{(j)}_{n+i}, \quad j > 1, \tag{7.6}
\]

which involves the boundary condition at \( x = b \). The reason for this choice of the above A.E. is that the matrix associated with the resulting system of linear equations becomes tridiagonal for \( k = 1, j > 1 \). The following Table V gives the values of the coefficients associated with the A.E. (7.6) for \( k = 1 \) and \( j = 3, 4 \) and 5.
TABLE V

COEFFICIENTS OF THE Δ.E. (7.6)

<table>
<thead>
<tr>
<th>j</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₀</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>a₁</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>b₀</td>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>b₁</td>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>c₀</td>
<td>1/10</td>
<td>3/28</td>
<td>1/9</td>
</tr>
<tr>
<td>c₁</td>
<td>-1/10</td>
<td>-3/28</td>
<td>-1/9</td>
</tr>
<tr>
<td>d₀</td>
<td>1/120</td>
<td>1/84</td>
<td>1/72</td>
</tr>
<tr>
<td>d₁</td>
<td>1/120</td>
<td>1/84</td>
<td>1/72</td>
</tr>
<tr>
<td>e₀</td>
<td>1/1,680</td>
<td>1/1,008</td>
<td></td>
</tr>
<tr>
<td>e₁</td>
<td>-1/1,680</td>
<td>-1/1,008</td>
<td></td>
</tr>
<tr>
<td>g₀</td>
<td></td>
<td></td>
<td>1/30,240</td>
</tr>
<tr>
<td>g₁</td>
<td></td>
<td></td>
<td>1/30,240</td>
</tr>
</tbody>
</table>

T.E. \[-h^7 y^{(7)}/100,800 \quad h^9 y^{(9)}/25,401,600 \quad -h^{11} y^{(11)}/21\times12!\]

The coefficients listed above for j = 3 and 4 have also been given by Lambert and Mitchell [18]. The Δ.E. (7.6) for k = 1, j = 3 and n = N-1 becomes

\[-y_{N-1} + y_N = (h/2)(y'_{N-1} + y'_N) + (h^2/10)(y''_{N-1} - y''_N) \quad \ldots (7.7)
+ (h^3/120)(y'''_{N-1} + y'''_N).\]

The numerical solution of the system (7.5) along with the equation (7.7) is assumed to approximate the solution of the boundary value problem (7.4).
7.3 ERROR ANALYSIS

The error equation is obtained in the usual manner in the form

\[ Ae = \Gamma(y) \], \hspace{1cm} \cdots (7.8)

where we now have

\[
\begin{bmatrix}
A_1 & B_2 \\
B_1 & A_2 & B_3 \\
B_2 & A_3 & B_4 \\
& & & \ldots & \ldots & \ldots \\
& & & & & & \ldots \\
& & & & & & \\
B_{N-2} & A_{N-1} & B_N \\
B^*_N & A_N \\
\end{bmatrix}
\begin{bmatrix}
Y_1 \\
Y_2 \\
Y_3 \\
Y_{N-1} \\
Y_N \\
\end{bmatrix}
\]

with

\[
A_n = 2 + (10h^2/12) F_n \hspace{1cm} (n = 1, 2, \ldots, N-1)
\]

\[
A_N = 1 - \frac{h}{2} F_N + \frac{h^2}{10} F_N - \frac{h^3}{120} R_N
\]

\[
B_n = -1 + \frac{h^2}{12} F_n \hspace{1cm} (n = 1, 2, \ldots, N)
\]

\[
B^*_{N-1} = -1 - \frac{h}{2} F_{N-1} + \frac{h^2}{10} F_{N-1} - \frac{h^3}{120} R_{N-1}
\]

\[
Y_n = -\frac{h^6}{240} y^{(6)} (x_n + \theta_n h) \hspace{1cm} (n = 1, 2, \ldots, N-1), |\theta_n| \leq 1
\]

\[
Y_N = \frac{h^6}{100,800} y^{(7)} (\xi), x_{N-1} \leq \xi \leq x_N
\]

Here \( y''' = R(x) y + S(x) \). Please note that the A.E. (7.7) also satisfies the generalized mean value theorem. This can be proved by using techniques discussed in Section 2.1. Now for all sufficiently small values of \( h \), we have

\[ ||\Gamma(y)|| \leq \frac{h^6 M_6}{240} \]

However, if \( f(x) \leq 0 \) and \( F(x) \geq 0 \) over \([a,b]\) (note that these two conditions will automatically be satisfied in case if \( f(x) \) is a real
negative constant), then the matrix $A$ has all the right kind of properties mentioned in (2.8) and it can be shown that

$$0 \leq A^{-1} \leq \bar{J}^{-1}$$

where $\bar{J}$ differs from $J$ in the last element on the main diagonal which is 1 rather than 2. In fact

$$\begin{bmatrix}
2 & -1 \\
-1 & 2 & -1 \\
-1 & 2 & -1 \\
\end{bmatrix}
\begin{bmatrix}
\ddots & \ddots & \ddots & \ddots \\
& \ddots & \ddots & \ddots \\
& & \ddots & \ddots & \ddots \\
& & & -1 & 2 & -1 \\
& & & -1 & 1 \\
\end{bmatrix}
\begin{bmatrix}
-1 \\
2 \\
-1 \\
1 \\
\end{bmatrix}$$

and $\bar{J}^{-1}$ can be determined explicitly (see Rutherford [23]) in the form

$$\begin{bmatrix}
1 & 1 & 1 & \ldots & 1 \\
1 & 2 & 2 & \ldots & 2 \\
1 & 2 & 3 & \ldots & 3 \\
& \ddots & \ddots & \ddots & \ddots \\
& & \ddots & \ddots & \ddots \\
1 & 2 & 3 & \ldots & N-1 & N-1 \\
1 & 2 & 3 & \ldots & N-1 & N \\
\end{bmatrix}$$

Now we prove the following lemma.

**Lemma 7.1**

Let $\bar{J}t = \omega$, where $t$ is a vector with components $t_1, t_2, \ldots, t_N$; $\omega$ being a column vector with each component unity, then

(i) $t_i = \frac{i(2N+1-i)}{2}$

(ii) $||t|| = \frac{N(N+1)}{2}$.
The proof of this lemma is trivial, since $t_i$ is just the sum of the elements of the $i$th row in $J^{-1}$. Now using (7.8), we get

$$e = A^{-1} \Gamma(y)$$

$$e \leq \frac{h^6 M_6}{240} J^{-1} \omega$$

$$|e_i| \leq \frac{h^6 M_6}{240} \times \frac{i(2N+1-i)}{2}$$

and

$$||e|| \leq \frac{h^4 M_6 N(N+1)}{480}$$

$$||e|| \leq \frac{h^4 M_6}{480} [(b-a)^2 + h(b-a)] . \quad \ldots(7.9)$$

Thus we have proved the following theorem:

**Theorem 7.2**

Given the boundary value problem

$$\begin{bmatrix} y'' = [f^2(x) + f'(x)] y + [f(x) g(x) + g'(x)] = F(x) y + G(x) , \\
y(a) = y_a, y'(b) = f(b) y(b) + g(b) \end{bmatrix}$$

(derived from the initial value problem $y' = f(x) y + g(x)$, $y(a) = y_a$), we replace it by the Δ.E.'s

$$y_n - 2y_{n+1} + y_{n+2} = \frac{h^2}{12} (y_n'' + 10y_{n+1}'' + y_{n+2}'') \quad (n = 0, 1, \ldots, N-2)$$

and

$$-y_{N-1} + y_N = \frac{h}{2} (y_{N-1}' + y_N') + \frac{h^2}{10} (y_{N-1}'' - y_N'') + \frac{h^3}{120} (y_{N-1}''' + y_N''') .$$

Then if

(i) $f(x) \leq 0, F(x) \geq 0$ over $[a,b]$$

(ii) $y(x) \in C^6, f(x), g(x) \in C^3$ ,

$$||e|| \leq \frac{h^4 M_6}{480} [(b-a)^2 + h(b-a)]$$

or equivalently $||e|| = 0(h^4)$. 
The boundary value technique described above to solve an initial value problem of the first order where the error estimate is given by Theorem 7.2 will henceforth be referred to as $M^{-1}$.

7.4 AN $O(h^6)$ AND $O(h^8)$ FINITE DIFFERENCE ANALOGUE FOR (7.4)

We approximate the boundary value problem (7.4) by the Δ.E. (4.9), viz.

$$y_n - 2y_{n+1} + y_{n+2} = h^2 \sum \beta_i y_{n+i} + h^3 \sum \gamma_i y_{n+i}^{(4)} \quad (n = 0, 1, \ldots, N-2)$$

and

$$-y_{N-1} + y_N = \frac{h}{2} (y_{N-1}^{(4)} + y_N^{(4)}) + \frac{h^2}{9} (y_{N-1}^{(2)} - y_N^{(2)}) + \frac{h^4}{72} (y_{N-1}^{(4)} + y_N^{(4)})$$

We prove in a manner analogous to the one used in previous section that

$$||e|| \leq \frac{29 h^6}{604,800} \left[ (b-a)^2 + h(b-a) \right]. \quad \ldots\ldots (7.10)$$

This method will be designated as $M^{-4}$ for reference purposes.

In order to develop an $O(h^8)$ finite difference analogue to solve (7.4) we use the Δ.E. (4.24) viz.

$$y_n - 2y_{n+1} + y_{n+2} = h^2 \sum \beta_i y_{n+i}^{(4)} + h^4 \sum \gamma_i y_{n+i}^{(4)} \quad (n = 0, 1, \ldots, N-2)$$

and

$$-y_{N-1} + y_N = \frac{h}{2} (y_{N-1}^{(4)} + y_N^{(4)}) + \frac{h^2}{9} (y_{N-1}^{(2)} - y_N^{(2)}) + \frac{h^4}{72} (y_{N-1}^{(4)} + y_N^{(4)})$$

$$\quad + \frac{h^4}{1,008} (y_{N-1}^{(4)} - y_N^{(4)}) + \frac{h^5}{30,240} (y_{N-1}^{(5)} - y_N^{(5)})$$

Under the same conditions as mentioned in Theorem 7.2, we prove that

$$||e|| \leq \frac{59 h^8}{152,409,600} \left[ (b-a)^2 + h(b-a) \right]. \quad \ldots\ldots (7.11)$$
This method will henceforth be referred to as $\mathcal{N}$-5. Experimental results to be given in the last chapter, will reveal that the method $\mathcal{N}$-5 is best so far to solve (7.4).

7.5 THE CASE WHEN $f(x) \equiv f$ (constant)

The error bounds (for the boundary value problem (7.4)) given by (7.9), (7.10) and (7.11) are obtained on the assumption that $f(x) \leq 0$, $F(x) > 0$ over $[a,b]$.

We have also remarked that the above conditions are automatically satisfied if $f(x)$ is a real negative constant. We will now show that error bounds of the form (7.9) and (7.11) could be obtained if the function $f(x)$ is constant $= f$ (say) over $[a,b]$ (regardless of the sign of $f$).

The matrix $A$ introduced in (7.8) could now be expressed as a product of $P_N^*$ ($\mu$) and $Q$. The matrix $Q$ has already been introduced in Section 2.4 and

$$P_N^*(\mu) = \begin{bmatrix}
\mu & -1 & & & \\
-1 & \mu & -1 & & \\
& & \ddots & \ddots & \ddots \\
& & & -1 & \mu & -1 \\
& & & & & -\alpha & \delta
\end{bmatrix} \tag{7.12}$$

Here

$$\begin{align*}
\lambda &= 1 - h^2 \beta_2 f^2 < 1, \\
\mu &= (2 + h^2 \beta_1 f^2)/\lambda > 2, \\
-\alpha &= (1 + \frac{h^2}{2} f + \frac{h^4}{10} f^2 + \frac{h^6}{120} f^3)/\lambda, \\
\text{and } \delta &= (1 - \frac{h^2}{2} f + \frac{h^4}{10} f^2 - \frac{h^6}{120} f^3)/\lambda. \tag{7.13}
\end{align*}$$
Let $D_n$ denote the determinant of the $n$th principal minor of the matrix $P_N^*(\mu)$, $(n = 1, 2, \ldots, N-1)$, then $D_n$ satisfy the A.E. (2.13a).

Hence

$$D_n = \frac{\sinh(n+1)\theta}{\sinh\theta} \quad \mu = 2 \cosh\theta > 2,$$

$(n = 0, 1, \ldots, N-1)$, see (2.13b).

Also, if $\phi_N$ denotes the determinant of the matrix $P_N^*(\mu)$, then

$$\phi_N = b D_{N-1} - a D_{N-2}.$$

Let

$$P_N^{*-1}(\mu) = (p_{ij})$$

then we have

$$
\begin{align*}
D_{ij} &= \begin{cases} 
D_{j-1} & \frac{\phi_{n-j}}{\phi_n}, \ j \leq i \\
p_{ji} & \phi_j, \ j > i 
\end{cases} \\
p_{N,j} &= \frac{a D_{j-1}}{\phi_N}, \quad (j = 1, 2, \ldots, N-1) \\
p_{i,N} &= \frac{D_{i-1}}{\phi_N}, \quad (i = 1, 2, \ldots, N)
\end{align*}
$$

...(7.15)

The above results can be obtained by following the usual analysis used in Section 2.4, for proving the corresponding results for the matrix $P_N(\mu)$. Now let $R_i$ denote the sum of the absolute values of the elements of the $i$th row in the matrix $P_N^{*-1}(\mu)$ i.e.

$$R_i = \sum_{j=1}^{N} |p_{ij}|,$$

(note that the inverse of the matrix $P_N^{*}(\mu)$ is not necessarily a monotone matrix unless $f \leq 0$), then we have
\[
\begin{align*}
R_1 &= \frac{1}{\phi_N} \sum_{j=0}^{N-1} |\phi_j| \\
R_i &= \frac{\phi_{N-i}}{\phi_N} \sum_{j=1}^{i} D_{j-1} + \frac{D_{i-1}}{\phi_N} \sum_{j=i+1}^{N} |\phi_{N-j}|, \quad (i = 2, 3, \ldots, N-1), \quad \phi_0 = 0, \\
R_N &= \left( a \sum_{j=0}^{N-2} D_j + D_{N-1} \right) / |\phi_N|.
\end{align*}
\]

We can also prove using (2.13e) that
\[
R_N = \frac{a (D_{N-1} - D_{N-2} - 1) + (\mu-2) D_{N-1}}{(\mu-2) |\phi_N|}.
\]

We finally obtain the norm of the matrix \( P_{N}^{*} \( u \) \) viz.
\[
\| P_{N}^{*} \( u \) \| = \max_{1 \leq i \leq N} \left( \frac{R_i}{\lambda} \right) = \tilde{N} \quad \text{(say)}.
\]

The quantity \( \tilde{N} \) can only be computed numerically, using the expressions for \( R_i \) \( (i = 1, 2, \ldots, N) \) given by (7.16). Now using (7.8), we obtain
\[
\| e \| \leq h^6 M_6 \tilde{N}/240.
\]

The boundary value technique described in Theorem 7.2 to solve an initial value problem of the first order where the error estimate is given by (7.18) will henceforth be referred to as \( M^{*}-1 \). Numerical results given in the next chapter reveal that the error estimate based on \( M^{*}-1 \) is better than the one based on \( M-1 \).
8.1 **LINEAR BOUNDARY VALUE PROBLEMS OF CLASS M**

In this section, we consider linear boundary value problems of class M. The methods M-1, M-2(c), M-3(c), M-4, M-5 and M-6 introduced earlier will be used for obtaining numerical solutions and theoretical error bounds. M-6 will only be used to solve linear problems with constant coefficients. The following D.E.'s are chosen to present the numerical results in order to study the comparative merits of methods from M-1 to M-5.

(A) \[ x^2 y'' - 2y = -x, \quad y(2) = y(3) = 0 \]
with \[ y(x) = \frac{1}{38} (19x - 5x^2 - \frac{36}{x}) \]

(B) \[ x^2 y'' - \frac{3}{4} y = x^2, \quad y(1) = -2.2, \quad y(2) = 3.2 \]
with \[ y(x) = 0.8 x^2 + (x^2-4)/\sqrt{x} \]

(C) \[ y'' - (1+x^2) y = 0, \quad y(0) = 1, \quad y(2) = e^2 \]
with \[ y(x) = \exp \left( \frac{x^2}{2} \right) \]

(D) \[ y'' - y = x^2 - 2, \quad y(0) = 0, \quad y(1) = 1 \]
with \[ y(x) = (2 \sinh x / \sinh 1) - x^2 \]

(E) \[ y'' - y = -4 x e^x, \quad y(0) = y(1) = 0 \]
with \[ y(x) = x(1-x) e^x \]

(F) \[ y'' + y = -x, \quad y(0) = y(2) = 0 \]
with \[ y(x) = (b \sin x / \sin b) - x \]

Please note that the problem (F) is not of class M, but has been included for the sake of comparison. We now discuss problem (A)
in detail. The true solution of (A) is used to check the accuracy of the numerical solution. We also have

\[
y^{(n)}(x) = \frac{18}{19} \frac{(-1)^{n+1}}{n!} x^{n+1}, \quad n \geq 3.
\]
Thus \( M_n = \frac{18 \times n!}{(19 \times 2^{n+1})} \), \( n \geq 3 \). Also experimentally it has been observed that the optimum choice of 'c' in M-2(c) and M-3(c) depends on 'h' to a very great extent. Sometimes the behaviour of this dependence varies from problem to problem. In some cases it has been observed that optimum 'c' remains constant while in others it either decreases or increases with decreasing 'h', till we hit the round-off region. Table VI, for instance, displays how the optimum choice of 'c' varies with 'h' in method M-3(c).

**TABLE VI**

<table>
<thead>
<tr>
<th>( \log_2 (1/2h) )</th>
<th>( A )</th>
<th>( B )</th>
<th>( C )</th>
<th>( D )</th>
<th>( E )</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4.1</td>
<td>5.9</td>
<td>5.8</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5.3</td>
<td>6.3</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4.6</td>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3.9</td>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td></td>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td></td>
<td></td>
<td>Round off Region</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
The computational results are summarized in Table VII for problem (A).

TABLE VII
EXPERIMENTS WITH PROBLEM (A) WITH h=1/8

<table>
<thead>
<tr>
<th>Method</th>
<th>Max. Abs. Error</th>
<th>Estimated Error</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-1</td>
<td>0.17 E-6</td>
<td>0.66 E-6</td>
<td>0.27</td>
</tr>
<tr>
<td>M-2(5)</td>
<td>0.49 E-7</td>
<td>0.14 E-5</td>
<td>0.04</td>
</tr>
<tr>
<td>M-3(4)</td>
<td>0.97 E-9</td>
<td>0.19 E-7</td>
<td>0.05</td>
</tr>
<tr>
<td>M-4</td>
<td>0.65 E-9</td>
<td>0.34 E-8</td>
<td>0.19</td>
</tr>
<tr>
<td>M-5</td>
<td>0.15 E-11</td>
<td>0.11 E-10</td>
<td>0.14</td>
</tr>
</tbody>
</table>

These computations were performed on IBM 7040 at the University of British Columbia, Vancouver. In order to reduce round-off error to a minimum, all calculations were made using double precision arithmetic. The tabulated results in this chapter were all based on double precision arithmetic.

We plotted curves of $\log_{10}|\text{Error}|$ versus $\log_2(1/2h)$ and also $\log_{10}|\text{Error}|$ versus $\log_2 \text{Cost}$ under two different assumptions given below by (8.1). The curves which we have plotted to study the variation of $\log_{10}|\text{Error}|$ versus $\log_2 (1/2h)$ (or $\log_{10}|\text{Error}|$ versus $\log_2 \text{Cost}$) are based on methods M-1, M-3(c) and M-5 respectively. Also in order to estimate the cost of computation, the following two different assumptions are made.

8.1 (i) Assume that the time taken by the machine in calculating the elements of the associated matrix and solving the system of linear algebraic equations is negligible as compared with the time taken in
evaluating \( g(x) \), \( s(x) \) and their higher derivatives at the grid points \( x_n \) (\( n = 0, 1, \ldots, N+1 \)).

or 8.1 (ii) Assume that the time taken by the machine in evaluating \( g(x) \), \( s(x) \) and their derivatives at the grid points is negligible as compared with the time taken in evaluating the elements of the associated matrix and solving the system of linear algebraic equations.

The accompanying graphs show the variation of \( \log_{10}|\text{Error}| \) versus \( \log_2 (1/2h) \) for the problems (A) and (D), (see graphs 1 and 7, respectively). The graphs of \( \log_{10}|\text{Error}| \) versus \( \log_2 \) Cost are also given for the same problems. In this case, the only difference is that the graphs of methods M-3(c) and M-5 are shifted to the left. The graphs for the other problems are similar to those of (A) and (D). These curves indicate that in all cases that we tried, the method M-5 with \( e = 0(h^8) \) is superior to the rest of the techniques.

Thus in comparison to M-1, M-2(c), M-3(c) or M-4 we will recommend M-5 to solve a linear boundary value problem provided analytic expressions are given for functions \( g(x) \) and \( s(x) \), so that the required higher derivatives of these functions at grid points may be computed easily. But if we do not have analytic expressions for \( g(x) \) and \( s(x) \), then the only choice of method will be M-1 or M-2(c).

For the sake of comparison, we also solved the above problems using single precision arithmetic. We have also plotted the curves for the problem (A), (see Figure 4, 5 and 6 respectively). The conclusions are slightly different from those obtained above using double precision arithmetic. Here the curves \( \log_{10}|\text{Error}| \) versus \( \log_2 (1/2h) \) show (see
Problem (A)

\[ 0 = (y')^2 - 2y + x = 0 \]
Problem (A)
Assumption 8.1(i)
Figure 3
FIGURE 4
Problem (A)
(Single Precision)

FIGURE 5
Assumption 8.1(i)
Problem (A)
Assumption (ii)
(Single Precision)
Problem (D)

\[ y'' - y = x^2 - 2 \]

\[ y(0) = 0, \ y(1) = 1 \]
FIGURE 8

Problem (D)

Assumption 8.1(i)
Figure 4 for problem (A)) that as \( h \) decreases, all methods become equally accurate. M-5 has no advantage either over M-3(c) or M-1. We hit the round-off region even before reaching \( h = 1/8 \), see Figure 4.

The minimum in all the three curves is attained for values of 
\[ \log_2 \left( \frac{1}{2h} \right) \leq 3. \]
However, the curves of \( \log_{10} |\text{Error}| \) versus \( \log_2 \text{Cost} \) (see Figures 5 and 6) indicate that M-5 is better than M-1 or M-3(c) under either of the two assumptions. It should be noted again that if functions \( g(x) \) and \( s(x) \) are given in tabular form, then the only choice of method will be either M-1 or M-2(c). In both these methods, the resulting error is \( O(h^4) \). Since M-1 is easy to use, so it is preferred in comparison to M-2(c).

Before we close this section, we would like to mention that the number of function evaluations is \( 2N+4 \), \( 2N+12 \) and \( 6N+12 \) in methods M-1, M-3(c) and M-5 respectively. In M-1 we evaluate \( g(x) \), \( s(x) \) at \( x_n \) \( (n = 0,1,...,N+1) \); in M-3(c) we evaluate \( g(x) \), \( s(x) \) at \( x_n \) \( (n = 0,1,...,N+1) \) and \( g'(x) \), \( s'(x) \) at the points \( x_j \) \( (j = 0,1,N,N+1) \); finally in M-5 we evaluate \( g(x) \), \( g'(x) \), \( g''(x) \), \( s(x) \), \( s'(x) \) and \( s''(x) \) at all the \( N+2 \) grid points.

8.2 LINEAR PROBLEMS OF CLASS M WITH CONSTANT COEFFICIENT

We choose problem (D) for illustration, since the true solution of (D) is
\[ y(x) = \left( \frac{2 \sinh x}{\sinh 1} \right) - x^2, \]
hence
\[ y^{(m)}(x) = \begin{cases} 
2 \sinh x/\sinh 1, & m \text{ even and } > 2 \\
2 \cosh x/\sinh 1, & m \text{ odd and } > 2 
\end{cases}. \]

Thus
\[ M_n = \begin{cases} 
2, & n \text{ even and } > 2 \\
2 \coth 1 = 2.626, & n \text{ odd and } > 2 
\end{cases}. \]
over the interval [0,1].

Table VIII shows the ratio between the maximum absolute error in numerical solution to its theoretical estimate.

**TABLE VIII**

<table>
<thead>
<tr>
<th>Method</th>
<th>Maximum Abs. Error</th>
<th>Estimated Error</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-1</td>
<td>0.11 E-6</td>
<td>0.23 E-6</td>
<td>0.47</td>
</tr>
<tr>
<td>M-2(3)</td>
<td>0.23 E-7</td>
<td>0.97 E-6</td>
<td>0.02</td>
</tr>
<tr>
<td>M-3(4)</td>
<td>0.66 E-10</td>
<td>0.51 E-9</td>
<td>0.13</td>
</tr>
<tr>
<td>M-4</td>
<td>0.38 E-10</td>
<td>0.91 E-10</td>
<td>0.42</td>
</tr>
<tr>
<td>M-5</td>
<td>0.50 E-14</td>
<td>0.12 E-13</td>
<td>0.44</td>
</tr>
</tbody>
</table>

The conclusions are obvious enough that in case of two point linear boundary value problem of class M with constant coefficients, we will recommend M-5 (see Figures 7, 8 and 9). In fact M-6 is superior to M-5 and yields very good results even for large values of step-size h. The resulting error in M-6 is $O(h^{12})$ while in M-5 it is $O(h^8)$. We solved the problem (D) by taking the other boundary at the point $b = 1, 2, 4, 8$ and 16 with $h = 1/2$ in each case using methods M-5 and M-6 respectively. The maximum absolute errors are tabulated in Table IX. The figures listed do justify the claim that method M-6 yields better results than M-5 for large step-size and over greater range of integration.
TABLE IX

EXPERIMENTS WITH PROBLEM (D) WITH $h = 1/2$ USING M-5 and M-6

<table>
<thead>
<tr>
<th>$b$</th>
<th>M-5</th>
<th>M-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.30 E-9</td>
<td>0.39 E-15</td>
</tr>
<tr>
<td>2</td>
<td>0.23 E-8</td>
<td>0.49 E-14</td>
</tr>
<tr>
<td>4</td>
<td>0.25 E-7</td>
<td>0.54 E-13</td>
</tr>
<tr>
<td>8</td>
<td>0.14 E-5</td>
<td>0.27 E-11</td>
</tr>
<tr>
<td>16</td>
<td>0.41 E-2</td>
<td>0.82 E- 8</td>
</tr>
</tbody>
</table>

8.3 We will now compare methods M-AH (developed by Aziz and Hubbard [2] and referred earlier in Section 2.4) and M-BH (developed by Bramble and Hubbard mentioned in Section 3.4) with our techniques M-1, M-2(c), M-3(c) and M-5. In M-AH the D.E. (2.21) is replaced by a set of $n$ simultaneous linear equations

\[(1 - \frac{h^2}{12} g_n)y_n + (2 + \frac{10h^2}{12} g_{n+1})y_{n+1} + (-1 + \frac{h^2}{12} g_{n+2})y_{n+2} = h^2 s_{i+1} + \frac{h^4}{12} s_{i+1}' \]

$(n = 0, 1, \ldots, N-1)$.

The resulting error is $O(h^4)$ as mentioned earlier, (2.22).

Now we tabulate maximum absolute errors in the entire range of integration for the problem (B) for a series of values of $h$ and $x_{N+2} = b$.

A glance at Table X shows that method M-5 should be preferred over M-1, M-AH, M-BH, M-2(c) or M-3(c). Even M-3(c) is better than M-1, M-AH, M-BH and M-2(c).

The methods M-1 and M-AH become identical if $s(x)$ is a constant. That is why in Table X the entries in columns M-1 and M-AH are the same.
However, M-AH gives better results than M-1 if \( s(x) \) is not a constant.

In Table XI we display the results obtained by solving the problem (E) for a series of values of \( b \) and \( h \) using methods M-1 and M-AH.

**TABLE X**

EXPERIMENTS WITH PROBLEM (B) VIZ. \( x^2 y'' - \frac{3}{4} y = x^2, y(1) = 2.2, \)

\( y(b) = .8 b^2 + (b^2 - 4)/\sqrt{b} \) WITH \( y(x) = .8 x^2 + (x^2 - 4)/\sqrt{x} \).

<table>
<thead>
<tr>
<th>b</th>
<th>h</th>
<th>M-1</th>
<th>M-AH</th>
<th>M-BH</th>
<th>M-2(c)</th>
<th>M-3(c)</th>
<th>M-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1/4</td>
<td>0.15 E-3</td>
<td>0.15 E-3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.18 E-6</td>
</tr>
<tr>
<td></td>
<td>1/8</td>
<td>0.10 E-4</td>
<td>0.10 E-4</td>
<td>0.26 E-2</td>
<td>0.42 E-5</td>
<td>0.16 E-6</td>
<td>0.83 E-9</td>
</tr>
<tr>
<td></td>
<td>1/16</td>
<td>0.65 E-6</td>
<td>0.65 E-6</td>
<td>0.27 E-3</td>
<td>0.10 E-6</td>
<td>0.31 E-8</td>
<td>0.34 E-11</td>
</tr>
<tr>
<td></td>
<td>1/32</td>
<td>0.40 E-7</td>
<td>0.40 E-7</td>
<td>0.23 E-4</td>
<td>0.22 E-8</td>
<td>0.53 E-10</td>
<td>0.18 E-13</td>
</tr>
<tr>
<td>3</td>
<td>1/4</td>
<td>0.22 E-3</td>
<td>0.22 E-3</td>
<td>0.18 E-1</td>
<td>0.14 E-3</td>
<td>0.85 E-5</td>
<td>0.21 E-6</td>
</tr>
<tr>
<td></td>
<td>1/8</td>
<td>0.14 E-4</td>
<td>0.14 E-4</td>
<td>0.27 E-2</td>
<td>0.44 E-5</td>
<td>0.21 E-6</td>
<td>0.99 E-9</td>
</tr>
<tr>
<td></td>
<td>1/16</td>
<td>0.89 E-6</td>
<td>0.89 E-6</td>
<td>0.28 E-3</td>
<td>0.11 E-6</td>
<td>0.40 E-8</td>
<td>0.41 E-11</td>
</tr>
<tr>
<td></td>
<td>1/32</td>
<td>0.56 E-7</td>
<td>0.56 E-7</td>
<td>0.24 E-4</td>
<td>0.86 E-8</td>
<td>0.68 E-10</td>
<td>0.11 E-12</td>
</tr>
<tr>
<td>5</td>
<td>1/4</td>
<td>0.25 E-3</td>
<td>0.25 E-3</td>
<td>0.19 E-1</td>
<td>0.14 E-3</td>
<td>0.97 E-5</td>
<td>0.23 E-6</td>
</tr>
<tr>
<td></td>
<td>1/8</td>
<td>0.16 E-4</td>
<td>0.16 E-4</td>
<td>0.27 E-2</td>
<td>0.45 E-5</td>
<td>0.24 E-6</td>
<td>0.11 E-8</td>
</tr>
<tr>
<td></td>
<td>1/16</td>
<td>0.10 E-5</td>
<td>0.10 E-5</td>
<td>0.29 E-3</td>
<td>0.11 E-6</td>
<td>0.45 E-8</td>
<td>0.44 E-11</td>
</tr>
<tr>
<td></td>
<td>1/32</td>
<td>0.65 E-7</td>
<td>0.65 E-7</td>
<td>0.24 E-4</td>
<td>0.35 E-7</td>
<td>0.76 E-10</td>
<td>0.13 E-11</td>
</tr>
<tr>
<td>9</td>
<td>1/4</td>
<td>0.27 E-3</td>
<td>0.27 E-3</td>
<td>0.19 E-1</td>
<td>0.14 E-3</td>
<td>0.10 E-4</td>
<td>0.24 E-6</td>
</tr>
<tr>
<td></td>
<td>1/8</td>
<td>0.17 E-4</td>
<td>0.17 E-4</td>
<td>0.28 E-2</td>
<td>0.45 E-5</td>
<td>0.25 E-6</td>
<td>0.11 E-8</td>
</tr>
<tr>
<td></td>
<td>1/16</td>
<td>0.11 E-5</td>
<td>0.11 E-5</td>
<td>0.29 E-3</td>
<td>0.16 E-6</td>
<td>0.47 E-8</td>
<td>0.43 E-11</td>
</tr>
<tr>
<td></td>
<td>1/32</td>
<td>0.69 E-7</td>
<td>0.69 E-7</td>
<td>0.24 E-4</td>
<td>0.13 E-6</td>
<td>0.78 E-10</td>
<td>0.19 E-10</td>
</tr>
<tr>
<td>17</td>
<td>1/4</td>
<td>0.27 E-3</td>
<td>0.27 E-3</td>
<td>0.19 E-1</td>
<td>0.14 E-3</td>
<td>0.10 E-4</td>
<td>0.24 E-6</td>
</tr>
<tr>
<td></td>
<td>1/8</td>
<td>0.18 E-4</td>
<td>0.18 E-4</td>
<td>0.28 E-2</td>
<td>0.45 E-5</td>
<td>0.25 E-6</td>
<td>0.11 E-8</td>
</tr>
<tr>
<td></td>
<td>1/16</td>
<td>0.11 E-5</td>
<td>0.11 E-5</td>
<td>0.29 E-3</td>
<td>0.51 E-6</td>
<td>0.48 E-8</td>
<td>0.60 E-10</td>
</tr>
<tr>
<td></td>
<td>1/32</td>
<td>0.71 E-7</td>
<td>0.71 E-7</td>
<td>0.24 E-4</td>
<td>0.49 E-6</td>
<td>0.12 E-9</td>
<td>0.18 E-9</td>
</tr>
</tbody>
</table>

**NOTE:** The '-' in Table X indicates that the method is not applicable for that value of \( h \).
### TABLE XI

**EXPERIMENTS WITH PROBLEM (E) VIZ.** $y'' - y = -4 xe^x$, $y(0) = 0$,

$y(b) = b(1-b)e^b$ WITH $y(x) = x(1-x)e^x$.

**Maximum Abs. Error in Methods**

<table>
<thead>
<tr>
<th>$b$</th>
<th>h'</th>
<th>M-1</th>
<th>M-AH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/4</td>
<td>0.93E-4</td>
<td>0.12E-5</td>
</tr>
<tr>
<td></td>
<td>1/8</td>
<td>0.59E-5</td>
<td>0.81E-7</td>
</tr>
<tr>
<td></td>
<td>1/16</td>
<td>0.37E-6</td>
<td>0.53E-8</td>
</tr>
<tr>
<td></td>
<td>1/32</td>
<td>0.23E-7</td>
<td>0.33E-9</td>
</tr>
<tr>
<td>2</td>
<td>1/4</td>
<td>0.69E-3</td>
<td>0.82E-4</td>
</tr>
<tr>
<td></td>
<td>1/8</td>
<td>0.43E-4</td>
<td>0.52E-5</td>
</tr>
<tr>
<td></td>
<td>1/16</td>
<td>0.27E-5</td>
<td>0.33E-6</td>
</tr>
<tr>
<td></td>
<td>1/32</td>
<td>0.17E-6</td>
<td>0.20E-7</td>
</tr>
<tr>
<td>4</td>
<td>1/4</td>
<td>0.11E-1</td>
<td>0.33E-2</td>
</tr>
<tr>
<td></td>
<td>1/8</td>
<td>0.68E-3</td>
<td>0.21E-3</td>
</tr>
<tr>
<td></td>
<td>1/16</td>
<td>0.42E-4</td>
<td>0.13E-4</td>
</tr>
<tr>
<td></td>
<td>1/22</td>
<td>0.27E-5</td>
<td>0.80E-6</td>
</tr>
</tbody>
</table>

8.4 We will now consider boundary value problems which are not of class M. In particular, we consider

$$y'' + E^2y = \sin x, \quad (E^2 \neq 0,1 \text{ and } E \neq \frac{mn}{b}) \quad \ldots (8.2)$$

where $m$ is an integer.

\[
\begin{aligned}
    y(0) &= 1 \\
    y(b) &= \sin(Eb) + \cos(Eb) + \frac{\sin b}{(E^2-1)}
\end{aligned}
\]
with \( y(x) = \sin (Ex) + \cos (Ex) + \frac{\sin x}{E^2-1} \).

It is easy to verify that
\[
M_n \leq \sqrt{2} E^n + \frac{1}{(E^2-1)} , \text{ over } [a,b].
\]

We solve (8.2) for \( E = 2, 3, 4, 5 \) and 6; \( b = 1, 2, 4, 8 \) and 16 with \( h = 1/4, 1/8, 1/16 \) and 1/32 respectively. We use methods M-1 and M-5. The error estimates are based on Theorem 5.1. The results of D.E. (8.2) for \( E = 2 \) and \( h = 1/8 \) are given in Table XII using method M-1, and for \( h = 1/4 \) using method M-5.

**TABLE XII**

**EXPERIMENTS WITH PROBLEM (8.2)**

<table>
<thead>
<tr>
<th>Method</th>
<th>h</th>
<th>b</th>
<th>Max. Abs. Error</th>
<th>Estimated Error</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-1</td>
<td>1/8</td>
<td>1</td>
<td>0.18 E-4</td>
<td>0.45 E-4</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.28 E-4</td>
<td>0.12 E-3</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0.64 E-4</td>
<td>0.18 E-3</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>0.44 E-3</td>
<td>0.13 E-2</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16</td>
<td>0.46 E-3</td>
<td>0.13 E-2</td>
<td>0.35</td>
</tr>
<tr>
<td>M-5</td>
<td>1/4</td>
<td>1</td>
<td>0.37 E-8</td>
<td>0.73 E-8</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.70 E-8</td>
<td>0.20 E-7</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>0.11 E-7</td>
<td>0.33 E-7</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>0.88 E-7</td>
<td>0.24 E-6</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16</td>
<td>0.86 E-7</td>
<td>0.25 E-6</td>
<td>0.34</td>
</tr>
</tbody>
</table>
Method M-5 gave better results as was expected even for large values of the step size $h$, therefore we have tabulated values of the max. abs. error in case of method M-5 for $h = 1/4, E = 2$.

Our conclusion is that the formula (5.6) gives a useful estimate of the discretization error. The ratio of "max. abs. error" to the estimated error is less than 1. Also this ratio stays fairly constant as $b$ increases, so that (5.6) can be regarded to reflect well the variation in the actual error (even in large range of integration). We have not tabulated $\mu = (24 - 10h^2g)/(12 + h^2g)$ because it is not sensitive to changes in 'h'. Also $\mu$ does not depend on 'b' at all, hence we are satisfied with tabulating $b$ and $h$ in Table XII. Thus it is clear that the inequality (5.6) for the estimation of $||e||$ in method M-1 is quite useful, and assists one in the selection of 'h' and we can rely upon it for a fairly large range of integration.

8.5 NONLINEAR PROBLEMS

Iterative procedures based on A.E.'s referred to in methods M-1, M-2(c) and M-4 will be studied for solving nonlinear boundary value problems of class M. The following D.E. (8.3) is chosen for experimentation.

$$y'' = \mu y^2, \quad y(0) = 4, \quad y(1) = 1, \quad \mu > 0 . \quad \ldots (8.3)$$

The closed solutions $y(x,\mu)$ of D.E. (8.3) are not known except for $\mu = 0$ and $\mu = 3/2$. We notice that

$$y(x,0) = 4-3x \text{ and } y(x,3/2) = 4/(1+x)^2 .$$

The solution $y(x,0)$ is taken to be an initial approximation i.e. $Q(x) = 4-3x$ according to the notations introduced in Chapter VI. We solve
problem (8.3) for \( \mu = 3/2 \). The criterion for stopping iterations is that the remainder as defined by (6.8) be such that

\[
||r(y^{(j)})|| < 10^{-10}, \text{ for } j = 1, 2, \ldots, N.
\]

The results are tabulated in Table XIII for \( h = 1/10 \). In method M-2(c) results are given for \( c = 3 \), although we tried the values of \( c \) in the range \(-5 \leq c \leq 18\) and varied \( c \) each time by an amount of .1. The best results were obtained for \( c = 3 \). The results based on M-2(3) are more or less identical to those based on M-1. Now we solve the same problem using method M-4, as described briefly in Section 6.5.

The results given in Table XIII also confirm that the order of the convergence of Newton's method is 2 i.e. the method is quadratically convergent or that the number of correct decimal places is doubled at each step. The figures given also indicate that the results are best using method M-4 in the sense that "max. abs. error" is least in that case. We notice that the maximum absolute error attains its minimum value and then does not change, however the quantity \( ||r(y)|| \), the norm of the remainder vector goes on decreasing as the number of iterations increases, till we hit the round-off region. In practical cases however, we will have no way of obtaining "max. abs. error", we have to stop iterations when \( ||r(y)|| \) becomes less than a preassigned small number \( \varepsilon > 0 \).

Finally we notice that M-2(c) has no advantage over M-1. In M-2(c), we have to decide the optimum value of the parameter \( c \) and then the amount of computation in M-2(c) is more than M-1. Therefore the method M-2(c) is dropped in comparison to M-1. Between M-1 and M-4, M-4 should be used to solve a nonlinear boundary value problem of class M,
because "max. abs. error" is smaller in that case, although the number of iterations required in both cases is 4 before \(||r(y)|| < 10^{-10}\). The only advantage which M-1 has over M-4 is that it is very easy to apply.

Besides the problem (8.3), we also tried the problem

\[ y'' = -2 + \mu \sinh y, \quad y(0) = y(1) = 0, \quad \mu > 0. \]

The closed solution \(y(x,\mu)\) of the above problem is not known except for \(\mu = 0\) and \(y(x,0) = x(1-x)\). This solution was taken to be an initial approximation to the actual solution of the problem. After only two steps the iterations were stopped in this case. The conclusions drawn were identical to those drawn for the problem (8.3).

**TABLE XIII**

| Method  | No. of Iterations | ||r(y)||       | Max. Abs. Error |
|---------|-------------------|---------------|-----------------|
| M-1     | 0                 | 0.21          |                 |
|         | 1                 | 0.76 E-2      | 0.43 E-1        |
|         | 2                 | 0.27 E-4      | 0.11 E-3        |
|         | 3                 | 0.35 E-9      | 0.64 E-4        |
|         | 4                 | 0.46 E-15     | 0.64 E-4        |
| M-2(3)  | 0                 | 0.87          |                 |
|         | 1                 | 0.37 E-1      | 0.43 E-1        |
|         | 2                 | 0.13 E-3      | 0.15 E-3        |
|         | 3                 | 0.17 E-8      | 0.32 E-4        |
|         | 4                 | 0.12 E-15     | 0.32 E-4        |
| M-4     | 0                 | 0.20          |                 |
|         | 1                 | 0.75 E-2      | 0.42 E-1        |
|         | 2                 | 0.26 E-4      | 0.15 E-3        |
|         | 3                 | 0.31 E-9      | 0.75 E-6        |
|         | 4                 | 0.16 E-14     | 0.75 E-6        |
8.6 INITIAL VALUE PROBLEMS

We finally present numerical results of initial value problems solved by boundary value techniques. The numerical techniques employed are M-1, M-4 and M-5 (introduced in Chapter VII). Most of the experiments we performed in this connection were based on D.E.

\[ y' = sy + \sin(tx) , \quad y(0) = \frac{t}{(s^2 + t^2)} , \quad \ldots (8.4) \]

with \( y(x) = -[s \sin(tx) + t \cos(tx)]/(s^2 + t^2) \). We also have

\[ M_n = |t|^{n}/\sqrt{s^2 + t^2} \]. We chose \( h = 3/32 \), and the "max. abs. error" given in Table XIV is over the range \( 0 \leq x \leq 7.5 \) for \( s > 0 \) and over the range \( 0 \leq x \leq 21 \) for \( s < 0 \). The results are listed in Table XIV below:

<table>
<thead>
<tr>
<th>Method</th>
<th>s</th>
<th>t</th>
<th>Max. Abs. Error</th>
<th>Estimated Error</th>
<th>Ratio × 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-1</td>
<td>-2</td>
<td>1</td>
<td>0.29 E-7</td>
<td>0.32 E-4</td>
<td>0.01</td>
</tr>
<tr>
<td>M-4</td>
<td></td>
<td></td>
<td>0.59 E-11</td>
<td>0.64 E-8</td>
<td>0.01</td>
</tr>
<tr>
<td>M-5</td>
<td></td>
<td></td>
<td>0.31 E-14</td>
<td>0.45 E-12</td>
<td>0.68</td>
</tr>
<tr>
<td>M-1</td>
<td>-1</td>
<td>2</td>
<td>0.27 E-5</td>
<td>0.20 E-2</td>
<td>0.13</td>
</tr>
<tr>
<td>M-4</td>
<td></td>
<td></td>
<td>0.22 E-8</td>
<td>0.16 E-5</td>
<td>0.13</td>
</tr>
<tr>
<td>M-5</td>
<td></td>
<td></td>
<td>0.63 E-12</td>
<td>0.46 E-9</td>
<td>0.13</td>
</tr>
<tr>
<td>M-1</td>
<td>2</td>
<td>1</td>
<td>0.13 E-1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M-4</td>
<td></td>
<td></td>
<td>0.32 E-2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M-5</td>
<td></td>
<td></td>
<td>0.26 E-2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M-1</td>
<td>1</td>
<td>2</td>
<td>3.725</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M-4</td>
<td></td>
<td></td>
<td>0.34 E-2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M-5</td>
<td></td>
<td></td>
<td>0.10 E-5</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Note when \( s > 0 \), the condition of the Theorem 7.2 viz. \( f(x) \leq 0 \), is not satisfied and the expressions for error bounds no longer hold good. That is why in Table XIV the corresponding columns for estimated error and ratio have been left blank. We notice that the max. abs. error in \( \tilde{M}-1 \) (being less accurate) for \( s = 1, t = 2 \) with \( h = 3/32 \) over the range \([0,7.5]\) is quite large so that the corresponding numerical results are of no practical importance. However the max. abs. error equals 0.32 E-3 over the range \([0,3]\). This difficulty can also be overcome by using more accurate techniques viz. \( \tilde{M}-4 \) or \( \tilde{M}-5 \), as is apparent from Table XIV. We mention that we have obtained results for a number of other situations besides those listed in Table XIV.

In addition to D.E. (8.4) we also experimented with the following D.E.'s:

\[
\begin{align*}
y' &= sy + e^{sx} , \quad y(0) = 1 \quad \text{with} \quad y(x) = (1+x) e^{sx}, \\
y' &= 12y - 11 e^x , \quad y(0) = 1 \quad \text{with} \quad y(x) = e^x, \\
x y' &= -y + x^3 , \quad y(1) \quad \text{with} \quad y(x) = (x^3 + x^{-1})/4.
\end{align*}
\]

...(8.5)  
...(8.6)  
...(8.7)

We now use method \( \tilde{M}^*-1 \) and \( \tilde{M}^*-5 \) to get the error bounds for the same problem viz. (8.4). We tabulate the numerical results in Table XV.

We have obtained error bounds for \( f(x) = \text{constant} \) (regardless of the sign of the function) and also the error estimates are better than those given in Table XIV. Note that the analysis of the Section 7.5 is not applicable to the method \( \tilde{M}-4 \) because the \( \Delta \)E.

\[
\Sigma a_i y_{n+i} = h^2 \Sigma b_i y''_{n+1} + h^3 \Sigma c_i y'''_{n+i}
\]

is such that \( \gamma_0 = 1/40, \quad \gamma_1 = 0, \quad \gamma_2 = -1/40 \), hence again the associated matrix \( A \) cannot be factored as a product of \( P_N^*(u) \) and \( Q_* \).
Now we proceed to compare the methods $M^{-1}$, $M^{-4}$ and $M^{-5}$ with some known techniques usually employed for solving (7.1). For the numerical solution of a linear first order D.E. of the form

$$y' = f(x) \ y + g(x), \ y(a) = y_a,$$

Allen and Severn introduce an auxilliary function $z(x)$, connected with $y$ by the equation

$$y = \psi(z,z').$$

The form of $\psi(z,z')$ is to some extent arbitrary, but is generally chosen to be a simple linear combination of $z$ and $z'$ such that the substitution $y = \psi(z,z')$ in the above linear initial value problem produces a second order D.E. for $z$ in which the coefficient of $z'$ vanishes. This second order D.E. is solved with boundary conditions.

### TABLE XV

**EXPERIMENTS WITH PROBLEM (8.4), USING $M^{-1}$ and $M^{-5}$ WITH $h = 3/32$.**

<table>
<thead>
<tr>
<th>Method</th>
<th>$s$</th>
<th>$t$</th>
<th>Max. Abs. Error</th>
<th>Estimated Error</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M^*-1$</td>
<td>-2</td>
<td>1</td>
<td>0.29 E-7</td>
<td>0.32 E-5</td>
<td>.009</td>
</tr>
<tr>
<td>$M^*-5$</td>
<td></td>
<td></td>
<td>0.31 E-14</td>
<td>0.46 E-13</td>
<td>.067</td>
</tr>
<tr>
<td>$M^*-1$</td>
<td>-1</td>
<td>2</td>
<td>0.27 E-5</td>
<td>0.80 E-3</td>
<td>.003</td>
</tr>
<tr>
<td>$M^*-5$</td>
<td></td>
<td></td>
<td>0.63 E-12</td>
<td>0.18 E-9</td>
<td>.0035</td>
</tr>
<tr>
<td>$M^*-1$</td>
<td>2</td>
<td>1</td>
<td>0.13 E-1</td>
<td>0.24</td>
<td>.055</td>
</tr>
<tr>
<td>$M^*-5$</td>
<td></td>
<td></td>
<td>0.26 E-2</td>
<td>0.25 E-1</td>
<td>.104</td>
</tr>
<tr>
<td>$M^*-1$</td>
<td>.1</td>
<td>2</td>
<td>3.725</td>
<td>0.19 E+3</td>
<td>.002</td>
</tr>
<tr>
<td>$M^*-5$</td>
<td></td>
<td></td>
<td>0.19 E-5</td>
<td>0.54 E-4</td>
<td>.002</td>
</tr>
</tbody>
</table>
\[ y_0 = \psi(z_0, z_0') \text{ and } z_n = \lambda \]

where \( \lambda \) is an arbitrary value of \( z \) at any other point \( x_n \) in the range. The value of \( \lambda \) does not affect the resulting solution for \( y \), which is finally recovered from \( y = \psi(z, z') \). As mentioned earlier in Section 7.1 that no indication was given by Allen and Severn regarding the choice of \( \psi(z, z') \), except when applied to certain partial D.E.'s.

Fox's method is direct and more convenient than that of Allen and Severn. He approximates the derived second order system (7.4) by the A.E. (2.10b) \( (n = 0, 1, \ldots, N-2) \) along with the A.E. which involves the boundary conditions at \( x = b \)

\[-y_{N-1} + y_{N+1} = (h/3) (y'_{N-1} + 4 y_N' + y'_{N+1}) \quad \ldots (8.8)\]

with a T.E. = \((-1/90) h^5 y^{(5)}\).

Obviously the A.E. (7.7) which we use to approximate the second boundary condition of (7.4) at \( x = b \) in \( \tilde{N}-1 \) is more accurate than A.E. (8.8). With the result \( \tilde{N}-1 \) turns out to be better than Fox's method. D.E. (8.6) was considered by Fox [10] to demonstrate the fact that initial value problems could be conveniently solved by boundary value techniques. The maximum absolute error for \( h = .2 \) was found to be 0.10 E-3. However in our case, the max. abs. error turns out to be 0.71 E-5, 0.45 E-7 and 0.39 E-9 using \( \tilde{N}-1, \tilde{N}-4 \) and \( \tilde{N}-5 \) respectively. Also in Fox's method, no error bounds are available, hence we drop it in favour of \( \tilde{N}-1 \). Now we will compare the results of the problem (8.5) for \( s = -1 \) with \( h = 0.1 \) using some of the standard initial value techniques.

The methods listed in Table XVI and referred to as \( w_k(c) \) for the solution of linear D.E. of the first order are based on
\[ y_{n+k} = \frac{\sum_{i=0}^{k-1} (a_i - h^\beta_i f_{n+i}) y_{n+i} + h \sum_{i=0}^{k} \beta_i g_{n+i}}{(1-h^\beta_k f_{n+k})} \] ... (8.9)

Thus we need \( k \) starting values viz. \( y_n, y_{n+1}, \ldots, y_{n+k-1} \) only. If we assume that the D.E. is such that the functions \( f(x) \) and \( g(x) \) are quite complicated, then the total cost of computation will be proportional to the \((2N+2)\) evaluations of functions \( f(x) \) and \( g(x) \) respectively in method \( W_k(c) \). However the example (8.5) whose results we summarize in Table XVI is such that we require only \((N+1)\) evaluations of the exponential function \( e^x \) in all the methods listed therein, except in the fourth order.

Runga-Kutta method, where a total of \( 4N \) evaluations of \( e^x \) are needed.

A close study of the Table XVI shows that the results obtained by using \( \hat{M} \)-1 are better than step-by-step methods \( w_3(\cdot.6) \) and \( w_2(1) \), but the fourth order Runga-Kutta method gives the best results. However, \( \hat{M} \)-4 yields better results than \( w_4(1) \). The methods that we propose viz. \( \hat{M} \)-1, \( \hat{M} \)-4 and \( \hat{M} \)-5 are direct methods while the remaining are step-by-step methods (or iterative in the sense that they are predictor-corrector methods provided we are solving a nonlinear differential system of the first order) and suffer from a disadvantage that they require formulas for starting the solution. Thus, in short, the methods that we propose compare very favourably with the standard initial value techniques. They are better in most cases and are very easy to apply.
<table>
<thead>
<tr>
<th>No.</th>
<th>Method</th>
<th>Order of the Resulting Error in the Method</th>
<th>Maximum Absolute Error</th>
<th>No. of function Evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(w_3(.6)), see Hull and Newberry [15]</td>
<td>(O(h^4))</td>
<td>.81 E-6</td>
<td>2N+2</td>
</tr>
<tr>
<td>2</td>
<td>(w_2(1)), Milne method</td>
<td>(O(h^4))</td>
<td>.82 E-6</td>
<td>2N+2</td>
</tr>
<tr>
<td>3</td>
<td>M-1</td>
<td>(O(h^4))</td>
<td>.22 E-6</td>
<td>2N+10</td>
</tr>
<tr>
<td>4</td>
<td>Fourth order Runge-Kutta method [13,pp.237]</td>
<td>(O(h^4))</td>
<td>.18 E-6</td>
<td>4N</td>
</tr>
<tr>
<td>5</td>
<td>(w_4(0)), Adams type</td>
<td>(O(h^5))</td>
<td>.25 E-6</td>
<td>2N+2</td>
</tr>
<tr>
<td>6</td>
<td>(w_4(1)), [13,(6.6.10)]</td>
<td>(O(h^6))</td>
<td>.53 E-8</td>
<td>2N+2</td>
</tr>
<tr>
<td>7</td>
<td>M-4</td>
<td>(O(h^6))</td>
<td>.73 E-10</td>
<td>4N+4</td>
</tr>
<tr>
<td>8</td>
<td>Method based on A.E. (7.6), (j = 4)</td>
<td>(O(h^8))</td>
<td>.93 E-10</td>
<td>3N+3</td>
</tr>
<tr>
<td>9</td>
<td>M-5</td>
<td>(O(h^8))</td>
<td>.81 E-14</td>
<td>4N+4</td>
</tr>
</tbody>
</table>


7. V.N. FADDEEVA, Computational Methods of Linear Algebra; Dover, New York, 1958.


