

PEARSON

Numerical Methods



Babu Ram

Numerical Methods

Babu Ram

Formerly Dean, Faculty of Physical Sciences, Maharshi Dayanand University, Rohtak



Delhi • Chennai • Chandigarh

Copyright © 2010 Dorling Kindersley (India) Pvt. Ltd

This book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior written consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser and without limiting the rights under copyright reserved above, no part of this publication may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording or otherwise), without the prior written permission of both the copyright owner and the above-mentioned publisher of this book.

ISBN 978-81-317-3221-2

10 9 8 7 6 5 4 3 2 1

Published by Dorling Kindersley (India) Pvt. Ltd., licensees of Pearson Education in South Asia

Head Office: 7th Floor, Knowledge Boulevard, A-8 (A), Sector 62, NOIDA, 201 309, UP, India.
Registered Office: 14 Local Shopping Centre, Panchsheel Park, New Delhi 110 017, India

Typeset by Integra Software Services Pvt. Ltd., Pondicherry, India.
Printed in India at Baba Barkha Nath Printers, New Delhi.

In the Memory of
MY PARENTS
Smt. Manohari Devi and Sri. Makhan Lal

This page is intentionally left blank

Contents

Preface ix

1 Preliminaries

1

- 1.1 Approximate Numbers and Significant Figures 1
- 1.2 Classical Theorems Used In Numerical Methods 2
- 1.3 Types of Errors 4
- 1.4 General Formula for Errors 5
- 1.5 Order of Approximation 7
- Exercises* 9

2 Non-Linear Equations

11

- 2.1 Classification of Methods 11
- 2.2 Approximate Values of the Roots 11
- 2.3 Bisection Method (Bolzano Method) 12
- 2.4 Regula–Falsi Method 15
- 2.5 Convergence of Regula–Falsi Method 16
- 2.6 Newton–Raphson Method 20
- 2.7 Square Root of a Number Using Newton–Raphson Method 23
- 2.8 Order of Convergence of Newton–Raphson Method 24
- 2.9 Fixed Point Iteration 25
- 2.10 Convergence of Iteration Method 26
- 2.11 Square Root of a Number Using Iteration Method 27
- 2.12 Sufficient Condition for the Convergence of Newton–Raphson Method 28
- 2.13 Newton’s Method for Finding Multiple Roots 29
- 2.14 Newton–Raphson Method for Simultaneous Equations 32
- 2.15 Graeffe’s Root Squaring Method 37
- 2.16 Muller’s Method 41
- 2.17 Bairstow Iterative Method 45
- Exercises* 49

3 Linear Systems of Equations

51

- 3.1 Direct Methods 51
- 3.2 Iterative Methods for Linear Systems 70
- 3.3 The Method of Relaxation 79
- 3.4 Ill-Conditioned System of Equations 82
- Exercises* 82

4 Eigenvalues

and Eigenvectors

85

- 4.1 Eigenvalues and Eigenvectors 85
- 4.2 The Power Method 88
- 4.3 Jacobi’s Method 94
- 4.4 Given’s Method 101
- 4.5 Householder’s Method 109
- 4.6 Eigenvalues of a Symmetric Tri-diagonal Matrix 115
- 4.7 Bounds on Eigenvalues (Gershgorin Circles) 117
- Exercises* 119

5 Finite Differences

and Interpolation

122

- 5.1 Finite Differences 122
- 5.2 Factorial Notation 130
- 5.3 Some More Examples of Finite Differences 132
- 5.4 Error Propagation 139
- 5.5 Numerical Instability 143
- 5.6 Interpolation 143
- 5.7 Use of Interpolation Formulae 162
- 5.8 Interpolation with Unequal-Spaced Points 163
- 5.9 Newton’s Fundamental (Divided Difference) Formula 164
- 5.10 Error Formulae 168
- 5.11 Lagrange’s Interpolation Formula 171

5.12	Error in Lagrange's Interpolation Formula	179
5.13	Hermite Interpolation Formula	180
5.14	Throwback Technique	185
5.15	Inverse Interpolation	188
5.16	Chebyshev Polynomials	195
5.17	Approximation of a Function with a Chebyshev Series	198
5.18	Interpolation by Spline Functions	200
5.19	Existence of Cubic Spline	202
	<i>Exercises</i>	209

6 Curve Fitting 213

6.1	Least Square Line Approximation	213
6.2	The Power Fit $y = ax^m$	219
6.3	Least Square Parabola (Parabola of Best Fit)	221
	<i>Exercises</i>	227

7 Numerical Differentiation 228

7.1	Centered Formula of Order $O(h^2)$	228
7.2	Centered Formula of Order $O(h^4)$	229
7.3	Error Analysis	230
7.4	Richardson's Extrapolation	231
7.5	Central Difference Formula of Order $O(h^4)$ for $f''(x)$	234
7.6	General Method for Deriving Differentiation Formulae	235
7.7	Differentiation of a Function Tabulated in Unequal Intervals	244
7.8	Differentiation of Lagrange's Polynomial	245
7.9	Differentiation of Newton Polynomial	246
	<i>Exercises</i>	250

8 Numerical Quadrature 252

8.1	General Quadrature Formula	252
8.2	Cote's Formulae	256
8.3	Error Term in Quadrature Formula	258
8.4	Richardson Extrapolation (or Deferred Approach to the Limit)	263
8.5	Simpson's Formula with End Correction	265
8.6	Romberg's Method	267

8.7	Euler–Maclaurin Formula	277
8.8	Double Integrals	279
	<i>Exercises</i>	285

9 Difference Equations 288

9.1	Definitions and Examples	288
9.2	Homogeneous Difference Equation with Constant Coefficients	289
9.3	Particular Solution of a Difference Equation	293
	<i>Exercises</i>	299

10 Ordinary Differential Equations 301

10.1	Initial Value Problems and Boundary Value Problems	301
10.2	Classification of Methods of Solution	301
10.3	Single-Step Methods	301
10.4	Multistep Methods	330
10.5	Stability of Methods	344
10.6	Second Order Differential Equation	349
10.7	Solution of Boundary Value Problems by Finite Difference Method	352
10.8	Use of the Formula $\delta^2 y_n = h^2 \left(1 + \frac{1}{12} - \frac{\delta^4}{240} + \dots \right) f_n$ to Solve Boundary Value Problems	355
10.9	Eigenvalue Problems	357
	<i>Exercises</i>	360

11 Partial Differential Equations 363

11.1	Formation of Difference Equation	363
11.2	Geometric Representation of Partial Difference Quotients	364
11.3	Standard Five Point Formula and Diagonal Five Point Formula	365
11.4	Point Jacobi's Method	366
11.5	Gauss–Seidel Method	366
11.6	Solution of Elliptic Equation by Relaxation Method	376
11.7	Poisson's Equation	379
11.8	Eigenvalue Problems	383

- 11.9 Parabolic Equations 389
11.10 Iterative Method to Solve Parabolic Equations 399
11.11 Hyperbolic Equations 402
Exercises 407
- 12.7 Control (Selection) Statements 419
12.8 Structure of a C Program 423
12.9 Programs of Certain Numerical Methods in C Language 425

12 Elements of C Language 412

- 12.1 Programming Language 412
12.2 C Language 412
12.3 C Tokens 412
12.4 Library Functions 416
12.5 Input Operation 417
12.6 Output Operation 418

Appendix

- Model Paper 1 A-1
Model Paper 2 A-11
Model Paper 3 A-18
Model Paper 4 A-28
Model Paper 5 A-38
Bibliography B-1
Index I-1

This page is intentionally left blank

Preface

The present text is intended to provide a fairly substantial ground in interpolation, numerical differentiation and integration, least square approximation, numerical solution of non-linear equations, system of linear equation, ordinary and partial differential equations, and eigenvalue problems. Various numerical methods have been described technically and their convergence and error propagation have been studied. Most of these methods are implemented efficiently in computers. Programs, in C language, for important and typical methods have been provided in the last chapter. Sufficient number of solved examples has been provided to make the matter understandable to students. As such, the text will meet the requirements of engineering and science students at the undergraduate and post-graduate level.

I wish to record my thanks to my family members for their encouragement and to Sushma S. Pradeep for assistance in the preparation of this manuscript. I am thankful to the editorial team and specially Anita Yadav, Thomas Mathew and Vamanan Namboodiri of Pearson Education for their continuous support at all levels.

BABU RAM

This page is intentionally left blank

1 Preliminaries

Numerical Analysis is a branch of mathematics in which we analyse and solve the problems that require calculations. The methods (techniques) used for this purpose are called *Numerical Methods* (techniques). These techniques are used to solve algebraic or transcendental equations, an ordinary or partial differential equations, integral equations, and to obtain functional value for an argument in some given interval where some values of the function are given. In numerical analysis we do not strive for exactness and try to device a method that will yield an approximate solution differing from the exact solution by less than a specified tolerance. The approximate calculation is one which involves approximate data, approximate methods, or both. The error in the computed result may be due to errors in the given data and errors of calculation. There is no remedy to the error in the given data but the second kind of error can usually be made as small as we wish. The calculations are carried out in such a way as to make the error of calculation negligible.

1.1 APPROXIMATE NUMBERS AND SIGNIFICANT FIGURES

The numbers of the type 3, 6, 2, 5/4, 7.35 are called *exact numbers* because there is no approximation associated with them. On the other hand, numbers like $\sqrt{2}$, π are exact numbers but cannot be expressed exactly by a finite number of digits when expressed in digital form. Numbers having finite number of digits approximate such numbers. An *approximate number* is a number that is used as an approximation to an exact number and differs only slightly from the exact number for which it stands. For example,

- (i) 1.4142 is an approximate number for $\sqrt{2}$
- (ii) 3.1416 is an approximate number for π
- (iii) 2.061 is an approximate number for 27/13.1.

A *significant figure* is any of the digits 1, 2, ..., 9, and 0 is a significant figure except when it is used to fix the decimal point or to fill the places of unknown or discarded digits. For example, 1.4142 contains five significant figures, whereas 0.0034 has only two significant figures: 3 and 4.

If we attempt to divide 22 by 7, we get

$$\frac{22}{7} = 3.142857\dots$$

In practical computation, we must cut it down to a manageable form such as 3.14 or 3.143. The process of cutting off superfluous digits and retaining as many digits as desired is called *rounding off*. Thus to round off a number, we retain a certain number of digits, counted from the left, and drop the others. However, the numbers are rounded off so as to cause the least possible error.

To round off a number to n significant figures,

- (i) Discard all digits to the right of the n th digit.
- (ii) (a) If the discarded number is less than half a unit in the n th place, leave the n th digit unchanged.
(b) If the discarded number is greater than half a unit in the n th place, increase the n th digit by 1.

- (c) If the discarded number is exactly half a unit in the n th place, increase the n th digit by 1 if it is odd, otherwise leave the n th digit unaltered. Thus, in this case, the n th digit shall be an even number. The reason for this step is that even numbers are more exactly divisible by many more numbers than are odd numbers and so there will be fewer leftover errors in computation when the rounded numbers are left even.

When a given number has been rounded off according to the above rule, it is said to be *correct to n significant figures*.

The rounding off procedure discussed in (i) and (ii) above is called *symmetric round off*. On the other hand, the process of dropping extra digits (without using symmetric round off) of a given number is called *chopping or truncation of number*. For example, if we are working on a computer with fixed word length of seven digits, then a number like 83.7246734 will be stored as 83.72467 by dropping extra digits 3 and 4. Thus, error in the approximation is 0.0000034.

EXAMPLE 1.1

Round off the following numbers correctly to *four significant figures*:

$$81.9773, 48.365, 21.385, 12.865, 27.553.$$

Solution. After rounding off,

$$\begin{aligned} 81.9773 &\text{ becomes } 81.98, \\ 48.365 &\text{ becomes } 48.36, \\ 21.385 &\text{ becomes } 21.38, \\ 12.865 &\text{ becomes } 12.86, \\ 27.553 &\text{ becomes } 27.55. \end{aligned}$$

1.2 CLASSICAL THEOREMS USED IN NUMERICAL METHODS

The following theorems will be used in the derivation of some of the numerical methods and in the study of error analysis of the numerical methods.

Theorem 1.1. (*Rolle's Theorem*). Let f be a function such that

- (i) f is continuous in $[a, b]$
- (ii) f is derivable in (a, b)
- (iii) $f(a) = f(b)$.

Then there exists at least one $\xi \in (a, b)$ such that $f'(\xi) = 0$.

The following version of the Rolle's Theorem will be used in error analysis of *Lagrange's interpolation formula*.

Theorem 1.2. (*Generalized Rolle's Theorem*). Let f be n times differentiable function in $[a, b]$. If f vanishes at $(n + 1)$ distinct points x_0, x_1, \dots, x_n in (a, b) , then there exists a number $\xi \in (a, b)$ such that $f^{(n)}(\xi) = 0$.

It follows from Theorem 1.2 that “between any two zeroes of a polynomial $f(x)$ of degree ≥ 2 , there lies at least one zero of the polynomial $f'(x)$.”

Theorem 1.3. (*Intermediate Value Theorem*). Let f be continuous in $[a, b]$ and $f(a) < k < f(b)$. Then there exists a number $\xi \in (a, b)$ such that $f(\xi) = k$.

Theorem 1.4. (*Mean Value Theorem*). If

- (i) f is continuous in $[a, b]$,
- (ii) f is derivable in (a, b) ,

then there exists at least one $\xi \in (a, b)$ such that

$$\frac{f(b) - f(a)}{b - a} = f'(\xi), \quad a < \xi < b.$$

The following theorem is useful in locating the roots of a given equation.

Theorem 1.5. If f is continuous in $[a, b]$ and if $f(a)$ and $f(b)$ are of opposite signs, then there exists at least one $\xi \in (a, b)$ such that $f(\xi) = 0$.

The following theorems of Taylor are frequently used in numerical methods.

Theorem 1.6. (Taylor's Theorem). Let f be continuous and possess continuous derivatives of order n in $[a, b]$. If $x_0 \in [a, b]$ is a fixed point, then for every $x \in [a, b]$, there exists a number ξ lying between x_0 and x such that

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2!}f''(x_0) + \cdots + \frac{(x - x_0)^{n-1}}{(n-1)!}f^{(n-1)}(x_0) + R_n(x),$$

where

$$R_n(x) = \frac{(x - x_0)^n}{n!}f^{(n)}(\xi), \quad x_0 < \xi < x.$$

If $x = x_0 + h$, then we get

$$\begin{aligned} f(x_0 + h) &= f(x_0) + hf'(x_0) + \frac{h^2}{2!}f''(x_0) + \cdots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(x_0) + \frac{h^n}{n!}f^{(n)}(\xi) \\ &= f(x_0) + hf'(x_0) + \frac{h^2}{2!}f''(x_0) + \cdots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(x_0) + O(h^n). \end{aligned}$$

As a corollary to Taylor's Theorem, we have

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \cdots + \frac{x^n}{n!}f^{(n)}(0) + \cdots$$

which is called *Maclaurin's Expansion* for the function f .

Theorem 1.7. (Taylor's Theorem for Function of Several Variables). If $f(x, y)$ and all its partial derivatives of order n are finite and continuous for all points (x, y) in the domain $a \leq x \leq a + h$, $b \leq y \leq b + k$, then

$$f(a + h, b + k) = f(a, b) + d f(a, b) + \frac{1}{2!}d^2 f(a, b) + \cdots + \frac{1}{(n-1)!}d^{n-1} f(a, b) + R_n$$

where

$$d = h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y}$$

and

$$R_n = \frac{1}{n!}d^n f(a + \theta h, b + \theta k), \quad 0 < \theta < 1.$$

Putting $a = b = 0$, $h = x$, $k = y$, we get

$$f(x, y) = f(0, 0) + df(0, 0) + \frac{1}{2!}d^2 f(0, 0) + \cdots + \frac{1}{(n-1)!}d^{n-1} f(0, 0) + R_n$$

where

$$R_n = \frac{1}{n!}d^n f(\theta x, \theta y), \quad 0 < \theta < 1.$$

This result is called *Maclaurin's Theorem* for functions of several variables.

Theorem 1.8. (*Fundamental Theorem of Integral Calculus*). If f is continuous over $[a, b]$, then there exists a function F , called the anti-derivative of f , such that

$$\int_a^b f(x)dx = F(b) - F(a),$$

where $F'(x) = f(x)$.

The second version of the above theorem is given below.

Theorem 1.9. If f is continuous over $[a, b]$ and $a < x < b$, then

$$\frac{d}{dx} \int_a^x f(t)dt = f(x) \quad \text{or} \quad F'(x) = f(x)$$

where

$$F(x) = \int_a^x f(t)dt.$$

1.3 TYPES OF ERRORS

In numerical computation, the quantity “True value – Approximate value” is called the *error*.

We come across the following types of errors in numerical computation:

1. *Inherent Error (initial error)*. Inherent error is the quantity which is already present in the statement (data) of the problem before its solution. This type of error arises due to the use of approximate value in the given data because there are limitations of the mathematical tables and calculators. This type of error can also be there due to mistakes by human. For example, one can write, by mistake, 67 instead of 76. The error in this case is called *transposing* error.
2. *Round-off Error*. This error arises due to rounding off the numbers during computation and occurs due to the limitation of computing aids. However, this type of error can be minimized by
 - (i) Avoiding the subtraction of nearly equal numbers or division by a small number.
 - (ii) Retaining at least one more significant figure at each step of calculation.
3. *Truncation Error*. It is the error caused by using approximate formulas during computation such as the one that arise when a function $f(x)$ is evaluated from an infinite series for x after truncating it at certain stage.

For example, we will see that in Newton–Raphson’s Method for finding the roots of an equation, if x is the true value of the root of $f(x) = 0$ and x_0 and h are approximate value and correction, respectively, then by Taylor’s Theorem,

$$f(x_0 + h) = f(x_0) + h' f'(x_0) + \frac{h^2}{2!} f''(x_0) + \dots = 0$$

To find the correction h , we truncate the series just after first derivative. Therefore, some error occurs due to this truncation.

4. *Absolute Error*. If x is the true value of a quantity and x_0 is the approximate value, then $|x - x_0|$ is called the *absolute error*.
5. *Relative Error*. If x is the true value of a quantity and x_0 is the approximate value, then $\left(\frac{x - x_0}{x}\right)$ is called the *relative error*.
6. *Percentage Error*. If x is the true value of quantity and x_0 is the approximate value, then $\left(\frac{x - x_0}{x}\right) \times 100$ is called the *percentage error*. Thus, percentage error is 100 times the relative error.

1.4 GENERAL FORMULA FOR ERRORS

Let

$$u = f(u_1, u_2, \dots, u_n) \quad (1.1)$$

be a function of u_1, u_2, \dots, u_n , which are subject to the errors $\Delta u_1, \Delta u_2, \dots, \Delta u_n$, respectively. Let Δu be the error in u caused by the errors $\Delta u_1, \Delta u_2, \dots, \Delta u_n$ in u_1, u_2, \dots, u_n , respectively. Then

$$u + \Delta u = f(u_1 + \Delta u_1, u_2 + \Delta u_2, \dots, u_n + \Delta u_n). \quad (1.2)$$

Expanding the right-hand side of equation (1.2) by Taylor's Theorem for a function of several variables, we have

$$u + \Delta u = f(u_1, u_2, \dots, u_n) + \left(\Delta u_1 \frac{\partial}{\partial u_1} + \dots + \Delta u_n \frac{\partial}{\partial u_n} \right) f + \frac{1}{2} \left(\Delta u_1 \frac{\partial}{\partial u_1} + \dots + \Delta u_n \frac{\partial}{\partial u_n} \right)^2 f + \dots$$

Since the errors are relatively small, we neglect the squares, product, and higher powers and have

$$u + \Delta u = f(u_1, u_2, \dots, u_n) + \left(\Delta u_1 \frac{\partial}{\partial u_1} + \dots + \Delta u_n \frac{\partial}{\partial u_n} \right) f \quad (1.3)$$

Subtracting equation (1.1) from equation (1.3), we have

$$\Delta u = \frac{\partial f}{\partial u_1} \Delta u_1 + \frac{\partial f}{\partial u_2} \Delta u_2 + \dots + \frac{\partial f}{\partial u_n} \Delta u_n$$

or

$$\Delta u = \frac{\partial u}{\partial u_1} \Delta u_1 + \frac{\partial u}{\partial u_2} \Delta u_2 + \dots + \frac{\partial u}{\partial u_n} \Delta u_n,$$

which is known as *general formula for error*. We note that the right-hand side is simply the total derivative of the function u .

For a relative error E_r of the function u , we have

$$E_r = \frac{\Delta u}{u} = \frac{\partial u}{\partial u_1} \frac{\Delta u_1}{u} + \frac{\partial u}{\partial u_2} \frac{\Delta u_2}{u} + \dots + \frac{\partial u}{\partial u_n} \frac{\Delta u_n}{u}.$$

EXAMPLE 1.2

If $u = 5xy^2/z^3$ and errors in x, y, z are 0.001, compute the relative maximum error $(E_r)_{\max}$ in u when $x = y = z = 1$.

Solution. We have $u = 5xy^2/z^3$. Therefore

$$\frac{\partial u}{\partial x} = \frac{5y^2}{z^3}, \frac{\partial u}{\partial y} = \frac{10xy}{z^3}, \frac{\partial u}{\partial z} = -\frac{15xy^2}{z^4}$$

and so

$$\Delta u = \frac{5y^2}{z^3} \Delta x + \frac{10xy}{z^3} \Delta y - \frac{15xy^2}{z^4} \Delta z.$$

But it is given that $\Delta x = \Delta y = \Delta z = 0.001$ and $x = y = z = 1$. Therefore,

$$(\Delta u)_{\max} \approx \left| \frac{5y^2}{z^3} \Delta x \right| + \left| \frac{10xy}{z^3} \Delta y \right| + \left| \frac{15xy^2}{z^4} \Delta z \right| = 5(0.001) + 10(0.001) + 15(0.001) = 0.03.$$

Thus, the relative maximum error $(E_r)_{\max}$ is given by

$$(E_r)_{\max} = \frac{(\Delta u)_{\max}}{u} = \frac{0.03}{u} = \frac{0.03}{5} = 0.006.$$

EXAMPLE 1.3

Given that

$$\begin{aligned}a &= 10.00 \pm 0.05 \\b &= 0.0356 \pm 0.0002 \\c &= 15300 \pm 100 \\d &= 62000 \pm 500.\end{aligned}$$

Find the maximum value of the absolute error in (i) $a + b + c + d$, and (ii) c^3 .

Solution. We are given that

$$\begin{aligned}a &= 10.00 \pm 0.05 \\b &= 0.0356 \pm 0.0002 \\c &= 15300 \pm 100 \\d &= 62000 \pm 500.\end{aligned}$$

If a_1, b_1, c_1 , and d_1 are true values of a, b, c , and d , respectively, then

$$\begin{aligned}|(a_1 + b_1 + c_1 + d_1) - (a + b + c + d)| &= |(a_1 - a) + (b_1 - b) + (c_1 - c) + (d_1 - d)| \\&\leq |a_1 - a| + |b_1 - b| + |c_1 - c| + |d_1 - d|, \\&= |0.05| + |0.0002| + |100| + |500| \\&= 600.0502,\end{aligned}$$

which is the required maximum value of the absolute error in $a + b + c + d$.

Further, if ϵ is the error in c , then

$$\begin{aligned}|(c + \epsilon)^3 - c^3| &= |\epsilon^3 + 3c\epsilon^2 + 3c^2\epsilon| \\&\leq |(100)^3| + |3(15300)(100)^2| + |3(15300)^2(100)| \\&= 10^6 + 459(10^4) + 3(153)^2(10^6) \\&= 10^6 + 459(10^4) + 70227(10^6) \\&= 10^{10}(0.0001 + 0.000459 + 7.0227) \\&= 10^{10}(7.023259),\end{aligned}$$

which is the required maximum absolute error.

EXAMPLE 1.4

Find the number of terms of the exponential series such that their sum gives the value of e^x correct to five decimal places for all values of x in the range $0 \leq x \leq 1$.

Solution. The remainder term in the expansion of e^x is

$$R_n(x) = \frac{x^n}{n!} e^\xi, \quad 0 < \xi < x.$$

Therefore, maximum absolute error is

$$e_{\max} = \left| \frac{x^n}{n!} \right| = \frac{1}{n!} \quad \text{at } x = 1.$$

Maximum relative error is

$$(e_r)_{\max} = \frac{x^n e^x}{n!} = \frac{x^n}{n!} = \frac{1}{n!} \quad \text{at } x=1.$$

For five-decimal accuracy at $x=1$, we have

$$\frac{1}{n!} < \frac{1}{2} 10^{-5},$$

which yields $n=9$. Therefore, the number of terms in the exponential series should be 9.

1.5 ORDER OF APPROXIMATION

A function $\phi(h)$ is said to approximate $f(h)$ with order of approximation $O(h^n)$ if

$$|f(h) - \phi(h)| \leq \mu |h|^n$$

or if

$$f(h) = \phi(h) + O(h^n).$$

For example, if

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \dots,$$

then we write

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + O(x^4)$$

to the fourth order of approximation.

Similarly,

$$\sin t = t - \frac{t^3}{3!} + \frac{t^5}{5!} - \frac{t^7}{7!} + \dots$$

can be written as

$$\sin t = t - \frac{t^3}{3!} + \frac{t^5}{5!} + O(t^7).$$

The number x_1 is said to approximate x to d significant digits if d is the largest positive integer such that

$$\frac{|x - x_1|}{|x|} < \frac{10^{-d}}{2}.$$

EXAMPLE 1.3

Consider the Taylor's expansions

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + O(x^7),$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + O(x^6).$$

Determine the order of approximation for their sum and product.

Solution. Since $O(x^6) + O(x^7) = O(x^6)$, we have

$$\begin{aligned}
\sin x + \cos x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} + O(x^7) + 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + O(x^6) \\
&= 1 + x - \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + O(x^6) + O(x^7) \\
&= 1 + x - \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + O(x^6).
\end{aligned}$$

Hence the order of approximation for the sum of the given expressions is $O(x^6)$. Further,

$$\begin{aligned}
\sin x \cos x &= \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} + O(x^7) \right) \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} + O(x^6) \right) \\
&= \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} \right) \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} \right) \\
&\quad + \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} \right) O(x^6) + \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} \right) O(x^7) + O(x^6)O(x^7) \\
&= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^3}{2!} + \frac{x^5}{2!3!} - \frac{x^7}{2!5!} \\
&\quad + \frac{x^5}{4!} - \frac{x^7}{3!4!} + \frac{x^9}{4!5!} + O(x^6) + O(x^7) + O(x^6)O(x^7).
\end{aligned}$$

Since $O(x^6) + O(x^7) = O(x^6)$ and $O(x^6)O(x^7) = O(x^{13})$, we have

$$\begin{aligned}
\sin x \cos x &= x - x^3 \left(\frac{1}{3!} + \frac{1}{2!} \right) + x^5 \left(\frac{1}{5!} + \frac{1}{2!3!} + \frac{1}{4!} \right) \\
&\quad - x^7 \left(\frac{1}{2!5!} - \frac{1}{3!4!} \right) + \frac{x^9}{4!5!} + O(x^6) + O(x^{13}) \\
&= x - \frac{2}{3}x^3 + \frac{2}{15}x^5 + O(x^6) + O(x^9) + O(x^{13}) \\
&= x - \frac{2}{3}x^3 + \frac{2}{15}x^5 + O(x^6).
\end{aligned}$$

Hence the order of approximation for the product of the given expressions is $O(x^6)$.

EXAMPLE 1.4

Find the order of the approximation for the sum and product of the following expansion:

$$\begin{aligned}
e^h &= 1 + h + \frac{h^2}{2!} + \frac{h^3}{3!} + O(h^4), \\
\cos h &= 1 - \frac{h^2}{2!} + \frac{h^4}{4!} + O(h^6).
\end{aligned}$$

Solution. Since $O(h^4) + O(h^6) = O(h^4)$, we have

$$\begin{aligned}
e^h + \cos h &= 1 + h + \frac{h^2}{2!} + \frac{h^3}{3!} + O(h^4) + 1 - \frac{h^2}{2!} + \frac{h^4}{4!} + O(h^6) \\
&= 2 + h + \frac{h^3}{3!} + \frac{h^4}{4!} + O(h^4) + O(h^6) \\
&= 2 + h + \frac{h^3}{3!} + O(h^4) + O(h^6) \\
&= 2 + h + \frac{h^3}{3!} + O(h^4).
\end{aligned}$$

Hence the order of approximation for the sum is $O(h^4)$.

On the other hand,

$$\begin{aligned}
e^h + \cos h &= \left(1 + h + \frac{h^2}{2!} + \frac{h^3}{3!} + O(h^4)\right) \left(1 - \frac{h^2}{2!} + \frac{h^4}{4!} + O(h^6)\right) \\
&= \left(1 + h + \frac{h^2}{2!} + \frac{h^3}{3!}\right) \left(1 - \frac{h^2}{2!} + \frac{h^4}{4!}\right) + \left(1 + h + \frac{h^2}{2!} + \frac{h^3}{3!}\right) O(h^6) \\
&\quad + \left(1 - \frac{h^2}{2!} + \frac{h^4}{4!}\right) O(h^4) + O(h^4) O(h^6) \\
&= 1 + h - \frac{h^3}{3} - \frac{5h^4}{24} - \frac{h^5}{24} + \frac{h^6}{48} + \frac{h^7}{144} + O(h^6) + O(h^4) + O(h^4) O(h^6) \\
&= 1 + h - h^3 + O(h^4) + O(h^6) + O(h^4) + O(h^4) O(h^6) \\
&= 1 + h - \frac{h^3}{3} + O(h^4) + O(h^{10}) \\
&= 1 + h - \frac{h^3}{3} + O(h^4).
\end{aligned}$$

Hence the order of approximation for the product is $O(h^4)$.

EXERCISES

- Round off the following number to three decimal places:

i) 498.5561	(ii) 52.2756
iii) 0.70035	(iv) 48.21416.

Ans. (i) 498.556 (ii) 52.276 (iii) 0.700 (iv) 48.214.
 - Round off to four significant figures

(i) 19.235101	(ii) 49.85561
(iii) 0.0022218	

Ans. (i) 19.24 (ii) 49.8600 (iii) 0.002222.
 - Find the number of term of the exponential series such that their sum gives the value of e^x correct to eight decimal places at $x = 1$

$$\text{Hint. } e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^{n-1}}{(n-1)!} + \frac{x^n}{n!} e^\xi, \quad 0 < \xi < x.$$

Thus,

Maximum absolute error at $\xi = x$ is equal to $x^n/n!$ and so

$$\text{Maximum relative error} = \left(\frac{x^n e^x}{n!} \right) / e^x = \frac{x^n}{n!} = \frac{1}{n!},$$

since $x = 1$. For eight-decimal accuracy at $x = 1$, we have

$$\frac{1}{n!} < \frac{1}{2} 10^{-8},$$

which yields $n = 12$.

4. If $n = 10x^3y^2z^2$ and error in x, y, z are, respectively, 0.03, 0.01, 0.02 at $x = 3, y = 1, z = 2$. Calculate the absolute error and percent relative error in the calculation of it.

Ans. 140.4, 13%.

5. What is the order of approximation of

$$\cos t = 1 - \frac{t^2}{2!} + \frac{t^4}{4!} + O(t^6)?$$

6. Find the order of approximation for the sum and product of the expansions

$$\frac{1}{1-h} = 1 + h + h^2 + h^3 + O(h^4),$$

$$\cos h = 1 + \frac{h^2}{2!} + \frac{h^4}{4!} + O(h^6).$$

Ans. $O(h^4)$.

2 Non-Linear Equations

The aim of this chapter is to discuss the most useful methods for finding the roots of any equation having numerical coefficients. Polynomial equations of degree ≤ 4 can be solved by standard algebraic methods. But no general method exists for finding the roots of the equations of the type $a \log x + bx = c$ or $ae^{-x} + b \tan x = 4$, etc. in terms of their coefficients. These equations are called transcendental equations. Therefore, we take help of numerical methods to solve such type of equations.

Let f be a continuous function. Any number ξ for which $f(\xi) = 0$ is called a root of the equation $f(x) = 0$. Also, ξ is called a zero of function $f(x)$.

A zero ξ is called of multiplicity p , if we can write

$$f(x) = (x - \xi)^p g(x),$$

where $g(x)$ is bounded at ξ and $g(\xi) \neq 0$. If $p = 1$, then ξ is said to be simple zero and if $p > 1$, then ξ is called a multiple zero.

2.1 CLASSIFICATION OF METHODS

The methods for finding roots numerically may be classified into the following two types:

1. **Direct Methods.** These methods require no knowledge of an initial approximation and are used for solving polynomial equations. The best known method is Graeffe's root squaring method.
2. **Iterative Methods.** There are many such methods. We shall discuss some of them in this chapter. In these methods, successive approximations to the solution are used. We begin with the first approximation and successively improve it till we get result to our satisfaction. For example, Newton-Raphson method is an iterative method.

Let $\{x_i\}$ be a sequence of approximate values of the root of an equation obtained by an iteration method and let x denote the exact root of the equation. Then the iteration method is said to be convergent if and only if

$$\lim_{n \rightarrow \infty} |x_n - x| = 0.$$

An iteration method is said to be of order p , if p is the smallest number for which there exists a finite constant k such that

$$|x_{n+1} - x| \leq k |x_n - x|^p.$$

2.2 APPROXIMATE VALUES OF THE ROOTS

Let

$$f(x) = 0 \quad (2.1)$$

be the equation whose roots are to be determined. If we take a set of rectangular co-ordinate axes and plot the graph of

$$y = f(x), \quad (2.2)$$

then the values of x where the graph crosses the x -axis are the roots of the given equation (2.1), because at these points y is zero and therefore equation (2.1) is satisfied.

However, the following fundamental theorem is more useful than a graph.

Theorem 2.1. If f is continuous on $[a,b]$ and if $f(a)$ and $f(b)$ are of opposite signs, then there is at least one real root of $f(x) = 0$ between a and b .

In many cases, the approximate values of the real roots of $f(x) = 0$ are found by writing the equation in the form

$$f_1(x) = f_2(x) \quad (2.3)$$

and then plotting the graphs, on the same axes, of two equations $y_1 = f_1(x)$ and $y_2 = f_2(x)$. The abscissas of the point of intersection of these two curves are the real roots of the given equation because at these points $y_1 = y_2$ and therefore $f_1(x) = f_2(x)$. Hence, equation (2.3) is satisfied and consequently $f(x) = 0$ is satisfied.

For example, consider the equation $x \log_{10} x = 1.2$. We write the equation in the form

$$f(x) = x \log_{10} x - 1.2 = 0.$$

It is obvious from the table given below that $f(2)$ and $f(3)$ are of opposite signs:

x	:	1	2	3	4
$f(x)$:	-1.2	-0.6	0.2	3 1.2 1

Therefore, a root lies between $x = 2$ and $x = 3$ and this is the only root.

The approximate value of the root can also be found by writing the equation in the form

$$\log_{10} x = \frac{1.2}{x}$$

and then plotting the graphs of $y_1 = \log_{10} x$ and $y_2 = 1.2/x$. The abscissa of the point of intersection of these graphs is the desired root.

2.3 BISECTION METHOD (BOLZANO METHOD)

Suppose that we want to find a zero of a continuous function f . We start with an initial interval $[a_0, b_0]$, where $f(a_0)$ and $f(b_0)$ have opposite signs. Since f is continuous, the graph of f will cross the x -axis at a root $x = \xi$ lying in $[a_0, b_0]$. Thus, the graph shall be as shown in Figure 2.1.

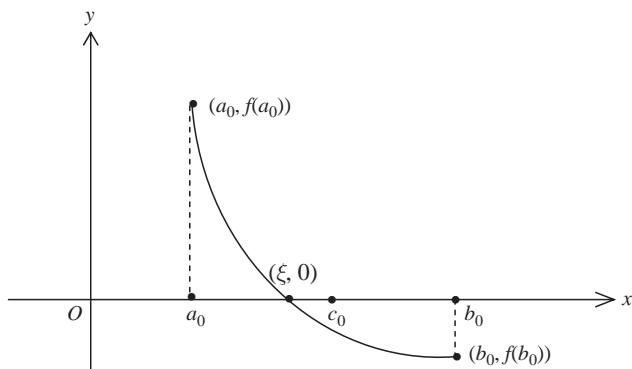


Figure 2.1

The bisection method systematically moves the endpoints of the interval closer and closer together until we obtain an interval of arbitrary small width that contains the root. We choose the midpoint $c_0 = (a_0 + b_0)/2$ and then consider the following possibilities:

- (i) If $f(a_0)$ and $f(c_0)$ have opposite signs, then a root lies in $[a_0, c_0]$.
- (ii) If $f(c_0)$ and $f(b_0)$ have opposite signs, then a root lies in $[c_0, b_0]$.
- (iii) If $f(c_0) = 0$, then $x = c_0$ is a root.

If (iii) happens, then nothing to proceed as c_0 is the root in that case. If anyone of (i) and (ii) happens, let $[a_1, b_1]$ be the interval (representing $[a_0, c_0]$ or $[c_0, b_0]$) containing the root, where $f(a_1)$ and $f(b_1)$ have opposite signs. Let $c_1 = (a_1 + b_1)/2$ and $[a_2, b_2]$ represent $[a_1, c_1]$ or $[c_1, b_1]$ such that $f(a_2)$ and $f(b_2)$ have opposite signs. Then the root lies between a_2 and b_2 . Continue with the process to construct an interval $[a_{n+1}, b_{n+1}]$, which contains the root and its width is half that of $[a_n, b_n]$. In this case $[a_{n+1}, b_{n+1}] = [a_n, c_n]$ or $[c_n, b_n]$ for all n .

Theorem 2.2. Let f be a continuous function on $[a, b]$ and let $\xi \in [a, b]$ be a root of $f(x) = 0$. If $f(a)$ and $f(b)$ have opposite signs and $\{c_n\}$ represents the sequence of the midpoints generated by the bisection process, then

$$|\xi - c_n| \leq \frac{b-a}{2^{n+1}}, \quad n = 0, 1, 2, \dots$$

and hence $\{c_n\}$ converges to the root $x = \xi$, that is, $\lim_{n \rightarrow \infty} c_n = \xi$.

Proof. Since both the root ξ and the midpoint c_n lie in $[a_n, b_n]$, the distance from c_n to ξ cannot be greater than half the width of $[a_n, b_n]$ as shown in Figure 2.2.

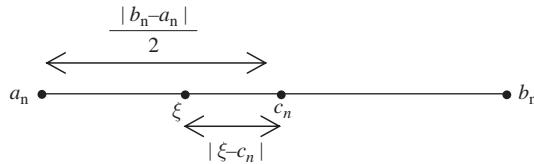


Figure 2.2

Thus,

$$|\xi - c_n| \leq \frac{|b_n - a_n|}{2} \quad \text{for all } n.$$

But, we note that

$$\begin{aligned} |b_1 - a_1| &= \frac{|b_0 - a_0|}{2}, \\ |b_2 - a_2| &= \frac{|b_1 - a_1|}{2} = \frac{|b_0 - a_0|}{2^2}, \\ |b_3 - a_3| &= \frac{|b_2 - a_2|}{2} = \frac{|b_0 - a_0|}{2^3} \\ &\dots \\ |b_n - a_n| &= \frac{|b_{n-1} - a_{n-1}|}{2} = \frac{|b_0 - a_0|}{2^n}. \end{aligned}$$

Hence,

$$|\xi - c_n| \leq \frac{|b_0 - a_0|}{2^{n+1}} \quad \text{for all } n$$

and so $\lim_{n \rightarrow \infty} |\xi - c_n| = 0$ or $\lim_{n \rightarrow \infty} c_n = \xi$.

EXAMPLE 2.1

Find a real root of the equation $x^3 + x^2 - 1 = 0$ using bisection method.

Solution. Let

$$f(x) = x^3 + x^2 - 1.$$

Then $f(0) = -1$, $f(1) = 1$. Thus, a real root of $f(x) = 0$ lies between 0 and 1. Therefore, we take $x_0 = 0.5$. Then $f(0.5) = (0.5)^3 + (0.5)^2 - 1 = 0.125 + 0.25 - 1 = -0.625$.

This shows that the root lies between 0.5 and 1, and we get

$$x_1 = \frac{1+0.5}{2} = 0.75.$$

Then $f(x_1) = (0.75)^3 + (0.75)^2 - 1 = 0.421875 + 0.5625 - 1 = -0.015625$.

Hence, the root lies between 0.75 and 1. Thus, we take

$$x_2 = \frac{1+0.75}{2} = 0.875$$

and then

$$f(x_2) = 0.66992 + 0.5625 - 1 = 0.23242 \text{ (+ve).}$$

It follows that the root lies between 0.75 and 0.875. We take

$$x_3 = \frac{0.75 + 0.875}{2} = 0.8125$$

and then

$$f(x_3) = 0.53638 + 0.66015 - 1 = 0.19653 \text{ (+ve).}$$

Therefore, the root lies between 0.75 and 0.8125. So, let

$$x_4 = \frac{0.75 + 0.8125}{2} = 0.781,$$

which yields

$$f(x_4) = (0.781)^3 + (0.781)^2 - 1 = 0.086 \text{ (+ve).}$$

Thus, the root lies between 0.75 and 0.781. We take

$$x_5 = \frac{0.750 + 0.781}{2} = 0.765$$

and note that

$$f(0.765) = 0.0335 \text{ (+ve).}$$

Hence, the root lies between 0.75 and 0.765. So, let

$$x_6 = \frac{0.750 + 0.765}{2} = 0.7575$$

and then

$$f(0.7575) = 0.4346 + 0.5738 - 1 = 0.0084 \text{ (+ve).}$$

Therefore, the root lies between 0.75 and 0.7575.

Proceeding in this way, the next approximations shall be

$$\begin{aligned}x_7 &= 0.7538, & x_8 &= 0.7556, & x_9 &= 0.7547, \\x_{10} &= 0.7551, & x_{11} &= 0.7549, & x_{12} &= 0.75486,\end{aligned}$$

and so on.

EXAMPLE 2.2

Find a root of the equation $x^3 - 3x - 5 = 0$ by bisection method.

Solution. Let $f(x) = x^3 - 3x - 5$. Then we observe that $f(2) = -3$ and $f(3) = 13$. Thus, a root of the given equation lies between 2 and 3. Let $x_0 = 2.5$. Then

$$f(2.5) = (2.5)^3 - 3(2.5) - 5 = 3.125 \text{ (+ve).}$$

Thus, the root lies between 2.0 and 2.5. Then

$$x_1 = \frac{2+2.5}{2} = 2.25.$$

We note that $f(2.25) = -0.359375$ (-ve). Therefore, the root lies between 2.25 and 2.5. Then we take

$$x_2 = \frac{2.25+2.5}{2} = 2.375$$

and observe that $f(2.375) = 1.2715$ (+ve). Hence, the root lies between 2.25 and 2.375. Therefore, we take

$$x_3 = \frac{2.25+2.375}{2} = 2.3125.$$

Now $f(2.3125) = 0.4289$ (+ve). Hence, a root lies between 2.25 and 2.3125. We take

$$x_4 = \frac{2.25+2.3125}{2} = 2.28125.$$

Now

$$f(2.28125) = 0.0281 \text{ (+ve).}$$

We observe that the root lies very near to 2.28125. Let us try 2.280. Then

$$f(2.280) = 0.0124.$$

Thus, the root is 2.280 approximately.

2.4 REGULA-FALSI METHOD

The Regula-Falsi method, also known as method of false position, chord method or secant method, is the oldest method for finding the real roots of a numerical equation. We know that the root of the equation $f(x) = 0$ corresponds to abscissa of the point of intersection of the curve $y = f(x)$ with the x -axis. In Regula-Falsi method, we replace the curve by a chord in the interval, which contains a root of the equation $f(x) = 0$. We take the point of intersection of the chord with the x -axis as an approximation to the root.

Suppose that a root $x = \xi$ lies in the interval (x_{n-1}, x_n) and that the corresponding ordinates $f(x_{n-1})$ and $f(x_n)$ have opposite signs. The equation of the straight line through the points $P(x_n, f(x_n))$ and $Q(x_{n-1}, f(x_{n-1}))$ is

$$\frac{f(x) - f(x_n)}{f(x_{n-1}) - f(x_n)} = \frac{x - x_n}{x_{n-1} - x_n}. \quad (2.4)$$

Let this straight line cut the x -axis at x_{n+1} . Since $f(x) = 0$ where the line (2.4) cuts the x -axis, we have, $f(x_{n+1}) = 0$ and so

$$x_{n+1} = x_n - \frac{x_{n-1} - x_n}{f(x_{n-1}) - f(x_n)} f(x_n). \quad (2.5)$$

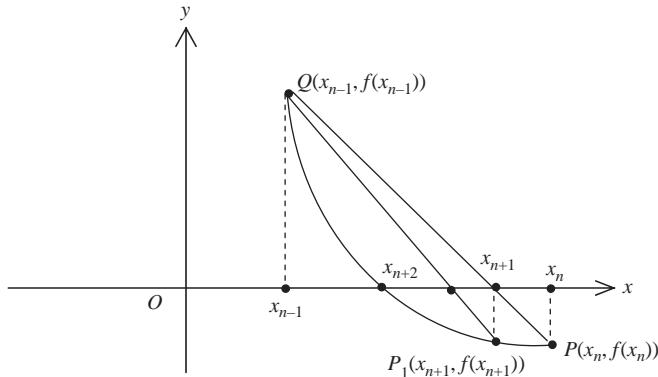


Figure 2.3

Now $f(x_{n-1})$ and $f(x_{n+1})$ have opposite signs. Therefore, it is possible to apply the approximation again to determine a line through the points Q and P_1 . Proceeding in this way we find that as the points approach ξ , the curve becomes more nearly a straight line. Equation (2.5) can also be written in the form

$$x_{n+1} = \frac{x_n f(x_{n-1}) - x_{n-1} f(x_n)}{f(x_{n-1}) - f(x_n)}, \quad n = 1, 2, \dots \quad (2.6)$$

Equation (2.5) or (2.6) is the required formula for Regula-Falsi method.

2.5 CONVERGENCE OF REGULA-FALSI METHOD

Let ξ be the actual root of the equation $f(x) = 0$. Thus, $f(\xi) = 0$. Let $x_n = \xi + \varepsilon_n$, where ε_n is the error involved at the n th step while determining the root. Using

$$x_{n+1} = \frac{x_n f(x_{n-1}) - x_{n-1} f(x_n)}{f(x_{n-1}) - f(x_n)}, \quad n = 1, 2, \dots,$$

we get

$$\xi + \varepsilon_{n+1} = \frac{(\xi + \varepsilon_n) f(\xi + \varepsilon_{n-1}) - (\xi + \varepsilon_{n-1}) f(\xi + \varepsilon_n)}{f(\xi + \varepsilon_{n-1}) - f(\xi + \varepsilon_n)}$$

and so

$$\begin{aligned}\varepsilon_{n+1} &= \frac{(\xi + \varepsilon_n)f(\xi + \varepsilon_{n-1}) - (\xi + \varepsilon_{n-1})f(\xi + \varepsilon_n)}{f(\xi + \varepsilon_{n-1}) - f(\xi + \varepsilon_n)} - \xi \\ &= \frac{\varepsilon_n f(\xi + \varepsilon_{n-1}) - \varepsilon_{n-1} f(\xi + \varepsilon_n)}{f(\xi + \varepsilon_{n-1}) - f(\xi + \varepsilon_n)}.\end{aligned}$$

Expanding the right-hand side by Taylor's series, we get

$$\begin{aligned}\varepsilon_{n+1} &= \frac{\varepsilon_n \left[f(\xi) + \varepsilon_{n-1} f'(\xi) + \frac{1}{2} \varepsilon_{n-1}^2 f''(\xi) + \dots \right] - \varepsilon_{n-1} \left[f(\xi) + \varepsilon_n f'(\xi) + \frac{1}{2} \varepsilon_n^2 f''(\xi) + \dots \right]}{f(\xi) + \varepsilon_{n-1} f'(\xi) + \frac{1}{2} \varepsilon_{n-1}^2 f''(\xi) + \dots - f(\xi) - \varepsilon_n f'(\xi) - \frac{1}{2} \varepsilon_n^2 f''(\xi) - \dots}\end{aligned}$$

that is,

$$\varepsilon_{n+1} = k \varepsilon_{n-1} \varepsilon_n + O(\varepsilon_n^2), \quad (2.7)$$

where

$$k = \frac{1}{2} \frac{f''(\xi)}{f'(\xi)}.$$

We now try to determine some number in m such that

$$\varepsilon_{n+1} = A \varepsilon_n^m \quad (2.8)$$

and

$$\varepsilon_n = A \varepsilon_{n-1}^m \quad \text{or} \quad \varepsilon_{n-1} = A^{-\frac{1}{m}} \varepsilon_n^{\frac{1}{m}}.$$

From equations (2.7) and (2.8), we get

$$\varepsilon_{n+1} = k \varepsilon_{n-1} \varepsilon_n = k A^{-\frac{1}{m}} \varepsilon_n^{\frac{1}{m}} \varepsilon_n = k A^{-\frac{1}{m}} \varepsilon_n^{1+\frac{1}{m}}.$$

and so

$$A \varepsilon_n^m = k A^{-\frac{1}{m}} \varepsilon_n^{\frac{1}{m}} \varepsilon_n = k A^{-\frac{1}{m}} \varepsilon_n^{1+\frac{1}{m}}.$$

Equating powers of ε_n on both sides, we get

$$m = \frac{m+1}{m} \quad \text{or} \quad m^2 - m - 1 = 0,$$

which yields $m = \frac{1 \pm \sqrt{5}}{2} = 1.618$ (+ve value). Hence,

$$\varepsilon_{n+1} = A \varepsilon_n^{1.618}.$$

Thus, Regula–Falsi method is of order 1.618.

EXAMPLE 2.3

Find a real root of the equation $x^3 - 5x - 7 = 0$ using Regula–Falsi method.

Solution. Let $f(x) = x^3 - 5x - 7 = 0$. We note that $f(2) = -9$ and $f(3) = 5$. Therefore, one root of the given equation lies between 2 and 3. By Regula–Falsi method, we have

$$x_{n+1} = \frac{x_n f(x_{n-1}) - x_{n-1} f(x_n)}{f(x_{n-1}) - f(x_n)}, \quad n = 1, 2, 3, \dots$$

We start with $x_0 = 2$ and $x_1 = 3$. Then

$$x_2 = \frac{x_1 f(x_0) - x_0 f(x_1)}{f(x_0) - f(x_1)} = \frac{3(-9) - 2(5)}{-9 - 5} = \frac{37}{14} \approx 2.6.$$

But $f(2.6) = -2.424$ and $f(3) = 5$. Therefore,

$$x_3 = \frac{x_2 f(x_1) - x_1 f(x_2)}{f(x_1) - f(x_2)} = \frac{(2.6) 5 + 3 (-2.424)}{5 + 2.424} = 2.73.$$

Now $f(2.73) = -0.30583$. Since we are getting close to the root, we calculate $f(2.75)$ which is found to be 0.046875. Thus, the next approximation is

$$\begin{aligned} x_4 &= \frac{2.75 f(2.73) - (2.73) f(2.75)}{f(2.73) - f(2.75)} \\ &= \frac{2.75(-0.30583) - 2.73(0.046875)}{-0.30583 - 0.046875} = 2.7473. \end{aligned}$$

Now $f(2.747) = -0.0062$. Therefore,

$$\begin{aligned} x_5 &= \frac{2.75 f(2.747) - 2.747 f(2.75)}{f(2.747) - f(2.75)} \\ &= \frac{2.75(-0.0062) - 2.747(0.046875)}{-0.0062 - 0.046875} = 2.74724. \end{aligned}$$

Thus, the root is 2.747 correct up to three places of decimal.

EXAMPLE 2.4

Solve $x \log_{10} x = 1.2$ by Regula–Falsi method.

Solution. We have $f(x) = x \log_{10} x - 1.2 = 0$. Then $f(2) = -0.60$ and $f(3) = 0.23$. Therefore, the root lies between 2 and 3. Then

$$x_2 = \frac{x_1 f(x_0) - x_0 f(x_1)}{f(x_0) - f(x_1)} = \frac{3(-0.6) - 2(0.23)}{-0.6 - 0.23} = 2.723.$$

Now $f(2.72) = 2.72 \log(2.72) - 1.2 = -0.01797$. Since we are getting closer to the root, we calculate $f(2.75)$ and have

$$f(2.75) = 2.75 \log(2.75) - 1.2 = 2.75(0.4393) - 1.2 = 0.00816.$$

Therefore,

$$x_3 = \frac{2.75(-0.01797) - 2.72(0.00816)}{-0.01797 - 0.00816} = \frac{-0.04942 - 0.02219}{-0.02613} = 2.7405.$$

Now $f(2.74) = 2.74 \log(2.74) - 1.2 = 2.74(0.43775) - 1.2 = -0.00056$.

Thus, the root lies between 2.74 and 2.75 and it is more close to 2.74. Therefore,

$$x_4 = \frac{2.75(-0.00056) - 2.74(0.00816)}{-0.00056 - 0.00816} = 2.7408.$$

Thus the root is 2.740 correct up to three decimal places.

EXAMPLE 2.5

Find by Regula–Falsi method the real root of the equation $\log x - \cos x = 0$ correct to four decimal places.

Solution. Let

$$f(x) = \log x - \cos x.$$

Then

$$f(1) = 0 - 0.54 = -0.54 \text{ (ve)}$$

$$f(1.5) = 0.176 - 0.071 = 0.105 \text{ (+ve).}$$

Therefore, one root lies between 1 and 1.5 and it is nearer to 1.5.

We start with $x_0 = 1$, $x_1 = 1.5$. Then, by Regula–Falsi method,

$$x_{n+1} = \frac{x_n f(x_{n-1}) - x_{n-1} f(x_n)}{f(x_{n-1}) - f(x_n)}$$

and so

$$x_2 = \frac{x_1 f(x_0) - x_0 f(x_1)}{f(x_0) - f(x_1)} = \frac{1.5(-0.54) - 1(0.105)}{-0.54 - 0.105} = 1.41860 \approx 1.42.$$

But, $f(x_2) = f(1.42) = 0.1523 - 0.1502 = 0.0021$. Therefore,

$$x_3 = \frac{x_2 f(x_1) - x_1 f(x_2)}{f(x_1) - f(x_2)} = \frac{1.42(0.105) - 1.5(0.0021)}{0.105 - 0.0021} = 1.41836 \approx 1.4184.$$

Now $f(1.418) = 0.151676 - 0.152202 = -0.000526$.

Hence, the next iteration is

$$x_4 = \frac{x_3 f(x_2) - x_2 f(x_3)}{f(x_2) - f(x_3)} = \frac{1.418(0.0021) - (1.42)(-0.000526)}{0.0021 + 0.000526} = 1.41840.$$

EXAMPLE 2.6

Find the root of the equation $\cos x - xe^x = 0$ by secant method correct to four decimal places.

Solution. The given equation is

$$f(x) = \cos x - xe^x = 0.$$

We note that $f(0) = 1$, $f(1) = \cos 1 - e = 0 - e = -e$ (ve). Hence, a root of the given equation lies between 0 and 1. By secant method, we have

$$x_{n+1} = x_n - \frac{x_{n-1} - x_n}{f(x_{n-1}) - f(x_n)} f(x_n).$$

So taking initial approximation as

$x_0 = 0$, $x_1 = 1$, $f(x_0) = 1$ and $f(x_1) = -e = -2.1780$, we have

$$x_2 = x_1 - \frac{x_0 - x_1}{f(x_0) - f(x_1)} f(x_1) = 1 - \frac{-1}{1 + 2.178} (-2.178) = 0.3147.$$

Further, $f(x_2) = f(0.3147) = 0.5198$. Therefore,

$$x_3 = x_2 - \frac{x_1 - x_2}{f(x_1) - f(x_2)} f(x_2) = 0.3147 - \frac{1 - 0.3147}{-2.178 - 0.5198} (0.5198) = 0.4467.$$

Further, $f(x_3) = f(0.4467) = 0.2036$. Therefore

$$x_4 = x_3 - \frac{x_2 - x_3}{f(x_2) - f(x_3)} f(x_3) = 0.4467 - \frac{0.3147 - 0.4467}{0.5198 - 0.2036} (0.2036) = 0.5318,$$

$$f(x_4) = f(0.5318) = -0.0432.$$

Therefore,

$$x_5 = x_4 - \frac{x_3 - x_4}{f(x_3) - f(x_4)} f(x_4) = 0.5318 - \frac{0.4467 - 0.5318}{0.2036 + 0.0432} (-0.0432) = 0.5168,$$

and

$$f(x_5) = f(0.5168) = 0.0029.$$

Now

$$x_6 = x_5 - \frac{x_4 - x_5}{f(x_4) - f(x_5)} f(x_5) = 0.5168 - \frac{0.5318 - 0.5168}{-0.0432 - 0.0029} (0.0029) = 0.5177,$$

and

$$f(x_6) = f(0.5177) = 0.0002.$$

The sixth iteration is

$$x_7 = x_6 - \frac{x_5 - x_6}{f(x_5) - f(x_6)} f(x_6) = 0.5177 - \frac{0.5168 - 0.5177}{0.0029 - 0.0002} (0.0002) = 0.51776.$$

We observe that $x_6 = x_7$ up to four decimal places. Hence, $x = 0.5177$ is a root of the given equation correct to four decimal places.

2.6 NEWTON-RAPHSON METHOD

If the derivative of a function f can be easily found and is a simple expression, then the real roots of the equation $f(x) = 0$ can be computed rapidly by Newton-Raphson method.

Let x_0 denote the approximate value of the desired root and let h be the correction which must be applied to x_0 to give the exact value of the root x . Thus, $x = x_0 + h$ and so the equation $f(x) = 0$ reduces to $f(x_0 + h) = 0$. Expanding by Taylor's Theorem, we have

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2!} f''(x_0 + \theta h), \quad 0 < \theta < 1.$$

Hence,

$$f(x_0) + xf'(x_0) + \frac{h^2}{2} f''(x_0 + \theta h) = 0.$$

If h is relatively small, we may neglect the term containing h^2 and have

$$f(x_0) + hf'(x_0) = 0.$$

Hence,

$$h = -\frac{f(x_0)}{f'(x_0)}$$

and so the improved value of the root becomes

$$x_1 = x_0 + h = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

If we use x_1 as the approximate value, then the next approximation to the root is

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

In general, the $(n + 1)$ th approximation is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, 3, \dots \quad (2.9)$$

Formula (2.9) is called Newton–Raphson method.

The expression $h = -\frac{f(x_0)}{f'(x_0)}$ is the fundamental formula in Newton–Raphson method. This formula

tells us that the larger the derivative, the smaller is the correction to be applied to get the correct value of the root. This means, when the graph of f is nearly vertical where it crosses the x -axis, the correct value of the root can be found very rapidly and with very little labor. On the other hand, if the value of $f'(x)$ is small in the neighborhood of the root, the value of h given by the fundamental formula would be large and therefore the computation of the root shall be a slow process. Thus, Newton–Raphson method should not be used when the graph of f is nearly horizontal where it crosses the x -axis. Further, the method fails if $f'(x) = 0$ in the neighborhood of the root.

EXAMPLE 2.7

Find the smallest positive root of $x^3 - 5x + 3 = 0$.

Solution. We observe that there is a root between -2 and -3 , a root between 1 and 2 , and a (smallest) root between 0 and 1 . We have

$$f(x) = x^3 - 5x + 3, \quad f'(x) = 3x^2 - 5.$$

Then taking $x_0 = 1$, we have

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 1 - \frac{f(1)}{f'(1)} = 1 - \frac{(-1)}{-2} = 0.5,$$

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 0.5 + \frac{5}{34} = 0.64,$$

$$x_3 = 0.64 + \frac{0.062144}{3.7712} = 0.6565,$$

$$x_4 = 0.6565 + \frac{0.000446412125}{3.70702325} = 0.656620,$$

$$x_5 = 0.656620 + \frac{0.00000115976975}{3.70655053} = 0.656620431.$$

We observe that the convergence is very rapid even though x_0 was not very near to the root.

EXAMPLE 2.8

Find the positive root of the equation

$$x^4 - 3x^3 + 2x^2 + 2x - 7 = 0$$

by Newton–Raphson method.

Solution. We have $f(0) = -7$, $f(1) = -5$, $f(2) = -3$, $f(3) = 17$. Thus, the positive root lies between 2 and 3. The Newton–Raphson formula becomes

$$x_{n+1} = x_n - \frac{x_n^4 - 3x_n^3 + 2x_n^2 + 2x_n - 7}{4x_n^3 - 9x_n^2 + 4x_n + 2}.$$

Taking $x_0 = 2.1$, the improved approximations are

$$x_1 = 2.39854269,$$

$$x_2 = 2.33168543,$$

$$x_3 = 2.32674082,$$

$$x_4 = 2.32671518,$$

$$x_5 = 2.32671518.$$

Since $x_4 = x_5$, the Newton–Raphson formula gives no new values of x and the approximate root is correct to eight decimal places.

EXAMPLE 2.9

Use Newton–Raphson method to solve the transcendental equation $e^x = 5x$.

Solution. Let $f(x) = e^x - 5x = 0$. Then $f'(x) = e^x - 5$. The Newton–Raphson formula becomes

$$x_{n+1} = x_n - \frac{e^{x_n} - 5x_n}{e^{x_n} - 5}, \quad n = 0, 1, 2, 3, \dots .$$

The successive approximations are

$$x_0 = 0.4, x_1 = 0.2551454079, x_2 = 0.2591682786,$$

$$x_3 = 0.2591711018, x_4 = 0.2591711018.$$

Thus, the value of the root is correct to 10 decimal places.

EXAMPLE 2.10

Find by Newton–Raphson method, the real root of the equation $3x = \cos x + 1$.

Solution. The given equation is

$$f(x) = 3x - \cos x - 1 = 0.$$

We have

$$f(0) = -2 \text{ (ve)} \text{ and } f(1) = 3 - 0.5403 - 1 = 1.4597 \text{ (+ve).}$$

Hence, one of the roots of $f(x) = 0$ lies between 0 and 1. The values at 0 and 1 show that the root is nearer to 1. So let us take $x = 0.6$. Further,

$$f'(x) = 3 + \sin x.$$

Therefore, the Newton–Raphson formula gives

$$\begin{aligned} x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{3x_n - \cos x_n - 1}{3 + \sin x_n} \\ &= \frac{3x_n + x_n \sin x_n - 3x_n + \cos x_n + 1}{3 + \sin x_n} = \frac{x_n \sin x_n + \cos x_n + 1}{3 + \sin x_n}. \end{aligned}$$

Hence,

$$\begin{aligned}x_1 &= \frac{x_0 \sin x_0 + \cos x_0 + 1}{3 + \sin x_0} = \frac{0.6(0.5646) + 0.8253 + 1}{3 + 0.5646} = 0.6071, \\x_2 &= \frac{x_1 \sin x_1 + \cos x_1 + 1}{3 + \sin x_1} = \frac{(0.6071)(0.5705) + 0.8213 + 1}{3 + 0.5705} = 0.6071.\end{aligned}$$

Hence the required root, correct to four decimal places, is 0.6071.

EXAMPLE 2.11

Using Newton–Raphson method, find a root of the equation $f(x) = x \sin x + \cos x = 0$ correct to three decimal places, assuming that the root is near to $x = \pi$.

Solution. We have

$$f(x) = x \sin x + \cos x = 0.$$

Therefore,

$$f'(x) = x \cos x + \sin x - \sin x = x \cos x.$$

Since the root is nearer to π , we take $x_0 = \pi$. By Newton–Raphson method

$$\begin{aligned}x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n \sin x_n + \cos x_n}{x_n \cos x_n} \\&= \frac{x_n^2 \cos x_n - x_n \sin x_n - \cos x_n}{x_n \cos x_n}\end{aligned}$$

Thus,

$$\begin{aligned}x_1 &= \frac{x_0^2 \cos x_0 - x_0 \sin x_0 - \cos x_0}{x_0 \cos x_0} \\&= \frac{\pi^2 \cos \pi - \pi \sin \pi - \cos \pi}{\pi \cos \pi} = \frac{1 - \pi^2}{\pi} = \frac{1 - 9.87755}{-3.142857} = 2.824, \\x_2 &= \frac{x_1^2 \cos x_1 - x_1 \sin x_1 - \cos x_1}{x_1 \cos x_1} \\&= \frac{(7.975)(-0.95) - (2.824)(0.3123) + (0.95)}{(2.824)(-0.95)} \\&= \frac{-7.576 - 0.8819 + 0.95}{-2.6828} = \frac{7.5179}{2.6828} = 2.8022, \\x_3 &= \frac{7.8512(-0.9429) - (2.8022)(0.3329) + 0.9429}{(2.8022)(-0.9429)} \\&= \frac{-7.4029 - 0.93285 + 0.9429}{-2.6422} = \frac{7.39285}{2.6422} = 2.797.\end{aligned}$$

Calculate x_4 and x_5 similarly.

2.7 SQUARE ROOT OF A NUMBER USING NEWTON–RAPHSOON METHOD

Suppose that we want to find the square root of N . Let

$$x = \sqrt{N} \quad \text{or} \quad x^2 = N.$$

We have

$$f(x) = x^2 - N = 0.$$

Then, Newton–Raphson method yields

$$\begin{aligned}x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^2 - N}{2x_n} \\&= \frac{1}{2} \left[x_n + \frac{N}{x_n} \right], \quad n = 0, 1, 2, 3, \dots\end{aligned}$$

For example, if $N = 10$, taking $x_0 = 3$ as an initial approximation, the successive approximations are

$$x_1 = 3.16666667, \quad x_2 = 3.162280702,$$

$$x_3 = 3.162277660, \quad x_4 = 3.162277660$$

correct up to nine decimal places.

However, if we take $f(x) = x^3 - Nx$ so that if $f(x) = 0$, then $x = \sqrt[N]{N}$. Now $f'(x) = 3x^2 - N$ and so the Newton–Raphson method gives

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^3 - Nx_n}{3x_n^2 - N} = \frac{2x_n^3}{3x_n^2 - N}.$$

Taking $x_0 = 3$, the successive approximations to $\sqrt{10}$ are

$$x_1 = 3.176, \quad x_2 = 3.1623, \quad x_3 = 3.16227, \quad x_4 = 3.16227$$

correct up to five decimal places.

Suppose that we want to find the p th root of N . Then consider $f(x) = x^p - N$. The Newton–Raphson formula yields

$$\begin{aligned}x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^p - N}{px_n^{p-1}} \\&= \frac{(p-1)x_n^p + N}{px_n^{p-1}}, \quad n = 0, 1, 2, 3, \dots\end{aligned}$$

For $p = 3$, the formula reduces to

$$x_{n+1} = \frac{2x_n^3 + N}{3x_n^2} = \frac{1}{3} \left(2x_n + \frac{N}{x_n^2} \right).$$

If $N = 10$ and we start with the approximation $x_0 = 2$, then

$$x_1 = \frac{1}{3} \left(4 + \frac{10}{4} \right) = 2.16666, \quad x_2 = 2.154503616,$$

$$x_3 = 2.154434692, \quad x_4 = 2.154434690, \quad x_5 = 2.154434690$$

correct up to eight decimal places.

2.8 ORDER OF CONVERGENCE OF NEWTON–RAPHSOON METHOD

Suppose $f(x) = 0$ has a simple root at $x = \xi$ and let ε_n be the error in the approximation. Then $x_n = \xi + \varepsilon_n$. Applying Taylor's expansion of $f(x_n)$ and $f'(x_n)$ about the root ξ , we have

$$f(x_n) = \sum_{r=1}^{\infty} a_r \varepsilon_n^r \quad \text{and} \quad f'(x_n) = \sum_{r=1}^{\infty} r a_r \varepsilon_n^{r-1},$$

where $a_r = \frac{f^{(r)}(\xi)}{r!}$. Then

$$\frac{f(x_n)}{f'(x_n)} = \varepsilon_n - \frac{a_2}{a_1} \varepsilon_n^2 + O(\varepsilon_n^3).$$

Therefore, Newton–Raphson formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

gives

$$\xi + \varepsilon_{n+1} = \xi + \varepsilon_n - \left[\varepsilon_n - \frac{a_2}{a_1} \varepsilon_n^2 + O(\varepsilon_n^3) \right]$$

and so

$$\varepsilon_{n+1} = \frac{a_2}{a_1} \varepsilon_n^2 = \frac{1}{2} \frac{f''(\xi)}{f'(\xi)} \varepsilon_n^2.$$

If $\frac{1}{2} \frac{f''(\xi)}{f'(\xi)} < 1$, then

$$\varepsilon_{n+1} < \varepsilon_n^2. \quad (2.10)$$

It follows therefore that Newton–Raphson method has a quadratic convergence (or second order convergence)

if $\frac{1}{2} \frac{f''(\xi)}{f'(\xi)} < 1$.

The inequality (2.10) implies that if the correction term $\frac{f(x_n)}{f'(x_n)}$ begins with n zeros, then the result is correct to about 2^n decimals. Thus, in Newton–Raphson method, the number of correct decimal roughly doubles at each stage.

2.9 FIXED POINT ITERATION

Let f be a real-valued function $f : \mathbb{R} \rightarrow \mathbb{R}$. Then a point $x \in \mathbb{R}$ is said to be a fixed point of f if $f(x) = x$.

For example, let $I : \mathbb{R} \rightarrow \mathbb{R}$ be an identity mapping. Then all points of \mathbb{R} are fixed points for I since $I(x) = x$ for all $x \in \mathbb{R}$. Similarly, a constant map of \mathbb{R} into \mathbb{R} has a unique fixed point.

Consider the equation

$$f(x) = 0. \quad (2.11)$$

The fixed point iteration approach to the solution of equation (2.11) is that it is rewritten in the form of an equivalent relation

$$x = \phi(x). \quad (2.12)$$

Then any solution of equation (2.11) is a fixed point of the iteration function ϕ . Thus, the task of solving the equation is reduced to find the fixed points of the iteration function ϕ .

Let x_0 be an initial solution (approximate value of the root of equation (2.11) obtained from the graph of f or otherwise). We substitute this value of x_0 in the right-hand side of equation (2.12) and obtain a better approximation x_1 given by

$$x_1 = \phi(x_0).$$

Then the successive approximations are

$$\begin{aligned} x_2 &= \phi(x_1), \\ x_3 &= \phi(x_2), \\ \dots &\dots \dots \\ \dots &\dots \dots \\ x_{n+1} &= \phi(x_n), \quad n=0,1,2,3,\dots \end{aligned}$$

The iteration

$$x_{n+1} = \phi(x_n), \quad n=0,1,2,3,\dots$$

is called fixed point iteration.

Obviously, Regula–Falsi method and Newton–Raphson method are iteration processes.

2.10 CONVERGENCE OF ITERATION METHOD

We are interested in determining the condition under which the iteration method converges, that is, for which x_{n+1} converges to the solution of $x = \phi(x)$ as $n \rightarrow \infty$. Thus, if $x_{n+1} = x$ up to the number of significant figures considered, then x_n is a solution to that degree of approximation. Let ξ be the true solution of $x = \phi(x)$, that is,

$$\xi = \phi(\xi). \quad (2.13)$$

The first approximation is

$$x_1 = \phi(x_0). \quad (2.14)$$

Subtracting equation (2.14) from equation (2.13), we get

$$\begin{aligned} \xi - x_1 &= \phi(\xi) - \phi(x_0) \\ &= (\xi - x_0)\phi'(\xi_0), \quad x_0 < \xi_0 < \xi, \end{aligned}$$

by Mean Value Theorem. Similar equations hold for successive approximations so that

$$\begin{aligned} \xi - x_2 &= (\xi - x_1)\phi'(\xi_1), \quad x_1 < \xi_1 < \xi \\ \xi - x_3 &= (\xi - x_2)\phi'(\xi_2), \quad x_2 < \xi_2 < \xi \\ \dots &\dots \dots \\ \xi - x_{n+1} &= (\xi - x_n)\phi'(\xi_n), \quad x_n < \xi_n < \xi. \end{aligned}$$

Multiplying together all the equations, we get

$$\xi - x_{n+1} = (\xi - x_0)\phi'(\xi_0)\phi'(\xi_1)\dots\phi'(\xi_n)$$

and so

$$|\xi - x_{n+1}| = |\xi - x_0| |\phi'(\xi_0)| \dots |\phi'(\xi_n)|.$$

If each of $|\phi'(\xi_0)|, \dots, |\phi'(\xi_n)|$ is less than or equal to $k < 1$, then

$$|\xi - x_{n+1}| \leq |\xi - x_0| k^{n+1} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Hence, the error $\xi - x_{n+1}$ can be made as small as we please by repeating the process a sufficient number of times. Thus, the condition for convergence is

$$|\phi'(x)| < 1$$

in the neighborhood of the desired root.

Consider the iteration formula $x_{n+1} = \phi(x_n)$, $n = 0, 1, 2, \dots$. If ξ is the true solution of $x = \phi(x)$, then $\xi = \phi(\xi)$. Therefore,

$$\begin{aligned}\xi - x_{n+1} &= \phi(\xi) - \phi(x_n) = (\xi - x_n)\phi'(\xi) \\ &= (\xi - x_n)k, \quad |\phi'(\xi)| \leq k < 1,\end{aligned}$$

which shows that the iteration method has a linear convergence. This slow rate of convergence can be accelerated in the following way: we write

$$\begin{aligned}\xi - x_{n+1} &= (\xi - x_n)k \\ \xi - x_{n+2} &= (\xi - x_{n+1})k.\end{aligned}$$

Dividing, we get

$$\frac{\xi - x_{n+1}}{\xi - x_{n+2}} = \frac{\xi - x_n}{\xi - x_{n+1}}$$

or

$$(\xi - x_{n+1})^2 = (\xi - x_{n+2})(\xi - x_n)$$

or

$$\xi = x_{n+2} - \frac{(x_{n+2} - x_{n+1})^2}{x_{n+2} - 2x_{n+1} + x_n} = x_{n+2} - \frac{(\Delta x_{n+1})^2}{\Delta^2 x_n}. \quad (2.15)$$

Formula (2.15) is called the Aitken's Δ^2 -method.

2.11 SQUARE ROOT OF A NUMBER USING ITERATION METHOD

Suppose that we want to find square root of a number, say N . This is equivalent to say that we want to find x such that $x^2 = N$, that is, $x = \frac{N}{x}$ or $x + x = x + \frac{N}{x}$. Thus,

$$x = \frac{x + \frac{N}{x}}{2}.$$

Thus, if x_0 is the initial approximation to the square root, then

$$x_{n+1} = \frac{x_n + \frac{N}{x_n}}{2}, \quad n = 0, 1, 2, \dots$$

Suppose $N = 13$. We begin with the initial approximation of $\sqrt{13}$ found by bisection method. The solution lies between 3.5625 and 3.625. We start with $x_0 = \frac{3.5625 + 3.6250}{2} \approx 3.59375$. Then, using the above iteration formula, we have

$$x_1 = 3.6055705, \quad x_2 = 3.6055513, \quad x_3 = 3.6055513$$

correct up to seven decimal places.

2.12 SUFFICIENT CONDITION FOR THE CONVERGENCE OF NEWTON-RAPHSON METHOD

We know that an iteration method $x_{n+1} = \phi(x_n)$ converges if $|\phi'(x)| < 1$. Since Newton-Raphson method is an iteration method, where $\phi(x) = x - \frac{f(x)}{f'(x)}$ and therefore it converges if $|\phi'(x)| < 1$, that is, if

$$\left| 1 - \frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2} \right| < 1,$$

that is, if

$$|f(x)f''(x)| < (f'(x))^2,$$

which is the required sufficient condition for the convergence of Newton-Raphson method.

EXAMPLE 2.12

Derive an iteration formula to solve $f(x) = x^3 + x^2 - 1 = 0$ and solve the equation.

Solution. Since $f(0)$ and $f(1)$ are of opposite signs, there is a root between 0 and 1. We write the equation in the form

$$x^3 + x^2 = 1, \text{ that is, } x^2(x+1) = 1, \text{ or } x^2 = \frac{1}{x+1},$$

or equivalently,

$$x = \frac{1}{\sqrt{1+x}}.$$

Then

$$x = \phi(x) = \frac{1}{\sqrt{1+x}}, \quad \phi'(x) = -\frac{1}{2(1+x)^{\frac{3}{2}}}$$

so that

$$|\phi'(x)| < 1 \quad \text{for } x < 1.$$

Hence, this iteration method is applicable. We start with $x_0 = 0.75$ and obtain the next approximations to the root as

$$\begin{aligned} x_1 &= \phi(x_0) = \frac{1}{\sqrt{1+x_0}} \approx 0.7559, \\ x_2 &= \phi(x_1) \approx 0.7546578, \\ x_3 &\approx 0.7549249, \\ x_4 &\approx 0.7548674, \\ x_5 &\approx 0.754880, \\ x_6 &\approx 0.7548772, \\ x_7 &\approx 0.75487767 \end{aligned}$$

correct up to six decimal places.

EXAMPLE 2.13

Find, by the method of iteration, a root of the equation $2x - \log_{10} x = 7$.

Solution. The fixed point form of the given equation is

$$x = \frac{1}{2}(\log_{10} x + 7).$$

From the intersection of the graphs $y_1 = 2x - 7$ and $y_2 = \log_{10} x$, we find that the approximate value of the root is 3.8. Therefore,

$$x_0 = 3.8, \quad x_1 = \frac{1}{2}(\log 3.8 + 7) \approx 3.78989,$$

$$x_2 = \frac{1}{2}(\log 3.78989 + 7) \approx 3.789313,$$

$$x_3 = \frac{1}{2}(\log 3.789313 + 7) \approx 3.78928026,$$

$$x_4 \approx 3.789278, \quad x_5 \approx 3.789278$$

correct up to six decimal places.

EXAMPLE 2.14

Use iteration method to solve the equation $e^x = 5x$.

Solution. The iteration formula for the given problem is

$$x_{n+1} = \frac{1}{5}e^{x_n}.$$

We start with $x_0 = 0.3$ and get the successive approximations as

$$x_1 = \frac{1}{5}(1.34985881) = 0.269972, \quad x_2 = 0.26198555,$$

$$x_3 = 0.25990155, \quad x_4 = 0.259360482,$$

$$x_5 = 0.259220188, \quad x_6 = 0.259183824,$$

$$x_7 = 0.259174399, \quad x_8 = 0.259171956,$$

$$x_9 = 0.259171323, \quad x_{10} = 0.259171159,$$

correct up to six decimal places.

If we use Aitken's Δ^2 -method, then

$$x_3 = x_2 - \frac{(\Delta x_1)^2}{\Delta^2 x_0} = x_2 - \frac{(x_2 - x_1)^2}{x_2 - 2x_1 + x_0} = 0.26198555 - \frac{0.000063783}{0.02204155} = 0.259091$$

and so on.

2.13 NEWTON'S METHOD FOR FINDING MULTIPLE ROOTS

If ξ is a multiple root of an equation $f(x) = 0$, then $f(\xi) = f'(\xi) = 0$ and therefore the Newton–Raphson method fails. However, in case of multiple roots, we proceed as follows:

Let ξ be a root of multiplicity m . Then

$$f(x) = (x - \xi)^m A(x) \quad (2.16)$$

We make use of a localized approach that in the immediate vicinity (neighborhood) of $x = \xi$, the relation (2.16) can be written as

$$f(x) = A(x - \xi)^m,$$

where $A = A(\xi)$ is effectively constant. Then

$$f'(x) = mA(x - \xi)^{m-1}$$

$$f''(x) = m(m-1)A(m - \xi)^{m-2}, \text{ and so on.}$$

We thus obtain

$$\frac{f'(x)}{f(x)} = \frac{m}{x - \xi}$$

or

$$\xi = x - \frac{mf(x)}{f'(x)},$$

where x is close to ξ , which is a modification of Newton's rule for a multiple root. Thus, if x_1 is in the neighborhood of a root ξ of multiplicity m of an equation $f(x) = 0$, then

$$x_2 = x_1 - m \frac{f(x_1)}{f'(x_1)}$$

is an even more close approximation to ξ . Hence, in general, we have

$$x_{n+1} = x_n - m \frac{f(x_n)}{f'(x_n)}. \quad (2.17)$$

Remark 2.1. (i) The case $m = 1$ of equation (2.17) yields Newton–Raphson method.

(ii) If two roots are close to a number, say x , then

$$f(x + \varepsilon) = 0 \text{ and } f(x - \varepsilon) = 0,$$

that is,

$$f(x) + \varepsilon f'(x) + \frac{\varepsilon^2}{2!} f''(x) + \dots = 0, \quad f(x) - \varepsilon f'(x) + \frac{\varepsilon^2}{2!} f''(x) - \dots = 0.$$

Since ε is small, adding the above expressions, we get

$$0 = 2f(x) + \varepsilon^2 f''(x) = 0$$

or

$$\varepsilon^2 = -2 \frac{f(x)}{f''(x)}$$

or

$$\varepsilon = \pm \sqrt{\frac{-2f(x)}{f''(x)}}.$$

So in this case, we take two approximations as $x + \varepsilon$ and $x - \varepsilon$ and then apply Newton–Raphson method.

EXAMPLE 2.15

The equation $x^4 - 5x^3 - 12x^2 + 76x - 79 = 0$ has two roots close to $x = 2$. Find these roots to four decimal places.

Solution. We have

$$f(x) = x^4 - 5x^3 - 12x^2 + 76x - 79$$

$$f'(x) = 4x^3 - 15x^2 - 24x + 76$$

$$f''(x) = 12x^2 - 30x - 24.$$

Thus

$$f(2) = 16 - 40 - 48 + 152 - 79 = 1$$

$$f''(2) = 48 - 60 - 24 = -36.$$

Therefore,

$$\varepsilon = \pm \sqrt{\frac{-2f(2)}{f''(2)}} = \pm \sqrt{\frac{-2}{-36}} = \pm 0.2357.$$

Thus, the initial approximations to the roots are

$$x_0 = 2.2357 \quad \text{and} \quad y_0 = 1.7643.$$

The application of Newton-Raphson method yields

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 2.2357 + 0.00083 = 2.0365.$$

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 2.2365 + 0.000459 = 2.24109.$$

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} = 2.24109 - 0.00019 = 2.2410.$$

Thus, one root, correct to four decimal places is 2.2410. Similarly, the second root correct to four decimal places will be found to be 1.7684.

EXAMPLE 2.16

Find a double root of the equation

$$x^3 - 5x^2 + 8x - 4 = 0$$

near 1.8.

Solution. We have

$$f(x) = x^3 - 5x^2 + 8x - 4$$

$$f'(x) = 3x^2 - 10x + 8$$

and $x_0 = 1.8$. Therefore,

$$f(x_0) = f(1.8) = 5.832 - 16.2 + 14.4 - 4 = 0.032$$

$$f'(x_0) = 9.72 - 18 + 8 = -0.28.$$

Hence,

$$x_1 = x_0 - 2 \frac{f(x_0)}{f'(x_0)} = 1.8 - 2 \frac{f(1.8)}{f'(1.8)} = 1.8 - 2 \frac{0.032}{-0.28} = 2.02857.$$

We take $x_1 = 2.028$. Then

$$f(x_1) = 8.3407 - 20.5639 + 16.224 - 4 = 0.0008$$

$$f'(x_1) = 12.3384 - 20.28 + 8 = 0.0584.$$

Therefore,

$$\begin{aligned} x_2 &= x_1 - 2 \frac{f(x_1)}{f'(x_1)} \\ &= 2.028 - \frac{2(0.0008)}{0.0584} = 2.0006, \end{aligned}$$

which is quite close to the actual double root 2.

EXAMPLE 2.17

Find the double root of $x^3 - x^2 - x + 1 = 0$ close to 0.8.

Solution. We have

$$\begin{aligned}f(x) &= x^3 - x^2 - x + 1 = 0 \\f'(x) &= 3x^2 - 2x - 1.\end{aligned}$$

We choose $x_0 = 0.8$. Then

$$x_{n+1} = x_n - m \frac{f(x_n)}{f'(x_n)}$$

and so

$$\begin{aligned}x_1 &= x_0 - 2 \frac{f(0.8)}{f'(0.8)} = 0.8 - 2 \left(\frac{(0.8)^3 - (0.8)^2 - 0.8 + 1}{3(0.8)^2 - 2(0.8) - 1} \right) = 1.01176 \\x_2 &= x_1 - \frac{2f(1.01176)}{f'(1.01176)} = 1.0118 - 0.0126 = 0.9992,\end{aligned}$$

which is very close to the actual double root 1.

2.14 NEWTON-RAPHSON METHOD FOR SIMULTANEOUS EQUATIONS

We consider the case of two equations in two unknowns. So let the given equations be

$$\phi(x, y) = 0, \quad (2.18)$$

$$\psi(x, y) = 0 \quad (2.19)$$

Now if x_0, y_0 be the approximate values of a pair of roots and h, k be the corrections, we have

$$x = x_0 + h \quad \text{and} \quad y = y_0 + k.$$

Then equations (2.18) and (2.19) become

$$\phi(x_0 + h, y_0 + k) = 0 \quad (2.20)$$

$$\psi(x_0 + h, y_0 + k) = 0. \quad (2.21)$$

Expanding equations (2.20) and (2.21) by Taylor's Theorem for a function of two variables, we have

$$\begin{aligned}\phi(x_0 + h, y_0 + k) &= \phi(x_0, y_0) + h \left(\frac{\partial \phi}{\partial x} \right)_{x=x_0} + k \left(\frac{\partial \phi}{\partial y} \right)_{y=y_0} + \dots = 0, \\\psi(x_0 + h, y_0 + k) &= \psi(x_0, y_0) + h \left(\frac{\partial \psi}{\partial x} \right)_{x=x_0} + k \left(\frac{\partial \psi}{\partial y} \right)_{y=y_0} + \dots = 0.\end{aligned}$$

Since h and k are relatively small, their squares, products, and higher powers can be neglected. Hence,

$$\phi(x_0, y_0) + h \left(\frac{\partial \phi}{\partial x} \right)_{x=x_0} + k \left(\frac{\partial \phi}{\partial y} \right)_{y=y_0} = 0 \quad (2.22)$$

$$\psi(x_0, y_0) + h \left(\frac{\partial \psi}{\partial x} \right)_{x=x_0} + k \left(\frac{\partial \psi}{\partial y} \right)_{y=y_0} = 0. \quad (2.23)$$

Solving the equations (2.22) and (2.23) by Cramer's rule, we get

$$h = \frac{\begin{vmatrix} -\phi(x_0, y_0) & \left(\frac{\partial \phi}{\partial y}\right)_{y=y_0} \\ -\psi(x_0, y_0) & \left(\frac{\partial \psi}{\partial y}\right)_{y=y_0} \end{vmatrix}}{D},$$

$$k = \frac{\begin{vmatrix} \left(\frac{\partial \phi}{\partial x}\right)_{x=x_0} & -\phi(x_0, y_0) \\ \left(\frac{\partial \psi}{\partial x}\right)_{x=x_0} & -\psi(x_0, y_0) \end{vmatrix}}{D},$$

where

$$D = \begin{vmatrix} \left(\frac{\partial \phi}{\partial x}\right)_{x=x_0} & \left(\frac{\partial \phi}{\partial y}\right)_{y=y_0} \\ \left(\frac{\partial \psi}{\partial x}\right)_{x=x_0} & \left(\frac{\partial \psi}{\partial y}\right)_{y=y_0} \end{vmatrix}$$

Thus,

$$x_1 = x_0 + h, \quad y_1 = y_0 + k.$$

Additional corrections can be obtained by repeated application of these formulae with the improved values of x and y substituted at each step.

Proceeding as in Section 2.10, we can prove that the iteration process for solving simultaneous equations $\phi(x, y) = 0$ and $\psi(x, y) = 0$ converges if

$$\left| \frac{\partial \phi}{\partial x} \right| + \left| \frac{\partial \psi}{\partial x} \right| < 1 \quad \text{and} \quad \left| \frac{\partial \phi}{\partial y} \right| + \left| \frac{\partial \psi}{\partial y} \right| < 1.$$

Remark 2.2. The Newton–Raphson method for simultaneous equations can be used to find complex roots. In fact the equation $f(z) = 0$ is $u(x, y) + iv(x, y) = 0$. So writing the equation as

$$u(x, y) = 0$$

$$v(x, y) = 0,$$

we can find x and y , thereby yielding the complex root.

EXAMPLE 2.18

Solve by Newton–Raphson method

$$\begin{aligned} x + 3 \log_{10} x - y^2 &= 0, \\ 2x^2 - xy + 5x + 1 &= 0. \end{aligned}$$

Solution. On plotting the graphs of these equations on the same set of axes, we find that they intersect at the points $(1.4, -1.5)$ and $(3.4, 2.2)$. We shall compute the second set of values correct to four decimal places. Let

$$\phi(x, y) = x + 3 \log_{10} x - y^2,$$

$$\psi(x, y) = 2x^2 - xy - 5x + 1.$$

Then

$$\frac{\partial \phi}{\partial x} = 1 + \frac{3M}{x}, \quad M = 0.43429$$

$$= 1 + \frac{1.30287}{x}$$

$$\frac{\partial \phi}{\partial y} = -2y$$

$$\frac{\partial \psi}{\partial x} = 4x - y - 5$$

$$\frac{\partial \psi}{\partial y} = -x.$$

Now $x_0 = 3.4, y_0 = 2.2$. Therefore,

$$\phi(x_0, y_0) = 0.1545, \quad \psi(x_0, y_0) = 0.72,$$

$$\left(\frac{\partial \phi}{\partial x} \right)_{x=x_0} = 1.383, \quad \left(\frac{\partial \phi}{\partial y} \right)_{y=y_0} = 4.4,$$

$$\left(\frac{\partial \psi}{\partial x} \right)_{x=x_0} = 6.4, \quad \left(\frac{\partial \psi}{\partial y} \right)_{y=y_0} = -3.1.$$

Putting these values in

$$\phi(x_0, y_0) + h_1 \left(\frac{\partial \phi}{\partial x} \right)_{x=x_0} + k_1 \left(\frac{\partial \phi}{\partial y} \right)_{y=y_0} = 0,$$

$$\psi(x_0, y_0) + h_1 \left(\frac{\partial \psi}{\partial x} \right)_{x=x_0} + k_1 \left(\frac{\partial \psi}{\partial y} \right)_{y=y_0} = 0,$$

we get

$$0.1545 + h_1(1.383) + k_1(4.4) = 0$$

$$-0.72 + h_1(6.4) + k_1(-3.1) = 0.$$

Solving these for h_1 and k_1 , we get

$$h_1 = 0.157 \quad \text{and} \quad k_1 = 0.085.$$

Thus,

$$x_1 = 3.4 + 0.517 = 3.557,$$

$$y_1 = 2.2 + 0.085 = 2.285.$$

Now

$$\phi(x_1, y_1) = 0.011, \quad \psi(x_1, y_1) = 0.3945,$$

$$\left(\frac{\partial \phi}{\partial x} \right)_{x=x_1} = 1.367, \quad \left(\frac{\partial \phi}{\partial y} \right)_{y=y_1} = -4.57,$$

$$\left(\frac{\partial \psi}{\partial x} \right)_{x=x_1} = 6.943, \quad \left(\frac{\partial \psi}{\partial y} \right)_{y=y_1} = -3.557.$$

Putting these values in

$$\begin{aligned}\phi(x_1, y_1) + h_2 \left(\frac{\partial \phi}{\partial x} \right)_{x=x_1} + k_2 \left(\frac{\partial \phi}{\partial y} \right)_{y=y_1} &= 0, \\ \psi(x_1, y_1) + h_2 \left(\frac{\partial \psi}{\partial x} \right)_{x=x_1} + k_2 \left(\frac{\partial \psi}{\partial y} \right)_{y=y_1} &= 0\end{aligned}$$

and solving the equations so obtained, we get

$$h_2 = -0.0685, k_2 = -0.0229.$$

Hence,

$$x_2 = x_1 + h_2 = 3.4885 \quad \text{and} \quad y_2 = y_1 + k_2 = 2.2621.$$

Repeating the process, we get

$$h_3 = -0.0013, k_3 = -0.000561.$$

Hence, the third approximations are

$$x_3 = 3.4872 \quad \text{and} \quad y_3 = 2.26154.$$

Finding the next approximation, we observe that the above approximation is correct to four decimal places.

EXAMPLE 2.19

Find the roots of $1 + z^2 = 0$, taking initial approximation as $(x_0, y_0) = \left(\frac{1}{2}, \frac{1}{2}\right)$.

Solution. We have

$$f(z) = 1 + (x + iy)^2 = 1 + x^2 - y^2 + 2ixy = u + iv,$$

where

$$u(x, y) = 1 + x^2 - y^2,$$

$$v(x, y) = 2xy.$$

Then

$$\frac{\partial u}{\partial x} = 2x, \quad \frac{\partial u}{\partial y} = -2y,$$

$$\frac{\partial v}{\partial x} = 2y, \quad \frac{\partial v}{\partial y} = -2x.$$

Taking initial approximation as $(x_0, y_0) = \left(\frac{1}{2}, \frac{1}{2}\right)$, we have

$$u(x_0, y_0) = u\left(\frac{1}{2}, \frac{1}{2}\right) = 1 + \frac{1}{4} - \frac{1}{4} = 1,$$

$$v(x_0, y_0) = v\left(\frac{1}{2}, \frac{1}{2}\right) = 2\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{2},$$

$$u_x(x_0, y_0) = u_x\left(\frac{1}{2}, \frac{1}{2}\right) = 2\left(\frac{1}{2}\right) = 1,$$

$$u_y(x_0, y_0) = u_y\left(\frac{1}{2}, \frac{1}{2}\right) = -2\left(\frac{1}{2}\right) = -1,$$

$$v_x(x_0, y_0) = v_x\left(\frac{1}{2}, \frac{1}{2}\right) = 2\left(\frac{1}{2}\right) = 1,$$

$$v_y(x_0, y_0) = v_y\left(\frac{1}{2}, \frac{1}{2}\right) = 2\left(\frac{1}{2}\right) = 1.$$

Putting these values in

$$u(x_0, y_0) + h_1 u_x(x_0, y_0) + k_1 u_y(x_0, y_0) = 0$$

and

$$v(x_0, y_0) + h_1 v_x(x_0, y_0) + k_1 v_y(x_0, y_0) = 0,$$

we get

$$1 + h_1 - k_1 = 0 \text{ and } \frac{1}{2} + h_1 - k_1 = 0.$$

Solving these equations for h_1 and k_1 , we get $h_1 = -\frac{3}{4}$, $k_1 = \frac{1}{4}$. Hence,

$$x_1 = x_0 + h_1 = \frac{1}{2} - \frac{3}{4} = -\frac{1}{4},$$

$$y_1 = y_0 + k_1 = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}.$$

Now

$$u(x_1, y_1) = 1 + \frac{1}{16} - \frac{9}{16} = \frac{1}{2},$$

$$v(x_1, y_1) = 2 \left(-\frac{1}{4} \right) \left(\frac{3}{4} \right) = -\frac{3}{8},$$

$$u_x(x_1, y_1) = 2 \left(-\frac{1}{4} \right) = -\frac{1}{2},$$

$$u_y(x_1, y_1) = -2 \left(\frac{3}{4} \right) = -\frac{3}{2},$$

$$v_x(x_1, y_1) = 2 \left(\frac{3}{4} \right) = \frac{3}{2},$$

$$v_y(x_1, y_1) = 2 \left(-\frac{1}{4} \right) = -\frac{1}{2}.$$

Putting these values in

$$u(x_1, y_1) + h_2 u_x(x_1, y_1) + k_2 u_y(x_1, y_1) = 0$$

and

$$v(x_1, y_1) + h_2 v_x(x_1, y_1) + k_2 v_y(x_1, y_1) = 0,$$

we get

$$\frac{1}{2} - \frac{1}{2} h_2 - \frac{3}{2} k_2 = 0 \text{ and } -\frac{3}{8} + \frac{3}{2} h_2 - \frac{1}{2} k_2 = 0.$$

Solving these equations, we get $h_2 = \frac{13}{40}$, $k_2 = \frac{9}{40}$. Hence,

$$x_2 = x_1 + h_2 = -\frac{1}{4} + \frac{13}{40} = \frac{3}{40} = 0.075,$$

$$y_2 = y_1 + k_2 = \frac{3}{4} + \frac{9}{40} = \frac{39}{40} = 0.975.$$

Proceeding in the same fashion, we get

$$x_3 = -0.00172 \text{ and } y_3 = 0.9973.$$

2.15 GRAEFFE'S ROOT SQUARING METHOD

Graeffe's root squaring method is applicable to polynomial equations only. The advantage of this method is that it does not require any prior information about the roots and it gives all the roots of the equation.

Let the given equation be

$$f(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n = 0.$$

If x_1, x_2, \dots, x_n are the roots of this equation, then we have

$$f(x) = a_0(x - x_1)(x - x_2)\dots(x - x_n) = 0. \quad (2.24)$$

Multiplying equation (2.24) by the function

$$(-1)^n f(-x) = a_0(x + x_1)(x + x_2)\dots(x + x_n), \quad (2.25)$$

we get

$$(-1)^n f(-x) f(x) = a_0^2 (x^2 - x_1^2)(x^2 - x_2^2)\dots(x^2 - x_n^2) = 0. \quad (2.26)$$

Putting $x^2 = y$, equation (2.26) becomes

$$\phi(y) = a_0^2 (y - x_1^2)(y - x_2^2)\dots(y - x_n^2) = 0. \quad (2.27)$$

The roots of equation (2.27) are $x_1^2, x_2^2, \dots, x_n^2$ and are thus the squares of the roots of the given equation. Since the relations between the roots x_1, x_2, \dots, x_n and coefficients a_0, a_1, \dots, a_n of the n th degree are

$$\frac{a_1}{a_0} = -(x_1 + x_2 + \dots + x_n),$$

$$\frac{a_2}{a_0} = (x_1 x_2 + x_1 x_3 + \dots),$$

$$\frac{a_3}{a_0} = -(x_1 x_2 x_3 + x_1 x_3 x_4 + \dots),$$

...

$$\frac{a_n}{a_0} = (-1)^n x_1 x_2 \dots x_n,$$

it follows that the roots $x_1^m, x_2^m, \dots, x_n^m$ and the coefficients b_0, b_1, \dots, b_n of the final transformed equation (after m squaring)

$$b_0(x^m)^n + b_1(x^m)^{n-1} + \dots + b_{n-1}x^m + b_n = 0$$

are connected by the corresponding relations

$$\frac{b_1}{b_0} = -(x_1^m + x_2^m + \dots + x_n^m) = -x_1^m \left(1 + \frac{x_2^m}{x_1^m} + \dots + \frac{x_n^m}{x_1^m} \right),$$

$$\frac{b_2}{b_0} = x_1^m x_2^m + x_1^m x_3^m + \dots = x_1^m x_2^m \left(1 + \frac{x_3^m}{x_2^m} + \frac{x_4^m}{x_2^m} + \dots \right),$$

$$\frac{b_3}{b_0} = -(x_1^m x_2^m x_3^m + x_1^m x_2^m x_4^m + \dots) = -x_1^m x_2^m x_3^m \left(1 + \frac{x_4^m}{x_3^m} + \dots \right),$$

...

$$\frac{b_n}{b_0} = (-1)^n x_1^m x_2^m \dots x_n^m.$$

Let the order of the magnitude of the roots be

$$|x_1| > |x_2| > |x_3| > \dots > |x_n|.$$

When the roots are sufficiently separated, the ratios $\frac{x_2^m}{x_1^m}, \frac{x_3^m}{x_2^m}$ etc., are negligible in comparison with unity. Hence, the relations between the roots and the coefficients in the final transformed equation after m squaring are

$$\frac{b_1}{b_0} = -x_1^m, \quad \frac{b_2}{b_0} = x_1^m x_2^m, \quad \frac{b_3}{b_0} = -x_1^m x_2^m x_3^m, \quad \dots, \quad \frac{b_n}{b_0} = (-1)^n x_1^m x_2^m x_3^m \dots x_n^m.$$

Dividing each of these equations after the first by the preceding equation, we obtain

$$\frac{b_2}{b_1} = -x_2^m, \quad \frac{b_3}{b_2} = -x_3^m, \dots, \quad \frac{b_n}{b_{n-1}} = -x_n^m. \quad (2.28)$$

From $\frac{b_1}{b_0} = -x_1^m$ and equations (2.28), we get

$$b_0 x_1^m + b_1 = 0, \quad b_1 x_2^m + b_2 = 0, \quad b_2 x_3^m + b_3 = 0, \quad \dots, \quad b_{n-1} x_n^m + b_n = 0.$$

The roots squaring process has thus broken up the original equation into n simple equations from which the desired roots can be found easily.

Remark 2.3. The multiplication by $(-1)^n f(-x)$ can be carried out as given below:

a_0	a_1	a_2	a_3	a_4	\dots
a_0	$-a_1$	a_2	$-a_3$	a_4	\dots
a_0^2	$-a_1^2$	a_2^2	$-a_3^2$	a_4^2	\dots
$2a_0 a_2$	$-2a_1 a_3$	$2a_2 a_4$	$-2a_3 a_5$	\dots	
$2a_0 a_4$	$-2a_1 a_5$	$2a_2 a_6$	\dots		
$2a_0 a_6$	$-2a_1 a_7$	\dots			
	$2a_0 a_8$	\dots			
b_0	b_1	b_2	b_3	b_4	\dots

EXAMPLE 2.20

Solve the equation $x^3 - 5x^2 - 17x + 20 = 0$ by Graeffe's root squaring method (squaring three times).

Solution. We have

$$f(x) = x^3 - 5x^2 - 17x + 20 = 0$$

and so

$$(-1)^3 f(-x) = x^3 + 5x^2 - 17x - 20.$$

Then, for the first squaring, we have

1	-5	-17	20
1	5	-17	-20
1	-25	289	-400
	-34	200	
1	-59	489	-400

and so the equation obtained after first squaring is

$$y^3 - 59y^2 + 489y - 400 = 0.$$

For the second squaring, we have

$$\begin{array}{cccc} 1 & -59 & 489 & -400 \\ 1 & 59 & -489 & 400 \\ \hline 1 & -3481 & 239121 & -16(10)^4 \\ & 978 & -47200 & \\ \hline 1 & -2503 & 191921 & -16(10)^4 \end{array}$$

Thus, the equation obtained after second squaring is

$$z^3 - 2503z^2 + 191921z - 16(10)^4 = 0$$

For the third squaring, we have

$$\begin{array}{cccc} 1 & -2503 & 191921 & -16(10)^4 \\ 1 & 2503 & -191921 & 16(10)^4 \\ \hline 1 & -6265009 & 36933670241 & -256(10)^6 \\ & 383842 & -800960000 & \\ \hline 1 & -5881167 & 36032710241 & -256(10)^8 \end{array}$$

Thus, the equation obtained after third squaring is

$$u_3 - 5881167u^2 + 36032710241u - 256(10)^8 = 0$$

Hence, the roots are given by

$$\begin{aligned} x_1^8 &= 5881167 \text{ and so } |x_1| = 7.0175, \\ x_2^8 &= \frac{36032710241}{5881167} \text{ and so } |x_2| = 2.9744, \\ x_3^8 &= \frac{256(10)^8}{36032710241} \text{ and so } |x_3| = 0.958170684. \end{aligned}$$

EXAMPLE 2.21

Apply Graeffe's root squaring method to find all the roots of the equation

$$x^3 - 2x^2 - 5x + 6 = 0.$$

Solution. We have

$$f(x) = x^3 - 2x^2 - 5x + 6 = 0 \text{ such that } (-1)f(-x) = x^3 + 2x^2 - 5x - 6.$$

Therefore, using Graeffe's root squaring method, we have three squaring as given below:

$$\begin{array}{cccc} 1 & -2 & -5 & 6 \\ 1 & 2 & -5 & -6 \\ \hline 1 & -4 & 25 & -36 \\ & -10 & 24 & \\ \hline \text{First sq.} & 1 & -14 & 49 & -36 \\ & 1 & 14 & 49 & 36 \\ \hline & 1 & -196 & 2401 & -1296 \\ & & 98 & -1008 & \\ \hline \text{Second sq.} & 1 & -98 & 1393 & -1296 \\ & 1 & 98 & 1393 & 1296 \\ \hline & 1 & -9604 & 1940449 & -1679616 \\ & & & 2796 & -254016 \\ \hline \text{Third sq.} & 1 & -6818 & 1686433 & -1679616 \end{array}$$

Therefore,

$$x_1^8 = \frac{6818}{1} \text{ and so } |x_1| = 3.0144433,$$

$$x_2^8 = \frac{1686433}{6818} \text{ and so } |x_2| = 1.9914253,$$

$$x_3^8 = \frac{1679616}{1686433} \text{ and so } |x_3| = 0.99949382.$$

The values of the roots are in good agreement with the actual values since the actual roots of the given equation are $x = 1, 2, 3$.

EXAMPLE 2.22

Find all the roots of the equation

$$x^3 - 4x^2 + 5x - 2 = 0$$

by Graeffe's root squaring method, squaring thrice.

Solution. The given equation is

$$f(x) = x^3 - 4x^2 + 5x - 2 = 0.$$

Then Graeffe's root squaring method yields

	1	-4	5	-2	
	1	4	5	2	
	<hr/>	1	-16	25	-4
		10	-16		
First sq.	1	-6	9	-4	
	1	6	9	4	
	<hr/>	1	-3	681	-16
		18	-48		
Second sq.	1	-18	33	-16	
	1	18	33	16	
	<hr/>	1	-324	1089	-256
		66	-864		
Third sq.	1	-258	225	-256	

Hence,

$$x_1^8 = \frac{258}{1} \text{ and so } |x_1| = 2.00194,$$

$$x_2^8 = \frac{225}{258} \text{ and so } |x_2| = 0.98304,$$

$$x_3^8 = \frac{256}{225} \text{ and so } |x_3| = 1.03280.$$

We further observe that magnitude of -18 in the second square is half of the square of the magnitude of -6 in the first squaring. Hence, the equation has a double root. Therefore, the double root is given by

$\left(\frac{256}{258}\right)^{\frac{1}{8}} = 0.999028$. Thus, the magnitudes of the roots are 2.00194 and 0.999028 . The actual roots of the equation are $2, 1, 1$.

2.16 MULLER'S METHOD

Muller's method is an iterative method in which we do not require derivative of the function. In this method, the function $f(x)$ is approximated by a second degree curve in the neighborhood of the root. This method is as fast as Newton's method and can be used to find real or complex zeros of a function.

Let x_{i-2} , x_{i-1} , and x_i be the approximations to a root of the equation $f(x) = 0$ and let $y_{i-2} = f(x_{i-2})$, $y_{i-1} = f(x_{i-1})$, and $y_i = f(x_i)$. Let

$$y = A(x - x_i)^2 + B(x - x_i) + y_i \quad (2.29)$$

be the parabola passing through (x_{i-2}, y_{i-2}) , (x_{i-1}, y_{i-1}) , and (x_i, y_i) .

Therefore, we have

$$y_{i-1} = A(x_{i-1} - x_i)^2 + B(x_{i-1} - x_i) + y_i$$

$$y_{i-2} = A(x_{i-2} - x_i)^2 + B(x_{i-2} - x_i) + y_i$$

and so

$$A(x_{i-1} - x_i)^2 + B(x_{i-1} - x_i) = y_{i-1} - y_i \quad (2.30)$$

$$A(x_{i-2} - x_i)^2 + B(x_{i-2} - x_i) = y_{i-2} - y_i. \quad (2.31)$$

Solving equations (2.30) and (2.31) for A and B , we get

$$A = \frac{(x_{i-2} - x_i)(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i-2} - y_i)}{(x_{i-1} - x_{i-2})(x_{i-1} - x_i)(x_{i-2} - x_i)}, \quad (2.32)$$

$$B = \frac{(x_{i-2} - x_i)^2(y_{i-1} - y_i) - (x_{i-1} - x_i)^2(y_{i-2} - y_i)}{(x_{i-2} - x_{i-1})(x_{i-1} - x_i)(x_{i-2} - x_i)}. \quad (2.33)$$

The quadratic equation $A(x_{i-1} - x_i)^2 + B(x_{i-1} - x_i) + y_i = 0$ with A and B given by equations (2.32) and (2.33) yields the next approximation x_{i+1} as

$$\begin{aligned} x_{i+1} - x_i &= \frac{-B \pm \sqrt{B^2 - 4Ay_i}}{2A} \\ &= \frac{-B \pm \sqrt{B^2 - 4Ay_i}}{2A} \left(\frac{B \pm \sqrt{B^2 - 4Ay_i}}{B \pm \sqrt{B^2 - 4Ay_i}} \right) \\ &= -\frac{2y_i}{B \pm \sqrt{B^2 - 4Ay_i}}. \end{aligned} \quad (2.34)$$

The sign in the denominator is chosen so that the denominator becomes largest in magnitude.

EXAMPLE 2.23

Using Muller's method find a root of the equation $x^3 - x - 1 = 0$.

Solution. We are given that

$$f(x) = x^3 - x - 1 = 0.$$

We note that

$$f(1) = -1, \quad f(1.2) = -0.472, \quad f(1.5) = 0.875.$$

Thus, one root of the equation lies between 1.2 and 1.5. Let $x_{i-2} = 1$, $x_{i-1} = 1.2$, and $x_i = 1.5$ be the initial approximations. Then $y_{i-2} = -1$, $y_{i-1} = -0.472$, and $y_i = 0.875$. Therefore,

$$\begin{aligned} A &= \frac{(x_{i-2} - x_i)(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i-2} - y_i)}{(x_{i-1} - x_{i-2})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{(-0.5)(-1.347) - (-0.3)(-1.875)}{(-0.2)(-0.3)(-0.5)} = 3.7. \end{aligned}$$

$$\begin{aligned} B &= \frac{(x_{i-2} - x_i)^2(y_{i-1} - y_i) - (x_{i-1} - x_i)^2(y_{i-2} - y_i)}{(x_{i-2} - x_{i-1})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{(-0.5)^2(-1.347) - (-0.3)^2(-1.875)}{(-0.2)(-0.3)(-0.5)} = 5.6. \end{aligned}$$

Thus the quadratic equation is

$$3.7(x - 1.5)^2 + 5.6(x - 1.5) + 0.875 = 0$$

and the next approximation to the root is

$$\begin{aligned} x_{i+1} &= 1.5 - \frac{2(0.875)}{5.6 + \sqrt{31.36 - 12.95}} \\ &= 1.5 - 0.1769 = 1.3231. \end{aligned}$$

We note that $f(1.3231) = -0.006890$. We repeat the process with $x_{i-2} = 1.2$, $x_{i-1} = 1.3231$, and $x_i = 1.5$ and get $x_{i+2} = 1.3247$.

EXAMPLE 2.24

Use Muller's method to find a root of the equation $f(x) = x^3 - x - 2 = 0$.

Solution. We have

$$f(x) = x^3 - x - 2 = 0.$$

We note that

$$f(1.4) = -0.656, f(1.5) = -0.125, f(1.6) = 0.496.$$

Thus, one root lies between 1.5 and 1.6. Let $x_{i-2} = 1.4$, $x_{i-1} = 1.5$, and $x_i = 1.6$ be the initial approximation. Then

$$y_{i-2} = -0.656, y_{i-1} = -0.125, y_i = 0.496.$$

Therefore,

$$\begin{aligned} A &= \frac{(x_{i-2} - x_i)(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i-2} - y_i)}{(x_{i-1} - x_{i-2})(x_{i-1} - x_i)(x_{i-2} - x_i)} = \frac{(-0.2)(-0.621) - (-0.1)(-1.152)}{(0.1)(-0.1)(-0.2)} = 4.5. \\ B &= \frac{(x_{i-2} - x_i)^2(y_{i-1} - y_i) - (x_{i-1} - x_i)^2(y_{i-2} - y_i)}{(x_{i-2} - x_{i-1})(x_{i-1} - x_{i-2})(x_{i-2} - x_i)} = \frac{(-0.2)^2(-0.621) - (0.1)^2(-1.152)}{(-0.1)(-0.1)(-0.1)} = 6.66. \end{aligned}$$

Therefore, the approximating quadratic equation is

$$4.5(x - 1.6)^2 + 6.66(x - 1.6) + 0.496 = 0$$

and the next approximation to the root is

$$x_{i+1} = 1.6 - \frac{2(0.496)}{6.66 + \sqrt{44.3556 - 8.928}} = 1.52134.$$

We note that

$$f(x_{i+1}) = f(1.52134) = -0.00199.$$

Hence, the approximate value of the root is quite satisfactory.

EXAMPLE 2.25

Apply Muller's method to find a root of the equation $\cos x - xe^x = 0$.

Solution. We are given that

$$f(x) = \cos x - xe^x = 0.$$

We note that

$$f(-1) = 0.540302305 + 0.3678794411 = 0.908181746$$

$$f(0) = 1$$

$$f(1) = -2.177979523.$$

Therefore, one root of the given equation lies between 0 and 1.

Let

$$x_{i-2} = -1, x_{i-1} = 0, x_i = 1$$

be the initial approximation of the root. Then

$$y_{i-2} = 0.908181746, y_{i-1} = 1, y_i = -2.177979523$$

Therefore,

$$\begin{aligned} A &= \frac{(x_{i-2} - x_i)(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i-2} - y_i)}{(x_{i-1} - x_{i-2})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{(-2)(3.177979523) - (-1)(3.086161269)}{(-1)(-1)(-2)} \approx -1.635, \end{aligned}$$

$$\begin{aligned} B &= \frac{(x_{i-2} - x_i)^2(y_{i-1} - y_i) - (x_{i-1} - x_i)^2(y_{i-2} - y_i)}{(x_{i-2} - x_{i-1})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{4(3.177979523) - (1)(3.086161269)}{(-1)(-1)(-2)} \approx -4.813. \end{aligned}$$

Then

$$\begin{aligned} x_{i+1} &= x_i - \frac{2y_i}{B \pm \sqrt{B^2 - 4Ay_i}} \\ &= 1 - \frac{2(-2.177979523)}{(-4.813) \pm \sqrt{23.164969 - 14.24398608}} \\ &= 1 + \frac{4.355959046}{(-7.799801453)} \approx 0.4415. \end{aligned}$$

Now we take the initial approximation as

$$x_{i-2} = 0, x_{i-1} = 0.4415, x_i = 1.$$

Then

$$y_{i-2} = 1, y_{i-1} = 0.217563, y_i = -2.177979523.$$

Therefore,

$$\begin{aligned} A &= \frac{(x_{i-2} - x_i)(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i-2} - y_i)}{(x_{i-1} - x_{i-2})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{(-1)(2.39554253) - (-0.5585)(3.177979523)}{(0.4415)(-0.5585)(-1)} \approx -2.5170, \\ B &= \frac{(x_{i-2} - x_i)^2(y_{i-1} - y_i) - (x_{i-1} - x_i)^2(y_{i-2} - y_i)}{(x_{i-2} - x_{i-1})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{(1)(2.39554253) - (0.31192225)(3.177979523)}{(-0.4415)(-0.5585)(-1)} \\ &= -5.694998867 \approx -5.6950. \end{aligned}$$

Then

$$\begin{aligned} x_{i+1} &= x_i - \frac{2y_i}{B \pm \sqrt{B^2 - 4Ay_i}}. \\ &= 1 - \frac{2(-2.177979523)}{(-5.695) \pm \sqrt{32.4330121 - 21.92789784}} \\ &= 1 + \frac{4.355959046}{-8.936159401} = 1 - 0.487453149 \approx 0.51255. \end{aligned}$$

Repeating this process twice, we get the approximate root 0.517 correct to three decimal places.

EXAMPLE 2.26

Apply Muller's method to find the real root of the equation

$$x^3 - x^2 - x - 1 = 0.$$

Solution. The given equation is

$$f(x) = x^3 - x^2 - x - 1 = 0.$$

We note that

$$f(0) = -1, \quad f(1) = -2, \quad f(2) = 1.$$

Thus, one root lies between 1 and 2. Let

$$\begin{aligned} x_{i-2} &= 0, \quad x_{i-1} = 1 \text{ and } x_i = 2 \\ y_{i-2} &= -1, \quad y_{i-1} = -2 \text{ and } y_i = 1. \end{aligned}$$

Therefore,

$$\begin{aligned} A &= \frac{(x_{i-2} - x_i)(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i-2} - y_i)}{(x_{i-1} - x_{i-2})(x_{i-1} - x_i)(x_{i-2} - x_i)} = \frac{(-2)(-3) - (-1)(-2)}{(1)(-1)(-2)} = \frac{6 - 2}{2} = 2, \\ B &= \frac{(x_{i-2} - x_i)^2(y_{i-1} - y_i) - (x_{i-1} - x_i)^2(y_{i-2} - y_i)}{(x_{i-2} - x_{i-1})(x_{i-1} - x_i)(x_{i-2} - x_i)} = \frac{(-2)^2(-3) - (-1)^2(-2)}{(-1)(-1)(-2)} = \frac{-12 + 2}{-2} = 5. \end{aligned}$$

Therefore,

$$x_{i+1} = 2 - \frac{2(1)}{5 + \sqrt{25 - 4(2)(1)}} = 2 - \frac{2}{5 + 4.123} = 2 - 0.2192 = 1.7808.$$

We note that $f(1.7808) = -0.53625$ (-ve). Thus, the root lies between 1.78 and 2. Therefore, for the second iteration, we set

$$x_{i-2} = 1, \quad x_{i-1} = 1.78 \text{ and } x_i = 2.$$

Then

$$y_{i-2} = -2, \quad y_{i-1} = -0.536, \quad y_i = 1.$$

Therefore,

$$\begin{aligned} A &= \frac{(x_{i-2} - x_i)(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i-2} - y_i)}{(x_{i-1} - x_{i-2})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{(1-2)(-0.536-1) - (1.78-2)(-2-1)}{(1.78-1)(1.78-2)(1-2)} \\ &= \frac{1.536 - 0.66}{(0.78)(-0.22)(-1)} = \frac{0.836}{0.1716} = 4.872, \\ B &= \frac{(-1)^2(-1.536) - (-0.22)^2(-3)}{(1-1.78)(1.78-2)(1-2)} \\ &= \frac{-1.536 + 0.1452}{(-0.78)(-0.22)(-1)} = \frac{1.3908}{0.1716} = 8.10. \end{aligned}$$

Hence,

$$x_{i+1} = 2 - \frac{2(1)}{8.1 + \sqrt{65.61 - 4(4.87)1}} = 2 - \frac{2}{81 + \sqrt{46.13}} = 2 - \frac{2}{8.1 + 6792} = 1.87.$$

We note $f(1.87) = 0.173$. Therefore, $x = 1.87$ is a satisfactory root.

2.17 BAIRSTOW ITERATIVE METHOD

The Bairstow iterative method is used to extract a quadratic factor from a given polynomial. After obtaining quadratic factors, the roots (real or complex) can be found by solving the quadratic equations. Let

$$f(x) = x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n = 0 \quad (2.35)$$

be a polynomial equation of degree n . We wish to find a quadratic factor $x^2 + px + q$ of equation (2.35). Dividing the polynomial $f(x)$ by $x^2 + px + q$, we get

$$f(x) = (x^2 + px + q) (x^{n-2} + b_1 x^{n-3} + \dots + b_{n-3} x + b_{n-2}) + Rx + S. \quad (2.36)$$

Obviously, $(x^2 + px + q)$ will be a factor of $f(x)$ if $R = S = 0$. Thus, our aim is to find p and q such that

$$R(p, q) = 0 \text{ and } S(p, q) = 0 \quad (2.37)$$

The two equations in (2.37) are non-linear equations in p and q . Therefore, Newton-Raphson method for simultaneous equations is applicable. Let p_1 and q_1 be the true values of p and q and Δp and Δq are the corrections, then

$$p_1 = p + \Delta p \text{ and } q_1 = q + \Delta q.$$

Therefore,

$$R(p_1, q_1) = R(p + \Delta p, q + \Delta q) = 0 \text{ and } S(p_1, q_1) = S(p + \Delta p, q + \Delta q) = 0.$$

Applying Taylor's series expansion for two variables, we get

$$R(p_1, q_1) = R(p, q) + \Delta p \frac{\partial R}{\partial p} + \Delta q \frac{\partial R}{\partial q} + \dots = 0$$

and

$$S(p_1, q_1) = S(p, q) + \Delta p \frac{\partial S}{\partial p} + \Delta q \frac{\partial S}{\partial q} + \dots = 0.$$

Neglecting the square and higher powers of Δp and Δq , we have

$$\Delta p \frac{\partial R}{\partial p} + \Delta q \frac{\partial R}{\partial q} = -R(p, q) \quad (2.38)$$

$$\Delta p \frac{\partial S}{\partial p} + \Delta q \frac{\partial S}{\partial q} = -S(p, q). \quad (2.39)$$

Solving equations (2.38) and (2.39) for Δp and Δq , we have

$$\Delta p = -\frac{RS_q - SR_q}{R_p S_q - R_q S_p} \text{ and } \Delta q = -\frac{SR_p - RS_p}{R_p S_q - R_q S_p}, \quad (2.40)$$

where $R_p = \frac{\partial R}{\partial p}$ etc.

We now determine the expression for R and S in terms of p and q . To do so, we equate coefficients of x^n, x^{n-1}, \dots on both sides of equation (2.36) to get

$$\begin{aligned} a_1 &= b_1 + p && \text{so that } b_1 = a_1 - p \\ a_2 &= b_2 + pb_1 + q && \text{so that } b_2 = a_2 - pb_1 - q \\ a_3 &= b_3 + pb_2 + qb_1 && \text{so that } b_3 = a_3 - pb_2 - qb_1 \\ \dots & && \dots \\ a_k &= b_k + pb_{k-1} + qb_{k-2} && \text{so that } b_k = a_k - pb_{k-1} - qb_{k-2} \\ \dots & && \dots \\ a_{n-1} &= R + pb_{n-2} + qb_{n-3} && \text{so that } R = a_{n-1} - pb_{n-2} - qb_{n-3} \\ a_n &= S + qb_{n-2} && \text{so that } S = a_n - qb_{n-2}. \end{aligned} \quad (2.41)$$

Thus, if we introduce the recurrence formula

$$b_k = a_k - pb_{k-1} - qb_{k-2} \text{ for } k = 1, 2, \dots, n \quad (2.42)$$

with $b_0 = 1$ and $b_{-1} = 0$, then the coefficient of the polynomial $x^{n-2} + b_1 x^{n-3} + \dots + b_{n-3} x + b_{n-2}$ of degree $n-2$, called deflated polynomial, can be determined. We observed that

$$R = a_{n-1} - pb_{n-2} - qb_{n-3} = b_{n-1} \quad (2.43)$$

and

$$S = a_n - qb_{n-2} = b_n + pb_{n-1}. \quad (2.44)$$

Therefore, R and S can be determined if b_n are known. To find Δp and Δq in equation (2.40), we require partial derivatives R_p , R_q , S_p , and S_q . From expression (2.42), we have

$$\frac{\partial b_k}{\partial p} = -b_{k+1} - p \frac{\partial b_{k-1}}{\partial p} - q \frac{\partial b_{k-2}}{\partial p} \quad \text{and} \quad \frac{\partial b_0}{\partial p} = \frac{\partial b_{-1}}{\partial p} = 0 \quad (2.45)$$

$$\frac{\partial b_k}{\partial q} = -b_{k-2} - p \frac{\partial b_{k-1}}{\partial q} - q \frac{\partial b_{k-2}}{\partial q} \quad \text{and} \quad \frac{\partial b_0}{\partial q} = \frac{\partial b_{-1}}{\partial q} = 0. \quad (2.46)$$

If $\frac{\partial b_k}{\partial p} = -c_{k-1}$ and $\frac{\partial b_k}{\partial q} = -c_{k-2}$, $k = 1, 2, \dots, n$, then equations (2.45) and (2.46) reduce to, respectively,

$$c_{k-1} = b_{k-1} - pc_{k-2} - qc_{k-3} \quad (2.47)$$

and

$$c_{k-2} = b_{k-2} - pc_{k-3} - qc_{k-4}. \quad (2.48)$$

Therefore, we can introduce a recurrence relation to find c_k in terms of b_k as

$$c_k = b_k - pc_{k-1} - qc_{k-2}, \quad k = 1, 2, \dots, n-1 \text{ with } c_0 = 1 \text{ and } c_{-1} = -0. \quad (2.49)$$

Hence, the relations (2.43) and (2.44) yield

$$\begin{aligned} R_p &= \frac{\partial b_{n-1}}{\partial p} = -c_{n-2}, \quad S_p = \frac{\partial b_n}{\partial p} + p \frac{\partial b_{n-1}}{\partial p} + b_{n-1} = b_{n-1} - c_{n-1} - pc_{n-2}, \\ R_q &= \frac{\partial b_{n-1}}{\partial q} = -c_{n-3}, \quad S_q = \frac{\partial b_n}{\partial q} + p \frac{\partial b_{n-1}}{\partial q} = -(c_{n-2} + pc_{n-3}). \end{aligned}$$

Putting these values of partial derivatives in equation (2.40), we get

$$\left. \begin{aligned} \Delta p &= -\frac{b_n c_{n-3} - b_{n-1} c_{n-2}}{c_{n-2}^2 - c_{n-3}(c_{n-1} - b_{n-1})}, \\ \Delta q &= -\frac{b_{n-1}(c_{n-1} - b_{n-1}) - b_n c_{n-2}}{c_{n-2}^2 - c_{n-3}(c_{n-1} - b_{n-1})} \end{aligned} \right\}. \quad (2.50)$$

Hence, if p_0 and q_0 are initial values of p and q , respectively, then the first improved values are

$$p_1 = p_0 + \Delta p \quad \text{and} \quad q_1 = q_0 + \Delta q$$

and so, in general, the improved values are given by

$$p_{k+1} = p_k + \Delta p \quad \text{and} \quad q_{k+1} = q_k + \Delta q,$$

where Δp and Δq are determined at $p = p_k$ and $q = q_k$. The process is repeated till the result, up to desired accuracy, is obtained.

Remark 2.4. (i). The values of b_k and c_k can be computed by the following scheme:

	1	a_1	a_2	...	a_k	...	a_{n-1}	a_n
	$-p_0$	$-p_0$	$-p_0 b_1$...	$-p_0 b_{k-1}$...	$-p_0 b_{n-2}$	$-p_0 b_{n-1}$
	$-q_0$		$-q_0$...	$-q_0 b_{k-2}$...	$-q_0 b_{n-3}$	$-q_0 b_{n-2}$
	1	b_1	b_2	...	b_k	...	b_{n-1}	b_n
	$-p_0$	$-p_0$	$-p_0 c_1$...	$-p_0 c_{k-1}$...	$-p_0 c_{n-2}$	
	$-q_0$		$-q_0$...	$-q_0 c_{k-2}$...	$-q_0 c_{n-3}$	
	1	c_1	c_2	...	c_k	...	c_{n-1}	

(ii) Since, we have used Newton–Raphson method to solve $R(p, q) = 0$ and $S(p, q) = 0$, Bairstow method is second order process.

EXAMPLE 2.27

Determine quadratic factors using Bairstow method to the equation $x^4 + 5x^3 + 3x^2 - 5x - 9 = 0$.

Solution. We take the initial values of p and q as $p_0 = 3, q_0 = -5$. Then

	1	5	3	-5	-9
	-3	-3	-6	-6	3
	5		5	10	10
	1	$2 = b_1$	$2 = 2b_2$	$-1 = b_3$	$4 = b_4$
	-3	-3	3	-30	
	5		5	-5	
	1	$-1 = c_1$	$10 = c_2$	$-36 = c_3$	

Therefore,

$$\Delta p = -\frac{b_4 c_1 - b_3 c_2}{c_2^2 - c_1(c_3 - b_3)} = -0.09$$

$$\Delta q = -\frac{b_3(c_3 - b_3) - b_4 c_2}{c_2^2 - c_1(c_3 - b_3)} = 0.08.$$

Hence, $p_1 = 3 - 0.09 = 2.91$ and $q_1 = -5 + 0.08 = -4.92$.

Repeating the computation with new values of p and q , we get

	1	5	3	-5	-9
	-2.91	-2.91	-6.08	-5.35	0.20
	4.92		4.92	10.28	9.05
	1	$2.09 = b_1$	1.84	-0.07	$0.25 = b_4$
	-2.91	-2.91	2.37	-26.57	
	4.92		4.92	-4.03	
	1	-0.82	9.13	-30.67	
		↓	↓	↓	
		c_1	c_2	c_3	

Then $\Delta p = -0.00745$, $\Delta q = 0.00241$, and so

$$\begin{aligned} p_2 &= p_1 + \Delta p = 2.902550 \\ q_2 &= q_1 + \Delta q = -4.91759. \end{aligned}$$

The next iterations will yield

$$p_3 = 2.902953, q_3 = -4.917736, p_4 = 2.902954, q_4 = -4.917738.$$

Hence, the approximate factorization is

$$x^4 + 5x^3 + 3x^2 - 5x - 9 = (x^2 + 2.90295x - 4.91773)(x^2 + 2.09704x + 1.83011).$$

EXERCISES

1. Find the root of the equation $x - \cos x = 0$ by bisection method.
Ans. 0.739.
 2. Find a positive root of equation $xe^x = 1$ lying between 0 and 1 using bisection method.
Ans. 0.567.
 3. Solve $x^3 - 4x - 9 = 0$ by Bolzano method.
Ans. 2.706.
 4. Use Regula–Falsi method to solve $x^3 + 2x^2 + 10x - 20 = 0$.
Ans. 1.3688.
 5. Use the method of false position to obtain a root of the equation $x^3 - x + 4 = 0$.
Ans. 1.796.
 6. Solve $e^x \sin x = 1$ by Regula–Falsi method.
Ans. 0.5885.
 7. Using Newton–Raphson method find a root of the equation $x \log_{10} x = 1.2$.
Ans. 2.7406.
 8. Use Newton–Raphson method to obtain a root of $x - \cos x = 0$.
Ans. 0.739.
 9. Solve $\sin x = 1 + x^3$ by Newton–Raphson method.
Ans. -1.24905.
 10. Find the real root of the equation $3x = \cos x + 1$ using Newton–Raphson method.
Ans. 0.6071.
 11. Derive the formula $x_{i+1} = \frac{1}{2} \left(x_i + \frac{N}{x_i} \right)$ to determine square root of N . Hence calculate the square root of 2.
Ans. 1.414214.
 12. Find a real root of the equation $\cos x = 3x - 1$ correct to three decimal places using iteration method.
- Hint:* Iteration formula is $x_n = \frac{1}{3}(1 + \cos x_n)$.
Ans. 0.607.
13. Using iteration method, find a root of the equation $x^3 + x^2 - 100 = 0$.
Ans. 4.3311.
 14. Find the double root of the equation $x^2 - x^2 - x + 1 = 0$ near 0.9.
Ans. 1.0001.

15. Use Newton's method to solve

$$x^2 - y^2 = 4, \quad x^2 + y^2 = 16$$

taking the starting value as (2.828, 2.828).

Ans. $x = 3.162, y = 2.450$.

16. Use Newton's method to solve

$$\begin{aligned} x^2 - 2x - y + 0.5 &= 0, \\ x^2 + 4y^2 - 4 &= 0, \end{aligned}$$

taking the starting value as (2.0, 0.25).

Ans. $x = 1.900677, y = 0.311219$.

17. Find the real roots of the equation $x^3 - 6x^2 + 11x - 6 = 0$ using Graeffe's root squaring method.

Ans. 3, 2, 1.

18. Find the roots of the equation $x^3 - 8x^2 + 17x - 10 = 0$ using Graeffe's root squaring method.

Ans. 5, 2.001, 0.9995.

19. Using Muller's method, find the root of the equation $x^3 - 2x - 5 = 0$ which lies between 2 and 3.

Ans. 2.094568.

20. Find root lying between 2 and 3 of the equation $x^3 - x^2 - x - 1 = 0$ using Muller's method.

Ans. 1.8393.

21. Solve the equation $x^4 + 4x^3 - 7x^2 - 22x + 24 = 0$ using Bairstow method.

Ans. quadratic factor: $(x^2 + 2.00004x - 8.00004)$ and $(x^2 + 2x - 3)$, roots are 1, 2, -3, -4.

22. Solve the equation $x^4 - 8x^3 + 39x^2 - 62x + 50 = 0$ using Bairstow method.

Ans. $1 \pm i, 3 \pm 4i$

3 Linear Systems of Equations

In this chapter, we shall study direct and iterative methods to solve linear system of equations. Among the direct methods, we shall study Gauss elimination method and its modification by Jordan, Crout, and triangularization methods. Among the iterative methods, we shall study Jacobi and Gauss–Seidel methods.

3.1 DIRECT METHODS

Matrix Inversion Method

Consider the system of n linear equations in n unknowns:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \dots &\quad \dots \quad \dots \quad \dots \quad \dots \\ \dots &\quad \dots \quad \dots \quad \dots \quad \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n \end{aligned} \tag{3.1}$$

The matrix form of the system (3.1) is

$$\mathbf{AX} = \mathbf{B}, \tag{3.2}$$

where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & \dots & a_{nn} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \dots \\ x_n \end{bmatrix}, \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \dots \\ b_n \end{bmatrix}.$$

Suppose \mathbf{A} is non-singular, that is, $\det \mathbf{A} \neq 0$. Then \mathbf{A}^{-1} exists. Therefore, premultiplying (3.2) by \mathbf{A}^{-1} , we get

$$\mathbf{A}^{-1}\mathbf{AX} = \mathbf{A}^{-1}\mathbf{B}$$

or

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}.$$

Thus, finding \mathbf{A}^{-1} we can determine \mathbf{X} and so x_1, x_2, \dots, x_n .

EXAMPLE 3.1

Solve the equations

$$\begin{aligned} x + y + 2z &= 1 \\ x + 2y + 3z &= 1 \\ 2x + 3y + z &= 2. \end{aligned}$$

Solution. We have

$$|A| = \begin{vmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 1 \end{vmatrix} = -4 \neq 0.$$

Also

$$A^{-1} = \frac{1}{4} \begin{bmatrix} 7 & -5 & 1 \\ -5 & 3 & 1 \\ 1 & 1 & -1 \end{bmatrix}.$$

Hence,

$$\begin{aligned} X &= \begin{bmatrix} x \\ y \\ z \end{bmatrix} = A^{-1}B = \frac{1}{4} \begin{bmatrix} 7 & -5 & 1 \\ -5 & 3 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \\ &= \frac{1}{4} \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

and so $x = 1, y = 0, z = 0$.

Gauss Elimination Method

This is the simplest method of step-by-step elimination and it reduces the system of equations to an equivalent upper triangular system, which can be solved by back substitution.

Let the system of equations be

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \dots &\quad \dots \quad \dots \quad \dots \quad \dots \\ \dots &\quad \dots \quad \dots \quad \dots \quad \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n \end{aligned}$$

The matrix form of this system is

$$AX = B,$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & \dots & a_{nn} \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ \dots \\ x_n \end{bmatrix}, \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ \dots \\ b_n \end{bmatrix}.$$

The augmented matrix is

$$[\mathbf{A} : \mathbf{B}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} & b_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} & b_n \end{array} \right].$$

The number a_{rr} at position (r, r) that is used to eliminate x_r in rows $r+1, r+2, \dots, n$ is called the r th pivotal element and the r th row is called the pivotal row. Thus, the augmented matrix can be written as

$$\text{pivot} \rightarrow \left[\begin{array}{cccc|c} \underline{a_{11}} & a_{12} & \cdots & \cdots & a_{1n} & b_1 \\ m_{2,1} = a_{21}/a_{11} & \underline{a_{22}} & \cdots & \cdots & a_{2n} & b_2 \\ a_{31} & a_{32} & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} & b_n \end{array} \right] \leftarrow \text{pivotal row}$$

The first row is used to eliminate elements in the first column below the diagonal. In the first step, the element a_{11} is pivotal element and the first row is pivotal row. The values $m_{k,1}$ are the multiples of row 1 that are to be subtracted from row k for $k = 2, 3, 4, \dots, n$. The result after elimination becomes

$$\text{pivot} \rightarrow \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & \cdots & a_{1n} & b_1 \\ \underline{c_{22}} & \cdots & \cdots & c_{2n} & d_2 & \cdots \\ m_{3,2} = c_{32}/c_{22} & c_{32} & \cdots & \cdots & c_{3n} & d_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ m_{n,2} = c_{n2}/c_{22} & c_{n2} & \cdots & \cdots & c_{nn} & d_n \end{array} \right] \leftarrow \text{pivotal row.}$$

The second row (now pivotal row) is used to eliminate elements in the second column that lie below the diagonal. The elements $m_{k,2}$ are the multiples of row 2 that are to be subtracted from row k for $k = 3, 4, \dots, n$.

Continuing this process, we arrive at the matrix:

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & \cdots & a_{1n} & b_1 \\ c_{22} & \cdots & \cdots & c_{2n} & d_2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ h_{nn} & & & & & p_n \end{array} \right].$$

Hence, the given system of equation reduces to

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ c_{22}x_2 + \cdots + c_{2n}x_n &= d_2 \\ \cdots & \cdots \cdots \\ \cdots & \cdots \cdots \\ h_{nn}x_n &= p_n. \end{aligned}$$

In the above set of equations, we observe that each equation has one lesser variable than its preceding equation. From the last equation, we have $x_n = \frac{p_n}{h_{nn}}$. Putting this value of x_n in the preceding equation, we can find x_{n-1} . Continuing in this way, putting the values of x_2, x_3, \dots, x_n in the first equation, x_1 can be determined. The process discussed here is called back substitution.

Remark 3.1. It may occur that the pivot element, even if it is different from zero, is very small and gives rise to large errors. The reason is that the small coefficient usually has been formed as the difference between two almost equal numbers. This difficulty is overcome by suitable permutations of the given equations. It is recommended therefore that the pivotal equation should be the equation which has the largest leading coefficient.

EXAMPLE 3.2

Express the following system in augmented matrix form and find an equivalent upper triangular system and the solution:

$$2x_1 + 4x_2 - 6x_3 = 4$$

$$x_1 + 5x_2 + 3x_3 = 10$$

$$x_1 + 3x_2 + 2x_3 = 5.$$

Solution. The augmented matrix for the system is

$$\begin{array}{l} \text{pivot} \rightarrow \left[\begin{array}{ccc|c} 2 & 4 & -6 & -4 \\ 1 & 5 & 3 & 10 \\ 1 & 3 & 2 & 5 \end{array} \right] \leftarrow \text{pivotal row} \\ m_{2,1} = 0.5 \\ m_{3,1} = 0.5 \end{array}$$

The result after first elimination is

$$\begin{array}{l} \text{pivot} \rightarrow \left[\begin{array}{ccc|c} 2 & 4 & -6 & -4 \\ 0 & 3 & 6 & 12 \\ 0 & 1 & 5 & 7 \end{array} \right] \leftarrow \text{pivotal row} \\ m_{3,2} = 1/3 \end{array}$$

The result after second elimination is

$$\left[\begin{array}{ccc|c} 2 & 4 & -6 & -4 \\ 0 & 3 & 6 & 12 \\ 0 & 0 & 3 & 3 \end{array} \right].$$

Therefore, back substitution yields

$$3x_3 = 3 \quad \text{and so } x_3 = 1,$$

$$3x_2 + 6x_3 = 12 \quad \text{and so } x_2 = 2,$$

$$2x_1 + 4x_2 - 6x_3 = -4 \quad \text{and so } x_1 = -3.$$

Hence, the solution is $x_1 = -3$, $x_2 = 2$, and $x_3 = 1$.

EXAMPLE 3.3

Solve by Gauss elimination method:

$$10x - 7y + 3z + 5u = 6$$

$$-6x + 8y - z - 4u = 5$$

$$3x + y + 4z + 11u = 2$$

$$5x - 9y - 2z + 4u = 7.$$

Solution. The augmented matrix for the given system is

$$\text{pivot} \rightarrow \left[\begin{array}{cccc|c} 10 & -7 & 3 & 5 & 6 \\ -6 & 8 & -1 & -4 & 5 \\ 3 & 1 & 4 & 11 & 2 \\ 5 & -9 & -2 & 4 & 7 \end{array} \right] \leftarrow \text{pivotal row}$$

$$m_{2,1} = -0.6$$

$$m_{3,1} = 0.3$$

$$m_{4,1} = 0.5$$

The first elimination yields

$$\text{pivot} \rightarrow \left[\begin{array}{cccc|c} 10 & -7 & 3 & 5 & 6 \\ 0 & 3.8 & 0.8 & -1 & 8.6 \\ 0 & 3.1 & 3.1 & 9.5 & 0.2 \\ 0 & -5.5 & -3.5 & 1.5 & 4 \end{array} \right] \leftarrow \text{pivotal row}$$

$$m_{3,2} = 0.81579$$

$$m_{4,2} = 1.4474$$

The result after second elimination is

$$\text{pivot} \rightarrow \left[\begin{array}{cccc|c} 10 & -7 & 3 & 5 & 6 \\ 0 & 3.8 & 0.8 & -1 & 8.6 \\ 0 & 0 & 2.4474 & 10.3158 & -6.8158 \\ 0 & 0 & -2.3421 & 0.0526 & 16.44764 \end{array} \right] \leftarrow \text{pivotal row}$$

$$m_{4,3} = -0.957$$

The result after third elimination is

$$\left[\begin{array}{cccc|c} 10 & -7 & 3 & 5 & 6 \\ 0 & 3.8 & 0.8 & -1 & 8.6 \\ 0 & 0 & 2.4474 & 10.3158 & -6.8158 \\ 0 & 0 & 0 & 9.9248 & 9.9249 \end{array} \right].$$

Therefore, back substitution yields

$$9.9248u = 9.9249 \text{ and so } u \approx 1$$

$$2.4474z + 10.3158u = -6.8158 \text{ and so } z = -6.9999 \approx -7$$

$$3.8y + 0.8z - u = 8.6 \text{ and so } y = 4$$

$$10x - 7y + 3z + 5u = 6 \text{ and so } x = 5.$$

Hence, the solution of the given system is $x = 5$, $y = 4$, $z = -7$, and $u = 1$.

EXAMPLE 3.4

Solve the following equations by Gauss elimination method:

$$2x + y + z = 10, 3x + 2y + 3z = 18, x + 4y + 9z = 16.$$

Solution. The given equations are

$$2x + y + z = 10, 3x + 2y + 3z = 18, x + 4y + 9z = 16.$$

The augmented matrix for given system of equations is

$$\text{pivot} \rightarrow \left[\begin{array}{cccc} 2 & 1 & 1 & 10 \end{array} \right] \leftarrow \text{pivotal row}$$

$$m_{2,1} = 3/2$$

$$m_{3,1} = 1/2$$

The result of first Gauss elimination is

$$\left[\begin{array}{cccc} 2 & 1 & 1 & 10 \\ 0 & \frac{1}{2} & \frac{3}{2} & 3 \\ 0 & \frac{7}{2} & \frac{17}{2} & 11 \end{array} \right] \leftarrow \text{pivotal row}$$

The second elimination yields

$$\left[\begin{array}{cccc} 2 & 1 & 1 & 10 \\ 0 & \frac{1}{2} & \frac{3}{2} & 3 \\ 0 & 0 & -2 & -1 \end{array} \right] 0$$

Thus, the given system equations reduces to

$$\begin{aligned} 2x + y + z &= 10 \\ 0.5y + 1.5z &= 3 \\ -2z &= -10. \end{aligned}$$

Hence, back substitution yields

$$z = 5, y = -9, x = 7.$$

Jordan Modification to Gauss Method

Jordan modification means that the elimination is performed not only in the equation below but also in the equation above the pivotal row so that we get a diagonal matrix. In this way, we have the solution without further computation.

Comparing the methods of Gauss and Jordan, we find that the number of operations is essentially $\frac{n^3}{3}$ for Gauss method and $\frac{n^3}{2}$ for Jordan method. Hence, Gauss method should usually be preferred over Jordan method.

To illustrate this modification we reconsider Example 3.2. The result of first elimination is unchanged and we have

$$\begin{aligned} m_{1,2} &= 4/3 \left[\begin{array}{ccc|c} 2 & 4 & -6 & -4 \\ 0 & 3 & 6 & 12 \end{array} \right] \leftarrow \text{pivotal row} \\ \text{pivot} \rightarrow & \\ m_{3,2} &= 1/3 \left[\begin{array}{ccc|c} 0 & 1 & 5 & 7 \end{array} \right] \end{aligned}$$

Now, the second elimination as per Jordan modification yields

$$\begin{aligned} m_{1,3} &= -14/3 \left[\begin{array}{ccc|c} 2 & 0 & -14 & -20 \\ 0 & 3 & 6 & 12 \end{array} \right] \\ \text{pivot} \rightarrow & \\ m_{2,3} &= 2 \left[\begin{array}{ccc|c} 0 & 0 & 3 & 3 \end{array} \right] \leftarrow \text{pivotal row} \end{aligned}$$

The third elimination as per Jordan modification yields

$$\left[\begin{array}{ccc|c} 2 & 0 & 0 & -6 \\ 0 & 3 & 0 & 6 \\ 0 & 0 & 3 & 3 \end{array} \right].$$

Hence,

$$\begin{aligned}2x_1 &= -6 \text{ and so } x_1 = -3, \\3x_2 &= 6 \text{ and so } x_2 = 2, \\3x_3 &= 3 \text{ and so } x_3 = 1.\end{aligned}$$

EXAMPLE 3.5

Solve

$$\begin{aligned}x + 2y + z &= 8 \\2x + 3y + 4z &= 2 \quad 0 \\4x + 3y + 2z &= 1 \quad 6\end{aligned}$$

by Gauss–Jordan method.

Solution. The augmented matrix for the given system of equations is

$$\begin{array}{l} \text{pivot} \rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & 8 \end{array} \right] \leftarrow \text{pivotal row} \\ m_{2,1} = 2 \left[\begin{array}{ccc|c} 1 & 2 & 1 & 8 \end{array} \right] \\ m_{3,1} = 4 \left[\begin{array}{ccc|c} 1 & 2 & 1 & 8 \end{array} \right] \end{array}$$

The result of first elimination is

$$\begin{array}{l} m_{1,2} = -2 \left[\begin{array}{ccc|c} 1 & 2 & 1 & 8 \end{array} \right] \\ \text{pivot} \rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & 8 \end{array} \right] \\ m_{3,2} = 5 \left[\begin{array}{ccc|c} 1 & 2 & 1 & 8 \end{array} \right] \end{array}$$

The second Gauss–Jordan elimination yields

$$\begin{array}{l} m_{1,3} = -5/12 \left[\begin{array}{ccc|c} 1 & 0 & 5 & 16 \end{array} \right] \\ m_{2,3} = -1/6 \left[\begin{array}{ccc|c} 1 & 0 & 5 & 16 \end{array} \right] \\ \text{pivot} \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 5 & 16 \end{array} \right] \end{array}$$

The third Gauss–Jordan elimination yields

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & -2 \\ 0 & 0 & -12 & -36 \end{array} \right].$$

Therefore, $x = 1$, $y = 2$, and $z = 3$ is the required solution.

EXAMPLE 3.6

Solve

$$\begin{aligned}10x + y + z &= 12 \\x + 10y + z &= 12 \\x + y + 10z &= 12\end{aligned}$$

by Gauss–Jordan method.

Solution. The augmented matrix for the given system is

$$\begin{array}{c} \text{pivot} \rightarrow \left[\begin{array}{ccc|c} 10 & 1 & 1 & 12 \end{array} \right] \leftarrow \text{pivotal row} \\ m_{2,1} = 1/10 \left[\begin{array}{ccc|c} 1 & 10 & 1 & 12 \end{array} \right] \\ m_{3,1} = 1/10 \left[\begin{array}{ccc|c} 1 & 1 & 10 & 12 \end{array} \right] \end{array}$$

The first Gauss–Jordan elimination yields

$$\begin{array}{c} m_{1,2} = 10/99 \left[\begin{array}{ccc|c} 10 & 1 & 1 & 12 \end{array} \right] \\ \text{pivot} \rightarrow \left[\begin{array}{ccc|c} 0 & 99/10 & 9/10 & 108/10 \end{array} \right] \leftarrow \text{pivotal row} \\ m_{3,2} = 1/11 \left[\begin{array}{ccc|c} 0 & 9/10 & 99/10 & 108/10 \end{array} \right] \end{array}$$

Now the Gauss–Jordan elimination gives

$$\begin{array}{c} m_{1,3} = 10/108 \left[\begin{array}{ccc|c} 10 & 0 & 10/11 & 120/11 \end{array} \right] \\ m_{2,3} = 11/120 \left[\begin{array}{ccc|c} 0 & 99/10 & 9/10 & 108/10 \end{array} \right] \\ \text{pivot} \rightarrow \left[\begin{array}{ccc|c} 0 & 0 & 108/11 & 108/11 \end{array} \right] \leftarrow \text{pivotal row} \end{array}$$

The next Gauss–Jordan elimination yields

$$\left[\begin{array}{ccc|c} 10 & 0 & 0 & 10 \\ 0 & 99/10 & 0 & 99/10 \\ 0 & 0 & 108/11 & 108/11 \end{array} \right].$$

Hence, the solution of the given system is $x = 1, y = 1, z = 1$.

EXAMPLE 3.7

Solve by Gauss–Jordan method

$$\begin{aligned} x + y + z &= 9 \\ 2x - 3y + 4z &= 13 \\ 3x + 4y + 5z &= 40. \end{aligned}$$

Solution. The augmented matrix for the given system is

$$\begin{array}{c} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 9 \end{array} \right] \leftarrow \text{pivotal row} \\ m_{21} = 2 \left[\begin{array}{ccc|c} 2 & -3 & 4 & 13 \end{array} \right] \\ m_{31} = 3 \left[\begin{array}{ccc|c} 3 & 4 & 5 & 40 \end{array} \right] \end{array}$$

The first Gauss–Jordan elimination yields

$$\begin{array}{c} m_{12} = -\frac{1}{5} \left[\begin{array}{cccc} 1 & 1 & 1 & 9 \\ 0 & -5 & 2 & -5 \end{array} \right] \leftarrow \text{pivotal row.} \\ m_{32} = -\frac{1}{5} \left[\begin{array}{cccc} 0 & 1 & 2 & 1 \\ 3 \end{array} \right] \end{array}$$

The second Gauss elimination yields

$$\begin{aligned} m_{13} &= 7/12 \begin{bmatrix} 1 & 0 & \frac{7}{5} & 8 \\ 0 & -5 & 2 & -5 \\ 0 & 0 & \frac{12}{5} & 12 \end{bmatrix} \\ m_{23} &= 10/12 \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & -5 & 0 & -15 \\ 0 & 0 & \frac{12}{5} & 12 \end{bmatrix} \leftarrow \text{pivotal row} \end{aligned}$$

The third Gauss elimination yields

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & -5 & 0 & -15 \\ 0 & 0 & \frac{12}{5} & 12 \end{bmatrix}.$$

Thus, we have attained the diagonal form of the system. Hence, the solution is

$$x = 1, y = \frac{15}{5} = 3, z = \frac{12(5)}{12} = 5.$$

Triangularization (Triangular Factorization) Method

We have seen that Gauss elimination leads to an upper triangular matrix, where all diagonal elements are 1. We shall now show that the elimination can be interpreted as the multiplication of the original coefficient matrix \mathbf{A} by a suitable lower triangular matrix. Hence, in three dimensions, we put

$$\begin{bmatrix} I_{11} & 0 & 0 \\ I_{21} & I_{22} & 0 \\ I_{31} & I_{32} & I_{33} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{bmatrix}.$$

In this way, we get nine equations with nine unknowns (six l elements and three u elements).

If the lower and upper triangular matrices are denoted by \mathbf{L} and \mathbf{U} , respectively, we have

$$\mathbf{LA} = \mathbf{U}$$

or

$$\mathbf{A} = \mathbf{L}^{-1}\mathbf{U}.$$

Since \mathbf{L}^{-1} is also a lower triangular matrix, we can find a factorization of \mathbf{A} as a product of one lower triangular matrix and one upper triangular matrix. Thus, a non-singular matrix \mathbf{A} is said to have a triangular factorization if it can be expressed as a product of a lower triangular matrix \mathbf{L} and an upper triangular matrix \mathbf{U} , that is, if $\mathbf{A} = \mathbf{LU}$. For the sake of convenience, we can choose $l_{ii} = 1$ or $u_{ii} = 1$. Thus, the system of equations $\mathbf{AX} = \mathbf{B}$ is resolved into two simple systems as follows:

$$\mathbf{AX} = \mathbf{B}$$

or

$$\mathbf{LUX} = \mathbf{B}$$

or

$$\mathbf{LY} = \mathbf{B} \text{ and } \mathbf{UX} = \mathbf{Y}.$$

Both the systems can be solved by back substitution.

EXAMPLE 3.8

Solve the following system of equations by triangularization method:

$$x_1 + 2x_2 + 3x_3 = 14$$

$$2x_1 + 5x_2 + 2x_3 = 18$$

$$3x_1 + x_2 + 5x_3 = 20.$$

Solution. The matrix form of the given system is

$$\mathbf{AX} = \mathbf{B},$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 3 & 1 & 5 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 14 \\ 18 \\ 20 \end{bmatrix}.$$

Let

$$\mathbf{A} = \mathbf{LU},$$

that is,

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 3 & 1 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

and so we have

$$1 = u_{11}$$

$$2 = l_{21}u_{11} \text{ and so } l_{21} = 2$$

$$3 = l_{31}u_{11} \text{ and so } l_{31} = 3$$

$$2 = u_{12}$$

$$5 = l_{21}u_{12} + u_{22} = 2(2) + u_{22} \text{ and so } u_{22} = 1$$

$$1 = l_{31}u_{12} + l_{32}u_{22} = 3(2) + l_{32}(1) \text{ and so } l_{32} = 5$$

$$3 = u_{13}$$

$$2 = l_{21}u_{13} + u_{23} = 2(3) + u_{23} \text{ and so } u_{23} = -4$$

$$5 = l_{31}u_{13} + l_{32}u_{23} + u_{33} = 3(3) + (-5)(-4) + u_{33} \text{ and so } u_{33} = -24.$$

Hence,

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & -5 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{U} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -4 \\ 0 & 0 & -24 \end{bmatrix}.$$

Now we have

$$\mathbf{AX} = \mathbf{B}$$

or

$$\mathbf{LUX} = \mathbf{B}$$

or

$$\mathbf{LY} = \mathbf{B} \text{ where } \mathbf{UX} = \mathbf{Y}.$$

But $\mathbf{LY} = \mathbf{B}$ yields

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & -5 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 18 \\ 20 \end{bmatrix}$$

and we have

$$y_1 = 14,$$

$$\begin{aligned} 2y_1 + y_2 &= 18 \text{ and so } y_2 = -10, \\ 3y_1 - 5y_2 + y_3 &= 20 \text{ and so } y_3 = -72. \end{aligned}$$

Then $\mathbf{U}\mathbf{X} = \mathbf{Y}$ yields

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -4 \\ 0 & 0 & -24 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ -10 \\ -72 \end{bmatrix}$$

and so

$$\begin{aligned} -24x_3 &= -72 \text{ which yields } x_3 = 3, \\ x_2 - 4x_3 &= -10 \text{ which yields } x_2 = 2, \\ x_1 + 2x_2 + x_3 &= 14 \text{ which yields } x_1 = 1. \end{aligned}$$

Hence, the required solution is $x_1 = 1$, $x_2 = 2$, and $x_3 = 3$.

EXAMPLE 3.9

Use Gauss elimination method to find triangular factorization of the coefficient matrix of the system

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 14 \\ 2x_1 + 5x_2 + 2x_3 &= 18 \\ 3x_1 + x_2 + 5x_3 &= 20 \end{aligned}$$

and hence solve the system.

Solution. In matrix form, we have

$$\mathbf{AX} = \mathbf{B},$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 3 & 1 & 5 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 14 \\ 18 \\ 20 \end{bmatrix}.$$

Write

$$\mathbf{A} = \mathbf{IA} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 3 & 1 & 5 \end{bmatrix} \leftarrow \begin{array}{l} \text{pivotal row} \\ m_{2,1} = 2 \\ m_{3,1} = 3 \end{array}$$

The elimination in the second member on the right-hand side is done by Gauss elimination method while in the first member l_{21} is replaced by m_{21} and l_{31} is replaced by m_{31} . Thus, the first elimination yields

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & \frac{1}{2} & -4 \\ 0 & -5 & -4 \end{bmatrix} \leftarrow \begin{array}{l} \text{pivotal row} \\ m_{3,2} = -5 \end{array}$$

Then the second elimination gives the required triangular factorization as

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & -5 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -4 \\ 0 & 0 & -24 \end{bmatrix} \\ &= \mathbf{LU}, \end{aligned}$$

where

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & -5 & 1 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -4 \\ 0 & 0 & -24 \end{bmatrix}.$$

The solution is then obtained as in Example 3.8.

EXAMPLE 3.10

Solve

$$\begin{aligned} 2x_1 + 4x_2 - 6x_3 &= -4 \\ x_1 + 5x_2 + 3x_3 &= 10 \\ x_1 + 3x_2 + 2x_3 &= 5. \end{aligned}$$

Solution. Write

$$\mathbf{A} = \mathbf{IA},$$

that is,

$$\begin{bmatrix} 2 & 4 & -6 \\ 1 & 5 & 3 \\ 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -6 \\ 1 & 5 & 3 \\ 1 & 3 & 2 \end{bmatrix} \leftarrow \begin{array}{l} \text{pivotal row} \\ m_{2,1} = 1/2 \\ m_{3,1} = 1/2 \end{array}$$

Using Gauss elimination method, discussed in Example 3.9, the first elimination yields

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 1/2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -6 \\ 0 & 3 & 6 \\ 0 & 1 & 5 \end{bmatrix} \leftarrow \begin{array}{l} \text{pivotal row} \\ m_{3,2} = 1/3 \end{array}$$

The second elimination yields

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 1/2 & 1/3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -6 \\ 0 & 3 & 6 \\ 0 & 0 & 3 \end{bmatrix} = \mathbf{LU}.$$

Therefore, $\mathbf{AX} = \mathbf{B}$ reduces to $\mathbf{LUX} = \mathbf{B}$ or $\mathbf{LY} = \mathbf{B}$, $\mathbf{UX} = \mathbf{Y}$.

Now $\mathbf{LY} = \mathbf{B}$ gives

$$\begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 1/2 & 1/3 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -4 \\ 10 \\ 5 \end{bmatrix}$$

and so

$$\begin{aligned} y_1 &= -4 \\ \frac{1}{2}y_1 + y_2 &= 10 \text{ which yields } y_2 = 12, \end{aligned}$$

$$\frac{1}{2}y_1 + \frac{1}{3}y_2 + y_3 = 5 \text{ which yields } y_3 = 3.$$

Then $\mathbf{UX} = \mathbf{Y}$ implies

$$\begin{bmatrix} 2 & 4 & -6 \\ 0 & 3 & 6 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -4 \\ 12 \\ 3 \end{bmatrix}$$

and so

$$\begin{aligned} 3x_3 &= 3 \text{ which yields } x_3 = 1, \\ 3x_2 + 6x_3 &= 12 \text{ which yields } x_2 = 2, \\ 2x_1 + 4x_2 - 6x_3 &= -4 \text{ which yields } x_1 = -3. \end{aligned}$$

Hence, the solution of the given system is $x_1 = -3$, $x_2 = 2$, and $x_3 = 1$.

EXAMPLE 3.11

Solve

$$\begin{aligned} x + 3y + 8z &= 4 \\ x + 4y + 3z &= -2 \\ x + 3y + 4z &= 1 \end{aligned}$$

by the method of factorization.

Solution. The matrix form of the system is $\mathbf{AX} = \mathbf{B}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 4 & 3 \\ 1 & 3 & 4 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix}.$$

Write

$$\mathbf{A} = \mathbf{IA} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 8 \\ 1 & 4 & 3 \\ 1 & 3 & 4 \end{bmatrix} \leftarrow \begin{array}{l} \text{pivotal row} \\ m_{2,1} = 1 \\ m_{3,1} = 1 \end{array}$$

Applying Gauss elimination method to the right member and replacing l_{21} by m_{21} and l_{31} by m_{31} in the left member, we get

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 8 \\ 0 & 1 & -5 \\ 0 & 0 & -4 \end{bmatrix} \leftarrow \begin{array}{l} \text{pivotal row} \\ \\ \end{array} \\ &= \mathbf{LU}. \end{aligned}$$

Then $\mathbf{AX} = \mathbf{B}$ reduces to $\mathbf{LUX} = \mathbf{B}$ or $\mathbf{LY} = \mathbf{B}$ and $\mathbf{UX} = \mathbf{Y}$. Now $\mathbf{LY} = \mathbf{B}$ gives

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix}$$

and so

$$y_1 = 4, y_2 = -6, y_1 + y_3 = 1 \text{ which implies } y_3 = -3.$$

Then $\mathbf{UX} = \mathbf{Y}$ gives

$$\begin{bmatrix} 1 & 3 & 8 \\ 0 & 1 & -5 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ -6 \\ -3 \end{bmatrix}.$$

Hence, the required solution is $x = \frac{19}{4}$, $y = -\frac{9}{4}$, $z = \frac{3}{4}$.

Triangularization of Symmetric Matrix

When the coefficient matrix of the system of linear equations is symmetric, we can have a particularly simple triangularization in the form

$$\mathbf{A} = \mathbf{LL}^T$$

or

$$\begin{bmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ a_{12} & a_{22} & \dots & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & \dots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & \dots & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & l_{n-1,n-1} & 0 \\ l_{n1} & l_{n2} & \dots & \dots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & \dots & \dots & l_{n1} \\ 0 & l_{22} & \dots & \dots & l_{n2} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & l_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} l_{11}^2 & l_{11}l_{21} & \dots & \dots & l_{11}l_{n1} \\ l_{11}l_{21} & l_{21}^2 + l_{22}^2 & \dots & \dots & l_{21}l_{n1} + l_{22}l_{n2} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ l_{n1}l_{11} & l_{n1}l_{21} + l_{n2}l_{22} & \dots & \dots & l_{n1}^2 + l_{n2}^2 + \dots + l_{nn}^2 \end{bmatrix}$$

Hence,

$$\begin{aligned} l_{11}^2 &= a_{11}, & l_{21}l_{31} + l_{22}l_{32} &= a_{23}, & l_{21}^2 + l_{22}^2 &= a_{22} \\ l_{11}l_{21} &= a_{12}, & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ l_{11}l_{n1} &= a_{1n}, & l_{21}l_{n1} + l_{22}l_{n2} &= a_{2n}, & l_{n1}^2 + \dots + l_{nn}^2 &= a_{nn} \end{aligned}$$

However, it may encounter with some terms which are purely imaginary but this does not imply any special complications. The matrix equation $\mathbf{AX} = \mathbf{B}$ reduces to $\mathbf{LL}^T\mathbf{X} = \mathbf{B}$ or $\mathbf{LZ} = \mathbf{B}$ and $\mathbf{L}^T\mathbf{X} = \mathbf{Z}$.

This method is known as the square root method and is due to Banachiewicz and Dwyer.

EXAMPLE 3.12

Solve by square root method:

$$\begin{aligned} 4x - y + 2z &= 12 \\ -x + 5y + 3z &= 10 \\ 2x + 3y + 6z &= 18. \end{aligned}$$

Solution. The matrix form of the given system is

$$\mathbf{AX} = \mathbf{B},$$

where

$$\mathbf{A} = \begin{bmatrix} 4 & -1 & 2 \\ -1 & 5 & 3 \\ 2 & 3 & 6 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 12 \\ 10 \\ 18 \end{bmatrix}.$$

The matrix \mathbf{A} is symmetric. Therefore, we have triangularization of the type $\mathbf{A} = \mathbf{L}\mathbf{L}^T$, that is,

$$\begin{bmatrix} 4 & -1 & 2 \\ -1 & 5 & 3 \\ 2 & 3 & 6 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix} \\ = \begin{bmatrix} l_{11}^2 & l_{11}l_{21} & l_{11}l_{31} \\ l_{11}l_{21} & l_{21}^2 + l_{22}^2 & l_{21}l_{31} + l_{22}l_{32} \\ l_{11}l_{31} & l_{21}l_{31} + l_{22}l_{32} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}.$$

Hence,

$$l_{11}^2 = 4 \text{ and so } l_{11} = 2,$$

$$l_{11}l_{21} = -1 \text{ and so } l_{21} = -\frac{1}{2},$$

$$l_{11}l_{31} = 2 \text{ and so } l_{31} = 1,$$

$$l_{21}^2 + l_{22}^2 = 5 \text{ and so } l_{22} = \sqrt{5 - \frac{1}{4}} = \sqrt{\frac{19}{4}},$$

$$l_{21}l_{31} + l_{22}l_{32} = 3 \text{ and so } -\frac{1}{2} + \sqrt{\frac{19}{4}}l_{32} = 3 \text{ or } l_{32} = \frac{7}{\sqrt{19}}.$$

$$l_{31}^2 + l_{32}^2 + l_{33}^2 = 6 \text{ and so } 1 + \frac{49}{19} + l_{33}^2 = 6 \text{ or } l_{33} = \sqrt{\frac{46}{19}}.$$

Thus,

$$\mathbf{L} = \begin{bmatrix} 2 & 0 & 0 \\ -\frac{1}{2} & \sqrt{\frac{19}{4}} & 0 \\ 1 & \frac{7}{\sqrt{19}} & \sqrt{\frac{46}{19}} \end{bmatrix}.$$

Then, $\mathbf{LZ} = \mathbf{B}$ yields

$$\begin{bmatrix} 2 & 0 & 0 \\ -\frac{1}{2} & \sqrt{\frac{19}{4}} & 0 \\ 1 & \frac{7}{\sqrt{19}} & \sqrt{\frac{46}{19}} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 12 \\ 10 \\ 18 \end{bmatrix}$$

and so

$$z_1 = 6 \\ -3 + \sqrt{\frac{19}{4}}z_2 = 10 \text{ which yields } z_2 = \frac{26}{\sqrt{19}}.$$

$$6 + \frac{7}{\sqrt{19}} \times \frac{26}{\sqrt{19}} + \sqrt{\frac{46}{19}} z_3 = 18, \text{ which yields } z_3 = \sqrt{\frac{46}{19}}.$$

Now $\mathbf{L}^T \mathbf{X} = \mathbf{Z}$ gives

$$\begin{bmatrix} 2 & -\frac{1}{2} & 1 \\ 0 & \sqrt{\frac{19}{4}} & \frac{7}{\sqrt{19}} \\ 0 & 0 & \sqrt{\frac{46}{19}} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ \frac{26}{\sqrt{19}} \\ \sqrt{\frac{46}{19}} \end{bmatrix}.$$

Hence,

$$\begin{aligned} z &= 1, \\ \sqrt{\frac{19}{4}} y + \frac{7}{\sqrt{19}} z &= \frac{26}{\sqrt{19}} \text{ or } y = \sqrt{19} \times \frac{\sqrt{4}}{\sqrt{19}} = 2, \\ 2x - \frac{1}{2} y + z &= 6 \text{ which gives } x = 3. \end{aligned}$$

Hence, the solution is $x = 3$, $y = 2$, and $z = 1$.

Crout's Method

Crout suggested a technique to determine systematically the entries of the lower and upper triangles in the factorization of a given matrix \mathbf{A} . We describe the scheme of the method stepwise.

Let the matrix form of the system (in three dimensions) be $\mathbf{AX} = \mathbf{B}$, where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

The augmented matrix is

$$[\mathbf{A} : \mathbf{B}] = \left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{array} \right].$$

The matrix of the unknowns (in factorization of \mathbf{A}), called the derived matrix or auxiliary matrix, is

$$\left[\begin{array}{ccc|c} l_{11} & u_{12} & u_{13} & y_1 \\ l_{21} & l_{22} & u_{23} & y_2 \\ l_{31} & l_{32} & l_{33} & y_3 \end{array} \right].$$

The entries of this matrix are calculated as follows:

Step 1. The first column of the auxiliary matrix is identical with the first column of the augmented matrix $[\mathbf{A} : \mathbf{B}]$.

Step 2. The first row to the right of the first column of the auxiliary matrix is obtained by dividing the corresponding elements in $[\mathbf{A} : \mathbf{B}]$ by the leading diagonal element a_{11} .

Step 3. The remaining entries in the second column of the auxiliary matrix are l_{22} and l_{32} . These entries are equal to corresponding element in $[\mathbf{A} : \mathbf{B}]$ minus the product of the first element in that row and in that column. Thus,

$$\begin{aligned}l_{22} &= a_{22} - l_{21}u_{12}, \\l_{32} &= a_{32} - l_{31}u_{12}.\end{aligned}$$

Step 4. The remaining elements of the second row of the auxiliary matrix are equal to:

[corresponding element in $[\mathbf{A} : \mathbf{B}]$ minus the product of the first element in that row and first element in that column]/leading diagonal element in that row. Thus,

$$u_{23} = \frac{a_{23} - l_{21}u_{13}}{l_{22}}$$

$$y_2 = \frac{b_2 - l_{21}y_1}{l_{22}}.$$

Step 5. The remaining elements of the third column of the auxiliary matrix are equal to:

corresponding element in $[\mathbf{A} : \mathbf{B}]$ minus the sum of the inner products of the previously calculated elements in the same row and column. Thus

$$l_{33} = a_{33} - (l_{31}u_{13} + l_{32}u_{23}).$$

Step 6. The remaining elements of the third row of the auxiliary matrix are equal to:

[corresponding element in $[\mathbf{A} : \mathbf{B}]$ minus the sum of inner products of the previously calculated elements in the same row and column]/leading diagonal element in that row. Thus,

$$y_3 = \frac{b_3 - (l_{31}y_1 + l_{32}y_2)}{l_{33}}.$$

Following this scheme, the upper and lower diagonal matrices can be found and then using

$$\mathbf{U}\mathbf{x} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix},$$

we can determine x_1, x_2, x_3 .

EXAMPLE 3.13

Solve by Crout's method:

$$\begin{aligned}x_1 + 2x_2 + 3x_3 &= 1 \\3x_1 + x_2 + x_3 &= 0 \\2x_1 + x_2 + x_3 &= 0.\end{aligned}$$

Solution. The augmented matrix of the system is

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 3 & 1 & 1 & 0 \\ 2 & 1 & 1 & 0 \end{array} \right].$$

Let the derived matrix be

$$\mathbf{M} = \begin{bmatrix} l_{11} & u_{12} & u_{13} & y_1 \\ l_{21} & l_{22} & u_{23} & y_2 \\ l_{31} & l_{32} & l_{33} & y_3 \end{bmatrix}.$$

Then

$$\begin{aligned} \mathbf{M} &= \begin{bmatrix} 1 & \frac{2}{1} & \frac{3}{1} & \frac{1}{1} \\ 3 & 1-3(2) & \frac{1-3(3)}{-5} & \frac{0-3(1)}{-5} \\ 2 & 1-2(2) & 1-\left[3(2)+(-3)\left(\frac{8}{5}\right)\right] & \frac{0-[2(1)+(-3)(3/5)]}{1-[6-(24/5)]} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 2 & 3 & 1 \\ 3 & -5 & 8/5 & 3/5 \\ 2 & -3 & -1/5 & 1 \end{bmatrix} \end{aligned}$$

Now $\mathbf{U}\mathbf{X} = \mathbf{Y}$ gives

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 8/5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3/5 \\ 1 \end{bmatrix}.$$

Hence,

$$\begin{aligned} x_3 &= 1 \\ x_2 + \frac{8}{5}x_3 &= \frac{3}{5} \text{ and so } x_2 = \frac{3}{5} - \frac{8}{5} = -1 \end{aligned}$$

$$x_1 + 2x_2 + 3x_3 = 1 \text{ and so } x_1 = 1 - 2x_2 - 3x_3 = 1 + 2 - 3 = 0.$$

Hence, the solution is $x_1 = 0$, $x_2 = -1$, and $x_3 = 1$.

EXAMPLE 3.14

Solve by Crout's method:

$$2x + y + 4z = 12$$

$$8x - 3y + 2z = 20$$

$$4x + 11y - z = 33.$$

Solution. The augmented matrix for the given system of equations is

$$\left[\begin{array}{ccc|c} 2 & 1 & 4 & 12 \\ 8 & -3 & 2 & 20 \\ 4 & 11 & -1 & 33 \end{array} \right].$$

Let the derived matrix be

$$\mathbf{M} = \left[\begin{array}{cccc} l_{11} & u_{12} & u_{13} & y_1 \\ l_{21} & l_{22} & u_{23} & y_2 \\ l_{31} & l_{32} & l_{33} & y_3 \end{array} \right].$$

Then

$$\begin{aligned}\mathbf{M} &= \begin{bmatrix} 2 & \frac{1}{2} & \frac{4}{2} & \frac{12}{2} \\ 8 & -3 - 8\left(\frac{1}{2}\right) & \frac{2 - [8(2)]}{-7} & \frac{20 - [8(6)]}{-7} \\ 4 & 11 - 4\left(\frac{1}{2}\right) & -1 - [4(2) + 9(2)] & \frac{3 - [6(4) + 9(4)]}{-27} \end{bmatrix} \\ &= \begin{bmatrix} 2 & 1/2 & 2 & 6 \\ 8 & -7 & 2 & 4 \\ 4 & 9 & -27 & 1 \end{bmatrix}.\end{aligned}$$

Now $\mathbf{U}\mathbf{X} = \mathbf{Y}$ gives

$$\begin{bmatrix} 1 & 1/2 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 1 \end{bmatrix}.$$

By back substitution, we get

$$\begin{aligned}z &= 1, \\ y + 2z &= 4 \text{ and so } y = 4 - 2z = 2, \\ x + \frac{1}{2}y + 2z &= 6 \text{ and so } x = 6 - 2z - \frac{1}{2}y = 3.\end{aligned}$$

Hence, the required solution is $x = 3, y = 2, z = 1$.

EXAMPLE 3.15

Using Crout's method, solve the system

$$x + 2y - 12z + 8v = 27$$

$$5x + 4y + 7z - 2v = 4$$

$$-3x + 7y + 9z + 5v = 11$$

$$6x - 12y - 8z + 3v = 49.$$

Solution. The augmented matrix of the given system is

$$\left[\begin{array}{cccc|c} 1 & 2 & -12 & 8 & 27 \\ 5 & 4 & 7 & -2 & 4 \\ -3 & 7 & 9 & 5 & 11 \\ 6 & -12 & -8 & 3 & 49 \end{array} \right].$$

Then the auxiliary matrix is

$$\mathbf{M} = \left[\begin{array}{ccccc} 1 & 2 & -12 & 8 & 27 \\ 5 & -6 & -67/6 & 7 & 131/6 \\ -3 & 13 & 709/6 & -372/709 & -1151/709 \\ 6 & -24 & -204 & 11319/709 & 5 \end{array} \right].$$

The solution of the equation is given by $\mathbf{UX} = \mathbf{Y}$, that is,

$$\begin{bmatrix} 1 & 2 & -12 & 8 \\ 0 & 1 & -67/6 & 7 \\ 0 & 0 & 1 & -372/709 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ v \end{bmatrix} = \begin{bmatrix} 27 \\ 131/6 \\ -1151/709 \\ 5 \end{bmatrix}$$

or

$$\begin{aligned} x + 2y - 12z + 8v &= 27 \\ y - \frac{67}{6}z + 7v &= \frac{131}{6} \\ z - \frac{372}{709}v &= \frac{1,151}{709} \\ v &= 5. \end{aligned}$$

Back substitution yields

$$x = 3, y = -2, z = 1, v = 5.$$

3.2 ITERATIVE METHODS FOR LINEAR SYSTEMS

We have seen that the direct methods for the solution of simultaneous linear equations yield the solution after an amount of computation that is known in advance. On the other hand, in case of iterative or indirect methods, we start from an approximation to the true solution and, if convergent, we form a sequence of closer approximations repeated till the required accuracy is obtained. The difference between direct and iterative method is therefore that in direct method the amount of computation is fixed, while in an iterative method, the amount of computation depends upon the accuracy required.

In general, we prefer a direct method for solving system of linear equations. But, in case of matrices with a large number of zero elements, it is economical to use iterative methods.

Jacobi Iteration Method

Consider the system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ a_{31}x_1 + a_{32}x_2 + \dots + a_{3n}x_n &= b_3 \\ \dots &\quad \dots \quad \dots \quad \dots \quad \dots \\ \dots &\quad \dots \quad \dots \quad \dots \quad \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n \end{aligned} \tag{3.3}$$

in which the diagonal coefficients a_{ii} do not vanish. If this is not the case, the equations should be rearranged so that this condition is satisfied. Equations (3.3) can be written as

$$\begin{aligned} x_1 &= \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}}x_2 - \dots - \frac{a_{1n}}{a_{11}}x_n \\ x_2 &= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}}x_1 - \dots - \frac{a_{2n}}{a_{22}}x_n \\ \dots &\quad \dots \quad \dots \quad \dots \\ x_n &= \frac{b_n}{a_{nn}} - \frac{a_{n1}}{a_{nn}}x_1 - \dots - \frac{a_{n-n}}{a_{nn}}x_{n-1} \end{aligned} \tag{3.4}$$

Suppose $x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}$ are first approximation to the unknowns x_1, x_2, \dots, x_n . Substituting in the right side of equation (3.4), we find a system of second approximations:

$$\begin{aligned} x_1^{(2)} &= \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}} x_2^{(1)} - \dots - \frac{a_{1n}}{a_{11}} x_n^{(1)} \\ x_2^{(2)} &= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1^{(1)} - \dots - \frac{a_{2n}}{a_{22}} x_n^{(1)} \\ &\dots \quad \dots \quad \dots \quad \dots \\ &\dots \quad \dots \quad \dots \quad \dots \\ x_n^{(2)} &= \frac{b_n}{a_{nn}} - \frac{a_{n1}}{a_{nn}} x_1^{(1)} - \dots - \frac{a_{n,n-1}}{a_{nn}} x_{n-1}^{(1)} \end{aligned}$$

In general, if $x_1^{(n)}, x_2^{(n)}, \dots, x_n^{(n)}$ is a system of n th approximations, then the next approximation is given by the formula

$$\begin{aligned} x_1^{(n+1)} &= \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}} x_2^{(n)} - \dots - \frac{a_{1n}}{a_{11}} x_n^{(n)} \\ x_2^{(n+1)} &= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1^{(n)} - \dots - \frac{a_{2n}}{a_{22}} x_n^{(n)} \\ &\dots \quad \dots \quad \dots \quad \dots \\ x_n^{(n+1)} &= \frac{b_n}{a_{nn}} - \frac{a_{n1}}{a_{nn}} x_1^{(n)} - \dots - \frac{a_{n,n-1}}{a_{nn}} x_{n-1}^{(n)} \end{aligned}$$

This method, due to Jacobi, is called the method of simultaneous displacements or Jacobi method.

Gauss–Seidel Method

A simple modification of Jacobi method yields faster convergence. Let $x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}$ be the first approximation to the unknowns x_1, x_2, \dots, x_n . Then the second approximations are given by:

$$\begin{aligned} x_1^{(2)} &= \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}} x_2^{(1)} - \dots - \frac{a_{1n}}{a_{11}} x_n^{(1)} \\ x_2^{(2)} &= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1^{(2)} - \frac{a_{23}}{a_{22}} x_3^{(1)} - \dots - \frac{a_{2n}}{a_{22}} x_n^{(1)} \\ x_3^{(2)} &= \frac{b_3}{a_{33}} - \frac{a_{31}}{a_{33}} x_1^{(2)} - \frac{a_{32}}{a_{33}} x_2^{(2)} - \dots - \frac{a_{3n}}{a_{33}} x_n^{(1)} \\ &\dots \quad \dots \quad \dots \quad \dots \quad \dots \\ x_n^{(2)} &= \frac{b_n}{a_{nn}} - \frac{a_{n1}}{a_{nn}} x_1^{(2)} - \frac{a_{n2}}{a_{nn}} x_2^{(2)} - \dots - \frac{a_{n,n-1}}{a_{nn}} x_{n-1}^{(2)} \end{aligned}$$

The entire process is repeated till the values of x_1, x_2, \dots, x_n are obtained to the accuracy required. Thus, this method uses an improved component as soon as available and so is called the method of successive displacements or Gauss–Seidel method.

Introducing the matrices

$$\mathbf{A}_1 = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix} \text{ and } \mathbf{A}_2 = \begin{bmatrix} 0 & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & 0 & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

it can be shown that the condition for convergence of Gauss–Seidel method is that the absolutely largest eigenvalue of $\mathbf{A}_1^{-1}\mathbf{A}_2$ must be absolutely less than 1. In fact, we have convergence if for $i=1, 2, \dots, n$, $|a_{ii}| > S_i$, where $S_i = \sum_{k \neq i} |a_{ik}|$. Thus, for convergence, the coefficient matrix should have a clear diagonal dominance.

It may be mentioned that Gauss–Seidel method converges twice as fast as the Jacobi's method.

EXAMPLE 3.16

Starting with $(x_0, y_0, z_0) = (0, 0, 0)$ and using Jacobi method, find the next five iterations for the system

$$5x - y + z = 10$$

$$2x + 8y - z = 11$$

$$-x + y + 4z = 3.$$

Solution. The given equations can be written in the form

$$x = \frac{y - z + 10}{5}, \quad y = \frac{-2x + z + 11}{8}, \quad \text{and} \quad z = \frac{x - y + 3}{4}, \text{ respectively.}$$

Therefore, starting with $(x_0, y_0, z_0) = (0, 0, 0)$, we get

$$x_1 = \frac{y_0 - z_0 + 10}{5} = 2$$

$$y_1 = \frac{-2x_0 + z_0 + 11}{8} = 1.375$$

$$z_1 = \frac{x_0 - y_0 + 3}{4} = 0.75.$$

The second iteration gives

$$x_2 = \frac{y_1 - z_1 + 10}{5} = \frac{1.375 - 0.75 + 10}{5} = 2.125$$

$$y_2 = \frac{-2x_1 + z_1 + 11}{8} = \frac{-4 + 0.75 + 11}{8} = 0.96875$$

$$z_2 = \frac{x_1 - y_1 + 3}{4} = \frac{2 - 1.375 + 3}{4} = 0.90625.$$

The third iteration gives

$$x_3 = \frac{y_2 - z_2 + 10}{5} = \frac{0.96875 - 0.90625 + 10}{5} = 2.0125$$

$$y_3 = \frac{-2x_2 + z_2 + 11}{8} = \frac{-4.250 + 0.90625 + 11}{8} = 0.95703125$$

$$z_3 = \frac{x_2 - y_2 + 3}{4} = \frac{2.125 - 0.96875 + 3}{4} = 1.0390625.$$

The fourth iteration yields

$$x_4 = \frac{y_3 - z_3 + 10}{5} = \frac{0.95703125 - 1.0390625 + 10}{5} = 1.98359375$$

$$y_4 = \frac{-2x_3 + z_3 + 11}{8} = \frac{-4.0250 + 1.0390625 + 11}{8} = 0.8767578$$

$$z_4 = \frac{x_3 - y_3 + 3}{4} = \frac{2.0125 - 0.95703125 + 3}{4} = 1.0138672,$$

whereas the fifth iteration gives

$$x_5 = \frac{y_4 - z_4 + 10}{5} = 1.9725781$$

$$y_5 = \frac{-2x_4 + z_4 + 11}{8} = \frac{-3.9671875 + 1.0138672 + 11}{8} = 1.005834963$$

$$z_5 = \frac{x_4 - y_4 + 3}{4} = \frac{1.98359375 - 0.8767578 + 3}{4} = 1.02670898.$$

One finds that the iterations converge to (2, 1, 1).

EXAMPLE 3.17

Using Gauss–Seidel iteration and the first iteration as (0, 0, 0), calculate the next three iterations for the solution of the system of equations given in Example 3.16.

Solution. The first iteration is (0, 0, 0). The next iteration is

$$x_1 = \frac{y_0 - z_0 + 10}{5} = 2$$

$$y_1 = \frac{-2x_1 + z_0 + 11}{8} = \frac{-4 + 0 + 11}{8} = 0.875$$

$$z_1 = \frac{x_1 - y_1 + 3}{4} = \frac{2 - 0.875 + 3}{4} = 1.03125.$$

Then

$$x_2 = \frac{y_1 - z_1 + 10}{5} = \frac{0.875 - 1.03125 + 10}{5} = 1.96875$$

$$y_2 = \frac{-2x_2 + z_1 + 11}{8} = \frac{-3.9375 + 1.03125 + 11}{8} = 1.01171875$$

$$z_2 = \frac{x_2 - y_2 + 3}{4} = \frac{1.96875 - 1.01171875 + 3}{4} = 0.989257812.$$

Further,

$$x_3 = \frac{y_2 - z_2 + 10}{5} = \frac{1.01171875 - 0.989257812 + 10}{5} = 2.004492188$$

$$y_3 = \frac{-2x_3 + z_2 + 11}{8} = \frac{-4.008984376 + 0.989257812 + 11}{8} = 0.997534179$$

$$z_3 = \frac{x_3 - y_3 + 3}{4} = \frac{2.004492188 - 0.997534179 + 3}{4} = 1.001739502.$$

The iterations will converge to (2, 1, 1).

Remark 3.2. It follows from Examples 3.16 and 3.17 that Gauss–Seidel method converges rapidly in comparison to Jacobi's method.

EXAMPLE 3.18

Solve

$$\begin{aligned} 54x + y + z &= 110 \\ 2x + 15y + 6z &= 72 \\ -x + 6y + 27z &= 85 \end{aligned}$$

by Gauss–Seidel method.

Solution. From the given equations, we have

$$x = \frac{110 - y - z}{54}, \quad y = \frac{72 - 2x - 6z}{15}, \quad \text{and} \quad z = \frac{85 + x - 6y}{27}.$$

We take the initial approximation as $x_0 = y_0 = z_0 = 0$. Then the first approximation is given by

$$\begin{aligned} x_1 &= \frac{110}{54} = 2.0370 \\ y_1 &= \frac{72 - 2x_1 - 6z_0}{15} = 4.5284 \\ z_1 &= \frac{85 + x_1 - 6y_1}{27} = 2.2173. \end{aligned}$$

The second approximation is given by

$$\begin{aligned} x_2 &= \frac{110 - y_1 - z_1}{54} = 1.9122 \\ y_2 &= \frac{72 - 2x_2 - 6z_1}{15} = 3.6581 \\ z_2 &= \frac{85 + x_2 - 6y_2}{27} = 2.4061. \end{aligned}$$

The third approximation is

$$\begin{aligned} x_3 &= \frac{110 - y_2 - z_2}{54} = 1.9247 \\ y_3 &= \frac{72 - 2x_3 - 6z_2}{15} = 3.5809 \\ z_3 &= \frac{85 + x_3 - 6y_3}{27} = 2.4237. \end{aligned}$$

The fourth approximation is

$$\begin{aligned} x_4 &= \frac{110 - y_3 - z_3}{54} = 1.9258 \\ y_4 &= \frac{72 - 2x_4 - 6z_3}{15} = 3.5738 \end{aligned}$$

$$z_4 = \frac{85 + x_4 - 6y_4}{27} = 2.4253.$$

The fifth approximation is

$$x_5 = \frac{110 - y_4 - z_4}{54} = 1.9259$$

$$y_5 = \frac{72 - 2x_5 - 6z_4}{15} = 3.5732$$

$$z_5 = \frac{85 + x_5 - 6y_5}{27} = 2.4254.$$

Thus, the required solution, correct to three decimal places, is

$$x = 1.926, y = 3.573, z = 2.425.$$

EXAMPLE 3.19

Solve

$$\begin{aligned} 28x + 4y - z &= 32 \\ 2x + 17y + 4z &= 35 \\ x + 3y + 10z &= 24 \end{aligned}$$

by Gauss–Seidel method.

Solution. From the given equations, we have

$$x = \frac{32 - 4y + z}{28}, \quad y = \frac{35 - 2x - 4z}{17}, \quad \text{and} \quad z = \frac{24 - x - 3y}{10}.$$

Taking first approximation as $x_0 = y_0 = z_0 = 0$, we have

$$\begin{array}{lll} x_1 = 1.1428571, & y_1 = 1.9243697, & z_1 = 1.7084034 \\ x_2 = 0.9289615, & y_2 = 1.5475567, & z_2 = 1.8428368 \\ x_3 = 0.9875932, & y_3 = 1.5090274, & z_3 = 1.8485325 \\ x_4 = 0.9933008, & y_4 = 1.5070158, & z_4 = 1.8485652 \\ x_5 = 0.9935893, & y_5 = 1.5069741, & z_5 = 1.8485488 \\ x_6 = 0.9935947, & y_6 = 1.5069774, & z_6 = 1.8485473. \end{array}$$

Hence the solution, correct to four decimal places, is

$$x = 0.9935, y = 1.5069, z = 1.8485.$$

EXAMPLE 3.20

Solve the equation by Gauss–Seidel method:

$$\begin{aligned} 20x + y - 2z &= 17 \\ 3x + 20y - z &= -18 \\ 2x - 3y + 20z &= 25. \end{aligned}$$

Solution. The given equation can be written as

$$x = \frac{1}{20}[17 - y + 2z]$$

$$y = \frac{1}{20}[-18 - 3x + z]$$

$$z = \frac{1}{20}[25 - 3x + 3y].$$

Taking the initial rotation as $(x_0, y_0, z_0) = (0, 0, 0)$, we have by Gauss–Seidal method,

$$x_1 = \frac{1}{20}[17 - 0 + 0] = 0.85$$

$$y_1 = \frac{1}{20}[-18 - 3(0.85) + 1] = -1.0275$$

$$z_1 = \frac{1}{20}[25 - 2(0.85) - 3(-1.0275)] = 1.0108$$

$$x_2 = \frac{1}{20}[17 + 1.0275 + 2(1.0108)] = 1.0024$$

$$y_2 = \frac{1}{20}[-18 - 3(1.0024) + 1.0108] = -0.9998$$

$$z_2 = \frac{1}{20}[25 - 2(1.0024) + 3(-0.9998)] = 0.9998$$

$$x_3 = \frac{1}{20}[17 + 0.9998 + 2(0.9998)] = 0.99997$$

$$y_3 = \frac{1}{20}[-18 - 3(0.99997) + 0.9998] = -1.00000$$

$$z_3 = \frac{1}{20}[25 - 2(0.99997) + 3(-1.00000)] = 1.00000.$$

The second and third iterations show that the solution of the given system of equations is $x = 1$, $y = -1$, $z = 1$.

Convergence of Iteration Method

(A) Condition of Convergence of Iteration Methods

We know (see Section 2.14) that conditions for convergence of the iteration process for solving simultaneous equations $f(x, y) = 0$ and $g(x, y) = 0$ is

$$\left| \frac{\partial f}{\partial x} \right| + \left| \frac{\partial g}{\partial x} \right| < 1$$

and

$$\left| \frac{\partial f}{\partial y} \right| + \left| \frac{\partial g}{\partial y} \right| < 1.$$

This result can be extended to any finite number of equations. For example, consider the following system of three equations:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3.$$

Then, in the fixed-point form, we have

$$\begin{aligned}x_1 &= f(x_1, x_2, x_3) = \frac{1}{a_{11}}(b_1 - a_{12}x_2 - a_{13}x_3) \\&= \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}}x_2 - \frac{a_{13}}{a_{11}}x_3,\end{aligned}\tag{3.5}$$

$$\begin{aligned}x_2 &= g(x_1, x_2, x_3) = \frac{1}{a_{22}}(b_2 - a_{21}x_1 - a_{23}x_3) \\&= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}}x_1 - \frac{a_{23}}{a_{33}}x_3,\end{aligned}\tag{3.6}$$

$$\begin{aligned}x_3 &= h(x_1, x_2, x_3) = \frac{1}{a_{33}}(b_3 - a_{31}x_1 - a_{32}x_2) \\&= \frac{b_3}{a_{33}} - \frac{a_{31}}{a_{33}}x_1 - \frac{a_{32}}{a_{33}}x_2.\end{aligned}\tag{3.7}$$

Then the conditions for convergence are

$$\left| \frac{\partial f}{\partial x_1} \right| + \left| \frac{\partial g}{\partial x_1} \right| + \left| \frac{\partial h}{\partial x_1} \right| < 1, \tag{3.8}$$

$$\left| \frac{\partial f}{\partial x_2} \right| + \left| \frac{\partial g}{\partial x_2} \right| + \left| \frac{\partial h}{\partial x_2} \right| < 1, \tag{3.9}$$

$$\left| \frac{\partial f}{\partial x_3} \right| + \left| \frac{\partial g}{\partial x_3} \right| + \left| \frac{\partial h}{\partial x_3} \right| < 1. \tag{3.10}$$

But partial differentiation of equations (3.5), (3.6), and (3.7) yields

$$\frac{\partial f}{\partial x_1} = 0, \quad \frac{\partial f}{\partial x_2} = -\frac{a_{12}}{a_{11}}, \quad \frac{\partial f}{\partial x_3} = -\frac{a_{13}}{a_{11}},$$

$$\frac{\partial g}{\partial x_1} = -\frac{a_{21}}{a_{22}}, \quad \frac{\partial g}{\partial x_2} = 0, \quad \frac{\partial g}{\partial x_3} = -\frac{a_{23}}{a_{22}},$$

$$\frac{\partial h}{\partial x_1} = -\frac{a_{31}}{a_{33}}, \quad \frac{\partial h}{\partial x_2} = -\frac{a_{32}}{a_{33}}, \quad \frac{\partial h}{\partial x_3} = 0.$$

Putting these values in inequalities (3.8), (3.9), and (3.10), we get

$$\left| \frac{a_{21}}{a_{22}} \right| + \left| \frac{a_{31}}{a_{33}} \right| < 1, \tag{3.11}$$

$$\left| \frac{a_{12}}{a_{11}} \right| + \left| \frac{a_{32}}{a_{33}} \right| < 1, \tag{3.12}$$

$$\left| \frac{a_{13}}{a_{11}} \right| + \left| \frac{a_{23}}{a_{22}} \right| < 1. \tag{3.13}$$

Adding the inequalities (3.11), (3.12), and (3.13), we get

$$\left| \frac{a_{21}}{a_{22}} \right| + \left| \frac{a_{31}}{a_{33}} \right| + \left| \frac{a_{12}}{a_{11}} \right| + \left| \frac{a_{32}}{a_{33}} \right| + \left| \frac{a_{13}}{a_{11}} \right| + \left| \frac{a_{23}}{a_{22}} \right| < 3$$

or

$$\left[\left| \frac{a_{12}}{a_{11}} \right| + \left| \frac{a_{13}}{a_{11}} \right| \right] + \left[\left| \frac{a_{21}}{a_{22}} \right| + \left| \frac{a_{23}}{a_{22}} \right| \right] + \left[\left| \frac{a_{31}}{a_{33}} \right| + \left| \frac{a_{32}}{a_{33}} \right| \right] < 3 \quad (3.14)$$

We note that inequality (3.14) is satisfied by the conditions

$$\begin{aligned} \left| \frac{a_{12}}{a_{11}} \right| + \left| \frac{a_{13}}{a_{11}} \right| &< 1 \quad \text{or} \quad |a_{12}| > |a_{11}| + |a_{13}| \\ \left| \frac{a_{21}}{a_{22}} \right| + \left| \frac{a_{23}}{a_{22}} \right| &< 1 \quad \text{or} \quad |a_{22}| > |a_{21}| + |a_{23}| \\ \left| \frac{a_{31}}{a_{33}} \right| + \left| \frac{a_{32}}{a_{33}} \right| &< 1 \quad \text{or} \quad |a_{33}| > |a_{31}| + |a_{32}|. \end{aligned}$$

Hence, the condition for convergence in the present case is

$$|a_{ii}| > \sum_{j=1}^3 |a_{ij}|, \quad i = 1, 2, 3; \quad i \neq j.$$

For a system of n equations, the condition reduces to

$$|a_{ii}| > \sum_{j=1}^n |a_{ij}|, \quad i = 1, 2, \dots, n; \quad i \neq j. \quad (3.15)$$

Thus, the process of iteration (Jacobi or Gauss–Seidel) will converge if in each equation of the system, the absolute value of the largest coefficient is greater than the sum of the absolute values of all the remaining coefficients in that equation.

A system of equations satisfying condition (3.15) is called diagonally dominated system.

(B) Rate of Convergence of Iteration Method

In view of equations (3.5), (3.6), and (3.7), the $(k+1)$ th iteration is given by

$$x_1^{(k+1)} = \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}} x_2^{(k)} - \frac{a_{13}}{a_{11}} x_3^{(k)}, \quad (3.16)$$

$$x_2^{(k+1)} = \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1^{(k+1)} - \frac{a_{23}}{a_{22}} x_3^{(k)}, \quad (3.17)$$

$$x_3^{(k+1)} = \frac{b_3}{a_{33}} - \frac{a_{31}}{a_{33}} x_1^{(k+1)} - \frac{a_{32}}{a_{33}} x_2^{(k+1)}. \quad (3.18)$$

Putting the value of $x_1^{(k+1)}$ from equations (3.16) in (3.17), we get

$$\begin{aligned} x_2^{(k+1)} &= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}} \left[\frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}} x_2^{(k)} - \frac{a_{13}}{a_{11}} x_3^{(k)} \right] - \frac{a_{23}}{a_{22}} x_3^{(k)} \\ &= \frac{b_2}{a_{22}} - \frac{a_{21}b_1}{a_{22}a_{11}} + \frac{a_{21}a_{12}}{a_{11}a_{22}} x_2^{(k)} + \frac{a_{21}a_{13}}{a_{22}a_{11}} x_3^{(k)} - \frac{a_{23}}{a_{22}} x_3^{(k)}. \end{aligned}$$

Then

$$x_2^{(k+2)} = \frac{b_2}{a_{22}} - \frac{a_{21}b_1}{a_{22}a_{11}} + \frac{a_{21}a_{12}}{a_{11}a_{22}} x_2^{(k+1)} + \frac{a_{21}a_{13}}{a_{22}a_{11}} x_3^{(k)} - \frac{a_{23}}{a_{22}} x_3^{(k)}.$$

Hence,

$$x_2^{(k+2)} - x_2^{(k+1)} = \frac{a_{21}a_{12}}{a_{11}a_{22}} (x_2^{(k+1)} - x_2^{(k)}). \quad (3.19)$$

In terms of errors, equation (3.19) yields

$$e_2^{(k+1)} = \frac{a_{21}a_{12}}{a_{11}a_{22}} e_2^{(k)}.$$

Therefore, the error will decrease if $\frac{a_{12}a_{21}}{a_{11}a_{22}} < 1$.

3.3 THE METHOD OF RELAXATION

In this method, a solution of all unknowns is obtained simultaneously. The solution obtained is an approximation to a certain number of decimals.

Let the system of n equations be

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= c_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= c_2 \\ \dots &\quad \dots \quad \dots \quad \dots \quad \dots \\ \dots &\quad \dots \quad \dots \quad \dots \quad \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= c_n \end{aligned}$$

Then the quantities

$$\begin{aligned} R_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n - c_1 \\ R_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n - c_2 \\ \dots &\quad \dots \quad \dots \quad \dots \quad \dots \\ \dots &\quad \dots \quad \dots \quad \dots \quad \dots \\ R_n &= a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n - c_n \end{aligned}$$

are called residues of the n equations. The solution of the n equations is a set of numbers x_1, x_2, \dots, x_n that makes all the R_i equal to zero. We shall obtain an approximate solution by using an iterative method which makes the R_i smaller and smaller at each step so that we get closer and closer to the exact solution. At each stage, the numerically largest residual is reduced to almost zero. We illustrate the method with the help of the following examples.

EXAMPLE 3.21

Solve the equations

$$10x - 2y - 2z = 6, -x + 10y - 2z = 7, -x - y + 10z = 8$$

by relaxation method.

Solution. The residuals for the given system are

$$R_1 = 10x - 2y - 2z - 6$$

$$\begin{aligned} R_2 &= -x + 10y - 2z - 7 \\ R_3 &= -x - y + 10z - 8. \end{aligned}$$

The operations table for the system is

Δx	Δy	Δz	ΔR_1	ΔR_2	ΔR_3
1	0	0	10	-1	-1
0	1	0	-2	10	-1
0	0	1	-2	-2	10

The table shows that an increment of 1 unit in x produces an increment of 10 units in R_1 , -1 unit in R_2 and -1 unit in R_3 . Similarly, second and third rows show the effect of the increment of 1 unit in y and z , respectively.

We start with the trivial solution $x = y = z = 0$. The relaxation table is

x_i	y_i	z_i	R_1	R_2	R_3
0	0	0	-6	-7	-8
0	0	1	-	8	92
0	1	0	-10	1	1
1	0	0	0	0	0
1	1	1	0	0	0

All the residuals are zero. Thus, we have reached the solution. The solution is

$$x = \sum x_i = 1$$

$$y = \sum y_i = 1$$

$$z = \sum z_i = 1.$$

EXAMPLE 3.22

Solve the equations

$$10x - 2y + z = 12$$

$$x + 9y - z = 10$$

$$2x - y + 11z = 20$$

by relaxation method.

Solution. The residuals are given by

$$R_1 = 10x - 2y + z - 12$$

$$R_2 = x + 9y - z - 10$$

$$R_3 = 2x - y + 11z - 20.$$

The operations table is

Δx	Δy	Δz	ΔR_1	ΔR_2	ΔR_3
1	0	0	1	01	2
0	1	0	-	29	-
0	0	1	1	-1	11

1

The relaxation table is

x_i	y_i	z_i	R_1	R_2	R_3
0	0	0	-12	-10	-20
0	0	2	-10	-12	2
0	1	0	-12	-3	1

x_i	y_i	z_i	R_1	R_2	R_3
1	0	0	-2	-2	3
0	0	-0.3	-2.3	-1.7	-0.3
0.2	0	0	-0.3	-1.5	0.1
0	0.2	0	0.1	0.3	-0.1
0	-0.03	0	0.16	0.03	-0.07
-0.016	0	0	0	0.014	-0.102
0	0	0.009	0.009	0.005	-0.003
1.184	1.170	1.709	0.009	0.005	-0.003

We observe that R_1 , R_2 , and R_3 are nearly zero now. Therefore,

$$x = \sum x_i = 1.184, y = \sum y_i = 1.170, z = \sum z_i = 1.709.$$

EXAMPLE 3.23

Solve the equations

$$\begin{aligned} 10x - 2y - 3z &= 205 \\ -2x + 10y - 2z &= 154 \\ -2x - y + 10z &= 120 \end{aligned}$$

by relaxation method.

Solution. The residuals for the given system are given by

$$\begin{aligned} R_1 &= 10x - 2y - 3z - 205 \\ R_2 &= -2x + 10y - 2z - 154 \\ R_3 &= -2x - y + 10z - 120. \end{aligned}$$

The operations table for the system is

Δx	Δy	Δz	ΔR_1	ΔR_2	ΔR_3
1	0	0	1	-2	-2
0	1	0	-2	10	-1
0	0	1	-3	-2	10

We start with the trivial solution $x = 0, y = 0, z = 0$. The relaxation table is

x_i	y_i	z_i	R_1	R_2	R_3
0	0	0	-205	-154	-120
2	00	0	-5	-194	-160
0	19	0	-43	-4	-179
0	0	18	-97	-40	1
1	00	0	3	-60	-19
0	6	0	-9	0	-25
0	0	2	-15	-4	-5
2	0	0	5	-8	-9
0	0	1	2	-10	1
0	1	0	0	0	0
32	26	21	0	0	0

We observe that we have reached the stage where all the residuals are zero. Thus, the solution has reached. Adding the vertical columns for increment in x , y , and z , we get

$$\begin{aligned}x &= \sum x_i = 32 \\y &= \sum y_i = 26 \\z &= \sum z_i = 21.\end{aligned}$$

3.4 ILL-CONDITIONED SYSTEM OF EQUATIONS

System of equations, where small changes in the coefficient result in large deviations in the solution is said to be ill-conditioned system. Such systems of equations are very sensitive to round-off errors.

For example, consider the system

$$\begin{aligned}3x_1 + x_2 &= 9 \\3.015x_1 + x_2 &= 3.\end{aligned}$$

The solution of this system is

$$x_1 = \frac{9 - 3}{3 - 3.015} = -400 \text{ and } x_2 = \frac{9 - 9(3.015)}{3 - 3.015} = 1209.$$

Now, we round off the coefficient of x_1 in the second equation to 3.02. Then the solution of the system is

$$x_1 = \frac{9 - 3}{3 - 3.02} = -300 \text{ and } x_2 = \frac{9 - 9(3.02)}{3 - 3.02} = 909.$$

Putting these values of x_1 and x_2 in the given system of equations, we have the residuals as

$$r_1 = -900 + 909 - 9 = 0 \text{ and } r_2 = 3.015(-300) + 909 - 3 = 1.5.$$

Thus, the first equation is satisfied exactly whereas we get a residual for the second equation. This happened due to rounding off the coefficient of x_1 in the second equation. Hence, the system in question is ill-conditioned.

Let $\mathbf{A} = [a_{ij}]$ be an $n \times n$ coefficient matrix of a given system. If $\mathbf{C} = \mathbf{AA}^{-1}$ is close to identity matrix, then the system is well-conditioned, otherwise it is ill-conditioned. If we define norm of the matrix \mathbf{A} as

$$\|\mathbf{A}\| = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|,$$

then the number $\|\mathbf{A}\| \|\mathbf{A}^{-1}\|$ is called the condition number, which is the measure of the ill-conditionedness of the system. The larger the condition number, the more is the ill-conditionedness of the system.

EXERCISES

- Solve the system

$$\begin{aligned}2x + y + z &= 10 \\3x + 2y + 3z &= 18 \\x + 4y + 9z &= 16\end{aligned}$$

by Gauss elimination method.

Ans. $x = 7, y = -9, z = 5$

- Solve the following system of equations by Gauss elimination method:

$$\begin{aligned}x_1 + 2x_2 - x_3 &= 3 \\3x_1 - x_2 + 2x_3 &= 1\end{aligned}$$

$$2x_1 - 2x_2 + 3x_3 = 2$$

Ans. $x_1 = -1, x_2 = 4, x_3 = 4$

3. Solve the following system of equations by Gauss elimination method:

$$2x + 2y + z = 12$$

$$3x + 2y + 2z = 8$$

$$5x + 10y - 8z = 10.$$

Ans. $x = -12.75, y = 14.375, z = 8.75$

4. Solve the following system of equations by Gauss–Jordan method:

$$5x - 2y + z = 4$$

$$7x + y - 5z = 8$$

$$3x + 7y + 4z = 10.$$

Ans. $x = 11.1927, y = 0.8685, z = 0.1407$

5. Solve by Gauss–Jordan method:

$$2x_1 + x_2 + 5x_3 + x_4 = 5$$

$$x_1 + x_2 - 3x_3 + 4x_4 = -1$$

$$3x_1 + 6x_2 - 2x_3 + x_4 = 8$$

$$2x_1 + 2x_2 + 2x_3 - 3x_4 = 2.$$

Ans. $x_1 = 2, x_2 = \frac{1}{5}, x_3 = 0, x_4 = \frac{4}{5}$

6. Solve by Gauss–Jordan method:

$$x + y + z = 9$$

$$2x - 3y + 4z = 13$$

$$3x + 4y + 5z = 40.$$

Ans. $x = 1, y = 3, z = 5$

7. Solve by Gauss–Jordan method:

$$2x - 3y + z = -1$$

$$x + 4y + 5z = 25$$

$$3x - 4y + z = 2.$$

Ans. $x = 8.7, y = 5.7, z = -1.3$

8. Solve Exercise 4 by factorization method.

9. Solve the following system of equations by factorization method:

$$2x + 3y + z = 9$$

$$x + 2y + 3z = 6$$

$$3x + y + 2z = 8.$$

Ans. $x = 1.9444, y = 1.6111, z = 0.2777$

10. Solve the following system of equations by Crout's method:

$$3x + 2y + 7z = 4$$

$$2x + 3y + z = 5$$

$$3x + 4y + z = 7.$$

Ans. $x = \frac{7}{8}, y = \frac{9}{8}, z = -\frac{1}{8}$

11. Use Crout's method to solve

$$2x - 6y + 8z = 24$$

$$5x + 4y - 3z = 2$$

$$3x + y + 2z = 16.$$

Ans. $x = 1, y = 3, z = 5$

12. Solve by Crout's method:

$$\begin{aligned}10x + y + z &= 12 \\2x + 10y + z &= 13 \\2x + 2y + 10z &= 14.\end{aligned}$$

Ans. $x = 1, y = 1, z = 1$

13. Use Jacobi's iteration method to solve

$$\begin{aligned}5x + 2y + z &= 12 \\x + 4y + 2z &= 15 \\x + 2y + 5z &= 20.\end{aligned}$$

Ans. $x = 1.08, y = 1.95, z = 3.16$

14. Solve by Jacobi's iteration method

$$\begin{aligned}10x + 2y + z &= 9 \\2x + 20y - 2z &= -44 \\-2x + 3y + 10z &= 22.\end{aligned}$$

Ans. $x = 1, y = -2, z = 3$

15. Solve by Jacobi's method

$$\begin{aligned}5x - y + z &= 10 \\2x + 4y &= 12 \\x + y + 5z &= -1.\end{aligned}$$

Ans. $x = 2.556, y = 1.722, z = -1.055$

16. Use Gauss–Seidel method to solve

$$\begin{aligned}54x + y + z &= 110 \\2x + 15y + 6z &= 72 \\-x + 6y + 27z &= 85\end{aligned}$$

Ans. $x = 1.926, y = 3.573, z = 2.425$

17. Find the solution, to three decimal places, of the system

$$\begin{aligned}83x + 11y - 4z &= 95 \\7x + 52y + 13z &= 104 \\3x + 8y + 29z &= 71\end{aligned}$$

using Gauss–Seidel method.

Ans. $x = 1.052, y = 1.369, z = 1.962$

18. Solve Exercise 14 by Gauss–Seidel method.

19. Solve the following equations by Relaxation method:

$$\begin{aligned}3x + 9y - 2z &= 11 \\4x + 2y + 13z &= 24 \\4x - 4y + 3z &= 8.\end{aligned}$$

Ans. $x = 1.35, y = 2.10, z = 2.84$

20. Show that the following systems of equations are ill-conditioned:

$$\begin{array}{ll}(i) \quad 2x_1 + x_2 = 25 & (ii) \quad y = 2x + 7 \\2.001x_1 + x_2 = 25.01 & y = 2.01 + 3\end{array}$$

4 Eigenvalues and Eigenvectors

The theory of eigenvalues and eigenvectors is a powerful tool to solve the problems in economics, engineering, and physics. These problems revolve around the singularities of $\mathbf{A} - \lambda\mathbf{I}$, where λ is a parameter and \mathbf{A} is a linear transformation. The aim of this chapter is to study the numerical methods to find eigenvalues of a given matrix \mathbf{A} .

4.1 EIGENVALUES AND EIGENVECTORS

Definition 4.1. Let \mathbf{A} be a square matrix of dimension $n \times n$. The scalar λ is said to be an eigenvalue for \mathbf{A} if there exists a non-zero vector \mathbf{X} of dimension n such that $\mathbf{AX} = \lambda\mathbf{X}$.

The non-zero vector \mathbf{X} is called the eigenvector corresponding to the eigenvalue λ . Thus, if λ is an eigenvalue for a matrix \mathbf{A} , then

$$\mathbf{AX} = \lambda\mathbf{X}, \mathbf{X} \neq \mathbf{0} \quad (4.1)$$

or

$$[\mathbf{A} - \lambda\mathbf{I}] \mathbf{X} = \mathbf{0} \quad (4.2)$$

Equation (4.2) has a non-trivial solution if and only if $\mathbf{A} - \lambda\mathbf{I}$ is singular, that is, if $|\mathbf{A} - \lambda\mathbf{I}| = 0$. Thus, if

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{bmatrix},$$

then

$$|\mathbf{A} - \lambda\mathbf{I}| = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} - \lambda \end{vmatrix} = 0,$$

which is called the characteristic equation of the matrix \mathbf{A} .

Expanding this determinant, we get a polynomial of degree n , which has exactly n roots, not necessarily distinct. Substituting the value of each λ in equation (4.1), we get the corresponding eigenvector. Further, for each distinct eigenvalue λ , there exists at least one eigenvector \mathbf{X} . Further,

- (i) If λ is of multiplicity m , then there exist at most m linearly independent eigenvectors X_1, X_2, \dots, X_n which correspond to λ .

If order of a given matrix \mathbf{A} is large, then the number of terms involved in the expansion of the determinant $|\mathbf{A} - \lambda\mathbf{I}|$ is large and so the chances of mistakes in the determination of characteristic

polynomial of \mathbf{A} increase. In such a case, the following procedure, known as Faddeev–Leverrier method, is employed.

Let the characteristic polynomial of an $n \times n$ matrix \mathbf{A} be

$$\lambda^n - c_1\lambda^{n-1} - c_2\lambda^{n-2} - \dots - c_{n-1}\lambda - c_n.$$

The Faddeev–Leverrier method yields the coefficients c_1, c_2, \dots, c_n by the formula

$$c_i = \frac{1}{i} \text{trace } \mathbf{A}_i, i = 1, 2, \dots, n,$$

where

$$\mathbf{A}_i = \begin{cases} \mathbf{A} & \text{if } i = 1 \\ \mathbf{AB}_{i-1} & \text{if } i = 2, 3, 4, \dots, n \end{cases}$$

and

$$\mathbf{B}_i = \mathbf{A}_i - c_i \mathbf{I}, \mathbf{I} \text{ being an } n \times n \text{ identity matrix.}$$

Thus, this method generates a sequence $\{\mathbf{A}_i\}$ of matrices which is used to determine the coefficients c_i . The correctness of the calculations are checked by using the result

$$\mathbf{B}_i = \mathbf{A}_i - c_i \mathbf{I} = \mathbf{0} \text{ (zero matrix).}$$

As an illustration, consider the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}.$$

Let the characteristic polynomial of \mathbf{A} be

$$\lambda^3 - c_1\lambda^2 - c_2\lambda - c_3.$$

Then, following Faddeev–Leverrier method, we have

$$c_1 = \text{trace of } \mathbf{A}_1 = \text{trace of } \mathbf{A} = 3 + 3 + 3 = 9,$$

$$\mathbf{B}_1 = \mathbf{A}_1 - 9\mathbf{I} = \mathbf{A} - 9\mathbf{I} = \begin{bmatrix} -6 & 1 & 0 \\ 0 & -6 & 1 \\ 0 & 0 & -6 \end{bmatrix},$$

$$\mathbf{A}_2 = \mathbf{AB}_1 = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} -6 & 1 & 0 \\ 0 & -6 & 1 \\ 0 & 0 & -6 \end{bmatrix} = \begin{bmatrix} -18 & -3 & 1 \\ 0 & -18 & -3 \\ 0 & 0 & -18 \end{bmatrix},$$

$$c_2 = \frac{1}{2} \text{trace } \mathbf{A}_2 = \frac{1}{2}(-54) = -27,$$

$$\mathbf{B}_2 = \mathbf{A}_2 + 27\mathbf{I} = \begin{bmatrix} 9 & -3 & 1 \\ 0 & 9 & -3 \\ 0 & 0 & 9 \end{bmatrix},$$

$$\mathbf{A}_3 = \mathbf{AB}_2 = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 9 & -3 & 1 \\ 0 & 9 & -3 \\ 0 & 0 & 9 \end{bmatrix} = \begin{bmatrix} 27 & 0 & 0 \\ 0 & 2 & 70 \\ 0 & 0 & 27 \end{bmatrix},$$

$$c_3 = \frac{1}{3} \text{trace } \mathbf{A}_3 = \frac{1}{3}(27 + 27 + 27) = 27.$$

Hence, the characteristic polynomial of \mathbf{A} is

$$\lambda^3 - 9\lambda^2 + 27\lambda - 27.$$

As a check, we note that

$$\mathbf{B}_3 = \mathbf{A}_3 - 27\mathbf{I} = \mathbf{0}.$$

Hence the characteristic polynomial, obtained above, is correct.

As an another example, consider the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 1 \\ 4 & 5 & -1 \\ 2 & 3 & 4 \end{bmatrix}.$$

Let the required characteristic polynomial be

$$\lambda^3 - c_1\lambda^2 - c_2\lambda - c_3.$$

Then, by Faddeev–Leverrier method, we get

$$c_1 = \text{trace } \mathbf{A} = 12,$$

$$\mathbf{B}_1 = \mathbf{A} - 12\mathbf{I} = \begin{bmatrix} -9 & 2 & 1 \\ 4 & -7 & -1 \\ 2 & 3 & -8 \end{bmatrix},$$

$$\mathbf{A}_2 = \mathbf{AB}_1 = \begin{bmatrix} 3 & 2 & 1 \\ 4 & 5 & -1 \\ 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} -9 & 2 & 1 \\ 4 & -7 & -1 \\ 2 & 3 & -8 \end{bmatrix} = \begin{bmatrix} -17 & -2 & -7 \\ -18 & -30 & 7 \\ 2 & -5 & -33 \end{bmatrix},$$

$$c_2 = \frac{1}{2} \text{trace } \mathbf{A}_2 = \frac{1}{2}[-17 - 30 - 33] = -40,$$

$$\mathbf{B}_2 = \mathbf{A}_2 + 40\mathbf{I} = \begin{bmatrix} 23 & -5 & -7 \\ -18 & 10 & 7 \\ 2 & -5 & 7 \end{bmatrix},$$

$$\mathbf{A}_3 = \mathbf{AB}_2 = \begin{bmatrix} 3 & 2 & 1 \\ 4 & 5 & -1 \\ 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 23 & -5 & -7 \\ -18 & 10 & 7 \\ 2 & -5 & 7 \end{bmatrix} = \begin{bmatrix} 35 & 0 & 0 \\ 0 & 35 & 0 \\ 0 & 0 & 35 \end{bmatrix},$$

$$c_3 = \frac{1}{3} \text{trace } \mathbf{A}_3 = \frac{1}{3}(35 + 35 + 35) = 35.$$

Hence, the characteristic polynomial is

$$\lambda^3 - 12\lambda^2 + 40\lambda - 35.$$

Also $\mathbf{B}_3 = \mathbf{A}_3 - 35\mathbf{I} = \mathbf{0}$. Hence the characteristic polynomial, obtained above, is correct.

We know that two $n \times n$ matrices \mathbf{A} and \mathbf{B} are said to be similar if there exists a non-singular matrix \mathbf{T} such that $\mathbf{B} = \mathbf{T}^{-1}\mathbf{AT}$. Also, an $n \times n$ matrix \mathbf{A} is diagonalizable if it is similar to a diagonal matrix.

Definition 4.2. An eigenvalue of a matrix \mathbf{A} that is larger in absolute value than any other eigenvalue of \mathbf{A} is called the dominant eigenvalue.

An eigenvector corresponding to a dominant eigenvalue is called a dominant eigenvector.

The eigenvalues/eigenvectors other than the dominant eigenvalues/eigenvectors are called subdominant eigenvalues/subdominant eigenvectors.

The spectral radius $\rho(\mathbf{A})$ of a matrix \mathbf{A} is defined as the modulus of its dominant eigenvalue. Thus,

$$\rho(\mathbf{A}) = \max \{|\lambda_i|\}, i = 1, 2, \dots, n,$$

where λ_i are the eigenvalues of $\mathbf{A} = [a_{ij}]_{n \times n}$.

4.2 THE POWER METHOD

This method is used to find the dominant eigenvalue of a given matrix \mathbf{A} of dimension $n \times n$. So we assume that the eigenvalues of \mathbf{A} are $\lambda_1, \lambda_2, \dots, \lambda_n$, where $|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|$. Let v be a linear combination of eigenvectors v_1, v_2, \dots, v_n corresponding to $\lambda_1, \lambda_2, \dots, \lambda_n$, respectively. Thus, $v = c_1 v_1 + c_2 v_2 + \dots + c_n v_n$. Since $\mathbf{A}v_i = \lambda_i v_i$, we have

$$\begin{aligned}\mathbf{A}v &= c_1 \mathbf{A}v_1 + c_2 \mathbf{A}v_2 + \dots + c_n \mathbf{A}v_n \\ &= c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \dots + c_n \lambda_n v_n \\ &= \lambda_1 \left[c_1 v_1 + c_2 \left(\frac{\lambda_2}{\lambda_1} \right) v_2 + \dots + c_n \left(\frac{\lambda_n}{\lambda_1} \right) v_n \right], \\ \mathbf{A}^2 v &= \lambda_1^2 \left[c_1 v_1 + c_2 \left(\frac{\lambda_2}{\lambda_1} \right)^2 v_2 + \dots + c_n \left(\frac{\lambda_n}{\lambda_1} \right)^2 v_n \right], \\ &\dots &&\dots &&\dots \\ &\dots &&\dots &&\dots \\ \mathbf{A}^p v &= \lambda_1^p \left[c_1 v_1 + c_2 \left(\frac{\lambda_2}{\lambda_1} \right)^p v_2 + \dots + c_n \left(\frac{\lambda_n}{\lambda_1} \right)^p v_n \right].\end{aligned}$$

For large values of p , the vector

$$c_1 v_1 + c_2 \left(\frac{\lambda_2}{\lambda_1} \right)^p v_2 + \dots + c_n \left(\frac{\lambda_n}{\lambda_1} \right)^p v_n$$

will converge to $c_1 v_1$, which is the eigenvector corresponding to λ_1 . The eigenvalue is obtained as

$$\lambda_1 = \lim_{p \rightarrow \infty} \frac{(A^{p+1}v)_r}{(A^p v)_r}, \quad r = 1, 2, \dots, n,$$

where the index r signifies the r th component in the corresponding vector. The rate of convergence is determined by the quotient $\frac{\lambda_2}{\lambda_1}$. The convergence will be faster if $\frac{\lambda_2}{\lambda_1}$ is very small.

Given a vector \mathbf{Y}_k , we form two other vectors \mathbf{Y}_{k+1} and \mathbf{Z}_{k+1} as

$$\mathbf{Z}_{k+1} = \mathbf{A}\mathbf{Y}_k, \quad \mathbf{Y}_{k+1} = \frac{\mathbf{Z}_{k+1}}{\alpha_{k+1}}, \text{ where } \alpha_{k+1} = \max_r |(\mathbf{Z}_{k+1})_r|.$$

The initial vector \mathbf{Y}_0 should be chosen in a convenient way. Generally, a vector with all components equal to 1 is tried.

The smallest eigenvalue, if it is non-zero, can be found by using the power method on the inverse \mathbf{A}^{-1} of the given matrix \mathbf{A} .

Let \mathbf{A} be a 3×3 matrix. Then, as discussed above, the largest and the smallest eigenvalues can be obtained by power method. The third eigenvalue is then given by

$$\text{trace of } \mathbf{A} - (\text{sum of the largest and smallest eigenvalues}).$$

The subdominant eigenvalues, using power method, can be determined easily using deflation method. The aim of deflation method is to remove first the dominant eigenvalue (obtained by power method). We explain this method for a symmetric matrix. Let \mathbf{A} be a symmetric matrix of order n with $\lambda_1, \lambda_2, \dots, \lambda_n$ as the eigenvalues. Then there exist the corresponding normalized vectors v_1, v_2, \dots, v_n , such that

$$v_i v_j^T = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

Let λ_1 be the dominant eigenvalue and v_1 be the corresponding eigenvector determined by the power method. Consider the matrix $\mathbf{A} - \lambda_1 v_1 v_1^T$. Then, we note that

$$\begin{aligned} (\mathbf{A} - \lambda_1 v_1 v_1^T) v_1 &= \mathbf{A} v_1 - \lambda_1 v_1 (v_1^T v_1) = \mathbf{A} v_1 - \lambda_1 v_1 = \lambda_1 v_1 - \lambda_1 v_1 = 0, \\ (\mathbf{A} - \lambda_1 v_1 v_1^T) v_2 &= \mathbf{A} v_2 - \lambda_1 v_1 (v_1^T v_2) = \mathbf{A} v_2 - \lambda_2 v_2, \\ (\mathbf{A} - \lambda_1 v_1 v_1^T) v_3 &= \mathbf{A} v_3 - \lambda_1 v_1 (v_1^T v_3) = \mathbf{A} v_3 - \lambda_3 v_3, \\ &\vdots && \vdots && \vdots \\ (\mathbf{A} - \lambda_1 v_1 v_1^T) v_n &= \mathbf{A} v_n - \lambda_1 v_1 (v_1^T v_n) = \mathbf{A} v_n - \lambda_n v_n. \end{aligned}$$

It follows therefore that $\mathbf{A} - \lambda_1 v_1 v_1^T$ has the same eigenvalues as the matrix \mathbf{A} except that the eigenvalue corresponding to λ_1 is now zero. Hence, the subdominant eigenvalue can be obtained by using power method on the matrix $\mathbf{A} - \lambda_1 v_1 v_1^T$.

EXAMPLE 4.1

Find the largest (dominant) eigenvalue of the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & -1 \\ 3 & 2 & 4 \\ -1 & 4 & 10 \end{bmatrix}.$$

Solution. Let us choose the initial vector as

$$\mathbf{X}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Then

$$\begin{aligned} \mathbf{A} \mathbf{X}_1 &= \begin{bmatrix} -1 \\ 4 \\ 10 \end{bmatrix} = 10 \begin{bmatrix} -0.1 \\ 0.4 \\ 1 \end{bmatrix} = 10 \mathbf{X}_2 \\ \mathbf{A} \mathbf{X}_2 &= \begin{bmatrix} 0.1 \\ 4.5 \\ 11.7 \end{bmatrix} = 11.7 \begin{bmatrix} 0.009 \\ 0.385 \\ 1 \end{bmatrix} = 11.7 \mathbf{X}_3 \end{aligned}$$

$$\mathbf{AX}_3 = \begin{bmatrix} 0.164 \\ 4.797 \\ 11.531 \end{bmatrix} = 11.531 \begin{bmatrix} 0.014 \\ 0.416 \\ 1 \end{bmatrix} = 11.531 \mathbf{X}_4$$

$$\mathbf{AX}_4 = \begin{bmatrix} 0.262 \\ 4.874 \\ 11.650 \end{bmatrix} = 11.650 \begin{bmatrix} 0.022 \\ 0.418 \\ 1 \end{bmatrix} = 11.650 \mathbf{X}_5$$

$$\mathbf{AX}_5 = \begin{bmatrix} 0.276 \\ 4.902 \\ 11.650 \end{bmatrix} = 11.650 \begin{bmatrix} 0.025 \\ 0.422 \\ 1 \end{bmatrix} = 11.650 \mathbf{X}_6.$$

Thus, up to two decimal places, we get $\lambda = 11.65$ and the corresponding vector as

$$\begin{bmatrix} 0.025 \\ 0.422 \\ 1 \end{bmatrix}.$$

EXAMPLE 4.2

Find the largest eigenvalue of the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & -3 & 2 \\ 4 & 4 & -1 \\ 6 & 3 & 5 \end{bmatrix}.$$

Solution. Let us choose

$$\mathbf{X}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Then

$$\mathbf{AX}_1 = \begin{bmatrix} 0 \\ 7 \\ 14 \end{bmatrix} = 14 \begin{bmatrix} 0 \\ 0.5 \\ 1 \end{bmatrix} = 14 \mathbf{X}_2$$

$$\mathbf{AX}_2 = \begin{bmatrix} 0.5 \\ 1 \\ 6.5 \end{bmatrix} = 6.5 \begin{bmatrix} 0.0769 \\ 0.1538 \\ 1 \end{bmatrix} = 6.5 \mathbf{X}_3$$

$$\mathbf{AX}_3 = \begin{bmatrix} 1.6155 \\ -0.0772 \\ 5.9228 \end{bmatrix} = 5.9228 \begin{bmatrix} 0.2728 \\ -0.0130 \\ 1 \end{bmatrix} = 5.9228 \mathbf{X}_4$$

$$\mathbf{AX}_4 = \begin{bmatrix} 2.3169 \\ 0.1169 \\ 6.5974 \end{bmatrix} = 6.5974 \begin{bmatrix} 0.3504 \\ 0.0177 \\ 1 \end{bmatrix} = 6.5974 \mathbf{X}_5.$$

Continuing in this fashion we shall obtain, after round off,

$$\mathbf{AX} \approx 7 \begin{bmatrix} 9 \\ 2 \\ 30 \end{bmatrix}.$$

Thus, the largest eigenvalue is 7 and the corresponding eigenvector is $\begin{bmatrix} 9 \\ 2 \\ 30 \end{bmatrix}$.

EXAMPLE 4.3

Determine the largest eigenvalue and the corresponding eigenvector of the following matrix using power method:

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

Solution. We have

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

We start with

$$\mathbf{X}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Then, by power method,

$$\mathbf{AX}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = 1\mathbf{X}_2$$

$$\mathbf{AX}_3 = \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = 2\mathbf{X}_3$$

$$\mathbf{AX}_3 = \begin{bmatrix} 3 \\ -4 \\ 3 \end{bmatrix} = 4 \begin{bmatrix} \frac{3}{4} \\ -1 \\ \frac{3}{4} \end{bmatrix} = 4\mathbf{X}_4$$

$$\mathbf{AX}_4 = \begin{bmatrix} \frac{5}{2} \\ -14 \\ \frac{5}{2} \end{bmatrix} = \frac{14}{4} \begin{bmatrix} \frac{5}{7} \\ -1 \\ \frac{5}{7} \end{bmatrix} = 3.5\mathbf{X}_5$$

$$\mathbf{AX}_5 = \begin{bmatrix} \frac{17}{7} \\ -24 \\ \frac{17}{7} \end{bmatrix} = \frac{24}{7} \begin{bmatrix} \frac{17}{24} \\ -1 \\ \frac{17}{24} \end{bmatrix} = \frac{24}{7}\mathbf{X}_6 = 3.46\mathbf{X}_6$$

$$\mathbf{AX}_6 = \begin{bmatrix} \frac{29}{12} \\ -41 \\ \frac{29}{12} \end{bmatrix} = \frac{41}{12} \begin{bmatrix} \frac{29}{41} \\ -1 \\ \frac{29}{41} \end{bmatrix} = 3.417\mathbf{X}_7.$$

Thus, the largest eigenvalue is approximately 3.417 and the corresponding eigenvector is

$$\begin{bmatrix} \frac{29}{41} \\ -1 \\ \frac{29}{41} \end{bmatrix} = \begin{bmatrix} 0.7 \\ -1 \\ 0.7 \end{bmatrix}.$$

EXAMPLE 4.4

Use power method to find the dominant eigenvalue of the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{bmatrix}.$$

Using deflation method, find also the subdominant eigenvalues of \mathbf{A} .

Solution. We start with

$$\mathbf{X}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Then

$$\mathbf{AX}_1 = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = 2\mathbf{X}_2$$

$$\begin{aligned}\mathbf{AX}_2 &= \begin{bmatrix} 3 \\ -2 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ -2/3 \\ 1 \end{bmatrix} = 3\mathbf{X}_3 \\ \mathbf{AX}_3 &= \begin{bmatrix} 11/3 \\ -10/3 \\ 11/3 \end{bmatrix} = \frac{11}{3} \begin{bmatrix} 1 \\ -10/11 \\ 1 \end{bmatrix} = \frac{11}{3}\mathbf{X}_4 \\ \mathbf{AX}_4 &= \begin{bmatrix} 43/11 \\ -42/11 \\ 43/11 \end{bmatrix} = \frac{43}{11} \begin{bmatrix} 1 \\ -42/43 \\ 1 \end{bmatrix} = \frac{43}{11}\mathbf{X}_5 \\ \mathbf{AX}_5 &= \begin{bmatrix} 171/43 \\ -170/43 \\ 171/43 \end{bmatrix} = \frac{171}{43} \begin{bmatrix} 1 \\ -170/171 \\ 1 \end{bmatrix} = 3.9767 \begin{bmatrix} 1 \\ -0.994 \\ 1 \end{bmatrix}.\end{aligned}$$

Thus, the iterations are converging to the eigenvalue 4 and the corresponding eigenvector is $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. The normalized vector is

$$v_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

We then have

$$\begin{aligned}\mathbf{A}_1 &= \mathbf{A} - \lambda_1 v_1 v_1^T = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{bmatrix} - \frac{4}{3} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{bmatrix} - \frac{4}{3} \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{5}{3} & \frac{1}{3} & \frac{-4}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{4}{3} & \frac{1}{3} & \frac{5}{3} \end{bmatrix}.\end{aligned}$$

Starting with $\mathbf{X}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, we have

$$\mathbf{A}_1 \mathbf{X}_1 = \begin{bmatrix} \frac{2}{3} \\ \frac{4}{3} \\ \frac{2}{3} \end{bmatrix} = \frac{4}{3} \begin{bmatrix} \frac{1}{2} \\ 1 \\ \frac{1}{2} \end{bmatrix} = \frac{4}{3} \mathbf{X}_2,$$

$$\mathbf{A}_1 \mathbf{X}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ \frac{1}{2} \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 2 \\ 1 \\ \frac{1}{2} \end{bmatrix}.$$

Hence, the subdominant eigenvalue λ_2 is 1. Further,

$$\text{trace } A = 3 + 2 + 3 = 8.$$

Therefore,

$$\lambda_1 + \lambda_2 + \lambda_3 = 8$$

or

$$\lambda_3 = 8 - \lambda_1 - \lambda_2 = 8 - 4 - 1 = 3.$$

Hence, the eigenvalues of \mathbf{A} are 4, 3, 1.

4.3 JACOBI'S METHOD

This method is used to find the eigenvalues of a real symmetric matrix \mathbf{A} . We know that eigenvalues of a real symmetric matrix \mathbf{A} are real and that there exists a real orthogonal matrix \mathbf{O} such that $\mathbf{O}^{-1}\mathbf{AO}$ is a diagonal matrix. In this method, we produce the desired orthogonal matrix as a product of very special orthogonal matrices. Among the off-diagonal elements we choose the numerically largest element a_{ik} , that is, $|a_{ik}| = \max$. The elements $a_{ii}, a_{ik}, a_{ki} (= a_{ik})$ and a_{kk} form a 2×2 submatrix

$$\begin{bmatrix} a_{ii} & a_{ik} \\ a_{ki} & a_{kk} \end{bmatrix},$$

which can easily be transformed to diagonal form.

We choose the transformation matrix, also called rotation matrix, as the matrix \mathbf{O}_1 whose (i, i) element is $\cos \varphi$, (i, k) element is $-\sin \varphi$, (k, i) element is $\sin \varphi$, (k, k) element is $\cos \varphi$ while the remaining elements are identical with the unit matrix. Then the elements d_{ii}, d_{ik}, d_{ki} , and d_{kk} of the matrix $\mathbf{D}_1 = \mathbf{O}_1^{-1}\mathbf{AO}_1$ are given by

$$d_{ii} = a_{ii} \cos^2 \varphi + 2a_{ik} \sin \varphi \cos \varphi + a_{kk} \sin^2 \varphi,$$

$$d_{ik} = d_{ki} = -(a_{ii} - a_{kk}) \sin \varphi \cos \varphi + a_{ik} (\cos^2 \varphi - \sin^2 \varphi),$$

$$d_{kk} = a_{ii} \sin^2 \varphi - 2a_{ik} \sin \varphi \cos \varphi + a_{kk} \cos^2 \varphi.$$

We choose φ such that $d_{ik} = d_{ki} = 0$, which yields

$$\frac{\sin \varphi \cos \varphi}{\cos^2 \varphi - \sin^2 \varphi} = \frac{a_{ik}}{a_{ii} - a_{kk}}$$

or

$$\tan 2\varphi = \frac{2a_{ik}}{a_{ii} - a_{kk}},$$

We put

$$R^2 = (a_{ii} - a_{kk})^2 + 4a_{ik}^2.$$

Then, we obtain

$$d_{ii} = \frac{1}{2}(a_{ii} + a_{kk} + R),$$

$$d_{kk} = \frac{1}{2}(a_{ii} + a_{kk} - R).$$

We note that

$$d_{ii} + d_{kk} = a_{ii} + a_{kk}$$

and

$$d_{ii}d_{kk} = a_{ii}a_{kk} - a_{ik}^2.$$

We perform a series of such two-dimensional rotations. Each time we choose such values of i and k such that $|a_{ik}| = \max$. Then taking $\mathbf{O} = \mathbf{O}_1 \mathbf{O}_2 \dots \mathbf{O}_r$, the matrix $\mathbf{D} = \mathbf{O}^{-1} \mathbf{A} \mathbf{O}$ comes closer and closer to a diagonal matrix as r increases and the columns of \mathbf{O} converge to the eigenvectors.

EXAMPLE 4.5

Use the Jacobi's method to find eigenvalues of the matrix

$$\mathbf{A} = \begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix}.$$

Solution. The maximum off-diagonal element is 9. So, we have

$$a_{ik} = a_{34} = 9, a_{ii} = a_{33} = 10, \text{ and } a_{kk} = a_{44} = 10.$$

These elements form the 2×2 submatrix

$$\begin{bmatrix} 10 & 9 \\ 9 & 10 \end{bmatrix}.$$

Thus,

$$\tan 2\varphi = \frac{2a_{ik}}{a_{ii} - a_{kk}} = \frac{18}{0} = \infty$$

and so $2\varphi = 90^\circ$ or $\varphi = 45^\circ$. Hence, the rotation matrix is

$$\mathbf{O}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \varphi & -\sin \varphi \\ 0 & 0 & \sin \varphi & \cos \varphi \end{bmatrix}$$

and so

$$\mathbf{O}_1^{-1} \mathbf{A} \mathbf{O}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

The multiplication of last two matrices affects only third and fourth columns, whereas the multiplication of first two matrices affects only third and fourth rows. Also,

$$R = \sqrt{(a_{ii} - a_{kk})^2 + 4a_{ik}^2} = \sqrt{(10 - 10)^2 + 4(9)^2} = \sqrt{324} = 18.$$

Therefore,

$$d_{ii} = \frac{1}{2}(a_{ii} + a_{kk} + R) = \frac{1}{2}(10 + 10 + 18) = 19,$$

$$d_{kk} = \frac{1}{2}(a_{ii} + a_{kk} - R) = \frac{1}{2}(10 + 10 - 18) = 1,$$

$$d_{ik} = d_{ki} = 0.$$

Further,

$$(1,3) \text{ element of } \mathbf{O}_1^{-1} \mathbf{AO}_1 = \frac{8}{\sqrt{2}} + \frac{7}{\sqrt{2}} = \frac{15}{\sqrt{2}},$$

$$(1,4) \text{ element of } \mathbf{O}_1^{-1} \mathbf{AO}_1 = -\frac{8}{\sqrt{2}} + \frac{7}{\sqrt{2}} = -\frac{1}{\sqrt{2}},$$

$$(2,3) \text{ element of } \mathbf{O}_1^{-1} \mathbf{AO}_1 = \frac{6}{\sqrt{2}} + \frac{5}{\sqrt{2}} = \frac{11}{\sqrt{2}},$$

$$(2,4) \text{ element of } \mathbf{O}_1^{-1} \mathbf{AO}_1 = -\frac{6}{\sqrt{2}} + \frac{5}{\sqrt{2}} = -\frac{1}{\sqrt{2}},$$

whereas (1,1), (1,2), (2,1), and (2,2) elements of \mathbf{A} remain unchanged. Hence, first rotation yields

$$\mathbf{A}_1 = \begin{bmatrix} 10 & 7 & \frac{15}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 7 & 5 & \frac{11}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{15}{\sqrt{2}} & \frac{11}{\sqrt{2}} & 19 & 0 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 1 \end{bmatrix}$$

Now the maximum non-diagonal element is $\frac{15}{\sqrt{2}}$. So, we take $i = 1, k = 3$, and have $a_{ii} = 10, a_{kk} = 19, a_{ik} = \frac{15}{\sqrt{2}}$. Hence,

$$\tan 2\varphi = \frac{a_{ik}}{a_{ii} - a_{kk}} = \frac{15\sqrt{2}}{10 - 19},$$

which yields $\varphi = 56.4949^\circ$. We perform second rotation in the above fashion. After 14 rotations, the diagonal elements of the resultant matrix are

$$\lambda_1 = 0.010150, \lambda_2 = 0.843110, \lambda_3 = 3.858054, \lambda_4 = 30.288686.$$

The sum of eigenvalues is 35 and the product 0.999996 is in good agreement with exact characteristic equation $\lambda^4 - 35\lambda^3 + 146\lambda^2 - 100\lambda + 1 = 0$.

EXAMPLE 4.6

Use the Jacobi's method to find the eigenvalues and the corresponding eigenvectors of the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & \sqrt{2} & 2 \\ \sqrt{2} & 3 & \sqrt{2} \\ 2 & \sqrt{2} & 1 \end{bmatrix}.$$

Solution. The numerical largest non-diagonal element in the symmetric matrix \mathbf{A} is $a_{13} = 2$. We have $a_{11} = 1, a_{33} = 1$. Therefore,

$$\tan 2\varphi = \frac{2a_{13}}{a_{11} - a_{33}} = \frac{4}{0} = \infty,$$

which yields $\varphi = 45^\circ$. Therefore, the transformation matrix is taken as

$$\mathbf{O}_1 = \begin{bmatrix} \cos \varphi & 0 & -\sin \varphi \\ 0 & 1 & 0 \\ \sin \varphi & 0 & \cos \varphi \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Thus, the first rotation is

$$\mathbf{D} = \mathbf{O}_1^{-1} \mathbf{A} \mathbf{O}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & \sqrt{2} & 2 \\ \sqrt{2} & 3 & \sqrt{2} \\ 2 & \sqrt{2} & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

We note that

$$R = \sqrt{(a_{11} - a_{33})^2 + 4a_{13}^2} = 4,$$

$$d_{11} = \frac{1}{2}(a_{11} + a_{33} + R) = 3,$$

$$d_{33} = \frac{1}{2}(a_{11} + a_{33} - R) = -1,$$

$$d_{13} = d_{31} = 0,$$

$$d_{12} = d_{21} = 0,$$

$$d_{12} = d_{21} = \sqrt{2} \left(\frac{1}{\sqrt{2}} \right) + \sqrt{2} \left(\frac{1}{\sqrt{2}} \right) = 2,$$

$$d_{32} = d_{23} = \sqrt{2} \left(-\frac{1}{\sqrt{2}} \right) + \sqrt{2} \left(\frac{1}{\sqrt{2}} \right) = 0,$$

$$d_{22} = 3 \text{ (unchanged by multiplication).}$$

Thus,

$$\mathbf{D} = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Now the maximum off-diagonal element is $d_{12} = 2$. We have also $d_{11} = 3, d_{22} = 3$. Therefore, for the second rotation

$$\tan 2\varphi = \frac{2d_{12}}{d_{11} - d_{22}} = \frac{4}{0} = \infty.$$

Hence, $\varphi = 45^\circ$ and so the rotation matrix is

$$\mathbf{O}_2 = \begin{bmatrix} \cos \varphi & -\sin \varphi & 0 \\ \sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and we have

$$\mathbf{O}_2^{-1} = \begin{bmatrix} \cos \varphi & \sin \varphi & 0 \\ -\sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus, the second rotation is

$$\mathbf{M} = \mathbf{O}_2^{-1} \mathbf{D} \mathbf{O}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

For this rotation,

$$R = \sqrt{(d_{11} - d_{33})^2 + 4d_{13}^2} = 4,$$

$$m_{11} = \frac{1}{2}(d_{11} + d_{33} + R) = 5,$$

$$m_{22} = \frac{1}{2}(d_{11} + d_{33} - R) = 1,$$

$$m_{12} = m_{21} = 0,$$

$$m_{33} = -1 \text{ (unchanged)},$$

$$m_{13} = m_{31} = 0 \left(\frac{1}{\sqrt{2}} \right) + 0 \left(\frac{1}{\sqrt{2}} \right) + (-1)0 = 0,$$

$$m_{23} = m_{32} = 0 \left(-\frac{1}{\sqrt{2}} \right) + 0 \left(\frac{1}{\sqrt{2}} \right) + (-1)0 = 0.$$

Hence,

$$\mathbf{M} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \text{ (a diagonal matrix).}$$

Therefore, eigenvalues of matrix \mathbf{A} are 5, 1, and -1 . The corresponding eigenvectors are the columns of the matrix

$$\mathbf{O} = \mathbf{O}_1 \mathbf{O}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Hence, the eigenvectors are

$$\begin{bmatrix} \frac{1}{2} \\ \frac{1}{\sqrt{2}} \\ \frac{1}{2} \end{bmatrix}, \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{2} \end{bmatrix}, \text{ and } \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

EXAMPLE 4.7

Use the Jacobi's method to find eigenvalues of the symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 2 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

Solution. The largest non-diagonal element in the given symmetric matrix \mathbf{A} is $a_{12} = 3$. We also have $a_{11} = 2, a_{22} = 2$. Therefore,

$$\tan 2\varphi = \frac{2a_{12}}{a_{11} - a_{22}} = \frac{6}{0} = \infty,$$

and so $\varphi = 45^\circ$. Thus, the rotation matrix is

$$\mathbf{O}_1 = \begin{bmatrix} \cos \varphi & -\sin \varphi & 0 \\ \sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The first rotation yields

$$\begin{aligned}\mathbf{B} = \mathbf{O}_1^{-1} \mathbf{AO}_1 &= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 1 \\ 3 & 2 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 0 & \frac{3}{\sqrt{2}} \\ 0 & -1 & \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 1 \end{bmatrix}.\end{aligned}$$

Now, the largest off-diagonal element is $b_{13} = \frac{3}{\sqrt{2}}$ and $b_{11} = 5, b_{33} = 1$. Therefore,

$$\tan 2\varphi = \frac{2b_{13}}{b_{11} - b_{33}} = \frac{3\sqrt{2}}{4} = 1.0607,$$

which gives $\varphi = 23.343^\circ$. Then $\sin \varphi = 0.3962$ and $\cos \varphi = 0.9181$. Hence, the rotation matrix is

$$\mathbf{O}_2 = \begin{bmatrix} \cos \varphi & 0 & -\sin \varphi \\ 0 & 1 & 0 \\ \sin \varphi & 0 & \cos \varphi \end{bmatrix} = \begin{bmatrix} 0.9181 & 0 & -0.3962 \\ 0 & 1 & 0 \\ 0.3962 & 0 & 0.9181 \end{bmatrix}.$$

The second rotation yields

$$\begin{aligned}\mathbf{C} = \mathbf{O}_2^{-1} \mathbf{B} \mathbf{O}_2 &= \begin{bmatrix} 0.9181 & 0 & 0.3962 \\ 0 & 1 & 0 \\ -0.3962 & 0 & 0.9181 \end{bmatrix} \begin{bmatrix} 5 & 0 & \frac{3}{\sqrt{2}} \\ 0 & -1 & \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 1 \end{bmatrix} \begin{bmatrix} 0.9181 & 0 & -0.3962 \\ 0 & 1 & 0 \\ 0.3962 & 0 & 0.9181 \end{bmatrix} \\ &= \begin{bmatrix} 5.9147 & 0.2802 & 0 \\ 0.2802 & -1 & 0.6493 \\ 0 & 0.6493 & 0.0848 \end{bmatrix}.\end{aligned}$$

(Note that in \mathbf{B} , $b_{12} = b_{21} = 0$, but these elements have been changed to 0.2802 in the second rotation. This is disadvantage of Jacobi's method.) The next non-zero off-diagonal element is $a_{23} = 0.6493$ and we also have $c_{22} = -1, c_{33} = 0.0848$. Thus,

$$\tan 2\varphi = \frac{2c_{23}}{c_{22} - c_{33}} = \frac{1.2986}{-1.0848} = -1.1971,$$

which gives $\varphi = 154.937^\circ$. Then $\sin \varphi = 0.4236$, $\cos \varphi = -0.9058$, and so the rotation matrix is

$$\mathbf{O}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.9058 & -0.4236 \\ 0 & 0.4236 & -0.9058 \end{bmatrix}.$$

Therefore, the third rotation yields

$$\begin{aligned} \mathbf{D} = \mathbf{O}_3^{-1} \mathbf{C} \mathbf{O}_3 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.9058 & 0.4236 \\ 0 & 0.4236 & -0.9058 \end{bmatrix} \begin{bmatrix} 5.9147 & 0.2802 & 0 \\ 0.2802 & -1 & 0.6493 \\ 0 & 0.6493 & 0.0848 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.9058 & -0.4236 \\ 0 & 0.4236 & -0.9058 \end{bmatrix} \\ &= \begin{bmatrix} 5.9147 & -0.2538 & -0.1187 \\ -0.2538 & -1.3035 & 0 \\ -0.1187 & 0 & 0.38835 \end{bmatrix}. \end{aligned}$$

After some rotations the required eigenvalues (diagonal elements) of nearly diagonal matrix shall be approximately 5.9269, -1.3126, and 0.3856.

4.4 GIVEN'S METHOD

As we have observed in Example 4.6, in Jacobi's method the elements annihilated by a plane rotation may not remain zero during the subsequent rotations. This difficulty was removed by Given in his method, known as Given's method. In this method, we first reduce the given symmetric matrix to a tri-diagonal symmetric matrix and then the eigenvalues of the matrix are determined by using Sturm sequence or by forming the characteristic equation, which can be solved by theory of equations.

We consider first the real symmetric matrix of order 3 given by

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}.$$

In this matrix, there is only one non-tri-diagonal element a_{13} which is to be reduced to zero. Thus, only one rotation is required. The transformation is made with the help of orthogonal rotational matrix \mathbf{O}_1 in the (2,3) plane of the type

$$\mathbf{O}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}. \quad (4.3)$$

Then (1,3) element in the matrix

$$\mathbf{B} = \mathbf{O}_1^{-1} \mathbf{A} \mathbf{O}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}$$

is $-a_{12} \sin \theta + a_{13} \cos \theta$. Thus, (1,3) element in \mathbf{B} will be zero if $a_{12} \sin \theta = a_{13} \cos \theta$, that is, if $\tan \theta = \frac{a_{13}}{a_{12}}$.

Finding the values of $\sin \theta$ and $\cos \theta$ from here gives us the transforming matrix \mathbf{O}_1 . Thus, we can find the tri-diagonal form \mathbf{B} of the given symmetric matrix \mathbf{A} of order 3.

Now consider the real symmetric matrix of order 4 given by

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12} & a_{22} & a_{23} & a_{24} \\ a_{13} & a_{23} & a_{33} & a_{34} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{bmatrix}.$$

In this matrix, there are three non-tri-diagonal elements: a_{13} , a_{14} , and a_{24} . Thus, three rotations are required to reduce the given matrix to tri-diagonal form. As discussed above, to annihilate a_{13} , we take orthogonal rotation matrix \mathbf{O}_1 in the (2,3) plane as given in expression (4.3) and obtain the matrix \mathbf{B} with zeros in (1,3) and (3,1) positions. To reduce the element (1,4) in \mathbf{B} to zero, we use the rotation in the (2,4) plane. The orthogonal rotation matrix \mathbf{O}_2 in (2,4) plane is then

$$\mathbf{O}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & 1 & -\sin \theta \\ 0 & 0 & 0 & 0 \\ 0 & \sin \theta & 0 & \cos \theta \end{bmatrix}.$$

Then (1,4) element in the matrix

$$\mathbf{C} = \mathbf{O}_2^{-1} \mathbf{B} \mathbf{O}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & 1 & \sin \theta \\ 0 & 0 & 0 & 0 \\ 0 & -\sin \theta & 0 & \cos \theta \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & 0 & b_{14} \\ b_{12} & b_{22} & b_{23} & b_{24} \\ 0 & b_{23} & b_{33} & b_{34} \\ b_{14} & b_{24} & b_{34} & b_{44} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & 1 & -\sin \theta \\ 0 & 0 & 0 & 0 \\ 0 & \sin \theta & 0 & \cos \theta \end{bmatrix}$$

is $-b_{12} \sin \theta + b_{14} \cos \theta$. Thus, (1,4) element in \mathbf{C} shall be zero if $\tan \theta = \frac{b_{14}}{b_{12}}$. Finding the value of $\sin \theta$ and $\cos \theta$, we get the transforming matrix \mathbf{O}_2 and so \mathbf{C} is obtained with zeros at (1,3), (3,1), (1,4), and (4,1) positions.

To annihilate the element at (2,4) position, we perform rotation in (3,4) plane by taking the orthogonal rotation matrix \mathbf{O}_3 as

$$\mathbf{O}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \theta & -\sin \theta \\ 0 & 0 & \sin \theta & \cos \theta \end{bmatrix}.$$

Then (2,4) element in the matrix

$$\mathbf{D} = \mathbf{O}_3^{-1} \mathbf{C} \mathbf{O}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \theta & \sin \theta \\ 0 & 0 & -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & 0 & 0 \\ c_{12} & c_{22} & c_{23} & c_{24} \\ 0 & c_{23} & c_{33} & c_{34} \\ 0 & c_{24} & c_{34} & c_{44} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \theta & -\sin \theta \\ 0 & 0 & \sin \theta & \cos \theta \end{bmatrix}$$

is $-c_{23} \sin \theta + c_{24} \cos \theta$. Thus, (2,4) element in \mathbf{D} shall be zero if $\tan \theta = \frac{c_{24}}{c_{23}}$. Putting the values of $\sin \theta$ and $\cos \theta$ in \mathbf{D} , we get the required tri-diagonal form of the matrix \mathbf{A} .

In case the matrix \mathbf{A} is of order n , the number of plane rotations required to reduce it to tri-diagonal form is $\frac{1}{2}(n-1)(n-2)$.

EXAMPLE 4.8

Use the Given's method to find eigenvalues of the symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 2 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

Solution. In the matrix \mathbf{A} , there is only one non-tri-diagonal element $a_{13} = 1$ which is to be reduced to zero. Thus, only one rotation is required. To annihilate a_{13} , we take orthogonal matrix in (2,3) plane as

$$\mathbf{O} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix},$$

where $\tan \theta = \frac{a_{13}}{a_{12}} = \frac{1}{3}$. Thus, $\sin \theta = \frac{1}{\sqrt{10}}$ and $\cos \theta = \frac{3}{\sqrt{10}}$. Then

$$\begin{aligned} \mathbf{B} = \mathbf{O}^{-1} \mathbf{A} \mathbf{O} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 2 & 3 & 1 \\ 3 & 2 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ 0 & -\frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} 2 & 3 & 1 \\ 3 & 2 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \\ 0 & \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix} \\ &= \begin{bmatrix} 2 & \sqrt{10} & 0 \\ \sqrt{10} & \frac{31}{10} & \frac{13}{10} \\ 0 & \frac{13}{10} & -\frac{1}{10} \end{bmatrix}, \end{aligned}$$

which is the required tri-diagonal form. The characteristic equation of this tri-diagonal matrix is

$$\begin{vmatrix} 2 - \lambda & \sqrt{10} & 0 \\ \sqrt{10} & \frac{31}{10} - \lambda & \frac{13}{10} \\ 0 & \frac{13}{10} & -\frac{1}{10} - \lambda \end{vmatrix} = 0,$$

which gives

$$-(2-\lambda) \left[\left(\frac{31}{10} - \lambda \right) \left(\frac{1}{10} + \lambda \right) + \frac{169}{100} \right] + \sqrt{10} \left[\sqrt{10} \left(\frac{1}{10} + \lambda \right) \right] = 0$$

or

$$(\lambda - 2)[(31 - 10\lambda)(1 + 10\lambda) + 169] + 100 + 1000\lambda = 0$$

or

$$\lambda^3 - 5\lambda^2 - 6\lambda + 3 = 0.$$

The approximate roots of this characteristic equation are 0.3856, -1.3126, and 5.9269.

EXAMPLE 4.9

Use the Given's method to reduce the symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

to tri-diagonal form and find its eigenvalues.

Solution. Let

$$\mathbf{O} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}$$

be the orthogonal matrix in the (2,3) plane, where $\tan \theta = \frac{a_{13}}{a_{12}} = \frac{1}{2}$. Thus, $\sin \theta = \frac{1}{\sqrt{5}}$ and $\cos \theta = \frac{2}{\sqrt{5}}$. Therefore,

$$\begin{aligned} \mathbf{B} = \mathbf{O}^{-1} \mathbf{A} \mathbf{O} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ 0 & -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \\ &= \begin{bmatrix} 3 & \sqrt{5} & 0 \\ \sqrt{5} & \frac{23}{5} & \frac{6}{5} \\ 0 & \frac{6}{5} & \frac{7}{5} \end{bmatrix}. \end{aligned}$$

The characteristic equation for this tri-diagonal matrix is

$$\begin{vmatrix} 3 - \lambda & \sqrt{5} & 0 \\ \sqrt{5} & \frac{23}{5} - \lambda & \frac{6}{5} \\ 0 & \frac{6}{5} & \frac{7}{5} - \lambda \end{vmatrix} = 0$$

or

$$\lambda^3 - 9\lambda^2 + 18\lambda - 8 = 0.$$

Clearly $\lambda = 2$ satisfies this equation. The reduced equation is

$$\lambda^2 - 7\lambda + 4 = 0,$$

which yields

$$\lambda = \frac{7 \pm \sqrt{49 - 16}}{2} = \frac{7 \pm \sqrt{33}}{2}.$$

Hence, the eigenvalues of the given matrix are 2 , $\frac{7 + \sqrt{33}}{2}$, and $\frac{7 - \sqrt{33}}{2}$.

EXAMPLE 4.10

Use the Given's method to reduce the symmetric matrix

$$\mathbf{C} = \begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix}$$

to tri-diagonal form and find its eigenvalues.

Solution. Let

$$\mathbf{O} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}$$

be the orthogonal matrix in the $(2,3)$ plane, where $\tan \theta = \frac{a_{13}}{a_{12}} = -\frac{1}{3}$. Thus, $\sin \theta = -\frac{1}{\sqrt{10}}$ and $\cos \theta = \frac{3}{\sqrt{10}}$.

Therefore,

$$\begin{aligned} \mathbf{B} = \mathbf{O}^{-1} \mathbf{A} \mathbf{O} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \\ 0 & \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ 0 & -\frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix} \\ &= \begin{bmatrix} 8 & -2\sqrt{10} & 0 \\ -2\sqrt{10} & 9 & -2 \\ 0 & -2 & 1 \end{bmatrix}. \end{aligned}$$

The characteristic equation of the above tri-diagonal matrix is

$$\begin{vmatrix} 8-\lambda & -2\sqrt{10} & 0 \\ -2\sqrt{10} & 9-\lambda & -2 \\ 0 & -2 & 1-\lambda \end{vmatrix} = 0$$

or

$$\lambda(-\lambda^2 + 18\lambda - 45) = 0.$$

Hence, $\lambda = 0, 3$, and 15 are the required eigenvalues.

EXAMPLE 4.11

Use the Given's method to reduce the symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$$

to tri-diagonal form and find its eigenvalues.

Solution. Let

$$\mathbf{O} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}$$

be the orthogonal matrix in the (2,3) plane, where $\tan \theta = \frac{a_{13}}{a_{12}} = 1$. Thus, $\sin \theta = \frac{1}{\sqrt{2}}$ and $\cos \theta = \frac{1}{\sqrt{2}}$.

Therefore,

$$\begin{aligned} \mathbf{B} = \mathbf{O}^{-1} \mathbf{A} \mathbf{O} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 2\sqrt{2} & 0 \\ 2\sqrt{2} & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \end{aligned}$$

The characteristic equation of this tri-diagonal matrix is

$$\begin{vmatrix} 1-\lambda & 2\sqrt{2} & 0 \\ 2\sqrt{2} & 3-\lambda & 0 \\ 0 & 0 & -1-\lambda \end{vmatrix} = 0$$

or

$$\lambda^3 - 3\lambda^2 - 9\lambda - 5 = 0.$$

Hence, the characteristic roots are $-1, -1$, and 5 .

EXAMPLE 4.12

Use the Given's method to reduce the symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 1 & 2 & 2 \\ 2 & 2 & 1 & 3 \\ 2 & 2 & 3 & 1 \end{bmatrix}$$

to tri-diagonal form.

Solution. In the given symmetric matrix there are three non-tri-diagonal elements a_{13}, a_{14} , and a_{24} . Thus, three rotations are required to reduce the matrix to tri-diagonal form. To annihilate a_{13} , we use the orthogonal rotation matrix

$$\mathbf{O}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta & 0 \\ 0 & \sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where $\tan \theta = \frac{a_{13}}{a_{12}} = 1$. Thus, $\sin \theta = \frac{1}{\sqrt{2}}$ and $\cos \theta = \frac{1}{\sqrt{2}}$. Hence,

$$\mathbf{O}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and so

$$\mathbf{B} = \mathbf{O}_1^{-1} \mathbf{A} \mathbf{O}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 1 & 2 & 2 \\ 2 & 2 & 1 & 3 \\ 2 & 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \frac{4}{\sqrt{2}} & 0 & 2 \\ \frac{4}{\sqrt{2}} & 3 & 0 & \frac{5}{\sqrt{2}} \\ 0 & 0 & -1 & \frac{1}{\sqrt{2}} \\ 2 & \frac{5}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 1 \end{bmatrix}.$$

To reduce the element (1,4) in \mathbf{B} to zero, we use the rotation \mathbf{O}_2 in (2,4) plane given by

$$\mathbf{O}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & 0 & -\sin \theta \\ 0 & 0 & 1 & 0 \\ 0 & \sin \theta & 0 & \cos \theta \end{bmatrix},$$

where $\tan \theta = \frac{b_{14}}{b_{12}} = \frac{1}{\sqrt{2}}$. Thus, $\cos \theta = \frac{\sqrt{2}}{\sqrt{3}}$ and $\sin \theta = \frac{1}{\sqrt{3}}$. Hence,

$$\mathbf{O}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{\sqrt{2}}{\sqrt{3}} & 0 & -\frac{1}{\sqrt{3}} \\ 0 & 0 & 1 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 & \frac{\sqrt{2}}{\sqrt{3}} \end{bmatrix}.$$

Hence, the second rotation yields

$$\begin{aligned} \mathbf{C} = \mathbf{O}_2^{-1} \mathbf{B} \mathbf{O}_2 &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{\sqrt{2}}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} \\ 0 & 0 & 1 & 0 \\ 0 & -\frac{1}{\sqrt{3}} & 0 & \frac{\sqrt{2}}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 1 & \frac{4}{\sqrt{2}} & 0 & 2 \\ \frac{4}{\sqrt{2}} & 3 & 0 & \frac{5}{\sqrt{2}} \\ 0 & 0 & -1 & \frac{1}{\sqrt{2}} \\ 2 & \frac{5}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{\sqrt{2}}{\sqrt{3}} & 0 & -\frac{1}{\sqrt{3}} \\ 0 & 0 & 1 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 & \frac{\sqrt{2}}{\sqrt{3}} \end{bmatrix} \\ &= \begin{bmatrix} 1 & \frac{6}{\sqrt{3}} & 0 & 0 \\ \frac{6}{\sqrt{3}} & \frac{17}{3} & \frac{1}{\sqrt{2}\sqrt{3}} & \frac{1}{3\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}\sqrt{3}} & -1 & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{5}{3} \end{bmatrix}. \end{aligned}$$

To annihilate (2,4) element in \mathbf{C} , we use the rotation matrix \mathbf{O}_3 in (3,4) plane given by

$$\mathbf{O}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \theta & -\sin \theta \\ 0 & 0 & \sin \theta & \cos \theta \end{bmatrix},$$

where $\tan \theta = \frac{c_{24}}{c_{23}} = \frac{1}{\sqrt{3}}$. Thus, $\sin \theta = \frac{1}{2}$ and $\cos \theta = \frac{\sqrt{3}}{2}$. Therefore,

$$\mathbf{O}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}.$$

Hence, the third rotation yields

$$\begin{aligned} \mathbf{D} = \mathbf{O}_3^{-1} \mathbf{C} \mathbf{O}_3 &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 1 & \frac{6}{\sqrt{3}} & 0 & 0 \\ \frac{6}{\sqrt{3}} & \frac{17}{3} & \frac{1}{\sqrt{2}\sqrt{3}} & \frac{1}{3\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}\sqrt{3}} & -1 & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{5}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \\ &= \begin{bmatrix} 1 & \frac{6}{\sqrt{3}} & 0 & 0 \\ \frac{6}{\sqrt{3}} & \frac{17}{3} & \frac{2}{3\sqrt{2}} & 0 \\ 0 & \frac{2}{3\sqrt{2}} & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}, \end{aligned}$$

which is the required tri-diagonal form.

4.5 HOUSEHOLDER'S METHOD

This method is used for finding eigenvalues of real symmetric matrices. The first step of the method consists of reducing the given matrix \mathbf{A} to a band matrix. This is carried out by orthogonal transformations. The orthogonal matrices, denoted by \mathbf{P}_r , are of the form

$$\mathbf{P} = \mathbf{I} - 2\omega\omega^T,$$

where ω is a column vector such that

$$\omega^T \omega = 1. \quad (4.4)$$

We note that

$$\begin{aligned} \mathbf{P}^T &= (\mathbf{I} - 2\omega\omega^T)^T \\ &= [\mathbf{I} - 2(\omega\omega^T)^T], \\ &= \mathbf{I} - 2(\omega\omega^T) = \mathbf{P} \end{aligned}$$

and so \mathbf{P} is symmetric. Further,

$$\begin{aligned}\mathbf{P}^T \mathbf{P} &= (\mathbf{I} - 2\omega\omega^T)^T(\mathbf{I} - 2\omega\omega^T) \\ &= (\mathbf{I} - 2\omega\omega^T)(\mathbf{I} - 2\omega\omega^T) \\ &= \mathbf{I} - 4\omega\omega^T + 4\omega\omega^T\omega\omega^T \\ &= \mathbf{I} - 4\omega\omega^T + 4\omega\mathbf{I}\omega^T \\ &= \mathbf{I} - 4\omega\omega^T + 4\omega\omega^T = \mathbf{I}\end{aligned}$$

and so \mathbf{P} is orthogonal. Thus, \mathbf{P} is symmetric orthogonal matrix.

The vectors ω are constructed with the first $(r-1)$ zero components. Thus,

$$\omega_r = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \\ x_r \\ x_{r+1} \\ \dots \\ x_n \end{bmatrix}.$$

With this choice of ω_r , we form

$$\mathbf{P}_r = \mathbf{I} - 2\omega_r\omega_r^T.$$

Then equation (4.4) implies

$$x_r^2 + x_{r+1}^2 + \dots + x_n^2 = 1.$$

Now, put $\mathbf{A} = \mathbf{A}_1$ and form successively

$$\mathbf{A}_r = \mathbf{P}_r \mathbf{A}_{r-1} \mathbf{P}_r, r = 2, 3, \dots, n-1.$$

At the first transformation, we get zeros in the positions $(1,3), (1,4), \dots (1,n)$ and in the corresponding places in the first column. In the second transformation, we get zeros in the positions $(2,4), (2,5), \dots, (2,n)$ and in the corresponding places in the second column. The final result will be a band matrix:

$$\mathbf{B} = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & 0 & \dots & \dots & \dots & 0 \\ \beta_1 & \alpha_2 & \beta_2 & 0 & \dots & \dots & \dots & 0 \\ 0 & \beta_2 & \alpha_3 & \beta_3 & 0 & \dots & \dots & 0 \\ 0 & 0 & \beta_3 & \alpha_4 & \beta_4 & \dots & \dots & 0 \\ \dots & \dots \\ \dots & \dots \\ \dots & \beta_{n-1} \\ 0 & 0 & 0 & 0 & \dots & \dots & \beta_{n-1} & \alpha_n \end{bmatrix}.$$

Then the characteristic equation $|\mathbf{B} - \lambda\mathbf{I}| = 0$ gives the required eigenvalues.

To describe the application of this method, we consider the real symmetric matrix, \mathbf{A} of order 3 given by

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}.$$

We wish to find a real symmetric orthogonal matrix \mathbf{P}_1 such that

$$\mathbf{P}_1 \mathbf{A} \mathbf{P}_1 = \begin{bmatrix} a_{11}^1 & a_{12}^1 & 0 \\ a_{12}^1 & a_{22}^1 & a_{23}^1 \\ 0 & a_{23}^1 & a_{33}^1 \end{bmatrix}. \quad (4.5)$$

We take

$$\boldsymbol{\omega} = \begin{bmatrix} 0 \\ \omega_2 \\ \omega_3 \end{bmatrix}$$

such that $\boldsymbol{\omega}\boldsymbol{\omega}^T = \mathbf{I}$ and $\omega_2^2 + \omega_3^2 = 1$. Then

$$\mathbf{P}_1 = \mathbf{I} - 2\boldsymbol{\omega}\boldsymbol{\omega}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 - 2\omega_2^2 & -2\omega_2\omega_3 \\ 0 & -2\omega_2\omega_3 & 1 - 2\omega_3^2 \end{bmatrix}. \quad (4.6)$$

Then

$$\mathbf{P}_1 \mathbf{A} \mathbf{P}_1 = \begin{bmatrix} a_{11} & a_{12}(1 - 2\omega_2^2) - 2a_{13}\omega_2\omega_3 & -2a_{12}\omega_2\omega_3 + a_{13}(1 - 2\omega_3^2) \\ a_{12} & a_{22}(1 - 2\omega_2^2) - 2a_{23}\omega_2\omega_3 & -2a_{22}\omega_2\omega_3 + a_{23}(1 - 2\omega_3^2) \\ a_{13} & a_{23}(1 - 2\omega_2^2) - 2a_{33}\omega_2\omega_3 & -2a_{23}\omega_2\omega_3 + a_{33}(1 - 2\omega_3^2) \end{bmatrix}. \quad (4.7)$$

Comparing equations (4.5) and (4.7), we get

$$\begin{aligned} a_{11}^1 &= a_{11} \\ a_{12}^1 &= a_{12}(1 - 2\omega_2^2) - 2a_{13}\omega_2\omega_3 \\ &= a_{12} - 2a_{12}\omega_2^2 - 2a_{13}\omega_2\omega_3 \\ &= a_{12} - 2\omega_2(a_{12}\omega_2 + a_{13}\omega_3) \\ &= a_{12} - 2\omega_2 q, \end{aligned} \quad (4.8)$$

$$\begin{aligned} 0 &= -2a_{12}\omega_2\omega_3 + a_{13}(1 - 2\omega_3^2) \\ &= -2a_{12}\omega_2\omega_3 + a_{13} - 2a_{13}\omega_3^2 \\ &= a_{13} - 2\omega_3(a_{12}\omega_2 + a_{13}\omega_3) \\ &= a_{13} - 2\omega_3 q, \end{aligned} \quad (4.9)$$

where

$$q = a_{12}\omega_2 + a_{13}\omega_3.$$

Squaring and adding equations (4.8) and (4.9), we get

$$\begin{aligned}(a_{12}^1)^2 &= a_{12}^2 + a_{13}^2 + 4q^2(\omega_2^2 + \omega_3^2) - 4q(a_{12}\omega_2 + a_{13}\omega_3) \\ &= a_{12}^2 + a_{13}^2 + 4q^2 - 4q^2 \\ &= a_{12}^2 + a_{13}^2\end{aligned}$$

and so

$$a_{12}^1 = \pm \sqrt{a_{12}^2 + a_{13}^2}.$$

Thus,

$$a_{12}^1 = a_{12} - 2q\omega_2 = \pm \sqrt{a_{12}^2 + a_{13}^2} = \pm S, \text{ say} \quad (4.10)$$

and

$$0 = a_{13} - 2\omega_3 q. \quad (4.11)$$

Multiplying equation (4.10) by ω_2 and (4.11) by ω_3 and adding we get

$$a_{12}\omega_2 - 2q\omega_2^2 + a_{13}\omega_3 - 2q\omega_3^2 = \pm S\omega_2$$

or

$$a_{12}\omega_2 + a_{13}\omega_3 - 2q(\omega_2^2 + \omega_3^2) = \pm S\omega_2$$

or

$$-q = \pm S\omega_2.$$

Therefore, equation (4.10) yields

$$a_{12} - 2\omega_2(\mp S\omega_2) = \pm S$$

and so

$$\omega_2^2 = \frac{1}{2} \left[1 \mp \frac{a_{12}}{S} \right]. \quad (4.12)$$

Now equation (4.11) gives

$$\begin{aligned}\omega_3 &= \frac{a_{13}}{2q} = \frac{a_{13}}{\mp 2S\omega_2} \\ &= \mp \frac{a_{13}}{2\omega_2 \sqrt{a_{12}^2 + a_{13}^2}}.\end{aligned} \quad (4.13)$$

The error in equation (4.13) will be minimum if ω_2 is large. Therefore, sign in equation (4.12) should be the same as that of a_{12} . Putting these values of ω_2 and ω_3 , we get ω and then $\mathbf{P}_1 = \mathbf{I} - 2\omega\omega^T$ which reduces $\mathbf{P}_1 \mathbf{A} \mathbf{P}_1$ in the tri-diagonal form.

Working Rule for Householder's Method

To find the vector $\omega = \begin{bmatrix} 0 \\ \omega_2 \\ \omega_3 \end{bmatrix}$, compute

$$S = \sqrt{a_{12}^2 + a_{13}^2},$$

$$\omega_2^2 = \frac{1}{2} \left[1 \mp \frac{a_{12}}{S} \right],$$

$$\omega_3 = \mp \frac{a_{13}}{2x_2 \sqrt{a_{12}^2 + a_{13}^2}},$$

where sign in ω_2^2 should be the same as that of a_{12} . Then find $\mathbf{P}_1 = \mathbf{I} - 2\omega\omega^T$ and compute $\mathbf{P}_1 \mathbf{A} \mathbf{P}_1$ and so on.

EXAMPLE 4.13

Use Householder's method to reduce the symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

to tri-diagonal form.

Solution. We have

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

Then

$$\begin{aligned} S &= \sqrt{a_{12}^2 + a_{13}^2} = \sqrt{5} \\ \omega_2^2 &= \frac{1}{2} \left[1 + \frac{a_{12}}{S} \right] = \frac{1}{2} \left[1 + \frac{2}{\sqrt{5}} \right] = 0.9472, \end{aligned}$$

or

$$\omega_2 = 0.9732.$$

Moreover,

$$\omega_3 = \frac{a_{13}}{2\omega_2 S} = \frac{1}{2(0.9732)\sqrt{5}} = 0.2298.$$

Hence,

$$\omega = \begin{bmatrix} 0 \\ 0.9732 \\ 0.2298 \end{bmatrix}$$

and so

$$\begin{aligned} \mathbf{P}_1 &= \mathbf{I} - 2 \begin{bmatrix} 0 \\ 0.9732 \\ 0.2298 \end{bmatrix} \begin{bmatrix} 0 & 0.9732 & 0.2298 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1.8942 & 0.4472 \\ 0 & 0.4472 & 0.1056 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.8942 & -0.4472 \\ 0 & -0.4472 & 0.8944 \end{bmatrix}. \end{aligned}$$

Therefore the first transformation yields

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{P}_1 \mathbf{A} \mathbf{P}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.8942 & -0.4472 \\ 0 & -0.4472 & 0.8944 \end{bmatrix} \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.8942 & -0.4472 \\ 0 & -0.4472 & 0.8944 \end{bmatrix} \\ &= \begin{bmatrix} 3 & -2.2356 & 0 \\ -2.2356 & 4.5983 & -1.1999 \\ 0 & -1.1998 & 1.4002 \end{bmatrix}, \end{aligned}$$

which is the required tri-diagonal form.

EXAMPLE 4.14

Reduce the symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 4 \\ 3 & 1 & 2 \\ 4 & 2 & 1 \end{bmatrix}$$

to tri-diagonal form by Householder's method.

Solution. We have

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 4 \\ 3 & 1 & 2 \\ 4 & 2 & 1 \end{bmatrix}.$$

Then

$$S = \sqrt{a_{12}^2 + a_{13}^2} = \sqrt{9 + 16} = 5,$$

$$\omega_2^2 = \frac{1}{2} \left[1 + \frac{a_{12}}{S} \right] = \frac{1}{2} \left[1 + \frac{3}{5} \right] = \frac{4}{5}$$

and so

$$\omega_2 = \frac{2}{\sqrt{5}},$$

$$\omega_3 = \frac{a_{13}}{2\omega_2 S} = \frac{1}{\sqrt{5}}.$$

Therefore,

$$\omega = \begin{bmatrix} 0 \\ 2 \\ \hline \sqrt{5} \\ 1 \\ \hline \sqrt{5} \end{bmatrix}$$

and so

$$\mathbf{P}_1 = \mathbf{I} - 2\omega\omega^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{3}{5} & -\frac{4}{5} \\ 0 & -\frac{4}{5} & \frac{3}{5} \end{bmatrix}.$$

Now, the first transformation yields

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{P}_1 \mathbf{A} \mathbf{P}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{3}{5} & -\frac{4}{5} \\ 0 & -\frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 1 & 3 & 4 \\ 3 & 1 & 2 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{3}{5} & -\frac{4}{5} \\ 0 & -\frac{4}{5} & \frac{3}{5} \end{bmatrix} \\ &= \begin{bmatrix} 1 & -5 & 0 \\ -5 & \frac{73}{25} & \frac{14}{25} \\ 0 & \frac{14}{25} & -\frac{23}{25} \end{bmatrix}, \end{aligned}$$

which is the required tri-diagonal form.

4.6 EIGENVALUES OF A SYMMETRIC TRI-DIAGONAL MATRIX

We have seen that Given's method and Householder's method reduce a given matrix to a tri-diagonal matrix

$$\mathbf{A}_1 = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{12} & a_{22} & a_{23} \\ 0 & a_{23} & a_{33} \end{bmatrix}.$$

Then the characteristic roots are given by

$$|\mathbf{A}_1 - \lambda \mathbf{I}| = \begin{vmatrix} a_{11} - \lambda & a_{12} & 0 \\ a_{12} & a_{22} - \lambda & a_{23} \\ 0 & a_{23} & a_{33} - \lambda \end{vmatrix} = 0,$$

that is, $f_3(\lambda) = 0$, where

$$\begin{aligned} f_3(\lambda) &= (a_{33} - \lambda) \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{12} & a_{22} - \lambda \end{vmatrix} - a_{23} \begin{vmatrix} a_{11} - \lambda & 0 \\ a_{12} & a_{23} \end{vmatrix} \\ &= (a_{33} - \lambda) f_2(\lambda) - a_{23}^2 (a_{11} - \lambda) \\ &= (a_{33} - \lambda) f_2(\lambda) - a_{23}^2 f_1(\lambda), \end{aligned}$$

where

$$\begin{aligned}f_1(\lambda) &= a_{11} - \lambda = (a_{11} - \lambda)f_0(\lambda), f_0(\lambda) = 1 \\f_2(\lambda) &= \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{12} & a_{22} - \lambda \end{vmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}^2 \\&= (a_{22} - \lambda)f_1(\lambda) - a_{12}^2 f_0(\lambda).\end{aligned}$$

Thus, we have the recursion formula

$$\begin{aligned}f_0(\lambda) &= 1 \\f_1(\lambda) &= (a_{11} - \lambda)f_0(\lambda) \\f_2(\lambda) &= (a_{22} - \lambda)f_1(\lambda) - a_{12}^2 f_0(\lambda) \\f_3(\lambda) &= (a_{33} - \lambda)f_2(\lambda) - a_{23}^2 f_1(\lambda),\end{aligned}$$

that is

$$f_k(\lambda) = (a_{kk} - \lambda)f_{k-1}(\lambda) - a_{k-1,k}^2 f_{k-2}(\lambda) \text{ for } 2 \leq k \leq n.$$

The sequence of functions $f_0(\lambda), f_1(\lambda), f_2(\lambda), \dots, f_k(\lambda)$ is known as Sturm sequence.

EXAMPLE 4.15

Using Sturm sequence, find the eigenvalues of the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}.$$

Solution. We have seen in Example 4.10 that the Given's method reduces the given matrix to the tri-diagonal form

$$\begin{bmatrix} 1 & 2\sqrt{2} & 0 \\ 2\sqrt{2} & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Then the Sturm sequence is

$$\begin{aligned}f_0(\lambda) &= 1 \\f_1(\lambda) &= (a_{11} - \lambda)f_0(\lambda) = (1 - \lambda) \\f_2(\lambda) &= (a_{22} - \lambda)f_1(\lambda) - a_{12}^2 f_0(\lambda) \\&= (3 - \lambda)(1 - \lambda) - 8 = \lambda^2 - 4\lambda - 5 \\f_3(\lambda) &= (a_{33} - \lambda)f_2(\lambda) - a_{23}^2 f_1(\lambda) \\&= (-1 - \lambda)(\lambda^2 - 4\lambda - 5) - 0 \\&= -(\lambda - 5)(\lambda + 1)(\lambda + 1).\end{aligned}$$

Hence, $f_3(\lambda) = 0$ yields the eigenvalues $-1, -1, 5$.

EXAMPLE 4.16

Using Sturm sequence, find the eigenvalues of the matrix

$$\begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix}.$$

Solution. The tri-diagonal form of the given matrix is (see Example 4.9)

$$\begin{bmatrix} 8 & -2\sqrt{10} & 0 \\ -2\sqrt{10} & 9 & -2 \\ 0 & -2 & 1 \end{bmatrix}.$$

Then the Sturm sequence is

$$\begin{aligned} f_0(\lambda) &= 1 \\ f_1(\lambda) &= (a_{11} - \lambda)f_0(\lambda) = (8 - \lambda) \\ f_2(\lambda) &= (a_{22} - \lambda)f_1(\lambda) - a_{12}^2 f_0(\lambda) \\ &= (9 - \lambda)(8 - \lambda) - 40 = \lambda^2 - 17\lambda + 32 \\ f_3(\lambda) &= (a_{33} - \lambda)f_2(\lambda) - a_{23}^2 f_1(\lambda) \\ &= (1 - \lambda)(\lambda^2 - 17\lambda + 32) - 4(8 - \lambda) \\ &= -\lambda^3 + 18\lambda^2 - 45\lambda \\ &= \lambda(\lambda - 3)(\lambda - 15). \end{aligned}$$

Hence, $\lambda = 0, 3, 15$ are the eigenvalues of the given matrix.

4.7 BOUNDS ON EIGENVALUES (GERSCHGORIN CIRCLES)

Some applications in engineering require only bounds on eigenvalues instead of their accurate approximations. These bounds can be obtained using the following two results, known as Gerschgorin Theorems.

Theorem 4.1. (First Gerschgorin Theorem). Every eigenvalue of an $n \times n$ matrix $\mathbf{A} = [a_{ij}]$ lies inside at least one of the circles, called Gerschgorin circles, in the complex plane with center a_{ii} and radii $r_i = \sum_{j=1}^n |a_{ij}|$, $i = 1, 2, \dots, n$. In other words, all the eigenvalues of the matrix \mathbf{A} lie in the union of the disks $|z - a_{ii}| \leq r_i = \sum_{j=1}^n |a_{ij}|$, $i = 1, 2, \dots, n$ in the complex plane.

Proof: Let λ be an eigenvalue of $\mathbf{A} = [a_{ij}]$ and \mathbf{X} be the corresponding eigenvector. Then $\mathbf{AX} = \lambda\mathbf{X}$, which yields

$$\left. \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = \lambda x_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = \lambda x_2 \\ \dots \\ a_{ii}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = \lambda x_i \\ \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = \lambda x_n \end{array} \right\} \quad (4.14)$$

Let x_i be the largest component of vector $\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$. Then $\left| \frac{x_m}{x_i} \right| \leq 1$ for $m = 1, 2, \dots, n$.

Dividing i th in equation (4.14) by x_i , we have

$$a_{i1} \frac{x_1}{x_i} + a_{i2} \frac{x_2}{x_i} + \dots + a_{i,i-1} \frac{x_{i-1}}{x_i} + a_{ii} + \dots + a_{in} \frac{x_n}{x_i} = \lambda$$

or

$$\lambda - a_{ii} = a_{i1} \frac{x_1}{x_i} + a_{i2} \frac{x_2}{x_i} + \dots + a_{i,i-1} \frac{x_{i-1}}{x_i} + \dots + a_{in} \frac{x_n}{x_i}.$$

Since $\left| \frac{x_m}{x_i} \right| \leq 1$, we get

$$|\lambda - a_{ii}| \leq |a_{i1}| + |a_{i2}| + \dots + |a_{i,i-1}| + |a_{i,i+1}| + \dots + |a_{in}| = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|.$$

This completes the proof of the theorem.

Since the disk $|z - a_{ii}| \leq r_i$ is contained within the disk

$$|z| \leq |a_{ii}| + r_i = |a_{ii}| + \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| = \sum_{j=1}^n |a_{ij}|,$$

centered at the origin, it follows that

"All the eigenvalues of the matrix \mathbf{A} lie within the disk $|z| \leq \max_i \left\{ \sum_{j=1}^n |a_{ij}| \right\}$, $i = 1, 2, \dots, n$ centered at the origin."

Theorem 4.2. (Second Gershgorin Theorem). If the union of m of the Gershgorin circles forms a connected region isolated from the remaining circles, then exactly m of the eigenvalues of \mathbf{A} lie within that region.

EXAMPLE 4.17

Determine the Gershgorin circles corresponding to the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 1 \end{bmatrix}.$$

Solution. The three Gershgorin circles are

$$(a) |z - 1| = |2| + |3| = 5$$

$$(b) |z - 4| = |2| + |6| = 8$$

$$(c) |z - 1| = |3| + |6| = 9.$$

Thus, one eigenvalue lies within the circle centered at (1,0) with radius 5, the second eigenvalue lies within the circle centered at (4,0) with radius 8, and the third lies within the circle with center (1,0) and radius 9.

Since disk (a) lies within (c) it follows that all the eigenvalues of \mathbf{A} lie within the region defined by disks (b) and (c). Hence,

$$-4 \leq \lambda \leq 12 \text{ and } -8 \leq \lambda \leq 10$$

and so

$$-8 \leq \lambda \leq 12.$$

EXERCISES

1. Find the largest eigenvalue and the corresponding eigenvector of the matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & 2 \\ 0 & 0 & 7 \end{bmatrix}.$$

Ans. 7, $\begin{bmatrix} \frac{37}{66} & \frac{2}{11} & 1 \end{bmatrix}^T$

2. Using power method, determine the largest eigenvalue and the corresponding eigenvector of the matrix

$$\begin{bmatrix} 1 & 6 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Ans. 4, $[2 \ 1 \ 0]^T$

3. Using power method, determine the largest eigenvalue and the corresponding eigenvector of the matrix

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

Ans. 3.41, $[0.74 \ -1 \ 0.67]^T$

4. Determine the largest eigenvalue and the corresponding eigenvector of the matrix

$$\begin{bmatrix} 10 & -2 & 1 \\ -2 & 10 & -2 \\ 1 & -2 & 10 \end{bmatrix}.$$

Ans. 9, $[1 \ 0 \ -1]^T$

5. Using Jacobi's method find the eigenvalues of the matrix

$$\begin{bmatrix} 5 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 5 \end{bmatrix}.$$

Ans. 4, -2, 6

6. Reduce the matrix

$$\begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & 2 \\ 3 & 2 & 3 \end{bmatrix}.$$

to tri-diagonal form by Given's method.

$$\text{Ans. } \begin{bmatrix} 2 & 3.16 & 0 \\ 3.16 & 4.3 & -1.9 \\ 0 & -1.9 & 3.9 \end{bmatrix}$$

7. Use Given's method to reduce the matrix

$$\begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

to tri-diagonal form and find its eigenvalues using Sturm's sequence.

$$\text{Ans. } \begin{bmatrix} 3 & \sqrt{2} & 0 \\ \sqrt{2} & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}, 1, 4 \pm \sqrt{3}$$

8. Use the Given's method to reduce the Hilbert matrix

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}$$

to tri-diagonal matrix.

$$\text{Ans. } \begin{bmatrix} 1 & \frac{\sqrt{13}}{6} & 0 \\ \frac{\sqrt{13}}{6} & \frac{34}{65} & \frac{9}{260} \\ 0 & \frac{9}{260} & \frac{2}{195} \end{bmatrix}$$

9. Reduce the matrix

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

to tri-diagonal form by Householder's method and use Sturm's sequence to find its eigenvalues.

$$\text{Ans. } 0, 3, 3$$

10. Reduce the matrix

$$\begin{bmatrix} 1 & 4 & 3 \\ 4 & 1 & 2 \\ 3 & 2 & 1 \end{bmatrix}$$

to tri-diagonal form by Householder's method.

$$\text{Ans. } \begin{bmatrix} 1 & -5 & 0 \\ -5 & \frac{73}{25} & \frac{-14}{25} \\ 0 & \frac{-14}{25} & \frac{-11}{25} \end{bmatrix}$$

11. Reduce the matrix

$$\begin{bmatrix} 1 & 3 & 4 \\ 3 & 2 & -1 \\ 4 & -1 & 1 \end{bmatrix}$$

to tri-diagonal form by Householder's method.

$$\text{Ans. } \begin{bmatrix} 1 & -5 & 0 \\ -5 & \frac{2}{5} & \frac{1}{5} \\ 0 & \frac{1}{5} & \frac{3}{5} \end{bmatrix}$$

12. Using Faddeev–Leverrier method find the characteristic equation of the matrix

$$(i) \begin{bmatrix} 1 & 1 & -2 \\ -1 & 2 & 1 \\ 0 & 1 & -1 \end{bmatrix}, \quad (ii) \begin{bmatrix} -1 & 0 & 0 \\ 1 & -2 & 3 \\ 0 & 2 & -3 \end{bmatrix}.$$

$$\text{Ans. (i)} \lambda^3 - 2\lambda^2 - \lambda + 2 = 0 \quad \text{(ii)} \lambda^3 + 6\lambda^2 + 5\lambda = 0$$

13. Using power method and deflation method, find the dominant and subdominant eigenvalues of the matrix

$$\begin{bmatrix} 2 & 2 & 0 \\ 2 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$\text{Ans. } 6, 3, 1.$$

14. Determine Gerschgorin circles corresponding to the matrix

$$\begin{bmatrix} 10 & -1 & 0 \\ -1 & 2 & 2 \\ 0 & 2 & 3 \end{bmatrix}.$$

$$\text{Ans. } |z - 10| = 1 \\ |z - 2| = 3 \\ |z - 3| = 2$$

15. Using Gerschgorin circles, show that the eigenvalues of the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

satisfy the inequality $0 \leq \lambda \leq 7$.

5 Finite Differences and Interpolation

Finite differences play a key role in the solution of differential equations and in the formulation of interpolating polynomials. The interpolation is the art of reading between the tabular values. Also the interpolation formulae are used to derive formulae for numerical differentiation and integration.

5.1 FINITE DIFFERENCES

Suppose that a function $y = f(x)$ is tabulated for the equally spaced arguments $x_0, \dots, x_n + 2h, \dots, x_0 + nh$ giving the functional values y_0, \dots, y_n . The constant difference between two consecutive values of x is called the interval of differencing and is denoted by h .

The operator Δ defined by

$$\begin{aligned}\Delta y_0 &= y_1 - y_0, \\ \Delta y_1 &= y_2 - y_1, \\ &\dots, \\ &\dots, \\ \Delta y_{n-1} &= y_n - y_{n-1}.\end{aligned}$$

is called the Newton's forward difference operator. We note that the first difference $\Delta y_n = y_{n+1} - y_n$ is itself a function of x . Consequently, we can repeat the operation of differencing to obtain

$$\begin{aligned}\Delta^2 y_n &= \Delta(\Delta y_n) = \Delta(y_{n+1} - y_n) = \Delta y_{n+1} - \Delta y_n, \\ &= y_{n+1} - y_n - (y_n - y_{n-1}) = y_{n+1} - 2y_n + y_{n-1},\end{aligned}$$

which is called the second forward difference. In general, the n th difference of f is defined by

$$\Delta^n y_r = \Delta^{n-1} y_{r+1} - \Delta^{n-1} y_r.$$

For example, let

$$f(x) = x^3 - 3x^2 + 5x + 7.$$

Taking the arguments as 0, 2, 4, 6, 8, 10, we have $h = 2$ and

$$\begin{aligned}\Delta f(x) &= (x+2)^3 - 3(x+2)^2 + 5(x+2) + 7 - (x^3 - 3x^2 + 5x + 7) = 6x^2 + 6, \\ \Delta^2 f(x) &= \Delta(\Delta f(x)) = \Delta(6x^2 + 6) = 6(x+2)^2 + 6 - (6x^2 + 6) = 24x + 24, \\ \Delta^3 f(x) &= 24(x+2) + 24 - (24x + 24) = 48, \\ \Delta^4 f(x) &= \Delta^5 f(x) = \dots = 0.\end{aligned}$$

In tabular form, we have

Difference Table

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$	$\Delta^4 f(x)$	$\Delta^5 f(x)$
0	7	6				
2	13	30	24	48		
4	43	102	72	0		
6	145	222	120	0		
8	367	390	168	48		
10	757					

Theorem 5.1. If $f(x)$ is a polynomial of degree n , that is,

$$f(x) = \sum_{i=0}^n a_i x^i,$$

then $\Delta^n f(x)$ is constant and is equal to $n! a_n h^n$

Proof: We shall prove the theorem by induction on n . If $n = 1$, then $f(x) = a_1 x + a_0$ and $\Delta f(x) = f(x+h) - f(x) = a_1 h$ and so the theorem holds for $n = 1$. Assume now that the result is true for all degrees 1, 2, ..., $n-1$. Consider

$$f(x) = \sum_{i=0}^n a_i x^i.$$

Then by the linearity of the operator Δ , we have

$$\Delta^n f(x) = \sum_{i=0}^n a_i \Delta^n x^i.$$

For $i < n$, $\Delta^n x^i$ is the n th difference of a polynomial of degree less than n and hence must vanish, by induction hypothesis. Thus,

$$\begin{aligned} \Delta^n f(x) &= a_n \Delta^n x^n = a_n \Delta^{n-1}(\Delta x^n) \\ &= a_n \Delta^{n-1}[(x+h)^n - x^n] \\ &= a_n \Delta^{n-1}[nhx^{n-1} + g(x)] \end{aligned}$$

where $g(x)$ is a polynomial of degree less than $n-1$. Hence, by induction hypothesis,

$$\Delta^n f(x) = a_n \Delta^{n-1}(nhx^{n-1}) = a_n (hn)(n-1) \dots n^{n-1} = a_n n! h^n.$$

Hence, by induction, the theorem holds.

Let y_0, y_1, \dots, y_n be the functional values of a function f for the arguments $x_0, x_1 + h, x_2 + 2h, \dots, x_n + nh$. Then the operator ∇ defined by

$$y_r = y_r - y_{r-1}$$

is called the Newton's backward difference operator.

The higher-order backward differences are

$$\begin{aligned}\nabla^2 y_r &= \nabla y_r - \nabla y_{r-1} \\ \nabla^3 y_r &= \nabla^2 y_r - \nabla^2 y_{r-1} \\ &\dots \\ \nabla^n y_r &= \nabla^{n-1} y_r - \nabla^{n-1} y_{r-1}.\end{aligned}$$

Thus, the backward difference table becomes

x	y	1st difference	2nd difference	3rd difference
x_0	y_0			
x_1	y_1	∇y_1	$\nabla^2 y_2$	
x_2	y_2	∇y_2	$\nabla^2 y_3$	$\nabla^3 y_3$
x_3	y_3	∇y_3		

EXAMPLE 5.1

Form the table of backward differences for the function

$$f(x) = x^3 - 3x^2 + 5x - 7$$

for $x = -1, 0, 1, 2, 3, 4$, and 5 .

Solution.

x	y	1st difference	2nd difference	3rd difference	4th difference
-1	-16	9			
0	-7	3	-6	6	0
1	-4	3	0	6	0
2	-1	9	6	6	0
3	8	21	12	6	0
4	29	39	18		
5	68				

An operator E , known as enlargement operator, displacement operator or shifting operator, is defined by

$$Ey_r = y_{r+1}.$$

Thus, shifting operator moves the functional value $f(x)$ to the next higher value $f(x+h)$. Further,

$$\begin{aligned} E^2 y_r &= E(Ey_r) = E(y_{r+1}) = y_{r+2} \\ E^3 y_r &= E(E^2 y_r) = E(y_{r+2}) = y_{r+3}. \\ \dots &\dots \\ E^n y_r &= y_{r+n} \end{aligned}$$

Relations between Δ , ∇ , and E

We know that

$$\Delta y_r = y_{r+1} - y_r = r y_r - y_r = (r - I)y_r,$$

where I is the identity operator. Hence,

$$\Delta = E - I \text{ or } E = I + \Delta. \quad (5.1)$$

Also, by definition,

$$y_r = y_r - y_{r-1} = y_r - E^{-1} y_r = y_r (I - E^{-1}),$$

and so

$$\nabla = I - E^{-1} \text{ or } E^{-1} = I - \nabla$$

or

$$E = \frac{I}{I - \nabla}. \quad (5.2)$$

From equations (5.1) and (5.2), we have

$$I + \Delta = \frac{1}{I - \nabla} \quad (5.3)$$

or

$$\Delta = \frac{I}{I - \nabla} - I = \frac{\nabla}{I - \nabla}. \quad (5.4)$$

From equations (5.3) and (5.4)

$$\nabla = I - \frac{I}{I + \Delta} = \frac{\Delta}{1 + \Delta} \quad (5.5)$$

Theorem 5.2. $f_{x+nh} = \sum_{k=0}^{\infty} \binom{n}{k} \Delta^k f_x$

Proof: We shall prove our result by mathematical induction. For $n = 1$, the theorem reduces to $f_{x+h} = f_x + \Delta f_x$ which is true. Assume now that the theorem is true for $n - 1$. Then

$$f_{x+nh} = r^n f_x = E(E^{n-1} f_x) = E \sum_{i=0}^{\infty} \binom{n-1}{i} \Delta^i f_x \text{ by induction hypothesis.}$$

But $E = I + \Delta$. So

$$\begin{aligned} E^n f_x &= (I + \Delta)E^{n-1}f_x = I^{n-1}f_x + \Delta^{n-1}f_x \\ &= \sum_{i=0}^{\infty} \binom{n-1}{i} \Delta^i f_x + \sum_{i=0}^{\infty} \binom{n-1}{i} \Delta^{i+1} f_x \\ &= \sum_{i=0}^{\infty} \binom{n-1}{i} \Delta^i f_x + \sum_{j=1}^{\infty} \binom{n-1}{j-1} \Delta^j f_x \end{aligned}$$

The coefficient of $\Delta^k f_x$ ($k = 0, 1, 2, \dots, n$) is given by

$$\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}.$$

Hence,

$$r_{x+h} = r_x^n f_x = \sum_{k=0}^{\infty} \binom{n}{k} \Delta^k f_x,$$

which completes the proof of the theorem.

As a special case of this theorem, we get

$$r_x = r_x^n f_0 = \sum_{k=0}^{\infty} \binom{x}{k} \Delta^k f_0,$$

which is known as Newton's advancing difference formula and expresses the general functional value f_x in terms of f_0 and its differences.

Let h be the interval of differencing. Then the operator δ defined by

$$\delta_x = r_{x+\frac{h}{2}} - r_{x-\frac{h}{2}}$$

is called the central difference operator.

We note that

$$\delta_x = r_{x+\frac{h}{2}} - r_{x-\frac{h}{2}} = \omega^{\frac{1}{2}} r_x - \omega^{-\frac{1}{2}} f_x = \left(E^{\frac{1}{2}} - E^{-\frac{1}{2}} \right) f_x.$$

Hence,

$$\delta = E^{\frac{1}{2}} - E^{-\frac{1}{2}}. \quad (5.6)$$

Multiplying both sides by $E^{\frac{1}{2}}$, we get

$$E - \delta^{\frac{1}{2}} - 1 = 0 \text{ or } \left(E^{\frac{1}{2}} - \frac{\delta}{2} \right)^2 - \frac{\delta^2}{4} - 1 = 0$$

or

$$E^{\frac{1}{2}} - \frac{\delta}{2} = \sqrt{1 + \frac{\delta^2}{4}} \text{ or } E^{\frac{1}{2}} = \frac{\delta}{2} + \sqrt{1 + \frac{\delta^2}{4}}$$

or

$$E = \frac{\delta^2}{4} + 1 + \frac{\delta^2}{4} + \delta \left(1 + \frac{\delta^2}{4} \right)^{\frac{1}{2}} = I + \frac{\delta^2}{2} + \delta \sqrt{1 + \frac{\delta^2}{4}}. \quad (5.7)$$

Also, using equation (5.7), we note that

$$\Delta = E - I = \frac{\delta^2}{2} + \delta \sqrt{I + \frac{\delta^2}{4}} \quad (5.8)$$

$$\begin{aligned} \nabla &= I - \frac{E}{E} = I - \left(I + \frac{\delta^2}{2} + \delta \sqrt{I + \frac{\delta^2}{4}} \right)^{-1} \\ &= -\frac{\delta^2}{2} + \delta \sqrt{I + \frac{\delta^2}{4}}. \end{aligned} \quad (5.9)$$

Conversely,

$$\begin{aligned} \delta &= E^{\frac{1}{2}} - E^{-\frac{1}{2}} = (I + \Delta)^{\frac{1}{2}} - \frac{I}{(I + \Delta)^{1/2}} \\ &= \frac{I + \Delta - I}{\sqrt{I + \Delta}} = \frac{\Delta}{\sqrt{I + \Delta}} \end{aligned} \quad (5.10)$$

and

$$\delta = E^{\frac{1}{2}} - E^{-\frac{1}{2}} = \frac{I}{\sqrt{1 - \nabla}} - \sqrt{1 - \nabla} = \frac{\nabla}{\sqrt{I - \nabla}}. \quad (5.11)$$

Let h be the interval of differencing. Then the operator μ defined by

$$\mu f_x = \frac{1}{2} \left[f_{x+\frac{h}{2}} + f_{x-\frac{h}{2}} \right]$$

is called the mean value operator or averaging operator. We have

$$\mu f_x = \frac{1}{2} \left[f_{x+\frac{h}{2}} + f_{x-\frac{h}{2}} \right] = \frac{1}{2} \left[E^{\frac{1}{2}}|_{x-h} + E^{-\frac{1}{2}}|_x \right].$$

Hence,

$$\mu = \frac{1}{2} \left[E^{\frac{1}{2}} + E^{-\frac{1}{2}} \right] \quad (5.12)$$

or

$$2\mu = E^{\frac{1}{2}} + E^{-\frac{1}{2}}. \quad (5.13)$$

Also, we know that

$$\delta = E^{\frac{1}{2}} - E^{-\frac{1}{2}}. \quad (5.14)$$

Adding equations (5.13) and (5.14), we get

$$2\mu + \delta = 2E^{\frac{1}{2}} \text{ or } E^{\frac{1}{2}} = \mu + \frac{\delta}{2}. \quad (5.15)$$

Also,

$$E^{\frac{1}{2}} = \frac{\delta}{2} + \sqrt{\frac{\delta^2}{4} + \frac{\delta^2}{4}}$$

Hence,

$$\mu + \frac{\delta}{2} = \frac{\delta}{2} + \sqrt{I + \frac{\delta^2}{4}} \text{ or } \mu = \sqrt{I + \frac{\delta^2}{4}} \quad (5.16)$$

The relation equation (5.16) yields

$$I + \frac{\delta^2}{4} = \mu^2 \text{ or } \delta = 2\sqrt{\mu^2 - I}. \quad (5.17)$$

Multiplying equation (5.13) throughout by $E^{\frac{1}{2}}$, we get

$$E + I = 2\mu E^{\frac{1}{2}} \text{ or } E - 2\mu E^{\frac{1}{2}} + I = 0$$

or

$$\left(E^{\frac{1}{2}} - \mu\right)^2 - \mu^2 + I = 0 \text{ or } E^{\frac{1}{2}} - \mu = \sqrt{\mu^2 - I}$$

or

$$E^{\frac{1}{2}} = \mu + \sqrt{\mu^2 - I} \text{ or } E = 2\mu^2 - I + 2\mu\sqrt{\mu^2 - I}. \quad (5.18)$$

Then

$$\Delta = E - I = 2\mu^2 - 2I + 2\mu\sqrt{\mu^2 - I} \quad (5.19)$$

and

$$\begin{aligned} \nabla &= I - \frac{I}{E} = I - (2\mu^2 + 2\mu\sqrt{\mu^2 - I} - I)^{-1} \\ &= \frac{2\mu(\mu + \sqrt{\mu^2 - I}) - 2I}{2\mu^2 + 2\sqrt{\mu^2 - I} - I} \end{aligned} \quad (5.20)$$

The differential operator D is defined by

$$\Delta f(x) = f'(x).$$

By Taylor's Theorem, we have

$$\begin{aligned} f(x+h) &= f(x) + h f'(x) + \frac{h^2}{2!} f''(x) + \dots \\ &= f(x) + h Df(x) + \frac{h^2}{2!} D^2 f(x) + \dots \\ &= \left(1 + hD + \frac{h^2}{2!} D^2 + \dots\right) f(x) \end{aligned}$$

and so

$$Ef(x) = f(x+h) = \left(1 + hD + \frac{h^2}{2!} D^2 + \dots\right) f(x).$$

Hence,

$$E = 1 + hD + \frac{h^2}{2!} D^2 + \dots = e^{hD}, \text{ where } \omega = hD. \quad (5.21)$$

Then

$$\Delta = E - I = e^U - I \text{ and } \nabla = I - e^{-U}.$$

We note that

$$\begin{aligned}\delta &= E^{\frac{1}{2}} - E^{-\frac{1}{2}} = e^{\frac{U}{2}} - e^{-\frac{U}{2}} \\ &= 2 \sinh \frac{U}{2}\end{aligned}\quad (5.22)$$

$$\mu = \frac{1}{2} \left(E^{\frac{1}{2}} + E^{-\frac{1}{2}} \right) = \frac{1}{2} \left(e^{\frac{U}{2}} + e^{-\frac{U}{2}} \right). \quad (5.23)$$

Conversely

$$e^{\frac{U}{2}} + e^{-\frac{U}{2}} = 2\mu$$

or

$$e^U + 1 = 2\mu^{\frac{U}{2}} \left(\text{quadratic in } e^{\frac{U}{2}} \right)$$

or

$$e^{\frac{U}{2}} = \mu + \sqrt{\mu^2 - I}$$

or

$$U = \log \left| 2\mu^2 + I + 2\mu\sqrt{\mu^2 - I} \right|. \quad (5.24)$$

Since, by equation (5.22),

$$\delta = 2 \sinh \frac{U}{2},$$

it follows that

$$U = \sinh^{-1} \frac{\delta}{2}. \quad (5.25)$$

From the above discussion, we obtain the following table for the relations among the finite difference operators:

	Δ	∇	δ	E	$U = hD$
Δ	Δ	$(I - \nabla)^{-1} - I$	$\frac{\delta^2}{2} + \delta \sqrt{I + \frac{\delta^2}{4}}$	$E - I$	$e^U - I$
∇	$I - \frac{I}{\Delta + I}$	∇	$\frac{\delta^2}{2} + \delta \sqrt{I + \frac{\delta^2}{4}}$	$I - \frac{1}{E}$	$I - e^{-U}$
δ	$\frac{\Delta}{\sqrt{I + \Delta}}$	$\frac{\nabla}{\sqrt{I - \nabla}}$	δ	$E^{\frac{1}{2}} - E^{-\frac{1}{2}}$	$2 \sinh \frac{U}{2}$
E	$I + \Delta$	$\frac{I}{I - \nabla}$	$I + \frac{\delta^2}{2} + \delta \sqrt{I + \frac{\delta^2}{4}}$	E	e^U
$U = hD$	$\log(I + \Delta)$	$\log \frac{I}{I - \nabla}$	$2 \sinh \frac{\delta}{2}$	$\log E$	U

EXAMPLE 5.2

The expression δy_0 cannot be computed directly from a difference scheme. Find its value expressed in known central differences.

Solution. We know that

$$\mu = \left(I + \frac{\delta^2}{4} \right)^{\frac{1}{2}} \text{ or } \mu \left(I + \frac{\delta^2}{4} \right)^{-\frac{1}{2}} = I$$

or

$$\begin{aligned} y_0 &= u\delta \left(1 + \frac{\delta^2}{4} \right)^{-\frac{1}{2}} y_0 \\ &= \mu\delta \left[1 - \frac{\delta^2}{8} + \frac{-\frac{1}{2}\left(-\frac{1}{2}-1\right)}{2!} \frac{1}{16} + \frac{-\frac{1}{2}\left(-\frac{1}{2}-1\right)\left(-\frac{1}{2}-2\right)}{3!} \frac{1}{64} + \dots \right] y_0 \\ &= \mu\delta \left[y_0 - \frac{\delta^2 y_0}{8} + \frac{3}{128} \delta^4 y_0 - \dots \right]. \end{aligned} \quad (5.26)$$

But

$$\begin{aligned} \mu\delta y_0 &= \delta\mu y_0 = \delta \left(\frac{y_{\frac{1}{2}} + y_{-\frac{1}{2}}}{2} \right) = \frac{1}{2} \left(\delta v_{\frac{1}{2}} + \delta y_{-\frac{1}{2}} \right) \\ &= \frac{1}{2} (y_1 - y_0 - y_0 - y_{-1}) = \frac{1}{2} (y_1 - y_{-1}). \end{aligned}$$

Hence, equation (5.26) reduces to

$$\delta y_0 = \frac{1}{2} [y_1 - y_0] - \frac{1}{16} [\delta^2 y_1 - \delta^2 y_{-1}] + \frac{3}{256} [\delta^4 v_1 - \delta^4 y_{-1}] - \dots$$

which is the required form.

5.2 FACTORIAL NOTATION

A product of the form $x(x-1)(x-2)\dots(x-r+1)$ is called a factorial and is denoted by $[x]^r$. Thus,

$$\begin{aligned} [x] &= x \\ [x]^2 &= x(x-1) \\ [x]^3 &= x(x-1)(x-2) \\ &\dots \\ &\dots \\ [x]^r &= x(x-1)(x-2)\dots(x-r+1). \end{aligned}$$

If h is the interval of differencing, then

$$[x]^r = x(x-h)(x-2h)\dots(x-(r-1)h).$$

We observe that

$$\begin{aligned}\Delta[x]^n &= [x+h]^n - [x]^n \\&= (x+h)(x+h-h)(x+h-2h)\dots(x+h-(n-1)h) \\&\quad - x(x-h)(x-2h)\dots(x-(n-1)h) \\&= x(x-h)[x-(n-2)h][x+h-(x-nh+h)] = nh[x]^{n-1} \\ \Delta^2[x]^n &= \Delta(\Delta[x]^n) = \Delta(nh[x]^{n-1}) = nh^2[x]^{n-1} \\&= nh[(n-1)h[x]^{n-2}] = n(n-1)h^2[x]^{n-2} \\&\quad \dots \dots \dots \dots \\&\quad \dots \dots \dots \dots \\ \Delta^{n-1}[x]^n &= n(n-1)\dots2h^{n-1}x \\ \Delta^n[x]^n &= n(n-1)\dots2h^{n-1}x = n(n-1)\dots2h^{n-1}(x+h-x) \\&= n(n-1)\dots2h^n = n!h^n \\ \Delta^{n+1}[x]^n &= n!n^n - n\cdot h^n = 0.\end{aligned}$$

If interval of differentiating is 1, then

$$\Delta^n[x]^n = n! \text{ and } \Delta^{n+1}[x]^n = 0.$$

Thus for $h = 1$, differencing $[x]^n$ is analogous to that of differentiating x^n .

EXAMPLE 5.3

Express $f(x) = x^3 - 2x^2 + x - 1$ into factorial notation and show that $\Delta^4 f(x) = 0$.

Solution. Suppose

$$f(x) = x^3 - 2x^2 + x - 1 = x(x-1)(x-2) + Bx(x-1) + Cx + D.$$

Putting $x = 0$, we get $-1 = D$. Putting $x = 1$, we get $-1 = D + C$ and so $C = -1 - D = 0$. Putting $x = 2$, we get $1 = 2B + 2C + D$ and so $2B = 1 - 2C - D = 1 - (-1) = 2$ and so $B = 1$. Hence,

$$\begin{aligned}f(x) &= x^3 - 2x^2 + x - 1 = x(x-1)(x-2) + x(x-1) - 1 \\&= [x]^3 + [x]^2 - 1.\end{aligned}$$

Now,

$$\begin{aligned}\Delta f(x) &= 3[x]^2 + 2[x] \\ \Delta^2 f(x) &= 6[x] + 2 \\ \Delta^3 f(x) &= 6 \\ \Delta^4 f(x) &= 0.\end{aligned}$$

EXAMPLE 5.4

Find the function whose first difference is $2x^3 + 3x^2 - 5x + 4$.

Solution. Let $f(x)$ be the required function. We are given that

$$\begin{aligned}\Delta f(x) &= 2x^3 + 3x^2 - 5x + 4 \\ &= 2x(x-1)(x-2) + Bx(x-1) + Cx + D.\end{aligned}$$

Putting $x = 0$, we have $4 = D$. Putting $x = 1$, we get $4 = C + D$ and so $C = 0$. Putting $x = 2$, we get $22 = 2B + 2C + D$ and so $2B = 22 - 2C - D = 22 - 4 = 18$ and so $B = 9$. Thus,

$$\begin{aligned}\Delta f(x) &= 2x(x-1)(x-2) + 9x(x-1) + 4 \\ &= 2[x]^3 + 9[x]^2 + 4.\end{aligned}$$

Integrating $\Delta f(x)$, we get

$$f(x) = \frac{2[x]^4}{4} + \frac{9[x]^3}{3} 4[x] + C = \frac{1}{2}[x]^4 + 3[x]^3 + 4[x] + C,$$

where C is constant of integration.

5.3 SOME MORE EXAMPLES OF FINITE DIFFERENCES**EXAMPLE 5.5**

Find the missing term in the following table:

x	0	1	2	3	4
$f(x)$	1	3	9	—	81

Solution. Since four entries y_0, y_1, y_2, y_3, y_4 are given, the given function can be represented by a third degree polynomial. The difference table is

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$	$\Delta^4 f(x)$
0	1	2			
1	3	6	4	$y_3 - 19$	
2	9	$y_3 - 9$	$y_3 - 15$	$105 - 3y_3$	$124 - 4y_3$
3	y_3	$81 - y_3$	$90 - 2y_3$		
4	81				

Since polynomial is of degree 3, $\Delta^4 f(x) = 0$ and so $124 - 4y_3 = 0$ and hence $y_3 = 31$.

EXAMPLE 5.6

If $y_0 = 3, y_1 = 12, y_2 = 81, y_3 = 2000$, and $y_4 = 100$, determine $\Delta^4 y_0$.

Solution. The difference table for the given data is

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
0	3				
1	12	9	60	1790	
2	81	69	1850	-5669	-7459
3	2000	1919	-3819		
4	100	-1900			

From the table, we have $\Delta^4 y_0 = -7459$.

EXAMPLE 5.7

Establish the relations

$$(i) \quad \Delta \nabla = \nabla \Delta = \Delta - \nabla = \delta^2;$$

$$(ii) \quad \mu \delta = \frac{1}{2}(\Delta + \nabla);$$

$$(iii) \quad \Delta = E \nabla = \nabla E = \delta E^{\frac{1}{2}}.$$

Solution. (i) We know that

$$\Delta = E - I \text{ and } \nabla = I - \frac{I}{E}$$

Therefore,

$$\Delta \nabla = (E - I) \left(I - \frac{I}{E} \right) = E + \frac{I}{E} - 2I = \nabla \Delta \quad (5.27)$$

and

$$\Delta - \nabla = E + \frac{I}{E} - 2I \quad (5.28)$$

Furthermore,

$$\delta = E^{\frac{1}{2}} - E^{-\frac{1}{2}} = E^{\frac{1}{2}} - \frac{I}{E^{1/2}}$$

and so

$$\delta^2 = E + \frac{I}{E} - 2I. \quad (5.29)$$

The result follows from equations (5.27), (5.28), and (5.29).

(ii) We have

$$\mu = \frac{1}{2}(E^{1/2} + E^{-1/2}), \quad \zeta = E^{1/2} - E^{-1/2}.$$

Therefore,

$$\begin{aligned}\mu\delta &= \frac{1}{2}(E^{1/2} + E^{-1/2})(E^{1/2} - E^{-1/2}) \\ &= \frac{1}{2}(E - E^{-1}) = \frac{1}{2}\left(E - \frac{I}{E}\right) \\ &= \frac{1}{2}\left(E - I + I - \frac{I}{E}\right) = \frac{1}{2}(\Delta + \nabla).\end{aligned}$$

(iii) We have

$$\begin{aligned}E\nabla &= E\left(I - \frac{I}{E}\right) = r - I = \Delta \\ \nabla E &= \left(I - \frac{1}{E}\right)E = r - I = \Delta \\ \delta E^{1/2} &= (E^{1/2} - E^{-1/2})E^{1/2} = E - I = \Delta.\end{aligned}$$

Hence,

$$E\nabla = \nabla E = \delta E^{\frac{1}{2}} = \Delta.$$

EXAMPLE 5.8

Show that

$$E^r = \frac{E^{r+1} - r^{-t}}{E - E^{-1}} = \frac{\sinh 2r\theta}{\sinh 2\theta} E + \frac{\sinh 2t\theta}{\sinh 2\theta},$$

where $t = 1-r$ and $\theta = \frac{hD}{2}$.

Solution. We have

$$\frac{E^{r+1} - r^{-t}}{E - E^{-1}} = \frac{E^{r+1} - r^{r-1}}{E - E^{-1}} = E^r \left(\frac{E - E^{-1}}{E - E^{-1}} \right) = E^r.$$

Also,

$$E - r^{-1} = e^{hD} - e^{-hD} = 2 \sinh 2\theta.$$

Therefore,

$$\begin{aligned}\frac{E^{r+1} - r^{-t}}{E - E^{-1}} &= \frac{\frac{1}{2}(E^{r+1} - r^{r-1})}{\frac{1}{2}(E - E^{-1})} = \frac{\frac{1}{2}[EE^r - E^{r-1}]}{\sinh 2\theta} \\ &= \frac{\frac{1}{2}[E(E^r - E^{-r}) + E^{-(r-1)} - E^{r-1}]}{\sinh 2\theta} \\ &= \frac{\frac{1}{2}[E(E^r - E^{-r})] - \frac{1}{2}(E^{1-r} - E^{r-1})}{\sinh 2\theta}\end{aligned}$$

$$\begin{aligned}
&= \frac{\frac{1}{2}[E(E^r - E^{-r})] - \frac{1}{2}(e^{-1-r} - e^{-(1-r)})}{\sinh 2\theta} \\
&= \frac{E \sinh 2r\theta + \sinh 2t\theta}{\sinh 2\theta} \\
&= \frac{\sinh 2r\theta}{\sinh 2\theta} E + \frac{\sinh 2t\theta}{\sinh 2\theta}.
\end{aligned}$$

EXAMPLE 5.9

Show that

$$\sum_{k=0}^{n-1} \Delta^2 f_k = \Delta f_n - \Delta f_0.$$

Solution. We have

$$\begin{aligned}
\Delta^2 f_0 &= \Delta(\Delta f_0) = \Delta(f_1 - f_0) = f_1 - f_0 \\
\Delta^2 f_1 &= \Delta(\Delta f_1) = \Delta(f_2 - f_1) = \Delta f_2 - \Delta f_1 \\
&\dots \\
&\dots \\
\Delta^2 f_{n-1} &= \Delta(\Delta f_{n-1}) = \Delta(f_n - f_{n-1}) = \Delta f_n - \Delta f_{n-1}.
\end{aligned}$$

Adding we get

$$\begin{aligned}
\sum_{k=0}^{n-1} \Delta^2 f_k &= (\Delta f_1 - \Delta f_0) + (\Delta f_2 - \Delta f_1) + \dots + (\Delta f_n - \Delta f_{n-1}) \\
&= f_n - \Delta f_0.
\end{aligned}$$

EXAMPLE 5.10

Show that $\sum_{k=0}^{n-1} \delta^2 f_{2k+1} = \tanh\left(\frac{U}{2}\right)(f_{2n} - f_0)$.

Solution. We have

$$\begin{aligned}
\delta^2 &= (E^{\frac{1}{2}} - E^{-\frac{1}{2}})^2 = \frac{(E^{\frac{1}{2}} - E^{-\frac{1}{2}})^2 (E^{\frac{1}{2}} + E^{-\frac{1}{2}})}{E^{\frac{1}{2}} + E^{-\frac{1}{2}}} = \frac{(E^{\frac{1}{2}} - E^{-\frac{1}{2}})(E - E^{-1})}{E^{\frac{1}{2}} + E^{-\frac{1}{2}}} \\
&= \frac{(e^{\frac{U}{2}} - e^{-\frac{U}{2}})(E - E^{-1})}{e^{\frac{U}{2}} + e^{-\frac{U}{2}}} = \tanh\left(\frac{U}{2}\right)(E - E^{-1}).
\end{aligned}$$

Thus,

$$\delta^2 f_{2k+1} = \tanh\left(\frac{U}{2}\right)[Ef_{2k+1} - E^{-1}f_{2k+1}] = \tanh\left(\frac{U}{2}\right)[f_{2k+2} - f_{2k}].$$

Therefore,

$$\sum_{k=0}^{n-1} \delta^2 f_{k+1} = \tanh\left(\frac{U}{2}\right) [(f_2 - f_1) + (f_4 - f_2) + \dots + (f_{2n} - f_{2n-2})] = \tanh\left(\frac{U}{2}\right) [f_{2n} - f_0].$$

EXAMPLE 5.11

Find the cubic polynomial $f(x)$ which takes on the values $f_0 = -5$, $f_1 = 1$, $f_2 = 9$, $f_3 = 25$, $f_4 = 55$, $f_5 = 105$.

Solution. The difference table for the given function is given below:

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$	$\Delta^4 f(x)$
0	-5				
1	1	6			
2	9	8	2		
3	25	16	8	6	
4	55	30	14	6	
5	105	50	20		

Now,

$$\begin{aligned}
 {}_x = r^x J_0 &= (I + \Delta)^x f_0 \\
 &= \left[1 + x\Delta + \frac{x(x-1)}{2!} \Delta^2 + \frac{x(x-1)(x-2)}{3!} \Delta^3 \right] f_0 \\
 &= f_0 + x\Delta f_0 + \frac{x^2 - x}{2} \Delta^2 f_0 + \frac{x^3 - 3x^2 + 2x}{6} \Delta^3 f_0 \\
 &= -5 + 6x + \frac{x^2 - x}{2} (2) + \frac{x^3 - 3x^2 + 2x}{6} (6) = x^3 - 2x^2 + 7x - 5,
 \end{aligned}$$

which is the required cubic polynomial.

EXAMPLE 5.12

Determine

$$\Delta^{10}[(1-ax)(1-bx^2)(1-cx^3)(1-dx^4)].$$

Solution. We have

$$\begin{aligned}
 &\Delta^{10}[(1-ax)(1-bx^2)(1-cx^3)(1-dx^4)] \\
 &= \Delta^{10}[abcd x^{10} + Ax^9 + Bx^8 + \dots + 1].
 \end{aligned}$$

The polynomial in the square bracket is of degree 10. Therefore, $\Delta^{10}f(x)$ is constant and is equal to $a_n n! h^n$. In this case, we have $a_n = abcd$, $n = 10$, $h = 1$. Hence,

$$\Delta^{10}f(x) = abcd(10)!.$$

EXAMPLE 5.13

Show that

$$\delta^2 y_5 = y_1 - 2y_5 + y_4.$$

Solution. We know that

$$\delta^2 = \Delta - \nabla.$$

Therefore,

$$\begin{aligned}\delta^2 y_5 &= (\Delta - \nabla) y_5 = \Delta y_5 - \nabla y_5 = v_6 - y_5 - (v_5 - y_4) \\ &= v_6 - 2y_5 + y_4.\end{aligned}$$

EXAMPLE 5.14

Show that

$$e^x = \left(\frac{\Delta^2}{E} \right) e^x \cdot \frac{Ee^x}{\Delta^2 e^x},$$

the interval of differencing being h .

Solution. We note that

$$\begin{aligned}Ee^x &= e^{x+h}, \quad \Delta e^x = e^{x+h} - e^x = e^x(e^h - 1) \\ \Delta^2 e^x &= e^x(e^h - 1)^2\end{aligned}$$

and

$$\left(\frac{\Delta^2}{E} \right) e^x = \Delta^2 E^{-1}(e^x) = \Delta^2 e^{x-h} = e^{-h} \Delta^2 e^x = e^{-h} e^x (e^h - 1)^2.$$

Hence,

$$\left(\frac{\Delta^2}{E} \right) e^x \cdot \frac{Ee^x}{\Delta^2 e^x} = e^{-h} e^x (e^h - 1)^2 \frac{e^{x+h}}{e^x (e^h - 1)^2} = e^x.$$

EXAMPLE 5.15

Show that

$$(i) \quad \delta^n y_x = \sum_{k=0}^n (-1)^k \frac{n!}{k!(n-k)!} y_{x+\frac{n}{2}-k}$$

$$(ii) \quad \Delta \binom{n}{i+1} = \binom{n}{i}.$$

Solution. (i) We have

$$\delta^n = \left(E^{\frac{1}{2}} - E^{-\frac{1}{2}} \right)^n = E^{-\frac{n}{2}} (E - I)^n$$

Therefore,

$$\delta^n y_x = (E - I)^n y_{x-\frac{n}{2}} = [E^n \binom{n}{1} E^{n-1} + \binom{n}{2} E^{n-2} - \dots + (-1)^n] y_{x-\frac{n}{2}}$$

$$\begin{aligned}
 &= \sum_{k=0}^n (-1)^k \frac{n!}{k!(n-k)!} E^{n-k} y_{x-\frac{n}{2}} \\
 &= \sum_{k=0}^n (-1)^k \frac{n!}{k!(n-k)!} y_{x+n-k-\frac{n}{2}} \\
 &= \sum_{k=0}^n (-1)^k \frac{n!}{k!(n-k)!} y_{x+\frac{n}{2}-k}.
 \end{aligned}$$

(ii) We have $\binom{n}{i+1} = \frac{n!}{(i+1)!(n-i-1)!}$.

Now,

$$\Delta \binom{n}{i+1} = \binom{n+1}{i+1} - \binom{n}{i+1} = \frac{n!(i+1)}{(i+1)!(n-i)!} = \frac{n!}{i!(n-1)!} = \binom{n}{i}.$$

EXAMPLE 5.16

Assuming that the following values of y belong to a polynomial of degree 4, find the missing values in the table:

x	0	1	2	3	4	5	6	7
y	1	-1	1	-1	1	-	-	-

Solution. The difference table of the given data is shown below:

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
0	1	-2			
1	-1	2	4	-8	
2	1	-2	-4	8	16
3	-1	2	4	$\Delta^3 y_2$	$\Delta^3 y_2 - 8$
4	1	Δy_4	$\Delta^2 y_3$	$\Delta^3 y_2$	$\Delta^3 y_3 - \Delta^3 y_2$
5	y_5	Δy_5	$\Delta^2 y_4$	$\Delta^3 y_3$	$\Delta^3 y_4 - \Delta^3 y_3$
6	y_6	Δy_6	$\Delta^2 y_5$	$\Delta^3 y_4$	
7	y_7				

Since the polynomial of the data is of degree 4, $\Delta^4 y$ should be constant. One of $\Delta^4 y$ is 16. Hence, all of the fourth differences must be 16. But then

$$\Delta^3 y_2 - 8 = 16 \text{ giving } \Delta^3 y_2 = 24$$

$$\Delta^3 y_3 - \Delta^3 y_2 = 16 \text{ giving } \Delta^3 y_3 = 40$$

$$\begin{aligned}
 \Delta^3 y_4 - \Delta^3 y_3 &= 16 \text{ giving } \Delta^3 y_4 = 56 \\
 \Delta^2 y_3 - 4 &= \Delta^3 y_2 = 24 \text{ and so } \Delta^2 y_3 = 28 \\
 \Delta^2 y_4 - \Delta^2 y_3 &= \Delta^3 y_3 = 40 \text{ and so } \Delta^2 y_4 = 68 \\
 \Delta^2 y_5 - \Delta^2 y_4 &= \Delta^3 y_4 = 56 \text{ and so } \Delta^2 y_5 = 124 \\
 y_4 - 2 &= \Delta^2 y_3 = 28 \text{ and so } \Delta y_4 = 30 \\
 \Delta y_5 - \Delta y_4 &= \Delta^2 y_4 = 68 \text{ and so } \Delta y_5 = 98 \\
 \Delta y_6 - \Delta y_5 &= \Delta^2 y_5 = 124 \text{ and so } \Delta y_6 = 222 \\
 y_5 - 1 &= \Delta y_4 = 30 \text{ which gives } y_5 = 31 \\
 y_6 - y_5 &= \Delta y_5 = 98 \text{ which gives } y_6 = 129 \\
 y_7 - y_6 &= \Delta y_6 = 222 \text{ which yields } y_7 = 351.
 \end{aligned}$$

Hence,

Hence, the missing terms are

$$y_5 = 31, y_6 = 129, y_7 = 351.$$

5.4 ERROR PROPAGATION

Let $y_0, y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8$ be the values of the function at the arguments $x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$, respectively. Suppose an error ε is committed in y_4 during tabulation. To study the error propagation, we use the difference table. For the sake of convenience, we construct difference table up to fourth difference only. If the error in y_4 is ε , then the value of the function at x_4 is $y_4 + \varepsilon$. The difference table of the data is as shown below.

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
x_0	y_0	Δy_0			
x_1	y_1	Δy_1	$\Delta^2 y_0$	$\Delta^3 y_0$	
x_2	y_2	Δy_2	$\Delta^2 y_1$	$\Delta^3 y_1 + \varepsilon$	$\Delta^4 y_0 + \varepsilon$
x_3	y_3	$\Delta y_3 + \varepsilon$	$\Delta^2 y_2 + \varepsilon$	$\Delta^3 y_2 - 3\varepsilon$	$\Delta^4 y_1 - 4\varepsilon$
x_4	$y_4 + \varepsilon$	$\Delta y_4 - \varepsilon$	$\Delta^2 y_3 - 2\varepsilon$	$\Delta^3 y_3 + 3\varepsilon$	$\Delta^4 y_2 + 6\varepsilon$
x_5	y_5	Δy_5	$\Delta^2 y_4 + \varepsilon$	$\Delta^3 y_4 - \varepsilon$	$\Delta^4 y_3 - 4\varepsilon$
x_6	y_6	Δy_6	$\Delta^2 y_5$	$\Delta^3 y_5$	$\Delta^4 y_4 + \varepsilon$
x_7	y_7	Δy_7	$\Delta^2 y_6$		
x_8	y_8				

We note that

- (i) Error propagates in a triangular pattern (shown by fan lines) and grows quickly with the order of difference.
- (ii) The coefficients of the error ε in any column are the binomial coefficients of $(1 - \varepsilon)^n$ with alternating signs. Thus, the errors in the third column are $\varepsilon, -3\varepsilon, 3\varepsilon, -\varepsilon$.
- (iii) The algebraic sum of the errors in any difference column is zero.
- (iv) If the difference table has even differences, then the maximum error lies on the same horizontal line on which the tabular value in error lies.

EXAMPLE 5.17

One entry in the following table of a polynomial of degree 4 is incorrect. Correct the entry by locating it

x	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
y	1.0000	1.5191	2.0736	2.6611	3.2816	3.9375	4.6363	5.3771	6.1776	7.0471	8.0

Solution. The difference table for the given data is shown below. Since the degree of the polynomial is four, the fourth difference must be constant. But we note that the fourth differences are oscillating for the larger values of x . The largest numerical fourth difference 0.0186 is at $x = 1.6$. This suggests that the error in the value of f is at $x = 1.6$. Draw the fan lines as shown in the difference table.

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
1.0	1.0000				
1.1	1.5191	0.5191			
1.2	2.0736	0.5545	0.0354	-0.0024	0.0024
1.3	2.6611	0.5875	0.0330	0	0.0024
1.4	3.2816	0.6205	0.0330	0.0024	0.0051 Fan line
1.5	3.9375	0.6559	0.0354	0.0075	-0.0084
1.6	4.6363	0.6988	0.0429	-0.0009	0.0186
1.7	5.3771	0.7408	0.0420	0.0177	-0.0084
1.8	6.1776	0.8005	0.0597	0.0093	0.0051
1.9	7.0471	0.8695	0.0690	0.0144	
2.0	8.0000	0.9529	0.0834		

Then taking 1.6 as x_0 , we have

$$\Delta^4 f_{-4} + \varepsilon = 0.0051$$

$$\Delta^4 f_{-3} - 4\varepsilon = -0.0084$$

$$\Delta^4 f_{-2} + 6\varepsilon = 0.0186$$

$$\Delta^4 f_{-1} - 4\varepsilon = -0.0084$$

$$\Delta^4 f_0 + \varepsilon = 0.0051$$

We want all fourth differences to be alike. Eliminating Δ^4 between any two of the compatible equations and solving for ε will serve our purpose. For example, subtracting the second equation from the first, we get

$$5\varepsilon = 0.0135 \text{ and so } \varepsilon = 0.0027.$$

Putting this value of ε in the above equations, we note that all the fourth differences become 0.0024. Further,

$$f(1.6) + \varepsilon = 4.6363,$$

which yields

$$f(1.6) = 4.6363 - \varepsilon = 4.6363 - 0.0027 = 4.6336.$$

Thus, the error was a transposing error, that is, writing 63 instead of 36 while tabulation.

EXAMPLE 5.18

Find and correct the error, by means of differences, in the data:

x	0	1	2	3	4	5	6	7	8	9	10
y	2	5	8	17	38	75	140	233	362	533	752

Solution. The difference table for the given data is shown below. The largest numerical fourth difference -12 is at $x = 5$. So there is some error in the value $f(5)$. The fan lines are drawn and we note from the table that

$$\Delta^4 f_{-4} + \varepsilon = -2$$

$$\Delta^4 f_{-3} - 4\varepsilon = 8$$

$$\Delta^4 f_{-2} + 6\varepsilon = -12$$

$$\Delta^4 f_{-1} - 4\epsilon = 8$$

$$\Delta^4 f_0 + \epsilon = -2$$

and

$$\Delta^3 f_{-3} + \epsilon = 4$$

$$\Delta^3 f_{-2} - 3\epsilon = 12$$

$$\Delta^3 f_{-1} + 3\epsilon = 0$$

$$\Delta^3 f_0 - \epsilon = 8.$$

Subtracting second equation from the first (for both sets shown above), we get

$5\epsilon = -10$ (for the first set) and $4\epsilon = -8$ (for the second set).

Hence, $\epsilon = -2$.

Difference table

x	y	Δ	Δ^2	Δ^3	Δ^4
0	2	3			
1	5	3	0	6	
2	8	9	6	6	0
3	17	21	12		-2 Fan line
4	38	16		4	8
5	75	37	28	12	-12
6	140	65	28	0	8
7	233	93	36	6	-2
8	362	129	42	6	0
9	533	171	48		
10	752	219			

We now have

$$f(5) + \epsilon = 75 \text{ and so } f(5) = 75 - \epsilon = 75 - (-2) = 77.$$

Therefore, the true value of $f(5)$ is 77.

5.5 NUMERICAL UNSTABILITY

Subtraction of two nearly equal numbers causes a considerable loss of significant digits and may magnify the error in the later calculations. For example, if we subtract 63.994 from 64.395, which are correct to five significant figures, their difference 0.401 is correct only to three significant figures.

A similar loss of significant figures occurs when a number is divided by a small divisor. For example, we consider

$$f(x) = \frac{1}{1-x^2}, x = 0.9.$$

Then true value of $f(0.9)$ is 0.526316×10 . If x is approximated to $x^* = 0.900005$, that is, if some error appears in the sixth figure, then $f(x^*) = 0.526341 \times 10$. Thus, an error in the sixth place has caused an error in the fifth place in $f(x)$.

We note therefore that every arithmetic operation performed during computation gives rise to some error, which once generated may decay or grow in subsequent calculations. In some cases, error may grow so large as to make the computed result totally redundant. We call such a process (procedure) numerically unstable.

Adopting the calculation procedure that avoids subtraction of nearly equal numbers or division by small numbers or retaining more digits in the mantissa may avoid numerical instability.

EXAMPLE 5.19

(Wilkinson): consider the polynomial

$$\begin{aligned} P_{20}(x) &= (x-1)(x-2)(x-20) \\ &\quad - x^{20} - 210^{-19} + +(20)! \end{aligned}$$

The zeros of this polynomial are 1, 2,..., 20. Let the coefficient of x^{19} be changed from 210 to $(210+2^{-23})$. This is a very small absolute change of magnitude 10^{-7} approximately. Most computers, generally, neglect this small change which occurs after 23 binary bits. But we note that smaller zeros of the new polynomial are obtained with good efficiency while the large roots are changed by a large amount. The largest change occurs in the roots 16 and 17. For example, against 16, we get $16.73\dots \pm i2.81$ where magnitude is 17 approximately.

5.6 INTERPOLATION

Interpolation is the process of finding the value of a function for any value of argument (independent variable) within an interval for which some values are given.

Thus, interpolation is the art of reading between the lines in a given table.

Extrapolation is the process of finding the value of a function outside an interval for which some values are given.

We now discuss interpolation processes for equal spacing.

(A) Newton's Forward Difference Formula

Let $\dots, f_{-2}, f_{-1}, f_0, f_1, f_2, \dots$ be the values of a function for $\dots, x_0 - 2h, x_0 - h, x_0, x_0 + h, x_0 + 2h, \dots$ Suppose that we want to compute the function value f_p for $x = x_0 + ph$, where in general $-1 < p < 1$. We have

$$f_p = f(x_0 + ph) \text{ and } p = \frac{x - x_0}{h},$$

where h is the interval of differencing. Then using shift operator and Binomial Theorem, we have

$$\begin{aligned} {}_x = r \cdot J_0 &= (I + \Delta)^p f_0 \\ &= \left[I + p\Delta + \frac{p(p-1)}{2!} \Delta^2 + \frac{(p(p-1)(p-2))}{3!} \Delta^3 + \dots \right] f_0 \\ &= f_0 + \binom{p}{1} \Delta f_0 + \binom{p}{2} \Delta^2 f_0 + \binom{p}{3} \Delta^3 f_0 + \dots \end{aligned} \quad (5.30)$$

The expression (5.30) is called Newton's forward difference formula for interpolation.

(B) Newton's Backward Difference Formula

Let $\dots, f_{-2}, f_{-1}, f_0, f_1, f_2, \dots$ be the values of a function for $\dots, x_0 - 2h, x_0 - h, x_0, x_0 + h, x_0 + 2h, \dots$. Suppose that we want to compute the function value f_p for $x = x_0 + ph$, $-1 < p < 1$. We have

$${}_p = t(x_0 + ph), \quad t = \frac{x - x_0}{h}.$$

Using Newton's backward differences, we have

$$\begin{aligned} {}_x = r \cdot J_0 &= (I - \nabla)^{-p} f_0 \\ &= \left[I + p\nabla + \frac{p(p+1)}{2!} \nabla^2 + \frac{p(p+1)(p+2)}{3!} \nabla^3 + \dots \right] f_0 \\ &= f_0 + p \cdot f_0 + \frac{p(p+1)}{2!} \nabla^2 f_0 + \frac{p(p+1)(p+2)}{3!} \nabla^3 f_0 + \dots, \end{aligned}$$

which is known as Newton's backward difference formula for interpolation.

Remark 5.1. It is clear from the differences used that

- (i) Newton's forward difference formula is used for interpolating the values of the function near the beginning of a set of tabulated values.
- (ii) Newton's backward difference formula is used for interpolating the values of the function near the end of a set of tabulated values.

EXAMPLE 5.20

Calculate approximate value of $\sin x$ for $x = 0.54$ and $x = 1.36$ using the following table:

x	0.5	0.7	0.9	1.1	1.3	1.5
$\sin x$	0.47943	0.64422	0.78333	0.89121	0.96356	0.99749

Solution.

We take

$$x_0 = 0.50, \quad x_p = 0.54, \quad \text{and} \quad p = \frac{0.54 - 0.50}{0.2} = 0.2.$$

Using Newton's forward difference method, we have

$$\begin{aligned}
{}_p &= t_0 + p\Delta f_0 + \frac{p(p-1)}{2!} \Delta^2 f_0 + \frac{p(p-1)(p-2)}{3!} \Delta^3 f_0 \\
&\quad + \frac{p(p-1)(p-2)(p-3)}{4!} \Delta^4 f_0 + \frac{p(p-1)(p-2)(p-3)(p-4)}{5!} \Delta^5 f_0 \\
&= 0.47943 + 0.2(0.16479) + \frac{0.2(0.2-1)}{2}(0.0268) \\
&\quad + \frac{0.2(0.2-1)(0.2-2)}{6}(-0.00555) + \frac{0.2(0.2-1)(0.2-2)(0.2-3)}{4!}(0.00125) \\
&\quad + \frac{0.2(0.2-1)(0.2-2)(0.2-3)(0.2-4)}{5!}(0.00016) \approx 0.51386.
\end{aligned}$$

Difference table

x	$\sin x$	1st difference	2nd difference	3rd difference	4th difference	5th difference
0.5	0.47943					
0.7	0.64422	0.16479				
0.9	0.78333	0.13911	-0.02568			
1.1	0.89121	0.10788	-0.03123	-0.00555		
1.3	0.96356	0.07235	-0.03553	-0.00430	0.00125	
1.5	0.99749	0.03393	-0.03842	-0.00289	0.00141	0.00016

Further, the point $x = 1.36$ lies toward the end of the tabulated values. Therefore, to find the value of the function at $x = 1.36$, we use Newton's backward differences method. We have

$$\begin{aligned}
x_p &= 1.36, \quad x_0 = 1.3, \quad \text{and} \quad p = \frac{1.36 - 1.30}{0.2} = 0.3, \\
{}_p &= t_0 + {}_p \nabla f_0 + \frac{p(p+1)}{2!} \nabla^2 f_0 + \frac{p(p+1)(p+2)}{3!} \nabla^3 f_0 + \frac{p(p+1)(p+2)(p+3)}{4!} \nabla^4 f_0 \\
&= 0.96356 + 0.3(0.07235) + \frac{0.3(0.3+1)}{2}(0.03553) \\
&\quad + \frac{0.3(0.3+1)(0.3+2)}{6}(-0.00430) + \frac{0.3(0.3+1)(0.3+2)(0.3+3)}{24}(0.00125) \\
&= 0.96356 + 0.021705 - 0.006128 - 0.000642 + 0.000154 \approx 0.977849.
\end{aligned}$$

EXAMPLE 5.21

Find the cubic polynomial $f(x)$ which takes on the values $f(0) = -4, f(1) = -1, f(2) = 2, f(3) = 11, f(4) = 32, f(5)$. Find $f(6)$ and $f(2.5)$.

Solution. The difference table for the given data is

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
0	-4			
1	-1	3	0	
2	2	3	6	6
3	11	9	12	6
4	32	21	18	
5	71	39		

Using Newton's forward difference formula, we have

$$\begin{aligned}
 f(x) &= f_0 + x\Delta f_0 + \frac{x(x-1)}{2!}\Delta^2 f_0 + \frac{x(x-1)(x-2)}{3!}\Delta^3 f_0 \\
 &= -4 + x(3) + \frac{x^2 - x}{2}(0) + \frac{x^3 - 3x^2 + 2x}{6}(6) \\
 &= x^3 - 3x^2 + 2x + 3x - 4 \\
 &= x^3 - 3x^2 + 5x - 4,
 \end{aligned}$$

which is the required cubic polynomial. Therefore,

$$\begin{aligned}
 f(6) &= 6^3 - 3(6^2) + 5(6) - 4 \\
 &= 216 - 108 + 30 - 4 = 134.
 \end{aligned}$$

On the other hand, if we calculate $f(6)$ using Newton's forward difference formula, then take $x_0 = 0$, $p = \frac{x - x_0}{h} = \frac{6 - 0}{1} = 6$ and have

$$\begin{aligned}
 f(6) &= f_6 = f_0 + 6\Delta f_0 + \frac{(6)(5)}{2}\Delta^2 f_0 + \frac{(6)(5)(4)}{6}\Delta^3 f_0 \\
 &= 4 + 6(3) + 15(0) + 20(6) = 134 \text{ (exact value of } f(6)).
 \end{aligned}$$

Again taking $x_0 = 2$, we have $p = \frac{x - x_0}{h} = \frac{2.5 - 2.0}{0.5} = 0.5$. Therefore,

$$\begin{aligned}
 f(2.5) &= f_0 + pf_0 + \frac{p(p-1)}{2!}\Delta^2 f_0 + \frac{p(p-1)(p-2)}{6}\Delta^3 f_0 \\
 &= 2 + 0.5(9) + \frac{(0.5)(0.5-1)}{2}(12) + \frac{0.5(0.5-1)(0.5-2)}{6}(6) \\
 &= 2 + 4.50 - 1.50 + 0.375 \\
 &= 6.875 - 1.500 = 5.375 \text{ (exact value of } f(2.5)).
 \end{aligned}$$

EXAMPLE 5.22

Find a cubic polynomial in x for the following data:

x	0	1	2	3	4	5
y	-3	3	11	27	57	107

Solution. The difference table for the given data is

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
0	-3	6		
1	3	8	2	6
2	11	16	8	6
3	27	30	14	6
4	57	50	20	
5	107			

Using Newton's forward difference formula, we have

$$\begin{aligned}
 {}_x &= t_0 + x\Delta f_r + \frac{x(x-1)}{2!} \Delta^2 f_0 + \frac{x(x-1)(x-2)}{3!} \Delta^3 f_0 \\
 &= 3 + 6x + \frac{x^2 - x}{2}(2) + \frac{x^3 - 3x^2 + 2x}{6}(6) \\
 &= x^3 - 3x^2 + 2x + x^2 - x + 6x - 3 \\
 &= x^3 - 2x^2 + 7x - 3.
 \end{aligned}$$

EXAMPLE 5.23

The area A of a circle of diameter d is given for the following values:

d	80	85	90	95	100
A	5026	5674	6362	7088	7854

Calculate the area of a circle of diameter 105.

Solution. The difference table for the given data is

d	A				
80	5026	648			
85	5674	688	40	-12	
90	6362	716	28	22	32
95	7088	766	50		
100	7854				

Letting $x_p = 105$, $x_0 = 100$, and $p = \frac{105 - 100}{5} = 1$, we shall use Newton's backward difference formula

$$p = t_0 + r \nabla f_0 + \frac{p(p+1)}{2} \nabla^2 f_0 + \frac{p(p+1)(p+2)}{3!} \nabla^3 f_0 + \frac{p(p+1)(p+2)(p+3)}{4!} \nabla^4 f_0.$$

Therefore,

$$f(105) = 7854 + 766 + 50 + 22 + 32 = 8724.$$

Remark 5.2. We note (in the above example) that if a tabulated function is a polynomial, then interpolation and extrapolation would give exact values.

(C) Central Difference Formula

Let ..., $f_{-2}, f_{-1}, f_0, f_1, f_2, \dots$ be the values of a function f for ..., $x_0 - 2h, x_0 - h, x_0, x_0 + h, x_0 + 2h, \dots$. Then

$$\delta = E^{\frac{1}{2}} - E^{-\frac{1}{2}}$$

and so

$$\delta E^{\frac{1}{2}} = r - I = \Delta, \quad \Delta^2 = \delta E \text{ and } \Delta^3 = \delta E^{\frac{3}{2}}.$$

Thus,

$$\Delta_{-2} = \delta E^{\frac{1}{2}} r_{-2} = \delta f_{-\frac{3}{2}}$$

$$\Delta_{-1} = \delta E^{\frac{1}{2}} r_{-1} = \delta f_{-\frac{1}{2}}$$

$$\Delta_0 = \delta E^{\frac{1}{2}} r_0 = \delta f_{\frac{1}{2}}$$

$$\Delta_1 = \delta E^{\frac{1}{2}} r_1 = \delta f_{\frac{3}{2}}$$

and so on. Similarly,

$$\Delta^2 f_{-2} = \delta^2 E f_{-2} = \delta^2 f_{-1}$$

$$\Delta^2 f_{-1} = \delta^2 E f_{-1} = \delta^2 f_0$$

$$\Delta^2 f_0 = \delta^2 E f_0 = \delta^2 f_1$$

and so on. Further,

$$\Delta^3 f_{-2} = \delta^3 E^{\frac{3}{2}} f_{-2} = \delta^3 f_{-\frac{1}{2}}$$

$$\Delta^3 f_{-1} = \delta^3 E^{\frac{3}{2}} f_{-1} = \delta^3 f_{\frac{1}{2}}$$

$$\Delta^3 f_0 = \delta^3 E^{\frac{3}{2}} f_0 = \delta^3 f_{\frac{3}{2}}$$

and so on. Hence, central difference table is expressed as

x	$f(x)$	δf_x	$\delta^2 f_x$	$\delta^3 f_x$	$\delta^4 f_x$
x_{-2}	f_{-2}	$\delta f_{-\frac{3}{2}}$			
x_{-1}	f_{-1}	$\delta f_{-\frac{1}{2}}$	$\delta^2 f_{-1}$	$\delta^3 f_{-\frac{1}{2}}$	
x_0	f_0	$\delta f_{\frac{1}{2}}$	$\delta^2 f_0$	$\delta^3 f_{\frac{1}{2}}$	$\delta^4 f_0$
x_1	f_1	$\delta f_{\frac{3}{2}}$	$\delta^2 f_1$		
x_2	f_2				

Now we are in a position to develop central difference interpolation formula.

(C₁) Gauss's Forward Interpolating Formula:

Let ..., $f_{-2}, f_{-1}, f_0, f_1, f_2, \dots$ be the values of the function f at ... $x_0 - 2h, x_0 - h, x_0, x_0 + h, x_0 + 2h, \dots$ Suppose that we want to compute the function value for $x = x_0 + ph$. In Gauss's forward formula, we use the differences $\delta f_{\frac{1}{2}}, \delta^2 f_{\frac{1}{2}}, \delta^3 f_{\frac{1}{2}}, \delta^4 f_{\frac{1}{2}}, \dots$ as shown by **boldface** letters in the table given below. The value f_p can be written as

$$f_p = t_0 + g_1 \delta f_{\frac{1}{2}} + g_2 \delta^2 f_{\frac{1}{2}} + g_3 \delta^3 f_{\frac{1}{2}} + g_4 \delta^4 f_{\frac{1}{2}} + \dots$$

where g_1, g_2, g_3, \dots are the constants to be determined. The above equation can be written as

$$E^p f_0 = f_0 + g_1 \delta E^{\frac{1}{2}} f_0 + g_2 \delta^2 f_0 + g_3 \delta^3 E^{\frac{1}{2}} f_0 + g_4 \delta^4 f_0 + \dots$$

Difference table

x	$f(x)$	δf_x	$\delta^2 f_x$	$\delta^3 f_x$	$\delta^4 f_x$
x_{-2}	f_{-2}	$\delta f_{-\frac{3}{2}}$			
x_{-1}	f_{-1}	$\delta f_{-\frac{1}{2}}$	$\delta^2 f_{-1}$	$\delta^3 f_{-\frac{1}{2}}$	
x_0	f_0	$\delta f_{\frac{1}{2}}$	$\delta^2 f_0$	$\delta^3 f_{\frac{1}{2}}$	$\delta^4 f_0$
x_1	f_1	$\delta f_{\frac{3}{2}}$	$\delta^2 f_1$		
x_2	f_2				

Hence,

$$E^p = I + g_1 \delta E^{\frac{1}{2}} + g_2 \delta^2 + g_3 \delta^3 E^{\frac{1}{2}} + g_4 \delta^4 + \dots$$

or

$$(1 + \Delta)^p = 1 + g_1 \Delta + g_2 \frac{\Delta^2}{1 + \Delta} + g_3 \frac{\Delta^3}{(1 + \Delta)^2} + g_4 \frac{\Delta^4}{(1 + \Delta)^3} + \dots$$

$$= I + g_1 \Delta + g_2 \Delta^2 (1 - \Delta + \Delta^2 - \dots) + g_3 \Delta^3 (1 - \Delta + \Delta^2 - \dots) + g_4 \Delta^4 (1 - 2\Delta - \dots).$$

The left-hand side equals

$$1 + p\Delta + \frac{p(p-1)}{2!} \Delta^2 + \frac{p(p-1)(p-2)}{3!} \Delta^3 + \frac{p(p-1)(p-2)(p-3)}{4!} + \dots$$

Comparing coefficients of the powers of Δ on both sides, we get

$$g_1 = p, g_2 = \frac{p(p-1)}{2!}, g_3 - g_2 = \frac{p(p-1)(p-2)}{6}$$

and so

$$\begin{aligned} g_3 &= \frac{p(p-1)(p-2)}{6} + \frac{p(p-1)}{2} = \frac{p(p-1)(p-2) + 3p(p-1)}{6} \\ &= \frac{p(p-1)(p-2+3)}{6} = \frac{(p+1)p(p-1)}{3!}, \end{aligned}$$

$$g_4 - g_3 + g_2 = \frac{p(p-1)(p-2)(p-3)}{4!},$$

and so

$$\begin{aligned} g_4 &= \frac{p(p-1)(p-2)(p-3)}{4!} + g_3 - g_2 \\ &= \frac{p(p-1)(p-2)(p-3)}{4!} + \frac{(p+1)p(p-1)}{3!} - \frac{p(p+1)}{2!} \\ &= \frac{p(p-1)[(p-2)(p-3) + 4(p+1) - 12]}{4!} \\ &= \frac{p(p-1)[p^2 - p - 2]}{4!} = \frac{(p+1)p(p-1)(p-2)}{4!}, \end{aligned}$$

and so on. Hence,

$$\begin{aligned} p &= f_0 + {}_x \delta f_{\frac{1}{2}} + \frac{(p-1)}{2!} \delta^2 f_0 + \frac{(p+1)p(p-1)}{3!} \delta^3 f_{\frac{1}{2}} + \frac{(p+1)p(p-1)(p-2)}{4!} \delta^4 f_0 + \dots \\ &= f_0 + \binom{p}{1} \delta f_{\frac{1}{2}} + \binom{p}{2} \delta^2 f_0 + \binom{p+1}{3} \delta^3 f_{\frac{1}{2}} + \binom{p+1}{4} \delta^4 f_0 + \binom{p+2}{5} \delta^5 f_{\frac{1}{2}} + \dots \end{aligned}$$

(C₂) Gauss's Backward Interpolation Formula:

The central difference table for this formula is shown below:

x	$f(x)$	δf_x	$\delta^2 f_x$	$\delta^3 f_x$	$\delta^4 f_x$
x_{-2}	f_{-2}	$\delta f_{-\frac{3}{2}}$			
x_{-1}	f_{-1}	$\delta f_{-\frac{1}{2}}$	$\delta^2 f_{-1}$	$\delta^3 f_{-\frac{1}{2}}$	
x_0	f_0	$\delta f_{\frac{1}{2}}$	$\delta^2 f_0$	$\delta^3 f_{\frac{1}{2}}$	$\delta^4 f_0$
x_1	f_1	$\delta f_{\frac{3}{2}}$	$\delta^2 f_1$		
x_2	f_2				

In Gauss's backward interpolation formula, we use the differences

$\delta f_{-\frac{1}{2}}, \delta^2 f_{-\frac{1}{2}}, \delta^3 f_{-\frac{1}{2}}, \dots$ Thus, f_p can be written as

$$f_p = t_0 + g_1 \delta f_{-\frac{1}{2}} + g_2 \delta^2 f_{-\frac{1}{2}} + g_3 \delta^3 f_{-\frac{1}{2}} + g_4 \delta^4 f_{-\frac{1}{2}} + \dots$$

where g_1, g_2, g_3, \dots are the constants to be determined. The above equation can be written as

$$E^p f_0 = f_0 + g_1 \delta E^{-\frac{1}{2}} f_r + g_2 \delta^2 E^{-\frac{1}{2}} f_r + g_3 \delta^3 E^{-\frac{1}{2}} f_r + g_4 \delta^4 f_0 + \dots$$

and so

$$E^p = I + g_1 \delta E^{-\frac{1}{2}} + g_2 \delta^2 + g_3 \delta^3 E^{-\frac{1}{2}} + g_4 \delta^4 + \dots$$

or

$$\begin{aligned} (I + \Delta)^p &= I + g_1 \frac{\Delta}{1 + \Delta} + g_2 \frac{\Delta^2}{(1 + \Delta)^2} + g_3 \frac{\Delta^3}{(1 + \Delta)^3} + g_4 \frac{\Delta^4}{(1 + \Delta)^4} + \dots \\ &= 1 + g_1 \Delta (1 - \Delta + \Delta^2 - \Delta^3 + \dots) + g_2 \Delta^2 (1 - \Delta + \Delta^2 - \Delta^3 + \dots) \\ &\quad + g_3 \Delta^3 (1 - 2\Delta + \dots) + g_4 \Delta^4 (\dots - 2\Delta + \dots). \end{aligned}$$

But

$$\begin{aligned} (I + \Delta)^p &= 1 + p\Delta + \frac{p(p-1)}{2!} \Delta^2 + \frac{p(p-1)(p-2)}{3!} \Delta^3 \\ &\quad + \frac{p(p-1)(p-2)(p-3)}{4!} \Delta^4 + \dots \end{aligned}$$

Therefore, comparing coefficients of the powers of Δ , we have

$$\begin{aligned} g_1 &= p, \quad g_2 - g_1 = \frac{p(p-1)}{2!} \text{ and so } g_2 = \frac{p(p-1)}{2!} + g_1 = \frac{(p+1)p}{2!}, \\ g_3 - g_2 + g_1 &= \frac{p(p-1)(p-2)}{3!} \text{ and so } g_3 = \frac{(p+1)p(p-1)}{3!}. \end{aligned}$$

Hence,

$$\begin{aligned} f_p &= t_0 + \delta f_{-\frac{1}{2}} + \frac{(p+1)p}{2!} \delta^2 f_0 + \frac{(p+1)p(p-1)}{3!} \delta^3 f_{-\frac{1}{2}} + \dots \\ &= f_0 + \binom{p}{1} \delta f_{-\frac{1}{2}} + \binom{p+1}{2} \delta^2 f_0 + \binom{p+1}{3} \delta^3 f_{-\frac{1}{2}} + \binom{p+2}{4} \delta^4 f_0 + \dots \end{aligned}$$

(C₃) Stirling's Interpolation Formula:

The central differences table for this formula is shown below. In this formula, we use $f_0, \delta f_{-\frac{1}{2}}, \delta^2 f_0, \delta^3 f_{-\frac{1}{2}}, \delta^3 f_{\frac{1}{2}}, \delta^4 f_0, \dots$

Difference table

x	$f(x)$	δf_x	$\delta^2 f_x$	$\delta^3 f_x$	$\delta^4 f_x$
x_{-2}	f_{-2}				
		$\delta f_{-\frac{3}{2}}$			
x_{-1}	f_{-1}		$\delta^2 f_{-1}$	$\delta^3 f_{-\frac{1}{2}}$	
		$\delta f_{-\frac{1}{2}}$			
x_0	f_0		$\delta^2 f_0$	$\delta^3 f_{\frac{1}{2}}$	$\delta^4 f_0$
		$\delta f_{\frac{1}{2}}$			
x_1	f_1		$\delta^2 f_1$		
		$\delta f_{\frac{3}{2}}$			
x_2	f_2				

By Gauss's forward interpolation formula, we have

$$\begin{aligned} {}_p = t_0 + \binom{p}{1} \delta f_{\frac{1}{2}} + \binom{p}{2} \delta^2 f_0 + \binom{p+1}{3} \delta^3 f_{\frac{1}{2}} + \binom{p+1}{4} \delta^4 f_0 \\ + \binom{p+2}{5} \delta^5 f_{\frac{1}{2}} + \dots \end{aligned} \quad (5.31)$$

and by Gauss's backward interpolation formula, we have

$${}_p = t_0 + \binom{p}{1} \delta f_{-\frac{1}{2}} + \binom{p+1}{2} \delta^2 f_0 + \binom{p+1}{3} \delta^3 f_{-\frac{1}{2}} + \binom{p+2}{4} \delta^4 f_0 + \dots \quad (5.32)$$

Adding equations (5.31) and (5.32), we get

$$\begin{aligned} {}_p &= t_0 + \frac{1}{2} \left[\binom{p}{1} \left[\delta f_{\frac{1}{2}} + \delta_{-\frac{1}{2}} \right] + \frac{1}{2} \left[\binom{p}{2} + \binom{p+1}{2} \right] \delta^2 f_0 \right. \\ &\quad \left. + \frac{1}{2} \left[\binom{p+1}{3} \right] \left[\delta^3 f_{\frac{1}{2}} + \delta^3 f_{-\frac{1}{2}} \right] \right] \\ &\quad + \frac{1}{2} \left[\binom{p+1}{4} + \binom{p+2}{4} \right] \delta^4 f_0 + \dots \\ &= f_0 \cdot \frac{p}{2} \left(\delta f_{\frac{1}{2}} + \delta_{-\frac{1}{2}} \right) + \frac{p^2}{2} \delta^2 f_0 \\ &\quad + \frac{(p+1)p(p-1)}{2(3!)} \left(\delta^3 f_{\frac{1}{2}} + \delta^3 f_{-\frac{1}{2}} \right) \\ &\quad + \frac{p(p+1)v(p-1)}{4(3!)} \delta^4 f_0 + \dots \end{aligned}$$

$$= f_0 \cdot \binom{p}{1} \mu \delta f_0 + \frac{p}{2} \binom{p}{1} \delta^2 f_0 + \binom{p+1}{3} \mu \delta^3 f_0, \\ + \frac{p}{4} \binom{p+1}{3} \delta^4 f_0 + \binom{p+2}{5} \mu \delta^5 f_0 + \dots$$

which is the required Stirling's formula.

Second Method: We have

$$_p = t_0 + S_1 \left(\delta_{\frac{1}{2}} + \delta f_{-\frac{1}{2}} \right) + S_2 \delta^2 f_0 + S_3 \left(\delta^3 f_{\frac{1}{2}} + \delta^3 f_{-\frac{1}{2}} \right) + S_4 \delta^4 f_0 + \dots, \quad (5.33)$$

where $S_1, S_2 \dots$ are the constants to be determined. Expression (5.33) can be written as

$$E^p f_0 = f_0 + S_1 (\delta E^{\frac{1}{2}} f_0 + \delta E^{-\frac{1}{2}} f_0) + S_2 \delta^2 f_0 + S_3 (\delta^3 E^{\frac{1}{2}} f_0 + \delta^3 E^{-\frac{1}{2}} f_0) + S_4 \delta^4 f_0 + \dots \\ = \left(I + S_1 \left(\Delta + \frac{\Delta}{1+\Delta} \right) + S_2 \frac{\Delta^2}{1+\Delta} + S_3 \left(\frac{\Delta^3}{1+\Delta} + \frac{\Delta^3}{(1+\Delta)^2} \right) + S_4 \frac{\Delta^4}{(1+\Delta)^4} + \dots \right) f_0.$$

Therefore, expression (5.33) gives

$$(I + \Delta)^p = I + [\Delta + \Delta(I - \Delta + \Delta^2 - \dots) + S_2 \Delta^2 (1 - \Delta + \Delta^2 - \dots)] \\ + S_3 [\Delta^3 (I - \Delta + \Delta^2 - \dots) + \Delta^3 (1 - 2\Delta + \dots)] + S_4 \Delta^4 (1 - 4\Delta + \dots).$$

The left-hand side is

$$(I + \Delta)^p = 1 + p\Delta + \frac{p(p-1)}{2!} \Delta^2 + \frac{p(p-1)(p-2)}{3!} \Delta^3 + \frac{p(p-1)(p-2)(p-3)}{4!} \Delta^4 + \dots$$

Comparing coefficients of the powers of Δ , we get

$$S_1 = \frac{p}{2},$$

$$S_2 = S_1 + \frac{p(p-1)}{2} = \frac{p^2}{2},$$

$$S_3 = \frac{p(p-1)(p+1)}{2(3!)},$$

$$S_4 = \frac{p}{4} \frac{(p+1)^2 p(p-1)}{3!}.$$

Thus,

$$_p = t_0 + \frac{p}{2} \left(\delta_{\frac{1}{2}} + \delta f_{-\frac{1}{2}} \right) + \frac{p^2}{2} \delta^2 f_0 + \frac{(p+1)p(p-1)}{2(3!)} \left(\delta^3 f_{\frac{1}{2}} + \delta^3 f_{-\frac{1}{2}} \right) \\ + \frac{p(p+1)p(p-1)}{4 \cdot 3!} \delta^4 f_0 + \dots,$$

which is the required Stirling's formula.

(C₄) Bessel's Interpolation Formula

Let $\dots, f_{-2}, f_{-1}, f_0, f_1, f_2, \dots$ be the values of a function at $\dots, x_0 - 2h, x_0 - h, x_0, x_0 + h, x_0 + 2h, \dots$. Suppose that we want to compute the function value f_p for $x = x_0 + ph$. In Bessel's formula, we use the differences as indicated in the table below. In this method $f_0, \delta f_{\frac{1}{2}}, \delta^2 f_{-\frac{1}{2}}, \delta^2 f_0, \delta^3 f_{-\frac{1}{2}}, \delta^3 f_1, \delta^4 f_0, \delta^4 f_1, \dots$ are used to approximate f_p . These values are shown in this difference table in boldface. Therefore, f_p can be written in the form

$$f_p = f_0 + \gamma_1 \delta f_{\frac{1}{2}} + B_2 (\delta^2 f_0 + \delta^2 f_1) + B_3 \delta^3 f_{\frac{1}{2}} + \gamma_4 (\delta^4 f_0 + \delta^4 f_1) + \dots$$

where $\gamma_1, \gamma_2, \dots$ are the constants to be determined.

x	$f(x)$	δf_x	$\delta^2 f_x$	$\delta^3 f_x$	$\delta^4 f_x$
x_{-2}	f_{-2}	$\delta f_{-\frac{3}{2}}$			
x_{-1}	f_{-1}	$\delta f_{-\frac{1}{2}}$	$\delta^2 f_{-1}$	$\delta^3 f_{-\frac{1}{2}}$	
x_0	f_0	$\boldsymbol{\delta f}_{\frac{1}{2}}$	$\delta^2 f_0$	$\delta^3 f_{\frac{1}{2}}$	$\delta^4 f_0$
x_1	f_1	$\delta f_{\frac{3}{2}}$	$\boldsymbol{\delta^2 f}_1$		$\boldsymbol{\delta^4 f}_1$
x_2	f_2				

The above equation can be written as

$$E^p f_0 = f_0 + \gamma_1 \delta E^{\frac{1}{2}} + B_2 (\delta^2 f_0 + \delta^2 E f_0) + B_3 \delta^3 E^{\frac{1}{2}} f_0 + \gamma_4 (\delta^4 f_0 + \delta^4 E f_0) + \dots$$

or

$$E^p = I + B_1 \delta E^{\frac{1}{2}} + B_2 (\delta^2 + \delta^2 E) + B_3 \delta^3 E^{\frac{1}{2}} + B_4 (\delta^4 + \delta^4 E) + \dots$$

or

$$(I + \Delta)^p = I + B_1 \Delta + B_2 \left(\frac{\Delta^2}{1 + \Delta} + \Delta^2 \right) + B_3 \frac{\Delta^3}{1 + \Delta} + B_4 [\Delta^2 (I + \Delta)^{-2} + \Delta^4 (I + \Delta)] + \dots \quad (5.34)$$

The left-hand side equals

$$I + p\Delta + \frac{p(p-1)}{2!} \Delta^2 + \frac{p(p-1)(p-2)}{3!} \Delta^3 + \frac{p(p-1)(p-2)(p-3)}{4!} \Delta^4 + \dots \quad (5.35)$$

Comparing coefficients of the powers of Δ in equations (5.34) and (5.35), we have

$$_1 = r,$$

$$2B_2 = \frac{p(p-1)}{2!} \text{ and so } B_2 = \frac{1}{2} \frac{p(p-1)}{2!},$$

$$-B_2 + B_3 = \frac{p(p-1)(p-2)}{3!},$$

and so

$$B_3 = \frac{p(p-1)(p-2)}{3!} + \frac{1}{2} \frac{p(p-1)}{2} = \frac{p(p-1) \left(p - 2 + \frac{3}{2} \right)}{3!} = \frac{p(p-1) \left(p - \frac{1}{2} \right)}{3!}$$

and

$$-B_2 + 2B_4 = \frac{p(p-1)(p-2)(p-3)}{4!},$$

which yields

$$\begin{aligned} B_4 &= \frac{1}{2} \left[\frac{p(p-1)(p-2)(p-3)}{4!} + \frac{p(p-1)(p-2)}{3!} \right] \\ &= \frac{1}{2} \left[\frac{p(p-1)(p-2)}{4!} (p - 3 + 4) \right] = \frac{1}{2} \frac{(p+1)p(p-1)(p-2)}{4!}. \end{aligned}$$

Similarly

$$B_5 = \frac{(p+1)p \left(p - \frac{1}{2} \right) (p-1)(p-2)}{5!}$$

and so on. Therefore,

$$\begin{aligned} t_p &= t_0 + p\delta_{\frac{1}{2}} + \frac{p(p-1)}{2!} \left[\frac{\delta^2 f_0 + \delta^2 f_1}{2} \right] + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{\frac{1}{2}} \\ &\quad + \frac{(p+1)p(p-1)(p-2)}{4!} \left[\frac{\delta^4 f_0 + \delta^4 f_1}{2} \right] + \dots \\ &= t_0 + p\delta f_{\frac{1}{2}} + \frac{p(p-1)}{2!} \mu \delta^2 f_{\frac{1}{2}} + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{\frac{1}{2}} \\ &\quad + \frac{(p+1)p(p-1)(p-2)}{4!} \mu \delta^4 f_{\frac{1}{2}} + \dots \\ &= t_0 + p\delta f_{\frac{1}{2}} + \binom{p}{2} \mu \delta^2 f_{\frac{1}{2}} + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{\frac{1}{2}} + \binom{p+1}{4} \mu \delta^4 f_{\frac{1}{2}} + \dots \end{aligned}$$

$$\begin{aligned}
&= f_0 + \frac{1}{2} \delta f_{\frac{1}{2}} + \left(p - \frac{1}{2} \right) \delta f_{\frac{1}{2}} + \binom{p}{2} \mu \delta^2 f_{\frac{1}{2}} + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{\frac{1}{2}} \\
&\quad + \binom{p+1}{4} \mu \delta^4 f_{\frac{1}{2}} + \dots \\
&= f_0 + \frac{1}{2} (f_1 - f_0) + \left(p - \frac{1}{2} \right) \delta f_{\frac{1}{2}} + \binom{p}{2} \mu \delta^2 f_{\frac{1}{2}} + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{\frac{1}{2}} \\
&\quad + \binom{p+1}{4} \mu \delta^4 f_{\frac{1}{2}} + \dots \\
&= \frac{1}{2} (f_0 + f_1) + \left(p - \frac{1}{2} \right) \delta f_{\frac{1}{2}} + \binom{p}{2} \mu \delta^2 f_{\frac{1}{2}} + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{\frac{1}{2}} \\
&\quad + \binom{p+1}{4} \mu \delta^4 f_{\frac{1}{2}} + \dots \\
&= \mu f_{\frac{1}{2}} + \left(p - \frac{1}{2} \right) \delta f_{\frac{1}{2}} + \binom{p}{2} \mu \delta^2 f_{\frac{1}{2}} + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{\frac{1}{2}} \\
&\quad + \binom{p+1}{4} \mu \delta^4 f_{\frac{1}{2}} + \dots \\
&= \binom{p}{0} \mu_{\frac{1}{2}} + \left(p - \frac{1}{2} \right) \delta f_{\frac{1}{2}} + \binom{p}{2} \mu \delta^2 f_{\frac{1}{2}} + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{\frac{1}{2}} \\
&\quad + \binom{p+1}{4} \mu \delta^4 f_{\frac{1}{2}} + \dots \\
&= \binom{p}{0} \mu_{\frac{1}{2}} + \left(p - \frac{1}{2} \right) \delta f_{\frac{1}{2}} + \binom{p}{2} \mu \delta^2 f_{\frac{1}{2}} + \frac{1}{3} \left(p - \frac{1}{2} \right) \binom{p}{2} \delta^3 f_{\frac{1}{2}} \\
&\quad + \binom{p+1}{4} \mu \delta^4 f_{\frac{1}{2}} + \dots
\end{aligned}$$

If we put $p = \frac{1}{2}$, we get

$$\frac{1}{2} = \mu f_{\frac{1}{2}} - \frac{1}{8} \mu \delta^2 f_{\frac{1}{2}} + \frac{3}{128} \mu \delta^4 f_{\frac{1}{2}} - \dots,$$

which is called formula for interpolating to halves or formula for halving an interval. It is used for computing values of the function midway between any two given values.

(C_s) Everett's Interpolation Formula

Let $\dots, f_{-2}, f_{-1}, f_0, f_1, f_2, \dots$ be the values of the function f at $\dots, x_0 - 2h, x_0 - h, x_0, x_0 + h, x_0 + 2h, \dots$. Suppose that we want to compute the function value for $x = x_0 + ph$. In Everett's formula, we use differences of even order only. Thus, we use the values $f_0, f_1, \delta^2 f_0, \delta^2 f_1, \delta^4 f_0, \delta^4 f_1, \dots$, which have been shown in boldface in the difference table below:

x	$f(x)$	δf_x	$\delta^2 f_x$	$\delta^3 f_x$	$\delta^4 f_x$
x_{-2}	f_{-2}		$\delta f_{-\frac{3}{2}}$		
x_{-1}	f_{-1}		$\delta^2 f_{-1}$		
		$\delta f_{-\frac{1}{2}}$		$\delta^3 f_{-\frac{1}{2}}$	
x_0	f_0		$\delta^2 f_0$		$\delta^4 f_0$
		$\delta f_{\frac{1}{2}}$		$\delta^3 f_{\frac{1}{2}}$	
x_1	f_1		$\delta^2 f_1$		$\delta^4 f_1$
		$\delta f_{\frac{3}{2}}$			
x_2	f_2				

By Bessel's formula, we have

$$\begin{aligned} f_p &= f_0 + p\delta_{\frac{1}{2}} + \frac{p(p-1)}{2!} \left[\frac{\delta^2 f_0 + \delta^2 f_1}{2} \right] + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{\frac{1}{2}} \\ &\quad + \frac{(p+1)p(p-1)p(p-2)}{4!} \left[\frac{\delta^4 f_0 + \delta^4 f_1}{2} \right] + \dots \end{aligned}$$

Since Everett's formula expresses f_p in terms of even differences lying on the horizontal lines through f_0 and f_1 , therefore we convert $\delta f_{\frac{1}{2}}, \delta^3 f_{\frac{1}{2}}, \dots$ in the Bessel's formula into even differences.

By doing so, we have

$$\begin{aligned} f_p &= f_0 + p(f_1 - f_0) + \frac{p(p-1)}{2!} \left[\frac{\delta^2 f_0 + \delta^2 f_1}{2} \right] + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} [\delta^2 f_1 - \delta^2 f_0] \\ &\quad + \frac{(p+1)p(p-1)(p-2)}{4!} \left[\frac{\delta^4 f_0 + \delta^4 f_1}{2} \right] + \dots \\ &= (1-p)f_0 - \frac{p(p-1)(p-2)}{3!} \delta^2 f_0 - \frac{(p+1)p(p-1)(p-2)(p-3)}{5!} \delta^4 f_0 - \dots \\ &\quad + pf_1 + \frac{(p+1)p(p-1)}{3!} \delta^2 f_1 + \frac{(p+2)(p+1)p(p-1)(p-2)}{5!} \delta^4 f_1 + \dots \end{aligned}$$

$$\begin{aligned}
&= (1-p)f_0 + \binom{2-p}{3} \delta^2 f_0 + \binom{3-p}{5} \delta^4 f_0 + \dots \\
&\quad + pf_1 + \binom{p+1}{3} \delta^2 f_1 + \binom{p+2}{5} \delta^4 f_1 + \dots \\
&= qf_0 + \binom{q+1}{3} \delta^2 f_0 + \binom{q+2}{5} \delta^4 f_0 + \dots \\
&\quad + pf_1 + \binom{p+1}{3} \delta^2 f_1 + \binom{p+2}{5} \delta^4 f_1 + \dots,
\end{aligned}$$

where $q = 1 - p$.

Second Method: Let $\dots, f_{-2}, f_{-1}, f_0, f_1, f_2, \dots$ be the values of a function f for $\dots, x_0 - 2h, x_0 - h, x_0, x_0 + h, x_0 + 2h, \dots$. We want to compute f_p , where $x = x_0 + ph$. We use the even differences lying on the horizontal lines through f_0 and f_1 . So let

$$\begin{aligned}
f_p &= f_{-2} + E_2 \delta^2 f_0 + E_4 \delta^4 f_0 + \dots \\
&\quad + F_0 f_1 + F_2 \delta^2 f_1 + F_4 \delta^4 f_1 + \dots
\end{aligned}$$

Therefore,

$$\begin{aligned}
E^p f_0 &= f_{-2} + E_2 \delta^2 f_0 + E_4 \delta^4 f_0 + \dots \\
&\quad + f_{-1} + F_2 \delta^2 f_0 + F_4 \delta^4 f_0 + \dots
\end{aligned}$$

or

$$\begin{aligned}
(I + \Delta)^p f_0 &= f_{-2} + E_2 \delta^2 f_0 + E_4 \delta^4 f_0 + \dots \\
&\quad + f_{-1} + F_2 \delta^2 f_0 + F_4 \delta^4 f_0 + \dots \\
&= f_{-2} + E_2 \frac{\Delta^2}{1 + \Delta} + E_4 \frac{\Delta^4}{(1 + \Delta)^2} + \dots \\
&\quad + F_0 (1 + \Delta) + F_2 \Delta^2 + F_4 \frac{\Delta^4}{1 + \Delta} + \dots
\end{aligned} \tag{5.36}$$

The left-hand side of equation (5.36) is

$$I + p\Delta + \frac{(p-1)}{2!} \Delta^2 + \frac{p(p-1)(p-2)}{3!} \Delta^3 + \frac{p(p-1)(p-2)(p-3)}{4!} \Delta^4 + \dots,$$

whereas the right-hand side is

$$\begin{aligned}
&f_{-2} + E_2 [\Delta^2 (I - \Delta + \Delta^2 - \Delta^3 + \dots)] + E_4 [\Delta^4 (I - 2\Delta + \dots)] + \dots \\
&\quad + F_0 (I + \Delta) + F_2 \Delta^2 + F_4 [\Delta^4 (I - \Delta + \Delta^2 - \Delta^3 + \dots)] + \dots
\end{aligned}$$

Comparing coefficients of $\Delta, \Delta^2, \Delta^3, \dots$ on both sides, we get

$$p = f_0, 1 = f_{-2} + f_0 \text{ and so } f_{-2} = 1 - p,$$

$$-E_2 = \frac{p(p-1)(p-2)}{3!}, -F_2 = \frac{p(p-1)}{2} \text{ and so } F_2 = \binom{p+1}{3}.$$

Similarly, other coefficients are obtained. Hence,

$$\begin{aligned}
 {}_p &= (1-p)f_0 + \binom{2-p}{3}\delta^2 f_0 + \binom{3-p}{5}\delta^4 f_0 + \dots \\
 &\quad + pf_0 + \binom{p+1}{3}\delta^2 f_1 + \binom{p+2}{5}\delta^4 f_1 + \dots \\
 &= qf_0 + \binom{q+1}{3}\delta^2 f_0 + \binom{q+2}{5}\delta^4 f_0 + \dots \\
 &\quad + pf_0 + \binom{p+1}{3}\delta^2 f_1 + \binom{p+2}{5}\delta^4 f_1 + \dots,
 \end{aligned}$$

where $q = 1 - p$.

Remark 5.3. The Gauss's forward, Gauss's backward, Stirling's, Bessel's, Everett's, Newton's forward and Newton's backward interpolation formulae are called classical formulae and are used for equal spacing.

EXAMPLE 5.24

The function y is given in the table below:

x	0.01	0.02	0.03	0.04	0.05
y	98.4342	48.4392	31.7775	23.4492	18.4542

Find y for $x = 0.0341$.

Solution. The central difference table is

x	y	δ	δ^2	δ^3	δ^4
0.01	98.4342	-49.9950			
0.02	48.4392	-16.6617	33.3333	-24.9999	
0.03	31.7775	-8.3283	8.3334	-5.0001	19.9998
0.04	23.4492	-4.9950	3.3333		
0.05	18.4542				

Letting $x_0 = 0.03$, we have $p = \frac{x - x_0}{h} = \frac{0.0341 - 0.030}{0.01} = 0.1$. Using Bessel's formula, we have

$$\begin{aligned}
 f(0.0341) &= t_c + \frac{p}{2}\delta f_0 + \frac{p(p-1)}{4}(\delta^2 f_0 + \delta^2 f_1) + \frac{p\left(p-\frac{1}{2}\right)(p-1)}{3!}\delta^3 f_{\frac{1}{2}} \\
 &\quad + \frac{(p+1)p(p-1)(p-2)}{4!}\left(\frac{\delta^4 f_0 + \delta^4 f_1}{2}\right) + \dots
 \end{aligned}$$

$$\begin{aligned}
&= 31.7775 + 0.41(8.3283) + \frac{11.6667}{4}(0.2419) \\
&= 27.475924 \text{ approximately.}
\end{aligned}$$

EXAMPLE 5.25

If third differences are constant, prove that

$$y_{x+\frac{1}{2}} = \frac{1}{2}(y_v + y_{x+1}) - \frac{1}{16}(\Delta^2 y_{x-1} + \Delta^2 y_v).$$

Solution. The Bessel's formula in this case becomes

$$\begin{aligned}
y_p &= \frac{y_0 + v_1}{2} + \left(p - \frac{1}{2}\right)\Delta_v^0 + \frac{p(p-1)}{2!} \frac{[\Delta^2 v_{-1} + \Delta^2 y_0]}{2} \\
&\quad + \frac{p\left(p - \frac{1}{2}\right)(p-1)}{3!} \Delta^3 y_{-1},
\end{aligned}$$

because the higher differences than that of third order will be equal to zero by the hypothesis. Putting $p = \frac{1}{2}$, we get

$$y_{\frac{1}{2}} = \frac{y_0 + v_1}{2} - \frac{1}{16}(\Delta^2 v_{-2} + \Delta^2 y_0).$$

Changing the origin to x , we have

$$y_{x+\frac{1}{2}} = \frac{1}{2}(y_v + y_{x+1}) - \frac{1}{16}(\Delta^2 y_{x-1} + \Delta^2 y_v).$$

EXAMPLE 5.26

Given $y_0, v_1, y_2, v_2, y_4, v_5$ (fifth difference constant), prove that

$$y_{\frac{5}{2}} = \frac{c}{2} + \frac{25(c-b) + 3(a-c)}{256},$$

where $a = v_0 + v_5, b = v_1 + v_4, c = v_2 + v_3$.

Solution. Putting $p = \frac{1}{2}$ in the Bessel's formula, we have

$$y_{\frac{5}{2}} = \frac{y_0 + v_1}{2} - \frac{1}{8} \left(\frac{\Delta^2 v_{-1} + \Delta^2 y_0}{2} \right) + \frac{3}{128} \left(\frac{\Delta^4 y_{-2} + \Delta^4 y_{-1}}{2} \right).$$

Shifting the origin to 2, we obtain

$$y_{\frac{5}{2}} = \frac{1}{2}(y_v + y_3) - \frac{1}{16}(\Delta^2 y_1 + \Delta^2 y_2) + \frac{3}{256}(\Delta^4 v_0 + \Delta^4 y_1). \quad (5.37)$$

But

$$\Delta^2 y_1 = y_3 - 2y_2 + y_1, \Delta^4 y_0 = y_v - 4y_2 + 6y_3 - 4y_4 + y_5 \text{ etc.}$$

Substituting these values in equation (5.37), we get the required result.

5.7 USE OF INTERPOLATION FORMULAE

We know that the Newton formulae with forward and backward differences are most appropriate for calculation near the beginning and the end, respectively, of tabulation, and their use is mainly restricted to such situations.

The Gaussian forward and backward formulae terminated with an even difference are equivalent to each other and to the Stirling's formula terminated with the same difference. The Gaussian forward formula terminated with an odd difference is equivalent to the Bessel formula terminated with the same difference. The Gaussian backward formula launched from x_0 and terminating with an odd difference is equivalent to the Bessel's formula launched from x_{-1} and terminated with the same difference. Thus, in place of using a Gaussian formula, we may use an equivalent formula of either Stirling or Bessel for which the coefficients are extensively tabulated.

To interpolate near the middle of a given table, Stirling's formula gives the most accurate result for $-\frac{1}{4} \leq p \leq \frac{1}{4}$ and Bessel's formula is most efficient near $p = \frac{1}{2}$, say $\frac{1}{4} \leq p \leq \frac{3}{4}$. When the highest difference to be retained is odd, Bessel's formula is recommended and when the highest difference to be retained is even, then Stirling's formula is preferred.

In case of Stirling's formula, the term containing the third difference, viz.,

$$\frac{p(p^2 - 1)}{6} \delta^3 f_{-\frac{1}{2}}$$

may be neglected if its contribution to the interpolation is less than half a unit in the last place. This means that

$$\left| \frac{p(p^2 - 1)}{6} \delta^3 f_{-\frac{1}{2}} \right| < \frac{1}{2} \text{ for all } p \text{ in the range } 0 \leq p \leq 1.$$

But the maximum value of $\frac{p(p^2 - 1)}{6}$ is 0.064 and so we have

$$\left| 0.064 \delta^3 f_{-\frac{1}{2}} \right| < \frac{1}{2} \text{ or } \left| \delta^3 f_{-\frac{1}{2}} \right| < 8.$$

If we consider Bessel's formula, the contribution from the term containing the third difference will be less than half a unit in the last place, provided that

$$\left| \frac{p \left(p - \frac{1}{2} \right) (p - 1)}{6} \delta^3 f_{\frac{1}{2}} \right| < \frac{1}{2}.$$

But the maximum value of $\frac{p \left(p - \frac{1}{2} \right) (p - 1)}{6}$ is 0.008 and so

$$\left| \delta^3 f_{\frac{1}{2}} \right| < 60.$$

Thus, if we neglect the third differences, Bessel's formula is about eight times more accurate than Stirling's formula. If the third differences need to be retained (when they are more than 60 in magnitude), then Everett's formula may be gainfully employed since Everett's formula with second difference is equivalent to Bessel's formula with third differences.

5.8 INTERPOLATION WITH UNEQUAL-SPACED POINTS

The classical polynomial interpolating formulae discussed so far are limited to the case in which intervals of independent variables were equally spaced. We shall now discuss interpolation formulae with unequally spaced values of the argument.

(A) Divided Differences

Let $f(x_0), f(x_1), \dots, f(x_n)$ be the values of a function f corresponding to the arguments x_0, x_1, \dots, x_n where the intervals $x_1 - x_0, x_2 - x_1, \dots, x_n - x_{n-1}$ are not necessarily equally spaced. Then the first divided differences of f for the arguments x_0, x_1, x_2, \dots are defined by

$$f(x_0, x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0},$$

$$f(x_1, x_2) = \frac{f(x_2) - f(x_1)}{x_2 - x_1},$$

and so on. The second divided difference (divided difference of order 2) of f for three arguments x_0, x_1, x_2 is defined by

$$f(x_0, x_1, x_2) = \frac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0}$$

and similarly, the divided difference of order n is defined by

$$f(x_0, x_1, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_n) - f(x_0, x_1, \dots, x_{n-1})}{x_n - x_0}.$$

Remark 5.4. Even if the arguments are equal, the divided difference may still have a meaning. For example, if we set $x_1 = x_0 + \varepsilon$, then

$$f(x_0, x_1) = f(x_0, x_0 + \varepsilon) = \frac{f(x_0 + \varepsilon) - f(x_0)}{\varepsilon}$$

and in the limit when $\varepsilon \rightarrow 0$, we have

$$f(x_0, x_0) = f'(x_0) \text{ if } f \text{ is derivable.}$$

Similarly,

$$f(x_0, x_1, \dots, x_r) = \frac{f^{(r)}(x_0)}{r!} \text{ for } r+1 \text{ equal arguments } x_0.$$

Further, we observe that

$$\begin{aligned} f(x_0, x_1) &= \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(x_0) - f(x_1)}{x_0 - x_1} = f(x_1, x_0) \\ f(x_0, x_1, x_2) &= \frac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0}, \\ &= \frac{1}{x_2 - x_0} \left[\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0} \right] \\ &= \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)} + \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)} \end{aligned}$$

and in general,

$$f(x_0, x_1, \dots, x_n) = \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)\dots(x_0 - x_n)} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)\dots(x_1 - x_n)} + \dots + \frac{f(x_n)}{(x_n - x_0)(x_n - x_1)\dots(x_n - x_{n-1})}.$$

Hence, the divided differences are symmetrical in their arguments. It follows therefore that for any function f , the value of the divided difference remains unaltered when any of the arguments involved are interchanged. Thus, the value of the divided difference depends only on the value of the arguments involved and not on the order in which they are taken. Thus,

$$\begin{aligned} f(x_0, \dots) &= f(x_1, x_0) \\ f(x_0, x_1, x_2, \dots) &= f(x_2, x_1, x_0) = f(x_1, x_0, x_2). \end{aligned}$$

Theorem 5.3. The n th divided differences of a polynomial of the n th degree are constant.

Proof: Consider the function $f(x) = x^n$. The first divided difference

$$\begin{aligned} f(x_r, x_{r+1}) &= \frac{f(x_{r+1}) - f(x_r)}{x_{r+1} - x_r} = \frac{x_{r+1}^n - x_r^n}{x_{r+1} - x_r} \\ &= x_{r+1}^{n-1} + x_r x_{r+1}^{n-2} + \dots + x_r^{n-2} x_{r+1} + \dots + x_r^{n-1} \end{aligned}$$

is a homogeneous polynomial of degree $n-1$ in x_r, \dots, x_{r+1} .

Similarly, it can be shown that second divided differences are homogeneous polynomials of degree $n-2$. Proceeding by mathematical induction, it can be shown that divided difference of n th order is a polynomial of degree $n-n=0$ and so is a constant.

For a polynomial of the n th degree with leading term $a_0 x^n$, the n th divided difference of all terms except the leading term are zero. So the n th divided differences of this polynomial are constant and of value a_0 .

Remark 5.5. Let the arguments be equally spaced so that $x_1 - x_0 = x_2 - x_1 = \dots = x_n - x_{n-1} = h$. Then

$$\begin{aligned} f(x_0, x_1) &= \frac{f(x_1) - f(x_0)}{h} = \frac{\Delta f_0}{h} \\ f(x_0, x_1, x_2) &= \frac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0} = \frac{1}{2h} \left[\frac{\Delta f_1}{h} - \frac{\Delta f_0}{h} \right] \\ &= \frac{1}{2h^2} \Delta^2 f_0 = \frac{1}{2!} \frac{1}{h^2} \Delta^2 f_0 \end{aligned}$$

and, in general,

$$f(x_0, \dots, x_n) = \frac{1}{n!} \frac{1}{h^n} \Delta^n f_0.$$

If the tabulated function is a polynomial of n th degree, then $\Delta^n f_0$ would be constant and hence the n th divided difference would also be a constant.

5.9 NEWTON'S FUNDAMENTAL (DIVIDED DIFFERENCE) FORMULA

Let $f(x_0), f(x_1), \dots, f(x_n)$ be the values of a function f corresponding to the arguments x_0, \dots, x_n , where the intervals $x_1 - x_0, x_2 - x_1, \dots, x_n - x_{n-1}$ are not necessarily equally spaced. By the definition of divided differences, we have

$$f(x, x_0) = \frac{f(x) - f(x_0)}{x - x_0}$$

and so

$$f(x) = t(x_0) + (x - x_0)f(x, x_0). \quad (5.38)$$

Further,

$$f(x, x_0, x_1) = \frac{f(x, x_0) - f(x_0, x_1)}{x - x_1},$$

which yields

$$f(x, x_0) = f(x_0, x_1) + (x - x_1)f(x, x_0, x_1). \quad (5.39)$$

Similarly,

$$f(x, x_0, x_1) = t(x_0, x_1, x_2) + (x - x_2)f(x, x_0, x_1, x_2) \quad (5.40)$$

and in general,

$$f(x, x_0, \dots, x_{n-1}) = f(x_0, x_1, \dots, x_n) + (x - x_n)f(x, x_0, x_1, \dots, x_n). \quad (5.41)$$

Multiplying equation (5.39) by $(x - x_0)$ (5.40) by $(x - x_0)(x - x_1)$, and so on and finally the last term (5.41) by $(x - x_0)(x - x_1)\dots(x - x_{n-1})$ and adding, we obtain

$$\begin{aligned} f(x) &= t(x_0) + (x - x_0)t(x_0, x_1) + (x - x_0)(x - x_1)f(x_0, x_1, x_2) \\ &\quad + \dots + (x - x_0)(x - x_1)\dots(x - x_{n-1})f(x_0, \dots, x_{n-1}, x_n) + R \end{aligned}$$

where

$$R = (x - x_0)(x - x_1)\dots(x - x_n)f(x, x_0, \dots, x_n).$$

This formula is called Newton's divided difference formula. The last term R is the remainder term after $(n + 1)$ terms.

Remark 5.6. If we consider the case of equal spacing, then we have

$$f(x_0, \dots, x_n) = \frac{1}{h^n n!} \Delta^n f_0$$

and so

$$\begin{aligned} f(x) &= t(x_0) + \frac{x - x_0}{h} \Delta f_0 + \frac{(x - x_0)(x - x_1)}{h^2 2!} \Delta^2 f_0 + \dots \\ &= f_0 + \frac{x_0 + ph - x_0}{h} \Delta f_0 + \frac{(x_0 + ph - x_0)(x_0 + ph - x_1)}{h^2 2!} \Delta^2 f_0 + \dots \\ &= f_0 + p \Delta f_0 + \frac{p(p-1)}{2!} \Delta^2 f_0 + \dots, \end{aligned}$$

which is nothing but Newton's forward difference formula.

EXAMPLE 5.27

Find a polynomial satisfied by $(-4, 1245)$, $(-1, 33)$, $(0, 5)$, $(2, 9)$, and $(5, 1335)$.

Solution. The divided difference table based on the given nodes is shown below:

x	y				
-4	1245	-404			
-1	33	-28	94	-14	
0	5	2	10	13	3
2	9	442			
5	1335				

In fact,

$$f(x_0, x_1) = \frac{f(x_0) - f(x_1)}{x_0 - x_1} = \frac{1245 - 33}{-3} = -404,$$

$$f(x_1, x_2) = \frac{f(x_1) - f(x_2)}{x_1 - x_2} = \frac{33 - 28}{-1} = -28,$$

$$f(x_2, x_3) = \frac{f(x_2) - f(x_3)}{x_2 - x_3} = \frac{28 - 5}{-2} = 2,$$

$$f(x_3, x_4) = \frac{f(x_3) - f(x_4)}{x_3 - x_4} = \frac{5 - 9}{-3} = 442,$$

$$f(x_0, x_1, x_2) = \frac{f(x_0) - f(x_1, x_2)}{x_0 - x_1} = \frac{-404 + 28}{-4} = 94,$$

$$f(x_1, x_2, x_3) = \frac{f(x_1) - f(x_2, x_3)}{x_1 - x_2} = \frac{28 - 5}{-3} = 10,$$

$$f(x_2, x_3, x_4) = \frac{f(x_2) - f(x_3, x_4)}{x_2 - x_3} = \frac{5 - 442}{-5} = 88,$$

$$f(x_0, x_1, x_2, x_3) = \frac{f(x_0, x_1, x_2) - f(x_1, x_2, x_3)}{x_0 - x_1} = \frac{94 - 10}{-6} = -14,$$

$$f(x_1, x_2, x_3, x_4) = \frac{f(x_1, x_2, x_3) - f(x_2, x_3, x_4)}{x_1 - x_2} = \frac{10 - 88}{-6} = 13,$$

$$f(x_0, x_1, x_2, x_3, x_4) = \frac{f(x_0, x_1, x_2, x_3) - f(x_1, x_2, x_3, x_4)}{x_0 - x_1} = \frac{-14 - 13}{-9} = 3.$$

Putting these values in Newton's fundamental formula, we have

$$\begin{aligned}
f(x) &= t(x_0 \dots x_0) t(x_0, x_1) + (x - x_0)(\dots - \dots) f(x_0, x_1, x_2) \\
&\quad + (x - x_0)(x - \dots_1)(\dots - x_2) f(x_0, \dots_1, \dots_2, x_3) \\
&\quad + (x - x_0)(x - x_1)(x - x_2)(x - x_3) f(x_0, \dots_1, x_2, x_3, x_4) \\
&= 1245 - 404(x+4) + 94(x+4)(x+1) - 14(x+4)(x+1)x \\
&\quad + 3(x+4)(x+1)x(x-2) \\
&= 3x^4 - 5x^3 + 6x^2 - 14x + 5.
\end{aligned}$$

EXAMPLE 5.28

Using the table given below, find $f(x)$ as a polynomial in x .

x	-1	0	3	6	7
$f(x)$	3	-6	39	822	1611

Solution. The divided difference table for the given data is shown below

	x	$f(x)$				
x_0	-1	3				
x_1	0	-6	-9			
x_2	3	39	15	6		
x_3	6	822	261	41	5	
x_4	7	1611	789	132	13	1

Putting these values in the Newton's divided difference formula, we have

$$\begin{aligned}
f(x) &= t(x_0 \dots x_0) t(x_0, x_1) + (x - x_0)(\dots - \dots) f(x_0, x_1, x_2) \\
&\quad + (x - x_0)(x - \dots_1)(\dots - x_2) f(x_0, \dots_1, \dots_2, x_3) \\
&\quad + (x - x_0)(x - x_1)(x - x_2)(x - x_3) f(x_0, \dots_1, x_2, x_3, x_4) \\
&= 3 - 9(x+1) + 6(x+1)x + 5(x+1)x(x-3) \\
&\quad + 1(x+1)x(x-3)(x-6) = x^4 - 3x^3 + 5x^2 - 6.
\end{aligned}$$

EXAMPLE 5.29

By means of Newton's divided difference formula, find the value of $f(8)$ and $f(15)$ from the following table:

x	4	5	7	10	11	13
$f(x)$	48	100	294	900	1210	2028

Solution. The divided difference table is

	x	$f(x)$					
x_0	4	48	52				
x_1	5	100	97	15	1	0	0
x_2	7	294	202	21	1	0	0
x_3	10	900	310	27	1	0	0
x_4	11	1210	409	33			
x_5	13	2028					

Using the formula

$$\begin{aligned} f(x) = & t(x_0 - \dots - x_0) t(x_0, x_1) + (x - x_0)(\dots - x_1) f(x_0, x_1, x_2) \\ & + (x - x_0)(x - x_1)(\dots - x_2) f(x_0, \dots, x_2, x_3), \end{aligned}$$

we obtain

$$f(8) = 48 + (8 - 4)(52) + (8 - 4 \cdot 8 - 5)15 + (8 - 4)(8 - 5)(8 - 7)(1) = 448 \text{ and}$$

$$f(15) = 48 + (15 - 4)(52) + (15 - 4 \cdot 15 - 5)(15) + (15 - 4)(15 - 5)(15 - 7)(1) = 3150.$$

5.10 ERROR FORMULAE

Let $f(x)$ be approximated by a polynomial $p(x)$ of degree n by Newton's divided difference formula. Then $f(x)$ and $p(x)$ coincide at $(n+1)$ distinct points x_0, \dots, x_n and the error $E(x) = r(x) - p(x)$ is given by

$$E(x) = \Pi(x) f(x, x_0, x_1, \dots, x_n), \quad (5.42)$$

where

$$\Pi(x) = (x - x_0)(x - x_1) \dots (x - x_n) \quad (5.43)$$

is a polynomial of degree $n+1$.

Assume that f possesses $n+1$ continuous derivatives in the relevant interval. Consider a linear combination of $f(x)$, $p(x)$, and $\Pi(x)$ as

$$F(x) = r(x) - p(x) - K \Pi(x), \quad (5.44)$$

where K is a constant to be determined in such a way that $F(x)$ vanishes not only at the $n+1$ points but also at an arbitrarily chosen point X which differs from all these points.

Let \bar{I} constitute the closed interval limited by the smallest and largest of $n+2$ values x_0, \dots, x_n, X . Then F vanishes at least $n+2$ times in the closed interval \bar{I} . By Rolle's Theorem $F'(x)$ vanishes at least $n+1$ times in \bar{I} , $F''(x)$ at least n times, and finally $F^{(n+1)}(x)$ vanishes at least once inside \bar{I} . Let ξ be the one such point. It follows from equation (5.44) that

$$0 = f^{(n+1)}(\xi) - p^{(n+1)}(\xi) - K \Pi^{(n+1)}(\xi). \quad (5.45)$$

But since $p(x)$ is a polynomial of degree n , its $(n+1)$ th derivative vanishes identically. Also, by equation (5.43), we have $\Pi^{(n+1)}(x) = (n+1)!$. Hence, equation (5.45) yields $K = \frac{1}{(n+1)!} f^{(n+1)}(\xi)$, and relation $F(x) = 0$ becomes

$$f(X) - p(X) = \frac{\Pi(x) f^{(n+1)}(\xi)}{(n+1)!}, \quad \xi \in I.$$

Even if X is taken any of the arguments x_0, \dots, x_n , both sides of this relation vanishes. Since X is arbitrary, we have

$$E(x) = f(x) - p(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi) \Pi(x) \quad (5.46)$$

for some $\xi \in I$, where ξ is in the interval limited by the largest and smallest of the numbers x_0, \dots, x_n . Since equations (5.42) and (5.46) must be equivalent, we have

$$f(x, x_0, \dots, x_n) = \frac{1}{(n+1)!} f^{(n+1)}(\xi)$$

for some argument in the interval I .

EXAMPLE 5.30

Find the maximum error in interpolating to find $\sin x$ for any x within the range of the table given below:

x	0°	15°	30°	45°	60°	75°	90°
$\sin x$	0	0.25882	0.5	0.70711	0.86603	0.96593	1.0

Solution. We have

$$f(x) = \sin x \text{ and } n+1 = 7.$$

Then $f^{(7)}(x) = -\cos x$. The formula for error will not yield the maximum error because we know nothing about ξ except that it lies in the range $0^\circ - 90^\circ$. But since $\cos x$ is bounded in that interval, the formula will give us an upper bound on the size of the error. Thus,

$$\left| f(x) - p(x) \right| \leq \frac{1}{7!} \left| (x-0) \left(x - \frac{\pi}{12} \right) \left(x - \frac{2\pi}{12} \right) \dots \left(x - \frac{6\pi}{12} \right) \right|.$$

For example, if we compute the value of $\sin \frac{5\pi}{24}$, then

$$\left| f\left(\frac{5\pi}{24}\right) - p\left(\frac{5\pi}{24}\right) \right| \leq \frac{1}{5040} (5)(3)(1)(1)(3)(5)(7) \frac{\pi^7}{(24)^7} = (2.06)10^{-7}.$$

EXAMPLE 5.31

The function $y = f(x)$ is supposed to be differentiable three times. Show that

$$\begin{aligned} f(x) &= -\frac{(x - x_1)(x - 2x_0 + x_1)}{(x_1 - x_0)^2} f(x_0) + \frac{(x - x_0)(x - x_1)}{x_0 - x_1} f'(x_0) \\ &\quad + \frac{(x - x_0)^2}{(x_1 - x_0)^2} f(x_1) + R(x), \end{aligned}$$

where

$$R(x) = \frac{1}{6} (x - x_0)^2 (x - x_1) f'''(\xi), x_0 < x, \xi < x_1.$$

Solution. We apply Newton's divided difference formula to three points x_0, x_1, x_2 and have

$$f(x) = f(x_0) + (x - x_0) f(x_0, x_1) + (x - x_0)(x - x_1) f(x_0, x_1, x_2) + R(x),$$

where

$$\begin{aligned} R(x) &= \frac{f'''(\xi)}{3!}(x - x_0)(x - x_1)(x - x_2) \\ &= \frac{f'''(\xi)}{6}(x - x_0)^2(x - x_1). \end{aligned}$$

But $f(x_0, \dots) = f'(x_0)$. Therefore,

$$\begin{aligned} f(x) &= f(x_0) + (x - x_0) f'(x_0) + (x - x_0)^2 \frac{f(x_0, x_1) - f(x_0, x_1)}{x_1 - x_0} + R(x) \\ &\quad \left. f'(x_0) - [f(x_0) - f(x_1)] \right/ (x_0 - x_1) + R(x) \\ &= f(x_0) + (x - x_0) f'(x_0) + (x - x_0)^2 \frac{(x - x_0)^2 f'(x_0) - (x - x_0)^2 f(x_0) + (x - x_0)^2 f(x_1)}{(x_0 - x_1)^2} + R(x) \\ &= f(x_0) + (x - x_0) f'(x_0) + \frac{(x - x_0)^2 f'(x_0)}{(x_0 - x_1)^2} - \frac{(x - x_0)^2 f(x_0)}{(x_0 - x_1)^2} + \frac{(x - x_0)^2 f(x_1)}{(x_0 - x_1)^2} + R(x) \\ &= -f(x_0) \left[\frac{(x - x_0)^2}{(x_0 - x_1)^2} - 1 \right] + f'(x_0) \left[\frac{(x - x_0)(x_0 - x_1 + x - x_0)}{x_0 - x_1} \right] + \frac{(x - x_0)^2}{(x_0 - x_1)^2} f(x_1) + R(x) \\ &= -\frac{(x - x_0)(x - 2x_0 + x_1)}{(x_1 - x_0)^2} f(x_0) + \frac{(x - x_0)(x - x_1)}{x_1 - x_0} f'(x_0) + \frac{(x - x_0)^2}{(x_1 - x_0)^2} f(x_1) + R(x). \end{aligned}$$

EXAMPLE 5.32

Find the missing term in the following table:

x	0	1	2	3	4
y	1	3	9	-	81

Explain, why the result differs from $3^3 = 27$.

Solution. The divided difference table is

x	y				
x_0	0	1	2		
x_1	1	3	6	2	2
x_2	2	9	36	10	
x_3	4	81			

Therefore, using Newton's divided difference formula, we have

$$\begin{aligned} f(3) &= t(x_0) + (x - x_0)t(x_0, x_1) + (x - x_0)(x - x_1)f(x_0, x_1, x_2) + (x - x_0)(x - x_1)(x - x_2)f(x_0, x_1, x_2, x_3) \\ &= 1 + (3-0)(2) + (3-0)(3-1)(2) + (3-0)(3-1)(3-2)(2) \\ &= 1 + 6 + 12 + 12 = 31. \end{aligned}$$

It differs from $3^3 = 27$ because of the error $E(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod(x)$.

Remark 5.7. Using Newton's divided difference formula, the polynomial satisfying the given data in the above example is

$$\begin{aligned} f(x) &= 1 + 2x + 2x(x-1) + 2x(x-1)(x-2) \\ &= 2x^3 - 4x^2 + 4x + 1. \end{aligned}$$

5.11 LAGRANGE'S INTERPOLATION FORMULA

Let f be continuous and differentiable $(n+1)$ times in an interval (a, b) and let $f_0, f_1, f_2, \dots, f_n$ be the values of f at $x_0, x_1, x_2, \dots, x_n$ where $x_0, x_1, x_2, \dots, x_n$ are not necessarily equally spaced. We wish to find a polynomial of degree n , say $P_n(x)$ such that

$$P_n(x_i) = f(x_i) = f_i, \quad i = 0, 1, \dots, n. \quad (5.47)$$

Let

$$P_n(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n, \quad (5.48)$$

be the desired polynomial. Substituting the condition (5.47) in equation (5.48), we obtain the following system of equations:

$$\left. \begin{array}{l} f_0 = a_0 + a_1x_0 + a_2x_0^2 + \dots + a_nx_0^n \\ f_1 = a_0 + a_1x_1 + a_2x_1^2 + \dots + a_nx_1^n \\ f_2 = a_0 + a_1x_2 + a_2x_2^2 + \dots + a_nx_2^n \\ \dots \\ \dots \\ f_n = a_0 + a_1x_n + a_2x_n^2 + \dots + a_nx_n^n \end{array} \right\} \quad (5.49)$$

This set of equations will have a solution if the determinant

$$\begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{vmatrix} \neq 0.$$

The value of this determinant, called Vandermonde's determinant, is $(x_0 - x_1)(x_0 - x_2)\dots(x_0 - x_n)(x_1 - x_2)\dots(x_1 - x_n)\dots(x_n - x_1)\dots(x_n - x_{n-1})$. Eliminating a_0, a_1, \dots, a_n from equations (5.48) and (5.49), we obtain

$$\begin{vmatrix} P_n(x) & 1 & x & x^2 & \dots & x^n \\ \alpha & 1 & x_0 & x_0^2 & \dots & x_0^n \\ \beta & 1 & x_1 & x_1^2 & \dots & x_1^n \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ f_n & 1 & x_n & x_n^2 & \dots & x_n^n \end{vmatrix} = 0, \quad (5.50)$$

which shows that $P_n(x)$ is a linear combination of $\alpha, \beta, \dots, f_n$. Hence, we write

$$P_n(x) = \sum_{i=0}^n L_i(x) f_i, \quad (5.51)$$

where $L_i(x)$ are polynomials in x of degree n . But $P_n(x_j) = f_j$ for $j = 0, 1, 2, \dots$. Therefore, equation (5.51) yields

$$\left. \begin{array}{l} L_i(x_j) = 0 \text{ for } i \neq j \\ L_i(x_j) = 1 \text{ for } i = j \end{array} \right\} \text{for all } j. \quad (5.52)$$

Hence, we may take $L_i(x)$ as

$$L_i(x) = \frac{(x - x_0)(x - x_1)\dots(x - x_{i-1})(x - x_{i+1})\dots(x - x_n)}{(x_i - x_0)(x_i - x_1)\dots(x_i - x_{i-1})(x_i - x_{i+1})\dots(x_i - x_n)} \quad (5.53)$$

which clearly satisfies the condition (5.52). Let

$$\Pi(x) = (x - x_0)(x - x_1)\dots(x - x_{i-1})(x - x_i)(x - x_{i+1})\dots(x - x_n). \quad (5.54)$$

Then

$$\Pi'(x_i) = \left[\frac{d}{dx} \Pi(x) \right]_{x=x_i} = (x_i - x_0)(x_i - x_1)\dots(x_i - x_{i-1})(x_i - x_{i+1})\dots(x_i - x_n)$$

and so equation (5.53) becomes

$$L_i(x) = \frac{\Pi(x)}{(x - x_i)\Pi'(x_i)}.$$

Hence, equation (5.51) becomes

$$P_n(x) = \sum_{i=0}^n \frac{\Pi(x)}{(x - x_i)\Pi'(x_i)} f_i, \quad (5.55)$$

which is called Lagrange's interpolation formula. The coefficients $L_i(x)$ defined in equation (5.53) are called Lagrange's interpolation coefficients.

Interchanging x and y in equation (5.55), we get the formula

$$P_n(y) = \sum_{i=0}^n \frac{\Pi(y)}{(y - y_i)\Pi'(y_i)} x_i, \quad (5.56)$$

which is useful for inverse interpolation.

Second Method: Let $f(x_0), f(x_1), \dots, f(x_n)$ be the values of the function f corresponding to the arguments x_0, \dots, x_n , not necessarily equally spaced. We wish to find a polynomial $P_n(x)$ in x of degree n such that

$$P_n(x_0) = f(x_0), P_n(x_1) = f(x_1), \dots, P_n(x_n) = f(x_n).$$

Suppose that

$$\begin{aligned} P_n(x) &= A_0(x - x_0)(x - x_1)\dots(x - x_n) + A_1(x - x_0)(x - x_1)\dots(x - x_n) \\ &\quad + A_2(x - x_0)(x - x_1)(x - x_2)\dots(x - x_n) + \dots \\ &\quad + A_n(x - x_0)(x - x_1)\dots(x - x_{n-1}). \end{aligned} \quad (5.57)$$

where A_0, A_1, \dots, A_n are the constants to be determined.

To determine A_0 , we put $x = x_0$ and $P_n(x_0) = f(x_0)$ and have

$$f(x_0) = A_0(x_0 - x_1)(x_0 - x_2)\dots(x_0 - x_n)$$

and so

$$A_0 = \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)\dots(x_0 - x_n)}.$$

Similarly, putting $x = x_1, \dots, x_n$, we get

$$A_1 = \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)\dots(x_1 - x_n)}$$

$$A_2 = \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)\dots(x_2 - x_n)}$$

.....

$$A_n = \frac{f(x_n)}{(x_n - x_0)(x_n - x_1)\dots(x_n - x_{n-1})}.$$

Substituting these values in equation (5.57), we get

$$P_n(x) = \sum_{i=0}^n L_i(x) f(x_i),$$

with

$$L_i(x) = \frac{(x - x_0)(x - x_1)\dots(x - x_{i-1})(x - x_{i+1})\dots(x - x_n)}{(x_i - x_0)(x_i - x_1)\dots(x_i - x_{i-1})(x_i - x_{i+1})\dots(x_i - x_n)},$$

which is Lagrange's interpolation formula.

Clearly,

$$L_i(x_j) = 0 \text{ for } i \neq j, \text{ and } L_i(x_i) = 1 \text{ for } i = j.$$

Remark 5.8. If f takes same value, say k , at each of the points x_0, \dots, x_n , we have

$$P_n(x) = \sum_{i=0}^n L_i(x) k = k \sum_{i=0}^n L_i(x).$$

This yields

$$\sum_{i=0}^n L_i(x) = 1,$$

which is an important check during calculations.

Further, dividing both sides of Lagrange's interpolation formula by $(x - x_0)(x - x_1)\dots(x - x_n)$, we obtain

$$\begin{aligned}\frac{P_n(x)}{(x - x_0)(x - x_1)\dots(x - x_n)} &= \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)\dots(x_0 - x_n)} \cdot \frac{1}{x - x_0} \\ &\quad + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)\dots(x_1 - x_n)} \cdot \frac{1}{x - x_1} \\ &\quad + \dots + \frac{f(x_n)}{(x_n - x_0)(x_n - x_1)\dots(x_n - x_{n-1})} \cdot \frac{1}{x - x_n}.\end{aligned}$$

Thus, $\frac{P_n(x)}{(x - x_0)(x - x_1)\dots(x - x_n)}$ has been expressed as the sum of partial fractions.

EXAMPLE 5.33

Use Lagrange's formula to express the function $\frac{x^2 + 6x - 1}{(x-1)(x+1)(x-4)(x-6)}$ as a sum of partial fractions.

Solution. We have

$$P_n(x) = x^2 + 6x - 1,$$

and so

$$P_n(x_0) = f(x_0) = f(1) = 6$$

$$P_n(x_1) = f(x_1) = f(-1) = -6$$

$$P_n(x_2) = f(x_2) = f(4) = 39$$

$$P_n(x_3) = f(x_3) = f(6) = 71.$$

Therefore,

$$\begin{aligned}\frac{x^2 + 6x - 1}{(x-1)(x+1)(x-4)(x-6)} &= \frac{6}{(x-1)(2)(-3)(-5)} \\ &\quad + \frac{-6}{(x+1)(-2)(-5)(-7)} + \frac{39}{(x-4)(3)(5)(-2)} + \frac{71}{(x-6)(5)(7)(2)} \\ &= \frac{1}{5(x-1)} + \frac{3}{35(x+1)} - \frac{13}{10(x-4)} + \frac{71}{70(x-6)}.\end{aligned}$$

EXAMPLE 5.34

Use Lagrange's interpolation formula to express the function

$$\frac{x^2 + x - 3}{x^3 - 2x^2 - x + 2}$$

as sum of partial functions.

Solution. We have

$$\frac{x^2 + x - 3}{x^3 - 2x^2 - x + 2} = \frac{x^2 + x - 3}{(x-1)(x+1)(x-2)}.$$

Let

$$P_n(x) = x^2 + x - 3,$$

and let $x_0 = 1, x_1 = -1, x_2 = 2$. Then

$$\begin{aligned} P_n(x_0) &= f(x_0) = f(1) = -1 \\ P_n(x_1) &= f(x_1) = f(-1) = -3 \\ P_n(x_2) &= f(x_2) = f(2) = 3. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{x^2 + x - 3}{(x-1)(x+1)(x-2)} &= \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)(x - x_0)} \cdot \frac{1}{(x-x_0)} \\ &\quad + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)(x - x_1)} \cdot \frac{1}{(x-x_1)} \\ &\quad + \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)(x - x_2)} \cdot \frac{1}{(x-x_2)} \\ &= \frac{-1}{(x-1)(2)(-1)} + \frac{-3}{(x+1)(-2)(-3)} + \frac{3}{(x-2)(1)(3)} \\ &= \frac{1}{2(x-1)} - \frac{1}{2(x+1)} + \frac{1}{(x-2)}. \end{aligned}$$

EXAMPLE 5.35

Using Lagrange's interpolation formula, prove that

$$32f(1) = -3f(-4) + 10f(-2) + 30f(2) - 5f(4).$$

Solution. We have

$$x_0 = -4, x_1 = -2, x_2 = 2, x_3 = 4 \text{ and } x = 1.$$

Then

$$L_0(x) = \frac{(x - x_1)(x - x_2)(x - x_3)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)} = \frac{(1+2)(1-2)(1-4)}{(-4+2)(-2-2)(-2-4)} = -\frac{3}{32}.$$

Similarly

$$L_1(x) = \frac{5}{16}, L_2(x) = \frac{15}{16}, L_3(x) = -\frac{5}{32}.$$

We observe that $\sum_{i=0}^3 L_i(x) = 1$. Therefore,

$$f(x) = \sum_{i=0}^3 L_i(x) f(x_i)$$

or

$$f(1) = \frac{5}{16} f_1 + \frac{15}{16} f_2 - \frac{5}{32} f_3 - \frac{3}{32} f_0$$

or

$$32f(1) = -3f(-4) + 10f(-2) + 30f(2) - 5f(4).$$

EXAMPLE 5.36

The function $y = f(x)$ is given in the points $(7,3)$, $(8,1)$, $(9,1)$, and $(10,9)$. Find the value of y for $x = 9.5$ using Lagrange's interpolation formula.

Solution. We have

	x	$y = f(x)$
x_0	7	3
x_1	8	1
x_2	9	1
x_3	10	9

By Lagrange's formula, we have

$$f(x) \approx P_n(x) = \sum_{i=0}^n L_i(x) f(x_i),$$

where

$$L_i(x) = \frac{(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}.$$

In the present problem, $x = 9.5$ and we have

$$\begin{aligned} L_0(x) &= \frac{(x - x_1)(x - x_3)}{(x_0 - x_1)(x_0 - x_3)} = \frac{(9.5 - 8)(9.5 - 9)(9.5 - 10)}{(7 - 8)(7 - 9)(7 - 10)} \\ &= \frac{0.375}{6} = 0.06250, \end{aligned}$$

$$\begin{aligned} L_1(x) &= \frac{(x - x_0)(x - x_3)}{(x_1 - x_0)(x_1 - x_3)} = \frac{(9.5 - 7)(9.5 - 9)(9.5 - 10)}{(8 - 7)(8 - 9)(8 - 10)} \\ &= -\frac{0.625}{2} = -0.3125, \end{aligned}$$

$$\begin{aligned} L_2(x) &= \frac{(x - x_0)(x - x_1)(x - x_3)}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)} = \frac{(9.5 - 7)(9.5 - 8)(9.5 - 10)}{(9 - 7)(9 - 8)(9 - 10)} \\ &= \frac{1.875}{2} = 0.9375, \end{aligned}$$

$$\begin{aligned} L_3(x) &= \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)} = \frac{(9.5 - 7)(9.5 - 8)(9.5 - 9)}{(10 - 7)(10 - 8)(10 - 9)} \\ &= \frac{1.875}{6} = 0.3125. \end{aligned}$$

We observe that $L_0(x) + L_1(x) + L_2(x) + L_3(x) = 1$ and therefore, so far, our calculations are correct. Hence,

$$\begin{aligned} P(x) = P(9.5) &= \sum_{i=0}^3 L_i(x) f(x_i) \\ &= L_{0,0} + L_1 f_1 + L_{2,2} + L_{3,3} \\ &= (0.06250)(3) - 0.3125(1) + 0.9395(1) + 0.3125(9) \\ &= 0.1875 - 0.3125 + 0.9375 + 2.8125 = 3.625. \end{aligned}$$

EXAMPLE 5.37

Find the interpolating polynomial for $(0, 2)$, $(1, 3)$, $(2, 12)$, and $(5, 147)$.

Solution. The given data is

x	0	1	2	5
$f(x)$	2	3	12	147

The Lagrange's formula reads

$$P_n(x) = \sum_{i=0}^n L_i(x) f(x_i),$$

where

$$L_i(x) = \frac{(x - x_0)(x - x_1)\dots(x - x_{i-1})(x - x_{i+1})\dots(x - x_n)}{(x_i - x_0)(x_i - x_1)\dots(x_i - x_{i-1})(x_i - x_{i+1})\dots(x_i - x_n)}.$$

Thus,

$$\begin{aligned} L_0(x) &= \frac{(x - 1)(x - 2)(x - 5)}{(x_0 - 0)(x_0 - 1)(x_0 - 2)(x_0 - 3)} = \frac{(x - 1)(x - 2)(x - 5)}{(0 - 1)(0 - 2)(0 - 5)} \\ &= -\frac{1}{10}(x^3 - 8x^2 + 17x - 10) \end{aligned}$$

$$\begin{aligned} L_1(x) &= \frac{(x - 0)(x - 2)(x - 5)}{(x_1 - 1)(x_1 - 2)(x_1 - 3)} = \frac{(x - 0)(x - 2)(x - 5)}{(1 - 0)(1 - 2)(1 - 5)} \\ &= \frac{1}{4}(x^3 - 7x^2 + 10x) \end{aligned}$$

$$\begin{aligned} L_2(x) &= \frac{(x - 0)(x - 1)(x - 5)}{(x_2 - 0)(x_2 - 1)(x_2 - 3)} = \frac{(x - 0)(x - 1)(x - 5)}{(2 - 0)(2 - 1)(2 - 5)} \\ &= -\frac{1}{6}(x^3 - 6x^2 + 5x) \end{aligned}$$

$$\begin{aligned} L_3(x) &= \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)} = \frac{(x - 0)(x - 1)(x - 2)}{(5 - 0)(5 - 1)(5 - 2)} \\ &= \frac{1}{60}(x^3 - 3x^2 + 2x). \end{aligned}$$

Putting these values in Lagrange's formula, we have

$$\begin{aligned} P(x) &= \sum_{i=0}^3 L_i(x)f(x_i) = -\frac{2}{10}(x^3 - 8x^2 + 17x - 10) + \frac{3}{4}(x^3 - 7x^2 + 10x) \\ &\quad - \frac{12}{6}(x^3 - 6x^2 + 5x) + \frac{147}{60}(x^3 - 3x^2 + 2x) = x^3 + x^2 - x + 2. \end{aligned}$$

EXAMPLE 5.38

Use Lagrange's interpolation formula to find the value of y when $x = 5$, if the following values of x and y are given:

x	1	2	3	4	7
y	2	4	8	16	128

Solution. Let $y_i = f(x_i)$ be the value of a function at x_i , $0 \leq i \leq n$. Then Lagrange's interpolating polynomial $P_n(x)$ is given by

$$P_n(x) = \sum_{i=0}^n L_i(x)f(x_i),$$

where

$$L_i(x) = \frac{(x - x_0)(x - x_1)\cdots(x - x_{i-1})(x - x_{i+1})\cdots(x - x_n)}{(x_i - x_0)(x_i - x_1)\cdots(x_i - x_{i-1})(x_i - x_{i+1})\cdots(x_i - x_n)}.$$

In the given problem, $x = 5$ and we have

$$\begin{aligned} L_0(x) &= \frac{(x - x_1)(x - x_2)(x - x_3)(x - x_4)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)(x_0 - x_4)} = \frac{(5 - 2)(5 - 3)(5 - 4)(5 - 7)}{(1 - 2)(1 - 3)(1 - 4)(1 - 7)} \\ &= -\frac{1}{3} \end{aligned}$$

$$\begin{aligned} L_1(x) &= \frac{(x - x_0)(x - x_2)(x - x_3)(x - x_4)}{(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)(x_1 - x_4)} = \frac{(5 - 1)(5 - 3)(5 - 4)(5 - 7)}{(2 - 1)(2 - 3)(2 - 4)(2 - 7)} \\ &= \frac{8}{5} \end{aligned}$$

$$\begin{aligned} L_2(x) &= \frac{(x - x_0)(x - x_1)(x - x_3)(x - x_4)}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)(x_2 - x_4)} = \frac{(5 - 1)(5 - 2)(5 - 4)(5 - 7)}{(3 - 1)(3 - 2)(3 - 4)(3 - 7)} \\ &= -3 \end{aligned}$$

$$L_3(x) = \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)} = \frac{(5-1)(5-2)(5-3)(5-4)(5-7)}{(4-1)(4-2)(4-3)(4-7)}$$

$$= \frac{8}{3}$$

$$L_4(x) = \frac{(x - x_0)(x - x_1)(x - x_2)(x - x_3)}{(x_4 - x_0)(x_4 - x_1)(x_4 - x_2)(x_4 - x_3)} = \frac{(5-1)(5-2)(5-3)(5-4)}{(7-1)(7-2)(7-3)(7-4)}$$

$$= \frac{1}{15}.$$

We note that $\sum_{i=0}^n L_i(x) = 1$. Hence, our calculations are correct up to this stage. By Lagrange's formula

$$\begin{aligned} P(x) &= r(5) = \sum_{i=0}^n L_i(x) f(x_i) \\ &= -\frac{1}{3}(2) + \frac{8}{5}(4) + (-3)(8) + \frac{8}{3}(16) + \frac{1}{15}(128) \\ &= -\frac{2}{3} + \frac{32}{5} - 24 + \frac{128}{3} + \frac{128}{15} \approx 32.933. \end{aligned}$$

5.12 ERROR IN LAGRANGE'S INTERPOLATION FORMULA

The error in this case is the difference between $f(x)$ and the Lagrange's polynomial $P_n(x)$ at a given point. Let $f(x) = P_n(x)$ at the $n+1$ points x_0, x_1, \dots, x_n . Suppose that the point X lies in the closed interval I bounded by the extreme points of (x_0, x_1, \dots, x_n) and further that $X \neq x_k$, $k = 0, 1, \dots, n$. Also, we assume that f can be differentiated $n+1$ times. We define the function

$$F(x) = f(x) - P_n(x) - R\Pi(x),$$

where $\Pi(x) = (x - x_0)(x - x_1)\dots(x - x_n)$ and R is a constant to be determined such that $F(X) = 0$. Obviously, $F(x) = 0$ for $x = x_0, x_1, \dots, x_n$. Using Rolle's Theorem repeatedly, we conclude that $F^{(n+1)}(\xi) = 0$, where $\xi \in I$. Since $P_n(x)$ is of degree n , we have

$$F^{(n+1)}(\xi) = f^{(n+1)}(\xi) - R(n+1)!$$

and so $F^{(n+1)}(\xi) = 0$ implies $R = \frac{f^{(n+1)}(\xi)}{(n+1)!}$. Thus, $F(X) = 0$ implies

$$f(X) - P_n(X) = R\Pi(X) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\Pi(X).$$

Replacing X by x , we get

$$f(x) - P_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\Pi(x)$$

or

$$\begin{aligned} f(x) &= P_n(x) + \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod(x) \\ &= \sum_{i=0}^n L_i(x) f(x_i) + \frac{f^{(n+1)}(\xi)}{(n+1)!} (x - x_0)(x - x_1) \dots (x - x_n). \end{aligned}$$

5.13 HERMITE INTERPOLATION FORMULA

So far, we considered interpolation formulae that make use only of a certain number of function values. We now derive one interpolation formula, called Hermite interpolation formula, in which both the function and its first derivative are to be assigned values at each point of interpolation. Interpolation of this kind is sometimes called osculating interpolation.

Suppose that the values of both f and f' are known for x_0, x_1, \dots, x_n . Since a polynomial of degree $2n+1$ is specified by $2n+2$ parameters, we can determine a polynomial $P_{2n+1}(x)$ in such a way that

$$\left. \begin{array}{l} P_{2n+1}(x_i) = f(x_i) \\ P'_{2n+1}(x_i) = f'(x_i) \end{array} \right\}, i = 0, 1, \dots, n. \quad (5.58)$$

Assume that $P_{2n+1}(x)$ is expressible in the form

$$P_{2n+1}(x) = \sum_{i=0}^n u_i(x) t_i(x_i) + \sum_{i=0}^n v_i(x) s_i'(x_i), \quad (5.59)$$

where $u_i(x)$ and $v_i(x)$ are polynomials in x of degree $2n+1$. The first part of condition (5.58) implies

$$u_i(x_j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

and

$$v_i(x_j) = 0.$$

The second part of condition (5.58) implies

$$u'_i(x_j) = 0 \text{ and } v'_i(x_j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (5.60)$$

Since

$$L_i(x) = \frac{(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}$$

is a polynomial of degree n , therefore $[L_i(x)]^2$ is of degree $2n$. Therefore,

$$\left. \begin{array}{l} u_i(x) = a_i(x)[L_i(x)]^2 \\ v_i(x) = b_i(x)[L_i(x)]^2 \end{array} \right\}, \quad (5.61)$$

where $a_i(x)$ and $b_i(x)$ are obviously linear functions. Using equations (5.60), we note that

$$u_i(x_i) = a_i(x_i)[L_i(x_i)]^2 \text{ implies } 1 = a_i(x_i) \cdot 1^2 = a_i(x_i)$$

$$\begin{aligned}
 v_i(x_i) &= r_i(x_i) L'_i(x_i)]^2 \text{ implies } 0 = b_i(x_i) - 1^2 = b_i(x_i) \\
 u'_i(x_i) &= a'_i(x_i)[L_i(x_i)]^2 + 2a_i(x_i)L_i(x_i)L'_i(x_i) \\
 &= a'_i(x_i) \cdot 2L'_i(x_i) \text{ implies } 0 = a'_i(x_i) \cdot 2r'_i(x_i) \\
 \text{or } a'_i(x_i) &= -2L'_i(x_i).
 \end{aligned}$$

Similarly,

$$r'_i(x_i) = 1.$$

It follows therefore that

$$a_i(x) = 1 - 2L'_i(x_i)(x - x_i)$$

and

$$r_i(x) = 1 - x - x_i.$$

Hence, equation (5.59) reduces to

$$\begin{aligned}
 P_{2n+1}(x) &= \sum_{i=0}^n \{[1 - 2L'_i(x_i)(x - x_i)][L_i(x)]^2 f_i\} \\
 &\quad + \sum_{i=0}^n \{(x - x_i)[L_i(x)]^2 f'_i\},
 \end{aligned}$$

which is the required Hermite's formula for interpolation.

If we have $n = 1$, that is, if we have two pairs of points $(x_0, y_0), (x_1, y_1)$, then Hermite's formula yields a cubic polynomial

$$P_3(x) = \sum_{i=0}^1 \{[1 - 2L_i(x_i)(x - x_i)][L_i(x)]^2 f(x_i)\} + \sum_{i=0}^1 \{(x - x_i)[L_i(x)]^2 f'(x_i)\}.$$

This formula gives rise to piecewise cubic Hermite's interpolation.

EXAMPLE 5.39

Determine the Hermite interpolating polynomial which fits the following data:

x	-1	0	1
y	-10	-4	-2
y'	10	3	2

Solution. We are given that

$$\begin{aligned}
 x_0 &= -1, \quad x_1 = 0, \quad x_2 = 1 \\
 f(x_0) &= -10, \quad f(x_1) = -4, \quad f(x_2) = -2 \\
 f'(x_0) &= 10, \quad f'(x_1) = 3, \quad f'(x_2) = 2.
 \end{aligned}$$

The Hermite interpolation formula is

$$P_{2n+1}(x) = \sum_{i=0}^n \{[1 - 2L'_i(x_i)(x - x_i)][L_i(x)]^2 f(x_i)\} + \sum_{i=0}^n \{(x - x_i)[L_i(x)]^2 f'(x_i)\},$$

where

$$L_i(x) = \frac{(x - x_0)(x - x_1)\dots(x - x_{i-1})(x - x_{i+1})\dots(x - x_n)}{(x_i - x_0)(x_i - x_1)\dots(x_i - x_{i-1})(x_i - x_{i+1})\dots(x_i - x_n)}.$$

In the present case, we have

$$L_0(x) = \frac{(x - x_0)(x - x_1)}{(x_0 - x_1)(x_0 - x_2)} = \frac{(x - 0)(x - 1)}{(-1 - 0)(-1 - 1)} = \frac{1}{2}x(x - 1),$$

$$L_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} = \frac{(x + 1)(x - 1)}{(1)(-1)} = -(x^2 - 1),$$

$$L_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} = \frac{(x + 1)(x)}{(2)(1)} = \frac{1}{2}x(x + 1),$$

$$L'_0(x) = x - \frac{1}{2} \text{ and so } L'_0(x_0) = -\frac{3}{2},$$

$$L'_1(x) = -2x \text{ and so } L'_1(x_1) = 0,$$

$$L'_2(x) = x + \frac{1}{2} \text{ and so } L'_2(x_2) = \frac{3}{2}.$$

Therefore, the Hermite interpolating polynomial is

$$\begin{aligned} P_{2n+1}(x) &= [1 - 2L'_0(x_0)(x - x_0)][L_0(x)]^2 f(x_0) \\ &\quad + [1 - 2L'_1(x_1)(x - x_1)][L_1(x)]^2 f(x_1) \\ &\quad + [1 - 2L'_2(x_2)(x - x_2)][L_2(x)]^2 f(x_2) \\ &\quad + (x - x_0)[L_0(x)]^2 f'(x_0) + (x - x_1)[L_1(x)]^2 f'(x_1) + (x - x_2)[L_2(x)]^2 f'(x_2) \\ &= -\frac{5}{2}(3x^5 - 2x^4 - 5x^3 + 4x^2) - 4(x^4 + 1 - 2x^2) \\ &\quad + \frac{1}{2}(5x^5 + 2x^4 - 5x^3 - 4x^2) + \frac{5}{2}(x^5 - x^4 - x^3 + x^2) \\ &\quad + 3(x^5 + x^4 - 2x^3) + \frac{1}{2}(x^5 + x^4 - x^3 - x^2) \\ &= x^5 - 2x^4 + 3x^3 - 4x^2 \end{aligned}$$

EXAMPLE 5.40

Determine the Hermite interpolating polynomial which fits the following data:

x	0	1	2
y	1	3	21
y'	0	6	36

Solution. It is given that

$$\begin{aligned}x_0 &= 0, \quad x_1 = 1, \quad x_2 = 2 \\f(x_0) &= 1, \quad f(x_1) = 3, \quad f(x_2) = 21 \\f'(x_0) &= 0, \quad f'(x_1) = 6, \quad f'(x_2) = 36.\end{aligned}$$

The Hermite interpolation formula is

$$\begin{aligned}P_{2n+1}(x) &= \sum_{i=0}^n \{[1 - 2L_i'(x_i)(x - x_i)][L_i(x)]^2 f(x_i)\} \\&\quad + \sum_{i=0}^n \{(x - x_i)[L_i(x)]^2 f'(x_i)\},\end{aligned}$$

where

$$L_i(x) = \frac{(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}.$$

For the given data, we have

$$L_0(x) = \frac{(x - x_0)(x - x_1)}{(x_0 - x_1)(x_0 - x_2)} = \frac{(x-1)(x-2)}{2} = \frac{1}{2}(x^2 - 3x + 2),$$

$$L_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} = \frac{x(x-2)}{-1} = -(x^2 - 2x),$$

$$L_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} = \frac{x(x-1)}{2} = \frac{1}{2}(x^2 - x),$$

$$L_0'(x) = \frac{1}{2}(x-3) \text{ and so } L_0'(x_0) = -\frac{3}{2},$$

$$L_1'(x) = -2x + 2 \text{ and so } L_1'(x_1) = 0,$$

$$L_2'(x) = \frac{1}{2}(x-1) \text{ and so } L_2'(x_2) = \frac{3}{2}.$$

Hence, the Hermite interpolating polynomial is

$$\begin{aligned}P_{2n+1}(x) &= [1 - 2L_0'(x_0)(x - x_0)][L_0(x)]^2 f(x_0) \\&\quad + [1 - 2L_1'(x_1)(x - x_1)][L_1(x)]^2 f(x_1) + [1 - 2L_2'(x_2)(x - x_2)][L_2(x)]^2 f(x_2) \\&\quad + (x - x_0)[L_0(x)]^2 f'(x_0) + (x - x_1)[L_1(x)]^2 f'(x_1) + (x - x_2)[L_2(x)]^2 f'(x_2) \\&= \frac{1}{4}[3x^5 - 11x^4 + 13x^3 - 23x^2 + 4] + 3[x^4 - 4x^3 + 4x^2] + \frac{21}{4}[-3x^5 + 13x^4 - 17x^3 + 7x^2] \\&\quad + 6[x^5 - 5x^4 + 8x^3 - 4x^2] + \frac{36}{4}[x^5 - 4x^4 + 5x^3 - 2x^2] = x^4 + x^2 + 1.\end{aligned}$$

EXAMPLE 5.41

Obtain piecewise cubic Hermite interpolation for the data of Example 5.40.

Solution. We first consider the interval $0 \leq x \leq 1$. So we have

$$x_0 = 0, x_1 = 1, f(x_0) = 1, f(x_1) = 3, f'(x_0) = 0, f'(x_1) = 6.$$

Further,

$$L_0(x) = \frac{x - x_1}{x_0 - x_1} = 1 - x,$$

$$L_1(x) = \frac{x - x_0}{x_1 - x_0} = x,$$

$$L_0'(x) = -1 \text{ and so } L_0'(x_0) = -1,$$

$$L_1'(x) = 1 \text{ and so } L_1'(x_1) = 1.$$

Therefore for this interval, we get

$$\begin{aligned} P_3(x) &= \sum_{i=0}^1 \{[1 - 2L_i'(x_i)(x - x_i)][L_i(x)]^2 f(x_i)\} \\ &\quad + \sum_{i=0}^1 \{(x - x_i)[L_i(x)]^2 f'(x_i)\} \\ &= [1 - 2(-1)(x - 0)][1 - x]^2(1) + [1 - (-1)(x - 1)][x - 1]^2(3) \\ &\quad + (x - 0)(1 - x)^2(0) + (x - 1)(x^2)(6) \\ &= (1 + 2x)(1 + x^2 - 2x) + 3(3 - 2x)x^2 + 6x^2(x - 1) = 2x^3 + 1. \end{aligned}$$

Now for the interval $1 \leq x \leq 2$, we have

$$x_0 = 1, x_1 = 2, f(x_0) = 3, f(x_1) = 21, f'(x_0) = 6, f'(x_1) = 36$$

and

$$L_0(x) = \frac{x - x_1}{x_0 - x_1} = \frac{x - 2}{-1} = 2 - x,$$

$$L_1(x) = \frac{x - x_0}{x_1 - x_0} = \frac{x - 1}{1} = x - 1,$$

$$L_0'(x) = -1 \text{ and so } L_0'(x_0) = -1,$$

$$L_1'(x) = 1 \text{ and so } L_1'(x_1) = 1.$$

Therefore, for this interval

$$\begin{aligned} P_3(x) &= \{1 - 2(-1)(x - 1)\}(2 - x)^2(3) + \{1 - (-1)(x - 2)\}\{(x - 1)^2\}(21) \\ &\quad + (x - 1)(2 - x)^2(6) + (x - 2)(x - 1)^2(36) \\ &= 6x^3 - 12x^2 + 12x - 3. \end{aligned}$$

Hence, the required piecewise cubic Hermite polynomial is

$$P_3(x) = \begin{cases} 2x^3 + 1, & 0 \leq x \leq 1 \\ 6x^2 - 12x^2 + 12x - 3, & 1 \leq x \leq 2. \end{cases}$$

5.14 THROWBACK TECHNIQUE

A general technique of converting higher-order differences into lower-order differences, thereby reducing the computational work on interpolation is called throwback technique.

Consider Everett's formula

$$\begin{aligned} y_p &= py_1 + \binom{p+1}{3} \delta^2 y_1 + \binom{p+2}{5} \delta^4 y_1 + \dots \\ &\quad + qy_0 + \binom{q+1}{3} \delta^2 y_0 + \binom{q+2}{5} \delta^4 y_0 + \dots \\ &= py_1 + \binom{p+1}{3} \left[\delta^2 y_1 + \frac{p^2 - 4}{20} \delta^4 y_1 \right] + \dots \\ &\quad + qy_0 + \binom{q+1}{3} \left[\delta^2 y_0 + \frac{q^2 - 4}{20} \delta^4 y_0 \right] + \dots \end{aligned}$$

When p varies from 0 to 1, the factor $\frac{p^2 - 4}{20}$ varies between 0.25 and 0.20. We replace this factor by a constant, which still remains to be chosen.

Define modified differences $\delta_m^2 y$ by

$$\delta_m^2 y_0 = \sigma^2 y_0 - C \delta^4 y_0, \quad \delta_m^2 y_1 = \sigma^2 y_1 - C \delta^4 y_1.$$

Thus, modified interpolation formula becomes

$$y_p = py_1 + \binom{p+1}{3} \delta_{m^2}^2 y_1 + \dots + qy_0 + \binom{q+1}{3} \delta_m^2 y_0 + \dots$$

If the sixth and higher-order terms are neglected and if the error is denoted by E , then

$$E(p) = \left[\binom{q+2}{5} + C \binom{q+1}{3} \right] \delta^4 y_0 + \left[\binom{p+2}{5} + C \binom{p+1}{3} \right] \delta^4 y_1.$$

Suppose that $\delta^4 y$ varies so slowly that $\delta^4 y_0$ and $\delta^4 y_1$ do not differ appreciably. Using $q = 1 - p$, we obtain

$$E(p) = \left[\frac{p(p-1)}{24} (p^2 - p + 12C - 2) \right] \delta^4 y = \varphi(p) \delta^4 y,$$

where

$$\begin{aligned} \varphi(p) &= \frac{p(p-1)}{24} (p^2 - p + 12C - 2) \\ &= \frac{1}{24} [(p^2 - p)(p^2 - p + 12C - 2)] \\ &= \frac{1}{24} [P(P + 2k)] = \frac{1}{24} [P^2 + 2Pk], \end{aligned}$$

where

$$P = n^2 - p \text{ and } k = 6c - 1.$$

Differentiating with respect to p and equating to zero, we obtain

$$(2P + 2k)(2p - 1) = 0.$$

Then either $p = \frac{1}{2}$ which gives $P(P+2k)$ a maximum value $\frac{1-8k}{16}$ or $P = -k$ which gives a minimum value equal to $-k^2$ at the two points given by $p^2 - n + 2k = 0$. Choose k so that the maximum and minimum values are equal in magnitude, that is,

$$\frac{1-8k}{16} = k^2 \text{ or } 16k^2 + 8k - 1 = 0,$$

which yields $k = \frac{-1 \pm \sqrt{2}}{4}$. Since we want the two values of p which give the minimum value to $\phi(p)$ to be between 0 and 1, it follows from $p^2 - n + 2k = 0$ that k is positive. Thus,

$$k = \frac{-1 + \sqrt{2}}{4}, \text{ that is, } 6C - 1 = \frac{-1 + \sqrt{2}}{4},$$

or

$$C = \frac{1 - \frac{1 + \sqrt{2}}{4}}{6} = \frac{3 + \sqrt{2}}{24} = 0.1839.$$

Hence,

$$\begin{aligned} y_p &= ny_1 + \binom{p+1}{3} \delta_m^2 y_1 + \dots \\ &\quad + qy_0 + \binom{p+1}{3} \delta_m^2 y_0 + \dots \end{aligned}$$

where

$$\begin{aligned} \delta_m^2 y_n &= \delta^2 y_n - 0.1839 \delta^4 y_0 \\ \delta_m^2 y_{n-1} &= \delta^2 y_{n-1} - 0.1839 \delta^4 y_0. \end{aligned}$$

Consider now the Bessel's formula and restrict ourselves to throw back the fourth difference on the second. The Bessel's formula is

$$\begin{aligned} y_p &= y_0 + p\delta_{\frac{1}{2}} y_{\frac{1}{2}} + \frac{p(p-1)}{2!} \left(\frac{\delta^2 y_0 + \delta^2 y_1}{2} \right) \\ &\quad + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 y_{\frac{1}{2}} + \frac{(p+1)p(p-1)(p-2)}{4!} \left(\frac{\delta^4 y_0 + \delta^4 y_1}{2} \right) + \dots \\ &= y_0 + p\delta_{\frac{1}{2}} y_{\frac{1}{2}} + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 y_{\frac{1}{2}} + \frac{p(p-1)}{2(2)!} (\delta^2 y_0 + \delta^2 y_1) \\ &\quad + \frac{(p+1)(p-2)}{2!} (\delta^4 y_0 + \delta^4 y_1) + \dots \end{aligned}$$

We now define modified differences by $\delta_m^2 y$ by

$$\begin{aligned}\delta_m^2 y_0 &= \delta^2 y_0 - C\delta^4 y_1 \\ \delta_m^2 y_1 &= \delta^2 y_1 - C\delta^4 y_0.\end{aligned}$$

The modified interpolation formula then takes the form

$$y_p = v_0 + \frac{\delta v_1}{2} + \frac{p\left(p-\frac{1}{2}\right)(p-1)}{3!} \delta^3 y_{\frac{1}{2}} + \frac{p(p-1)}{2(2)!} (\delta_m^2 y_0 + \delta_m^2 y_1) + \dots$$

If the fifth and higher-order terms are neglected, then the error term is

$$E(p) = \frac{p(p-1)}{2!} \left[\frac{(p+1)(p-2)}{12} + C \right] \delta^4 y,$$

under the assumption that $\Delta^4 y_0$ and $\Delta^4 y_1$ differ slightly. Thus,

$$E(p) = \phi(p) \frac{\delta^4 y}{24},$$

where

$$\phi(p) = p(p-1)[p^2 - p + 12C - 2].$$

The next steps are similar to those for Everett's formula discussed above.

EXAMPLE 5.42

Determine the constants a, b, c , and d in such a way that the formula

$$y_p = av_0 + bv_1 + h^2(cy''_0 + dy''_1)$$

becomes correct to the highest possible order.

Solution. We have

$$y_p = av_0 + bv_1 + h^2(cy''_0 + dy''_1).$$

Converting both sides to shifting operator E , we get

$$E^p y_0 = ay_0 + bE y_0 + ch^2 D^2 v_0 + h^2 D^2 E y_0$$

and so

$$\begin{aligned}E^p &= a + bE + ch^2 D^2 + h^2 D^2 E \\ &= a + bE + c(\log E)^2 + d(\log E)^2 E.\end{aligned}$$

Then

$$\begin{aligned}\text{L.H.S.} &= (1 + \Delta)^p = 1 + p\Delta + \frac{p(p-1)}{2} \Delta^2 + \frac{p(p-1)(p-2)}{3!} \Delta^3 \\ &\quad + \frac{p(p-1)(p-2)(p-3)}{4!} \Delta^4 + \dots\end{aligned}\tag{5.62}$$

and

$$\begin{aligned}
 \text{R.H.S.} &= a + b(I + \Delta) + c[\log(1 + \Delta)]^2 + d[\log(1 + \Delta)]^2(1 + \Delta) \\
 &= a + b(1 + \Delta) + c\left[\Delta - \frac{\Delta^2}{2} + \dots\right]^2 + d\left[\Delta - \frac{\Delta^2}{2} + \dots\right]^2(1 + \Delta) \\
 &= (a + b) + b\Delta + c\left[\Delta^2 + \frac{\Delta^4}{4} - \Delta^3 + \dots\right] \\
 &\quad + d(1 + \Delta)\left(\Delta^2 + \frac{\Delta^4}{4} - \Delta^3 + \dots\right)
 \end{aligned} \tag{5.63}$$

Comparing equations (5.62) and (5.63), we get

$$\begin{aligned}
 a + b &= 1, b = n \text{ and so } a = 1 - n = q \\
 \binom{p}{2} &= c + d \text{ and } \binom{p}{3} = -c - a + d = -c.
 \end{aligned}$$

Now

$$\binom{p}{3} = -c \text{ implies } c = \frac{-pq(q+1)}{6}.$$

and

$$\begin{aligned}
 c + d &= \binom{p}{2} \text{ implies } d = \binom{p}{2} + \binom{p}{3} \\
 &= -\frac{pq}{6}(p+1).
 \end{aligned}$$

Hence,

$$a = \alpha = 1 - p, \quad b = p, \quad c = \frac{-pq(q+1)}{6}, \quad d = \frac{-pq}{6}(p+1).$$

5.15 INVERSE INTERPOLATION

Inverse interpolation is the process of finding the value of the argument to a given value of the function when the latter is intermediate between two tabulated values.

(A) Inverse Interpolation Using Newton's Forward Difference Formula

Let $\dots, f_{-3}, f_{-2}, f_{-1}, f_0, f_1, f_2, f_3, \dots$ be the functional values of a function f at $\dots, x_{-3}, x_{-2}, x_{-1}, x_0, x_1, x_2, x_3, \dots$. Then Newton's forward difference formula reads as

$$f_p = f_0 + p\Delta f_0 + \frac{p(p-1)}{2!} \Delta^2 f_0 + \frac{p(p-1)(p-2)}{3!} \Delta^3 f_0 + \dots \tag{5.64}$$

We want to find value of x between x_0 and $x_1 = x_0 + h$ such that $f(x) = f_p$, where f_p is a given value. As before, we denote $\frac{x - x_0}{h} = p$, that is, $x = x_0 + ph$. Thus, our aim is to find p to get the value of x . From equation (5.64), we have

$$p = \frac{1}{\Delta f_0} \left[f_p - f_0 - \frac{p(p-1)}{2!} \Delta^2 f_0 - \frac{p(p-1)(p-2)}{3!} \Delta^3 f_0 - \dots \right].$$

We use iteration technique to find p . So, we first neglect the second and higher differences and find first approximation to p as

$$p_1 = \frac{1}{\Delta f_0} (\tau_p - \tau_0).$$

Now the approximate value p_1 of p is inserted in the second difference term to get

$$p_2 = \frac{1}{\Delta f_0} \left[\tau_p - \tau_0 - \frac{p_1(p_1-1)}{2!} \Delta^2 f_0 \right].$$

Next retaining the term with third difference, we get

$$p_3 = \frac{1}{\Delta f_0} \left[\tau_p - \tau_0 - \frac{p_2(p_2-1)}{2!} \Delta^2 f_0 - \frac{p_2(p_2-1)(p_2-2)}{3!} \Delta^3 f_0 \right].$$

The process is carried out till two successive approximations of p agree with each other up to desired accuracy. Then

$$x = x_0 + p_n h.$$

EXAMPLE 5.43

Find the value of x for $f(x) = 10$ using the following table:

x	2	3	4	5
$f(x)$	8	27	64	125

Solution. The difference table for the given data is

x	$f(x)$	Δf	$\Delta^2 f$	$\Delta^3 f$
2	8	19		
3	27	37	18	6
4	64	61	24	
5	125			

We are given that $f_p = 10$, $h = 1$, $f_0 = 8$, $x_0 = 2$, $\frac{x - x_0}{h} = p$. Using Newton's forward differences, the first approximation to p is

$$p_1 = \frac{1}{\Delta f_0} [\tau_p - \tau_0] = \frac{1}{19}(10 - 8) = 0.1.$$

The second approximation to p is

$$\begin{aligned} p_2 &= \frac{1}{\Delta f_0} \left[\tau_p - \tau_0 - \frac{p_1(p_1-1)}{2!} \Delta^2 f_0 \right] \\ &= \frac{1}{19} \left[10 - 8 - \frac{0.1(0.1-1)}{2} (18) \right] = 0.15 \end{aligned}$$

The third approximation to p is

$$\begin{aligned}
p_3 &= \frac{1}{\Delta f_0} \left[f_p - f_0 - \frac{p_2(p_2-1)}{2!} \Delta^2 f_0 - \frac{p_2(p_2-1)(p_2-2)}{3!} \Delta^3 f_0 \right] \\
&= \frac{1}{19} \left[10 - 8 - \frac{0.15(0.15-1)}{2} (18) - \frac{0.15(0.15-\cdot)(0.15-2)}{6} (6) \right] \\
&= 0.1532.
\end{aligned}$$

The fourth approximation (using the same available differences but replacing p_2 by p_3) is

$$\begin{aligned}
p_4 &= \frac{1}{\Delta f_0} \left[f_p - f_0 - \frac{p_3(p_3-1)}{2} \Delta^2 f_0 - \frac{p_3(p_3-1)(p_3-2)}{6} \Delta^3 f_0 \right] \\
&= \frac{1}{4} \left[10 - 8 - \frac{0.1532(0.1532)}{2} (18) - \frac{0.1532(0.1532-1)(0.1532-2)}{6} (6) \right] \\
&= 0.1541.
\end{aligned}$$

The next approximation is

$$p_5 = 0.1542.$$

Thus, $p = 0.154$ correct to three decimal places.

Hence,

$$x = x_0 + ph = 2 + 0.154(1) = 2.154.$$

(B) Inverse Interpolation Using Everett's Formula

Let a function f be tabulated with $\delta^2 f$ and $\delta^4 f$. We want to find a value x between x_0 and $x_1 = x_0 + h$ such that $f(x) = f_p$ where f_p is a given value. If $p = \frac{x - x_0}{h}$, then by Everett's formula, we have

$$\begin{aligned}
f_p &= p f_1 + \binom{p+1}{3} \delta^2 f_1 + \binom{p+2}{5} \delta^4 f_1 \\
&\quad + (1-p) f_0 + \binom{2-p}{3} \delta^2 f_0 + \binom{3-p}{5} \delta^4 f_0.
\end{aligned}$$

To determine the first approximation to p , we have

$$p_1 s_1 + (1-p_1) s_0 = f_p.$$

This approximated value is inserted into $\delta^2 f$ terms and we have second approximation p_2 given by

$$p_2 s_1 + (1-p_2) s_0 = f_p - \binom{p_1+1}{3} \delta^2 f_1 - \binom{2-p_1}{3} \delta^2 f_0.$$

Next, we obtain p_3 from

$$\begin{aligned}
p_3 s_1 + (1-p_3) s_0 &= f_p - \binom{p_2+1}{3} \delta^2 f_1 - \binom{2-p_2}{3} \delta^2 f_0 \\
&\quad - \binom{p_2+2}{5} \delta^4 f_1 - \binom{3-p_2}{5} \delta^4 f_0.
\end{aligned}$$

If necessary, the process is repeated until we get value to the required accuracy.

EXAMPLE 5.44

The function $y = \log(x!)$ has a minimum between 0 and 1. Find the abscissa from the data below:

x	$\frac{d}{dx} \log(x!)$	δ^2	δ^4
0.46	-0.0015805620	-0.0000888096	-0.0000000396
0.47	0.0080664890	-0.0000872716	-0.0000000383

Solution. The problem is clearly of inverse interpolation. We are provided with even differences and therefore Everett's formula is to be used. The relevant terms in the Everett's formula are

$$\begin{aligned} p &= (1-p)f_0 + \binom{2-p}{3}\delta^2 f_0 + \binom{3-p}{5}\delta^4 f_0 \\ &\quad + pf_1 + \binom{p+1}{3}\delta^2 f_1 + \binom{p+2}{5}\delta^4 f_1. \end{aligned}$$

For minimum, $\frac{d}{dx} \log(x!) = 0$. We choose $x_0 = 0.46$, $x_1 = 0.47$, $f_0 = -0.0015805620$, and $f_1 = 0.0080664890$. Therefore,

$$\begin{aligned} 0 &= (1-p)f_0 + pf_1 + \binom{2-p}{3}\delta^2 f_0 + \binom{p+1}{3}\delta^2 f_1 \\ &\quad + \binom{3-p}{5}\delta^4 f_0 + \binom{p+2}{5}\delta^4 f_1. \end{aligned}$$

First we determine a value p_1 from the equation

$$p_{1,1} + (1-p_1)_{0,0} = 0$$

and get

$$\begin{aligned} p_1 &= -\frac{f_0}{f_1 - f_0} = \frac{0.0015805620}{0.0096470510} \\ &= 0.16383887677. \end{aligned}$$

This value is inserted in the $\delta^2 f$ terms while $\delta^4 f$ terms are neglected. Then we obtain a value p_2 from

$$p_{2,1} + (1-p_2)_{0,0} = -\binom{p_1+1}{3}\delta^2 f_1 - \binom{2-p_1}{3}\delta^2 f_0$$

and so

$$\begin{aligned} p_2 &= \frac{1}{0.0096470510} \left[0.0015805620 - \frac{1}{6}(p_1^3 - p_1)\delta^2 f_0 + \frac{1}{6}(-p_1^3 + 3p_1^2 - 2p_1)\delta^2 f_1 \right] \\ &= 0.163219205537. \end{aligned}$$

Next inserting the value of p_2 in $\Delta^4 f$ terms, we obtain $p_3 = 0.16321441$. The value is correct to five decimal places. We have $h = 0.01$. Hence,

$$x = x_0 + ph = 0.46 + 0.00163321 = 0.46163321.$$

(C) Inverse Interpolation Using Lagrange's Interpolation Formula

While deriving Lagrange's formula, we observed that it is a relation between two variables either of which may be taken as independent variables. Therefore, interchanging x and f in the Lagrange's formula, we have

$$x = \sum_{i=0}^n L_i(f)x_i,$$

where

$$L_i(f) = \frac{(f - f_0)(f - f_1)\dots(f - f_{i-1})(f - f_{i+1})\dots(f - f_n)}{(f_i - f_0)(f_i - f_1)\dots(f_i - f_{i-1})(f_i - f_{i+1})\dots(f_i - f_n)}.$$

EXAMPLE 5.45

Apply Lagrange's formula inversely to obtain the root of the equation $f(x) = 0$ given that

$$f(30) = -30, f(34) = -13, f(38) = 3 \text{ and } f(42) = 180.$$

Solution. Since $f(34) = -13$ and $f(38) = 3$, the root lies between 34 and 38. We have

	x	$f(x)$
x_0	30	-30
x_1	34	-13
x_2	38	3
x_3	42	18

In the present case $f(x) = 0$. Therefore,

$$_0(f) = \frac{(f - f_1)(f - f_2)(f - f_3)}{(f_0 - f_1)(f_0 - f_2)(f_0 - f_3)} = \frac{(0+13)(0-3)(0-18)}{(-30+13)(-30-3)(-3-18)}$$

$$= \frac{702}{-26928} = -0.0261$$

$$_1(f) = \frac{(f - f_0)(f - f_2)(f - f_3)}{(f_1 - f_0)(f_1 - f_2)(f_1 - f_3)} = \frac{(30)(-3)(-18)}{(17)(-16)(-31)} = \frac{1620}{8432} = 0.192$$

$$_2(f) = \frac{(f - f_0)(f - f_1)(f - f_3)}{(f_2 - f_0)(f_2 - f_1)(f_2 - f_3)} = \frac{(30)(13)(-18)}{(33)(16)(-15)} = \frac{7020}{7920} = 0.8864$$

$$_3(f) = \frac{(f - f_0)(f - f_1)(f - f_2)}{(f_3 - f_0)(f_3 - f_1)(f_3 - f_2)} = \frac{(30)(13)(-3)}{(48)(31)(15)} = -\frac{1170}{22320} = -0.0524.$$

Therefore,

$$x = L_0x_0 + L_1x_1 + L_2x_2 + L_3x_3 = -0.7830 \cdot 6.5314 + 33.6832 \cdot -0.0261 + 0.192 \cdot 0.8864 - 0.0524 \cdot 37.231 = 37.231.$$

EXAMPLE 5.46

A function f is known in three points x_1 , x_2 , and x_3 in the vicinity of an extreme point x_0 . Show that

$$x_0 \approx \frac{x_1 + 2x_2 + x_3}{4} - \frac{f(x_1, x_2) + f(x_2, x_3)}{4f(x_1, x_2, x_3)}.$$

Use this formula to find x_0 when the following values are known:

x	3.00	3.6	3.8
f	0.13515	0.83059	0.26253

Solution. By Newton's divided difference formula, we have

$$f(x) = f(x_1) + (x - x_1) f(x_1, x_2) + (x - x_1)(x - x_2) f(x_1, x_2, x_3).$$

Now x_0 is given to be an extreme point, therefore derivative at x_0 vanishes. Therefore,

$$\begin{aligned} 0 &= [f'(x)]_{x=x_0} = [0 + f(x_1, x_2) + 2x f(x_1, x_2, x_3) - (x_1 + x_2) f(x_1, x_2, x_3)]_{x=x_0} \\ &= f(x_1, x_2) + 2x_0 f(x_1, x_2, x_3) - (x_1 + x_2) f(x_1, x_2, x_3) \end{aligned}$$

and so

$$2x_0 f(x_1, x_2, x_3) = (x_1 + x_2) f(x_1, x_2, x_3) - f(x_1, x_2)$$

which yields

$$\begin{aligned} x_0 &= \frac{x_1 + x_2}{2} - \frac{f(x_1, x_2)}{2f(x_1, x_2, x_3)} \\ &= \frac{x_1 + 2x_2 + x_3}{4} - \frac{f(x_1, x_2)}{2f(x_1, x_2, x_3)} - \frac{x_3 - x_1}{4} \\ &= \frac{x_1 + 2x_2 + x_3}{4} - \frac{2f(x_1, x_2) + f(x_1, x_2, x_3)(x_3 - x_1)}{4f(x_1, x_2, x_3)} \\ &= \frac{x_1 + 2x_2 + x_3}{4} - \frac{2f(x_1, x_2) + (x_3 - x_1) \left[\frac{f(x_2, x_3) - f(x_1, x_2)}{x_3 - x_1} \right]}{4f(x_1, x_2, x_3)} \\ &= \frac{x_1 + 2x_2 + x_3}{4} - \frac{f(x_1, x_2) + f(x_2, x_3)}{4f(x_1, x_2, x_3)}. \end{aligned} \tag{5.65}$$

Further, we have

$$f(x_1, x_2) = \frac{f(x_2) - f(x_1)}{x_2 - x_1} = \frac{0.83059 - 0.13515}{0.6} = 1.15906$$

$$f(x_2, x_3) = \frac{f(x_3) - f(x_2)}{x_3 - x_2} = \frac{0.26253 - 0.83059}{0.2} = -2.84030$$

$$f(x_1, x_2, x_3) = \frac{f(x_2, x_3) - f(x_1, x_2)}{x_3 - x_1} = \frac{-2.84030 - 1.15906}{0.8} = -4.99920.$$

Putting these values in equation (5.65), we have

$$x_0 = 3.4915925.$$

EXAMPLE 5.47

The equation $x^3 - 15x + 4 = 0$ has a root close to 0.3. Obtain this root with six decimal places using inverse interpolation (for example, with Bessel's interpolation formula).

Solution. Taking h to be 0.02, we tabulate the values as below:

	x	$f(x)$	δ	δ^2	δ^3	δ^4
x_2	0.22	0.710648				
x_{-1}	0.24	0.413824	-0.296824		0.000576	
x_0	0.26	0.117576	-0.296248		0.000624	0.000048
x_1	0.28	-0.178048	-0.295624		0.000672	0.000048
x_2	0.30	-0.47300	-0.294952		0.000720	0.000048
x_3	0.32	-0.767232	-0.294232		0.000768	0.000048
x_4	0.34	-1.060696	-0.293464		0.000816	0.000048
x_5	0.36	-1.353344	-0.292648		0.000864	0.000048
x_6	0.38	-1.645128	-0.291784			

It is visible from the table that the root lies between 0.26 and 0.28. Therefore, we take 0.26 to be x_0 . The Bessel's formula reads as

$$p = t_0 + p\delta_{\frac{1}{2}} + \frac{p(p-1)}{2!} \left(\frac{\delta^2 f_0 + \delta^2_{-1}}{2} \right) + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{\frac{1}{2}} + \dots$$

where $p = \frac{x - x_0}{h}$. Using first order difference, we get the first approximation p_1 from the equation

$$0 = f_0 + p_1 \delta_{\frac{1}{2}}$$

which yields

$$p_1 = -\frac{f_0}{\delta f_{\frac{1}{2}}} = \frac{0.117576}{0.295624} = 0.39772.$$

This value is inserted in $\delta^2 f$ terms and p_2 is obtained from the equation

$$\begin{aligned} 0 &= f_0 + p_2 \delta f_{\frac{1}{2}} + \frac{p_1^2 - p_2}{4} \left(\frac{\delta^2 f_0 + \delta^2 f_1}{2} \right) \\ &= 0.117576 + p_2(-0.295624) + \frac{0.001296}{4}(-0.23915) \end{aligned}$$

and so

$$p_2 = \frac{-0.117576 - 0.00007755}{-0.295624} = 0.39746$$

Inserting the value of p_2 in $\delta^3 y$ term, we have the next approximation p_3 given by

$$\begin{aligned} 0 &= f_0 + p_3 \delta f_{\frac{1}{2}} + \frac{p_2^2 - p_3}{4}(-0.001296) \\ &\quad + \frac{(p_2^2 - p_3)}{3!} \left(p_2 - \frac{1}{2} \right) (0.000048) \end{aligned}$$

which yields

$$p_3 = 0.39753.$$

Thus,

$$x = x_0 + ph = 0.26 + 0.3975(0.02) = 0.26 + 0.00796 = 0.267950.$$

5.16 CHEBYSHEV POLYNOMIALS

The aim of this section is to study polynomial interpolation for a function $f(x)$ over $[-1, 1]$ based on the nodes (arguments) $-1 \leq x_0 < x_1 < \dots < x_n \leq 1$.

We know that both the Newton's polynomial and Lagrange's polynomials satisfy

$$f(x) = P_n(x) + E_n(x),$$

where

$$E_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi) \Pi(x),$$

and

$$\Pi(x) = (x - x_0)(x - x_1) \dots (x - x_n)$$

is a polynomial of degree $n+1$. We have (for $[-1, 1]$)

$$|E_n(x)| \leq |\Pi(x)| \max_{-1 \leq x \leq 1} \frac{|f^{(n+1)}(x)|}{(n+1)!}.$$

The aim of our study is to select the set of nodes $\{x_k\}$ which minimizes $\max_{-1 \leq x \leq 1} |\Pi(x)|$. For the purpose, we introduce the concept of Chebyshev polynomial.

The Chebyshev polynomial of degree n over the interval $[-1, 1]$ is defined by the relation

$$T_n(x) = \cos(n \cos^{-1} x). \quad (5.66)$$

We note that

$$T_n(x) = T_{-n}(x). \quad (5.67)$$

Putting $\cos^{-1} x = \theta$ in equation (5.66), we have

$$T_n(x) = \cos n\theta, \quad (5.68)$$

which yields

$$T_0(x) = 1 \text{ and } T_1(x) = \cos \theta = \cos(\cos^{-1} x) = x.$$

Further, using equation (5.68), we have

$$\begin{aligned} T_{n-1}(x) &= \cos(n-1)\theta \\ T_{n+1}(x) &= \cos(n+1)\theta \end{aligned}$$

and so

$$\begin{aligned} T_{n-1}(x) + T_{n+1}(x) &= \cos(n-1)\theta + \cos(n+1)\theta \\ &= 2\cos n\theta \cos \theta \\ &= 2x \cos n\theta = 2x T_n(x). \end{aligned}$$

Thus, the recurrence relation to compute the values of $T_n(x)$ is

$$T_{n+1}(x) = 2x T_n(x) - T_{n-1}(x), \quad n \geq 1. \quad (5.69)$$

The first eight Chebyshev polynomials are

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x \\ T_4(x) &= 8x^4 - 8x^2 + 1 \\ T_5(x) &= 16x^5 - 20x^3 + 5x \\ T_6(x) &= 32x^6 - 48x^4 + 18x^2 - 1 \\ T_7(x) &= 64x^7 - 112x^5 + 56x^3 - 7x. \end{aligned}$$

We note that

- (i) Coefficient of x^n in $T_n(x)$ is 2^{n-1} when $n \geq 1$.
- (ii) When $n = 2m$, then T_{2m} is an even function. For example, $T_2(x) = 2x^2 - 1$ and so $T_2(-x) = 2x^2 - 1 = T_2(x)$. On the other hand, if $n = 2m+1$, then T_{2m+1} is an odd function.
- (iii) $T_n(x)$ has n distinct zeros x_k in the interval $[-1, 1]$ given by

$$x_k = \cos\left(\frac{(2k+1)\pi}{2n}\right), \quad k = 0, 1, \dots, n-1.$$

These zeros are called the Chebyshev abscissas (nodes).

- (iv) Relation (5.68) yields

$$|T_n(x)| \leq 1 \text{ for } -1 < x \leq 1.$$

(v) If

$$y = T_n(x) = \cos n\theta,$$

then

$$\begin{aligned}\frac{dy}{dx} &= \frac{dy}{d\theta} \cdot \frac{d\theta}{dx} = -n \sin n\theta \left(\frac{1}{-\sin \theta} \right) \\ &= \frac{n \sin n\theta}{\sin \theta}\end{aligned}$$

and

$$\begin{aligned}\frac{d^2y}{dx^2} &= \frac{-n^2 \cos n\theta + n \sin n\theta \cot \theta}{\sin^2 \theta} \\ &= -\frac{n^2 y + x \frac{dy}{dx}}{1 - x^2}.\end{aligned}$$

Hence, the Chebyshev polynomial satisfies the differential equation

$$(1 - x^2) \cdot \frac{d^2y}{dx^2} - x \frac{dy}{dx} + n^2 y = 0. \quad (5.70)$$

(vi) The expressions for $T_0(x)$, $T_1(x)$, ... yield

$$x^0 = 1 = T_0(x)$$

$$x = T_1(x)$$

$$x^2 = \frac{1}{2}[T_0(x) + T_2(x)]$$

$$x^3 = \frac{1}{4}[3T_1(x) + T_3(x)]$$

$$x^4 = \frac{1}{8}[3T_0(x) + 4T_2(x) + T_4(x)]$$

$$x^5 = \frac{1}{16}[10T_1(x) + 5T_3(x) + T_5(x)]$$

$$x^6 = \frac{1}{32}[10T_0(x) + 15T_2(x) + 6T_4(x) + T_6(x)]$$

$$x^7 = \frac{1}{64}[35T_1(x) + 21T_3(x) + 7T_5(x) + T_7(x)]$$

and so on. Thus, the powers of x can be expressed in terms of Chebyshev polynomial.

(vii) The polynomials $T_n(x)$ are orthogonal with the weight function $\frac{1}{\sqrt{1-x^2}}$. In fact, putting $x = \cos \theta$ in

$$\int_{-1}^1 \frac{T_m(x)T_n(x)}{\sqrt{1-x^2}} dx,$$

we get

$$\begin{aligned} \int_0^\pi T_m(\cos\theta) T_n(\cos\theta) d\theta &= \int_0^\pi \cos m\theta \cos n\theta d\theta \\ &= \left[\frac{\sin(m+n)\theta}{2(m+n)} + \frac{\sin(m-n)\theta}{2(m-n)} \right]_0^\pi \\ &= \begin{cases} 0, & m \neq n \\ \pi/2, & m = n \neq 0 \\ \pi, & m = n = 0 \end{cases}. \end{aligned}$$

- (viii) For $-1 \leq x \leq 1$, let $p_n(x)$ be all polynomials of degree n and with the coefficient of x^n equal to 1. Putting $\alpha_n = \underset{-1 \leq x \leq 1}{\text{sup}} |p_n(x)|$, we seek the polynomial $P_n(x)$ for which α_n is as small as possible. Since the coefficient of x^n in $T_n(x)$ is 2^{n-1} , we can take

$$p_n(x) = 2^{1-n} T_n(x), n \geq 1. \quad (5.71)$$

We claim that equation (5.71) is the polynomial for which α_n is the least. In this respect, we use the fact that

$$T_n(x) = 0 \text{ when } x = x_k = \frac{\cos(2k+1)\pi}{2n}, \quad k = 0, 1, 2, \dots, n-1,$$

and

$$T_n(x) = (-1)^k \text{ for } x = x_k = \frac{\cos k\pi}{n}, \quad k = 0, 1, 2, \dots, n.$$

Now suppose that $|p_n(x)| < 2^{n-1}$ everywhere in the interval $-1 \leq x \leq 1$. Then we have

$$2^{1-n} T_n(x_0) - p_n(x_0) > 0,$$

$$2^{1-n} T_n(x_1) - p_n(x_1) < 0,$$

.....

.....

that is, the polynomial $p_{n-1}(x) = T_n(x) - p_n(x)$ of degree $(n-1)$ would have an alternating sign in $(n+1)$ points x_0, \dots, x_n . Hence, $p_{n-1}(x)$ has n roots in the interval, which is possible only if $p_{n-1}(x) \equiv 0$ (identically zero) and so

$$p_n(x) = T_n(x),$$

which is called minimum polynomial.

5.17 APPROXIMATION OF A FUNCTION WITH A CHEBYSHEV SERIES

Suppose we want to approximate a function $f(x)$ with a Chebyshev series

$$f(x) = \frac{1}{2} c_0 + c_1 T_1(x) + c_2 T_2(x) + \dots + c_{n-1} T_{n-1}(x) + R_n(x),$$

where

$$R_n(x) = c_n T_n(x) + c_{n+1} T_{n+1}(x) + \dots$$

If the convergence is sufficiently fast, we have

$$R_n(x) \approx c_n T_n(x)$$

and so the error oscillates between $-c_n$ and $+c_n$. The approximation can be performed using the expression for the powers of x in terms of Chebyshev polynomials.

EXAMPLE 5.48

Approximate $x^3 + 2x^2$ using Chebyshev polynomial.

Solution. We know that

$$\begin{aligned} x &= T_1(x), \\ x^3 &= \frac{1}{4}[3T_1(x) + T_3(x)]. \end{aligned}$$

Therefore,

$$\begin{aligned} 4x^3 + 2x^2 &= [3T_1(x) + T_3(x)] + 2x^2 \\ &= 3x + T_3(x) + 2x^2 \\ &= 2x^2 + 3x + T_3(x). \end{aligned}$$

Thus, $2x^2 + 3x$ approximates $x^3 + 2x^2$ and the error oscillates between $\pm c_3 \equiv \pm 1$ since coefficient c_3 of $T_3(x) = 1$.

Chebyshev showed that the minimum value of the error bound in Lagrange interpolation is achieved if the nodes $\{x_k\}$ are Chebyshev abscissas of $T_{n+1}(x)$.

For example, if f is to be approximated by a polynomial of degree at most 3, then Chebyshev nodes are given by

$$x_k = \cos\left(\frac{(2k+1)\pi}{2n}\right), \quad k = 0, 1, 2, \dots, n-1,$$

and so

$$x_3 = \cos\left(\frac{7\pi}{8}\right)$$

$$x_2 = \cos\left(\frac{5\pi}{8}\right)$$

$$x_1 = \cos\left(\frac{3\pi}{8}\right)$$

$$x_0 = \cos\frac{\pi}{8}.$$

EXAMPLE 5.49

Economize the series

$$f(x) = 1 - \frac{1}{2}x - \frac{1}{8}x^2 - \frac{1}{16}x^3$$

in the interval $[-1, 1]$.

Solution. We have

$$\begin{aligned} f(x) &= 1 - \frac{1}{2}x - \frac{1}{8}x^2 - \frac{1}{16}x^3 \\ &= 1 - \frac{1}{2}T_1(x) - \frac{1}{8}\left[\frac{1}{2}\{(T_0(x) + T_2(x))\}\right] - \frac{1}{16}\left[\frac{1}{4}\{(3T_1(x) + T_3(x))\}\right] \\ &= 1 - \frac{1}{2}T_1(x) - \frac{1}{16} - \frac{1}{16}T_2(x) - \frac{3}{64}T_1(x) - \frac{1}{64}T_3(x) \\ &= \frac{15}{16} - \frac{35}{64}T_1(x) - \frac{1}{16}T_2(x) - \frac{1}{64}T_3(x). \end{aligned}$$

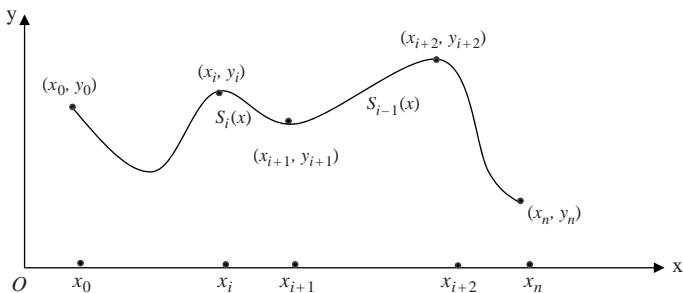
Since $\frac{1}{64} = 0.016$, we have

$$\begin{aligned} f(x) &= \frac{15}{16} - \frac{35}{64}x - \frac{1}{16}(2x^2 - 1) \\ &= \frac{15}{16} - \frac{35}{64}x - \frac{1}{8}x^2 + \frac{1}{16} = 1 - \frac{35}{64}x - \frac{1}{8}x^2. \end{aligned}$$

5.18 INTERPOLATION BY SPLINE FUNCTIONS

In polynomial interpolation, we construct a single polynomial to the given tabulated points. The results obtained are not satisfactory if the tabulated points do not belong to the polynomial. Another method of interpolation is to “piece together” the graphs of lower-degree polynomials $S_i(x)$ and interpolate between the successive nodes (x_i, y_i) and (x_{i+1}, y_{i+1}) .

The two adjacent portions of the curve $y = S_i(x)$ and $y = S_{i+1}(x)$ which lie above $[x_i, x_{i+1}]$ and $[x_{i+1}, x_{i+2}]$, respectively, pass through the common node (x_{i+1}, y_{i+1}) . The two portions of the graph are tied together at the node (x_{i+1}, y_{i+1}) and the set of functions $\{S_i(x)\}$ form a piecewise polynomial curve which is denoted by $S(x)$.



A function $S(x)$ defined on $[x_0, x_n]$ is called a spline of degree p (or of type $p + 1$) if it is

- (i) a polynomial of degree $p \geq 2$ on every segment $[x_i, x_{i+1}], i = 0, 1, \dots, n$
- (ii) is $p - 1$ times continuously differentiable on $[x_0, x_n]$.
- (iii) $S(x_i) = y_i$, that is, the spline passes through each point (x_i, y_i) .

If we use a polynomial of degree 1, then we get a polygonal path consisting of line segments passing through the points (x_i, y_i) . Using Lagrange's polynomial, we have

$$S_i(x) = L_i(x) f(x_i) + L_{i+1}(x) f(x_{i+1}),$$

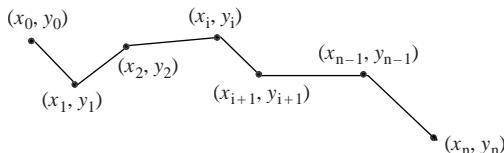
where

$$L_i(x) = \frac{x - x_{i+1}}{x_i - x_{i+1}}, \quad L_{i+1}(x) = \frac{x - x_i}{x_{i+1} - x_i}$$

and so

$$S_i(x) = \frac{x - x_{i+1}}{x_i - x_{i+1}} f(x_i) + \frac{x - x_i}{x_{i+1} - x_i} f(x_{i+1}), \quad x_i \leq x \leq x_{i+1}.$$

The resulting curve looks like a “broken line” as shown below:



This type of interpolation is called a piecewise linear interpolation and $\{S_i(x)\}$ is called a linear spline. Thus, the linear spline function can be written in the form

$$S(x) = \begin{cases} \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1) & \text{for } x \in [x_0, x_1] \\ \frac{x - x_2}{x_1 - x_2} f(x_1) + \frac{x - x_1}{x_2 - x_1} f(x_2) & \text{for } x \in [x_1, x_2] \\ \dots \\ \dots \\ \frac{x - x_{i+1}}{x_i - x_{i+1}} f(x_i) + \frac{x - x_i}{x_{i+1} - x_i} f(x_{i+1}) & \text{for } x \in [x_i, x_{i+1}] \end{cases}$$

Similarly, if we draw a quadratic curve through the points (x_i, y_i) and (x_{i+1}, y_{i+1}) and another quadratic curve through (x_{i+2}, y_{i+2}) and (x_{i+3}, y_{i+3}) such that the slopes of the two quadratic curves match at x_{i+1} , the resulting curve is also not so smooth.

However, it is possible to construct cubic functions $S_i(x)$ on each interval $[x_i, x_{i+1}]$ so that the resulting piecewise curve $S(x) = \{S_i(x)\}$ and its first and second derivatives are all continuous on the interval $[x_0, x_n]$.

The continuity of $S'(x)$ ensures that the graph of $S(x)$ will not have sharp corners, whereas the continuity of $S''(x)$ means that the radius of curvature is defined at each point.

Let $x_0 < x_1 < \dots < x_n$ be the partition of $[x_0, x_n]$ and $[(x_i, x_{i+1})]$, $i = 0, 1, \dots, n$ are $n + 1$ points. Then the function $S(x)$ is called a cubic spline if there exist n cubic polynomials $S_i(x)$ such that

- (i) $S_i(x)$ is a cubic polynomial in each $[x_i, x_{i+1}]$
- (ii) $S'(x)$ and $S''(x)$ are continuous on $[x_0, x_n]$.
- (iii) $S'(x_i) = y_i$, that is, the spline passes through each point (x_i, y_i) .

5.19 EXISTENCE OF CUBIC SPLINE

Let $S(x) = \{S_i(x_i)\}$ be piecewise cubic in each of the subintervals $[x_i, x_{i+1}]$. Thus, $S''(x)$ is piecewise linear on $[x_0, x_n]$. The Lagrange's interpolation formula gives the following representation for $S''(x)$:

$$S''(x) = \frac{x - x_{i+1}}{x_i - x_{i+1}} S''(x_i) + \frac{x - x_i}{x_{i+1} - x_i} S''(x_{i+1}). \quad (5.72)$$

Put

$$m_i = S''(x_i), \quad m_{i+1} = S''(x_{i+1}), \quad h_i = x_{i+1} - x_i$$

to get

$$S''(x) = \frac{m_i}{h_i}(x_{i+1} - x) + \frac{m_{i+1}}{h_i}(x - x_i), \quad x_i \leq x \leq x_{i+1}$$

for $i = 0, 1, \dots, n - 1$. Integrating twice we get

$$S_i(x) = \frac{m_i}{6h_i}(x_{i+1} - x)^3 + \frac{m_{i+1}}{6h_i}(x - x_i)^3 + c_i(x_{i+1} - x) + d_i(x - x_i), \quad (5.73)$$

where c_i and d_i are constants of integration. Substituting x_i and x_{i+1} for x in equation (5.73) and using the values $y_i = S_i(x_i)$ and $y_{i+1} = S_i(x_{i+1})$, we get

$$y_i = \frac{m_i}{6} h_i^2 + c_i h_i \text{ and } y_{i+1} = \frac{m_{i+1}}{6} h_i^2 + c_i h_i. \quad (5.74)$$

Solving these two equations for c_i and d_i and putting these values of constants in equation (5.73), we get

$$\begin{aligned} S_i(x) &= \frac{m_i}{6h_i}(x_{i+1} - x)^3 + \frac{m_{i+1}}{6h_i}(x - x_i)^3 \\ &\quad + \left(\frac{y_i}{h_i} - \frac{m_i h_i}{6} \right)(x_{i+1} - x) + \left(\frac{y_{i+1}}{h_i} - \frac{m_{i+1} h_i}{6} \right)(x - x_i), \end{aligned} \quad (5.75)$$

which contains only two unknown coefficients m_i and m_{i+1} . To find these coefficients, we differentiate (5.75) and get

$$\begin{aligned} S'_i(x) = & -\frac{m_i}{2h_i}(x_{i+1} - x_i)^2 + \frac{m_{i+1}}{2h_i}(x - x_i)^2 \\ & - \left(\frac{y_i}{h_i} - \frac{m_{i-1}}{6} \right) + \left(\frac{y_{i+1}}{h_i} - \frac{m_{i+1}}{6} \right). \end{aligned} \quad (5.76)$$

Replacing i by $i-1$ in equation (5.76), we get

$$\begin{aligned} S'_{i-1}(x) = & -\frac{m_{i-1}}{2h_{i-1}}(x_i - x)^2 + \frac{m_i}{2h_{i-1}}(x - x_{i-1})^2 \\ & - \left(\frac{y_{i-1}}{h_{i-1}} - \frac{m_{i-1}}{6} \right) + \left(\frac{y_i}{h_{i-1}} - \frac{m_{i-1}}{6} \right). \end{aligned} \quad (5.77)$$

From equation (5.76), we have

$$\begin{aligned} S'_i(x_i) = & -\frac{m_i}{2}h_i - \frac{y_i}{h_i} + \frac{m_i}{6}h_i + \frac{y_{i+1}}{h_i} - \frac{m_{i+1}}{6}h_i \\ = & -\frac{m_i}{3}h_i - \frac{m_{i+1}}{6}h_i + p_i, \end{aligned} \quad (5.78)$$

where

$$p_i = \frac{y_{i+1} - y_i}{h_i}.$$

Similarly, from equation (5.77), we have

$$S'_{i-1}(x_i) = \frac{m_i}{3}h_{i-1} + \frac{m_{i-1}}{6}h_{i-1} + r_{i-1},$$

where

$$p_{i-1} = \frac{y_i - y_{i-1}}{h_{i-1}}. \quad (5.79)$$

But continuity of $S'(x)$ implies

$$S'_i(x_i) = S'_{i-1}(x_i).$$

Hence, equations (5.78) and (5.79) yield

$$-\frac{m_i}{3}h_i - \frac{m_{i+1}}{6}h_i + p_i = \frac{m_i}{3}h_{i-1} + \frac{m_{i-1}}{6}h_{i-1} + r_{i-1}$$

or

$$h_{i-1} + 2(h_{i-1} + h_i) + h.m_{i+1} = 6(p_i - p_{i-1}) \text{ for } i = 1, 2, \dots, n-1. \quad (5.80)$$

Thus, we get a relation between m_{i-1} , m_i , and m_{i+1} .

If we consider natural cubic spline in which the graph is linear for $x < x_0$ and $x > x_n$, we have $m_0 = 0$ and $m_n = 0$. Thus, we have $(n+1)$ equations in $(n+1)$ unknowns m_i and so m_i can be found. Substituting the values of m_i in equation (5.75), we get the required cubic spline.

Remark 5.9. For equal intervals, we have $h_i = n_{i-1} = h$ and so formula (5.80) becomes

$$hm_{i-1} + 4m_i h + m_{i+1} = \frac{6}{h}(y_{i+1} - 2y_i + y_{i-1})$$

or

$$m_{i-1} + 4m_i + m_{i+1} = \frac{6}{h^2}(y_{i+1} - 2y_i + y_{i-1}) \text{ for } i = 1, 2, \dots, n-1.$$

EXAMPLE 5.50

Fit a linear spline to the data

x	1	2	3
y	-1	4	21

Solution. We know that the piecewise linear interpolation polynomial is given by

$$S_i(x) = \frac{x - x_i}{x_{i-1} - x_i} f(x_{i-1}) + \frac{x - x_{i-1}}{x_i - x_{i-1}} f(x_i), \quad x_{i-1} \leq x \leq x_i.$$

We have,

$$x_0 = 1, x_1 = 2, x_2 = 3, f(x_0) = -1, f(x_1) = 4, f(x_2) = 21.$$

Thus, for $1 \leq x \leq 2$

$$S_1(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1) = \frac{(x-2)}{1-2}(-1) + \frac{(x-1)(4)}{2-1} = 5x - 6.$$

Similarly, for $2 \leq x \leq 3$, we have

$$S_2(x) = \frac{(x - x_2)}{x_1 - x_2} f(x_1) + \frac{(x - x_1)}{x_2 - x_1} f(x_2) = \frac{(x-3)(4)}{2-3} + \frac{(x-2)(21)}{3-2} = 17x - 30.$$

Hence, the required linear spline is

$$S(x) = \begin{cases} 5x - 6, & 1 \leq x \leq 2. \\ 17x - 30, & 2 \leq x \leq 3. \end{cases}$$

EXAMPLE 5.51

Find the natural cubic splines passing through the points $(1, -6)$, $(2, -1)$, and $(3, 16)$. Hence, evaluate $y(1.5)$ and $y'(2)$.

Solution: The given tabular values are

x	1	2	3
y	-6	-1	16

We note that the arguments are equispaced with $h = 1$ and $n = 2$.

Further,

$$p_i - p_{i-1} = \frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}} = \frac{y_{i+1} - 2y_i + y_{i-1}}{1}.$$

Then, since $h_i = 1$, the expression

$$h_{i-1} \cdots h_{i-1} + 2(h_{i-1} + \cdots + h_i) \cdots h_{i+1} = 6(p_i - p_{i-1})$$

yields

$$m_{i-1} + 4m_i + m_{i+1} = 6(y_{i+1} - 2y_i + y_{i-1})$$

for $i = 1, \dots, n-1$, that is for $i = 1$ in this problem. Thus, we have

$$m_0 + 4m_1 + m_2 = 6(y_2 - 2y_1 + y_0).$$

Since $m_0 = 0$, $m_2 = 0$, $y_2 = 16$, $y_1 = -1$, $y_0 = -6$, we have

$$4m_1 = 6(16 + 2 - 6) = 72$$

and so $m_1 = 18$. Hence,

$$\begin{aligned} S_0(x) &= \frac{18}{6}(x-1)^3 - 6(2-x) + (-1-3)(x-1) \\ &= 3(x^3 - 3x^2 + 3x - 1) - 12 + 6x - 4x + 4 \\ &= 3x^3 - 9x^2 + 11x - 11, \\ S_1(x) &= \frac{18}{6}(3-x)^3 + (1-3)(3-x) + (16)(x-2) \\ &= 3(-x^3 + 9x^2 - 27x + 27) - 12 + x + 16x - 32 \\ &= -3x^3 + 27x^2 - 16x + 37. \end{aligned}$$

Hence, the cubic spline is

$$S(x) = \begin{cases} 3x^3 - 9x^2 + 11x - 11, & 1 \leq x \leq 2 \\ -3x^3 + 27x^2 - 61x + 37, & 2 \leq x \leq 3. \end{cases}$$

Then

$$y(1.5) = 3(3.375) - 9(2.25) + 11(1.5) - 11 = 4.625.$$

Now since 2 lies in both the intervals $[1, 2]$ and $[2, 3]$,

$$y'(2) = \begin{cases} [9x^2 - 18x + 11]_{x=2} = 11 \\ [9x^2 + 54x - 61]_{x=2} = 11 \end{cases}$$

from first and second spline respectively

Thus $S'_0(2) = S'_1(2)$ and so the spline is smooth.

EXAMPLE 5.52

Show that the function

$$f(x) = \begin{cases} 18 - \frac{75x}{2} + 26x^2 - \frac{11x^3}{2}, & 1 \leq x \leq 2 \\ 70 + \frac{189x}{2} - 40x^2 + \frac{11x^3}{2}, & 2 \leq x \leq 3 \end{cases}$$

is a cubic spline.

Solution: We observe that

(i) $f(x)$ is a cubic in both intervals $[1, 2]$ and $[2, 3]$

$$(ii) S_0(x) = 18 - \frac{75x}{2} + 26x^2 - \frac{11x^3}{2}, \quad 1 \leq x \leq 2,$$

$$S_1(x) = 70 + \frac{189x}{2} - 40x^2 + \frac{11x^3}{2}. \quad 2 \leq x \leq 3,$$

and

$$S_0(2) = 18 - \frac{75}{2}(2) + 26(4) - \frac{11(2)^3}{2} = 3$$

$$S_1(2) = 70 + 189 - 160 + 44 = 3$$

$$S'_0(2) = \left[-\frac{75}{2} + 52x - \frac{33x^2}{2} \right]_{x=2} = -\frac{75}{2} + 104 - 66 = 0.50$$

$$S'_1(2) = \left[\frac{189}{2} - 80x + \frac{33x^2}{2} \right]_{x=2} = \frac{189}{2} - 160 + 66 = 0.50$$

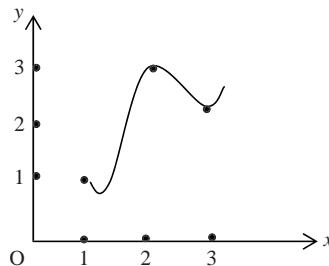
$$S''_0(2) = \left[52 - \frac{66}{2} x \right]_{x=2} = -14,$$

$$S''_1(2) = \left[-80 + \frac{66}{2} x \right]_{x=2} = -14.$$

Therefore, both $S_0(x)$ and $S_1(x)$ and their first two derivatives agree at $x = 2$. Hence, $f(x)$ is a cubic spline. Further, the values at $x = 1, 2, 3$ are

x	1	2	3
y	1	3	2

Therefore, the graph of the cubic spline is



EXAMPLE 5.53

Fit a natural cubic spline to the data

x	1	2	3
y	-3	4	23

and compute $y(1.5)$ and $y'(1)$.

Solution: The arguments are equally spaced with $h = 1$. Further, $n = 2$ and $m_0 = m_2 = 0$. Then m_1 is given by

$$m_{i-1} + 4m_i + m_{i+1} = 6(y_{i+1} - 2y_i + y_{i-1}),$$

by putting $i = 1$. Thus,

$$m_0 + 4m_1 + m_2 = 6(y_2 - 2y_1 + y_0),$$

which yields

$$m_1 = \frac{6}{4}(23 - 8 - 3) = 18.$$

Hence,

$$\begin{aligned} S_0(x) &= \frac{m_1}{6}(x - x_0)^3 + y_0(x_1 - x) + \left(y_1 - \frac{m_1}{6} \right)(x - x_0) \\ &= \frac{18}{6}(x-1)^3 + (-3)(2-x) + \left(4 - \frac{18}{6} \right)(x - x_0) \\ &= 3(x^3 - 3x^2 + 3x - 1) - 6 + 3x + x - 1 \\ &= 3x^3 - 9x^2 + 9x - 3 - 6 + 3x + x - 1 = 3x^3 - 9x^2 + 13x - 10, \\ S_1(x) &= \frac{18}{6}(3-x)^3 + \left(4 - \frac{18}{6} \right)(3-x) + (23)(x-2) \\ &= 3(-x^3 + 9x^2 - 27x + 27) + (3-x) + 23x - 46 \\ &= -3x^3 - 27x^2 - 81x + 81 + 3 - x + 23x - 46 \\ &= 3x^3 - 27x^2 - 59x + 38. \end{aligned}$$

Then

$$\begin{aligned} y(1.5) &= S_0(1.5) = 3(1.5)^3 - 9(1.5)^2 + 13(1.5) - 10 \\ &= 10.125 - 20.250 + 19.5 - 10 = 0.625 \end{aligned}$$

and

$$\begin{aligned} y'(1) &= S'(1) = [9x^2 - 18x + 13]_{x=1} \\ &= 9 - 18 + 13 = 4. \end{aligned}$$

The tabulated function is $f(x) = x^3 - 4$. Therefore, actual value of $f(1.5) = 3.375 - 4 = -0.625$ and actual value of $f'(1) = 3$.

EXAMPLE 5.54:

Develop cubic splines for the data given below and predict $f(1.5)$

x	0	1	2	3
$f(x)$	1	-1	-1	0

Solution: The given tabular values are

x	0	1	2	3
y	1	-1	-1	0

We note that the arguments are equispaced with $h = 1$ and $n = 3$. The splines will be developed by the formula

$$\begin{aligned} s_i(x) &= \frac{m_i}{6h_i}(x_{i+1} - x)^3 + \frac{m_{i+1}}{6h_i}(x - x_i)^3 \\ &+ \left(\frac{y_i}{h_i} - \frac{m_i h_i}{6} \right)(x_{i+1} - x) + \left(\frac{y_{i+1}}{h_i} - \frac{m_{i+1} h_i}{6} \right)(x - x_i). \end{aligned} \quad (5.81)$$

For equal spacing $h_i = 1$, we have

$$m_{i-1} + 4m_i + m_{i+1} = 6(y_{i+1} - 2y_i + y_{i-1}).$$

Putting $i = 1$ and 2 , we get

$$m_0 + 4m_1 + m_2 = 6(y_2 - 2y_1 + y_0) = 6(0 - 2(-1) + 1) = 12$$

and

$$m_1 + 4m_2 + m_3 = 6(y_3 - 2y_2 + y_1) = 6(0 - 2(-1) - 1) = 6.$$

But for $n = 3$, the natural cubic spline requires $m_0 = m_3 = 0$. Therefore,

$$4m_1 + m_2 = 12 \text{ and } m_1 + 4m_2 = 6.$$

Solving these equations, we get $m_1 = \frac{14}{5}$ and $m_2 = \frac{4}{5}$. Hence, (5.81) yields the cubic splines as

$$\begin{aligned} s_0(x) &= \frac{14}{6(5)}(x-0)^3 + \left(\frac{1}{1} - 0 \right)(1-x) + \left(\frac{-1}{1} - \frac{14}{5(6)} \right)(x-0) \\ &= \frac{14}{30}x^3 + (1-x)^3 - \frac{44}{30}x = \frac{1}{30}[14x^3 - 74x + 30], \\ s_1(x) &= \frac{14}{6(5)}(2-x)^3 + \frac{4}{6(5)}(x-1)^3 + \left(\frac{-1}{1} - \frac{14}{6(5)} \right)(2-x) + \left(\frac{-1}{1} - \frac{4}{5(6)} \right)(x-1) \\ &= \frac{14}{30}(2-x)^3 + \frac{4}{30}(x-1)^3 - \frac{44}{30}(2-x) - \frac{34}{30}(x-1), \\ s_2(x) &= \frac{14}{6(5)}(3-x)^3 + \left(\frac{-1}{1} - \frac{14}{6(5)} \right)(3-x) + (0-0)(x-x_2) \\ &= \frac{14}{30}(3-x)^3 - \frac{44}{30}(3-x) = \frac{1}{30}[14(3-x)^3 - 44(3-x)]. \end{aligned}$$

EXERCISES

1. Evaluate

(i) $\Delta^2 \cos 2x$ (ii) $\Delta^n \left(\frac{1}{x} \right)$.

Ans. (i) $-4 \sin^2 h \cos (2x + 2h)$

(ii)
$$\frac{(-1)^n n!}{x(x+1)(x+2)\dots(x+n)}$$

2. Show that $\Delta + \nabla = \frac{\Delta}{\nabla} - \frac{\nabla}{\Delta}$.

3. Show that $\Delta^3 y_i = y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i$.

4. Find the function whose first difference is $9x^2 + 11x + 5$.

Ans. $3x^3 + x^2 + x + k$

5. Find the missing values in the following data:

x	45	50	55	60	65
y	3.0	—	2.0	—	-2.4

Ans. $f(50) = 2.925, f(60) = 0.225$

6. Express $3x^4 - 4x^3 + 6x^2 + 2x + 1$ as a factorial polynomial and find fourth order difference

Ans. $3[x]^4 + 14[x]^3 + 15[x]^2 + 7[x] + 1, \Delta^4 y = 72$

7. Form a difference table to fourth differences

x	1	2	3	4	5	6	7	8
f_x	7.93	10.05	12.66	15.79	19.47	23.73	28.60	34.11

Repeat the procedure for the same table when $f_5 = 19.47 + \epsilon$, where ϵ represents an error. How many $\Delta^n f_x$ are affected?8. If $f(x)$ is a cubic polynomial, use the difference table to locate and correct the error in the data:

x	0	1	2	3	4	5	6	7
$f(x)$	25	21	18	18	27	45	76	123

Ans. $f(3)$ is in error, true value is 199. If $f(x)$ is a polynomial of degree 4, locate and correct the error in the table

x	1	2	3	4	5	6	7	8
y	3010	3424	3802	4105	4472	4771	5051	5315

10. The function y is given in the table below:

x	20	24	28	32
y	2854	3162	3544	3992

Find y for $x = 25$ using Bessel's interpolation formula.

Ans. 3250.875 approx.

11. Evaluate $f(3.75)$ from the table

x	2.5	3.0	3.5	4.0	4.5	5.0
y	24.145	22.043	20.225	18.644	17.262	16.047

(Hint: Use Gauss's forward formula).

Ans. 19.40746093

12. Use Stirling's interpolation formula to find $f(35)$ from the table

x	20	30	40	50
y	512	439	346	243

Ans. 395

13. Using Newton's divided difference formula find $f(x)$ as a polynomial in x for the table:

x	0	1	2	4	5	6
y	1	14	15	5	6	19

Ans. $x^3 - 9x^2 + 21x + 1$

14. Let $f(x) = x^3 - 4x$. Construct the divided difference table based on the nodes $x_0 = 1, x_1 = 2, \dots, x_5 = 6$ and find the Newton's polynomial $P_3(x)$ based on x_0, x_1, x_2, x_3 .

Ans. $P_3(x) = 3 + 3(x-1) + 6(x-1)(x-2) + (x-1)(x-2)(x-3)$

15. Using Lagrange's interpolation formula, find the value of t for $A = 85$ using the table

t	2	5	8	14
A	94.8	87.9	81.3	68.7

Ans. 6.5928

16. Use Lagrange's interpolation formula to find the value of y for $x = 10$ using the table given below:

x	5	6	9	11
y	12	13	14	16

Ans. 14.3

17. Find the Lagrange's interpolating polynomial for $(1, -3), (3, 9), (4, 30)$, and $(6, 132)$.

Ans. $x^3 - 3x^2 + 5x - 6$

18. Approximate $2x^3 + 3x^2$ using Chebyshev polynomials.

Ans. $3x^2 + \frac{3}{2}$, error oscillates between $\pm \frac{1}{2}$

19. Economize the power series

$$x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{5040}$$

on the interval $[-1, 1]$ allowing for a tolerance of 0.0005.

Ans. $\frac{383}{384}x - \frac{5}{32}x^3$

20. Find the cubic polynomial which takes the values given in the table:

x	0	1	2	3
$f(x)$	1	2	1	10

Also evaluate $f(4)$ from the difference table and compare it with exact value.

Ans. $f(x) = 2x^3 - 7x^2 + 6x + 1$
 $f(4)$ (by Newton's backward formula) = 41 = exact value 41.

21. Use Everett's formula to obtain $f(1.15)$ from the data:

x	1	1.1	1.2	1.3
$f(x)$	1.000	1.049	1.096	1.140

Ans. 1.073

22. Find $f(x)$ from the table:

x	0	1	3	5	6	9
$f(x)$	-18	0	0	-248	0	13104

Ans. $x^5 - 9x^4 + 18x^3 - x^2 + 9x - 18$

23. Determine the Hermite interpolating polynomial which fits the following data and hence evaluate approximately $y(2.7)$:

x	2.0	2.5	3.0
y	0.69315	0.91629	1.09861
y'	0.50000	0.40000	0.33333

Ans. $y(2.7) = 0.993252$

24. Obtain piecewise cubic Hermite polynomial for the following data:

x	0	1	2
y	1	3	35
y'	1	6	81

Ans. $p_3(x) = \begin{cases} 3x^3 - 2x^2 + x + 1, & 0 \leq x \leq 1 \\ 23x^3 - 66x^2 + 69x - 23, & 1 \leq x \leq 2 \end{cases}$

25. Using Lagrange's interpolation formula, express the function

$$\frac{3x^2 + x + 1}{(x-1)(x-2)(x-3)}$$

as a sum of partial function.

Ans. $\frac{5}{2(x-1)} - \frac{15}{x-2} + \frac{31}{2(x-3)}$

26. Fit a linear spline to the data

x	0	1	2
y	1	3	35

$$\text{Ans. } S(x) = \begin{cases} 2x + 1, & 0 \leq x \leq 1 \\ 32x - 29, & 1 \leq x \leq 2 \end{cases}$$

27. Fit a natural cubic spline to the data

x	1	2	3	4
y	1	2	5	11

and compute $y(1.5)$ and $y'(3)$. Is this spline smooth?

$$\text{Ans. } S(x) = \begin{cases} \frac{1}{3}(x^3 - 3x^2 + 5x), & 1 \leq x \leq 2 \\ \frac{1}{3}(x^3 - 3x^2 + 5x), & 2 \leq x \leq 3 \\ \frac{1}{3}(-2x^3 - 24x^2 + 76x + 81), & 3 \leq x \leq 4 \end{cases}$$

$$y(1.5) = S_0(1.5) = \frac{11}{8}$$

$$y'(3) = S'_1(3) = S'_2(3) = \frac{14}{3} \text{ and so it is a smooth spline.}$$

28. Show that the function

$$f(x) = \begin{cases} 13 - 31x + 23x^2 - 5x^3, & 1 \leq x \leq 2 \\ -35 + 51x - 22x^2 + 3x^3, & 2 \leq x \leq 3 \end{cases}$$

is not a cubic spline.

6 Curve Fitting

So far we have considered the construction of a polynomial, which approximates a given function and takes the same values as the function at certain given points. This is called the method of collocation and the conditions are satisfied by the approximate Lagrange's interpolation polynomial. When the given points are equally spaced, we can form a difference table and find the polynomial using Newton's forward difference formula.

For example, the polynomial $4x - 4x^2$ agrees with the function $\sin \pi x$ for $x = 0, \frac{1}{2}, 1$, but this approximation is not very satisfactory because the polynomial $4x - 4x^2$ is larger than $\sin \pi x$ in the range $(0,1)$ except at point $x = \frac{1}{2}$. Similarly, the Lagrangian interpolation polynomials constructed for the function $\frac{1}{x^2 + 1}$ in the interval $[-5,5]$ with uniformly distributed nodes give rise to arbitrary large deviations for increasing degree n .

If the functional values at the given points (nodes) are the result of experiments or if they are rounded values or if the nodes are subject to error, then the advantages of the method of collocation are to some extent lost. In such a case, Weierstrass's Approximation Theorem is of remarkable utility:

Theorem 6.1. (Weierstrass's Approximation Theorem). If f is a continuous function in the interval $[a,b]$, then to each $\varepsilon > 0$ there exists a polynomial $p(x)$ such that

$$|f(x) - p(x)| < \varepsilon \text{ for all } x \in [a, b].$$

Weierstrass's Theorem allows us to consider other methods of approximations. We discuss these methods one by one.

6.1 LEAST SQUARE LINE APPROXIMATION

Suppose that we have an empirical data in the form of n pairs of values $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, where the experimental errors are associated with the functional values y_1, y_2, \dots, y_n only. Then we seek a linear function

$$y = f(x) = a + bx \quad (6.1)$$

fitting the given points as well as possible. Equation (6.1) will not in general be satisfied by any of the n pairs. Substituting in equation (6.1) each of the n pairs of values in turn, we get

$$\left. \begin{aligned} e_1 &= y_1 - a - bx_1 \\ e_2 &= y_2 - a - bx_2 \\ \cdots &\quad \cdots \quad \cdots \\ e_n &= y_n - a - bx_n \end{aligned} \right\}, \quad (6.2)$$

where e_k , $k = 1, \dots, n$ are measurement errors, called residuals or deviations. To know how far the curve $y = f(x)$ lies from the given data, the following errors are considered:

(i) Maximum error

$$e_{\infty}(f) = \max_{1 \leq k \leq n} \{ |y_k - a - bx_k| \}$$

(ii) Average error

$$e_A(f) = \frac{1}{n} \sum_{k=1}^n |y_k - a - bx_k|$$

(iii) Root mean square (RMS) error

$$e_{\text{rms}}(f) = \left[\frac{e_1^2 + \dots + e_n^2}{n} \right]^{\frac{1}{2}}.$$

The least square line $y = f(x) = a + bx$ is the line that minimizes the root mean square error $e_{\text{rms}}(f)$. But the quantity $e_{\text{rms}}(f)$ is minimum if and only if $\sum_{k=1}^n (y_k - a - bx_k)^2 = \sum_{k=1}^n e_k^2$ is minimum. Thus, in case of least square line we are looking for a linear function $a + bx$ as an approximation to a function $y = f(x)$ when we are given the values of y at the points x_1, \dots, x_n . We aim at minimizing the sum of the squared errors

$$e(a, b) = \sum_{i=1}^n (y_i - a - bx_i)^2. \quad (6.3)$$

Geometrically, if d_i is the vertical distance from the data point (x_i, y_i) to the point $(x_i, a + bx_i)$ on the line, then $d_i = y_i - a - bx_i$ (see Figure 6.1). We must minimize the sum of the squares of the vertical distances d_i ,

that is, the sum $\sum_{i=1}^n d_i^2$.

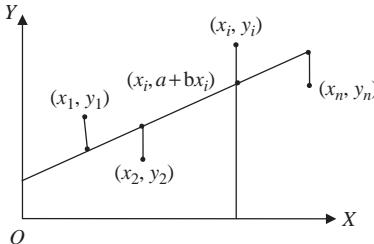


Figure 6.1

To minimize $e(a, b)$, we equate to zero the partial derivative of equation (6.3) with respect to a and with respect to b . Thus,

$$\frac{\partial e(a, b)}{\partial a} = \sum_{i=1}^n 2(y_i - a - bx_i) = 0$$

and

$$\frac{\partial e(a, b)}{\partial b} = \sum_{i=1}^n 2x_i(y_i - a - bx_i) = 0,$$

which are known as normal equations. We write these equations in the form

$$na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad (6.4)$$

and

$$a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i. \quad (6.5)$$

The normal equations (6.4) and (6.5) can be solved for a and b using Cramer's rule or by some other method.

EXAMPLE 6.1

Show that, according to the principle of least squares, the best fitting linear function for the points (x_i, y_i) , $i = 1, 2, \dots, n$ may be expressed in the form

$$\begin{vmatrix} x & y & 1 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n y_i & n \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n y_i^2 & \sum_{i=1}^n x_i \end{vmatrix} = 0.$$

Solution. Eliminating a and b from equations (6.4), (6.5), and $y = a + bx$, we get the required result.

We have supposed in the above derivation that the errors in x values can be neglected compared with the errors in the y values. Now we suppose that the x values as well as the y values are subject to errors of about the same order of magnitude. Now we minimize the sum of the squares of the perpendicular distances to the line. Thus, if $y = a + bx$ is the equation of the line, then

$$e(a, b) = \frac{1}{1+b^2} \sum_{i=1}^n (y_i - a - bx_i)^2.$$

For minimum, partial derivatives with respect to a and b should vanish. Thus,

$$\frac{\partial e(a, b)}{\partial a} = \frac{2}{1+b^2} \sum_{i=1}^n (y_i - a - bx_i) = 0$$

and

$$\begin{aligned} \frac{\partial e(a, b)}{\partial b} &= -2(1+b^2) \sum_{i=1}^n (y_i - a - bx_i)x_i - 2b \sum_{i=1}^n (y_i - a - bx_i)^2 = 0, \text{ that is,} \\ &\sum_{i=1}^n (y_i - a - bx_i) = 0 \end{aligned} \quad (6.6)$$

and

$$(1+b^2) \sum_{i=1}^n (y_i - a - bx_i)x_i = b \sum_{i=1}^n (y_i - a - bx_i)^2. \quad (6.7)$$

From equation (6.6), we get

$$a = y_0 - bx_0, \quad (6.8)$$

where

$$x_0 = \frac{1}{n} \sum_{i=1}^n x_i \text{ and } y_0 = \frac{1}{n} \sum_{i=1}^n y_i.$$

After simplification, equation (6.7) yields

$$b^2 + \frac{A - C}{B} b - 1 = 0, \quad (6.9)$$

where

$$A = \sum_{i=1}^n x_i^2 - nx_0^2,$$

$$B = \sum_{i=1}^n x_i y_i - nx_0 y_0,$$

$$C = \sum_{i=1}^n y_i^2 - ny_0^2.$$

Finding the value of b from equation (6.9), we obtain the corresponding value of a from equation (6.8).

EXAMPLE 6.2

The points (2,2), (5,4), (6,6), (9,9), and (11,10) should be approximated by a straight line. Perform this assuming

- (a) the error in x values can be neglected
- (b) that the errors in x and y values are of the same order of magnitude.

Solution. (a) The sum table for the given problem is

n	x	x^2	y	xy	y^2
1	2	4	2	4	4
1	5	25	4	20	16
1	6	36	6	36	36
1	9	81	9	81	81
1	11	121	10	110	100
5	33	267	31	251	237

Let the least square line be $y = a + bx$. Therefore, the normal equations are

$$5a + 33b = 31 \quad (6.10)$$

$$33a + 267b = 251. \quad (6.11)$$

Multiplying equation (6.10) by 33 and equation (6.11) by 5, we obtain

$$165a + 1089b = 1023$$

$$165a + 1335b = 1255.$$

Subtracting, we get

$$246b = 232 \text{ and so } b = \frac{116}{123} = 0.9431.$$

Then equation (6.10) yields

$$a = \frac{31 - 33(0.9431)}{5} = -0.0244.$$

Hence, the least square line is

$$y = 0.9431x - 0.0244.$$

(b) We have

$$x_0 = \frac{1}{n} \sum_{i=1}^n x_i = \frac{33}{5}$$

$$y_0 = \frac{1}{n} \sum_{i=1}^n y_i = \frac{31}{5}$$

$$A = \sum_{i=1}^n x_i^2 - nx_0^2 = 267 - 5\left(\frac{33}{5}\right)^2 = \frac{246}{5} = 49.2$$

$$B = \sum_{i=1}^n x_i y_i - nx_0 y_0 = 251 - \frac{5(33)(31)}{25} = 46.4$$

$$C = \sum_{i=1}^n y_i^2 - ny_0^2 = 237 - 5\left(\frac{31}{5}\right)^2 = 44.8.$$

Therefore, equation in b

$$b^2 + \frac{A-C}{B} b - 1 = 0$$

becomes

$$b^2 + \frac{4.4}{46.4} b - 1 = 0$$

or

$$b^2 + 0.0948b - 1 = 0.$$

Hence,

$$b = \frac{-0.948 \pm \sqrt{4.0089}}{2} = 0.9537 \text{ (+ve).}$$

Then $a = y_0 - bx_0$ yields

$$a = -0.0944.$$

Hence,

$$y = 0.9537x - 0.0944$$

is the required least square line.

EXAMPLE 6.3

In the following data, x and y are subject to error of the same order of magnitude:

$x:$	1	2	3	4	5	6	7	8
$y:$	3	3	4	5	5	6	6	7

Find a straight line approximation using the least square method.

Solution. The sum table for the given problem is

n	x	x^2	y	xy	y^2
1	1	1	3	3	9
1	2	4	3	6	9
1	3	9	4	12	16
1	4	16	5	20	25
1	5	25	5	25	25
1	6	36	6	36	36
1	7	49	6	42	36
1	8	64	7	56	49
8	36	204	39	200	205

Let the equation be $y = a + bx$. Then

$$a = y_0 - bx_0, \quad (6.12)$$

where

$$x_0 = \frac{1}{n} \sum_{i=1}^n x_i = \frac{36}{8},$$

$$y_0 = \frac{1}{n} \sum_{i=1}^n y_i = \frac{39}{8}.$$

Further,

$$A = \sum_{i=1}^n x_i^2 - nx_0^2 = 204 - 8 \left(\frac{36}{8} \right)^2 = \frac{408 - 324}{2} = 42,$$

$$B = \sum_{i=1}^n x_i y_i - nx_0 y_0 = 200 - 8 \frac{(36)(39)}{8^2} = \frac{400 - 351}{2} = \frac{49}{2} = 24.5,$$

$$C = \sum_{i=1}^n y_i^2 - ny_0^2 = 205 - 8 \left(\frac{39}{8} \right)^2 = \frac{1640 - 1521}{8} = \frac{119}{8} = 14.87.$$

Then the value of b is given by

$$b^2 + \frac{A - C}{B} b - 1 = 0$$

or

$$b^2 + \frac{42.0 - 14.87}{24.5} b - 1 = 0$$

or

$$b^2 + 1.107b - 1 = 0,$$

which yields

$$b = \frac{-1.107 \pm \sqrt{5.225}}{2} = 0.5895 \text{ (+ve).}$$

Then equation (6.12) gives $a = 2.225$. Hence, the least square line is

$$y = 0.59x + 2.22.$$

6.2 THE POWER FIT $y = a x^m$

Suppose we require ax^m as an approximation to a function y , where m is a known constant. We must find the value of a such that the equation

$$y = ax^m \quad (6.13)$$

is satisfied as nearly as possible by each of the n pairs of observed values $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Using least square technique, we should minimize the error function

$$e(a) = \sum_{i=1}^n (ax_i^m - y_i)^2. \quad (6.14)$$

For this purpose, partial derivative of equation (6.14) with respect to a must vanish. So, we have

$$0 = 2 \sum_{i=1}^n (ax_i^m - y_i)(x_i^m)$$

and so

$$0 = a \sum_{i=1}^n x_i^{2m} - \sum_{i=1}^n x_i^m y_i,$$

which yields

$$a = \frac{\sum_{i=1}^n x_i^m y_i}{\sum_{i=1}^n x_i^{2m}}.$$

Putting the value of a in equation (6.13), we get the required equation.

Alternative method: Taking logarithms of both sides of equation (6.13) yields

$$\log y = \log a + m \log x,$$

which is of the form $Y = A + BX$, where $Y = \log y$, $A = \log a$, $B = m$, and $X = \log x$. Now the least square line can be found. Then a and m are found.

EXAMPLE 6.4

Find the gravitational constant g using the data below and the relation $h = \frac{1}{2} g t^2$, where h is distance in meters and t is the time in seconds.

t	0.200	0.400	0.600	0.800	1.000
h	0.1960	0.7850	1.7665	3.1405	4.9075

Solution. The sum table for the given problem is

t	h	$t^{2m}(m = 2)$	ht^2
0.200	0.1960	0.0016	0.00784
0.400	0.7850	0.0256	0.12560
0.600	1.7665	0.1296	0.63594
0.800	3.1405	0.4096	2.00992
1.000	4.9075	1.0000	4.90750
		1.5664	7.68680

Then using the formula $y = ax^m$ for power fit, we have

$$\frac{1}{2}g = \frac{\sum_{k=1}^n h_k t_k^m}{\sum_{k=1}^n t_k^{2m}} = \frac{7.68680}{1.5664} = 4.9073$$

and so the gravitational constant $g = 9.8146 \text{ m/sec}^2$.

EXAMPLE 6.5

Find the power fits $y = ax^2$ and $y = bx^3$ for the data given below and determine which curve fits best:

x	2.0	2.3	2.6	2.9	3.2
y	5.1	7.5	10.6	14.4	19.0

Solution. The sum table for the given problem is

x	x^2	x^3	x^4	x^6	y	yx^2	yx^3
2	4	8	16	64	5.1	20.4	40.8
2.3	5.29	12.167	27.984	148.035	7.5	39.675	91.252
2.6	6.76	17.576	45.698	308.918	10.6	71.656	186.306
2.9	8.41	24.389	70.729	594.831	14.4	121.104	351.202
3.2	10.24	32.768	104.858	1073.746	19.0	194.560	622.592
			265.269	2189.530		447.395	1292.152

Then for $y = ax^2$, we have

$$a = \frac{\sum y_i x_i^2}{\sum x_i^4} = \frac{447.395}{265.269} = 1.6866.$$

Hence, the power fit is

$$y = 1.6866x^2.$$

On the other hand, for $y = bx^3$, we have

$$b = \frac{\sum y_i x_i^3}{\sum x_i^6} = \frac{1292.152}{2189.530} = 0.5902.$$

Hence, the power fit is

$$y = 0.5902x^3.$$

To know which of these is best fit, we calculate the corresponding errors. For the first power fit, we have

$$\begin{aligned} e_{\text{rms}} &= \left[\frac{1}{5} \left\{ (ax_1^2 - y_1)^2 + (ax_2^2 - y_2)^2 + (ax_3^2 - y_3)^2 + (ax_4^2 - y_4)^2 + (ax_5^2 - y_5)^2 \right\} \right]^{\frac{1}{2}} \\ &= \left[\frac{1}{5} \left\{ (1.646)^2 + (1.4330)^2 + (0.8014)^2 + (0.2157)^2 + (-1.7292)^2 \right\} \right]^{\frac{1}{2}} \\ &= \left[\frac{1}{5} (2.704 + 2.053 + 0.642 + 0.046 + 2.990) \right]^{\frac{1}{2}} \approx 1.3. \end{aligned}$$

Similarly for the second curve, we have

$$e_{\text{rms}} \approx 0.29.$$

Hence, the power fit curve $y = 0.5902x^3$ is the best.

EXAMPLE 6.6

By using the methods of least squares, find a relation of the form $y = ax^b$ that fits the data:

x	2	3	4	5
y	27.8	62.1	110	161

Solution. The sum table for the given problem is

x	x^2	x^4	y	yx^2
2	4	16	27.8	111.2
3	9	81	62.1	558.9
4	16	256	110	1760
5	25	625	161	4025
		978		6455.1

Then for $y = ax^2$, we have

$$a = \frac{\sum y_i x_i^2}{\sum x_i^4} = \frac{6455.1}{978} = 6.60.$$

Hence, the power fit is

$$y = 6.6x^2$$

6.3 LEAST SQUARE PARABOLA (PARABOLA OF BEST FIT)

Suppose that we want to approximate a given function $y = f(x)$ by a quadratic $a + bx + cx^2$. We must find the values of a , b , and c such that the equation

$$y = a + bx + cx^2 \quad (6.15)$$

is satisfied as nearly as possible by each of the n pairs of observed values $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. The equation will not in general be satisfied exactly by any of the n pairs. Substituting in equation (6.15) each of the n pairs of values in turn, we get the following residual equations:

$$\begin{aligned} e_1 &= a + bx_1 + cx_1^2 - y_1 \\ e_2 &= a + bx_2 + cx_2^2 - y_2 \\ &\dots \quad \dots \quad \dots \quad \dots \\ &\dots \quad \dots \quad \dots \quad \dots \\ e_n &= a + bx_n + cx_n^2 - y_n \end{aligned}$$

The principle of least square says that the best values of the unknown constants a , b , and c are those which make the sum of the squares of the residuals a minimum, that is,

$$\sum_{i=1}^n e_i^2 = e_1^2 + e_2^2 + \dots + e_n^2$$

must be minimum. Thus,

$$e(a, b, c) = \sum_{i=1}^n (a + bx_i + cx_i^2 - y_i)^2$$

should be minimum. For this, the partial derivatives of $e(a, b, c)$ with respect to a , b , and c should be zero. We therefore have

$$\begin{aligned}\frac{\partial e(a, b, c)}{\partial a} &= 2(a + b x_1 + c x_1^2 - y_1) + 2(a + b x_2 + c x_2^2 - y_2) + \dots + 2(a + b x_n + c x_n^2 - y_n) = 0 \\ \frac{\partial e(a, b, c)}{\partial b} &= 2(a + b x_1 + c x_1^2 - y_1)x_1 + 2(a + b x_2 + c x_2^2 - y_2)x_2 + \dots + 2(a + b x_n + c x_n^2 - y_n)x_n = 0 \\ \frac{\partial e(a, b, c)}{\partial c} &= 2(a + b x_1 + c x_1^2 - y_1)x_1^2 + 2(a + b x_2 + c x_2^2 - y_2)x_2^2 + \dots + 2(a + b x_n + c x_n^2 - y_n)x_n^2 = 0.\end{aligned}$$

Hence, the normal equations are

$$\begin{aligned}(a + b x_1 + c x_1^2 - y_1) + (a + b x_2 + c x_2^2 - y_2) + \dots + (a + b x_n + c x_n^2 - y_n) &= 0 \\ (a + b x_1 + c x_1^2 - y_1)x_1 + (a + b x_2 + c x_2^2 - y_2)x_2 + \dots + (a + b x_n + c x_n^2 - y_n)x_n &= 0 \\ (a + b x_1 + c x_1^2 - y_1)x_1^2 + (a + b x_2 + c x_2^2 - y_2)x_2^2 + \dots + (a + b x_n + c x_n^2 - y_n)x_n^2 &= 0.\end{aligned}$$

These equations can further be written as

$$\begin{aligned}na + b(x_1 + x_2 + \dots + x_n) + c(x_1^2 + x_2^2 + \dots + x_n^2) &= y_1 + y_2 + \dots + y_n, \\ a(x_1 + x_2 + \dots + x_n) + b(x_1^2 + x_2^2 + \dots + x_n^2) + c(x_1^3 + x_2^3 + \dots + x_n^3) &= x_1 y_1 + x_2 y_2 + x_3 y_3, \\ a(x_1^2 + x_2^2 + \dots + x_n^2) + b(x_1^3 + x_2^3 + \dots + x_n^3) + c(x_1^4 + x_2^4 + \dots + x_n^4) &= x_1^2 y_1 + x_2^2 y_2 + \dots + x_n^2 y_n.\end{aligned}$$

The above normal equations are solved by ordinary methods of algebra for solving simultaneous equations of first degree in two or more unknowns.

Remark 6.1. The number of normal equations is always the same as the number of unknown constants, whereas the number of residual equations is equal to the number of observations. The number of observations must always be greater than the number of undetermined constants if the method of least square is to be used in the solution.

EXAMPLE 6.7

Find the parabola of best fit (with equation of the form $a + bx + cx^2$) for the data in the following table:

x	0	1	2	3	4
y	-2.1	-0.4	2.1	3.6	9.9

Solution. We establish the following sum table:

n	x	x^2	x^3	x^4	y	xy	x^2y
1	0	0	0	0	-2.1	0	0
1	1	1	1	1	-0.4	-0.4	-0.4
1	2	4	8	16	2.1	4.2	8.4
1	3	9	27	81	3.6	10.8	32.4
1	4	16	64	256	9.9	39.6	158.4
5	10	30	100	354	13.1	54.2	198.8

The normal equations are

$$5a + 10b + 30c = 13.1 \quad (6.16)$$

$$10a + 30b + 100c = 54.2 \quad (6.17)$$

$$30a + 100b + 354c = 198.8. \quad (6.18)$$

Multiplying equation (6.16) by 2 and then subtracting from equation (6.17), we get

$$10b + 40c = 28. \quad (6.19)$$

Multiplying equation (6.17) by 3 and then subtracting from equation (6.18), we get

$$10b + 54c = 36.2. \quad (6.20)$$

Subtracting equation (6.19) from equation (6.20), we get

$$c = 0.58571.$$

Then equation (6.19) yields

$$b = 0.45716$$

and then equation (6.16) yields

$$a = -1.80858.$$

Hence, the parabola of best fit is

$$y = -1.80858 + 0.45716x + 0.58571x^2.$$

EXAMPLE 6.8

Find the least square polynomial of degree two for the following data:

x	0.78	1.56	2.34	3.12	3.81
y	2.50	1.20	1.12	2.25	4.28

Solution. Let the required polynomial be $a + bx + cx^2$. To make the calculations simple, we use the substitution

$$X = \frac{x - 2.34}{0.78}$$

making use of the equal spacing of the arguments. The sum table then becomes

n	X	X^2	X^3	X^4	y	Xy	X^2y
1	-2	4	-8	16	2.50	-5.00	10.00
1	-1	1	-1	1	1.20	-1.20	1.20
1	0	0	0	0	1.12	0	0
1	1	1	1	1	2.25	2.25	2.25
1	1.88	3.53	6.64	12.49	4.28	8.05	15.13
5	-0.12	9.53	-1.36	30.49	11.35	4.10	28.58

The normal equations are

$$5a - 0.12b + 9.53c = 11.35$$

$$-0.12a + 9.53b - 1.36c = 4.10$$

$$9.53a - 1.36b + 30.49c = 2858.$$

Solving these equations by Cramer's rule, we get

$$\begin{aligned}a &= 1.1155021, \\b &= 0.5316061, \\c &= 0.612401.\end{aligned}$$

Hence, the parabola of best fit is

$$y = 1.1155 + 0.5316X + 0.6124X^2,$$

$$\text{where } X = \frac{x - 2.34}{0.78}.$$

EXAMPLE 6.9

Find the least square fit $y = a + bx + cx^2$ for the data

x	-3	-1	1	3
y	15	5	1	5

Solution. The sum table for the given problem is

n	x	x^2	x^3	x^4	y	xy	x^2y
1	-3	9	-27	81	15	-45	135
1	-1	1	-1	1	5	-5	5
1	1	1	1	1	1	1	1
1	3	9	27	81	5	15	45
4	0	20	0	164	26	-34	186

The normal equations are

$$4a + 20c = 26$$

$$20b = -34$$

$$20a + 164c = 186.$$

Solving these equations, we have

$$b = -\frac{34}{20} = -1.70, c = 0.875, a = 2.125.$$

Hence, the least square parabola is

$$y = 2.125 - 1.700x + 0.875x^2.$$

EXAMPLE 6.10

Fit a parabola to the following data

x	1	2	3	4
y	0.30	0.64	1.32	5.40

Solution. The sum table for the given problem is

n	x	x^2	x^3	x^4	y	xy	x^2y
1	1	1	1	1	0.30	0.30	0.30
1	2	4	8	16	0.64	1.28	2.56
1	3	9	27	81	1.32	3.96	11.88
1	4	16	64	256	5.40	21.60	86.40
4	10	30	100	354	7.66	27.14	101.14

The normal equations are

$$4a + 10b + 30c = 7.66,$$

$$10a + 30b + 100c = 27.14,$$

$$30a + 100b + 354c = 101.14,$$

Solving these equations by Gauss elimination method or Cramer's rule, we get

$$a = -1.09, \quad b = 0.458, \quad c = 0.248.$$

Hence, the parabola of fit is

$$y = -1.09 + 0.458x + 0.248x^2.$$

EXAMPLE 6.11

Fit a parabola $y = a + bx + x^2$ to the following data:

x	2	4	6	8	10	0
y	3.07	12.85	31.47	57.38	91.29	

Solution. The sum table for the given problem is

n	x	x^2	x^3	x^4	y	xy	x^2y
1	2	4	8	16	3.07	6.14	12.28
1	4	16	64	256	12.85	51.4	205.6
1	6	36	216	1,296	31.47	188.82	1,132.92
1	8	64	512	4,096	57.38	459.04	3,672.32
1	10	100	1,000	1,000	91.29	912.9	9,129.00
5	30	220	1,800	15,664	196.06	1,618.3	14,152.12

The normal equations are

$$5a + 30b + 220c = 196.06$$

$$30a + 220b + 1800c = 1618.30$$

$$220a + 1800b + 15644c = 14152.12.$$

These equations yield

$$40b + 480c = 44.94$$

and

$$480b + 5984c = 5525.48.$$

This last pair of equations give $b = -0.859$ and $c = 0.992$. Putting these values in the first normal equation, we get $a = 0.720$.

Hence, the least square parabola is

$$y = 0.72 - 0.859x + 0.992x^2$$

EXAMPLE 6.12

If x (km/hr) and y (kg/tonne) are related by a relation of the type $y = a + bx^2$, find by the method of least squares a and b with the help of the following table:

x	10	20	30	40	50
y	8	10	15	21	30

Solution. The normal equations for the curve fitting of the type $y = a + bx^2$ are

$$na + b(x_1^2 + x_2^2 + \dots + x_n^2) = y_1 + y_2 + \dots + y_n$$

$$a(x_1^2 + x_2^2 + \dots + x_n^2) + b(x_1^4 + x_2^4 + \dots + x_n^4) = x_1^2 y_1 + x_2^2 y_2 + \dots + x_n^2 y_n.$$

So we establish the following table:

n	x	x^2	x^4	y	$x^2 y$
1	10	100	10,000	8	800
1	20	400	160,000	10	4,000
1	30	900	810,000	15	13,500
1	40	1,600	2,560,000	21	33,600
1	50	2,500	6,250,000	30	75,000
5	150	5,500	9,790,000	84	126,900

The normal equations are

$$5a + 5500b = 84. \quad (1)$$

and

$$5500a + 9790000b = 126900,$$

that is,

$$5a + 5500b = 84$$

and

$$55a + 97900b = 1269,$$

or

$$55a + 60500b = 924 \quad (2)$$

and

$$55a + 97900b = 1269 \quad (3)$$

Subtracting equation (2) from (3), we get

$$37400b = 345 \text{ which yields } b = 0.00924.$$

Putting this value in equation (1), we get

$$5a + 50.82 = 84 \text{ which yields } a = 6.76.$$

Hence, $a = 6.76$, $b = 0.00924$ and the parabola of best fit is

$$y = 6.76 + 0.00924x^2$$

EXERCISES

1. Find the least square line for the data given below:

x	-	0	1	2	3	4	5	6
y	1	0	7	5	4	3	0	-1

Ans. $y = -1.60714x + 8.64286$

2. The result of measurement of electric resistance R of a copper wire at various temperatures is listed below.

t	19	25	30	36	40	45	50
R	76	77	79	80	82	83	85

Using the method of least square, find the straight line $R = a + bt$ that fits best in the data.

Ans. $R = 70.052 + 0.290t$

3. The points (1, 14), (2, 27), (3, 40), (4, 55), and (5, 68) should be approximated by a straight line. Find the line assuming that the error in the x values can be neglected.

Ans. $y = 13.6x$

4. Find the power fit $y = ax$ (straight line through the origin) for the data

x	1	2	3	4	5
y	1.6	2.8	4.7	6.4	8.0

Ans. $y = 1.58x, e_2(f) = 0.1720$

5. Find the power fit $y = ax^m$ for the data

x	1	2	3	4	5
y	0.5	2	4.5	8	12.5

Hint: Taking \log we have $\log y = \log a + m \log x$ that is, $Y = A + BX$, where $Y = \log y$, $A = \log a$ and $X = \log x$. Form table in X and Y and find A and B . Then take anti-logarithm to find a and m .

Ans. $y = 0.5012 x^{1.998}$

6. Find the least square parabolic fit $y = a + bx + cx^2$ for the data

x	1	2	3	4
y	1.7	1.8	2.3	3.2

Ans. $y = 1.53 + 0.063x + 0.074x^2$

7 Numerical Differentiation

Let $p(x)$ be an interpolation polynomial approximating satisfactorily a given function $f(x)$ over a certain interval I . We may hope that the result of differentiating $p(x)$ will also satisfactorily approximate the corresponding derivative of $f(x)$. However, if we observe a curve representing the polynomial approximating and oscillating about the curve representing $f(x)$, we may anticipate the fact that even though the deviation between $p(x)$ and $f(x)$ be small throughout the interval, still the slope of the two curves representing them may differ quite appreciably. Also it is seen that the round-off errors of alternating sign in consecutive ordinates could affect the calculation of the derivative quite strongly if those ordinates were fairly closely spaced. That is why, numerical differentiation is considered the weakest concept in the subject of numerical analysis.

7.1 CENTERED FORMULA OF ORDER $O(h^2)$

Let f be a function defined in $[a,b]$. The derivative of f is defined by

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

Suppose further that f has continuous derivatives of order 1, 2, and 3 and that $x-h, x, x+h \in [a, b]$. Then, by Taylor's expansion, we have

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \frac{h^3}{3!} f'''(c_1) \quad (7.1)$$

and

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2!} f''(x) - \frac{h^3}{3!} f'''(c_2). \quad (7.2)$$

Subtracting equation (7.2) from equation (7.1), we get

$$f(x+h) - f(x-h) = 2hf'(x) + \frac{h^3}{3!} [f'''(c_1) + f'''(c_2)].$$

Since $f'''(x)$ is continuous, by intermediate value theorem, there exists a value c such that

$$\frac{f'''(c_1) + f'''(c_2)}{2} = f'''(c).$$

Therefore,

$$f(x+h) - f(x-h) = 2hf'(x) + 2 \frac{h^3}{3!} [f'''(c)]$$

and so

$$\begin{aligned} f'(x) &= \frac{f(x+h) - f(x-h)}{2h} - \frac{f'''(c)h^2}{3!} \\ &= \frac{f(x+h) - f(x-h)}{2h} + E_{\text{trunc}}(f, h) \end{aligned} \quad (7.3)$$

where

$$E_{\text{trunc}}(f, h) = \frac{-h^2}{6} f'''(c) = O(h^2)$$

is called truncation error. Expression (7.3) for the derivative of f is called the centered formula of order $O(h^2)$.

If the third derivative $f'''(c)$ does not change too rapidly, that is, $f'''(c)$ is bounded, then the truncation error in equation (7.3) tends to zero along with h^2 .

7.2 CENTERED FORMULA OF ORDER $O(h^4)$

It is not desirable to choose h too small when computer is used for calculation of derivative. For this reason, a formula for approximating $f'(x)$ and having a truncation error term of order $O(h^4)$ is used.

Suppose f has continuous derivatives of order 1, 2, 3, 4, 5 and $x - 2h, x - h, x, x + h, x + 2h$ be the points in (a, b) . Then, by fifth degree Taylor's expansion, we have

$$f(x+h) - f(x-h) = 2h f'(x) + \frac{2f'''(x)h^3}{3!} + \frac{2f^{(v)}(c_1)h^5}{5!}. \quad (7.4)$$

If we use step size $2h$ instead of h , then

$$f(x+2h) - f(x-2h) = 4h f'(x) + \frac{16f'''(x)h^3}{3!} + \frac{64f^{(v)}(c_2)h^5}{5!}. \quad (7.5)$$

Multiplying both sides of equation (7.4) by 8 and subtracting equation (7.5) from it, we get

$$-f(x+2h) + 8f(x+h) - 8f(x-h) + f(x-2h) = 12h f'(x) + \frac{[16f^{(v)}(c_1) - 64f^{(v)}(c_2)]h^5}{120}.$$

If the sign and magnitude of $f^{(v)}(x)$ does not change rapidly, we can find a value c in $[x - 2h, x + 2h]$ so that

$$16f^{(v)}(c_1) - 64f^{(v)}(c_2) = -48f^{(v)}(c)$$

and so

$$\begin{aligned} f'(x) &\approx \frac{-f(x+2h) + 8f(x+h) - 8f(x-h) + f(x-2h)}{12h} + \frac{f^{(v)}(c)}{30} h^4 \\ &= \frac{-f(x+2h) + 8f(x+h) - 8f(x-h) + f(x-2h)}{12h} + E_{\text{trunc}}(f, h), \end{aligned}$$

where

$$E_{\text{trunc}}(f, h) = \frac{f^{(v)}(c)}{30} h^4 = O(h^4)$$

is the truncation error.

EXAMPLE 7.1

Approximate the derivative of $f(x) = \sin x$ at $x = 0.50$ by

(i) centered formula of order $O(h^2)$

(ii) centered formula of order $O(h^4)$

and determine which of the two yields a better approximation.

Solution. (i) Using centered formula of order $O(h^2)$ and taking spacing $h = 0.01$, we get

$$f'(0.50) \approx \frac{f(0.51) - f(0.49)}{2(0.01)} = \frac{0.488177 - 0.470626}{0.02} = 0.87755.$$

(ii) Using centered formula of order $O(h^4)$, we have

$$\begin{aligned} f'(0.50) &\approx \frac{-f(0.52) + 8f(0.51) - 8f(0.49) + f(0.48)}{0.12} \\ &= \frac{-0.496880 + 3.905418 - 3.765007 + 0.461779}{0.12} = 0.877583. \end{aligned}$$

Since $f'(x) = \cos x$, we have

$$f'(0.50) = 0.877582.$$

Hence, formula of $O(h^4)$ yields better result.

7.3 ERROR ANALYSIS

(A) Error for Centered Formula of Order $O(h^2)$

Suppose $f(x_0 - h)$ and $f(x_0 + h)$ are approximated by y_{-1} and y_1 , and e_{-1} and e_1 are the associated round-off errors, respectively. Then

$$f'(x_0) = \frac{y_1 - y_{-1}}{2h} + E(f, h),$$

where the total error $E(f, h)$ is given by

$$\begin{aligned} E(f, h) &= e_{\text{round}}(f, h) + E_{\text{trunc}}(f, h) \\ &= \frac{e_1 - e_{-1}}{2h} - \frac{h^2}{6} f'''(c). \end{aligned}$$

If $|e_{-1}| \leq \varepsilon$ and $|e_1| \leq \varepsilon$ and $M = \max_{x \in [a, b]} \{|f'''(x)|\}$, then

$$|E(f, h)| \leq \frac{\varepsilon}{h} + \frac{Mh^2}{6} = \frac{6\varepsilon + Mh^3}{6h}. \quad (7.6)$$

The derivative of equation (7.6) is

$$\frac{h(3Mh^2) - (6\varepsilon + Mh^3)}{h^2}.$$

Equating this derivative to zero, we get $2Mh^3 = 6\varepsilon$ and so the value of h that minimizes the right-hand side of equation (7.6) is

$$h = \left(\frac{3\epsilon}{M} \right)^{\frac{1}{3}}.$$

(B) Error for Centered Formula of Order $O(h^4)$

If $f(x_0 + kh) = y_k + e_k$, then

$$f'(x_0) = \frac{-y_2 + 8y_1 - 8y_{-1} + y_{-2}}{12h} + E(f, h),$$

where

$$E(f, h) = \frac{-e_2 + e_1 - e_{-1} + e_{-2}}{12h} + \frac{h^4}{30} f^{(v)}(c).$$

If $|e_k| \leq \epsilon$ and $M = \max_{x \in [a, b]} \{|f^{(v)}(x)|\}$, then

$$|E(f, h)| \leq \frac{3\epsilon}{2h} + \frac{Mh^4}{30} \quad (7.7)$$

and the value of h that minimizes the right-hand side of equation (7.7) is $h = \left(\frac{45\epsilon}{4M} \right)^{\frac{1}{5}}$.

7.4 RICHARDSON'S EXTRAPOLATION

The method of obtaining a formula for $f'(x_0)$ of higher order from a formula of lower order is called Richardson's extrapolation.

Let $D_0(h)$ and $D_0(2h)$ denote the approximations to $f'(x_0)$ obtained from centered formula of order $O(h^2)$ with step size h and $2h$, respectively. Then

$$f'(x_0) \approx D_0(h) + ch^2 \quad (7.8)$$

and

$$f'(x_0) \approx D_0(2h) + 4ch^2. \quad (7.9)$$

Multiplying equation (7.8) by 4 and subtracting equation (7.9) from the product, we get

$$\begin{aligned} 3f'(x_0) &\approx 4D_0(h) - D_0(2h) \\ &= 4 \frac{f_1 - f_{-1}}{2h} - \frac{f_2 - f_{-2}}{4h} \end{aligned}$$

and so

$$\begin{aligned} f'(x_0) &\approx \frac{4D_0(h) - D_0(2h)}{3} \\ &= \frac{-f_2 + 8f_1 - 8f_{-1} + f_{-2}}{12h}, \end{aligned}$$

which is nothing but centered formula of order $O(h^4)$.

Similarly, if $D_1(h)$ and $D_1(2h)$ denote the approximation to $f'(x_0)$, obtained from centered formula of order $O(h^4)$ with step size h and $2h$, respectively, then

$$\begin{aligned} f'(x_0) &= \frac{-f_2 + 8f_1 - 8f_{-1} + f_{-2}}{12h} + \frac{h^4}{30} f^{(v)}(c_1) \\ &\approx D_1(h) + ch^4 \end{aligned} \quad (7.10)$$

and

$$\begin{aligned} f'(x_0) &= \frac{-f_4 + 8f_2 - 8f_{-2} + f_{-4}}{12h} + \frac{16h^4}{30} f^{(v)}(c_2) \\ &\approx D_1(2h) + 16ch^4. \end{aligned} \quad (7.11)$$

Multiplying equation (7.10) by 16 and subtracting equation (7.11) from the product, we get

$$f'(x_0) = \frac{16D_1(h) - D_1(2h)}{15}.$$

In general, if two approximations of order $O(h^{2k})$ for $f'(x_0)$ are $D_{k-1}(h)$ and $D_{k-1}(2h)$ and if

$$f'(x_0) = D_{k-1}(h) + c_1 h^{2k} + c_2 h^{2k+2} + \dots$$

and

$$f'(x_0) = D_{k-1}(2h) + 4^k c_1 h^{2k} + 4^{k+1} c_2 h^{2k+2} + \dots,$$

then an improved approximation is of the form

$$\begin{aligned} f'(x_0) &= D_k(h) + O(h^{2k+2}) \\ &= \frac{4^k D_{k-1}(h) - D_{k-1}(2h)}{4^k - 1} + O(h^{2k+2}). \end{aligned}$$

This result is known as Richardson's extrapolation.

EXAMPLE 7.2

The voltage $E(t)$ in an electrical circuit obeys the equation

$$E(t) = L \frac{dI}{dt} + RI(t),$$

where L is the inductance and R is the resistance. If $L = 0.05$, $R = 2$ and $I(t)$ at time t is given by the table

$t:$	1.0	1.1	1.2	1.3	1.4
$I(t):$	8.2277	7.2428	5.9908	4.5260	2.9122

Find $I'(1.2)$ by numerical differentiation and compute $E(1.2)$.

Solution. Using centered formula of order $O(h^2)$, we have

$$\begin{aligned} I'(1.2) &\approx \frac{I(1.2) - I(1.0)}{2h} = \frac{5.9908 - 8.2277}{2(0.1)} \\ &= -13.5840 \end{aligned}$$

and then

$$\begin{aligned} E(1.2) &\approx 0.05(-13.5840) + 2(5.9908) \\ &= -0.6792 + 11.9816 = 11.3024. \end{aligned}$$

If we use centered formula of order $O(h^4)$, then

$$\begin{aligned} I'(1.2) &\approx \frac{-I(x+2h) + 8I(x+h) - 8I(x-h) + I(x-2h)}{12h} \\ &= \frac{-2.9122 + 8(4.5260) - 8(7.2428) + 8.2277}{12(0.1)} \\ &= \frac{-2.9122 + 36.2080 - 57.9424 + 8.2277}{1.2} \\ &= \frac{-60.8546 + 49.4357}{1.2} = -13.6824 \end{aligned}$$

and so

$$\begin{aligned} E(1.2) &\approx 0.05(-13.6824) + 2(5.9908) \\ &= -0.6841 + 11.9816 = 11.2975. \end{aligned}$$

EXAMPLE 7.3

Find $I'(1.2)$ in the Example 7.2 using Richardson's extrapolation.

Solution. We have

$$\begin{aligned} D_0(h) &\approx \frac{I(x+h) - I(x-h)}{2h} = \frac{I(1.3) - I(1.1)}{2(0.1)} \\ &= \frac{4.5260 - 7.2428}{0.2} = -13.5840 \\ D_0(2h) &\approx \frac{I(x+2h) - I(x-2h)}{4h} = \frac{I(1.4) - I(1.0)}{0.4} \\ &= \frac{2.9122 - 8.227}{0.4} = -13.28875. \end{aligned}$$

Therefore,

$$\begin{aligned} I'(1.2) &\approx \frac{4D_0(h) - D_0(2h)}{3} \\ &= \frac{4(-13.5840) + 13.28875}{3} \\ &= \frac{-54.3360 + 13.28875}{3} = -13.6824. \end{aligned}$$

We observe that this value is exactly that we found by centered formula of order $O(h^4)$.

EXAMPLE 7.4

From the following table, find $f'(1.4)$.

$x:$	1.2	1.3	1.4	1.5	1.6
$f(x):$	1.5095	1.6984	1.9043	2.1293	2.3756

Solution. Using centered formula of order (h^4) and proceeding as in Example 7.2, we have

$$\begin{aligned}
f'(1.4) &\approx \frac{-f(x+2h)+8f(x+h)-8f(x-h)+f(x-2h)}{12h} \\
&= \frac{-2.3756+8(2.1293)-8(1.6984)+1.5095}{12(0.1)} \\
&= \frac{-2.3756+17.0344-13.5872+1.5095}{1.2} = 2.1509.
\end{aligned}$$

7.5 CENTRAL DIFFERENCE FORMULA OF ORDER $O(h^4)$ FOR $f''(x)$

By Taylor's expansion

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \frac{h^3}{3!} f'''(x) + \frac{h^4 f^{(iv)}(x)}{4!} + \dots \quad (7.12)$$

and

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2!} f''(x) - \frac{h^3}{3!} f'''(x) + \frac{h^4}{4!} f^{(iv)}(x) + \dots \quad (7.13)$$

Adding equations (7.12) and (7.13), we have

$$f(x+h) + f(x-h) = 2f(x) + \frac{2h^2}{2!} f''(x) + \frac{2h^4}{4!} f^{(iv)}(x) + \dots,$$

which yields

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} - \frac{2h^2}{24} f^{(iv)}(x) - \dots$$

Truncating at the fourth derivative, we get

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} - \frac{h^2}{12} f^{(iv)}(c)$$

and hence the desired formula is

$$f''(x_0) \approx \frac{f_1 - 2f_0 + f_{-1}}{h^2}. \quad (7.14)$$

Let $f_k = y_k + e_k$, where e_k is the error in computing. Then the total error in equation (7.14) is

$$E(f, h) = \frac{e_1 - 2e_0 + e_{-1}}{h^2} - \frac{h^2}{12} f^{(iv)}(c).$$

If $|e_k| \leq \varepsilon$ and $|f^{(iv)}(c)| \leq M$, then

$$|E(f, h)| \leq \frac{4\varepsilon}{h^2} + \frac{Mh^2}{12}.$$

Differentiating the right-hand side and equating the differential to zero gives

$$h = \left(\frac{48\varepsilon}{M} \right)^{\frac{1}{4}}$$

for minimum error.

The central difference formulae of order $O(h^2)$ for $f'''(x_0)$ and $f^{(iv)}(x_0)$ are, respectively,

$$f'''(x_0) \approx \frac{f_2 - 2f_1 + 2f_{-1} - f_{-2}}{2h^3},$$

$$f^{(iv)}(x_0) \approx \frac{f_2 - 4f_1 + 6f_0 - 4f_{-1} + f_{-2}}{h^4}.$$

On the other hand, central difference formulae of order $O(h^4)$ for $f''(x_0)$, $f'''(x_0)$, and $f^{(iv)}(x_0)$ are, respectively,

$$f''(x_0) \approx \frac{-f_2 + 16f_1 - 30f_0 + 16f_{-1} - f_{-2}}{12h^2},$$

$$f'''(x_0) \approx \frac{-f_3 + 8f_2 - 13f_1 + 13f_{-1} - 8f_2 + f_{-3}}{8h^3},$$

$$f^{(iv)}(x_0) \approx \frac{-f_3 + 12f_2 - 39f_1 + 56f_0 - 39f_{-1} + 12f_{-2} - f_{-3}}{6h^4}.$$

7.6 GENERAL METHOD FOR DERIVING DIFFERENTIATION FORMULAE

Suppose the function f is analytic and tabulated at equidistant points. We know that

$$e^{hD} = E = I + \Delta.$$

Therefore,

$$hD = \log(I + \Delta) = \Delta - \frac{\Delta^2}{2} + \frac{\Delta^3}{3} - \frac{\Delta^4}{4} + \dots$$

and so

$$D = \frac{1}{h} \left[\Delta - \frac{\Delta^2}{2} + \frac{\Delta^3}{3} - \frac{\Delta^4}{4} + \dots \right].$$

Hence,

$$f'(x) = \frac{1}{h} \left[\Delta f(x) - \frac{\Delta^2}{2} f(x) + \frac{\Delta^3}{3} f(x) - \frac{\Delta^4}{4} f(x) + \dots \right]. \quad (7.15)$$

To find second derivative, we have

$$\begin{aligned} h^2 D^2 &= (\log E)^2 = (\log(1 + \Delta))^2 \\ &= \left[\Delta - \frac{\Delta^2}{2} + \frac{\Delta^3}{3} - \frac{\Delta^4}{4} + \dots \right]^2 \end{aligned}$$

and so

$$D^2 = \frac{1}{h^2} \left[\Delta - \frac{\Delta^2}{2} + \frac{\Delta^3}{3} - \frac{\Delta^4}{4} + \dots \right]^2.$$

Hence,

$$f''(x) = \frac{1}{h^2} \left[\Delta - \frac{\Delta^2}{2} + \frac{\Delta^3}{3} - \frac{\Delta^4}{4} + \dots \right]^2 f(x). \quad (7.16)$$

In terms of central differences, we know that

$$U = hD = 2 \sinh^{-1} \left(\frac{\delta}{2} \right).$$

If we put $f(x) = \sinh^{-1} x$, then

$$f'(x) = (1 + x^2)^{-\frac{1}{2}} = 1 - \frac{1}{2}x^2 + \frac{3}{8}x^4 - \dots,$$

which on integration yields

$$f(x) = x - \frac{x^3}{6} + \frac{3}{40}x^5 - \dots$$

Thus,

$$f\left(\frac{\delta}{2}\right) = \sinh^{-1}\left(\frac{\delta}{2}\right) = \frac{\delta}{2} - \frac{\delta^3}{48} + \frac{3}{1280}\delta^5 - \dots$$

and so

$$hD = \delta - \frac{\delta^3}{24} + \frac{3}{640}\delta^5 - \frac{5}{7168}\delta^7 + \dots \quad (7.17)$$

Hence,

$$D = \frac{1}{h} \left[\delta - \frac{\delta^3}{24} + \frac{3}{640}\delta^5 - \frac{5}{7168}\delta^7 + \dots \right]$$

and we have

$$f'(x) = \frac{1}{h} \left[\delta f(x) - \frac{\delta^3}{24} f(x) + \frac{3}{640} \delta^5 f(x) - \frac{5}{7168} \delta^7 f(x) + \dots \right]. \quad (7.18)$$

Squaring equation (7.17), we have

$$h^2 D^2 = \delta^2 - \frac{1}{12}\delta^4 + \frac{1}{90}\delta^6 - \dots$$

or

$$D^2 = \frac{1}{h^2} \left[\delta^2 - \frac{1}{12}\delta^4 + \frac{1}{90}\delta^6 - \dots \right]$$

and so we have

$$f''(x) = \frac{1}{h^2} \left[\delta^2 f(x) - \frac{1}{12} \delta^4 f(x) + \frac{1}{90} \delta^6 f(x) - \frac{1}{560} \delta^8 f(x) + \dots \right]. \quad (7.19)$$

The above derived formulae (7.15), (7.16), (7.18), and (7.19) yield derivatives at the nodes. We now seek derivatives at interior points. Let $x = x_0 + ph$. Then $dx = hdp$. Therefore,

$$\frac{df}{dx} = \frac{df}{hdp}.$$

Newton's forward difference formula states that

$$f_p = f_0 + p\Delta f_0 + \frac{p(p-1)}{2!} \Delta^2 f_0 + \frac{p(p-1)}{3!} \Delta^3 f_0 + \dots$$

Therefore,

$$\frac{df_p}{dx} = \frac{df_p}{hdp} = \frac{1}{h} \left[\Delta f_0 + \frac{2p-1}{2!} \Delta^2 f_0 + \frac{3p^2-6p+2}{3!} \Delta^3 f_0 + \dots \right]. \quad (7.20)$$

Usually, we are interested in the derivative at a tabular point x_0 or at a mid-interval $x_{1/2}$. These are obtained by putting $p=0$ and $\frac{1}{2}$, respectively. Thus, we get

$$f'_0 = \frac{1}{h} \left[\Delta f_0 - \frac{1}{2} \Delta^2 f_0 + \frac{1}{3} \Delta^3 f_0 + \dots \right] \quad (7.21)$$

and

$$f'_{1/2} = \frac{1}{h} \left[\Delta f_0 + \frac{1}{12} \Delta^3 f_0 + \dots \right]. \quad (7.22)$$

When the point is midway the table, then we use central differences formulae. For example, if we take Bessel's formula

$$f_p = f_0 + p\delta f_{1/2} + \frac{p(p-1)}{2!} \left(\frac{\delta^2 f_0 + \delta^2 f_1}{2} \right) + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{1/2} + \dots$$

then

$$\frac{df_p}{dx} = \frac{df_p}{hdp} = \frac{1}{h} \left[\delta f_{1/2} + \frac{2p-1}{(2)2!} (\delta^2 f_0 + \delta^2 f_1) + \frac{3p^2-3p+\frac{1}{2}}{3!} \delta^3 f_{1/2} + \dots \right]. \quad (7.23)$$

If we put $p=0$, then $x=x_0$ and we have

$$f'_0 = \frac{1}{h} \left[\delta f_{1/2} - \frac{1}{4} (\delta^2 f_0 + \delta^2 f_1) + \frac{1}{12} \delta^3 f_{1/2} + \dots \right]. \quad (7.24)$$

If we put $p=\frac{1}{2}$, the coefficients of even differences become zero and we have

$$f'_{1/2} = \frac{1}{h} \left[\delta f_{1/2} - \frac{1}{24} \delta^3 f_{1/2} + \frac{3}{640} \delta^5 f_{1/2} - \dots \right]. \quad (7.25)$$

If we use Everett's formula, then we have

$$f_p = (1-p)f_0 - \frac{p(p-1)(p-2)}{3!} \delta^2 f_0 + \dots + pf_1 + \frac{p(p-1)(p-2)}{3!} \delta^2 f_1 + \dots$$

Therefore,

$$f'_p = \frac{1}{h} \left[f_1 - f_0 + \frac{3p^2 - 1}{3!} \delta^2 f_1 - \frac{3p^2 - 6p + 2}{3!} \delta^2 f_0 + \dots \right]. \quad (7.26)$$

Similarly, Stirling's formula reads that

$$f_p = f_0 + p\mu\delta f_0 + \frac{p^2}{2} \delta^2 f_0 + \frac{(p+1)p(p-1)}{3!} \mu\delta^3 f_0 + \dots + \frac{p^2(p+1)(p-1)}{(4)3!} \delta^4 f_0 + \dots.$$

Therefore,

$$f'_p = \frac{1}{h} \left[\mu\delta f_0 + p\delta^2 f_0 + \frac{3p^2 - 1}{6} \mu\delta^3 f_0 + \frac{4p^3 - 2p}{24} \delta^4 f_0 + \dots \right]. \quad (7.27)$$

Putting $p = 0$, we get

$$\begin{aligned} f'_0 &= \frac{1}{h} \left[\mu\delta f_0 - \frac{1}{6} \mu\delta^3 f_0 + \frac{1}{30} \mu\delta^5 f_0 - \dots \right] \\ &= \frac{1}{2h} \left[(f_1 - f_{-1}) - \frac{1}{6} (\delta^2 f_1 - \delta^2 f_{-1}) + \frac{1}{30} (\delta^4 f_1 - \delta^4 f_{-1}) - \dots \right]. \end{aligned} \quad (7.28)$$

Also,

$$f''_p = \frac{1}{h^2} \left[\delta^2 f_0 + p\mu\delta^3 f_0 + \frac{p^2 - 1}{12} \delta^4 f_0 + \dots \right] \quad (7.29)$$

Putting $p = 0$, we have

$$f''_0 = \frac{1}{h^2} \left[\delta^2 f_0 - \frac{1}{12} \delta^4 f_0 + \frac{1}{90} \delta^6 f_0 - \dots \right]. \quad (7.30)$$

From equation (7.30), we have

$$h^2 D^2 = \delta^2 = \left(1 - \frac{1}{12} \delta^2 + \frac{1}{90} \delta^4 - \dots \right)$$

or

$$\begin{aligned} \delta^2 &= h^2 \left(1 - \frac{\delta^2}{12} + \frac{1}{90} \delta^4 - \dots \right)^{-1} D^2 \\ &= h^2 \left(1 + \frac{\delta^2}{12} - \frac{1}{240} \delta^4 + \dots \right) D^2 \end{aligned}$$

and so

$$\delta^2 f_0 = h^2 = \left(f''_0 + \frac{1}{12} \delta^2 f_0 - \frac{1}{240} \delta^4 f_0'' + \dots \right),$$

which expresses the second difference of f in terms of second and higher differences of f'' .

EXAMPLE 7.5

The function $y = \sin x$ is tabulated below. Find the derivative at the point $x = 1$.

$x:$	0.7	0.8	0.9	1.0	1.1	1.2	1.3
$y:$	0.644218	0.717356	0.783327	0.841471	0.891207	0.932039	0.963558

Solution. The difference table for the given data is

x	y	δ	δ^2	δ^3	δ^4
0.7	0.644218	0.073138			
0.8	0.717356	0.065971	-0.007167	-0.00660	
0.9	0.783327	0.058144	-0.007827	-0.00581	0.000079
1.0	0.841471	0.049736	-0.008408	-0.00496	0.000085
1.1	0.891207	0.040832	-0.008904	-0.00409	0.000087
1.2	0.932039	0.031519	-0.009313		
1.3	0.963558				

Since $x=1$ is tabulated argument, we have $p=0$ and it will be better to use Stirling's formula. Using

$$\begin{aligned}
 f'_p &= \frac{1}{h} \left[\mu \delta f_0 - \frac{1}{6} \mu \delta^3 f_0 + \frac{1}{30} \mu \delta^5 f_0 - \dots \right] \\
 &= \frac{1}{2h} \left[\left(\delta f_{\frac{1}{2}} + \delta f_{-\frac{1}{2}} \right) - \frac{1}{6} \left(\delta^3 f_{\frac{1}{2}} + \delta^3 f_{-\frac{1}{2}} \right) + \frac{1}{30} \left(\delta^5 f_{\frac{1}{2}} + \delta^5 f_{-\frac{1}{2}} \right) + \dots \right] \\
 &= \frac{1}{2h} \left[(f_1 - f_{-1}) - \frac{1}{6} (\delta^2 f_1 - \delta^2 f_{-1}) + \frac{1}{30} (\delta^4 f_1 - \delta^4 f_{-1}) + \dots \right] \\
 &\approx \frac{1}{0.2} \left[(0.891207 - 0.783327) - \frac{1}{6} (0.008904 - 0.007827) + \frac{1}{30} (0.000087 - 0.000079) \right] \\
 &= \frac{1}{0.2} [0.107880 - 0.0001795 + 0.0000003] = 0.538504.
 \end{aligned}$$

The tabulated value of $\cos 1$ is 0.540302. Thus, the computed value of the derivative is in good agreement with the tabulated value.

EXAMPLE 7.6

The function $y=f(x)$ has a minimum in the interval $0.2 < x < 1.4$. Find the x coordinate of the minimum point.

$x:$	0.2	0.4	0.6	0.8	1.0	1.2	1.4
$y=f(x):$	2.10022	1.98730	1.90940	1.86672	1.85937	1.88737	1.95063

Solution.

x	y	δ	δ^2	δ^3	δ^4	δ^5
0.2	2.10022	-0.11292				
0.4	1.98730	-0.07790	0.03502	0.00020		
0.6	1.90940	-0.04268	0.03522	0.00011	-0.00009	
0.8	1.86672	-0.00735	0.03533	0.00002	-0.00009	0
1.0	1.85937	0.02800	0.03535	-0.00009	-0.00011	-0.00002
1.2	1.88737	0.06326	0.03526			
1.4	1.95063					

Taking $x_0 = 0.80$, we shall use Everett's formula

$$f_p = (1-p)f_0 - \frac{p(p-1)(p-2)}{3!} \delta^2 f_0 + \dots + pf_1 + \frac{(p-1)p(p-2)}{3!} \delta^2 f_1 + \dots$$

and obtain

$$\begin{aligned} f'_p &= \frac{1}{h} \left[f_1 - f_0 + \frac{3p^2 - 1}{3!} \delta^2 f_1 - \frac{3p^2 - 6p + 2}{3!} \delta^2 f_0 + \dots \right] \\ &= \frac{1}{h} \left[-0.00735 + \frac{3p^2 - 1}{3!} (0.03535) - \frac{3p^2 - 6p + 2}{3!} (0.03533) + \dots \right] \\ &\approx \frac{1}{h} \left[-0.00735 + \frac{0.03535}{6} (3p^2 - 1 - 3p^2 + 6p - 2) \right] \end{aligned}$$

since $0.03533 \approx 0.03535$. Now for minimum, $f'_p = 0$ and so we get

$$-0.00735 + \frac{0.03535}{6} (6p - 3) = 0,$$

which yields

$$p = \frac{0.025025}{0.03535} = 0.707921.$$

Therefore,

$$\begin{aligned} x &= x_0 + ph = 0.80 + (0.707921)(0.2) \\ &= 0.80 + 0.1415842 = 0.9416. \end{aligned}$$

EXAMPLE 7.7

y is a function of x satisfying the differential equation $xy'' + ay' + (x - b)y = 0$, where a and b are known to be integers. Find the constants a and b from the table below:

$x:$	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2
$y:$	1.73036	1.95532	2.19756	2.45693	2.73309	3.02549	3.33334	3.65563

Solution.

x	y	δ	δ^2	δ^3	δ^4
0.8	1.73036				
1.0	1.95532	0.22496	0.01728	-0.00015	
1.2	2.19756	0.24224	0.01713	-0.00034	-0.00019
1.4	2.45693	0.25937	0.01679	-0.00055	-0.00021
1.6	2.73309	0.27616	0.01624	-0.00079	-0.00024
1.8	3.02549	0.29240	0.01545	-0.00101	-0.00022
2.0	3.33334	0.30785	0.01444		
2.2	3.65563	0.32229			

We have to find two constants a and b and so we require two equations. So, we evaluate y' and y'' at two points say 1.4 and 1.6. Thus, $p = 0$ in this case and the formula to be used will be that derived from Stirling's formula. We have

$$y'_p = \frac{1}{2h} \left[y_1 - y_{-1} - \frac{1}{6}(\delta^2 y_1 - \delta^2 y_{-1}) + \frac{1}{30}(\delta^4 y_1 - \delta^4 y_{-1}) + \dots \right]$$

and

$$y''_p = \frac{1}{h^2} \left[\delta^2 y_0 - \frac{1}{12} \delta^4 y_0 + \frac{1}{90} \delta^6 y_0 + \dots \right].$$

Thus,

$$\begin{aligned} y'(1.4) &\approx \frac{1}{0.4} \left[(2.73309 - 2.19756) - \frac{1}{6}(0.01624 - 0.01713) + \frac{1}{30}(-0.00024 + 0.00019) + \dots \right] \\ &= \frac{1}{0.4} [0.53553 + 0.000148 - 0.000002] \\ &= 1.33919, \end{aligned}$$

$$y''(1.4) \approx \frac{1}{0.04} \left[(0.01679 - \frac{1}{12}(-0.00021)) \right] = 0.419325,$$

$$\begin{aligned} y'(1.6) &\approx \frac{1}{0.4} \left[(3.02549 - 2.45693) - \frac{1}{6}(0.01545 - 0.01679) + \frac{1}{30}(-0.00022 + 0.00021) \right] \\ &= \frac{1}{0.4} [0.56856 + 0.00022 - 0.0000003] = 1.42195, \end{aligned}$$

$$y''(1.6) \approx \frac{1}{0.04} \left[0.01624 - \frac{1}{12}(-0.00024) \right] = 0.4065.$$

Thus, we have two equations

$$(1.4)(0.419325) + a(1.33919) + (1.4 - b)(2.45693) = 0$$

$$(1.6)(0.40650) + a(1.42195) + (1.6 - b)(2.73309) = 0$$

or

$$1.33919a - 2.45693b = -4.026757$$

$$1.42195a - 2.73309b = -5.023344.$$

We shall use Cramer's rule to find a and b . We have

$$\Delta = 3.6601268 + 3.493632 = 0.1664948$$

$$\Delta_1 = 11.005489 - 12.34200 = 1.336515$$

$$\Delta_2 = -1.001365$$

and so

$$a = \frac{\Delta_1}{\Delta} = 8.0273 \text{ and } b = \frac{\Delta_2}{\Delta} = 6.0143.$$

The true values of a and b are 8 and 6, respectively.

EXAMPLE 7.8

A function $y = f(x)$ is given in the table below. The function is a solution of the equation $x^2 y'' + xy' + (x^2 - n^2)y = 0$, where n is a positive integer. Find n .

$x:$	85	85.01	85.02	85.03	85.04
$y:$	0.0353878892	0.0346198696	0.0338490002	0.0330753467	0.032298975

Solution. The difference table for the given data is

x	y	δ	δ^2	δ^3	δ^4
85.0	0.0353878892				
85.01	0.0346198696	-0.0007680196		-0.0000028498	
85.02	0.0338490002	-0.0007708694	-0.0000027841	0.00000000657	
85.03	0.0330753467	-0.0007736535	-0.0000027182	0.0000000659	$\frac{2}{10^9}$
85.04	0.0322989750	-0.0007763717	-0.0000027182		

Differentiating Stirling's formula, the value of f' at $p = 0$ and the value of f'' at $p = 0$ are

$$f'_0 = \frac{1}{2h} \left[(f_1 - f_{-1}) - \frac{1}{6}(\delta^2 f_1 - \delta^2 f_{-1}) + \frac{1}{30}(\delta^4 f_1 - \delta^4 f_{-1}) - \dots \right]$$

$$f''_0 = \frac{1}{h^2} \left[\delta^2 f_0 - \frac{1}{12} \delta^4 f_0 + \dots \right].$$

We calculate y'_0 and y''_0 at $x = 85.02$. We have

$$\begin{aligned}
 y'_0 &\approx \frac{1}{0.02} [0.0330753467 - 0.0346198676] - \frac{1}{6} \{(-0.0000027182) + 0.0000028498\} + \dots \\
 &= \frac{1}{0.02} [-0.0015445229 - 0.0000002193] = -0.077227, \\
 y''_0 &= \frac{1}{0.001} \left[\delta^2 f_0 - \frac{1}{12} \delta^4 f_0 \right] = -0.02784116.
 \end{aligned}$$

Putting these values in $x^2 y'' + xy' + (x^2 - n^2)y = 0$, we have

$$0.033849002n^2 = -201.238667136 - 6.56583954 + 224.6741116.$$

Thus, $n^2 = 1089.23$ and so $n \approx \pm 33.003$. Hence, $n = 33$ is the required value.

EXAMPLE 7.9

Given that

$x:$	1.0	1.1	1.2	1.3	1.4	1.5
$y:$	7.989	8.403	8.781	9.129	9.451	9.750

Find $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ at $x = 1.6$.

Solution. Differentiating Newton's backward formula, we get

$$\left[\frac{dy}{dx} \right]_{x=x_n} = \frac{1}{h} \left[\nabla y_n + \frac{1}{2} \nabla^2 y_n + \frac{1}{3} \nabla^3 y_n + \dots \right]$$

and

$$\left(\frac{d^2y}{dx^2} \right)_{x=x_n} = \frac{1}{h^2} \left[\nabla^2 y_n + \nabla^3 y_n + \frac{11}{12} \nabla^4 y_n + \dots \right].$$

The difference table is

x	y						
1.0	7.989	0.414	-0.036	0.006	-0.002	0.001	0.002
1.1	8.403	0.378	-0.030	0.004	-0.001	0.003	
1.2	8.781	0.348	-0.026	0.003	0.002		
1.3	9.129	0.322	-0.023				
1.4	9.451						

x	y						
1.5	9.750	0.299	-0.018	0.005			
1.6	10.031	0.281					

Therefore, for the given spacing $h = 0.1$, we have

$$\left(\frac{dy}{dx} \right)_{x=1.6} = \frac{1}{0.1} \left[0.281 + \frac{1}{2}(-0.018) + \frac{1}{3}(0.005) + \frac{1}{4}(0.002) \right] = 2.7416.$$

and

$$\left(\frac{d^2y}{dx^2} \right)_{x=1.6} = \frac{1}{0.01} \left[-0.018 + 0.005 + \frac{11}{12}(0.02) \right] = -1.117.$$

7.7 DIFFERENTIATION OF A FUNCTION TABULATED IN UNEQUAL INTERVALS

Let f be a function continuously differentiable in the interval $[c,d]$. If x_0, x_1, \dots, x_n are distinct points in $[c,d]$, then

$$f(x) = P_n(x) + f[x_0, x_1, \dots, x_n, x] \Psi_n(x), \quad (7.31)$$

where $P_n(x)$ is a polynomial of degree $\leq n$ which interpolates $f(x)$ at x_0, x_1, \dots, x_n , and

$$\Psi_n(x) = \prod_{i=0}^n (x - x_i).$$

Also,

$$\frac{d}{dx} f[x_0, x_1, \dots, x_n, x] \neq [x_0, x_1, \dots, x_n, x, x].$$

Therefore, differentiating equation (7.31), we get

$$f'(x) = P'_n(x) + f[x_0, x_1, \dots, x_n, x, x] \Psi_n(x) + f[x_0, x_1, \dots, x_n, x] \Psi'_n(x).$$

Thus, if $a \in [c,d]$, then

$$f'(a) = P'_n(a) + f[x_0, x_1, \dots, x_n, a, a] \Psi_n(a) + f[x_0, x_1, \dots, x_n, a] \Psi'_n(a)$$

and so if we approximate $f''(a)$ by $P''_n(a)$, the error in the approximation is

$$\begin{aligned} E(f) &= f[x_0, x_1, \dots, x_n, a, a] \Psi_n(a) + f[x_0, x_1, \dots, x_n, a] \Psi'_n(a) \\ &= \frac{f^{(n+2)}(\xi) \Psi_n(a)}{(n+2)!} + \frac{f^{(n+1)}(\eta) \Psi'_n(a)}{(n+1)!} \end{aligned}$$

for some $\xi, \eta \in [c,d]$.

In light of the above discussion, we can derive derivative formulae differentiating Lagrange's polynomial.

7.8 DIFFERENTIATION OF LAGRANGE'S POLYNOMIAL

Consider the Lagrange's interpolation polynomial for $f(x)$ based on three points x_0, x_1 , and x_2 . We have

$$f(x) \approx f_0 \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + f_1 \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} + f_2 \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}.$$

Differentiating, we get

$$f'(x_0) \approx f_0 \frac{(x - x_1) + (x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + f_1 \frac{(x - x_0) + (x - x_2)}{(x_1 - x_0)(x_1 - x_2)} + f_2 \frac{(x - x_0) + (x - x_1)}{(x_2 - x_0)(x_2 - x_1)}$$

and so

$$f'(x_0) \approx f_0 \frac{(x_0 - x_1) + (x_0 - x_2)}{(x_0 - x_1)(x_0 - x_2)} + f_1 \frac{(x_0 - x_0) + (x_0 - x_2)}{(x_1 - x_0)(x_1 - x_2)} + f_2 \frac{(x_0 - x_0) + (x_0 - x_1)}{(x_2 - x_0)(x_2 - x_1)}.$$

But $x_1 - x_0 = h$, $x_2 - x_0 = 2h$. Therefore,

$$\begin{aligned} f'(x_0) &\approx f_0 \frac{(-h) + (-2h)}{(-h)(-2h)} + f_1 \frac{(-2h)}{(h)(-h)} + f_2 \frac{(-h)}{(2h)(h)} \\ &= \frac{f_0}{2h^2} (-3h) + f_1 \left(\frac{2h}{h^2} \right) + f_2 \frac{(-1)}{2h} \\ &= \frac{-3f_0 + 4f_1 - f_2}{2h} \end{aligned} \tag{7.32}$$

which is first order differential formula.

If we consider Lagrange's interpolation polynomial for $f(x)$ based on four points x_0, x_1, x_2 , and x_3 , then

$$\begin{aligned} f(x) &\approx f_0 \frac{(x - x_1)(x - x_2)(x - x_3)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)} + f_1 \frac{(x - x_0)(x - x_2)(x - x_3)}{(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)} \\ &\quad + f_2 \frac{(x - x_0)(x - x_1)(x - x_3)}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)} + f_3 \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)}. \end{aligned}$$

Differentiating twice, we get

$$\begin{aligned} f''(x) &\approx f_0 \frac{2[(x - x_1) + (x - x_2) + (x - x_3)]}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)} + f_1 \frac{2[(x - x_0) + (x - x_2) + (x - x_3)]}{(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)} \\ &\quad + f_2 \frac{2[(x - x_0) + (x - x_1) + (x - x_3)]}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)} + f_3 \frac{2[(x - x_0) + (x - x_1) + (x - x_2)]}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)}. \end{aligned}$$

Putting $x = x_0$ and taking $x_i - x_j = (i - j)h$, we get

$$\begin{aligned} f''(x_0) &\approx f_0 \frac{2[(-h) + (-2h) + (-3h)]}{(-h)(-2h)(-3h)} + f_1 \frac{2[0 + (-2h) + (-3h)]}{h(-h)(-2h)} + f_2 \frac{2[0 + (-h) + (-3h)]}{(2h)h(-h)} + f_3 \frac{2[0 + (-h) + (-2h)]}{(3h)(2h)h} \\ &= f_0 \left(\frac{-12h}{-6h^3} \right) + f_1 \left(\frac{-10h}{2h^3} \right) + f_2 \left(\frac{-8h}{-2h^3} \right) + f_3 \left(\frac{-6h}{6h^3} \right) \\ &= \frac{2f_0 - 5f_1 + 4f_2 - f_3}{h^2} \end{aligned} \tag{7.33}$$

which is second order derivative formula.

EXAMPLE 7.10

Find the approximations to $f'(x_n)$ of order $O(h^2)$ at $x = 0$ and $x = 0.1$ in the following table:

$x:$	0.0	0.1	0.2	0.3
$f(x):$	0.989992	0.999135	0.998295	0.987480

Solution. Using first order differential formula (7.32), we have

$$\begin{aligned} f'(0) &\approx \frac{-3(0.989992) + 4(0.999135) - 0.998295}{2(0.1)} \\ &= \frac{-2.969976 + 3.99654 - 0.998295}{0.2} \\ &= \frac{3.99654 - 3.968271}{0.2} = \frac{0.028269}{0.2} \\ &= 0.141345 \end{aligned}$$

and

$$\begin{aligned} f'(0.1) &\approx \frac{-3(0.999135) + 4(0.998295) - 0.987480}{0.2} \\ &= \frac{-2.997405 + 3.99318 - 0.987480}{0.2} \\ &= 0.041475. \end{aligned}$$

7.9 DIFFERENTIATION OF NEWTON POLYNOMIAL

Consider the Newton polynomial $P(x)$ based on the three nodes x_0 , x_1 , and x_2 . We have

$$P(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1), \quad (7.34)$$

where

$$\begin{aligned} a_0 &= f(x_0), \quad a_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0}, \\ a_2 &= \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0}. \end{aligned}$$

Differentiating equation (7.34) with respect to x , we get

$$P'(x) = a_1 + a_2[(x - x_0) + (x - x_1)] \quad (7.35)$$

and so

$$P'(x_0) = a_1 + a_2(x_0 - x_1).$$

Thus,

$$f'(x_0) \approx P'(x_0) = a_1 + a_2(x_0 - x_1). \quad (7.36)$$

If we set $x_0 = x$, $x_1 = x + h$, and $x_2 = x_0 + 2h$, then

$$\begin{aligned} a_1 &= \frac{f(x_1) - f(x)}{x_1 - x} = \frac{f(x+h) - f(x)}{h}, \\ a_2 &= \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x)}{x_1 - x}}{x_2 - x} \\ &= \frac{\frac{f(x+2h) - f(x+h)}{h} - \frac{f(x+h) - f(x)}{h}}{2h} \\ &= \frac{f(x) - 2f(x+h) + f(x+2h)}{2h^2}, \end{aligned}$$

and so equation (7.35) becomes

$$\begin{aligned} P'(x) &= \frac{f(x+h) - f(x)}{h} + \frac{f(x) - 2f(x+h) + f(x+2h)}{2h^2}(x - x_1) \\ &= \frac{f(x+h) - f(x)}{h} + \frac{f(x) - 2f(x+h) + f(x+2h)}{2h^2}(-h) \\ &= \frac{-3f(x) + 4f(x+h) - f(x+2h)}{2h} \end{aligned}$$

and so

$$f'(x) \approx P'(x) = \frac{3f(x) + 4f(x+h) - f(x+2h)}{2h},$$

which is second order forward difference formula for $f'(x)$ (see equation 7.32).

EXAMPLE 7.11

Find the maximum value of $f(x)$ using the table given below:

$x:$	-	1	2	3
$f(x):$	-21	15	12	3

Solution. We note that the arguments given are not equispaced. Therefore, we shall use Newton's divided difference formula. The divided difference table is

x	$f(x)$	$f(x_0, x_1)$	$f(x_0, x_1, x_2)$	$f(x_0, x_1, x_2, x_3)$
-1	-21			
1	15	18		
2	12	-3	-7	
3	3	-9	-3	1

The divided difference formula yields

$$\begin{aligned} f(x) &= f_0 + (x - x_0)f(x_0, x_1) + (x - x_0)(x - x_1)f(x_0, x_1, x_2) + (x - x_0)(x - x_1)(x - x_2)f(x_0, x_1, x_2, x_3) \\ &= -21 + (x+1)(18) + (x+1)(x-1)(-7) + (x+1)(x-1)(x-2)(1) \\ &= x^3 - 9x^2 + 17x + 6. \end{aligned}$$

Therefore,

$$f'(x) = 3x^2 - 18x + 17.$$

For maximum value, we should have $3x^2 - 18x + 17 = 0$. This equation yields $x = 4.8257$ or 1.1743 . The value 1.1743 is the value in the given range. Then

$$\text{Maximum value of } f(x) = (1.1743)^3 - 9(1.1743)^2 + 17(1.1743) + 6 \approx 15.1716.$$

EXAMPLE 7.12

Evaluate the first derivative at $x = -3$ and $x = 0$ from the following table:

$x:$	-3	-2	-1	0	1	2	3
$y:$	-33	-12	-3	0	3	12	33

Solution. The difference table for the given problem is

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
-3	-33				
-2	-12	21			
-1	-3	9	-12	6	
0	0	3	-6	6	0
1	3	3	0	6	0
2	12	9	6	6	0
3	33	21	12		

We know that (see formula (7.15) or (7.21))

$$f'(x) = \frac{1}{h} \left[\Delta f(x) - \frac{\Delta^2}{2} f(x) + \frac{\Delta^3}{3} f(x) - \frac{\Delta^4}{4} f(x) + \dots \right].$$

Therefore,

$$f'(-3) = \frac{1}{1} \left[21 - \frac{1}{2}(-12) + \frac{1}{3}(6) \right] = 29$$

and

$$f'(0) = \frac{1}{1} \left[3 - \frac{1}{2}(6) + \frac{1}{3}(6) \right] = 2.$$

EXAMPLE 7.13

Find the first and second derivatives of $f(x)$ at $x = 1.5$ using the following data:

$x:$	1.5	2.0	2.5	3.0	3.5	4.0
$f(x):$	3.375	7.000	13.625	24.000	38.875	59.000

Solution. The difference table for the given problem is

x	$f(x)$					
1.5	3.375	3.625				
2.0	7.000	6.625	3	0.750	0	0
2.5	13.625	10.375	3.750	0.750	0	0
3.0	24.000	14.875	4.500	0.750	0	0
3.5	38.875	20.125	5.250			
4.0	59.000					

Since the tabular point $x = 1.5$ lies in the beginning of the table, we use the differentiation formula obtained by differentiating Newton's forward difference formula. Thus,

$$f'(x_0) = \frac{1}{h} \left[\Delta f(x_0) - \frac{1}{2} \Delta^2 f(x_0) + \frac{1}{3} \Delta^3 f(x_0) - \frac{1}{4} \Delta^4 f(x_0) + \dots \right].$$

Here $h = 0.5$. Therefore, we have

$$f'(1.5) = \frac{1}{0.5} [3.625 - 1.5 + 0.250] = 4.750.$$

For the second derivative, we have

$$f''(x_0) = \frac{1}{h^2} \left[\Delta^2 f(x_0) - \Delta^3 f(x_0) + \frac{11}{12} \Delta^4 f(x_0) \right],$$

which implies

$$f''(1.5) = \frac{1}{0.25} [3 - 0.750 + 0] = 9.$$

EXAMPLE 7.14

Find $f'(10)$ from the following table:

$x:$	3	5	11	27	34
$f(x):$	-13	23	899	17315	35606

Solution. Since the spacing is unequal, we use differentiation formula derived from Newton's divided difference formula. The formula is (see Section 7.9, expression 7.35)

$$f'(x) \approx P'(x) = a_1 + a_2 [(x - x_0) + (x - x_1)], \quad (1)$$

where

$$a_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

and

$$a_2 = \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0}.$$

In the given data, we have

$$x_0 = 5, x_1 = 11, x_2 = 27, x = 10.$$

Therefore,

$$\begin{aligned} a_1 &= \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{899 - 23}{11 - 5} = 146, \\ a_2 &= \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0} = \frac{\frac{17315 - 899}{16} - \frac{899 - 23}{6}}{22} \\ &= \frac{1026 - 146}{22} = 40. \end{aligned}$$

Therefore equation (1) yields

$$f'(10) = 146 + 40[(10 - 5) + (10 - 11)] = 306.$$

EXERCISES

- Approximate the derivative of $f(x) = \cos x$ at $x = 0.1$ by central formula of order $O(h^2)$.
Ans. -0.93050
- The data given below give the distance covered by a body at a specified period. Calculate the velocity of the body at 0.3 second using Stirling's formula

$t:$	0	0.1	0.2	0.3	0.4	0.5	0.6
$x:$	30.13	31.62	32.87	33.64	33.95	33.81	33.24

Ans. -5.33 units

- Find the value of $\frac{dy}{dx}$ at $x = 2.03$ using the following table:

$x:$	1.96	1.98	2.0	2.02	2.04
$y:$	0.7825	0.7739	0.7651	0.7563	0.7473

Ans. -0.06

- Find $f'(x)$ at $x = 1.5$ using the following table:

$x:$	1.5	2.0	2.5	3.0	3.5	4.0
$f(x):$	3.375	7.000	13.625	24.000	38.875	59.000

Hint: Taking $x_0 = 1.5$, use $= f'_0 = \frac{1}{h} \left[\Delta f_0 - \frac{1}{2} \Delta^2 f_0 + \frac{1}{3} \Delta^3 f_0 + \dots \right]$.

Ans. 4.75

5. Find the value of $\cos 1.74$ using the table given below:

$x:$	1.70	1.74	1.78	1.82	1.86
$\sin x:$	0.9916	0.9857	0.9781	0.9691	0.9584

Ans. 0.175

6. Using the table given below, determine the value of x for which y is maximum. Also find the maximum value of y .

$x:$	1.2	1.3	1.4	1.5	1.6
$y:$	0.9320	0.9636	0.9855	0.9975	0.9996

Hint: Use Everett's formula, put the derivative f'_p equal to zero and find p . Then $x_p = x_0 + ph$. Everett's formula then gives the maximum value.

Ans. $x = 1.58$, $y \approx 1.00$

7. Using the table given below, find the value of x for which y is maximum:

$x:$	3	4	5	6	7	8
$y:$	0.205	0.240	0.259	0.262	0.250	0.224

Ans. $x = 5.6875$

8. Using Bessel's formula and the table given below, find $f'(0.04)$:

$x:$	0.01	0.02	0.03	0.04	0.05	0.06
$f(x):$	0.1023	0.1047	0.1071	0.1096	0.1122	0.1148

Ans. $f'(0.04) = 0.25625$

9. For the values of x and y given below, find $f'(4)$:

$x:$	1	2	4	8	1	0
$y:$	0	1	5	21	27	

Ans. 2.833

10. Find the maximum value of $f(x)$ using the table given below:

$x:$	-1	1	2	3
$f(x):$	7	5	19	51

Hint: Arguments not equispaced, so use Newton's divided difference formula. Polynomial is $x^3 + 3x^2 - 2x + 3$, maximum value is at $x = 0.291$ and it is 2.6967.

8 Numerical Quadrature

Numerical integration is the process of computing the approximate value of a definite integral using a set of numerical values of the integrand. If the integrand is a function of single variable, the process is called mechanical quadrature. If the integrand is a function of two independent variables, the process of computing double integral is called mechanical cubature.

Like numerical differentiation, the numerical integration is performed by representing the integrand by an interpolation formula and then integrating the interpolation formula between the given limits. Thus, to find $\int_a^b f(x) dx$, we replace the function f by an interpolation formula involving differences and then integrate this formula between the limits a and b .

8.1 GENERAL QUADRATURE FORMULA

In equidistant interpolation formulae, the relation between x and p is

$$x = x_0 + ph, \quad (8.1)$$

where h is the equidistance between the given nodes. Then

$$x = h p. \quad (8.2)$$

We integrate Newton's forward difference formula over n equidistant intervals of width h . Let the limit of integration for x be x_0 and $x_0 + nh$. Then equation (8.1) yields the corresponding limits of p as 0 and n . Therefore, integration of Newton's forward difference formula

$$f(x) = t_0 + p\Delta f_0 + \frac{(p-1)}{2!} \Delta^2 f_0 + \frac{p(p-1)(p-2)}{3!} \Delta^3 f_0 + \frac{p(p-1)(p-2)(p-3)}{4!} \Delta^4 f_0 + \dots$$

yields

$$\begin{aligned} \int_{x_0}^{x_0+nh} f(x) dx &= h \int_0^n [t_0 + p\Delta f_0 + \binom{p}{2} \Delta^2 f_0 + \binom{p}{3} \Delta^3 f_0 + \binom{p}{4} \Delta^4 f_0 + \dots] dp \\ &= h \left[n f_0 + \frac{n^2}{2} \Delta f_0 + \left(\frac{n^3}{3} - \frac{n^2}{2} \right) \frac{\Delta^2 f_0}{2} + \left(\frac{n^4}{4} - n^3 + n^2 \right) \frac{\Delta^3 f_0}{3!} + \dots \right]. \end{aligned} \quad (8.3)$$

From this general formula, we obtain the distinct quadrature formulae by putting $n = 1, 2, 3, \dots$

(A) Trapezoidal Rule: Setting $n = 1$ in the general formula (8.3), we get the differences $\Delta^2, \Delta^3, \dots$ to be zero and therefore for the interval $[x_0, x_1]$, we have

$$\int_{x_0}^{x_1} f(x) dx = h \left[t_0 + \frac{1}{2} \Delta f_0 \right] = h \left[f_0 + \frac{1}{2} (f_1 - f_0) \right] = \frac{h}{2} (f_0 + f_1),$$

which is called trapezoidal rule.

For the next intervals $[x_1, x_2], [x_2, x_3], \dots [x_{n-1}, x_n]$, we have

$$\int_{x_1}^{x_2} f(x)dx = \frac{h}{2}(f_1 + f_2)$$

$$\dots \quad \dots \quad \dots$$

$$\dots \quad \dots \quad \dots$$

$$\int_{x_{n-1}}^{x_n} f(x)dx = \frac{h}{2}(f_{n-1} + f_n).$$

Adding all these expressions, we get

$$\int_{x_0}^{x_n} f(x)dx = \frac{h}{2}[f_0 + 2(f_1 + f_2 + \dots + f_{n-1}) + f_n],$$

which is known as the composite trapezoidal rule.

(B) Simpson's one-third Rule: Setting $n = 2$ in the general formula (8.3), the differences $\Delta^3, \Delta^4, \dots$ are all zero. The interval of integration is from x_0 to $x_0 + 2h$ and the functional values available to us are f_0, f_1 , and f_2 . Thus we have, from general formula (8.3),

$$\begin{aligned} \int_{x_0}^{x_0+2h} f(x)dx &= h \left[2f_0 + 2\Delta_0 + \left(\frac{8}{3} - 2 \right) \frac{\Delta^2 f_0}{2} \right] \\ &= h \left[2f_0 + 2(f_1 - f_0) + \frac{1}{3}(f_2 - 2f_1 + f_0) \right] \\ &= \frac{h}{3}[f_0 + 4f_1 + f_2], \end{aligned}$$

which is known as Simpson's one-third rule.

Similarly,

$$\begin{aligned} \int_{x_2}^{x_4} f(x)dx &= \frac{h}{3}[f_2 + 4f_3 + f_4] \\ \int_{x_4}^{x_6} f(x)dx &= \frac{h}{3}[f_4 + 4f_5 + f_6] \\ \dots &\quad \dots \quad \dots \quad \dots \\ \dots &\quad \dots \quad \dots \quad \dots \\ \int_{x_{n-2}}^{x_n} f(x)dx &= \frac{h}{3}[f_{n-2} + 4f_{n-1} + f_n]. \end{aligned}$$

Thus for even n , adding the above expressions gives

$$\int_{x_0}^{x_0+nh} f(x)dx = \frac{h}{3}[(f_0 + f_n) + 4(f_1 + f_3 + \dots + f_{n-1}) + 2(f_2 + f_4 + \dots + f_{n-2})],$$

which is known as composite Simpson's rule or parabolic rule and is probably the most useful formula for mechanical quadrature. Obviously, to use this formula, we divide the interval of integration into an even

number of subintervals of width h . The geometric significance of Simpson's rule is that we replace the graph of the given function by $\frac{n}{2}$ arcs of the second degree polynomials or parabolas with vertical axis.

(C) Simpson's three-eight Rule: If we put $n = 3$ in the general formula (8.3), then the values available are f_0, f_1, f_2, f_3 and so the differences $\Delta^4, \Delta^5, \dots$ are all zero. Then we shall obtain

$$\begin{aligned} \int_{x_0}^{x_3} f(x)dx &= \int_{x_0}^{x_0+3h} f(x)dx = \frac{3h}{8} [f_0 + 3f_1 + 3f_2 + f_3] \\ \int_{x_3}^{x_6} f(x)dx &= \frac{3h}{8} [f_3 + 3f_4 + 3f_5 + f_6], \\ \dots &\quad \dots \quad \dots \quad \dots \quad \dots \\ \dots &\quad \dots \quad \dots \quad \dots \quad \dots \\ \int_{x_{n-3}}^{x_n} f(x)dx &= \frac{3h}{8} [f_{n-3} + 3f_{n-2} + 3f_{n-1} + f_n]. \end{aligned}$$

Thus, if n is a multiple of 3, then adding the above expressions, we get

$$\int_{x_0}^{x_0+nh} f(x)dx = \frac{3h}{8} [(f_0 + f_n) + 3(f_1 + f_2 + f_4 + f_5 + \dots + f_{n-1}) + 2(f_3 + f_6 + \dots + f_{n-3})],$$

which is called Simpson's three-eight rule. Thus, in this method, we divide the interval of integration into multiple of 3 subintervals.

(D) Boole's Rule: If $n = 4$, the available values of f are f_0, f_1, f_2, f_3, f_4 and therefore $\Delta^5, \Delta^6, \dots$ are zero. So, putting $n = 4$ in the general quadrature formula, we get

$$\begin{aligned} \int_{x_0}^{x_0+4h} f(x)dx &= \int_{x_0}^{x_4} f(x)dx \\ &= h \left[4f_0 + 8\Delta f_0 + \frac{20}{3}\Delta^2 f_0 + \frac{8}{3}\Delta^3 f_0 + \frac{28}{90}\Delta^4 f_0 \right] \\ &= \frac{2h}{45} [7f_0 + 32f_1 + 12f_2 + 32f_3 + 7f_4], \\ \int_{x_4}^{x_8} f(x)dx &= \frac{2h}{45} [7f_4 + 32f_5 + 12f_6 + 32f_7 + 7f_8] \\ \dots &\quad \dots \quad \dots \\ \dots &\quad \dots \quad \dots \end{aligned}$$

Adding these integrals, we get

$$\int_{x_0}^{x_0+nh} f(x)dx = \frac{2h}{45} [7f_0 + 32f_1 + 12f_2 + 32f_3 + 14f_4 + 32f_5 + 12f_6 + 32f_7 + 14f_8 + \dots],$$

where n is a multiple of 4. This formula is known as Boole's rule.

(E) Weddle's Rule: If $n = 6$, then $\Delta^7, \Delta^8, \dots$ are zero and we have

$$\int_{x_0}^{x_0+6h} f(x)dx = h \left[6f_0 + 18\Delta f_0 + 27\Delta^2 f_0 + 24\Delta^3 f_0 + \frac{123}{10}\Delta^4 f_0 + \frac{33}{10}\Delta^5 f_0 + \frac{41}{140}\Delta^6 f_0 \right].$$

The coefficient of $\Delta^6 f_0$ differs from $\frac{3}{10}$ by a small fraction $\frac{1}{140}$. Therefore, if we replace this coefficient by $\frac{3}{10}$, we commit an error of only $\frac{h}{140} \Delta^6 f_0$. For small values of h , this error is negligible. Making this change, we get

$$\int_{x_0}^{x_0+6h} f(x)dx = \frac{3h}{10} [t_0 + 5f_1 + f_2 + 6f_3 + t_4 + 5f_5 + f_6].$$

Similarly,

$$\begin{aligned} \int_{x_0+6h}^{x_0+12h} f(x)dx &= \frac{3h}{10} [t_6 + 5f_7 + f_8 + 6f_9 + f_{10} + 5f_{11} + f_{12}] \\ &\quad \dots \quad \dots \quad \dots \\ &\quad \dots \quad \dots \quad \dots \end{aligned}$$

So, if n is a multiple of 6, adding all such above expressions, we get

$$\begin{aligned} \int_{x_0}^{x_0+nh} f(x)dx &= \frac{3h}{10} [t_0 + 5f_1 + f_2 + 6f_3 + t_4 + 5f_5 + 2f_6 + 5f_7 + \dots + 5t_{n-1} + f_n] \\ &= \frac{3h}{10} \sum_{i=0}^n Kf_i, \end{aligned}$$

where

$$K = 1, 5, 1, 6, 1, 5, 2, 5, 16, 15, 2 \text{ etc.}$$

This formula is known as Weddle's rule. It is more accurate, in general, than Simpson's rule but requires at least seven consecutive values of the function. The geometric meaning of Weddle's rule is that we replace the graph of the given function by $\frac{n}{6}$ arcs of sixth degree polynomials.

If we integrate Newton's backward difference formula

$${}_{p-} = t_0 + {}_r \nabla f_0 + \frac{p(p+1)}{2!} \nabla^2 f_0 + \frac{p(p+1)(p+2)}{3!} \nabla^3 f_0 + \dots,$$

then we get

$$\begin{aligned} \int_{x_0}^{x_1} f(x)dx &= \int_{x_0}^{x_0+h} f(x)dx \\ &= h [{}_{-1} + \frac{1}{2} \nabla f_0 + \frac{5}{12} \nabla^2 f_0 + \frac{3}{8} \nabla^3 f_0 + \frac{251}{720} \nabla^4 f_0 + \dots]. \end{aligned} \tag{8.4}$$

If we multiply the right-hand side of equation (8.4) by the identity operator $(I - \nabla)E$, we get

$$\int_{x_0}^{x_1} f(x)dx = h [{}_{-1} - \frac{1}{2} \nabla f_1 - \frac{1}{12} \nabla^2 f_1 - \frac{1}{24} \nabla^3 f_1 - \frac{19}{720} \nabla^4 f_1 - \dots] \tag{8.5}$$

The above two formulae are used for the numerical solution of differential equations. Formula (8.4) is an extrapolation formula because it uses the ordinates at $x_0, x_{-1}, x_{-2}, \dots$ to find the integral up to x_1 . For this reason, it is called a predictor, whereas formula (8.5) is called corrector and is more accurate as its coefficients are smaller, which make it more rapidly convergent than the predictor.

We now integrate Bessel's formula

$$\begin{aligned} p = t_0 + p\delta_{1/2} + \frac{p(p-1)}{2!} \left(\frac{\delta^2 f_0 + \delta^2 f_1}{2} \right) + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{1/2} \\ + \frac{(p+1)p(p-1)(p-2)}{4!} \left(\frac{\delta^4 f_0 + \delta^4 f_1}{2} \right) + \dots \end{aligned}$$

Integration yields

$$\begin{aligned} \int_{x_0}^{x_1} f = \int_{x_0}^{x_0+h} f(x) dx &= h \int_0^1 \left[p\delta_{1/2} + \frac{p(p-1)}{2!} \left(\frac{\delta^2 f_0 + \delta^2 f_1}{2} \right) + \frac{p \left(p - \frac{1}{2} \right) (p-1)}{3!} \delta^3 f_{1/2} \right. \\ &\quad \left. + \frac{(p+1)p(p-1)(p-2)}{4!} \left(\frac{\delta^4 f_0 + \delta^4 f_1}{2} \right) + \dots \right] dp \\ &= h \left[pf_0 + \frac{p^2}{2} \delta f_{1/2} + \frac{\frac{p^2}{2} - \frac{p^2}{2}}{2!} \left(\frac{\delta^2 f_0 + \delta^2 f_1}{2} \right) + \frac{\frac{p^4}{4} - \frac{p^3}{2} + \frac{p^2}{4}}{3!} \delta^3 f_{1/2} \right. \\ &\quad \left. + \frac{\frac{p^5}{5} - \frac{p^4}{2} - \frac{p^3}{3} + p^2}{4!} \left(\frac{\delta^4 f_0 + \delta^4 f_1}{2} \right) + \dots \right]_0^1 \\ &= h \left[f_0 + \frac{f_1 - f_0}{2} - \frac{1}{12} \left(\frac{\delta^2 f_0 + \delta^2 f_1}{2} \right) + \frac{11}{720} \left(\frac{\delta^4 f_0 + \delta^4 f_1}{2} \right) + \dots \right] \\ &= h \left[\mu J_{1/2} - \frac{1}{2} \delta^2 f_{1/2} + \frac{11}{720} \mu \delta^4 f_{1/2} - \frac{191}{60480} \mu \delta^6 f_{1/2} + \dots \right]. \end{aligned}$$

This formula is much more powerful than the integration formula using forward or backward differences, but it cannot be used at the two ends of a table.

8.2 COTE'S FORMULAE

Let the function values of a function f be available at equidistant points $x_0, x_1, x_2, \dots, x_n$, where $x_n = x_0 + nh$. We replace $f(x)$ by a suitable function

$$P(x) = \sum_{k=0}^n L_k(x) f_k,$$

where

$$L_k(x) = \frac{(x - x_0)(x - x_1) \dots (x - x_{k-1})(x - x_{k+1}) \dots (x - x_n)}{(x_k - x_0)(x_k - x_1) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_n)},$$

and f_k is the function value at x_k . But $x_k = x_0 + kh$. Also, we have $x = x_0 + ph$, and so $dx = hdp$. Since

$$x - x_0 = ph, (-\bar{x}_1) = x - (x_0 + h) = (x - x_0) - h = ph - n = (-1)h, \dots,$$

and

$$x_k - x_0 = kh, \dots, x_1 = (k-1)h, \dots,$$

we get,

$$L_k(x) = \frac{p(p-1)\dots(p-k+1)(p-k-1)\dots(p-n)}{k(k-1)\dots(1)(-1)\dots(k-n)}.$$

Therefore,

$$\begin{aligned} \int_{x_0}^{x_n} P(x) dx &= h \int_0^n \left[\sum_{k=0}^n L_k(x) \right] dp \\ &= nh \left(\frac{1}{n} \sum_{k=0}^n f_k \int_0^n L_k dp \right) \\ &= nh \sum_{k=0}^n C_k^n f_k, \end{aligned}$$

where

$$C_k^n = \frac{1}{n} \int_{-k}^n dp \quad (0 \leq k \leq n).$$

are called Cote's numbers. It can be seen that

$$C_k^n = C_{n-k}^n \text{ and } \sum_{k=0}^n C_k^n = 1.$$

Case I. Let $n = 1$. Then

$$\int_{x_0}^{x_1} P(x) dx = n \sum_{k=0}^1 C_k^1 f_k = h[C_0^1 f_0 + C_1^1 f_1]$$

and we have

$$\begin{aligned} C_0^1 &= \frac{1}{1} \int_0^1 \frac{(p-1)}{-1} dp = \left[\frac{-p^2}{2} + p \right]_0^1 = \frac{1}{2}, \\ C_1^1 &= \frac{1}{1} \int_0^1 p dp = \left[\frac{p^2}{2} \right]_0^1 = \frac{1}{2}. \end{aligned}$$

Therefore,

$$\begin{aligned} \int_{x_0}^{x_1} f(x) dx &\approx \int_0^1 P(x) dx = h \left[\frac{f_0 + f_1}{2} \right] \\ &= \frac{h}{2} (f_0 + f_1) \end{aligned}$$

which is nothing but trapezoidal rule.

Case II. Setting $n = 2$, we get

$$\begin{aligned} C_0^2 &= \frac{1}{2} \int_0^2 \frac{(p-1)(p-2)}{(-1)(-2)} dp = \frac{1}{6} \\ C_1^2 &= \frac{1}{2} \int_0^2 \frac{p(p-2)}{1(-1)} dp = \frac{4}{6} \\ C_2^2 &= \frac{1}{2} \int_0^2 \frac{p(p-1)}{(2)(1)} dp = \frac{1}{6}. \end{aligned}$$

Hence,

$$\int_{x_0}^{x_2} f(x) dx \approx nh \left[\frac{1}{6} f_0 + \frac{4}{6} f_1 + \frac{1}{6} f_2 \right] = \frac{h}{3} [f_0 + 4 f_1 + f_2],$$

which is Simpson's formula.

Case III. Setting $n = 3$, we obtain

$$\begin{aligned} C_0^3 &= \frac{1}{3} \int_0^3 \frac{(p-1)(p-2)(p-3)}{(-1)(-2)(-3)} dp = -\frac{1}{18} \left[\frac{p^4}{4} - \frac{6p^3}{3} + 11\frac{p^2}{2} - 6p \right]_0^3 = \frac{1}{8} \\ C_1^3 &= \frac{1}{3} \int_0^3 \frac{p(p-2)(p-3)}{1(-1)(-2)} dp = \frac{1}{6} \left[\frac{p^4}{4} - \frac{5p^3}{3} + 6\frac{p^2}{2} \right]_0^3 = \frac{3}{8} \\ C_2^3 &= \frac{1}{3} \int_0^3 \frac{p(p-1)(p-3)}{2(1)(-1)} dp = \frac{3}{8} \\ C_3^3 &= \frac{1}{3} \int_0^3 \frac{p(p-1)(p-2)}{3(2)(1)} dp = \frac{1}{8}. \end{aligned}$$

Hence,

$$\int_{x_0}^{x_3} f(x) dx = \frac{3h}{8} [f_0 + 3 f_1 + 3 f_2 + f_3],$$

which is Simpson's three-eight rule or four point formula.

8.3 ERROR TERM IN QUADRATURE FORMULA

We have seen that in any quadrature formula, the function f is approximated by a polynomial of degree n , say. Thus,

$$f(x) \approx P_n(x)$$

and quadrature formula becomes

$$\int_{x_0}^{x_n} f(x) dx \approx \int_{x_0}^{x_n} P_n(x) dx. \quad (8.6)$$

If $R_n(x)$ is the difference between $f(x)$ and $P_n(x)$ at a point belonging to the interval bounded by the extreme points of (x_0, x_1, \dots, x_n) . Then

$$f(x) = P_n(x) + R_n(x).$$

Therefore,

$$\int_{x_0}^{x_n} f(x) dx = \int_{x_0}^{x_n} P_n(x) dx + \int_{x_0}^{x_n} R_n(x) dx.$$

Therefore, the error between the true value of the integral and the value given by the quadrature formula (8.6) is

$$\int_{x_0}^{x_n} R_n(x) dx.$$

But for a polynomial of degree n , we have

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} F(x),$$

where $F(x) = \prod_{i=0}^n (x - x_i)$.

If we consider equispaced ordinates, then $x = x_0 + ph$ and $x_n = x_0 + nh$.
Therefore,

$$\begin{aligned} x - x_0 &= ph \\ x - x_1 &= (x_0 + ph) - (x_0 + h) = (p-1)h \\ x - x_2 &= (x_0 + ph) - (x_0 + 2h) = (p-2)h \\ &\dots && \dots && \dots \\ &\dots && \dots && \dots \\ x - x_n &= (x_0 + ph) - (x_0 + nh) = (p-n)h \end{aligned}$$

and so

$$F(x) = p^{n+1} [p(p-1)(p-2)\dots(p-n)].$$

Putting this value of $F(x)$ in $R_n(x)$, we get

$$R_n(x) = h^{n+1} \frac{f^{(n+1)}(\xi)}{(n+1)!} p^{(n+1)}, \quad \xi \in (x_0, x_n),$$

where

$$p^{(n+1)} = p(p-1)\dots(p-n).$$

Therefore, the error in the quadrature formula is

$$\begin{aligned} E_n(x) &= \int_{x_0}^{x_n} R_n(x) dx \\ &= \frac{h}{(n+1)!} \int_0^n h^{n+1} p^{(n+1)}(\xi) p^{(n+1)} dp \\ &= \frac{h^{n+2}}{(n+1)!} \int_0^n p^{(n+1)}(\xi) dp. \end{aligned}$$

Since ξ is independent of p , this integral cannot be evaluated directly. It can be shown, however, that this integral takes one of the following two forms:

$$\begin{aligned} E_n(x) &= \frac{h^{n+2}}{(n+1)!} f^{(n+1)}(\xi) \int_0^n p^{(n+1)} dp && (n \text{ odd}), \\ E_n(x) &= \frac{h^{n+3}}{(n+1)!} f^{(n+2)}(\xi) \int_0^n \left(-\frac{n}{2} \right) p^{(n+1)} dp && (n \text{ even}). \end{aligned}$$

For example,

(i) In trapezoidal rule, $n = 1$ (odd). Therefore,

$$\begin{aligned} E_1(x) &= \frac{h^3 f''(\xi)}{2!} \int_0^1 p(p-1) dp \\ &= \frac{h^3 f''(\xi)}{2} \left[\frac{p^3}{3} - \frac{p^2}{2} \right]_0^1 = -\frac{h^3 f''(\xi)}{12}, \quad x_0 < \xi < x_1. \end{aligned}$$

Summing over n intervals, we get

$$E_n = -\frac{nh^3 f''(\xi)}{12} = -\frac{h^2}{12} (x_n - x_0) f''(\xi),$$

since $nh = x_n - x_0$. Thus, the error in trapezoidal rule is of order 2.

(ii) In the case of Simpson's formula, $n = 2$ (even). Therefore, the error term is given by

$$\begin{aligned} E_2(x) &= \frac{h^5 f^{(iv)}(\xi)}{4!} \int_0^2 (p-1)p(p-1)(p-2) dp \\ &= -\frac{h^5}{90} f^{(iv)}(\xi), \quad x_0 < \xi < x_2. \end{aligned}$$

Summing up for $\frac{n}{2}$ intervals, we get

$$E_n = -\frac{n}{2} \frac{h^5}{90} f^{(iv)}(\xi) = -\frac{(x_n - x_0)}{180} h^4 f^{(iv)}(\xi), \quad x_0 < \xi < x_n.$$

Thus, the error is of order 4 in case of Simpson's rule. Therefore, the complete Simpson's formula is

$$\int_{x_0}^{x_n} f(x) dx = \frac{h}{3} \left[f_0 + 4(f_1 + \dots + f_{n-1}) + 2(f_2 + f_4 + \dots + f_{n-2}) + f_n \right] - \frac{(x_n - x_0)}{180} h^4 f^{(iv)}(\xi).$$

The value obtained for E_n shows that E_n is zero when $f^{(iv)}(x) = 0$. Hence, when $f(x)$ is a polynomial of the first, second, or third degree, Simpson's rule yields the exact value of $\int_{x_0}^{x_n} f(x) dx$.

Similarly, we can show that error is of order 8 in case of Weddle's rule.

Taylor's Series Method for Finding Error

Trapezoidal Rule: Let f be a finite continuous function in the interval $x = x_0$ to $x = x_0 + h$ and have continuous first and second derivatives in the said interval. Let

$$F(x) = \int_0^x f(t) dt.$$

Then, by fundamental theorem of integral calculus,

$$F'(x) = f(x), F''(x) = f'(x) \dots$$

and so

$$\int_{x_0}^{x_0+h} f(x) dx = F(x_0 + \dots) - F(x_0).$$

On the other hand, by trapezoidal rule,

$$\int_{x_0}^{x_0+h} f(x) dx = \frac{h}{2} [f(x_0) + f(x_0 + h)].$$

Therefore, the error is given by

$$\begin{aligned} E(x) &= F(x_0 + n) - F(x_0) - \frac{h}{2} [f(x_0) + f(x_0 + h)] \\ &= F(x_0) + hf(x_0) + \frac{h^2}{2} f'(x_0) + \frac{h^3}{3!} f''(x_0) + \dots - F(x_0) - \frac{h}{2} \left[f(x_0) + f(x_0) + hf'(x_0) + \frac{h^2}{2!} f''(x_0) \right. \\ &\quad \left. + \frac{h^3}{3!} f'''(x_0) + \dots \right] \\ &= \left[\frac{h^3}{3!} f''(x_0) - \frac{h^3}{4} f''(x_0) \right] - \dots \\ &= -\frac{h^3}{12} f''(x_0) - \dots \end{aligned}$$

Summing over n intervals, we get

$$E_n(x) \leq -\frac{nh^3}{12} f''(x_m),$$

where $f''(x_m)$ denotes the greatest value of $f''(x_0), f''(x_1), \dots, f''(x_{n-1})$. Thus,

$$E_n(x) \leq -\frac{(x_n - x_0)}{12} h^2 f''(x_m),$$

which shows that error is of order 2.

Simpson's Rule: Let f be a finite and continuous function in the interval $x = x_0 - h$ to $x = x_0 + h$ and have continuous derivatives of all orders up to and including the fourth in the said interval. Let

$$F(x) = \int_0^x f(t) dt.$$

Then, by fundamental theorem of integral calculus

$$F'(x) = f(x) \text{ and so } F''(x) = f'(x), F'''(x) = f''(x) \dots$$

Also, by the same theorem,

$$\int_{x_0-h}^{x_0+h} f(x) dx = F(x_0 + h) - F(x_0 - h).$$

But, by Simpson's rule,

$$\int_{x_0-h}^{x_0+h} f(x) dx = \frac{h}{3} [f(x_0 - h) + 4f(x_0) + f(x_0 + h)].$$

Therefore, the error in Simpson's rule is given by

$$E(x) = F(x_0 + h) - F(x_0 - h) - \frac{h}{3} [f(x_0 - h) + 4f(x_0) + f(x_0 + h)].$$

By Taylor's expansion

$$\begin{aligned} F(x_0 + h) &= f(x_0) + hf'(x_0) + \frac{h^2}{2} f''(x_0) + \frac{h^3}{3!} f'''(x_0) + \dots \\ F(x_0 - h) &= f(x_0) - hf'(x_0) + \frac{h^2}{2} f''(x_0) - \frac{h^3}{3!} f'''(x_0) + \dots \\ f(x_0 + h) &= f(x_0) + hf'(x_0) + \frac{h^2}{2} f''(x_0) + \frac{h^3}{3!} f'''(x_0) + \dots \\ f(x_0 - h) &= f(x_0) - hf'(x_0) + \frac{h^2}{2} f''(x_0) - \frac{h^3}{3!} f'''(x_0) + \dots \end{aligned}$$

Substituting these values in the error term, we get

$$\begin{aligned} E(x) &= \left[2hf(x_0) + \frac{2h^3}{3!} f''(x_0) + \frac{2h^5}{5!} f^{(iv)}(x_0) + \dots \right] \\ &\quad - \frac{h}{3} \left[6f(x_0) + h^2 f''(x_0) + \frac{2h^4}{4!} f^{(iv)}(x_0) + \dots \right] \\ &= -\frac{h^5}{90} [f^{(iv)}(x_0) + \dots]. \end{aligned}$$

Summing over $\frac{n}{2}$ intervals, we get

$$E(x) = \frac{h^5}{90} [f^{(iv)}(x_0) + f^{(iv)}(\dots_2) + \dots + f^{(iv)}(x_{n-2})].$$

If $f^{(iv)}(x_n)$ denotes the greatest value of any of the $\frac{n}{2}$ values $f^{(iv)}(x_0), f^{(iv)}(\dots_2), \dots, f^{(iv)}(x_{n-2})$, then

$$E(x) \leq -\frac{nh^5}{180} f^{(iv)}(x_m) = -\frac{(x_n - x_0)}{180} h^4 f^{(iv)}(x_m),$$

since $x_n - x_0 = nh$. Hence, error in Simpson's rule is of order 4.

8.4 RICHARDSON EXTRAPOLATION (OR DEFERRED APPROACH TO THE LIMIT)

Knowing the order of the error, one can get fairly accurate estimate of the true value Q of the approximate values of the derivative or integral as soon as two approximate values Q_1 and Q_2 of Q have been obtained by means of different spacing, say h_1 and h_2 . Thus, if the order of error is n , truncating the error series after its first term, we obtain

$$\left. \begin{aligned} Q - \epsilon'_1 &\approx Ch_1^n \\ Q - \epsilon'_2 &\approx Ch_2^n \end{aligned} \right\}, \quad (8.7)$$

where in differentiation formulae the constant C depends on the pivotal point at which the derivative is evaluated. From equation (8.7), we get

$$\frac{Q - \epsilon_1}{h_1^n} = \frac{Q - \epsilon_2}{h_2^n} \approx C$$

or

$$Q \left(\frac{1}{h_2^n} - \frac{1}{h_1^n} \right) = \frac{Q_2}{h_2^n} - \frac{Q_1}{h_1^n}$$

or

$$Q_{12} \approx \frac{\left(\frac{h_1}{h_2} \right)^n Q_2 - Q_1}{\left(\frac{h_1}{h_2} \right)^n - 1}, \quad (8.8)$$

which is called h^n extrapolation formula of Richardson. This formula gives the approximate value Q_{12} of Q .

We, generally, consider the cases where $\frac{h_1}{h_2} = 2$. In trapezoidal rule, order of error is 2. Therefore, the extrapolation formula with $\frac{h_1}{h_2} = 2$ becomes

$$Q_{12} \approx \frac{2^2 Q_2 - Q_1}{2^2 - 1} = \frac{4}{3} \epsilon'_2 - \frac{1}{3} Q_1,$$

which is called $\frac{1}{3} h^2$ extrapolation formula.

In Simpson's rule, the order of error is 4. Therefore, the extrapolation formula becomes $\frac{1}{15} h^4$ extrapolation formula given by

$$Q_{12} \approx \frac{16}{15} \epsilon'_2 - \frac{1}{15} Q_1 \text{ with error } O(h^4) \text{ and } \frac{h_1}{h_2} = 2.$$

In Boole's rule, the order of error is 6 and so the expression (8.8) yields the following $\frac{1}{63} h^6$ extrapolation formula:

$$Q_{12} \approx \frac{64}{63} \epsilon'_2 - \frac{1}{63} Q_1.$$

EXAMPLE 8.1

Dividing the range into 10 equal parts, apply Simpson's one-third rule to evaluate the integral $\int_0^5 \frac{dx}{4x+5}$ correct to four decimal places. Hence, find the approximate value of $\log_e 5$.

Solution. The values of the integrand for $h = \frac{1}{2}$ are.

$x:$	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{5}{2}$	3	$\frac{7}{2}$	4	$\frac{9}{2}$	5
$f(0):$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{9}$	$\frac{1}{11}$	$\frac{1}{13}$	$\frac{1}{15}$	$\frac{1}{17}$	$\frac{1}{19}$	$\frac{1}{21}$	$\frac{1}{23}$	$\frac{1}{25}$

Therefore, by Simpson's one-third rule,

$$\begin{aligned} \int_0^5 \frac{dx}{4x+5} &= \frac{h}{3}[f_0 + 4(f_1 + f_3 + f_5 + f_7 + f_9) + 2(f_2 + f_4 + f_6 + f_8) + f_{10}] \\ &= \frac{1}{6} \left[\left(\frac{1}{5} + \frac{1}{25} \right) + 4 \left(\frac{1}{7} + \frac{1}{11} + \frac{1}{15} + \frac{1}{19} + \frac{1}{23} \right) + 2 \left(\frac{1}{9} + \frac{1}{13} + \frac{1}{17} + \frac{1}{21} \right) \right] \\ &= \frac{1}{6} \left[\frac{6}{25} + 4(0.142857 + 0.09090 + 0.066666 + 0.05263 + 0.04348) + 2(0.11111 + 0.07692 \right. \\ &\quad \left. + 0.05882 + 0.04761) \right] \\ &= \frac{1}{6} [0.24 + 1.58613 + 0.58892] = 0.4025. \end{aligned}$$

Also,

$$\begin{aligned} \int_0^5 \frac{dx}{4x+5} &= \frac{1}{4} [\log(4x+5)]_0^5 = \frac{1}{4} [\log 25 - \log 5] \\ &= \frac{1}{4} \left[\log \frac{25}{5} \right] = \frac{1}{4} \log_e 5. \end{aligned}$$

Hence,

$$\log_e 5 = 4 \int_0^5 \frac{dx}{4x+5} = 4(0.4025) = 1.61.$$

The actual value is 1.6094.

EXAMPLE 8.2

Calculate $\int_1^2 \frac{dx}{x}$ by

- (i) Simpson's rule with $h = 0.50$,
- (ii) Simpson's rule with $h = 0.25$,
- (iii) Richardson's extrapolation.

Compare the results with exact value.

Solution. (i) With $h = 0.50$, we have

$$f_0 = 1, f_1 = \frac{1}{1.5} = \frac{2}{3}, f_2 = \frac{1}{2}$$

and so by Simpson's rule

$$\int_1^2 \frac{dx}{x} \approx \frac{h}{3} [f_0 + 4f_1 + f_2] = \frac{0.5}{3} \left[1 + 4 \cdot \frac{2}{3} + \frac{1}{2} \right] = \frac{12.5}{18} = 0.69444.$$

(ii) With $h = 0.25$, we have

$$f_0 = 1, f_1 = \frac{1}{1.25}, f_2 = \frac{1}{1.50}, f_3 = \frac{1}{1.75}, f_4 = \frac{1}{2}.$$

Then, by Simpson's rule, we have

$$\begin{aligned} \int_1^2 \frac{dx}{x} &\approx \frac{h}{3} [(f_0 + f_4) + 4(f_1 + f_3) + 2f_2] \\ &= \frac{0.25}{3} [(1 + 0.5) + 4(0.5 + 0.5714) + 1.3333] = 0.69324. \end{aligned}$$

(iii) We have

$$\frac{h_1}{h_2} = \frac{0.50}{0.25} = 2.$$

Also Simpson's rule is of order 4. Therefore,

$$Q_{12} \approx \frac{16}{15}(0.69324) - \frac{1}{15}(0.69444) = 0.69316,$$

which is in good agreement with the exact value

$$\log 2 - \log 1 = \log 2 = 0.69315.$$

8.5 SIMPSON'S FORMULA WITH END CORRECTION

We now improve usual Simpson's formula by allowing derivatives in the endpoints. Let

$$\int_{x_0-h}^{x_0+h} f(x) dx \approx h[af_{-1} + bf_0 + af_1] + h^2[f'_{-1} - f'_1]. \quad (8.9)$$

Let

$$F(x) = \int f(x) dx$$

and so

$$\int_{x_0-h}^{x_0+h} f(x) dx = F(x_0+h) - F(x_0-h).$$

Therefore formula (8.9) becomes

$$F(x_0+h) - F(x_0-h) \approx h[af_{-1} + bf_0 + af_1] + h^2[f'_{-1} - f'_1] \quad (8.10)$$

Expanding by Taylor's Theorem, we get

$$\begin{aligned} F(x_0 + h) - f(x_0) &= \left[F(x_0) + hf'(x_0) + \frac{h^2}{2} f''(x_0) + \frac{h^3}{3!} f'''(x_0) + \dots \right] \\ &= \left[F(x_0) - hf(x_0) + \frac{h^2}{2} f'(x_0) - \frac{h^3}{3!} f''(x_0) + \dots \right] \\ &= 2hf_0 + \frac{2h^3}{3!} f_0''' + \frac{2h^5}{5!} f_0'''' + \dots \end{aligned}$$

Also,

$$\begin{aligned} f_{-1} &= f(x_0 - h) = f(x_0) - hf'_0 + \frac{h^2}{2!} f''_0 - \frac{h^3}{3!} f'''_0 + \frac{h^4}{4!} f^{(iv)}_0 - \dots \\ f_0 &= f(x_0 + h) = f(x_0) + hf'_0 + \frac{h^2}{2!} f''_0 + \frac{h^3}{3!} f'''_0 + \frac{h^4}{4!} f^{(iv)}_0 + \dots \\ f_{-1}' &= f_0' - hf''_0 + \frac{h^2}{2!} f'''_0 - \frac{h^3}{3!} f^{(iv)}_0 + \dots \\ f_1' &= f_0' + hf''_0 + \frac{h^2}{2!} f'''_0 + \frac{h^3}{3!} f^{(iv)}_0 + \dots \end{aligned}$$

Therefore, the right-hand side of equation (8.10) is

$$bhf_0 + ah \left[2f_0 + \frac{2h^2}{2!} f''_0 + \frac{2h^4}{4!} f^{(iv)}_0 + \dots \right] + ch^2 \left[-2hf''_0 - \frac{2h^3}{3!} f^{(iv)}_0 - \dots \right].$$

Comparing the coefficients of f_0 , f_0'' , and $f_0^{(iv)}$ and on both sides, we get

$$\begin{aligned} 2a + b &= 2, \\ a - 2c &= \frac{1}{3}, \\ a - 4c &= \frac{1}{5}, \end{aligned}$$

which yield

$$a = \frac{7}{15}, \quad b = \frac{16}{15}, \quad \text{and} \quad c = \frac{1}{15}.$$

Hence,

$$\int_{x_0-h}^{x_0+h} f(x) dx \approx \frac{h}{15} [7f_{-1} + 16f_0 + 7f_1] + \frac{h^2}{15} [f'_{-1} - f'_1]$$

or, we can write

$$\int_{x_0}^{x_0+2h} f(x) dx \approx \frac{h}{15} [7f_0 + 16f_1 + 7f_2] + \frac{h^2}{15} [f'_0 - f'_2].$$

Similarly,

$$\int_{x_2}^{x_4} f(x)dx \approx \frac{h}{15}[7f_2 + 16f_3 + 7f_4] + \frac{h^2}{15}[f'_2 - f'_4]$$

$$\dots \quad \dots \quad \dots$$

$$\dots \quad \dots \quad \dots$$

$$\int_{x_{n-2}}^{x_n} f(x)dx \approx \frac{h}{15}[7f_{n-2} + 16f_{n-1} + 7f_n] + \frac{h^2}{15}[f'_{n-2} - f'_n]$$

Adding, we get

$$\int_{x_0}^{x_n} f(x)dx \approx \frac{h}{15}[7f_0 + 16f_1 + 14f_2 + 16f_3 + \dots + 7f_n] + \frac{h^2}{15}[f'_0 - f'_n] + O(h^6),$$

where $x_n - x_0 = nh$ and f_0, f_1, \dots, f_n are function values at the points x_0, x_1, \dots, x_n .

8.6 ROMBERG'S METHOD

This method makes use of Richardson's extrapolation in a systematic way. We have seen that for trapezoidal rule, the error is $O(h^2)$ and with the spacing ratio $\frac{h_1}{h_2} = 2$ Richardson extrapolation yields

$$R = \frac{4}{3} R_2 - \frac{1}{3} R_1 = Q_2 + \frac{Q_2 - Q_1}{3}.$$

We observe that R is the same result as obtained by Simpson's rule.

Since error in R is of order 4 (Simpson's rule), taking again $\frac{h_1}{h_2} = 2$, we get

$$S = \frac{16}{15} R_2 - \frac{1}{15} R_1 = R_2 + \frac{R_2 - R_1}{15}.$$

As a matter of fact, S is the same result as obtained by Cote's formula for $n = 4$. Now, the error is of order 6.

So taking $\frac{h_1}{h_2} = 2$, we obtain

$$T = \frac{64}{63} S_2 - \frac{1}{63} S_1 = S_2 + \frac{S_2 - S_1}{63}.$$

The process is repeated till two successive values are sufficiently close to each other.

EXAMPLE 8.3

One wants to construct a quadrature formula of the type

$$\int_0^h f(x) dx = \frac{h}{2}(f_0 + f_1) + ah^2(f'_0 - f'_1) + R.$$

Calculate the constant a and find the order of the remainder term R .

Solution. Let $F(x) = \int f(x)dx$. Then

$$\begin{aligned}
 \int_0^h f(x)dx &= r \cdot h - F(0) \\
 &= f(0) + hF'(0) + \frac{h^2}{2} F''(0) + \frac{h^3}{3!} F'''(0) + \frac{h^4}{4!} F^{(iv)}(0) + \frac{h^5}{5!} F^{(v)}(0) + \dots - F(0) \\
 &= hf'(0) + \frac{h^2}{2} f''(0) + \frac{h^3}{3!} f'''(0) + \frac{h^4}{4!} f^{(iv)}(0) + \frac{h^5}{5!} f^{(v)}(0) + \dots \\
 &= hf_0 + \frac{h^2}{2} f'_0 + \frac{h^3}{3!} f''_0 + \frac{h^4}{4!} f'''_0 + \frac{h^5}{5!} f^{(iv)}_0 + \dots
 \end{aligned} \tag{8.11}$$

Also,

$$\begin{aligned}
 f_1 &= t_0 + hf'_0 + \frac{h^2}{2!} f''_0 + \frac{h^3}{3!} f'''_0 + \frac{h^4}{4!} f^{(iv)}_0 + \frac{h^5}{5!} f^{(v)}_0 + \dots \\
 f'_1 &= f'_0 + hf''_0 + \frac{h^2}{2!} f'''_0 + \frac{h^3}{3!} f^{(iv)}_0 + \frac{h^4}{4!} f^{(v)}_0 + \frac{h^5}{5!} f^{(vi)}_0 + \dots
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \frac{h}{2}(f_0 + f_1) + ah^2(f'_0 - f'_1) &= \frac{h}{2} \left[2t_0 + f'_0 + \frac{h^2}{2!} f''_0 + \frac{h^3}{3!} f'''_0 + \frac{h^4}{4!} f^{(iv)}_0 + \frac{h^5}{5!} f^{(v)}_0 + \dots \right] \\
 &\quad + ah^2 \left[-hf''_0 - \frac{h^2}{2} f'''_0 - \frac{h^3}{3!} f^{(iv)}_0 - \frac{h^4}{4!} f^{(v)}_0 - \frac{h^5}{5!} f^{(vi)}_0 - \dots \right]
 \end{aligned} \tag{8.12}$$

Comparing the coefficients of f''_0 in equations (8.11) and (8.12), we obtain

$$\frac{h^3}{3!} = \frac{1}{4} - ah^3$$

and so $a = \frac{1}{12}$. Hence, the formula is

$$\int_0^h f(x)dx = hf_0 + \frac{h^2}{2} f'_0 + \frac{h^3}{3!} f''_0 + \frac{h^4}{4!} f'''_0 + \frac{h^5}{5!} f^{(iv)}_0 + \dots,$$

which clearly shows that R is of order h^5 .

EXAMPLE 8.4

If $f(x) = a + bx + cx^2$, find the quadrature formulae of the form

$$\int_0^1 f(x)dx \approx A_1 f(-1) + A_2 f(1) + A_3 f(2) \tag{8.13}$$

and

$$\int_0^1 f(x)dx \approx B_1 f(0) + B_2 f(1) + C_2 f(2). \tag{8.14}$$

By finding out the truncation error in both the cases, point out which of the two formulae is more accurate.

Solution. The

$$\text{L.H.S. of equations (8.13) and (8.14)} = \int_0^1 (a + x + cx^2) dx = a + \frac{b}{2} + \frac{c}{3}.$$

Now,

$$\text{R.H.S. of equation (8.13)} = A(a - b + c) + B_1(a + \dots + c) + \dots (a + 2\dots + 4c).$$

Comparing coefficients on both sides, we get

$$\begin{aligned} A_1 + B_1 + C_1 &= 1, \\ A_1 + B_1 + 2C_1 &= \frac{1}{2}, \\ 3A_1 + 3B_1 + 12C_1 &= 1. \end{aligned}$$

Solving these equations, we get

$$A_1 = \frac{5}{36}, B_1 = \frac{13}{12}, \text{ and } C_1 = -\frac{2}{9}.$$

Therefore, the first formula is

$$\int_0^1 f(x) dx = \frac{5}{36} f(-1) + \frac{13}{12} f(1) - \frac{2}{9} f(2) + R_1 \quad (8.15)$$

Further,

$$\text{R.H.S. of equation (8.14)} = A_2(a) + B_2(a + b + c) + C_2(a + 2b + 4c).$$

Comparing coefficient on both sides, we get

$$\begin{aligned} A_2 + B_2 + C_2 &= 1, \\ B_2 + 2C_2 &= \frac{1}{2}, \\ B_2 + 4C_2 &= \frac{1}{3}. \end{aligned}$$

Solving these equations, we get

$$A_2 = \frac{5}{12}, B_2 = \frac{2}{3}, \text{ and } C_2 = -\frac{1}{12}.$$

Therefore, the second formula is

$$\int_0^1 f(x) dx = \frac{5}{12} f(0) + \frac{2}{3} f(1) - \frac{1}{12} f(2) + R_2. \quad (8.16)$$

To find the error, we put $F(x) = \int f(x) dx$. Then the left-hand side of equation (8.15) is equal to

$$\begin{aligned} F(1) - F(0) &= F(0) + F'(0) + \frac{1}{2} F''(0) + \frac{1}{3!} F'''(0) + \frac{1}{4!} F^{(iv)}(0) + \frac{1}{5!} F^{(v)}(0) - F(0) \\ &= f(0) + \frac{1}{2} f'(0) + \frac{1}{3!} f''(0) + \frac{1}{4!} f'''(0) + \frac{1}{5!} f^{(v)}(0) + \dots \end{aligned} \quad (8.17)$$

Also,

$$\begin{aligned}f(1) &= f(0) + f'(0) + \frac{1}{2!} f''(0) + \frac{1}{3!} f'''(0) + \frac{1}{4!} f^{(iv)}(0) + \frac{1}{5!} f^{(v)}(0) + \dots \\f(-1) &= f(0) - f'(0) + \frac{1}{2!} f''(0) - \frac{1}{3!} f'''(0) + \frac{1}{4!} f^{(iv)}(0) - \frac{1}{5!} f^{(v)}(0) + \dots \\f(2) &= f(0) + 2f'(0) + 2f''(0) + \frac{8}{3!} f'''(0) + \frac{16}{4!} f^{(iv)}(0) + \frac{32}{5!} f^{(v)}(0) + \dots\end{aligned}$$

Putting these values in equation (8.15), the right-hand side becomes $f(0) + \frac{1}{2} f'(0) + \frac{1}{3!} f''(0) - \frac{30}{216} f'''(0) - \dots$. Therefore, the order of the remainder term is $Cf'''(0)$, where

$$C = \frac{1}{24} + \frac{30}{216} = \frac{13}{72}.$$

Similarly, the right-hand side of equation (8.16) is equal to

$$f(0) + \frac{1}{2} f'(0) + \frac{1}{6} f''(0) + 0 + \left(-\frac{1}{36} \right) f^{(iv)}(0) + \dots$$

Therefore, the order of the remainder is $\frac{1}{24} f'''(0)$. Comparing the two errors, we observe that the second formula is better.

EXAMPLE 8.5

Using method of undetermined coefficients, derive Simpson's one-third rule.

Solution. Let

$$\int_{x_0-h}^{x_0+h} f(x) dx = af_{-1} + cf_0 + bf_1, \quad (8.18)$$

where the coefficients a , b , c are to be determined. Put $F(x) = \int f(x) dx$. Then the left-hand side of equation (8.18) is equal to

$$\begin{aligned}F(x_0+h) - F(x_0-h) &= F(x_0) + hF'(x_0) + \frac{h^2}{2!} F''(x_0) - \frac{h^3}{3!} F'''(x_0) + \frac{h^4}{4!} F^{(iv)}(x_0) + \dots \\&\quad - \{F(x_0) - hF'(x_0) + \frac{h^2}{2!} F''(x_0) - \frac{h^3}{3!} F'''(x_0) + \frac{h^4}{4!} F^{(iv)}(x_0) + \dots\} \\&= 2hF'(x_0) + 2\frac{h^3}{3!} F'''(x_0) + 2\frac{h^5}{5!} F^{(v)}(x_0) + \dots \\&= 2hf(x_0) + \frac{2h^3}{6} f''(x_0) + \frac{2h^5}{5!} f^{(iv)}(x_0).\end{aligned}$$

Also,

$$\begin{aligned}f_{-1} &= f(x_0-h) = f(x_0) - hf'(x_0) + \frac{h^2}{2!} f''(x_0) - \frac{h^3}{3!} f'''(x_0) + \frac{h^4}{4!} f^{(iv)}(x_0) - \frac{h^5}{5!} f^{(v)}(x_0) + \dots \\f_1 &= f(x_0+h) = f(x_0) + hf'(x_0) + \frac{h^2}{2!} f''(x_0) + \frac{h^3}{3!} f'''(x_0) + \frac{h^4}{4!} f^{(iv)}(x_0) + \frac{h^5}{5!} f^{(v)}(x_0) + \dots\end{aligned}$$

Therefore the right-hand side of equation (8.18) is equal to

$$\begin{aligned} & a[f(x_0) - \frac{h^2}{2!} f''(x_0) + \frac{h^3}{3!} f'''(x_0) - \frac{h^4}{4!} f^{(iv)}(x_0) + \frac{h^5}{5!} f^{(v)}(x_0) + \dots] \\ & + b f(x_0) + c [f(x_0) + h f'(x_0) + \frac{h^2}{2!} f''(x_0) + \frac{h^3}{3!} f'''(x_0) + \dots] \end{aligned}$$

Comparing coefficients of $f(x_0)$, $f'(x_0)$, $f''(x_0)$, $f^{(iv)}(x_0)$, we get

$$2h = a + b + c, \quad 0 = -ah + ch, \quad \frac{2h^3}{6} = \frac{ah^2}{2} + \frac{ch^2}{2},$$

which yield $c = a = \frac{h}{3}$ and $b = \frac{4h}{3}$. Hence,

$$\int_{x_0-h}^{x_0+h} f(x) dx = \frac{h}{3} f_{-1} + \frac{4h}{3} f_0 + \frac{h}{3} f_1 = \frac{h}{3} [f_{-1} + 4f_0 + f_1],$$

which is Simpson's one-third rule.

EXAMPLE 8.6

The integral equation

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

where f is a given function, can be solved by forming a sequence of functions y_0, y_1, y_2, \dots according to

$$y_{n+1}(x) = 1 + \int_0^x f(t)y_n(t)dt.$$

Find the first five functions for $x = 0, 0.25, 0.50$, when $f(x)$ is given in the table below. Start with $y_0 = 1$ and use Bessel's interpolation formula and Simpson's rule.

x	0	0.25	0.50	0.75	1
f	0.5000	0.4794	0.4594	0.4398	0.4207

Solution. Bessel's quadrature formula reads

$$\int_{x_0}^{x_1} f(x)dx = h \left[\frac{f_0 + f_1}{2} - \frac{1}{12} \left(\frac{\delta^2 f_0 + \delta^2 f_1}{2} \right) + \frac{11}{720} \left(\frac{\delta^4 f_0 + \delta^4 f_1}{2} \right) + \dots \right]$$

and Simpson's formula reads

$$\int_0^{x_1} f(x) dx \approx \frac{h}{3} (f_0 + 4f_1 + f_2).$$

We have

$$y_{n+1}(x) = 1 + \int_0^x f(t)y_n(t)dt$$

and we start with $y_0 = 1$. For $x = 0$, we have clearly

$$y_1 = 1, \quad y_2 = \dots, \quad y_3 = 1, \quad y_4 = \dots, \quad y_5 = 1.$$

Now for $x = 0.25$, we use Bessel's formula and get

$$y_1 = 1 + \int_0^{0.25} f(t) v_0(t) dt = 1 + h \left[\frac{f_0 + f_1}{2} \right],$$

since $y_0(t) = 1$ and higher-order differences contribution is very small. Thus,

$$\begin{aligned} y_1 &= 1 + 0.25 \left[\frac{0.5 + 0.4794}{2} \right] = 0.1224, \\ y_2 &= 1 + \int_0^{0.25} f(t) y_1(t) dt = 1 + h \left[\frac{f_0(t)y_1(0) + f_1(t)y_1(0.25)}{2} \right] \\ &= 1 + 0.25 \left[\frac{0.5 + 1.1224(0.4794)}{2} \right] = 1.1298, \\ y_3 &= 1 + \int_0^{0.25} f(t) v_2(t) dt = 1 + h \left[\frac{f_0(t)v_2(0) + f_1(t)v_2(0.25)}{2} \right] \\ &= 1 + 0.25 \left[\frac{0.5 + 1.1298(0.4794)}{2} \right] = 1.1302, \\ y_4 &= 1 + \int_0^{0.25} f(t) y_3(t) dt = 1 + h \left[\frac{f_0(t)y_3(0) + f_1(t)v_3(0.25)}{2} \right] \\ &= 1 + 0.25 \left[\frac{0.5 + 1.1302(0.4794)}{2} \right] = 1.1302 \\ y_5 &= 1 + \int_0^{0.25} f(t) y_4(t) dt = 1 + h \left[\frac{f_0(t)y_4(0) + f_1(t)y_4(0.25)}{2} \right], \\ &= 1 + 0.25 \left[\frac{0.5 + 1.1302(0.4794)}{2} \right] = 1.1302. \end{aligned}$$

Now for $x = 0.50, h = 0.25$, we use Simpson's rule

$$\begin{aligned} y_1 &= 1 + \int_0^{0.50} f(t) v_0(t) dt = 1 + \frac{0.25}{3} [0.5 + 4(0.4794) + 0.4594] = 1.2398, \\ y_2 &= 1 + \int_0^{0.50} f(t) v_1(t) dt = 1 + \frac{h}{3} [v_0(t)y_1(0) + 4f_1(t)y_1(0.25) + v_2(t)y_1(0.50)] \\ &= 1 + \frac{0.25}{3} [0.5(1) + 4(0.4794)(1.1224) + 0.4594(1.2398)] = 1.2685, \\ y_3 &= 1 + \int_0^{0.50} f(t) v_2(t) dt = 1 + \frac{h}{3} [v_0(t)v_2(0) + 4f_1(t)y_2(0.25) + v_2(t)y_2(0.50)] \\ &= 1 + \frac{0.25}{3} [0.5(1) + 4(0.4794)(1.1298) + 0.4594(1.2685)] = 1.2708, \end{aligned}$$

$$\begin{aligned}
 y_4 &= 1 + \int_0^{0.50} f(t) v_3(t) dt = 1 + \frac{h}{3} [f_0(t) v_3(0) + 4f_1(t) v_3(0.25) + f_2(t) v_3(0.50)] \\
 &= 1 + \frac{0.25}{3} [0.5(1) + 4(0.4794)(1.1302) + 0.4594(1.2708)] = 1.2709, \\
 y_5 &= 1 + \int_0^{0.50} f(t) v_4(t) dt = 1 + \frac{h}{3} [f_0(t) y_4(0) + 4f_1(t) y_4(0.25) + f_2(t) y_4(0.50)] \\
 &= 1 + \frac{0.25}{3} [0.5(1) + 4(0.4794)(1.1302) + 0.4594(1.2709)] = 1.2709.
 \end{aligned}$$

EXAMPLE 8.7

The prime number theorem states that the number of primes in the interval $a < x < b$ is approximately $\int_a^b \frac{dx}{\log x}$. Use this for $a = 100$ and $b = 200$ and compare with the exact value.

Solution. We know that

$$\log_e x = (\log_{10} x) \cdot \log_e 10 = (2.302585) \log_{10} x.$$

Therefore,

$$\int_{100}^{200} \frac{dx}{\log x} = \int_{100}^{200} \frac{x}{(2.3025) \log_{10} x}$$

We have the following table:

x	100	150	200
f	1	1	1
	$2(2.302585)$	$2.1760(2.302585)$	$2.3010(2.302585)$

Here $h = 50$. We use Simpson's rule and get

$$\begin{aligned}
 \int_{100}^{200} \frac{dx}{\log x} &= \frac{h}{3} (f_0 + 4f_1 + f_2) \\
 &= \frac{50}{3} \left(\frac{1}{4.60517} + \frac{4}{5.0104} + \frac{1}{5.282} \right) \\
 &= 16.6667 (0.2171 + 0.7983 + 0.1887) = 20.068.
 \end{aligned}$$

If $h = 25$, then the table is

x	100	125	150	175	200
f	0.2171	0.2071	0.1996	0.1936	0.1887

and therefore Simpson's formula now yields

$$\begin{aligned}
 \int_{100}^{200} \frac{dx}{\log x} &= \frac{h}{3} [(f_0 + f_4) + 4(f_1 + f_3) + 2f_2] \\
 &= \frac{25}{3} [0.4058 + 4(0.4007) + 2(0.1996)] = 20.065.
 \end{aligned}$$

The exact number of primes between 100 and 200 is 21.

EXAMPLE 8.8

Use Romberg's method to compute

$$\int_0^1 \frac{1}{1+x} dx$$

correct to four decimal places and hence find the value of $\log_e 2$.

Solution. Let $h = 0.5$. Then the values of the integrand $f(x) = \frac{1}{1+x}$ are

x	0	0.5	1.0
$f(x)$	1	0.6667	0.5

Therefore, by trapezoidal rule, we have

$$Q_1 = \int_0^1 \frac{1}{1+x} dx = \frac{0.5}{2}[1 + 2(0.6667) + 0.5] = 0.70835 \approx 0.7084.$$

Now, let $h = 0.25$. Then the values of the integrand are

x	0	0.25	0.5	0.75	1.0
$f(x)$	1	0.8	0.6667	0.5714	0.5

Therefore,

$$Q_2 = \frac{0.25}{2}[1 + 2(0.8 + 0.6667 + 0.5714) + 0.5] = 0.6970.$$

Then

$$R_1 = \frac{4}{3}Q_2 - \frac{1}{3}Q_1 = 0.9293 - 0.236 = 0.6932.$$

Now, let $h = 0.125$. Then the values of the integrand are

x	0	0.125	0.25	0.375	0.5	0.625	0.75	0.875	1.0
$f(x)$	1	0.8889	0.8	0.7272	0.6667	0.6154	0.5714	0.5333	0.5

Then

$$\begin{aligned} R_2 &= \frac{0.125}{3}[1 + 4(0.8889 + 0.7272 + 0.6154 + 0.5333) + 2(0.8 + 0.6667 + 0.5714) + 0.5] \\ &= \frac{0.125}{3}[1 + 11.0592 + 4.0762 + 0.5] = 0.6931. \end{aligned}$$

Then

$$S = \frac{16}{15}R_2 - \frac{1}{15}R_1 = 0.7393 - 0.0462 = 0.6931.$$

Also,

$$\int_0^1 \frac{dx}{1+x} = \left[\log(1+x) \right]_0^1 = \log_e 2.$$

Hence,

$$\log_e 2 \approx 0.6931.$$

EXAMPLE 8.9

Use Romberg's method to compute

$$\int_4^{5.2} \log x \, dx$$

from the data

x	4	4.2	4.4	4.6	4.8	5.0	5.2
$\log_e x$	1.3863	1.4351	1.4816	1.526	1.5686	1.6094	1.6486

Solution. Let $h = 0.4$. Then by trapezoidal rule,

$$Q_1 = \frac{0.4}{2} [1.3863 + 2(1.4816 + 1.5686) + 1.6486] = 1.8271.$$

Now, let $h = 0.2$. Then, again by trapezoidal rule,

$$Q_2 = \frac{0.2}{2} [1.3863 + 2(1.4351 + 1.4816 + 1.526 + 1.5686 + 1.6094) + 1.6486] = 1.8237.$$

Then

$$R_1 = \frac{4}{3} Q_2 - \frac{1}{3} Q_1 = 2.4316 - 0.6090 = 1.8226.$$

EXAMPLE 8.10

Use Romberg's method to compute $\int_0^1 \frac{dx}{1+x^2}$ correct to four decimal places.

Solution. Let $h = 0.5$. The values of the integrand $f(x) = \frac{1}{1+x^2}$ are

$x:$	0	0.5	1
$f(x):$	1	0.8	0.5

Therefore, by trapezoidal rule, we have

$$Q_1 = \int_0^1 \frac{dx}{1+x^2} = \frac{h}{2} [f_0 + 2(f_1 + f_2) + f_3] = \frac{1}{4} [1 + 2(0.8) + 0.5] = 0.775.$$

Now, let $h = 0.25$. Then, the values of the integrand are

$x:$	0	0.25	0.5	0.75	1.0
$f(x):$	1	0.9412	0.8	0.64	0.5

Therefore, by trapezoidal rule,

$$\begin{aligned} Q_2 &= \int_0^1 \frac{dx}{1+x^2} = \frac{h}{2} [f_0 + 2(f_1 + f_2 + f_3) + f_4] \\ &= \frac{1}{8} [1 + 2(0.9412 + 0.8 + 0.64) + 0.5] = 0.7828. \end{aligned}$$

Then, by Romberg's method, we get

$$R_1 = \frac{4}{3}Q_2 - \frac{1}{3}Q_1 = \frac{4}{3}(0.7828) - \frac{1}{3}(0.775) = 0.7854.$$

Now, let $h = 0.125$. Then the values of the integrand are

$x:$	0	0.125	0.25	0.375	0.5	0.625	0.75	0.875	1.0
$f(x):$	1	0.9846	0.9412	0.8767	0.8	0.7191	0.64	0.5664	0.5

Then, by Simpson's one-third rule, we have

$$\begin{aligned} R_2 &= \frac{0.125}{3}[1.4(0.9846 + 0.8767 + 0.7191 + 0.5664) + 2(0.9412 + 0.8 + 0.64) + 0.5] \\ &= \frac{0.125}{3}[1 + 12.5872 + 4.7624 + 0.5] = 0.7854. \end{aligned}$$

Therefore, by Romberg's method,

$$\begin{aligned} S &= \frac{16}{15}R_2 - \frac{1}{15}R_1 = \frac{16(0.7854)}{15} - \frac{1}{15}(0.7854) \\ &= 0.83776 - 0.05236 = 0.7854. \end{aligned}$$

EXAMPLE 8.11

Evaluate $\int_0^1 \frac{dx}{1+x^2}$ using

- (i) Trapezoidal rule taking $h = \frac{1}{4}$,
- (ii) Simpson's one-third rule taking $h = \frac{1}{4}$,
- (iii) Simpson's three-eighth rule taking $h = \frac{1}{6}$,
- (iv) Weddle's rule taking $h = \frac{1}{6}$.

Solution. The value of $f(x) = \frac{1}{1+x^2}$ for first two cases are

$x:$	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$f(x):$	1	0.9412	0.8000	0.6400	0.5000

Case (i): By trapezoidal rule, we have

$$\begin{aligned} \int_0^1 \frac{dx}{1+x^2} &= \frac{h}{2}[f_0 + f_4 + 2(f_1 + f_2 + f_3)] \\ &= \frac{1}{8}[1 + 2(0.9412 + 0.8000 + 0.6400) + 0.5000] \\ &= 0.7828. \end{aligned}$$

Case (ii): Using Simpson's one-third rule, we have

$$\begin{aligned}\int_0^1 \frac{dx}{1+x^2} &= \frac{h}{3}[f_0 + 4(f_1 + f_3) + 2f_2 + f_4] \\ &= \frac{1}{12}[1 + 4(0.9412 + 0.6400) + 2(0.8000) + 0.5000] \\ &= 0.7854.\end{aligned}$$

The values of $f(x)$ for the cases (iii) and (iv) are:

$x:$	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{5}{6}$	1
$f(x):$	1	0.9730	0.9000	0.8000	0.6923	0.5902	0.5000

Case (iii): By Simpson's three-eighth rule, we have

$$\begin{aligned}\int_0^1 \frac{dx}{1+x^2} &= \frac{3h}{8}[(f_0 + f_6) + 3(f_1 + f_3 + f_5) + 2f_4] \\ &= \frac{1}{16}[(1+0.5) + 3(0.9730 + 0.9000 + 0.6923 + 0.5902) + 2(0.8)] \\ &= 0.78541.\end{aligned}$$

Case (iv): By Weddle's rule, we have

$$\begin{aligned}\int_0^1 \frac{dx}{1+x^2} &= \frac{3h}{10}[f_0 + 5f_1 + f_2 + 6f_3 + f_4 + 5f_5 + f_6] \\ &= \frac{1}{20}[1 + 5(0.9730) + 0.9000 + 6(0.8) + 0.6923 + 5(0.5902) + 0.5000] \\ &= 0.78542.\end{aligned}$$

8.7 EULER-MACLAURIN FORMULA

We know that

$$\frac{1}{e^x - 1} = \frac{1}{x} - \frac{1}{2} + Ax + Bx^3 + Cx^5 + \dots, \quad (8.19)$$

where

$$A = \frac{1}{12}, \quad B = \frac{1}{720}, \quad C = \frac{1}{30240}, \dots$$

Also, we have seen that the relation between shift operator E and the differential operator D is $E = e^{hD}$. Putting $x = hD$ in equation (8.19), we get

$$\frac{1}{E - I} = \frac{1}{hD} - \frac{1}{2} + AhD + Bh^3 D^3 + Ch^5 D^5 + \dots$$

Multiplying throughout by $E^n - I$, we get

$$\frac{E^n - I}{E - I} = \frac{1}{hD} (E^n - I) - \frac{1}{2} (E^n - I) + AhD(E^n - I) + Bh^3 D^3(E^n - I) + Ch^5 D^5(E^n - I) + \dots$$

and so

$$\begin{aligned} \left(\frac{E^n - I}{E - I} \right) (y_n) &= \frac{1}{hD} (E^n - I) y_0 - \frac{1}{2} (E^n - I) y_0 + AhD(E^n - I) y_0 + Bh^3 D^3(E^n - I) y_0 + Ch^5 D^5(E^n - I) y_0 + \dots \\ &= \frac{1}{hD} (y_n - y_0) - \frac{1}{2} (y_n - y_0) + AhD(y_n - y_0) + Bh^3 D^3(y_n - y_0) + Ch^5 D^5(y_n - y_0) + \dots \\ &= \frac{1}{hD} (y_n - y_0) - \frac{1}{2} (y_n - y_0) + Ah(y_n' - y_0') + Bh^3 (y_n''' - y_0''') + Ch^5 (y_n^{(v)} - y_0^{(v)}) + \dots \end{aligned} \quad (8.20)$$

But

$$\left(\frac{E^n - I}{E - I} \right) y_0 = y_0 + y_1 + \dots + y_{n-1}$$

and

$$\frac{1}{hD} (y_n - y_0) = \frac{1}{h} \left[\frac{1}{D} (y_n - y_0) \right] = \frac{1}{h} \int_{x_0}^{x_n} f(x) dx.$$

Hence equation (8.20) reduces to

$$\int_{x_0}^{x_n} f(x) dx = \frac{h}{2} [y_0 + 2(y_1 + y_2 + \dots + y_{n-1}) + y_n] - \frac{h^2}{12} (y_n' - y_0') + \frac{h^4}{720} (y_n''' - y_0''') - \frac{h^6}{30240} (y_n^{(v)} - y_0^{(v)}) + \dots,$$

which is known as Euler–Maclaurin formula. We observe that the term

$$\frac{h}{2} [y_0 + 2(y_1 + y_2 + \dots + y_{n-1}) + y_n]$$

represents the approximate value of the integral obtained from trapezoidal rule, while the other terms in (2) represent the successive corrections to this value.

EXAMPLE 8.12

Use Euler–Maclaurin formula to evaluate

$$I = \int_0^1 \sin \pi x dx.$$

Solution. We have

$$\begin{aligned} f(x) &= \sin \pi x \\ f'(x) &= \pi \cos \pi x \\ f''(x) &= -\pi^2 \sin \pi x \\ f'''(x) &= -\pi^3 \cos \pi x \\ f^{(iv)}(x) &= \pi^4 \sin \pi x \\ f^{(v)}(x) &= \pi^5 \cos \pi x \end{aligned}$$

Therefore, Euler–Maclaurin formula in the present case gives

$$\begin{aligned} I &= \frac{h}{2}[y_0 + 2(y_1 + y_2 + \dots + y_{n-1}) + y_n] - \frac{h^2}{12}(-\pi - \pi) - \frac{\pi^4}{720}(\pi + \pi) - \frac{h^6}{30240}(-\pi - \pi^5) + \dots \\ &= \frac{h}{2}[y_0 + 2(y_1 + y_2 + \dots + y_{n-1}) + y_n] + \frac{\pi h^2}{6} - \frac{-\pi^3 h}{360} + \frac{\pi^5 h^6}{15120} + \dots \end{aligned}$$

We take $h = 0.5$, then the table for the function values is

x	0	0.5	1
$\sin \pi x$	0	1	0

Thus,

$$\begin{aligned} I &= \frac{0.5}{2}[0 + 2(1) + 0] + \frac{\pi(0.25)}{6} - \frac{\pi^3(0.0625)}{360} + \frac{\pi^5(0.0156)}{15120} + \dots \\ &= 0.5 + 0.1309 - 0.0054 + 0.00032 \\ &= 0.6258. \end{aligned}$$

Now taking $h = 0.25$, the functional values are

x	0	0.25	0.5	0.75	1
$\sin \pi x$	0	0.7071	1	0.7071	0

Therefore, Euler–Maclaurin formula gives

$$\begin{aligned} I &= \frac{0.25}{2}[0 + 2(0.7071 + 1 + 0.7071) + 0] + \frac{0.0625\pi}{6} - \frac{\pi^3}{360}(0.0039) + \frac{\pi^5}{15120}(0.00024) \\ &= 0.6035 + 0.0327 - 0.0003 + 0.000004 = 0.635904. \end{aligned}$$

The exact value of I is $\frac{2}{\pi} = 0.63661978$.

8.8 DOUBLE INTEGRALS

(A) Trapezoidal Rule

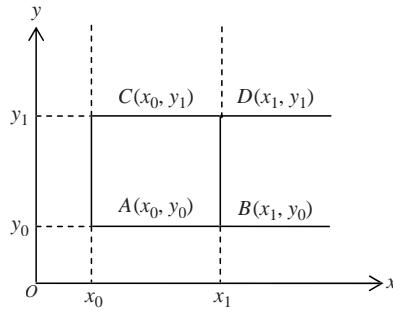
Consider the double integral

$$I = \int_{x_0}^{x_n} \int_{y_0}^{y_n} f(x, y) dx dy,$$

where $x_n = x_0 + nh$, $y_n = y_0 + nk$, h is equidistance between the nodes x_0, \dots, x_n and k is the equidistance between the nodes y_0, \dots, y_n and the region of integration is the area bounded by the lines $x = x_0$, $x = x_n$, $y = y_0$, $y = y_n$ in the xy plane. We evaluate first the integral

$$I_1 = \int_{x_0}^{x_1} \int_{y_0}^{y_1} f(x, y) dx dy.$$

The rectangular area represented by I_1 is bounded by $x = x_0$, $x = x_1$, $y = y_0$ and $y = y_1$ as shown below by ABCD in Figure 8.1.

**Figure 8.1**

Using the ordinary trapezoidal rule for the interval $[x_0, x_1]$, we have

$$\begin{aligned}
 I_1 &= \int_{x_0}^{x_1} \int_{y_0}^{y_1} f(x, y) dx dy \\
 &= \int_{y_0}^{y_1} \frac{h}{2} [f(x_0, y) + f(x_1, y)] dy \\
 &= \frac{h}{2} \int_{y_0}^{y_1} f(x_0, y) dy + \frac{h}{2} \int_{y_0}^{y_1} f(x_1, y) dy . \tag{8.21}
 \end{aligned}$$

Now using the ordinary trapezoidal rule for the interval $[y_0, y_1]$, we have

$$\begin{aligned}
 I_1 &= \frac{hk}{4} [f(x_0, y_0) + f(x_1, y_0) + f(x_0, y_1) + f(x_1, y_1)] \\
 &= \frac{hk}{4} [f(x_0, y_0) + f(x_1, y_0) + f(x_0, y_1) + f(x_1, y_1)] \\
 &= \frac{hk}{4} \times [\text{sum of the values of } f(x, y) \text{ at the four corner points of ABCD}] .
 \end{aligned}$$

Now consider,

$$\begin{aligned}
 I_2 &= \int_{x_0}^{x_1} \int_{y_0}^{y_2} f(x, y) dx dy = \int_{x_0}^{x_1} \left[\int_{y_0}^{y_1} + \int_{y_1}^{y_2} \right] \\
 &= \int_{x_0}^{x_1} \int_{y_0}^{y_1} + \int_{x_0}^{x_1} \int_{y_1}^{y_2} = \left(\int_{x_0}^{x_1} + \int_{x_1}^{x_2} \right) y_0 + \left(\int_{x_0}^{x_1} + \int_{x_1}^{x_2} \right) y_1 \\
 &= \int_{x_0}^{x_1} \int_{y_0}^{y_1} + \int_{x_1}^{x_2} \int_{y_0}^{y_1} + \int_{x_0}^{x_1} \int_{y_1}^{y_2} + \int_{x_1}^{x_2} \int_{y_1}^{y_2} . \tag{8.22}
 \end{aligned}$$

But, as in equation (8.21), we have

$$\int_{x_0}^{x_1} \int_{y_0}^{y_1} f(x, y) dx dy = \frac{hk}{4} [f(x_0, y_0) + f(x_0, y_1) + f(x_1, y_0) + f(x_1, y_1)] ,$$

$$\int_{x_0}^{x_1} \int_{y_0}^{y_1} f(x, y) dx dy = \frac{hk}{4} [f(x_0, y_0) + f(x_1, y_0) + f(x_0, y_1) + f(x_1, y_1)],$$

$$\int_{x_0}^{x_1} \int_{y_1}^{y_2} f(x, y) dx dy = \frac{hk}{4} [f(x_0, y_1) + f(x_1, y_1) + f(x_0, y_2) + f(x_1, y_2)],$$

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x, y) dx dy = \frac{hk}{4} [f(x_1, y_1) + f(x_2, y_1) + f(x_1, y_2) + f(x_2, y_2)].$$

Hence, equation (8.22) reduces to

$$I_2 = \frac{hk}{4} [\{f(x_0, y_0) + f(x_2, y_0) + f(x_0, y_2) + f(x_2, y_2)\} + 2\{f(x_1, y_0) + f(x_1, y_2)\} + 4\{f(x_1, y_1)\}]. \quad (8.23)$$

The region of integration is now as shown in Figure 8.2.

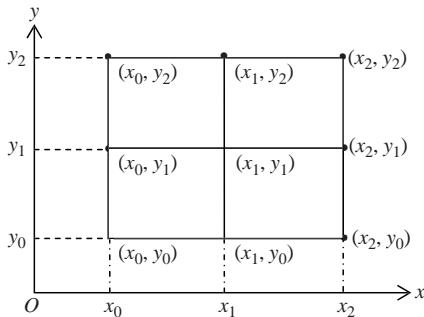


Figure 8.2

Hence,

$$I_2 = \frac{hk}{4} \times [\{\text{sum of values of } f(x, y) \text{ at four corners of the region of integration}\} \\ + 2\{\text{sum of the values of } f(x, y) \text{ at the remaining nodes on the boundary of the region of integration}\} \\ + 4\{\text{value of } f(x, y) \text{ at the internal node } (x_1, y_1)\}].$$

Continuing in the same way, we get

$$I_n = \int_{x_0}^{x_n} \int_{y_0}^{y_n} f(x, y) dx dy \\ = \frac{hk}{4} [\{\text{sum of values of } f(x, y) \text{ at the four corners of the region of integration}\} \\ + 2\{\text{sum of the values of } f(x, y) \text{ at the remaining nodes of the region of integration}\} \\ + 4\{\text{sum of the values of } f(x, y) \text{ at the internal nodes of the region of integration}\}]. \quad (8.24)$$

The expression (8.24) is called the trapezoidal rule for double integration.

(B) Simpson's Rule

To apply Simpson's one-third rule, the number of intervals must be even. So let us consider first the integral

$$I_1 = \int_{x_0}^{x_2} \int_{y_0}^{y_2} f(x, y) dx dy.$$

The region of integration of I_1 is shown in Figure 8.3

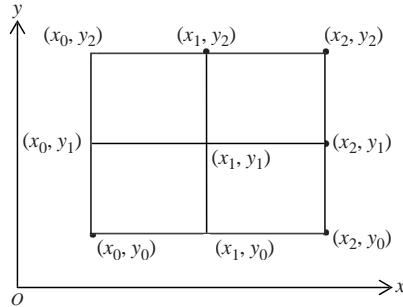


Figure 8.3

Therefore, using ordinary Simpson's one-third rule for the interval I_1 , we get

$$\begin{aligned} I_1 &= \int_{y_0}^{y_2} \left[\int_{x_0}^{x_2} f(x, y) dx \right] dy \\ &= \int_{y_0}^{y_2} \frac{h}{3} [f(x_0, y) + 4f(x_1, y) + f(x_2, y)] dy \\ &= \frac{h}{3} \int_{y_0}^{y_2} f(x_0, y) dy + \frac{4h}{3} \int_{y_0}^{y_2} f(x_1, y) dy + \frac{h}{3} \int_{y_0}^{y_2} f(x_2, y) dy. \end{aligned} \quad (8.25)$$

Now applying ordinary Simpson's one-third rule to equation (8.25) for the interval $[y_0, y_2]$, we get

$$\begin{aligned} I_1 &= \frac{hk}{9} [f(x_0, y_0) + 4f(x_0, y_1) + f(x_0, y_2)] + \frac{4hk}{9} [f(x_1, y_0) + 4f(x_1, y_1) + f(x_1, y_2)] + \frac{hk}{9} [f(x_2, y_0) + 4f(x_2, y_1) + f(x_2, y_2)] \\ &= \frac{hk}{9} [\{f(x_0, y_0) + f(x_0, y_2) + f(x_2, y_0) + f(x_2, y_2)\} + 4\{f(x_0, y_1) + f(x_1, y_0) + f(x_1, y_2) + f(x_2, y_1)\} + 16f(x_1, y_1)] \\ &= \frac{hk}{9} \times [\{\text{sum of the values of } f(x, y) \text{ at the four corners of the region of integration}\} \\ &\quad + 4\{\text{sum of the values of } f(x, y) \text{ at the remaining nodes on the boundary of the region of integration}\} \\ &\quad + 16\{\text{value of } f(x, y) \text{ at the central point } (x_1, y_1)\}]. \end{aligned} \quad (8.26)$$

Now consider

$$\begin{aligned} I_2 &= \int_{x_0}^{x_4} \int_{y_0}^{y_4} f(x, y) dx dy \\ &= \int_{x_0}^{x_4} \left[\int_{y_0}^{y_2} + \int_{y_2}^{y_4} \right] f(x, y) dx dy = \int_{x_0}^{x_4} \int_{y_0}^{y_2} f(x, y) dx dy + \int_{x_0}^{x_4} \int_{y_2}^{y_4} f(x, y) dx dy \end{aligned}$$

$$\begin{aligned}
&= \left(\int_{x_0}^{x_2} + \int_{x_2}^{x_4} \right) \int_{y_0}^{y_2} + \left(\int_{x_0}^{x_2} + \int_{x_2}^{x_4} \right) \int_{y_2}^{y_4} \\
&= \int_{x_0}^{x_2} \int_{y_0}^{y_2} + \int_{x_2}^{x_4} \int_{y_0}^{y_2} + \int_{x_0}^{x_2} \int_{y_2}^{y_4} + \int_{x_2}^{x_4} \int_{y_2}^{y_4}. \tag{8.27}
\end{aligned}$$

Evaluating each of the four double integrals in equation (8.27), as in equation (8.26), we get

$$\begin{aligned}
I_2 &= \frac{hk}{9} [\{f(x_0, y_0) + f(x_4, y_0) + f(x_0, y_4) + f(x_4, y_4)\} + 2\{f(x_2, y_0) \\
&\quad + f(x_4, y_2) + f(x_0, y_4) + f(x_0, y_2)\} + 4\{f(x_1, y_0) + f(x_3, y_0) \\
&\quad + f(x_4, y_1) + f(x_4, y_3) + f(x_3, y_4) + f(x_1, y_4) + f(x_0, y_3) + f(x_0, y_1)\} \\
&\quad + 4\{f(x_1, y_2) + f(x_2, y_2)\} + 8\{f(x_1, y_1) + f(x_2, y_1)\} + 8\{f(x_2, y_1) + f(x_3, y_1)\} \\
&\quad + 16\{f(x_1, y_1) + f(x_2, y_1) + f(x_3, y_1) + f(x_4, y_1)\}].
\end{aligned}$$

Continuing in this way, we get

$$\begin{aligned}
I_n &= \int_{x_0}^{x_n} \int_{y_0}^{y_n} t(x, y) dx dy \\
&= \frac{hk}{9} \times [\text{sum of the values of } f(x, y) \text{ at four corners of the region of integration} \\
&\quad + 2\{\text{sum of the values of } f(x, y) \text{ at the odd positions (except corners) on the boundary of the region of integration}\} \\
&\quad + 4\{\text{sum of the values of } f(x, y) \text{ at the even positions on the boundary } y\} + 4\{\text{the value of } f(x, y) \text{ at the central node}\} \\
&\quad + 8\{\text{sum of the values of } f(x, y) \text{ at the even positions on the central line}\} \\
&\quad + 8\{\text{sum of the values of } f(x, y) \text{ at the even positions on the vertical central line}\} + 16\{\text{sum of the values of } f(x, y) \\
&\quad \text{at the even positions on the even horizontal lines of the region of integration}\}]. \tag{8.28}
\end{aligned}$$

The formula (8.28) is called Simpson's rule for double integral.

EXAMPLE 8.13

Taking differencing as $h = \kappa = 0.1$, evaluate $\int_{-2}^2 \int_{-4}^4 xy dx dy$ using (i) trapezoidal rule and (ii) Simpson's rule.

Solution. Taking $h = \kappa = 0.1$, we get the following table of functional values:

$x \backslash y$	4	4.1	4.2	4.3	4.4
2	8	8.20	8.40	8.60	8.80
2.1	8.40	8.61	8.82	9.03	9.24
2.2	8.80	9.02	9.24	9.46	9.68
2.3	9.20	9.43	9.66	9.89	10.12
2.4	9.60	9.84	10.08	10.32	10.56

- (i) Evaluation of the double integral by trapezoidal rule yields

$$\begin{aligned}
 \int_2^2 \int_{-4}^4 xy \, dx \, dy &= \frac{(0.1)(0.1)}{4} [\{8 + 8.80 + 10.56 + 9.60\} \\
 &\quad + 2\{9.84 + 10.08 + 10.32 + 10.12 + .68 + 9.24 + .60 \\
 &\quad + 8.40 + 8.20 + 8.40 + .80 + 9.20\} + 4\{8.61 + 8.82 \\
 &\quad + 9.33 + 9.02 + 9.24 + 9.46 + 9.43 + 9.66 + 9.89\}] \\
 &= (0.0025)[36.96 + 221.76 + 332.64] = 1.4784.
 \end{aligned}$$

(ii) Simpson's rule yields

$$\begin{aligned}
 \int_2^2 \int_{-4}^4 xy \, dx \, dy &= \frac{(0.1)(0.1)}{9} [\{8 + 8.80 + 10 + 56 + 9.60\} \\
 &\quad + 2\{8.40 + .68 + 10.8 + 8.80\} \\
 &\quad + 4\{9.84 + 10.32 + 10.12 + .24 \\
 &\quad + .60 + 8.20 + .40 + 9.20\} + 4(9.24) \\
 &\quad + 8\{9.02 + 9.46\} + 8\{9.66 + 8.82\} + 16\{9.43 + 9.89 + 8.1 + 9.03\}] \\
 &= 0.0011[36.96 + 73.92 + 295.68 + 36.96 + 147.84 + 147.84 + 591.36] \\
 &= 1.4746.
 \end{aligned}$$

The actual value of the given double integral is 1.4784.

EXAMPLE 8.14

Taking $h = \kappa = 0.25$, evaluate $\int_1^2 \int_1^2 \frac{xy \, dy}{x+y}$ by trapezoidal rule.

Solution. Taking $h = \kappa = 0.25$, the table for the values of $f(x, y)$ is as follows:

$x \backslash y$	1	1.25	1.50	1.75	2.0
1	0.5	0.444	0.4	0.3636	0.3333
1.25	0.444	0.4	0.3636	0.5	0.30769
1.50	0.4	0.3636	0.3333	0.30769	0.28571
1.75	0.3636	0.5	0.30769	0.28571	0.26666
2.0	0.3333	0.30769	0.28571	0.26666	0.50

Therefore, by trapezoidal rule,

$$\begin{aligned}
 \int_1^2 \int_1^2 \frac{xy \, dy}{x+y} &= \frac{(0.25)(0.25)}{4} [\{0.5 + 0.3333 + 0.3333 + 0.50\} \\
 &\quad + 2\{0.30769 + 0.28571 + 0.26666 \\
 &\quad + 0.26666 + 0.28571 + 0.30769 \\
 &\quad + 0.3636 + .4 + 0.444 + 0.444 + 0.4 + 0.3636\} \\
 &\quad + 4\{0.4 + 0.3636 + 0.5 + 0.3636 + 0.3333 \\
 &\quad + 0.30769 + .5 + 0.30769 + 0.28571\}] \\
 &= 0.01562[1.6666 + 8.27064 + 13.44636] = 0.36525.
 \end{aligned}$$

EXERCISES

1. Using Simpson's rule, find the volume of the solid of revolution formed by rotating about x -axis the area between the x -axis, the lines $x = 0$ and $x = 1$ and a curve through the points $(0, 1)$, $(0.25, 0.9896)$, $(0.50, 0.9589)$, $(0.75, 0.9089)$, and $(1, 0.8415)$.

Hint: $\text{Volume} = \int_0^1 \pi y^2 dx = \pi \int_0^1 y^2 dx$

$$= \pi \frac{h}{3} [y_0^2 + 4(y_1^2 + y_3^2) + 2y_2^2 + y_4^2]$$

Ans. 2.8192

2. Find the approximate value of

$$\int_0^{\frac{\pi}{2}} \sqrt{\cos \theta} d\theta$$

by dividing the interval into six parts.

Ans. 1.1873

3. Evaluate

$$\int_1^2 \frac{dx}{x}$$

by Simpson's rule and compare the approximate value obtained with the exact solution

Ans. 0.6932

Exact value: $\log_2 2 = 0.693147$

4. Evaluate

$$\int_0^{\frac{\pi}{2}} \sin x dx$$

by Simpson's one-third rule using 11 ordinates.

Ans. 0.9985

5. The velocity v of a particle at distance s from a point on its path is given by the table:

s ft.:	0	10	20	30	40	50	60
v ft/sec.:	47	58	6	465	61	52	38

Using Simpson's one-third rule, determine the time taken by the particle to travel 60 ft.

Hint: $v = \frac{ds}{dt}$ and so $dt = \frac{1}{v} s$. So find $\int_0^{60} \frac{1}{v} ds$.

Ans. 1.063 sec

6. For the case of six known ordinates, show that

$$\int_0^5 f(x) dx = \frac{5}{288} [19(f_0 + f_5) + 75(f_1 + f_4) + 50(f_2 + f_3)]$$

7. The velocity v km/min of a moped started from rest is given at fixed intervals of time t (minutes) as follows:

$t:$	2	4	6	8	10	12	14	16	18	20
$v:$	10	18	25	29	32	20	11	5	2	0

Using Simpson's rule, find the distance covered in 20 minutes.

Hint: $v = \frac{ds}{dt}$ and so $ds = v dt$. So find $\int_0^{20} dt$. Take interval length equal to 2 and use Simpson's formula.

Ans. 309.33 km

8. Obtain an estimate of the number of subintervals that should be chosen so as to guarantee that the error committed in evaluating $\int_1^2 \frac{1}{x} dx$ by trapezoidal rule is less than 0.001.

Hint: $E_n(x) < -\frac{nh^3}{12} f''(\xi)$,

Ans. $n = 8$

9. Compute the value of

$$\int_0^1 \frac{dx}{1+x^2}$$

using trapezoidal rule with $h = 0.5, 0.25$, and 0.125 . Then use Romberg's method to get better approximation. Compare the result obtained with the true value.

Ans. 0.77500, 0.78279, 0.78475, 0.7854

10. Use Euler–Maclaurin formula to find the value of $\log 2$ from $\int_0^1 \frac{dx}{1+x}$.

Hint : $\int_0^1 \frac{dx}{1+x} = [\log_e(1+x)]_0^1 = \log_e 2$. So find $\int_0^1 \frac{dx}{1+x}$ by Euler–Maclaurin formula.

Ans. 0.693149

11. Calculate by Simpson's rule an approximate value of $\int_{-3}^3 x^4 dx$ by taking seven equidistant ordinates.

Compare it with exact value and the estimate obtained by using trapezoidal rule.

Ans. by Simpson's rule: 98

Exact value: 97.2

by trapezoidal rule: 115

So Simpson's rule yields better results

12. Calculate $\int_2^{10} \frac{dx}{1+x}$ by dividing the range into eight equal parts.

Ans. 1.299

13. If $e^0 = 1, e^1 = 2.72, e^2 = 7.39, e^3 = 20.09, e^4 = 54.60$, find $\int_0^4 e^x$ by Simpson's rule.

Ans. 2.97049

14. A river is 80 feet wide. The depth d (in feet) of the river at a distance x from one bank is given by the following table :

$x:$	0	10	20	30	40	50	60	70	80
$d:$	0	4	7	9	12	15	14	8	3

Find approximately the area of the cross-section of the river.

Hint : Since $A = \int y dx$ and $h = 10$ we have by Simpson's rule,

$$A = \frac{10}{3} [(0+3) + 4(4+9+15+8) + 2(7+12+14)] = 710 \text{ sq. feet.}$$

15. Show that

$$\int_{-1}^1 f(x) dx = \frac{13}{12} [f(1) + f(-1) - f(3) - f(-3)].$$

16. Taking $h = \kappa = 0.5$ evaluate $\int_0^1 \int_0^1 \frac{1}{1+x+y} dx dy$ by

(i) trapezoidal rule, (ii) Simpson's rule.

Ans. (i) 0.5356 (ii) 0.5229

17. Taking $h = \kappa = 2$, evaluate $\int_1^5 \left(\int_1^5 \frac{dx}{\sqrt{x^2 + y^2}} \right) dy$ by trapezoidal rule.

Ans. 3.952

9 Difference Equations

The concept of difference equations is useful in the study of electrical networks and to solve differential equations. In fact, the difference equation is the discrete counterpart of the differential equations. We shall note that the partial differential equations are first converted into difference equations and then solved by numerical methods.

9.1 DEFINITIONS AND EXAMPLES

Definition 9.1. A difference equation is an equation which involves an independent variable, a dependent variable, and the successive differences of the dependent variables.

For example,

$$\Delta y_{n+1} + y_n = 2n^2, \quad (9.1)$$

$$y_{n+1} + \Delta^2 y_{n-1} = 2, \quad (9.2)$$

$$y_{n+3} - 5\Delta y_n = 2^n \quad (9.3)$$

are difference equations.

Since successive differences of a dependent variable can be expressed in terms of the successive values of dependent variable, the difference equation may be defined as an equation which involves independent variable and successive values of the dependent variable.

For example, equation (9.1) can be written as

$$y_{n+2} - y_{n+1} + y_n = 2n^2$$

and is a difference equation.

Difference equations are also called recurrence relations. Thus, a recurrence relation for a sequence $y_0, y_1, \dots, y_n, \dots$ is an equation that relates every term y_n to some of its predecessors y_0, y_1, \dots, y_{n-1} .

EXAMPLE 9.1

Find the sequence represented by the recurrence formula

$$y_1 = 5, \quad y_n = 2y_{n-1}, \quad 2 \leq n \leq 6.$$

Solution. We have

$$\begin{aligned} y_1 &= 5, \\ y_2 &= 2y_1 = 10, \\ y_3 &= 2y_2 = 20, \\ y_4 &= 2y_3 = 40, \\ y_5 &= 2y_4 = 80, \\ y_6 &= 2y_5 = 160. \end{aligned}$$

Therefore, the given recurrence formula defines the finite sequence 5, 10, 20, 40, 80, 160.

EXAMPLE 9.2

Find the recurrence formula for the sequence 87, 82, 77, 72, 67.

Solution. The recurrence formula for the given sequence is

$$y_1 = 87, \quad y_n = y_{n-1} - 5, \quad 2 \leq n \leq 5.$$

Definition 9.2. A difference equation (linear recurrence relation) of order k with constant coefficients is a difference equation of the form

$$y_n = c_1 y_{n-1} + c_2 y_{n-2} + \dots + c_k y_{n-k}, \quad c_k \neq 0.$$

For example,

- (i) The order of the difference equation $y_n = -2y_{n-1}$ is one;
- (ii) The difference equation $y_{n+2} - y_{n+1} + y_n = 3$ is of second order.

Definition 9.3. Solution of a difference equation is an expression for y_n that satisfies the given difference equation.

Definition 9.4. The general solution of a difference equation is that in which the number of arbitrary constants is equal to the order of the difference equation.

Definition 9.5. A particular solution (or particular integral) is that solution which is obtained from the general equation by giving particular values to the constants.

Definition 9.6. A difference equation in which $y_n, y_{n+1}, y_{n+2}, \dots$ occur to the first degree only and are not multiplied together is called a linear difference equation.

Thus, a linear difference equation of order k with constant coefficients is of the form

$$y_{n+k} + c_1 y_{n+k-1} + c_2 y_{n+k-2} + \dots + c_k y_n = f(n), \quad (9.4)$$

where c_1, c_2, \dots, c_k are constants.

9.2 HOMOGENEOUS DIFFERENCE EQUATION WITH CONSTANT COEFFICIENTS

The homogeneous difference equation of order k of the form

$$y_{n+k} + c_1 y_{n+k-1} + c_2 y_{n+k-2} + \dots + c_k y_n = 0 \quad (9.5)$$

is called a homogeneous difference equation with constant coefficients.

Further,

- (i) if $y_1(n)$ is a solution of equation (9.5), then so is $c_1 y_1(n)$.
- (ii) if $y_1(n), y_2(n), \dots, y_k(n)$ are solutions of equation (9.5), then $u(n) = c_1 y_1(n) + c_2 y_2(n) + \dots + c_k y_k(n)$ is also a solution and in fact is its complete solution.
- (iii) if $v(n)$ is a particular solution of equation (9.4), then the total solution of equation (9.4) is $y_n = u(n) + v(n)$.

The part $u(n)$ is called the complementary function (C.F.) and the part $v(n)$ is called the particular integral (P.I.). Thus, the total solution of equation (9.1) is $y_n = C.F. + P.I.$

Consider the homogeneous difference equation

$$y_{n+k} + a_1 y_{n+k-1} + a_2 y_{n+k-2} + \dots + a_k y_n = 0 \quad (9.6)$$

of order k . We seek its solutions of the form $y_n = x^n$ for all n . Substituting $y_n = x^n$ into equation (9.6), we get

$$x^{n+k} + a_1 x^{n+k-1} + \dots + a_k x^n = 0.$$

Dividing by x^n yields

$$p(x) = x^k + a_1 x^{k-1} + \dots + a_k = 0 \quad (9.7)$$

The equation $p(x)$ of degree k in equation (9.7) is called the characteristic equation of the difference equation (9.6).

Suppose that the roots $\alpha_1, \alpha_2, \dots, \alpha_k$ of the characteristic equation (9.7) are all distinct. Then $\alpha_1^n, \alpha_2^n, \dots, \alpha_k^n$ are all solutions of equation (9.6). Hence, by linearity, it follows that

$$y_n = c_1 \alpha_1^n + c_2 \alpha_2^n + \dots + c_k \alpha_k^n \quad (9.8)$$

for arbitrary constants c_1, c_2, \dots, c_k and for all n is also a solution of equation (9.6). Hence, equation (9.8) is a general solution of equation (9.6).

Suppose now that the characteristic equation (9.7) has a double root α_1 . Then

$$p(\alpha_1) = 0 \text{ and } p'(\alpha_1) = 0.$$

Putting $y_n = n\alpha_1^n$ in left-hand side of equation (9.6), we get

$$\begin{aligned} & (n+k)\alpha_1^{n+k} + a_1(n+k-1)\alpha_1^{n+k-1} + \dots + a_k n \alpha_1^n \\ &= \alpha_1^n \{n(\alpha_1^k + a_{k-1}\alpha_1^{k-1} + \dots + a_1) + \alpha_1 [k\alpha_1^{k-1} + a_{k-1}(k-1)\alpha_1^{k-2} + \dots + a_1]\} \\ &= \alpha_1^n [np(\alpha_1) + \alpha_1 p'(\alpha_1)] = 0, \text{ since } p(\alpha_1) = p'(\alpha_1) = 0. \end{aligned}$$

Hence, $y_n = n\alpha_1^n$ is also a solution of equation (9.6). These two solutions α_1^n and $n\alpha_1^n$ are linearly independent. Therefore, the general solution of the difference equation (9.6) is

$$y_n = (c_1 + c_2 n) \alpha_1^n + c_3 \alpha_2^n + \dots + c_k \alpha_k^n.$$

Remark 9.1. If the multiplicity of the root of equation (9.6) is m , then the general solution of the difference equation is

$$y_n = (c_1 + nc_2 + \dots + n^{m-1}c_m) \alpha_1^n + c_2 \alpha_2^n + \dots + c_k \alpha_k^n.$$

Suppose now that the characteristic equation (9.7) has a pair of conjugate complex roots. Let these roots be $\alpha_1 = \alpha + i\beta$ and $\alpha_2 = \alpha - i\beta$. Then $\alpha_1 = re^{i\theta}$, $\alpha_2 = re^{-i\theta}$, where

$$r = \sqrt{\alpha^2 + \beta^2} \text{ and } \theta = \tan^{-1} \left(\frac{\beta}{\alpha} \right).$$

The solution of equation (9.6) corresponding to this pair of roots is

$$\begin{aligned} c_1 \alpha_1^n + c_2 \alpha_2^n &= c_1 r^n e^{in\theta} + c_2 r^n e^{-in\theta} \\ &= r^n [c_1 (\cos n\theta + i \sin n\theta) + c_2 (\cos n\theta - i \sin n\theta)] \\ &= r^n [A \cos n\theta + B \sin n\theta], \end{aligned}$$

where $A = c_1 + c_2$ and $B = i(c_1 - c_2)$.

EXAMPLE 9.3

Solve the difference equation

$$16y_{n+2} - 8y_{n+1} + y_n = 0.$$

Solution. The characteristic equation for the given difference equation is

$$16x^2 - 8x + 1 = 0.$$

The characteristic equation has a double root at $x = \frac{1}{4}$. Therefore, the solution of the difference equation is

$$y_n = (c_1 + c_2 n) \left(\frac{1}{4} \right)^n = \frac{1}{4^n} (c_1 + c_2 n).$$

EXAMPLE 9.4

Solve the difference equation

$$y_{n+3} - 2y_{n+2} - y_{n+1} + 2y_n = 0.$$

Solution. The characteristic equation of the given difference equation is

$$x^3 - 2x^2 - x + 2 = 0.$$

The roots of this equation are 1, -1, and 2. Hence, the general solution of the difference equation is

$$y_n = c_1 (1)^n + c_2 (-1)^n + c_3 (2)^n.$$

EXAMPLE 9.5

Solve the difference equation

$$y_n = 2y_{n-1} - y_{n-2}$$

with initial conditions $y_1 = 1.5$, $y_2 = 3$.

Solution. The characteristic equation for the given difference equation is

$$x^2 - 2x + 1 = 0,$$

which yields

$$x = \frac{2 \pm \sqrt{4-4}}{2} = 1.$$

Hence,

$$y_n = (c_1 + c_2 n)1^n.$$

Therefore,

$$y_1 = c_1 + c_2 = 1.5 \text{ (given)}$$

$$y_2 = c_1 + 2c_2 = 3 \text{ (given)}$$

Solving for c_1 and c_2 , we get $c_1 = 0$, $c_2 = 1.5$.

Hence, the required solution is

$$y_n = 1.5 \dots$$

EXAMPLE 9.6

Solve

$$y_{n+2} - 2y_{n+1} + 2y_n = 0.$$

Solution. The characteristic equation for the given difference equation is

$$x^2 - 2x + 2 = 0,$$

whose roots are $1 \pm i$. Thus, $r = \sqrt{2}$, $\theta = \frac{\pi}{4}$ and so the general solution is

$$y_n = (\sqrt{2})^n \left(c_1 \cos \frac{n\pi}{4} + c_2 \sin \frac{n\pi}{4} \right)$$

EXAMPLE 9.7

Solve the second order difference equation

$$y_{n+1} - 2y_n \cos \alpha + y_n = 0.$$

Solution. The characteristic equation is

$$x^2 - 2x \cos \alpha + 1 = 0$$

and so

$$\begin{aligned} x &= \frac{2 \cos \alpha \pm \sqrt{4 \cos^2 \alpha - 4}}{2} \\ &= \cos \alpha \pm i \sin \alpha. \end{aligned}$$

Hence, the general solution of the given difference equation is

$$y_n = c_1 \cos n\alpha + c_2 \sin n\alpha.$$

EXAMPLE 9.8

Solve

$$(\Delta^2 - 3\Delta + 2)f(n) = 0,$$

where Δ is the forward difference operator.

Solution. In terms of the shift operator, $\Delta = E - I$ and so the given difference equation transforms to

$$[(E - \cdot)^2 - 3(E - I) + 2]f(n) = 0$$

or

$$(E^2 - 5E + 6I)f(n) = 0$$

or

$$f_{n+2} - 5f_{n+1} + 6f_n = 0.$$

Therefore, the characteristic equation of the given difference equation is

$$x^2 - 5x + 6 = 0,$$

whose roots are 3 and 2. Hence, the required solution is

$$f(n) = c_1 3^n + c_2 2^n.$$

9.3 PARTICULAR SOLUTION OF A DIFFERENCE EQUATION

Consider the difference equation

$$y_{n+k} + c_1 y_{n+k-1} + \dots + c_n y_n = f(n).$$

We know that the total solution of this equation is the sum of two parts: the homogeneous solution satisfying the difference equation with right-hand side equal to 0 and the particular solution, which satisfies the difference equation with $f(n)$ on the right-hand side. To find the particular solution, we discuss the following cases:

Case I. If $f(n)$ is a polynomial in n of degree m , then we take

$$_1 n^m + _2 n^{m-1} + \dots + P_{m+1}$$

as the particular solution of the difference equation. Putting this solution in the given difference equation, the values of $_1, _2, \dots, P_{m+1}$ are determined.

EXAMPLE 9.9

Find the total solution of the difference equation

$$y_n - y_{n-1} - 2y_{n-2} = 2n^2.$$

Solution. The characteristic equation of the given difference equation is

$$x^2 - 1 - 2 = 0,$$

which yields

$$x = \frac{1 \pm \sqrt{1+8}}{2} = 2, -1.$$

Therefore, the complementary function is

$$\text{C.F.} = c_1 2^n + c_2 (-1)^n.$$

Suppose the particular solution is of the form $_1 n^2 + _2 n + P_3$, where $_1, _2$, and P_3 are the constants to be determined. Substituting in the given equation, we get

$$(_1 n^2 + P_2 n + P_3) - [P_1(n-1)^2 + P_2(n-1) + P_3] - 2[P_1(n-2)^2 + P_2(n-2) + P_3] = 2n^2$$

or

$$-2P_1 n^2 + n(10P_1 - 2P_2) + (-9P_1 + P_2 - 2P_3) = 2n^2.$$

Comparing coefficients of the powers of n , we have

$$-2P_1 = 2,$$

$$10P_1 - 2P_2 = 0,$$

$$9P_1 - 5P_2 + 2P_3 = 0,$$

which yield

$$P_1 = -1, P_2 = -2, \text{ and } P_3 = -8.$$

Thus, the particular solution is

$$-n^2 - 5n - 8.$$

Hence, the total solution of the difference equation is

$$y_n = \text{C.F.} + \text{P.I.} = c_1 2^n + c_2 (-1)^n - n^2 - 5n - 8.$$

EXAMPLE 9.10

Solve the difference equation

$$y_{n+2} - 4y_n = 9n^2.$$

Solution. The characteristic equation of the given difference equation is

$$x_2 - 4 = 0,$$

and so $x = \pm 2$. Therefore,

$$\text{C.F.} = c_1 2^n + c_2 (-2)^n.$$

Suppose the particular solution is

$$P_1 n^2 + P_2 n + P_3.$$

Putting in the given difference equation, we have

$$[P_1(n+2)^2 + P_2(n+2) + P_3] - 4(P_1 n^2 + P_2 n + P_3) = 9n^2$$

or

$$n^2(-3P_1 - 9) + n(4P_1 - 3P_2) + (4P_1 + 2P_2 - 3P_3) = 0.$$

Therefore,

$$\begin{aligned} -3P_1 - 9 &= 0 \\ 4P_1 - 3P_2 &= 0 \\ 4P_1 + 2P_2 - 3P_3 &= 0. \end{aligned}$$

Hence,

$$P_1 = -3, P_2 = -\frac{4}{3}, \text{ and } P_3 = -\frac{20}{3}.$$

Thus,

$$\text{P.I.} = -3n^2 - 4n - \frac{20}{3}.$$

Hence, the total solution is

$$\begin{aligned} y_n &= \text{C.F.} + \text{P.I.} \\ &= c_1 2^n + c_2 (-2)^n - 3n^2 - 4n - \frac{20}{3}. \end{aligned}$$

EXAMPLE 9.11

Solve

$$y_{n+2} - 4y_n = n^2 + n - 1.$$

Solution. From Example 9.10, we have

$$\text{C.F.} = c_1 2^n + c_2 (-2)^n.$$

Further, by the same example, we have

$$-3P_1 n^2 + n(4P_1 - 3P_2) + (4P_1 + 2P_2 - 3P_3) = n^2 + n - 1.$$

Therefore,

$$-3P_1 = 1 \text{ which yeilds } P_1 = -\frac{1}{3}$$

$$4P_1 - 3P_2 = 1 \text{ which yeilds } P_2 = -\frac{7}{9}$$

$$4P_1 + 2P_2 - 3P_3 = -1 \text{ which yeilds } P_3 = -\frac{17}{27}.$$

Thus,

$$\text{P.I.} = -\frac{1}{3}n^2 - \frac{7}{9}n - \frac{17}{27}.$$

Case II. If $f(n)$ is a constant then the particular solution of the difference equation will be a constant P provided that 1 is not a characteristic root of the difference equation.

EXAMPLE 9.12

Solve

$$y_n - 4y_{n-1} + y_{n-2} = 2.$$

Solution. The characteristic equation of the given difference equation is

$$x^2 - 4x + 5 = 0$$

and so $x = 2 \pm i$. Thus, the homogeneous solution is

$$c_1(2+i)^n + c_2(2-i)^n, n \geq 0.$$

We observe that 1 is not the characteristic root. Therefore, the particular integral shall be a constant P . Putting it into the given difference equation, we get

$$P - 4P + 5P = 2$$

and so $P = 1$. Therefore,

$$\text{P.I.} = 1.$$

Hence, the total solution is

$$y_n = c_1(2+i)^n + c_2(2-i)^n + 1.$$

Case III. If $f(n)$ is of the form α^n , the particular solution of the difference equation shall be of the form $P\alpha^n$ provided that α is not a root of the characteristic equation of the difference equation.

EXAMPLE 9.13

Solve

$$y_{n+2} - 7y_{n+1} + 10y_n = (12)4^n.$$

Solution. The characteristic equation is

$$x^2 - 7x + 10 = 0$$

and so $x = 2, 5$. Hence,

$$\text{C.F.} = c_1 2^n + c_2 5^n.$$

Since 4 is not the characteristic root, the particular integral shall be of the form $P \cdot 4^n$. Substituting in the given difference equation yields

$$P \cdot 4^{n+2} - 7P \cdot 4^{n+1} + 10P \cdot 4^n = (12)4^n$$

or

$$P[4^{n+2} - (7)4^{n+1} + 10 \cdot 4^n] = (12)4^n$$

or

$$P(16 - 28 + 10) = 12.$$

Thus, $P = -6$ and so

$$\text{P.I.} = -(6)4^n.$$

Hence, the total solution is

$$y_n = \text{C.F.} + \text{P.I.} = c_1 2^n + c_2 5^n - (6)4^n.$$

EXAMPLE 9.14

Solve

$$y_n + 5y_{n-1} + y_{n-2} = (56)3^n.$$

Solution. The characteristic equation of the difference equation is

$$x^2 + 5x + 4 = 0,$$

whose roots are -4 and -1 . Therefore,

$$\text{C.F.} = c_1(-4)^n + c_2(-1)^n.$$

Since 3 is not a root of the characteristic equation, the particular integral is of the form $P \cdot 3^n$. Putting in the given equation, we have

$$P \cdot 3^n + 5P \cdot 3^{n-1} + P \cdot 3^{n-2} = (56)3^n,$$

which yields $P = 18$. Therefore,

$$\text{P.I.} = (18)3^n.$$

Hence, the total solution is

$$y_n = \text{C.F.} + \text{P.I.} = c_1(-4)^n + c_2(-1)^n + (18)3^n.$$

Case IV. If $f(n)$ is of the form

$$(c_1 n^m + c_2 n^{m-1} + \dots + c_{m+1}) \alpha^n$$

and α is not a characteristic root of the difference equation, then particular solution shall be of the form

$$(P_1 n^m + P_2 n^{m-1} + \dots + P_{m+1}) \alpha^n.$$

EXAMPLE 9.15

Solve $y_n - 2y_{n-1} - 2y_{n-2} = 3 \cdot n \cdot 4^n$.

Solution. The characteristic equation for the given difference equation is $x^2 - 2 - 2 = 0$ and so $x = 2, -1$ are the characteristic roots. Therefore,

$$\text{C.F.} = c_1 2^n + c_2 (-1)^n.$$

Since $f(n) = 3^n \cdot 4^n$ and 4 is not a characteristic root, the particular integral is of the form $(nP_1 + P_2)4^n$. Substituting this expression in the given equation, we have

$$(nP_1 + P_2)4^n - [(n-1)P_1 + P_2]4^{n-1} - [(n-2)P_1 + P_2]4^{n-2} = 3n \cdot 4^n$$

or

$$n\left(\frac{5P_1}{8}\right) + \left(\frac{P_1}{2} + \frac{5P_2}{8}\right) = 3n.$$

Therefore,

$$\frac{5P_1}{8} = 3, \quad \frac{P_1}{2} + \frac{5P_2}{8} = 0$$

which gives $P_1 = \frac{24}{5}$, $P_2 = -\frac{96}{25}$. Thus,

$$\text{P.I.} = \left(\frac{24}{5}n - \frac{96}{25}\right)4^n.$$

Hence, the total solution of the given difference equation is

$$y_n = \text{C.F.} + \text{P.I.} = c_1 2^n + c_2 (-1)^n + \left(\frac{24}{5}n - \frac{96}{25}\right)4^n.$$

EXAMPLE 9.16

Solve $y_{n+2} - 2y_{n+1} + y_n = n^2 2^n$.

Solution. The characteristic equation of the given difference equation is $x^2 - 2x + 1 = 0$ and so $x = 1, 1$ are the characteristic roots. Therefore,

$$\text{C.F.} = c_1 + c_2 n.$$

Now $f(n) = n^2 2^n$ and 2 is not a characteristic root, therefore the particular integral shall be of the form

$$(n^2 P_1 + nP_2 + P_3)2^n.$$

Substituting this value in the given difference equation, we have

$$[(n+2)^2 P_1 + (n+2)P_2 + P_3]2^{n+2} - 2[(n+1)^2 P_1 + (n+1)P_2 + P_3]2^{n+1} + (n^2 P_1 + nP_2 + P_3)2^n = n^2 \cdot 2^n$$

or

$$n^2 P_1 + n(8P_1 + P_2) + (12P_1 + 4P_2 + P_3) = n^2.$$

Comparing coefficients of the power of n , we have

$$P_1 = 1,$$

$$8P_1 + P_2 = 0 \text{ which yields } P_2 = -8,$$

$$12P_1 + 4P_2 + P_3 = 0 \text{ which yields } P_3 = 20.$$

Thus, the particular solution is

$$\text{P.I.} = n^2 - 8n + 20.$$

Therefore, the total solution of the given equation is

$$y_n = \text{C.F.} + \text{P.I.} = c_1 + c_2 n + n^2 - 8n + 20.$$

Case V. If α is a characteristic root of multiplicity $m-1$ and $f(n)$ is of the form

$$(c_1 n^p + c_2 n^{p-1} + \dots + c_{p+1}) \alpha^n,$$

then the particular solution of the recurrence relation will be of the form

$$n^{m-1} (P_1 n^p + P_2 n^{p-1} + \dots + P_{p+1}) \alpha^n.$$

EXAMPLE 9.17

Solve

$$y_{n+2} - 3y_{n+1} + 2y_n = 2n.$$

Solution. The characteristic equation of the given difference equation is

$$x^2 - 3x + 2 = 0,$$

whose roots are $x = 1, 2$. Therefore,

$$\text{C.F.} = c_1 + c_2 \cdot 2^n.$$

We note that $f(n) = 2n$ is of the form

$$(c_1 n^p + c_2 n^{p-1} + \dots + c_{p+1}) \alpha^n,$$

and 2 is also a characteristic root of multiplicity 1. Therefore, particular solution shall be of the form

$$n(P_1 n + P_2) = 2n^2 + 2nP_2.$$

Putting it in the given equation, we get

$$2(n+2)^2 P_1 + 2(n+2)P_2 - 3[2(n+1)^2 P_1 + 2(n+1)P_2] + 2[2n^2 P_1 + 2nP_2] = 2^n$$

or

$$n(-4P_1) + (2P_1 - 2P_2) = 2n.$$

Comparing coefficients of the powers of n , we get

$$-4P_1 = 2 \text{ and so } P_1 = -\frac{1}{2}$$

$$2P_1 - 2P_2 = 0 \text{ and so } P_1 = P_2 = -\frac{1}{2}.$$

Thus,

$$\text{P.I.} = -n^2 - n.$$

Hence, the total solution of the difference equation is

$$\begin{aligned}y_n &= \text{C.F.} + \text{P.I.} \\&= c_1 + c_2 \cdot 2^n - n^2 - n.\end{aligned}$$

EXAMPLE 9.18

Solve

$$y_n - 4y_{n-1} = (6)4^n.$$

Solution. The characteristic equation of the given difference equation is

$$x - 4 = 0$$

and so the characteristic root is $x = 4$. Therefore,

$$\text{C.F.} = c_1 4^n.$$

Since 4 is a root of multiplicity 1, we have the particular integral of the form

$$(nP_1 + P_2)4^n.$$

Substituting this in the given difference equation, we have

$$(nP_1 + P_2)4^n - 4[(n-1)P_1 + P_2]4^{n-1} = (6)4^n$$

or

$$(nP_1 + P_2) - [(n-1)P_1 + P_2] = 6.$$

This yields $P_1 = 6$, $P_2 = \dots$. Therefore,

$$\text{P.I.} = 6n \dots^n.$$

Hence, the total solution of the given difference equation is

$$y_n = \text{C.F.} + \text{P.I.} = c_1 4^n + 6 \dots 4^n = 4^n(c_1 + 6n).$$

EXERCISES

1. Solve the difference equation

$$y_{n+2} - 2y_{n+1} + 4y_n = 0.$$

Ans. $y_n = 2^n \left[c_1 \cos \frac{n\pi}{2} + c_2 \sin \frac{n\pi}{3} \right]$

2. Solve $y_{n+2} - 4y_n = 0$.

Ans. $y_n = c_1 (-2)^n + c_2 2^n$

3. Solve the difference equation

$${}^2 f(x) = \alpha f(x),$$

where δ is the central operator.Hint. Since $\delta = E^{\frac{1}{2}} - E^{-\frac{1}{2}}$, the equation reduces to

$$[E^2 - (2 + \alpha)E + 1]t(x) = 0.$$

Ans. $f(x) = c_1 \alpha^x + c_2 \alpha^{-x}$,

where $\alpha_1 = \frac{1}{2}(\alpha + 2 + \sqrt{\alpha^2 + 4\alpha})$

$$\alpha_2 = \frac{1}{2}(\alpha + 2 - \sqrt{\alpha^2 + 4\alpha})$$

4. Solve $y_{n+2} - 2y_{n+1} + 4y_n = 0$.

Ans. $y_n = 2^n \left[c_1 \cos \frac{\pi}{3} + c_2 \sin \frac{\pi}{3} \right]$

5. Solve the difference equation

$$(E^2 + E + 1)y_n = n(n+1),$$

where E is the shift operator.

Hint. The difference equation is $y_{n+2} + y_{n+1} + y_n = n(n+1)$.

Ans. $c_1 \cos \frac{n\pi}{3} + c_2 \sin \frac{n\pi}{3} + \frac{1}{3} \left(n^2 - \dots - \frac{2}{3} \right)$

6. Solve $y_{n+2} - 3y_{n+1} + 2y_n = 2^n$.

Ans. $c_1 + c_2 \cdot 2^n + n2^{n-1}$

7. Solve $y_{n+2} - 3y_{n+1} - 4y_n = m^n + n$.

Ans. $c_1 4^n + c_2 (-1)^n - \frac{n}{6} - \frac{1}{36} + \frac{m^n}{n^2 - 3n - 4}$

8. Solve the difference equation

$$y_{n+2} - 2y_{n+1} + y_n = 3n + 5.$$

Ans. $y_n = c_1 + nc_2 + \frac{1}{2}n(n-1)(n+3)$

9. Solve $y_{n+2} + y_{n+1} - 56y_n = 2^n(n^2 - 3)$.

Ans. $y_n = c_1 \cdot 7^n + c_2 (-8)^n - \frac{2^{n-1}}{25}n^2 + \frac{2n}{5} - \frac{64}{25}$

10. Solve $(E^2 - 5E + 6)y = n + 2^n$.

Ans. $y_n = c_1 \cdot 2^n + c_2 \cdot 3^n + \frac{1}{4}(2n+2) - n2^{n-1}$

10 Ordinary Differential Equations

An ordinary differential equation is an equation containing one independent variable and one dependent variable and at least one of its derivatives with respect to the independent variable. We know that a differential equation of n th order has n independent arbitrary constants in its general solution. Therefore, we need n conditions to compute the numerical solution of an n th order differential equation.

10.1 INITIAL VALUE PROBLEMS AND BOUNDARY VALUE PROBLEMS

Problems in which all the initial conditions are specified at the initial point only are called initial value problems or marching problems. Thus, in an initial value problem, all the auxiliary conditions are specified at a point, for example, value of $y, y', \dots, y^{(n-1)}$ at the point x_0 .

As an illustration, we note that the equation

$$y' = x - y^2, \quad y(0) = 1$$

is an initial value problem.

Problems involving second and higher order differential equations in which auxiliary conditions are specified at two or more points are called boundary value problems or jury problems.

As an illustration, we note that the equation

$$y'' = xy, \quad y(0) = 0, \quad y(2) = 1$$

is a boundary value problem.

10.2 CLASSIFICATION OF METHODS OF SOLUTION

Consider first order differential equation $y' = f(x, y)$. Let $x_n = x_0 + nh$ and let y_n be the value of y obtained from a particular method. If the value y_{n+1} appears as a function of just one y -value y_n , then the method is called a single-step method. On the other hand, if the value y_{n+1} appears as a function of several values $y_n, y_{n-1}, \dots, y_{n-p}$, then the method is called a multistep method. Thus, a single-step method is a method that requires only one preceding value of y , while a multistep method requires two or more preceding values of y .

10.3 SINGLE-STEP METHODS

1. Taylor Series Method

Let $f(x, y)$ be a function that is differentiable for sufficient number of times and let

$$\frac{dy}{dx} = y' = f(x, y), \quad y(x_0) = y_0 \tag{10.1}$$

be the initial value problem. We expand $y(x)$ into Taylor series about the point x_0 . Thus,

$$y(x_0 + h) = y_0 + hy'_0 + \frac{h^2}{2!} y''_0 + \dots + \frac{h^p}{p!} y_0^{(p)} + \frac{h^{p+1}}{(p+1)!} y^{(p+1)}(\xi), \tag{10.2}$$

where ξ is a point in $[x_0, x]$. Since the solution is not known, the derivatives in the expansion are not known. However, they can be obtained by taking total derivative of the differential equation (10.1). Therefore,

$$\begin{aligned}y' &= f(x, y), \\y'' &= f_x + f_y y' = f_x + ff_y, \\y''' &= f_{xx} + f_{xy}f + f_{yx}f + f_{yy}f^2 + f_y f_x + f_y^2 f \\&= f_{xx} + 2f_{xy}f + f^2 f_{yy} + f_y^2 f,\end{aligned}$$

and so on.

The number of terms to be included in equation (10.2) is fixed by permissible error. If the permissible error is ε and the series in equation (10.2) is truncated after the term in $y^{(p)}$, then we have

$$\frac{h^{p+1}}{(p+1)!} |y^{(p+1)}(\xi)| < \varepsilon$$

or

$$\frac{h^{p+1}}{(p+1)!} |f^p(\xi)| < \varepsilon.$$

For a given h , we can find p and obtain an upper bound on h . For computational purposes $|f^p(\xi)|$ is replaced by $\max |f^p(\xi_n)|$ in $[x_0, x_n]$.

Advantages:

- (i) A large interval can be used by increasing the number of terms.
- (ii) No special starting procedure is required.
- (iii) The values computed can be checked by applying Taylor's expansion equally on either side of the point x_n . Thus corresponding to y_{n+1} , we may also compute y_{n-1} from the series

$$\begin{aligned}y_{n+1} &= y_n + hy'_n + \frac{h^2}{2!} y''_n + \frac{h^3}{3!} y'''_n + \dots, \\y_{n-1} &= y_n - hy'_n + \frac{h^2}{2!} y''_n - \frac{h^3}{3!} y'''_n + \dots.\end{aligned}$$

Disadvantages:

- (i) The necessity of calculating the higher derivatives makes this method completely unsuitable on high-speed computers.
- (ii) The method is labourious and so is not recommended except for a few equations.

EXAMPLE 10.1

Solve by Taylor series method:

$$y' = y - \frac{2x}{y}, \quad y(0) = 1, \text{ for } x = 0.1 \text{ and } x = -0.1.$$

Solution. The given equation is

$$y' = y - \frac{2x}{y}, \quad y(0) = 1.$$

Therefore,

$$y'' = \frac{y(2yy' - 2) - (y^2 - 2x)y'}{y^2} = \frac{2yy' - 2 - y'^2}{y},$$

$$y''' = \frac{2yy'' - 3y'y'' + 2y'^2}{y},$$

.....

so that

$$y'(0) = y(0) - \frac{2(0)}{y(0)} = y(0) = 1,$$

$$y''(0) = \frac{2y(0)y'(0) - y'^2(0) - 2}{y(0)} = \frac{2 - 1 - 2}{1} = -1,$$

$$y'''(0) = \frac{2y(0)y''(0) - 3y'(0)y''(0) + 2y'^2(0)}{y(0)},$$

$$= \frac{2(1)(-1) - 3(1)(-1) + 2}{1} = 3,$$

.....

.....

Therefore,

$$\begin{aligned} y(0.1) &= y(0) + (0.1)y'(0) + \frac{(0.1)^2}{2!}y''(0) + \frac{(0.1)^3}{3!}y'''(0) + \dots, \\ &= 1 + 0.1 + \frac{0.01}{2}(-1) + \frac{0.001}{3!}(3) + \dots = 1.0955. \end{aligned}$$

Similarly,

$$\begin{aligned} y(-0.1) &= y(0) - (0.1)y'(0) + \frac{(0.1)^2}{2!}y''(0) - \frac{(0.1)^3}{3!}y'''(0) + \dots, \\ &= 1 - 0.1 + \frac{0.01}{2}(-1) + \frac{0.001}{6}(3) + \dots = 0.8955. \end{aligned}$$

EXAMPLE 10.2

Solve the differential equation $y' = x - y^2$, by series expansion, for $x = 0.2(0.2)1$ under the initial condition $y(0) = 1$.

Solution. We have

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

$$y'' = 1 - 2yy' = 1 - 2y(x - y^2) = 1 - 2xy + 2y^3,$$

$$y''' = -2yy'' - 2y'^2 = -2(y - 4xy^2 + 3y^4 + x^2),$$

$$y^{(iv)} = -2yy''' - 2y'y'' - 4y'y'' = -2yy''' - 6y'y'',$$

.....

.....

Using the initial condition $y(0) = 1$, we get

$$y'(0) = 0 - (y(0))^2 = -1,$$

$$y''(0) = 1 - 2y(0)y'(0) = 1 - 2(1)(-1) = 3,$$

$$y'''(0) = -2y(0)y''(0) - 2y'^2 = 2(1)(3) - 2(-1)^2 = -8,$$

$$y^{(iv)}(0) = -2y(0)y'''(0) - 6y'(0)y''(0) = -2(1)(-8) - 6(-1)(3) = 34.$$

Therefore,

$$\begin{aligned} y_1 &= y(0.2) \approx y(0) + 0.2y'(0) + \frac{(0.2)^2}{2!} y''(0) + \frac{(0.2)^3}{3!} y'''(0) + \frac{(0.2)^4}{4!} y^{(iv)}(0) + \dots, \\ &= 1 - 0.2 + 0.06 - 0.01066 + 0.002266 = 0.8516. \end{aligned}$$

Now

$$y_2 = y(0.4) = y_1 + 0.2y'_1 + \frac{(0.2)^2}{2!} y''_1 + \frac{(0.2)^3}{3!} y'''_1 + \frac{(0.2)^4}{4!} y^{(iv)}_1 + \dots$$

But

$$y'_1 = x_1 - y_1^2 = 0.2 - (0.8516)^2 = -0.5252,$$

$$y''_1 = 1 - 2y_1y'_1 = 1 - 2(0.8516)(-0.5252) = 1.8945,$$

$$y'''_1 = -2y_1y''_1 - 2y'^2_1$$

$$= -2(0.8516)(1.8945) - 2(0.5252)^2$$

$$= -3.2267 - 0.5517 = -3.7784,$$

$$y^{(iv)}_1 = -2y_1y'''_1 - 6y'y''_1$$

$$= -2(0.8516)(-3.7784) - 6(-0.5252)(1.8945)$$

$$= 6.43537 + 5.96995 = 12.40532.$$

Therefore,

$$\begin{aligned} y_2 &\approx 0.8516 + 0.2(-0.5252) + \frac{0.04}{2}(1.8945) + \frac{0.008}{6}(-3.7784) + \frac{0.0016}{24}(12.40532) \\ &= 0.8516 - 0.10504 + 0.03789 - 0.00504 + 0.000827 = 0.7802. \end{aligned}$$

Similarly, we can calculate $y(0.6)$, $y(0.8)$, and $y(1)$.

EXAMPLE 10.3

Solve the differential equation $y'' = xy$ for $x = 0.5$ and $x = 1$ by Taylor series method. Initial values: $x = 0$, $y = 0$, $y' = 1$.

Solution. We have

$$y'' = xy,$$

$$y''' = xy' + y,$$

$$y^{(iv)} = xy'' + y' + y' = xy'' + 2y',$$

$$y^{(v)} = xy''' + y'' + 2y'' = xy''' + 3y''.$$

Initial conditions are $y(0) = 0$, $y'(0) = 1$. Further,

$$y''(0) = 0,$$

$$y'''(0) = 0 + y(0) = 0,$$

$$y^{(iv)}(0) = 0 + 2y'(0) = 2(1) = 2,$$

$$y^{(v)}(0) = 0 + 3y''(0) = 0.$$

Hence,

$$\begin{aligned}y_1 &= y(0.5) = y(0) + 0.5y'(0) + \frac{(0.5)^2}{2!} y''(0) + \frac{(0.5)^3}{3!} y'''(0) + \frac{(0.5)^4}{4!} y^{(iv)}(0) + \frac{(0.5)^5}{5!} y^{(v)}(0) + \dots \\&= 0 + 0.5(1) + \frac{0.0625}{24}(2) = 0.5 + 0.00521 = 0.50521.\end{aligned}$$

Now we find $y_2 = y(1)$. We have

$$y_2 = y_1 + 0.5y'_1 + \frac{(0.5)^2}{2!} y''_1 + \frac{(0.5)^3}{3!} y'''_1 + \frac{(0.5)^4}{4!} y^{(iv)}_1 + \dots$$

But

$$\begin{aligned}y'_1 &= y'_0 + hy''_0 + \frac{h^2}{2!} y'''_0 + \frac{h^2}{3!} y^{(iv)}_0 + \dots \\&= 1 + 0.5(0) + \frac{0.25}{2}(0) + \frac{0.125}{6}(2) + \dots \\&= 1.04167.\end{aligned}$$

$$y''_1 = x_1 y'_1 = 0.5(0.50521) = 0.25261.$$

$$y'''_1 = x_1 y'_1 + y''_1 = 0.5(1.04167) + 0.50521 = 1.02604.$$

$$y^{(iv)}_1 = x_1 y''_1 + 2y'_1 = 0.5(0.25261) + 2(1.04167) = 2.2096.$$

Hence,

$$\begin{aligned}y_2 &= y(1) = 0.50521 + 0.5(1.04167) + \frac{0.25}{2}(0.25261) + \frac{0.125}{6}(1.02604) + \frac{0.0625}{24}(2.2096) + \dots \\&\approx 0.50521 + 0.52084 + 0.03157 + 0.021376 = 0.00575 = 1.08475.\end{aligned}$$

2. Euler's Method

Consider the initial value problem

$$y' = \frac{dy}{dx} = f(x, y), y(x_0) = y_0. \quad (10.3)$$

The Euler method is based on the property that in a small interval, a curve is nearly a straight line. Thus, if $x \in [x_0, x_1]$, a small interval, we approximate the curve by the tangent at the point (x_0, y_0) . But the equation of the tangent at (x_0, y_0) is

$$y - y_0 = \left. \frac{dy}{dx} \right|_{(x_0, y_0)} (x - x_0) = f(x_0, y_0)(x - x_0), \quad \text{using equation (10.3)}$$

or

$$y = y_0 + (x - x_0)f(x_0, y_0).$$

Therefore, the value of y corresponding to x_1 is

$$y_1 = y_0 + (x_1 - x_0)f(x_0, y_0).$$

If $x_n = x_0 + nh$, then we get

$$y_1 = y_0 + hf(x_0, y_0).$$

Similarly, approximating the curve by the tangent in $[x_1, x_2]$ at the point (x_1, y_1) with slope $f(x_1, y_1)$, we have

$$y_2 = y_1 + hf(x_1, y_1),$$

and so, in general

$$y_{n+1} = y_n + hf(x_n, y_n). \quad (10.4)$$

The Euler's method is very slow if h is very small. On the other hand, if h is not small, then this method is too inaccurate. These drawbacks suggest further modifications of Euler's method.

Geometric Interpretation: The Euler method has a very simple geometric interpretation. The integral of equation (10.3) yields y as a function of x . Let $y = F(x)$. Then the graph of $y = F(x)$ is a curve in the xy -plane. Since the curve is nearly a straight line in a small interval, we approximate the curve by the tangent at the point (x_0, y_0) . Then, as shown in Figure 10.1, the true value Δy (equal to AQ in the figure) is approximated by $\Delta x \tan \theta$ (equal to AB in the figure). Thus,

$$\Delta y \approx \Delta x \tan \theta = \left(\frac{dy}{dx} \right)_{(x_0, y_0)} \Delta x.$$

Therefore

$$y_1 \approx y_0 + \left(\frac{dy}{dx} \right)_{(x_0, y_0)} (x_1 - x_0)$$

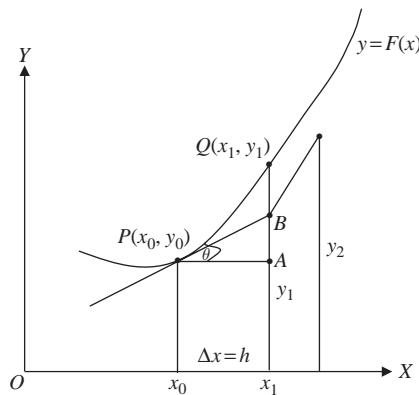


Figure 10.1

In the interval $x_n \leq x \leq x_{n+1}$, the solution is assumed to follow the line tangent to $y(x)$ at (x_n, y_n) . Therefore,

$$y_{n+1} = y_n + hf(x_n, y_n).$$

When this method is applied repeatedly across several intervals in sequence, the numerical solutions traces a polygon segment with sides of slope $f(x_n, y_n)$, $n = 0, 1, 2, \dots$. That is why, this method is also called polygon method.

Error Analysis of Euler's Method

Let $y(x_n)$ be exact value of y at $x = x_n$ and let y_{n+1} be the computed value of y at $x = x_{n+1}$. Then the truncation error after one step, called the local truncation error is given by

$$\begin{aligned} T_{n+1} &= y_{n+1} - y(x_{n+1}) \\ &= y_n + hy'(x_n) - y(x_{n+1}) \quad (\text{by Euler's formula}) \\ &= y_n + hy'(x_n) \left[y_n + hy'_n(x_n) + \frac{h^2}{2} y''_n(\xi) \right], \xi \in [x_n, x_{n+1}] \\ &= -\frac{h^2}{2} y''_n(\xi). \end{aligned}$$

Hence, the local truncation error is $O(h^2)$.

The total truncation error is

$$e_n = y_n - y(x_n).$$

We assume that (i) y_0 is exact so that $e_0 = 0$ and y_i are the values of y computed by Euler's method (ii) Lipschitz condition

$$|f(x, y) - f(x, y^*)| \leq L |y - y^*|$$

is satisfied, and (iii) $|y''(\xi)| \leq M$ in the given interval. By Euler's method, we have

$$y_{n+1} = y_n + hf(x_n, y_n) \quad (10.5)$$

and by Taylor's expansion, we have

$$y(x_{n+1}) = y(x_n) + h f(x_n, y(x_n)) + \frac{h^2}{2!} y''(\xi). \quad (10.6)$$

Subtracting equation (10.6) from equation (10.5), we have

$$e_{n+1} = e_n + h[f(x_n, y_n) - f(x_n, y(x_n))] - \frac{h^2}{2!} y''(\xi).$$

Hence,

$$|e_{n+1}| \leq |e_n| + hL |y_n - y(x_n)| + \frac{h^2}{2!} M$$

or

$$|e_{n+1}| \leq (1 + hL) |e_n| + \frac{h^2}{2} M.$$

Putting $1 + hL = A$ and $\frac{h^2}{2} M = B$, we get

$$|e_{n+1}| \leq A |e_n| + B, n = 0, 1, 2, \dots, N-1.$$

Thus,

$$\begin{aligned} |e_1| &\leq A |e_0| + B \\ |e_2| &\leq A |e_1| + B \leq A[A |e_0| + B] \\ &= A^2 |e_0| + (A+1)B = \frac{A^2 - 1}{A-1} B + A_2 |e_0|, \\ |e_3| &\leq A |e_2| + B = A^3 |e_0| + \frac{A^3 - 1}{A-1} B, \\ &\dots \\ |e_N| &\leq A^N |e_0| + \frac{A^N - 1}{A-1} B. \end{aligned}$$

But $e_0 = 0$ and

$$A^N = (1 + hL)^N \leq e^{NhL} = e^{L(x_N - x_0)}.$$

Hence,

$$|e_N| \leq \frac{1}{2} hM \frac{e^{L(x_N - x_0)} - 1}{L} = O(h).$$

The error tends to zero as $h \rightarrow 0$ in such a way that $nh = x_n - x_0$ remains constant. From this computation it follows that the Euler method is convergent.

Improved Euler's Method

In this method, the curve in the interval $[x_0, x_1]$ is approximated by a line through (x_0, y_0) whose slope is the average of the slopes at (x_0, y_0) and $(x_1, y_1^{(1)})$ such that

$$y_1^{(1)} = y_0 + hf(x_0, y_0).$$

Thus, the equation of the line becomes

$$y - y_0 = (x - x_0) \left[\frac{1}{2} \left\{ f(x_0, y_0) + f(x_1, y_1^{(1)}) \right\} \right]$$

and so the line through (x_0, y_0) and (x_1, y_1) is

$$y - y_0 = (x - x_0) \frac{1}{2} [f(x_0, y_0) + f(x_1, y_1^{(1)})]$$

or

$$\begin{aligned} y_1 &= y_0 + \frac{h}{2} \left[f(x_0, y_0) + f(x_1, y_1^{(1)}) \right] \\ &= y_0 + \frac{h}{2} \left[f(x_0, y_0) + f(x_0 + h, y_0 + hf(x_0, y_0)) \right]. \end{aligned}$$

Hence, the general formula becomes

$$y_{n+1} = y_n + \frac{h}{2} \left[f(x_n, y_n) + f(x_n + h, y_n + hf(x_n, y_n)) \right],$$

where $x_n - x_{n-1} = h$.

Geometrical Interpretations: Let Δy computed by Euler's method is represented by AB. If PC is drawn parallel to the tangent at $Q(x_1, y_1)$, then Δy computed by using the slope at Q is represented by AC. On the other hand, if we take the average of the slopes, we have

$$\begin{aligned} \Delta y &= \frac{\left(\frac{dy}{dx} \right)_{(x_0, y_0)} + \left(\frac{dy}{dx} \right)_{(x_1, y_1)}}{2} h \\ &= \frac{1}{2} (AB + AC) = \frac{1}{2} (AB + AB + BC) \\ &= AB + \frac{1}{2} BC, \end{aligned}$$

which is very close to the true value AQ (Figure 10.2).

Therefore,

$$\begin{aligned} y_1 &= y_0 + \frac{h}{2} \left[\left(\frac{dy}{dx} \right)_{(x_0, y_0)} + \left(\frac{dy}{dx} \right)_{(x_1, y_1)} \right] \\ &= y_0 + \frac{h}{2} [f(x_0, y_0) + f(x_0 + h, y_0 + hf(x_0, y_0))]. \end{aligned}$$

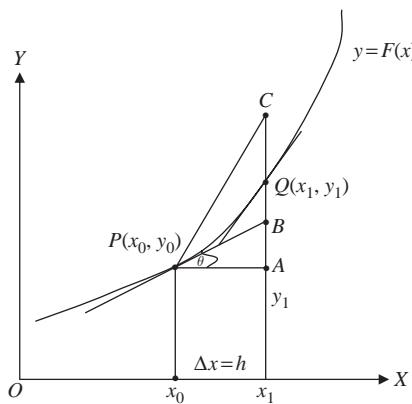


Figure 10.2

Hence, the general formula becomes

$$y_{n+1} = y_n + \frac{h}{2} [f(x_n, y_n) + f(x_n + h, y_n + h f(x_n, y_n))].$$

Modified Euler's Method

In this method, the curve in the interval $[x_0, x_1]$ is approximated by the line through (x_0, y_0) with slope $f\left(x_0 + \frac{h}{2}, y_0 + \frac{h}{2} f(x_0, y_0)\right)$, that is, the slope at the midpoint whose abscissa is the average of x_0 and x_1 , that is, the slope at $x_0 + \frac{h}{2}$. Thus, the equation of the line is

$$y - y_0 = (x - x_0) \left\{ f\left(x_0 + \frac{h}{2}, y_0 + \frac{h}{2} f(x_0, y_0)\right) \right\}.$$

Taking $x = x_1$, we have

$$y_1 = y_0 + h \left\{ f\left(x_0 + \frac{h}{2}, y_0 + \frac{h}{2} f(x_0, y_0)\right) \right\}.$$

Hence, the general formula becomes

$$y_{n+1} = y_n + h \left\{ f\left(x_n + \frac{h}{2}, y_n + \frac{h}{2} f(x_n, y_n)\right) \right\}.$$

EXAMPLE 10.4

Solve, by Euler's method, the initial value problem

$$\frac{dy}{dx} = \frac{x-y}{2}, \quad y(0) = 1$$

over $[0, 3]$, using step size $\frac{1}{2}$.

Solution. By Euler's method,

$$y_{n+1} = y_n + h f(x_n, y_n).$$

We are given that $h = \frac{1}{2}$ and $f(x, y) = \frac{x - y}{2}$. Therefore,

$$y_{n+1} = y_n + 0.5 \left(\frac{x_n - y_n}{2} \right) = 0.25x_n + 0.75y_n.$$

Thus,

$$\begin{aligned} y_1 &= 0.25x_0 + 0.75y_0 = 0.25(0) + 0.75(1) \\ &= 0.75, \\ y_2 &= 0.25x_1 + 0.75y_1 = 0.25(0.5) + 0.75(0.75) \\ &= 0.125 + 0.5625 = 0.6875, \\ y_3 &= 0.25(1) + 0.75(0.6875) \\ &= 0.25 + 0.515625 = 0.765625, \\ y_4 &= 0.25(1.5) + 0.75(0.765625) = 0.375 + 0.57421875 \\ &= 0.94921875, \\ y_5 &= 0.25(2) + 0.75(0.94921875) = 0.50 + 0.711914062 \\ &= 1.211914063, \\ y_6 &= 0.25(2.5) + 0.75(1.211914063) \\ &= 0.625 + 0.908935546 \\ &= 1.533935547 \approx 1.533936. \end{aligned}$$

EXAMPLE 10.5

Solve the initial value problem

$$\frac{dy}{dx} = \frac{y-x}{y+x}, \quad y(0) = 1 \text{ for } x = 0.1 \text{ by Euler's method.}$$

Solution. By Euler's method

$$y_{n+1} = y_n + hf(x_n, y_n).$$

We take $h = 0.02$. Therefore,

$$y_{n+1} = y_n + 0.02 \left(\frac{y_n - x_n}{y_n + x_n} \right)$$

and so

$$\begin{aligned} y_1 &= y_0 + 0.02 \left(\frac{y_0 - x_0}{y_0 + x_0} \right) \\ &= 1 + 0.02 \left(\frac{1 - 0}{1 + 0} \right) = 1.02, \\ y_2 &= y_1 + 0.02 \left(\frac{y_1 - x_1}{y_1 + x_1} \right) \\ &= 1.02 + 0.02 \left(\frac{1.02 - 0.02}{1.02 + 0.02} \right) = 1.0392, \end{aligned}$$

$$\begin{aligned}
y_3 &= y_2 + 0.02 \left(\frac{y_2 - x_2}{y_2 + x_2} \right) \\
&= 1.0392 + 0.02 \left(\frac{1.0392 - 0.04}{1.0392 + 0.04} \right) = 1.05918, \\
y_4 &= y_3 + 0.02 \left(\frac{y_3 - x_3}{y_3 + x_3} \right) \\
&= 1.05918 + 0.02 \left(\frac{1.05918 - 0.06}{1.05918 + 0.06} \right) = 1.07917, \\
y_5 &= y_4 + 0.02 \left(\frac{y_4 - x_4}{y_4 + x_4} \right) \\
&= 1.07917 + 0.02 \left(\frac{1.07917 - 0.08}{1.07917 + 0.08} \right) = 1.09916.
\end{aligned}$$

Hence, the required solution is 1.09916.

EXAMPLE 10.6

Use Euler's method and its modified form to obtain $y(0.2)$, $y(0.4)$, and $y(0.6)$ correct to three decimal places, given that $y' = y - x^2$ with initial condition $y(0) = 1$.

Solution. By Euler's method,

$$y_{n+1} = y_n + hf(x_n, y_n).$$

Here $f(x, y) = y - x^2$ and $h = 0.2$. Therefore,

$$y_{n+1} = y_n + 0.2(y_n - x_n^2) = 1.2y_n - 0.2x_n^2.$$

Thus,

$$\begin{aligned}
y_1 &= 1.2y_0 - 0.2x_0^2 = 1.2(1) = 1.2, \\
y_2 &= 1.2y_1 - (0.2)x_1^2 = (1.2)^2 - (0.2)^3 = 1.44 - 0.008 = 1.4320, \\
y_3 &= 1.2y_2 - (0.2)x_2^2 = (1.2)(1.432) - (0.2)(0.4)^2 = 1.6864.
\end{aligned}$$

Modified Euler's formula is

$$y_{n+1} = y_n + hf\left(x_n + \frac{h}{2}, y_n + \frac{h}{2}f(x_n, y_n)\right).$$

Taking $h = 0.2$, we have

$$\begin{aligned}
y_1 &= y_0 + 0.2 \left[y_0 + \frac{0.2}{2}(y_0 - x_0^2) - \left(x_0 + \frac{0.2}{2} \right)^2 \right] \\
&= 1 + 0.2[1 + 0.1(1 - 0) - (0 + 0.1)^2] \\
&= 1 + 0.2(1 + 0.1 - 0.01) = 1.218, \\
y_2 &= y_1 + 0.2[y_1 + 0.1(y_1 - x_1^2) - (x_1 + 0.1)^2] \\
&= 1.218 + 0.2[1.218 + 0.1(1.218 + (0.2)^2) - (0.2 + (0.1)^2)] \\
&= 1.218 + 0.2[1.218 + 0.1178 - 0.09] = 1.4672,
\end{aligned}$$

$$\begin{aligned}
y_3 &= y_2 + 0.2[y_2 + 0.1(y_2 - x_2^2) - (x_2 + 0.1)^2] \\
&= 1.4672 + 0.2[1.4672 + 0.1(1.4672 - 0.4^2) - (0.4 + 0.1)^2] \\
&= 1.4672 + 0.2[1.4672 + 0.13072 - 0.25] = 1.7368.
\end{aligned}$$

EXAMPLE 10.7

Using Euler modified method, obtain a solution of $\frac{dy}{dx} = x + \sqrt{|y|}$, $y(0) = 1$ for the range $0 \leq x \leq 0.6$ in steps of 0.2.

Solution. The given differential equation is

$$\frac{dy}{dx} = x + \sqrt{|y|}, \quad y(0) = 1.$$

The modified Euler's formula is

$$y_{n+1} = y_n + h f\left(x_n + \frac{h}{2}, y_n + \frac{h}{2} f(x_n, y_n)\right).$$

Taking $h = 0.2$, we have

$$\begin{aligned}
y_1 &= y_0 + 0.2 \left[x_0 + \frac{0.2}{2} + |y_0| + \frac{0.2}{2} \left(x_0 + \sqrt{|y_0|} \right) \right] \\
&= 1 + 0.2[0.1 + 1 + 0.1(1)] = 1.240 \\
y_2 &= y_1 + 0.2 \left[x_1 + \frac{0.2}{2} + |y_1| + \frac{0.2}{2} \left(x_1 + \sqrt{|y_1|} \right) \right] \\
&= 1.24 + 0.2[0.2 + 0.1 + 1.24 + 0.1(0.2 + 1.114)] \\
&= 1.24 + 0.33428 = 1.574. \\
y_3 &= y_2 + 0.2 \left[x_2 + \frac{0.2}{2} + |y_2| + \frac{0.2}{2} \left(x_2 + \sqrt{|y_2|} \right) \right] \\
&= 1.574 + 0.2[0.4 + 0.1 + 1.547 + 0.1(0.4 + 1.255)] = 2.0219.
\end{aligned}$$

3. Picard's Method of Successive Integration

Consider the initial value problem

$$y'(x) = f(x, y(x)) \text{ over } [a, b] \text{ with } y(x_0) = y_0. \quad (10.5)$$

Using fundamental theorem of calculus, we have

$$\int_{x_0}^{x_1} f(x, y(x)) dx = \int_{x_0}^{x_1} y'(x) dx = y(x_1) - y(x_0). \quad (10.6)$$

Thus,

$$y(x_1) = y(x_0) + \int_{x_0}^{x_1} f(x, y(x)) dx. \quad (10.7)$$

Thus, if we start with the approximation $y(x_0)$, then

$$y_1 = y_0 + \int_{x_0}^{x_1} f(x, y_0) dx;$$

$$y_2 = y_0 + \int_{x_0}^{x_1} f(x, y_1) dx;$$

...

$$y_{n+1} = y_0 + \int_{x_0}^{x_1} f(x, y_n) dx.$$

We stop the process when $y_{n+1} = y_n$ upto desired decimal places.

The Picard's method of successive integration fails if the function is not easily integrable.

EXAMPLE 10.8

Using Picard's method, solve

$$\frac{dy}{dx} = x^2 - y, \quad y(0) = 1 \quad \text{for } x = 0.2.$$

Solution. We start with the approximation $y(0) = 1$. Then

$$\begin{aligned} y_1 &= y_0 + \int_0^{0.2} (x^2 - y_0) dx = 1 + \int_0^{0.2} (x^2 - 1) dx \\ &= 1 + \left[\frac{x^3}{3} - x \right]_0^{0.2} = 1 + \left[\frac{(0.2)^3}{3} - 0.2 \right] = 0.8027 \\ y_2 &= 1 + \int_0^{0.2} (x^2 - y_1) dx = 1 + \int_0^{0.2} (x^2 - 0.8027) dx \\ &= 1 + \left[\frac{x^3}{3} - 0.8027x \right]_0^{0.2} = 1 + [0.00267 - 0.16054] = 0.8421, \\ y_3 &= 1 + \int_0^{0.2} (x^2 - y_2) dx = 1 + \left[\frac{x^3}{3} - y_2 x \right]_0^{0.2} \\ &= 1 + [0.00267 - (0.8421)(0.2)] = 0.8342, \\ y_4 &= 1 + [0.00267 - (0.8342)(0.2)] = 0.8358, \\ y_5 &= 1 + [0.00267 - (0.8358)(0.2)] = 0.8355. \end{aligned}$$

Hence $y(0.2) = 0.835$ upto three decimal places.

EXAMPLE 10.9

Solve

$$y' = x^2 + 2xy, \quad y(0) = 0.$$

Solution. We take first approximation to be $y(0) = 0$. Then,

$$\begin{aligned}y_1 &= y_0 + \int_0^x (x^2 + 2xy(0))dx = 0 + \int_0^x x^2 dx = \frac{x^3}{3}, \\y_2 &= 0 + \int_0^x \left(x^2 + 2x \left(\frac{x^3}{3} \right) \right) dx = \frac{x^3}{3} + \frac{2x^5}{15}, \\y_3 &= 0 + \int_0^x \left[x^2 + 2x \left(\frac{x^3}{3} + \frac{2x^5}{15} \right) \right] dx = \frac{x^3}{3} + \frac{2x^5}{3(5)} + \frac{4x^7}{3(5)(7)}, \\y_4 &= 0 + \int_0^x \left[x^2 + 2x \left(\frac{x^3}{3} + \frac{2x^5}{3(5)} + \frac{4x^7}{3(5)(7)} \right) \right] dx \\&= \frac{x^3}{3} + \frac{2x^5}{3(5)} + \frac{4x^7}{3(5)(7)} + \frac{8x^9}{3(5)(7)(9)}.\end{aligned}$$

EXAMPLE 10.10

Solve by Picard's method,

$$\frac{dy}{dx} = 1 + xy, \quad y(0) = 1 \text{ for } x = 0.1.$$

Solution. We take first approximation to be $y(0) = 1$. Then,

$$\begin{aligned}y_1 &= y_0 + \int_0^x f(x, y(0))dx \\&= 1 + \int_0^x (1+x)dx = 1 + x + \frac{x^2}{2}, \\y_2 &= 1 + \int_0^x \left[1+x \left(1+x + \frac{x^2}{2} \right) \right] dx \\&= 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{8}, \\y_3 &= 1 + \int_0^x \left[1+x \left(1+x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{8} \right) \right] dx \\&= 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{8} + \frac{x^5}{15} + \frac{x^6}{48}.\end{aligned}$$

Thus,

$$\begin{aligned}y_3(0.1) &= 1 + 0.1 + \frac{(0.1)^2}{2} + \frac{(0.1)^3}{3} + \frac{(0.1)^4}{8} + \frac{(0.1)^5}{15} + \frac{(0.1)^6}{48} \\&= 1 + 0.1 + \frac{0.01}{2} + \frac{0.001}{3} + \frac{0.0001}{8} + \frac{0.00001}{15} + \frac{0.000001}{48} \\&= 1.105346.\end{aligned}$$

Further,

$$\begin{aligned} y_4 &= 1 + \int_0^x \left[1 + x \left(1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{8} + \frac{x^5}{15} + \frac{x^6}{48} \right) \right] dx \\ &= 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{8} + \frac{x^5}{15} + \frac{x^6}{48} + \frac{x^7}{105} + \frac{x^8}{384}. \end{aligned}$$

Thus,

$$\begin{aligned} y_4(0.1) &= 1 + 0.1 + \frac{0.01}{2} + \frac{0.001}{3} + \frac{0.0001}{8} + \frac{0.00001}{15} + \frac{0.000001}{48} + \frac{0.0000001}{105} + \frac{0.00000001}{384} \\ &= 1.1053465. \end{aligned}$$

Hence,

$$y(0.1) = 1.1053465.$$

4. Heun's Method

Consider the initial value problem

$$y'(x) = f(x, y(x)), y(x_0) = y_0 \quad (10.8)$$

over the interval $[a, b]$. By fundamental theorem of calculus, we have

$$\int_{x_0}^{x_1} f(x, y(x)) dx = \int_{x_0}^{x_1} y'(x) dx = y(x_1) - y(x_0). \quad (10.9)$$

Hence,

$$y(x_1) = y(x_0) + \int_{x_0}^{x_1} f(x, y(x)) dx. \quad (10.10)$$

Using trapezoidal rule with $h = x_1 - x_0$, equation (10.10) reduces to

$$y(x_1) \approx y(x_0) + \frac{h}{2} [f(x_0, y(x_0)) + f(x_1, y(x_1))]. \quad (10.11)$$

We observe that $y(x_1)$ appears on both sides of equation (10.11). We replace $y(x_1)$ on right-hand side of equation (10.11) by Euler's method. Thus, $y(x_1)$ on the right-hand side is replaced by

$$y_0 + hf(x_0, y_0).$$

Hence,

$$y(x_1) = y(x_0) + \frac{h}{2} [f(x_0, y_0) + f(x_1, y_0 + h f(x_0, y_0))], \quad (10.12)$$

which is called Heun's method.

The process is repeated to get closer and closer approximation. At each step, Euler's method is used as a predictor and trapezoidal rule is used as corrector. Thus, the general step for Heun's method can be expressed as

$$\begin{aligned} p_{n+1} &= y_n + hf(x_n, y_n) \\ y_{n+1} &= y_n + \frac{h}{2} [f(x_n, y_n) + f(x_{n+1}, p_{n+1})]. \end{aligned}$$

EXAMPLE 10.11

Use Heun's method to solve the initial value problem $y'(x) = \frac{x-y}{2}$, $y(0) = 1$ over $[0, 2]$ using step size $\frac{1}{2}$.

Solution. By Example 10.4, we have

$$p_1 = 0.75.$$

Therefore,

$$\begin{aligned} y_1 &= y_0 + \frac{h}{2}[f(x_0, y_0) + f(x_1, p_1)] \\ &= 1 + \frac{1}{4} \left[-\frac{1}{2} + \frac{(1/2) - 0.75}{2} \right] \\ &= 1 + \frac{1}{4} \left[-\frac{1}{2} - \frac{0.25}{2} \right] = 1 - \frac{1.25}{8} = 0.84375. \end{aligned}$$

Again, by Example 10.4,

$$p_2 = 0.6875.$$

Therefore,

$$\begin{aligned} y_2 &= y_1 + \frac{h}{2}[f(x_1, y_1) + f(x_2, p_2)] \\ &= 0.84375 + \frac{1}{4} \left[\frac{(1/2) - 0.84375}{2} + \frac{1 - 0.6875}{2} \right] \\ &= 0.84375 + \frac{1}{8}[-0.34375 - 0.3125] = 0.83984. \end{aligned}$$

Further,

$$p_3 = 0.765625$$

and so

$$\begin{aligned} y_3 &= y_2 + \frac{h}{2}[f(x_2, y_2) + f(x_3, p_3)] \\ &= 0.83984 + \frac{1}{4} \left[\frac{1 - 0.83984}{2} + \frac{1.5 - 0.765625}{2} \right] \\ &= 0.83984 + \frac{1}{8}[0.16016 + 0.73475] = 0.95170375. \end{aligned}$$

Now,

$$p_4 = 0.94921875$$

and so

$$\begin{aligned} y_4 &= y_3 + \frac{h}{2}[f(x_3, y_3) + f(x_4, p_4)] \\ &= 0.95170375 + \frac{1}{8}[(1.5 - 0.9517035) + (2 - 0.94921875)] \\ &= 0.95170375 + \frac{1}{8}[0.5482965 + 1.05078125] = 1.151588. \end{aligned}$$

5. Runge–Kutta Method

Consider the initial value problem

$$y'(x) = f(x, y), \quad y(x_0) = y_0.$$

We note that

$$\begin{aligned} y'(x) &= f(x, y), \\ y''(x) &= f_x + f_y y' = f_x + f_y f, \\ y'''(x) &= f_{xx} + f_{xy} y' + f_y y'' + y'(f_{yx} + f_{yy} y') \\ &= f_{xx} + 2f_{xy} y' + f_y y'' + f_{yy} (y')^2 \\ &= f_{xx} + 2f_{xy} y' + f_y (f_x + f_y f) + f_{yy} f^2 \\ &= f_{xx} + 2f_{xy} y' + f_{yy} f^2 + f_y (f_x + f_y f), \end{aligned}$$

and is general

$$y^{(n)}(x) = d^{(n-1)} f(x, y(x)),$$

where

$$d = \left(\frac{\partial}{\partial x} + f \frac{\partial}{\partial y} \right).$$

If we put

$$\begin{aligned} F_1 &= f_x + f f_y, \\ F_2 &= f_{xx} + 2f f_{xy} + f^2 f_{yy}, \\ F_3 &= f_{xxx} + 3f f_{xxy} + 3f^2 f_{xyy} + f^3 f_{yyy}, \end{aligned}$$

then

$$\begin{aligned} y'(x) &= f(x, y), \\ y''(x) &= F_1, \\ y'''(x) &= F_2 + F_1 f_y, \\ y^{(iv)}(x) &= F_3 + F_2 f_y + 3F_1 f_y + F_1 f_y^2. \end{aligned}$$

Using Taylor's series expansion, we have

$$\begin{aligned} y_{r+1} &= y(x_r + h) = y_r + h y'_r + \frac{h^2}{2!} y''_r + \frac{h^3}{3!} y'''_r + \frac{h^4}{4!} y^{(iv)}_r + O(h^5) \\ &= y_r + hf(x_r, y_r) + \frac{h^2}{2} F_1 + \frac{h^3}{3!} (F_2 + F_1 f_y) + \frac{h^4}{4!} (F_3 + F_2 f_y + 3F_1 f_y + F_1 f_y^2) + O(h^5). \end{aligned} \quad (10.13)$$

On the other hand, by Taylor's theorem for function of two variables, we have

$$f(x_r + h, y_r + k) = f(x_r, y_r) \left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right) f + \frac{1}{2!} \left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right)^2 f + \dots \quad (10.14)$$

A Runge–Kutta method of order n is a formula which expresses $y_{r+1} - y_r$ in terms of n values of the function $f(x, y)$ in such a manner that the values obtained coincide with equation (10.13) as far as the terms involving h^n .

Second Order Runge–Kutta Method

Define

$$\left. \begin{aligned} K_1 &= hf(x_r, y_r) \\ K_2 &= hf(x_r + mh, y_r + mK_1). \end{aligned} \right\} \quad (10.15)$$

Our aim is to obtain an expression of the form

$$y_{r+1} = y_r + aK_1 + bK_2. \quad (10.16)$$

If we put

$$\begin{aligned} F_1 &= f_x + ff_y \\ F_2 &= f_{xx} + 2ff_{xy} + f^2 ff_{yy}, \end{aligned}$$

then the left-hand side of equation (10.16) becomes

$$\begin{aligned} y_{r+1} &= y_r + hy'_r + \frac{h^2}{2!} y''_r + \frac{h^3}{3!} y'''_r + \frac{h^4}{4!} y^{(iv)}_r + O(h^5) \\ &= y_r + hf(x_r, y_r) + \frac{h^2}{2!} F_1 + \frac{h^3}{3!} (F_2 + F_1 f_y + \dots) \end{aligned} \quad (10.17)$$

Now expanding K_2 by Taylor's series for a function of two variables, we have

$$\begin{aligned} K_2 &= \left[f(x_r, y_r) + \left(mh \frac{\partial}{\partial x} + mK_1 \frac{\partial}{\partial y} \right) f + \frac{1}{2!} \left(mh \frac{\partial}{\partial x} + mK_1 \frac{\partial}{\partial y} \right)^2 f + O(h^3) \right] \\ &= h \left[f(x_r, y_r) + mh f_x + mK_1 f_y + \frac{m^2 h^2}{2} f_{xx} + m^2 h K_1 f_{xy} + \frac{m^2 K_1^2}{2} f_{yy} + O(h^3) \right] \end{aligned}$$

Putting the value of K_1 in K_2 , we get

$$K_2 = h \left[f(x_r, y_r) + mh f_x + mh f_y f(x_r, y_r) + \frac{m^2 h^2}{2} f_{xx} + m^2 h^2 f_{xy} f(x_r, y_r) + \frac{m^2 h^2}{2} f_{yy} f(x_r, y_r) + O(h^3) \right].$$

Thus, putting the values of K_1 and K_2 from equation (10.15) into equation (10.16), we get

$$\begin{aligned} y_{r+1} &= y_r + ahf(x_r, y_r) \\ &\quad + bh \left[f(x_r, y_r) + mh f_x + mh f_y f(x_r, y_r) + \frac{m^2 h^2}{2} f_{xx} + m^2 h^2 f_{xy} f(x_r, y_r) + \frac{m^2 h^2}{2} f_{yy} f(x_r, y_r) + O(h^3) \right] \\ &= y_r + (a+b)hf + mbh^2(f_x + ff_y) + \frac{m^2 bh^3}{2}(f_{xx} + 2ff_{xy} + f^2 ff_{yy}) + O(h^4) \\ &= y_r + (a+b)hf + mbh^2 F_1 + \frac{m^2 bh^3}{2} F_2 + O(h^4). \end{aligned} \quad (10.18)$$

Comparing coefficients of h and h^2 in equations (10.17) and (10.18), we get

$$\left. \begin{aligned} a+b &= 1 \\ mb &= \frac{1}{2} \end{aligned} \right\} \quad (10.19)$$

Thus, there are two equations in three unknowns and so there are many solutions to equation (10.19). We choose $a = 0$, $b = 1$, and $m = \frac{1}{2}$ as one of the solution. Then equation (10.16) reduces to

$$\begin{aligned}
y_{r+1} &= y_r + K_2 \\
&= y_r + hf \left(x_r + \frac{h}{2}, y_r + \frac{K_1}{2} \right) \\
&= y_r + hf \left(x_r + \frac{h}{2}, y_r + \frac{h}{2} f(x_r, y_r) \right),
\end{aligned} \tag{10.20}$$

which is the required second order Runge–Kutta method. We observe that this formula is nothing but modified Euler's method. If we choose $a = b = \frac{1}{2}$, $m = 1$ as the solution, then we have

$$\begin{aligned}
y_{r+1} &= y_r + \frac{1}{2}(K_1 + K_2) \\
&= y_r + \frac{h}{2}[f(x_r, y_r) + f(x_r + h, y_r + hf(x_r, y_r))],
\end{aligned}$$

which is nothing but improved Euler's method.

Third Order Runge–Kutta Method

We define

$$\left. \begin{array}{l} K_1 = hf(x_r, y_r), \\ K_2 = hf(x_r + mh, y_r + mK_1), \\ K_3 = hf(x_r + nh, y_r + nK_2). \end{array} \right\} \tag{10.21}$$

We want to obtain an expression of the form

$$y_{r+1} = y_r + aK_1 + bK_2 + cK_3. \tag{10.22}$$

If we put

$$\begin{aligned}
F_1 &= f_x + ff_y, \\
F_2 &= f_{xx} + 2ff_{xy} + f^2 f_{yy},
\end{aligned}$$

then the left-hand side of equation (10.22) is

$$y_{r+1} = y_r + hf + \frac{h^2}{2!} F_1 + \frac{h^3}{3!} (F_2 + F_1 f_y) + O(h^4). \tag{10.23}$$

Also,

$$\begin{aligned}
K_2 &= h \left[f(x_r, y_r) + mh f_x + mh f_y f + \frac{m^2 h^2}{2} f_{xx} + m^2 h^2 f_{xy} f + \frac{m^2 h^2}{2} f_{yy} f + O(h^3) \right], \\
K_3 &= h \left[f(x_r, y_r) + nh f_x + nK_2 f_y + \frac{n^2 h^2}{2} f_{xx} + n^2 h K_2 f_{xy} + \frac{n^2 K_2^2}{2} f_{yy} + O(h^3) \right] \\
&= h [f(x_r, y_r) + nh f_x + nh f_y \{f(x_r, y_r) + mh f_x + mh f_y f + \frac{m^2 h^2}{2} f_{xx} + m^2 h^2 f_{xy} f \\
&\quad + \frac{m^2 h^2}{2} f_{yy} f + O(h^3)\} + \frac{n^2 h^2}{2} f_{xx} + n^2 h^2 f_{xy} \{f(x_r, y_r) + mh f_x + mh f_y f + \frac{m^2 h^2}{2} f_{xx} \\
&\quad + m^2 h^2 f_{xy} f + \frac{m^2 h^2}{2} f_{yy} f + O(h^3)\} + \frac{n^2 h^2}{2} f_{yy} \{f(x_r, y_r) + mh f_x + mh f_y f + \frac{m^2 h^2}{2} f_{xx} \\
&\quad + m^2 h^2 f_{xy} f + \frac{m^2 h^2}{2} f_{yy} f + O(h^3)\}].
\end{aligned}$$

Therefore, putting the values of K_1 , K_2 , and K_3 in equation (10.22), we get

$$y_{r+1} = y_r + h(a+b+c)f + h^2(bm+cn)F_1 + \frac{h^3}{2}[(bm^2+cn^2)F_2 + 2cmnF_1f_y] + O(h^4). \quad (10.24)$$

Comparing the coefficients of like powers of h in equations (10.23) and (10.24), we get

$$a+b+c=1,$$

$$bm+cn=\frac{1}{2},$$

$$bm^2+cn^2=\frac{1}{3},$$

$$cmn=\frac{1}{6}.$$

Thus, we obtain four equations for five unknowns. Therefore, there exist many solutions. If we take $a=\frac{1}{4}$, $b=0$, $c=\frac{3}{4}$, $n=\frac{2}{3}$, $m=\frac{1}{3}$ as the solution, then equation (10.22) becomes

$$\begin{aligned} y_{r+1} &= y_r + \frac{1}{4}(K_1 + 3K_3), \\ K_1 &= hf(x_r, y_r), \\ K_2 &= hf\left(x_r + \frac{h}{3}, y_r + \frac{K_1}{3}\right), \\ K_3 &= hf\left(x_r + \frac{2h}{3}, y_r + \frac{2K_2}{3}\right). \end{aligned} \quad (10.25)$$

Formula (10.25) is called Heun's third order formula.

If we set $a=\frac{2}{9}$, $b=\frac{1}{3}$, $c=\frac{4}{9}$, $m=\frac{1}{2}$, $n=\frac{3}{4}$, then

$$\begin{aligned} y_{n+1} &= y_r + \frac{1}{9}(2K_1 + 3K_2 + 4K_3), \\ K_1 &= hf(x_r, y_r), \\ K_2 &= hf\left(x_r + \frac{h}{2}, y_r + \frac{K_1}{2}\right), \\ K_3 &= hf\left(x_r + \frac{3h}{4}, y_r + \frac{3K_2}{4}\right) \end{aligned}$$

If we set $a=c=\frac{1}{6}$, $b=\frac{2}{3}$, $m=\frac{1}{2}$, $n=1$, then

$$\begin{aligned} y_{r+1} &= y_r + \frac{1}{6}(K_1 + 4K_2 + K_3), \\ K_1 &= hf(x_r, y_r), \\ K_2 &= hf\left(x_r + \frac{h}{2}, y_r + \frac{K_1}{2}\right), \\ K_3 &= hf(x_r + h, y_r + K_2), \end{aligned}$$

which is the most popular third order Runge–Kutta method. It is also known as Kutta's third order rule.

Fourth Order Runge-Kutta Method

Consider the initial value problem

$$y'(x) = f(x, y), y(x_0) = y_0.$$

We define

$$\begin{aligned} K_1 &= hf(x_r, y_r), \\ K_2 &= hf(x_r + mh, y_r + mK_1), \\ K_3 &= hf(x_r + nh, y_r + nK_2), \\ K_4 &= hf(x_r + ph, y_r + pK_3). \end{aligned} \quad (10.26)$$

We wish to obtain a formula of the type

$$y_{r+1} = y_r + aK_1 + bK_2 + cK_3 + dK_4. \quad (10.27)$$

Let

$$\begin{aligned} F_1 &= f_x + ff_y, \\ F_2 &= f_{xx} + 2ff_{xy} + f^2f_{yy}, \\ F_3 &= f_{xxx} + 3ff_{xxy} + 3f^2f_{xyy} + f^3f_{yyy}. \end{aligned}$$

Expanding y_{r+1} in the series, we obtain

$$y_{r+1} = y_r + hf + \frac{h^2}{2}F_1 + \frac{h^3}{3!}(F_2 + F_1f_y) + \frac{h^4}{4!}(F_3 + F_1f_y^2 + 3F_1(f_{xy} + f_{yy}f)) + O(h^5) \quad (10.28)$$

Further, using Taylor's Theorem for two variables, we have

$$\begin{aligned} K_1 &= hf(x_r, y_r), \\ K_2 &= h \left[f(x_r, y_r) + mhF_1 + \frac{m^2h^2}{2}F_2 + \frac{m^3h^3}{3!}F_3 + \dots \right], \\ K_3 &= h \left[f(x_r, y_r) + nhF_1 + \frac{h^2}{2}(n^2F_2 + 2mnF_1f_y) + \frac{h^3}{6}(n^3F_3 + 3m^2nF^2f_y + 6mn^2F_1f_y') + \dots \right], \\ K_4 &= h \left[f(x_r, y_r) + phF_1 + \frac{h^2}{2}(p^2F_2 + 2npF_1f_y) + \frac{h^3}{6}(p^3F_3 + 3n^2pF_2f_y + 6np^2F_1f_y' + 6mnpF_1f_y^2) + \dots \right]. \end{aligned}$$

Putting these values of K_1 , K_2 , K_3 , and K_4 in equation (10.27) and equating the like powers of h in the corresponding expressions for y_{r+1} , we obtain

$$\begin{aligned} a + b + c + d &= 1, \quad cmn + dnp = \frac{1}{6}, \\ bm + cn + dp &= \frac{1}{2}, \quad cmn^2 + dnp^2 = \frac{1}{8}, \\ bm^2 + cn^2 + dp^2 &= \frac{1}{3}, \quad cm^2n + dn^2p = \frac{1}{12}, \\ bm^3 + cn^3 + dp^3 &= \frac{1}{4}, \quad dmnp = \frac{1}{24}. \end{aligned} \quad (10.29)$$

Any solution of equation (10.29) will serve our purpose. Let us take $m = n = \frac{1}{2}$, $p = 1$, $a = d = \frac{1}{6}$, $b = c = \frac{1}{3}$. Then

$$\begin{aligned} K_1 &= hf(x_r, y_r), \\ K_2 &= hf\left(x_r + \frac{h}{2}, y_r + \frac{K_1}{2}\right), \\ K_3 &= hf\left(x_r + \frac{h}{2}, y_r + \frac{K_2}{2}\right), \\ K_4 &= hf(x_r + h, y_r + K_3), \end{aligned}$$

and

$$y_{r+1} = y_r + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4),$$

which is the required fourth order Runge–Kutta method.

Remark 10.1. Whenever we mention only Runge–Kutta method, we mean the Runge–Kutta method of order 4.

EXAMPLE 10.12

Apply third order Runge–Kutta method to the initial value problem

$$\frac{dy}{dx} = x^2 - y, \quad y(0) = 1,$$

over $[0, 0.2]$ taking $h = 0.1$.

Solution. Taking $h = 0.1$, we have

$$\begin{aligned} K_1 &= hf(x_0, y_0) = 0.1(x_0^2 - y_0) = 0.1(0 - 1) = -0.1, \\ K_2 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{K_1}{2}\right) = 0.1\left[\left(\frac{0.1}{2}\right)^2 - \left(1 - \frac{0.1}{2}\right)\right] \\ &= 0.1[0.0025 - 0.95] = -0.09475, \\ K_3 &= hf(x_0 + h, y_0 + K_2) = 0.1\left[(0.1)^2 - (1 - 0.09475)\right] \\ &= 0.1[0.01 - 0.90525] = -0.089525. \end{aligned}$$

Then, by third order Runge–Kutta method,

$$\begin{aligned} y_1 &= y(0.1) = y_0 + \frac{1}{6}[K_1 + 4K_2 + K_3] \\ &= 1 + \frac{1}{6}[-0.1 - 4(-0.09475) - 0.089525] \\ &= 0.905245833. \end{aligned}$$

To find $y(0.2)$, we have

$$\begin{aligned} K_1 &= hf(x_1, y_1) = 0.1(x_1^2 - y_1) \\ &= 0.1[(0.1)^2 - 0.905245833] = -0.089524, \\ K_2 &= hf\left(x_1 + \frac{h}{2}, y_1 + \frac{K_1}{2}\right) \\ &= 0.1\left[\left(0.1 + \frac{0.1}{2}\right)^2\right] - \left(0.905245833 - \frac{0.089524}{2}\right) \end{aligned}$$

$$\begin{aligned}
&= 0.083798383, \\
K_3 &= hf(x_1 + h, y_1 + K_2) \\
&= 0.1[(0.1 + 0.1)^2 - (0.905245833 - 0.083798383)] \\
&= -0.078144745.
\end{aligned}$$

Then,

$$\begin{aligned}
y_2 &= y(0.2) = y_1 + \frac{1}{6}[K_1 + 4K_2 + K_3] \\
&= 0.905245833 + \frac{1}{6}[-0.089524 - 4(0.083798383) - 0.078144745] \\
&= 0.821435453.
\end{aligned}$$

EXAMPLE 10.13

Use Runge–Kutta method to solve $y' = x + y$, $y(0) = 1$, for $x = 0.1$.

Solution. Taking $h = 0.1$, we obtain

$$\begin{aligned}
K_1 &= hf(x_0, y_0) = 0.1(x_0 + y_0) = 0.1(0 + 1) = 0.1, \\
K_2 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{K_1}{2}\right) = 0.1\left(x_0 + \frac{h}{2} + y_0 + \frac{K_1}{2}\right) \\
&= 0.1(0 + 0.05 + 1 + 0.05) = 0.11, \\
K_3 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{K_2}{2}\right) = 0.1\left(0 + 0.05 + 1 + \frac{0.11}{2}\right) = 0.1105, \\
K_4 &= hf(x_0 + h, y_0 + K_3) = 0.1(0 + 0.1 + 1 + 0.1105) = 0.12105.
\end{aligned}$$

Therefore,

$$\begin{aligned}
y_1 &= y(0.1) = y_0 + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4) \\
&= 1 + \frac{1}{6}(0.1 + 0.22 + 0.2210 + 0.12105) = 0.11034167.
\end{aligned}$$

EXAMPLE 10.14

Apply fourth order Runge–Kutta method to

$$\frac{dy}{dx} = 3x + \frac{1}{2}y, \quad y(0) = 1$$

to determine $y(0.1)$ and $y(0.2)$ correct to four decimal places.

Solution. Taking $h = 0.1$, we have

$$\begin{aligned}
K_1 &= hf(x_0, y_0) = 0.1\left(0 + \frac{1}{2}\right) = 0.05, \\
K_2 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{K_1}{2}\right) = 0.1\left[3(0 + 0.05) + \frac{1}{2}(1 + 0.025)\right] = 0.06625,
\end{aligned}$$

$$\begin{aligned}
 K_3 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{K_2}{2}\right) = 0.1\left[3(0 + 0.05) + \frac{1}{2}\left(1 + \frac{0.06625}{2}\right)\right] = 0.06665625, \\
 K_4 &= hf(x_0 + h, y_0 + K_3) = 0.1[3(0 + 0.1) + \frac{1}{2}(1 + 0.06665625)] \\
 &= 0.1[0.3 + 0.533328125] = 0.0833328125.
 \end{aligned}$$

Hence,

$$\begin{aligned}
 y_1 &= y_0 + \frac{1}{6}[K_1 + 2K_2 + 2K_3 + K_4] \\
 &= 1 + \frac{1}{6}[0.05 + 2(0.0625) + 2(0.06665625) + 0.0833328125] \\
 &= 1.06652421875 \approx 1.0665.
 \end{aligned}$$

To find $y(0.2)$, we note that

$$\begin{aligned}
 K_1 &= hf(x_1, y_1) = 0.1[3(0.1) + \frac{1}{2}(1.066524)] = 0.0833262, \\
 K_2 &= hf\left(x_1 + \frac{h}{2}, y_1 + \frac{K_1}{2}\right) \\
 &= 0.1\left[3(0.1 + 0.05) + \frac{1}{2}\left(1.066524 + \frac{0.0833262}{2}\right)\right] \\
 &= 0.100409515, \\
 K_3 &= hf\left(x_1 + \frac{h}{2}, y_1 + \frac{K_2}{2}\right) \\
 &= 0.1\left[3(0.1 + 0.05) + \frac{1}{2}\left(1.066524 + \frac{0.100409515}{2}\right)\right] \\
 &= 0.100836437, \\
 K_4 &= hf(x_1 + h, y_1 + K_3) \\
 &= 0.1\left[3(0.1 + 0.1) + \frac{1}{2}\left(1.066524 + \frac{0.100836437}{2}\right)\right] \\
 &= 0.11584711.
 \end{aligned}$$

Hence,

$$\begin{aligned}
 y(0.2) &= y_1 + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4) \\
 &= 1.06652422 + \frac{1}{6}[0.0833262 + 2(0.100409515) + 2(0.100836437) + 0.11584711] \\
 &= 1.166801756 \approx 1.1668.
 \end{aligned}$$

EXAMPLE 10.15

Apply the fourth order Runge–Kutta method to solve

$$\frac{dy}{dx} = x^2 + y^2, \quad y(0) = 1.$$

Take step size $h = 0.1$ and determine approximations to $y(0.1)$ and $y(0.2)$ correct to four decimal places.

Solution. Taking $h = 0.1$, we have

$$\begin{aligned} K_1 &= hf(x_0, y_0) = 0.1(0+1) = 0.1, \\ K_2 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{K_1}{2}\right) = 0.1\left[\left(0.05^2 + \left(1 + \frac{0.1}{2}\right)^2\right)\right] \\ &= 0.1105, \\ K_3 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{K_2}{2}\right) = 0.1\left[\left(0.05^2 + \left(1 + \frac{0.1105}{2}\right)^2\right)\right] \\ &= 0.111605256, \\ K_4 &= hf(x_0 + h, y_0 + K_3) = 0.1[(0.1)^2 + (1 + 0.111605256)^2] \\ &= 0.124566624. \end{aligned}$$

Therefore,

$$\begin{aligned} y_1 &= y(0.1) = y_0 + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4) \\ &= 1 + \frac{1}{6}(0.1 + 2(0.1105) + 2(0.111605256) + 0.124566624) \\ &= 1.111462856 \approx 1.11146. \end{aligned}$$

To find $y(0.2)$, we have

$$\begin{aligned} K_1 &= hf(x_1, y_1) = 0.1[x_1^2 + y_1^2] \\ &= 0.1[(0.1)^2 + (1.1114628)^2] = 0.124534956, \\ K_2 &= hf\left(x_1 + \frac{h}{2}, y_1 + \frac{K_1}{2}\right) \\ &= 0.1\left[\left(0.1 + \frac{0.1}{2}\right)^2 + \left(1.1114628 + \frac{0.124534956}{2}\right)^2\right] \\ &= 0.1400142, \\ K_3 &= hf\left(x_1 + \frac{h}{2}, y_1 + \frac{K_2}{2}\right) \\ &= 0.1\left[0.0225 + \left(1.1114628 + \frac{0.1400142}{2}\right)^2\right] \\ &= 0.1418371125, \\ K_4 &= hf(x_1 + h, y_1 + K_3) \\ &= 0.1[(0.2)^2 + (1.1114628 + 0.1418371125)^2] \\ &= 0.161076063. \end{aligned}$$

Hence,

$$\begin{aligned}y_2 &= y(0.2) = y_1 + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4) \\&= 1.11142856 + \frac{1}{6}[0.124534956 + 2(0.1400142) + 2(0.1418371125) + 0.161076063] \\&= 1.2529808 \approx 1.2530.\end{aligned}$$

EXAMPLE 10.16

Using Runge–Kutta method of fourth order solve for y at $x = 1.2, 1.4$ from the equation $\frac{dy}{dx} = \frac{2xy + e^x}{x^2 + xe^x}$ with $x_0 = 1, y_0 = 0$.

Solution. We have

$$\frac{dy}{dx} = f(x, y) = \frac{2xy + e^x}{x^2 + xe^x}, \quad y(1) = 0.$$

Thus, $x_0 = 1, y_0 = 0$ and we take $h = 0.2$. Then,

$$\begin{aligned}k_1 &= hf(x_0, y_0) = 0.2 \left(\frac{e^1}{1+e^1} \right) = 0.2 \left(\frac{2.71828}{3.71828} \right) = 0.1462 \\k_2 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{k_1}{2}\right) = 0.2[f(1.1, 0.0731)] = 0.2 \left[\frac{0.161 + e^{1.1}}{1.21 + 1.1(e^{1.1})} \right] = 0.2 \left[\frac{3.1652}{4.5146} \right] = 0.1402 \\k_3 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{k_2}{2}\right) = hf(1.1, 0.0701) \\&= 0.2 \left[\frac{0.15422 + 3.0042}{1.21 + 1.1(e^{1.1})} \right] = 0.2 \left[\frac{3.1584}{4.5146} \right] = 0.1399 \\k_4 &= hf(x_0 + h, y_0 + k_3) = hf(1.2, 0.1399) \\&= 0.2 \left[\frac{0.3358 + 3.3201}{1.44 + 1.2(e^{1.2})} \right] = 0.2 \left[\frac{3.6559}{5.4241} \right] = 0.1348.\end{aligned}$$

Therefore,

$$\begin{aligned}y(1.2) &= y_0 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\&= 0 + \frac{1}{6}[0.146 + 2(0.1402) + 2(0.1399) + 0.1348] = 0.1402.\end{aligned}$$

Now $x_0 = 1.2, y_0 = 0.1402, h = 0.2$. Calculate as above k_1, k_2, k_3 , and k_4 and then find

$$\begin{aligned}y(1.4) &= y_0 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\&= 0.1402 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4).\end{aligned}$$

It will be approximately 0.264.

6. Runge–Kutta Method for System of First Order Equations

Consider the system of equations

$$y' = F(x, y, z), \quad z' = G(x, y, z),$$

with the initial conditions $y(x_0) = y_0$ and $z(x_0) = z_0$. Then the Runge–Kutta method for the system becomes

$$y_{r+1} = y_r + \frac{1}{6}[K_1 + 2K_2 + 2K_3 + K_4],$$

$$z_{r+1} = z_r + \frac{1}{6}[L_1 + 2L_2 + 2L_3 + L_4],$$

where

$$\begin{aligned} K_1 &= hF(x_r, y_r, z_r), & L_1 &= hG(x_r, y_r, z_r) \\ K_2 &= hF\left(x_r + \frac{h}{2}, y_r + \frac{K_1}{2}, z_r + \frac{L_1}{2}\right), & L_2 &= hG\left(x_r + \frac{h}{2}, y_r + \frac{K_1}{2}, z_r + \frac{L_1}{2}\right) \\ K_3 &= hF\left(x_r + \frac{h}{2}, y_r + \frac{K_2}{2}, z_r + \frac{L_2}{2}\right), & L_3 &= hG\left(x_r + \frac{h}{2}, y_r + \frac{K_2}{2}, z_r + \frac{L_2}{2}\right) \\ K_4 &= hF(x_r + h, y_r + K_3, z_r + L_3), & L_4 &= hG(x_r + h, y_r + K_3, z_r + L_3). \end{aligned}$$

EXAMPLE 10.17

Solve

$$\begin{aligned} y' &= x + z, & y(0) &= 0 \\ z' &= x - y, & z(0) &= 1 \end{aligned}$$

for $x = 0.1$ and $x = 0.2$ by Runge–Kutta method.

Solution. We have $h = 0.1$ and

$$F(x, y, z) = x + z, \quad G(x, y, z) = x - y.$$

Then,

$$\begin{aligned} K_1 &= hF(x_0, y_0, z_0) = 0.1(0 + 1) = 0.1, & L_1 &= hG(x_0, y_0, z_0) = 0.1(0 - 0) = 0 \\ K_2 &= hF\left(x_0 + \frac{h}{2}, y_0 + \frac{K_1}{2}, z_0 + \frac{L_1}{2}\right), & L_2 &= hG\left(x_0 + \frac{h}{2}, y_0 + \frac{K_1}{2}, z_0 + \frac{L_1}{2}\right) \\ &= 0.1\left[\frac{0.1}{2} + 1 + \frac{0}{2}\right] = 0.105, & &= 0.1\left[\frac{0.1}{2} - \frac{0.1}{2}\right] = 0 \\ K_3 &= hF\left(x_0 + \frac{h}{2}, y_0 + \frac{K_2}{2}, z_0 + \frac{L_2}{2}\right), & L_3 &= hG\left(x_0 + \frac{h}{2}, y_0 + \frac{K_2}{2}, z_0 + \frac{L_2}{2}\right) \\ &= 0.1\left[\frac{0.1}{2} + 1\right] = 0.105, & &= 0.1\left[\frac{0.1}{2} - \frac{0.105}{2}\right] = 0.00025 \\ K_4 &= hF(x_0 + h, y_0 + K_3, z_0 + L_3), & L_4 &= hG(x_0 + h, y_0 + K_3, z_0 + L_3) \\ &= 0.1[0.1 + 1 + 0.00025], & &= 0.1[0.1 - 0.105] \\ &= 0.110025., & &= 0.0205. \end{aligned}$$

Thus,

$$\begin{aligned} y(0.1) &= y_0 + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4) \\ &= \frac{1}{6}[0.1 + 2(0.105) + 2(0.105) + 0.110025] = 0.105004, \end{aligned}$$

$$\begin{aligned} z(0.1) &= z_0 + \frac{1}{6}(L_1 + 2L_2 + 2L_3 + L_4) \\ &= 1 + \frac{1}{6}[2(0.00025) - 0.0205] = 0.99667. \end{aligned}$$

Now to find $y(0.2)$ and $z(0.2)$, we have

$$\begin{aligned} K_1 &= hF(x_1, y_1, z_1), & L_1 &= hG(x_1, y_1, z_1) \\ &= 0.1[0.1 + 0.99667] = 0.109667, & &= 0.1[0.1 - 0.105004] = -0.0005004 \\ K_2 &= hF\left(x_1 + \frac{h}{2}, y_1 + \frac{K_1}{2}, z_1 + \frac{L_1}{2}\right), & L_2 &= hG\left(x_1 + \frac{h}{2}, y_1 + \frac{K_1}{2}, z_1 + \frac{L_1}{2}\right) \\ &= 0.1[0.150 + (0.99667 - 0.00025002)], & &= 0.1[0.150 - (0.105004 + 0.054816)] \\ &= 0.114642, & &= -0.0009837 \\ K_3 &= hF\left(x_1 + \frac{h}{2}, y_1 + \frac{K_2}{2}, z_1 + \frac{L_2}{2}\right), & L_3 &= hG\left(x_1 + \frac{h}{2}, y_1 + \frac{K_2}{2}, z_1 + \frac{L_2}{2}\right) \\ &= 0.1[0.150 + 0.99667 - 0.0004918], & &= 0.1[0.15 - (0.105004 + 0.057321)] \\ &= 0.101167, & &= -0.0012325 \\ K_4 &= hF(x_1 + h, y_1 + K_3, z_1 + L_3), & L_4 &= hG(x_1 + h, y_1 + K_3, z_1 + L_3) \\ &= 0.1[0.150 + 0.99667 - 0.0012325], & &= 0.1[0.15 - (0.105004 + 0.101167)] \\ &= 0.11454375. & &= -0.0056174. \end{aligned}$$

Hence,

$$\begin{aligned} y(0.2) &= y_1 + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4) \\ &= 0.105004 + \frac{1}{6}[0.109667 + 2(0.114642) + 2(0.101167) + 0.11454375] \\ &= 0.2143073 \end{aligned}$$

and

$$\begin{aligned} z(0.2) &= z_1 + \frac{1}{6}(L_1 + 2L_2 + 2L_3 + L_4) \\ &= 0.99667 + \frac{1}{6}[-0.0005004 + 2(-0.0009837) + 2(-0.0012325) + (-0.0056174)] \\ &= 0.99591. \end{aligned}$$

7. Runge–Kutta Method for Higher Order Differential Equations

Since the higher order differential equations can be converted into a set of first order differential equations, therefore these equations can be solved by Runge–Kutta method. We illustrate the method in the form of the following example:

EXAMPLE 10.18

Using Runge–Kutta method, solve the differential equation

$$y'' = \frac{x^2 - y^2}{1 + y'^2}, \quad y(0) = 1, \quad y'(0) = 0$$

for $x = 0.5$ and $x = 1$.

Solution. We are given that

$$y'' = \frac{x^2 - y^2}{1 + y'^2}, \quad y(0) = 1, y'(0) = 0.$$

Putting $y' = z$, the given equation is equivalent to

$$y' = z = F(x, y, z),$$

$$z' = \frac{x^2 - y^2}{1 + y'^2} = G(x, y, z).$$

Now y and z can be determined by

$$y_{r+1} = y_r + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4),$$

$$z_{r+1} = z_r + \frac{1}{6}(L_1 + 2L_2 + 2L_3 + L_4).$$

Taking $h = 0.5$, we have

$$\begin{aligned} K_1 &= hF(x_0, y_0, z_0) = 0.5(0) = 0, & L_1 &= hG(x_0, y_0, z_0) = 0.5\left(\frac{-1}{1}\right) = 0.5 \\ K_2 &= hF\left(x_0 + \frac{h}{2}, y_0 + \frac{K_1}{2}, z_0 + \frac{L_1}{2}\right), & L_2 &= hG\left(x_0 + \frac{h}{2}, y_0 + \frac{K_1}{2}, z_0 + \frac{L_1}{2}\right) \\ &= 0.5\left(-\frac{0.5}{2}\right) = -0.125, & &= 0.5\left[\frac{(0.25)^2 - 1}{1 + \left(-\frac{0.5}{2}\right)^2}\right] = -0.4412 \\ K_3 &= hF\left(x_0 + \frac{h}{2}, y_0 + \frac{K_2}{2}, z_0 + \frac{L_2}{2}\right), & L_3 &= hG\left(x_0 + \frac{h}{2}, y_0 + \frac{K_2}{2}, z_0 + \frac{L_2}{2}\right) \\ &= 0.5\left[\frac{-0.4412}{2}\right] = -0.110294, & &= \frac{0.5[0.0625 - (1 - 0.0625)^2]}{1 + (0.2206)^2} = -0.3892 \\ K_4 &= hF(x_0 + h, y_0 + K_3, z_0 + L_3), & L_4 &= hG(x_0 + h, y_0 + K_3, z_0 + L_3) \\ &= 0.5(-0.3892) = -0.19460, & &= 0.5\left[\frac{0.25 + (1 - 0.11029)^2}{1 + (-0.3892)^2}\right] = 0.2351. \end{aligned}$$

Therefore,

$$\begin{aligned} y(0.5) &= y_0 + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4) \\ &= 1 + \frac{1}{6}(0 - 0.250 - 0.22058 - 0.1946) = -0.88914, \\ z(0.5) &= z_0 + \frac{1}{6}(L_1 + 2L_2 + 2L_3 + L_4) \\ &= \frac{1}{6}(0.5 - 0.8826 - 0.7784 - 0.2351) = -0.39935. \end{aligned}$$

To find $y(0.2)$, we note that $x_1 = 0.5$, $y_1 = 0.88914$, and $z_1 = 0.39935$. Therefore, proceeding as above, we have

$$\begin{aligned}
K_1 &= hF(x_1, y_1, z_1) = -0.199675, & L_1 &= hG(x_1, y_1, z_1) = 0.23314 \\
K_2 &= -0.257936, & L_2 &= -0.0238965 \\
K_3 &= -0.205632, & L_3 &= -0.00656937 \\
K_4 &= -0.2029425, & L_4 &= 0.228729.
\end{aligned}$$

Thus,

$$y_2 = y(1) = y_1 + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4) = 0.667517.$$

10.4 MULTISTEP METHODS

As we have already pointed out that in a multistep method, y_{n+1} appears as a function of several values $y_n, y_{n-1}, \dots, y_{n-p}$. We first derive explicit multistep methods.

1. Explicit Multistep Methods (Adams–Basforth and Nystrom's Methods)

Suppose that the function $f(x, y)$ is approximated by Newton's backward difference formula of degree p . Thus, taking $x = x_n + th$, we get

$$f(x, y) = f_n + t\nabla f_n + \frac{t(t+1)}{2!} \nabla^2 f_n + \frac{t(t+1)(t+2)}{3!} \nabla^3 f_n + \dots$$

and so

$$\begin{aligned}
\int_{x_{n-i}}^{x_{n+1}} f(x, y) dx &= h \int_{-i}^1 \left[f_n + t\nabla f_n + \frac{t(t+1)}{2!} \nabla^2 f_n + \frac{t(t+1)(t+2)}{3!} \nabla^3 f_n + \dots \right] dt \\
&= h \left[tf_n + \frac{t^2}{2} \nabla f_n + \frac{t^2}{2} \left(\frac{t}{3} + \frac{1}{2} \right) \nabla^2 f_n + \frac{t^2}{3} \left(\frac{t^2}{4} + t + 1 \right) \nabla^3 f_n + \dots \right]_{-i}^1
\end{aligned} \tag{10.30}$$

with an error term given by

$$T = h^{p+2} \int_{-i}^1 \frac{t(t+1)\dots(t+p)}{(p+1)!} f^{(p+1)}(\xi) dt$$

for $x_{n-i} < \xi < x_{n+1}$. But

$$\begin{aligned}
\int_{x_{n-i}}^{x_{n+1}} f(x, y) dx &= \int_{x_{n-i}}^{x_{n+1}} y'(x) dx = y(x_{n+1}) - y(x_{n-i}) \\
&= y_{n+1} - y_{n-i}.
\end{aligned}$$

Hence, equation (10.30) reduces to

$$y_{n+1} = y_{n-i} + h \left[tf_n + \frac{t^2}{2} \nabla f_n + \frac{t^2}{2} \left(\frac{t}{3} + \frac{1}{2} \right) \nabla^2 f_n + \frac{t^2}{3} \left(\frac{t^2}{4} + t + 1 \right) \nabla^3 f_n + \dots \right]_{-i}^1 \tag{10.31}$$

For $i = 0$, expression (10.31) becomes

$$y_{n+1} = y_n + h \left[f_n + \frac{1}{2} \nabla F_n + \frac{5}{12} \nabla^2 f_n + \frac{3}{8} \nabla^3 f_n + \frac{251}{720} \nabla^4 f_n + \dots \right]. \tag{10.32}$$

The procedure using formula (10.32) is called Adams–Basforth method. Formula (10.32) requires the values of $f(x, y)$ at $(p+1)$ points $x_n, x_{n-1}, \dots, x_{n-p}$. The most important case is $p = 3$ for which

$$y_{n+1} = y_n + h \left[f_n + \frac{1}{2} \nabla f_n + \frac{5}{12} \nabla^2 f_n + \frac{3}{8} \nabla^3 f_n \right] + T, \quad (10.33)$$

where

$$T = \frac{251}{720} h^5 f^{(iv)}(\xi).$$

The scheme for the application of formula (10.33) is therefore

x	y	f	∇	∇^2	∇^3
x_{n-3}	y_{n-3}	f_{n-3}			
x_{n-2}	y_{n-2}	f_{n-2}	∇f_{n-2}	$\nabla^2 f_{n-1}$	$\nabla^3 f_n$
x_{n-1}	y_{n-1}	f_{n-1}	∇f_{n-1}	$\nabla^2 f_n$	
x_n	y_n	f_n	∇f_n		

Since

$$\begin{aligned}\nabla f_n &= f_n - f_{n-1}, \\ \nabla^2 f_n &= f_n - 2f_{n-1} + f_{n-2}, \\ \nabla^3 f_n &= f_n - 3f_{n-1} + 3f_{n-2} - f_{n-3},\end{aligned}$$

formula (10.33) reduces to

$$y_{n+1} = y_n + \frac{h}{24} [55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3}] + T. \quad (10.34)$$

For $i = 1$, expression (10.31) yields

$$\begin{aligned}y_{n+1} &= y_{n-1} + h \left[2f_n + \frac{1}{3} \nabla^2 f_n + \frac{1}{3} \nabla^3 f_n + \frac{29}{90} \nabla^4 f_n + \frac{14}{45} \nabla^5 f_n + \dots \right] \\ &= y_{n-1} + h \left[2f_n + \frac{1}{3} (\nabla^2 f_n + \nabla^3 f_n + \nabla^4 f_n + \nabla^5 f_n) - \frac{1}{90} (\nabla^4 f_n + 2\nabla^5 f_n + \dots) \right]\end{aligned} \quad (10.35)$$

which is known as Nystrom's formula. For $p = 3$, equation (10.35) becomes

$$y_{n+1} = y_{n-1} + h \left[2f_n + \frac{1}{3} (\nabla^2 f_n + \nabla^3 f_n) \right] + T. \quad (10.36)$$

The scheme for the application of this formula is same as given for formula (10.33).

For $i = 3, p = 3$, expression (10.31) yields

$$y_{n+1} = y_{n-3} + \frac{4h}{3} (2f_n - f_{n-1} + f_{n-2}) + T, \quad (10.37)$$

where

$$T = \frac{14}{15} h^5 f^{(iv)}(\xi).$$

The formulae obtained above are called open type formulae or extrapolation formulae.

2. Implicit Multistep Methods (Adams–Moulton and Milne–Simpson Formulae)

If $f(x, y)$ is approximated by Newton's backward difference formula based at x_{n+1} instead of x_n so that $x = x_{n+1} + (t - 1)h$, then we get

$$\begin{aligned} y_{n+1} &= y_{n-i} + h \int_{-i}^1 [f_{n+1} + (t-1)\nabla f_{n+1} + \frac{(t-1)t}{2!} \nabla^2 f_{n+1} + \dots] dt \\ &= y_{n-i} + h \left[tf_{n+1} + t \left(\frac{t}{2} - 1 \right) \nabla f_{n+1} + \frac{t^2}{2!} \left(\frac{t}{3} - \frac{1}{2} \right) \nabla^2 f_{n+1} + \frac{t^3}{3!} \left(\frac{t^2}{4} - \frac{1}{2} \right) \nabla^3 f_{n+1} + \dots \right]_{-i}^1 + T, \end{aligned} \quad (10.38)$$

where

$$T = h^{p+2} \int_{-i}^1 \frac{(t-1)t(t+1)\dots(t+p-1)}{(p+1)!} f^{(p+1)}(\xi) d\xi, \quad x_{n-1} < \xi < x_{n+1}.$$

For $i = 0$, expression (10.38) yields

$$y_{n+1} = y_n + h \left[f_{n+1} - \frac{1}{2} \nabla f_{n+1} - \frac{1}{12} \nabla^2 f_{n+1} - \frac{1}{24} \nabla^3 f_{n+1} - \frac{19}{720} \nabla^4 f_{n+1} - \frac{3}{160} \nabla^5 f_{n+1} + \dots \right] + T, \quad (10.39)$$

which is known as Adams–Moulton formula.

For $p = 1$, we get

$$y_{n+1} = y_n + \frac{h}{2} [f_n + f_{n+1}]. \quad (\text{trapezoidal rule})$$

For $p = 3$, the formula reduces to

$$y_{n+1} = y_n + h \left[f_{n+1} - \frac{1}{2} \nabla f_{n+1} - \frac{1}{12} \nabla^2 f_{n+1} - \frac{1}{24} \nabla^3 f_{n+1} + T \right], \quad (10.40)$$

where

$$T = -\frac{19}{720} h^5 f^{(iv)}(\xi).$$

If we substitute the values of ∇f_{n+1} , $\nabla^2 f_{n+1}$, $\nabla^3 f_{n+1}$ in terms of ordinates, we get

$$y_{n+1} = y_n + \frac{h}{24} [9f_{n+1} + 19f_n - 5f_{n-1} + f_{n-2}] + T, \quad (10.41)$$

where

$$T = -\frac{19}{720} h^5 f^{(iv)}(\xi).$$

For $i = 1$, expression (10.38) yields

$$y_{n+1} = y_{n-1} + h \left[2f_{n+1} - 2\nabla f_{n+1} + \frac{1}{3} \nabla^2 f_{n+1} - \frac{1}{90} \nabla^4 f_{n+1} + \dots \right] + T, \quad (10.42)$$

which is called Milne–Simpson formula. When $p = 3$, this reduces to

$$y_{n+1} = y_{n-1} + h \left[2f_{n+1} - 2\nabla f_{n+1} + \frac{1}{3} \nabla^2 f_{n+1} \right] + T, \quad (10.43)$$

where

$$T = -\frac{1}{90} h^5 f^{(iv)}(\xi).$$

In terms of ordinates, equation (10.43) takes the form

$$y_{n+1} = y_{n-1} + \frac{h}{3} [f_{n+1} + 4f_n + f_{n-1}] + T, \quad (10.44)$$

where

$$T = -\frac{1}{90} h^5 f^{(iv)}(\xi).$$

The formulae obtained above are called closed type formulae or interpolation formulae.

EXAMPLE 10.19

Use Adams–Basforth formula to find $y(0.4)$ for the equation $\frac{dy}{dx} = \frac{1}{2}xy$ using the data

x:	0	0.1	0.2	0.3
y:	1	1.0025	1.0101	1.0228

Improve the value of $y(0.4)$ using Adams–Moulton formula.

Solution. We have

$$\frac{dy}{dx} = \frac{1}{2}xy.$$

Also $y(0) = 1$ (given) and $f(x, y) = \frac{1}{2}xy$. So taking $x_0 = 0, x_1 = 0.1, x_2 = 0.2, x_3 = 0.3, x_4 = 0.4$, we have $y_0 = 1$,

$y_1 = 1.0025, y_2 = 1.0101, y_3 = 1.0228$.

Further,

$$\begin{aligned} f_0 &= 0, \\ f_1 &= \frac{1}{2}(0.1)(1.0025) = 0.05012, \\ f_2 &= \frac{1}{2}(0.2)(1.0101) = 0.10101, \\ f_3 &= \frac{1}{2}(0.3)(1.0228) = 0.15342. \end{aligned}$$

Therefore, the Adams–Basforth formula

$$y_{n+1} = y_n + \frac{h}{24} [55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3}]$$

gives

$$\begin{aligned} y_4 &= y_3 + \frac{0.1}{24} [55f_3 - 59f_2 + 37f_1 - 9f_0] \\ &= 1.0228 + \frac{0.1}{24} [55(0.15342) - 59(0.10101) + 37(0.05012) - 9(0)] \\ &= 1.040853. \end{aligned}$$

Therefore,

$$f(x_4, y_4) = \frac{1}{2}x_4 y_4 = \frac{1}{2}(0.4)(1.040853) = 0.20817.$$

Now, by Adams–Moulton formula, we have

$$\begin{aligned}
 y_4 &= y_3 + \frac{h}{24} [9f_4 + 19f_3 - 5f_2 + f_1] \\
 &= 1.0228 + \frac{0.1}{24} [9(0.20817) + 19(0.15342) - 5(0.10101) + 0.05012] \\
 &= 1.040856.
 \end{aligned}$$

EXAMPLE 10.20

Using Adams–Moulton method, obtain the solution of $\frac{dy}{dx} = x^2 y + x^2$ at $x=1.4$, given the values

$x:$	1	1.1	1.2	1.3
$y:$	1	1.233	1.548488	1.978921

Solution. We are given that

$$\frac{dy}{dx} = f(x, y) = x^2 y + x^2.$$

Taking $x_0=1$, $x_1=1.1$, $x_2=1.2$, and $x_3=1.3$, we have

$$y_0=1, y_1=1.233, y_2=1.548488, \text{ and } y_3=1.978921.$$

Further,

$$\begin{aligned}
 f_0 &= f(x_0, y_0) = x_0^2 y_0 + x_0^2 = 1+1=2, \\
 f_1 &= f(x_1, y_1) = x_1^2 y_1 + x_1^2 = (1.1)^2(1.233) + (1.1)^2 = 2.70193, \\
 f_2 &= f(x_2, y_2) = x_2^2 y_2 + x_2^2 = (1.2)^2(1.548488) + (1.2)^2 = 3.669822, \\
 f_3 &= f(x_3, y_3) = x_3^2 y_3 + x_3^2 = (1.3)^2(1.978921) + (1.3)^2 = 5.034376.
 \end{aligned}$$

To use Adams–Moulton method, we shall need f_4 and to find f_4 we require y_4 . We calculate y_4 by using Adams–Basforth formula

$$y_{n+1} = y_n + \frac{h}{24} [55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3}].$$

We have,

$$\begin{aligned}
 y_4 &= y_3 + \frac{0.1}{24} [55f_3 - 59f_2 + 37f_1 - 9f_0] \\
 &= 1.978921 + \frac{0.1}{24} [55(5.034376) - 59(3.669822) + 37(2.70193) - 9(2)] \\
 &= 2.57202.
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 f_4 &= f(x_4, y_4) = x_4^2 y_4 + x_4^2 \\
 &= (1.4)^2(2.57202) + (1.4)^2 = 7.0012.
 \end{aligned}$$

Now using Adams–Moulton formula, we have

$$\begin{aligned}
 y_4 &= y_3 + \frac{0.1}{24} [9f_4 + 19f_3 - 5f_2 + f_1] \\
 &= 1.978921 + \frac{0.1}{24} [9(7.00120) + 19(5.034376) - 5(3.669822) + 2.70193] \\
 &= 2.57482.
 \end{aligned}$$

EXAMPLE 10.21

Solve the initial value problem

$$\frac{dy}{dx} = x - y^2, \quad y(0) = 1$$

to find $y(0.4)$ by Adam's method. Starting solutions required are to be obtained using Runge–Kutta method of order 4, using step value $h = 0.1$.

Solution. We have

$$\frac{dy}{dx} = x - y^2, \quad y(0) = 1.$$

Therefore, taking step value $h = 0.1$, we get

$$\begin{aligned} k_1 &= hf(x_0, y_0) = 0.1(0 - 1) = -0.1, \\ k_2 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{k_1}{2}\right) = 0.1f(0.05, 0.95) = -0.08525, \\ k_3 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{k_2}{2}\right) = 0.1f(0.05, 0.9574) = -0.0867, \\ k_4 &= h(x_0 + h, y_0 + k_3) = 0.1f(0.1, 0.9137) = -0.07341. \end{aligned}$$

Hence,

$$\begin{aligned} y_1 &= y(0.1) = y_0 + \frac{1}{6}[k_1 + 2k_2 + 2k_3 + k_4] \\ &= 1 + \frac{1}{6}[-0.1 + 2(-0.08525) + 2(-0.0867) + 0.07341] \\ &= 0.9117. \end{aligned}$$

Similarly,

$$\begin{aligned} y_2 &= y(0.2) = 0.8494, \\ y_3 &= y(0.3) = 0.8061. \end{aligned}$$

Thus,

$$\begin{aligned} f_0 &= x_0 - y_0^2 = -1 \\ f_1 &= x_1 - y_1^2 = 0.1 - (0.9117)^2 = -0.7312, \\ f_2 &= x_2 - y_2^2 = 0.2 - (0.8494)^2 = -0.5215, \\ f_3 &= x_3 - y_3^2 = 0.3 - (0.8061)^2 = -0.3498. \end{aligned}$$

Therefore, Adams–Basforth formula

$$y_{n+1} = y_n + \frac{h}{24} [55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3}]$$

yields

$$\begin{aligned} y_4 &= y_3 + \frac{h}{24} [55f_3 - 59f_2 + 37f_1 - 9f_0] \\ &= 0.8061 + \frac{0.1}{24} [55(-0.3498) - 59(-0.5215) + 37(-0.7312) - 9(-1)] \\ &= 0.8061 - 0.02718 = 0.779. \end{aligned}$$

Therefore,

$$f_4 = x_4 - y_4^2 = 0.4 - (0.779)^2 = -0.2068.$$

Now, by Adams–Moulton formula, we have

$$\begin{aligned} y_4 &= y_3 + \frac{h}{24} [9f_4 + 19f_3 - 5f_2 + f_1] \\ &= 0.8061 + \frac{0.1}{24} [9(-0.2068) + 19(-0.3498) - 5(-0.5215) - 0.7312] \\ &= 0.77847. \end{aligned}$$

3. Milne–Simpson's Method

Consider the first order differential equation $y' = f(x, y)$. In the Milne–Simpson's method, we first extrapolate an approximate value for y_{n+1} using a rather coarse method called predictor and then we improve this value by use of a more accurate method called corrector.

We use the open type formula

$$\begin{aligned} y_{n+1} - y_{n-3} &= \frac{4h}{3}(2f_n - f_{n-1} + 2f_{n-2}) + \frac{14}{45}h^5 f^{(iv)}(\xi) \\ &= \frac{4h}{3}(2y'_n - y'_{n-1} + 2y'_{n-2}) + \frac{14}{45}h^5 f^{(v)}(\xi) \end{aligned}$$

as predictor formula and the closed type formula

$$\begin{aligned} y_{n+1} - y_{n-1} &= \frac{h}{3}(f_{n+1} + 4f_n + f_{n-1}) - \frac{h^5}{90} f^{(iv)}(\xi) \\ &= \frac{h}{3}(y'_{n+1} + 4y'_n + y'_{n-1}) - \frac{h^5}{90} y^{(v)}(\xi) \end{aligned}$$

as corrector formula. For sufficiently small value of h , the remainder term $-\frac{h^5}{90} y^{(v)}(\xi)$ is much smaller than the error term in the predictor formula and is negligible. Therefore, the corrector formula is

$$y_{n+1} = y_{n-1} + \frac{h}{3}(y'_{n+1} + 4y'_n + y'_{n-1}).$$

We find y_{n+1} from the predictor formula. From this predicted value of y_{n+1} and the differential equation, y_{n+1} is computed using a corrector formula which comes out to be a better value for y_{n+1} . We again recalculate y_{n+1} and then correct the value by using corrector. The process is repeated till the table of values is complete.

EXAMPLE 10.22

Solve the differential equation

$$y' = x^2 + y^2 - 2, \quad y(0) = 1$$

for $x = 0.3$ and 0.4 using Milne's method. The values of y for $x = -0.1, 0.1$, and 0.2 should be computed by series method.

Solution. We have

$$\begin{aligned} y' &= x^2 + y^2 - 2, \\ y'' &= 2x + 2yy', \\ y''' &= 2 + 2yy'' + 2y'^2, \\ y^{(iv)} &= 2yy''' + 6y'y''. \end{aligned}$$

Then

$$y(0) = 1 \text{ (given)}, y'_0 = -1, y''_0 = 2(-1) = -2, y'''_0 = 0, y^{(iv)}_0 = 12.$$

Thus, by series method,

$$\begin{aligned} y_1 &= y_0 + hy'_0 + \frac{h^2}{2} y''_0 + \frac{h^3}{3!} y'''_0 + \frac{h^4}{4!} y^{(iv)}_0 + \dots \\ &= 1 - 0.1 + 0.0 + \frac{0.0001}{2} + \dots \approx 0.91005, \\ y_{-1} &= y_0 - hy'_0 + \frac{h^2}{2} y''_0 - \frac{h^3}{3!} y'''_0 + \frac{h^4}{4!} y^{(iv)}_0 - \dots \\ &= 1 + 0.1 + 0.01 + 0.00005 + \dots \approx 1.11005. \end{aligned}$$

Also,

$$y_2 = y_1 + hy'_1 + \frac{h^2}{2} y''_1 + \frac{h^3}{3!} y'''_1 + \frac{h^4}{4!} y^{(iv)}_1 + \dots$$

But

$$\begin{aligned} y'_1 &= (0.1)^2 + (0.91005)^2 - 2 = -1.1618, \\ y''_1 &= 2(0.1) + 2(0.91005)(-1.1618) = -1.9146, \\ y'''_1 &= 2 + 2(0.91005)(-1.9146) + 2(-1.1618)^2 = 1.2148, \\ y^{(iv)}_1 &= 2(0.91005)(1.2148) + 6(-1.1618)(-1.9146) = 15.5573. \end{aligned}$$

Therefore,

$$y_2 = 0.91005 + 0.1(-1.1618) + \frac{0.01}{2}(-1.9146) + \frac{0.001}{6}(1.2148) + \frac{0.0001}{24}(15.5573) + \dots \approx 0.7846.$$

Then

$$y'_2 = (0.2)^2 + (0.7846)^2 - 2 = 0.04 + 0.6156 - 2 = -1.3444.$$

Using predictor, we have

$$\begin{aligned} y(0.3) &= y_3 = y_{-1} + \frac{4h}{3}(2y'_0 - y'_1 + 2y'_2) \\ &= 1.11005 + \frac{0.4}{3}[-2 + 1.1618 + 2(-1.3444)] = 0.64024. \end{aligned}$$

Now,

$$y'_3 = (0.3)^2 + (0.64024)^2 - 2 = -1.5000.$$

Using corrector, we have

$$\begin{aligned} y_3 &= y_1 + \frac{h}{3}[y'_1 + 4y'_2 + y'_3] \\ &= 0.91005 + \frac{0.1}{3}[-1.1618 + 4(-1.3444) - 1.5000] = 0.6421. \end{aligned}$$

Now using predictor, we have

$$\begin{aligned} y_4 &= y_0 + \frac{4h}{3}[2y'_1 - y'_2 + 2y'_3] \\ &= 1 + \frac{0.4}{3}[2(-1.1618) - (-1.3444) + 2(-1.5000)] = 0.4694. \end{aligned}$$

Now,

$$y'_4 = (0.4)^2 + (0.4694)^2 - 2 = -1.6197.$$

The use of corrector formula yields

$$\begin{aligned} y_4 &= y_2 + \frac{h}{3} [y'_2 + 4y'_3 + y'_4] \\ &= 0.7846 + \frac{0.1}{3} [-1.3444 + 4(-1.5000) + (-1.6197)] \approx 0.4858. \end{aligned}$$

EXAMPLE 10.23

Using Milne–Simpson's predictor–corrector formula solve $y' = 2y - y^2$, $y(0) = 1$ for $x = 0.2$ and $x = 0.25$ if

$$y(0.05) = 1.0499584, y(0.10) = 1.0996680, y(0.15) = 1.1488850.$$

Solution. We are given that

$$y_1 = 1.0499584, y_2 = 1.0996680, y_3 = 1.1488850.$$

Thus,

$$\begin{aligned} y'_1 &= 2(1.0499584) - (1.0499584)^2 = 0.9975, \\ y'_2 &= 2(1.0996680) - (1.0996680)^2 = 0.8794, \\ y'_3 &= 2(1.1488850) - (1.1488850)^2 = 0.9778. \end{aligned}$$

Therefore using predictor, we have

$$\begin{aligned} y(0.20) &= y_4 = y_0 + \frac{4h}{3} [2y'_3 - y'_2 + 2y'_1] \\ &= 1 + \frac{4(0.05)}{3} [2(0.9778) - 0.8794 + 2(0.9975)] = 1.20475. \end{aligned}$$

Now

$$y'_4 = 2(1.20475) - (1.20475)^2 = 0.9581.$$

Then corrector formula yields

$$\begin{aligned} y_4 &= y_2 + \frac{h}{3} [y'_2 + 4y'_3 + y'_4] \\ &= 1.0996680 + \frac{0.05}{3} [0.8794 + 4(0.9778) + 0.9581] = 1.19548. \end{aligned}$$

But now,

$$y'_4 = 2(1.19548) - (1.19548)^2 = 0.9618.$$

Again using predictor, we have

$$\begin{aligned} y_5 &= y_1 + \frac{4h}{3} [2y'_4 - y'_3 + 2y'_2] \\ &= 1.0499584 + \frac{4(0.05)}{3} [2(0.9618) - 0.9778 + 2(0.8794)] \\ &= 1.23026. \end{aligned}$$

But

$$y'_5 = 2(1.23026) - (1.23026)^2 = 0.9470.$$

Therefore, corrector formula yields

$$\begin{aligned}y_5 &= y_3 + \frac{h}{3} [y'_3 + 4y'_4 + y'_5] \\&= 1.1488850 + \frac{0.05}{3} [0.9778 + 4(0.9618) + 0.9470] \\&= 1.24505.\end{aligned}$$

EXAMPLE 10.24

Solve the initial value problem $\frac{dy}{dx} = 1 + xy^2$, $y(0) = 1$ for $x = 0.4$ by Milne's method, it is given that

$x:$	0.1	0.2	0.3
$y:$	1.105	1.223	1.355

Solution. The given initial value problem is

$$\begin{aligned}\frac{dy}{dx} &= 1 + xy^2, \quad y(0) = 1 \\y(0.1) &= 1.105, \quad y(0.2) = 1.223 \quad \text{and } y(0.3) = 1.355.\end{aligned}$$

We have

$$y' = 1 + xy^2.$$

Therefore,

$$\begin{aligned}y'_1 &= 1 + (0.1)(1.105)^2 = 1.1221, \\y'_2 &= 1 + (0.2)(1.223)^2 = 1.2991, \\y'_3 &= 1 + (0.3)(1.355)^2 = 1.5508.\end{aligned}$$

Using predictor, we have

$$\begin{aligned}y_4 &= y_0 + \frac{4h}{3} [2y'_1 - y'_2 + 2y'_3] \\&= 1 + \frac{4(0.1)}{3} [2(1.1221) - 1.2991 + 2(1.5508)] \\&= 1 + \frac{0.4}{3} (2.2442 - 1.2991 + 3.1016) = 1.53956.\end{aligned}$$

Now

$$y'_4 = 1 + (0.4)(1.53956)^2 = 1.9481.$$

Then the corrector formula yields

$$\begin{aligned}y_4 &= y(0.4) = y_2 + \frac{h}{3} [y'_2 + 4y'_3 + y'_4] \\&= 1.223 + \frac{0.1}{3} [1.2991 + 4(1.5508) + 1.9481] \\&= 1.223 + \frac{0.1}{3} [9.4504] = 1.5380.\end{aligned}$$

EXAMPLE 10.25

Using Runge–Kutta method of order 4, find y for $x = 0.1, 0.2, 0.3$, given that $\frac{dy}{dx} = xy + y^2$, $y(0) = 1$. Applying Milne–Simpson's method find y for $x = 0.4$.

Solution. We have

$$\frac{dy}{dx} = xy + y^2, y(0) = 1.$$

We have $x_0 = 0$, $y_0 = 1$, $h = 0.1$. Therefore,

$$\begin{aligned} K_1 &= hf(x_0, y_0) = (0.1)f(0, 1) = 0.1, \\ K_2 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{K_1}{2}\right) = (0.1)f(0.05, 1.05) \\ &= 0.1[0.05(1.05) + (1.05)^2] = 0.1155, \\ K_3 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{K_2}{2}\right) = (0.1)f(0.05, 1.0578) \\ &= 0.1[0.05(1.0578) + (1.0578)^2] = 0.1172, \\ K_4 &= hf(x_0 + h, y_0 + K_3) = (0.1)f(0.1, 1.1172) \\ &= 0.1[0.1(1.1172) + (1.1172)^2] = 0.13598. \end{aligned}$$

Therefore,

$$\begin{aligned} y(0.1) &= y_0 + \frac{1}{6}[K_1 + 2K_2 + 2K_3 + K_4] \\ &= 1 + \frac{1}{6}[0.1 + 2(0.1155) + 2(0.1172) + 0.13598] = 1.1169. \end{aligned}$$

Now $x_1 = 0.1$, $y_1 = 1.1169$, $h = 0.1$. Therefore,

$$\begin{aligned} K_1 &= hf(x_1, y_1) = (0.1)f(0.1, 1.1169) \\ &= (0.1)[0.1(1.1169) + (1.1169)^2] = 0.1359, \\ K_2 &= hf\left(x_1 + \frac{h}{2}, y_1 + \frac{K_1}{2}\right) = (0.1)f(0.15, 1.1848) \\ &= 0.1[0.15(1.1848) + (1.1848)^2] = 0.1581, \\ K_3 &= hf\left(x_1 + \frac{h}{2}, y_1 + \frac{K_2}{2}\right) = (0.1)f(0.15, 1.1959) \\ &= 0.1[0.15(1.1959) + (1.1959)^2] = 0.1610, \\ K_4 &= hf(x_1 + h, y_1 + K_3) = (0.1)f(0.2, 1.2779) \\ &= 0.1[0.2(1.2779) + (1.2779)^2] = 0.1887. \end{aligned}$$

Hence,

$$\begin{aligned} y(0.2) &= y_1 + \frac{1}{6}[K_1 + 2K_2 + 2K_3 + K_4] \\ &= 1.1169 + \frac{1}{6}[0.1359 + 2(0.1581) + 2(0.1610) + 0.1887] = 1.2774. \end{aligned}$$

Now $x_2 = 0.2$, $y_2 = 1.2774$, $h = 0.1$. Therefore,

$$\begin{aligned} K_1 &= hf(x_2, y_2) = (0.1)f(0.2, 1.2774) \\ &= (0.1)[0.2(1.2774) + (1.2774)^2] = 0.1887, \\ K_2 &= hf\left(x_2 + \frac{h}{2}, y_2 + \frac{K_1}{2}\right) = (0.1)f(0.25, 1.3718), \\ &= 0.1[0.25(1.3718) + (1.3718)^2] = 0.22248, \\ K_3 &= hf\left(x_2 + \frac{h}{2}, y_2 + \frac{K_2}{2}\right) = (0.1)f(0.25, 1.3886) \\ &= 0.1[0.25(1.3886) + (1.3886)^2] = 0.2275, \\ K_4 &= hf(x_2 + h, y_2 + K_3) = (0.1)f(0.3, 1.5049) \\ &= 0.1[0.3(1.5049) + (1.5049)^2] = 0.2716. \end{aligned}$$

Hence,

$$\begin{aligned} y(0.3) &= y_2 + \frac{1}{6}[K_1 + 2K_2 + 2K_3 + K_4] \\ &= 1.2774 + \frac{1}{6}[0.1887 + 2(0.22248) + 2(0.2275) + 0.2716] = 1.5041. \end{aligned}$$

We have thus

$$\begin{aligned} x_1 &= 0.1, y_1 = 1.1169, f_1 = 0.1(1.1169) + (1.1169)^2 = 1.3592, \\ x_2 &= 0.2, y_2 = 1.2774, f_2 = 0.2(1.2774) + (1.2774)^2 = 1.8874, \\ x_3 &= 0.3, y_3 = 1.5041, f_3 = 0.3(1.5041) + (1.5041)^2 = 2.714. \end{aligned}$$

Using the predictor, we have

$$\begin{aligned} y(0.4) &= y_4 = y_0 + \frac{4h}{3}[2f_3 - f_2 + 2f_1] \\ &= 1 + \frac{0.4}{3}[2(2.714) - 1.8872 + 2(1.3592)] = 1.8346. \end{aligned}$$

Then

$$f_4 = 0.4(1.8346) + (1.8346)^2 = 4.0996.$$

Using corrector, we have

$$\begin{aligned} y(0.4) &= y_2 + \frac{h}{3}[f_4 + 4f_3 + f_2] \\ &= 1.2774 + \frac{0.1}{3}[4.0996 + 4(2.714) + 1.8872] = 1.8388. \end{aligned}$$

EXAMPLE 10.26

Apply Milne's method to find a solution of the differential equation $\frac{dy}{dx} = x - y^2$ in the range $0 \leq x \leq 1$ for the boundary condition $y = 0$ at $x = 0$.

Solution. We have

$$y' = x - y^2, y = 0 \text{ at } x = 0.$$

By Picard's method

$$\begin{aligned}y_1 &= y_0 + \int_0^x f(x, y_0) dx = 0 + \int_0^x x dx = \frac{x^2}{2}, \\y_2 &= y_0 + \int_0^x f(x, y_1) dx = 0 + \int_0^x \left(x - \frac{x^4}{4} \right) dx = \frac{x^2}{2} - \frac{x^5}{20} \\y_3 &= y_0 + \int_0^x f(x, y_2) dx = 0 + \int_0^x \left[x - \left(\frac{x^2}{2} - \frac{x^5}{20} \right)^2 \right] dx \\&= \frac{x^2}{2} - \frac{x^5}{20} + \frac{x^8}{160} - \frac{x^{11}}{4400}.\end{aligned}$$

Taking $h = 0.2$, we have

$$\begin{aligned}y_0 &= 0 \text{ which gives } y'_0 = 0 - 0^2 = 0, \\y_1 &= \frac{(0.2)^2}{2} = 0.02 \text{ which implies } y'_1 = 0.2 - (0.02)^2 = 0.1996 \\y_2 &= \frac{(0.4)^2}{2} - \frac{(0.4)^5}{20} = 0.0795, \text{ which yields} \\y'_2 &= 0.4 - (0.0795)^2 = 0.3937, \\y_3 &= \frac{(0.6)^2}{2} - \frac{(0.6)^5}{20} + \frac{(0.6)^8}{160} - \frac{(0.6)^{11}}{4400} = 0.1762, \text{ which yields} \\y'_3 &= 0.5689.\end{aligned}$$

Using the predictor, we have

$$\begin{aligned}y_4 &= y(0.8) = y_0 + \frac{4h}{3} [2y'_1 - y'_2 + 2y'_3] \\&= 0 + \frac{4(0.2)}{3} [2(0.1996) - 0.3937 + 2(0.1762)] \\&= 0.3049.\end{aligned}$$

Then

$$y'_4 = 0.8 - (0.3049)^2 = 0.7070.$$

Therefore using corrector, we have

$$\begin{aligned}y_4 &= y_2 + \frac{h}{3} [y'_2 + 4y'_3 + y'_4] \\&= 0.0795 + \frac{0.2}{3} [0.3937 + 4(0.1762) + 0.7070] \\&= 0.3046.\end{aligned}$$

Now using predictor, we have

$$\begin{aligned}y_5 &= y(1) = y_1 + \frac{4h}{3} [2y'_2 - y'_3 + 2y'_4] \\&= 0.02 + \frac{4(0.2)}{3} [2(0.3937) - 0.5689 + 2(0.7070)] \\&= 0.4554.\end{aligned}$$

Then

$$y'_5 = 1 - (0.4554)^2 = 0.7926.$$

Therefore, the corrector gives

$$y(1) = y_5 = y_3 + \frac{h}{3}(y'_3 + 4y'_4 + y'_5) = 0.4555.$$

EXAMPLE 10.27

Using the formula

$$y_{n+1} = Dy_{n-3} + h[Ay'_n + By'_{n-1} + Cy'_{n-2}],$$

find the predictor of Milne's method.

Solution. The given formula is

$$y_{n+1} = Dy_{n-3} + h[Ay'_n + By'_{n-1} + Cy'_{n-2}].$$

Expanding the left-hand side and right-hand side in terms of y_{n-1} , we get

$$\begin{aligned} \text{L.H.S.} &= y_{n-1} + 2hy'_{n-1} + \frac{(2h)^2}{2} y''_{n-1} + \frac{(2h)^3}{3!} y'''_{n-1} + \frac{(2h)^4}{4!} y^{(iv)}_{n-1} + \dots \\ \text{R.H.S.} &= D \left[y_{n-1} - 2hy'_{n-1} + \frac{(2h)^2}{2} y''_{n-1} - \frac{(2h)^3}{3!} y'''_{n-1} + \frac{(2h)^4}{4!} y^{(iv)}_{n-1} + \dots \right] \\ &\quad + h \left[A \left(y'_{n-1} + hy''_{n-1} + \frac{h^2}{2} y'''_{n-1} + \frac{h^3}{3!} y^{(iv)}_{n-1} + \frac{h^4}{4!} y^{(v)}_{n-1} + \dots \right) \right. \\ &\quad \left. + By'_{n-1} + C \left(y'_{n-1} - hy''_{n-1} + \frac{h^2}{2} y'''_{n-1} - \frac{h^3}{3!} y^{(iv)}_{n-1} + \frac{h^4}{4!} y^{(v)}_{n-1} + \dots \right) \right] \end{aligned}$$

Comparing coefficients, we get

$$D = 1,$$

$$-2D + A + B + C = 2 \text{ or } A + B + C = 4,$$

$$-2D + A - C = 2 \text{ or } A - C = 0 \text{ or } A = C,$$

$$-\frac{8D}{6} + \frac{A}{2} + \frac{C}{2} = \frac{8}{6} \text{ or } A + C = \frac{32}{6}.$$

Hence, $A = C = \frac{8}{3}$, $B = -\frac{4}{3}$, $D = 1$. Putting these values in the given formula, we get

$$y_{n+1} = y_{n-3} + \frac{4h}{3} [2y'_n - y'_{n-1} + 2y'_{n-2}],$$

which is nothing but predictor formula of Milne–Simpson's method.

EXAMPLE 10.28

The equation $y' = f(x, y)$ is to be solved by using the following formula

$$y_{n+1} = Ay_n + By_{n-1} + h[Cy'_{n+1} + Dy'_n + Ey'_{n-1}] + R.$$

Find the coefficients A, B, C, D , and E .

Solution. By series expansion, we have

$$\begin{aligned}\text{L.H.S.} &= y_n + hy'_n + \frac{h^2}{2} y''_n + \frac{h^3}{3!} y'''_n + \frac{h^4}{4!} y^{(iv)}_n + \dots, \\ \text{R.H.S.} &= Ay_n + B \left(y_n - hy'_n + \frac{h^2}{2} y''_n - \frac{h^3}{3!} y'''_n + \frac{h^4}{4!} y^{(iv)}_n - \dots \right) \\ &\quad + Ch \left(y'_n + hy''_n + \frac{h^2}{2!} y'''_n + \frac{h^3}{3!} y^{(iv)}_n + \frac{h^4}{4!} y^{(v)}_n + \dots \right) \\ &\quad + Dhy'_n + Eh \left(y'_n - hy''_n + \frac{h^2}{2} y'''_n - \frac{h^3}{3!} y^{(iv)}_n + \frac{h^4}{4!} y^{(v)}_n - \dots \right),\end{aligned}$$

Comparing the coefficients on both sides, we obtain

$$\begin{aligned}A + B &= 1, \\ -B + C + D + E &= 1, \\ B + 2C - 2E &= 1, \\ -B + 3C + 3E &= 1, \\ B + 4C - 4E &= 1.\end{aligned}$$

Solving these equations, we get

$$A = 0, B = 1, C = \frac{1}{3}, D = \frac{4}{3}, E = \frac{1}{3}.$$

Hence, the formula becomes

$$y_{n+1} = y_{n-1} + \frac{h}{3} [y'_{n+1} + 4y'_n + y'_{n-1}] + O(h^5),$$

which is the corrector formula of Milne–Simpson’s method.

10.5 STABILITY OF METHODS

While solving initial value problem for ordinary differential equations numerically, an error is introduced into computation at each step due to the inaccuracy of the formula. The magnitude of this error is called the local truncation error. This local truncation error is a measure of the accuracy of the integration formula. The magnitude of the total error occurred in a particular method may become large due to accumulation and amplification of the local errors. This growth phenomenon of error is called numerical instability.

A computed solution to a problem is said to be stable if it behaves like the true solution of the problem. The stability of the solution depends on three factors:

- (i) the method used,
- (ii) the differential equation, and
- (iii) the step-length h .

If there is stability for small h but not for large h , then the solution is said to be partially instable. We observe that one-step methods like those of Runge–Kutta do not have numerical instability for sufficiently small h .

1. Stability of Multistep Method

To determine the stability of a multistep method, we proceed as follows:

Suppose that the multistep method yields a difference equation of order k . Suppose that $\beta_1, \beta_2, \dots, \beta_k$ are the roots of the characteristic equation corresponding to the homogeneous difference equation. The general solution of the homogeneous difference equation is then

$$y_n = c_1 \beta_1^n + c_2 \beta_2^n + \dots + c_k \beta_k^n. \quad (10.45)$$

One of these solutions, say β_1^n , will tend to exact solution of the difference equation as $h \rightarrow 0$. All the other solutions are extraneous. A multistep method is called strongly stable if the extraneous roots satisfy the condition

$$|\beta_i| < 1, \quad i = 2, 3, \dots, k$$

as $h \rightarrow 0$. Under these conditions, any error introduced in the computation decays as n increases. But if $|\beta_i| > 1$ for any $i = 2, 3, \dots, k$, then the error will grow exponentially.

Since it is impossible to obtain the roots β_i of the characteristic equation for the differential equation $y' = f(x, y)$, we consider special equation $y' = Ay$, where A is a constant, to get an indication of the stability of a method.

2. Stability of Milne's Method

In Milne's method, the error in the solution may be considered only due to corrector formula and so the stability depends only on the corrector formula:

$$y_{n+1} = y_{n-1} + \frac{h}{3} [y'_{n+1} + 4y'_n + y'_{n-1}].$$

Therefore, for the differential equation $y' = Ay$, the difference equation is

$$y_{n+1} = y_{n-1} + \frac{h}{3} [Ay_{n+1} + 4Ay_n + Ay_{n-1}]$$

or

$$(1 - \gamma)y_{n+1} - 4\gamma y_n - (1 + \gamma)y_{n-1} = 0,$$

where $\gamma = \frac{Ah}{3}$. The characteristic equation of this difference equation is

$$(1 - \gamma)\beta^2 - 4\gamma\beta - (1 + \gamma)\beta = 0,$$

whose roots are

$$\begin{aligned} \beta_1 &= \frac{2\gamma + \sqrt{1 + 3\gamma^2}}{1 - \gamma} = \frac{\frac{2Ah}{3} + \sqrt{1 + \frac{A^2h^2}{3}}}{1 - \frac{Ah}{3}} \\ \beta_2 &= \frac{2\gamma - \sqrt{1 + 3\gamma^2}}{1 - \gamma} = \frac{\frac{2}{3}Ah - \sqrt{1 + \frac{A^2h^2}{3}}}{1 - \frac{Ah}{3}}. \end{aligned}$$

When h is small,

$$\beta_1 = 1 + Ah + O(h^2)$$

$$\beta_2 = -(1 - Ah) + O(h^2).$$

Therefore, the general solution is

$$y_n = c_1(1 + Ah + O(h^2))^n + c_2(-1)^n(1 - Ah + O(h^2))^n.$$

Since $n = \frac{x_n}{h}$ and $\lim_{\varepsilon \rightarrow 0} (1 + \varepsilon)^{\frac{1}{\varepsilon}} = e$,

we have

$$\begin{aligned} (1 + Ah)^n &= (1 + Ah)^{\frac{x_n}{h}} = (1 + Ah)^{\frac{Ax_n}{Ah}} \\ &= (1 + Ah)^{\frac{1}{Ah} \cdot Ax_n} = e^{Ax_n} \text{ as } h \rightarrow 0. \end{aligned}$$

Hence, as $h \rightarrow 0$, the solution approaches

$$y_n = c_1 e^{Ax_n} + c_2 (-1)^n e^{-Ax_n}. \quad (10.46)$$

Also, the true solution of $y' = Ay$ is $y_0 e^{Ax_n}$. It follows therefore that the first term in equation (10.46) arises from the true solution, whereas the second term has no relationship, whatsoever, to the true solution of the differential equation. Such solutions are called parasitic solutions or extraneous solutions and arise only because we have replaced a first order differential equation by a second order difference equation.

Expression (10.46) shows that the stability of the method depends upon the sign of A . If $A > 0$, the desired solution $c_1 e^{Ax_n}$ increases exponentially and extraneous solution decreases exponentially as n increases. Thus, for $A > 0$, the errors introduced at various stages of computation will not amplify and so the Milne's method will be stable. On the other hand if $A < 0$, the extraneous solution increases exponentially, whereas the true solution decays. Hence for $A < 0$, the Milne's method is unstable.

The methods whose stability depends upon the sign of A are also known as weakly unstable. Thus, Milne's method is weakly unstable.

Remark 10.2. Almost all predictor–corrector methods that are commonly used are subject to instability for some range of values of the step-size h .

A method said to be absolutely stable if $|\beta_i| \leq 1$, $i = 1, 2, \dots$, and all β_i are simple roots of the characteristic equation of the difference equation of the method.

EXAMPLE 10.29

Find the stability interval of the trapezoidal formula

$$y_{n+1} = y_n + \frac{h}{2}(f_n + f_{n+1}).$$

Solution. We set $y' = f(x, y) = Ay$. Then the difference equation reduces to

$$y_{n+1} = y_n + \frac{h}{2}(Af_n + Af_{n+1}).$$

or

$$\left(1 - \frac{Ah}{2}\right)y_{n+1} - \left(1 + \frac{Ah}{2}\right)y_n = 0.$$

The characteristic equation of this difference equation is

$$\left(1 - \frac{Ah}{2}\right)\beta - \left(1 + \frac{Ah}{2}\right) = 0,$$

which has only one root β given by

$$\beta = \frac{1 + \frac{Ah}{2}}{1 - \frac{Ah}{2}}.$$

Therefore, the stability requires

$$|\beta| \left| \frac{1 + \frac{Ah}{2}}{1 - \frac{Ah}{2}} \right| \leq 1$$

and so the stability interval of the trapezoidal formula is $(-\infty, 0)$.

EXAMPLE 10.30

Discuss the stability of Adams–Moulton method

$$y_{n+1} = y_n + \frac{h}{24} [9f_{n-1} + 19f_n - 5f_{n-1} + f_{n-2}] + \text{Error term.}$$

Solution. Setting $y' = f(x, y) = Ay$ and rearranging, we obtain the following difference equation of order three:

$$\left(1 - \frac{9Ah}{24}\right)y_{n+1} - \left(1 + \frac{19Ah}{24}\right)y_n + \frac{5Ah}{24}y_{n-1} - \frac{Ah}{24}y_{n-2} = 0.$$

The characteristic equation for this difference equation is

$$(1 - 9\gamma)\beta^3 + (1 + 19\gamma)\beta^2 + 5\gamma\beta - \gamma = 0,$$

where $\gamma = \frac{Ah}{24}$. The roots of this equation are functions of h . But as $h \rightarrow 0$, the above characteristic equation reduces to $\beta^3 - \beta^2 = 0$, whose roots are $\beta_1 = 1$, $\beta_2 = \beta_3 = 0$. For small h , it can be shown that β_1^n approximates the desired solution of the differential equation, while β_2^n and β_3^n represent extraneous solutions whose magnitudes are less than 1. It follows therefore that the Adams–Moulton method is strongly stable for sufficiently small h .

EXAMPLE 10.31

Discuss the stability of the multistep method

$$y_{n+1} = y_{n-1} + 2hf_n.$$

Solution. Setting $y' = Ay$, the given difference equation of the multistep method reduces to

$$y_{n+1} - 2Ah y_n - y_{n-1} = 0.$$

The characteristic equation of this difference equation is

$$\beta^2 - 2Ah\beta - 1 = 0,$$

which yields

$$\beta = Ah \pm \sqrt{1 + A^2 h^2} = \beta_1, \beta_2 \text{ say.}$$

When h is small,

$$\begin{aligned}\beta_1 &= Ah + 1 + \frac{1}{2} A^2 h^2 + \dots \\ &= 1 + Ah + O(h^2), \\ \beta_2 &= -(1 - Ah) + O(h^2).\end{aligned}$$

Hence, the general solution is

$$y_n = c_1(1 + Ah + O(h^2))^n + c_2(-1)^n(1 - Ah + O(h^2))^n.$$

Since $n = \frac{x_n}{h}$, as $h \rightarrow 0$, the general solution becomes of the form

$$y_n = c_1 e^{Ax_n} + c_2 (-1)^n e^{-Ax_n}.$$

The true solution of the equation is $y_0 e^{Ax_0}$. Hence the stability depends on the sign of A and so this method is weakly unstable.

EXAMPLE 10.32

Discuss the stability analysis of the formula

$$y_{n+1} = y_n + h \left[\frac{3}{2} f_n - \frac{1}{2} f_{n-1} \right].$$

Solution. Setting $y' = f(x, y) = Ay$, the difference equation for the given formula is

$$y_{n+1} = y_n + h \left[\frac{3A}{2} y_n - \frac{A}{2} y_{n-1} \right]$$

or

$$y_{n+1} - \left(1 + \frac{3Ah}{2} \right) y_n + \frac{Ah}{2} y_{n-1} = 0.$$

The characteristic equation of this difference equation is

$$\beta^2 - \left(1 + \frac{3Ah}{2} \right) \beta + \frac{Ah}{2} = 0.$$

The roots of this equation are

$$\begin{aligned}\beta &= \frac{1 + \frac{3Ah}{2} \pm \sqrt{1 + \frac{9A^2 h^2}{4} + \frac{6Ah}{2} - \frac{4Ah}{2}}}{2} \\ &= \frac{1 + \frac{3Ah}{2} \pm \sqrt{1 + \frac{9A^2 h^2}{4} + Ah}}{2} = \beta_1, \beta_2 \text{ say.}\end{aligned}$$

Simplifying, we get

$$\beta_1 = 1 + Ah + O(h^2),$$

$$\beta_2 = \frac{Ah}{2} + O(h^2).$$

The general solution is, therefore,

$$y_n = c_1(1 + Ah + O(h^2))^n + \frac{1}{2^n}(Ah + O(h^2))^n.$$

Since $n = \frac{x_n}{h}$, as $h \rightarrow 0$, the first term tends to $c_1 e^{Ax_n}$ which is the true solution of the equation but the second term (extraneous solution) decays. Hence the given method is stable.

10.6 SECOND ORDER DIFFERENTIAL EQUATION

Consider second order differential equation of the type

$$y'' = f(x, y) \quad (y_0 \text{ and } y'_0 \text{ given}),$$

which frequently occurs in applied mathematics. We discuss Numerov's method to solve such equations. We know that

$$\frac{\delta^2}{U^2} = 1 + \frac{\delta^2}{12} - \frac{\delta^4}{240} + \dots, \quad U = hD.$$

Therefore,

$$\begin{aligned} \delta^2 y_n &= \left(1 + \frac{\delta^2}{12} - \frac{\delta^4}{240} + \dots\right) h^2 D^2 y_n, \\ &= h^2 \left(1 + \frac{\delta^2}{12} - \frac{\delta^4}{240} + \dots\right) y''_n, \\ &= h^2 \left(1 + \frac{\delta^2}{12} - \frac{\delta^4}{240} + \dots\right) f_n. \end{aligned}$$

Also,

$$\begin{aligned} \delta^2 y_n &= \delta(\delta y_n) = \delta \left(y_{\frac{n+1}{2}} - y_{\frac{n-1}{2}} \right) \\ &= \delta y_{\frac{n+1}{2}} - \delta y_{\frac{n-1}{2}} = (y_{n+1} - y_n) - (y_n - y_{n-1}) \\ &= y_{n+1} - 2y_n + y_{n-1}. \end{aligned}$$

Hence,

$$\begin{aligned} y_{n+1} - 2y_n + y_{n-1} &= h^2 \left(1 + \frac{\delta^2}{12} - \frac{\delta^4}{240} + \dots\right) f_n \\ &= h^2 \left(f_n + \frac{1}{12}(f_{n+1} - 2f_n + f_{n-1})\right) - \frac{h^6}{240} y_n^{(vi)} + O(h^8) \\ &= \frac{h^2}{12}(f_{n+1} + 10f_n + f_{n-1}) - \frac{h^6}{240} y_n^{(vi)} + O(h^8), \end{aligned}$$

since $\delta^4 \approx U^4 = h^4 D^4$ and $y''_n = f_n$. Neglecting terms of order h^6 and higher, we get

$$y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{12}(f_{n+1} + 10f_n + f_{n-1}),$$

with a local truncation error $O(h^6)$.

In the special case when $f(x, y) = y \cdot g(x)$, we get

$$y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{12}(y_{n+1}g_{n+1} + 10y_ng_n + y_{n-1}g_{n-1})$$

or

$$\left(1 - \frac{h^2}{12}g_{n+1}\right)y_{n+1} = \left(2 + \frac{5h^2}{6}g_n\right)y_n - \left(1 - \frac{h^2}{12}g_{n-1}\right)y_{n-1}$$

or

$$y_{n+1} = \frac{\left(2 + \frac{5h^2}{6}g_n\right)y_n - \left(1 - \frac{h^2}{12}g_{n-1}\right)y_{n-1}}{1 - \frac{h^2}{12}g_{n+1}}.$$

EXAMPLE 10.33

Solve

$y'' = -y$, $y(0) = 0$ and $y(h) = k$ by Numerov's method. Compute y_6 when $h = \frac{\pi}{6}$ and $k = \frac{1}{2}$.

Solution. We have

$$y'' = -y = yg(x), \quad y(0) = 0, \quad y(h) = k,$$

where $g(x) = -1$. By Numerov's method, we have

$$y_{n+1} = \frac{\left(2 + \frac{5h^2}{6}g_n\right)y_n - \left(1 - \frac{h^2}{12}g_{n-1}\right)y_{n-1}}{1 - \frac{h^2}{12}g_{n+1}}.$$

But $g_n = -1$, $y(0) = 0$, $y_1 = y(h) = k$. Therefore,

$$y_2 = \frac{\left(2 - \frac{5h^2}{6}\right)y_1}{1 + \frac{h^2}{12}} = \frac{2(12 - 5h^2)}{12 + h^2}k.$$

We consider

$$y_n = \frac{k \sin n\phi}{\sin \phi}. \quad (10.47)$$

Then y_n satisfies the initial conditions $y(0) = 0$, $y(h) = k$. Also,

$$y_2 = \frac{k \sin 2\phi}{\sin \phi} = \frac{2(12 - 5h^2)}{12 + h^2}k$$

and so

$$\frac{2 \sin \phi \cos \phi}{\sin \phi} = \frac{2(12 - 5h^2)}{12 + h^2}$$

or

$$\cos \phi = \frac{12 - 5h^2}{12 + h^2}$$

or

$$\phi = \cos^{-1} \frac{12 - 5h^2}{12 + h^2}.$$

Substituting this value of ϕ in equation (10.47) and taking $n = 6$, $k = \frac{1}{2}$, $h = \frac{\pi}{6}$, we get

$$y_6 = \frac{1}{2} \frac{\sin 6\phi}{\sin \phi} \quad \text{with } \phi \cos^{-1} \frac{12 - 5(\pi/6)^2}{12 + (\pi/6)^2} \\ = -0.005.$$

EXAMPLE 10.34

Solve the differential equation

$$y'' = (x-1)(x-2)y, \quad y(0) = 0.5, \quad y(0.5) = 0.42$$

by Numerov's method for $x = 1$ and $x = 1.5$.

Solution. By Numerov's method

$$y_{n+1} = \frac{\left(2 + \frac{5h^2}{6} g_n\right) y_n - \left(1 - \frac{h^2}{12} g_{n-1}\right) y_{n-1}}{1 - \frac{h^2}{12} g_{n+1}}.$$

In this problem

$$g(x) = (x-1)(x-2)$$

and so

$$g_0 = (0-1)(0-2) = 2,$$

$$g_1 = (0.5-1)(0.5-2) = 0.75$$

$$g_2 = 0,$$

$$g_3 = (1.5-1)(1.5-2) = -0.25$$

Also,

$$y_0 = 0.5, \quad y_1 = 0.42$$

Hence,

$$y_2 = y(1) = \frac{\left[2 + \frac{5}{6}(0.25)(0.75)\right](0.42) - \left(1 - \frac{0.25}{12}(2)\right)(0.5)}{1-0} \\ = 0.426465,$$

and

$$y_3 = y(1.5) = \frac{\left[2 + \frac{5}{6}(0.25)(0)\right](0.426465) - \left[1 - \frac{0.25}{12}(0.75)\right](0.42)}{1 - \frac{0.25}{12}(-0.25)} \\ = 0.437213.$$

EXAMPLE 10.35

Expanding the expression

$$y_{n+1} + Ah^2 y''_{n+1} = By_n + Ch^2 y''_n + Dy_{n-1} + Eh^2 y''_{n-1} + R,$$

obtain Numerov's method for solving differential equations of second order.

Solution. Expanding the left-hand side and the right-hand side of the given expression, we get

$$\begin{aligned} \text{L.H.S.} &= y_n + hy'_n + \frac{h^2}{2} y''_n + \frac{h^3}{3!} y'''_n + \frac{h^4}{4!} y^{(iv)}_n + \frac{h^5}{5!} y^{(v)}_n + \dots \\ &\quad + Ah^2 \left(y''_n + hy'''_n + \frac{h^2}{2} y^{(iv)}_n + \frac{h^3}{3!} y^{(v)}_n + \frac{h^4}{4!} y^{(vi)}_n + \frac{h^5}{5!} y^{(vii)}_n + \dots \right), \\ \text{R.H.S.} &= By_n + Ch^2 y''_n + D \left(y_n - hy'_n + \frac{h^2}{2} y''_n - \frac{h^3}{3!} y'''_n + \frac{h^4}{4!} y^{(iv)}_n - \dots \right) \\ &\quad + Eh^2 \left(y''_n - hy'''_n + \frac{h^2}{2!} y^{(iv)}_n - \frac{h^3}{3!} y^{(v)}_n + \frac{h^4}{4!} y^{(vi)}_n - \dots \right) + R. \end{aligned}$$

Comparison of coefficients yields

$$1 = B + D,$$

$$-1 = D,$$

$$\frac{1}{2} + A = C + \frac{D}{2} + E \text{ and so } A - C - E = -1,$$

$$\frac{1}{6} + A = -\frac{D}{6} - E \text{ and so } A = -E,$$

$$\frac{1}{24} + \frac{A}{2} = \frac{D}{24} + \frac{E}{2}.$$

Solving these equations, we get

$$A = -\frac{1}{12}, B = 2, C = \frac{5}{6}, D = -1, E = \frac{1}{12}.$$

We have used the terms up to $h^4 y^{(iv)}$. But the coefficient of $h^5 y^{(v)}$ is also zero. Hence, the error term is $-\frac{h^6}{240} h^6 y^{(vi)}$. Therefore, the formula becomes

$$y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{12} (y''_{n+1} + 10y''_n + y''_{n-1}) + O(h^6).$$

10.7 SOLUTION OF BOUNDARY VALUE PROBLEMS BY FINITE DIFFERENCE METHOD

We discuss here finite difference method to solve a boundary value problem. In this method, every derivative appearing in the differential equation or the boundary conditions is replaced by an appropriate difference approximation. Some useful difference approximations are given below:

By Taylor's series approximation, we have

$$y_{n+1} = y_n + hy'_n + \frac{h^2}{2!} y''_n + \frac{h^3}{3!} y'''_n + \dots \tag{10.48}$$

and

$$y_{n-1} = y_n - hy'_n + \frac{h^2}{2!} y''_n - \frac{h^3}{3!} y'''_n + \dots \quad (10.49)$$

Subtracting equation (10.49) from equation (10.48), we obtain

$$y_{n+1} - y_{n-1} = 2hy'_n + O(h^3).$$

Hence,

$$y'_n = \frac{y_{n+1} - y_{n-1}}{2h} + O(h^3). \quad (10.50)$$

On the other hand, addition of equations (10.48) and (10.49) yields

$$y_{n+1} + y_{n-1} = 2y_n + h^2 y''_n + O(h^4). \quad (10.51)$$

Therefore, we have

$$y''_n = \frac{y_{n+1} - 2y_n + y_{n-1}}{h^2}. \quad (10.52)$$

We illustrate the method of finding a solution of a boundary value problem with the help of the following examples:

EXAMPLE 10.36

Solve the boundary value problem

$$y'' = x + y, \quad y(0) = 1, \quad y(1) = 1$$

by finite difference method.

Solution. Taking step size $h = 0.2$, the given differential equation may be approximated by

$$\frac{y_{n+1} - 2y_n + y_{n-1}}{h^2} = y''_n = x_n + y_n.$$

or

$$y_{n+1} - 2y_n + y_{n-1} = 0.04(x_n + y_n).$$

Taking $n = 1$, we get

$$y_0 - 2y_1 + y_2 = 0.04(x_1 + y_1).$$

Since $y_0 = 0$, $x_1 = x_0 + h = 0.2$, this yields

$$-2.04y_1 + y_2 = 0.008. \quad (10.53)$$

Similarly, taking $n = 2, 3, 4$, we get

$$y_1 - 2.04y_2 + y_3 = 0.016 \quad (10.54)$$

$$y_2 - 2.04y_3 + y_4 = 0.024 \quad (10.55)$$

$$y_3 - 2.04y_4 = 0.968, \quad (10.56)$$

since $y_5 = y(1) = 1$. The solution of the above system is

$$y_1 = 0.1428, \quad y_2 = 0.2993, \quad y_3 = 0.4838, \quad y_4 = 0.7117.$$

EXAMPLE 10.37

Solve

$$y'' + \frac{y}{1+x^2} = 7x, \quad y(0) = 0, \quad y(1) = 2$$

by finite difference method.

Solution. The given equation is approximated by

$$y'' = \frac{y_{n+1} - 2y_n + y_{n-1}}{h^2}.$$

Therefore taking $h = 0.2$, the given differential equation is approximated by

$$\frac{y_{n+1} - 2y_n + y_{n-1}}{h^2} + \frac{y_n}{1+x_n^2} = 7x_n$$

or

$$y_{n+1} + y_n \left(\frac{h^2 - 2(1+x_n)^2}{1+x_n^2} \right) + y_{n-1} = 7h^2 x_n$$

or

$$y_{n+1} + y_n \left(\frac{0.0 - 42 - 2x_n^2}{1+x_n^2} \right) + y_{n-1} = 0.28x_n$$

or

$$y_{n+1} - y_n \left(\frac{1.96 + 2x_n^2}{1+x_n^2} \right) + y_{n-1} = 0.28x_n.$$

This is a linear system of equations with four unknowns denoted by $y(0.2) = y_1$, $y(0.4) = y_2$, $y(0.6) = y_3$, and $y(0.8) = y_4$. Also, $y(1) = y_5 = 2$ (given).

Therefore, the system of equations becomes

$$\begin{aligned} -1.9615y_1 + y_2 &= 0.056 \\ y_1 - 1.9655y_2 + y_3 &= 0.112 \\ y_2 - 1.9706y_3 + y_4 &= 0.168 \\ y_3 - 1.9756y_4 &= 1.776. \end{aligned}$$

Solving this system, we get

$$y_1 = 0.208, \quad y_2 = 0.464, \quad y_3 = 0.816, \quad y_4 = 1.312.$$

The solution is in good agreement with the exact solution $y = x^3 + x$.

EXAMPLE 10.38

Find the solution of the differential equation

$$y'' + x^2 y = 0, \quad y(0) = 0, \quad y'(0) = 1$$

at the points 0.25, 0.50, and 0.75 using finite difference method.

Solution. Approximating y'' by

$$\frac{y_{n+1} - 2y_n + y_{n-1}}{h^2},$$

the given equation turns into the difference equation

$$\frac{y_{n+1} - 2y_n + y_{n-1}}{h^2} + x_n^2 y_n = 0$$

or

$$y_{n+1} - 2y_n + y_{n-1} + h^2 x_n^2 y_n = 0$$

or

$$y_{n+1} - (2 - h^2 x_n^2) y_n + y_{n-1} = 0. \quad (10.57)$$

Taking $h = 0.25$, $y(0.25) = y_1$, $y(0.50) = y_2$, $y(0.75) = y_3$ and putting $n = 1, 2, 3$ in equation (10.57), we get the following system of equations:

$$\begin{aligned} -1.99609375 y_1 + y_2 &= 0, \\ y_1 - 1.984375 y_2 + y_3 &= 0, \\ y_2 - 1.96484375 y_3 + 1 &= 0, \end{aligned}$$

since $y_4 = y(1) = 1$. Solving these equations, we get

$$y_1 = 0.2617, \quad y_2 = 0.5223 \text{ and } y_3 = 0.7748.$$

EXAMPLE 10.39

Solve the boundary value problem

$$y'' - 64y + 10 = 0, \quad y(0) = y(1) = 0$$

by finite difference method.

Solution. Approximating y'' by $\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2}$, the given equation is represented by the difference equation

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} - 64y_i + 10 = 0.$$

Taking $h = 0.25$, $y(0.25) = y_1$, $y(0.50) = y_2$, $y(0.75) = y_3$ and $i = 1, 2, 3$, we get

$$\begin{aligned} -48y_1 + 8y_2 + 5 &= 0 \\ 8y_1 - 48y_2 + 8y_3 + 5 &= 0 \\ 8y_2 - 48y_3 + 5 &= 0. \end{aligned}$$

Solving these equations, we get

$$y_1 = 0.1287, \quad y_2 = 0.1471, \quad y_3 = 0.1287.$$

10.8 USE OF THE FORMULA $\delta^2 y_n = h^2 \left(1 + \frac{\delta^2}{12} - \frac{\delta^4}{240} + \dots \right) f_n$ TO SOLVE BOUNDARY VALUE PROBLEMS

A boundary value problem can also be solved using the above-mentioned formula. We illustrate the use of the formula in the form of the following example:

EXAMPLE 10.40

Solve the boundary value problem:

$$y'' = xy, \quad y(0) = 0, \quad y(1) = 1$$

taking step size $h = 0.2$.

Solution. We have

$$\delta^2 y_n = h^2 \left(1 + \frac{\delta^2}{12} - \frac{\delta^4}{240} + \dots \right) f_n.$$

The first approximate solution is obtained from the relation

$$\delta^2 y_n = h^2 f_n = h^2 (x_n y_n). \quad (10.58)$$

But

$$\begin{aligned} \delta^2 y_n &= \delta(\delta y_n) = \delta \left(y_{\frac{n+1}{2}} - y_{\frac{n-1}{2}} \right) \\ &= \delta y_{\frac{n+1}{2}} - \delta y_{\frac{n-1}{2}} \\ &= (y_{n+1} - y_n) - (y_n - y_{n-1}) \\ &= y_{n+1} - 2y_n + y_{n-1}. \end{aligned}$$

Thus, equation (10.58) reduces to

$$y_{n+1} - 2y_n + y_{n-1} = h^2 x_n y_n. \quad (10.59)$$

Putting $n = 1, 2, 3, 4$ in equation (10.59), we have

$$\begin{aligned} y_2 - 2y_1 + y_0 &= h^2 x_1 y_1 \\ y_3 - 2y_2 + y_1 &= h^2 x_2 y_2 \\ y_4 - 2y_3 + y_2 &= h^2 x_3 y_3 \\ y_5 - 2y_4 + y_3 &= h^2 x_4 y_4. \end{aligned}$$

But $y_0 = 0$, $y_5 = 1$ and $h = 0.2$. Therefore, the above set of equations reduces to

$$\begin{aligned} -2.008y_1 + y_2 &= 0 \\ y_1 - 2.016y_2 + y_3 &= 0 \\ y_2 - 2.024y_3 + y_4 &= 0 \\ y_3 - 2.032y_4 + 1 &= 0. \end{aligned}$$

The solution of this system is

$$y_1 = 0.18496, \quad y_2 = 0.37139, \quad y_3 = 0.56374, \quad y_4 = 0.76956.$$

We use these values to find the values of xy and get the difference table for $y'' = f(x, y) = xy$ given below:

f_0	0.0000		
f_1	0.0370	0.370	0.0746
f_2	0.1486	0.1116	0.0780
f_3	0.3382	0.1896	0.0878
f_4	0.6156	0.2774	0.1070
f_5	1.0000	0.3844	

Then the correction terms required (up to two significant figures) are

$$\frac{h^2}{12} \delta^2 f_1 = \frac{0.04}{12} (0.0746) = 0.00025,$$

$$\frac{h^2}{12} \delta^2 f_2 = \frac{0.04}{12} (0.0780) = 0.00026,$$

$$\frac{h^2}{12} \delta^2 f_3 = \frac{0.04}{12} (0.0878) = 0.00029,$$

$$\frac{h^2}{12} \delta^2 f_4 = \frac{0.04}{12} (0.1070) = 0.00036.$$

The system of equations is now

$$y_2 - 2.008 y_1 = 0.00025$$

$$y_3 - 2.016 y_2 + y_1 = 0.00026$$

$$y_4 - 2.024 y_3 + y_2 = 0.00029$$

$$1 - 2.032 y_4 + y_3 = 0.00036.$$

We therefore deduce that the corrections z_1, z_2, z_3 , and z_4 to the approximate values already found for y_1, y_2, y_3 , and y_4 will satisfy the system of equations

$$z_2 - 2.008 z_1 = 0.00025$$

$$z_3 - 2.016 z_2 + z_1 = 0.00026$$

$$z_4 - 2.024 z_3 + z_2 = 0.00029$$

$$1 - 2.032 z_4 + z_3 = 0.00036.$$

Since z_1, z_2, z_3 , and z_4 are small, these equations can be simplified to

$$-2z_1 + z_2 = 0.00025$$

$$z_1 - 2z_2 + z_3 = 0.00026$$

$$z_2 - 2z_3 + z_4 = 0.00029$$

$$z_3 - 2z_4 = 0.99964.$$

Solving these equations, we get

$$z_1 = -0.00055, \quad z_2 = -0.00084, \quad z_3 = -0.00087, \quad z_4 = -0.00061.$$

Therefore, the revised approximations for y_1, y_2, y_3 , and y_4 are

$$y_1 = 0.18496 - 0.00055 = 0.18441$$

$$y_2 = 0.37139 - 0.00084 = 0.37055$$

$$y_3 = 0.56374 - 0.00087 = 0.56287$$

$$y_4 = 0.76956 - 0.00061 = 0.76895.$$

The above process is repeated till we get the values correct to required decimal places.

10.9 EIGENVALUE PROBLEMS

Consider the boundary value problem

$$y'' + \lambda^2 y = 0, \quad y(0) = y(1) = 0. \tag{10.60}$$

This differential equation has the solution

$$y = A \cos \lambda x + B \sin \lambda x.$$

The condition $y(0) = 0$ implies $A = 0$, whereas the condition $y(1) = 0$ yields $B \sin \lambda = 0$. Now if $\sin \lambda \neq 0$, we have $B = 0$ and so the equation has the trivial solution $y(x) = 0$. On the other hand, if $\sin \lambda = 0$, that is, $\lambda = n\pi$; n an integer, then B can be chosen arbitrarily. The special values $\lambda^2 = n^2\pi^2$ are called eigenvalues and the corresponding solutions eigenfunctions.

The above discussion shows that a differential equation with boundary values corresponds exactly to a linear homogeneous system of equations. If the coefficient matrix is regular (non-singular), we have only one trivial solution. But if the matrix is singular, we have an infinite number of non-trivial solutions. We encounter with such problems in vibrations of mechanical system.

To solve equation (10.60) numerically, we use finite difference method. So, we replace y'' by $\frac{y_{n+1} - 2y_n + y_{n-1}}{h^2}$ and take $h = 0.25$. Thus,

$$\frac{y_{n+1} - 2y_n + y_{n-1}}{h^2} + \lambda^2 y_n = 0$$

or

$$y_{n+1} - 2y_n + y_{n-1} + \lambda^2 h^2 y_n = 0$$

or

$$y_{n+1} - (2 - \lambda^2 h^2) y_n + y_{n-1} = 0.$$

We are given that $y(0) = 0$ and $y(1) = y_4 = 0$. We further write $y(0.25) = y_1$, $y(0.50) = y_2$, $y(0.75) = y_3$. Thus putting $n = 1, 2, 3$, we get

$$\begin{aligned} &-[2 - \lambda^2(0.0625)]y_1 + y_2 = 0 \\ &y_1 - [2 - \lambda^2(0.0625)]y_2 + y_3 = 0 \\ &y_2 - [2 - \lambda^2(0.0625)]y_3 = 0. \end{aligned}$$

For non-trivial solution, we must have

$$\begin{vmatrix} 0 & 1 & -(2 - 0.0625\lambda^2) \\ 1 & -(2 - 0.0625\lambda^2) & 1 \\ -(2 - 0.0625\lambda^2) & 1 & 0 \end{vmatrix} = 0$$

or

$$(2 - 0.0625\lambda^2)[-2 + (2 - 0.0625\lambda^2)^2] = 0.$$

The first factor yields

$$2 - 0.0625\lambda^2 = 0$$

or

$$\lambda^2 = \frac{2}{0.0625} = 32 \text{ and so } \lambda^2 = 4\sqrt{2}.$$

The second factor yields

$$(2 - 0.0625\lambda^2)^2 = 2$$

or

$$2 - 0.0625\lambda^2 = \pm\sqrt{2}$$

or

$$\lambda^2 = \frac{2 \mp \sqrt{2}}{0.0625} = 16(2 \mp \sqrt{2}) \text{ and so } \lambda = 4\sqrt{2 \pm \sqrt{2}}.$$

EXAMPLE 10.41

Determine approximate value of the smallest characteristic value of λ for the problem

$$y'' + \lambda y = 0, \quad y(0) = y(1) = 0.$$

Solution. The eigenvalue problem is

$$y'' + \lambda y = 0, \quad y(0) = y(1) = 0.$$

Replacing y'' by the approximation $\frac{y_{n+1} - 2y_n + y_{n-1}}{h^2}$ and taking $h = 0.25$, we obtain the system of equations:

$$y_{n+1} - 2y_n + y_{n-1} + \lambda h^2 y_n = 0$$

or

$$y_{n+1} - (2 - 0.0625\lambda)y_n + y_{n-1} = 0 \quad (10.61)$$

We put $y(0.25) = y_1$, $y(0.50) = y_2$, and $y(0.75) = y_3$. Also, we are given that $y(0) = 0$ and $y(1) = y_4 = 0$. Thus putting $n = 1, 2, 3$ in equation (10.61), we obtain

$$\begin{aligned} -(2 - 0.0625\lambda)y_1 + y_2 &= 0 \\ y_1 - (2 - 0.0625\lambda)y_2 + y_3 &= 0 \\ y_2 - (2 - 0.0625\lambda)y_3 + 0 &= 0. \end{aligned}$$

For non-trivial solution, we must have

$$\begin{vmatrix} -(2 - 0.0625\lambda) & 1 & 0 \\ 1 & -(2 - 0.0625\lambda) & 1 \\ 0 & 1 & -(2 - 0.0625\lambda) \end{vmatrix} = 0$$

or

$$(2 - 0.0625\lambda)[2 - (2 - 0.0625\lambda)^2] = 0.$$

The first factor yields

$$\lambda = \frac{2}{0.0625} = 32.$$

The second factor yields

$$(2 - 0.0625\lambda)^2 = 2$$

or

$$2 - 0.0625\lambda = \pm\sqrt{2}$$

and so

$$\lambda = \frac{2 \mp \sqrt{2}}{0.0625} = 9.37258, 54.6274.$$

Therefore the smallest characteristic value is 9.37258. The exact solution is $y = \sin\sqrt{\lambda}x$. The condition $y(1) = 0$ gives $\sqrt{\lambda} = m\pi$, $m = 1, 2, 3, \dots$. For $m = 1$, we have $\sqrt{\lambda} = \pi$ or $\lambda = \pi^2 = 9.869604$.

EXERCISES

1. Solve $\frac{dy}{dx} = 1 - 2xy$, $y(0) = 0$ by Taylor's series method for $x = 0.2$.

Ans. 0.1947

2. Using Taylor's series method, obtain the values of y at $x = 0.1, 0.2, 0.3$ if y satisfies the equation $\frac{d^2y}{dx^2} + xy = 0$ and $y(0) = 1, y'(0) = 0.5$.

Ans. $y(0.1) = 1.050, y(0.2) = 1.099, y(0.3) = 1.145$

3. Solve $\frac{dy}{dx} = -xy$, $y(0) = 1$ over $[0, 0.1]$ with $h = 0.05$ using Taylor's series method.

Ans. $y(0.05) = 0.9987508, y(0.1) = 0.9950125$

4. Solve $\frac{dy}{dx} = 1 - y$, $y(0) = 0$ in $[0, 0.3]$ by modified Euler's method taking $h = 0.1$.

Ans. $y(0.1) = 0.095, y(0.2) = 0.180975, y(0.3) = 0.2587823$

5. Solve $\frac{dy}{dx} = x + y^2$, $y(0) = 1$ for $x = 0.5$ by modified Euler's method.

Ans. 2.2352

6. Solve $\frac{dy}{dx} = y - \frac{2x}{y}$, $y(0) = 1$ in $[0, 0.2]$ using Euler's method and taking $h = 0.1$.

Ans. $y(0.1) = 1.095909, y(0.2) = 1.184097$

7. Use Picard's method to solve $\frac{dy}{dx} = x - y^2$, $y(0) = 1$.

Ans. 0.9138

8. Use Picard's method to solve $y'' + 2xy' + y = 0$, $y(0) = 0.5, y'(0) = 0.1$ for $x = 0.1$.

Ans. 0.5075

9. Use Picard's method to solve $\frac{dy}{dx} = x^2 + y^2$, $y(0) = 0$ for $x = 0.4$.

Ans. 0.0214

10. Solve for $x = 0.1$, the equation $\frac{dy}{dx} = 3x + y^2$, $y(0) = 1$ by Picard's method.

Ans. $y(0.1) = 1.127$

11. Use Runge–Kutta method of order four to solve the differential equation $\frac{dy}{dx} = \frac{y^2 - x^2}{y^2 + x^2}$, $y(0) = 1$ at $x = 0.2$.

Ans. $y(0.2) = 1.196$

12. Use fourth order Runge–Kutta method to find $y(0.2)$ for the equation $\frac{dy}{dx} = \frac{y-x}{y+x}$, $y(0) = 1$.

Ans. $y(0.2) = 1.1749$

13. Solve $y' = \frac{y^2 - 2x}{y^2 + x}$, $y(0) = 1$ for $x = 0.1$ and $x = 0.2$ using Runge–Kutta method.

Ans. $y(0.1) = 1.091, y(0.2) = 1.168$

14. Using Runge–Kutta method, solve $y' = x + y$, $y(0) = 1$ for $x = 0.2$.

$$\text{Ans. } y(0.2) = 0.2428$$

15. Use Runge–Kutta method to solve $y' = -xy$, $y(0) = 1$ for $x = 0.2$.

$$\text{Ans. } 0.9801987$$

16. Use Runge–Kutta method to solve $\frac{dy}{dx} = 1 + xz$, $y(0) = 0$; $\frac{dz}{dx} = -xy$, $z(0) = 1$ for $x = 0.3$ and $x = 0.6$.

$$\text{Ans. } y(0.3) = 0.3448, z(0.3) = 0.99; y(0.6) = 0.7738, z(0.6) = 0.9121$$

17. Solve $y' = x + z$, $y(0) = 2$; $z' = x - y^2$, $z(0) = 1$ for $x = 0.1$ by Runge–Kutta method.

$$\text{Ans. } y(0.1) = 2.0845, z(0.1) = 0.586$$

18. Use Runge–Kutta method to solve $y'' = y^3$, $y(0) = 10$, $y'(0) = 5$ for $x = 0.1$.

$$\text{Ans. } y(0.1) = 17.42$$

19. Use fourth order Runge–Kutta method to solve $y'' = y + xy'$, $y(0) = 1$, $y'(0) = 0$ for $x = 0.2$.

$$\text{Ans. } y(0.2) = 0.9802$$

20. Solve $y'' = xy'^2 - y^2$, $y(0) = 1$, $y'(0) = 0$ for $x = 0.2$ using Runge–Kutta method.

$$\text{Ans. } y(0.2) = 0.9801$$

21. Use Milne's method to solve $y' = x^2 + y^2$, $y(0) = 1$ for $x = 0.3$. The values of y for $x = -0.1$, 0.1 , and 0.2 should be calculated by series method.

$$\text{Ans. } y(0.3) = 1.4392$$

22. Use Milne–Simpson's method to solve $y' = x - y^2$, at $x = 0.8$ under the conditions $y(0) = 0$, $y(0.2) = 0.02$, $y(0.4) = 0.0795$, $y(0.6) = 0.1762$.

$$\text{Ans. } y(0.8) = 0.3049$$

23. Use Milne–Simpson's method to solve $y' = 2e^x - y$ for $x = 0.4$ under the conditions $y(0) = 2$, $y(0.1) = 2.010$, $y(0.2) = 2.040$, $y(0.3) = 2.090$.

$$\text{Ans. } y(0.4) = 2.1621$$

24. Using Adams–Basforth formula find $y(1.4)$ for $y' = x^2(1 + y)$ subject to the given data

$x:$	1	1.1	1.2	1.3
$y:$	1	1.233	1.548	1.979

$$\text{Ans. } 2.575$$

25. Using Adams–Basforth formula find the solution of $y' = x - y^2$ at $x = 0.8$, given the data

$x:$	0	0.2	0.4	0.6
$y:$	0	0.0200	0.0795	0.1762

$$\text{Ans. } y(0.8) = 0.2416$$

26. Use Heun's method to solve the initial value problem $y' = -ty$, $y(0) = 1$, over $[0, 0.2]$ taking $h = 0.1$.

$$\text{Ans. } y(0.1) = 0.995, y(0.2) = 0.980175$$

27. Solve the boundary value problem $y'' = y$, $y(0) = 0$, $y(2) = 3.627$ by finite difference method taking $h = 0.5$.

$$\text{Ans. } y(0.5) = 0.5262, y(1.0) = 1.1843, y(1.5) = 2.1382$$

28. Solve the boundary value problem $y'' - 64y + 10 = 0$, $y(0) = y(1) = 0$, by finite difference method for $x = 0.5$.

$$\text{Ans. } y(0.5) = 0.1389$$

29. Solve the boundary value problem $y'' + y + 1 = 0$, $y(0) = y(1) = 0$ for $x = 0.5$, by finite difference method, taking $h = 0.25$.

Ans. $y(0.5) = 0.14031$

30. Determine approximate value of the smallest characteristic value of λ for the problem $y'' + \lambda xy = 0$, $y(0) = 0$, $y(1) = 0$ taking $h = 0.25$, 0.50, and 0.75.

Ans. $\lambda \approx 17.9$

11 Partial Differential Equations

Partial differential equations appear in the description of physical processes in applied sciences and engineering. A differential equation that involves more than one independent variable is called a partial differential equation. We restrict ourselves to second order partial differential equations. The general second order linear partial differential equation is of the form

$$Au_{xx} + Bu_{xy} + Cu_{yy} + Du_x + Eu_y + Fu = G,$$

where A, B, C, D, E, F, G are all functions of x and y . Equations of the above form can be classified into three types:

- (i) If $B^2 - 4AC < 0$ at a point in the (x, y) plane, then the equation is called elliptic. For example, the equation $u_{xx} + u_{yy} = 0$, known as Laplace equation, is elliptic.
- (ii) If $B^2 - 4AC = 0$ at a point in the (x, y) plane, then the equation is called parabolic. For example, the equation $u_{xx} - u_t = 0$, called the heat conduction equation, is parabolic.
- (iii) If $B^2 - 4AC > 0$ at a point in the (x, y) plane, then the equation is called hyperbolic. For example, the equation $u_{xx} - \frac{1}{c^2} u_{tt} = 0$, known as the wave equation, is hyperbolic.

The most popular method for solving partial differential equations is finite-difference method. This method is based on formulae for approximating the first and second derivatives of a function.

11.1 FORMATION OF DIFFERENCE EQUATION

To get finite difference analog of a partial differential equation, we replace the derivatives in the equation by their corresponding difference approximations.

We first derive difference formula for approximating u_x . By Taylor's series about the point (x_0, y_0) , we have

$$u(x_0 + h, y_0) = u(x_0, y_0) + h u'_x(x_0, y_0) + \frac{h^2}{2} u''_{xx}(\xi, y_0), \quad x_0 \leq \xi \leq x_0 + h$$

and so

$$u'_x(x_0, y_0) = \frac{u(x_0 + h, y_0) - u(x_0, y_0)}{h} + O(h) \text{ (error)}$$

Thus, the finite difference formula for the first derivative is

$$u'_x(x_0, y_0) = \frac{u(x_0 + h, y_0) - u(x_0, y_0)}{h} + O(h).$$

Dropping the term $O(h)$ and using $u_{i,j}$ for $u(x_i, y_j)$, $i = 0, 1, 2, \dots$, we get

$$u'_x \approx \frac{u_{i+1,j} - u_{i,j}}{h}, \quad (11.1)$$

which is called forward difference approximation to u_x . Similarly, expanding $u(x_0 - h, y_0)$ by Taylor series, we get

$$u_x(x_0, y_0) \approx \frac{u(x_0, y_0) - u(x_0 - h, y_0)}{h}$$

or

$$u_x \approx \frac{u_{i,j} - u_{i-1,j}}{h}, \quad (11.2)$$

which is known as backward difference approximation to u_x .

Now using formula (11.2), we get by Taylor's series,

$$u_{xx} \approx \frac{u(x_0 + h, y_0) - 2u(x_0, y_0) + u(x_0 - h, y_0)}{h^2}$$

or

$$u_{xx} \approx \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} \quad (11.3)$$

as the difference approximation to u_{xx} .

Similarly, we have the approximation with $y = ik$, $k = 0, 1, 2, \dots$,

$$u_y \approx \frac{u_{i,j+1} - u_{i,j}}{k}, \quad (11.4)$$

$$u_y \approx \frac{u_{i,j} - u_{i,j-1}}{k}, \quad (11.5)$$

and

$$u_{yy} \approx \frac{u(x_0, y_0 + k) - 2u(x_0, y_0) + u(x_0, y_0 - k)}{k^2},$$

or

$$u_{yy} \approx \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{k^2}.$$

11.2 GEOMETRIC REPRESENTATION OF PARTIAL DIFFERENCE QUOTIENTS

Let (x, y) plane be partitioned into a network of rectangles of sides $\Delta x = h$ and $\Delta y = k$ by drawing the sets of lines

$$x = ih, \quad i = 0, 1, 2, 3, \dots,$$

$$y = jk, \quad j = 0, 1, 2, 3, \dots$$

The points of intersection of these families of lines are called mesh points, grid points, or lattice points. Thus, the points (x, y) , $(x + h, y)$, $(x + 2h, y)$, $(x - h, y)$, $(x - 2h, y)$, \dots , $(x, y - k)$, $(x, y - 2k)$, \dots are the grid points as shown in Figure 11.1.

If we represent (x, y) by (i, j) , then $(x + h, y) = (i + 1, j)$, $(x + 2h, y) = (i + 2, j)$ and so on.

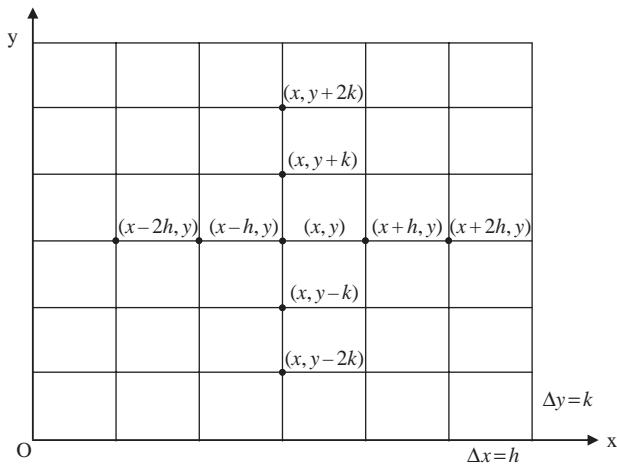


Figure 11.1

11.3 STANDARD FIVE POINT FORMULA AND DIAGONAL FIVE POINT FORMULA

Consider the Laplace equation in two dimensions,

$$u_{xx} + u_{yy} = 0.$$

Its finite difference analog is

$$\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{k^2} = 0. \quad (11.6)$$

If we consider square mesh, that is, $h = k$, then equation (11.6) yields

$$u_{i,j} = \frac{1}{4}(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}). \quad (11.7)$$

Equation (11.7) shows that the value of u at any point is the mean of its values at the four neighboring points as shown in Figure 11.2.

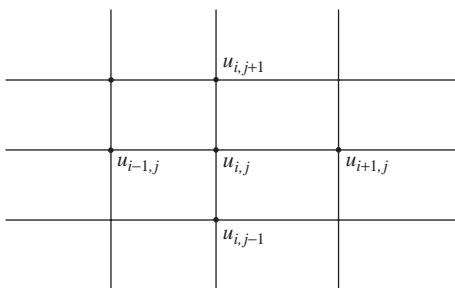
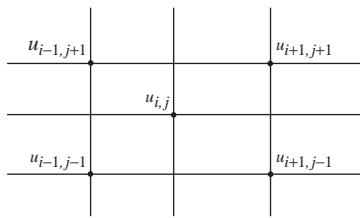


Figure 11.2 (Standard five point formula)

Formula (11.7) is called standard five point formula.

If we rotate the coordinate axes through 45° , then the Laplace equation remains invariant. In fact, $X = x \cos \theta + y \sin \theta$, $Y = x \sin \theta - y \cos \theta$, where $\theta = 45^\circ$, then $u_{xx} + u_{yy} = 0$. Therefore, we may use the function values at the diagonal points (Figure 11.3) in place of the neighboring points. Then we may use the formula

**Figure 11.3 (Diagonal Five Point Formula)**

$$u_{i,j} = \frac{1}{4}(u_{i-1,j-1} + u_{i+1,j-1} + u_{i+1,j+1} + u_{i-1,j+1})$$

in place of formula (11.7). This formula is called the diagonal five point formula.

11.4 POINT JACOBI'S METHOD

Let $u_{i,j}^{(n)}$ be the n th iterative value of $u_{i,j}$. Then the iterative procedure to solve (11.7) is

$$u_{i,j}^{(n+1)} = \frac{1}{4}(u_{i-1,j}^{(n+1)} + u_{i+1,j}^{(n+1)} + u_{i,j-1}^{(n+1)} + u_{i,j+1}^{(n)})$$

for the interior mesh points. This procedure is called the point Jacobi method.

11.5 GAUSS–SEIDEL METHOD

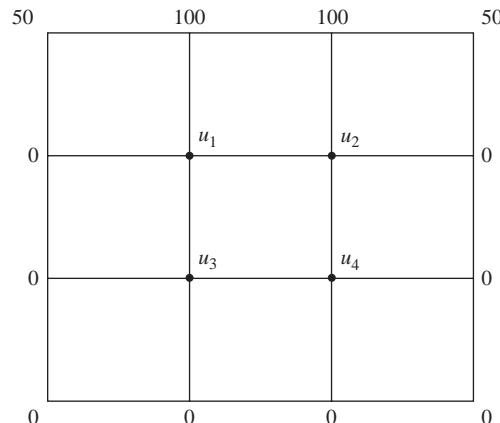
This method uses the latest iterative values available and scans the mesh points systematically from left to right along successive rows. The formula is

$$u_{i,j}^{(n+1)} = \frac{1}{4} \left[u_{i-1,j}^{(n+1)} + u_{i+1,j}^{(n+1)} + u_{i,j-1}^{(n+1)} + u_{i,j+1}^{(n)} \right].$$

It can be shown that the Gauss–Seidel method converges twice as fast as Jacobi's method.

EXAMPLE 11.1

Solve Laplace equation $u_{xx} + u_{yy} = 0$ for the following square meshes with boundary conditions exhibited in Figure 11.4.

**Figure 11.4**

Solution. Using diagonal five point formula, we have

$$u_1 = \frac{1}{4}[0 + 100 + 50 + u_4], \quad (11.8)$$

$$u_2 = \frac{1}{4}[100 + 0 + u_3 + 50], \quad (11.9)$$

$$u_3 = \frac{1}{4}[0 + u_2 + 0 + 0], \quad (11.10)$$

$$u_4 = \frac{1}{4}[0 + 0 + 0 + u_1]. \quad (11.11)$$

From equations (11.10) and (11.11), we have

$$u_3 = \frac{1}{4}u_2,$$

$$u_4 = \frac{1}{4}u_1.$$

Then equation (11.8) yields

$$u_1 = \frac{1}{4}\left(150 + \frac{1}{4}u_1\right),$$

or

$$\frac{15}{16}u_1 = \frac{150}{4}$$

or

$$u_1 = 40 \text{ and so } u_4 = 10.$$

Similarly, equation (11.9) yields

$$u_2 = \frac{1}{4}[150 + u_3] = \frac{1}{4}\left(150 + \frac{1}{4}u_2\right) = \frac{150}{4} - \frac{1}{16}u_2.$$

Therefore,

$$u_2 = 40 \text{ and so } u_3 = \frac{1}{4}u_2 = 10.$$

Hence, the first approximation is

$$u_1 = 40, u_2 = 40, u_3 = 10, u_4 = 10.$$

Now using Jacobi's method, we have

$$u_1^{(1)} = \frac{1}{4}[0 + u_2 + u_3 + 100] = \frac{1}{4}[0 + 40 + 10 + 100] = \frac{150}{4} = 37.5,$$

$$u_2^{(1)} = \frac{1}{4}[40 + 10 + 0 + 100] = \frac{150}{4} = 37.5,$$

$$u_3^{(1)} = \frac{1}{4}[0 + 0 + 40 + 10] = \frac{50}{4} = 12.5,$$

$$u_4^{(1)} = \frac{1}{4}[10 + 0 + 0 + 40] = \frac{50}{4} = 12.5.$$

The next approximation is

$$u_1^{(2)} = \frac{1}{4}[0 + 12.5 + 37.5 + 100] = \frac{150}{4} = 37.5,$$

$$u_2^{(2)} = \frac{1}{4}[37.5 + 0 + 12.5 + 100] = \frac{150}{4} = 37.5,$$

$$u_3^{(2)} = \frac{1}{4}[0 + 0 + 1.25 + 37.5] = \frac{50}{4} = 12.5,$$

$$u_4^{(2)} = \frac{1}{4}[0 + 0 + 1.25 + 37.5] = \frac{50}{4} = 12.5.$$

Hence, the solution is

$$u_1 = 37.5, u_2 = 37.5, u_3 = 12.5, u_4 = 12.5.$$

EXAMPLE 11.2

Solve Laplace equation

$$u_{xx} + u_{yy} = 0$$

for the square meshes with the boundary values shown in Figure 11.5.

Solution. Using diagonal five point formula, we have

$$u_1 = \frac{1}{4}[2 + 2 + u_4 + u_5], \quad (11.12)$$

$$u_2 = \frac{1}{4}[1 + 5 + u_3 + u_6], \quad (11.13)$$

$$u_3 = \frac{1}{4}[1 + 5 + u_2 + u_7], \quad (11.14)$$

$$u_4 = \frac{1}{4}[4 + 4 + u_1 + u_8]. \quad (11.15)$$

If we use standard five point formula, we have

$$u_1 = \frac{1}{4}[1 + u_2 + u_3 + 1], \quad (11.16)$$

$$u_2 = \frac{1}{4}[4 + 2 + u_1 + u_4], \quad (11.17)$$

$$u_3 = \frac{1}{4}[2 + 4 + u_4 + u_1], \quad (11.18)$$

$$u_4 = \frac{1}{4}[5 + 5 + u_2 + u_3]. \quad (11.19)$$

Expressions (11.17) and (11.18) show that $u_2 = u_3$. Therefore, expressions (11.16), (11.17), (11.18), and (11.19) reduce to

$$u_1 = \frac{1}{4}[2 + 2u_2],$$

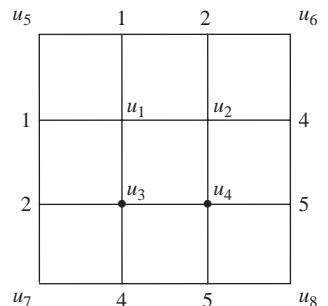


Figure 11.5

$$u_2 = \frac{1}{4}[6 + u_1 + u_4],$$

$$u_3 = \frac{1}{4}[6 + u_1 + u_4],$$

$$u_4 = \frac{1}{4}[10 + 2u_2].$$

If we start with the approximation $u_2 = 0$, then

$$u_1 = \frac{1}{2}, u_2 = 0, u_3 = 0, u_4 = \frac{5}{2}.$$

Then, by Gauss–Seidel's method, we have

$$u_1^{(1)} = \frac{1}{4}[1 + 0 + 1 + 0] = \frac{1}{2} = 0.5,$$

$$u_2^{(1)} = \frac{1}{4}\left[\frac{1}{2} + \frac{5}{2} + 4 + 2\right] = \frac{9}{4} = 2.25,$$

$$u_3^{(1)} = \frac{1}{4}\left[2 + 4 + \frac{5}{2} + \frac{1}{2}\right] = \frac{9}{4} = 2.25,$$

$$u_4^{(1)} = \frac{1}{4}\left[5 + 5 + \frac{9}{4} + \frac{9}{4}\right] = \frac{5}{2} = 2.50,$$

$$u_1^{(2)} = \frac{1}{4}[1 + 2.25 + 1 + 2.25] = 1.625,$$

$$u_2^{(2)} = \frac{1}{4}[1.625 + 4 + 2.50 + 2] = 2.53125,$$

$$u_3^{(2)} = \frac{1}{4}[2 + 4 + 1.625 + 2.50] = 2.53125,$$

$$u_4^{(2)} = \frac{1}{4}[5 + 5 + 2.53125 + 2.53125] = 3.765625,$$

$$u_1^{(3)} = \frac{1}{4}[1 + 1 + 2.53125 + 2.53125] = 1.765625,$$

$$u_2^{(3)} = \frac{1}{4}[4 + 2 + 1.765625 + 3.765625] = 2.8828125,$$

$$u_3^{(3)} = 2.8828125, \text{ since } u_2 = u_3.$$

$$u_4^{(3)} = \frac{1}{4}[5 + 5 + 2.8828125 + 2.8828125] = 3.9414031,$$

$$u_1^{(4)} = \frac{1}{4}[1 + 1 + 2.8828125 + 2.8828125] = 1.94140625,$$

$$u_2^{(4)} = \frac{1}{4}[2 + 4 + 1.94140625 + 3.9414031] = 2.970702338,$$

$$u_3^{(4)} = 2.970702338, \text{ since } u_2 = u_3.$$

$$u_4^{(4)} = \frac{1}{4}[5 + 5 + 2.970702338 + 2.970702338] = 3.985351169,$$

$$u_1^{(5)} = \frac{1}{4}[1 + 1 + 2.970702338 + 2.970702338] = 1.985351169,$$

$$u_2^{(5)} = \frac{1}{4}[2 + 4 + 1.985351169 + 3.985351169] = 2.992675585,$$

$$u_3^{(5)} = 2.992675585, \text{ since } u_2 = u_3.$$

$$u_4^{(5)} = \frac{1}{4}[5 + 5 + 2.992675585 + 2.992675585] = 3.996337793,$$

$$u_1^{(6)} = \frac{1}{4}[1 + 1 + 2.992675585 + 2.992675585] = 1.996337793,$$

$$u_2^{(6)} = \frac{1}{4}[2 + 4 + 1.996337793 + 3.996337793] = 2.998168897,$$

$$u_3^{(6)} = 2.998168897, \text{ since } u_2 = u_3.$$

$$u_4^{(6)} = \frac{1}{4}[5 + 5 + 2.998168897 + 2.998168897] = 3.999084449,$$

$$u_1^{(7)} = \frac{1}{4}[1 + 1 + 2.998168897 + 2.998168897] = 1.999084449,$$

$$u_2^{(7)} = \frac{1}{4}[2 + 4 + 1.999084449 + 3.999084449] = 2.999542224,$$

$$u_3^{(7)} = 2.999542224, \text{ since } u_2 = u_3.$$

$$u_4^{(7)} = \frac{1}{4}[5 + 5 + 2.999542224 + 2.999542224] = 3.999771112.$$

We observe that the values of 6th and 7th iterations agree up to two decimal places. Hence,

$$u_1 = 1.99, \quad u_2 = 2.99,$$

$$u_3 = 2.99, \quad u_4 = 3.99.$$

EXAMPLE 11.3

Using the given boundary values, solve the Laplace equation $\nabla^2 u = 0$ at the nodal points of the square grid shown in Figure 11.6.

Solution. We assume that $u_4 = 0$. Then the initial approximation is

$$u_1 = \frac{1}{4}[20 + 60 + 60 + 0] = 35 \text{ (diagonal five point formula),}$$

$$u_2 = \frac{1}{4}[35 + 60 + 50 + 0] = 36.25 \text{ (standard five point formula),}$$

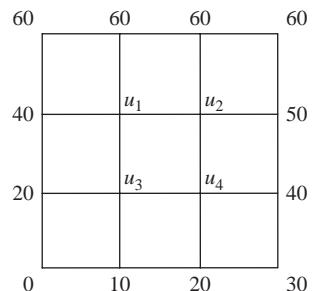


Figure 11.6

$$u_3 = \frac{1}{4}[35 + 20 + 10 + 0] = 16.25 \text{ (standard five point formula),}$$

$$u_4 = \frac{1}{4}[36.25 + 16.25 + 20 + 40] = 28.125 \text{ (standard five point formula).}$$

Now using Gauss–Seidel's method, we have

$$u_1^{(1)} = \frac{1}{4}[60 + 40 + u_2 + u_3] = \frac{1}{4}[100 + 36.25 + 16.25] = 38.125,$$

$$u_2^{(1)} = \frac{1}{4}[60 + 50 + u_1^{(1)} + u_4] = \frac{1}{4}[110 + 38.125 + 28.125] = 44.0625,$$

$$u_3^{(1)} = \frac{1}{4}[20 + 10 + u_1^{(1)} + u_4] = \frac{1}{4}[30 + 38.125 + 28.125] = 24.0625,$$

$$u_4^{(1)} = \frac{1}{4}[40 + 20 + u_2^{(1)} + u_3^{(1)}] = \frac{1}{4}[60 + 44.0625 + 24.0625] = 32.03125,$$

$$u_1^{(2)} = \frac{1}{4}[100 + 44.0625 + 24.0625] = 42.03125,$$

$$u_2^{(2)} = \frac{1}{4}[110 + 42.03125 + 32.03125] = 46.015625,$$

$$u_3^{(2)} = \frac{1}{4}[30 + 42.03125 + 32.03125] = 26.015625,$$

$$u_4^{(2)} = \frac{1}{4}[60 + 46.015625 + 26.015625] = 33.0078125,$$

$$u_1^{(3)} = \frac{1}{4}[100 + 46.015625 + 26.015625] = 43.0078125,$$

$$u_2^{(3)} = \frac{1}{4}[110 + 43.0078125 + 33.0078125] = 46.50390625,$$

$$u_3^{(3)} = \frac{1}{4}[30 + 43.0078125 + 33.0078125] = 26.50390625,$$

$$u_4^{(3)} = \frac{1}{4}[60 + 46.50390625 + 26.50390625] = 33.25195311,$$

$$u_1^{(4)} = \frac{1}{4}[100 + 46.50390625 + 26.50390625] = 43.25195311,$$

$$u_2^{(4)} = \frac{1}{4}[110 + 43.25195311 + 33.25195311] = 46.62597655,$$

$$u_3^{(4)} = \frac{1}{4}[30 + 43.25195311 + 33.25195311] = 26.625997656,$$

$$u_4^{(4)} = \frac{1}{4}[60 + 46.62597655 + 26.62597656] = 33.31298827,$$

$$u_1^{(5)} = \frac{1}{4}[100 + 46.62597655 + 26.62597656] = 43.31298827,$$

$$u_2^{(5)} = \frac{1}{4}[110 + 43.31298827 + 33.31298827] = 46.65649412,$$

$$u_3^{(5)} = \frac{1}{4}[30 + 43.31298827 + 33.31298827] = 26.65649414,$$

$$u_4^{(5)} = \frac{1}{4}[60 + 46.65649412 + 26.65649414] = 33.32824706.$$

Hence $u_1 = 43.313$, $u_2 = 46.656$, $u_4 = 33.328$.

EXAMPLE 11.4

Solve the Laplace equation $u_{xx} + u_{yy} = 0$ for the square mesh with boundary values shown in Figure 11.7.

Solution. We assume that $u_4 = 0$. Then the first approximation is

$$u_1 = \frac{1}{4}[1 + 2 + 0 + 0] = 0.75 \text{ (diagonal five point formula),}$$

$$u_2 = \frac{1}{4}[0.75 + 2 + 2 + 0] = 1.1875 \text{ (standard five point formula),}$$

$$u_3 = \frac{1}{4}[0 + 0 + 0.75 + 0] = 0.1875 \text{ (standard five point formula),}$$

$$u_4 = \frac{1}{4}[0.1875 + 2 + 0 + 1.1875] = 0.84375 \text{ (standard five point formula).}$$

Now using Gauss–Seidel's method, we have

$$u_1^{(1)} = \frac{1}{4}[0 + 1.1875 + 0.1875 + 2] = 0.84375,$$

$$u_2^{(1)} = \frac{1}{4}[0.84375 + 2 + 2 + 0.84375] = 1.421875,$$

$$u_3^{(1)} = \frac{1}{4}[0 + 0.84375 + 0 + 0.84375] = 0.421875,$$

$$u_4^{(1)} = \frac{1}{4}[0.421875 + 2 + 0 + 1.421875] = 0.9609375,$$

$$u_1^{(2)} = \frac{1}{4}[0 + 2 + 1.421875 + 0.421875] = 0.9609375,$$

$$u_2^{(2)} = \frac{1}{4}[0.9609375 + 2 + 2 + 0.9609375] = 1.48046875,$$

$$u_3^{(2)} = \frac{1}{4}[0 + 0.9609375 + 0 + 0.9609375] = 0.48046875,$$

$$u_4^{(2)} = \frac{1}{4}[0.4804675 + 2 + 0 + 1.48046875] = 0.990234375,$$

$$u_1^{(3)} = \frac{1}{4}[0 + 1.48046875 + 0.48046875 + 2] = 0.990234375,$$

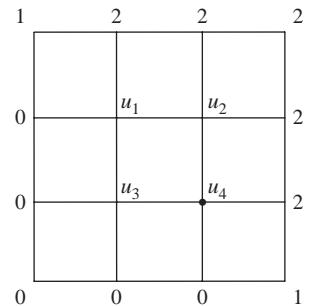


Figure 11.7

$$u_2^{(3)} = \frac{1}{4}[0.990234375 + 2 + 0.990234375 + 2] = 1.495117188,$$

$$u_3^{(3)} = \frac{1}{4}[0 + 0.990234375 + 0 + 0.990234375] = 0.495117188,$$

$$u_4^{(3)} = \frac{1}{4}[0.495117187 + 2 + 0 + 1.495117187] = 0.997558593,$$

$$u_1^{(4)} = \frac{1}{4}[0 + 1.495117188 + 2 + 0.495117188] = 0.997558594,$$

$$u_2^{(4)} = \frac{1}{4}[0.997558594 + 2 + 2 + 0.997558593] = 1.498779297,$$

$$u_3^{(4)} = \frac{1}{4}[0 + 0.997558593 + 0 + 0.997558594] = 0.498779296,$$

$$u_4^{(4)} = \frac{1}{4}[0.498779296 + 2 + 0 + 1.498779297] = 0.999389648.$$

Hence, up to two decimal places, we have

$$u_1 = 0.99, u_2 = 1.49, u_3 = 0.49, u_4 = 0.99.$$

EXAMPLE 11.5

Solve the elliptic equation $u_{xx} + u_{yy} = 0$ for the square mesh with boundary values shown in Figure 11.8.

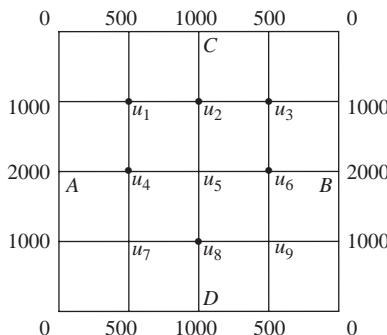


Figure 11.8

Solution. We observe that the figure is symmetrical about AB and so

$$u_1 = u_7, u_2 = u_8, u_3 = u_9.$$

Similarly, the figure's symmetry about CD yields

$$u_1 = u_3, u_4 = u_6, u_7 = u_9.$$

Thus,

$$u_1 = u_3 = u_7 = u_9, u_2 = u_8, u_4 = u_6.$$

Hence, it is sufficient to find u_1, u_2, u_4 , and u_5 . To get initial values, we have

$$u_5 = \frac{1}{4}[2000 + 2000 + 1000 + 1000] = 1500 \text{ (standard five point formula),}$$

$$u_1 = \frac{1}{4}[0 + 1000 + 2000 + 1500] = 1125 \text{ (diagonal five point formula),}$$

$$u_2 = \frac{1}{4}[1125 + 1125 + 1000 + 1500] = 1187.5 \text{ (standard five point formula),}$$

$$u_4 = \frac{1}{4}[2000 + 1500 + 1125 + 1125] = 1437.5 \text{ (standard five point formula).}$$

Now we use Gauss–Seidel's method to improve the values and get

$$u_1^{(1)} = \frac{1}{4}[1000 + 1187.5 + 500 + 1437.5] = 1031.25,$$

$$u_2^{(1)} = \frac{1}{4}[1031.25 + 1000 + 1500 + 1031.25] = 1140.625,$$

$$u_4^{(1)} = \frac{1}{4}[2000 + 1500 + 1031.25 + 1031.25] = 1390.625,$$

$$u_5^{(1)} = \frac{1}{4}[1390.625 + 1390.625 + 1140.625 + 1140.625] = 1265.625.$$

After nine iterations, we shall obtain

$$u_1 = u_3 = u_7 = u_9 = 938.05,$$

$$u_2 = u_8 = 1000.55,$$

$$u_4 = u_6 = 1250.55,$$

$$u_5 = 1125.55.$$

EXAMPLE 11.6

Determine the system of four equations in four unknowns p_1, p_2, p_3 , and p_4 for computing approximation for the harmonic function $u(x, y)$ in the rectangle $R = \{(x, y) : 0 \leq x \leq 3, 0 \leq y \leq 3\}$ shown in Figure 11.9, under the conditions

$$u(x, 0) = 10, \quad u(x, 3) = 90 \text{ for } 0 < x < 3$$

$$u(0, y) = 70, \quad u(3, y) = 0 \text{ for } 0 < y < 3.$$

Hence find p_1, p_2, p_3 , and p_4 .

Solution. We want to solve $u_{xx} + u_{yy} = 0$ under the conditions given in the problem. Taking $h = k = 1$, the square mesh is shown in Figure 11.9.

Therefore, using standard five point formula, we have

$$-4p_1 + p_2 + p_3 + 0p_4 = -80,$$

$$p_1 - 4p_2 + 0p_3 + p_4 = -10,$$

$$p_1 - 0p_2 - 4p_3 + p_4 = -160,$$

$$0p_1 + p_2 + p_3 - 4p_4 = -90.$$

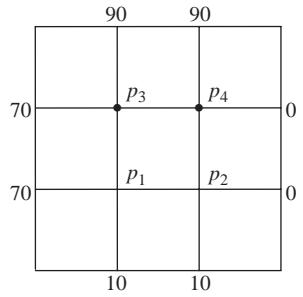


Figure 11.9

We solve this system by Gaussian elimination method. The augmented matrix is

$$\text{Pivot} \rightarrow \left[\begin{array}{cccc|c} -4 & 1 & 1 & 0 & -80 \\ 1 & -4 & 0 & 1 & -10 \\ 1 & 0 & -4 & 1 & -160 \\ 0 & 1 & 1 & -4 & -90 \end{array} \right]$$

$$m_{21} = -1/4$$

$$m_{31} = -1/4$$

The result after first elimination is

$$\text{Pivot} \rightarrow \left[\begin{array}{cccc|c} -4 & 1 & 1 & 0 & -80 \\ 0 & \frac{-15}{4} & \frac{1}{4} & 1 & -30 \\ 1 & \frac{1}{4} & \frac{-15}{4} & 1 & -180 \\ 0 & 1 & 1 & -4 & -90 \end{array} \right]$$

$$m_{32} = -1/15$$

$$m_{42} = -4/15$$

The second elimination yields

$$\text{Pivot} \rightarrow \left[\begin{array}{cccc|c} -4 & 1 & 1 & 0 & -80 \\ 0 & \frac{-15}{4} & \frac{1}{4} & 1 & -30 \\ 0 & 0 & \frac{-56}{15} & \frac{16}{15} & -182 \\ 0 & 0 & \frac{16}{15} & \frac{-56}{15} & -98 \end{array} \right]$$

$$m_{43} = -2/7$$

The third elimination yields

$$\left[\begin{array}{cccc|c} -4 & 1 & 1 & 0 & -80 \\ 0 & \frac{-15}{4} & \frac{1}{4} & 1 & -30 \\ 0 & 0 & \frac{-56}{15} & \frac{16}{15} & -182 \\ 0 & 0 & 0 & \frac{-24}{7} & \frac{-1050}{7} \end{array} \right]$$

Back substitution yields

$$p_4 = \frac{1050}{24} = 43.75.$$

$$-\frac{56}{15} p_3 + \frac{16}{15} p_4 = -182 \text{ and so } p_3 = 61.16.$$

$$-\frac{15}{4} p_2 + \frac{1}{4} p_3 + p_4 = -30 \text{ and so } p_2 = 23.61.$$

and

$$-4 p_1 + p_2 + p_3 = -80 \text{ and so } p_1 = 41.19.$$

11.6 SOLUTION OF ELLIPTIC EQUATION BY RELAXATION METHOD

Consider the Laplace equation $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$. We want to solve it in a square region with mesh size h . Let u_0 be the value of u at a grid point A and let u_1, u_2, u_3, u_4 be the values of u at four adjacent grid points (Figure 11.10). The difference equation of the given Laplace equation is

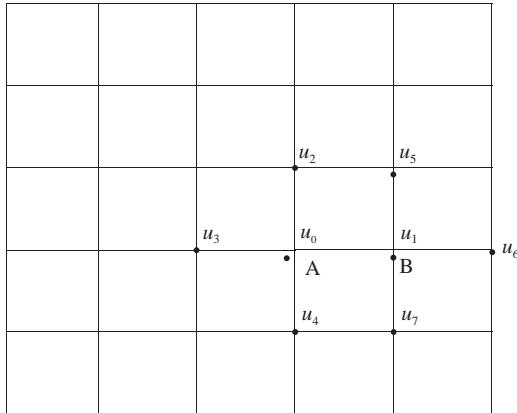


Figure 11.10

$$\frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} + \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{h^2} + O(h^2) = 0.$$

Thus, for the grid point A , we have

$$u_1 + u_2 + u_3 + u_4 - 4u_0 \approx 0.$$

Therefore, the residual at A is

$$r_A = u_1 + u_2 + u_3 + u_4 - 4u_0$$

Similarly, the residue at B is

$$r_B = u_0 + u_6 + u_5 + u_7 - 4u_1$$

and so on for other grid points.

The aim of relaxation method is to reduce all the residuals to zero or a very small quantity. To do so we try to annihilate the values of u at the internal mesh points.

When the value of u is changed at a mesh point, the value of the residuals at the adjacent interior point will also change. For example, when an increment of 1 is given to u_0 , then r_A is changed by -4 , while each residual at the adjacent mesh points are changed by 1 unit.

The working procedure of relaxation method is demonstrated in the following example.

EXAMPLE 11.7

Solve by relaxation method, the Laplace equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

inside a square region bounded by the lines $x = 0, x = 4, y = 0, y = 4$ given that $u = x^2 y^2$ on the boundary.

Solution. Take $h = k = 1$. The boundary values using $u = x^2 y^2$ are shown in Figure 11.11.

By standard five point formula, the value of u at E is

$$\frac{0+64+0+6}{4} = 32.$$

Then using diagonal five point formula, the value of u at G is

$$\frac{0+0+0+32}{4} = 8.$$

In a similar way, using standard five point formula or diagonal five point formula, we find the values at A, B, C, D, E, F, G, H, and I to be 24, 56, 104, 16, 32, 56, 8, 16, and 24, respectively. Then residuals at these points are

$$r_A = 0 + 56 + 16 + 16 - 4(24) = -8,$$

$$r_B = 24 + 104 + 64 + 32 - 4(56) = 0,$$

$$r_C = -16, r_D = 0, r_E = 16, r_F = 0, r_G = 0, r_H = 0, r_I = 8.$$

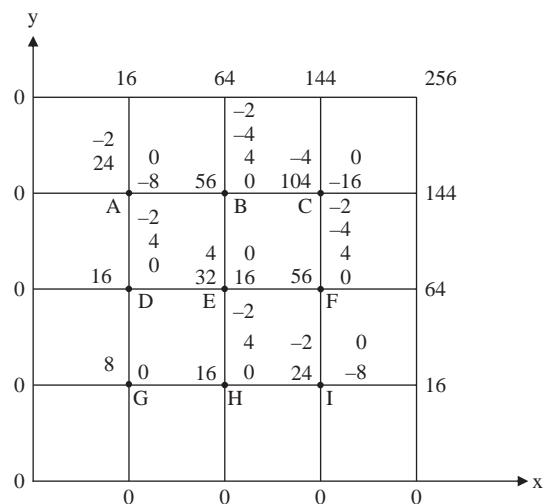


Figure 11.11

The numerically largest residual is at E and $r_E = 16$. To annihilate it, we increase u by 4 so that the residual becomes zero at E. The residuals at the adjoining nodes are increased by 4.

Now, the numerically largest residual is -16 at C. So we increase u by -4 so that the residual becomes zero. The residuals at the adjoining nodes are increased by -4 . Next the numerically largest residual is -8 at A. So we increase u by -2 so that residual at A is zero and the residuals at the adjoining nodes are increased by -2 .

Finally, the largest residual is -8 at I. So we increase u by -2 so that residual at I becomes zero and the residuals at the adjoining nodes is increased by -2 .

We may now stop the process since the numerically largest residual at this stage is only 2. The answer is shown in Figure 11.12 below.

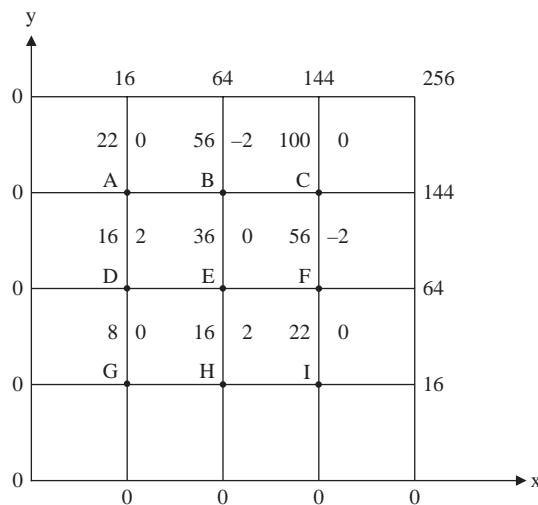


Figure 11.12

Thus, we have

$$u_A = 22, u_B = 56, u_C = 100,$$

$$u_D = 16, u_E = 36, u_F = 56,$$

$$u_G = 8, u_H = 16, u_I = 22.$$

EXAMPLE 11.8

Solve by relaxation method, the Laplace equation $u_{xx} + u_{yy} = 0$ in the square region $x = 0$ to $x = 1$ and $y = 0$ to $y = 1$ shown in Figure 11.13 starting with the values $u_1 = u_2 = u_3 = u_4 = 1$.

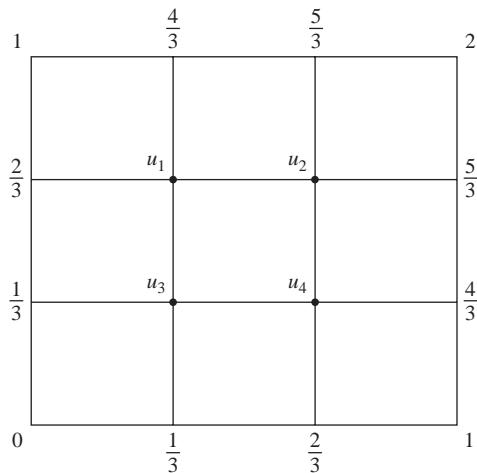


Figure 11.13

Solution. The grid points and residues are shown in Figure 11.14. We have

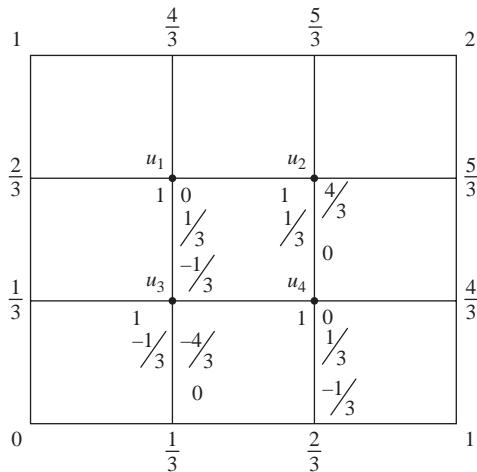


Figure 11.14

$$\begin{aligned}
 r_{u_1} &= \frac{2}{3} + 1 + 1 + \frac{4}{3} - 4(1) = 0, \\
 r_{u_2} &= 1 + 1 + \frac{5}{3} + \frac{5}{3} - 4(1) = \frac{4}{3}, \\
 r_{u_3} &= \frac{1}{3} + \frac{1}{3} + 1 + 1 - 4(1) = -\frac{4}{3}, \\
 r_{u_4} &= 1 + \frac{4}{3} + \frac{2}{3} + 1 - 4(1) = 0.
 \end{aligned}$$

Using relaxation method, the answer is shown in Figure 11.15.

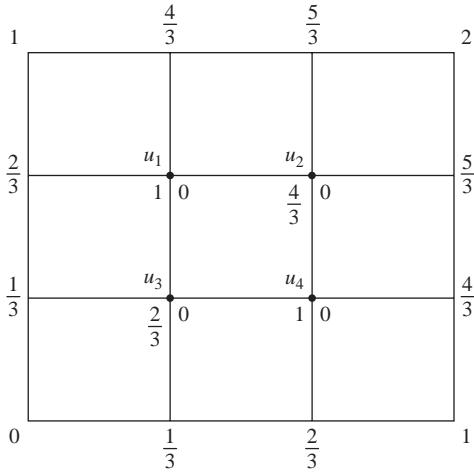


Figure 11.15

Hence, the solution is

$$u_1 = 1, \quad u_2 = \frac{4}{3} = 1.333, \quad u_3 = \frac{2}{3} = 0.666, \quad u_4 = 1.$$

11.7 POISSON'S EQUATION

The elliptic partial differential equation

$$u_{xx} + u_{yy} = f(x, y), \quad (11.20)$$

where $f(x, y)$ is a given function of x and y , is called the Poisson's equation.

The Poisson's equation is solved numerically by replacing the derivatives by difference expressions at the points $x = ih$, $y = jh$. Thus, we have

$$\frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} + \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{h^2} = f(ih, jh)$$

or

$$u_{i-1,j} - 4u_{i,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1} = h^2 f(ih, jh). \quad (11.21)$$

The error involved in equation (11.21) is $O(h^2)$.

EXAMPLE 11.9

Solve the Poisson's equation

$$u_{xx} + u_{yy} = -10(x^2 + y^2 + 10)$$

over the square with sides $x = 0 = y$, $x = 3 = y$ with $u = 0$ on the boundary and mesh length 1.

Solution. Since mesh length is 1 and side of the square is 3, Figure 11.16 of the problem is

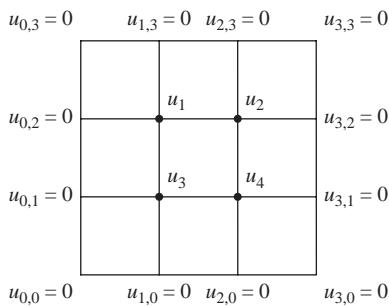


Figure 11.16

By standard formula (11.21), we have

$$u_{i-1,j} - 4u_{i,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1} = h^2 f(ih, jh).$$

For u_1 , we have $i = 1$ and $j = 2$ and so the formula gives

$$u_{0,2} - 4u_{1,2} + u_{2,2} + u_{1,1} + u_{1,3} = h^2 f(h, 2h)$$

or

$$0 - 4u_1 + u_2 + u_3 + 0 = f(1, 2) = -10(1 + 4 + 10)$$

or

$$u_1 = \frac{1}{4}(u_2 + u_3 + 150).$$

Now, for u_2 , we have $i = 2$ and $j = 2$. Therefore, the formula yields

$$u_2 = \frac{1}{4}(u_1 + u_4 + 180).$$

For u_3 , we have $i = 1$ and $j = 1$ and so the formula yields

$$u_3 = \frac{1}{4}(u_1 + u_4 + 120).$$

For u_4 , we have $i = 2$ and $j = 1$ and so the formula yields

$$u_4 = \frac{1}{4}(u_2 + u_3 + 150).$$

We observe that $u_1 = u_4$. Therefore,

$$u_1 = \frac{1}{4}(u_2 + u_3 + 150),$$

$$u_2 = \frac{1}{4}(u_1 + u_4 + 180),$$

$$u_3 = \frac{1}{4}(u_1 + u_4 + 120).$$

We start with $u_2 = u_3 = 0$ and use Gauss–Seidel's method to improve the values. We have

$$u_1^{(1)} = \frac{1}{4}(0 + 0 + 150) = 37.5,$$

$$u_2^{(1)} = \frac{1}{4} [2(37.5) + 180] = 63.75,$$

$$u_3^{(1)} = \frac{1}{4} [2(37.5) + 120] = 48.75,$$

$$u_1^{(2)} = \frac{1}{4} [63.75 + 48.75 + 150] = 65.625,$$

$$u_2^{(2)} = \frac{1}{4} [2(65.625) + 180] = 77.8125,$$

$$u_3^{(2)} = \frac{1}{4} [2(65.625) + 120] = 62.8125,$$

$$u_1^{(3)} = \frac{1}{4} [77.8125 + 62.8125 + 150] = 72.65625,$$

$$u_2^{(3)} = \frac{1}{4} [2(72.65625) + 180] = 81.328125,$$

$$u_3^{(3)} = \frac{1}{4} [2(72.65625) + 120] = 66.328125,$$

$$u_1^{(4)} = \frac{1}{4} [81.328125 + 66.328125 + 150] = 74.4140625,$$

$$u_2^{(4)} = \frac{1}{4} [2(74.4140625) + 180] = 82.20703125,$$

$$u_3^{(4)} = \frac{1}{4} [2(74.4140625) + 120] = 67.20703125,$$

$$u_1^{(5)} = \frac{1}{4} [82.2070125 + 67.20703125 + 150] = 74.8535,$$

$$u_2^{(5)} = \frac{1}{4} [2(74.8535) + 180] = 82.4268,$$

$$u_3^{(5)} = \frac{1}{4} [2(74.8535) + 120] = 67.4268,$$

$$u_1^{(6)} = \frac{1}{4} [82.4268 + 67.4268 + 150] = 74.9634,$$

$$u_2^{(6)} = \frac{1}{4} [2(74.9634) + 180] = 82.4817,$$

$$u_3^{(6)} = \frac{1}{4} [2(74.9634) + 120] = 67.4817.$$

The values obtained by fifth and sixth iteration are nearly equal and so the solution is

$$u_1 \approx 74.9, u_2 \approx 82.5, u_3 \approx 67.5, u_4 = u_1 = 74.9.$$

EXAMPLE 11.10

The function ϕ satisfies the equation

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + 2 = 0$$

at every point inside the square bounded by the straight lines $x = \pm 1$, $y = \pm 1$, and is zero on the boundary. Calculate a finite difference solution using a square mesh of side $\frac{1}{2}$. Assuming that error is $O(h^2)$, calculate the improved value of ϕ at $(0,0)$.

(The example is the non-dimensional form of the torsion problem for a solid elastic cylinder with a square cross-section.)

Solution. The mesh points and the boundary values are shown in Figure 11.17.

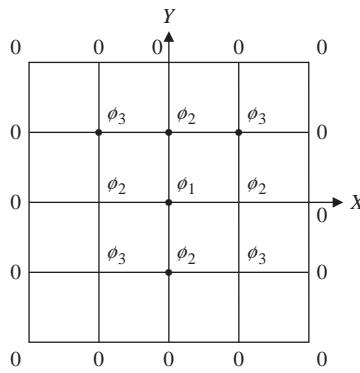


Figure 11.17

Because of the symmetry w.r.t. x -axis, y -axis, and the diagonals, there are only three unknowns, ϕ_1 at $(0,0)$, ϕ_2 at $\left(\frac{1}{2}, 0\right)$ and ϕ_3 at $\left(\frac{1}{2}, \frac{1}{2}\right)$. The difference equation for the given problem is

$$\frac{\phi(x_0 - h, y_0) - 2\phi(x_0, y_0) + \phi(x_0 + h, y_0)}{h^2} + \frac{\phi(x_0, y_0 - h) - 2\phi(x_0, y_0) + \phi(x_0, y_0 + h)}{h^2} + 2 = 0$$

or

$$\phi(x_0 - h, y_0) - 4\phi(x_0, y_0) + \phi(x_0 + h, y_0) + \phi(x_0, y_0 - h) + \phi(x_0, y_0 + h) + 2h^2 = 0.$$

Taking $h = \frac{1}{2}$, the above formula yields

$$\begin{aligned} 2\phi_2 - 8\phi_1 + 2\phi_2 + 2\phi_2 + 1 &= 0, \\ 2\phi_1 - 8\phi_2 + 0 + 2\phi_3 + 2\phi_3 + 1 &= 0, \\ 2\phi_2 - 8\phi_3 + 0 + 2\phi_2 + 0 + 1 &= 0. \end{aligned}$$

Thus, we have

$$\begin{aligned} 8\phi_2 - 8\phi_1 + 1 &= 0, \\ 4\phi_3 + 2\phi_1 - 8\phi_2 + 1 &= 0, \\ 4\phi_2 - 8\phi_3 + 1 &= 0. \end{aligned}$$

Solving these three equations, we get

$$\phi_1 = 0.562, \phi_2 = 0.438, \phi_3 = 0.344.$$

On the other hand, if we use coarse mesh of side $h = 1$, then the figure becomes as shown below (Figure 11.18).

The finite difference equation now is

$$-4\phi + 2 = 0$$

and so $\phi = 0.5$. If we take $h_1 = 1$ and $h_2 = \frac{1}{2}$, then $\frac{h_1}{h_2} = 2$. Therefore, by deferred

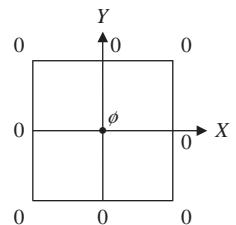


Figure 11.18

approach to the limit method, we have the improved value of ϕ as

$$\phi^* = \phi_1 + \frac{1}{3}(\phi_1 - \phi) = 0.562 + \frac{1}{3}(0.562 - 0.500) = 0.583,$$

which is very close to the exact value 0.589 of ϕ at $(0,0)$.

11.8 EIGENVALUE PROBLEMS

Problems which contain a parameter and can be solved only for certain values of this parameter are called eigenvalue problems.

EXAMPLE 11.11

Solve the eigenvalue problem

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \lambda u = 0$$

inside a square with corners $(0,0)$, $(3,0)$, $(3,3)$, and $(0,3)$, assuming that $u = 0$ on the sides of the square.

(This equation describes a vibrating membrane.)

Solution. We first take the mesh size $h = 1$. Thus, the meshes are as shown in Figure 11.19.

By symmetry, all values are equal in this case. Let these values be U .

We have $h^2 = 1$. Therefore, the difference equation for the differential equation becomes

$$\frac{u(x_0 - h, y_0) - 2u(x_0, y_0) + u(x_0 + h, y_0)}{1} + \frac{u(x_0, y_0 - h) - 2u(x_0, y_0) + u(x_0, y_0 + h)}{1} + \lambda u = 0$$

Therefore, we have

$$0 - 4U + U + U + 0 + \lambda U = 0,$$

which yields $\lambda = 2$.

Now we use $h = \frac{3}{4}$. The meshes in this case are shown in Figure 11.20.

By symmetry there are only three different values U , V , and W . In this case $h^2 = \frac{9}{16}$ and so the difference equations of the given eigenvalue problem are

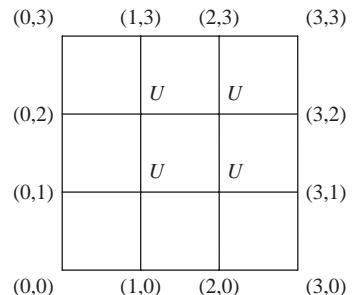


Figure 11.19

	(0,3)					(3,3)
$\begin{pmatrix} 0, \frac{9}{4} \end{pmatrix}$	U	V	U			$\begin{pmatrix} 3, \frac{9}{4} \end{pmatrix}$
$\begin{pmatrix} 0, \frac{6}{4} \end{pmatrix}$	V	W	V			$\begin{pmatrix} 3, \frac{6}{4} \end{pmatrix}$
$\begin{pmatrix} 0, \frac{3}{4} \end{pmatrix}$	U	V	U			$\begin{pmatrix} 3, \frac{3}{4} \end{pmatrix}$
(0,0)						(3,0)
	$\begin{pmatrix} \frac{3}{4}, 0 \end{pmatrix}$	$\begin{pmatrix} \frac{6}{4}, 0 \end{pmatrix}$	$\begin{pmatrix} \frac{9}{4}, 0 \end{pmatrix}$			

Figure 11.20

$$0 - 4U + V + V + 0 + \frac{9}{16}\lambda U = 0,$$

$$U - 4V + U + W + 0 + \frac{9}{16}\lambda V = 0,$$

$$V - 4W + V + V + \frac{9}{16}\lambda W = 0.$$

With $\mu = \frac{9}{16}\lambda$, these equations simplify to

$$(4 - \mu)U - 2V + 0W = 0,$$

$$-2U + (4 - \mu)V - W = 0,$$

$$0U - 4V + (4 - \mu)W = 0.$$

The condition for non-trivial solution is

$$\begin{vmatrix} 4 - \mu & -2 & 0 \\ -2 & 4 - \mu & -1 \\ 0 & -4 & 4 - \mu \end{vmatrix} = 0,$$

that is,

$$\mu^3 - 12\mu^2 + 40\mu - 32 = 0.$$

The smallest positive root is $\mu = 4 - \sqrt{8} = 1.17157$ and so $\frac{9}{16}\lambda = 1.1715$. Thus,

$$\lambda = 2.0827 \approx 2.083.$$

We know that the order of error is $O(h^2)$. Also $h_1 = 1$, $h_2 = \frac{3}{4}$ and so $\frac{h_1}{h_2} = \frac{4}{3}$. Also $Q_1 = 2$, $Q_2 = 2.083$. Therefore, Richardson's extrapolation yields

$$Q_{12} = \frac{\left(\frac{h_1}{h_2}\right)^2 Q_2 - Q_1}{\left(\frac{h_1}{h_2}\right)^2 - 1} = \frac{\frac{16}{9}Q_2 - Q_1}{\frac{16}{9} - 1}$$

$$= \frac{\frac{16}{9}(2.083) - 2}{\frac{16}{9} - 1} = \frac{16(2.083) - 18}{7} \approx 2.1897.$$

EXAMPLE 11.12

Solve the eigenvalue problem

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \lambda u = 0$$

in a triangular domain bounded by the lines $x = 0$, $y = 0$ and $x + y = 1$ assuming that $u = 0$ on the boundary.

Solution. We replace the given differential equation by a second order difference equation with mesh size h . Thus we have,

$$\frac{u(x_0 - h, y_0) - 2u(x_0, y_0) + u(x_0 + h, y_0)}{h^2} + \frac{u(x_0, y_0 - h) - 2u(x_0, y_0) + u(x_0, y_0 + h)}{h^2} + \lambda u(x_0, y_0) = 0.$$

We first take the mesh size $h = \frac{1}{4}$. Owing to the symmetry, only two functional values U and V have to

be considered as shown in Figure 11.21.

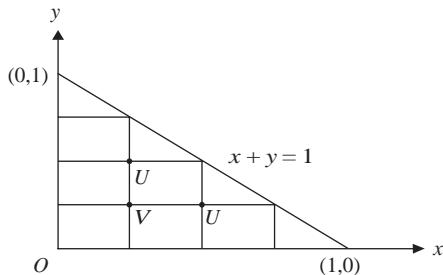


Figure 11.21

The difference equation yields two equations

$$V - 4U + 0 + 0 + 0 + \frac{1}{16}\lambda U = 0,$$

and

$$0 - 4V + U + U + 0 + \frac{1}{16}\lambda V = 0,$$

that is,

$$(\lambda - 64)U + 16V = 0$$

$$32U + V(\lambda - 64) = 0.$$

For non-trivial solution of these equations, we must have

$$\begin{vmatrix} \lambda - 64 & 16 \\ 32 & \lambda - 64 \end{vmatrix} = 0,$$

that is,

$$\lambda^2 - 128\lambda + 3584 = 0.$$

The solution of this quadratic in λ is

$$\lambda = 64 \pm 16\sqrt{2}.$$

The smallest eigenvalue is $64 - 16\sqrt{2} = 41.3726$. Now we take $h = \frac{1}{5}$ so that $h^2 = \frac{1}{25}$. By symmetry, as shown in Figure 11.22, we have to consider only three values U , V , and W .

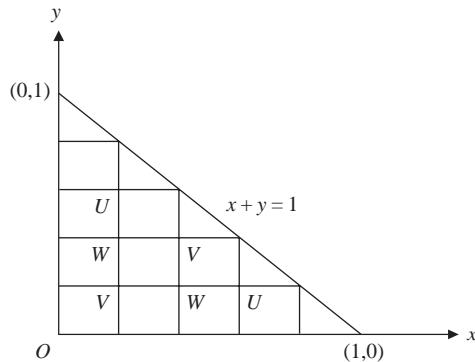


Figure 11.22

The difference equation in this case yields

$$W + 0 + 0 + 0 - 4U + \frac{1}{25} \lambda U = 0,$$

$$0 - 4V + W + W + 0 + \frac{1}{25} \lambda V = 0,$$

$$V - 4W + U + V + 0 + \frac{1}{25} \lambda W = 0,$$

that is,

$$(\lambda - 100)U + 0V + 25W = 0,$$

$$0U + (\lambda - 100)V + 50W = 0,$$

$$25U + 50V + (\lambda - 100)W = 0.$$

For a non-trivial solution, we must have

$$\begin{vmatrix} \lambda - 100 & 0 & 25 \\ 0 & \lambda - 100 & 50 \\ 25 & 50 & \lambda - 100 \end{vmatrix} = 0.$$

or

$$(\lambda - 100) \left[625 - (\lambda - 100)^2 + 2500 \right] = 0.$$

$$\text{Therefore, } \lambda = 100 \text{ and } \lambda = \frac{200 \pm \sqrt{40000 - 27500}}{2} = \frac{200 \pm 111.803}{2}.$$

The smaller value is therefore $\lambda = 44.0985$. We now have

$$\frac{h_1}{h_2} = \frac{1/4}{1/5} = \frac{5}{4}$$

and $Q_1 = 41.3726$, $Q_2 = 44.0985$. Therefore, Richardson's extrapolation yields

$$\begin{aligned} Q_{12} &= \frac{\left(\frac{h_1}{h_2}\right)^2 Q_2 - Q_1}{\left(\frac{h_1}{h_2}\right)^2 - 1} = \frac{\frac{25}{16} Q_2 - Q_1}{\frac{25}{16} - 1} = \frac{25Q_2 - 16Q_1}{9} \\ &= \frac{25(44.0985) - 16(41.3726)}{9} = 48.945. \end{aligned}$$

EXAMPLE 11.13

Find the smallest eigenvalue of the equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \lambda(x^2 + y^2)u = 0$$

for a triangular domain with corners $(-1,0)$, $(0,1)$, and $(1,0)$, where $u = 0$ on the boundary.

Solution. We replace the given differential equation by a second order difference equation with mesh size h given below:

$$\frac{u(x_0 - h, y_0) - 2u(x_0, y_0) + u(x_0 + h, y_0)}{h^2} + \frac{u(x_0, y_0 - h) - 2u(x_0, y_0) + u(x_0, y_0 + h)}{h^2} + \lambda u(x_0, y_0)(x_0^2 + y_0^2) = 0.$$

Taking $h = \frac{1}{2}$, we note that there is only one interior point $\left(0, \frac{1}{2}\right)$ (Figure 11.23)

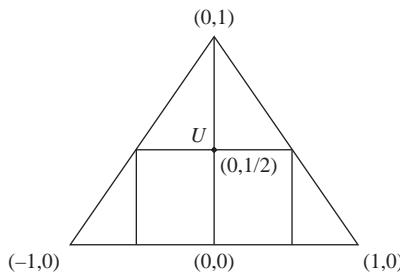
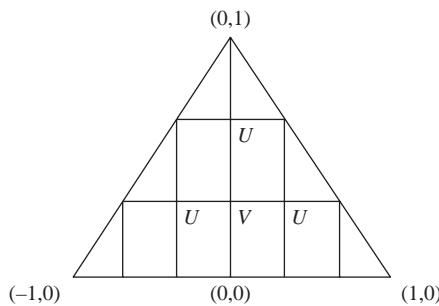


Figure 11.23

At the interior point $\left(0, \frac{1}{2}\right)$, we get $0 - 4U + 0 + 0 + 0 + \lambda \left(0 + \left(\frac{1}{2}\right)^2\right)U = 0$ and so $\lambda = 64$. We now take $h = \frac{1}{3}$. Figure 11.24 of the problem now becomes

**Figure 11.24**

By symmetry, there are only two values, U at $\left(-\frac{1}{3}, \frac{1}{3}\right)$ and V at $\left(0, \frac{1}{3}\right)$. We have

$$V - 4U + \frac{\lambda}{9} \left(\frac{1}{9} + \frac{1}{9} \right) U = 0,$$

and

$$\begin{aligned} U - 4V + U + U + \frac{\lambda}{9} \left(0 + \frac{1}{9} \right) V &= 0, \\ (2\lambda - 324)U + 81V &= 0, \\ 243U + (\lambda - 324)V &= 0. \end{aligned}$$

For non-trivial solution, we must have

$$\begin{vmatrix} 2\lambda - 324 & 81 \\ 243 & \lambda - 324 \end{vmatrix} = 0,$$

that is,

$$2\lambda^2 - 972\lambda + 85293 = 0.$$

Therefore,

$$\begin{aligned} &= \frac{972 \pm \sqrt{944784 - 682344}}{4} \\ &= \frac{972 \pm 512.289}{4} = 114.93 \text{ and } 371.07. \end{aligned}$$

The smaller eigenvalue is 114.93. Now $h_1 = \frac{1}{2}$ and $h_2 = \frac{1}{3}$. Therefore,

$$\frac{h_1}{h_2} = \frac{3}{2} \text{ and so } \left(\frac{h_1}{h_2} \right)^2 = \frac{9}{4}.$$

Also $Q_1 = 64$, $Q_2 = 114.93$. Therefore, Richardson's extrapolation yields

$$\begin{aligned}
Q_{12} &= \frac{\left(\frac{h_1}{h_2}\right)^2 Q_2 - Q_1}{\left(\frac{h_1}{h_2}\right)^2 - 1} = \frac{\frac{9}{4}Q_2 - Q_1}{\frac{9}{4} - 1} \\
&= \frac{9Q_2 - 4Q_1}{5} \\
&= \frac{9(114.93) - 4(64)}{5} = 155.674.
\end{aligned}$$

11.9 PARABOLIC EQUATIONS

The simplest example of parabolic equation is one-dimensional heat equation

$$\frac{\partial u}{\partial t} = c^2 \frac{\partial^2 u}{\partial x^2}. \quad (11.22)$$

Its solution gives the temperature u at a distance x units of length from one end of a thermally insulated bar after t seconds of heat conduction. In this problem, the temperatures at the ends of a bar of length L are often known for all time. Thus, the boundary conditions are known. Also the temperature distribution along the bar is known at some particular instant. This instant is usually taken as zero time and the temperature distribution is called the initial condition. The solution gives u for all values of x between 0 and L and values of t from 0 to ∞ .

Let the (x, t) plane be divided into smaller rectangles with sides $\Delta x = h$ and $\Delta t = k$. Our aim is to develop a difference formula for the solution of the problem. The difference formulae used for $u_t(x, t)$ and $u_{xx}(x, t)$ are

$$u_t(x, t) = \frac{u(x, t+k) - u(x, t)}{k} + O(k) \quad (11.23)$$

and

$$u_{xx}(x, t) = \frac{u(x-h, t) - 2u(x, t) + u(x+h, t)}{h^2} + O(h^2). \quad (11.24)$$

Since grid spacing is uniform, we have

$$x_{i+1} = x_i + h \text{ and } t_{j+1} = t_j + k.$$

Dropping the terms $O(k)$ and $O(h^2)$ and using $u_{i,j}$ for $u(x_i, t_j)$, and putting the values from equations (11.23) and (11.24) in equation (11.22), we get

$$\frac{u_{i,j+1} - u_{i,j}}{k} = c^2 \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2}.$$

Putting $r = \frac{c^2 k}{h^2}$, we get

$$u_{i,j+1} = u_{i,j} + r[u_{i-1,j} - 2u_{i,j} + u_{i+1,j}]. \quad (11.25)$$

Equation (11.25) creates the $(j+1)$ th row across the grid assuming that approximations in the j th row are known. This formula is called the explicit formula. However, it can be shown that this formula is valid only for $0 < r \leq \frac{1}{2}$.

For $r = \frac{1}{2}$, the formula (11.25) reduces to

$$u_{i,j+1} = \frac{u_{i-1,j} + u_{i+1,j}}{2},$$

which is called Bender–Schmidt method.

Crank–Nicholson Method

This formula is based on numerical approximations for the solution of the equation (11.22) at the point $\left(x, t + \frac{k}{2}\right)$, which lies between the rows in the grid. The approximation used for $u_t\left(x, t + \frac{k}{2}\right)$ is obtained from the central difference formula

$$u_t\left(x, t + \frac{k}{2}\right) = \frac{u(x, t + k) - u(x, t)}{k} + O(k^2). \quad (11.26)$$

The approximation for $u_{xx}\left(x, t + \frac{k}{2}\right)$ is the average of $u_{xx}(x, t)$ and $u_{xx}(x, t + k)$. Thus

$$u_{xx}\left(x, t + \frac{k}{2}\right) = \frac{1}{2} \left[\frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} + \frac{u_{i-1,j+1} - 2u_{i,j+1} + u_{i+1,j+1}}{h^2} \right] + O(h^2). \quad (11.27)$$

Thus using equations (11.26) and (11.27), the difference equation for the heat equation (11.22) becomes

$$\frac{u_{i,j+1} - u_{i,j}}{k} = \frac{c^2}{2h^2} \left[u_{i-1,j} - 2u_{i,j} + u_{i+1,j} + u_{i-1,j+1} - 2u_{i,j+1} + u_{i+1,j+1} \right].$$

Putting $\frac{c^2 k}{h^2} = r$, we get

$$-ru_{i-1,j+1} + (2 + 2r)u_{i,j+1} - ru_{i+1,j+1} = ru_{i-1,j} + (2 - 2r)u_{i,j} + ru_{i+1,j}. \quad (11.28)$$

On the left-hand side of equation (11.28), we have three unknowns and on the right-hand side all the three quantities are known. The implicit formula (11.28) is called Crank–Nicolson formula which is convergent for all finite values of r . If we have m internal mesh points on each row, then Crank–Nicolson formula gives m simultaneous equations in m unknowns in terms of the given boundary values. Thus, the solution at each interval point on all rows can be obtained.

EXAMPLE 11.14

Solve the heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2},$$

subject to the conditions $u(x, 0) = 0$, $u(0, t) = 0$ and $u(1, t) = t$.

Solution. The given equation is $\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$.

- (i) Here $c^2 = 1$. We first choose $k = \frac{1}{8}$ and $h = \frac{1}{2}$ so that $r = \frac{c^2 k}{h^2} = \frac{1}{2}$. The Crank–Nicolson formula becomes

$$-u_{i-1,j+1} + 6u_{i,j+1} - u_{i+1,j+1} = u_{i-1,j} - 2u_{i,j} + u_{i+1,j}. \quad (11.29)$$

The grid for the solution is shown in Figure 11.25.

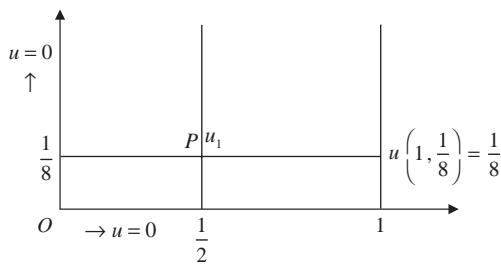


Figure 11.25

Suppose that u_1 is the value of u at the mesh point $P\left(\frac{1}{2}, \frac{1}{8}\right)$. Then formula (11.29) yields

$$0 + 6u_1 - \frac{1}{8} = 0 \text{ and so } u_1 = \frac{1}{48} = 0.02083.$$

- (ii) We now choose $k = \frac{1}{8}$, $h = \frac{1}{4}$ so that $r = 2$. For this value of r , the Crank–Nicolson formula takes the form

$$-u_{i-1,j+1} + 3u_{i,j+1} - u_{i+1,j+1} = u_{i-1,j} - u_{i,j} + u_{i+1,j}. \quad (11.30)$$

The grid for the solution is now as shown in Figure 11.26.

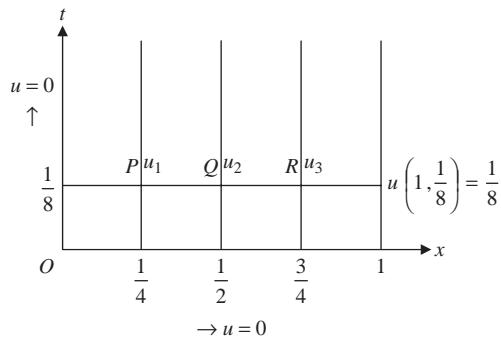


Figure 11.26

Let u_1, u_2, u_3 be the values of u at $P\left(\frac{1}{4}, \frac{1}{8}\right)$, $Q\left(\frac{1}{2}, \frac{1}{8}\right)$, and $R\left(\frac{3}{4}, \frac{1}{8}\right)$. Then equation (11.30) yields

$$0 + 3u_1 - u_2 = 0,$$

$$-u_1 + 3u_2 - u_3 = 0, \text{ and}$$

$$-u_2 + 3u_3 - \frac{1}{8} = 0.$$

Solving these equations, we get

$$u_1 = 0.00595, u_2 = 0.01785, \text{ and } u_3 = 0.04760.$$

- (iii) We now choose $k = \frac{1}{16}, h = \frac{1}{4}$ so that $r = 1$. Thus, we want to find our solution for $t = \frac{1}{8}$ in two steps instead of one as in (i) and (ii). For $r = 1$, the Crank–Nicolson formula becomes

$$-u_{i-1,j+1} + 4u_{i,j+1} - u_{i+1,j+1} = u_{i-1,j} + 0 + u_{i+1,j}. \quad (11.31)$$

The grid for the solution in this case is shown in Figure 11.27.

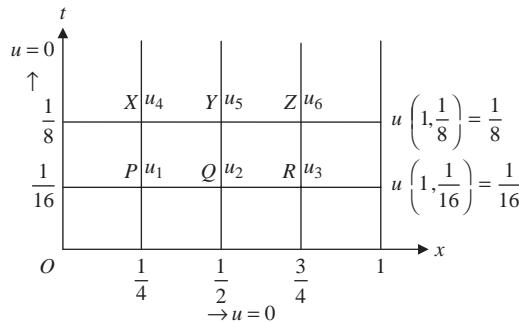


Figure 11.27

Let $u_1, u_2, u_3, u_4, u_5, u_6$ be the values of u at the points

$$P\left(\frac{1}{4}, \frac{1}{16}\right), Q\left(\frac{1}{2}, \frac{1}{16}\right), R\left(\frac{3}{4}, \frac{1}{16}\right), X\left(\frac{1}{4}, \frac{1}{8}\right), Y\left(\frac{1}{2}, \frac{1}{8}\right), \text{ and } Z\left(\frac{3}{4}, \frac{1}{8}\right).$$

Then equation (11.31) yields the following equations for u_1, u_2, u_3 :

$$4u_1 - u_2 = 0, \quad (11.32)$$

$$-u_1 + 4u_2 - u_3 = 0, \quad (11.33)$$

$$-u_2 + 4u_3 - \frac{1}{16} = 0. \quad (11.34)$$

Solving these equations, we get

$$u_1 = \frac{1}{56(16)}, \quad u_2 = \frac{1}{56(4)}, \quad u_3 = \frac{15}{56(16)}.$$

Also equation (11.31) yields the following equations for u_4, u_5, u_6 :

$$\begin{aligned} 4u_4 - u_5 &= \frac{1}{4(56)}, \\ -u_4 + 4u_5 - u_6 &= \frac{1}{56}, \\ -u_5 + 4u_6 - \frac{1}{8} &= \frac{1}{4(56)} + \frac{1}{16}. \end{aligned}$$

Solving these equations, we get

$$u_4 = 0.005899, u_5 = 0.019132, u_6 = 0.052771.$$

The exact solution of the problem by Fourier series method is

$$u(x, t) = \frac{1}{6}(x^3 - x + 6xt) + \frac{2}{\pi^3} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^3} e^{-n^2 \pi^2 t} \sin n\pi x,$$

which yields

$$u\left(\frac{1}{4}, \frac{1}{8}\right) = 0.00541, u\left(\frac{1}{2}, \frac{1}{8}\right) = 0.01878, \text{ and } u\left(\frac{3}{4}, \frac{1}{8}\right) = 0.5240.$$

EXAMPLE 11.15

Solve, by Bender–Schmidt method, the parabolic equation

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2}$$

subject to the condition $u(0, t) = u(4, t) = 0$ and $u(x, 0) = x(4 - x)$.

Solution. We have $c^2 = \frac{1}{2}$. We first choose $k = 1$ and $h = 1$. Then $r = \frac{c^2 k}{h^2} = \frac{1}{2}$. The Bender–Schmidt method is applicable and we have

$$u_{i,j+1} = \frac{u_{i-1,j} + u_{i+1,j}}{2}. \quad (11.35)$$

The grid for the solution is shown in Figure 11.28.

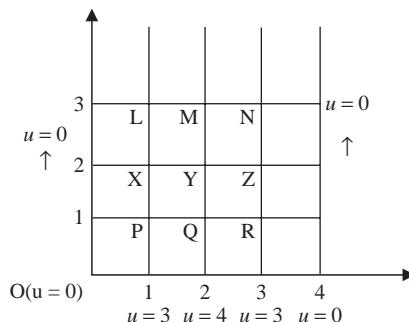


Figure 11.28

We have

$$u(1,0) = 1(4-1) = 3, \quad u(2,0) = 2(4-2) = 4$$

and

$$u(3,0) = 3(4-3) = 3.$$

Let u_1, u_2 , and u_3 be the values of u at $P(1,1)$, $Q(2,1)$, and $R(3,1)$, respectively. Then equation (11.35) yields

$$u_1 = \frac{0+4}{2} = 2,$$

$$u_2 = \frac{3+3}{2} = 3, \text{ and}$$

$$u_3 = \frac{4+0}{2} = 2.$$

Similarly, if u_4, u_5 , and u_6 are the values of u at $X(1,2), Y(2,2)$, and $Z(3,2)$, respectively, then

$$u_4 = \frac{0+u_2}{2} = \frac{0+3}{2} = 1.5,$$

$$u_5 = \frac{u_1+u_3}{2} = \frac{2+2}{2} = 2, \text{ and}$$

$$u_6 = \frac{u_2+0}{2} = \frac{3+0}{2} = 1.5.$$

Similarly, the values u_7, u_8, u_9 at L, M , and M are, respectively,

$$u_7 = \frac{0+u_5}{2} = \frac{0+2}{2} = 1,$$

$$u_8 = \frac{u_4+u_6}{2} = \frac{1.5+1.5}{2} = 1.5, \text{ and}$$

$$u_9 = \frac{u_5+0}{2} = \frac{2+0}{2} = 1.$$

EXAMPLE 11.16

Use Crank–Nicolson method to solve

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

subject to the conditions

$$u(x,0) = \sin \pi x, \quad 0 \leq x \leq 1, \quad u(0,t) = u(1,t) = 0.$$

Solution. We first take $k = \frac{1}{8}$ and $h = \frac{1}{4}$ so that $r = 2$. The Crank–Nicolson scheme corresponding to $r = 2$

is given by

$$-u_{i-1,j+1} + 3u_{i,j+1} - u_{i+1,j+1} = u_{i-1,j} - u_{i,j} + u_{i+1,j}. \quad (11.36)$$

The grid corresponding to these values of h , k , and r is shown in Figure 11.29.

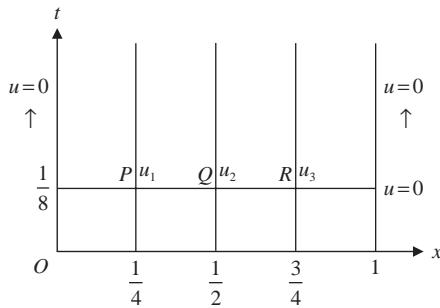


Figure 11.29

Let u_1 , u_2 , and u_3 be the values at $P\left(\frac{1}{4}, \frac{1}{8}\right)$, $Q\left(\frac{1}{2}, \frac{1}{8}\right)$, and $R\left(\frac{3}{4}, \frac{1}{8}\right)$, respectively. Then equation (11.36) yields

$$\begin{aligned} 0 + 3u_1 - u_2 &= 0 - \sin \frac{\pi}{4} + \sin \frac{\pi}{2} = -0.7071 + 1 = 0.2929, \\ -u_1 + 3u_2 - u_3 &= \sin \frac{\pi}{4} - \sin \frac{\pi}{2} + \sin \frac{3\pi}{4} \\ &= 0.7071 - 1 + 0.7071 = 0.4142, \end{aligned}$$

and

$$\begin{aligned} -u_2 + 3u_3 - 0 &= \sin \frac{\pi}{2} - \sin \frac{3\pi}{4} + \sin \pi \\ &= 1 + 0.7071 = 1.7071. \end{aligned}$$

Solving these equations, we obtain

$$u_1 = 0.252042, u_2 = 0.463228, \text{ and } u_3 = 0.7234426.$$

Now we choose $k = \frac{1}{16}$ and $h = \frac{1}{4}$ so that $r = 1$. The Crank–Nicolson scheme corresponding to this value of r is

$$-u_{i-1,j+1} + 4u_{i,j+1} - u_{i+1,j+1} = u_{i-1,j} + 0 + u_{i+1,j}. \quad (11.37)$$

The grid corresponding to these values of h , k , and r is shown in Figure 11.30.

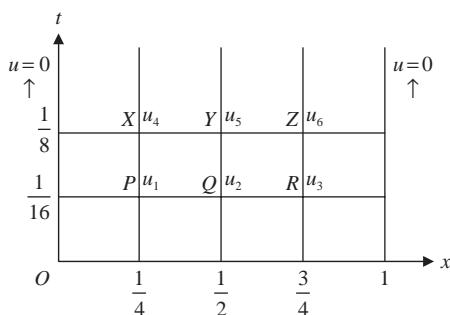


Figure 11.30

Applying scheme (11.37) at the grid points P , Q , and R , we get

$$\begin{aligned} 0 + 4u_1 - u_2 &= 0 + 0 + \sin \frac{\pi}{2} = 1, \\ -u_1 + 4u_2 - u_3 &= \sin \frac{\pi}{4} + \sin \frac{3\pi}{4} + 0 = 1.4142, \\ -u_2 + 4u_3 + 0 &= \sin \frac{\pi}{2} + 0 + \sin \pi - 0.7071. \end{aligned}$$

Solving these equations, we get

$$u_1 = 0.381497, u_2 = 0.52599 \text{ and } u_3 = 0.30827.$$

Now using the scheme (11.37) at each of the points X , Y , and Z , we have

$$\begin{aligned} 4u_4 - u_5 &= u_2 = 0.52599, \\ -u_4 + 4u_5 - u_6 &= u_1 + 0 - u_3 = 0.073227, \\ -u_5 + 4u_6 - 0 &= u_2 + 0 - 0 = 0.52599. \end{aligned}$$

The solution is

$$u_4 = 0.15551, u_5 = 0.09606, \text{ and } u_6 = 0.15551.$$

The analytical solution of the problem is $u = e^{-\pi^2 t} \sin \pi x$. Putting $x = \frac{1}{4}, t = \frac{1}{16}$, we observe that

$$u_1 = \left(\frac{1}{4}, \frac{1}{16} \right) = \frac{\sin \pi/4}{e \pi^2/16} = \frac{0.7071}{1.853} = 0.38159,$$

which is in good agreement with the calculated value.

EXAMPLE 11.17

Solve $\frac{\partial u}{\partial t} = 5 \frac{\partial^2 u}{\partial x^2}$ under the condition

$$u(0, t) = 0, u(5, t) = 60 \text{ and } u(x, 0) = \begin{cases} 20x & \text{for } 0 < x \leq 3 \\ 60 & \text{for } 3 < x \leq 5 \end{cases}$$

for five time steps having $h = 1$, by Schmidt's method.

Solution. We have

$$\frac{\partial u}{\partial t} = 5 \frac{\partial^2 u}{\partial x^2}, \quad u(0, t) = 0, \quad u(5, t) = 60,$$

$$u(x, 0) = \begin{cases} 20x & \text{for } 0 < x \leq 3 \\ 6 & \text{for } 3 < x \leq 5. \end{cases}$$

Here $c^2 = 5, h = 1$. We choose $k = \frac{1}{10}$ so that $r = \frac{c^2 k}{h} = \frac{5}{10} = \frac{1}{2}$.

Therefore, Bender–Schmidt method is applicable and we have

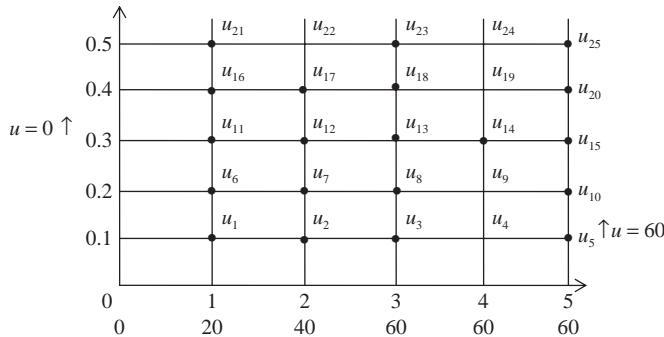
$$u_{i,j+1} = \frac{u_{i-1,j} - u_{i+1,j}}{2}. \tag{11.38}$$

We have

$$u(1,0) = 20(1) = 20, u(2,0) = 20(2) = 40, u(3,0) = 20(3) = 60.$$

$$u(4,0) = 60, \quad u(5,0) = 60.$$

The grid for the solution is



Using equation (11.38), the values of u at the grid points $(1,0,1)$, $(2,0,1)$, $(3,0,1)$, $(4,0,1)$, $(5,0,1)$ are, respectively,

$$u_1 = \frac{0+40}{2} = 20, \quad u_2 = \frac{60+20}{2} = 40,$$

$$u_3 = \frac{60+40}{2} = 50, \quad u_4 = \frac{60+60}{2} = 60, \quad u_5 = 60.$$

Further, the second row of the solution is

$$u_6 = \frac{0+u_2}{2} = \frac{0+40}{2} = 20, \quad u_7 = \frac{u_1+u_3}{2} = \frac{20+50}{2} = 35,$$

$$u_8 = \frac{u_2+u_4}{2} = \frac{40+60}{2} = 50, \quad u_9 = \frac{u_3+u_5}{2} = \frac{50+60}{2} = 55, \quad u_{10} = 60.$$

For the third row, we have

$$u_{11} = \frac{0+u_7}{2} = \frac{0+35}{2} = 17.5, \quad u_{12} = \frac{u_6+u_8}{2} = \frac{20+50}{2} = 35,$$

$$u_{13} = \frac{u_7+u_9}{2} = \frac{35+55}{2} = 45, \quad u_{14} = \frac{u_8+u_{10}}{2} = \frac{50+60}{2} = 55, \quad u_{15} = 60.$$

Proceeding in the same fashion, we get the following table:

$i \backslash j$	0	1	2	3	4	5
0	0	20	40	60	60	60
1	0	20	40	50	60	60
2	0	20	35	50	55	60
3	0	17.5	35	45	55	60
4	0	17.5	31.25	45	52.5	60
5	0	15.625	31.25	41.875	52.5	60

EXAMPLE 11.18

Find the value of $u(x,t)$ satisfying the parabolic equation $\frac{\partial u}{\partial t} = 4 \frac{\partial^2 u}{\partial x^2}$ and the boundary conditions $u(0,t) = 0 = u(8,t)$ and $u(x,0) = 4x - \frac{1}{2}x^2$ at the points $x = i$, $i = 0, 1, 2, \dots, 7$ and $t = \frac{1}{8}j$, $j = 0, 1, 2, \dots, 5$.

Solution. The given equation is

$$\frac{\partial u}{\partial t} = 4 \frac{\partial^2 u}{\partial x^2}, \quad u(0,t) = u(8,t) = 0$$

and

$$u(x,0) = 4x - \frac{1}{2}x^2 \text{ at } x = i, \quad i = 0, 1, 2, \dots, 7; \quad t = \frac{1}{8}j, \quad j = 0, 1, 2, \dots, 5.$$

Comparing with the standard form $\frac{\partial u}{\partial t} = c^2 \frac{\partial^2 u}{\partial x^2}$, we have $c^2 = 4$. Also $h = 1$, $k = \frac{1}{8}$. Therefore, $r = \frac{c^2 k}{h^2} = \frac{1}{2}$. Hence, Bredre-Schmidt formula

$$u_{i,j+1} = \frac{1}{2}[u_{i-1,j} + u_{i+1,j}]. \quad (11.39)$$

is applicable. The boundary conditions imply that

$$u_{0,t} = 0 \text{ for any } t \text{ and } u_{8,t} = 0 \text{ for any } t.$$

Further,

$$u(x,0) = 4x - \frac{1}{2}x^2$$

implies

$$u_{i,0} = 4i - \frac{1}{2}x^2.$$

Putting $i = 0, 1, \dots, 7$, we get the entries of the first row as

$$0 \quad 3.5 \quad 6 \quad 7.5 \quad 8 \quad 7.5 \quad 6 \quad 3.5 \quad 0.$$

Putting $j = 0$ in equation (11.39), we get

$$u_{i,1} = \frac{1}{2}[u_{i-1,0} + u_{i+1,0}]$$

and so taking $i = 1, 2, \dots, 7$, we get the values of the second row as

$$u_{11} = \frac{1}{2}(u_{0,0} + u_{2,0}) = \frac{1}{2}(0 + 6) = 3,$$

$$u_{21} = \frac{1}{2}(u_{1,0} + u_{3,0}) = \frac{1}{2}(3.5 + 7.5) = 5.5,$$

and so on.

Thus, the entries in the second row are

$$0 \ 3 \ 5.5 \ 7 \ 7.5 \ 7 \ 5.5 \ 3 \ 0.$$

Then the entries in the third row are obtained by putting $j = 1$ in equation (1).

These are

$$0 \ 2.7 \ 5.5 \ 6.5 \ 7 \ 6.5 \ 5 \ 2.7 \ 50.$$

Similarly, the entries in the fourth, fifth, and sixth rows are, respectively,

$$0 \ 2.5 \ 4.625 \ 6 \ 6.5 \ 6 \ 4.625 \ 2.5 \ 0,$$

$$0 \ 2.3125 \ 4.25 \ 5.5625 \ 6 \ 5.5625 \ 4.25 \ 2.3125 \ 0,$$

$$0 \ 2.125 \ 3.9375 \ 5.125 \ 5.5625 \ 5.125 \ 3.9375 \ 2.125 \ 0.$$

11.10 ITERATIVE METHOD TO SOLVE PARABOLIC EQUATIONS

Consider the heat equation

$$\frac{\partial u}{\partial t} = c^2 \frac{\partial^2 u}{\partial x^2}.$$

The Crank–Nicolson scheme for the solution of this differential equation is

$$(1+r)u_{i,j+1} = u_{i,j} + \frac{1}{2}r \left[u_{i-1,j+1} + u_{i+1,j} + u_{i+1,j+1} + u_{i-1,j} - 2u_{i,j} \right], \quad (11.40)$$

where $r = \frac{c^2 k}{h^2}$. In this scheme, $u_{i,j+1}, u_{i-1,j+1}, u_{i+1,j+1}$, are unknowns, while $u_{i,j}, u_{i+1,j}$, and $u_{i-1,j}$ have already been calculated at the j th step. Let

$$c_i = u_{i,j} + \frac{r}{2} \left[u_{i-1,j} - 2u_{i,j} + u_{i+1,j} \right].$$

Then equation (11.40) reduces to

$$u_i = \frac{r}{2(1+r)} \left[u_{i-1} + u_{i+1} \right] + \frac{c_i}{(1+r)}. \quad (11.41)$$

Equation (11.41) yields the iteration formula

$$u_i^{(n+1)} = \frac{r}{2(1+r)} \left[u_{i-1}^{(n)} + u_{i+1}^{(n)} \right] + \frac{c_i}{1+r}, \quad (11.42)$$

which expresses the $(n+1)$ th iterate in terms of the n th iterate only and is known as Jacobi's iteration formula.

Since, while computing $u_i^{(n+1)}$, the latest value $u_{i-1}^{(n+1)}$ of u_{i-1} is available, equation (11.42) can be written as

$$u_i^{(n+1)} = \frac{r}{2(1+r)} \left[u_{i-1}^{(n+1)} + u_{i+1}^{(n)} \right] + \frac{c_i}{1+r}, \quad (11.43)$$

which is known as Gauss–Seidel iteration formula and converges twice as fast as Jacobi's formula.

EXAMPLE 11.19

Solve by Gauss–Seidel method the parabolic equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2},$$

subject to the conditions

$$u(x, 0) = \sin \pi x, \quad 0 \leq x \leq 1$$

$$u(0, t) = u(1, t) = 0, \quad t > 0.$$

Solution. We choose $h = 0.2$ and $k = 0.02$ so that $r = \frac{c^2 k}{h^2} = \frac{1}{2}$. Thus, the Gauss–Seidel iteration formula takes the form

$$u_i^{(n+1)} = \frac{1}{6} [u_{i-1}^{(n+1)} + u_{i+1}^{(n)}] + \frac{2}{3} c_i. \quad (11.44)$$

Let u_1, u_2, u_3, u_4 be the values of u on the row corresponding to $t = 0.02$. Therefore, grid for the solution is as shown in Figure 11.31.

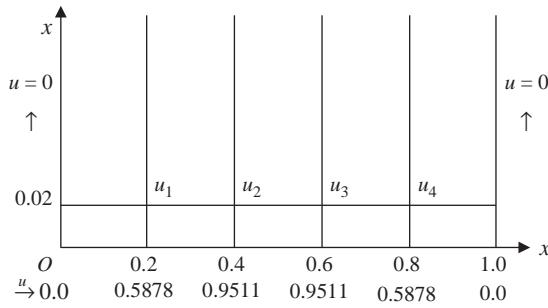


Figure 11.31

Using formula (11.44) at the four grid points, we have

$$\begin{aligned} u_1^{(n+1)} &= \frac{1}{6} [0 + u_2^{(n)}] + \frac{2}{3} \left[0.5878 + \frac{1}{4} \{0 - 2(0.5878) + 0.9511\} \right] \\ &= \frac{1}{6} u_2^{(n)} + 0.3544, \end{aligned} \quad (11.45)$$

$$\begin{aligned} u_2^{(n+1)} &= \frac{1}{6} [u_1^{(n+1)} + u_3^{(n)}] + \frac{2}{3} \left[0.9511 + \frac{1}{4} \{0.5878 - 2(0.9511) + 0.9511\} \right] \\ &= \frac{1}{6} [u_1^{(n+1)} + u_3^{(n)}] + 0.5736, \end{aligned} \quad (11.46)$$

$$\begin{aligned} u_3^{(n+1)} &= \frac{1}{6} [u_2^{(n+1)} + u_4^{(n)}] + \frac{2}{3} \left[0.9511 + \frac{1}{4} \{0.9511 - 2(0.9511) + 0.5878\} \right] \\ &= \frac{1}{6} [u_2^{(n+1)} + u_4^{(n)}] + 0.5736, \end{aligned} \quad (11.47)$$

$$\begin{aligned} u_4^{(n+1)} &= \frac{1}{6} [u_3^{(n+1)} + 0] + \frac{2}{3} \left[0.5878 + \frac{1}{4} \{0.9511 - 2(0.5878) + 0\} \right] \\ &= \frac{1}{6} u_3^{(n+1)} + 0.3544. \end{aligned} \quad (11.48)$$

Putting $n = 0$, we have

$$\begin{aligned} u_1^{(1)} &= \frac{1}{6}u_2^{(0)} + 0.3544 = \frac{1}{6}(0.9511) + 0.3544 = 0.51291, \\ u_2^{(1)} &= \frac{1}{6}[u_1^{(1)} + u_3^{(0)}] + 0.5736 \\ &= \frac{1}{6}[0.51291 + 0.9511] + 0.5736 = 0.81760, \\ u_3^{(1)} &= \frac{1}{6}[u_2^{(1)} + u_4^{(0)}] + 0.5736 \\ &= \frac{1}{6}[0.81760 + 0.5878] + 0.5736 = 0.80783, \\ u_4^{(1)} &= \frac{1}{6}u_3^{(1)} + 0.3544 \\ &= \frac{1}{6}(0.80783) + 0.3544 = 0.48904. \end{aligned}$$

Putting $n = 1$ in equations (11.45), (11.46), (11.47), and (11.48), we have

$$\begin{aligned} u_2^{(1)} &= \frac{1}{6}u_2^{(1)} + 0.3544 = \frac{1}{6}(0.81760) + 0.3544 = 0.49067, \\ u_2^{(1)} &= \frac{1}{6}[u_1^{(2)} + u_3^{(1)}] + 0.5736 \\ &= \frac{1}{6}[0.49067 + 0.80783] + 0.5736 = 0.79002, \\ u_3^{(2)} &= \frac{1}{6}[u_2^{(2)} + u_4^{(1)}] + 0.5736 \\ &= \frac{1}{6}[0.79002 + 0.48904] + 0.5736 = 0.78677, \\ u_4^{(2)} &= \frac{1}{6}u_3^{(2)} + 0.3544 = \frac{1}{6}[0.78678] + 0.3544 = 0.48523. \end{aligned}$$

Proceeding with the process we shall observe that fourth and fifth iterates are equal up to four decimal places. We have then

$$\begin{aligned} u_1 &= 0.4853, u_2 = 0.7854, \\ u_3 &= 0.7854, u_4 = 0.4853. \end{aligned}$$

The analytical solution of the problem is $u = e^{-\pi^2 t} \sin \pi x$. Putting $x = 0.2$ and $t = 0.02$, we get

$$u_1 = \frac{\sin \pi/5}{e \pi^2/50} = \frac{0.58779}{1.2182} = 0.48251,$$

which is in good agreement with the calculated value 0.4853.

11.11 HYPERBOLIC EQUATIONS

A simple example of an hyperbolic partial differential equation is the wave equation

$$u_{tt}(x, t) = c^2 u_{xx}(x, t). \quad (11.49)$$

We divide the (x, t) plane into grids consisting of small rectangles with sides $\Delta x = h$ and $\Delta t = k$. We shall use a difference equation method to compute approximations $\{u_{i,j} : i = 1, 2, \dots, n\}$ in successive rows for $j = 2, 3, \dots$. The true solution at the grid point (x_i, t_j) is $u(x_i, t_j)$.

We know that the central difference formulae for approximating $u_{tt}(x, t)$ and $u_{xx}(x, t)$ are

$$u_{tt}(x, t) = \frac{u(x, t+k) - 2u(x, t) + u(x, t-k)}{k^2} + O(k^2), \quad (11.50)$$

$$u_{xx}(x, t) = \frac{u(x+h, t) - 2u(x, t) + u(x-h, t)}{h^2} + O(h^2). \quad (11.51)$$

Since grid spacing is uniform, we have

$$\begin{aligned} x_{i+1} &= x_i + h, x_{i-1} = x_i - h, \\ t_{j+1} &= t_j + k, t_{j-1} = t_j - k. \end{aligned}$$

Dropping the terms $O(h^2), O(k^2)$, using $u_{i,j}$ for $u(x_i, t_j)$ and putting the values from equations (11.50) and (11.51) in equation (11.49), we get

$$\frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{k^2} = c^2 \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2}.$$

Putting $r = \frac{ck}{h}$, this equation reduces to

$$u_{i,j-1} - 2u_{i,j} + u_{i,j+1} = r^2 [u_{i-1,j} - 2u_{i,j} + u_{i+1,j}]. \quad (11.52)$$

Thus, using equation (11.52), we can find row $j+1$ across the grid assuming that approximations in both rows j and $j-1$ are known. Therefore,

$$u_{i,j+1} = (2 - 2r^2)u_{i,j} + r^2(u_{i+1,j} + u_{i-1,j}) - u_{i,j-1} \quad (11.53)$$

for $i = 2, 3, \dots, n-1$. We observe that the four known values on the right-hand side of equation (11.53), which are used to find $u_{i,j+1}$ can be shown as in Figure 11.32.

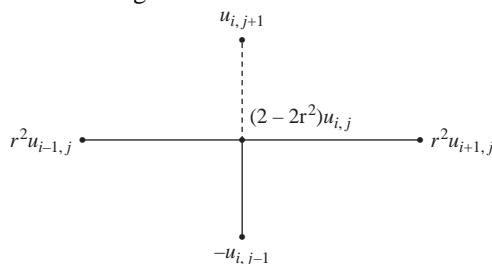


Figure 11.32

EXAMPLE 11.20

Use the finite difference method to solve the wave equation for a vibrating string

$$u_{tt}(x,t) = 4u_{xx}(x,t) \text{ for } 0 \leq x \leq 4 \text{ and } 0 \leq t \leq 2$$

with the boundary conditions

$$u(0,t) = u(4,t) = 0, \quad t > 0$$

and the initial conditions

$$\left. \begin{array}{l} u(x,0) = x(4-x) \\ u_t(x,0) = 0 \end{array} \right\} \quad 0 \leq x \leq 4.$$

Solution. We have $c^2 = 4$. We take $h = 1$ and $k = 0.5$. Then $r^2 = \frac{c^2 k^2}{h^2} = 1$. Therefore, the difference equation for the problem is

$$u_{i,j+1} = u_{i+1,j} + u_{i-1,j} - u_{i,j-1}. \quad (11.54)$$

Since $u(0,t) = u(4,t) = 0$, we have

$$u_{0,j} = 0, \quad u_{4,j} = 0.$$

Further, since $u(x,0) = x(4-x)$, we have

$$\begin{aligned} u_{i,0} &= i(4-i) && \text{for } t = 0 \\ &= 3, 4, 3 && \text{for } i = 1, 2, 3 \text{ at } t = 0, \end{aligned}$$

which are entries for the first row.

Finally, since $u_t(x,0) = 0$, $0 \leq x \leq 4, t = 0$, we have

$$\frac{u_{i,j+1} - u_{i,j}}{k} = 0 \quad \text{for } t = 0 \text{ and } j = 0$$

and so

$$u_{i,1} = u_{i,0},$$

which shows that the entries in the second row are same as those of the first row. Thus, the grid for the solution is as shown in Figure 11.33.

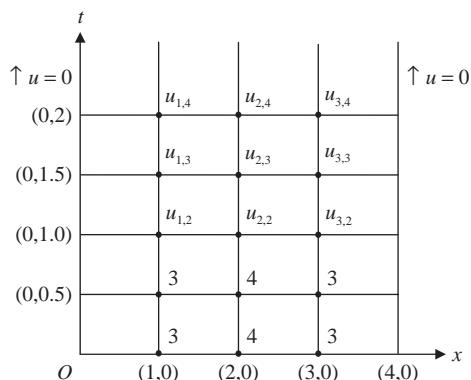


Figure 11.33

Using formula (11.54), we have

$$\left. \begin{array}{l} u_{1,2} = 4 + 0 - 3 = 1 \\ u_{2,2} = 3 + 3 - 4 = 2 \\ u_{3,2} = 4 + 0 - 3 = 1 \end{array} \right\} \quad \text{third row}$$

$$\left. \begin{array}{l} u_{1,3} = 0 + 2 - 3 = -1 \\ u_{2,3} = 1 + 1 - 4 = -2 \\ u_{3,3} = 2 + 0 - 3 = -1 \end{array} \right\} \quad \text{fourth row}$$

$$\left. \begin{array}{l} u_{1,4} = 0 - 2 - 1 = -3 \\ u_{2,4} = -1 - 1 - 2 = -4 \\ u_{3,4} = -2 + 0 - 1 = -3 \end{array} \right\} \quad \text{fifth row}$$

EXAMPLE 11.21

Solve the wave equation

$$u_{tt}(x, t) = 16 u_{xx}(x, t), \quad 0 \leq x \leq 5, \quad 0 \leq t \leq 1.25$$

subject to the conditions

$$\left. \begin{array}{l} u(0, t) = u(5, t) = 0, \quad t > 0 \\ u(x, 0) = x^2(5 - x) \\ u_t(x, 0) = 0 \end{array} \right\} \quad 0 \leq x \leq 5.$$

Solution. In the given problem, we have $c^2 = 16$. We take $h = 1$, $k = 0.25$. Then $r^2 = \frac{c^2 k^2}{h^2} = 16(0.0625) = 1$. Therefore, the difference equation for the problem becomes

$$u_{i,j+1} = u_{i+1,j} + u_{i-1,j} - u_{i,j-1} \quad (11.55)$$

and the scheme for calculation becomes as shown in Figure 11.34.

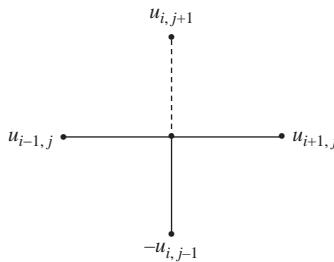


Figure 11.34

Since $u(0, t) = u(5, t) = 0$, we have

$$u_{0,j} = 0 \text{ and } u_{5,j} = 0 \text{ for all } j.$$

Hence, the entries in the first and last columns are zero. Since $u(x, 0) = x^2(5 - x)$, we have

$$u_{i,0} = i^2(5 - i),$$

which yields 4, 12, 18, and 16 for $i = 1, 2, 3, 4$, and $t = 0$. Thus, values of u on the first row are 0, 4, 12, 18, 16, and 0. Also $u_t(x, 0) = 0$ and so

$$\frac{u_{i,j+1} - u_{i,j}}{k} = 0 \text{ for } j = 0.$$

This implies $u_{i,1} = u_{i,0}$ and so the entries in the second row are the same as those of first row. Thus, the grid of the solution of the given equation is as shown in Figure 11.35.

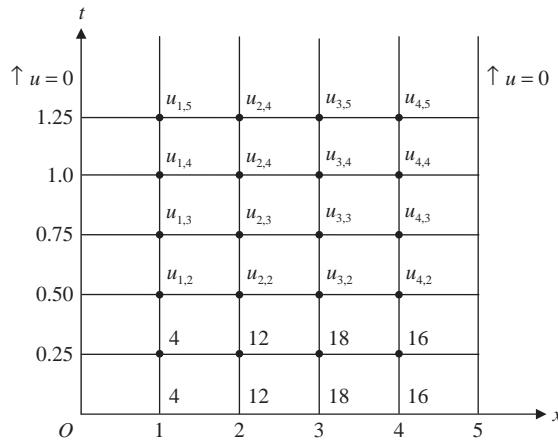


Figure 11.35

Using equation (11.55) and the scheme given above, we have

$$\left. \begin{array}{l} u_{1,2} = 0 + 12 - 4 = 8 \\ u_{2,2} = 4 + 18 - 12 = 10 \\ u_{3,2} = 12 + 16 - 18 = 10 \\ u_{4,2} = 18 + 0 - 16 = 2 \\ u_{1,3} = 0 + 10 - 4 = 6 \\ u_{2,3} = 8 + 10 - 12 = 6 \\ u_{3,3} = 10 + 2 - 18 = -6 \\ u_{4,3} = 10 + 0 - 16 = -6 \\ u_{1,4} = 0 + 6 - 8 = -2 \\ u_{2,4} = 6 - 6 - 10 = -10 \\ u_{3,4} = 6 - 6 - 10 = -10 \\ u_{4,4} = -6 + 0 - 2 = -8 \\ u_{1,5} = 0 - 10 - 6 = -16 \\ u_{2,5} = -2 - 10 - 6 = -18 \\ u_{3,5} = -10 - 8 + 6 = -12 \\ u_{4,5} = -10 + 0 + 6 = -4 \end{array} \right\} \begin{array}{l} \text{third row} \\ \text{fourth row} \\ \text{fifth row} \\ \text{sixth row.} \end{array}$$

EXAMPLE 11.22

Solve the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$$

subject to initial condition $u = f(x)$, $\frac{\partial u}{\partial t} = g(x)$, $0 \leq x \leq 1$ at $t = 0$ and the boundary conditions: $u(0, t) = \phi(t)$, $u(1, t) = \psi(t)$.

Solution. From article 11.11, we have $\left(\text{taking } r = \frac{ck}{h} \right)$,

$$u_{i,j+1} + (2 - 2r^2)u_{i,j} + r^2(u_{i+1,j} + u_{i-1,j}) - u_{i,j-1}. \quad (11.56)$$

But, we are given that

$$\frac{\partial u}{\partial i} = \frac{u_{i,j+1} - u_{i,j-1}}{2k} = g(x) \text{ at } t = 0$$

or

$$u_{i,j+1} = u_{i,j-1} + 2kg(x) \text{ at } t = 0$$

or

$$u_{i,1} = u_{i,-1} + 2kg(x) \text{ at } t = 0 \text{ and } j = 0. \quad (11.57)$$

Also, $u(x, 0) = f(x)$. Therefore,

$$u_{i,-1} = f(x).$$

Thus, equation (11.57) reduces to

$$u_{i,1} = f(x) + 2kg(x). \quad (11.58)$$

Further, the boundary conditions reduce to

$$u_{0,j} = \phi(t) \text{ and } u_{i,j} = \psi(t).$$

Thus, we have the grid as shown in the Figure 11.36

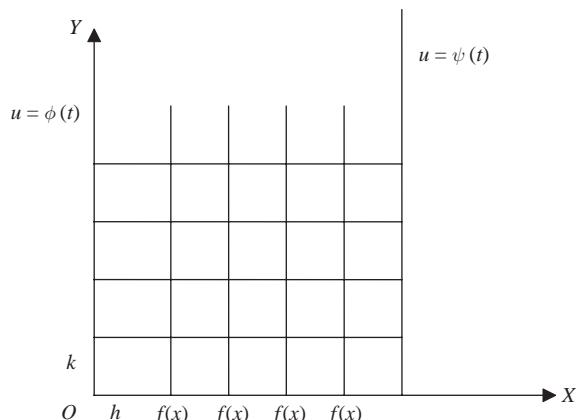


Figure 11.36

The entries in the first row of the solution are all $f(x)$. The entries in the second row are given by equation (11.58). The entries in the third row are given by formula (11.56) and so on. Thus formula (11.56) is three level time formula.

EXERCISES

- Solve the elliptic equation in the square region $0 \leq x \leq 4$, $0 \leq y \leq 4$ subject to the conditions

$$u(0, y) = 0, \quad u(4, y) = 8 + 2y$$

$$u(x, 0) = \frac{x^2}{2}, \quad u(x, y) = x^2$$

and taking $h = k = 1$.

Hint. Use standard five point formula and diagonal five point formula to find initial value at the mesh point and then use iteration method.

$$\text{Ans. } u_1 \approx 1.99, u_2 \approx 4.91, u_3 \approx 8.99, u_4 \approx 2.06,$$

$$u_5 \approx 4.69, u_6 \approx 8.06, u_7 \approx 1.57, u_8 \approx 3.71, u_9 \approx 6.57$$

- Solve Laplace equation $u_{xx} + u_{yy} = 0$ in the domain of Figure 11.37.

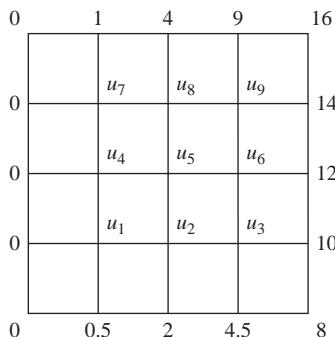
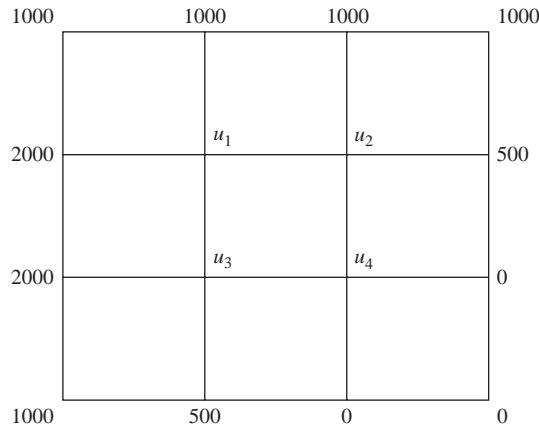


Figure 11.37

$$\text{Ans. } u_1 \approx 1.57, u_2 \approx 3.71, u_3 \approx 6.57, u_4 \approx 2.06,$$

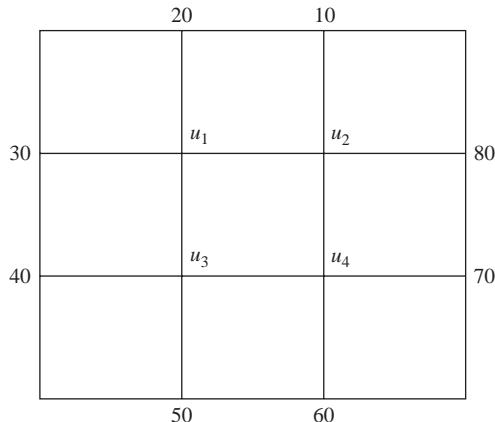
$$u_5 \approx 4.69, u_6 \approx 8.06, u_7 \approx -2, u_8 \approx -2$$

3. Solve Laplace equation $u_{xx} + u_{yy} = 0$ at the internal mesh points of the square region with the boundary values shown in Figure 11.38.

**Figure 11.38**

Ans. $u_1 = 1208.3, u_2 = 791.7, u_3 = 1041.7, u_4 = 458.4$

4. Solve the Laplace equation $\nabla^2 u = 0$ in the square region with mesh points and boundary conditions shown in Figure 11.39.

**Figure 11.39**

Ans. $u_1 = 34.99, u_2 = 44.99, u_3 = 44.99, u_4 = 55.0$

5. Solve Exercise 1 by relaxation method.

6. Solve $\nabla^2 u = -400$ by relaxation method in the square bounded by $x = 0, x = 4, y = 0, y = 4$ and under the condition that u is zero on the boundary of the square.

Ans. $u_1 = 275, u_2 = 350, u_3 = 275, u_4 = 350, u_5 = 450,$
 $u_6 = 350, u_7 = 275, u_8 = 350, u_9 = 275$

7. Solve the Poisson's equation $u_{xx} + u_{yy} - 8x^2y^2 = 0$ for the square mesh shown in Figure 11.40 under the condition that $u = 0$ on the boundary and that mesh length $h = k = 1$.

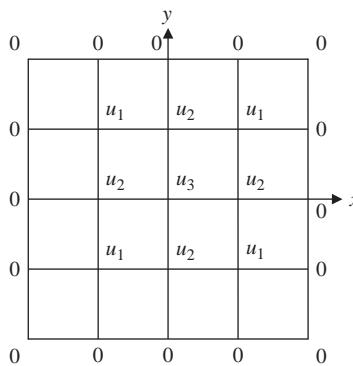


Figure 11.40

Hint: Use symmetry first, so there are only three values to be determined.

Ans. $u_1 = -3$, $u_2 = -2$, $u_3 = -2$

8. Determine the system of four equations in four unknowns p_1, p_2, p_3 , and p_4 for solving the Laplace equation $\nabla^2 u = 0$ on the 4×4 grid shown in Figure 11.41.

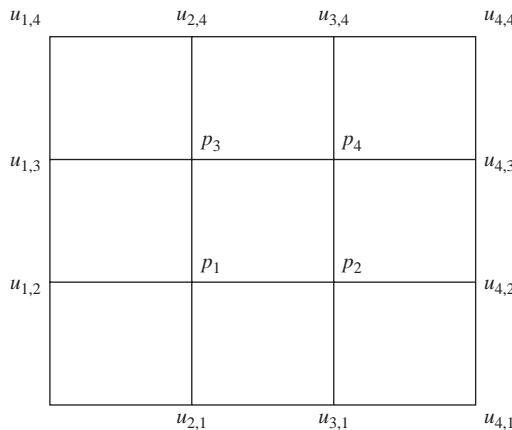


Figure 11.41

Ans. $-4p_1 + u_{1,2} + p_2 + p_3 + u_{2,1} = 0$

$p_1 - 4p_2 + p_4 + u_{4,2} + u_{3,1} = 0$

$p_1 + u_{1,3} - 4p_3 + p_4 + u_{2,4} = 0$

$p_2 + p_3 - 4p_4 + u_{4,3} + u_{3,4} = 0$

9. Solve the heat equation $\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$ subject to the conditions

$u(0,t) = 0, u(x,0) = x(1-x), u(1,t) = 0$. Take $h = 0.1$ and $t = 0, 1, 2$.

Ans.

0	0.09	0.16	0.21	0.24	0.25	0.24	0.21	0.16	0.09	0
0	0.08	0.15	0.20	0.23	0.24	0.23	0.20	0.15	0.08	0

10. Solve the heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \text{ for } 0 < x < 1, 0 < t < 0.04$$

subject to the conditions

$$u(x,0) = 4x - 4x^2, 0 \leq x \leq 1$$

$$u(0,t) = 0, 0 \leq t \leq 0.04$$

$$u(1,t) = 0 \text{ for } 0 \leq t \leq 0.04.$$

Ans.

0	0.64	0.96	0.96	0.64	0
0	0.48	0.80	0.80	0.48	0
0	0.40	0.64	0.64	0.40	0

11. Use Crank–Nicolson method for solving the heat equation $u_t = u_{xx}$ for $0 < x < 1$ and $0 < t < 0.03$ subject to the conditions

$$u(x,0) = \sin(\pi x) + \sin(2\pi x), 0 \leq x \leq 1$$

$$u(x,t) = 0, u(1,t) = 0 \text{ for } 0 \leq t \leq 0.03.$$

Use $h = 0.1$, $k = 0.01$ and $r = 1$.

Ans.

0.897	1.539	1.760	1.539	1.0	0.363	-0.142	-0.363	-0.279
0.679	1.179	1.379	1.262	0.907	0.463	0.087	-0.113	-0.119
0.5525	0.922	1.104	1.053	0.822	0.511	0.226	0.0444	-0.017

12. Solve the heat equation $u_t(x,t) = u_{xx}(x,t), 0 < x < 5, t > 0$ by Crank–Nicolson method subject to the conditions

$$u(x,0) = 20, u(0,t) = 0, u(5,t) = 100 \text{ and taking } h = 1, k = 1.$$

Ans.

0	20	20	20	20	100
0	9.80	20.19	30.72	59.92	100

13. Solve the wave equation $u_{tt} = u_{xx}$ up to $t = 0.2$ with spacing 0.1 subject to the conditions

$$u(0,t) = 0, u(1,t) = 0$$

$$u_t(x,0) = 0, u(x,0) = 10 + x(1-x).$$

Ans.

0	10.09	10.16	10.21	10.24	10.25	10.24	10.21	10.16	10.09	0
0	10.09	10.16	10.21	10.24	10.25	10.24	10.21	10.16	10.09	0
0	0.07	10.14	10.19	10.22	10.23	10.22	10.19	10.17	0.7	0

14. Solve the wave equation $u_{tt} = u_{xx}$ for $x = 0, 0.1, 0.2, 0.3, 0.4$ and $0 \leq t \leq 0.2$ subject to the conditions

$$u(x, 0) = \frac{1}{8} \sin \pi x, u_t(x, 0) = 0, \quad 0 \leq x \leq 1$$

$$u(0, t) = u(1, t) = 0, \quad t \leq 0.$$

Ans.

0	0.037	0.070	0.096	0.113	0.119
0	0.031	0.059	0.082	0.096	0.101
0	0.023	0.043	0.059	0.07	0.074

15. Solve $u_{xx} + u_{yy} = 0$ for the following square meshes with the boundary values shown in the Figure 11.42

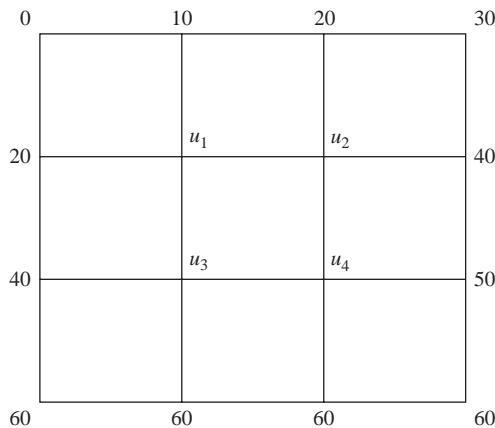


Figure 11.42

Ans. $u_1 = 26.65, u_2 = 33.32, u_3 = 43.31, u_4 = 46.65$

12 Elements of C Language

12.1 PROGRAMMING LANGUAGE

A language that processes some kind of data and provides some meaningful result known as information is called a programming language. The given data and the information obtained are represented by the character set of the language. There are two types of programming languages:

- (a) **Low-Level Languages.** Machine language and assembly language are called low-level languages. These languages are machine oriented and control the computer's internal circuitry. As such these languages require extensive knowledge of electronic circuits of a computer. In machine language, the instructions are very detailed and are written in binary codes to which the computer responds directly. However, a machine language program written for one type of computer may not run on another type of computer without altering it significantly. On the other hand, a language in which the instructions are written using symbolic names for machine operations and operands is called assembly language. For example, ADD, READ, STORE are symbolic instructions in assembly language.
- (b) **High-Level Languages.** In high-level languages, the instruction set is more compatible with human languages and human thought processes. In these languages, we use English keywords, selection, variables, constants, and iterations. The advantage of high-level language over the low-level language is that a high level language is simple, uniform, and portable. As such it is machine independent and so can be used on any type of computer. A program written in high-level language is converted into machine language before it can be executed. This task of conversion is called compilation or interpretation and is performed by compiler or interpreter. The translation is carried out automatically within the computer. The original high-level program, which is input to a compiler, is called the source program, whereas the resulting machine language program is called the object program.

12.2 C LANGUAGE

C language, developed by Dennis M. Ritchie in 1970 at Bell Telephone Laboratories, is a general-purpose, structured computer programming language. It is an outgrowth of two languages BCPL and B developed at Bell Laboratories. C language is a high-level language, but it also contains additional features that allow it to be used at low level. Therefore, it is also called middle-level language and bridges the gap between machine language and the high-level languages. Thus, C language can be used for writing both system programs and application programs. Every C program has one or more modules. Any module can be modified without much difficulty and without affecting the remaining modules.

12.3 C TOKENS

The smallest individual units of C program are known as tokens. The C language provides the following tokens:

- (i) **Keywords.** Keywords are the reserved words with fixed meaning that convey a special meaning to the compiler. The keywords are written in lowercase. The standard keywords in C are

ada	extern	signed
asm	far	static
auto	for	struct
break	float	switch
case	goto	typedef
char	huge	union
const	it	unsigned
continue	int	void
default	long	volatile
do	near	while
double	register	
enum	return	
entry	short	
else	size of	

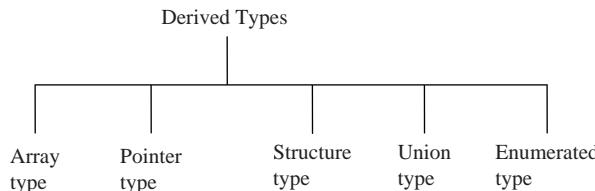
- (ii) **Identifiers.** The names given to program elements, like variables, symbolic constants, function, and array are called identifiers. The following rules should be followed while naming the identifiers:
- (a) The first character of the identifier must be a letter followed by digits, lowercase letters or underscore (_). For example, *y*, *x34*, *table*, *chair*, *x_12* are identifiers but *3rd* is not an identifier because its first letter is not a letter. Similarly, *tax rate* is not an identifier because the blank space between its elements cannot be used. The underscore is considered as a letter and is used for better understanding of the identifier's name.
 - (b) C language is case sensitive and so uppercase and lowercase letters are different. Thus, the names *tax-rate* and *Tax-rate* are different identifiers. In fact, uppercase letters are used for symbolic constants, whereas the lowercase letters are used for variable names.
- (iii) **Constants.** The data items which do not change their values during a program execution are called constants or literally. The constants are further classified as: (a) character constants comprising of single character constants and string constants, and (b) numeric constants comprising integer constants and real constants. For example, 7, 41, 346 are integer constants, 'A', '7', 'x' are character constants, "red", "bell", "London" are string constants. Note that a character constant is enclosed in single quotation marks, while a string constant is enclosed in double quotation marks. Integer constants are exact quantities, while floating point constant is a base-10 number that contains either a decimal point or an exponent or both. For example, 0.000846, 18.36, 2378.0012, 3e5 are floating point constants and thus are approximations. The floating point constants cannot be used for counting/indexing since in such cases exact values are required.
- (iv) **Variables.** A variable is a data item that may be used to store a data value. It takes different values at different times during the execution of a program. For example, if we come across *int a, b, c*, in a C program, then *a*, *b*, *c* represent integer variables, while *a = 7*, *b = 6*, *c = 2*, etc., represent values given to these variables.
- (v) **Data Types in C.** The following types of data appear in C programming:

1. Basic Data Types

Data Type	Description	Memory required
char	Single character	1 byte
int	Integer quantity	2 bytes

Data Type	Description	Memory required
float	Floating point number	4 bytes
double	Floating point number (larger in magnitude)	8 bytes

2. Derived Data Types



Array: An array is a kind of variable that refers to a collection of data items that all have the same name. All the data items must be of the same type (e.g., all integers, all characters etc.). Consider the following statement:

`int a[10];`

This describes an array of 10 integer values. The first value of array element can be referred as `a[1]`.

Category of arrays:

1. One-dimensional array
 2. Two-dimensional array
 3. Three-dimensional array
- } Multidimensional array

One-dimensional arrays. One-dimensional arrays are suitable for processing of lists of items of identical types. They are useful for problems that require the same operation to be performed on a group of data. For example,

`int a[10];`

Multidimensional arrays. An array having more than one subscript is known as multidimensional array.

A two-dimensional array is suitable for table processing or matrix manipulations. For this purpose, one can use two subscript enclosed in square brackets. The first subscript represents row and second subscript represents column. For example,

`int matrix [3][2];`

The above example represents a two-dimensional array having three rows and two columns. The elements of the array can be referred as `a[i][j]`, where `i` represents the row and `j` represents column.

- (vi) **Operators.** The symbols representing specific operations are called operators. The data items on which an operator acts are called operands. The operators used in C language are as follows:

(A) Arithmetic Operators:

Operator	Action	Example
+	Addition	$c = a+b$
-	subtraction	$c = a-b$
*	multiplication	$c = a*b$
/	Division	$c = a/b$
% (modulus operator)	Remainder after division	$c = a \% b$

For example, let the operands a and b be 18 and 7, respectively. Then

$$\begin{aligned} a + b &= 18 + 7 = 25, \quad a - b = 18 - 7 = 11, \\ a * b &= 18 * 7 = 126, \quad a/b = 18/7 = 2.571, \\ a \% b &= 18 \% 7 = 4. \end{aligned}$$

Similarly, if a and b are floating point variables with values 80.24 and 4.0, respectively, then

$$\begin{aligned} a + b &= 84.25, \quad a - b = 76.24, \\ a * b &= 160.48, \quad a/b = 20.06, \text{ and} \quad a \% b = 0. \end{aligned}$$

Remarks:

- (i) C language does not have exponentiation operator. To carry out exponentiation, the library function (pow) is used.
- (ii) If a is an integer variable and b is a floating point variable, then their sum $a + b$ is floating point. Therefore, the expression $(a + b)\%c$, where c is an integer, is not valid. But the expression

$$(\text{int})(a + b)\%c$$
- is valid because $(\text{int})(a + b)$ is integral value of $a + b$. Similarly, for a floating point variable, the expression $((\text{int})a)\%c$, where c is an integer is valid.
- (iii) The hierarchy of the arithmetic operators is
 $\ast, /, \%, +, -$.

(B) Unary Operator: An operator that operates upon a single operand to produce a new value is called an unary operator.

In C language, we have the following commonly used unary operators:

Unary operator	Operand	Value obtained
$-$ (minus)	a	$-a$
$++$ (increment operator)	a	$a = a+1$ (operand increased by 1)
$--$ (decrement operator)	a	$a = a-1$ (operand decreased by 1)

Remark: If we use $a ++$ then current value will be displayed first and then the value is increased by 1.

(C) Relational Operators: The operators that are used to compare two variables or constants are called relational operators.

C language has the following relational operators (hierarchy wise):

Operator	Meaning
$<$	less than
\leq	less than or equal to
$>$	greater than
\geq	greater than or equal to
$=$	equal to
\neq	not equal to

(D) Logical Operators. The operators used to combine logical conditions are called logical operators. The following logical operators are used in C language:

Operator	Meaning
!	logical negation NOT
&&	logical AND in a complex expression
	logical OR in a complex expression

For example, if 1 and 0 are two operands, then $1 \&\& 1 = 1$, $1 \&\& 0 = 0$, $0 \&\& 1 = 0$, $0 \&\& 0 = 0$, $1 || 1 = 1$, $1 || 0 = 1$, $0 || 1 = 1$, $0 || 0 = 0$, $!1 = 0$, $!0 = 1$.

(E) **Conditional Operators.** The conditional operator (?: :) is used to carry out simple condition operations. The general form of the expression that use conditional operators is

(text – expression) ? T: expression:F – expression,

or

expression 1 ? expression 2: expression 3,

where expression 1 is evaluated first; if expression 1 is true, then expression 2 is evaluated else expression 3 is evaluated. Since such expression contains three parts, therefore conditional operator is also called ternary operator. The expression is called ternary expression or conditional expression. The conditional expressions are used in place of if–else statements.

For example, consider the expression

$(i == 0) ? 0; 1$

This expression means that if i is equal to 0, then the expression takes the value 0, otherwise the entire conditional expression takes the value 1.

(F) **Assignment Operators.** In C programs, the following assignment operators are used:

Operator	Action
=	simple assignment
+=	assign sum
-=	assign difference
*=	assign product
/=	assign quotient
%=	assign remainder

For example, let i and j be integers with values 8 and 7, respectively, then we have

Expression	Meaning	Value
$i += 6$	$i = i + 6$	$8 + 6 = 14$
$i -= j$	$i = i - j$	$8 - 7 = 1$
$j *= (i - 4)$	$j = j * (i - 4)$	$7(4) = 28$
$i /= 2$	$i = i / 2$	$8 / 2 = 4$
$i \% = (j - 3)$	$i = i \% (j - 3)$	$8 \% 4 = 0$

12.4 LIBRARY FUNCTIONS

C compilers contain library functions to carry out various calculations or operations. Some commonly used library functions are listed below:

Function	Purpose
1. abs (<i>i</i>)	Returns the absolute value of the integer <i>i</i>
2. getchar ()	Enter a character from the standard input device
3. printf (---)	Send data item to the standard output device
4. putchar	Send a character to the standard output device
5. scanf (---)	Enter data items from the standard input device
6. cos (<i>d</i>)	Returns the cosine of <i>d</i>
7. sin (<i>d</i>)	Returns sine of <i>d</i>
8. tan (<i>d</i>)	Returns the tangent of <i>d</i>
9. sinh (<i>d</i>)	Returns the hyperbolic sine of <i>d</i>
10. cosh (<i>d</i>)	Returns the hyperbolic cosine of <i>d</i>
11. tanh (<i>d</i>)	Returns the hyperbolic tangent of <i>d</i>
12. sqrt (<i>d</i>)	Returns the square root of <i>d</i>
13. ceil (<i>d</i>)	Round up to the next integer that is greater than or equal to <i>d</i>
14. floor (<i>d</i>)	Round down to the next integer value that is largest integer that does not exceed <i>d</i>
15. pow(<i>d</i> ₁ , <i>d</i> ₂)	Returns <i>d</i> ₁ raised to the power <i>d</i> ₂
16. fab(<i>d</i>)	Returns the absolute value of <i>d</i>

All mathematical functions in the above list are available in a header file math.h, which is included in the beginning of the program as #include <math.h>. The remaining functions are available in a header file stdio.h which is included in the beginning of the C program as #include <stdio.h>. Thus, standard input–output file is included in the program. In the above notations, h stands for “header file.”

12.5 INPUT OPERATION

The scanf() is used to enter input data into the computer from the input device (keyboard). The syntax of the function scanf() is

scanf(“format specification string”, list of addresses of variable); where the format specification string (also called control string) begins with % (percentage sign) followed by character which indicates the type of corresponding data item. The conversion characters and their meaning are listed below:

Conversion character	Meaning
c	the data item (variable) is character
d	the variable is a decimal integer
f	the data item is a floating point value
s	the data item is a floating point value followed by a white space character (string terminator)
u	the data item is an unsigned decimal integer

The list of addresses of variables is separated by commas and the sign & (ampersand) is put before each variable.

For example, consider

```
scanf ("%c %d %f", &ch, &d, &f);
```

The control string is “%c %d %f”. The conversion characters used are c, d, and f.

The consecutive nonwhite space characters that define a data item collectively define a field. An unsigned integer indicating the field width is placed within the format specification string between the % sign and the conversion character.

For example, consider

```
scanf("%2d %2d", &a, &b);
```

Suppose that the input data items entered are

3 4 5,

then the assignment shall be

a = 3, b = 4, and c = 5.

If the data items entered are

34 56 78,

then the assignment shall be

a = 34, b = 56, c = 78.

Similarly, if the data items entered are

345678,

then the assignment will be

a = 34, b = 56, c = 78.

If the data items entered are

345 678 92,

then the assignment will be

a = 34, b = 5, c = 67 (ignoring other digits)

If the data items entered are

3 453 56

then the assignment will be

a = 3, b = 45, c = 3.

12.6 OUTPUT OPERATION

The printf() function is used to format out to standard output device (screen). The syntax of printf() function is given by

```
printf("control string", list of variables);
```

The control string begins with % followed by a conversion character. The conversion characters for output along with their meaning are listed below:

Conversion character	Meaning
c	the data item displayed as a character
d	the data item is displayed as a signed decimal integer
f	the data item is displayed as a floating point without exponent
e	the data item is displayed as a floating point with exponent
s	the data item is displayed as a string until a null character is encountered

The list of variables is comma separated.

For example, consider

```
printf("%d %d", i, j, sqrt (i + j));
```

If $i = 25$, $j = 24$, then execution of the program will yield

25	24	7
----	----	---

12.7 CONTROL (SELECTION) STATEMENTS

C language has the following statements, called control or selection statements, which allow to choose a set of instructions for the execution depending upon the test condition.

- (a) ***if statement.*** The *if* statement tests a condition. If the condition is true, then the statement (s) associated with *if* is (are) executed otherwise the statement (s) is (are) neglected or ignored. The syntax of *if* statement is

```
if (condition)
    statement (s);
```

For example, consider

```
void main ( )
{
    int a, b, c;
    float d, x1, x2;
    printf("Enter the value of a, b and c");
    scanf("%d %d", &a, &b, &c);
    d = b*b - 4*a*c;
    if (d>0)
    {
        printf("Roots are real and distinct");
        x1 = (-b + sqrt(d))/(2*a);
        x2 = (-b - sqrt(d))/(2*a);
        printf("The roots are x1 = %f, x2 = %f", x1, x2);
    }
}
```

In this program, we note that if $d = b^2 - 4ac > 0$, then the roots of the equation $ax^2 + bx + c = 0$ are real and distinct.

- (b) ***if-else statement.*** The *if-else* statement allows to choose any one set of statements out of two sets of statements depending upon the test condition. The syntax of *if-else* statement is given below:

```
if (condition)
    statement 1;
else
    statement 2;
```

For example, consider

```
void main ( )
{
    int a, b, c;
    float d, x1, x2;
```

```

printf("Enter the value of a, b, c");
scanf("%d %d", &a, &b, &c);
d = b*b - 4*a*c;
if (d>0)
{
    printf("Roots are real and distinct");
    x1 = (-b + sqrt(d))/(2*a);
    x2 = (-b - sqrt(d))/(2*a);
    printf("The roots are x1 = %f, x2 = %f, x1, x2");
}
if (d == 0)
{
    printf("The roots are real and equal");
    x1 = -b/(2*a);
    printf("The roots is %f", x1);
}
else
{
    printf("Roots are complex");
}

```

- (c) **Nested if–else statement.** The if–else statements in a C program may form a nest. For example, consider the following part of a program in C, written to find the type of a given triangle.

```

main()
{
    float a, b, c;
    printf("\n Enter three sides \n \n");
    scanf("%f %f %f", &a, &b, &c);
    printf("\n The given sides are \n \n");
    printf("%8.2f %8.2f \n \n", a, b, c);
    if ((a + b) > c ) && ((a + c ) > b) && ((b + c ) > a)
    {
        if (a == b) && (a == c))
        {
            printf("\n Equilateral triangle \n");
        }
    }
    else
    {
        if ((a == b) || (a == c) || (b == c))
            printf("\n Isosceles triangle \n");
    }
}

```

```

{
    printf("\n Scalene triangle \n");
}
}
else
{
    printf("\n Triangle not possible \n");
}
}

```

In this part of the program the if–else statements are nested.

- (d) **The switch statement.** The switch statement is a multiple branch solution statement. It successively tests the value of an expression against a list of integer or character constants. If an appropriate match is found, then the statement(s) associated with that statement is (are) executed. The syntax of the switch statement is as given below:

```

switch (expression or variable)
{
    case constant 1:
        statement  ) – 1  (s
        break;
    case constant 2:
        statement  ) – 2  (s
        break;
    -----
    -----
    case constant n:
        statement (s) – n
        break;
    default:
        default tatement
        break;
}

```

- (e) **Loops.** The statements that allow a set of instructions to be performed until some specified condition is satisfied, are called loops or looping statements. There exist three types of loops in C language namely, *for* loop, *while* loop, and *do while* loop.

- (i) ***for* loop.** The *for* loop is used when the number of iterations is known in advance. The syntax of the *for* loop is

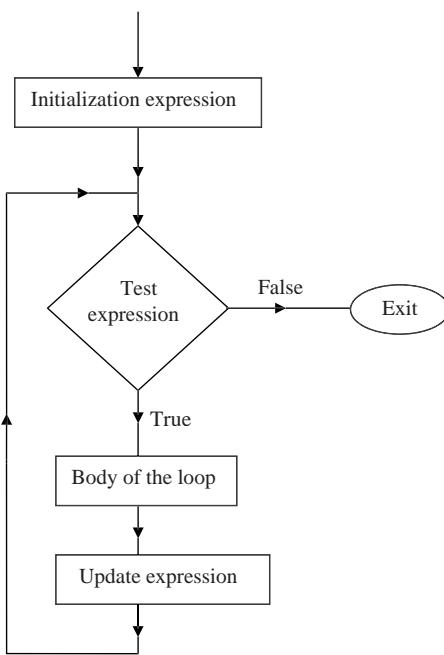
```

for (initialization, test expression, update expression (s))
{
    body of the loop
}

```

We note that the body of the *for* loop is enclosed by braces and not by semicolon.

The flow chart for the *for* loop is shown below:



Suppose that we want to calculate factorial of an integer n . Then the main() shall be as shown below:

```

main ()
{
    int f, i, n
    printf("Enter the value of n");
    scanf("%d", &n);
    for (f=1, i=1, i<=n, i++)
    {
        f=f*i;
    }
    printf("%d", f);
}
  
```

In this part of the program, the initialization statements are $f = 1$, $i = 1$, whereas the test expression is $i \leq n$. The body of the loop is $f = f * i$ and the update expression (also called re-initialization) is $i = i + 1$. Thus, i is the loop control variable whose value will decide the termination of the loop.

- (ii) **While loop.** The while loops are used when the number of iterations is not known. The syntax of the while loop is

```

initialization statements;
while (test expression)
{
    body of the loop;
    update expression;
}
  
```

We note that *for* loop and *while* loop are quite similar. The only difference, we observe, is that initialization statement exists outside *while* ().

- (iii) **do while loop.** These loops are also used when the number of iterations is not known. The syntax of the do while loop is

```
initialization statements;
do
{
    body of the loop
}
while (test expression);
```

Thus, for calculating factorial of a number n , the main() will be as given below:

```
main( )
{
    int f, i, n;
    printf("Enter the value of n");
    scanf("%d", &n);
    i = 1;
    f = 1;
    do
    {
        f = f*i;
        i = i+1;
    }
    while (i <= n);
    printf("%d", f);
}
```

12.8 STRUCTURE OF A C PROGRAM

A C program consists of one or more modules called functions. The general structure of a C program is as shown below:

```
documentation
header files
symbolic constants
global variables
main( )
{
declarations
executable statements
}
user-defined function (function subprograms)
fun c1( )
{
}
fun c2( )
{
```

In a C program, the main() function is a must because the program execution always starts with main().

The documentation section contains one or more comments specifying the name, date, and objective of the program. This section is optional and so the program may or may not have it. For a single line, comment begins with //. For a comment of more than one line, the comment is placed between /* and */.

For example, to write a program to solve a non-linear equation by Newton–Raphson method, the documentation may be

```
//Newton–Raphson method
```

or

```
/*Newton–Raphson method*/
```

The Header files are used to include functions from C language library. For example, for writing C program for Newton–Raphson method, the header filers are

```
#include<stdio.h>
#include<math.h>
#include<conio.h>
```

To write C program for calculating area of a circle, the header file will be

```
#include<stdio.h>
```

Similarly, to write C program for finding roots of a quadratic equation, the header file will be

```
#include<stdio.h>
#include<math.h>
```

Symbolic constants are the constants that do not change throughout the program. For example, define PI 3.14159.

Global variables are the variables that are used in more than one function. For example, if we want to calculate both simple interest and compound interest on a principal amount p for time t and at rate r in the same program, then the variables p, r, t are global variables since these are used in both functions used for simple interest and compound interest. Therefore these variables are declared outside of all the functions.

The main() function section is a must in every C program. In the declaration section, all the variables used in the execution part are declared. In executable section, all executable statements are kept. All the statements in this section must be terminated by a semicolon (;). For example, the main() for the roots of a quadratic equation $ax^2 + bx + c = 0$ shall be

```
main()
{
    float a, b, c, d, x1, x2;
    printf("a=");
    scanf("%f", &a);
    printf("b=");
    scanf("%f", &b);
    printf("c=");
    scanf("%f", &c);
    d = sqrt(b*b - 4*a*c);
    x1 = (-b + d)/(2*a);
    x2 = (-b - d)/(2*a);
    printf("\n x1=%e, x2=%e, x1, x2);
}
```

12.9 PROGRAMS OF CERTAIN NUMERICAL METHODS IN C LANGUAGE

1. Program in C Demonstrating Bisection Method

Problem: To find a root of the equation $x^3 - 3x - 5 = 0$ using bisection method, where the root lies between 2 and 3.

Variables: Let

x_1, x_2 be the initial approximations enclosing the root,
 err be the allowed error,
 x_3 be the new approximation to the root in each iteration,
 i be the count for iteration.

Program:

```
/* Bisection Method*/
```

```
#include <stdio.h>
#include <math.h>
#include <conio.h>
#include <process.h>
float f(float x)
{
    return(x*x*x-3*x-5); //defining the function which is to be evaluated
}
int main(int argc, char *argv[])
{
    int i = 0;
    float x1,x2,x3,err;
    printf("\n Enter the initial approximation x1:");
    scanf("%f",&x1);
    printf("\n Enter the initial approximation x2:");
    scanf("%f",&x2);
    if (f(x1)*f(x2)<0.0) //checking whether root lies within x1 & x2
        printf("\n the initial approximations are correct");
    else
    {
        printf("\n The initial approximations are incorrect");
        getch();
        exit(0);
    }
    printf("\n Enter the allowed error in the solution:");
    scanf("%f",&err);
    while(fabs(x2-x1)>err) //loop for the required number of iterations
    {
        i++;
        x3 = (x1+x2)/2;
        printf("\n Iteration=%d has f(x3)=%f and x=%f",i,f(x3),x3);
        if(f(x3)==0.0)
        {
```

```

printf("\n Solution converges");
printf("\n The solution is %f",x3);
exit(0);
}
if((f(x2)*f(x3)<0))
x1=x3;
else
x2=x3;
}

printf("\n Solution converges in iteration %d",i);
printf("\n The solution is %f",x3);
getch();

return 0;
}

```

The execution of the program on the given equation yields the following results:

Enter the initial approximation x1: 2

Enter the initial approximation x1: 2.5

The initial approximations are correct.

Enter the allowed error in the solution: 0.0001

Iteration=1 has f(x3)=-0.359375 and x3=2.25

Iteration=2 has f(x3)=1.271484 and x3=2.375

Iteration=3 has f(x3)=0.428955 and x3=2.3125

Iteration=4 has f(x3)=-0.028107 and x3=2.28125

Iteration=5 has f(x3)=-0.167294 and x3=2.2265625

Iteration=6 has f(x3)=-0.070010 and x3=2.273438

Iteration=7 has f(x3)=-0.021056 and x3=2.277344

Iteration=8 has f(x3)=0.003499 and x3=2.279297

Iteration=9 has f(x3)=-0.008785 and x3=2.278320

Iteration=10 has f(x3)=-0.002644 and x3=2.278809

Iteration=11 has f(x3)=-0.000427 and x3=2.279053

Iteration=12 has f(x3)=-0.001109 and x3=2.278931

Iteration=13 has f(x3)=-0.000341 and x3=2.278992

Solution converges in 13 iterations.

The solution is 2.278992.

2. Program in C Demonstrating Newton–Raphson Method

Problem: Apply Newton-Raphson method to solve the transcendental equation

$$e^x - 5x = 0$$

Formula Used: The Newton-Raphson formula is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Program:

```
/* Newton-Raphson Method */
#include <stdio.h>
#include <math.h>
#include <conio.h>
float f(float x)
{
    return (exp(x)-5*x);
}
float df(float x)
{
    return (exp(x)-5);
}
int main(int argc, char *argv[])
{
    int i=0;
    float x0,x1,e,p;
    printf("\n Enter the initial approximation x0:");
    scanf("%f",&x0);
    printf("\n Enter the Allowed error e in the solution:");
    scanf("%f",&e);
    do
    {
        i++;
        if(df(x0) == 0)
        {
            printf("\n Newton-Raphson method fails");
            getch();
            exit(0);
        }
        x1 = x0-f(x0)/df(x0);
        p = x0;
        x0 = x1;
    }while(fabs(x1-p)>e);
    printf("\n The solution converges in %d Iterations",i);
    printf("\n The solution is % 1.10f",x1);
    return 0;
}
```

The execution of the above program yields

Enter the initial approximation x0: 0.4

Enter the allowed error e in the solution: 0.00000000001

The solution converges in 4 iterations

The solution is 0.2591710985.

3. Program in C Demonstrating Gauss Elimination Method

Problem: Solve by Gaussian elimination method the following system of equations:

$$10x_1 - 7x_2 + 3x_3 + 5x_4 = 6$$

$$-6x_1 + 8x_2 - x_3 - 4x_4 = 5$$

$$3x_1 + x_2 + 4x_3 + 11x_4 = 2$$

$$5x_1 - 2x_2 - 2x_3 + 4x_4 = 7$$

Program:

```
/* Gauss elimination method */
```

```
#include<stdio.h>
#include<conio.h>
#include<math.h>
#define n 4

int main(int argc, char *argv[])
{
float a[n][n+1],x[n],multi,s;
int i,j,k;
printf("enter the elements of aug. matrix row wise\n");
for (i=0;i<n;i++)
for(j=0;j<n+1;j++)
scanf("%f",&a[i][j]); //reading the elements of augmented matrix
for(j=0;j<n-1;j++)
for(i=j+1;i<n;i++) // loops for obtaining upper triangular matrix
{
multi=a[i][j]/a[j][j];
for(k=0;k<n+1;k++)
a[i][k]-=a[j][k]*multi;// row operation is applied
}
printf("the upper triangular matrix\n");
for(i=0;i<n;i++)
{
for(j=0;j<n+1;j++)
printf("%f ",a[i][j]); // printing the upper triangular matrix
printf("\n");
}
for(i=n-1;i>=0;i--) // loop for back substitution
{
s=0;
for(j=i+1;j<n;j++)
s+=a[i][j]*x[j];
x[i]=(a[i][n]-s)/a[i][i];
}
for(i=0;i<n;i++)
printf("x[%d]= %f\n",i+1,x[i]); // printing the values for 4 unknowns.
getch();
```

```
return 0;
}
```

The execution of the program for the given problem yields:

Enter the elements of aug. matrix row wise

```
10 -7 3 5 6
-6 8 -1 -4 5
3 1 4 11 2
5 -9 -2 4 7
```

The upper triangular matrix is

10	-7	3	5	6
0	3.8	0.8	-1	8.6
0	0	2.447368	10.315789	-6.815791
0	0	0	9.924731	9.924732

Hence

$$\begin{aligned}x_1 &= 5.000001 \\x_2 &= 4.000001 \\x_3 &= -7.000001 \\x_4 &= 1.000000.\end{aligned}$$

4. Program in C Demonstrating Gauss–Jordan Method

Problem: Solve by Gaussian–Jordan method the following system of equations:

$$\begin{aligned}10x_1 - 7x_2 + 3x_3 + 5x_4 &= 6 \\-6x_1 + 8x_2 - x_3 - 4x_4 &= 5 \\3x_1 + x_2 + 4x_3 + 11x_4 &= 2 \\5x_1 - 9x_2 - 2x_3 + 4x_4 &= 7\end{aligned}$$

Program:

```
/* Gauss–Jordan Method */
#include<stdio.h>
#include<conio.h>
#include<math.h>
#define n 4

int main(int argc, char *argv[])
{
float a[n][n+1],multi,x[n];
int i,j,k;

printf("enter the elements of aug. matrix row wise\n");
for(i=0;i<n;i++)
for(j=0;j<n+1;j++)
scanf("%f",&a[i][j]);
for(j=0;j<n;j++)
for(i=0;i<n;i++)
```

```

if(i!=j)
{
multi= a[i][j]/a[j][j];
for(k=0;k<n+1;k++)
a[i][k]=a[j][k]*multi;
}
for(i=0;i<n;i++)
{
for(j=0;j<n+1;j++)
printf("%f",a[1][j]);
printf("\n");
}
printf("solution of the equations:\n");
for(i=0;i<n;i++)
printf("x(%d)=%f\n",i+1,a[i][n]/a[i][i]);
getch();
return 0;
}

```

The execution of the program to the given problem yields:

Enter the elements of aug. matrix row wise

10 -7 3 5 6
-6 8 -1 -4 5
3 1 4 11 2
5 -9 -2 4 7

The upper triangular matrix is

10.000001	0	0	0	50.000011
0	3.8	0	0	15.200003
0	0	2.447368	0	-17.131580
0	0	0	9.924731	9.924732

Hence

$$\begin{aligned}
x_1 &= 5.000001, \\
x_2 &= 4.000001, \\
x_3 &= -7.000001, \\
x_4 &= 1.000000.
\end{aligned}$$

5. Program in C demonstrating Gauss–Seidel Method

Problem: Solve the following system of equations by Gauss–Seidel method:

$$5x - y + z = 10$$

$$2x + 8y - z = 11$$

$$-x + y + 4z = 3.$$

(The coefficient matrix of the given system is already diagonally dominant)

Program:

```

/* Gauss-Seidel Method */
#include <stdio.h>
#include<conio.h>
#include <math.h>
#define n 3
main()
{
    int i,j,itr,maxitr;
    float a[n][n+1],x[n],multi,aerr,maxerr,err,s;
    clrscr();
    for(i=0;i<n;i++)
        x[i]=0;
    printf("The Execution of the program for the given problem yields:\n");
    printf("\n Enter the elements of augmented matrix row wise");
    for(i=0;i<n;i++)
        for(j=0;j<n+1;j++)
            scanf("%f",&a[i][j]);
    printf("\n Enter the allowed error,maximum iterations \n");
    scanf("%f",&aerr);
    scanf("%d", &maxitr);
    printf("\n iteration      x[1]      [x2]      [x3]\n");
    for(itr=1;itr<=maxitr;itr++)
    {
        maxerr=0;
        for(i=0;i<n;i++)
        {
            s=0;
            for(j=0;j<n;j++)
                if(i!=j)
                    s+=a[i][j]*x[j];
            multi=(a[i][n]-s)/a[i][i];
            err=fabs(x[i]-multi);
            if (err>maxerr)
                maxerr=err;
            x[i]=multi;
        }
        printf("t%d/t",itr);
        for(i=0;i<n;i++)
            printf("%f\t",x[i]);
        printf("\n");
        if(maxerr<aerr)
        {
            printf("\n Convergence in %d iterations\n",itr);
            for(i=0;i<n;i++)
                printf("\n x[%d]=%f\n",i+1,x[i]);
        }
    }
}

```

```

    return 0;
}
}
printf("\n Solution does not converge,iterations not sufficient \n");
return 1;
}

```

The execution of the program for the given problem yields:

Enter the elements of augmented matrix row wise

```

5
-1
1
10
2
8
-1
11
-1
1
4
3

```

Enter the allowed error, maximum iterations

```

.0001
8

```

iteration	x[1]	x[2]	x[3]
1	2.000000	0.875000	1.031250
2	1.968750	1.011719	0.989258
3	2.004492	0.997534	1.001740
4	1.999159	1.000428	0.999683
5	2.000149	0.999923	1.000057
6	1.999973	1.000014	0.999990
7	2.000005	0.999998	1.000002

Convergence in 7 iterations

x[1]=2.000005,
x[2]=0.999998,
x[3]=1.000002.

6. Program in C Demonstrating Lagrange's Interpolation Method

Problem: The function $y = f(x)$ is given in the points (7,3), (8,1), (9,1), and (10,9). Find the value of y for $x = 9.5$ using Lagrange's interpolation formula.

Formula Used: For the points $(x_0, y_0), (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, the Lagrange's interpolation formula is

$$\begin{aligned}
y(x) = & \frac{(x - x_1)(x - x_2) \dots (x - x_n)}{(x_0 - x_1)(x_0 - x_2) \dots (x_0 - x_n)} y_0 + \frac{(x - x_0)(x - x_2) \dots (x - x_n)}{(x_1 - x_0)(x_1 - x_2) \dots (x_1 - x_n)} y_1 \\
& + \dots + \frac{(x - x_0)(x - x_1) \dots (x - x_{n-1})}{(x_n - x_0)(x_n - x_1) \dots (x_n - x_{n-1})} y_n
\end{aligned}$$

Program:

```

/* Lagrange's Interpolation Method */
#include <conio.h>
#define MAX 6
int main(int argc, char *argv[])
{
int i,j,n;
float ax[MAX+1],ay[MAX+1],num,denom,x,y=0;
printf("\n Enter the number of intervals:");
scanf("%d",&n);
printf("Enter the values x and y for the pair:\n");
for(i=0;i <= n;i++)
    scanf("%f %f",&ax[i],&ay[i]);
printf("Enter the value of x for which value of y is required:");
scanf("%f",&x);
for(i=0; i <= n; i++)
{
    num = denom = 1;
    for(j=0; j <= n; j++)
    {
        if(i != j)
        {
            num *= x - ax[j];
            denom *= ax[i] - ax[j];
        }
    }
    y += (num/denom)*ay[i];
}
printf("\n The value of y for x=%f is %f",x,y);
getch();
return 0;
}

```

The execution of the program for the given data yields:

Enter the number of intervals: 3

Enter the value of the x and y for the pair:

7	3
8	1
9	1
10	9

Enter the value of x for which the value of y is required 9.5.

The value of y for x = 9.500000 is 3.62500.

7. Program in C Demonstrating Least Square Method to Fit a Straight Line to a Given Data

Problem: Find a straight line approximation to the points (2,2), (5,4), (6,6), (9,9), and (11,10) using the Least Square Method.

Formula Used: The equation of the straight line is $y = a + bx$. The normal equations are

$$\begin{aligned} na + b \sum x_i &= \sum y_i \\ a \sum x_i + b \sum x_i^2 &= \sum x_i y_i \end{aligned}$$

Program:

```
/*
C code for fitting a straight line to a given data
*/
#include<stdio.h>
#include<conio.h>
#include<math.h>
int main(int argc, char *argv[])
{
    float aug[2][3]={{0,0,0},{0,0,0}};
    float multi,a,b,x,y,xsq;
    int i,j,k,n;
    printf("Enter the number of pairs of observed values:");
    scanf("%d",&n);
    aug[0][0]=n;
    for(i=0;i<n;i++)
    {
        printf("pair no.%d:",i+1);
        scanf("%f %f",&x,&y);
        xsq=x*x;
        //summation of x,y,xsq,x*y
        aug[0][1]+=x;
        aug[0][2]+=y;
        aug[1][1]+=xsq;
        aug[1][2]+=x*y;
    }
    aug[1][0]=aug[0][1];
    printf("The Entries of aug.matrix are: \n");
    for(i=0;i<2;i++)
    {
        for( j=0;j<3;j++)
            printf("%f ",aug[i][j]);
        printf("\n");
    }
    printf("\n");
/*
Application of Gauss-Jordan method to reduce the coefficient matrix to diagonal matrix
*/
    for(j=0;j<2;j++)
        for(i=0;i<2;i++)
```

```

if (i!=j)
{
    multi=aug[i][j]/aug[j][j];
    for(k=0;k<3;k++)
        aug[i][k]-=aug[j][k]*multi;
}
a=aug[0][2]/aug[0][0];
b=aug[1][2]/aug[1][1];
printf("a=%f b=%f",a,b);
printf("\n\nThe Least Square Line is:\n");
printf(" y = %f x",b);
printf(a>0?" +%f":" %f",a);
getch();
return 0;
}

```

The execution of the program to the given data yields:

Enter the number of pairs of observed values: 5

Pair no. 1: 2 2
 Pair no. 2: 5 4
 Pair no. 3: 6 6
 Pair no. 4: 9 9
 Pair no. 5: 11 10

The entries of the augmented matrix are

5.00000	33.00000	31.00000
33.00000	267.00000	251.00000
$a = -0.024390$,	$b = 0.943089$	

The Least Square Line is:

$$y = 0.943089x - 0.024390.$$

8. Program in C Demonstrating Least Square Method to Fit a Parabola to a Given Data

Problem: Fit a parabola by least square approximation method to the pair of points (0,-2.1), (1,-0.4), (2,2.1), (3,3.6), (4,9.9).

Normal Equations: The equation of the parabola is $y = a + bx + cx^2$. The normal equations are

$$\begin{aligned} na + b \sum x_i + c \sum x_i^2 &= \sum y_i \\ a \sum x_i + b \sum x_i^2 + c \sum x_i^3 &= \sum x_i y_i \\ a \sum x_i^2 + b \sum x_i^3 + c \sum x_i^4 &= \sum x_i^2 y_i \end{aligned}$$

Program:

```

/*
To fit a parabola to a given data
*/

```

```
#include<stdio.h>
#include<conio.h>
#include<math.h>

int main(int argc, char *argv[])
{
    float augm[3][4]={ (0,0,0,0), (0,0,0,0), (0,0,0,0) };
    float t,a,b,c,x,y,xsq;
    int i,j,k,n;
    printf ("Enter the no. of pairs of observed values:");
    scanf ("%d", &n);
    augm[0][0] = n;
    for (i=0;i<n;i++)
    {
        printf ("\n");
        printf ("pair no.%d:", i+1);
        scanf ("%f %f", &x, &y);
        xsq=x*x;
        augm[0][1]+=x;
        augm[0][2]+=xsq;
        augm[1][2]+=x*xsq;
        augm[2][2]+=xsq*xsq;
        augm[0][3]+=y;
        augm[2][3]+=xsq*y;
        augm[1][3]+=x*y;
    }
    augm[1][1]=augm[0][2]; augm[2][1]=augm[1][2];
    augm[1][0]=augm[0][1]; augm[2][0]=augm[1][1];
    printf ("\n");
    printf ("The augmented matrix is:\n");
    for(i=0;i<3;i++)
    {
        for(j=0;j<4;j++)
        {
            printf ("%f ", augm[i][j]);
        }
        printf ("\n");
    }
    printf ("\n");
    // solve by Gauss-Jordan method
    for (j=0;j<3;j++)
    for (i=0;i<3;i++)
    if(i!=j)
    {
        t=augm[i][j]/augm[j][j];
        for(k=0;k<4;k++)
        augm[i][k]-=augm[j][k]*t;
    }
}
```

```

a=augm[0][3]/augm[0][0];
b=augm[1][3]/augm[1][1];
c=augm[2][3]/augm[2][2];
printf(" a=%f \n b=%f\n c=%f \n",a,b,c);
printf("The equation of the parabola of best fit is:\n");
printf("y = %f",a);
printf((b > 0)?"+ %f x":"%f",b);
printf((c > 0)?"+ %f x2":"%f",c);
getch();
return 0;
}

```

The execution of the program to the given data yields:

Enter the number of pairs of observed values: 5

Pair no. 1: 0 -2.1
 Pair no. 2: 1 -0.4
 Pair no. 3: 2 2.1
 Pair no. 4: 3 3.6
 Pair no. 5: 4 9.9

The entries of the augmented matrix are

5.000000	10.000000	30.000000	13.099999
10.000000	30.000000	100.000000	54.199997
30.000000	100.000000	354.000000	198.799988

$$a = -1.808572, \quad b = 0.457143, \quad c = 0.585714.$$

The equation of the parabola of best fit is:

$$y = -1.808572 + 0.457143x + 0.585714x^2.$$

9. Program in C Demonstrating Trapezoidal Rule

Problem: Evaluate the integral $\int_0^6 (1+x^2)dx$ using trapezoidal rule by dividing the interval of integration into six equal parts.

Formula Used: The trapezoidal rule for integration is

$$\int_{x_0}^{x_n} f(x)dx = \frac{h}{2} [f(x_0) + f(x_n) + 2(f(x_1) + f(x_2) + \dots + f(x_{n-1}))]$$

Program:

```

/*
Trapezoidal Rule
*/
#include<stdio.h>
#include<conio.h>
#include<math.h>
float f(float x)

```

```

{
return (1+x*x);
}

int main(int argc, char *argv[])
{
    int i,n;
    float x0,xn,h,s;
    printf("Enter the values of x0,xn,n:");
    scanf("%f %f %d",&x0,&xn,&n);
    h=(xn-x0)\n;
    s=f(x0)+f(xn);
    for(i=1;i<=n-1;i++)
        s+=2*f(x0+i*h);
    printf("The value of the integral is: %f", (h/2)*s);
    getch();
    return 0;
}

```

The execution of the program for the given problem yields:

```

Enter x0, ..., n: 0 6 6
The value of the integral is: 79.00000

```

10. Program in C Demonstrating Simpson's 1/3 Rule

Problem: Evaluate the integral $\int_0^6 (1+x^2) dx$ using Simpson's 1/3 rule by dividing the interval of integration into six equal parts.

Formula Used: The Simpson's 1/3 rule for integration is

$$\int_{x_0}^{x_n} f(x) dx = \frac{h}{3} [f(x_0) + f(x_n) + 2(f(x_2) + f(x_4) + \dots + f(x_{n-2})) + 4(f(x_1) + f(x_3) + \dots + f(x_{n-1}))]$$

Program:

```

/*
Simpson's 1/3 rule
*/
#include<stdio.h>
#include<conio.h>
#include<math.h>
float f(float x)
{
    return((1+x*x));
}
int main(int argc, char *argv[])
{
    int i,n;
    float x0,xn,h,s;

```

```

printf("Enter the value of x0,xn,h:");
scanf("%f %f %d",&x0,&xn,&n);
h=(xn-x0)\n;
s=f(x0)+f(xn)+4*f(x0+h);
for(i=3;i<=n-1;i+=2)
    s+= 4* f(x0+i*h)+ 2 * f(x0+(i-h)*h);
printf("The value of the integral is equal to:%f", (h*s)/3);
getch();
return 0;
}

```

The execution of the program for the given problem yields:

Enter $x_0, x_n, n: 0 \ 6 \ 6$

The value of the integral is : 78.00000

11. Program in C Demonstrating Simpson's 3/8 Rule

Problem: Evaluate the integral $\int_0^6 (1+x^2) dx$ using Simpson's 3/8 Rule by dividing the interval of integration into six equal parts

Formula Used: The Simpson's 3/8 rule for integration is

$$\int_{x_0}^{x_n} f(x) dx = \frac{3h}{8} \left[f(x_0) + f(x_n) + 3(f(x_1) + f(x_2) + f(x_4) + f(x_5) + \dots + f(x_{n-1})) + 2(f(x_3) + f(x_6) + \dots + f(x_{n-3})) \right]$$

Program:

```

/*
Simpson's 3/8 Rule
*/
#include<stdio.h>
#include<conio.h>
#include<math.h>

float f(float x)
{
    return((1+x*x));
}

int main(int argc, char *argv[])
{
    int i,n;
    float x0,xn,h,sum;
    printf("Enter the value of x0,xn,n:");
    scanf("%f %f %d",&x0,&xn,&n);
    h = (xn-x0)\n;
    sum = f(x0)+f(xn);
    for(i=1;i<=n-1;i++)
    {
        if(i%3 != 0)
            sum += 3 * f(x0+i*h);
    }
}

```

```

    else
        sum += 2* f(x0+i*h);
}
printf("The value of integral is:%f", (3*h/8)*sum);
getch();
return 0;
}

```

The execution of the program for the given problem yields:

Enter x_0 , x_n , n : 0 6 6

The value of the integral is: 78.00000.

12. Program in C Demonstrating Euler's Method

Problem: Solve the initial value problem

$$\frac{dy}{dx} = \frac{y-x}{y+x}, \quad y(0) = 1,$$

for $x = 0.1$ by Euler's method.

Program:

```

/*Euler's Method*/
#include <stdio.h>
#include <conio.h>
#include <math.h>

float f(float x, float y)
{
    return ((y-x)/(y+x));
}

int main(int argc, char *argv[])
{
    float x0,y0,h,x1,y1,x;
    printf("\n Enter the value of x0, y0,h,x:");
    scanf("%f %f %f %f", &x0, &y0, &h, &x);
    x1=x0;
    y1=y0;
    if(h > x)
    {
        printf("Error:\nPlease enter valid values:\n");
        return;
    }
    while(1)
    {
        y1+=h* f(x1,y1);
        x1+=h;
        if(x1 > x)
        {

```

```

        return;
    }
    printf("\n The value of y at x=%3.2f is %4.6f\n",x1,y1);
}
getch();
return 0;
}

```

The execution of the program for the given problem yields:

Enter the values of x0, y0, h, x : 0 1 0.02 0.1

The value of y at 0.02 is 1.020000
The value of y at 0.04 is 1.039231
The value of y at 0.06 is 1.057748
The value of y at 0.08 is 1.075601
The value of y at 0.1 is 1.092832

13. Program in C Demonstrating Runge–Kutta Method

Problem: Apply fourth order Runge–Kutta method to solve

$$\frac{dy}{dx} = x^2 + y^2, \quad y(0) = 1$$

correct to four decimal places for $x = 0.1$ and $x = 0.2$ taking step size $h = 0.1$.

Formula Used: The formula for Runge–Kutta method is

$$y_{n+1} = y_n + \frac{1}{6}(K_1 + 2(K_2 + K_3) + K_4),$$

where

$$\begin{aligned}
K_1 &= hf(x_n, y_n) \\
K_2 &= hf\left(x_n + \frac{h}{2}, y_n + \frac{K_1}{2}\right) \\
K_3 &= hf\left(x_n + \frac{h}{2}, y_n + \frac{K_2}{2}\right) \\
K_4 &= hf(x_n + h, y_n + K_3)
\end{aligned}$$

Program:

```

/* Program in C demonstrating Runge-Kutta Method*/
#include <stdio.h>
#include <conio.h>
#include <math.h>

float f(float x,float y)
{
    return(x*x+y*y);
}

int main(int argc, char *argv[])
{

```

```

float x0,y0,h,xn,x,y,k1,k2,k3,k4,k;
printf("Enter the values of x0,y0,h,xn:");
scanf("%f %f %f %f",&x0,&y0,&h,&xn);
x=x0;y=y0;
printf("\n\nThe execution of the program to the given problem yields:\n");
while(1)
{
    if(x==xn)
    {
        return;
    }
    k1=h*f(x,y);
    k2=h*f(x+h/2,y+k1/2);
    k3=h*f(x+h/2,y+k2/2);
    k4=h*f(x+h,y+k3);
    k=(k1+(k2+k3)*2+k4)/6;
    x+=h;
    y+=k;
    printf("when x=%f, then y=%f \n\n",x,y);
}
getch();
return 0;
}

```

The execution of the program to given problem yields:

Enter the values of x0, y0, h, x: 0 1 0.1 0.2

The value of y at 0.1 is 1.111463

The value of y at 0.2 is 1.253012

14. Program in C Demonstrating Milne–Simpson’s Method

Problem. Solve the differential equation

$$\frac{dy}{dx} = x^2 + y^2 - 2, \quad y(0) = 1$$

for $x = 0.4$ using Milne–Simpson’s method.

The starting four values found by Runge–Kutta method are $y_0 = 1, y_1 = 0.91005, y_2 = 0.7846, y_3 = 0.6421$.

Program:

/*Milne predictor corrector method */

#include <stdio.h>

#include<conio.h>

#include<math.h>

float f (float x, float y)

{

return ((x*x) + (y*y) - 2);

}

void main()

{

```

int i,n;
float x0,h,x,t;
float f1,f2,f3,f4,y[10],z,e=0.0005;
clrscr();
i=4;
printf("\n enter the value of x0,y0 and x");
scanf("%f %f %f",&x0,&y[0],&x);
printf("\n enter the value of y1,y2 and y3");
for(i=1;i<=3;i++)
scanf("%f",&y[i]);
printf("\n enter the number of subdivisions of interval (x,x0)");
scanf("%d",&n);
h=(x-x0)/n;
f1=f(x0+h,y[1]);
f2=f(x0+2*h,y[2]);
f3=f(x0+3*h,y[3]);
y[i]=y[i-4]+(4*h*(2*f1-f2+2*f3))/3;
f4=f(x0+i*h,y[4]);
y[i]=y[i-2]+h*(f2+4*f3+f4)/3;
printf("\n the value of y at x=%2f is %5f",x0+i*h,y[4]);
getch();
}

```

The execution of the program for the given problem yields:

enter the value of x0,y0 and x

0

1

0.4

enter the value of y1,y2 and y3

0.91005

0.7846

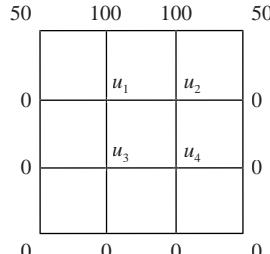
0.6421

enter the number of subdivisions of interval (x,x0)4

the value of y at x=0.400000 is 0.486124

15. Program Demonstrating solution of Laplace's Equation

Problem: Solve Laplace's equation $u_{xx} + u_{yy} = 0$ for the following square meshes with boundary conditions exhibited in the figure given below:



Program:

```

// Solution of Laplace's Equation by Five Point Formula
#include<stdio.h>
#include<conio.h>
#include<math.h>
#define sqr 4
typedef float array[sqr+1][sqr+1];
void getrow(int i, array u)
{
int j;
printf("enter the values of u[%d,j],j=1,%d\n",i,sqr);
for(j=1;j<=sqr;j++)
scanf("%f",&u[i][j]);
}
void getcol(int j, array u)
{
int i;
printf("enter the values of u[i,%d],i=2,%d\n",j,sqr-1);
for(i=2;i<=sqr-1;i++)
scanf("%f",&u[i][j]);
}
void printarr(array u)
{
int i,j;
for(i=1;i<=sqr;i++)
{
for(j=1;j<=sqr;j++)
printf("%f\t",u[i][j]);
printf("\n");
}
}
main()
{
array u;
float maxerr,aerr,err,t;
int i,j,itr,maxitr;
clrscr();
for(i=1;i<=sqr;i++)
for(j=1;j<=sqr;j++)
u[i][j]=0;
printf("enter the boundary cond.");
getrow(1,u);
getrow(sqr,u);
getcol(1,u);
getcol(sqr,u);
printf("enter allowed error,max iteration");

```

```

scanf("%f%d",&aerr,&maxitr);
for(itr=1;itr<=maxitr;itr++)
{
maxerr=0;
for(i=2;i<=sqr-1;i++)
for(j=2;j<=sqr-1;j++)
{
t=(u[i-1][j]+u[i+1][j]+u[i][j+1]+u[i][j-1])/4;
err=fabs(u[i][j]-t);
if(err>maxerr)
maxerr=err;
u[i][j]=t;
}
printf("iteration no %d\n",itr);
printarr(u);
if(maxerr<=aerr)
{
printf("after %d iteration \n" " the solution is \n",itr);
printarr(u);
return 0;
}
}
printf("iterations are not sufficient");
return 1;
}

```

The execution of the program for the given problem yields:

```

enter the values of u[i,1],i=2,3
0
0
enter the values of u[i,4],i=2,3
0
0
enter allowed error,max iteration.01
10

```

iteration no 1

50.000000	100.000000	100.000000	50.000000
0.000000	25.000000	31.250000	0.000000
0.000000	6.250000	9.375000	0.000000
0.000000	0.000000	0.000000	0.000000

iteration no 2

50.000000	100.000000	100.000000	50.000000
0.000000	34.375000	35.937500	0.000000
0.000000	10.937500	11.718750	0.000000
0.000000	0.000000	0.000000	0.000000

iteration no 3

50.000000	100.000000	100.000000	50.000000
0.000000	36.718750	37.109375	0.000000
0.000000	12.109375	12.304688	0.000000
0.000000	0.000000	0.000000	0.000000

iteration no 4

50.000000	100.000000	100.000000	50.000000
0.000000	37.304688	37.402344	0.000000
0.000000	12.402344	12.451172	0.000000
0.000000	0.000000	0.000000	0.000000

iteration no 5

50.000000	100.000000	100.000000	50.000000
0.000000	37.451172	37.475586	0.000000
0.000000	12.475586	12.487793	0.000000
0.000000	0.000000	0.000000	0.000000

iteration no 6

50.000000	100.000000	100.000000	50.000000
0.000000	37.487793	37.493896	0.000000
0.000000	12.493896	12.496948	0.000000
0.000000	0.000000	0.000000	0.000000

iteration no 7

50.000000	100.000000	100.000000	50.000000
0.000000	37.496948	37.498474	0.000000
0.000000	12.498474	12.499237	0.000000
0.000000	0.000000	0.000000	0.000000

after 7 iterations the solution is

50.000000	100.000000	100.000000	50.000000
0.000000	37.496948	37.498474	0.000000
0.000000	12.498474	12.499237	0.000000
0.000000	0.000000	0.000000	0.000000

16. Program in C Demonstrating Bender–Schmidt Method to Solve One-Dimensional Heat Equation

Problem. Solve, by Bender–Schmidt method, the one-dimensional heat equation $\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2}$ subject to the conditions $u(0, t) = u(4, t) = 0$, $u(x, 0) = x(4 - x)$.

Program:

```
/* Solution of One-Dimensional Heat Equation by Bender-Schmidt Method */
```

```
#include<stdio.h>
#include<math.h>
#include<conio.h>

float cal(float x)
{

```

```

return(4*x-x*x);
}
void main()
{
    int i,j,r,xe,te;
    float h,k,u[6][6],a,csq,uxbt,uxet;
    printf("\n enter the value of xe and te");
    scanf("%d %d",&xe,&te);
    printf("\n enter the value of uxbt and uxet");
    scanf("%f %f",&uxbt,&uxet);
    printf("\n enter the value of h,k and csq");
    scanf("%f %f %f",&h,&k,&csq);
    for(j=0;j<=te;j++)
    {
        u[0][j]=uxbt;
        u[xe][j]=uxet;
    }
    for(i=0;i<xe;i++)
        u[i][0]=cal(i);
    for(j=0;j<te;j++)
        for(i=1;i<xe;i++)
            u[i][j+1]=(u[i-1][j]+u[i+1][j])/2;
        printf("\n the value of uij are");
    for(j=0;j<=te;j++)
    {
        printf("\n \n");
        for(i=0;i<=xe;i++)
            printf("\t %.2f",u[i][j]);
    }
    getch();
}

```

The execution of the program for the given problem yields:

enter the value of xe and te

4

3

enter the value of uxbt and uxet

0

0

enter the value of h, k and csq

1

1

0.5

The value of uij are

0.00	3.00	4.00	3.00	0.00
0.00	2.00	3.00	2.00	0.00
0.00	1.50	2.00	1.50	0.00
0.00	1.00	1.50	1.00	0.00

This page is intentionally left blank

Appendix: Model Question Papers

Model Paper I

1. (a) Determine $f(x)$ as a polynomial in x for the following data:

$x:$	-4	-1	0	2	5
$f(x):$	1245	33	5	9	1355

- (b) Fit a parabola $y = a + bx + x^2$ to the following data:

$x:$	2	4	6	8	1
$y:$	3.07	12.85	31.47	57.38	91.29

0

2. (a) Find by Newton–Raphson method, the real root of the equation

$$3x = \cos x + 1.$$

- (b) Apply Muller's method to find a root of the equation

$$x^3 - x^2 - x - 1 = 0.$$

3. (a) Solve the equations by Gauss–Seidel's method:

$$20x + y - 2z = 17,$$

$$3x + 20y - z = -18,$$

$$2x - 3y + 20z = 25.$$

- (b) Solve by Gauss–Jordan method

$$x + y + z = 9$$

$$2x - 3y + 4z = 13$$

$$3x + 4y + 5z = 40.$$

4. (a) Find $f'(10)$ from the following table:

$x:$	3	5	11	27	34
$f(x):$	-13	23	899	17315	35606

- (b) Use Romberg's method to compute $\int_0^1 \frac{dx}{1+x^2}$ correct to four decimal places.

5. Using Euler modified method, obtain a solution of $\frac{dy}{dx} = x + \sqrt{|y|}$, $y(0) = 1$ for the range $0 \leq x \leq 0.6$ in steps of 0.2.

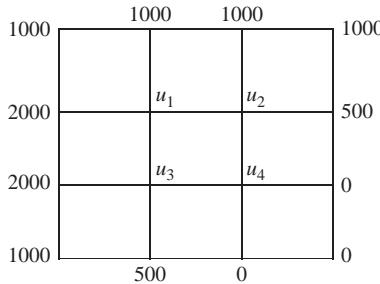
6. Solve the initial value problem

$$\frac{dy}{dx} = x - y^2, y(0) = 1$$

to find $y(0.4)$ by Adam's method. Starting solutions required are to be obtained using Runge–Kutta method of order 4, using step value $h = 0.1$.

A-2 ■ Numerical Methods

7. Given the values of $u(x,y)$ on the boundary of the square in the figure below, evaluate the function $u(x,y)$ satisfying the Laplace equation $\nabla^2 u = 0$ at the pivotal point of the figure.



8. Find the value of $u(x,t)$ satisfying the parabolic equation $\frac{\partial u}{\partial t} = 4 \frac{\partial^2 u}{\partial x^2}$ and the boundary conditions $u(0,t) = 0 = u(8,t)$ and $u(x,0) = 4x - \frac{1}{2}x^2$ at the points $x = i$, $i = 0, 1, 2, \dots, 7$ and $t = \frac{1}{8}j$, $j = 0, 1, 2, \dots, 5$.

SOLUTIONS

1. (a) Example 5.27.

- (b) The sum table for the given problem is

n	x	x^2	x^3	x^4	y	xy	x^2y
1	2	4	8	16	3.07	6.14	12.28
1	4	16	64	256	12.85	51.4	205.6
1	6	36	216	1296	31.47	188.82	1132.92
1	8	64	512	4096	57.38	459.04	3672.32
1	10	100	1000	1000	91.29	912.9	9129.00
5	30	220	1800	15664	196.06	1618.3	14152.12

The normal equations are

$$5a + 30b + 220c = 196.06,$$

$$30a + 220b + 1800c = 1618.30,$$

$$220a + 1800b + 15644c = 14152.12.$$

These equations yield

$$40b + 480c = 44.94$$

and

$$480b + 5984c = 5525.48.$$

This last pair of equations gives $b = -0.859$ and $c = 0.992$. Putting these values in the first normal equation, we get $a = 0.720$.

Hence, the least square parabola is

$$y = 0.72 - 0.859x + 0.992x^2.$$

2. (a) The given equation is

$$f(x) = 3x - \cos x - 1 = 0.$$

We have

$$f(0) = -2(-\text{ve}) \text{ and } f(1) = 3 - 0.5403 - 1 = 1.4597(+\text{ve}).$$

Hence, one of the roots of $f(x) = 0$ lies between 0 and 1. The values at 0 and 1 show that root is nearer to 1. So let us take $x = 0.6$. Further,

$$f'(x) = 3 + \sin x.$$

Therefore, the Newton-Raphson formula gives

$$\begin{aligned} x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{3x_n - \cos x_n - 1}{3 + \sin x_n} \\ &= \frac{3x_n + x_n \sin x_n - 3x_n + \cos x_n + 1}{3 + \sin x_n} = \frac{x_n \sin x_n + \cos x_n + 1}{3 + \sin x_n}. \end{aligned}$$

Hence,

$$\begin{aligned} x_1 &= \frac{x_0 \sin x_0 + \cos x_0 + 1}{3 + \sin x_0} = \frac{0.6(0.5646) + 0.8253 + 1}{3 + 0.5646} = 0.6071, \\ x_2 &= \frac{x_1 \sin x_1 + \cos x_1 + 1}{3 + \sin x_1} = \frac{(0.6071)(0.5705) + 0.8213 + 1}{3 + 0.5705} = 0.6071. \end{aligned}$$

Hence the required root, correct to four decimal places, is 0.6071.

(b) The given equation is

$$f(x) = x^3 - x^2 - x - 1 = 0.$$

We note that

$$f(0) = -1, f(1) = -2, f(2) = 1.$$

Thus, one root lies between 1 and 2. Let

$$x_{i-2} = 0, x_{i-1} = 1, \text{ and } x_i = 2,$$

$$y_{i-2} = -1, y_{i-1} = -2, \text{ and } y_i = 1.$$

Therefore,

$$\begin{aligned} A &= \frac{(x_{i-2} - x_i)(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i-2} - y_i)}{(x_{i-1} - x_{i-2})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{(-2)(-3) - (-1)(-2)}{(1)(-1)(-2)} = \frac{6 - 2}{2} = 2. \end{aligned}$$

$$\begin{aligned} B &= \frac{(x_{i-2} - x_i)^2(y_{i-1} - y_i) - (x_{i-1} - x_i)^2(y_{i-2} - y_i)}{(x_{i-2} - x_{i-1})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{(-2)^2(-3) - (-1)^2(-2)}{(-1)(-1)(-2)} = \frac{-12 + 2}{-2} = 5. \end{aligned}$$

Therefore,

$$\begin{aligned} x_{i+1} &= 2 - \frac{2(1)}{5 + \sqrt{25 - 4(2)(1)}} \\ &= 2 - \frac{2}{5 + 4.123} = 2 - 0.2192 = 1.7808. \end{aligned}$$

A-4 ■ Numerical Methods

We note that $f(1.7808) = -0.53625$ (–ve). Thus, the root lies between 1.78 and 2. Therefore, for the second iteration, we set

$$x_{i-2} = 1, x_{i-1} = 1.78, \text{ and } x_i = 2.$$

Then

$$y_{i-2} = -2, y_{i-1} = -0.536, y_i = 1.$$

Therefore,

$$\begin{aligned} A &= \frac{(x_{i-2} - x_i)(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i-2} - y_i)}{(x_{i-1} - x_{i-2})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{(1-2)(-0.536-1) - (1.78-2)(-2-1)}{(1.78-1)(1.78-2)(1-2)} \\ &= \frac{1.536 - 0.66}{(0.78)(-0.22)(-1)} = \frac{0.836}{0.1716} = 4.872, \\ B &= \frac{(-1)^2(-1.536) - (-0.22)^2(-3)}{(1-1.78)(1.78-2)(1-2)} \\ &= \frac{-1.536 + 0.1452}{(-0.78)(-0.22)(-1)} = \frac{1.3908}{0.1716} = 8.10. \end{aligned}$$

Hence,

$$\begin{aligned} x_{i+1} &= 2 - \frac{2(1)}{8.1 + \sqrt{65.61 - 4(4.87)1}} \\ &= 2 - \frac{2}{8.1 + \sqrt{46.13}} \\ &= 2 - \frac{2}{8.1 + 6792} = 1.87. \end{aligned}$$

We note $f(1.87) = 0.173$. Therefore, $x = 1.87$ is a satisfactory root.

3. (a) The given equation can be written as

$$x = \frac{1}{20}[17 - y + 2z],$$

$$y = \frac{1}{20}[-18 - 3x + z],$$

$$z = \frac{1}{20}[25 - 3x + 3y].$$

Taking the initial notation as $(x_0, y_0, z_0) = (0, 0, 0)$, we have by Gauss–Seidal method,

$$x_1 = \frac{1}{20}[17 - 0 + 0] = 0.85,$$

$$y_1 = \frac{1}{20}[-18 - 3(0.85) + 1] = -1.0275,$$

$$z_1 = \frac{1}{20}[25 - 2(0.85) - 3(1.0275)] = 1.0108,$$

$$x_2 = \frac{1}{20}[17 + 1.0275 + 2 \pm (1.0108)] = 1.0024,$$

$$y_2 = \frac{1}{20}[-18 - 3(1.0024) + 1.0108] = -0.9998,$$

$$z_2 = \frac{1}{20}[25 - 2(1.0024) + 3(-0.9998)] = 0.9998,$$

$$x_3 = \frac{1}{20}[17 + 0.9998 + 2(0.9998)] = 0.99997,$$

$$y_3 = \frac{1}{20}[-18 - 3(0.99997) + 0.9998] = -1.00000,$$

$$z_3 = \frac{1}{20}[25 - 2(0.99997) + 3(-1.00000)] = 1.00000.$$

The second and third iterations show that the solution of the given system of equations is $x = 1$, $y = -1$, $z = 1$.

(b) The augmented matrix for the given system is

$$\begin{aligned} m_{21} &= 2 \begin{bmatrix} 1 & 1 & 1 & 9 \\ 2 & -3 & 4 & 13 \end{bmatrix} \leftarrow \text{Pivotal row} \\ m_{31} &= 3 \begin{bmatrix} 1 & 1 & 1 & 9 \\ 2 & -3 & 4 & 13 \\ 3 & 4 & 5 & 40 \end{bmatrix} \end{aligned}$$

The first Gauss–Jordan elimination yields

$$\begin{aligned} m_{12} &= -\frac{1}{5} \begin{bmatrix} 1 & 1 & 1 & 9 \\ 0 & -5 & 2 & -5 \end{bmatrix} \leftarrow \text{Pivotal row.} \\ m_{32} &= -\frac{1}{5} \begin{bmatrix} 0 & 1 & 2 & 1 \\ 0 & 1 & 2 & 1 \end{bmatrix} 3 \end{aligned}$$

The second Gauss–Jordan elimination yields

$$\begin{aligned} m_{13} &= \frac{7}{12} \begin{bmatrix} 1 & 0 & \frac{7}{5} & 8 \\ 0 & -5 & 2 & -5 \end{bmatrix} \\ m_{23} &= \frac{10}{12} \begin{bmatrix} 0 & 0 & \frac{12}{5} & 12 \end{bmatrix} \leftarrow \text{Pivotal row} \end{aligned}$$

The third Gauss–Jordan elimination yields

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & -5 & 0 & -15 \\ 0 & 0 & \frac{12}{5} & 12 \end{bmatrix}.$$

A-6 ■ Numerical Methods

Thus, we have attained the diagonal form of the system. Hence the solution is

$$x = 1, y = \frac{15}{5} = 3, z = \frac{12(5)}{12} = 5.$$

4. (a) Since the spacing is unequal, we use differentiation formula derived from Newton's divided difference formula. The formula is (see Article 7.9, Expression 7.35).

$$f'(x) \approx P'(x) = a_1 + a_2[(x - x_0) + (x - x_1)], \quad (1)$$

where

$$a_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0},$$

and

$$a_2 = \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0}.$$

In the given data, we have

$$x_0 = 5, x_1 = 11, x_2 = 27, x = 10.$$

Therefore,

$$\begin{aligned} a_1 &= \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{899 - 23}{11 - 5} = 146, \\ a_2 &= \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0} = \frac{\frac{17315 - 899}{16} - \frac{899 - 23}{6}}{22} \\ &= \frac{1.26 - 146}{22} = 40. \end{aligned}$$

Therefore, equation (1) yields

$$f'(10) = 146 + 40[(10 - 5) + (10 - 11)] = 306.$$

- (b) Let $h = 0.5$. The values of the integrand $f(x) = \frac{1}{1+x^2}$ are

$x:$	0	0.5	1
$f(x)$	1	0.8	0.5

Therefore, by trapezoidal rule, we have

$$Q_1 = \int_0^1 \frac{dx}{1+x^2} = \frac{h}{2} [f_0 + 2f_1 + f_2] = \frac{1}{4} [1 + 2(0.8) + 0.5] = 0.775.$$

Now, let $h = 0.25$. Then, the values of the integrand are

$x:$	0	0.25	0.5	0.75	1.0
$f(x):$	1	0.9412	0.8	0.64	0.5

Therefore, by trapezoidal rule,

$$\begin{aligned} Q_2 &= \int_0^1 \frac{dx}{1+x^2} = \frac{h}{2} [f_0 + 2(f_1 + f_2 + f_3) + f_4] \\ &= \frac{1}{8} [1 + 2(0.9412 + 0.8 + 0.64) + 0.5] = 0.7828. \end{aligned}$$

Then, by Romberg's method, we get

$$R_1 = \frac{4}{3}Q_2 - \frac{1}{3}Q_1 = \frac{4}{3}(0.7828) - \frac{1}{3}(0.775) = 0.7854.$$

Now, let $h = 0.125$. Then the values of the integrand are

$x:$	0	0.125	0.25	0.375	0.5	0.625	0.75	0.875	1.0
$f(x):$	1	0.9846	0.9412	0.8767	0.8	0.7191	0.64	0.5664	0.5

Then, by Simpson's $\frac{1}{3}$ rule, we have

$$\begin{aligned} R_2 &= \frac{0.125}{3} [1.4(0.9846 + 0.8767 + 0.7191 + 0.5664) + 2(0.9412 + 0.8 + 0.64) + 0.5] \\ &= \frac{0.125}{3} [1 + 12.5872 + 4.7624 + 0.5] = 0.7854. \end{aligned}$$

Therefore, by Romberg's method,

$$\begin{aligned} S &= \frac{16}{15}R_2 - \frac{1}{15}R_1 = \frac{16(0.7854)}{15} - \frac{1}{15}(0.7854) \\ &= 0.83776 - 0.05236 = 0.7854. \end{aligned}$$

5. The given differential equation is

$$\frac{dy}{dx} = x + |\sqrt{y}|, y(0) = 1.$$

The modified Euler's formula is

$$y_{n+1} = y_n + hf\left(x_n + \frac{h}{2}, y_n + \frac{h}{2}f(x_n, y_n)\right).$$

Taking $h = 0.2$, we have

$$\begin{aligned} y_1 &= y_0 + 0.2 \left[x_0 + \frac{0.2}{2} + |y_0| + \frac{0.2}{2} \left(x_0 + |\sqrt{y_0}| \right) \right] \\ &= 1 + 0.2 [0.1 + 1 + 0.1(1)] = 1.240 \\ y_2 &= y_1 + 0.2 \left[x_1 + \frac{0.2}{2} + |y_1| + \frac{0.2}{2} \left(x_1 + |\sqrt{y_1}| \right) \right] \\ &= 1.24 + 0.2 [0.2 + 0.1 + 1.24 + 0.1(0.2 + 1.11)] \\ &= 1.24 + 0.33428 = 1.574, \\ y_3 &= y_2 + 0.2 \left[x_2 + \frac{0.2}{2} + |y_2| + \frac{0.2}{2} \left(x_2 + |\sqrt{y_2}| \right) \right] \\ &= 1.574 + 0.2(0.4 + 0.1 + 1.54 + 0.1(0.4 + 1.25)) \\ &= 2.0219. \end{aligned}$$

6. We have

$$\frac{dy}{dx} = x - y^2, \quad y(0) = 1.$$

Therefore, taking step value $h = 0.1$, we get

$$\begin{aligned} k_1 &= hf(x_0, y_0) = 0.1(0 - 1) = -0.1, \\ k_2 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{k_1}{2}\right) = 0.1f(0.05, 0.95) = -0.08525, \\ k_3 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{k_2}{2}\right) = 0.1f(0.05, 0.9574) = -0.0867, \\ k_4 &= hf(x_0 + h, y_0 + k_3) = 0.1f(0.1, 0.9137) = -0.07341. \end{aligned}$$

Hence,

$$\begin{aligned} y_1 &= y(0.1) = y_0 + \frac{1}{6}[k_1 + 2k_2 + 2k_3 + k_4] \\ &= 1 + \frac{1}{6}[-0.1 + 2(-0.08525) + 2(-0.0867) + 0.07341] \\ &= 0.9117 \end{aligned}$$

Similarly,

$$\begin{aligned} y_2 &= y(0.2) = 0.8494. \\ y_3 &= y(0.3) = 0.8061 \end{aligned}$$

Thus,

$$\begin{aligned} f_0 &= x_0 - y_0^2 = -1. \\ f_1 &= x_1 - y_1^2 = 0.1 - (0.9117)^2 = -0.7312, \\ f_2 &= x_2 - y_2^2 = 0.2 - (0.8494)^2 = -0.5215, \\ f_3 &= x_3 - y_3^2 = 0.3 - (0.8061)^2 = -0.3498. \end{aligned}$$

Therefore, Adams–Basforth formula

$$y_{n+1} = y_n + \frac{h}{24}[55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3}]$$

yields

$$\begin{aligned} y_4 &= y_3 + \frac{h}{24}[55f_3 - 59f_2 + 37f_1 - 9f_0] \\ &= 0.8061 + \frac{0.1}{24}[55(-0.3498) - 59(-0.5215) + 37(-0.7312) - 9(-1)] \\ &= 0.8061 - 0.02718 = 0.779. \end{aligned}$$

Therefore,

$$f_4 = x_4 - y_4^2 = 0.4 - (0.779)^2 = -0.2068.$$

Now, by Adams–Moulton formula, we have

$$\begin{aligned}y_4 &= y_3 + \frac{h}{24}[9f_4 + 19f_3 - 5f_2 + f_1] \\&= 0.8061 + \frac{0.1}{24}[9(-0.2068) + 19(-0.3498) - 5(-0.5215) - 0.7312] \\&= 0.77847.\end{aligned}$$

7. Exercise 3, Chapter 11. From Figure 11.38, we have on setting $u_4 = 0$,

$$\begin{aligned}u_1 &= \frac{1}{4}[1000 + 2000 + 1000 + 0] = 1000 \text{ (diagonal five point formula),} \\u_2 &= \frac{1}{4}[1000 + 0 + 500 + 1000] = 625 \text{ (standard five point formula),} \\u_3 &= \frac{1}{4}[1000 + 500 + 2000 + 0] = 875 \text{ (standard five point formula),} \\u_4 &= \frac{1}{4}[0 + 875 + 625 + 0] = 375 \text{ (standard five point formula).}\end{aligned}$$

Now using Gauss–Seidel's formula, we have

$$\begin{aligned}u_1^{(1)} &= \frac{1}{4}[2000 + 625 + 875 + 1000] = 1125, \\u_2^{(1)} &= \frac{1}{4}[1125 + 500 + 375 + 1000] = 750, \\u_3^{(1)} &= \frac{1}{4}[2000 + 500 + 1125 + 375] = 1000, \\u_4^{(1)} &= \frac{1}{4}[1000 + 0 + 0 + 750] = 437.5, \\u_1^{(2)} &= \frac{1}{4}[2000 + 1000 + 750 + 1000] = 1187.5, \\u_2^{(2)} &= \frac{1}{4}[1000 + 500 + 1187.5 + 437.5] = 781.25, \\u_3^{(2)} &= \frac{1}{4}[2000 + 500 + 437.5 + 1187.5] = 1031.25, \\u_4^{(2)} &= \frac{1}{4}[0 + 0 + 1031.25 + 781.25] = 453.125.\end{aligned}$$

Continuing with the process, we shall get

$$u = 1208.3, u_2 = 791.7, u_3 = 1041.7, u_4 = 458.4.$$

8. The given equation is

$$\frac{\partial u}{\partial t} = 4 \frac{\partial^2 u}{\partial x^2}, \quad u(0, t) = u(8, t) = 0$$

and

$$u(x, 0) = 4x - \frac{1}{2}x^2 \text{ at } x = i, i = 0, 1, 2, \dots, 7, t = \frac{1}{8}j, j = 0, 1, 2, \dots, 5.$$

Comparing with the standard form $\frac{\partial u}{\partial t} = c^2 \frac{\partial^2 u}{\partial x^2}$, we have $c^2 = 4$. Also $h = 1$, $k = \frac{1}{8}$. Therefore $r = \frac{c^2 k}{h^2} = \frac{1}{2}$. Hence, Bentre-Schmidt formula

$$u_{i,j+1} = \frac{1}{2}[u_{i-1,j} + u_{i+1,j}] \quad (1)$$

is applicable.

The boundary conditions imply that

$$u_{0,t} = 0 \text{ for any } t \text{ and } u_{8,t} = 0 \text{ for any } t.$$

Further,

$$u(x, 0) = 4x - \frac{1}{2}x^2$$

implies

$$u_{i,0} = 4i - \frac{1}{2}x^2.$$

Putting $i = 0, 1, \dots, 7$, we get the entries of the first row as

$$0, 3.5, 6, 7.5, 8, 7.5, 6, 3.5, 0.$$

Putting $j = 0$ in equation (1), we get

$$u_{i,1} = \frac{1}{2}(u_{i-1,0} + u_{i+1,0})$$

and so taking $i = 1, 2, \dots, 7$, we get the values of the second row as

$$u_{11} = \frac{1}{2}(u_{0,0} + u_{2,0}) = \frac{1}{2}(0 + 6) = 3,$$

$$u_{21} = \frac{1}{2}(u_{1,0} + u_{3,0}) = \frac{1}{2}(3.5 + 7.5) = 5.5,$$

and so on.

Thus, the entries in the second row are

$$0, 3, 5.5, 7, 7.5, 7, 5.5, 3, 0.$$

Then the entries in the third row are obtained by putting $j = 1$ in equation (1).

These are

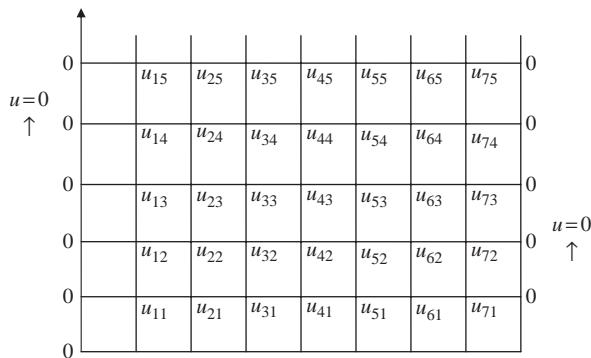
$$0, 2.75, 5, 6.5, 7, 6.5, 5, 2.75, 0.$$

Similarly, the entries in the fourth, fifth, and sixth rows are

$$0, 2.5, 4.625, 6, 6.5, 6, 4.625, 2.5, 0$$

$$0, 2.3125, 4.25, 5.5625, 6, 5.5625, 4.25, 2.3125, 0$$

$$0, 2.125, 3.9375, 5.125, 5.5625, 5.125, 3.9375, 2.125, 0.$$



Model Paper II

1. (a) By using the method of least squares, find a relation of the form $y = ax^b$ that fits the data:

$x:$	2	3	4	5
$y:$	27.8	62.1	110	161

- (b) Find a cubic polynomial in x which takes on the following values $-3, 3, 11, 27, 57$ and 107 , when $x = 0, 1, 2, 3, 4$ and 5 respectively.
2. (a) Find by Newton-Raphson method, the real root of the equation $3x = \cos x + 1$ correct to four decimal places.
- (b) Find by Regula-Falsi method, the real root of the equation $\log x - \cos x = 0$ correct to four decimal places,
3. (a) Solve: by Relaxation method, the following equations:

$$\begin{aligned} 3x + 9y - 2z &= 11; \\ 4x + 2y + 13z &= 24; \\ 4x - 4y + 3z &= -8. \end{aligned}$$

- (b) Solve the following equations by Gauss-Seidal method.

$$\begin{aligned} 20x + y - 2z &= 17; \\ 3x + 20y - z &= -18; \\ 2x - 3y + 20z &= 25. \end{aligned}$$

4. (a) Evaluate: $\int_0^1 \frac{dx}{1+x^2}$
by using:
(i) Simpson's $\frac{1}{3}$ rd rule taking $h = \frac{1}{4}$
(ii) Simpson's $\frac{3}{8}$ th rule taking $h = \frac{1}{6}$. And compare the results with actual values.
(b) Given that:

$x:$	1.0	1.1	1.2	1.3	1.4	1.5	1.6
$y:$	7.989	8.403	8.781	9.129	9.451	9.750	10.031

Find $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ at $x = 1.6$

5. Given $\frac{dy}{dx} = \frac{y-x}{y+x}$ with boundary conditions $y = 1$ when $x = 0$, find approximately y for $x = 0.1$ by:
- Euler's method
 - Modified Euler's method.
6. (a) Solve the boundary value problem defined by $y'' - x = 0$ and $y(0) = 0$, $y'(1) = -\frac{1}{2}$ by Reyleigh-Ritzmethod.
- (b) Determine the largest eigenvalue and the corresponding eigen vector of the matrix using power method.

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

7. (a) Solve the equation $\nabla^2 u = -10(x^2 + y^2 + 10)$ over the square with sides $x = 0 = y$, $x = 3 = y$ with $u = 0$ on the boundary and mesh length = 1
- (b) Derive a difference equation to represent a Poisson's equation.

8. Solve the wave equation:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$$

subject to initial condition

$$u = f(x), \frac{\partial u}{\partial t} = g(x), 0 \leq x \leq 1 \text{ at } t = 0$$

and the boundary conditions:

$$u(0, t) = \phi(t), u(1, t) = \psi(t)$$

SOLUTIONS

1. (a) The sum table for the given problem is

x	x^2	x^4	y	yx^2
2	4	16	27.8	111.2
3	9	81	62.1	558.9
4	16	256	110	1760
5	25	625	161	4025
		978		6455.1

Then for $y = ax^2$, we have

$$a = \frac{\sum y_i x_i^2}{\sum x_i^4} = \frac{6455.1}{978} = 6.60.$$

Hence, the power fit is

$$y = 6.6x^2.$$

- (b) The difference table for the given data is

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
0	-3	6		
1	3	8	2	6
2	11	16	8	6
3	27	30	14	6
4	57	50	20	6
5	107			

Using Newton's forward difference formula, we have

$$\begin{aligned}
 f_x &= f_0 + x\Delta f_0 + \frac{x(x-1)}{2!}\Delta^2 f_0 + \frac{x(x-1)(x-2)}{3!}\Delta^3 f_0 \\
 &= -3 + 6x + \frac{x^2 - x}{2}(2) + \frac{x^3 - 3x^2 + 2x}{6}(6) \\
 &= x^3 - 3x^2 + 2x + x^2 - x + 6x - 3 \\
 &= x^3 - 2x^2 + 7x - 3.
 \end{aligned}$$

2. (a) The given equation is

$$f(x) = 3x - \cos x - 1 = 0.$$

We have

$$f(0) = -2 \text{ (-ve)} \text{ and } f(1) = 3 - 0.5403 - 1 = 1.4597 \text{ (+ve).}$$

Hence, one of the roots of $f(x) = 0$ lies between 0 and 1. The values at 0 and 1 show that root is nearer to 1. So let us take $x = 0.6$. Further,

$$f'(x) = 3 + \sin x.$$

Therefore, the Newton-Raphson formula gives

$$\begin{aligned}
 x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{3x_n - \cos x_n - 1}{3 + \sin x_n} \\
 &= \frac{3x_n + x_n \sin x_n - 3x_n + \cos x_n + 1}{3 + \sin x_n} = \frac{x_n \sin x_n + \cos x_n + 1}{3 + \sin x_n}.
 \end{aligned}$$

Hence,

$$x_1 = \frac{x_0 \sin x_0 + \cos x_0 + 1}{3 + \sin x_0} = \frac{0.6(0.5646) + 0.8253 + 1}{3 + 0.5646} = 0.6071,$$

$$x_2 = \frac{x_1 \sin x_1 + \cos x_1 + 1}{3 + \sin x_1} = \frac{(0.6071)(0.5705) + 0.8213 + 1}{3 + 0.5705} = 0.6071.$$

Hence the required root, correct to four decimal places, is 0.6071.

(b) Let

$$f(x) = \log x - \cos x.$$

Then

$$\begin{aligned} f(1) &= 0 - 0.54 = -0.54 \text{(-ve)} \\ f(1.5) &= 0.176 - 0.071 = 0.105 \text{(+ve).} \end{aligned}$$

Therefore, one root lies between 1 and 1.5 and it is nearer to 1.5.

We start with $x_0 = 1, x_1 = 1.5$. Then, by Regula-Falsi method,

$$x_{n+1} = \frac{x_n f(x_{n-1}) - x_{n-1} f(x_n)}{f(x_{n-1}) - f(x_n)}$$

and so

$$\begin{aligned} x_2 &= \frac{x_1 f(x_0) - x_0 f(x_1)}{f(x_0) - f(x_1)} = \frac{1.5(-0.54) - 1(0.105)}{-0.54 - 0.105} \\ &= 1.41860 \approx 1.42. \end{aligned}$$

But $f(x_2) = f(1.42) = 0.1523 - 0.1502 = 0.0021$. Therefore,

$$\begin{aligned} x_3 &= \frac{x_2 f(x_1) - x_1 f(x_2)}{f(x_1) - f(x_2)} = \frac{1.42(0.105) - 1.5(0.0021)}{0.105 - 0.0021} \\ &= 1.41836 \approx 1.4184. \end{aligned}$$

Now $f(1.418) = 0.151676 - 0.152202 = -0.000526$.

The next iteration is

$$\begin{aligned} x_4 &= \frac{x_3 f(x_2) - x_2 f(x_3)}{f(x_2) - f(x_3)} \\ &= \frac{1.418(0.0021) - (1.42)(-0.000526)}{0.0021 + 0.000526} \\ &= 1.41840. \end{aligned}$$

Since $x_3 = x_4$ upto four decimal places, the required root is $x = 1.4184$

3. (a) The residuals for the given system are

$$R_1 = 3x + 9y - 2z - 11$$

$$R_2 = 4x + 2y + 13z - 24$$

$$R_3 = 4x - 4y + 3z + 8.$$

The operations table for the system is

Δx	Δy	Δz	ΔR_1	ΔR_2	ΔR_3
1	0	0	3	4	4
0	1	0	9	2	-4
0	0	1	-2	13	3

We start with the trivial solution $x = y = z = 0$. The relaxation table is

x_i	y_i	z_i	R_1	R_2	R_3
0	0	0	-11	-24	8
0	2	0	7	-20	0
0	0	1	5	-7	3
1	0	0	8	-3	7
0	0	1	6	10	10
-2	0	0	0	2	2
-0.5	0	0	-1.5	0	0
0	0.1	0	-0.6	0.2	-0.4
0.1	0	0	-0.3	0.6	0

$\sum x_i = -1.6$, $\sum y_i = 2.1$, $\sum z_i = 2$. Hence the approximate solution is $x = -1.6$, $y = 2.1$, $z = 2$.

(b) See question 3(a) of model paper I

4. (a) Example 8.11

Actual value is $\tan^{-1} 1 = \frac{\pi}{4} = 0.7857$.

(b) Differentiating Newton's backward formula, we get

$$\left(\frac{dy}{dx} \right)_{x=x_n} = \frac{1}{h} \left(\nabla y_n + \frac{1}{2} \nabla^2 y_n + \frac{1}{3} \nabla^3 y_n + \dots \right)$$

and

$$\left(\frac{d^2 y}{dx^2} \right)_{x=x_n} = \frac{1}{h^2} \left(\nabla^2 y_n + \nabla^3 y_n + \frac{11}{12} \nabla^4 y_n + \dots \right).$$

The difference table is

x	y						
1.0	7.989	0.414	-0.036				
1.1	8.403	0.378	-0.030	0.006			
1.2	8.781	0.348	-0.026	0.004	-0.002	0.001	
1.3	9.129	0.322	-0.023	0.003	-0.001	0.003	0.002
1.4	9.451	0.299	-0.018	0.005	0.002		
1.5	9.750	0.281					
1.6	10.031						

Therefore, for the given spacing $h = 0.1$, we have

$$\left(\frac{dy}{dx} \right)_{x=1.6} = \frac{1}{0.1} \left[0.281 + \frac{1}{2}(-0.018) + \frac{1}{3}(0.005) + \frac{1}{4}(0.002) \right] = 2.7416.$$

and

$$\left(\frac{d^2y}{dx^2} \right)_{x=1.6} = \frac{1}{0.01} \left[-0.018 + 0.005 + \frac{11}{12}(0.02) \right] = -1.117.$$

5. Example 10.5.

For modified Euler's method, we have

$$y_{n+1} = y_n + h \left[\frac{y_n + \frac{h}{2} \left(\frac{y_n - x_n}{y_n + x_n} \right) - \left(x_n + \frac{h}{2} \right)}{y_n + \frac{h}{2} \left(\frac{y_n - x_n}{y_n + x_n} \right) + \left(x_n + \frac{h}{2} \right)} \right].$$

Thus,

$$\begin{aligned} y(0.02) &= y_1 = y_0 + 0.02 \left[\frac{y_0 + \frac{0.02}{2} \left(\frac{1}{1} \right) - \left(0 + \frac{0.02}{2} \right)}{y_0 + \frac{0.02}{2} \left(\frac{1}{1} \right) + \left(0 + \frac{0.02}{2} \right)} \right] \\ &= 1 + 0.02 \left[\frac{1}{1 + 0.01 + 0.01} \right] = 1 + \frac{0.02}{1.02} = 1.0196. \end{aligned}$$

Proceed likewise to obtain y_2, y_3, y_4, y_5 .

6. (a) Out of scope.

(b) We have

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

We start with

$$x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Then, by power method,

$$AX_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = 1X_2,$$

$$AX_2 = \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = 2X_3,$$

$$AX_3 = \begin{bmatrix} 3 \\ -4 \\ 3 \end{bmatrix} = 4 \begin{bmatrix} \frac{3}{4} \\ -1 \\ \frac{3}{4} \end{bmatrix} = 4X_4,$$

$$AX_4 = \begin{bmatrix} \frac{5}{2} \\ -\frac{14}{4} \\ \frac{5}{2} \end{bmatrix} = \frac{14}{4} \begin{bmatrix} \frac{5}{7} \\ -1 \\ \frac{5}{7} \end{bmatrix} = 3.5X_5,$$

$$AX_5 = \begin{bmatrix} \frac{17}{7} \\ -\frac{24}{7} \\ \frac{17}{7} \end{bmatrix} = \frac{24}{7} \begin{bmatrix} \frac{17}{24} \\ -1 \\ \frac{17}{24} \end{bmatrix} = \frac{24}{7} X_6 = 3.46X_6,$$

$$AX_6 = \begin{bmatrix} \frac{29}{12} \\ -\frac{41}{12} \\ \frac{29}{12} \end{bmatrix} = \frac{41}{12} \begin{bmatrix} \frac{29}{41} \\ -1 \\ \frac{29}{41} \end{bmatrix} = 3.417X_7.$$

Thus, the largest eigenvalue is approximately 3.417 and the corresponding eigenvector is

$$\begin{bmatrix} \frac{29}{41} \\ -1 \\ \frac{29}{41} \end{bmatrix} = \begin{bmatrix} 0.7 \\ -1 \\ 0.7 \end{bmatrix}.$$

7. (a) Example 11.9.

(b) Article 11.7.

8. From Article 11.11, we have $\left(\text{taking } r = \frac{ck}{h} \right)$,

$$u_{i,j+1} + (2 - 2r^2)u_{i,j} + r^2(u_{i+1,j} + u_{i-1,j}) - u_{i,j-1}. \quad (1)$$

But, we are given that

$$\frac{\partial u}{\partial i} = \frac{u_{i,j+1} - u_{i,j-1}}{2k} = g(x) \text{ at } t=0$$

or

$$u_{i,j+1} = u_{i,j-1} + 2k \ g(x) \text{ at } t=0$$

or

$$u_{i,1} = u_{i,-1} + 2k \ g(x) \text{ at } t=0 \text{ and } j=0. \quad (2)$$

Also, $u(x,0) = f(x)$. Therefore,

$$u_{i,-1} = f(x).$$

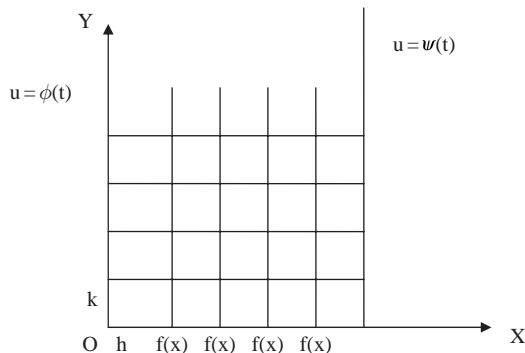
Thus equation (2) reduces to

$$u_{i,1} = f(x) + 2k \ g(x). \quad (3)$$

Further, the boundary conditions reduce to

$$u_{0,j} = \phi(t) \text{ and } u_{i,j} = \psi(t).$$

Thus, we have the figure



The entries in the first row of the solution are all $f(x)$. The entries in the second row are given by equation (3). The entries in the third row are given by formula (1) and so on. Thus, formula (1) is three level time formula.

Model Paper III

1. (a) If $V(\text{km/hr})$ and $R(\text{kg/tonne})$ are related by a relation of the type $R = a + bV^2$, find by the method of least squares a and b with the help of the following table:

$V:$	10	20	30	40	50
$R:$	8	10	15	21	30

- (b) The area A of a circle of diameter d is given for the following values:

$d:$	80	85	90	95	100
$A:$	5026	5674	6362	7088	7854

Calculate the area of a circle of diameter 105.

2. (a) Using Newton's iterative method, find a root of the equation $x^4 + x^3 - 7x^2 - x + 5 = 0$ correct to the four decimal places.
- (b) Using Muller's method, find a root of the equation $\log x = x - 3$, taking $x_0 = 0.25$, $x_1 = 0.5$ and $x_2 = 1$.
3. (a) Solve the following system of equations of using Gauss-Jordan method:

$$x_1 + 2x_2 + x_3 = 8; 2x_1 + 3x_2 + 4x_3 = 20; 4x_1 + 3x_2 + 2x_3 = 16.$$
- (b) Solve the equations by Jacobi's method:

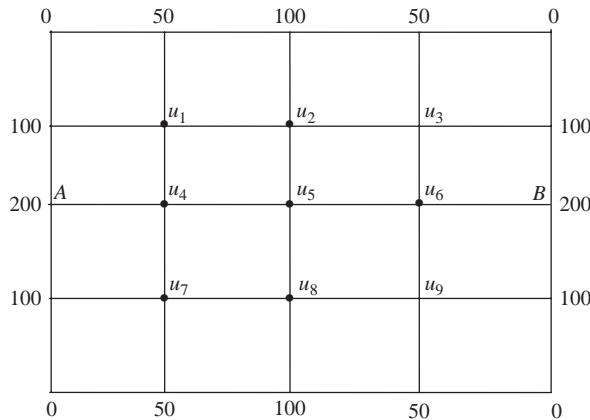
$$5x + 2y + z = 12; x + 4y + 2z = 15; x + 2y + 5z = 20$$
4. (a) Evaluate the first derivative at $x = -3$ and $x = 0$ from the following table:

$x:$	-3	-2	-1	0	1	2	3
$y:$	-33	-12	-3	0	3	12	33

- (b) Use Simpson's method with $n = 4$ to estimate:

$$\int_0^1 \frac{dx}{1+x^2}$$

5. Using Runge-Kutta method of fourth order solve for y at $x = 1.2, 1.4$ from the equation $\frac{dy}{dx} = \frac{2xy + e^x}{x^2 + x \cdot e^x}$ with $x_0 = 1, y_0 = 0$.
6. Apply Milne's method to find a solution of the differential equation $\frac{dy}{dx} = x - y^2$ in the range $0 \leq x \leq 1$ for the boundary condition $y = 0$ at $x = 0$.
7. Solve the equation $\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$, subject to the conditions $u(x, 0) = \sin \pi x$, $0 \leq x \leq 1$; $u(0, t) = u(1, t) = 0$ by using Schmidt method and compare this with the $\Delta \supset /^n$ obtained by using Crank-Nicolson formula. Carry out computation for two levels taking $h = \frac{1}{3}$, $k = \frac{1}{36}$.
8. Solve the elliptic equation $u_{xx} + u_{yy}$ for the following square mesh with boundary values as shown:



SOLUTIONS

9. (a) The normal equation for the curve fitting of the type $y = a + bx^2$ is

$$na + b(x_1^2 + x_2^2 + \dots + x_n^2) = y_1 + y_2 + \dots + y_n$$

$$a(x_1^2 + x_2^2 + \dots + x_n^2) + b(x_1^4 + x_2^4 + \dots + x_n^4) = x_1^2 y_1 + x_2^2 y_2 + \dots + x_n^2 y_n.$$

So we establish the following table:

n	x	x^2	x^4	y	x^2y
1	10	100	10000	8	800
1	20	400	160000	10	4000
1	30	900	810000	15	13500
1	40	1600	2560000	21	33600
1	50	2500	6250000	30	75000
5	150	5500	9790000	84	126900

The normal equations are

$$5a + 5500b = 84 \quad (1)$$

and

$$5500a + 9790000b = 126900,$$

that is,

$$5a + 5500b = 84$$

and

$$55a + 97900b = 1269,$$

or

$$55a + 60500b = 924 \quad (2)$$

and

$$55a + 97900b = 1269. \quad (3)$$

Subtracting equation (2) from (3), we get

$$37400b = 345 \text{ which yields } b = 0.00924.$$

Putting this value in equation (1), we get

$$5a + 50.82 = 84 \text{ which yields } a = 6.76.$$

Hence, $a = 6.76, b = 0.00924$ and the parabola of best fit is

$$R = 6.76 + 0.00924V^2.$$

1. (b) The difference table for the given data is

d	A				
80	5026	648			
85	5674	688	40	-12	
90	6362	716	28	22	32
95	7088	766	50		
100	7854				

Letting $x_p = 105$, $x_0 = 100$, and $p = \frac{105-100}{5} = 1$, we shall use Newton's backward difference formula

$$f_p = f_0 + p\bar{V}f_0 + \frac{p(p+1)}{2}\bar{V}^2f_0 + \frac{p(p+1)(p+2)}{3!}\bar{V}^3f_0 + \frac{p(p+1)(p+2)(p+3)}{4!}\bar{V}^4f_0.$$

Therefore,

$$f(105) = 7854 + 766 + 50 + 22 + 32 = 8724.$$

2. (a) We have

$$f(x) = x^4 + x^3 - 7x^2 - x + 5 = 0$$

and

$$f'(x) = 4x^3 + 3x^2 - 14x - 1.$$

We note that

$$f(1) = -1, f(2) = -1, \text{ and } f(3) = 47.$$

Therefore, one of the roots of $f(x) = 0$ lies between 2 and 3 and it is nearer to 2. So, let $x_0 = 2$. By Newton-Raphson method

$$\begin{aligned} x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^4 + x_n^3 - 7x_n^2 - x_n + 5}{4x_n^3 + 3x_n^2 - 14x_n - 1} \\ &= \frac{4x_n^4 + 3x_n^3 - 14x_n^2 - x_n - x_n^4 - x_n^3 + 7x_n^2 + x_n - 5}{4x_n^3 + 3x_n^2 - 14x_n - 1} \\ &= \frac{3x_n^4 + 2x_n^3 - 7x_n^2 - 5}{4x_n^3 + 3x_n^2 - 14x_n - 1}. \end{aligned}$$

Therefore,

$$\begin{aligned} x_1 &= \frac{3x_0^4 + 2x_0^3 - 7x_0^2 - 5}{4x_0^3 + 3x_0^2 - 14x_0 - 1} = \frac{3(16) + 2(8) - 7(4) - 5}{4(8) + 3(4) - 14(2) - 1} \\ &= \frac{31}{15} = 2.066, \end{aligned}$$

$$\begin{aligned} x_2 &= \frac{3x_1^4 + 2x_1^3 - 7x_1^2 - 5}{4x_1^3 + 3x_1^2 - 14x_1 - 1} = \frac{3(18.219) + 2(8.767) - 7(4.268) - 5}{4(8.767) + 3(4.268) - 14(2.066) - 1} \\ &= \frac{37.315}{17.948} = 2.079, \end{aligned}$$

$$\begin{aligned} x_3 &= \frac{3(18.68) + 2(8.986) - 7(4.322) - 5}{4(8.986) + 3(4.322) - 14(2.079) - 1} \\ &= \frac{38.758}{18.804} = 2.061. \end{aligned}$$

Proceed further to get the answer correct up to four decimal places.

2. (b) We are given that

$$f(x) = x - \log x - 3.$$

We take initial approximations as

$$x_{i-2} = 0.25, x_{i-1} = 0.5, \text{ and } x_i = 1.$$

Then,

$$y_{i-2} = 0.25 - \log 0.25 - 3 = 0.25 - (-0.602) - 3 = -2.148,$$

$$y_{i-1} = 0.5 - \log 0.5 - 3 = 0.5 - (-0.301) - 3 = -2.199,$$

$$y_i = 1 - \log 1 - 3 = -2.$$

Therefore,

$$\begin{aligned} A &= \frac{(x_{i-2} - x_i)(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i-2} - y_i)}{(x_{i-1} - x_{i-2})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{(0.25 - 1)(-2.199 + 2) - (0.5 - 1)(-2.148 + 2)}{(0.5 - 0.2)(0.5 - 1)(0.2 - 1)} \\ &= \frac{(-0.75)(-0.199) - (-0.5)(-0.148)}{(0.25)(-0.5)(-0.7)} \\ &= \frac{0.14925 - 0.074}{0.09375} = \frac{0.07525}{0.09375} = 0.8026. \end{aligned}$$

$$\begin{aligned} B &= \frac{(x_{i-2} - x_i)^2(y_{i-1} - y_i) - (x_{i-1} - x_i)^2(y_{i-2} - y_i)}{(x_{i-2} - x_{i-1})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{(-0.75)^2(-0.199) - (-0.5)^2(-0.148)}{(-0.2)(-0.5)(-0.7)} \\ &= \frac{(0.5625)(-0.199) - (0.25)(-0.148)}{-0.09375} \\ &= \frac{-0.11193 + 0.03700}{-0.09375} = \frac{-0.07493}{-0.09375} = 0.7992. \end{aligned}$$

Then

$$\begin{aligned} x_{i+1} &= x_i - \frac{2y_i}{B \pm \sqrt{B^2 - 4Ay_i}} = 1 - \frac{2(-2)}{0.7992 \pm \sqrt{0.6387 + 6.4208}} \\ &= 1 + \frac{4}{0.7992 + 2.657} = 2.1573. \end{aligned}$$

The process is repeated taking $x_{i-2} = 0.5$, $x_{i-1} = 1$, and $x_i = 2.1573$ and so on.

The approximate value of the root is 3.55.

3. (a) Example 3.4.

(b) From the given equations, we have

$$x = \frac{1}{5}[12 - 2y - z],$$

$$y = \frac{1}{4}[15 - x - 2z],$$

$$z = \frac{1}{5}[20 - x - 2y].$$

Starting with $(x_0, y_0, z_0) = (0, 0, 0)$, we have

$$x_1 = \frac{1}{3}(12 - 0 - 0) = \frac{12}{5} = 2.4,$$

$$y_1 = \frac{1}{4}[15 - 0 - 0] = \frac{15}{4} = 3.75,$$

$$z_1 = \frac{1}{5}[2 - 0 - 0] = \frac{20}{5} = 4.$$

The next iteration is

$$x_2 = \frac{1}{5}[12 - 2(3.75) - 4] = 0.1,$$

$$y_2 = \frac{1}{4}[15 - 2.40 - 8] = 1.15,$$

$$z_2 = \frac{1}{5}[20 - 2.4 - 2(3.75)] = 2.02.$$

Proceed and calculate up to x_9, y_9, z_9 . The answer is 1.09, 2.10, and 3.09. The exact roots are 1, 2, and 3.

4. (a) The difference table for the given problem is

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
-3	-33				
-2	-12	21			
-1	-3	9	-12	6	
0	0	3	-6	6	0
1	3	3	0	6	0
2	12	9	6	6	0
3	33	21	12		

We know that [see formula (7.15) or (7.21)]

$$f'(x) = \frac{1}{h} \left[\Delta f(x) - \frac{\Delta^2}{2} f(x) + \frac{\Delta^3}{3} f(x) - \frac{\Delta^4}{4} f(x) + \dots \right].$$

Therefore,

$$f'(-3) = \frac{1}{1} \left[21 - \frac{1}{2}(-12) + \frac{1}{3}(6) \right] = 29$$

and

$$f'(0) = \frac{1}{1} \left[3 - \frac{1}{2}(6) + \frac{1}{3}(6) \right] = 2.$$

4. (b) Example 8.10.

5. We have

$$\frac{dy}{dx} = f(x, y) = \frac{2xy + e^x}{x^2 + xe^x}, \quad y(1) = 0.$$

Thus, $x_0 = 1$, $y_0 = 0$ and we take $h = 0.2$. Then

$$k_1 = hf(x_0, y_0) = 0.2 \left(\frac{e^1}{1+e^1} \right) = 0.2 \left(\frac{2.71828}{3.71828} \right) = 0.1462,$$

$$\begin{aligned} k_2 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{k_1}{2}\right) = 0.2[f(1.1, 0.0731)] \\ &= 0.2 \left[\frac{0.161 + e^{1.1}}{1.21 + 1.1(e^{1.1})} \right] = 0.2 \left[\frac{3.1652}{4.5146} \right] = 0.1402, \end{aligned}$$

$$\begin{aligned} k_3 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{k_2}{2}\right) = hf(1.1, 0.0701) \\ &= 0.2 \left[\frac{0.15422 + 3.0042}{1.21 + 1.1(e^{1.1})} \right] = 0.2 \left[\frac{3.1584}{4.5146} \right] = 0.1399, \end{aligned}$$

$$\begin{aligned} k_4 &= hf(x_0 + h, y_0 + k_3) = hf(1.2, 0.1399) \\ &= 0.2 \left[\frac{0.3358 + 3.3201}{1.44 + 1.2(e^{1.2})} \right] = 0.2 \left[\frac{3.6559}{5.4241} \right] = 0.1348. \end{aligned}$$

Therefore,

$$\begin{aligned} y(1.2) &= y_0 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\ &= 0 + \frac{1}{6}[0.146 + 2(0.1402) + 2(0.1399) + 0.1348] = 0.1402. \end{aligned}$$

Now $x_0 = 1.2$, $y_0 = 0.1402$, and $h = 0.2$. Calculate as above k_1 , k_2 , k_3 , and k_4 , and then find

$$\begin{aligned} y(1.4) &= y_0 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\ &= 0.1402 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4). \end{aligned}$$

It will be approximately 0.264.

6. We have

$$y' = x - y^2, \quad y = 0 \text{ at } x = 0.$$

By Picard's method

$$y_1 = y_0 + \int_0^x f(x, y_0) dx = 0 + \int_0^x x dx = \frac{x^2}{2},$$

$$\begin{aligned}y_2 &= y_0 \int_0^x f(x, y_1) dx = 0 + \int_0^x \left(x - \frac{x^4}{4} \right) dx = \frac{x^2}{2} - \frac{x^5}{20}, \\y_3 &= y_0 \int_0^x f(x, y_2) dx = 0 + \int_0^x \left[x - \left(\frac{x^2}{2} - \frac{x^5}{20} \right)^2 \right] dx \\&= \frac{x^2}{2} - \frac{x^5}{20} + \frac{x^8}{160} - \frac{x^{11}}{4400}.\end{aligned}$$

Taking $h = 0.2$, we have

$$\begin{aligned}y_0 &= 0 \text{ which gives } y'_0 = 0 - 0^2 = 0, \\y_1 &= \frac{(0.2)^2}{2} = 0.02 \text{ which implies } y'_1 = 0.2 - (0.02)^2 = 0.1996, \\y_2 &= \frac{(0.4)^2}{2} - \frac{(0.4)^5}{20} = 0.0795, \text{ which yields} \\y'_2 &= 0.4 - (0.0795)^2 = 0.3937, \\y_3 &= \frac{(0.6)^2}{2} - \frac{(0.6)^5}{20} + \frac{(0.6)^8}{160} - \frac{(0.6)^{11}}{4400} = 0.1762, \text{ which yields} \\y'_3 &= 0.5689.\end{aligned}$$

Using the predictor, we have

$$\begin{aligned}y_4 &= y(0.8) = y_0 + \frac{4h}{3}[2y'_1 - y'_2 + 2y'_3] \\&= 0 + \frac{4(0.2)}{3}[2(0.1996) - 0.3937 + 2(0.1762)] \\&= 0.3049.\end{aligned}$$

Then

$$y'_4 = 0.8 - (0.3049)^2 = 0.7070.$$

Therefore using corrector, we have

$$\begin{aligned}y_4 &= y_2 + \frac{h}{3}[y'_2 + 4y'_3 + y'_4] \\&= 0.0795 + \frac{0.2}{3}[0.3937 + 4(0.1762) + 0.7070] = 0.3046.\end{aligned}$$

Now using predictor, we have

$$\begin{aligned}y_5 &= y(1) = y_1 + \frac{4h}{3}[2y'_2 - y'_3 + 2y'_4] \\&= 0.02 + \frac{4(0.2)}{3}[2(0.3937) - 0.5689 + 2(0.7070)] = 0.4554.\end{aligned}$$

Then

$$y'_5 = 1 - (0.4554)^2 = 0.7926.$$

Therefore, the corrector gives

$$y(1) = y_5 = y_3 + \frac{h}{3}(y'_3 + 4y'_4 + y'_5) = 0.4555.$$

7. Similar to Example 11.16.

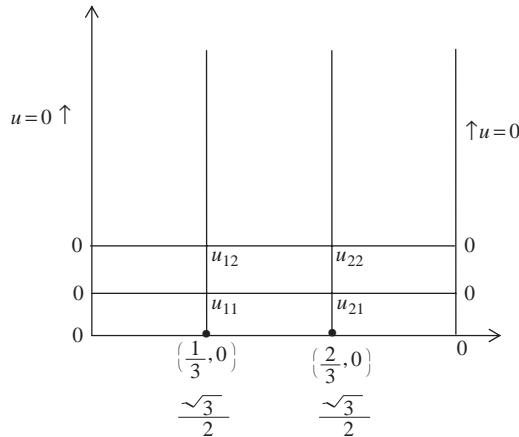
We have

$$\begin{aligned}\frac{\partial u}{\partial t} &= \frac{\partial^2 u}{\partial x^2}, \quad u(x, 0) = \sin \pi x, \quad 0 \leq x \leq 1 \\ u(0, t) &= u(1, t) = 0.\end{aligned}$$

Since $c^2 = 1$, taking $h = \frac{1}{3}$, $k = \frac{1}{36}$, we have $r = \frac{c^2 k}{h^2} = \frac{1}{4}$. Hence, the explicit formula

$$u_{i,j+1} = u_{i,j} + r[u_{i-1,j} - 2u_{i,j} + u_{i+1,j}] = \frac{1}{4}[u_{i-1,j} + 2u_{i,j} + u_{i+1,j}]$$

is valid. The corresponding grid is shown below:



$$u\left(\frac{1}{3}, 0\right) = \sin \frac{\pi}{3} = \frac{\sqrt{3}}{2}, \quad u\left(\frac{2}{3}, 0\right) = \sin \frac{2\pi}{3} = \frac{\sqrt{3}}{2},$$

$$u_{11} = \frac{1}{4}[u_{00} + 2u_{10} + u_{20}] = \frac{1}{4}\left[0.2\left(\frac{\sqrt{3}}{2}\right) + \frac{\sqrt{3}}{2}\right] = 0.65,$$

$$u_{21} = \frac{1}{4}[u_{10} + 2u_{20} + u_{30}] = \frac{1}{4}\left[\frac{\sqrt{3}}{2} + 2\left(\frac{\sqrt{3}}{2}\right) + 0\right] = 0.65,$$

$$u_{12} = \frac{1}{4}[u_{01} + 2u_{11} + u_{21}] = \frac{1}{4}[0 + 2(0.65) + 0.65] = 0.49,$$

$$u_{22} = \frac{1}{4}[u_{11} + 2u_{21} + u_{31}] = \frac{1}{4}[0.65 + 2(0.65) + 0] = 0.49.$$

The Crank–Nicolson formula for $r = \frac{1}{4}$ becomes

$$-\frac{1}{4}u_{i-1,j+1} + \frac{5}{2}u_{i,j+1} - \frac{1}{4}u_{i+1,j+1} = \frac{1}{4}u_{i-1,j} + \frac{3}{2}u_{i,j} + \frac{1}{4}u_{i+1,j}$$

or

$$-u_{i-1,j+1} + 10u_{i,j+1} - u_{i+1,j+1} = u_{i-1,j} + 6u_{i,j} + u_{i+1,j}.$$

Therefore, taking $i = 1, j = 0$, and $i = 2, j = 0$, we have

$$-u_{01} + 10u_{11} - u_{21} = u_{00} + 6u_{10} + u_{20}$$

and

$$-u_{11} + 10u_{21} - u_{31} = u_{10} + 6u_{20} + u_{30}$$

or

$$10u_{11} - u_{21} = \frac{6\sqrt{3}}{2} + \frac{\sqrt{3}}{2} = \frac{7\sqrt{3}}{2}$$

and

$$-u_{11} + 10u_{21} = \frac{\sqrt{3}}{2} + 6 \cdot \frac{\sqrt{3}}{2} = 7 \frac{\sqrt{3}}{2}.$$

Solving these equations, we get

$$u_{11} = u_{21} = 0.67.$$

Similarly taking $i = 1, 2$, and $j = 1$, we get

$$10u_{12} - u_{22} = 4.69$$

and

$$-u_{12} + 10u_{22} = 4.69$$

and so

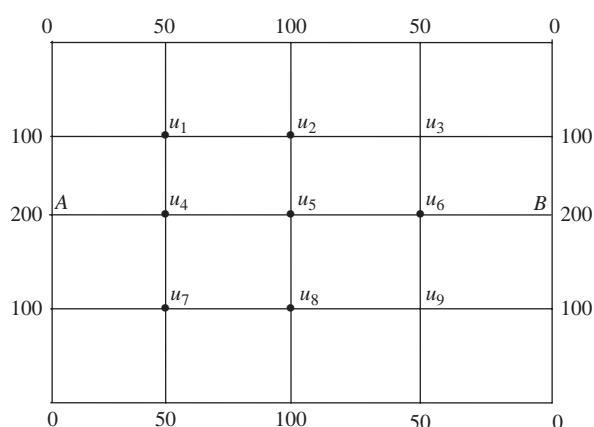
$$u_{12} = u_{22} = 0.52.$$

8. Similar to Example 11.5 (each boundary value has been divided by 10).

The figure is symmetrical about AB and CD. Therefore,

$$u_1 = u_3 = u_7 = u_9 \text{ and } u_2 = u_8, u_4 = u_6.$$

Hence, it sufficient to find u_1, u_2, u_4, u_5 .



We have

$$u_5 = \frac{1}{4}[200 + 200 + 100 + 100] = 150 \text{ (standard five point formula),}$$

$$u_1 = \frac{1}{4}[0 + 100 + 200 + 150] = 112.50 \text{ (diagonal five point formula),}$$

$$u_2 = \frac{1}{4}[112.5 + 112.5 + 100 + 150] = 118.75 \text{ (standard five point formula),}$$

$$u_4 = \frac{1}{4}[200 + 150 + 112.50 + 112.50] = 143.75 \text{ (standard five point formula).}$$

We now use Gauss–Seidal's method to improve these values

$$u_1^{(1)} = \frac{1}{4}[100 + 118.75 + 50 + 143.75] = 103.125,$$

$$u_2^{(1)} = \frac{1}{4}[103.125 + 100 + 150 + 103.125] = 114.0625,$$

$$u_4^{(1)} = \frac{1}{4}[200 + 150 + 103.125 + 103.125] = 139.0625,$$

$$u_5^{(1)} = \frac{1}{4}[139.0625 + 139.0625 + 114.0625 + 114.0625] = 126.5625.$$

After nine iterations, we have

$$u_1 = u_3 = u_7 = u_9 = 93.805,$$

$$u_2 = u_8 = 100.055,$$

$$u_4 = u_6 = 125.055,$$

$$u_5 = 112.555.$$

Model Paper IV

1. (a) State Lagrange's Interpolation formula. Use this formula to find the value of y , when $x = 5$, if the following values of x and y are given:

x	1	2	3	4	7
y	2	4	8	16	128

- (b) Develop cubic splines for the data given below and predict $f(1.5)$:

x	0	1	2	3
$f(x)$	1	-1	-1	0

2. (a) Using Muller's method find root of the equation $\log x = x - 3$, taking $x_0 = 0.25$, $x_1 = 0.5$ and $x_2 = 1.0$
- (b) Using Newton-Raphson method find a root of the equation correct to 3 decimal places. $x \sin x + \cos x = 0$, which is near $x = \pi$.

3. (a) Solve the following equations by Gauss-Jordan method
 $x_1 + 2x_2 + x_3 = 8; 2x_1 + 3x_2 + 4x_3 = 20; 4x_1 + 3x_2 + 2x_3 = 16$

- (b) Solve the following equations by Relaxation method
 $10x - 2y - 2z = 6; -x + 10y - 2z = 7; -x - y + 10z = 8.$

4. (a) From the following table find $f'(1.4)$

x	1.2	1.3	1.4	1.5	1.6
$f(x)$	1.5095	1.6984	1.9043	2.1293	2.3756

- (b) Find the approximate value of $\log_e 5$ by calculating to 4 decimal places, by Simpson's $\frac{1}{3}$ Rule,
 $\int_0^5 \frac{dx}{4x+5}$, dividing the range into 10 equal parts

Section-B

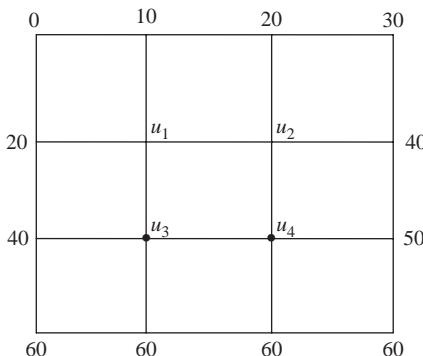
5. Solve the initial value problem $\frac{dy}{dx} = 1 + xy^2, y(0) = 1$, for $x = 0.4$ by Milne's method, it is given that

x	0.1	0.2	0.3
y	1.105	1.223	1.355

6. Using Runge Kutta method of order 4, find $y(0.2)$ given that

$$\frac{dy}{dx} = 3x + \frac{1}{2}y, y(0) = 1, \text{ taking } h = 0.1$$

7. Solve $u_{xx} + u_{yy} = 0$ for the following square meshes with the boundary values shown:



8. Solve $\frac{\partial u}{\partial t} = 5 \frac{\partial^2 u}{\partial x^2}$ with $u(0, t) = 0, u(5, t) = 60$

$$\text{and } u(x, 0) = \begin{cases} 20x & \text{for } 0 < x \leq 3 \\ 60 & \text{for } 3 < x \leq 5 \end{cases}$$

for five time steps having $h = 1$, by Schmidt method.

SOLUTIONS

1. (a) Let $y_i = f(x_i)$ be the value of a function at $x_i, 0 \leq i \leq n$. Then Lagrange's interpolating polynomial $P_n(x)$ is given by

$$P_n(x) = \sum_{i=0}^n L_i(x) f(x_i),$$

where

$$L_i(x) = \frac{(x - x_0)(x - x_1) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n)}{(x_i - x_0)(x_i - x_1) \cdots (x_i - x_{i-1})(x_i - x_{i+1}) \cdots (x_i - x_n)}.$$

In the given problem, $x = 5$ and we have

$$\begin{aligned} L_0(x) &= \frac{(x - x_1)(x - x_2)(x - x_3)(x - x_4)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)(x_0 - x_4)} = \frac{(5-2)(5-3)(5-4)(5-7)}{(1-2)(1-3)(1-4)(1-7)} \\ &= -\frac{1}{3}, \\ L_1(x) &= \frac{(x - x_0)(x - x_2)(x - x_3)(x - x_4)}{(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)(x_1 - x_4)} = \frac{(5-1)(5-3)(5-4)(5-7)}{(2-1)(2-3)(2-4)(2-7)} \\ &= \frac{8}{5}, \\ L_2(x) &= \frac{(x - x_0)(x - x_1)(x - x_3)(x - x_4)}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)(x_2 - x_4)} = \frac{(5-1)(5-2)(5-4)(5-7)}{(3-1)(3-2)(3-4)(3-7)} \\ &= -3, \\ L_3(x) &= \frac{(x - x_0)(x - x_1)(x - x_2)(x - x_4)}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)(x_3 - x_4)} = \frac{(5-1)(5-3)(5-4)(5-7)}{(4-1)(4-2)(4-3)(4-7)} \\ &= \frac{8}{3}, \\ L_4(x) &= \frac{(x - x_0)(x - x_1)(x - x_2)(x - x_3)}{(x_4 - x_0)(x_4 - x_1)(x_4 - x_2)(x_4 - x_3)} = \frac{(5-1)(5-2)(5-3)(5-4)}{(7-1)(7-2)(7-3)(7-4)} \\ &= \frac{1}{15}. \end{aligned}$$

We note that $\sum_{i=0}^n L_i(x) = 1$. Hence, our calculations are correct up to this stage. By Lagrange's formula

$$\begin{aligned} P(x) &= P(5) = \sum_{i=0}^n L_i(x) f(x_i) \\ &= -\frac{1}{3}(2) + \frac{8}{5}(4) + (-3)(8) + \frac{8}{3}(16) + \frac{1}{15}(128) \\ &= -\frac{2}{3} + \frac{32}{5} - 24 + \frac{128}{3} + \frac{128}{15} \approx 32.933. \end{aligned}$$

(b) The given tabular values are

x	0	1	2	3
y	1	-1	-1	0

We note that the arguments are equispaced with $h = 1$ and $n = 3$. The splines will be developed by the formula

$$s_i(x) = \frac{m_i}{6h_i}(x_{i+1} - x)^3 + \frac{m_{i+1}}{6h_i}(x - x_i)^3 + \left(\frac{y_i}{h_i} - \frac{m_i h_i}{6} \right)(x_{i+1} - x) + \left(\frac{y_{i+1}}{h_i} - \frac{m_{i+1} h_i}{6} \right)(x - x_i). \quad (1)$$

For equal spacing $h_i = 1$, we have

$$m_{i-1} + 4m_i + m_{i+1} = 6(y_{i+1} - 2y_i + y_{i-1}).$$

Putting $i = 1$ and 2 and we get

$$m_0 + 4m_1 + m_2 = 6(y_2 - 2y_1 + y_0) = 6(-1 - 2(-1) + 1) = 12$$

and

$$m_1 + 4m_2 + m_3 = 6(y_3 - 2y_2 + y_1) = 6[0 - 2(-1) - 1] = 6.$$

But for $n = 3$, the natural cubic spline requires $m_0 = m_3 = 0$. Therefore,

$$4m_1 + m_2 = 12 \text{ and } m_1 + 4m_2 = 6.$$

Solving these equations, we get $m_1 = \frac{14}{5}$ and $m_2 = \frac{4}{5}$. Hence (1) yields the cubic splines as

$$\begin{aligned} s_0(x) &= \frac{14}{6(5)}(x - 0)^3 + \left(\frac{1}{1} - 0 \right)(1 - x) + \left(\frac{-1}{1} - \frac{14}{5(6)} \right)(x - 0) \\ &= \frac{14}{30}x^3 + (1 - x) - \frac{44}{30}x = \frac{1}{30}[14x^3 - 74x + 30], \\ s_1(x) &= \frac{14}{6(5)}(2 - x)^3 + \frac{4}{6(5)}(x - 1)^3 + \left(\frac{-1}{1} - \frac{14}{6(5)} \right)(2 - x) + \left(\frac{-1}{1} - \frac{4}{5(6)} \right)(x - 1) \\ &= \frac{14}{30}(2 - x)^3 + \frac{4}{30}(x - 1)^3 - \frac{44}{30}(2 - x) - \frac{34}{30}(x - 1), \\ s_2(x) &= \frac{14}{6(5)}(3 - x)^3 + \left(\frac{-1}{1} - \frac{14}{6(5)} \right)(3 - x) + (0 - 0)(x - x_2) \\ &= \frac{14}{30}(3 - x)^3 - \frac{44}{30}(3 - x) = \frac{1}{30}[14(3 - x)^3 - 44(3 - x)]. \end{aligned}$$

2. (a) We are given that

$$f(x) = x - \log x - 3.$$

We take initial approximations as

$$x_{i-2} = 0.25, x_{i-1} = 0.5 \text{ and } x_i = 1.$$

Then

$$y_{i-2} = 0.25 - \log 0.25 - 3 = 0.25 - (-0.602) - 3 = -2.148,$$

$$y_{i-1} = 0.5 - \log 0.5 - 3 = 0.5 - (-0.301) - 3 = -2.199,$$

$$y_i = 1 - \log 1 - 3 = -2.$$

Therefore,

$$\begin{aligned} A &= \frac{(x_{i-2} - x_i)(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i-2} - y_i)}{(x_{i-1} - x_{i-2})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\ &= \frac{(0.25 - 1)(-2.199 + 2) - (0.5 - 1)(-2.148 + 2)}{(0.5 - 0.25)(0.5 - 1)(0.25 - 1)} \end{aligned}$$

$$\begin{aligned}
&= \frac{(-0.75)(-0.199) - (-0.5)(-0.148)}{(0.25)(-0.5)(-0.75)} \\
&= \frac{0.14925 - 0.074}{0.09375} = \frac{0.07525}{0.09375} = 0.8026. \\
B &= \frac{(x_{i-2} - x_i)^2(y_{i-1} - y_i) - (x_{i-1} - x_i)^2(y_{i-2} - y_i)}{(x_{i-2} - x_{i-1})(x_{i-1} - x_i)(x_{i-2} - x_i)} \\
&= \frac{(-0.75)^2(-0.199) - (-0.5)^2(-0.148)}{(-0.25)(-0.5)(-0.75)} \\
&= \frac{(0.5625)(-0.199) - (0.25)(-0.148)}{-0.09375} \\
&= \frac{-0.11193 + 0.03700}{-0.09375} = \frac{-0.07493}{-0.09375} = 0.7992.
\end{aligned}$$

Then

$$\begin{aligned}
x_{i+1} &= x_i - \frac{2y_i}{B \pm \sqrt{B^2 - 4Ay_i}} = 1 - \frac{2(-2)}{0.7992 \pm \sqrt{0.6387 + 6.4208}}. \\
&= 1 + \frac{4}{0.7992 + 2.657} = 2.1573.
\end{aligned}$$

The process is repeated taking $x_{i-2} = 0.5$, $x_{i-1} = 1$, and $x_i = 2.1573$ and so on.

The approximate value of the root is 3.55.

(b) We have

$$f(x) = x \sin x + \cos x = 0.$$

Therefore,

$$f'(x) = x \cos x + \sin x - \sin x = x \cos x.$$

Since the root is nearer to π , we take $x_0 = \pi$. By Newton-Raphson method

$$\begin{aligned}
x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n \sin x_n + \cos x_n}{x_n \cos x_n} \\
&= \frac{x_n^2 \cos x_n - x_n \sin x_n - \cos x_n}{x_n \cos x_n}.
\end{aligned}$$

Thus,

$$\begin{aligned}
x_1 &= \frac{x_0^2 \cos x_0 - x_0 \sin x_0 - \cos x_0}{x_0 \cos x_0} \\
&= \frac{\pi^2 \cos \pi - \pi \sin \pi - \cos \pi}{\pi \cos \pi} = \frac{1 - \pi^2}{\pi} = \frac{1 - 9.87755}{-3.142857} \\
&= 2.824,
\end{aligned}$$

$$x_2 = \frac{x_1^2 \cos x_1 - x_1 \sin x_1 - \cos x_1}{x_1 \cos x_1}$$

$$\begin{aligned}
&= \frac{(7.975)(-0.95) - (2.824)(0.3123) + (0.95)}{(2.824)(-0.95)} \\
&= \frac{-7.576 - 0.8819 + 0.95}{-2.6828} = \frac{7.5179}{2.6828} = 2.8022, \\
x_3 &= \frac{7.8512(-0.9429) - (2.8022)(0.3329) + 0.9429}{(2.8022)(-0.9429)} \\
&= \frac{-7.4029 - 0.93285 + 0.9429}{-2.6422} = \frac{7.39285}{2.6422} = 2.797.
\end{aligned}$$

Calculate x_4 and x_5 similarly.

3. (a) Example 3.5.
(b) Example 3.21.
4. (a) Using centered formula of order (h^4) , we have

$$\begin{aligned}
f'(1.4) &\approx \frac{-f(x+2h) + 8f(x+h) - 8f(x-h) + f(x-2h)}{12h} \\
&= \frac{-2.3756 + 8(2.1293) - 8(1.6984) + 1.5095}{12(0.1)} \\
&= \frac{-2.3756 + 17.0344 - 13.5872 + 1.5095}{1.2} = 2.1509.
\end{aligned}$$

- (b) The values of the integrand for $h = \frac{1}{2}$ are

x	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{5}{2}$	3	$\frac{7}{2}$	4	$\frac{9}{2}$	5
$f(0)$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{9}$	$\frac{1}{11}$	$\frac{1}{13}$	$\frac{1}{15}$	$\frac{1}{17}$	$\frac{1}{19}$	$\frac{1}{21}$	$\frac{1}{23}$	$\frac{1}{25}$

Therefore, by Simpson's $\frac{1}{3}$ rule,

$$\begin{aligned}
\int_0^5 \frac{dx}{4x+5} &= \frac{h}{3}[f_0 + 4(f_1 + f_3 + f_5 + f_7 + f_9) + 2(f_2 + f_4 + f_6 + f_8) + f_{10}] \\
&= \frac{1}{6} \left[\left(\frac{1}{5} + \frac{1}{25} \right) + 4 \left(\frac{1}{7} + \frac{1}{11} + \frac{1}{15} + \frac{1}{19} + \frac{1}{23} \right) + 2 \left(\frac{1}{9} + \frac{1}{13} + \frac{1}{17} + \frac{1}{21} \right) \right] \\
&= \frac{1}{6} \left[\frac{6}{25} + 4(0.142857 + 0.09090 + 0.066666 + 0.05263 + 0.04348) \right. \\
&\quad \left. + 2(0.11111 + 0.07692 + 0.05882 + 0.04761) \right] \\
&= \frac{1}{6} \left[\frac{6}{25} + 4(0.142857 + 0.09090 + 0.066666 + 0.05263 + 0.04348) \right. \\
&\quad \left. + 2(0.11111 + 0.07692 + 0.05882 + 0.04761) \right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{6} \left[\frac{6}{25} + 4(0.142857 + 0.09090 + 0.066666 + 0.05263 + 0.04348) \right. \\
&\quad \left. + 2(0.11111 + 0.07692 + 0.05882 + 0.04761) \right] \\
&= \frac{1}{6} [0.24 + 1.58613 + 0.58892] = 0.4025.
\end{aligned}$$

Also

$$\begin{aligned}
\int_0^5 \frac{dx}{4x+5} &= \frac{1}{4} [\log(4x+5)]_0^5 = \frac{1}{4} [\log 25 - \log 5] \\
&= \frac{1}{4} \left[\log \frac{25}{5} \right] = \frac{1}{4} \log_e 5.
\end{aligned}$$

Hence,

$$\log_e 5 = 4 \int_0^5 \frac{dx}{4x+5} = 4(0.4025) = 1.61.$$

The actual value is 1.6094.

5. The given initial value problem is

$$\begin{aligned}
\frac{dy}{dx} &= 1 + xy^2, \quad y(0) = 1, \\
y(0.1) &= 1.105, \quad y(0.2) = 1.223, \text{ and } y(0.3) = 1.355.
\end{aligned}$$

We have

$$y' = 1 + xy^2.$$

Therefore,

$$y'_1 = 1 + (0.1)(1.105)^2 = 1.1221,$$

$$y'_2 = 1 + (0.2)(1.223)^2 = 1.2991,$$

$$y'_3 = 1 + (0.3)(1.355)^2 = 1.5508.$$

Using predictor, we have

$$\begin{aligned}
y_4 &= y_0 + \frac{4h}{3} [2y'_1 - y'_2 + 2y'_3] \\
&= 1 + \frac{4(0.1)}{3} [2(1.1221) - 1.2991 + 2(1.5508)] \\
&= 1 + \frac{0.4}{3} (2.2442 - 1.2991 + 3.1016) = 1.53956.
\end{aligned}$$

Now,

$$y'_4 = 1 + (0.4)(1.53956)^2 = 1.9481.$$

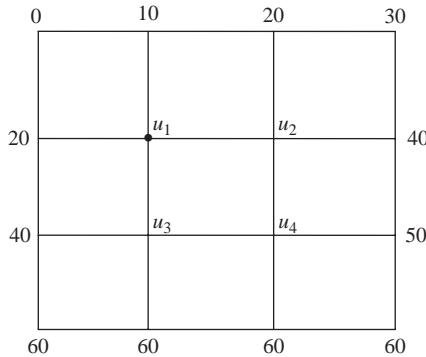
Then the corrector formula yields

$$y_4 = y(0.4) = y_2 + \frac{h}{3} [y'_2 + 4y'_3 + y'_4]$$

$$\begin{aligned}
 &= 1.223 + \frac{0.1}{3}[1.2991 + 4(1.5508) + 1.9481] \\
 &= 1.223 + \frac{0.1}{3}[9.4504] = 1.5380.
 \end{aligned}$$

6. Example 10.14.

7. (This question is nothing but Example 11.3. The figure has been rotated about the x -axis so that $u_1 \leftrightarrow u_3$ and $u_2 \leftrightarrow u_4$.) The given square meshes are shown below.



We assume that $u_4 = 0$. Then the initial approximation is

$$u_1 = \frac{1}{4}[40 + 20 + 0 + 0] = 15 \text{ (diagonal five point formula)}$$

$$u_2 = \frac{1}{4}[15 + 40 + 20 + 0] = \frac{75}{4} = 18.75 \text{ (standard five point formula)}$$

$$u_3 = \frac{1}{4}[15 + 60 + 40 + 0] = \frac{115}{4} = 28.75 \text{ (standard five point formula)}$$

$$u_4 = \frac{1}{4}[50 + 60 + 18.75 + 28.75] = 39.375 \text{ (standard five point formula).}$$

Now using Gauss-Seidel's method, we have

$$u_1^{(1)} = \frac{1}{4}[20 + 10 + u_2 + u_3] = \frac{1}{4}[20 + 10 + 18.75 + 28.75] = 19.375$$

$$u_2^{(1)} = \frac{1}{4}[40 + 20 + u_1^{(1)} + u_4] = \frac{1}{4}[60 + 19.375 + 39.375] = 29.6875$$

$$u_3^{(1)} = \frac{1}{4}[40 + 60 + u_1^{(1)} + u_4] = \frac{1}{4}[100 + 19.375 + 39.375] = 39.6875$$

$$u_4^{(1)} = \frac{1}{4}[50 + 60 + u_2^{(1)} + u_3^{(1)}] = \frac{1}{4}[110 + 29.6875 + 39.6875] = 44.84375$$

$$u_1^{(2)} = \frac{1}{4}[30 + u_2^{(1)} + u_3^{(1)}] = \frac{1}{4}[30 + 29.6875 + 39.6875] = 24.843$$

$$u_2^{(2)} = \frac{1}{4}[60 + u_1^{(1)} + u_4^{(1)}] = \frac{1}{4}[60 + 24.843 + 44.84375] = 32.4219$$

$$u_3^{(2)} = \frac{1}{4}[100 + u_1^{(2)} + u_4^{(1)}] = \frac{1}{4}[100 + 24.843 + 44.84375] = 42.4217$$

$$u_4^{(2)} = \frac{1}{4}[110 + u_2^{(2)} + u_3^{(2)}] = \frac{1}{4}[110 + 32.4219 + 42.4217] = 46.2109$$

$$u_1^{(3)} = \frac{1}{4}[30 + u_2^{(2)} + u_3^{(2)}] = \frac{1}{4}[30 + 32.4219 + 42.4217] = 26.2110$$

$$u_2^{(3)} = \frac{1}{4}[60 + u_1^{(3)} + u_4^{(2)}] = \frac{1}{4}[60 + 26.2110 + 46.2109] = 33.1055$$

$$u_3^{(3)} = \frac{1}{4}[100 + u_1^{(3)} + u_4^{(2)}] = \frac{1}{4}[100 + 26.2110 + 46.2109] = 43.1055$$

$$u_4^{(3)} = \frac{1}{4}[110 + u_2^{(3)} + u_3^{(3)}] = \frac{1}{4}[110 + 33.1055 + 43.1055] = 46.5527.$$

Further applications of Gauss–Seidel method yields

$$u_1 = 26.65, \quad u_2 = 33.32, \quad u_3 = 43.31, \quad u_4 = 46.65.$$

8. We have

$$\frac{\partial u}{\partial t} = 5 \frac{\partial^2 u}{\partial x^2}, \quad u(0, t) = 0, \quad u(5, t) = 60,$$

$$u(x, 0) = \begin{cases} 20x & \text{for } 0 < x \leq 3 \\ 60 & \text{for } 3 < x \leq 5. \end{cases}$$

Here $c^2 = 5$, $h = 1$. We choose $k = \frac{1}{10}$ so that $r = \frac{c^2 k}{h} = \frac{5}{10} = \frac{1}{2}$.

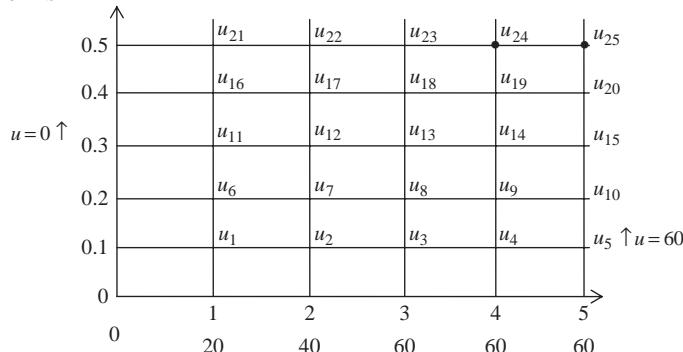
Therefore Bender–Schmidt method is applicable and we have

$$u_{i,j+1} = \frac{u_{i-1,j} - u_{i+1,j}}{2}. \quad (1)$$

We have

$$\begin{aligned} u(1, 0) &= 20(1) = 20, \quad u(2, 0) = 20(2) = 40, \quad u_3(3, 0) = 20(3) = 60 \\ u(4, 0) &= 60, \quad u(5, 0) = 60. \end{aligned}$$

The grid for the solution is



Using equation (1), the values of u at the grid points $(1,0.1)$, $(2,0.1)$, $(3,0.1)$, $(4,0.1)$, $(5,0.1)$ are, respectively,

$$u_1 = \frac{0+40}{2} = 20,$$

$$u_2 = \frac{60+20}{2} = 40,$$

$$u_3 = \frac{60+40}{2} = 50,$$

$$u_4 = \frac{60+60}{2} = 60,$$

$$u_5 = 60.$$

Further, the second row of the solution is

$$u_6 = \frac{0+u_2}{2} = \frac{0+40}{2} = 20,$$

$$u_7 = \frac{u_1+u_3}{2} = \frac{20+50}{2} = 35,$$

$$u_8 = \frac{u_2+u_4}{2} = \frac{40+60}{2} = 50,$$

$$u_9 = \frac{u_3+u_5}{2} = \frac{50+60}{2} = 55,$$

$$u_{10} = 60.$$

For the third row, we have

$$u_{11} = \frac{0+u_7}{2} = \frac{0+35}{2} = 17.5,$$

$$u_{12} = \frac{u_6+u_8}{2} = \frac{20+50}{2} = 35,$$

$$u_{13} = \frac{u_7+u_9}{2} = \frac{35+55}{2} = 45,$$

$$u_{14} = \frac{u_8+u_{10}}{2} = \frac{50+60}{2} = 55,$$

$$u_{15} = 60.$$

Proceeding in the same fashion, we get the following table:

$j \backslash i$	0	1	2	3	4	5
0	0	20	40	60	60	60
1	0	20	40	50	60	60
2	0	20	35	50	55	60
3	0	17.5	35	45	55	60
4	0	17.5	31.25	45	52.5	60
5	0	15.625	31.25	41.875	52.5	60

Model Paper V

1. (a) Define the term absolute error. Given that

$$\begin{aligned}a &= 10.00 \pm 0.05 \\b &= 0.0356 \pm 0.0002 \\c &= 15300 \pm 100 \\d &= 62000 \pm 500\end{aligned}$$

Find the maximum value of the absolute error in

(i) $a + b + c + d$, and (ii) c^3 .

- (b) Find the number of terms of the exponential series such that their sum gives the values of e^x correct to five decimal places for all values of x in the range $0 \leq x \leq 1$.
 2. (a) By the method of least squares, find the straight line that best fits the following data:

x:	1	2	3	4	5
y:	14	27	40	55	68

- (b) Following values of x and y are given

x:	1	2	3	4	
y:	1	2	5	1	1

Find the cubic splines and evaluate $y(1.5)$ and $y'(3)$

3. (a) Find the first and second derivatives of $f(x)$ at $x = 1.5$, if:

x:	1.5	2.0	2.5	3.0	3.5	4.0
$f(x)$:	3.375	7.000	13.625	24.000	38.875	59.000

- (b) Evaluate $\int_0^1 \frac{dx}{1+x^2}$ by using:

- (i) Trapezoidal rule.
 (ii) Simpson's $\frac{1}{3}$ re rule.
 (iii) Simpson's $\frac{3}{8}$ th rule

and compare the results with its actual value.

4. (a) Find, by Newton's method, the real root of the equation $3x = \cos x + 1$.
 (b) Find the root of the equation

$$xe^x = \cos x,$$

Using the secant method correct to four decimal places.

5. (a) Solve the following equations by, Gauss elimination method
 $2x + y + z = 10$; $3x + 2y + 3z = 18$; $x + 4y + 9z = 16$.

- (b) Apply Gauss-Seidal iteration method to solve the following equations
 $20x + y - 2z = 17$; $3x + 20y - z = -18$; $2x - 3y + 20z = 25$.

6. (a) Using Jacobi's method, find all the eigenvalues and eigenvectors of the matrix.

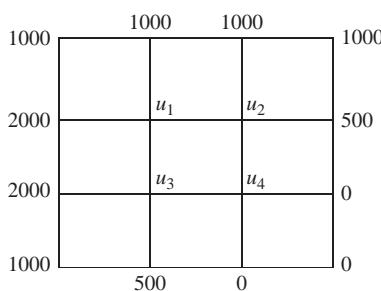
$$\begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & 2 \\ 3 & 2 & 3 \end{bmatrix}$$

(b) Using Given's method, reduce the following matrix to the tri-diagonal form.

$$\begin{bmatrix} 2 & 3 & 1 \\ 3 & 2 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

7. Using Runge-Kutta method of order 4, find y for $x = 0.1, 0.2, 0.3$, given that $\frac{dy}{dx} = xy + y^2$, $y(0) = 1$. Continue the solution at: $x = 0.4$, using Milne's method.

8. Solve the Laplace's equation $\nabla^2 u = 0$ for the following square mesh with boundary values as shown below:



SOLUTIONS

1. (a) If x is the true value of a quantity and x_0 is the approximate value, then $|x - x_0|$ is called the absolute error. We are given that

$$a = 10.00 \pm 0.05,$$

$$b = 0.0356 \pm 0.000$$

$$c = 15300 \pm 100,$$

$$d = 62000 \pm 500.$$

If a_1, b_1, c_1 , and d_1 are true values of a, b, c , and d , respectively. Then

$$\begin{aligned} & |(a_1 + b_1 + c_1 + d_1) - (a + b + c + d)| \\ &= |(a_1 - a) + (b_1 - b) + (c_1 - c) + (d_1 - d)| \\ &\leq |a_1 - a| + |b_1 - b| + |c_1 - c| + |d_1 - d| \\ &= |0.05| + |0.0002| + |100| + |500| \\ &= 600.0502, \end{aligned}$$

which is the required maximum value of the absolute error in $a + b + c + d$.

Further, if ϵ is the error in c , then

$$\begin{aligned} & |(c + \epsilon)^3 - c^3| = |\epsilon^3 + 3c\epsilon^2 + 3c^2\epsilon| \\ &\leq |(100)^3| + |3(15300)(100)^2| + |3(15300)^2(100)| \\ &= 10^6 + 459(10^4) + 3(153)^2(10^6) \\ &= 10^6 + 459(10^4) + 70227(10^6) \\ &= 10^{10}(0.0001 + 0.000459 + 70227) \\ &= 10^{10}(7.023259), \end{aligned}$$

which is the required maximum absolute error.

(b) The remainder term in the expansion of e^x is

$$R_n(x) = \frac{x^n}{n!} e^\xi, \quad 0 < \xi < x.$$

Therefore, the maximum absolute error is

$$e_{\max} = \left| \frac{x^n}{n!} \right| = \frac{1}{n!} \text{ at } x = 1.$$

Maximum relative error is

$$(e_r)_{\max} = \frac{\frac{x^n e^x}{n!}}{e^x} = \frac{x^n}{n!} = \frac{1}{n!} \text{ at } x = 1.$$

For a five decimal accuracy at $x = 1$, we have

$$\frac{1}{n!} < \frac{1}{2} 10^{-5},$$

which yields $n = 9$. Therefore, the number of terms in the exponential series should be nine.

2. (a) Exercise 3, Chapter 6.

The sum table for the given problem is

N	x	x^2	y	xy
1	1	1	14	14
1	2	4	27	54
1	3	9	40	120
1	4	16	55	220
1	5	25	68	340
5	15	55	204	748

Let the least square line be $y = a + bx$. Therefore, the normal equations are

$$5a + 15b = 204,$$

$$15a + 55b = 748.$$

Solving these equations, we get $a = 0$, $b = 13.6$. Hence, the least square line is $y = 13.6x$.

(b) Exercise 27, Chapter 5.

We note that the arguments are equispaced with $h = 1$ and $n = 3$. The spline will be formed by the formula:

$$s_i(x) = \frac{m_i}{h_i}(x_{i+1} - x)^3 + \frac{m_{i+1}}{6h_i}(x - x_i)^3 + \left(\frac{y_i}{h_i} - \frac{m_i h_i}{6} \right)(x_{i+1} - x) + \left(\frac{y_{i+1}}{h_i} - \frac{m_{i+1} h_i}{6} \right)(x - x_i)$$

For equal spacing $h_i = 1$, we have

$$m_{i-1} + 4m_i + m_{i+1} = 6(y_{i+1} - 2y_i + y_{i-1}).$$

Putting $i = 1$ and $i = 2$, we get

$$m_0 + 4m_1 + m_2 = 6(y_2 - 2y_1 + y_0) = 6[5 - 2(2) + 1] = 12$$

and

$$m_1 + 4m_2 + m_3 = 6(y_3 - 2y_2 + y_1) = 6[11 - 2(5) + 2] = 18.$$

But for $n = 3$, the natural spline requires $m_0 = m_3 = 0$. Therefore,

$$4m_1 + m_2 = 12 \quad \text{and} \quad m_1 + 4m_2 = 18.$$

Solving these equations, we have $m_1 = 2$, $m_2 = 4$. Hence the cubic splines are

$$\begin{aligned}s_0(x) &= \frac{2}{6}(x-1)^3 + \left(\frac{1}{1}-0\right)(2-x) + \left(\frac{2}{1}-\frac{2}{3}\right)(x-1) \\&= \frac{1}{3}(x-1)^3 + (2-x) + \frac{5}{3}(x-1) \\&= \frac{1}{3}(x^3 - 3x^2 + 5x),\end{aligned}$$

$$\begin{aligned}s_1(x) &= \frac{2}{6}(3-x)^3 + \frac{4}{6}(x-2)^3 + \left(\frac{2}{1}-\frac{2}{6}\right)(3-x) + \left(\frac{5}{1}-\frac{4}{6}\right)(x-2) \\&= \frac{1}{3}(27+9x^2-27x-x^3) + \frac{2}{3}(x^3-6x^2+12x-8) + \frac{5}{3}(3-x) + \frac{13}{3}(x-2) \\&= \frac{1}{3}(x^3 - 3x^2 + 5x),\end{aligned}$$

$$\begin{aligned}s_2(x) &= \frac{4}{6}(4-x)^3 + \left(\frac{5}{1}-\frac{4}{6}\right)(4-x) + \frac{11}{1}(x-3) \\&= \frac{2}{3}(64+12x^2-48x-x^3) + \frac{13}{3}(4-x) + 11x - 33 \\&= \frac{1}{3}(-2x^3 + 24x^2 - 76x + 81)\end{aligned}$$

Now

$$\begin{aligned}y(1.5) &= s_0(1.5) = \frac{1}{3}[(1.5)^3 - 3(1.5)^2 + 5(1.5)] \\&= \frac{1}{3}[3.375 - 6.750 + 7.50] = 1.375,\end{aligned}$$

$$y'(3) = s'_1(3) = \left[\frac{1}{3}(3x^2 - 6x + 5) \right]_{x=3} = \frac{14}{3}.$$

Also,

$$s'_2(3) = \left[\frac{1}{3}(-6x^2 + 48x - 76) \right]_{x=3} = \frac{14}{3}.$$

Hence the spline is smooth.

3. (a) The difference table for the given problem is

x	$f(x)$					
1.5	3.375	3.625				
2.0	7.000	6.625	3	0.750	0	0
2.5	13.625	10.375	3.750	0.750	0	0
3.0	24.000	14.875	4.500	0.750		
3.5	38.875	20.125	5.250			
4.0	59.000					

Since the tabular point $x = 1.5$ lies in the beginning of the table, we use the differentiation formula obtained by differentiating Newton's forward difference formula. Thus,

$$f'(x_0) = \frac{1}{h} \left[\Delta f(x_0) - \frac{1}{2} \Delta^2 f(x_0) + \frac{1}{3} \Delta^3 f(x_0) - \frac{1}{4} \Delta^4 f(x_0) + \dots \right]$$

Here $h = 0.5$. Therefore, we have

$$f'(1.5) = \frac{1}{0.5} [3.625 - 1.5 + 0.250] = 4.750.$$

For the second derivative, we have

$$f''(x_0) = \frac{1}{h^2} \left[\Delta^2 f(x_0) - \Delta^3 f(x_0) + \frac{11}{12} \Delta^4 f(x_0) \right],$$

which implies

$$f''(1.5) = \frac{1}{0.25} [3 - 0.750 + 0] = 9.$$

(b) Example 8.11

The actual value is 0.7857.

4. (a) The given equation is

$$f(x) = 3x - \cos x - 1 = 0.$$

We have

$$f(0) = -2 \text{ (ve)} \text{ and } f(1) = 3 - 0.5403 - 1 = 1.4597 \text{ (+ve).}$$

Hence, one of the roots of $f(x) = 0$ lies between 0 and 1. The values at 0 and 1 show that root is nearer to 1. So let us take $x = 0.6$. Further,

$$f'(x) = 3 + \sin x.$$

Therefore, the Newton-Raphson formula gives

$$\begin{aligned}x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{3x_n - \cos x_n - 1}{3 + \sin x_n} \\&= \frac{3x_n + x_n \sin x_n - 3x_n + \cos x_n + 1}{3 + \sin x_n} = \frac{x_n \sin x_n + \cos x_n + 1}{3 + \sin x_n}.\end{aligned}$$

Hence,

$$x_1 = \frac{x_0 \sin x_0 + \cos x_0 + 1}{3 + \sin x_0} = \frac{0.6(0.5646) + 0.8253 + 1}{3 + 0.5646} = 0.6071.$$

$$x_2 = \frac{x_1 \sin x_1 + \cos x_1 + 1}{3 + \sin x_1} = \frac{(0.6071)(0.5705) + 0.8213 + 1}{3 + 0.5705} = 0.6071.$$

Hence the required root, correct to four decimal places, is 0.6071.

(b) The given equation is

$$f(x) = \cos x - x e^x = 0.$$

We note that $f(0) = 1$, $f(1) = \cos 1 - e = 0 - e = -e$ (–ve). Hence, a root of the given equation lies between 0 and 1. By secant method, we have

$$x_{n+1} = x_n - \frac{x_{n-1} - x_n}{f(x_{n-1}) - f(x_n)} f(x_n).$$

So taking initial approximation as $x_0 = 0$, $x_1 = 1$, $f(x_0) = 1$ and $f(x_1) = -e = -2.1780$, we have

$$\begin{aligned}x_2 &= x_1 - \frac{x_0 - x_1}{f(x_0) - f(x_1)} f(x_1) \\&= 1 - \frac{-1}{1 + 2.178} (-2.178) = 0.3147.\end{aligned}$$

Further, $f(x_2) = f(0.3147) = 0.5198$. Therefore,

$$\begin{aligned}x_3 &= x_2 - \frac{x_1 - x_2}{f(x_1) - f(x_2)} f(x_2) \\&= 0.3147 - \frac{1 - 0.3147}{-2.178 - 0.5198} (0.5198) = 0.4467.\end{aligned}$$

Further, $f(x_3) = f(0.4467) = 0.2036$. Therefore,

$$\begin{aligned}x_4 &= x_3 - \frac{x_2 - x_3}{f(x_2) - f(x_3)} f(x_3) \\&= 0.4467 - \frac{0.3147 - 0.4467}{0.5198 - 0.2036} (0.2036) = 0.5318,\end{aligned}$$

$$f(x_4) = f(0.5318) = -0.0432.$$

Therefore,

$$\begin{aligned}x_5 &= x_4 - \frac{x_3 - x_4}{f(x_3) - f(x_4)} f(x_4) \\&= 0.5318 - \frac{0.4467 - 0.5318}{0.2036 + 0.0432} (-0.0432) = 0.5168,\end{aligned}$$

and

$$f(x_5) = f(0.5168) = 0.0029.$$

Now

$$\begin{aligned}x_6 &= x_5 - \frac{x_4 - x_5}{f(x_4) - f(x_5)} f(x_5) \\&= 0.5168 - \frac{0.5318 - 0.5168}{-0.0432 - 0.0029} (0.0029) = 0.5177,\end{aligned}$$

and

$$f(x_6) = f(0.5177) = 0.0002.$$

The sixth iteration is

$$\begin{aligned}x_7 &= x_6 - \frac{x_5 - x_6}{f(x_5) - f(x_6)} f(x_6) \\&= 0.5177 - \frac{0.5168 - 0.5177}{0.0029 - 0.0002} (0.0002) = 0.51776.\end{aligned}$$

We observe that $x_6 = x_7$ up to four decimal places. Hence $x = 0.5177$ is a root of the given equation correct to four decimal places.

5. (a) The given equations are

$$2x + y + z = 10, 3x + 2y + 3z = 18, x + 4y + 9z = 16.$$

The augmented matrix for given system of equations is

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 10 \\ 3 & 2 & 3 & 18 \\ 1 & 4 & 9 & 16 \end{array} \right] \begin{matrix} \leftarrow \text{Pivotal row} \\ m_{21} = \frac{3}{2} \\ m_{31} = \frac{1}{2} \end{matrix}$$

The result of first Gauss elimination is

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 10 \\ 0 & \frac{1}{2} & \frac{3}{2} & 3 \\ 0 & \frac{7}{2} & \frac{17}{2} & 11 \end{array} \right] \begin{matrix} \leftarrow \text{Pivotal row} \\ m_{32} = 7 \end{matrix}$$

The second elimination yields

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 10 \\ 0 & \frac{1}{2} & \frac{3}{2} & 3 \\ 0 & 0 & -2 & -10 \end{array} \right]$$

Thus, the given system of equations reduces to

$$\begin{aligned}2x + y + z &= 10, \\0.5y + 1.5z &= 3, \\-2z &= -10.\end{aligned}$$

Hence, back substitution yields

$$z = 5, y = -9, x = 7.$$

(b) The given equation can be written as

$$\begin{aligned}x &= \frac{1}{20}[17 - y + 2z], \\y &= \frac{1}{20}[-18 - 3x + z], \\z &= \frac{1}{20}[25 - 3x + 3y].\end{aligned}$$

Taking the initial rotation as $(x_0, y_0, z_0) = (0, 0, 0)$, we have by Gauss–Seidal's method,

$$\begin{aligned}x_1 &= \frac{1}{20}[17 - 0 + 0] = 0.85, \\y_1 &= \frac{1}{20}[-18 - 3(0.85) + 1] = -1.0275, \\z_1 &= \frac{1}{20}[25 - 2(0.85) - 3(-1.0275)] = 1.0108, \\x_2 &= \frac{1}{20}[17 + 1.0275 + 2(1.0108)] = 1.0024, \\y_2 &= \frac{1}{20}[-18 - 3(1.0024) + 1.0108] = -0.9998, \\z_2 &= \frac{1}{20}[25 - 2(1.0024) + 3(-0.9998)] = 0.9998, \\x_3 &= \frac{1}{20}[17 + 0.9998 + 2(0.9998)] = 0.99997, \\y_3 &= \frac{1}{20}[-18 - 3(0.99997) + 0.9998] = -1.00000, \\z_3 &= \frac{1}{20}[25 - 2(0.99997) + 3(-1.00000)] = 1.00000.\end{aligned}$$

The second and third iterations show that the solution of the given system of equations is

$$x = 1, y = -1, z = 1$$

6. (a) Exercise 6, Chapter 4.

Proceed as in Example 4.6. The answer is

$$\begin{bmatrix} 2 & 3.16 & 0 \\ 3.16 & 4.3 & -1.9 \\ 0 & -1.9 & 3.9 \end{bmatrix}$$

(b) Example 4.8

7. We have

$$\frac{dy}{dx} = xy + y^2, y(0) = 1.$$

We have $x_0 = 0$, $y_0 = 1$, $h = 0.1$. Therefore,

$$\begin{aligned} K_1 &= hf(x_0, y_0) = (0.1)f(0, 1) = 0.1, \\ K_2 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{K_1}{2}\right) = (0.1)f(0.05, 1.05) \\ &= 0.1[0.05(1.05) + (1.05)^2] = 0.1155, \\ K_3 &= hf\left(x_0 + \frac{h}{2}, y_0 + \frac{K_2}{2}\right) = (0.1)f(0.05, 1.0578) \\ &= 0.1[0.05(1.0578) + (1.0578)^2] = 0.1172, \\ K_4 &= hf(x_0 + h, y_0 + K_3) = (0.1)f(0.1, 1.1172) \\ &= 0.1[0.1(1.1172) + (1.1172)^2] \\ &= 0.13598. \end{aligned}$$

Therefore,

$$\begin{aligned} y(0.1) &= y_0 + \frac{1}{6}[K_1 + 2K_2 + 2K_3 + K_4] \\ &= 1 + \frac{1}{6}[0.1 + 2(0.1155) + 2(0.1172) + 0.13598] = 1.1169. \end{aligned}$$

Now $x_1 = 0.1$, $y_1 = 1.1169$, $h = 0.1$. Therefore,

$$\begin{aligned} K_1 &= hf(x_1, y_1) = (0.1)f(0.1, 1.1169) \\ &= (0.1)[0.1(1.1169) + (1.1169)^2] = 0.1359, \\ K_2 &= hf\left(x_1 + \frac{h}{2}, y_1 + \frac{K_1}{2}\right) = (0.1)f(0.15, 1.1848) \\ &= 0.1[0.15(1.1848) + (1.1848)^2] = 0.1581, \\ K_3 &= hf\left(x_1 + \frac{h}{2}, y_1 + \frac{K_2}{2}\right) = (0.1)f(0.15, 1.1959) \\ &= 0.1[0.15(1.1959) + (1.1959)^2] = 0.1610, \end{aligned}$$

$$\begin{aligned} K_4 &= hf(x_1 + h, y_1 + K_3) = (0.1)f(0.2, 1.2779) \\ &= 0.1[0.2(1.2779) + (1.2779)^2] = 0.1887. \end{aligned}$$

Hence,

$$\begin{aligned} y(0.2) &= y_1 + \frac{1}{6} [K_1 + 2K_2 + 2K_3 + K_4] \\ &= 1.1169 + \frac{1}{6} [0.1359 + 2(0.1581) + 2(0.1610) + 0.1887] = 1.2774. \end{aligned}$$

Now $x_2 = 0.2$, $y_2 = 1.2774$, $h = 0.1$. Therefore,

$$\begin{aligned} K_1 &= hf(x_2, y_2) = (0.1)f(0.2, 1.2774) \\ &= (0.1)[0.2(1.2774) + (1.2774)^2] = 0.1887, \end{aligned}$$

$$\begin{aligned} K_2 &= hf\left(x_2 + \frac{h}{2}, y_2 + \frac{K_1}{2}\right) = (0.1)f(0.25, 1.3718) \\ &= 0.1[0.25(1.3718) + (1.3718)^2] = 0.22248, \end{aligned}$$

$$\begin{aligned} K_3 &= hf\left(x_2 + \frac{h}{2}, y_2 + \frac{K_2}{2}\right) = (0.1)f(0.25, 1.3886) \\ &= 0.1[0.25(1.3886) + (1.3886)^2] = 0.2275, \end{aligned}$$

$$\begin{aligned} K_4 &= hf(x_2 + h, y_2 + K_3) = (0.1)f(0.3, 1.5049) \\ &= 0.1[0.3(1.5049) + (1.5049)^2] = 0.2716. \end{aligned}$$

Hence,

$$\begin{aligned} y(0.3) &= y_2 + \frac{1}{6} [K_1 + 2K_2 + 2K_3 + K_4] \\ &= 1.2774 + \frac{1}{6} [0.1887 + 2(0.22248) + 2(0.2275) + 0.2716] = 1.5041. \end{aligned}$$

We have thus

$$\begin{aligned} x_1 &= 0.1, y_1 = 1.1169, f_1 = 0.1(1.1169) + (1.1169)^2 = 1.3592, \\ x_2 &= 0.2, y_2 = 1.2774, f_2 = 0.2(1.2774) + (1.2774)^2 = 1.8874, \\ x_3 &= 0.3, y_3 = 1.5041, f_3 = 0.3(1.5041) + (1.5041)^2 = 2.714. \end{aligned}$$

Using the predictor, we have

$$\begin{aligned} y(0.4) &= y_4 = y_0 + \frac{4h}{3}[2f_3 - f_2 + 2f_1] \\ &= 1 + \frac{0.4}{3}[2(2.714) - 1.8872 + 2(1.3592)] = 1.8346. \end{aligned}$$

Then,

$$f_4 = 0.4(1.8346) + (1.8346)^2 = 4.0996.$$

Using corrector, we have

$$\begin{aligned}y(0.4) &= y_2 + \frac{h}{3}[f_4 + 4f_3 + f_2] \\&= 1.2774 + \frac{0.1}{3}[4.0996 + 4(2.714) + 1.8872] = 1.8388.\end{aligned}$$

8. Exercise 3, Chapter 11. From Figure 11.38, we have on setting $u_4 = 0$,

$$\begin{aligned}u_1 &= \frac{1}{4}[1000 + 2000 + 1000 + 0] = 1000 \text{ (diagonal five point formula),} \\u_2 &= \frac{1}{4}[1000 + 0 + 500 + 1000] = 625 \text{ (standard five point formula),} \\u_3 &= \frac{1}{4}[1000 + 500 + 2000 + 0] = 875 \text{ (standard five point formula),} \\u_4 &= \frac{1}{4}[0 + 875 + 625 + 0] = 375 \text{ (standard five point formula).}\end{aligned}$$

Now using Gauss–Seidel's formula, we have

$$\begin{aligned}u_1^{(1)} &= \frac{1}{4}[2000 + 625 + 875 + 1000] = 1125, \\u_2^{(1)} &= \frac{1}{4}[1125 + 500 + 375 + 1000] = 750, \\u_3^{(1)} &= \frac{1}{4}[2000 + 500 + 1125 + 375] = 1000, \\u_4^{(1)} &= \frac{1}{4}[1000 + 0 + 0 + 750] = 437.5, \\u_1^{(2)} &= \frac{1}{4}[2000 + 1000 + 750 + 1000] = 1187.5, \\u_2^{(2)} &= \frac{1}{4}[1000 + 500 + 1187.5 + 437.5] = 781.25, \\u_3^{(2)} &= \frac{1}{4}[2000 + 500 + 437.5 + 1187.5] = 1031.25, \\u_4^{(2)} &= \frac{1}{4}[0 + 0 + 1031.25 + 781.25] = 453.125.\end{aligned}$$

Continuing with the process, we shall get

$$u = 1208.3, u_2 = 791.7, u_3 = 1041.7, u_4 = 458.4.$$

Bibliography

- Atkinson, K.E. *An Introduction to Numerical Analysis*, John Wiley & Sons, Inc., New York, 1989.
- Froberg, C.E. *Introduction to Numerical Analysis*, Addison Wesley, 1968.
- Hartee, D.R. *Numerical Analysis*, Clarendon Press, Oxford, 1952.
- Hildebrand, F.B. *Introduction to Numerical Analysis*, Tata McGraw-Hill Publishing Company Ltd. New Delhi, 1956.
- Jordan, Charles, *Calculus of Finite Differences*, Chelsea Publishing Company, New York, 1947.
- Mathews, J.H. *Numerical Methods for Mathematics, Science and Engineering*, Prentice-Hall of India Pvt. Ltd. New Delhi, 2003.
- Salvadori, M.G. and M.L. Baron, *Numerical Methods in Engineering*, Prentice-Hall, Inc., New York, 1952.
- Scarborough, J.B. *Numerical Mathematical Analysis*, Oxford & IBH Publishing Co., New Delhi, 1974.
- Smith, G.D. *Numerical Solution of Partial Differential Equations*, Oxford University Press, London, 1965.
- Stanton, R.G. *Numerical Methods for Science and Engineering*, Prentice-Hall of India Pvt. Ltd. New Delhi, 1967.

This page is intentionally left blank

Index

A

Adams-Basforth method, 330
Adams-Moulton method, 334, 347
Approximate number, 1–2
Approximate values of roots, 11–12
Approximation of a function by chebyshev series, 198–200
Average error, 214

B

Bairstow iterative method, 45–49
Bender-Schmidt method, 446–447
Bessel's interpolation formula, 155–157, 194, 271
Bisection (Bolzano), 12–15, 27, 425
Boole's rule, 254, 263
Boundary value problems, 301, 352–357

C

Centered formula of order $O(h^2)$, 228–232, 234, 246
Centered formula of order $O(h^4)$, 229–231, 233–235
Central difference operator, 126
Characteristic equation, 85, 96, 101, 103–104, 106, 110, 290–299, 345–348
Convergence of iteration method, 26–27, 76–78
Chebyshev polynomial, 195–198
Closed type formulae, 333
Corrector, 256, 315, 336–339, 341–346, 442
Cote's formulae, 256–258, 267
Cote's numbers, 257
Crank-Nicholson formula, 390–392, 394
Crout's method, 66–70
Cubic spline, 202–208

D

Deviations, 82, 213–214
Diagonal five point formula, 366

Diagonalizable matrix, 87, 434

Difference equation, 288–299, 345–348, 355

Differentiation of function in unequal intervals, 244

Differentiation of Lagrange's polynomial, 179, 195, 201, 244–246

Differentiation of Newton polynomial, 246–250

Divided differences, 163–164

Dominant eigenvalue, 87–88, 92, 94

Dominant eigenvector, 88

E

Eigenvalue, 72, 85–118, 357–359
Eigenvalue problems, 357–359
Eigenvalues of symmetric tri-diagonal matrix, 115–117
Eigenvector, 85–118
Error
 absolute, 4
 general formula for, 5–7
 inherent (initial), 4
 percentage, 4
 relative, 4
 round off, 4
 truncation, 4

Error formulas (interpolation), 168–171

Error in Lagrange's interpolation formula, 179–180

Error term in quadrature formula, 258–262

Errors in centered formulae, 230–231

Euler-Maclaurin's formula, 277–278

Euler's method, 305–312

 Error analysis of, 305–308

 Improved, 308–309

 Modified, 309–312

Everett's interpolation formula, 158–161

Exact number, 1, 273

Explicit multi-step methods, 330–331

 Adams-Basforth, 330–331, 333–335

 Nystrom's, 330–331

F

Finite differences, 122–208
Finite difference method, 352–355
Fixed point iteration, 25–26
Formation of difference equation, 363–364
Fundamental theorem of integral calculus, 4, 261–262

G

Gauss-elimination method, 51–56
Gauss's backward difference formula, 151–152
Gauss's forward difference formula, 149–151
Gauss-Seidel iteration formula, 71–72, 399–401
Gauss-Seidel method, 71–76
General formula for errors, 5–7
General method for differentiation formulas, 235–244
General solution, 285–292, 301, 345–346, 348–349
Gershgorin Circles, 117–118
Graeffe's root squaring method, 37–40
Given's method, 101–109
Grid points, 364, 376, 378, 396–397, 400, 402

H

Hermite interpolation formula, 180–185
Heun's method, 315–316
Homogeneous difference equation, 289–292
Householder's method, 109–115
Hyperbolic equation, 402–407

I

Ill-conditioned system of equations, 82
Implicit multistep methods
 Adam-Moulton, 332–336
 Milne-Simpson, 332, 336–344

Intermediate value theorem, 2, 228
 Interval of differencing, 122,
 126–127, 131, 137, 144
 Interpolation, 122, 143–164
 Inverse interpolation, 172, 188
 Using Everett’s formula, 190–191
 Using Lagrange’s interpolation
 formula, 191–195
 Using Newton’s forward
 difference formula, 188–190
 Interpolation by spline function,
 200–202
 Initial value problems, 301, 305,
 309–310, 312, 315–317,
 321–322, 335, 339, 344, 440
 Iterative methods, 11, 70–79, 399–401

J

Jacobi iteration method, 70–71, 399
 Jacobi’s method for eigenvalues,
 94–101
 Jordon’s modification to Gauss’s
 method, 56–59
 Jury problems, 301

L

Lagrange’s interpolation coefficients,
 172
 Lagrange’s interpolation formula, 2,
 171–180
 Lattice points, 364
 Least square line approximation,
 213–218
 Least square parabola, 221–226

M

Maclaurin’s expansion, 3
 Marching problems, 301
 Maximum error, 5, 140, 169, 214
 Mean value operator, 127
 Mean value theorem, 2, 26
 Mechanical quadrature, 252–253
 Mechanical cubature, 252
 Method of collocation, 213
 Method of relaxation, 79–82
 Milne-Simpson formula, 332
 Milne-Simpson method, 336–344,
 442–443
 Multiplicity of zero, 11
 Muller’s method, 41–45
 Multistep methods, 330–344

N
 Natural cubic spline, 203–204,
 206, 208
 Newton-backward difference
 operator, 123
 Newton-forward difference operator,
 122, 292
 Newton-Raphson method
 fundamental formula in, 21
 order of convergence of,
 24–25
 for simultaneous equations,
 32–36
 square root of a number using,
 23–24
 sufficient condition for
 convergence of, 27–29
 Newton’s advancing difference
 formula, 126
 Newton’s forward difference
 formula, 143–144, 146–147,
 165, 188, 213, 237, 247, 249,
 252
 Newton’s backward difference
 formula, 144–148, 255, 330,
 332
 Newton’s divided difference formula,
 164–168, 170–171, 193,
 247, 249
 Normal equations, 214–216,
 222–226, 434–435
 Numerical differentiation, 122,
 228–250, 252
 Numerical instability, 143, 344
 Numerical unstability, 143
 Nystom’s method, 330–331

O
 Order of approximation, 7–9

P
 Partial differential equation, 1, 288,
 363–364
 elliptic, 363, 376–379
 Laplace equation, 443–444
 heat conduction equation, 363,
 389–399
 hyperbolic, 402–407
 parabolic, 363, 389–399
 wave equation, 402–407

Partially instable method, 344

Particular solution, 289,
 293–299
 Picard’s method of successive
 integration, 312–315, 342
 Pivotal element, 53
 Pivotal row, 53–54, 57–59,
 61–63
 Point Jacobi method, 94, 366
 Poisson’s equation, 379–383
 Power method, 88–94
 Power fit, 219–221
 Predictor, 255, 315, 336–339,
 341–343, 346, 442

R

Recurrence relations, 288
 Regula-Falsi method, 15–20
 convergence of, 16–20
 Relaxation method (Partial diff.eq.),
 70, 80–81
 Residuals, 79–82, 214, 221–222
 Richardson’s extrapolation, 231–234,
 263–265
 Rounding off, 1–2, 4, 82
 Rolle’s theorem, 2, 168, 179
 Romberg’s method, 262–277
 Root mean square error, 214
 Rotation matrix, 94–95, 98–102,
 107–108
 Runge-Kutta method, 317–330
 fourth order, 321–326
 for higher order differential
 equation, 328–330
 second order, 318–319
 for system of first order
 equation, 326–328
 third order, 319–320

S

Second order differential equations,
 349–352
 Shifting (enlargement) operator,
 125, 187
 Significant figure, 1–2, 4, 26,
 143, 357
 Simpson’s formula with end
 correction, 265–267
 Simpson’s one-third rule, 253–254,
 264, 270–271, 276–277, 282,
 438–439
 Single step methods, 301–330

Square root of a number by
iteration, 27

Square root method, 64

Stability of methods, 344–349

Standard five point formula,
365–366, 444

Stirling's interpolation formula,
152–154, 160, 162, 238–239,
241–242

T

Taylor series method, 301–305

Taylor's theorem, 3–5, 20, 128, 266
for function of two variable, 32,

317, 321

Trapezoidal rule, 252–253, 257, 260–
261, 263, 267, 274–276, 278–281,
283–284, 315, 332, 437–438

Triangular factorization, 59–63

Triangularization method, 51, 60

Triangularization of symmetric
matrix, 64–66

Throwback technique, 185–188

W

Weddle's rule, 254–256, 260, 276–277

Weierstrass's approximation
theorem, 213