Numerical Methods

AIMA 203

Department of Artificial Intelligence,
Kathmandu University, Dhulikhel, Kavre

Lecture notes by

Sandesh Thakuri

Sandesh.775509@cdmath.tu.edu.np

Text Book: S.S. Sastary, Introductory Methods of Numerical Analysis, Prentice Hall of India.

2025

Contents

1 Errors in Numerical Computing **4**

1.1 Introduction . . . 4
1.2 Approximations . . . 5
1.3 Error . . . 6
1.4 General Error Formula . . . 8

2 Root Finding **9**

2.1 Introduction . . . 9
2.2 Characteristic of Numerical Methods . . . 10
2.3 Bisection Method . . . 10
2.4 Iteration Method . . . 11
2.5 Newton-Rapshon's Method . . . 12
2.6 Secant Method . . . 12
2.7 System of Non-linear equations . . . 13
2.8 Rate of Convergence . . . 16
2.9 Exercise . . . 16
2.10 Lab Work . . . 16

3 System of Linear Equations **18**

3.1 Introduction . . . 18
3.2 LU Factorization Method . . . 19
3.3 Vector and Matrix Norms . . . 20
3.4 Ill-conditioned Linear Systems . . . 21
3.5 Iterative Method . . . 22

4 Interpolation **24**

4.1 Introduction . . . 24
4.2 Finite Differences . . . 24
4.3 Newton's formula for Interpolation . . . 26
4.4 Interpolation with Unevenly spaced points . . . 27
4.5 Divided Differences . . . 29

4.6 Inverse Interpolation 30
4.7 Introduction 31
4.8 Cubic Splines 32

5 Least Square Problems **35**
5.1 Introduction 35
5.2 Curve Fitting by Polynomials 36
5.3 Weighted Least Square Approximation 37
5.4 Linearization of Nonlinear Laws 37

6 Numerical Differentiation and Numerical Integration **38**
6.1 Numerical Differentiation 38
6.2 Numerical Integration 39
6.3 Double Integration 43

7 Solving Ordinary Differential Equations **46**
7.1 Introduction 46
7.2 Taylor's series method 46
7.3 Picard's Method of Successive Approximation 47
7.4 Euler's Method 48
7.5 Runge-Kutta Methods 50
7.6 Simultaneous and Higher Order Equations 51
7.7 Boundary Value Problems 52
7.8 Finite-Difference Method 52

1

Errors in Numerical Computing

1.1 Introduction

In practical applications, an engineer would finally obtain results in a numerical form. The aim of numerical analysis is to provide efficient methods for obtaining numerical answers to such problems.

1.1.1 Approximate Value

There are certain numbers whose exact value cannot be written. For the famous number π, we can only write value of π to certain degree of accuracy. For example, π is 3.1416 or 3.14159265. This values of π are called the approximate values of π. The exact value of π, we cannot write. Another such number is the *Euler's number 'e'*. One scenario is using the value of π in calculating the area of circle? *Can you come up with another scenario where approximate value of a number is used instead of its exact value?*

1.1.2 Significant Digits

(Significant Figures) The digits that are significant (important) in a number expressed as digits, are called significant digits.

Rules to identify significant figures in a number:

1. All non-zero digits are significant. The number 21.11 has four significant digits.

2. Zeros between two significant non-zero digits are significant. The number 20001 has five significant digits.

3. Zeros to the left of the first non-zero digit (leading zeros) are not significant. The number 0.0085 has two significant digits.

4. Zeros to the right of the last non-zero digit (trailing zeros) in a number with the decimal point are significant. The number 320. has three significant digits, and 320.00 has five significant digits.

5. When the decimal point is not written, the trailing zeros are not significant. The number 4500 may be written as 45×10^2 has two significant digits. However, the number 4500.0 has five significant digits.

6. Integers with trailing zeros may be written in scientific notation to specify the significant digits.

7.56×10^4	has 3 significant digits
7.560×10^4	has 4 significant digits
7.5600×10^4	has 5 significant digits

The concept of accuracy and precision are closely related to significant digits.

1. **Accuracy**

   This refers to the number of significant digits in a value. For example, the number 57.396 is accurate to five significant digits.

2. **Precision**

   This refers to the number of decimal positions, i.e. the order of magnitude of the last digit in a value. For example, the number 57.396 has a precision of 0.001 or 10^{-3}.

1.1.3 Exercise

1. Which of the following numbers has the greatest precision?

   a). 4.3201 b). 4.32 c). 4.320106

2. What is the accuracy of the following numbers?

   a). 95.763 b). 0.008472 c). 0.0456000 d). 36 e). 3600 f). 3600.00

1.2 Approximations

In numerical computation, we come across numbers which have large number of digits and it will be necessary to bring them to the required number of significant figures. For instance, the finite precision of computer storage, (using fixed number of bits), does not allow us to store the infinite digits of certain fractions like 1/3. So, the representation of 1/3 in the computers is going to an approximation.

1.2.1 Rounding off

Rounding off is the method of approximation used for the numbers. There are two types of rounding off.

Chopping

A number is written up to its certain digits and remaining digits are simply discarded. For example the number 1/3 is chopped to 0.3333.

Symmetric Rounding

A number is adjusted to the nearest representable value. For example 2.678 is rounded to 2.68, and 2.674 rounded to 2.67. Rules for rounding off.

1. To round-off a number to n significant digits , discard all digits to the right of the nth digit, and if the first digit of this discared number is,
2. less than half a unit in the nth place, leave the nth digit unaltered;
3. greater than half a unit in the nth place, increase the nth digit by unity;
4. exactly half a unit in the nth place, increase the nth digit by unity if it is odd; otherwise, leave it unchanged.

1.2.2 Truncation

While rounding off is the approximation in a number, truncation is the approximation in a mathematical procedure. Especially an infinite mathematical procedure or a complicated mathematical procedure is approximate by a finite mathematical procedure. The function $sinx$ is representation in a computer by its $series$; $sinx = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + ...$ But due to finite precision of computer, we represent sin function by only the finite terms of its series. Another scenario of truncation error is representing a continuous function in a computer. Digital systems like computer cannot represent a continuous phenomenon like a continuous function, because digital system is a discrete system. While approximating a function with a step-size h, the truncation error depends upon the step size. Forward difference approximation of a derivative of function has truncation error of order, $O(h)$.

1.3 Error

Both the rounding off and truncation causes error as they reduce a given number to an approximate value. Using approximate instead of exact value of number gives rise to the error.

1.3.1 Types of Error

Absolute Error

The absolute error E_A is the difference between the exact-value X and the approximate-value X_1 of a number.

$$E_A = X - X_1 = \delta X \tag{1.1}$$

Relative Error

The relative error E_R is the ratio of the absolute error to the exact-value.

$$E_R = \frac{E_A}{X} = \frac{X - X_1}{X} \tag{1.2}$$

Percentage Error

The percentage error E_P is

$$E_P = 100(E_R) = \frac{X - X_1}{X} \times 100 \tag{1.3}$$

The number 2.146879 is rounded to three significant digits. Find its errors.

1.3.2 Absolute errors of Sum, Product and Product

Suppose a_1 and a_2 are approximate values of two numbers, with their absolute errors E_A^1 and E_A^2 respectively.

Sum

If E_A is the absolute error of $a_1 + a_2$ then,

$$E_A = (a_1 + E_A^1) + (a_2 + E_A^2) - (a_1 + a_2) = E_A^1 + E_A^2 \tag{1.4}$$

Product

If E_A is the absolute error of a_1a_2 then,

$$E_A = (a_1 + E_A^1)(a_2 + E_A^2) - (a_1a_2) = a_1E_A^2 + a_2E_A^1 + E_A^1E_A^2 \approx a_1E_A^2 + a_2E_A^1 \tag{1.5}$$

Quotient

If E_A is the absolute error of a_1/a_2 then,

$$E_A = \frac{a_1 + E_A^1}{a_2 + E_A^2} - \frac{a_1}{a_2} = \frac{a_2E_A^1 - a_1E_A^2}{a_2(a_2 + E_A^2)} = \frac{a_2E_A^1 - a_1E_A^2}{a_2a_2(1 + E_A^2/a_2)} \approx \frac{a_2E_A^1 - a_1E_A^2}{(a_2)^2}$$

This implies,

$$E_A = \frac{a_1}{a_2}\left[\frac{E_A^1}{a_1} - \frac{E_A^2}{a_2}\right] \tag{1.6}$$

1.3.3 Operations on numbers of different absolute accuracies

While dealing with several numbers of different number of significant digits, the following procedure may be adopted:

1. Isolate the number with the greatest absolute error,
2. Round-off all other numbers retaining in them one digit more than in the isolated number,
3. Add up, and
4. Round-off the sum by discarding one digit.

1.3.4 Upper limit of Absolute Error

The number ΔX such that

$$|X_1 - X| \leq \Delta X. \tag{1.7}$$

Then ΔX is an upper limit on the magnitude of the absolute error and is said to measure *absolute accuracy.*

Theorem 1. If the number X is rounded to N decimal places, then $\Delta X = \frac{1}{2}(10^{-N})$.

Verify this theorem by taking $1.23x$*; x varies from 1 to 9, to two decimal places.*

1.4 General Error Formula

Let us consider a function u that depends upon the variables: x, y, z, i.e $u = f(x, y, z)$. Let $\Delta x, \Delta y, \Delta z$ be the errors in x, y, and z, respectively. Then the total error in u is Δu, which is approximated by du as follows: We have the total derivative:

$$du = \frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy + \frac{\partial u}{\partial z}\,dz \tag{1.8}$$

Approximating dx by Δx, dy by Δy and dz by Δz we get,

$$\Delta u \approx \frac{\partial u}{\partial x}\,\Delta x + \frac{\partial u}{\partial y}\,\Delta y + \frac{\partial u}{\partial z}\,\Delta z \tag{1.9}$$

The relation 1.9 gives the absolute general error formula. Then the relative error formula is given by $E_R = \frac{\Delta u}{u}$.

1.4.1 Exercise

1. Round off the following numbers to two decimal places.
   48.21461, 2.3742, 52.275

2. Round off the following numbers to four significant figures:
   38.46235, 0.70029, 0.0022218, 19.235101, 2.36425

3. Two numbers 3.1425 and 34.5851 are rounded to 2 decimal places. Find the error in their sum, product and quotient.

4. Find the absolute error in the sum of the numbers 105.6, 27.28, 5.63, 0.1467, 0.000523, 208.5, 0.0235, 0.432, 0.0467, where each number is correct to the digits given.

5. Find the absolute error in the product uv, where $u = 4.536$ and $v = 1.32$, the numbers being correct up to the digits given. Find the relative error of the quotient u/v as well.

6. Find the percentage error in $z = (1/8)\,xy^3$ when $x = 3014 \pm 0.0016$ and $y = 4.5 \pm 0.05$.

7. Prove that in a product of three nonzero numbers, the relative error does not exceed the sum of the relative errors of the numbers.

2

Root Finding

2.1 Introduction

Roots of an equation

$$f(x) = 0 \tag{2.1}$$

are the zeros, of f, which means, the values of x that makes the value of f zero. Basically equations are categorized into two. If f is a polynomial then the Equation-2.1 is a **polynomial equation** and if f is a non-polynomial then Equation-2.1 is a **transcendental equation**. For the polynomial equations following results hold:

1. Every polynomial equation of degree n has at most n real roots. These roots may be:

    (a) Real and distinct

    (b) Real and same

    (c) Complex

2. If n is even and the constant term is negative, then the equation has at-least one positive root and at-least one negative root.

3. The imaginary roots occurs in a pair (conjugate-pair). If the coefficients of f are rationals then, the irrational roots occurs in pairs (conjugate-pair).

4. Since imaginary roots occurs in a pair, if n is odd then, the polynomial equation has at least one real root and this root has its sign opposite to that of the last term.

5. **Descartes' Rule of Signs**

    a). A polynomial equation cannot have more number of positive real roots than the number of changes of signs in the coefficients of $f(x)$.

    b). A polynomial equation cannot have more number of negative real roots than the number of changes of signs in the coefficients of $f(-x)$.

2.2 Characteristic of Numerical Methods

1. What do you understand by numerical methods?

2. *Can you name a method that is against numerical method?*

The numerical methods are characterized as follows:

1. **Initial guess**

   It begins with an approximate value of the solution, called as the initial guess.

2. **Iteration**

   Then the initial guess is successively corrected by iterations.

3. **Stopping Criteria**

   The iterations are done until certain stopping criteria is meet. If we do not specify the stopping criteria then the iterations will run forever. Generally, the stopping criteria is that the solution has reached the required accuracy. The following tests may be used for that purpose.

1.	$\lvert x_{i+1} - x_i \rvert \leq E_\alpha$	E_α is the absolute error in x.
2.	$\dfrac{\lvert x_{i+1} - x_i \rvert}{x_{i+1}} \leq E_r$	E_r is the relative error in x.
3.	$\lvert f(x_{i+1}) \rvert \leq E$	E is the value of f at root.
4.	$\lvert f(x_{i+1}) - f(x_i) \rvert \leq E$	E is the difference in function values.

2.3 Bisection Method

Theorem 2 (Bolzano's Theorem)**.** If f(x) is continuous in $[a, b]$, and if $f(a)$ and $f(b)$ are of opposite signs, then $f(c) = 0$ for at least one number $c \in (a, b)$.

The Bisection method is based on Theorem-2. The word "bisection" means "half". Using this method the root c of f is given by $c \approx \dfrac{a+b}{2}$. Let $x_1 = \dfrac{a+b}{2}$. If $f(x_1) \neq 0$ then, the root, c lies either in $[a, x_1]$ or in $[x_1, b]$. If $f(a)f(x_1) < 0$ then, c lies in $[a, x_1]$ else, it lies in $[x_1, b]$.

At each step of this method, the given interval is bisected, so the length of the interval is halfed. At nth step the length of the interval is $\dfrac{|b-a|}{2^n}$. If the tolerance of the given approximation is ϵ then we must have $\dfrac{|b-a|}{2^n} \leq \epsilon$. And the number of steps required to reach this accuracy is $n \geq log_2(|b-a|/\epsilon) = \dfrac{log_e(|b-a|/\epsilon)}{log_e 2}$.

2.3.1 Procedure

1. Choose two real numbers a and b such that $f(a)f(b) < 0$.
2. Set $x_0 = 0$ and $x_1 = \dfrac{a+b}{2}$.
3. Do
   $\epsilon_r = \left|\dfrac{x_0 - x_1}{x_0}\right|$
   If $\epsilon_r < tolerance$ then $root = x_1$,
   else
   $x_0 = x_1$ and if $f(a)f(x_1) < 0$ then $x_1 = \dfrac{a + x_1}{2}$
   if $f(x_1)f(b) < 0$ then $x_1 = \dfrac{x_1 + b}{2}$.

2.3.2 Exercise

Using Bisection Method:

1. Find a root of $f(x) = x^3 - x - 1 = 0$, correct to 4 decimal places.
2. Obtain a root correct up to three decimal places:

   a). $x^3 - 4x - 9 = 0$
   b). $5x\,log_{10}^x - 6 = 0$
   c). $x^2 + x - cosx = 0$
   d). $x = e^{-x}$

2.4 Iteration Method

Steps:

1. Re-write the given equation $f(x) = 0$ in the form $x = \phi(x)$. This equation is of **iterative-type**. Meaning we can substitute a value of x in $\phi(x)$ to get another value of x, and continue this process to get the desired value of x, if the iteration is of convergent one.
2. Choose an initial root of f, x_0.
3. $x_1 = \phi(x_0)$, $x_2 = \phi(x_1)$ and so on.

The sequence $x_0, x_1, x_2, ...$ may not converge to a definite number. But if the sequence converges to a definite number ζ, then ζ is a root of the given equation. The sufficient condition for the sequence of the approximations $x_0, x_1, x_2, ...$ by iteration method to converge is that $|\phi'(x)| < 1$ in some neighborhood (interval) of x_0.

2.4.1 Exercise

Using Iteration Method:

1. Find a root of $2x - 3 - cosx = 0$, correct to 3 decimal places lying in $[3/2, \pi/2]$.

2.5 Newton-Rapshon's Method

Steps:

1. Choose an initial guess solution of the given equation $f(x) = 0$, x_0.

2. Let x_1 be a solution, which is more close to the exact solution of $f(x) = 0$. Then Using Taylor's expansion of f about x_0:

$$f(x_1) = f(x_0) + (x_1 - x_0)f'(x_0) + (x_1 - x_0)^2 f''(x_0) + ... = 0$$

   Neglecting the second and higher order derivatives, we get

$$f(x_0) + (x_1 - x_0)f'(x_0) = 0 \tag{2.2}$$

   The equation-2.8 is a linear equation, so this is an linear approximation. This equation is infact the tangent to the curve of the function $f(x)$ at $(x_0, f(x_0))$. And it is the point x_1 where the tangent meets the $x-axis$. So, the next approximation after x_0 by Newton-Rapshon's method is the point on $x - axis$, where the tangent to the f at x_0 meets the $x - axis$. This point can be solved as follows:

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

3. Successive approximation are given by $x_2, x_3, x_4, ...$, where $x_{n+1} = x_n - \dfrac{f(x_n)}{f'(x_n)}$

2.5.1 Exercise

Using Newton-Rapshon's Method:

1. Find a root of $f(x) = xe^x - 1 = 0$, correct to 4 decimal places.

2. Find a non-zero root of the equation $x^2 + 4sinx = 0$ choosing $x_0 = \pi$.

3. Using the Newton-Raphson method, derive a formula for finding the kth root of a positive number N and hence compute the value of $(25)^{1/4}$.

2.6 Secant Method

In Newton-Rapshon's method we use a tangent to the curve to get close to the root of the function. So, Newton-Rapshon's method requires the evaluation of derivatives of the function, which may not always exit. So we replace the tangent, with a secant to approximate the root of the function.

Steps:

1. Choose two initial guess solutions of the given equation $f(x) = 0$, x_{-1} and x_0.

2. The slope of the secant is $\dfrac{f(x_0) - f(x_{-1})}{x_0 - x_{-1}}$.

3. Then equation of the line passing through the points of given by the two initial guesses is $f(x) - f(x_0) = \dfrac{f(x_0) - f(x_{-1})}{x_0 - x_{-1}}(x - x_0)$.

4. x_1 is the point where the secant meets the $x - axis$ so, $f(x_1) = 0$. This gives,

$$\begin{aligned} 0 - f(x_0) &= \frac{f(x_0) - f(x_{-1})}{x_0 - x_{-1}}(x_1 - x_0) \\ x_1 - x_0 &= -\frac{x_0 - x_{-1}}{f(x_0) - f(x_{-1})} f(x_0) \\ x_1 &= x_0 - \frac{x_0 - x_{-1}}{f(x_0) - f(x_{-1})} f(x_0) \end{aligned}$$

5. This generalizes to

$$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} f(x_n) \tag{2.3}$$

You can get this relation-2.3 just by plugging $f'(x_n) = \frac{f(x_n)-f(x_{n-1})}{x_n - x_{n-1}}$ as the slope of the tangent in Newton Rapshon's method is just approximated by the slope of the secant in Secant method.

Exercise

1. Find a real root of the equation $x^3 - 2x - 5 = 0$ with three iterations.

2.7 System of Non-linear equations

For now we consider only a system of two equations. Let a system of two equations be

$$f(x, y) = 0, \qquad g(x, y) = 0 \tag{2.4}$$

2.7.1 Method of Iteration

First we assume that the sytem of equations 2.4 may be written in the form

$$x = F(x, y), \qquad y = G(x, y) \tag{2.5}$$

where the function F and G satisty the following conditions in a **closed** neighborhood of R of the root (α, β):

i) F and G and their firt partial derivatives are continuous in R, and

ii) $\left|\dfrac{\partial F}{\partial x}\right| + \left|\dfrac{\partial F}{\partial y}\right| < 1$ and $\left|\dfrac{\partial G}{\partial x}\right| + \left|\dfrac{\partial G}{\partial y}\right| < 1$, for all (x, y) in R.

If (x_0, y_0) is an initial approximation to the root (α, β), then Equations 2.5 give the sequence

$$\begin{aligned} x_1 &= F(x_0, y_0), & y_1 &= G(x_0, y_0) \\ x_2 &= F(x_1, y_1), & y_2 &= G(x_1, y_1) \\ &\dots \\ x_{n+1} &= F(x_n, y_n), & y_{n+1} &= G(x_n, y_n) \end{aligned} \tag{2.6}$$

For faster convergence, recently computed values of x_i may be used in the evaluation of y_i in Equations. Above conditions are sufficient for convergence and in the limit we obtain,

$$\alpha = F(\alpha, \beta) \quad and \quad \beta = G(\alpha, \beta) \tag{2.7}$$

Hence (α, β) is the root of the system 2.4.

Exampale

Exmaple: Find a real root of the equations $y^2 - 5y + 4 = 0$ and $3yx^2 - 10x + 7 = 0$ correct to 4 decimal places using initial approximation $(0, 0)$

Rewriting the equations in the form:

$$x = \frac{3yx^2 + 7}{10}, y = \frac{y^2 + 4}{5} \quad OR \quad x = \sqrt{\frac{10x - 7}{3y}}, y = \sqrt{5y - 4}$$

n	x	y
1	0.70000	0.80000
2	0.81760	0.92800
3	0.88610	0.97224
4	0.92901	0.98905
5	0.95608	0.99564
6	0.97303	0.99826
7	0.98354	0.99931

n	x	y
8	0.99001	0.99972
9	0.99395	0.99989
10	0.99635	0.99996
11	0.99780	0.99998
12	0.99868	0.99999
13	0.99920	1.00000
14	0.99952	1.00000

Exercise: solve the system: $x^2 + y = 11, \quad x + y^2 = 7$.

2.7.2 Newton-Raphson Method

Let (x_0, y_0) be an initial approximation to the root of the system of equations in two variables 2.4. If $(x_0 + h, y_0 + k)$ is the root of the system, then we must have

$$f(x_0 + h, y_0 + k) = 0 \qquad g(x_0 + h, y_0 + k) = 0$$

Assuming that f and g are sufficiently differentiable, we expand both of these functions by Taylor's series to obtain

$$f_0 + h\frac{\partial f}{\partial x_0} + k\frac{\partial f}{\partial y_0} ... = 0$$
$$g_0 + h\frac{\partial g}{\partial x_0} + k\frac{\partial g}{\partial y_0} ... = 0$$

where, $\dfrac{\partial f}{\partial x_0} = \left[\dfrac{\partial f}{\partial x}\right]_{x=x_0}$, $f_0 = f(x_0, y_0)$, etc

Neglating the second and higher-order derivatives terms, we get,

$$
\begin{aligned}
h\frac{\partial f}{\partial x_0} + k\frac{\partial f}{\partial y_0} ... &= -f_0 \\
h\frac{\partial g}{\partial x_0} + k\frac{\partial g}{\partial y_0} ... &= -g_0
\end{aligned} \tag{2.8}
$$

The system of equations 2.8 possesses a unique solution if

$$
D = \begin{vmatrix} \frac{\partial f}{\partial x_0} & \frac{\partial f}{\partial y_0} \\ \frac{\partial g}{\partial x_0} & \frac{\partial g}{\partial y_0} \end{vmatrix} \neq 0
$$

By Cramer's rule

$$
h = \frac{1}{D}\begin{vmatrix} -f_0 & \frac{\partial f}{\partial y_0} \\ -g_0 & \frac{\partial g}{\partial y_0} \end{vmatrix} \quad and \quad k = \frac{1}{D}\begin{vmatrix} \frac{\partial f}{\partial x_0} & -f_0 \\ \frac{\partial g}{\partial x_0} & -g_0 \end{vmatrix} \tag{2.9}
$$

The new approximations are, therefore

$$
x_1 = x_0 + h \quad and \quad y_1 = y_0 + k \tag{2.10}
$$

Exmaple: Find a real root of the equations $3yx^2 - 10x + 7 = 0$ and $y^2 - 5y + 4 = 0$ correct to 4 decimal places using initial approximation $(0, 0)$

Given, $f(x, y) = 3yx^2 - 10x + 7 = 0, \quad g(x, y) = y^2 - 5y + 4 = 0$

Then, $\frac{\partial f}{\partial x} = 6yx - 10, \quad \frac{\partial f}{\partial y} = 3x^2, \quad \frac{\partial g}{\partial x} = 0, \quad \frac{\partial g}{\partial y} = 2y - 5$

x	y	f	g	$\frac{\partial f}{\partial x}$ a	$\frac{\partial f}{\partial y}$ b	$\frac{\partial g}{\partial x}$ c	$\frac{\partial g}{\partial y}$ d	D $ad - bc$	D_1 $-fd + gb$	D_2 $-ga + fc$	h $\frac{D_1}{D}$	k $\frac{D_2}{D}$
0.000	0.000	7.000	4.000	-10.000	0.000	0.000	-5.000	50.000	35.000	40.000	0.700	0.800
0.700	0.800	1.176	0.640	-6.640	1.470	0.000	-3.400	22.576	4.939	4.250	0.219	0.188
0.919	0.988	0.315	0.035	-4.552	2.532	0.000	-3.024	13.764	1.042	0.161	0.076	0.012
0.994	1.000	0.022	0.000	-4.033	2.967	0.000	-3.000	12.101	0.067	0.001	0.006	0.000
1.000	1.000	0.000	0.000	-4.000	3.000	0.000	-3.000	12.000	0.000	0.000	0.000	0.000

Exercise: solve the system: $x^2 - y^2 = 4, \quad x^2 + y^2 = 16.$

2.7.3 Exercise

1. Find a real root of the system: $y^2 - 5y + 4 = 0$ and $3x^2y - 10x + 7 = 0$ correct to 4 decimal places using initial approximation $(0, 0)$.

2. Solve the system: $x^2 + y = 11$, $x + y^2 = 7$.

3. Solve the system: $x^2 - y^2 = 4$, $x^2 + y^2 = 16$.

2.8 Rate of Convergence

If a sequence $x_1, x_2, x_3, \ldots$ converges to a value α then, the rate of convergence of the sequence (x_n) to α is p, if p is the **largest** possible number such that:

$$\lim_{n\to\infty} \frac{|x_{n+1} - \alpha|}{|x_n - \alpha|^p} = constant \tag{2.11}$$

In another words, if $\epsilon_n = x_n - \alpha$ is the error at nth step then, the sequence of the error (ϵ_n) is a decreasing sequence and the positive integer p defines the rate at which the sequence of error is decreasing and converging to the correct value α. So we have, e

$$\lim_{n\to\infty} \frac{|\epsilon_{n+1}|}{|\epsilon_n|^p} = constant$$

When $p = 1$, we say the sequence or the method converges **linearly**, when $p = 2$ we say the method converges **quadratically**, when $1 < p < 2$, we say the method converges **superlinearly**.

2.9 Exercise

1. Find the rate of convergence of Bisection Method.
2. The rate of convergence of Newton Rapshon's method is quadratic. Does it converge faster than Bisection Method?
3. Write the rate of convergence of secant method.

2.10 Lab Work

1. Try a suitable data structure for storing the sequence of outputs in a iteration.
2. Find a way of representing a function.
3. Find an efficient way of evaluating a function or a polynomial.
4. Draw a flow chart for the working of Bisection Method.
5. Writing a function in Python.

```
def bisection_method(f, a, b, tol=1e-6, max_iter=100):
    """
    Find a root of a function f(x) in the interval [a, b] using the bisection method.

    Parameters:
    f (function): The function for which to find a root.
    a (float): Left endpoint of the interval.
    b (float): Right endpoint of the interval.
    tol (float): Tolerance (stopping criterion).
```

```
    max_iter (int): Maximum number of iterations.

    Returns:
    float: Approximation of the root.
    int: Number of iterations performed.
    """
    # Check if the interval contains a root
    if f(a) * f(b) >= 0:
    raise ValueError("Function must have opposite signs at the
    endpoints of the interval.")

    iter_count = 0
    c = a

    while (b - a) / 2 > tol and iter_count < max_iter:
        c = (a + b) / 2  # Midpoint

        if f(c) == 0:
            break  # Exact solution found
        elif f(a) * f(c) < 0:
            b = c  # Root is in left half
        else:
            a = c  # Root is in right half

        iter_count += 1

        return c, iter_count

# Example usage:
if __name__ == "__main__":
    # Define the function whose root we want to find
    def f(x):
        return x**3 - 2*x - 5

    # Initial interval
    a, b = 2, 3

    # Apply bisection method
    root, iterations = bisection_method(f, a, b)

    print(f"Approximate root: {root:.6f}")
    print(f"Found in {iterations} iterations")
    print(f"f(root) = {f(root):.6f}")
```

3 System of Linear Equations

3.1 Introduction

A completely general system of m linear equations in n unknowns is of the following form:

$$
\begin{gathered}
a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1 \\
a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2 \\
. \\
. \\
. \\
a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = b_m
\end{gathered}
$$

The a_{1j} is the coefficient of x_j in the ith equation. The data for this system of equations are all the numbers a_{ij} and b_i. Now consider the four matrices.

$$
A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ . & & & \\ . & & & \\ . & & & \\ a_{m1} & a_{m2} & ... & a_{mn} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{bmatrix}, \qquad [A|b] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & ... & a_{1n} & b_1 \\ a_{21} & a_{22} & ... & a_{2n} & b_2 \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ a_{m1} & a_{m2} & ... & a_{mn} & b_n \end{array}\right]
$$

In this context of a system of equations; A is called the coefficient matrix, x is called vector of unknowns, b is called righthandside vector, $[A|b]$ is called the augmented matrix. So, the **matrix notation** of the system of linear equations is $Ax = b$.

3.2 LU Factorization Method

Theorem 3. Let $A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ . & & & \\ . & & & \\ . & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$ be a singular matrix. Then if A is a non-singular matrix then A can be factorized as $A = LU$, where

- L is a lower triangular matrix, and
- U is an upper triangular matrix

It is the standard result of linear algebra that such a factorization, when exists, is *unique*.

Procedure for finding L and U

Let $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$ Multiplying the matrices on the right side and equating the coefficients, we get

$$\begin{aligned} u_{11} &= a_{11}, & u_{12} &= a_{12}, & u_{13} &= a_{13} \\ l_{21}u_{11} &= a_{21}, & l_{21}u_{22} + u_{22} &= a_{22}, & l_{21}u_{23} + u_{23} &= a_{23} \\ l_{31}u_{11} &= a_{31}, & l_{31}u_{22} + l_{32}u_{22} &= a_{32}, & l_{31}u_{23} + l_{32}u_{33} &= a_{33} \end{aligned}$$

From these equations we obtain

$$l_{21} = \frac{a_{21}}{a_{11}}, \qquad l_{31} = \frac{a_{31}}{a_{11}}, \qquad u_{22} = a_{22} - a_{12}\frac{a_{21}}{a_{11}}, \dots \tag{3.1}$$

This procedure is a systematic one. First we determine first row of U and first column of L, then we determine second row of U and second column of L, and finally third row of U.

3.2.1 Method

Let $Ax = b$ be a system of linear equations. Let $A = LU$ be the LU factorization of A. Then we have,

$$Ax = b \implies LUx = B \implies Ly = b \tag{3.2}$$

where, $y = Ux$.
Since L is an lower triangular matrix y can be solved conveniently by forward substituion in 3.2. Then x can be solved using backward substituion in $Ux = y$, as U is an upper triangular matrix.

3.2.2 Tridiagonal Systems

The matrix of the form

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 & ... & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & ... & 0 & 0 & 0 & 0 \\ 0 & a_{32} & a_{33} & a_{34} & ... & 0 & 0 & 0 & 0 \\ ... \\ 0 & 0 & 0 & 0 & ... & 0 & a_{(n-1)(n-2)} & a_{(n-1)(n-1)} & a_{(n-1)n} \\ 0 & 0 & 0 & 0 & ... & 0 & 0 & a_{n(n-1)} & a_{nn} \end{bmatrix}$$

is called tridiagonal matrix. It is a square matrix with nonzero elements only on the main diagonal, subdiagonal, superdiagonal. The method of LU factorization can conveniently apllied to the system having a tridiagonal matrix as the coefficient matirx.

3.2.3 Exercise

Solve the system:

1. $2x - y = 0, \quad -x + 2y - z = 0, \quad -y + 2z - u = 0, \quad -z + 2u = 1.$
2. $3x_1 - x_2 = -1, \quad -x_1 + 3x_2 - x_3 = 7, \quad -x_2 + 3x_3 = 7$

3.3 Vector and Matrix Norms

Vector Norm

The distance between a vector and the null vector is a measure of the *length* of the vector. This is called a norm of the vector. Mathematically, a norm of a vector in a vector space V is defined as a function $\|.\| : V \mapsto \mathbb{R}$ that satisfies the following conditions:

i) $\|x\| \geq 0$

ii) $\|x\| = 0$ if and only if $x = 0$.

iii) $\alpha\|x\| = |\alpha|\|x\|$ for any real α.

iv) $\|x + y\| \leq \|x\| + \|y\|$.

There are different types of length of a vector. These gives rise to different types of norms. Some useful norms are as follows:

1. $L1$ norm also called
   Rectangular norm or Manhattan norm: $\|x\|_1 = |x_1| + |x_2| + ...|x_n|$.
2. $L2$ norm also called **Euclidean norm**: $\|x\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + ...|x_n|^2}$.
3. $L\infty$ norm also called
   Maximum norm or Uniform norm: $\|x\|_\infty = max\{|x_1|, |x_2|, ..., |x_n|\}$.

Matrix Norm

Corresponding to the vector norms we have the three matrix norms defned as follows:

1. **Column Norm**: $\|A\|_1 = \max_j \sum_i |a_{ij}|$.

2. **Euclidean Norm**: $\|A\|_e = \left[\sum_{i,j} |a_{ij}|^2\right]^{1/2}$.

3. **Row Norm**: $\|A\|_\infty = \max_i \sum_j |a_{ij}|$.

3.3.1 Exercise

1. Find $L1$, $L2$, $L\infty$ norms of $x = (1, -13, 5, 3, -4)$.

2. Find the Column, the Euclidean and the Row norm of $A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 4 & 3 \\ 1 & 3 & 2 \end{bmatrix}$

3.4 Ill-conditioned Linear Systems

1. In practical applications, one usually encouners system of equations in which small changes in the coefficients of the system produce large changes in th solution. Such system are said to be *ill-conditioned*.

2. On the other hand, if the corespoinding changes in the solution is also small, then the system is *well-conditioned*.

Definition 1 (Condition Number)**.** The quantity $c(A) = \|A\|\|A^{-1}\|$, gives the measure of the condition of the matrix A, and is called the condition number. Here $\|.\|$ is any matrix norm.

Ill-conditioning can usually be expected when $det(A)$, in the system $Ax = b$, is small. Let $A = [a_{ij}]$ and $s_i = \left[a_{i1}^2 + a_{i2}^2 + ... + a_{in}^2\right]^{1/2}$. Let $k = \dfrac{det(A)}{s_1 s_2 ... s_n}$. Then the system whose coefficient matrix is A is ill-conditioned if k is very small compared to unity. Otherwise, it is well-conditioned.

3.4.1 Exercise

1. Show that the matrix $A = \begin{bmatrix} 25 & 24 & 10 \\ 66 & 78 & 37 \\ 92 & -73 & -80 \end{bmatrix}$ is ill-conditioned.

3.4.2 Method for Ill-conditioned Systems

In general, the accuracy of an approximate solution can be improved upon by an iterative procedure.

3.5 Iterative Method

The general system of linear equations with n unknowns and n linear equations is:

$$a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2$$
$$.$$
$$.$$
$$.$$
$$a_{n1}x_1 + a_{n2}x_2 + ... + a_{nn}x_n = b_n$$

For the iterative this system has to be rearranged as following

$$x_1 = \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}}x_2 - \frac{a_{13}}{a_{11}}x_3 - ... - \frac{a_{1n}}{a_{11}}x_n$$
$$x_2 = \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}}x_1 - \frac{a_{23}}{a_{22}}x_3 ... - \frac{a_{2n}}{a_{22}}x_n$$
$$.$$
$$.$$
$$.$$
$$x_n = \frac{b_n}{a_{nn}} - \frac{a_{n1}}{a_{nn}}x_1 - \frac{a_{n2}}{a_{nn}}x_2 - ... - \frac{a_{n(n-1)}}{a_{nn}}x_{n-1}$$

We can write this system as $x = Bx + C$. This equation is a iterative in nature as x is a function of x-itself. So the iteration formula is

$$x^{n+1} = Bx^n + C \tag{3.3}$$

The necessary condition for the convergence of the iterative method is that the diagonal elements a_{ii} do not vanish. If this is not the case then, the equations should be rearranged so that, the system satisfies this condition.

1. **Method of simultaneous displacements**

   This method 7.11 is due to *Jacobi* and is called method of simultaneous displacements, because $x_1, ..., x_n$ values are displaced simultaneous in 7.11. It can be shown that a sufficient condition for the convergence of this method is $\|B\| < 1$.

2. **Method of successive displacements** But if the values $x_1, ..., x_n$ are displaced successively; obtain x_1 first from the initialization, then use this value to obtain x_2 and so on, then this medthod is called method of successive displacements which is also commonly known as *Gauss-Seidel method*.

The Jacobi and Gauss-Seidel methods converge for any choice of the first approximation $(x_j^{(1)})$ $j = 1, 2, ..., n$, if every equation of the system satisfies the condition that

$$\sum_{j=1, j\neq i}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| \leq 1, \qquad i = 1, 2, ..., n \tag{3.4}$$

It is found that the Gauss-Seidel method converges **twice as fast** as the Jacobi method.

3.5.1 Exercise

1. $10x - 2y - z - u = 3, \quad -2x + 10y - z - u = 15, \quad -x - y + 10z - 2u = 27,$
   $-x - y - 2z + 10u = 9$

2. $2x - y = 1, \quad -x + 3y - z = 8, \quad -y + 2z = -5$

3. An approximate solution of the system $10x_1 + x_2 + x_3 = 12, \ x_1 + 10x_2 + x_3 = 12, \ x_1 + x_2 + 10x_3 = 12$ is given as $x_1^{(0)} = 0.4, \ x_2^{(0)} = 0.6, \ x_3^{(0)} = 0.8$. Use the iterative method to improve this solution.

4. Solve the system: $10x + 2y + z = 9, \quad 2x + 20y - 2z = -44, \quad -2x + 3y + 10z = 22.$

5. Apply upto six iterations of Gauss-Seidel to solve: $28x + 4y - z = 32, \quad 2x + 17y + 4z = 35, \quad x + 3y + 10z = 24.$

6. **Cholesky's method** A matrix A is called a symmetric matrix if $a_{ij} = a_{ji}$. For symmetric matrix the LU decomposition is more conveniently obtained, since $U = L^T$. This method is called Cholesky's method. Solve the system by cholesky's method : $5x_1 + x_3 = 8, \quad -2x_2 = -4, \quad x_1 + 5x_3 = 16$

4

Interpolation

4.1 Introduction

The central problem of numerical analysis is: given a set of tabular values $(x_0, y_0), (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ satisfying the realtion $y = f(x)$, where the explict nature of $f(x)$ is not known, is to find a simpler function $\phi(x)$ such that $\phi(x)$ and $f(x)$ agree at the set of the tabulated points. Such process is called **interpolation**. If $\phi(x)$ is a polynomial, then the process is called *polynomial interpolation*, and $\phi(x)$ a interpolating polynomial. In this chapter we are concerned with only polynomial interpolation.

Theorem 4 (Weiestrass Theorem)**.** If $f(x)$ is continuous in $x_0 \geq x \geq x_n$, then given $\epsilon > 0$, there exists a polynomial $P(x)$ such that $|f(x) - P(x)| < \epsilon$, for all $x \in (x_0, x_n)$.

So, this gives the justification for approximating a function with a polynomial function.

4.2 Finite Differences

Assume that we have a table of values $(x_i, y_i),\ i = 0, 1, ..., n$ of any function $y = f(x)$. Suppose the values of x are equally spaced, i.e., $x_i = x_0 + ih,\ i = 0, 1, ..., n$ for some constant h.
Then, the values $y_1 - y_0$, $y_2 - y_1$, ..., $y_n - y_{n-1}$, are called the differences of y. If $y_{i+1} - y_i$ is taken as the i^{th} difference, then we are considering it as a forward difference, if it is taken as the $(i+1)^{th}$ differnce then, we are considering it as a backward differnce, and if it is taken as $((i+1)/2)^{th}$ then, we are considering it as a central differnce. More on these topics in coming subsections.

4.2.1 Forward Differences

The forward differnce Δy_i is defined as $\Delta y_i = y_{i+1} - y_i$, where Δ is called the *forward difference operator*. We have,

$$\Delta y_0 = y_1 - y_0, \quad \Delta y_1 = y_2 - y_1, \quad ..., \quad \Delta y_{n-1} = y_n - y_{n-1} \tag{4.1}$$

These differnces $\Delta y_i,\ i = 0, 1, ..., (n-1)$ are called *first forward differences.*

- For $(n+1)$ points, there are n forward differnces.
- These n-forward differnces begin at Δy_0, and end at Δy_{n-1}. There is no Δy_n.

Second Forward Differnces

The forward differences of the first forward differences are called second forward differences. They are denoted by $\Delta^2 y_i,\ i = 0, 1, ..., (n-2)$. That is,

$$\Delta^2 y_0 = \Delta y_1 - \Delta y_0, \quad \Delta^2 y_1 = \Delta y_2 - \Delta y_1, \quad ..., \quad \Delta^2 y_{n-2} = \Delta y_n - \Delta y_{n-1} \tag{4.2}$$

Here, $\Delta^2 y_0 = \Delta y_1 - \Delta y_0 = y_2 - y_1 - (y_1 - y_0) = y_2 - 2y_1 - y_0$.

Higer Order Forward Differnces

Here, $\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0 = \Delta y_2 - \Delta y_1 - (\Delta y_1 - \Delta y_0) = ... = y_3 - 3y_2 + 3y_1 - y_0$.
The coefficients occuring are the **binomial coefficients**.

Programming

For practical computations, the following notation is useful: $y_j = DEL(0, j)$, and

$$\Delta y_j = DEL(0, j+1) - DEL(0, j)$$
$$\Delta^i y_j = DEL(i-1, j+1) - DEL(i, j)$$

4.2.2 Backward Differences

The backward differnce ∇y_{i+1} is defined as $\nabla y_{i+1} = y_{i+1} - y_i$, where ∇ is called the *backward difference operator*. We have,

$$\nabla y_1 = y_1 - y_0, \quad \nabla y_2 = y_2 - y_1, \quad ..., \quad \nabla y_n = y_n - y_{n-1} \tag{4.3}$$

These differnces $\nabla y_{i+1},\ i = 0, 1, ..., (n-1)$ are called *first backward differences*.

- For $(n+1)$ points, there are n backward differnces.
- These n-backward differnces begin at ∇y_1, and end at ∇y_n. There is no ∇y_0.

Second Backward Differnces

The backward differences of the first backward differences are called second backward differences. They are denoted by $\nabla^2 y_{i+2},\ i = 0, 1, ..., (n-2)$. That is,

$$\nabla^2 y_2 = \nabla y_2 - \nabla y_1, \quad \nabla^2 y_3 = \nabla y_3 - \nabla y_2, \quad ..., \quad \nabla^2 y_n = \nabla y_n - \nabla y_{n-1} \tag{4.4}$$

Here, $\nabla^2 y_2 = \nabla y_2 - \nabla y_1 = y_2 - y_1 - (y_1 - y_0) = y_2 - 2y_1 - y_0$.

Higer Order Backward Differnces

Here, $\nabla^3 y_3 = \nabla^2 y_3 - \nabla^2 y_2 = \nabla y_3 - \nabla y_2 - (\nabla y_2 - \nabla y_1) = ... = y_3 - 3y_2 + 3y_1 - y_0$.
The coefficients occuring are the **binomial coefficients**.

4.2.3 Central Differences

The central differnce $\delta y_{i+1/2}$ is defined as $\delta y_{i+1/2} = y_{i+1} - y_i$, where δ is called the *central difference operator*. We have,

$$\delta y_{1/2} = y_1 - y_0, \quad \delta y_{3/2} = y_2 - y_1, \quad ..., \quad \delta y_{(2n-1)/2} = y_n - y_{n-1} \tag{4.5}$$

These differnces $\delta y_{i+1/2}$, $i = 0, 1, ..., (n-1)$ are called *first central differences*.

- For $(n+1)$ points, there are n first central differnces.
- These n-central first differnces begin at $\delta y_{1/2}$, and end at $\delta y_{(2n-1)/2}$.

Second Central Differnces

The central differences of the first central differences are called second central differences. They are denoted by $\delta^2 y_i$, $i = 1, ..., (n-1)$ $\delta^2 y_i = \delta y_{i+1/2} - \delta y_{(i-1)+1/2}$. That is,

$$\delta^2 y_1 = \delta y_{3/2} - \delta y_{1/2}, \quad \delta^2 y_2 = \delta y_{5/2} - \delta y_{3/2}, \quad ..., \quad \delta^2 y_{n-1} = \delta y_{(2n-1)/2} - \delta y_{(2n-3)/2} \tag{4.6}$$

Here, $\delta^2 y_1 = \delta y_{3/2} - \delta y_{1/2} = y_2 - y_1 - (y_1 - y_0) = y_2 - 2y_1 - y_0$.

Higer Order Central Differnces

Here, $\delta^3 y_{3/2} = \delta^2 y_2 - \delta^2 y_1 = \delta y_{5/2} - \delta y_{3/2} - (\delta y_{3/2} - \delta y_{1/2}) = ... = y_3 - 3y_2 + 3y_1 - y_0$.

4.2.4 Equivallency of the differences

From the above subsections we can see that, the same numbers occurs at the same positions whether we use forward backward or central differences. That is,

$$\Delta y_0 = \nabla y_1 = \delta y_{1/2}, \qquad \Delta^2 y_0 = \nabla^2 y_1 = \delta^2 y_1, \qquad \Delta^3 y_0 = \nabla^3 y_3 = \delta^3 y_{3/2}$$

4.3 Newton's formula for Interpolation

Given $(n+1)$ points, (x_0, y_0), (x_1, y_1), ..., (x_n, y_n) of x and $y(x)$, with x values equally spaced, we find a polynomial $y_n(x)$ of nth degree such that, $y(x)$ and $y_n(x)$ agree at the given points. If we consider $y_n(x)$ to be the polynomial:

$$\begin{aligned} y_n x(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + a_3(x - x_0)(x - x_1)(x - x_2) \\ +... + a_n(x - x_0)(x - x_1)...(x - x_{n-1}). \end{aligned} \tag{4.7}$$

Then, $a_0 = y_0, \quad a_1 = \dfrac{y - y_1}{x_1 - x_0} = \dfrac{\Delta y_0}{h}, \quad a_2 = \dfrac{\Delta^2 y_0}{2!\, h^2}, \quad a_3 = \dfrac{\Delta^3 y_0}{3!\, h^3}, \quad ..., \quad a_n = \dfrac{\Delta^n y_0}{n!\, h^n}$

where, h is the $x_i - x_j, \quad i \neq j$.
Setting $x = x_0 + ph$, we get,

$$\begin{aligned} y_n(x) = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!}\Delta^2 y_0 + \frac{p(p-1)(p-2)}{3!}\Delta^3 y_0 + ... \\ + \frac{p(p-1)(p-2)...(p-(n-1))}{n!}\Delta^n y_0 \end{aligned} \tag{4.8}$$

This formula 4.8 is the **Newton's forward difference interpolation** formula.

4.3.1 Newton's backward difference Interpolation

If we consider $y_n(x)$ to be the polynomial:

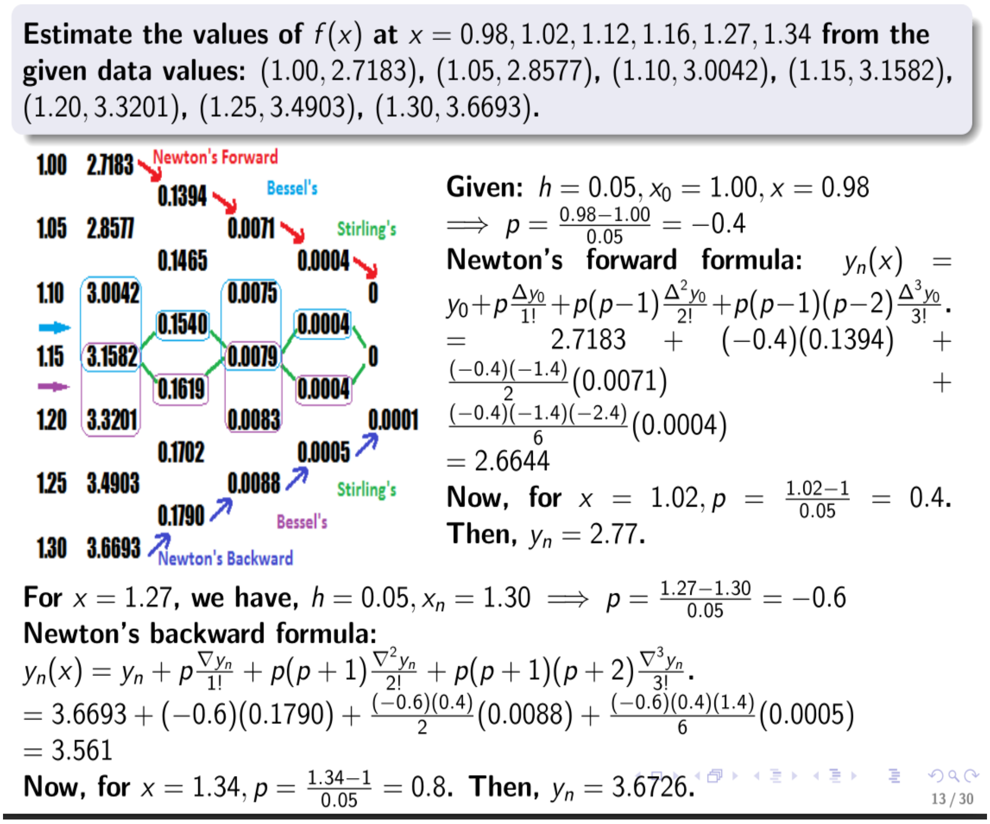
$$\begin{aligned} y_n x(x) &= a_0 + a_1(x - x_n) + a_2(x - x_n)(x - x_{n-1}) + a_3(x - x_n)(x - x_{n-1})(x - x_{n-2}) \\ &\quad + ... + a_n(x - x_n)(x - x_{n-1})(x - x_{n-2})...(x - x_1). \end{aligned} \tag{4.9}$$

Then we get the following formula for $y_n(x)$.

$$\begin{aligned} y_n(x) &= y_n + p\nabla y_n + \frac{p(p+1)}{2!}\nabla^2 y_n + \frac{p(p+1)(p+2)}{3!}\nabla^3 y_n + ... \\ &\quad + \frac{p(p+1)(p+2)...(p+(n-1))}{n!}\nabla^n y_n \end{aligned} \tag{4.10}$$

This formula 4.10 is the **Newton's backward difference interpolation** formula.

Remark. The Newton's forward difference interpolation formula is useful for interpolating near the begining of the tabular values and the Newton's backward difference interpolation formula is useful for interpolating near the end of the tabular values.



1. Find the cubic polynomial which takes the following values: $y(1) = 24$, $y(3) = 120$, $y(5) = 336$, $y(7) = 720$. Then obtain the value of $y(8)$.

2. Find a cubic polynomial which fits the data: $(-2, -12)$ $(-1, -8)$, $(2, 3)$, $(3, 5)$.

4.4 Interpolation with Unevenly spaced points

This section discusses the inerpolation method for unequally spaced values of the argument.

4.4.1 Lagrange's Interpolation Formula

Let $y(x)$ be continuous and differentiable $(n+1)$ times in the interval (a,b). Given $(n+1)$ points, (x_0, y_0), (x_1, y_1), ..., (x_n, y_n) of x and $y(x)$, where x values not necessarily equally spaced, we find a polynomial $L_n(x)$ of nth degree such that, $y(x)$ and $L_n(x)$ agree at the given points.
The formula $\sum_{i=0}^{n} l_i(x)\, y_i$, where $l_i(x) = \dfrac{(x-x_0)(x-x_1)...(x-x_{i-1})(x-x_{i+1})...(x-x_n)}{(x_i-x_0)(x_i-x_1)...(x_i-x_{i-1})(x_i-x_{i+1})...(x_i-x_n)}$ is called the Lagrange's Interpolating formula.

Example

Estimate the value of $f(301)$ from the following data values:
$(300, 2.4771), (304, 2.4829), (305, 2.4843), (307, 2.4871)$.

Since, we have 4 data points, Using Lagrange's third degree polynomial,

$$L_3(x) = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}y_0 + \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}y_1$$
$$+\frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}y_2 + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}y_3$$
$$L_3(301) = \frac{(-3)(-4)(-6)}{(-4)(-5)(-7)}2.4771 + \frac{(1)(-4)(-6)}{(4)(-1)(-3)}2.4829$$
$$+\frac{(1)(-3)(-6)}{(5)(1)(-2)}2.4843 + \frac{(1)(-3)(-4)}{(7)(3)(2)}2.4871$$
$$= 1.2739 + 4.9658 - 4.4717 + 0.7106$$
$$= 2.4786$$

Exercise

1. Estimate the value of $f(0.7)$ from the following data values:
$(0, 1), (0.4, 1.8556), (0.9, 2.5868), (1.2, 2.1786), (1.5, 0.4167)$
2. Find $f(x)$ as a polynomial in x from the following data:
$(0, -12), (1, 0), (3, 12), (4, 24)$

4.4.2 Disadvantage of Lagrange's Interpolation

The Lagrange's Interpolation formula has the disadvantage that if another interpolation point were added, then interpolation coefficients $l_i(x)$ will have to be recomputed. We therefore see an interpolation polynomial which has the property that a polynomial of higer degree may be derived from it by simply adding new terms. Newton's general interpolation formula is one such formula.

4.5 Divided Differences

Given $(n+1)$ points, (x_0, y_0), (x_1, y_1), ..., (x_n, y_n) of x and $y(x)$, with x values equally spaced, the divided differences of order $1, 2, ..., n$ are defined by the relations:

$$[x_0, x_1] = \frac{y_1 - y_0}{x_1 - x_0}$$

$$[x_0, x_1, x_2] = \frac{[x_1, x_2] - [x_0, x_1]}{x_2 - x_0}$$

$$\cdots$$

$$[x_0, x_1, x_2, ..., x_n] = \frac{[x_1, x_2, ..., x_n] - [x_0, x_1, ..., x_{n-1}]}{x_n - x_0}$$

It is easy to see that, $[x_0, x_1] = \dfrac{y_1 - y_0}{x_1 - x_0} = \dfrac{y_0}{x_0 - x_1} + \dfrac{y_1}{x_1 - x_0} = [x_1, x_0]$. Again,

$$\begin{aligned}
[x_0, x_1, x_2] &= \frac{[x_2, x_0] - [x_0, x_1]}{x_2 - x_0} \\
&= \frac{1}{x_2 - x_0}\left[\frac{y_2 - y_1}{x_2 - x_1} - \frac{y_1 - y_0}{x_1 - x_0}\right] \\
&= \frac{1}{x_2 - x_0}\left[\frac{y_2}{x_2 - x_1} - y_1\left(\frac{1}{x_2 - x_1} + \frac{1}{x_1 - x_0}\right) + \frac{y_0}{x_1 - x_0}\right] \\
&= \frac{y_0}{(x_0 - x_1)(x_0 - x_2)} + \frac{y_1}{(x_1 - x_0)(x_1 - x_2)} + \frac{y_2}{(x_2 - x_0)(x_2 - x_1)}
\end{aligned}$$

$$\begin{array}{llllll}
x_0 & y_0 & & & & \\
 & & [x_0,x_1] = \dfrac{y_1 - y_0}{x_1 - x_0} & & & \\
x_1 & y_1 & & [x_0,x_1,x_2] = \dfrac{[x_1,x_2] - [x_0,x_1]}{x_2 - x_0} & & \\
 & & [x_1,x_2] = \dfrac{y_2 - y_1}{x_2 - x_1} & & [x_0,x_1,x_2,x_3] = \dfrac{[x_1,x_2,x_3] - [x_0,x_1,x_2]}{x_3 - x_0} & \\
x_2 & y_2 & & [x_1,x_2,x_3] = \dfrac{[x_2,x_3] - [x_1,x_2]}{x_3 - x_1} & & \\
 & & [x_2,x_3] = \dfrac{y_3 - y_2}{x_3 - x_2} & & & \\
x_3 & y_3 & & & &
\end{array}$$

4.5.1 Newton's General Interpolation Formula

We have, $[x, x_0] = \dfrac{y - y_0}{x - x_0} \implies y = y_0 + (x - x_0)[x, x_0]$

And $[x, x_0, x_1] = \dfrac{[x, x_0] - [x_0, x_1]}{x - x_1} \implies [x, x_0] = [x_0, x_1] + (x - x_1)[x, x_0, x_1]$

From these two relations we get,

$$y = y_0 + (x - x_0)[x_0, x_1] + (x - x_0)(x - x_1)[x, x_0, x_1] \tag{4.11}$$

Similarly we can get $[x, x_0, x_1] = [x_0, x_1, x_2] + (x - x_2)[x, x_0, x_1, x_2]$ and so on... Proceeding this way we get,

$$
\begin{aligned}
y = y_0 + (x - x_0)[x_0, x_1] + (x - x_0)(x - x_1)[x_0, x_1, x_2] + \\
(x - x_0)(x - x_1)(x - x_2)[x_0, x_1, x_2, x_3] + \ldots \\
+(x - x_0)(x - x_1)(x - x_2)\ldots(x - x_n)[x, x_0, x_1, \ldots, x_n]
\end{aligned}
\tag{4.12}
$$

This Equation 4.12 is called the Newton's general interpolation formula with divided differences.

Example

Estimate the value of $f(301)$ from the following data values:
$(300, 2.4771), (304, 2.4829), (305, 2.4843), (307, 2.4871)$.

The divided difference table is

300	2.4771		
		0.00145	
304	2.4829		−0.00001
		0.00140	
305	2.4843		0
		0.00140	
307	2.4871		

From Newton's divided difference formula, we have

$y = y_0 + (x - x_0)[x_0, x_1] + (x - x_0)(x - x_1)[x_0, x_1, x_2] + (x - x_0)(x - x_1)(x - x_2)[x_0, x_1, x_2, x_3] + \cdots$

$f(301) = 2.4771 + (301 - 300)(0.00145) + (301 - 300)(301 - 304)(-0.00001)$

$f(301) = 2.4786$

Exercise

1. Estimate the value of $f(0.7)$ from the following data values:
$(0, 1), (0.4, 1.8556), (0.9, 2.5868), (1.2, 2.1786), (1.5, 0.4167)$
2. Find $f(x)$ as a polynomial in x from the following data:
$(-1, 3), (0, -6), (3, 39), (6, 822), (7, 1611)$

4.6 Inverse Interpolation

Given a set of values of x and y, the process of finding the value of x for a certain value of y is called inverse inerpolation.

- When the values of x are at unequal intervals, the most obvious way of performing this process is by interchanging x and y in Lagrange's method.
- When the values of x are equally spaced, the method of successive approximations on Newton's method should be used.

Chapter 4 Cont.. Spline Functions

4.7 Introduction

Earlier we discussed the methods of finding an nth order polynomial passing through $(n+1)$ points. In certain cases, it was found that these polynomials give **erroneous** result. Furthermore, it was found that a low-order polynomial approximation in each sub-interval provides a better approximation to the function than fitting a single high-order polynomial to the entire interval. Such piece-wise connecting polynomials are called **spline functions**. The points at which two connecting splines meet are called knots.
The **cubic** spline is the most popular in engineering applications. Before starting cubic splines, we discuss linear and quadratic splines since such a discussion will eventually justify the development of cubic splines.

4.7.1 Linear Splines

Suppose the given data points are $(x_i, y_i),\ i = 0, 1, ..., n$ and $h_i = x_i - x_{i-1},\ i = 1, 2, ..., n$. Let $s_i(x)$ be a straight line from x_{i-1} to x_i. Then, the slope of $s_i(x)$ is
$m_i = \dfrac{y_i - y_{i-1}}{x_i - x_{i-1}}$ and $s_i(x) = y_{i-1} + m_i(x - x_{i-1})$.
From the discussion above the $s_i(x)$ are the **linear splines**.

Drawback

The linear splines derived above are continuous in $[x_0, x_n]$, but their slopes are discontinuous, i.e their first derivatives are discontinuous.

4.7.2 Quadratic Splines

Let $s_i(x)$ be a quadratic approximation of the data points in the sub-interval $[x_{i-1} - x_i]$ satisfying the following conditions:

1. $s_i(x)$ and $s_i'(x)$ are continuous on $[x_0, x_n]$,

2. $s_i(x_i) = y_i, \qquad i = 0,\ 1,\ 2,\ ...,\ n$

3. $s_i'(x) = \frac{1}{h_i}[(x_i - x)m_{i-1} + (x - x_{i-1})m_i]$ as $s_i'(x)$ is linear, where $m_i = s_i'(x)$.

Integrating $s_i'(x)$ with respect to x, we obtain

$$s_i(x) = \frac{1}{h_i}\left[-\frac{(x_i - x)^2}{2}m_{i-1} + \frac{(x - x_{i-1})^2}{2}m_i\right] + c_i \tag{4.13}$$

Putting $x = x_{i-1}$ we get,

$$c_i = y_{i-1} + \frac{h_i}{2}m_{i-1} \tag{4.14}$$

Imposing the continuity condition on the spline functions $s_i(x)$ we get,

$$m_{i-1} + m_i = \frac{2}{h_i}(y_i - y_{i-1}), \qquad i = 1,\ 2,\ ...,\ n \tag{4.15}$$

Imposing the natural spline condition $s_1''(x_1) = 0$ we obtain, $m_0 = m_1$

Drawbacks

The second derivative of the quadratic splines derived above are discontinuous which is an obvious disadvantages. This drawback is removed in the cubic splines.

4.8 Cubic Splines

When computers were not available, the draftsman used a device to draw a smooth curve through a given set of points such that the slope and the curvature are also continuous along the curve, that is $f(x),\ f'(x),\ f''(x)$ are continuous on the curve. Such a device was called a **spline** and plotting of the curve was called **spline fitting**.

Let $s_i(x)$ be a cubic approximation of the data points in the sub-interval $[x_{i-1} - x_i]$ satisfying the following conditions:

1. $s_i(x)$ is atmost a cubic for $i = 1,\ 2,\ ...,\ n$,
2. $s_i(x)$, $s_i'(x)$ and $s_i''(x)$ are continuous on $[x_0, x_n]$,
3. $s_i(x) = y_i, \qquad i = 0,\ 1,\ 2,\ ...,\ n$
4. $\mathbf{s_i''(x) = \frac{1}{h_i}[(x_i - x)M_{i-1} + (x - x_{i-1})M_i]}$ as $s_i''(x)$ is linear, where $M_i = s_i''(x)$.

Integrating the condition-4, twice with respect to x: we get,

$$s_i(x) = \frac{1}{h_i}\left[-\frac{(x_i - x)^3}{6}M_{i-1} + \frac{(x - x_{i-1})^3}{6}M_i\right] + c_i(x_i - x) + d_i(x - x_{i-1}) \tag{4.16}$$

Using the condition: $s_i(x_{i-1}) = y_{i-1}$ and $s_i(x_i) = y_i$

$$c_i = \frac{1}{h_i}\left[y_{i-1} - \frac{h_i^2}{6}M_{i-1}\right], \qquad d_i = \frac{1}{h_i}\left[y_i - \frac{h_i^2}{6}M_i\right] \tag{4.17}$$

Imposing all these conditions we get,

$$\boxed{\frac{h_i}{6}\,M_{i-1} + \frac{1}{3}(h_i + h_{i+1})\,M_i + \frac{h_{i+1}}{6}\,M_{i+1} = \frac{y_{i+1} - y_i}{h_{i+1}} - \frac{y_i - y_{i-1}}{h_i}, \qquad (i = 1, 2, ..., n-1) \qquad (4.18)}$$

For subintervals of equal lengths:

$$M_{i-1} + 4M_i + M_{i+1} = \frac{6}{h^2}\,(y_{i+1} - 2y_i + y_{i-1}), \qquad (i = 1, 2, ..., n-1) \tag{4.19}$$

Imposing **natural spline** condition $s_i''(x_0) = s_i''(x_n) = 0$ in 4.18, we get,

$$2(h_1 + h_2)M_1 + h_2 M_2 = 6\left[\frac{y_2 - y_1}{h_2} - \frac{y_1 - y_0}{h_1} - h_1 M_0\right]$$
$$h_2 M_1 + 2(h_2 + h_3)M_2 + h_3 M_3 = 6\left[\frac{y_3 - y_2}{h_3} - \frac{y_2 - y_1}{h_2}\right]$$
$$h_3 M_2 + 2(h_3 + h_4)M_3 + h_4 M_4 = 6\left[\frac{y_4 - y_3}{h_4} - \frac{y_3 - y_2}{h_3}\right]$$
$$...$$
$$h_{n-1}M_{n-2} + 2(h_{n-1} + h_n)M_{n-1} = 6\left[\frac{y_n - y_{n-1}}{h_n} - \frac{y_{n-1} - y_{n-2}}{h_{n-1}}\right] - h_n M_n$$

This system is called **tridiagonal system** and there an efficient and an accurate method for solving it.

Example 1. Obtain the natural cubic spline approximation for the function defined by the data: $(0, 1)$, $(1, 2)$, $(2, 33), (3, 244)$. Hence find an estimate of $y(2.5)$.

Solution

For the equally x-spaced data we obtain:

$$M_0 + 4M_1 + M_2 = 6(y_2 - 2y_1 + y_0) \tag{4.20}$$
$$M_1 + 4M_2 + M_3 = 6(y_3 - 2y_2 + y_1) \tag{4.21}$$

Using $M_0 = 0 = M_3$, we get, $4M_1 + M_2 = 180, \quad M_1 + 4M_2 = 1080$. Then,

$$s_3(x) = \frac{1}{h_3}\left[-\frac{(x_3 - x)^3}{6}\,M_2 + \frac{(x - x_2)^3}{6}\,M_3\right] + \frac{1}{h_3}\left[y_2 - \frac{h_3^2}{6}M_2\right](x_3 - x) + \frac{1}{h_3}\left[y_3 - \frac{h_3^2}{6}M_3\right](x - x_2)$$
$$= -46x^3 + 414x^2 - 985x + 725$$
$$s(2.5) = -46(2.5)^3 = 414(2.5)^2 - 982(2.5) + 715 = 121.25$$

4.8.1 Exercise

1. For the data points: $(0, 0)$, $(\pi/2, 1)$, $(\pi, 0)$, determine the following:

   (a) natural quadratic splines

   (b) natural cubic splines

(c) $y(\pi/6)$ using natural cubic spline

2. Determine $y(\pi/6)$ using the natural cubic splines from the data points:
$(0, 0)$, $(\pi/4, 1/\sqrt{2})$, $(\pi/2, 1)$, $(3\pi/4, 1/\sqrt{2})$, $(\pi, 0)$.

3. What do you understand by natural spline? Explain why the natural cubic spline condition have $M_0 = 0$ and $M_n = 0$.

5

Least Square Problems

5.1 Introduction

Let the set of data points be (x_i, y_i), $\quad i = 1, 2, ..., m$, and let the curve given by $y = f(x)$ be fitted to this data. If e_i is the error of the approximation at $x = x_i$ due to this fitting then, $e_i = y_i - f(x_i)$. If we write $S = \sum_{i=1}^{m} e_i^2 = \sum_{i=1}^{m} [y_i - f(x_i)]^2$. Then the method of minimizing this error which is the sum of the square of errors is called the method of least squares.

5.1.1 Fitting a Straight Line

Let $Y = a_0 + a_1 x$ be the straight line to be fitted to the given data by the method of least squares. Then the sum of errors is $S = \sum_{i=1}^{m} [y_i - a_0 - a_1 x_i]^2$. S is a function of a_0 and a_1. To minimize S, $\frac{\partial S}{\partial a_0} = 0$ and $\frac{\partial S}{\partial a_1} = 0$. We have,

$$\frac{\partial S}{\partial a_0} = -2\sum_{i=1}^{m} [y_i - a_0 - a_1 x_i], \qquad \frac{\partial S}{\partial a_1} = -2x_i \sum_{i=1}^{m} [y_i - a_0 - a_1 x_i] \tag{5.1}$$

$$\frac{\partial S}{\partial a_0} = 0 \implies -2\sum_{i=1}^{m} [y_i - a_0 - a_1 x_i] = 0$$

$$\implies \sum_{i=1}^{m} y_i = \sum_{i=1}^{m} a_0 + a_1 \sum_{i=1}^{m} x_i$$

$$\implies \sum_{i=1}^{m} y_i = a_1 \sum_{i=1}^{m} x_i + m a_0 \tag{5.2}$$

$$\frac{\partial S}{\partial a_1} = 0 \implies -2x_i \sum_{i=1}^{m} [y_i - a_0 - a_1 x_i] = 0$$

$$\implies \sum_{i=1}^{m} x_i y_i = a_0 \sum_{i=1}^{m} x_i + a_1 \sum_{i=1}^{m} x_i^2 \tag{5.3}$$

The equations 5.2 and 5.3 are called normal equations which can be solved for a_0 and a_1. From the equation 5.2, it can be easily shown that the fitted line passes through the *means* of x_i and y_i,i.e the line satisfies $\overline{y} = a_0 + a_1 \overline{x}$.

5.1.2 Multiple Linear Least Square

Let $z = a_0+a_1x+a_2y$ be a linear equation to be fitted to the given data $(x_1, y_1, z_1), (x_2, y_2, z_2), ..., (x_m, y, z_m)$ by the method of least squares. Then the sum of errors is $S = \sum_{i=1}^{m} [z_i - a_0 - a_1x_i - a_2y_i]^2$. S is a function of a_0, a_1 and a_2. To minimize S:

$$\frac{\partial S}{\partial a_0} = -2\sum_{i=1}^{m} [y_i - a_0 - a_1x_i - a_2y_i] = 0$$

$$\frac{\partial S}{\partial a_1} = -2x_i\sum_{i=1}^{m} [y_i - a_0 - a_1x_i - a_2y_i] = 0$$

$$\frac{\partial S}{\partial a_2} = -2y_i\sum_{i=1}^{m} [y_i - a_0 - a_1x_i - a_2y_i] = 0$$

So, the normal equations are:

$$\begin{aligned}
\sum_{i=1}^{m} z_i &= a_1\sum_{i=1}^{m} x_i + a_2\sum_{i=1}^{m} y_i + ma_0 \\
\sum_{i=1}^{m} z_i\, x_i &= a_1\sum_{i=1}^{m} x_i^2 + a_2\sum_{i=1}^{m} x_iy_i + a_0\sum_{i=1}^{m} x_i \\
\sum_{i=1}^{m} z_i\, y_i &= a_1\sum_{i=1}^{m} x_iy_i + a_2\sum_{i=1}^{m} y_i^2 + a_0\sum_{i=1}^{m} y_i
\end{aligned} \tag{5.4}$$

5.2 Curve Fitting by Polynomials

Let $Y = a_0 + a_1x + a_2x^2 + ... + a_nx^n$ be a polynomial to be fitted to the given data $(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)$ by the method of least squares. Then the sum of errors is $S = \sum_{i=1}^{m} [z_i - a_0 - a_1x_i - a_2y_i]^2$. Then the normal equations are:

$$\begin{aligned}
\sum_{i=1}^{m} y_i &= a_1\sum_{i=1}^{m} x_i + a_2\sum_{i=1}^{m} x_i^2 + ... + ma_0 \\
\sum_{i=1}^{m} x_i\, y_i &= a_0\sum_{i=1}^{m} x_i + a_1\sum_{i=1}^{m} x_i^2 + a_2\sum_{i=1}^{m} x_i^3 + ... + a_n\sum_{i=1}^{m} x_i^{n+1} \\
&... \\
\sum_{i=1}^{m} x_i^n\, y_i &= a_0\sum_{i=1}^{m} x_i^n + a_1\sum_{i=1}^{m} x_i^{n+1} + a_2\sum_{i=1}^{m} x_i^{n+2} + ... + a_n\sum_{i=1}^{m} x_i^{2n}
\end{aligned} \tag{5.5}$$

This system constitutes of $(n + 1)$ equations in $(n + 1)$ unknowns. For larger value of n the system is *ill conditioned*, so orthogonal polynomials are used to fit such data points.

5.3 Weighted Least Square Approximation

For the errors are weighted by W_i, then the sum of weighed errors is:
$S = \sum_{i=1}^{m} W_i\, e_i^2 = \sum_{i=1}^{m} W_i\, [y_i - f(x_i)]^2$. So that the normal equations of the linear curve fitting $Y = a_0 + a_1 x$ becomes:

$$\begin{aligned} \sum_{i=1}^{m} W_i\, y_i &= a_1 \sum_{i=1}^{m} W_i\, x_i + m a_0 \\ \sum_{i=1}^{m} W_i\, x_i y_i &= a_0 \sum_{i=1}^{m} x W_{i\,i} + a_1 \sum_{i=1}^{m} W_i\, x_i^2 \end{aligned} \tag{5.6}$$

5.4 Linearization of Nonlinear Laws

1. $y = ax + \dfrac{b}{x}$
   This can be written as $xy = b + ax^2$. Put $xy = Y$, $b = A_0$, $a = A_1$ and $x^2 = X$. Then the given nonlinear equation becomes the linear equation $Y = A_0 + A_1 X$.

2. $y = ae^{bx}$
   This can be written as $xy = b + ax^2$. Put $lny = Y$, $lna = A_0$, $b = A_1$ and $x = X$. Then the given nonlinear equation becomes the linear equation $Y = A_0 + A_1 X$.

5.4.1 Exercise

1. Fit a linear equation to the data points with the given weights.
   $(0, -1)$, $(2, 5)$, $(5, 12)$, $(7, 20)$ and $W = (1, 1, 10, 1)$

2. Fit the a second-degree polynomial to the given data points:
   $(0, 71)$, $(1, 89)$, $(2, 67)$, $(3, 43)$, $(4, 31)$, $(5, 18)$, $(6, 9)$

3. Determine the normal equations for the cubic polynomial fitting.

4. What is the difference between interpolation and curve fitting for a given data points?

5. Convert the following nonlinear equations into suitable linear equations for the curve fitting using linear least square fitting.

   a). $xy^a = b$ b). $y = ab^x$ c). $y = ax^b$

6. Fit a function of the form $y = ax^b$ to the data points: $(61, 350)$, $(26, 400)$, $(7, 500)$, $(2.6, 600)$

6

Numerical Differentiation and Numerical Integration

In this chapter we shall be concerned with the problems of numerical differentiation and integration. We shall derive the formula to compute the following when only tabulated values of the function are known but the explicitly nature of the function is not known. Such scenario occurs in engineering in case of experimental data:

- $\dfrac{dy}{dx}, \dfrac{d^2y}{dx^2}, \ldots$ for any value of x in $[x_0, x_n]$, and
- $\displaystyle\int_{x_0}^{x_n} y\, dx$

6.1 Numerical Differentiation

The general method for deriving the numerical differentiation formulae is to **differentiate the interpolating polynomial**. Hence, corresponding to each of the Interpolating formula derived, we may derive a formula for the derivative.

6.1.1 Newton's forward difference formulae

The Newton's forward difference formula is:

$$\begin{aligned} y_n(x) = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!}\Delta^2 y_0 + \frac{p(p-1)(p-2)}{3!}\Delta^3 y_0 + \ldots \\ + \frac{p(p-1)(p-2)\ldots(p-(n-1))}{n!}\Delta^n y_0 \end{aligned} \tag{6.1}$$

where, $\mathbf{x} = \mathbf{x_0} + \mathbf{ph}$ and $h = x_{i+1} - x_i$. Differentiating 6.1 with respect to x,

$$\frac{dy}{dx} = \frac{dy}{dp}\frac{dp}{dx} = \frac{1}{h}\left[\Delta y_0 + \frac{2p-1}{2}\Delta^2 y_0 + \frac{3p^2-6p+2}{6}\Delta^3 y_0 + \ldots\right] \tag{6.2}$$

Differentiating 6.2 with respect to x, we get,

$$\frac{d^2y}{dx^2} = \frac{1}{h^2}\left[\Delta^2 y_0 + \frac{6p-6}{6}\Delta^3 y_0 + \frac{12p^2-36p+22}{24}\Delta^4 y_0 + \ldots\right] \tag{6.3}$$

These formulae are used for *non-tabular* values of x. For tabular values of x, the formulae take simpler form. For $x = x_0$, we have $p = 0$, and using this we can find the relations of $\left[\dfrac{dy}{dx}\right]_{x=x_0}$ *and* $\left[\dfrac{d^2y}{dx^2}\right]_{x=x_0}$ which is in a simpler form.

$$\left[\frac{dy}{dx}\right]_{x=x_n} = \frac{1}{h}\left[\Delta\, y_n - \frac{1}{2}\Delta^2\, y_n + \frac{1}{3}\Delta^3\, y_n - ...\right] \tag{6.4}$$

and

$$\left[\frac{d^2y}{dx^2}\right]_{x=x_n} = \frac{1}{h^2}\left[\Delta^2\, y_n - \Delta^3\, y_n + \frac{11}{12}\Delta^4\, y_n - \frac{5}{6}\Delta^5\, y_n + ...\right] \tag{6.5}$$

6.1.2 Newton's backward difference formula

In a similar way, different formulae can be derived by starting with other interpolation formulae. Thus, Newton's backward difference formula gives: where $\mathbf{x} = \mathbf{x_n} + \mathbf{uh}$

$$\left[\frac{dy}{dx}\right]_{x=x_n} = \frac{1}{h}\left[\nabla\, y_n + \frac{1}{2}\nabla^2\, y_n + \frac{1}{3}\nabla^3\, y_n + ...\right] \tag{6.6}$$

and

$$\left[\frac{d^2y}{dx^2}\right]_{x=x_n} = \frac{1}{h^2}\left[\nabla^2\, y_n + \nabla^3\, y_n + \frac{11}{12}\nabla^4\, y_n + \frac{5}{6}\nabla^5\, y_n + ...\right] \tag{6.7}$$

Remark. If the x-values of the data points are equally spaced then it is common to use Newton's interpolation.

6.1.3 Exercise

1. Find the dy/dx and d^2y/dx^2 at $x = 1.2,\ x = 1.6,\ x = 2.2$ from the data:

x	1	1.2	1.4	1.6	1.8
y	2.7183	3.3201	4.0522	4.9530	6.60496

2. The distance traversed by a particle at different times are given below. Find the velocity and the acceleration of the particle at $t = 0.1$ seconds.

t	0.0	0.1	0.2	0.3
x	3.01	3.16	3.29	3.36

3. From the given data points (x, y), find $\dfrac{dy}{dx}$ at $x = 2$ using the natural cubic spline method. $(2, 11),\ (3, 49),\ (4, 123)$.

6.2 Numerical Integration

The general problem of numerical integration may be stated as: Given a set of data points of a function $y = f(x)$, where $f(x)$ is not known explicitly, it is required compute the value of the definite integral

$$I = \int_a^b y\, dx$$

As with the case of numerical differentiation, one replaces $f(x)$ by an interpolating polynomial $\phi(x)$ and obtains, on integration, an approximate value of the definite integral. Thus, different integration formulae can be obtained depending upon the type of the interpolation formula used. Approximating y by Newton's forward difference formula, we obtained

$$I = \int_{x_0}^{x_n} \left[y_0 + p\Delta y_0 + \frac{p(p-1)}{2}\,\Delta^2 y_0 + \frac{p(p-1)(p-2)}{6}\,\Delta^3 y_0 + ...\right] dx$$

Since $x = x_0 + ph$, $dx = hdp$. When $x = x_0$, $p = 0$ and $x = x_n$, $p = n$. Then,

$$I = \int_{0}^{n} \left[y_0 + p\Delta y_0 + \frac{p(p-1)}{2}\,\Delta^2 y_0 + \frac{p(p-1)(p-2)}{6}\,\Delta^3 y_0 + ...\right] h\,dp$$

Integrating we get,

$$\begin{aligned} I &= h\left[py_0 + \frac{p^2}{2}\,\Delta y_0 + \left(\frac{p^3}{3} - \frac{p^2}{2}\right)\frac{\Delta^2 y_0}{2} + \left(\frac{p^4}{4} - p^3 + p^2\right)\frac{\Delta^3\, y_0}{6} + ...\right]_0^n \\ &= nh\left[y_0 + \frac{n}{2}\,\Delta y_0 + \left(\frac{2n^2 - 3n}{12}\right)\Delta^2\, y_0 + \left(\frac{n^3 - 4n^2 + 4n}{24}\right)\Delta^3\, y_0 + ...\right] \qquad (6.8) \end{aligned}$$

This relation 6.8 is considered to be a general formula in the variable n. For a particular value of n we get a particular formula. For example for $n = 1$ we get the famous Trapezoidal rule and for $n = 2$ and $n = 3$ we get Simpson's 1/3 rule and Simpson's 3/8 rule respectively.

6.2.1 Trapezoidal Rule

Setting $n = 1$ in the general formula 6.8, all differences higher than the first order will become zero and we obtained

$\int_{x_0}^{x_1} y\,dx = h\left[y_0 + \frac{1}{2}\Delta y_0\right] = h\left[y_0 + \frac{1}{2}(y_1 - y_0)\right] = \frac{h}{2}\,(y_0 + y_1)$. Similarly we can obtain

$\int_{x_1}^{x_2} y\,dx, ... \int_{x_{n-1}}^{n} y\,dx$. Then we have,

$$\begin{aligned} \int_{x_0}^{x_n} y\,dx &= \int_{x_0}^{x_1} y\,dx + \int_{x_1}^{x_2} y\,dx + ... + \int_{x_{n-1}}^{x_n} y\,dx \\ &= \frac{h}{2}\,(y_0 + y_1) + \frac{h}{2}\,(y_1 + y_2) + ... + \frac{h}{2}\,(y_{n-1} + y_n) \\ &= \frac{h}{2}\,\{y_0 + 2(y_1 + y_2 + ... + y_{n-1}) + y_n\} \qquad (6.9) \end{aligned}$$

Question: Illustrate the trapezoidal method geometrically.

6.2.2 Simpson's 1/3-Rule

Setting $n = 2$ in the general formula 6.8, all differences higher than the second will become zero and we obtained

$\int_{x_0}^{x_2} y\,dx = 2h\left[y_0 + \Delta\, y_0 + \frac{1}{6}\Delta^2\, y_0\right] = \frac{h}{3}\,(y_0 + 4y_1 + y_2)$. Similarly we can obtain

$\int_{x_2}^{x_4} y\,dx, ... \int_{x_{n-2}}^{n} y\,dx$. Then we have,

$$\int_{x_0}^{x_n} y\,dx = \int_{x_0}^{x_2} y\,dx + \int_{x_2}^{x_4} y\,dx + ... + \int_{x_{n-2}}^{x_n} y\,dx$$
$$= \frac{h}{3}(y_0 + 4y_1 + y_2) + \frac{h}{3}(y_2 + 4y_3 + y_4) + ... + \frac{h}{3}(y_{n-2} + 4y_{n-1} + y_n)$$
$$= \frac{h}{3}\{y_0 + 4(y_1 + y_3 + ... + y_{n-1}) + 2(y_2 + y_4 ... + y_{n-2}) + y_n\} \tag{6.10}$$

Remark. It should be noted that this rule requires the division of the whole range into a number of sub-intervals that is a multiple of 2, of width h.

6.2.3 Simpson's 3/8-Rule

Setting $n = 3$ in the general formula 6.8, all differences higher than the third will become zero and we obtained

$$\int_{x_0}^{x_3} y\,dx = 3h\left[y_0 + \frac{3}{2}\Delta\, y_0 + \frac{3}{4}\Delta^2\, y_0 + \frac{1}{8}\Delta^3\, y_0\right] = \frac{3h}{8}(y_0 + 3y_1 + 3y_2 + y_3).$$

Similarly we can obtain

$\int_{x_3}^{x_6} y\,dx, ... \int_{x_{n-3}}^{n} y\,dx$. Then we have,

$$\int_{x_0}^{x_n} y\,dx = \int_{x_0}^{x_3} y\,dx + \int_{x_3}^{x_6} y\,dx + ... + \int_{x_{n-3}}^{x_n} y\,dx$$
$$= \frac{3h}{8}(y_0 + 3y_1 + 3y_2 + y_3) + \frac{3h}{8}(y_3 + 3y_4 + 3y_5 + y_6) + ... + \frac{3h}{8}(y_{n-3} + 3y_{n-2} + 3y_{n-1} + y_n)$$
$$= \frac{3h}{8}\{y_0 + 3(y_1 + y_2 + y_4 + y_5 + ... + y_{n-1}) + 2(y_3 + y_6 ... + y_{n-3}) + y_n\} \tag{6.11}$$

Remark. It should be noted that this rule requires the division of the whole range into a number of sub-intervals that is a multiple of 3, of width h.

6.2.4 Exercise

1. For the given (x, y) data points: Evaluate $\int_0^2 ydx$ using the trapezoidal rule. $(0.0, 0.399)$, $(0.5, 0.352)$, $(1.0, 0.242)$, $(1.5, 0.129)$, $(2.0, 0.054)$.

2. Evaluate using Simpson's 1/3 rule:

   (a) $\displaystyle\int_1^3 \frac{1}{x}\,dx$ with 8 stripes.

   (b) $\displaystyle\int_0^{\pi/2} \sqrt{sinx}\,dx$ with $h = \pi/12$.

   (c) $\displaystyle\int_3^7 x^2 logx\,dx$ with $h = 1$.

   (d) $\displaystyle\int_0^1 e^{-x^2}\,dx$ with $h = 0.1$

3. Evaluate using Simpson's 3/8 rule: $\displaystyle\int_0^1 \frac{dx}{1+x}$ with $h = 1/6$.

6.2.5 Boole's and Weddle's Rules, *Extra material*

To incorporate the fourth order difference as well we need 5 data points $x_0, ..., x_4$ and we need to integrate from x_0 to x_4. Then we obtain Boole's formula:

$$\int_{x_0}^{x_4} y\,dx = \frac{2h}{45}\left(7y_0 + 32y_1 + 12y_2 + 32y_3 + 7y_4\right) \tag{6.12}$$

If we incorporate differences u to six order then we obtain Weddle's formula as follows:

$$\int_{x_0}^{x_6} y\,dx = \frac{3h}{10}\left(y_0 + 5y_1 + y_2 + 6y_3 + y_4 + 5y_5 + y_6\right) \tag{6.13}$$

6.2.6 Romberg Integration

This method improves the approximation obtained from the other methods. Here we describe the Romberge method to improve the results obtained from the Trapezoidal method. Suppose the integral $\int_a^b y\,dx$ is evaluated by the Trapezoidal rule with two different subintervals of widths h_1 and h_2, to obtain the approximate values I_1 and I_2, respectively. Romberg's method uses this two approximations: I_1 and I_2 to obtain the next better approximations which is given by

$$I_3 = \frac{I_1 h_2^2 - I_2 h_1^2}{h_2^2 - h_1^2} \tag{6.14}$$

6.2.7 Gaussian Quadrature

The methods in the previous sections are based upon the values of function a equally-spaced points of the interval. Gauss derived a formula for which the points in the interval **need not be equally spaced**. The Gaussian formula is expressed as:

$$\int_{-1}^{1} F(u)\,du = \sum_{i=1}^{n} W_i\,F(u_i) \tag{6.15}$$

where W_i and u_i are called the *weights* and *abscissae*, respectively. It can be shown that the u_i are the zeros of the $(n+1)^{th}$ **Legendre** polynomial $P_{n+1}(u)$ which can be generated usingthe recurrence relation:

$$(n+1)\,P_{n+1}(u) = (2n+1)\,u\,P_n(u) - n\,P_{n-1}(u)$$

where $P_0(u) = 1$ and $P_1(u) = u$. It can also be shown that corresponding weights W_i are given by

$$W_i = \int_{-1}^{1} \prod_{j=0, j\neq i}^{n} \frac{u - u_j}{u_i - u_j}\,du$$

However the weights and the abscissae are extensively tabulated for different values of n. We list the weights and the abscissae for values of n up to $n = 6$.

Change of Limits In the general case, the limits of the integral have to be changed to that of 6.15 by means of the following variable transformation:

$$x = \frac{1}{2}\,u(b-a) + \frac{1}{2}\,(a+b)$$

Table 5.1 Abscissae and Weights for Gaussian Integration

n	$\pm u_i$	W_i
2	0.57735 02692	1.0
3	0.0	0.88888 88889
	0.77459 66692	0.55555 55556
4	0.33998 10436	0.65214 51549
	0.86113 63116	0.34785 48451
5	0.0	0.56888 88889
	0.53846 93101	0.47862 86705
	0.90617 98459	0.23692 68851
6	0.23861 91861	0.46791 39346
	0.66120 93865	0.36076 15730
	0.93246 95142	0.17132 44924

6.2.8 Exercise

1. Compute the values of $I = \int_0^1 \frac{dx}{1+x^2}$ using the trapezoidal rule with $h = 0.5,\ 0.25,\ 0.125$. Then obtain a better estimates using **Romberg's method**.

2. Compute the values of $I = \int_0^{\pi/2} cos^2 x\, dx$ using the trapezoidal rule with $h = \pi/4,\ \pi/8$. Then obtain a better estimates using **Romberg's method**.

3. Derive the Gauss quadrature formula for $n = 2$ and apply it to evaluate the integral $\int_{-1}^1 \frac{dx}{1+x^2}$.

4. Use the three point Gauss formula to evaluate the integral $\int_0^1 \frac{dx}{1+x}$. Compare this result with that obtained by Simpson's 1/3-rule with $h = 0.125$.

6.3 Double Integration

Formulae for the evaluation of double integral can be obtained by repeatedly applying the trapezoidal and Simpson's rules derived earlier.

6.3.1 Trapezoidal Rule

$$I = \int_{y_j}^{y_{j+1}} \int_{x_i}^{x_{i+1}} f(x, y)\, dx\, dy$$

where, $x_{i+1} = x_i + h$ and $y_{j+1} = y_j + k$. By repeated application of trapezoidal rule to we get,

$$\begin{aligned} I &= \frac{h}{2}\int_{y_j}^{y_{j+1}} [f(x_i, y) + f(x_{i+1}, y)]\, dx\, dy \\ &= \frac{hk}{4}[f(x_i, y_j) + f(x_{i+1}, y_j) + f(x_i, y_{j+1}) + f(x_{i+1}, y_{j+1})]\, dx\, dy \\ &= \frac{hk}{4}[f_{i,j} + f_{i+1,j} + f_{i,j+1} + f_{i+1,j+1}]\, dx\, dy \end{aligned} \tag{6.16}$$

6.3.2 Simpson's Rule

$$I = \int_{y_{j-1}}^{y_{j+1}} \int_{x_{i-1}}^{x_{i+1}} f(x, y)\, dx\, dy$$

where, $x_{i+1} = x_i + h$ and $y_{j+1} = y_j + k$. By repeated application of Simpson's rule to we get,

$$\begin{aligned} I &= \frac{h}{3}\int_{y_{j-1}}^{y_{j+1}} [f(x_{i-1}, y) + 4f(x_i, y) + f(x_{i+1}, y)]\, dx\, dy \\ &= \frac{hk}{9}[f(x_{i-1}, y_{j-1}) + 4f(x_{i-1}, y_j) + f(x_{i-1}, y_{j+1}) \\ &\quad 4\{f(x_i, y_{j-1}) + 4f(x_i, y_j) + f(x_i, y_{j+1})\} + f(x_{i+1}, y_{j-1}) + 4f(x_{i+1}, y_j) + f(x_{i+1}, y_{j+1})] \end{aligned} \tag{6.17}$$

Numerical double integration

Evaluate:
$I = \int_a^b \int_c^d f(x, y)dxdy = \int_0^2 \int_0^1 e^{x+y} dxdy$

Given: $a = 0, b = 2, c = 0, d = 1$, $f(x, y) = e^{x+y}$. **Let us choose** $n = 4$ **such that** $h = \frac{b-a}{n} = \frac{2-0}{4} = 0.5$ **and** $k = \frac{d-c}{n} = \frac{1-0}{4} = 0.25$ **The corresponding functional value table is:**

f(x,y)	0	0.5	1	1.5	2
0	1.0000	1.6487	2.7183	4.4817	7.3891
0.25	1.2840	2.1170	3.4903	5.7546	9.4877
0.5	1.6487	2.7183	4.4817	7.3891	12.1825
0.75	2.1170	3.4903	5.7546	9.4877	15.6426
1	2.7183	4.4817	7.3891	12.1825	20.0855

Now, the solution is
$I = \int_0^2 \int_0^1 e^{x+y} dxdy = \frac{h}{2}\frac{k}{2}$**[sum of all multiples of corresponding elements of functional value table and multiplication table]**

Multiplication table (Trapezoidal rule):

×	1	2	2	2	1
1	1	2	2	2	1
2	2	4	4	4	2
2	2	4	4	4	2
2	2	4	4	4	2
1	1	2	2	2	1

$\therefore I = \left(\frac{0.5}{2}\right)\left(\frac{0.25}{2}\right)[1 \times 1 + 1.6487 \times 2 + 2.7183 \times 2 + \cdots + 20.0855 \times 1] = 11.2643$

Multiplication table (Simpson's rule):

×	1	4	2	4	1
1	1	4	2	4	1
4	4	16	8	16	4
2	2	8	4	8	2
4	4	16	8	16	4
1	1	4	2	4	1

$\therefore I = \left(\frac{0.5}{3}\right)\left(\frac{0.25}{3}\right)[1 \times 1 + 1.6487 \times 4 + 2.7183 \times 2 + \cdots + 20.0855 \times 1] = 10.9821$

Exercise: Evaluate
$\int_{-2}^2 \int_0^4 (x^2 - xy + y^2)dxdy$

30 / 30

Questions: Using Trapezoidal rule and Simpson's 1/3 rule evaluate the double integral

$$\int_2^2 \int_0^4 (x^2 - xy + y^2)\, dx\, dy.$$

Example

Find, from the data, the area bounded by the curve and the x-axis:
$(7.47, 1.93), (7.48, 1.95), (7.49, 1.98), (7.50, 2.01), (7.51, 2.03), (7.52, 2.06)$

Solution: We know that: Area=$\int_{7.47}^{7.52} f(x)dx$

Here, $h = 0.01$**, total data points:** 6**, Total subintervals:** $n = 5$**.** $\therefore$ **we can use trapezoidal rule only.**

Trapezoidal rule: $\int_{x_0}^{x_n} ydx = \frac{h}{2}(y_0 + 2(y_1 + y_2 + \cdots + y_{n-1}) + y_n)$

Area=$\int_{7.47}^{7.52} f(x)dx = \frac{0.01}{2}[1.93 + 2(1.95 + 1.98 + 2.01 + 2.03) + 2.06] = 0.0996$

Exercise

Find the volume of solid of revolution formed by rotating about the x-axis the area between the $x - axis$**, the lines** $x = 0$ **and** $x = 1$**, and a curve through the points:**
$(0, 1), (0.25, 0.9896), (0.5, 0.9589), (0.75, 0.9089), (1, 0.8415)$.

Hint: Volume=$\pi \int_a^b y^2 dx$

Example

Integrate: $I = \int_0^{1.5} \frac{e^x + x}{\sin x + 1} dx$

Given, $a = 0, b = 1.5, f(x) = \frac{e^x+x}{\sin x+1}$**. Let us choose** $n = 6$**, so that** $h = \frac{b-a}{n} = \frac{1.5-0}{6} = 0.25$**. The functional values are:**

x	0	0.25	0.5	0.75	1	1.25	1.5
f(x)	1	1.2298	1.4524	1.7049	2.0192	2.4322	2.9946

Simpson's 1/3 rule:
$I = \frac{h}{3}(y_0 + 4(y_1 + y_3 + \cdots + y_{n-1}) + 2(y_2 + y_4 + \cdots + y_{n-2}) + y_n)$

$= \frac{0.25}{3}(1 + 4(1.2298 + 1.7049 + 2.4322) + 2(1.4524 + 2.0192) + 2.9946)$

$= 2.7004$

Simpson's 3/8 rule:
$I = \frac{3h}{8}(y_0 + 3y_1 + 3y_2 + 2y_3 + 3y_4 + 3y_5 + 2y_6 + \cdots + 2y_{n-3} + 3y_{n-2} + 3y_{n-1} + y_n)$

$I = \frac{3h}{8}(1 + 3(1.2298 + 1.4524 + 2.0192 + 2.4322) + 2(1.7049) + 2.9946)$
$= 2.7005$

Exercise

Evaluate: $I = \int_0^1 \frac{1}{1+x} dx$ **using** $h = 0.125$.

7 Solving Ordinary Differential Equations

7.1 Introduction

The analytical methods of solution can be applied to solve only a selected class of differential equations. Those equations which govern physical systems do not possess, in general **closed-form** solutions, and hence recourse must be made to numerical methods for solving such differential equations. To describe various numerical methods for the solution of ordinary differential equations, we consider the general first order initial value problem. The methods so developed can, in general, be applied to the solution of first order equations.

Initial Value Problem: The general first order initial value problem is of the form:

$$y' = f(x, y), \quad y(x_0) = y_0 \tag{7.1}$$

In general there are two classes of the methods for solving 7.1.

1. A series solution.
   (a) Taylor series method
   (b) Picard method
2. A set of tabulated values of x and y.
   (a) Euler's methods
   (b) Runge-Kutta methods

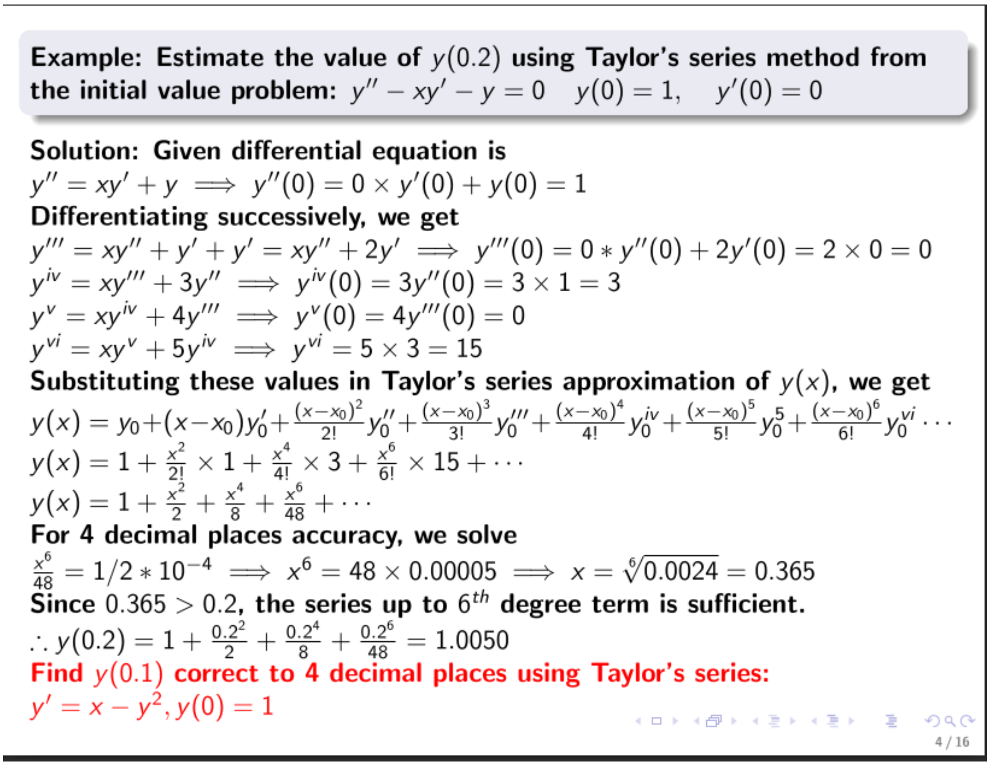
7.2 Taylor's series method

We consider the IVP 7.1. If $y(x)$ is the exact solution of this IVP, then the Taylor's series for $y(x)$ around $x = x_0$ is given by

$$y(x) = y_0 + (x - x_0)y_0' + \frac{(x - x_0)^2}{2!} y_0'' + \ldots \tag{7.2}$$

Differentiating 7.1 successively, with respect to x, we obtain, $y'', y''', \ldots$. Since 7.2 is an infinite series, for $y(x_0)$ correct to N decimal places we solve the following equation

n^{th} term of Taylor series $\leq$ Upper limiting error $\implies \dfrac{(x-x_0)^n}{n!} f^{(n)}(x_0) \leq \dfrac{1}{2}10^{-N}$.
If this holds then, the n number of terms is sufficient otherwise, we need to include additional terms.



7.3 Picard's Method of Successive Approximation

Integrating the above differential equation 7.1, as follows, we obtain:

$$\int_{x_0}^{x} y'\,dx = \int_{x_0}^{x} f(x,y)\,dx$$
$$y = y_0 + \int_{x_0}^{x} f(x,y)\,dx \tag{7.3}$$

This equation in which the unknown function y appears under the integral sign is called an *integral equation*. Such equation can be solved by the method of successive approximations as follows:

$$y^{(1)} = y_0 + \int_{x_0}^{x} f(x,y_0)\,dx \tag{7.4}$$

Proceeding in this way, we obtain:

$$y^{(n)} = y_0 + \int_{x_0}^{x} f(x,y^{(n-1)})\,dx \quad \text{with} \quad y^{(0)} = y_0 \tag{7.5}$$

It has been proved that if $f(x,y)$ is bounded in some region about the point (x_0,y_0) and if $f(x,y)$ satisfies the *Lipschitz condition*, viz.,

$$|f(x,y) - f(x,\overline{y})| \leq K|y - \overline{y}|$$

for some constant K, then the sequence $y^{(n)}$ converges to the solution of the equation 7.1.

Example: Solve the equation $y' = x + y^2, \quad y(0) = 1$ using Picard's method correct to 4 decimal places and estimate the value of $y(0.2)$

Given equation: $y' = x + y^2, \quad y(0) = 1$. **Integrating w.r.t.** x, **we get**
$y = 1 + \int_0^x (x + y^2)dx$
Approximating y on r.h.s by $y(0) = 1$, **we get**
$y_1 = 1 + \int_0^x (x+1)dx = 1 + [\frac{x^2}{2} + x]_0^x dx = 1 + \frac{x^2}{2} + x = 1 + x + \frac{x^2}{2}$
Testing for 4 decimal places accuracy using last term of y_1,
$\frac{x^2}{2} = 0.00005 \implies x = \sqrt{2 \times 0.00005} \implies x = 0.01$
$\therefore$ **The terms in** y_1 **is not sufficient.**
Again approximating y on r.h.s by y_1, **we get**
$y_2 = 1 + \int_0^x (x + (1 + x + \frac{x^2}{2})^2)dx = 1 + \int_0^x (x + (1 + x^2 + \frac{x^4}{4} + 2x + x^2 + x^3))$
$y_2 = 1 + \int_0^x (1 + 3x + 2x^2 + x^3 + \frac{x^4}{4}) = 1 + x + \frac{3}{2}x^2 + \frac{2}{3}x^3 + \frac{1}{4}x^4 + \frac{1}{20}x^5$
Testing for 4 decimal places accuracy using last term of y_2,
$\frac{1}{20}x^5 = 0.00005 \implies x = \sqrt[5]{20 \times 0.00005} \implies x = 0.2511$
Since $0.2511 > 0.2$, **The series** y_2 **is sufficient for 4 decimal places accuracy.**
Substituting $x = 0.2$ **in** y_2, **we get**
$y(0.2) = 1 + 0.2 + \frac{3}{2}0.2^2 + \frac{2}{3}0.2^3 + \frac{1}{4}0.2^4 + \frac{1}{20}0.2^5 = 1.2657$
Exercise: Find $y(0.2)$ using Picard's method correct to 4 decimal places:
$y' = x + yx^4, y(0) = 3.$

7.3.1 Exercise

1. Find by Taylor's series method:

   (a) the value of $y(0.1)$ given that $y' = x - y^2$, $y(0) = 1$.

   (b) the value of $y(0) = 1$ given that $\dfrac{dy}{dx} = 1 + xy$ correct up to four decimal places.

2. Use Picard's method to obtain:

   (a) $y(0.1)$ and $y(0.2)$ given that $\dfrac{dy}{dx} = x + yx^4, \quad y(0) = 3.$

   (b) $y(0.1)$ given that $\dfrac{dy}{dx} = \dfrac{y-x}{y+x}, \quad y(0) = 1.$

7.4 Euler's Method

We have so far discussed the methods which yield the solution of a differential equation in the form of a power series. We will now describe the methods which give the solution in the form of a set of tabulated values. The approximation of the IVP 7.1 by the following iteration formula is called Euler's method.

$$y_{n+1} = y_n + h\, f(x_n, y_n) \qquad n = 0, 1, 2, ... \tag{7.6}$$

This formula is obtain from the forward difference $\left[\dfrac{y_{n+1} - y_n}{h} = y_n'\right]$.

7.4.1 Backward Euler's Method

Using the backward difference $\left[\dfrac{y_{n+1} - y_n}{h} = y'_{n+1}\right]$; we obtain, Backward Euler's Method as follows:

$$y_{n+1} = y_n + h\, f(x_{n+1}, y_{n+1}) \qquad n = 0, 1, 2, ... \tag{7.7}$$

We obtain y_{n+1} in $f(x_{n+1}, y_{n+1})$ from the Euler's forward method. Actually, $y^1_{n+1} = y_n + h\, f(x_n, y_n) \quad n = 0, 1, 2, ...$ and $y_{n+1} = y_n + h\, f(x_{n+1}, y^1_{n+1}) \qquad n = 0, 1, 2, ...$

Example: Find $y(1)$ using Euler's method: $y' = \frac{y^2 - x^2}{y^2 + x^2}, y(0) = 1$.

Given, $a = 0, b = 1, y_0 = 1$. **Let us choose** $n = 5$ **so that** $h = \frac{b-a}{n} = \frac{1-0}{5} = 0.2$
Now using Euler's formula on $[x_0, x_1] = [0, 0.2]$, **we obtain**
$y_1 = y_0 + hf(x_0, y_0) = y_0 + h\frac{y_0^2 - x_0^2}{y_0^2 + x_0^2} = 1 + 0.2 \times \frac{1^2 - 0^2}{1^2 + 0^2} = 1 + 0.2 \times 1 = 1.2$
Similarly, on $[x_1, x_2] = [0.2, 0.4]$, **we obtain**
$y_2 = y_1 + hf(x_1, y_1) = y_1 + h\frac{y_1^2 - x_1^2}{y_1^2 + x_1^2} = 1.2 + 0.2 \times \frac{1.2^2 - 0.2^2}{1.2^2 + 0.2^2}$
$y_2 = 1.2 + 0.2 \times 0.9459 = 1.3892$
Similarly, on $[x_2, x_3] = [0.4, 0.6]$, **we obtain**
$y_3 = y_2 + hf(x_2, y_2) = y_2 + h\frac{y_2^2 - x_2^2}{y_2^2 + x_2^2} = 1.3892 + 0.2 \times \frac{1.3892^2 - 0.4^2}{1.3892^2 + 0.4^2}$
$y_3 = 1.3892 + 0.2 \times 0.8469 = 1.5586$
Similarly, on $[x_3, x_4] = [0.6, 0.8]$, **we obtain**
$y_4 = y_3 + hf(x_3, y_3) = y_3 + h\frac{y_3^2 - x_3^2}{y_3^2 + x_3^2} = 1.5586 + 0.2 \times \frac{1.5586^2 - 0.6^2}{1.5586^2 + 0.6^2}$
$y_4 = 1.5586 + 0.2 \times 0.7419 = 1.7069$
Finally, on $[x_4, x_5] = [0.8, 1]$, **we obtain**
$y_5 = y_4 + hf(x_4, y_4) = y_4 + h\frac{y_4^2 - x_4^2}{y_4^2 + x_4^2} = 1.7069 + 0.2 \times \frac{1.7069^2 - 0.8^2}{1.7069^2 + 0.8^2}$
$y_5 = 1.7069 + 0.2 \times 0.6398 = 1.8349$

Exercise: Find $y(2)$ using Euler's method: $y' = y(1 + x^2), y(0) = 1$.

7.4.2 Modified Euler's Method

In the Modified Euler's Method the formula for iteration is

$$y^1_n = y_n + h\, f(x_n, y_n) \qquad n = 0, 1, 2, ...$$

$$y_{n+1} = y_n + \frac{h}{2}\left[f(x_n, y_n) + f(x_{n+1} + y^1_n)\right] \tag{7.8}$$

7.4.3 Exercise

1. Use Euler's method to find $y(0.5)$ using $h = 0.1$:

   a). $\dfrac{dy}{dx} = \dfrac{3}{5}x^3 y, \quad y(0) = 1$ b). $\dfrac{dy}{dx} = 1 + y^2, \quad y(0) = 0$

2. Use Euler's modified method to find $y(0.2)$ and $y(0.4)$ using $h = 0.2$:
   $\dfrac{dy}{dx} = x + y, \quad y(0) = 0.$

Example: Find $y(1)$ using Modified Euler's method: $y' = \frac{y^2-x^2}{y^2+x^2}, y(0) = 1$.

Given, $a = 0, b = 1, y_0 = 1$. **Let us choose** $n = 5$ **so that** $h = \frac{b-a}{n} = \frac{1-0}{5} = 0.2$
Now using Modified Euler's formula on $[x_0, x_1] = [0, 0.2]$, **we obtain**

$y_1^0 = y_0 + hf(x_0, y_0) = y_0 + h\frac{y_0^2-x_0^2}{y_0^2+x_0^2} = 1 + 0.2 \times \frac{1^2-0^2}{1^2+0^2} = 1 + 0.2 \times 1 = 1.2$

$y_1 = y_0 + \frac{h}{2}[f(x_0, y_0) + f(x_1, y_1)] = 1 + \frac{h}{2}\left[\frac{y_0^2-x_0^2}{y_0^2+x_0^2} + \frac{(y_1^0)^2-x_1^2}{(y_1^0)^2+x_1^2}\right]$

$y_1 = 1 + \frac{0.2}{2}\left[\frac{1^2-0^2}{1^2+0^2} + \frac{1.2^2-0.2^2}{1.2^2+0.2^2}\right] = 1 + 0.1 \times [1 + 0.9459] = 1.1946$

Similarly, on $[x_1, x_2] = [0.2, 0.4]$, **we obtain**

$y_2^0 = y_1 + hf(x_1, y_1) = y_1 + h\frac{y_1^2-x_1^2}{y_1^2+x_1^2} = 1.1946 + 0.2 \times \frac{1.1946^2-0.2^2}{1.1946^2+0.2^2} = 1.3837$

$y_2 = y_1 + \frac{h}{2}[f(x_1, y_1) + f(x_2, y_2)] = 1.1946 + \frac{h}{2}\left[\frac{y_1^2-x_1^2}{y_1^2+x_1^2} + \frac{(y_2^0)^2-x_2^2}{(y_2^0)^2+x_2^2}\right]$

$y_2 = 1.1946 + \frac{0.2}{2}\left[\frac{1.1946^2-0.2^2}{1.1946^2+0.2^2} + \frac{1.3837^2-0.4^2}{1.3837^2+0.4^2}\right] = 1.3737$

Similarly, on $[x_2, x_3] = [0.4, 0.6]$, **we obtain** $y_3^0 = 1.5425, y_3 = 1.5318$
Similarly, on $[x_3, x_4] = [0.6, 0.8]$, **we obtain** $y_4^0 = 1.6786, y_4 = 1.6682$
Finally, on $[x_4, x_5] = [0.8, 1]$, **we obtain** $y_5^0 = 1.7934, y_5 = 1.7833$

Exercise: Find $y(2)$ using Euler's method: $y' = y(1 + x^2), y(0) = 1$.

Example

Find $y(1)$ using Runge-Kutta second order (RK2) method:
$y' = \frac{y^2-x^2}{y^2+x^2}, y(0) = 1$.

Given, $a = 0, b = 1, y_0 = 1$. **Let us choose** $n = 5$ **so that** $h = \frac{b-a}{n} = \frac{1-0}{5} = 0.2$
Now using Runge-Kutta second order formula on $[x_0, x_1] = [0, 0.2]$, **we get**

$k_1 = hf(x_0, y_0) = h\frac{y_0^2-x_0^2}{y_0^2+x_0^2} = 0.2 \times \frac{1^2-0^2}{1^2+0^2} = 0.2$

$k_2 = hf(x_0 + h, y_0 + k_1) = 0.2f(0.2, 1.2) = 0.2 \times \frac{1.2^2-0.2^2}{1.2^2+0.2^2} = 0.1892$

$y_1 = y_0 + \frac{1}{2}[k_1 + k_2] = 1 + 0.5 \times (0.2 + 0.1892) = 1.1946$

Similarly, on $[x_1, x_2] = [0.2, 0.4]$, **we obtain**

$k_1 = hf(x_1, y_1) = h\frac{y_1^2-x_1^2}{y_1^2+x_1^2} = 0.2 \times \frac{1.1946^2-0.2^2}{1.1946^2+0.2^2} = 0.1891$

$k_2 = hf(x_1 + h, y_1 + k_1) = 0.2f(0.4, 1.3837) = 0.2 \times \frac{1.3837^2-0.4^2}{1.3837^2+0.4^2} = 0.1692$

$y_2 = y_1 + \frac{1}{2}[k_1 + k_2] = 1.1946 + 0.5 \times (0.1891 + 0.1692) = 1.3737$

Similarly, on $[x_2, x_3] = [0.4, 0.6]$, **we get** $k_1 = 0.1687, k_2 = 0.1474, y_3 = 1.5318$
Similarly, on $[x_3, x_4] = [0.6, 0.8]$, **we get** $k_1 = 0.1468, k_2 = 0.1260, y_4 = 1.6682$
Finally, on $[x_4, x_5] = [0.8, 1]$, **we get** $k_1 = 0.1252, k_2 = 0.1051, y_5 = 1.7833$

Exercise:Find $y(2)$ using Runge-Kutta fourth order (RK4) method:
$y' = y(1 + x^2), y(0) = 1$.

7.5 Runge-Kutta Methods

Euler's method is less efficient in practical problems. It requires h to be small for obtaining reasonable accuracy. To overcome this issue, Runge-Kutta methods are designed. Let $k_0 = hf(x_0, y_0)$ and $k_1 = hf(x_1, y_0 + hf(x_0, y_0))$. The special case of second order Runge-Kutta

Method gives $y_1 = y_0 + \frac{h}{2}(k_0 + k_1)$, which is equivalent to Modified Euler's Method. So, the iteration formula for the **second-order Runge-Kutta Method** is:

$$\begin{aligned} k_n &= f(x_n, y_n) \\ k_{n+1} &= f(x_{n+1}, y_n + hk_n) \\ y_{n+1} &= y_n + \frac{h}{2}\left[k_n + k_{n+1}\right] \end{aligned} \tag{7.9}$$

7.5.1 Fourth-Order Runge-Kutta Method

The iteration formula for the Fourth-Order Runge-Kutta method is as follows:

$$\begin{aligned} k_n &= f(x_n, y_n) \\ k_{n+1} &= f(x_n + (h/2), y_n + (h/2)k_n) \\ k_{n+2} &= f(x_n + (h/2), y_n + (h/2)k_{n+1}) \\ k_{n+3} &= hf(x_n + h, y_n + hk_{n+2}) \\ y_{n+1} &= y_n + \frac{h}{6}\left[k_n + 2k_{n+1} + 2k_{n+2} + k_{n+3}\right] \end{aligned} \tag{7.10}$$

7.5.2 Exercise

1. Use Runge-Kutta fourth order formula

   (a) to find $y(0.2)$ and $y(0.4)$ given $y' = \dfrac{y^2 - x^2}{y^2 + x^2}, \quad y(0) = 1.$

   (b) to find $y(1.2)$ and $y(1.4)$ given $y' = \dfrac{3x + y}{x + 2y}, \quad y(1) = 1.$

   (c) to find $y(0.2)$ given $y'' + y = 0, \quad y(0) = 1, \quad y'(0) = 0.$

2. Given $\dfrac{dy}{dx} = y(1 + x^2), \quad y(0) = 1$ find the values for $x = 0.2, 0.4, 0.6$ using the Euler, the modified Euler, and the fourth order Runge-Kutta methods. Compare the results thus obtained.

3. Solve the boundary value problems by finite difference method taking $y(0) = 0, \quad y(1) = 1$ for $y = 0.5, 0.25$: a). $y'' - y = 0$ b).$y'' - x^2y' - e^xy = 1/(1 + x)$

7.6 Simultaneous and Higher Order Equations

Consider a pair simultaneous equations:

$$\frac{dy}{dt} = f(t, x, y) \quad \text{and} \quad \frac{dx}{dt} = g(t, x, y) \tag{7.11}$$

with the initial conditions $x = x_0$ and $y = y_0$ when $t = t0$. Assuming that $\Delta t = h$, $\Delta x = k$, and $\Delta y = l$, the fourth order Runge-Kutta method gives,

$$\begin{aligned} k_n &= f(t_n, x_n, y_n) \\ l_n &= g(t_n, x_n, y_n) \\ &\ldots \\ y_{n+1} &= y_n + \frac{h}{6}\left[k_n + 2k_{n+1} + 2k_{n+2} + k_{n+3}\right] \\ x_{n+1} &= x_n + \frac{h}{6}\left[l_n + 2l_{n+1} + 2l_{n+2} + l_{n+3}\right] \end{aligned} \tag{7.12}$$

We now consider the second-order differential equation:

$$y'' = F(x, y, y'), \quad y(x_0) = y_0, \quad y'(x_0) = y'_0 \tag{7.13}$$

By setting $z = y'$, the above equation 7.13 becomes a system of first order equations in two equations.

$$y' = z, \quad z' = F(x, y, z), \quad y(x_0) = y_0, \quad z(x_0) = y'_0 \tag{7.14}$$

7.7 Boundary Value Problems

The second order ordinary differential equation consists of two boundary equations as follows:

$$F(x, y, y', y'') = 0, \quad y(x_0) = a, \quad y(x_n) = b$$

One of the popular method of solution to this boundary value problem is the method of finite difference.

7.8 Finite-Difference Method

The finite difference method for the solution of a two-point boundary value problem consists in replacing the derivatives occurring in the differential equation (and in the boundary conditions as well) by means of their finite-difference approximations and then solving the resulting linear system of equations by a standard procedure. We have the following finite difference formula for the derivatives.

1. **Forward Difference**

   a). $y'(x) \approx \dfrac{y(x+h) - y(x)}{h}$ b). $y''(x) \approx \dfrac{y(x+2h) - 2y(x+h) + y(x)}{h^2}$

   Backward Difference

   a). $y'(x) \approx \dfrac{y(x) - y(x-h)}{h}$ b). $y''(x) \approx \dfrac{y(x) - 2y(x-h) + y(x-2h)}{h^2}$

2. **Central Difference**

a). $y'(x) \approx \dfrac{y(x+h) - y(x-h)}{2h}$ b). $y''(x) \approx \dfrac{y(x+h) - 2y(x) + y(x-h)}{h^2}$

In principle the central difference formula should be $y'(x) \approx \dfrac{y(x+h/2) - y(x-h/2)}{h}$ but for easier computation during program execution the above formula has been adopted.

These formulae can be easily derived by expanding $y(x)$ in the neighborhood of x using Taylor's series. In general for higher order derivatives central difference formula gives better accuracy so,this formula is commonly used.

7.8.1 General Method:

To solve the boundary-value problem 7.15:

$$y'' + f(x)y' + g(x)y = r(x), \quad y(x_0) = a, \quad y(x_n) = b \tag{7.15}$$

1. First we divide the range $[x_0, x_n]$ into n equal subintervals of width h so that $x_i = x_0 + ih, \;\; i = 1, 2, ..., n.$
2. Then $y(x_i) = y_i$ and $y'_i = \dfrac{y_{i+1} - y_{i-1}}{2h}$ and $y''_i = \dfrac{y_{i+1} - 2y_i + y_{i-1}}{h^2}$.
3. Substituting these values in 7.15 we get,
$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} + f_i\frac{y_{i+1} - y_{i-1}}{2h} + g_iy_i = r_i, \qquad i = 1, 2, ..., n-1$$
4. Multiplying through $2h^2$, and gathering the like terms, we get,
$$(2 - hf_i)y_{i-1} + 2(g_ih^2 - 4)y_i + (2 + hf_i)y_{i+1} = 2r_ih^2, \qquad i = 1, 2, ..., n-1$$
with $y_0 = a, \quad y_n = b$. This comprise of a system of linear equations which can be solved with a suitable method to obtain the solution of the original differential equation 7.15.

Example 2. Solve: $y'' - cosx\, y' - sinx\, y = e^x, \qquad y(0) = 0, \quad y(1) = 1$, taking $n = 4$.
Solution Steps:

1. Here, $x_0 = 0$ and $x_n = 1$. Choosing $n = 4$ we get $h = (1-0)/4 = 0.25$.
2. At $x = x_i$ we have $\dfrac{y_{i+1} - 2y_i + y_{i-1}}{h^2} - cosx_i\,\dfrac{y_{i+1} - y_i}{h} - sinx_i\; y_i = e^{x_i}$
Substituting $h = 0.25$, we get, $\dfrac{y_{i+1} - 2y_i + y_{i-1}}{(0.25)^2} - cosx_i\,\dfrac{y_{i+1} - y_i}{0.25} - sinx_i\; y_i = e^{x_i}$
3. Simplifying: $16y_{i-1} + (-32 + 4cosx_i - sinx_i)y_i + (16 - 4cosx_i)y_{i+1} = e^{x_i}$
4. Substituting $x_0 = 0, \quad x_1 = 0.25, \quad x_2 = 0.5, \quad x_3 = 0.75$ and $y_0 = 1, \quad y_4 = 1$:
$$(16)0 + (-32 + 4cos(0.25) - sin(0.25))(16 - 4cos(0.25))y_2 = e^{0.25}$$
$$16y_1 + (-32 + 4cos(0.5) - sin(0.5))y_2 + (16 - 4cos(0.5))y_3 = e^{0.5}$$
$$16y_2 + (-32 + 4cos(0.75) - sin(0.75))y_3 + (16 - 4cos(0.75))1 = e^{0.75}$$

5. Simplifying:

$$\begin{aligned} -28.371\,y_1 + 12.124\,y_2 &= 1.284 \\ 16\,y_1 - 28.969\,y_2 + 12.489\,y_3 &= 1.648 \\ 16\,y_2 - 29.754\,y_3 + 14.54 &= -10.956 \end{aligned} \tag{7.16}$$

6. Solving: $y_1 = 0.016, \quad y_2 = 0.144, \quad y_3 = 0.445.$

x	y
0	0
0.25	0.016
0.5	0.144
0.75	0.454
1	1