MA/MSc MT-08

Vardhaman Mahaveer Open University, Kota

Numerical Analysis

MA/MSc MT-08

Vardhaman Mahaveer Open University, Kota

Numerical Analysis

Course Development Committee

Chairman
Prof. (Dr.) Naresh Dadhich
Vice-Chancellor
Vardhaman Mahaveer Open University, Kota

Co-ordinator/Convener and Members

Subject Convener
Prof. D.S. Chauhan
Department of Mathematics
University of Rajasthan, Jaipur

Co-ordinator
Dr. Anuradha Sharma
Assistant Professor
Department of Botany, V.M.O.U., Kota

Members :

1. **Prof. V.P. Saxena**
Ex Vice-Chancellor
Jiwaji University,
Gwalior (MP)
2. **Prof. S.C. Rajvanshi**
Department of Mathematics
Institute of Eng. & Tech.
Bhaddal, Ropar (Punjab)
3. **Prof. P.K. Banerjee**
Emeritus Fellow (UGC)
Department of Mathematics
J.N.V. University, Jodhpur
4. **Prof. S.P. Goyal**
Emeritus Scientist (CSIR)
Department of Mathematics
University of Rajasthan, Jaipur
5. **Dr. A.K. Mathur**
Associate Prof. (Retired)
Department of Mathematics
University of Rajasthan, Jaipur
6. **Dr. K.N. Singh**
Associate Prof. (Retired)
Department of Mathematics
University of Rajasthan, Jaipur
7. **Dr. Paresh Vyas**
Assistant Professor
Department of Mathematics
University of Rajasthan, Jaipur
8. **Dr. Vimlesh Soni**
Lecturer
Department of Mathematics
Govt. PG College, Kota (Raj.)
9. **Dr. K.K. Mishra**
Lecturer
Department of Mathematics
M.S.J. College, Bharatpur (Raj.)
10. **Dr. K.S. Shekhawat**
Lecturer, Department of Mathematics
Govt. Shri Kalyan College, Sikar (Raj.)

Editing and Course Writing

Editor
Prof. R.C. Choudhary
D-299, Sarvanand Marg
Malviya Nagar, Jaipur

Writers

1. **Dr. Vimlesh Soni**
Department of Mathematics
Goverment College, Kota
2. **Dr. I.K. Dadhich**
Birla Institute of Technology (Ranchi)
Jaipur Campus
Malviya Industrial Area, Jaipur
3. **Dr. Paresh Vyas**
Department of Mathematics
University of Rajasthan, Jaipur

Academic and Administrative Management

Prof. (Dr.) Naresh Dadhich
Vice-Chancellor
Vardhaman Mahveer Open University,
Kota

Prof. B.K. Sharma
Director (Academic)
Vardhaman Mahveer Open University,
Kota

Mr. Yogendra Goyal
Incharge
Material Production and
Distribution Department

Course Material Production

Mr. Yogendra Goyal
Assistant Production Officer
Vardhaman Mahaveer Open University, Kota

PREFACE

The present book entitled **"Numerical Analysis"** has been designed so as to cover the unit-wise syllabus of MA/MSc MT-08 course for M.A./M.Sc. Mathematics (Final) students of Vardhaman Mahaveer Open University, Kota. It can also be used for competitive examinations. The basic principles and theory have been explained in a simple, concise and lucid manner. Adequate number of illustrative examples and exercises have also been included to enable the students to grasp the subject easily. The units have been written by various experts in the field. The unit writers have consulted various standard books on the subject and they are thankful to the authors of these reference books.

Unit - 1 : Iterative Methods

Structure of the Unit

1.0 Objectives

In this unit, we shall study the methods for finding the roots of the equation of the type $f(x)=0$. We shall also study the methods for finding solution of system of non-linear equations having two varibales, which can be generalized for more than two variables.

1.1 Introduction

An equation, $f(x)=0$, is said to be **algebric** if it is purely a polynomial in x and is said to be transcendental if $f(x)$ contains trigonometric, logarithmic or exponential function. For example, $x^3+2x^2+5x-7=0$ is an algebric equation while $x^3+\sin x=0$ and $e^x\log x+\tan x+x^2=0$ are transcendental equations.

The value $x=\alpha$ is said to be a root or solution of the equation $f(x)=0$ if it satisfies the equation, i.e., if $f(\alpha)=0$.

The root $x=\alpha$ is a **simple root** if $f(x)$ contains a factor $(x-\alpha)$ only once, i.e., if $f(x)=(x-\alpha)g(x)$, where $g(\alpha)\neq 0$, and $f'(\alpha)\neq 0$.

The root $x = \alpha$ is a **multiple root** with **multiplicity m** if

$f(x) = (x-\alpha)^m g(x)$, where $g(\alpha) \neq 0$. In this case, $f'(\alpha) = f''(\alpha) = ... = f^{(m-1)}(\alpha) = 0$ and $f^{(m)}(\alpha) \neq 0$.

We shall study iterative methods to find the root of the equation $f(x) = 0$. To get the solution using iterative method, first we take an initial approximation x_0 for the root of the equation and using the required iterative method, we get an improved approximate value for the root x_1 (say), called first approximation. Using x_1 in the method or formula, we get next improved approximate value x_2, called second approximation. Continuing the process, we obtain a sequence $\{x_n\}$ of approximations. If

$$\lim_{n\to\infty} x_n = \alpha$$

then we say that this sequence converges to the root α.

After some finite steps of the above process, we get an approximate root with some **error tolerance ε.** If $|x_{n+1} - x_n| < \varepsilon$, for some n, and if ε is the required error tolerance then we stop the process.

Let $x_n - \alpha = \varepsilon_n$ and $x_{n+1} - \alpha = \varepsilon_{n+1}$ be the error in nth and $(n+1)th$ approximations respectively, then method or process of approximations is said to be **convergent**, if

$$|\varepsilon_{n+1}| = c|\varepsilon_n|^p, \; p \geq 1 \qquad ...(1)$$

This is called error equation for the method. Here p is said to be the **order** of iterative method. If $p = 1$, then $c < 1$ and we say that method is of linear convergence. In comparison of two methods, that method is faster which has greater value of p.

For any iterative method, initial approxmiation is required. To get an initial approxmiation, mostly the following theorem is used :

"If $f(x)$ is continuous on the interval $[a,b]$ and $f(a)$ and $f(b)$ have opposite signs then one root of the equation $f(x) = 0$ lies in the interval (a,b)."

Now we shall study some of the iterative methods for finding the root of the given equation $f(x) = 0$.

1.2 Bisection Method

Let the given equation be $f(x) = 0$ and a and b are two points such that $f(a)f(b) < 0$, i.e., $f(a)$ and $f(b)$ are of opposite signs. Also, let $f(x)$ be continuous on the interval $[a,b]$. Then the root lies in (a,b).

Let

$$x_0 = \frac{a+b}{2}$$

If $f(x_0) = 0$, then it is the root of the equation, otherwise $f(x_0)$ will have either positive or negative sign. If, $f(x_0)f(a) < 0$, then the root lies in the interval (a, x_0) and we shall take next approximation as

$$x_1 = \frac{a + x_0}{2}$$

otherwise, $f(x_0)$ and $f(b)$ will have opposite signs and the next approximation x_1 will be given by

$$x_1 = \frac{x_0 + b}{2}$$

Let this interval be $[a_1, b_1]$ (say), i.e., x_1 is the midpoint of $[a_1, b_1]$. The length of this interval will be half of the original interval $[a, b]$. Now, the next approximation x_2 will be the mid point of the interval, either $[a_1, x_1]$ or $[x_1, b_1]$, according to the condition $f(a_1)f(x_1) < 0$ or $f(x_1)f(b_1) < 0$, respectively.

Continuing the process, we obtain the sequence $x_1, x_2, x_3.....$ and we stop the process when we obtain required error tolerance ε, i.e.,

$$|x_{n+1} - x_n| < \varepsilon$$

In above process, length of the new interval will be exactly half of the length of the previous one. At each step length is reduced by a factor of one-half. At the end of nth step, the new interval will be $[a_n, b_n]$ such that its length is $(b-a)/2^n$. Then, we have

$$\frac{(b-a)}{2^n} \le \varepsilon$$

$$\Rightarrow \qquad n \ge \frac{\log_e\left(\frac{b-a}{\varepsilon}\right)}{\log_e 2}$$

This method has linear convergence and it always works. This method is alos known as **Interval halving method.**

Example 1.1 Find a real root of the equation $x^3 - 9x + 1 = 0$ by bisection method.

Solution : Given that $f(x) = x^3 - 9x + 1$, then $f(2) = -9$ and $f(3) = 1$, so that $f(2)f(3) < 0$ and hence root lies in the interval (2,3), then

$$x_0 = \frac{2+3}{2} = 2.5, \qquad f(2.5) = -5.88$$

so that, $f(2.5)f(3) < 0$, therefore

$$x_1 = \frac{2.5+3}{2} = 2.75, \ f(2.75) = -2.9531$$

Thus, $f(2.75)f(3) < 0$, and

$$x_2 = \frac{2.75+3}{2} = 2.875, \ f(2.875) = -1.1113$$

Proceeding similarly, we obtain a sequence of approximations as follows :

$x_3 = 2.9375$, $\quad f(x_3) < 0$

$x_4 = 2.96875$, $\quad f(x_4) > 0$

$x_5 = 2.95313$, $\quad f(x_5) > 0$

$x_6 = 2.94532$, $\quad f(x_6) > 0$

$x_7 = 2.94141$, $\quad f(x_9) < 0$

$x_8 = 2.94337$, $\quad f(x_8) > 0$

$x_9 = 2.94239$, $\quad f(x_9) < 0$

$x_{10} = 2.94288$,

Thus, we can take approximate value of the root as 2.942 correct upto four significant digit.

1.3 Regula Falsi Method

This method is also known as method of **false position.** In this method, we take two points x_0 and x_1 such that $f(x_0)f(x_1) < 0$ and $f(x)$ is continuous on the interval $[x_0, x_1]$. Part of the curve $y = f(x)$ between the points $(x_0, f(x_0))$ and $(x_1, f(x_1))$ is replaced by a chord joining these two points and the point of intersection of the chord and x-axis is considered as first approximation of the root.

The equation of chord joining $(x_0, f(x_0))$ and $(x_1, f(x_1))$ is given by

$$y - f(x_0) = \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) \qquad \text{...(2)}$$

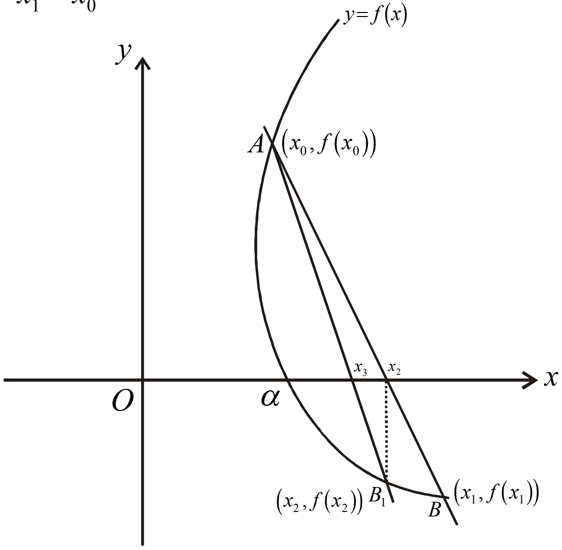

Figure 1.1 : Regula Falsi Method

Intersection of equation (2) and x-axis can be obtained by substituting $y = 0$ in equation (2). Thus, we have

$$x = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)} \qquad ...(3)$$

This value of x is considered as a next approximation x_2 (say). Now, if $f(x_0)f(x_2) < 0$ then we consider the chord joining $(x_0, f(x_0))$ and $(x_2, f(x_2))$ to replace the part of the curve between this two points otherwise points $(x_1, f(x_1))$ and $(x_2, f(x_2))$ will be considered. We repeat this process till required accuracy is obtained. In general, the sequence of approximations can be obtained by using the formula :

$$x_{n+1} = \frac{x_{n-1} f(x_n) - x_n f(x_{n-1})}{f(x_n) - f(x_{n-1})},\ f(x_{n-1})f(x_n) < 0, \qquad ...(4)$$

$n = 1,2,3.....$

In this method, one point remains fixed. Let fixed point be $(x_0, f(x_0))$. Let α be the exact root of the equation $f(x) = 0$ and let $x_n - \alpha = \varepsilon_n$ be the error in nth approximation, then using

$$x_{n+1} = \frac{x_0 f(x_1) - x_n f(x_0)}{f(x_n) - f(x_0)}$$

and Taylor's series expansion of the functions $f(\alpha + \varepsilon_0)$ and $f(\alpha + \varepsilon_n)$, we get (after neglecting higher powers of ε_n),

$$\varepsilon_{n+1} = \frac{1}{2}\frac{f''(\alpha)}{f'(\alpha)}\varepsilon_0\,\varepsilon_n$$

$$\Rightarrow \qquad \varepsilon_{n+1} \le \varepsilon_n$$

This shows that Regula-Falsi method has linear convergence.

1.4 Secant Method

This method is also known as **chord method.** The technique is similar to Regula-Falsi method. Only difference is that the condition $f(x_{n-1})f(x_n) < 0$ is dropped and we use last two consecutive points to get the next approximation.

Thus, the formula is same as that of the Regula-Falsi method (equation (4) of 1.3) without the condition $f(x_{n-1})f(x_n) < 0$.

Using $x_n - \alpha = \varepsilon_n$, $x_{n+1} - \alpha = \varepsilon_{n+1}$, $x_{n-1} - \alpha = \varepsilon_{n-1}$ and $f(\alpha) = 0$ in formula (4) of (1.3) and expanding the functions by taylor's series, we get

$$\varepsilon_n = \frac{1}{2}\frac{f''(\alpha)}{f'(\alpha)}\varepsilon_n\varepsilon_{n-1} \quad \text{(neglecting higher powers of } \varepsilon_n \text{ and } \varepsilon_{n-1}\text{)},$$

or $\varepsilon_n = c\,\varepsilon_n\,\varepsilon_{n-1}$, where $c = \frac{1}{2}\frac{f''(\alpha)}{f'(\alpha)}$ ...(5)

Let error-equation be

$$\varepsilon_{n+1} = a\,\varepsilon_n^p \qquad ...(6)$$

Solving (5) and (6), we get $p = 1.618$, i.e.

$$\varepsilon_{n+1} = a\,\varepsilon_n^{1.618}$$

Thus, convergence of this method is 1.618. Hence secant method is faster than Regula-Falsi method but it may fail when $f(x_n) = f(x_{n-1})$.

Example 1.2 Find the real root of the equation $x^3 - 2x - 5 = 0$ using

(A) Regula-Falsi method

(B) Secant method.

Solution : Here $f(x) = x^3 - 2x - 5$, then

$f(z) = -1$ and $f(3) = 16$, so the root lies in the interval (2, 3). Also, function is continuous in this inteval. Let $x_0 = 2$ and $x_1 = 3$.

(A) By Regula Falsi Method : The next iteration is given by

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}$$

$$= \frac{(2)(16) - (3)(-1)}{16 - (-1)} = \frac{35}{17}$$

$$= 2.0588$$

Since, $f(2.0588) = -0.3911$, so the root lies in the interval $(2.0588, 3)$. Thus, the next iteration is given by

$$x_3 = \frac{x_1 f(x_2) - x_2 f(x_1)}{f(x_2) - f(x_1)}$$

$$= \frac{(3)(-0.3911) - (2.0588)(16)}{-0.3911 - 16} = \frac{-34.1141}{-16.3911}$$

$$= 2.0813$$

Since, $f(2.0813) = -0.1468$, so the root lies in the interval $(2.0813, 3)$. Thus the next iteration is given by

$$x_4 = \frac{x_1 f(x_3) - x_3 f(x_1)}{f(x_3) - f(x_1)}$$

$$= \frac{(3)(-0.1468) - (2.0813)(16)}{-0.1468 - 16} = \frac{-33.7412}{-16.1468}$$

$$= 2.0897$$

Now, $f(2.0897) = -0.0540$, so the root lies in the interval $(2.0897, 3)$. Thus, the next iteration is

$$x_5 = \frac{x_1 f(x_4) - x_4 f(x_1)}{f(x_4) - f(x_1)}$$

$$= \frac{(3)(-0.0540) - (2.0897)(16)}{-0.0540 - 16} = \frac{-33.5972}{-16.0540}$$

$$= 2.0928, \; f(2.0928) = -0.0195$$

Similarly, we get

$$x_6 = 2.0939$$

Thus, approximate value of the root can be taken as 2.09 correct to two decimal places.

(B) By Secant Method : First and second step of the part (A) will be same here. The next iteration x_4 will be obtained using consecutive points x_2 and x_3 (where $x_2 = 2.0588$, $x_3 = 2.0813$) instead of using x_1 and x_3. Thus

$$x_4 = \frac{x_2 f(x_3) - x_3 f(x_2)}{f(x_3) - f(x_2)}$$

$$= \frac{(2.0588)(-0.1468) - (2.0813)(-0.3911)}{(-0.1468) - (-0.3911)}$$

$$= \frac{0.51176459}{0.2443} = 2.0948$$

Similarly, the next iteration x_5 can be obtained using the points x_3 and x_4, which is give by

$$x_5 = \frac{x_3 f(x_4) - x_4 f(x_3)}{f(x_4) - f(x_3)}$$

$$= \frac{(2.0813)(0.0028) - (2.0948)(-0.1468)}{0.0028 - (-0.1468)}$$

$$= \frac{0.31319788}{0.1496} = 2.0936$$

Thus, the root is 2.09 correct to two decimal places.

1.5 Newton-Raphson Method

Let x_n be an approximate root of the equation $f(x)=0$ and let x_{n+1} be the correct root such that

$$x_{n+1} = x_n + h \qquad ...(7)$$

Then, $f(x_{n+1}) = 0$

$$\Rightarrow \qquad f(x_n + h) = 0$$

$$\Rightarrow \qquad f(x_n) + h f'(x_n) + \frac{h^2}{2!} f''(x_n) + ... = 0$$

Neglecting higher powers of h (if h is small), we get

$$f(x_n) + h f'(x_n) = 0$$

which gives

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

Then by (7), we have

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \; n = 0,1,2,.... \qquad ...(8)$$

Geometrically, we draw a tangent to the curve $y = f(x)$ at initial point $(x_0, f(x_0))$ and the point of intersection of this tangent with x-axis is taken as next approximation, say x_1. Then, we again draw a tangent at $(x_1, f(x_1))$ and get the next iteration. Thus, we replace the part of the curve between the point $(x_n f(x_n))$ and the x-axis by the tangent to the curve at that point. This method is also known as **tangent method.**

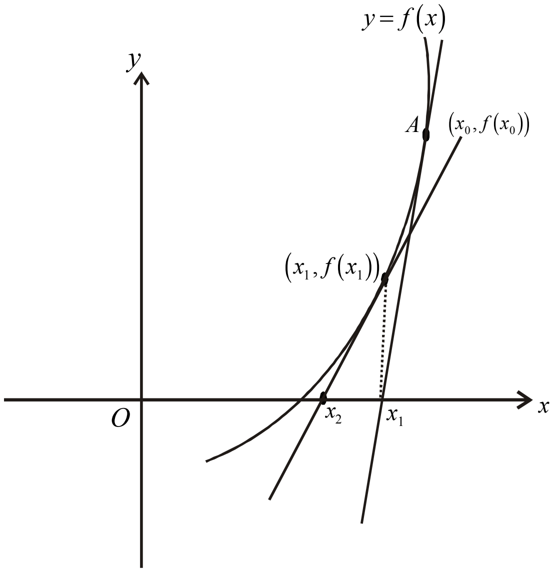

Figure 1.2 : Newton-Raphson Method

For convergence of the method, initial approximation x_0 must be chosen sufficiently close to the exact root, i.e, h must be small.

Substituting $x_n - \alpha = \varepsilon_n$ in (8) and using Taylor's series expansion with the fact $f(\alpha) = 0$, we get error equation as

$$\varepsilon_{n+1} = c\,\varepsilon_n^2, \text{ where } c = \frac{1}{2}\frac{f''(\alpha)}{f'(\alpha)},$$

which shows that this method has quadratic convergence.

Example 1.3 Find the root of the equation $\sin x - x^3 = 1$ using Newton-Raphson method.

Solution : Given that, $f(x) = \sin x - x^3 - 1$

Then $f'(x) = \cos x - 3x^2$

Now $f(-1) = -0.8415$, $f(-2) = 6.0907$

Therefore, the root lies in the interval $(-2,-1)$ and since $f(x)$ is continuous in this interval, we can take mid point -1.5 as initial approximation. Using Newton-Raphson method, with $x_0 = -1.5$, we get

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

$$= x_0 - \frac{\sin x_0 - x_0^3 - 1}{\cos x_0 - 3x_0^2}$$

$$= (-1.5) - \frac{\sin(-1.5) - (-1.5)^3 - 1}{\cos(-1.5) - 3(-1.5)^2}$$

$$= -1.2938$$

The second approximation is given by

$$x_2 = (-1.2938) - \frac{\sin(-1.2938) - (-1.2938)^3 - 1}{\cos(-1.2938) - 3(-1.2938)^2}$$

$$= -1.2509$$

Proceeding similarly, we can get approximate value of the root upto required accuracy.

1.5.1 Newton-Raphson method for nearly equal roots

When $f'(x)$ is very small or nearly equal to zero at $x = a$ then there will be two nearly equal roots in the neighbourhood of $x = a$ if $f(a) \neq 0$ and $f(a)f''(a) < 0$.

Expanding $f(x)$ about $x = a$, we get

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \ldots..,$$

neglecting the terms containing third and higher order of $\left(x-a\right)$ and using the fact that $f\left(x\right)=0$ and $f'\left(a\right)=0$, we get

$$f\left(a\right)+\frac{\left(x-a\right)^2}{2}f''\left(a\right)=0$$

which gives

$$x=a\pm\sqrt{\frac{-2f\left(a\right)}{f''\left(a\right)}} \qquad \text{...(9)}$$

Thus, one root is in left of a and the other is in right. These two values, further can be improved separately by Newton-Raphson scheme taking these values as initial approximations.

Example 1.4 Find two nearly equal roots of the equation $x^3-4.8x^2+6.5x-2.7=0$ in the neighbourhood of $x=1$.

Solution : Here $f\left(x\right)=x^3-4.8x^2+6.56x-2.7$,

$$f'\left(x\right)=3x^2-9.6x+6.56$$

$$f''\left(x\right)=6x-9.6$$

Now, $f'\left(1\right)=-0.04$, which is very small, also $f\left(1\right)\neq 0$ and $f\left(1\right)f''\left(1\right)<0$, therefore, there exist two roots in the neighbourhood of $x=1$, which are nearly equal. Taking $a=1$, we get $f\left(1\right)=0.06$ and $f''\left(1\right)=-3.6$ and by scheme (9), we have

$$x=1\pm\sqrt{\frac{-2\times 0.06}{-3.6}}$$

$$=1\pm\sqrt{0.03333}$$

$$=1\pm 0.18257$$

which gives,

$$x_1=1+0.18257$$

$$=1.18257$$

and $\quad x_2=1-0.18257$

$$=0.81743$$

These are approximate values of two roots in the neighbourhood of $x=1$, which can be improved by Newton-Raphson method.

1.5.2 Newton-Raphson method fro pth root of a number

We can find pth root of a given number N using Newton-Raphson method as follows :

Let

$$(\mathrm{N})^{1/p} = x$$

Which can be written as

$$x^p - N = 0$$

Thus, root of this equation will be the required solution.

Taking $f(x) = x^p - N$,

We get $f'(x) = p\,x^{p-1}$,

Then, by scheme (8), we have

$$x_{n+1} = x_n - \frac{(x_n^p - N)}{p\,x_n^{p-1}}$$

or $$x_{n+1} = \frac{(p-1)\,x_n^p + N}{p\,x_n^{p-1}} \qquad ...(11)$$

Example 1.6 Find square root of 10 using Newton-Raphson method.

Solution : Since $\sqrt{9} < \sqrt{10} < \sqrt{16}$,

that is, $3 < \sqrt{10} < 4$,

therefore square root of 10 lies in the interval $(3,4)$. Let us take $x_0 = 3.2$ as a initial approximation.

Now, from scheme (11), we have

$$x_{n+1} = \frac{(p-1)\,x_n^p + N}{p\,x_n^{p-1}}, \quad n = 0,1,2,......$$

Here $p = 2$, $N = 10$, so that

$$x_{n+1} = \frac{x_n^2 + 10}{2\,x_n}, \quad n = 0,1,2,.......$$

Thus,

$$x_1 = \frac{x_0^2 + 10}{2\,x_0} = \frac{(3.2)^2 + 10}{2(3.2)}$$

$$= \frac{20.24}{6.4} = 3.1625$$

Again, $x_2 = \dfrac{(3.1625)^2 + 10}{2(3.1625)} = \dfrac{20.00140625}{6.325}$

$= 3.162278$

Similarly, the next iteration is given by

$$x_3 = \frac{(3.162278)^2 + 10}{2(3.162278)} = 3.162278$$

thus, we can take $x = 3.162278$ as the square root of 10 correct upto six decimal places.

1.6 Iteration Method

The equation $f(x) = 0$ can be rewritten as

$$x = \phi(x) \qquad ...(12)$$

then the recurrence relation

$$x_{n+1} = \phi(x_n),\ n = 0,1,2,... \qquad ...(13)$$

can be taken as iterative scheme to get the required root, provided that

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

where x is in the neighbourhood of exact root α. This condition is necessary for the convergence of the scheme (13).We can verify this condition as follows :

Let α be the exact root of the equation $f(x) = 0$, then by (12), we have

$$\alpha = \phi(\alpha) \qquad ...(14)$$

and if $x_n - \alpha = \varepsilon_n$, $x_{n+1} - \alpha = \varepsilon_{n+1}$ are errors in x_n and x_{n+1} approximations respectively, then by (13), we have

$$x_{n+1} - \alpha = \phi(x_n) - \alpha$$

$$= \phi(x_n) - \phi(\alpha) \qquad \text{(by (14))}$$

$$= \frac{\phi(x_n) - \phi(\alpha)}{(x_n - \alpha)}(x_n - \alpha)$$

$\Rightarrow$ $\varepsilon_{n+1} = \varepsilon_n \phi'(\alpha_n)$, where $\alpha < \alpha_n < x_n$

or $\varepsilon_{n+1} = c\,\varepsilon_n$, where $c = \phi'(\alpha_n)$ ...(15)

This shows that scheme (13) has linear convergence if $|C| < 1$, j.e., $|\phi'(\alpha_n)| < 1$, where α_n is in the neighbourhood of α. Hence the condition is verified.

Example 1.7 Find a root of the equation

$$3x-\sqrt{1+\sin x}=0$$

using iteration method.

Solution : Here $f(x)=3x-\sqrt{1+\sin x}$, then

$$f(0)=-1 \text{ and } f(1)=\left(3-\sqrt{1+\sin 1}\right)=1.932$$

therefore root lies in the interval $(0,1)$. Let us take initial approximation $x_0=0.5$. The given equation can be written as

$$x=\frac{\sqrt{1+\sin x}}{3}=\phi(x) \qquad \text{(say)}$$

then $$\phi'(x)=\frac{\cos x}{6\sqrt{1+\sin x}}$$

Now, since $|\cos x|\le 1$, therefore

$|\phi'(x)|<1$, in the neighbourhood of $x_0=0.5$.

Thus, the iteration scheme is given by

$$x_{n+1}=\frac{\sqrt{1+\sin x_n}}{3},\ n=0,1,2,\ldots\ldots$$

so that

$$x_1=\frac{\sqrt{1+\sin x_0}}{3}=\frac{\sqrt{1+\sin(0.5)}}{3}=0.40544$$

the next iteration is given by

$$x_2=\frac{\sqrt{1+\sin(0.40544)}}{3}=0.39362$$

Similarly, we can get other approximations given by

$$x_3=0.39208$$

$$x_4=0.39188$$

$$x_5=0.39185$$

$$x_6=0.39185$$

Thus, the root is 0.39185 correct upto five decimal places.

1.6.1 Aitken's Δ^2-method to accelerate the convergence

We know that the convergence of iteration method is linear which can be accelerated by the method known as Aitken's Δ^2-method. By the error equation (15), we have

$$x_{n+1}-\alpha=C(x_n-\alpha)$$

and $$x_n-\alpha=C(x_{n-1}-\alpha)$$

which gives

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

or $$\alpha=x_{n+1}-\frac{(\Delta x_n)^2}{\Delta^2 x_{n-1}}, \qquad \text{...(16)}$$

where Δ is forward difference operator. Forward differences Δx_n and $\Delta^2 x_{n-1}$ can be obtained by constructing forward difference table for three consecutive values of approximations of the root.

Example 1.8 Apply Aitken's Δ^2-method to find a root of the equation

$$\sin^2 x=x^2-1$$

Solution : Given equation can be written as

$$\sin^2 x-x^2+1=0, \text{ so that}$$

$$f(x)=\sin^2 x-x^2+1$$

then $f(1)=0.78$ and $f(2)=-2.173$.

Thus, root of the equation lies in the interval $(1,2)$. Let us take initial approximation $x_0=1.5$. Now the equation can be written as

$$x=\sqrt{1+\sin^2 x}=\phi(x) \qquad \text{(say)}$$

then, $$\phi'(x)=\frac{\sin x\cos x}{\sqrt{1+\sin^2 x}}$$

Since $|\phi'(x)|<1$ in the neighbourhood of $x_0=1.5$, we can take iterative scheme as

$$x_{n+1}=\sqrt{1+\sin^2 x_n},\ n=0,1,2,\ldots\ldots\ldots$$

then

$$x_1=\sqrt{1+\sin^2 x_o}=\sqrt{1+\sin^2(1.5)}=1.41244,$$

$$x_2 = \sqrt{1+\sin^2 x_1} = \sqrt{1+\sin^2(1.41244)} = 1.405394$$

similarly, $x_3 = 1.404596$

After getting three consecutive approximations we can apply Aitken's Δ^2-method. Let us construct forward difference table.

$$\begin{array}{lll} x_1 = 1.412443 & & \\ & \Delta x_1 = -0.007049 & \\ x_2 = 1.405394 & & \Delta^2 x_1 = 0.006251 \\ & \Delta x_2 = -0.000798 & \\ x_3 = 1.404596 & & \end{array}$$

Now, the next iteration is given by

$$x_4^* = x_3 - \frac{(\Delta x_2)^2}{\Delta^2 x_1}$$

$$= 1.404596 - \frac{(-0.000798)^2}{(0.006251)}$$

$$= 1.404494$$

Again using iteration scheme, we get

$$x_5 = \sqrt{1+\sin^2 x_4^*} = \sqrt{1+\sin^2(1.404494)} = 1.404492$$

Thus, the root is 1.40449 correct upto six significant digits.

1.7 System of Non-Linear Equations

Now we shall study the methods for finding the solution of a system of non-linear equations, particularly, system of two equations in two varibales x and y, which can be written as

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

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

The methods can be generalized to a system of n equations in n varibles.

1.7.1 Newton-Raphson method

Let (α, β) be the solution of the system

$$f(x,y) = 0, \; g(x,y) = 0$$

therefore, $f(\alpha,\beta) = 0$, $g(\alpha,\beta) = 0$ ...(17)

Let (x_0, y_0) be an initial approximation to the solution of the system and let

$$\alpha = x_0 + h\,,\ \beta = y_0 + k$$

then by (17), we have

$$f(x_0 + h, y_0 + k) = 0 \text{ and } g(x_0 + h, y_0 + k) = 0 \qquad ...(18)$$

Let us assume that f and g are differentiable. Expanding (18), using Taylor's series, we have

$$f_0 + h\,f_{x_0} + k\,f_{y_0} + ... = 0$$

and $$g_0 + h\,g_{x_0} + k\,g_{y_0} + ... = 0 \qquad ...(19)$$

where $f_0 = f(x_0, y_0)$, $f_{x_0} = \left(\frac{\partial f}{\partial x}\right)_{(x_0, y_0)}$, $f_{y_0} = \left(\frac{\partial f}{\partial y}\right)_{(x_0, y_0)}$ etc.

Neglecting the second and higher order terms in (19), we get

$$h\,f_{x_0} + k\,f_{y_0} = -f_0$$

and $$h\,g_{x_0} + k\,g_{y_0} = -g_0 \qquad ...(20)$$

Solving these two equations for h and k, we get

$$h = \frac{A}{J} \text{ and } k = \frac{B}{J}$$

Where $A = \begin{vmatrix} -f_0 & f_{y_o} \\ -g_0 & g_{y_o} \end{vmatrix}$, $B = \begin{vmatrix} f_{x_0} & -f_0 \\ g_{x_0} & -g_0 \end{vmatrix}$

and $$J = \begin{vmatrix} f_{x_0} & f_{y_0} \\ g_{x_0} & g_{y_0} \end{vmatrix} \neq 0 \qquad ...(21)$$

The next approximation (x_1, y_1) to the solution is given by

$$x_1 = x_0 + h,\ y_1 = y_0 + k \qquad ...(22)$$

Proceeding as above, finding the values A, B and J at (x_1, y_1), we can get increment h and k in x_1 and y_1, so that the new approximation will be given by

$$x_2 = x_1 + h \text{ and } y_2 = y_1 + k$$

This process is to be continued till the required accuracy is obtained.

Example 1.9 : Solve the following system of equations by Newton-Raphson method :

$$y - \sin(x + y) = 0$$

$$x - \cos(y - x) = 0$$

taking initial approximation $x_0 = 1, y_0 = 1$.

Solution : Given that

$$f(x, y) = y - \sin(x + y)$$

and $\quad g(x, y) = x - \cos(y - x)$

then,

$$f_x = \frac{\partial f}{\partial x} = -\cos(x + y), \quad f_y = \frac{\partial f}{\partial y} = 1 - \cos(x + y)$$

$$g_x = \frac{\partial g}{\partial x} = 1 - \sin(y - x), \quad g_y = \frac{\partial g}{\partial y} = \sin(y - x)$$

Here $(x_0, y_0) = (1,1)$, then

$$f_0 = f(x_0, y_0) = 1 - \sin(1 + 1) = 0.090703$$

$$g_0 = g(x_0, y_0) = 1 - \cos(1 - 1) = 0$$

$$f_{x_0} = f_x(x_0, y_0) = -\cos(1 + 1) = 0.416147$$

$$f_{y_0} = f_y(x_0, y_0) = 1 - \cos(1 + 1) = 1.416147$$

$$g_{x_0} = g_x(x_0, y_0) = 1 - \sin(1 - 1) = 1$$

$$g_{y_0} = g_y(x_0, y_0) = \sin(1 - 1) = 0$$

Now,

$$J = \begin{vmatrix} f_{x_0} & f_{y_0} \\ g_{x_0} & g_{y_0} \end{vmatrix} = \begin{vmatrix} 0.416147 & 1.0416147 \\ 1 & 0 \end{vmatrix}$$

$$= -1.416147 \,(\neq 0)$$

$$A = \begin{vmatrix} -f_0 & f_{y_0} \\ -g_0 & g_{y_0} \end{vmatrix} = \begin{vmatrix} -0.090703 & 1.416147 \\ 0 & 0 \end{vmatrix}$$

$$= 0$$

$$B = \begin{vmatrix} f_{x_0} & -f_0 \\ g_{x_0} & -g_0 \end{vmatrix} = \begin{vmatrix} 0.416147 & -0.090703 \\ 1 & 0 \end{vmatrix}$$

$$= 0.090703$$

Then,

$$h = \frac{A}{J} = \frac{0}{-1.416147} = 0$$

and $$k = \frac{B}{J} = \frac{0.090703}{-1.416147} = -0.064049$$

Thus, the first approximation (x_1, y_1) is given by

$$x_1 = x_0 + h = 1 + 0 = 1$$

$$y_1 = y_0 + k = 1 - 0.064049 = 0.935951$$

Now, we have

$$f_1 = f(x_1, y_1) = 0.001882$$

$$g_1 = g(x_1, y_1) = 0.002050$$

$$f_{x_1} = f_x(x_1, y_1) = 0.357094$$

$$f_{y_1} = f_y(x_1, y_1) = 1.357094$$

$$g_{x_1} = g_x(x_1, y_1) = 1.064005$$

$$g_{y_1} = g_y(x_1, y_1) = -0.064005$$

Now,

$$J = \begin{vmatrix} f_{x_1} & f_{y_1} \\ g_{x_1} & g_{y_1} \end{vmatrix} = \begin{vmatrix} 0.357094 & 1.357094 \\ 1.064005 & -0.064005 \end{vmatrix}$$

$$= -1.466811$$

$$A = \begin{vmatrix} -f_1 & f_{y_1} \\ -g_1 & g_{y_1} \end{vmatrix} = \begin{vmatrix} -0.001882 & 1.357094 \\ -0.002050 & -0.064005 \end{vmatrix}$$

$$= 0.002902$$

$$B = \begin{vmatrix} f_{x_1} & -f_1 \\ g_{x_1} & -g_1 \end{vmatrix} = \begin{vmatrix} 0.357094 & -0.001882 \\ 1.0064005 & -0.002050 \end{vmatrix}$$

$$= 0.001270$$

Then,

$$h = \frac{A}{J} = \frac{0.002902}{-1.466811} = -0.001978$$

and $$k=\frac{B}{J}=\frac{0.001270}{-1.466811}=-0.000866$$

Thus, the second approximation (x_2, y_2) is given by

$$x_2 = x_1 + h \; = \; 1-0.001978=0.998022$$

$$y_2 = y_1 + k \; = \; 0.935951=0.000866 \; = \; 0.935085$$

The approximate solution can be taken as

$$x=0.998022, \; y=0.935085$$

1.7.2 Iteration Method

This method is similar to the iteration method (1.6) used for single equation in one variable. The given system of equation $f(x,y)=0$ and $g(x,y)=0$ can be rewritten as

$$x=F(x,y) \text{ and } y=G(x,y) \qquad ...(23)$$

If, in the neighbourhood of the solution of th system,

$$\left|\frac{\partial F}{\partial x}\right|+\left|\frac{\partial F}{\partial y}\right|<1,$$

and $$\left|\frac{\partial G}{\partial x}\right|+\left|\frac{\partial G}{\partial y}\right|<1, \qquad ...(24)$$

then the scheme (23) will be convergent and hence

$$x_{n+1}=F(x_n,y_n), \; n=0,1,2,...$$

$$y_{n+1}=G(x_n,y_n), \; n=0,1,2,... \qquad ...(25)$$

can be taken as iterative scheme to find the solution of the given system.

Example 1.10 Find a real solution of the equations

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

$$3xy^2-10y+7=0$$

taking initial approximation as (0.5, 0.5).

Solution : The given system can be written as

$$x=\frac{1}{5}(x^2+4)=F(x,y)$$

$$y=\frac{1}{10}(3xy^2+7)=G(x,y)$$

so that, $\left|\frac{\partial F}{\partial x}\right|+\left|\frac{\partial F}{\partial y}\right|=\left|\frac{2x}{5}\right|+|0|$

$$= 0.2 \text{ at } x_0 = 0.5, y_0 = 0.5$$

and $\left|\frac{\partial G}{\partial x}\right|+\left|\frac{\partial G}{\partial y}\right|=\left|\frac{3y^2}{10}\right|+\left|\frac{6xy}{10}\right|$

$$= 0.075 + 0.15$$

$$= 0.225 \text{ at } x_0 = 0.5,\ y_0 = 0.5$$

Thus, the condition of convergence is satisfied, so the iterative scheme is

$$x_{n+1} = \frac{1}{5}\left(x_n^2 + 4\right), \qquad n = 0, 1, 2, ...$$

$$y_{n+1} = \frac{1}{10}\left(3x_n y_n^2 + 7\right),\ n = 0,1,2,....$$

Now, the first approximation is given by

$$x_1 = \frac{1}{5}\left(x_0^2 + 4\right) = \frac{1}{5}\left[(0.5)^2 + 4\right] = 0.85$$

$$y_1 = \frac{1}{10}\left(3x_0 y_0^2 + 7\right) = \frac{1}{10}\left[3(0.5)(0.5)^2 + 7\right] = 0.7375$$

Second approximation is given by

$$x_2 = \frac{1}{5}\left(x_1^2 + 4\right) = \frac{1}{5}\left[(0.85)^2 + 4\right] = 0.9445$$

$$y_2 = \frac{1}{10}\left(3x_1 y_1^2 + 7\right) = \frac{1}{10}\left[3(0.85)(0.7375)^2 + 7\right]$$

$$= 0.8387$$

Similarly, we can obtain other approximations as

$x_3 = 0.9784$, $\qquad y_3 = 0.8993$

$x_4 = 0.9915$, $\qquad y_4 = 0.9374$

$x_5 = 0.9966$, $\qquad y_5 = 0.9614$

The solution will converge to $x = 1$, $y = 1$.

Self-Learning Exercise

1. If α is a simple root of the equation $f(x) = 0$, then

(a) $f(x) = (x-\alpha)g(x),\ g(\alpha) = 0$ (b) $f(x-\alpha)g(x),\ g(\alpha) \neq 0$

(c) $f(x) = (x-\alpha)^2 g(x),\ g(\alpha) \neq 0$ (d) $f(x) = (x-\alpha)^2 g(x),\ g(\alpha) = 0$

2. If α is a multiple root of the equation $f(x)=0$ with multiplicity m, then

(a) $f(\alpha)=f'(\alpha)=......=f^{(m-1)}(\alpha)=0$ and $f^{(m-1)}(\alpha)\neq 0$

(b) $f(\alpha)=f'(\alpha)=......=f^{(m-1)}(\alpha)=0$

(c) $f(\alpha)=0$ but no derivate of $f(x)$ at $x=\alpha$ is zero

(d) $f(\alpha)\neq 0$ but all derivatives of $f(x)at$ $x=\alpha$ are zero

3. For an iterative method error equation is given by

(a) $\varepsilon_{n+1}\leq\varepsilon_n^p,\ p>1$ (b) $\varepsilon_{n+1}=\varepsilon_n^p,\ p=1$

(c) $\varepsilon_{n+1}\leq\varepsilon_n^p,\ p<1$ (d) $\varepsilon_{n+1}=\varepsilon_n^p,\ p<1$

4. Secont method is also known as ....

5. Newton-Raphson method is known as ....

6. Write the condition for Newton-Raphson method to be convergent.

7. In Secant method the condition $f(x_n)f(x_{n-1})<0$ is necessary. (True/False)

8. What is the condition for iterative scheme $x=\phi(x)$ so that it becomes convergent?

1.8 Summary

In this unit, we have studied different methods for finding the solution of a non-linear equation and solution of a system of non-linear equations. All the methods were iterative in nature. We have also discussed about the choice of initial approximations and different conditions of convergence for these iterative methods.

1.9 Answers of Self-Learning Exercise

1. (b) 2. (a) 3 (a)

4. Chord method

5. Tangent method

6. Initial approximation should be very close to the exact root.

7. False.

8. $|\phi'(x)|<1$, in the neighbourhood of the root.

1.10 Exercises

1. Find the root of the equation $\log x-\cos x=0$ by Bisection method.

(Ans. 1.42 correct upto two decimal places**)**

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

(Ans. 1.328125 after fifth iteration**)**

3. Find the root of the equation $4\sin x + x^2 = 0$ by secant method.

(Ans. -1.93375 **)**

4. Solve the equation $x\log_{10} x = 1.2$ by Regula-Falsi Method.

(Ans. 2.7406 **)**

5. Find a real root of the equation $x^3 + x^2 - 1 = 0$ by iterative method.

(Ans. 0.755 **)**

6. Find the root of the equation $x^2 - 5x + 2 = 0$ correct to four decimal places by Newton-Raphson method.

(Ans. 0.4384 **)**

7. Find the square root of 8.

(Ans. 2.8284 **)**

8. Find cube root of 10.

(Ans. 2.1544 **)**

9. Find double root of the equation

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

Taking initial approximation $x_0 = 0.9$

(Ans. 1 **)**

10. Find the solution of the system

$$x^2 + y^2 + xy - 7 = 0$$

$$x^3 + y^3 - 9 = 0$$

by taking $(x_0, y_0) = (1.5, 0.5)$ using

(a) Newton-Raphson method

(b) Iteration method.

(Ans. $x = 2.0013$, $y = 0.9987$ **)**

□□□□

Unit - 2 : Chebyshev Method, Muller's Method, Methods for Multiple and Complex Roots

Structure of the Unit

2.0 Objectives

In this unit we shall study Chebyshev method and Muller's method for finding a root of the equation $f(x)=0$. We shall also study Newton-Raphson method for finding multiple root and complex roots of the equation.

2.1 Introduction

Chebyshev method is also known as third order method. Its convergence is three, there fore it is faster than Newton-Raphson method which has convergnence two. In previous unit, we studied the methods in which function $f(x)$, in the neighbourhood of the root, was approximated by a straight line. Muller's method is based on approximating the function in the neighbourhood of the root by a quadratic polynomial. Then the root is approximated by this quadratic polynomial. With the help of Newton-Rahson method, studied in unit-1, we can also find multiple roots and complex roots.

2.2 Chebyshev Method

Let x_n and x_{n+1} be two consecutive approximations to the root $x=\alpha$ of the equation $f(x)=0$. Let

$$x_{n+1}=x_n+h \qquad ...(1)$$

x_{n+1} will be exact root if

$$f(x_{n+1})=0$$

or $$f(x_n+h)=0 \qquad ...(2)$$

Expanding, using Taylor's series, we have

$$f(x_n)+h\,f'(x_n)+\frac{h^2}{2!}f''(x_n)+....=0 \qquad ...(3)$$

Neglecting third and higher power of h, we get

$$f(x_n)+h\,f'(x_n)+\frac{h^2}{2}f''(x_n)=0$$

which gives

$$h=-\frac{f(x_n)}{f'(x_n)}-\frac{h^2}{2}\frac{f''(x_n)}{f'(x_n)} \qquad ...(4)$$

If we neglect the term containing h^2 also, then from (4), we have

$$h=-\frac{f(x_n)}{f'(x_n)} \qquad ...(5)$$

using this value of h in R.H.S. of (4), we have

$$h=-\frac{f(x_n)}{f'(x_n)}-\frac{1}{2}\left[\frac{f(x_n)}{f'(x_n)}\right]^2\frac{f''(x_n)}{f'(x_n)}$$

then by (1), we have

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}-\frac{1}{2}\left[\frac{f(x_n)}{f'(x_n)}\right]^2\frac{f''(x_n)}{f'(x_n)} \qquad ...(6)$$

This is known as **Chebyshev method** of third order or **Newton-Raphson extended formula.** Error equation of this method is given by

$$\varepsilon_{n+1}\le M\varepsilon_n^3$$

where, $M=\left[\frac{f''(\alpha)}{f'(\alpha)}\right]^2-\frac{2}{3}\frac{f'''(\alpha)}{f'(\alpha)}$,

which shows that this method has third order convergence.

Example 2.1 Find the root of the equation $x^4-x-10=0$ using chebyshev method.

Solution : Here, $f(x)=x^4-x-10$,

then $f'(x)=4x^3-1$,

and $f''(x)=12x^2$

Also, since $f(1)=-10$ and $f(2)=4$, root lies in the interval (1,2). Let us take initial approximation as $x_0=1.5$, then

$$f(x_0)=x_0^4-x_0-10=(1.5)^4-(1.5)-10=-6.4375,$$

$$f'(x_0)=4x_0^3-1=4(1.5)^3-1=12.5,$$

and $$f''(x_0)=12x_0^2=12(1.5)^2=27,$$

then by Chebyshev scheme, we have

$$x_1=x_0-\frac{f(x_0)}{f'(x_0)}-\frac{1}{2}\left[\frac{f(x_0)}{f'(x_0)}\right]^2\frac{f''(x_0)}{f'(x_0)}$$

$$=1.5-\frac{(-6.4375)}{12.5}-\frac{1}{2}\left[\frac{-6.4375}{12.5}\right]^2\left(\frac{27}{12.5}\right)$$

$$=1.5+0.515-0.2864$$

$$=1.7286$$

For second iteration, we have

$$f(x_1)=(1.7286)^4-(1.7286)-10=-2.8001,$$

$$f'(x_1)=4(1.7286)^3-1=19.6606,$$

$$f''(x_1)=12(1.7286)^2=35.8567$$

thus,

$$x_2=1.7286-\frac{(-2.8001)}{19.6606}-\frac{1}{2}\left[\frac{-2.8001}{19.6606}\right]^2\left(\frac{35.8567}{19.6606}\right)$$

$$=1.7286+0.1424-0.0185$$

$$=1.8525$$

The third approximation can be obtained as follows

$$f(x_2)=-0.0755,$$

$$f'(x_2)=24.4293$$

$$f''(x_2)=41.1810,$$

Hence, $x_3=1.8525+0.0031-0$

$$=1.8556$$

Similarly, the fourth approximation is $x_4=1.8556$, Thus, the approximate value of the root is 1.8556 correct upto four decimal places.

Example 2.2 Find teh square root of 13 using

(a) Newton-Raphson method

(b) Chebyshev method

Solution : Let $\sqrt{13} = x$, then $x^2 = 13$, that is

$$x^2 - 13 = 0$$

Thus, $f(x) = x^2 - 13$,

$$f'(x) = 2x$$

$$f''(x) = 2$$

Also, $\sqrt{9} < \sqrt{13} < \sqrt{16}$

$\Rightarrow$ $3 < \sqrt{13} < 4$

Thus, we can take initial approximation as $x_0 = 3.5$

(a) By Newton-Raphson Method :

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

$$= x_n - \frac{x_n^2 - 13}{2x_n}$$

$$= \frac{2\,x_n^2 - x_n^2 + 13}{2x_n}$$

Thus,

$$x_{n+1} = \frac{x_n^2 + 13}{2x_n}$$

so that,

$$x_1 = \frac{x_0^2 + 13}{2x_0} = \frac{(3.5)^2 + 13}{2(3.5)} = 3.607143$$

$$x_2 = \frac{x_1^2 + 13}{2x_1} = \frac{(3.607143)^2 + 13}{2(3.607143)} = 3.605551$$

Similarly, third approximation is $x_3 = 3.605551$, so square root of 13 is takes as 3.605551 correct upto six decimal places.

(b) By Chebyshew Method :

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} - \frac{1}{2}\left[\frac{f(x_n)}{f'(x_n)}\right]^2 \frac{f''(x_n)}{f'(x_n)}$$

then, for the first iteration, we have

$$f(x_0) = x_0^2 - 13 = -0.75$$

$$f'(x_0) = 2x_0 = 7$$

$$f''(x_0) = 2$$

and $$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} - \frac{1}{2}\left[\frac{f(x_0)}{f'(x_0)}\right]^2 \frac{f''(x_0)}{f'(x_0)}$$

$$= 3.5 - \frac{(-0.75)}{7} - \frac{1}{2}\left[\frac{(-0.75)}{7}\right]^2 \left(\frac{2}{7}\right)$$

$$= 3.5 + 0.107143 - 0.0016400$$

$$= 3.605503$$

Again, for the second iteration, we have

$$f(x_1) = x_1^2 - 13 = -0.000348$$

$$f'(x_1) = 2x_1^2 = 7.211006$$

$$f''(x_1) = 2$$

Then, $$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} - \frac{1}{2}\left[\frac{f(x_1)}{f'(x_1)}\right]^2 \frac{f''(x_1)}{f'(x_0)}$$

$$= 3.605503 - \frac{(-0.000348)}{7.211006} - \frac{1}{2}\left[\frac{(-0.000348)}{7.211006}\right]^2 \left(\frac{2}{7.211006}\right)$$

$$= 3.605503 + 0.000048 - 0$$

$$= 3.605551$$

Similarly, the third iteration is given by

$x_3 = 3.605551$, hence square root of 13 is 3.605551 correct upto six decimal places.

Example 2.3 Find the root of the equation

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

using chebyshev method and Newton-Raphson method. Compare the results.

Solution : By Chebyshev method :

Here, $f(x) = x^3 - x^2 - x - 1$,

$f'(x) = 3x^2 - 2x - 1$,

$$f''(x) = 6x - 2$$

Also, $f(1) = -2$ and $f(2) = 1$, hence root lies in the interval $(1,2)$. Let us take initial approximation $x_0 = 1.5$.

The first iteration x_1 can be obtained as follows

$$f(x_0) = x_0^3 - x_0^2 - x_0 - 1 = (1.5)^3 - (1.5)^2 - 1.5 - 1 = -1.375$$

$$f'(x_0) = 3x_0^2 - 2x_0 - 1 = 3(1.5)^2 - 2(1.5) - 1 = 2.75$$

$$f''(x_0) = 6x_0 - 2 = 6(1.5) - 2 = 7$$

Substituting the values, we get

$$x_1 = 1.5 + 0.5 - 0.318182$$

$$= 1.681818$$

For second iteration, we have

$$f(x_1) = (1.681818)^3 - (1.681818)^2 - (1.681818) - 1$$

$$= -0.753288$$

$$f'(x_1) = 3(1.681818)^2 - 2(1.681818) - 1$$

$$= 4.121899$$

$$f''(x_1) = 6(1.681818) - 2$$

$$= 8.090908$$

then,

$$x_2 = 1.681818 + 0.182753 - \frac{1}{2}(-0.182753)^2(1.962908)$$

$$= 1.831792$$

Proceeding similarly, we get following approximations

$$x_3 = 1.839287\ ,$$

$$x_4 = 1.839287$$

Therefore, the root is 1.839287 correct upto six decimal places.

By Newton-Raphson Method : Proceeding as above and taking initial approximation as $x_0 = 1.5$, the first iteration is given by

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 1.5 - \frac{(-1.375)}{2.75}$$

$$= 1.5 + 0.5 = 2$$

For second iteration, we have

$$f(x_1) = (2)^3 - (2)^2 - 2 - 1 = 1$$

$$f'(x_1) = 3(2)^2 - 2(2) - 1 = 7$$

then, $x_2 = x_1 - \dfrac{f(x_1)}{f'(x_1)} = 2 - \dfrac{1}{7} = 1.857143$

For third iteration, we have

$$f(x_2) = (1.857143)^3 - (1.857143)^2 - (1.857143) - 1$$

$$= 0.099126$$

$$f'(x_2) = 3(1.857143)^2 - 2(1.857143) - 1$$

$$= 5.632654$$

then,

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} = 1.857143 - \frac{0.099126}{5.632654}$$

$$= 1.839545$$

Proceeding similarly, we get following approximations :

$$x_4 = 1.839287$$

$$x_5 = 1.839287$$

Thus, the root can be takes as 1.839287 correct upto six decimal palces.

We observe that, by chebyshev method we obtain the root in three iterations while by the Newton-Raphson method root of same accuracy is obtained in four iterations.

2.3 Muller's Method

In this method, function $f(x)$ is approximated by a quadratic polynomial in the neighbourhood of the root of the equation $f(x) = 0$. Then the root is approximated by the zero of this interpolating quadratic curve.

To find the quadratic curve we need three points, say (x_i, y_i), (x_{i-1}, y_{i-1}) and (x_{i-2}, y_{i-2}) where $y_i = f(x_i)$ etc. Let the interpolating quadratic curve be

$$f(x) = A(x - x_i)^2 + B(x - x_i) + y_i \qquad ...(7)$$

passing through above three points. Then we have

$$y_{i-1} = A(x_{i-1} - x_i)^2 + B(x_{i-1} - x_i) + y_i \qquad ...(8)$$

and $$y_{i-2} = A(x_{i-2} - x_i)^2 + B(x_{i-2} - x_i) + y_i \qquad ...(9)$$

Solving equation (8) and equation (9), we get

$$A = \frac{A_1}{D} \text{ and } B = \frac{B_1}{D}$$

where, $$A_1 = (x_i - x_{i-2})(y_i - y_{i-1}) - (x_i - x_{i-1})(y_i - y_{i-2})$$

$$A_2 = (x_i - x_{i-1})^2 (y_i - y_{i-1}) - (x_i - x_{i-1})^2 (y_i - y_{i-2})$$

and $$D = (x_i - x_{i-1})(x_i - x_{i-2})(x_{i-1} - x_{i-2})$$

The equation (7) has roots

$$(x - x_i) = \frac{-B \pm \sqrt{B^2 - 4Ay_i}}{2A}$$

when $f(x) = 0$ ...(10)

so the next approximation is given by

$$x_{i+1} = x_i + \frac{-B \pm \sqrt{B^2 - 4Ay_i}}{2A}$$

or $$x_{i+1} = x_i - \frac{2y_i}{B \pm \sqrt{B^2 - 4Ay_i}}$$

or $$x_{i+1} = x_i - \frac{2y_i}{B \pm C} \qquad ...(11)$$

where $C = \sqrt{B^2 - 4Ay_i}$

We select that root which is nearer to x_i. For this, the sign before the radical in (11) is chosen which gives larger magnitude for the denominator. That is, if B is positive we should take positive sign otherwise negative sign should be taken.

The rate of convergence of the method is 1.84.

Example 2.4 Find the root of the equation $x^3 - 2x - 5 = 0$ by Muller's method. Take 1, 2 and 3 as initial approximations.

Solution : Let $y = x^3 - 2x - 5$ and $x_0 = 1$, $x_1 = 2$, $x_2 = 3$

Then $y_0 = -6$, $y_1 = -1$, $y_2 = 16$.

First Iteration :

$$x_2 - x_1 = 1,\ x_2 - x_0 = 2,\ x_1 - x_0 = 1$$

and $y_2 - y_1 = 17$, $y_2 - y_0 = 22$

so that,

$$D = (x_2 - x_0)(x_2 - x_1)(x_1 - x_0)$$

$$= (2)(1)(1) = 2$$

$$A_1 = (x_2 - x_0)(y_2 - y_1) - (x_2 - x_1)(y_2 - y_0)$$

$$= (2)(17) - (1)(22)$$

$$= 12$$

$$A_2 = (x_2 - x_0)^2 (y_2 - y_1) - (x_2 - x_1)^2 (y_2 - y_0)$$

$$= (2)^2 (17) - (1)^2 (22)$$

$$= 46$$

$$A = \frac{A_1}{D} = \frac{12}{2} = 6,\ B = \frac{A_2}{D} = \frac{46}{2} = 23$$

and $C = \sqrt{B^2 - 4Ay_2} = \sqrt{(23)^2 - (4)(6)(16)} = 12.041595$

Since $B > 0$, so

$$x_3 = x_2 - \frac{2y_2}{B + C} = 2.086800$$

Second Iteration : Now, $x_0 = 2$, $x_1 = 3$ and $x_2 = 2.086800$, then $y_0 = -1$, $y_1 = 16$ and $y_2 = -0.086141$.

Also, $x_2 - x_1 = -0.913200$, $x_2 - x_0 = 0.086800$,

$x_1 - x_0 = 1$, $y_2 - y_1 = -16.086141$,

$y_2 - y_0 = 0.913859$

so that

$D = -0.079266$, $A_1 = -0.561741$,

$A_2 = -0.883295$, $A = 7.086784$

$B = 11.143428$, $C = 11.252459$

Since $B > 0$, so

$$x_3 = x_2 - \frac{2y_2}{B+C} = 2.094493$$

Third Iteration : $x_0 = 3$, $x_1 = 2.086800$, $x_2 = 2.094493$

then, $y_0 = 16$, $y_1 = -0.086141$, $y_2 = -0.000653$

and $x_2 - x_1 = 0.007693$, $x_2 - x_0 = -0.905507$,

$x_1 - x_0 = -0.913200$, $y_2 - y_1 = 0.085488$,

$y_2 - y_0 = -16.000653$

so that $D = 0.006361$, $A_1 = 0.045683$

$A_2 = 0.071042$, $A = 7.181732$, $B = 11.168370$

$C = 11.169210$

hence $x_3 = 2.094551$

Thus the approximate root of the equation is 2.0945

Example 2.5 Perform two iterations of Muller's method to find the root of the equation

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

Take $x_0 = -1$, $x_1 = 0.5$, $x_2 = 1$ as initial approximations.

Solution : Let $y = x^3 - x - 1$ and $x_0 = -1$, $x_1 = 0.5$, $x_2 = 1$

then $y_0 = -1$, $y_1 = -1.375$, $y_2 = -1$

First Iteration :

$x_2 - x_1 = 0.5$, $x_2 - x_0 = 2$, $x_1 - x_0 = 1.5$

$y_2 - y_1 = 0.375$, $y_2 - y_0 = 0$

Now,

$$D = (x_2 - x_0)(x_2 - x_1)(x_1 - x_0) = (2)(0.5)(1.5)$$

$$= 1.5,$$

$$A_1 = (x_2 - x_0)(y_2 - y_1) - (x_2 - x_1)(y_2 - y_0)$$

$$= (2)(0.375) - (0.5)(0)$$

$$= 0.750$$

$$A_2 = (x_2 - x_0)^2(y_2 - y_1) - (x_2 - x_1)^2(y_2 - y_0)$$

$$= (4)(0.375) - (0.25)(0)$$

$$= 1.5$$

so that,

$$A = \frac{A_1}{D} = \frac{0.750}{1.5} = 0.5,$$

$$B = \frac{A_2}{D} = \frac{1.5}{1.5} = 1,$$

$$C = \sqrt{B^2 - 4Ay_2} = \sqrt{(1)^2 - (4)(0.5)(-1)}$$

$$= 1.732051$$

Since $B > 0$, therefore x_3 is obtained by the formula

$$x_3 = x_2 - \frac{2y_2}{B+C}$$

$$= 1 - \frac{(2)(-1)}{1 + 1.732051} = 1.732051$$

Second Iteration :

Now, $x_0 = 0.5$, $x_1 = 1$, $x_2 = 1.732051$

and $y_0 = -1.375$, $y_1 = -1$, $y_2 = 2.464103$

$x_2 - x_1 = 0.732051$, $x_2 - x_0 = 1.232051$

$x_1 - x_0 = 0.5$, $y_2 - y_1 = 3.464103$, $y_2 - y_0 = 3.839103$

$D = 0.450962$, $A_1 = 1.457532$, $A_2 = 3.200964$

$A = 3.232051$, $B = 7.098079$ and $C = 4.304219$

hence,

$$x_3 = 1.299839$$

Thus approximate root is 1.299839 after two iterations.

2.4 Newton-Raphson Method for Multiple Root

The multiple root of the equation $f(x) = 0$ with multiplicity m can be obtained using the formula

$$x_{n+1} = x_n - m\frac{f(x_n)}{f'(x_n)} \qquad ...(12)$$

This is called **generalised Newton-Raphson Method.** Multiple root can also be obtained using Newton-Raphson method for simple root, but it will have linear convergence. If the multiple root is obtained by scheme (12), then the convergence will be quadratic as in usual case

Example 2.6 Show that $x = 1$ is a multiple root of equation

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

with multiplicity three.

Solution : Here $f(x) = x^3 - 3x^2 + 3x - 1$,

then $f'(x) = 3x^2 - 6x + 3$,

$f''(x) = 6x - 6$, and $f''(x) = 6$

We observe that, $f(1) = 0$, $f'(1) = 0$, $f''(1) = 0$

but $f'''(1) \neq 0$

Thus, $x = 1$ is a multiple root with multiplicity three.

Example 2.7 : Find a root of the equation

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

with multiplicity 2, taking initial approximation as $x_0 = -0.9$.

Solution : Here $f(x) = x^3 + x^2 - x - 1$

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

Generalized Newton's scheme is

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

for multiplicity $m = 2$. On simplification, we get

$$x_{n+1} = \frac{x_n^3 + x_n + 2}{3x_n^2 + 2x_n - 1},\ n = 0,1,2,\ldots\ldots\ldots$$

Taking $x_0 = -0.9$, we get

$$x_1 = \frac{(-0.9)^3 + (-0.9) + 2}{3(-0.9)^2 + 2(-0.9) - 1} = \frac{0.371}{-0.370}$$

$$= -1.00270$$

The second approximation is given by

$$x_2 = \frac{(-1.0027)^3 + (-1.0027) + 2}{3(-1.0027)^2 + 2(-1.0027) - 1}$$

$$= \frac{-0.010821889}{0.010821870} = -1.00002$$

The third approximation is given by

$$x_3 = \frac{(-1.00002)^3 + (-1.00002) + 2}{3(-1.00002)^2 + 2(-1.00002) - 1}$$

$$= -1$$

Thus, the root is $x = -1$.

Example 2.8 Find double root of the equation

$$x^3 - 0.75x + 0.25 = 0$$

taking initial approximation $x_0 = 0.3$

Solution : Here $f(x) = x^3 - 0.75x + 0.25$

then $f'(x) = 3x^2 - 0.75$

For multiplicity 2, Newton's scheme is

$$x_{n+1} = x_n - \frac{2f(x_n)}{f'(x_n)},\ n = 0,1,2,...........$$

Substituting the values of $f(x)$ and $f'(x)$, on simplification, we get

$$x_{n+1} = \frac{x_n^3 + 0.75x_n - 0.5}{3x_n^2 - 0.75}$$

taking $x_0 = 0.3$, we get

$$x_1 = \frac{(0.3)^3 + (0.75)(0.3) - 0.5}{3(0.3)^2 - (0.75)}$$

$$= 0.516667$$

The second approximation is

$$x_2 = \frac{(0.516667)^3 + (0.75)(0.516667) - 0.5}{3(0.516667)^2 - 0.75}$$

$$= 0.500091$$

Similarly, the third approximation can be obtained which is $x_3 = 0.5$. This value is the correct double root of the given equation.

2.5 Newton-Raphson Method for Complex Roots

A non-linear equation may have complex roots even if all coefficients of the equation are real. Complex roots always occur in pair. The iterative methods like Newton-Raphson method or secant method discussed in unit-1 are applicable to find complex roots provided that complex initial approximation and complex arithmetic are used.

Let $f(z)=0$ is a non-linear equation where z is a complex variable, then

$$f(z)=f(x+iy)=u(x,y)+iv(x,y)=0$$

where u and v are real functions. Thus, the problem of finding a complex root of $f(z)=0$ is equivalent to finding real values x and y by solving system of two non-linear equations

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

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

Methods for solving this sytem have been discussed in unit-1 (section 1.7).

Example 2.9 Find complex root of the equation

$$z^2+1=0$$

By Newton-Raphson method. Use $z_0=\frac{1}{2}(1+i)$ as an initial approximation.

Solution : Here $z=x+iy$, then $z^2+1=0$ gives

$$(x+iy)^2+1=0$$

or $$(x^2-y^2+1)+(2xy)i=0$$

that is, $x^2-y^2+1=0$ and $2xy=0$

Let $f(x,y)=x^2-y^2+1$ and $g(x,y)=2xy$

and initial approximation is

$$z_0=x_0+i\,y_0=\frac{1}{2}(1+i)$$

which gives

$$x_0=0.5\,,\ y_0=0.5$$

that is,

$$(x_o,y_o)=(0.5,0.5)$$

Now $f_x=\dfrac{\partial f}{\partial x}=2x$, $f_y=\dfrac{\partial f}{\partial y}=-2y$

$$g_x = \frac{\partial g}{\partial x} = 2y \text{ , } g_y = \frac{\partial g}{\partial y} = 2x$$

so that,

$$f_0 = f(x_0, y_0) = f(0.5,0.5) = 1$$

$$g_0 = g(x_0, y_0) = g(0.5,0.5) = 0.5$$

similarly,

$$f_{x_0} = f_x(x_0, y_0) = 2(0.5) = 1$$

$$f_{y_0} = f_y(x_0, y_0) = -2(0.5) = -1$$

$$g_{x_0} = g_x(x_0, y_0) = 2(0.5) = 1$$

$$g_{y_0} = g_y(x_0, y_0) = 2(0.5) = 1$$

Now,

$$J = \begin{vmatrix} f_{x_0} & f_{y_0} \\ g_{x_0} & g_{y_0} \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix} = 2$$

$$A = \begin{vmatrix} -f_0 & f_{y_0} \\ -g_0 & g_{y_0} \end{vmatrix} = \begin{vmatrix} -1 & -1 \\ -0.5 & 1 \end{vmatrix} = -1.5$$

and $$B = \begin{vmatrix} f_{x_0} & -f_0 \\ g_{x_0} & -g_0 \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 1 & -0.5 \end{vmatrix} = 0.5$$

then,

$$h = \frac{A}{J} = \frac{-1.5}{2} = -0.75,$$

$$k = \frac{B}{J} = \frac{0.5}{2} = 0.25$$

So the first approximation (x_1, y_1) is given by

$$x_1 = x_0 + h = 0.5 - 0.75 = -0.25$$

$$y_1 = y_0 + k = 0.5 + 0.25 = 0.75$$

Now, for the second approximation, we have

$$f_1 = f(x_1, y_1) = f(-0.25,0.75) = 0.5$$

$$g_1 = g(x_1, y_1) = g(-0.25, 0.75) = -0.375$$

$$f_{x_1} = f_x(x_1, y_1) = 2(-0.25) = -0.5$$

$$f_{y_1} = f_y(x_1, y_1) = -2(-0.75) = -1.5$$

$$g_{x_1} = g_x(x_1, y_1) = 2(0.75) = 1.5$$

$$g_{y_1} = g_y(x_1, y_1) = 2(-0.25) = -0.5$$

Now, $J = \begin{vmatrix} f_{x_1} & f_{y_1} \\ g_{x_1} & g_{y_1} \end{vmatrix} = \begin{vmatrix} -0.5 & -1.5 \\ 1.5 & -0.5 \end{vmatrix} = 2.5$

$$A = \begin{vmatrix} -f_1 & f_{y_1} \\ -g_1 & g_{y_1} \end{vmatrix} = \begin{vmatrix} -0.5 & -1.5 \\ 0.375 & -0.5 \end{vmatrix} = 0.8125$$

$$B = \begin{vmatrix} f_{x_1} & -f_1 \\ g_{x_1} & -g_1 \end{vmatrix} = \begin{vmatrix} -0.5 & -0.5 \\ 1.5 & 0.375 \end{vmatrix} = 0.5625$$

Then,

$$h = \frac{A}{J} = \frac{0.8125}{2.5} = 0.325$$

$$k = \frac{B}{J} = \frac{0.5625}{2.5} = 0.225$$

The second approximation (x_2, y_2) is given by

$$x_2 = x_1 + h = -0.25 + 0.325 = 0.075$$

$$y_2 = y_1 + k = 0.75 + 0.225 = 0.975$$

Proceeding similarly we can obtain the solution (0, 1), that is, exact root $0+i$ or i and its conjugate $-i$ will also be a root.

Self-Learning Exercise

1. Convergence of the chebyshev method is

   (a) 2 (b) 3 (c) 1 (d) 1.62

2. In muller method, root of the equation is approximated by

   (a) Tangent (b) Chord

   (c) Quadratic polynomial (d) Cubic polynomial

3. If multiple root is obtained by Newton-Raphson method for simple root, then the convergence of the method will be

(a) reduced (b) increased (c) remain same (d) oscillates

4. If $1+i$ is a root of some equation then the another root of that equation will be

(a) 1 (b) i (c) $1-i$ (d) $2(1+i)$

2.6 Summary

In this unit we studied some methods to find simple root, multiple root and complex root of a non-linear equation in one variable. For finding a complex root we can use the methods discussed in previous unit to find the solution of system of non-linear equations. Complex roots always occur in pair.

2.7 Answers of Self-Learning Exercise

1. b 2. c

3. a 4. c

2.8 Exercises

1. Find a root of equation $3x-\cos x=1$ by chebyshev method.

(Ans. 0.60710165 **)**

2. Find a root of equation $\cos x - xe^{x}=0$ by chebyshev method.

(Ans. 0.517757**)**

3. Find the root of the equation $x^3-x^2-x-1=0$ using Muller's method, taking initial approximation as $x_0=0$, $x_1=1$, $x_2=2$.

(Ans. 1..839287 **)**

4. Find a double root of the equation $x^3-x^2-x+1=0$ in the neighbourhood of 0.8.

(Ans. 1 **)**

5. Find complex roots of the equation $z^3-2z^2+z-2=0$ taking initial approximation $z_0=0.5+0.5i$.

(Ans. $\pm i$ **)**

Unit - 3 : Solution of Polynomial Equations

Structure of the Unit

3.0 Objectives

In this unit we shall study iterative methods to find the root of a polynomial equation. Birge-Vieta, Bairstow and Graeffe's root squaring methods are applicable only on polynomial equations. We shall also study synthetic division, which is also applicable only on polynomial-equations.

3.1 Introduction

In unit-1, we defined the polynomial equation and the transcendental equation. An equation $f(x)=0$ is said to be algebric or polynomial equation if $f(x)$ is purely a polynomial in varibale x. In previous units, methods studied, were applicable on both algebric and transcendental equations. In this unit we consider the polynomial of degree n in varibale x,

$$f(x)=P_n(x)=a_0x^n+a_1x^{n-1}+a_2x^{n-2}+.....+a_{n-1}x+a_n=0$$

where $a_0\neq 0$ and $a_0,a_1,a_2,.........,a_n$ are real numbers. **Fundamental theorem** states that algebraic or polynomial equation of degree n has exactly n roots.

3.2 Synthetic Division

With the help of synthetic division algorithm we can obtain quotient and remainder on division of a polynomial of degree n by a linear factor. This algorithm can also be used to evaluate polynomials and thier derivatives at given value of x . The process of division is as follows :

Let $P_n(x)$ be a polynomial of degree n such that

$$P_n(x)=a_0x^n+a_1x^{n-1}+a_2x^{n-2}+........a_{n-1}x+a_n \qquad ...(1)$$

with $a_0\neq 0$ and $a_1,a_2,.....a_n$ are real numbers. Let, we have to divide this polynomial by a linear factor $(x-\alpha)$. In this case remainder R (say) will be constant and quotient will be a polynomial of degree

n, say, $Q_{n-1}(x)$. Then

$$P_n(x) = Q_{n-1}(x).(x-\alpha)+R \qquad ...(2)$$

at $x=\alpha$, we have

$$P_n(\alpha) = R \qquad ...(3)$$

This verifies remainder theorem, which states that, when $P_n(x)$ is divided by $(x-\alpha)$, then remainder will be $P_n(\alpha)$

Let $Q_{n-1}(\alpha) = b_o x^{n-1} + b_1 x^{n-2} + \ldots\ldots + b_{n-2}x + b_{n-1}$ ...(4)

then, from (2), we have

$$a_o x^n + a_1 x^{n-1} + a_2 x^{n-2} + \ldots\ldots + a_{n-1}x + a_n$$

$$= \left(b_o x^{n-1} + b_1 x^{n-2} + \ldots\ldots + b_{n-2}x + b_{n-1}\right)(x-\alpha)+R$$

Equating coefficients of like terms in x, we have

$a_0 = b_0$

$a_1 = b_1 - b_0\alpha$,

$a_2 = b_2 - b_1\alpha$, $\qquad (= R)$

....

$a_{n-1} = b_{n-1} - b_{n-2}\,\alpha$

and $\quad a_n = R - b_{n-1}\,\alpha$

on simplification, we get coefficients of $Q_{n-1}(x)$ as follows :

$b_0 = a_0$,

$b_1 = a_1 + a_0\,\alpha$,

$b_2 = a_2 + b_1\,\alpha$,

....

$b_{n-1} = a_{n-1} + b_{n-2}\,\alpha$

and $\quad R = a_n + b_{n-1}\,\alpha$ ...(5)

The above values of b_i and R can be written in tabular form as follows :

α	a_0	a_1	a_2	$\ldots$	a_{n-1}	a_n
		$\alpha\, b_0$	$\alpha\, b_1$	$\ldots$	$\alpha\, b_{n-2}$	$\alpha\, b_{n-1}$
	a_0	$a_1 + \alpha\, b_0$	$a_2 + \alpha\, b_1$	$\ldots$	$a_{n-1} + \alpha\, b_{n-2}$	$a_n + \alpha\, b_{n-1}$
	$(= b_0)$	$(= b_1)$	$(= b_2)$		$(= b_{n-1})$	$(= R)$

Now, we shall derive the process to obtain derivatives of $P_n(x)$ of different order at $x = \beta$. Using definition of $P_n(x)$, we have

$$P_0(x) = a_0,$$

$$P_1(x) = a_0 x + a_1 = x\,P_0(x) + a_1,$$

$$P_2(x) = a_0 x^2 + a_1 x + a_2 = (a_0 x + a_1)x + a_2$$

$$= x\,P_1(x) + a_2,$$

$$\ldots \qquad \ldots. \qquad \ldots.$$

$$P_{n-1}(x) = x\,P_{n-2}(x) + a_{n-1}$$

and $\quad P_n(x) = x\,P_{n-1}(x) + a_n \qquad \ldots(6)$

values of $P_0(x),\ P_1(x), \ldots\ldots\ldots, P_n(x)$ at $x = \beta$ can be obtained by synthetic division algorithm, using above tabular scheme.

Now, differentiating (6) with respect to x, we get

$$P_0'(x) = 0,$$

$$P_1'(x) = x\,P_0'(x) + P_0(x) = P_0(x)$$

$$P_2'(x) = x\,P_1'(x) + P_1(x),$$

$$\ldots \qquad \ldots \qquad \ldots$$

$$P_{n-1}'(x) = x\,P_{n-2}'(x) + P_{n-2}(x)$$

and $\quad P_n'(x) = x\,P_n'(x) + P_{n-1}(x) \qquad \ldots(7)$

This scheme is same as that of synthetic division. Thus $P_1'(x)$, $P_2'(x)$, ..., $P_n'(x)$ at $x = \beta$ can be calculated using the results obtained in (6) and synthetic division algorithm in tabular form. Proceeding in similar manner, we obtain

$$P_0^{(n)}(x) = P_1^{(n)}(x) = \ldots\ldots. = P_{n-1}^{(n)}(x) = 0,$$

and $\quad P_n^{(n)}(x) = n\,P_{n-1}^{(n-1)}(x)$

or $\quad \dfrac{1}{n!} P_n^{(n)}(x) = \dfrac{1}{(n-1)!} P_{n-1}^{(n-1)}(x) \qquad \ldots(8)$

Thus, we obtain following table consisting above results :

β	a_0	a_1	a_2	$\cdots$	a_{n-2}	a_{n-1}	a_n
		βP_0	βP_1	$\cdots$	βP_{n-3}	βP_{n-2}	βP_{n-1}
β	$a_0 = P_0$	P_1	P_2	$\cdots$	P_{n-2}	P_{n-1}	$P_n(\beta)$
		$\beta P_1'$	$\beta P_2'$	$\cdots$	$\beta P_{n-2}'$	$\beta P_{n-1}'$	
β	$P_0 = P_1'$	P_2'	P_3'	$\cdots$	P_{n-1}'	$P_n'(\beta)$	
		$\frac{\beta}{2!} P_2''$	$\frac{\beta}{2!} P_3''$	$\cdots$	$\frac{\beta}{2!} P_{n-1}''$		
β	$P_1' = \frac{1}{2!} P_2''$	$\frac{1}{2!} P_3''$	$\frac{1}{2!} P_4''$	$\cdots$	$\frac{1}{2!} P_n''(\beta)$		
	$\cdots$	$\cdots$	$\cdots.$				
β	$\cdots$	$\cdots$	$\cdots$	$\frac{1}{3!} P_n'''(\beta)$			
	$\cdots$	$\cdots$	$\cdots$				
	$\frac{1}{(n-1)!} P_{n-1}^{(n-1)}(\beta) = \frac{1}{n!} P_n^{(n)}(\beta)$						

Above procedure can be understood easily after going through the following examples.

Example 3.1 Find quotient and remainder on division of polynomial $x^4 - 5x^3 + 6x^2 + 4x - 18$ by a linear factor $(x-2)$. Also verify the result.

Solution : Given polynomial is $x^4 - 5x^3 + 6x^2 + 4x - 18$ so that $a_0 = 1$, $a_1 = -5$, $a_2 = 6$, $a_3 = 6$, $a_4 = 4$, $a_5 = -18$ and linear factor is $(x-2)$, so $\alpha = 2$. The synthetic division procedure will be as follows

2	1	-5	6	4	-18
		2	-6	0	8
	$b_0 = 1$	$b_1 = -3$	$b_2 = 0$	$b_3 = 4$	$-10 = R$

Thus, quotient polynomial $Q_3(x) = x^3 - 3x^2 + 4$ and remainder $R = -10$.

Verification : $Q_3(x)(x-2) + R$

$$= (x^3 - 3x^2 + 4)\,(x-2) - 10$$

$$= x^4 - 3x^3 + 4x - 2x^3 + 6x^2 - 8 - 10$$

$$= x^4 - 5x^3 + 6x^2 + 4x - 18 = \text{given polynomial}$$

Example 3.2 Find all the derivatives of $x^4 - 4x^3 + 8x^2 - 8x + 4$ at $x = 3$, using synthetic division.

Solution : Given that $a_0 = 1$, $a_1 = -4$, $a_2 = 8$, $a_3 = -8$, $a_4 = 4$ and $\beta = 3$. Then the procedure is as follows :

3	1	-4	8	-8	4
		3	-3	15	21
3	1	-1	5	7	$25 = P_4(3)$
		3	6	33	
3	1	2	11	$40 = P_4'(3)$	
		3	15		
3	1	5	$26 = \frac{1}{2!} P_4''(3)$		
		3			
3	1	$8 = \frac{1}{3!} P_4'''(3)$			
	$1 = \frac{1}{4!} P_4^{(iv)}(3)$				

Thus,

$$P_4(3) = 25,$$

$$P_4'(3) = 40$$

$$P_4''(3) = 52,$$

$$P_4'''(3) = 48$$

and $\quad P_4^{(iv)}(3) = 24$

This reulst can be easily verified by actually differentiating the given polynomial.

3.3 Birge-Vieta Method

Let $f(x) = 0$ be the given polynomial equation. Let α_0 be the initial approximation of the root of the equation $f(x) = 0$. To improve the value α_0, we use Newton-Raphosn method as

$$\alpha_1 = \alpha_0 - \frac{P_n(\alpha_0)}{P_n'(\alpha_0)} \qquad \text{or} \qquad \alpha_0 - \frac{f(\alpha)}{f'(\alpha)} \qquad ...(9)$$

where $f(x) = P_n(x)$ and α_1 is improved value of α_0. To obtain the value of $P_n(\alpha_0)$ and $P_n'(\alpha_0)$, we use synthetic division algorithm as discussed earlier. To improve the value α_1, we repeat the process. This process is continued till the required accuracy is achieved or remainder $R = P_n(\alpha)$ becomes zero where R can be obtained on division of $P_n(x)$ by $(x-\alpha)$.

Example 3.3 Find the root of the equation $x^4 - x - 10 = 0$ using Birge-Vieta method. Perform three iterations.

Solution : Here, $P_4(x) = x^4 - x - 10 = f(x)$, then $f(1) = -10$ and $f(2) = 4$, so the root lies between 1 and 2. Let $x_0 = 2$ be the initial approximation, then $f(2)$ and $f'(2)$ can be obtained using synthetic division algorithm as follows :

2	1	0	0	-1	-10
		2	4	8	14
2	1	2	4	7	$f(2)=4$
		2	8	24	
	1	4	12	31 = $f'(2)$	

so $f(2) = 4$ and $f'(2) = 31$ then the first approximation is given by Newton-Raphson method as

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

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

$$= 2 - \frac{4}{31} = 1.871$$

Now, to get $f(1.871)$ and $f'(1.871)$, we use synthetic division as follows :

1.871	1	0	0	– 1	– 10
		1.871	3.501	6.550	10.384
1.871	1	1.871	3.501	5.550	0.384
					$= f(1.871)$
		1.871	7.001	19.649	
	1	3.742	10.502	25.199 $= f'(1.871)$	

The next approximation is given by

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

$$= 1.871 - \frac{0.384}{25.199}$$

$$= 1.856$$

Again, by synthetic division, we have

1.856	1	0	0	– 1	– 10
		1.856	3.445	6.394	10.011
1.856	1	1.856	3.445	5.394	0.011
					$= f(1.856)$
		1.856	6.889	19.180	
	1	3.712	10.334	24.574 $= f'(1.856)$	

Then, the next approximation is given by

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}$$

$$= 1.856 - \frac{0.011}{24.574}$$

$$= 1.856$$

Thus, the root is 1.856 correct upto three decimal places.

Example 3.4 Find a real root of the equation $x^4 + 7x^3 + 24x^2 - 15 = 0$, using Birge-Vieta method. Perform two iterations.

Solution : Hence $P_4(x) = f(x) = x^4 + 7x^3 + 24x^2 - 15$, then $f(0) = -15$ and $f(1) = 17$, so root lies between 0 and 1. Let $x_0 = 0.5$, then

0.5	1	7	24	0	– 15
		0.5	3.75	13.875	6.936
0.5	1	7.5	27.75	13.875	– 8.064
		0.5	4	15.875	
	1	8	31.75	29.75	

So, the next approximation is given by

$$x_1 = 0.5 - \frac{(-8.064)}{29.75}$$

$$= 0.771$$

Again,

0.771	1	7	24	0	– 15
		0.771	5.991	23.123	17.883
0.771	1	7.771	29.991	23.123	2.833
		0.771	6.586	28.201	
	1	8.542	36.577	51.324	

Then, the next approximation is given by

$$x_2 = 0.771 - \frac{2.833}{51.324}$$

$$= 0.716$$

Thus, approximate value of the root, after two iterations, is 0.716.

Example 3.5 Using synthetic division and chebyshev method find a root of the equation $x^3 + x^2 + 3x + 4 = 0$. Perform two iterations.

Solution : Here $P_3(x) = x^3 + x^2 + 3x + 4 = f(x)$, then $f(-1) = 1$, $f(-2) = -6$ so root lies in the interval $(-2,-1)$. Let us take initial approximation $x_0 = -1.5$. We have to obtain $f(-1.5)$, $f'(-1.5)$ and $f''(-1.5)$ for using Chebyshev method. Synthetic division process, to get these values, is as follows :

– 1.5	1	1	3	4
		– 1.5	0.75	– 5.625
– 1.5	1	– 0.5	3.75	– 1.625 $= f(-1.5)$
		– 1.5	3.00	
– 1.5	1	– 2.0	6.75 $= f'(-1.5)$	
		– 1.5		
	1	– 3.5 $= \frac{1}{2!} f''(-1.5)$		

Thus, $f(-1.5) = -1.625$, $f'(-1.5) = 6.75$ and $f''(-1.5) = -7$

The next approximation to the root is given by chebyshev method as follows

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} - \frac{1}{2}\left[\frac{f(x_0)}{f'(x_0)}\right]^2 \left[\frac{f''(x_0)}{f'(x_0)}\right]$$

$$= -1.5 - \frac{(-1.625)}{6.75} - \frac{1}{2}\left[\frac{-1.625}{6.75}\right]^2 \left[\frac{-7}{6.75}\right]$$

$$= -1.5 + 0.24074074 + \frac{1}{2} \times 0.05795610 \times 1.03703704$$

$$= -1.22920795 \approx -1.229$$

Proceeding similarly, we get

-1.229	1	1	3	4
		-1.229	0.281	-4.032
-1.229	1	-0.229	3.281	$-0.032 = f(-1.229)$
		-1.229	1.792	
-1.229	1	-1.458	$5.073 = f'(-1.229)$	
		-1.229		
	1	$-2.687 = \frac{1}{2!} f''(-1.229)$		

Thus, $f(-1.229) = -0.032$, $f'(-1.229) = 5.073$

and $f''(-1.229) = -5.375$,

Then, the next approximation is given by

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} - \frac{1}{2}\left[\frac{f(x_1)}{f'(x_1)}\right]^2 \left[\frac{f''(x_1)}{f'(x_1)}\right]$$

$$= -1.229 - \frac{(-0.032)}{5.073} - \frac{1}{2}\left[\frac{0.032}{5.073}\right]^2 \left[\frac{-5.375}{5.073}\right]$$

$$= -1.229 + 0.00630790 + 0.5 \times 0.00003979 \times 1.05953085$$

$$= -1.22267102 \approx -1.223$$

Thus, the root is -1.22 correct upto two decimal places.

3.4 Bairstow Method

Bairstow method extracts a quadratic factor from polynomial $P_n(x)$, which gives two real roots or a pair of complex roots. Thus, to get complex roots, complex arithmetic can be avoided. In this method we need to extract quadratic factor $(x^2 + px + q)$ from the polynomial $P_n(x)$, where the given equation is $f(x) = 0$ and $f(x) = P_n(x)$. When $P_n(x)$ is divided by two degree polynomial, then quotient will be a polynomial of degree $(n-2)$, say, $Q_{n-2}(x)$ and remainder will be a linear polynomial, say, $Rx + s$. Then

$$P_n(x) = (x^2 + px + q)\, Q_{n-2}(x) + (Rx + s) \qquad ...(10)$$

where $Q_{n-2}(x) = b_0 x^{n-2} + b_1 x^{n-3} + ... + b_{n-3} x + b_{n-2}$

Remainder $(Rx + s)$ will vanish if $(x^2 + px + q)$ is a factor of $P_n(x)$. So we choose p and q such that R and S becomes zero or very small. Thus value of R and S depend on parameters p

and q.

Let

$$R\left(p+\Delta p,\, q+\Delta q\right)=0=S\left(p+\Delta p,\, q+\Delta q\right)$$

Using Taylor's series expansion, we get

$$R\left(p,q\right)+\left(\Delta p\frac{\partial}{\partial p}+\Delta q\frac{\partial}{\partial q}\right)R+\left(\Delta p\frac{\partial}{\partial p}+\Delta q\frac{\partial}{\partial q}\right)^{2}R+\ldots=0$$

and $$S\left(p,q\right)+\left(\Delta p\frac{\partial}{\partial p}+\Delta q\frac{\partial}{\partial q}\right)S+\left(\Delta p\frac{\partial}{\partial p}+\Delta q\frac{\partial}{\partial q}\right)^{2}S+\ldots=0$$

neglecting second and higher order terms, we get

$$R\left(p,q\right)+\frac{\partial R}{\partial p}\,\Delta p+\frac{\partial R}{\partial q}\Delta\, q=0$$

and $$S\left(p,q\right)+\frac{\partial S}{\partial p}\,\Delta p+\frac{\partial S}{\partial q}\Delta\, q=0$$

solving these two equations for Δp and Δq, we get

$$\Delta p=-\frac{R\left(\frac{\partial S}{\partial q}\right)-S\left(\frac{\partial R}{\partial q}\right)}{\left(\frac{\partial R}{\partial p}\right)\left(\frac{\partial S}{\partial q}\right)-\left(\frac{\partial R}{\partial q}\right)\left(\frac{\partial S}{\partial p}\right)}$$

and $$\Delta q=-\frac{S\left(\frac{\partial R}{\partial p}\right)-R\left(\frac{\partial S}{\partial p}\right)}{\left(\frac{\partial S}{\partial q}\right)\left(\frac{\partial R}{\partial p}\right)-\left(\frac{\partial R}{\partial q}\right)\left(\frac{\partial S}{\partial p}\right)} \qquad \ldots(12)$$

Now, by equation (10), we have

$$\left(a_0x^n+a_1x^{n-1}+a_2x^{n-2}+\ldots+a_{n-1}x+a_n\right)$$

$$=\left(x^2+px+q\right)\left(b_0x^{n-2}+b_1x^{n-3}+\ldots+b_{n-3}x+b_{n-2}\right)+\left(R\,x+S\right)$$

comparing the coefficients of like power of x, we get

$a_0=b_0$	or	$b_0=a_0$
$a_1=b_1+p\,b_0$	or	$b_1=a_1-p\,b_0$
$a_2=b_2+p\,b_1+q\,b_0$	or	$b_2=a_2-p\,b_1+q\,b_0$

$$\cdots \qquad \cdots \qquad \cdots$$

$$a_i = b_i + p\,b_{i-1} + q\,b_{i-2} \quad \text{or} \quad b_i = a_i - p\,b_{i-1} + q\,b_{i-2}$$

$$\cdots \qquad \cdots \qquad \cdots$$

$$a_{n-1} = R + p\,b_{n-2} + q\,b_{n-3} \quad \text{or} \quad R = a_{n-1} - p\,b_{n-2} + q\,b_{n-3}$$

and $$a_n = S + q\,b_{n-2} \quad \text{or} \quad S = a_n - q\,b_{n-2} \qquad \text{...(13)}$$

Let us introduce the recursion formula as follows :

$$b_r = a_r - p\,b_{r-1} - q\,b_{r-2},\ b_0 = a_0 \text{ and } b_{-1} = 0 \qquad \text{...(14)}$$

$$r = 1,2,\ldots,n.$$

From equation (13), we get

$$R = b_{n-1} \text{ and } S = b_n + p\,b_{n-1} \qquad \text{...(15)}$$

Now, differentiating (14) and (15) with respect to p and q, we get

$$-\frac{\partial b_r}{\partial p} = b_{r-1} + p\,\frac{\partial b_{r-1}}{\partial p} + q\,\frac{\partial b_{r-2}}{\partial p},$$

$$\frac{\partial b_0}{\partial p} = \frac{\partial b_{-1}}{\partial p} = 0,$$

$$-\frac{\partial b_r}{\partial q} = b_{r-2} + p\,\frac{\partial b_{r-1}}{\partial q} + q\,\frac{\partial b_{r-2}}{\partial q},$$

and $$\frac{\partial b_0}{\partial q} = \frac{\partial b_{-1}}{\partial q} = 0 \qquad \text{...(16)}$$

$$\frac{\partial R}{\partial p} = \frac{\partial b_{n-1}}{\partial p},$$

$$\frac{\partial R}{\partial q} = \frac{\partial b_{n-1}}{\partial q},$$

$$\frac{\partial S}{\partial p} = \frac{\partial b_n}{\partial p} + b_{n-1} + p\,\frac{\partial b_{n-1}}{\partial p},$$

and $$\frac{\partial S}{\partial q} = \frac{\partial b_n}{\partial q} + p\frac{\partial b_{n-1}}{\partial q} \qquad \text{...(17)}$$

From above results, we observe that

$$\frac{\partial b_r}{\partial p} = \frac{\partial b_{r+1}}{\partial q} \qquad \text{...(18)}$$

Let $$-\frac{\partial b_r}{\partial p}=c_{r-1}, \qquad \text{...(19)}$$

then by (18), we have

$$-\frac{\partial b_r}{\partial q}=c_{r-2}, \qquad \text{...(20)}$$

where $r=1,2,....n$.

The value of c_r can be obtained using following recurrence relations

$$c_r=b_r-p\,c_{r-1}-q\,c_{r-2},$$

$$c_{-1}=0,\ c_0=-\frac{\partial b_1}{\partial p}=-\frac{\partial}{\partial p}\left(a_1-pb_0\right)=b_0 \qquad \text{...(21)}$$

using (20), equation (17) gives

$$\frac{\partial R}{\partial p}=\frac{\partial b_{n-1}}{\partial p}=-c_{n-2},\ \frac{\partial R}{\partial q}=\frac{\partial b_{n-1}}{\partial q}=-c_{n-3}$$

$$\frac{\partial S}{\partial q}=-c_{n-1}+b_{n-1}+p\left(-c_{n-2}\right)$$

$$=b_{n-1}-c_{n-1}-p\,c_{n-2}$$

$$\frac{\partial S}{\partial q}=-c_{n-2}+p\left(-c_{n-3}\right)$$

$$=c_{n-2}-p\,c_{n-3} \qquad \text{...(22)}$$

Using relations (12) and (22) values of Δp and Δq are given by

$$\Delta p=\frac{b_{n-1}\,c_{n-2}-b_n\,c_{n-3}}{c_{n-2}^2-c_{n-3}\left(c_{n-1}-b_{n-1}\right)},$$

$$\Delta q=\frac{b_n\,c_{n-2}-b_{n-1}\left(c_{n-1}-b_{n-1}\right)}{c_{n-2}^2-c_{n-3}\left(c_{n-1}-b_{n-1}\right)} \qquad \text{...(23)}$$

Then, the next approximate values of p and q are

$$p_1=p+\Delta p,\ q_1=q+\Delta q \qquad \text{...(24)}$$

These values can be further improved by repeating the process.

The polynomial $Q_{n-2}(x)$ is called deflated polynomial if p and q obtained, are of desired accuracy.

The process of division of a polynomial by a quadratic factor and getting values of b_i and c_i is similar to synthetic division studied earlier and scheme is as follows :

	a_0	a_1	a_2	$\cdots$	a_{n-2}	a_{n-1}	a_n
$-p$	$-$	$-pb_0$	$-pb_1$	$\cdots$	$-pb_{n-3}$	$-pb_{n-2}$	$-pb_{n-1}$
$-q$	$-$	$-$	$-qb_0$	$\cdots$	$-qb_{n-4}$	$-qb_{n-3}$	$-qb_{n-2}$
	b_0	b_1	b_2	$\cdots$	b_{n-2}	b_{n-1}	b_n
$-p$	$-$	$-pc_0$	$-pc_1$	$\cdots$	$-pc_{n-3}$	$-pc_{n-2}$	
$-q$	$-$	$-$	$-qc_0$	$\cdots$	$-qc_{n-4}$	$-qc_{n-3}$	
	c_0	c_1	c_2	$\cdots$	c_{n-2}	c_{n-1}	

Example 3.6 Peform two iterations of Bristow-method to find two roots of the equation

$$x^4 - 3x^3 + 20x^2 + 44x + 54 = 0$$

use $(2, 2)$ as initial approximation.

Solution : Let $p_0 = 2$ and $q_0 = 2$. To find the quadratic factor $(x^2 + px + q)$, we shall find $b_i's$ and $c_i's$ as follows :

	1	-3	20	44	54
-2	$-$	-2	10	-56	4
-2	$-$	$-$	-2	10	-56
	$1 (= b_0)$	$-5 (= b_1)$	$28 (= b_2)$	$-2 (= b_3)$	$2 (= b_4)$
-2	$-$	-2	14	-80	
-2	$-$	$-$	-2	14	
	$1 (= c_0)$	$-7 (= c_1)$	$40 (= c_2)$	$-68 (= c_3)$	

First approximations p_1 and q_1 are given by

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

where, $\Delta p = \dfrac{b_3c_2 - b_4c_1}{c_2^2 - c_1(c_3 - b_3)}$ and $\Delta q = \dfrac{b_4c_2 - b_3(c_3 - b_3)}{c_2^2 - c_1(c_3 - b_3)}$

substituting the values of b_i and c_i, we get

$$\Delta p = \frac{(-2)(40) - (2)(-7)}{(40)^2 - (-7)(-68+2)}$$

$$= \frac{-80 + 14}{1600 - 462}$$

$$= -0.058$$

and $$\Delta q = \frac{2 \times 40 - (-2)(-68+2)}{(40)^2 - (7)(-68+2)}$$

$$= \frac{80 - 132}{1600 - 462}$$

$$= -0.046$$

Thus, $p_1 = p_0 + \Delta p = 2 - 0.058 = 1.942$

$q_1 = q_0 + \Delta q = 2 - 0.046 = 1.954$

Using new values of p and q, performing above scheme, we get

	1	– 3	20	44	54
– 1.942	–	– 1.942	9.597	– 53.683	0.050
– 1.954	–	–	– 1.954	9.657	– 54.014
	1	– 4.492	27.643	– 0.026	0.036
– 1.942	–	– 1.942	13.369	– 75.851	
– 1.954	–	–	– 1.954	13.451	
	1	– 6.884	39.058	– 62.426	

Now,

$$\Delta p = \frac{(-0.026)(39.058) - (0.036)(-6.884)}{(39.058)^2 - (-6.884)(-62.426 + 0.026)}$$

$$= \frac{-1.015508 + 0.247824}{1525.527364 - 429.5616}$$

$$= -0.000700$$

and $$\Delta q = \frac{(0.036)(39.058) - (0.026)(-62.426 + 0.026)}{(39.058)^2 - (-6.884)(-62.426 + 0.026)}$$

$$= \frac{1.406088 - 1.6224}{1525.527364 - 429.5616}$$

$$= -0.000197$$

Thus, the next approximate values of p and q are

$$p_2 = p_1 + \Delta p = 1.942 - 0.000700$$

$$= 1.941300$$

and $$q_2 = q_1 + \Delta q = 1.954 - 0.000197$$

$$= 1.953803$$

Hence, after two iterations, quadratic factor is $x^2+1.9413x+1.953803$. Solving this equation, we get

$$x=\frac{-1.9413\pm\sqrt{(1.9413)^2-4\times1.953803}}{2}$$

$$=\frac{-1.9413\pm\sqrt{-4.04656631}}{2}$$

$$=-0.97065\pm1.005804i$$

Thus, $(-0.97065+1.005804i)$ and $(-0.97065-1.005804i)$ are two roots of the given equation.

Example 3.7 Extract quadratic factor from the equation $x^3-2x+x-2=0$ using Bairstow method and hence find the roots of the equation. Perform only two iterations and use $(-0.5,\,1)$ as initial approximation.

Solution : Let required quadratic factor be (x^2+px+q) and let $p_0=-0.5$, $q_0=1$.

	1	− 2	1	− 2
0.5	–	0.5	− 0.75	− 0.375
− 1	–	–	− 1	1.5
	1	− 1.5	− 0.75	− 0.875
0.5	–	0.5	− 0.5	
− 1	–	–	− 1	
	1	− 1	− 2.25	

Now,

$$\Delta p=\frac{b_2c_1-b_3c_0}{c_1^2-c_0(c_2-b_2)}$$

$$=\frac{(-0.75)(-1)-(-0.875)\times1}{(-1)^2-1\,(-2.25+0.75)}$$

$$=\frac{0.75+0.875}{1+1.5}$$

$$=0.65$$

$$\Delta q=\frac{b_3c_1-b_2(c_2-b_2)}{c_1^2-c_0(c_2-b_2)}$$

$$= \frac{(-0.875)(-1)+(0.75)(-2.25+0.75)}{(-1)^2-1(-2.25+0.75)}$$

$$= \frac{0.875-1.125}{1+1.5} = -0.1$$

The next approximation is given by

$$p_1 = p_0 + \Delta\, p = -0.5 + 0.65 = 0.15$$

and $\quad q_1 = q_0 + \Delta\, q = 1 - 0.1 = 0.9$

For next approximation, we have

	1	−2	1	−2
−0.15	–	−0.15	0.323	−0.063
−0.9	–	–	−0.9	1.935
	1	−2.15	0.423	−0.128
−0.15	–	−0.15	0.345	
−0.9	–	–	−0.9	
	1	−2.30	−0.132	

then

$$\Delta\, p = \frac{(0.423)(-2.30)-(-0.128)(1)}{(-2.30)^2-1(-0.132-0.423)}$$

$$= \frac{-0.9729+0.128}{5.29+0.555} = -0.144551$$

$$\Delta\, q = \frac{(-0.128)(-2.30)-(0.423)(-0.132-0.423)}{(-2.30)^2-1(-0.132-0.423)}$$

$$= \frac{0.2944+0.234765}{5.29+0.555} = 0.090532$$

Thus, the next arppoximation is given by

$$p_2 = p_1 + \Delta\, p = 0.15 - 0.144551$$

$$= 0.005449$$

$$q_2 = q_1 + \Delta\, q = 0.9 + 0.090532$$

$$= 0.990532$$

Thus, the required quadratic factor is $\left(x^2 + 0.005449x + 0.990532\right)$

3.5 Graeffe's Root Squaring Method

This method is also applicable for the numerical solution of polynomial equations. In this method given polynomial is transformed into another polynomial of same degree but whose roots are the squares of the roots of the original polynomial. This process of squaring the roots, is repeated m times so that the roots of the new polynomial are the 2^m power of the roots of the original polynomial equation. Also, roots of new polynomial are widely separated for large m provided the roots of the original polynomial are real and distinct.

Advantage of this method is that all the roots of the given polynomial equation can be obtained at a time and no initial approximation is required.

Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the real and distinct roots of the given polynomial equation

$$a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \ldots + a_{n-1} x + a_n = 0 \qquad \ldots(25)$$

where $a_0 \neq 0$ and $a_1, a_2, \ldots, a_n$ are real.

equation (25) can be re-written as

$$\left(a_0 x^n + a_2 x^{n-2} + a_4 x^{n-4} + \ldots\right)^2 = \left(a_1 x^{n-1} + a_3 x^{n-3} + a_5 x^{n-5} + \ldots\right)^2$$

on simplification, we get

$$a_0^2 x^{2n} - \left(a_1^2 - 2a_0 a_2\right) x^{2n-2} + \left(a_2^2 - 2a_1 a_3 + 2a_0 a_4\right) x^{2n-4} - \ldots + (-1)^n a_n^2 = 0$$

Let $y = -x^2$, then we have

$$Q_n(y) = b_0 y^n + b_1 y^{n-1} + b_2 y^{n-2} + \ldots + b_n = 0 \qquad \ldots(26)$$

where,

$$b_0 = a_0^2$$

$$b_1 = a_1^2 - 2a_0 a_2$$

$$b_2 = a_2^2 - 2a_1 a_3 + 2a_0 a_4$$

$$\ldots \qquad \ldots \qquad \ldots$$

$$b_n = a_n^2$$

Root of the equation (26) are $-\alpha_1^2, -\alpha_2^2, \ldots, -\alpha_n^2$.

This procedure is repeated m times and the polynomial obtained after m steps be

$$A_0 z^n + A_1 z^{n-1} + A_2 z^{n-2} + \ldots + A_{n-1} z + A_n = 0 \qquad \ldots(27)$$

Roots of this equation will be $q_1, q_2, \ldots, q_n$ such that

$$\alpha_i = |q_i|^{1/2^m}, \; i = 1, 2, \ldots, n$$

By the theory of equations, we have

$$q_i = -\frac{A_i}{A_{i-1}},\ i = 1,2,\ldots,n$$

Thus, roots α_i can be estimated by the value

$$\left|\frac{A_i}{A_{i-1}}\right|^{1/2^m},\ i = 1,2,\ldots n \qquad \ldots(28)$$

The sign of the root of the original equation can be determined by substituting in the original equation.

Working table to obtained coefficients of new polynomial is as follows :

a_0	a_1	a_2	a_3	a_4	...	a_n
a_0^2	a_1^2	a_2^2	a_3^2	a_4^2	...	a_n^2
	$-2\,a_0\,a_2$	$-2\,a_1\,a_3$	$-2\,a_2\,a_4$	$-2\,a_3\,a_5$	...	
		$2\,a_0\,a_4$	$2\,a_1\,a_5$	$2\,a_2\,a_6$	...	
			$-2\,a_0\,a_6$	$-2\,a_1\,a_7$	...	
			....	....	....	
b_0	b_1	b_2	b_3	b_4	....	b_n

Complext Roots : Roots α_k and α_{k+1} of the given polynomial equation will be a pair of complex root if the coefficient of x^{n-k} in the successive squaring fluctuate in sign and magnitude.

Let this pair be $\alpha \pm i\,\beta$, then from (28), we have

$$\alpha_k = \alpha + i\,\beta = \left|\frac{A_k}{A_{k-1}}\right|^{1/2^m}$$

and $$\alpha_{k+1} = \alpha - i\,\beta = \left|\frac{A_{k+1}}{A_k}\right|^{1/2^m}$$

then, $$\alpha^2 + \beta^2 = \left|\frac{A_{k+1}}{A_{k-1}}\right|^{1/2^m} \qquad \ldots(29)$$

and sum of all roots, when there is only one pair of complex root, is

$$\alpha_1 + \alpha_2 + \ldots + (\alpha + i\,\beta) + (\alpha - i\,\beta) + \alpha_{k+2} + \ldots + \alpha_n = -\frac{a_1}{a_0}$$

or $$\alpha_1+\alpha_2+...+2\alpha+\alpha_{k+2}+...+\alpha_n=-\frac{a_1}{a_0} \qquad ...(30)$$

thus α and β can be obtained from equations (29) and (30).

Double root : The root α_k is a double root of the given equation if the mangitude of the coefficient A_k is nearly equal to half the square of the magnitude of the corresponding coefficient in the previous equation, thus,

$$q_k=-\frac{A_k}{A_{k-1}}$$

and $$q_{k+1}=-\frac{A_{k+1}}{A_k}$$

then $$q_k.q_{k+1}\approx q_k^2$$

$$\approx\left|\frac{A_{k+1}}{A_{k-1}}\right|$$

Hence, $$|\alpha_k|=\left|\frac{A_{k+1}}{A_{k-1}}\right|^{1/2^{m+1}} \qquad ...(31)$$

sign of this root can be determined by substituting it in the given equation.

Example 3.8 Find all the roots of the equation $x^3-6x^2+11x-6=0$ using Graeffe's root squaring method.

Solution : The coefficients of the successive squaring can be obtained as follows :

m	A_0	A_1	A_2	A_3
0	1	−6	11	−6
	1	36 −22	121 −72	36
1	1	14	49	36
	1	196 −98	2401 −1008	1296
2	1	98	1393	1296
	1	9604 −2786	1940449 −254016	1679616
	1	6818	1686433	1679616

After three squarings, we get

$$A_0 = 1,\ A_1 = 6818,\ A_2 = 1686433,\ A_3 = 1679616$$

then, for $m = 3$, we have

$$|\alpha_1| = \left|\frac{A_1}{A_0}\right|^{1/2^3}$$

$$= \left|\frac{6818}{1}\right|^{1/8}$$

$$= 3.014443,$$

$$|\alpha_2| = \left|\frac{A_2}{A_1}\right|^{1/2^3}$$

$$= \left|\frac{1686433}{6818}\right|^{1/8}$$

$$= 1.991425$$

and $$|\alpha_3| = \left|\frac{A}{A_2}\right|^{1/2^3}$$

$$= \left|\frac{1679616}{1686433}\right|^{1/8}$$

$$= 0.999494$$

Sign of the roots can be determined by substitution of these values in the given equation, which gives

$$f(3.014443) \approx 0,\ f(1.991425) \approx 0 \text{ and } f(0.999494) \approx 0$$

Thus approximate value of the roots are

$$3.014443,\ 1.991425 \text{ and } 0.999494$$

These values converges to the exacts roots 3, 2 and 1 after some more squarings.

Example 3.9 Find all the roots of the equation $x^4 - 3x + 1 = 0$

using Graeffe's root squaring method. Use four squaring to estimate roots.

Solution : The coefficients of successive squarings can be obtained as follows :

m	A_0	A_1	A_2	A_3	A_4
0	1	0	0	−3	1
	1	0 0	0 0 2	9 0	1
1	1	0	2	9	1
	1	0 −4	4 0 2	81 −4	1
2	1	−4	6	77	1
	1	16 −12	36 616 2	5929 −12	1
3	1	4	654	5917	1
	1	16 −1308	427716 − 47336 2	35010889 −1308	1
4	1	−1292	380382	35009581	1

The coefficients given in column A_1, change sign alternatively, therefore there exists a pair of complex roots, α_1 and α_2, such that

$$\alpha_1 = \alpha + i\beta, \quad \alpha_2 = \alpha - i\beta$$

then $$\alpha^2 + \beta^2 = \left|\frac{A_2}{A_0}\right|^{1/2m}$$

$$= \left|\frac{380382}{1}\right|^{1/16}$$

or $$\alpha^2 + \beta^2 = 2.232358 \qquad \text{...(i)}$$

other roots are given by

$$|\alpha_3| = \left|\frac{A_3}{A_2}\right|^{1/2m}$$

$$= \left| \frac{35009581}{380382} \right|^{1/16}$$

$$= 1.326624$$

and $|\alpha_4| = \left| \frac{A_4}{A_3} \right|^{1/2m}$

$$= \left| \frac{1}{35009581} \right|^{1/16}$$

$$= 0.337667$$

substituting these values in given equation, we get

$$f(1.326624) \approx 0 \text{ and } f(0.337667) \approx 0$$

thus $\alpha_3 = 1.326624$, $\alpha_4 = 0.337667$

Now, sum of roots is given by

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = -\frac{a_1}{a_0}$$

$\Rightarrow \quad (\alpha + i\beta) + (\alpha - i\beta) + 1.326624 + 0.337667 = 0$

$\Rightarrow \quad 2\alpha + 1.664291 = 0$

$\Rightarrow \quad \alpha = -0.8321455$ ...(ii)

Using relation (i), we get

$$(-0.8321455)^2 + \beta^2 = 2.232358$$

$\Rightarrow \quad \beta^2 = 1.539892$

$\Rightarrow \quad \beta = 1.240924$

Thus, pair of complex roots are

$$(-0.8321455 \pm 1.240924\, i)$$

Self-Learning Exercise

1. Synthetic division can be applied on transcendental equations. (True/False)

2. Quadratic factor can be extracted from the given polynomial using the method :

(a) Birge-Vieta method (b) Graeffe's root squaring method

(c) Bairstow method (d) Chebyshev method

3. All roots of the given polynomial equation can be obtained at a time by

(a) Birge-Vieta method (b) Newton-Raphson method

(c) Graeffe's root squaring method (d) Bairstow method

4. By Graeffe's root squaring method, we can find complex root also. (True/False)

5. Which of the following method, we do not need any information about initial approximation

(a) Newton-Raphson method (b) Graeffe's root squaring method

(c) Bairstow method (d) Birge-Vieta method

3.6 Summary

In this unit, we have studied the methods which are applicable only on polynomial or algebraic equations. All the methods require initial approximation to the root except Graeffe's root squaring method. In Graeffe's root squaring method, we must carefully observe the successive coefficients of each column, so that we can determine the nature of roots, whether they are real or complex and simple or doule root.

3.7 Answers of Self-Learning Exercise

1. False 2. (c) 3. (c)

4. True 5. (b)

3.8 Exercises

1. Divide $x^5 - 2x^4 + 2x^2 + 4x - 1$ by $(x-3)$ using synthetic division and find quotient polynomial and remainder.

[Ans. $Q_4(x) = x^4 + x^3 + 3x^2 + 12x + 40$ and $R = 119$]

2. Use synthetic division and perform two iterations of the Birge-Vieta method to find the smallest positive root of the equation

$$2x^3 - 5x + 1 = 0$$

Take initial approximation as 0.5.

[Ans. 0.202630]

3. Find a real root of $x^3 - x^2 - x - 1 = 0$ near $x = 2$ using Birge-Vieta method.

[Ans. 1.839 correct upto three decimal places]

4. Find a quadratic factor of the polynomial

$$P_4(x) \equiv x^4 + 5x^3 + 3x^2 - 5x - 9 = 0$$

by Bairstow method. Take initial approximation $(3,-5)$

[Ans : $x^2+2.90255x+4.91759$, after two iterations]

5. Solve $x^4-5x^3+20x^2-40x+60=0$ using Bairstow method, taking initial approximations as $(-4,8)$

   [Ans : $x^2-3.83x+7.3064$ and roots are $1.9149 \pm 1.9077i$]

6. Find all the roots of the polynomial equation

   $$x^3-3x^2-6x+8=0$$

   using Graeffe's root squaring method.

   [Ans : 4, 2, 1]

7. Using Graeffe's root squaring method, find all the roots of the equation

   $$x^4-3x^3-3x^2+11x-6=0$$

   [Ans : $3,2,-1,-1$, Hint : After fourth squaring A_3 is half of the square of the corresponding coefficient at third squaring, therefore there exists a pair of double root, i.e., $\alpha_3=\alpha_4$]

8. Find all roots of the equation $x^3-2x^2-5x+6=0$ by Graeffe's root squaring method.

   [Ans : $3,-2,1$ after three squaring]

□□□

Unit - 4 : System of Simultaneous Equations

Structure of the Unit

4.0 Objectives

The objective of this unit is to find the solutions of linear system of equations which is the most important use of matrices. These systems of equations are frequently used in frameworks, electrical networks, traffic flow, production and consumption, assignment of jobs of workers, population growth, statistics and many others. The practical problems can be modeled to the system of equations and hence solved by the methods discussed in this unit.

4.1 Introduction

Some important methods are discussed in this chapter to solve the system of linear equations. This unit is basically consisting of two parts viz. direct methods and iterative methods. In direct methods the coefficient matrix is reduced either in diagonal forms or upper and lower triangular forms and hence the system can easily be solved. In iterative methods initial approximations are proposed and hence get the better approximations. Ideas of basic cocepts of scalar multiplication, determinants, Gauss elimination method are the prerequisities for better understanding of this matter.

4.2 Direct Methods

The first classifications is the direct method in which we proceed through a finite number of steps and produce an exact solution.

Let the given system of linear equations having the form

$$a_{11}x_1 + a_{12}x_2 + \ldots\ldots\ldots\ldots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \ldots\ldots\ldots\ldots + a_{2n}x_n = b_2$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a_{n1}x_1 + a_{n2}x_2 + \ldots\ldots\ldots\ldots + a_{nn}x_n = b_n \qquad \ldots(1)$$

This system of n equations cosisting n number of unknowns i.e. $x_1, x_2, \ldots x_n$. The elements a_{ij} and b_i are prescribed real numbers. The above system of equations can be written in the matrix notation as

$$\begin{bmatrix} a_{11} & a_{12}\ldots\ldots\ldots\ldots a_{1n} \\ a_{21} & a_{22}\ldots\ldots\ldots\ldots a_{2n} \\ \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots \\ a_{n1} & a_{n2}\ldots\ldots\ldots\ldots a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ \ldots \\ x_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \ldots \\ \ldots \\ b_n \end{bmatrix} \qquad \ldots(2)$$

The first matrix is known as coefficient matrix and denoted by A, the second, matrix of unknowns is denoted by X and RHS column matrix is denoted by B.So that the system now simply written as

$$\mathrm{AX} = \mathrm{B} \qquad \ldots(3)$$

4.2.1 Method of Determinants

This method is introduced by Gabriel Cramer, so it is also known as Cramer's rule. The method is quite easy. Let us consider the set of equations

$$a_1x + b_1y + c_1z = d_1$$

$$a_2x + b_2y + c_2z = d_2$$

$$a_3x + b_3y + c_3z = d_3 \qquad \ldots(4)$$

This system can also be written as

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

Now, the determinant of the coefficient matrix

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \qquad \ldots(5)$$

Then

$$x\,\Delta = \begin{vmatrix} xa_1 & b_1 & c_1 \\ xa_2 & b_2 & c_2 \\ xa_3 & b_3 & c_3 \end{vmatrix}, \text{(Operating } C_1 + yC_2 + zC_3\text{)}$$

$$= \begin{vmatrix} a_1x+b_1y+c_1z & b_1 & c_1 \\ a_2x+b_2y+c_2z & b_2 & c_2 \\ a_3x+b_3y+c_3z & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}$$

This implies

$$x = \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix} \div \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}, \text{ Provided } \Delta \neq 0 \qquad ...(6)$$

Similarly

$$y = \begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix} \div \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \qquad ...(7)$$

and

$$z = \begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix} \div \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \qquad ...(8)$$

These three equations giving the values of x, y, z constitute the Cramer's rule, which reduces the solution of the linear equations (4) to a problem in evaluation of determinants.

When matrix of coefficients is singular i.e. $\Delta = 0$, the above method is failed.

Example 4.1 Solve the given system of the equations using the method of determinants

$$3x + y + 2z = 3$$
$$2x - 3y - z = -3$$
$$x + 2y + z = 4$$

Solution : The given system can again be written as the matrix form

$$\begin{bmatrix} 3 & 1 & 2 \\ 2 & -3 & -1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ -3 \\ 4 \end{bmatrix}$$

Hence the determinant of the coefficient matrix is

$$\Delta = \begin{vmatrix} 3 & 1 & 2 \\ 2 & -3 & -1 \\ 1 & 2 & 1 \end{vmatrix} = 8$$

Now, as we discussed above

$$x = \frac{1}{8}\begin{vmatrix} 3 & 1 & 2 \\ -3 & -3 & -1 \\ 4 & 2 & 1 \end{vmatrix}$$

$$x = \frac{1}{8}\left[3(-3+2) - (-3+4) + 2(-6+12)\right] = 1$$

Similarly

$$y = \frac{1}{8}\begin{vmatrix} 3 & 3 & 2 \\ 2 & -3 & -1 \\ 1 & 4 & 1 \end{vmatrix} = 2 \quad \text{and} \quad z = \frac{1}{8}\begin{vmatrix} 3 & 1 & 3 \\ 2 & -3 & -1 \\ 1 & 2 & 4 \end{vmatrix} = -1$$

Hence $x = 1;\ y = 2;\ z = -1$

Self-Learning Exercise - 1

1. Using the method of determinant solve the given system of equations

$$x + y + z = 3$$
$$2x - y + z = 2$$
$$x - 2y + 3z = 2$$

2. Using the method of determinant solve the given system of equations

$$2x - 3y + 5z = 11$$
$$3x + 2y - 4z = -5$$
$$x + y - 2z = -3$$

4.2.2 Gauss Jordan Method

This method is a modified form of Gauss elimination method. In this method, the coefficient matrix is reduced to a diagonal matrix. Although this method involves more arithmetical operations, the back substitution is not required, as each of the final equivalent system of equations will contain only one unknown. Let us understand the method better by the following illustration.

Example 4.2 Solve the following linear equations

$$2x_1 + 8x_2 + 2x_3 = 14$$
$$6x_1 + 6x_2 - x_3 = 13$$
$$2x_1 + x_2 + 2x_3 = 5$$

using Gauss-Jordan method.

Solution : The given system is

$$2x_1 + 8x_2 + 2x_3 = 14$$
$$6x_1 + 6x_2 - x_3 = 13$$
$$2x_1 - x_2 + 2x_3 = 5$$

In matrix notation the system can be written as

$$\begin{bmatrix} 2 & 8 & 2 \\ 6 & 6 & -1 \\ 2 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 13 \\ 5 \end{bmatrix} \qquad \text{...(1)}$$

Eliminating x_1 from second and third rows of(1), we get

$$\begin{bmatrix} 2 & 8 & 2 \\ 0 & -18 & -7 \\ 0 & -9 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ -29 \\ -9 \end{bmatrix} \qquad \text{...(2)}$$

Eliminating x_2 from first and third rows using the second row, we get

$$\begin{bmatrix} 2 & 0 & -10/9 \\ 0 & -18 & -7 \\ 0 & 0 & 7/2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10/9 \\ -29 \\ 11/2 \end{bmatrix} \qquad \text{...(3)}$$

Eliminating x_3 from first and second rows, we get

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & -18 & 0 \\ 0 & 0 & 7/2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 20/7 \\ -18 \\ 11/2 \end{bmatrix} \qquad \text{...(4)}$$

From equation (4), we get the solution as

$$x_1 = \frac{10}{7},\ x_2 = 1,\ x_3 = \frac{11}{7}.$$

Example 4.3 Using the Gauss-Jordan method solve the following linear equations

$$\begin{aligned} &10x + y + z = 12 \\ &2x + 10y + z = 13 \\ &x + y + 5z = 7 \end{aligned}$$

Solution : In matrix form the given system can be written as

$$\begin{bmatrix} 10 & 1 & 1 \\ 2 & 10 & 1 \\ 1 & 1 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 12 \\ 13 \\ 7 \end{bmatrix}$$

and the augmented matrix will be

$$[A|B] = \begin{bmatrix} 10 & 1 & 1:12 \\ 2 & 10 & 1:13 \\ 1 & 1 & 5:7 \end{bmatrix}$$

Operating $R_{13}(-9)$

$$\approx\begin{bmatrix}1 & -8 & 44 & -51\\ 2 & 10 & 1 & 13\\ 1 & 1 & 5 & 7\end{bmatrix}$$

Operating $R_{21}(-2)$ and $R_{31}(-1)$

$$\approx\begin{bmatrix}1 & -8 & -44 & -51\\ 0 & 26 & 89 & 115\\ 0 & 9 & 49 & 58\end{bmatrix}$$

Operating $R_2 - 3R_3$

$$\approx\begin{bmatrix}1 & -8 & -44 & -51\\ 0 & -1 & -58 & -59\\ 0 & 9 & 49 & 58\end{bmatrix}$$

Operating $R_2(-1)$

$$\approx\begin{bmatrix}1 & -8 & -44 & -51\\ 0 & 1 & 58 & 59\\ 0 & 9 & 49 & 58\end{bmatrix}$$

Operating $R_{32}(-9)$ and $R_{12}(8)$

$$\approx\begin{bmatrix}1 & 0 & 420 & 421\\ 0 & 1 & 58 & 59\\ 0 & 9 & -473 & -473\end{bmatrix}$$

Operating $R_3(-1/473)$

$$\approx\begin{bmatrix}1 & 0 & 420 & 421\\ 0 & 1 & 58 & 59\\ 0 & 0 & 1 & 1\end{bmatrix}$$

Operating $R_{13}(-420)$ and $R_{23}(-58)$

$$\approx\begin{bmatrix}1 & 0 & 0 & 1\\ 0 & 1 & 0 & 1\\ 0 & 0 & 1 & 1\end{bmatrix}$$

Therefore the equivalent system of equation is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Obviously the solution will be

$$x = 1,\ y = 1,\ z = 1$$

Self-Learning Exercise - 2

1. Solve the system of simultaneous equations using Gauss-Jordan method

$$\begin{aligned} 10x + 2y + z &= 9 \\ 2x + 20y - 2z &= -44 \\ -2x + 3y + 10z &= 22 \end{aligned}$$

2. Use Gauss-Jordan method to solve the system

$$\begin{aligned} 10x + y + z &= 12 \\ x + 10y + z &= 12 \\ x + y + 10z &= 12 \end{aligned}$$

4.2.3 Methods of Decomposition (LU Method)

This method is also known as the method of factorization or Triangularization method. In this method the coefficient matrix A of the system of the equation $\mathrm{AX} = \mathrm{B}$ is decomposed into the product of a lower triangular matrix L and upper triangular matrix U so that

$$\mathrm{A} = \mathrm{LU} \qquad ...(9)$$

where

$$\mathrm{L} = \begin{bmatrix} l_{11} & 0 & 0 & \Lambda & 0 \\ l_{12} & l_{22} & 0 & \Lambda & 0 \\ l_{13} & l_{23} & l_{33} & \Lambda & 0 \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} \\ l_{1n} & l_{2n} & l_{3n} & \Lambda & l_{nn} \end{bmatrix} \text{ and } \mathrm{U} = \begin{bmatrix} u_{11} & u_{21} & u_{31} & \Lambda & u_{n1} \\ 0 & u_{22} & u_{32} & \Lambda & u_{n2} \\ 0 & 0 & u_{33} & \Lambda & u_{n3} \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} \\ 0 & 0 & 0 & \Lambda & u_{nn} \end{bmatrix}$$

Multiplying the matrices L and U and comparing the elements of the resulting matrix with those of A, we obtain a system of equation in unknowns l_{ij} and u_{ij}, where $l_{ij} = 0$, $j > i$ and $u_{ij} = 0$, $i > j$, in other words

$$l_{i1}u_{1j} + l_{i2}u_{2j} + \ldots\ldots + l_{in}u_{nj} = a_{ij},\ 1 \le j \le n$$

where $l_{ij} = 0$, $j > i$ and $u_{ij} = 0$, $i > j$

To produce a unique solution, it is convenient to choose either $u_{ii} = 1$ or $l_{ii} = 1$; $1 \le i \le n$.

(i) When we choose $l_{ii} = 1$, the method is called the **Doolittle's method.**

(ii) When we choose $u_{ii} = 1$, the method is called the **Crout's method.**

(iii) When $U = L^T$ so that $l_{ii} = u_{ii}$; $1 \le i \le n$, then it is called **Cholesky's factorization.**

The given system of equation is

$$\mathrm{AX} = \mathrm{B}$$

$$\Rightarrow \quad \mathrm{LUX} = \mathrm{B} \qquad ...(10)$$

Let $\quad \mathrm{UX} = \mathrm{Y} \qquad ...(11)$

Then equation (10) becomes

$$\mathrm{LY} = \mathrm{B} \qquad ...(12)$$

The unknowns $y_1, y_2, y_3, ..., y_n$ in (12) are determined by forward substitution and the unknowns $x_1, x_2, x_3, ..., x_n$ in $\mathrm{UX} = \mathrm{Y}$, are obtained by back substitution.

We know that solving these two triangular sytem is simple. Finally, if we need, the inverse of the matrix A can also be determined using the following relation

$$\mathrm{A}^{-1} = \mathrm{U}^{-1}\mathrm{L}^{-1} \qquad ...(13)$$

Note : The method fails if any of the diagonal elements l_{ii} or u_{ii} is zero.

Let us illustrate the above LU decomposition method by taking some examples.

Example 4.4 Solve the system of equations by LU factorization method :

$$2x + 3y + z = 9$$
$$x + 2y + 3z = 6$$
$$3x + y + 2z = 8$$

Solution : The given system can be written as

$$\mathrm{AX} = \mathrm{B}, \qquad \text{i.e.}$$

$$\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 9 \\ 6 \\ 8 \end{bmatrix}$$

Let us choose $l_{ii} = 1$ **(Doolittle method)** and write the coefficients matrix as

$$\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

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

Equating, we get

$$u_{11} = 2, \qquad u_{12} = 3, \qquad u_{13} = 1$$

$$l_{21}u_{11} = 1 \quad \Rightarrow \quad l_{21} = 1/2$$

$$l_{31}u_{11} = 3 \quad \Rightarrow \quad l_{31} = 3/2$$

$$l_{21}u_{12} + u_{22} = 2 \quad \Rightarrow \quad u_{22} = 1/2$$

$$l_{21}u_{13} + u_{23} = 3 \quad \Rightarrow \quad u_{23} = 5/2$$

$$l_{31}u_{12} + l_{32}u_{22} = 1 \quad \Rightarrow \quad l_{32} = -7$$

$$l_{31}u_{13} + l_{32}u_{23} + u_{33} = 2 \quad \Rightarrow \quad u_{33} = 18$$

Thus, we get

$$A = LU$$

$$\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 3/2 & -7 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 1 \\ 0 & 1/2 & 5/2 \\ 0 & 0 & 18 \end{bmatrix}$$

The given system is

$$AX = B$$

$$\Rightarrow \qquad LUX = B$$

Let $UX = Y$, so that the system becomes

$$LY = B, \text{ where } Y = [y_1, y_2, y_3]^T$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 3/2 & -7 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 6 \\ 8 \end{bmatrix}$$

Which gives (by forward substitution)

$$y_1 = 9$$

$$\frac{1}{2}y_1 + y_2 = 6$$

$$\frac{3}{2}y_1 - 7y_2 + y_3 = 8$$

$\Rightarrow \quad y_1 = 9,\ y_2 = 3/2,\ y_3 = 5$

Now, $\quad \mathrm{UX} = \mathrm{Y}$

$$\begin{bmatrix} 2 & 3 & 1 \\ 0 & 1/2 & 5/2 \\ 0 & 0 & 18 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 9 \\ 3/2 \\ 5 \end{bmatrix}$$

Which gives (by backward subtitution)

$$18z = 5$$

$$\frac{1}{2}y + \frac{5}{2}z = \frac{3}{2}$$

$$2x + 3y + z = 9$$

$$\Rightarrow \quad x = \frac{35}{18},\ y = \frac{29}{18},\ z = \frac{5}{18}$$

Example 4.5 Using Cholesky (square root) method solve the system of equations

$$4x - y = 1$$
$$-x + 4y - z = 0$$
$$-y + 4z = 0$$

Solution : The given system can be written as

$$\mathrm{AX} = \mathrm{B}$$

Where

$$\mathrm{A} = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{bmatrix},\ \mathrm{X} = \begin{bmatrix} x \\ y \\ z \end{bmatrix},\ \mathrm{B} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

The coefficients matrix A can be written as

$$\mathrm{A} = \mathrm{LL}^{\mathrm{T}}$$

Where L is a lower triangular matrix and L^{T} is the transpose of the same.

Then we have

$$\begin{bmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

$$= \begin{bmatrix} l_{11}^2 & l_{11}l_{21} & l_{11}l_{31} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{21}l_{31} + l_{22}l_{32} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

Equating, we get

$$l_{11}^2 = 4 \qquad \Rightarrow \qquad l_{11} = 2$$

$$l_{11}l_{21} = -1 \qquad \Rightarrow \qquad l_{21} = -1/2$$

$$l_{11}l_{31} = 0 \qquad \Rightarrow \qquad l_{31} = 0$$

$$l_{21}^2 + l_{22}^2 = 4 \qquad \Rightarrow \qquad l_{22} = \sqrt{15}/2$$

$$l_{21}l_{31} + l_{22}l_{32} = -1 \qquad \Rightarrow \qquad l_{32} = -2/\sqrt{15}$$

$$l_{31}^2 + l_{32}^2 + l_{33}^2 = 4 \qquad \Rightarrow \qquad l_{33} = \sqrt{56/15}$$

Thus we have

$$L = \begin{bmatrix} 0 & 0 & 0 \\ -1/2 & \sqrt{15}/2 & 0 \\ 0 & -2/\sqrt{15} & \sqrt{56/15} \end{bmatrix}$$

Now the system can be written as

$$\mathrm{AX} = \mathrm{LL^TX} = \mathrm{B}$$

Let $\mathrm{L^TX} = \mathrm{Y}$

Then $\mathrm{LY} = \mathrm{B}$

$$\begin{bmatrix} 0 & 0 & 0 \\ -1/2 & \sqrt{15}/2 & 0 \\ 0 & -2/\sqrt{15} & \sqrt{56/15} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

On solving this by forward substitution, we get

$$y_1 = \frac{1}{2},\ y_2 = \frac{1}{2\sqrt{15}},\ y_3 = \frac{1}{\sqrt{840}}$$

Now using these values in $\mathrm{L^TX} = \mathrm{Y}$, we have

$$\begin{bmatrix} 0 & -1/2 & 0 \\ 0 & \sqrt{15}/2 & -2/\sqrt{15} \\ 0 & 0 & \sqrt{56/15} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2\sqrt{15} \\ 1/\sqrt{840} \end{bmatrix}$$

On solving this by back substitution, we get

$$x = \frac{1}{56},\ y = \frac{1}{14},\ z = \frac{15}{56}$$

Self-Learning Exercise - 3

1. Show that the decomposition method fails to solve the system of equations

$$\begin{aligned} x+y-z&=2 \\ 2x+2y+5z&=-3 \\ 3x+2y-3z&=6 \end{aligned}$$

2. Solve the system of equations

$$\begin{aligned} x+2y+3z&=14 \\ 2x+5y+2z&=18 \\ 3x+y+5z&=20 \end{aligned}$$

Using (i) Doolittle's method, (ii) Crout's method, (iii) Cholesky method.

4.2.4 Partition Method

The partition method is basically used to obtain the inverse of the given matrix of higher order. Let A be the non singular coefficient matrix of order n and it is partitioned like

$$A=\left[\begin{array}{c:c} A_1 & A_2 \\ \hline A_4 & A_3 \end{array}\right] \qquad ...(14)$$

Let r and s be positive integers such that $n=r+s$ then distribution of the order in partitioned A will take place like A_1 is a matrix of order $r\times r$, A_2 is matrix of $r\times s$, A_3 is matrix of $s\times s$ and A_4 is matrix of $s\times r$.

Let the inverse of the same matrix is partitioned as

$$A^{-1}=\left[\begin{array}{c:c} B_1 & B_2 \\ \hline B_4 & B_3 \end{array}\right] \qquad ...(15)$$

Here the order of B_1, B_2, B_3 and B_4 are the same as orders of A_1, A_2, A_3 and A_4 respectively. Hence we have

$$A\ A^{-1}=\left[\begin{array}{c:c} A_1 & A_2 \\ \hline A_4 & A_3 \end{array}\right]\left[\begin{array}{c:c} B_1 & B_2 \\ \hline B_4 & B_3 \end{array}\right]=\begin{bmatrix} I_1 & O \\ O & I_2 \end{bmatrix} \qquad ...(16)$$

Where I_1 and I_2 are identity matrices of order r and s respectively. Now equation (10) gives

$$A_1B_1+A_2B_4=I_1$$

$$A_1B_2+A_2B_3=O$$

$$A_4B_1+A_3B_4=O$$

$$A_4B_2+A_3B_3=I_2 \qquad ...(17)$$

On solving all these we obtain

$$B_3 = \left(A_3 - A_4 A^{-1} A_2\right)^{-1}$$

$$B_2 = -A_1^{-1} A_2 B_3$$

$$B_4 = B_3 A_4 A_1^{-1}$$

$$B_1 = A_1^{-1}\left(I_1 - A_2 B_4\right) \qquad \text{...(18)}$$

One should remember the expression (18).

Example 4.6 Solve the following system of the equation using partition method

$$\begin{aligned} 3x + 2y + z &= 11.6 \\ 2x + 3y + 2z &= 15.9 \\ x + 2y + 2z &= 12.2 \end{aligned}$$

Solution : Let in the given system the coefficient matrix is A and it can be partitioned as

$$A = \left[\begin{array}{cc:c} 3 & 2 & 1 \\ 2 & 3 & 2 \\ \hline 1 & 2 & 2 \end{array}\right] = \left[\begin{array}{c} A_1 \vdots A_2 \\ \hline A_4 \vdots A_3 \end{array}\right]$$

Therefore,

$$A_1 = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad A_3 = [2], \quad A_4 = [1 \quad 2]$$

Now, we have

$$A_1^{-1} = \frac{1}{5}\begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix},$$

$$A_3 - A_4\, A_1^{-1} A_2 = [2] - [1 \quad 2]\frac{1}{5}\begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$= [2] - \frac{1}{5}[7] = [3/5]$$

And

$$\left(A_3 - A_4\, A_1^{-1} A_2\right)^{-1} = [5/3].$$

Let

$$A^{-1} = \left[\begin{array}{c} B_1 \vdots B_2 \\ \hline B_4 \vdots B_3 \end{array}\right],$$

Hence using the above expression we have

$$B_3 = \left(A_3 - A_4\, A_1^{-1} A_2\right)^{-1} = [5/3],$$

$$B_2 = -A_1^{-1} A_2 B_3 = -\frac{1}{5}\begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix}[5/3]$$

$$= \begin{bmatrix} 1/3 \\ -4/3 \end{bmatrix},$$

$$B_4 = B_3\, A_4\, A_1^{-1} = [5/3]\,[1 \quad 2]\,\frac{1}{5}\begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}$$

$$= [1/3 \quad -4/3]$$

And

$$B_1 = A_1^{-1}\left(I_1 - A_2 B_4\right)$$

$$= \frac{1}{5}\begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1/3 & -4/3 \\ 2/3 & -8/3 \end{bmatrix}\right)$$

$$= \frac{1}{5}\begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 2/3 & 4/3 \\ -2/3 & 11/3 \end{bmatrix}$$

$$= \begin{bmatrix} 2/3 & -2/3 \\ -2/3 & 5/3 \end{bmatrix}$$

Thus the inverse of the coefficient martix is given by

$$A^{-1} = \begin{bmatrix} B_1 \vdots B_2 \\ \hline B_4 \vdots B_3 \end{bmatrix} = \begin{bmatrix} 2/3 & -2/3 & 1/3 \\ -2/3 & 5/3 & -4/3 \\ 1/3 & -4/3 & 5/3 \end{bmatrix}$$

Hence the solution of the system can be obtained as

$$X = A^{-1}B = \frac{1}{3}\begin{bmatrix} 2 & -2 & 1 \\ -2 & 5 & -4 \\ 1 & -4 & 5 \end{bmatrix}\begin{bmatrix} 11.6 \\ 15.9 \\ 12.2 \end{bmatrix}$$

Therefore

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1.2 \\ 2.5 \\ 3.0 \end{bmatrix}$$

Hence the solution be

$$x = 1.2\,,\; y = 2.5\,,\; z = 3.0$$

Self-Learning Exercise - 4

Using the partition method solve the system of equation

$$\begin{bmatrix} 2 & 1 & 1 & 2 \\ 4 & 0 & 2 & 1 \\ 3 & 2 & 2 & 0 \\ 1 & 3 & 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ -1 \\ -4 \end{bmatrix}$$

4.3 Method of Successive Approximation

The direct method discussed so far involve many subtractions. When the terms involved in subtraction is nearly equal, their difference is nearly zero and hence causes inaccuracies. The inaccuracies due to this inherent weakness of the direct methods can not be completely avoided, whereas the iterative methods (method of successive approximation) are free from such inaccuracies. Moreover these methods are self correcting; viz. errors made at any stage in the computation are corrected automatically in the subsequent stage of iteration. When the coefficient matrix has more zeros, the iterative methods are rapid than the direct methods. In order for the iteration procedure to give the solution of system of equations, each equation of the system must contain one coefficient much larger in magnitude than the others in that equation and large coefficient must be that of a different unknown in each equation. In other word, after rearranging the equations if necessary, the large coefficients must be along the leading diagonal of the coefficient matrix.

Among the iterative methods we wil desicuss the Conjugate Gradient method and Relaxation method.

4.3.1 Conjugate Gradient Method

The conjugate gradient method or more briefly, the CG method, is a successive technique for solving large system of linear equations $AX = B$, when the coefficient matrix A is symmetric $\left(A^T = A\right)$ and positive definite $\left(X^T AX > 0\right)$. It terminates in at most n steps if no rounding-off errors are encountered. Starting with an initial estimate x_0 (arbitrary) of the solution h and hence one can obtain successive newly estimates x_1, x_2, x_3..... At each step the residual $R_i = k - Ax_i$ is computed. Normally this vector can be used as a measure of the "goodness" of the estimate of x_i. Experience indicate that frequently x_{n+1} are considerably better than x_n. One should not continue too far beyond x_n but should start a new with the last estimate obtained as the initial estimate, so as to diminish the effect of round-off error. Infact one can start with a new at any iteration. This flexibility is one of the principal advantages of this method. In case matrix of coefficients is symmetric and positive definite, the following formulae are used in the conjugate gradient method :

$$p_0 = R_0 = B - Ax_0 \qquad ...(19)$$

$$\alpha_1 = \frac{R_i^T R_i}{p_i^T Ap_i} \qquad ...(20)$$

$$x_{i+1} = x_i + \alpha_i p_i \qquad ...(21)$$

$$R_{i+1} = R_i - \alpha_i A p_i \qquad ...(22)$$

$$\beta_i = \frac{R_{i+1}^T R_{i+1}}{R_i^T R_i} \qquad ...(23)$$

$$p_{i+1} = R_i + \beta_i p_i \qquad ...(24)$$

Select an estimate x_0 and compute the residual R_0 and the direction p_0 using (19). After getting these initial values get the routine values of x_i residual R_i, direction p_i and hence compute x_{i+1}, R_{i+1}, p_{i+1} using the given forumulae. On the same lines find the better approximations to the solution of the given system.

Example 4.7 Solve the given system of equation using CG method

$$4x + y = 1$$
$$x + 3y = 2$$

Solution : Considering the given system as $AX = B$, the system is

$$\begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

To start the conjugate gradient method let us take initial guess as

$X_0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, in order to find an approximate solution of the system.

Now we have to calculate the first residual R_0 corresponding to guessed solution

$$R_0 = B - Ax_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -8 \\ -3 \end{bmatrix},$$

Since this the first iteration we can take the first residual as our initial direction p_0.

Now calculating the scalar α_0 using the relation (20)

$$\alpha_0 = \frac{R_0^T R_0}{p_0^T A p_0} = \frac{\begin{bmatrix} -8 & -3 \end{bmatrix} \begin{bmatrix} -8 \\ -3 \end{bmatrix}}{\begin{bmatrix} -8 & -3 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} -8 \\ -3 \end{bmatrix}} = \frac{73}{331}.$$

Now we calculate the first approximate solution X_1 using (21)

$$X_1 = X_0 + \alpha_0 p_0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{73}{331} \begin{bmatrix} -8 \\ -3 \end{bmatrix} = \begin{bmatrix} 0.2356 \\ 0.3384 \end{bmatrix}$$

This solution is improved one than we guessed; similarly we can find the next improvement and so on.

Compute the next residual R_1 using the formula (22)

$$R_1 = R_0 - \alpha_0\, Ap_0 = \begin{bmatrix} -8 \\ -3 \end{bmatrix} - \frac{73}{331}\begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} -8 \\ -3 \end{bmatrix} = \begin{bmatrix} -0.2810 \\ 0.7492 \end{bmatrix}$$

Let us compute now β_0, which will be used to obtain next search direction p_1.

$$\beta_0 = \frac{R_1^T R_0}{R_0^T R_0} = \frac{\begin{bmatrix} -0.2810 & 0.7492 \end{bmatrix}\begin{bmatrix} -0.2810 \\ 0.7492 \end{bmatrix}}{\begin{bmatrix} -8 & 3 \end{bmatrix}\begin{bmatrix} -8 \\ -3 \end{bmatrix}} = 0.0088$$

Now

$$p_1 = R_1 + \beta_0 p_0 = \begin{bmatrix} -0.2810 \\ 0.7492 \end{bmatrix} + 0.0088\begin{bmatrix} -8 \\ -3 \end{bmatrix} = \begin{bmatrix} -0.3511 \\ 0.7229 \end{bmatrix}$$

Now again using (20) we get

$$\alpha_1 = \frac{R_1^T R_1}{p_1^T A p_1} = \frac{\begin{bmatrix} -0.2810 & 0.7492 \end{bmatrix}\begin{bmatrix} -0.2810 \\ 0.7492 \end{bmatrix}}{\begin{bmatrix} -0.3511 & 0.7229 \end{bmatrix}\begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} -0.3511 \\ 0.7229 \end{bmatrix}} = 0.4122$$

Now find the next approximation

$$X_2 = X_1 + \alpha_1 p_1 = \begin{bmatrix} 0.2356 \\ 0.3384 \end{bmatrix} + 0.4122\begin{bmatrix} -0.3511 \\ 0.7229 \end{bmatrix} = \begin{bmatrix} 0.0909 \\ 0.6364 \end{bmatrix}$$

The result x_2 is better that x_1 and one can go further to get more improvement.

$x = 0.0909$, $y = 0.6364$

Self-Learning Exercise - 5

Solve the given system using conjugate Gradient method (two iterations only)

$$2x - z = 1$$
$$-2x - 10y = -12$$
$$-x - y + 4z = 3$$

Using the initial vector $(x, y, z)^T = (0,0,0)^T$

4.3.2 Relaxation Method

This method is a generalization of the Gauss Seidel method (which we already studied in previous classes). This method is the most powerful acceleration technique to find the solution of the given system and has a rapid convergence. This particular shceme permits one to select the best equation to be used for maximum rate of convergence. Initially we assume the values of unknowns, which further improved by reducing the so called *residuals* to zero or as close as possible to zero.

We first make the given system diagonally dominated, and hence take all the terms to the one side. The residual of *i th* equation is denoted by R_i and given by

$$R_i = b_i - a_{i1}x_1 - a_{i2}x_2 - \ldots\ldots - a_{in}x_n,\ i = 1,2,3,\ldots.,n.$$

The largest residual in magnitude (say R_{ik}) tells us that the *k th* equation is most in the error and should be improved first. We then select the new residual of greatest magnitude and relax it to zero. We continue until all residues are zero and when it is obtained the values of x_i, $i = 1,2,3,\ldots\ldots,n$ will givs us the solution of the system.

Example 4.8 Solve the system of the following equations using relaxation method

$$\begin{aligned} 2x + y - 8z &= -15 \\ x - 7y + z &= 10 \\ 6x - 3y + z &= 11 \end{aligned}$$

Solution : First we reorder the equations so that system will convert to diagonal dominated (the largest coefficients in the equations appear on the diagonal) and transfer all the terms on one side.

Therefore

$$\begin{aligned} R_1 &= 11 - 6x + 3y - z \\ R_2 &= 10 - x + 7y - z \\ R_3 &= -15 - 2x - y + 8z \end{aligned} \qquad \ldots(1)$$

If we start with the initial values $x = 0$, $y = 0$, $z = 0$, the residuals are $R_1 = 11$, $R_2 = 10$, $R_3 = -15$. In which the largest residual in magnitude is R_3. Since the third equation has more error we have to improve this first.

Let us introduce the variation in z as

$$\Delta = -\frac{R_3}{a_{33}} = -\frac{(-15)}{8} = 1.87$$

Put this value in equation (1) in place of z and keep the rest values as same. Again we find the residues as $R_1 = 9.125$, $R_2 = 8.125$, $R_3 = 0$. In which the largest is R_1. Similarly we can find the further residuals (which are largest in the magnitude) and applying the same strategy and use this improved value for the variables. This process will be continued untill all the residues come down to zero or near to that. Let us show the process and values in tabulated form

1	Residuals			Variations	Variables		
	R_1	R_2	R_3	Δ	x	y	z
0	11	10	**−15**	$-\frac{(-15)}{8}=1.875$	0	0	0
1	**9.125**	8.125	0	$-\frac{(9.125)}{(-6)}=1.5288$	0	0	1.875
2	0.002	**6.6042**	−3.0416	$-\frac{(6.6042)}{7}=-0.9434$	1.5208	0	1.875
3	**−2.83**	0.0004	−2.0982	$-\frac{(-2.83)}{(-6)}=-0.4716$	1.5208	−0.9434	1.875
4	−0.004	0.472	**−1.155**	$-\frac{(-1.155)}{8}=0.1443$	1.0492	−0.9434	1.875
5	−0.1447	**0.3277**	−0.0006	$-\frac{(0.3277)}{7}=-0.0468$	1.0492	−0.9902	2.0193
6	**−0.2851**	0.0001	0.0462	$-\frac{(-0.2851)}{(-6)}=-0.0475$	1.0492	−0.9902	2.0193
7	−0.0001	0.0476	**0.1412**	$-\frac{(0.1412)}{8}=-0.176$	1.007	−0.9902	2.0193
8	0.0175	**0.0652**	0.0004	$-\frac{(0.0652)}{7}=-0.0093$	1.007	−0.9902	2.0017
9	−0.0104	0.0001	0.0097	--	1.007	−0.9995	2.0017

The residuals maximum in magnitude, at each iteration, are showed boldly. At this final stage one can see that all the residues are sufficiently small so the corresponding values of x, y, z at this position can be taken as a solution. Hence

$$x = 1.0017,\ y = -0.9995,\ z = 2.0017$$

Or in round figures the solution is

$$x = 1.0,\ y = -1.0,\ z = 2.0$$

Self-Learning Exercise - 6

1. Solve the following system of equations using relaxation method

$$5x - 2y + z = 13$$
$$3x + 7y - 11z = 2$$
$$x + 20y - 2z = 8$$

2. Solve the following system of equations using relaxation method

$$8x + y - z = 8$$
$$2x + y + 9z = 12$$
$$x - 7y + 2z = -4$$

4.4 Summary

By the study of this unit one can solve the system of simultaneous linear equations using both the iterative and direct methods.

In particular the method of determinant is good enough in solving the system of equations up to third order after that it may be laborious. Before adapting the method of determinant reader must check the consistency of the given system. The Gauss Jordan method is better than the previous one. This method is basically the modified form of Guass Elimination method. The Gauss elemination method gives

$$\begin{bmatrix} A & \vdots & B \end{bmatrix} \xrightarrow{Gauss\ E\lim ination} \begin{bmatrix} A & \vdots & B \end{bmatrix}$$

Where $\left[A \vdots B \right]$ is augmented matrix.

Whereas the Gauss Jordan method gives

$$\begin{bmatrix} A & \vdots & B \end{bmatrix} \xrightarrow{Gauss\ Jordon} \begin{bmatrix} I & \vdots & d \end{bmatrix}$$

Generally it is seen that the Gauss Jordan method is more laborious than Gauss Elimination method.

Next direct method is LU decomposition method. In which we bifurcate the coefficient matrix in lower and upper triangular matrices and after two steps we can find the solution. Let given system of equations be

AX = B, then we convert it into

LUX = B, where L is lower and U is upper triangular matrices. Again let UX = Y, then the previous equation will take the form LY = B

The unknowns $y_1, y_2, y_3, ..., y_n$ of Y are determined by forward substitution and the unknowns $x_1, x_2, x_3, ..., x_n$ in UX = Y, are obtained by back substitution. The next direct method is Partition method, which is promptly used to obtain the inverse of the given matrix of higher order. Let A be the non singular coefficient matrix of order n and it is partitioned like

$$A = \left[\frac{A_1 \vdots A_2}{A_4 \vdots A_3} \right]$$

Let r and s be positive integers such that $n = r + s$ then distribution of the order in partitioned A will take place like A_1 is a matrix of order $r \times r$, A_2 is matrix of $r \times s$, A_3 is matrix of $s \times s$ and A_4 is matrix of $s \times r$.

The inverse of the same matrix is partitioned as

$$A^{-1} = \left[\frac{B_1 \vdots B_2}{B_4 \vdots B_3} \right]$$

Further by applying the certain formulae given above, one can proceed to obtain the desired solution of the system of the linear equations.

Now the iterative methods start from an approximation to the true solution and if convergent, derive the sequence of closer approximations. The cycle of the commputation is repeated till the desired accuracy is attained.

Applyin the two or more methods on one problem we can check which method gives the solution faster. If it is desired one can find the convergence of the methods and also find a comparative study of the methods.

4.5 Answer of Self-Learning Exercise

Self-Learning Exercise - 1

1. $x=1,\ y=1,\ z=1$ 2. $x=1,\ y=2,\ z=3$

Self-Learning Exercise - 2

1. $x=1,\ y=-2,\ z=3$ 2. $x=1,\ y=1,\ z=1$

Self-Learning Exercise - 3

1. Try yourself 2. $x=1,\ y=2,\ z=3$

Self-Learning Exercise - 4

$x=1,\ y=-1,\ z=-1,\ w=1$

Self-Learning Exercise - 5

$x=0.195751,\ y=1.109993,\ z=0.948462$

Self-Learning Exercise - 6

1. $x=2.6,\ y=0.3432,\ z=0.741$ 2. $x=1,\ y=1,\ z=1$

4.6 Exercises

1. Using Gauss Jordan method solve the following system of equations

(i) $3x+2y+z=10$
$2x+3y+2z=14$
$x+2y+3z=14$

(ii) $3x+y+z=6$
$x+2y+3z=8$
$2x+y+4z=8$

2. Solve the following simultaneous linear equations using Crout's method

(i) $x+y+z=1$
$4x+3y-z=6$
$3x+5y+3z=4$

(ii) $10x+y+z=12$
$2x+10y+z=13$
$2x+2y+10z=14$

3. Solve the given system of equations using Choleskey method

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 8 & 22 \\ 3 & 22 & 82 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}=\begin{bmatrix} 5 \\ 6 \\ -10 \end{bmatrix}$$

4. Find the Doolittle, Crout and Cholesky factorizations of the matrix

$$A = \begin{bmatrix} 60 & 30 & 20 \\ 30 & 20 & 15 \\ 20 & 15 & 12 \end{bmatrix}$$

5. Solve the following system of equations by the Relaxation method

$$\begin{aligned} x + 9y - z &= 10 \\ 2x - y + z &= 20 \\ 10x - 2y + z &= 12 \end{aligned}$$

6. Solve the following system of equations by the CG method

$$\begin{bmatrix} 2 & -1 & 0 \\ 1 & 6 & -2 \\ 4 & -3 & 8 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \\ 5 \end{bmatrix}$$

Take the initial approximation $x = y = z = 0$.

□□□

Unit - 5 : Eigen Value Problems-I

Structure of the Unit

5.0 Objectives

The Eigen value problems are of greatest importance to the engineer and physicist. For example, in solid mechanics, when we consider an element in a continuum, subject to a normal and shear stresses, we usually find principal stresses which are the maximum and minimum stresses in an element. When we solve this kind of problems we have to go through matrices, Eigen values and vectors etc.

5.1 Introduction

The Eigen values and vectors of a matrix are important in numerical analysis. The numerical solution involves finding Eigen value of the coefficient matrix of the set of difference equations. More precisely, if eigenvalue problems involving a differential equation or an integral equation, it is known as algebraic eigenvalue problem. In some problems, we may be interested in only the eigenvalues with the largest magnitude or with large matrices which usually arise from discretisation of differential equations; we are normally interested in only a few eigenvalues and vectors. Here in this unit we are interested to understand the two important methods, first is Power method and another is Jacobi method.

5.2 Eigen Values and Vectors

Let $A=\left[a_{ij}\right]$ be a given $n \times n$ matrix and consider the vector equation of the form

$$Ax = \lambda x \qquad ...(1)$$

Where A is a given square matrix x is unknown vector and λ is an unknown scalar. Our aim is to solve this equation, obviously if $x=0$ is a solution then $0=0$, but this has no practical interest. A value of λ for which (1) has a solution $x \neq 0$ is called an **Eigenvalue** or **Characteristic value** or **Latent root** of the matrix A. The corresponding solution $x \neq 0$ of the same equation is called the **Eigenvectors** or **Characteristic vectors** of A corresponding to that eigenvalue λ. The set of eigenvalues is called the **spectrum** of A. The largest eigenvalue (in magnitude) is called the **spectral radius** of A.

Again taking the above equation

$Ax = \lambda\, x$, which implies

$$\left(A - \lambda I\right)X = 0 \qquad ...(2)$$

This equation is a homogeneous system of n linear equations and shall have a non trivial solution if $\left|A - \lambda I\right| = 0$, when it is simplified gives the polynomial equation

$$\left(-1\right)^n \lambda^n + a_1\lambda^{n-1} + ...+ a_n = 0 \qquad ...(3)$$

Equation (3) is called the characteristic equation and has n roots, say $\lambda_1, \lambda_2, \lambda_3, ..., \lambda_n$, these values of λ are called eigenvalues of A. Also the values of the vector X, say $X_1, X_2, X_3, ..., X_n$, corresponding to $\lambda_1, \lambda_2, \lambda_3, ..., \lambda_n$ are called the eigenvectors of A.

Let us see how to find the eigenvalues and vectors by taking an example of following matrix

$$A = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix} \qquad ...(1)$$

Using the equation (1), we have

$$\begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad ...(2)$$

or can be written as

$$-5x_1 + 2x_2 = \lambda\, x_1$$

$$2x_1 - 2x_2 = \lambda\, x_2$$

or $$\left(-5 - \lambda\right)x_1 + 2x_2 = 0$$

$$2x_1 + \left(-2 - \lambda\right)x_2 = 0$$

Which can be written in the matrix notation is given in (2)

$$\left(A - \lambda I\right)X = 0 \qquad ...(3)$$

This equation is a homogeneous system of equations and shall have a non trivial solution if $\left|A - \lambda I\right| = 0$

Hence

$$\left|A - \lambda I\right| = \begin{vmatrix} -5 - \lambda & 2 \\ 2 & -2 - \lambda \end{vmatrix} = 0$$

$$= \left(-5 - \lambda\right)\left(-2 - \lambda\right) - 4 = \lambda^2 + 7\lambda + 6 = 0$$

On solving this characteristic equation of A, we have

$$\lambda_1 = -1,\ \lambda_2 = -6$$

These are the eigenvalues of the matrix A. Now to determine the Eigen vector corresponding to eigenvalue $\lambda_1 = -1$, using this valuse in (3) we have

$$\begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

or

$$-4x_1 + 2x_2 = 0$$

$$2x_1 - x_2 = 0$$

Solving this we find

$X_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} k \\ 2k \end{bmatrix}$, where k is arbitrary constant. If we put k = 1, we have

$$X_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Now we can check

$$A\,X_1 = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix} = (-1)\begin{bmatrix} 1 \\ 2 \end{bmatrix} = (-1)X_1 = \lambda_1 X_1.$$

Similarly we can find the second corresponding vector to the eigenvalues $\lambda_2 = -6$,

which will be $X_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

Example 5.1 Find the eigenvalues and vectors of the following matrix A

$$A = \begin{bmatrix} 5 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 5 \end{bmatrix}$$

Solution : The characteristic equation for the given matrix will be

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

$$\Rightarrow \quad \begin{vmatrix} 5-\lambda & 0 & 1 \\ 0 & -2-\lambda & 0 \\ 1 & 0 & 5-\lambda \end{vmatrix} = 0$$

$$\Rightarrow \quad (5-\lambda)(-2-\lambda)(5-\lambda) + (2+\lambda) = 0$$

$\Rightarrow \qquad (\lambda+2)(\lambda^2-10\lambda+24)=0$

$\Rightarrow \qquad \lambda=-2,4,6$

These are the eigenvalues of A.

Now let us find the corresponding eigenvectors.

When $\lambda=-2$,

Let the corresponding vector be $X_1=(x_1,x_2,x_3)^T$

Then we have

$$(A+2I)X_1=0$$

$$\Rightarrow \qquad \begin{bmatrix} 7 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$\Rightarrow \qquad 7x_1+x_2=0$

$\qquad\qquad x_1+7x_2=0$

$\Rightarrow \qquad x_1=0=x_3$

Let $x_2=k_1$, then

$$X_1=\begin{bmatrix} 0 \\ k_1 \\ 0 \end{bmatrix} = k_1\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Choosing $k_1=1$, we get $X_1=\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$.

When $\lambda=4$

Let the corresponding vector be $X_2=(x_1,x_2,x_3)^T$

Then, we have

$$(A-4I)X_2=0$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & -6 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

On solving this, we get

$$x_1+x_3=0 \ ; \quad -6x_2=0$$

$\Rightarrow \qquad x_1 = -x_1 \quad ; \quad x_2 = 0$

Now randomly we can take $x_1 = k_2$ which implies $x_3 = -k_2$.

Hence the corresponding vector is

$$X_2 = \begin{bmatrix} k_2 \\ 0 \\ k_2 \end{bmatrix}$$

Note : In particular, if we choose $x_1 = 1/\sqrt{2}$ and $x_3 = 1/\sqrt{2}$ so $x_1^2 + x_2^2 + x_3^3 = 1$.

Then the vector, choosen in this way is said to be normalized.

When $\lambda = 6$

Let the corresponding vector by $X_3 = (x_1, x_2, x_3)^T$, we have

$$(A - 6I)X_3 = 0$$

$$\begin{bmatrix} -1 & 0 & 1 \\ 0 & -8 & 0 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Solving this, we have

$$x_3 - x_1 = 0 \ ; \ -8x_2 = 0 \ ; \ x_1 - x_3 = 0$$

$\Rightarrow \qquad x_1 = x_3 \ ; \ x_2 = 0$

Now let us choose $x_1 = k = 1/\sqrt{2}$

The corresponding vector will be

$$X_1 = \begin{bmatrix} k \\ 0 \\ k \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}.$$

5.3 Basic Properties

1. For a given square matrix the eigenvalues are unique but the eigen vectors are not unique.
2. A square matrix and its transpose have the same eigenvalues.
3. The eigenvalues of a triangular matrix are exactly the diagonal elements of the matrix.
4. If $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \ldots\ldots.$ are the eigenvalues of square matrix A then

   (a) $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \frac{1}{\lambda_3}, \frac{1}{\lambda_4} \cdots\cdots$ are the eigenvalues of A^{-1}

(b) $k\lambda_1, k\lambda_2, k\lambda_3, k\lambda_4..........$ are the eigenvalues of kA.

(c) $\lambda_1^k, \lambda_2^k, \lambda_3^k, \lambda_4^k......$ are the eigenvaluse of A^k.

5. The eigenvalues of a Harmitian matrix and a real symmetric matrix are real.

6. The set of all eigenvalues of matrix A is called spectrum of A and the largest eigenvalue (in magnitude) is called the spectral radius of A.

7. The sum of all the eigenvalues of a square matrix is equal to the trace (sum of diagonal elements) of the matrix.

8. The product of the eigenvalues is equal to the determinant of the matrix.

9. Statement of Cayley Hamilton theorem: Every square matrix satisfies its own characteristic equation.

10. Eigen vector X is said to be orthonormal or normalized if $XX^T = 1$.

Self-Learning Exercise - 1

Find all eigenvalues and vectors of the following matrices

1. $\begin{bmatrix} 4 & 1 & 1 \\ 2 & 4 & 1 \\ 0 & 1 & 4 \end{bmatrix}$ 2. $\begin{bmatrix} 2 & \sqrt{2} \\ \sqrt{2} & 1 \end{bmatrix}$

5.4 Power Method

In some practical problems only the largest eigenvalue and the corresponding vector are required. The power method is a simple iterative method which is designed to compute the dominant eigenvalue or largest latent root in magnitude and the corresponding vector of a matrix.

This procedure is applicable if the latent roots are real and distinct so the corresponding vectors are linearly independent. Let $\lambda_1, \lambda_2, \lambda_3, \lambda_4, ..., \lambda_n$ be real and distinct eigenvalues of the given matrix A of order n, such that

$$|\lambda_1| > |\lambda_2| > |\lambda_3| > ......> |\lambda_n| \tag{4}$$

Let $X_1, X_2, X_3, ..., X_n$ be the corresponding vectors, then any vector X can be written as their linear combination,

$$X = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + ...+ \alpha_n X_n \tag{5}$$

Now we have,

$$AX = A\left(\alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + ...+ \alpha_n X_n\right)$$

$$= \alpha_1 AX_1 + \alpha_2 AX_2 + \alpha_3 AX_3 + ...+ \alpha_n AX_n$$

$$= \alpha_1 \lambda_1 X_1 + \alpha_2 \lambda_2 X_2 + \alpha_3 \lambda_3 X_3 + ...+ \alpha_n \lambda_n X_n$$

Again,

$$A(AX) = \alpha_1 \lambda_1^2 X_1 + \alpha_2 \lambda_2^2 X_2 + \alpha_3 \lambda_3^2 X_3 + ...+ \alpha_n \lambda_n^2 X_n$$

$$A^2X = \alpha_1\lambda_1^2X_1 + \alpha_2\lambda_2^2X_2 + \alpha_3\lambda_3^2X_3 + ... + \alpha_n\lambda_n^2X_n$$

On the same steps if the above expression is multiplied by A, we get

$$A^3X = \alpha_1\lambda_1^3X_1 + \alpha_2\lambda_2^3X_2 + \alpha_3\lambda_3^3X_3 + ... + \alpha_n\lambda_n^3X_n$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$

$$A^rX = \alpha_1\lambda_1^rX_1 + \alpha_2\lambda_2^rX_2 + \alpha_3\lambda_3^rX_3 + ... + \alpha_n\lambda_n^rX_n$$

$$\Rightarrow \qquad A^rX = \alpha_1\lambda_1^rX_1 + \lambda_1^r\sum_2^n \alpha_i\left(\frac{\lambda_i}{\lambda_1}\right)^r X_1$$

When $r \to \infty$, $\left(\frac{\lambda_i}{\lambda_1}\right)^r \to 0$; $(\lambda_i < \lambda_1)$; $i = 2,3,.....n$.

therefore, $A^rX = \alpha_1\lambda_1^rX_1$

The same process can be extended to

$$A^{r+1}X = \alpha_1\lambda_1^{r+1}X_1$$

Now the dominant eigenvalue λ_1 can be obtained as

$$\lambda_1 = \frac{\lambda_1^{r+1}}{\lambda_1^r} = \overset{\lim}{r \to \infty}\frac{\left(A^{r+1}X\right)^k}{\left(A^rX\right)^k} \; ; \; k = 1,2,3,.....,n. \qquad ...(6)$$

where k denote the $k\,th$ component in the corresponding vector.

The convergence of the method is depending on the ratio $\frac{|\lambda_i|}{|\lambda_1|}$.

The initial vector X is selected suitably if no other approximation is given.

The least eigenvalue of A can be obtained using the fact the inverse matrix has a set of eigenvalues which are the reciprocals of the eigenvalues of A. Thus to obtain the least eigenvalues we have to apply the power method to the inverse of A. The main advantage of the power method is its simplicity, only matrix multiplication is required for computation. To illustrate power method, let us take an arbitrary vector and multiply it by given matrix and normalized. Repeat the process untill the normalized product converges.

Let the matrix

$$A = \begin{bmatrix} 3 & -1 & 0 \\ -2 & 4 & -3 \\ 0 & -1 & 1 \end{bmatrix}$$

And let the initial vector be

$X = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, at each step we normalized the vectory by the making its largest component equal to unity.

$$\begin{bmatrix} 3 & -1 & 0 \\ -2 & 4 & -3 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ -0.5 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 3 & -1 & 0 \\ -2 & 4 & -3 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -0.5 \\ 0 \end{bmatrix} = \begin{bmatrix} 3.5 \\ -4 \\ 0.5 \end{bmatrix} = -4 \begin{bmatrix} -0.875 \\ 1 \\ -0.125 \end{bmatrix}$$

Continue the process until the required accuracy is achieved, finally we get

$$\begin{bmatrix} 3 & -1 & 0 \\ -2 & 4 & -3 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} -0.4037 \\ 1 \\ -0.2233 \end{bmatrix} = \begin{bmatrix} -2.2111 \\ 5.4774 \\ -1.2233 \end{bmatrix} = 5.4774 \begin{bmatrix} -0.4037 \\ 1 \\ -0.2233 \end{bmatrix}$$

which shows the $AX = \lambda X$ form.

Therefore we obtain the dominant eigenvalue of A as $\lambda_1 = 5.4774$ and the corresponding vector $X_1 = (-0.4037, 1, -0.2233)^T$.

Example 5.2 Compute largest eigenvalue in magnitude and corresponding eigenvector of the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$$

Solution : Let $X^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ be the initial eigenvector. Then

$$AX^{(0)} = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix} = 5 \begin{bmatrix} 3/5 \\ 1 \end{bmatrix} = 5X^{(1)}$$

Hence an approximate eigenvalue is 5 and corresponding eigenvector is $X^{(1)}$.

Now we have

$$AX^{(1)} = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 3/5 \\ 1 \end{bmatrix} = \begin{bmatrix} 13/5 \\ 19/5 \end{bmatrix} = \frac{19}{5} \begin{bmatrix} 13/19 \\ 1 \end{bmatrix} = \frac{19}{5} X^{(2)}$$

Now $X^{(2)} = \begin{bmatrix} 13/19 \\ 1 \end{bmatrix}$ corresponding to eigenvalue $\frac{19}{5}$.

Repeating the above process finally we get the dominant eigenvalue as 4.000203 and corresponding eigenvector $X = \begin{bmatrix} 0.66664 \\ 1 \end{bmatrix}$.

Self-Learning Exercise - 2

1. Compute the dominant eigenvalue and eigenvector of the following matrix

$$\begin{bmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{bmatrix}$$

2. Compute the dominant latent root and eigenvector of the following matrix

$$\begin{bmatrix} 1 & 6 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$, also find its other two latent roots.

5.5 Jacobi Method

Jacobi method is highly recommended when we need to compute all the eigenvalues and vectors of a real symmetric matrix. In fact if A is a real symmetric matrix then its eigenvalues are real and corresponding vectors to distinct eigenvalues are orthogonal. In other words, for a real symmetric matrix A, there exist a real orthogonal matrix P such that $P^{-1}AP$ is a diagonal matrix. The diagonalization is carried out by applying a series of orthogonal transformation $P_1, P_2, ..., P_n$ as follows :

Let a_{ij} be the largest element in magnitude amongst the off diagonal elements of A. Then a 2×2 sub matrix of A is formed as

$$A_1 = \begin{bmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{bmatrix} \qquad ...(7)$$

which can be transformed to a diagonal form. We construct an orthogonal matrix

as $$P_1 = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad ...(8)$$

We find θ such that A_1 is diagonalized, we have

$$P_1^{-1}A_1P_1 = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$= \begin{bmatrix} a_{ii}\cos^2\theta + a_{ij}\sin 2\theta + a_{jj}\sin^2\theta & \left(a_{jj} - a_{ii}\right)\sin\theta\cos\theta + a_{ij}\cos 2\theta \\ \left(a_{jj} - a_{ii}\right)\sin\theta\cos\theta + a_{ij}\cos 2\theta & a_{ii}\sin^2\theta - a_{ij}\sin 2\theta + a_{jj}\cos^2\theta \end{bmatrix}$$

For this expression to be diagonal we put

$$\frac{1}{2}\left(a_{jj}-a_{ii}\right)\sin 2\theta + a_{ij}\cos 2\theta = 0$$

$$\Rightarrow \qquad \tan 2\theta = \frac{2a_{ij}}{a_{ii}-a_{jj}} \qquad \text{...(9)}$$

We choose $-\frac{\pi}{4} \le \theta \le \frac{\pi}{4}$ in order to get last possible rotation.

Now we construct P_1 as

$$P_1 = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \dots & & & \dots & & & \dots & \\ 0 & 0 & \dots & \cos\theta & \dots & -\sin\theta & \dots & 0 \\ \dots & & & \dots & & & \dots & \\ 0 & 0 & \dots & \sin\theta & \dots & \cos\theta & \dots & 0 \\ \dots & & & \dots & & & \dots & \\ 0 & 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix}$$

Where $\cos\theta$, $-\sin\theta$, $\sin\theta$, $\cos\theta$ are inserted at (i,i), (i,j), (j,i), (j,j) positions respectively and elsewhere it is identical with a unit matrix. With the value of θ given by (9) the first step is now completed by computing $P_1^{-1}AP_1$. Next the largest off diagonal element in this matrix is found and the procedure is repeated, until the matrix A is diagonalized, with the eigenvalue on the main diagonal. The corresponding eigenvectors are the columns of P.

A disadvantage is there of the method that when elements replaced by another one through a plane rotation, they not necessarily remain zero during the trasformation. So we have to check that the value of $\left|\sin^2\theta + \cos^2\theta - 1\right|$ is sufficiently small.

In particular if the element a_{12} is largest in the magnitude among the off diagonal elements of the matrix of order 3×3, we use the transformation matrix

$$P = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Similarly for the other positions we have different transformation matrices.

Further it is noted that $\frac{n(n-1)}{2}$ is the minimum required number to transform the given $n \times n$ real symmetric matrix into a diagonal form.

Example 5.3 Use Jacobi method to compute eigenvalues of given matrix (two iterations only)

$$A=\begin{bmatrix}2 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 2\end{bmatrix}$$

Solution : The given matrix A is real symmetric in which all the off diagonal elements are of the same magnitude, therefore we can choose any one of them. Let us select the position a_{12}, then

$$\tan 2\theta=\frac{2a_{12}}{a_{11}-a_{22}}=\frac{2(-1)}{2-2}=-\infty$$

$$\Rightarrow \qquad \theta=-\pi/4$$

Hence $\cos\theta=1/\sqrt{2}$; $\sin\theta=-1/\sqrt{2}$

Then $P_1=\begin{bmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix}=\begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2} & 0\\ -1/\sqrt{2} & 1/\sqrt{2} & 0\\ 0 & 0 & 1\end{bmatrix}$

Now the first rotation is

$$A_1=P_1^{-1}AP_1$$

$$=\begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2} & 0\\ -1/\sqrt{2} & 1/\sqrt{2} & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}2 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 2\end{bmatrix}\begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2} & 0\\ -1/\sqrt{2} & 1/\sqrt{2} & 0\\ 0 & 0 & 1\end{bmatrix}$$

$$=\begin{bmatrix}3 & 0 & 1/\sqrt{2}\\ 0 & 1 & -1/\sqrt{2}\\ 1/\sqrt{2} & -1/\sqrt{2} & 2\end{bmatrix}$$

Now it is pretty obvious to consider a_{13} position of A_1, so that

$$\tan 2\theta=\frac{2a_{13}}{a_{11}-a_{33}}=\frac{2\left(1/\sqrt{2}\right)}{3-2}=\sqrt{2}$$

$\sin\theta=0.45970$; $\cos\theta=0.88807$

Now the second rotation will be

$$A_2=P_2^{-1}A_1P_2$$

$$= \begin{bmatrix} 0.8887 & 0 & 0.45970 \\ 0 & 1 & 0 \\ -0.45970 & 0 & 0.8887 \end{bmatrix} \begin{bmatrix} 3 & 0 & 1/\sqrt{2} \\ 0 & 1 & -1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} & 2 \end{bmatrix} \begin{bmatrix} 0.8887 & 0 & -0.45970 \\ 0 & 1 & 0 \\ -0.45970 & 0 & 0.8887 \end{bmatrix}$$

$$= \begin{bmatrix} 3.366 & -0.325057 & -0.000002 \\ -0.325057 & 1 & -0.627961 \\ -0.000002 & -0.627961 & 1.633962 \end{bmatrix}$$

Thus approximation to eigenvalues after two iterations are

3.366, 1 and 1.633962.

Example 5.4 Using Jacobi's method to find all the eigenvalues and eigenvectors of the following matrix A (perform three iterations)

$$A = \begin{bmatrix} 1 & 1 & 0.5 \\ 1 & 1 & 0.25 \\ 0.5 & 0.25 & 2 \end{bmatrix}$$

Solution : The numerically largest off diagonal values is $a_{12} = 1$

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2a_{12}}{a_{11} - a_{22}}\right)$$

Therefore

$$= \frac{1}{2}\tan^{-1}\left(\frac{2}{1-1}\right) = \frac{\pi}{4}$$

$$\Rightarrow \qquad \sin\theta = \frac{1}{\sqrt{2}} \quad ; \qquad \cos\theta = \frac{1}{\sqrt{2}}$$

Now

$$P_1 = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\therefore \qquad P_1 = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Now for the first rotation

$$A_1 = P_1^{-1} A P_1$$

$$=\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1/2 \\ 1 & 1 & 1/4 \\ 1/2 & 1/4 & 2 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$=\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{2} & 0 & 1/2 \\ \sqrt{2} & 0 & 1/4 \\ 3\sqrt{2}/8 & -\sqrt{2}/8 & 2 \end{bmatrix}$$

$$=\begin{bmatrix} 2 & 0 & 3\sqrt{2}/8 \\ 0 & 0 & -\sqrt{2}/8 \\ 3\sqrt{2}/8 & -\sqrt{2}/8 & 2 \end{bmatrix}$$

Again in this new matrix we have largest off diagonal element is $a_{13} = \frac{3}{8}\sqrt{2}$

Therefore $\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\,a_{13}}{a_{11}-a_{33}}\right) = \frac{1}{2}\tan^{-1}\left(\frac{\frac{3}{4}\sqrt{2}}{2-2}\right) = \frac{\pi}{4}$

Now, we have $\cos\theta = \frac{1}{\sqrt{2}}$; $\sin\theta = \frac{1}{\sqrt{2}}$

Hence $P_2 = \begin{bmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}$

Now for the second rotation

$$A_2 = P_2^{-1} A_1 P_2$$

$$=\begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 2 & 0 & 3\sqrt{8} \\ 0 & 0 & -\sqrt{2}/8 \\ 3\sqrt{2}/8 & -\sqrt{2}/8 & 2 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}$$

$$=\begin{bmatrix} 2+\frac{3}{8}\sqrt{2} & -\frac{1}{8} & 0 \\ -\frac{1}{8} & 0 & -\frac{1}{8} \\ 0 & -\frac{1}{8} & 2-\frac{3}{8}\sqrt{2} \end{bmatrix}$$

$$=\begin{bmatrix} 2.5303 & -0.125 & 0 \\ -0.125 & 0 & -0.125 \\ 0 & -0.125 & 1.4697 \end{bmatrix}$$

Now let us move to third iteration using this matrix A_2 in which the numerically largest off diagonal element is $a_{12} = 0.125$,

Therefore $\theta = \frac{1}{2}\tan^{-1}\left(\frac{2a_{33}}{a_{11}-a_{22}}\right) = \frac{1}{2}\tan^{-1}\left(\frac{-0.250}{2.5303}\right) = -0.0492$ (radian)

Now we have $\cos\theta = 0.9988$; $\sin\theta = -0.0492$

Hence $P_3 = \begin{bmatrix} 0.9988 & 0.0492 & 0 \\ -0.0492 & 0.9988 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Now for the third rotation

$$A_3 = P_3^{-1} A_2 P_3$$

$$=\begin{bmatrix} 0.9988 & -0.0492 & 0 \\ 0.0492 & 0.9988 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2.5303 & -0.125 & 0 \\ -0.125 & 0 & -0.125 \\ 0 & -0.125 & 1.4697 \end{bmatrix}\begin{bmatrix} 0.9988 & 0.0492 & 0 \\ -0.0492 & 0.9988 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$=\begin{bmatrix} 2.5365 & -0.0001 & 0.0062 \\ -0.0001 & -0.0062 & -0.1249 \\ 0.0062 & -0.1249 & 1.4697 \end{bmatrix}$$

Hence the eigenvalues approximately

2.5365, -0.0062 and 1.4697.

Self-Learning Exercise - 3

1. Use Jacobi method to estimate the eigenvalues of the following matrix

$$\begin{bmatrix} 1 & -2 & 4 \\ -2 & 5 & -2 \\ 4 & -2 & 1 \end{bmatrix}$$

2. Perform two iterations of Jacobi method to estimate the eigenvalues of the given matrix

$$\begin{bmatrix} 1 & \sqrt{2} & 2 \\ \sqrt{2} & 3 & \sqrt{2} \\ 2 & \sqrt{2} & 1 \end{bmatrix}$$

5.6 Summary

By the study of this particular unit one can understand how to find the eigenvalues and corresponding vectors by using two great methods namely Power method and Jacobi method. Besides this one can learn the properties of the eigenvalues and thier vactors, which can be used in further studies, especially in engineering courses.

Here described, the power method is basically designed to obtain the largest eigenvalues and corresponding vector for the given matrix. And also one can find the lowest eigenvalue through the same method by using the properties of the eigenvalues. Whereas the Jacobi method is recommended for real symmetric matrix.

If only one eigenvalue (the dominant or the least dominant) and the corresponding vectors are required, the Power method and the inverse Power method are suited. The convergence rate of the Power method is poor, when two largest eigenvalues are nearly equal in the magnitude. When all the eigenvalues are required Jacobi's method is to be used, but only in the case when the matrix is symmetric. Jacobi's method takes many rotations to reduce the matrix to the diagonal form but still it is reliable and converges with accuracy.

5.7 Answers of Self-Learning Exercise

Self-Learning Exercise - 1

1. Eigenvalues are 6, 3, 3 and vectors are $\begin{pmatrix}3 & 2 & 4\end{pmatrix}^T$ and $\begin{pmatrix}0 & 1 & -1\end{pmatrix}^T$

2. Eigenvalues are 0, 3 and vectors are $\left(-1/\sqrt{3} \quad \sqrt{2/3}\right)^T$ and $\left(\sqrt{2/3} \quad 1/\sqrt{3}\right)^T$

Self-Learning Exercise - 2

1. Eigenvalue is 4 and Eigenvector is $\begin{pmatrix}1 & -1 & 1\end{pmatrix}^T$.

2. Eigenvalue is 4 and Eigenvector is $\begin{pmatrix}1 & 0.5 & 0\end{pmatrix}^T$, other roots are -land 3.

Self-Learning Exercise - 3

1. Eigenvalues are $5-2\sqrt{2}$, $5+\sqrt{2}$, -3.

2. Eigenvalues are 5, 1, −1.

5.8 Exercises

1. Find all the eigenvalues of the matrix

$$A = \begin{bmatrix} -5 & 2 & 1 \\ 1 & -9 & -1 \\ 2 & -1 & 7 \end{bmatrix}$$

2. Obtain the largest eigenvalue and corresponding vector of the matrix

$$A = \begin{bmatrix} 1 & 3 & -1 \\ 3 & 2 & 4 \\ -1 & 4 & 10 \end{bmatrix}$$

3. Using the Power method obtain the dominant eigenvalue and corresponding vector of the matrix

$$A = \begin{bmatrix} 2 & 3 & 2 \\ 4 & 3 & 5 \\ 3 & 2 & 9 \end{bmatrix}$$

4. Find the dominant eigenvalue and corresponding vector of the matrix

$$A = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 20 & 1 \\ 0 & 0 & 4 \end{bmatrix}$$, using Power method (four iterations) with intial vector

$(0, 0, 1)^T$.

5. Find all the eigenvalue and eigenvectors of the matrix

$$A = \begin{bmatrix} 1 & \sqrt{2} & 2 \\ \sqrt{2} & 3 & \sqrt{2} \\ 2 & \sqrt{2} & 1 \end{bmatrix}$$, using Jacobi method.

6. Find all the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 1 & \sqrt{3} & 4 \\ \sqrt{3} & 4 & \sqrt{3} \\ 4 & \sqrt{3} & 1 \end{bmatrix}$$, using Jacobi method.

□□□

Unit - 6 : Eigen Value Problems - II

Structure of the Unit

6.0 Objectives

In this particular unit let us introduce two new methods say Given's method and Rutishuaser method to obtain eigenvalues and corresponding vectors. Definitely thses two methods are more efficient than them, we learned in previous unit. By studying this unit one can learn to reduce a given matrix to a tridiagonal form and also about the eigenvalues of the complex matrices.

6.1 Introduction

In the previous unit we have learnt two important methods to obtain the eigenvalues and their corresponding vectors. In this section we are going to learn two more methods to obtain the same things. First is the Given's method, which converts the given matrix into a tridiagonal matrix and hence open an algorithm to find the required eigenvalues. Tridiagonal matrix is the matrix having non-zero entries only in the leading diagonal, sub diagonal and super diagonal. For example

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix} \qquad \text{...(1)}$$

be a symmetrical tridiagonal matrix.

Beside this one more method is there for finding out appropriate similarity transformation are based on matrix decomposition. The Rutishauser method proposed a LU decomposition of matrix, where L is a lower triangular matrix and U is upper triangular matrix.

Also in this unit we are going to deal with complex eigenvalues or eigenvalues of complex matrices.

6.2 Given's Method

Let A be a real symmetric matrix. For such matrices, it is natural to apply transformation which preserve symmetry. The Jacobi method gives a sequence of orthogonal transformations, which diagonalizes a given real symmetric matrix iteratively. This transformation does not preserve the zeros already present in the matrix, however the matrix can be reduced to a symmetrical tridiagonal form and in this case it is possible to arrange the sequence of transformation such that zero elements introduced in the previous steps are preserved.

To avoid this Given's proposed an algorithm using plane rotation which preserves zeros in the off diagonal elements once they are created. This method reduces the given matrix into a tridiagonal matrix using plane rotation and form a sturm sequence which determines the eigenvalues and hence eigenvectors obtained. In this case we start with the subspace having the elements $a_{22}, a_{23}, a_{32}, a_{33}$. Perform a plane rotation $P_1^{-1}AP_1$ using the orthogonal matrix

$$P_1^* = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad ...(2)$$

If we choose $\tan\theta = \frac{a_{13}}{a_{12}}$ ...(3)

Taking this value of θ we obtain zeros in the (1, 3) and (3, 1) positions, after performing the plane rotation. Then we perform rotation in (2, 4) subspace, putting $a_{14} = a_{41} = 0$. This transformation will not affect zeros already obtained earlier. Proceeding similarly, by performing rotation in the subspace (2, 5),......, (2, n); (3, 4),........, (3, n) etc, we obtain the tridiagonal matrix. If the order of matrix is n, and then it requires $\frac{(n-1)(n-2)}{2}$ rotations to reduce it to tridiagonal form D. matrices A and D have same eigenvalues.

Let $f_n(\lambda) = |\lambda I - D|$ ...(4)

So that $f_n(\lambda) = 0$, is characteristic equation.

Expanding the equation (4), we find

$$f_0 = 1,\ f_1 = \lambda - d_1,$$

$$f_r = (\lambda - d_r)f_{r-1} - c_{r-1}^2 f_{r-2}\ ;\ 2 \le r \le n \qquad ...(5)$$

where

$$D = \begin{bmatrix} d_1 & c_1 & & 0 \\ c_1 & d_2 & c_2 & \\ & & d_{n-1} & c_{n-1} \\ 0 & & c_{n-1} & d_n \end{bmatrix} \qquad ...(6)$$

The eigenvalues can be determined by finding zeros of the determinants of $(\lambda I - D)$. It turns out that in this case, the sequence of $f_r(\lambda)$ forms a sturm sequence, which can be effectively used to determine any eigenvalue.

Example 6.1 Transform the following matrix to tridiagonal form by Given's method

$$A=\begin{bmatrix}1 & 2 & 3\\ 2 & 1 & -1\\ 3 & -1 & 1\end{bmatrix}$$

Solution : Here $\tan\theta=\dfrac{a_{13}}{a_{12}}=\dfrac{3}{2},$

So $\sin\theta=\dfrac{3}{\sqrt{13}}$; $\cos\theta=\dfrac{2}{\sqrt{13}}$

To create zero at (1, 3) position we write

$$P_1=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{bmatrix}=\begin{bmatrix}1 & 0 & 0\\ 0 & 2/\sqrt{13} & -3/\sqrt{13}\\ 0 & 3/\sqrt{13} & 2/\sqrt{13}\end{bmatrix}$$

Now

$$P_1^T AP_1=\begin{bmatrix}1 & 0 & 0\\ 0 & 2/\sqrt{13} & 3/\sqrt{13}\\ & -3/\sqrt{13} & 2/\sqrt{13}\end{bmatrix}\begin{bmatrix}1 & 2 & 3\\ 2 & 1 & -1\\ 3 & -1 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 2/\sqrt{13} & -3/\sqrt{13}\\ 0 & 3/\sqrt{13} & 2/\sqrt{13}\end{bmatrix}$$

$$=\begin{bmatrix}1 & \sqrt{13} & 0\\ \sqrt{13} & 1/13 & 5/13\\ 0 & 5/13 & 25/13\end{bmatrix}$$

This is the required tridiagonal form.

Example 6.2 Using the Given's method reduce the following matrix to tridiagonal form and use sturm sequence to find eigenvalues

$$A=\begin{bmatrix}1 & 2 & 2\\ 2 & 1 & -1\\ 2 & -1 & 1\end{bmatrix}$$

Solution : Performing the orthogonal rotation with respect to $a_{22}, a_{23}, a_{32}, a_{33}$, we get

$$\tan\theta=\frac{a_{13}}{a_{12}}=\frac{2}{2}=1$$

So $\sin\theta=\dfrac{1}{\sqrt{2}}$; $\cos\theta=\dfrac{1}{\sqrt{2}}$

Then we have orthogonal matrix P in the plane (2, 3) as

$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Now we have

$$A_1 = P_1^{-1} A P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & -1 \\ 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 2\sqrt{2} & 0 \\ 2 & 0 & -\sqrt{2} \\ 2 & 0 & \sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2\sqrt{2} & 0 \\ 2\sqrt{2} & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

This is the required tridiagonal form. The strum sequence is

$$f_0 = 1,\ f_1 = \lambda - 1$$

$$f_2 = (\lambda - 1) f_1 - \left(2\sqrt{2}\right)^2 f_0$$

$$= \lambda^2 - \lambda - 8$$

$$f_3 = (\lambda - 2) f_2 - 0 f_1$$

$$= (\lambda - 2)\left(\lambda^2 - \lambda - 8\right)$$

$$= \lambda^3 - 3\lambda^2 - 6\lambda + 16$$

It can be observed that $f_3(2) = 0$, so that $\lambda = 2$ is an eigenvalue.

Now we get

λ	f_0	f_1	f_2	f_3	$v(\lambda)$
-3	$+$	$-$	$+$	$-$	3
-2	$+$	$-$	$-$	$+$	2
-1	$+$	$-$	$-$	$+$	2
0	$+$	$-$	$-$	$+$	2
1	$+$	$0(+)$	$-$	$+$	2
2	$+$	$+$	$-$	$0(-)$	1
3	$+$	$+$	$-$	$-$	1
4	$+$	$+$	$+$	$+$	0

$v(\lambda)$ shows the changes of sign. Now there is an eigenvalue in the interval $(-3, -2)$ and $(3,4)$. Now better estimates of the eigenvalue can be found by Newton Raphson method. Let $\lambda = -2.5$ be initial approximation in the interval $(-3, -2)$. By the definition we have

$$\lambda_{n+1} = \lambda_n - \frac{f_3(\lambda_n)}{f_3'(\lambda_n)}, \; n = 0,1,2,3,.....$$

$$= \lambda_n - \frac{\lambda_n^3 - 3\lambda_n^2 - 6\lambda_n + 16}{3\lambda_n^2 - 6\lambda_n - 6}$$

First approximation :

$$\lambda_1 = \lambda_0 - \frac{\lambda_0^3 - 3\lambda_0^2 - 6\lambda_0 + 16}{3\lambda_0^2 - 6\lambda_0 - 6}$$

$$= -2.5 - \frac{(-2.5)^3 - 3\,(-2.5)^2 - 6\,(-2.5) + 16}{3(-2.5)^2 - 6(-2.5) - 6}$$

$$= -2.5 + 0.12162162$$

$$= -2.3783783$$

Second approximation :

$$\lambda_2 = \lambda_1 - \frac{\lambda_1^3 - 3\lambda_1^2 - 6\lambda_1 + 16}{3\lambda_1^2 - 6\lambda_1 - 6}$$

$$= -2.3783783 - \frac{(-2.3783783)^3 - 3(-2.3783783)^2 - 6(-2.3783783) + 16}{3(-2.3783783)^2 - 6(-2.3783783) - 6}$$

$$= -2.372301615 \cdot$$

Hence $\lambda = -2.37$ is the least eigenvalue correct to two decimal places. Similarly we can find the remaining eigenvalue and it is $\lambda = 3.37$.

The exact eigenvalue are $\lambda = 2, \; \frac{1 \pm \sqrt{33}}{2}$.

Example 6.3 Find all the eigenvalues and eigenvectors of the following matrix using Given's method

$$A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 5 & 1 \\ 2 & 1 & 6 \end{bmatrix}$$

Solution : According to the Given's method as per required we have

$$\theta = \tan^{-1}\left(\frac{a_{13}}{a_{12}}\right) = \tan^{-1}\left(\frac{2}{2}\right) = \frac{\pi}{4}$$

$$\therefore \qquad P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Now we have

$$A_1 = P_1^{-1}AP_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 4 & 2 & 2 \\ 2 & 5 & 1 \\ 2 & 1 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 4 & 2\sqrt{2} & 0 \\ 2 & 3\sqrt{2} & -2\sqrt{2} \\ 2 & 7\sqrt{2}/2 & 5\sqrt{2}/2 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 2\sqrt{2} & 0 \\ 2\sqrt{2} & 13/2 & 1/2 \\ 0 & 1/2 & 9/2 \end{bmatrix}$$

This is the tridiagonal matrix. The characteristic equation of A_1 is

$$= \begin{vmatrix} 4-\lambda & 2\sqrt{2} & 0 \\ 2\sqrt{2} & \frac{13}{2}-\lambda & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{9}{2}-\lambda \end{vmatrix} = 0$$

Now the sturm sequence is given by

$$f_0(\lambda) = 1 \ ; \ f_1(\lambda) = 4-\lambda$$

$$f_2(\lambda) = \left(\frac{13}{2}-\lambda\right) f_1(\lambda)$$

$$f_3(\lambda) = \left(\frac{9}{2}-\lambda\right) f_2(\lambda) - \frac{1}{4} f_1(\lambda)$$

Let's now consider the changes in the sign in strum sequence given as

λ	f_0	f_1	f_2	f_3	$v(\lambda)$
0	+	+	+	+	0
1	+	+	+	+	0
2	+	+	+	+	0
3	+	+	−	−	1
4	+	0	−	−	1
5	+	−	−	+	2
6	+	−	−	+	2
7	+	−	−	+	2
8	+	−	−	+	2
9	+	−	+	−	3

As the table of changes of sign indicates, the roots of the equation $f_3(\lambda)=0$ lie on each in (2, 3), (4, 5) and (8, 9).

Now $f_3(\lambda)=\left(\frac{9}{2}-\lambda\right)\left\{\left(\frac{13}{2}-\lambda\right)(4-\lambda)-8\right\}-\frac{1}{4}(4-\lambda)=0$

i.e. $f_3(\lambda)=\lambda^3-15\lambda^2+65\lambda-80$

Now, choosing the first domain i.e. $2<\lambda_1<3$; $\lambda_1^{(0)}=2$

We use the Newton Raphson method to find λ_1

$$\lambda_1^{(1)}=2-\frac{f(2)}{f'(2)}=2+\frac{2}{17}=2.1176$$

$$\lambda_1^{(2)}=2.1176-\frac{f(2.1176)}{f'(2.1176)}=2.1258$$

$$\lambda_1^{(3)}=2.1258-\frac{f(1258)}{f'(1258)}=2.1259$$

$$\lambda_1^{(4)}=2.1259-\frac{f(1259)}{f'(1259)}=2.1259$$

$\therefore$ $\lambda_1=2.1259$

Similarly we can find the other two values as

$\lambda_2=4.4867$; $\lambda_3=8.3874$.

The eigenvector $(y_1, y_2, y_3)^T$ of A_1 corresponding to $\lambda_1 = 2.1259$ is given by

$$1.8741y_1 + 2.8284y_2 = 0$$

$$2.8284y_1 + 4.3741y_2 + 0.5y_3 = 0$$

$$0.5y_2 + 2.3741y_3 = 0$$

Solving these equations we get

$$\frac{y_1}{1.4142} = \frac{y_2}{-0.9371} = \frac{y_3}{0.1977}$$

Now the eigenvector $(x_1, x_2, x_3)^T$ of A corresponding to $\lambda_1 = 2.1259$ is given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = A_1 \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1.4142 \\ -0.9371 \\ 0.1977 \end{bmatrix}$$

$$\Rightarrow \qquad X_1 = \begin{bmatrix} 1.4142 \\ -0.8024 \\ -0.5228 \end{bmatrix} = \begin{bmatrix} 1 \\ -0.5674 \\ -0.3697 \end{bmatrix}$$

Similarly we can get other eigenvectors X_1 and X_2 of A as

$$X_2 = \begin{bmatrix} 0.2170 \\ 1 \\ -0.9473 \end{bmatrix} \quad ; \quad X_3 = \begin{bmatrix} 0.8077 \\ 0.7720 \\ 1 \end{bmatrix}$$

Self-Learning Exercise - 1

1. Transform the following matrix to tridiagonal form applying Given's method

$$A = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}$$

2. Reduce the following matrix A to the tridiagonal form using Given's method. Use sturm sequence to locate the eigenvalues

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

6.3 Rutishauser Method

In the previous unit and section we already studied the methods depending on rotation and reflection transformations. The Rutishauser method is different method which depends on the LU decomposition. Where L stands for lower triangular matrix and U for upper triangular matrix.

Let the given matrix be A, then we write

$$A = A_1$$

And split A_1 into two triangular matrices

$$A_1 = L_1 U_1 \qquad ...(7)$$

Taking $l_{ii} = 1$.

Now form

$$A_2 = U_1 L_1 \qquad ...(8)$$

We find that A_1 and A_2 have the same eigenvalues, since

$$A_2 = U_1 L_1 = U_1 A_1 U_1^{-1}$$

Again split

$$A_2 = L_2 U_2 \qquad ...(9)$$

Taking $l_{ii} = 1$ and form

$$A_3 = U_2 L_2 \qquad ...(10)$$

Where A_2 and A_3 have the same eigenvalues. Proceeding in the same manner we obtain a sequence of matrices, which in general reduces to an upper triangular matrix and its leading diagonal entries are the eigenvalues of A.

Example 6.1 Using the Rutishauser method, find all the eigenvalues of the matrix

$$A = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$$

Solution : By the procedure of the method let us start with the matrix $A = A_1$, we split it into two triangular matrices

$$A = A_1 = L_1 U_1 = \begin{bmatrix} 1 & 0 \\ l_{21} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix}$$

$$= \begin{bmatrix} u_{11} & u_{12} \\ l_{21} u_{11} & l_{21} u_{12} + u_{22} \end{bmatrix}$$

Now by solving this matrix we have

$$u_{11} = 4\,,\; u_{12} = 3\,,\; l_{21} = \frac{1}{4}\,,\; u_{22} = \frac{5}{4}$$

Hence we have

$$U_1 = \begin{bmatrix} 4 & 3 \\ 0 & 5/4 \end{bmatrix} \text{ and } L_1 = \begin{bmatrix} 1 & 0 \\ 1/4 & 1 \end{bmatrix}$$

At the second stage let us form

$$A_2 = U_1 L_1 = \begin{bmatrix} 4 & 3 \\ 0 & 5/4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1/4 & 1 \end{bmatrix} = \begin{bmatrix} 19/4 & 3 \\ 5/16 & 5/4 \end{bmatrix}$$

Now again decomposing the matrix we have,

$$A_2 = U_2 L_2$$

$$\begin{bmatrix} 19/4 & 3 \\ 5/16 & 5/4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ l_{21} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix}$$

Again in the same manner multiply the RHS matrices and hence compare with the values of the matrix A_2, it yields

$$l_{21} = \frac{5}{76}\,,\; u_{11} = \frac{19}{4}\,,\; u_{21} = 3\,,\; u_{22} = \frac{20}{19}$$

On the same lines let us again form

$$A_3 = U_2 L_2 = \begin{bmatrix} 19/4 & 3 \\ 0 & 20/4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 5/76 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 4.9473 & 3 \\ 0.6925 & 1.0526 \end{bmatrix}$$

Similarly repeating the same process we find

$$A_4 = \begin{bmatrix} 4.9893 & 3 \\ 0.1415 & 1.0106 \end{bmatrix}$$

$$A_5 = \begin{bmatrix} 4.9978 & 3 \\ 0.0028 & 1.0021 \end{bmatrix}$$

One can see easily that the value at the position a_{21} is tending to zero i.e. the above sequence at last will converge to an upper triangular matrix and the diagonal elements will be the eigenvalues of the given matrix A. the exact eigenvalues will be 5 and 1.

Example 6.2 Using the Rutishauser method to compute all the eigenvalues of the matrix

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

Solution : Starting with the matrix $A = A_1$, we split it into two triangular matrices as

$$A_1 = L_1 U_1 = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

On solving this we obtain

$$l_{21} = -0.5,\ l_{31} = 0,\ l_{32} = -0.6667,$$

$$u_{11} = 2,\ u_{12} = -1,\ u_{13} = 0,\ u_{23} = -1$$

$$u_{22} = 1.5,\ u_{33} = 0.3333.$$

Then we form

$$A_1 = U_1 L_1 = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1.5 & -1 \\ 0 & 0 & 0.3333 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -0.5 & 1 & 0 \\ 0 & -0.6667 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2.5 & -1 & 0 \\ -0.75 & 2.1667 & -1 \\ 0 & -0.222 & 0.333 \end{bmatrix}$$

Decomposing again as $A_2 = L_2 U_2$, proceeding exactly same as the previous step we have

$$A_2 = \begin{bmatrix} 1 & 0 & 0 \\ -0.3 & 1 & 0 \\ 0 & -0.119 & 1 \end{bmatrix} \begin{bmatrix} 2.5 & -1 & 0 \\ 0 & 1.8667 & -1 \\ 0 & 0 & 0.2143 \end{bmatrix}$$

Then we form

$$A_3 = U_2 L_2 = \begin{bmatrix} 2.5 & -1 & 0 \\ 0 & 1.8667 & -1 \\ 0 & 0 & 0.2143 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -0.3 & 1 & 0 \\ 0 & -0.119 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2.8 & -1 & 0 \\ -0.56 & 1.9857 & -1 \\ 0 & -0.00255 & 0.2143 \end{bmatrix}$$

Decomposing now $A_3 = L_3U_3$, we have

$$A_3 = \begin{bmatrix} 1 & 0 & 0 \\ -0.2 & 1 & 0 \\ 0 & -0.0143 & 1 \end{bmatrix}\begin{bmatrix} 2.8 & -1 & 0 \\ 0 & 1.7857 & -1 \\ 0 & 0 & 0.2 \end{bmatrix}$$

Now let us form $A_4 = U_3L_3$, hence

$$A_4 = U_3L_3 = \begin{bmatrix} 2.8 & -1 & 0 \\ 0 & 1.7857 & -1 \\ 0 & 0 & 0.2 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -0.2 & 1 & 0 \\ 0 & -0.143 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & -1 & 0 \\ -0.3571 & 1.8 & -1 \\ 0 & -0.0028 & 0.1981 \end{bmatrix}$$

Proceeding similarly we find

$$A_{15} = \begin{bmatrix} 3.2469 & -1 & 0 \\ -0.0001 & 1.5550 & -1 \\ 0 & 0 & 0.1981 \end{bmatrix}$$

Hence the approximate eigenvalues of the given matrix are

3.2469 ; 1.5550 ; 0.1981

Self-Learning Exercise - 2

1. Using Rutishauser method, compute all the eigenvalues of the matrix

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}$$

2. Find approximately the eigenvalues of the following matrix, using Rutishauser method

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 3 & 2 \end{bmatrix}$$

6.4 Complex Eigen Values

So far, we have considered matrices whose elements were real numbers. The elements of a matrix can be complex numbers also; such matrix is known as complex matrix. If the elements of matrix $A = [a_{rs}]$ are complex numbers $\alpha_{rs} + \beta_{rs}$, α_{rs} and β_{rs} being real, then the matix $\overline{A} = [\overline{\alpha}_{rs}] = [\alpha_{rs} - i\beta_{rs}]$ is called the conjugate matrix of A. the transpose of a conjugate of a matrix A is denoted as A^*. Let us define now important definitions.

Hermitian matrix : A square matrix A such that $A^{/} = \overline{A}$ (transporse of the matrix is equal to the conjugate of the same) is said to be a Hermitian matrix. The elements of the leading diagonal of this matrix are real, while every other element is the complex conjugate of the element in the transpose position.

For example $A = \begin{bmatrix} 3 & 2+4i \\ 2-4i & -8 \end{bmatrix}$ is a Hermitian matrix.

Since $A^{/} = \begin{bmatrix} 3 & 2-4i \\ 2+4i & -8 \end{bmatrix} = \overline{A}$.

But in case if $A^{/} = -\overline{A}$, then matrix A is said to be a skew-Hermitian matrix. This implies that the principal diagonal elements of skew-Hermitian matrix are either all zero or all are purely imaginary.

Unitary Matrix : A square matrix U such that $\overline{U}^{/} = U^{-1}$ is called unitary matrix. For a unitary matrix U, $U.U^{*} = U^{*}.U = I$.

Now to determine the eigenvalues of the matrix A with complex elements we try to reduce the problem of determining the eigenvalues of a real matrix.

For this let us consider a matrix

$$A = B + iC$$

And let λ is an eigenvalues of the matrix A and x is the eigenvector, then

$$(B+iC)x = \lambda x$$

$$\Rightarrow \qquad (C-iB)x = -i\lambda x$$

Or it can be written as

$$\begin{bmatrix} B & -C \\ C & B \end{bmatrix}\begin{bmatrix} x \\ -ix \end{bmatrix} = \lambda\begin{bmatrix} x \\ -ix \end{bmatrix}$$

It can be seen easily that λ is also an eigenvalues of the real matrix

$$E = \begin{bmatrix} B & -C \\ C & B \end{bmatrix}, \text{with eigenvector } \begin{bmatrix} x & -ix \end{bmatrix}^{T}.$$

One can observe easily that this method doubles the order of the matrix.

Further if A is Hermitian matrix then it can be diagonalized using similarity transformation involving unitary matrices. Thus the Jacobi method can be applied by replacing the orthogonal matrix Q by unitary matrix U.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \text{ be a Hermitian matrix.}$$

Here a, d are real and $b = \overline{c}$ or $\overline{b} = c$ so that $A = A^{*}$.

Let $U = \begin{bmatrix} p & -\bar{q} \\ q & p \end{bmatrix}$

Where p is real and $UU^* = 1$, i.e. $U^* = U^{-1}$

We choose the unitary matrix U such that the matrix $U^{-1}AU$ is diagonalized.

This unitary matrix is obtained by evaluating p and q as

$$p = \left(1 + |k|^2\right)^{-1/2},$$

$$q = kp,$$

$$k = \frac{1}{2b}\left[(d-a) \pm \sqrt{(d-a)^2 + 4bc}\right]$$

We chose the sign that makes k small.

Now, consider the given Hermitian matrix

$$A = \begin{bmatrix} 1 & 1-i \\ 1+i & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Here $a = 1$; $b = 1-i$; $c = 1+i$; $d = 1$, then

$$k = \frac{1}{2(1-i)}\left[0 \pm \sqrt{0 + 4(1-i)(1+i)}\right]$$

$$= \frac{\pm\sqrt{2}}{1-i} = \pm\frac{(1+i)}{\sqrt{2}}$$

$$|k| = 1.$$

Now, unitary matrix

$$U = \begin{bmatrix} p & -q \\ q & p \end{bmatrix}$$

Where $p = (1+1)^{-1/2} = \dfrac{1}{\sqrt{2}}$; $q = \dfrac{1+i}{2}$

Therefore $U = \begin{bmatrix} 1/\sqrt{2} & -(1-i)/2 \\ (1+i)/2 & 1/\sqrt{2} \end{bmatrix}$

Hence,

$$U^{-1}AU = \begin{bmatrix} 1/\sqrt{2} & (1-i)/2 \\ -(1+i)/2 & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 1 & (1-i) \\ (1+i) & 1 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & -(1-i)/2 \\ (1+i)/2 & 1/\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} 1+\sqrt{2} & 0 \\ 0 & 1-\sqrt{2} \end{bmatrix}$$

Thus the eigenvalues of the Hermitian matrix A are $\left(1 \pm \sqrt{2}\right)$.

Self-Learning Exercise - 3

1. Find the eigenvalues of the following Hermitian matrix

$$A = \begin{bmatrix} 2 & -4i & 0 \\ 4i & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

2. Find the eigenvalues of the following Hermitian matrix

$$A = \begin{bmatrix} 2 & 1-2i \\ 1+2i & -2 \end{bmatrix}$$

6.5 Summary

In this present unit we learnt two different methods to obtain the eigenvalues. The first method is Given's method in which it is proposed an algorithm using plane rotation which preserves zeros in the off diagonal elements once they are created. This method reduces the given matrix into a tridiagonal matrix using plane rotation and form a sturm sequence which determines the eigenvalues and hence eigenvectors obtained.

If the order of matrix is n, then it requires $\frac{(n-1)(n-2)}{2}$ rotations to reduce it to tridiagonal form D. matrices A and D have same eigenvalues.

Let $f_n(\lambda) = |\lambda I - D|$

So that $f_n(\lambda) = 0$, is characteristic equation.

Expanding the equation (4) we find

$$f_0 = 1,\ f_1 = \lambda - d_1,$$

$$f_r = (\lambda - d_r) f_{r-1} - c_{r-1}^2 f_{r-2} \ ;\ 2 \le r \le n$$

where

$$D = \begin{bmatrix} d_1 & c_1 & & 0 \\ c_1 & d_2 & c_2 & \\ & & d_{n-1} & c_{n-1} \\ 0 & & c_{n-1} & d_n \end{bmatrix}$$

The eigenvalues can be determined by finding zeros of the determinants of $(\lambda I - D)$. It turns out that in this case, the sequence of $f_r(\lambda)$ forms a sturm sequence, which can be effectively used to determie any eigenvalue.

Another method is the Rutishauser method, a different method which depends on the LU decomposition. Where L stands for lower triangular matrix and U for upper trianguler matrix. Let the given matrix be A, then we write

$$A = A_1$$

And split A_1 into two triangular matrices

$$A_1 = L_1U_1 \text{ ; Taking } l_{ii} = 1.$$

Now form $A_2 = U_1L_1$

Where, A_1 and A_2 have the same eigenvalues. Proceeding in the same manner we obtain a sequence of matrices, which in general reduces to an upper triangular matrix and its leading diagonal entries are the eigenvalues of A. Next we studied a brief introduction of the complex matrices and their eigenvalues, espacialy in the case of that the given matrix is Hermitian. This matrix basically has a property that transpose of the matrix is equal to the conjugate of the same. The elements of the leading diagonal of this matrix are real, while every other element is the complex conjugate of the element in the transpose position. In this case the matrix can be diagonalized using similarity transformation involving unitary matrices. Thus the Jacobi method can be applied by replacing the orthogonal matrix by unitary matrix.

6.6 Answer of Self-Learning Exercise

Self-Learning Exercise - 1

1. $$A = \frac{1}{13}\begin{bmatrix} 1 & 13/6 & 0 \\ 13/6 & 34/5 & 9/20 \\ 0 & 9/20 & 2/15 \end{bmatrix}$$

2. $\lambda = 1$ and another two lie in intervals (2, 3) and (5, 6).

Self-Learning Exercise - 2

1. $\lambda_1 = 3.413792$; $\lambda_2 = 0.586207$

2. $\lambda_1 = 4.7912$; $\lambda_2 = -0.9998$; $\lambda_3 = 0.2085$

Self-Learning Exercise - 3

1. $\lambda_1 = 6$; $\lambda_2 = -2$; $\lambda_3 = 4$

2. $\lambda_1 = 3$; $\lambda_2 = -3$

6.7 Exercises

1. Transform the matrix

$$A = \begin{bmatrix} 2 & 1 & \sqrt{3} \\ 1 & 2 & \sqrt{3} \\ \sqrt{3} & \sqrt{3} & 3 \end{bmatrix}$$

to tridiagonal form using Given's method. Using sturm sequence, obtain eigenvalues.

2. Using Given's method transform the following matrix A to the tridiagonal form and compute the largest eigenvalue

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$$

3. Compute the eigenvalues using Rutishauser method of the following matrix

$$A = \begin{bmatrix} 6 & 4 & 4 & 1 \\ 4 & 6 & 1 & 4 \\ 4 & 1 & 6 & 4 \\ 1 & 4 & 4 & 6 \end{bmatrix}$$

4. Compute the eigenvalues using Rutishauser method of the following matrix

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}$$

5. Find the eigenvalues of the following Hermitian matrix

$$A = \begin{bmatrix} 2 & -i & 0 \\ i & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

□□□

Unit - 7 : Curve Fitting and Function Approximations

Structure of the Unit

7.0 Objectives

In this unit, we shall study about a principle called **'Least-Squares Principle'** which gives us a way to fit a desired curve to a set of given discrete data. We shall also study least-square approximation for a continuous function on an given interval.

7.1 Introduction

In this unit, we shall consider the process of approximating a function when the function is known only in the form of a table values. We shall also study the process to obtain polynomial approximation to the given continuous function on the given interval. Using least-square approximations. In this unit, approximations are obtained by minimising sum or integral of the squares of the error.

7.2 Least-Squares Principle

Let $\{x_i, y_i\}$ be an approximate set of given data. Now we have to fit a curve to this given data. To find the approximate equation of this curve, which passes through as many data (point) as possible, is called **curve-fitting**. Let this curve be $Y = f(x)$ be fitted to the given data $\{x_i, y_i\}$, $i = 1,2,.......m$.

Let $f(x_i) = Y_i$ be the value obtained by substituting $x = x_i$ in the equation of the curve, then $y_i - Y_i = e_i$ (say), will be the error of approximation at $x = x_i$ for $i = 1,2,.....,m$.

Let $$S = \sum_{i=1}^{m} e_i^2 = \sum_{i=1}^{m} (y_i - Y_i)^2 \qquad ...(1)$$

The least-squares principle requires S to be minimum. If all the point lie on the approximated curve $Y = f(x)$, then S will be zero. Thus, the curve $Y = f(x)$ will be the best approximation, for the given data, if S is least. This is known as **least-squares principle.**

7.3 Linear Regression or Fitting a Straight Line

Let $\{(x_i, y_i)|i = 1,2,......,m\}$ be a set of observations. We have to fit a straight line

$$Y = a + bx \qquad ...(2)$$

to the given data. Let at $x = x_i$, Y_i be the expected value, then

$$Y_i = a + bx_i$$

and corresponding observed value is y_i. Let e_i be the error at $x = x_i$, then

$$e_i = y_i - Y_i, \qquad i = 1,2,.......,m$$

or $$e_i = y_i - (a + bx_i), \qquad i = 1,2,......m.$$

The sum of squares S (say) of this error is given by

$$S = \sum_{i=1}^{m}\left[y_i - (a + bx_i)^2\right] \qquad ...(3)$$

Least squares principle requires that S be minimum. From (3), it is clear that S depends an a and b, that is, S is a function of a and b. Thus, we have to find the value of a and b so that S become minimum. By the theory of **maxima-minima,** the necessary conditions for S to be minimum are

$$\frac{\partial S}{\partial a} = 0 = \frac{\partial S}{\partial b},$$

From (3), we have

$$-\sum_{i=1}^{m} 2\left[y_i - (a + bx_i)\right] = 0$$

and $$-\sum_{i=1}^{m} 2x_i\left[y_i - (a + bx_i)\right] = 0$$

On simplification of these two equations, we have

$$\sum_{i=1}^{m} y_i = ma + b\sum_{i=1}^{m} x_i \qquad \left(\therefore \sum_{i=1}^{m} a = ma\right)$$

or $$\sum y_i = ma + b\sum x_i \qquad ...(4)$$

and $$\sum_{i=1}^{m} x_i y_i = a\sum_{i=1}^{m} x_i + b\sum_{i=1}^{m} x_i^2$$

or $$\sum x_i y_i = a\sum x_i + b\sum x_i^2 \qquad ...(5)$$

Equation (4) and (5) are said to be **normal equations**. Solving these two equations, we can determine the value of a and b. It can be easily verfied that for the obtained values of a and b, S will be minimum.

Example 7.1 Using the method of least-squares find a straight line that fits the following data :

x	71	68	73	69	67	65	66	67
y	69	72	70	70	68	67	68	64

Also find the value of y at $x = 68.5$.

Solution : Let the required straight line be

$$y = a + bx \qquad \text{...(i)}$$

the normal equations are

$$\sum y_i = ma + b\sum x_i \qquad \text{...(ii)}$$

and $$\sum x_i y_i = a\sum x_i + b\sum x_i^2 \qquad \text{...(iii)}$$

Now, to get the values of $\sum y_i$, $\sum x_i$, $\sum x_i y_i$ and $\sum x_i^2$, we construct following table :

i	x_i	y_i	$x_i y_i$	x_i^2
1	71	69	4899	5041
2	68	72	4896	4624
3	73	70	5110	5329
4	69	70	4830	4761
5	67	68	4556	4489
6	65	67	4355	4225
7	66	68	4488	4356
8	67	64	4288	4489
Sum	546	548	37422	37314

Hence, $\sum x_i = 546$, $\sum y_i = 548$

$$\sum x_i y_i = 37422, \ \sum x_i^2 = 37314$$

and total number of given data $m = 8$,

substituting these values in (ii) and (iii), we get

$$548 = 8a + 546b$$

$$37422 = 546a + 37314b.$$

Solving these two equations for a and b, we get

$a = 39.545484$ and $b = 0.424242$

Thus, the required straight line is

$$y = 39.545484 + 0.424242\,x$$

Now, at $x = 68.5$, value of y is given by

$$y = 39.545484 + 0.424242 \times 68.5$$

$$= 68.606061$$

Example 7.2 Fit a straight line to the given data

x	1	2	3	4	5	6
y	2.6	2.7	2.9	3.025	3.2	3.367

Also find value of y at $x = 5.5$.

Solution : Let the required straight line be

$$y = a + bx$$

then, the normal equation are

$$\sum y_i = ma + b\sum x_i$$

and $$\sum x_i y_i = a\sum x_i + b\sum x_i^2$$

Now, from the given data, we have following table :

i	x_i	y_i	$x_i y_i$	x_i^2
1	1	2.6	2.6	1
2	2	2.7	5.4	4
3	3	2.9	8.7	9
4	4	3.025	12.1	16
5	5	3.2	16	25
6	6	3.367	20.202	36
SUM	21	17.792	65.002	91

Hence,

$$\sum x_i = 21,\ \sum y_i = 17.792,$$

$$\sum x_i y_i = 65.002,\ \sum x_i^2 = 91$$

and $m = 6$

then, the normal equations become

$$17.792 = 6a + 21b$$

and $\quad 65.002 + 21a + 91b$,

Solving these equations, we get

$$a = 2.419333 \text{,} \; b = 0.156$$

Hence, required straight line is given by the equation

$$y = 2.419333 + 0.156x$$

Now, at $x = 5.5$, value of y is given by

$$y = 2.419333 + 0.156 \times 5.5$$

$$= 3.277333$$

Example 7.3 Fit a straight line to the given data

x	-1	0	1	2	3	4	5	6
y	10	9	7	5	4	3	0	-1

Also find the value of y at $x = 3.5$

Solution : Let the required straight line be

$$y = a + bx,$$

then the normal equations are given by

$$\sum y_i = ma + b\sum x_i$$

and $\quad \sum x_i y_i = a\sum x_i + b\sum x_i^2$

Now, we construct following table, using the given data :

i	x_i	y_i	$x_i y_i$	x_i^2
1	-1	10	-10	1
2	0	9	0	0
3	1	7	7	1
4	2	5	10	4
5	3	4	12	9
6	4	3	12	16
7	5	0	5	25
8	6	-1	-6	36
SUM	20	37	30	92

From the table we obtained following values,

$$\sum x_i = 20, \; \sum y_i = 37,$$

$$\sum x_i y_i = 30, \ \sum x_i^2 = 92$$

and $\quad m = 8$

from the normal equations, we have

$$37 = 8a + 20b$$

and $\quad 30 = 20a + 92b,$

solving above two equations, we get the values of a and b as

$$a = 8.345238, \ b = -1.488095$$

Thus, the equation of the straight line is

$$y = 8.345238 - 1.488095x$$

Now, at $x = 3.5$, we have

$$y = 8.345238 - 1.488095 \times 3.5$$

$$= 6.857143$$

Example 7.4 Fit a curve of the form $y = ax + bx^2$ to the given data :

x	1	1.5	2	2.5	3	3.5	4
y	1.1	1.95	3.2	5	8.1	11.9	16.4

Solution : Equation of the required curve is

$$y = ax + bx^2$$

which can be written as

$$\frac{y}{x} = a + bx \qquad \text{...(i)}$$

let $\frac{y}{x} = Y$, then the above equation becomes

$$Y = a + bx \qquad \text{...(ii)}$$

Normal equation for this curve are given by

$$\sum Y_i = ma + b\sum x_i$$

and $\quad \sum x_i Y_i = a\sum x_i + b\sum x_i^2$

Using given data, we construct following table :

i	x	y	$Y=\frac{y}{x}$	xY	x^2
1	1	1.1	1.1	1.1	1
2	1.5	1.95	1.3	1.95	2.25
3	2	3.2	1.6	3.2	4
4	2.5	5	2.0	5	6.25
5	3	8.1	2.7	8.1	9
6	3.5	11.9	3.4	11.9	12.25
7	4	16.4	4.1	16.4	16
SUM	17.5	-	16.2	47.65	50.75

From the table we have

$$\sum x_i = 17.5,\ \sum Y_i = 16.2$$

$$\sum x_i Y_i = 47.65,\ \sum x_i^2 = 50.75$$

and $m = 7$

Substituting these values in normal equations, we get

$$16.2 = 7a + 17.5b$$

$$47.65 = 17.5a + 50.75b \cdot$$

Solving these equations, we get

$$a = -0.239287,\ b = 1.021429$$

Thus, from (ii), we have

$$Y = -0.239287 + 1.021429x$$

Substituting $Y = \frac{y}{x}$, we have

$$y = -0.239287x + 1.021429x^2$$

which is the required equation.

7.4 Fitting a Polynomial of Degree n

We can fit a polynomial of degree n, to the given data using least-squares principle.

Let

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$$

be an polynomial of degree n which is to be fitted to the given data (x_i, y_i), $i = 1,2,\ldots\ldots,m$.

Then,

$$S = \sum_{i=1}^{m} \left(y_i - Y_i \right),$$

where

$$Y_i = Y\left(x_i\right)$$

$$= a_0 + a_1 x_i + a_2 x_i^2 + \ldots. + a_n x_i^n$$

so $$S = \sum_{i=1}^{m} \left[y_i - \left(a_0 + a_1 x_i + a_2 x_i^2 + \ldots. + a_n x_i^n \right) \right]$$

For S to be minimum, we must have

$$\frac{\partial S}{\partial a_0} = 0, \ \frac{\partial S}{\partial a_1} = 0, \ldots., \ \frac{\partial S}{\partial a_n} = 0,$$

that is,

$$\frac{\partial S}{\partial a_0} = \sum_{i=1}^{m} -2 \left[y_i - \left(a_0 + a_1 x_i + a_2 x_i^2 + \ldots. + a_n x_i^n \right) \right] = 0,$$

$$\frac{\partial S}{\partial a_1} = \sum_{i=1}^{m} -2 x_i \left[y_i - \left(a_0 + a_1 x_i + a_2 x_i^2 + \ldots. + a_n x_i^n \right) \right] = 0,$$

$$\frac{\partial S}{\partial a_2} = \sum_{i=1}^{m} -2 x_i^2 \left[y_i - \left(a_0 + a_1 x_i^2 + a_2 x_i^2 + \ldots. + a_n x_i^n \right) \right] = 0,$$

$$\ldots. \qquad\qquad \ldots. \qquad\qquad \ldots.$$

$$\frac{\partial S}{\partial a_n} = \sum_{i=1}^{m} -2 x_i^n \left[y_i - \left(a_0 + a_1 x_i^2 + a_2 x_i^2 + \ldots. + a_n x_i^n \right) \right] = 0.$$

Simplifying above $(n+1)$ equations, we get following normal equations,

$$\sum y_i = a_0 m + a_1 \sum x_i + a_2 \sum x_i^2 + \ldots. + a_n \sum x_i^n$$

$$\sum x_i y_i = a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i^3 + \ldots. + a_n \sum x_i^{n+1},$$

$$\sum x_i^2 y_i = a_0 \sum x_i^2 + a_1 \sum x_i^3 + a_2 \sum x_i^4 + \ldots. + a_n \sum x_i^{n+2},$$

$$\ldots. \qquad\qquad \ldots. \qquad\qquad \ldots.$$

$$\sum x_i^n y_i = a_0 \sum x_i^n + a_1 \sum x_i^{n+1} + a_2 \sum x_i^{n+2} + \ldots. + a_n \sum x_i^{2n} \qquad \ldots(6)$$

These $(n+1)$ equations can be solved for $(n+1)$ unknowns $a_0, a_1, a_2, \ldots\ldots a_n$.

Particular Case : For $n=3$, polynomial will be a parabola. Let equation of this parabola be

$$y = a + bx + cx^2$$

values of a, b and c can be obtained using normal equations given by

$$\sum y_i = ma + b\sum x_i + c\sum x_i^2 ,$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3$$

$$\sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 \qquad ...(7)$$

Example 7.5 Fit a second degree polynomial to th data :

x	– 4	– 3	– 2	– 1	0	1	2	3	4
y	– 5	– 1	0	1	3	4	4	3	2

Solution : Let the required equation of the curve be

$$y = a + bx + cx^2$$

Normal equations for this curve are

$$\sum y_i = ma + b\sum x_i + c\sum x_i^2$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3$$

$$\sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4$$

From the given data we construct following table :

i	x_i	y_i	$x_i y_i$	x_i^2	$x_i^2 y$	x_i^3	x_i^4
1	– 4	– 5	20	16	– 80	– 64	256
2	– 3	– 1	3	9	– 9	– 27	81
3	– 2	0	0	4	0	– 8	16
4	– 1	1	– 1	1	1	– 1	1
5	0	3	0	0	0	0	0
6	1	4	4	1	4	1	1
7	2	4	8	4	16	8	16
8	3	3	9	9	27	27	81
9	4	2	8	16	32	64	256
SUM	0	11	51	60	– 9	0	708

Thus,

$$\sum x_i = 0, \ \sum y_i = 11, \ \sum x_i y_i = 51$$

$\sum x_i^2 = 60$, $\sum x_i^2 y_i = -9$, $\sum x_i^3 = 0$,

$\sum x_i^4 = 708$ and $m = 9$,

Substituting above values in normal equations, we get

$$11 = 9a + b.0 + 60c \quad \text{or} \quad 11 = 9a + 60c,$$

$$51 = a.0 + b.60 + c.0 \quad \text{or} \quad 51 = 60b,$$

$$-9 = a.60 + b.0 + c.708 \quad \text{or} \quad -9 = 60a + 708c$$

Solving above equations for a, b and c, we get

$$a = 3.004329,\ b = 0.85,\ c = -0.267316$$

so, the required equation is given by

$$y = 3.004329 + 0.85x - 0.267316x^2$$

Example 7.6 Population of a city in different years are given in the following table :

x	1970	1980	1990	2000	2010
y (in thousands)	1450	1600	1850	2150	2500

Fit a parabola to the given data, using least squares principle. Also estimate the population of the city in 2005.

Solution : Since the magnitude of given data is large and values of x are given at equal intervals, therefore we reduce it by shift of origin and scale. Let $x_0 = 1990$ be origin of $x-$values and $y_0 = 1850$ be origin of $y-$values.

Then, let

$$X = \frac{x - 1990}{10} \quad \text{and} \quad Y = \frac{y - 1850}{50} \qquad \text{...(i)}$$

Let required curve be $y = a + bx + cx^2$, after change of origin and scale, it will be

$$Y = a + bX + cX^2 \qquad \text{...(ii)}$$

Now, we construct following table :

x	X	y	Y	XY	X^2	X^2Y	X^3	X^4
1970	–2	1450	–8	16	4	–32	–8	16
1980	–1	1600	–5	5	1	–5	–1	1
1990	0	1850	0	0	0	0	0	0
2000	1	2150	6	6	1	6	6	1
2010	2	2500	13	26	4	52	8	16
SUM	0	–	6	53	10	21	0	34

Normal equations, in new variables, will be

$$\sum Y_i = ma + b\sum X_i + c\sum X_i^2 ,$$

$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2 + c\sum X_i^3 ,$$

$$\sum X_i^2 Y_i = a\sum X_i^2 + b\sum X_i^3 + c\sum X_i^4$$

From the table, we have

$$\sum X_i = 0,\ \sum Y_i = 6,\ \sum X_i Y_i = 53$$

$$\sum X_i^2 = 10,\ \sum X_i^2\, Y_i = 21,\ \sum X_i^3 = 0,$$

$\sum X_i^4 = 34$, and $m = 5$,

substituting these values in normal equations, we have

$6 = 5a + b.0 + c.10$

or $6 = 5a + 10c$,

$53 = a.0 + b.10 + c.0$

or $53 = 10b$

and $21 = a.10 + b.0 + c.34$

or $21 = 10a + 34c$

Solving above equations for a, b and c, we get

$a = -0.085714$, $b = 5.3$, $c = 0.642857$

Now, from (ii), we have

$$Y = -0.085714 + 5.3X + 0.642857X^2 \qquad \text{...(iii)}$$

From (i), we have

$$\frac{y-1850}{50} = -0.085714 + 5.3\left(\frac{x-1990}{10}\right) + 0.642857\left(\frac{x-1990}{10}\right)^2$$

on simplification, we have

$$y = 1222008.286 - 1252.78543x + 0.3214285x^2$$

which is the required equation of parabola. Now, in the year 2005, population of the city will be given by

$$y = 1222008.286 - 1252.78543(2005) + 0.3214285(2005)^2$$

$= 2324.10456$

≈ 2324 thousands, approximately.

7.5 Fitting a Curve of the Form $y = ax^b$

Let, for the given data (x_i, y_i), $i = 1,2,.....,m$, we have to fit a curve of the form

$$y = ax^b \qquad ...(8)$$

taking logarithm of both sides, we get

$$\log_e y = \log_e a + b\log_e x \qquad ...(9)$$

Let $\log_e y = Y$, $\log_e x = X$, $\log_e a = A$, then equation (9) becomes

$$Y = A + bX$$

which is a straight line so the normal equations are given by

$$\sum Y_i = mA + b\sum X_i$$

and $$\sum X_iY_i = A + \sum X_i + b\sum X_i^2 \qquad ...(10)$$

where m is total number of given data.

Solving these equations, we can get values of A and b. Using the relation $\log_e a = A$, we can get value of a.

Example 7.7 Fit a curve of the form $y = ax^b$ to the given data :

x	2	3	4	5	6
y	144	172.8	207.4	248.8	298.5

Solution : The curve to be fitted is

$$y = ax^b \qquad ...(i)$$

Taking logarithm of both sides of equation (i), we get

$$\log_e y = \log_e a + b\log_e x$$

or $$Y = A + bX \qquad ...(ii)$$

where $Y = \log_e y$, $A = \log_e a$ and $\log_e x = X$

i	x	$X = \log_e x$	y	$Y = \log_e y$	XY	X^2
1	2	0.6932	144	4.9698	3.4451	0.4808
2	3	1.0986	172.8	5.1521	5.6601	1.2069
3	4	1.3863	207.4	5.3346	7.3954	1.9218
4	5	1.6094	248.8	5.5166	8.8784	2.5902
5	6	1.7918	298.5	5.6988	10.2111	3.2105
SUM	-	6.5793	-	26.6719	35.5901	9.4099

From the table, we obtained

$$\sum X_i = 6.5793,\ \sum Y_i = 26.6719,$$

$$\sum X_i Y_i = 35.5901,\ \sum X_i^2 = 9.4099,$$

and $m = 5$

Normal equations, for the curve (ii), are given by

$$\sum Y_i = mA + b\sum X_i$$

$$\sum X_i Y_i = A\sum X_i + b\sum X_i^2$$

Substituting values, obtained from the table, we get

$$26.6719 = 5A + 6.5793b,$$

$$35.5901 = 6.5793A + 9.4099b$$

Solving these equations, we get

$$A = 4.471176,\ b = 0.656$$

From the relation $\log_e a = A$, we get

$$a = 87.459515$$

Thus, the required curve is

$$y = 87.459515(x)^{0.656}$$

Example 7.8 Fit a curve $y = ax^b$ to the following data :

x	1	2	3	4
y	5	7	9	10

Also estimate the value of y at $x = 2.5$

Solution : For the required curve

$$y = ax^b$$

normal equations are given by

$$\sum Y_i = mA + b\sum X_i$$

$$\sum X_i Y_i = A\sum X_i + b\sum X_i^2$$

where, $Y = \log_e y$, $X = \log_e x$ and $A = \log_e a$

From the given data, we construct following table :

i	x	$X=\log_e x$	y	$Y=\log_e y$	XY	X^2
1	1	0	5	1.6094	0	0
2	2	0.6931	7	1.9459	1.3487	0.4804
3	3	1.0986	9	2.1972	2.4138	1.2069
4	4	1.3863	10	2.3026	3.1921	1.9218
SUM	-	3.1780	-	8.0551	6.9546	3.6091

From the table, we obtained

$$\sum X_i = 3.178\,,\ \sum Y_i = 8.0551\,,$$

$$\sum X_i Y_i = 6.9546\,,\ \sum X_i^2 = 3.6091$$

and $m = 4$

Substituting these values in normal equations, we get

$$8.0551 = 4A + 3.178b$$

$$6.9546 = 3.178A + 3.6091b$$

Solving these equations, we get

$$A = 1.607194\,,\ b = 0.511745$$

by the relation $\log_e a = A$, we get $a = 4.988793$

hence, the required equation is

$$y = 4.988793(x)^{0.511745}$$

Now, at $x = 2.5\,,\ y = 4.988793(2.5)^{0.511745}$

$$= 7.973322$$

7.6 Fitting a Curve of the Form $y = ae^{bx}$

Let the curve of the form

$$y = ae^{bx} \qquad ...(11)$$

to be fitted to the given data (x_i, y_i), $i = 1,2,.......m$.

Taking logarithm both sides of the above equation, we get

$$\log_e y = \log_e a + bx$$

or $Y = A + bx$ ...(12)

where $\log_e y = Y$ and $\log_e a = A$

Normal equations, for the equation (12), are given by

$$\sum Y_i = mA + b\sum x_i ,$$

$$\sum x_i Y_i = A\sum x_i + b\sum x_i^2$$

solving these equations, we can get required values of unknowns.

Example 7.9 Fit a exponential curve of the form $y = ae^{bx}$ to the given data :

x	1	2	3	4	5	6
y	1.6	4.5	13.8	40.2	125	300

Also find the value of y at $x = 4.5$

Solution : The required curve is

$$y = ae^{bx}$$

Normal equations are given by

$$\sum Y_i = mA + b\sum x_i$$

and $$\sum x_i Y_i = A\sum x_i + b\sum x_i^2$$

where, $Y = \log_e y$ and $A = \log_e a$

From the given data, we construct the table as follows :

i	x	y	$Y = \log_e y$	xy	x^2
1	1	1.6	0.4700	0.4700	1
2	2	4.5	1.5041	3.0082	4
3	3	13.8	2.6247	7.8741	9
4	4	40.2	3.6939	14.7756	16
5	5	125	4.8283	24.1415	25
6	6	300	5.7038	34.2228	36
SUM	21	-	18.8248	84.4922	91

From the table, we obtain following values

$$\sum x_i = 21, \sum Y_i = 18.8248,$$

$$\sum x_i Y_i = 84.4922, \sum x_i^2 = 91$$

and $m = 6$

substituting these values in above normal equations, we get

$$18.8248 = 6A + 21b$$

and $$84.4922 = 21A + 91b$$

solving these equations, we get

$$A = -0.583614 , \ b = 1.063166$$

From the relation $\log_e a = A$, we get

$$a = 0.557879$$

Thus, the required curve is

$$y = (0.557879)\ e^{(1.063166)x}$$

Now, at $x = 4.5$, value of y is given by

$$y = (0.557879)\ e^{(1.063166)\,(4.5)}$$

$$= 66.728611$$

7.7 Least-Squares Principle for Continuous Functions

We have studied least-squres principle for discrete data. A continuous function $f(x)$ can also be approximated by a polynomial of degree n on the given interval $[a,b]$, using least-squres principle.

Let desired polynomial of degree n be

$$P_n(x) = a_0 + a_1 x + a_2 x^2 + ...+ a_n x^n$$

where $a_0, a_1, a_2, ..., a_n$ are arbitrary parameters to be determined.

Let

$$S = \int_a^b w(x)\,[f(x) - P_n(x_i)]^2\,dx \qquad ...(13)$$

where, $w(x)$ is the weight function such that $w(x) > 0$ and S is a function of $a_0, a_1, a_2, ..., a_n$.

Also, $P_n(x_i) = a_0 + a_1 x_i + a_2 x_i^2 + ...+ a_n x_i^n$

Now, according to least-squares principle criterion, S must be minimum. The necessary conditions for S to be minimum, are given by

$$\frac{\partial S}{\partial a_0} = -2\int_a^b w(x)\left[f(x) - \sum_{r=0}^{n} a_r x^r\right]dx = 0,$$

$$\frac{\partial S}{\partial a_1} = -2\int_a^b w(x)\left[f(x) - \sum_{r=0}^{n} a_r x^r\right]x\,dx = 0,$$

$$\frac{\partial S}{\partial a_2} = -2\int_a^b w(x)\left[f(x) - \sum_{r=0}^{n} a_r x^r\right] x^2 dx = 0,$$

....

$$\frac{\partial S}{\partial a_n} = -2\int_a^b w(x)\left[f(x) - \sum_{r=0}^{n} a_r x^r\right] x^n dx = 0$$

Simplifying above equations, we get

$$a_0 \int_a^b w(x)dx + a_1 \int_a^b x\, w(x)dx + a_2 \int_a^b x^2\, w(x)dx + ...$$

$$+ a_n \int_a^b x^n\, w(x)dx = \int_a^b w(x) f(x)dx$$

$$a_0 \int_a^b x\, w(x)dx + a_1 \int_a^b x^2\, w(x)dx + a_2 \int_a^b x^3\, w(x)dx + ...$$

$$+ a_n \int_a^b x^{n+1}\, w(x)dx = \int_a^b x\, w(x) f(x)dx$$

....

$$a_0 \int_a^b x^n\, w(x)dx + a_1 \int_a^b x^{n+1}\, w(x)dx + a_2 \int_a^b x^{n+2}\, w(x)dx + ...$$

$$+ a_n \int_a^b x^{2n}\, w(x)dx = \int_a^b x^n\, w(x) f(x)dx$$

This is a system of $(n+1)$ equations, called normal equations and this system can be solved for $(n+1)$ unknown $a_0, a_1, a_2, ..., a_n$.

Example 7.10 Obtain a second degree polynomial approximation to the function $f(x) = x^3$, on the interval $[0,1]$, using least-squares principle. Take weight function $w(x) = 1$.

Solution : Let, the required polynomial be

$$y = a + bx + cx^2$$

Then, we have

$$S(a,b,c) = \int_0^1 \left[x^3 - \left(a + bx + cx^2\right)\right]^2 dx$$

Normal-equations are given by

$$\frac{\partial S}{\partial a} = -2\int_0^1 \left[x^3 - \left(a + bx + cx^2\right)\right] dx = 0,$$

$$\frac{\partial S}{\partial b} = -2\int_0^1 \left[x^3 - \left(a + bx + cx^2\right)\right] x\, dx = 0,$$

$$\frac{\partial S}{\partial c} = -2\int_0^1 \left[x^3 - \left(a + bx + cx^2\right)\right] x^2\, dx = 0$$

Simplifying above equations, we get

$$a\int_0^1 dx + b\int_0^1 x\, dx + c\int_0^1 x^2 dx = \int_0^1 x^3 dx\ ,$$

$$a\int_0^1 x\, dx + b\int_0^1\ x^2 dx + c\int_0^1 x^3 dx = \int_0^1 x^4 dx\ ,$$

$$a\int_0^1\ x^2\ dx + b\int_0^1\ x^3 dx + c\int_0^1 x^4 dx = \int_0^1 x^5 dx$$

Simplifying these equations, we get

$$a + \frac{b}{2} + \frac{c}{3} = \frac{1}{4},$$

$$\frac{a}{2} + \frac{b}{3} + \frac{c}{4} = \frac{1}{5},$$

$$\frac{a}{3} + \frac{b}{4} + \frac{c}{5} = \frac{1}{5},$$

solving these equations, we get

$$a = \frac{1}{20},\ b = -\frac{6}{10},\ c = \frac{15}{10}$$

Thus, the required polynomial is given by

$$f(x) = y = \frac{1}{20} - \frac{6}{10}x + \frac{15}{10}x^2$$

or $\qquad x^3 = \frac{1}{20}\left(1 - 12x + 30x^2\right).$

As a check, at $x = \frac{1}{2}$

$$y = \frac{1}{20}\left[1 - 12 \times \frac{1}{2} + 30 \times \frac{1}{4}\right]$$

$$= \frac{1}{20}\left[1 - 6 + 7.5\right]$$

$$= 0.125$$

The true value is given by

$$f(x) = x^3$$

which gives, $f\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^3$

$$= \frac{1}{8}$$

$$= 0.125$$

Thus, error at $x = \frac{1}{2}$ is zero.

Self-Learning Exercise

1. In least-squares principle sum of errors e_i^n is minimised. Here value of n is

(a) 1 (b) 2 (c) 3 (d) $1/2$

2. Using least-squares principle, we can approximate a polynomial of n degree, fit to discrete data. (Ture/False)

3. To fit a parabola, unknown parameters can be obtained by solving normal-equations consistings of

(a) two equatons (b) three equations

(c) four equations (d) five equations

7.8 Summary

In this unit we studied least-squares principle. This principle provides us a technique to approximate a curve or polynomial of best fit to the given descrete data or a given continuous function.

7.9 Answers of Self-Learning Exercise

1. (b) 2. True 3. (b)

7.10 Exercises

1. The temperatures θ and length l of a heated rod are given below. Establish a relation between θ and l of the form $l = a + b\theta$ using least-squares principle.

$\theta(^0c)$	20	30	40	50	60	70
$l(mm)$	800.3	800.4	800.6	800.7	800.9	800.10

[Ans. $a = 800$, $b = 0.0146$ **]**

2. Fit a curve of the form $y = ax + bx^2$ to the given data :

x	1	2	3	4	5	6
y	2.6	5.4	8.7	12.1	16	20.2

[Ans. $y = 2.41973x + 0.15589x^2$]

3. Fit a straight line to the following data :

x	1	2	3	4	5	8
y	2.4	3	3.6	4	5	6

Also find y at $x = 3.5$.

[Ans. $y = 1.976 + 0.506x$, $y = 3.747$ at $x = 3.5$]

4. Compute the constants α and γ^{β} such that the curve $y = \alpha\gamma^{\beta x}$ fits the given data :

x	1	2	3	4	5	6
y	151	100	61	50	20	8

[Ans. $\alpha = 309$, $\gamma^{\beta} = 0.5754$]

5. Fit a curve of the form $y = ax^b$ to the data given below :

x	2	4	7	10	20	40	60	80
y	43	25	18	13	8	5	3	2

[Ans. $a = 4.36$, $b = -0.7975$]

6. Fit the curve $pV^r = k$ to the data given in the table :

p	0.5	1	1.5	2	2.5	3
V	1.62	1	0.75	0.62	0.52	0.46

[Ans. $r = 1.4224$, $k = 0.9970$]

7. Fit a curve $y = ae^{bx}$ to the following data :

x	2	4	6	8	10
y	4.077	11.084	30.128	81.897	222.62

Also estimate y at $x = 7$.

[Ans. $a = 1.499$, $b = 0.5$, $c = 49.6401$]

8. Fit a second degree polynomial to the following data, taking x as independent variable :

x	1	1.5	2	2.5	3	3.5	4
y	1.1	1.3	1.6	2.0	2.7	3.4	4.1

[Ans. $y = 1.0368 - 0.1932x + 0.2429x^2$]

9. Fit a second degree parabola to the given data

x	1929	1930	1931	1932	1933	1934	1935	1936	1937
y	352	356	357	358	360	361	361	360	359

[Ans. $y = -1010135 + 1044.67x - 0.27x^2$]

10. Obtain a least-squares quadratic approximation to the function $y(x) = \sqrt{x}$ on $[0,1]$, with respect to the weight function $w(x) = 1$.

[Ans. $y = \frac{1}{35}\left(6 + 48x - 20x^2\right)$]

11. Construct a least-squares quadratic approximation to the function $y(x) = \sin x$ on $[0, \pi/2]$, with respect to weight function $w(x) = 1$.

[Ans. $y \sin x = -0.03511 + 1.235x - 0.362x^2$]

□□□

Unit - 8 : Approximation of Functions by Taylor Series and Chebyshev Polynomials

Structure of the Unit

8.0 Objectives

In this unit we shall study about chebyshev polynomials and their application in approximation of functions and economization of power series. We shall also study about approximation of a function using Taylor series.

8.1 Introduction

We have already studied function approximation through interpolation. In previous unit, we studied curve fitting based on **least-squares principle**. This technique is also used to approximate the function. In least-squares principle, we minimize $S = \sum e_i^2$, where e_i is difference between observed and expected value of y at $x = x_i$. If we minimize maximum component of errors e_i, we get a technique to approximate the function using chebyshev polynomials. Chebyshev polynomials are orthogonal polynomials.

We are quite familier with Taylor series, which is also an important tool to approximate the function.

8.2 Taylor Series Expansion of a Function

A Taylor series expansion for a function $f(x)$ about a point x_0 is given by

$$f(x) = f(x_0) + (x - x_0) f'(x_0) + \frac{1}{2!}(x - x_0)^2 f''(x_0) + \ldots$$

$$+ \frac{1}{n!}(x - x_0)^n f^{(n)}(x_0) + \frac{1}{(n+1)!}(x - x_0)^{n+1} f^{(n+1)}(x_0) + \ldots \qquad \ldots(1)$$

If $P_n(x)$ is the polynomial of degree n, approximating the given function, obtained by truncating the terms containing $(n+1)th$ and higher order of x in (i), then

$$P_n(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2!}(x - x_0)^2 f''(x_0) + ...$$

$$+\frac{1}{n!}(x - x_0)^n f^{(n)}(x_0) \qquad ...(2)$$

where

$$P_n^{(k)}(x_0) = f^{(k)}(x_0), \quad k = 0,1,2,......n \qquad ...(3)$$

and remainder term

$$R_n = \frac{1}{(n+1)!}(x - x_0)^{n+1} f^{(n+1)}(\theta), \quad x_0 < \theta < x \qquad ...(4)$$

is the truncation term.

If the truncation error is $\in > 0$, then we have

$$\frac{1}{(n+1)!}\left|x - x_0\right|^{n+1}\left|f^{(n+1)}(x)\right| \le \in \qquad ...(5)$$

Using this relation we can find number of terms to be retained in (2), for the given **error tolerance** $\in$.

Example 8.1 Obtain Taylor series expansion of the function $f(x) = e^x$, about $x = 0$. Find the number of terms of the exponential series such that their sum gives the value of e^x correct to six decimal places at $x = 1$.

Solution : Given that $f(x) = e^x$

then $f'(x) = e^x = f''(x) = ... = f^{(n)}(x)$ and so on,

Now, at $x = 0$, $f(0) = 1 = f'(0) = ...f^{(n)}(0)$ and so on.

Thus, from (2), we get

$$P_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ... + \frac{x^n}{n!}$$

Here, $R_n = \frac{x^{n+1}}{(x+1)!}e^\theta, \quad 0 < \theta < x$

For, accuracy of six decimal places at $x = 1$, we have [from (5)]

$$\frac{1}{(n+1)!} < \frac{1}{2} \times 10^{-6} \qquad (\because \text{ at } x = 1,\ f^{(n+1)}(x) = 1)$$

that is $(n+1)! > 2 \times 10^{6}$

which gives $n = 9$, so total no. of terms $= n + 1 = 10$

that is, we need 10 terms for required accuracy.

Example 8.2 Obtain polynomial approximation $P_n(x)$ for the function $f(x) = e^{-x}$ using Taylor series expansion about $x_0 = 0$ and find the value of x when the error in $P_n(x)$ obtained from the first five terms only is to be less than 10^{-7} after rounding.

Solution : Given that $f(x) = e^{-x}$

then, $f'(x) = -e^{-x}$,

$f''(x) = e^{-x}$,

$f'''(x) = -e^{-x}$, and so on

Now, at $x = 0$,

$f(0) = 1$, $f'(0) = -1$, $f''(0) = 1$ and so on.

Thus,

$$f^{(r)}(0) = (-1)^{r},\ r = 0,1,2,.......$$

So, Taylor series expansion is given by

$$P_n(x) = 1 - x + \frac{x^2}{2} - \frac{x^3}{6} + \frac{x^4}{24} - \frac{x^5}{120} + ...$$

Now, from (5), we have

$$\frac{1}{(n+1)!}\left|(x-0)^{n+1}\right|\left|f^{(n+1)}(x)\right| < \frac{1}{2} \times 10^{-7}$$

If we retain first five terms, then

$$\frac{1}{5!} x^5 \max_{0 \le x \le 1}\left|e^{-x}\right| < 5 \times 10^{-8}$$

or $\quad x^5 < 120 \times 5 \times 10^{-8} \qquad \left(\because \max_{0 \le x \le 1}\left|e^{-x}\right| = 1\right)$

which gives

$$x^5 < 6 \times 10^{-6}$$

or $\quad x < 0.0903$

Example 8.3 Obtain a second degree polynomial approximation to the function

$$f(x)=\frac{1}{1+x^2}, \quad x\in[1,\ 1.2]$$

using Taylor series expansion about $x=1$. Find a bound on the truncation error.

Solution : Given that, $f(x)=\dfrac{1}{1+x^2}$

The Taylor series expansion is given by

$$f(x)=f(1)+(x-1)f'(1)+\frac{1}{2!}(x-1)^2 f''(1)+\frac{1}{3!}(x-1)^3 f'''(\theta),\ 1\le\theta\le x$$

Substituting the values of derivatives of $f(x)$ at $x=1$, we get

$$P_2(x)=\frac{1}{2}-\frac{1}{2}(x-1)+\frac{1}{4}(x-1)^2$$

Truncation errors is given by

$$\frac{(x-1)^3}{6}f'''(\theta), \quad 1\le\theta\le x$$

Error bound is given by

$$|R_2|\le\frac{(x-1)^3}{6}\max_{1\le x\le 1.2}|f'''(x)|$$

$$\le\frac{(x-1)^3}{6}(0.3575076)$$

$$\le 0.0595846\,(x-1)^3$$

Maximum absolute truncation error at $x=1.2$ is given by 0.0004767.

8.3 Orthogonal Polynomials and Least-Squares Approximations

A set of functions $\{\phi_i(x)\}$ is said to be orthogonal on an interval $[a,\ b]$, with respect to the weight function $w(x)$, if

$$\int_a^b w(x)\ \phi_j(x)\phi_k(x)=0, \text{ if } j\ne k \qquad ...(6)$$

For descrete data, a set of function $\{\phi_i(x)\}$ is said to be **orthogonal** over a set of points $\{x_i\}$, with respect to the weight function $w(x)$, if

$$\sum_{i=1}^{m} w(x_i)\phi_j(x_i)\phi_k(x_i)=0, \text{ if } j\ne k \qquad ...(7)$$

In approximating the function, using least-squares principle, we obtain a system of linear equation, which may possess problem of **ill-conditioning**. This problem can be avoided by the use of orthogonal functions or polynomials. One more advantage is that we can determine parameters directly.

Let, the approximation to the given function $y = f(x)$, is of the form

$$Y(x) = a_0\phi_0(x) + a_1\phi_1(x) + ... + a_n\phi_n(x) \qquad ...(8)$$

where, the set of polynomials $\{\phi_j(x)\}$ are orthogonal on an interval $[a, b]$ and $\phi_j(x)$ is a polynomial of degree j in variable x.

Now, by the weighted least-squares approximation criterion.

$$S(a_0, a_1, ..., a_n) = \int_a^b w(x)\left[f(x) - \sum_{j=0}^{n} a_j\phi_j(x)\right]^2 dx \qquad ...(9)$$

should be minimum. The necessary conditions for S to be minimum, are given by

$$\frac{\partial S}{\partial a_i} = 0,\ i = 0,1,2,......., n$$

or $$\int_a^b w(x)\left[f(x) - \sum_{j=0}^{n} a_j\phi_j(x)\right]\phi_k(x)\,dx = 0 \qquad ...(10)$$

where $k = 0,1,2,...., n$

These are $(n+1)$ equations in $(n+1)$ unknowns and can be solved, using the property of orthogonal functions.

Thus, equation (10) gives

$$a_j \int_a^b w(x)\phi_j^2(x)dx = \int_a^b w(x) f(x)\phi_j(x)dx$$

hence, $$a_j = \frac{\int_a^b w(x) f(x)\phi_j(x)dx}{\int_a^b w(x)\phi_j^2(x)dx} \qquad ...(11)$$

8.4 Gram-Schmidt Orthogonalizing Process

Every set of linearly independent polynomials is not orthogonal, but it can be orthogonalized using Gram-Schmidt process.

Let, the orthogonal monic polynomial $\phi_j(x)$ be of degree j then it has leading term x^j. Thus,

$$\phi_0(x) = 1 \qquad ...(12)$$

Let

$$\phi_1(x) = x + a_{10}\phi_0(x) \qquad ...(13)$$

such that $\phi_0(x)$ and $\phi_1(x)$ are orthogonal and $\phi_1(x)$ is of degree one therefore it has leading term x.

By the condition (6) of orthogonality, we have

$$\int_a^b w(x)\,\phi_0(x)\,\phi_1(x)\,dx = 0$$

or $$\int_a^b x\,w(x)\,\phi_0(x)\,dx + a_{10}\int_a^b w(x)\,\phi_0^2(x)\,dx = 0$$

which gives

$$a_{10} = -\frac{\int_a^b x\,w(x)\,\phi_0(x)\,dx}{\int_a^b w(x)\,\phi_0^2(x)\,dx}$$

$$= -\frac{\int_a^b x\,w(x)\,dx}{\int_a^b w(x)\,dx} \qquad (\because\ \phi_0(x) = 1) \qquad ...(14)$$

Using the value of a_{10} in (13) we can get the value of $\phi_1(x)$. Now, let

$$\phi_2(x) = x^2 + a_{20}\,\phi_0 + a_{21}\,\phi_1(x) \qquad ...(15)$$

where $\phi_2(x)$ is a polynomial orthogonal to both $\phi_0(x)$ and $\phi_1(x)$ having x^2 as leading term. Now, to determine a_{20} and a_{21}, we use the condition of orthogonality as follows :

$$\int_a^b w(x)\,\phi_0(x)\,\phi_2(x)\,dx = 0 \qquad (\because\ \phi_0(x) \text{ and } \phi_2(x) \text{ are orthogonal})$$

and $$\int_a^b w(x)\,\phi_1(x)\,\phi_2(x)\,dx = 0 \qquad (\because \phi_1(x) \text{ and } \phi_2(x) \text{ are orthogonal})$$

using (15), we get

$$\int_a^b w(x)\,\phi_0(x)\left[x^2 + a_{20}\,\phi_0(x) + a_{21}\,\phi_1(x)\right]dx = 0$$

and $$\int_a^b w(x)\,\phi_1(x)\left[x^2 + a_{20}\,\phi_0(x) + a_{21}\,\phi_1(x)\right]dx = 0$$

Solving these equations, we get

$$a_{20} = -\frac{\int_a^b x^2\,w(x)\,dx}{\int_a^b w(x)\,dx}$$

$$a_{21} = -\frac{\int_a^b x^2\,w(x)\,\phi_1(x)\,dx}{\int_a^b w(x)\,\phi_1^2(x)\,dx} \qquad ...(16)$$

using values of a_{20} and a_{21} in equation (15), we get the value of $\phi_2(x)$.

Proceeding in similar manner, we get

$$\phi_j(x) = x^j + b_{j0}\,\phi_0(x) + b_{j1}\,\phi_1(x) + ... + b_{j,j-1}\,\phi_{j-1}(x) \qquad ...(17)$$

where b_{ji} is given by

$$b_{ji} = -\frac{\int_a^b x^j\, w(x)\,\phi_i(x)\,dx}{\int_a^b w(x)\,\phi_i^2(x)\,dx}, \qquad \text{where } i = 0,1,2,....,j-1 \qquad ...(18)$$

Particular Cases : 1. If take weight function $w(x)=1$ and interval $[-1,1]$, then we get

$$\phi_0(x) = 1 = P_0(x),$$

$$\phi_1(x) = x = P_1(x),$$

$$\phi_2(x) = \frac{1}{3}(3x^2-1) = \frac{2}{3}P_2(x),$$

$$\phi_3(x) = \frac{1}{5}\left[5x^3-3x\right] = \frac{2}{5}P_3(x) \text{ and so on.}$$

where $P_0(x), P_1(x), P_2(x), P_3(x), ...$ are the **Legendre Polynomials,** orthogonal with respect to weight function $w(x)=1$ on the interval $[-1,1]$.

2. If we take weight function $w(x) = (1-x^2)^{-\frac{1}{2}}$ on the interval $[-1, 1]$, then we obtain

$$\phi_0(x) = 1 = T_0(x),$$

$$\phi_1(x) = x = T_1(x),$$

$$\phi_2(x) = x^2 - \frac{1}{2} = \frac{1}{2}(2x^2-1) = \frac{1}{2}T_2(x) \text{ and so on,}$$

here $T_0, T_1, T_2, ...$ are the chebyshev polynomials which are also orthgonal polynomials.

8.5 Chebyshev Polynomials and its Properties

The chebyshev polynomials of the first kind of degree n over the interval $[-1, 1]$ is defined by

$$T_n(x) = \cos(n\cos^{-1}x) \qquad ...(19)$$

Let $\cos^{-1}x = \theta$, then we have

$$T_n(x) = \cos n\,\theta$$

Hence, for $n=0$, $T_0(x)=1$,

for $\quad n=1$, $T_1(x)=\cos\theta = x$

for $\quad n=2$, $T_2(x)=\cos 2\theta$

$$=2\cos^2\theta - 1$$

$$=2x^2-1, \text{ and so on}$$

From (19), it is obvious that

$$T_n(x)=T_{-n}(x) \qquad \text{...(20)}$$

Recurrence Relation :

By the definition of $T_n(x)$, we have

$$T_{n+1}(x)=\cos(n+1)\theta \text{ and } T_{n-1}(x)=\cos(n-1)\theta,$$

adding these, we get

$$T_{n+1}(x)+T_{n-1}(x)=\cos(n+1)\theta+\cos(n-1)\theta$$

$$=2\cos n\theta.\cos\theta$$

$$=2.T_n(x).x$$

$$=2x\,T_n(x)$$

Thus, we have

$$T_{n+1}(x)=2x\,T_n(x)-T_{n-1}(x) \qquad \text{...(21)}$$

Using this relation we can find successively all $T_n(x)$. Some of these are given below :

$$T_0(x)=1,$$

$$T_1(x)=x,$$

$$T_2(x)=2x^2-1,$$

$$T_3(x)=4x^3-3x,$$

$$T_4(x)=8x^4-8x^2+1,$$

$$T_5(x)=16x^5-20x^3+5x,$$

$$T_6(x)=32x^6-48x^4+18x^2-1,$$

$$T_7(x)=64x^7-112x^5+56x^3-7x,$$

$$T_8(x) = 128x^8 - 256x^6 + 160x^4 - 32x^2 + 1, \text{ and so on.} \qquad ...(22)$$

It is to be noted that coefficient of leading term in $T_n(x)$ is always 2^{n-1}.

Expansion for Power of x terms of Chebyshev Polynomials :

$$x^0 = 1 = T_0(x),$$

$$x = T_1(x),$$

$$x^2 = \frac{1}{2}\left[T_0(x) + T_2(x)\right],$$

$$x^3 = \frac{1}{4}\left[3T_1(x) + T_3(x)\right],$$

$$x^4 = \frac{1}{8}\left[3T_0(x) + 4T_2(x) + T_4(x)\right],$$

$$x^5 = \frac{1}{16}\left[10T_1(x) + 5T_3(x) + T_5(x)\right],$$

$$x^6 = \frac{1}{32}\left[10T_0(x) + 15T_2(x) + 6T_4(x) + T_6(x)\right], \text{ and so on.} \qquad ...(23)$$

Orthogonal Properties of Chebyshev Polynomials :

$$\int_{-1}^{1} \frac{T_m(x)T_n(x)}{\sqrt{1-x^2}}\,dx = \begin{cases} 0, & m \neq n \\ \frac{\pi}{2}, & m = n \neq 0 \\ \pi, & m = n = 0 \end{cases} \qquad ...(24)$$

Minimax Property : An important property of chebyshev polynomials, called minimax property, is that, of all polynomials of degree n where the coefficient of leading term x^n is unity, the polynomial $2^{1-n}T_n(x)$ has the smallest least upper bound for its absolute value in the interval $[-1,1]$, that is,

$$\max_{-1 \le x \le 1} \left|2^{1-n}T_n(x)\right| \le \max_{-1 \le x \le 1} \left|P_n(x)\right| \qquad ...(25)$$

Here $P_n(x)$ is any monic polynomial of degree n.

8.6 Chebyshev Approximation (Uniform-Minimax Polynomial Approximation)

Chebyshev polynomials are very useful in **minimax approximations** as these polynomials have minimax property. In chebyshev approximation, the maximum error is kept down to minimal. This is called **minimax principle** and polynomial $T_n(x)$ is referred to as **minimax polynomial**. This process is used for **lower-order** approximation and called **minimax approximation.**

Let a function $f(x)$ is approximated by the polynomial

$$P_n(x) = a_0 + a_1 x + a_2 x^2 + \ldots\ldots + a_n x^n$$

where $f(x)$ is continuous on the interval $[a,b]$, then the minimax polynomial approximation, we shall determine constants $a_0, a_1, a_2, \ldots\ldots, a_n$, such that

$$\max_{a \le x \le b} |\varepsilon(x)| = \min_{a \le x \le b} |\varepsilon(x)|, \qquad \ldots(26)$$

where error $\varepsilon(x)$ is given by

$$\varepsilon(x) = f(x) - P_n(x)$$

If $P_n(x)$ is the best uniform approximation, following (26), and if

$$E_n = \max_{a \le x \le b} |\varepsilon(x)|$$

then there are atleast $(n+2)$ points $a = x_0 < x_1 < x_2 < \ldots\ldots < x_n < x_{n+1} = b$ such that

(i) error at these points must alternate in sings

(ii) $\varepsilon(x_i) = \pm E_n,\ i = 0,1,2,\ldots\ldots,n+1$

(iii) $\varepsilon(x_i) = -\varepsilon(x_{i+1})$ for $i = 0,1,2,\ldots\ldots,n$

(iv) $\varepsilon'(x_i) = 0,\ i = 1,2,\ldots\ldots\ldots,n$ ...(27)

The best uniform approximation can be found using (27).

8.7 Chebyshev Series Expansion

A function $f(x)$ can be expanded in a series of chebyshev polynomials as

$$f(x) = \frac{1}{2}c_0 + \sum_{j=1}^{\infty} c_j T_j(x) \qquad \ldots(28)$$

where $f(x)$ is continous on $[-1,1]$.

The partial sum of (23) is given by

$$P_n(x) = \frac{1}{2}c_0 + \sum_{j=1}^{\infty} c_j T_j(x) \qquad \ldots(29)$$

This is the truncated chebyshev series expansion for the function $f(x)$, which is nearly the best uniform approximation to $f(x)$.

Coefficient c_j in the above expansion can be obtained using the orthogonal property of chebyshev polynomials and it is given by

$$c_j = \frac{2}{\pi}\int_{-1}^{1}\frac{f(x)T_j(x)}{\sqrt{1-x^2}}dx, \quad j = 0,1,2,.....n \qquad ...(30)$$

First we express $f(x)$ in a power series in x then using relation (23), it is expressed in terms of $T_n(x)$.

8.8 Economization of the Power Series

In section 8.6, we studied the process of expan a function in a series of chebyshev polynomials. Let, the truncated chebyshev polynomial expansion for a given function $f(x)$ be given by

$$P_n(x) = \sum_{i=0}^{n} a_i T_i(x) \qquad ...(31)$$

then,

$$\max_{-1\le x\le 1}|f(x) - P_n(x)| \le |a_{n+1}| + |a_{n+2}| + ... \le \in \qquad (\because |T_n(x)| \le 1) \qquad ...(32)$$

where $\in$ is error tolerance. Number of terms retained in (31) depends on the given error tolerance $\in$. Replacing each $T_i(x)$ by its polynomials form using (23), we get economized polynomial approximation. This process is known as **Lanczos Economization.**

Example 8.4 Express $2 - x^2 + 3x^4$ as a sum of chebyshev polynomials.

Solution : We know that,

$$1 = T_0(x),$$

$$x^2 = \frac{1}{2}[T_0(x) + T_2(x)],$$

and $$x^4 = \frac{1}{8}[3T_0(x) + 4T_2(x) + T_4(x)]$$

then

$$2 - x^2 + 3x^4 = 2.T_0(x) - \frac{1}{2}[T_0(x) + T_2(x)] + \frac{3}{8}[T_0(x) + 4T_2(x) + T_4(x)]$$

$$= \left(2 - \frac{1}{2} + \frac{9}{8}\right)T_0(x) + \left(-\frac{1}{2} + \frac{3}{2}\right)T_2(x) = \frac{3}{8}T_4(x)$$

$$= \frac{21}{8}T_0(x) + T_2(x) + \frac{3}{8}T_4(x)$$

$$= \frac{1}{8}[21\,T_0(x) + 8\,T_2(x) + 3\,T_4(x)]$$

Example 8.5 Express $2\,T_0(x)+T_1(x)+2\,T_2(x)$ as a polynomial in x.

Solution : We know that,

$$T_0(x)=1\,,$$

$$T_1(x)=x$$

and $\quad T_2(x)=2x^2-1\,,$

then,

$$2\,T_0(x)+T_1(x)+2\,T_2(x)=2.1+x+2\left(2x^2-1\right)$$

$$=2+x+4x^2-4$$

$$=4x^2+x-2$$

Example 8.6 Find the best lower order approximation to the polynomial $2x^3+5x^2$.

Solution : We know that

$$x^3=\frac{1}{4}\left[3T_1(x)+T_3(x)\right]$$

Therefore,

$$2x^3+5x^2=\frac{2}{4}\left[3\,T_1(x)+T_3\,(x)\right]+5x^2$$

$$=5x^2+\frac{3}{2}T_1(x)+\frac{1}{2}T_3(x)$$

$$=5x^2+\frac{3}{2}x+\frac{1}{2}T_3(x) \qquad \left[\because\ T_1(x)=x\right]$$

Thus, the lower order approximation for the given polynomial is given by

$$5x^2+\frac{3}{2}x \qquad \text{or} \qquad \frac{1}{2}\left[10x^2+3x\right]$$

Example 8.7 Using the chebyshev polynomials, obtain the least-squares approximations of second degree for the function $f(x)=x^3+x^2+3$, where $x\in[-1,1]$.

Solution : Let the second degree approximation is given by

$$P_2(x)=c_0T_0(x)+c_1T_1(x)+c_2T_2(x) \qquad \text{...(i)}$$

then

$$S(c_0,c_1,c_2)=\int_{-1}^{1}\frac{1}{\sqrt{1-x^2}}\left[f(x)-P_2(x)\right]dx \qquad \text{...(ii)}$$

For S to be minimum, the necessary conditions are

$$\frac{\partial S}{\partial c_0} = 0 = \frac{\partial S}{\partial c_1} = \frac{\partial S}{\partial c_2} \qquad \text{...(iii)}$$

Using (1), (2) and (3), we get normal equations as follows :

$$\int_{-1}^{1} \frac{T_0(x)}{\sqrt{1-x^2}} \left[x^3 + x^2 + 3 - c_0 T_0(x) - c_1 T_1(x) - c_2 T_2(x)\right] dx = 0$$

$$\int_{-1}^{1} \frac{T_1(x)}{\sqrt{1-x^2}} \left[x^3 + x^2 + 3 - c_0 T_0(x) - c_1 T_1(x) - c_2 T_2(x)\right] dx = 0$$

$$\int_{-1}^{1} \frac{T_2(x)}{\sqrt{1-x^2}} \left[x^3 + x^2 + 3 - c_0 T_0(x) - c_1 T_1(x) - c_2 T_2(x)\right] dx = 0$$

Using orthogonal property of chebyshev polynomial, we get

$$c_0 = \frac{1}{\pi} \int_{-1}^{1} \frac{\left(x^3 + x^2 + 3\right) T_0(x)}{\sqrt{1-x^2}} dx$$

$$= \frac{1}{\pi} \left[\int_{-1}^{1} \frac{x^3}{\sqrt{1-x^2}} dx + \int_{-1}^{1} \frac{x^2}{\sqrt{1-x^2}} dx + \int_{-1}^{1} \frac{3}{\sqrt{1-x^2}} dx\right] \qquad \left(\because T_0(x) = 1\right)$$

$$= \frac{1}{\pi} \left[I_1 + I_2 + I_3\right] \qquad \text{...(iv)}$$

where, $I_1 = \int_{-1}^{1} \frac{x^3}{\sqrt{1-x^2}} dx = 0$ ($\because$ integrand is odd function)

$$I_2 = \int_{-1}^{1} \frac{x^2}{\sqrt{1-x^2}} dx = 2\int_{0}^{1} \frac{x^2}{\sqrt{1-x^2}} dx$$ ($\because$ integrand is even function)

substituting $x = \cos\theta$, $dx = -\sin\theta\, d\theta$, we get

$$I_2 = 2\int_{\frac{\pi}{2}}^{0} \frac{\cos^2\theta}{\sqrt{1-\cos^2\theta}} \left(-\sin\theta\, d\,\theta\right)$$

$$= 2\int_{0}^{\frac{\pi}{2}} \cos^2\theta\, d\,\theta$$

$$= 2.\frac{\pi}{4} = \frac{\pi}{2}$$

and $I_3 = \int_{-1}^{1} \frac{3}{\sqrt{1-x^2}} dx = 6\int_{0}^{1} \frac{1}{\sqrt{1-x^2}} dx$

$$= 6\left[\sin^{-1} x\right]_0^1$$

$$= 6.\frac{\pi}{2} = 3\pi$$

Thus, $c_0 = \frac{1}{\pi}\left[\frac{\pi}{2} + 3\pi\right] = 7$ [from (iv)]

Similarly,

$$c_1 = \frac{2}{\pi}\int_{-1}^{1} \frac{\left(x^3 + x^2 + 3\right) T_1(x)}{\sqrt{1-x^2}}\, dx = \frac{3}{4},$$

$$c_2 = \frac{2}{\pi}\int_{-1}^{1} \frac{\left(x^3 + x^2 + 3\right) T_2(x)}{\sqrt{1-x^2}}\, dx = \frac{1}{2}$$

Thus, the least-squares approximation of second degree is given by

$$P_2(x) = \frac{7}{2} T_0(x) + \frac{3}{4} T_1(x) + \frac{1}{2} T_2(x)$$

$$= \frac{1}{4}\left[14\, T_0(x) + 3\, T_1(x) + 2\, T_2(x)\right]$$

Example 8.8 Obtain the chebyshev polynomial approximation of second degree (best minimax approximation) to $f(x) = x^3$ on the interval $[0, 1]$

Solution : Let the required approximation be

$$P_2(x) = a_0 + a_1 x + a_2 x^2 \qquad ...(i)$$

Let, in the given interval

$$x_0 = 0,\ x_1 = \alpha,\ x_2 = \beta,\ x_3 = 1$$

Now, $\varepsilon(x) = f(x) - P_2(x)$, then

$$\varepsilon(x) = x^3 - a_0 - a_1 x - a_2 x^2 \qquad ...(ii)$$

then, by the property (27), we have

$$\varepsilon(0) + \varepsilon(\alpha) = 0,$$

$$\varepsilon(\alpha) + \varepsilon(\beta) = 0$$

$$\varepsilon(\beta) + \varepsilon(1) = 0$$

From (ii), we have

$$\alpha^3 - 2a_0 - a_1\alpha - a_2\alpha^2 = 0$$

$$\left(\alpha^3+\beta^3\right)-2a_0-a_1(\alpha+\beta)-a_2\left(\alpha^2+\beta^2\right)=0$$

$$\left(1+\beta^3\right)-2a_0-a_1(1+\beta)-a_2\left(\beta^2+1\right)=0$$

Solving this system, we get

$$a_0=\frac{\alpha(1-\beta)(\beta-\alpha)}{2(1+\alpha-\beta)}$$

$$a_1=\frac{\beta^2-\beta(1+\alpha)+\alpha\beta(\beta-\alpha)}{(1+\alpha-\beta)}$$

$$a_2=\frac{1+\alpha+\alpha^2-\beta^2}{1+\alpha-\beta} \qquad \text{...(iii)}$$

Again, using (27) $\left(\varepsilon'(x_i)=0,\ i=1,2\right)$, we get

$$\varepsilon'(\alpha)=0\Rightarrow 3\alpha^2-2a_2\alpha-a_1=0$$

and $\varepsilon'(\beta)=0\Rightarrow 3\beta^2-2a_2\beta-a_1=0$

Solving these equations, we get

$$a_1=-3\alpha\beta,\ a_2=\frac{3}{2}(\alpha+\beta) \qquad \text{...(iv)}$$

from (iii) and (iv), we get

$$\alpha=\frac{1}{4},\ \beta=\frac{3}{4}$$

and $a_0=\frac{1}{32},\ a_1=-\frac{9}{16},\ a_2=\frac{3}{2}$

Thus, required approximation is

$$P_2(x)=\frac{1}{32}-\frac{9}{16}x+\frac{3}{2}x^2$$

$$=\frac{1}{32}\left(1-18x+48x^2\right)$$

Example 8.9 Determine the best minimax approximation to the function $f(x)=x^2$ on $[0,\ 1]$ with a straight line.

Solution : Let, required straight line be given by

$$P_1(x)=a_0+a_1(x)$$

Let, $x = 0$, $x_1 = \alpha$, $x_2 = 1$

then, $\varepsilon(x) = f(x) - P_1(x)$, which gives

$$\varepsilon(x) = x^2 - a_0 - a_1 x$$

By the property (27), we have

$$\varepsilon(0) + \varepsilon(\alpha) = 0$$

$$\varepsilon(\alpha) + \varepsilon(1) = 0$$

which gives

$$\alpha^2 - \alpha a_1 - 2a_0 = 0$$

and $\alpha^2 - (1+\alpha)a_1 - 2a_0 + 1 = 0$ ...(i)

Also $\varepsilon'(x_i) = 0$, for $i = 1$, that is

$$\varepsilon'(x_1) = 0$$

which gives

$$\varepsilon'(\alpha) = 0 \Rightarrow 2\alpha - a_1 = 0 \qquad \text{...(ii)}$$

Solving (i) and (ii), we get

$$\alpha = \frac{1}{2},\ a_0 = -\frac{1}{8},\ a_1 = 1$$

Thus, requires straight line is given by

$$P_1(x) = -\frac{1}{8} + x$$

$$= \frac{1}{8}[8x - 1]$$

Example 8.10 Find a uniform polynomial approximation of degree four or less to $f(x) = \sin^{-1}(x)$ on the interval $[-1, 1]$, using Lanczos economization with error tolerance of 0.05.

Solution : Given that

$$f(x) = \sin^{-1} x$$

$$= x + \frac{x^3}{6} + \frac{3x^5}{40} + \frac{15}{336}x^7 + \ldots. \qquad \text{...(i)}$$

Since $\left|\frac{15}{336}x^7\right| \leq 0.0446$, on the interval $[-1, 1]$ which is less than the given error tolerance.

Hence, we can truncate series (i) from fourth term, thus

$$\sin^{-1} x = x + \frac{x^3}{6} + \frac{3x^5}{40}$$

$$= T_1 + \frac{1}{6}.\frac{1}{4}\left[3T_1(x) + T_3(x)\right] + \frac{3}{40}.\frac{1}{16}\left[10T_1(x) + 5T_3(x) + T_5(x)\right]$$

$$= \frac{75}{64}T_1 + \frac{25}{384}T_3 + \frac{3}{640}T_5 \qquad \text{...(ii)}$$

Again, $\left|\frac{3}{640}\right| = 0.0047$

so that, $\left|\frac{3}{640}T_5\right| \le 0.0047 \qquad \left(\because \left|T_5\right| \le 1,\ \forall n \in [-1,1]\right)$

therefore, we may truncate this term, as total error is

$$0.0446 + 0.0047 = 0.0493$$

which is less than the given error tolerance 0.05. So, we can truncate the term $3T_5/640$ from (ii). Thus, the required approximation is given by

$$\sin^{-1} x = \frac{75}{64}T_1 + \frac{25}{384}T_3$$

$$= \frac{75}{64}(x) + \frac{25}{384}\left(4x^3 - 3x\right)$$

$$= \frac{25}{96}x^4 + \frac{125}{128}x$$

Example 8.11 Find a uniform polynomial approximation of degree four or less to the function $f(x) = e^x$ on the interval $[-1,\ 1]$ using Lanczos economization with error tolerance 0.02.

Solution : Given function is

$$f(x) = e^x$$

$$= 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \ldots \qquad \text{...(i)}$$

Since $\left|\frac{x^5}{120}\right| \le 0.0083$, on the interval $[-1,\ 1]$, which is less than the given error tolerance 0.02, so we can truncate series (i) from sixth term. Hence,

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24}$$

$$= T_0(x) + T_1(x) + \frac{1}{2}.\frac{1}{2}\left[T_0(x) + T_1(x)\right]$$

$$+\frac{1}{6}.\frac{1}{4}\left[3\,T_1(x) + T_3(x)\right] + \frac{1}{24}.\frac{1}{8}\left[3\,T_0(x) + 4\,T_2(x) + T_4(x)\right]$$

$$= \frac{81}{64}T_0 + \frac{9}{8}T_1 + \frac{13}{48}T_2 + \frac{1}{24}T_3 + \frac{1}{192}T_4 \qquad \text{...(ii)}$$

Again, $\left|\frac{1}{192}\right| = 0.0052$, so that

$$\left|\frac{1}{192}T_4\right| \le 0.0052\,, \qquad \left(\because \left|T_4\right| < 1,\ \forall n \in [-1,\ 1]\right)$$

Thus, the total error is

$$0.0083 + 0.0052 = 0.0135$$

which is less than the given error tolerance 0.02, so we can truncate the term $T_4/_{192}$ from (iii) Thus, the required approximation is given by

$$e^x = \frac{81}{64}T_0 + \frac{9}{8}T_1 + \frac{13}{48}T_2 + \frac{1}{24}T_3$$

$$= \frac{81}{64}.1 + \frac{9}{8}.x + \frac{13}{48}\left(2x^2 - 1\right) + \frac{1}{24}\left(4x^3 - 3x\right)$$

$$= \frac{x^3}{6} + \frac{13}{24}x^2 + x + \frac{191}{192}$$

Self-Learning Exercise

1. Chebyshev polynomials are .... polynomials. (orthogonal/orhtonormal)

2. Chebyshev polynomial can be orthogonalized using the weight function $w(x)$ equal to

   (a) 1 (b) x

   (c) $\frac{1}{\sqrt{1-x^2}}$ (d) $\sqrt{1-x^2}$

3. Chebyshev polynomials have minimax property. (True/False)

4. Using Chebyshev polynomials, we can economize the initial poower series for the given function. (True/False)

5. Recurrence relation for the chebyshev polynomials is

$(a)\quad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$

$(b)\quad 2T_{n+1}(x) = xT_n(x) - T_{n-1}(x)$

$(c)\quad 2T_{n+1}(x) = xT_n(x) - 2T_{n-1}(x)$

$(d)\quad T_{n+1}(x) = 2xT_n(x) + T_{n-1}(x)$

8.9 Summary

In this unit, we studied two important techniques to approximate the given function, namely Taylor series and chebyshev polynomials. We also studied uniform-minimax property of chebyshev polynomial, approximation to the function using this property and economization of the power series of the given function.

8.10 Answers of Self-Learning Exercise

1.	Orthogonal	2.	(c)	3.	True
4.	True	5.	(a)		

8.11 Exercises

1. Using the Gram Schmidt orthogonalization process compute the first three orthogonal polynomials on the interval $[-1,1]$ with weight function $w(x) = (1-x^2)^{-1/2}$.

   [Ans. $1,\ x,\ \frac{1}{2}(2x^2-1)$ **]**

2. Express $T_0(x) + 2T_1(x) + T_2(x)$ as a polynomial in x.

   [Ans. $2x + 2x^2$]

3. Express $1 - x^2 + 2x^4$ as a sum of chebyshev polynomials.

   [Ans. $\frac{1}{4}[5T_0(x) + 2T_2(x) + T_4(x)]$ **]**

4. Obtain least-square approximation of second degree for x^4 on $[-1,1]$, using chebysehv polynomials.

   [Ans. $\frac{1}{8}[3T_0(x) + 4T_2(x)]$

5. Obtain the best lower degree approximation to $x^3 + 2x^2$.

   [Ans. $\frac{1}{4}(8x^2 + 2x)$]

6. Obtain the chebysehv linear polynomials (best minimax approximation) to the function x^3 on $[0,1]$.

[Ans. $\frac{1}{9}\left(9x-\sqrt{3}\right)$]

7. Determine the best minimax approximation (chebyshev polynomial approximation) to $f(x) = {}^{1}\!/_{x^2}$ on $[1,2]$ with a straight line $y = a_0 + a_1 x$.

   [Ans. $y = 1.66 - 0.75x$]

8. Use chebyshev polynomials to find the best uniform approximation of degree four or less to x^5 on $[-1,1]$

   [Ans. $\frac{5}{16}\left(4x^3 - x\right)$]

9. Compute $\sin x$ correct to three significant digits, by the economization of the power series

$$\sin x = x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{840} + \ldots$$

   [Ans. $\sin x = 0.9974x - 0.1562x^3$]

□□□

Unit - 9 : Numerical Solutions to Ordinary Differential Equations

Structure of the Unit

9.0 Objectives

After studying this unit you will be able to -

1. Understand the idea of numerical solutions to ordinary differential equations.
2. Derive various numerical methods for solving differential equations equipped with initial conditions.
3. Solve initial value problems numerically.
4. Distinguish features of different numerical methods.

9.1 Introduction

Numerical techniques for solving differential equations are of immense significance. Their utility seems to be paramount when one come across with non linear differential equations. Since, non linear differential equations don't alow analytic solutony, hence numerical techniques are resorted to. Here it should be noted that numerical techniques are "recursive formulas" which provide solutions in steps. This disertization of the differented equations make them computer friendly.

Before embarking on the different numerical methods, we first go through the basic idea of such techniques.

Initial Value Problem (IVP)

A first order differential equation

$$\frac{dy}{dt} = f(t, y), \quad t \in [t_0, b]$$

$$y(t_0) = y_0 \qquad ...(1)$$

is called an initial value problem since the condition "$y(t_0) = y_0$ i.e. at $t = t_0$, $y(t) = y_0$" is prescribed at the initial point t_0 of the solution space $[t_0, b]$. The solution of the above IVP requires to determine $y(t)$ at $t = b$.

Numerical Solution V/s Analytical Solution

Analytical solution of a differential equation is a functional relationshop between dependent variable and independent variable. For the IVP given in equation (1), the analytical solution may be of the form

$$y(t) = F(t) + C.$$

where *C* is a constant of integration whose value is determined by the initial condition $y(t_0) = y_0$ and it is

$$C = y_0 - F(t_0)$$

Thus $\quad y(t) = F(t) - F(t_0) + y_0 \qquad ...(2)$

From (2) we can find

$$y(b) = F(b) - F(t_0) + y_0$$

Contrary to above analytical approach, in numerical methods, the solution space $[t_0, b]$ is described in number of mesh (grid) points $t_0, t_1, ..., t_n$ such that

$$t_j = t_0 + jh, \; i = 0,1,2,.....,n$$

where $t_n = b$ and h is called the step size the numerical method employed provides solutions at the point $t_1, t_2, ..., t_n$, where in the solution obtained at the previous step is used to compute the solution at the next step. Here, it is worth to remind that every numerical method has some error and consequently their accuracy may differ. However, one may attain desired accuracy by prescribing the error tolerance and ultimately solution can be found to converge to the exact solution.

This existence and uniqueness of the solution of the IVP given in (1) is narrated in the following theorem.

Theorem 1 : Let $f(t, y)$ be real and continous function in the interval $[t_0, b]$, where $y \in (-\infty, \infty)$ and there exists a solution L such that for any y_1 and y_2, $|f(t, y_1) - f(t, y_2)| \le L|y_1 - y_2|$, where L is called the Lipschit constant. Then for any y_0, the IVP (1) has unique solution.

9.2 Taylor's Series Method

This method is used to solve an inital value problem numerically. The method is useful when the dependent variable give rise to convergent Taylor's series. To illustrate the method, let us consider an IVP

$$\frac{dy}{dt} = f(t, y), \; t \in [t_0, b]$$

$$y(t_0) = y_0 \qquad ...(1)$$

Then $y(t)$ can be expanded by Taylor's series about point $t = t_0$ as follows

$$y(t)=y(t_0)+(t-t_0)y'(t_0)+\frac{(t-t_0)^2}{2!}y''(t_0)+...+\frac{(t-t_0)^p}{P!}y^{(p)}(t_0)+R_p \qquad ...(2)$$

where $R_p=\frac{(t-t_0)^{p+1}}{(p+1)!}y^{(p+1)}(\xi), \qquad t_0<\xi<t$

is the truncation error.

The derivation appearing in (2) are computed manually as

$$y'=f(t,y) \quad \text{(given)}$$

$$\Rightarrow \qquad y'(t_0)=y'_0=f(t_0,y_0)$$

$$y''_0=\left[\frac{df}{dt}\right]_{t_0}=\left[\frac{\partial f}{\partial t}+\frac{\partial f}{\partial y}\frac{dy}{dt}\right]=\left[f_t+f_yf\right]_{t=t_0} \qquad [\because y'=f]$$

$$y'''_0=\left[\frac{d^2f}{dt^2}\right]_{t=t_0}=\frac{d}{dt}\left[\frac{df}{dt}\right]=\left\{\frac{d}{dt}\left[f_t+f_yf\right]\right\}_{t=t_0}=\left[\frac{d}{dt}(f_t)+\frac{d}{dt}(f_yf)\right]_{t=t_0}$$

$$=\left[f_{tt}+2f\,f_{ty}+f_tf_y+f^2f_{yy}+f\,f_y^2\right]_{t=t_0}, \text{ etc.}$$

Then the solution at the point $t=t_0+h=t_1$ is obtained as

$$y(t_1)=y_1=y_0+h\,y'_0+\frac{h^2}{2}y''_0+\frac{h^3}{6}y'''_0+...+\frac{h^p}{p!}y_0^{(p)}+R_p \qquad ...(3)$$

Now, two important issues require attention

(i) How many terms are to included in the expansion (3) so as to ensure prescribed accuracy.

The number of terms to be included is decided by the permissible error. Let his error be $\in$, then we must have

$$\left|\frac{(t-t_0)^{p+1}}{(p+1)!}y^{(p+1)}(\xi)\right|<\in, \qquad t_0<\xi<t$$

or $$\left|\frac{h^{p+1}}{(p+1)!}y^{(p+1)}(\xi)\right|<\in,\ t_0<\xi_1<t, \qquad \text{where } t-t_0=h$$

This inequality contains three unknowns $\in, h$ and p. If any two are given, then third can be determined. Again note that ξ is not known, therefore $y^{(p+1)}(\xi)$ is replaced by its estimate $\max_{t_0\le\xi\le t} y^{(p+1)}(\xi)$

(ii) The method requires manual computation of derivatives which may be time consuming and tedious. The manual computations sometimes may defeat the advantages of computing competenc while working on computer.

Example 9.1 Solve the initial value problem by Taylor's series method,

$$\frac{dy}{dt} = -[y+2t],\ t \in [0, 0.2]$$

$$y(0) = -1$$

Solution : Given that $t_0 = 0$, $y(t_0) = y_0 = -1$.

Let $h = 0.1$, so that $t_1 = 0.2$ and $y(t_1) = y_1 = ?$

Expanding $y(t)$ by Taylor's series about point $t = t_0$, we have,

$$y(t) = y(t_0) + (t-t_0)y'(t_0) + \frac{(t-t_0)^2}{2!}y''(t_0) + \frac{(t-t_0)^3}{3!}y'''(t_0) + \ldots.$$

Thus,

$$y(t_1) = y_1 = y(0.2) = y_0 + (t_1 - t_0)y'_0 + \frac{(t_1-t_0)^2}{2!}y'' + \ldots.$$

since $\quad t_1 - t_0 = h = 0.2$

Thus, $\quad y(t_1) = y_0 + (0.1)y_0^{i} + \frac{(0.1)^2}{2!}y_0^{ii} + \frac{(0.1)^3}{3!}y_0^{iii} + \frac{(0.1)^4}{4!}y_0^{iv} + \ldots.$

Now, given that,

$$y' = -y - 2t, \quad \Rightarrow y'_0 = -y_0 - 2t_0 = -(-1) - 0 = 1$$

$$y'' = -y' - 2, \quad \Rightarrow y''_0 = -y'_0 - 2 = -1 - 2 = -3$$

$$y''' = -y'', \quad \Rightarrow y'''_0 = -y''_0 = 3$$

$$y^{iv} = -y''', \quad \Rightarrow y_0^{iv} = -y'''_0 = -3 \qquad \text{etc.}$$

Thus, we have

$$y_1 = y(0.2) = -1 + (0.1)(1) + \frac{(0.1)^2}{2}(-3) + \frac{(0.1)^3}{3!}(3) + \frac{(0.1)^4}{4!}(-3) + \ldots.$$

$$= -0.91451$$

Thus, we have

$$y = -0.91451 \text{ at } t = 0.1$$

The derivatives of y at $t = 0.1$ are evaluated as,

$$y'_0(0.1) = y'_1 = -2(0.1) - (-0.91451) = 0.71451$$

$$y_0''(0.1) = y_1'' = -y_1' - 2 = -0.71451 - 2 = -2.71451$$

$$y_0''(0.1) = y_1''' = -y_1'' = 2.71451$$

$$y^{iv}(0.1) = y_1^{iv} = -y_1''' = -2.71451 \qquad \text{etc.}$$

Substituting the values of y_1 and respective derivatives in the Taylor's series expansion of $y(t)$ about $t = t_1$, we get

$$y_2 = y(0.2) = y_1 + (t - t_1)y_1' + \frac{(t-t_1)^2}{2!}y_1'' + \frac{(t-t_1)^3}{3!}y_1''' + \frac{(t-t_1)^4}{4!}y_1^{iv} + ...$$

Thus,

$$y_2 = -0.91451 + (0.1)(0.71451) + \frac{(0.1)^2}{2}(-2.71451)$$

$$+ \frac{(0.1)^3}{6}(2.71451) + \frac{(0.1)^4}{24}(-2.71451) + ...$$

$$= -0.856190$$

Verification :

The analytical solution to the given IVP is

$$y(t) = -3e^{-t} - 2t + 2$$

Thus $\quad y(0.2) = -3e^{-0.2} - 2(0.2) + 2$

$$= -0.8561923$$

Thus, we see that numerically computed result is in good aggrement with the analytical solution.

Example 9.2 Compute $y(0.2)$ by Taylor's series, where $y(t)$ is the solution of the IVP,

$$\frac{dy}{dt} = t + y, \qquad y(0) = 1$$

Solution : Given that

$$t_0 = 0,\ y_0 = 1.$$

Let the step size $h = 0.2$.

Then the grid points are $t_0 = 0$, $t_1 = 0.2$

Solution at $t = 0.2$

We have

$$y(t) = y(t_0) + (t - t_0)y'(t_0) + \frac{(t-t_0)^2}{2}y''(t_0) + \frac{(t-t_0)^3}{6}y'''(t_0) + ... \qquad ...(1)$$

Now

$$y' = t + y \qquad \Rightarrow y_0' = t_0 + y_0 = 0 + 1 = 1$$

$$y'' = 1 + y' \qquad \Rightarrow y_0'' = 1 + y_0' = 1 + 1 = 2$$

$$y''' = y'' \qquad \Rightarrow y_0''' = y_0'' = 2$$

$$y^{iv} = y''' \qquad \Rightarrow y_0^{iv} = y_0''' = 2 \qquad \text{etc.}$$

Using these derivatives in (1). we get

$$y(t) = 1 + t + t^2 + \frac{t^3}{3} + \frac{t^4}{12} + \frac{t^5}{60} + ...$$

Now,

$$y_1 = y(t_1) = y(0.2) = 1 + 0.2 + (0.2)^2 + \frac{(0.2)^3}{3} + \frac{(0.2)^4}{12} + \frac{(0.5)^5}{60} \text{ (using terms up to } t^5\text{)}$$

$$= 1.24280533$$

Example 9.3 Find the value of y at $t = 0.2$ by using seven terms Taylor's series, where $y(t)$ is the solution of the second order initial value problem,

$$\frac{d^2y}{dt^2} = 4 - t + y^2, \; y(0) = 1, \; y'(0) = -1$$

Solution : We have,

$$y'' = 4 - t + y^2, \qquad y(0) = 1, \qquad y'(0) = -1$$

$$\Rightarrow \qquad y''(0) = 4 - 0 + 1 = 5$$

$$y''' = -1 + 2yy' \Rightarrow y'''(0) = -1 + 2(1)(-1) = -3$$

$$y^{iv} = 2y'^2 + 2yy'' \Rightarrow y^{iv}(0) = 2(-1)^2 + 2(1)(5) = 12$$

$$y^{v} = 6y'y'' + 2yy''' \Rightarrow y^{v}(0) = 6(-1)(5) + 2(1)(-3) = -36$$

$$y^{vi} = 6y''^2 + 8y'y''' + 2yy^{iv}$$

$$y^{vi}(0) = 6(5)^2 + 8(-1)(-3) + 2(1)(12) = 198$$

Thus, the required seven terms series is

$$y(t) = y(0) + t\,y'(0) + \frac{t^2}{2}y''(0) + \frac{t^3}{6}y'''(0)$$

$$+ \frac{t^4}{24}y^{iv}(0) + \frac{t^5}{120}y^{v}(0) + \frac{t^6}{720}y^{vi}(0)$$

and $\quad y(0.2)=1+(0.2)(-1)+\frac{(0.2)^2}{2}(5)+\frac{(0.2)^3}{6}(-3)$

$$+\frac{(0.2)^4}{24}(12)+\frac{(0.2)^5}{120}(-36)+\frac{(0.2)^6}{720}(198)=0.8967216$$

9.3 Picard's Method

Picard's method provides polynomial solution to an initial value problem by taking successive approximations to the dependent variable. To elaborate the procedure let us consider the IVP

$$\frac{dy}{dt}=f(t,y), \quad t\in[t_0,b]$$

$$y(t_0)=y_0 \qquad \text{...(1)}$$

on integration, we have

$$\int_{y_0}^{y} dy=\int_{t_0}^{t} f(t,y)\,dt$$

or $\quad y=y_0+\int_{t_0}^{t} f(t,y)\,dt \qquad$ ...(2)

The integrand $f(t,y)$ in (2) involves unknown function y, therefore to integrate (2), we may take approximations for y.

As a first approximation to y, y in $f(t,y)$ is replaced by y_0 and then integration is done. Hence first approximation to y is

$$y^{(1)}=y_0+\int_{t_0}^{t} f(t,y_0)\,dt$$

Similarly second approximation to y is

$$y^{(2)}=y_0+\int_{t_0}^{t} f\left(t,y^{(1)}\right)dt \qquad \text{and so on}$$

Consequently, we can have a recursive scheme,

$$y^{(n)}=y_0+\int_{t_0}^{t} f\left(t,y^{(n-1)}\right)dt,$$

where $\quad y^{(0)}=y_0$

Obviously, the Picard's method would generate a polynomial in t to provide solution $y(t)$. Though, the method is quite simple but has some short comings. Firstly, the method may not proceed at the initial step or at the subsequent steps if the integrand is not integrable. Secondly, it requires manual computations to integrate.

Example 9.4 Use Picard's method to compute $y(0.5)$, where $y(t)$ is the solution to the given IVP

$$\frac{dy}{dt} = 1 + y$$

$$y(0) = 1$$

Solution : Given that

$$\frac{dy}{dt} = 1 + y$$

Let $f(t,y) = 1 + y$, $y_0 = 1$, $t_0 = 0$

First approximation to y, by Picard's method, is given by

$$y^{(1)} = y_0 + \int_{t_0}^{t} f(t, y_0)\,dt$$

$$= y_0 + \int_0^t (1 + y_0)\,dt = 1 + \int_0^t (1+1)\,dt$$

$$= 1 + 2t$$

To compute second approximation, we put $y = y^{(1)}$ in $f(t,y)$ and have

$$y^{(2)} = y_0 + \int_{t_0}^{t} f\left(t, y^{(1)}\right)dt$$

$$= 1 + \int_{t_0}^{t} (1 + 1 + 2t)\,dt$$

$$= 1 + (1+t)^2 = 2 + 2t + t^2$$

The third approximation is

$$y^{(3)} = y_0 + \int_{t_0}^{t} f\left(t, y^{(2)}\right)dt$$

$$= 1 + \int_0^t \left\{1 + \left(2 + 2t + t^2\right)\right\}dt$$

$$= 1 + 3t + t^2 + \frac{t^3}{3}$$

Similarly, the fourth approximation is

$$y^{(4)} = y_0 + \int_{t_0}^{t} f\left(t, y^{(3)}\right)\,dt$$

$$= 1 + \int_0^t \left[1 + \left(1 + 3t + t^2 + \frac{t^3}{3}\right)\right]dt$$

$$= 1 + 2t + \frac{3}{2}t^2 + \frac{t^3}{3} + \frac{t^4}{12}$$

Thus the value of $y(t)$ at $t = 0.5$ considering fourth approximation is obtained as

$$y(0.5) = 1 + 2(0.5) + \frac{3}{2}(0.5)^2 + \frac{(0.5)^3}{3} + \frac{(0.5)^4}{12}$$

$$= 2.421875$$

Example 9.5 Use Picard's method to compute $y(t)$ given that

$$\frac{dy}{dt} = \frac{e^{-t}}{y}$$

$$y(0) = 2$$

Solution : Given,

$$\frac{dy}{dt} = \frac{e^{-t}}{y} = f(t, y)$$

$$t_0 = 0,\ y(t_0) = y_0 = 2\,.$$

First approximation $y^{(1)}$ to y, by Picard's method, is given by

$$y^{(1)} = y_0 + \int_{t_0}^{t} f(t, y_0)\ dt$$

$$= 2 + \int_0^t \frac{e^{-t}}{2} = 2 + \left(\frac{e^{-t}}{2}\right)_0^t$$

$$= 2 - \frac{e^{-t}}{2} + \frac{1}{2} = \frac{5}{2} - \frac{e^{-t}}{2}$$

The second approximation $y^{(2)}$ is given by

$$y^{(2)} = y_0 + \int_{t_0}^{t} f\left(t, y^{(1)}\right)\ dt$$

$$= 2 + \int_0^t \frac{e^{-t}dt}{\left(5 - e^{-t}\right)}$$

$$= 2 + \left[\log\left(5 - e^{-t}\right)\right]_0^t$$

$$= 2 + \log\left(5 - e^{-t}\right) - \log 5$$

Similarly

$$y^{(3)} = y_0 + \int_{t_0}^{t} f\left(t, y^{(2)}\right)\, dt$$

$$= 2 + \int_0^t \frac{e^{-t} dt}{\log\left(5 - e^{-t}\right) + 2 - \log 5}$$

The intitgrand on RHS is too complex to solve. This is the practical problem with Picard's method where is one may find complicated integrand at any stage.

9.4 Runge-Kutta Methods

Runge-Kutta methods are extensions of the idea Euler's modified method which considers the average of the slopes at the end points of the subinterval (solution space for the time being) to approximate the exact solution curve. In Runge Kutta methods, the weighted average of the slopes at the end points as well as at the interior points of the interval is considered. These methods are single step method and Runge-Kutta methods of order k is comparable to the Taylor series methods of order *k*. Thus solution at $t = t_{j+1}$ is approximated as

$$y_{j+1} = y_j + \text{ (weighted average of the slopes)}$$

The order of the method is determined by the number of slopes used.

Runge-Kutta method of order two

Let us consider the initial value problem

$$\frac{dy}{dt} = f\left(t, y\right), \quad t \in \left[t_0, b\right]$$

$$y\left(t_0\right) = y_0$$

Let the mesh points are

$$t_0 < t_1 < t_2 ..... < t_n = b$$

and the spacing $h = t_j - t_{j-1}$, $j = 1, 2, ... n$

Let us consider the subinterval $\left[t_0, t_1\right]$.

The solution of y at $t = t_1$ by Runge-Kutta method is approximated as

$$y_1 = y\left(t_1\right) = y_0 + w_1 k_1 + w_2 k_2 \qquad ...(1)$$

where w_1, w_2 are weights and the slopes k_1, k_2 are given as

$$k_1 = h\, f\left(t_0, y_0\right) \qquad ...(2)$$

$$k_2 = h\, f\left(t_0 + ah,\; y_0 + b\, k_1\right) \qquad ...(3)$$

where a, b are parameters.

Here, we have to determined a, b, w_1 and w_2. Recall that, every Runge-Kutta method of order k is comparable to Taylor's series method of order *k*, therefore unknowns will be determined by having second order Taylor's series for $y(t)$ at $t = t_1$ and its comparision with (1).

Taylor's series solution for $y(t)$ about point $t = t_0$ is given by

$$y(t) = y(t_0) + (t - t_0)\,y_0' + \frac{(t-t_0)^2}{2!}\,y_0''$$

or $$y(t_1) = y_0 + h\,y_0' + \frac{h^2}{2}\,y_0'' \qquad [\because t_1 - t_0 = h] \qquad \text{...(4)}$$

Now,

$\because$ $$y' = f(t, y) \Rightarrow y_0' = f(t_0, y_0) = f_0 \text{ (say)}$$

$$y'' = \frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial y}\frac{dy}{dt} = f_t + f_y f$$

$\Rightarrow$ $$y_0'' = f_{t_0} + f_0 f_{y_0}, \qquad \text{where } f_{t_0} = (f_t)_{t=t_0} \text{ etc.}$$

On putting the values of the derivatives in (4), we get

$$y_1 = y_0 + h\,f_0 + \frac{h^2}{2}\left[f_{t_0} + f_0 f_{y_0}\right] \qquad \text{...(5)}$$

Now, again consider (1), in views of (2), (3)

$$y_1 = y_0 + w_1 h\,f_0 + w_2 h\,f\left(t_0 + ah,\ y_0 + bhf_0\right) \qquad \text{...(6)}$$

$$= y_0 + w_1 h\,f_0 + w_2 h\left[f(t_0, y_0) + \left(ah\,\frac{\partial f}{\partial t} + bhf_0\,\frac{\partial f}{\partial y}\right)_{t=t_0}\right.$$

$$\left.+\frac{1}{2}\left(a^2h^2\,\frac{\partial^2 f}{\partial t^2} + abh^2 f_0\,\frac{\partial^2 f}{\partial t\,\partial y} + b^2h^2 f_0^2\,\frac{\partial^2 f}{\partial y^2}\right)_{t=t_0}\right]$$

$$= y_0 + w_1 hf_0 + w_2 h\Big[f_0 + \left(ah\,f_{t_0} + bhf_0 f_{y_0}\right)$$

$$+\frac{1}{2}\left(a^2h^2 f_{t_0t_0} + 2abh^2 f_0\,f_{t_0y_0} + b^2h^2 f_0^2 f_{y_0}^2\right) + \ldots\Big]$$

thus, $$y_1 = y_0 + h\,f_0(w_1 + w_2) + h^2 w_2\left(a\,f_{t_0} + bf_{y_0} f_0\right) + o\left(h^3\right) \qquad \text{...(7)}$$

Comparing (7) and (5), we find that

$$w_1 + w_2 = 1, \; w_2\left(a\, f_{t_0} + b\, f_{y_0} f_0\right) = \frac{1}{2}\left(f_{t_0} + f_0 f_{y_0}\right) \qquad \text{...(8)}$$

(8) implies that

$$w_1 + w_2 = 1, \; a\, w_2 = \frac{1}{2}, \; b\, w_2 = \frac{1}{2} \qquad \text{...(9)}$$

(9) constitutes three equations in four unknown namely w_1, w_2, a and b, therefore we can choose one unknown arbitrarily.

Let us take $w_1 = \frac{1}{3}$, then

$$w_2 = \frac{2}{3}, \;\; a = \frac{3}{4}, \; b = \frac{3}{4}$$

Thus the method becomes

$$y_1 = y_0 + \frac{1}{3}k_1 + \frac{2}{3}k_2$$

where $k_1 = h\, f\left(t_0, y_0\right)$

$$k_2 = h\, f\left(t_0 + \frac{3}{4}h, \; y_0 + \frac{3}{4}k_1\right)$$

Again, if we take $w_2 = \frac{1}{2}$, then

$$w_1 = \frac{1}{2}, \;\; a = b = 1$$

and the second order Runge-Kutta method leads

$$y_1 = y_0 + \frac{\left(k_1 + k_2\right)}{2}$$

where $k_1 = h\, f\left(t_0, y_0\right)$

$$k_2 = h\, f\left(t_0 + h, \; y_0 + k_1\right)$$

Thus, it is very much clear that for different choices of the parameters, one can have different set of second order Runge-Kutta schemes. Further, note that after computing $y\left(t_1\right)$ i.e. y_1, we can compute $y\left(t_2\right)$ i.e. y_2 by the formula taking y_1 as initial condition and so forth.

Note : The expression $f\left(t_0 + ah, \; y_0 + bhf_0\right)$ in (6) has been expanded by Taylor's series for two variables. For your ready reference note that it is given as,

$$f\left(x+\Delta x,\, y+\Delta y\right)=f\left(x,y\right)+\left(\Delta x\,\frac{\partial f}{\partial x}+\Delta y\,\frac{\partial f}{\partial y}\right)$$

$$+\frac{1}{2}\left(\left(\Delta x\right)^2\,\frac{\partial^2 f}{\partial x^2}+2\,\Delta x.\Delta y\,\frac{\partial^2 f}{\partial x\,\partial y}+\left(\Delta y\right)^2\,\frac{\partial^2 f}{\partial y^2}\right)+\ldots$$

Example 9.6 Compute $y\left(0.2\right)$, using second order Runge-Kutta method with two different schemes, where $y\left(t\right)$ is the solution of the IVP

$$\frac{dy}{dt}=t+y, \qquad y\left(0\right)=1$$

Solution : $\dfrac{dy}{dt}=t+y=f\left(t,y\right)$ say,

$t_0=0$, $y\left(0\right)=y_0=1$ (given), Let $h=0.2$

then $t_1=t_0+h=0.2$, to find $y\left(t_1\right)$

Scheme 1 : $y_1=y\left(t_1\right)=y_0+\dfrac{k_1+k_2}{2}$

where $k_1=h\,f\left(t_0,y_0\right)=h\left(t_0+y_0\right)$

$=\left(0.2\right)\left[0+1\right]=0.2$

$k_2=h\,f\left(t_0+h,\,y_0+k_1\right)$

$=h\left[\left(t_0+h\right)+\left(y_0+k_1\right)\right]$

$=\left(0.2\right)\left[\left(0+0.2\right)+\left(1+0.2\right)\right]$

$=0.28$

Thus, $y_1=y\left(0.2\right)=y_0+\dfrac{\left(k_1+k_2\right)}{2}$

$=1+\dfrac{\left(0.2+0.28\right)}{2}$

$=1.24$

Scheme 2 : $y_1=y_0+\dfrac{1}{3}k_1+\dfrac{2}{3}k_2$

$k_1=h\,f\left(t_0,y_0\right)$

$$k_2 = h\,f\left(t_0 + \frac{3}{4}h,\ y_0 + \frac{3}{4}k_1\right)$$

Hence $k_1 = 0.2$, $k_2 = (0.2)\left[\left\{0 + \frac{3}{4}(0.2)\right\} + \left\{1 + \frac{3}{4}(0.2)\right\}\right]$

$$= 0.26$$

Thus $y_1 = 1 + \frac{1}{3}\,(0.2) + \frac{2}{3}(0.26)$

$$= 1.24$$

Higher order Runge-Kutta methods are also obtained by following the same procedure.

Third order Runge-Kutta Scheme

$$y(t_{j+1}) = y_{j+1} = y_j + \frac{1}{6}(k_1 + 4\,k_2 + k_3)$$

where $k_1 = h\,f\,(t_j, y_j)$

$$k_2 = h\,f\left(t_j + \frac{h}{2},\ y_j + \frac{1}{2}k_1\right)$$

$$k_3 = h\,f\,(t_j + h,\ y_j - k_1 + 2k_2)$$

Fourth order Runge-Kutta method

$$y(t_{j+1}) = y_{j+1} = y_j + \frac{1}{6}\,(k_1 + 2k_2 + 2k_3 + k_4)$$

where $k_1 = h\,f\,(t_j, y_j)$

$$k_2 = h\,f\left(t_j + \frac{h}{2},\ y_j + \frac{k_1}{2}\right)$$

$$k_3 = h\,f\left(t_j + \frac{h}{2},\ y_j + \frac{k_2}{2}\right)$$

$$k_4 = h\,f\,(t_j + h,\ y_j + k_3)$$

Exmaple 9.7 Compute $y(1.4)$, using fourth order Runge-Kutta method, given that

$$\frac{dy}{dt} = \frac{t}{y},\ y(1) = 2$$

Solution : Let the step size $h = 0.2$, then the partition of the solution space $\left[1,\ 1.4\right]$ is given by the mesh points

$$t_0 = 1, \quad t_1 = 1.2, \quad t_2 = 1.4$$

given that $y\left(t_0\right) = y_0 = 2$ and to find $y_1,\ y_2$

Computation of $y\left(t_1\right)$, i.e, y_1 :

$$y_1 = y_0 + \frac{1}{6}\left[k_1 + 2k_2 + 2k_3 + k_4\right] \qquad ...(1)$$

where $k_1 = h\,f\left(t_0, y_0\right) = h\left(\frac{t_0}{y_0}\right) = \left(0.2\right)\left[\frac{1}{2}\right]$

$$= 0.1$$

$$k_2 = h\,f\left(t_0 + \frac{h}{2},\ y_0 + \frac{k_1}{2}\right)$$

$$= h\left[\frac{t_0 + \frac{h}{2}}{y_0 + \frac{k_1}{2}}\right] = \left(0.2\right)\left[\frac{1 + \frac{0.2}{2}}{2 + \frac{0.1}{2}}\right]$$

$$k_3 = h\,f\left(t_0 + \frac{h}{2},\ y_0 + \frac{k_2}{2}\right)$$

$$= 0.10712589$$

$$k_4 = h\,f\left(t_0 + h,\ y_0 + k_3\right)$$

$$= 0.11389922$$

Using the values of the slopes in (1), we get

$$y_1 = 2 + \frac{1}{6}\left[0.1 + 2\left(0.10731707 + 0.10712589\right) + 0.11389922\right]$$

$$= 2.10713086$$

Computation of y_2 :

$$y\left(t_2\right) = y_2 = y\left(1.4\right) = y_1 + \frac{1}{6}\left[k_1 + 2\left(k_2 + k_3\right) + k_4\right] \qquad ...(2)$$

where,

$$k_1 = h\,f\left(t_1, y_1\right)$$

$$= h\left[\frac{t_1}{y_1}\right] = (0.2)\left[\frac{1.2}{2.10713086}\right]$$

$$= 0.11389895$$

$$k_2 = h\,f\left(t_1 + \frac{h}{2},\, y_1 + \frac{k_1}{2}\right)$$

$$= h\left[\frac{t_1 + h/2}{y_1 + \frac{k_1}{2}}\right] = (0.2)\left[\frac{1.2 + \frac{0.2}{2}}{2.10713086 + \frac{0.11389895}{2}}\right]$$

$$= 0.12014341$$

Similarly,

$$k_3 = h\,f\left(t_1 + \frac{h}{2},\, y_1 + \frac{k_2}{2}\right)$$

$$= 0.11997033$$

$$k_4 = h\,f\left(t_1 + h,\, y_1 + k_3\right)$$

$$= 0.12572397$$

Using these values of the slopes in (2), we get

$$y_2 = y(1.4) = 2.22710593$$

Verification : The analytic solution to the given IVP is

$$y(t) = \sqrt{3 + t^2}$$

Hence exact value of y at $t_1 = 1.2$, 1.4 are

$$y(1.2) = \sqrt{3 + (1.2)^2} = 2.10713075$$

$$y(1.4) = \sqrt{3 + (1.4)^2} = 2.22710575$$

which are in good match with the numerical solutions.

Example 9.8 Compute $y(1.2)$ by using Runge-Kutta fourth order method, where $y(t)$ is the solution of the IVP

$$\frac{dy}{dt} = ty\,, \quad y(1) = 2$$

Solution : Given that

$$t_0 = 1,\; y(t_0) = y_0 = 2$$

Let the step size $h = 0.2$. Then we have to compute $y_1 = y(t_1) = y(1.2)$, where $t_1 = t_0 + h$

further let $\frac{dy}{dt} = ty = f(t, y)$

Now fourth order Runge-Kutta scheme is

$$y_1 = y_0 + \frac{1}{6}\left[k_1 + 2k_2 + 2k_3 + k_4\right] \qquad \text{...(1)}$$

where $k_1 = h\,f(t_0, y_0) = (0.2)\,[t_0 y_0]$

$$= (0.2)\,[(1)(2)] = 0.4$$

$$k_2 = h\,f\left(t_0 + \frac{h}{2},\, y_0 + \frac{k_1}{2}\right)$$

$$= (0.2)\left[\left(t_0 + \frac{h}{2}\right)\left(y_0 + \frac{k_1}{2}\right)\right]$$

$$= (0.2)\left[\left(1 + \frac{0.2}{2}\right)\left(2 + \frac{0.4}{2}\right)\right]$$

$$= 0.484$$

$$k_3 = h\,f\left(t_0 + \frac{h}{2},\, y_0 + \frac{k_2}{2}\right)$$

$$= (0.2)\left[\left(1 + \frac{0.2}{2}\right)\left(2 + \frac{0.484}{2}\right)\right]$$

$$= 0.49324$$

$$k_4 = h\,f(t_0 + h,\, y_0 + k_3)$$

$$= (0.2)\left[(1 + 0.2)(2 + 0.49324)\right]$$

$$= 0.598377$$

Using these values of the slopes in (1), we get

$$y_1 = y(1.2) = 2 + \frac{1}{6}\left[0.4 + 2(0.484 + 0.49324) + 0.598377\right]$$

$$= 2.492142$$

Verification : We can verify the accuracy of the computed numerical results by comparing it with the exact analytical solution.

The analytical solution of the given IVP is

$$y(t) = (1.21306)\, e^{t^2/2}$$

Thus $\quad y(1.2) = (1.21306)\, e^{(1.2)^2/2}$

$$= 2.4921507$$

Thus, we see that the numerical and analytical solutions do match upto five significant places.

9.5 Numerical Solution to Higher Order Differential Equations

Higher order differential equations can be solved numerically by the previously disscussed methods with a change that they are applied to system of simultaneous first order differential equations. For this, the higher order differential equation is reduced to first order differential equations system.

For example, consider

$$\left.\begin{aligned} &\frac{d^2y}{dx^2} = f(x,y) \\ &y(a) = A,\ y'(a) = B \end{aligned}\right. \qquad \text{...(1)}$$

Let $\quad \dfrac{dy}{dx} = z$ so that $\dfrac{d^2y}{dx^2} = \dfrac{dz}{dx}$

Then (1) can be written as

$$\frac{dz}{dx} = f(x,y) = F(x,y,z) \qquad \text{(say)}$$

$$\frac{dy}{dx} = z \qquad = G(x,y,z) \qquad \text{(say)}$$

and $\quad y(a) = A, \qquad z(a) = B \qquad [\because\ y' = z] \qquad$ ...(2)

observe that (2) is system of initial value problems which can easily be solved by any method discussed previously.

We can extend earlier discussed methods to solve system of initial value problems.

As an illustration, let us consider

$$\left.\begin{aligned} &\frac{dy}{dt} = F(t,y,z) \\ &\frac{dz}{dt} = G(t,y,z) \\ &y(t_0) = y_0,\ z(t_0) = z_0 \end{aligned}\right. \qquad \text{...(1)}$$

Then the fourth order Runge-Kutta method for the system (1) is given by

$$y_1 = y_0 + \frac{1}{6}\left[k_1 + 2k_2 + 2k_3 + k_4\right]$$

$$z_1 = z_0 + \frac{1}{6}\left[l_1 + 2l_2 + 2l_3 + l_4\right]$$

where

$$k_1 = h\,F\left(t_0, y_0, z_0\right) \qquad l_1 = h\,G\left(t_0, y_0, z_0\right)$$

$$k_2 = h\,F\left(t_0 + \frac{h}{2},\, y_0 + \frac{k_1}{2},\, z_0 + \frac{l_1}{2}\right); \quad l_2 = h\,G\left(t_0 + \frac{h}{2},\, y_0 + \frac{k_1}{2},\, z_0 + \frac{l_1}{2}\right)$$

$$k_3 = h\,F\left(t_0 + \frac{h}{2},\, y_0 + \frac{k_2}{2},\, z_0 + \frac{l_2}{2}\right); \quad l_3 = h\,G\left(t_0 + \frac{h}{2},\, y_0 + \frac{k_2}{2},\, z_0 + \frac{l_2}{2}\right)$$

$$k_4 = h\,F\left(t_0 + h,\, y_0 + k_3,\, z_0 + l_3\right); \qquad l_4 = h\,G\left(t_0 + h,\, y_0 + k_3,\, z_0 + l_3\right)$$

Similarly, one can extend the Taylor's series method, Picard method to higher order initial value problem by converting it into first order initial value system.

Example 9.9 Solve the following initial value problem

$$\frac{d^2y}{dt^2} + 2\frac{dy}{dt} + y = 0 \qquad t \in [0,\ 0.1]$$

$$y(0) = 0,\ y'(0) = 1$$

Solution : Let $\frac{dy}{dt} = z$, then $\frac{d^2y}{dt^2} = \frac{dz}{dt}$

Thus the given IVP becomes

$$\begin{aligned} \frac{dz}{dt} &= -2z - y = F(t, y, z) \\ \frac{dy}{dt} &= z \qquad\quad = G(t, y, z) \end{aligned} \qquad ...(1)$$

with the conditions $y(0) = 0,\ z(0) = 1$

The fourth order Runge-Kutta method for (1) is

$$z_1 = z_0 + \frac{1}{6}\left[k_1 + 2k_2 + 2k_3 + k_4\right]$$

$$y_1 = y_0 + \frac{1}{6}\left[l_1 + 2l_2 + 2l_3 + l_4\right]$$

where $k_1 = h\,F\left(t_0, y_0, z_0\right) \qquad l_1 = h\,G\left(t_0, y_0, z_0\right)$

$$k_2 = h\,F\left(t_0 + \frac{h}{2},\, y_0 + \frac{l_1}{2},\, z_0 + \frac{k_1}{2}\right), \quad l_2 = h\,G\left(t_0 + \frac{h}{2},\, y_0 + \frac{l_1}{2},\, z_0 + \frac{k_1}{2}\right),$$

$$k_3 = h\,F\left(t_0 + \frac{h}{2},\, y_0 + \frac{l_2}{2},\, z_0 + \frac{k_2}{2}\right), \quad l_3 = h\,G\left(t_0 + \frac{h}{2},\, y_0 + \frac{l_2}{2},\, z_0 + \frac{k_2}{2}\right),$$

$$k_4 = h\,F\left(t_0 + h,\, y_0 + l_3,\, z_0 + k_3\right), \qquad l_4 = h\,G\left(t_0 + h,\, y_0 + l_3,\, z_0 + k_3\right).$$

Given that $t_0 = 0$, $y(0) = y_0 = 0$, $y'(0) = z(0) = 1$

Let $h = 0.1$

Then $k_1 = (0.1)\,[-2z_0 - y_0] = (0.1)\,[-2(1) - 0] = -0.2$

$$l_1 = (0.1)\,[z_0] = (0.1)\,[1] = 0.1$$

$$k_2 = (0.1)\left[-2\left(z_0 + \frac{k_1}{2}\right) - \left(y_0 + \frac{l_1}{2}\right)\right]$$

$$= (0.1)\left[-2\left(1 + \left(\frac{0.2}{2}\right)\right) - \left(0 + \frac{0.1}{2}\right)\right] = -0.185$$

$$l_2 = (0.1)\left[z_0 + \frac{k_1}{2}\right] = (0.1)\left[1 - \frac{0.2}{2}\right] = 0.09$$

$$k_3 = (0.1)\left[-2\left(z_0 + \frac{k_2}{2}\right) - \left(y_0 + \frac{l_2}{2}\right)\right]$$

$$= (0.1)\left[-2\left(1 + \left(-\frac{0.185}{2}\right)\right) - \left(0 + \frac{0.09}{2}\right)\right]$$

$$= -0.186$$

$$l_3 = (0.1)\left[z_0 + \frac{k_2}{2}\right] = (0.1)\left[1 - \frac{0.185}{2}\right] = 0.09075$$

$$k_4 = (0.1)\left[-2\left(z_0 + \frac{k_3}{2}\right) - (y_0 + l_3)\right]$$

$$= (0.1)\left[-2(1 - 0.186) - (0 + 0.09075)\right]$$

$$= -0.17187$$

$$l_4 = (0.1)\,[z_0 + k_3]$$

$$= (0.1)[1 - 0.186] = 0.0814$$

Thus, $y_1 = y(0.1) = y_0 + \frac{1}{6}\,[l_1 + 2l_2 + 2l_3 + l_4]$

$$= 0 + \frac{1}{6}\left[0.1 + 2\left(0.09 + 0.09075\right) + 0.0814\right]$$

$$= 0.09048$$

Verification : The analytical solution of the given IVP is

$$y(t) = t\,e^{-t}$$

Thus $y(0.1) = (0.1)e^{-0.1} = 0.0904837$

Thus we see that the numerical solution and analytical solution agree upto five places of decimals.

Example 9.10 Compute $x(0.1)$, $y(0.1)$ by Taylor's series method where $x(t), y(t)$ satisfy the following system of initial valur problems,

$$\frac{dx}{dt} = xy + 2t\,, \quad \frac{dy}{dt} = 2ty + x$$

$$x(0) = 1,\; y(0) = 2$$

Solution : We expand $x(t)$, $y(t)$ about point $t = 0$ by Taylor's series. Let $t_0 = 0$ and we denote $x(t_0) = x_0$, $y(t_0) = y_0$. Thus, $x_0 = 1$, $y_0 = 2$

Hence,

$$x(t) = x_0 + (t-0)x_0' + \frac{(t-0)^2}{2}\,x_0'' + \ldots \qquad \ldots(1)$$

$$y(t) = y_0 + (t-0)y_0' + \frac{(t-0)^2}{2}\,y_0'' + \ldots \qquad \ldots(2)$$

The derivatives appearing in (1), (2) are computed as given below :

Given that $x' = xy + 2t$, $\quad x(0) = 1,\; y(0) = 2$

$$y' = 2ty + x \quad \Rightarrow x_0' = x_0y_0 + 2t_0 = 2\,, \; y_0' = 2t_0y_0 + x_0 = 1$$

Thus on differentiating successively with respect to t and using given values, we find,

$$x'' = x'y + xy' + 2 \Rightarrow x_0'' = x_0'y_0 + x_0y_0' + 2 = 2(2) + 1(1) + 2 = 7$$

$$y'' = 2y + 2ty' + x' \Rightarrow y_0'' = 2y_0 + 2t_0y_0' + x_0' = 2(1) + 0 + 2 = 4$$

$$x''' = x''y + x'y' + x'y' + xy'' \Rightarrow x_0''' = 2x_0'y_0' + 2x_0y_0'' + y_0x_0'' = 4 + 4 + 14 = 22$$

$$y''' = 2y' + 2y' + 2ty'' + x'' \Rightarrow y_0''' = 4y_0' + 2t_0y_0'' + x_0'' = 4 + 0 + 7 = 11$$

etc.

Thus, putting the values of the derivatives in (1), (2), we get

$$x(t) = 1 + 2t + \frac{7}{2}t^2 + \frac{22}{6}t^3 + \ldots$$

$$y(t) = 2 + t + \frac{4t^2}{2} + \frac{11}{6}t^3 + \ldots$$

Thus, $x(0.1)$, $y(0.1)$ using four terms Taylor's series are

$$x(0.1) = 1 + 2(0.1) + \frac{7}{2}(0.1)^2 + \frac{22}{6}(0.1)^3$$

$$= 1.238666$$

$$y(0.1) = 2 + (0.1) + 2(0.1)^2 + \frac{11}{6}(0.1)^3$$

$$= 2.1218333$$

Self-Learning Exercise

1. Using Taylor's series method, solve

$$\frac{dy}{dt} = y\sin t + \cos t$$

for some t, given that $y(0) = 0$

2. Use Picard's method to compute $y(0.1)$, given that

$$\frac{dy}{dt} = 3t + y^2,\ y(0) = 1$$

3. Use fourth order Runge-Kutta method to compute $y(0.4)$, given that

$$\frac{dy}{dt} = -2t - y,\ y(0) = -1 \qquad \text{[Take step size } h = 0.1\text{]}$$

4. Solve the following system of equations

$$\frac{dx}{dt} = x + 2y,\ \frac{dy}{dt} = 3x + 2y$$

$$x(0) = 6,\ y(0) = 4$$

by Runge-Kutta method over the interval $[0.02,\ 0.04]$ with setp size $h = 0.02$.

9.6 Summary

In this unit you have studied numerical methods to solve initial value problems. Taylor's series method and Picard's method involve series computations. Runge-Kutta methods use weighted average of the slopes at the end points and at the internal points of the solution space to approximate the solution. Recall that we may have Runga-Kutta method of any order where it must be remembered that *k th* order Runge-Kutta method is comparable to *k th* order Taylor's series method.

All the methods dicussed in this unit are single step methods since these require information at the preceeding one point only to predict the value at the next point.

Further recall that these methods are recursive formulae which provide solutions in steps at different mesh points of the solution space.

9.7 Answers of Self-Learning Exercises

1. $y(t)=1+t+\frac{3}{2}t^2+\frac{5}{6}t^3+\frac{7}{12}t^4$

2. $y(0.1)=1.127$

3. $y(0.4)=-0.811$

4. $x(0.02)=6.2935$ $x(0.04)=6.6156$

   $y(0.02)=4.5393$ $y(0.04)=5.1195$

9.8 Exercises

1. Use Picard's method to compute $y(2.1)$, where $y(t)$ is the solution of the following IVP

$$\frac{dy}{dt}=1+ty$$

$$y(2)=0$$

[Ans. $y_{(t)}^{(3)}=-\frac{22}{15}+t-\frac{2}{3}t^2+\frac{t^3}{3}-\frac{t^4}{4}+\frac{t^5}{15}$

Thus, $y(2.1)=-\frac{22}{15}+2.1-\frac{2}{3}(2.1)^2+\frac{(2.1)^3}{3}-\frac{(2.1)^4}{4}+\frac{(2.1)^5}{15}$

2. Compute $y(1.4)$, where $y(t)$ satisfies $\frac{dy}{dt}=ty$, $y(1)=2$ {Take step size $h=0.2$]

by Runge-Kutta method

[Ans. $y(1.4)=3.2321$**]**

3. Compute $y(1.5)$, where $y(t)$ is the solution of the following IVP

$$\frac{dy}{dt}=\frac{1}{t+y}$$

$$y(0)=1$$

[Ans. $y(0.5)=1.3571$, $y(1.0)=1.5837$ $y(1.5)=1.7555$**]**

4. Compute $y(2.2)$ by Taylor's series method given that

$$\frac{dy}{dt} = 1 - \frac{y}{t}, \; y(2) = 2$$

[Ans. $y(2.2) = 2.0091$**]**

5. Given that

$$\frac{dy}{dt} = ty + x, \quad \frac{dx}{dt} = xy + t$$

$$y(0) = -1, \; x(0) = 1$$

compute $y(0.2), \; x(0.2)$ by fourth order Runge-Kutta method [Take step size $h = 0.2$]

[Ans. $x(0.2) = 0.8522$**,** $y(0.2) = -0.8341$**]**

6. Compute $y(0.2)$ by second order Runge-Kutta method, given that

$$\frac{d^2y}{dt^2} - \frac{3dy}{dt} + 2y = t + e^{3t}$$

$$y(0) = 4.3 \qquad y'(0) = 5.8$$

[Ans. $y(0.2) = 5.6841$**]**

7. Use Taylor's series method to compute $x(0.1), \; y(0.1)$, given that

$$\frac{dx}{dt} = x + y + t, \qquad \frac{d^2y}{dt^2} = x - t$$

[Ans. $x(0.1) = 1.3105$, $y(0.1) = 1.0853$]

8. Compute $x(0.05), \; y(0.05)$, given that

$$\frac{dx}{dt} = xy + t, \quad \frac{dy}{dt} = ty + x$$

[use Taylor's series method]

[Ans. $x(0.05) = 1.7264$**,** $y(0.05) = -2.91068$**]**

□□□

Unit - 10 : Numerical Solutions to Initial Value Problems

Structure of the Unit

10.0 Objectives

After studying this unit you will be able to

1. Distinguish single step method and multistep method.
2. Derive formulae for multistep methods, namely Milne's method and Adams Moulton method.
3. Understand stability of a numerical method.

10.1 Introduction

The methods discussed in previous sections are called single step methods simply because they require information at only one preceeding point t_j to predict the value at t_{j+1}. Contrary to this, in a multistep method we require values of $y(t)$ at the preceeding points $t_j, t_{j-1}, t_{j-2} \ldots$ to evaluate $y\left(t_{j+1}\right)$. These values are required to be used in suitable formulas. The predictor corrector methods (which involve a predictor formula to predict the value $y\left(t_{j+1}\right)$ and a corrector formula to improvise $y\left(t_{j+1}\right)$ are multistep methods. You may recall that Euler's modified method is an example of predictor corrector method where $y\left(t_{j+1}\right)$ obtained from Euler's method is improved by Euler's modified method. Here, Euler's method may be termed predictor formula and the Euler's modified method may be termed corrector formula.

However, it is important to note that we can derive higher order predictor-corrector strategies to solve an initial value problem with better accuracy.

10.2 Milne's Method

This method is a multistep method which requires information for the dependent variable at past four equidistant points. The method involves predictor-corrector formulae which are derived by using Newton's forward interpolation formula. Note that since four points are used to interpolate the solution, therefore it infers that a third degree polynomial is used to interpolate the solution. This means that differences upto third order will be taken into account. We now proceed to derive the predictor-corrector formula to solve the initial value problem (IVP)

$$\frac{dy}{dt} = f(t,y), \quad t \in [t_0, b] \qquad ...(1)$$

Predictor Formula :

We integrate (1) over the interval $[t_0, t_0+4h]$ where h being the step size. Thus, we have

$$\int_{t_0}^{t_0+4h} \frac{dy}{dt}\, dt = \int_{t_0}^{t_0+4h} f(t,y)\, dt$$

or $\quad y(t_0+4h) - y(t_0) = \int_{t_0}^{t_0+4h} f(t,y)\, dt$

or $\quad y_4 = y_0 + \int_{t_0}^{t_0+4h} f(t,y)\, dt \qquad$ [Note that $t_j = t_0 + jh$, $j = 0,1,2,....$] $\qquad ...(2)$

We now use Newton's forward interpolation to expand $f(t,y)$ about point $t = t_0$. We use the notation $f_j = f(t_j, y_j)$

Thus

$$f(t,y) = f_0 + u\Delta f_0 + \frac{u(u-1)}{2}\Delta^2 f_0 + \frac{u(u-1)(u-2)}{6}\Delta^3 f_0 + \frac{u(u-1)(u-2)(u-3)}{24}\Delta^4 f_0 + ...$$

where $\quad t = t_0 + uh$, $dt = h\,du \qquad ...(3)$

using (3) in (2), we get

$$y_4 = y_0 + h\int_0^4 \left[f_0 + u\Delta f_0 + \frac{u(u-1)}{2}\Delta^2 f_0 + \frac{u(u-1)(u-2)}{6}\Delta^3 f_0 + \frac{u(u-1)(u-2)(u-3)}{24}\Delta^4 f_0 + ... \right] du$$

$$= y_0 + h\left[f_0 u + \frac{u^2}{2}\Delta f_0 + \left(\frac{u^3}{6} - \frac{u^2}{4}\right)\Delta^2 f_0 + \frac{1}{6}\left(\frac{u^4}{4} - u^3 + u^2\right)\Delta^3 f_0 + \frac{1}{24}\left(\frac{u^5}{5} - \frac{3}{2}u^4 + \frac{11}{3}u^3 - 3u^2\right)\Delta^4 f_0 + ...\right]_0^4$$

$$= y_0 + h\left[4f_0 + 8\Delta f_0 + \frac{20}{3}\Delta^2 f_0 + \frac{8}{3}\Delta^3 f_0 + \frac{28}{90}\Delta^4 f_0 \right] \qquad ...(4)$$

The various order forward differences appearing in (4) are evaluated as

$$\Delta f_0 = f_1 - f_0$$

$$\Delta^2 f_0 = f_2 - 2f_1 + f_0$$

$$\Delta^3 f_0 = f_3 - 3f_2 + 3f_1 - f_0$$

on putting these expression in (4) and simplifying, we get

$$y_4 = y_0 = \frac{4h}{3}\left[2f_1 - f_2 + 2f_3\right] + \frac{28h}{90}\Delta^4 f_0 \qquad ...(5)$$

The term $\frac{28h}{90}\Delta^4 f_0$ is the truncation error of the above formula and the formula in working form is

$$y_4 = y_0 + \frac{4h}{3}\left[2f_1 - f_2 + 2f_3\right]$$

or $$y_4 = y_0 + \frac{4h}{3}\left[2y_1' - y_2' + 2y_3'\right] \qquad \left[\because y' = f\right] \qquad ...(6)$$

The formula (6) is called Milne's predictor formula. The value y_4 predicted in (6) is improved by employing corrector formula.

Corrector Formula :

Integrating (1) over the interval $\left[t_0,\, t_0 + 2h\right]$ and proceeding above we obtain

$$y_2 = y_0 + \frac{h}{3}\left[f_0 + 4f_1 + f_2\right] - \frac{h}{90}\Delta^4 f_0$$

or $$y_4 = y_2 + \frac{h}{3}\left[f_2 + 4f_3 + f_4\right] - \frac{h}{90}\Delta^4 f_0 \qquad \text{[Note]}$$

or $$y_4 = y_2 + \frac{h}{3}\left[y_2' + 4y_3' + y_4'\right] - \frac{h}{90}\Delta^4 f_0 \qquad ...(7)$$

The term $\left(-\frac{h}{90}\Delta^4 f_0\right)$ in (7) is the truncation error of the corrector formula.

Thus, predictor-corrector formula in general form can be written as

Predictor : $$y_{j+1} = y_{j-3} + \frac{4h}{3}\left[2y_{j-2}' - y_{j-1}' + 2y_j'\right]$$

Corrector : $$y_{j+1} = y_{j-1} + \frac{h}{3}\left[y_{j-1}' + 4y_j' + y_{j+1}'\right]$$

Example 10.1 Solve the following IVP by Milne's method, given that

$$\frac{dy}{dt} = t + y, \qquad t \in [0,\, 0.4]$$

$$t_0 = 0,\; y_0 = 1$$

Solution : We know that to compute the value of dependent variable y at certain point in the solution space, we require information at past four points. Here, we are given information at the initial point only.

We suppose the step size $h = 0.1$

Then the grid points are

$$t_0 = 0,\ t_1 = 0.1,\ t_2 = 0.2,\ t_3 = 0.3,\ t_4 = 0.4$$

Thus in order to compute $y(t_4) = y_4$, we require the estimates for y_1, y_2 and y_3. These can be computed by any method discussed earlier.

We use Taylor's series method to compute these values and obtain

$$y_1 = 1.1103,\ y_2 = 1.2428,\ y_3 = 1.3997$$

Milne's predictor formula is

$$y_4^{(p)} = y_0 + \frac{4h}{3}\left[2y_1' - y_2' + 2y_3'\right] \qquad ...(1)$$

the derivatives appearing in (1) are computed as

$$y' = t + y \qquad \text{[given]} \qquad \Rightarrow y_1' = t_1 + y_1 = 0.1 + 1.1103 = 1.2103$$

Thus

$$y_2' = t_2 + y_2 = 0.2 + 1.2428 = 1.4428$$

$$y_3' = t_3 + y_3 = 0.3 + 1.3997 = 1.6997$$

Using these values in (1), we obtain,

$$y_4^{(p)} = 1 + \frac{4(0.1)}{3}\left[2(1.2103) - (1.4428) + 2(1.6997)\right]$$

$$= 1.58362$$

Thus $y_4^p = 1.58362$ is the predicted value of y at $t = t_4 = 0.4$. This can be improved by employing the Milne's corrector formula.

$$y_4^{(c)} = y_2 + \frac{h}{3}\left[y_2' + 4y_3' + y_4'\right]$$

Now $y_4' = t_4 + y_4 = 0.4 + 1.58362$

$$= 1.98362$$

Thus $y_4^{(c)} = 1.2428 + \frac{(0.1)}{3}\left[1.4428 + 4(1.6997) + 1.98362\right]$

$$= 1.583640$$

Verification : Ths exact solution to the given IVP is

$$y(t) = 2e^t - (t-1)$$

$$y(0.4) = 2e^{0.4} - (0.4+1)$$

$$= 1.5836493$$

Thus we see that $y(0.4)_{\text{exact}}$ and $y(0.4)_{\text{numerical}}$ do match excellenty upto five decimal places.

Example 10.2 Solve by Milne's method

$$\frac{dy}{dt} = \frac{t}{y},\ y(1) = 2,\ t \in [1, 1.4]$$

Solution : Milne's method requires information at past four points to predict the value of dependent variable y, but in the question given the values are not supplied. These may be computed by any intial value method and then Milne's method is applied.

We will use Taylor's series method to compute these values. Let the step size $h = 0.1$ then the mesh points are

$$t_0 = 1,\ t_1 = 1.1,\ t_2 = 1.2,\ t_3 = 1.3,\ t_4 = 1.4$$

given that $y(t_0) = y_0 = 1$.

We will find $y_1 = y(t_1),\ y_2 = y(t_2),\ y_3 = y(t_3)$

Now, $y' = \frac{t}{y}$ (given) $\Rightarrow y'(t_0) = y'_0 = \frac{1}{2} = 0.5$

$$\Rightarrow y'' = \frac{1}{y} - \frac{t}{y^2}y' = \frac{1}{y} - \frac{t^2}{y^3}$$

$$\Rightarrow y''_0 = \frac{1}{2} - \frac{1}{8} = \frac{3}{8}$$

Similarly, $y'''_0 = -\frac{9}{32}$ etc.

Now Taylor's series solution at $t = t_1$

$$y_1 = y(t_1) = y_0 + h\,y'_0 + \frac{h^2}{2}\,y''_0 + \frac{h^3}{6}\,y'''_0$$

$$= 2 + 0.1(0.5) + \frac{(0.1)^2}{2}\left(\frac{3}{8}\right) + \frac{(0.1)^3}{6}\left(-\frac{9}{32}\right)$$

$$= 2.051828$$

Similarly,

$$y_2 = y(t_2) = y_1 + h\,y'_1 + \frac{h^2}{2}\,y''_1 + \frac{h^3}{6}\,y'''_1$$

$$y_1^{'} = \frac{t_1}{y_1} = \frac{1.1}{2.051828}$$

$$y_1'' = \frac{1}{y_1} - \frac{t_1^2}{y_1^3} = \frac{1}{2.051828} - \frac{(1.1)^2}{(2.051828)^3} \qquad \text{etc}$$

Thus, $y_2 = y(t_2) = 2.051828 + 0.536107 + 0.001736 - 0.000045$

$= 2.107130$

Similarly, $y_3 = 2.16564$

Thus we have

t_0	t_1	t_2	t_3	t_4
$t:1$	1.1	1.2	1.3	1.4
$y:2$	2.051828	2.107130	2.16564	?

Milne's predictor formula is

$$y_4^{(p)} = y_0 + \frac{4h}{3}\left[2y_1' - y_2' + 2y_3'\right] \qquad ...(1)$$

Since $y' = \frac{t}{y} \Rightarrow y_1' = \frac{t_1}{y_1} = 0.53607$

$$y_2' = \frac{t_2}{y_2} = 0.569495$$

$$y_3' = \frac{t_3}{y_3} = 0.600284$$

Using these values in (1), we get

$$y_4^{(p)} = 2.227095$$

Thus the predicted value of y at $t_4 = 1.4$ is 2.227095

Now, we use corrector formula to improve this value.

Corrector formula is given by

$$y_4^{(c)} = y_2 + \frac{h}{3}\left[y_2' + 4y_3' + y_4'\right] \qquad ...(2)$$

Now $y_4' = 2.10713 + \frac{0.1}{3}\left[0.569495 + 4(0.600284) + 0.628622\right]$

$= 2.227105$

Verification : The exact solution to the given IVP is

$$y(t) = \sqrt{3+t^2}$$

Thus $y(1.4) = \sqrt{3+(1.4)^2}$

$$= 2.22710575$$

We see that $y(1.4)_{\text{exact}}$ and $y(1.4)_{\text{numerical}}$ do match upto six decimal places.

Example 10.3 Compute $y(0.5)$ by Milne's method, given that

$$\frac{dy}{dt} = 2e^t - y$$

and the corresponding values of t and y are given as

$t:0$	0.1	0.2	0.3
$y:2$	2.01	2.04	2.09

Solution : Obviously step size $h = 0.1$

Thus the mesh points $t_n = t_0 + nh$ are

$$t_0 = 0,\ t_1 = 0.1,\ t_2 = 0.2,\ t_3 = 0.3,\ t_4 = 0.4,\ t_5 = 0.5$$

First we compute $y_4 = y(t_4) = y(0.4)$

Then this value of y_4 is used to compute y_5.

Milne's predictor formula is given by

$$y_4^{(p)} = y_0 + \frac{4h}{3}[2y_1' - y_2' + 2y_3']$$

Now $y_1' = 2e^{t_1} - y_1 \qquad [\because y' = 2e^t - y]$

$$= 2e^{0.1} - 2.01 = 0.2004$$

$$y_2' = 2e^{t_2} - y_2 = 2e^{0.2} - 2.04 = 0.4028$$

$$y_3' = 2e^{t_3} - y_3 = 2e^{0.3} - 2.09 = 0.6098$$

Thus

$$y_4^{(p)} = 2 + \frac{4(0.1)}{3}[2(0.2004) - (0.4028) + 2(0.6098)]$$

$$= 2.1623$$

Therefore

$$y_4' = 2\,e^{t_4} - y_4 = 0.8213$$

Now, corrector formula is given by

$$y_4^{(c)} = y_2 + \frac{h}{3}\left[y_2' + 4y_3' + y_4'\right]$$

$$= 2.04 + \frac{0.1}{3}\left[0.4028 + 4\left(0.6098\right) + 0.8213\right]$$

$$= 2.162$$

Thus, $y\left(t_4\right) = y_4 = y\left(0.4\right) = 2.162$

Now, Milne's Predictor formula for $y\left(t_5\right)$ is

$$y_4^{(p)} = y_1 + \frac{4h}{3}\left[2y_2' - y_3' + 2y_4'\right]$$

$$= 2.01 + \frac{4\left(0.1\right)}{3}\left[2\left(0.4028\right) - 0.6098 + 2\left(0.8213\right)\right]$$

$$= 2.2551$$

Further, $y_5' = 2e^{t_5} - y_5 = 2e^{0.5} - 2.2551 = 1.0423$

Thus, corrector formula

$$y_5^{(c)} = y_3 + \frac{h}{3}\left[y_3' + 4y_4' + y_5'\right]$$

$$= 2.09 + \frac{0.1}{3}\left[0.6098 + 4\left(0.8213\right) + 1.0423\right]$$

$$= 2.25458$$

Hence $y_5 = y\left(0.5\right) = 2.25458$

10.3 Adams-Moulten Method

This is another multistep method which involves generation of predictor-corrector formulae using the backward interpolation formula. The method is used to solve an initial value problem and requires information at past four equidistant points t_{j-3}, t_{j-2}, t_{j-1}, t_j (say) to compute the solution at $t = t_{j+1}$.

Let us consider the initial value problem

$$\frac{dy}{dt} = f\left(t, y\right),\ y\left(t_j\right) = y_j \qquad ...(1)$$

Predictor Formula :

We integrate (1) on the interval $\left[t_j, t_{j+1}\right]$

$$\int_{t_j}^{t_{j+1}} \frac{dy}{dt} dt = \int_{t_j}^{t_{j+1}} f(t,y)\, dt$$

This gives

$$(y)_{t_j}^{t_{j+1}} = \int_{t_j}^{t_{j+1}} f(t,y)\, dt$$

or $$y_{j+1} - y_j = \int_{t_j}^{t_{j+1}} f(t,y)\, dt$$

or $$y_{j+1} = y_j + \int_{t_j}^{t_{j+1}} f(t,y)\, dt \qquad ...(2)$$

The function $f(t,y)$ in (2) is replaced by backward difference formula and we get

$$y_{j+1} = y_j + \int_{t_j}^{t_{j+1}} \left[f(t_j) + u\nabla f(t_j) + \frac{u(u+1)}{2} \nabla^2 f(t_j) \right.$$

$$\left. + \frac{u(u+1)(u+2)}{6} \nabla^3 f(t_j) + \frac{u(u+1)(u+2)(u+3)}{24} \nabla^4 f(t_j) + \ldots \right] dt \qquad ...(3)$$

where $t = t_j + hu \Rightarrow dt = h\,du$

Now note that when $t = t_{j+1} = t_j + h$ $\qquad [\because t_{j+1} = t_j + h]$

then $t_j + hu = t_j + h$

or $u = 1$

and when $t = t_j$, then $t_j + hu = t_j$

$\Rightarrow$ $u = 0$

Thus we have got the limits

when $t = t_{j+1}$ then $u = 1$

$t = t_j$ then $u = 0$

Now on changing the variable of integration from t to u, (3) becomes

$$y_{j+1} = y_j + h \int_0^1 \left[f_j + u\nabla f_j + \frac{u(u+1)}{2} \nabla^2 f_j \right.$$

$$\left. + \frac{u(u+1)(u+2)}{6} \nabla^3 f_j + \frac{u(u+1)(u+2)(u+3)}{24} \nabla^4 f_j + \ldots. \right] du$$

where $f_j = f\left(t_j, y_j\right)$

On simplification, we get

$$y_{j+1} = y_j + h\left[f_j + \frac{1}{2}\nabla f_j + \frac{5}{12}\nabla^2 f_j + \frac{3}{8}\nabla^3 f_j + \frac{251}{724}\nabla^4 f_j\right] \qquad ...(4)$$

Recall that the backward differences are given by

$$\nabla f_j = f\left(t_j\right) - f\left(t_j - h\right)$$

$$= f_j - f_{j-1}$$

$$\nabla^2 f_j = \nabla\left[\nabla f_j\right] = \nabla\left[f_j - f_{j-1}\right]$$

$$= \nabla f_j - \nabla f_{j-1}$$

$$= f_j - f_{j-1} - \left(f_{j-1} - f_{j-2}\right)$$

$$= f_j - 2f_{j-1} + f_{j-2}$$

Similarly, $\nabla^3 f_j = f_j - 3f_{j-1} + 3f_{j-2} - f_{j-3}$ etc.

Substituting the above differences in (4) and simplifying, we obtain

$$y_{j+1} = y_j + \frac{h}{24}\left[55f_j - 59f_{j-1} + 37f_{j-2} - 9f_{j-3}\right] + h\left(\frac{251}{720}\nabla^4 f_j\right) \qquad ...(5)$$

or $$y_{j+1} = y_j + \frac{h}{24}\left[55y'_j - 59y'_{j-1} + 37y'_{j-2} - 9y'_{j-3}\right] + \frac{251}{720}\nabla^4 y'_j \quad \left[\because y' = f\right] \qquad ...(6)$$

The formula given by (5) or (6) is called Adams-Moulton formula.

Corrector Fromula :

In order to get the corrector formula the function $f(t, y)$ is replaced by Newton's backward interpolation formula about point $t = t_{j+1}$, that is, we take $t = t_{j+1} + hu \Rightarrow dt = h\,du$

Recall equation (2)

$$y_{j+1} = y_j + \int_{t_j}^{t_{j+1}} f(t, y)\,dt$$

We expand $f(t, y)$ about $t = t_{j+1}$ by Newton's backward formula to get

$$f(t, y) = f\left(t_{j+1}\right) + u\nabla f\left(t_{j+1}\right) + \frac{u(u+1)}{2}\nabla^2 f\left(t_{j+1}\right) + ...$$

Using this expression in (2), we get

$$y_{j+1} = y_j + \int_{t_j}^{t_{j+1}} \Bigg[f_{j+1} + u\nabla f_{j+1} + \frac{u(u+1)}{2}\nabla^2 f_{j+1}$$

$$+\frac{u(u+1)(u+2)}{6}\nabla^3 f_{j+1} + \frac{u(u+1)(u+2)(u+3)}{24}\nabla^4 f_{j+1} + \ldots \Bigg] dt \qquad \ldots(7)$$

Note that since $t = t_{j+1} + hu$

Hence, when $t = t_{j+1}$ then $u = 0$

when $t = t_j$ then $u = -1$

Thus on changing variable of integration from t to u, (7) becomes

$$y_{j+1} = y_j + h\int_{-1}^{0} \Bigg[f_{j+1} + u\nabla f_{j+1} + \frac{u(u+1)}{2}\nabla^2 f_{j+1}$$

$$+\frac{u(u+1)(u+2)}{6}\nabla^3 f_{j+1} + \frac{u(u+1)(u+2)(u+3)}{24}\nabla^4 f_{j+1} + \ldots \Bigg] du \qquad \ldots(8)$$

Again, we have backward differences as

$$\nabla f_{j+1} = f_{j+1} - f_j$$

$$\nabla^2 f_{j+1} = f_{j+1} - 2f_j + f_{j-1}$$

$$\nabla^3 f_{j+1} = f_{j+1} - 3f_j + 3f_{j-1} - f_{j-2} \qquad \text{etc.}$$

On putting the values of the backward differences in (8) and integrating, we get

$$y_{j+1} = y_j + \frac{h}{24}\left[9 f_{j+1} + 19 f_j - 5f_{j-1} + f_{j-2}\right] + \left(-\frac{19}{720}\right) h\nabla^4 f_j \qquad \ldots(9)$$

or $$y_{j+1} = y_j + \frac{h}{24}\left[9 y'_{j+1} + 19 y'_j - 5y'_{j-1} + y'_{j-2}\right] + \left(-\frac{19}{720}\right) h\nabla^4 y'_j \qquad \ldots(10)$$

Formula given in (9) or (10) is called the Adams-Moultan corrector formula.

Note that, fourth order difference term in both the predictor and corrector formula is truncation error implying that $f(t,y)$ has been interpolated by third degree interpolating polynomial. This means we require information at past four points. Thus neglecting the truncation error, the Adams-Moultan predictor-corrector formulae read

Predictor : $$y_{j+1}^{(p)} = y_j + \frac{h}{24}\left[55 y'_j - 59 y'_{j-1} + 37 y'_{j-2} - 9 y'_{j-3}\right]$$

Corrector : $$y_{j+1}^{(c)} = y_j + \frac{h}{24}\left[9 y'_{j+1} + 19 y'_j - 5 y'_{j-1} + y'_{j-2}\right]$$

Example 10.4 Evaluate $y(1.5)$ by Adams-Bashfourth method of order four, given that

$$\frac{dy}{dt} = t^2(1+y)$$

$$y(1.1) = 1.233,\ y(1.2) = 1.548,\ y(1.3) = 1.979$$

$$y(1.4) = 2.575$$

Solution : The Adams-Bashfourth method of order four is given by

$$y_{j+1} = y_j + \frac{h}{24}\left[55y'_j - 59y'_{j-1} + 37\,y'_{j-2} - 9\,y'_{j-3}\right] \qquad ...(1)$$

Let the mesh points are

$$t_0 = 1.1,\ t_1 = 1.2,\ t_2 = 1.3,\ t_3 = 1.4,\ t_4 = 1.5$$

and $y_j = y(t_j)$

Then to compute $y_4 = y(t_4) = y(1.5)$.

Further, let us assume that

$$\frac{dy}{dt} = f(t,y) \qquad\qquad y(t)$$

Then from (1), we have on putting $j = 3$.

$$y_4 = y_3 + \frac{h}{24}\left[55f_3 - 59f_2 + 37f_1 - 9f_0\right]$$

We compute

$$f_3 = t_3(1+y_3) = (1.4)^2[1+2.575] = 7.007$$

$$f_2 = t_2(1+y_2) = (1.3)^2[1+1.979] = 5.03451$$

$$f_1 = t_1(1+y_1) = (1.2)^2[1+1.548] = 3.66912$$

$$f_0 = t_0(1+y_0) = (1.1)^2[1+1.233] = 2.70193$$

Thus, $y_4 = 2.575 + \frac{0.1}{24}\left[55(7.007) - 59(5.03451) + 37(3.66912) - 9(2.70193)\right]$

$$= 2.575 + \frac{0.1}{24}[199.78898] = 3.407454$$

Thus, $y_4 = y(1.5) = 3.4074540$ is the predicted value of y at $t = 1.5$, Now, this value will be improved by corrector formula given by

$$y_4^{(c)} = y_3 + \frac{h}{24}\left[9f_4 + 19f_3 - 5f_2 + f_1\right]$$

Now $f_4 = t_4^2\left(1+y_4\right) = \left(1.5\right)^2\left[1+3.407454\right]$

$$= 9.916771$$

Thus, $y_4^{(c)} = y^{(c)}\left(1.5\right) = 2.575 + \frac{0.1}{24}\left[9\left(9.916771\right) + 19\left(7.007\right) - 5\left(5.03451\right) + \left(3.66912\right)\right]$

$$= 3.412002$$

Verification :

The exact solution to the given IVP is

$$y\left(t\right) = \left(1.432869\right)e^{t^3/3} - 1 \qquad \text{[using initial condition } y\left(1.1\right) = 1.233\text{]}$$

Now $y\left(1.5\right) = 3.413547$

Thus we see that exact solution and numerical solution do match upto two decimal places.

Example 10.5 Use Adams-Moultan Predictor corrector formula to compute $y\left(0.4\right)$, given that

$$\frac{dy}{dt} = ty$$

$y\left(0\right) = 1$, $y\left(0.1\right) = 1.01$, $y\left(0.2\right) = 1.022$, $y\left(0.3\right) = 1.023$

Solution : Given that

$$y' = ty$$

Here $t_0 = 0$, $t_1 = 0.1$, $t_2 = 0.2$, $t_3 = 0.3$, $t_4 = 0.4$

and the step size $h = 0.1$

Predictor-corrector pair of Adams-Moulton method is

$$y_{j+1}^{(p)} = y_j + \frac{h}{24}\left[55y'_j - 59y'_{j-1} + 37y'_{j-2} - 9y'_{j-3}\right]$$

$$y_{j+1}^{(c)} = y_j + \frac{h}{24}\left[9y'_{j+1} + 19y'_j - 5y'_{j-1} + y'_{j-2}\right]$$

Now, using the predictor formula for the values

$$y_0 = y\left(0\right) = 1,\ y_1 = y\left(0.1\right) = 1.01,\ y_2 = y\left(0.2\right) = 1.022,\ y_3 = y\left(0.3\right) = 1.023$$

we obtain

$$y_4^{(p)} = y_3 + \frac{h}{24}\left[55y'_3 - 59y'_2 + 37y'_1 - 9y'_0\right]$$

$$= 1.023 + \frac{(0.1)}{24}\left[55(0.3)(1.023) - 59(0.2)(1.022) + 37(0.1)(1.01) - 9(0)(1)\right]$$

$$= 1.023 + \frac{0.1}{24}[8.5569]$$

$$= 1.058653$$

This prediced value is corrected by the corrector forumula

$$y_4^{(c)} = y_3 + \frac{h}{24}\left[9y_4' + 19y_3' - 5y_2' + y_1'\right]$$

Now $y_4' = t_4 y_4 = (0.4)(1.058653)$

$$= 0.4234612$$

$$y_4^{(c)} = 1.023 + \frac{0.1}{24}\left[9(0.4234612) + 19(0.3069) - 5(0.2044) + (0.101)\right]$$

$$= 1.05933$$

Verification :

The exact solution to given IVP subject to condition $y(0) = 1$ is

$$y(t) = e^{t^2/2}$$

Thus $y(0.4) = e^{(0.4)^2/2} = 1.08328$

Thus numerical solution do match to the exact solution upto one decimal place only.

Self-Learning Exercise

1. Compute $y(0.4)$ by Milne's method, given that

$$\frac{dy}{dt} = 2e^t - y$$

$y(0) = 2$, $y(0.1) = 2.01$, $y(0.2) = 2.04$, $y(0.3) = 2.09$

2. Use Runge-Kutta method of order four to compute $y(1.5)$, where given that

$$\frac{dy}{dt} = \frac{1}{t+y}, \quad y(0) = 1$$

3. Solve the following initial value problem by Taylor's series method,

$$\frac{dy}{dt} = -y - 2t, \quad t \in [0, 0.2]$$

(take step size $h = 0.1$)

10.4 Stability Analysis

Recall that numerical solution of differential equation is obtained in steps at different points of the solution space. For this, the solution space $\left[t_0, b\right]$ is partitioned into number of mesh points given as

$$t_0 < t_1 < t_2 \ldots\ldots. t_n = b$$

In general, spacing between these points is considered to be uniform and is given by

$$h = t_{j+1} - t_j, \quad j = 1,2,\ldots.n$$

In numerical solution of the differential equation $\frac{dy}{dt} = f(t, y)$, constituting on inital value problem, we determine $y(t_j)$ or y_j which provides numerical solution of $y(t)$ at $t = t_j$. The exactness of y_j may differ from method to method as per the prescribed accuracy and the accuracy of the method itself. Infact, every numerical method may involve some errors and obviously one is curious to contain it so as to obtain desired accurate results.

Local truncation error and convergence :

If $y(t_j)$ at $t = t_j$ denote exact value of $y(t)$ at $t = t_j$ and y_j is numerically computed value of $y(t)$ at $t = t_j$ by employed numerical method, then the local truncation T_j is given by

$$T_j = y(t_j) - y_j, \qquad j = 1,2,\ldots n$$

Note that, a numerical method is a recursive scheme, wherein previously computed value y_j is used to computed y_{j+1}. Recall that every solution y_j, $j = 1,2,\ldots n$ contains errors and the error of previous step propagates to next step too.

Convergence :

The error in solutions, y_j, $j = 1,2,\ldots.n$ may be cuntailed by taking small step size i.e. by taking more mesh points. A method is said to be convergent if step size is decreased, the numerical solution converges to the exact solution in the absence of any round off errors.

Stabiltiy :

Stability of a method is a vital concept. A method is said to be stable if the total effect of all errors (including round off erros) is bounded and is independent of the number of mesh points. A method may be **absolutely stable** and **relatively stable.**

To understand this, we consider a test equation

$$\frac{dy}{dt} = \lambda y, \; t \in [t_0, b]$$

$$y(t_0) = y_0 \qquad \ldots(1)$$

The differential equation given by (1) has the exact solution

$$y(t) = y_0\, e^{\lambda(t-t_0)} \qquad ...(2)$$

We now establish a relationship between solutions at $t = t_j$, $t = t_{j+1}$

Now $$y(t_j) = y_0\, e^{\lambda(t_j - t_0)} \qquad ...(3)$$

$$y(t_{j+1}) = y_0\, e^{\lambda(t_{j+1} - t_0)} \qquad ...(4)$$

From (3) and (4), we get

$$y(t_{j+1}) = y(t_j)\, e^{\lambda h} \qquad [\text{Note } h = t_{j+1} - t_j] \qquad ...(5)$$

Equation (5) provides exact solution.

Now, we can devise numerical method by having approximation to $e^{\lambda h}$ as

$e^{\lambda h} = 1 + \lambda h$ First order

$e^{\lambda h} = 1 + \lambda h + \dfrac{\lambda^2 h^2}{2}$ Second order

Let $E(\lambda h)$ denote approximation to $e^{\lambda h}$, then the numerical method is written as

$$y_{j+1} = y_j\, E(\lambda h) \qquad ...(6)$$

Let $\in_j$ is errors given by

$$\in_j = y_j - y(t_j)$$

Then $\in_{j+1} = y_{j+1} - y(t_{j+1})$

$$= y_j\, E(\lambda h) - y(t_j) e^{\lambda h}$$

$$= E(\lambda h)\left[\in_j + y(t_j)\right] - y(t_j) e^{\lambda h}$$

$$\in_{j+1} = \left[E(\lambda h) - e^{\lambda h}\right] y(t_j) + \in_j E(\lambda h) \qquad ...(7)$$

Note that in (7), the $\in_{j+1}$ accounts for the error at the $(j+1)^{th}$ step. From the very equation, we can see that the first term on the right hand side is the truncation errors and the second term is the propagation error. The propagation error is the most cause of worry since it is beyond one control and infact depends on the equation given.

Similarly,

$$\in_{j+2} = \left[E^2(\lambda h) - e^{2\lambda h}\right] y(t_j) + E^2(\lambda h) \in_j$$

$$\in_{j+l} = \left[E^l(\lambda h) - e^{l\lambda h}\right] y(t_j) + E^l(\lambda h) \in_j \qquad ...(8)$$

Thus we conclude that accurate results are obtained if the propagation error decays or bounded for this

$$\left|E(\lambda h)\right| < 1 \qquad \text{...(9)}$$

so that the second term in (8) has insignificant contribution in the numerical procedure.

Absolute Stable Method :

If the criterion given in (9) is met, then the method is called absolutely stable.

Relatively Stable Method :

If $E(\lambda h) < e^{\lambda h}$, then the method is called relatively stable.

Stability Analysis of Single Step Methods :

(i) Euler Method :

Consider the test equation

$$\frac{dy}{dt} = \lambda y, \qquad y(t_0) = y_0$$

The Euler method to the above test equation reads

$$y_{j+1} = y_j + h f(t_j, y_j)$$

$$= y_j + h(\lambda y_j) = y_j [1 + \lambda h]$$

$$= y_j E(\lambda h), \qquad \text{where } E(\lambda h) = 1 + \lambda h$$

When,

$$\left|E(\lambda h)\right| = \left|1 + \lambda h\right| < 1$$

This gives

$$-1 < 1 + \lambda h < 1$$

or $\qquad \lambda h \in (-2, 0)$

(ii) Runge-Kutta Method of Order Two

The Runge-Kutta method of order two is given as

$$y_{j+1} = y_j + \left[1 - \frac{1}{2a}\right] k_1 + \frac{1}{2a} k_2$$

where $\quad k_1 = h f(t_j, y_j)$

$$k_2 = h f(t_j + ah, y_j + ak_1)$$

on applying the above method to the test equation $\frac{dy}{dt} = \lambda y$, we find

$$k_1 = h\lambda y_j, \qquad k_2 = \lambda h\left(y_j + ak_1\right)$$

$$= \lambda h\left[1 + a h \lambda\right] y_j$$

Using these value of k_1, k_2 in the Runge-Kutta formula, we obtain

$$y_{j+1} = y_j + \left[1 - \frac{1}{2a}\right] h\lambda y_j + \frac{\lambda h y_j}{2a}\left[1 + a h\lambda\right]$$

$$= y_j\left[1 + h\lambda\left(1 - \frac{1}{2a}\right) + \frac{h\lambda}{2a}\left(1 + a h\lambda\right)\right]$$

$$= y_j\left[1 + h\lambda + \frac{h^2\lambda^2}{2}\right]$$

$$= y_j E\left(\lambda h\right), \qquad \text{where} \quad E\left(\lambda h\right) = 1 + h\lambda + \frac{h^2\lambda^2}{2}$$

The above relation indicates that the propagation factor $E\left(\lambda h\right)$ does not depend on the parameter *a*.

Thus the condition

$$\left|E\left(\lambda h\right)\right| < 1 \quad \text{gives}$$

$$\left|1 + h\lambda + \frac{\lambda^2 h^2}{2}\right| < 1$$

i.e. $\quad h\lambda \in (-2, 0)$

Runge-Kutta method of order four

The fourth order Runge-Kutta method is given as

$$y_{j+1} = y_j + w_1 k_1 + w_2 k_2 + w_3 k_3 + w_4 k_4$$

where $k_1 = h f\left(t_j, y_j\right)$

$$k_2 = h f\left(t_j + ah, y_j + d_{21} k_1\right)$$

$$k_3 = h f\left(t_j + bh, y_j + d_{31} k_1 + d_{32} k_2\right)$$

$$k_4 = h f\left(t_j + ch, y_j + d_{41} k_1 + d_{42} k_2 + d_{43} k_3\right)$$

on applying Runge-Kutta method to the test equation $\frac{dy}{dt} = \lambda y$, the above scheme becomes,

$$y_{j+1} = y_j + w_1 \lambda h y_j + w_2 \lambda h \left(1 + a \lambda h\right) + w_3 \lambda h \left(1 + b\lambda h + d_{32} a \lambda^2 h^2\right) y_j$$

$$+ w_4 \lambda h \left\{1 + c\lambda h + \left(d_{42} a + d_{43} b\right) \lambda^2 h^2 + d_{43} d_{32} a h^3 \lambda^3\right\} y_j$$

On simplification and using the parameter values obtained earlier, we get

$$y_{j+1} = \left[1 + \lambda h + \frac{\lambda^2 h^2}{2} + \frac{\lambda^3 h^3}{6} + \frac{\lambda^4 h^4}{24}\right] y_j$$

or $\quad y_{j+1} = E\left(\lambda h\right) y_j$

This shows that the propagation factor $E\left(\lambda h\right)$ is free from the abitrary parameters.

The condition for absolute stability

$$\left|E\left(\lambda h\right)\right| < 1$$

i.e. $\quad \left|1 + h\lambda + \frac{h^2 \lambda^2}{2} + \frac{h^3 \lambda^3}{6} + \frac{h^4 \lambda^4}{24}\right| < 1$

gives $\quad \lambda h \in \left(-2.78,\ 0\right)$

10.5 Summary

In this unit Predictor-corrector strategies for solving differential equation numerically were discussed. These strategies are mutlistep methods which require information at more than one preceeding points to predict the value of dependent variable at the next mesh point. This unit discussed Milne's method and Adams-Moulton method which have distinct predictor-corrector formulae based on the forward and backward intrpolation formulae. Further, the stability of the different numerical methods were also discussed.

10.6 Answers of Self-Learning Exercise

1. $y^{(p)}\left(0.4\right) = 2.1623, \; y^{(c)}\left(0.4\right) = 2.162$

2. $y\left(1.5\right) = 1.7555$

3. $y\left(0.1\right) = -0.94145125, \; y\left(0.2\right) = -0.8561904$

10.7 Exercises

1. Compute $y\left(1\right)$ by Adams-Moulton method, given that

$$\frac{dy}{dt} = y - t^2, \quad y\left(0\right) = 1$$

$$y(0.2) = 1.2859, \quad y(0.4) = 1.46813, \; y(0.8) = 1.73779$$

[Ans. $\overset{(p)}{y}(0.8) = 2.01451$, $\quad \overset{(c)}{y}(0.8) = 2.01434$

$\overset{(p)}{y}(1) = 2.28178 \qquad \overset{(c)}{y}(1) = 2.28393$]

2. Compute $y(0.5)$ by Adams-Moulton method, given that

$$\frac{dy}{dt} = y - \frac{2t}{y}$$

$$y(0) = 1, \quad y(0.1) = 1.0954$$

$$y(0.2) = 1.1832, \quad y(0.3) = 1.2649$$

[Ans. $y^{(p)}(0.4) = 1.3415$, $\; y^{(c)}(0.4) = 1.3416$

$y^{(p)}(0.5) = 1.4150 \quad y^{(c)}(0.5) = 1.4142$]

3. Use Taylor's series method to compute x and y at $t = 0.05$, where x and y are the solutions of the following system of differential equations

$$\frac{dx}{dt} = xy + t$$

$$\frac{dy}{dt} = yt + x$$

given that $x(0) = 2$, $y(0) = -3$

[Ans. Using four term Taylor's series, we obtain $x(0.05) = 1.7264$, $y(0.05) = -2.91068$]

□□□

Unit - 11 : Boundary Value Problem - I

Structure of the Unit

11.0 Objectives

After studying this unit you will be able to

1. Understand the notion of a boundary value problem and its relevance
2. Learn the idea of shooting method to solve boundary valur problems.

11.1 Introduction

Boundary value problems are the problems where the conditions are prescribed at the end points of the solution space. You may recall that in an initial value problem, the conditions are prescribed at the initial point of the solution space. A higher order differential equation may give rise to a boundary value problem. For example, an nth order differential equation's solution involves n arbitrary constants which require n conditions. If these conditions are prescribed at the end points t_0, b of the solution space $[t_0, b]$ then the differential equation, together with the boundary conditions, constitute a boundary value problem. Many phenomenon cutting across different disciplines are described mathematically by boundary value problems whose analytical solutions may not be possible, hence numerical solutions are resorted to. The finite difference method and the shooting method are amongest the methods which are frequently used to solve boundary value problems. This unit is aimed to focus on shooting method.

11.2 Boundary Value Problems

A boundary value problem as well as its boundary conditions are classified as homogeneous or inhomogeneous.

A homogeneous boundary value problem is that which involves a homogeneous differential equation (i.e. the equation that contains dependent variable and its derivative) and the homogeneous boundary conditions (i.e. $\alpha = \beta = 0$). A homogeneous BVP has trivial solution $y(x) = 0$. Thus, such BVP is not of fundamental importance for our present study.

A BVP which is not homogeneous i.e. inhomogeneous is our focus of study. Hence we define,

Eigenvalue Problem :

A BVP which involves a parameter μ (say) in the differential equations or in the boundary conditions is called an eigenvalue problem and the values μ take are called eigenvalues.

Solutions of Boundary Value Problem :

BVP given in (1) possesses a unique solution if it satisfies certain conditions. However we are not inclined to narrate them here. For the sake of our study, we presume that the solution exists uniquely.

A two point boundary value problem is the simplest case to understand the numerical solution methods.

Let us consider a two point boundary value problem,

$$\frac{d^2y}{dx^2} = f(x, y, y'), \qquad x \in [a, b] \qquad \text{...(1)}$$

Obviously the general solution of (1) will involve two arbitrary constants which require two end conditions for their determination. These boundary conditions may take the following forms :

(A) Boundary Condition of First Kind :

$$y(a) = \alpha, \qquad y(b) = \beta \qquad \text{...(2)}$$

i.e. the values of the dependent variable y are prescribed at the end points a and b.

(B) Boundary Condition of Second Kind :

$$y'(a) = \alpha, \qquad y'(b) = \beta \qquad \text{...(3)}$$

i.e. the values of the derivatives are prescribed at the end points.

(C) Boundary Condition of the Mixed Kind :

This condition involves the dependent variable y and its derivatives y' such as

$$a_1 y(a) - a_2 y'(a) = \alpha \qquad \text{...(4)}$$

$$a_3 y(b) + a_4 y'(b) = \beta \qquad \text{...(5)}$$

Where the constants a_1, a_2, a_3, a_4 are such that

$$a_1 a_2 \geq 0, \qquad |a_1| + |a_2| \neq 0$$

$$a_3 a_4 \geq 0, \qquad |a_3| + |a_4| \neq 0, \quad |a_1| + |a_3| \neq 0$$

11.3 Shooting Method

Shooting method is used to solve a boundary value problem where in the given BVP is converted into a system of initial value problems. This method requires a systematic guessing of unknown quantities at one end of the solution space such that the condition at the other end is satisfied. The guessing of these unknown quantities is done purely on hit and trial basis. To illustrate the method let us consider the BVP

$$\frac{d^2y}{dx^2} = f(x, y), \qquad x \in [a, b] \qquad \text{...(1)}$$

$$y(a) = \alpha, \qquad y(b) = \beta$$

Let $\dfrac{dy}{dx} = z$ then $\dfrac{d^2y}{dx^2} = \dfrac{dz}{dx}$

Thus the given BVP can be written as

$$\left.\begin{array}{l} \dfrac{dz}{dx} = f(x,y) = F(x,y,z) \\ \dfrac{dy}{dx} = z = G(x,y,z) \\ y(a) = \alpha,\ z(a) = y'(a) = ? \end{array}\right\} \qquad ...(2)$$

Equation (2) constitutes a system of initial value problems and can be solved if we choose proper choice for $y'(a)$ such that the condition $y(b) = \beta$ is satisfied.

The proper values of $y'(a)$ is obtained by testing different values.

In order to solve (2), we require $y'(a)$. Let us assume that the exact value of $y'(a)$ is m. Let m_1 and m_2 be two initial guesses for m. For these m_1 and m_2, we determine the value of y at $x = b$ from the system (2). Let the values of y at $x = b$ corresponding to m_1 and m_2 are deonted as $y(m_1;b)$ and $y(m_2;b)$ respectively. Then using linear interpolation, the better choice m_3 for m is given by

$$\frac{m_3 - m_1}{y(b) - y(m_1;b)} = \frac{m_2 - m_1}{y(m_2;b) - y(m_1;b)}$$

on solving

$$m_3 = m_1 + \frac{y(b) - y(m_1;b)}{y(m_1;b) - y(m_1;b)}(m_2 - m_1)$$

Thus the system of IVP's given by (2) is solved by considering $y'(a) = m_3$ for this m_3, we compute y at $x = b$. If this is in good agreement with end condition $y = \beta$ at $x = b$, then the procedure is finished otherwise we seek better choice for m by using m_2, m_3 as discussed above and so forth until the end condition $y(b) = \beta$ is satisfied.

Though the method seems to be easy to apply, but the speed of convergence very much depends on the "good guessings" of the unknown quantity. Further more, it should be noted that the method is not that much handy while applying for higher order boundary value problem or BVP's involving higher non-linear differential equation simply because these may be quite sensitive to the choices of initial guesess. Besides this, it is pertinent to record that manual working of shooting method is very tedious in view of guessing involved. However, the illustrations given below would help keep into the real intricacies of the procedure.

Example 11.1 Solve the boundary value problem

$$\frac{d^2y}{dx^2} = y$$

$$y(0) = 0, \qquad y(1) = 1.2$$

by employing shooting method, take $y'(0) = 0.85,\ 0.95$ as initial guesses.

Solution : Taylor series method for the given

$$y(x) = \left[x + \frac{x^3}{6} + \frac{x^5}{120} + \frac{x^7}{5040} + \frac{x^9}{362800} + \ldots\right] y'(0) \qquad \ldots(1)$$

when $y'(0) = 0.85 = m_1$ (say)

Then $y(m_1;1) = \left[1 + \frac{1}{6} + \frac{1}{120} + \frac{1}{5040} + \frac{1}{362800}\right](0.85)$

$$= (1.1752)\ (0.85) = 0.99892$$

Similarly

$$y(m_2;1) = (1.1752)(0.95)$$

$$= 1.11644$$

Thus better approximation m_3 for $y'(0)$ is obtained as

$$m_3 = m_1 + (m_2 - m_1)\left[\frac{y(1) - y(m_1;1)}{y(m_2;1) - y(m_1;1)}\right]$$

$$= 0.85 + 0.1\left[\frac{1.2 - 0.99892}{1.11644 - 0.98892}\right]$$

$$= 0.85 + 0.1\left[\frac{0.20108}{0.11752}\right]$$

$$= 0.85 + 0.15 = 1.0211$$

Thus, $y'(0) = m_3 = 1.0211$

Then, $y(m_3;1) = (1.1752)(1.0211)$

$$= 1.1999967$$

$$= 1.2$$

Thus we see how $y'(0)$ have been improvised so that the end condition is satisfied. Once the unknown quantity is properly guessed then the solution in the solution space with proper step size may be obtained.

Example 11.2 Solve the BVP

$$\frac{d^2y}{dt^2} = y, \quad y(0) = 0, \quad y(1) = 1.1752$$

by shooting method together with Runge-Kutta method.

Solution : Let $y' = z$ then $y'' = z'$

Thus the given BVP reduces to

$$\left.\begin{aligned} &\frac{dz}{dt} = y = F(t,y,z) \quad (say) \\ &\frac{dy}{dt} = z = G(t,y,z) \quad (say) \\ &y(0) = 0,\ z(0) = y'(0) = ? \end{aligned}\right\} \qquad ...(1)$$

We have to find $y'(0)$ such that (1) gives $y(1) = 1.1752$

Let $y'(0) = m$ and m_1, m_2 be two initial guesses for m.

Let $\quad m_1 = 0.9 = z_0 = z(0)$

Then the fourth order Runge-Kutta scheme for the system of initial value problems given by (1) is

$$z_{j+1} = z_j + \frac{1}{6}\left[k_1 + 2k_2 + 2k_3 + k_4\right]$$

$$y_{j+1} = y_j + \frac{1}{6}\left[l_1 + 2l_2 + 2l_3 + l_4\right]$$

where

$$k_1 = h\,F(t_0, y_0, z_0) \qquad l_1 = h\,G(t_0, y_0, z_0)$$

$$k_2 = h\,F\left(t_0 + \frac{h}{2},\ y_0 + \frac{l_1}{2},\ z_0 + \frac{k_1}{2}\right) \qquad l_2 = h\,G\left(t_0 + \frac{h}{2},\ y_0 + \frac{l_1}{2},\ z_0 + \frac{k_1}{2}\right)$$

$$k_3 = h\,F\left(t_0 + \frac{h}{2},\ y_0 + \frac{l_2}{2},\ z_0 + \frac{k_2}{2}\right) \qquad l_3 = h\,G\left(t_0 + \frac{h}{2},\ y_0 + \frac{l_2}{2},\ z_0 + \frac{k_2}{2}\right)$$

$$k_4 = h\,F(t_0 + h,\ y_0 + l_3,\ z_0 + k_3) \qquad l_4 = h\,G(t_0 + h,\ y_0 + l_3,\ z_0 + k_3)$$

given that

$$t_0 = 0, \quad y_0 = y(0) = 0 \quad \text{and} \quad z_0 = z(0) = 0.9$$

Let $\quad h = 1$

Then the slopes $k's$ and $l's$ are obtained as

$$k_1 = h\,F\left(t_0, y_0, z_0\right) = 1(0) = 0$$

$$l_1 = h\,G\left(t_0, y_0, z_0\right) = 1(0.9) = 0.9$$

$$k_2 = h\,F\left(t_0 + \frac{h}{2},\, y_0 + \frac{l_1}{2},\, z_0 + \frac{k_1}{2}\right) = 0.45$$

$$l_2 = h\,G\left(t_0 + \frac{h}{2},\, y_0 + \frac{l_1}{2},\, z_0 + \frac{k_1}{2}\right) = 0.9$$

similarly,

$$k_3 = 0.45$$

$$l_3 = 1.125$$

$$k_4 = 1.125$$

$$l_4 = 1.35$$

Thus, $y(m_1;1) = y(1) = y_0 + \frac{1}{6}\left[l_1 + 2l_2 + 2l_3 + l_4\right]$

$$= 0 + \frac{1}{6}\left[0.9 + 2(0.9 + 1.125) + 1.35\right]$$

$$= 1.05$$

Let $m_2 = 1$ be another guess for $z(0) = y'(0)$

Then, on repeating the above procedure for $m_2 = 1,\ t_0 = 0,\ y_0 = 0,\ h = 1$, we get

$$k_1 = 0,\ l_1 = 1,\ k_2 = 0.5,\ l_2 = 1,\ k_3 = 0.5,\ l_3 = 1.25,\ k_4 = 1.25,\ l_4 = 1.5$$

Thus, we get

$$y(m_2;1) = y(1) = 0 + \frac{1}{6}\left[1 + 2(1 + 1.25) + 1.5\right]$$

$$= 1.1666$$

with these two slopes $m_1,\ m_2$ we have computed the values of y at $x = 1$ which are not far from the given end condition $y(1) = 1.1752$

Let m_3 be better slope, then

$$m_3 = m_1 + (m_2 - m_1)\left[\frac{y(1) - y(m_1;1)}{y(m_2;1) - y(m_1;1)}\right]$$

$$= 0.9 + (0.1)\left[\frac{1.1752 - 1.05}{1.1666 - 1.05}\right]$$

$= 1.0073756$

Thus if we take $m_3 = y'(0) = 1.0073756$ and solve (1), then we will find

$y(1) = 1.175271$

which is correct upto four decimal to the given end condition. The above working explains the procedure how missing guesses are made. However, the above question was solved with $h = 1$. But in practice we require solution in the solution space ($[0,\ 1]$ in the present question) for which we partition the interval by specifying small step size. That is to say, the above question requires value of y at internal point of $[0,\ 1]$. Hence, if we take $h = 0.2$, the shooting method would be requiring lot of computations and if it is done manually, then certainly it is quite time consuming and tedious.

11.4 Summary

In this unit, you came across with the idea of boundary value problems, their genesis and types. This unit explained shooting method which provides solution to BVP by converting it into a system of IVP's where in the estimates for unknown quantities at one end is assumed in such a manner that the condition at the other end is satisfied.

11.5 Exercises

1. Solve the boundary value problem

$$\frac{d^2y}{dx^2} = y$$

$$y(0) = 0, \quad y(0.6) = 0.7$$

by shooting method [Use Taylor's series method]

2. Solve the boundary value problem

$$\frac{d^2y}{dx^2} = y$$

$$y(0) = 0, \; y(0.4) = 0.4$$

by shooting method [Use Runge-Kutta method]

3. Solve the boundary value problem

$$\frac{d^2y}{dx^2} = 64y - 10$$

$y(0) = 0, \qquad y(1) = 0, \qquad$ by shooting method.

□□□

Unit - 12 : Boundary Value Problem - II

Structure of the Unit

12.0 Objectives

After reading this unit you will be able to

1. Understand the notion of finite difference methods to solve boundary value problem.
2. Work out various types of boundary value problem.

12.1 Introduction

In the last unit you studied shooting method to solve boundary value problems $(BVPs)$. This unit discusses another method "finite difference methods" to solve $BVPs$. These methods are based on the idea of converting differential equation giving rise to BVP into system of difference equations. For this, various order derivatives are replaced by their difference approximations. Similarly, other variables appearing in the differential equation are also discretized in the solution space. This all leads to linear or non-linear system of difference equation together with end conditions which can be solved by previously discussed methods. The advantage of finit difference method over shooting method is that no "guessing" is required as we do in shooting method. Finite difference methods are useful in solving complex differential equations and infact serve as a basis to many other advanced numerical methods. For example, these methods are founding stone of computational fluid dynamics which itself is a fundamental tool of industrial engineering.

Finite Difference Methods :

In finite difference methods, the differential equation is approximated by corresponding difference approximations. The solution is obtained by the resultant difference equations at each mesh point. This system of difference equations may give rise to system of linear or non-linear equations which can be solved by the previously discussed methods.

We now illustrate the various difference approximations to the derivatives.

By Taylor's series expansion, we have

$$y\left(x_j+h\right)=y\left(x_j\right)+h\,y'\left(x_j\right)+\frac{h^2}{2}y''\left(x_j\right)+\frac{h^3}{6}y'''\left(x_j\right)+....$$

or $$y_{j+1}=y_j+h\,y_j'+\frac{h^2}{2}\,y_j''+\frac{h^3}{6}\,y_j'''+... \qquad ...(1)$$

[Note that $x_{j+1}=x_j+h$ and $y\left(x_j+1\right)=y_{j+1}$ etc]

Similarly

$$y\left(x_j-h\right)=y\left(x_j\right)-h\,y'\left(x_j\right)+\frac{h^2}{2}y''\left(x_j\right)-\frac{h^3}{6}y'''\left(x_j\right)+....$$

or $$y_{j-1}=y_j-h\,y_j'+\frac{h^2}{2}\,y_j''-\frac{h^3}{6}\,y_j'''+... \qquad ...(2)$$

From (1), we have

$$\frac{y_{j+1}-y_j}{h}=y_j'+0(h) \qquad ...(3)$$

Here $0(h)$ indicates the terms which contain h or its higher degree.

From (2), we have

$$\frac{y_j-y_{j-1}}{h}=y_j'+0(h) \qquad ...(4)$$

Again from (1) and (2),

$$\frac{y_{j+1}-y_{j-1}}{2h}=y_j'+0\left(h^2\right) \qquad ...(5)$$

Thus we see that finite difference approximations to y' at $x=x_j$ in (3), (4) are of $0(h)$ and the aproximation given is (5) is better as it is of $0\left(h^2\right)$ accuracy.

Again, from (1) and (2), we have

$$\frac{y_{j+1}-2y_j+y_{j-1}}{h^2}=y_j''+0\left(h^2\right) \qquad ...(6)$$

Thus, we see that

$y_j'=\dfrac{y_{j+1}-y_j}{h}$ (forward difference approximations with truncation error of $0(h)$)

$y_j'=\dfrac{y_j-y_{j-1}}{h}$ (Backward difference approximation with truncation error of $0(h)$)

$$y'_j = \frac{y_{j+1} - y_{j-1}}{2h}$$ (Central difference approximations with truncation error of $0(h^2)$)

$$y''_j = \frac{y_{j+1} - 2y_j + y_{j-1}}{h^2}$$ (forward difference approximation with truncation errror of $0(h^2)$)

Note that when the differential equation is replaced by difference equation then each differnece approximation chosen is to be of same order (i.e. same order of truncation error) with $y_0 = \alpha$, $y_n = \beta$

[Note that the above approximation contains the term y_{j+1} and the end condition $y_n = \beta$. Thus the range for j can not exceed $n-1$, otherwise the system would exceed the solution space. That is the range for j must be decided in view of the difference approximaition and the boundary conditions.]

The above equation together with conditions $y_0 = \alpha$, $y_n = \beta$ would give rise to $(n-1)$ equations in $(n-1)$ unknowns $y_1, y_2, ..., y_{n-1}$. This system can be solved by any previously discussed method.

The local truncation error of this approximation is of $0(h^2)$ i.e., the finite difference scheme has second-order accuracy and the method can be termed as second-order method.

We now illustrate the finite difference methods for various types of BVPs.

12.2.1 Type $y'' = f(x, y)$:

Let us consider BVP

$$y'' = f(x, y),\ x \in [a, b]$$

$$y(a) = A,\ y(b) = B \qquad ....(1)$$

We can have second order and fourth order methods for the BVP given by (1). Let the solution space $[a, b]$ is discretized by the grid points

$$x_j = a + jh,\ j = 1, 2, ..., N+1$$

where $x_0 = a$, $x_{N+1} = b$

and $h = \frac{b-a}{N+1}$

Let y_j denote the approximate value of y at x_j.

Second Order Method :

The derivative y'' at x_j is approximated as

$$y''(x_j) = \frac{y(x_{j+1}) - 2y(x_j) + y(x_j - 1)}{h^2}$$

or $$y_j'' = \frac{y_{j+1} - 2y_j + y_{j-1}}{h^2} \qquad \dots(2)$$

Thus the given BVP can be written as the following system of difference equations:

$$y_{j+1} - 2y_j + y_{j-1} = h^2 f\left(x_j, y_j\right), \quad j = 1,2,\dots,N$$

$$y_0 = a\,, \qquad y_{N+1} = b \qquad \dots(3)$$

Note that j runs upto N and not upto $N+1$. When j runs upto N then y_{j+1} appearing in (3) amounts to y_{N+1} (which is the end condition). If we taken j running upto N, then we would have, y_{N+2}, which is out of the solution space. This understanding must be taken care of.

System (3) constitutes a difference equation for every $1 \le j \le N$ and with y_0 and y_{N+1} known the system of N equation with N variables $y_1, y_2, \dots, y_N$ can be solved. Thus value of y_j, $j = 1,2,\dots,N$ computed provide y at mesh point x_j.

Fourth Order Method :

A fourth order method (Numerov method) can also be written down for the BVP given by (1). We have Numerov method as under

$$y_{j+1} - 2y_j + y_{j+1} = \frac{h^2}{12}\left[f_{j+1} + 10 f_j + f_{j-1}\right]$$

where $f_j = f\left(x_j, y_j\right)$ etc. $j = 1,2,\dots,N$

with end conditions. $y_0 = A$, $y_{N+1} = B$ ...(4)

Derivative Boundary Conditions for $y'' = f(x,y)$:

We now discuss the case where BVP of the type $y'' = f(x,y)$ involves derivative boundary conditions.

consider,

$$y'' = f(x,y)$$

$$\alpha\, y(a) - \beta\, y'(a) = A$$

$$\alpha_1\, y(b) - \beta_1\, y'(b) = B \qquad \dots(5)$$

The difference approximation to $y'' = f(x,y)$ is taken of $0\left(h^2\right)$ as given in (3). Which have $N+2$ unknowns $y_0, y_1, y_2, \dots, y_{N+1}$ in N equations. In order to solve the system we require two more equations in view of (5).

We now approximate y' by central difference approximation (which is of $0\left(h^2\right)$. Thus, from the boundary condition given in (5), we have at $\quad x = x_0 \qquad \alpha\, y_0 - \beta\left(\dfrac{y_1 - y_{-1}}{2h}\right) = A$

This gives, $\quad y_{-1} = \dfrac{2h}{\beta} A - \dfrac{2h\alpha}{\beta} y_0 + y_1 \qquad$ ...(6)

and at $x = x_{N+1}$, $\quad \alpha_1\, y_{N+1} + \beta_1\left[\dfrac{y_{N+2} - y_N}{2h}\right] = B$

This gives $\quad y_{N+2} = \dfrac{2hB}{\beta_1} - \dfrac{2h\alpha_1}{\beta_1} + y_N \qquad$ ...(7)

Note that y_{-1}, y_{N+2} are the values of $y(x)$ at x_{-1} and x_{N+2}. Since x_{-1}, x_{N+2} don't belong to the discretization of $[a,b]$ and infact lie outside of $[a,b]$, therefore these are fictitious nodes. The values of y_{-1}, y_{N+2} can be determined by taking $j = 0$, $j = N+1$ in the equation (3)

12.2.2 BVP of the type $y'' = f(x, y, y')$:

The BVP involving y' are also solved by applying difference approximation to y'. Here it should be noted that order of approximation to y' is same as that of y''. That is, if y'' is taken to be of $0(h^2)$ then y' is also of $0(h^2)$.

Let us consider

$$y'' = f(x, y, y'),$$

$$y(a) = A,\ y(b) = B$$

The second order approximation to the given differential equation is given by

$$\frac{y_{j-1} - 2y_j + y_{j+1}}{h^2} = f\left(x_j, y_j, \frac{y_{j+1} - y_{j-1}}{2h}\right)$$

12.2.3 BVP of $y^{iv} = f(x, y)$:

The second order finite difference approximation to the BVP of the type $y^{iv} = f(x, y)$ is given by

$$y_{j-2} - 4y_{j-1} + 6y_j - 4y_{j+1} + y_{j+2} = h^4 f(x_j, y_j)$$

Example 12.1 Solve the boundary value problem

$$\frac{d^2y}{dx^2} + (1 + x^2)y + 1 = 0,\ x \in [0,1].$$

by a second order finite difference method with step size $h = \dfrac{1}{4}$

Solution : We discretize the solution space $[0,\ 1]$ by taking mesh points

$$x_j = x_0 + x_j h\,,\ x_j = 0,\ 1,\ 2,\ldots,N\,,\ \text{where }\ Nh = 1$$

When $h = 0.25$, then $N = 4$ and the mesh point are

$$x_0 = 0\,,\ x_1 = 0.25\,,\ x_2 = 0.5\,,\ x_3 = 0.75\,,\ x_4 = 1$$

The second order method for the given BVP

$$\left(\frac{y_{j+1} - 2\,y_j + y_{j-1}}{h^2}\right) + \left(1 + x_j^2\right).\ y_j + 1 = 0$$

or $$-y_{j-1} + \left[2 - h^2\left(1 + x_j^2\right)\right] y_j - y_{j+1} = h^2\,, \qquad f = 1{,}2{,}3 \qquad \ldots(1)$$

together with boundary conditions

$$y_0 = y(0) = 0\,,\ y_4 = y(1) = 0 \qquad \ldots(2)$$

System (1) for $j = 1{,}2{,}3$ gives

$$-y_0 + \left[2 - \frac{1}{4^2}\left(1 + \left(\frac{1}{4}\right)\right)^2\right] y_1 - y_2 = \left(\frac{1}{4}\right)^2$$

$$-y_1 + \left[2 - \frac{1}{4^2}\left(1 + \left(\frac{1}{2}\right)^2\right)\right] y_2 - y_3 = \left(\frac{1}{4}\right)^2$$

$$-y_2 + \left[2 - \frac{1}{4^2}\left(1 + \left(\frac{3}{4}\right)^2\right)\right] y_3 - y_4 = \left(\frac{1}{4}\right)^2$$

and $y_0 = 0\,,\ y_4 = 0$

on simplification, we get

$$-y_0 + \left[\frac{495}{256}\right] y_1 - y_2 = \frac{1}{16}$$

$$-y_1 + \left[\frac{123}{64}\right] y_2 - y_3 = \frac{1}{16}$$

$$-y_2 + \left[\frac{487}{256}\right] y_3 - y_4 = \frac{1}{16}$$

with $y_0 = 0\,,\ y_4 = 0$

on solving the above system of linear equations, we obtain

$$y_1 = 0.10744093\,,\ y_2 = 0.14524711\,,\ y_3 = 0.1092058$$

Example 12.2 Solve the BVP by Numerov method

$$\frac{d^2y}{dx^2} = x + y$$

$$y(0) = 0, \; y(1) = 0$$

with step size $h = \frac{1}{4}$

Solution : The interval $[0, 1]$ is discretized as $x_j = x_0 + jh$, $j = 0, 1, 2, N$, where $Nh = 1$ when $h = \frac{1}{4}$. Then $N = 4$. Hence the mesh point are $x_0 = 0$, $x_1 = \frac{1}{4}$, $x_2 = \frac{1}{2}$, $x_3 = \frac{3}{4}$, $x_4 = 1$. The number scheme for the given differential equation is

$$y_{j-1} - 2y_j + y_{j+1} = \frac{h^2}{12}\left[\left(y_{j+1} + x_{j+1}\right) + 10\left(y_j + x_j\right) + \left(y_{j-1} + x_{j-1}\right)\right], \; j = 1, 2, 3$$

Putting $h = \frac{1}{4}$ and on simplification, the above scheme becomes,

$$y_{j-1} - \frac{394}{191}y_j + y_{j+1} = \frac{1}{191}\left[x_{j+1} + 10x_j + x_{j-1}\right], \; j = 1, 2, 3$$

Thus for $j = 1, 2, 3$ and $x_1 = \frac{1}{4}$, $x_2 = \frac{1}{2}$, $x_3 = \frac{3}{4}$,

we have following system of equations

$$y_0 - \frac{394}{191}y_1 + y_2 = \frac{3}{191}$$

$$y_1 - \frac{394}{191}y_2 + y_3 = \frac{6}{191}$$

$$y_2 - \frac{394}{191}y_3 + y_4 = \frac{9}{191}$$

with the conditions $y_0 = 0 = y_4$

which yields

$$y_1 = y(0.25) = -0.035048$$

$$y_2 = y(0.5) = -0.056591$$

$$y_3 = y(0.75) = -0.050276$$

Example 12.3 Solve the BVP

$$\frac{d^2y}{dx^2} = x\,y$$

$$y(0) + y'(0) = 1,\; y(1) = 1$$

with step size $h = \frac{1}{3}$.

Solution : The solution space $[0,\, 1]$ is discretized as $x_j = x_0 + j\,h,\; j = 0,\, 1,\, 2, \ldots, N$, where $N\,h = 1$ when $h = \frac{1}{3}$ then $N = 3$, hence the mesh points are

$$x_0 = 0,\; x_1 = \frac{1}{3},\; x_2 = \frac{2}{3},\; x_3 = 1$$

The second order method for the given differential equation is

$$y_{j-1} - 2\,y_j + y_{j+1} = h^2\left(x_j\, y_j\right),\; j = 0,\, 1,\, 2 \qquad \text{...(1)}$$

Note that the given BVP involves derivative boundary conditions. Since the method used is second order method therefore $y'(0)$ in boundary condition is replaced by its second order difference approximation i.e.,

$$y_j' = \frac{y_{j+1} - y_{j-1}}{2\,h}$$

We take $y'(0) = \dfrac{y_1 - y_{-1}}{2\,h}$

In view of the boundary conditions become

$$2\,y_0 + 3\,y_1 - 3\,y_{-1} = 2,\; y_3 = 1 \qquad \text{...(2)}$$

Thus for $j = 0,\, 1,\, 2,$ the system (1) yields

$$y_{-1} - 2\,y_0 + y_1 = 0$$

$$y_0 - 2\,y_1 + y_2 = \frac{1}{9}\left(\frac{1}{3}\right)y_1$$

$$y_1 - 2\,y_2 + y_3 = \frac{1}{9}\left(\frac{2}{3}\right)y_2 \qquad \text{...(3)}$$

Solving (3) in view of (2), we obtain

$$y_0 = y(0) = -\,0.987951$$

$$y_1 = y\left(\frac{1}{3}\right) = -\ 0.325301$$

$$y_2 = y\left(\frac{2}{3}\right) = 0.325301$$

Example 12.4 Solve the BVP

$$\frac{d^2y}{dx^2} = \frac{3}{2}\,y^2$$

$$y(0) = 4\,,\ y(1) = 1$$

with step size $h = \frac{1}{3}$, using second order method.

Solution : The interval $[0,\ 1]$ is discretized as

$$x_0 = \frac{1}{3}\,.\ x_1 = \frac{1}{3}\,,\ x_2 = \frac{2}{3}\,,\ x_3 = 1\,.$$

The second order finite difference approximation to the given differential equation is given by

$$y_{j-1} - 2\,y_j + y_{j+1} = h^2\left(\frac{3}{2}\,y_j^2\right), \qquad j = 1,2$$

for $j = 1,\ 2$, we have

$$\left.\begin{aligned} y_0 - 2\,y_1 + y_2 &= \frac{y_1^2}{6} \\ y_1 - 2\,y_2 + y_3 &= \frac{y_2^2}{6} \end{aligned}\right\} \qquad ...(1)$$

with conditions $y_0 = 4\,,\ y_3 = 1$

System (1) yields following non linear equations in y_1 and y_2

$$\left.\begin{aligned} y_1^2 + 12\,y_1 - 6\,y_2 - 24 = 0 \\ -6\,y_1 + y_2^2 + 12\,y_2 - 6 = 0 \end{aligned}\right\} \qquad ...(2)$$

Above system of non-linear equations can be solved by method of iteration, Newton-Raphson method. We have used Newton-Raphson method and obtained

$$y_1 = 2.29504\,.\ y_2 = 1.46794$$

after three iterations.

Example 12.5 Solve the BVP

$$y^{iv} = 2$$

$$y(0) = y'(0) = y(1) = y'(1) = 0$$

Solution : The interval $[0, 1]$ is discretized

as $\quad x_0 = 0,\ x_1 = \frac{1}{4},\ x_2 = \frac{1}{2},\ x_3 = \frac{3}{4},\ x_4 = 1$

on taking the step size $h = \frac{1}{4}$

The second order method for the given differential equation reads

$$y_{j-2} - 4y_{j-1} + 6y_j - 4y_{j+1} + y_{j+2} = h^4(2),\ j = 1, 2, 3$$

The end conditions are

$$y_0 = 0,\ y_{-1} = y_1,\ y_4 = 0,\ y_3 = y_5$$

Thus for $j = 1, 2, 3$ and using and conditions we obtain the following system of equations

$$7y_1 - 4y_2 + y_3 = \frac{2}{4^4}$$

$$-4y_1 + 6y_2 - 4y_3 = \frac{2}{4^4}$$

$$y_1 - 4y_2 + 7y_3 = \frac{2}{4^4}$$

on solving the above system of equations, we obtain

$$y_1 = y(0.25) = \frac{5}{8}\left(\frac{2}{4^4}\right),\ y_2 = y(0.5) = \left(\frac{2}{4^4}\right),\ y_3 = y(0.75) = \frac{5}{8}\left(\frac{2}{4^4}\right)$$

Self-Learning Exercise

1. Solve the following BVP by second order method

$$y'' = -y - 1$$

$y(0) = y(1) = 0$, take step size $h = 0.25$

2. Solve the following BVP by second order method

$$y'' = x + y$$

$y(0) = 0,\ y(1) = 0$, step size $h = 0.25$

12.3 Summary

In this unit solution of boundary value problem comprising differential equations by finite difference approximations have been explained. The finite difference methods explained. The finite difference methods replace the differential equation by its difference approximations which amount to linear or non-linear system of algebraic equations. which can be solved numercially.

12.4 Answers of Self-Learning Exercise

1. $y(0.25) = 0.10467$, $y(0.5) = 0.14031$, $y(0.75) = 0.10475$

2. $y(0.25) = -0.03488$, $y(0.5) = -0.05632$, $y(0.75) = -0.05003$

12.5 Exercises

1. Solve the BVP

$$y'' = -8\left(\sin^2 \pi x\right) y$$

$$y(0) = y(1) = 1$$

by second order method, with step size $h = \frac{1}{4}$

(Ans. $y(0.25) = 2.4$**,** $y(0.5) = 3.2$**,** $y(0.75) = 2.4$**)**

2. Compute $y(0.5)$, given that

$$y'' = y$$

$y(0) = 0$, $y(2) = 3.63$, with step size $h = 0.5$

(Ans. $y(0.5) = 0.5262$**,** $y(1) = 1.1843$**,** $y(1.5) = 2.1382$**)**

3. Solve the following BVP by finite difference method

$$y'' = -y$$

$$y(o) + y'(o) = 2$$

$$y\left(\frac{\pi}{2}\right) + y'\left(\frac{\pi}{2}\right) = -1$$

(Ans. $y(o) = 1.509, y\left(\frac{\pi}{8}\right) = 1.586 \;\; y\left(\frac{\pi}{4}\right) = 1.417$

$y\left(\frac{3\pi}{8}\right) = 1.031 \;\; y\left(\frac{\pi}{2}\right) = 0.485$**)**

4. Solve the BVP by Number method

$$x^2 y - 2y + x = 0$$

$y(2) = y(3) = 0$ with step size $\mathrm{h} = 0.25$

(Ans. $y(2.25) = 0.03783$, $y(2.5) = 0.048686$,

$y(2.75) = 0.035438$**)**

□□□