...ant and Regula-Falsi Methods

... and x_1 are two approximations to the root, then we determine a_0 and a_1 in (2.8) by using the ...ditions

$$f_{k-1} = a_0 x_{k-1} + a_1$$

$$f_k = a_0 x_k + a_1$$

$$f_{k-1} = f(x_{k-1}) \text{ and } f_k = f(x_k).$$

...re

...olving, we obtain

$$a_0 = (f_k - f_{k-1})/(x_k - x_{k-1})$$

$$a_1 = (x_k f_{k-1} - x_{k-1} f_k)/(x_k - x_{k-1}). \quad (2.10)$$

...m the equations (2.9) and (2.10), the next approximation x_{k+1} to the root is given by

$$x_{k+1} = \frac{x_{k-1} f_k - x_k f_{k-1}}{f_k - f_{k-1}} \quad (2.11)$$

...ich may also be written as

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f_k - f_{k-1}} f_k, \; k = 1, 2, \cdots \quad (2.12)$$

...is is called the **secant** or the **chord method.**

Geometrically, in this method we replace the function $f(x)$ by a straight line or a chord passing ...rough the points (x_k, f_k) and (x_{k-1}, f_{k-1}) and take the point of intersection of the straight line with ... x-axis as the next approximation to the root (Fig. 2.2a). If the approximations are such that ... $f_{k-1} < 0$, then the method (2.11) or (2.12) is known as the **Regula-Falsi method.** The method is ...own graphically in Fig. 2.2b. Since (x_{k-1}, f_{k-1}), (x_k, f_k) are known before the start of the iteration, ... secant and the Regular-Falsi methods require one function evaluation per iteration.

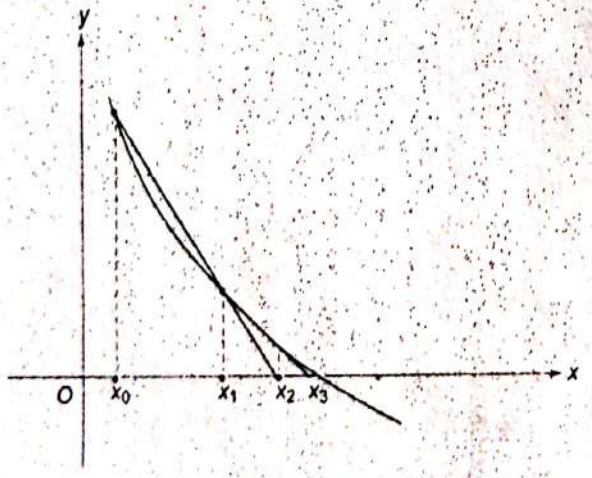

Fig. 2.2 (a). Secant method.

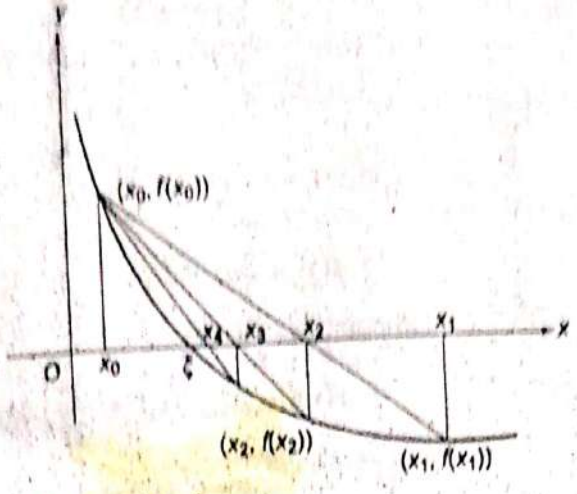

Fig. 2.2 (b). The Regula-Falsi method.

Example 2.5 A real root of the equation

$$f(x) = x^3 - 5x + 1 = 0$$

lies in the interval (0, 1). Perform four iterations of the secant method and the Regula-Falsi me to obtain this root.

We have $x_0 = 0,\ x_1 = 1,\ f_0 = f(x_0) = 1,\ f_1 = f(x_1) = -3.$

Secant method

$$x_2 = x_1 - \left[\frac{x_1 - x_0}{f_1 - f_0}\right] f_1 = 0.25,\ f_2 = f(x_2) = -0.234375.$$

$$x_3 = x_2 - \left[\frac{x_2 - x_1}{f_2 - f_1}\right] f_2 = 0.186441,\ f_3 = f(x_3) = 0.074276.$$

$$x_4 = x_3 - \left[\frac{x_3 - x_2}{f_3 - f_2}\right] f_3 = 0.201736,\ f_4 = f(x_4) = -0.000470.$$

$$x_5 = x_4 - \left[\frac{x_4 - x_3}{f_4 - f_3}\right] f_4 = 0.201640.$$

Regula-Falsi method

$$x_2 = x_1 - \left[\frac{x_1 - x_0}{f_1 - f_0}\right] f_1 = 0.25,\ f_2 = f(x_2) = -0.234375.$$

Since $f(x_0)\, f(x_2) < 0,\ \xi \in (x_0, x_2)$. Therefore,

$$x_3 = x_2 - \left[\frac{x_2 - x_0}{f_2 - f_0}\right] f_2 = 0.202532,\ f_3 = f(x_3) = -0.004352.$$

Since $f(x_0)\, f(x_3) < 0$, $\xi \in (x_0, x_3)$. Therefore,

$$x_4 = x_3 - \left[\frac{x_3 - x_0}{f_3 - f_0}\right] f_3 = 0.201654, \; f_4 = f(x_4) = -0.000070.$$

Since $f(x_0)\, f(x_4) < 0$, $\xi \in (x_0, x_4)$. Therefore,

$$x_5 = x_4 - \left[\frac{x_4 - x_0}{f_4 - f_0}\right] f_4 = 0.201640.$$

Example 2.6 Use the secant and Regula-Falsi methods to determine the root of the equation

$$\cos x - x e^x = 0.$$

Using the initial approximations as $x_0 = 0$, $x_1 = 1$, we obtain for the secant method

$$f(0) = 1, \quad f(1) = \cos 1 - e = -2.177979523$$

$$x_2 = x_1 - \left[\frac{x_1 - x_0}{f_1 - f_0}\right] f_1 = 0.3146653378$$

$$f_2 = f(x_2) = 0.519871175$$

$$x_3 = x_2 - \left[\frac{x_2 - x_1}{f_2 - f_1}\right] f_2 = 0.4467281466$$

$$f_3 = f(x_3) = 0.203544710$$

$$x_4 = x_3 - \left[\frac{x_3 - x_2}{f_3 - f_2}\right] f_3 = 0.5317058606.$$

Now, for the Regula-Falsi method, we get

$$x_2 = x_1 - \left[\frac{x_1 - x_0}{f_1 - f_0}\right] f_1 = 0.3146653378,$$

$$f_2 = f(x_2) = 0.519871175.$$

Since $f(x_1)\, f(x_2) < 0$, $\xi \in (x_1, x_2)$. Therefore,

$$x_3 = x_2 - \left[\frac{x_2 - x_1}{f_2 - f_1}\right] f_2 = 0.4467281466$$

$$f_3 = f(x_3) = 0.203544710.$$

Since $f(x_1)\, f(x_3) < 0$, $\xi \in (x_1, x_3)$. Therefore,

$$x_4 = x_3 - \left[\frac{x_3 - x_1}{f_3 - f_1}\right] f_3 = 0.4940153366.$$

The computed results are tabulated in Table 2.6.

Table 2.6 Approximations to the Root by the Secant and the Regula-Falsi Methods

k	*Secant Method*		*Regula-Falsi Method*	
	x_{k+1}	$f(x_{k+1})$	x_{k+1}	$f(x_{k+1})$
1	0.3146653378	0.519871	0.3146653378	0.519871
2	0.4467281466	0.203545	0.4467281446	0.203545
3	0.5317058606	−0.429311(− 01)	0.4940153366	0.708023(− 01)
4	0.5169044676	0.259276(− 02)	0.5099461404	0.236077(− 01)
5	0.5177474653	0.301119(− 04)	0.5152010099	0.776011(− 02)
6	0.5177573708	−0.215132(− 07)	0.5169222100	0.253886(− 02)
7	0.5177573637	0.178663(− 12)	0.5174846768	0.829358(− 03)
8	0.5177573637	0.222045(− 15)	0.5176683450	0.270786(− 03)
10	—	—	0.5177478783	0.288554(− 04)
20	—	—	0.5177573636	0.396288(− 09)

The numbers within the parentheses denote exponentiation.

Newton-Raphson Method

We determine a_0 and a_1 in (2.8) using the conditions

$$f_k = a_0 x_k + a_1$$

$$f'_k = a_0$$

where a prime denotes differentiation with respect to x.

On substituting a_0 and a_1 from (2.13) in (2.9) and representing the approximate value of x by x_{k+1} we obtain

$$x_{k+1} = x_k - \frac{f_k}{f'_k}, \quad k = 0, 1, \cdots \tag{2.14}$$

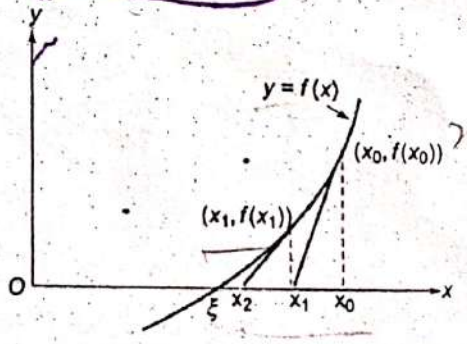

Fig. 2.3. The Newton-Raphson method.

This method is called the **Newton-Raphson** method. The method (2.14) may also be obtained directly from (2.12) by taking the limit $x_{k-1} \to x_k$. In the limit when $x_{k-1} \to x_k$, the chord passing through th

(x_k, f_k) and (x_{k-1}, f_{k-1}) becomes the tangent at the point (x_k, f_k). Thus, in this case the problem ...ding the root of the equation (2.1) is equivalent to finding the point of intersection of the tangent ...urve $y = f(x)$ at the point (x_k, f_k) with the x-axis. The method is shown graphically in Fig. 2.1 ...ewton-Raphson method requires two evaluations f_k, f'_k for each iteration.

...native

... be an approximation to the root of the equation $f(x) = 0$. Let Δx be an increment in x such ... Δx is an exact root. Therefore,

$$f(x_k + \Delta x) = 0.$$

...ding in Taylor series about the point x_k, we get

$$f(x_k) + \Delta x\, f'(x_k) + \frac{1}{2!}(\Delta x)^2 f''(x_k) + \cdots = 0.$$

...lecting the second and higher powers of Δx, we obtain

$$f(x_k) + \Delta x\, f'(x_k) = 0$$

$$\Delta x = -\frac{f(x_k)}{f'(x_k)}.$$

...ence, we obtain the iteration method

$$x_{k+1} = x_k + \Delta x = x_k - \frac{f(x_k)}{f'(x_k)}, \quad k = 0, 1, \cdots$$

...hich is same as (2.14).

...ample 2.7 Perform four iterations of the Newton-Raphson method to find the smallest positive ... of the equation

$$f(x) = x^3 - 5x + 1 = 0.$$

The smallest positive root lies in the interval (0, 1). Take the initial approximation as $x_0 = 0.5$. We ...ave

$$f(x) = x^3 - 5x + 1, \; f'(x) = 3x^2 - 5.$$

Using the Newton-Raphson method

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)},$$

we get

$$x_{k+1} = x_k - \frac{x_k^3 - 5x_k + 1}{3x_k^2 - 5} = \frac{2x_k^3 - 1}{3x_k^2 - 5}, \quad k = 0, 1, \cdots$$

Starting with $x_0 = 0.5$, we obtain

$$x_1 = 0.176471, \; x_2 = 0.201568.$$

$$x_3 = 0.201640, \; x_4 = 0.201640.$$

The exact value correct to six decimal places is 0.201640.

Example 2.8 Perform four iterations of the Newton-Raphson method to obtain the approximate value of $(17)^{1/3}$ starting with the initial approximation $x_0 = 2$.

Let $x = (17)^{1/3}$. We obtain $x^3 = 17$ and $f(x) = x^3 - 17 = 0$. Using the Newton-Raphson method

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad k = 0, 1, \cdots$$

we get

$$x_{k+1} = x_k - \frac{x_k^3 - 17}{3x_k^2} = \frac{2x_k^3 + 17}{3x_k^2}, \quad k = 0, 1, \cdots$$

Starting with $x_0 = 2$, we obtain

$$x_1 = \frac{2x_0^3 + 17}{3x_0^2} = 2.75, \qquad x_2 = \frac{2x_1^3 + 17}{3x_1^2} = 2.582645$$

$$x_3 = \frac{2x_2^3 + 17}{3x_2^2} = 2.571332, \quad x_4 = \frac{2x_3^3 + 17}{3x_3^2} = 2.571282.$$

The exact value correct to six decimal places is 2.571282.

Example 2.9 Apply Newton-Raphson's method to determine a root of the equation

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

such that $|f(x^*)| < 10^{-8}$, where x^* is the approximation to the root. Take the initial approximation as $x_0 = 1$.

We write (2.14) in the form

$$x_{k+1} = x_k - \Delta x_k, \; k = 0, 1, 2, \cdots$$

where $$\Delta x_k = \frac{f(x_k)}{f'(x_k)} = \frac{\left(\cos x_k - x_k e^{x_k}\right)}{\left(-\sin x_k - x_k e^{x_k} - e^{x_k}\right)}$$

Starting with $x_0 = 1$, we get

$$\Delta x_0 = \frac{\cos x_0 - x_0 e^{x_0}}{-\sin x_0 - x_0 e^{x_0} - e^{x_0}} = \frac{-2.17797952}{-6.27803464} = 0.34692060$$

$$x_1 = x_0 - \Delta x_0 = 1 - 0.34692060 = 0.65307940$$

$$\Delta x_1 = \frac{\cos x_1 - x_1 e^{x_1}}{-\sin x_1 - x_1 e^{x_1} - e^{x_1}} = \frac{-0.46064211}{-3.78394215} = 0.12173603$$

$$x_2 = x_1 - \Delta x_1 = 0.53134337.$$

The results obtained are given in Table 2.7.

where a_0, a_1 and a_2 are arbitrary parameters to be determined by prescribing three appropri
conditions on $f(x)$ and /or its derivatives.

Muller Method

If x_{k-2}, x_{k-1} and x_k are three approximations to the root ξ of $f(x) = 0$, then we may determine a_0,
and a_2 in (2.15) by using the conditions

$$\begin{aligned} f_{k-2} &= a_0 x_{k-2}^2 + a_1 x_{k-2} + a_2 \\ f_{k-1} &= a_0 x_{k-1}^2 + a_1 x_{k-1} + a_2 \\ f_k &= a_0 x_k^2 + a_1 x_k + a_2. \end{aligned} \tag{2.16}$$

Eliminating a_0, a_1 and a_2 from (2.15) and (2.16), we get

$$\begin{vmatrix} f(x) & x^2 & x & 1 \\ f_{k-2} & x_{k-2}^2 & x_{k-2} & 1 \\ f_{k-1} & x_{k-1}^2 & x_{k-1} & 1 \\ f_k & x_k^2 & x_k & 1 \end{vmatrix} = 0 \tag{2.17}$$

which we may simplify and obtain

$$f(x) = \frac{(x-x_{k-1})(x-x_k)}{(x_{k-2}-x_{k-1})(x_{k-2}-x_k)} f_{k-2} + \frac{(x-x_{k-2})(x-x_k)}{(x_{k-1}-x_{k-2})(x_{k-1}-x_k)} f_{k-1} + \frac{(x-x_{k-2})(x-x_{k-1})}{(x_k-x_{k-2})(x_k-x_{k-1})} f_k = 0. \tag{2.18}$$

The equation (2.18) may also be written as

$$\frac{h(h+h_k)}{h_{k-1}(h_{k-1}+h_k)} f_{k-2} - \frac{h(h+h_k+h_{k-1})}{h_k h_{k-1}} f_{k-1} + \frac{(h+h_k)(h+h_k+h_{k-1})}{h_k(h_k+h_{k-1})} f_k = 0 \tag{2.19}$$

where $h = x - x_k$, $h_k = x_k - x_{k-1}$, $h_{k-1} = x_{k-1} - x_{k-2}$.

We further define $\lambda = h/h_k$, $\lambda_k = h_k/h_{k-1}$ and $\delta_k = 1 + \lambda_k$

and express (2.19) in the form

$$\lambda^2 \epsilon_k + \lambda g_k + \delta_k f_k = 0 \tag{2.20}$$

where

$$g_k = \lambda_k^2 f_{k-2} - \delta_k^2 f_{k-1} + (\lambda_k + \delta_k) f_k$$

$$c_k = \lambda_k (\lambda_k f_{k-2} - \delta_k f_{k-1} + f_k)$$

(2.20) for λ, we obtain

$$\lambda = \frac{-g_k \pm \sqrt{g_k^2 - 4\delta_k f_k c_k}}{2c_k} \quad \text{or} \quad \lambda = \frac{-2\delta_k f_k}{g_k \pm \sqrt{g_k^2 - 4\delta_k f_k c_k}} = \lambda_{k+1}$$

sign in the denominator in (2.21) is chosen such that λ_{k+1} has the smallest absolute value

$$\lambda_{k+1} = \frac{h}{h_k} = \frac{x - x_k}{x_k - x_{k-1}}$$

$$x = x_k + (x_k - x_{k-1})\,\lambda_{k+1}.$$

placing x on the left hand side of (2.22) by x_{k+1}, we obtain the method

$$x_{k+1} = x_k + (x_k - x_{k-1})\,\lambda_{k+1}$$

is called the **Muller method**. The graphical representation of this method is shown in Fig.

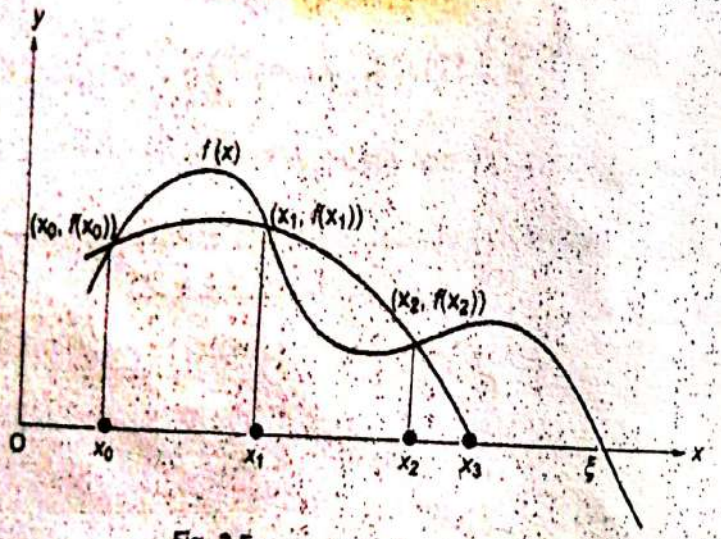

Fig. 2.5. The Muller method.

lternative

assume for $f(x)$ a polynomial of degree two in the form

$$f(x) = a_0(x - x_k)^2 + a_1(x - x_k) + a_2 = 0,\ a_0 \neq 0. \qquad (2.24)$$

bstituting $x = x_k, x_{k-1}$ and x_{k-2}, we determine a_0, a_1 and a_2 from the equations

$$f_k = a_2$$

$$f_{k-1} = a_0(x_{k-1} - x_k)^2 + a_1(x_{k-1} - x_k) + a_2$$

$$f_{k-2} = a_0(x_{k-2} - x_k)^2 + a_1(x_{k-2} - x_k) + a_2.$$

We obtain

$$a_2 = f_k$$

$$a_1 = \frac{1}{D}\left[(x_k - x_{k-2})^2 (f_k - f_{k-1}) - (x_k - x_{k-1})^2 (f_k - f_{k-2})\right]$$

$$a_0 = \frac{1}{D}\left[(x_k - x_{k-2}) (f_k - f_{k-1}) - (x_k - x_{k-1}) (f_k - f_{k-2})\right] \quad (2.2$$

where

$$\begin{aligned} D &= (x_{k-1} - x_k)^2 (x_{k-2} - x_k) - (x_{k-2} - x_k)^2 (x_{k-1} - x_k) \\ &= (x_{k-1} - x_k) (x_{k-2} - x_k) (x_{k-1} - x_k - x_{k-2} + x_k) \\ &= (x_k - x_{k-1}) (x_k - x_{k-2}) (x_{k-1} - x_{k-2}). \end{aligned}$$

Solving the equation (2.24) for $(x - x_k)$ and replacing x by x_{k+1}, we obtain

$$x_{k+1} = x_k + \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0 a_2}}{2a_0}$$

$$= x_k - \frac{2a_2}{a_1 \pm \sqrt{a_1^2 - 4a_0 a_2}}, \; k = 2, 3, \cdots \quad (2.2$$

The sign in the denominator in equation (2.25 *b*) is chosen as that of a_1, so that the denominato the maximum absolute value, that is, x_k changes by a smaller value.

Note that both the methods given in equations (2.23) and (2.25 *b*) are identical, though they different.

Generally, the method converges for all initial approximations. If no better approximations are kn then we choose $x_0 = -1$, $x_1 = 0$ and $x_2 = 1$. This method is an extension of the secant method. I the next approximation x_{k+1} is found as the zero of the second degree curve passing through the p (x_{k-2}, f_{k-2}), (x_{k-1}, f_{k-1}) and (x_k, f_k) where x_{k-2}, x_{k-1} and x_k are the initial approximations to the The method requires one function evaluation per iteration.

Example 2.11 Perform three iterations of the Muller method to find the smallest positive root of equation

$$f(x) = x^3 - 5x + 1 = 0.$$

The smallest positive root of the given equation lies in the interval (0, 1). We take the in approximations as $x_0 = 0$, $x_1 = 0.5$ and $x_2 = 1.0$. Therefore, $f_0 = 1$, $f_1 = -1.375$ and $f_2 = -3$. Us the method given in equation (2.23), we obtain

First iteration

$$h_2 = x_2 - x_1 = 0.5, \; h_1 = x_1 - x_0 = 0.5,$$

$$\lambda_2 = \frac{h_2}{h_1} = 1, \; \delta_2 = 1 + \lambda_2 = 2,$$

$$g_2 = \lambda_2^2 f_0 - \delta_2^2 f_1 + (\lambda_2 + \delta_2) f_2 = -2.5$$
$$c_2 = \lambda_2 (\lambda_2 f_0 - \delta_2 f_1 + f_2) = 0.75.$$

$g_2 < 0$, we get

$$\lambda_3 = \frac{-2\delta_2 f_2}{g_2 - \sqrt{g_2^2 - 4\delta_2 f_2 c_2}} = -\frac{12}{2.5 + \sqrt{24.25}} = -1.616286$$
$$x_3 = x_2 + (x_2 - x_1)\lambda_3 = 0.191857 \text{ and}$$
$$f_3 = 0.047777.$$

nd iteration

$$h_3 = x_3 - x_2 = -0.808143$$
$$\lambda_3 = \frac{h_3}{h_2} = -1.616286, \ \delta_3 = 1 + \lambda_3 = -0.616286$$
$$g_3 = \lambda_3^2 f_1 - \delta_3^2 f_2 + (\lambda_3 + \delta_3) f_3 = -2.559263$$
$$c_3 = \lambda_3 (\lambda_3 f_1 - \delta_3 f_2 + f_3) = -0.680961.$$

$g_3 < 0$, we get

$$\lambda_4 = \frac{-2\delta_3 f_3}{g_3 - \sqrt{g_3^2 - 4\delta_3 f_3 c_3}} = -\frac{0.058889}{2.559263 + \sqrt{6.469625}}$$
$$= -0.011541$$
$$x_4 = x_3 + (x_3 - x_2)\lambda_4 = 0.201184 \text{ and}$$
$$f_4 = 0.002223.$$

rd iteration

$$h_4 = x_4 - x_3 = 0.009327$$
$$\lambda_4 = \frac{h_4}{h_3} = -0.011541, \ \delta_4 = 1 + \lambda_4 = 0.988459$$
$$g_4 = \lambda_4^2 f_2 - \delta_4^2 f_3 + (\lambda_4 + \delta_4) f_4 = -0.044908$$
$$c_4 = \lambda_4(\lambda_4 f_2 - \delta_4 f_3 + f_4) = 0.000120.$$

$g_4 < 0$, we get

$$\lambda_5 = -\frac{2\delta_4 f_4}{g_4 - \sqrt{g_4^2 - 4\delta_4 f_4 c_4}} = \frac{0.004395}{0.044908 + \sqrt{0.002016}}$$
$$= 0.048940.$$
$$x_5 = x_4 + (x_4 - x_3)\lambda_5 = 0.201640.$$

we use the alternative method given in equation (2.25), we get

11

First iteration

$$a_2 = f_2 = -3,\ D = (x_2 - x_1)(x_2 - x_0)(x_1 - x_0) = 0.25$$

$$a_1 = \frac{1}{D}[(f_2 - f_1)(x_2 - x_0)^2 - (f_2 - f_0)(x_2 - x_1)^2] = -2.5$$

$$a_0 = \frac{1}{D}[(f_2 - f_1)(x_2 - x_0) - (f_2 - f_0)(x_2 - x_1)] = 1.5.$$

Since $a_1 < 0$, we get

$$x_3 = x_2 - \frac{2a_2}{a_1 - \sqrt{a_1^2 - 4a_0a_2}} = 1 - \frac{6}{2.5 + \sqrt{24.25}} = 0.191857.$$

Second iteration

$$a_2 = f_3 = 0.047777,\ D = (x_3 - x_1)(x_3 - x_2)(x_2 - x_1) = 0.124512$$

$$a_1 = \frac{1}{D}[(f_3 - f_2)(x_3 - x_1)^2 - (f_3 - f_1)(x_3 - x_2)^2] = -5.138588$$

$$a_0 = \frac{1}{D}[(f_3 - f_2)(x_3 - x_1) - (f_3 - f_1)(x_3 - x_2)] = 1.691854.$$

Since $a_1 < 0$, we get

$$x_4 = x_3 - \frac{2a_2}{a_1 - \sqrt{a_1^2 - 4a_0a_2}}$$

$$= 0.191857 + \frac{0.095554}{5.138588 + \sqrt{26.081760}} = 0.201183.$$

Third iteration

$$a_2 = f_4 = 0.002228,\ D = (x_4 - x_2)(x_4 - x_3)(x_3 - x_2) = 0.006020$$

$$a_1 = \frac{1}{D}[(f_4 - f_3)(x_4 - x_2)^2 - (f_4 - f_2)(x_4 - x_3)^2] = -4.871483$$

$$a_0 = \frac{1}{D}[(f_4 - f_3)(x_4 - x_2) - (f_4 - f_2)(x_4 - x_3)] = 1.393112.$$

Since $a_1 < 0$, we get

$$x_5 = x_4 - \frac{2a_2}{a_1 - \sqrt{a_1^2 - 4a_0a_2}}$$

$$= 0.201183 + \frac{0.004456}{4.871483 + \sqrt{23.718931}} = 0.201640.$$

Example 2.12 Perform five iterations of the Muller method to find the root of the equation

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

Use the initial approximations $x_0 = -1.0$, $x_1 = 0.0$ and $x_2 = 1.0$.

We have $f_0 = 0.90818175$, $f_1 = 1$, $f_2 = -2.17797952$.

Chebyshev Method

We determine a_0, a_1 and a_2 in (2.15) using the conditions

$$f_k = a_0 x_k^2 + a_1 x_k + a_2$$
$$f_k' = 2a_0 x_k + a_1$$
$$f_k'' = 2a_0. \qquad (2.$$

On eliminating a_i's from (2.15) and (2.26) we obtain

$$f_k + (x - x_k) f_k' + \frac{1}{2} (x - x_k)^2 f_k'' = 0 \qquad (2.$$

which is the Taylor series expansion of $f(x)$ about $x = x_k$ such that the terms of order $(x - x_k)^3$ higher powers are neglected.

The equation (2.27) is a quadratic equation and can be solved easily. Only one of the two r converges to the correct root. In order to get the next approximation to the correct root we write (2 as

$$x_{k+1} - x_k = -\frac{f_k}{f_k'} - \frac{1}{2} (x_{k+1} - x_k)^2 \frac{f_k''}{f_k'} \qquad (2$$

We substitute for $(x_{k+1} - x_k)$ from (2.14) by $(-f_k/f_k')$ on the right side of (2.28) and obtain

$$x_{k+1} = x_k - \frac{f_k}{f_k'} - \frac{1}{2} \frac{f_k^2}{f_k'^3} f_k'' \qquad (2$$

which is called the **Chebyshev method**. This method requires three evaluations for each iteratio $(x_{k+1} - x_k)$ in the right hand side of (2.28) is replaced by the secant or Regula-Falsi method, the c of the method is reduced.

Example 2.13 Perform two iterations of the Chebyshev method to find the smallest positive of the equation

$$f(x) = x^3 - 5x + 1 = 0.$$

Take the initial approximation as $x_0 = 0.5$.

We have

$$f(x) = x^3 - 5x + 1, f'(x) = 3x^2 - 5, f''(x) = 6x.$$

Using $x_0 = 0.5$, we get

$$f(x_0) = -1.375, f'(x_0) = -4.25, f''(x_0) = 3$$

$$x_1 = x_0 - \frac{f_0}{f_0'} - \frac{1}{2}\left(\frac{f_0}{f_0'}\right)^2 \left(\frac{f_0''}{f_0'}\right)$$
$$= 0.5 - 0.323529 - 0.5(0.104671)(-0.705882)$$
$$= 0.213414$$
$$f(x_1) = -0.057350, f'(x_1) = -4.863363, f''(x_1) = 1.280484$$

$$x_2 = x_1 - \frac{f_1}{f_1'} - \frac{1}{2}\left(\frac{f_1}{f_1'}\right)^2\left(\frac{f_1''}{f_1'}\right)$$
$$= 0.213414 - 0.011792 - 0.5\ (0.000139)\ (-\ 0.263292)$$
$$= 0.201640.$$

Example 2.14 Perform two iterations of the Chebyshev method to find an approximate value of $\frac{1}{7}$. Take the initial approximation as $x_0 = 0.1$.

Let $x = \frac{1}{7}$. We get $\frac{1}{x} = 7$. Define $f(x) = \frac{1}{x} - 7$. We get

$$f'(x) = -\frac{1}{x^2},\ f''(x) = \frac{2}{x^3}.$$

Using $x_0 = 0.1$, we get $f(x_0) = 3$, $f'(x_0) = -\ 100$, $f''(x_0) = 2000$.

$$x_1 = x_0 - \frac{f_0}{f_0'} - \frac{1}{2}\left(\frac{f_0}{f_0'}\right)^2\left(\frac{f_0''}{f_0'}\right)$$
$$= 0.1 + 0.03 - 0.5\ (0.0009)\ (-\ 20) = 0.139$$
$$f(x_1) = 0.194245,\ f'(x_1) = -\ 51.757155,\ f''(x_1) = 744.707272$$
$$x_2 = x_1 - \frac{f_1}{f_1'} - \frac{1}{2}\left(\frac{f_1}{f_1'}\right)^2\left(\frac{f_1''}{f_1'}\right)$$
$$= 0.139 + 0.003753 - 0.5\ (0.000014)\ (-\ 14.388489)$$
$$= 0.142854.$$

Example 2.15 Using Chebyshev method, find the root of the equation

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

correct to six decimal places. Take the initial approximation as $x_0 = 1.0$.

We write (2.29) in the form

$$x_{k+1} = x_k - \Delta x_k - \Delta^* x_k$$

where $\Delta x_k = \frac{f_k}{f_k'}$ and $\Delta^* x_k = \frac{1}{2}\left(\frac{f_k}{f_k'}\right)^2 \frac{f_k''}{f_k'}$.

We have

$$f(x) = \cos x - xe^x$$
$$f'(x) = -\sin x - (x + 1)e^x$$
$$f''(x) = -\cos x - (x + 2)e^x.$$

Example 2.34 Find the number of real and complex roots of the polynomial

$$P_4(x) = x^4 - 4x^3 + 3x^2 + 4x - 4$$

using Sturm sequence.

We have

$$f(x) = x^4 - 4x^3 + 3x^2 + 4x - 4,$$
$$f_1(x) = f'(x) = 4x^3 - 12x^2 + 6x + 4, \text{ or as } 2x^3 - 6x^2 + 3x + 2,$$
$$f_2(x) = 3x^2 - 9x + 6, \text{ or as } x^2 - 3x + 2,$$
$$f_3(x) = x - 2.$$

When we divide $f_2(x)$ by $f_3(x)$, we get zero as the remainder. Therefore, $f_3(x)$ is the last ele of the Sturm sequence. Hence, $x = 2$ is a double root of the polynomial. We divide each eleme the Sturm sequence by $(x - 2)$ and obtain the new sequence as

$$f^*(x) = x^3 - 2x^2 - x + 2, f_1^*(x) = 2x^2 - 2x - 1, f_2^*(x) = x - 1, f_3^*(x) = 1$$

We construct the following table of sign changes in the Sturm sequence.

x	$f^*(x)$	$f_1^*(x)$	$f_2^*(x)$	$f_3^*(x)$	$V(x)$
$-\infty$	−	+	−	+	3
− 1.5	−	+	−	+	3
0	+	−	−	+	2
1.5	−	+	+	+	1
2.5	+	+	+	+	0
∞	+	+	+	+	0

We find that the polynomial has three real roots in the intervals (– 1.5, 0), (0, 1.5) and (1.5, 2 Note that the root $x = 2$ which lies in the interval (1.5, 2.5) is a double root.

Hence, the polynomial has two simple roots in the intervals (– 1.5, 0) and (0, 1.5) and a dou root in the interval (1.5, 2.5). The exact roots of the polynomial are – 1, 1, 2, 2.

ITERATIVE METHODS

Birge-Vieta Method

In this method, we seek to determine a real number p such that $(x - p)$ is a factor of the polynom equation (2.88). If we divide $P_n(x)$ by the factor $(x - p)$ then we get a quotient $Q_{n-1}(x)$ of deg $(n - 1)$

$$Q_{n-1}(x) = b_0 x^{n-1} + b_1 x^{n-2} + \cdots + b_{n-2} x + b_{n-1} \quad (2.$$

and a remainder R. Thus we have

$$P_n(x) = (x - p)\, Q_{n-1}(x) + R. \quad (2.$$

The value of R depends on p. Starting with an initial approximation p_0 to p, we use some iterati method to improve the value of p such that

$$P_n(p) = R(p) = 0. \quad (2.$$

is a single equation in one unknown and the Newton-Raphson method or any
can be applied to improve the assumed value p_0. The Newton-Raphson method

$$p_{k+1} = p_k - \frac{P_n(p_k)}{P_n'(p_k)}, \quad k = 0, 1, 2, \cdots$$

obtaining a multiple root, we use the modified Newton-Raphson method.
polynomial equations, the computation of $P_n(p_0)$ and $P_n'(p_0)$ can be computed
thetic division. On comparing the coefficient of like powers of x on both sides of

$$\begin{aligned} a_0 &= b_0, & b_0 &= a_0 \\ a_1 &= b_1 - bp_0, & b_1 &= a_1 + pb_0 \\ a_2 &= b_2 - pb_1, & b_2 &= a_2 + pb_1 \\ &\vdots & &\vdots \\ a_k &= b_k - pb_{k-1}, & b_k &= a_k + pb_{k-1} \\ &\vdots & &\vdots \\ a_n &= R - pb_{n-1}, & R &= a_n + pb_{n-1}. \end{aligned}$$

introduce a quantity b_n and define the following recurrence relation

$$b_k = a_k + pb_{k-1}, \; k = 1, 2, \cdots, n$$

$= a_0$.

the equation (2.94) we have

$$P_n(p) = R = b_n.$$

mine $P_n'(p)$, we differentiate (2.98) with respect to p and obtain

$$\frac{db_k}{dp} = b_{k-1} + p\,\frac{db_{k-1}}{dp}.$$

put

$$\frac{db_k}{dp} = c_{k-1}$$

the equation (2.100) becomes

$$c_{k-1} = b_{k-1} + pc_{k-2}$$

can also be written as

$$c_k = b_k + pc_{k-1}, \; k = 1, 2, \cdots, n-1$$

$$c_0 = \frac{db_1}{dp} = \frac{d}{dp}(a_1 + pb_0) = b_0.$$

tiating (2.99) and using (2.101), we get

$$P_n'(p) = \frac{dR}{dp} = \frac{db_n}{dp} = c_{n-1}.$$

Bairstow Method

The **Bairstow method** extracts a quadratic factor of the form $x^2 + px + q$ from the polynomial (2.
which may give a pair of complex roots or a pair of real roots. If we divide the polynomial (2.8
the quadratic factor $x^2 + px + q$, then we obtain a quotient polynomial $Q_{n-2}(x)$ of degree $n-2$
a remainder term which is a polynomial of degree one, i.e. $Rx + S$.

Thus

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

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

The problem is then to find p and q, such that

$$R(p, q) = 0,\ S(p, q) = 0. \tag{2.}$$

The above equations are two simultaneous equations in two unknowns p and q. Suppose (p_0, q_0) is an initial approximation and that $(p_0 + \Delta p, q_0 + \Delta q)$ is the true solution. Following Newton-Raphson method, we obtain

$$\Delta p = -\frac{RS_q - SR_q}{R_p S_q - R_q S_p},\quad \Delta q = -\frac{R_p S - RS_p}{R_p S_q - R_q S_p} \tag{2.}$$

where R_p, R_q, S_p, S_q are the partial derivatives of R and S with respect to p and q respectively. quantities and R, S are evaluated at p_0, q_0.

The coefficients b_i, R and S can be determined by comparing the like powers of x in (2.105), obtain

$$\begin{aligned}
a_0 &= b_0, & b_0 &= a_0\\
a_1 &= b_1 + pb_0, & b_1 &= a_1 - pb_0\\
a_2 &= b_2 + pb_1 + qb_0, & b_2 &= a_2 - pb_1 - qb_0\\
&\vdots & &\vdots\\
a_k &= b_k + pb_{k-1} + qb_{k-2}, & b_k &= a_k - pb_{k-1} - qb_{k-2} \qquad (2.\\
&\vdots & &\vdots\\
a_{n-1} &= R + pb_{n-2} + qb_{n-3}, & R &= a_{n-1} - pb_{n-2} - qb_{n-3}\\
a_n &= S + qb_{n-2}, & S &= a_n - qb_{n-2}.
\end{aligned}$$

We now introduce the recursion formula

$$b_k = a_k - pb_{k-1} - qb_{k-2},\ k = 1, 2, \cdots, n$$

where $$b_0 = a_0,\ b_{-1} = 0.$$

Comparing the last two equations with those of (2.108), we get

$$\begin{aligned}
R &= b_{n-1}\\
S &= b_n + pb_{n-1}.
\end{aligned} \tag{2.}$$

For computing b_i's and c_i's we use the following scheme:

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

Note that the polynomial $P_n(x)$ is complete.

When p and q have been obtained to the desired accuracy, the polynomial

$$Q_{n-2}(x) = P_n(x)/(x^2 + px + q)$$
$$= b_0 x^{n-2} + b_1 x^{n-4} + \cdots + b_{n-2}$$

is called the **deflated** polynomial. The coefficients b_i, $i = 0, 1, 2, \cdots, n-2$ are known from synthetic division procedure. The next quadratic factor is obtained using this deflated polynomi

Example 2.37 Perform two iterations of the Bairstow method to extract a quadratic f $x^2 + px + q$ from the polynomial

$$P_3(x) = x^3 + x^2 - x + 2 = 0.$$

Use the initial approximations $p_0 = -0.9$, $q_0 = 0.9$.

Starting with $p_0 = -0.9$ and $q_0 = 0.9$, we obtain

0.9	1	1	-1	2
-0.9		0.9	1.71	-0.171
			-0.9	-1.71
	$1 = b_0$	1.9	$-0.19 = b_2$	$0.119 = b_3$
0.9		0.9	2.52	
-0.9			-0.9	
	$1 = c_0$	$2.8 = c_1$	$1.43 = c_2$	

$$\Delta p = -\frac{-b_3 c_0 - b_2 c_1}{c_1^2 - c_0(c_2 - b_2)} = -\frac{0.651}{6.22} = -0.1047$$

$$\Delta q = -\frac{b_2(c_2 - b_2) - b_3 c_1}{c_1^2 - c_0(c_2 - b_2)} = \frac{0.6410}{6.22} = 0.1031$$

$$p_1 = p_0 + \Delta p = -0.9 - 0.1047 = -1.0047$$
$$q_1 = q_0 + \Delta q = 0.9 + 0.1031 = 1.0031$$

		1	− 1	2
1.0047	1	1.0047	2.0141	0.0111
1.0031			− 1.0031	− 2.0109
	$1 = b_0$	2.0047	$0.0110 = b_2$	$0.0002 = b_3$
		1.0047	3.0235	
			− 1.0031	
	$1 = c_0$	$3.0094 = c_1$	$2.0314 = c_2$	

$$\Delta p = -\frac{b_3 c_0 - b_2 c_1}{c_1^2 - c_0(c_2 - b_2)} = \frac{0.0329}{7.0361} = 0.0047$$

$$\Delta q = -\frac{b_2(c_2 - b_2) - b_3 c_1}{c_1^2 - c_0(c_2 - b_2)} = -\frac{0.0216}{7.0361} = -0.0031$$

$$p_2 = p_1 + \Delta p = -1.0047 + 0.0047 = -1.0000$$

$$q_2 = q_1 + \Delta q = 1.0031 - 0.0031 = 1.0000.$$

Hence, the extracted quadratic factor is $x^2 + p_2 x + q^2 = x^2 - x + 1$.

The exact factor is $x^2 - x + 1$.

Example 2.38 Perform one iteration of the Bairstow method to extract a quadratic factor $x^2 + px + q$ from the polynomial

$$x^4 + x^3 + 2x^2 + x + 1 = 0.$$

Use the initial approximation $p_0 = 0.5$, $q_0 = 0.5$.

Starting with $p_0 = 0.5$, $q_0 = 0.5$ we obtain

0.5	1	1	2	1	1
0.5		− 0.5	− 0.25	− 0.625	− 0.0625
		0	− 0.5	− 0.25	− 0.625
	1	0.5	1.25	$0.125 = b_3$	$0.3125 = b_4$
		− 0.5	0.0	− 0.375	
			− 0.5	0.0	
	1	$0.0 = c_1$	$0.75 = c_2$	$-0.25 = c_3$	

$$\Delta p = -\frac{b_4 c_1 - b_3 c_2}{c_2^2 - c_1(c_3 - b_3)} = 0.1667$$

$$\Delta q = -\frac{b_3(c_3 - b_3) - b_4 c_2}{c_2^2 - c_1(c_3 - b_3)} = 0.5.$$

Therefore, $p_1 = p_0 + \Delta p = 0.6667$, $q_1 = q_0 + \Delta q = 1.0$.

The exact values of p and q are 1.0.

Unit finish.

$$-E B^{-1} C = \begin{bmatrix} 2 & 0 \\ 2 & -1 \end{bmatrix} - \frac{1}{4}\begin{bmatrix} 3 & 2 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 4 & -2 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 2 & 17 \\ 6 & 25 \end{bmatrix}$$

$$= (D - E B^{-1} C)^{-1} = \frac{1}{13}\begin{bmatrix} -25 & 17 \\ 6 & -2 \end{bmatrix}$$

$$= -B^{-1} C V = -\frac{1}{52}\begin{bmatrix} 0 & 1 \\ 4 & -2 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} -25 & 17 \\ 6 & -2 \end{bmatrix} = \frac{1}{13}\begin{bmatrix} 11 & -8 \\ 15 & -5 \end{bmatrix}$$

$$= -V E B^{-1} = -\frac{1}{52}\begin{bmatrix} -25 & 17 \\ 6 & -2 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 4 & -2 \end{bmatrix} = \frac{1}{13}\begin{bmatrix} -1 & 15 \\ -6 & -1 \end{bmatrix}$$

$$X = B^{-1} - B^{-1} C Z = \frac{1}{4}\begin{bmatrix} 0 & 1 \\ 4 & -2 \end{bmatrix} - \frac{1}{52}\begin{bmatrix} 0 & 1 \\ 4 & -2 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} -1 & 15 \\ -6 & -1 \end{bmatrix}$$

$$= \frac{1}{13}\begin{bmatrix} 2 & -4 \\ -2 & -9 \end{bmatrix}$$

Hence, we obtain

$$\mathbf{A}^{-1} = \frac{1}{13}\begin{bmatrix} 2 & -4 & 11 & -8 \\ -2 & -9 & 15 & -5 \\ -1 & 15 & -25 & 17 \\ -6 & -1 & 6 & -2 \end{bmatrix}$$

$$\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{b} = \frac{1}{13}\begin{bmatrix} 2 & -4 & 11 & -8 \\ -2 & -9 & 15 & -5 \\ -1 & 15 & -25 & 17 \\ -6 & -1 & 6 & -2 \end{bmatrix}\begin{bmatrix} -10 \\ 8 \\ 7 \\ -5 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \\ -10 \\ 8 \end{bmatrix}$$

UNIT-II Start

3.3 ERROR ANALYSIS FOR DIRECT METHODS

The methods discussed in the previous section involve only a finite number of arithmetic oper
The number of divisions and multiplications involved in solving the system of equations is u
called the operational count for that method. In the following, we obtain the operational cou
Gauss elimination procedure.

Operational count for Gauss elimination

Number of divisions:

first step [first equation (division by first pivot)] : n

second step [second equation (division by second pivot)] : $n - 1$

.....

nth step [nth equation (division by nth pivot)] : 1

introduce round-off errors in the computation. Because of this, the methods used will produce which will differ considerably from the exact solution. The exact solution x and the corresp approximate solution $\hat{\mathbf{x}}$ will satisfy respectively the equations

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$
$$(\mathbf{A} + \delta\mathbf{A})\,\hat{\mathbf{x}} = \mathbf{b} + \delta\mathbf{b}$$

where $\delta\mathbf{A}$ and $\delta\mathbf{b}$ are the changes in A and b respectively, due to round-off error. From (3.4) obtain

$$\hat{\mathbf{x}} - \mathbf{x} = (\mathbf{A} + \delta\mathbf{A})^{-1}(\mathbf{b} + \delta\mathbf{b}) - \mathbf{A}^{-1}\mathbf{b}$$
$$= \{(\mathbf{A} + \delta\mathbf{A})^{-1} - \mathbf{A}^{-1}\}\,\mathbf{b} + (\mathbf{A} + \delta\mathbf{A})^{-1}\,\delta\mathbf{b}$$

which may be called the error equation. In order to estimate the error vector $\varepsilon = \hat{\mathbf{x}} - \mathbf{x}$, we recall concept of a norm of a vector x and a matrix A.

Vector Norm

The non-negative quantity $\|\mathbf{x}\|$ is a measure of the size or length of a vector satisfying:

(i) $\|\mathbf{x}\| > 0$, for $\mathbf{x} \neq \mathbf{0}$ and $\|\mathbf{0}\| = 0$,

(ii) $\|c\mathbf{x}\| = |c|\,\|\mathbf{x}\|$, for an arbitrary complex number c,

(iii) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

The most commonly used norms are

(i) *Absolute norm* (l_1 *norm*)

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

(ii) *Euclidean norm*

$$\|\mathbf{x}\|_2 = (\mathbf{x}^*\mathbf{x})^{1/2} = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}.$$

(iii) *Maximum norm* (l_∞ *norm*)

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

Matrix Norm

The matrix norm, $\|\mathbf{A}\|$, is a non-negative number which satisfies the properties:

(i) $\|\mathbf{A}\| > 0$, if $\mathbf{A} \neq \mathbf{0}$ and $\|\mathbf{0}\| = 0$,

(ii) $\|c\mathbf{A}\| = |c|\,\|\mathbf{A}\|$, for an arbitrary complex number c,

(iii) $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$,

(iv) $\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\|\,\|\mathbf{B}\|$.

...t commonly used norms are:

...robenius or Euclidean norm

$$F(\mathbf{A}) = \left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{1/2}.$$

Maximum norm

$$\|\mathbf{A}\| = \|\mathbf{A}\|_\infty$$

$$= \max_i \sum_k |a_{ik}| \text{ (maximum absolute row sum).}$$

$$\|\mathbf{A}\| = \|\mathbf{A}\|_1$$

$$= \max_k \sum_i |a_{ik}| \text{ (maximum absolute column sum).}$$

...Hilbert norm or Spectral norm

$$\|\mathbf{A}\|_2 = \sqrt{\lambda}, \text{ where } \lambda = \rho(\mathbf{A}^* \mathbf{A}).$$

If $\mathbf{A}$ is Hermitian or real and symmetric, then

$$\lambda = \rho(\mathbf{A}^2) = \rho^2(\mathbf{A})$$

so that $\|\mathbf{A}\|_2 = \rho(\mathbf{A})$.

The matrix norm must be consistent with the vector norm that we are using for any vector $\mathbf{x}$ and ...trix $\mathbf{A}$, i.e.,

$$\|\mathbf{Ax}\| \le \|\mathbf{A}\| \, \|\mathbf{x}\|. \tag{3.54}$$

...may be verified that the norm

$$\|\mathbf{A}\| = \max_i \sum_{j=1}^{n} |a_{ij}| \tag{3.55}$$

...consistent with maximum norm $\|\mathbf{x}\|$.

...rror Estimate

...m (3.50) we obtain

$$\|\hat{\mathbf{x}} - \mathbf{x}\| \le \|(\mathbf{A} + \delta\mathbf{A})^{-1} - \mathbf{A}^{-1}\| \, \|\mathbf{b}\| + \|(\mathbf{A} + \delta\mathbf{A})^{-1}\| \, \|\delta\mathbf{b}\| \tag{3.56}$$

But
$$\|(\mathbf{A} + \delta\mathbf{A})^{-1}\| = \|(\mathbf{A} + \delta\mathbf{A})^{-1} - \mathbf{A}^{-1} + \mathbf{A}^{-1}\|$$

$$\le \|(\mathbf{A} + \delta\mathbf{A})^{-1} - \mathbf{A}^{-1}\| + \|\mathbf{A}^{-1}\| \tag{3.57}$$

and
$$\|(\mathbf{A} + \delta\mathbf{A})^{-1} - \mathbf{A}^{-1}\| = \|\mathbf{A}^{-1} - (\mathbf{A} + \delta\mathbf{A})^{-1}\|$$

$$\le \|\mathbf{A}^{-1}\| \, \|\mathbf{I} - (\mathbf{I} + \mathbf{A}^{-1} \delta\mathbf{A})^{-1}\|$$

$$= \|\mathbf{A}^{-1}\| \, \|(\mathbf{I} + \mathbf{A}^{-1} \delta\mathbf{A})^{-1} (\mathbf{I} + \mathbf{A}^{-1} \delta\mathbf{A} - \mathbf{I})\|$$

$$\mathbf{x}^{(k+1)} = \mathbf{H}\,\mathbf{x}^{(k)} + \mathbf{c},\ k = 0, 1, 2,... \tag{3.64}$$

re $\mathbf{x}^{(k+1)}$ and $\mathbf{x}^{(k)}$ are the approximations for $\mathbf{x}$ at the $(k+1)$th and kth iterations, respectively. H alled the iteration matrix depending on A and c is a column vector. In the limiting case when k , $\mathbf{x}^{(k)}$ converges to the exact solution

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \tag{3.65}$$

the iteration equation (3.64) becomes, by substitution from (3.65)

$$\mathbf{A}^{-1}\mathbf{b} = \mathbf{H}\,\mathbf{A}^{-1}\mathbf{b} + \mathbf{c}. \tag{3.66}$$

m (3.66), the column vector c is given by

$$\mathbf{c} = (\mathbf{I} - \mathbf{H})\,\mathbf{A}^{-1}\mathbf{b}. \tag{3.67}$$

now determine the iteration matrix H and the column vector c for a few well known iteration hods.

obi Iteration Method

assume that the quantities a_{ii} in (3.3) are pivot elements. The equations (3.3) may be written as

$$\begin{aligned} a_{11}x_1 &= -(a_{12}x_2 + a_{13}x_3 + \ldots + a_{1n}x_n) + b_1 \\ a_{22}x_2 &= -(a_{21}x_1 + a_{23}x_3 + \ldots + a_{2n}x_n) + b_2 \\ &\vdots \\ a_{nn}x_n &= -(a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{n,n-1}x_{n-1}) + b_n \end{aligned} \tag{3.68}$$

Jacobi iteration method or **Gauss-Jacobi** iteration method may now be defined as

$$\begin{aligned} x_1^{(k+1)} &= -\frac{1}{a_{11}}(a_{12}x_2^{(k)} + a_{13}x_3^{(k)} + \cdots + a_{1n}x_n^{(k)} - b_1) \\ x_2^{(k+1)} &= -\frac{1}{a_{22}}(a_{21}x_1^{(k)} + a_{23}x_3^{(k)} + \cdots + a_{2n}x_n^{(k)} - b_2) \\ &\vdots \\ x_n^{(k+1)} &= -\frac{1}{a_{nn}}(a_{n1}x_1^{(k)} + a_{n2}x_2^{(k)} + \cdots + a_{n,n-1}x_{n-1}^{(k)} - b_n) \end{aligned} \tag{3.69}$$

$$k = 0, 1, 2,...$$

ince, we replace the complete vector $\mathbf{x}^{(k)}$ in the right side of (3.69) at the end of each iteration, this ethod is also called the *method of simultaneous displacement.*

In matrix form, the method can be written as

$$\begin{aligned} \mathbf{x}^{(k+1)} &= -\mathbf{D}^{-1}(\mathbf{L} + \mathbf{U})\,\mathbf{x}^{(k)} + \mathbf{D}^{-1}\mathbf{b} \\ &= \mathbf{H}\,\mathbf{x}^{(k)} + \mathbf{c},\ k = 0, 1, 2, \ldots \end{aligned} \tag{3.70}$$

ere $\mathbf{H} = -\mathbf{D}^{-1}(\mathbf{L} + \mathbf{U}),\ \mathbf{c} = \mathbf{D}^{-1}\mathbf{b}$

Chapter 3

and L and U are respectively lower and upper triangular matrices with zero diagonal entries, D is a diagonal matrix such that $A = L + D + U$.

Equation (3.70) can alternatively be written as

$$\begin{aligned} x^{(k+1)} &= x^{(k)} - [I + D^{-1}(L + U)]\, x^{(k)} + D^{-1} b \\ &= x^{(k)} - D^{-1}[D + L + U]\, x^{(k)} + D^{-1} b \\ &= x^{(k)} + D^{-1}\,[b - Ax^{(k)}] \end{aligned}$$

or

$$v^{(k)} = D^{-1}\, r^{(k)}. \tag{3.7}$$

where $v^{(k)} = x^{(k+1)} - x^{(k)}$, is the error in the approximation and $r^{(k)} = b - Ax^{(k)}$ is the residual vector. We may rewrite the above equation as

$$D\, v^{(k)} = r^{(k)}.$$

We solve for $v^{(k)}$ and find $x^{(k+1)} = x^{(k)} + v^{(k)}$. These equations describe the **Jacobi iteration method** in an error format.

Example 3.21 Solve the system of equations

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

$$x_1 + 5x_2 + 2x_3 = -6$$

$$x_1 + 2x_2 + 3x_3 = -4$$

using the Jacobi iteration method given in equation (3.70) and its error format given in equation (3.71). Take the initial approximation as $x^{(0)} = [0.5, -0.5, -0.5]^T$ and perform three iterations in each case. The exact solution is $x_1 = 1$, $x_2 = -1$, $x_3 = -1$.

(i) We have

$$L = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 2 & 0 \end{bmatrix}, \; D = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \; U = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

$$H = -D^{-1}(L + U) = -\begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix}$$

$$= -\begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1/4 & -1/4 \\ -1/5 & 0 & -2/5 \\ -1/3 & -2/3 & 0 \end{bmatrix}$$

$$c = D^{-1} b = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} 2 \\ -6 \\ -4 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -6/5 \\ -4/3 \end{bmatrix}.$$

efore, Jacobi iteration method (3.70) becomes

$$\mathbf{x}^{(k+1)} = \begin{bmatrix} 0 & -1/4 & -1/4 \\ -1/5 & 0 & -2/5 \\ -1/3 & -2/3 & 0 \end{bmatrix} \mathbf{x}^{(k)} + \begin{bmatrix} 1/2 \\ -6/5 \\ -4/3 \end{bmatrix}$$

rting with $\mathbf{x}^{(0)} = [0.5, -0.5, -0.5]^T$, we obtain

$$\mathbf{x}^{(1)} = \begin{bmatrix} 0.75 \\ -1.1 \\ -1.1667 \end{bmatrix}, \mathbf{x}^{(2)} = \begin{bmatrix} 1.0667 \\ -0.8833 \\ -0.8500 \end{bmatrix}, \mathbf{x}^{(3)} = \begin{bmatrix} 0.9333 \\ -1.0733 \\ -1.1000 \end{bmatrix}$$

ternately, we may write directly

$$x_1^{(k+1)} = \frac{1}{4}\left[2 - x_2^{(k)} - x_3^{(k)}\right], \quad x_2^{(k+1)} = \frac{1}{5}\left[-6 - x_1^{(k)} - 2x_3^{(k)}\right]$$

$$x_3^{(k+1)} = \frac{1}{3}\left[-4 - x_1^{(k)} - 2x_2^{(k)}\right].$$

rting with $x_1^{(0)} = 0.5$, $x_2^{(0)} = -0.5$, $x_3^{(0)} = -0.5$, we get

$$\mathbf{x}^{(1)} = [0.75, -1.1, -1.1667]^T, \quad \mathbf{x}^{(2)} = [1.0667, -0.8833, -0.8500]^T$$

$$\mathbf{x}^{(3)} = [0.9333, -1.0733, -1.1000]^T$$

) Using (3.71), we get for $\mathbf{x}^{(0)} = [0.5, -0.5, -0.5]^T$

$k = 0$:
$$\mathbf{r}^{(0)} = \mathbf{b} - \mathbf{A}\,\mathbf{x}^{(0)} = \begin{bmatrix} 2 \\ -6 \\ -4 \end{bmatrix} - \begin{bmatrix} 1 \\ -3 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \\ -2 \end{bmatrix}$$

$$\mathbf{v}^{(0)} = \mathbf{D}^{-1}\,\mathbf{r}^{(0)} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \\ -2 \end{bmatrix} = \begin{bmatrix} 0.25 \\ -0.6 \\ -0.6667 \end{bmatrix}$$

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \mathbf{v}^{(0)} = [0.75, -1.1, -1.1667]^T$$

$k = 1$:
$$\mathbf{r}^{(1)} = \mathbf{b} - \mathbf{A}\,\mathbf{x}^{(1)} = \begin{bmatrix} 2 \\ -6 \\ -4 \end{bmatrix} - \begin{bmatrix} 0.7333 \\ -7.0834 \\ -4.9501 \end{bmatrix} = \begin{bmatrix} 1.2667 \\ 1.0834 \\ 0.9501 \end{bmatrix}$$

$$\mathbf{v}^{(1)} = \mathbf{D}^{-1}\,\mathbf{r}^{(1)} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} 1.2667 \\ 1.0834 \\ 0.9501 \end{bmatrix} = \begin{bmatrix} 0.3167 \\ 0.2167 \\ 0.3167 \end{bmatrix}$$

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \mathbf{v}^{(1)} = [1.0667, -0.8833, -0.85]^T.$$

$k = 2$:
$$\mathbf{r}^{(2)} = \mathbf{b} - \mathbf{A}\,\mathbf{x}^{(2)} = \begin{bmatrix} 2 \\ -6 \\ -4 \end{bmatrix} - \begin{bmatrix} 2.5335 \\ -5.0498 \\ -3.2499 \end{bmatrix} = \begin{bmatrix} -0.5335 \\ -0.9502 \\ -0.7501 \end{bmatrix}$$

$$\mathbf{v}^{(2)} = \mathbf{D}^{-1}\,\mathbf{r}^{(2)} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} -0.5335 \\ -0.9502 \\ -0.7501 \end{bmatrix} = \begin{bmatrix} -0.1 \\ -0.1 \\ -0.2 \end{bmatrix}$$

$$\mathbf{x}^{(3)} = \mathbf{x}^{(2)} + \mathbf{v}^{(2)} = [0.9333, -1.0733, -1.1000]^T.$$

we obtain the same result from both the techniques.

Seidel Iteration Method

use on the right hand side of (3.69), all the available values from the present iteration
e Gauss-Seidel method as

$$x_1^{(k+1)} = -\frac{1}{a_{11}}\,(a_{12}x_2^{(k)} + a_{13}x_3^{(k)} + \cdots + a_{1n}x_n^{(k)}) + \frac{b_1}{a_{11}}$$

$$x_2^{(k+1)} = -\frac{1}{a_{22}}\,(a_{21}x_1^{(k+1)} + a_{23}x_3^{(k)} + \cdots + a_{2n}x_n^{(k)}) + \frac{b_2}{a_{22}}$$

$$\vdots$$

$$x_n^{(k+1)} = -\frac{1}{a_{nn}}\,(a_{n1}x_1^{(k+1)} + a_{n2}x_2^{(k+1)} + \cdots + a_{n,n-1}x_{n-1}^{(k+1)}) + \cdots$$

may be rearranged in the form

$$a_{11}x_1^{(k+1)} = -\sum_{i=2}^{n} a_{1i}\,x_i^{(k)} + b_1$$

$$a_{21}x_1^{(k+1)} + a_{22}x_2^{(k+1)} = -\sum_{i=3}^{n} a_{2i}x_i^{(k)} + b_2 \qquad (3$$

$$\vdots$$

$$a_{n1}x_1^{(k+1)} + \cdots + a_{nn}x_n^{(k+1)} = b_n.$$

we replace the vector $\mathbf{x}^{(k)}$ in the right side of (3.69) element by element, this method is
d the *method of successive displacement.*

atrix notation, (3.72) becomes

$$(\mathbf{D} + \mathbf{L})\,\mathbf{x}^{(k+1)} = -\mathbf{U}\,\mathbf{x}^{(k)} + \mathbf{b}$$

$$\mathbf{x}^{(k+1)} = -(\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{U}\,\mathbf{x}^{(k)} + (\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{b} \qquad (3$$

$$= \mathbf{H}\,\mathbf{x}^{(k)} + \mathbf{c},\ k = 0, 1, 2, \ldots$$

$$\mathbf{H} = -(\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{U} \text{ and } \mathbf{c} = (\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{b}.$$

re

ation (3.73) can alternatively be written as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - [\mathbf{I} + (\mathbf{D} + \mathbf{L})^{-1}\mathbf{U}]\,\mathbf{x}^{(k)} + (\mathbf{D} + \mathbf{L})^{-1}\mathbf{b}$$

$$= \mathbf{x}^{(k)} - (\mathbf{D} + \mathbf{L})^{-1}\,(\mathbf{D} + \mathbf{L} + \mathbf{U})\,\mathbf{x}^{(k)} + (\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{b}$$

$$= \mathbf{x}^{(k)} - (\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{A}\mathbf{x}^{(k)} + (\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{b}$$

$$= \mathbf{x}^{(k)} + (\mathbf{D} + \mathbf{L})^{-1}\,(\mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}).$$

write

$$\mathbf{v}^{(k)} = (\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{r}^{(k)}$$

$\mathbf{v}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$ and $\mathbf{r}^{(k)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}$ is the residual vector.
rewrite the above equations as

$$(\mathbf{D} + \mathbf{L})\,\mathbf{v}^{(k)} = \mathbf{r}^{(k)} \tag{3.74}$$

ve for $\mathbf{v}^{(k)}$ by forward substitution. The solution is then found from

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{v}^{(k)}.$$

quations describe the Gauss-Seidel method in an error format.

le 3.22 Solve the system of equations

$$2x_1 - x_2 + 0x_3 = 7$$
$$-x_1 + 2x_2 - x_3 = 1$$
$$0x_1 - x_2 + 2x_3 = 1$$

the Gauss-Seidel method given in equations (3.73) and its error format given in equations
Take the initial approximation as $\mathbf{x}^{(0)} = \mathbf{0}$ and perform three iterations.
have

$$\mathbf{D} + \mathbf{L} = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

The Gauss-Seidel method gives

$$\mathbf{x}^{(k+1)} = -(\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{U}\,\mathbf{x}^{(k)} + (\mathbf{D} + \mathbf{L})^{-1}\mathbf{b}.$$

We get

$$(\mathbf{D} + \mathbf{L})^{-1} = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 1/2 & 0 & 0 \\ 1/4 & 1/2 & 0 \\ 1/8 & 1/4 & 1/2 \end{bmatrix}$$

$$(\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{U} = \begin{bmatrix} 1/2 & 0 & 0 \\ 1/4 & 1/2 & 0 \\ 1/8 & 1/4 & 1/2 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1/2 & 0 \\ 0 & -1/4 & -1/2 \\ 0 & -1/8 & -1/4 \end{bmatrix}$$

$$(\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{b} = \begin{bmatrix} 1/2 & 0 & 0 \\ 1/4 & 1/2 & 0 \\ 1/8 & 1/4 & 1/2 \end{bmatrix} \begin{bmatrix} 7 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7/2 \\ 9/4 \\ 13/8 \end{bmatrix}$$

Therefore, we obtain the iteration scheme

$$\mathbf{x}^{(k+1)} = \begin{bmatrix} 0 & 1/2 & 0 \\ 0 & 1/4 & 1/2 \\ 0 & 1/8 & 1/4 \end{bmatrix} \mathbf{x}^{(k)} + \begin{bmatrix} 7/2 \\ 9/4 \\ 13/8 \end{bmatrix}$$

Starting with zero initial vector, we get

$$\mathbf{x}^{(1)} = \begin{bmatrix} 3.5 \\ 2.25 \\ 1.625 \end{bmatrix}, \mathbf{x}^{(2)} = \begin{bmatrix} 4.625 \\ 3.625 \\ 2.3125 \end{bmatrix}, \text{ and } \mathbf{x}^{(3)} = \begin{bmatrix} 5.3125 \\ 4.3125 \\ 2.6563 \end{bmatrix}$$

The exact solution is $\mathbf{x} = [6, 5, 3]^T$.

(ii) Using (3.74), we get for $\mathbf{x}^{(0)} = 0$

$k = 0$: $\mathbf{r}^{(0)} = \mathbf{b} - \mathbf{A}\,\mathbf{x}^{(0)} = [7, 1, 1]^T$

$$\mathbf{v}^{(0)} = (\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{r}^{(0)} = \begin{bmatrix} 1/2 & 0 & 0 \\ 1/4 & 1/2 & 0 \\ 1/8 & 1/4 & 1/2 \end{bmatrix} \begin{bmatrix} 7 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3.5 \\ 2.25 \\ 1.625 \end{bmatrix}$$

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \mathbf{v}^{(0)} = [3.5, 2.25, 1.625]^T$$

$k = 1$: $$\mathbf{r}^{(1)} = \mathbf{b} - \mathbf{A}\,\mathbf{x}^{(1)} = \begin{bmatrix} 7 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 3.5 \\ 2.25 \\ 1.625 \end{bmatrix} = \begin{bmatrix} 2.25 \\ 1.625 \\ 0 \end{bmatrix}$$

$$\mathbf{v}^{(1)} = (\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{r}^{(1)} = \begin{bmatrix} 1/2 & 0 & 0 \\ 1/4 & 1/2 & 0 \\ 1/8 & 1/4 & 1/2 \end{bmatrix} \begin{bmatrix} 2.25 \\ 1.625 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.125 \\ 1.375 \\ 0.6875 \end{bmatrix}$$

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \mathbf{v}^{(1)} = [4.625, 3.625, 2.3125]^T$$

$k = 2$: $$\mathbf{r}^{(2)} = \mathbf{b} - \mathbf{A}\,\mathbf{x}^{(2)} = \begin{bmatrix} 7 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 4.625 \\ 3.625 \\ 2.3125 \end{bmatrix} = \begin{bmatrix} 1.375 \\ 0.6875 \\ 0 \end{bmatrix}$$

$$\mathbf{v}^{(2)} = (\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{r}^{(2)} = \begin{bmatrix} 1/2 & 0 & 0 \\ 1/4 & 1/2 & 0 \\ 1/8 & 1/4 & 1/2 \end{bmatrix} \begin{bmatrix} 1.375 \\ 0.6875 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.6875 \\ 0.6875 \\ 0.3438 \end{bmatrix}$$

$$\mathbf{x}^{(3)} = \mathbf{x}^{(2)} + \mathbf{v}^{(2)} = [5.3125, 4.3125, 2.6563]^T$$

Note that we obtain identical results by both the techniques.

Successive Over Relaxation (SOR) Method

This method is a generalization of the Gauss-Seidel method. This method is often used when the coefficient matrix of the system of equations is symmetric and has 'property A'. We define an auxiliary vector $\hat{x}$ as

$$\hat{\mathbf{x}}^{(k+1)} = -\mathbf{D}^{-1}\mathbf{L}\,\mathbf{x}^{(k+1)} - \mathbf{D}^{-1}\mathbf{U}\,\mathbf{x}^{(k)} + \mathbf{D}^{-1}\mathbf{b}.$$

final solution is now written as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + w\,(\hat{\mathbf{x}}^{(k+1)} - \mathbf{x}^{(k)})$$

$$\mathbf{x}^{(k+1)} = (1 - w)\,\mathbf{x}^{(k)} + w\,\hat{\mathbf{x}}^{(k+1)}.$$

stituting (3.75) into (3.76) and simplifying we obtain

$$\mathbf{x}^{(k+1)} = (\mathbf{D} + w\mathbf{L})^{-1}\,[(1 - w)\,\mathbf{D} - w\mathbf{U}]\,\mathbf{x}^{(k)} + w(\mathbf{D} + w\mathbf{L})^{-1}\mathbf{b}$$

$$= \mathbf{H}\mathbf{x}^{(k)} + \mathbf{c},\ k = 0, 1, 2,...$$

$$\mathbf{H} = (\mathbf{D} + w\mathbf{L})^{-1}\,[(1 - w)\mathbf{D} - w\mathbf{U}]$$

$$\mathbf{c} = w(\mathbf{D} + w\mathbf{L})^{-1}\mathbf{b}.$$

ation (3.77) can alternatively be written as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (\mathbf{D} + w\mathbf{L})^{-1}\,[(\mathbf{D} + w\mathbf{L}) - (1 - w)\mathbf{D} + w\mathbf{U}]\,\mathbf{x}^{(k)} + w(\mathbf{D} + w\mathbf{L})^{-1}\mathbf{b}$$

$$= \mathbf{x}^{(k)} + w(\mathbf{D} + w\mathbf{L})^{-1}\,\mathbf{r}^{(k)}$$

re $\mathbf{r}^{(k)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}$ is the residual.

may write

$$\mathbf{v}^{(k)} = w(\mathbf{D} + w\mathbf{L})^{-1}\,\mathbf{r}^{(k)}$$

$$(\mathbf{D} + w\mathbf{L})\,\mathbf{v}^{(k)} = w\mathbf{r}^{(k)}.$$

his equation describes the SOR method in its error format. For computational purpo venient to use this equation.

When $w = 1$, equation (3.78) reduces to the Gauss-Seidel method. The quantity w is ca **laxation parameter** and $\mathbf{x}^{(k+1)}$ is a weighted mean of $\hat{\mathbf{x}}^{(k+1)}$ and $\mathbf{x}^{(k)}$. From the equation (3 nd that the weights are non-negative for $0 \le w \le 1$. If $w > 1$, then the method is called a laxation method and if $w < 1$, then it is called an **under relaxation** method.

Convergence Analysis of Iterative Methods

To discuss the convergence of the iteration method (3.64), we study the behaviour of the differ tween the exact solution $\mathbf{x}$ and an approximation $\mathbf{x}^{(k)}$. The exact solution $\mathbf{x}$ will satisfy

$$\mathbf{x} = \mathbf{H}\,\mathbf{x} + \mathbf{c}. \qquad (3.$$

ubtracting (3.79) from (3.64) and substituting $\boldsymbol{\varepsilon}^{(k)} = \mathbf{x}^{(k)} - \mathbf{x}$, we get

$$\boldsymbol{\varepsilon}^{(k+1)} = \mathbf{H}\,\boldsymbol{\varepsilon}^{(k)},\ k = 0, 1, 2,... \qquad (3.80$$

om which we obtain

$$\boldsymbol{\varepsilon}^{(k)} = \mathbf{H}^k\,\boldsymbol{\varepsilon}^{(0)},\ k = 0, 1, 2,... \qquad (3.81$$

where we have assumed that the iteration matrix $\mathbf{H}$ remains constant for each iteration. We now give few results which we require for proving the convergence of the iterative methods.

Theorem 3.1 *Let* $\mathbf{A}$ *be a square matrix. Then*

$$\lim_{m \to \infty} \mathbf{A}^m = 0$$

$\|\mathbf{A}\| < 1$, *or iff* $\rho(\mathbf{A}) < 1$.

Proof. If $\|\mathbf{A}\| < 1$, then we have

$$\|\mathbf{A}^m\| \leq \|\mathbf{A}\|^m$$

$$\left\| \lim_{m \to \infty} A^m \right\| \leq \lim_{m \to \infty} \|\mathbf{A}\|^m = 0.$$

For simplicity, assume that all the eigenvalues of $\mathbf{A}$ are distinct. Then, there exists a similarity transformation $\mathbf{S}$, such that

$$\mathbf{A} = \mathbf{S}^{-1} \mathbf{D} \mathbf{S}$$

where $\mathbf{D}$ is the diagonal matrix having the eigenvalues of $\mathbf{A}$ on the diagonal.

Therefore,

$$\mathbf{A}^m = \mathbf{S}^{-1} \mathbf{D}^m \mathbf{S}$$

$$\mathbf{D}^m = \begin{bmatrix} \lambda_1^m & & 0 \\ & \lambda_2^m & \\ 0 & & \lambda_n^m \end{bmatrix}$$

where

Obviously, $\lim_{m \to \infty} \mathbf{A}^m = 0$, iff all the eigenvalues satisfy $|\lambda_i| < 1$, that is, $\rho(\mathbf{A}) < 1$.

Theorem 3.2 *The infinite series*

$$\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots$$

converges iff $\lim_{m \to \infty} \mathbf{A}^m = 0$. *The series converges to* $(\mathbf{I} - \mathbf{A})^{-1}$.

Proof. Only if, is obvious.

If $\lim_{m \to \infty} \mathbf{A}^m = 0$, then by Theorem 3.1, $\rho(\mathbf{A}) < 1$.

Hence $|\mathbf{I} - \mathbf{A}| \neq 0$ and $(\mathbf{I} - \mathbf{A})^{-1}$ exists.

Consider the identity

$$(\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^m)(\mathbf{I} - \mathbf{A}) = \mathbf{I} - \mathbf{A}^{m+1}.$$

Post multiplying by $(\mathbf{I} - \mathbf{A})^{-1}$, we have

$$(\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^m) = (\mathbf{I} - \mathbf{A}^{m+1})(\mathbf{I} - \mathbf{A})^{-1}$$

As $m \to \infty$, we get

$$\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots = (\mathbf{I} - \mathbf{A})^{-1}.$$

Theorem 3.3 *No eigenvalue of a matrix* $\mathbf{A}$ *exceeds the norm of a matrix, i.e.,* $\|\mathbf{A}\| \geq \rho(\mathbf{A})$.

Proof. We have

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

$$\|\lambda \mathbf{x}\| = \|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \|\mathbf{x}\|$$

$$|\lambda| \|\mathbf{x}\| \le \|\mathbf{A}\| \|\mathbf{x}\|, \qquad \|\mathbf{x}\| \neq 0$$

$$|\lambda| \le \|\mathbf{A}\|.$$

ence the result.

Theorem 3.4 *The iteration method of the form (3.64) for the solution of (3.3) converges to the exact solution for any initial vector, if* $\|\mathbf{H}\| < 1$.

Proof. Without loss of generality, we take initial vector $\mathbf{x}^{(0)} = \mathbf{0}$. We have

$$\begin{aligned}
\mathbf{x}^{(1)} &= \mathbf{c} \\
\mathbf{x}^{(2)} &= \mathbf{H}\mathbf{x}^{(1)} + \mathbf{c} = (\mathbf{H} + \mathbf{I})\,\mathbf{c} \\
\mathbf{x}^{(3)} &= \mathbf{H}\mathbf{x}^{(2)} + \mathbf{c} = (\mathbf{H}^2 + \mathbf{H} + \mathbf{I})\,\mathbf{c} \\
&\vdots \\
\mathbf{x}^{(k+1)} &= (\mathbf{H}^k + \mathbf{H}^{k-1} + \cdots + \mathbf{H} + \mathbf{I})\mathbf{c} \\
\lim_{k\to\infty} \mathbf{x}^{(k+1)} &= \lim_{k\to\infty} (\mathbf{H}^k + \mathbf{H}^{k-1} + \cdots + \mathbf{H} + \mathbf{I})\mathbf{c} \\
&= (\mathbf{I} - \mathbf{H})^{-1}\mathbf{c}
\end{aligned} \tag{3.82}$$

if $\|\mathbf{H}\| < 1$ or iff $\rho(\mathbf{H}) < 1$.

In the case of the Jacobi method, we have

$$\begin{aligned}
(\mathbf{I} - \mathbf{H})^{-1}\mathbf{c} &= [\,\mathbf{I} + \mathbf{D}^{-1}(\mathbf{L} + \mathbf{U})\,]^{-1}\,\mathbf{D}^{-1}\mathbf{b} \\
&= [\,\mathbf{D}^{-1}(\mathbf{D} + \mathbf{L} + \mathbf{U})\,]^{-1}\,\mathbf{D}^{-1}\mathbf{b} \\
&= (\mathbf{D} + \mathbf{L} + \mathbf{U})^{-1}\,\mathbf{D}\,\mathbf{D}^{-1}\mathbf{b} \\
&= \mathbf{A}^{-1}\mathbf{b} = \mathbf{x}.
\end{aligned}$$

Similar result can be proved for the Gauss-Seidel and SOR methods.

Theorem 3.5 *A necessary and sufficient condition for convergence of an iterative method of the form (3.64) is that the eigenvalues of the iteration matrix satisfy*

$$|\lambda_i(\mathbf{H})| < 1, \quad i = 1(1)n.$$

Proof. We prove the result for the case when the iteration matrix $\mathbf{H}$ has n independent eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively. The error vector $\boldsymbol{\varepsilon}^{(0)}$ can be written as

$$\boldsymbol{\varepsilon}^{(0)} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n$$

Using (3.81) we get

$$\boldsymbol{\varepsilon}^{(k)} = c_1\lambda_1^k\mathbf{x}_1 + c_2\lambda_2^k\mathbf{x}_2 + \cdots + c_n\lambda_n^k\mathbf{x}_n \tag{3.83}$$

(*i*) *Necessity.* If $\lim_{k\to\infty} \boldsymbol{\varepsilon}^{(k)} = \mathbf{0}$ for any arbitrary initial vector $\mathbf{x}^{(0)}$ and thus for any arbitrary error vector $\boldsymbol{\varepsilon}^{(0)}$, then by (3.83), the magnitudes of the eigenvalues, $|\lambda_i|$, $i = 1(1)n$, must necessarily be less than unity.

(*ii*) *Sufficiency.* For $|\lambda_i| < 1$, $i = 1(1)n$, the convergence of $\boldsymbol{\varepsilon}^{(k)}$ towards the zero vector follows from (3.83).

***Definition* 3.1** *The rate of convergence of an iterative method is given by*

$$\upsilon = -\log_{10}[\rho(\mathbf{H})] \text{ or also as } \upsilon = -\ln[\rho(\mathbf{H})] \tag{3.84}$$

where $\rho(\mathbf{H})$ is the spectral radius of $\mathbf{H}$.

Chapter 3

from (3.86)

$$|\lambda||a_{ii}| \le \sum_{j=i+1}^{n} |a_{ij}| + |\lambda| \sum_{j=1}^{i-1} |a_{ij}|$$

$$|\lambda|\left[|a_{ii}| - \sum_{j=1}^{i-1} |a_{ij}|\right] \le \sum_{j=i+1}^{n} |a_{ij}|$$

$$|\lambda| \le \frac{\sum_{j=i+1}^{n} |a_{ij}|}{|a_{ii}| - \sum_{j=1}^{i-1} |a_{ij}|} < 1$$

true, since **A** is strictly diagonally dominant.

nal Relaxation Parameter for the SOR Method

laxation method (3.77) can be written as

$$\begin{aligned}\mathbf{x}^{(k+1)} &= (\mathbf{I} + w\,\mathbf{D}^{-1}\mathbf{L})^{-1}\left[(1-w)\mathbf{I} - w\,\mathbf{D}^{-1}\mathbf{U}\right]\mathbf{x}^{(k)} \\ &\quad + w\,(\mathbf{I} + w\mathbf{D}^{-1}\mathbf{L})^{-1}\,(\mathbf{D}^{-1}\mathbf{b}) \\ &= (\mathbf{I} + w\,\mathbf{R})^{-1}\left[(1-w)\mathbf{I} - w\,\mathbf{C}\right]\mathbf{x}^{(k)} \\ &\quad + w\,(\mathbf{I} + w\mathbf{R})^{-1}\,(\mathbf{D}^{-1}\mathbf{b})\end{aligned}$$

$\mathbf{R} = \mathbf{D}^{-1}\,\mathbf{L}$ and $\mathbf{C} = \mathbf{D}^{-1}\,\mathbf{U}$.

A has 'property *A*', there exists a permutation matrix **P** such that

$$\mathbf{M} = \mathbf{PAP}^{T} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \qquad (3$$

re $\mathbf{A}_{11}$, $\mathbf{A}_{22}$ are diagonal matrices. Hence, **A** and **M** have the same eigenvalues. The proble determine a value of $w = w_{opt}$ such that $\rho(\mathbf{H})$ is minimized. It is sufficient to consider envalues of **M**. Without loss of generality, assume **A** is in the form (3.87). The Jacobi method new system is

$$\begin{aligned}\mathbf{x}^{(k+1)} &= -\,\mathbf{D}^{-1}(\mathbf{L} + \mathbf{U})\,\mathbf{x}^{(k)} + \mathbf{D}^{-1}\mathbf{b} \\ &= -\,(\mathbf{R} + \mathbf{C})\,\mathbf{x}^{(k)} + \mathbf{D}^{-1}\mathbf{b} \\ &= \mathbf{B}\,\mathbf{x}^{(k)} + \mathbf{D}^{-1}\mathbf{b}\end{aligned}$$

here $\mathbf{B} = -\,(\mathbf{R} + \mathbf{C})$.

ote that $\mu\mathbf{I} - \mathbf{B}$ is of the form

$$\mu\,\mathbf{I} - \mathbf{B} = \begin{pmatrix} \mu\mathbf{I}_1 & \overline{\mathbf{A}}_{12} \\ \overline{\mathbf{A}}_{21} & \mu\mathbf{I}_2 \end{pmatrix}$$

ince **B** is of the form

$$\mathbf{B} = \begin{pmatrix} \mathbf{N}_1 & -\overline{\mathbf{A}}_{12} \\ -\overline{\mathbf{A}}_{21} & \mathbf{N}_2 \end{pmatrix}, \text{ where } \overline{\mathbf{A}}_{12} = \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \text{ and } \overline{\mathbf{A}}_{21} = \mathbf{A}_{22}^{-1}\mathbf{A}_{21}$$

I_1, I_2, N_1, N_2 are identity and null matrices respectively, of required orders. The characteristic equation of $\mathbf{B}$ is

$$|\mu\mathbf{I} - \mathbf{B}| = \begin{vmatrix} \mu\mathbf{I}_1 & \bar{\mathbf{A}}_{12} \\ \bar{\mathbf{A}}_{21} & \mu\mathbf{I}_2 \end{vmatrix}$$

following properties of this determinant are of interest:

The characteristic equation is of the form

$$\mu^n + a_1\mu^{n-2} + a_2\mu^{n-4} + \cdots = 0$$

If all the elements of $\bar{\mathbf{A}}_{12}$ are multiplied by a factor k, and all elements of $\bar{\mathbf{A}}_{21}$ are divided by the same factor k, then the value of the determinant is unchanged.

μ be an eigenvalue of $\mathbf{B}$ and λ be an eigenvalue of $\mathbf{H} = \mathbf{H}_{SOR}$. Then

$$|\mathbf{H} - \lambda\mathbf{I}| = 0$$

$$|(\mathbf{I} + w\mathbf{R})^{-1}[(1-w)\mathbf{I} - w\mathbf{C}] - \lambda\mathbf{I}| = 0$$

$$|(\mathbf{I} + w\mathbf{R})^{-1}|\,|(1-w)\mathbf{I} - w\mathbf{C} - \lambda(\mathbf{I} + w\mathbf{R})| = 0$$

$$|(1-w)\mathbf{I} - w\mathbf{C} - \lambda(\mathbf{I} + w\mathbf{R})| = 0$$

$|(\mathbf{I} + w\mathbf{R})^{-1}| \neq 0$. We have

$$|w(\mathbf{C} + \lambda\mathbf{R}) + (\lambda + w - 1)\mathbf{I}| = 0.$$

divide $\mathbf{R}$ by $\lambda^{1/2}$ and multiply $\mathbf{C}$ by $\lambda^{1/2}$, then the value of the above determinant is unchanged

$$|(C\lambda^{1/2} + \lambda^{1/2}\mathbf{R}) + \frac{\lambda + w - 1}{w}\mathbf{I}| = 0$$

$$|(\mathbf{C} + \mathbf{R}) + \frac{\lambda + w - 1}{\lambda^{1/2}w}\mathbf{I}| = 0$$

$$|\mathbf{B} - \frac{\lambda + w - 1}{\lambda^{1/2}w}\mathbf{I}| = 0. \tag{3.88}$$

μ is an eigenvalue of $\mathbf{B}$, we get

$$\mu = \frac{\lambda + w - 1}{\lambda^{1/2}w}.$$

$$\lambda - \mu w\lambda^{1/2} + (w-1) = 0$$

$$\lambda^{1/2} = \frac{1}{2}\left[\mu w \pm \sqrt{\mu^2 w^2 - 4(w-1)}\right]$$

$$= \frac{1}{2}\mu w \pm \frac{\mu}{2}\sqrt{(w - w_1)(w - w_2)}$$

$$w_{1,2} = \frac{2}{\mu^2}\left[1 \mp \sqrt{1 - \mu^2}\right].$$

For $w < w_1$ or $w > w_2$, λ is real.

For $w_1 < w < w_2$, λ is complex and

$$|\lambda^{1/2}|^2 = |\lambda| = w - 1.$$

In the real case, the larger eigenvalue is of interest. The graph of $|\lambda|$ versus w is given in Fig. 3.1.

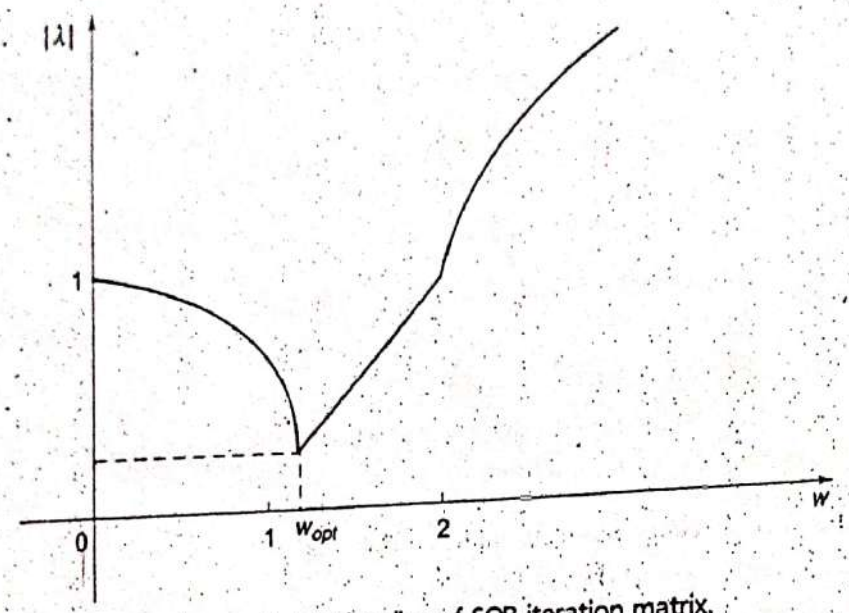

Fig. 3.1. Spectral radius of SOR iteration matrix.

When $w = w_1$, $\rho(\mathbf{H})$ is smallest. Hence, the optimal relaxation factor is given as

$$w_{opt} = \frac{2}{\mu^2}\left[1 - \sqrt{1-\mu^2}\right] = \frac{2}{\left[1+\sqrt{1-\mu^2}\right]}. \tag{3.89}$$

When $w = w_1$, we get

$$\lambda = \frac{1}{4}\mu^2 w^2 = w - 1. \tag{3.90}$$

For convergence, we require

$$|\lambda| < 1$$

$$|w - 1| < 1 \text{ or } 0 < w < 2. \tag{3.91}$$

For $0 < w < 1$, it is called under-relaxation and for $1 < w < 2$ it is called over-relaxation. The rate of convergence of the SOR scheme is $-\log(w - 1)$. The relaxation factor w_{opt} should be rounded to the next digit, for, when $w \to w_{opt}-$, the slope is infinite. When $w = 1$, we have from (3.88),

$$\mu = \lambda^{1/2} \text{ or } \lambda = \mu^2$$

$$\rho(\mathbf{H}_G) = [\rho(\mathbf{H}_J)]^2 \tag{3.92}$$

Therefore, the rate of convergence of Gauss-Seidel scheme is twice that of the Jacobi scheme.

From (3.89), we find that w_{opt} is real if $|\mu| < 1$. Therefore, the necessary condition for the SOR method to converge is that the corresponding Jacobi iteration method is convergent.

where

$$H_{SOR} = (D + wL)^{-1}\,[(1 - w)\,D - wU]$$

$$= \begin{bmatrix} 2 & 0 & 0 \\ -w & 2 & 0 \\ 0 & -w & 2 \end{bmatrix}^{-1} \begin{bmatrix} 2(1-w) & w & 0 \\ 0 & 2(1-w) & w \\ 0 & 0 & 2(1-w) \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 & 0 & 0 \\ w/4 & 1/2 & 0 \\ w^2/8 & w/4 & 1/2 \end{bmatrix} \begin{bmatrix} 2(1-w) & w & 0 \\ 0 & 2(1-w) & w \\ 0 & 0 & 2(1-w) \end{bmatrix}$$

$$= \begin{bmatrix} 1-w & w/2 & 0 \\ w(1-w)/2 & 1-w+(w^2/4) & w/2 \\ w^2(1-w)/4 & (w(1-w)/2)+(w^3/8) & 1-w+(w^2/4) \end{bmatrix}$$

$$\mathbf{C} = w\,(D + wL)^{-1}\,\mathbf{b}$$

$$= w \begin{bmatrix} 1/2 & 0 & 0 \\ w/4 & 1/2 & 0 \\ w^2/8 & w/4 & 1/2 \end{bmatrix} \begin{bmatrix} 7 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7w/2 \\ w(7w+2)/4 \\ w(7w^2+2w+4)/8 \end{bmatrix}$$

Hence, we have

$$\mathbf{x}^{(k+1)} = \begin{bmatrix} 1-w & w/2 & 0 \\ w(1-w)/2 & 1-w+(w^2/4) & w/2 \\ w^2(1-w)/4 & (w(1-w)/2+(w^3/8) & 1-w+w^2/4 \end{bmatrix} \mathbf{x}^{(k)}$$

$$+ \begin{bmatrix} 7w/2 \\ w(7w+2)/4 \\ w(7w^2+2w+4)/8 \end{bmatrix}$$

The iteration matrix associated with the Jacobi method is given by

$$H_J = -D^{-1}(L+U) = -\begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}\begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{bmatrix}$$

$$\text{Now, } |H_J - \lambda I| = \begin{vmatrix} -\lambda & 1/2 & 0 \\ 1/2 & -\lambda & 1/2 \\ 0 & 1/2 & -\lambda \end{vmatrix} = -\lambda\left(\lambda^2 - \frac{1}{2}\right) = 0$$

$$\lambda = 0, \pm 1/\sqrt{2}.$$

The spectral radius of the Jacobi iteration matrix is $\mu = 1/\sqrt{2}$. The optimal relaxation factor for the SOR scheme is

$$w_{opt} = \frac{2}{\mu^2}\left(1-\sqrt{1-\mu^2}\right) = 4\left(1-\frac{1}{\sqrt{2}}\right) \approx 1.171573$$

For $w = w_{opt}$, we have

$$\rho(H_{SOR}) = \frac{\mu^2 w^2}{4} = w - 1 = 0.171573.$$

The rate of convergence of the SOR method is

$$v = -\log(0.171573) = 0.7656.$$

Substituting the value of $w = w_{opt} = 1.1716$, the SOR iteration scheme becomes

$$\mathbf{x}^{(k+1)} = \begin{bmatrix} -0.1716 & 0.5858 & 0 \\ -0.1005 & 0.1716 & 0.5858 \\ -0.0589 & 0.1005 & 0.1716 \end{bmatrix}\mathbf{x}^{(k)} + \begin{bmatrix} 4.1006 \\ 2.9879 \\ 2.3361 \end{bmatrix}, \quad k = 0$$

Starting with $\mathbf{x}^{(0)} = \mathbf{0}$, we obtain

$$\mathbf{x}^{(1)} = \begin{bmatrix} 4.1006 \\ 2.9879 \\ 2.3361 \end{bmatrix}, \quad \mathbf{x}^{(2)} = \begin{bmatrix} 5.1472 \\ 4.4570 \\ 2.7957 \end{bmatrix}, \quad \mathbf{x}^{(3)} = \begin{bmatrix} 5.8283 \\ 4.8731 \\ 2.9606 \end{bmatrix}$$

Write the SOR scheme in the form

$$(D + wL)\,\mathbf{v}^{(k)} = w\,\mathbf{r}^{(k)}, \quad w = w_{opt} = 1.1716$$

$$\text{or, } \begin{bmatrix} 2 & 0 & 0 \\ -1.1716 & 2 & 0 \\ 0 & -1.1716 & 2 \end{bmatrix}\mathbf{v}^{(k)} = 1.1716\,\mathbf{r}^{(k)}$$

$b - A x^{(k)}$, $x^{(k+1)} = x^{(k)} + r^{(k)}$, $k = 0, 1, \ldots$

$x^{(0)} = 0$, we obtain

$$r^{(0)} = \begin{pmatrix} 7 \\ 1 \\ 1 \end{pmatrix}, \quad v^{(0)} = \begin{pmatrix} 4.1006 \\ 2.9897 \\ 2.3361 \end{pmatrix}, \quad x^{(1)} = \begin{pmatrix} 4.1006 \\ 2.9897 \\ 2.3361 \end{pmatrix}$$

$$r^{(1)} = \begin{pmatrix} 1.7867 \\ 1.4609 \\ -0.6843 \end{pmatrix}, \quad v^{(1)} = \begin{pmatrix} 1.0466 \\ 1.4689 \\ 0.4596 \end{pmatrix}, \quad x^{(2)} = \begin{pmatrix} 5.1472 \\ 4.4568 \\ 2.7957 \end{pmatrix}$$

$$r^{(2)} = \begin{pmatrix} 1.1624 \\ 0.0293 \\ -0.1346 \end{pmatrix}, \quad v^{(2)} = \begin{pmatrix} 0.6809 \\ 0.4160 \\ 0.1648 \end{pmatrix}, \quad x^{(3)} = \begin{pmatrix} 5.8281 \\ 4.8728 \\ 2.9605 \end{pmatrix}$$

Method to determine A^{-1}

hat we know the approximate inverse B of a matrix A. Then $AB \neq I$. The error matrix is

$$E = AB - I$$

$$AB = I + E. \tag{3.93}$$

93) we get

$$A^{-1} = B(I + E)^{-1}$$
$$= B(I - E + E^2 - \cdots) \tag{3.94}$$

< 1. Approximating (3.94) we have

$$A^{-1} \approx B(I - E) = B(2I - AB).$$

es an iterative method

$$B^{(k+1)} = B^{(k)}(2I - AB^{(k)}), \quad k = 0, 1, 2, \ldots \tag{3.95}$$

plying (3.95) by A we have

$$AB^{(k+1)} = 2AB^{(k)} - (AB^{(k)})^2 \tag{3.96}$$

$$I - AB^{(k+1)} = I - 2AB^{(k)} + (AB^{(k)})^2$$
$$= (I - AB^{(k)})^2$$

the convergence of the iterative method is quadratic.

ple 3.26 Find the inverse of the matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0.007612 & 1 & 0 \\ 0.000018 & -0.009231 & 1 \end{bmatrix} \begin{bmatrix} 4.791209 & 0.007633 & 1 \\ 0 & -0.997374 & 0.344036 \\ 0 & 0 & 0.209265 \end{bmatrix} = \mathbf{L}_5 \mathbf{U}_5$$

$$A_5 = U_5 L_5 = \begin{bmatrix} 4.791285 & -0.001598 & 1 \\ -0.007586 & -1.000550 & 0.344036 \\ 0.000004 & -0.001932 & 0.209265 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ -0.001583 & 1 & 0 \\ 0.000001 & 0.001931 & 1 \end{bmatrix} \begin{bmatrix} 4.791285 & -0.001598 & 1 \\ 0 & -1.000553 & 0.345619 \\ 0 & 0 & 0.208597 \end{bmatrix} = \mathbf{L}_6 \mathbf{U}_6$$

$$L_6 = \begin{bmatrix} 4.791289 & 0.000333 & 1 \\ 0.001584 & -0.999886 & 0.345619 \\ 0 & 0.000403 & 0.208597 \end{bmatrix}$$

the eigenvalues are approximately 4.791289, −0.999886 and 0.208597. The exact

ues are

$\lambda = (5 + \sqrt{21})/2 = 4.791288$, $\lambda = -1$ and $\lambda = (5 - \sqrt{21})/2 = 0.208712$.

POWER METHOD

method is normally used to determine the largest eigenvalue (in magnitude) and the ding eigenvector of the system

$A x = \lambda x$.

$\lambda_2, \ldots, \lambda_n$ be the distinct eigenvalues such that

$$|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n| \tag{3.152}$$

$\mathbf{v}_2, \ldots, \mathbf{v}_n$ be the corresponding eigenvectors. The procedure is applicable if a complete system endent eigenvectors exists, even though some of the eigenvalues $\lambda_2, \ldots, \lambda_n$ may not be Then, any eigenvector $\mathbf{v}$ in the space of eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ can be written as

$$\mathbf{v} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \ldots + c_n \mathbf{v}_n. \tag{3.153}$$

plying by A and substituting $A \mathbf{v}_1 = \lambda_1 \mathbf{v}_1$, $A \mathbf{v}_2 = \lambda_2 \mathbf{v}_2$ etc., we obtain

$$A\mathbf{v} = c_1 \lambda_1 \mathbf{v}_1 + c_2 \lambda_2 \mathbf{v}_2 + \ldots + c_n \lambda_n \mathbf{v}_n$$

$$= \lambda_1 \left[c_1 \mathbf{v}_1 + c_2 \left(\frac{\lambda_2}{\lambda_1} \right) \mathbf{v}_2 + \cdots + c_n \left(\frac{\lambda_n}{\lambda_1} \right) \mathbf{v}_n \right]$$

plying by A again and simplifying we get

$$A^2 \mathbf{v} = \lambda_1^2 \left[c_1 \mathbf{v}_1 + c_2 \left(\frac{\lambda_2}{\lambda_1} \right)^2 \mathbf{v}_2 + \cdots + c_n \left(\frac{\lambda_n}{\lambda_1} \right)^2 \mathbf{v}_n \right]$$

$$\vdots$$

$$A^k \mathbf{v} = \lambda_1^k \left[c_1 \mathbf{v}_1 + c_2 \left(\frac{\lambda_2}{\lambda_1} \right)^k \mathbf{v}_2 + \cdots + c_n \left(\frac{\lambda_n}{\lambda_1} \right)^k \mathbf{v}_n \right] \tag{3.154}$$

$$\mathbf{A}^{k+1}\mathbf{v} = \lambda_1^{k+1}\left[c_1\mathbf{v}_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^{k+1}\mathbf{v}_2 + \cdots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^{k+1}\mathbf{v}_n\right] \tag{3.155}$$

As $k \to \infty$, the right hand sides of (3.154) and (3.155) tend to $\lambda_1^k c_1\mathbf{v}_1$ and $\lambda_1^{k+1}c_1\mathbf{v}_1$, since $|\lambda_i/\lambda_1| < 1$, $i = 2, 3, \ldots, n$. The vector $c_1\mathbf{v}_1 + c_2(\lambda_2/\lambda_1)^k\mathbf{v}_2 + \ldots + c_n(\lambda_n/\lambda_1)^k\mathbf{v}_n$ tends to $c_1\mathbf{v}_1$ which is the eigenvector corresponding to λ_1. The eigenvalue λ_1 is obtained as the ratio of the corresponding components of $\mathbf{A}^{k+1}\mathbf{v}$ and $\mathbf{A}^k\mathbf{v}$

$$\lambda_1 = \lim_{k\to\infty}\frac{\left(\mathbf{A}^{k+1}\mathbf{v}\right)_r}{\left(\mathbf{A}^k\mathbf{v}\right)_r}, \quad r = 1, 2, \ldots, n \tag{3.156}$$

where the suffix r denotes the rth component of the vector.

The iteration is stopped when the magnitudes of the differences of the ratios are less than the given error tolerance.

If $|\lambda_2| << |\lambda_1|$, then faster convergence is obtained. In order to keep the round-off error in control, we normalize (such that the largest element is unity) the vector before premultiplying by $\mathbf{A}$. We can write the method as follows. Let $\mathbf{v}_0$ be an non-zero arbitrary initial vector (non-orthogonal to $\mathbf{v}_1$) and find

$$\mathbf{y}_{k+1} = \mathbf{A}\mathbf{v}_k \tag{3.157}$$

$$\mathbf{v}_{k+1} = \mathbf{y}_{k+1}/m_{k+1} \tag{3.158}$$

where m_{k+1} is the largest element in magnitude of $\mathbf{y}_{k+1}$. Then

$$\lambda_1 = \lim_{k\to\infty}\frac{\left(\mathbf{y}_{k+1}\right)_r}{\left(\mathbf{v}_k\right)_r}, \quad r = 1, 2, \ldots, n \tag{3.159}$$

and $\mathbf{v}_{k+1}$ is the required eigenvector. It may be noted that as $k \to \infty$, m_{k+1} also gives λ_1. The initial vector is usually chosen as a vector with all components equal to unity (non-orthogonal to $\mathbf{v}_1$), if no suitable approximation is available.

Example 3.42 Find the largest eigenvalue in modulus and the corresponding eigenvector of the matrix

$$\mathbf{A} = \begin{bmatrix} -15 & 4 & 3 \\ 10 & -12 & 6 \\ 20 & -4 & 2 \end{bmatrix}$$

using the power method.

We start the iteration using the unit vector as the initial vector

$$\mathbf{v}_0 = [1, 1, 1]^T.$$

Chapter 3

We find

$$y_1 = [-8, 4, 18]^T, \quad v_1 = \left[-\frac{4}{9}, \frac{2}{9}, 1\right]^T$$

$$y_2 = \left[\frac{95}{9}, -\frac{10}{9}, -\frac{70}{9}\right]^T, \quad v_2 = \left[1, -\frac{2}{19}, -\frac{14}{19}\right]^T, \ldots,$$

$y_7 = [-19.8674, 9.7072, 19.8524]^T$, $v_7 = [-1.0, 0.4886, 0.9992]^T$

$y_8 = [19.952, -9.868, -19.956]^T$, $v_8 = [0.9998, -0.494, -1]^T$

$y_9 = [-19.973, 9.926, 19.988]^T$

At this step, the approximations to the largest eigenvalue in modulus are

$$|\lambda| = 19.977, 20.093, 19.988.$$

If we round-off to 3 digits, we have $|\lambda| \doteq 20$.

The approximate eigenvector is $[0.9998, -0.494, -1]^T$.

The exact eigenvalue is -20 and its eigenvector is $[1, -0.5, -1]^T$.

Shift of Origin

Power method can be used with a shift of origin. We know that $\mathbf{A}$ and $\mathbf{A} - k\mathbf{I}$ have the same set of eigenvectors and for each eigenvalue λ_i of $\mathbf{A}$, we have, for $\mathbf{A} - k\mathbf{I}$, the eigenvalue $\lambda_i - k$.

$$\therefore (\mathbf{A} - k\mathbf{I})\mathbf{v} = \mathbf{A}\mathbf{v} - k\mathbf{v}$$

$$= \lambda\mathbf{v} - k\mathbf{v} = (\lambda - k)\mathbf{v}.$$

Therefore, if we subtract k from the diagonal elements of $\mathbf{A}$, then each eigenvalue is reduced by the same factor and the eigenvectors are not changed. The power method can now be used as

$$y_{k+1} = (\mathbf{A} - k\mathbf{I})v_k \tag{3.160}$$

$$v_{k+1} = y_{k+1}/m_{k+1}. \tag{3.161}$$

The dominant eigenvalue is then obtained using (3.159). Regardless of the value we choose, the dominant eigenvalue of $\mathbf{A} - k\mathbf{I}$ will always be either $\lambda_1 - k$ or $\lambda_n - k$. For example, if 20, 10, 1 are the eigenvalues of a system, then if we choose, say $k = 15$, then the new eigenvalues are 5, -5 and -14. Hence, the power method applied on the new system would give the largest eigenvalue of the new system and the corresponding eigenvector, which is the smallest eigenvalue and its eigenvector respectively of the original system. If we choose $k = (\lambda_2 + \lambda_n)/2$ we have maximum rate of convergence to $\lambda_1 - k$, if we use $\mathbf{A} - k\mathbf{I}$ as the iteration matrix. If we choose $k = (\lambda_1 + \lambda_{n-1})/2$, then we have maximum rate of convergence to $\lambda_n - k$. In the above example, if we choose $k = (10 + 1)/2 = 5.5$, then the eigenvalues of the new system are 14.5, 4.5, -4.5 and the ratios of $|\lambda_2/\lambda_1|$ and $|\lambda_3/\lambda_1|$ are approximately 0.31, while these ratios for the original system are 0.5, 0.05. All other choices of k would produce larger values for the ratio $|\lambda_2/\lambda_1|$. The convergence is faster for the new system than the original system, both the systems converging to the same eigenvalue. If we choose $k = (20 + 10)/2 = 15$, then the eigenvalues for the new system are 5,

and −14. The ratios of $|\lambda_2/\lambda_1|$ and $|\lambda_3/\lambda_1|$ for the new system are approximately
convergence to the largest eigenvalue is maximum for this value of k and this eigenvalue corresp
the smallest eigenvalue of the original system. Of course, the choice of k is difficult unless we
a priori an estimate of the eigenvalues.

3.12 INVERSE POWER METHOD

The inverse power method is usually more powerful than the power method. The method has
advantage that it can give approximation to any eigenvalue rather than only to λ_1 or λ_n. If λ
eigenvalue of A and v is the corresponding eigenvector, then $1/\lambda$ is an eigenvalue of
corresponding to the same eigenvector. Choose any nonzero eigenvector $\mathbf{y}_0 \in R^n$ and express it
linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. Applying the power method on $\mathbf{A}^{-1}$, we have

$$\mathbf{z}_{k+1} = \mathbf{A}^{-1}\,\mathbf{y}_k \tag{3.16}$$

$$\mathbf{y}_{k+1} = \mathbf{z}_{k+1}/m_{k+1}. \tag{3.16}$$

This gives an approximation to the dominant eigenvalue in modulus of $\mathbf{A}^{-1}$, that is, the smal
eigenvalue of A in modulus. However, one need not find $\mathbf{A}^{-1}$ to find the smallest eigenvalue
modulus) of A. We write equation (3.162) as

$$\mathbf{A}\mathbf{z}_{k+1} = \mathbf{y}_k. \tag{3.16}$$

We find $\mathbf{z}_{k+1}$ by solving the linear system of algebraic equations (3.164). Normalization is done
according to (3.163). The coefficient matrix for all iterations is the same.

If we introduce a shift of the origin we have

$$\mathbf{z}_{k+1} = (\mathbf{A} - k\,\mathbf{I})^{-1}\,\mathbf{y}_k. \tag{3.16}$$

The ratio of the corresponding components tends to $1/(\lambda_i - k)$, where λ_i are the eigenvalues of A.

$$\frac{1}{\lambda_i - k} = \lim_{k\to\infty} \frac{(\mathbf{z}_{k+1})_r}{(\mathbf{y}_k)_r}. \tag{3.16}$$

By carefully selecting k, one can find an approximation to any eigenvalue of A. For example, again, let 20,
10, 1 be the eigenvalues of A and choose, say $k = 9$. Then, the eigenvalues of $\mathbf{A} - k\,\mathbf{I}$ are 11, 1,
8 and that of $(\mathbf{A} - k\,\mathbf{I})^{-1}$ are 1/11, 1, − 1/8. The power method applied to $(\mathbf{A} - k\,\mathbf{I})^{-1}$ gives the
dominant eigenvalue of the new system which is 1.

The system (3.165) can be written as

$$(\mathbf{A} - k\,\mathbf{I})\,\mathbf{z}_{k+1} = \mathbf{y}_k. \tag{3.167}$$

We find $\mathbf{z}_{k+1}$ by solving the linear system of equations with the right hand side changing at each
iteration. We normalize the vectors at each stage of iteration. It is known that this inverse iteration
is the most powerful and accurate of all methods for computing eigenvectors.

Example 3.43 Find the eigenvalue nearest to 3 for the matrix

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

I
11-compute.

x	$f(x)$	x	$f(x)$
0.1	20.02502	0.6	3.48692
0.2	10.05013	0.7	3.03787
0.3	6.74211	0.8	2.70861
0.4	5.10105	0.9	2.45959
0.5	4.12706	1.0	2.26712

(Bergen Univ., Sweden, BIT 24(1984), 397)

The graph of a function f is almost parabolic segment attaining its extreme values in an interval (x_1, x_2). The function values $f_i = f(x_i)$ are known at the equidistant abscissas x_0, x_1, x_2. The extreme value is searched. Use the quadratic interpolation to derive x coordinate of the extremum.

(Royal Inst. Tech., Stockholm, Sweden, BIT 26(1986), 135)

In the following problems, find the maximum value of the uniform mesh size h that can be used to tabulate $f(x)$ on $[a, b]$, using cubic interpolation such that $|\text{Error}| \leq \varepsilon$.

(i) $f(x) = e^x$, $[a, b] = [1, 2.5]$, $\varepsilon = 1 \times 10^{-4}$.

(ii) $f(x) = \cos 2x$, $[a, b] = [0, \pi/4]$, $\varepsilon = 1 \times 10^{-6}$.

(iii) $f(x) = xe^x$, $[a, b] = [1, 2]$, $\varepsilon = 5 \times 10^{-5}$.

UNIT - III start.

HERMITE INTERPOLATION

Chapter 4

The Hermite interpolating polynomial interpolates not only the function $f(x)$ but also its (certain order) derivatives at a given set of tabular points. The simple interpolating conditions are given in (4.6). We now give an explicit expression for the interpolating polynomial satisfying (4.6), that is

$$P(x_i) = f(x_i)$$
$$P'(x_i) = f'(x_i), \; i = 0, 1, \cdots, n.$$

Since there are $2n + 2$ conditions to be satisfied, $P(x)$ must be a polynomial of degree $\leq 2n + 1$. The required polynomial may be written as

$$P(x) = \sum_{i=0}^{n} A_i(x) f(x_i) + \sum_{i=0}^{n} B_i(x) f'(x_i) \tag{4.55}$$

where $A_i(x)$ and $B_i(x)$ are polynomials of degree $\leq 2n + 1$ and satisfy

(i) $A_i(x_j) = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$

(ii) $A_i'(x_j) = 0$ for all i and j

(iii) $B_i(x_j) = 0$ for all i and j

(iv) $B_i'(x_j) = \begin{cases} 0, & i \neq j \\ 1, & i = j. \end{cases}$ (4.56)

Using the Lagrange fundamental polynomials $l_i(x)$, we write

$$\begin{aligned} A_i(x) &= \gamma_i(x)\ l_i^2(x) \\ B_i(x) &= \delta_i(x)\ l_i^2(x). \end{aligned} \tag{4}$$

Since $l_i^2(x)$ is a polynomial of degree $2n$, $\gamma_i(x)$ and $\delta_i(x)$ must be linear polynomials. Let

$$\begin{aligned} \gamma_i(x) &= a_i x + b_i \\ \delta_i(x) &= c_i x + d_i. \end{aligned} \tag{4}$$

Using the conditions (4.56), we obtain

$$\begin{aligned} a_i &= -2\, l_i'(x_i), \\ b_i &= 1 + 2x_i\, l_i'(x_i) \\ c_i &= 1 \text{ and } d_i = -x_i. \end{aligned} \tag{4}$$

Substituting (4.59) into (4.58) and using (4.57), the equation (4.55) becomes

$$P(x) = \sum_{i=0}^{n} [1 - 2(x - x_i)\ l_i'(x_i)]\ l_i^2(x)\ f(x_i) + \sum_{i=0}^{n} (x - x_i)\ l_i^2(x)\ f'(x_i) \tag{4}$$

which is called the **Hermite interpolating polynomial**. It is easy to verify that

$$l_i'(x_i) = \frac{w''(x_i)}{2w'(x_i)}.$$

Alternative

From the conditions (4.56), we find that the Lagrange fundamental polynomial $l_i^2(x)$ is a factor of $A_i(x)$ and $B_i(x)$. Since $l_i^2(x)$ is a polynomial of degree $2n$, we write

$$\begin{aligned} A_i(x) &= [a_i + b_i\ (x - x_i)]\ l_i^2(x) \\ B_i(x) &= [c_i + d_i\ (x - x_i)]\ l_i^2(x). \end{aligned}$$

We have

$$\begin{aligned} A_i'(x) &= b_i\ l_i^2(x) + [a_i + b_i(x - x_i)]\ [2l_i(x)\ l_i'(x)] \\ B_i'(x) &= d_i\ l_i^2(x) + [c_i + d_i(x - x_i)]\ [2l_i(x)\ l_i'(x)] \end{aligned}$$

Substituting in the given conditions (4.56), we obtain

$A_i(x_i) = 1$, gives $a_i = 1$

$A_i'(x_j) = 0$, is satisfied for all $i \neq j$,

$A_i'(x_i) = 0$, gives $b_i + 2a_i\ l_i'(x_i) = 0$ or $b_i = -2l_i'(x_i)$,

$B_i(x_j) = 0$, is satisfied for all $i \neq j$,

$B_i(x_i) = 0$, gives $c_i = 0$

$B_i'(x_j) = 0$, is satisfied for all $i \neq j$

$B_i'(x_i) = 1$, gives $d_i = 1$.

$$\boxed{\begin{aligned} A_i(x) &= [1 - 2(x - x_i)\, l_i'(x_i)]\, l_i^2(x) \\ B_i(x) &= (x - x_i)\, l_i^2(x) \end{aligned}}$$

which are the same expressions as given in Eq. (4.60).

The truncation error associated with (4.60) can be written as

$$E_{2n+1}(f; x) = \frac{w^2(x)}{(2n+2)!} f^{(2n+2)}(\xi), \; x_0 < \xi < x_n.$$

Example 4.19 Determine the parameters in the formula

$$P(x) = a_0(x - a)^3 + a_1(x - a)^2 + a_2(x - a) + a_3$$

such that

$$P(a) = f(a), \; P'(a) = f'(a)$$
$$P(b) = f(b), \; P'(b) = f'(b).$$

Using the interpolatory conditions in the given formula we obtain

$$P(a) = f(a) = a_3$$
$$P'(a) = f'(a) = a_2$$
$$P(b) = f(b) = a_0(b - a)^3 + a_1(b - a)^2 + a_2(b - a) + a_3$$
$$P'(b) = f'(b) = 3a_0(b - a)^2 + 2a_1(b - a) + a_2.$$

Solving the above system of equations, we get

$$a_3 = f(a)$$
$$a_2 = f'(a)$$
$$a_1 = \frac{3}{(b-a)^2}[f(b) - f(a)] - \frac{1}{(b-a)}[2f'(a) + f'(b)]$$
$$a_0 = \frac{2}{(b-a)^3}[f(a) - f(b)] + \frac{1}{(b-a)^2}[f'(a) + f'(b)]$$

Example 4.20 Given the following values of $f(x)$ and $f'(x)$

x	$f(x)$	$f'(x)$
-1	1	-5
0	1	1
1	3	7

Estimate the values of $f(-0.5)$ and $f(0.5)$ using the Hermite interpolation. The exact values are $f(-0.5) = 33/64$ and $f(0.5) = 97/64$.

Here

$$n = 2, \; x_0 = -1, \; x_1 = 0 \text{ and } x_2 = 1.$$

$$P(x) = \sum_{i=0}^{2} A_i(x) f(x_i) + \sum_{i=0}^{2} B_i(x) f'(x_i)$$

$$A_i(x) = [1 - 2(x - x_i) l_i'(x_i)]\, l_i^2(x)$$
$$B_i(x) = (x - x_i)\, l_i^2(x)$$

$$A_0(x) = [1 - 2(x - x_0)\, l'_0(x_0)]\, l_0^2(x)$$
$$A_1(x) = [1 - 2(x - x_1)\, l'_1(x_1)]\, l_1^2(x)$$
$$A_2(x) = [1 - 2(x - x_2)\, l'_2(x_2)]\, l_2^2(x)$$
$$B_0(x) = (x - x_0)\, l_0^2(x)$$
$$B_1(x) = (x - x_1)\, l_1^2(x)$$
$$B_2(x) = (x - x_2)\, l_2^2(x)$$

$$l_0(x) = \frac{(x-0)(x-1)}{(-1-0)(-1-1)} = \frac{x(x-1)}{2}, \quad l'_0(-1) = -\frac{3}{2}$$

$$l_1(x) = \frac{(x+1)(x-1)}{(0+1)(0-1)} = -(x^2 - 1), \quad l'_1(0) = 0$$

$$l_2(x) = \frac{(x+1)(x-0)}{(1+1)(1-0)} = \frac{x(x+1)}{2}, \quad l'_2(1) = \frac{3}{2}$$

$$A_0(x) = [1 + 3(x+1)]\, \frac{x^2(x-1)^2}{4}$$
$$= \frac{1}{4}(3x^5 - 2x^4 - 5x^3 + 4x^2)$$

$$A_1(x) = [1 - 2(x-0)(0)]\, (x^2 - 1)^2$$
$$= x^4 - 2x^2 + 1$$

$$A_2(x) = [1 - 3(x-1)]\, \frac{x^2(x+1)^2}{4}$$
$$= \frac{1}{4}(-3x^5 - 2x^4 + 5x^3 + 4x^2)$$

$$B_0(x) = \frac{(x+1)x^2(x-1)^2}{4}$$
$$= \frac{1}{4}(x^5 - x^4 - x^3 + x^2)$$

$$B_1(x) = x(x^2 - 1)^2 = x^5 - 2x^3 + x$$

$$B_2(x) = \frac{(x-1)x^2(x+1)^2}{4}$$
$$= \frac{1}{4}(x^5 + x^4 - x^3 - x^2).$$

Thus, we obtain

$$P(x) = \frac{1}{4}(3x^5 - 2x^4 - 5x^3 + 4x^2)(1) + (x^4 - 2x^2 + 1)(1)$$
$$+ \frac{1}{4}(-3x^5 - 2x^4 + 5x^3 + 4x^2)(3)$$

$$+ \frac{1}{4}(x^5 - x^4 - x^3 + x^2)(-5) + (x^5 - 2x^3 + x)(1)$$

$$\vdots$$

$$+ \frac{1}{4}(x^5 + x^4 - x^3 - x^2)(7)$$

$$= 2x^4 - x^2 + x + 1$$

ing $x = -0.5$ and 0.5, we get

$f(-0.5) = 3/8$, exact value is 33/64

$f(0.5) = 11/8$, exact value is 97/64.

PIECEWISE AND SPLINE INTERPOLATION

ain reasonably accurate results using interpolation, we may have to use polynomials of high . With polynomials of high degrees, beyond a certain order, not only the computation becomes , but also the computed results become unreliable because of roundoff errors. In order to keep e of interpolating polynomials small and also to achieve accurate results, we use **piecewise** olation. We subdivide the given interval $[a, b]$ into a number of **subintervals** I_i, $i = 1, 2, \cdots, n$ and approximate the function by some lower degree polynomial in each val.

, we subdivide the given interval $[a, b]$, where $a = x_0 < x_1 < x_2 < \cdots < x_n = b$, into a number overlapping subintervals each containing 2 or 3 or 4 nodal points. Then, we construct the onding linear or quadratic or cubic interpolating polynomials fitting the given data on each rval. These polynomials define the piecewise linear or quadratic or cubic interpolating mials respectively. For example, for the data $(x_i, f(x_i))$, $i = 0, 1, \cdots, 6$ we can construct the ing piecewise linear or quadratic or cubic polynomials.

x_0 x_1 x_2 x_3 x_4 x_5 x_6

$f(x_0)$ $f(x_1)$ $f(x_2)$ $f(x_3)$ $f(x_4)$ $f(x_5)$ $f(x_6)$ Piecewise linear polynomial

Piecewise quadratic polynomial

Piecewise cubic polynomial

ewise Linear Interpolation

have $n + 1$ distinct nodal points $x_0, x_1, \cdots, x_n$ and we want to determine an interpolating ynomial as shown in Fig. 4.2. The interpolating polynomial is linear in each subinterval $_{-1}, x_i)$ and it agrees with the function $f(x)$ at the $n + 1$ nodal points. The subintervals or the line ments are called **finite elements** in one space dimension and the nodal points are called knots. ing the linear Lagrange interpolating polynomial (4.14), we have for $x \in [x_{i-1}, x_i]$, the piecewise ar interpolating polynomial

Chapter 4

$$P_{i,1}(x) = \frac{x - x_i}{x_{i-1} - x_i} f(x_{i-1}) + \frac{x - x_{i-1}}{x_i - x_{i-1}} f(x_i), \quad i = 1, 2, \cdots, n$$

For $x \in [x_i, x_{i+1}]$, we have

$$P_{i+1,1}(x) = \frac{x - x_{i+1}}{x_i - x_{i+1}} f(x_i) + \frac{x - x_i}{x_{i+1} - x_i} f(x_{i+1}).$$

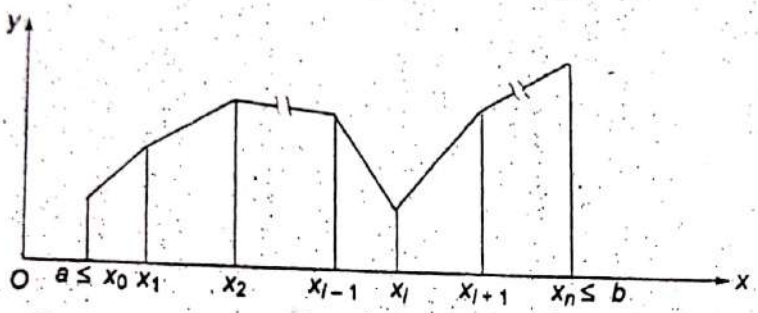

Fig. 4.2. A piecewise linear Lagrange interpolation.

Define

$$N_i(x) = \begin{cases} 0, & x \le x_{i-1} \\ \dfrac{x - x_{i-1}}{x_i - x_{i-1}}, & x_{i-1} \le x \le x_i \\ \dfrac{x_{i+1} - x}{x_{i+1} - x_i}, & x_i \le x \le x_{i+1} \\ 0, & x \ge x_{i+1}. \end{cases}$$

Note that the non-zero terms in $N_i(x)$ are the coefficients of $f(x_i)$ in $P_{i,1}(x)$ and $P_{i+1,1}(x$ respectively. Then, the interpolating polynomial

$$P(x) = \sum_{i=0}^{n} P_{i,1}(x)$$

which agrees with $f(x)$ at x_i, $i = 0, 1, \cdots, n$ and is linear in each subinterval $[x_{i-1}, x_i]$ can be

$$P_1(x) = \sum_{i=0}^{n} N_i(x) f(x_i).$$

The function $N_i(x)$ is called a **shape function** and it is shown in Fig. 4.3.

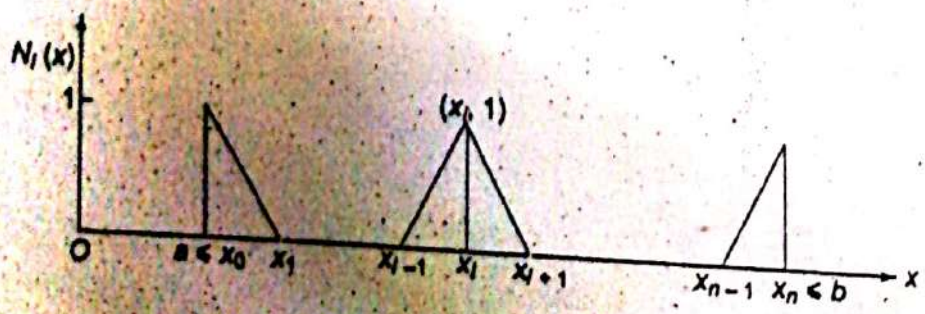

Fig. 4.3. Shape function $N_i(x)$.

...or in the piecewise linear interpolation is given by

$$f(x) - P_{i,1}(x) = \frac{1}{2!}(x - x_{i-1})(x - x_i) f''(\xi_i), \quad x_{i-1} < \xi_i < x_i.$$

...ple 4.21 Obtain the piecewise linear interpolating polynomials for the function $f(x)$ def...

x	1	2	4	8
$f(x)$	3	7	21	73

...ence, estimate the values of $f(3)$ and $f(7)$.

...the interval [1, 2], we have

$$P_1(x) = \frac{x-2}{(-1)}(3) + (x-1)(7) = 4x - 1.$$

...the interval [2, 4], we have

$$P_1(x) = \frac{x-4}{-2}(7) + \frac{(x-2)}{2}(21) = 7x - 7.$$

...the interval [4, 8], we have

$$P_1(x) = \frac{x-8}{-4}(21) + \frac{x-4}{4}(73) = 13x - 31.$$

Hence, the piecewise linear interpolating polynomials are given as

$$P_1(x) = \begin{cases} 4x-1, & 1 \le x \le 2 \\ 7x-7, & 2 \le x \le 4 \\ 13x-31, & 4 \le x \le 8. \end{cases}$$

Using the polynomial in the interval [2, 4], we obtain $f(3) = 21 - 7 = 14$. Using the polynomial ...the interval [4, 8], we obtain $f(7) = 91 - 31 = 60$.

...ecewise Quadratic Interpolation

...t the number of distinct nodal points be $2n + 1$ with $a = x_0 < x_1 < x_2 < \cdots < x_{2n} = b$. We consider ...groups of three consecutive nodal points as $[x_0, x_2]$, $[x_2, x_4]$, $\cdots$, $[x_{i-1}, x_{i+1}]$, $\cdots$, $[x_{2n-2}, x_{2n}]$. On ...of the subintervals, we write the quadratic interpolating polynomial. For $x \in [x_{i-1}, x_{i+1}]$, we have ...quadratic interpolating polynomial

$$P_{i,2}(x) = \frac{(x - x_i)(x - x_{i+1})}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} f(x_{i-1}) + \frac{(x - x_{i-1})(x - x_{i+1})}{(x_i - x_{i-1})(x_i - x_{i+1})} f(x_i) + \frac{(x - x_{i-1})(x - x_i)}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)} f(x_{i+1}) \tag{4.67}$$

the piecewise quadratic interpolation is given by

$$= \frac{1}{3!}(x - x_{i-1})(x - x_i)(x - x_{i+1}) f'''(\xi_i),\ x_{i-1} < \xi_i < x_{i+1}.$$

Obtain the piecewise quadratic interpolating polynomials for the functi

the data

-3	-2	-1	1	3	6	7
369	222	171	165	207	990	1779

and an approximate value of $f(-2.5)$ and $f(6.5)$.

sider the groups of nodal points as $\{-3, -2, -1\}$, $\{-1, 1, 3\}$, $\{3, 6, 7\}$. On each of these we write the quadratic interpolating polynomial. We have the follo

ions.

$-1]$, we have

$$= \frac{(x+2)(x+1)}{(-3+2)(-3+1)}(369) + \frac{(x+3)(x+1)}{(-2+3)(-2+1)}(222)$$

$$+ \frac{(x+3)(x+2)}{(-1+3)(-1+2)}(171)$$

$$= \frac{369}{2}(x^2 + 3x + 2) - 222(x^2 + 4x + 3) + \frac{171}{2}(x^2 + 5x + 6)$$

$$= 48x^2 + 93x + 4$$

$-1, 3]$, we have

$$P_{1,2}(x) = \frac{(x-1)(x-3)}{(-1-1)(-1-3)}(171) + \frac{(x+1)(x-3)}{(1+1)(1-3)}(165) + \frac{(x+1)(x-1)}{(3+1)(3-1)}(207)$$

$$= \frac{171}{8}(x^2 - 4x + 3) - \frac{165}{4}(x^2 - 2x - 3) + \frac{207}{8}(x^2 - 1)$$

$$= 6x^2 - 3x + 162.$$

$3, 7]$, we have

$$P_{3,3}(x) = \frac{(x-6)(x-7)}{(3-6)(3-7)}(207) + \frac{(x-3)(x-7)}{(6-3)(6-7)}(990) + \frac{(x-3)(x-6)}{(7-3)(7-6)}(1779)$$

$$= \frac{207}{12}(x^2 - 13x + 42) - \frac{990}{3}(x^2 - 10x + 21) + \frac{1779}{4}(x^2 - 9x + 18)$$

$$= 132x^2 - 927x + 1800.$$

$x = -2.5$ lies in the interval $[-3, -1]$. Hence, using $P_{2,1}(x)$ we obtain

$$f(-2.5) \approx P_{2,1}(-2.5) = 48(-2.5)^2 + 93(-2.5) + 216 = 283.5.$$

The point 6.5 lies in the interval $[3, 7]$, we obtain

$$f(6.5) \approx P_{2,3}(6.5) = 132(6.5)^2 - 927(6.5) + 1800 = 1351.5.$$

Cubic Interpolation

Let the number of distinct nodal points be $3n + 1$, with $a = x_0 < x_1 < x_2 < \cdots < x_{3n} = b$. We consider the set of four nodal points as $[x_0, \cdots, x_3]$, $[x_3, \cdots, x_6]$, $\cdots$, $[x_{3n-3}, \cdots, x_{3n}]$. On each of the intervals, we write the cubic interpolating polynomial. For $x \in [x_i, x_{i+3}]$, we have the cubic interpolating polynomial

$$P_{i,3}(x) = l_{i,0} f(x_i) + l_{i,1} f(x_{i+1}) + l_{i,2} f(x_{i+2}) + l_{i,3} f(x_{i+3}) \tag{4.68}$$

$$l_{i,0} = \frac{(x - x_{i+1})(x - x_{i+2})(x - x_{i+3})}{(x_i - x_{i+1})(x_i - x_{i+2})(x_i - x_{i+3})},$$

$$l_{i,1} = \frac{(x - x_i)(x - x_{i+2})(x - x_{i+3})}{(x_{i+1} - x_i)(x_{i+1} - x_{i+2})(x_{i+1} - x_{i+3})},$$

$$l_{i,2} = \frac{(x - x_i)(x - x_{i+1})(x - x_{i+3})}{(x_{i+2} - x_i)(x_{i+2} - x_{i+1})(x_{i+2} - x_{i+3})},$$

$$l_{i,3} = \frac{(x - x_i)(x - x_{i+1})(x - x_{i+2})}{(x_{i+3} - x_i)(x_{i+3} - x_{i+1})(x_{i+3} - x_{i+2})}$$

The error in the piecewise cubic interpolation is given by

$$f(x) - P_{i,3}(x) = \frac{1}{4!}(x - x_i)(x - x_{i+1})(x - x_{i+2})(x - x_{i+3}) f^{iv}(\xi_i),$$

$$x_i < \xi_i < x_{i+3}.$$

Alternately, we can use the Newton's divided difference interpolation to obtain the interpolating polynomial.

Example 4.23 Using the data given in Example 4.22, obtain the approximate values of $f(-2.5)$ and $f(6.5)$ using the piecewise cubic interpolation.

We consider the groups of nodal points as $\{-3, -2, -1, 1\}$ and $\{1, 3, 6, 7\}$.

Chapter 4

We shall use the Newton's divided difference interpolation for these two sets of points.

x	$f(x)$	first d.d.	second d.d.	third d.d.
− 3	369			
		− 147		
− 2	222		48	
		− 51		− 8
− 1	171		16	
		− 3		
1	165			

Therefore, on [− 3, 1], we have

$$P_{1,3} = 369 + (x + 3)(-147) + (x + 3)(x + 2)(48) + (x + 3)(x + 2)(x + 1)(-8)$$
$$= -8x^3 + 5x + 168$$

and $f(-2.5) \approx P_{1,3}(-2.5) = -8(-2.5)^3 + 5(-2.5) + 168 = 280.5$.

On [1, 7], we have the following difference table.

x	$f(x)$	first d.d.	second d.d.	third d.d.
1	165			
		21		
3	207		48	
		261		14
6	990		132	
		789		
7	1779			

Hence, $P_{2,3} = 165 + (x - 1)(21) + (x - 1)(x - 3)(48) + (x - 1)(x - 3)(x - 6)(14)$

$$= 14x^3 - 92x^2 + 207x + 36$$

and $f(6.5) = 14(6.5)^3 - 92(6.5)^2 + 207(6.5) + 36 = 1339.25$.

Piecewise Cubic Interpolation using Hermite Type Data

Let the following Hermite type of data be given on each subinterval $[x_{i-1}, x_i]$, $i = 1, 2, \cdots, n$

$$P_{i,3}(x_{i-1}) = f_{i-1}, \qquad P_{i,3}(x_i) = f_i$$
$$P'_{i,3}(x_{i-1}) = f'_{i-1}, \qquad P'_{i,3}(x_i) = f'_i. \qquad (4$$

Then, we can construct a cubic polynomial $P_{i,3}(x)$ on each of the sub intervals. The polyno thus obtained is called **piecewise cubic Hermite interpolating polynomial**. Using (4.60), we write this polynomial in the form

$$P_{i,3}(x) = A_{i-1}(x) f_{i-1} + A_i(x) f_i + B_{i-1}(x) f'_{i-1} + B_i(x) f'_i \qquad (4$$

Hence, we obtain

$$P_1(x) = -4x^3 + 5x + 1, \quad 0 \le x \le 1$$

$$P_2(x) = 50x^3 - 162x^2 + 167x - 53, \quad 1 \le x \le 2$$

$$P_3(x) = -46x^3 + 414x^2 - 985x + 715, \quad 2 \le x \le 3.$$

BIVARIATE INTERPOLATION

The problem of polynomial interpolation for functions of several independent variables is quite important. For the sake of simplicity, we shall only consider functions of two variables, the extension to higher dimensions is straightforward.

Lagrange Bivariate Interpolation

Let $f(x, y)$ be defined at $(m + 1)(n + 1)$ distinct points (x_i, y_j), $i = 0, 1, \cdots, m$, $j = 0, 1, \cdots, n$ and denote $f(x_i, y_j)$ by $f_{i,j}$. We want to obtain a polynomial $P(x, y)$ of degree at most m in x and n in y, such that

$$P(x_i, y_j) = f_{i,j}, \quad i = 0(1)m, \; j = 0(1)n. \tag{4.117}$$

Using the Lagrange fundamental polynomials (4.28) of a single variable, we define

$$X_{m,i}(x) = \frac{w(x)}{(x - x_i)\, w'(x_i)}, \qquad i = 0, 1, \cdots, m$$

$$Y_{n,j}(y) = \frac{w^*(y)}{(y - y_j)\, w^{*\prime}(y_j)}, \qquad j = 0, 1, \cdots, n$$

where

$$w(x) = (x - x_0)(x - x_1) \cdots (x - x_m)$$
$$w^*(y) = (y - y_0)(y - y_1) \cdots (y - y_n).$$

Obviously, $X_{m,i}(x)$ and $Y_{n,j}(y)$ are polynomials of degree m in x, and n in y respectively. These polynomials satisfy the following properties

$$X_{m,i}(x_k) = \delta_{ik}, \; Y_{n,j}(y_k) = \delta_{jk}.$$

Thus, the polynomial which satisfies (4.117) can be written as

$$P_{m,n}(x, y) = \sum_{i=0}^{m} \sum_{j=0}^{n} X_{m,i}(x)\, Y_{n,j}(y)\, f_{i,j}. \tag{4.118}$$

This polynomial is called the **Lagrange bivariate interpolating** polynomial. It may also be interpreted as double application of the Lagrange interpolating polynomial in a single variable.

Interpolation for Equispaced Points

...points, with spacing h in x and k in y, we define

$$\Delta_x f(x, y) = f(x + h, y) - f(x, y)$$
$$= (E_x - 1) f(x, y)$$

$$\Delta_y f(x, y) = f(x, y + k) - f(x, y)$$
$$= (E_y - 1) f(x, y)$$

$$\Delta_{xx} f(x, y) = \Delta_x f(x + h, y) - \Delta_x f(x, y)$$
$$= (E_x - 1)^2 f(x, y)$$

$$\Delta_{yy} f(x, y) = \Delta_y f(x, y + k) - \Delta_y f(x, y)$$
$$= (E_y - 1)^2 f(x, y)$$

$$\Delta_{xy} f(x, y) = \Delta_x [f(x, y + k) - f(x, y)] = \Delta_x \Delta_y f(x, y)$$
$$= (E_x - 1)(E_y - 1) f(x, y)$$
$$= (E_y - 1)(E_x - 1) f(x, y)$$
$$= \Delta_y \Delta_x f(x, y) = \Delta_{yx} f(x, y)$$

$$f(x_0 + mh, y_0 + nk) = E_x^m E_y^n f(x_0, y_0)$$
$$= (1 + \Delta_x)^m (1 + \Delta_y)^n f(x_0, y_0)$$
$$= \left[1 + \binom{m}{1}\Delta_x + \binom{m}{2}\Delta_{xx} + \cdots\right] \times$$
$$\left[1 + \binom{n}{1}\Delta_y + \binom{n}{2}\Delta_{yy} + \cdots\right] f(x_0, y_0)$$
$$= \left[1 + \binom{m}{1}\Delta_x + \binom{n}{1}\Delta_y + \binom{m}{2}\Delta_{xx} + \binom{m}{1}\binom{n}{1}\Delta_{xy} + \binom{n}{2}\Delta_{yy} + \cdots\right] f(x_0, y_0)$$

Let $x = x_0 + mh$ and $y = y_0 + nk$. Hence, $m = (x - x_0)/h$ and $n = (y - y_0)/k$.
Then, from (4.119) we have the interpolating polynomial

$$P(x, y) = f(x_0, y_0) + \left[\frac{1}{h}(x - x_0)\Delta_x + \frac{1}{k}(y - y_0)\Delta_y\right] f(x_0, y_0)$$
$$+ \frac{1}{2!}\left[\frac{1}{h^2}(x - x_0)(x - x_1)\Delta_{xx} + \frac{2}{hk}(x - x_0)(y - y_0)\Delta_{xy}\right.$$
$$\left. + \frac{1}{k^2}(y - y_0)(y - y_1)\Delta_{yy}\right] f(x_0, y_0) + \cdots$$

it called the Newton's bivariate interpolating polynomial for equispaced points.

4.28 The following data for a function $f(x, y)$ is given:

y\x	0	1	2
0	1	1.414214	
1	1.732051		

0.25, 0.75), using linear interpolation.

ar interpolating polynomial is given by

$$P(x, y) = f(x_0, y_0) + \frac{1}{h}(x - x_0)\Delta_x f(x_0, y_0) + \frac{1}{k}(y - y_0)\Delta_y f(x_0, y_0)$$

$$\Delta_x f(x_0, y_0) = f(x_0 + h, y_0) - f(x_0, y_0)$$
$$= 1.414214 - 1 = 0.414214$$

$$\Delta_y f(x_0, y_0) = f(x_0, y_0 + k) - f(x_0, y_0)$$
$$= 1.732051 - 1 = 0.732051.$$

find with $h = k = 1$

$$P(0.25, 0.75) = 1 + 0.25\ (0.414214) + 0.75\ (0.732051)$$
$$= 1.652592.$$

4.29 The following data for a function $f(x, y)$ is given

y\x	0	1	3
0	1	2	10
1	2	4	14
3	10	14	28

truct the bivariate interpolating polynomial and hence find $f(0.5, 0.5)$.

have

$$X_{2,i}(x) = \frac{(x - x_0)(x - x_1)(x - x_2)}{(x - x_i)w'(x_i)},$$

$$Y_{2,j}(y) = \frac{(y - y_0)(y - y_1)(y - y_2)}{(y - y_j)w^{*\prime}(x_j)}$$

$i = 0, 1, 2$ and $j = 0, 1, 2$.

$$X_{2,0}(x) = \frac{(x-1)(x-3)}{(-1)(-3)} = \frac{1}{3}(x^2 - 4x + 3),$$

$$X_{2,1}(x) = \frac{x(x-3)}{(1)(-2)} = -\frac{1}{2}(x^2 - 3x),$$

$$X_{2,2}(x) = \frac{x(x-1)}{(3)(2)} = \frac{1}{6}(x^2 - x),$$

$$Y_{2,0}(y) = \frac{(y-1)(y-3)}{(-1)(-3)} = \frac{1}{3}(y^2 - 4y + 3),$$

$$Y_{2,1}(y) = \frac{y(y-3)}{(1)(-2)} = -\frac{1}{2}(y^2 - 3y),$$

$$Y_{2,2}(y) = \frac{y(y-1)}{3(2)} = \frac{1}{6}(y^2 - y).$$

From (4.118), the second degree interpolating polynomial that fits the data is given by

$$\begin{aligned}P_{2,2}(x, y) &= X_{2,0}(x)\,[Y_{2,0}(y)\,f_{0,0} + Y_{2,1}\,f_{0,1} + Y_{2,2}\,f_{0,2}] \\ &\quad + X_{2,1}(x)\,[Y_{2,0}(y)\,f_{1,0} + Y_{2,1}\,f_{1,1} + Y_{2,2}\,f_{1,2}] \\ &\quad + X_{2,2}(x)\,[Y_{2,0}(y)\,f_{2,0} + Y_{2,1}\,f_{2,1} + Y_{2,2}\,f_{2,2}]\end{aligned}$$

$$\begin{aligned}&= \frac{1}{3}(x^2 - 4x + 3)\left[\frac{1}{3}(y^2 - 4y + 3)(1) - \frac{1}{2}(y^2 - 3y)(2)\right. \\ &\quad \left. + \frac{1}{6}(y^2 - y)(10)\right] - \frac{1}{2}(x^2 - 3x)\left[\frac{1}{3}(y^2 - 4y + 3)(2)\right. \\ &\quad \left. - \frac{1}{2}(y^2 - 3y)(4) + \frac{1}{6}(y^2 - y)(14)\right] \\ &\quad + \frac{1}{6}(x^2 - x)\left[\frac{1}{3}(y^2 - 4y + 3)(10) - \frac{1}{2}(y^2 - 3y)(14)\right. \\ &\quad \left. + \frac{1}{6}(y^2 - y)(28)\right] \\ &= \frac{1}{3}(x^2 - 4x + 3)(y^2 + 1) - \frac{1}{2}(x^2 - 3x)(y^2 + y + 2) \\ &\quad + \frac{1}{6}(x^2 - x)(y^2 + 3y + 10) \\ &= \frac{x^2}{6}(2y^2 + 2 - 3y^2 - 3y - 6 + y^2 + 3y + 10) \\ &\quad - \frac{x}{6}(8y^2 + 8 - 9y^2 - 9y - 18 + y^2 + 3y + 10) + y^2 + 1 \\ &= x^2 + xy + y^2 + 1.\end{aligned}$$

Hence, $f(0.5, 0.5) = 1.75.$

EXERCISE 4.2

1. A function $f(x)$ is approximated by the interpolating polynomial

$$P(x) = c_0 + c_1(x - 1) + c_2(x - 1)^2 + c_3(x - 1)^3,\ 1 \le x \le$$

Determine the parameters c_0, c_1, c_2 and c_3 such that

$$P(1) = f(1),\ P(2) = f(2),\ P'(1) = f'(1) \text{ and } P'(2) = f'(2).$$

$$= \int_0^1 [1 + \sin(8\pi x) + \frac{1}{8}\{1 - \cos(16\pi x)\}]dx$$

$$= \left[\frac{9}{8}x - \frac{1}{8\pi}\cos(8\pi x) - \frac{1}{128\pi}\sin(16\pi x)\right]_0^1 = \frac{9}{8}$$

or $$\|f_2\| = \frac{3}{2\sqrt{2}}.$$

$$\|f_3\|^2 = \int_0^1 \left[1 + \alpha e^{-\alpha^2 x}\right]^2 dx = \int_0^1 \left[1 + 2\alpha e^{-\alpha^2 x} + \alpha^2 e^{-2\alpha^2 x}\right] dx$$

$$= \left[x - \frac{2}{\alpha}e^{-\alpha^2 x} - \frac{1}{2}e^{-2\alpha^2 x}\right]_0^1 = 1 - \frac{2}{\alpha}(e^{-\alpha^2} - 1) - \frac{1}{2}(e^{-2\alpha^2} - 1)$$

$$= \frac{3}{2} + \frac{2}{\alpha} - \frac{2}{\alpha}e^{-\alpha^2} - \frac{1}{2}e^{-2\alpha^2} \approx 1.5 + 0.02 = 1.52, \text{ for } \alpha = 100.$$

or $$\|f_3\| \approx 1.233$$

L_∞ norm: $\|f_1\| = 1$.

$$\|f_2\| = \max_{0 \le x \le 1} \left|1 + \frac{1}{2}\sin(8\pi x)\right| = \frac{3}{2}$$

$$\|f_3\| = \max_{0 \le x \le 1} |1 + \alpha e^{-\alpha^2 x}| = 101, \text{ for } \alpha = 100.$$

4.9 LEAST SQUARES APPROXIMATION

Least squares approximations are the most commonly used approximations for approximat function $f(x)$ which may be given in tabular form or known explicitly over a given interval. I case, we use the Euclidean norm (4.124) or (4.127). The best approximation in the least square is defined as that for which the constants c_i, $i = 0, 1, \cdots, n$ are determined so that the aggrega $W(x)\,E^2$ over a given domain D is as small as possible, where $W(x) > 0$ is the weight function functions whose values are given at $N + 1$ points $x_0, x_1, \cdots, x_N$, we have

$$I(c_0, c_1, \cdots, c_n) = \sum_{k=0}^{N} W(x_k)\left[f(x_k) - \sum_{i=0}^{n} c_i \phi_i(x_k)\right]^2$$

$$= \text{minimum.} \qquad (4$$

For functions which are continuous on $[a, b]$ and are given explicitly, we have

$$I(c_0, c_1, \cdots, c_n) = \int_a^b W(x)\left[f(x) - \sum_{i=0}^{n} c_i \phi_i(x)\right]^2 dx$$

$$= \text{minimum.} \qquad (4$$

The coordinate functions $\phi_i(x)$ are usually chosen as

$$\phi_i(x) = x^i,\ i = 0, 1, \cdots, n$$

and $W(x) = 1$. The necessary conditions for (4.129) or (4.130) to have a minimum value is th

$$\frac{\partial I}{\partial c_i} = 0,\ i = 0,\ 1,\ \cdots,\ n. \tag{4.131}$$

This gives a system of $n + 1$ linear equations in $n + 1$ unknowns $c_0, c_1, \cdots, c_n$. These equations are called **normal equations**. The normal equations for (4.129) and (4.130) become, respectively

$$\sum_{k=0}^{N} W(x_k)\left[f(x_k) - \sum_{i=0}^{n} c_i\,\phi_i(x_k)\right]\phi_j(x_k) = 0,\ j = 0(1)n \tag{4.132}$$

$$\int_a^b W(x)\left[f(x) - \sum_{i=0}^{n} c_i\,\phi_i(x)\right]\phi_j(x)\,dx = 0,\ j = 0(1)n. \tag{4.133}$$

Example 4.31 Obtain a linear polynomial approximation to the function $f(x) = x^3$ on the interval $[0, 1]$ using the least squares approximation with $W(x) = 1$.

Consider a linear polynomial

$$P(x) = a_0 x + a_1$$

where a_0 and a_1 are arbitrary parameters.

Using (4.130), we get

$$I(a_0, a_1) = \int_0^1 [x^3 - (a_0 x + a_1)]^2\,dx = \text{minimum}$$

$$I(a_0, a_1) = \frac{1}{7} - 2\left(\frac{a_0}{5} + \frac{a_1}{4}\right) + \frac{a_0^2}{3} + a_0 a_1 + a_1^2 = \text{minimum}.$$

The necessary conditions for $I(a_0, a_1)$ to be minimum are given by

$$\frac{\partial I}{\partial a_0} = -\frac{2}{5} + \frac{2}{3}a_0 + a_1 = 0$$

$$\frac{\partial I}{\partial a_1} = -\frac{1}{2} + a_0 + 2a_1 = 0$$

The solution is $a_0 = 9/10$ and $a_1 = -1/5$. The desired linear polynomial approximation is $P(x) = (9x - 2)/10$. The value of $I(a_0, a_1)$, that is, the minimum least squares error is $9/700$.

If we take the linear polynomial approximation through the origin, then we get $P(x) = 3x/5$ and $I(a_0, a_1) = 16/700$. The approximations are plotted in Fig. 4.4. It may be noted that the approximating polynomial $P(x)$ may or may not have common values with $f(x)$.

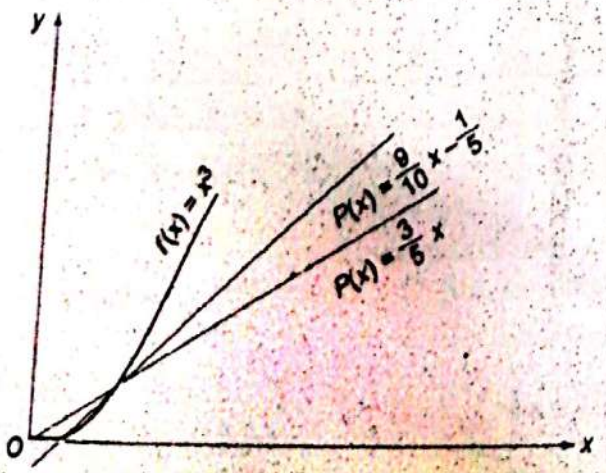

Fig. 4.4. Linear approximation.

Example 4.32 Obtain the least squares polynomial approximation of degree one and two for $x^{1/2}$ on [0, 1].

For $n = 1$, we have

$$I(c_0, c_1) = \int_0^1 (x^{1/2} - c_0 - c_1 x)^2 \, dx = \text{minimum}.$$

The normal equations are

$$\frac{\partial I}{\partial c_0} = -2 \int_0^1 (x^{1/2} - c_0 - c_1 x) \, dx = -2\left(\frac{2}{3} - c_0 - \frac{c_1}{2}\right) = 0$$

$$\frac{\partial I}{\partial c_1} = -\int_0^1 2(x^{1/2} - c_0 - c_1 x)\, x \, dx = -2\left(\frac{2}{5} - \frac{c_0}{2} - \frac{c_1}{3}\right) = 0.$$

We obtain $c_0 = 4/15$ and $c_1 = 4/5$. Thus, the first degree least square approximation to $x^{1/2}$ on [0, 1] is

$$P(x) = 4(1 + 3x)/15.$$

For $n = 2$, we have

$$I(c_0, c_1, c_2) = \int_0^1 (x^{1/2} - c_0 - c_1 x - c_2 x^2)^2 \, dx = \text{minimum}.$$

The normal equations are

$$\frac{\partial I}{\partial c_0} = -2 \int_0^1 (x^{1/2} - c_0 - c_1 x - c_2 x^2) \, dx = 0$$

$$\frac{\partial I}{\partial c_1} = -2 \int_0^1 (x^{1/2} - c_0 - c_1 x - c_2 x^2)\, x \, dx = 0$$

$$\frac{\partial I}{\partial c_2} = -2 \int_0^1 (x^{1/2} - c_0 - c_1 x - c_2 x^2)\, x^2 \, dx = 0$$

which give

$$c_0 + \frac{1}{2}c_1 + \frac{1}{3}c_2 = \frac{2}{3}$$

$$\frac{1}{2}c_0 + \frac{1}{3}c_1 + \frac{1}{4}c_2 = \frac{2}{5}$$

$$\frac{1}{3}c_0 + \frac{1}{4}c_1 + \frac{1}{5}c_2 = \frac{2}{7}.$$

The solution of this system is

$$c_0 = \frac{6}{35}, \; c_1 = \frac{48}{35} \; c_2 = -\frac{20}{35}.$$

...ulted approximation is

$$P(x) = \frac{1}{35}(6 + 48x - 20x^2).$$

...ple 4.33 Derive the least squares straight line and quadratic fits for the discrete data (x_i, f_i), $i = 0, 1, \ldots, N$.

...) $= c_0 + c_1 x$ be a straight line approximation. We have

$$I(c_0, c_1) = \sum_{i=0}^{N} [f(x_i) - (c_0 + c_1 x_i)]^2 = \text{minimum}.$$

...rmal equations are

$$\frac{\partial I}{\partial c_0} = -\sum_{i=0}^{N} 2[f(x_i) - (c_0 + c_1 x_i)] = 0$$

$$\frac{\partial I}{\partial c_1} = -\sum_{i=0}^{N} 2[f(x_i) - (c_0 + c_1 x_i)]\, x_i = 0.$$

...equations simplify to

$$c_0(N+1) + c_1 \sum x_i = \sum f(x_i)$$

$$c_0 \sum x_i + c_1 \sum x_i^2 = \sum x_i f(x_i).$$

...the second degree least squares approximation $P(x) = a + bx + cx^2$, the normal equations ...ify to

$$a(N+1) + b \sum x_i + c \sum x_i^2 = \sum f(x_i)$$

$$a \sum x_i + b \sum x_i^2 + c \sum x_i^3 = \sum x_i f(x_i)$$

$$a \sum x_i^2 + b \sum x_i^3 + c \sum x_i^4 = \sum x_i^2 f(x_i).$$

...ple 4.34 Obtain the least squares straight line fit to the following data

x	0	0.2	0.4	0.6	0.8	1
$f(x)$	0.447	0.632	0.775	0.894	1	

...have $\sum x_i = 3$, $\sum x_i^2 = 2.2$, $\sum f(x_i) = 3.748$ and $\sum x_i f(x_i) = 2.5224$. The normal equations ...fitting a straight line $P_1(x) = c_0 + c_1 x$, are

$$5c_0 + 3c_1 = 3.748, \quad 3c_0 + 2.2c_1 = 2.5224.$$

The solution of this system is $c_0 = 0.3392$ and $c_1 = 0.684$. The required approximation is ...(x) = 0.3392 + 0.684 x.

...Least squares error $= \sum_{i=0}^{4} [f(x_i) - (0.3392 + 0.684 x_i)]^2 = 0.00245.$

Example 4.35 Find the least squares approximation of second degree for the discrete data

x	-2	-1	0	1	2
$f(x)$	15	1	1	3	19

We have $\sum x_i = 0$, $\sum x_i^2 = 10$, $\sum x_i^3 = 0$, $\sum x_i^4 = 34$, $\sum f(x_i) = 39$, $\sum x_i f(x_i) =$ 10 and $\sum x_i^2 f(x_i) = 140$.

Therefore, the normal equations for fitting a second degree polynomial

$$P_2(x) = c_0 + c_1 x + c_2 x^2$$

are $5c_0 + 10c_2 = 39$

$$10c_1 = 10$$

$$10c_0 + 34c_2 = 140.$$

The solution of this system is

$$c_0 = -\frac{37}{35},\ c_1 = 1,\ c_2 = \frac{31}{7}.$$

The required approximation is

$$P_2(x) = \frac{1}{35}(-37 + 35x + 155x^2).$$

Example 4.36 Use the method of least squares to fit the curve $f(x) = c_0 x + (c_1/\sqrt{x})$ for the following data

x	0.2	0.3	0.5	1	2
$f(x)$	16	14	11	6	3

Find the least squares error.

From the condition that

$$I(c_0, c_1) = \sum \left[f(x_i) - c_0 x_i - \frac{c_1}{\sqrt{x_i}} \right]^2 = \text{minimum}$$

we obtain the normal equations

$$c_0 \sum x_i^2 + c_1 \sum x_i^{1/2} = \sum x_i f(x_i)$$

$$c_0 \sum x_i^{1/2} + c_1 \sum (1/x_i) = \sum [f(x_i)/x_i^{1/2}]$$

We have

$$\sum x_i^{1/2} = 4.1163,\ \sum (1/x_i) = 11.8333,\ \sum x_i^2 = 5.38,$$

$$\sum x_i f(x_i) = 24.9,\ \sum [f(x_i)/x_i^{1/2}] = 85.0151.$$

The normal equations are given by

$$5.38c_0 + 4.1163c_1 = 24.9$$

$$4.1163c_0 + 11.8333c_1 = 85.0151$$

... solution is $c_0 = -1.1836$, $c_1 = 7.5961$. Therefore, the least squares fit is given as

$$f(x) = (7.5961/x^{1/2}) - 1.1836x.$$

... squares error $= \sum \left[f(x_i) - \left\{ \frac{7.5961}{x_i^{1/2}} - 1.1836 x_i \right\} \right]^2 = 1.6887.$

... We are given the following values of a function of the variable t:

t	0.1	0.2	0.3	0.4
f	0.76	0.58	0.44	0.35

... a least squares fit of the form $f = ae^{-3t} + be^{-2t}$.

(Royal Inst. Tech., Stockholm, Sweden, BIT 17(1977), 11...

... the method of least squares, we have

$$I = \sum_{k=1}^{4} [f_k - (ae^{-3t_k} + be^{-2t_k})]^2 = \text{minimum}$$

... obtain the normal equations

$$\frac{\partial I}{\partial a} = \sum_{k=1}^{4} (f_k - ae^{-3t_k} - be^{-2t_k})\, e^{-3t_k} = 0$$

$$\frac{\partial I}{\partial b} = \sum_{k=1}^{4} (f_k - ae^{-3t_k} - be^{-2t_k})\, e^{-2t_k} = 0$$

$$a \sum_{k=1}^{4} e^{-6t_k} + b \sum_{k=1}^{4} e^{-5t_k} - \sum_{k=1}^{4} f_k e^{-3t_k} = 0$$

$$a \sum_{k=1}^{4} e^{-5t_k} + b \sum_{k=1}^{4} e^{-4t_k} - \sum_{k=1}^{4} f_k e^{-2t_k} = 0.$$

Using the table of values, we get the system of equations

$$1.106023a + 1.332876b - 1.165642 = 0$$

$$1.332876a + 1.622740b - 1.409764 = 0$$

which have the solution $a = 0.6853$, $b = 0.3058$. The least squares fit is

$$f = 0.6853\, e^{-3t} + 0.3058\, e^{-2t}.$$

For large n, the normal equations become ill-conditioned, which cause large errors in the parameters c_i, $i = 0, 1, \cdots, n$. This difficulty can be avoided, if the functions $\phi_i(x)$ are so chosen that they are orthogonal with respect to the weight function $W(x)$ on an interval $[a, b]$.

III - Unit finish

$$= \frac{1}{6}\,[2(x_0^3 + 3x_0^2 h + 3x_0 h^2 + h^3) - 2x_0^3 - 3x_0^2 h - 3h(x_0^2 + 2x_0 h + h^2)]$$

$$= -\frac{h^3}{6}.$$

The truncation error becomes

$$R_1 = \frac{C}{2} f''(\xi) = -\frac{h^3}{12} f''(\xi), \quad x_0 < \xi < x_1.$$

IV- Unit Start

Simpson's method

We have $n = 2$, $x_0 = a$, $x_1 = x_0 + h$, $x_2 = x_0 + 2h = b$, $h = (b - a)/2$. We write

$$\int_{x_0}^{x_2} f(x)\,dx = \lambda_0 f(x_0) + \lambda_1 f(x_1) + \lambda_2 f(x_2).$$

The rule can be made exact for polynomials of degree upto two.

For $f(x) = 1, x, x^2$, we get the following system of equations.

$$f(x) = 1: x_2 - x_0 = \lambda_0 + \lambda_1 + \lambda_2, \text{ or } 2h = \lambda_0 + \lambda_1 + \lambda_2 \qquad (5.$$

$$f(x) = x: \frac{1}{2}(x_2^2 - x_0^2) = \lambda_0 x_0 + \lambda_1 x_1 + \lambda_2 x_2 \qquad (5.$$

$$f(x) = x^2: \frac{1}{3}(x_2^3 - x_0^3) = \lambda_0 x_0^2 + \lambda_1 x_1^2 + \lambda_2 x_2^2. \qquad (5.8$$

From (5.81 *b*), we get

$$\frac{1}{2}(x_2 - x_0)(x_2 + x_0) = \lambda_0 x_0 + \lambda_1(x_0 + h) + \lambda_2(x_0 + 2h)$$

$$\frac{1}{2}(2h)(2x_0 + 2h) = (\lambda_0 + \lambda_1 + \lambda_2)x_0 + (\lambda_1 + 2\lambda_2)h$$

$$= 2h\,x_0 + (\lambda_1 + 2\lambda_2)h \quad \text{using (5.81 } a\text{)}$$

$$2h = \lambda_1 + 2\lambda_2. \qquad (5.81$$

From (5.81 *c*), we get

$$\frac{1}{3}\,[(x_0^3 + 6x_0^2 h + 12x_0 h^2 + 8h^3) - x_0^3] = \lambda_0 x_0^2 + \lambda_1 (x_0^2 + 2x_0 h + h^2) + \lambda_2(x_0^2 + 4x_0 h + 4h^2)$$

$$2x_0^2 h + 4x_0 h^2 + \frac{8}{3} h^3 = (\lambda_0 + \lambda_1 + \lambda_2)x_0^2 + 2(\lambda_1 + 2\lambda_2)x_0 h + (\lambda_1 + 4\lambda_2)h^2$$

$$= 2h\,x_0^2 + 4x_0 h^2 + (\lambda_1 + 4\lambda_2)h^2$$

$$\frac{8}{3}h = \lambda_1 + 4\lambda_2. \tag{5.81 c}$$

(c) and using (5.81 a), we obtain $\lambda_0 = h/3$, $\lambda_1 = 4h/3$, $\lambda_2 = h/3$.

by

$$\int_{x_0}^{x_2} f(x)\,dx = \frac{h}{3}\,[f(x_0) + 4f(x_1) + f(x_2)].$$

error constant is given by

$$C = -\frac{(b-a)^5}{120} = -\frac{4}{15}h^5.$$

$$R_2 = \frac{C}{4!}f^{iv}(\eta) = -\frac{h^5}{90}f^{iv}(\eta), \quad x_0 < \eta < x_2.$$

undetermined coefficients can be used to derive quadrature formulas of a given type derivations through the following examples.

Determine a, b and c such that the formula

$$\int_0^h f(x)\,dx = h\left\{af(0) + bf\left(\frac{h}{3}\right) + cf(h)\right\}$$

polynomials of as high order as possible, and determine the order of the truncation error.

(Uppsala Univ., Sweden, BIT 13(1973), 123)

method exact for polynomials of degree upto 2, we obtain

$$f(x) = 1: \quad h = h(a + b + c), \text{ or } a + b + c = 1.$$

$$f(x) = x: \quad \frac{h^2}{2} = h\left(\frac{bh}{3} + ch\right), \text{ or } \frac{1}{3}b + c = \frac{1}{2}.$$

$$f(x) = x^2: \quad \frac{h^3}{3} = h\left(\frac{bh^2}{9} + ch^2\right), \text{ or } \frac{1}{9}b + c = \frac{1}{3}.$$

the above equations, we get

$$a = 0, \quad b = 3/4, \text{ and } \quad c = 1/4.$$

ration error of the formula is given by

$$TE = \frac{C}{3!}f'''(\xi), \quad 0 < \xi < h$$

$$C = \int_0^h x^3\,dx - h\left[\frac{bh^3}{27} + ch^3\right] = -\frac{h^4}{36}.$$

Hence, we have

$$TE = -\frac{h^4}{216} f'''(\xi) = O(h^4).$$

Example 5.15 Find the quadrature formula

$$\int_0^1 f(x) \frac{dx}{\sqrt{x(1-x)}} = \alpha_1 f(0) + \alpha_2 f\left(\frac{1}{2}\right) + \alpha_3 f(1)$$

which is exact for polynomials of highest possible degree. Then use the formula on

$$\int_0^1 \frac{dx}{\sqrt{x - x^3}}$$

and compare with the exact value.

(Oslo Univ., Norway, BIT 7 (1967), 170)

Making the method exact for polynomials of degree upto 2, we obtain

for $f(x) = 1$: $I_1 = \int_0^1 \frac{dx}{\sqrt{x(1-x)}} = \alpha_1 + \alpha_2 + \alpha_3$

for $f(x) = x$: $I_2 = \int_0^1 \frac{x\,dx}{\sqrt{x(1-x)}} = \frac{1}{2}\alpha_2 + \alpha_3$

for $f(x) = x^2$: $I_3 = \int_0^1 \frac{x^2\,dx}{\sqrt{x(1-x)}} = \frac{1}{4}\alpha_2 + \alpha_3$

where

$$I_1 = \int_0^1 \frac{dx}{\sqrt{x(1-x)}} = 2\int_0^1 \frac{dx}{\sqrt{1-(2x-1)^2}} = \int_{-1}^1 \frac{dt}{\sqrt{1-t^2}} = \left[\sin^{-1} t\right]_{-1}^1 = \pi$$

$$I_2 = \int_0^1 \frac{x\,dx}{\sqrt{x(1-x)}} = 2\int_0^1 \frac{x\,dx}{\sqrt{1-(2x-1)^2}} = \int_{-1}^1 \frac{t+1}{2\sqrt{1-t^2}}\,dt$$

$$= \frac{1}{2}\int_{-1}^1 \frac{t\,dt}{\sqrt{1-t^2}} + \frac{1}{2}\int_{-1}^1 \frac{dt}{\sqrt{1-t^2}} = \frac{\pi}{2}.$$

$$I_3 = \int_0^1 \frac{x^2\,dx}{\sqrt{x(1-x)}} = 2\int_0^1 \frac{x^2\,dx}{\sqrt{1-(2x-1)^2}} = \frac{1}{4}\int_{-1}^1 \frac{(t+1)^2}{\sqrt{1-t^2}}\,dt$$

$$= \frac{1}{4}\int_{-1}^1 \frac{t^2}{\sqrt{1-t^2}}\,dt + \frac{1}{2}\int_{-1}^1 \frac{t}{\sqrt{1-t^2}}\,dt + \frac{1}{4}\int_{-1}^1 \frac{dt}{\sqrt{1-t^2}}$$

$$= \frac{3\pi}{8}.$$

Hence, we have the equations

$$\alpha_1 + \alpha_2 + \alpha_3 = \pi$$

$$\frac{1}{2}\alpha_2 + \alpha_3 = \frac{\pi}{2}$$

$$\frac{1}{4}\alpha_2 + \alpha_3 = \frac{3\pi}{8}$$

which gives $\alpha_1 = \pi/4, \quad \alpha_2 = \pi/2, \quad \alpha_3 = \pi/4.$

The quadrature formula is given by

$$\int_0^1 \frac{f(x)\,dx}{\sqrt{x(1-x)}} = \frac{\pi}{4}\left[f(0) + 2f\left(\frac{1}{2}\right) + f(1)\right]$$

We now use this formula to evaluate

$$I = \int_0^1 \frac{dx}{\sqrt{x - x^3}} = \int_0^1 \frac{dx}{\sqrt{1+x}\,\sqrt{x(1-x)}} = \int_0^1 \frac{f(x)\,dx}{\sqrt{x(1-x)}}$$

where $f(x) = 1/\sqrt{1+x}$.

We obtain $$I = \frac{\pi}{4}\left[1 + \frac{2\sqrt{2}}{\sqrt{3}} + \frac{\sqrt{2}}{2}\right] \approx 2.62331.$$

The exact value is $I = 2.62205755$.

Gauss Quadrature Methods

In the integration method (5.70), the nodes x_k's and the weights λ_k's, $k = 0(1)n$ can also be obtained by making the formula exact for polynomials of degree upto m. When the nodes are known, the $m = n$, the corresponding methods are called Newton-Cotes methods. When the nodes are also determined, we have $m = 2n + 1$ and the methods are called **Gaussian integration methods.** Since any finite interval $[a, b]$ can always be transformed to $[-1, 1]$, using the transformation

$$x = \frac{b-a}{2}t + \frac{b+a}{2}$$

we consider the integral in the form

$$\int_{-1}^{1} w(x)\,f(x)\,dx = \sum_{k=0}^{n} \lambda_k f_k$$

where $w(x) > 0$, $-1 \le x \le 1$, is the weight function.

...-Legendre Integration Methods

...the weight function be $w(x) = 1$. Then, the method (5.82) reduces to

$$\int_{-1}^{1} f(x)\,dx = \sum_{k=0}^{n} \lambda_k f(x_k).$$

...this case, all the nodes x_k and weights λ_k are unknown. Consider the following cases.

...*point formula* $n = 0$. The formula is given by

$$\int_{-1}^{1} f(x)\,dx = \lambda_0 f(x_0).$$

...method has two unknowns λ_0, x_0. Making the method exact for $f(x) = 1, x$, we get

$$f(x) = 1: \quad 2 = \lambda_0$$

$$f(x) = x: \quad 0 = \lambda_0 x_0 \text{ or } x_0 = 0.$$

...ce, the method is given by

$$\int_{-1}^{1} f(x)\,dx = 2f(0)$$

...ch is same as the mid-point formula. The error constant is given by

$$C = \int_{-1}^{1} x^2\,dx - 2[0] = \frac{2}{3}.$$

...nce,

$$R_1 = \frac{C}{2!} f''(\xi) = \frac{1}{3} f''(\xi), \quad -1 < \xi < 1.$$

...*o-point formula* $n = 1$. The formula is given by

$$\int_{-1}^{1} f(x)\,dx = \lambda_0 f(x_0) + \lambda_1 f(x_1).$$

...e method has four unknowns, x_0, x_1, λ_0 and λ_1. Making the method exact for $f(x) = 1, x, x^2, x^3$. (5.86)

...get

$$f(x) = 1 \;:\; 2 = \lambda_0 + \lambda_1$$

$$f(x) = x \;:\; 0 = \lambda_0 x_0 + \lambda_1 x_1 \qquad (5.87\,a)$$

$$f(x) = x^2 \;:\; \frac{2}{3} = \lambda_0 x_0^2 + \lambda_1 x_1^2 \qquad (5.87\,b)$$

$$f(x) = x^3 \;:\; 0 = \lambda_0 x_0^3 + \lambda_1 x_1^3. \qquad (5.87\,c)$$

$$(5.87\,d)$$

(5.8...liminating λ_0 from (5.87 *b*), (5.87 *d*), we get

$$\lambda_1 x_1^3 - \lambda_1 x_1 x_0^2 = 0, \text{ or } \lambda_1 x_1 (x_1 - x_0)(x_1 + x_0) = 0.$$

$\neq x_1$, we get $x_1 + x_0 = 0$ or $x_1 = -x_0$. Note that if $x_1 = 0$, then from (5.87 *b*), we
$\neq \lambda_0 \neq 0$. Therefore, $x_1 \neq 0$.
(5.87 *b*), we get $\lambda_0 - \lambda_1 = 0$, or $\lambda_0 = \lambda_1$.
(5.87 *a*), we get $\lambda_0 = \lambda_1 = 1$.
we get $x_0^2 = 1/3$, or $x_0 = \pm 1/\sqrt{3}$, and $x_1 = \mp 1/\sqrt{3}$. Therefore, the two-point Gauss
thod is given by

$$\int_{-1}^{1} f(x)\,dx = f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right). \qquad (5.87)$$

onstant is given by

$$C = \int_{-1}^{1} x^4\,dx - \left[\frac{1}{9} + \frac{1}{9}\right] = \frac{2}{5} - \frac{2}{9} = \frac{8}{45}.$$

term R_4 becomes

$$R_4 = \frac{C}{4!} f^{(4)}(\xi) = \frac{1}{135} f^{(4)}(\xi), \quad -1 < \xi < 1. \qquad (5.89)$$

int formula $n = 2$. The method is given by

$$\int_{-1}^{1} f(x)\,dx = \lambda_0 f(x_0) + \lambda_1 f(x_1) + \lambda_2 f(x_2).$$

re six unknowns in the method and it can be made exact for polynomials of degree upto five.
$= x^i$, $i = 0(1)5$, we get the system of equations

$$f(x) = 1 \;:\; \lambda_0 + \lambda_1 + \lambda_2 = 2 \qquad (5.90\,a)$$
$$f(x) = x \;:\; \lambda_0 x_0 + \lambda_1 x_1 + \lambda_2 x_2 = 0 \qquad (5.90\,b)$$
$$f(x) = x^2 \;:\; \lambda_0 x_0^2 + \lambda_1 x_1^2 + \lambda_2 x_2^2 = \frac{2}{3} \qquad (5.90\,c)$$
$$f(x) = x^3 \;:\; \lambda_0 x_0^3 + \lambda_1 x_1^3 + \lambda_2 x_2^3 = 0 \qquad (5.90\,d)$$
$$f(x) = x^4 \;:\; \lambda_0 x_0^4 + \lambda_1 x_1^4 + \lambda_2 x_2^4 = \frac{2}{5} \qquad (5.90\,e)$$
$$f(x) = x^5 \;:\; \lambda_0 x_0^5 + \lambda_1 x_1^5 + \lambda_2 x_2^5 = 0. \qquad (5.90\,f)$$

minating λ_0 from (5.90 *b*), (5.90 *d*) and (5.90 *d*), (5.90 *f*), we get

$$\lambda_1 x_1(x_1^2 - x_0^2) + \lambda_2 x_2(x_2^2 - x_0^2) = 0$$
$$\lambda_1 x_1^3(x_1^2 - x_0^2) + \lambda_2 x_2^3(x_2^2 - x_0^2) = 0.$$

Eliminating the first term from these two equations, we get

$$\lambda_2 x_2^3(x_2^2 - x_0^2) - \lambda_2 x_2 x_1^2(x_2^2 - x_0^2) = 0$$
$$\lambda_2 x_2(x_2^2 - x_0^2)(x_2^2 - x_1^2) = 0.$$

or

Since x_0, x_1, x_2 are distinct, we get on cancelling the terms $(x_2 - x_0)$ and $(x_2 - x_1)$

$$\lambda_2 x_2(x_2 + x_0)(x_2 + x_1) = 0.$$

We have $\lambda_2 \neq 0$ and let $x_2 \neq 0$. Then, we have either $x_2 = -x_0$ or $x_2 = -x_1$. Let $x_2 = -x_0$. Then, from (5.90 *b*), (5.90 *d*), we get

$$(\lambda_0 - \lambda_2)x_0 + \lambda_1 x_1 = 0$$

$$(\lambda_0 - \lambda_2)x_0^3 + \lambda_1 x_1^3 = 0.$$

Eliminating the first term, we get $\lambda_1 x_1 (x_1^2 - x_0^2) = 0$. Since, $\lambda_1 \neq 0$, $x_1 \neq x_0$, $x_1 \neq -x_0$ (otherwise $x_1 = x_2$), we get $x_1 = 0$.

Hence, $(\lambda_0 - \lambda_2)\, x_0 = 0$, or $\lambda_0 = \lambda_2$ since $x_0 \neq 0$.

Now, (5.90 *c*), (5.90 *e*) give

$$2\lambda_0 x_0^2 = \frac{2}{3}, \quad 2\lambda_0 x_0^4 = \frac{2}{5}.$$

Dividing, we get $x_0^2 = 3/5$, or $x_0 = \pm\sqrt{3/5}$. Then $x_2 = \mp\sqrt{3/5}$.

Now, $\lambda_0 x_0^2 = 1/3$ gives $\lambda_0 = 5/9$ and $\lambda_2 = \lambda_0 = 5/9$. From (5.90 *a*), we get $\lambda_1 = 2 - 2\lambda_2 = 8/9$.

Therefore, the three-point Gauss-Legendre method is given by

$$\int_{-1}^{1} f(x)\, dx = \frac{1}{9}\left[5f\left(-\sqrt{\frac{3}{5}}\right) + 8f(0) + 5f\left(\sqrt{\frac{3}{5}}\right)\right] \tag{5.91}$$

If we take $x_2 = -x_1$, then we get $x_0 = 0$ and $x_2 = \pm\sqrt{3/5}$ giving the same method. The nodes are symmetrically placed about $x = 0$.

The error constant is given by

$$C = \int_{-1}^{1} x^6\, dx - \frac{1}{9}\left[5\left(-\sqrt{\frac{3}{5}}\right)^6 + 0 + 5\left(\sqrt{\frac{3}{5}}\right)^6\right]$$

$$= \frac{2}{7} - \frac{6}{25} = \frac{8}{175}.$$

The error in the method becomes

$$R_6 = \frac{C}{6!} f^{(6)}(\xi) = \frac{8}{(6!)175} f^{(6)}(\xi) = \frac{1}{15750} f^{(6)}(\xi), \quad -1 < \xi < 1.$$

In the later part of this section, we shall prove that the abscissas of the above formulas are the zeros of the Legendre polynomials of the corresponding order. Hence, they are called the Gauss-Legendre quadrature methods.

The nodes and the corresponding weights for the **Gauss-Legendre integration method** (5.83) for $n = 1(1)5$ are given in Table 5.3.

Table 5.3 Nodes and Weights for Gauss-Legendre Integration Method (5.83).

n	nodes x_k	weights λ_k
1	±0.5773502692	1.0000000000
	0.0000000000	0.8888888889
2	±0.7745966692	0.5555555556
	±0.3399810436	0.6521451549
3	±0.8611363116	0.3478548451
	0.0000000000	0.5688888889
4	±0.5384693101	0.4786286705
	±0.9061798459	0.2369268851
	±0.2386191861	0.4679139346
5	±0.6612093865	0.3607615730
	±0.9324695142	0.1713244924

Example 5.16 Evaluate the integral

$$I = \int_0^1 \frac{dx}{1+x}$$

using Gauss-Legendre three-point formula.

First we transform the interval $[0, 1]$ to the interval $[-1, 1]$. Let $t = ax + b$. We have

$$-1 = b, \quad 1 = a + b$$

or

$$a = 2, \quad b = -1, \text{ and } t = 2x - 1.$$

$$I = \int_0^1 \frac{dx}{1+x} = \int_{-1}^1 \frac{dt}{t+3}.$$

Using Gauss-Legendre three-point rule (corresponding to $n = 2$), we get

$$I = \frac{1}{9}\left[8\left(\frac{1}{0+3}\right) + 5\left(\frac{1}{3+\sqrt{3/5}}\right) + 5\left(\frac{1}{3-\sqrt{3/5}}\right)\right]$$

$$= \frac{131}{189} = 0.693122.$$

The exact solution is $I = \ln 2 = 0.693147$.

Example 5.17 Evaluate the integral $I = \int_1^2 \frac{2x\,dx}{1+x^4}$, using the Gauss-Legendre 1-point, 2-point, 3-point quadrature rules. Compare with the exact solution

$$I = \tan^{-1}(4) - (\pi/4).$$

the Gauss-Legendre rules, the interval $[1, 2]$ is to be reduced to $[-1, 1]$. Writing $x = at + b$

$$1 = -a + b, \quad 2 = a + b$$

solution is $b = 3/2$, $a = 1/2$. Therefore, $x = (t + 3)/2$, $dx = dt/2$ and

$$I = \int_{-1}^{1} \frac{8(t+3)\,dt}{\left[16 + (t+3)^4\right]} = \int_{-1}^{1} f(t)dt.$$

the 1-point rule, we get

$$I = 2f(0) = 2\left[\frac{24}{16+81}\right] = 0.4948.$$

the 2-point rule, we get

$$I = f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right) = 0.3842 + 0.1592 = 0.5434.$$

the 3-point rule, we get

$$I = \frac{1}{9}\left[5f\left(-\sqrt{\frac{3}{5}}\right) + 8f(0) + 5f\left(\sqrt{\frac{3}{5}}\right)\right]$$

$$= \frac{1}{9}\left[5(0.4393) + 8(0.2474) + 5(0.1379)\right] = 0.5406.$$

exact solution is $I = 0.5404$.

...s-Chebyshev Integration Methods

the weight function be $w(x) = 1/\sqrt{1-x^2}$. Then, the method (5.82) reduces to

$$\int_{-1}^{1} \frac{f(x)}{\sqrt{1-x^2}}\,dx = \sum_{k=0}^{n} \lambda_k f(x_k). \tag{5.92}$$

abscissas x_k and weights λ_k are unknown. Consider the following cases.

...point formula $n = 0$. The formula is given by

$$\int_{-1}^{1} \frac{f(x)}{\sqrt{1-x^2}}\,dx = \lambda_0 f(x_0). \tag{5.93}$$

method has two unknowns λ_0, x_0. Making it exact for $f(x) = 1, x$, we get

$$f(x) = 1: \quad \int_{-1}^{1} \frac{dx}{\sqrt{1-x^2}} = \lambda_0, \text{ or } \left[\sin^{-1}(x)\right]_{-1}^{1} = \lambda_0, \text{ or } \lambda_0 = \pi$$

$$f(x) = x: \quad \int_{-1}^{1} \frac{x\,dx}{\sqrt{1-x^2}} = \lambda_0 x_0, \text{ or } \lambda_0 x_0 = 0, \text{ or } x_0 = 0.$$

...ace, the method is given by

$$\int_{-1}^{1} \frac{f(x)}{\sqrt{1-x^2}}\,dx = \pi f(0). \tag{5.94}$$

We have $f(x) = 1/(x^2 + x + 1)$. Using the two-point formula, we get

$$I = \frac{\sqrt{\pi}}{2}\left[f\left(-\frac{1}{\sqrt{2}}\right) + f\left(\frac{1}{\sqrt{2}}\right)\right]$$

$$= \frac{\sqrt{\pi}}{2}\,[1.26120 + 0.45308] = 1.51924.$$

Using the three-point formula, we get

$$I = 0.29541\,[f(-1.22474) + f(1.22474)] + 1.18164\,f(0)$$

$$= 0.29541\,[0.78416 + 0.26848] + 1.18164 = 1.49260.$$

Integration Methods of Gaussian Type with Preassigned Abscissas

The Lobatto and Radau integration methods are of Gauss type with two nodes and one preassigned respectively. We consider the quadrature formula

$$\int_{-1}^{1} w(x)\, f(x)\, dx \approx \sum_{k=1}^{m} a_k\, f(\eta_k) + \sum_{k=1}^{n} \lambda_k\, f(x_k). \qquad (5.$$

In the case of the Lobatto formula, we preassign $w(x) = 1$, $\eta_1 = -1$, $\eta_m = 1$, $a_i = 0$, $i = 2, 3$ $m - 1$.

In the case of the Radau formula, we preassign $w(x) = 1$, $\eta_1 = -1$, $a_i = 0$, $i = 2, 3, \cdots, m$.

We now, derive these formulas.

Lobatto Integration Methods

In this case, $w(x) = 1$ and the end points -1 and 1 are always taken as nodes. The remaining n nodes are to be determined. The integration formula can be written as

$$\int_{-1}^{1} f(x)\, dx = \lambda_0\, f(-1) + \lambda_n\, f(1) + \sum_{k=1}^{n-1} \lambda_k\, f(x_k). \qquad (5.$$

Since there are $2n$ unknowns ($n - 1$ nodes and $n + 1$ weights), this method can be made exact for polynomials of degree upto $2n - 1$.

For $n = 2$, we have the method as

$$\int_{-1}^{1} f(x)\, dx = \lambda_0\, f(-1) + \lambda_1\, f(x_1) + \lambda_2\, f(1). \qquad (5.$$

Making the formula exact for $f(x) = 1$, x, x^2 and x^3, we get

$$f(x) = 1: \quad \lambda_0 + \lambda_1 + \lambda_2 = 2 \qquad (5.12$$

$$f(x) = x: \quad -\lambda_0 + \lambda_1 x_1 + \lambda_2 = 0 \qquad (5.12$$

$$f(x) = x^2: \quad \lambda_0 + \lambda_1 x_1^2 + \lambda_2 = \frac{2}{3} \qquad (5.12$$

$$f(x) = x^3: \quad -\lambda_0 + \lambda_1 x_1^3 + \lambda_2 = 0. \qquad (5.12$$

Subtracting (5.122 *b*) from (5.122 *d*), we get

$$\lambda_1 x_1(x_1^2 - 1) = 0.$$

$x_1 \neq \pm 1$, we get $x_1 = 0$. Substituting $x_1 = 0$ in (5.122 *b*) and (5.122 *c*) and solving, w
$\lambda_2 = 1/3$. From (5.122 *a*), we get $\lambda_1 = 4/3$.
Lobatto method is given by

$$\int_{-1}^{1} f(x)\,dx = \frac{1}{3}\,[f(-1) + 4f(0) + f(1)]. \qquad (5.$$

error constant is given by

$$C = \int_{-1}^{1} x^4\,dx - \frac{1}{3}\,[1 + 0 + 1] = \frac{2}{5} - \frac{2}{3} = -\frac{4}{15}.$$

refore, the error in the method is

$$R_4 = \frac{C}{4!}\,f^{(4)}(\xi) = -\frac{1}{90}\,f^{(4)}(\xi). \qquad (5.12$$

be noted that (5.123) is the Simpson rule with the step length $h = 1$.
$n = 3$, we have the method as

$$\int_{-1}^{1} f(x)\,dx = \lambda_0 f(-1) + \lambda_1 f(x_1) + \lambda_2 f(x_2) + \lambda_3 f(1). \qquad (5.125$$

method has six unknowns and it can be made exact for polynomials of degree upto 5. For
$= x^i$, $i = 0, 1, 2, 3, 4, 5$, we get the system of equations

$$f(x) = 1: \quad \lambda_0 + \lambda_1 + \lambda_2 + \lambda_3 = 2 \qquad (5.126\,a)$$

$$f(x) = x: \quad -\lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 = 0 \qquad (5.126\,b)$$

$$f(x) = x^2: \quad \lambda_0 + \lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 = \frac{2}{3} \qquad (5.126\,c)$$

$$f(x) = x^3: \quad -\lambda_0 + \lambda_1 x_1^3 + \lambda_2 x_2^3 + \lambda_3 = 0 \qquad (5.126\,d)$$

$$f(x) = x^4: \quad \lambda_0 + \lambda_1 x_1^4 + \lambda_2 x_2^4 + \lambda_3 = \frac{2}{5} \qquad (5.126\,e)$$

$$f(x) = x^5: \quad -\lambda_0 + \lambda_1 x_1^5 + \lambda_2 x_2^5 + \lambda_3 = 0. \qquad (5.126\,f)$$

racting (5.126 *b*) from (5.126 *d*), we obtain

$$\lambda_1 x_1(x_1^2 - 1) + \lambda_2 x_2(x_2^2 - 1) = 0. \qquad (5.127\,a)$$

racting (5.126 *d*) from (5.126 *f*), we obtain

$$\lambda_1 x_1^3(x_1^2 - 1) + \lambda_2 x_2^3(x_2^2 - 1) = 0. \qquad (5.127\,b)$$

ing the second terms of (5.127 *a*), (5.127 *b*) to the right hand side and dividing the two equations,
get $x_1^2 = x_2^2$.
e $x_1 \neq x_2$, we get $x_2 = -x_1$.
stituting in (5.127 *a*), we get $(\lambda_1 - \lambda_2)\,x_1\,(x_1^2 - 1) = 0$.
e, $x_1 \neq \pm 1$ and $x_1 \neq 0$ (otherwise $x_2 = 0$, $x_1 = 0$), we get $\lambda_1 = \lambda_2$.
g (5.126 *b*), we get $-\lambda_0 + \lambda_3 = 0$, or $\lambda_0 = \lambda_3$. (5.127 *c*)
g (5.126 *a*), we get $\lambda_0 + \lambda_1 = 1$.

(5.126 c) from (5.126 a), we get

$$\lambda_1(1 - x_1^2) + \lambda_2(1 - x_2^2) = \frac{4}{3}, \text{ or } \lambda_1(1 - x_1^2) = \frac{2}{3}.$$

(5.126 e) from (5.126 c), we get (5.127

$$\lambda_1 x_1^2(1 - x_1^2) + \lambda_2 x_2^2(1 - x_2^2) = \frac{2}{3} - \frac{2}{5} = \frac{4}{15},$$

$$\lambda_1 x_1^2(1 - x_1^2) = \frac{2}{15}.$$

the last two equations, we get $x_1^2 = 1/5$. Hence, we have

$$x_1 = -1/\sqrt{5} \text{ and } x_2 = -x_1 = 1/\sqrt{5}.$$

(5.127 d), we get $\lambda_1 = \left(\frac{2}{3}\right)\left(\frac{5}{4}\right) = \frac{5}{6}$ and $\lambda_2 = \lambda_1 = \frac{5}{6}$.

(5.127 c), we get $\lambda_0 = \frac{1}{6} = \lambda_3$.

bove, the method is given by

$$\int_{-1}^{1} f(x)\,dx = \frac{1}{6}\left[f(-1) + 5f\left(-\frac{\sqrt{5}}{5}\right) + 5f\left(\frac{\sqrt{5}}{5}\right) + f(1)\right]. \quad (5.12$$

error constant is given by

$$C = \int_{-1}^{1} x^6\,dx - \frac{1}{6}\left[1 + 5\left(\frac{1}{125}\right) + 5\left(\frac{1}{125}\right) + 1\right]$$

$$= \frac{2}{7} - \frac{26}{75} = -\frac{32}{525}.$$

error in the method becomes

$$R_6 = \frac{C}{6!} f^{(6)}(\xi) = -\frac{32}{525(6!)} f^{(6)}(\xi), \quad -1 < \xi < 1. \quad (5.12$$

nodes and the corresponding weights for the Lobatto integration method (5.120) for $n = 2(1)5$ a
n in Table 5.6.

Table 5.6. Nodes and Weights for Lobatto Integration Method (5.120).

n	nodes x_k	weights λ_k
2	±1.00000000	0.33333333
	0.00000000	1.33333333
3	±1.00000000	0.16666667
	±0.44721360	0.83333333
4	±1.00000000	0.10000000
	±0.65465367	0.54444444
	0.00000000	0.71111111
5	±1.00000000	0.06666667
	±0.76505532	0.37847496
	±0.28523152	0.55485837

IV - unit finish

Simplifying, we have

$$[y(t_{j+1}) - y(t_j)]\,[-\sin(t_{j+1})\cos t_j + \cos(t_{j+1})\sin t_j]$$
$$- y'(t_{j+1})\,[\cos t_j\,\{\cos(t_{j+1}) - \cos t_j\} + \sin t_j\,\{\sin(t_{j+1}) - \sin t_j\}]$$
$$+ y'(t_j)\,[\cos(t_{j+1})\,\{\cos(t_{j+1}) - \cos t_j\} + \sin(t_{j+1})\,\{\sin(t_{j+1}) - \sin t_j\}] = 0$$

or $(y_{j+1} - y_j)(-\sin h) - y'_{j+1}[\cos h - 1] + y'_j[1 - \cos h] = 0$

or $$y_{j+1} = y_j + \frac{1 - \cos h}{\sin h}(y'_{j+1} + y'_j).$$

6.4 SINGLESTEP METHODS

The methods for the solution of the initial value problem

$$u' = f(t, u),\ u(t_0) = \eta_0,\ t \in [t_0, b] \qquad (6.$$

can be classified mainly into two types. They are (i) singlestep methods and (ii) multistep meth

In singlestep methods, the solution at any point is obtained using the solution at only the previ point. Thus, a general singlestep method can be written as

$$u_{j+1} = u_j + h\phi(t_{j+1}, t_j, u_{j+1}, u_j, h) \qquad (6.$$

where ϕ is a function of the arguments $t_j, t_{j+1}, u_j, u_{j+1}, h$ and also depends on f. We often writ as $\phi(t, u, h)$. This function ϕ is called the **increment function**. If u_{j+1} can be obtained simply evaluating the right hand side of (6.71), then the method is called an **explicit method**. In this c the method is of the form

$$u_{j+1} = u_j + h\phi(t_j, u_j, h). \qquad (6.$$

If the right hand side of (6.71) depends on u_{j+1} also, then it is called an **implicit method**. The gen form in this case is as given in (6.71).

Local Truncation Error or Discretization Error

The true (exact) value $u(t_j)$ satisfies the equation

$$u(t_{j+1}) = u(t_j) + h\phi(t_{j+1}, t_j, u(t_{j+1}), u(t_j), h) + T_{j+1}$$

where T_{j+1} is called the local truncation error or discretization error of the method. Therefore, truncation error is given by

$$T_{j+1} = u(t_{j+1}) - u(t_j) - h\phi(t_{j+1}, t_j, u(t_{j+1}), u(t_j), h) \qquad (6.$$

Order of a Method

The order of a method is the largest integer p for which

$$\left|\frac{1}{h}T_{j+1}\right| = O(h^p). \qquad (6.$$

We now derive singlestep methods which have different increment functions.

Taylor Series Method

The fundamental numerical method for the solution of (6.70) is the Taylor series method. Assume that the function $u(t)$ can be expanded in Taylor series about any point t_j, that is,

$$u(t) = u(t_j) + (t - t_j)\, u'(t_j) + \frac{1}{2!}\,(t - t_j)^2\, u''(t_j) + \cdots$$

$$+ \frac{1}{p!}\,(t - t_j)^p\, u^{(p)}(t_j) + \frac{1}{(p+1)!}\,(t - t_j)^{p+1}\, u^{(p+1)}(t_j + \theta h). \tag{6.75}$$

This expansion holds for $t \in [t_0, b]$ and $0 < \theta < 1$.

Substituting $t = t_{j+1}$ in (6.75), we get

$$u(t_{j+1}) = u(t_j) + hu'(t_j) + \frac{h^2}{2!}\,u''(t_j) + \cdots + \frac{1}{p!}\,h^p u^{(p)}(t_j) + \frac{1}{(p+1)!}\,h^{p+1} u^{(p+1)}(t_j + \theta h)$$

$$= u(t_j) + h\phi(t_j, u(t_j), h) + \frac{1}{(p+1)!}\,h^{p+1}\, u^{(p+1)}(t_j + \theta h)$$

where $h\phi(t_j, u(t_j), h) = hu'(t_j) + \frac{h^2}{2!}u''(t_j) + \cdots + \frac{h^p}{p!}u^{(p)}(t_j)$.

We denote by $h\phi(t_j, u_j, h)$, the value obtained from $h\phi(t_j, u(t_j), h)$ by using an approximate value u_j in place of the exact value $u(t_j)$. Neglecting the error term, we have the method

$$u_{j+1} = u_j + h\phi(t_j, u_j, h),\ j = 0, 1, \cdots, N - 1 \tag{6.76}$$

to approximate $u(t_{j+1})$. The error or the truncation error of the method is given by

$$T_{j+1} = \frac{1}{(p+1)!}\,h^{p+1}\, u^{p+1}(t_j + \theta h). \tag{6.77}$$

The method (6.76) is called the **Taylor series method** of order p. Substituting $p = 1$, in (6.76) we get

$$u_{j+1} = u_j + hu'_j = u_j + hf(t_j, u_j)$$

which is the Euler method. Therefore, Euler method can also be called as the Taylor series method of order 1.

To apply (6.76), it is necessary to know $u(t_j), u'(t_j), \cdots, u^{(p)}(t_j)$. If t_j and $u(t_j)$ are known, then the derivatives can be calculated as follows:

First, the known values t_j and $u(t_j)$ are substituted into the differential equation to give

$$u'(t_j) = f(t_j, u(t_j)).$$

Next, the differential equation $u' = f(t, u)$ is differentiated to obtain expressions for the higher order derivatives of $u(t)$. Thus, we have

$$u' = f(t, u)$$

$$u'' = f_t + ff_u$$
$$u''' = f_{tt} + 2ff_{tu} + f^2 f_{uu} + f_u(f_t + ff_u)$$
. . . . . .

represent the partial derivatives of f with respect to t and u and so on. The $u^{(p)}$ can be computed by substituting $t = t_j$. Therefore, if t_j and $u(t_j)$ are known exactly be used to compute u_{j+1} with an error

$$\frac{h^{p+1}}{(p+1)!} u^{(p+1)}(t_j + \theta h).$$

of terms to be included in (6.76) is fixed by the permissible error. If this error is to be truncated at the term $u^{(p)}(t_j)$ then

$$h^{p+1} |u^{(p+1)}(t_j + \theta h)| < (p+1)!\, \varepsilon$$

$$h^{p+1} |f^{(p)}(t_j + \theta h)| < (p+1)!\, \varepsilon. \qquad (6.$$

does an estimate of $|f^{(p)}(t_j + \theta h)|$ is known.

and ε, (6.78) will determine p, and if p and ε are specified, then it will give an upper

θh is not known, $|f^{(p)}(t_j + \theta h)|$ in (6.78) is replaced by its maximum value in $[t_0$ determining this value is as follows. Write one more non-vanishing term in the series th and then differentiate this series p times. The maximum value of this quantity in $[t_0$ ough required bound.

Ex. 6.15 Given the initial value problem

$$u' = t^2 + u^2, \; u(0) = 0$$

the first three non-zero terms in the Taylor series for $u(t)$ and hence obtain the value of determine t when the error in $u(t)$ obtained from the first two non-zero terms is to be less after rounding.

$$u(0) = 0, \; u'(0) = 0$$
$$u'' = 2t + 2uu', \; u''(0) = 0$$
$$u''' = 2 + 2(uu'' + (u')^2), \; u'''(0) = 2$$
$$u^{(4)} = 2(uu''' + 3u'u''), \; u^{(4)}(0) = 0$$
$$u^{(5)} = 2[uu^{(4)} + 4u'u''' + 3(u'')^2], \; u^{(5)}(0) = 0$$
$$u^{(6)} = 2(uu^{(5)} + 5u'u^{(4)} + 10u''u'''), \; u^{(6)}(0) = 0$$
$$u^{(7)} = 2(uu^{(6)} + 6u'u^{(5)} + 15u''u^{(4)} + 10(u''')^2), \; u^{(7)}(0) = 80$$
$$u^{(8)}(0) = u^{(9)}(0) = u^{(10)}(0) = 0$$
$$u^{(11)} = 2[uu^{(10)} + 10u'u^{(9)} + 45u''\, u^{(8)} + 120u'''u^{(7)} + 210u^{(4)}\, u^{(6)} + 126(u^{(5)})^2], \; u^{(11)}(0) = 38400.$$

The Taylor series for $u(t)$ becomes

$$u(t) = \frac{1}{3}t^3 + \frac{1}{63}t^7 + \frac{2}{2079}t^{11}.$$

Approximate value of $u(1)$ is given by

$$u(1) = \frac{1}{3} + \frac{1}{63} + \frac{2}{2079} \approx 0.350168.$$

If the first two terms are used, then the value of t is obtained from

$$\left|\frac{2}{2079}t^{11}\right| < 0.5 \times 10^{-7}$$

Solving, we get $t \approx 0.41$.

Example 6.16 Find the three term Taylor series solution for the third order initial value problem

$$W''' + WW'' = 0, \quad W(0) = 0,$$
$$W'(0) = 0, \quad W''(0) = 1.$$

Find the bound on the error for $t \in [0, 0.2]$.

Solution. $W''' = -WW''$, $W'''(0) = 0$

$$W^{(4)} = -(WW''' + W'W''), \; W^{(4)}(0) = 0$$
$$W^{(5)} = -[WW^{(4)} + 2W'W''' + (W'')^2], \; W^{(5)}(0) = -1$$
$$W^{(6)}(0) = 0, \; W^{(7)}(0) = 0, \; W^{(8)}(0) = 11$$
$$W^{(9)}(0) = W^{(10)}(0) = 0, \; W^{(11)}(0) = -375.$$

The Taylor series solution is

$$W(t) = \frac{t^2}{2!} - \frac{t^5}{5!} + \frac{11}{8!}t^8 + E_8$$

where

$$|E_8| \le \max |W^{(9)}(t)| \frac{t^9}{9!}.$$

Omitting the next term, we have

$$W(t) = \frac{t^2}{2!} - \frac{t^5}{5!} + \frac{11}{8!}t^8 - \frac{375}{11!}t^{11}.$$

We find

$$W^{(9)}(t) = -\frac{375}{2}t^2$$

and

$$\max_{0 \le t \le 0.2} |W^{(9)}(t)| = 7.5.$$

Hence, $|E_8| \le \frac{7.5\,(0.2)^9}{9!} \le (1.06)10^{-11}$.

Taylor series expansion of K_1, K_2 given in this recursive form is very difficult. Since, K_1, K_2 can be expanded in powers of h, we can write

$$K_1 = hA_1 + h^2B_1 + h^3C_1 + \cdots$$

$$K_2 = hA_2 + h^2B_2 + h^3C_2 + \cdots \qquad (6.1$$

Substitute (6.126) in (6.125), expand in Taylor series and compare the coefficients of h, h^2, h^3 and Solving the resulting equations, we obtain the parameter values as

$$W_1 = 1/2, \quad W_2 = 1/2, \quad c_1 = (3 - \sqrt{3})/6, \quad c_2 = (3 + \sqrt{3})/6,$$

$$a_{11} = 1/4, \quad a_{12} = (3-2\sqrt{3})/12, \quad a_{21} = (3 + 2\sqrt{3})/12, \quad a_{22} = 1/4.$$

Since, the truncation error is of $O(h^5)$, the order of the method is 4. The method is given by

$$u_{j+1} = u_j + \frac{1}{2}(K_1 + K_2) \qquad (6.12$$

$$K_1 = hf\left(t_j + \frac{3-\sqrt{3}}{6}h,\ u_j + \frac{1}{4}K_1 + \frac{3-2\sqrt{3}}{12}K_2\right)$$

$$K_2 = hf\left(t_j + \frac{3+\sqrt{3}}{6}h,\ u_j + \frac{3+2\sqrt{3}}{12}K_1 + \frac{1}{4}K_2\right).$$

For obtaining the values of K_1, K_2 we need to solve a system of two nonlinear algebraic equations two unknowns K_1, K_2.

Example 6.21 Solve the initial value problem

$$u' = -2tu^2,\ u(0) = 1$$

with $h = 0.2$ on the interval $[0, 0.4]$. Use the second order implicit Runge-Kutta method.

The second order implicit Runge-Kutta method is given by

$$u_{j+1} = u_j + K_1,\ j = 0, 1 \qquad (6.12$$

$$K_1 = hf\left(t_j + \frac{h}{2},\ u_j + \frac{1}{2}K_1\right)$$

which gives

$$K_1 = -h(2t_j + h)\left(u_j + \frac{1}{2}K_1\right)^2.$$

This is an implicit equation in K_1 and can be solved by using an iterative method. We generally the Newton-Raphson method. We write

$$F(K_1) = K_1 + h(2t_j + h)\left(u_j + \frac{1}{2}K_1\right)^2 = K_1 + 0.2(2t_j + 0.2)\left(u_j + \frac{1}{2}K_1\right)^2$$

have

$$F'(K_1) = 1 + h(2t_j + h)\left(u_j + \frac{1}{2}K_1\right) = 1 + 0.2\,(2t_j + 0.2)\left(u_j + \frac{1}{2}K_1\right).$$

Newton-Raphson method gives

$$K_1^{(s+1)} = K_1^{(s)} - \frac{F_1\left(K_1^{(s)}\right)}{F'\left(K_1^{(s)}\right)}, \; s = 0, 1, \cdots \tag{6.129}$$

assume $K_1^{(0)} = h f(t_j, u_j)$, $j = 0, 1$.

we obtain from (6.128) and (6.129)

$j = 0$: $t_0 = 0$, $u_0 = 1$, $K_1^{(0)} = -h(2t_0 u^2_0) = 0$,

$F(K_1^{(0)}) = 0.04$, $F'(K_1^{(0)}) = 1.04$, $K_1^{(1)} = -0.03846150$.

$F(K_1^{(1)}) = 0.00001483$, $F'(K_1^{(1)}) = 1.03923077$, $K_1^{(2)} = -0.03847567$

$F(K_1^{(2)}) = 0.30 \times 10^{-8}$.

Therefore, $K_1 = K_1^{(2)} = -0.03847567$,

and $u(0.2) = u_1 = u_0 + K_1 = 0.96152433$.

$j = 1$; $t_1 = 0.2$, $u_1 = 0.96152433$, $K_1^{(0)} = -h(2t_1 u_1^2) = -0.07396231$,

$F(K_1^{(0)}) = 0.02861128$, $F'(K_1^{(0)}) = 1.11094517$, $K_1^{(1)} = -0.09971631$,

$F(K_1^{(1)}) = 0.00001989$, $F'(K_1^{(1)}) = 1.10939993$, $K_1^{(2)} = -0.09973423$,

$F(K_1^{(2)}) = 0.35 \times 10^{-7}$, $F'(K_1^{(2)}) = 1.10939885$, $K_1^{(3)} = -0.099773420$.

Therefore, $K_1 = K_1^{(3)} = -0.09973420$,

and $u(0.4) = u_2 = u_1 + K_1 = 0.86179013$.

Second Order Equations

The second order and higher order equations can be solved by considering an equivalent system of first order equations. However, we can also derive single step methods to solve second order or higher order equations directly. Such methods are useful when we consider oscillatory systems, which are usually governed by second order equations.

Consider the second order initial value problem

$$u'' = f(t, u, u'), \; t_0 \le t \le b$$

$$u(t_0) = u_0, \; u'(t_0) = u'_0. \tag{6.130}$$

We present a few methods to solve (6.130) directly.

, we find

$$A = \begin{bmatrix} 1 & E(h) - 1 \\ 0 & E(h) \end{bmatrix}.$$

matrix A for $h = 0.1$ becomes

$$A = \begin{bmatrix} 1 & 0.1051708 \\ 0 & 1.1051708 \end{bmatrix}.$$

$j = 0$, we have

$u_0 = 1$, $u_0' = 1$, $t_1 = 0.1$

$$\begin{bmatrix} u_1 \\ u_1' \end{bmatrix} = \begin{bmatrix} 1 & 0.1051708 \\ 0 & 1.1051708 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.1051708 \\ 1.1051708 \end{bmatrix}.$$

$j = 1$, we have

$u_1 = 1.1051708$, $u_1' = 1.1051708$, $t_2 = 0.2$

$$\begin{bmatrix} u_2 \\ u_2' \end{bmatrix} = \begin{bmatrix} 1 & 0.1051708 \\ 0 & 1.1051708 \end{bmatrix} \begin{bmatrix} 1.1051708 \\ 1.1051708 \end{bmatrix} = \begin{bmatrix} 1.2214025 \\ 1.2214025 \end{bmatrix}.$$

exact solution is $u(t) = e^t$ and the exact values at $t = 0.2$ are given by

$$\begin{bmatrix} u_2 \\ u_2' \end{bmatrix} = \begin{bmatrix} 1.2214028 \\ 1.2214028 \end{bmatrix}.$$

Example 6.23 Solve the initial value problem

$$u'' = (1 + t^2)u,\ u(0) = 1,\ u'(0) = 0,\ t \in [0, 0.4]$$

the Runge-Kutta-Nyström method with $h = 0.2$. Compare with the exact solution $u(t) = e^{t^2/2}$.

$j = 0$, we have

$t_0 = 0$, $u_0 = 1$, $u_0' = 0$

$$K_1 = \frac{h^2}{2} f(t_0, u_0) = \frac{h^2}{2}(1 + t_0^2)u_0 = \frac{(0.2)^2}{2}(1 + 0)1 = 0.02$$

$$K_2 = \frac{h^2}{2} f\left(t_0 + \frac{2}{5}h, u_0 + \frac{2}{5}h\,u_0' + \frac{4}{25}K_1\right)$$

$$= \frac{h^2}{2}\left[1 + \left(t_0 + \frac{2}{5}h\right)^2\right]\left[u_0 + \frac{2}{5}hu_0' + \frac{4}{25}K_1\right]$$

$$= \frac{(0.2)^2}{2}\left[1 + (0.08)^2\right]\left[1 + 0 + \frac{4}{25} \times 0.02\right]$$

$$= (0.02)\ (1.0064)\ (1.0032) = 0.0201924$$

$$K_3 = \frac{h^2}{2} f\left(t_0 + \frac{2}{3}h, u_0 + \frac{2}{3}hu_0' + \frac{4}{9}K_1\right)$$

Chapter 6

$$K_2 = \frac{h^2}{2} f\left(t_1 + \frac{2}{5}h, u_1 + \frac{2}{5}hu_1' + \frac{4}{25}K_1\right)$$

$$= \frac{h^2}{2}\left[1 + \left(t_1 + \frac{2}{5}h\right)^2\right]\left[u_1 + \frac{2}{5}hu_1' + \frac{4}{25}K_1\right]$$

$$= \frac{(0.2)^2}{2}[1 + (0.28)^2][1.0202010 + (0.08)(0.2040329) + (0.16)(0.0212202)]$$

$$= (0.02)\ (1.0784)\ (1.0399189) = 0.0224290$$

$$K_3 = \frac{h^2}{2} f\left(t_1 + \frac{2}{3}h, u_1 + \frac{2}{3}hu_1' + \frac{4}{9}K_1\right)$$

$$= \frac{h^2}{2}\left[1 + \left(t_1 + \frac{2}{3}h\right)^2\right]\left[u_1 + \frac{2}{3}hu_1' + \frac{4}{9}K_1\right]$$

$$= \frac{(0.2)^2}{2}[1 + (0.3333333)^2][1.0202010 + (0.1333333)(0.2040329) + (0.4444444)(0.0212202)]$$

$$= (0.02)\ (1.1111111)\ (1.0568366) = 0.0234853$$

$$K_4 = \frac{h^2}{2} f\left(t_1 + \frac{4}{5}h, u_1 + \frac{4}{5}hu_1' + \frac{8}{25}(K_1 + K_2)\right)$$

$$= \frac{h^2}{2}\left[1 + \left(t_1 + \frac{4h}{5}\right)^2\right]\left[u_1 + \frac{4}{5}hu_1' + \frac{8}{25}(K_1 + K_2)\right]$$

$$= \frac{(0.2)^2}{2}[1 + (0.36)^2][1.0202010 + (0.16)(0.2040329) + (0.32)\ (0.0436492)]$$

$$= (0.02)\ (1.1296)\ (1.0668140) = 0.0241015$$

$$u_2 = u_1 + hu_1' + \frac{1}{96}(23K_1 + 75K_2 - 27K_3 + 25K_4)$$

$$= 1.0202010 + (0.2)(0.2040329) + \frac{1}{96}[23(0.0212202) + 75(0.0224290) - 27\ (0.0234853) + 25\ (0.0241015)]$$

$$= 1.0202010 + 0.0408066 + \frac{1}{96}(2.1386740) = 1.0832855$$

$$u_2' = u_1' + \frac{1}{96h}(23K_1 + 125K_2 - 81K_3 + 125K_4)$$

$$= 0.2040329 + \frac{1}{19.2}[23(0.0212202) + 125(0.0224290) - 81(0.0234853) + 125(0.0241015)]$$

$$= 0.2040329 + \frac{1}{19.2}(4.4020678) = 0.4333073.$$

we have

$$u(0.2) = u_1 = 1.0202010,\ u'(0.2) \approx u_1' = 0.2040329,$$
$$u(0.4) \approx u_2 = 1.0832855,\ u'(0.4) \approx u_2' = 0.4333073.$$

act solution is $u(t) = e^{t^2/2}$. The exact values are given by

$$u(0.2) = 1.02020134, \quad u'(0.2) = 0.20404027,$$
$$u(0.4) = 1.0832871, \quad u'(0.4) = 0.43331484.$$

STABILITY ANALYSIS OF SINGLE STEP METHODS

nalytical solution $u(t_j)$ of the differential equation, the difference solution u_j of the differ... tion and the numerical solution $\bar{u}_j$ can be related by a relation of the form

$$|u(t_j) - \bar{u}_j| \le |u(t_j) - u_j| + |u_j - \bar{u}_j|. \tag{6.1}$$

ctice, we would like the difference between the analytical and numerical solution to be sm... (6.142), we find that this difference depends on the values $|u(t_j) - u_j|$ and $|u_j - \bar{u}_j|$. The v... $- u_j|$ is the truncation error which arises because the differential equation is replaced by ... rence equation. A method, is said to be **consistent** if it is at least of order 1. For a consis... hod, the truncation error tends to zero as h approaches zero. The **numerical error** $|u_j - \bar{u}_j|$ ar... use in actual computation, we cannot compute the difference solution exactly as we are faced w... round-off errors. In fact, in some cases the numerical solution may differ considerably from ... erence solution. If the effect of the total error including the round-off error remains bou... $\to\infty$ with fixed step size, then the difference method is said to be **stable**, otherwise **unstable**. ... w study the stability of the single step methods when applied to the test equation (6.10),

$$u' = \lambda u, \quad u(t_0) = u_0$$

here λ may be a real or complex number. In the previous section, it was shown that the analy... lution of the test equation satisfies the equation (see (6.46))

$$u(t_{j+1}) = e^{\lambda h}\, u(t_j). \tag{6.}$$

we apply any single step method to solve the test equation $u' = \lambda u$, then we get a first o... ifference equation of the form (see also (6.48))

$$u_{j+1} = E(\lambda h)\, u_j,\ j = 0, 1, 2, \cdots \tag{6.}$$

Let $\varepsilon_j = u_j - u(t_j)$. Then, we obtain (see (6.51 a))

$$\varepsilon_{j+1} = [E(\lambda h) - e^{\lambda h}]\, u(t_j) + E(\lambda h)\, \varepsilon_j \tag{6.}$$

first term on the right hand side is the local truncation error and the second term on the right hand is the error propagated from the step t_j to t_{j+1}. The error on the next step t_{j+2} satisfies the equation (6.51 b))

$$\varepsilon_{j+2} = [E^2(\lambda h) - e^{2\lambda h}]\, u(t_j) + E^2(\lambda h)\, \varepsilon_j \tag{6.146}$$

again the first term on the right hand side is the local truncation error and the second term is propagated error. The errors in computations do not grow, if the propagated error tends to zero at least bounded.

all a singlestep method, when applied to the test equation $u' = \lambda u$

solutely stable	if	$\lvert E(\lambda h)\rvert \le 1,\ \lambda < 0,$
atively stable	if	$\lvert E(\lambda h)\rvert \le e^{\lambda h},\ \lambda > 0,$
odically stable	if	$\lvert E(\lambda h)\rvert = 1,\ \lambda$ pure imaginary.

(6.147)

ymptotically stable (A-stable) if $u_j \to 0$ as $j \to \infty$. This implies that the stability interval is h $(-\infty, 0)$, that is the entire left half $h\lambda$ plane.

$\lambda < 0$, the exact solution decreases as t increases and the necessary condition is absolute ility, since the numerical solution must also decrease with t. When $\lambda > 0$, the exact solution ases with t and we do not need the condition $\lvert E(\lambda h)\rvert \le 1$, so that the relative stability is the sary condition to be satisfied.

λ is pure imaginary and $\lvert E(\lambda h)\rvert = 1$, the absolute stability is called periodic stability ability).

r Method

ying the Euler method to the test equation $u' = \lambda u$, we obtain

$$u_{j+1} = u_j + hf_j = (1 + \lambda h)u_j = E(\lambda h)u_j$$
$$E(\lambda h) = 1 + \lambda h.$$

λ is real and $\lambda < 0$, we get the condition $-2 < h\lambda < 0$; (see (6.53)).

λ is complex with $R_e(\lambda) < 0$, let $\bar{h} = \lambda h = x + iy$, then

$$\lvert 1 + \lambda h\rvert = \lvert 1 + x + iy\rvert = \sqrt{(1+x)^2 + y^2}\,.$$

$\lvert 1 + \lambda h\rvert < 1$ gives $(x + 1)^2 + y^2 < 1$, which is the region inside the circle with centre at $(-1, 0)$ and radius 1 (Fig. 6.3).

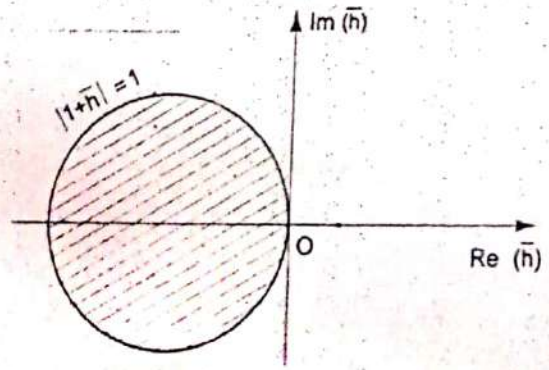

Fig. 6.3. Stability region for Euler's method, $\bar{h} = h\lambda$.

Backward Euler Method

Applying the backward Euler method $u_{j+1} = u_j + hf_{j+1}$, to the test equation $u' = \lambda u$, we get

$$u_{j+1} = u_j + \lambda h u_{j+1}$$

or
$$u_{j+1} = \frac{1}{1-\lambda h} u_j = E(\lambda h) u_j$$

where
$$E(\lambda h) = 1/(1 - \lambda h).$$

When λ is real and $\lambda < 0$, we find that the condition

$$\frac{1}{|1-\lambda h|} < 1$$

is always satisfied. Therefore, the method is absolutely stable for $-\infty < \lambda h < 0$.
When λ is complex with $\mathrm{Re}(\lambda) < 0$, let $\lambda h = x + iy$. Then

$$\frac{1}{|1-\lambda h|} < 1 \text{ or } |1 - \lambda h| > 1 \text{ gives } (1 - x)^2 + y^2 > 1 \tag{6.}$$

which the region outside the circle with centre at (1, 0) and radius 1. Since, $\mathrm{Re}(\lambda h) = x < 0$, condition is always satisfied. Hence, the region of stability is the entire left half of λh-plane.

Runge-Kutta Second Order Method

Applying the Runge-Kutta second order methods (see (6.99))

$$u_{j+1} = u_j + \left(1 - \frac{1}{2c_2}\right) K_1 + \frac{1}{2c_2} K_2$$

where
$$K_1 = hf(t_j, u_j)$$
$$K_2 = hf(t_j + c_2 h, u_j + c_2 K_1)$$

to the test equation $u' = \lambda u$, we get

$$K_1 = h\lambda u_j, \; K_2 = h\lambda(u_j + c_2 K_1) = h\lambda[1 + c_2 h\lambda]u_j$$

$$u_{j+1} = u_j + \left(1 - \frac{1}{2c_2}\right) h\lambda u_j + \frac{1}{2c_2} h\lambda (1 + c_2 h\lambda) u_j$$

$$= \left[1 + h\lambda\left(1 - \frac{1}{2c_2}\right) + \frac{h\lambda}{2c_2}(1 + c_2 h\lambda)\right] u_j$$

$$= \left[1 + h\lambda + \frac{h^2\lambda^2}{2}\right] u_j = E(\lambda h) u_j \tag{6.14}$$

I - unit finish