MAR GREGORIOS COLLEGE OF ARTS & SCIENCE

Block No.8, College Road, Mogappair West, Chennai – 37

Affiliated to the University of Madras
Approved by the Government of Tamil Nadu
An ISO 9001:2015 Certified Institution

DEPARTMENT OF MATHEMATICS

SUBJECT NAME: NUMERICAL METHODS

SUBJECT CODE: TAG3B

SEMESTER: III

PREPARED BY: PROF.R.VASUKI

NUMERICAL METHODS SYLLABUS

UNIT I

Interpolation: Finite differences – operators Δ,δ**,E,D –** relation between operators – linear interpolation – interpolation with equal intervals – Newtons forward interpolation formula – Newton backward interpolation formula.

UNIT II

Numerical solutions of Algebraic, Transcendental and Differential equations: Bisection method – Regula falsi method- Newton Raphson method – Horner's method – Solution of ordinary differential equation – Euler's method (Only Basic)

UNIT III

Simultaneous Linear Algebraic Equations: Method of triangularisation – Gauss elimination method – Inverse of a matrix – Gauss Jordan method.

UNIT IV

Methods of curve fitting: Principles of Least squares – fitting a straight line – linear regression – fitting an exponential curve.

UNIT V

Numerical integration: General Quadrature formula – Trapizoidal rule, Simpson's 1/3 rule and 3/8 rule –Applications – Weddle's rule.

FINITE DIFFERENCES OPERATORS

For a function y=f(x), it is given that $y_0, y_1, ..., y_n$ are the values of the variable y corresponding to the equidistant arguments, $x_0, x_1, ..., x_n$, where $x_1 = x_0 + h, x_2 = x_0 + 2h,\ x_3 = x_0 + 3h, ..., x_n = x_0 + nh$. In this case, even though Lagrange and divided difference interpolation polynomials can be used for interpolation, some simpler interpolation formulas can be derived. For this, we have to be familiar with some finite difference operators and finite differences, which were introduced by Sir Isaac Newton. Finite differences deal with the changes that take place in the value of a function f(x) due to finite changes in x. Finite difference operators include, forward difference operator, backward difference operator, shift operator, central difference operator and mean operator.

- **Forward difference operator (Δ) :**

For the values $y_0, y_1, ..., y_n$ of a function y=f(x), for the equidistant values $x_0, x_1, x_2, ..., x_n$, where $x_1 = x_0 + h, x_2 = x_0 + 2h,\ x_3 = x_0 + 3h, ..., x_n = x_0 + nh$, the forward difference operator Δ is defined on the function f(x) as,

$$\Delta f(x_i) = f(x_i + h) - f(x_i) = f(x_{i+1}) - f(x_i)$$

That is,

$$\Delta y_i = y_{i+1} - y_i$$

Then, in particular

$$\Delta f(x_0) = f(x_0 + h) - f(x_0) = f(x_1) - f(x_0)$$
$$\Rightarrow \quad \Delta y_0 = y_1 - y_0$$

$$\Delta f(x_1) = f(x_1 + h) - f(x_1) = f(x_2) - f(x_1)$$
$$\Rightarrow \quad \Delta y_1 = y_2 - y_1$$

etc.,

$\Delta y_0, \Delta y_1, ..., \Delta y_i, ...$ are known as the **first forward differences.**

The second forward differences are defined as,

$$\begin{aligned}\Delta^2 f(x_i) &= \Delta\left[\Delta f(x_i)\right] = \Delta\left[f(x_i+h) - f(x_i)\right] \\ &= \Delta f(x_i+h) - \Delta f(x_i) \\ &= f(x_i+2h) - f(x_i+h) - \left[f(x_i+h) - f(x_i)\right] \\ &= f(x_i+2h) - 2f(x_i+h) + f(x_i) \\ &= y_{i+2} - 2y_{i+1} + y_i\end{aligned}$$

In particular,

$$\Delta^2 f(x_0) = y_2 - 2y_1 + y_0 \quad or \quad \Delta^2 y_0 = y_2 - 2y_1 + y_0$$

The third forward differences are,

$$\begin{aligned}\Delta^3 f(x_i) &= \Delta\left[\Delta^2 f(x_i)\right] \\ &= \Delta\left[f(x_i+2h) - 2f(x_i+h) + f(x_i)\right] \\ &= y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i\end{aligned}$$

In particular,

$$\Delta^3 f(x_0) = y_3 - 3y_2 + 3y_1 - y_0 \quad or \quad \Delta^3 y_0 = y_3 - 3y_2 + 3y_1 - y_0$$

In general the n[th] forward difference,

$$\Delta^n f(x_i) = \Delta^{n-1} f(x_i+h) - \Delta^{n-1} f(x_i)$$

The differences $\Delta y_0, \Delta^2 y_0, \Delta^3 y_0 \ldots$ are called the **leading differences.**

Forward differences can be written in a tabular form as follows:

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
x_0	$y_0 = f(x_o)$			
		$\Delta y_0 = y_1 - y_0$		
x_1	$y_1 = f(x_1)$		$\Delta^2 y_0 = \Delta y_1 - \Delta y_0$	
		$\Delta y_1 = y_2 - y_1$		$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0$
x_2	$y_2 = f(x_2)$		$\Delta^2 y_1 = \Delta y_2 - \Delta y_1$	
		$\Delta y_2 = y_3 - y_2$		
x_3	$y_3 = f(x_3)$			

Example Construct the forward difference table for the following x values and its corresponding f values.

x	0.1	0.3	0.5	0.7	0.9	1.1	1.3
f	0.003	0.067	0.148	0.248	0.370	0.518	0.697

x	f	Δf	$\Delta^2 f$	$\Delta^3 f$	$\Delta^4 f$	$\Delta^5 f$
0.1	0.003					
		0.064				
0.3	0.067		0.017			
		0.081		0.002		
0.5	0.148		0.019		0.001	
		0.100		0.003		0.000
0.7	0.248		0.022		0.001	
		0.122		0.004		0.000
0.9	0.370		0.026		0.001	
		0.148		0.005		
1.1	0.518		0.031			
		0.179				
1.3	0.697					

Example Construct the forward difference table, where $f(x) = \frac{1}{x}$, x = 1(0.2)2, 4D.

x	$f(x) = \frac{1}{x}$	Δf first difference	$\Delta^2 f$ second difference	$\Delta^3 f$	$\Delta^4 f$	$\Delta^5 f$
1.0	1.000					
		-0.1667				
1.2	0.8333		0.0477			
		-0.1190		-0.0180		
1.4	0.7143		0.0297		0.0082	-0.0045
		-0.0893		-0.0098		
1.6	0.6250		0.0199		0.0037	
		-0.0694		-0.0061		
1.8	0.5556		0.0138			
		-0.0556				
2.0	0.5000					

Example Construct the forward difference table for the data

$$\begin{array}{lccccc} x: & -2 & 0 & 2 & 4 \\ y=f(x): & 4 & 9 & 17 & 22 \end{array}$$

The forward difference table is as follows:

x	y=f(x)	Δy	$\Delta^2 y$	$\Delta^3 y$
-2	4			
		Δy_0 =5		
0	9		$\Delta^2 y_0$ =3	
		Δy_1 =8		$\Delta^3 y_0$ =-6
2	17		$\Delta^2 y_1$ =-3	
		Δy_2 =5		
4	22			

Properties of Forward difference operator (Δ):

(i) Forward difference of a constant function is zero.

Proof: Consider the constant function $f(x)=k$

Then, $\Delta f(x)=f(x+h)-f(x)=k-k=0$

(ii) For the functions $f(x)$ *and* $g(x)$; $\Delta\big(f(x)+g(x)\big)=\Delta f(x)+\Delta g(x)$

Proof: By definition,

$$\begin{aligned}\Delta\big(f(x)+g(x)\big)&=\Delta\big((f+g)(x)\big)\\&=(f+g)(x+h)-(f+g)(x)\\&=f(x+h)+g(x+h)-\big(f(x)+g(x)\big)\\&=f(x+h)-f(x)+g(x+h)-g(x)\\&=\Delta f(x)+\Delta g(x)\end{aligned}$$

(iii)Proceeding as in (ii), for the constants a and b,

$$\Delta\big(af(x)+bg(x)\big)=a\Delta f(x)+b\Delta g(x).$$

(iv)Forward difference of the product of two functions is given by,

$$\Delta\big(f(x)g(x)\big)=f(x+h)\Delta g(x)+g(x)\Delta f(x)$$

Proof:

$$\begin{aligned}\Delta\big(f(x)g(x)\big) &= \Delta\big((fg)(x)\big)\\ &= (fg)(x+h)-(fg)(x)\\ &= f(x+h)g(x+h)-f(x)g(x)\end{aligned}$$

Adding and subtracting $f(x+h)g(x)$, the above gives

$$\Delta\big(f(x)g(x)\big) = f(x+h)g(x+h)-f(x+h)g(x)+f(x+h)g(x)-f(x)g(x)$$

$$\begin{aligned}&= f(x+h)\big[g(x+h)-g(x)\big]+g(x)\big[f(x+h)-f(x)\big]\\ &= f(x+h)\Delta g(x)+g(x)\Delta f(x)\end{aligned}$$

Note : Adding and subtracting $g(x+h)f(x)$ instead of $f(x+h)g(x)$, it can also be proved that

$$\Delta\big(f(x)g(x)\big) = g(x+h)\Delta f(x)+f(x)\Delta g(x)$$

(v) Forward difference of the quotient of two functions is given by

$$\Delta\left(\frac{f(x)}{g(x)}\right)=\frac{g(x)\Delta f(x)-f(x)\Delta g(x)}{g(x+h)g(x)}$$

Proof:

$$\begin{aligned}\Delta\left(\frac{f(x)}{g(x)}\right) &= \frac{f(x+h)}{g(x+h)}-\frac{f(x)}{g(x)}\\ &= \frac{f(x+h)g(x)-f(x)g(x+h)}{g(x+h)g(x)}\\ &= \frac{f(x+h)g(x)-f(x)g(x)+f(x)g(x)-f(x)g(x+h)}{g(x+h)g(x)}\\ &= \frac{g(x)\big[f(x+h)-f(x)\big]-f(x)\big[g(x+h)-g(x)\big]}{g(x+h)g(x)}\\ &= \frac{g(x)\Delta f(x)-f(x)\Delta g(x)}{g(x+h)g(x)}\end{aligned}$$

Following are some results on forward differences:

Result 1: The n^{th} forward difference of a polynomial of degree n is constant when the values of the independent variable are at equal intervals.

Result 2: If n is an integer,

$$f(a+nh) = f(a) + {}^nC_1\Delta f(a) + {}^nC_2\Delta^2 f(a) + \cdots + \Delta^n f(a)$$

for the polynomial f(x) in x.

Forward Difference Table

x	f	Δf	$\Delta^2 f$	$\Delta^3 f$	$\Delta^4 f$	$\Delta^5 f$	$\Delta^6 f$
x_0	f_0						
x_1	f_1	Δf_0	$\Delta^2 f_0$				
x_2	f_2	Δf_1	$\Delta^2 f_2$	$\Delta^3 f_0$	$\Delta^4 f_0$		
x_3	f_3	Δf_2	$\Delta^2 f_2$	$\Delta^3 f_1$	$\Delta^4 f_1$	$\Delta^5 f_0$	$\Delta^6 f_0$
x_4	f_4	Δf_3	$\Delta^2 f_3$	$\Delta^3 f_2$	$\Delta^4 f_2$	$\Delta^5 f_1$	
x_5	f_5	Δf_4	$\Delta^2 f_4$	$\Delta^3 f_3$			
x_6	f_6	Δf_5					

Example Express $\Delta^2 f_0$ and $\Delta^3 f_0$ in terms of the values of the function f.

$$\Delta^2 f_0 = \Delta f_1 - \Delta f_0 = f_2 - f_1 - (f_1 - f_0) = f_2 - 2f_1 + f_0$$

$$\Delta^3 f_0 = \Delta^2 f_1 - \Delta^2 f_0 = \Delta f_2 - \Delta f_1 - (\Delta f_1 - \Delta f_0)$$

$$= (f_3 - f_2) - (f_2 - f_1) - (f_2 - f_1) + (f_1 - f_0)$$

$$= f_3 - 3f_2 + 3f_1 - f_0$$

In general,

$$\Delta^n f_0 = f_n - {}^nC_1 f_{n-1} + {}^nC_2 f_{n-2} - {}^nC_3 f_{n-3} + \ldots + (-1)^n f_0 \ .$$

If we write y_n to denote f_n the above results takes the following forms:

$$\Delta^2 y_0 = y_2 - 2y_1 + y_0$$

$$\Delta^3 y_0 = y_3 - 3y_2 + 3y_1 - y_0$$

$$\Delta^n y_0 = y_n - {}^nC_1\, y_{n-1} + {}^nC_2\, y_{n-2} - {}^nC_3\, y_{n-3} + \ldots + (-1)^n y_0$$

Example Show that the value of y_n can be expressed in terms of the leading value y_0 and the leading differences $\Delta y_0, \Delta^2 y_0, \ldots, \Delta^n y_0$.

Solution

(For notational convenience, we treat y_n as f_n and so on.)

From the forward difference table we have

$$\left.\begin{aligned} \Delta f_0 = f_1 - f_0 \quad &\text{or} \quad f_1 = f_0 + \Delta f_0 \\ \Delta f_1 = f_2 - f_1 \quad &\text{or} \quad f_2 = f_1 + \Delta f_1 \\ \Delta f_2 = f_3 - f_2 \quad &\text{or} \quad f_3 = f_2 + \Delta f_2 \end{aligned}\right\}$$

and so on. Similarly,

$$\left.\begin{aligned} \Delta^2 f_0 = \Delta f_1 - \Delta f_0 \ \text{or}\ \Delta f_1 = \Delta f_0 + \Delta^2 f_0 \\ \Delta^2 f_1 = \Delta f_2 - \Delta f_1 \ \text{or}\ \Delta f_2 = \Delta f_1 + \Delta^2 f_1 \end{aligned}\right\}$$

and so on. Similarly, we can write

$$\left.\begin{aligned} \Delta^3 f_0 = \Delta^2 f_1 - \Delta^2 f_0 \ \text{or}\ \Delta^2 f_1 = \Delta^2 f_0 + \Delta^3 f_0 \\ \Delta^3 f_1 = \Delta^2 f_2 - \Delta^2 f_1 \ \text{or}\ \Delta^2 f_2 = \Delta^2 f_1 + \Delta^3 f_1 \end{aligned}\right\}$$

and so on. Also, we can write f_2 as

$$\begin{aligned} f_2 &= (f_0 + \Delta f_0) + (\Delta f_0 + \Delta^2 f_0) \\ &= f_0 + 2\Delta f_0 + \Delta^2 f_0 \\ &= (1+\Delta)^2 f_0 \end{aligned}$$

Hence

$$\begin{aligned} f_3 &= f_2 + \Delta f_2 \\ &= (f_1 + \Delta f_1) + \Delta f_0 + 2\Delta^2 f_0 + \Delta^3 f_0 \\ &= f_0 + 3\Delta f_0 + 3\Delta^2 f_0 + \Delta^3 f_0 \\ &= (1+\Delta)^3 f_0 \end{aligned}$$

That is, we can symbolically write

$$f_1 = (1+\Delta) f_0 \ , \ f_2 = (1+\Delta)^2 f_0 \ , \ f_3 = (1+\Delta)^3 f_0 .$$

Continuing this procedure, we can show, in general

$$f_n = (1+\Delta)^n f_0 .$$

Using binomial expansion, the above is

$$f_n = f_0 + {}^nC_1\Delta f_0 + {}^nC_2\Delta^2 f_0 + \ldots + \Delta^n f_0$$

Thus

$$f_n = \sum_{i=0}^{n} {}^nC_i\, \Delta^i f_0.$$

Backward Difference Operator

For the values $y_0, y_1, \ldots, y_n$ of a function y=f(x), for the equidistant values $x_0, x_1, \ldots, x_n$, where $x_1 = x_0 + h, x_2 = x_0 + 2h,\ x_3 = x_0 + 3h, \ldots, x_n = x_0 + nh$, the **backward difference operator** ∇ is defined on the function f(x) as,

$$\nabla f(x_i) = f(x_i) - f(x_i - h) = y_i - y_{i-1},$$

which is the **first backward difference.**

In particular, we have the first backward differences,

$$\nabla f(x_1) = y_1 - y_0;\ \nabla f(x_2) = y_2 - y_1 \ etc$$

The second backward difference is given by

$$\begin{aligned}\nabla^2 f(x_i) &= \nabla\left(\nabla f(x_i)\right) = \nabla\left[f(x_i) - f(x_i - h)\right] = \nabla f(x_i) - \nabla f(x_i - h)\\ &= \left[f(x_i) - f(x_i - h)\right] - \left[f(x_i - h) - f(x_i - 2h)\right]\\ &= (y_i - y_{i-1}) - (y_{i-1} - y_{i-2})\\ &= y_i - 2y_{i-1} + y_{i-2}\end{aligned}$$

Similarly, the third backward difference, $\nabla^3 f(x_i) = y_i - 3y_{i-1} + 3y_{i-2} - y_{i-3}$ and so on.

Backward differences can be written in a tabular form as follows:

x	Y	∇y	$\nabla^2 y$	$\nabla^3 y$
x_o	$y_0 = f(x_o)$			
		$\nabla y_1 = y_1 - y_0$		
x_1	$y_1 = f(x_1)$		$\nabla^2 y_2 = \nabla y_2 - \nabla y_1$	
		$\nabla y_2 = y_2 - y_1$		$\nabla^3 y_3 = \nabla^2 y_3 - \nabla^2 y_2$
x_2	$y_2 = f(x_2)$		$\nabla^2 y_3 = \nabla y_3 - \nabla y_2$	
		$\nabla y_3 = y_3 - y_2$		
x_3	$y_3 = f(x_3)$			

Relation between backward difference and other differences:

1. $\Delta y_0 = y_1 - y_0 = \nabla y_1;\ \ \Delta^2 y_0 = y_2 - 2y_1 + y_0 = \nabla^2 y_2$ etc.

2. $\Delta - \nabla = \Delta\nabla$

Proof: Consider the function f(x).

$$\Delta f(x) = f(x+h) - f(x)$$
$$\nabla f(x) = f(x) - f(x-h)$$

$$\begin{aligned}(\Delta - \nabla)(f(x)) &= \Delta f(x) - \nabla f(x) \\ &= [f(x+h) - f(x)] - [f(x) - f(x-h)] \\ &= \Delta f(x) - \Delta f(x-h) \\ &= \Delta[f(x) - f(x-h)] \\ &= \Delta[\nabla f(x)]\end{aligned}$$

$$\Rightarrow \quad \Delta - \nabla = \Delta\nabla$$

3. $\nabla = \Delta E^{-1}$

Proof: Consider the function f(x).

$$\nabla f(x) = f(x) - f(x-h) = \Delta f(x-h) = \Delta E^{-1} f(x) \quad \Rightarrow \quad \nabla = \Delta E^{-1}$$

4. $\nabla = 1 - E^{-1}$

Proof: Consider the function f(x).

$$\nabla f(x) = f(x) - f(x-h) = f(x) - E^{-1} f(x) = \left(1 - E^{-1}\right) f(x) \quad \Rightarrow \quad \nabla = 1 - E^{-1}$$

Problem: Construct the backward difference table for the data

x	-2	0	2	4
$y = f(x)$	-8	3	1	12

Solution: The backward difference table is as follows:

x	Y=f(x)	∇y	$\nabla^2 y$	$\nabla^3 y$
-2	-8			
		∇y_1 =3-(-8)=11		
0	3		$\nabla^2 y_2$ =-2-11= -13	
		∇y_2 =1-3=-2		$\nabla^3 y_3$ =13-(-13)=26
2	1		$\nabla^2 y_3$ =11-(2)=13	
		∇y_3 =12-1=11		
4	12			

Backward Difference Table

x	f	∇f	$\nabla^2 f$	$\nabla^3 f$	$\nabla^4 f$	$\nabla^5 f$	$\nabla^6 f$
x_0	f_0						
x_1	f_1	∇f_1	$\nabla^2 f_2$				
x_2	f_2	∇f_2	$\nabla^2 f_3$	$\nabla^3 f_3$	$\nabla^4 f_4$	$\nabla^5 f$	
x_3	f_3	∇f_3	$\nabla^2 f_4$	$\nabla^3 f_4$	$\nabla^4 f_5$	5	$\nabla^6 f_6$
x_4	f_4	∇f_4	$\nabla^2 f_5$	$\nabla^3 f_5$	$\nabla^4 f_6$	$\nabla^5 f$	
x_5	f_5	∇f_5	$\nabla^2 f_6$	$\nabla^3 f_6$		6	
x_6	f_6	∇f_6					

Example Show that any value of f (or y) can be expressed in terms of f_n (or y_n) and its backward differences.

Solution

$\nabla f_n = f_n - f_{n-1}$ implies $f_{n-1} = f_n - \nabla f_n$

and $\nabla f_{n-1} = f_{n-1} - f_{n-2}$ implies $f_{n-2} = f_{n-1} - \nabla f_{n-1}$

$\nabla^2 f_n = \nabla f_n - \nabla f_{n-1}$ implies $\nabla f_{n-1} = \nabla f_n - \nabla^2 f_n$

From equations (1) to (3), we obtain

$$f_{n-2} = f_n - 2\nabla f_n + \nabla^2 f_n .$$

Similarly, we can show that

$$f_{n-3} = f_n - 3\nabla f_n + 3\nabla^2 f_n - \nabla^3 f_n .$$

Symbolically, these results can be rewritten as follows:

$$f_{n-1} = (1-\nabla) f_n \ , \ f_{n-2} = (1-\nabla)^2 f_n \ , \ f_{n-3} = (1-\nabla)^3 f_n .$$

Thus, in general, we can write

$$f_{n-r} = (1-\nabla)^r f_n .$$

i.e., $f_{n-r} = f_n - {}^rC_1 \nabla f_n + {}^rC_2 \nabla^2 f_n - \ \ldots + \ (-1)^r \nabla^r f_n$

If we write y_n to denote f_n the above result is:

$$y_{n-r} = y_n - {}^rC_1 \nabla y_n + {}^rC_2 \nabla^2 y_n - \ \ldots + \ (-1)^r \nabla^r y_n$$

Central Differences

Central difference operator ¢ for a function f(x) at x_i is defined as,

$$¢ f(x_i) = f\left(x_i + \frac{h}{2}\right) - f\left(x_i - \frac{h}{2}\right)$$, where *h* being the interval of differencing.

Let $y_{\frac{1}{2}} = f\left(x_0 + \frac{h}{2}\right)$. Then,

$$¢ y_{\frac{1}{2}} = ¢ f\left(x_0 + \frac{h}{2}\right) = f\left(x_0 + \frac{h}{2} + \frac{h}{2}\right) - f\left(x_0 + \frac{h}{2} - \frac{h}{2}\right)$$
$$= f(x_0 + h) - f(x_0) = f(x_1) - f(x_0) = y_1 - y_0$$
$$\Rightarrow \quad ¢ y_{\frac{1}{2}} = \Delta y_0$$

Central differences can be written in a tabular form as follows:

x	y	¢ y	¢ ² y	¢ ³ y
x_o	$y_0 = f(x_o)$			
		$¢ y_{\frac{1}{2}} = y_1 - y_0$		
x_1	$y_1 = f(x_1)$		$¢^2 y_1 = ¢ y_{\frac{3}{2}} - ¢ y_{\frac{1}{2}}$	
		$¢ y_{\frac{3}{2}} = y_2 - y_1$		$¢^3 y_{\frac{3}{2}} = ¢^2 y_2 - ¢^2 y_1$
x_2	$y_2 = f(x_2)$		$¢^2 y_2 = ¢ y_{\frac{5}{2}} - ¢ y_{\frac{3}{2}}$	
		$¢ y_{\frac{5}{2}} = y_3 - y_2$		
x_3	$y_3 = f(x_3)$			

Central Difference Table

x	f	δf	$\delta^2 f$	$\delta^3 f$	$\delta^4 f$
x_0	f_0				
		$\delta f_{1/2}$			
x_1	f_1		$\delta^2 f_1$		
		$\delta f_{3/2}$		$\delta^3 f_{3/2}$	
x_2	f_2		$\delta^2 f_2$		$\delta^4 f_2$
		$\delta f_{5/2}$		$\delta^3 f_{5/2}$	
x_3	f_3		$\delta^2 f_3$		
		$\delta f_{7/2}$			
x_4	f_4				

Example Show that

(a) $\Delta^2 f_m = f_{m+1} - 2f_m + f_{m-1}$

(b) $\Delta^3 f_{m+\frac{1}{2}} = f_{m+2} - 3f_{m+1} + 3f_m - f_{m-1}$

(a) $\delta^2 f_m = \delta f_{m+1/2} - \delta f_{m-1/2} = (f_{m+1} - f_m) - (f_m - f_{m-1})$

$= f_{m+1} - 2f_m + f_{m-1}$

(b) $\delta^3 f_{m+1/2} = \delta^2 f_{m+1} - \delta^2 f_m = \left(f_{m+2} - 2f_{m+1} + f_m\right) -$

$\left(f_{m+1} - 2f_m + f_{m-1}\right) = f_{m+2} - 3f_{m+1} + 3f_m - f_{m-1}$

Shift operator, E

Let $y = f(x)$ be a function of x, and let x takes the consecutive values x, $x + h$, $x + 2h$, etc. We then define an operator E, called **the shift operator** having the property

$$E f(x) = f(x + h) \qquad \ldots(1)$$

Thus, when E operates on $f(x)$, the result is the next value of the function. If we apply the operator twice on $f(x)$, we get

$$E^2 f(x) = E[E f(x)] = f(x + 2h).$$

Thus, in general, if we apply the shift operator n times on $f(x)$, we arrive at

$$E^n f(x) = f(x + nh) \qquad \ldots(2)$$

for all real values of n.

If f_0 $(= y_0)$, f_1 $(= y_1)\ldots$ are the consecutive values of the function

$y = f(x)$, then we can also write

$E f_0 = f_1$ (or $E y_0 = y_1$), $E f_1 = f_2$ (or $E y_1 = y_2)\ldots$

$E^2 f_0 = f_2$ (or $E^2 y_0 = y_2$), $E^2 f_1 = f_3$ (or $E y_1 = y_3)\ldots$

$E^3 f_0 = f_3$ (or $E^2 y_0 = y_3$), $E^3 f_1 = f_4$ (or $E y_1 = y_4)\ldots$

and so on. The **inverse operator** E^{-1} is defined as:

$$E^{-1} f(x) = f(x - h) \qquad \ldots(3)$$

and similarly

$$E^{-n} f(x) = f(x - nh) \qquad \ldots(4)$$

Average Operator μ

The **average operator** μ is defined as

$$\mu f(x) = \frac{1}{2}\left[f\left(x + \tfrac{h}{2}\right) + f\left(x - \tfrac{h}{2}\right)\right]$$

Differential operator D

The **differential operator** D has the property

$$Df(x) = \frac{d}{dx} f(x) = f'(x)$$

$$D^2 f(x) = \frac{d^2}{dx^2} f(x) = f''(x)$$

Relations between the operators:

Operators $\Delta, \nabla, \delta, \mu$ and D in terms of E

From the definition of operators Δ and E, we have

$$\Delta f(x) = f(x + h) - f(x) = E f(x) - f(x) = (E - 1) f(x).$$

Therefore,

$$\Delta = E - 1$$

From the definition of operators ∇ and E^{-1}, we have

$$\nabla f(x) = f(x) - f(x - h) = f(x) - E^{-1} f(x) = (1 - E^{-1}) f(x).$$

Therefore,

$$\nabla = 1 - E^{-1} = \frac{E - 1}{E}.$$

The definition of the operators δ and E gives

$$\delta f(x) = f(x + h/2) - f(x - h/2) = E^{1/2} f(x) - E^{-1/2} f(x)$$

$$= (E^{1/2} - E^{-1/2}) f(x).$$

Therefore,

$$\delta = E^{1/2} - E^{-1/2}$$

The definition of the operators « and E yields

$$\mu f(x) = \frac{1}{2}\left[f\left(x + \frac{h}{2}\right) + f\left(x - \frac{h}{2}\right)\right] = \frac{1}{2}\left[E^{1/2} + E^{-1/2}\right] f(x).$$

Therefore,

$$\mu = \frac{1}{2}\left(E^{1/2} + E^{-1/2}\right).$$

It is known that

$$E f(x) = f(x + h).$$

Using the Taylor series expansion, we have

$$Ef(x) = f(x) + h\ f'(x) + \frac{h^2}{2!} f''(x) + \ldots$$

$$= f(x) + h\, Df(x) + \frac{h^2}{2!} D^2(x) + \ldots$$

$$= \left(1 + \frac{h\, D}{1!} + \frac{h^2 D^2}{2!} + \ldots\right) f(x) = e^{hD} f(x).$$

Thus $E = e^{hD}$. Or,

$$hD = \log E.$$

Example If Δ, ∇, δ denote forward, backward and central difference operators, E and « respectively the shift operator and average operators, in the analysis of data with equal spacing h, prove the following:

$(i)\ 1 + \delta^2 «^2 = \left(1 + \frac{\delta^2}{2}\right)^2$ $(ii)\ E^{1/2} = « + \frac{\delta}{2}$

$(iii)\ \Delta = \frac{\delta^2}{2} + \delta\sqrt{1 + \left(\delta^2/4\right)}$

$(iv)\ \mu\delta = \frac{\Delta E^{-1}}{2} + \frac{\Delta}{2}$ $(v)\ \mu\delta = \frac{\Delta + \nabla}{2}.$

Solution

(i) From the definition of operators, we have

$$\mu\delta = \frac{1}{2}\left(E^{1/2} + E^{-1/2}\right)\left(E^{1/2} - E^{-1/2}\right) = \frac{1}{2}\left(E - E^{-1}\right).$$

Therefore

$$1 + \mu^2\delta^2 = 1 + \frac{1}{4}\left(E^2 - 2 + E^{-2}\right) = \frac{1}{4}\left(E + E^{-1}\right)^2$$

Also,

$$1 + \frac{\delta^2}{2} = 1 + \frac{1}{2}\left(E^{1/2} - E^{-1/2}\right)^2 = \frac{1}{2}\left(E + E^{-1}\right)$$

From equations (1) and (2), we get

$$1 + \delta^2\mu^2 = \left(1 + \frac{\delta^2}{2}\right)^2.$$

(*ii*) $\mu + \frac{\delta}{2} = \frac{1}{2}\left(E^{1/2} + E^{-1/2} + E^{1/2} - E^{-1/2}\right) = E^{1/2}.$

(*iii*) We can write

$$\frac{\delta^2}{2} + \delta\sqrt{1 + \left(\delta^2/4\right)} = \frac{\left(E^{1/2} - E^{-1/2}\right)^2}{2} + \left(E^{1/2} - E^{-1/2}\right)\sqrt{1 + \frac{1}{4}\left(E^{1/2} - E^{-1/2}\right)^2}$$

$$= \frac{E - 2 + E^{-1}}{2} + \frac{1}{2}\left(E^{1/2} - E^{-1/2}\right)\left(E^{1/2} + E^{-1/2}\right)$$

$$= \frac{E - 2 + E^{-1}}{2} + \frac{E - E^{-1}}{2}$$

$$= E - 1$$

$$= \Delta$$

(*iv*) We write

$$\mu\delta = \frac{1}{2}\left(E^{1/2} + E^{-1/2}\right)\left(E^{1/2} - E^{-1/2}\right) = \frac{1}{2}\left(E - E^{-1}\right)$$

$$= \frac{1}{2}\left(1 + \Delta - E^{-1}\right) = \frac{\Delta}{2} + \frac{1}{2}\left(1 - E^{-1}\right) = \frac{\Delta}{2} + \frac{1}{2}\left(\frac{E - 1}{E}\right) = \frac{\Delta}{2} + \frac{\Delta}{2E}.$$

(*v*) We can write

$$\mu\delta = \frac{1}{2}\left(E^{1/2} + E^{-1/2}\right)\left(E^{1/2} - E^{-1/2}\right) = \frac{1}{2}\left(E - E^{-1}\right)$$

$$= \frac{1}{2}\left(1 + \Delta - (1 - \nabla)\right) = \frac{1}{2}(\Delta + \nabla).$$

Example Prove that

$$hD = \log(1+\Delta) = -\log(1-\nabla) = \sinh^{-1}(\mu\delta).$$

Using the standard relations given in boxes in the last section, we have

$$hD = \log E = \log(1+\Delta) = \log E = -\log E^{-1} = -\log(1+\nabla)$$

Also,

$$\mu\delta = \frac{1}{2}\left(E^{1/2} + E^{-1/2}\right)\left(E^{1/2} - E^{-1/2}\right) = \frac{1}{2}\left(E + E^{-1}\right)$$

$$= \frac{1}{2}\left(e^{hD} - e^{-hD}\right) = \sin(hD)$$

Therefore

$$hD = \sinh^{-1}(\mu\delta).$$

Example Show that the operations « and E commute.

Solution

From the definition of operators « and E, we have

$$\mu E f_0 = \mu f_1 = \frac{1}{2}\left(f_{3/2} + f_{1/2}\right)$$

and also

$$E\mu f_0 = \frac{1}{2}E\left(f_{1/2} + f_{-1/2}\right) = \frac{1}{2}\left(f_{3/2} + f_{1/2}\right)$$

Hence

$$\mu E = E\mu.$$

Therefore, the operators « and E commute.

Example Show that

$$e^x\left(u_0 + x\Delta u_0 + \frac{x^2}{2!}\Delta^2 u_0 + \ldots\right) = u_0 + u_1 x + u_2\frac{x^2}{2!} + \ldots$$

$$e^x\left(u_0 + x\Delta u_0 + \frac{x^2}{2!}\Delta^2 u_0 + \ldots\right) = e^x\left(1 + x\Delta + \frac{x^2\Delta^2}{2!} + \ldots\right)u_0$$

$$= e^x e^{x\Delta} u_0 = e^{x(1+\Delta)} u_0$$

$$= e^{xE} u_0$$

$$= \left(1 + xE + \frac{x^2E^2}{2!} + \ldots\right)u_0$$

$$= u_0 + xu_1 + \frac{x^2}{2!}u_2 + \ldots,$$

as desired.

Example Using the method of separation of symbols, show that

$$\Delta^n u_{x-n} = u_x - nu_{x-1} + \frac{n(n-1)}{2}u_{x-2} + \cdots + (-1)^n u_{x-n}.$$

To prove this result, we start with the right-hand side. Thus,

$$\text{R.H.S} \quad = u_x - nu_{x-1} + \frac{n(n-1)}{2}u_{x-2} + \cdots + (-1)^n u_{x-n}.$$

$$= u_x - nE^{-1}u_x + \frac{n(n-1)}{2}E^{-2}u_x + \cdots + (-1)^n E^{-n}u_x$$

$$= \left[1 - nE^{-1} + \frac{n(n-1)}{2}E^{-2} + \cdots + (-1)^n E^{-n}\right]u_x$$

$$= \left(1 - E^{-1}\right)^n u_x$$

$$= \left(1 - \frac{1}{E}\right)^n u_x$$

$$= \left(\frac{E-1}{E}\right)^n u_x$$

$$= \frac{\Delta^n}{E^n}u_x$$

$$= \Delta^n E^{-n}u_x$$

$$= \Delta^n u_{x-n},$$

$$= \text{L.H.S}$$

Differences of a Polynomial

Let us consider the polynomial of degree n in the form

$$f(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n,$$

where $a_0 \neq 0$ and $a_0, a_1, a_2, \ldots, a_{n-1}, a_n$ are constants. Let h be the interval of differencing. Then

$$f(x+h) = a_0(x+h)^n + a_1(x+h)^{n-1} + a_2(x+h)^{n-2} + \ldots \quad + a_{n-1}(x+h) + a_n$$

Now the difference of the polynomials is:

$$\Delta f(x) = f(x+h) - f(x) = a_0\left[(x+h)^n - x^n\right] + a_1\left[(x+h)^{n-1} - x^{n-1}\right] + \ldots \qquad + a_{n-1}(x+h-x)$$

Binomial expansion yields

$$\begin{aligned}\Delta f(x) \;=\; & a_0\left[x^n + {}^nC_1 x^{n-1}h + {}^nC_2 x^{n-2}h^2 + \ldots + h^n - x^n\right] \\ & + a_1[x^{n-1} + {}^{(n-1)}C_1 x^{n-2}h + {}^{(n-1)}C_2 x^{n-3}h^2 \\ & \qquad + \ldots + h^{n-1} - x^{n-1}] + \ldots + a_{n-1}h \\ = \; & a_0 nhx^{n-1} + \left[a_0\,{}^nC_2 h^2 + a_1\,{}^{(n-1)}C_1 h\right]x^{n-2} + \ldots + a_{n-1}h.\end{aligned}$$

Therefore,

$$\Delta f(x) \;=\; a_0 nhx^{n-1} + b'\,x^{n-2} + c'\,x^{n-3} + \ldots + k'x + l',$$

where bà, cà, . . . , kà, là are constants involving h but not x. Thus, the first difference of a polynomial of degree n is another polynomial of degree (n J 1). Similarly,

$$\begin{aligned}\Delta^2 f(x) = \Delta(\Delta f(x)) \;&= \Delta f(x+h) - \Delta f(x) \\ &= a_0 nh\left[(x+h)^{n-1} - x^{n-1}\right] + b'\left[(x+h)^{n-2} - x^{n-2}\right] \\ &\qquad + \ldots + k'(x+h-x)\end{aligned}$$

On simplification, it reduces to the form

$$\Delta^2 f(x) = a_0 n(n-1)h^2 x^{n-2} + b''x^{n-3} + c''x^{n-4} + \ldots + q''.$$

Therefore, $\Delta^2 f(x)$ is a polynomial of degree (n J 2) in x. Similarly, we can form the higher order differences, and every time we observe that the degree of the polynomial is reduced by 1. After differencing n times, we are left with only the first term in form

$$\begin{aligned}\Delta^n f(x) &= a_0 n(n-1)(n-2)(n-3)\ldots(2)(1)h^n \\ &= a_0(n!)h^n = \text{constant}.\end{aligned}$$

This constant is independent of x. Since $\Delta^n f(x)$ is a constant $\Delta^{n+1} f(x) = 0$. Hence the (n + 1)th and higher order differences of a polynomial of degree n are 0.

Conversely, if the nth differences of a tabulated function are constant and the $(n+1)$th, $(n+2)$th,..., differences all vanish, then the tabulated function represents a polynomial of degree n. It should be noted that these results hold good only if the values of x are equally spaced. The converse is important in numerical analysis since it enables us to approximate a function by a polynomial if its differences of some order become nearly constant.

Theorem (*Differences of a polynomial*)The nth differences of a polynomial of degree n is a constant, when the values of the independent variable are given at equal intervals.

Exercises

1. Calculate $f(x)=\frac{1}{x+1}$, $x=0(0.2)1$ to (a) 2 decimal places, (b) 3 decimal places and (c)4 decimal places. Then compare the effect of rounding errors in the corresponding difference tables.
2. Express $\Delta^2 y_1$ (i.e. $\Delta^2 f_1$) and $\Delta^4 y_0$ (i.e. $\Delta^4 f_0$) in terms of the values of the function $y = f(x)$.
3. Set up a difference table of $f(x)=x^2$ for $x=0(1)10$. Do the same with the calculated value 25 of $f(5)$ replaced by 26. Observe the spread of the error.
4. Calculate $f(x)=\frac{1}{x+1}$, $x=0(0.2)1$ to (a)2 decimal places, (b)3 decimal places and (c)4 decimal places. Then compare the effect of rounding errors in the corresponding difference tables.
5. Set up a forward difference table of $f(x) = x^2$ for $x = 0(1)10$. Do the same with the calculated value 25 of $f(5)$ replaced by 26. Observe the spread of the error.
6. Construct the difference table based on the following table.

x	0.0	0.1	0.2	0.3	0.4	0.5
cos$x$	1.000 00	0.995 00	0.980 07	0.955 34	0.921 06	0.877 58

7. Construct the difference table based on the following table.

x	0.0	0.1	0.2	0.3	0.4	0.5
sin$x$	0.000 00	0.099 83	0.198 67	0.295 52	0.389 42	0.479

8. Construct the backward difference table, where

$$f(x)=\sin x,\ x = 1.0(0.1)1.5,\ 4D.$$

9. Show that $E\nabla = \Delta = \delta E^{1/2}$.

10. Prove that

11. $(i)\ \delta = 2\sinh(hD/2)$ and $(ii)\ \mu = 2\cosh(hD/2)$.

12. Show that the operators δ, « , E, Δ and ∇ commute with each other.

13. Construct the backward difference table based on the following table.

x	0.0	0.1	0.2	0.3	0.4	0.5
cos x	1.000 00	0.995 00	0.980 07	0.955 34	0.921 06	0.877 58

Construct the difference table based on the following table.

x	0.0	0.1	0.2	0.3	0.4	0.5
sin x	0.000 00	0.099 83	0.198 67	0.295 52	0.389 42	0.479 43

6. Construct the backward difference table, where

$f(x) = \sin x,\ \ x = 1.0(0.1)1.5$, 4D.

7. Evaluate $(2a + 3)(E + 2)(3x^2 + 2)$, interval of differencing being unity.

8. Compute the missing values of y_n and Δy_n in the following table:

y_n	Δy_n	$\Delta^2 y_n$
-		
-	-	1
-	-	4
6	5	13
-	-	18
-	-	24
-	-	

NUMERICAL INTERPOLATION

Consider a single valued continuous function $y = f(x)$ defined over [a,b] where $f(x)$ is known explicitly. It is easy to find the values of '*y*' for a given set of values of '*x*' in [a,b]. i.e., it is possible to get information of all the points (x, y) where $a \le x \le b$.

But the converse is not so easy. That is, using only the points $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$ where $a \le x_i \le b, i = 0,1,2,...,n$, it is not so easy to find the relation between x and y in the form $y = f(x)$ explicitly. That is one of the problem we face in numerical differentiation or integration.

Now we have first to find a simpler function, say $g(x)$, such that $f(x)$ and $g(x)$ agree at the given set of points and accept the value of $g(x)$ as the required value of $f(x)$ at some point x in between a and b. Such a process is called **interpolation.** If $g(x)$ is a polynomial, then the process is called polynomial interpolation.

When a function *f*(*x*) is not given explicitly and only values of $f(x)$ are given at a set of distinct points called *nodes* or *tabular points*, using the interpolated function $g(x)$ to the function *f*(*x*), the required operations intended for $f(x)$, like determination of roots, differentiation and integration etc. can be carried out. The approximating polynomial $g(x)$ can be used to predict the value of $f(x)$ at a non- tabular point. The deviation of $g(x)$ from $f(x)$, that is $|f(x) - g(x)|$ is called the *error of approximation*.

Consider a continuous single valued function $f(x)$ defined on an interval [*a, b*]. Given the values of the function for n + 1 distinct tabular points $x_0, x_1, ..., x_n$ such that $a \le x_0 \le x_1 \le ... \le x_n \le b$. The problem of polynomial interpolation is to find a polynomial *g(x) or* $p_n(x)$, of degree *n*, which fits the given data. The interpolation polynomial fitted to a given data is unique.

If we are given two points satisfying the function such as $(x_0, y_0); (x_1, y_1)$, where $y_0 = f(x_0)$ and $y_1 = f(x_1)$ it is possible to fit a unique polynomial of degree 1. If three distinct points are given, a polynomial of degree not greater than two can be fitted uniquely. In general, if n+ *1 distinct points* are given, a polynomial of degree not greater than *n* can be fitted uniquely.

Interpolation fits a real function to discrete data. Given the set of tabular values $(x_0, y_0), (x_1, y_1) \ldots . (x_n, y_n)$ satisfying the relation $y = f(x)$, where the explicit nature of

$f(x)$ is not known, and it is required to find the values of $f(x)$ corresponding to certain given values of x in between x_0 and x_n. To do this we have first to find a simpler function, say $g(x)$, such that $f(x)$ and $g(x)$ agree at the set of tabulated points and accept the value of $g(x)$ as the required value of $f(x)$ at some point x in between x_0 and x_n. Such a process is called **interpolation**. If $g(x)$ is a polynomial, then the process is called **polynomial interpolation**.

In interpolation, we have to determine the function $g(x)$, in the case that $f(x)$ is difficult to be obtained, using the **pivotal values** $f_0 = f(x_0),\ f_1 = f(x_1), \ldots,\ f_n = f(x_n)$.

Linear interpolation

In linear interpolation, we are given with two pivotal values $f_0 = f(x_0)$ and $f_1 = f(x_1)$, and we approximate the curve of f by a chord (straight line) P_1 passing through the points $(x_0,\ f_0)$ and $(x_1,\ f_1)$. Hence the approximate value of f at the intermediate point $x = x_0 + rh$ is given by the **linear interpolation formula**

$$f(x) \approx P_1(x) = f_0 + r(f_1 - f_0) = f_0 + r\Delta f_0$$

where $r = \dfrac{x - x_0}{h}$ and $0 \le r \le 1$.

Example Evaluate $\ln 9.2$, given that $\ln 9.0 = 2.197$ and $\ln 9.5 = 2.251$.

Here $x_0 = 9.0$, $x_1 = 9.5$, $h = x_1 - x_0 = 9.5 - 9.0 = 0.5$, $f_0 = f(x_0) = \ln 9.0 = 2.197$ and $f_1 = f(x_1) = \ln 9.5 = 2.251$. Now to calculate $\ln 9.2 = f(9.2)$, take $x = 9.2$, so that

$$r = \frac{x - x_0}{h} = \frac{9.2 - 9.0}{0.5} = \frac{0.2}{0.5} = 0.4 \text{ and hence}$$

$$\ln 9.2 = f(9.2) \approx P_1(9.2) = f_0 + r(f_1 - f_0) = 2.197 + 0.4\,(2.251 - 2.197) = 2.219$$

Example Evaluate $f(15)$, given that $f(10) = 46$, $f(20) = 66$.

Here $x_0 = 10$, $x_1 = 20$, $h = x_1 - x_0 = 20 - 10 = 10$,

$f_0 = f(x_0) = 46$ and $f_1 = f(x_1) = 66$.

Now to calculate $f(15)$, take $x = 15$, so that

$$r = \frac{x - x_0}{h} = \frac{15 - 10}{10} = \frac{5}{10} = 0.5$$

and hence $f(15) \approx P_1(15) = f_0 + r(f_1 - f_0) = 46 + 0.5\,(66 - 46) = 56$

Example Evaluate $e^{1.24}$, given that $e^{1.1} = 3.0042$ and $e^{1.4} = 4.0552$.

Here $x_0 = 1.1$, $x_1 = 1.4$, $h = x_1 - x_0 = 1.4 - 1.1 = 0.3$, $f_0 = f(x_0) = 1.1$ and $f_1 = f(x_1) = 1.24$. Now to calculate $e^{1.24} = f(1.24)$, take $x = 1.24$, so that $r = \frac{x - x_0}{h} = \frac{1.24 - 1.1}{0.3} = \frac{0.14}{0.3} = 0.4667$ and hence

$e^{1.24} \approx P_1(1.24) = f_0 + r(f_1 - f_0) = 3.0042 + 0.4667(4.0552 - 3.0042) = 3.4933$, while the exact value of $e^{1.24}$ is 3.4947.

Quadratic Interpolation

In quadratic interpolation we are given with three pivotal values $f_0 = f(x_0)$, $f_1 = f(x_1)$ and $f_2 = f(x_2)$ and we approximate the curve of the function f between x_0 and $x_2 = x_0 + 2h$ by the quadratic parabola P_2, which passes through the points $(x_0, f_0), (x_1, f_1), (x_2, f_2)$ and obtain the quadratic interpolation formula

$$f(x) \approx P_2(x) = f_0 + r\Delta f_0 + \frac{r(r-1)}{2}\Delta^2 f_0$$

where $r = \frac{x - x_0}{h}$ and $0 \le r \le 2$.

Example Evaluate ln 9.2, using quadratic interpolation, given that

ln 9.0 = 2.197, ln 9.5 = 2.251 and ln10.0 = 2.3026.

Here $x_0 = 9.0$, $x_1 = 9.5$, $x_1 = 10.0$, $h = x_1 - x_0 = 9.5 - 9.0 = 0.5$, $f_0 = f(x_0) = \ln 9.0 = 2.197$, $f_1 = f(x_1) = \ln 9.5 = 2.251$ and $f_2 = f(x_2) = \ln 10.0 = 2.3026$. Now to calculate $\ln 9.2 = f(9.2)$, take $x = 9.2$, so that $r = \frac{x - x_0}{h} = \frac{9.2 - 9.0}{0.5} = \frac{0.2}{0.5} = 0.4$ and

$$\ln 9.2 = f(9.2) \approx P_2(x) = f_0 + r\Delta f_0 + \frac{r(r-1)}{2}\Delta^2 f_0$$

To proceed further, we have to construct the following forward difference table.

x	f	Δf	$\Delta^2 f$
9.0	2.1972		
9.5	2.2513	0.0541	-0.0028
10.0	2.3026	0.0513	

Hence,

$\ln 9.2 = f(9.2) \approx P_2(9.2) = 2.1972 + 0.4(0.0541) + \frac{0.4(0.4-1)}{2}(-0.0028)$ = 2.2192, which exact to 4D to the exact value of $\ln 9.2 = 2.2192$.

Example Using the values given in the following table, find cos0.28 by linear interpolation and by quadratic interpolation and compare the results with the value 0.96106 (exact to 5D)

x	$f(x) = \cos x$	First difference	Second difference
0.0	1.00000		
		-0.01993	
0.2	0.98007		-0.03908
		-0.05901	
0.4	0.92106		

Here $f(x)$, where $x_0 = 0.28$ is to determined. In linear interpolation, we need two consecutive x values and their corresponding f values and first difference. Here, since x=0.28 lies in between 0.2 and 0.4, we take x_0 = 0.2, x_1 = 0.4. (**Attention!** Choosing $x_0 = 0.2$, $x_1 = 0.4$ is very important; taking $x_0 = 0.0$ would give wrong answer). Then $h = x_1 - x_0$ =0.4–0.2 = 0.2, $f_0 = f(x_0)$ =0.98007 and $f_1 = f(x_1)$ =0.92106.

Also $r = \frac{x - x_0}{h} = \frac{0.28 - 0.2}{0.2} = \frac{0.08}{0.2} = 0.4$ and

$$\cos 0.28 = f(0.28) \approx P_1(0.28) = f_0 + r(f_1 - f_0)$$

$$= 0.98007 + 0.4(0.92106 - 0.98007)$$

= 0.95647, correct to 5 D.

In quadratic interpolation, we need three consecutive (equally spaced) x values and their corresponding f values, first differences and second difference. Here x_0 = 0.0 , x_1 = 0.2, x_1 = 0.4, $h = x_1 - x_0$ =0.2 – 0.0 = 02, f_0 = 1.00000, f_1 = 0.98007 and f_2 = 0.92106, Δf_0=-0.01993, $\Delta^2 f_0$=-0.03908 $r = \frac{x - x_0}{h} = \frac{0.28 - 0.00}{0.2} = 1.4$ and

$$\cos 0.28 \approx P_2(0.28) = f_0 + r\Delta f_0 + \frac{r(r-1)}{2}\Delta^2 f_0$$

$$= 1.00 + 1.4(-0.-1993) + \frac{1.4(1.4-1)}{2}(-0.03908) = 0.96116 \text{ to 5D.}$$

From the above, it can be seen that quadratic interpolation gives more accurate value.

Newton's Forward Difference Interpolation Formula

Using Newton's forward difference interpolation formula we find the n degree polynomial P_n which approximates the function $f(x)$ in such a way that P_n and f agrees at n+1 equally spaced x values, so that $P_n(x_0) = f_0, P_n(x_1) = f_1, \ldots, P_n(x_n) = f_n$, where $f_0 = f(x_0)$, $f_1 = f(x_1)$, …, $f_n = f(x_n)$ are the values of f in the table.

Newton's forward difference interpolation formula is

$$f(x) \approx P_n(x) =$$

$$= f_0 + r\Delta f_0 + \frac{r(r-1)}{2!}\Delta^2 f_0 + \ldots + \frac{r(r-1)\ldots(r-n+1)}{n!}\Delta^n f_0$$

where $x = x_0 + rh,\ r = \frac{x - x_0}{h},\ 0 \le r \le n$.

Derivation of Newton's forward Formulae for Interpolation

Given the set of $(n+1)$ values, viz., $(x_0, f_0), (x_1, f_1), (x_2, f_2), \ldots, (x_n, f_n)$

of x and f, it is required to find $p_n(x)$, a polynomial of the nth degree such that $f(x)$ and $p_n(x)$ agree at the tabulated points. Let the values of x be equidistant, i.e., let

$$x_i = x_0 + rh, \qquad r = 0, 1, 2, \ldots, n$$

Since $p_n(x)$ is a polynomial of the nth degree, it may be written as

$$\left.\begin{aligned} p_n(x) = a_0 &+ a_1(x - x_0) + a_2(x - x_0)(x - x_1) \\ &+ a_3(x - x_0)(x - x_1)(x - x_2) + \ldots \\ &+ a_n(x - x_0)(x - x_1)(x - x_2)\ldots(x - x_{n-1}) \end{aligned}\right\}$$

Imposing now the condition that $f(x)$ and $p_n(x)$ should agree at the set of tabulated points, we obtain

$$a_0 = f_0;\ a_1 = \frac{f_1 - f_0}{x_1 - x_0} = \frac{\Delta f_0}{h};\ a_2 = \frac{\Delta^2 f_0}{h^2 2!};\ a_3 = \frac{\Delta^3 f_0}{h^3 3!};\ldots;\ a_n = \frac{\Delta^n f_0}{h^n n!};$$

Setting $x = x_0 + rh$ and substituting for $a_0, a_1, \ldots, a_n$, we obtain the expression.

Remark 1:

Newton's forward difference formula has the permanence property. If we add a new set of value (x_{n+1}, y_{n+1}), to the given set of values, then the forward difference table gets a new column of $(n+1)^{th}$ forward difference. Then the Newton's Forward difference

Interpolation Formula with the already given values will be added with a new term at the end, $(x-x_0)(x-x_1).....(x-x_n)\frac{1}{(n+1)!h^{n+1}}\left[\Delta^{n+1}y_0\right]$ to get the new interpolation formula with the newly added value.

Remark 2:

Newton's forward difference interpolation formula is useful for interpolation near the beginning of a set of tabular values and for extrapolating values of y a short distance backward, that is left from y_0.The process of finding the value of y for some value of x outside the given range is called *extrapolation.*

Example Using Newton's forward difference interpolation formula and the following table evaluate $f(15)$.

x	$f(x)$	Δf	$\Delta^2 f$	$\Delta^3 f$	$\Delta^4 f$
10	46				
		20			
20	66		-5	2	
		15			
30	81		-3	-1	-3
		12			
40	93		-4		
		8			
50	101				

Here $x = 15$, $x_0 = 10$, $x_1 = 20$, $h = x_1 - x_0 = 20 - 10 = 10$, $r = (x - x_0)/h = (15-10)/10 = 0.5$, $f_0 = -46$, $\Delta f_0 = 20$, $\Delta^2 f_0 = -5$, $\Delta^3 f_0 = 2$, $\Delta^4 f_0 = -3$.

Substituting these values in the Newton's forward difference interpolation formula for $n = 4$, we obtain

$$f(x) \approx P_4(x) = f_0 + r\Delta f_0 + \frac{r(r-1)}{2!}\Delta^2 f_0 + \ldots + \frac{r(r-1)\ldots(r-4+1)}{4!}\Delta^4 f_0,$$

so that

$$f(15) \approx 46 + (0.5)(20) + \frac{(0.5)(0.5-1)}{2!}(-5) + \frac{(0.5)(0.5-1)(0.5-2)}{3!}(2)$$

$$+\frac{(0.5(0.5-1)(0.5-2)(0.5-3)}{4!}(-3)$$

= 56.8672, correct to 4 decimal places.

Example Find a cubic polynomial in x which takes on the values -3, 3, 11, 27, 57 and 107, when x=0, 1, 2, 3, 4 and 5 respectively.

x	$f(x)$	Δ	Δ^2	Δ^3
0	-3			
		6		
1	3		2	
		8		6
2	11		8	
		16		6
3	27		14	
		30		6
4	57		20	
		50		
5	107			

Now the required cubic polynomial (polynomial of degree 3) is obtained from Newton's forward difference interpolation formula

$$f(x) \approx P_3(x) = f_0 + r\Delta f_0 + \frac{r(r-1)}{2!}\Delta^2 f_0 + \frac{r(r-1)(r-3+1)}{3!}\Delta^3 f_0,$$

where $r=(x - x_0)/h = (x - 0)/1 = x$, so that

$$f(x) \approx P_3(x) = -3 + x(6) + \frac{x(x-1)}{2!}(2) + \frac{x(x-1)(x-3+1)}{3!}(6)$$

or $f(x) = x^3 - 2x^2 + 7x - 3$

Example Using the Newton's forward difference interpolation formula evaluate $f(2.05)$ where $f(x) = \sqrt{x}$, using the values:

x	2.0	2.1	2.2	2.3	2.4
$\sqrt{x}$	1.414 214	1.449 138	1.483 240	1.516 575	1.549 193

The forward difference table is

x	$\sqrt{x}$	Δ	Δ^2	Δ^3	Δ^4
2.0	1.414 214				
		0.034 924			
2.1	1.449 138		-0.000 822		
		0.034 102		0.000055	
2.2	1.483 240		-0.000 767		–0.000 005
		0.033 335		0.000050	
2.3	1.516 575		-0.000 717		
		0.032 618			
2.4	1.549 193				

Here $r = \frac{x - x_0}{h}$ =(2.05–2.00)/0.1=0.5, so by substituting the values in Newton's formula (for 4 degree polynomial), we get

$$f(2.05) \approx P_4(2.05) = 1.414214 + (0.5)(0.034924) + \frac{(0.5)(0.5-1)}{2!}(-0.000822)$$

$$+ \frac{(0.5)(0.5-1)(0.5-2)}{3!}(0.000055)$$

$$+ \frac{(0.5(0.5-1)(0.5-2)(0.5-3)}{4!}(0.000005) = 1.431783.$$

Example Find the cubic polynomial which takes the following values; $f(1) = 24, f(3) = 120, f(5) = 336,$ and $f(7) = 720$. Hence, or otherwise, obtain the value of $f(8)$.

We form the difference table:

x	y	Δ	Δ^2	Δ^3
1	24			
		96		
3	120		120	
		216		48
5	336		168	
		384		
7	720			

Here $h = 2$ with $x_0 = 1$, we have $x = 1 + 2p$ or $r = (x-1)/2$. Substituting this value of r, we obtain

$$f(x) = 24 + \frac{x-1}{2}(96) + \frac{\left(\frac{x-1}{2}\right)\left(\frac{x-1}{2} - 1\right)}{2}(120)$$

$$+\frac{\left(\frac{x-1}{2}\right)\left(\frac{x-1}{2}-1\right)\left(\frac{x-1}{2}-2\right)}{6}(48) = x^3+6x^2+11x+6.$$

To determine $f(9)$, we put $x=9$ in the above and obtain $f(9)=1320.$

With $x_0=1,\ x_r=9,$ and $h=2,$ we have $r=\frac{x_r-x_0}{h}=\frac{9-1}{2}=4$. Hence

$$f(9)\approx p(9)=f_0+r\Delta f_0+\frac{r(r-1)}{2!}\Delta^2 f_0+\frac{r(r-1)\,(r-2)}{3!}\Delta^3 f_0$$

$$=24+4\times 96+\frac{4\times 3}{2}\times 120+\frac{4\times 3\times 2}{3\times 2}\times 48=1320$$

Example Using Newton's forward difference formula, find the sum

$$S_n=1^3+2^3+3^3+\ldots+n^3.$$

Solution

$$S_{n+1}=1^3+2^3+3^3+\ldots+n^3+(n+1)^3$$

and hence

$$S_{n+1}-S_n=(n+1)^3,$$

or

$$\Delta S_n=(n+1)^3.$$

it follows that

$$\Delta^2 S_n=\Delta S_{n+1}-\Delta S_n=(n+2)^3-(n+1)^3=3n^2+9n+7$$

$$\Delta^3 S_n=3(n+1)+9n+7-(3n^2+9n+7)=6n+12$$

$$\Delta^4 S_n=6(n+1)+12-(6n+12)=6$$

Since $\Delta^5 S_n=\Delta^6 S_n=\ldots=0, S_n$ is a fourth-degree polynomial in the variable n.

Also,

$$S_1=1,\qquad \Delta S_1=(1+1)^3=8,\qquad \Delta^2 S_1=3+9+7=19,$$

$$\Delta^3 S_1=6+12=18,\qquad \Delta^4 S_1=8.$$

formula (3) gives (with $f_0=S_1$ and $r-n-1$)

$$S_n=1+(n-1)(8)+\frac{(n-1)(n-2)}{2}(19)+\frac{(n-1)(n-2)(n-3)}{6}(18)$$

$$+\frac{(n-1)(n-2)(n-3)(n-4)}{24}(6)$$

$$=\frac{1}{4}n^4+\frac{1}{2}n^3+\frac{1}{4}n^2$$

$$=\left[\frac{n(n+1)}{2}\right]^2$$

Problem: The population of a country for various years in millions is provided. Estimate the population for the year 1898.

Year x:	1891	1901	1911	1921	1931
Population y:	46	66	81	93	101

Solution: Here the interval of difference among the arguments *h=10.* Since 1898 is at the beginning of the table values, we use Newton's forward difference interpolation formula for finding the population of the year 1898.

The forward differences for the given values are as shown here.

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
1891	46				
		$\Delta y_0 = 20$			
1901	66		$\Delta^2 y_0 = -5$		
		$\Delta y_1 = 15$		$\Delta^3 y_0 = 2$	
1911	81		$\Delta^2 y_1 = -3$		$\Delta^4 y_0 = -3$
		$\Delta y_2 = 12$		$\Delta^3 y_1 = -1$	
1921	93		$\Delta^2 y_2 = -4$		
		$\Delta y_3 = 8$			
1931	101				

Let x=1898. Newton's forward difference interpolation formula is,

$$f(x) = y_0 + (x-x_0)\frac{1}{h}[\Delta y_0] + (x-x_0)(x-x_1)\frac{1}{2!h^2}[\Delta^2 y_0]$$
$$+(x-x_0)(x-x_1)(x-x_2)\frac{1}{3!h^3}[\Delta^3 y_0]+....+$$
$$(x-x_0)(x-x_1).....(x-x_{n-1})\frac{1}{n!h^n}[\Delta^n y_0]$$

Now, substituting the values, we get,

$$f(1898) = 46 + (1898-1891)\frac{1}{10}[20] + (1898-1891)(1898-1901)\frac{1}{2!10^2}[-5]$$
$$+ (1898-1891)(1898-1901)(1898-1911)\frac{1}{3!10^3}[2] +$$
$$(1898-1891)(1898-1901)(1898-1911)(1898-1921)\frac{1}{4!10^4}[-3]$$

$$\Rightarrow f(1898) = 46 + 14 + \frac{21}{40} + \frac{91}{500} + \frac{18837}{40000} = 61.178$$

Example Values of x (in degrees) and $\sin x$ are given in the following table:

x (in degrees)	$\sin x$
15	0.2588190
20	0.3420201
25	0.4226183
30	0.5
35	0.5735764
40	0.6427876

Determine the value of $\sin 38^0$.

Solution

The difference table is

x	$\sin x$	Δ	Δ^2	Δ^3	Δ^4	Δ^5
15	0.2588190					
		0.0832011				
20	0.3420201		–0.0026029			
		0.0805982		–0.0006136		
25	0.4226183		–0.0032165		0.0000248	
		0.0773817		–0.0005888		0.0000041
30	0.5		–0.0038053		0.0000289	
		0.0735764		–0.0005599		
35	0.5735764		–0.0043652			
		0.0692112				
40	0.6427876					

As 38 is closer to $x_n = 40$ than $x_0 = 15$, we use Newton's backward difference formula with $x_n = 40$ and $x = 38$. This gives

$$r = \frac{x - x_n}{h} = \frac{38 - 40}{5} = -\frac{2}{5} = -0.4$$

Hence, using formula, we obtain

$$f(38) = 0.6427876 - 0.4(0.0692112) + \frac{-0.4(-0.4-1)}{2}(-0.0043652)$$
$$+ \frac{(-0.4)(-0.4+1)(-0.4+2)}{6}(-0.0005599)$$
$$+ \frac{(-0.4)(-0.4+1)(-0.4+2)(-0.4+3)}{24}(0.0000289)$$
$$+ \frac{(-0.4)(-0.4+1)(-0.4+2)(-0.4+3)(-0.4+4)}{120}(0.0000041)$$
$$= 0.6427876 - 0.02768448 + 0.00052382 + 0.00003583$$
$$-0.00000120$$
$$= 0.6156614$$

Example Find the missing term in the following table:

x	$y = f(x)$
0	1
1	3
2	9
3	–
4	81

Explain why the result differs from $3^3 = 27$?

Since four points are given, the given data can be approximated by a third degree polynomial in x. Hence $\Delta^4 f_0 = 0$. Substituting $\Delta = E - 1$ we get, $(E-1)^4 f_0 = 0$, which on simplification yields

$$E^4 f_0 - 4E^3 f_0 + 6E^2 f_0 - 4Ef_0 + f_0 = 0.$$

Since $E^r f_0 = f_r$ the above equation becomes

$$f_4 - 4f_3 + 6f_2 - 4f_1 + f_0 = 0$$

Substituting for f_0, f_1, f_2 and f_4 in the above, we obtain

$$f_3 = 31$$

By inspection it can be seen that the tabulated function is 3^x and the exact value of $f(3)$ is 27. The error is due to the fact that the exponential function 3^x is approximated by means of a polynomial in x of degree 3.

Example The table below gives the values of $\tan x$ for $0.10 \le x \le 0.30$

x	$y = \tan x$
0.10	0.1003
0.15	0.1511
0.20	0.2027
0.25	0.2553
0.30	0.3093

Find: (a) $\tan 0.12$ (b) $\tan 0.26$. (c) $\tan 0.40$ (d) $\tan 0.50$

The table difference is

x	$y = f(x)$	Δ	Δ^2	Δ^3	Δ^4
0.10	0.1003				
		0.0508			
0.15	0.1511		0.0008		
		0.0516		0.0002	
0.20	0.2027		0.0010		0.0002
		0.0526		0.0004	
0.25	0.2553		0.0014		
		0.0540			
0.30	0.3093				

a) To find $\tan(0.12)$, we have $r = 0.4$ Hence Newton's forward difference interpolation formula gives

$$\tan(0.12) = 0.1003 + 0.4(0.0508) + \frac{0.4(0.4-1)}{2}(0.0008)$$

$$+ \frac{0.4(0.4-1)(0.4-2)}{6}(0.0002)$$

$$+ \frac{0.4(0.4-1)(0.4-2)(0.4-3)}{24}(0.0002)$$

$$= 0.1205$$

b) To find $\tan(0.26)$, we use Newton's backward difference interpolation formula with

$$r = \frac{x - x_n}{n}$$

$$= \frac{0.26 - 03}{0.05}$$

$$= -0.8$$

which gives

$$\tan(0.26) = 0.3093 - 0.8(0.0540) + \frac{-0.8(-0.8+1)}{2}(0.0014)$$

$$+ \frac{-0.8(-0.8+1)(-0.8+2)}{6}(0.0004)$$

$$+ \frac{-0.8(-0.8+1)(-0.8+2)(-0.8+3}{24}(0.0002) \quad = 0.2662$$

Proceeding as in the case (i) above, we obtain

(c) $\tan 0.40 = 0.4241,$ and

(d) $\tan 0.50 = 0.5543$

The actual values, correct to four decimal places, of tan (0.12), tan(0.26) are respectively 0.1206 and 0.2660. Comparison of the computed and actual values shows that in the first two cases (i.e., of interpolation) the results obtained are fairly accurate whereas in the last-two cases (i.e., of extrapolation) the errors are quite considerable. The example therefore demonstrates the important results that if a tabulated function is other than a polynomial, then extrapolation very far from the table limits would be dangerous-although interpolation can be carried out very accurately.

Exercises

1. Using the difference table in exercise 1, compute cos0.75 by Newton's forward difference interpolating formula with $n = 1, 2, 3, 4$ and compare with the 5D-value 0.731 69.

2. Using the difference table in exercise 1, compute cos0.28 by Newton's forward difference interpolating formula with $n = 1, 2, 3, 4$ and compare with the 5D-value

3. Using the values given in the table, find cos0.28 (in radian measure) by linear interpolation and by quadratic interpolation and compare the results with the value 0.961 06 (exact to 5D).

x	$f(x)=\cos x$	First difference	Second difference
0.0	1.000 00		
		-0.019 93	
0.2	0.980 07		-0.03908
		-0.059 01	
0.4	0.921 06		-0.03671
		-0.095 72	
0.6	0.825 34		-0.03291
		-0.128 63	
0.8	0.696 71		-0.02778
		-0.156 41	
1.0	0.540 30		

4. Find Lagrangian interpolation polynomial for the function f having $f(4)=1,\ f(6)=3,\ f(8)=8,\ f(10)=16$. Also calculate $f(7)$.

5. The sales in a particular shop for the last ten years is given in the table:

Year	1996	1998	2000	2002	2004
Sales (in lakhs)	40	43	48	52	57

Estimate the sales for the year 2001 using Newton's backward difference interpolating formula.

6. Find $f(3)$, using Lagrangian interpolation formula for the function f having $f(1)=2,\ f(2)=11,\ f(4)=77$.

7. Find the cubic polynomial which takes the following values:

x	0	1	2	3
$f(x)$		1	2	1	10

8. Compute sin0.3 and sin0.5 by Everett formula and the following table.

	sin$x$	δ^2
0.2	0.198 67	-0.007 92
0.4	0.389 42	-0.015 53
.6	0.564 64	-0.022 50

9. The following table gives the distances in nautical miles of the visible horizon for the given heights in feet above the earth's surface:

x =height :	100	150	200	250	300	350	400
y = distance :	10.63	13.03	15.04	16.81	18.42	19.90	21.27

Find the value of y when x = 218 ft (*Ans*: 15.699)

10. Using the same data as in exercise 9, find the value of y when x = 410ft.

UNIT 2

FIXED POINT ITERATION METHOD

Nature of numerical problems

Solving mathematical equations is an important requirement for various branches of science. The field of numerical analysis explores the techniques that give approximate solutions to such problems with the desired accuracy.

Computer based solutions

The major steps involved to solve a given problem using a computer are:

1. Modeling: Setting up a mathematical model, i.e., formulating the problem in mathematical terms, taking into account the type of computer one wants to use.
2. Choosing an appropriate numerical method (algorithm) together with a preliminary error analysis (estimation of error, determination of steps, size etc.)
3. Programming, usually starting with a flowchart showing a block diagram of the procedures to be performed by the computer and then writing, say, a C++ program.
4. Operation or computer execution.
5. Interpretation of results, which may include decisions to rerun if further data are needed.

Errors

Numerically computed solutions are subject to certain errors. Mainly there are three types of errors. They are inherent errors, truncation errors and errors due to rounding.

1. ***Inherent errors or experimental errors*** arise due to the assumptions made in the mathematical modeling of problem. It can also arise when the data is obtained from certain physical measurements of the parameters of the problem. i.e., errors arising from measurements.
2. ***Truncation errors*** are those errors corresponding to the fact that a finite (or infinite) sequence of computational steps necessary to produce an exact result is "truncated" prematurely after a certain number of steps.
3. ***Round of errors*** are errors arising from the process of rounding off during computation. These are also called *chopping*, i.e. discarding all decimals from some decimals on.

Error in Numerical Computation

Due to errors that we have just discussed, it can be seen that our numerical result is an approximate value of the (sometimes unknown) exact result, except for the rare case where the exact answer is sufficiently simple rational number.

If $\tilde{a}$ is an approximate value of a quantity whose exact value is a, then the difference $\varepsilon = \tilde{a} - a$ is called the absolute error of $\tilde{a}$ or, briefly, the error of $\tilde{a}$. Hence, $\tilde{a} = a + \varepsilon$, i.e.

Approximate value = True value + Error.

For example, if $\tilde{a} = 10.52$ is an approximation to $a = 10.5$, then the error is $\varepsilon = 0.02$. The relative error, ε_r, of $\tilde{a}$ is defined by

$$|\varepsilon_r| = \frac{|\varepsilon|}{|a|} = \frac{|\text{Error}|}{|\text{Truevalue}|}$$

For example, consider the value of $\sqrt{2}(=1.414213...)$ up to four decimal places, then

$$\sqrt{2} = 1.4142 + \text{Error}\ .$$

$$|\text{Error}| = |1.4142 - 1.41421| = .00001,$$

taking 1.41421 as true or exact value. Hence, the relative error is

$$\varepsilon_r = \frac{0.00001}{1.4142}\ .$$

We note that

$\varepsilon_r \approx \dfrac{\varepsilon}{\tilde{a}}$ if $|\varepsilon|$ is much less than $|\tilde{a}|$.

We may also introduce the quantity $\gamma = a - \tilde{a} = -\varepsilon$ and call it the correction, thus, $a = \tilde{a} + \gamma$, i.e.

True value = Approximate value + Correction.

Error bound for $\tilde{a}$ is a number β such that $|\ \tilde{a} - a\ | \le \beta$ i.e., $|\varepsilon| \le \beta$.

Number representations

Integer representation

Floating point representation

Most digital computers have two ways of representing numbers, called **fixed point** and **floating point**. In a fixed point system the numbers are represented by a fixed number of decimal places e.g. 62.358, 0.013, 1.000.

In a floating point system the numbers are represented with a fixed number of significant digits, for example

0.6238×10^{3} 0.1714×10^{-13} -0.2000×10^{1}

also written as 0.6238 E03 0.1714 E –13 –0.2000 E01

or more simply 0.6238 +03 0.1714 –13 –0.2000 +01

Significant digits

Significant digit of a number c is any given digit of c, except possibly for zeros to the left of the first nonzero digit that serve only to fix the position of the decimal point. (Thus, any other zero is a significant digit of c). For example, each of the number 1360, 1.360, 0.01360 has 4 significant digits.

Round off rule to discard the k + 1th and all subsequent decimals

(*a*) ***Rounding down*** If the number at $(k + 1)^{th}$ decimal to be discarded is less than half a unit in the k^{th} place, leave the k^{th} decimal unchanged. For example, rounding of 8.43 to 1 decimal gives 8.4 and rounding of 6.281 to 2 decimal places gives 6.28.

(*b*) ***Rounding up*** If the number at $(k + 1)^{th}$ decimal to be discarded is greater than half a unit in the k^{th} place, add 1 to the k^{th} decimal. For example, rounding of 8.48 to 1 decimal gives 8.5 and rounding of 6.277 to 2 decimals gives 6.28.

(*c*) If it is exactly half a unit, round off to the nearest even decimal. For example, rounding off 8.45 and 8.55 to 1 decimal gives 8.4 and 8.6 respectively. Rounding off 6.265 and 6.275 to 2 decimals gives 6.26 and 6.28 respectively.

Example Find the roots of the following equations using 4 significant figures in the calculation.

(*a*) $x^2 - 4x + 2 = 0$ and (*b*) $x^2 - 40x + 2 = 0$.

Solution

A formula for the roots x_1, x_2 of a quadratic equation $ax^2 + bx + c = 0$ is

(*i*) $x_1 = \frac{1}{2a}(-b + \sqrt{b^2 - 4ac})$ and $x_2 = \frac{1}{2a}(-b - \sqrt{b^2 - 4ac})$.

Furthermore, since $x_1x_2 = c/a$, another formula for these roots is

(*ii*) $x_1 = \frac{1}{2a}(-b + \sqrt{b^2 - 4ac})$, and $x_2 = \frac{c}{ax_1}$

For the equation in (*a*), formula (*i*) gives,

$x_1 = 2 + \sqrt{2} = 2 + 1.414 = 3.414,$

$x_2 = 2 - \sqrt{2} = 2 - 1.414 = 0.586$

and formula (*ii*) gives,

$x_1 = 2 + \sqrt{2} = 2 + 1.414 = 3.414,$

$x_2 = 2.000/3.414 = 0.5858.$

For the equation in (*b*), formula (*i*) gives,

$x_1 = 20 + \sqrt{398} = 20 + 19.95 = 39.95,$

$x_2 = 20 - \sqrt{398} = 20 - 19.95 = 0.05$

and formula (*ii*) gives,

$x_1 = 20 + \sqrt{398} = 20 + 19.95 = 39.95,$

$x_2 = 20.000/39.95 = 0.05006.$

Example Convert the decimal number (which is in the base 10) 81.5 to its binary form (of base 2).

Solution Note that $(81.5)_{10} = 8 \cdot 10^1 + 1 \cdot 10^0 + 5 \cdot 10^{-1}$

Now $81.5 = 64 + 16 + 1 + 0.5 = 2^6 + 2^4 + 2^0 + 2^{-1} = (1010001.1)_2$.

		Remainder		Product	Integer part	
2	81		0.5 × 2	1.0	1	↓
2	40	1				
2	20	0				
2	10	0				
2	5	0				
2	2	1				
2	1	0				
	0	1				

Example Convert the binary number 1010.101 to its decimal form.

Solution

$$(1010.101)_2 = 1\cdot 2^3 + 1\cdot 2^1 + 1\cdot 2^{-1} + 1\cdot 2^{-3}$$

$$= 8 + 2 + 0.5 + 0.125 = (10.625)_{10}$$

Numerical Iteration Method

A **numerical iteration method** or simply **iteration method** is a mathematical procedure that generates a sequence of improving approximate solutions for a class of problems. A specific way of implementation of an iteration method, including the termination criteria, is called an algorithm of the iteration method. In the problems of finding the solution of an equation an iteration method uses an initial guess to generate successive approximations to the solution.

Since the iteration methods involve repetition of the same process many times, computers can act well for finding solutions of equation numerically. Some of the iteration methods for finding solution of equations involves (1) Bisection method, (2) Method of false position (Regula-falsi Method), (3) Newton-Raphson method.

A numerical method to solve equations may be a long process in some cases. If the method leads to value close to the exact solution, then we say that the method is convergent. Otherwise, the method is said to be divergent.

Solution of Algebraic and Transcendental Equations

One of the most common problem encountered in engineering analysis is that given a function f (x), find the values of x for which f(x) = 0. The solution (values of x) are known

as the **roots** of the equation f(x) = 0, or the **zeroes** of the function f (x). The roots of equations may be real or complex.

In general, an equation may have any number of (real) roots, or no roots at all. For example, sin x – x = 0 has a single root, namely, x = 0, whereas tan x – x = 0 has infinite number of roots (x = 0, ± 4.493, ± 7.725, …).

Algebraic and Transcendental Equations

$f(x) = 0$ is called an **algebraic equation** if the corresponding $f(x)$ is a polynomial. An example is $7x^2 + x - 8 = 0$. $f(x) = 0$ is called **transcendental equation** if the $f(x)$ contains trigonometric, or exponential or logarithmic functions. Examples of transcendental equations are sin x – x = 0, $\tan x - x = 0$ and $7x^3 + \log(3x - 6) + 3e^x \cos x + \tan x = 0$.

There are two types of methods available to find the roots of algebraic and transcendental equations of the form f (x) = 0.

1. Direct Methods: Direct methods give the exact value of the roots in a finite number of steps. We assume here that there are no round off errors. Direct methods determine all the roots at the same time.

2. Indirect or Iterative Methods: Indirect or iterative methods are based on the concept of successive approximations. The general procedure is to start with one or more initial approximation to the root and obtain a sequence of iterates x_k which in the limit converges to the actual or true solution to the root. Indirect or iterative methods determine one or two roots at a time. The indirect or iterative methods are further divided into two categories: bracketing and open methods. The bracketing methods require the limits between which the root lies, whereas the open methods require the initial estimation of the solution. Bisection and False position methods are two known examples of the bracketing methods. Among the open methods, the Newton-Raphson is most commonly used. The most popular method for solving a non-linear equation is the

Newton-Raphson method and this method has a high rate of convergence to a solution.

In this chapter and in the coming chapters, we present the following indirect or iterative methods with illustrative examples:

1. Fixed Point Iteration Method

2. Bisection Method

3. Method of False Position (Regula Falsi Method)

4. Newton-Raphson Method (Newton's method)

Fixed Point Iteration Method

Consider

$$f(x) = 0 \qquad \ldots (1)$$

Transform (1) to the form,

$$x = \phi(x). \qquad \ldots(2)$$

Take an arbitrary x_0 and then compute a sequence x_1, x_2, x_3, . . . recursively from a relation of the form

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

A **solution** of (2) is called **fixed point** of ϕ. To a given equation (1) there may correspond several equations (2) and the behaviour, especially, as regards speed of convergence of iterative sequences x_0, x_1, x_2, x_3, . . . may differ accordingly.

Example Solve $f(x) = x^2 - 3x + 1 = 0$, by fixed point iteration method.

Solution

Write the given equation as

$x^2 = 3x - 1$ or $x = 3 - 1/x$.

Choose $\phi(x) = 3 - \frac{1}{x}$. Then $\phi'(x) = \frac{1}{x^2}$ and $|\phi'(x)| < 1$ on the interval (1, 2).

Hence the iteration method can be applied to the Eq. (3).

The iterative formula is given by

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

Starting with, $x_0 = 1$, we obtain the sequence

x_0=1.000, x_1 =2.000, x_2 =2.500, x_3 = 2.600, x_4 =2.615, . . .

Question : Under what assumptions on ϕ and x_0, does Algorithm 1 converge ? When does the sequence (x_n) obtained from the iterative process (3) converge ?

We answer this in the following theorem, that is a sufficient condition for convergence of iteration process

Theorem Let $x = ¶$ be a root of $f(x) = 0$ and let I be an interval containing the point $x = ¶$. Let $¤(x)$ be continuous in I, where $¤(x)$ is defined by the equation $x = ¤(x)$ which is equivalent to $f(x) = 0$. Then if $|¤'(x)| < 1$ for all x in I, the sequence of approximations $x_0, x_1, x_2, \cdots, x_n$ defined by

$$x_{n+1} = ¤(x_n) \quad (n = 0, 1, \ldots)$$

converges to the root ¶, provided that the initial approximation x_0 is chosen in I.

Example Find a real root of the equation $x^3 + x^2 - 1 = 0$ on the interval $[0, 1]$ with an accuracy of 10^{-4}.

To find this root, we rewrite the given equation in the form

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

Take

$$¤(x) = \frac{1}{\sqrt{x+1}}. \text{ Then } ¤(x) = -\frac{1}{2}\frac{1}{(x+1)^{\frac{3}{2}}}$$

$$\max_{[0,1]} |¤'(x)| = \left|\frac{1}{2\sqrt{8}}\right| = k = 0.17678 < 0.2.$$

Choose $¤(x) = 3 - \frac{1}{x}$. Then $¤'(x) = \frac{1}{x^2}$ and $|¤'(x) < 1$ on the interval $(1, 2)$.

Hence the iteration method gives:

n	x_n	$\sqrt{x_n + 1}$	$x_{n+1} = 1/\sqrt{x_n + 1}$
0	0.75	1.3228756	0.7559289
1	0.7559289	1.3251146	0.7546517
2	0.7546617	1.3246326	0.7549263

At this stage,

$$|x_{n+1} - x_n| = 0.7549263 - 0.7546517 = 0.0002746,$$

which is less than 0.0004. The iteration is therefore terminated and the root to the required accuracy is 0.7549.

Example Use the method of iteration to find a positive root, between 0 and 1, of the equation $xe^x = 1$.

Writing the equation in the form

$$x = e^{-x}$$

We find that $¤(x) = e^{-x}$ and so $¤'(x) = -e^{-x}$.

Hence $|¤'(x)| < 1$ for $x < 1$, which assures that the iterative process defined by the equation $x_{n+1} = ¤(x_n)$ will be convergent, when $x < 1$.

The iterative formula is

$$x_{n+1} = \frac{1}{e^{x_n}} \qquad (n = 0, 1, \dots)$$

Starting with $x_0 = 1$, we find that the successive iterates are given by

$$x_1 = 1/e = 0.3678794, \; x_2 = \frac{1}{ex_1} = 0.6922006,$$

$$x_3 = 0.5004735, \qquad x_4 = 0.6062435,$$

$$x_5 = 0.5453957, \qquad x_6 = 0.5796123,$$

We accept 6.5453957 as an approximate root.

Example Find the root of the equation $2x = \cos x + 3$ correct to three decimal places.

We rewrite the equation in the form

$$x = \frac{1}{2}(\cos x + 3)$$

so that

$$¤ = \frac{1}{2}(\cos x + 3),$$

and

$$|¤'(x)| = \left|\frac{\sin x}{2}\right| < 1.$$

Hence the iteration method can be applied to the eq. (3) and we start with $x_0 = ®/2$. The successive iterates are

$$x_1 = 1.5, \quad x_2 = 1.535, \quad x_3 = 1.518,$$
$$x_4 = 1.526, \quad x_5 = 1.522, \quad x_6 = 1.524,$$
$$x_7 = 1.523, \quad x_8 = 1.524.$$

We accept the solution as 1.524 correct to three decimal places.

Example Find a solution of $f(x) = x^3 + x - 1 = 0$, by fixed point iteration.

$x^3 + x - 1 = 0$ can be written as $x(x^2+1)=1$, or $x=\dfrac{1}{x^2+1}$.

Note that

$$|¤'(x)| = \frac{2|x|}{\left(1+x^2\right)^2} < 1 \text{ for any real } x,$$

so by the Theorem we can expect a solution for any real number x_0 as the starting point.

Choosing $x_0 = 1$, and undergoing calculations in the iterative formula

$$x_{n+1} = ¤(x_n) = \frac{1}{1+x_n^2} \quad (n = 0, 1, \ldots), \qquad \ldots(4)$$

we get the sequence

x_0=1.000 , x_1=0.500, x_2=0.800, x_3 =0.610 ,

x_4= 0.729, x_5=0.653, x_6=0.701, ...

and we choose 0.701 as an (approximate) solution to the given equation.

Example Solve the equation $x^3 = \sin x$. Considering various $¤(x)$, discuss the convergence of the solution.

How do the functions we considered for $¤(x)$ compare? Table shows the results of several

iterations using initial value $x_0 = 1$ and four different functions for $¤(x)$. Here x_n is the value of x

on the nth iteration .

Answer:

When $¤(x) = \sqrt[3]{\sin x}$, we have:

$x_1 = 0.94408924124306;$ $x_2 = 0.93215560685805$

$x_3 = 0.92944074461587$; $x_4 = 0.92881472066057$

When $¤(x) = \dfrac{\sin x}{x^2}$, we have:

$x_1 = 0.84147098480790;$ $x_2 = 1.05303224555943$

$x_3 = 0.78361086350974$; $x_4 = 1.14949345383611$

Referring to Theorem, we can say that for $¤(x)=\frac{\sin x}{x^2}$, the iteration doesn't converge.

When $¤(x)=x+\sin x-x^3$, we have:

$x_1 = 0.84147098480790;$ $x_2 = 0.99127188988250$

$x_3 = 0.85395152069647\ ;$ $x_4 = 0.98510419085185$

When $¤(x)=x-\frac{\sin x-x^3}{\cos x-3x^2}$, we have:

$x_1 = 0.93554939065467;$ $x_2 = 0.92989141894368$

$x_3 = 0.92886679103170;$ $x_4 = 0.92867234089417$

Example Give all possible transpositions to $x=¤(x)$, and solve $f(x)=x^3+4x^2-10=0.$

Possible Transpositions to $x=¤(x)$, are

$x=¤_1(x)=x-x^3-4x^2+10,$

$x=¤_2(x)=\sqrt{\frac{10}{x}-4x},$

$x=¤_3(x)=\frac{1}{2}\sqrt{10-x^3}$

$x=¤_4(x)=\sqrt{\frac{10}{4+x}}$

$x=¤_5(x)=x-\frac{x^3+4x^2-10}{3x^2+8x}$

For $x=¤_1(x)=x-x^3-4x^2+10,$ numerical results are:

$x_0=1.5;$ $x_2=-0.875$

$x_3=6.732;$ $x_4=-469.7;$

Hence doesn't converge.

For $x=¤_2(x)=\sqrt{\frac{10}{x}-4x},$ numerical results are:

$x_0=1.5;$ $x_2=0.8165$

$x_3=2.9969;$ $x_4=(-8.65)^{1/2};$

For $x = ¤_3(x) = \frac{1}{2}\sqrt{10 - x^3}$, numerical results are:

$x_0 = 1.5;$ $x_2 = 1.2869$
$x_3 = 1.4025;$ $x_4 = 1.3454$;

Exercises

Solve the following equations by iteration method:

- $\sin x = \frac{x+1}{x-1}$
- $3x - \cos x - 2 = 0$
- $x^3 + x + 1 = 0$
- $3x = 6 + \log_{10} x$
- $2x - \log_{10} x = 7$
- $2\sin x = x$
- $x^3 + x^2 = 100$
- $x^4 = x + 0.15$
- $x^3 - 5x + 3 = 0,$
- $x = \frac{1}{6}\left(x^3 + 3\right)$
- $x = \frac{1}{5}\left(x^3 + 3\right)$
- $x^3 = 2x^2 + 10x = 20$
- $\cos x = 3x - 1$
- $3x + \sin x = e^x$

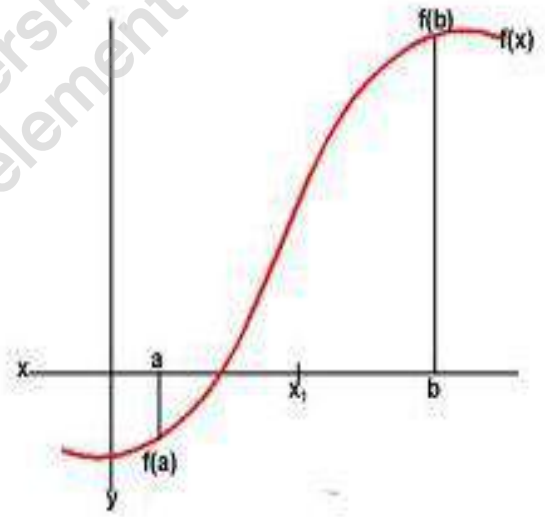
BISECTION AND REGULA FALSI METHODS

Bisection Method

The bisection method is one of the bracketing methods for finding roots of an equation. For a given a function $f(x)$, guess an interval which might contain a root and perform a number of iterations, where, in each iteration the interval containing the root is get halved.

The **bisection method** is based on the intermediate value theorem for continuous functions.

Intermediate value theorem for continuous functions: If f is a continuous function and $f(a)$ and $f(b)$ have opposite signs, then at least one root lies in between a and b. If the interval (a, b) is small enough, it is likely to contain a single root.



i.e., an interval [a, b] must contain a zero of a continuous function f if the product $f(a)f(b) < 0$. Geometrically, this means that if $f(a)f(b) < 0$, then the curve f has to cross the x-axis at some point in between a and b.

Algorithm : Bisection Method

Suppose we want to find the solution to the equation $f(x) = 0$, where f is continuous.

Given a function $f(x)$ continuous on an interval [a_0, b_0] and satisfying $f(a_0)f(b_0) < 0$.

For n = 0, 1, 2, … until termination do:

Compute $x_n = \frac{1}{2}(a_n + b_n)$.

If $f(x_n) = 0$, accept x_n as a solution and stop.

Else continue.

If $f(a_n)\,f(x_n) < 0$, a root lies in the interval (a_n, x_n).

Set $a_{n+1} = a_n,\ b_{n+1} = x_n$.

If $f(a_n)\,f(x_n) > 0,$ a root lies in the interval (x_n, b_n).

Set $a_{n+1} = x_n,\ b_{n+1} = b_n$.

Then $f(x) = 0$ for some x in $[a_{n+1}, b_{n+1}]$.

Test for termination.

Criterion for termination

A convenient criterion is to compute the percentage error $£_r$ defined by

$$£_r = \left| \frac{x_r' - x_r}{x_r'} \right| \times 100\%.$$

where x_r' is the new value of x_r. The computations can be terminated when $£_r$ becomes less than a prescribed tolerance, say $£_p$. In addition, the maximum number of iterations may also be specified in advance.

Some other termination criteria are as follows:

- Termination after N steps (N given, fixed)
- Termination if $| x_{n+1} - x_n | \leq \varepsilon$ ($\varepsilon > 0$ given)
- Termination if $|f(x_n)| \leq \alpha$ ($\alpha > 0$ given).

In this chapter our criterion for termination is terminate the iteration process after some finite steps. However, we note that this is generally not advisable, as the steps may not be sufficient to get an approximate solution.

Example Solve $x^3 - 9x + 1 = 0$ for the root between $x = 2$ and $x = 4$, by bisection method.

Given $f(x) = x^3 - 9x + 1$. Now $f(2) = -9,\ f(4) = 29$ so that $f(2)\,f(4) < 0$ and hence a root lies between 2 and 4.

Set $a_0 = 2$ and $b_0 = 4$. Then

$$x_0 = \frac{(a_0 + b_0)}{2} = \frac{2+4}{2} = 3 \quad \text{and} \quad f(x_0) = f(3) = 1.$$

Since $f(2)f(3) < 0$, a root lies between 2 and 3, hence we set $a_1 = a_0 = 2$ and $b_1 = x_0 = 3$. Then

$$x_1 = \frac{(a_1 + b_1)}{2} = \frac{2+3}{2} = 2.5 \quad \text{and} \quad f(x_1) = f(2.5) = -5.875$$

Since $f(2)f(2.5) > 0$, a root lies between 2.5 and 3, hence we set $a_2 = x_1 = 2.5$ and $b_2 = b_1 = 3$.

Then $x_2 = \frac{(a_2 + b_2)}{2} = \frac{2.5+3}{2} = 2.75$ and $f(x_2) = f(2.75) = -2.9531.$

The steps are illustrated in the following table.

n	x_n	$f(x_n)$
0	3	1.0000
1	2.5	−5.875
2	2.75	− 2.9531
3	2.875	− 1.1113
4	2.9375	− 0.0901

Example Find a real root of the equation $f(x) = x^3 - x - 1 = 0.$

Since $f(1)$ is negative and $f(2)$ positive, a root lies between 1 and 2 and therefore we take $x_0 = 3/2 = 1.5.$ Then

$f(x_0) = \frac{27}{8} - \frac{3}{2} = \frac{15}{8}$ is positive and hence $f(1)\ f(1.5) < 0$ and Hence the root lies between 1 and 1.5 and we obtain

$$x_1 = \frac{1+1.5}{2} = 1.25$$

$f(x_1) = -19/64,$ which is negative and hence $f(1)\ f(1.25) > 0$ and hence a root lies between 1.25 and 1.5. Also,

$$x_2 = \frac{1.25+1.5}{2} = 1.375$$

The procedure is repeated and the successive approximations are

$x_3 = 1.3125,\quad x_4 = 1.34375,\quad x_5 = 1.328125,$ etc.

Example Find a positive root of the equation $xe^x = 1,$ which lies between 0 and 1.

Let $f(x) = xe^x - 1.$ Since $f(0) = -1$ and $f(1) = 1.718,$ it follows that a root lies between 0 and 1. Thus,

$$x_0 = \frac{0+1}{2} = 0.5.$$

Since $f(0.5)$ is negative, it follows that a root lies between 0.5 and 1. Hence the new root is 0.75, i.e.,

$$x_1 = \frac{.5+1}{2} = 0.75.$$

Since $f(x_1)$ is positive, a root lies between 0.5 and 0.75 . Hence

$$x_2 = \frac{.5+.75}{2} = 0.625$$

Since $f(x_2)$ is positive, a root lies between 0.5 and 0.625. Hence

$$x_3 = \frac{.5+.625}{2} = 0.5625.$$

We accept 0.5625 as an approximate root.

Merits of bisection method

a) The iteration using bisection method always produces a root, since the method brackets the root between two values.

b) As iterations are conducted, the length of the interval gets halved. So one can guarantee the convergence in case of the solution of the equation.

c) the Bisection Method is simple to program in a computer.

Demerits of bisection method

a) The convergence of the bisection method is slow as it is simply based on halving the interval.

b) Bisection method cannot be applied over an interval where there is a discontinuity.

c) Bisection method cannot be applied over an interval where the function takes always values of the same sign.

d) The method fails to determine complex roots.

e) If one of the initial guesses a_0 or b_0 is closer to the exact solution, it will take larger number of iterations to reach the root.

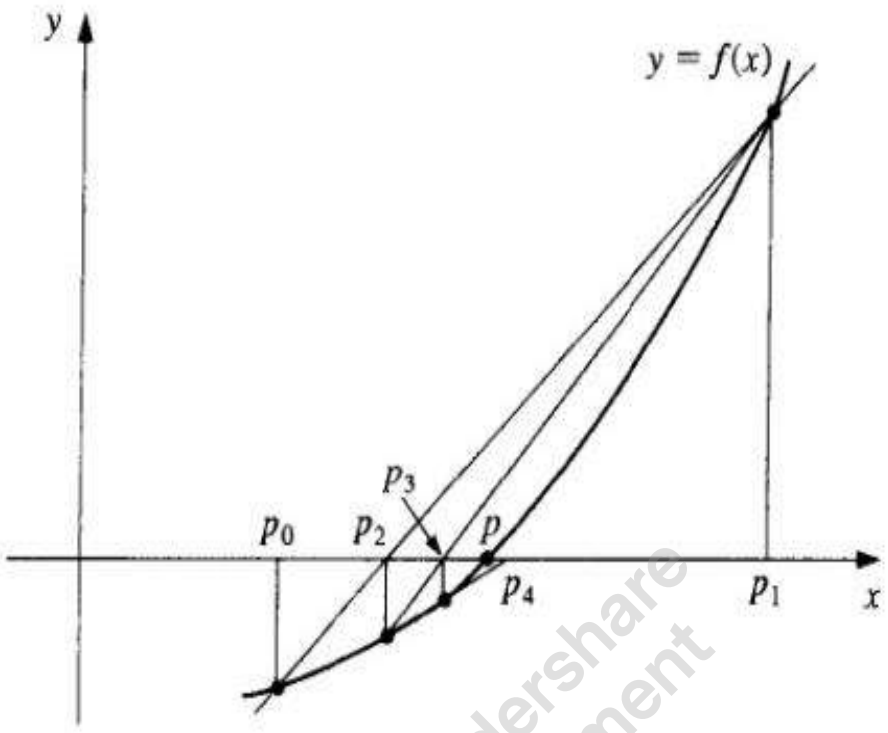
Exercises

Find a real root of the following equations by bisection method.

1. $3x = \sqrt{1+\sin x}$
2. $x^3 + 1.2x^2 - = 4x + 48$
3. $e^x = 3x$
4. $x^3 - 4x - 9 = 0$
5. $x^3 + 3x - 1 = 0$
6. $3x = \cos x + 1$
7. $x^3 + x^2 - 1 = 0$
8. $2x = 3 + \cos x$
9. $x^4 = 3$
10. $x^3 - 5x = 6$
11. $\cos x = \sqrt{x}$
12. $x^3 - x^2 - x - 3 = 0,$
13. $x^4 = x + 0.15$ near $x = 0$.

Regula Falsi method or Method of False Position

This method is also based on the intermediate value theorem. In this method also, as in bisection method, we choose two points a_n and b_n such that $f(a_n)$ and $f(b_n)$ are of opposite signs (i.e., $f(a_n) f(b_n) < 0$). Then, intermediate value theorem suggests that a zero of f lies in between a_n and b_n, if f is a continuous function.



Algorithm: Given a function $f(x)$ continuous on an interval $[a_0, b_0]$ and satisfying $f(a_0)f(b_0) < 0$.

For n = 0, 1, 2, … until termination do:

Compute

$$x_n = \frac{\begin{vmatrix} a_n & b_n \\ f(a_n) & f(b_n) \end{vmatrix}}{f(b_n) - f(a_n)}.$$

If $f(x_n) = 0$, accept x_n as a solution and stop.

Else continue.

If $f(a_n)f(x_n) < 0$, set $a_{n+1} = a_n, b_{n+1} = x_n$. Else set $a_{n+1} = x_n, b_{n+1} = b_n$.

Then $f(x) = 0$ for some x in $[a_{n+1}, b_{n+1}]$.

Example Using regula-falsi method, find a real root of the equation,

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

Here note that $f(0)$ = -1 and $f(0)=-1$. Hence $f(0)f(1)<0$, so by intermediate value theorem a root lies in between 0 and 1. We search for that root by regula falsi method and we will get an approximate root.

Set a_0 = 0 and b_0 = 1. Then

$$x_0=\frac{\begin{vmatrix} a_0 & b_0 \\ f(a_0) & f(b_0)\end{vmatrix}}{f(b_0)-f(a_0)}=\frac{\begin{vmatrix} 0 & 1 \\ -1 & 1\end{vmatrix}}{1-(-1)}=0.5$$

and $f(x_0)=f(0.5)=-0.375$.

Since $f(0)f(0.5)>0$, a root lies between 0.5 and 1. Set $a_1=x_0=0.5$ and $b_1=b_0=1$.

Then

$$x_1=\frac{\begin{vmatrix} a_1 & b_1 \\ f(a_1) & f(b_1)\end{vmatrix}}{f(b_1)-f(a_1)}=\frac{\begin{vmatrix} 0.5 & 1 \\ -0.375 & 1\end{vmatrix}}{1-(-0.375)}=0.6364$$

and $f(x_1)=f(0.6364)=-0.1058.$

Since $f(0.6364)f(x_1)>0$, a root lies between x_1 and 1 and hence we set $a_2=x_1=0.6364$ and $b_2=b_1=1$. Then

$$x_2=\frac{\begin{vmatrix} a_2 & b_2 \\ f(a_2) & f(b_2)\end{vmatrix}}{f(b_2)-f(a_2)}=\frac{\begin{vmatrix} 0.6364 & 1 \\ -0.1058 & 1\end{vmatrix}}{1-(-0.1058)}=0.6712$$

and $f(x_2)=f(0.6712)=-0.0264$

Since $f(0.6712)f(0.6364)>0,$ a root lies between x_2 and 1, and hence we set $a_3=x_2=0.6364$ and $b_3=b_1=1.$

Then $$x_3=\frac{\begin{vmatrix} a_3 & b_3 \\ f(a_3) & f(b_3)\end{vmatrix}}{f(b_3)-f(a_3)}=\frac{\begin{vmatrix} 0.6712 & 1 \\ -0.0264 & 1\end{vmatrix}}{1-(-0.0264)}=0.6796$$

and $f(x_3)=f(0.6796)=-0.0063\approx 0$.

Since $f(0.6796) \approx 0.0000$ we accept 0.6796 as an (approximate) solution of $x^3 - x - 1 = 0$.

Example Given that the equation $x^{2.2} = 69$ has a root between 5 and 8. Use the method of regula-falsi to determine it.

Let $f(x) = x^{2.2} - 69$. We find

$$f(5) = -3450675846 \text{ and } f(8) = -28.00586026.$$

$$x_1 = \frac{\begin{vmatrix} 5 & 8 \\ f(5) & f(8) \end{vmatrix}}{f(8) - f(5)} = \frac{5(28.00586026) - 8(-34.50675846)}{28.00586026 + 34.50675846)} = 6.655990062\,.$$

Now, $f(x_1) = -4.275625415$ and therefore, $f(5)\, f(x_1) > 0$ and hence the root lies between 6.655990062 and 8.0. Proceeding similarly,

$$x_2 = 6.83400179, \quad x_3 = 6.850669653,$$

The correct root is $x_3 = 6.8523651\cdots$, so that x_3 is correct to these significant figures. We accept 6.850669653 as an approximate root.

Theoretical Exercises with Answers:

1. What is the difference between algebraic and transcendental equations?

   Ans: An equation $f(x) = 0$ is called an algebraic equation if the corresponding $f(x)$ is a polynomial, while, $f(x) = 0$ is called transcendental equation if the $f(x)$ contains trigonometric, or exponential or logarithmic functions.

2. Why we are using numerical iterative methods for solving equations?

   Ans: As analytic solutions are often either too tiresome or simply do not exist, we need to find an approximate method of solution. This is where numerical analysis comes into the picture.

3. Based on which principle, the bisection and regula-falsi method is developed?

   Ans: These methods are based on the *intermediate value theorem for continuous functions*: stated as , "If f is a continuous function and $f(a)$ and $f(b)$ have opposite signs, then at least one root lies in between a and b. If the interval (a, b) is small enough, it is likely to contain a single root. "

4. What are the advantages and disadvantages of the bracketing methods like bisection and regula-falsi?

Ans: (i) The bisection and regula-falsi method is always convergent. Since the method brackets the root, the method is guaranteed to converge. The main disadvantage is, if it is not possible to bracket the roots, the methods cannot applicable. For example, if $f(x)$ is such that it always takes the values with same sign, say, always positive or always negative, then we cannot work with bisection method. Some examples of such functions are

- $f(x) = x^2$ which take only non-negative values and
- $f(x) = -x^2$, which take only non-positive values.

Exercises

Find a real root of the following equations by false position method:

1. $x^3 - 5x = 6$
2. $4x = e^x$
3. $x \log_{10} x = 1.2$
4. $\tan x + \tanh x = 0$
5. $e^{-x} = \sin x$
6. $x^3 - 5x - 7 = 0$
7. $x^3 + 2x^2 + 10x - 20 = 0$
8. $2x - \log_{10} x = 7$
9. $xe^x = \cos x$
10. $x^3 - 5x + 1 = 0$
11. $e^x = 3x$
12. $x^2 - \log_e x = 12$
13. $3x - \cos x = 1$
14. $2x - 3\sin x = 5$
15. $2x = \cos x + 3$
16. $xe^x = 3$
17. $\cos x = \sqrt{x}$
18. $x^3 - 5x + 3 = 0$

Ramanujan's Method

We need the following Theorem:

Binomial Theorem: If n is any rational number and $|x| < 1$, then

$$(1+x)^n = 1 + \frac{n}{1}x + \frac{n(n-1)}{1\cdot 2}x^2 + \ldots + \frac{n(n-1)\ldots(n-(r-1))}{1\cdot 2\cdot \ldots \cdot r}x^r + \ldots$$

In particular,

$$(1+x)^{-1} = 1 - x + x^2 - x^3 + \ldots + (-1)^n x^n + \ldots$$

and $$(1-x)^{-1} = 1 + x + x^2 + x^3 + \ldots + x^n + \ldots$$

Indian Mathematician Srinivasa Ramanujan (1887-1920) described an iterative method which can be used to determine the smallest root of the equation

$$f(x) = 0,$$

where $f(x)$ is of the form

$$f(x) = 1 - (a_1 x + a_2 x^2 + a_3 x^2 + a_4 x^4 + \cdots).$$

For smaller values of x, we can write

$$[1 - (a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \cdots)]^{-1} = b_1 + b_2 x + b_3 x^2 + \cdots$$

Expanding the left-hand side using binomial theorem , we obtain

$$1 + (a_1 x + a_2 x^2 + a_3 x^3 + \cdots) + (a_1 x + a_2 x^2 + a_3 x^3 + \cdots)^2 + \cdots$$

$$= b_1 + b_2 x + b_3 x^2 + \cdots$$

Comparing the coefficients of like powers of x on both sides of we obtain

$$\left.\begin{aligned} & b_1 = 1, \\ & b_2 = a_1 = a_1 b_1, \\ & b_3 = a_1^2 + a_2 = a_1 b_2 + a_2 b_1, \\ & \vdots \\ & b_n = a_1 b_{n-1} + a_2 b_{n-2} + \cdots + a_{n-1} b_1 \quad n = 2, 3, \cdots \end{aligned}\right\}$$

Then b_n / b_{n+1} approach a root of the equation $f(x) = 0$.

Example Find the smallest root of the equation

$$f(x) = x^3 - 6x^2 + 11x - 6 = 0.$$

Solution

The given equation can be written as $f(x)$

$$f(x) = 1 - \frac{1}{6}(11x - 6x^2 + x^3)$$

Comparing,

$$a_1 = \frac{11}{6},\quad a_2 = -1,\qquad a_3 = \frac{1}{6},\qquad a_4 = a_5 = \cdots = 0$$

To apply Ramanujan's method we write

$$1 - \left(\frac{11x - 6x^2 + x^3}{6}\right)^{-1} = b_1 + b_2 x + b_3 x^2 + \cdots$$

Hence,

$$b_1 = 1;$$

$$b_2 = a_1 = \frac{11}{6};$$

$$b_3 = a_1 b_2 + a_2 b_1 = \frac{121}{36} - 1 = \frac{85}{36};$$

$$b_4 = a_1 b_3 + a_2 b_2 + a_3 b_1 = \frac{575}{216};$$

$$b_5 = a_1 b_4 + a_2 b_3 + a_3 b_2 + a_4 b_1 = \frac{3661}{1296};$$

$$b_6 = a_1 b_5 + a_2 b_4 + a_3 b_3 + a_4 b_2 + a_5 b_1 = \frac{22631}{7776};$$

Therefore,

$$\frac{b_1}{b_2} = \frac{6}{11} = 0.54545\,; \qquad \frac{b_2}{b_3} = \frac{66}{85} = 0.7764705$$

$$\frac{b_3}{b_4} = \frac{102}{115} = 0.8869565\,; \qquad \frac{b_4}{b_5} = \frac{3450}{3661} = 0.9423654$$

$$\frac{b_5}{b_6} = \frac{3138}{3233} = 0.9706155$$

By inspection, a root of the given equation is unity and it can be seen that the successive convergents $\frac{b_n}{b_{n+1}}$ approach this root.

Example Find a root of the equation $xe^x = 1$.

Let $xe^x = 1$

Recall $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$

Hence,

$$f(x) = 1 - \left(x + x^2 + \frac{x^3}{2} + \frac{x^4}{6} + \frac{x^5}{24} + \cdots \right) = 0$$

$$a_1 = 1, \quad a_2 = 1, \quad a_3 = \frac{1}{2}, \quad a_4 = \frac{1}{6}, \quad a_5 = \frac{1}{24}, \cdots$$

We then have

$b_1 = 1;$

$b_2 = a_2 = 1;$

$b_3 = a_1 b_2 + a_2 b_1 = 1 + 1 = 2;$

$b_4 = a_1 b_3 + a_2 b_2 + a_3 b_1 = 2 + 1 + \frac{1}{2} = \frac{7}{2};$

$b_5 = a_1 b_4 + a_2 b_3 + a_3 b_2 + a_4 b_1 = \frac{7}{2} + 2 + \frac{1}{2} + \frac{1}{6} = \frac{37}{6};$

$b_6 = a_1 b_5 + a_2 b_4 + a_3 b_3 + a_4 b_2 + a_5 b_1 = \frac{37}{6}; + \frac{7}{2} + 1 + \frac{1}{6} + \frac{1}{24} = \frac{261}{24};$

Therefore,

$$\frac{b_2}{b_3} = \frac{1}{2} = 0.5; \qquad \frac{b_3}{b_4} = \frac{4}{7} = 0.5714;$$

$$\frac{b_4}{b_5} = \frac{21}{37} = 0.56756756; \qquad \frac{b_5}{b_6} = \frac{148}{261} = 0.56704980.$$

Example Using Ramanujan's method, find a real root of the equation

$$1 - x + \frac{x^2}{(2!)^2} - \frac{x^3}{(3!)^2} + \frac{x^4}{(4!)^2} - \cdots = 0.$$

Solution

Let $$f(x) = 1 - \left[x - \frac{x^2}{(2!)^2} + \frac{x^3}{(3!)^2} - \frac{x^4}{(4!)^2} + \cdots \right] = 0.$$

Here

$$a_1 = 1, \quad a_2 = -\frac{1}{(2!)^2}, \quad a_3 = \frac{1}{(3!)^2}, \quad a_4 = -\frac{1}{(4!)^2},$$

$$a_5 = \frac{1}{(5!)^2}, \qquad a_6 = -\frac{1}{(6!)^2}, \cdots$$

Writing

$$\left\{1 - \left[x - \frac{x^2}{(2!)} + \frac{x^3}{(3!)^2} - \frac{x^4}{(4!)^2} + \cdots\right]\right\}^{-1} = b_1 + b_2 x + b_3 x^2 + \cdots,$$

we obtain

$$b_1 = 1,$$

$$b_2 = a_1 = 1,$$

$$b_3 = a_1 b_2 + a_2 b_1 = 1 - \frac{1}{(2!)^2} = \frac{3}{4};$$

$$b_4 = a_1 b_3 + a_2 b_2 + a_3 b_1 = \frac{3}{4} - \frac{1}{(2!)^2} + \frac{1}{(3!)^2} = \frac{3}{4} - \frac{1}{4} + \frac{1}{36} = \frac{19}{36},$$

$$b_5 = a_1 b_4 + a_2 b_3 + a_3 b_2 + a_4 b_1$$

$$= \frac{19}{36} - \frac{1}{4} \times \frac{3}{4} + \frac{1}{36} \times 1 - \frac{1}{576} = \frac{211}{576}.$$

It follows

$$\frac{b_1}{b_2} = 1; \qquad \frac{b_2}{b_3} = \frac{4}{3} = 1.333\cdots;$$

$$\frac{b_3}{b_4} = \frac{3}{4} \times \frac{36}{19} = \frac{27}{19} = 1.4210\cdots, \qquad \frac{b_4}{b_5} = \frac{19}{36} \times \frac{576}{211} = 1.4408\cdots,$$

where the last result is correct to three significant figures.

Example Find a root of the equation $\sin x = 1 - x$.

Using the expansion of $\sin x$, the given equation may be written as

$$f(x) = 1 - \left(x + x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots\right) = 0.$$

Here

$$a_1 = 2, \quad a_2 = 0, \quad a_3 = \frac{1}{6}, \quad a_4 = 0,$$

$$a_5 = \frac{1}{120}, \quad a_6 = 0, \quad a_7 = -\frac{1}{5040}, \cdots$$

we write

$$\left[1 - \left(2x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{5040} + \cdots\right)\right]^{-1} = b_1 + b_2 x + b_3 x^2 + \cdots$$

We then obtain

$$b_1 = 1;$$

$$b_2 = a_1 = 2;$$

$$b_3 = a_1 b_2 + a_2 b_1 = 4;$$

$$b_4 = a_1 b_3 + a_2 b_2 + a_3 b_1 = 8 - \frac{1}{6} = \frac{47}{6};$$

$$b_5 = a_1 b_4 + a_2 b_3 + a_3 b_2 + a_4 b_1 = \frac{46}{3};$$

$$b_6 = a_1 b_5 + a_2 b_4 + a_3 b_3 + a_4 b_2 + a_5 b_1 = \frac{3601}{120};$$

Therefore,

$$\frac{b_1}{b_2} = \frac{1}{2}; \qquad \frac{b_2}{b_3} = \frac{1}{2};$$

$$\frac{b_3}{b_4} = \frac{24}{27} = 0.5106382 \qquad \frac{b_4}{b_5} = \frac{47}{92} = 0.5108695$$

$$\frac{b_5}{b_6} = \frac{1840}{3601} = 0.5109691.$$

The root, correct to four decimal places is 0.5110

Exercises

1. Using Ramanujan's method, obtain the first-eight convergents of the equation

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

2. Using Ramanujan's method, find the real root of the equation $x + x^3 = 1.$

NEWTON RAPHSON ETC..

The Newton-Raphson method, or Newton Method, is a powerful technique for solving equations numerically. Like so much of the differential calculus, it is based on the simple idea of linear approximation.

Newton – Raphson Method

Consider $f(x)=0$, where f has continuous derivative f'. From the figure we can say that at $x=a,\ y=f(a)=0;$ which means that a is a solution to the equation $f(x)=0$. In order to find the value of a, we start with any arbitrary point x_0. From figure we can see that, the tangent to the curve f at $(x_0, f(x_0))$ (with slope $f'(x_0)$) touches the x-axis at x_1.

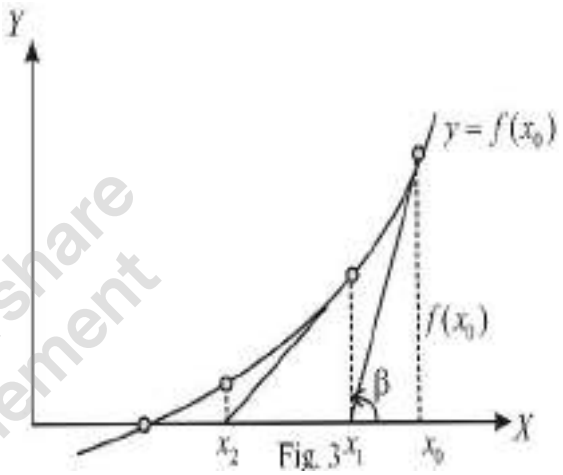


Fig. 3

$$\text{Now, } \tan\ = f'(x_0)=\frac{f(x_0)-f(x_1)}{x_0-x_1},$$

As $f(x_1)=0,$ the above simplifies to

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

In the second step, we compute

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

in the third step we compute

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

and so on. More generally, we write x_{n+1} in terms of x_n, $f(x_n)$ and $f'(x_n)$ for $n=1, 2, \ldots$ by means of the **Newton-Raphson** formula

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

The refinement on the value of the root x_n is terminated by any of the following conditions.

(i) Termination after a pre-fixed number of steps

(ii) After *n* iterations where, $|x_{n+1} - x_n| \le \varepsilon$ (*for a given* $\varepsilon > 0$), or

(iii) After *n* iterations, where $f(x_n) \le \alpha$ (*for a given* $\alpha > 0$).

Termination after a fixed number of steps is not advisable, because a fine approximation cannot be ensured by a fixed number of steps.

Algorithm: The steps of the Newton-Raphson method to find the root of an equation $f(x) = 0$ are

1. Evaluate $f'(x)$
2. Use an initial guess of the root, x_i, to estimate the new value of the root, x_{i+1}, as

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

3. Find the absolute relative approximate error $|\in_a|$ as

$$|\in_a| = \left|\frac{x_{i+1} - x_i}{x_{i+1}}\right| \times 100$$

4. Compare the absolute relative approximate error with the pre-specified relative error tolerance, $\in_s$. If $|\in_a| > \in_s$ then go to Step 2, else stop the algorithm. Also, check if the number of iterations has exceeded the maximum number of iterations allowed. If so, one needs to terminate the algorithm and notify the user.

The method can be used for both algebraic and transcendental equations, and it also works when coefficients or roots are complex. It should be noted, however, that in the case of an algebraic equation with real coefficients, a complex root cannot be reached with a real starting value.

Example Set up a Newton iteration for computing the square root of a given positive number. Using the same find the square root of 2 exact to six decimal places.

Let *c* be a given positive number and let *x* be its positive square root, so that $x = \sqrt{c}$. Then $x^2 = c$ or

$$f(x) = x^2 - c = 0$$

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

Using the Newton's iteration formula we have

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

or $$x_{n+1} = \frac{x_n}{2} + \frac{c}{2x_n}$$

or $$x_{n+1} = \frac{1}{2}\left(x_n + \frac{c}{x_n}\right), n = 0.1, 2, \cdots,$$

Now to find the square root of 2, **let** $c = 2$, so that

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{2}{x_n}\right), n = 0, 1, 2, \cdots$$

Choose $x_0 = 1$. Then

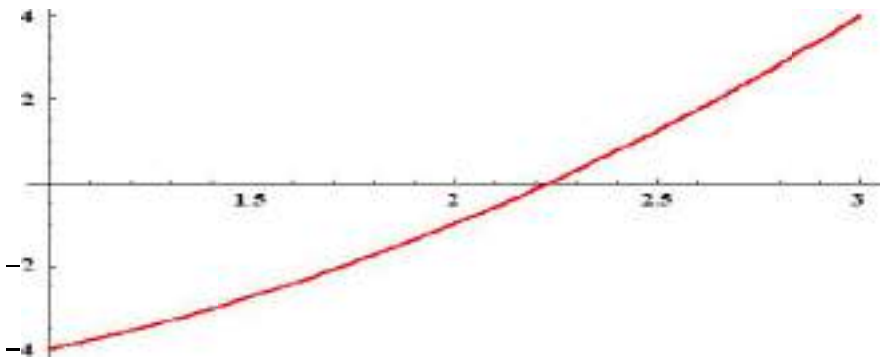
$x_1 = 1.500000$, $x_2 = 1.416667$, $x_3 = 1.414216$, $x_4 = 1.414214$, …

and accept 1.414214 as the square root of 2 exact to 6D.

Historical Note: Heron of Alexandria (60 CE?) used a pre-algebra version of the above recurrence. It is still at the heart of computer algorithms for finding square roots.

Example. Let us find an approximation to $\sqrt{5}$ to ten decimal places.

Note that $\sqrt{5}$ is an irrational number. Therefore the sequence of decimals which defines $\sqrt{5}$ will not stop. Clearly $\sqrt{5}$ is the only zero of $f(x) = x^2 - 5$ on the interval [1, 3]. See the Picture.



Let (x_n) be the successive approximations obtained through Newton's method. We have

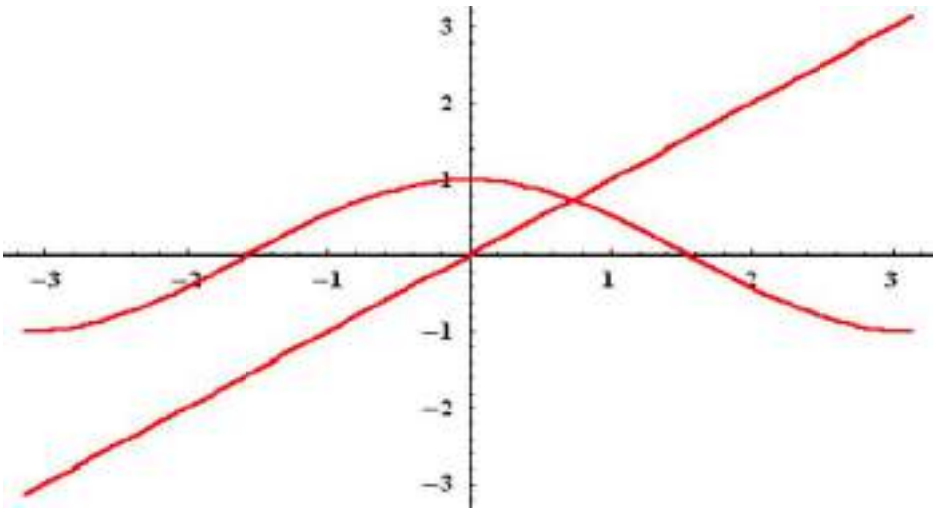
$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^2 - 5}{2x_n}.$$

Let us start this process by taking $x_1 = 2$.

$$\begin{aligned}
x_1 &= 2 \\
x_2 &= 2.25 \\
x_3 &= 2.23611111111111111111111111111111 \\
x_4 &= 2.23606797791580400276052449965493 4 \\
x_5 &= 2.23606797749978969644787282832711 0 \\
x_6 &= 2.23606797749978969640917366873127 6
\end{aligned}$$

Example. Let us approximate the only solution to the equation $x = \cos x$

In fact, looking at the graphs we can see that this equation has one solution.



This solution is also the only zero of the function $f(x) = x - \cos x$. So now we see how Newton's method may be used to approximate *r*. Since *r* is between 0 and ®/2 , we set x_1 = 1. The rest of the sequence is generated through the formula

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

We have

$$\begin{aligned}
x_1 &= 1. \\
x_2 &= 0.750363867840243893034942306682177 \\
x_3 &= 0.739112890911361670360585290904890 \\
x_4 &= 0.739085133385283969760125120856804 \\
x_5 &= 0.739085133215160641661702625685026 \\
x_6 &= 0.739085133215160641655312087673873 \\
x_7 &= 0.739085133215160641655312087673873 \\
x_8 &= 0.739085133215160641655312087673873
\end{aligned}$$

Example Apply Newton's method to solve the algebraic equation $f(x) = x^3 + x - 1 = 0$ correct to 6 decimal places. (Start with x_0=1)

$$f(x) = x^3 + x - 1,$$

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

and substituting these in Newton's iterative formula, we have

$$x_{n+1} = x_n - \frac{x_n^3 + x_n - 1}{3x_n^2 + 1} \quad \text{or} \quad x_{n+1} = \frac{2x_n^3 + 1}{3x_n^2 + 1}, \; n = 0,1,2,\ldots.$$

Starting from x_0=1.000 000,

$x_1 + 0.750000$, $x_2 = 0.686047$, $x_3 = 0.682340$, $x_4 = 0.682328$, $\cdots$ and we accept 0.682328 as an approximate solution of $f(x) = x^3 + x - 1 = 0$ correct to 6 decimal places.

Example Set up Newton-Raphson iterative formula for the equation

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

Solution

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

Noting that $\log_{10} x = \log_e x \cdot \log_{10} e \approx 0.4343 \log_e x,$

we obtain $f(x) = 0.4343x \log_e x - 1.2.$

$$f'(x) = 0.4343 \log_e x + 0.4343x \times \frac{1}{x} = \log_{10} x + 0.4343$$

and hence the Newton's iterative formula for the given equation is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{0.4343x \log_e x_n - 1.2}{\log_{10} x + 0.4343}.$$

Example Find the positive solution of the transcendental equation

$$2 \sin x = x.$$

Here $f(x) = x - 2 \sin x,$

so that $f'(x) = 1 - 2 \cos x$

Substituting in Newton's iterative formula, we have

$$x_{n+1} = x_n - \frac{x_n - 2 \sin x_n}{1 - 2 \cos x_n}, \quad n = 0.1, 2, \cdots \quad \text{or}$$

$$x_{n+1} = \frac{2(\sin x_n - x_n \cos x_n)}{1 - 2 \cos x_n} = \frac{N_n}{D_n}, \quad n = 0.1, 2, \cdots$$

where we take $N_n = 2(\sin x_n - x_n \cos x_n)$ and $D_n = 1 - 2 \cos x_n$, to easy our calculation. Values calculated at each step are indicated in the following table (Starting with $x_0 = 2$).

n	x_n	N_n	D_n	x_{n+1}
0	2.000	3.483	1.832	1.901
1	1.901	3.125	1.648	1.896
2	1.896	3.107	1.639	1.896

1.896 is an approximate solution to $2 \sin x = x.$

Example Use Newton-Raphson method to find a root of the equation $x^3 - 2x - 5 = 0.$

Here $f(x) = x^3 - 2x - 5$ and $f'(x) = 3x^2 - 2.$ Hence Newton's iterative formula becomes

$$x_{n+1} = x_n - \frac{x_n^{3`} - 2x_n - 5}{3x_n^2 - 2}$$

Choosing $x_0 = 2$, we obtain $f(x_0) = -1$ and $f'(x_0) = 10$.

$$x_1 = 2 - \left(-\frac{1}{10}\right) = 2.1$$

$$f(x_1) = (2.1)^3 - 2(2.1) - 5 = 0.06,$$

and $$f'(x_1) = 3(2.1)^2 - 2 = 11.23.$$

$$x_2 = 2.1 - \frac{0.061}{11.23} = 2.094568.$$

2.094568 is an approximate root.

Example Find a root of the equation $x \sin x + \cos x = 0$.

We have

$$f(x) = x \sin x + \cos x \text{ and } f'(x) = x \cos x.$$

Hence the iteration formula is

$$x_{n+1} = x_n - \frac{x_n \sin x_n + \cos x_n}{x_n \cos x_n}$$

With $x_0 = ®$, the successive iterates are given below:

n	x_n	$f(x_n)$	x_{n+1}
0	3.1416	–1.0	2.8233
1	2.8233	–0.0662	2.7986
2	2.7986	–0.0006	2.7984
3	2.7984	0.0	2.7984

Example Find a real root of the equation $x = e^{-x}$, using the Newton – Raphson method.

$$f(x) = xe^x - 1 = 0$$

Let $x_0 = 1$. Then

$$x_1 = 1 - \frac{e-1}{2e} = \frac{1}{2}\left(1 + \frac{1}{e}\right) = 0.6839397$$

Now $f(x_1) = 0.3553424$, and $f'(x_1) = 3.337012$,

$$x_2 = 0.6839397 - \frac{0.3553424}{3.337012} = 0.5774545.$$

$x_3 = 0.5672297$ and $x_4 = 0.5671433$.

Example f(x) = x−2+lnx has a root near x = 1.5. Use the Newton-Raphson formula to obtain a better estimate.

Here x0 = 1.5, f(1.5)= −0.5 + ln(1.5)= −0.0945

$$f'(x) = 1 + \frac{1}{x};\ f'(1.5) = \frac{5}{3};\ x_1 = 1.5 - \frac{(-0.0945)}{1.6667} = 1.5567$$

The Newton-Raphson formula can be used again: this time beginning with 1.5567 as our initial

$$x_2 = 1.5567 - \frac{(-0.0007)}{1.6424} = 1.5571$$

This is in fact the correct value of the root to 4 d.p.

Generalized Newton's Method

If ¶ is a root of $f(x) = 0$ with multiplicity p, then the generalized Newton's formula is

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

Since ¶ is a root of $f(x) = 0$ with multiplicity p, it follows that ¶ is a root of $f'(x) = 0$ with multiplicity $(p-1)$, of $f''(x) = 0$ with multiplicity $(p-2)$, and so on. Hence the expressions

$$x_0 - p\frac{f(x_0)}{f'(x_0)},\ x_0 - (p-1)\frac{f'(x_0)}{f''(x_0)},\ x_0 - (p-2)\frac{f'''(x_0)}{f'''(x_0)}$$

must have the same value if there is a root with multiplicity p, provided that the initial approximation x_0 is chosen sufficiently close to the root.

Example Find a double root of the equation

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

Here $f'(x) = 3x^2 - 2x - 1$, and $f''(x) = 6x - 2$. With $x_0 = 0.8$, we obtain

$$x_0 - 2\frac{f(x_0)}{f'(x_0)} = 0.8 - 2\frac{0.072}{-(0.68)} = 1.012,$$

and

$$x_0 - \frac{f'(x_0)}{f''(x_0)} = 0.8 - \frac{-(0.68)}{2.8} = 1.043,$$

The closeness of these values indicates that there is a doublel root near to unity. For the next approximation, we choose $x_1 = 1.01$ and obtain

$$x_1 - 2\frac{f(x_1)}{f'(x_1)} = 1.01 - 0.0099 = 1.0001,$$

and $$x_1 - \frac{f'(x_1)}{f''(x_1)} = 1.01 - 0.0099 = 1.0001,$$

Hence we conclude that there is a double root at $x = 1.0001$ which is sufficiently close to the actual root unity.

On the other hand, if we apply Newton-Raphson method with $x_0 = 0.8,$ we obtain $x_1 = 0.8 + 0.106 \approx 0.91,$ and $x_2 = 0.91 + 0.046 \approx 0.96.$

Exercises

1. Approximate the real root to two four decimal places of $x^3 + 5x - 3 = 0$
2. Approximate to four decimal places $\sqrt[3]{3}$
3. Find a positive root of the equation $x^4 + 2x + 1 = 0$ correct to 4 places of decimals. (Choose $x_0 = 1.3$)
4. Explain how to determine the square root of a real number by $N - R$ method and using it determine $\sqrt{3}$ correct to three decimal places.
5. Find the value of $\sqrt{2}$ correct to four decimals places using Newton Raphson method.
6. Use the Newton-Raphson method, with 3 as starting point, to find a fraction that is within 10^{-8} of $\sqrt{10}$.
7. Design Newton iteration for the cube root. Calculate $\sqrt[3]{7}$, starting from $x_0 = 2$ and performing 3 steps.
8. Calculate $\sqrt{7}$ by Newton's iteration, starting from $x_0 = 2$ and calculating x_1, x_2, x_3. Compare the results with the value $\sqrt{7} = 2.645751$

9. Design a Newton's iteration for computing k^{th} root of a positive number c.

10. Find all real solutions of the following equations by Newton's iteration method.

(a) $\sin x = \frac{x}{2}$. (b) $\ln x = 1 - 2x$ (c) $\cos x = \sqrt{x}$

11. Using Newton-Raphson method, find the root of the equation $x^3 - x^2 - x - 3 = 0$, correct to three decimal places

12. Apply Newton's method to the equation

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

starting from the given $x_0 = 2$ and performing 3 steps.

13. Apply Newton's method to the equation

$$x^4 - x^3 - 2x - 34 = 0$$

starting from the given $x_0 = 3$ and performing 3 steps.

14. Apply Newton's method to the equation

$$x^3 - 3.9x^2 + 4.79x - 1.881 = 0$$

starting from the given $x_0 = 1$ and performing 3 steps.

Ramanujan's Method

We need the following Theorem:

Binomial Theorem: If n is any rational number and $|x| < 1$, then

$$(1+x)^n = 1 + \frac{n}{1}x + \frac{n(n-1)}{1\cdot 2}x^2 + \ \ldots\ + \frac{n(n-1)\ \ldots\ (n-(r-1))}{1\cdot 2\cdot\ \ldots\ \cdot r}x^r + \ \ldots$$

In particular,

$$(1+x)^{-1} = 1 - x + x^2 - x^3 + \ldots + (-1)^n x^n + \ldots$$

and $$(1-x)^{-1} = 1 + x + x^2 + x^3 + \ldots + x^n + \ldots$$

Indian Mathematician Srinivasa Ramanujan (1887-1920) described an iterative method which can be used to determine the smallest root of the equation

$$f(x)=0,$$

where $f(x)$ is of the form

$$f(x)=1-(a_1x+a_2x^2+a_3x^2+a_4x^4+\cdots).$$

For smaller values of x, we can write

$$[1-(a_1x+a_2x^2+a_3x^3+a_4x^4+\cdots)]^{-1}=b_1+b_2x+b_3x^2+\cdots$$

Expanding the left-hand side using binomial theorem , we obtain

$$1+(a_1x+a_2x^2+a_3x^3+\cdots)+(a_1x+a_2x^2+a_3x^3+\cdots)^2+\cdots$$

$$=b_1+b_2x+b_3x^2+\cdots$$

Comparing the coefficients of like powers of x on both sides of we obtain

$$\left.\begin{aligned} &b_1=1,\\ &b_2=a_1=a_1b_1,\\ &b_3=a_1^2+a_2=a_1b_2+a_2b_1,\\ &\vdots\\ &b_n=a_1b_{n-1}+a_2b_{n-2}+\cdots+a_{n-1}b_1 \quad n=2,3,\cdots \end{aligned}\right\}$$

Then b_n/b_{n+1} approach a root of the equation $f(x)=0$.

Example Find the smallest root of the equation

$$f(x)=x^3-6x^2+11x-6=0.$$

Solution

The given equation can be written as $f(x)$

$$f(x)=1-\frac{1}{6}(11x-6x^2+x^3)$$

Comparing,

$$a_1=\frac{11}{6},\quad a_2=-1,\quad a_3=\frac{1}{6},\quad a_4=a_5=\cdots=0$$

To apply Ramanujan's method we write

$$1-\left(\frac{11x-6x^2+x^3}{6}\right)^{-1}=b_1+b_2x+b_3x^2+\cdots$$

Hence,

$$b_1 = 1;$$

$$b_2 = a_1 = \frac{11}{6};$$

$$b_3 = a_1 b_2 + a_2 b_1 = \frac{121}{36} - 1 = \frac{85}{36};$$

$$b_4 = a_1 b_3 + a_2 b_2 + a_3 b_1 = \frac{575}{216};$$

$$b_5 = a_1 b_4 + a_2 b_3 + a_3 b_2 + a_4 b_1 = \frac{3661}{1296};$$

$$b_6 = a_1 b_5 + a_2 b_4 + a_3 b_3 + a_4 b_2 + a_5 b_1 = \frac{22631}{7776};$$

Therefore,

$$\frac{b_1}{b_2} = \frac{6}{11} = 0.54545\,; \qquad \frac{b_2}{b_3} = \frac{66}{85} = 0.7764705$$

$$\frac{b_3}{b_4} = \frac{102}{115} = 0.8869565\,; \qquad \frac{b_4}{b_5} = \frac{3450}{3661} = 0.9423654$$

$$\frac{b_5}{b_6} = \frac{3138}{3233} = 0.9706155$$

By inspection, a root of the given equation is unity and it can be seen that the successive convergents $\frac{b_n}{b_{n+1}}$ approach this root.

Example Find a root of the equation $xe^x = 1$.

Let $xe^x = 1$

Recall $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$

Hence,

$$f(x) = 1 - \left(x + x^2 + \frac{x^3}{2} + \frac{x^4}{6} + \frac{x^5}{24} + \cdots \right) = 0$$

$$a_1 = 1, \quad a_2 = 1, \quad a_3 = \frac{1}{2}, \quad a_4 = \frac{1}{6}, \quad a_5 = \frac{1}{24}, \cdots$$

We then have

$$b_1 = 1;$$

$$b_2 = a_2 = 1;$$

$$b_3 = a_1 b_2 + a_2 b_1 = 1 + 1 = 2;$$

$$b_4 = a_1 b_3 + a_2 b_2 + a_3 b_1 = 2 + 1 + \frac{1}{2} = \frac{7}{2};$$

$$b_5 = a_1 b_4 + a_2 b_3 + a_3 b_2 + a_4 b_1 = \frac{7}{2} + 2 + \frac{1}{2} + \frac{1}{6} = \frac{37}{6};$$

$$b_6 = a_1 b_5 + a_2 b_4 + a_3 b_3 + a_4 b_2 + a_5 b_1 = \frac{37}{6}; + \frac{7}{2} + 1 + \frac{1}{6} + \frac{1}{24} = \frac{261}{24};$$

Therefore,

$$\frac{b_2}{b_3} = \frac{1}{2} = 0.5\,; \qquad \frac{b_3}{b_4} = \frac{4}{7} = 0.5714\,;$$

$$\frac{b_4}{b_5} = \frac{21}{37} = 0.56756756\,; \qquad \frac{b_5}{b_6} = \frac{148}{261} = 0.56704980\,.$$

Example Using Ramanujan's method, find a real root of the equation

$$1 - x + \frac{x^2}{(2!)^2} - \frac{x^3}{(3!)^2} + \frac{x^4}{(4!)^2} - \cdots = 0.$$

Solution

Let $$f(x) = 1 - \left[x - \frac{x^2}{(2!)^2} + \frac{x^3}{(3!)^2} - \frac{x^4}{(4!)^2} + \cdots \right] = 0.$$

Here

$$a_1 = 1, \quad a_2 = -\frac{1}{(2!)^2}, \quad a_3 = \frac{1}{(3!)^2}, \quad a_4 = -\frac{1}{(4!)^2},$$

$$a_5 = \frac{1}{(5!)^2}, \quad a_6 = -\frac{1}{(6!)^2}, \cdots$$

Writing

$$\left\{1-\left[x-\frac{x^2}{(2!)}+\frac{x^3}{(3!)^2}-\frac{x^4}{(4!)^2}+\cdots\right]\right\}^{-1}=b_1+b_2x+b_3x^2+\cdots,$$

we obtain

$$b_1=1,$$

$$b_2=a_1=1,$$

$$b_3=a_1b_2+a_2b_1=1-\frac{1}{(2!)^2}=\frac{3}{4};$$

$$b_4=a_1b_3+a_2b_2+a_3b_1=\frac{3}{4}-\frac{1}{(2!)^2}+\frac{1}{(3!)^2}=\frac{3}{4}-\frac{1}{4}+\frac{1}{36}=\frac{19}{36},$$

$$b_5=a_1b_4+a_2b_3+a_3b_2+a_4b_1$$

$$=\frac{19}{36}-\frac{1}{4}\times\frac{3}{4}+\frac{1}{36}\times1-\frac{1}{576}=\frac{211}{576}.$$

It follows

$$\frac{b_1}{b_2}=1; \qquad \frac{b_2}{b_3}=\frac{4}{3}=1.333\cdots;$$

$$\frac{b_3}{b_4}=\frac{3}{4}\times\frac{36}{19}=\frac{27}{19}=1.4210\cdots, \qquad \frac{b_4}{b_5}=\frac{19}{36}\times\frac{576}{211}=1.4408\cdots,$$

where the last result is correct to three significant figures.

Example Find a root of the equation $\sin x=1-x.$

Using the expansion of $\sin x,$ the given equation may be written as

$$f(x)=1-\left(x+x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots\right)=0.$$

Here

$$a_1=2, \quad a_2=0, \quad a_3=\frac{1}{6}, \quad a_4=0,$$

$$a_5=\frac{1}{120}, \quad a_6=0, \quad a_7=-\frac{1}{5040},\cdots$$

we write

$$\left[1-\left(2x-\frac{x^3}{6}+\frac{x^5}{120}-\frac{x^7}{5040}+\cdots\right)\right]^{-1}=b_1+b_2x+b_3x^2+\cdots$$

We then obtain

$$b_1=1;$$

$$b_2=a_1=2;$$

$$b_3=a_1b_2+a_2b_1=4;$$

$$b_4=a_1b_3+a_2b_2+a_3b_1=8-\frac{1}{6}=\frac{47}{6};$$

$$b_5=a_1b_4+a_2b_3+a_3b_2+a_4b_1=\frac{46}{3};$$

$$b_6=a_1b_5+a_2b_4+a_3b_3+a_4b_2+a_5b_1=\frac{3601}{120};$$

Therefore,

$$\frac{b_1}{b_2}=\frac{1}{2};\qquad \frac{b_2}{b_3}=\frac{1}{2};$$

$$\frac{b_3}{b_4}=\frac{24}{27}=0.5106382\qquad \frac{b_4}{b_5}=\frac{47}{92}=0.5108695$$

$$\frac{b_5}{b_6}=\frac{1840}{3601}=0.5109691.$$

The root, correct to four decimal places is 0.5110

Exercises

1. Using Ramanujan's method, obtain the first-eight convergents of the equation

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

2. Using Ramanujan's method, find the real root of the equation $x+x^3=1.$

The Secant Method

We have seen that the Newton-Raphson method requires the evaluation of derivatives of the function and this is not always possible, particularly in the case of functions arising in practical problems. In the secant method, the derivative at x_n is approximated by the formula

$$f'(x_n) \approx \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}},$$

which can be written as

$$f_n' = \frac{f_n - f_{n-1}}{x_n - x_{n-1}},$$

where $f_n = f(x_n)$. Hence, the Newton-Raphson formula becomes

$$x_{n+1} = x_n - \frac{f_n(x_n - x_{n-1}}{f_n = f_{n-1}} = \frac{x_{n+1} f_n - x_n f_{n-1}}{f_n = f_{n-1}}.$$

It should be noted that this formula requires two initial approximations to the root.

Example Find a real root of the equation $x^3 - 2x - 5 = 0$ using secant method.

Let the two initial approximations be given by $x_{-1} = 2$ and $x_0 = 3$.

We have

$$f(x_{-1}) = f_1 = 8 - 9 = -1, \quad \text{and} \quad f(x_0) = f_0 = 27 - 11 = 16.$$

$$x_1 = \frac{2(16) - 3(-1)}{17} = \frac{35}{17} = 2.058823529.$$

Also,

$$f(x_1) = f_1 = -0.390799923.$$

$$x_2 = \frac{x_0 f_1 - x_1 f_0}{f_1 - f_0} = \frac{3(-0.390799923) - 2.058823529(16)}{-16.390799923} = 2.08126366.$$

Again

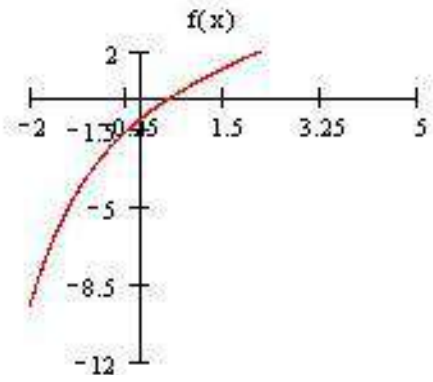
$$f(x_2) = f_2 = -0.147204057.$$

$$x_3 = 2.094824145.$$

***Example*:** Find a real root of the equation $x - e^{-x} = 0$ using secant method.

Solution

The graph of $f(x) = x - e^{-x}$ is as shown here.



Let us assume the initial approximation to the roots as 1 and 2. That is consider $x_{-1}=1$ and $x_0=2$

$$f(x_{-1})=f_{-1}=1-e^{-1}=1-0.367879441=0.632120559 \quad \text{and}$$

$$f(x_0)=f_0=2-e^{-2}=2-0.135335283=1.864664717.$$

Step 1: Putting $n=0$, we obtain $x_1=\dfrac{x_{-1}f_0-x_0f_{-1}}{f_0-f_{-1}}$

$$\text{Here, } x_1=\frac{1(1.864664717)-2(0.632120559)}{1.864664717-0.632120559}=\frac{0.600423599}{1.232544158}=0.487142.$$

Also,

$$f(x_1)=f_1=0.487142-e^{-0.487142}=\text{-}0.12724.$$

Step 2: Putting $n=1$, we obtain

$$x_2=\frac{x_0f_1-x_1f_0}{f_1-f_0}=\frac{2(\text{-}0.12724)-0.487142(1.864664717)}{\text{-}0.12724-1.864664717}=\frac{\text{-}1.16284}{\text{-}1.99190}=0.58378$$

Again

$$f(x_2)=f_2=0.58378-e^{-0.58378}=0.02599.$$

Step 3: Setting $n=2$,

$$x_3=\frac{x_1f_2-x_2f_1}{f_2-f_1}=\frac{0.487142(0.02599)-0.58378(\text{-}0.12724)}{0.02599-(\text{-}0.12724)}=\frac{0.08694}{0.15323}=0.56738$$

$$f(x_3)=f_3=0.56738-e^{-0.56738}=0.00037.$$

Step 4: Setting $n=3$ in (*),

$$x_4 = \frac{x_2 f_3 - x_3 f_2}{f_3 - f_2} = \frac{0.58378(0.00037) - 0.56738(0.02599)}{0.00037 - 0.02599} = \frac{-0.01453}{-0.02562} = 0.5671$$

Approximating to three digits, the root can be considered as 0.567.

Exercises

1. Determine the real root of the equation $xe^x = 1$ using the secant method. Compare your result with the true value of $x = 0.567143\cdots$.
2. Use the secant method to determine the root, lying between 5 and 8, of the equation $x^{2.2} = 69$.

Objective Type Questions

(a) The Newton-Raphson method formula for finding the square root of a real number C from the equation $x^2 - C = 0$ is,

(i) $x_{n+1} = \frac{x_n}{2}$ (ii) $x_{n+1} = \frac{3x_n}{2}$ (iii) $x_{n+1} = \frac{1}{2}\left(x_n + \frac{C}{x_n}\right)$ (iv) None of these

(b) The next iterative value of the root of $2x^2 - 3 = 0$ using the Newton-Raphson method, if the initial guess is 2, is

(i) 1.275 (ii) 1.375 (iii) 1.475 (iv) None of these

(c) The next iterative value of the root of $2x^2 - 3 = 0$ using the secant method, if the initial guesses are 2 and 3, is

(i) 1 (ii) 1.25 (iii) 1.5 (iv) None of these

(d) In secant method,

(i) $x_{n+1} = \frac{x_{n-1} f_n - x_n f_{n-1}}{f_n - f_{n-1}}$ (ii) $x_{n+1} = \frac{x_n f_n - x_{n-1} f_{n-1}}{f_n - f_{n-1}}$ (iii) $x_{n+1} = \frac{x_{n-1} f_{n-1} - x_n f_n}{f_{n-1} - f_n}$

(iv) None of these

Answers

(a) (iii) $x_{n+1} = \frac{1}{2}\left(x_n + \frac{C}{x_n}\right)$

(b) (ii) 1.375

(c) (iii) 1.5

(d) (i) $x_{n+1} = \dfrac{x_{n-1} f_n - x_n f_{n-1}}{f_n - f_{n-1}}$

Theoretical Questions with Answers:

1. What is the difference between bracketing and open method?

   Ans: For finding roots of a nonlinear equation $f(x) = 0$, bracketing method requires two guesses which contain the exact root. But in open method initial guess of the root is needed without any condition of bracketing for starting the iterative process to find the solution of an equation.

2. When the Generalized Newton's methods for solving equations is helpful?

   Ans: To solve the find the oot of $f(x) = 0$ with multiplicity p, the generalized Newton's formula is required.

3. What is the importance of Secant method over Newton-Raphson method?

   Ans: Newton-Raphson method requires the evaluation of derivatives of the function and this is not always possible, particularly in the case of functions arising in practical problems. In such situations Secant method helps to solve the equation with an approximation to the derivative.

****

FINITE DIFFERENCES OPERATORS

For a function y=f(x), it is given that $y_0, y_1, ..., y_n$ are the values of the variable y corresponding to the equidistant arguments, $x_0, x_1, ..., x_n$, where $x_1 = x_0 + h, x_2 = x_0 + 2h,\ x_3 = x_0 + 3h, ..., x_n = x_0 + nh$. In this case, even though Lagrange and divided difference interpolation polynomials can be used for interpolation, some simpler interpolation formulas can be derived. For this, we have to be familiar with some finite difference operators and finite differences, which were introduced by Sir Isaac Newton. Finite differences deal with the changes that take place in the value of a function f(x) due to finite changes in x. Finite difference operators include, forward difference operator, backward difference operator, shift operator, central difference operator and mean operator.

- **Forward difference operator (Δ) :**

For the values $y_0, y_1, ..., y_n$ of a function y=f(x), for the equidistant values $x_0, x_1, x_2, ..., x_n$, where $x_1 = x_0 + h, x_2 = x_0 + 2h,\ x_3 = x_0 + 3h, ..., x_n = x_0 + nh$, the forward difference operator Δ is defined on the function f(x) as,

$$\Delta f(x_i) = f(x_i + h) - f(x_i) = f(x_{i+1}) - f(x_i)$$

That is,

$$\Delta y_i = y_{i+1} - y_i$$

Then, in particular

$$\Delta f(x_0) = f(x_0 + h) - f(x_0) = f(x_1) - f(x_0)$$
$$\Rightarrow \quad \Delta y_0 = y_1 - y_0$$

$$\Delta f(x_1) = f(x_1 + h) - f(x_1) = f(x_2) - f(x_1)$$
$$\Rightarrow \quad \Delta y_1 = y_2 - y_1$$

etc.,

$\Delta y_0, \Delta y_1, ..., \Delta y_i, ...$ are known as the **first forward differences.**

The second forward differences are defined as,

$$\begin{aligned}\Delta^2 f(x_i) &= \Delta\left[\Delta f(x_i)\right] = \Delta\left[f(x_i+h) - f(x_i)\right] \\ &= \Delta f(x_i+h) - \Delta f(x_i) \\ &= f(x_i+2h) - f(x_i+h) - \left[f(x_i+h) - f(x_i)\right] \\ &= f(x_i+2h) - 2f(x_i+h) + f(x_i) \\ &= y_{i+2} - 2y_{i+1} + y_i\end{aligned}$$

In particular,

$$\Delta^2 f(x_0) = y_2 - 2y_1 + y_0 \quad or \quad \Delta^2 y_0 = y_2 - 2y_1 + y_0$$

The third forward differences are,

$$\begin{aligned}\Delta^3 f(x_i) &= \Delta\left[\Delta^2 f(x_i)\right] \\ &= \Delta\left[f(x_i+2h) - 2f(x_i+h) + f(x_i)\right] \\ &= y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i\end{aligned}$$

In particular,

$$\Delta^3 f(x_0) = y_3 - 3y_2 + 3y_1 - y_0 \quad or \quad \Delta^3 y_0 = y_3 - 3y_2 + 3y_1 - y_0$$

In general the n[th] forward difference,

$$\Delta^n f(x_i) = \Delta^{n-1} f(x_i+h) - \Delta^{n-1} f(x_i)$$

The differences $\Delta y_0, \Delta^2 y_0, \Delta^3 y_0 \ldots$ are called the **leading differences.**

Forward differences can be written in a tabular form as follows:

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
x_0	$y_0 = f(x_o)$			
		$\Delta y_0 = y_1 - y_0$		
x_1	$y_1 = f(x_1)$		$\Delta^2 y_0 = \Delta y_1 - \Delta y_0$	
		$\Delta y_1 = y_2 - y_1$		$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0$
x_2	$y_2 = f(x_2)$		$\Delta^2 y_1 = \Delta y_2 - \Delta y_1$	
		$\Delta y_2 = y_3 - y_2$		
x_3	$y_3 = f(x_3)$			

Example Construct the forward difference table for the following x values and its corresponding f values.

x	0.1	0.3	0.5	0.7	0.9	1.1	1.3
f	0.003	0.067	0.148	0.248	0.370	0.518	0.697

x	f	Δf	$\Delta^2 f$	$\Delta^3 f$	$\Delta^4 f$	$\Delta^5 f$
0.1	0.003					
		0.064				
0.3	0.067		0.017			
		0.081		0.002		
0.5	0.148		0.019		0.001	
		0.100		0.003		0.000
0.7	0.248		0.022		0.001	
		0.122		0.004		0.000
0.9	0.370		0.026		0.001	
		0.148		0.005		
1.1	0.518		0.031			
		0.179				
1.3	0.697					

Example Construct the forward difference table, where $f(x) = \frac{1}{x}$, x = 1(0.2)2, 4D.

x	$f(x) = \frac{1}{x}$	Δf first difference	$\Delta^2 f$ second difference	$\Delta^3 f$	$\Delta^4 f$	$\Delta^5 f$
1.0	1.000					
		-0.1667				
1.2	0.8333		0.0477			
		-0.1190		-0.0180		
1.4	0.7143		0.0297		0.0082	-0.0045
		-0.0893		-0.0098		
1.6	0.6250		0.0199		0.0037	
		-0.0694		-0.0061		
1.8	0.5556		0.0138			
		-0.0556				
2.0	0.5000					

Example Construct the forward difference table for the data

$$\begin{array}{lccccc} x: & -2 & 0 & 2 & 4 \\ y=f(x): & 4 & 9 & 17 & 22 \end{array}$$

The forward difference table is as follows:

x	y=f(x)	Δy	$\Delta^2 y$	$\Delta^3 y$
-2	4			
		Δy_0 =5		
0	9		$\Delta^2 y_0$ =3	
		Δy_1 =8		$\Delta^3 y_0$ =-6
2	17		$\Delta^2 y_1$ =-3	
		Δy_2 =5		
4	22			

Properties of Forward difference operator (Δ):

(i) Forward difference of a constant function is zero.

Proof: Consider the constant function $f(x)=k$

Then, $\Delta f(x)=f(x+h)-f(x)=k-k=0$

(ii) For the functions $f(x)$ *and* $g(x)$; $\Delta\big(f(x)+g(x)\big)=\Delta f(x)+\Delta g(x)$

Proof: By definition,

$$\begin{aligned}\Delta\big(f(x)+g(x)\big)&=\Delta\big((f+g)(x)\big)\\&=(f+g)(x+h)-(f+g)(x)\\&=f(x+h)+g(x+h)-\big(f(x)+g(x)\big)\\&=f(x+h)-f(x)+g(x+h)-g(x)\\&=\Delta f(x)+\Delta g(x)\end{aligned}$$

(iii)Proceeding as in (ii), for the constants a and b,

$$\Delta\big(af(x)+bg(x)\big)=a\Delta f(x)+b\Delta g(x).$$

(iv)Forward difference of the product of two functions is given by,

$$\Delta\big(f(x)g(x)\big)=f(x+h)\Delta g(x)+g(x)\Delta f(x)$$

Proof:

$$\begin{aligned}\Delta\big(f(x)g(x)\big) &= \Delta\big((fg)(x)\big)\\ &= (fg)(x+h)-(fg)(x)\\ &= f(x+h)g(x+h)-f(x)g(x)\end{aligned}$$

Adding and subtracting $f(x+h)g(x)$, the above gives

$$\Delta\big(f(x)g(x)\big) = f(x+h)g(x+h)-f(x+h)g(x)+f(x+h)g(x)-f(x)g(x)$$

$$\begin{aligned}&= f(x+h)\big[g(x+h)-g(x)\big]+g(x)\big[f(x+h)-f(x)\big]\\ &= f(x+h)\Delta g(x)+g(x)\Delta f(x)\end{aligned}$$

Note : Adding and subtracting $g(x+h)f(x)$ instead of $f(x+h)g(x)$, it can also be proved that

$$\Delta\big(f(x)g(x)\big) = g(x+h)\Delta f(x)+f(x)\Delta g(x)$$

(v) Forward difference of the quotient of two functions is given by

$$\Delta\left(\frac{f(x)}{g(x)}\right)=\frac{g(x)\Delta f(x)-f(x)\Delta g(x)}{g(x+h)g(x)}$$

Proof:

$$\begin{aligned}\Delta\left(\frac{f(x)}{g(x)}\right) &= \frac{f(x+h)}{g(x+h)}-\frac{f(x)}{g(x)}\\ &= \frac{f(x+h)g(x)-f(x)g(x+h)}{g(x+h)g(x)}\\ &= \frac{f(x+h)g(x)-f(x)g(x)+f(x)g(x)-f(x)g(x+h)}{g(x+h)g(x)}\\ &= \frac{g(x)\big[f(x+h)-f(x)\big]-f(x)\big[g(x+h)-g(x)\big]}{g(x+h)g(x)}\\ &= \frac{g(x)\Delta f(x)-f(x)\Delta g(x)}{g(x+h)g(x)}\end{aligned}$$

Following are some results on forward differences:

Result 1: The n^{th} forward difference of a polynomial of degree n is constant when the values of the independent variable are at equal intervals.

Result 2: If n is an integer,

$$f(a+nh) = f(a) + {}^nC_1\Delta f(a) + {}^nC_2\Delta^2 f(a) + \cdots + \Delta^n f(a)$$

for the polynomial f(x) in x.

Forward Difference Table

x	f	Δf	$\Delta^2 f$	$\Delta^3 f$	$\Delta^4 f$	$\Delta^5 f$	$\Delta^6 f$
x_0	f_0						
x_1	f_1	Δf_0	$\Delta^2 f_0$				
x_2	f_2	Δf_1	$\Delta^2 f_2$	$\Delta^3 f_0$	$\Delta^4 f_0$		
x_3	f_3	Δf_2	$\Delta^2 f_2$	$\Delta^3 f_1$	$\Delta^4 f_1$	$\Delta^5 f_0$	$\Delta^6 f_0$
x_4	f_4	Δf_3	$\Delta^2 f_3$	$\Delta^3 f_2$	$\Delta^4 f_2$	$\Delta^5 f_1$	
x_5	f_5	Δf_4	$\Delta^2 f_4$	$\Delta^3 f_3$			
x_6	f_6	Δf_5					

Example Express $\Delta^2 f_0$ and $\Delta^3 f_0$ in terms of the values of the function f.

$$\Delta^2 f_0 = \Delta f_1 - \Delta f_0 = f_2 - f_1 - (f_1 - f_0) = f_2 - 2f_1 + f_0$$

$$\Delta^3 f_0 = \Delta^2 f_1 - \Delta^2 f_0 = \Delta f_2 - \Delta f_1 - (\Delta f_1 - \Delta f_0)$$

$$= (f_3 - f_2) - (f_2 - f_1) - (f_2 - f_1) + (f_1 - f_0)$$

$$= f_3 - 3f_2 + 3f_1 - f_0$$

In general,

$$\Delta^n f_0 = f_n - {}^nC_1 f_{n-1} + {}^nC_2 f_{n-2} - {}^nC_3 f_{n-3} + \ldots + (-1)^n f_0 \ .$$

If we write y_n to denote f_n the above results takes the following forms:

$$\Delta^2 y_0 = y_2 - 2y_1 + y_0$$

$$\Delta^3 y_0 = y_3 - 3y_2 + 3y_1 - y_0$$

$$\Delta^n y_0 = y_n - {}^nC_1\, y_{n-1} + {}^nC_2\, y_{n-2} - {}^nC_3\, y_{n-3} + \ldots + (-1)^n y_0$$

Example Show that the value of y_n can be expressed in terms of the leading value y_0 and the leading differences $\Delta y_0, \Delta^2 y_0, \ldots, \Delta^n y_0$.

Solution

(For notational convenience, we treat y_n as f_n and so on.)

From the forward difference table we have

$$\left.\begin{aligned} \Delta f_0 = f_1 - f_0 \quad &\text{or} \quad f_1 = f_0 + \Delta f_0 \\ \Delta f_1 = f_2 - f_1 \quad &\text{or} \quad f_2 = f_1 + \Delta f_1 \\ \Delta f_2 = f_3 - f_2 \quad &\text{or} \quad f_3 = f_2 + \Delta f_2 \end{aligned}\right\}$$

and so on. Similarly,

$$\left.\begin{aligned} \Delta^2 f_0 = \Delta f_1 - \Delta f_0 \ \text{or}\ \Delta f_1 = \Delta f_0 + \Delta^2 f_0 \\ \Delta^2 f_1 = \Delta f_2 - \Delta f_1 \ \text{or}\ \Delta f_2 = \Delta f_1 + \Delta^2 f_1 \end{aligned}\right\}$$

and so on. Similarly, we can write

$$\left.\begin{aligned} \Delta^3 f_0 = \Delta^2 f_1 - \Delta^2 f_0 \ \text{or}\ \Delta^2 f_1 = \Delta^2 f_0 + \Delta^3 f_0 \\ \Delta^3 f_1 = \Delta^2 f_2 - \Delta^2 f_1 \ \text{or}\ \Delta^2 f_2 = \Delta^2 f_1 + \Delta^3 f_1 \end{aligned}\right\}$$

and so on. Also, we can write f_2 as

$$\begin{aligned} f_2 &= (f_0 + \Delta f_0) + (\Delta f_0 + \Delta^2 f_0) \\ &= f_0 + 2\Delta f_0 + \Delta^2 f_0 \\ &= (1+\Delta)^2 f_0 \end{aligned}$$

Hence

$$\begin{aligned} f_3 &= f_2 + \Delta f_2 \\ &= (f_1 + \Delta f_1) + \Delta f_0 + 2\Delta^2 f_0 + \Delta^3 f_0 \\ &= f_0 + 3\Delta f_0 + 3\Delta^2 f_0 + \Delta^3 f_0 \\ &= (1+\Delta)^3 f_0 \end{aligned}$$

That is, we can symbolically write

$$f_1 = (1+\Delta) f_0 \ , \ f_2 = (1+\Delta)^2 f_0 \ , \ f_3 = (1+\Delta)^3 f_0 .$$

Continuing this procedure, we can show, in general

$$f_n = (1+\Delta)^n f_0 .$$

Using binomial expansion, the above is

$$f_n = f_0 + {}^nC_1\Delta f_0 + {}^nC_2\Delta^2 f_0 + \ldots + \Delta^n f_0$$

Thus

$$f_n = \sum_{i=0}^{n} {}^nC_i\, \Delta^i f_0.$$

Backward Difference Operator

For the values $y_0, y_1, \ldots, y_n$ of a function y=f(x), for the equidistant values $x_0, x_1, \ldots, x_n$, where $x_1 = x_0 + h, x_2 = x_0 + 2h,\ x_3 = x_0 + 3h, \ldots, x_n = x_0 + nh$, the **backward difference operator** ∇ is defined on the function f(x) as,

$$\nabla f(x_i) = f(x_i) - f(x_i - h) = y_i - y_{i-1},$$

which is the **first backward difference.**

In particular, we have the first backward differences,

$$\nabla f(x_1) = y_1 - y_0;\ \nabla f(x_2) = y_2 - y_1 \ etc$$

The second backward difference is given by

$$\begin{aligned}\nabla^2 f(x_i) &= \nabla\left(\nabla f(x_i)\right) = \nabla\left[f(x_i) - f(x_i - h)\right] = \nabla f(x_i) - \nabla f(x_i - h)\\ &= \left[f(x_i) - f(x_i - h)\right] - \left[f(x_i - h) - f(x_i - 2h)\right]\\ &= (y_i - y_{i-1}) - (y_{i-1} - y_{i-2})\\ &= y_i - 2y_{i-1} + y_{i-2}\end{aligned}$$

Similarly, the third backward difference, $\nabla^3 f(x_i) = y_i - 3y_{i-1} + 3y_{i-2} - y_{i-3}$ and so on.

Backward differences can be written in a tabular form as follows:

x	Y	∇y	$\nabla^2 y$	$\nabla^3 y$
x_o	$y_0 = f(x_o)$			
		$\nabla y_1 = y_1 - y_0$		
x_1	$y_1 = f(x_1)$		$\nabla^2 y_2 = \nabla y_2 - \nabla y_1$	
		$\nabla y_2 = y_2 - y_1$		$\nabla^3 y_3 = \nabla^2 y_3 - \nabla^2 y_2$
x_2	$y_2 = f(x_2)$		$\nabla^2 y_3 = \nabla y_3 - \nabla y_2$	
		$\nabla y_3 = y_3 - y_2$		
x_3	$y_3 = f(x_3)$			

Relation between backward difference and other differences:

1. $\Delta y_0 = y_1 - y_0 = \nabla y_1;\ \ \Delta^2 y_0 = y_2 - 2y_1 + y_0 = \nabla^2 y_2$ etc.

2. $\Delta - \nabla = \Delta\nabla$

Proof: Consider the function f(x).

$$\Delta f(x) = f(x+h) - f(x)$$
$$\nabla f(x) = f(x) - f(x-h)$$

$$\begin{aligned}(\Delta - \nabla)(f(x)) &= \Delta f(x) - \nabla f(x) \\ &= [f(x+h) - f(x)] - [f(x) - f(x-h)] \\ &= \Delta f(x) - \Delta f(x-h) \\ &= \Delta[f(x) - f(x-h)] \\ &= \Delta[\nabla f(x)]\end{aligned}$$

$$\Rightarrow \quad \Delta - \nabla = \Delta\nabla$$

3. $\nabla = \Delta E^{-1}$

Proof: Consider the function f(x).

$$\nabla f(x) = f(x) - f(x-h) = \Delta f(x-h) = \Delta E^{-1} f(x) \quad \Rightarrow \quad \nabla = \Delta E^{-1}$$

4. $\nabla = 1 - E^{-1}$

Proof: Consider the function f(x).

$$\nabla f(x) = f(x) - f(x-h) = f(x) - E^{-1} f(x) = \left(1 - E^{-1}\right) f(x) \quad \Rightarrow \quad \nabla = 1 - E^{-1}$$

Problem: Construct the backward difference table for the data

x :	−2	0	2	4
$y = f(x)$:	−8	3	1	12

Solution: The backward difference table is as follows:

x	Y=f(x)	∇y	$\nabla^2 y$	$\nabla^3 y$
-2	-8			
		∇y_1 =3-(-8)=11		
0	3		$\nabla^2 y_2$ =-2-11= -13	
		∇y_2 =1-3=-2		$\nabla^3 y_3$ =13-(-13)=26
2	1		$\nabla^2 y_3$ =11-(2)=13	
		∇y_3 =12-1=11		
4	12			

Backward Difference Table

x	f	∇f	$\nabla^2 f$	$\nabla^3 f$	$\nabla^4 f$	$\nabla^5 f$	$\nabla^6 f$
x_0	f_0						
x_1	f_1	∇f_1	$\nabla^2 f_2$				
x_2	f_2	∇f_2	$\nabla^2 f_3$	$\nabla^3 f_3$	$\nabla^4 f_4$	$\nabla^5 f$	
x_3	f_3	∇f_3	$\nabla^2 f_4$	$\nabla^3 f_4$	$\nabla^4 f_5$	5	$\nabla^6 f_6$
x_4	f_4	∇f_4	$\nabla^2 f_5$	$\nabla^3 f_5$	$\nabla^4 f_6$	$\nabla^5 f$	
x_5	f_5	∇f_5	$\nabla^2 f_6$	$\nabla^3 f_6$		6	
x_6	f_6	∇f_6					

Example Show that any value of f (or y) can be expressed in terms of f_n (or y_n) and its backward differences.

Solution

$\nabla f_n = f_n - f_{n-1}$ implies $f_{n-1} = f_n - \nabla f_n$

and $\nabla f_{n-1} = f_{n-1} - f_{n-2}$ implies $f_{n-2} = f_{n-1} - \nabla f_{n-1}$

$\nabla^2 f_n = \nabla f_n - \nabla f_{n-1}$ implies $\nabla f_{n-1} = \nabla f_n - \nabla^2 f_n$

From equations (1) to (3), we obtain

$$f_{n-2} = f_n - 2\nabla f_n + \nabla^2 f_n .$$

Similarly, we can show that

$$f_{n-3} = f_n - 3\nabla f_n + 3\nabla^2 f_n - \nabla^3 f_n .$$

Symbolically, these results can be rewritten as follows:

$$f_{n-1} = (1-\nabla) f_n \ , \ f_{n-2} = (1-\nabla)^2 f_n \ , \ f_{n-3} = (1-\nabla)^3 f_n .$$

Thus, in general, we can write

$$f_{n-r} = (1-\nabla)^r f_n .$$

i.e., $f_{n-r} = f_n - {}^rC_1 \nabla f_n + {}^rC_2 \nabla^2 f_n - \ \ldots + \ (-1)^r \nabla^r f_n$

If we write y_n to denote f_n the above result is:

$$y_{n-r} = y_n - {}^rC_1 \nabla y_n + {}^rC_2 \nabla^2 y_n - \ \ldots + \ (-1)^r \nabla^r y_n$$

Central Differences

Central difference operator ¢ for a function f(x) at x_i is defined as,

$$¢ f(x_i) = f\left(x_i + \frac{h}{2}\right) - f\left(x_i - \frac{h}{2}\right)$$, where *h* being the interval of differencing.

Let $y_{\frac{1}{2}} = f\left(x_0 + \frac{h}{2}\right)$. Then,

$$¢ y_{\frac{1}{2}} = ¢ f\left(x_0 + \frac{h}{2}\right) = f\left(x_0 + \frac{h}{2} + \frac{h}{2}\right) - f\left(x_0 + \frac{h}{2} - \frac{h}{2}\right)$$

$$= f(x_0 + h) - f(x_0) = f(x_1) - f(x_0) = y_1 - y_0$$

$$\Rightarrow \quad ¢ y_{\frac{1}{2}} = \Delta y_0$$

Central differences can be written in a tabular form as follows:

x	y	¢ y	¢ ² y	¢ ³ y
x_o	$y_0 = f(x_o)$			
		$¢ y_{\frac{1}{2}} = y_1 - y_0$		
x_1	$y_1 = f(x_1)$		$¢^2 y_1 = ¢ y_{\frac{3}{2}} - ¢ y_{\frac{1}{2}}$	
		$¢ y_{\frac{3}{2}} = y_2 - y_1$		$¢^3 y_{\frac{3}{2}} = ¢^2 y_2 - ¢^2 y_1$
x_2	$y_2 = f(x_2)$		$¢^2 y_2 = ¢ y_{\frac{5}{2}} - ¢ y_{\frac{3}{2}}$	
		$¢ y_{\frac{5}{2}} = y_3 - y_2$		
x_3	$y_3 = f(x_3)$			

Central Difference Table

x	f	δf	$\delta^2 f$	$\delta^3 f$	$\delta^4 f$
x_0	f_0				
		$\delta f_{1/2}$			
x_1	f_1		$\delta^2 f_1$		
		$\delta f_{3/2}$		$\delta^3 f_{3/2}$	
x_2	f_2		$\delta^2 f_2$		$\delta^4 f_2$
		$\delta f_{5/2}$		$\delta^3 f_{5/2}$	
x_3	f_3		$\delta^2 f_3$		
		$\delta f_{7/2}$			
x_4	f_4				

Example Show that

$$(a)\ \Delta^2 f_m = f_{m+1} - 2f_m + f_{m-1}$$

$$(b)\ \Delta^3 f_{m+\frac{1}{2}} = f_{m+2} - 3f_{m+1} + 3f_m - f_{m-1}$$

(a) $\delta^2 f_m = \delta f_{m+1/2} - \delta f_{m-1/2} = (f_{m+1} - f_m) - (f_m - f_{m-1})$

$$= f_{m+1} - 2f_m + f_{m-1}$$

(b) $\delta^3 f_{m+1/2} = \delta^2 f_{m+1} - \delta^2 f_m = \left(f_{m+2} - 2f_{m+1} + f_m\right) -$

$$\left(f_{m+1} - 2f_m + f_{m-1}\right) = f_{m+2} - 3f_{m+1} + 3f_m - f_{m-1}$$

Shift operator, E

Let $y = f(x)$ be a function of x, and let x takes the consecutive values x, $x + h$, $x + 2h$, etc. We then define an operator E, called **the shift operator** having the property

$$E f(x) = f(x + h) \qquad \ldots(1)$$

Thus, when E operates on $f(x)$, the result is the next value of the function. If we apply the operator twice on $f(x)$, we get

$$E^2 f(x) = E[E f(x)] = f(x + 2h).$$

Thus, in general, if we apply the shift operator n times on $f(x)$, we arrive at

$$E^n f(x) = f(x + nh) \qquad \ldots(2)$$

for all real values of n.

If $f_0\ (= y_0), f_1\ (= y_1)\ldots$ are the consecutive values of the function

$y = f(x)$, then we can also write

$$E f_0 = f_1 \text{ (or } E y_0 = y_1), \quad E f_1 = f_2 \text{ (or } E y_1 = y_2)\ldots$$

$$E^2 f_0 = f_2 \text{ (or } E^2 y_0 = y_2), \quad E^2 f_1 = f_3 \text{ (or } E y_1 = y_3)\ldots$$

$$E^3 f_0 = f_3 \text{ (or } E^2 y_0 = y_3), \quad E^3 f_1 = f_4 \text{ (or } E y_1 = y_4)\ldots$$

and so on. The **inverse operator** E^{-1} is defined as:

$$E^{J1} f(x) = f(x \text{ J } h) \qquad \ldots(3)$$

and similarly

$$E^{Jn} f(x) = f(x \text{ J } nh) \qquad \ldots(4)$$

Average Operator «

The **average operator** « is defined as

$$« f(x) = \frac{1}{2}\left[f(x+\tfrac{h}{2}) + f(x-\tfrac{h}{2}) \right]$$

Differential operator D

The **differential operator** D has the property

$$Df(x) = \frac{d}{dx} f(x) = f'(x)$$

$$D^2 f(x) = \frac{d^2}{dx^2} f(x) = f''(x)$$

Relations between the operators:

Operators $\Delta, \nabla, \delta, «$ **and** D **in terms of** E

From the definition of operators Δ and E, we have

$$\Delta f(x) = f(x+h) - f(x) = E f(x) - f(x) = (E \text{ J } 1) f(x).$$

Therefore,

$$\Delta = E \text{ J } 1$$

From the definition of operators ∇ and E^{J1}, we have

$$\nabla f(x) = f(x) - f(x \text{ J } h) = f(x) - E^{J1} f(x) = (1 \text{ J } E^{J1}) f(x).$$

Therefore,

$$\nabla = 1 - E^{-1} = \frac{E-1}{E}.$$

The definition of the operators δ and E gives

$$\delta f(x) = f(x + h/2) - f(x \text{ J } h/2) = E^{1/2} f(x) - E^{J1/2} f(x)$$

$$= (E^{1/2} - E^{J1/2}) f(x).$$

Therefore,

$$\delta = E^{1/2} - E^{-1/2}$$

The definition of the operators « and E yields

$$\mu f(x) = \frac{1}{2}\left[f\left(x + \frac{h}{2}\right) + f\left(x - \frac{h}{2}\right)\right] = \frac{1}{2}\left[E^{1/2} + E^{-1/2}\right] f(x).$$

Therefore,

$$\mu = \frac{1}{2}\left(E^{1/2} + E^{-1/2}\right).$$

It is known that

$$E f(x) = f(x + h).$$

Using the Taylor series expansion, we have

$$Ef(x) = f(x) + h\ f'(x) + \frac{h^2}{2!} f''(x) + \ldots$$

$$= f(x) + h\, Df(x) + \frac{h^2}{2!} D^2(x) + \ldots$$

$$= \left(1 + \frac{h\, D}{1!} + \frac{h^2 D^2}{2!} + \ldots\right) f(x) = e^{hD} f(x).$$

Thus $E = e^{hD}$. Or,

$$hD = \log E.$$

Example If Δ, ∇, δ denote forward, backward and central difference operators, E and « respectively the shift operator and average operators, in the analysis of data with equal spacing h, prove the following:

$(i)\ 1 + \delta^2 «^2 = \left(1 + \frac{\delta^2}{2}\right)^2$ $\qquad$ $(ii)\ E^{1/2} = « + \frac{\delta}{2}$

$(iii)\ \Delta = \frac{\delta^2}{2} + \delta\sqrt{1 + \left(\delta^2/4\right)}$

$(iv)\ \mu\delta = \frac{\Delta E^{-1}}{2} + \frac{\Delta}{2}$ $\qquad$ $(v)\ \mu\delta = \frac{\Delta + \nabla}{2}.$

Solution

(i) From the definition of operators, we have

$$\mu\delta = \frac{1}{2}\left(E^{1/2} + E^{-1/2}\right)\left(E^{1/2} - E^{-1/2}\right) = \frac{1}{2}\left(E - E^{-1}\right).$$

Therefore

$$1 + \mu^2\delta^2 = 1 + \frac{1}{4}\left(E^2 - 2 + E^{-2}\right) = \frac{1}{4}\left(E + E^{-1}\right)^2$$

Also,

$$1 + \frac{\delta^2}{2} = 1 + \frac{1}{2}\left(E^{1/2} - E^{-1/2}\right)^2 = \frac{1}{2}\left(E + E^{-1}\right)$$

From equations (1) and (2), we get

$$1 + \delta^2\mu^2 = \left(1 + \frac{\delta^2}{2}\right)^2.$$

(*ii*) $\mu + \frac{\delta}{2} = \frac{1}{2}\left(E^{1/2} + E^{-1/2} + E^{1/2} - E^{-1/2}\right) = E^{1/2}.$

(*iii*) We can write

$$\frac{\delta^2}{2} + \delta\sqrt{1 + \left(\delta^2/4\right)} = \frac{\left(E^{1/2} - E^{-1/2}\right)^2}{2} + \left(E^{1/2} - E^{-1/2}\right)\sqrt{1 + \frac{1}{4}\left(E^{1/2} - E^{-1/2}\right)^2}$$

$$= \frac{E - 2 + E^{-1}}{2} + \frac{1}{2}\left(E^{1/2} - E^{-1/2}\right)\left(E^{1/2} + E^{-1/2}\right)$$

$$= \frac{E - 2 + E^{-1}}{2} + \frac{E - E^{-1}}{2}$$

$$= E - 1$$

$$= \Delta$$

(*iv*) We write

$$\mu\delta = \frac{1}{2}\left(E^{1/2} + E^{-1/2}\right)\left(E^{1/2} - E^{-1/2}\right) = \frac{1}{2}\left(E - E^{-1}\right)$$

$$= \frac{1}{2}\left(1 + \Delta - E^{-1}\right) = \frac{\Delta}{2} + \frac{1}{2}\left(1 - E^{-1}\right) = \frac{\Delta}{2} + \frac{1}{2}\left(\frac{E - 1}{E}\right) = \frac{\Delta}{2} + \frac{\Delta}{2E}.$$

(*v*) We can write

$$\mu\delta = \frac{1}{2}\left(E^{1/2} + E^{-1/2}\right)\left(E^{1/2} - E^{-1/2}\right) = \frac{1}{2}\left(E - E^{-1}\right)$$

$$= \frac{1}{2}\left(1 + \Delta - (1 - \nabla)\right) = \frac{1}{2}(\Delta + \nabla).$$

Example Prove that

$$hD = \log(1+\Delta) = -\log(1-\nabla) = \sinh^{-1}(\mu\delta).$$

Using the standard relations given in boxes in the last section, we have

$$hD = \log E = \log(1+\Delta) = \log E = -\log E^{-1} = -\log(1+\nabla)$$

Also,

$$\mu\delta = \frac{1}{2}\left(E^{1/2} + E^{-1/2}\right)\left(E^{1/2} - E^{-1/2}\right) = \frac{1}{2}\left(E + E^{-1}\right)$$

$$= \frac{1}{2}\left(e^{hD} - e^{-hD}\right) = \sin(hD)$$

Therefore

$$hD = \sinh^{-1}(\mu\delta).$$

Example Show that the operations « and E commute.

Solution

From the definition of operators « and E, we have

$$\mu E f_0 = \mu f_1 = \frac{1}{2}\left(f_{3/2} + f_{1/2}\right)$$

and also

$$E\mu f_0 = \frac{1}{2}E\left(f_{1/2} + f_{-1/2}\right) = \frac{1}{2}\left(f_{3/2} + f_{1/2}\right)$$

Hence

$$\mu E = E\mu.$$

Therefore, the operators « and E commute.

Example Show that

$$e^x\left(u_0 + x\Delta u_0 + \frac{x^2}{2!}\Delta^2 u_0 + \ldots\right) = u_0 + u_1 x + u_2\frac{x^2}{2!} + \ldots$$

$$e^x\left(u_0 + x\Delta u_0 + \frac{x^2}{2!}\Delta^2 u_0 + \ldots\right) = e^x\left(1 + x\Delta + \frac{x^2\Delta^2}{2!} + \ldots\right)u_0$$

$$= e^x e^{x\Delta}u_0 = e^{x(1+\Delta)}u_0$$

$$= e^{xE}u_0$$

$$= \left(1 + xE + \frac{x^2E^2}{2!} + \ldots\right)u_0$$

$$= u_0 + xu_1 + \frac{x^2}{2!}u_2 + \ldots,$$

as desired.

Example Using the method of separation of symbols, show that

$$\Delta^n u_{x-n} = u_x - nu_{x-1} + \frac{n(n-1)}{2}u_{x-2} + \cdots + (-1)^n u_{x-n}.$$

To prove this result, we start with the right-hand side. Thus,

$$\text{R.H.S} \quad = u_x - nu_{x-1} + \frac{n(n-1)}{2}u_{x-2} + \cdots + (-1)^n u_{x-n}.$$

$$= u_x - nE^{-1}u_x + \frac{n(n-1)}{2}E^{-2}u_x + \cdots + (-1)^n E^{-n}u_x$$

$$= \left[1 - nE^{-1} + \frac{n(n-1)}{2}E^{-2} + \cdots + (-1)^n E^{-n}\right]u_x$$

$$= \left(1 - E^{-1}\right)^n u_x$$

$$= \left(1 - \frac{1}{E}\right)^n u_x$$

$$= \left(\frac{E-1}{E}\right)^n u_x$$

$$= \frac{\Delta^n}{E^n}u_x$$

$$= \Delta^n E^{-n}u_x$$

$$= \Delta^n u_{x-n},$$

$$= \text{L.H.S}$$

Differences of a Polynomial

Let us consider the polynomial of degree n in the form

$$f(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n,$$

where $a_0 \neq 0$ and $a_0, a_1, a_2, \ldots, a_{n-1}, a_n$ are constants. Let h be the interval of differencing. Then

$$f(x+h)=a_0(x+h)^n+a_1(x+h)^{n-1}+a_2(x+h)^{n-2}+\ldots\quad+a_{n-1}(x+h)+a_n$$

Now the difference of the polynomials is:

$$\Delta f(x)=f(x+h)-f(x)=a_0\left[(x+h)^n-x^n\right]+a_1\left[(x+h)^{n-1}-x^{n-1}\right]+\ldots\qquad+a_{n-1}(x+h-x)$$

Binomial expansion yields

$$\begin{aligned}\Delta f(x) &= a_0\left[x^n+{}^nC_1x^{n-1}h+{}^nC_2x^{n-2}h^2+\ldots+h^n-x^n\right]\\ &\quad+a_1[x^{n-1}+{}^{(n-1)}C_1x^{n-2}h+{}^{(n-1)}C_2x^{n-3}h^2\\ &\quad+\ldots+h^{n-1}-x^{n-1}]+\ldots+a_{n-1}h\\ &= a_0nhx^{n-1}+\left[a_0{}^nC_2h^2+a_1{}^{(n-1)}C_1h\right]x^{n-2}+\ldots+a_{n-1}h.\end{aligned}$$

Therefore,

$$\Delta f(x) = a_0nhx^{n-1}+b'x^{n-2}+c'x^{n-3}+\ldots+k'x+l',$$

where bà, cà, $\ldots$, kà, là are constants involving h but not x. Thus, the first difference of a polynomial of degree n is another polynomial of degree (n J 1). Similarly,

$$\begin{aligned}\Delta^2 f(x)=\Delta(\Delta f(x)) &= \Delta f(x+h)-\Delta f(x)\\ &= a_0nh\left[(x+h)^{n-1}-x^{n-1}\right]+b'\left[(x+h)^{n-2}-x^{n-2}\right]\\ &\quad+\ldots+k'(x+h-x)\end{aligned}$$

On simplification, it reduces to the form

$$\Delta^2 f(x)=a_0n(n-1)h^2x^{n-2}+b''x^{n-3}+c''x^{n-4}+\ldots+q''.$$

Therefore, $\Delta^2 f(x)$ is a polynomial of degree (n J 2) in x. Similarly, we can form the higher order differences, and every time we observe that the degree of the polynomial is reduced by 1. After differencing n times, we are left with only the first term in form

$$\begin{aligned}\Delta^n f(x) &= a_0n(n-1)(n-2)(n-3)\ldots(2)(1)h^n\\ &= a_0(n!)h^n=\text{constant}.\end{aligned}$$

This constant is independent of x. Since $\Delta^n f(x)$ is a constant $\Delta^{n+1}f(x)=0$. Hence the (n + 1)th and higher order differences of a polynomial of degree n are 0.

Conversely, if the nth differences of a tabulated function are constant and the $(n+1)$th, $(n+2)$th,..., differences all vanish, then the tabulated function represents a polynomial of degree n. It should be noted that these results hold good only if the values of x are equally spaced. The converse is important in numerical analysis since it enables us to approximate a function by a polynomial if its differences of some order become nearly constant.

Theorem (*Differences of a polynomial*)The nth differences of a polynomial of degree n is a constant, when the values of the independent variable are given at equal intervals.

Exercises

1. Calculate $f(x)=\frac{1}{x+1}$, $x=0(0.2)1$ to (a) 2 decimal places, (b) 3 decimal places and (c)4 decimal places. Then compare the effect of rounding errors in the corresponding difference tables.
2. Express $\Delta^2 y_1$ (i.e. $\Delta^2 f_1$) and $\Delta^4 y_0$ (i.e. $\Delta^4 f_0$) in terms of the values of the function $y = f(x)$.
3. Set up a difference table of $f(x)=x^2$ for $x=0(1)10$. Do the same with the calculated value 25 of $f(5)$ replaced by 26. Observe the spread of the error.
4. Calculate $f(x)=\frac{1}{x+1}$, $x=0(0.2)1$ to (a)2 decimal places, (b)3 decimal places and (c)4 decimal places. Then compare the effect of rounding errors in the corresponding difference tables.
5. Set up a forward difference table of $f(x) = x^2$ for $x = 0(1)10$. Do the same with the calculated value 25 of $f(5)$ replaced by 26. Observe the spread of the error.
6. Construct the difference table based on the following table.

x	0.0	0.1	0.2	0.3	0.4	0.5
cos$x$	1.000 00	0.995 00	0.980 07	0.955 34	0.921 06	0.877 58

7. Construct the difference table based on the following table.

x	0.0	0.1	0.2	0.3	0.4	0.5
sin$x$	0.000 00	0.099 83	0.198 67	0.295 52	0.389 42	0.479

8. Construct the backward difference table, where

$$f(x)=\sin x,\ x = 1.0(0.1)1.5,\ 4D.$$

9. Show that $E\nabla = \Delta = \delta E^{1/2}$.

10. Prove that

11. $(i)\ \delta = 2\sinh(hD/2)$ and $(ii)\ \mu = 2\cosh(hD/2)$.

12. Show that the operators δ, « , E, Δ and ∇ commute with each other.

13. Construct the backward difference table based on the following table.

x	0.0	0.1	0.2	0.3	0.4	0.5
cos x	1.000 00	0.995 00	0.980 07	0.955 34	0.921 06	0.877 58

Construct the difference table based on the following table.

x	0.0	0.1	0.2	0.3	0.4	0.5
sin x	0.000 00	0.099 83	0.198 67	0.295 52	0.389 42	0.479 43

6. Construct the backward difference table, where

$f(x) = \sin x,\ \ x = 1.0(0.1)1.5$, 4D.

7. Evaluate $(2a + 3)(E + 2)(3x^2 + 2)$, interval of differencing being unity.

8. Compute the missing values of y_n and Δy_n in the following table:

y_n	Δy_n	$\Delta^2 y_n$
-		
-	-	1
-	-	4
6	5	13
-	-	18
-	-	24
-	-	

NUMERICAL INTERPOLATION

Consider a single valued continuous function $y = f(x)$ defined over [a,b] where $f(x)$ is known explicitly. It is easy to find the values of '*y*' for a given set of values of '*x*' in [a,b]. i.e., it is possible to get information of all the points (x, y) where $a \le x \le b$.

But the converse is not so easy. That is, using only the points $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$ where $a \le x_i \le b, i = 0,1,2,...,n$, it is not so easy to find the relation between x and y in the form $y = f(x)$ explicitly. That is one of the problem we face in numerical differentiation or integration.

Now we have first to find a simpler function, say $g(x)$, such that $f(x)$ and $g(x)$ agree at the given set of points and accept the value of $g(x)$ as the required value of $f(x)$ at some point x in between a and b. Such a process is called **interpolation.** If $g(x)$ is a polynomial, then the process is called polynomial interpolation.

When a function *f*(*x*) is not given explicitly and only values of $f(x)$ are given at a set of distinct points called *nodes* or *tabular points*, using the interpolated function $g(x)$ to the function *f*(*x*), the required operations intended for $f(x)$, like determination of roots, differentiation and integration etc. can be carried out. The approximating polynomial $g(x)$ can be used to predict the value of $f(x)$ at a non- tabular point. The deviation of $g(x)$ from $f(x)$, that is $|f(x) - g(x)|$ is called the *error of approximation*.

Consider a continuous single valued function $f(x)$ defined on an interval [*a, b*]. Given the values of the function for n + 1 distinct tabular points $x_0, x_1, ..., x_n$ such that $a \le x_0 \le x_1 \le ... \le x_n \le b$. The problem of polynomial interpolation is to find a polynomial *g(x)* *or* $p_n(x)$, of degree *n*, which fits the given data. The interpolation polynomial fitted to a given data is unique.

If we are given two points satisfying the function such as $(x_0, y_0); (x_1, y_1)$, where $y_0 = f(x_0)$ and $y_1 = f(x_1)$ it is possible to fit a unique polynomial of degree 1. If three distinct points are given, a polynomial of degree not greater than two can be fitted uniquely. In general, if n+ *1 distinct points* are given, a polynomial of degree not greater than *n* can be fitted uniquely.

Interpolation fits a real function to discrete data. Given the set of tabular values $(x_0, y_0), (x_1, y_1) \ldots . (x_n, y_n)$ satisfying the relation $y = f(x)$, where the explicit nature of

$f(x)$ is not known, and it is required to find the values of $f(x)$ corresponding to certain given values of x in between x_0 and x_n. To do this we have first to find a simpler function, say $g(x)$, such that $f(x)$ and $g(x)$ agree at the set of tabulated points and accept the value of $g(x)$ as the required value of $f(x)$ at some point x in between x_0 and x_n. Such a process is called **interpolation**. If $g(x)$ is a polynomial, then the process is called **polynomial interpolation**.

In interpolation, we have to determine the function $g(x)$, in the case that $f(x)$ is difficult to be obtained, using the **pivotal values** $f_0 = f(x_0),\ f_1 = f(x_1), \ldots,\ f_n = f(x_n)$.

Linear interpolation

In linear interpolation, we are given with two pivotal values $f_0 = f(x_0)$ and $f_1 = f(x_1)$, and we approximate the curve of f by a chord (straight line) P_1 passing through the points $(x_0,\ f_0)$ and $(x_1,\ f_1)$. Hence the approximate value of f at the intermediate point $x = x_0 + rh$ is given by the **linear interpolation formula**

$$f(x) \approx P_1(x) = f_0 + r(f_1 - f_0) = f_0 + r\Delta f_0$$

where $r = \dfrac{x - x_0}{h}$ and $0 \le r \le 1$.

Example Evaluate $\ln 9.2$, given that $\ln 9.0 = 2.197$ and $\ln 9.5 = 2.251$.

Here $x_0 = 9.0$, $x_1 = 9.5$, $h = x_1 - x_0 = 9.5 - 9.0 = 0.5$, $f_0 = f(x_0) = \ln 9.0 = 2.197$ and $f_1 = f(x_1) = \ln 9.5 = 2.251$. Now to calculate $\ln 9.2 = f(9.2)$, take $x = 9.2$, so that

$$r = \frac{x - x_0}{h} = \frac{9.2 - 9.0}{0.5} = \frac{0.2}{0.5} = 0.4 \text{ and hence}$$

$$\ln 9.2 = f(9.2) \approx P_1(9.2) = f_0 + r(f_1 - f_0) = 2.197 + 0.4\,(2.251 - 2.197) = 2.219$$

Example Evaluate $f(15)$, given that $f(10) = 46$, $f(20) = 66$.

Here $x_0 = 10$, $x_1 = 20$, $h = x_1 - x_0 = 20 - 10 = 10$,

$f_0 = f(x_0) = 46$ and $f_1 = f(x_1) = 66$.

Now to calculate $f(15)$, take $x = 15$, so that

$$r = \frac{x - x_0}{h} = \frac{15 - 10}{10} = \frac{5}{10} = 0.5$$

and hence $f(15) \approx P_1(15) = f_0 + r(f_1 - f_0) = 46 + 0.5\,(66 - 46) = 56$

Example Evaluate $e^{1.24}$, given that $e^{1.1} = 3.0042$ and $e^{1.4} = 4.0552$.

Here $x_0 = 1.1$, $x_1 = 1.4$, $h = x_1 - x_0 = 1.4 - 1.1 = 0.3$, $f_0 = f(x_0) = 1.1$ and $f_1 = f(x_1) = 1.24$. Now to calculate $e^{1.24} = f(1.24)$, take $x = 1.24$, so that $r = \frac{x - x_0}{h} = \frac{1.24 - 1.1}{0.3} = \frac{0.14}{0.3} = 0.4667$ and hence

$e^{1.24} \approx P_1(1.24) = f_0 + r(f_1 - f_0) = 3.0042 + 0.4667(4.0552 - 3.0042) = 3.4933$, while the exact value of $e^{1.24}$ is 3.4947.

Quadratic Interpolation

In quadratic interpolation we are given with three pivotal values $f_0 = f(x_0)$, $f_1 = f(x_1)$ and $f_2 = f(x_2)$ and we approximate the curve of the function f between x_0 and $x_2 = x_0 + 2h$ by the quadratic parabola P_2 , which passes through the points $(x_0, f_0), (x_1, f_1), (x_2, f_2)$ and obtain the quadratic interpolation formula

$$f(x) \approx P_2(x) = f_0 + r\Delta f_0 + \frac{r(r-1)}{2}\Delta^2 f_0$$

where $r = \frac{x - x_0}{h}$ and $0 \le r \le 2$.

Example Evaluate ln 9.2, using quadratic interpolation, given that

ln 9.0 = 2.197, ln 9.5 = 2.251 and ln10.0 = 2.3026.

Here $x_0 = 9.0$, $x_1 = 9.5$, $x_1 = 10.0$, $h = x_1 - x_0 = 9.5 - 9.0 = 0.5$, $f_0 = f(x_0) = \ln 9.0 = 2.197$, $f_1 = f(x_1) = \ln 9.5 = 2.251$ and $f_2 = f(x_2) = \ln 10.0 = 2.3026$. Now to calculate $\ln 9.2 = f(9.2)$, take $x = 9.2$, so that $r = \frac{x - x_0}{h} = \frac{9.2 - 9.0}{0.5} = \frac{0.2}{0.5} = 0.4$ and

$$\ln 9.2 = f(9.2) \approx P_2(x) = f_0 + r\Delta f_0 + \frac{r(r-1)}{2}\Delta^2 f_0$$

To proceed further, we have to construct the following forward difference table.

x	f	Δf	$\Delta^2 f$
9.0	2.1972		
9.5	2.2513	0.0541	-0.0028
10.0	2.3026	0.0513	

Hence,

$\ln 9.2 = f(9.2) \approx P_2(9.2) = 2.1972 + 0.4(0.0541) + \frac{0.4(0.4-1)}{2}(-0.0028)$ = 2.2192, which exact to 4D to the exact value of $\ln 9.2 = 2.2192$.

Example Using the values given in the following table, find cos0.28 by linear interpolation and by quadratic interpolation and compare the results with the value 0.96106 (exact to 5D)

x	$f(x) = \cos x$	First difference	Second difference
0.0	1.00000		
		-0.01993	
0.2	0.98007		-0.03908
		-0.05901	
0.4	0.92106		

Here $f(x)$, where $x_0 = 0.28$ is to determined. In linear interpolation, we need two consecutive x values and their corresponding f values and first difference. Here, since x=0.28 lies in between 0.2 and 0.4, we take x_0 = 0.2, x_1 = 0.4. (**Attention!** Choosing $x_0 = 0.2$, $x_1 = 0.4$ is very important; taking $x_0 = 0.0$ would give wrong answer). Then $h = x_1 - x_0$ =0.4–0.2 = 0.2, $f_0 = f(x_0)$ =0.98007 and $f_1 = f(x_1)$ =0.92106.

Also $r = \frac{x - x_0}{h} = \frac{0.28 - 0.2}{0.2} = \frac{0.08}{0.2} = 0.4$ and

$$\cos 0.28 = f(0.28) \approx P_1(0.28) = f_0 + r(f_1 - f_0)$$

$$= 0.98007 + 0.4(0.92106 - 0.98007)$$

= 0.95647, correct to 5 D.

In quadratic interpolation, we need three consecutive (equally spaced) x values and their corresponding f values, first differences and second difference. Here x_0 = 0.0 , x_1 = 0.2, x_1 = 0.4, $h = x_1 - x_0$ =0.2 – 0.0 = 02, f_0 = 1.00000, f_1 = 0.98007 and f_2 = 0.92106, Δf_0=-0.01993, $\Delta^2 f_0$=-0.03908 $r = \frac{x - x_0}{h} = \frac{0.28 - 0.00}{0.2} = 1.4$ and

$$\cos 0.28 \approx P_2(0.28) = f_0 + r\Delta f_0 + \frac{r(r-1)}{2}\Delta^2 f_0$$

$$= 1.00 + 1.4(-0.-1993) + \frac{1.4(1.4-1)}{2}(-0.03908) = 0.96116 \text{ to 5D.}$$

From the above, it can be seen that quadratic interpolation gives more accurate value.

Newton's Forward Difference Interpolation Formula

Using Newton's forward difference interpolation formula we find the n degree polynomial P_n which approximates the function $f(x)$ in such a way that P_n and f agrees at n+1 equally spaced x values, so that $P_n(x_0) = f_0, P_n(x_1) = f_1, \ldots, P_n(x_n) = f_n$, where $f_0 = f(x_0)$, $f_1 = f(x_1)$, …, $f_n = f(x_n)$ are the values of f in the table.

Newton's forward difference interpolation formula is

$$f(x) \approx P_n(x) =$$

$$= f_0 + r\Delta f_0 + \frac{r(r-1)}{2!}\Delta^2 f_0 + \ldots + \frac{r(r-1)\ldots(r-n+1)}{n!}\Delta^n f_0$$

where $x = x_0 + rh,\ r = \frac{x - x_0}{h},\ 0 \le r \le n$.

Derivation of Newton's forward Formulae for Interpolation

Given the set of $(n+1)$ values, viz., $(x_0, f_0), (x_1, f_1), (x_2, f_2), \ldots, (x_n, f_n)$

of x and f, it is required to find $p_n(x)$, a polynomial of the nth degree such that $f(x)$ and $p_n(x)$ agree at the tabulated points. Let the values of x be equidistant, i.e., let

$$x_i = x_0 + rh, \qquad r = 0, 1, 2, \ldots, n$$

Since $p_n(x)$ is a polynomial of the nth degree, it may be written as

$$\left.\begin{aligned} p_n(x) = a_0 &+ a_1(x - x_0) + a_2(x - x_0)(x - x_1) \\ &+ a_3(x - x_0)(x - x_1)(x - x_2) + \ldots \\ &+ a_n(x - x_0)(x - x_1)(x - x_2)\ldots(x - x_{n-1}) \end{aligned}\right\}$$

Imposing now the condition that $f(x)$ and $p_n(x)$ should agree at the set of tabulated points, we obtain

$$a_0 = f_0;\ a_1 = \frac{f_1 - f_0}{x_1 - x_0} = \frac{\Delta f_0}{h};\ a_2 = \frac{\Delta^2 f_0}{h^2 2!};\ a_3 = \frac{\Delta^3 f_0}{h^3 3!};\ldots;\ a_n = \frac{\Delta^n f_0}{h^n n!};$$

Setting $x = x_0 + rh$ and substituting for $a_0, a_1, \ldots, a_n$, we obtain the expression.

Remark 1:

Newton's forward difference formula has the permanence property. If we add a new set of value (x_{n+1}, y_{n+1}), to the given set of values, then the forward difference table gets a new column of $(n+1)^{th}$ forward difference. Then the Newton's Forward difference

Interpolation Formula with the already given values will be added with a new term at the end, $(x-x_0)(x-x_1).....(x-x_n)\frac{1}{(n+1)!h^{n+1}}\left[\Delta^{n+1}y_0\right]$ to get the new interpolation formula with the newly added value.

Remark 2:

Newton's forward difference interpolation formula is useful for interpolation near the beginning of a set of tabular values and for extrapolating values of y a short distance backward, that is left from y_0.The process of finding the value of y for some value of x outside the given range is called *extrapolation.*

Example Using Newton's forward difference interpolation formula and the following table evaluate $f(15)$.

x	$f(x)$	Δf	$\Delta^2 f$	$\Delta^3 f$	$\Delta^4 f$
10	46				
		20			
20	66		-5	2	
		15			
30	81		-3	-1	-3
		12			
40	93		-4		
		8			
50	101				

Here $x = 15$, $x_0 = 10$, $x_1 = 20$, $h = x_1 - x_0 = 20 - 10 = 10$, $r = (x - x_0)/h = (15-10)/10 = 0.5$, $f_0 = -46$, $\Delta f_0 = 20$, $\Delta^2 f_0 = -5$, $\Delta^3 f_0 = 2$, $\Delta^4 f_0 = -3$.

Substituting these values in the Newton's forward difference interpolation formula for $n = 4$, we obtain

$$f(x) \approx P_4(x) = f_0 + r\Delta f_0 + \frac{r(r-1)}{2!}\Delta^2 f_0 + \ldots + \frac{r(r-1)\ldots(r-4+1)}{4!}\Delta^4 f_0,$$

so that

$$f(15) \approx 46 + (0.5)(20) + \frac{(0.5)(0.5-1)}{2!}(-5) + \frac{(0.5)(0.5-1)(0.5-2)}{3!}(2)$$

$$+\frac{(0.5(0.5-1)(0.5-2)(0.5-3)}{4!}(-3)$$

= 56.8672, correct to 4 decimal places.

Example Find a cubic polynomial in x which takes on the values -3, 3, 11, 27, 57 and 107, when x=0, 1, 2, 3, 4 and 5 respectively.

x	$f(x)$	Δ	Δ^2	Δ^3
0	-3			
		6		
1	3		2	
		8		6
2	11		8	
		16		6
3	27		14	
		30		6
4	57		20	
		50		
5	107			

Now the required cubic polynomial (polynomial of degree 3) is obtained from Newton's forward difference interpolation formula

$$f(x) \approx P_3(x) = f_0 + r\Delta f_0 + \frac{r(r-1)}{2!}\Delta^2 f_0 + \frac{r(r-1)(r-3+1)}{3!}\Delta^3 f_0,$$

where $r=(x - x_0)/h = (x - 0)/1 = x$, so that

$$f(x) \approx P_3(x) = -3 + x(6) + \frac{x(x-1)}{2!}(2) + \frac{x(x-1)(x-3+1)}{3!}(6)$$

or $f(x) = x^3 - 2x^2 + 7x - 3$

Example Using the Newton's forward difference interpolation formula evaluate $f(2.05)$ where $f(x) = \sqrt{x}$, using the values:

x	2.0	2.1	2.2	2.3	2.4
$\sqrt{x}$	1.414 214	1.449 138	1.483 240	1.516 575	1.549 193

The forward difference table is

x	$\sqrt{x}$	Δ	Δ^2	Δ^3	Δ^4
2.0	1.414 214				
		0.034 924			
2.1	1.449 138		-0.000 822		
		0.034 102		0.000055	
2.2	1.483 240		-0.000 767		–0.000 005
		0.033 335		0.000050	
2.3	1.516 575		-0.000 717		
		0.032 618			
2.4	1.549 193				

Here $r = \frac{x - x_0}{h}$ =(2.05–2.00)/0.1=0.5, so by substituting the values in Newton's formula (for 4 degree polynomial), we get

$$f(2.05) \approx P_4(2.05) = 1.414214 + (0.5)(0.034924) + \frac{(0.5)(0.5-1)}{2!}(-0.000822)$$

$$+ \frac{(0.5)(0.5-1)(0.5-2)}{3!}(0.000055)$$

$$+ \frac{(0.5(0.5-1)(0.5-2)(0.5-3)}{4!}(0.000005) = 1.431783.$$

Example Find the cubic polynomial which takes the following values; $f(1) = 24, f(3) = 120, f(5) = 336,$ and $f(7) = 720$. Hence, or otherwise, obtain the value of $f(8)$.

We form the difference table:

x	y	Δ	Δ^2	Δ^3
1	24			
		96		
3	120		120	
		216		48
5	336		168	
		384		
7	720			

Here $h = 2$ with $x_0 = 1$, we have $x = 1 + 2p$ or $r = (x-1)/2$. Substituting this value of r, we obtain

$$f(x) = 24 + \frac{x-1}{2}(96) + \frac{\left(\frac{x-1}{2}\right)\left(\frac{x-1}{2} - 1\right)}{2}(120)$$

$$+\frac{\left(\frac{x-1}{2}\right)\left(\frac{x-1}{2}-1\right)\left(\frac{x-1}{2}-2\right)}{6}(48) = x^3+6x^2+11x+6.$$

To determine $f(9)$, we put $x=9$ in the above and obtain $f(9)=1320.$

With $x_0=1,\ x_r=9,$ and $h=2,$ we have $r=\frac{x_r-x_0}{h}=\frac{9-1}{2}=4$. Hence

$$f(9)\approx p(9)=f_0+r\Delta f_0+\frac{r(r-1)}{2!}\Delta^2 f_0+\frac{r(r-1)\,(r-2)}{3!}\Delta^3 f_0$$

$$=24+4\times 96+\frac{4\times 3}{2}\times 120+\frac{4\times 3\times 2}{3\times 2}\times 48=1320$$

Example Using Newton's forward difference formula, find the sum

$$S_n=1^3+2^3+3^3+\ldots+n^3.$$

Solution

$$S_{n+1}=1^3+2^3+3^3+\ldots+n^3+(n+1)^3$$

and hence

$$S_{n+1}-S_n=(n+1)^3,$$

or

$$\Delta S_n=(n+1)^3.$$

it follows that

$$\Delta^2 S_n=\Delta S_{n+1}-\Delta S_n=(n+2)^3-(n+1)^3=3n^2+9n+7$$

$$\Delta^3 S_n=3(n+1)+9n+7-(3n^2+9n+7)=6n+12$$

$$\Delta^4 S_n=6(n+1)+12-(6n+12)=6$$

Since $\Delta^5 S_n=\Delta^6 S_n=\ldots=0, S_n$ is a fourth-degree polynomial in the variable n.

Also,

$$S_1=1,\qquad \Delta S_1=(1+1)^3=8,\qquad \Delta^2 S_1=3+9+7=19,$$

$$\Delta^3 S_1=6+12=18,\qquad \Delta^4 S_1=8.$$

formula (3) gives (with $f_0=S_1$ and $r-n-1$)

$$S_n=1+(n-1)(8)+\frac{(n-1)(n-2)}{2}(19)+\frac{(n-1)(n-2)(n-3)}{6}(18)$$

$$+\frac{(n-1)(n-2)(n-3)(n-4)}{24}(6)$$

$$=\frac{1}{4}n^4+\frac{1}{2}n^3+\frac{1}{4}n^2$$

$$=\left[\frac{n(n+1)}{2}\right]^2$$

Problem: The population of a country for various years in millions is provided. Estimate the population for the year 1898.

Year x:	1891	1901	1911	1921	1931
Population y:	46	66	81	93	101

Solution: Here the interval of difference among the arguments *h=10.* Since 1898 is at the beginning of the table values, we use Newton's forward difference interpolation formula for finding the population of the year 1898.

The forward differences for the given values are as shown here.

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
1891	46				
		$\Delta y_0 = 20$			
1901	66		$\Delta^2 y_0 = -5$		
		$\Delta y_1 = 15$		$\Delta^3 y_0 = 2$	
1911	81		$\Delta^2 y_1 = -3$		$\Delta^4 y_0 = -3$
		$\Delta y_2 = 12$		$\Delta^3 y_1 = -1$	
1921	93		$\Delta^2 y_2 = -4$		
		$\Delta y_3 = 8$			
1931	101				

Let x=1898. Newton's forward difference interpolation formula is,

$$f(x) = y_0 + (x-x_0)\frac{1}{h}[\Delta y_0] + (x-x_0)(x-x_1)\frac{1}{2!h^2}[\Delta^2 y_0]$$
$$+(x-x_0)(x-x_1)(x-x_2)\frac{1}{3!h^3}[\Delta^3 y_0] + ... +$$
$$(x-x_0)(x-x_1).....(x-x_{n-1})\frac{1}{n!h^n}[\Delta^n y_0]$$

Now, substituting the values, we get,

$$f(1898) = 46 + (1898-1891)\frac{1}{10}[20] + (1898-1891)(1898-1901)\frac{1}{2!10^2}[-5]$$
$$+ (1898-1891)(1898-1901)(1898-1911)\frac{1}{3!10^3}[2] +$$
$$(1898-1891)(1898-1901)(1898-1911)(1898-1921)\frac{1}{4!10^4}[-3]$$

$$\Rightarrow f(1898) = 46 + 14 + \frac{21}{40} + \frac{91}{500} + \frac{18837}{40000} = 61.178$$

Example Values of x (in degrees) and $\sin x$ are given in the following table:

x (in degrees)	$\sin x$
15	0.2588190
20	0.3420201
25	0.4226183
30	0.5
35	0.5735764
40	0.6427876

Determine the value of $\sin 38^0$.

Solution

The difference table is

x	$\sin x$	Δ	Δ^2	Δ^3	Δ^4	Δ^5
15	0.2588190					
		0.0832011				
20	0.3420201		–0.0026029			
		0.0805982		–0.0006136		
25	0.4226183		–0.0032165		0.0000248	
		0.0773817		–0.0005888		0.0000041
30	0.5		–0.0038053		0.0000289	
		0.0735764		–0.0005599		
35	0.5735764		–0.0043652			
		0.0692112				
40	0.6427876					

As 38 is closer to $x_n = 40$ than $x_0 = 15$, we use Newton's backward difference formula with $x_n = 40$ and $x = 38$. This gives

$$r = \frac{x - x_n}{h} = \frac{38 - 40}{5} = -\frac{2}{5} = -0.4$$

Hence, using formula, we obtain

$$f(38) = 0.6427876 - 0.4(0.0692112) + \frac{-0.4(-0.4-1)}{2}(-0.0043652)$$
$$+ \frac{(-0.4)(-0.4+1)(-0.4+2)}{6}(-0.0005599)$$
$$+ \frac{(-0.4)(-0.4+1)(-0.4+2)(-0.4+3)}{24}(0.0000289)$$
$$+ \frac{(-0.4)(-0.4+1)(-0.4+2)(-0.4+3)(-0.4+4)}{120}(0.0000041)$$
$$= 0.6427876 - 0.02768448 + 0.00052382 + 0.00003583$$
$$-0.00000120$$
$$= 0.6156614$$

Example Find the missing term in the following table:

x	$y = f(x)$
0	1
1	3
2	9
3	–
4	81

Explain why the result differs from $3^3 = 27$?

Since four points are given, the given data can be approximated by a third degree polynomial in x. Hence $\Delta^4 f_0 = 0$. Substituting $\Delta = E - 1$ we get, $(E-1)^4 f_0 = 0$, which on simplification yields

$$E^4 f_0 - 4E^3 f_0 + 6E^2 f_0 - 4Ef_0 + f_0 = 0.$$

Since $E^r f_0 = f_r$ the above equation becomes

$$f_4 - 4f_3 + 6f_2 - 4f_1 + f_0 = 0$$

Substituting for f_0, f_1, f_2 and f_4 in the above, we obtain

$$f_3 = 31$$

By inspection it can be seen that the tabulated function is 3^x and the exact value of $f(3)$ is 27. The error is due to the fact that the exponential function 3^x is approximated by means of a polynomial in x of degree 3.

Example The table below gives the values of $\tan x$ for $0.10 \le x \le 0.30$

x	$y = \tan x$
0.10	0.1003
0.15	0.1511
0.20	0.2027
0.25	0.2553
0.30	0.3093

Find: (a) $\tan 0.12$ (b) $\tan 0.26$. (c) $\tan 0.40$ (d) $\tan 0.50$

The table difference is

x	$y = f(x)$	Δ	Δ^2	Δ^3	Δ^4
0.10	0.1003				
		0.0508			
0.15	0.1511		0.0008		
		0.0516		0.0002	
0.20	0.2027		0.0010		0.0002
		0.0526		0.0004	
0.25	0.2553		0.0014		
		0.0540			
0.30	0.3093				

a) To find $\tan(0.12)$, we have $r = 0.4$ Hence Newton's forward difference interpolation formula gives

$$\tan(0.12) = 0.1003 + 0.4(0.0508) + \frac{0.4(0.4-1)}{2}(0.0008)$$

$$+ \frac{0.4(0.4-1)(0.4-2)}{6}(0.0002)$$

$$+ \frac{0.4(0.4-1)(0.4-2)(0.4-3)}{24}(0.0002)$$

$$= 0.1205$$

b) To find $\tan(0.26)$, we use Newton's backward difference interpolation formula with

$$r = \frac{x - x_n}{n}$$

$$= \frac{0.26 - 03}{0.05}$$

$$= -0.8$$

which gives

$$\tan(0.26) = 0.3093 - 0.8(0.0540) + \frac{-0.8(-0.8+1)}{2}(0.0014)$$

$$+ \frac{-0.8(-0.8+1)(-0.8+2)}{6}(0.0004)$$

$$+ \frac{-0.8(-0.8+1)(-0.8+2)(-0.8+3}{24}(0.0002) \quad = 0.2662$$

Proceeding as in the case (i) above, we obtain

(c) $\tan 0.40 = 0.4241,$ and

(d) $\tan 0.50 = 0.5543$

The actual values, correct to four decimal places, of tan (0.12), tan(0.26) are respectively 0.1206 and 0.2660. Comparison of the computed and actual values shows that in the first two cases (i.e., of interpolation) the results obtained are fairly accurate whereas in the last-two cases (i.e., of extrapolation) the errors are quite considerable. The example therefore demonstrates the important results that if a tabulated function is other than a polynomial, then extrapolation very far from the table limits would be dangerous-although interpolation can be carried out very accurately.

Exercises

1. Using the difference table in exercise 1, compute cos0.75 by Newton's forward difference interpolating formula with $n = 1, 2, 3, 4$ and compare with the 5D-value 0.731 69.

2. Using the difference table in exercise 1, compute cos0.28 by Newton's forward difference interpolating formula with $n = 1, 2, 3, 4$ and compare with the 5D-value

3. Using the values given in the table, find cos0.28 (in radian measure) by linear interpolation and by quadratic interpolation and compare the results with the value 0.961 06 (exact to 5D).

x	$f(x)=\cos x$	First difference	Second difference
0.0	1.000 00		
		-0.019 93	
0.2	0.980 07		-0.03908
		-0.059 01	
0.4	0.921 06		-0.03671
		-0.095 72	
0.6	0.825 34		-0.03291
		-0.128 63	
0.8	0.696 71		-0.02778
		-0.156 41	
1.0	0.540 30		

4. Find Lagrangian interpolation polynomial for the function f having $f(4)=1,\ f(6)=3,\ f(8)=8,\ f(10)=16$. Also calculate $f(7)$.

5. The sales in a particular shop for the last ten years is given in the table:

Year	1996	1998	2000	2002	2004
Sales (in lakhs)	40	43	48	52	57

Estimate the sales for the year 2001 using Newton's backward difference interpolating formula.

6. Find $f(3)$, using Lagrangian interpolation formula for the function f having $f(1)=2,\ f(2)=11,\ f(4)=77$.

7. Find the cubic polynomial which takes the following values:

x	0	1	2	3
$f(x)$		1	2	1	10

8. Compute sin0.3 and sin0.5 by Everett formula and the following table.

	$\sin x$	δ^2
0.2	0.198 67	-0.007 92
0.4	0.389 42	-0.015 53
.6	0.564 64	-0.022 50

9. The following table gives the distances in nautical miles of the visible horizon for the given heights in feet above the earth's surface:

x =height :	100	150	200	250	300	350	400
y = distance :	10.63	13.03	15.04	16.81	18.42	19.90	21.27

Find the value of y when x = 218 ft (*Ans*: 15.699)

10. Using the same data as in exercise 9, find the value of y when x = 410ft.

Consider the initial value problem of first order

$$y' = f(x, y), \quad y(x_0) = y_0. \qquad \ldots(1)$$

Starting with given x_0 and the value of h is chosen so small, we suppose $x_0, x_1, x_2, \ldots$ be equally spaced x values (called *mesh points*) with interval h.

i.e., $x_1 = x_0 + h, \quad x_2 = x_1 + h, \ldots$

Also denote $y_0 = y(x_0), \; y_1 = y(x_1), \; y_2 = y(x_2), \ldots$

By separating variables, the differential equation in (1) becomes

$$dy = f(x, y)dx. \qquad \ldots(1A)$$

Integrating (1A) from x_0 to x_1 with respect to x, (at the same time y changes from y_0 to y_1) we get

$$\int_{y_0}^{y_1} dy = \int_{x_0}^{x_1} f(x, y)dx$$

or $$y_1 - y_0 = \int_{x_0}^{x_1} f(x, y)dx$$

or $$y_1 = y_0 + \int_{x_0}^{x_1} f(x, y)dx \qquad \ldots(2)$$

Assuming that $f(x, y) \approx f(x_0, y_0)$ in $x_0 \le x \le x_1$, (2) gives

$$y_1 \approx y_0 + f(x_0, y_0)(x_1 - x_0)$$

or $$y_1 \approx y_0 + hf(x_0, y_0).$$

Similarly, for the range $x_1 \le x \le x_2$, we have

$$y_2 = y_1 + \int_{x_1}^{x_2} f(x, y)dx \qquad \ldots(3)$$

Assuming that $f(x, y) \approx f(x_1, y_1)$ in $x_1 \le x \le x_2$, (3) gives

$$y_2 \approx y_1 + hf(x_1, y_1).$$

Proceeding in this way, we obtain the general formula

$$y_{n+1} = y_n + hf(x_n, y_n) \qquad (n = 0, 1, \cdots) \qquad \ldots(4)$$

The above is called the **Euler method** or **Euler-Cauchy method.**

Working Rule (Euler method)

Given the initial value problem (1). Suppose $x_0, x_1, x_2, \ldots$ be equally spaced x values with interval h. i.e., $x_1 = x_0 + h, \quad x_2 = x_1 + h, \ldots$ Also denote $y_0 = y(x_0), \quad y_1 = y(x_1), \quad y_2 = y(x_2), \ldots$

Then the iterative formula of **Euler method** is:

$$y_{n+1} = y_n + hf(x_n, y_n) \qquad (n = 0, 1, \cdots) \qquad \ldots(5)$$

Example Use Euler's method with h = 0.1 to solve the initial value problem $\frac{dy}{dx} = x^2 + y^2$ with $y(0) = 0$ in the range $0 \le x \le 0.5$.

Here $f(x, y) = x^2 + y^2, \; x_0 = 0, \; y_0 = 0, \; h = 0.1.$

Hence

$$x_1 = x_0 + h = 0.2, \quad x_2 = x_1 + h = 0.2, \quad x_3 = x_2 + h = 0.3, \quad x_4 = x_3 + h = 0.4, \quad x_5 = x_4 + h = 0.5.$$

We determine $y_1, \; y_2, \; y_3, \; y_4, \; y_5$ using the Euler formula (5). Substituting the given value in

$$y_{n+1} = y_n + hf(x_n, y_n)$$

we obtain

$$y_{n+1} = y_n + 0.1(x_n^2 + y_n^2) \qquad (n = 0, 1, \cdots)$$

$$y_1 = y_0 + 0.1(x_0^2 + y_0^2) = 0 + 0.1(0 + 0) = 0.$$

$$y_2 = y_1 + 0.1(x_1^2 + y_1^2) = 0 + 0.1\left[(0.1)^2 + 0^2\right] = 0.001.$$

$$y_3 = y_2 + 0.1(x_2^2 + y_2^2) = 0.001 + 0.1\left[(0.2)^2 + (0.001)^2\right] = 0.005.$$

$$y_4 = y_3 + 0.1(x_3^2 + y_3^2) = 0.005 + 0.1\left[(0.3)^2 + (0.005)^2\right] = 0.014.$$

$$y_5 = y_4 + 0.1(x_4^2 + y_4^2) = 0.014 + 0.1\left[(0.4)^2 + (0.014)^2\right] = 0.0300196.$$

Hence

$y(0) = 0$ $\qquad$ $y(0.1) = 0$ $\qquad$ $y(0.2) = 0.001$

$y(0.3) = 0.005$ $\qquad$ $y(0.4) = 0.014$ $\qquad$ $y(0.5) = 0.0300196.$

Example Using Euler method solve the equation $y' = 2xy + 1$ with $y(0) = 0, h = 0.02$ for $x = 0.1$.

Here $f(x, y) = 2xy + 1, x_0 = 0, y_0 = 0, h = 0.02$. Hence

$$x_1 = x_0 + h = 0.02, \quad x_2 = x_1 + h = 0.04, \quad x_3 = x_2 + h = 0.06, \quad x_4 = x_3 + h = 0.08, \quad x_5 = x_4 + h = 0.1.$$

We determine y_1, y_2, y_3, y_4, y_5 using the Euler formula (5). Substituting the given value in

$$y_{n+1} = y_n + hf(x_n, y_n)$$

we obtain

$$y_{n+1} = y_n + 0.02(2x_n y_n + 1) \qquad (n = 0, 1, \cdots)$$

$$y_1 = y_0 + 0.02(2x_0 y_0 + 1) = 0 + 0.02(0 + 1) = 0.02.$$

$$y_2 = y_1 + 0.02(2x_1 y_1 + 1) = 0.02 + 0.02(2 \times 0.02 \times 0.02 + 1) = 0.04,$$

approximate to 2 places of decimals

$$y_3 = y_2 + 0.02(2x_2 y_2 + 1) = 0.04 + 0.02(2 \times 0.04 \times 0.04 + 1) = 0.06$$

$$y_4 = y_3 + 0.02(2x_3 y_3 + 1) = 0.06 + 0.02(2 \times 0.06 \times 0.06 + 1) = 0.08$$

$$y_5 = y_4 + 0.02(2x_4 y_4 + 1) = 0.08 + 0.02(2 \times 0.08 \times 0.08 + 1) = 0.1$$

Hence

$y(0) = 0$ $\quad y(0.02) = 0.02$ $\quad y(0.04) = 0.04$

$y(0.06) = 0.06$ $\quad y(0.08) = 0.08$ $\quad y(0.1) = 0.1.$

That is the approximate value of $y(0.1)$ is 0.1.

Example Given the initial value problem $y' = x + y, y(0) = 0$. Find the value of y approximately for $x = 1$ by Euler method in five steps. Compare the result with the exact value.

Here $f(x, y) = x + y, \quad x_0 = 0, \quad y_0 = y(x_0) y(0) = 0.$ As we have to calculate the value of y in five steps, we have to take $h = \frac{x_n - x_0}{n} = \frac{1 - 0}{5} = 0.2.$ Hence

$$x_1 = x_0 + h = 0.2, \quad x_2 = x_1 + h = 0.4, \quad x_3 = x_2 + h = 0.6, \quad x_4 = x_3 + h = 0.8, \quad x_5 = x_4 + h = 1.0.$$

We determine y_1, y_2, y_3, y_4, y_5 using the Euler formula (5). Substituting the given value in (5), we obtain

$$y_{n+1} = y_n + 0.2(x_n + y_n) \qquad (n = 0, 1, \cdots)$$

The steps are given in the following Table.

Also the exact solution to the linear differential equation $y' = x + y$ with the initial condition $y(0) = 0$ can be found out to be

$$y = e^x - x - 1. \qquad \ldots(6)$$

The exact values of y can be evaluated from (6) by substituting the corresponding x values, in particular,

$y_1 = y(x_1) = e^{x_1} - x_1 - 1 = e^{0.2} - 0.2 - 1 = 0.000,$ approximately.

The other exact values are also shown in the following table.

n	x_n	approximate value of y_n	$0.2(x_n + y_n)$	Exact values	Absolute value of Error
0	0.0	0.000	0.000	0.000	0.000
1	0.2	0.000	0.040	0.021	0.021
2	0.4	0.040	0.088	0.092	0.052
3	0.6	0.128	0.146	0.222	0.094
4	0.8	0.274	0.215	0.426	0.152
5	1.0	0.489		0.718	0.229

The approximate value of $y(1.0)$ by Euler's method is 0.489, while exact value is 0.718.

Exercises

In Exercises 1-11, solve the initial value problem using Euler's method for value of y at the given point of x with given (h is given in brackets)

1. $\frac{dy}{dx} = 1 - y,\ y(0) = 0$ at the point $x = 0.2\ (h = 0.1).$
2. $\frac{dy}{dx} = \frac{y - x}{1 + x},\ y(0) = 1$ at the point $x = 0.1\ (h = 0.02).$
3. $yy' = x,\ y(0) = 1.5$ at the point $x = 0.2\ (h = 0.1).$
4. $\frac{dy}{dx} = 3x + \frac{1}{2}y,\ y(0) = 1$ at the point $x = 0.2\ (h = 0.05).$
5. $y' = x + y + xy,\ y(0) = 1$ at the point $x = 0.1 (\ h = 0.02).$

6. $\frac{dy}{dx} = 1 + y^2,\ y(0) = 0$ at the point $x = 0.4$ $(h = 0.2)$.

7. $\frac{dy}{dx} = xy,\ y(0) = 1$ at the point $x = 0.4$ $(h = 0.2)$.

8. $\frac{dy}{dx} = 1 + \ln(x + y),\ y(0) = 1$ at the point $x = 0.2$ $(h = 0.1)$.

9. $y' = x^2 + y,\ y(0) = 1$ at the point $x = 0.1$ $(h = 0.05)$.

10. $y' = 2xy,\ y(0) = 1$ at the point $x = 0.5$ $(h = 0.1)$.

11. $y' = -y,\ y(0) = 1$ at the point $x = 0.04$ $(h = 0.01)$.

In Exercises 12-15, apply Euler's method. Do 10 steps. Also solve the problem exactly. Compute the errors to see that the method is too inaccurate for practical purposes.

12. $y' + 0.1y = 0,\ y(0) = 2,\ h = 0.1$

13. $y' = \frac{1}{2}\sqrt{1 - y^2},\ y(0) = 0,\ h = 0.1$

14. $y' + 5x^4y^2 = 0,\ y(0) = 1,\ h = 0.2$

15. $y' = (y + x)^2,\ y(0) = 1,\ h = 0.1$

16. Solve using Euler's method $y'(x + y) = y - x$ with $y(0) = 2$ for the range $0.00(0.02)0.06$.

17. Solve using Euler's method $y' = y - \frac{2x}{y}$ with $y = 1$ at $x = 0$ for $h = 0.5$ on the interval $[0, 1]$.

18. Using Euler's method find $y(0.2)$ of the initial value problem $y' = x + 2y,\ y(0) = 1$, taking $h = 0.1$.

19. Using Euler's method find the value of y at the point $x = 2$ in steps of 0.2 of the initial value problem $\frac{dy}{dx} = 2 + \sqrt{xy},\ y(1) = 1$.

Modified Euler Method

Modified Euler method is given by the iteration formula

$$y_1^{(n+1)} = y_0 + \frac{h}{2}[f(x_0, y_0) + f(x_1, y_1^{(n)})],\quad n = 0, 1, 2, \cdots$$

where $y_1^{(n)}$ is the nth approximation to y_1. The iteration formula can be started by choosing $y_1^{(0)}$ from Euler's formula

$$y_1^{(0)} = y_0 + hf(x_0, y_0).$$

Example Using modified Euler's method, determine the value of y when $x = 0.1$ given that

$$y' = x^2 + y;\ y(0) = 1. \quad (\text{Take } h = 0.05)$$

Here $f(x, y) = x^2 + y;\ x_0 = 0,\ y_0 = 1.$

$$y_1^{(0)} = y_0 + hf(x_0, y_0) = 1 + 0.05(1) = 1.05$$

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

$$= 1 + \frac{0.05}{2}[f(0, 1) + f(0.05, 1.05)]$$

$$= 1 + 0.025[1 + (0.05)^2 + 1.05]$$

$$= 1.0513$$

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

$$= 1 + \frac{0.05}{2}[f(0, 1) + f(0.05, 1.0513)]$$

$$= 1 + 0.025[1 + (0.05)^2 + 1.0513]$$

$$= 1.0513$$

Hence we take $y_1 = 1.0513,$ which is correct to four decimal places.

Formula takes the form

$$y_2^{(n+1)} = y_1 + \frac{h}{2}[f(x_1, y_1) + f(x_2, y_2^{(n)})] \quad n = 0, 1, 2, \cdots$$

where we first evaluate $y_2^{(0)}$ using the Euler formula

$$y_2^{(0)} = y_1 + hf(x_1, y_1).$$

$$= 1.0513 + 0.05\left[(0.05)^2 + 1.0513\right] = 1.1040$$

$$y_2^{(1)} = y_1 + \frac{h}{2}[f(x_1, y_1) + f(x_2, y_2^{(0)})]$$

$$= 1 + \frac{0.05}{2}\left\{\left[(0.05)^2 + 1.0513\right] + \left[(0.1)^2 + 1.1040\right]\right\}$$

$$= 1.1055$$

$$y_2^{(2)} = y_1 + \frac{h}{2}[f(x_1, y_1) + f(x_2, y_2^{(1)})]$$

$$= 1 + \frac{0.05}{2}\left\{\left[(0.05)^2 + 1.0513\right] + \left[(0.1)^2 + 1.1055\right]\right\}$$

$$= 1.1055$$

Hence we take $y_2 = 1.1055$.

Hence the value of y when $x = 0.1$ is 1.1055 correct to four decimal places.

Example Using modified Euler's method, determine the value of y when $x = 0.2$ given that

$$\frac{dy}{dx} = x + \sqrt{y};\ y(0) = 1. \quad \text{(Take } h = 0.2\text{)}$$

Here $f(x, y) = x + \sqrt{y};\ x_0 = 0,\ y_0 = 1.$

$$y_1^{(0)} = y_0 + hf(x_0, y_0) = 1 + 0.2(0 + 1) = 1.2$$

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

$$= 1 + \frac{0.2}{2}[1 + (0.2 + \sqrt{1.2}] = 1.2295.$$

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

$$= 1 + \frac{0.2}{2}[1 + (0.2 + \sqrt{1.2295}] = 1.2309.$$

$$y_1^{(3)} = y_0 + \frac{h}{2}[f(x_0, y_0) + f(x_1, y_1^{(2)})]$$

$$= 1 + \frac{0.2}{2}[1 + (0.2 + \sqrt{1.2309}] = 1.2309.$$

Hence we take $y(0.2) = y_1 = 1.2309..$

Exercises

In Exercises 1-11, solve the initial value problem using modified Euler's method for value of y at the given point of x with given (h is given in brackets)

1. $\frac{dy}{dx} = 1 - y,\ y(0) = 0$ at the point $x = 0.2$ $(h = 0.1)$.

2. $\frac{dy}{dx} = \frac{y - x}{1 + x},\ y(0) = 1$ at the point $x = 0.1$ $(h = 0.02)$.

3. $yy' = x,\ y(0) = 1.5$ at the point $x = 0.2$ $(h = 0.1)$.

4. $\frac{dy}{dx} = 3x + \frac{1}{2}y,\ y(0) = 1$ at the point $x = 0.2$ $(h = 0.05)$.

5. $y' = x + y + xy,\ y(0) = 1$ at the point $x = 0.1$ $(h = 0.02)$.

6. $\frac{dy}{dx} = 1 + y^2,\ y(0) = 0$ at the point $x = 0.4$ $(h = 0.2)$.

7. $\frac{dy}{dx} = xy,\ y(0) = 1$ at the point $x = 0.4$ $(h = 0.2)$.

8. $\frac{dy}{dx} = 1 + \ln(x + y),\ y(0) = 1$ at the point $x = 0.2$ $(h = 0.1)$.

9. $y' = x^2 + y,\ y(0) = 1$ at the point $x = 0.1$ $(h = 0.05)$.

10. $y' = 2xy,\ y(0) = 1$ at the point $x = 0.5$ $(h = 0.1)$.

11. $y' = -y,\ y(0) = 1$ at the point $x = 0.04$ $(h = 0.01)$.

UNIT 3

SOLUTION OF SYSTEMS OF LINEAR EQUATIONS

Solution of system of linear equations

A system of m linear equations in n unknowns $x_1, x_2, \ldots, x_n$ is a set of equations of the form

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

where the coefficients a_{jk} and the b_j are given numbers. The system is said to be **homogeneous** if all the b_j are zero; otherwise, it is said to be **non-homogeneous**.

The system of linear equations is equivalent to the matrix equation (or the single vector equation)

$$Ax = b$$

where the **coefficient matrix** $A = [a_{ij}]$ is the $m \times n$ matrix and **x** and **b** are the column matrices (vectors) given by:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdot & \cdot & \ldots & \cdot \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \text{ and } \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_m \end{bmatrix}$$

A **solution** of the system is a set of numbers $x_1, x_2, \ldots, x_n$ which satisfy all the m equations, and a **solution vector** of (1) is a column matrix whose components constitute a solution of system. The method of solving such a system using methods like Cramer's rule is impracticable for large systems. Hence, we use other methods like Gauss elimination.

Gauss Elimination Method

In the Gauss elimination method, the solution to the system of equations is obtained in two stages. In the first stage, the given system of equations is reduced to an equivalent upper triangular form using elementary transformations. In the second stage, the upper triangular system is solved using back substitution procedure by which we obtain the solution in the order $x_n, x_{n-1}, x_{n-2}, \cdots, x_2, x_1$.

Example Solve the system

$$2x_1+x_2+2x_3+x_4=6 \qquad \ldots(1)$$

$$6x_1-6x_2+6x_3+12x_4=36 \qquad \ldots(2)$$

$$4x_1+3x_2+3x_3-3x_4=-1 \qquad \ldots(3)$$

$$2x_1+2x_2-x_3+x_4=10 \qquad \ldots(4)$$

To eliminate x_1 from equations (2), (3) and (4), we subtract suitable multiples of equation (1) and we get the following system of equations:

$(2)-3\cdot(1)\rightarrow \quad -9x_2+0x_3+9x_4=18 \qquad \ldots(5)$

$(3)-2\cdot(1)\rightarrow \quad x_2-x_3-5x_4=-13 \qquad \ldots(6)$

$(4)-1\cdot(1)\rightarrow \quad x_2-3x_3+0x_4=4 \qquad \ldots(7)$

To eliminate x_2 from equations (6) and (7), subtract suitable multiples of equation (5) and get the following system of equations:

$(6)-(-1/9)(5)\rightarrow -x_3-4x_4=-11 \qquad \ldots(8)$

$(7)-(-1/9)(5)\rightarrow -3x_3+x_4=6 \qquad \ldots(9)$

To eliminate x_2 from equation (9), subtract $3\times(8)$ and get the following equation:

$$13\,x_4=39 \qquad \ldots(10)$$

From equation (10), $x_4 = 39/13 = 3$; using this value of x_4, (9) gives $x_3 = -1$; using these values of x_4 and x_3, (7) gives $x_2 = 1$; using all these values (1) gives $x_1 = 2$. Hence the solution to the system is $x_1 = 2$, $x_2 = 1$, $x_3 = -1$, $x_4 = 3$.

Note: The above method can be simplified using the matrix notation. The given system of equations can be written as

$$Ax=b$$

and the augmented matrix is

$$\begin{bmatrix} 2 & 1 & 2 & 1 & 6 \\ 6 & -6 & 6 & 12 & 36 \\ 4 & 3 & 3 & -3 & -1 \\ 2 & 2 & -1 & 1 & 10 \end{bmatrix}$$

which on successive row transformations give

$$\begin{bmatrix} 2 & 1 & 2 & 1 & 6 \\ 0 & -9 & 0 & 9 & 18 \\ 0 & 0 & -1 & -4 & -11 \\ 0 & 0 & 0 & 13 & 39 \end{bmatrix}.$$

Hence

$$\begin{bmatrix} 2 & 1 & 2 & 1 \\ 0 & -9 & 0 & 9 \\ 0 & 0 & -1 & -4 \\ 0 & 0 & 0 & 13 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 6 \\ 18 \\ -11 \\ 39 \end{bmatrix}$$

Back substitution gives

$$x_1 = 2, \quad x_2 = 1, \quad x_3 = -1, \quad x_4 = 3$$

In the example, we had $a_{11} \neq 0$. Otherwise we would not have been able to eliminate x_1 by using the equations in the given order. Hence if $a_{11} \neq 0$ in the system of equations we have to reorder the equations (and perhaps even the unknowns in each equation) in a suitable fashion; similarly, in the further steps. Such a situation can be seen in the following Example.

Example Using Gauss elimination solve:

$$\begin{aligned} y + 3z &= 9 \\ 2x + 2y - z &= 8 \\ -x + 5z &= 8 \end{aligned}$$

Here the leading coefficient (i.e., coefficient of x) is 0. Hence to proceed further we have to interchange rows 1 and 2, so that

$$\begin{aligned} 2x + 2y - z &= 8 \quad \ldots(1) \\ y + 3z &= 9 \quad \ldots(2) \\ -x + 5z &= 8 \quad \ldots(3) \end{aligned}$$

Elimination of x from last two equations:

$$\begin{aligned} 2x + 2y - z &= 8 \\ y + 3z &= 9 \\ (3) + \frac{1}{2}(1) \rightarrow \quad y + \frac{9}{2}z &= 12 \quad \ldots(4) \end{aligned}$$

Elimination of y from last equation:

$$\begin{aligned} 2x + 2y - z &= 8 \\ y + 3z &= 9 \end{aligned}$$

(4) – (2) → $\frac{3}{2}z = 3$ …(5)

Hence $z = 2,\ y = 9 - 6 = 3,\ x = 2.$

Hence

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}.$$

Partial and Full Pivoting

In each step in the Gauss elimination method, the coefficient of the first unknown in the first equation is called **pivotal coefficient**. By the above Example, the Gauss elimination method fails if any one of the pivotal coefficients becomes zero. In such a situation, we rewrite the equations in a different order to avoid zero pivotal coefficient. Changing the order of equations is called **pivoting**.

In **partial pivoting**, if the pivotal coefficient a_{ii} happens to be zero or near to zero, the i^{th} column elements are searched for the numerically largest element. Let the j^{th} row ($j>i$) contains this element, then we interchange the i^{th} equation with the j^{th} equation and proceed for elimination. This process is continued whenever pivotal coefficients become zero during elimination.

In **total pivoting**, we look for an absolutely largest coefficient in the entire system and start the elimination with the corresponding variable, using this coefficient as the pivotal coefficient (for this we have to interchange *rows* and *columns*, if necessary); similarly in the further steps. Total pivoting, in fact, is more complicated than the partial pivoting. Partial pivoting is preferred for hand calculation.

Example Solve the system

$$0.0004x_1 + 1.402x_2 = 1.406 \quad \ldots(1)$$

$$0.4003x_1 - 1.502x_2 = 2.501 \quad \ldots(2)$$

by Gauss elimination (*a*) without pivoting (*b*) with partial pivoting.

(*a*) without pivoting (choosing the first equation as the pivotal equation)

$$0.0004x_1 + 1.402x_2 = 1.406 \quad \ldots(1a)$$

$$(2) - \frac{0.40031}{0.0001} \times (1a) \rightarrow \quad -1405x_2 = -1404 \quad \ldots(2a)$$

and so $x_2 = \frac{1404}{1405} = 0.9993$

and hence from (1*a*),

$$x_1 = \frac{1}{0.0004}(1.406 - 1.402 \times 0.9993) = \frac{0.005}{0.0004} = 12.5.$$

(*b*) (with partial pivoting)

Since $|a_{11}|$ is small and is nearer to zero as compared with $|a_{21}|$, we accept a_{21} as the pivotal coefficient (i.e. second equation becomes the pivotal equation). To start with we rearrange the given system as follows:

$$0.4003x_1 - 1.502x_2 = 2.501 \qquad \ldots(3)$$

$$0.0004x_1 + 1.402x_2 = 1.406 \qquad \ldots(4)$$

Now by Gauss elimination the system becomes,

$$0.4003x_1 - 1.502x_2 = 2.501 \qquad \ldots(3a)$$

$$(4) - \frac{\cdot 0004}{\cdot 4003}(3) \quad 1.404x_2 = 1.404 \qquad \ldots(4a)$$

and so $$x_2 = \frac{1.404}{1.404} = 1$$

and from (3*a*) $$x_1 = \frac{1}{0.4003}(2.501 + 1.502 \times 1) = 10.$$

Example Solve the following system (i) without pivoting (ii) with pivoting

$$0.0002x + 0.3003y = 0.1002 \qquad \ldots (1)$$

$$2.0000x + 3.0000y = 2.0000. \qquad \ldots (2)$$

(i) without pivoting

$$0.0002x + 0.3003y = 0.1002$$

$$(2) - \frac{2}{\cdot 0002}(1) \rightarrow \left(3.000 - \frac{0.3003 \times 2}{0.0002}\right)y = 2.0000 - \frac{0.1002 \times 2}{0.0002}$$

i.e., $$1498.5y = 499.$$

Now by back substitution, the solution to the system is given by $y = 0.3330$ and $x = 0.5005$;

(ii) With pivoting:

Since $|a_{11}|$ is small and is nearer to zero as compared with $|a_{21}|$, we accept a_{21} as the pivotal coefficient (i.e. second equation becomes the pivotal equation). To start with we rearrange the given system as follows:

$$2.0000x + 3.0000y = 2.0000 \qquad \ldots (3)$$

$$0.0002x + 0.3003y = 0.1002 \qquad \ldots (4)$$

$$(4)-\frac{\cdot 0002}{2}(3)\rightarrow\left(0.3003-\frac{3.0000\times 0.0002}{2}\right)y=0.1002-\frac{2\times 0.0002}{2}$$

which simplifies to

$$0.3000\,y=0.1000.$$

Hence by bank substitution, the solution is

$$y=\frac{1}{3}\quad\text{and}\quad x=\frac{1}{2}.$$

Cholesky Method (Modification of the Gauss method)

Cholesky method, which is a modification of the Gauss method, is based on the result that any positive definite square matrix **A** can be represented in the form **A = LU**, where **L** and **U** are the unique lower and upper triangular matrices. The method is illustrated through the following examples.

Example Using Cholesky's method, solve the system:

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 14\\ 2x_1 + 3x_2 + 4x_3 &= 20\\ 3x_1 + 4x_2 + x_3 &= 14 \end{aligned}$$

(**LU** *decomposition of the coefficient matrix* **A**)

$$\mathbf{A}=\begin{bmatrix}1&2&3\\2&3&4\\3&4&1\end{bmatrix}\qquad \begin{aligned} R_2 &\rightarrow R_2+(-2)R_1 \quad m_{21}=-2\\ R_3 &\rightarrow R_3+(-3)R_1 \quad m_{31}=-3 \end{aligned}$$

$$\sim\begin{bmatrix}1&2&3\\0&-1&-2\\0&-2&-8\end{bmatrix}$$

$$\sim\begin{bmatrix}1&2&3\\0&-1&-2\\0&0&-4\end{bmatrix}\qquad R_3\rightarrow R_3+(-2)R_2 \quad m_{32}=-3$$

We take $U=\begin{bmatrix}1&2&3\\0&-1&-2\\0&0&-4\end{bmatrix}$ as the upper triangular matrix.

Using the multipliers $m_{21}=-2,\ m_{31}=-3,\ m_{32}=-2$, we get the lower triangular matrix as follows:

$$L=\begin{bmatrix}1&0&0\\-m_{21}&1&0\\-m_{31}&-m_{32}&1\end{bmatrix}=\begin{bmatrix}1&0&0\\2&1&0\\3&2&1\end{bmatrix}.$$

(*Solution of the system*)

The given system of equations can be written as

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 20 \\ 14 \end{bmatrix} \quad \ldots (1)$$

The above can be written as

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 20 \\ 14 \end{bmatrix} \quad \ldots (2)$$

where

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \quad \ldots (3)$$

Solving the system in (2) by forward substitution, we get

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 14 \\ -8 \\ -12 \end{bmatrix}$$

With these values of y_1, y_2, y_3, Eq. (3) can now be solved by back substitution and we obtain

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Example Solve the equations

$$2x + 3y + z = 9$$

$$x + 2y + 3z = 6$$

$$3x + y + 2z = 8$$

by LU decomposition.

(**LU** decomposition of the coefficient matrix **A**)

Proceeding as in the above example,

$$U = \begin{bmatrix} 2 & 3 & 1 \\ 0 & \frac{1}{2} & \frac{5}{2} \\ 0 & 0 & 18 \end{bmatrix} \text{ and } L = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{3}{2} & -7 & 1 \end{bmatrix}$$

(*Solution of the system*)

The given system of equations can be written as

$$\begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 3/2 & -7 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 1 \\ 0 & 1/2 & 5/2 \\ 0 & 0 & 18 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 9 \\ 6 \\ 8 \end{bmatrix} \qquad \ldots \text{(iv)}$$

or, as

$$\begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 3/2 & -7 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 6 \\ 8 \end{bmatrix}, \qquad \ldots \text{(v)}$$

where

$$\begin{bmatrix} 2 & 3 & 1 \\ 0 & 1/2 & 5/2 \\ 0 & 0 & 18 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}. \qquad \ldots \text{(vi)}$$

Solving the system in (v) by forward substitution, we get

$$y_1 = 9, \qquad y_2 = \frac{3}{2}, \qquad y_3 = 5.$$

With these values of $y_1, y_2, y_3,$ eq. (vi) can now be solved by the back substitution process and we obtain

$$x = \frac{35}{18}, \qquad y = \frac{29}{18}, \qquad z = \frac{5}{18}.$$

Gauss Jordan Method

The method is based on the idea of reducing the given system of equations **Ax = b,** to a diagonal system of equations **Ix = d,** where **I** is the identity matrix, using elementary row operations. We know that the solutions of both the systems are identical. This reduced system gives the solution vector **x**. This reduction is equivalent to finding the solution as $x = A^{-1}b$.

In this case, a system of 3 equations in 3 unknowns

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_2 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_2 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned}$$

is written as

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \quad ---(*)$$

After some linear transformations, we obtain the 3 × 3 system as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \quad --- \quad (**)$$

To obtain the system as given in (**), first we augment the matrices given is (*) as,

$\begin{bmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{bmatrix}$ and after some elementary operations, it is written as,

$\begin{bmatrix} 1 & 0 & 0 & d_1 \\ 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & d_3 \end{bmatrix} --- (***)$, this helps us to write the given system as given in (**). Then it is easy to get the solution of the system as $x_1 = d_1, x_2 = d_2$ *and* $x_3 = d_3$.

Elimination procedure: The first step is same as in Gauss elimination method, which is, we make the elements below the first pivot in the augmented matrix as zeros, using the elementary row transformations. From the second step onwards, we make the elements below and above the pivots as zeros using the elementary row transformations. Lastly, we divide each row by its pivot so that the final matrix is of the form (***). Partial pivoting can also be used in the solution. We may also make the pivots as 1 before performing the elimination.

Problem: Solve the following system of equations

$$x_1 + x_2 + x_3 = 1$$
$$4x_1 + 3x_2 - x_3 = 6$$
$$3x_1 + 5x_2 + 3x_3 = 4$$

using the Gauss-Jordan method without partial pivoting

Solution:

We have the matrix form as

$\begin{bmatrix} 1 & 1 & 1 \\ 4 & 3 & -1 \\ 3 & 5 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 4 \end{bmatrix}$. Then the augmented matrix is,

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & 3 & -1 & 6 \\ 3 & 5 & 3 & 4 \end{bmatrix}$$

(*i*) To do the eliminations follow the operations,

$R2 = R2 - 4R1$, and $R3 = R3 - 3R1$. This gives,

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -1 & -5 & 2 \\ 0 & 2 & 0 & 1 \end{bmatrix}$$

Then, $R1 = R1 + R2$ and $R3 = R3 + 2R2$ gives,

$$\begin{bmatrix} 1 & 0 & -4 & 3 \\ 0 & -1 & -5 & 2 \\ 0 & 0 & -10 & 5 \end{bmatrix}$$

$R1 = R1 - (4/10)\,R3$, $R2 = R2 - (5/10)\,R3$ gives,

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & -\frac{1}{2} \\ 0 & 0 & -10 & 5 \end{bmatrix}$$

Now, making the pivots as 1, $R2 = ((-R2)$ and $R3 = (R3/(-10)))$, we get

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

Hence, $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$

Therefore, the solution of the system is,

$$x_1 = 1,\ x_2 = \frac{1}{2},\ x_3 = -\frac{1}{2}.$$

Note: The Gauss-Jordan method looks very elegant as the solution is obtained directly. However, it is computationally more expensive than Gauss elimination. For large n, the total number of divisions and multiplications for Gauss-Jordan method is almost 1.5 times the total number of divisions and multiplications required for Gauss elimination. Hence, we do not normally use this method for the solution of the system of equations.

The most important application of this method is to find the inverse of a non-singular matrix. To obtain inverse of a matrix, we start with the augmented matrix of **A** with the identity matrix **I** of the same order.

When the Gauss-Jordan procedure is completed, we obtain, the matrix A augmented with I, $\left[A|I\right]$ in the form $\left[I|A^{-1}\right]$, since $AA^{-1} = I$.

Example Using Gauss Jordan method solve the system of equations:

$$x + 2y + z = 8 \quad \ldots (1)$$

$$2x + 3y + 4z = 20 \quad \ldots (2)$$

$$4x + 3y + 2z = 16 \quad \ldots (3)$$

[*Elimination of* x *from Eqs.* (2) *and* (3), using (1)]

$$x + 2y + z = 8 \quad \ldots (1a)$$

$$-y + 2z = 4 \quad \ldots (2a)$$

$$-5y + 2z = -16 \quad \ldots (3a)$$

[*Elimination of* y *from* (1a) *and* (3a), using (2a)]

$$x + 5z = 16 \quad \ldots (1b)$$

$$-y + 2z = 4 \quad \ldots (2b)$$

$$-12z = -36 \quad \ldots (3b)$$

[*Elimination of* z *from* (1b) *and* (2b), using (3b)]

$$x = 1 \quad \ldots (1c)$$

$$-y = -2 \quad \ldots (2c)$$

$$-12z = -36 \quad \ldots (3c)$$

Hence, $x = 1, y = 2, z = 3$.

Assignments

1. Apply Gauss elimination method to solve the equations:

$$2x + 3y - z = 5$$

$$4x + 4y - 3z = 3$$

$$-2x + 3y - z = 1$$

2. Apply Gauss elimination method to solve the equations:

$$3x_1 + 6x_2 + x_3 = 16$$

$$2x_1 + 4x_2 + 3x_3 = 13$$

$$x_1 + 3x_2 + 2x_3 = 9$$

3. Apply Gauss elimination method to solve the equations:

$$10x + 2y + z = 9$$

$$2x + 20y - 2z = -44$$

$$-2x + 3y + 10z = 22$$

4. Apply Gauss elimination method to solve the equations:

$$x + y + z = 10$$
$$2x + y + 2z = 17$$
$$3x + 2y + z = 17$$

5. Solve the system, using Gauss elimination method:

$$5x_1 + x_2 + x_3 + x_4 = 4$$
$$x_1 + 7x_2 + x_3 + x_4 = 12$$
$$x_1 + x_2 + 6x_3 + x_4 = -5$$
$$x_1 + x_2 + x_3 + 4x_4 = -6$$

6. Apply Gauss elimination method to solve the equations:

$$x + 4y - z = -5$$
$$x + y - 6z = -12$$
$$3x + y - z = 4$$

7. Solve the following system, using Cholesky method

$$10x + y + z = 12$$
$$2x + 10y + z = 13$$
$$2x + 2y - 10z = 14$$

8. Solve the following system, using Cholesky method

$$2x + 3y - z = 5$$
$$4x + 4y - 3z = 3$$
$$-2x + 3y - z = 1$$

9. Solve the following system, using Cholesky method

$$2x + 3y + z = 9$$
$$x + 2y + 3z = 6$$
$$3x + y + 2z = 8$$

10. Solve the following using Cholesky method:

$$\begin{bmatrix} 3 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 4 \end{bmatrix}.$$

11. Find the inverse of the following matrix using Cholesky method:

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & -2 & 4 \\ 1 & 2 & 2 \end{bmatrix}.$$

12. Solve the following system using Gauss Jordan method:

$$2x - 3y + z = -1$$

$$x + 4y + 5z = 25$$

$$3x - 4y + z = 2$$

13. Solve the following system using Gauss Jordan method:

$$2x - 3y + 4z = 7$$

$$5x - 2y + 2z = 7$$

$$6x - 3y + 10z = 23$$

MATRIX INVERSION USING GAUSS ELIMINATION

We know that X will be the inverse of an n-square non-singular matrix A if

$$AX = I, \qquad \text{... (1)}$$

where I is the $n \times n$ identity matrix.

Every square non-singular matrix will have an inverse. Gauss elimination and Gauss-Jordan methods are popular among many methods available for finding the inverse of a non-singular matrix.

For the third order matrices, (1) may be written as

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Clearly the above equation is equivalent to the three equations

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_{13} \\ x_{23} \\ x_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

We can therefore solve each of these systems using Gaussian elimination method and the result in each case will be the corresponding column of $X = A^{-1}$. We solve all the three equations simultaneously as illustrated in the following examples.

Example Using Gaussian elimination, find the inverse of the matrix $A = \begin{bmatrix} 2 & 1 & 1 \\ 3 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix}$.

In this method, we place an identity matrix, whose order is same as that of A, adjacent to A which we call *augmented matrix*. Then the inverse of A is computed in two stages. In the first stage, A is converted into an upper triangular form, using Gaussian elimination method.

We write the augmented system first and then apply low transformations:

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 3 & 2 & 3 & 0 & 1 & 0 \\ 1 & 4 & 9 & 0 & 0 & 1 \end{array}\right] \sqcup \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{3}{2} & -\frac{3}{2} & 1 & 0 \\ 0 & \frac{7}{2} & \frac{17}{2} & -\frac{1}{2} & 0 & 1 \end{array}\right] \begin{array}{l} \text{by } R_2 \to R_2 - \frac{3}{2}R_1 \\ \text{by } R_3 \to R_3 - \frac{1}{2}R_1 \end{array}$$

$$\sqcup \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{3}{2} & -\frac{3}{2} & 1 & 0 \\ 0 & 0 & -2 & 10 & -7 & 1 \end{array}\right] \text{by } R_3 \to R_3 - 7R_{21}$$

The above is equivalent to the following three systems:

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & \frac{1}{2} & \frac{3}{2} & -\frac{3}{2} \\ 0 & 0 & -2 & 10 \end{array}\right] \qquad \ldots (1)$$

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{3}{2} & 1 \\ 0 & 0 & -2 & -7 \end{array}\right] \qquad \ldots (2)$$

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{3}{2} & 1 \\ 0 & 0 & -2 & 1 \end{array}\right] \qquad \ldots (3)$$

Now the matrix equation of the system of equations corresponding to (1) is

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & \frac{1}{2} & \frac{3}{2} \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix} = \begin{bmatrix} 1 \\ -\frac{3}{2} \\ 10 \end{bmatrix}$$

which on back substitution gives $x_{31} = -5,\ x_{21} = 12,\ x_{11} = -3.$

Similarly using the other two systems other x values are determined and hence the inverse is given by

$$A^{-1} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} -3 & \frac{5}{2} & -\frac{1}{2} \\ 12 & -\frac{17}{2} & \frac{3}{2} \\ -5 & \frac{7}{2} & -\frac{1}{2} \end{bmatrix}.$$

All these operations are also performed on the adjacently placed identity matrix.

Example Use the Gaussian elimination method to find the inverse of the matrix

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

At first, we place an identity matrix of the same order adjacent to the given matrix. Thus, the augmented matrix can be written as

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 4 & 3 & -1 & 0 & 1 & 0 \\ 3 & 5 & 3 & 0 & 0 & 1 \end{array}\right] \qquad \ldots (1)$$

In order to increase the accuracy of the result, it is essential to employ partial pivoting. We look for an absolutely largest coefficient *in the first column* and we use this coefficient as the pivotal coefficient (for this we have to interchange *rows* if necessary)

In first column of matrix (1), 4 is the largest element, and hence is the pivotal element. In order to bring 4 in the first row we interchange the first and second rows and obtain the augmented matrix in the form

$$\left[\begin{array}{ccc|ccc} 4 & 3 & -1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 3 & 5 & 3 & 0 & 0 & 1 \end{array}\right] \qquad \ldots (2)$$

$$\sim \left[\begin{array}{ccc|ccc} 1 & \frac{3}{4} & -\frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 3 & 5 & 3 & 0 & 0 & 1 \end{array}\right] \text{ by } R_1 \to \frac{1}{4}R_1$$

$$\sim \left[\begin{array}{ccc|ccc} 1 & \frac{3}{4} & -\frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{5}{4} & 1 & -\frac{1}{4} & 0 \\ 0 & \frac{11}{4} & \frac{15}{4} & 0 & -\frac{3}{4} & 1 \end{array}\right] \begin{array}{l} \text{by } R_2 \to R_2 - R_1 \\ \text{by } R_3 \to R_3 - 3R_1 \end{array}$$

We now search for an absolutely largest coefficient *in the second column* (and not in the first row) and we use this coefficient as the pivotal coefficient. The pivot element is the max (1/4, 11/4) and is 11/4. Therefore, we interchange second and third rows of the above.

$$\left[\begin{array}{ccc|ccc} 1 & \frac{3}{4} & -\frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{11}{4} & \frac{15}{4} & 0 & -\frac{3}{4} & 1 \\ 0 & \frac{1}{4} & \frac{5}{4} & 1 & -\frac{1}{4} & 0 \end{array}\right]$$

Now, divide R_2 by the pivot element $a_{22} = 11/4$, and obtain

$$\left[\begin{array}{ccc|ccc} 1 & \frac{3}{4} & -\frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 1 & \frac{15}{11} & 0 & -\frac{3}{11} & \frac{4}{11} \\ 0 & \frac{1}{4} & \frac{5}{4} & 1 & -\frac{1}{4} & 0 \end{array}\right]$$

In order to make the entries below 1 in the second column we perform

$R_3 \rightarrow R_3 - (1/4)R_1$ in the above matrix and obtain

$$\left[\begin{array}{ccc|ccc} 1 & \frac{3}{4} & -\frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 1 & \frac{15}{11} & 0 & -\frac{3}{11} & \frac{4}{11} \\ 0 & 0 & \frac{10}{11} & 1 & -\frac{2}{11} & -\frac{1}{11} \end{array}\right]$$

This is equivalent to the following three matrices

$$\left[\begin{array}{ccc|c} 1 & \frac{3}{4} & -\frac{1}{4} & 0 \\ 0 & 1 & \frac{15}{11} & 0 \\ 0 & 0 & \frac{10}{11} & 1 \end{array}\right]; \quad \left[\begin{array}{ccc|c} 1 & \frac{3}{4} & -\frac{1}{4} & \frac{1}{4} \\ 0 & 1 & \frac{15}{10} & -\frac{3}{11} \\ 0 & 0 & \frac{10}{11} & -\frac{2}{11} \end{array}\right]; \quad \left[\begin{array}{ccc|c} 1 & \frac{3}{4} & -\frac{1}{4} & 0 \\ 0 & 1 & \frac{15}{11} & \frac{4}{11} \\ 0 & 0 & \frac{10}{11} & -\frac{1}{11} \end{array}\right]$$

Thus we have

$$A^{-1} == \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} \frac{7}{5} & \frac{1}{5} & -\frac{2}{5} \\ -\frac{3}{2} & 0 & \frac{1}{2} \\ \frac{11}{10} & -\frac{1}{5} & -\frac{1}{10} \end{bmatrix}$$

Matrix Inversion using Gauss-Jordan method

This method is similar to Gaussian elimination method for matrix inversion, starting with the augmented matrix $[AI]$ and reducing A to the identity matrix using elementary row transformations. The method is illustrated in the following example.

Example Find the inverse of the following matrix A by Gauss-Jordan method.

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

The augmented matrix is given by

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 4 & 3 & -1 & 0 & 1 & 0 \\ 3 & 5 & 3 & 0 & 0 & 1 \end{array}\right]$$

$$\sim \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -5 & -4 & 1 & 0 \\ 0 & 2 & 0 & -3 & 0 & 1 \end{array}\right] \begin{array}{l} \text{by } R_2 \to R_2 - 4R_1 \\ \text{by } R_3 \to R_3 - 3R_1 \end{array}$$

$$\sim \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 5 & 4 & -1 & 0 \\ 0 & 2 & 0 & -3 & 0 & 1 \end{array}\right] \text{by } R_2 \to -R_2$$

$$\sim \left[\begin{array}{ccc|ccc} 1 & 0 & -4 & -3 & 1 & 0 \\ 0 & 1 & 5 & 4 & -1 & 0 \\ 0 & 0 & -10 & -11 & 2 & 1 \end{array}\right] \begin{array}{l} \text{by } R_1 \to R_1 - R_2 \\ \text{by } R_3 \to R_3 - 2R_2 \end{array}$$

$$\sim \left[\begin{array}{ccc|ccc} 1 & 0 & -4 & -3 & 1 & 0 \\ 0 & 1 & 5 & 4 & -1 & 0 \\ 0 & 0 & 1 & 11/10 & -1/5 & -1/10 \end{array}\right] \text{by } R_3 \to -\frac{1}{10}R_3$$

$$\sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 7/5 & 1/5 & -2/5 \\ 0 & 1 & 0 & -3/2 & 0 & 1/2 \\ 0 & 0 & 1 & 11/10 & -1/5 & -1/10 \end{array}\right] \begin{array}{l} \text{by } R_1 \to R_1 + 4R_3 \\ \text{by } R_2 \to R_2 - 5R_1 \end{array}$$

Thus we have

$$A^{-1} = \begin{bmatrix} \frac{7}{5} & \frac{1}{5} & -\frac{2}{5} \\ -\frac{3}{2} & 0 & \frac{1}{2} \\ \frac{11}{10} & -\frac{1}{5} & -\frac{1}{10} \end{bmatrix}.$$

- **Triangulation Method (LU Decomposition Method):**

In linear algebra, **LU decomposition** (also called **LU factorization**) factorizes a matrix as the product of a lower triangular matrix and an upper triangular matrix

Let A be a non-singular square matrix. LU **decomposition** is a decomposition of the form

$$A=LU$$

where L is a lower triangular matrix and U is an upper triangular matrix. This means that L has only zeros above the diagonal and U has only zeros below the diagonal. For example, for a 3-by-3 matrix A, its LU decomposition looks like this:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

Consider a system of linear equations,

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1$$
$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2$$
$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3$$

This can be written in the form,

$$Ax=b,$$

where $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ b_m \end{bmatrix}$

To solve the system of equations by LU decomposition, first we decompose A as LU, where,

$$L = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \quad and \quad U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

This gives,

$$LUx = b.$$

Let Ux=y. This implies, Ly=b.

That is,

$$\begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Thus,

$$y_1 = b_1$$
$$l_{21}y_1 + y_2 = b_2$$
$$l_{31}y_1 + l_{32}y_2 + y_3 = b_3$$

This gives the y values by forward substitution, which means, substitute the value of y_1 given by the first equation in the second and solve y_2, then use these values of y_1 *and* y_2 in the third and solve y_3.

Then the system of equations

$$Ux = y \text{ ; that is } \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

gives the required values of x_1, x_2 *and* x_3 as the solution of the original system of linear equations by backward substitution.

To decompose a matrix $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, in the form

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$, we proceed as follows.

On multiplying $\begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix}$ *and* $\begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$, we get,

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} \\ l_{21}u_{11} & l_{21}u_{12} + u_{22} & l_{21}u_{13} + u_{23} \\ l_{31}u_{11} & l_{31}u_{12} + l_{32}u_{22} & l_{31}u_{13} + l_{32}u_{23} + u_{33} \end{bmatrix}$$

Equating it with the corresponding terms of A, we get,

$$u_{11} = a_{11}; \quad u_{12} = a_{12}; \quad u_{13} = a_{13}$$

$$l_{21}u_{11} = a_{21} \Rightarrow l_{21} = \frac{a_{21}}{u_{11}} \; ; \quad l_{31}u_{11} = a_{31} \Rightarrow l_{31} = \frac{a_{31}}{u_{11}}$$

$$l_{21}u_{12} + u_{22} = a_{22} \Rightarrow u_{22} = a_{22} - l_{21}u_{12};$$

$$l_{21}u_{13} + u_{23} = a_{23} \Rightarrow u_{23} = a_{23} - l_{21}u_{13};$$

similililarly,

$$l_{31}u_{12} + l_{32}u_{22} = a_{32}, \quad l_{31}u_{13} + l_{32}u_{23} + u_{33} = a_{33} \text{ gives } l_{32} \text{ and } u_{33}$$

Example: Solve the following system of equations by LU decomposition.

$2x+3y+z=9$

$x+2y+3z=6$

$3x+y+2z=8.$

Solution:

The above system of equations is written as,

$$\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}=\begin{bmatrix} 9 \\ 6 \\ 8 \end{bmatrix}$$

To decompose the matrix $\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}$ in the form of LU, we equate the corresponding terms of A and LU as already illustrated, and obtain

$$u_{11}=2;\quad u_{12}=3;\quad u_{13}=1$$

$$l_{21}=\frac{a_{21}}{u_{11}}=\frac{1}{2}\ ;\quad l_{31}=\frac{a_{31}}{u_{11}}=\frac{3}{2}$$

$$u_{22}=a_{22}-l_{21}u_{12}=2-\frac{1}{2}\times 3=\frac{1}{2};$$

$$u_{23}=a_{23}-l_{21}u_{13}=3-\frac{1}{2}\times 1=\frac{5}{2};$$

$$l_{32}=\frac{a_{32}-l_{31}u_{12}}{u_{22}}=\frac{1-\frac{3}{2}\times 3}{\frac{1}{2}}=-7\quad \textit{and}$$

$$u_{33}=u_{33}=a_{33}-\left(l_{31}u_{13}+l_{32}u_{23}\right)=2-\left(\frac{3}{2}\times 1+\left(-7\right)\times\frac{5}{2}\right)=2-\left(\frac{3}{2}-\frac{35}{2}\right)=18$$

Hence,

$$\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}=\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{3}{2} & -7 & 1 \end{bmatrix}\begin{bmatrix} 2 & 3 & 1 \\ 0 & \frac{1}{2} & \frac{5}{2} \\ 0 & 0 & 18 \end{bmatrix}$$

This implies,

$$\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{3}{2} & -7 & 1 \end{bmatrix}\begin{bmatrix} 2 & 3 & 1 \\ 0 & \frac{1}{2} & \frac{5}{2} \\ 0 & 0 & 18 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}=\begin{bmatrix} 9 \\ 6 \\ 8 \end{bmatrix}$$

Consider

$$\begin{bmatrix} 2 & 3 & 1 \\ 0 & \frac{1}{2} & \frac{5}{2} \\ 0 & 0 & 18 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \text{ then } \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{3}{2} & -7 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 6 \\ 8 \end{bmatrix},$$

Solving these, we get, $\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 9 \\ \frac{3}{2} \\ 5 \end{bmatrix}$

That is,

$$\begin{bmatrix} 2 & 3 & 1 \\ 0 & \frac{1}{2} & \frac{5}{2} \\ 0 & 0 & 18 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 9 \\ \frac{3}{2} \\ 5 \end{bmatrix}$$

Now, solving the above expression we obtain the values of x, y and z as a solution of the given system of equations as,

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{35}{18} \\ \frac{29}{18} \\ \frac{5}{18} \end{bmatrix}.$$

Assignments

1. Using Gauss-Jordan method, find the inverse of the following matrices:

(i) $A = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 3 & -3 \\ -2 & -4 & -4 \end{bmatrix}$ (ii) $B = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 4 \\ 2 & 4 & 7 \end{bmatrix}$

2. Using Gaussian elimination method, find the inverse of the following matrices:

(i) $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 3 & 1 & 1 \end{bmatrix}$ (ii) $B = \begin{bmatrix} 2 & 0 & 1 \\ 3 & 2 & 5 \\ 1 & -1 & 0 \end{bmatrix}$

SOLUTION BY ITERATIONS

SOLUTION BY ITERATION: Jacobi's iteration method and Gauss Seidel iteration method

The methods discussed in the previous section belong to the **direct methods** for solving systems of linear equations; these are methods that yield solutions after an amount of computations that can be specified in advance.

In this section, we discuss **indirect** or **iterative methods** in which we start from an initial value and obtain better and better approximations from a computational cycle repeated as often as may be necessary, for achieving a required accuracy, so that the amount of arithmetic depends upon the accuracy required.

Jacobi's iteration method and Gauss Seidel iteration method

Consider a linear system of n linear equations in n unknowns $x_1, x_2, \ldots, x_n$ of the form

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n &= b_3 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \cdots + a_{nn}x_n &= b_n \end{aligned}\right\} \quad \ldots (1)$$

in which the diagonal elements a_{ii} do not vanish.

Now the system (1) can be written as

$$\left.\begin{aligned} x_1 &= \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}}x_2 - \frac{a_{13}}{a_{11}}x_3 - \cdots - \frac{a_{1n}}{a_{11}}x_n \\ x_2 &= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}}x_1 - \frac{a_{23}}{a_{22}}x_3 - \cdots - \frac{a_{2n}}{a_{22}}x_n \\ x_3 &= \frac{b_3}{a_{33}} - \frac{a_{31}}{a_{33}}x_1 - \frac{a_{32}}{a_{33}}x_2 - \cdots - \frac{a_{2n}}{a_{33}}x_n \\ &\vdots \\ x_n &= \frac{b_n}{a_{nn}} - \frac{a_{n1}}{a_{nn}}x_1 - \frac{a_{n2}}{a_{nn}}x_2 - \cdots - \frac{a_{n,n-1}}{a_{nn}}x_{n-1} \end{aligned}\right\} \quad \ldots (2)$$

Suppose we start with $x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}$ as initial values to the variables $x_1, x_2, \ldots, x_n$. Then we can find better approximations to $x_1, x_2, \ldots, x_n$ using the following two iterative methods:

(i) Jacobi's iteration method

Jacobi's iteration method, also called the *method of simultaneous displacements*, is as follows:

Step 1: Determination of first approximation $x_1^{(1)},\ x_2^{(1)},\ \ldots, x_n^{(1)}$ using $x_1^{(0)},\ x_2^{(0)},\ \ldots, x_n^{(0)}$.

$$\left.\begin{aligned}
x_1^{(1)} &= \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}}x_2^{(0)} - \frac{a_{13}}{a_{11}}x_3^{(0)} - \ \ldots\ - \frac{a_{1n}}{a_{11}}x_n^{(0)} \\
x_2^{(1)} &= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}}x_1^{(0)} - \frac{a_{23}}{a_{22}}x_3^{(0)} - \ \ldots\ - \frac{a_{2n}}{a_{22}}x_n^{(0)} \\
x_3^{(1)} &= \frac{b_3}{a_{33}} - \frac{a_{31}}{a_{33}}x_1^{(0)} - \frac{a_{32}}{a_{33}}x_2^{(0)} - \ \ldots\ - \frac{a_{2n}}{a_{33}}x_n^{(0)} \\
&\ \ \vdots \\
x_n^{(1)} &= \frac{b_n}{a_{nn}} - \frac{a_{n1}}{a_{nn}}x_1^{(0)} - \frac{a_{n2}}{a_{nn}}x_2^{(0)} - \ \ldots\ - \frac{a_{n,n-1}}{a_{nn}}x_{n-1}^{(0)}
\end{aligned}\right\} \quad \ldots\ (3)$$

Step 2: Similarly, $x_1^{(2)},\ x_2^{(2)},\ \ldots, x_n^{(2)}$ are evaluated by just replacing $x_r^{(0)}$ in the right hand sides equations in (3) by $x_r^{(1)}$.

Step $n+1$: In general, if $x_1^{(n)},\ x_2^{(n)},\ \ldots, x_n^{(n)}$ are a system of nth approximations, then the next approximation is given by the formula

$$\left.\begin{aligned}
x_1^{(n+1)} &= \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}}x_2^{(n)} - \frac{a_{13}}{a_{11}}x_3^{(n)} - \ \ldots\ - \frac{a_{1n}}{a_{11}}x_n^{(n)} \\
x_2^{(n+1)} &= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}}x_1^{(n)} - \frac{a_{23}}{a_{22}}x_3^{(n)} - \ \ldots\ - \frac{a_{2n}}{a_{22}}x_n^{(n)} \\
x_3^{(n+1)} &= \frac{b_3}{a_{33}} - \frac{a_{31}}{a_{33}}x_1^{(n)} - \frac{a_{32}}{a_{33}}x_2^{(n)} - \ \ldots\ - \frac{a_{2n}}{a_{22}}x_n^{(n)} \\
&\ \ \vdots \\
x_n^{(n+1)} &= \frac{b_n}{a_{nn}} - \frac{a_{n1}}{a_{nn}}x_1^{(n)} - \frac{a_{n2}}{a_{nn}}x_2^{(n)} - \ \ldots\ - \frac{a_{n,n-1}}{a_{nn}}x_{n-1}^{(n)}
\end{aligned}\right\} \quad \ldots\ (4)$$

The system in (4) can also be briefly described as follows:

$$x_i^{(r+1)} = \frac{b_i}{a_{ii}} - \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{a_{ij}}{a_{ii}} x_j^{(r)} \qquad (r = 0,1,2,\ldots,\quad i = 1, 2, \ldots, n)$$

A sufficient condition for obtaining a solution by Jacobi's iteration method is the diagonal dominance,

i.e., $$|a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij}, \quad i = 1, 2, \ldots, n.$$

i.e., in each row of **A** the modulus of the diagonal element exceeds the sum of the off diagonal elements and also the diagonal elements $a_{ii} \neq 0$. If any diagonal element is 0, the equations can always be re-arranged to satisfy this condition.

(ii) Gauss Seidel iteration method

A simple modification to Jacobi's iteration method is given by *Gauss-Seidel* method.

Step 1 (*Gauss-Seidel method*): Determination of first approximation $x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}$ using $x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}$.

$$\left.\begin{aligned}
x_1^{(1)} &= \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}}x_2^{(0)} - \frac{a_{13}}{a_{11}}x_3^{(0)} - \quad \cdots \quad - \frac{a_{1n}}{a_{11}}x_n^{(0)} \\
x_2^{(1)} &= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}}x_1^{(1)} - \frac{a_{23}}{a_{22}}x_3^{(0)} - \quad \cdots \quad - \frac{a_{2n}}{a_{22}}x_n^{(0)} \\
x_3^{(1)} &= \frac{b_3}{a_{33}} - \frac{a_{31}}{a_{33}}x_1^{(1)} - \frac{a_{32}}{a_{33}}x_2^{(1)} - \quad \cdots \quad - \frac{a_{2n}}{a_{33}}x_n^{(0)} \\
&\vdots \\
x_n^{(1)} &= \frac{b_n}{a_{nn}} - \frac{a_{n1}}{a_{nn}}x_1^{(1)} - \frac{a_{n2}}{a_{nn}}x_2^{(1)} - \quad \cdots \quad - \frac{a_{n,n-1}}{a_{nn}}x_{n-1}^{(1)}
\end{aligned}\right\} \quad \ldots (5)$$

Step $n+1$: In general, if $x_1^{(n)}, x_2^{(n)}, \ldots, x_n^{(n)}$ are a system of nth approximations, then the next approximation is given by the formula

$$\left.\begin{aligned}
x_1^{(n+1)} &= \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}}x_2^{(n)} - \frac{a_{13}}{a_{11}}x_3^{(n)} - \quad \cdots \quad - \frac{a_{1n}}{a_{11}}x_n^{(n)} \\
x_2^{(n+1)} &= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}}x_1^{(n+1)} - \frac{a_{23}}{a_{22}}x_3^{(n)} - \quad \cdots \quad - \frac{a_{2n}}{a_{22}}x_n^{(n)} \\
x_3^{(n+1)} &= \frac{b_3}{a_{33}} - \frac{a_{31}}{a_{33}}x_1^{(n+1)} - \frac{a_{32}}{a_{33}}x_2^{(n+1)} - \quad \cdots \quad - \frac{a_{2n}}{a_{33}}x_n^{(n)} \\
&\vdots \\
x_n^{(n+1)} &= \frac{b_n}{a_{nn}} - \frac{a_{n1}}{a_{nn}}x_1^{(n+1)} - \frac{a_{n2}}{a_{nn}}x_2^{(n+1)} - \quad \cdots \quad - \frac{a_{n,n-1}}{a_{nn}}x_{n-1}^{(n+1)}
\end{aligned}\right\} \quad \ldots (6)$$

(6) can be briefly described as follows:

$$x_i^{(r+1)} = \frac{b_i}{a_{ii}} - \sum_{j=1}^{i-1}\frac{a_{ij}}{a_{ii}}x_j^{(r+1)} - \sum_{j=i+1}^{n}\frac{a_{ij}}{a_{ii}}x_j^{(r)} \quad (r = 0,1,2,\ldots, \quad i = 1, 2, \ldots, n).$$

Remark We note the difference between Jacobi's method and *Gauss-Seidel* method.

(**Attention!** In the following the bold face letters must be carefully noted):

Jacobi's method: In the first equation of (3), we substitute the initial approximations $x_2^{(0)}, x_3^{(0)}, \ldots, x_n^{(0)}$ into the right-hand side and denote the result as $x_1^{(1)}$. In the second equation, we substitute $\boldsymbol{x_1^{(0)}}, \boldsymbol{x_3^{(0)}}, \ldots, x_n^{(0)}$ and denote the result as $x_2^{(1)}$. In third, we substitute $\boldsymbol{x_1^{(0)}}, \boldsymbol{x_2^{(0)}}, \ldots, x_n^{(0)}$ and call the result as $x_3^{(1)}$. The process is repeated in this manner.

Gauss-Seidel method: In the first equation of (3), we substitute the initial approximation $x_2^{(0)}, \ldots, x_n^{(0)}$ into the right-hand side and denote the result as $x_1^{(1)}$. In the second equation, we substitute $\boldsymbol{x_1^{(1)}}, x_3^{(0)}, \ldots, x_n^{(0)}$ and denote the result as $x_2^{(1)}$. In third, we substitute $\boldsymbol{x_1^{(1)}}, \boldsymbol{x_2^{(1)}}, \ldots, x_n^{(0)}$ and call the result as $x_3^{(1)}$. The process is repeated in this manner and illustrated below:

***Example* 11** Solve the following system of equations using (a) Jacobi's iteration method and (b) Gauss-Seidel iteration method.

$$10x_1 - 2x_2 - x_3 - x_4 = 3$$

$$-2x_1 + 10x_2 - x_3 - x_4 = 15$$

$$-x_1 - x_2 + 10x_3 - 2x_4 = 27$$

$$-x_1 - x_2 - 2x_3 + 10x_4 = -9\,.$$

Solution

To solve these equations by the iterative methods, we re-write them as follows:

$$x_1 = 0.3 + 0.2x_2 + 0.1x_3 + 0.1x_4$$

$$x_2 = 1.5 + 0.2x_1 + 0.1x_3 + 0.1x_4$$

$$x_3 = 2.7 + 0.1x_1 + 0.1x_2 + 0.2x_4$$

$$x_4 = -0.9 + 0.1x_1 + 0.1x_2 + 0.2x_3$$

It can be verified that these equations satisfy the diagonal dominance condition. The process and given in the following Tables.

Table 1. Jacobi's Method

n	x_1	x_2	x_3	x_4
1	0.3	1.56	2.886	−0.1368
2	0.8869	1.9523	2.9566	−0.0248
3	0.9836	1.9899	2.9924	−0.0042
4	0.9968	1.9982	2.9987	−0.0008
5	0.9994	1.9997	2.9998	−0.0001
6	0.9999	1.9999	3.0	0.0
7	1.0	2.0	3.0	0.0

Table 2. Gauss-Seidel method

n	x_1	x_2	x_3	x_4
1	0.3	1.5	2.7	−0.9
2	0.78	1.74	2.7	−0.18
3	0.9	1.908	2.916	−0.108
4	0.9624	1.9608	2.9592	−0.036
5	0.9845	1.9848	2.9851	−0.0158
6	0.9939	1.9938	2.9938	−0.006
7	0.9975	1.9975	2.9976	−0.0025
8	0.9990	1.9990	2.9990	−0.0010
9	0.9996	1.9996	2.9996	−0.0004
10	0.9998	1.9998	2.9998	−0.0002
11	0.9999	1.9999	2.9999	−0.0001
12	1.0	2.0	3.0	0.0

From Tables 1 and 2, it is clear that twelve iterations are required by Jacobi's method to achieve the same accuracy as seven Gauss-Seidel iterations.

***Example* 12** Solve by Jacobi's iteration method, the system of equations

$$20x_1 + x_2 - 7x_3 = 17$$

$$3x_1 + 20x_2 - x_3 = -18$$

$$2x_1 - 3x_2 + 20x_3 = 25$$

Solution The given system of equations can be written as

$$\left.\begin{aligned} x_1 &= \frac{17}{20} - \frac{1}{20}x_2 + \frac{7}{20}x_3 \\ x_2 &= -\frac{18}{20} - \frac{3}{20}x_1 + \frac{1}{20}x_3 \\ x_3 &= \frac{25}{20} - \frac{2}{20}x_1 + \frac{3}{20}x_2 \end{aligned}\right\} \quad (3)$$

We start from an approximation $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = 0$ to x_1, x_2, x_3 respectively. Substituting these values on the right sides of equations in (3), we get the first approximation values $x_1^{(1)} = \frac{17}{20} = 0.85$, $x_2^{(1)} = -\frac{18}{20} = -0.90$ and $x_3^{(1)} = \frac{25}{20} = 1.25$

Putting these values on the right side of the equations in (2), we obtain the second approximation values, $x_1^{(2)} = 1.02$, $x_2^{(2)} = -0.965$ and $x_3^{(2)} = 1.03$. Similarly, third approximation values are $x_1^{(3)} = 1.00125$, $x_2^{(3)} = -1.0015$ and $x_3^{(3)} = 1.004$ and fourth approximation values are $x_1^{(4)} = 1.000475$, $x_2^{(4)} = -0.9999875$ and $x_3^{(4)} = 0.99965$. It can be seen that the values approach the exact solution $x_1 = 1$, $x_2 = -1$, $x_3 = 1$.

Example 13 Solve, using Gauss-Seidel iteration method, the system:

$$\begin{aligned} x_1 - 0.25x_2 - 0.25x_3 &= 50 \\ -0.25x_1 + x_2 - 0.25x_4 &= 50 \\ -0.25x_1 + x_3 - 0.25x_4 &= 25 \\ -0.25x_2 - 0.25x_3 + x_4 &= 25 \end{aligned}$$

Solution

The given system of equations can be written as

$$\left.\begin{aligned} x_1 &= 50 + 0.25x_2 + 0.25x_3 \\ x_2 &= 50 + 0.25x_1 + 0.25x_4 \\ x_3 &= 25 + 0.25x_1 + 0.25x_4 \\ x_4 &= 25 + 0.25x_2 + 0.25x_3 \end{aligned}\right\} \quad \ldots(2)$$

We start from an approximation $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = 100$ to x_1, x_2, x_3 respectively. Then we get approximation values as follows:

$$x_1^{(1)} = 50 + 0.25\, x_2^{(0)} + 0.25x_3^{(0)} = 100.00$$

$$x_2^{(1)} = 50 + 0.25x_1^{(1)} + 0.25x_4^{(0)} = 100.00 \qquad x_3^{(1)} = 50 + 0.25x_1^{(1)} + 0.25x_4^{(0)} = 75.00$$

$$x_4^{(1)} = 25 + 0.25x_2^{(1)} + 0.25x_3^{(1)} = 68.75$$

Now second approximation values are given by:

$$x_1^{(2)} = 50 + 0.25x_2^{(1)} + 0.25x_3^{(1)} = 93.75$$

$$x_2^{(2)} = 50 + 0.25x_1^{(2)} + 0.25x_4^{(1)} = 90.62 \qquad x_3^{(2)} = 50 + 0.25x_1^{(2)} + 0.25x_4^{(1)} = 65.62$$

$$x_4^{(2)} = 25 + 0.25x_2^{(2)} + 0.25x_3^{(2)} = 64.06\,.$$

Note that the exact solution to the system is

$$x_1 = x_2 = 87.5, \quad x_3 = x_4 = 62.5$$

Example 14 Using Gauss Siedel iteration solve the following system of equations, in three steps starting from 1, 1, 1.

$$10x + y + z = 6$$

$$x + 10y + z = 6$$

$$x + y + 10z = 6$$

Solution

$$x = 0.6 - 0.1y - 0.1z$$

$$y = 0.6 - 0.1x - 0.1z$$

$$z = 0.6 - 0.1x - 0.1y$$

Step 1 Using $x^{(0)} = y^{(0)} = z^{(0)} = 1$, we have

$x^{(1)} = 0.6 - 0.1\ y^{(0)} - 0.1\ z^{(0)} = 0.6 - 0.1 - 0.1 = 0.4$

$y^{(1)} = 0.6 - 0.1\ x^{(1)} - 0.1\ z^{(0)} = 0.6 - 0.1\times0.4 - 0.1 = 0.46$

$z^{(1)} = 0.6 - 0.1\ x^{(1)} - 0.1\ y^{(1)} = 0.6 - 0.1\times0.4 - 0.1\times0.46 = 0.514$

Step 2 Using $x^{(1)} = 0.4$, $y^{(1)} = 0.46$, $z^{(1)} = 0.514$, we have

$x^{(2)} = 0.6 - 0.1\ y^{(1)} - 0.1\ z^{(1)} = 0.6 - 0.1 \times 0.46 - 0.1\times0.514 = 0.5026$

$y^{(2)} = 0.6 - 0.1\ x^{(2)} - 0.1\ z^{(1)} = 0.6 - 0.1\times0.5026 - 0.1\times0.514 = 0.49834$

$z^{(2)} = 0.6 - 0.1\ x^{(2)} - 0.1\ y^{(2)}$

$= 0.6 - 0.1\times0.5026 - 0.1\times0.49834 = 0.499906$

Step 3 Using $x^{(2)} = 0.5026$, $y^{(2)} = 0.49834$, $z^{(2)} = 0.499906$, we have

$x^{(3)} = 0.6 - 0.1\ y^{(2)} - 0.1\ z^{(2)} = 0.6 - 0.1 \times 0.49834 - 0.1\times0.499906 = 0.5001754$

$y^{(3)} = 0.6 - 0.1\ x^{(3)} - 0.1\ z^{(2)}$

$= 0.6 - 0.1\times0.5001754 - 0.1\times0.499906 = 0.49999186$

$z^{(3)} = 0.6 - 0.1\ x^{(3)} - 0.1\ y^{(3)}$

$= 0.6 - 0.1\times0.5001754 - 0.1\times0.5001754 = 0.49996492$

We take $x \approx 5$, $y \approx 5$, $z \approx 5$ as the solution of the given system of equations.

Exercises

1. Apply Gauss Seidel iteration method to solve:

$$10x + 2y + z = 9$$

$$2x + 20y - 2z = -44$$

$$-2x + 3y + 10z = 22$$

2. Apply Gauss Seidel iteration method to solve:

$$1.2x + 2.1y + 4.2z = 9.9$$

$$5.3x + 6.1y + 4.7z = 21.6$$

$$9.2x + 8.3y + z = 15.2$$

3. Apply Jacobi's iteration method to solve:

$$5x - y + z = 10$$

$$2x - y + z = 10$$

$$x + y + 5z = -1$$

4. Apply Jacobi's iteration method to solve:

$$5x + 2y + z = 12$$

$$x + 4y + 2z = 15$$

$$x + 2y + 5z = 20$$

Answers

1. $x = 1.013, y = -1.996, z = 3.001$

2. $x = 2, y = 3, z = 4$ (Approximately)

3. $x = -13.223,\ y = 16.766,\ z = -2.306$

4. $x = 2.556, y = 1.722,\ z = -1.005$

5. $x = 1.08,\ y = 1.95,\ z = 3.16$

Curve Fitting

- *Curve fitting* describes techniques to fit curves at points between the discrete values to obtain intermediate estimates.
- Two general approaches for curve fitting:

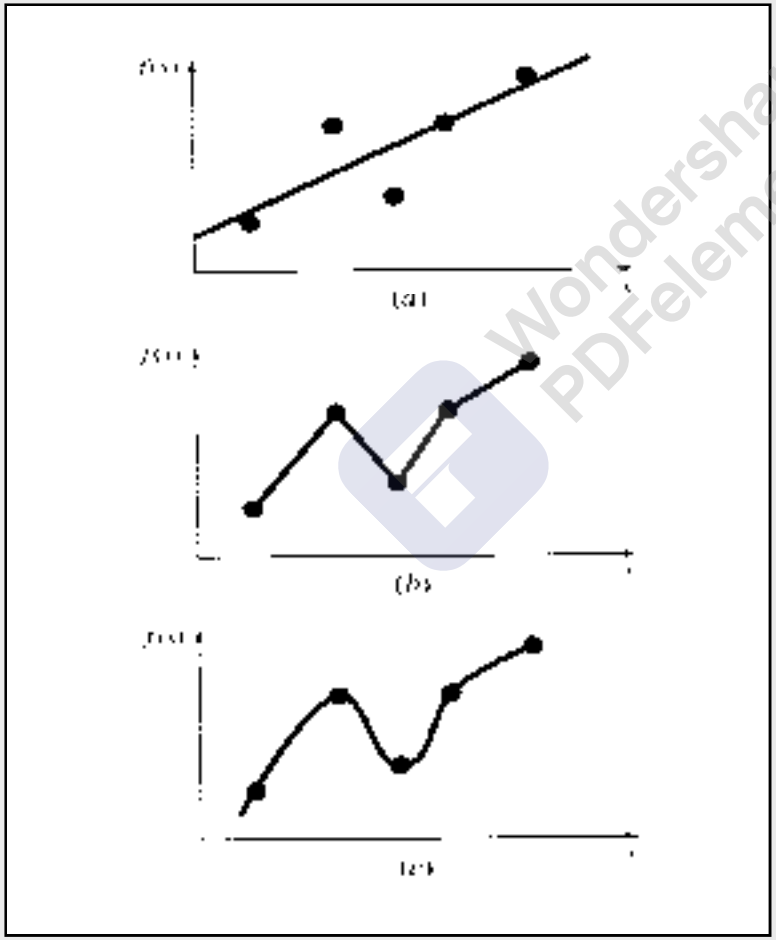
a) Least –Squares Regression - to fits the shape or general trend by sketch a best line of the data without necessarily matching the individual points (figure PT5.1, pg 426).

- 2 types of fitting:

i) Linear Regression

ii) Polynomial Regression

Figure shows sketches developed from same set of data by 3 engineers.

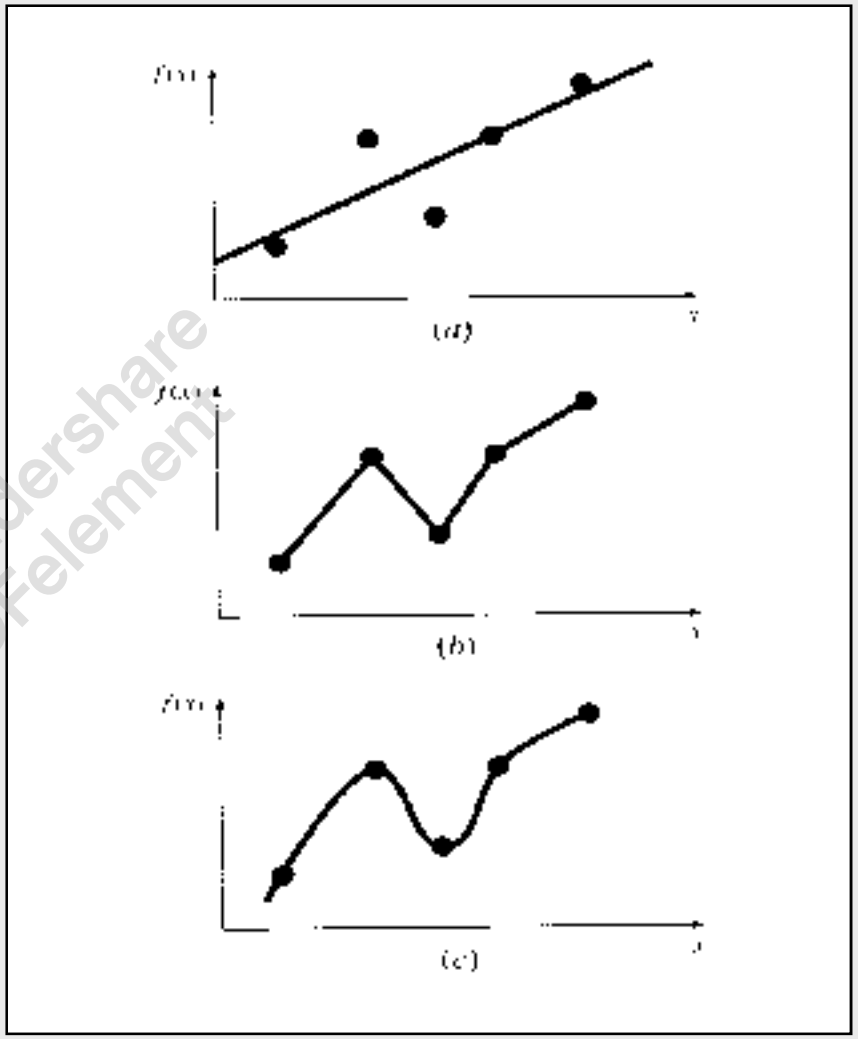


a) least-squares regression - did not attempt to connect the point, but characterized the general upward trend of the data with a straight line

b) Linear interpolation - Used straight-line segments or linear interpolation to connect the points. Very common practice in engineering. If the values are close to being linear, such approximation provides estimates that are adequate for many engineering calculations. However, if the data is widely spaced, significant errors can be introduced by such linear interpolation.

c) Curvilinear interpolation - Used curves to try to capture suggested by the data.

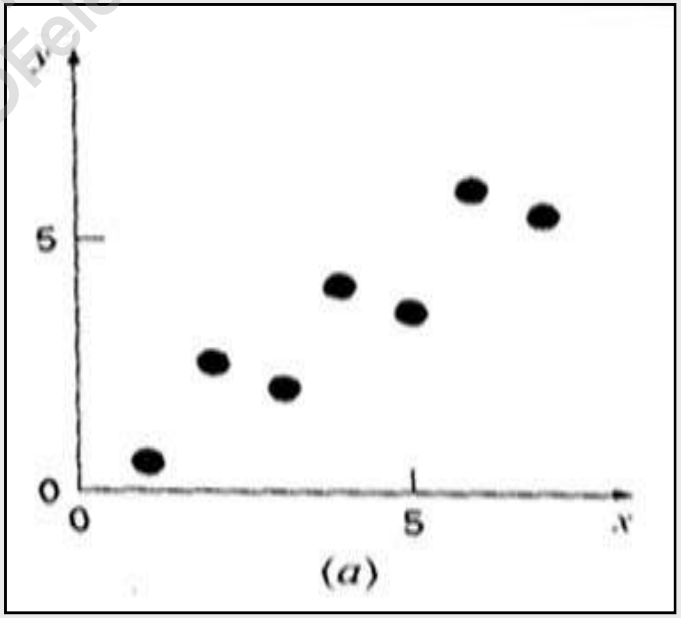
❖ **Our goal here to develop systematic and objective method deriving such curves.**



a) Least-square Regression : i) Linear Regression

- Is used to minimize the discrepancy/differences between the data points and the curve plotted. Sometimes, polynomial interpolation is inappropriate and may yield unsatisfactory results when used to predict intermediate values (see Fig. 17.1, pg 455).

Fig. 17.1 a): shows 7 experimentally derived data points exhibiting significant variability. Data exhibiting significant error.



Curve Fitting

- Linear Regression is fitting a '*best*' straight line through the points.
- The mathematical expression for the straight line is:

$$y = a_0 + a_1 x + e \qquad \text{Eq 17.1}$$

where, a_1- slope

a_0- intercept

e - error, or residual, between the model and the observations

- Rearranging the eq. above as:

$$e = y - a_0 - a_1 x$$

- Thus, the error or residual, is the discrepancy between the true value y and the approximate value, $a_0 + a_1 x$, predicted by the linear equation.

Criteria for a 'best' Fit

- To know how a "best" fit line through the data is by **minimize the sum of residual error**, given by ;

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i) \qquad \text{----- Eq 17.2}$$

where; n : total number of points

- A strategy to overcome the shortcomings: The sum of the squares of the errors between the measured y and the y calculated with the linear model is shown in Eq 17.3;

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_{i,measured} - y_{i,\mathrm{model}})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \qquad \text{----- Eq 17.3}$$

Least-squares fit for a straight line

- To determine values for a_o and a_1, i) differentiate equation 17.3 with respect to each coefficient, ii) setting the derivations equal to zero (minimize S_r), iii) set $\boldsymbol{\Sigma a_o = n.a_o}$ to give equations 17.4 and 17.5, called as *normal equations,* *(refer text book)* which can be solved simultaneously for a_1 and a_o;

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

----- Eq 17.6

$$a_0 = \bar{y} - a_1 \bar{x}$$

----- Eq 17.7

Example 1

Use least-squares regression to fit a straight line to:

x	1	2	3	4	5	6	7
y	0.5	2.5	2.0	4.0	3.5	6.0	5.5

- Two criteria for least-square regression will provide the best estimates of a_o and a_1 called *maximum likelihood principle* in statistics:
  i. The spread of the points around the line of similar magnitude along the entire range of the data.
  ii. The distribution of these points about the line is normal.

- If these criteria are met, a "*standard deviation*" for the regression line is given by equation:

---------- Eq. 17.9

$$S_{y/x} = \sqrt{\frac{S_r}{n-2}}$$

$s_{y/x}$: ***standard error of estimate***

"y/x" : predicted value of y corresponding to a particular value of x

n -2 : two data derived estimates a_o and a_1 were used to compute S_r (we have lost 2 degree of freedom)

- Equation 17.9 is derived from *Standard Deviation (S_y)* about the mean :

$$S_y = \sqrt{\frac{S_t}{n-1}} \qquad \text{-------- (PT5.2, pg 442)}$$

$$S_t = \sum (y_i - \bar{y})^2 \qquad \text{-------- (PT5.3, pg 442)}$$

S_t : total sum of squares of the residuals between data points and the mean.

- Just as the case with the standard deviation, the standard error of the estimate quantifies the spread of the data.

Estimation of error in summary

1. Standard Deviation

$$S_y = \sqrt{\frac{S_t}{n-1}}$$ ----- (PT5.2, pg 442)

$$S_t = \sum (y_i - \bar{y})^2$$ ----- (PT5.3, pg 442)

2. Standard error of the estimate

$$S_r = \sum_{i=1}^{n} ei^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x)^2$$ ----- Eq 17.8

$$S_{y/x} = \sqrt{\frac{S_r}{n-2}}$$ ----- Eq 17.9

where, y/x designates that the error is for a predict value of y corresponding to a particular value of x.

3. Determination coefficient

$$r^2 = \frac{S_t - S_r}{S_t}$$ ----- Eq 17.10

4. Correlation coefficient

$$r = \sqrt{\frac{S_t - S_r}{S_t}} \; @ \; r = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$ ----- Eq 17.11

Example 2

Use least-squares regression to fit a straight line to:

x	1	2	3	4	5	6	7
y	0.5	2.5	2.0	4.0	3.5	6.0	5.5

Compute the standard deviation (S_y), the standard error of estimate ($S_{y/x}$) and the correlation coefficient (r) for data above (use Example 1 result)

Work with your buddy and lets do Quiz 1

Use least-squares regression to fit a straight line to:

x	1	2	3	4	5	6	7	8	9
y	1	1.5	2	3	4	5	8	10	13

Compute the standard error of estimate ($S_{y/x}$) and the correlation coefficient (r)

Quiz 2

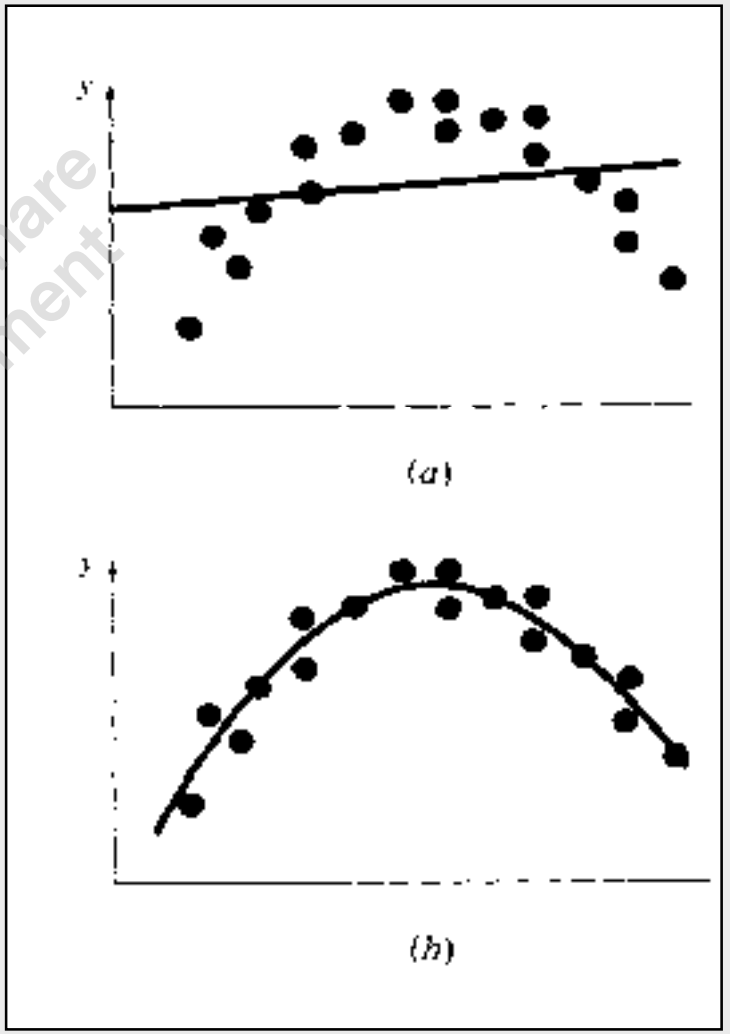
Compute the standard error of the estimate and the correlation coefficient.

x	0.25	0.75	1.25	1.50	2.00
y	-0.45	-0.60	0.70	1.88	6.00

Linearization of Nonlinear Relationships

- Linear regression provides a powerful technique for fitting the *best line* to data, where the relationship between the dependent and independent variables is linear.
- But, this is not always the case, thus first step in any regression analysis should be to plot and visually inspect whether the data is a linear model or not.

Figure 17.8: a) data is ill-suited for linear regression, b) parabola is preferable.



Nonlinear Relationships

- Linear regression is predicated on the fact that the relationship between the dependent and independent variables is linear - this is not always the case.
- Three common examples are:

$$\text{exponential}: \quad y = \alpha_1 e^{\beta_1 x}$$

$$\text{power}: \quad y = \alpha_2 x^{\beta_2}$$

$$\text{saturation-growth-rate}: \quad y = \alpha_3 \frac{x}{\beta_3 + x}$$

Linearization of Nonlinear Relationships

- One option for finding the coefficients for a nonlinear fit is to linearize it. For the three common models, this may involve taking logarithms or inversion:

Model	Nonlinear	Linearized
exponential :	$y = \alpha_1 e^{\beta_1 x}$	$\ln y = \ln \alpha_1 + \beta_1 x$
power :	$y = \alpha_2 x^{\beta_2}$	$\log y = \log \alpha_2 + \beta_2 \log x$
saturation - growth - rate :	$y = \alpha_3 \frac{x}{\beta_3 + x}$	$\frac{1}{y} = \frac{1}{\alpha_3} + \frac{\beta_3}{\alpha_3} \frac{1}{x}$

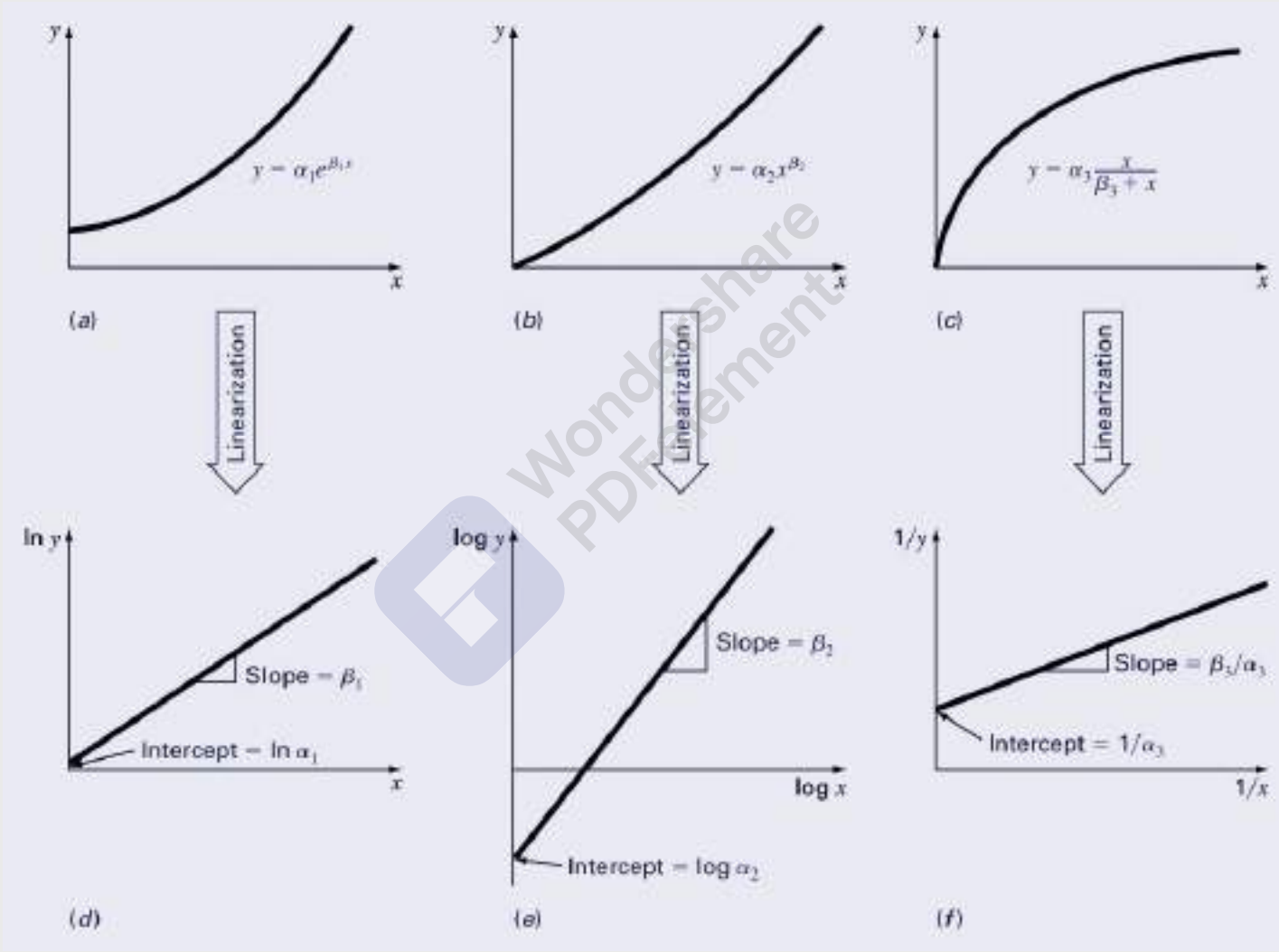
Linearization of Nonlinear Relationships

- After linearization, Linear regression can be applied to determine the linear relation.
- For example, the linearized exponential equation:

$$\ln y = \ln \alpha_1 + \beta_1 x$$

$$y \qquad a_0 \qquad a_1 x$$

Figure 17.9: Type of polynomial equations and their linearized versions, respectively.



- Fig. 17.9, pg 453 shows population growth of radioactive decay behavior.

Fig. 17.9 (a) : the **exponential model**

$$y = \alpha_1 e^{\beta_1 x} \qquad \text{------ (17.12)}$$

α_1 , β_1 : constants, $\beta_1 \neq 0$

This model is used in many fields of engineering to characterize quantities.

Quantities increase : β_1 positive

Quantities decrease : β_1 negative

Example 2

Fit an exponential model $y = a\, e^{bx}$ to:

x	0.4	0.8	1.2	1.6	2.0	2.3
y	750	1000	1400	2000	2700	3750

Solution

- Linearized the model into;

$$\underbrace{\ln y}_{y} = \underbrace{\ln a}_{a_0} + \underbrace{bx}_{a_1 x}$$

$$y = a_0 + a_1 x \qquad \text{----- (Eq. 17.1)}$$

- Build the table for the parameters used in eqs 17.6 and 17.7, as in example 17.1, pg 444.

x_i	y_i	$\ln y_i$	x_i^2	$(x_i)(\ln y_i)$
0.4	750	6.620073	0.16	2.648029
0.8	1000	6.900775	0.64	5.520620
1.2	1400	7.244228	1.44	8.693074
1.6	2000	7.600902	2.56	12.161443
2.0	2700	7.901007	4.00	15.802014
2.3	3750	8.229511	5.29	18.927875
Σ 8.3		**44.496496**	**14.09**	**63.753055**

$$n = 6$$

$$\sum_{i=1}^{n} x_i = 8.3 \qquad \sum_{i=1}^{n} \ln y_i = 44.496496$$

$$\sum_{i=1}^{n} x_i^2 = 14.09 \qquad \sum_{i=1}^{n} (x_i)(\ln y_i) = 63.753055$$

$$\bar{x} = \frac{8.3}{6} = 1.383333 \qquad \overline{\ln y} = \frac{44.496496}{6} = 7.416083$$

$$a_0 = \ln a = \ln \bar{y} - b\bar{x} = 7.416083 - (0.843)(1.383333)$$

$$\ln a = 6.25$$

$$a_1 = b = \frac{n\Sigma(x_i)(\ln y_i) - \Sigma x_i \Sigma(\ln y_i)}{n\Sigma x_i^2 - (\Sigma x_i)^2}$$

$$b = \frac{(6)(63.753055) - (8.3)(44.496496)}{(6)(14.09) - (8.3)^2} = 0.843$$

Straight-line:

$$\ln y = \ln a + bx$$

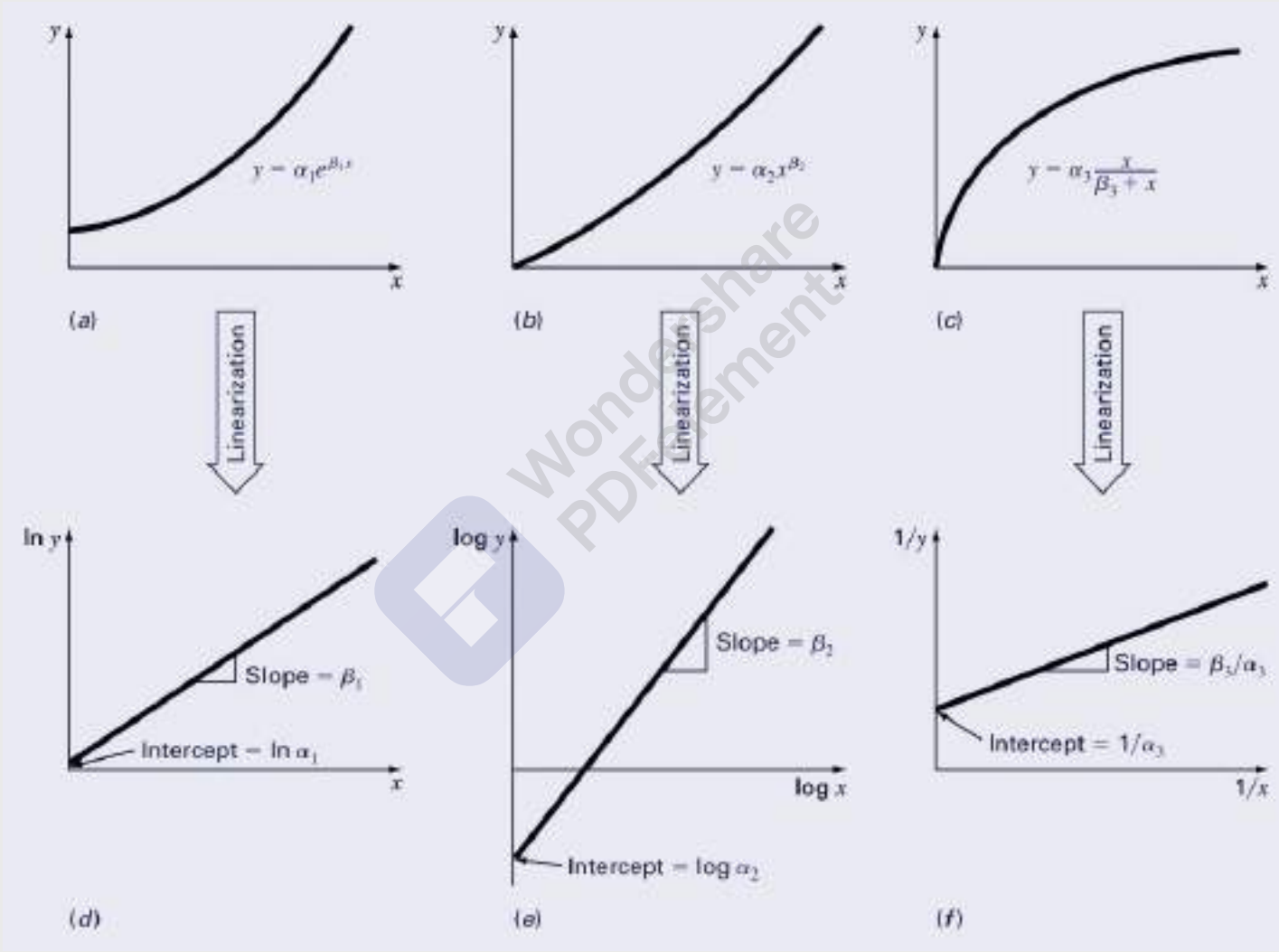
$$\therefore \ln y = 6.25 + 0.843x$$

Exponential: $y = a e^{bx}$

$$\ln a = 6.25 \quad \Rightarrow \quad a = e^{6.25} = 518$$

$$\therefore \quad y = a e^{bx} = 518 e^{0.843x}$$

Figure 17.9: Type of polynomial equations and their linearized versions, respectively.



Power Equation

- Equation (17.13) can be linearized by taking base-10 logarithm to yield:

$$y = \alpha_2 x^{\beta_2} \qquad \text{-------- (17.13)}$$

$$\log y = \log \alpha_2 + \beta_2 \log x \qquad \text{-------- (17.16)}$$

- A plot of ***log y*** versus ***log x*** will yield a straight line with slope of β_2 and an intercept of ***log*** α_2.

Example 4

Linearization of a Power equation and fit equation (17.13) to the data in table below using a logarithmic transformation of the data.

x	1	2	3	4	5
y	0.5	1.7	3.4	5.7	8.4

x_i	y_i	$\log x_i$	$\log y_i$	$(\log x_i)^2$	$(\log x_i)(\log y_i)$
1	0.5	0	-0.301	0	0
2	1.7	0.301	0.226	0.090601	0.068026
3	3.4	0.477	0.534	0.227529	0.254718
4	5.7	0.602	0.753	0.362404	0.453306
5	8.4	0.699	0.922	0.488601	0.644478
Σ		**2.079**	**2.134**	**1.169135**	**1.420528**

$$n = 5$$

$$\sum_{i=1}^{n} \log x_i = 2.079 \qquad \sum_{i=1}^{n} \log y_i = 2.134$$

$$\sum_{i=1}^{n} (\log x_i)^2 = 1.169135 \qquad \sum_{i=1}^{n} (\log x_i)(\log y_i) = 1.420528$$

$$\overline{\log x} = \frac{2.079}{5} = 0.4158 \qquad \overline{\log y} = \frac{2.134}{5} = 0.4268$$

$$b = \frac{n\Sigma(\log\ x_i)(\log\ y_i) - (\Sigma \log\ x_i)(\Sigma \log\ y_i)}{n\Sigma(\log\ x_i)^2 - (\Sigma \log\ x_i)^2}$$

$$b = \frac{(5)(1.420528\ ) - (2.079\ )(2.134\ )}{(5)(1.169135\ ) - (2.079\ )^2} = 1.75$$

$$\log a = \log \bar{y} - b(\log \bar{x}) = 0.4268 - (1.75)(0.4158)$$

$$\ln a = -0.3$$

Straight-line:

$$\log y = \log a + b \log x$$

$$\therefore \quad \log y = -0.3 + 1.75 \log x$$

Power: $y = a x^b$

$$\log a = -0.3 \quad \Rightarrow \quad a = 10^{-0.3} = 0.5$$

$$\therefore \quad y = a x^b = 0.5 x^{1.75}$$

- Fig. 17.10 a), pg 455, is a plot of the original data in its untransformed state, while fig. 17.10 b) is a plot of the transformed data.
- The intercept, log α_2 = -0.300, and by taking the antilogarithm, α_2 = $10^{-0.3}$ = 0.5.
- The slope is β_2 = 1.75, consequently, the power equation is : ***y = 0.5x^1.75***

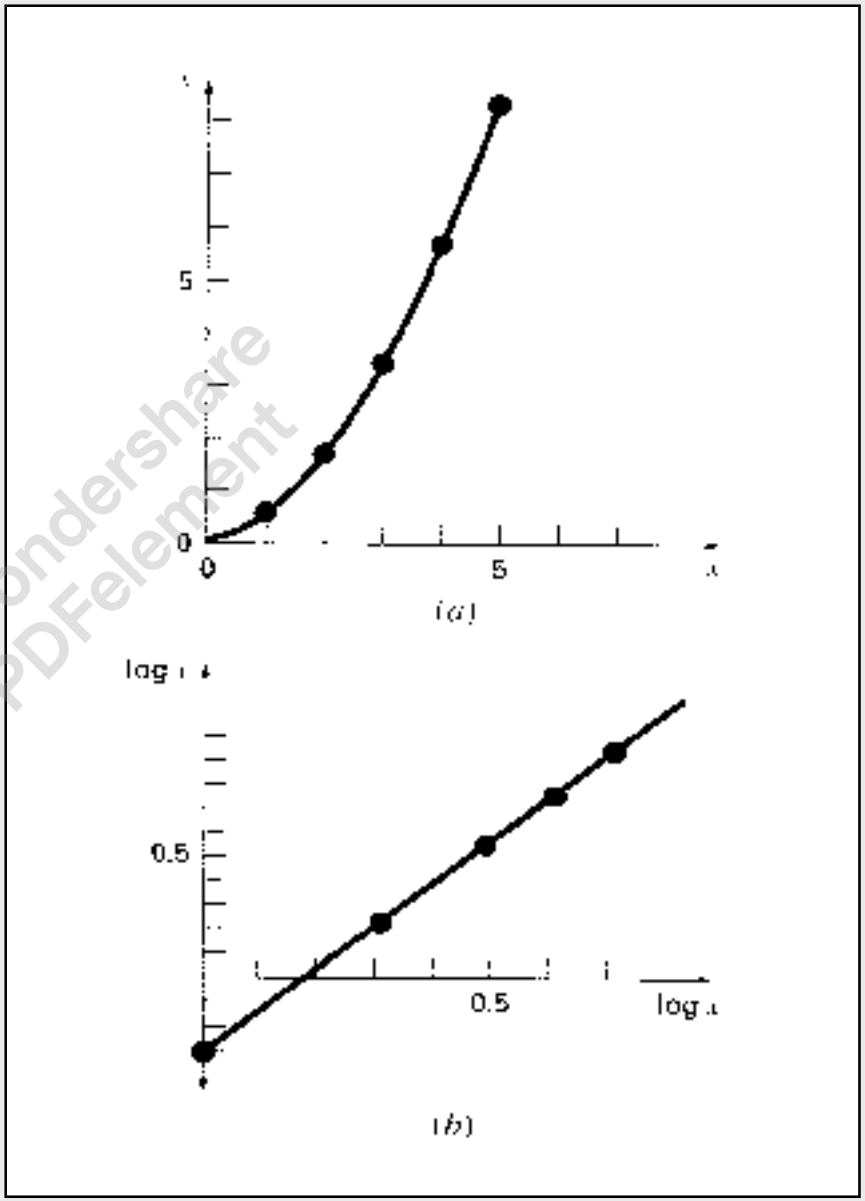
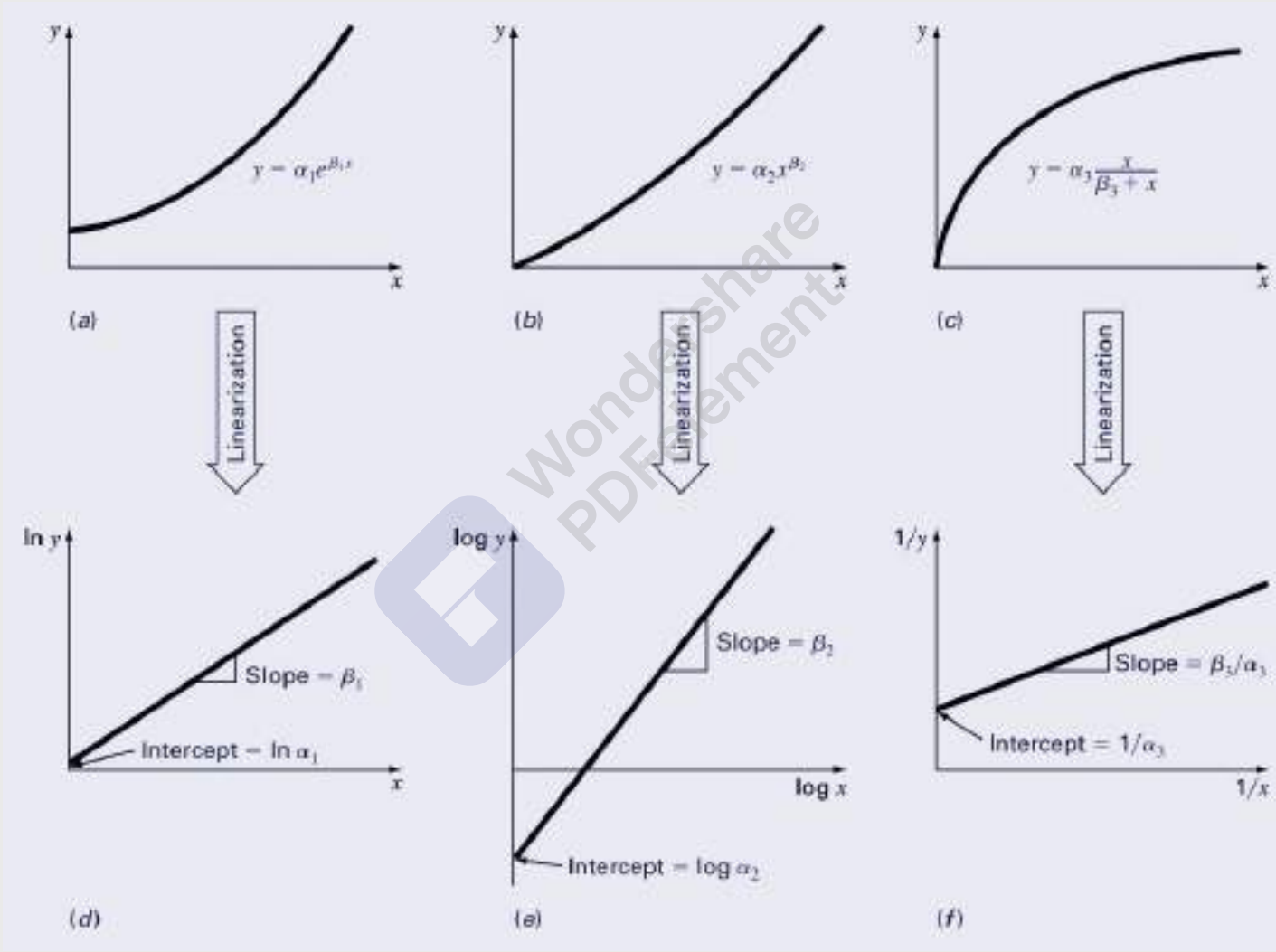


Figure 17.9: Type of polynomial equations and their linearized versions, respectively.



Saturation growth rate equation

- Equation (17.14) can be linearized by inverting it to yield:

$$y = \alpha_3 \left[\frac{x}{\beta_3 + x} \right] \qquad \text{------- (17.14)}$$

$$\frac{1}{y} = \frac{\beta_3}{\alpha_3} \frac{1}{x} + \frac{1}{\alpha_3} \qquad \text{------- (17.17)}$$

- A plot of $1/y$ versus $1/x$ will yield a straight line with slope of β_3/α_3 and an intercept of $1/\alpha_3$
- In their transformed forms, these models are fit using linear regression in order to evaluate the constant coefficients.
- This model well-suited for characterizing population growth under limiting conditions.

Example 5

Linearization of a saturation-growth rate equation to the data in table below.

x	0.75	2	2.5	4	6	8	8.5
y	0.8	1.3	1.2	1.6	1.7	1.8	1.7

$$n = 7$$

$$\sum_{i=1}^{n} \frac{1}{x_i} = 2.8926 \qquad \sum_{i=1}^{n} \frac{1}{y_i} = 5.2094$$

$$\sum_{i=1}^{n} \left(\frac{1}{x_i}\right)^2 = 2.3074 \qquad \sum_{i=1}^{n} \left(\frac{1}{x_i}\right)\left(\frac{1}{y_i}\right) = 2.8127$$

$$\overline{\left(\frac{1}{x}\right)} = \frac{2.8926}{7} = 0.4132 \qquad \overline{\left(\frac{1}{y}\right)} = \frac{5.2094}{7} = 0.7442$$

x_i	y_i	$1/x_i$	$1/y_i$	$(1/x_i)^2$	$(1/x_i)(1/y_i)$
0.75	0.8	1.33333	1.25000	1.7777	1.6666
2	1.3	0.50000	0.76923	0.2500	0.3846
2.5	1.2	0.40000	0.83333	0.1600	0.3333
4	1.6	0.25000	0.62500	0.0625	0.1562
6	1.7	0.16667	0.58823	0.0278	0.0981
8	1.8	0.12500	0.55555	0.0156	0.0694
8.5	1.7	0.11765	0.58823	0.0138	0.1045
Σ		**2.89260**	**5.20940**	**2.3074**	**2.8127**

$$\frac{b}{a} = \frac{n\Sigma\left(\frac{1}{x_i}\right)\left(\frac{1}{y_i}\right) - \Sigma\left(\frac{1}{x_i}\right)\Sigma\left(\frac{1}{y_i}\right)}{n\Sigma\left(\frac{1}{x_i}\right)^2 - \left(\Sigma\left(\frac{1}{x_i}\right)\right)^2}$$

$$\frac{b}{a} = \frac{(7)(2.8127) - (2.8926)(5.2094)}{(7)(2.3074) - (2.8926)^2} = 0.5935$$

$$\left(\frac{1}{a}\right) = \overline{\left(\frac{1}{y}\right)} - \frac{b}{a}\overline{\left(\frac{1}{x}\right)} = 0.7442 - (0.5935)(0.4132)$$

$$\left(\frac{1}{a}\right) = 0.4990$$

Straight-line: $\frac{1}{y} = \frac{1}{a} + \frac{b}{a}\frac{1}{x}$

$$\therefore \frac{1}{y} = 0.4990 + 0.5935\frac{1}{x}$$

Saturation-growth: $y = a\left[\frac{x}{b+x}\right]$

$$\frac{1}{a} = 0.4990 \quad \Rightarrow \quad \therefore \; a = 2$$

$$\frac{b}{a} = 0.5935 \quad \Rightarrow \quad \therefore \; b = (0.5935)(2) = 1.187$$

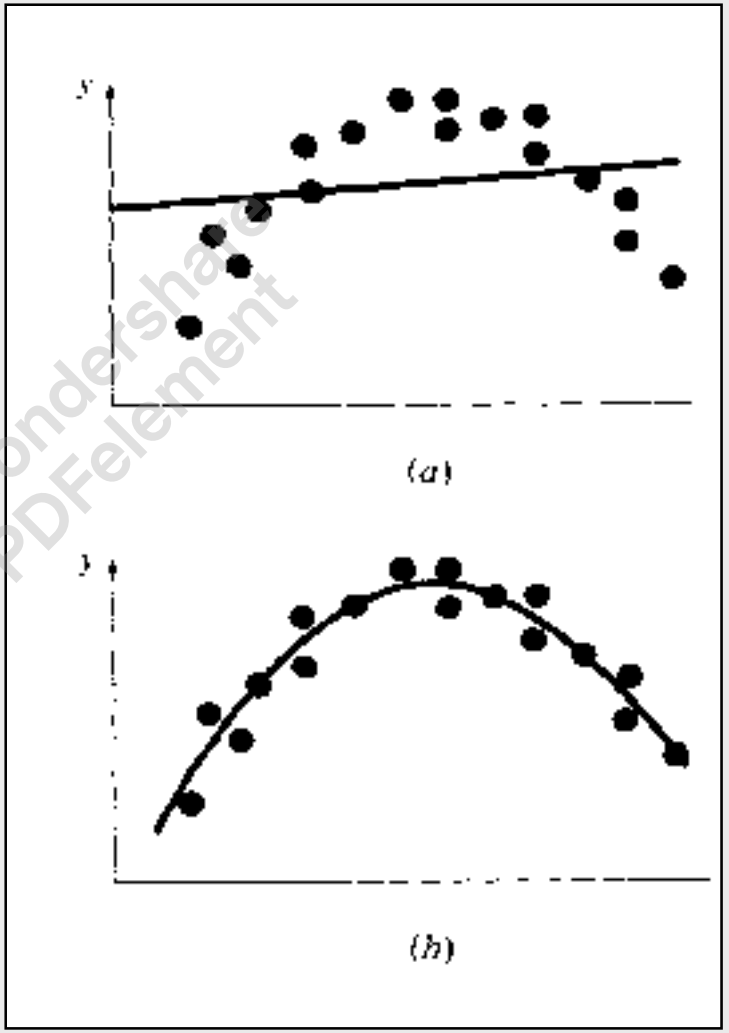
$$\therefore \quad y = 2\left[\frac{x}{1.187 + x}\right]$$

Lets do Quiz 3

Fit a power equation and saturation growth rate equation to:

x	1	2	3	4	5	6	7
y	2.1	2.2	2.3	2.4	2.5	2.6	2.7

Figure 17.8: a) data is ill-suited for linear regression, b) parabola is preferable.



Polynomial Regression

- Another alternative is to fit polynomials to the data using *polynomial regression.*

- The least-squares procedure can be readily extended to fit the data to a higher-order polynomial.

- For example, to fit a second–order polynomial or quadratic:

$$y = a_0 + a_1 x + a_2 x^2 + e$$

- The sum of the squares of the residual is:

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2$$

where n= total number of points

- Then, taking the derivative of equation (17.18) with respect to each of the unknown coefficients, a_o , a_1 ,and, a_2 of the polynomial, as in:

$$\frac{\delta S_r}{\delta a_0} = -2\sum (y_i - a_0 - a_1 x_i - a_2 x_i^2)$$

$$\frac{\delta S_r}{\delta a_1} = -2\sum x_i (y_i - a_0 - a_1 x_i - a_2 x_i^2)$$

$$\frac{\delta S_r}{\delta a_2} = -2\sum x_i^2 (y_i - a_0 - a_1 x_i - a_2 x_i^2)$$

- Setting the equations equal to zero and rearrange to develop set of normal equations and by setting $\Sigma a_o = n.a_o$

$$(n)a_0 + (\sum x_i)a_1 + (\sum x_i^2)a_2 = \sum y_i$$

$$(\sum x_i)a_0 + (\sum x_i^2)a_1 + (\sum x_i^3)a_2 = \sum x_i y_i \qquad \text{----- 17.19}$$

$$(\sum x_i^2)a_0 + (\sum x_i^3)a_1 + (\sum x_i^4)a_2 = \sum x_i^2 y_i$$

- The above 3 equations are linear with 3 unknowns coefficients (a_o , a_1 ,*and,* a_2) which can be calculated directly from observed data.
- In matrix form:

$$\begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{bmatrix} \times \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{bmatrix}$$

- The two-dimensional case can be easily extended to an ***m***th-order polynomial as:

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m + e$$

- Thus, standard error for ***m***th-order polynomial :

$$S_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}} \qquad \text{----- 17.20}$$

Example 6

Fit a second order polynomial to the data in the first 2 columns of table 17.4:

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
0	2.1	0	0	0	0	0
1	7.7	1	1	1	7.7	7.7
2	13.6	4	8	16	27.2	54.4
3	27.2	9	27	81	81.6	244.8
4	40.9	16	64	256	163.6	654.4
5	61.1	25	125	625	305.5	1527.5
Σ 15	**152.6**	**55**	**225**	**979**	**585.6**	**2488.8**

- From the given data:

$m = 2$ $\quad \Sigma x_i = 15$ $\quad \Sigma x_i^4 = 979$ $\quad \bar{y} = 25.433$

$n = 6$ $\quad \Sigma y_i = 152.6$ $\quad \Sigma x_i y_i = 585.6$ $\quad \Sigma x_i^3 = 225$

$\bar{x} = 2.5$ $\quad \Sigma x_i^2 = 55$ $\quad \Sigma x_i^2 y_i = 2488.8$

$$\begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{bmatrix} \times \begin{bmatrix} a_0 \\ a_1 \\ a_3 \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{bmatrix}$$

- Therefore, the simultaneous linear equations are:

$$\begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \begin{Bmatrix} a_o \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 152.6 \\ 585.6 \\ 2488.8 \end{Bmatrix}$$

- Solving these equations through a technique such as Gauss elimination gives:

$a_o = 2.47857$, $a_1 = 2.35929$, and $a_2 = 1.86071$

- Therefore, the least-squares quadratic equation for this case is:

$$y = 2.47857 + 2.35929x + 1.86071x^2$$

- To calculate s_t and s_r , build table 17.4 for columns 3 and 4.

x_i	y_i	$(y_i - \bar{y})^2$	$(y_i - a_o - a_1 x_i - a_2 x_i^2)^2$
0	2.1	544.44	0.14332
1	7.7	314.47	1.00286
2	13.6	140.03	1.08158
3	27.2	3.12	0.80491
4	40.9	239.22	0.61951
5	61.1	1272.11	0.09439
Σ	**152.6**	**2513.39**	**3.74657**

$$S_t = \Sigma(y_i - \bar{y}) = 2513.39$$

$$S_r = \Sigma(y_i - a_0 - a_1 x_i - a_2 x_i^{\,2})^2 = 3.74657$$

The standard error (regression polynomial):

$$S_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}} = \sqrt{\frac{3.74657}{6-(2+1)}} = 1.12$$

- The correlation coefficient can be calculated by using equations 17.10 and 17.11, respectively:

$$r^2 = \frac{S_t - S_r}{S_t}$$

$$r = \sqrt{\frac{S_t - S_r}{S_t}} \; @ \; r = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

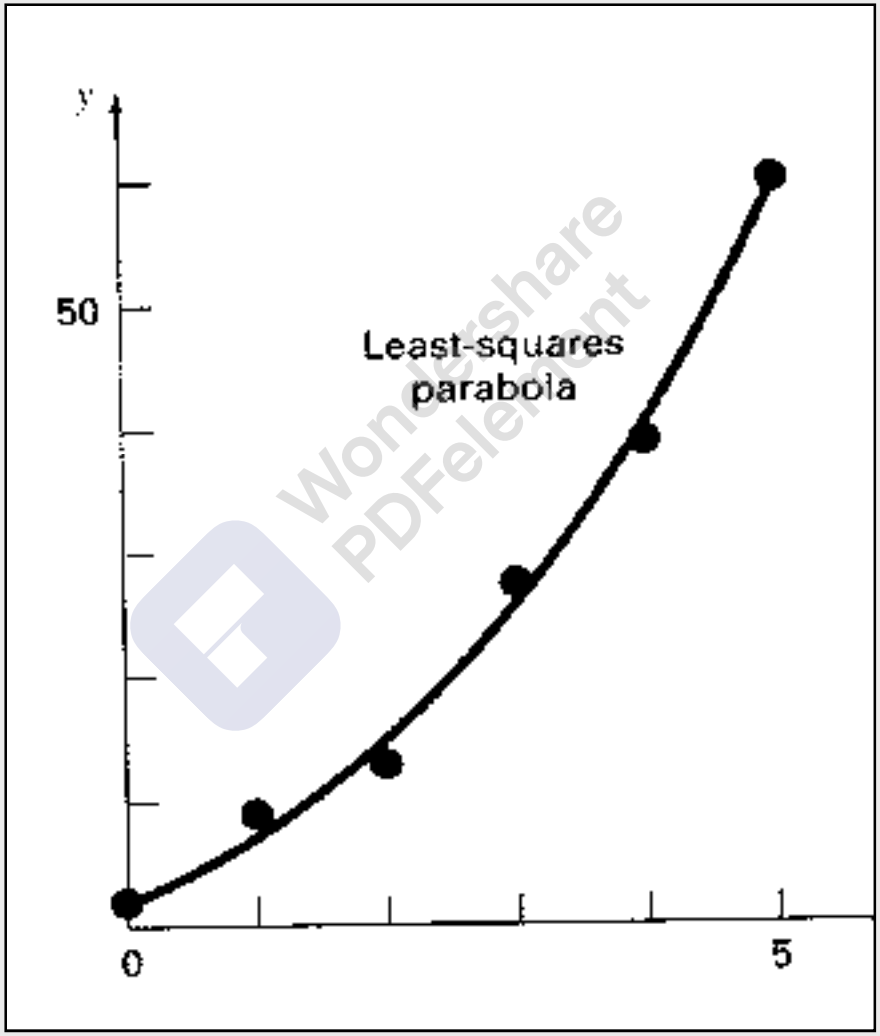
Therefore, $r^2 = (S_t - S_r) / S_t = (2513.39 - 3.74657) / 2513.39$

$$\mathbf{r^2 = 0.99851}$$

$\therefore$ The correlation coefficient is, $\mathbf{r = 0.99925}$

- The results indicate that 99.851% of the original uncertainty has been explained by the model. This result supports the conclusion that the quadratic equation represents an excellent fit, as evident from Fig.17.11.

Figure 17.11: fit of a second-order polynomial



UNIT 5

NUMERICAL INTEGRATION

Introduction

The problem of numerical integration is to find an approximate value of the integral

$$I = \int_a^b w(x)\, f(x)\, dx$$

where $w(x) > 0$ in (a, b) is called the *weight function*. The function $f(x)$ may be given explicitly or as a tabulated data. We assume that $w(x)$ and $w(x)\, f(x)$ are integrable on $[a, b]$. The limits of integration may be finite, semi-infinite or infinite. The integral is approximated by a linear combination of the values of $f(x)$ at the tabular points as

$$I = \int_a^b w(x)\, f(x)\, dx = \sum_{k=0}^{n} \lambda_k\, f(x_k)$$

$$= \lambda_0 f(x_0) + \lambda_1 f(x_1) + \lambda_2 f(x_2) + \ldots + \lambda_n f(x_n).$$

The tabulated points x_k's are called *abscissas*, $f(x_k)$'s are called the ordinates and λ_k's are called the weights of the *integration rule* or *quadrature formula* (3.26).

We define the error of approximation for a given method as

$$R_n(f) = \int_a^b w(x)\, f(x)\, dx - \sum_{k=0}^{n} \lambda_k\, f(x_k).$$

Order of a method An integration method of the form (3.26) is said to be of order p, if it produces exact results, that is $R_n = 0$, for all polynomials of degree less than or equal to p. That is, it produces exact results for $f(x) = 1, x, x^2, \ldots, x^p$. This implies that

$$R_n(x^m) = \int_a^b w(x)\, x^m\, dx - \sum_{k=0}^{n} \lambda_k\, x_k^{\,m} = 0, \quad \text{for } m = 0, 1, 2, \ldots, p.$$

The error term is obtained for $f(x) = x^{p+1}$. We define

$$c = \int_a^b w(x)\, x^{p+1}\, dx - \sum_{k=0}^{n} \lambda_k\, x_k^{p+1}$$

where c is called the *error constant*. Then, the error term is given by

$$R_n(f) = \int_a^b w(x)\, f(x)\, dx - \sum_{k=0}^{n} \lambda_k f(x_k)$$

$$= \frac{c}{(p+1)!}\, f^{(p+1)}(\xi), \quad a < \xi < b.$$

The bound for the error term is given by

$$| R_n(f) | \leq \frac{|c|}{(p+1)!} \max_{a \leq x \leq b} | f^{(p+1)}(x) |.$$

If $R_n(x^{p+1})$ also becomes zero, then the error term is obtained for $f(x) = x^{p+2}$.

Integration Rules Based on Uniform Mesh Spacing

When $w(x) = 1$ and the nodes x_k's are prescribed and are equispaced with $x_0 = a$, $x_n = b$, where $h = (b - a)/n$, the methods (3.26) are called *Newton-Cotes integration rules.* The weights λ_k's are called *Cotes numbers*.

We shall now derive some Newton-Cotes formulas. That is, we derive formulas of the form

$$I = \int_a^b f(x)\, dx = \sum_{k=0}^{n} \lambda_k f(x_k)$$

$$= \lambda_0 f(x_0) + \lambda_1 f(x_1) + \lambda_2 f(x_2) + \ldots + \lambda_n f(x_n).$$

We note that, $\int_a^b f(x)\, dx$ defines the area under the curve $y = f(x)$, above the x-axis, between the lines $x = a$, $x = b$.

Trapezium Rule

This rule is also called the *trapezoidal rule*. Let the curve $y = f(x)$, $a \leq x \leq b$, be approximated by the line joining the points $P(a, f(a))$, $Q(b, f(b))$ on the curve (see Fig. 3.1).

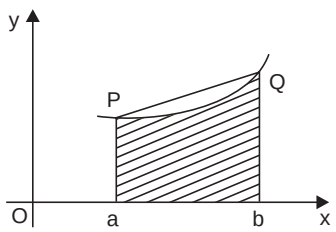


Fig. 5.1 Trapezium rule.

Using the Newton's forward difference formula, the linear polynomial approximation to $f(x)$, interpolating at the points $P(a, f(a))$, $Q(b, f(b))$, is given by

$$f(x) = f(x_0) + \frac{1}{h}(x - x_0)\, \Delta f(x_0) \qquad (3.32)$$

where $x_0 = a$, $x_1 = b$ and $h = b - a$. Substituting in (3.31), we obtain

$$I = \int_a^b f(x)\, dx = \int_{x_0}^{x_1} f(x) dx = f(x_0) \int_{x_0}^{x_1} dx + \frac{1}{h}\left[\int_{x_0}^{x_1} (x - x_0) dx\right] \Delta f_0$$

$$= (x_1 - x_0)\, f(x_0) + \frac{1}{h}\left[\frac{1}{2}(x - x_0)^2\right]_{x_0}^{x_1} \Delta f_0$$

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

$$= hf(x_0) + \frac{h}{2}\, [f(x_1) - f(x_0)]$$

$$= \frac{h}{2}\, [f(x_1) + f(x_0)] = \frac{(b-a)}{2}\, [f(b) + f(a)].$$

The trapezium rule is given by

$$I = \int_a^b f(x)\, dx = \frac{h}{2}\, [f(x_1) + f(x_0)] = \frac{(b-a)}{2}[f(b) + f(a)].$$

Remark Geometrically, the right hand side of the trapezium rule is the area of the trapezoid with width $b-a$, and ordinates $f(a)$ and $f(b)$, which is an approximation to the area under the curve $y = f(x)$ above the x-axis and the ordinates $x = a$ and $x = b$.

Error term in trapezium rule We show that the trapezium rule integrates exactly polynomial of degree ≤ 1. That is, using the definition of error given in (3.27), we show that

$$R_1(f, x) = 0 \text{ for } f(x) = 1, x.$$

Substituting $f(x) = 1, x$ in (3.27), we get

$$f(x) = 1: \quad R_1(f, x) = \int_a^b dx - \frac{(b-a)}{2}\, (2) = (b-a) - (b-a) = 0.$$

$$f(x) = x: \quad R_1(f, x) = \int_a^b x\, dx - \frac{(b-a)}{2}\, (b+a) = \frac{1}{2}\, (b^2 - a^2) - \frac{1}{2}\, (b^2 - a^2) = 0.$$

Hence, the trapezium rule integrates exactly polynomial of degree ≤ 1, and the method is of order 1.

Let $f(x) = x^2$. From (3.28), we get

$$c = \int_a^b x^2\, dx - \frac{(b-a)}{2}\, (b^2 + a^2) = \frac{1}{3}\, (b-a)^3 - \frac{1}{2}\, (b^3 + a^2 b - ab^2 - a^3)$$

$$= \frac{1}{6}\, (a^3 - 3a^2 b + 3ab^2 - b^3) = -\frac{1}{6}\, (b-a)^3.$$

Using (3.29), the expression for the error is given by

$$R_1(f, x) = \frac{c}{2!}\, f''(\xi) = -\frac{(b-a)^3}{12}\, f''(\xi) = -\frac{h^3}{12}\, f''(\xi)$$

where $a \leq \xi \leq b$.

The bound for the error is given by

$$|\, R_1\,(f, x)\,| \leq \frac{(b-a)^3}{12}\, M_2 = \frac{h^3}{12}\, M_2, \text{ where } M_2 = \max_{a \leq x \leq b} |\, f''(x)\,|.$$

If the length of the interval $[a, b]$ is large, then $b - a$ is also large and the error expression given (3.35) becomes meaningless. In this case, we subdivide $[a, b]$ into a number of subintervals of equal length and apply the trapezium rule to evaluate each integral. The rule is then called the *composite trapezium rule.*

Composite trapezium rule Let the interval $[a, b]$ be subdivided into N equal parts of length h. That is, $h = (b - a)/N$. The nodal points are given by

$$a = x_0, x_1 = x_0 + h, x_2 = x_0 + 2h, \ldots, x_N = x_0 + Nh = b.$$

We write

$$\int_a^b f(x)\,dx = \int_{x_0}^{x_N} f(x)\,dx = \int_{x_0}^{x_1} f(x)\,dx + \int_{x_1}^{x_2} f(x)\,dx + \ldots + \int_{x_{N-1}}^{x_N} f(x)dx\,.$$

There are N integrals. Using the trapezoidal rule to evaluate each integral, we get the *composite trapezoidal rule* as

$$\int_a^b f(x)\,dx = \frac{h}{2}\,[\{f(x_0) + f(x_1)\} + \{f(x_1) + f(x_2)\} + \ldots + \{f(x_{N-1}) + f(x_N)\}]$$

$$= \frac{h}{2}\,[f(x_0) + 2\{f(x_1) + f(x_2) + \ldots + f(x_{N-1})\} + f(x_N)].$$

The composite trapezium rule is also of order 1.

The error expression (3.34) becomes

$$R_1(f, x) = -\frac{h^3}{12}[f''(\xi_1) + f''(\xi_2) + \ldots + f''(\xi_N)], \quad x_{N-1} < \xi_N < x_N.$$

The bound on the error is given by

$$|\,R_1(f, x)\,| \le \frac{h^3}{12}[\,|f''(\xi_1)| + |f''(\xi_2)| + \ldots + |f''(\xi_N)|\,]$$

$$\le \frac{Nh^3}{12} M_2 = \frac{(b-a)h^2}{12} M_2$$

or $$|\,R_1(f, x)\,| \le \frac{(b-a)^3}{12N^2} M_2$$

where $$M_2 = \max_{a \le x \le b} |f''(x)| \quad \text{and} \quad Nh = b - a.$$

This expression is a true representation of the error in the trapezium rule. As we increase the number of intervals, the error decrases.

Remark Geometrically, the right hand side of the composite trapezium rule is the sum of areas of the N trapezoids with width h, and ordinates $f(x_{i-1})$ and $f(x_i)$, $i = 1, 2, \ldots, N$. This sum is an approximation to the area under the curve $y = f(x)$ above the x-axis and the ordinates $x = a$ and $x = b$.

Remark We have noted that the trapezium rule and the composite trapezium rule are of order 1. This can be verified from the error expressions given in (3.34) and (3.37). If $f(x)$ is a polynomial of degree ≤ 1, then $f''(x) = 0$. This result implies that error is zero and the trapezium rule produces exact results for polynomials of degree ≤ 1.

Example 1 *Derive the trapezium rule using the Lagrange linear interpolating polynomial.*

Solution The points on the curve are $P(a, f(a))$, $Q(b, f(b))$ (see Fig. 3.1). Lagrange linear interpolation gives

$$f(x) = \frac{(x-b)}{(a-b)} f(a) + \frac{(x-a)}{(b-a)} f(b)$$

$$= \frac{1}{(b-a)} [\{f(b) - f(a)\} x + \{bf(a) - af(b)\}].$$

Substituting in the integral, we get

$$I = \int_a^b f(x)\, dx = \frac{1}{(b-a)} \int_a^b [\{f(b) - f(a)\}x + \{bf(a) - af(b)\}]\, dx$$

$$= \frac{1}{(b-a)} \left[\frac{1}{2} \{f(b) - f(a)\} (b^2 - a^2) + \{bf(a) - af(b)\}(b-a) \right]$$

$$= \frac{1}{2} (b+a)[f(b) - f(a)] + bf(a) - af(b)$$

$$= \frac{(b-a)}{2} [f(a) + f(b)]$$

which is the required trapezium rule.

Example 2 *Find the approximate value of* $I = \int_0^1 \frac{dx}{1+x}$, *using the trapezium rule with* 2, 4 *and* 8 *equal subintervals. Using the exact solution, find the absolute errors.*

Solution With $N = 2$, 4 and 8, we have the following step lengths and nodal points.

$N = 2$: $h = \frac{b-a}{N} = \frac{1}{2}$. The nodes are 0, 0.5, 1.0.

$N = 4$: $h = \frac{b-a}{N} = \frac{1}{4}$. The nodes are 0, 0.25, 0.5, 0.75, 1.0.

$N = 8$: $h = \frac{b-a}{N} = \frac{1}{8}$. The nodes are 0, 0.125, 0.25, 0.375, 0.5, 0.675, 0.75, 0.875, 1.0.

We have the following tables of values.

$N = 2$:

x	0	0.5	1.0
$f(x)$	1.0	0.666667	0.5

$N = 4$: We require the above values. The additional values required are the following:

x	0.25	0.75
$f(x)$	0.8	0.571429

$N = 8$: We require the above values. The additional values required are the following:

x	0.125	0.375	0.625	0.875
$f(x)$	0.888889	0.727273	0.615385	0.533333

Now, we compute the value of the integral.

$N = 2$: $$I_1 = \frac{h}{2}[f(0) + 2f(0.5) + f(1.0)]$$
$$= 0.25\,[1.0 + 2(0.666667) + 0.5] = 0.708334.$$

$N = 4$: $$I_2 = \frac{h}{2}[f(0) + 2\{f(0.25) + f(0.5) + f(0.75)\} + f(1.0)]$$
$$= 0.125\,[1.0 + 2\,\{0.8 + 0.666667 + 0.571429\} + 0.5] = 0.697024.$$

$N = 8$: $$I_3 = \frac{h}{2}[f(0) + 2\{f(0.125) + f(0.25) + f(0.375) + f(0.5)$$
$$+ f(0.625) + f(0.75) + f(0.875)\} + f(1.0)]$$
$$= 0.0625[1.0 + 2\{0.888889 + 0.8 + 0.727273 + 0.666667 + 0.615385$$
$$+ 0.571429 + 0.533333\} + 0.5] = 0.694122.$$

The exact value of the integral is $I = \ln 2 = 0.693147$.

The errors in the solutions are the following:

$$|\ \text{Exact} - I_1\ | = |\ 0.693147 - 0.708334\ | = 0.015187$$
$$|\ \text{Exact} - I_2\ | = |\ 0.693147 - 0.697024\ | = 0.003877$$
$$|\ \text{Exact} - I_3\ | = |\ 0.693147 - 0.694122\ | = 0.000975.$$

Example 3 *Evaluate* $I = \int_1^2 \frac{dx}{5+3x}$ *with* 4 *and* 8 *subintervals using the trapezium rule. Compare with the exact solution and find the absolute errors in the solutions. Comment on the magnitudes of the errors obtained. Find the bound on the errors.*

Solution With $N = 4$ and 8, we have the following step lengths and nodal points.

$N = 4$: $h = \frac{b-a}{N} = \frac{1}{4}$. The nodes are 1, 1.25, 1.5, 1.75, 2.0.

$N = 8$: $h = \dfrac{b-a}{N} = \dfrac{1}{8}$. The nodes are 1, 1.125, 1.25, 1.375, 1.5, 1.675, 1.75, 1.875, 2.0.

We have the following tables of values.

$N = 4$:

x	1.0	1.25	1.5	1.75	2.0
$f(x)$	0.125	0.11429	0.10526	0.09756	0.09091

$N = 8$: We require the above values. The additional values required are the following.

x	1.125	1.375	1.625	1.875
$f(x)$	0.11940	0.10959	0.10127	0.09412

Now, we compute the value of the integral.

$N = 4$: $I_1 = \dfrac{h}{2}\,[f(1) + 2\,\{f(1.25) + f(1.5) + f(1.75)\} + f(2.0)]$

$= 0.125\,[0.125 + 2\,\{0.11429 + 0.10526 + 0.09756\} + 0.09091]$

$= 0.10627.$

$N = 8$: $I_2 = \dfrac{h}{2}\,[f(1) + 2\{f(1.125) + f(1.25) + f(1.375) + f(1.5)$

$+ f(1.625) + f(1.75) + f(1.875)\} + f(2.0)]$

$= 0.0625\,[0.125 + 2\{0.11940 + 0.11429 + 0.10959 + 0.10526 + 0.10127$

$+ 0.09756 + 0.09412\} + 0.09091]$

$= 0.10618.$

The exact value of the integral is

$$I = \frac{1}{3}\Big[\ln(5+3x)\Big]_1^2 = \frac{1}{3}[\ln 11 - \ln 8] = 0.10615.$$

The errors in the solutions are the following:

$$|\ \text{Exact} - I_1\ | = |\ 0.10615 - 0.10627\ | = 0.00012.$$

$$|\ \text{Exact} - I_2\ | = |\ 0.10615 - 0.10618\ | = 0.00003.$$

We find that $|\ \text{Error in } I_2\ | \approx \dfrac{1}{4}\ |\ \text{Error in } I_1\ |$.

Bounds for the errors

$$|\ \text{Error}\ | \le \frac{(b-a)h^2}{12} M_2, \text{ where } M_2 = \max_{[1,\,2]}\ |\ f''(x)\ |.$$

We have $f(x) = \dfrac{1}{5+3x}, f'(x) = -\dfrac{3}{(5+3x)^2}, f''(x) = \dfrac{18}{(5+3x)^3}.$

$$M_2 = \max_{[1,2]} \left| \frac{18}{(5+3x)^3} \right| = \frac{18}{512} = 0.03516.$$

$$h = 0.25: \ | \text{ Error } | \le \frac{(0.25)^2}{12} (0.03516) = 0.00018.$$

$$h = 0.125: \ | \text{ Error } | \le \frac{(0.125)^2}{12} (0.03516) = 0.000046.$$

Actual errors are smaller than the bounds on the errors.

Example 4 *Using the trapezium rule, evaluate the integral* $I = \int_0^1 \frac{dx}{x^2 + 6x + 10}$, *with* 2 *and* 4 *subintervals. Compare with the exact solution. Comment on the magnitudes of the errors obtained.*

Solution With $N = 2$ and 4, we have the following step lengths and nodal points.

$N = 2$: $h = 0.5$. The nodes are 0.0, 0.5, 1.0.

$N = 4$: $h = 0.25$. The nodes are 0.0, 0.25, 0.5, 0.75, 1.0.

We have the following tables of values.

$N = 2$:

x	0.0	0.5	1.0
$f(x)$	0.1	0.07547	0.05882

$N = 4$: We require the above values. The additional values required are the following.

x	0.25	0.75
$f(x)$	0.08649	0.06639

Now, we compute the value of the integral.

$N = 2$:
$$I_1 = \frac{h}{2} [f(0.0) + 2 f(0.5) + f(1.0)]$$
$$= 0.25 \ [0.1 + 2(0.07547) + 0.05882] = 0.07744.$$

$N = 4$:
$$I_2 = \frac{h}{2} [f(0.0) + 2\{f(0.25) + f(0.5) + f(0.75)\} + f(1.0)]$$
$$= 0.125[0.1 + 2(0.08649 + 0.07547 + 0.06639) + 0.05882] = 0.07694.$$

The exact value of the integral is

$$I = \int_0^1 \frac{dx}{(x+3)^2 + 1} = \left[\tan^{-1}(x+3) \right]_0^1 = \tan^{-1}(4) - \tan^{-1}(3) = 0.07677.$$

The errors in the solutions are the following:

$$| \text{ Exact} - I_1 | = | \ 0.07677 - 0.07744 \ | = 0.00067$$
$$| \text{ Exact} - I_2 | = | \ 0.07677 - 0.07694 \ | = 0.00017.$$

We find that

$$| \text{ Error in } I_2 | \approx \frac{1}{4} | \text{ Error in } I_1 | .$$

Example 5 *The velocity of a particle which starts from rest is given by the following table.*

t (sec)	0	2	4	6	8	10	12	14	16	18	20
v (ft/sec)	0	16	29	40	46	51	32	18	8	3	0

Evaluate using trapezium rule, the total distance travelled in 20 *seconds.*

Solution From the definition, we have

$$v = \frac{ds}{dt} \text{ , or } s = \int v \, dt.$$

Starting from rest, the distance travelled in 20 seconds is

$$s = \int_0^{20} v \, dt.$$

The step length is $h = 2$. Using the trapezium rule, we obtain

$$\begin{aligned} s = \frac{h}{2} [f(0) + 2\{f(2) + f(4) + f(6) + f(8) + f(10) + f(12) + f(14) \\ + f(16) + f(18)\} + f(20)] \\ = 0 + 2\{16 + 29 + 40 + 46 + 51 + 32 + 18 + 8 + 3\} + 0 = 486 \text{ feet.} \end{aligned}$$

Simpson's 1/3 Rule

In the previous section, we have shown that the trapezium rule of integration integrates exactly polynomials of degree ≤ 1, that is, the order of the formula is 1. In many science and engineering applications, we require methods which produce more accurate results. One such method is the Simpson's 1/3 rule.

Let the interval $[a, b]$ be subdivided into two equal parts with step length $h = (b - a)/2$. We have three abscissas $x_0 = a$, $x_1 = (a + b)/2$, and $x_2 = b$.

Then, $P(x_0, f(x_0))$, $Q(x_1 f(x_1))$, $R(x_2, f(x_2))$ are three points on the curve $y = f(x)$. We approximate the curve $y = f(x)$, $a \leq x \leq b$, by the parabola joining the points P, Q, R, that is, we approximate the given curve by a polynomial of degree 2. Using the Newton's forward difference formula, the quadratic polynomial approximation to $f(x)$, interpolating at the points $P(x_0, f(x_0))$, $Q(x_1 f(x_1))$, $R(x_2, f(x_2))$, is given by

$$f(x) = f(x_0) + \frac{1}{h}(x - x_0)\,\Delta f(x_0) + \frac{1}{2h^2}(x - x_0)(x - x_1)\Delta^2 f(x_0).$$

Substituting in (3.31), we obtain

$$\int_a^b f(x)dx = \int_{x_0}^{x_2} f(x)dx = \int_{x_0}^{x_2} \left[f(x_0) + \frac{1}{h}(x - x_0)\,\Delta f(x_0) + \frac{1}{2h^2}(x - x_0)(x - x_1)\Delta^2 f(x_0) \right] dx$$

$$= (x_2 - x_0)\, f(x_0) + \frac{1}{h}\left[\frac{1}{2}(x - x_0)^2\right]_{x_0}^{x_2} \Delta f(x_0) + I_1 = 2hf\,(x_0) + 2h\Delta f(x_0) + I_1.$$

Evaluating I_1, we obtain

$$I_1 = \frac{1}{2h^2}\left[\frac{x^3}{3} - (x_0 + x_1)\frac{x^2}{2} + x_0 x_1 x\right]_{x_0}^{x_2} \Delta^2 f(x_0)$$

$$= \frac{1}{12h^2}\left[2(x_2^3 - x_0^3) - 3(x_0 + x_1)(x_2^2 - x_0^2) + 6x_0x_1(x_2 - x_0)\right] \Delta^2 f(x_0)$$

$$= \frac{1}{12h^2}(x_2 - x_0)\left[2(x_2^2 + x_0x_2 + x_0^2) - 3(x_0 + x_1)(x_2 + x_0) + 6x_0x_1\right] \Delta^2 f(x_0).$$

Substituting $x_2 = x_0 + 2h$, $x_1 = x_0 + h$, we obtain

$$I_1 = \frac{1}{6h}\left[2(3x_0^2 + 6hx_0 + 4h^2) - 3(4x_0^2 + 6hx_0 + 2h^2) + 6x_0^2 + 6hx_0\right] \Delta^2 f(x_0)$$

$$= \frac{1}{6h}(2h^2)\,\Delta^2\, f(x_0) = \frac{h}{3}\Delta^2 f(x_0).$$

Hence

$$\int_a^b f(x)\,dx = \int_{x_0}^{x_2} f(x)\,dx = 2hf(x_0) + 2h\Delta f(x_0) + \frac{h}{3}\Delta^2 f(x_0)$$

$$= \frac{h}{3}\left[6f(x_0) + 6\{f(x_1) - f(x_0)\} + \{f(x_0) - 2f(x_1) + f(x_2)\}\right]$$

$$= \frac{h}{3}\left[f(x_0) + 4f(x_1) + f(x_2)\right]$$

In terms of the end points, we can also write the formula as

$$\int_a^b f(x)dx = \frac{(b-a)}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(x_2)\right]$$

This formula is called the *Simpson's* 1/3 *rule*.

We can also evaluate the integral $\int_{x_0}^{x_2} f(x)\,dx$, as follows. We have

$$\int_{x_0}^{x_2} f(x)\,dx = \int_{x_0}^{x_2}\left[f(x_0) + \frac{1}{h}(x - x_0)\,\Delta f(x_0) + \frac{1}{2h^2}(x - x_0)(x - x_1)\Delta^2 f(x_0)\right]dx.$$

Let $[(x - x_0)/h] = s$. The limits of integration become:

for $x = x_0$, $s = 0$, and for $x = x_2$, $s = 2$.

We have $dx = h\,ds$. Hence,

$$\int_{x_0}^{x_2} f(x)\,dx = h\int_0^2 \left[f(x_0) + s\Delta f(x_0) + \frac{1}{2}s(s-1)\Delta^2 f(x_0)\right] ds$$

$$= h\left[s\,f(x_0) + \frac{s^2}{2}\Delta f(x_0) + \frac{1}{2}\left(\frac{s^3}{3} - \frac{s^2}{2}\right)\Delta^2 f(x_0)\right]_0^2$$

$$= h\left[2f(x_0) + 2\Delta f(x_0) + \frac{1}{3}\Delta^2 f(x_0)\right]$$

$$= \frac{h}{3}\,[6f(x_0) + 6\{f(x_1) - f(x_0)\} + \{f(x_0) - 2f(x_1) + f(x_2)\}]$$

$$= \frac{h}{3}\,[f(x_0) + 4f(x_1) + f(x_2)]$$

which is the same formula as derived earlier.

Error term in Simpson 1/3 rule. We show that the Simpson's rule integrates exactly polynomials of degree ≤ 3. That is, using the definition of error given in (3.27), we show that

$$R_2(f, x) = 0 \text{ for } f(x) = 1, x, x^2, x^3.$$

Substituting $f(x) = 1, x, x^2, x^3$ in (3.27), we get

$f(x) = 1$: $$R_2(f, x) = \int_a^b dx - \frac{(b-a)}{6}(6) = (b-a) - (b-a) = 0.$$

$f(x) = x$: $$R_2(f, x) = \int_a^b x\,dx - \frac{(b-a)}{6}\left[a + 4\left(\frac{a+b}{2}\right) + b\right]$$

$$= \frac{1}{2}(b^2 - a^2) - \frac{1}{2}(b^2 - a^2) = 0.$$

$f(x) = x^2$: $$R_2(f, x) = \int_a^b x^2\,dx - \frac{(b-a)}{6}\left[a^2 + 4\left(\frac{a+b}{2}\right)^2 + b^2\right]$$

$$= \frac{1}{3}(b^3 - a^3) - \frac{(b-a)}{3}[a^2 + ab + b^2]$$

$$= \frac{1}{3}(b^3 - a^3) - \frac{1}{3}(b^3 - a^3) = 0.$$

$f(x) = x^3$: $$R_2(f, x) = \int_a^b x^3\,dx - \frac{(b-a)}{6}\left[a^3 + 4\left(\frac{a+b}{2}\right)^3 + b^3\right]$$

$$= \frac{1}{4}(b^4 - a^4) - \frac{(b-a)}{4}[a^3 + a^2b + ab^2 + b^3]$$

$$= \frac{1}{4}(b^4 - a^4) - \frac{1}{4}(b^4 - a^4) = 0.$$

Hence, the Simpson's rule integrates exactly polynomials of degree ≤ 3. Therefore, the method is of order 3. It is interesting to note that the method is one order higher than expected, since we have approximated $f(x)$ by a polynomial of degree 2 only.

Let $f(x) = x^4$. From (3.28), we get

$$c = \int_a^b x^4\,dx - \frac{(b-a)}{6}\left[a^4 + 4\left(\frac{a+b}{2}\right)^4 + b^4\right]$$

$$= \frac{1}{5}(b^5 - a^5) - \frac{(b-a)}{24}(5a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + 5b^4)$$

$$= \frac{1}{120}[24(b^5 - a^5) - 5(b-a)(5a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + 5b^4)]$$

$$= -\frac{(b-a)}{120}[b^4 - 4ab^3 + 6a^2b^2 - 4a^3b + a^4)]$$

$$= -\frac{(b-a)^5}{120}.$$

Using (3.29), the expression for the error is given by

$$R(f, x) = \frac{c}{4!}f^{(4)}(\xi) = -\frac{(b-a)^5}{2880}f^{(4)}(\xi) = -\frac{h^5}{90}f^{(4)}(\xi)$$

since $h = (b-a)/2$, and $a \leq \xi \leq b$.

Since the method produces exact results, that is, $R_2(f, x) = 0$, when $f(x)$ is a polynomial of degree ≤ 3, the method is of order 3.

The bound for the error is given by

$$|R(f, x)| \leq \frac{(b-a)^5}{2880}M_4 = \frac{h^5}{90}M_4, \text{ where } M_4 = \max_{a \leq x \leq b}|f^{(4)}(x)|.$$

As in the case of the trapezium rule, if the length of the interval $[a, b]$ is large, then $b - a$ is also large and the error expression given in (3.41) becomes meaningless. In this case, we subdivide $[a, b]$ into a number of subintervals of equal length and apply the Simpson's 1/3 rule to evaluate each integral. The rule is then called the *composite Simpson's* 1/3 *rule.*

Composite Simpson's 1/3 rule We note that the Simpson's rule derived earlier uses three nodal points. Hence, we subdivide the given interval $[a, b]$ into even number of subintervals of equal length h. That is, we obtain an *odd number* of nodal points. We take the even number of intervals as $2N$. The step length is given by $h = (b-a)/(2N)$. The nodal points are given by

$$a = x_0,\ x_1 = x_0 + h,\ x_2 = x_0 + 2h,\ \ldots,\ x_{2N} = x_0 + 2N\,h = b.$$

The given interval is now written as

$$\int_a^b f(x)\,dx = \int_{x_0}^{x_{2N}} f(x)\,dx = \int_{x_0}^{x_2} f(x)\,dx + \int_{x_2}^{x_4} f(x)\,dx + \ldots + \int_{x_{2N-2}}^{x_{2N}} f(x)\,dx.$$

Note that there are N integrals. The limits of each integral contain three nodal points. Using the Simpson's 1/3 rule to evaluate each integral, we get the *composite Simpson's* 1/3 *rule* as

$$\begin{aligned}\int_a^b f(x)dx = \frac{h}{3}\ &[\{f(x_0) + 4f(x_1) + f(x_2)\} + \{f(x_2) + 4\,f(x_3) + f(x_4)\} + \ldots \\ &+ \{f(x_{2N-2}) + 4f(x_{2N-1}) + f(x_{2N})\}] \\ = \frac{h}{3}\ &[f(x_0) + 4\{f(x_1) + f(x_3) + \ldots + f(x_{2N-1})\} + 2\{f(x_2) + f(x_4) + \ldots \\ &+ f(x_{2N-2})\} + f(x_{2N})]\end{aligned}$$

The composite Simpson's 1/3 rule is also of order 3.

The error expression (3.34) becomes

$$R(f, x) = -\frac{h^5}{90}\,[f^{(4)}(\xi_1) + f^{(4)}(\xi_2) + \ldots + f^{(4)}(\xi_N)],$$

where $x_0, < \xi_1 < x_2,\ x_2 < \xi_2 < x_4$, etc.

The bound on the error is given by

$$\begin{aligned}|\ R(f, x)\ | &\le \frac{h^5}{90}\left[|\,f^{(4)}(\xi_1)\,| + |\,f^{(4)}(\xi_2)\,| + \ldots + |\,f^{(4)}(\xi_N)\,|\right] \\ &\le \frac{Nh^5}{90}M_4 = \frac{(b-a)h^4}{180}M_4\end{aligned}$$

or $$|\ R(f, x)\ | \le \frac{(b-a)^5}{2880N^4}M_4$$

where $$M_4 = \max_{a \le x \le b} |\,f^{(4)}(x)\ |\ \text{ and } N\,h = (\,b - a)/2.$$

This expression is a true representation of the error in the Simpson's 1/3 rule. We observe that as N increases, the error decreases.

Remark We have noted that the Simpson 1/3 rule and the composite Simpson's 1/3 rule are of order 3. This can be verified from the error expressions given in (3.41) and (3.45). If $f(x)$ is a polynomial of degree ≤ 3, then $f^{(4)}(x) = 0$. This result implies that error is zero and the composite Simpson's 1/3 rule produces exact results for polynomials of degree ≤ 3.

Remark Note that the number of subintervals is $2N$. We can also say that the number of subintervals is $n = 2N$ and write $h = (b - a)/n$, where n is even.

Example 6 *Find the approximate value of* $I = \int_0^1 \frac{dx}{1+x}$, *using the Simpson's* 1/3 *rule with* 2, 4 *and* 8 *equal subintervals. Using the exact solution, find the absolute errors.*

Solution With $n = 2N = 2$, 4 and 8, or $N = 1$, 2, 4 we have the following step lengths and nodal points.

$N = 1$: $h = \frac{b-a}{2N} = \frac{1}{2}$. The nodes are 0, 0.5, 1.0.

$N = 2$: $h = \frac{b-a}{2N} = \frac{1}{4}$. The nodes are 0, 0.25, 0.5, 0.75, 1.0.

$N = 4$: $h = \frac{b-a}{2N} = \frac{1}{8}$. The nodes are 0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1.0.

We have the following tables of values.

$n = 2N = 2$:

x	0	0.5	1.0
$f(x)$	1.0	0.666667	0.5

$n = 2N = 4$: We require the above values. The additional values required are the following.

x	0.25	0.75
$f(x)$	0.8	0.571429

$n = 2N = 8$: We require the above values. The additional values required are the following.

x	0.125	0.375	0.625	0.875
$f(x)$	0.888889	0.727273	0.615385	0.533333

Now, we compute the value of the integral.

$n = 2N = 2$: $I_1 = \frac{h}{3}[f(0) + 4f(0.5) + f(1.0)]$

$= \frac{1}{6}[1.0 + 4(0.666667) + 0.5] = 0.674444.$

$n = 2N = 4$: $I_2 = \frac{h}{3}[f(0) + 4\{f(0.25) + f(0.75)\} + 2f(0.5) + f(1.0)]$

$= \frac{1}{12}[1.0 + 4\{0.8 + 0.571429\} + 2(0.666667) + 0.5] = 0.693254.$

$n = 2N = 8$: $I_3 = \frac{h}{3}[f(0) + 4\{f(0.125) + f(0.375) + f(0.625) + f(0.875)\}$

$+ 2\{f(0.25) + f(0.5) + f(0.75)\} + f(1.0)]$

$$= \frac{1}{24}\,[1.0 + 4\,\{0.888889 + 0.727273 + 0.615385 + 0.533333\}$$

$$+ 2\,\{0.8 + 0.666667 + 0.571429\} + 0.5]$$

$$= 0.693155.$$

The exact value of the integral is $I = \ln 2 = 0.693147$.

The errors in the solutions are the following:

$$|\ \text{Exact} - I_1\ | = |\ 0.693147 - 0.694444\ | = 0.001297.$$

$$|\ \text{Exact} - I_2\ | = |\ 0.693147 - 0.693254\ | = 0.000107.$$

$$|\ \text{Exact} - I_3\ | = |\ 0.693147 - 0.693155\ | = 0.000008.$$

Example 7 *Evaluate* $I = \int_1^2 \frac{dx}{5+3x}$, *using the Simpson's* 1/3 *rule with* 4 *and* 8 *subintervals. Compare with the exact solution and find the absolute errors in the solutions.*

Solution With $N = 2N = 4, 8$ or $N = 2, 4$, we have the following step lengths and nodal points.

$N = 2$: $h = \frac{b-a}{2N} = \frac{1}{4}$. The nodes are 1, 1.25, 1.5, 1.75, 2.0.

$N = 4$: $h = \frac{b-a}{2N} = \frac{1}{8}$. The nodes are 1, 1.125, 1.25, 1.375, 1.5, 1.675, 1.75, 1.875, 2.0.

We have the following tables of values.

$n = 2N = 4$:

x	1.0	1.25	1.5	1.75	2.0
$f(x)$	0.125	0.11429	0.10526	0.09756	0.09091

$n = 2N = 8$: We require the above values. The additional values required are the following.

x	1.125	1.375	1.625	1.875
$f(x)$	0.11940	0.10959	0.10127	0.09412

Now, we compute the value of the integral.

$n = 2N = 4$: $I_1 = \frac{h}{3}\,[f(1) + 4\{(1.25) + f(1.75)\} + 2f(1.5) + f(2.0)]$

$$= \frac{0.25}{3}\,[0.125 + 4\{0.11429 + 0.09756\} + 2(0.10526) + 0.09091]$$

$$= 0.10615.$$

$n = 2N = 8$: $I_2 = \frac{h}{3}\,[f(1) + 4\{f(1.125) + f(1.375) + f(1.625) + f(1.875)\}$

$$+ 2\{f(1.25) + f(1.5) + f(1.75)\} + f(2.0)]$$

$$= \frac{0.125}{3}\,[0.125 + 4\{0.11940 + 0.10959 + 0.10127 + 0.09412\}$$

$$+ 2\{0.11429 + 0.10526 + 0.09756\} + 0.09091]$$

$$= 0.10615.$$

The exact value of the integral is $I = \frac{1}{3}\,[\ln 11 - \ln 8] = 0.10615$.

The results obtained with $n = 2N = 4$ and $n = 2N = 8$ are accurate to all the places.

Example 8 *Using Simpson's* 1/3 *rule, evaluate the integral* $I = 4\int_0^1 \frac{dx}{x^2 + 6x + 10}$, *with* 2 *and subintervals. Compare with the exact solution.*

Solution With $n = 2N = 2$ and 4, or $N = 1, 2$, we have the following step lengths and nodal points.

$N = 1$: $h = 0.5$. The nodes are 0.0, 0.5, 1.0.

$N = 2$: $h = 0.25$. The nodes are 0.0, 0.25, 0.5, 0.75, 1.0.

We have the following values of the integrand.

$n = 2N = 2$:

x	0.0	0.5	1.0
$f(x)$	0.1	0.07547	0.05882

$n = 2N = 4$: We require the above values. The additional values required are the following.

x	0.25	0.75
$f(x)$	0.08649	0.06639

Now, we compute the value of the integral.

$n = 2N = 2$: $I_1 = \frac{h}{3}\,[f(0.0) + 4\,f(0.5) + f(1.0)]$

$$= \frac{0.5}{3}\,[0.1 + 4(0.07547) + 0.05882] = 0.07678.$$

$n = 2N = 4$: $I_2 = \frac{h}{3}\,[f(0.0) + 4\,\{f(0.25) + f(0.75)\} + 2\,f(0.5) + f(1.0)]$

$$= \frac{0.25}{3}\,[0.1 + 4(0.08649 + 0.06639) + 2(0.07547) + 0.05882] = 0.07677.$$

The exact value of the integral is

$$I = \int_0^1 \frac{dx}{(x+3)^2 + 1} = \left[\tan^{-1}(x+3)\right]_0^1 = \tan^{-1}(4) - \tan^{-1}(3) = 0.07677\,.$$

The errors in the solutions are the following:

$$| \text{Exact} - I_1 | = | 0.07677 - 0.07678 | = 0.00001.$$

$$| \text{Exact} - I_2 | = | 0.07677 - 0.07677 | = 0.00000.$$

Example 9 *The velocity of a particle which starts from rest is given by the following table.*

t (*sec*)	0	2	4	6	8	10	12	14	16	18	20
v (*ft*/*sec*)	0	16	29	40	46	51	32	18	8	3	0

Evaluate using Simpson's 1/3 *rule, the total distance travelled in* 20 *seconds.*

Solution From the definition, we have

$$v = \frac{ds}{dt}, \quad \text{or} \quad s = \int v\, dt.$$

Starting from rest, the distance travelled in 20 seconds is

$$s = \int_0^{20} v\, dt.$$

The step length is $h = 2$. Using the Simpson's rule, we obtain

$$s = \frac{h}{3}\,[f(0) + 4\{f(2) + f(6) + f(10) + f(14) + f(18)\} + 2\{f(4) + f(8) + f(12) + f(16)\} + f(20)]$$

$$= \frac{2}{3}\,[0 + 4\{16 + 40 + 51 + 18 + 3\} + 2\{29 + 46 + 32 + 8\} + 0]$$

$$= 494.667 \text{ feet.}$$

Simpson's 3/8 Rule

To derive the Simpson's 1/3 rule, we have approximated $f(x)$ by a quadratic polynomial. To derive the Simpson's 3/8 rule, we approximate $f(x)$ by a cubic polynomial. For interpolating by a cubic polynomial, we require four nodal points. Hence, we subdivide the given interval $[a, b]$ into 3 equal parts so that we obtain four nodal points. Let $h = (b - a)/3$. The nodal points are given by

$$x_0 = a,\ x_1 = x_0 + h,\ x_2 = x_0 + 2h,\ x_3 = x_0 + 3h.$$

Using the Newton's forward difference formula, the cubic polynomial approximation to $f(x)$, interpolating at the points

$$P(x_0, f(x_0)),\ Q(x_1, f(x_1)),\ R(x_2, f(x_2)),\ S(x_3, f(x_3))$$

is given by

$$f(x) = f(x_0) + \frac{1}{h}(x - x_0)\,\Delta f(x_0) + \frac{1}{2h^2}(x - x_0)(x - x_1)\,\Delta^2 f(x_0)$$

$$+ \frac{1}{6h^3}(x - x_0)(x - x_1)(x - x_2)\,\Delta^3 f(x_0).$$

Substituting in (3.31), and integrating, we obtain the Simpson's 3/8 rule as

$$\int_a^b f(x)dx = \int_{x_0}^{x_3} f(x)dx = \frac{3h}{8}[f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)].$$

The error expression is given by

$$R_3(f, x) = -\frac{3}{80}h^5 f^{(4)}(\xi) = \frac{(b-a)^5}{6480} f^{(4)}(\xi), \quad x_0 < \xi < x_3.$$

Since the method produces exact results, that is, $R_3(f, x) = 0$, when $f(x)$ is a polynomial of degree ≤ 3, the method is of order 3.

As in the case of the Simpson's 1/3 rule, if the length of the interval $[a, b]$ is large, then $b - a$ is also large and the error expression given in (3.47) becomes meaningless. In this case, we subdivide $[a, b]$ into a number of subintervals of equal length such that the number of subintervals is divisible by 3. That is, the number of intervals must be 6 or 9 or 12 etc., so that we get 7 or 10 or 13 nodal points etc. Then, we apply the Simpson's 3/8 rule to evaluate each integral. The rule is then called the *composite Simpson's 3/8 rule*. For example, if we divide $[a, b]$ into 6 parts, then we get the seven nodal points as

$$x_0 = a, x_1 = x_0 + h, x_2 = x_0 + 2h, x_3 = x_0 + 3h, \ldots, x_6 = x_0 + 6h.$$

The Simpson's 3/8 rule becomes

$$\int_a^b f(x)\,dx = \int_{x_0}^{x_3} f(x)\,dx + \int_{x_3}^{x_6} f(x)\,dx$$

$$= \frac{3h}{8}[\{f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)\} + \{f(x_3) + 3f(x_4) + 3f(x_5) + f(x_6)\}]$$

$$= \frac{3h}{8}[f(x_0) + 3f(x_1) + 3f(x_2) + 2f(x_3) + 3f(x_4) + 3f(x_5) + f(x_6)]$$

The error in this composite Simpson's 3/8 rule becomes

$$R_3(f, x) = -\frac{3}{80}h^5[f^{(4)}(\xi_1) + f^{(4)}(\xi_2)], \quad x_0 < \xi_1 < x_3, x_3 < \xi_2 < x_6.$$

In the general case, the bound for the error expression is given by

$$|R(f, x)| \leq C h^4 M_4$$

where $$M_4 = \max_{a \leq x \leq b} |f^{(4)}(x)|.$$

If $f(x)$ is a polynomial of degree ≤ 3, then $f^{(4)}(x) = 0$. This result implies that error expression given in (3.47) or (3.48) is zero and the composite Simpson's 3/8 rule produces exact results for polynomials of degree ≤ 3. Therefore, the formula is of order 3, which is same as the order of the Simpson's 1/3 rule.

Remark In Simpson's 3/8th rule, the number of subintervals is $n = 3N$. Hence, we have

$$h = \frac{b-a}{3N}, \text{ or } h = \frac{b-a}{n}$$

where n is a multiple of 3.

Remark Simpson's 3/8 rule has some disadvantages. They are the following: (i) The number of subintervals must be divisible by 3. (ii) It is of the same order as the Simpson's 1/3 rule, which only requires that the number of nodal points must be odd. (iii) The error constant c in the case of Simpson's 3/8 rule is $c = 3/80$, which is much larger than the error constant $c = 1/90$, in the case of Simpson's 1/3 rule. Therefore, the error in the case of the Simpson's 3/8 rule is larger than the error in the case Simpson 1/3 rule. Due to these disadvantages, Simpson's 3/8 rule is not used in practice.

Example 10 *Using the Simpson's* 3/8 *rule, evaluate* $I = \int_1^2 \frac{dx}{5+3x}$ *with* 3 *and* 6 *subintervals. Compare with the exact solution.*

Solution With $n = 3N = 3$ and 6, we have the following step lengths and nodal points.

$n = 3N = 3$: $h = \frac{b-a}{3N} = \frac{1}{3}$. The nodes are 1, 4/3, 5/3, 2.0.

$n = 3N = 6$: $h = \frac{b-a}{3N} = \frac{1}{6}$. The nodes are 1, 7/6, 8/6, 9/6, 10/6, 11/6, 2.0

We have the following tables of values.

$n = 3N = 3$:

x	1.0	4/3	5/3	2.0
$f(x)$	0.125	0.11111	0.10000	0.09091

$n = 3N = 6$: We require the above values. The additional values required are the following.

x	7/6	9/6	11/6
$f(x)$	0.11765	0.10526	0.09524

Now, we compute the value of the integral.

$n = 3N = 3$:

$$I_1 = \frac{3h}{8}[f(1) + 3f(4/3) + 3f(5/3) + f(2.0)]$$
$$= 0.125[0.125 + 3\{0.11111 + 0.10000\} + 0.09091] = 0.10616.$$

$n = 3N = 6$:

$$I_2 = \frac{3h}{8}[f(1) + 3\{f(7/6) + f(8/6) + f(10/6) + f(11/6)\} + 2f(9/6)\} + f(2.0)]$$

$$= \frac{1}{16} [0.125 + 3 \{0.11765 + 0.11111 + 0.10000 + 0.09524\}$$
$$+ 2(0.10526) + 0.09091] = 0.10615.$$

The exact value of the integral is $I = \frac{1}{3} [\log 11 - \log 8] = 0.10615$.

The magnitude of the error for $n = 3$ is 0.00001 and for $n = 6$ the result is correct to all places.

Romberg Method (Integration)

In order to obtain accurate results, we compute the integrals by trapezium or Simpson's rules for a number of values of step lengths, each time reducing the step length. We stop the computation, when convergence is attained (usually, the magnitude of the difference in successive values of the integrals obtained by reducing values of the step lengths is less than a given accuracy). Convergence may be obtained after computing the value of the integral with a number of step lengths. While computing the value of the integral with a particular step length, the values of the integral obtained earlier by using larger step lengths were not used. Further, convergence may be slow.

Romberg method is a powerful tool which uses the method of extrapolation.

We compute the value of the integral with a number of step lengths using the same method. Usually, we start with a coarse step length, then reduce the step lengths and recompute the value of the integral. The sequence of these values converges to the exact value of the integral. Romberg method uses these values of the integral obtained with various step lengths, to refine the solution such that the new values are of higher order. That is, as if the results are obtained using a higher order method than the order of the method used. The extrapolation method is derived by studying the error of the method that is being used.

Let us derive the Romberg method for the trapezium and Simpson's rules.

Romberg method for the trapezium rule

Let the integral

$$I = \int_a^b f(x)dx$$

be computed by the composite trapezium rule. Let I denote the exact value of the integral and I_T denote the value obtained by the composite trapezium rule.

The error, $I - I_T$, in the composite trapezium rule in computing the integral is given by

$$I - I_T = c_1h^2 + c_2h^4 + c_3h^6 + \ldots$$

or

$$I = I_T + c_1h^2 + c_2h^4 + c_3h^6 + \ldots$$

where c_1, c_2, c_3, ... are independent of h.

To illustrate the extrapolation procedure, first consider two error terms.

$$I = I_T + c_1h^2 + c_2h^4.$$

Let I be evaluated using two step lengths h and qh, $0 < q < 1$. Let these values be denoted by $I_T(h)$ and $I_T(qh)$. The error equations become

$$I = I_T(h) + c_1h^2 + c_2h^4. \tag{3.51}$$

$$I = I_T(qh) + c_1q^2h^2 + c_2q^4h^4. \tag{3.52}$$

From (3.51), we obtain

$$I - I_T(h) = c_1h^2 + c_2h^4. \tag{3.53}$$

From (3.52), we obtain

$$I - I_T\ (qh) = c_1q^2h^2 + c_2q^4h^4. \tag{3.54}$$

Multiply (3.53) by q^2 to obtain

$$q^2\ [I - I_T\ (h)] = c_1q^2h^2 + c_2q^2h^4. \tag{3.55}$$

Eliminating $c_1q^2h^2$ from (3.54) and (3.55), we obtain

$$(1 - q^2)I - I_T\ (qh) + q^2I_T(h) = c_2q^2h^4(q^2 - 1).$$

Solving for I, we obtain

$$I = \frac{I_T(qh) - q^2I_T(h)}{(1-q^2)} - c_2q^2h^4.$$

Note that the error term on the right hand side is now of order $O(h^4)$.

Neglecting the $O(h^4)$ error term, we obtain the new approximation to the value of the integral as

$$I \approx I_T^{(1)}(h) = \frac{I_T(qh) - q^2I_T(h)}{(1-q^2)}. \tag{3.56}$$

We note that this value is obtained by suitably using the values of the integral obtained with step lengths h and qh, $0 < q < 1$. This computed result is of order, $O(h^4)$, which is higher than the order of the trapezium rule, which is of $O(h^2)$.

For $q = 1/2$, that is, computations are done with step lengths h and $h/2$, the formula (3.56) simplifies to

$$I_T^{(1)}(h) \approx \frac{I_T(h/2) - (1/4)\,I_T(h)}{1 - (1/4)}$$

$$= \frac{4I_T(h/2) - I_T(h)}{4 - 1} = \frac{4\,I_T(h/2) - I_T(h)}{3}. \tag{3.57}$$

In practical applications, we normally use the sequence of step lengths h, $h/2$, $h/2^2$, $h/2^3$, ...

Suppose, the integral is computed using the step lengths h, $h/2$, $h/2^2$. Using the results obtained with the step lengths $h/2$, $h/2^2$, we get

$$I_T^{(1)}(h/2) \approx \frac{I_T(h/4) - (1/4)\, I_T(h/2)}{1 - (1/4)}$$

$$= \frac{4\, I_T(h/4) - I_T(h/2)}{4 - 1} = \frac{4\, I_T(h/4) - I_T(h/2)}{3}.$$

Both the results $I_T^{(1)}(h), I_T^{(1)}(h/2)$ are of order, $O(h^4)$. Now, we can eliminate the $O(h^4)$ terms of these two results to obtain a result of next higher order, $O(h^6)$. The multiplicative factor is now $(1/2)^4 = 1/16$. The formula becomes

$$I_T^{(2)}(h) \approx \frac{16 I_T^{(1)}(h/2) - I_T^{(1)}(h)}{16 - 1} = \frac{16 I_T^{(1)}(h/2) - I_T^{(1)}(h)}{15}.$$

Therefore, we obtain the Romberg extrapolation procedure for the composite trapezium rule as

$$I_T^{(m)}(h) \approx \frac{4^m\, I_T^{(m-1)}(h/2) - I_T^{(m-1)}(h)}{4^m - 1}, \quad m = 1, 2, \ldots$$

where $I_T^{(0)}(h) = I_T(h)$.

The computed result is of order $O(h^{2m+2})$.

The extrapolations using three step lengths h, $h/2$, $h/2^2$, are given in Table 3.1.

Table 1. Romberg method for trapezium rule.

Step Length	*Value of I* $O(h^2)$	*Value of I* $O(h^4)$	*Value of I* $O(h^6)$
h	$I(h)$	$I^{(1)}(h) = \frac{4I(h/2) - I(h)}{3}$	
$h/2$	$I(h/2)$	$I^{(1)}(h/2) = \frac{4I(h/4) - I(h/2)}{3}$	$I^{(2)}(h) = \frac{16I^{(1)}(h/2) - I^{(1)}(h)}{15}$
$h/4$	$I(h/4)$		

Note that the most accurate values are the values at the end of each column.

Romberg method for the Simpson's 1/3 rule We can apply the same procedure as in trapezium rule to obtain the Romberg's extrapolation procedure for the Simpson's 1/3 rule.

Let I denote the exact value of the integral and I_S denote the value obtained by the composite Simpson's 1/3 rule.

The error, $I - I_S$, in the composite Simpson's 1/3 rule in computing the integral is given by

$$I - I_S = c_1 h^4 + c_2 h^6 + c_3 h^8 + \ldots$$

or

$$I = I_S + c_1 h^4 + c_2 h^6 + c_3 h^8 + \ldots$$

As in the trapezium rule, to illustrate the extrapolation procedure, first consider two error terms.

$$I = I_S + c_1 h^4 + c_2 h^6.$$

Let I be evaluated using two step lengths h and qh, $0 < q < 1$. Let these values be denoted by $I_S(h)$ and $I_S(qh)$. The error equations become

$$I = I_S(h) + c_1 h^4 + c_2 h^6.$$

$$I = I_S(qh) + c_1 q^4 h^4 + c_2 q^6 h^6.$$

From (3.63), we obtain

$$I - I_S(h) = c_1 h^4 + c_2 h^6.$$

From (3.64), we obtain

$$I - I_S(qh) = c_1 q^4 h^4 + c_2 q^6 h^6.$$

Multiply (3.65) by q^4 to obtain

$$q^4[I - I_S(h)] = c_1 q^4 h^4 + c_2 q^4 h^6.$$

Eliminating $c_1 q^4 h^4$ from (3.66) and (3.67), we obtain

$$(1 - q^4)I - I_S(qh) + q^4 I_S(h) = c_2 q^4 h^6 (q^2 - 1).$$

Note that the error term on the right hand side is now of order $O(h^6)$. Solving for I, we obtain

$$I = \frac{I_S(qh) - q^4 I_S(h)}{(1 - q^4)} - \frac{c_2 q^4}{1 + q^2} h^6.$$

Neglecting the $O(h^6)$ error term, we obtain the new approximation to the value of the integral as

$$I \approx I_S^{(1)}(h) = \frac{I_S(qh) - q^4 I_S(h)}{(1 - q^4)}.$$

Again, we note that this value is obtained by suitably using the values of the integral obtained with step lengths h and qh, $0 < q < 1$. This computed result is of order, $O(h^6)$, which is higher than the order of the Simpson's 1/3 rule, which is of $O(h^4)$.

For $q = 1/2$, that is, computations are done with step lengths h and $h/2$, the formula (3.68) simplifies to

$$I_S^{(1)}(h) \approx \frac{I_S(h/2) - (1/16) I_S(h)}{1 - (1/16)}$$

$$= \frac{16\, I_S(h/2) - I_S(h)}{16-1} = \frac{16\, I_S(h/2) - I_S(h)}{15}.$$

In practical applications, we normally use the sequence of step lengths h, $h/2$, $h/2^2$, $h/2^3$, ...

Suppose, the integral is computed using the step lengths h, $h/2$, $h/2^2$. Using the results obtained with the step lengths $h/2$, $h/2^2$, we get

$$I_S^{(1)}(h/2) \approx \frac{I_S(h/4) - (1/16)\, I_S(h/2)}{1-(1/16)}$$

$$= \frac{16\, I_S(h/4) - I_S(h/2)}{16-1} = \frac{16 I_S(h/4) - I_S(h/2)}{15}.$$

Both the results $I_T^{(1)}(h), I_T^{(1)}(h/2)$ are of order, $O(h^6)$. Now, we can eliminate the $O(h^6)$ terms of these two results to obtain a result of next higher order, $O(h^8)$. The multiplicative factor is now $(1/2)^6 = 1/64$. The formula becomes

$$I_S^{(2)}(h) \approx \frac{64\, I_S^{(1)}(h/2) - I_S^{(1)}(h)}{64-1} = \frac{64 I_S^{(1)}(h/2) - I_S^{(1)}(h)}{63}.$$

Therefore, we obtain the Romberg extrapolation procedure for the composite Simpson's 1/3 rule as

$$I_S^{(m)}(h) \approx \frac{4^{m+1} I_S^{(m-1)}(h/2) - I_S^{(m-1)}(h)}{4^{m+1}-1},\ m = 1, 2, \ldots$$

where $I_S^{(0)}(h) = I_S(h)$.

The computed result is of order $O(h^{2m+4})$.

The extrapolations using three step lengths h, $h/2$, $h/2^2$, are given in Table 3.2.

Table 2. Romberg method for Simpson's 1/3 rule.

Step Length	*Value of I* $O(h^4)$	*Value of I* $O(h^6)$	*Value of I* $O(h^8)$
h	$I(h)$	$I^{(1)}(h) = \frac{16I(h/2) - I(h)}{15}$	
$h/2$	$I(h/2)$	$I^{(1)}(h/2) = \frac{16I(h/4) - I(h/2)}{15}$	$I^{(2)}(h) = \frac{64I^{(1)}(h/2) - I^{(1)}(h)}{63}$
$h/4$	$I(h/4)$		

Note that the most accurate values are the values at the end of each column.

Example 11 *The approximations to the values of the integrals in Examples* 3.12 *and* 3.13 *were obtained using the trapezium rule. Apply the Romberg's method to improve the approximations to the values of the integrals.*

Solution In Example 3.12, the given integral is

$$I = \int_0^1 \frac{dx}{1+x}$$

The approximations using the trapezium rule to the integral with various values of the step lengths were obtained as follows.

$h = 1/2$, $N = 2$: $I = 0.708334$; $h = 1/4$, $N = 4$: $I = 0.697024$.

$h = 1/8$, $N = 8$: $I = 0.694122$.

We have $$I^{(1)}(1/2) = \frac{4I(1/4) - I(1/2)}{3} = \frac{4(0.697024) - 0.708334}{3} = 0.693254$$

$$I^{(1)}(1/4) = \frac{4I(1/8) - I(1/4)}{3} = \frac{4(0.694122) - 0.697024}{3} = 0.693155.$$

$$I^{(2)}(1/2) = \frac{16I^{(1)}(1/4) - I^{(1)}(1/2)}{15} = \frac{16(0.693155) - 0.693254}{15} = 0.693148.$$

The results are tabulated in Table 3.3.

Magnitude of the error is

$$| I - 0.693148 | = | 0.693147 - 0.693148 | = 0.000001.$$

Table 3. Romberg method. Example 3.21.

Step Length	*Value of I* $O(h^2)$	*Value of I* $O(h^4)$	*Value of I* $O(h^6)$
1/2	0.708334		
		0.693254	
1/4	0.697024		0.693148
		0.693155	
1/8	0.694122		

In Example 3.13, the given integral is

$$I = \int_1^2 \frac{dx}{5+3x}.$$

The approximations using the trapezium rule to the integral with various values of the step lengths were obtained as follows.

$h = 1/4$, $N = 4$: $I = 0.10627$; $h = 1/8$, $N = 8$: $I = 0.10618$.

We have $I^{(1)}(1/4) = \dfrac{4I(1/8) - I(1/4)}{3} = \dfrac{4(0.10618) - 0.10627}{3} = 0.10615.$

Since the exact value is $I = 0.10615$, the result is correct to all places.

Example 12 *The approximation to the value of the integral in Examples* 3.16 *was obtained using the Simpson's* 1/3 *rule. Apply the Romberg's method to improve the approximation to the value of the integral.*

Solution In Example 3.16, the given integral is

$$I = \int_0^1 \frac{dx}{1+x}.$$

The approximations using the Simpson's 1/3 rule to the integral with various values of the step lengths were obtained as follows.

$h = 1/2$, $n = 2N = 2$: $I = 0.694444$; $h = 1/4$, $n = 2N = 4$: $I = 693254$;

$h = 1/8$, $n = 2N = 8$: $I = 693155$.

We have $$I^{(1)}(1/2) = \frac{16I(1/4) - I(1/2)}{15} = \frac{16\,(0.693254) - 0.694444}{15} = 0.693175$$

$$I^{(1)}(1/4) = \frac{16I(1/8) - I(1/4)}{15} = \frac{16\,(0.693155) - 0.693254}{15} = 0.693148$$

$$I^{(2)}(1/2) = \frac{64I^{(1)}(1/4) - I^{(1)}(1/2)}{63} = \frac{64(0.693148) - 0.693175}{63} = 0.693148.$$

The results are tabulated in Table 3.4.

Magnitude of the error is

$$| I - 0.693148 | = | 0.693147 - 0.693148 | = 0.000001.$$

Table 3 Romberg method. Example 3.22.

Step Length	*Value of I* $O(h^4)$	*Value of I* $O(h^6)$	*Value of I* $O(h^8)$
1/2	0.694444		
		0.693175	
1/4	0.693254		0.693148
		0.693148	
1/8	0.693155		

REVIEW QUESTIONS

1. What is the order of the trapezium rule for integrating $\int_a^b f(x)\,dx$? What is the expression for the error term?

Solution The order of the trapezium rule is 1. The expression for the error term is

$$\text{Error} = -\frac{(b-a)^3}{12} f''(\xi) = -\frac{h^3}{12} f''(\xi), \quad \text{where } a \le \xi \le b.$$

2. When does the trapezium rule for integrating $\int_a^b f(x)dx$ gives exact results?

Solution Trapezium rule gives exact results when $f(x)$ is a polynomial of degree ≤ 1.

3. What is the restriction in the number of nodal points, required for using the trapezium rule for integrating $\int_a^b f(x)dx$?

Solution There is no restriction in the number of nodal points, required for using the trapezium rule.

4. What is the geometric representation of the trapezium rule for integrating $\int_a^b f(x)dx$?

Solution Geometrically, the right hand side of the trapezium rule is the area of the trapezoid with width $b - a$, and ordinates $f(a)$ and $f(b)$, which is an approximation to the area under the curve $y = f(x)$ above the x-axis and the ordinates $x = a$, and $x = b$.

5. State the composite trapezium rule for integrating $\int_a^b f(x)dx$, and give the bound on the error.

Solution The composite trapezium rule is given by

$$\int_a^b f(x)dx = \frac{h}{2}\,[f(x_0) + 2\{f(x_1) + f(x_2) + \ldots + f(x_{n-1})\} + f(x_n)]$$

where $nh = (b - a)$. The bound on the error is given by

$$|\text{ Error }| \le \frac{nh^3}{12} M_2 = \frac{(b-a)h^2}{12} M_2$$

where $$M_2 = \max_{a \le x \le b} |f''(x)| \text{ and } nh = b - a.$$

6. What is the geometric representation of the composite trapezium rule for integrating $\int_a^b f(x)dx$?

Solution Geometrically, the right hand side of the composite trapezium rule is the sum of areas of the n trapezoids with width h, and ordinates $f(x_{i-1})$ and $f(x_i)$ $i = 1, 2, \ldots, n$. This

sum is an approximation to the area under the curve $y = f(x)$ above the x-axis and the ordinates $x = a$ and $x = b$.

7. How can you deduce that the trapezium rule and the composite trapezium rule produce exact results for polynomials of degree less than or equal to 1?

Solution The expression for the error in the trapezium rule is given by

$$R_1(f, x) = -\frac{h^3}{12} f''(\xi)$$

and the expression for the error in the composite trapezium rule is given by

$$R_1(f, x) = -\frac{h^3}{12} [f''(\xi_1) + f''(\xi_2) + \ldots + f''(\xi_n)], \quad x_{n-1} < \xi_n < \xi_n.$$

If $f(x)$ is a polynomial of degree ≤ 1, then $f''(x) = 0$. This result implies that error is zero and the trapezium rule produces exact results for polynomials of degree ≤ 1.

8. When does the Simpson's 1/3 rule for integrating $\int_a^b f(x)dx$ gives exact results?

Solution Simpson's 1/3 rule gives exact results when $f(x)$ is a polynomial of degree ≤ 3.

9. What is the restriction in the number of nodal points, required for using the Simpson's 1/3 rule for integrating $\int_a^b f(x)dx$?

Solution The number of nodal points must be odd for using the Simpson's 1/3 rule or the number of subintervals must be even.

10. State the composite Simpson's 1/3 rule for integrating $\int_a^b f(x)dx$, and give the bound on the error.

Solution Let $n = 2N$ be the number of subintervals. The composite Simpson's 1/3 rule is given by

$$\begin{aligned}\int_a^b f(x)dx &= \frac{h}{3} [\{f(x_0) + 4f(x_1) + f(x_2)\} + \{f(x_2) + 4f(x_3) + f(x_4)\} + \ldots \\ &\quad + \{f(x_{2N-2}) + 4\, f(x_{2N-1}) + f(x_{2N})\}] \\ &= \frac{h}{3} [f(x_0) + 4\{f(x_1) + f(x_3) + \ldots + f(x_{2N-1})\} \\ &\quad + 2\{f(x_2) + f(x_4) + \ldots + f(x_{2N-2})\} + f(x_{2N})]\end{aligned}$$

The bound on the error is given by

$$\begin{aligned}|\, R(f, x)\, | &\leq \frac{h^5}{90} \left[|\, f^{(4)}\, (\xi_1)\, | + |\, f^{(4)}\, (\xi_2)\, | + \ldots + |\, f^{(4)}\, (\xi_N)\, | \right] \\ &\leq \frac{Nh^5}{90} M_4 = \frac{(b-a)h^4}{180} M_4\end{aligned}$$

where $x_0 < \xi_1 < x_2$, $x_2 < \xi_2 < x_4$, etc., $M_4 = \max_{a \le x \le b} |f^{(4)}(x)|$ and $N h = (b - a)/2$.

11. How can you deduce that the Simpson's 1/3 rule and the composite Simpson's 1/3 rule produce exact results for polynomials of degree less than or equal to 3?

Solution The expression for the error in the Simpson's 1/3 rule is given by

$$R(f, x) = \frac{c}{4!} f^{(4)}(\xi) = -\frac{(b-a)^5}{2880} f^{(4)}(\xi) = -\frac{h^5}{90} f^{(4)}(\xi)$$

where $h = (b - a)/2$, and $a \le \xi \le b$.

The expression for the error in the composite Simpson's 1/3 rule is given by

$$R(f, x) = -\frac{h^5}{90} [f^{(4)}(\xi_1) + f^{(4)}(\xi_2) + \ldots + f^{(4)}(\xi_N)]$$

where $x_0 < \xi_1 < x_2$, $x_2 < \xi_2 < x_4$, etc.

If $f(x)$ is a polynomial of degree ≤ 3, then $f^{(4)}(x) = 0$. This result implies that error is zero and the Simpson 1/3 rule produces exact results for polynomials of degree ≤ 3.

12. What is the restriction in the number of nodal points, required for using the Simpson's 3/8 rule for integrating $\int_a^b f(x)dx$?

Solution The number of subintervals must be divisible by 3.

13. What are the disadvantages of the Simpson's 3/8 rule compared with the Simpson's 1/3 rule?

Solution The disadvantages are the following: (*i*) The number of subintervals must be divisible by 3. (*ii*) It is of the same order as the Simpson's 1/3 rule, which only requires that the number of nodal points must be odd. (*iii*) The error constant c in the case of Simpson's 3/8 rule is $c = 3/80$, which is much larger than the error constant $c = 1/90$, in the case of Simpson's 1/3 rule. Therefore, the error in the case of the Simpson's 3/8 rule is larger than the error in the case Simpson 1/3 rule.

14. Explain why we need the Romberg method.

Solution In order to obtain accurate results, we compute the integrals by trapezium or Simpson's rules for a number of values of step lengths, each time reducing the step length. We stop the computation, when convergence is attained (usually, the magnitude of the difference between successive values of the integrals obtained with the reducing values of the step lengths is less than a given accuracy). Convergence may be obtained after computing the value of the integral with a number of step lengths. While computing the value of the integral with a particular step length, the values of the integral obtained earlier by using larger step lengths were not used. Further, convergence may be slow. Romberg method is a powerful tool which uses the method of extrapolation. Romberg method uses these computed values of the integrals obtained with various step lengths, to refine the solution such that the new values are of higher order. That is, as if they are obtained using a higher order method than the order of the method used.