CHAPTER 3

A BRIEF ABOUT MATHEMATICAL ASPECTS OF NONLINEAR MODELS
CHAPTER 3

A BRIEF ABOUT MATHEMATICAL ASPECTS OF NONLINEAR MODELS

3.1 INTRODUCTION:

Mathematical models involving nonlinearity are becoming increasingly popular in recent years. In practice nonlinear mathematical models have a wide variety of applications. A mathematical model is said to be nonlinear model, if the derivatives of the model with respect to the model parameters depend on one or more parameters. This aspect is essential to distinguish nonlinear from curve linear model. Generally, the direct interpretation of the process under study can be described by the model parameters.

Several theorems in mathematical analysis are stated as preliminaries for nonlinear models. Certain numerical analysis techniques to obtain solutions of algebraic and transcendental equations have been discussed to use in iterative estimation procedures for estimating parameters of nonlinear models.

3.2 MATHEMATICAL PRELIMINARIES FOR NONLINEAR MODELS:

The following mathematical preliminaries will be useful in the present study:

Intermediate Value Theorems:

(i) If $f(x)$ is continuous in $a \leq x \leq b$ and if $f(a)$ and $f(b)$ are of opposite signs then $f(\xi) = 0$ for atleast one number $\xi$ such that $a < \xi < b$

(ii) Let $f(x)$ be continuous in $[a, b]$ and $k$ be any number between $f(a)$ and $f(b)$. Then there exists atleast one number $\xi$ in $(a, b)$ such that $f(\xi) = k$. 
Rolle’s Theorem:

If
1. \( f(x) \) is continuous in \( a \leq x \leq b \)
2. \( f'(x) \) exists in \( a < x < b \) and
3. \( f(a) = f(b) \)

then there exists at least one value of \( x \), say \( \xi \) such that \( f'(\xi) = 0 \), \( a < \xi < b \)

Generalized Rolle’s Theorem:

Let \( f(x) \) be a function which is \( n \) times differentiable on \( [a, b] \). If \( f(x) \) vanishes at \( (n+1) \) distinct points \( x_0, x_1, \ldots, x_n \) in \( (a, b) \) then there exists a number \( \xi \) in \( (a, b) \) such that \( f^{(n)}(\xi) = 0 \)

Mean-value theorem for derivatives:

If
1. \( f(x) \) is continuous in \( [a, b] \) and
2. \( f'(x) \) exists in \( (a, b) \)

then there exists at least one value of \( x \) say \( \xi \) between \( a \) and \( b \) such that

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

By assuming \( b = a + h \) one can find

\[
 f(a + h) = f(a) + hf'(a + \theta h), \quad 0 < \theta < 1
\]

Taylor’s series for a function of one variable:

If \( f(x) \) is continuous and possesses continuous derivatives of order \( n \) in an interval that includes \( x = a \) then in that interval

\[
 f(x) = f(a) + (x - a)f'(a) + \frac{(x - a)^2}{2}f''(a) + \ldots + \frac{(x - a)^n}{n!}f^{(n)}(a) + R_n(x)
\]

where \( R_n(x) \), the remainder term can be expressed in the form

\[
 R_n(x) = \frac{(x - a)^n}{n!}f^{(n)}(\xi), \quad a < \xi < x
\]
Maclaurin’s Expansion:

\[ f(x) = f(0) + xf'(0) + \frac{x^2}{2} f''(0) + \ldots + \frac{x^n}{n!} f^{(n)}(0) + \ldots \] ........................(3.2.1)

Taylor’s series for a function of several variables:

\[
f (x_1 + \Delta x_1, x_2 + \Delta x_2) = f(x_1, x_2) + \frac{\partial f}{\partial x_1} \Delta x_1 + \frac{\partial f}{\partial x_2} \Delta x_2 +
\]

\[
\frac{1}{2} \left[ \frac{\partial^2 f}{\partial x_1^2} (\Delta x_1)^2 + 2 \frac{\partial^2 f}{\partial x_1 \partial x_2} \Delta x_1 \Delta x_2 + \frac{\partial^2 f}{\partial x_2^2} (\Delta x_2)^2 \right] + \ldots \] ........................(3.2.2)

Taylor’s series for a function of several variables:

\[
f (x_1 + \Delta x_1, x_2 + \Delta x_2, \ldots, x_n + \Delta x_n) = f(x_1, x_2, \ldots, x_n) \\
+ \frac{\partial f}{\partial x_1} \Delta x_1 + \frac{\partial f}{\partial x_2} \Delta x_2 + \ldots + \frac{\partial f}{\partial x_n} \Delta x_n +
\]

\[
\frac{1}{2} \left[ \frac{\partial^2 f}{\partial x_1^2} (\Delta x_1)^2 + \ldots + \frac{\partial^2 f}{\partial x_n^2} (\Delta x_n)^2 + 2 \frac{\partial^2 f}{\partial x_1 \partial x_2} \Delta x_1 \Delta x_2 + \ldots + 2 \frac{\partial^2 f}{\partial x_{n-1} \partial x_n} \Delta x_{n-1} \Delta x_n \right] + \ldots \] ........................(3.2.3)

3.3. SOLUTION OF ALGEBRAIC AND TRANSCENDENTAL EQUATIONS:

In research work, a frequently occurring problem is to find the roots of the equation of the form

\[ f(x) = 0 \] .................................(3.3.1)

If \( f(x) \) is quadratic, cubic or a biquadratic expression then algebraic formulae are available for expressing the roots in terms of the coefficients. On the other hand when \( f(x) \) is a polynomial of higher degree or an expression involving transcendental functions, algebraic methods are not available and approximate methods are to be performed.
Here some numerical methods for the solution of (3.2.1) will be discussed where \( f(x) \) is algebraic or transcendental or a combination of both.

**Bisection Method:**

This method is based on Intermediate Value Theorem which states that if a function \( f(x) \) is continuous in \([a, b]\) and \( f(a)f(b)<0 \) then there exists at least one root for \( f(x)=0 \) between \( a \) and \( b \). For definiteness let \( f(a) \) be positive and \( f(b) \) be negative. Then the root lies between \( a \) and \( b \) and let its approximate value be given by \( x_0 = \frac{a+b}{2} \). If \( f(x_0) = 0 \) then obviously \( x_0 \) is a root of (3.3.1). If not then the root lies either between \( x_0 \) and \( b \) or between \( x_0 \) and \( a \) depending on the sign of \( f(x_0) \).

Let this new interval be \([a_1, b_1]\) whose length is \( \frac{1}{2}(b-a) \). As above this is bisected at \( x_1 \) and the new interval possesses length \( \frac{1}{4}(b-a) \). This process is repeated until the latest interval which contains the root is as small as desired say \( \varepsilon \). At each step the length of the interval is reduced by a factor of one-half and at the end of the \( n^a \) step the new interval will be \([a_n, b_n]\) of length \( \frac{1}{2^n}(b-a) \). Now

\[
\frac{b-a}{2^n} \leq \varepsilon
\]

Multiplying by \( \frac{2^n}{\varepsilon} \)

\[
\frac{b-a}{\varepsilon} \leq 2^n
\]

\[
\frac{\log \frac{b-a}{\varepsilon}}{\log 2} \leq n \text{ ..........................................................(3.3.2)}
\]

This inequation gives the number of iterations required to get an accuracy \( \varepsilon \).
The Iteration Method:

This method solves \( x = F(x) \) by the recursion \( x_n = F(x_{n-1}) \) and converges to a root if \( |F'(x)| \leq L < 1 \). The error \( e_n = r - x_n \) where \( r \) is the exact root has the property

\[
e_n = F'(r)e_{n-1}
\]

so that each iteration reduces the error by a factor near \( F'(r) \). If \( F'(r) \) is near 1 this is slow convergence.

Regula Falsi Method:

Let two points \( x_0 \) and \( x_1 \) be such that \( f(x_0) \) and \( f(x_1) \) are of opposite signs and \( f \) is continuous on \([x_0, x_1]\). Since the graph of \( y = f(x) \) crosses the \( x \)-axis between these two points, a root must lie between these points. Now the equation of the chord joining the two points \((x_0, f(x_0))\) and \((x_1, f(x_1))\) is

\[
(x - x_0)(f(x_0) - f(x_1)) = (y - f(x_0))(x_0 - x_1)
\]

The method consists in replacing the part of the curve between the points \((x_0, f(x_0))\) and \((x_1, f(x_1))\) by means of the chord joining these points and taking the point of intersection of the chord with the \( x \)-axis as an approximation to the root. The point of intersection in the present case is given by putting \( y = 0 \) in (3.3.4).

Hence the second approximation to the root of \( f(x) = 0 \) is given by

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

If now \( f(x_1) \) and \( f(x_0) \) are of opposite signs then the root lies between \( x_0 \) and \( x_2 \) and replace \( x_0 \) by \( x_1 \) in (3.3.5) to get the next approximation. Otherwise
replace \( x_0 \) by \( x_1 \) and generate the next approximation. The procedure is repeated
till the root is obtained to the desired accuracy.

Newton - Raphson Method:

Let \( x_0 \) be an approximation to the desired root of \( f(x) = 0 \) and let

\[ x_1 = x_0 + h \]

be the correct root so that \( f(x_1) = 0 \). Expanding \( f(x_0 + h) \) by Taylor's
series one can get

\[
f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \ldots = 0 \tag{3.3.6}
\]

Neglecting the derivatives form second order onwards, one may find

\[
f(x_0) + hf'(x_0) = 0 \tag{3.3.7}
\]

This gives

\[
h = -\frac{f(x_0)}{f'(x_0)} \tag{3.3.8}
\]

A better approximation than \( x_0 \) is therefore given by \( x_1 \), where

\[
x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \tag{3.3.9}
\]

Successive approximations are given by \( x_1, x_2, \ldots, x_n \),

where \( x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (n=0,1,2,\ldots) \tag{3.3.10}
\]

This is known as Newton - Raphson formula.

If \( f'(x) \) is complicated the previous iterative method may be preferable, but

Newton's method converges such more rapidly and usually gets the nod. The error

\( e_n \) here satisfies

\[
e_n = \frac{f''(x)}{2f'(x)} e_{n-1} \tag{3.3.1}
\]
This is known as quadratic convergence, each error roughly proportional to the square of the previous error. The number of correct digits almost doubles with each iteration. The square root iteration

$$x_n = \frac{1}{2} \left( x_{n-1} + \frac{Q}{x_{n-1}} \right) \tag{3.3.12}$$

is a special case of Newton's method corresponding to $f(x) = x^2 - Q$. It converges quadratically to the positive square root of $Q$, for $Q > 0$.

The more general root finding formula

$$x_n = x_{n-1} - \frac{x_n^p - Q}{p x_n^{p-1}} \tag{3.3.13}$$

is a special case of Newton's method. It produces a $p$th root of $Q$.

**Generalized Newton's Method:**

If $\xi$ is a multiple root of order $p$ of $f(x) = 0$ then the iteration formula corresponding to (3.3.10) is written as

$$x_{n+1} = x_n - \frac{f(x_n)}{p f'(x_n)} \tag{3.3.14}$$

This means that $\frac{1}{p} f'(x_n)$ is the slope of the straight line passing through $(x_n, y_n)$ and intersecting the $x$-axis at the point $(x_{n+1}, 0)$.

**Muller's Method:**

In this method $f(x)$ is approximated by a second degree curve in the nbd of a root. The roots of the quadratic are then assumed to be the approximations to the roots of the equation $f(x) = 0$. The method is iterative, converges almost quadratically and can be used to obtain non real complex roots.
Let \( x_{i-2}, x_{i-1}, x_i \) be three different approximations to a root of \( f(x) = 0 \) and let \\
\( y_{i-2}, y_{i-1}, \) and \( y_i \) be the corresponding values of \( y = f(x) \).

Let \( P(x) = A(x - x_i)^2 + B(x - x_i) + y_i \) be the parabola passing through the points \\
\((x_{i-2}, y_{i-2}), (x_{i-1}, y_{i-1}), \) and \((x_i, y_i)\). Then

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

and 

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

From equations (3.3.16) and (3.3.17), one may get

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

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

Solutions of equations (3.3.18) and (3.3.19) give

\[
A = (x_{i-2} - x_i)(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i-2} - y_i)
\]

\[
B = \frac{(x_{i-2} - x_i)^2(y_{i-1} - y_i) - (x_{i-1} - x_i)^2(y_{i-2} - y_i)}{(x_{i-1} - x_i)(x_{i-2} - x_i)(x_{i-2} - x_i)}
\]

With the above values of \( A \) and \( B \) the quadratic equation 

\[
A(x - x_i)^2 + B(x - x_i) + y_i = 0
\]

gives the next approximation

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

Ramanujan’s Method:

Srinivasa Ramanujan described an iterative method which can be used to compute the smallest root of the equation

\[
f(x) = 0
\]

where 

\[
f(x) = 1 - (a_1x + a_2x^2 + a_3x^3 + a_4x^4 + ...)
\]
For smaller values of $x$, one can write

$$\left[ 1 - \left( a_1 x + a_2 x^2 + a_3 x^3 + \ldots \right) \right]^{-1} = b_1 + b_2 x + b_3 x^2 + \ldots \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.3.25)$$

Expanding LHS by binomial theorem for negative integer gives

$$1 + \left( a_1 x + a_2 x^2 + a_3 x^3 + \ldots \right) + \left( a_1 x + a_2 x^2 + a_3 x^3 + \ldots \right)^2 + \ldots = b_1 + b_2 x + b_3 x^2 + \ldots \quad (3.3.26)$$

Comparing the coefficients of like powers of $x$, one can get

\[
\begin{align*}
    b_1 &= 1 \\
    b_2 &= a_1 = a_1 (1) = a_1 b_1 \\
    b_3 &= a_1^2 + a_2 = a_1 a_1 + a_2 (1) = a_1 b_2 + a_2 b_1 \\
    \ldots \ldots \\
    b_n &= a_1 b_{n-1} + a_2 b_{n-2} + \ldots + a_n b_1 \\
    n &= 2, 3, \ldots \ldots
\end{align*}
\]

\[\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (3.3.27)\]

Ramanujan states that the successive convergents namely $\frac{b_n}{b_{n-1}}$ approach a root of the equation $f(x) = 0$.

Graeffe’s Root-Squaring method:

Let $P_n(x)$ be a polynomial of degree $n$. Graeffe’s method consists in transforming $P_n(x)$ into another polynomial say $Q_n(z)$ of the same degree but whose roots are the squares of the roots of the original polynomial. The process is repeated so that the roots of the new polynomial are distributed more spaciously.

This is possible provided that the roots of the original polynomial are all real and distinct. The roots are finally computed directly from the coefficients.

Lin-Balrastow’s method:

This is often used in estimating the factors of polynomials. Let the polynomial be given by

44
Let $x^2 + Rx + S$ be a quadratic factor of $f(x)$ and let an approximate factor be $x^2 + rx + s$. Usually first approximations to $r$ and $s$ can be obtained from the last three terms of the given polynomial. Thus

$$r = \frac{A_1}{A_2} \quad \text{and} \quad s = \frac{A_3}{A_2} \quad \text{(3.3.29)}$$

Let

$$f(x) = (x^2 + rx + s)(B_1x + B_2) + Cx + D \quad \text{(3.3.30)}$$

where the constants $B_1, B_2, C$ and $D$ are to be determined. Equating the coefficients of like powers of $x$ in equations (3.3.28), (3.3.30) one may get

$$\begin{align*}
B_1 &= A_1, \\
B_2 &= A_1 - rB_1, \\
C &= A_1 - rB_1 - sB_2, \\
D &= A_1 - sB_1
\end{align*} \quad \text{(3.3.31)}$$

From (3.3.31) it is clear that the coefficients $B_1, B_2$ of the factored polynomial and also the coefficients $C$ and $D$ are functions of $r$ and $s$. Since $x^2 + Rx + S$ is a factor of the given polynomial it follows that

$$\begin{align*}
C(R,S) &= 0, \\
\text{and} \quad D(R,S) &= 0
\end{align*} \quad \text{(3.3.32)}$$

Letting $R = r + \Delta r$ and $S = s + \Delta s \quad \text{(3.3.33)}$
equations (3.3.32) can be expanded by Taylor’s series

\[
\begin{align*}
C(R, S) &= C(r, s) + \Delta r \frac{\partial C}{\partial r} + \Delta s \frac{\partial C}{\partial s} = 0 \\
D(R, S) &= D(r, s) + \Delta r \frac{\partial D}{\partial r} + \Delta s \frac{\partial D}{\partial s} = 0
\end{align*}
\]  

(3.3.34)

where the derivatives are to be calculated at \( r \) and \( s \). Equations (3.3.34) can then be solved for \( \Delta r \) and \( \Delta s \). Using these values in (3.3.33) will give the next approximation to \( R \) and \( S \). The process can be repeated until successive values of \( R \) and \( S \) show no significant change.

Quotient – Difference method:

Let the given cubic equation be

\[ f(x) = a_0 x^3 + a_1 x^2 + a_2 x + a_3 = 0 \]  

(3.3.35)

and let \( x_1, x_2 \) and \( x_3 \) be its roots such that \( 0 < |x_1| < |x_2| < |x_3| \)

Then \( f(x) \) can be expressed as given below

\[
(f(x))^{-1} = \sum_{i=1}^{3} \frac{P_i}{x - x_i} = \sum_{i=1}^{3} \frac{P_i}{x - x_i} \left( \frac{1 - x}{x_i} \right)
\]

\[ = \sum_{i=1}^{3} \frac{P_i}{x_i} \left( 1 + \frac{x}{x_i} + \frac{x^2}{x_i^2} + \ldots \right) = \sum_{i=1}^{3} \frac{P_i}{x_i} \left( \frac{x}{x_i} + \frac{x^2}{x_i^2} + \frac{x^3}{x_i^3} + \ldots \right) = \sum_{i=0}^{\infty} \alpha_i x_i
\]

(3.3.36)

where \( \alpha_i = -\sum_{i=1}^{3} \frac{P_i}{x_i^{i+1}} \)  

(3.3.37)

Quotients and differences are defined by the relations
Substituting (3.3.37) in (3.3.38), one can get

\[ q_i^{(i)} = \frac{p_1 + p_2}{x_1^{(i)} + x_2^{(i)}} = \frac{1}{x_1} \cdot \frac{p_1 + p_2 \left( \frac{x_1}{x_2} \right)^{i} + p_3 \left( \frac{x_1}{x_3} \right)^{i}}{p_1 + p_2 \left( \frac{x_1}{x_2} \right)^{i} + p_3 \left( \frac{x_1}{x_3} \right)^{i}} \]

Since \( \left| \frac{x_1}{x_2} \right| < 1 \) and \( \left| \frac{x_1}{x_3} \right| < 1 \), one can have

\[ \lim_{i \to \infty} q_i^{(i)} = \frac{1}{x_1} \] (3.3.40)

Thus the quotients \( q_i^{(i)} \) approach, in the limit, the reciprocal of a root of the given cubic equation (3.3.35).

In a similar way it can be shown that

\[ \lim_{i \to \infty} \frac{1}{x_1} - q_i^{(i)} = \frac{1}{x_1} \cdot \frac{p_2 \left( 1 - \frac{x_1}{x_2} \right)}{x_1 p_1 \left( 1 - \frac{x_1}{x_2} \right)} \] (3.3.41)

and

\[ \lim_{i \to \infty} \frac{1}{x_1} - q_i^{(i)} = \frac{1}{x_1} \cdot \frac{p_2 \left( 1 - \frac{x_1}{x_2} \right)}{x_1 p_1 \left( 1 - \frac{x_1}{x_2} \right)} \] (3.3.42)
Subtracting (3.3.42) from (3.3.41) and using (3.3.39) one can get

$$\lim_{x \to x_0} \frac{\Delta^{(0)}}{x_1} = \left(1 - \frac{1}{x_1} \right) \left( \frac{x_1}{x_2} \right)$$

...(3.3.43)

Similarly

$$\lim_{x \to x_0} \frac{\Delta^{(m)}}{x_1} = \left(1 - \frac{1}{x_1} \right) \left( \frac{x_1}{x_2} \right)$$

...(3.3.44)

Form (3.3.43) and (3.3.44) one can get

$$\frac{\Delta^{(0)}}{\Delta^{(m)}} = \frac{x_1}{x_1} \left( \frac{x_1}{x_2} \right)$$

...(3.3.45)

Put

$$\frac{1}{x_2} = \lim_{x \to x_2} q_{(1)}^{(0)}$$

...(3.3.46)

and using (3.3.40), eq (3.3.45) gives

$$\Delta_1^{(0)} q_{1}^{(0)} = \Delta_1^{(m)} q_{1}^{(m)}$$

...(3.3.47)

Following (3.3.39) define

$$\Delta_2^{(0)} - \Delta_1^{(m)} = q_{1}^{(m)} - q_{1}^{(0)}$$

...(3.3.48)

In the similar way with (3.3.45) it can be shown that

$$\lim_{x \to x_0} \frac{\Delta^{(0)}}{\Delta^{(m)}} \frac{1}{x_1}$$

...(3.3.49)

The relations (3.3.47), (3.3.48) can be easily generalized as

$$\Delta_k^{(0)} q_{k+1}^{(0)} = \Delta_k^{(m)} q_{k}^{(m)}$$

...(3.3.50)
3.4. SOLUTION OF SYSTEMS OF NONLINEAR EQUATIONS:

(i) The method of iteration:

Let the equations be given by
\[
\begin{align*}
    f(x, y) &= 0 \\
    g(x, y) &= 0
\end{align*}
\] .................................(3.4.1)

whose real roots are required within a specified accuracy. Put the equations in the form
\[
\begin{align*}
    x &= F(x, y) \\
    y &= G(x, y)
\end{align*}
\] .................................(3.4.2)

where the functions F and G satisfy the equations
\[
\begin{align*}
    \left| \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} \right| < 1 \\
    \text{and} \quad \left| \frac{\partial G}{\partial x} + \frac{\partial G}{\partial y} \right| < 1
\end{align*}
\] .................................(3.4.3)

in the neighbourhood of the root.

Let \((x_0, y_0)\) be the initial approximation to a root \((\alpha, \beta)\) of the system (3.4.1). Then construct the successive approximations according to the formulae
\[ \begin{align*}
  x_1 &= F(x_0, y_0), & y_1 &= G(x_0, y_0) \\
  x_2 &= F(x_1, y_1), & y_2 &= G(x_1, y_1) \\
  &\vdots & &\vdots \\
  x_n &= F(x_{n-1}, y_{n-1}), & y_n &= G(x_{n-1}, y_{n-1}) \\
  x_{n+1} &= F(x_n, y_n), & y_{n+1} &= G(x_n, y_n)
\end{align*} \]  

\[(3.4.4)\]

For faster convergence recently computed values of \( x \) may be used in the evaluation of \( y_i \) in (3.4.4). If the iteration process (3.4.4) converges then one can get

\[\begin{align*}
  \alpha &= F(\alpha, \beta) \\
  \beta &= G(\alpha, \beta)
\end{align*}\]  

\[(3.4.5)\]

in the limit. Thus \( \alpha \) and \( \beta \) are the roots of the system (3.4.2).

**Theorem:**

Let \( x = \alpha \) and \( y = \beta \) be one pair of roots of the system \( x = F(x, y) \) and \( y = G(x, y) \) in the closed neighbourhood \( R \).

If (a) the functions \( F \) and \( G \) and their first partial derivatives are continuous in \( R \)

\[\begin{align*}
  \left| \frac{\partial F}{\partial x} \right| + \left| \frac{\partial F}{\partial y} \right| < 1 \\
  \left| \frac{\partial G}{\partial x} \right| + \left| \frac{\partial G}{\partial y} \right| < 1
\end{align*}\]  

\[(3.4.6)\]

for all \( x, y \) in \( R \) and

(c) the initial approximation \( (x_0, y_0) \) is chosen in \( R \) then the sequence of approximations given by (3.4.4) converges to the roots \( x = \alpha \) and \( y = \beta \) of the system (3.4.1).
Newton-Raphson method:

Let \((x_0, y_0)\) be an initial approximation to the root of the system

\[
\begin{align*}
 f(x, y) &= 0 \\
 g(x, y) &= 0
\end{align*}
\]

\------------------ (3.4.7) \------------------

If \((x_0 + h, y_0 + k)\) is the root of the system then

\[
\begin{align*}
 f(x_0 + h, y_0 + k) &= 0 \\
 g(x_0 + h, y_0 + k) &= 0
\end{align*}
\]

\------------------ (3.4.8) \------------------

Let \(f\) and \(g\) be sufficiently differentiable. Expanding (3.4.8) by Taylor's series gives

\[
\begin{align*}
 f_0 + h \frac{\partial f}{\partial x_0} + k \frac{\partial f}{\partial y_0} + \ldots &= 0 \\
 g_0 + h \frac{\partial g}{\partial x_0} + k \frac{\partial g}{\partial y_0} + \ldots &= 0
\end{align*}
\]

\------------------ (3.4.9) \------------------

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

Neglecting the second and higher order terms one can obtain the system of linear equations

\[
\begin{align*}
 h \frac{\partial f}{\partial x_0} + k \frac{\partial f}{\partial y_0} &= -f_0 \\
 h \frac{\partial g}{\partial x_0} + k \frac{\partial g}{\partial y_0} &= -g_0
\end{align*}
\]

\------------------ (3.4.10) \------------------
If the Jacobian

\[ J(f, g) = \begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{vmatrix} \] (3.4.11)

does not vanish then the linear equations (3.4.10) possess a unique solution given by

\[ h = \begin{bmatrix} -f \\ -g \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{bmatrix} \] and \[ k = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial g}{\partial x} \end{bmatrix} \begin{bmatrix} -f \\ -g \end{bmatrix} \] (3.4.12)

The new approximations are then given by

\[ \begin{align*}
  x_i &= x_0 + h \\
  y_i &= y_0 + k
\end{align*} \] (3.4.13)

The process is to be repeated till one can obtain the roots to the desired accuracy. This method, if possible, possesses quadratic convergence. The following gives the conditions which are sufficient for convergence.

Theorem:

Let \((x_0, y_0)\) be an approximation to a root \((\xi, \eta)\) of the system (3.4.1) in the closed neighbourhood \(R\) containing \((\xi, \eta)\). If \(f, g\) and all their first and second
derivatives are continuous and bounded in \( R \) and (b) \( J(f,g) = 0 \) in \( R \) then the sequence of approximations given by

\[
\begin{align*}
x_{i+1} &= x_i - \left( \frac{f}{\partial y} \frac{g}{\partial y} - \frac{1}{J(f,g)} \right) \\
y_{i+1} &= y_i - \left( \frac{g}{\partial x} \frac{f}{\partial x} - \frac{1}{J(f,g)} \right)
\end{align*}
\]

converge to the root \((\xi, \eta)\) of the system (3.4.1)

3.5. TAYLOR POLYNOMIAL AND FINITE DIFFERENCE OPERATORS:

The Taylor polynomial is the ultimate in osculation. For a single argument \( x_0 \) the values of the polynomial and its first \( n \) derivatives are required to match those of a given function \( y(x) \). That is

\[
p^{(i)}(x_0) = y^{(i)}(x_0) \quad \text{for} \quad i = 0, 1, 2, \ldots, n \quad \ldots \quad (3.5.1)
\]

The Taylor formula settles the existence issue directly by exhibiting such a polynomial in the form

\[
p(x) = \sum_{i=0}^{n} \frac{y^{(i)}(x_0)}{i!} (x-x_0)^i \quad \ldots \quad (3.5.2)
\]

The error of the Taylor polynomial when viewed as an approximation to \( y(x) \) can be expressed by the integral formula

\[
y(x) - p(x) = \frac{1}{i!} \int_{x_0}^{x} y^{(i+1)}(x_0)(x-x_0)^i \, dx_0 \quad \ldots \quad (3.5.3)
\]

Lagrange's error formula may be deduced by applying a mean value theorem to the integral formula. It is
\[ y(x) - p(x) = \frac{y^{(n+1)}(\xi)}{n+1}(x - x_0)^{n+1} \]  \hspace{1cm} (3.5.4)

and clearly resembles the error formulae of collocation ad osculation.

If the derivatives of \( y(x) \) are bounded independently of \( n \) then either error formula serves to estimate the degree \( n \) required to reduce \( |y(x) - p(x)| \) below a prescribed tolerance over a given interval of arguments \( x \).

Analytic functions have the property that for \( n \) tending to infinity the above error of approximation has limit zero for all arguments \( x \) in a given interval. Such functions are then represented by the Taylor series

\[ y(x) = \sum_{i=0}^{\infty} \frac{y^{(i)}(x_0)}{i!}(x - x_0)^{i} \]  \hspace{1cm} (3.5.5)

The binomial series is an especially important case of the Taylor series. For \(-1 < x < 1\) one can have

\[ (1+x)^n = \sum_{i=0}^{\infty} \binom{n}{i} x^i \]  \hspace{1cm} (3.5.6)

### Differentiation Operator \( D \):

The differentiation operator \( D \) is defined by

\[ D = h \frac{d}{dx} \]  \hspace{1cm} (3.5.7)

The exponential operator may then be defined by

\[ e^{D} = \sum_{i=0}^{\infty} \frac{k!D^i}{i!} \]  \hspace{1cm} (3.5.8)

and the Taylor series in operator form becomes

\[ y(x_0) = e^{D}y_0(x_0) \]  \hspace{1cm} (3.5.9)

The relationship between \( D \) and \( \Delta \) may be expressed in either of the forms
\[ \Delta + 1 = e^\Delta \] 
\[ D = \Delta - \frac{1}{2} \Delta^2 + \frac{1}{3} \Delta^3 \] .................................(3.5.10)

both of which involve “infinite series” operator.

The Euler transformation is another useful relationship between two infinite series operators. It may be written as

\[ (1 + E)^{-1} = \frac{1}{2} \left( 1 - \frac{1}{2} \Delta + \frac{1}{4} \Delta^2 - \frac{1}{8} \Delta^3 + \ldots \right) \] .................................(3.5.11)

by using the binomial series.

The Bernoulli numbers \( B_i \) are defined by

\[ \frac{x}{e^x - 1} = \sum_{i=0}^{\infty} \frac{B_i}{i!} x^i \] .................................(3.5.12)

Actually expanding the LHS into its Taylor series one can find \( B_0 = 1, B_1 = -\frac{1}{2}, B_2 = \frac{1}{6} \), and so on.

These numbers occur in various operator equations. For example the indefinite summation operator \( \Delta^{-1} \) is defined by

\[ \Delta^{-1} F_k = y_k \quad F_k = \Delta^{-1} y_k \]

and is related to \( D \) by

\[ \Delta^{-1} = D^{-1} \sum_{i=0}^{\infty} \frac{1}{i!} B_i D^i \] .................................(3.5.13)

where \( B_i \) are Bernoulli numbers. The operator \( D^{-1} \) is the familiar indefinite integral operator.

The Euler-Maclaurin formula may be deduced from the previous relationship

\[ \sum_{i=0}^{\infty} y_i = \sum_{i=0}^{\infty} y_i \Delta x - \frac{1}{2} (y_s - y_e) + \frac{h}{12} (y_s' - y_e') + \ldots \] .................................(3.5.14)
and is often used for the evaluation of either sums or integrals. The powers of \( D \) may be expressed in terms of the central difference operator \( \delta \) by using Taylor series. Some examples are the following

\[
D = \mu \left( \delta - \frac{1}{3!} \delta^3 + \frac{1^2}{5!} \delta^5 - \frac{1^2 \cdot 2^2}{7!} \delta^7 + \ldots \right) \quad \text{...................................(3.5.15)}
\]

\[
D^2 = \delta^2 - \frac{1}{12} \delta^4 + \frac{1}{90} \delta^6 - \frac{1}{560} \delta^8 + \frac{1}{3150} \delta^{10} \quad \text{...................................(3.5.16)}
\]

3.6. LEAST-SQUARES POLYNOMIAL APPROXIMATION:

The Least-Squares Principle:

The basic idea of choosing a polynomial approximation \( p(x) \) to a given function \( y(x) \) in a way which minimizes the squares of the errors (in some sense) was developed by Gauss. There are several variations depending on the set of arguments involved and the error measure to be used.

First of all when the data are discrete one may minimize the sum

\[
S = \sum_{i=0}^{N} \left( y_i - a_0 - a_1 x_i - \ldots - a_m x_i^m \right)^2 \quad \text{...................................(3.6.1)}
\]

for the given data \( x_i, y_i \) and \( m < N \). The condition \( m < N \) makes it unlikely that the polynomial

\[
p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m \quad \text{...................................(3.6.2)}
\]

can collocate at all \( N \) data points. So \( S \) probably cannot be made zero. The idea of Gauss is to make as small as one can. Standard techniques of Calculus then lead to the normal equations which determine the coefficients \( a_i \). These equations are

\[ \]
\[
\begin{align*}
\begin{cases}
s_0a_0 + s_1a_1 + \ldots + s_na_n = t_0 \\
s_0a_1 + s_1a_1 + \ldots + s_na_n = t_1 \\
\vdots \\
s_0a_n + s_1a_1 + \ldots + s_na_n = t_n
\end{cases}
\end{align*}
\]

(3.6.3)

where \( s_i = \sum_{i=0}^{N} x_i^i \), \( t_i = \sum_{i=0}^{N} y_i x_i^i \). This system of linear equations does determine the \( a_i \) uniquely and the resulting \( a_i \) do actually produce the minimum possible value of \( S \). For the case of a linear polynomial

\[
p(x) = Mx + B
\]

(3.6.4)

The normal equations are easily solved and yield

\[
M = \frac{s_0x_1 - s_1x_1}{s_0x_2 - s_1x_1}
\]

(3.6.5)

\[
B = \frac{s_1x_0 - s_0x_1}{s_0x_2 - s_1x_1}
\]

(3.6.6)

In order to provide a unifying treatment of the various least-squares methods to be presented, including this first method just described a general problem of minimization in a vector space is considered. The solution is easily found by an algebraic argument using the idea of orthogonal projection. Naturally the general problem produces \( p(x) \) and normal equations. It will be reinterpreted to solve other variations of the least-squares principle as one proceed. In most cases a duplicate argument for the special case in hand will also be provided.

Except for very low degree polynomials the above system of normal equations proves to be ill-conditioned. This means that although it does define the coefficients \( a_i \), uniquely in practice it may prove to be impossible to extricate these
Standard methods for solving linear systems may either produce no solution at all or else badly magnify data errors. As a result orthogonal polynomials are introduced. (This amounts to choosing an orthogonal basis for the abstract vector space). For the case of discrete data these are polynomials \( P_{m,n}(t) \) of degree \( m=0,1,2,\ldots \) with property

\[
\sum_{i=0}^{N} P_{m,n}(t) P_{m,n}(t) = 0 \quad \text{..............................................(3.6.7)}
\]

This is the orthogonal property. The explicit representation

\[
P_{m,n}(t) = \sum_{i=0}^{m} (-1)^i \binom{m}{i} \binom{m+i}{i} t^i \binom{N}{i} \quad \text{..............................................(3.6.8)}
\]

will be obtained in which binomial coefficients and factorial polynomials are prominent.

An alternate form of our least squares polynomial now becomes convenient namely

\[
p(t) = \sum_{i=0}^{N} a_i P_{i,n}(t) \quad \text{..............................................(3.6.9)}
\]

with new coefficients \( a_i \). The equations determining these \( a_i \) prove to be extremely easy to solve. In fact

\[
a_i = \frac{\sum_{i=0}^{N} y_i P_{i,n}(t)}{\sum_{i=0}^{N} P_{i,n}(t)} \quad \text{..............................................(3.6.10)}
\]

These \( a_i \) do minimize the error sum \( S \), the minimum being

\[
S_{min} = \sum_{i=0}^{N} y_i^2 - \sum_{i=0}^{N} W_i a_i^2 \quad \text{..............................................(3.6.11)}
\]

where \( W_k \) is the denominator sum in the expression for \( a_k \).
Applications

There are two major applications of least-squares polynomials for discrete data.

1) Data Smoothing: By accepting the polynomial

\[ p(x) = a_0 + a_1 x + \ldots + a_n x^n \]  

in place of the given \( y(x) \), one can obtain a smooth line, parabola or other curve in place of the original, probably irregular data function. What degree \( p(x) \) should have depends on the circumstances. Frequently a five-point least-squares parabola is used corresponding to points \((x_i, y_i)\) with \(i = k-2, k-1, \ldots, k+2\). It leads to the smoothing formula

\[ y(x_i) = p(x_i) = y_i - \frac{3}{35} \delta^4 y_i \]  

This formula blends together the five values \( y_{k-2}, \ldots, y_{k+2} \) to provide a new estimate to the unknown exact value \( y(x_k) \). Near the ends of a finite data supply, minor modifications are required.

The root mean square error of a set of approximations \( A_i \) to corresponding true values \( T_i \) is defined as

\[ \text{RMS error} = \left[ \frac{1}{N} \sum_{i=0}^{N} (T_i - A_i)^2 \right]^{1/2} \]  

In various test cases where the \( T_i \) are known we shall use this measure to estimate the effectiveness of least-squares smoothing.
(ii) Approximate differentiation:

Fitting a collocation polynomial to irregular data leads to very poor estimates of derivatives. Even small errors in the data are magnified to troublesome size. But a least – squares polynomial does not collocate. It passes between the data values and provides smoothing. This smoother function usually brings better estimates of derivatives, namely, the values of \( p'(x) \). The five point parabola just mentioned leads to the formula

\[
y'(x_i) = p'(x_i) = \frac{1}{10}(-2y_{i-2} - y_{i-1} + y_{i+1} + 2y_{i+2}) \quad \text{..................(3.6.15)}
\]

Near the ends of a finite data supply this also requires modification. This formula usually produces results much superior to those obtained by differentiating collocation polynomials. However reapplying it to the \( p'(x_i) \) values in an effort to estimate \( y''(x_i) \) again leads to questionable accuracy.

Continuous Data:

For continuous data \( y(x) \) one may minimize the integral.

\[
I = \int [y(x) - a_0 P_0(x) - \ldots - a_n P_n(x)]^2 \, dx \quad \text{..................(3.6.16)}
\]

The \( P_j(x) \) being Legendre polynomials. This means that one can have chosen to represent our least – squares polynomial \( p(x) \) from the start in terms of orthogonal polynomials in the form

\[
p(x) = a_0 P_0(x) + \ldots + a_n P_n(x) \quad \text{..................(3.6.17)}
\]

The coefficients prove to be
\[ a_n = \frac{2k+1}{2} \int_{-1}^{1} y(x) P_k(x) \, dx \] \hspace{1cm} (3.6.18)

For convenience in using the Legendre polynomials the interval over which the data \( y(x) \) are given in first normalized to \((-1, 1)\). Occasionally it is more convenient to use the interval \((0,1)\). In this case the Legendre polynomials must also be subjected to a change of argument. The new polynomials are called shifted Legendre polynomials.

Some type of discretization is usually necessary when \( y(x) \) is of complicated structure. Either the integrals which give the coefficients must be computed by approximation methods or the continuous agreement set must be discretized at the outset and a sum minimized rather than an integral. Plainly there are several alternate approaches and the computer must decide which to use for a particular problem. Smoothing and approximate differentiation of the given continuous data function \( y(x) \) are again the foremost applications of least squares polynomial \( p(x) \). One can simply accept \( p(x) \) and \( p'(x) \) as substitutes for the more irregular \( y(x) \) and \( y'(x) \).

A generalization of the least - squares principle involves minimizing the integral

\[ I = \int w(x) \left[ y(x) - a_0 Q_0(x) - \ldots - a_n Q_n(x) \right]^2 \, dx \] \hspace{1cm} (3.6.19)

where \( w(x) \) is a negative weight function. The \( Q_k(x) \) are orthogonal polynomials in the generalized sense.
\[ \int_{x_j}^{x_k} w(x) \mathcal{Q}_j(x) \mathcal{Q}_k(x) dx = 0 \] .................................(3.6.20)

for \( j \neq k \). The details parallel those for the case \( w(x) = 1 \) already mentioned, the coefficients \( a_k \) being given by

\[ a_k = \frac{\int_{x_j}^{x_k} w(x) y(x) \mathcal{Q}_k(x) dx}{\int_{x_j}^{x_k} w(x) \mathcal{Q}_k^2(x) dx} \] .................................(3.6.21)

The minimum value of \( I \) can be expressed as

\[ I_{\min} = \int_{x_j}^{x_k} w(x) y^2(x) dx - \sum_{i=1}^{m} W_i a_i^2 \] .................................(3.6.22)

where \( W_k \) is the denominator integral in the expression for \( a_k \). This leads to the Bessel's inequality.

\[ \sum_{i=1}^{m} W_i a_i^2 \leq \int_{x_j}^{x_k} w(x) y^2(x) dx \] .................................(3.6.23)

and to the fact that for \( m \) tending to infinity the series \( \sum_{i=1}^{m} W_i a_i^2 \) is convergent. If the orthogonal family involved has a property known as completeness and if \( y(x) \) is sufficiently smooth then the series actually converges to the integral which appears \( I_{\min} \). This means that the error of approximation tends to zero as the degree of \( p(x) \) is increased.

**Chebyshev Polynomials:**

Approximation using Chebyshev polynomials is the important special case

\[ w(x) = \frac{1}{\sqrt{1-x^2}} \]

of the generalized least squares method, the interval of integration
being normalized to \((-1,1)\). In this case the orthogonal polynomials \(Q_n(x)\) are the Chebyshev polynomials

\[ T_n(x) = \cos(n \arccos x) \]

The first few prove to be

\[ T_0(x) = 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1, \quad T_3(x) = 4x^3 - 3x \]

Properties of the Chebyshev polynomials include

\[ T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x) \]

\[ \int_{-1}^{1} \frac{T_m(x)T_n(x)}{\sqrt{1-x^2}} \, dx = \begin{cases} 0 & \text{if } m \neq n \\ \frac{\pi}{2} & \text{if } m = n 
eq 0 \\ \pi & \text{if } m = n = 0 \end{cases} \]

\[ T_n(x) = 0 \quad \text{for } x = \cos \left( \frac{(2i+1)\pi}{2n} \right), \quad i = 0, 1, 2, \ldots, n - 1 \]

\[ T_n(x) = (-1)^i \quad \text{for } x = \cos \left( \frac{i\pi}{n} \right), \quad i = 0, 1, 2, \ldots, n \]

An especially attractive property is the equal-error property which refers to the oscillation of the Chebyshev polynomials between extreme values of \(\pm 1\) reaching these extremes at \(n+1\) arguments inside the interval \((-1,1)\). As a consequence of this property, the error \(y(x) - p(x)\) is frequently found to oscillate between maxima and minima of approximately \(\pm E\). Such an almost-equal-error is desirable since it implies that our approximation has almost uniform accuracy across the entire interval.
The powers of $x$ may be expressed in terms of Chebyshev polynomials by simple manipulations. For example

$$1 = T_0 \quad x = T_1 \quad x^2 = \frac{1}{2}(T_0 + T_2) \quad x^3 = \frac{1}{4}(3T_0 + T_3)$$

This has suggested a process known as economization of polynomials by which each power of $x$ in a polynomial is replaced by the corresponding combination of Chebyshev polynomials. It is often found that a number of the higher-degree Chebyshev polynomials may then be dropped, the terms retained then constituting a least-squares approximation to the original polynomial of sufficient accuracy for many purposes. The result obtained will have the almost-equal-errors property. This process of economization may be used as an approximate substitute for direct evaluation of the coefficient integrals of an approximation by Chebyshev polynomials. The unpleasant weight factor $w(x)$ makes these integrals formidable for most $y(x)$.

Another variation of the least-squares principle is to minimize the sum

$$\sum_{i=0}^{N-1} \left[ y(x_i) - a_0 T_0(x_i) - \cdots - a_n T_n(x_i) \right]^2$$

the arguments being $x_i = \cos \left( \frac{(2i+1)\pi}{2N} \right)$. These arguments may be recognized as the zeros of $T_N(x)$. The coefficients are easily determined using a second orthogonality property of the Chebyshev polynomials.
and prove to be

\[ a_n = \frac{1}{N} \sum_{i=0}^{N-1} y(x_i) \quad a_s = \frac{2}{N} \sum_{i=0}^{N-1} y(x_i)T_s(x_i) \]

The approximating polynomial is then of course

\[ p(x) = a_0 T_0(x) + \ldots + a_n T_n(x) \]

This polynomial also has an almost-equal error.

The L_2 Norm:

The underlying theme of the above discussion is to minimize the norm

\[ \| y - p \|_2 \]

where \( y \) represents the given data and \( p \) the approximating polynomial.

3.7 ROOTS OF EQUATIONS:

The problem treated here is the ancient problem of finding roots of equations or of systems of equations. The long list of available methods shows the long history of this problem and its continuing importance. Which method to use depends upon whether one needs all the roots of a particular equation or only a few, whether the roots are real or complex, simple or multiple, whether one has a ready first approximation or not and so on.

(i) Interpolation Methods:

These methods use two or more approximations usually some too small and some too large to obtain improved approximations to a root by use of collocation polynomials. The most ancient of these is based on linear interpolation between
two previous approximations. It is called regula falsi and solves \( f(x) = 0 \) by the iteration

\[
x_n = x_{n-1} - \frac{(x_{n-1} - x_{n-2})f(x_{n-1})}{f(x_{n-1}) - f(x_{n-2})}
\]  

\[(3.7.1)\]

The rate of convergence is between those of the previous two methods. A method based on quadratic interpolation between three successive approximations \( x_0, x_1, x_2 \) uses the formula

\[
x_1 = x_2 - \frac{2C}{B \pm \sqrt{B^2 - 4AC}}
\]  

\[(3.7.2)\]

(ii) Bernoulli's Method:

This method produces the dominant root of a real polynomial equation

\[
a_0x^n + a_1x^{n-1} + \ldots + a_n = 0
\]  

\[(3.7.3)\]

provided a single dominant root exists by computing a solution sequence of the difference equation

\[
a_0x_k + a_1x_{k-1} + \ldots + a_nx_{k-n} = 0
\]  

\[(3.7.4)\]

and taking \( \lim \left( \frac{x_{k+1}}{x_k} \right) \). The initial values \( x_{-n+1} = \ldots = x_{-1} = 0, x_0 = 1 \) are usually used. If a complex conjugate pair of roots is dominant then the solution sequence is still computed but the formulas

\[
r_1^n \approx \frac{x_k^2 - x_{k+1}x_{k-1}}{x_{k-1} - x_kx_{k-2}}
\]  

\[(3.7.5)\]

\[-2\text{Re}\phi = \frac{x_{k+1}x_{k-2} - x_{k+1}x_k}{x_{k-1} - x_kx_{k-2}}
\]  

\[(3.7.6)\]

serve to determine the roots as \( r_1, r_2 = r(\cos \phi \pm \text{isin}\phi) \).
Methods:

These are based upon the idea that the system $F=0$ or $f_i=0$ for $i=1,2...n$ is solved whenever the function

$$S = f_1^2 + f_2^2 + ... + f_n^2$$

is minimized since the minimum clearly occurs when all the $f_i$ are zero. Direct methods for seeking this minimum or descent methods have been developed. For example the two-dimensional problem (with a familiar change of notation)

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

is equivalent to minimizing this sum

$$S(x,y) = f^2 + g^2$$

Beginning at an initial approximation $(x_0, y_0)$ we select the next approximation in the form

$$x_1 = x_0 - tS_x, \quad y_1 = y_0 - tS_y,$$

where $S_x$ and $S_y$ are the components of the gradient vector of $S$ at $(x_0, y_0)$. Thus progress is in the direction of steepest descent and the algorithm is known as the steepest descent algorithm. The number $t$ may be chosen to minimize $S$ in this direction though alternatives have been proposed. Similar steps then follow. The method is often used to provide initial approximations to the Newton Method.

The above equivalence is of course often exploited in the opposite way. To optimize a function $f(x_1, x_2, ..., x_n)$ one looks for places where the gradient of $f$ is zero.
\[
\text{grad}(f) = (f_1, f_2, \ldots, f_n) = (0, 0, 0) \quad \text{...........................................(3.7.9)}
\]

Here, \( f \) denotes the partial derivative of \( f \) relative to \( x_i \). The optimization is then attempted through the solution of the system of \( n \) nonlinear equations.

(iv) Ballistow’s Method:

This method produces complex roots of a real polynomial equation \( p(x) = 0 \) by applying the Newton method to a related system. More specifically, division of \( p(x) \) by a quadratic polynomial suggests the identity

\[
p(x) = (x^2 - ux - v)q(x) + r(x) \quad \text{...........................................(3.7.10)}
\]

where \( r(x) \) is a linear remainder

\[
r(x) = b_{a-1}(u, v)(x-u) + b_a(u, v) \quad \text{...........................................(3.7.11)}
\]

The quadratic divisor will be a factor of \( p(x) \) if one can choose \( u \) and \( v \) so that

\[
b_{a-1}(u, v) = 0 \quad b_a(u, v) = 0
\]

This is the system to which Newton’s method is now applied. Once \( u \) and \( v \) are known, a complex pair of roots may be found by solving

\[
x^2 - ux - v = 0 \quad \text{...........................................(3.7.12)}
\]