Optimization

Unit-5

Optimization :- Optimization Problem requires us to determine maximum and Minimum value of a function.

There are Two types of Optimization

(1) Constrained optimization

(2) UnConstrained optimization

(1) **UnConstrained Optimization** :-

There are two types of Problems

They are (1) Profit maximization

(2) Cost minimization

Practical Computational task of finding maxima (or) minima of a function of many variables

Method :-

Step (1) Find the derivative of a function with respect to x and y then put it equals to Zero to find the values of x and y Points which is Called as stationary points.

$$\frac{df}{dx} = 0 \longrightarrow (1)$$

$$\frac{df}{dy} = 0 \longrightarrow (2)$$

By solving (1) & (2) we have Points (x, y) $\longrightarrow$ stationary points.

Step-2 :- Second step is to find Second Order derivative of a function with respect to x and y and xy, then solve the equation $AC - B^2$.

w.r.to $x \longrightarrow A$ put
w.r.to $y \longrightarrow C$ put
w.r.to.to $xy \longrightarrow B$ put

Step-3 :- If $AC - B^2 > 0$, then its the Case of extreme Points

$$\frac{\partial^2 f}{\partial x^2} = A$$

$$\frac{\partial^2 f}{\partial y^2} = C$$

$$\frac{\partial^2 f}{\partial x \partial y} = B$$

(1) $AC - B^2 > 0$ ✓ (extreme points)

(2) $AC - B^2 < 0$
(3) $AC - B^2 = 0$
No extreme ✗ Points.

Step-4 :- we Can find maxima (or) minima, according to the Value of A.

Step-5 :- Last step is to find the Maximum (or) Minimum value of that function through the extreme Value.

Problem.

(1) Find the Extreme Value of
$f(x) = x^3 + y^3 - 6xy$ and determine whether they are Maximum (or) Minimum.

Sol:- given $f(x) = x^3 + y^3 - 6xy$

$$\frac{\partial f}{\partial x} = 3x^2 + 0 - 6y = 0$$

$$= \cancel{3(x^2)} = 3x^2 - 6y = 0$$

$$= 3(x^2 - 2y) = 0$$

$$x^2 - 2y = 0$$

$$-2y = -x^2$$

$$\Rightarrow \boxed{y = \frac{x^2}{2}} \longrightarrow (1).$$

$$\frac{\partial f}{\partial y} = 3y^2 - 6x = 0$$

$$= 3(y^2 - 2x) = 0$$

$$= y^2 - \cancel{2x} = 0 \longrightarrow (2)$$

By Solving (1) & (2)

Putting y in (2)

$$y^2 - 2x = 0$$

$$\left(\frac{x^2}{2}\right)^2 - 2x = 0$$

$$\frac{x^4}{4} - 2x = 0$$

$$= \frac{x^4 - 8x}{4} = 0$$

$$x^4 - 8x = 0$$

$$x(x^3 - 8) = 0$$

$\boxed{x=0}$ $\quad x^3 - 8 = 0$

$$x^3 = 8$$

$$x = \sqrt[3]{8} = 2^3$$

$\boxed{x=2}$ $\quad x^3 = 2^3$

$\boxed{x=2}$

$\therefore$ Putting x in ①

$$y = \frac{x^2}{2}$$

when $\boxed{x=0} \Rightarrow y = \frac{0}{2} = 0$

when $\boxed{x=2} \Rightarrow y = \frac{2^2}{2} = \frac{4}{2} = 2$

$\therefore$ Stationary Points are $(0,0)$ $(2,2)$.

Second Order Condition :

$$\frac{\partial f}{\partial x} = 3x^2 - 6y \qquad \frac{\partial f}{\partial y} = 3y^2 - 6x$$

$$A = \frac{\partial^2 f}{\partial x^2} = 6x$$

$$B = \frac{\partial^2 f}{\partial x \partial y} = -6$$

$$C = \frac{\partial^2 f}{\partial y^2} = 6y$$

$A = 6x \quad , \quad B = -6, \quad C = 6y$

a) At Point $(\underset{x}{0}, \underset{y}{0})$:-

$A = 6(0) = 0 \Rightarrow \boxed{A = 0}$

$B = -6 \Rightarrow \boxed{B = -6}$

$C = 6(0) = 0 \Rightarrow \boxed{C = 0}$

$AC - B^2$

$\Rightarrow (0)(0) - (-6)^2 = -(36) = -36 < 0$

$\therefore$ No Extreme Point

b) At Point $(\underset{x}{2}, \underset{y}{2})$:-

$A = 6(2) = 12 \Rightarrow \boxed{A = 12}$

$B = -6 \Rightarrow \boxed{B = -6}$

$C = 6(2) = 12 \Rightarrow \boxed{C = 12}$

$AC - B^2$

$\Rightarrow (12)(12) - (-6)^2 = 144 - 36 = 108 > 0$

It is an Extreme Point.

$A = 12 > 0$, Positive minima Point:

Extreme Value at $f(\underset{x}{2}, \underset{y}{2})$

$f(x, y) = x^3 + y^3 - 6xy$

$= 2^3 + 2^3 - 6(2)(2)$

$= 8 + 8 - 6(4) = 16 - 24 = -8$

So, -8 is Minimum value at $x = 2$ & $y = 2$

So, function is Minimum

Constrained Optimization

→ The general Constrained Optimization task is to Maximize (or) Minimize a function $f(x)$ by varying x, given certain constraints on x.

→ for example :-

Find Minimum of

$$f(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + x_3^2$$

where $\|x\|_2 \geq 1$

→ Very common to encounter this in engineering practice

for example :- designing the fastest vehicle with a constraint on fuel efficiency.

→ All constraints can be converted to two types of constraints

→ Equality Constraints :-

Ex:- Minimize $f(x_1, x_2, x_3)$

Subject to $x_1 + x_2 + x_3 = 1$ ⟶ (1)

→ Inequality Constraints :-

Ex:- Minimize $f(x_1, x_2, x_3)$

Subject to $x_1 + x_2 + x_3 < 1$ ⟶ (2)

→ **Canonical form** :

All optimization Problems Can be written as

Minimize $f(x)$ Subject to Constraint that $x \in S$ → feasible points

$$S = \{ x / \underbrace{\forall_i}_{\text{multiple}} , \underbrace{g^{(i)}(x) = 0}_{\text{equality Constraints}} \text{ and } \underbrace{\forall_j}_{\text{multiple}}, \underbrace{h^{(j)}(x) \le 0}_{\text{Inequality Constraints}} \}$$

from eqn (1)

⟹ Equality Constraints Can be written as

equality Constraint

$$g(x) = x_1 + x_2 + x_3 - 1 = 0$$

from eqn (2) Inequality Constraints Can be written as

$$h(x) = x_1 + x_2 + x_3 - 1 \le 0$$

feasibility set is a Combination of equality Constraint & Inequality Constraints.

- **Generalized Lagrange function :-**

→ The Constrained Optimization Problem requires us to minimize the function $f(x)$ while ensuring that the Point discovered belongs to the feasible set.

→ There are several techniques that acheive this but it is, in general, a difficult Problem.

→ A very Common approach is to define a new function called the generalized Lagrangian.

$$\underbrace{L}_{\text{Lagrangian}}(x, \underbrace{\lambda, \alpha}_{\substack{\text{two vectors} \\ (\text{variables } \lambda, \alpha)}}) = \underbrace{f(x)}_{\text{original function}} + \sum_i \lambda_i g^{i}(x) + \sum_j \alpha_j h^{(j)}(x)$$

→ Then the Constrained minimum is given by

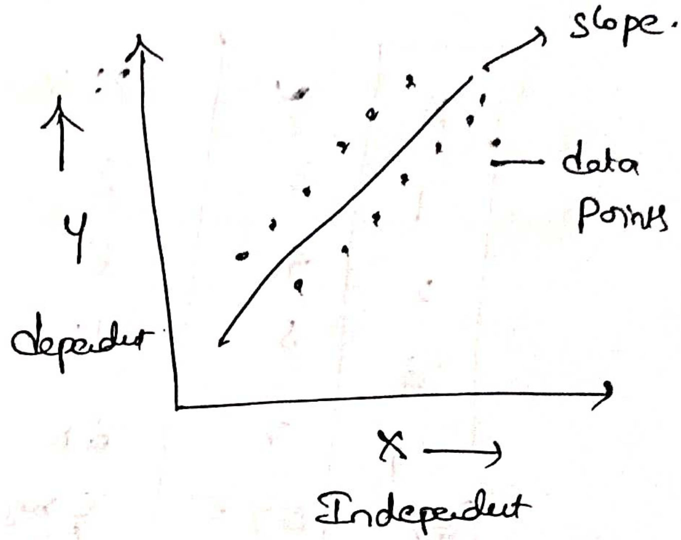
$$\min_{x \in S} f(x) = \min_{x} \max_{\lambda} \max_{\alpha, \alpha \geq 0} L(x, \lambda, \alpha)$$

→ we will the Proof and details of this when we come to later weeks.

Linear Regression

→ Dependent Variable is Continuous in nature.

Simple Linear Regression



$$\boxed{y = \alpha_0 + \alpha_1 x_1}$$

$$y = c + mx$$

x_1 is '1' Independent Variable

y is dependent Variable

α Values are Co-efficients of Regression

Linear in nature.

This Can be easily written as

$$\boxed{y = c + mx}$$

x is independent Variable

m is a slope.

c is Intercept

y is dependent Variable.

multiple linear regression

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m$$

α_i = Regression Co-eff

x_i = Independent Variables

y = dependent Variables

$\alpha_3 \geq \alpha_4$

ex: $y = 0.9 + 1.2x_1 + 2x_2 + 4x_3 + 1x_4$

α_3 is more

x_3 is more

(1) Given $x = 1, 2, 3, 4$.
$y = 3, 4, 5, 7$

$$\boxed{y = bx + a}$$

x	y	xy	x^2
1	3	3	1
2	4	8	4
3	5	15	9
4	7	28	16
$\Sigma x = 10$	$\Sigma y = 19$	Σxy 54	$\Sigma x^2 = 30$.

$$\boxed{a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}}$$

$$a = \frac{(19)(30) - (10)(54)}{(4)(30) - (100)}$$

a & b are Unknown.

$$= \frac{570 - 540}{120 - 100} = \frac{30}{20} = \frac{3}{2} = \underline{1.5}$$

$$\therefore \boxed{a = 1.5}$$

$$\therefore \boxed{a = 1.5}$$
$$\boxed{b = 1.3}$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$\therefore \boxed{y = 1.3x + 1.5}$$

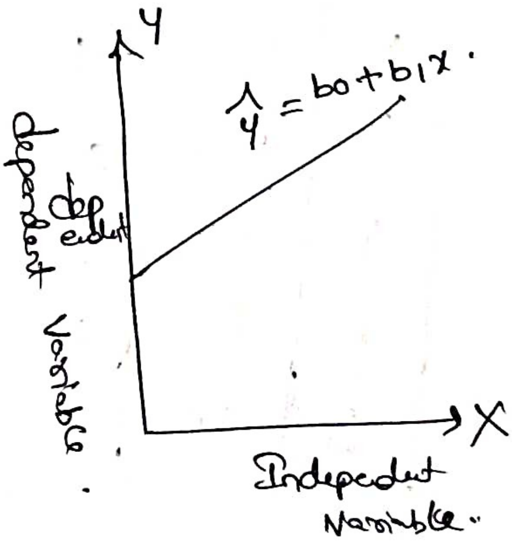
$$b = \frac{(4)(54) - (10)(19)}{(4)(30) - 100} = \frac{216 - 190}{120 - 100} = \frac{26}{20} = \frac{13}{10}$$
$$= \underline{1.3}$$

Linear Regression Analysis

→ This analysis is Used in Understanding the relationship between two (or) more Variables (Multiple regression)

→ when in case of understanding two variables One is Independent Variable (Input) and the Other variable is dependent variable (Predicted variable)

$$\boxed{\hat{y} = b_0 + b_1 x}$$



→ In this analysis we try to find a straight line which fits most of the Observations

b_0 – y-Intercept
b_1 – slope

→ This line is Called as Linear Regression Line and it is Obtained by Least Square method

→ when $x \uparrow y \uparrow$ Slope is +ve $(\hat{y} = b_0 + b_1 x)$
→ when $x \uparrow y \downarrow$ slope is −ve $(\hat{y} = b_0 - b_1 x)$

Problem

x	1	2	3	4	5
y	2	4	5	4	5

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$	$\hat{y}$	$\hat{y}-\bar{y}$	$(\hat{y}-\bar{y})^2$
1	2	-2	-2	4	4	2.8	0.8	0.64
2	4	-1	0	1	0	3.4	0.6	0.36
3	5	0	1	0	0	4	-1	1
4	4	1	0	1	0	4.6	0.6	0.36
5	5	2	1	4	2	5.2	0.2	0.04
15	20			10	6			2.4

$\bar{x} = \frac{15}{5} = \underline{\underline{3}}$ $\quad \therefore \bar{x} = 3$

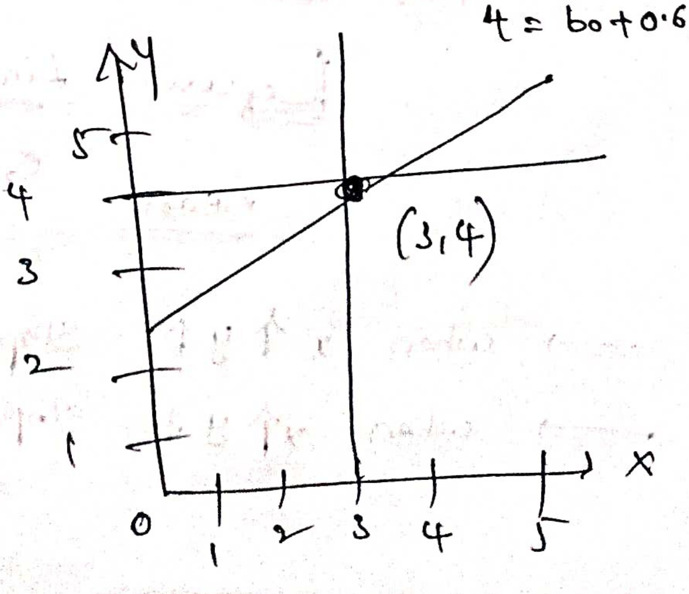
$\bar{y} = \frac{20}{5} = \underline{\underline{4}}$ $\quad \bar{y} = 4$

$\hat{y} = b_0 + b_1 x$

$4 = b_0 + 0.6 \cdot 3$

$$b_1 = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

$$= \frac{6}{10} = \underline{\underline{0.6}}$$



$$\therefore b_1 = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2} = \frac{6}{10} = \underline{0.6}$$

b_0 is calculated using the mean coordinate

(3, 4)

$$y = b_0 + b_1 x$$

$$4 = b_0 + (0.6)3$$

$$b_0 = 4 - (0.6)3$$

$$\boxed{b_0 = 2.2}$$

$$\therefore \boxed{\hat{y} = 2.2 + 0.6x}$$ (Regression Line)

Standard error

$$= \sqrt{\frac{\Sigma(\hat{y}-\bar{y})^2}{n-2}} = \sqrt{\frac{2.4}{5-2}} = \underline{0.89}$$

Calculates the difference between actual + estimated.

Assumptions of Linear Regression :-

1) Linear Relation

2) very low / No multi Collinearity

3) Heterostochastycity

4) No Auto Correlation of errors.

5) Normal distribution of errors.

6) All the Observations are Independent to each other

Linear Regression

1) ~~A~~ Simple example of a Regression equation to Predict the Glucose level given the age

Subject	Age(X)	Glucose level (Y)
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81
7	55	?

X is Independent
Y is dependent
(Here 6 data points are given & 7th one is asked)

Sol:

The Simple Linear Regression Equation Provides an Estimate of the Population Regression Line.

$$\boxed{\hat{Y}_i = b_0 + b_1 X_i}$$

X_i: Value of X for Observation i

$\hat{Y}_i$: Estimated (or Predicted) Y value for ~~Obse~~ Observation i

b_0: Estimate of the Regression Intercept

b_1: Estimate of the Regression slope.

Y is dependent
X is Independent.

A simple example of a Regression equation to Predict the glucose level given the age.

$$b_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Subject	Age (x)	Glucose level (y)	xy	x^2	y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
	$\Sigma x = 247$	$\Sigma y = 486$	$\Sigma xy = 20485$	$\Sigma x^2 = 11409$	$\Sigma y^2 = 40022$

$$\boxed{b_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}}$$

$$b_0 = \frac{(486)(11409) - (247)(20485)}{6(11409) - (247)^2}$$

$$b_0 = \frac{4848979}{7445} = \underline{\underline{65.14}}$$

$$\therefore \boxed{b_0 = 65.14}$$

$$\boxed{b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}}$$

$$b_1 = \frac{6(20485) - (247)(486)}{6(11409) - (247)^2}$$

$$b_1 = \frac{2868}{7445} = \underline{\underline{0.385335}}$$

$$\therefore \boxed{b_1 = 0.385335}$$

$$\boxed{\hat{y} = b_0 + b_1 x}$$

$$\therefore \boxed{\hat{y} = b_0 + b_1 x.}$$

$$\boxed{\hat{y} = 65.14 + 0.385225x}$$

The value of Y for given value of X = 55

$$\hat{y} = 65.14 + (0.385225)(55)$$

$$\boxed{\hat{y} = 86.327}$$

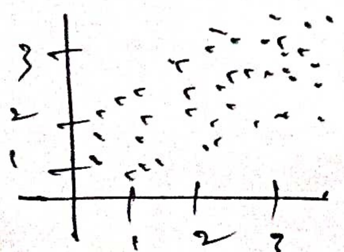
Hence the glucose level for the given age 55 is <u>86.327</u>

→ what is the Problem of Heteroscedasticity?

<u>Sol</u>: The Problem of Heteroscedasticity refers to a situation when the residuals in a Regression <u>do not have Uniform Variance</u>.

→ Arises when variation is uneven across Observations

→ Tends to give inefficient regression results



— Incorrect standared error.

→ Heteroscedasticity.

LN4

Unit-5

→ In this Chapter we shall Concern ourselves with the Classical theory of Optimization.

→ This theory deals with the use of differential Calculus to determine the Points of Maxima & Minima for both UnConstrained and Constrained Continuous functions.

→ In this Chapter the topics include the development of necessary and sufficient Conditions for locating extreme points for UnConstrained Problems

→ The treatement of the Constrained Problems Using the Lagrangian methods and the development of the Kuhn-Tucker Conditions for the general Problem with inequality Constraints.

Un-Constrained Problems of Maxima & Minima.

→ Here we shall discuss the Problem of determining the extreme points (the Points of Maxima & Minima) of an Unconstrained type of Continious functions.

Mathematically

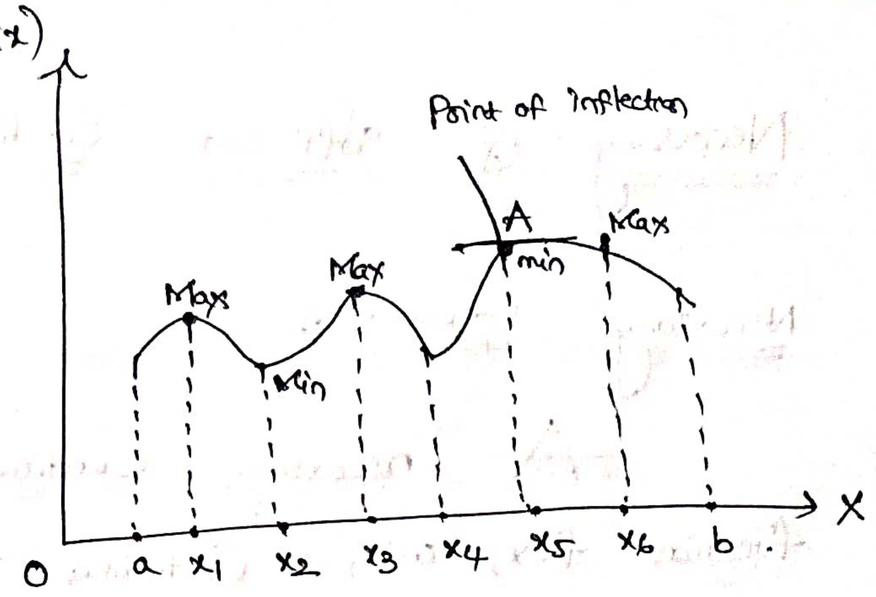
→ A function $f(x)$ has a Maximum at a Point 'x_0' If

$$\boxed{f(x_0+h) - f(x_0) < 0}$$

→ IIIy a function $f(x)$ has a Minimum at a Point 'x_0' If

$$\boxed{f(x_0+h) - f(x_0) > 0}$$

Consider a Continuous function $f(x)$ defined on in interval (a, b)



Here the points x_1, x_2, x_3, x_4 & x_6 (not x_5) represent all the Points of Maxima & Minima.

(Called the stationary (or) Critical points) of $f(x)$

These includes $x_1, x_3,$ & x_6 as Points of Maxima & x_2 & x_4 as Points of Minima.

Global (absolute) Maximum :-

since $f(x_6) = \max\{f(x_1), f(x_3), f(x_6)\}$,

$f(x_6)$ is Called a global (or) absolute maximum

Local (relative) maxima :-

On the Other hand $f(x_1)$ & $f(x_3)$ are Called local (or) relative maxima.

||ly $f(x_4)$ is a local Minimum while $f(x_2)$ is a global minimum.

→ It should be noted that the point 'A' Corresponding to $f(x_5)$ is Called Point of Inflection.

Necessary & Sufficient Conditions for Optima.

Necessary Condition :-

A necessary Condition for a Continious function $f(x)$ with Continious first and second Partial derivatives to have an extreme Point at 'x_0' is that each first Partial derivative of $f(x)$ evaluate at x_0, Vanish that is

$$\boxed{\nabla f(x_0) = 0}$$

where $\nabla \equiv \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3} \cdots \frac{\partial f}{\partial x_n}\right)$ is the gradient vector.

Sufficient Condition :-

A sufficient Condition for a stationary Point 'x_0' to be an extreme Point is that the Hession matrix H evaluated at 'x_0' is

(1) Negative - definite when 'x_0' is a Maximum Point &

(2) Positive - definite when 'x_0' is a Minimum Point.

Proof for Conditions :-

1) **Proof for necessary Condition** :-

By Taylors theorem, for $0 < \theta < 1$

$$f(x_0+h) - f(x_0) = \nabla f(x_0)h + \frac{1}{2} h' H h \Big|_{x_0+\theta h} \longrightarrow (1)$$

where $h = (h_1 h_2 \cdots h_j \cdots h_n)'$ &

$|h_j|$ is small enough $\forall\ j = 1, 2 \cdots n$.

for small $|h_j|$ the remainder term $\frac{1}{2}(h' H h)$ is of Order h_j^2 & hence it will tend to Zero as $h_j \to 0$

$$f(x_0+h) - f(x_0) = \nabla f(x_0)h + O(h_j^2) \longrightarrow (2)$$

$$\nabla f(x_0)h \equiv \left[h_1 \frac{\partial f(x)}{\partial x_1} + h_2 \frac{\partial f(x)}{\partial x_2} + \cdots + h_p \frac{\partial f(x)}{\partial x_p} + \cdots + h_n \frac{\partial f(x)}{\partial x_n} \right]_{x = x_0}$$

Suppose that x_0 is an extreme point, now we shall Prove the theorem by Contradiction.

If possible, let us suppose that one of the partial derivatives, say pth, does not vanish,

ie:- $\dfrac{\partial f(x_0)}{\partial x_p} \neq 0$

then eqn (2) becomes $f(x_0+h) - f(x_0) = h_p \dfrac{\partial f(x_0)}{\partial x_p} \longrightarrow (3)$

Since $\frac{\partial f(x_0)}{\partial x_p} \neq 0$, either

$$\frac{\partial f(x_0)}{\partial x_p} < 0 \quad (\text{or}) \quad \frac{\partial f(x_0)}{\partial x_p} > 0$$

Now suppose $\frac{\partial f(x_0)}{\partial x_p} > 0$ then $f(x_0+h) - f(x_0)$ will have the same sign as hp

ie:- (i) $f(x_0+h) - f(x_0) > 0$ when $hp > 0$ &

(ii) $f(x_0+h) - f(x_0) < 0$ when $hp < 0$.

This contradicts the assumption that 'x_0' is an extreme point

The argument when $\frac{\partial f(x_0)}{\partial x_p} < 0$ is similar to the given above.

Thus we may conclude that when any of the partial derivatives are not identically equal to zero at 'x_0', the point 'x_0' is not an extreme point

Thus, it follows that for 'x_0' to be an extreme point it is necessary that

$$\boxed{\nabla f(x_0) = 0}$$

This completes the proof of the theorem.

2) Proof for Sufficient Condition :-

Proof :- By Taylor's theorem for $0 < \theta < 1$

we have

$$f(x_0+h) - f(x_0) = \nabla f(x_0) h + \frac{1}{2} h' H h \Big|_{x_0+\theta h}.$$

Since 'x_0' is a stationary Point, then by Preceding theorem (necessary Condition theorem) we have

$$\boxed{\nabla f(x_0) = 0}$$

Thus $f(x_0+h) - f(x_0) = \frac{1}{2} h' H h \Big|_{x_0+\theta h}$

Let x_0 be a Maximum Point then by definition

$$f(x_0+h) < f(x_0)$$

for all non-null h.

This implies that for x_0 ~~be~~ to be a Maximum.

$$\frac{1}{2} h' H h \Big|_{x_0+\theta h} < 0 \quad \text{(or)}$$

$$h' H h \Big|_{x_0+\theta h} < 0 \longrightarrow (1)$$

writing the quadratic form $h' H h$ in expanded form we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \Bigg|_{x = x_0+\theta h} < 0$$

However, since the second Partial derivative

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \text{ is Continuous in the neighbourhood of } x_0$$

$$\left.\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right|_{x=x_0} \text{ will have the Same sign as }$$

$$\left|\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right|_{x=x_0+\theta h}.$$

Consequently $h'Hh$ must yield the Same sign when evaluated at both x_0 & $x_0+\theta h$.

Thus from eqn (1) we have

$$h'Hh\Big|_{x=x_0} < 0$$

Since $h'Hh\Big|_{x=x_0}$ defines a quadratic form, this expression (and hence $h'Hh\Big|_{x=x_0+\theta h}$) is negative $\Longleftrightarrow$ the Hessian matrix H is negative-definite at x_0

This Completes the Proof for maximization Case.

A similar Proof Can be established for minimization Case to show the Corresponding Hessian matrix H is Positive definite at x_0

Problem

(1) Find the Maximum (or) Minimum of the function

$$f(x) = x_1^2 + x_2^2 + x_3^2 - 4x_1 - 8x_2 - 12x_3 + 56$$

Sol:- Applying the necessary Condition

$$\nabla f(x_0) = 0 \quad \text{(or)}$$

$$\left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right) f(x) = (0, 0, 0)$$

this gives

$$\frac{\partial f}{\partial x_1} = 2x_1 - 4 = 0$$

$$\frac{\partial f}{\partial x_2} = 2x_2 - 8 = 0$$

$$\frac{\partial f}{\partial x_3} = 2x_3 - 12 = 0$$

The Solution of these simultaneous equations is given by $x_0 = (2, 4, 6)$ which is the only point that satisfies the necessary Condition.

Now by checking the sufficient Condition, we must determine whether this point is a Maximum or minimum

The Hessian matrix, evaluated at $(2, 4, 6)$

is given by

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$\therefore$ The Principal minors determinants of H

$$|2|, \quad \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} \quad \& \quad \begin{vmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{vmatrix}$$

have the values $2, 4$ & 8 respectively

Thus each of the Principal minors determinant is

Positive

Hence H is Positive-definite

$\therefore$ the Point $(2, 4, 6)$ yields a ~~maximum~~ minimum

of $f(x)$

Optimization View of Machine Learning

i) why do we need Optimization for Machine learning?

A) Optimization is an Advanced topic, we can build many models, experiments etc

a) what do you learn in optimization?

1) **Calculus**:- Understanding different types of functions how to find maxima & Minima of functions

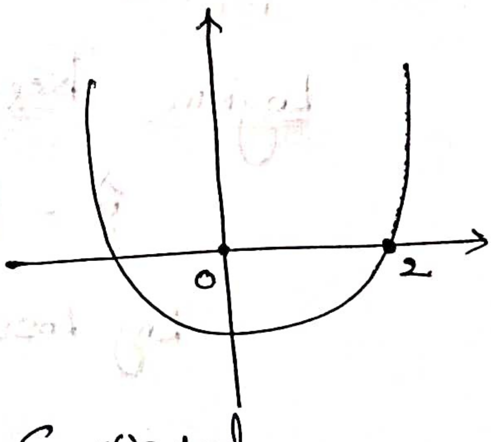
ex:- $y = f(x) = x^2 - 2x$

minima:- $\frac{dy}{dx} = f'(x) = 0$

$= 2x - 2 = 0$

$\Rightarrow 2x = 2$

$\Rightarrow \boxed{x = 1}$



→ functions Could get more Complicated (with "loss functions" of ML Models)

usually no Closed form solutions.

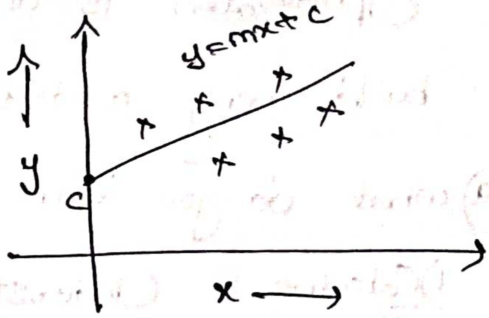
Optimization: Can we Come up with algorithms to find maxima / minima of these functions? in an efficient and effective way.

(1) **Understanding Loss functions :-**

most ML algorithms are driven by

→ writing an Objective function / loss function

→ finding the best Parameters that minimize the loss .



Ex :- Linear Regression :-

$$\boxed{y = mx + c}$$

straight line

$$\sum_{i=1}^{n} [y_i - (mx_i + c)]^2$$

Logistic Regression :-

$$\hat{y} = p(y=1) = \sigma(\omega^T x) = \frac{1}{1+e^{-\omega^T x}}$$

$$\text{Log Loss} = \sum_{i=1}^{N} - y_i \log(\hat{y}_i) - (1-y_i) \log(1-\hat{y}_i)$$

(2) **Understanding what Solver to use :-**

→ Do you want to use

RMSProp vs ADAM vs Momentum ?

→ Do you want to use

Batch Gradient Descent vs Stochastic Gradient Descent vs Mini-Batch Gradient Descent ?

(3) **Writing Loss functions :-**

You have a new Problem that does not fit in the usual setting

Easier in the Context of deep Learning :-

You can write a Custom loss function, leave the Solving to underlying platform.

Optimization : Learn to write Loss functions.

→ Optimization gives the understanding of how you can write Loss function.

4) **Writing updates / solves for an Optimization Problem :-**

Understand writing the updates for different Kinds of solvers

Gradient Descent :

$$W_j^{new} = W_j - \alpha \frac{\partial}{\partial w_j} \text{loss}(w)$$

Summarize :- **Optimization for Machine Learning**

Why Optimization :-

1) Understanding Loss functions
2) Picking the right Solver
3) writing new Loss functions
4) Implementing solves for Custom loss function.

Non-Linear Programming

Un Constrained Optimization Techniques :-

(i) Direct Search methods

(ii) Descent methods (or) Gradient methods.

~~(i)~~ steepest descent (Cauchy) method

(ii) Newton's method

(iii) Fletcher Reeves method

(iv) Marquardt methods

(v) Quasi - Newton Methods.

(i) Steepest descent (Cauchy) method :-

Procedure :-

(1) Start with the arbitrary initial point X_1 set the iteration number $i=1$

(2) Find the search direction S_i as

$$S_i = -\nabla f = -\nabla f(X_i)$$

(3) Find the optimal step length λ_i^* in the direction S_i set

$$X_{i+1} = X_i + \lambda^* S_i = X_i - \lambda_i^* \nabla f_i$$

(4) Test X_{i+1} for Optimality.

If X_{i+1} is Optimum, stop, Otherwise go to step-5

(5) Set the new iteration number $i=i+1$ and go to step-2

This method looks to be a very effective Unconstrained Optimization technique.

But since steepest descent direction is a local property, the method is not very effective in most of the problems.

Problems

(1) Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ and starting from the point $X_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ using descent method.

Sol :- Iteration-1 :-

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{bmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{bmatrix}$$

Diff Partial w.r.to x_1 & x_2

$$\nabla f_1 = \nabla f(x_1) = \begin{bmatrix} +1 \\ -1 \end{bmatrix} \longrightarrow \begin{pmatrix} x_1 = 0 \\ x_2 = 0 \end{pmatrix}$$

Substitute.

$$\therefore S_1 = -\nabla f_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

For X_2, we need λ^* (the optimal step length)

so we minimize

$$f(X_1 + \lambda_1 S_1) = f(-\lambda_1, \lambda_1) = \lambda_1^2 - 2\lambda_1$$

w.r.t from λ_1

since $\frac{\partial f}{\partial \lambda_1} = 0$ gives $\lambda_1^* = 1$,

we get $X_2 = X_1 + \lambda_1^* S_1$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} + 1\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\nabla f_2 = \nabla f(x_2) = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$\implies X_2$ <u>is not</u> Optimal.

<u>Iteration-2</u> :-

$$S_2 = -\nabla f_2 = (1, 1)$$

$$f(x_2 + \lambda_2 S_2) = f(-1 + \lambda_2,\ 1 + \lambda_2)$$

$$= 5\lambda_2^2 - 2\lambda_2 - 1$$

$$\frac{\partial f}{\partial \lambda_2} = 0 \implies \lambda_2^* = \frac{1}{3}$$

$$\implies X_3 = X_2 + \lambda_2^* S_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \frac{1}{5}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.8 \\ 1.2 \end{pmatrix}$$

$$\nabla f_3 = \nabla f(x_3) = \begin{pmatrix} 0.2 \\ -0.2 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$\implies X_3$ <u>is not</u> Optimal

<u>Iteration -3</u> :-

$$S_3 = -\nabla f_3 = \begin{pmatrix} -0.2 \\ 0.2 \end{pmatrix}$$

$$f(x_3 + \lambda_3 S_3) = f(-0.8 - 0.2\lambda_3,\ 1.2 + 0.2\lambda_3)$$

$$= 0.04\lambda_3^2 - 0.08\lambda_3 - 1.2$$

$$\frac{\partial f}{\partial \lambda_3} = 0 \implies \lambda_3^* = 1.0$$

$$\implies X_4 = X_3 + \lambda_3^* S_3 = \begin{pmatrix} -1.0 \\ 1.4 \end{pmatrix}$$

$$\nabla f_4 = \nabla f(X_4) = \begin{pmatrix} -0.2 \\ -0.2 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\implies X_4 \text{ is not Optimal}$$

Iteration - 4

$$S_4 = -\nabla f_4 = \begin{pmatrix} 0.2 \\ 0.2 \end{pmatrix}$$

$$f(x_4 + \lambda_4 S_4) = f(-1 + 0.2\lambda_4,\ 1.4 + 0.2\lambda_4)$$

and so on.

we Continue the Process Until the Optimum Point

$$x^* = \begin{pmatrix} -1.0 \\ 1.5 \end{pmatrix} \text{ is found}$$

(2) Newton's Method :-

$$f(x) = f(x_i) + \nabla f_i^T (x - x_i) + \frac{1}{2}(x - x_i)^T [J_i] (x - x_i) \longrightarrow (1)$$

where $[J_i] = (J)\big|_{x_i}$ is the Hessian matrix (matrix of second order Partial derivatives) of f evaluated at the Point x_i

for the Minimum of $f(x)$,

evaluate the Partial derivatives of $f(x) = 0$

$f(x)$ to Zero

ie:- $\frac{\partial f(x)}{\partial x_j} = 0$, $j = 1, 2, 3 \ldots n$ $\longrightarrow (2)$

from eqns (1) & (2) we get

$$\nabla f = \nabla f_i + [J_i][x - x_i] = 0 \longrightarrow (3)$$

If $[J_i]$ is non-singular the above equation (3) can be used to improve the approximation of x as x_{i+1} given by

$$x_{i+1} = x_i - [J_i]^{-1} \nabla f_i \longrightarrow (4)$$

Problem.

1) Minimize

$$f(x_1,x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$$ by taking the starting Point $X_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ Use Newtons Method.

Sol :-

$$J_1 = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}_{X_1} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

$$\therefore J_1^{-1} = \frac{1}{4}\begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}$$

$$\nabla f_1 = g_1 = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1+4x_1+2x_2 \\ -1+2x_1+3x_2 \end{bmatrix}_{(0\ 0)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

by eqn (4) $(x_{i+1} = x_i - [J_i]^{-1} \nabla f_i)$

$$x_2 = x_1 - J^{-1} \nabla f_1$$

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

$$= \begin{bmatrix} -1 \\ 3/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1.5 \end{bmatrix}$$

$$\nabla f_2 = g_2 = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}_{X_2} = \begin{bmatrix} 1+4x_1+2x_2 \\ -1+2x_1+2x_2 \end{bmatrix}_{X_2} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Since $\nabla f_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow X_2$ is Optimal Point

Constrained Optimization :-

The general Constrained Optimization task is to maximize (or) Minimize a function $f(x)$ by varying x given Certain Constraint on X.

→ for example,

find minimum of $f(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + x_3^2$
where $\|x\|_2 \geq 1$

→ Very Common to encounter this in engineering Practice.

→ For example, designing the fastest vehicle with a Constraint on fuel efficiency.

→ All Constraints Can be Converted to two types of Constraints

1) Equality Constraints
2) Inequality Constraints.

1) **Equality Constraints** :-

Ex :- Minimize $f(x_1, x_2, x_3)$

Subject to $x_1 + x_2 + x_3 = 1 \Rightarrow x_1 + x_2 + x_3 - 1 = 0$

$\Rightarrow g(x) = x_1 + x_2 + x_3 - 1 = 0$.

2) **Inequality Constraints** :-

Minimize $f(x_1, x_2, x_3)$

Subject to $x_1 + x_2 + x_3 < 1 \Rightarrow h(x) = x_1 + x_2 + x_3 - 1 < 0$

Canonical form :-

All optimization Problems Can be written as

$$S = \{ x / \forall_i ,\ g^{(i)}(x) = 0 \ \& \ \forall_j \ h^{(j)}(x) < 0 \}$$

minimize $f(x)$ subject to the Constraint that $x \in S$ is the feasible Point.

Generalized Lagrange function :-

→ The Constrained Optimization Problem requires us to minimize the function $f(x)$, while ensuring that the Point discovered belongs to the feasible state

→ There are Several techniques that achieve this but it is in general, a difficult Problem

→ A very Common approach is to define a new function Called the generalized Lagrangian

$$L(x, \lambda, \alpha) = f(x) + \sum_i \lambda_i g^i(x) + \sum_j a_j h^{(j)}(x)$$

where

$L(x, \lambda, \alpha)$ Lagrangian

$f(x)$ = given function

→ Then the Constrained minimum is given by

$$\min_{x \in S} f(x) = \min_{x} \max_{\lambda} \max_{\alpha > 0} L(x, \lambda, \alpha)$$

- KKT Conditions / necessary and sufficient Conditions for Optima :-

In Mathematical Optimization, the Karush–Kuhn–Tucker (KKT) Conditions, also known as the Kuhn–Tucker Conditions, are first derivative tests (sometimes called First Order necessary Conditions) for a solution in non-linear Programming to be Optimal, Provided that some regularity Conditions are Satisfied.

The Necessary and Sufficient Conditions for solving the Non-Linear Programming Problem with Inequality Constraints

→ We know that when Non-Linear Programming Problem with equality Constraints

Maximize / Minimize $f(x)$ such that $g_i(x) = b_i$

⟹ We can solve the Problem with Lagrangian multiplier method:

→ Now, Consider the Non-Linear Programming Problem with Inequality Constraints

Maximize / Minimize $f(x)$ such that $g_i(x) \leq b_i$

(or) $g_i(x) \geq b_i$

$\Longrightarrow$ We Can solve the Problem with KKT Method.

Conditions :-

Maximization Problem :-

Consider the NLPP maximize $f(x)$ such that

$$g_i(x) \leq b_i$$

Convert each ith inequality Constraints into equations by adding the non-negative slack variables s_i^2

Slack Variables :- $x_1 + x_2 \leq 1$

$$\Longrightarrow x_1 + x_2 + s_1 = 1$$

$$\therefore g_i(x) + s_i^2 = b_i$$

Consider $h_i(x) = g_i(x) + s_i^2 - b_i = 0 \longrightarrow (1)$

Thus given NLPP reduces to maximize $f(x)$

such that $h_i(x) = 0$

$\therefore$ Now equality Constraint so we Can use Lagrangian method.

Formulate the Lagrangian function as

$$L(x_i, s, \lambda) = f(x) - \sum_i \lambda_i h_i(x)$$

$$= f(x) - \sum_i \lambda_i (g_i(x) + s_i^2 - b_i)$$

$$(\because \text{from } (1))$$

- The necessary Conditions for stationary Points are

$$\frac{\partial L}{\partial x} = 0 \Rightarrow \frac{\partial f}{\partial x} - \sum_i \lambda_i \frac{\partial g_i}{\partial x} = 0 \longrightarrow (2)$$

$$\frac{\partial L}{\partial \lambda_i} = 0 \Rightarrow g_i(x) + S_i^2 - b_i = 0 \longrightarrow (3)$$

$$S_i^2 = b_i - g_i(x)$$

$$\frac{\partial L}{\partial S_i} = 0 \Rightarrow -2\lambda_i S_i = 0 \longrightarrow (4)$$

solve (2) (3) & (4) we get Stationary Points

multiply equations (4) by S_i & get

$$\lambda_i S_i^2 = 0$$

$$\Rightarrow \lambda_i (b_i - g_i(x)) = 0$$

$$\underline{\lambda_i = 0} \quad (\text{or}) \quad b_i - g_i(x) = 0$$

$$\Rightarrow \underline{b_i = g_i(x)}$$

Since λ_i measures the rate of variance of f wrto b_i

$$\underline{\text{ie}}: \frac{\partial f}{\partial b_i} = \lambda_i$$

from equation (4) we have either $\boxed{\lambda_i = 0}$ (or) $\boxed{S_i = 0}$ (or) both Vanish at Optimal Conditions.

Case-1 :- when $S_i \neq 0$

It means Constraint is Satisfied as strict Inequality $(\because S_i \lambda_i = 0)$

If we relaxed the Constraint (make b_i Larger) the stationary point will not be affected

$$\therefore \boxed{\lambda_i = 0}$$

Case-2 :- when $\lambda_i \neq 0$.

This implies $S_i = 0$

ie :- Constraint Satisfy as equality.

ie :- $g_i(x) = b_i$

Let $\lambda_i < 0 \implies \frac{\partial f}{\partial b_i} < 0$

This implies that as b_i is increased, the Objective function decreases.

However as b_i increases more space become feasible and the Optimal value of the Objective function $f(x)$, clearly cannot decrease.

Hence an Optimal solution

ie :- $\boxed{\lambda_i \geq 0}$

||ly for case of minimization as b_i increases $f(x)$ cannot increase which implies that

$$\boxed{\lambda_i \leq 0}$$

- **Remarks** :-

If the Constraints are equations

i.e :- $g_i(x) = b_i$

then λ_i becomes Unrestricted in sign.

Conclusion :-

Hence for Non-Linear Programming Problem

Maximize $f(x)$

Such that $g_i(x_i) \leq b_i$

The necessary Conditions

$$\frac{\partial f}{\partial x} - \sum_i \lambda_i \frac{\partial g_i}{\partial x} = 0. \quad \left(\because \frac{\partial L}{\partial x} = 0\right)$$

$$\lambda_i (g_i(x_i) - b_i) = 0$$

$$\therefore g_i(x) \leq b_i$$

$$\boxed{\lambda_i \geq 0}$$

Minimize $f(x)$

Such that $g_i(x_i) \leq b_i$

The necessary Conditions

$$\frac{\partial f}{\partial x} - \sum_i \lambda_i \frac{\partial g_i}{\partial x} = 0$$

$$\lambda_i (g_i(x_i) - b_i) = 0$$

$$g_i(x_i) \leq b_i$$

$$\boxed{\lambda_i \leq 0}$$

(1) Solve the NLPP

Maximize $Z = 3.6x_1 - 0.4x_1^2 + 1.6x_2 - 0.2x_2^2$

Such that $2x_1 + x_2 \leq 10$

$x_1, x_2 \geq 0$

Sol: For the KKT Conditions to be necessary and sufficient for Z to a maximum f(x) should be Concave

and $g(x) \leq 0$ is Convex.

for $f(x) = 3.6x_1 - 0.4x_1^2 + 1.6x_2 - 0.2x_2^2$ to be Concave we Construct the Hessian matrix as

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} -0.8 & 0 \\ 0 & -0.4 \end{bmatrix}$$

The Principle minors $D_1 = -0.8 < 0$

$$D_2 = \begin{vmatrix} -0.8 & 0 \\ 0 & -0.4 \end{vmatrix} = \underline{\underline{0.32}}$$

Thus $D_1 < 0$, $D_2 > 0$

ie:- Opposite sign with $<$ & hence it is Concave

Also the Constraint $2x_1 + x_2 \leq 10$ is Linear form

and we know every Linear function is Convex

Hence the KKT Conditions are sufficient Conditions for the maximum.

Define the Lagrangian function as

$$L = f(x) - \lambda g(x)$$

$$= (3.6x_1 - 0.4x_1^2 + 1.6x_2 - 0.2x_2^2) - \lambda(2x_1 + x_2 - 10)$$

The necessary Conditions are

$$\frac{\partial L}{\partial x} = 0, \quad \lambda g = 0, \quad \lambda \geq 0, \quad g \leq 0, \quad x \geq 0$$

$$\frac{\partial L}{\partial x_1} = 0 \Rightarrow 3.6 - 0.8x_1 - 2\lambda = 0 \longrightarrow (1)$$

$$\frac{\partial L}{\partial x_2} = 0 \Rightarrow 1.6 - 0.4x_2 - \lambda = 0 \longrightarrow (2)$$

$$\lambda g = 0 \Rightarrow \lambda(2x_1 + x_2 - 10) = 0 \longrightarrow (3)$$

$$\lambda \geq 0 \Rightarrow \lambda \geq 0 \longrightarrow (4)$$

$$g \leq 0 \Rightarrow 2x_1 + x_2 \leq 10 \longrightarrow (5)$$

$$x \geq 0 \Rightarrow x_1, x_2 \geq 0 \longrightarrow (6)$$

We have the following Cases:

Case-1: When $\boxed{\lambda = 0}$

from (1) & (2) we have $x_1 = 4.5$, $x_2 = 4$

which does not satisfy eqn (5) ~~&~~

hence this Case is discarded.

• **Case - II** :- when $\boxed{\lambda \neq 0}$

from (3) we get

$$2x_1 + x_2 = 10$$

from (1) & (2) $\quad x_1 = \dfrac{3.6 - 2\lambda}{0.8} \quad x_2 = \dfrac{1.6 - \lambda}{0.4}$

$$\therefore 2\left(\frac{3.6 - 2\lambda}{0.8}\right) + \left(\frac{1.6 - \lambda}{0.4}\right) = 10$$

$$\boxed{\lambda = 0.4} \quad (\because \lambda \geq 0)$$

Hence $x_1 = 3.5$, $x_2 = 3$ $\quad$ ($\because$ x_1 x_2 Sub values in (5) Satisfy)

(2) Solve the following NLPP

$$\text{Minimize } Z = -\log x_1 - \log x_2$$

$$\text{such that } x_1 + x_2 \leq 2$$

$$x_1 \; x_2 \geq 0$$

Sol :- For the KKT Conditions to be for Z to be minimum

$f(x)$ should be Convex & $g(x) \leq 0$ is Convex.

for $f(x)$ to be Convex, we Construct the Hessian matrix as

$$H = \begin{bmatrix} 1/x_1^2 & 0 \\ 0 & 1/x_2^2 \end{bmatrix} \quad \because H = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

Minors are $D_1 = \frac{1}{x_1^2} > 0$

$$D_2 = \begin{vmatrix} 1/x_1^2 & 0 \\ 0 & 1/x_2^2 \end{vmatrix} = \frac{1}{x_1^2 x_2^2} > 0$$

Also the Constraint $x_1 + x_2 \leq 2$ is Linear function

and hence it is Convex also

Thus KKT Conditions will be minimum

Define a Lagrangian function as

$$L = (-\log x_1 - \log x_2) - \lambda (x_1 + x_2 - 2)$$

The necessary Conditions are

$$\frac{\partial L}{\partial x} = 0, \quad \lambda g \geq 0, \quad \lambda \leq 0, \quad g \leq 0, \quad x \geq 0.$$

$$\frac{\partial L}{\partial x_1} = 0 \Rightarrow \frac{-1}{x_1} - \lambda = 0 \longrightarrow (1)$$

$$\frac{\partial L}{\partial x_2} = 0 \Rightarrow \frac{-1}{x_2} - \lambda = 0 \longrightarrow (2)$$

$$\lambda g = 0 \Rightarrow \lambda (x_1 + x_2 - 2) = 0 \longrightarrow (3)$$

$$\lambda \leq 0 \Rightarrow \lambda \leq 0 \longrightarrow (4)$$

$$g \leq 0 \Rightarrow x_1 + x_2 \leq 2 \longrightarrow (5)$$

$$x \geq 0 \Rightarrow x_1, x_2 \geq 0 \longrightarrow (6)$$

We have the following Cases.

- Case-1 :- $\boxed{\lambda = 0}$

from (1) & (2) we have

$$x_1 = \infty, \quad x_2 = \infty$$

which violate eqn (5) and hence this case is discarded.

Case-2 :- when $\boxed{\lambda \neq 0}$

from eqn (3) ; $x_1 + x_2 = 2$

from (1) & (2) we get

$$x_1 = -\frac{1}{\lambda}, \quad x_2 = -\frac{1}{\lambda}$$

Hence $-\frac{1}{\lambda} - \frac{1}{\lambda} = 2$

$$\boxed{\lambda = -1}$$

Further we have $x_1 = 1, \; x_2 = 1$

which satisfy all the necessary conditions

Hence the stationary point is

$$(x_1, x_2, \lambda) = (1, 1, -1)$$

is the optimal solution and value is

$$Z = -\log 1 - \log 1$$

$$Z = 0$$

(3) Solve the following NLPP

Maximize $Z = 8x_1 + 10x_2 - x_1^2 - x_2^2$

Such that $3x_1 + 2x_2 \leq 6$

$x_1, x_2 \geq 0$

Sol:- For the KKT Conditions to be NC & SC

for Z to be Maximum

$f(x)$ should be Concave &

$g(x) \leq 0$ is Convex.

for $f(x)$ to be Concave we Construct the

Hessian Matrix

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$$

Its minors are $D_1 = -2 < 0$

$$D_2 = \begin{vmatrix} -2 & 0 \\ 0 & -2 \end{vmatrix} = 4 > 0.$$

which is Opposite sign starting with < 0

Thus it is Concave.

Also the Constraint $3x_1 + 2x_2 \leq 6$ is a

Linear Inequality function &

hence it is Convex function

• Thus the KKT Conditions will be Necessary & sufficient for a Minimum

Define a Lagrangian function as

$$L = (8x_1 + 10x_2 - x_1^2 - x_2^2) - \lambda(3x_1 + 2x_2 - 6)$$

The necessary Conditions are

$$\frac{\partial L}{\partial x} = 0, \quad \lambda g = 0, \quad \lambda \geq 0, \quad g \leq 0, \quad x \geq 0.$$

The necessary Conditions are

$$8 - 2x_1 - 9\lambda = 0 \longrightarrow (1)$$

$$10 - 2x_1 - 2\lambda = 0 \longrightarrow (2)$$

$$\lambda(3x_1 + 2x_2 - 6) = 0 \longrightarrow (3)$$

$$\lambda \geq 0 \longrightarrow (4)$$

$$3x_1 + 2x_2 \leq 6 \longrightarrow (5)$$

$$x_1, x_2 \geq 0 \longrightarrow (6)$$

The following Cases arises.

Case-1 :- when $\boxed{\lambda = 0}$

from (1) & (2) we get $x_1 = 4, \; x_2 = 5$

which violates eqn (2) & (4) and

hence this Case is discarded

Case-2 :- when $\boxed{\lambda \neq 0}$

from eqn (3) we get.

$$3x_1 + 2x_2 = 6.$$

from ① & ② we have

$$x_1 = \frac{8-3\lambda}{2}, \quad x_2 = 5-\lambda$$

Thus $3\left(\frac{8-3\lambda}{2}\right) + 2(5-\lambda) = 6$

$$\boxed{\lambda = \frac{32}{13}}$$

Hence $x_1 = \frac{4}{13}, \quad x_2 = \frac{33}{13}$.

which satisfy all the necessary conditions.

Hence the stationary point is

$$(x_1, x_2, \lambda) = \left(\frac{4}{13}, \frac{33}{13}, \frac{32}{13}\right)$$

is optimal solution & value is

$$Z = 8\left(\frac{4}{13}\right) + 10\left(\frac{33}{13}\right) - \left(\frac{4}{13}\right)^2 - \left(\frac{33}{13}\right)^2$$

$$Z = \frac{277}{13}$$

(4) Solve the NLPP

$$\max f(x) = 4x_1 + 6x_2 - x_1^2 - x_2^2 - x_3^2$$

such that $x_1 + x_2 \leq 2$,

$2x_1 + 3x_2 \leq 12$

$x_1, x_2 \geq 0$

Sol:- For KKT Conditions to be necessary and sufficient for Z to be maximum $f(x)$ should be Concave and $g(x) \leq 0$ is Convex

for $f(x)$ to be Concave

we Construct the Hessian matrix as

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2} \end{bmatrix} \quad H = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Its minors are

$$D_1 = -1 < 0$$

$$D_2 = \begin{vmatrix} -1 & 0 \\ 0 & -1 \end{vmatrix} = 1 > 0$$

$$D_3 = \begin{vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{vmatrix} = -1 < 0$$

which is Opposite sign starting with < 0

Thus is Concave.

Also the Constraint $x_1 + x_2 \leq 2$, $2x_1 + 3x_2 \leq 12$ is Linear inequality function

and hence it is Convex function

Thus the KKT Condition will be necessary & sufficient for a Maximum.

Define the Lagrangian function L as

$$L = (4x_1 + 6x_2 - x_1^2 - x_2^2 - x_3^2) - \lambda_1 (x_1 + x_2 - 2)$$
$$- \lambda_2 (2x_1 + 3x_2 - 12)$$

The necessary Conditions are

$$\frac{\partial L}{\partial x} = 0, \quad \lambda_i g_i = 0, \quad \lambda_i \geq 0, \quad g_i \leq 0, \quad x \geq 0$$

$$\frac{\partial L}{\partial x_1} = 0 \Rightarrow 4 - 2x_1 - \lambda_1 - 2\lambda_2 = 0 \longrightarrow (1)$$

$$6 - 2x_2 - \lambda_1 - 3\lambda_2 = 0 \longrightarrow (2)$$

$$\frac{\partial L}{\partial x_2} = 0 \Rightarrow \longrightarrow (3)$$

$$\frac{\partial L}{\partial x_3} = 0 \Rightarrow -2x_3 = 0$$

$$\lambda_2 (2x_1 + 3x_2 - (2)) = 0 \longrightarrow (5)$$

$$\lambda_1, g_1 = 0 \Rightarrow \lambda_1 \lambda_2 \geq 0 \longrightarrow (6)$$

$$\lambda_1 g_1 = 0 \Rightarrow \cancel{-2x_3 = 0} \longrightarrow (4)$$

$$\lambda_1 (x_1 + x_2 - 2) = 0$$

$$\longrightarrow (7)$$

$$g_1 \leq 0 \Rightarrow x_1 + x_2 \leq 2 \longrightarrow (8)$$

$$g_2 \leq 0 \Rightarrow 2x_1 + 3x_2 \leq 12$$

$$x_1, x_2 \geq 0 \longrightarrow (9)$$

Case-1:- when $\lambda_1 = 0$, $\lambda_2 = 0$

from (1) & (2) we get $x_1 = 2$, $x_2 = 3$

This does not satisfy (7) & (8)

and hence discarded.

Case-2: when $\lambda_1 \neq 0$, $\lambda_2 = 0$

from (1) & (2)

$$-2x_1 + 4 = \lambda_1$$
$$-2x_2 + 6 = \lambda_1$$

from (4) we get $x_1 + x_2 = 2$

$$\Rightarrow \frac{4-\lambda_1}{2} + \frac{6-\lambda_1}{2} = 2$$

$$\Rightarrow \lambda_1 = 3$$

Hence $x_1 = \frac{1}{2}$, $x_2 = \frac{3}{2}$

which also satisfy (7), (8) & (9)

Hence the stationary Point is

$x_1 = \frac{1}{2}$, $x_2 = \frac{3}{2}$, $x_3 = 0$, $\lambda_1 = 3$, $\lambda_2 = 0$

The Optimal solution and value is

$$Z = 4\left(\frac{1}{2}\right) + 6\left(\frac{3}{2}\right) - \left(\frac{1}{2}\right)^2 - \left(\frac{3}{2}\right)^2 - (0)^2$$

$$\boxed{Z = \frac{17}{2}}$$

Using Hessian Matrix H :-

1) $f(x)$ is Convex $\Longleftrightarrow$ H is Positive semi-definite (Convex - minimum point)

$\Rightarrow D_1 > 0,\ D_2 \geq 0,\ D_3 \geq 0$ and atleast one of $D_i = 0$

($\because$ D = Principle minors)

(2) $f(x)$ is strictly Convex $\Longleftrightarrow$ H is Positive definite (Convex - minimum point)

$\Rightarrow D_1 > 0,\ D_2 > 0,\ D_3 > 0$

(3) $f(x)$ is Concave $\Longleftrightarrow$ $-f(x)$ is Convex.

(Convex $\Rightarrow$ Maximum Point)

$\Rightarrow D_1 < 0,\ D_2 > 0,\ D_3 < 0 \ldots$

ie: alternative sign, but first Principle minor should be -ve.

GRADIENT METHOD

(or)

Steepest Ascent method

(1) The Maximizing function is

$$f(x_1, x_2) = 4x_1 + 6x_2 - 2x_1^2 - 2x_1x_2 - 2x_2^2 \longrightarrow (1)$$

absolute optimum occurs at

$$(x_1^*, x_2^*) = (1/3, 4/3)$$

$\left(\begin{matrix} 1/3 & 4/3 \\ 0.333 & 1.333 \end{matrix}\right)$

Sol: Let the initial point be given by

$$X^0 = (1, 1)$$

$$\nabla f(x) = (4 - 4x_1 - 2x_2,\ 6 - 2x_1 - 4x_2) \longrightarrow (2)$$

(on differentiating $x_1 \cdot x_2$)

First Iteration $X^0 = (1,1)$ in (2)

$$\nabla f(x^0) = (-2, 0)$$

$$X^1 = X^0 + r\nabla f(x^0)$$

$$= (1,1) + r(-2,0)$$

$$= (1,1) + (-2r, 0)$$

$$X^1 = (\underset{x_1}{1-2r}, \underset{x_2}{1}) \longrightarrow (3)$$

$$\boxed{h(r) = f(x^1)}$$

$$= 4(1-2r) + 6(1) - 2(1-2r)^2 + 2(1-2r)(1) - 2(1)^2$$

$$h(r) = 4 - 8r^2 + 4r$$

Let $h'(r) = 0$

$-16r + 4 = 0$

$-16r = -4$

$\boxed{r = 1/4}$

Substitute $r = \frac{1}{4}$ in eqn (3)

$x' = (1 - 2r, 1)$

$= (1 - 2(1/4), 1)$

$\boxed{x' = (1/2, 1)}$

Second Iteration :

Sub $x' = (1/2, 1)$ in eqn (2)

$\nabla f(x') = (0, 1)$

$\boxed{x'' = x' + r \nabla f(x')}$

$= (1/2, 1) + r(0, 1)$

$= (1/2, 1) + (0, r)$

$\boxed{x'' = \left(\frac{1}{2}, 1 + r\right)}$

$h(r) = f(x'')$

$= 4\left(\frac{1}{2}\right) + 6(1+r) - 2\left(\frac{1}{2}\right)^2 - 2\left(\frac{1}{2}\right)(1+r) - 2(1+r)^2$

$= \cancel{2} + 6 + 6r - \frac{1}{2} - 1 - r - \cancel{2} - 2r^2 - 4r$

$= 6 + r - 2r^2 - \frac{1}{2} - 1$

$= 5 + \gamma - 2\gamma^2 - \frac{1}{2}$

$= \gamma - 2\gamma^2 + \frac{9}{2}$

$$\boxed{h(\gamma) = \gamma - 2\gamma^2 + \frac{9}{2}}$$

Let $h'(\gamma) = 0$

$1 - 4\gamma = 0$

$-4\gamma = -1$

$$\boxed{\gamma = \frac{1}{4}}$$

Sub $\gamma = 1/4$ in x''

$x'' = (1/2, 1 + \frac{1}{4})$

$x'' = (1/2, 5/4)$

Third Iteration :-

Sub $x'' = (1/2, 5/4)$ in eqn (2)

$\nabla f(x'') = (-1/2, 0)$

$$\boxed{x''' = x'' + \gamma \nabla f(x'')}$$

$= (1/2, 5/4) + \gamma(-1/2, 0)$

$= (1/2, 5/4) + (-\gamma/2, 0)$

$$\boxed{x''' = (1/2 - \gamma/2, 5/4)}$$

$$h(r) = f(x^{III})$$

$$= 4\left(1/2 - r/2\right) + 6\left(5/4\right) - 2\left(1/2 - \frac{r}{2}\right)^2 -$$

$$2\left(\frac{1}{2} - \frac{r}{2}\right)\left(\frac{5}{4}\right) - 2\left(\frac{5}{4}\right)^2$$

$$= 4\left(\frac{1-r}{2}\right) + \frac{15}{2} - 2\left(\frac{1-r}{2}\right)^2 - 2\left(\frac{1-r}{2}\right)\left(\frac{5}{4}\right)$$

$$-2\left(\frac{25}{16}\right)$$

$$= 4\left(\frac{1-r}{2}\right) + \frac{15}{2} - 2\left(\frac{1-r}{2}\right)^2 - \frac{2}{2}(1-r)\left(\frac{5}{2}\right)$$

$$-\left(\frac{25}{8}\right)$$

$$= 2(1-r) + \frac{15}{2} - 2\left(\frac{1}{4} + \frac{r^2}{4} - \frac{r}{2}\right) - \left(\frac{5}{2} - \frac{5r}{2}\right) - \frac{25}{8}$$

$$= 2 - 2r + \frac{15}{2} - \frac{1}{2} - \frac{r^2}{2} + r - \frac{5}{2} - \frac{5r}{2} - \frac{25}{8}$$

$$= -\frac{r^2}{2} - r + \frac{5r}{2} + \frac{37}{8}$$

$$\boxed{h(r) = -\frac{r^2}{2} + \frac{1}{4}r + \frac{37}{8}}$$

Let $h'(r) = 0$

$$-\frac{2r}{2} + \frac{1}{4} = 0$$

$$-r + \frac{1}{4} = 0$$

$$\boxed{r = \frac{1}{4}}$$

Sub $r = \frac{1}{4}$ in eqn x^{III}

$$x^{III} = \left(1/2 - \frac{1/4}{2}, 5/4\right)$$

$$= \left(\frac{1}{2} - \frac{1}{8}, \frac{5}{4}\right)$$

$$\boxed{x^{III} = \left(3/8, 5/4\right)}$$

$\therefore \boxed{r = \frac{1}{4}}$ $\boxed{x^{III} = \left(3/8,\ 5/4\right)}$

Fourth Iteration :-

Sub $x^{III} = \left(3/8,\ 5/4\right)$ in eqn (2)

$$\nabla f(x^{III}) = f(0, 1/4)$$

$$\boxed{x^{IV} = x^{III} + r \nabla f(x^{III})}$$

$$= \left(3/8,\ 5/4\right) + r\left(0,\ 1/4\right)$$

$$= \left(3/8,\ 5/4\right) + \left(0,\ r/4\right)$$

$$\boxed{x^{IV} = \left(\underset{x_1}{3/8},\ \underset{x_2}{5/4 + \frac{r}{4}}\right)}$$

$$h(r) = f(x^{IV})$$

$$= 4(3/8) + 6\left(\frac{5}{4} + \frac{r}{4}\right) - 2(3/8)^2 - 2(3/8)\left(\frac{5}{4} + \frac{r}{4}\right) - 2\left(\frac{5}{4} + \frac{r}{4}\right)^2$$

$$= \frac{3}{2} + \frac{30}{4} + \frac{6r}{4} - \frac{18}{64} - \left(\frac{6}{8}\right)\left(\frac{10}{4} + \frac{2r}{4}\right) - 2\left(\frac{25}{16} + \frac{r^2}{16} + \frac{20r}{4}\right)$$

$$= \frac{3}{2} + \frac{30}{4} + \frac{6r}{4} - \frac{18}{64} - \frac{60}{32} + \frac{12r}{32} - \frac{50}{16} - \frac{2r^2}{16} - \frac{20r}{4}$$

$$= -\frac{1}{8}r^2 + \frac{3r}{2} + \frac{3}{8}r - \frac{5r}{4} - \frac{60}{32} + \frac{30}{4} - \frac{50}{16} - \frac{18}{64}$$

$$= -\frac{1}{8}\gamma^2 + \frac{3\gamma}{2} + \frac{3}{8}\gamma - \frac{5\gamma}{4} - \frac{60}{32} + \frac{30}{4} - \frac{50}{16} - \frac{18}{64}$$

$$= -\frac{1}{8}\gamma^2 + \frac{24\gamma + 6\gamma - 20\gamma}{16} - \frac{149}{32}$$

$$\boxed{h(\gamma) = \frac{-1}{8}\gamma^2 + \frac{1}{16}\gamma + \frac{149}{32}}$$

Let $h'(\gamma) = 0$

$$-\frac{1}{4}\gamma + \frac{1}{16} = 0$$

$$-\frac{1}{4}\gamma = -\frac{1}{16}$$

$$\boxed{\gamma = \frac{1}{4}}$$

Sub $\gamma = \frac{1}{4}$ in eqn x^{IV}

$$x^{IV} = \left(3/8,\ 5/4 + \frac{\gamma}{4}\right)$$

$$= \left(3/8,\ \frac{5}{4} + \frac{(1/4)}{4}\right)$$

$$\boxed{x^{IV} = \left(3/8,\ \frac{21}{16}\right)}$$

Fifth Iteration: Sub $x^{IV} = (3/8,\ 21/16)$ in eqn (2)

$$\nabla f(x^{IV}) = (-1/8,\ 0)$$

$$x^{V} = x^{IV} + \gamma \nabla f(x^{IV})$$

$$= (3/8,\ 21/16) + \gamma(-1/8,\ 0)$$

$$= (3/8,\ 21/16) + (-\gamma/8,\ 0)$$

$$\boxed{h(\gamma) = \frac{-2\gamma^2}{64} + \frac{1}{64}\gamma + \frac{597}{38}}$$

$$\boxed{x^{V} = \left(\frac{3}{8} - \frac{\gamma}{8},\ \frac{21}{16}\right)}$$

$$h(\gamma) = f(x^{V})$$

$$= 4\left(\frac{3-\gamma}{8}\right) + 6\left(\frac{21}{16}\right) - 2\left(\frac{3-\gamma}{8}\right)^2 - 2\left(\frac{3-\gamma}{8}\right)\left(\frac{21}{16}\right) - 2\left(\frac{21}{16}\right)^2$$

$$\boxed{h(\gamma) = \frac{-2\gamma^2}{64} + \frac{1}{64}\gamma + \frac{597}{38}}$$

Let $h'(\gamma) = 0$

$$\frac{-4\gamma}{64} + \frac{1}{64} = 0$$

$$-\frac{\gamma}{16} + \frac{1}{64} = 0$$

$$-\frac{\gamma}{16} = -\frac{1}{64}$$

$$\boxed{\gamma = \frac{1}{4}}$$

Sub $\gamma = \frac{1}{4}$ in eqn X^v

$$X^v = \left(\frac{3-\gamma}{8}, \frac{21}{16}\right)$$

$$= \left(\frac{3-\left(\frac{1}{4}\right)}{8}, \frac{21}{16}\right)$$

$$\boxed{X^v = \left(11/32, \frac{21}{16}\right)}$$

Sixth Iteration :- sub $X^v = \left(\frac{11}{32}, \frac{21}{16}\right)$ in eqn 2

$$\nabla f(X^v) = \left(0, \frac{1}{16}\right)$$

Because $\nabla f(X^v) \approx 0$, the Process can be terminated at this point.

$\therefore$ The approximate maximum Point is given by

$$X^v = \left(0.34375, 1.3125\right) \quad (\text{or})$$

$$X^v = \left(\frac{11}{32}, \frac{21}{16}\right)$$

$\therefore$ The exact Optimum is

$$X^* = \left(0.3333, 1.3333\right)$$