Unit 18. Convex Optimization

Convex Optimization - Introduction

This course is useful for the students who want to solve non-linear optimization problems that arise in various engineering and scientific applications. This course starts with basic theory of linear programming and will introduce the concepts of convex sets and functions and related terminologies to explain various theorems that are required to solve the non linear programming problems. This course will introduce various algorithms that are used to solve such problems. These type of problems arise in various applications including machine learning, optimization problems in electrical engineering, etc. It requires the students to have prior knowledge of high school maths concepts and calculus.

In this course, the students will learn to solve the optimization problems like \( \min f(x) \) subject to some constraints. These problems are easily solvable if the function \( f(x) \) is a linear function and if the constraints are linear. Then it is called a linear programming problem (LPP). But if the constraints are non-linear, then it is difficult to solve the above problem. Unless we can plot the functions in a graph, then try to analyse the optimization can be one way, but we can't plot a function if it's beyond three dimensions. Hence there comes the techniques of non-linear programming or convex programming to solve such problems. In these tutorial, we will focus on learning such techniques and in the end, a few algorithms to solve such problems. first we will bring the notion of convex sets which is the base of the convex programming problems. Then with the introduction of convex functions, we will some important theorems to solve these problems and some algorithms based on these theorems.

Terminologies

- The space \( \mathbb{R}^n \) – It is an n-dimensional vector with real numbers, defined as follows
  \[
  \mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_1, x_2, \ldots, x_n \in \mathbb{R}\}
  \]
• The space $\mathbb{R}^{m \times n}$ – It is a set of all real values matrices of order $m \times n$.  

Convex Optimization - Linear Programming

Methodology

Linear Programming also called Linear Optimization, is a technique which is used to solve mathematical problems in which the relationships are linear in nature. The basic nature of Linear Programming is to maximize or minimize an **objective function** with subject to some **constraints**. The objective function is a linear function which is obtained from the mathematical model of the problem. The constraints are the conditions which are imposed on the model and are also linear.

• From the given question, find the objective function.  
• find the constraints.  
• Draw the constraints on a graph.  
• find the feasible region, which is formed by the intersection of all the constraints.  
• find the vertices of the feasible region.  
• find the value of the objective function at these vertices.  
• The vertex which either maximizes or minimizes the objective function (according to the question) is the answer.

**Examples**

**Step 1** – Maximize $5x + 3y$ subject to  
$x + y \leq 2x + y \leq 2$,  
$3x + y \leq 3x + y \leq 3$,  
$x \geq 0$ and $y \geq 0$  

**Solution** –

The first step is to find the feasible region on a graph.
Clearly from the graph, the vertices of the feasible region are 

\[(0,0),(0,2),(1,0),(12,32),(0,0),(0,2),(1,0),(12,32)\]

Let \(f(x,y) = 5x + 3y\)

Putting these values in the objective function, we get –

\[\begin{align*}
    f(0,0) &= 0 \\
    f(0,2) &= 6 \\
    f(1,0) &= 5 \\
    f(12,32) &= 7 \\
\end{align*}\]

Therefore, the function maximizes at \((12,32)\)

**Step 2** – A watch company produces a digital and a mechanical watch. Long-term projections indicate an expected demand of at least 100 digital and 80 mechanical watches each day. Because of limitations on production capacity, no more than 200 digital and 170 mechanical watches can be made daily. To satisfy a shipping contract, a total of at least 200 watches must be shipped each day.

If each digital watch sold results in a $2 loss, but each mechanical watch produces a $5 profit, how many of each type should be made daily to maximize net profits?

**Solution** –
Let $x$ be the number of digital watches produced
$y$ be the number of mechanical watches produced
According to the question, at least 100 digital watches are to be made daily and maximum 200 digital watches can be made.

$\Rightarrow 100 \leq x \leq 200$  

Similarly, at least 80 mechanical watches are to be made daily and maximum 170 mechanical watches can be made.

$\Rightarrow 80 \leq y \leq 170$  

Since at least 200 watches are to be produced each day.

$\Rightarrow x + y \leq 200$  

Since each digital watch sold results in a $2\text{ loss}$, but each mechanical watch produces a $5\text{ profit}$,

Total profit can be calculated as

Profit = $-2x + 5y$

And we have to maximize the profit, Therefore, the question can be formulated as

Maximize $-2x + 5y$ subject to

$100 \leq x \leq 200$
$80 \leq y \leq 170$

Plotting the above equations in a graph, we get,
The vertices of the feasible region are
(100, 170)(200, 170)(200, 180)(120, 80) and (100, 100)(100, 170)(200, 180)(120, 80) and (100, 100)
The maximum value of the objective function is obtained at (100, 170) (100, 170)
Thus, to maximize the net profits, 100 units of digital watches and 170 units of mechanical watches should be produced.

Convex Optimization - Norm

A norm is a function that gives a strictly positive value to a vector or a variable.

Norm is a function \( f: \mathbb{R}^n \rightarrow \mathbb{R} \)
The basic characteristics of a norm are −

Let \( X \) be a vector such that \( X \in \mathbb{R}^n \)
- \( \|x\| \geq 0 \) and \( \|x\| = 0 \iff x = 0 \)
- \( \|\alpha x\| = |\alpha| \|x\| \) for all \( \alpha \in \mathbb{R} \) and \( x \in \mathbb{R}^n \)
- \( \|x + y\| \leq \|x\| + \|y\| \) for all \( x, y \in \mathbb{R}^n \)
- \( \|x - y\| \geq \|x\| - \|y\| \) for all \( x, y \in \mathbb{R}^n \)

By definition, norm is calculated as follows −
Let \( \langle x, y \rangle \) be the inner product of vectors \( x \) and \( y \).

Inner product is a function which gives a scalar to a pair of vectors.

\[ \langle x, y \rangle = x^T y \]

Hence, norm is a continuous function.

\[ \| x \| = \sqrt{\langle x, x \rangle} \]

\[ \| y \| = \sqrt{\langle y, y \rangle} \]

\[ \| x \| = \sqrt{\sum_{i=1}^{n} x_i^2} \]

\[ \| x \| = \sum_{i=1}^{n} | x_i | \]

Therefore, \( \| f(x_n) - f(x) \| \leq \| x_n - x \| \) for all \( x_n \) and \( x \) in \( X \).

Since \( X_n \rightarrow x \) in \( X \),

Therefore \( \| f(x_n) - f(x) \| \leq 0 \) \( \Rightarrow \| x_n - x \| \leq 0 \)

Hence, norm is a continuous function.

Convex Optimization - Inner Product

Inner product is a function which gives a scalar to a pair of vectors.

Inner Product \( f: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{K} \) where \( \mathbb{K} \) is a scalar.

The basic characteristics of inner product are as follows:

Let \( X \in \mathbb{R}^n \)

- \( \langle x, x \rangle \geq 0, \forall x \in X \)
- \( \langle x, x \rangle = 0 \iff x = 0, \forall x \in X \)
- \( \langle ax, y \rangle = a \langle x, y \rangle, \forall a \in \mathbb{K} \) and \( \forall x \in X \)
- \( \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle, \forall x, y, z \in X \)
- \( \langle y, x \rangle = \langle x, y \rangle, \forall x, y \in X \)

Note:

- Relationship between norm and inner product: \( \| x \| = \sqrt{\langle x, x \rangle} \)

\[ \forall x, y \in \mathbb{R}^n, \langle x, y \rangle = x_1y_1 + x_2y_2 + \ldots + x_ny_n \] 

Examples

1. Find the inner product of \( x = (1, 2, 1) \) and \( y = (3, -1, 3) \)

Solution

\( \langle x, y \rangle = x_1y_1 + x_2y_2 + x_3y_3 = 1 \cdot 3 + 2 \cdot (-1) + 1 \cdot 3 = 3 + 2 + 3 = 8 \)

2. If \( x = (4, 9, 1), y = (-3, 5, 1) \) and \( z = (2, 4, 1) \), find \( \langle x + y, z \rangle \)

THANKS FOR READING – VISIT OUR WEBSITE www.educatererindia.com
Solution
As we know, \((x+y,z) = (x,z) + (y,z)\)
\((x+y,z) = (x_1 z_1 + x_2 z_2 + x_3 z_3) + (y_1 z_1 + y_2 z_2 + y_3 z_3)\)
\((x+y,z) = (x_1 z_1 + x_2 z_2 + x_3 z_3) + (y_1 z_1 + y_2 z_2 + y_3 z_3)\)
\((x+y,z) = (4\times 2) + (9\times 4) + (1\times 1)\) + \((x+y,z) = (4\times 2) + (9\times 4) + (1\times 1)\) +
\((-3\times 2) + (5\times 4) + (1\times 1)\) + \((-3\times 2) + (5\times 4) + (1\times 1)\)
\((x+y,z) = (8+36+1) + (-6+20+1)\) \((x+y,z) = (8+36+1) + (-6+20+1)\)
\((x+y,z) = 45+15\) \((x+y,z) = 45+15\)
\((x+y,z) = 60\) \((x+y,z) = 60\)

Convex Optimization - Minima and Maxima

Local Minima or Minimize
\(x \in S x \in S\) is said to be local minima of a function \(f\) if \(f(x) \leq f(y), \forall x \in N_e(x), f(x) \leq f(x)\), \(\forall x \in N_e(x)\) where \(N_e(x)\) means neighbourhood of \(x\), \(x\), i.e., \(N_e(x)\) means \(\|x-x\| < \varepsilon\)

Local Maxima or Maximizer
\(x \in S x \in S\) is said to be local maxima of a function \(f\) if \(f(x) \geq f(y), \forall x \in N_e(x), f(x) \geq f(x)\), \(\forall x \in N_e(x)\) where \(N_e(x)\) means neighbourhood of \(x\), \(x\), i.e., \(N_e(x)\) means \(\|x-x\| < \varepsilon\)

Global minima
\(x \in S x \in S\) is said to be global minima of a function \(f\) if \(f(x) \leq f(x), \forall x \in S f(x) \leq f(x), \forall x \in S\)

Global maxima
\(x \in S x \in S\) is said to be global maxima of a function \(f\) if \(f(x) \geq f(x), \forall x \in S f(x) \geq f(x), \forall x \in S\)

Examples
Step 1 - find the local minima and maxima of \(f(x) = \|x-2\|\) \(f(x) = \|x-2\|\)

Solution -
From the graph of the above function, it is clear that the local minima occurs at \( x = \pm 2 \) and local maxima at \( x = 0 \).

**Step 2** – find the global minima of the function \( f(x) = |4x^3 - 3x^2 + 7| \).

**Solution** –

From the graph of the above function, it is clear that the global minima occurs at \( x = -1 \).

**Convex Optimization - Convex Set**
Let \( S \subseteq R^n \), \( S \subseteq R^n \). A set \( S \) is said to be convex if the line segment joining any two points of the set \( S \) also belongs to the set \( S \), i.e., if \( x_1, x_2 \in S \), then \( \lambda x_1 + (1-\lambda)x_2 \in S \) for \( \lambda \in (0,1) \).

**Note**

- The union of two convex sets may or may not be convex.
- The intersection of two convex sets is always convex.

**Proof**

Let \( S_1 \) and \( S_2 \) be two convex set.

Let \( S_3 = S_1 \cap S_2 \) and \( x_1, x_2 \in S_3 \).

Since \( S_3 = S_1 \cap S_2 \) then \( x_1, x_2 \in S_1 \) and \( x_1, x_2 \in S_2 \).

Thus \( \lambda x_1 + (1-\lambda)x_2 \in S_1 \cap S_2 \), where \( \lambda \in (0,1) \).

Therefore, \( \lambda x_1 + (1-\lambda)x_2 \in S_3 \) implies \( \lambda x_1 + (1-\lambda)x_2 \in S_1 \cap S_2 \).

Hence, \( S_3 \) is a convex set.

**Examples**

**Step 1** – Prove that the set \( S = \{ x \in R^n : Cx \leq a \} \) is a convex set.

**Solution**

Let \( x_1, x_2 \in S \).

\( \Rightarrow Cx_1 \leq a \) and \( Cx_2 \leq a \).

To show: \( y = (\lambda x_1 + (1-\lambda)x_2) \in S \), where \( \lambda \in (0,1) \).

\( Cy = C(\lambda x_1 + (1-\lambda)x_2) = \lambda Cx_1 + (1-\lambda)Cx_2 \)

\( \Rightarrow Cy \leq \lambda a + (1-\lambda)a \)

\( \Rightarrow Cy \leq \lambda a \) and \( Cy \leq \lambda a \).

\( y \in S \).

Therefore, \( S \) is a convex set.

**Step 2** – Prove that the set \( S = \{ (x_1, x_2) \in R^2 : x_2 \leq 8x_2 \} \) is a convex set.
Solution

Let \( x, y \in S \times x, y \in S \)

Let \( x = (x_1, x_2) \) and \( y = (y_1, y_2) \)

\[ x_2 \leq x_2 = x_1 \leq x_2 \text{ and } y_2 \leq y_2 y_1 \leq x_2 \]

To show

\[ \lambda x + (1-\lambda)y \in S \iff \lambda (x_1, x_2) + (1-\lambda)(y_1, y_2) \in S \]

Now, \( \lambda x_1 + (1-\lambda)y_1 \leq \lambda x_2 + (1-\lambda)y_2 \)

But \( x_1 y_1 \leq x_2 y_1 \)

Therefore,

\[ \lambda x_1 + (1-\lambda)y_1 \leq \lambda x_2 + (1-\lambda)y_2 \]

Step 3 – Show that a set \( S \in \mathbb{R}^n \) is convex if and only if for each integer \( k \), every convex combination of any \( k \) points of \( S \) is in \( S \).

Solution

Let \( S \) be a convex set. Then, to show;

\[ c_1 x_1 + c_2 x_2 + \cdots + c_k x_k \in S, \sum_{i=1}^{k} c_i = 1, c_i \geq 0, \forall i \in 1, 2, \ldots, k \]

Proof by induction

For \( k = 1, x_1 \in S \), \( c_1 = 1 \Rightarrow c_1 x_1 \in S 

For \( k = 2, x_1, x_2 \in S \), \( c_1 + c_2 = 1 \Rightarrow c_1 x_1 + c_2 x_2 \in S \) and Since \( S \) is a convex set

\[ c_1 x_1 + c_2 x_2 \in S \Rightarrow c_1 x_1 + c_2 x_2 \in S \]

Let the convex combination of \( m \) points of \( S \) be in \( S \), i.e.,

\[ c_1 x_1 + c_2 x_2 + \cdots + c_m x_m \in S, \sum_{i=1}^{m} c_i = 1, c_i \geq 0, \forall i \in 1, 2, \ldots, m \]

Now, Let \( x_1, x_2, \ldots, x_m, x_m + 1 \in S \times x_1, x_2, \ldots, x_m, x_m + 1 \in S \)

Let \( x = \mu_1 x_1 + \mu_2 x_2 + \cdots + \mu_m x_m + 1 \in S \times x_1, x_2, \ldots, x_m, x_m + 1 \in S \)

Let \( x = (\mu_1 + \mu_2 + \cdots + \mu_m) x_1 + (\mu_2 + \cdots + \mu_m) x_2 + \cdots + (\mu_1 + \mu_2 + \cdots + \mu_m) x_m + 1 \in S \times x_1, x_2, \ldots, x_m, x_m + 1 \in S \)

Now \( y \in S \times y \in S \) because the sum of the coefficients is 1.
\[ x \in S \Rightarrow x \in S \text{ since } S \text{ is a convex set and } y, x_{m+1} \in S, y, x_{m+1} \in S \]

Hence proved by induction.

**Convex Optimization - affine Set**

A set \( AA \) is said to be an affine set if for any two distinct points, the line passing through these points lie in the set \( AA \).

*Note –*

- \( SS \) is an affine set if and only if it contains every affine combination of its points.
- Empty and singleton sets are both affine and convex set.

For example, solution of a linear equation is an affine set.

**Proof**

Let \( S \) be the solution of a linear equation.

By definition, \( S = \{ x \in \mathbb{R}^n : Ax = b \} \) \( S = \{ x \in \mathbb{R}^n : Ax = b \} \)

Let \( x_1, x_2 \in S \) \( \Rightarrow Ax_1 = b, x_1, x_2 \in S \)

To prove: \( A[\theta x_1 + (1-\theta)x_2] = b, \forall \theta \in (0, 1) \)

\( A[\theta x_1 + (1-\theta)x_2] = \theta Ax_1 + (1-\theta)Ax_2 = \theta b + (1-\theta)b = \theta b + (1-\theta)b = b \)

Thus \( S \) is an affine set.

**Theorem**

If \( CC \) is an affine set and \( x_0 \in C, x_0 \in C \), then the set \( V = C - x_0 = \{ x - x_0 : x_0 \in C \} \) \( V = C - x_0 = \{ x - x_0 : x_0 \in C \} \) is a subspace of \( C \).

**Proof**

Let \( x_1, x_2 \in V = x_1, x_2 \in V \)

To show: \( \alpha x_1 + \beta x_2 \in V \) \( \alpha x_1 + \beta x_2 \in V \) for some \( \alpha, \beta, \alpha \), \( \beta \)

Now, \( x_1 + x_0 \in C \) and \( x_2 + x_0 \in C \) \( x_2 + x_0 \in C \) by definition of \( V \)

Now, \( \alpha x_1 + \beta x_2 + x_0 = \alpha (x_1 + x_0) + \beta (x_2 + x_0) + (1 - \alpha - \beta) x_0 \) \( \alpha (x_1 + x_0) + \beta (x_2 + x_0) + (1 - \alpha - \beta) x_0 \)

But \( \alpha (x_1 + x_0) + \beta (x_2 + x_0) + (1 - \alpha - \beta) x_0 \in C \) \( \alpha (x_1 + x_0) + \beta (x_2 + x_0) + (1 - \alpha - \beta) x_0 \in C \) because \( C \) is an affine set.

Therefore, \( \alpha x_1 + \beta x_2 \in V = x_1 + \beta x_2 \in V \)

Hence proved.

**Convex Optimization - Hull**

The convex hull of a set of points in \( S \) is the boundary of the smallest convex region that contain all the points of \( S \) inside it or on its boundary.
OR

Let \( S \subseteq \mathbb{R}^n \) \( \subseteq \mathbb{R}^n \) The convex hull of \( S \), denoted \( \text{Co}(S) \), is the collection of all convex combinations of \( S \), i.e., \( x \in \text{Co}(S) \) if and only if \( x = \sum_{i=1}^{n} \lambda_i x_i \), where \( \sum_{i=1}^{n} \lambda_i = 1 \) and \( \lambda_i \geq 0 \) for all \( x_i \in S \).

Remark – Convex hull of a set of points in \( S \) in the plane defines a convex polygon and the points of \( S \) on the boundary of the polygon defines the vertices of the polygon.

Theorem \( \text{Co}(S) = \{ x : x = \sum_{i=1}^{n} \lambda_i x_i, x_i \in S, \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geq 0 \} \). Show that a convex hull is a convex set.

Proof

Let \( x_1, x_2 \in \text{Co}(S) \) \( \subseteq \mathbb{R}^n \). Then \( x = \sum_{i=1}^{n} \lambda_i x_i \) and \( x = \sum_{i=1}^{n} \lambda_i x_i \) where \( \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geq 0 \).

For \( \theta \in (0, 1) \), \( \theta x_1 + (1 - \theta) x_2 = \theta \sum_{i=1}^{n} \lambda_i x_i + (1 - \theta) \sum_{i=1}^{n} \lambda_i x_i = \sum_{i=1}^{n} \lambda_i \theta x_i + \sum_{i=1}^{n} \lambda_i (1 - \theta) x_i \).

Consider the coefficients,

\[ \sum_{i=1}^{n} \lambda_i \theta + \theta (1 - \theta) = \theta \sum_{i=1}^{n} \lambda_i + (1 - \theta) \sum_{i=1}^{n} \lambda_i = 1 \]

Hence, \( \theta x_1 + (1 - \theta) x_2 \in \text{Co}(S) \).

Thus, a convex hull is a convex set.

Caratheodory Theorem

Let \( S \) be an arbitrary set in \( \mathbb{R}^n \). If \( x \in \text{Co}(S) \), then \( x = \sum_{i=1}^{n} \lambda_i x_i \) \( \subseteq \mathbb{R}^n \).

Proof

Since \( x \in \text{Co}(S) \), then \( x \) is represented by a convex combination of a finite number of points of \( S \), i.e., \( x = \sum_{i=1}^{k} \lambda_i x_i \) \( \subseteq \mathbb{R}^n \) \( \subseteq \mathbb{R}^n \) \( \subseteq \mathbb{R}^n \) and \( x_j \in S, \forall j \in \{1, k\} \).

If \( k \leq n + 1 \), the result obtained is obviously true.

If \( k \geq n + 1 \), then \( (x_2 - x_1)(x_3 - x_1), \ldots, (x_k - x_1)(x_2 - x_1)(x_3 - x_1), \ldots, (x_k - x_1) \) are linearly dependent.

\( \Rightarrow \exists \mu \in \mathbb{R}, 2 \leq j \leq k \Rightarrow \exists \mu \in \mathbb{R}, 2 \leq j \leq k \) such that \( \sum_{j=2}^{k} \mu_j (x_j - x_1) = 0 \)
Define \( \mu_1 = -\sum_{j}^2 k\mu_j \mu_1 = -\sum_{j}^2 k\mu_j \).

Then \( \sum_{j}^1 k\mu_j x_j = 0, \sum_{j}^1 k\mu_j = 0 \sum_{j}^1 k\mu_j x_j = 0, \sum_{j}^1 k\mu_j = 0 \)

where not all \( \mu_j \)’s are equal to zero. Since \( \sum_{j}^1 k\mu_j = 0 \sum_{j}^1 k\mu_j = 0 \), at least one of the \( \mu_j > 0, 1 \leq j \leq k \).

Thus, \( x \) can be represented as a convex combination of at most \( (k-1) \) points.

This reduction process can be repeated until \( x \) is represented as a convex combination of \( (n+1) \) elements.

Convex Optimization - Weierstrass Theorem

Let \( S \) be a non-empty, closed and bounded set (also called compact set) in \( \mathbb{R}^n \) and let \( f:S \to \mathbb{R}:f:S \to \mathbb{R} \) be a continuous function on \( S \), then the problem \( \min \{ f(x):x \in S \}|f(x):x \in S \} \) attains its minimum.

Proof

Since \( S \) is non-empty and bounded, there exists a lower bound.

\( a = \inf \{ f(x):x \in S \} = \inf \{ f(x):x \in S \} \)

Now let \( S_j = \{ x \in S: \alpha \leq f(x) \leq \alpha + \delta j \} \forall j = 1,2, \ldots S_j = \{ x \in S: \alpha \leq f(x) \leq \alpha + \delta j \} \forall j = 1,2, \ldots \) and \( \delta \in (0,1) \delta \in (0,1) \)

By the definition of infimum, \( S_j \) is non-empty, for each \( j \).

Choose some \( x_j \in S_j \) to get a sequence \( \{ x_j \} \) for \( j = 1,2, \ldots j = 1,2, \ldots \)

Since \( S \) is bounded, the sequence is also bounded and there is a convergent subsequence \( \{ y_j \} \) of \( \{ y_j \} \), which converges to \( x^\wedge x^\wedge \). Hence \( x^\wedge x^\wedge \) is a limit point and \( S \) is closed, therefore, \( x^\wedge \in Sx^\wedge \in S \). Since \( f \) is continuous, \( f(y_j) \to f(x^\wedge) \).

Since \( a \leq f(y_j) \leq \alpha + \delta k, a = \lim k \to \infty f(y_i) = f(x^\wedge) \alpha \leq f(y_j) \leq \alpha + \delta k, a = \lim k \to \infty f(y_i) = f(x^\wedge) \)

Thus, \( x^\wedge \) is the minimizing solution.

Remarks
There are two important necessary conditions for Weierstrass Theorem to hold. These are as follows –

- **Step 1** – The set $S$ should be a bounded set.

  Consider the function $f: \mathbb{R} \rightarrow x$.

  It is an unbounded set and it does have a minima at any point in its domain.

  Thus, for minima to obtain, $S$ should be bounded.

- **Step 2** – The set $S$ should be closed.

  Consider the function $f(x) = 1x$, $1x$ in the domain \left (0,1 \right )$.

  This function is not closed in the given domain and its minima also does not exist.

  Hence, for minima to obtain, $S$ should be closed.

### Convex Optimization - Closest Point Theorem

Let $S$ be a non-empty closed convex set in $\mathbb{R}^n$ and let $y \in S \in S$, then $\exists \exists$ a point $x \in S \in S$ such that the minimum distance of $S$ from $y$ is less than or equal to $\|y-x\| \subseteq y-x \|$. There are two important necessary conditions for Weierstrass Theorem to hold. These are as follows –

- **Step 1** – The set $S$ should be a bounded set.

  Consider the function $f: \mathbb{R} \rightarrow x$.

  It is an unbounded set and it does have a minima at any point in its domain.

  Thus, for minima to obtain, $S$ should be bounded.

- **Step 2** – The set $S$ should be closed.

  Consider the function $f(x) = 1x$, $1x$ in the domain \left (0,1 \right )$.

  This function is not closed in the given domain and its minima also does not exist.

  Hence, for minima to obtain, $S$ should be closed.

### Existence of closest point

Since $S \neq \emptyset, \exists S \neq \emptyset, \exists$ a point $x^\in S \in S$ such that the minimum distance of $S$ from $y$ is less than or equal to $\|y-x\| \subseteq y-x \|$. Define $S^\in S \in S \subseteq S \subseteq \{x: \|y-x\| \subseteq y-x \|\}$.

Since $S^\in S \in S$ is closed and bounded, and since norm is a continuous function, then by Weierstrass theorem, there exists a minimum point $x^\in S \in S$ such that $\|y-x\| = \inf \{\|y-x\|, x \in S\} \|y-x\| = \inf \{\|y-x\|, x \in S\}$

### Uniqueness

Suppose $x^\in S \in S$ such that $\|y-x\| = \|y-x\| = \alpha |y-x| = \|y-x\| = \alpha$

Since $S$ is convex, $x^\in S \in S$ and $\bar{y}$ are in $S$.

But, $\|y-x\| + \|y-x\| = \alpha |y-x| + \alpha |y-x| = \alpha |y-x| + \alpha |y-x|$

It can’t be strict inequality because $x^\in S \in S$ is closest to $y$.

Therefore, $\|y-x\| = \|y-x\| = \|y-x\| = \|y-x\| = \|y-x\|$, for some $\mu \mu$

Now $\|\mu\| = 1, \|\mu\| = 1$.

If $\mu = -1 \mu = -1$,

then $(y-x^\in S \in S)$ \Rightarrow $y = x^\in S \in S$ and $(y-x^\in S \in S)$ \Rightarrow $y = x^\in S \in S$.

But $y \in S \in S$. Hence contradiction. Thus $\mu = 1 \Rightarrow x^\in S \in S$ and $\mu = 1 \Rightarrow x^\in S \in S$

Thus, minimizing point is unique.
For the second part of the proof, assume \((y - x^\lambda)^T(x - x^\lambda) \leq 0\) for all \(x \in S\).

Now,

\[
\|y - x\|^2 = \|y - x^\lambda + x^\lambda - x\|^2 = \|y - x\|^2 + \|x^\lambda - x\|^2 + 2(y - x^\lambda)^T(y - x) = \|y - x\|^2 + \|x^\lambda - x\|^2 + 2(y - x^\lambda)^T(y - x)
\]

Since \(S\) is convex set.

\[
\Rightarrow \|y - x\|^2 = \|y - x\|^2 + \|x^\lambda - x\|^2 + 2(y - x^\lambda)^T(y - x) \geq 0
\]

Thus, \(x^\lambda x^\lambda\) is minimizing point.

Conversely, assume \(x^\lambda x^\lambda\) is minimizing point.

\[
\Rightarrow \|y - x\|^2 \geq \|y - x\|^2 \forall x \in S \Rightarrow \|y - x\|^2 \geq \|y - x\|^2 \forall x \in S
\]

Since \(S\) is convex set.

\[
\Rightarrow \lambda x + (1 - \lambda)x^\lambda = x^\lambda + \lambda(x - x^\lambda) \in S \Rightarrow \lambda x + (1 - \lambda)x^\lambda = x^\lambda + \lambda(x - x^\lambda) \in S, \text{ for } x \in S \forall x \in S \text{ and } \lambda \in (0, 1)
\]

Now, \(\|y - x\|^2 = \|y - x\|^2 \lambda x + (1 - \lambda)x^\lambda \geq \|y - x\|^2 \lambda x + (1 - \lambda)x^\lambda \leq 0\)

And

\[
\|y - x\|^2 - \lambda(x - x^\lambda) = 2 = \|y - x\|^2 + \lambda 2\|x - x^\lambda\|^2 = 2(y - x^\lambda)^T(x - x^\lambda)
\]

\[
\Rightarrow \|y - x\|^2 + \lambda 2\|x - x^\lambda\|^2 = \|y - x\|^2 + \lambda 2\|x - x^\lambda\|^2 - 2\lambda(y - x^\lambda)^T(x - x^\lambda)
\]

\[
\Rightarrow \|y - x\|^2 + \lambda 2\|x - x^\lambda\|^2 = \|y - x\|^2 + \lambda 2\|x - x^\lambda\|^2 - 2\lambda(y - x^\lambda)^T(x - x^\lambda)
\]

\[
\Rightarrow 2\lambda(y - x^\lambda)^T(x - x^\lambda) \leq \lambda 2\|x - x^\lambda\|^2 \Rightarrow 2\lambda(y - x^\lambda)^T(x - x^\lambda) \leq \lambda 2\|x - x^\lambda\|^2
\]

\[
\Rightarrow (y - x^\lambda)^T(x - x^\lambda) \leq 0 \Rightarrow (y - x^\lambda)^T(x - x^\lambda) \leq 0
\]

Hence Proved.

**Fundamental Separation Theorem**

Let \(S\) be a non-empty closed, convex set in \(\mathbb{R}^n\) and \(y \in S\). Then, there exists a non zero vector \(p\) and scalar \(\beta\) such that 

\[
py > \beta pTy \quad \text{and} \quad px < \beta pTx < \beta \text{ for each } x \in S
\]

**Proof**

Since \(S\) is non empty closed convex set and \(y \in S\) thus by closest point theorem, there exists a unique minimizing point \(x^\lambda \in S\) such that 

\[
(x - x^\lambda)^T(y - x^\lambda) \leq 0 \forall x \in S \quad \text{and} \quad (x - x^\lambda)^T(y - x^\lambda) \leq 0 \forall x \in S
\]

Let \(p = (y - x^\lambda) \neq 0\) and \(\beta = x^\lambda^T(y - x^\lambda) = px^\lambda \leq y^T(y - x^\lambda) = pTy^\lambda.

Then 

\[
(x - x^\lambda)^T(y - x^\lambda) \leq 0 \Rightarrow (x - x^\lambda)^T(x - x^\lambda) \leq 0
\]

\[
\Rightarrow (y - x^\lambda)^T(x - x^\lambda) \leq 0 \Rightarrow (y - x^\lambda)^T(x - x^\lambda) \leq 0
\]

\[
\Rightarrow (y - x^\lambda)^TX(x - x^\lambda) = x^\lambda^T(y - x^\lambda) \Rightarrow (y - x^\lambda)^TX(x - x^\lambda) = x^\lambda^T(y - x^\lambda) \quad \text{i.e.,} \quad pTx \leq \beta
\]

Also, \(pTy \geq \beta (y - x^\lambda)\)

\[
(y - x^\lambda)^T(y - x^\lambda) \geq 0 \Rightarrow (y - x^\lambda)^T(y - x^\lambda) \geq 0
\]

\[
\Rightarrow py > \beta pTy \geq \beta
\]
This theorem results in separating hyperplanes. The hyperplanes based on the above theorem can be defined as follows –

Let $S_1$ and $S_2$ are be non-empty subsets of $\mathbb{R}$ and $H = \{x : A^T x = b\}$. Let $H = \{x : A^T x = b\}$ be a hyperplane.

- The hyperplane $H$ is said to separate $S_1$ and $S_2$ if $A^T x \leq b \forall X \in S_1 A^T x \leq b \forall X \in S_1$ and $A^T x \geq b \forall X \in S_2 A^T x \geq b \forall X \in S_2$.
- The hyperplane $H$ is said to strictly separate $S_1$ and $S_2$ if $A^T x < b \forall X \in S_1 A^T x < b \forall X \in S_1$ and $A^T x > b \forall X \in S_2 A^T x > b \forall X \in S_2$.
- The hyperplane $H$ is said to strongly separate $S_1$ and $S_2$ if $A^T x \leq b \forall X \in S_1 A^T x \leq b \forall X \in S_1$ and $A^T x \geq b + \epsilon \forall X \in S_2 A^T x \geq b + \epsilon \forall X \in S_2$, where $\epsilon$ is a positive scalar.

**Convex Optimization - Cones**

A non-empty set $C$ in $\mathbb{R}^n$ is said to be a cone with vertex 0 if $x \in C \Rightarrow \lambda x \in C \forall \lambda \geq 0 x \in C \Rightarrow \lambda x \in C \forall \lambda \geq 0$.

A set $C$ is a convex cone if it is convex as well as cone.

For example, $y = \|x\| |y| = |x|$ is not a convex cone because it is not convex.

But, $y \geq \|x\| |y| = |x|$ is a convex cone because it is convex as well as cone.

**Note** – A cone $C$ is convex if and only if for any $x, y \in C, x + y \in C, y \in C, x + y \in C$.

**Proof**

Since $C$ is a cone, for $x, y \in C \Rightarrow \lambda x \in C, y \in C \Rightarrow \lambda x \in C$ and $\mu y \in C \forall \lambda, \mu \geq 0 \mu y \in C \forall \lambda, \mu \geq 0$.

$C$ is convex if $\lambda x + (1 - \lambda) y \in C \forall \lambda \in (0, 1), (1 - \lambda) y \in C \forall \lambda \in (0, 1)$.

Since $C$ is a cone, $\lambda x \in C, \lambda x \in C$ and $(1 - \lambda)y \in C \Rightarrow x, y \in C (1 - \lambda) y \in C \Rightarrow x, y \in C$.

Thus $C$ is convex if $x + y \in C x + y \in C$.

In general, if $x_1, x_2 \in C$, then, $\lambda_1 x_1 + \lambda_2 x_2 \in C, \forall \lambda_1, \lambda_2 \geq 0 \lambda_1 x_1 + \lambda_2 x_2 \in C, \forall \lambda_1, \lambda_2 \geq 0$.

**Examples**

- The conic combination of infinite set of vectors in $\mathbb{R}^n$ is a convex cone.
- Any empty set is a convex cone.
- Any linear function is a convex cone.
- Since a hyperplane is linear, it is also a convex cone.
- Closed half spaces are also convex cones.

**Note** – The intersection of two convex cones is a convex cone but their union may or may not be a convex cone.
Convex Optimization - Polar Cone

Let $S$ be a non empty set in $\mathbb{R}^n \mathbb{R}^n$ Then, the polar cone of $S$ denoted by $S^*S^*$ is given by $S^*=\{p \in \mathbb{R}^n, ptx \leq 0 \forall x \in S\}S^*=\{p \in \mathbb{R}^n, pTx \leq 0 \forall x \in S\}$.

**Remark**

- Polar cone is always convex even if $S$ is not convex.
- If $S$ is empty set, $S^*=\mathbb{R}nS^*=\mathbb{R}n$.
- Polarity may be seen as a generalisation of orthogonality.

Let $C \subseteq \mathbb{R}nC \subseteq \mathbb{R}n$ then the orthogonal space of $C$, denoted by $C^\perp = \{y \in \mathbb{R}n: (x, y) = 0 \forall x \in C\}C^\perp = \{y \in \mathbb{R}n: (x, y) = 0 \forall x \in C\}$.

**Lemma**

Let $S, S^1S^1, S^2S^2$ be non empty sets in $\mathbb{R}n \mathbb{R}n$ then the following statements are true

- $S^*S^*$ is a closed convex cone.
- $S^1S^1 \subseteq S^*S^*$ where $S^*S^*$ is a polar cone of $S^*S^*$.
- $S^1S^1 \subseteq S^*S^* \subseteq S^1S^1 \subseteq S^*S^* \Rightarrow S^2S^2 \subseteq S^*S^*$.

**Proof**

**Step 1** - $S^*=\{p \in \mathbb{R}n, ptx \leq 0 \forall x \in S\}S^*=\{p \in \mathbb{R}n, pTx \leq 0 \forall x \in S\}$

- Let $x_1, x_2 \in S \Rightarrow xt_1x \leq 0 x_1, x_2 \in S \Rightarrow x_1Tx \leq 0$ and $xT_2x \leq 0, \forall x \in SxT_2x \leq 0, \forall x \in S$

  For $\lambda \in (0, 1), [\lambda x_1+(1-\lambda)x_2]x=[(\lambda x_1)+(1-\lambda)x_2]x, \forall x \in Sx \lambda \in (0, 1), [\lambda x_1+(1-\lambda)x_2]x=[(\lambda x_1)+(1-\lambda)x_2]x, \forall x \in S$

  $=[\lambda x_1x+(1-\lambda)x_2x]=\lambda x_1x+(1-\lambda)x_2x \leq 0=[\lambda x_1x+(1-\lambda)x_2x]x=\lambda x_1x+(1-\lambda)x_2x \leq 0$

  Thus $\lambda x_1+(1-\lambda)x_2 \in S \Rightarrow \lambda x_1+(1-\lambda)x_2 \in S$

  Therefore $S^*S^*$ is a convex set.

- For $\lambda \geq 0, ptx \leq 0, \forall x \in S, \lambda x_1 \geq 0, pT \leq 0, \forall x \in S$

  Therefore, $\lambda xT \leq 0, \lambda pT \leq 0, \forall x \in S$

  $\Rightarrow (\lambda p)x \leq 0 \Rightarrow (\lambda p)xT \leq 0$

  $\Rightarrow \lambda p \in S \Rightarrow \lambda p \in S$

  Thus, $S^*S^*$ is a cone.

  - To show $S^*S^*$ is closed, i.e., to show if $p_n \rightarrow ppn \rightarrow p$ as $n \rightarrow \infty n \rightarrow \infty$, then $p \in S^*S^*$

    $\forall x \in S, pTnx-pTx=(p_n-p)x \forall x, pTnx-pTx=(p_n-p)x$

    $\Rightarrow (p_n-p)x \rightarrow (p_n-p)x \Rightarrow (p_n-p)x \rightarrow (p_n-p)x$

    Therefore $pTnx-pTnpx \rightarrow pTnpx \rightarrow pTnpx \rightarrow pTnpx$. But $pTnx \leq 0, \forall x \in SpTnx \leq 0, \forall x \in S$

    Thus, $pTnx \leq 0, \forall x \in S$ $\Rightarrow p \in S^*S^*$ $\Rightarrow p \in S^*S^*$

    Hence, $S^*S^*$ is closed.
Step 2 – S* = {q ∈ R^n : qtTp ≤ 0, ∀p ∈ S \} S** = {q ∈ R^n : qtTp ≤ 0, ∀p ∈ S*}
Let x ∈ Sx ∈ S, then ∀p ∈ S*, ptx ≤ 0 ⇒ xtp ≤ 0 ⇒ x ∈ S*⇒ ∀p ∈ S*, pTx ≤ 0 ⇒ xtp ≤ 0 ⇒ x ∈ S**
Thus, S* ∈ S**

Step 3 – S*2 = {p ∈ R^n : ptx ≤ 0, ∀x ∈ S2} S*2 = {p ∈ R^n : ptx ≤ 0, ∀x ∈ S2}
Since S1 ⊆ S2 ⇒ ∀x ∈ S2 ⇒ ∀x ∈ S1 S1 ⊆ S2 ⇒ ∀x ∈ S2 ⇒ ∀x ∈ S1
Therefore, if p^T ∈ S*2, p^T ∈ S*2, then p^T ≤ 0, ∀x ∈ S2 ⇒ p^T ≤ 0, ∀x ∈ S1
⇒ p^T ∈ S*1 ⇒ p^T ∈ S1
⇒ S*2 ⊆ S*1 ⇒ S2* ⊆ S1*

Theorem
Let C be a non empty closed convex cone, then C = C** = C*

Proof
C = C** = C* by previous lemma.
To prove : x ∈ C** ⊆ Cx ∈ C* ⊆ C
Let x ∈ C** and let x ∈ Cx ∈ C
Then by fundamental separation theorem, there exists a vector p ≠ 0, p ≠ 0 and a scalar α such that
\[ p^T y ≤ α, \forall y ∈ CpTy ≤ α, \forall y ∈ C \]
Therefore, ptx > αptx > α
But since \( y = 0 \) ∈ C \( y = 0 \) ∈ C and p^T y ≤ α, \forall y ∈ C ⇒ α ≥ 0 pTy ≤ α, \forall y ∈ C ⇒ α ≥ 0 and p^T x > 0 p^T x > 0
If p ∈ C* p ∈ C*, then there exists some \( y^- \) ∈ C \( y^- \) ∈ C such that
\[ p^T y^- > 0 p^T y^- > 0 \] and p^T(λy^-)p^T(λy^-) can be made arbitrarily large by taking λ sufficiently large.
This contradicts with the fact that p^T y ≤ α, \forall y ∈ CpTy ≤ α, \forall y ∈ C
Therefore, p ∈ C* p ∈ C*
Since x ∈ C = {q : qtq ≤ 0, ∀p ∈ C*} x ∈ C = {q : qtq ≤ 0, ∀p ∈ C*}
Therefore, xtp ≤ 0 ⇒ ptx ≤ 0 xtp ≤ 0 ⇒ ptx ≤ 0
But p^T x > αptx > α
Thus is contardiction.

Thus, x ∈ Cx ∈ C
Hence C = C** = C*

Convex Optimization - Conic Combination

A point of the form \( \alpha_1x_1 + \alpha_2x_2 + \ldots + \alpha_nx_n + x_1 + \alpha_2x_2 + \ldots + \alpha_nx_n \) with \( \alpha_1, \alpha_2, \ldots, \alpha_n ≥ 0 \) \( \alpha_1, \alpha_2, \ldots, \alpha_n ≥ 0 \) is called conic combination of \( x_1, x_2, \ldots, x_n, x_1, x_2, \ldots, x_n \).
- If \( x_1, x_2, \ldots, x_n \) are in convex cone C, then every conic combination of \( x_1, x_2, \ldots, x_n \) is also in C.
- A set C is a convex cone if it contains all the conic combination of its elements.
Conic Hull
A conic hull is defined as a set of all conic combinations of a given set S and is denoted by coni(S).
Thus, \(\text{coni}(S) = \{\sum_{i=1}^{k} \lambda_i x_i : x_i \in S, \lambda_i \in \mathbb{R}, \lambda_i \geq 0, i = 1, 2, \ldots\}\) coni(S) = \(\{\sum_{i=1}^{k} \lambda_i x_i : x_i \in S, \lambda_i \in \mathbb{R}, \lambda_i \geq 0, i = 1, 2, \ldots\}\)
- The conic hull is a convex set.
- The origin always belong to the conic hull.

Convex Optimization - Polyhedral Set
A set in \(\mathbb{R}^n\) is said to be polyhedral if it is the intersection of a finite number of closed half spaces, i.e.,
\[ S = \{ x \in \mathbb{R}^n : p^T x \leq a_i, i = 1, 2, \ldots, n \} \]
S = \{ x \in \mathbb{R}^n : p^T x \leq a_i, i = 1, 2, \ldots, n \}
For example,
- \(\{ x \in \mathbb{R}^n : AX = b \}\) \(\{ x \in \mathbb{R}^n : AX = b \}\)
- \(\{ x \in \mathbb{R}^n : AX \leq b \}\) \(\{ x \in \mathbb{R}^n : AX \leq b \}\)
- \(\{ x \in \mathbb{R}^n : AX \geq b \}\) \(\{ x \in \mathbb{R}^n : AX \geq b \}\)

Polyhedral Cone
A set in \(\mathbb{R}^n\) is said to be polyhedral cone if it is the intersection of a finite number of half spaces that contain the origin, i.e.,
\[ S = \{ x \in \mathbb{R}^n : p^T x \leq 0, i = 1, 2, \ldots \} \]
S = \{ x \in \mathbb{R}^n : p^T x \leq 0, i = 1, 2, \ldots \}

Polytope
A polytope is a polyhedral set which is bounded.

Remarks
- A polytope is a convex hull of a finite set of points.
- A polyhedral cone is generated by a finite set of vectors.
- A polyhedral set is a closed set.
- A polyhedral set is a convex set.

Extreme point of a convex set
Let \(S\) be a convex set in \(\mathbb{R}^n\). A vector \(x \in S\) is said to be a extreme point of \(S\) if \(x = \lambda x_1 + (1-\lambda)x_2\) with \(x_1, x_2 \in S\) and \(\lambda \in (0, 1)\) \(\Rightarrow x = x_1 = x_2 \lambda \in (0, 1)\) \(\Rightarrow x = x_1 = x_2\).

Example
**Step 1** – \(S = \{ (x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \leq 1\} S = \{ (x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \leq 1\} \)
Extreme point, \(E = \{ (x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 = 1\} E = \{ (x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 = 1\} \)
Step 2 – \( S=\{(x_1,x_2)\in \mathbb{R}^2: x_1+x_2<2, -x_1+2x_2\leq 2, x_1,x_2\geq 0\} \)

Extreme points: \( E=\{(0,0),(2,0),(0,1),(23,43)\} \)

Step 3 – \( S \) is the polytope made by the points \( \{(0,0),(1,1),(1,3),(-2,4),(0,2)\};\{(0,0),(1,1),(1,3),(-2,4),(0,2)\} \)

Extreme point: \( E=\{(0,0),(1,1),(1,3),(-2,4)\} \)

Remarks

- Any point of the convex set \( S \), can be represented as a convex combination of its extreme points.
- It is only true for closed and bounded sets in \( \mathbb{R}^n \).
- It may not be true for unbounded sets.

**k extreme points**

A point in a convex set is called \( k \) extreme if and only if it is the interior point of a \( k \)-dimensional convex set within \( S \), and it is not an interior point of a \( (k+1) \)-dimensional convex set within \( S \). Basically, for a convex set \( S \), \( k \) extreme points make \( k \)-dimensional open faces.

**Convex Optimization - Direction**

Let \( S \) be a closed convex set in \( \mathbb{R}^n \). A non zero vector \( d\in \mathbb{R}^n \) is called a direction of \( S \) if for each \( x\in \mathbb{R}^n \), \( x+\lambda d\in \mathbb{R}^n \) for all \( \lambda \geq 0 \).

- Two directions \( d_1 \) and \( d_2 \) of \( S \) are called distinct if \( d\neq ad_2 \) for \( a>0 \).
- A direction \( dd \) of \( SS \) is said to be extreme direction if it cannot be written as a positive linear combination of two distinct directions, i.e., if \( d=\lambda_1 d_1 + \lambda_2 d_2 \) for \( \lambda_1, \lambda_2 > 0 \), then \( d=ad_1=ad_2 \) for some \( a \).
- Any other direction can be expressed as a positive combination of extreme directions.
- For a convex set \( SS \), the direction \( d \) such that \( x+\lambda d\in SS \) for some \( x\in SS \) and all \( \lambda \geq 0 \) is called *recessive* for \( SS \).
- Let \( E \) be the set of the points where a certain function \( f:S\rightarrow \mathbb{R} \) over a non-empty convex set \( S \) in \( \mathbb{R}^n \) attains its maximum, then \( EE \) is called exposed face of \( SS \). The directions of exposed faces are called exposed directions.
- A ray whose direction is an extreme direction is called an extreme ray.

**Example**

Consider the function \( f(x)=y=|x| \) if \( f(x)=y=|x| \), where \( x\in \mathbb{R}^n \). Let \( d \) be unit vector in \( \mathbb{R}^n \)

Then, \( d \) is the direction for the function \( f \) because for any \( \lambda \geq 0, x+\lambda d\in f(x) \lambda \geq 0, x+\lambda d\in f(x) \).
Convex and Concave Function

Let \( f:S\rightarrow R \), where \( S \) is non empty convex set in \( R^n \), then \( f(x) \) is said to be convex on \( S \) if \( f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2), \forall \lambda \in (0,1) \).

On the other hand, Let \( f:S\rightarrow R \), where \( S \) is non empty convex set in \( R^n \), then \( f(x) \) is said to be concave on \( S \) if \( f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda)f(x_2), \forall \lambda \in (0,1) \).

Let \( f:S\rightarrow R \) where \( S \) is non empty convex set in \( R^n \), then \( f(x) \) is said to be strictly convex on \( S \) if \( f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda)f(x_2), \forall \lambda \in (0,1) \).

Let \( f:S\rightarrow R \) where \( S \) is non empty convex set in \( R^n \), then \( f(x) \) is said to be strictly concave on \( S \) if \( f(\lambda x_1 + (1-\lambda)x_2) > \lambda f(x_1) + (1-\lambda)f(x_2), \forall \lambda \in (0,1) \).

Examples

- A linear function is both convex and concave.
- \( f(x) = |x| \) is a convex function.
- \( f(x) = 1x \) is a convex function.

Theorem

Let \( f_1,f_2,\ldots,f_k:R^n\rightarrow R \) be convex functions. Consider the function \( f(x) = \sum_{j=1}^{k} a_j f_j(x) \) where \( a_j > 0, j = 1, 2, \ldots, k \), then \( f(x) \) is a convex function.

Proof

Since \( f_1,f_2,\ldots,f_k \) are convex functions

Therefore, \( f_i(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f_i(x_1) + (1-\lambda)f_i(x_2), \forall \lambda \in (0,1) \).

Consider the function \( f(x) = \sum_{j=1}^{k} a_j f_j(x) \).

Therefore,

\[
\sum_{j=1}^{k} a_j f_j(\lambda x_1 + (1-\lambda)x_2) \leq \sum_{j=1}^{k} a_j \lambda f_j(x_1) + (1-\lambda)f_j(x_2) = \sum_{j=1}^{k} a_j (\lambda f_j(x_1) + (1-\lambda)f_j(x_2))
\]

\[
\sum_{j=1}^{k} a_j (\sum_{j=1}^{k} a_j f_j(x_1)) \leq \lambda (\sum_{j=1}^{k} a_j f_j(x_1)) + (1-\lambda)(\sum_{j=1}^{k} a_j f_j(x_2))
\]

THANKS FOR READING – VISIT OUR WEBSITE: www.educatererindia.com
\[ f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_2) \leq (1-\lambda)f(x_2) \Rightarrow f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(\lambda x_2) \leq (1-\lambda)f(\lambda x_2) \]

Hence, \( f(x)f(x) \) is a convex function.

**Theorem**

Let \( f(x)f(x) \) be a convex function on a convex set \( S \subset \mathbb{R}^n \) then a local minima of \( f(x)f(x) \) on \( S \) is a global minima.

**Proof**

Let \( x^\ast x^\ast \) be a local minima for \( f(x)f(x) \) and \( x^\ast x^\ast \) is not global minima.

therefore, \( \exists x^\ast x^\ast \in S \exists x^\ast x^\ast \in S \) such that \( f(\lambda x^\ast) < f(x^\ast)f(x^\ast) < f(x^\ast) \)

Since \( x^\ast x^\ast \) is a local minima, there exists neighbourhood \( N_\varepsilon(x^\ast)N_\varepsilon(x^\ast) \) such that \( f(x^\ast) \leq f(x), \forall x \in N_\varepsilon(x^\ast) \cap Sf(x^\ast) \leq f(x), \forall x \in N_\varepsilon(x^\ast) \cap S \)

But \( f(x)f(x) \) is a convex function on \( S \), therefore, for \( \lambda \in (0,1) \lambda \in (0,1) \)

we have \( \lambda x^\ast + (1-\lambda)x^\ast \leq \lambda f(x^\ast) + (1-\lambda)f(x^\ast) \lambda x^\ast + (1-\lambda)x^\ast \leq \lambda f(x^\ast) + (1-\lambda)f(x^\ast) \)

\[ \Rightarrow \lambda x^\ast + (1-\lambda)x^\ast \leq \lambda f(x^\ast) + (1-\lambda)f(x^\ast) \]

\[ \Rightarrow \lambda x^\ast + (1-\lambda)x^\ast \leq f(x^\ast), \forall x \in (0,1) \Rightarrow \lambda x^\ast + (1-\lambda)x^\ast \leq f(x^\ast), \forall x \in (0,1) \]

But for some \( \lambda < 1 \lambda < 1 \) but close to 1, we have

\[ \lambda x^\ast + (1-\lambda)x^\ast \in N_\varepsilon(x^\ast) \cap S \lambda x^\ast + (1-\lambda)x^\ast \in N_\varepsilon(x^\ast) \cap S \] and \( f(\lambda x^\ast + (1-\lambda)x^\ast) < f(x^\ast) f(\lambda x^\ast + (1-\lambda)x^\ast) < f(x^\ast) \)

which is a contradiction.

Hence, \( x^\ast x^\ast \) is a global minima.

**Epigraph**

let \( S \) be a non-empty subset in \( \mathbb{R}^n \mathbb{R}^n \) and let \( f:S \rightarrow \mathbb{R}f:S \rightarrow \mathbb{R} \) then the epigraph of \( f \) denoted by \( \text{epi}(f) \) \( \text{epi}(f) \) is a subset of \( \mathbb{R}^{n+1} \mathbb{R}^{n+1} \) defined by \( \text{Ei} = \{(x,\alpha) : x \in \mathbb{R}^n, \alpha \in \mathbb{R}, f(x) \leq \alpha \} \)

**Hypograph**

let \( S \) be a non-empty subset in \( \mathbb{R}^n \mathbb{R}^n \) and let \( f:S \rightarrow \mathbb{R}f:S \rightarrow \mathbb{R} \), then the hypograph of \( f \) denoted by \( \text{hypi}(f) \) or \( \text{Hf} = \{(x,\alpha) : x \in \mathbb{R}^n, \alpha \in \mathbb{R}, f(x) \geq \alpha \} \)

**Theorem**

Let \( S \) be a non-empty convex set in \( \mathbb{R}^n \mathbb{R}^n \) and let \( f:S \rightarrow \mathbb{R}f:S \rightarrow \mathbb{R} \), then \( f \) is convex if and only if its epigraph \( \text{Ei} \) \( \text{Ei} \) is a convex set.

**Proof**

Let \( f \) is a convex function.

To show \( \text{Ei} \) \( \text{Ei} \) is a convex set.

Let \( (x_1,\alpha_1), (x_2,\alpha_2) \in \text{Ei} \)

To show \( \lambda(x_1,\alpha_1) + (1-\lambda)(x_2,\alpha_2) \in \text{Ei} \)

\[ \Rightarrow [\lambda x_1 + (1-\lambda)x_2, \lambda \alpha_1 + (1-\lambda)\alpha_2] \in \text{Ei} \]

\[ f(x_1) \leq \alpha_1, f(x_2) \leq \alpha_2 \]
Therefore, \( f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \)
\( \Rightarrow f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda a_1 + (1-\lambda)a_2 = f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda a_1 + (1-\lambda)a_2 \)

**Converse**

Let \( \mathbb{E}f \) is a convex set.

To show \( f \) is convex.

i.e., to show if \( x_1, x_2 \in S, \lambda \in (0, 1) \),
\( f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \)

Let \( x_1, x_2 \in S, \lambda \in (0, 1), f(x_1), f(x_2) \in \mathbb{R}, x_2 \in S, \lambda \in (0, 1), f(x_1), f(x_2) \in \mathbb{R} \)

Since \( \mathbb{E}f \) is a convex set, \( (\lambda x_1 + (1-\lambda)x_2, \lambda f(x_1) + (1-\lambda)f(x_2)) \in \mathbb{E}f \)

Therefore, \( f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \)

**Convex Optimization - Jensen's Inequality**

Let \( S \) be a non-empty convex set in \( \mathbb{R}^n \) and \( f:S \rightarrow \mathbb{R}^n \rightarrow \mathbb{R}^n \). Then \( f \) is convex if and only if for each integer \( k > 0 \) \( k > 0 \)

\( x_1, x_2, \ldots, x_k \in S, \sum_{i=1}^{k} \lambda_i x_i = 1, \lambda_i \geq 0, \forall i = 1, 2, s, k, x_1, x_2, \ldots, x_k \in S, \sum_{i=1}^{k} \lambda_i x_i = 1, \lambda_i \geq 0, \forall i = 1, 2, s, k, \)

we have \( f(\sum_{i=1}^{k} \lambda_i x_i) \leq \sum_{i=1}^{k} \lambda_i f(x_i) = f(\sum_{i=1}^{k} \lambda_i x_i) \)

**Proof**

By induction on \( k \).

\( k=1: x \in S \) Therefore \( f(\lambda_1 x_1) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) \) because \( \lambda_1 = 1, \lambda_2 = 1 \).

\( k=2: \lambda_1 + \lambda_2 = 2: \lambda_1 + \lambda_2 = 1 \) and \( x_1, x_2 \in S \)

Therefore, \( \lambda_1 x_1 + \lambda_2 x_2 \in S, \lambda_1 x_1 + \lambda_2 x_2 \in S \)

Hence by definition, \( f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) \frac{f(\lambda_1 x_1 + \lambda_2 x_2)}{\lambda_1 f(x_1)} \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) \)

Let the statement is true for \( n < k \)

Therefore,

\[ f(\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \ldots + \lambda_k f(x_k) \]

\( k=n+1: k=n+1 \) Let \( x_1, x_2, \ldots, x_n, x_{n+1} \in S, \sum_{i=1}^{n+1} x_i = 1, \sum_{i=1}^{n+1} \lambda_i x_i = 1 \)

Therefore \( \mu x_1 + \mu x_2 + \ldots \mu_{n+1} x_{n+1} \in S, \mu x_1 + \mu x_2 + \ldots \mu_{n+1} x_{n+1} \in S \)

thus, \( f(\mu x_1 + \mu x_2 + \ldots + \mu_{n+1} x_{n+1}) f(\mu x_1 + \mu x_2 + \ldots + \mu_{n+1} x_{n+1}) \)

\( = f(\mu_{n+1} + \mu_{n+2} + \ldots + \mu_{n+1} x_{n+1}) = f(\mu_{n+2} + \ldots + \mu_{n+1} x_{n+1}) = f(\mu_{n+1} + \mu_{n+2} + \ldots + \mu_{n+1} x_{n+1}) \)

where \( \mu = 1, \mu_{n+1} = 1, y \in S, \mu = 1, \mu_{n+1} = 1, y \in S \)
\( f(\mu_1 x_1 + \mu_2 x_2 + \ldots + \mu_n x_n + \mu_1 x_{n+1}) \leq \mu f(y) + \mu_n + 1 f(x_{n+1}) \Rightarrow f(\mu_1 x_1 + \mu_2 x_2 + \ldots + \mu_n x_n + \mu_1 x_{n+1}) \leq f(\mu_1 x_1 + \mu_2 x_2 + \ldots + \mu_n x_n + \mu_1 x_{n+1}) \)
\( \Rightarrow f(\mu_1 x_1 + \mu_2 x_2 + \ldots + \mu_n x_n + \mu_1 x_{n+1}) \leq f(\mu_1 x_1 + \mu_2 x_2 + \ldots + \mu_n x_n + \mu_1 x_{n+1}) \leq (\mu_1 + \mu_2 + \ldots + \mu_n) f(x_1) + \mu_n + 1 f(x_{n+1}) \)
\( \Rightarrow f(\mu_1 x_1 + \mu_2 x_2 + \ldots + \mu_n x_n + \mu_1 x_{n+1}) \leq (\mu_1 + \mu_2 + \ldots + \mu_n) f(x_1) + \mu_n + 1 f(x_{n+1}) \)

Hence Proved.

**Convex Optimization - Differentiable Function**

Let \( S \) be a non-empty open set in \( \mathbb{R}^n \), then \( f:S \rightarrow \mathbb{R} \) is said to be differentiable at \( x^* \in S \) if there exist a vector \( \nabla f(x^*) \) called gradient vector and a function \( \alpha:S \rightarrow \mathbb{R}^n \) such that

\[ f(x) = f(x^*) + \nabla f(x^*)(x-x^*) + \|x-x^*\| \alpha(x^*, x-x^*), \forall x \in S \text{ if } f(x^*) = f(x) + \nabla f(x^*)^T(x-x^*) \]

\[ \alpha(x^*, x-x^*) \rightarrow 0 \text{ as } x \rightarrow x^* \]

Theorem

Let \( S \) be a non-empty, open convexset in \( \mathbb{R}^n \) and let \( f:S \rightarrow \mathbb{R} \) be differentiable on \( S \). Then, \( f \) is convex if and only if for \( x_1, x_2 \in S \),

\[ \nabla f(x_2)T(x_1-x_2) \leq f(x_1) - f(x_2) \]

**Proof**

Let \( f \) be a convex function. i.e., for \( x_1, x_2 \in S \), \( \lambda \in (0,1) \),

\[ f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \]

\[ \Rightarrow f(x_1) - f(x_2) \leq f(\lambda x_1 + (1-\lambda) x_2) - f(x_2) = f(\lambda x_1 + (1-\lambda) x_2) \leq (\lambda x_1 + (1-\lambda) x_2) f(x_1) - f(x_2) \]

Converse

Let for \( x_1, x_2 \in S \),

\[ \nabla f(x_2)T(x_1-x_2) \leq f(x_1) - f(x_2) \]

\[ \Rightarrow f(x_1) - f(x_2) \geq \nabla f(x_2)T(x_1-x_2) \]
To show that \( f \) is convex.

Since \( S \) is convex, \( x_3 = \lambda x_1 + (1-\lambda)x_2 \in S, \lambda \in (0,1) \)
\( x_3 = \lambda x_1 + (1-\lambda)x_2 \in S, \lambda \in (0,1) \)
Since \( x_1, x_2 \in S \), we have
\[
f(x_1) - f(x_2) \geq \nabla f(x_2) \cdot (x_1 - x_2) \]
\[
f(x_1) - f(x_2) \geq \nabla f(x_2) \cdot (x_1 - x_2) \]
Thus, combining the above equations, we get
\[
\lambda (f(x_1) - f(x_3)) + (1-\lambda)(f(x_2) - f(x_3)) \geq 0 \]
\[
\Rightarrow f(x_3) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \]

**Theorem**

Let \( S \) be a non-empty open convex set in \( \mathbb{R}^n \) and let \( f:S \rightarrow \mathbb{R} \) be differentiable on \( S \),
Then \( f \) is convex on \( S \) if and only if for any \( x_1, x_2 \in S \),
\( \nabla f(x_2) - \nabla f(x_1) \) is non-negative.

**Proof**

Let \( f \) be a convex function, then using the previous theorem
\[
\nabla f(x_2) \cdot (x_1 - x_2) \leq f(x_1) - f(x_2) \]
\[
\nabla f(x_1) \cdot (x_2 - x_1) \leq f(x_2) - f(x_1) \]
Adding the above two equations, we get
\[
\nabla f(x_2) \cdot (x_1 - x_2) + \nabla f(x_1) \cdot (x_2 - x_1) \leq 0 \]
\[
\Rightarrow (\nabla f(x_2) - \nabla f(x_1)) \cdot (x_1 - x_2) \leq 0 \]
\[
\Rightarrow (\nabla f(x_2) - \nabla f(x_1)) \cdot (x_2 - x_1) \leq 0 \]

**Converse**

Let for any \( x_1, x_2 \in S \),
\( \nabla f(x_2) - \nabla f(x_1) \) is non-negative.

To show that \( f \) is convex.

Let \( x_1, x_2 \in S \), thus by mean value theorem, \( f(x) - f(x_2) \) is convex on \( S \).
\[
f(x) \in (x_1 - x_2) \Rightarrow x = \lambda x_1 + (1-\lambda)x_2 \]
\[
f(x) \in (x_1 - x_2) \Rightarrow x = \lambda x_1 + (1-\lambda)x_2 \]
for \( x, x_1, x_2 \), we know
\[
(\nabla f(x) - \nabla f(x_1)) \cdot (x_1 - x_2) \geq 0 \]
\[
(\nabla f(x) - \nabla f(x_1)) \cdot (x_1 - x_2) \geq 0 \]

THANKS FOR READING – VISIT OUR WEBSITE: www.educatererindia.com
\[ (\nabla f(x) - \nabla f(x_1)) + (1-\lambda)(x-x_1) \geq 0 \Rightarrow (\nabla f(x) - \nabla f(x_1))T(1-\lambda)(x-x_1) \geq 0 \]
\[ \Rightarrow \nabla f(x)T(x-x_1) \geq \nabla f(x_1)T(x-x_1) \]

Combining the above equations, we get
\[ \Rightarrow \nabla f(x_1)T(x-x_1) \leq f(x_2) - f(x_1) \Rightarrow \nabla f(x_1)T(x-x_1) \leq f(x_2) - f(x_1) \]

Hence using the last theorem, f is a convex function.

**Twice Differentiable function**

Let S be a non-empty subset of \( \mathbb{R}^n \) and let \( f:S \rightarrow \mathbb{R}^n \rightarrow \mathbb{R} \) then f is said to be twice differentiable at \( x^- \in S \) if there exists a vector \( \nabla f(x^-) \), an \( n \times n \) matrix \( H(x) \) called the Hessian matrix and a function \( \alpha: \mathbb{R}^n \rightarrow \mathbb{R}^n \) such that
\[ f(x) = f(x^-) + \alpha(x-x^-) \]
where \( \alpha(x^-) \rightarrow OaSx \rightarrow x^- \alpha(x^-) \rightarrow OaSx \rightarrow x^- \)

**Sufficient & Necessary Conditions for Global Optima**

**Theorem**

Let f be twice differentiable function. If \( x^- \) is a local minima, then \( \nabla f(x^-) = 0, \nabla f(x^-) = 0 \) and the Hessian matrix \( H(x^-)H(x^-) \) is a positive semidefinite.

**Proof**

Let \( d \in \mathbb{R}^n \). Since f is twice differentiable at \( x^- \).

Therefore,
\[ f(x^+ + \lambda d) = f(x^-) + \lambda \nabla f(x^-)Td + \lambda^2 dT H(x^-) d + \lambda dT(H(x^-)) d + f(x^- + \lambda d) = f(x^-) + \lambda \nabla f(x^-)Td + \lambda^2 dT(H(x^-)) d + \lambda dT(H(x^-)) d \]

But \( \nabla f(x^-) = 0, \nabla f(x^-) = 0 \) and \( \beta(x^- + \lambda d) \rightarrow 0, \beta(x^- + \lambda d) \rightarrow 0 \) as \( \lambda \rightarrow 0 \)
\[ \Rightarrow f(x^- + \lambda d) \leq f(x^-) \]

Since \( x^- \) is a local minima, there exists a \( \delta > 0 \) such that
\[ f(x) \leq f(x^- + \lambda d), \forall \lambda \in (0, \delta) \]

**Theorem**

Let \( f:S \rightarrow \mathbb{R}^n \) such that \( S \subset \mathbb{R}^n \) be twice differentiable over S. If \( \nabla f(x) = 0, \nabla f(x) = 0 \) and \( H(x^-)H(x^-) \) is positive semi-definite, for all \( x \in S x \in S \), then \( x^- \) is a global optimal solution.

**Proof**

Since \( H(x^-)H(x^-) \) is positive semi-definite, f is convex function over S. Since f is differentiable and convex at \( x^- \)
\[ \nabla f(x^-) + \alpha(x^- - x^-) + \|x^- - x^-\| \alpha(x^- - x^-) = f(x^-) + \alpha(x^- - x^-) \]

Since \( \nabla f(x^-) = 0 \), if \( f(x^-) = f(x^-) \)

Hence, \( x^- x^- \) is a global optima.

**Theorem**

Suppose \( x^- \in S \) is a local optimal solution to the problem \( f:S \rightarrow R \) where \( S \) is a non-empty subset of \( R^n \) and \( f \) is convex. \( \min f(x) = \min f(x) \) where \( x \in S \).

Then:

- \( x^- x^- \) is a global optimal solution.
- If either \( x^- x^- \) is strictly local minima or \( f \) is strictly convex function, then \( x^- x^- \) is the unique global optimal solution and is also strong local minima.

**Proof**

Let \( x^- x^- \) be another global optimal solution to the problem such that \( x^- \neq x^- x^- \) and \( f(x^-) = f(x^-) \).

Since \( x^- x^- \in S \) and \( S \) is convex, then \( x^- + x^- \in S \) and \( f(x^-) = f(x^-) \).

\[ \Rightarrow f(x^- + x^-) < 12f(x^-) + 12f(x^-) = f(x^-) \Rightarrow f(x^- + x^-) < 12f(x^-) + 12f(x^-) = f(x^-) \]

This is contradiction.

Hence, \( x^- x^- \) is a unique global optimal solution.

**Corollary**

Let \( f:S \subseteq R^n \rightarrow R \) be a differentiable convex function where \( \phi \neq S \subseteq R^n \) is a convex set. Consider the problem \( \min f(x), x \in S \) \( \min f(x), x \in S \), then \( x^- x^- \) is an optimal solution if \( \nabla f(x^-) = 0 \), \( \forall x \in S \).

**Proof**

Let \( x^- x^- \) be an optimal solution, i.e., \( f(x^-) = f(x), \forall x \in S \) \( f(x^-) = f(x) \), \( \forall x \in S \).

\[ \Rightarrow f(x^-) = f(x^-) \geq 0 \Rightarrow f(x^-) = f(x^-) \geq 0 \]

\[ f(x^-) = f(x^-) + \nabla f(x^-) + \|x^- - x^-\| f(x^-) = f(x^-) + \nabla f(x^-) + \|x^- - x^-\| f(x^-) \]

where \( \alpha(x^- - x^-) = 0 \), \( \alpha(x^- - x^-) = 0 \) as \( x \rightarrow x^- x^- \)

\[ \Rightarrow f(x^-) = f(x^-) + \nabla f(x^-) + \|x^- - x^-\| f(x^-) \geq 0 \Rightarrow f(x^-) = f(x^-) + \nabla f(x^-) + \|x^- - x^-\| f(x^-) \geq 0 \]

**Corollary**

Let \( f \) be a differentiable convex function at \( x^- x^- \), then \( x^- x^- \) is global minimum if \( \nabla f(x^-) = 0 \), \( \nabla f(x^-) = 0 \).
\[ f(±1)=0, f(0)=-1 \forall x \in \mathbb{R}, \forall x \in \mathbb{R} \]

Hence, \( f(x) = -1 \) if \( f(0) = -1 \) for \( x \neq 0 \).

- \( f(x) = x \log f(x) \) for \( x > 0 \) defined on \( S = \{x \in \mathbb{R}, x > 0\} \).

Thus, this function is strictly convex.

- \( f(x) = ex, x \in R \), \( f(x) = ex, x \in R \) is strictly convex.

**Quasiconvex and Quasiconcave functions**

Let \( f : S \rightarrow R \), \( f : S \rightarrow R \) where \( S \subseteq R^n \subseteq R^n \) is a non-empty convex set. The function \( f \) is said to be quasiconvex if for each \( x_1, x_2 \in S \), \( x_1, x_2 \in S \), we have \( f(\lambda x_1 + (1-\lambda)x_2) \leq \max\{f(x_1), f(x_2)\}, \lambda \in (0,1) \) for each \( \lambda \) and \( \lambda \in (0,1) \).

For example, \( f(x) = x^3f(x) = x^3 \)

Let \( f : S \rightarrow R \), \( f : S \rightarrow R \) where \( S \subseteq R^n \subseteq R^n \) is a non-empty convex set. The function \( f \) is said to be quasiconcave if for each \( x_1, x_2 \in S \), \( x_1, x_2 \in S \), we have \( f(\lambda x_1 + (1-\lambda)x_2) \geq \min\{f(x_1), f(x_2)\}, \lambda \in (0,1) \) for each \( \lambda \) and \( \lambda \in (0,1) \).

**Remarks**

- Every convex function is quasiconvex but the converse is not true.
- A function which is both quasiconvex and quasiconcave is called quasimonotone.

**Theorem**

Let \( f : S \rightarrow R \), \( f : S \rightarrow R \) and \( S \) is a non-empty convex set in \( R^n \). The function \( f \) is quasiconvex if and only if \( S_A = \{x \in S : f(x) \leq a\} \subseteq S_A = \{x \in S : f(x) \leq a\} \) is convex for each real number \( \alpha \).

**Proof**

Let \( f \) is quasiconvex on \( S \).

Let \( x_1, x_2 \in S \), \( x_1, x_2 \in S \) therefore \( x_1, x_2 \in S \), \( x_2 \in S \), \( x_2 \in S \) and \( \max\{f(x_1), f(x_2)\} \leq \alpha \) for each \( \alpha \) and \( \alpha \) for each \( \alpha \).

Let \( \lambda \in (0,1) \) \( \lambda \in (0,1) \) and let \( x = \lambda x_1 + (1-\lambda)x_2 \leq \max\{f(x_1), f(x_2)\} \Rightarrow x \in S \), \( x \in S \), \( x \in S \), \( x \in S \) and \( \lambda \in (0,1) \) \( \lambda \in (0,1) \).

Thus, \( f(\lambda x_1 + (1-\lambda)x_2) \geq \max\{f(x_1), f(x_2)\} \leq \alpha f(\lambda x_1 + (1-\lambda)x_2) \leq \max\{f(x_1), f(x_2)\} \leq \alpha \)

Therefore, \( S_A \subseteq S_A \) is convex.

**Converse**

Let \( S \subseteq S \subseteq S \) for each \( \alpha \).

For \( \alpha \subseteq S \subseteq S \), \( \alpha \subseteq S \subseteq S \), \( \alpha \subseteq S \subseteq S \), \( \lambda \subseteq (0,1) \), \( \lambda \subseteq (0,1) \), \( \lambda \subseteq (0,1) \), \( x = \lambda x_1 + (1-\lambda)x_2 \leq \lambda x_1 + (1-\lambda)x_2 \)

Let \( x = \lambda x_1 + (1-\lambda)x_2 \leq \lambda x_1 + (1-\lambda)x_2 \)

For \( x_1, x_2 \in S \), \( x_1, x_2 \in S \), \( x_1, x_2 \in S \), \( \alpha = \max\{f(x_1), f(x_2)\} \leq \alpha = \max\{f(x_1), f(x_2)\} \leq \alpha = \max\{f(x_1), f(x_2)\}

THANKS FOR READING – VISIT OUR WEBSITE: www.educatenetindia.com
Proof by contradiction

Therefore, \( f(x) \leq f(x_1) \).

But since \( f \) is quasiconvex,

\[
f(x') \leq f(x_1) \leq f(x_2) \leq f(x')
\]

Hence proved.

**Theorem**

Let \( f: S \to \mathbb{R} \) be quasiconvex and \( S \) be a non-empty convex set in \( \mathbb{R}^n \). The function \( f \) is quasiconvex if and only if \( S_\alpha = \{ x \in S : f(x) \geq \alpha \} \) is convex for each real number \( \alpha \).

**Theorem**

Let \( f: S \to \mathbb{R} \) be quasiconvex and \( S \) be a non-empty convex set in \( \mathbb{R}^n \). The function \( f \) is quasiconvex if and only if \( S_\alpha = \{ x \in S : f(x) = \alpha \} \) is convex for each real number \( \alpha \).

**Differentiable Quasiconvex Function**

**Theorem**

Let \( S \) be a non-empty convex set in \( \mathbb{R}^n \) and \( f: S \to \mathbb{R} \) be differentiable on \( S \), then \( f \) is quasiconvex if and only if for any \( x_1, x_2 \in S \), \( x_2 \in S \) and \( f(x_1) \leq f(x_2) \), we have \( \nabla f(x_2) \leq \nabla f(x_1) \) and \( \nabla f(x_2) \leq \nabla f(x_1) \).

**Proof**

Let \( f \) be a quasiconvex function.

Let \( x_1, x_2 \in S \) such that \( f(x_1) \leq f(x_2) \).

By differentiability of \( f \) at \( x_2 \),

\[
f(\lambda x_1 + (1 - \lambda) x_2) - f(x_2) = \frac{d}{d\lambda} f(\lambda x_1 + (1 - \lambda) x_2) \bigg|_{\lambda = 1}
\]

But since \( f \) is quasiconvex,

\[
\nabla f(x) \leq \nabla f(x')
\]

Therefore, \( \nabla f(x_2) \leq \nabla f(x_1) \).

**Converse**

Let for \( x_1, x_2 \in S \) and \( f(x_1) \leq f(x_2) \), \( \nabla f(x_2) \leq \nabla f(x_1) \), \( \nabla f(x_2) \leq \nabla f(x_1) \).

To show that \( f \) is quasiconvex, ie, \( f(\lambda x_1 + (1 - \lambda) x_2) \leq \min \{ f(x_1), f(x_2) \} \).

**Proof by contradiction**

Suppose there exists an \( x_3 = \lambda x_1 + (1 - \lambda)x_2 \) such that \( f(x_2) < f(x_3) \).
For $x_2x_2$ and $x_3$, $\nabla f(x_3)T(x_2-x_3)\leq 0 \Rightarrow \nabla f(x_3)T(x_2-x_3)\leq 0$.

$\Rightarrow \nabla f(x_3)T(x_1-x_2)\geq 0 \Rightarrow \nabla f(x_3)T(x_1-x_2)\geq 0$. 

Thus, the above equations, $\nabla f(x_3)T(x_1-x_2)=0 \Rightarrow f(x_2)=f(x_3)$.

Define $U=\{x:f(x)\leq f(x_2), x=\mu x_2+(1-\mu)x_3, \mu \in (0,1)\} \subseteq \Omega_2$. 

Thus, we can find $x_0 \in U \times x_0 \in U$ such that $x_0=x_0x_2=\mu_1x_2+(1-\mu)x_3$ for some $\mu_0 \in (0,1) \mu_0 \in (0,1)$ which is nearest to $x_3x_3$ and $x^\in \Omega_2x_1x^\in \Omega_2x_1, x_1 \in (0,1)$ such that by mean value theorem, 

$f(x_3)-f(x_0)=\nabla f(x_3)x_3-x_0=\nabla f(x_3)x_3-x_0=\nabla f(x_3)$.

$\Rightarrow f(x_3)=f(x_0)+\nabla f(x_3)x_3-x_0=\nabla f(x_3)x_3-x_0=\nabla f(x_3)$.

$\Rightarrow f(x_3)=f(x_0)+\nabla f(x_3)x_3-x_0=\nabla f(x_3)x_3-x_0=\nabla f(x_3)$.

Since $x_0x_0$ is a combination of $x_1x_1$ and $x_2x_2$ and $f(x_2)<f(x_2)f(x_2)<f(x_2)$.

By repeating the starting procedure, $\nabla f(x_3)=f(x_3)x_3-x_2=0 \Rightarrow f(x_3)x_3-x_2=0$.

Thus, combining the above equations, we get:

$f(x_3)=f(x_0)+f(x_2)f(x_3)=f(x_0)+f(x_2)$.

Hence, it is a contradiction.

**Examples**

**Step 1** - $f(x)= x_3x_3(x_3)\leq x_3$.

Let $f(x_1)=f(x_2)\leq f(x_2)$

$\Rightarrow x_1\leq x_3 \Rightarrow x_1\leq x_2 \Rightarrow x_1\leq x_2 \Rightarrow x_1\leq x_2$.

$\nabla f(x_2)(x_1-x_2)=3x_2(3x_1-x_2)\leq 0 \Rightarrow f(x_2)(x_1-x_2)=3x_2(3x_1-x_2)\leq 0$.

Thus, $f(x) \in quasiconvex$.

**Step 2** - $f(x)=x_3x_3+x_32x_3(x_3)\leq x_3+\mu_3$.

Let $x_1^0=(2,2) \in (2,2) \leq x_2^0=(1,0) \in (0,1)$.

thus, $f(x_1^0)=0, f(x_2^0)=1 \Rightarrow f(x_1^0) \leq f(x_2^0)$.

Thus, $\nabla f(x_2^0)x_2(3,0)=3 \Rightarrow \nabla f(x_2)(3,0)T(3,0)=3 \Rightarrow \nabla f(x_2)(3,0)T(3,0)=3$.

Hence $f(x) \in quasiconvex$.

**Strictly Quasiconvex Function**
Let \( f: S \rightarrow \mathbb{R}^n : S \rightarrow \mathbb{R}^n \) and \( S \) be a non-empty convex set in \( \mathbb{R}^n \). Then, \( f \) is said to be strictly quasiconvex function if for each \( x_1, x_2 \in S \) with \( f(x_1) \neq f(x_2) \), we have \( f(\lambda x_1 + (1-\lambda)x_2) \leq \max\{ f(x_1), f(x_2) \} \) for \( \lambda \in [0,1) \).

**Remarks**

- Every strictly quasiconvex function is strictly convex.
- Strictly quasiconvex function does not imply quasiconvexity.
- Strictly quasiconvex function may not be strongly quasiconvex.
- Pseudoconvex function is a strictly quasiconvex function.

**Theorem**

Let \( f: S \rightarrow \mathbb{R}^n : S \rightarrow \mathbb{R}^n \) be strictly quasiconvex function and \( S \) be a non-empty convex set in \( \mathbb{R}^n \). Consider the problem: \( \min f(x), x \in S \min f(x), x \in S \). If \( x^* \) is local optimal solution, then \( x^* \) is global optimal solution.

**Proof**

Let there exists \( x^- \in S \) such that \( f(x^-) \leq f(x^\lambda) \leq f(x) \).

Since \( x^-, x^\lambda \in S \), and \( S \) is convex set, therefore, \( \lambda x^- + (1-\lambda)x^\lambda \in S \) for all \( \lambda \in [0,1) \).

Since \( x^\lambda \) is local minima, \( f(\lambda x^- + (1-\lambda)x^\lambda) \leq f(\lambda x^- + (1-\lambda)x^\lambda) \) for all \( \lambda \in (0,\delta) \).

Since \( f \) is strictly quasiconvex, \( f(\lambda x^- + (1-\lambda)x^\lambda) \leq \max\{ f(x^-), f(x^-) \} = f(x^\lambda) = f(\lambda x^- + (1-\lambda)x^\lambda) \).

Hence, it is contradiction.

**Strictly quasiconvex function**

Let \( f: S \rightarrow \mathbb{R}^n : S \rightarrow \mathbb{R}^n \) and \( S \) be a non-empty convex set in \( \mathbb{R}^n \), then \( f \) is said to be strictly quasiconvex function if for each \( x_1, x_2 \in \mathbb{R}^n \) with \( f(x_1) \neq f(x_2) \), we have

\[
f(\lambda x_1 + (1-\lambda)x_2) > \min\{ f(x_1), f(x_2) \} = f(\lambda x_1 + (1-\lambda)x_2) > \min\{ f(x_1), f(x_2) \}
\]

**Examples**

- \( f(x) = x^2 - 2f(x) = x^2 - 2 \)

  It is a strictly quasiconvex function because if we take any two points \( x_1, x_2 \in \mathbb{R}^n \) in the domain that satisfy the constraints in the definition \( f(\lambda x_1 + (1-\lambda)x_2) < \max\{ f(x_1), f(x_2) \} = f(\lambda x_1 + (1-\lambda)x_2) < \max\{ f(x_1), f(x_2) \} \). As the function is decreasing in the negative \( x \)-axis and is increasing in the positive \( x \)-axis (since it is a parabola).
• $f(x) = -x^2f(x) = -x^2$

It is not a strictly quasiconvex function because if we take $x_1 = 1, x_1 = 1$ and $x_2 = -1, x_2 = -1$ and $\lambda = 0.5, \lambda = 0.5$, then $f(x_1) = -1 = f(x_2)f(x_1) = -1 = f(x_2)$ but $f(\lambda x_1 + (1 - \lambda)x_2) = 0 = f(\lambda x_1 + (1 - \lambda)x_2)$ Therefore it does not satisfy the conditions stated in the definition. But it is a quasiconcave function because if we take any two points in the domain that satisfy the constraints in the definition $f(\lambda x_1 + (1 - \lambda)x_2) > \min\{f(x_1), f(x_2)\}, \forall \lambda \in (0, 1)$.

Strongly Quasiconvex Function

Let $f : S \to \mathbb{R}^n : S \to \mathbb{R}^n$ and $S$ be a non-empty convex set in $\mathbb{R}^n$ then $f$ is strongly quasiconvex function if for any $x_1, x_2 \in S$ with $x_1 \neq x_2, x_2 \neq x_2$, we have $f(\lambda x_1 + (1 - \lambda)x_2) = \max\{f(x_1), f(x_2)\}, \forall \lambda \in (0, 1) f(\lambda x_1 + (1 - \lambda)x_2) = \max\{f(x_1), f(x_2)\}, \forall \lambda \in (0, 1)$

Theorem

A quasiconvex function $f : S \to \mathbb{R}^n : S \to \mathbb{R}^n$ on a non-empty convex set $S$ in $\mathbb{R}^n$ is strongly quasiconvex function if it is not constant on a line segment joining any points of $S$.

Proof

Let $f$ is quasiconvex function and it is not constant on a line segment joining any points of $S$.

Suppose $f$ is not strongly quasiconvex function.

There exist $x_1, x_2 \in S$ with $x_1 \neq x_2, x_2 \neq x_2$ such that $f(z) = \max\{f(x_1), f(x_2)\}, \forall z = \lambda x_1 + (1 - \lambda)x_2, \lambda \in (0, 1) f(z) = \max\{f(x_1), f(x_2)\}, \forall z = \lambda x_1 + (1 - \lambda)x_2, \lambda \in (0, 1)$

$\Rightarrow f(x_1) \leq f(z) \Rightarrow f(x_1) \leq f(z)$ and $f(x_2) \leq f(z) f(x_2) \leq f(z)$

Since $f$ is not constant in $[x_1, z] \in [x_1, z]$ and $[z, x_2] \in [z, x_2]$.

So there exists $u \in[x_1, z] u \in[x_1, z]$ and $v = [z, x_2] v = [z, x_2]$

$\Rightarrow u = \mu_1 x_1 + (1 - \mu_1)z, v = \mu_2 z + (1 - \mu_2)x_2 \Rightarrow u = \mu_1 x_1 + (1 - \mu_1)z, v = \mu_2 z + (1 - \mu_2)x_2$

Since $f$ is quasiconvex,

$\Rightarrow f(u) \leq \max\{f(x_1), f(z)\} = f(z) \text{ and } f(v) \leq \max\{f(z), f(x_2)\} \Rightarrow f(u) \leq \max\{f(x_1), f(z)\} = f(z) \text{ and } f(v) \leq \max\{f(x_2), f(z)\}$

$\Rightarrow f(u) \leq f(z) \text{ and } f(v) \leq f(z) \Rightarrow f(u) \leq f(z) \text{ and } f(v) \leq f(z)$

$\Rightarrow \max\{f(u), f(v)\} \leq f(z) \Rightarrow \max\{f(u), f(v)\} \leq f(z)$

But $z$ is any point between $u$ and $v$, if any of them are equal, then $f$ is constant.

Therefore, $\max\{f(u), f(v)\} \leq f(z) \text{ and } f(u, f(v)) \leq f(z)$
which contradicts the quasiconvexity of \( f \) as \( z \in [u,v]z \in [u,v] \).

Hence \( f \) is strongly quasiconvex function.

**Theorem**

Let \( f:S \rightarrow \mathbb{R}n:S \rightarrow \mathbb{R}n \) and \( S \) be a non-empty convex set in \( \mathbb{R}n \). If \( x^\star x^\star \) is local optimal solution, then \( x^\star x^\star \) is unique global optimal solution.

**Proof**

Since a strong quasiconvex function is also strictly quasiconvex function, thus a local optimal solution is global optimal solution.

**Uniqueness** – Let \( f \) attains global optimal solution at two points \( u,v \in S \).

\[ f(u) \leq f(x) \forall x \in S \text{ and } f(v) \leq f(x) \forall x \in S \]

If \( u \) is global optimal solution, \( f(u) \leq f(v) \) and \( f(v) \leq f(u) \), then \( f(u) = f(v) \) which contradicts the quasiconvexity of \( f \) as \( a_{x^\star x^\star} = a_{x^\star x^\star} \).

Hence there exists only one global optimal solution.

**Remarks**

- A strongly quasiconvex function is also strictly quasiconvex function.
- A strictly convex function may or may not be strongly quasiconvex.
- A differentiable strictly convex is strongly quasiconvex.

**Pseudoconvex Function**

Let \( f:S \rightarrow \mathbb{R}f:S \rightarrow \mathbb{R} \) be a differentiable function and \( S \) be a non-empty convex set in \( \mathbb{R}n \), then \( f \) is said to be pseudoconvex if for each \( x_1,x_2 \in S \), \( x_2 \in S \) with \( \nabla f(x_1)T(x_2-x_1) \geq 0 \) \( \nabla f(x_1)T(x_2-x_1) \geq 0 \), we have \( f(x_2) \geq f(x_1) \) or equivalently if \( f(x_1) > f(x_2) \) then \( \nabla f(x_1)T(x_2-x_1) < 0 \).

**Pseudoconcave function**

Let \( f:S \rightarrow \mathbb{R}f:S \rightarrow \mathbb{R} \) be a differentiable function and \( S \) be a non-empty convex set in \( \mathbb{R}n \), then \( f \) is said to be pseudoconcave if for each \( x_1,x_2 \in S \), \( x_2 \in S \) with \( \nabla f(x_1)T(x_2-x_1) \geq 0 \) \( \nabla f(x_1)T(x_2-x_1) \geq 0 \), we have \( f(x_2) \leq f(x_1) \) or equivalently if \( f(x_1) > f(x_2) \), then \( \nabla f(x_1)T(x_2-x_1) > 0 \).

**Remarks**

- If a function is both pseudoconvex and pseudoconcave, then it is called pseudolinear.
- A differentiable convex function is also pseudoconvex.
- A pseudoconvex function may not be convex. For example,
If \( f \) is differentiable pseudoconvex function over \( S \), then \( f \) is both strictly quasiconvex as well as pseudoconvex function.

Let \( f \) be a pseudoconvex function and suppose \( \nabla f(x^k) = 0 \) for some \( x^k \in S \) with \( x^k \neq x \). Then we have \( f(x^k) \geq f(x) \) for all \( x \in S \), or equivalently if \( f(x^k) > f(x) \) for some \( x \in S \), then \( \nabla f(x^k) = 0 \).

**Theorem**

Let \( f \) be a pseudoconvex function and suppose \( \nabla f(x^k) = 0 \) for some \( x^k \in S \) with \( x^k \neq x \). Then \( x^k \) is global optimal solution of \( f \) over \( S \).

**Proof**

Let \( x^k \) be a critical point of \( f \), i.e., \( \nabla f(x^k) = 0 \). Since \( f \) is pseudoconvex function, for \( x \in S \), we have \( \nabla f(x^k)(x - x^k) = 0 \Rightarrow f(x^k) \leq f(x), \forall x \in S \) and \( \nabla f(x^k)(x - x^k) = 0 \Rightarrow f(x^k) \leq f(x), \forall x \in S \).

**Remark**

If \( f \) is strictly pseudoconvex function, \( x^k \) is unique global optimal solution.

**Theorem**

If \( f \) is differentiable pseudoconvex function over \( S \), then \( f \) is both strictly quasiconvex as well as quasiconvex function.

**Remarks**

- The sum of two pseudoconvex functions defined on an open set \( S \) of \( \mathbb{R}^n \) may not be pseudoconvex.
- Let \( f:S \rightarrow \mathbb{R}; S \rightarrow \mathbb{R} \) be a quasiconvex function and \( S \) be a non-empty convex subset of \( \mathbb{R}^n \) then \( f \) is pseudoconvex if and only if every critical point is a global minima of \( f \) over \( S \).
- Let \( S \) be a non-empty convex subset of \( \mathbb{R}^n \) and \( f:S \rightarrow \mathbb{R}; S \rightarrow \mathbb{R} \) be a function such that \( \nabla f(x) \neq 0 \) for every \( x \in S \). Then \( f \) is pseudoconvex if and only if it is a quasiconvex function.
Convex Optimization - Programming Problem

There are four types of convex programming problems

**Step 1** - minf(x)minf(x), where x∈Sx∈S and S be a non-empty convex set

**Step 2** - minf(x),x∈Rnminf(x),x∈Rn subject to

**Step 3** - maxf(x)maxf(x) where x∈Sx∈S and S be a non-empty convex set

**Step 4** - minf(x)minf(x), where x∈Rnx∈Rn subject to

Cone of feasible direction

Let S be a non-empty set in RnRn and let x^∈Closure(S)x^∈Closure(S), then the cone of feasible direction of S at x^x^, denoted by D, is defined as D={d:d≠0,x^+λd∈S,λ∈(0,δ),δ>0}D={d:d≠0,x^+λd∈S,λ∈(0,δ),δ>0}

Each non-zero vector d∈Dd∈D is called feasible direction.

For a given function f:Rn⇒Rff:Rn⇒R the cone of improving direction at x^x^ is denoted by F and is given by

F={d:f(x^+λd)<f(x^),∀λ∈(0,δ),δ>0}F={d:f(x^+λd)<f(x^),∀λ∈(0,δ),δ>0}

Each direction d∈Fd∈F is called an improving direction or descent direction of f at x^x^

**Theorem**

**Necessary Condition**

Consider the problem minf(x)minf(x) such that x∈Sx∈S where S be a non-empty set in RnRn. Suppose f is differentiable at a point x^x^∈Sx^x^∈S. If x^x^ is a local optimal solution, then F₀∩D=∅F₀∩D=∅ where F₀={d:▽f(x^)Td<0}F₀={d:▽f(x^)Td<0} and D is a cone of feasible direction.

**Sufficient Condition**

If F₀∩D=∅F₀∩D=∅ f is a pseudoconvex function at x^x^ and there exists a neighbourhood of x^,ε(x^),ε>0x^,ε(x^),ε>0 such that d=x−x^∈Dd=x−x^∈D for any x∈S∩Nε(x^)x∈S∩Nε(x^), then x^x^ is local optimal solution.

THANKS FOR READING – VISIT OUR WEBSITE www.educatererindia.com
Proof

Necessary Condition

Let $F_0 \cap D \neq \emptyset$, i.e., there exists a $d \in F_0 \cap D \neq \emptyset$ such that $d \in F_0 \cap D$ and $d \in D \neq \emptyset$. Since $d \in D \neq \emptyset$, therefore there exists $\delta_1 > 0 \delta_1 > 0$ such that $x^+ + \lambda d \in S, \lambda \in (0, \delta_1)]. x^+ + \lambda d \in S, \lambda \in (0, \delta_1)$. Let $d \in F_0 \neq F_0$, therefore $\nabla f(x^+)T d < 0 \nabla f(x^+)Td < 0$. Thus, there exists $\delta_2 > 0 \delta_2 > 0$ such that $f(x^+ + \lambda d) < f(x^+), \forall \lambda \in (0, \delta_2) f(x^+ + \lambda d) < f(x^+), \forall \lambda \in (0, \delta_2)$. Let $\delta = \min \{\delta_1, \delta_2\} = \min \{\delta_1, \delta_2\}$ Then $x^+ + \lambda d \in S, f(x^+ + \lambda d) < f(x^+), \forall \lambda \in (0, \delta) x^+ + \lambda d \in S, f(x^+ + \lambda d) < f(x^+), \forall \lambda \in (0, \delta)$. But $x^+ x^+$ is local optimal solution. Hence it is contradiction. Thus $F_0 \cap D = \emptyset$. $F_0 \cap D = \emptyset$.

Sufficient Condition

Let $F_0 \cap D \neq \emptyset$, nd let $f$ be a pseudoconvex function. Let there exists a neighbourhood of $x^+, N_\varepsilon(x^+)x^+, N_\varepsilon(x^+)$ such that $d = x^+ - x^+, \forall x \in S \cap N_\varepsilon(x^+)d = x^+ - x^+, \forall x \in S \cap N_\varepsilon(x^+)$ Let $x^+ x^+$ is not a local optimal solution. Thus, there exists $x^- \in S \cap N_\varepsilon(x^+)x^- \in S \cap N_\varepsilon(x^+)$ such that $f(x^-) < f(x^+) f(x^-) < f(x^+)$. By assumption on $S \cap N_\varepsilon(x^+), d = (x^- - x^+) \in D S \cap N_\varepsilon(x^+), d = (x^- - x^+) \in D$. By pseudoconvex of $f$, $f(x^+) > f(x^-) \Rightarrow \nabla f(x^+)T (x^- - x^+) < 0 f(x^+) > f(x^-) \Rightarrow \nabla f(x^+)T (x^- - x^+) < 0$ $\Rightarrow \nabla f(x^+)Td < 0 \Rightarrow \nabla f(x^+)Td < 0$ $\Rightarrow d \in F_0 \Rightarrow d \in F_0$. Hence $F_0 \cap D = \emptyset$. $F_0 \cap D = \emptyset$. which is a contradiction. Hence, $x^+ x^+$ is local optimal solution.

Consider the following problem: $\min f(x) \min f(x)$ where $x \in X x \in X$ such that $g_i(x) \leq 0, i = 1, 2, \ldots, m \max g(x) \leq 0, i = 1, 2, \ldots, m$. $f: X \rightarrow \mathbb{R}, g_i: X \rightarrow \mathbb{R} f: X \rightarrow \mathbb{R}, g_i: X \rightarrow \mathbb{R}$ and $X$ is an open set in $\mathbb{R}^n \mathbb{R}^n$. Let $S = \{x: g_i(x) \leq 0, \forall i\} S = \{x: g_i(x) \leq 0, \forall i\}$ Let $x^+ x^+ \in X x^+ x^+ \in X$ then $M = \{1, 2, \ldots, m\} M = \{1, 2, \ldots, m\}$. Let $I = \{i: g_i(x^+) = 0, i \in M\} I = \{i: g_i(x^+) = 0, i \in M\}$ where $I$ is called index set for all active constraints at $x^+ x^+$. 

THANKS FOR READING – VISIT OUR WEBSITE www.educatererindia.com
Let \( J = \{ i : g_i(x^+) < 0, i \in M \} \) and \( J = \{ i : g_i(x^-) < 0, i \in M \} \) where \( J \) is called index set for all active constraints at \( x^+x^- \).

Let \( F_0 = \{ d \in \mathbb{R}^m : \nabla f(x^+) T d < 0 \} \) and \( G_0 = \{ d \in \mathbb{R}^m : \nabla g_i(x^+) T d < 0 \} \) if \( G_0 \cap d \in \mathbb{R}^m : \nabla g_i(x^+) T d < 0, \forall i \in I \).

**Lemma**

If \( S = \{ x \in X : g_i(x) \leq 0, \forall i \in I \} \) is open and non-empty open set in \( \mathbb{R}^n \mathbb{R}^n \). Let \( x^+x^- \in \mathbb{S}x^+x^- \) and \( g_i(x) \) are differentiable at \( x^+, i \in I \) and \( g_i(x) \) is non-empty open set in \( \mathbb{R}^n \mathbb{R}^n \). Let \( \lambda \in (0, \delta_1) \) and \( \lambda \in (0, \delta_2) \). Then there exists \( \delta > 0 \) such that \( g_i(x^+\lambda d) < 0, \lambda \in (0, \delta) \) and \( x^+ \in \mathbb{M}x^+\lambda d \mathbb{E}X^+\lambda d \).

**Proof**

Let \( d \in G_o d \in G_0 \)

Since \( x^+Xx^+X \) is an open set, thus there exists \( \delta > 0 \) such that \( x^+\lambda d \) is contained in \( g_i(x) \leq 0, \lambda \in (0, \delta) \).

Also since \( g_i(x) < 0 \) and \( g_i(x) \) is continuous at \( x^+, i \in J \mathbb{J} x^+, i \in J \), there exists \( \delta > 0, \delta > 0 \) such that \( g_i(x^+\lambda d) < 0, \lambda \in (0, \delta) \) and \( x^+ \in \mathbb{M}x^+\lambda d \mathbb{E}X^+\lambda d \).

Since \( d \in G_0 d \in G_0 \), therefore \( \nabla g_i(x^+) T d < 0, \forall i \in I \) and \( g_i(x^+) T d < 0, \forall i \in I \) since \( \delta > 0 \) and \( g_i(x^+\lambda d) < 0, \lambda \in (0, \delta) \).

Let \( \delta = \min \{ \delta_1, \delta_2, \delta_3 \} \).

Therefore, \( x^+\lambda d \in X^+, g_i(x^+\lambda d) < 0, i \in Mx^+\lambda d \mathbb{E}X^+, g_i(x^+\lambda d) < 0, i \in M \).

Hence, \( G_0 \) is a local minimizer of \( S \).

**Theorem**

**Necessary Condition**

Let \( f_i \) and \( g_i, i \in I \) be differentiable at \( x^+x^- \in \mathbb{S}x^+x^- \) and \( g_i(x) \) is continuous at \( x^+x^- \in \mathbb{S}x^+x^- \). Let \( x^+x^- \) be a feasible point such that \( \nabla f(x^+) = 0, \nabla f(x^-) = 0 \), then \( f_0 = f_0 = \phi \).

**Sufficient Condition**

If \( f_0 = f_0 \) and \( f \) is a pseudo-convex function at \( (x^+, g(x)) \), \( i \in I \) \( g(x) \) is strictly pseudo-convex functions over some \( \epsilon \mathbb{E} \) - neighbourhood of \( x^+, x^+x^+, x^- \) is a local optimal solution.

**Remarks**

- Let \( x^+x^- \) be a feasible point such that \( \nabla f(x^+) = 0, \nabla f(x^-) = 0 \), then \( f_0 = f_0 = \phi \).
- But if \( \nabla g(x^+) = 0, \nabla g(x^-) = 0 \), then \( G_0 = G_0 = \phi \), thus \( f_0 G_0 = f_0 G_0 = \phi \).
- Consider the problem: \( \min f(x) \) such that \( g(x) = 0, g(x) = 0 \).
Since
\[ g(x) = 0, \]

thus \( g(x) < 0 \) and \( g_2(x) = -g(x) \leq 0 \).

Let \( x^* \in S \) and \( g_1(x^*) = 0 \) and \( g_2(x^*) = 0 \).

But \( \nabla g_1(x^*) = -\nabla g_2(x^*) \nabla g_1(x^*) = -\nabla g_2(x^*) \)

Thus, \( G_0 = \emptyset \) and \( F_0 \cap G_0 = \emptyset \).

**Convex Optimization - Fritz-John Conditions**

**Theorem**

Consider the problem \(-\inf_{x \in X} \min f(x)\) such that \( x \in X \) where \( X \) is an open set in \( \mathbb{R}^n \) and \( g_i(x) \leq 0, i=1,2,\ldots, m \).

Let \( f : X \rightarrow \mathbb{R} \) and \( g_i : X \rightarrow \mathbb{R} \)

Let \( x^* \) be a feasible solution and \( g_i, i \in I \) are differentiable at \( x^* \) and \( g_i, i \in J \) are continuous at \( x^* \).

If \( x^* \) solves the above problem locally, then there exists \( u_0, u_i \in \mathbb{R} \) such that

\[ u_0 \nabla f(x^*) + \sum_{i \in I} u_i \nabla g_i(x^*) = 0, \]

\[ u_0 \nabla g_i(x^*) = 0, \]

\( u_0, u_i \geq 0, i = 1,2,\ldots, m \)

such that \( u_0, u_i \) are called Lagrangian multipliers.

**Remarks**

- \( u_i \) are called Lagrangian multipliers.
- The condition that \( x^* \) be feasible to the given problem is called primal feasible condition.
- The requirement \( u_0 \nabla f(x^*) + \sum_{i=1}^m u_i \nabla g_i(x) = 0, u_0 \nabla g_i(x) = 0 \) is called dual feasibility condition.
- The condition \( u_i g_i(x^*) = 0, i = 1,2,\ldots, m \) is called complementary slackness condition. This condition requires \( u_i = 0, i \in J \).
- Together the primal feasible condition, dual feasibility condition and complimentary slackness are called Fritz-John Conditions.

**Sufficient Conditions**

**Theorem**

If there exists an \( \varepsilon \)-neighbourhood of \( x^* \), \( \varepsilon > 0 \) such that \( f \) is pseudoinvex over \( N_{\varepsilon}(x^*) \cap S \) and \( g_i, i \in I \) are strictly pseudoinvex.
over $N_\varepsilon(x^\wedge) \cap S_\varepsilon(x^\wedge) \cap S$, then $x^\wedge$ is local optimal solution to problem described above. If $f$ is pseudoconvex at $x^\wedge$ and if $g_i, i \in I, i \in I$ are both strictly pseudoconvex and quasiconvex function at $x^\wedge, x^\wedge x^\wedge, x^\wedge$ is global optimal solution to the problem described above.

Example

- $\min f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2 \min f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2$
  such that $x_21 + x_22 \leq 5, x_1 + 2x_2 \leq 4, x_1, x_2 \geq 0 x_1 + x_2 \leq 5, x_1 + 2x_2 \leq 4, x_1, x_2 \geq 0$ And $x^\wedge = (2, 1)^x^\wedge = (2, 1)$
  Let $g_1(x_1, x_2) = x_21 + x_22 - 5, g_1(x_1, x_2) = x_1 + x_2 - 5,$
  $g_2(x_1, x_2) = x_1 + 2x_2 - 4, g_2(x_1, x_2) = x_1 + 2x_2 - 4,$
  $g_3(x_1, x_2) = -x_1g_3(x_1, x_2) = -x_1$ and $g_4(x_1, x_2) = -x_1g_4(x_1, x_2) = -x_1.$

Thus the above constraints can be written as

$g_1(x_1, x_2) \leq 0, g_1(x_1, x_2) \leq 0,$
$g_2(x_1, x_2) \leq 0, g_2(x_1, x_2) \leq 0,$
$g_3(x_1, x_2) \leq 0, g_3(x_1, x_2) \leq 0$ and
$g_4(x_1, x_2) \leq 0, g_4(x_1, x_2) \leq 0$
Thus, $I = \{1, 2\} I = \{1, 2\}$ therefore, $u_3 = 0, u_4 = 0 u_3 = 0, u_4 = 0$
$\nabla f(x^\wedge) = (2, -2), \nabla g_1(x^\wedge) = (4, 2) \nabla f(x^\wedge) = (2, -2), \nabla g_1(x^\wedge) = (4, 2)$
and
$\nabla g_2(x^\wedge) = (1, 2) \nabla g_2(x^\wedge) = (1, 2)$
Thus putting these values in the first condition of Fritz-John conditions, we get –

$u_0 = 32u_2, u_1 = 12u_2, u_0 = 32u_2, u_1 = 12u_2,$
and let $u_2 = 1u_2 = 1,$
therefore $u_0 = 32, u_1 = 12u_0 = 32, u_1 = 12$
Thus Fritz John conditions are satisfied.

- $\min f(x_1, x_2) = -x_1 \min f(x_1, x_2) = -x_1$
  such that $x_2 - (1 - x_1)3 \leq 0 x_2 - (1 - x_1)3 \leq 0,$
  $-x_2 \leq 0 - x_2 \leq 0$ and $x^\wedge = (1, 0) x^\wedge = (1, 0)$
  Let $g_1(x_1, x_2) = x_2 - (1 - x_1)3 g_1(x_1, x_2) = x_2 - (1 - x_1)3,$
  $g_2(x_1, x_2) = -x_2g_2(x_1, x_2) = -x_2$

Thus the above constraints can be written as

$g_1(x_1, x_2) \leq 0, g_1(x_1, x_2) \leq 0,$
$g_2(x_1, x_2) \leq 0, g_2(x_1, x_2) \leq 0,$
Thus, $I = \{1, 2\} I = \{1, 2\}$
$\nabla f(x^\wedge) = (-1, 0) \nabla f(x^\wedge) = (-1, 0)$
$\nabla g_1(x^\wedge) = (0, 1) \nabla g_1(x^\wedge) = (0, 1)$ and $g_2(x^\wedge) = (0, -1) g_2(x^\wedge) = (0, -1)$
Thus putting these values in the first condition of Fritz-John conditions, we get –

$u_0 = 0, u_1 = u_2 = a > 0 u_0 = 0, u_1 = u_2 = a > 0$
Thus Fritz John conditions are satisfied.
Karush-Kuhn-Tucker Optimality Necessary Conditions

Consider the problem –

\[
\min f(x) \text{ such that } x \in X, \quad x \in X
\]

where \( X \) is an open set in \( \mathbb{R}^n \) and \( g_i(x) \leq 0, i = 1, 2, \ldots, m \)

Let \( S = \{ x \in X : g_i(x) \leq 0, \forall i \} \)

Let \( x^* \in S \) and let \( f \) and \( g_i, i \in I, i \in I \) be differentiable at \( x^* \) and \( g_i, i \in J, g_i \in J \)

are also differentiable at \( x^* \). Furthermore, \( \nabla g(x^*), i \in I \) \( \nabla g(x^*), i \in I \) are linearly independent.

If \( x^* \) solves the above problem locally, then there exists \( u_i, i \in I u_i, i \in I \) such that

\[
\nabla f(x^*) + \sum_{i \in I} u_i \nabla g_i(x^*) = 0
\]

\( i \in I \)

\[
\nabla f(x^*) + \sum_{i = 1}^m u_i \nabla g_i(x^*) = 0
\]

Therefore, \( u_i g_i(x^*) = 0, \forall i = 1, 2, \ldots, m \)

\[
u_i \geq 0, \forall i = 1, 2, \ldots, m, u_i \geq 0, \forall i = 1, 2, \ldots, m
\]

Example

Consider the following problem –

\[
\min f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2 \min f(x_1, x_2) = (x_1 - 3)^2 + (x_2 - 2)^2
\]

such that \( x_1 + x_2 \leq 5, x_2 \leq 5 \)

\( x_1, x_2 \geq 0 \) and \( x^* = (2, 1) \)

Let \( g_1(x_1, x_2) = x_1 + x_2 - 5, g_1(x_1, x_2) = x_1 + x_2 - 5 \)

\( g_2(x_1, x_2) = x_1 + 2x_2 - 4, g_2(x_1, x_2) = x_1 + 2x_2 - 4 \)

\( g_3(x_1, x_2) = -x_1g_3(x_1, x_2) = -x_1 \quad \text{and} \quad g_4(x_1, x_2) = -x_2g_4(x_1, x_2) = -x_2 \)

Thus the above constraints can be written as –

\[
g_1(x_1, x_2) \leq 0, g_2(x_1, x_2) \leq 0, g_3(x_1, x_2) \leq 0, g_4(x_1, x_2) \leq 0
\]

Thus, \( I = \{ 1, 2 \} \) and \( u_1 = 13, u_2 = 13 \)

Thus Karush-Kuhn-Tucker conditions are satisfied.

Algorithms for Convex Problem

Method of Steepest Descent
This method is also called Gradient method or Cauchy's method. This method involves the following terminologies:

\[ x_{k+1} = x_k + \alpha_k d_k x_k + 1 = x_k + \alpha_k d_k \]

\[ d_k = -\nabla f(x_k) \] or \[ d_k = -\nabla f(x_k) \parallel \nabla f(x_k) \parallel d_k = -\nabla f(x_k) \parallel f(x_k) \]

Let \[ \phi(\alpha) = f(x_k + \alpha d_k) \phi(\alpha) = f(x_k + \alpha d_k) \]

By differentiating \[ \phi \phi \] and equating it to zero, we can get \[ \alpha \alpha \].

So the algorithm goes as follows:

- Initialize \( x_0, \varepsilon_1, \varepsilon_2 \) and set \( k=0 \).
- Set \( d_k = -\nabla f(x_k) \) or \( d_k = -\nabla f(x_k) \parallel \nabla f(x_k) \parallel d_k = -\nabla f(x_k) \parallel f(x_k) \).
- Find \( \alpha \) such that \( \phi(\alpha) = f(x_k + \alpha d_k) \).
- Set \( x_{k+1} = x_k + \alpha_k d_k x_k + 1 = x_k + \alpha_k d_k \).
- If \( \|x_{k+1} - x_k\| < \varepsilon \) or \( \|f(x_{k+1})\| < \varepsilon \), go to step 6, otherwise set \( k = k + 1 \) and go to step 2.
- The optimal solution is \( x^* = x_k + 1 x^* = x_k + 1 \).

**Newton Method**

Newton Method works on the following principle:

\[ f(x) = y(x) = f(x_k) + (x - x_k) T \nabla f(x_k) + (x - x_k) T H(x_k) T (x - x_k) f(x) = f(x_k) + (x - x_k) T \nabla f(x_k) + (x - x_k) T H(x_k) T (x - x_k) \]

\[ \nabla y(x) = \nabla f(x_k) + H(x_k) (x - x_k) \nabla y(x) = \nabla f(x_k) + H(x_k) (x - x_k) \]

At \( x_{k+1} \), \( \nabla y(x_{k+1}) = \nabla f(x_k) + H(x_k) (x_{k+1} - x_k) x_k + 1, \nabla y(x_{k+1}) = \nabla f(x_k) + H(x_k) (x_{k+1} - x_k) \)

For \( x_{k+1} = x_k + 1 H(x_k) - 1 \nabla f(x_k) x_k + 1 = x_k - H(x_k) - 1 \nabla f(x_k) \)

Here \( H(x_k) H(x_k) \) should be non-singular.

Hence the algorithm goes as follows:

- **Step 1** – Initialize \( x_0, \varepsilon_0, \varepsilon \) and set \( k=0 k=0 \).
- **Step 2** – Find \( H(x_k) \nabla f(x_k) \).
- **Step 3** – Solve for the linear system \( H(x_k) h(x_k) = \nabla f(x_k) h(x_k) = \nabla f(x_k) \) for \( h(x_k) h(x_k) \).
- **Step 4** – Find \( x_{k+1} = x_k - h(x_k) x_k + 1 = x_k - h(x_k) \).
- **Step 5** – If \( \|x_{k+1} - x_k\| < \varepsilon \) or \( \|f(x_{k+1})\| < \varepsilon \), then go to step 6, else set \( k = k + 1 k + 1 \) and go to step 2.
- **Step 6** – The optimal solution is \( x^* = x_k + 1 x^* = x_k + 1 \).

**Conjugate Gradient Method**

This method is used for solving problems of the following types:

\[ \min f(x) = 12 T Q x - b x \min f(x) = 12 T Q x - b x \]

where \( Q \) is a positive definite nXn matrix and \( b \) is constant.
Given $x_0, \varepsilon, x_0', \varepsilon$, compute $g_0 = Qx_0 - bg_0 = Qx_0 - b$
Set $d_0 = -g_0d_0 = -g_0$ for $k = 0, 1, 2, ..., k = 0, 1, 2, ...$
Set $a_k = g_kg_kd_kQd_kak = g_kTg_kd_kTQd_k$
Compute $x_{k+1} = x_k + a_kd_kx_k + 1 = x_k + a_kd_k$
Set $g_{k+1} = g_k + a_kd_kg_k + 1 = g_k + a_kd_k$
Compute $\beta_k = g_kg_kd_kQd_k\beta_k = g_kTg_kd_kTQd_k$
Compute $x_{k+1} = x_k + a_kd_kx_k + 1 = x_k + a_kd_k$
Set $g_{k+1} = x_k + a_kQd_kg_k + 1 = x_k + a_kQd_k$
Compute $\beta_k = g_{k+1}g_{k+1}g_kg_k\beta_k = g_{k+1}Tg_{k+1}g_kTg_k$
Set $d_{k+1} = -g_{k+1} + \beta_kd_kd_k + 1 = -g_{k+1} + \beta_kd_k$. 

THANKS FOR READING – VISIT OUR WEBSITE www.educatererindia.com