Optimization for Machine Learning (CSL4010)

Dr. Md Abu Talhamainuddin Ansary
Department of Mathematics, IIT Jodhpur

PPT 2

Convex Set and Convex Function

- For $\in \mathbb{R}^n$, p-norm of x is defined as

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

  for $p \geq 1$.
- $\|x\|_p$ satisfies following properties
  - $\|x\|_p \geq 0$, $\|x\| = 0$ if and only if $x = 0$
  - $\|x - y\|_p = \|y - x\|_p$
  - $\|x - z\|_p \leq \|x - y\|_p + \|y - z\|_p$ (triangular inequality)

- For $\in \mathbb{R}^n$, p-norm of x is defined as

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

  for $p \geq 1$.
- $\|x\|_p$ satisfies following properties
  - $\|x\|_p \geq 0$, $\|x\| = 0$ if and only if $x = 0$
  - $\|x - y\|_p = \|y - x\|_p$
  - $\|x - z\|_p \leq \|x - y\|_p + \|y - z\|_p$ (triangular inequality)
- For $p = 1$, $\|x\|_1 = \sum_{i=1}^{n} |x_i|$.

- For $\in \mathbb{R}^n$, p-norm of x is defined as

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

  for $p \geq 1$.
- $\|x\|_p$ satisfies following properties
  - $\|x\|_p \geq 0$, $\|x\| = 0$ if and only if $x = 0$
  - $\|x - y\|_p = \|y - x\|_p$
  - $\|x - z\|_p \leq \|x - y\|_p + \|y - z\|_p$ (triangular inequality)
- For $p = 1$, $\|x\|_1 = \sum_{i=1}^{n} |x_i|$.
- For $p = 2$, $\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}}$. $\|x\|_2$ is known as Euclidian norm.

- For $\in \mathbb{R}^n$, p-norm of x is defined as

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$$

for $p \geq 1$.
- $\|x\|_p$ satisfies following properties
    - $\|x\|_p \geq 0$, $\|x\| = 0$ if and only if $x = 0$
    - $\|x - y\|_p = \|y - x\|_p$
    - $\|x - z\|_p \leq \|x - y\|_p + \|y - z\|_p$ (triangular inequality)
- For $p = 1$, $\|x\|_1 = \sum_{i=1}^{n} |x_i|$.
- For $p = 2$, $\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}}$. $\|x\|_2$ is known as Euclidian norm.
- For $p = \infty$ $\|x\|_\infty = \max\{|x_i|, \ \ i = 1, 2, \ldots, n\}$.
- We consider $\|x\| = \|x\|_2$ until it is specified.

- For $x^0 \in \mathbb{R}^n$, closed ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| \leq r\}$.

- For $x^0 \in \mathbb{R}^n$, closed ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| \leq r\}$.
- For $x^0 \in \mathbb{R}^n$, open ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| < r\}$.

- For $x^0 \in \mathbb{R}^n$, closed ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| \leq r\}$.
- For $x^0 \in \mathbb{R}^n$, open ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| < r\}$.
- A neighbourhood of $x^0 \in \mathbb{R}^n$ is defined as $N(x^0; \epsilon) = \{x \in \mathbb{R}^n | \|x - x^0\| < \epsilon\}$ where $\epsilon > 0$ is very small.

- For $x^0 \in \mathbb{R}^n$, closed ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| \leq r\}$.
- For $x^0 \in \mathbb{R}^n$, open ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| < r\}$.
- A neighbourhood of $x^0 \in \mathbb{R}^n$ is defined as $N(x^0; \epsilon) = \{x \in \mathbb{R}^n | \|x - x^0\| < \epsilon\}$ where $\epsilon > 0$ is very small.
- A set $S \subseteq \mathbb{R}^n$ is said to be an open set if for every $x \in S$, there exists $N(x, \epsilon) \subset S$

- For $x^0 \in \mathbb{R}^n$, closed ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| \leq r\}$.
- For $x^0 \in \mathbb{R}^n$, open ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| < r\}$.
- A neighbourhood of $x^0 \in \mathbb{R}^n$ is defined as $N(x^0; \epsilon) = \{x \in \mathbb{R}^n | \|x - x^0\| < \epsilon\}$ where $\epsilon > 0$ is very small.
- A set $S \subseteq \mathbb{R}^n$ is said to be an open set if for every $x \in S$, there exists $N(x, \epsilon) \subset S$
- For example: $(0, 1)$, $\|x \in \mathbb{R}^2 | x_1^2 + x_2^2 < 1\}$ etc..

- For $x^0 \in \mathbb{R}^n$, closed ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| \leq r\}$.
- For $x^0 \in \mathbb{R}^n$, open ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| < r\}$.
- A neighbourhood of $x^0 \in \mathbb{R}^n$ is defined as $N(x^0; \epsilon) = \{x \in \mathbb{R}^n | \|x - x^0\| < \epsilon\}$ where $\epsilon > 0$ is very small.
- A set $S \subseteq \mathbb{R}^n$ is said to be an open set if for every $x \in S$, there exists $N(x, \epsilon) \subset S$
- For example: $(0, 1)$, $\|x \in \mathbb{R}^2 | x_1^2 + x_2^2 < 1\}$ etc..
- A set $S \in \mathbb{R}^n$ is said to be a closed set if $\{x^n\} \subset S$ and $x^n \to x^*$ for $n \to \infty$ then $x^* \in S$.

- For $x^0 \in \mathbb{R}^n$, closed ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| \leq r\}$.
- For $x^0 \in \mathbb{R}^n$, open ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| < r\}$.
- A neighbourhood of $x^0 \in \mathbb{R}^n$ is defined as $N(x^0; \epsilon) = \{x \in \mathbb{R}^n | \|x - x^0\| < \epsilon\}$ where $\epsilon > 0$ is very small.
- A set $S \subseteq \mathbb{R}^n$ is said to be an open set if for every $x \in S$, there exists $N(x, \epsilon) \subset S$
- For example: $(0, 1)$, $\|x \in \mathbb{R}^2 | x_1^2 + x_2^2 < 1\}$ etc..
- A set $S \in \mathbb{R}^n$ is said to be a closed set if $\{x^n\} \subset S$ and $x^n \to x^*$ for $n \to \infty$ then $x^* \in S$.
- For example: $[0, 1]$, $\|x \in \mathbb{R}^2 | x_1^2 + x_2^2 \leq 1\}$ etc..

- For $x^0 \in \mathbb{R}^n$, closed ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| \leq r\}$.
- For $x^0 \in \mathbb{R}^n$, open ball centred at x^0 with radius r is the set $\{x \in \mathbb{R}^n | \|x - x_0\| < r\}$.
- A neighbourhood of $x^0 \in \mathbb{R}^n$ is defined as $N(x^0; \epsilon) = \{x \in \mathbb{R}^n | \|x - x^0\| < \epsilon\}$ where $\epsilon > 0$ is very small.
- A set $S \subseteq \mathbb{R}^n$ is said to be an open set if for every $x \in S$, there exists $N(x, \epsilon) \subset S$
- For example: $(0, 1)$, $\|x \in \mathbb{R}^2 | x_1^2 + x_2^2 < 1\}$ etc..
- A set $S \in \mathbb{R}^n$ is said to be a closed set if $\{x^n\} \subset S$ and $x^n \to x^*$ for $n \to \infty$ then $x^* \in S$.
- For example: $[0, 1]$, $\|x \in \mathbb{R}^2 | x_1^2 + x_2^2 \leq 1\}$ etc..
- If S is an open set then $\mathbb{R}^n \setminus S$ is a closed set and vice versa.

- A set $C \subseteq \mathbb{R}^n$ is said to be affine if the line through two distinct points in C lies in C. i.e if $x^1, x^2 \in C$ then $\theta x^1 + (1-\theta)x^2 \in C$ for any real number θ.

- A set $C \subseteq \mathbb{R}^n$ is said to be affine if the line through two distinct points in C lies in C. i.e if $x^1, x^2 \in C$ then $\theta x^1 + (1-\theta)x^2 \in C$ for any real number θ.
- A linear combination of the form $\theta_1 x^1 + \theta_2 x^2 + \cdots + \theta_n x^n$, $\theta_1 + \theta_2 + \cdots + \theta_n = 1$ is said to be an affine combination of $x^1, x^2, \ldots, x^n$.

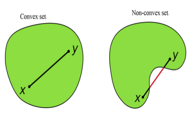
- A set $C \subseteq \mathbb{R}^n$ is said to be affine if the line through two distinct points in C lies in C. i.e if $x^1, x^2 \in C$ then $\theta x^1 + (1-\theta)x^2 \in C$ for any real number θ.
- A linear combination of the form $\theta_1 x^1 + \theta_2 x^2 + \cdots + \theta_n x^n$, $\theta_1 + \theta_2 + \cdots + \theta_n = 1$ is said to be an affine combination of $x^1, x^2, \ldots, x^n$.
- The set of all affine combinations of some set C is called affine hull of C and denoted by *aff* C. i.e.

$$\textit{aff } C = \{\theta_1 x^1 + \theta_2 x^2 + \cdots + \theta_n x^n | x^1, x^2, \ldots, x^n \in C, \theta_1 + \theta_2 + \cdots + \theta_n = 1.$$

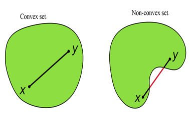
- Definition: A set $S \subseteq \mathbb{R}^n$ is said to be a convex set of for any $x^1, x^2 \in S$, $\lambda \geq 0$, $\lambda x^1 + (1-\lambda)x^2 \in S$.

- Definition: A set $S \subseteq \mathbb{R}^n$ is said to be a convex set of for any $x^1, x^2 \in S$, $\lambda \geq 0$, $\lambda x^1 + (1-\lambda)x^2 \in S$.
- **Geometrical interpretation:** If S is a convex set then for any two points of S, the line segment joining these two points lie inside the set.

- Definition: A set $S \subseteq \mathbb{R}^n$ is said to be a convex set of for any $x^1, x^2 \in S$, $\lambda \geq 0$, $\lambda x^1 + (1-\lambda)x^2 \in S$.
- **Geometrical interpretation:** If S is a convex set then for any two points of S, the line segment joining these two points lie inside the set.
- In first picture we can see that entire line segment joining x and y lies inside S. So S is a convex set.



- Definition: A set $S \subseteq \mathbb{R}^n$ is said to be a convex set of for any $x^1, x^2 \in S$, $\lambda \geq 0$, $\lambda x^1 + (1-\lambda)x^2 \in S$.
- **Geometrical interpretation:** If S is a convex set then for any two points of S, the line segment joining these two points lie inside the set.
- In first picture we can see that entire line segment joining x and y lies inside S. So S is a convex set.



- In the second picture, we can see that certain portion of the line segment joining x and y does not lies inside S_1. So S_1 is not a convex set.

Examples of convex set

- *One dimensional problem:* $S = \{x | x \geq, <, \leq, > a\}$, $S = [a, b]$, $S = (a, b)$.

Examples of convex set

- *One dimensional problem:* $S = \{x | x \geq, <, \leq, > a\}$, $S = [a, b]$, $S = (a, b)$.
- *Two dimension:* $S = \{(x_1, x_2) | a_1 x_1 + a_2 x_2 = b\}$ is a convex set.

Examples of convex set

- *One dimensional problem:* $S = \{x | x \geq, <, \leq, > a\}$, $S = [a, b]$, $S = (a, b)$.
- *Two dimension:* $S = \{(x_1, x_2) | a_1 x_1 + a_2 x_2 = b\}$ is a convex set. Proof: Suppose $x = (x_1, x_2)^T$ and $y = (y_1, y_2)^T$ be two points on S. Then $a_1 x_1 + a_2 x_2 = c$ and $a_1 y_1 + a_2 y_2 = c$. Now

$$\lambda x + (1 - \lambda) y = (\lambda x_1 + (1 - \lambda) y_1, \lambda x_2 + (1 - \lambda) y_2)$$

This implies

$$\begin{aligned} & a_1(\lambda x_1 + (1 - \lambda) y_1) + a_2(\lambda x_2 + (1 - \lambda) y_2) \\ &= \lambda(a_1 x_1 + a_2 x_2) + (1 - \lambda)(a_1 y_1 + a_2 y_2) \\ &= \lambda b + (1 - \lambda) b \qquad = b \end{aligned}$$

Hence $\lambda x + (1 - \lambda) y \in S$ So S is a convex set.

Examples of convex set

- *One dimensional problem:* $S = \{x | x \geq, <, \leq, > a\}$, $S = [a, b]$, $S = (a, b)$.
- *Two dimension:* $S = \{(x_1, x_2) | a_1 x_1 + a_2 x_2 = b\}$ is a convex set. Proof: Suppose $x = (x_1, x_2)^T$ and $y = (y_1, y_2)^T$ be two points on S. Then $a_1 x_1 + a_2 x_2 = c$ and $a_1 y_1 + a_2 y_2 = c$. Now

  $$\lambda x + (1-\lambda)y = (\lambda x_1 + (1-\lambda)y_1, \lambda x_2 + (1-\lambda)y_2)$$

  This implies

  $$\begin{aligned} & a_1(\lambda x_1 + (1-\lambda)y_1) + a_2(\lambda x_2 + (1-\lambda)y_2) \\ &= \lambda(a_1 x_1 + a_2 x_2) + (1-\lambda)(a_1 y_1 + a_2 y_2) \\ &= \lambda b + (1-\lambda)b \qquad = b \end{aligned}$$

  Hence $\lambda x + (1-\lambda)y \in S$ So S is a convex set.
- Similarly we can show that $S = \{x \in \mathbb{R}^n | a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = c\}$ i.e. $S = \{x \in \mathbb{R}^n | a^T x = b\}$ is a convex set. The set $S = \{x \in \mathbb{R}^n | a^T x = b\}$ is called a hyperplane.

Examples of convex set contd...

- The hyperplane $S = \{x \in \mathbb{R}^n | a^T x = b\}$ divides the space $\mathbb{R}^n$ into two parts

$$\begin{aligned} S_1 = &\quad \{x \in \mathbb{R}^n | a^T x \geq b\} \\ and \;\; S_2 = &\quad \{x \in \mathbb{R}^n | a^T x \leq b\} \end{aligned}$$

Each part is called a closed half space.

Examples of convex set contd...

- The hyperplane $S = \{x \in \mathbb{R}^n | a^T x = b\}$ divides the space $\mathbb{R}^n$ into two parts

$$
\begin{aligned}
S_1 = &\quad \{x \in \mathbb{R}^n | a^T x \geq b\} \\
and \;\; S_2 = &\quad \{x \in \mathbb{R}^n | a^T x \leq b\}
\end{aligned}
$$

  Each part is called a closed half space.
- The set $S_1 = \{x \in \mathbb{R}^n | a^T x > b\}$is called an open half space.

Examples of convex set contd...

- The hyperplane $S = \{x \in \mathbb{R}^n | a^T x = b\}$ divides the space $\mathbb{R}^n$ into two parts

$$\begin{aligned} S_1 = & \quad \{x \in \mathbb{R}^n | a^T x \geq b\} \\ and \;\; S_2 = & \quad \{x \in \mathbb{R}^n | a^T x \leq b\} \end{aligned}$$

  Each part is called a closed half space.
- The set $S_1 = \{x \in \mathbb{R}^n | a^T x > b\}$is called an open half space.
- Both half spaces (closed & open) are convex set (prove it).
- Intersection of m closed half space $A_{m \times n} x \leq b$, $b \in \mathbb{R}^m$ is a convex set.

Examples of convex set contd...

- The set$S = \{x \in \mathbb{R}^2 | x_1^2 + x_2^2 \leq 4\}$ is a convex set.
  Proof: Suppose $x = (x_1, x_2)^T$ and $y = (y_1, y_2)^T$ be two points on S.
  Then $x_1^2 + x_2^2 \leq 4$ and $y_1^2 + y_2^2 \leq 4$. Now

$$
\begin{aligned}
& (\lambda x_1 + (1-\lambda)y_1)^2 + (\lambda x_2 + (1-\lambda)y_2)^2 \\
= \; & \lambda^2 x_1^2 + 2\lambda(1-\lambda)x_1 y_1 + (1-\lambda)^2 y_1^2 + \lambda^2 x_2^2 - 2\lambda(1-\lambda)x_2 y_2 \\
& +(1-\lambda)^2 y_2^2 - \lambda x_1^2 - \lambda x_2^2 - (1-\lambda)y_1^2 - (1-\lambda)y_2^2 + \lambda x_1^2 + x_2^2 + (1-\lambda)y \\
= \; & \lambda(x_1^2 + x_2^2) + (1-\lambda)(y_1^2 + y_2^2) - \lambda(1-\lambda)(x_1 - y_1)^2 - \lambda(1-\lambda)(x_2 - y_2)^2 \\
= \; & 4
\end{aligned}
$$

  This implies $\lambda x + (1-\lambda)y \in S$. Hence S is a convex set.

Examples of convex set contd...

- The set $S = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 4\}$ is not a convex set since $x = (2, 0)^T$ and $y = (0, 2)^T$ belong to S but $0.5x + (1 - 0.5)y = (1, 1)^T$ does not belong to S.

Examples of convex set contd...

- The set $S = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 4\}$ is not a convex set since $x = (2,0)^T$ and $y = (0,2)^T$ belong to S but $0.5x + (1-0.5)y = (1,1)^T$ does not belong to S.
- The set $S = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \geq 4\}$ is not a convex set since $x = (-3,0)^T$ and $y = (3,0)^T$ belong to S but $0.5x + (1-0.5)y = (0,0)^T$ does not belong to S.

Examples of convex set contd...

- The set $S = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 4\}$ is not a convex set since $x = (2, 0)^T$ and $y = (0, 2)^T$ belong to S but $0.5x + (1 - 0.5)y = (1, 1)^T$ does not belong to S.
- The set $S = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \geq 4\}$ is not a convex set since $x = (-3, 0)^T$ and $y = (3, 0)^T$ belong to S but $0.5x + (1 - 0.5)y = (0, 0)^T$ does not belong to S.
- If $f : \mathbb{R}^n \to \mathbb{R}$ is any nonlinear function then $\{x \in \mathbb{R}^n : f(x) = \alpha\}$ is not a convex set but $\{x \in \mathbb{R}^n : f(x) \geq \alpha\}$ and $\{x \in \mathbb{R}^n : f(x) \leq \alpha\}$ may/may-not be convex.

Examples of convex set contd...

- The set $S = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 4\}$ is not a convex set since $x = (2,0)^T$ and $y = (0,2)^T$ belong to S but $0.5x + (1-0.5)y = (1,1)^T$ does not belong to S.
- The set $S = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \geq 4\}$ is not a convex set since $x = (-3,0)^T$ and $y = (3,0)^T$ belong to S but $0.5x + (1-0.5)y = (0,0)^T$ does not belong to S.
- If $f : \mathbb{R}^n \to \mathbb{R}$ is any nonlinear function then $\{x \in \mathbb{R}^n : f(x) = \alpha\}$ is not a convex set but $\{x \in \mathbb{R}^n : f(x) \geq \alpha\}$ and $\{x \in \mathbb{R}^n : f(x) \leq \alpha\}$ may/may-not be convex.

Properties of convex sets

- Intersection of two convex set is a convex set.

Properties of convex sets

- Intersection of two convex set is a convex set.
- Union of two convex set is not necessarily convex.

Properties of convex sets

- Intersection of two convex set is a convex set.
- Union of two convex set is not necessarily convex.
- If S_1, S_2 are two convex sets then $S_1 + S_2$ is a convex set.(prove it).

Properties of convex sets

- Intersection of two convex set is a convex set.
- Union of two convex set is not necessarily convex.
- If S_1, S_2 are two convex sets then $S_1 + S_2$ is a convex set.(prove it).
- Discrete set is not convex.

Properties of convex sets

- Intersection of two convex set is a convex set.
- Union of two convex set is not necessarily convex.
- If S_1, S_2 are two convex sets then $S_1 + S_2$ is a convex set.(prove it).
- Discrete set is not convex.
- $\mathbb{R}^n$, any singleton set, and ϕ are convex set.
- Suppose S is a convex set and f is an affine function (i.e. $f(x) = Ax + b$). Then image of S under f, i.e.$f(S) = \{f(x)|x \in S\}$ is a convex set.

Properties of convex sets

- Intersection of two convex set is a convex set.
- Union of two convex set is not necessarily convex.
- If S_1, S_2 are two convex sets then $S_1 + S_2$ is a convex set.(prove it).
- Discrete set is not convex.
- $\mathbb{R}^n$, any singleton set, and ϕ are convex set.
- Suppose S is a convex set and f is an affine function (i.e. $f(x) = Ax + b$). Then image of S under f, i.e.$f(S) = \{f(x)|x \in S\}$ is a convex set.
Proof Suppose $s^1, s^2 \in f(S)$. Then there exists $x^1, x^2 \in S$ such that $s^1 = f(x^1) = Ax^1 + b$ and $s^2 = f(x^2) = Ax^2 + b$. Now for $0 \leq \lambda \leq 1$,

$$\begin{aligned} \lambda s^1 + (1-\lambda)s^2 &= \lambda(Ax^1 + b) + (1-\lambda)(Ax^2 + b) \\ &= A(\lambda x^1 + (1-\lambda)x^2) + b \\ &= Ax^3 + b = f(x^3) \end{aligned}$$

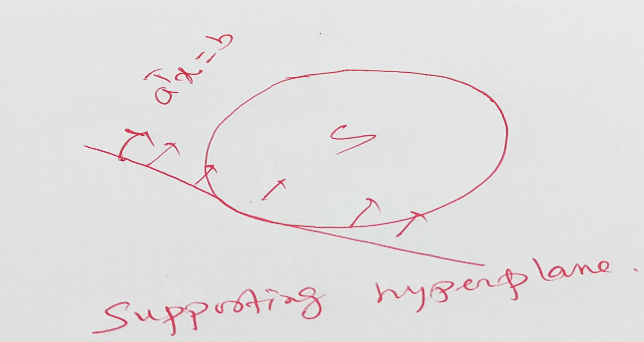
where $x^3 = \lambda x^1 + (1-\lambda)x^2$. Since S is a convex set $x^3 \in S$. Hence $\lambda s^1 + (1-\lambda)s^2 = f(x^3)$ for some $x^3 \in S$. i.e. $\lambda s^1 + (1-\lambda)s^2 \in f(S)$. So $f(S)$ is a convex set.

Properties of convex sets

- Suppose S be a non-empty convex set. A hyperplane $H = \{x \in \mathbb{R}^n | a^T x = b\}$ is said to be a supporting hyperplane of S if $a^T y \geq b$ or $a^T y \leq b$ for all $y \in S$.

Properties of convex sets

- Suppose S be a non-empty convex set. A hyperplane $H = \{x \in \mathbb{R}^n | a^T x = b\}$ is said to be a supporting hyperplane of S if $a^T y \geq b$ or $a^T y \leq b$ for all $y \in S$.



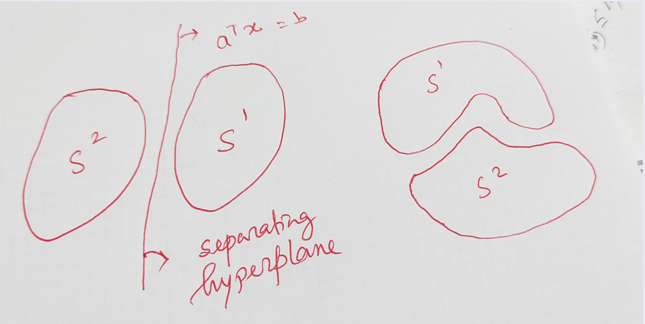
- If S is a convex set then there exists a supporting hyperplane at every boundary point of S .
- This result does not hold for non-convex sets.

Properties of convex sets

- Suppose S^1 and S^2 be two disjoint convex sets. Then there exists hyperplane $H = \{x|a^T x = b\}$ such that $a^T x \geq b$ for all $x \in S^1$ and $a^T y \leq b$ for all $y \in S^2$. The hyperplane H is called a separating hyperplane.

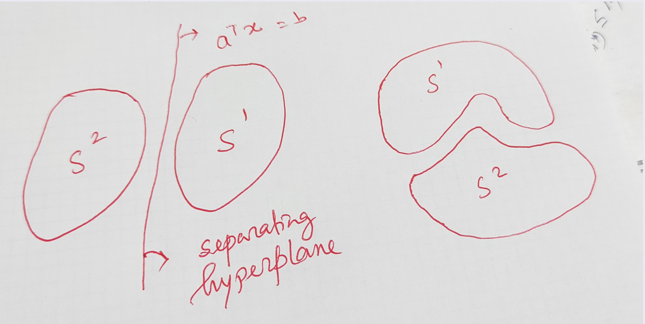
Properties of convex sets

- Suppose S^1 and S^2 be two disjoint convex sets. Then there exists hyperplane $H = \{x|a^T x = b\}$ such that $a^T x \geq b$ for all $x \in S^1$ and $a^T y \leq b$ for all $y \in S^2$. The hyperplane H is called a separating hyperplane.



Properties of convex sets

- Suppose S^1 and S^2 be two disjoint convex sets. Then there exists hyperplane $H = \{x | a^T x = b\}$ such that $a^T x \geq b$ for all $x \in S^1$ and $a^T y \leq b$ for all $y \in S^2$. The hyperplane H is called a separating hyperplane.



- H is a strict separating hyperplane if $a^T x > b$ for all $x \in S^1$ and $a^T y < b$ for all $y \in S^2$.

- **Convex Combination**: Suppose $\{x^1, x^2, \ldots, x^n\}$ The linear combination $\lambda_1 x^1 + \lambda_2 x^2 + \cdots + \lambda_n x^n$, $0 \leq \lambda \leq 1$, $\sum_{i=0}^{n} \lambda_i = 1$ is said to be a convex combination of $x^1, x^2, \ldots, x^n$.

- **Convex Combination**: Suppose $\{x^1, x^2, \dots, x^n\}$ The linear combination $\lambda_1 x^1 + \lambda_2 x^2 + \cdots + \lambda_n x^n$, $0 \le \lambda \le 1$, $\sum_{i=0}^{n} \lambda_i = 1$ is said to be a convex combination of $x^1, x^2, \dots, x^n$.
- Suppose $S = \{x^1, x^2, \dots, x^n\}$. Collection of all convex combinations of S is said to be the convex hull of S and denoted by $Conv(S)$ or $Co(S)$. i.e.

$$S = \{x \in \mathbb{R}^n | x = \sum_{i=0}^{n} \lambda_i x^i,\ 0 \le \lambda \le 1,\ \sum_{i=0}^{n} \lambda_i = 1\}$$

- **Convex Combination**: Suppose $\{x^1, x^2, \ldots, x^n\}$ The linear combination $\lambda_1 x^1 + \lambda_2 x^2 + \cdots + \lambda_n x^n$, $0 \leq \lambda \leq 1$, $\sum_{i=0}^{n} \lambda_i = 1$ is said to be a convex combination of $x^1, x^2, \ldots, x^n$.
- Suppose $S = \{x^1, x^2, \ldots, x^n\}$. Collection of all convex combinations of S is said to be the convex hull of S and denoted by $Conv(S)$ or $Co(S)$. i.e.

$$S = \{x \in \mathbb{R}^n | x = \sum_{i=0}^{n} \lambda_i x^i,\ 0 \leq \lambda \leq 1,\ \sum_{i=0}^{n} \lambda_i = 1\}$$

- Suppose$S = \{x^1, x^2\}$ then $Co(S)$ is the line segment joining x^1 and x^2.

- **Convex Combination**: Suppose $\{x^1, x^2, \ldots, x^n\}$ The linear combination $\lambda_1 x^1 + \lambda_2 x^2 + \cdots + \lambda_n x^n$, $0 \leq \lambda \leq 1$, $\sum_{i=0}^{n} \lambda_i = 1$ is said to be a convex combination of $x^1, x^2, \ldots, x^n$.
- Suppose $S = \{x^1, x^2, \ldots, x^n\}$. Collection of all convex combinations of S is said to be the convex hull of S and denoted by $Conv(S)$ or $Co(S)$. i.e.

$$S = \{x \in \mathbb{R}^n | x = \sum_{i=0}^{n} \lambda_i x^i,\ 0 \leq \lambda \leq 1,\ \sum_{i=0}^{n} \lambda_i = 1\}$$

- Suppose$S = \{x^1, x^2\}$ then $Co(S)$ is the line segment joining x^1 and x^2.
- Suppose $S = \{x^1, x^2, x^3\}$ then $Co(S)$ is the triangle formed by x^1, x^2 and x^3 (if x^1, x^2, x^3 don't line in a hyperplane).

- **Convex Combination**: Suppose $\{x^1, x^2, \dots, x^n\}$ The linear combination $\lambda_1 x^1 + \lambda_2 x^2 + \cdots + \lambda_n x^n$, $0 \le \lambda \le 1$, $\sum_{i=0}^{n} \lambda_i = 1$ is said to be a convex combination of $x^1, x^2, \dots, x^n$.
- Suppose $S = \{x^1, x^2, \dots, x^n\}$. Collection of all convex combinations of S is said to be the convex hull of S and denoted by $Conv(S)$ or $Co(S)$. i.e.

$$S = \{x \in \mathbb{R}^n | x = \sum_{i=0}^{n} \lambda_i x^i,\ 0 \le \lambda \le 1,\ \sum_{i=0}^{n} \lambda_i = 1\}$$

- Suppose$S = \{x^1, x^2\}$ then $Co(S)$ is the line segment joining x^1 and x^2.
- Suppose $S = \{x^1, x^2, x^3\}$ then $Co(S)$ is the triangle formed by x^1, x^2 and x^3 (if x^1, x^2, x^3 don't line in a hyperplane).
- Suppose S be a convex set. $x \in S$ is said to be a vertex of S if x can not be expressed as convex combination of any other points S.

- **Convex Combination**: Suppose $\{x^1, x^2, \dots, x^n\}$ The linear combination $\lambda_1 x^1 + \lambda_2 x^2 + \cdots + \lambda_n x^n$, $0 \le \lambda \le 1$, $\sum_{i=0}^{n} \lambda_i = 1$ is said to be a convex combination of $x^1, x^2, \dots, x^n$.
- Suppose $S = \{x^1, x^2, \dots, x^n\}$. Collection of all convex combinations of S is said to be the convex hull of S and denoted by $Conv(S)$ or $Co(S)$. i.e.

$$S = \{x \in \mathbb{R}^n | x = \sum_{i=0}^{n} \lambda_i x^i,\ 0 \le \lambda \le 1,\ \sum_{i=0}^{n} \lambda_i = 1\}$$

- Suppose$S = \{x^1, x^2\}$ then $Co(S)$ is the line segment joining x^1 and x^2.
- Suppose $S = \{x^1, x^2, x^3\}$ then $Co(S)$ is the triangle formed by x^1, x^2 and x^3 (if x^1, x^2, x^3 don't line in a hyperplane).
- Suppose S be a convex set. $x \in S$ is said to be a vertex of S if x can not be expressed as convex combination of any other points S.
- Suppose $S = \{x^1, x^2, \dots, x^n\}$. Then $Co(S)$ is the convex hull of all vertexes of S.

- **Convex Combination**: Suppose $\{x^1, x^2, \ldots, x^n\}$ The linear combination $\lambda_1 x^1 + \lambda_2 x^2 + \cdots + \lambda_n x^n$, $0 \leq \lambda \leq 1$, $\sum_{i=0}^{n} \lambda_i = 1$ is said to be a convex combination of $x^1, x^2, \ldots, x^n$.
- Suppose $S = \{x^1, x^2, \ldots, x^n\}$. Collection of all convex combinations of S is said to be the convex hull of S and denoted by $Conv(S)$ or $Co(S)$. i.e.

$$S = \{x \in \mathbb{R}^n | x = \sum_{i=0}^{n} \lambda_i x^i,\ 0 \leq \lambda \leq 1,\ \sum_{i=0}^{n} \lambda_i = 1\}$$

- Suppose$S = \{x^1, x^2\}$ then $Co(S)$ is the line segment joining x^1 and x^2.
- Suppose $S = \{x^1, x^2, x^3\}$ then $Co(S)$ is the triangle formed by x^1, x^2 and x^3 (if x^1, x^2, x^3 don't line in a hyperplane).
- Suppose S be a convex set. $x \in S$ is said to be a vertex of S if x can not be expressed as convex combination of any other points S.
- Suppose $S = \{x^1, x^2, \ldots, x^n\}$. Then $Co(S)$ is the convex hull of all vertexes of S.

Theorem 1

Suppose $S = \{x^1, x^2, \ldots, x^n\}$*. Then* $Co(S)$ *is a convex set.*

Theorem 1

Suppose $S = \{x^1, x^2, \ldots, x^n\}$. Then $Co(S)$ is a convex set.

Proof: Suppose $y^1, y^2 \in Co(S)$. Then there exists $\mu = (\mu_1, \mu_2, \ldots, \mu_n)^T, \theta = (\theta_1, \theta_2, \ldots, \theta_n)^T$ such that

$$\begin{aligned} y^1 &= \mu_1 x^1 + \mu_2 x^2 + \cdots + \mu_n x^n, \quad 0 \le \mu_i \le 1, \quad \sum_{i=1}^{n} \mu_i = 1 \\ y^2 &= \theta_1 x^1 + \theta_2 x^2 + \cdots + \theta_n x^n, \quad 0 \le \theta_i \le 1, \quad \sum_{i=1}^{n} \theta_i = 1 \end{aligned}$$

Now for $0 \le \lambda \le 1$,

$$\begin{aligned} \lambda y^1 + (1-\lambda) y^2 &= \lambda(\mu_1 x^1 + \mu_2 x^2 + \cdots + \mu_n x^n) + (1-\lambda)(\theta_1 x^1 + \theta_2 x^2 + \cdots + \theta_n \\ &= \sum_{i=1}^{n} (\lambda \mu_i + (1-\lambda)\theta_i)\, x^i \end{aligned}$$

Clearly $\lambda \mu_i + (1-\lambda)\theta_i \ge 0$ for all i and

$$\sum_{i=1}^{n} (\lambda \mu_i + (1-\lambda)\theta_i) = \lambda \sum_{i=1}^{n} \mu_i + (1-\lambda) \sum_{i=1}^{n} \theta_i = \lambda + (1-\lambda) = 1$$

Hence $\lambda y^1 + (1-\lambda) y^2 \in Co(S)$. So $Co(S)$ is convex set.

Convex combination contd...

Theorem 2

Suppose S be a non-empty convex set and $A = \{x^1, x^2, \dots, x^n\} \subseteq S$. Then $Co(A) \subseteq S$.

Proof: We prove the result by induction.
For n=2, the result holds from the definition of convex set.
Suppose the result holds for $n-1$, i.e. $Co\{x^1, x^2, \dots, x^{n-1}\} \subseteq S$.
Let

$$y = \lambda_1 x^1 + \lambda_2 x^2 + \cdots + \lambda_n x^n \in Co(A) \tag{1}$$

Then $\lambda_i \geq 0$ and $\sum_{i=1}^{n} \lambda_i = 1$. Then $\sum_{i=1}^{n-1} \lambda_i = 1 - \lambda_n$. This implies $\frac{\lambda_i}{1-\lambda_n} \geq 0$ and $\sum_{i=1}^{n-1} \frac{\lambda_i}{1-\lambda_n} = 1$.
Hence $z = \sum_{i=1}^{n-1} \frac{\lambda_i}{1-\lambda_n} x^i \in Co\{x^1, x^2, \dots, x^{n-1}\} \subseteq S$ and

$$(1-\lambda_n)z = \lambda_1 x^1 + \lambda_2 x^2 + \cdots + \lambda_{n-1} x^{n-1}.$$

Then from (**??**), $(1-\lambda_n)z = y - \lambda_n x^n$. This implies $y = (1-\lambda_n)z + \lambda_n x^n$.
Since S is a convex set, $z, x^n \in S$ implies $y = (1-\lambda_n)z + \lambda_n x^n \in S$. Hence $Co\{x^1, x^2, \dots, x^n\} \subseteq S$.

- A set $C \subseteq \mathbb{R}^n$ is said to be a cone if $x \in C$, $\lambda > 0$ implies $\lambda x \in C$. For example $C = \{x \in \mathbb{R}^2 | x_1, x_2 \geq 0\}$.

- A set $C \subseteq \mathbb{R}^n$ is said to be a cone if $x \in C$, $\lambda > 0$ implies $\lambda x \in C$. For example $C = \{x \in \mathbb{R}^2 | x_1, x_2 \geq 0\}$.
- A cone is said to be convex if it is convex and cone. i.e. $x^1, x^2 \in C$, $\lambda_1, \lambda_2 \geq 0$ $\lambda_1^2 + \lambda_2^2 > 00$ imply $\lambda_1 x^1 + \lambda_2 x^2 \in C$.

- A set $C \subseteq \mathbb{R}^n$ is said to be a cone if $x \in C$, $\lambda > 0$ implies $\lambda x \in C$. For example $C = \{x \in \mathbb{R}^2 | x_1, x_2 \geq 0\}$.
- A cone is said to be convex if it is convex and cone. i.e. $x^1, x^2 \in C$, $\lambda_1, \lambda_2 \geq 0$ $\lambda_1^2 + \lambda_2^2 > 00$ imply $\lambda_1 x^1 + \lambda_2 x^2 \in C$.
- A linear combination of the form $\theta_1 x^1 + \theta_2 x^2 + \cdots + \theta_n x^n$, $\theta_1, \theta_2, \ldots, \theta_n \geq 0$ is said to be a conic combination of $x^1, x^2, \ldots, x^n$.

- A set $C \subseteq \mathbb{R}^n$ is said to be a cone if $x \in C$, $\lambda > 0$ implies $\lambda x \in C$. For example $C = \{x \in \mathbb{R}^2 | x_1, x_2 \geq 0\}$.
- A cone is said to be convex if it is convex and cone. i.e. $x^1, x^2 \in C$, $\lambda_1, \lambda_2 \geq 0$ $\lambda_1^2 + \lambda_2^2 > 00$ imply $\lambda_1 x^1 + \lambda_2 x^2 \in C$.
- A linear combination of the form $\theta_1 x^1 + \theta_2 x^2 + \cdots + \theta_n x^n$, $\theta_1, \theta_2, \ldots, \theta_n \geq 0$ is said to be a conic combination of $x^1, x^2, \ldots, x^n$.
- The set of all conic combinations of some set C is called conic hull of C and denoted by $cone(C)$. i.e.

$$cone(C) = \{\theta_1 x^1 + \theta_2 x^2 + \cdots + \theta_n x^n | x^1, x^2, \ldots, x^n \in C, \theta_1, \theta_2, \ldots, \theta_n \geq 0.$$

- A set $C \subseteq \mathbb{R}^n$ is said to be a cone if $x \in C$, $\lambda > 0$ implies $\lambda x \in C$. For example $C = \{x \in \mathbb{R}^2 | x_1, x_2 \geq 0\}$.
- A cone is said to be convex if it is convex and cone. i.e. $x^1, x^2 \in C$, $\lambda_1, \lambda_2 \geq 0$ $\lambda_1^2 + \lambda_2^2 > 0$0 imply $\lambda_1 x^1 + \lambda_2 x^2 \in C$.
- A linear combination of the form $\theta_1 x^1 + \theta_2 x^2 + \cdots + \theta_n x^n$, $\theta_1, \theta_2, \ldots, \theta_n \geq 0$ is said to be a conic combination of $x^1, x^2, \ldots, x^n$.
- The set of all conic combinations of some set C is called conic hull of C and denoted by $cone(C)$. i.e.

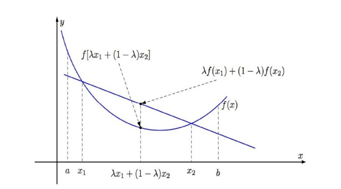
  $$cone(C) = \{\theta_1 x^1 + \theta_2 x^2 + \cdots + \theta_n x^n | x^1, x^2, \ldots, x^n \in C, \theta_1, \theta_2, \ldots, \theta_n \geq 0.$$

- A cone C is said to be a pointed cone if $C \cap -C \subseteq \{0\}$.

- A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be a convex function iff for any $x, y \in X$, $0 \leq \lambda \leq 1$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y). \tag{2}$$

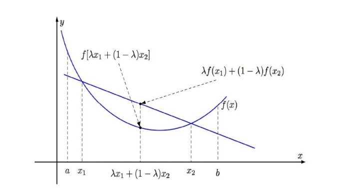
- Geometrically this inequality implies that the graph of f between x and y lies below the line segment joining $(x, f(x))$ and $(y, f(y))$.



- A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be a convex function iff for any $x, y \in X$, $0 \leq \lambda \leq 1$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y). \quad (2)$$

- Geometrically this inequality implies that the graph of f between x and y lies below the line segment joining $(x, f(x))$ and $(y, f(y))$.

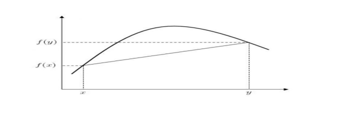


- A function is strictly convex if strict inequality holds in (**??**).

- A function $f : X\mathbb{R}^n \to \mathbb{R}$ is said to be a convex function iff for any $x, y \in X$, $0 \leq \lambda \leq 1$,

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y). \tag{3}$$

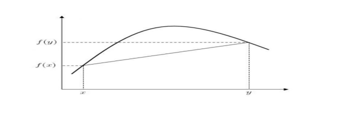
- Geometrically this inequality implies that the graph of f between x and y lies above the line segment joining $(x, f(x))$ and $(y, f(y))$.



- A function $f : X\mathbb{R}^n \to \mathbb{R}$ is said to be a convex function iff for any $x, y \in X$, $0 \leq \lambda \leq 1$,

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y). \tag{3}$$

- Geometrically this inequality implies that the graph of f between x and y lies above the line segment joining $(x, f(x))$ and $(y, f(y))$.



- A function is strictly concave if strict inequality holds in (**??**).

- **Example of convex functions** $y = \sum_{i=1}^{n} x_i^2$, $y = e^{\sum_{i=1}^{n} x_i}$ are (strictly) convex functions in $\mathbb{R}^n$. $y = \sum_{i=1}^{n} -\log(x_i)$, $y = \sum_{i=1}^{n} x_i \log(x_i)$ are (strictly) convex function in $\mathbb{R}^n_{++} = \{x \in \mathbb{R}^n | x_i > 0, \forall\ i\}$.
- If f is a convex function then $-f$ is a concave function and vice-versa.
- If $f_1, f_2, \ldots, f_n$ two convex functions then $\sum_{i=1}^{n} f_i$,αf_i ($\alpha > 0$), $\max\{f_1, f_2, \ldots, f_n\}$ are also convex functions.
- If f_1, f_2 are two convex functions then $\min\{f_1, f_2\}$ is not necessarily convex. For example suppose $f_1(x) = x^2$ and $f_2(x) = (x-1)^2$. Define $f(x) = \min\{x^2, (x-1)^2\}$. Then $f(1/2 * 0 + 1/2 * 1) > 1/2 f(0) + 1/2 f(1)$
- Affine functions $f(x) = a^T x + b$ are both convex and concave function.

- If f is a convex function then the level set $L = \{x : f(x) \leq \alpha\}$ $\alpha \in \mathbb{R}$ is a convex set.
  Proof: Suppose $x^1, x^2 \in L$. Then $f(x^1) \leq \alpha$ and $f(x^2) \leq \alpha$. Now for $0 \leq \lambda \leq 1$,

$$\begin{aligned} f(\lambda x^1 + (1-\lambda)x^2) &\leq \lambda f(x^1) + (1-\lambda) f(x^2) \\ &\leq \lambda\alpha + (1-\lambda)\alpha \\ &= \alpha \end{aligned}$$

  The first inequality holds, since f is a convex function. Hence $f(\lambda x^1 + (1-\lambda)x^2) \leq \alpha$ i.e. $\lambda x^1 + (1-\lambda)x^2 \in L$. So L is a convex set.
- If f is a concave function then the set $S = \{x | f(x) \geq \alpha\}$ is a convex set. (prove it).
- If f is a convex function then the set $L = \{x : f(x) \geq \alpha\}$ $\alpha \in \mathbb{R}$ is not necessarily a convex set (provide counter example).
- Every convex function is continuous in the interior of its domain.
- $f(x) = |x|$ is an example of non-differentiable convex function.

- Suppose $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Then gradient of f at x^* is defined as $\nabla f(x^*) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x^*) \\ \frac{\partial f}{\partial x_2}(x^*) \\ \cdots \\ \frac{\partial f}{\partial x_n}(x^*) \end{bmatrix}$.
- Suppose $f(x) = a^T x = x^T a = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$. Then $\nabla f(x) = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{bmatrix} = a$.
- Suppose $Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix}$ is an 2×2 symmetric matrix and $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x) = \frac{1}{2} x^T Q x = \frac{1}{2}(x_1, x_2) Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{2}\{q_{11} x_1^2 + 2 q_{12} x_1 x_2 + q_{22} x_2^2$$

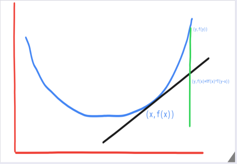
Then $\nabla f(x) = \begin{bmatrix} q_1 1 x_1 + q_{12} x_2 \\ q_{12} x_1 + q_{22} x_2 \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = Qx.$

- Similarly for n dimensional square matrix Q, if $f(x) = \frac{1}{2} x^T Q x$ then $\nabla f(x) = Qx$

- If a convex function is differentiable at x, then for any y,

$$f(y) \geq f(x) + \nabla f(x^*)^T (y - x) \tag{4}$$

- Geometrically this inequality implies that the entire graph of f lies above the tangent plane at x.



- Suppose f is a convex function and $\nabla f(x^*) = 0$ for some $x^* \in X$, then from (**??**) $f(y) \geq f(x^*)$ for all $y \in X$. Hence x^* is a minima of f.
- Consider the problem $\min\limits_{x \in \mathbb{R}^n} f(x)$. If f is convex, then $\nabla f(x^*) = 0$ implies x^* is the solution of this problem.

- Consider the least square problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|^2 = \frac{1}{2} \left(x^T A^T A x - 2x^T A^T b + b^T b \right\} .$$

Here the objective function is convex. Hence $\nabla \left(\frac{1}{2} \|Ax - b\|^2 \right) = 0$ gives the optimal solution. Now $\nabla \left(\frac{1}{2} \|Ax - b\|^2 \right) = 0$ implies $A^T A x - A^T b = 0$, i.e.

$$x = (A^T A)^{-1} A^T b,$$

which is the least square solution.

- An $n \times n$ matrix H is said to be a
  - a positive semi-definite matrix if $x^T Hx \geq 0$ for all $x \in \ mathbbR^n$.
  - a positive definite matrix if $x^T Hx > 0$ for all non-zeros $x \in \ mathbbR^n$.
  - a negative semi-definite matrix if $x^T Hx \leq 0$ for all $x \in \ mathbbR^n$.
  - a negative definite matrix if $x^T Hx < 0$ for all non-zeros $x \in \ mathbbR^n$.
- Diagonal elements of a
  - a positive semi-definite matrix is greater than equal to 0.
  - a positive definite matrix is strictly greater then zero.
  - a negative semi-definite matrix is less than equal to zeros.
  - a negative definite matrix is less than zero.
- Converse of the above statements are not always true. For example if diagonal elements of a matrix is > 0 then it is not necessarily a positive definite matrix. For example $H = \begin{bmatrix} 2 & 4 \\ 4 & 1 \end{bmatrix}$ has strictly positive diagonal elements but H is not positive definite since $(1, 2)H \begin{bmatrix} 1 \\ 2 \end{bmatrix} = -10 < 0$.

- Eigen values of a symmetric
  - positive definite matrix are > 0.
  - positive semi definite matrix are ≥ 0.
  - negative definite matrix are < 0.
  - negative semi definite matrix are ≤ 0.
- **Definition:** Suppose H be an $n \times n$ matrix. A leading principal minor of H of order k is the minor of order k obtained by deleting the last $n - k$ rows and columns.
- If H is a positive (semi) definite matrix then leading principal minors of H of order $1, 2, \ldots, n$ $(\geq) > 0$.
- If H is a negative (semi) definite matrix then leading principal minors of H of order k is $(\leq) < 0$ if k is odd and $(\geq) > 0$ if k is even for $k = 1, 2, \ldots, n$.
- Suppose $H = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 2 & -1 \\ 2 & -1 & 4 \end{bmatrix}$. For any $x \in \mathbb{R}^3$, we have

$$x^T Hx = (x_1 + x_2)^2 + (\sqrt{2}x_1 + \sqrt{2}x_3)^2 + (x_2 - x_3)^2 + x_3^2 > 0$$

  for any $x \neq 0$. Hence H is a positive definite matrix.
- One can verify eigenvalues and leading principle minors of H are > 0.

- Suppose $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function. Then Hessian of f at x is defined as

$$H(x) = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 x_n}(x) \\ \frac{\partial^2 f}{\partial x_1 x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \dots & \frac{\partial^2 f}{\partial x_2 x_n}(x) \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial x_n x_2}(x) & \frac{\partial^2 f}{\partial x_n x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}$$

- Since f is continuous, Hessian of f is a symmetric matrix.
- f is a convex function in $X \subseteq \mathbb{R}^n$ if and only $\nabla^f(x)$ is positive semi-definite for every $x \in X$.
- If Hessian of f is positive definite then f is strictly convex.

- Suppose $f(x) = 2x_1^2 + 3x_2^2 + 2x_3^2 + 2x_1x_2 + 3x_1x_3 + 4x_2x_3$. Then $\nabla f(x) = \begin{bmatrix} 4x_1 + 2x_2 + 3x_3 \\ 2x_1 + 6x_2 + 4x_3 \\ 3x_1 + 4x_2 + 4x_3 \end{bmatrix}$ and $\nabla^2 f(x) = \begin{bmatrix} 4 & 2 & 3 \\ 2 & 6 & 4 \\ 3 & 4 & 4 \end{bmatrix}$.
- For any $x \in \mathbb{R}^3$, leading principle minors of $\nabla^2 f(x)$ are $4, 20, 10$. This implies $\nabla^2 f(x)$ is positive definite for every $x \in \mathbb{R}^3$. Hence f is a strictly convex function.
- If f is strictly convex then $\nabla^2 f(x)$ is not necessarily positive definite. For example $f(x) = x_1^4 + x_2^4$. (verify)

- A function $f : X \to \mathbb{R}$ is said to be a σ-strongly convex for some $\sigma > 0$ if for any $x, y \in X$ and $0 \le \lambda \le 1$

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) - \frac{\sigma}{2}\lambda(1-\lambda)\|x-y\|^2.$$

  $\sigma(> 0)$ is said to be the modulus of strong convexity. Sometimes, a σ-strongly convex function is called as 'a strongly convex function with modulus σ.
- If f is a σ-strongly convex then $g(x) = f(x) - \frac{\sigma}{2}\|x\|^2$ is a strictly convex function.
- Every strongly convex function is a strictly convex function but not converse. For example $f(x) = e^x, -\log(x)$ are strictly convex but not strongly convex function.
- If f is a σ-strongly convex function then $d^T\nabla^2 f(x)d \ge \frac{\sigma}{2}\|d\|^2$ for every $x \in X$ and $d \in \mathbb{R}^n$.
- The modulus of convexity $\sigma = \min_{x\in X} eig\_min \nabla^2 f(x)$, where *eig_min* of a matrix H is the minimum eigenvalue of H.

- Boyd, S. and Vandenberghe, L.: Convex optimization. Cambridge university press, 2004.
- Boyd, S. and Vandenberghe L.: Introduction to applied linear algebra: vectors, matrices, and least squares. Cambridge university press, 2018.