THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics
MATH4230 2024-25
Lecture 11
February 18, 2025 (Tuesday)

1 Recall

We learned separation last week and proved the direction " $\implies$ " for the Theorem 1. Today, let us complete the proof for the direction " $\impliedby$ ".

Definition 1. Let X_1, X_2 be two non-empty subsets. We say X_1 and X_2 can be properly separated if there exists $w \neq 0$ such that

$$\begin{cases} \sup\limits_{x_1 \in X_1} w^T x_1 \leq \inf\limits_{x_2 \in X_2} w^T x_2 \\ \inf\limits_{x_1 \in X_1} w^T x_1 < \sup\limits_{x_2 \in X_2} w^T x_2 \end{cases}$$

Theorem 1. *Let X_1, X_2 be non-empty convex subsets. Then*

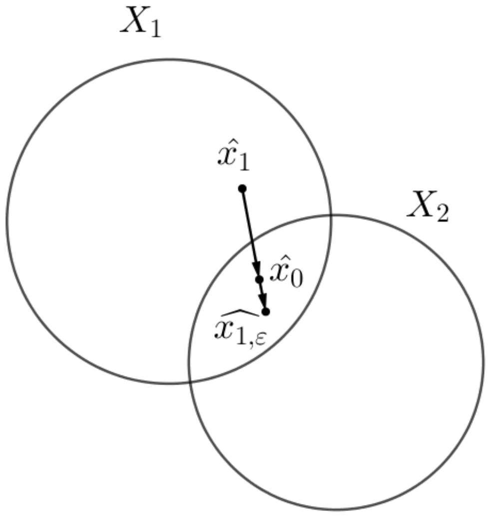
$$\mathrm{ri}(X_1) \cap \mathrm{ri}(X_2) = \emptyset \iff X_1 \text{ and } X_2 \text{ can be properly separated.}$$

Proof. ($\implies$): Proved last week.
($\impliedby$): Assume that $\mathrm{ri}(X_1) \cap \mathrm{ri}(X_2) \neq \emptyset$, then there exists $\hat{x}_0 \in \mathrm{ri}(X_1) \cap \mathrm{ri}(X_2)$.
Since X_1 and X_2 can be properly separated, so by definition, there exists $w \neq 0$ such that

$$\begin{cases} \sup\limits_{x_1 \in X_1} w^T x_1 \leq \inf\limits_{x_2 \in X_2} w^T x_2 \\ \inf\limits_{x_1 \in X_1} w^T x_1 < \sup\limits_{x_2 \in X_2} w^T x_2 \end{cases}$$

Then, we have $\sup\limits_{x_1 \in X_1} w^T x_1 = w^T \hat{x}_0 = \inf\limits_{x_2 \in X_2} w^T x_2$.
If there exists $\hat{x}_1 \in X_1$ such that $w^T \hat{x}_1 < w^T \hat{x}_0$, then we put

$$\widehat{x_{1,\varepsilon}} := \hat{x}_0 + (\hat{x}_0 - \hat{x}_0)\varepsilon \in X_1, \quad \text{for } \varepsilon > 0 \text{ small enough}$$



Since $\hat{x}_0 \in \mathrm{ri}(X_1)$, it follows that

$$w^T \widehat{x_{1,\varepsilon}} = w^T \hat{x}_0 + \underbrace{\varepsilon \left(w^T \hat{x}_0 - w^T \hat{x}_0\right)}_{>0} > w^T \hat{x}_0$$

This contradicts the fact that $\sup_{x_1 \in X_1} w^T x_1 = w^T \hat{x}_0$.
Therefore, there is no point $\hat{x}_0 \in X_1$ such that $w^T \hat{x}_0 < w^T \hat{x}_0$. That means

$$w^T x_1 = w^T \hat{x}_0 \quad \text{for all } x_1 \in X_1$$

$$\implies \inf_{x_1 \in X_1} w^T x_1 = w^T \hat{x}_0.$$

Similarly, we also have $\sup_{x_2 \in X_2} w^T x_2 = w^T x_0$.
Therefore, it follows that $\inf_{x_1 \in X_1} w^T x_1 = w^T \hat{x}_0 = \sup_{x_2 \in X_2} w^T x_2$ which contradicts to the proper separation property. Thus, we have $\mathrm{ri}(X_1) \cap \mathrm{ri}(X_2) = \emptyset$. □

Immediately, we have the following corollary.

Corollary 2. *Let $X \subseteq \mathbb{R}^n$ be a non-empty subset and $y \in X \setminus \mathrm{ri}(X)$ (the relative boundary of X). Then, there exists $w \neq 0$ such that*

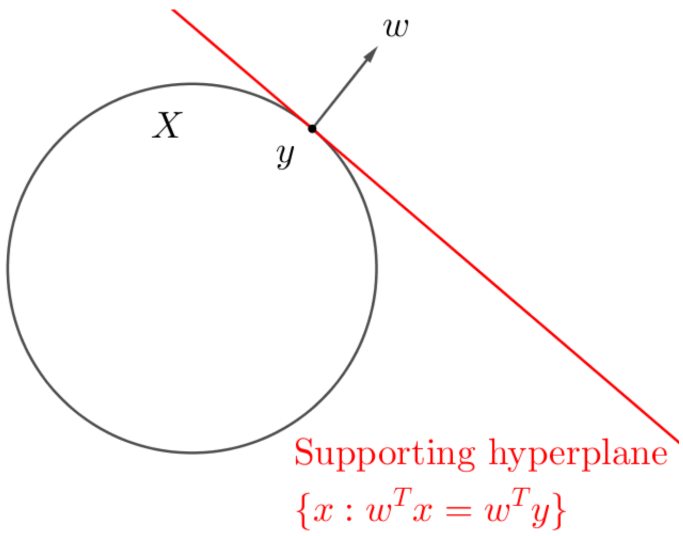
$$\begin{cases} w^T y \geq \sup_{x \in X} w^T x \\ w^T y > \inf_{x \in X} w^T x \end{cases}$$

Proof. We notice that $\{y\}$ and $\mathrm{ri}(X)$ has no intersection, i.e. $\{y\} \cap \mathrm{ri}(X) = \emptyset$, and $\{y\}$ is a point, so we may replace $\{y\} = \mathrm{ri}(\{y\})$. By the previous theorem, $\{y\}$ and X can be properly separated, so there exists $w \neq 0$ such that

$$\begin{cases} \sup_{x \in X} w^T x \leq w^T y \\ \inf_{x \in X} w^T x < w^T y \end{cases}$$

and the proof is thus completed. □

Remarks. Let $y \in X \setminus \mathrm{ri}(X)$, and w be the separation vector.
Then the hyperplane $\{x \in \mathbb{R}^n : w^T x = w^T y\}$ is called a **supporting hyperplane** of X at y.



It is the end of our discussion on separation of convex sets. Let us introduce the new section - convex function.

2 Convex Function

Definition 2. Let X be a nonempty convex subset of $\mathbb{R}^n$, and

1. the function $f : X \to \mathbb{R}$ is called a **convex** function if
$$\lambda f(x) + (1-\lambda)f(y) \geq f\left(\lambda x + (1-\lambda)y\right)$$
for all $x, y \in X$ and $\lambda \in (0,1)$.
2. the function $f : X \to \mathbb{R}$ is called a **strictly convex** function if
$$\lambda f(x) + (1-\lambda)f(y) > f\left(\lambda x + (1-\lambda)y\right)$$
for all $x \neq y$, $x, y \in X$ and $\lambda \in (0,1)$.
3. the function $f : X \to \mathbb{R}$ is called a **(strictly) concave** if $-f(x)$ is (strictly) convex.

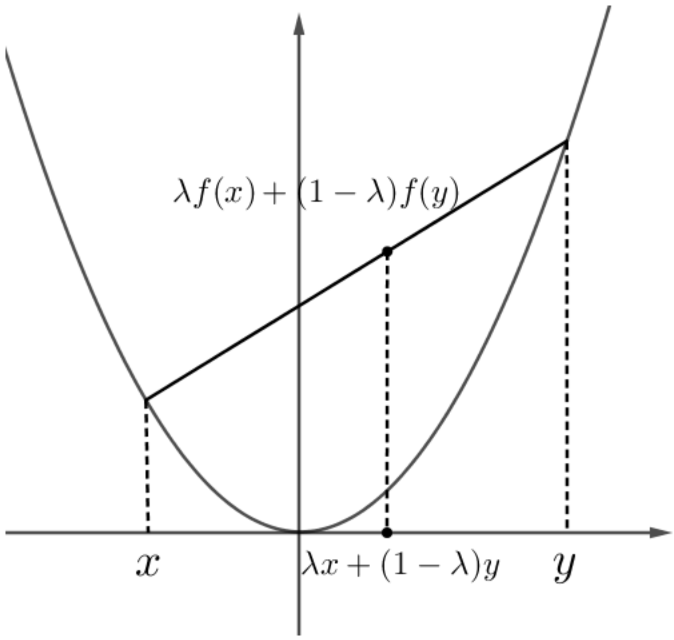


Figure 1: Convex Function

Lemma 3. *Let $X \subseteq \mathbb{R}^n$ be non-empty convex set.*

1. *If $f : X \to \mathbb{R}$ is **differentiable**, then*
$$f(x) \ \textit{is convex} \iff f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad \forall x, y \in X$$
2. *If $f : X \to \mathbb{R}$ is **twice differentiable**, then*
$$f(x) \ \textit{is convex} \iff \nabla^2 f(x) \geq 0, \ \forall x \in X$$
where $\nabla f^2(x) \geq 0$ denotes the Hessian matrix of f is positive semidefinite.

Proof. 1. " $\implies$ " By the convexity of f, then

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$$

for any $x, y \in X,\ \lambda \in [0,1]$. Then

$$\begin{aligned} f(y) &\ge \frac{f(\lambda x + (1-\lambda)y) - \lambda f(x)}{1-\lambda} - f(x) + f(x) \\ &= f(x) + \frac{f(\lambda x + (1-\lambda)y) - (\lambda + (1-\lambda)) f(x)}{1-\lambda} \\ &= f(x) + \frac{f(x + (1-\lambda)(y-x)) - f(x)}{1-\lambda} \end{aligned}$$

Taking limit $\lambda \to 1^-$, we have

$$f(y) \ge f(x) + \lim_{\lambda \to 1^-} \frac{f(x + (1-\lambda)(y-x)) - f(x)}{1-\lambda} = f(x) + \langle \nabla f(x), y - x \rangle$$

" $\impliedby$ " Let $x, y \in X$, $\lambda \in (0,1)$. We define $z := \lambda x + (1-\lambda) y \in X$.
Then, we have

$$\begin{cases} f(x) \ge f(z) + \langle \nabla f(z), x - z \rangle & (1) \\ f(y) \ge f(z) + \langle \nabla f(z), y - z \rangle & (2) \end{cases}$$

Multiplying (1) by λ and (2) by $(1-\lambda)$ then sum together yields

$$\begin{aligned} \lambda f(x) + (1-\lambda) f(y) &\ge f(z) + \lambda \langle \nabla f(z), x - z \rangle + (1-\lambda) \langle \nabla f(z), y - z \rangle \\ &= f(z) + \left\langle \nabla f(z), \underbrace{\lambda(x-z) + (1-\lambda)(y-z)}_{=0} \right\rangle \\ &= f(z) \\ &= f(\lambda x + (1-\lambda) y) \end{aligned}$$

Thus by definition, f is convex.

2. " $\implies$ " Suppose that f is convex, then we assume that $\nabla^2 f(x) \ge 0$ for all $x \in X$ is not true.
Then there exists $x \in X$ and $y \in \mathbb{R}^n$ such that $y^T \nabla^2 f(x) y < 0$ (negative definite).
Then, for any $|\varepsilon| > 0$ be small enough, we compute

$$\begin{aligned} f(x + \varepsilon y) &= f(x) + \langle \nabla f(x), \varepsilon y \rangle + \frac{1}{2} \varepsilon^2 \underbrace{y^T \nabla^2 f(x) y}_{<0} + \underbrace{O(\varepsilon^2)}_{\text{small}} \\ &< f(x) + \langle \nabla f(x), \varepsilon y \rangle \end{aligned}$$

for $\varepsilon \to 0$. Contradicts to the statement 1 of the Lemma.
Thus, $\nabla^2 f(x) \ge 0$ for any $x \in X$.
" $\impliedby$ " Suppose that $\nabla^2 f(x) \ge 0$ for any $x \in X$.
Let $x, y \in X$. By the Taylor's expansion, we compute

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} (y-x)^T \nabla^2 f(z) (y-x)$$

for some $z = \theta x + (1-\theta) y$ for some $\theta \in (0,1)$.
By the assumption, so $f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle$.
Thus f is convex by applying statement 1 of the Lemma.

□

Example 1. The following are examples of convex/concave functions.

1. Quadratic $f(x) = x^2$ satisfies $f''(x) = 2 > 0$, so it is a convex function.
2. Exponential $f(x) = e^{ax}$ satisfies $f''(x) = a^2 e^{ax} \geq 0$, so it is convex function.
3. Logarithm $f(x) = \log(x)$ is concave on $\mathbb{R}_+$ because $f''(x) = -1/x^2 < 0$.
4. Affine function $f(x) = w^T x + b$ is concave and convex (by definition) with $\nabla^2 f(x) = 0$.
5. Quadratic forms $f(x) = \frac{1}{2} x^T A x + b^T x$, where A is $n \times n$ positive semidefinite, then $\nabla^2 f(x) = A \succeq 0$ and thus f is convex.
6. Norms $f(x) = \|x\|$ is convex since
$$\|\lambda x + (1-\lambda) y\| \leq \lambda \|x\| + (1-\lambda)\|y\|.$$
7. Sum of convex functions is still convex:
If f_1 and f_2 are convex, then
$$\begin{cases} f_1(\lambda x + (1-\lambda)y) \leq \lambda f_1(x) + (1-\lambda) f_1(y) \\ f_2(\lambda x + (1-\lambda)y) \leq \lambda f_2(x) + (1-\lambda) f_2(y) \end{cases}$$
This follows that
$$(f_1 + f_2)(\lambda x + (1-\lambda)y) \leq \lambda (f_1 + f_2)(x) + (1-\lambda)(f_1 + f_2)(y).$$

2.1 Non-differentiable Convex Function

Note that convex functions are not always differentiable everywhere over its domain. For example, the absolute value function $f(x) = |x|$ is not differentiable when $x = 0$. In this subsection, we will introduce an important notion about convex functions, i.e., subgradients, to generalize the gradients for differentiable convex functions.

Definition 3. Let X be a convex set and $f : X \to \mathbb{R}$ be a function. A vector $w \in \mathbb{R}^n$ is called a **subgradient** of f at point $x \in X$ if
$$f(y) \geq f(x) + w^T (y - x), \quad \forall y \in X$$
We denote $\partial f(x) = \{\text{all subgradient of } f \text{ at } x\}$, and it is also called the **subdifferential**.

Example 2. Consider $f(x) = |x|$ and $n = 1$.

- When $x > 0$, then $\partial f(x) = \{1\}$.
- When $x < 0$, then $\partial f(x) = \{-1\}$.
- When $x = 0$, then $\partial f(x) = [-1, 1]$. (it is possible that all subgradient may not be unique.)

— End of Lecture 11 —