Lecture 18

§1 Convex functions (凸函数)

1、definition

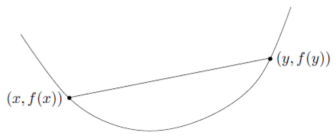
一个函数 $f(x): R^n \to R$ 被称作 convex，若

1° f 的定义域为凸集

2° 对 $\forall \vec{x}, \vec{y} \in \mathrm{dom}(f)$ ($\vec{x}, \vec{y}$ 表示平面(空间)内的两个点)，且 $0 \le \lambda \le 1$，有

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

Geometrically, this means that the line segment between $(x, f(x))$ and $(y, f(y))$ lies **above** the graph of f. **The region above the graph is a convex set.**



例：
- $f(x) = ax + b$ (仿射函数 (affine function) 同时也是凹函数)
- $f(x) = x^2$
- $f(x) = e^x$
- $f(x) = |x|$

2、性质一

$C = \{\vec{x} : f(\vec{x}) \le r\}$ is a convex set if $f(\vec{x})$ is a convex function

(若 $f(\vec{x})$ 为凸函数，则满足 $f(\vec{x}) \le r$ 的所有定义域内的点 $\vec{x}$ 构成的集合 C 为凸集)

证明：

- 在集合 C 中任选两点 $\vec{x}_1, \vec{x}_2$，有

  $\vec{x}_1 : f(\vec{x}_1) \le r$

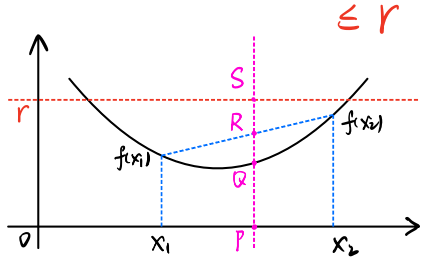
  $\vec{x}_2 : f(\vec{x}_2) \le r$

- 在点 $\vec{x}_1, \vec{x}_2$ 形成的线段上任取一点

  $\lambda \vec{x}_1 + (1-\lambda)\vec{x}_2$

- 由 $f(\vec{x})$ 为凸函数得：

$$f(\lambda \vec{x}_1 + (1-\lambda)\vec{x}_2) \le \lambda f(\vec{x}_1) + (1-\lambda) f(\vec{x}_2) \le r$$



S: r

R: $\lambda f(x_1) + (1-\lambda) f(x_2)$

Q: $f(\lambda x_1 + (1-\lambda) x_2)$

P: $\lambda x_1 + (1-\lambda) x_2$

3、性质二

$C = \{(\vec{x}, y) : y \ge f(\vec{x})\}$ is a convex set if $f(\vec{x})$ is a convex function

(若 $f(\vec{x})$ 为凸函数，则满足 $y \ge f(\vec{x})$ 的所有值域内的点 $(\vec{x}, y)$ 构成的集合 C 为凸集)

证明：

- 在集合C中任选两点 $(\vec{x}_1, y_1)$，$(\vec{x}_2, y_2)$，有
  $(\vec{x}_1, y_1)$：$y_1 \geq f(\vec{x}_1)$
  $(\vec{x}_2, y_2)$：$y_2 \geq f(\vec{x}_2)$
- 在点 $(\vec{x}_1, y_1)$，$(\vec{x}_2, y_2)$ 形成的线段上任取一点
  $\lambda(\vec{x}_1, y_1) + (1-\lambda)(\vec{x}_2, y_2)$
  即 $\langle \lambda \vec{x}_1 + (1-\lambda)\vec{x}_2,\ \lambda y_1 + (1-\lambda) y_2 \rangle$
- 由 $f(x)$ 为凸函数得：
  $$\lambda y_1 + (1-\lambda) y_2 \geq \lambda f(\vec{x}_1) + (1-\lambda) f(\vec{x}_2)$$
  $$\geq f(\lambda \vec{x}_1 + (1-\lambda)\vec{x}_2)$$

§2 Concave functions（凹函数）

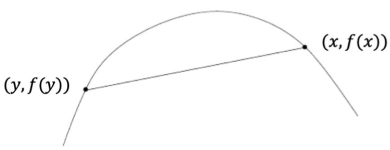
1、definition

一个函数 $f(x)$：$R^n \to R$ 被称作 concave，若

1° f 的定义域为凸集

2° 对 $\forall \vec{x}, \vec{y} \in dom(f)$（$\vec{x}, \vec{y}$ 表示平面(空间)内的两个点），且 $0 \leq \lambda \leq 1$，有
$$f(\lambda \vec{x} + (1-\lambda)\vec{y}) \geq \lambda f(\vec{x}) + (1-\lambda) f(\vec{y})$$

Geometrically, this means that the line segment between $(x, f(x))$ and $(y, f(y))$ lies **below** the graph of f. **The region below the graph is a convex set.**



例：
- $f(x) = -x^2$
- $f(x) = \log(x)$ on $(0, +\infty)$
- $f(x) = \sin(x)$ on $[0, \pi]$

2、性质1

$C = \{(\vec{x}, y) : y \leq f(\vec{x})\}$ is a convex set if $f(\vec{x})$ is a concave function

（若 $f(\vec{x})$ 为凹函数，则满足 $y \leq f(\vec{x})$ 的所有值域内的点 $(\vec{x}, y)$ 构成的集合 C 为凸集）

3、性质2

一个负的凹函数为凸函数，反之亦然

证明：

- 若 f 为一个 concave function，则
  $$f(\lambda \vec{x} + (1-\lambda)\vec{y}) \geq \lambda f(\vec{x}) + (1-\lambda) f(\vec{y})$$
- 令 $g(x) = -f(x)$，则有
  $$g(\lambda \vec{x} + (1-\lambda)\vec{y}) \leq \lambda g(\vec{x}) + (1-\lambda) g(\vec{y})$$

§3 Convex optimization: local minimizer is also global minimizer

1、convex optimization 的前提

1° feasible set 为一个 convex set (凸集)

2° objective function 为一个 convex function (凸函数)

2、convex optimization 的性质

任一 local minimizer 同时也是 global minimizer

Theorem 1. *Consider an optimization problem*

$$\min.\ f(x)$$
$$s.t.\ x \in \Omega,$$

where f is a convex function and Ω is a convex set. Then, any local minimum is also a global minimum.

证明 (反证法):

- Let $\bar{x}$ be a local minimum.

  $\Rightarrow \bar{x} \in \Omega$ and $\exists \epsilon > 0$ s.t. $f(\bar{x}) \le f(x), \forall x \in B(\bar{x}, \epsilon)$

- Suppose for the sake of contradiction that $\exists z \in \Omega$ with $f(z) < f(\bar{x})$
- Because of convexity of Ω, we have

  $$\lambda \bar{x} + (1-\lambda) z \in \Omega, \ \forall \lambda \in [0,1]$$

- By convexity of f, we have

  $$\begin{aligned} f(\lambda \bar{x} + (1-\lambda) z) &\le \lambda f(\bar{x}) + (1-\lambda) f(z) \\ &< \lambda f(\bar{x}) + (1-\lambda) f(\bar{x}) \\ &= f(\bar{x}) \end{aligned}$$

- But, as $\lambda \to 1$, $\lambda \bar{x} + (1-\lambda) z \to \bar{x}$ and the previous inequality contradicts local optimality of $\bar{x}$.

§4 判断 convex function 的方法

1、First order condition (FOC)

若 f 可微,则 f 为凸函数,当且仅当:

1° 定义域为凸集

2° 对定义域内的任意两点 $\vec{x}, \vec{y}$,

- 令 $\vec{e}$ 为 $\vec{x}$ 指向 $\vec{y}$ 的单位方向向量
- 令 $\vec{y} = \vec{x} + s\vec{e}$

有 $$f(\vec{y}) \ge f(\vec{x}) + s \cdot \left.\frac{d f(\vec{x} + \theta \vec{e})}{d\theta}\right|_{\theta=0}$$

* one dimension: $f(y) \ge f(x) + (y-x) f'(x)$

证明:

Necessity

$$\boldsymbol{y} = \boldsymbol{x} + s\boldsymbol{e}$$
$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + s\frac{d\, f(\boldsymbol{x}+\theta \boldsymbol{e})}{d\theta}|_{\theta=0}$$

For any $\boldsymbol{x}, \boldsymbol{y} \in$ **dom f,**

- denote $\boldsymbol{e}$ as the unit vector from $\boldsymbol{x}$ to $\boldsymbol{y}$
- let $\boldsymbol{y} = \boldsymbol{x} + s\,\boldsymbol{e}$

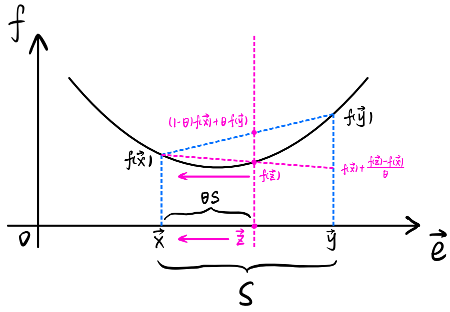
Let $\boldsymbol{z} = (1-\theta)\boldsymbol{x} + \theta\boldsymbol{y} = \boldsymbol{x} + \theta\, s\, \boldsymbol{e}$, by convexity of f, we have

$$(1-\theta)\, f(\boldsymbol{x}) + \theta f(\boldsymbol{y}) \geq f(\boldsymbol{z})$$

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \frac{f(\boldsymbol{Z})-f(\boldsymbol{x})}{\theta} = f(\boldsymbol{x}) + \boxed{\frac{f(\boldsymbol{x}+\theta\, s\, \boldsymbol{e})-f(\boldsymbol{x})}{\theta}}$$

When $\theta \longrightarrow 0$,

$$s\frac{d\, f(\boldsymbol{x}+\theta \boldsymbol{e})}{d\theta}|_{\theta=0}$$



Sufficiency

$$\boldsymbol{y} = \boldsymbol{x} + s\boldsymbol{e}$$
$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + s\frac{d\, f(\boldsymbol{x}+\theta \boldsymbol{e})}{d\theta}|_{\theta=0}$$

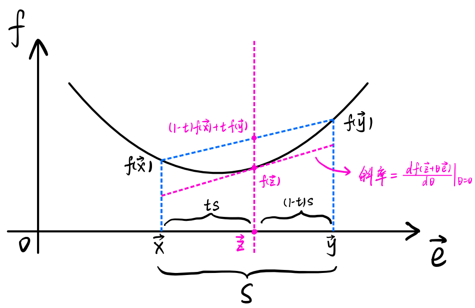
- Denote $\boldsymbol{e}$ as the unit vector from $\boldsymbol{x}$ to $\boldsymbol{y}$
- Let $\boldsymbol{y} = \boldsymbol{x} + s\,\boldsymbol{e}$
- Let $\boldsymbol{z} = (1-t)\boldsymbol{x} + t\,\boldsymbol{y} = \boldsymbol{x} + t\, s\, \boldsymbol{e}$,

As a result, $\boldsymbol{x} = \boldsymbol{z} - t\, s\, \boldsymbol{e}$, $\boldsymbol{y} = \boldsymbol{z} + (s - t\, s)\, \boldsymbol{e}$,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{z}) - ts\frac{d\, f(\boldsymbol{z}+\theta\boldsymbol{e})}{d\theta}|_{\theta=0} \text{ and } f(\boldsymbol{y}) \geq f(\boldsymbol{z}) + s(1-t)\frac{d\, f(\boldsymbol{z}+\theta\boldsymbol{e})}{d\theta}|_{\theta=0}$$

Multiplying the first inequality by 1-t, the second by t, and adding them yields

$$(1\text{-}t)\, f(\boldsymbol{x}) + t\, f(\boldsymbol{y}) \geq f(\boldsymbol{z})$$



2、Second order condition (SDC)

若f可微，则f为凸函数，当且仅当:

1° 定义域为凸集

2° 对定义域内的任意一点 $\vec{x}$，和任意 θ

· 令 $\vec{e}$ 为任意单位方向向量

有 $\dfrac{d^2 f(\vec{x}+\theta\vec{e})}{d\theta\, d\theta} \geq 0$

* one dimension: $f''(x) \geq 0$ for all x

证明:

Necessity

$$\boldsymbol{y} = \boldsymbol{x} + s\boldsymbol{e}$$
$$g(\theta; \boldsymbol{x}, \boldsymbol{e}) = f(\boldsymbol{x}+\theta\boldsymbol{e})$$
$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + s g'(0; \boldsymbol{x}, \boldsymbol{e})$$

For any $\boldsymbol{x}$ and θ,

- let $\boldsymbol{y} = \boldsymbol{x} + (s+\theta)\boldsymbol{e}$ and $\boldsymbol{z} = \boldsymbol{x} + \theta\, \boldsymbol{e}$ with s >0
- $\boldsymbol{y} = \boldsymbol{z} + s\, \boldsymbol{e}$ and $\boldsymbol{z} = \boldsymbol{y} - s\, \boldsymbol{e}$

By FOC of f, we have

$$f(\boldsymbol{y}) \geq f(\boldsymbol{z}) + s\, g'(0; \boldsymbol{z}, \boldsymbol{e}),\, f(\boldsymbol{z}) \geq f(\boldsymbol{y}) - s g'(0; \boldsymbol{y}, \boldsymbol{e})$$

Sum both sides together,

$$g'(0; \boldsymbol{y}, \boldsymbol{e}) - g'(0; \boldsymbol{z}, \boldsymbol{e}) \geq 0, \text{ then } \frac{g'(0; \boldsymbol{y}, \boldsymbol{e}) - g'(0; \boldsymbol{z}, \boldsymbol{e})}{s} \geq 0$$

As $g(t; \boldsymbol{y}, \boldsymbol{e}) = f(\boldsymbol{y}+t\boldsymbol{e}) = f(\boldsymbol{x}+(t+s+\theta)\boldsymbol{e}) = g(t+s+\theta; \boldsymbol{x}, \boldsymbol{e})$, $g'(t; \boldsymbol{y}, \boldsymbol{e}) = g'(t+s+\theta; \boldsymbol{x}, \boldsymbol{e})$. Similarly, $g'(t; \boldsymbol{z}, \boldsymbol{e}) = g'(t+\theta; \boldsymbol{x}, \boldsymbol{e})$. Accordingly,

$$0 \leq \frac{g'(0; \boldsymbol{y}, \boldsymbol{e}) - g'(0; \boldsymbol{z}, \boldsymbol{e})}{s} = \frac{g'(s+\theta; \boldsymbol{x}, \boldsymbol{e}) - g'(\theta; \boldsymbol{x}, \boldsymbol{e})}{s}.$$

When s is approaching 0, $g''(\theta; \boldsymbol{x}, \boldsymbol{e}) \geq 0$

Sufficiency

$\boldsymbol{y} = \boldsymbol{x} + s\boldsymbol{e}$
$g(\theta; \boldsymbol{x}, \boldsymbol{e}) = f(\boldsymbol{x} + \theta\boldsymbol{e})$
$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + sg'(0; \boldsymbol{x}, \boldsymbol{e})$

- Denote $\boldsymbol{e}$ as the unit vector from $\boldsymbol{x}$ to $\boldsymbol{y}$ (chosen arbitrarily)
- Let $\boldsymbol{y} = \boldsymbol{x} + s\,\boldsymbol{e}$, then

$$0 \leq \int_0^s (s-t)\, g''(t; \boldsymbol{x}, e)\, dt$$
$$= -s\, g'(0; \boldsymbol{x}, e) + \int_0^s g'(t; \boldsymbol{x}, e)\, dt$$
$$= -s\, g'(0; \boldsymbol{x}, e) + g(s; \boldsymbol{x}, e) - g(0; \boldsymbol{x}, e)$$
$$= -s\, g'(0; \boldsymbol{x}, e) + f(\boldsymbol{y}) - f(\boldsymbol{x})$$

$$\int_a^b u(x)v'(x)\,dx = \Big[u(x)v(x)\Big]_a^b - \int_a^b u'(x)v(x)\,dx$$
$$= u(b)v(b) - u(a)v(a) - \int_a^b u'(x)v(x)\,dx.$$

Integration by parts

- Then $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + s\, g'(0; \boldsymbol{x}, e)$

3、其他表示方法

In the reference book, it uses the matrix representation. Since you haven't learned linear algebra, no need to know.

FOC: Suppose f is differentiable (*i.e.*, its gradient ∇f exists at each point in **dom** f, which is open). Then f is convex if and only if **dom** f is convex and

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \qquad (3.2)$$

holds for all $x,\ y \in$ **dom** f. This inequality is illustrated in figure 3.2.

SOC: We now assume that f is twice differentiable, that is, its *Hessian* or second derivative $\nabla^2 f$ exists at each point in **dom** f, which is open. Then f is convex if and only if **dom** f is convex and its Hessian is positive semidefinite: for all $x \in$ **dom** f,

$$\nabla^2 f(x) \succeq 0.$$

The proofs of FOC and SOC are not required, you can just remember their connections with convexity.