a. 证明若函数 f 二次可微，则 f 为凸函数的充要条件为:

- $\operatorname{dom} f$ 为凸集
- $\nabla^2 f(x) \succeq 0$, for all $x \in \operatorname{dom} f$

1. 必要性 $\forall d \in \mathbb{R}^n$

$$f(x+td) \geq f(x) + \nabla f(x)^T(td)$$

$$f(x+td) = \cancel{f(x)} + \nabla f(x)^T(td) + \frac{1}{2}(td)^T \nabla^2 f(x)(td) + o(\|td\|^2)$$

$$\geq \cancel{f(x)} + \nabla f(x)^T(td)$$

$$\Rightarrow \frac{1}{2}(td)^T \nabla^2 f(x)(td) + o(\|td\|^2) \geq 0$$

$$\frac{1}{2} d^T \nabla^2 f(x) d + \frac{o(\|td\|^2)}{t^2} \geq 0$$

$$\Rightarrow \lim_{t\to 0}\left(\frac{1}{2} d^T \nabla^2 f(x) d + \frac{o(\|td\|^2)}{t^2}\right) \geq 0$$

$$\Rightarrow d^T \nabla^2 f(x) d \geq 0$$

$\nabla^2 f(x)$ 是对称半正定矩阵

且 ∵ f 凸函数 ∴ $\operatorname{dom} f$ 凸集

2. 充分性

$$\nabla^2 f(x) \in S_+^n$$

$$g(t) = f(x+t(y-x))$$

$$g'(t) = \nabla f(x+t(y-x))^T(y-x)$$

$$g''(t) = (y-x)^T \nabla^2 f(x+t(y-x))(y-x)$$

∴ $g''(t) \geq 0$

对 $\forall x_1 < x_2$

$$\exists \xi \in [x_1, x_2] \quad g'(\xi) = \frac{g(x_1) - g(x_2)}{x_1 - x_2}$$

$g''(t) \geq 0$ ∴ $g'(\xi) \leq g'(x_2)$

∴ $g(x_1) - g(x_2) \geq g'(x_2)(x_1 - x_2)$

$\Leftrightarrow g(x_1) \geq g(x_2) + g'(x_2)(x_1 - x_2)$

∴ g 是凸函数

g 凸性与 f 等价 ∴ f 也是凸函数

b. 闭集与凸集

b.1. 证明多面体 $\{x : Ax \le b\}, A \in \mathbb{R}^{mn}, b \in \mathbb{R}^m$ 为闭的凸集

b.2. 举例：$\mathbb{R}^2$ 上闭集的凸包不一定是闭的

b.3. 若 $A \in \mathbb{R}^{m\times n}, S \subseteq \mathbb{R}^m$ 为凸集，证明集合 S 在 A 下的原像 $\{x \in \mathbb{R}^n : Ax \in S\}$ 是凸集

b.4. 若 $A \in \mathbb{R}^{m\times n}, S \subseteq \mathbb{R}^n$ 为凸集，证明集合 S 在 A 下的像 $\{Ax : x \in S\}$ 是凸集

b.5. 举例：存在 $A \in \mathbb{R}^{m\times n}$ 及闭凸集 $S \subseteq \mathbb{R}^n$，使得 $A(S)$ 不是闭集

b.1

令 $C=\{x: Ax\le b\}$

$\forall x, y \in C$, $\theta\in[0,1]$

$Ax\le b$ $Ay\le b$

说 $A(\theta x+(1-\theta)y)$

$=\theta Ax+(1-\theta)Ay$

$\le \theta b+(1-\theta)b=b$

且集合所有聚点属于 C 说明闭集

b.2

这两个闭集的并组成的闭集

$S=\{(0,0)\}\cup\{(x,y): xy\ge 1, x\ge 0, y\ge 0\}$

$\mathrm{conv}(S)=\{(0,0)\}\cup \mathbb{R}^2_{++}$

b.3

对 x s.t. $Ax\in S$

y $Ay\in S$

则 $\theta Ax+(1-\theta)Ay\in S$

$A(\theta x+(1-\theta)y)\in S$

说 $\theta x+(1-\theta)y$ 也是 S 在 A 下原像

b.4 令 $C=\{Ax: x\in S\}$

$Ax\in C$ $Ay\in C$ $y\in S$

则 $\theta x+(1-\theta)y\in S$

$A(\theta x+(1-\theta)y)=\theta Ax+(1-\theta)Ay\in C$

b.5 $A=(1,\alpha)$ α 无理数 $A(S)=\{x_1+\alpha x_2 : x_1\in[0,1], x_2\in\mathbb{Z}\}$

$S=\{(x_1,x_2)\in\mathbb{R}^2 : 0\le x_1\le 1, x_2\in\mathbb{Z}\}$

c. 考虑二次不等式

$$x^\top Ax + b^\top x + c \le 0, \tag{1}$$

其中 A 为 n 阶对称矩阵，设 C 为上述不等式的解集.

(a) 证明：当 A 正定时，C 为凸集；

(b) 设 C' 是 C 和超平面 $g^\top x + h = 0$ 的交集 ($g \neq 0$)，若存在 $\lambda \in \mathbb{R}$，使得 $A + \lambda gg^\top$ 正定，证明 C' 为凸集.

仿射函数

(a) $f(x) = x^\top A x + \underbrace{b^\top x + c}_{\text{仿射函数}}$ 是凸函数 ∵ $x^\top A x = \mathrm{Tr}(x^\top A x) = \mathrm{Tr}(Axx^\top) = \langle A, xx^\top \rangle$

$f(x)$ 映射 ≤ 0 像是凸集 ∴ 原像也是凸集

(b) $\lambda(x^\top gg^\top x + hx^\top g) = 0$ $\quad$ $x^\top(A+\lambda gg^\top)x + (b^\top + \lambda h g^\top)x + c \le 0$ 由(1)知 满足这个的 C' 凸集

c. 多面体

c.1. 证明：若 $P \subseteq \mathbb{R}^n$ 为多面体，$A \in \mathbb{R}^{m\times n}$，则 $A(P)$ 为多面体，提示：可使用以下事实：

$$P \subseteq \mathbb{R}^{m+n} \text{ 为多面体} \Rightarrow \{x \in \mathbb{R}^n : (x, y) \in P \text{ for some } y \in \mathbb{R}^m\} \text{ 是多面体}$$

c.2. 证明若 $Q \subseteq \mathbb{R}^m$ 为多面体，$A \in \mathbb{R}^{m\times n}$，则 $A^{-1}(Q)$ 为多面体

c.1 仿射变换有保凸性，∴ 像是多面体

c.2 由 c.1 知 逆映射下为多面体 (凸集在凸函数下原像)

凹函数

a. 证明熵函数:

$$f(x) = -\sum_{i=1}^{n} x_i \log(x_i), \quad \mathrm{dom}\, f = \{x \in \mathbb{R}^n_{++} : \sum_{i=1}^{n} x_i = 1\}$$

是严格凹的。

证: $g(x)=x\ln x$ $\quad g'(x)=\ln x+1$ $\quad g''(x)=\frac{1}{x}$ $\quad x\in(0,1]$ $\quad \therefore g''(x)>0$

$g(x)$ 严格凸 $\quad \therefore f(x)$ 严格凹

b. 若 f 为二次可微函数且 $\mathrm{dom}\, f$ 为凸集，证明 f 为凸函数的充要条件为:

$$(\nabla f(x) - \nabla f(y))^\top (x-y) \geq 0, \quad \forall x, y$$

这被称为梯度 ∇f 的单调性。

必要性:

$f(y) \geq f(x) + \nabla f(x)^T (y-x)$ ①

$f(x) \geq f(y) + \nabla f(y)^T (x-y)$ ②

①+② 得 $(\nabla f(x) - \nabla f(y))^T (x-y) \geq 0$

充分性:

定义 g: $g(t)=f(x+t(y-x))$, $t\in[0,1]$

$g'(t) = \nabla f(x+t(y-x))^T (y-x)$

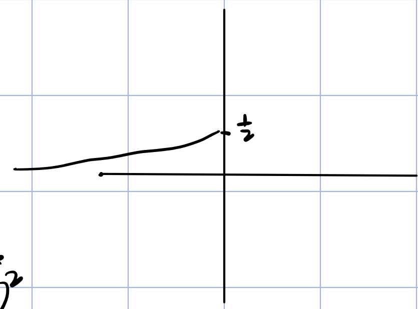
由 $g'(1)-g'(0)\geq 0$ 且 $g'(0)-g'(0)=0$ 得

$g'(t)-g'(0)\geq 0$

$\therefore f(y)=g(1)=g(0)+\int_0^1 g'(t)\,dt \geq g(0)+g'(0)$

$= f(x)+\nabla f(x)^T (y-x)$

c. 举例：严格凸函数并不一定能达到其极小值。



$$f(x)=\frac{1}{1+e^{-x}}$$

$$f'(x)=\frac{-e^{-x}}{(1+e^{-x})^2}=\frac{-e^{-x}-1+1}{(1+e^{-x})^2}=-\frac{1}{1+e^{-x}}+\frac{1}{(1+e^{-x})^2}$$

$$f''(x)=\frac{e^{-x}}{(1+e^{-x})^2}+\frac{2e^{-x}}{(1+e^{-x})^4}$$

在 $[-\infty,0]$ ≥ 0

无法到达极小值

c. 举例：严格凸函数并不一定能达到其极小值。

d. 函数 $f:\mathbb{R}^n\to\mathbb{R}$ 被称为强制的（coercive），如果当 $\|x\|_2\to\infty$ 时，有 $f\to\infty$。强制函数的一个关键事实为其可以达到极小值。证明对于一个二次可微的强凸函数是强制的（coercive），并因此可达到极小值。

设 f 的 Hessian 矩阵为 $\nabla^2 f(x)$ 则 $\exists m>0$ s.t. 对所有 $x\in\mathbb{R}^n$

$\nabla^2 f(x)\succeq mI$ $\quad f(y)\approx f(x)+\nabla f(x)^T(y-x)+\frac{1}{2}(y-x)^T\nabla^2 f(x)(y-x)$

$\therefore f(y)\geq f(x)+\nabla f(x)^T(y-x)+\frac{m}{2}\|y-x\|^2$ $\quad \|y\|\to\infty$

$f(y)\to\infty$

$\therefore$ 有极小值