

1. 定义^①: 一个函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为凸函数 \Leftrightarrow

o $\text{dom } f$ 为凸集.

o 且对所有的 $x, y \in \text{dom } f$, $0 \leq \theta \leq 1$, 有 $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$

定义^②: (适用于高维情况) 证明西瓜皮是凸函数, 在西瓜上切无数刀(很冗余)

$\forall x \in \text{dom } f$, $\forall v$ ^{方向}, $g(t) = f(x + tv)$. $\text{dom } g = \{t \mid x + tv \in \text{dom } f\}$

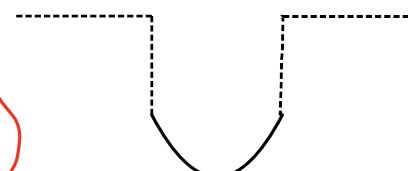
2. 凸函数的扩展

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为凸, $\text{dom } f = C \subseteq \mathbb{R}^n$

$$\hat{f}(x) = \begin{cases} f(x) & x \in \text{dom } f \\ +\infty & x \notin \text{dom } f \end{cases}$$

\downarrow 仍然是凸函数

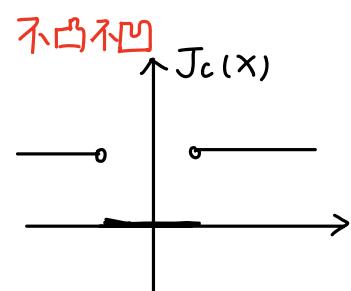
$$\begin{aligned} \hat{f}: \mathbb{R}^n &\rightarrow \mathbb{R} \\ \text{dom } \hat{f} &= \mathbb{R}^n \end{aligned}$$



eg: 凸集的亦性函数是凸函数. 不要扩展到比较小的数
要扩展到 $+\infty$

凸集 $C \subseteq \mathbb{R}^n$. 亦性函数 $f_C(x) = \begin{cases} \text{无定义} & x \notin C \\ 0 & x \in C \end{cases}$

$$I_C(x) = \begin{cases} \infty & x \notin C \\ 0 & x \in C \end{cases}, \quad J_C(x) = \begin{cases} 1 & x \notin C \\ 0 & x \in C \end{cases}$$



3. 一阶条件. ~~凸~~ 凸函数最重要的性质. (定义^③)

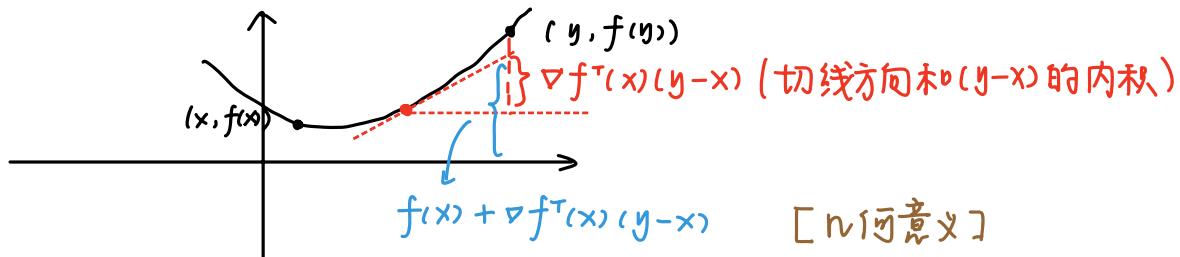
闭集. 不能是开集

设 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 可微, 即梯度 ∇f 在 $\text{dom } f$ 上均存在. 则 f 为凸当且仅

① $\text{dom } f$ 为凸.

~~一阶偏导(切线方向)~~

② $f(y) \geq f(x) + \nabla f^\top(x)(y-x) \quad \forall x, y \in \text{dom } f$



eg: $\exists x_0, \nabla f(x_0) = 0, \forall y, f(y) \geq f(x_0) + \nabla f^\top(x_0)(y-x_0) = f(x_0)$

直观解释：如果凸函数可微，某点一阶偏导为0，则该点为最小值点

构成最优解集.

证明：(一阶条件)

考虑一维情况， $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为凸 $\Leftrightarrow \text{dom } f$ 为凸，且 $f(y) \geq f(x) + f'(x)(y-x)$

证明“ \Rightarrow ” ① 由于 f 为凸函数，则其定义域为凸函数.

② f 为凸， $\forall x, y \in \text{dom } f$ 为凸.

$\forall t, 0 < t < 1, x + t(y-x) \in \text{dom } f$.

$$\text{then } f(x + t(y-x)) \leq (1-t)f(x) + tf(y)$$

$$tf(y) \geq tf(x) + f(x + t(y-x)) - f(x)$$

$$f(y) \geq f(x) + \frac{f(x + t(y-x)) - f(x)}{t}$$

取 $\lim_{t \rightarrow 0^+}$, $f(y) \geq f(x) + f'(x)(y-x)$

出现偏导，则尝试
取极限的技巧

证明“ \Leftarrow ” 设 $\forall x \neq y, x, y \in \text{dom } f, 0 \leq \theta \leq 1$, 构造 $z = \theta x + (1-\theta)y \in \text{dom } f$

$$\begin{cases} f(x) \geq f(z) + f'(z)(x-z) \\ f(y) \geq f(z) + f'(z)(y-z) \end{cases}$$

$$\theta f(x) + (1-\theta)f(y) \geq f(z) + (\theta(x-z) + (1-\theta)(y-z))f'(z) = f(z)$$

考虑高维情况

证明“ \Rightarrow ” 设 f 为凸，考虑 $x, y \in \text{dom } f$,

构造低维函数 $(g(t)) = f(ty + (1-t)x) = f(x + t(y-x))$

方向： $x \rightarrow y$

$$g'(t) = \nabla f^\top(ty + (1-t)x)(y-x)$$

$$g(t_1) \geq g(t_2) + g'(t_2)(t_1 - t_2)$$

$$\text{取 } t_1=0, t_2=1, \underline{\underline{g(1)}} \geq \underline{\underline{g(0)}} + g'(0)$$

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

证明 " \Leftarrow " $\forall x, y \in \text{dom}f, ty + (1-t)x \in \text{dom}f, \tilde{t}y + (1-\tilde{t})x \in \text{dom}f$

$$f(ty + (1-t)x) \geq f(\tilde{t}y + (1-\tilde{t})x)$$

$$+ \nabla f(\tilde{t}y + (1-\tilde{t})x)(\underline{ty + (1-t)x - \tilde{t}y - (1-\tilde{t})x})$$

$(y-x)(t-\tilde{t}) / t(1-t)$

构造低维函数

$$\text{定义 } g(t) = f(ty + (1-t)x), g(\tilde{t}) = f(\tilde{t}y + (1-\tilde{t})x)$$

$$g'(\tilde{t}) = \nabla f^T(\tilde{t}y + (1-\tilde{t})x)(y-x)$$

$g(t) \geq g(\tilde{t}) + g'(\tilde{t})(t-\tilde{t})$. 则 $f(x)$ 为凸函数.

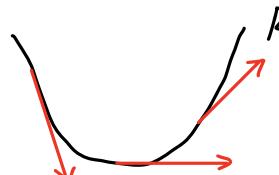
4.(定义④) 二阶条件. 使用最多的条件.

or 一阶偏导单调不减.

Hessian 矩阵半正定

若 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 二阶可微, 则 f 为凸 $\Leftrightarrow \text{dom}f$ 为凸, $\nabla^2 f(x) \geq 0, \forall x \in \text{dom}f$

(铁锅的形状)



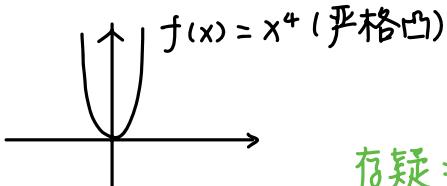
几何意义: 一阶偏导单调不减.

注意: $\nabla^2 f(x) > 0 \Rightarrow$ 严格凸.

\Leftarrow 1 不一定成立, 但是在定义①②③成立)

$$\text{eg: } f(x) = x^4.$$

$$f''(x) = 12x^2$$



存疑: 从图象判断凸或严格凸?

$x=0$ 时, $f''(x)|_{x=0} = 0$, 但是 $f(x) = x^4$ 是严格凸的.

eg: 二次函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $\text{dom } f = \mathbb{R}^n$,

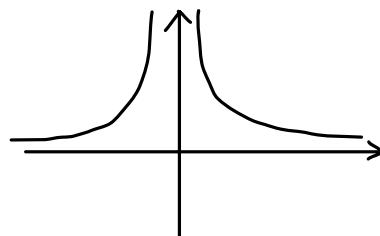
$$f(x) = \frac{1}{2} x^T P x + q^T x + r, \quad P \in \mathbb{S}^n, \quad q \in \mathbb{R}^n, \quad r \in \mathbb{R}$$

$\nabla^2 f(x) = P$ 正定, 则严格凸; 半正定, 则 f 凸函数.

$$\text{eg: } f(x) = \frac{1}{x^2}, \quad x \neq 0, \quad x \in \mathbb{R}$$

$$f''(x) = 6x^{-4} > 0,$$

但定义域不是凸集，故不是凸函数。



eg: 判断函数的凸性。

① 仿射函数 $f(x) = Ax + b$, $\nabla^2 f(x) = 0$ 既凸又凹。

② 指数函数 $f(x) = e^{-x}$, $x \in \mathbb{R}$ ($f(x) = e^{ax}$, $\forall a$ 都是凸函数)

$$f'(x) = ae^{ax}, f''(x) = a^2e^{ax} \geq 0 \Rightarrow \text{凸函数}.$$

③ 幂函数 $f(x) = x^\alpha$, $x \in \mathbb{R}_{++}$

$$f'(x) = \alpha x^{\alpha-1}, f''(x) = \alpha(\alpha-1)x^{\alpha-2}$$

$$\nabla^2 f(x) = \begin{cases} \geq 0, & \alpha \geq 1 \text{ 或 } \alpha \leq 0 \quad \text{凸函数} \\ \leq 0, & 0 < \alpha < 1 \quad \text{凹函数} \end{cases}$$

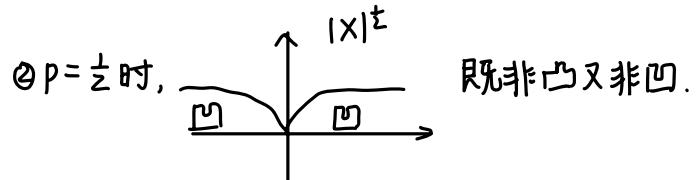
④ 绝对值的幂函数 $f(x) = |x|^p$, $x \in \mathbb{R}$ p 不能等于 1, 因为影响可微性。

$$f'(x) = \begin{cases} p x^{p-1}, & x \geq 0 \\ -p(-x)^{p-1}, & x < 0 \end{cases}, \quad f''(x) = \begin{cases} p(p-1)x^{p-2}, & x \geq 0 \\ p(p-1)x^{p-2}, & x < 0 \end{cases}$$

$p > 2$ 时, 则为凸函数。($p \geq 2$ 时, 2 阶导才有意义)

$p=1$ 时, $f(x) = |x|$  虽然不可微但是凸函数。

$p < 1$ 时, 具体分析: ① $p=0$ 时, 定义 $0^0=1$. 则既凸又凹。



因此在 $p < 1$ 的情况下没有统一的结论。

⑤ 对数函数 $f(x) = \log(x)$, $x \in \mathbb{R}_{++}$

$$f'(x) = \frac{1}{x}, f''(x) = -\frac{1}{x^2} < 0 \Rightarrow \text{严格凹函数}.$$

⑥ 负熵 $f(x) = x \log x$, $x \in R^+$

$$f'(x) = \log x + 1, f''(x) = \frac{1}{x} > 0 \Rightarrow \text{严格凸函数}.$$

⑦ 范数 R^n 空间的范数 $P(x)$, $x \in R^n$.

I. $P(\alpha x) = |\alpha| P(x)$

II. $P(x+y) \leq P(x) + P(y)$

III. $P(x)=0 \Leftrightarrow x=0$

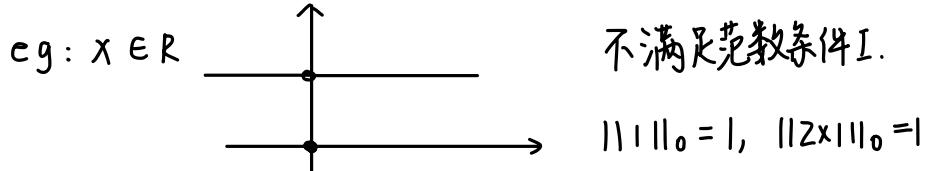
证明: $\forall x, y \in R^n$, $\forall 0 \leq \theta \leq 1$,

$$P(\theta x + (1-\theta)y) \leq P(\theta x) + P((1-\theta)y)$$

$$= \theta P(x) + (1-\theta)P(y)$$

∴ 范数一定是凸函数.

⑧ 零范数 $\|x\|_0$ = 非零元素的数目. (不是凸函数)



⑨ 极大值函数 $f(x) = \max\{x_1, \dots, x_n\}$, $x \in R^n$. (不能用二阶条件)

$$\forall x, y \in R^n, \forall 0 \leq \theta \leq 1$$

$$f(\theta x + (1-\theta)y) = \max\{\theta x_i + (1-\theta)y_i, i=1, \dots, n\}$$

$$\leq \theta \max\{x_i, i=1, \dots, n\} + (1-\theta) \max\{y_i, i=1, \dots, n\}$$

$$= \theta f(x) + (1-\theta)f(y) \Rightarrow \text{凸函数}.$$

引申: 极小-极大问题 $\min_x \max_y f(x, y)$ 凸, 多项式时间可解.

由于不可导, 所以可导近似 (解析逼近) log-sum-exp

$$f(x) = \log(e^{x_1} + \dots + e^{x_n}), x \in R^n$$

近似的最大误差

$$\max\{x_1, \dots, x_n\} \leq f(x) \leq \max\{x_1, \dots, x_n\} + \underbrace{\log n}_{\text{误差}}$$

$$\frac{\partial f}{\partial x_i} = \frac{e^{x_i}}{e^{x_1} + \dots + e^{x_n}}, \quad \text{Hessian} = \begin{bmatrix} \cdots & H_{ij} & \cdots \end{bmatrix}$$

定义 $z = [e^{x_1}, \dots, e^{x_n}]^T$

$$i \neq j \text{ 时}, \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{-e^{x_i} e^{x_j}}{(e^{x_1} + \dots + e^{x_n})^2}$$

$$i=j \text{ 时}, \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{-e^{x_i} e^{x_j} + e^{x_i} (e^{x_1} + \dots + e^{x_n})}{(e^{x_1} + \dots + e^{x_n})^2} I^T z$$

(erg 都是对角线元素)

$$\text{Hessian Matrix} = \frac{1}{(e^{x_1} + \dots + e^{x_n})^2} \left\{ \begin{bmatrix} -e^{x_1(x_1+\dots+x_n)} & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \ddots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & e^{x_n(x_1+\dots+x_n)} \end{bmatrix} - \begin{bmatrix} e^{x_1} \\ \vdots \\ e^{x_n} \end{bmatrix} \begin{bmatrix} e^{x_1} & \dots & e^{x_n} \end{bmatrix}^T \right\}$$

$$\text{定义 } z = [e^{x_1}, \dots, e^{x_n}]^T, \quad H = \frac{1}{(I^T z)^2} (I^T z \text{ diag}\{z\} - z z^T)$$

$$\text{定义 } K = I^T z \text{ diag}\{z\} - z z^T, \quad \text{验证 if } K \in R^{n \times n}$$

$$\text{if } \forall v \in R^n, v^T K v \geq 0$$

$$v^T K v = (I^T z) v^T \text{ diag}\{z\} v - v^T z z^T v$$

$$= (\sum_i z_i) (\sum_i v_i z_i) - (\sum_i v_i z_i)^2$$

Cauchy-Schwarz 不等式.

$$\text{定义 } a_i = v_i \sqrt{z_i}, \quad b_i = \sqrt{z_i}, \quad \text{则 } v^T K v = (b^T b)(a^T a) - (a^T b)^2 \geq 0$$

(遇到范数想三角不等式, 遇到平方想 Cauchy-Schwarz 不等式)

$$\text{⑩ 几何平均函数 } f(x) = (x_1 + \dots + x_n)^{\frac{1}{n}}, \quad x \in R_{++}^n$$

(需证明 $f(x)$ 为凸函数, or $f(x)$ 为凸函数)

$$\text{⑪ 行列式的对数 } f(x) = \log \det(x), \quad \text{dom } f = S_{++}^n$$

(补充: 确保 $x > 0 \Leftrightarrow$ 在对称正定矩阵中)

行列式 = 特征值乘积, 特征值 = 奇异值.

切面的方向 v ,
距离 t , 起点 \bar{z}

证明: $n=1$ 时, 凸函数

\leftarrow (定义) $n>1$ 时, $\forall \bar{z} \in S_{++}^n, \forall t \in \mathbb{R}, \forall v \in \mathbb{R}^{n \times n}$

对称矩阵 (转置等于自身) (特征值分解)

$\lambda_i = \text{矩阵 } (\bar{z}^{-\frac{1}{2}} v \bar{z}^{-\frac{1}{2}})^{-1}$ 的第 i 个特征值

分解: $Q \Lambda Q^T$

① Λ , 对角矩阵, 对角元素 = 特征值

② Q 满足 $Q Q^T = I$ (正交矩阵)

$$\det(I + t \bar{z}^{-\frac{1}{2}} v \bar{z}^{-\frac{1}{2}})$$

$$= \det(Q \Lambda Q^T + Q \Lambda Q^T)$$

$$= \det(Q) \overset{\Lambda^T}{\underset{v}{\det}} (I + \Lambda)$$

$$= \det(I + \Lambda) \text{ 第 } i \text{ 个特征值: } 1 + \lambda_i$$

$$\bar{z} + t v \in S_{++}^n = \text{dom } f, \text{ 故 } v \in \mathbb{R}^{n \times n}$$

$$g(t) = f(\bar{z} + t v) = \log \det(\bar{z} + t v)$$

$$= \log \det \left\{ \bar{z}^{\frac{1}{2}} (I + t \bar{z}^{-\frac{1}{2}} v \bar{z}^{-\frac{1}{2}}) \bar{z}^{\frac{1}{2}} \right\}$$

$$= 2 \log \det \{ \bar{z}^{\frac{1}{2}} \} + \log \det \{ I + t \bar{z}^{-\frac{1}{2}} v \bar{z}^{-\frac{1}{2}} \}$$

$$= \log \det \{ \bar{z} \} + \log \det \{ I + t \bar{z}^{-\frac{1}{2}} v \bar{z}^{-\frac{1}{2}} \}$$

$$= \log \det \{ \bar{z} \} + \sum_{i=1}^n \log(1 + t \lambda_i)$$

$$g'(t) = \sum_i \frac{\lambda_i}{1 + t \lambda_i}, \quad g''(t) = \sum_i \frac{-\lambda_i^2}{(1 + t \lambda_i)^2} \leq 0$$

$\therefore g(x)$ 是凸函数

$\therefore f(x)$ 是凸函数.

5. 保凸运算

(1) 非负加权和

① f_1, \dots, f_m 为凸, 若 $w_i \geq 0, \forall i$, 则 $f = \sum_{i=1}^m w_i f_i$ 为凸

不一定是凸函数

② 若 $f(x, y)$, 对任意 $y \in A$, 则 $f(x, y)$ 均为凸.

(x, y) joint convex

设 $w(y) \geq 0, \forall y \in A, g(x) = \int_{y \in A} w(y) f(x, y) dy$ 为凸.

(2) 仿射映射

$f: \mathbb{R}^n \rightarrow \mathbb{R}, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, g(x) = f(Ax + b), \text{ dom } g = \{x \mid Ax + b \in \text{dom } f\}$

证明: $\forall x, y \in \text{dom } g, 0 \leq \theta \leq 1$

$$g(\theta x + (1-\theta)y) = f(\theta Ax + (1-\theta)Ay + b) = f(\theta(Ax + b) + (1-\theta)(Ay + b))$$

$$\leq \theta f(Ax+b) + (1-\theta) f_1(Ax+b) = \theta g(x) + (1-\theta) g(y)$$

• $f_i: \mathbb{R}^n \rightarrow \mathbb{R}$, $i=1, \dots, m$ 为凸, $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, $g(x) = A^\top [f_1(x) \cdots f_m(x)]^\top + b$

\downarrow
实质是非负加权和. 只有 $A^\top \geq 0$ 才能保持凸性.

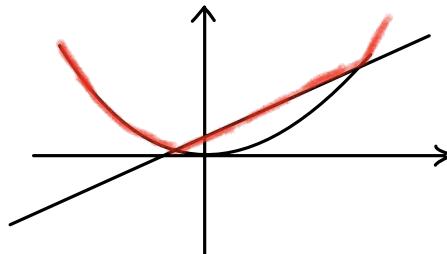
(3) 两个函数的极值函数

f_1, f_2 为凸, 定义 $f(x) = \max\{f_1(x), f_2(x)\}$, $\text{dom } f = \text{dom } f_1 \cap \text{dom } f_2$

证明: $x, y \in \text{dom } f$, $0 \leq \theta \leq 1$ $\max\{a+b, c+d\} \leq \max\{a, c\} + \max\{b, d\}$

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max\{f_1(\theta x + (1-\theta)y), f_2(\theta x + (1-\theta)y)\} \\ &\leq \max\{\theta f_1(x) + (1-\theta)f_1(y), \theta f_2(x) + (1-\theta)f_2(y)\} \\ &\leq \max\{\theta f_1(x), \theta f_2(x)\} + \max\{(1-\theta)f_1(y), (1-\theta)f_2(y)\} \\ &= \theta f(x) + (1-\theta)f(y) \end{aligned}$$

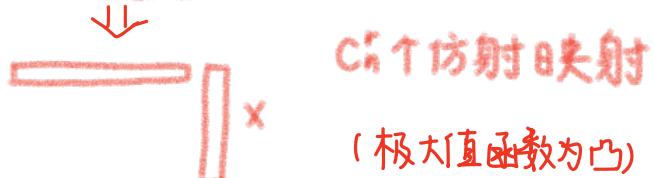
eg: $f(x) = \max\{x^2, x\}$



eg: 向量中 r 个最大元素的和. $x \in \mathbb{R}^n$. 记 $x_{[i]}$ 为第 i 大的元素.

$$x_{[1]} \geq x_{[2]} \geq \dots \geq x_{[r]} \geq \dots \geq x_{[n]}, f(x) = \sum_{i=1}^r x_{[i]}$$

举个例子: $f(x) = \max\{x_1 + \dots + x_r \mid 1 \leq i_1 \dots \leq i_r \leq n\}$



无限个凸函数, 极大值.

$f(x, y)$ 对于 x 为凸, $\forall y \in A$, $g = \sup_{y \in A} f(x, y)$

eg: 实对称矩阵的最大特征值. (结论: 关于 x 的凸函数)

$$f(x) = \lambda_{\max}(x), \text{dom } f = S^n, xy = \lambda y, y^\top x y = y^\top \lambda y$$

$$y^T \times y = \lambda \|y\|_2^2 \Leftrightarrow \lambda = \frac{y^T \times y}{\|y\|_2^2}$$

将y理解成线性项 \Rightarrow 凸函数

$$\lambda = y^T \times y, \|y\|_2 = 1, \lambda_{\max}(x) = \sup \{ \underline{y^T \times y} \mid \|y\|_2 = 1 \}$$

(4) 函数的组合

$h: R^k \rightarrow R, g: R^n \rightarrow R^k$, 组合 $f = h \circ g: R^n \rightarrow R$. $f(x) = h(g(x))$

$$\text{dom } f = \{x \in \text{dom } g \mid g(x) \in \text{dom } h\}$$

① - 维度: $k=n=1$; $\text{dom } g = \text{dom } h = \text{dom } f = R$; h, g 均为二阶可微.

f 为凸 $\Leftrightarrow f''(x) \geq 0$

$$f'(x) = h'(g(x))g'(x), f''(x) = \frac{h''(g(x))g'(x)^2}{\geq 0} + \frac{h'(g(x))g''(x)}{\geq 0} \geq 0$$

I. h 为凸 不降, g 为凸, 则 f 为凸. 假设 $g(x)$ 为凸, $g''(x) \geq 0$

II. h 为凸 不增, g 为凹, 则 f 为凸.

III. h 为凹 不降, g 为凹, 则 f 为凹.

IV. h 为凹 不增, g 为凸, 则 f 为凹.

② 高维 $n, k \geq 1$; $\text{dom } g, \text{dom } h, \text{dom } f \neq R^n, R^k, R^l$; h, k 均不二阶可微

举个例子 (函数的扩展) $h(x) = -\log x$, $\text{dom } h = R^+$, $H(x) = \begin{cases} -\log x, & x > 0 \\ +\infty, & x \leq 0 \end{cases}$

I. h 为凸, \tilde{h} 不降, g 为凸, 则 f 为凸.

II. h 为凸, \tilde{h} 不增, g 为凹, 则 f 为凸.

III. h 为凹, \tilde{h} 不降, g 为凹, 则 f 为凹.

IV. h 为凹, \tilde{h} 不增, g 为凸, 则 f 为凹.

证明: $\forall x, y \in \text{dom } f, 0 \leq \theta \leq 1$ ($f = h \circ g \Rightarrow x, y \in \text{dom } g, g(x), g(y) \in \text{dom } h$)

g 为凸, 故 $\text{dom } g$ 为凸, $\theta x + (1-\theta)y \in \text{dom } g$

$$g(\theta x + (1-\theta)y) \leq \theta g(x) + (1-\theta)g(y)$$

h 为凸, 故 $\text{dom } h$ 为凸, $\theta g(x) + (1-\theta)g(y) \in \text{dom } h$

$$\begin{aligned} f(\theta x + (1-\theta)y) &\leq h(\theta g(x) + (1-\theta)g(y)) \leq \\ &= h(g(\theta x + (1-\theta)y)) = \theta h(g(x)) + (1-\theta)h(g(y)) \end{aligned}$$

证明: $g(\theta x + (1-\theta)y) \in \text{dom } h$

(反证) $g(\theta x + (1-\theta)y) \notin \text{dom } h$, 由于 \hat{h} 不降

$$\hat{h}(g(\theta x + (1-\theta)y)) \leq \hat{h}(\theta g(x) + (1-\theta)g(y))$$

→ (可以用 h 的单调性)

$$h(g(\theta x + (1-\theta)y)) \leq h(\theta g(x) + (1-\theta)g(y))$$

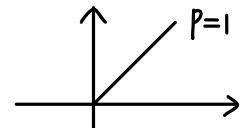
所以 $f(x)$ 一定 是 凸 函数.

eg: 若 g 为凸, $\exp\{g(x)\}$ 为凸 $h(z) = \exp(z)$

若 g 为凹, $g > 0$, $\log\{g(x)\}$ 为凹.

若 g 为凹, $g > 0$, $1/g(x)$ 为凸. $h(z) = \frac{1}{z}$

若 g 为凸, $g \geq 0$, $p \geq 1$, $g^p(x)$ 为凸. $h(z) = z^p$



☆乔布斯在斯坦福大学的演讲. Make it great.

(5) 函数的透视. 丰透视函数 $P: R^{n+1} \rightarrow R^n$, $\text{dom } P \subseteq R^n \times R_{++}$, $P(z, t) = \frac{z}{t}$.

$f: R^n \rightarrow R$, $g: R^n \times R_{++} \rightarrow R$, $g(x, t) = t f(\frac{x}{t})$ (g 是 f 的透视)

$$\text{dom } g = \left\{ (x, t) \mid t > 0, \frac{x}{t} \in \text{dom } f \right\}$$

如果 f 为凸, 则 g 为凸; 如果 f 为凹, 则 g 为凹.

(课下证明: ① 定义域是凸集 ② 对 (x, t) 进行凸组合的不等式)

x, t joint convex

eg: 欧氏范数的平方. $f(x) = x^\top x$, $\text{dom } f = R^n$.

$$g(x, t) = t \left(\frac{x}{t} \right)^\top \left(\frac{x}{t} \right) = \frac{x^\top x}{t} \quad (\text{因为 } t \text{ 是 梯量})$$

eg: 负的对数. $f(x) = -\log x$, $\text{dom } f = \mathbb{R}_{++}$

$$g(x, t) = t(-\log \frac{x}{t}) = t \log \frac{t}{x}, \quad \text{dom } g = \mathbb{R}_{++}^2$$

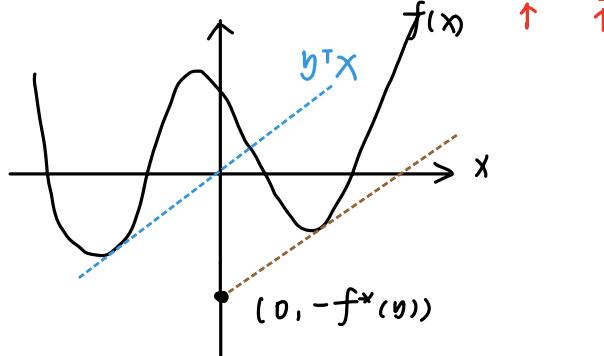
扩展: $u, v \in \mathbb{R}_{++}^n$, $g(u, v) = \sum_{i=1}^n u_i \log \frac{u_i}{v_i}$ 凸.

(不能用非负加权和考虑 $\sum_{i=1}^n g_i(u_i, v_i)$, $\sum_{i=1}^n g_i(u, v)$ 才可以)

扩展: $D_{KL}(u, v) \triangleq \sum_{i=1}^n (u_i \log \frac{u_i}{v_i} - u_i + v_i)$ 凸(非负加权和)

6. 函数的共轭 Conjugate

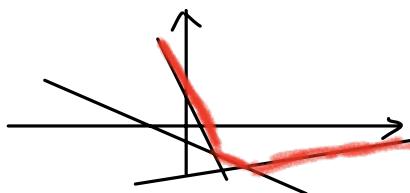
$$f: \mathbb{R}^n \rightarrow \mathbb{R}, \quad f^*: \mathbb{R}^n \rightarrow \mathbb{R}, \quad f^*(y) = \sup_{x \in \text{dom } f} (y^T x - f(x)) \quad \text{选定 } y, \text{ 遍历 } x \text{ 求最大值.}$$



性质: (1) $f(x)$ 若可微, 则 $f^*(y)$ 对应的 x 必是 $f'(x) = y$ 的一点. ($y - f'(x) = 0$)

(2) 函数的共轭一定为凸函数. (极大值的分段线性函数)

↙
(对偶原理)



分割线

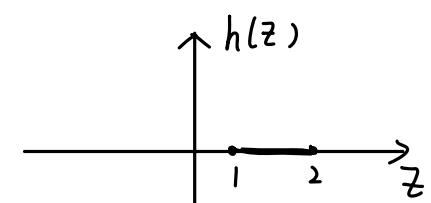
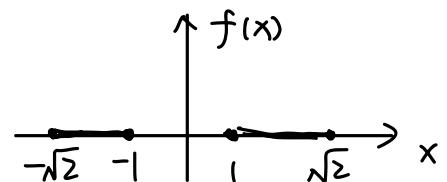
NOTES: (函数的组合-补充) \hat{h} 的单调性.

$$g(x) = x^2, \quad \text{dom } g \in \mathbb{R}, \quad \text{凸}$$

$$h(z) = 0, \quad \text{dom } h = [1, 2], \quad \text{凸. 不增不降}$$

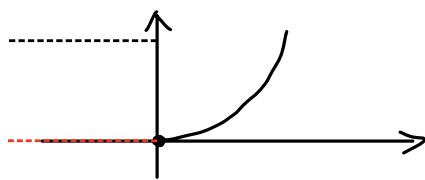
$$f = g(h(x)) = \begin{cases} 0 & x \in [-\sqrt{2}, -1] \cup [1, \sqrt{2}] \\ \text{不存在} & \text{else} \end{cases}$$

非凸非凹(看定义域)



\hat{h} 在 \mathbb{R}^n 都带有单调性.

INCORRECT (函数的组合-补充) eg: 若 g 为凸, $P \geq 1$, 则 $g^P(x)$ 为凸. $h(z) = z^P$, $\text{dom}h = \mathbb{R}_+$



$$g \geq 0$$

$$h(z) = \begin{cases} z^P, & z \in \mathbb{R}_+ \\ 0, & z \in \mathbb{R}_{-} \end{cases}$$

单调不降, 使用第一条规则.

分割线

$$\text{eg: } f(x) = Ax + b, \text{ dom}f = \mathbb{R}. \text{ 求 } f^*(y) = \sup_{x \in \text{dom}f} (yx - (ax + b))$$

$$f^*(y) = \sup_{x \in \text{dom}f} (yx - (ax + b)) = \sup_{x \in \text{dom}f} ((y-a)x - b) = \begin{cases} -b & y=a \\ +\infty & y \neq a \end{cases}$$

$y + \frac{1}{x} = 0$
 $x = -\frac{1}{y}$

$$\text{eg: } f(x) = -\log x, \text{ dom}f = \mathbb{R}_{++}, f^*(y) = \sup_{x>0} (yx + \log x) = \begin{cases} -1 - \log(-y), & y < 0 \\ +\infty & y \geq 0 \end{cases}$$

$$\text{eg: } f(x) = \frac{1}{2} x^T Q x, Q \in S^n_+, \text{ dom}f = \mathbb{R}^n. \quad \text{正定, 奇异值=特征值, 可逆} \Rightarrow x = Q^{-1}y$$

$$f^*(y) = \sup_x (y^T x - \frac{1}{2} x^T Q x) \xrightarrow{\frac{\partial(y^T x - \frac{1}{2} x^T Q x)}{\partial x}} = y^T Q^{-1} y \quad (\text{书附录矩阵计算})$$

$$= y^T Q^{-1} y - \frac{1}{2} y^T (Q^{-1})^T Q Q^{-1} y = y^T Q^{-1} y - \frac{1}{2} y^T Q^{-1} y = \frac{1}{2} y^T Q^{-1} y$$

REMINDER 若 f 非凸, $f^{**} \neq f$; 若 f 凸, 且为闭函数, 则 $f^{**} = f$.

7. 凸集与凸函数的关系 δ -sublevel set

若 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 定义其 δ -sublevel set 为 $C_\delta = \{x \in \text{dom}f \mid f(x) \leq \delta\}$

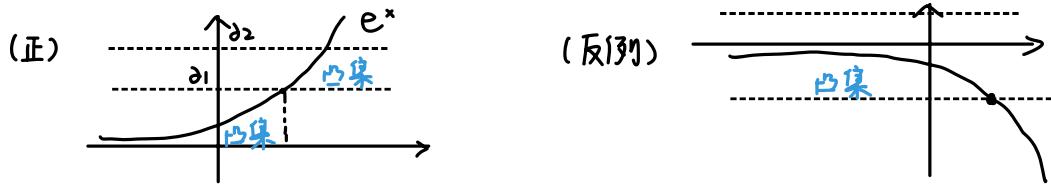
性质: 凸函数的 δ -sublevel set 都是凸集.

证明: $\forall x, y \in C_\delta, f(x) \leq \delta, f(y) \leq \delta, x \in \text{dom}f, y \in \text{dom}f$

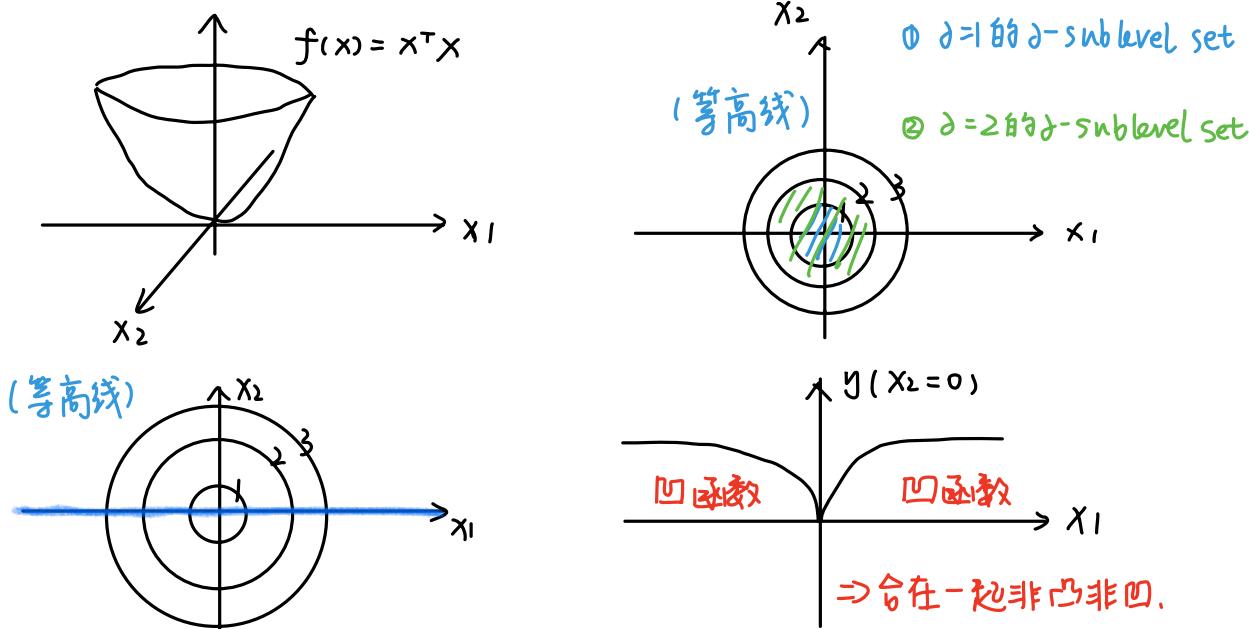
$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \leq \theta \delta + (1-\theta)\delta = \delta$$

$\therefore \theta x + (1-\theta)y \in C_\delta$, 故 C_δ 是凸集.

若函数的 δ -sublevel set都是凸集，则不一定 \Rightarrow 是凸函数.



二维: $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $x \in \mathbb{R}^2$. $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$



\Rightarrow unimodal function 单模态.

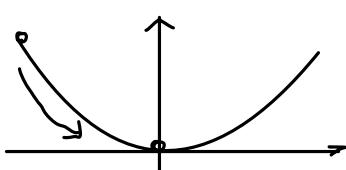
8. 拟凸函数 Quasi Convex Function (拟凸不一定凸, 但是很好优化)

(1) $f: \mathbb{R}^n \rightarrow \mathbb{R}$, Quasi Convex: $S_\delta = \{x \in \text{dom } f \mid f(x) \leq \delta\}$ 凸, $\forall \delta$

Quasi Concave: $S_\delta' = \{x \in \text{dom } f \mid f(x) \geq \delta\}$ 凸, $\forall \delta$

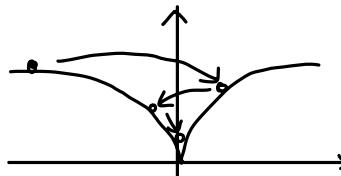
Quasi linear: $S_\delta'' = \{x \in \text{dom } f \mid f(x) = \delta\}$ 凸. $\forall \delta$

e.g.: $y = e^x$ 是拟凸, 拟凹, 拟线性

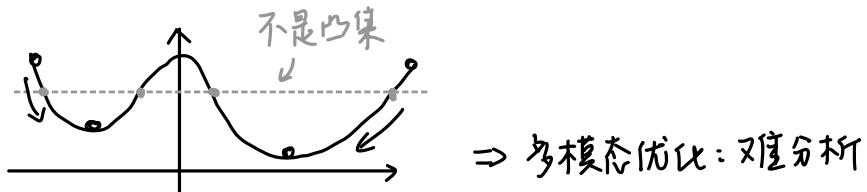


凸, 梯度法: 小球沿切线滚

凸函数 \Rightarrow 拟凸, 拟凸为凸



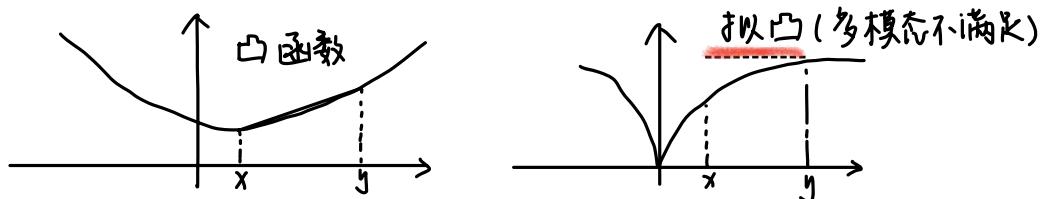
单模态优化: 设计停止准则.



(2) (定义推广) $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $\text{dom } f$ 为凸, $\forall x, y \in \text{dom } f$, $0 \leq \theta \leq 1$

凸函数: $\forall f(x) + (1-\theta)f(y) \geq f(\theta x + (1-\theta)y)$

拟凸: $\max\{f(x), f(y)\} \geq f(\theta x + (1-\theta)y)$



eg: 向量的长度 $x \in \mathbb{R}^n$. (x 中最后一个非零元素的位置)

$$f(x) = \begin{cases} \max\{i, x_i \neq 0\} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

$\{f(x) \leq \alpha\} \Rightarrow$ 对所有 $i = 1, 2, \dots, n$, $x_i = 0$ (n 维空间的子空间)

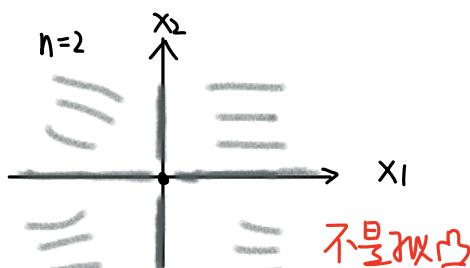
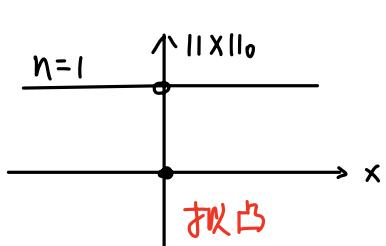
$\forall \alpha$, 都是子空间. 拟凸得证.

eg: 线性分式函数 $f(x) = \frac{c^T x + b}{c^T x + d}$, $\text{dom } f = \{x \mid c^T x > d\}$

$$S_\alpha = \left\{ x \mid c^T x + d > 0, \frac{c^T x + b}{c^T x + d} \leq \alpha \right\}$$

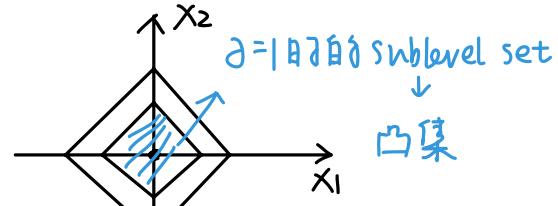
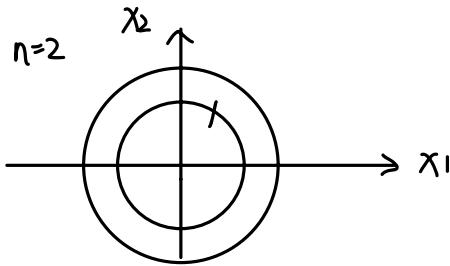
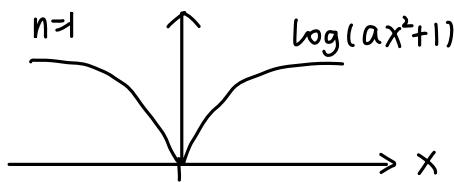
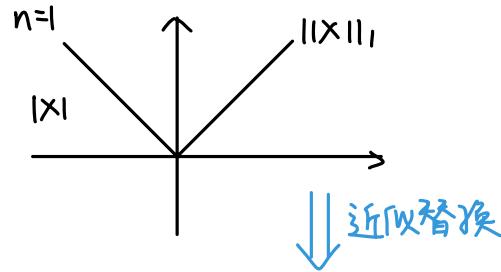
$= \left\{ x \mid c^T x + d > 0, \alpha c^T x + b \leq \alpha(c^T x + d) \right\}$ 多面体 \Rightarrow 凸集 \Rightarrow 拟凸函数

※: 向量的零范数 $x \in \mathbb{R}^n$, $f(x) = \|x\|_0$



$$\min \|x\|_0 \Leftrightarrow (\text{非零元素尽可能少}) \xrightarrow{\text{松弛}} \min \|x\|_1$$

s.t. $x \in C$ (找到最稀疏的 x)



$$\begin{aligned} & \min \log(x^T x + 1) \\ & \text{s.t. } x \in C \end{aligned} \quad \left. \right\} \text{拟凸}$$

拟凸函数

(3) 可微拟凸函数 - 阶条件

凸: $\text{dom } f$ 为凸, $f(y) \geq f(x) + \nabla f^T(x)(y-x)$, $\forall x, y \in \text{dom } f$

拟凸: $\text{dom } f$ 为凸, $f(y) \leq f(x) \Rightarrow \nabla f^T(x)(y-x) \leq 0$

证明: 考虑一维情况

$$\textcircled{1} \Rightarrow x, y \in \text{dom } f, 0 \leq \theta \leq 1, \max\{f(x), f(y)\} \geq f(\theta x + (1-\theta)y)$$

$$\text{设 } f(y) \leq f(x) . f(x) \geq f(\theta x + (1-\theta)y)$$

$$f(\theta x + (1-\theta)y) - f(x) \leq 0$$

$$\Leftrightarrow f(\theta x + (1-\theta)y) - f(\theta x + (1-\theta)x) \leq 0$$

$$\Leftrightarrow \frac{f(\theta x + (1-\theta)y) - f(\theta x + (1-\theta)x)}{(1-\theta)(y-x)} (1-\theta)(y-x) \leq 0$$

$$\stackrel{\theta \rightarrow 1}{\Leftrightarrow} f'(x)(y-x) \leq 0$$

不能用极限

$$\textcircled{2} \Leftarrow \forall x, y \in \text{dom } f, \text{均有 } f(y) \leq f(x) \quad \forall \theta \in (0, 1), z = \theta x + (1-\theta)y$$

$$(\text{证明 } f(x) \geq f(z)) \quad f(x) \leq f(z), f(y) \leq f(z) \Rightarrow f'(z)(x-z) \leq 0$$

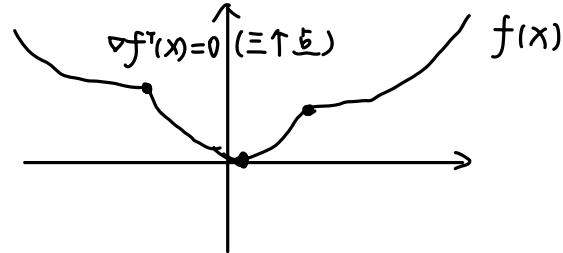
$$\begin{cases} f'(z)\theta(y-x) \leq 0 \\ f'(z)(1-\theta)(x-y) \leq 0 \end{cases} \Rightarrow f'(z)$$

if $f(x) \leq f(z)$, then $f(x) = f(z)$

又 $f(x) > f(z)$, 所以 $f(x) \geq f(z)$

凸: 若 $\nabla^T f'(x) = 0$, 则 $\forall y, f(y) \geq f(x)$

拟凸: 若 $\nabla^T f'(x) = 0$, 则 $\forall y, f(y) \leq f(x) \Rightarrow 0 \leq 0$



(4) 二阶条件

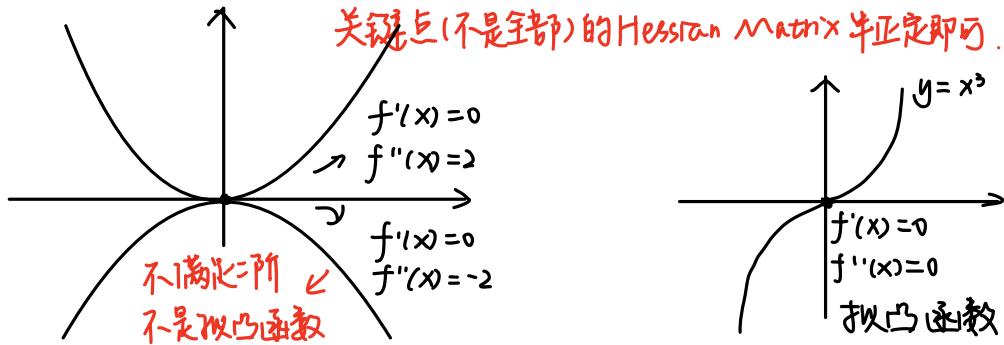
凸: $\text{dom } f$ 为凸, 且 $\nabla^T f'(x) \geq 0, \forall x \in \text{dom } f$

拟凸: $\text{dom } f$ 为凸, 且 $y^T \nabla^T f'(x) \geq 0 \Rightarrow y^T \nabla^T f'(x) y \geq 0$

$n=1$ 时, $y^T f'(x) \geq 0 \Rightarrow y^T f''(x) y \geq 0$

$$\begin{cases} y=0, 0 \geq 0 \\ f'(x)=0, y \neq 0, f''(x) \geq 0 \end{cases}$$

关键点(不是全部)的Hessian Matrix半正定即可.



9. logconvex function & logconcave function

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为 logconcave, 若 $f(x) > 0, \forall x \in \text{dom } f$, 且 $\log f$ 为凹函数.

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为 logconvex, 若 $f(x) > 0, \forall x \in \text{dom } f$, 且 $\log f$ 为凸函数.

若 f 为 logconvex, 则 f 为 convex. ($f = e^{\log f}$) \Rightarrow 函数组合的四条规则

若 f 为 concave, $f > 0$, 则 f 为 log concave

作业： Chapter2 , 1、2、5、7、10、16、18、19

Chapter3 1、2、5、13、18、21、32、33、36、43