1. 定义①：一个函数 $f: R^n \to R$ 为凸函数 $\Leftrightarrow$

○ dom f 为凸集.

○ 且对所有的 $x, y \in \mathrm{dom} f$, $0 \le \theta \le 1$, 有 $f(\theta x+(1-\theta)y) \le \theta f(x)+(1-\theta)f(y)$

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)

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

2. 凸函数的扩展

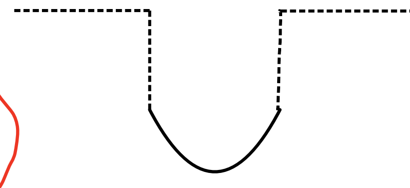
$f: R^n \to R$ 为凸, $\mathrm{dom} f = C \subseteq R^n$

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

仍然是凸函数

$\tilde{f}: R^n \to R$

$\mathrm{dom} \tilde{f} = R^n$



eg: 凸集的示性函数是凸函数.

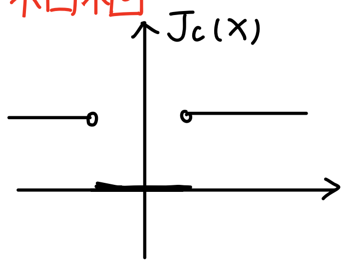
不要扩展到比较小的数

要扩展到 $+\infty$

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

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

不凸不凹



3. 一阶条件. ☆☆☆ 凸函数最重要的性质. (定义③)

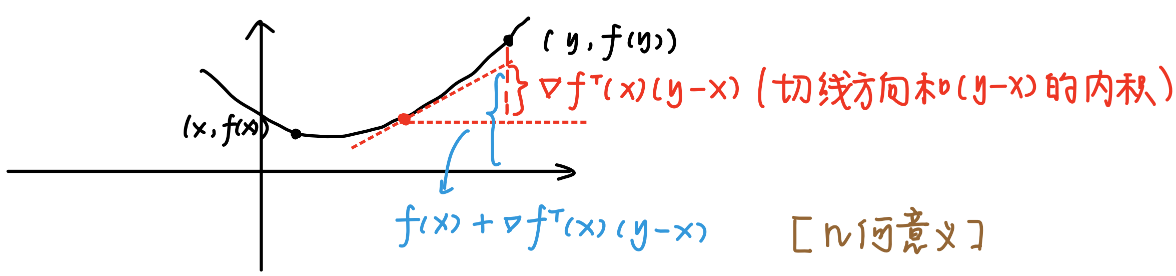
闭集. 不能是开集

设 $f: R^n \to R$ 可微, 即梯度 ∇f 在 dom f 上均存在. 则 f 为凸等价于

① dom f 为凸.

一阶偏导(切线方向)

② $f(y) \ge f(x) + \nabla f^T(x)(y-x) \quad \forall x, y \in \mathrm{dom} f$



[几何意义]

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

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，$f: R^n \to R$ 为凸 $\Leftrightarrow$ 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.

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

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

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

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

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

取 $\lim\limits_{t \to 0^+}$，$f(y) \geq f(x) + f'(x)(y-x)$

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 $z = \theta x + (1-\theta) y \in \text{dom} f$

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

$\theta x + (1-\theta) y - z = 0$

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

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：$x \to y$

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

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

取 $t_1 = 0,\ t_2 = 0$，$\underset{f(y)}{g(1)} \geq \underset{f(x)}{g(0)} + g'(0)$

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

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

$$f(ty+(1-t)x)\geq f(\tilde t y+(1-\tilde t)x)$$
$$+\nabla f(\tilde t y+(1-\tilde t)x)\underbrace{(ty+(1-t)x-\tilde t y-(1-\tilde t)x)}_{(y-x)(t-\tilde t)\ /tuta}$$

构造低维函数

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

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

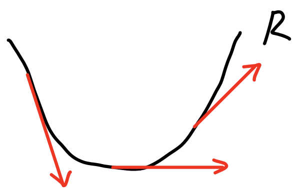
$g(t)\geq g(\tilde t)+g'(t)(t-\tilde t)$, 则 $f(x)$ 为凸函数.

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 $f:R^n\to R$ 二阶可微, 则 f 为凸 ⇔ $\mathrm{dom}f$ 为凸, $\nabla f^2(x)\geq 0$, $\forall x\in \mathrm{dom}f$

(铁锅的形状)

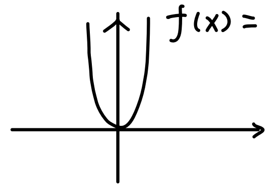


几何意义: 一阶偏导单调不减.

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

$\not\Leftarrow$ (不一定成立, 但是在定义①②③成立)

eg: $f(x)=x^4$.



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

存疑: 从图象判断凸or严格凸?

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

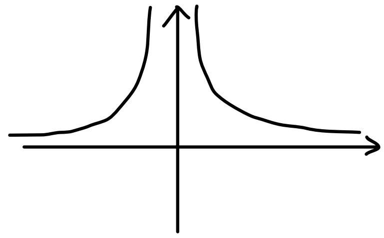
eg: 二次函数 $f:R^n\to R$. $\mathrm{dom}f=R^n$.

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

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

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

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



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 $f(x) = Ax+b$ ，$\nabla f^2(x) = 0$ 既凸又凹.

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

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

③ 幂函数 $f(x) = x^a$，$x \in R_{++}$

$f'(x) = ax^{a-1}$，$f''(x) = a(a-1)x^{a-2}$

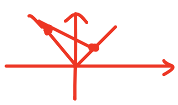
$$\nabla^2 f(x) = \begin{cases} \geq 0, & a \geq 1 \text{ 或 } a \leq 0 \quad \text{凸函数} \\ \leq 0, & 0 \leq a \leq 1 \quad \text{凹函数} \end{cases}$$

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

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

$p \geq \cancel{1}2$ 时，则为凸函数.（$p \geq 2$ 时，2阶导才有意义）

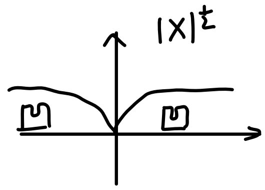
$p=1$ 时，$f(x) = |x|$



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.

② $p=\frac{1}{2}$ 时，



既非凸又非凹.

因此在 $p<1$ 的情况下没有统一的结论.

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

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

⑥ 负熵 $f(x)=x\log x,\ x\in R_{++}$

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

⑦ 范数 R^n空间的范数 $p(x),\ x\in R^n$.

Ⅰ. $p(ax)=|a|\,p(x)$

Ⅱ. $p(x+y)\leq p(x)+p(y)$

Ⅲ. $p(x)=0 \Leftrightarrow x=0$

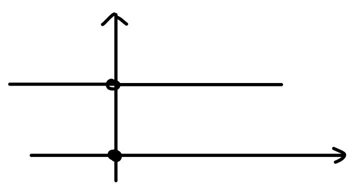
证明: $\forall x,y\in R^n,\ \forall 0\leq\theta\leq 1$,

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

∴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)

eg: $x\in R$



不满足范数条件Ⅰ.

$\|1\|_0=1,\ \|2\times 1\|_0=1$

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

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

$$f(\theta x+(1-\theta)y)=\max\{\theta x_i+(1-\theta)y_i,\ i=1,\cdots,n\}$$
$$\leq\theta\max\{x_i,\ i=1,\cdots,n\}+(1-\theta)\max\{y_i,\ i=1,\cdots,n\}$$
$$=\theta f(x)+(1-\theta)f(y)\ \Rightarrow \text{凸函数}.$$

引申: 极小-极大问题 $\min\limits_x \max\limits_y f(x,y)$ 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

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

近似的最大误差

$$\max\{x_1,\cdots,x_n\}\leq f(x)\leq\max\{x_1,\cdots,x_n\}+\log n$$

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

定义 $z = [e^{x_1}, \cdots, e^{x_n}]^T$

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

$i = j$时，$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{-e^{x_i}e^{x_j} + e^{x_i}(e^{x_1}+\cdots+e^{x_n})}{(e^{x_1}+\cdots+e^{x_n})^2}$ $1^T z$

(erg都是对角线元素)

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

定义 $z = [e^{x_1}, \cdots, e^{x_n}]^T$，$H = \frac{1}{(1^T z)^2}(1^T z \; \mathrm{diag}\{z\} - zz^T)$

定义 $K = 1^T z \; \mathrm{diag}\{z\} - zz^T$，验证 if $K \in R^{n\times n}$

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

$$v^T K v = (1^T z) v^T \mathrm{diag}\{z\} v - v^T z z^T v$$

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

Cachy-Schwart 不等式.

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

(故遇到范数想三角不等式，遇到平方想 Cachy-Schwart 不等式)

⑩ n何平均函数 $f(x) = (x_1 + \cdots + x_n)^{\frac{1}{n}}$，$x \in R^n_{++}$

(需证明 $f(x)$ 为凹函数，or $f(x)$ 为凸函数)

⑪ 行列式的对数 $f(x) = \log \underset{\sim}{\overset{\text{行列式.}}{\det(x)}}$，$\mathrm{dom} f = S^n_{++}$

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

$\downarrow$

行列式 = 特征值乘积，特征值 = 奇异值.

切曲面的方向v，
距离t，起点Z

证明：$n=1$时，凹函数

（定义二）$n>1$时，$\forall Z\in S_{++}^n$，$\forall t\in R$，$\forall v\in R^{n\times n}$

$Z+tv\in S_{++}^n=\mathrm{dom}f$，故$v\in S^{n\times n}$

$g(t)=f(Z+tv)=\log\det(Z+tv)$

$=\log\det\{Z^{\frac12}(I+tZ^{-\frac12}vZ^{-\frac12})Z^{\frac12}\}$

$=2\log\det\{Z^{\frac12}\}+\log\det\{I+tZ^{-\frac12}vZ^{-\frac12}\}$

$=\log\det\{Z\}+\log\det\{I+tZ^{-\frac12}vZ^{-\frac12}\}$

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

对称矩阵（转置等于自身）（特征值分解）

λ_i：矩阵$tZ^{-\frac12}vZ^{-\frac12}$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$QQ^T=I$（正交矩阵）

$\det(I+tZ^{-\frac12}vZ^{-\frac12})$

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

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

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

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

$\therefore g(x)$是凹函数

$\therefore f(x)$是凹函数.

5. 保凸运算

(1) 非负加权和

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

- $f:R^n\to R$，$A\in R^{m\times n}$，$b\in R^n$，$g(x)=f(Ax+b)$．$\mathrm{dom}\,g=\{x\mid Ax+b\in\mathrm{dom}f\}$

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

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

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

- $f_i: R^n \to R$, $i=1,\cdots,m$ 为凸，$A \in R^n$，$b \in R$，$g(x) = A^T[f_1(x) \cdots f_m(x)]^T + b$

实质是非负加权和. 只有 $A^T \geq 0$ 才能保持凸性.

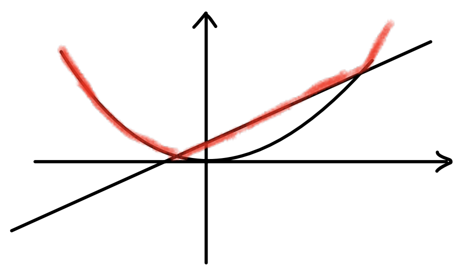
(3) 两个函数的极大值函数

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

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

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

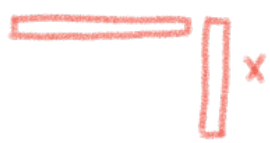
eg：$f(x) = \max\{x^2, x\}$



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

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

举个栗子：$f(x) = \max\{x_{i_1} + \cdots + x_{i_r} \mid 1 \leq i_1 \cdots \leq i_r \leq n\}$



C_n^r 个仿射映射

(极大值函数为凸)

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）

$f(x) = \lambda_{max}(X)$，$\mathrm{dom} f = S^m$. $xy = \lambda y$，$y^T x y = y^T \lambda y$

$$y^T X y = \lambda \|y\|_2^2 \iff \lambda = \frac{y^T X y}{\|y\|_2^2}$$

将y理解成线性项⇒凸函数

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

(4) 函数的组合

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

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

① 一维: $k=n=1$; $\mathrm{dom}\, g = \mathrm{dom}\, h = \mathrm{dom} f = R$; h、g均为二阶可微.

f 为凸 $\iff f''(x) \geq 0$

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

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 $n, k \geq 1$; $\mathrm{dom}\, g$, $\mathrm{dom}\, h$, $\mathrm{dom} f \neq R^n, R^k, R^h$; h、k均不二阶可微

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

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

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

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

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

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

$\|$

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

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

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

$\tilde{h}(\overset{+\infty}{\underline{g(\theta x+(1-\theta)y)}}) \le \tilde{h}(\overset{h(\cdots)}{\theta g(x)+(1-\theta)g(y)})$

(可以用h的单调性)

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

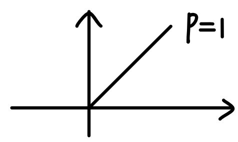
所以 $f(x)$ 一定是凸函数.

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

若g为凹, $g>0$, $\log\{g(x)\}$为凹.

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

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



✿乔布斯在斯坦福大学的演讲. Make it great.

(5)函数的透视. 十透视函数 $P: R^{n+1}\to R^n$, $\mathrm{dom}\,P \in R^n\times R_{++}$, $P(z,t)=\frac{z}{t}$.

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

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

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. $f(x)=x^T x$, $\mathrm{dom}\,f = R^n$.

$g(x,t) = t(\frac{x}{t})^T(\frac{x}{t}) = \frac{x^T x}{t}$ (因为t是标量)

eg：负的对数. $f(x) = -\log x$, $\mathrm{dom} f = R_{++}$

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

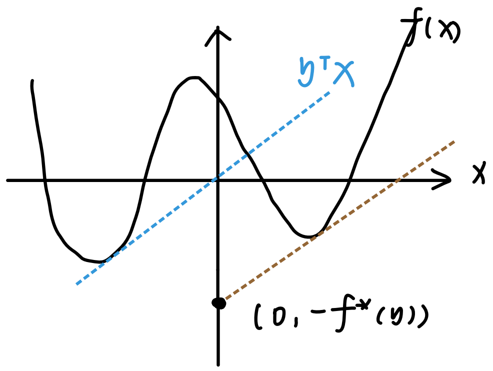
扩展：$u, v \in R^n_{++}$, $g(u,v) = \sum_{i=1}^{n} u_i \log\frac{u_i}{v_i}$ 凸.

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

6. 函数的共轭 Conjugate

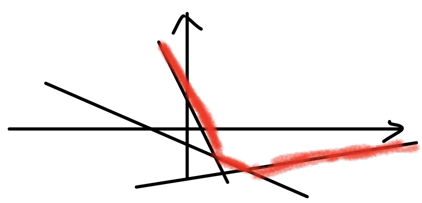
$f: R^n \to R$, $f^*: R^n \to R$, $f^*(y) = \sup_{x\in \mathrm{dom} f}(y^Tx - f(x))$ 选定y，遍历x求最大值.



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

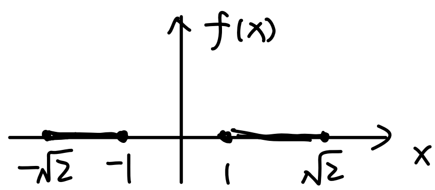
(2) 函数的共轭一定为凸函数. (极大值的分段线性函数)

(对偶原理)



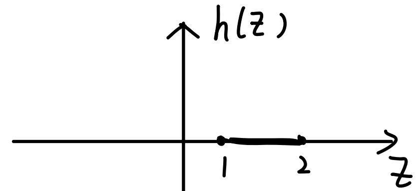
分割线

NOTES: (函数的组合-补充) $\tilde{h}$的单调性.



$g(x) = x^2$, dom g ∈ R, 凸

$h(z) = 0$, dom h = [1, 2]. 凸. 不增不降

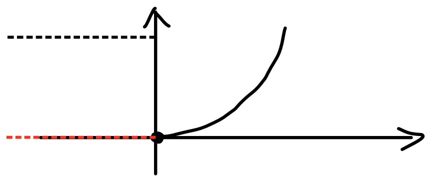


$$f = g(h(x)) = \begin{cases} 0 & x\in[-\sqrt{2},-1]\cup[1,\sqrt{2}] \\ \text{不存在} & \text{else} \end{cases}$$

非凸非凹(看定义域)

$\tilde{h}$在R^n都需有单调性.

INCORRECT (函数的组合-补充) eg: 若 g 为凸, $g\geq 0$, $p\geq 1$, 则 $g^p(x)$ 为凸. $h(z)=z^p$, $\mathrm{dom}\,h=R_+$



$$h(z)=\begin{cases} z^p, & z\in R_+ \\ 0, & z\in R_{--}\end{cases}$$

单调不降，使用第一条规则.

分割线

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

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

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

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

eg: $f(x)=-\log x$, $\mathrm{dom}f=R_{++}$, $f^*(y)=\sup\limits_{x>0}(yx+\log x)=\begin{cases}-1-\log(-y), & y<0\\ +\infty & y>0\end{cases}$

eg: $f(x)=\frac{1}{2}x^TQx$, $Q\in S^n_{++}$, $\mathrm{dom}f=R^n$.

正定, 奇异值=特征值, 可逆 $\Rightarrow x=Q^{-1}y$

$$f^*(y)=\sup_x(y^Tx-\frac{1}{2}x^TQx) \rightarrow \frac{\partial(y^Tx-\frac{1}{2}x^TQx)}{\partial x}=y-Qx$$ (书附录矩阵计算)

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

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 α-sublevel set

若 $f: R^n\rightarrow R$, 定义其 α-sublevel set 为 $C_\alpha=\{x\in \mathrm{dom}f \mid f(x)\leq\alpha\}$

性质: 凸函数的 α-sublevel set 都是凸集.

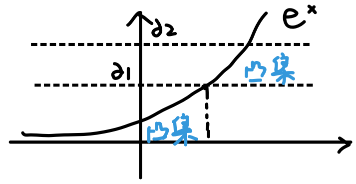
证明: $\forall x,y\in C_\alpha$, $f(x)\leq\alpha$, $f(y)\leq\alpha$, $x\in\mathrm{dom}f$, $y\in\mathrm{dom}f$

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

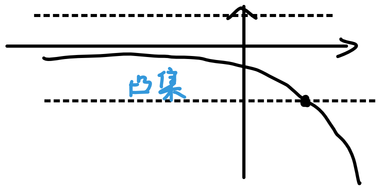
$\therefore \theta x+(1-\theta)y\in C_\alpha$, 故 C_α 是凸集.

若函数的α-sublevel set都是凸集，则f不一定是凸函数.

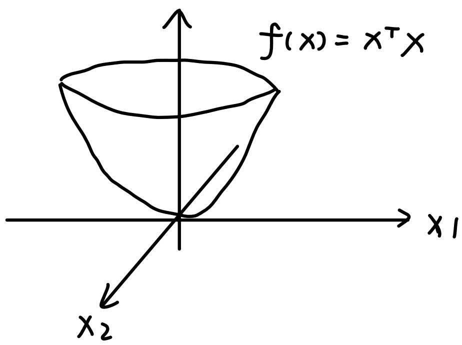
(正)



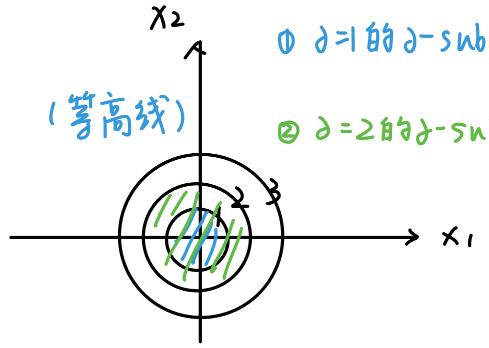
(反例)



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



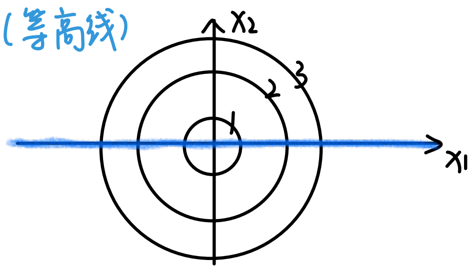
(等高线)

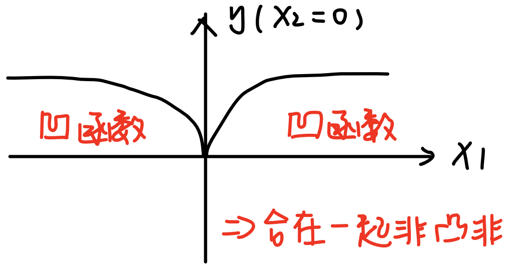


① $\alpha=1$的α-sublevel set

② $\alpha=2$的α-sublevel set

(等高线)





⇒合在一起非凸非凹.

= unimodel function 单模态.

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

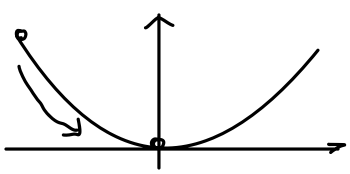
(1) $f: \mathbb{R}^n \to \mathbb{R}$，Quasi Convex：$S_\alpha = \{x \in \mathrm{dom} f \mid f(x) \le \alpha\}$ 凸，$\forall \alpha$

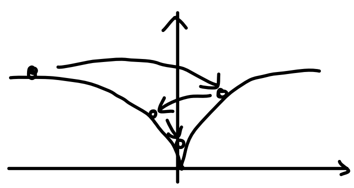
Quasi Concave：$S_\alpha' = \{x \in \mathrm{dom} f \mid f(x) \ge \alpha\}$ 凸，$\forall \alpha$

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

eg：$y=e^x$ 是拟凸，拟凹，拟线性

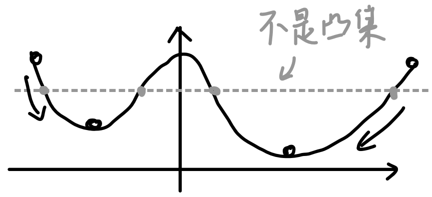
凸函数 ⇒ 拟凸． 拟凸 ⇏ 凸





凸．梯度法：小球沿切线滚

单模态优化：设计停止准则.

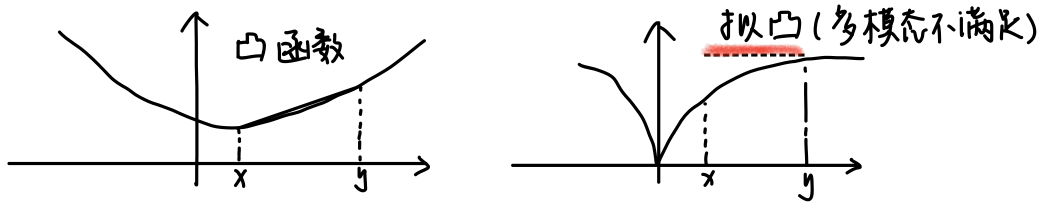


$\Rightarrow$ 多模态优化：难分析

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

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

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



eg：向量的长度 $x \in R^n$.（x 中最后一个非零元素的位置）

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

$\{f(x) \le \alpha\} \Rightarrow$ 对所有 $i = \lfloor \alpha \rfloor + 1 \cdots\cdots n$，$x_i = 0$（$n$ 维空间的子空间）

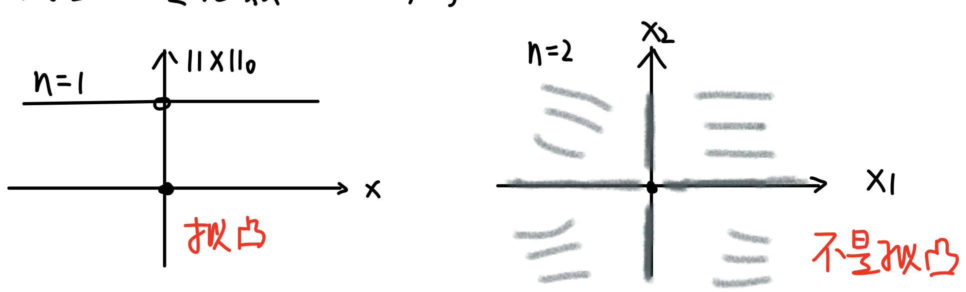
$\forall \alpha$，都是子空间．拟凸得证．

eg：线性分数函数 $f(x) = \dfrac{a^T x + b}{c^T x + d}$，$\text{dom} f = \{x \mid c^T x > d\}$

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

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

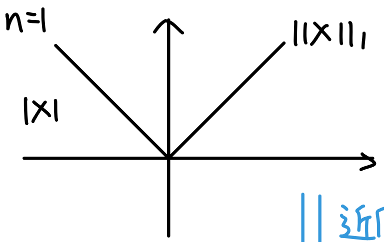
~~eg~~：向量的零范数 $x \in R^n$，$f(x) = \|x\|_0$

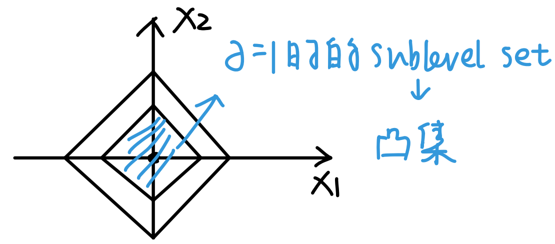


$$\min \|x\|_0 \quad \text{s.t. } x \in C \xrightarrow{\text{松弛}} \min \|x\|_1 \quad \text{s.t. } x \in C$$

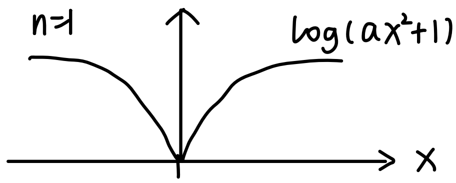
⇔非零元素尽可能少

(找到最稀疏的x)

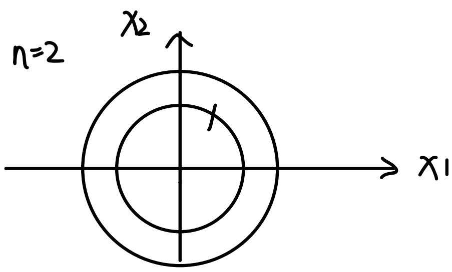




⇓近似替换



$$\min \log(x^Tx+1) \quad \text{s.t. } x \in C$$ 拟凸



拟凸函数

(3) 可微拟凸函数一阶条件

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

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

证明：考虑一维情况

① "⇒" $x, y \in \mathrm{dom}f$，$0 \leq \theta \leq 1$，$\max\{f(x), f(y)\} \geq f(\theta x + (1-\theta)y)$

设 $f(y) \leq f(x)$，$f(x) \geq f(\theta x + (1-\theta)y)$

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

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

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

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

② "⇐" $\forall x, y \in \mathrm{dom}f$，均有 $f(y) \leq f(x)$，$\forall \theta \in (0,1)$，$z = \theta x + (1-\theta)y$

不能用极限

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

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

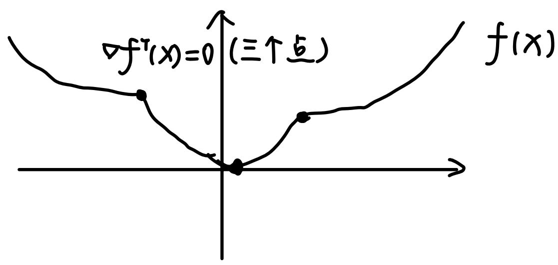
if $f(x) \le f(z)$, then $f(x) = f(z)$

又 $f(x) > f(z)$，所以 $f(x) \ge f(z)$

凸：若 $\nabla f^T(x) = 0$，则 $\forall y$, $f(y) \ge f(x)$

没有意义的点

拟凸：若 $(\nabla f^T(x) = 0)$，则 $\forall y$, $f(y) \le f(x) \Rightarrow 0 \le 0$



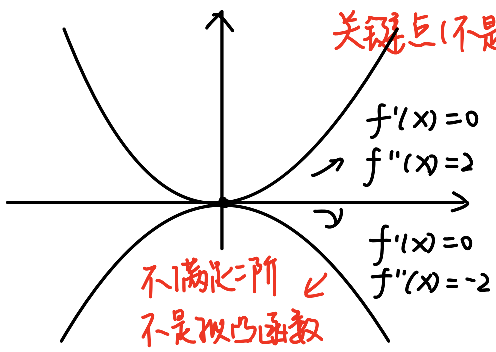
(4) 二阶条件

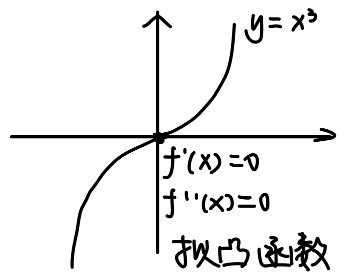
凸：dom$f$ 为凸，且 $\nabla^2 f(x) \ge 0$，$\forall x \in$ dom$f$

拟凸：dom$f$ 为凸，且 $y^T\nabla f(x) \ge 0 \Rightarrow y^T\nabla^2 f(x) y \ge 0$

$n=1$时，$yf'(x) \ge 0 \Rightarrow y^2 f''(x) \ge 0$ $\begin{cases} y=0, & 0 \ge 0 \\ f'(x)=0, & y \ne 0,\ f''(x) \ge 0 \end{cases}$

关键点(不是全部)的Hessian Matrix半正定即可.





9. log convex function 及 log concave function

$f: R^n \to R$ 为 log concave，若 $f(x)>0$，$\forall x \in$ dom$f$，且 $\log f$ 为凹函数.

$f: R^n \to R$ 为 log convex，若 $f(x)>0$，$\forall x \in$ dom$f$，且 $\log f$ 为凸函数.

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

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

$g(x) = \frac{M}{2}\|x\|^2 - f(x)$

$g(\theta x + (1-\theta)y) = \frac{M}{2}\|\theta x + (1-\theta)y\|^2 - f(\theta x + (1-\theta)y)$

$\theta g(x) + (1-\theta)g(y) = \theta(\frac{M}{2}\|x\|^2 - f(x)) + (1-\theta)(\frac{M}{2}\|y\|^2 - f(y))$

$\nabla g(x) = M\|x\| - f'(x)$

$\nabla^2 g(x) = MI - f''(x)$