Essence of the lecture (9(缺)/10/11)

凸函数的性质：

- 凸函数的第一定义
- 凸函数的第二定义
- 凸函数的扩展
- 凸函数的第三定义：一阶条件
- 凸函数的第四定义：二阶条件

凸函数的第一定义

凸函数：f 是凸函数，当且仅当其定义域为凸集，

$$\text{且 } \forall x, y \in S,\ \forall \theta \in [0,1],\ f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$$

严格凸函数：定义域为凸集，且 $\forall x, y \in S,\ \forall \theta \in [0,1],\ f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y)$

类似定义有凹函数和严格凹函数

凸函数的第二定义

$f:\ R^n \to R$ 为凸函数 $\Leftrightarrow$ $dom\ f$ 为凸，且 $\forall\ x \in dom\ f,\ \forall\ v,\ g(t) = f(x+tv)$ 为凸函数，$dom\ g = \{t \mid x + tv \in dom\ f\}$【g(t) 是关于 t 的函数，相当于从高维降到一维，可以通过证明该一维函数 g(t) 为凸来证明高维的 f(x) 为凸】

凸函数的扩展

$f:\ R^n \to R$ 为凸函数，$dom\ f:\ C \subseteq R^n$

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

，$\tilde{f}:\ R^n \to R$，$dom\ \tilde{f}:\ R$，即将 f 的定义域由 C 扩展到 R

示性函数是凸函数，凸集 $C \subseteq R^n$，同理将定义域由 C 扩展到 R^n

$$I_c(x) = \begin{cases} \infty & x \notin C \\ 0 & x \in C \end{cases}$$

，注意 $x \notin C$ 时，取值必须为 ∞，否则总能够找到特定的值使其非凸非凹

凸函数的一阶条件

设 $f:\ R^n \to R$ 可微（这表示 $dom\ f$ 一定是开集，因为对于扩展的 $\tilde{f}$，其边界不可微），即梯度 ∇f 在 $dom\ f$ 上均存在，则 f 等价于

①$dom\ f$ 为凸集 ②$f(y) \geq f(x) + \nabla f^T(x)(y-x),\ \forall x, y \in dom\ f$

一阶条件的证明

一维情况下

即证 $f:\ R \to R$为凸函数 $\iff dom\ f$为凸集，且$f(y) \geq f(x) + f'(x)(y-x)$

证明：（充分性条件）首先由 f 为凸函数，由定义得，其定义域为凸集（虽然显然，但必须指出）

因此有 $x + t(y-x),\ 0 < t \leq 1 \in dom\ f$（注意是对 0 是开区间）

$$
\begin{aligned}
&f(x + t(y-x)) \leq (1-t)f(x) + tf(y) \\
\Rightarrow &\lim_{t\to 0^+} f(y) \geq \lim_{t\to 0^+} f(x) + \frac{f(x+t(y-x)) - f(x)}{t} \\
\Rightarrow &f(y) \geq f(x) + f'(x)(y-x)
\end{aligned}
$$

（必要性条件）设 $\forall x \neq 0,\ x, y \in dom\ f,\ \theta \in [0,1]$, 构造$z = \theta x + (1-\theta)y \in dom\ f$

由 $f(x) \geq f(x) + f'(x)(x-z),\ f(y) \geq f(y) + f'(y)(y-z)$

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

高维情况下

证明：（充分性条件）设 f 为凸函数，则其定义域为凸集，考虑 $x, y \in dom\ f$

由定义二知，$g(t) = f(x + t(y-x))$ 为凸函数，有 $g'(x) = \nabla f^T(ty + (1-t)x)(y-x)$

利用上面证完的一维情况下的一阶条件，有 $g(t_1) \geq g(t_2) + g'(t_2)(t_1 - t_2)$

取 $t_1 = 1,\ t_2 = 0$，回代得到 $f(y) \geq f(x) + \nabla f^T(x)(y-x)$，得证

（必要性条件）$\forall x, y \in dom\ f$，取 $ty + (1-t)x \in dom\ f,\ \tilde{t}y + (1-\tilde{t})x \in dom\ f$

根据条件有 $f(ty + (1-t)x) \geq f(\tilde{t}y + (1-\tilde{t})x) + \nabla f^T(\tilde{t}y + (1-\tilde{t})x)(y-x)(t-\tilde{t})$

取 $g(t) = f(ty + (1-t)x),\ g(\tilde{t}) = f(\tilde{t}y + (1-\tilde{t})x)$，又 $g'(\tilde{t}) = f(\tilde{t}y + (1-\tilde{t})x)(y-x)$

原式可化为 $g(t) \geq g(\tilde{t}) + g'(\tilde{t})(t-\tilde{t})$，为凸函数，由定义二得f 也为凸函数

凸函数的二阶条件

若 $f:\ R^n \to R$ 二阶可微，则 f 为凸 $\Leftrightarrow dom\ f$ 为凸，且 $\nabla^2 f(x) \succeq 0,\ \forall\ x \in dom\ f$，

其中 $\nabla^2 f(x)$ 为二阶偏导矩阵（Hessian 矩阵）

$\nabla^2 f(x) \succ 0 \Rightarrow$ 严格凸函数，但是 $\nabla^2 f(x) \not\Leftarrow$ 严格凸函数（反例为 x^4，其二阶导在 0 处取 0）

Essence of the lecture (12/13/14)

常见的凸函数和凹函数：

1. 常见的凸函数
    - 仿射函数
    - 指数函数
    - 幂函数，绝对值幂函数
    - 负熵：$f(x) = xlogx$
    - 范数（范数的三个条件，零范数不是范数）
    - 极大值函数，其解析逼近（log-sum-up）
2. 常见的凹函数
    - 对数函数
    - 几何平均
    - 行列式的对数

仿射函数：$f(x) = Ax + b\ \nabla^2 f(x) = 0 \Rightarrow$ 既是半正定又是半负定，既凸又凹

指数函数：$f(x) = e^{ax},\ x \in R,\ \nabla^2 f(x) = a^2 e^{ax} \geq 0$，为凸

幂函数：$f(x) = x^a,\ x \in R_{++}$，防止出现负数开根号或除 0 的情况

$$\nabla^2 f(x) = a(a-1)x^{a-2} = \begin{cases} \geq 0 & a \geq 1,\ a \leq 0 \quad 凸 \\ \leq 0 & a \in [0,1] \quad\quad\ \ 凹 \end{cases}$$

Q：考虑 $f(x) = \frac{1}{x^2}$，该函数在 R 上是否是凸函数

A：非凸，因为在 0 处无定义，定义域非凸，即不能只通过指数判断凸性，首先要考察定义域

绝对值幂函数：$f(x) = |x|^p,\ x \in R$

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

先说结论，当 $p \geq 1$，函数为凸，当 $p < 1$，凸性则需具体讨论，如 $p = 0$ 时既凸又凹，而当 $p = \frac{1}{2}$ 时非凸非凹。

对此的证明，在 $p \geq 2$ 时可以通过二阶条件证明，在 $p = 1$ 可以通过第一定义证明，而 $p \in [1, 2]$ 需要通过另外的方法进行证明，因为此时二阶不可微，其一阶导函数不连续

负熵： $f(x) = xlogx,\ x \in R_{++},\ f''(x) = \frac{1}{x} > 0$

范数： R^n 空间里的范数 $P(x),\ x \in R^n$，其满足如下三个条件

① $P(ax) = |a|P(x)$ ②$P(x+y) \leq P(x) + P(y)$ ③$P(x) = 0 \Leftrightarrow x = 0$

证明：$\forall x, y \in R^n,\ \forall \theta \in [0, 1] \quad P(\theta x + (1-\theta)y) \leq P(\theta x) + P((1-\theta)y) = \theta P(x) + (1-\theta)P(y)$

需要注意的是，零范数并非范数，其不满足①，其非凸

极大值函数： $f(x) = \max\{x_1, \ldots, x_n\},\ x \in R^n$

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

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

由此可知，极小极大问题（$\min\limits_{x} \max\limits_{y} f(x, y)$）相当于最优化一个凸函数

由于极大值函数是离散不可导的，对不可导的函数作可导的近似称为**解析逼近**，对极大值函数进行的解析逼近为 log-usm-up

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

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

讨论该函数的 Hessian 矩阵（$H = [H_{ij}]$）

$$H_{ij} = \begin{cases} \dfrac{\partial^2 f}{\partial x_i \partial x_j} = \dfrac{-e^{x_i} e^{x_j}}{(e^{x_1} + \cdots + e^{x_n})^2} & i \neq j \\ \dfrac{\partial^2 f}{\partial x_i \partial x_i} = \dfrac{-e^{x_i} e^{x_i} + e^{x_i}(e^{x_1} + \cdots + e^{x_n})}{(e^{x_1} + \cdots + e^{x_n})^2} & i = j \end{cases}$$

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

令 $z = [e^{x_1}, \ldots, e^{x_n}]$，上式可以化为 $H = \dfrac{1}{(1 \cdot z)^2}\left((1^T \cdot z)\, diag\{z\} - z \cdot z^T\right)$，取其后半部分为 K

证明 log-sum-up 为凸利用二阶条件只需要证明 Hessian 矩阵的半正定性 ($\forall v \in R^n$, $v^T K v \geq 0$)

$$
\begin{aligned}
v^T K v &= (1^T \cdot z)\, v^T diag\{z\} v - v^T z \cdot z^T v (\text{看作}(z^T v)^T (z^T v)) \\
&= (\sum_i z_i)(\sum_i v_i^2\, z_i) - (\sum_i v_i\, z_i)^2 \\
&\text{令 } a_i = v_i \sqrt{z_i},\ b_i = \sqrt{z_i} \\
&= (b^T b)(a^T a) - (a^T b)^2 \geq 0 \quad [Cauchy - Schwarz\text{不等式，得证}]
\end{aligned}
$$

看到范数就要想到三角不等式，看到平方就要想到 Cauchy-Schwarz 不等式

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

证明：当 $n = 1$ 时，$f(x) = log\ x$ 为凹

当 $n > 1$ 时，$\forall z \in S_{++}^n$, $\forall t \in R$, $\forall v \in R^{n \times n}$, $z + tv \in S_{++}^n\ =\ dom\ f$

$$
\begin{aligned}
g(t) = f(z + tv) = log\ det(z + tv) & \\
&= log\ det\{z^{\frac{1}{2}}(I + tz^{-\frac{1}{2}} v z^{-\frac{1}{2}}) z^{\frac{1}{2}}\} \\
&= log\ det\{z\} + log\ det(I + tz^{-\frac{1}{2}} v z^{-\frac{1}{2}}) \\
&\text{对后半部分做对称分解，有 } log\ det(QQ^T + tQ\Lambda Q^T) = log\ det(I + t\Lambda) \\
&= log\ det\{z\} + \sum_{i=1}^{n} log(1 + t\lambda_i) \quad [\lambda_i\text{是}tz^{-\frac{1}{2}} v z^{-\frac{1}{2}}\text{的第 i 个特征根}]
\end{aligned}
$$

又 $g'(t) = \sum_i \frac{\lambda_i}{1+t\lambda_i}$, $g''(x) = \sum_i \frac{-\lambda_i}{(1+t\lambda_i)^2} \leq 0$，由第二定义可以推出 f 为凹

几何平均函数： $f(x) = (x_1 \ldots x_n)^{\frac{1}{n}}$ 是凹函数，待证

Essence of the lecture (14/15/16/17)

常见的保凸变换：

- 非负加权和
- 仿射映射
- 凸函数的极大值函数
- 函数的组合
- 函数的透视
- 函数的共轭

非负加权和：若 $f_1 \ldots f_m$ 为凸且 $w_i \geq 0$，则 $f = \sum_{i=1}^{m} w_i f_i$ 为凸

证明：①定义域为各凸函数的交集，为凸　②由于各组份满足满足定义一，其组合也满足

若 $f(x,y)$ 对 $\forall y \in A, f(x,y)$ 均为凸（注意在 $y \in A$ 下为凸 $\neq$ 在 (x,y) 下均为凸【jointly convex】）

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

Q：$f_i:\ R^n \to R,\ i = 1, 2, \ldots m$ 为凸，$A \in R^n,\ b \in R$，$g(x) = A^T \left[f_1(x)m \ldots f_m(x)\right]^T + b$ 是否为凸

A：其本质是加权和，但是不一定非负，所以不一定是凸函数

仿射映射：$f:\ R^n \to R$，$A \in R^{n \times n},\ b \in R^n$，$g(x) = f(Ax+b),\ dom\ g = \{x \mid Ax + b \in dom\ f\}$

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

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

凸函数的极大值函数：f_1, f_2 为凸函数，定义 $f(x) = \max\{f_1(x), f_2(x)\}$，$dom\ f = dom\ f_1 \cap dom\ f_2$

证明：令 $x, y \in dom\ f,\ 0 \leq \theta \leq 1$

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

推广到无穷个凸函数的极大值情况： $y \in A,\ f(x,y)$ 对 x 为凸,$g(x) = \sup\limits_{y \in A} f(x,y)$ 为凸

Q：向量中 r 个最大元素的和为凸函数

A：$x \in R^n$，对 x 进行排序，$x[i]$ 表示第 i 大的元素，有

$f(x) = \sum_{i=1}^{r} x[i] = \max\{x_{i1} + \cdots + x_{ir} \mid 1 \le i1 \cdots \le ir \le n\}$

由于 $x_{i1} + \cdots + x_{ir}$ 是 x 的线性组合，保凸，同时 max 保凸，因此整个函数保凸，结果为凸函数

Q：实对称矩阵的最大特征值 λ，即 $f(x) = \lambda_{max}(x),\ dom\ f = S^m$

A：$Xy = \lambda y \Rightarrow y^T X y = \lambda y^T y = \lambda \|y\|_2^2 \Rightarrow \lambda = \frac{y^T X y}{\|y\|_2^2}$

令 $\|y\|_2^2 = 1$，则 $\lambda = y^T X y$

$f(x) = \lambda_{max}(x) = \sup\{y^T X y \mid \|y\|_2^2 = 1\}$

注意到 $y^T X y$ 是关于 x 的线性组合（将 y^2 看作系数），同时 sup 保凸，所以该函数为凸函数

函数的组合： $h:\ R^h \to R,\ g:\ R^n \to R^k$，$f = h \circ g,\ R^n \to R$，f 为 h 和 g 的函数组合

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

考虑满足如下三个条件的简单情况：

1. 一维：$k = n = 1$
2. 实空间：$dom\ g = dom\ h = dom\ f = R$
3. 二阶可微：h,g 均二阶可微

$$f'(x) = h'(g(x))g'(x) \quad f''(x) = h''(g(x))g'^2(x) + h'(g(x))g''(x)$$

利用 $f''(x) \ge 0\ or\ \le 0$，由此可以得到四条推论

① h 为凸，不降，g 为凸，则 f 为凸

② h 为凸，不增，g 为凹，则 f 为凸

③ h 为凹，不降，g 为凹，则 f 为凹

④ h 为凹，不增，g 为凸，则 f 为凹

对简单的情况进行扩展

1. 高维情况下：$n, k \geq 1$
2. 非实空间：$dom\ g$，$dom\ h$，$dom\ f \neq R^n$，R^k，R^n
3. 非二阶可微：h，g 均二阶不可微

上述四条结论中的单调性条件对 h 的扩展成立即可

① h 为凸，$\tilde{h}$ 不降，g 为凸，则 f 为凸

② h 为凸，$\tilde{h}$ 不增，g 为凹，则 f 为凸

③ h 为凹，$\tilde{h}$ 不降，g 为凹，则 f 为凹

④ h 为凹，$\tilde{h}$ 不增，g 为凸，则 f 为凹

一般定义 $\tilde{h}$ 为保持 h 函数凸性的在全空间的扩展，当 h 为凸函数时，一般定义为

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

扩展条件下，对①，h 为凸，$\tilde{h}$ 不降，g 为凸，则 f 为凸的证明 (其余条件同理)

证明：$\forall\ x, y \in dom\ f$，$\theta \in [0,1]$，即 $x, y \in dom\ g$，$g(x), g(y) \in dom\ h$

目标是证明 $h(g(\theta x + (1-\theta)y)) \leq \theta h(g(x)) + (1-\theta)h(g(y))$

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

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

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

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

因此只需证：$h(g(\theta x + (1-\theta)y)) \leq h(\theta g(x) + (1-\theta)g(y))$

$\Rightarrow$ 求证 $g(\theta x + (1-\theta)y)$ 在 $dom\ h$ 中，而后利用 h 的单调性即可

对 $g(\theta x + (1-\theta)y) \in dom\ h$ 的证明

假设 $g(\theta x + (1-\theta)y) \notin dom\ h$，利用 $\tilde{h}$ 在全空间的定义

$\tilde{h}$ 不降，则 $\tilde{h}(g(\theta x + (1-\theta)y)) \leq \tilde{h}(\theta g(x) + (1-\theta)g(y))$

不等式前半部分趋于 ∞，后半部分为 h，无意义，所以 $g(\theta x + (1-\theta)y) \in dom\ h$

由 h 不降可知，原不等式成立，f 为凸函数

函数的透视 $f:\ R^n \to R,\ g:\ R^n \times R_{++} \to R$

$$g(x,t)=tf(\frac{x}{t}) \quad dom\ g=\{(x,t)\mid t\in R_{++},\frac{x}{t}\in dom\ f\}$$

函数的透视有一个很重要的性质，若 f 为凸函数，则 g 也为凸函数，且对 (x,t) 是联合凸的，若 f 为凹函数，则 g 对 (x,t) 联合凹

注意与透视函数不同，透视函数 $P(z,t)=\frac{z}{t},\ P:R^{n+1}\to R^n$

例：欧几里得范数的平方

$f(x)=x^Tx,\ dom\ f=R^n,\ g(x,t)=t\frac{x}{t}^T\frac{x}{t}=\frac{1}{t}x^Tx$

例：负对数

$f(x)=-logx,\ dom\ f=R_{++}\ g(x,t)=t(-log\frac{x}{t})=tlog\frac{t}{x}$

考虑到 $x\geq 0,\ dom\ g=R^2_{++}$

对负对数的扩展：

取 $u,v\in R^n_{++},\ g(u,v)=\sum_{i=1}^n u_ilog\frac{u_i}{v_i}$，考虑到每一个分项都是凸的，因此整个函数是凸的，注意与非负加权和区分，这个函数是 $\sum_{i=1}^n g_i(u_i,v_i)$，非负加权和是 $\sum_{i=1}^n g_i(u,v)$

进一步扩展：KL-Divergence

$D_{KL}(u,v)\triangleq\sum_{i=1}^n u_ilog\frac{u_i}{v_i}-u_i+v_i$，注意到第一项为凸，后两项为仿射项，因此整个函数是凸的

进一步扩展：Bregman-Divergence

对于 $f:\ R\to R$ 的凸函数，$D_B(u,v)\triangleq f(u)f(v)-\nabla f(v)(u-v)$

由于 Bregman 散度并不能保证是凸的，常采用 KL 散度

KL 散度是取 $f(u)=\sum_{i=1}^n u_ilogu_i-\sum_{i=1}^n u_i$ 的特殊情况

函数共轭 (conjugate)：

$f: R^n \to R,\ f^*: R^n \to R,\ f^*(y) = \sup\limits_{x \in dom\ f} y^T x - f(x)$

①若 f(x) 可微，则 $f^*(y)$ 对应的 x 必是 $f'(x) = y$ 上的点 ($[f^*(y)]'_x = y - f'(x) = 0$)

②不论 f(x) 是否为凸函数，$f^*(y)$ 一定为凸函数

相当于无数条线取 max，对 y 而言是共线性项，因此恒为凸

对于函数而言常常说共轭，对于问题而言常常说对偶。

例：求 $f(x) = ax + b,\ dom\ f = R$ 的共轭函数

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

由定义一知，该函数为凸函数

例：求 $f(x) = -logx,\ dom\ f = R_{++}$ 的共轭函数

利用 $[f^*(x)]'_x = y + \frac{1}{x} = 0$，将 $x = -\frac{1}{y}$ 回代有 $-1 - log(-y)$

$$f^*(y) = \sup_{x>0} (yx + logx) = \begin{cases} -1 - log(-y) & y < 0 \\ +\infty & y \geq 0 \end{cases}$$

同理，由定义一知该函数为凸函数

例：求 $f(x) = \frac{1}{2} X^T \theta X,\ Q \in S^n_{++},\ dom\ f = R^n$ 的共轭函数

$f^*(y) = \sup\ (y^T x - \frac{1}{2} X^T Q X),\ [f^*(y)]'_x = y - QX = 0$

回代有 $y^T Q^{-1} y - \frac{1}{2} y^T Q^{-1} Q^T Q^{-1} y = \frac{1}{2} y^T Q^{-1} y$

对比原函数，x 变为 y，Q 变为 Q^{-1} 成了共轭，这是二次项共轭函数的性质

对于数而言，共轭的共轭为其自身，而这对函数并不成立，因为函数的共轭一定是凸函数，凹函数的共轭的共轭一定不为自身

只有在 f 为凸函数，并且为闭函数时，f 的共轭的共轭才为自身

Essence of the lecture (17/18/20)

拟凸函数：

- 凸集与凸函数的关系（α-sublevel set）
- 拟凸函数（Quasi Convex Function）
- 可微拟凸函数的一阶条件
- 可微拟凸函数的二阶条件
- 对数凸函数及对数凹函数

α-sublevel set：

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

性质：凸函数的所有 α-sublevel set 为凸集

$$\forall\ x, y \in C_\alpha,\ \theta \in [0,1],\ x, y \in dom\ f,\ f(x) \le \alpha,\ f(y) \le \alpha$$

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

注意：凸函数 $\underset{\nLeftarrow}{\Rightarrow}$ α-sublevel set 为凸集，例如 e^x 与 $-e^x$

拟凸函数（Quasi Convex Function）：

定义一：对于任意的 α，其 α-sublevel set 均为凸集的函数为拟凸函数

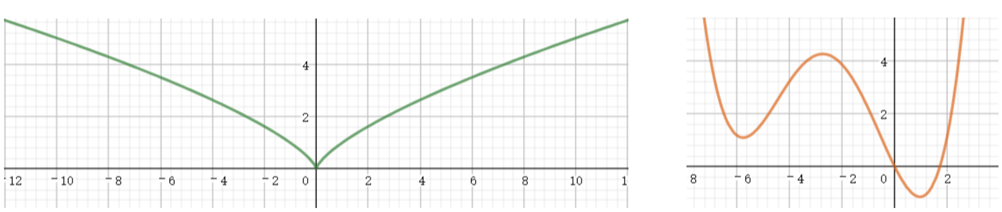
Quasi Convex $S_\alpha = \{x = dom\ f \mid f(x) \le \alpha\}$

Quasi Concave $S_\alpha^{'} = \{x = dom\ f \mid f(x) \ge \alpha\}$

Quasi Linear $S_\alpha^{"} = \{x = dom\ f \mid f(x) = \alpha\}$

注意：凸函数 $\underset{\nLeftarrow}{\Rightarrow}$ 拟凸函数

定义二：$\forall\ x, y \in dom\ f,\ \theta \in [0,1],\ \max\{f(x), f(y)\} \ge f(\theta x + (1-\theta)y)$



(a) 非凸的拟凸函数（单模态函数） (b) 多模态函数

图 1: 拟凸函数又称单模态函数

例：向量长度 $x \in R^n$，x 中最后一个非零元素的位置

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

$S_\alpha = \{f(x) \leq \alpha\} \Rightarrow\ \forall i = \lfloor i \rfloor + 1 \ldots n,\ x_i = 0$

相当于取子空间，一些轴上取零，其余轴上的值任意，为凸集，因此 f(x) 为拟凸函数

例：线性分数函数 $f(x) = \dfrac{a^T x + b}{c^T x + d}$, $dom\ f = \{x \mid c^T x + d > 0\}$

$S_\alpha = \{x \mid c^T x + d > 0,\ \dfrac{a^T x + b}{c^T x + d} < \alpha\} = \{x \mid c^T x + d > 0,\ ax + b \leq \alpha(c^T x + d)\}$

该集合表示一个多面体，为凸集，因此 f(x) 为拟凸函数

可微拟凸函数的一阶条件：$dom\ f$ 为凸，$\forall x, y \in dom\ f,\ f(y) \leq f(x) \Rightarrow \nabla^T f(x)(y - x) \leq 0$

20 中只证明了 $\theta \to 1$ 情况，具体的证明见 21（不好描述）

凸函数的一阶条件告诉我们局部最小即为全局最小：若 $\nabla f^T(x) = 0,\ \forall y,\ f(y) \geq f(x)$

拟凸函数的一阶条件并没有告诉我们什么，若 $\nabla f^T(x) = 0,\ \forall y, f(y) \leq f(x) \Rightarrow 0 \leq 0$ 无意义

这也就是凸函数和拟凸函数一阶条件的最大不同，拟凸函数不能保证一阶导数为 0 的点有意义

可微拟凸函数的二阶条件：$dom\ f$ 为凸，且 $y^T \nabla f(x) \geq 0 \Rightarrow y^T \nabla^2 f(x) y \geq 0$

考虑 $n = 1$ 的情况，$yf'(x) \geq 0 \Rightarrow y^2 f''(x) \geq 0$

$y = 0$ 时情况成立，当 $y \neq 0$，只需 $f'(x) = 0$，一定有 $yf'(x) = 0 \Rightarrow f''(x) \geq 0$

推广到高维情况，即对凸函数而言，所有点的二阶 Hessian 矩阵半正定，而对于拟凸函数而言，只需要部分关键点，$f'(x) = 0$ 这些点半正定

由此可以用该二阶条件判断函数是否是拟凸函数，考虑一阶为 0 的点二阶是否大于等于 0

log concave/log convex：

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

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

f 为凹函数$\underset{\not\Leftarrow}{\Rightarrow}$log f 为凹函数，f 为凸函数 $\underset{\Leftarrow}{\not\Rightarrow}$log f 为凸函数

可以将 f 看做 $e^{log\ f}$，利用函数组合的规则理解