

Part 1

证明: 当 $n > 1$ 时,

$$\forall X \in S_n^+, \forall t \in \mathbb{R}, \exists X+tV \in S_n^+,$$

$$f(t) = \log \det(X+tV) = \log \det(X^{\frac{1}{2}}(I+tX^{-\frac{1}{2}}VX^{-\frac{1}{2}})X^{\frac{1}{2}})$$

$$= \log \det X + \log \det(I+tX^{-\frac{1}{2}}VX^{-\frac{1}{2}})$$

记 λ_i 是 $X^{-\frac{1}{2}}VX^{-\frac{1}{2}}$ 的特征值, 存在正交阵 $Q \in \mathbb{R}^{n \times n}$

$$\text{对角阵 } \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n), X^{-\frac{1}{2}}VX^{-\frac{1}{2}} = Q\Lambda Q^T$$

$$= \log \det X + \log \det(I+t\Lambda Q^T)$$

\$\square\$

$$\text{几何平均 } f(x) = \left(\prod_{i=1}^n x_i \right)^{1/n}, x \in \mathbb{R}_+^n$$

行列表的对数

$$f(t) = \log \det(X), \quad \text{dom}(f) = S_n^+, \quad f: S_n^+ \rightarrow \mathbb{R}$$

$$\begin{aligned} &= \log \det X + \log \det(Q\Lambda Q^T + t\Lambda Q^T) \\ &= \log \det X + \log \det((I+t\Lambda)Q^T) \\ &= \log \det X + \log(\det(Q) \det(I+t\Lambda) \det(Q^T)) \\ &= \log \det X + \log \det(I+t\Lambda) \\ &= \log \det X + \log \det(L+t\Lambda) \\ &= \log \det X + \sum_{i=1}^n \log(1+t\lambda_i) \\ &= \log \det X + \sum_{i=1}^n \lambda_i (1+t\lambda_i) \end{aligned}$$

此处照片没拍到, 可以用基本定义证明, 将 x 用凸组合代入

答: 不能保证函数 g 为凸, 因为 A 中的元素不一定全部满足 1 中的 $w_i \geq 0$.

1. 非负加权和: f_1, \dots, f_m 为凸, 对 $w_i \geq 0$, $f = \sum_i w_i f_i$ 为凸

2. 若 $f(x,y)$ 对于任意 $y \in A$, $f(x,y)$ 为凸, 则 $w_i y_i$ 为凸, 有 $g(x) = \int_{y \in A} w_i y_i f(x,y) dy$ 也为凸

3. 仿射映射: $f: \mathbb{R}^m \rightarrow \mathbb{R}$, $A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$, 令 $x \in \mathbb{R}^n$ 时, $g(x) = f(Ax+b)$, 其中 $\text{dom}(g) = \{x | Ax+b \in \text{dom}(f)\}$ 且 f 为凸

(2). 若 $f_i: \mathbb{R}^n \rightarrow \mathbb{R}$ 凸, $A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$, $\theta: \mathbb{R}^m \rightarrow \mathbb{R}$, $0 \leq \theta_i \leq 1$, $\sum_i \theta_i = 1$ 时

$$g(x) = A^T \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix} + b \quad \text{不} \quad \text{若 } f_i, f_j \text{ 为凸时}$$

4. 两个函数的极大值函数: $f = \max\{f_1(x), f_2(x)\}$, 由 f_1, f_2 为凸时, f 也为凸, $\text{dom}(f) = \text{dom}(f_1) \cap \text{dom}(f_2)$

向量 $x \in \mathbb{R}^n$ 中两个最大元素之和

$$\begin{aligned} &\forall x, y \in \text{dom}(f), \forall \theta \in [0, 1] \\ &f(\theta x + (1-\theta)y) \\ &= \max\{f_1(\theta x + (1-\theta)y), f_2(\theta x + (1-\theta)y)\} \\ &\leq \max\{f_1(\theta x_1 + (1-\theta)y_1), f_2(\theta x_2 + (1-\theta)y_2)\} \\ &\leq \max\{f_1(x_1) + (1-\theta)f_1(y_1), f_2(x_2) + (1-\theta)f_2(y_2)\} \\ &\leq \theta \max\{f_1(x_1), f_2(x_2)\} + (1-\theta) \max\{f_1(y_1), f_2(y_2)\} \end{aligned}$$

$$\begin{aligned} x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} & \quad x_1 \geq x_2 \geq \dots \geq x_n, i \in \mathbb{N} \\ & \quad f(x) = \max\{x_1, \dots, x_n \mid 1 \leq i_1 < i_2 < \dots < i_n \leq n\} \end{aligned}$$

根据 4 得证函数 f 为凸函数

5. $f(x,y)$ 对于任意 $y \in A$, $f(x,y)$ 凸, $g(x) = \sup_{y \in A} f(x,y)$ 为凸

例: 実对称矩阵的最特征值

$$f(x) = \lambda_{\max}(x), \quad \text{dom}(f) = S_n^+$$

证: $y \in \mathbb{R}^n$, $xy = y^T x$, y 特征向量, λ 特征值

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

令 $\|y\|_2 = 1$, 则 $\lambda = y^T xy$

$$f(x) = \lambda_{\max}(x) = \sup\{y^T xy \mid \|y\|_2 = 1, y \text{ 为 } x \text{ 的特征向量}\}$$

Part 3

复合函数

给定函数 $h: \mathbb{R}^k \rightarrow \mathbb{R}$, $g: \mathbb{R}^k \rightarrow \mathbb{R}$. 令 复合函数 $f = hg: \mathbb{R}^k \rightarrow \mathbb{R}$

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

一维情况 $\left\{ \begin{array}{l} k=n=1 \\ \text{dom}(f) = \text{dom}(g) = \text{dom}(h) = \mathbb{R} \end{array} \right.$

f 为凸 $\Leftrightarrow f''(x) \geq 0 \quad f(x) = h(g(x))$

$$f'(x) = h'(g(x))g'(x)$$

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

- ① 若 h 为凸且单调不减, g 凸, 则 f 也凸
- ② 若 h 不增, g 凸, 则 f 为凸
- ③ 若 h 凸且单调不减, g 减, 则 f 为凸
- ④ 若 h 凸且单调不增, g 减, 则 f 为凸

① 若 h 为凸且单调不减, g 凸, 则 f 为凸

$h: \mathbb{R} \rightarrow \mathbb{R}$, $h(x) \geq h(y), x > y$

$g: \mathbb{R} \rightarrow \mathbb{R}$, $(-\infty, a], (-\infty, a]$

Part 1

例：若 f 为凸，则 e^f 也必凸

例：若 f 为凹， $\lambda > 0$ ，则 $\lambda f(x)$ 为凹

$\lambda = \lambda_1 \lambda_2$

例：若 f 为凸， $\lambda > 0$ ，则 $\frac{1}{\lambda} f(x)$ 为凸

例：若 f 为凸，则 $f(x) + c$ 为凸

例：若 f 为凹，则 $f(x) + c$ 为凹

高维 $k > 1$

$\dim(f) = \dim(\lambda f) = \dim(\lambda_1 f_1 + \dots + \lambda_k f_k)$

$\lambda_1, \dots, \lambda_k$ 不需要同时为正数

① 若 f 为凸， $\lambda_1, \dots, \lambda_k > 0$ ， $\lambda_1 + \dots + \lambda_k = 1$ ，则 $\lambda_1 f_1 + \dots + \lambda_k f_k$ 为凸

② 若 f 为凹， $\lambda_1, \dots, \lambda_k > 0$ ， $\lambda_1 + \dots + \lambda_k = 1$ ，则 $\lambda_1 f_1 + \dots + \lambda_k f_k$ 为凹

③ 一凹，一凸，一凹，一凸

④ 一凹，一凸，一凹，一凸

$f(x) = x^3$, $\dim(g) = 1$, $\lambda \in \mathbb{R}$, $\lambda \neq 0$, $\dim(\lambda g) = 1$

$f = h(g(x)) = \begin{cases} 0, & x \in \mathbb{R} \setminus [-1, 1] \\ \text{无定义}, & \text{else} \end{cases}$

不单通性

Part 2

函数的共轭
定义: $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $f^*: \mathbb{R}^m \rightarrow \mathbb{R}$

$$f^*(y) = \sup_{x \in \text{dom}(f)} (y^T x - f(x))$$

共轭的共轭 $\Rightarrow (f^*)^* = f$

①若 $f(x)$ 可微, 则 $f'(x_0) = 0$.

②函数的某点 x_0 为 f(x) 对应的 y 值是 $f(x_0) = y_0$.

例: $f(x) = ax + b$, $\text{dom}(f) = \mathbb{R}$

$$f^*(y) = \sup_{x \in \text{dom}(f)} (y-x - (\alpha x + b)) = \sup_{\substack{x \in \text{dom}(f) \\ \lambda \geq 0}} (y-\alpha x - b) = \begin{cases} -b & y < b \\ +\infty & y \geq b \end{cases}$$

$$f(x) = -\log x, \text{dom}(f) = \mathbb{R}_+$$

$$f^*(y) = \sup_{x \in \text{dom}(f)} (yx + \log x) \quad y > -\frac{1}{e} \quad y \cdot \frac{1}{x} = 0 \\ = \begin{cases} -1 - \log(-y), & y < 0 \\ +\infty, & y \geq 0 \end{cases}$$

$$\text{例: } f(x) = \frac{1}{2}x^T \beta_2, \quad \theta \in \mathbb{C}^n.$$

$$\begin{aligned}
 f(y) &= y^T Q y - \frac{1}{2} y^T Q^T Q y, \quad Q \in S_{++}, \det(Q) = 1, R^T = R \\
 f'(y) &= \sup_x (y^T x - \frac{1}{2} x^T Q x) \quad \text{subject to } y^T x = 1 \\
 &= y^T Q^{-1} y - \frac{1}{2} y^T (Q^{-1})^T Q y \quad \text{subject to } x = R^T y \\
 &= y^T Q^{-1} y - \frac{1}{2} y^T Q y = \frac{1}{2} y^T (Q - Q^{-1}) y
 \end{aligned}$$

Part 3

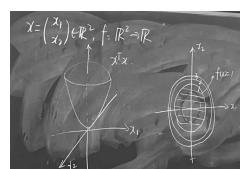
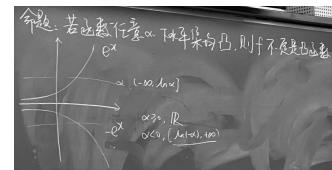
α -下半集 (α -sublevel set) 对 $f: P^n \rightarrow P$ 其 α -下半集 $C = \{x \in \text{dom}(f) \mid f(x) \leq \alpha\}$

对于 $T: \mathbb{R} \rightarrow \mathbb{R}$, 其下木至半某 ($\alpha = \{x \in \text{dom}(T) | T$
全数。且函数的任一 $x \in \text{dom}(T)$ 为 $T(x)$ 的像

证：对于 $\forall x, y \in C_x$, 则 $f(x) \leq \alpha, f(y) \leq \alpha, x, y \in \text{dom}(f)$

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

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



九

定义域和所有上水平集 $S'_d = \{x \in \text{dom}(f) | f(x) = d\}$

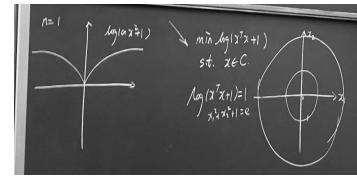
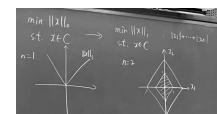
$$\text{水平集 } S_\lambda'' = \left\{ x \in \text{dom}(f) \mid f(x) \leq \lambda \right\} \text{ 为 } P_\lambda$$

Part 1

凸与拟凸一些例子

$$\text{例 1: } f(x) = \frac{ax+b}{cx+d}, \text{dom}(f) = \{x \mid cx+d > 0\}, a, x \in \mathbb{R}^n, b, d \in \mathbb{R}$$

$$S_{\infty} = \{x \mid c^T x + d > 0, \frac{ax+b}{cx+d} \leq \infty\} = \{x \mid cx+d > 0, ax+b \leq cx+d\}$$



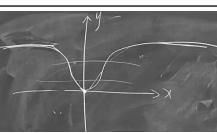
凸与拟凸一些定义

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

拟凸: 若 $f(x) = 0$ 且 $\forall y, f(y) \geq f(x)$

半凸: 若 $f(x) = 0$ 且 $\forall y, f(y) \geq f(x)$

半拟凸: 若 $\nabla f(x) = 0$, 则若真有 $f(y) \leq f(x)$, 则 $\nabla^2 f(x) \geq 0$



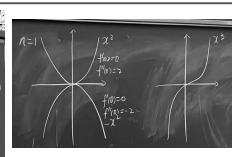
若 f 凸: $\text{dom}(f) \subset \mathbb{R}^n, \nabla^2 f(x) \geq 0, \forall x \in \text{dom}(f)$

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

$n=1$: 若 $f'(x) = 0$ 则 $f''(x) \geq 0$

$y=0, 0 \geq 0$

$y \neq 0, f'(x) = 0, \nabla^2 f(x) \geq 0$



Part 3

对数凹、对数凸

对数凹、对数凸
 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 若 $f(x) > 0, \forall x \in \text{dom}(f)$ 且 f 凸
 对数凸: $\log f$ 凸

1. $f \circ 0$ 凸, 它是 对数凸? 不是.
2. 若 f 对数凸, 则 f 凸. $f = e^{h(x)}$ 且 h 凸
3. 若 f 对数凹, f 凹? 不是.
4. 若 f 凹, $f > 0$. 则 f 对数凹. $\log f$ 凸.

Part 4

凸优化问题

凸优化问题
 $\min f(x)$
 $s.t. f_i(x) \leq 0, i=1, \dots, m$
 $h_i(x) = 0, i=1, \dots, p$

$x \in \mathbb{R}^n$, 优化变量 optimization variable
 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 目标函数 (objective function)/损失函数 (loss function)
 $\max u(x)$ 效用函数 (utility function)
 $f_i(x) \leq 0$ 不等式约束 (inequality constraint)
 $h_i(x) = 0$
 若 $m=p=0$, 则为未约束问题

优化问题的域 domain $D = (\bigcap_{i=1}^m \text{dom}(f_i)) \cap (\bigcap_{i=1}^p \text{dom}(h_i))$
 可行解集, feasible set $X_f = \{x \mid x \text{ 为可行解}\}$
 $f_i(x) \leq 0, i=1, \dots, m$
 $h_i(x) = 0, i=1, \dots, p$
 问题的最优值 optimal value $p^* = \inf_{x \in X_f} f(x)$, 若 $x^* \in D$, 则 $p^* = f(x^*)$
 最优解 optimal point/solution, 若 x^* 可行且 $f(x^*) = p^*$
 最优解集 optimal set $X_{opt} = \{x \mid x \in X_f, f(x) = p^*\}$

次优解集 $X_{\text{sub}} = \{x \mid x \in X_f, f(x) < p^*\}$
 $X_{\text{loc}} = \{x \mid x \in X_f, f(x) \leq p^*\}$
 局部最优解 locally optimal
 衍行 x 为局部最优解, 若 $\exists r > 0$, 使得
 $f_0(x) = \inf_{z \in B_r(x)} f(z) \quad \forall i, f_i(z) \leq 0, i=1, \dots, m$
 $h_i(z) = 0, i=1, \dots, p$
 $\|z - x\| \leq r$

Part 1

$$\begin{aligned} & \min f_1(x) \\ \text{s.t. } & f_i(x) \leq 0, i=1, \dots, m \\ & h_i(x) = 0, i=1, \dots, p \end{aligned}$$

定义: 若 $x \in X$, $f_i(x) \leq 0$, 则 $f_i(x) \leq 0$ 在处活动(起作用)(active)

$f_i(x) > 0$ 不活动(inactive)

$$\begin{aligned} & \min -\text{效用} \\ \text{s.t. } & \text{效用} \leq 100 \rightarrow \min -\text{效用} \leq 100 \\ & \min -3 \leq \text{效用} \\ \text{s.t. } & \text{效用} \leq 100 \\ & -1 \log(100 - \text{效用}) \leq 0 \end{aligned}$$

可行性优化问题(feasibility problems)

find x

s.t. $f_i(x) \leq 0, i=1, \dots, m$ $\rightarrow \min 0$

$h_i(x) = 0, i=1, \dots, p$

(1) Box constraint

$$\begin{aligned} & \min f_0(x) \\ \text{s.t. } & -1 \leq x_1 \leq 2, -1 \leq x_2 \leq 1 \\ & x_1 + x_2 \leq 0 \\ & x_1 - x_2 \leq 0 \end{aligned}$$

(2) $m \times n$ 线性等式约束 $Ax = b, A \in \mathbb{R}^{m \times n}$

$$\begin{aligned} & \min f_0(x) \\ \text{s.t. } & f_1(x) \leq 0, i=1, \dots, m \\ & f_{n+1}(x) = 0, i=1, \dots, p \\ & 0 \geq x_1 \geq 2, x_2 \geq 0, x_3 \geq 0 \end{aligned}$$

(3) 满足线性等式约束 $Ax = b, A \in \mathbb{R}^{m \times n}$

$$\begin{aligned} & \min f_0(x) \\ \text{s.t. } & Ax = b \\ & x \in \mathbb{R}^n \end{aligned}$$

(4) $x \neq 0$ 为 $F + x_0$, F : 非零向量 $\in \mathbb{R}^n$

$$\begin{aligned} & x \in \mathbb{R}^n \\ & x_0 \in \mathbb{R}^n \\ & z = x + x_0 \end{aligned}$$

(5) x 为非单调变量

$$\begin{aligned} & \min f_0(x) \\ \text{s.t. } & f_1(x) \leq 0, i=1, \dots, m \\ & f_{n+1}(x) = 0, i=1, \dots, p \\ & \frac{d}{dx} f_n(x) \geq 0, i=1, \dots, m \\ & \frac{d}{dx} f_{n+1}(x) \geq 0, i=1, \dots, p \end{aligned}$$

(6) x 为 \mathbb{R} , 单调变量

$$\begin{aligned} & \min f_0(x) \\ \text{s.t. } & f_1(x) \leq 0, i=1, \dots, m \\ & f_{n+1}(x) = 0, i=1, \dots, p \end{aligned}$$

若 $f'(x) \neq 0$, 则 $x \in \mathbb{R}$, 且 $x \in [f_1(x), f_{n+1}(x)]$

Part 2

$$\begin{aligned} & \min f_0(x) \\ \text{s.t. } & f_i(x) \leq 0, i=1, \dots, m \\ & h_i(x) = 0, i=1, \dots, p \end{aligned}$$

狭义凸优化: $f''(x) \geq 0$

拟凸优化: $f''(x) \leq 0$

例、 $\min f_0(x) = x_1^2 + x_2^2$

$$\begin{aligned} \text{s.t. } & f_1(x) = \frac{x_1}{x_2} \leq 0 \\ & h_1(x) = (x_1 + x_2)^2 = 0 \end{aligned}$$

若 $f_1(x) \geq 0$, 则 $x_1 = 0$

(1) $\min f_0(x)$

$$\begin{aligned} \text{s.t. } & f_1(x) \leq 0, i=1, \dots, m \\ & f_0(x) = x_1^2 + x_2^2 \geq 0 \\ & x_1 = 0, i=1, \dots, m \end{aligned}$$

当优化目标为凸函数

重要性质: 局部最优 = 全局最优

\Leftrightarrow 若 f 为凸函数, $\exists \theta > 0$, 使得 $|f(\theta x) - f(x)| \leq \theta |x|$

假设 f 不是全局最优, $\exists y \neq x$, 且 $f(y) < f(x)$

$\Rightarrow \|y - x\|_2 > r$, 记 $Z = (1-\theta)x + \theta y$, $\theta = \frac{\|y - x\|_2}{r}$

计算 $\|Z - x\|_2 = \|\theta y - \theta x\|_2 = \theta \|y - x\|_2 = \theta \|y - x\|_2 = \frac{r}{2} < r$, 得 $f(Z) < f(x)$

图 $\|y - x\|_2 > r$, $\forall \theta \in (0, 1)$, f 凸, 则

$$f_0(y) \leq (1-\theta)f_0(x) + \theta f_0(y) \leq (1-\theta)f_0(x) + \theta f_0(x) = f_0(x)$$

(其中第二步等式使用了 $f_0(y) \leq f_0(x)$)

$f_0(y) \geq f_0(x) + \nabla f_0^T(x)(y-x)$

目标函数引数到微时的单侧性准则

设凸问题可行解集 $X_F = \{x \mid f_i(x) \leq 0, i=1, \dots, m, h_i(x) = 0, i=1, \dots, p\}$

则 $x^* \in X_F^{\text{最外}} \Leftrightarrow \nabla f_0^T(x^*)(y-x^*) \geq 0, \forall y \in X_F$

例: 约束仅为等式约束 $\min f_0(x)$, $\text{dom}(f_0) = \mathbb{R}^n$

若存在 $x, Ax = b$ 且 $x^* \Leftrightarrow \nabla f_0^T(x)(y-x) \geq 0$

$\begin{cases} Ax = b \\ Ay = b \end{cases} \Rightarrow y = x + v, v \in N(A) = \{z \mid Az = 0\}$

于是 $\nabla f_0^T(x)v \geq 0, \forall v \in N(A)$

例: 约束仅为非负约束

若存在 $x, x \geq 0 \Leftrightarrow \nabla f_0^T(x)(y-x) \geq 0$

由 y 的非负性和 $\nabla f_0^T(x) \geq 0$

(否则 $\exists y \in N(\nabla f_0^T(x))$, $y = -\frac{1}{\|\nabla f_0^T(x)\|} \nabla f_0^T(x)$, $\nabla f_0^T(x)^T \nabla f_0^T(x) \geq 0$)

$\nabla f_0^T(x)^T \nabla f_0^T(x) \geq 0$

若 $\nabla f_0^T(x)v = 0, v \in N(\nabla f_0^T(x))$

则 $\nabla f_0^T(x)x \geq 0$

再由 $y = 0 \rightarrow \nabla f_0^T(x)x \geq 0$, $\nabla f_0^T(x)x = 0$

另在 $x \geq 0$ 若 $\begin{cases} \nabla f_0^T(x) \geq 0 \\ \nabla f_0^T(x)x = 0 \end{cases} \Rightarrow \forall y \geq 0, \nabla f_0^T(x)y \geq 0$

设 y 为最优点 $\begin{cases} y \geq 0 \\ \nabla f_0^T(x)y \geq 0 \end{cases} \Rightarrow \begin{cases} y \geq 0 \\ (\nabla f_0^T(x))^T y = 0 \end{cases}$

Part 1

线性规划

线性规划 $\min c^T x + d$, $x \in \mathbb{R}^n$, $d \in \mathbb{R}$
 s.t. $Gx \leq h$, $G \in \mathbb{R}^{m_1 \times n}$, $h \in \mathbb{R}^{m_1}$
 $Ax = b$, $A \in \mathbb{R}^{m_2 \times n}$, $b \in \mathbb{R}^{m_2}$

目标和约束方程组

线性规划的图解法
 $\min c^T x + d$
 s.t. $Gx \leq h$
 $Ax = b$
 $x \geq 0$

$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$

线性规划的标量形式
 $\min c^T x + d$
 s.t. $Gx \leq h$
 $Ax = b$
 $x \geq 0$

$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$

线性分式规划 (Linear fractional programming)
 (P0) $\min f(x)$
 s.t. $Gx \leq h$, $f(x) = \frac{c^T x + d}{e^T x + f}$, $e^T x + f > 0$
 $Ax = b$, $A \in \mathbb{R}^{m_1 \times n}$, $b \in \mathbb{R}^{m_1}$

(P1) $\min c^T x + d$
 s.t. $Gx \leq h \leq 0$
 $Ax = b$
 $c^T x + d = 1$
 $x \geq 0$

Part 2

二次规划和二次约束二次规划

二次规划 (Quadratic Programming)
 $\min \frac{1}{2} x^T P x + q^T x + r$, $P \in \mathbb{S}_+$
 s.t. $Gx \leq h$, $G \in \mathbb{R}^{m_1 \times n}$, $h \in \mathbb{R}^{m_1}$
 $Ax = b$, $A \in \mathbb{R}^{m_2 \times n}$, $b \in \mathbb{R}^{m_2}$

二次约束二次规划
 $\min \frac{1}{2} x^T P x + q^T x + r$, $P \in \mathbb{S}_+$
 s.t. $\frac{1}{2} x^T P x + q^T x + r \leq 0$, $Ax = b$

例. 带噪声的测量系统 $b = Ax + e$
 $\hat{x} = \arg \min_x \|b - Ax\|_2^2$
 $\hat{x} = \arg \min_x \|b - Ax\|_2^2 = \arg \min_x x^T A^T A x - 2b^T A x + b^T b$ ($\nabla x = Ab$)
 若 $A^T A$ 正定, $(A^T A)^{-1} A^T b$

(2). 若 x 稳定, $\hat{x} = \arg \min_x \|b - Ax\|_2^2 + \lambda_1 \|x\|_1$, LASSO
 $\hat{x} = \arg \min_x \|b - Ax\|_2^2 + \lambda_1 \|x\|_1$, L_1 -regularized least squares

$x = x^+ - x^-$
 $\hat{x} = \arg \min_{x^+, x^-} \|b - Ax^+ + Ax^-\|_2^2 + \lambda_1 \|x^+ - x^-\|_1$
 $\hat{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, x^+ = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, x^- = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

例. 若 x 中元素相等类似
 $\hat{x} = \arg \min_x \|b - Ax\|_2^2 + \lambda_2 \|x\|_2$, Ridge regression

$\hat{x} = \arg \min_x \|b - Ax\|_2^2$, QCQP
 s.t. $\|x\|_2 \leq 0$

一般的优化问题
 $\min f(x)$
 s.t. $f'(x) = 0$, $f''(x) \geq 0$
 $x \in \mathbb{R}^n$, $D = \bigcap_{i=1}^m \text{dom}(f_i) \bigcap \bigcap_{i=1}^n \text{dom}(h_i)$, f 为凸函数

Part 3

拉格朗日函数

拉格朗日函数 (Lagrange Function)
 $L(x, \lambda, v) = f(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^n v_i h_i(x)$

对偶函数 / 拉格朗日对偶函数 (Dual Function/Lagrange Dual Function)
 $g(\lambda, v) = \inf_{x \in D} L(x, \lambda, v)$, λ, v 称为拉格朗日乘子

重要性质 1. 对偶函数为凸函数
 2. $\forall \lambda \geq 0, \forall v, g(\lambda, v) \leq p^*$

证明: 设 x^* 为原问题的最优解, 则 x^* 可行, 那么
 $f_i(x^*) \leq 0, i=1, \dots, m, h_i(x^*) = 0, i=1, \dots, n$
 于是 $\forall \lambda \geq 0, \forall v$ 有 $\sum_{i=1}^m \lambda_i f_i(x^*) + \sum_{i=1}^n v_i h_i(x^*) \leq 0$
 从而 $L(x^*, \lambda, v) = f_0(x^*) + \sum_{i=1}^m \lambda_i f_i(x^*) + \sum_{i=1}^n v_i h_i(x^*) \leq p^*$
 于是 $g(\lambda, v) \leq p^*$

例. $\min x^T x$
 s.t. $Ax = b$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$

拉格朗日函数 $L(x, v) = x^T x + v^T (Ax - b)$
 对偶函数 $g(v) = \inf_{x \in D} L(x, v) = \inf_{x \in D} (x^T x + v^T Ax - v^T b)$

$2x + A^T v = 0 \Rightarrow x = -\frac{1}{2} A^T v$, $g(v) = \frac{v^T A A^T v}{2} - v^T b - \frac{1}{2} v^T A A^T v - b^T v$

例. $\min c^T x$
 s.t. $Ax = b$, $x \in \mathbb{R}^n$

拉格朗日函数 $L(x, \lambda, v) = c^T x + \sum_{i=1}^m \lambda_i (A_i x - b_i) + \sum_{i=1}^n v_i (x_i - 1)$
 对偶函数 $g(v) = \inf_{x \in D} L(x, \lambda, v) = \inf_{x \in D} (c^T x + \sum_{i=1}^m \lambda_i (A_i x - b_i) + \sum_{i=1}^n v_i (x_i - 1))$

例. $\min x^T W x$ 此处限制条件修改为: $x_i^2 - 1 = 0$

拉格朗日函数 $L(x, \lambda, v) = x^T W x + \sum_{i=1}^n v_i (x_i^2 - 1)$
 对偶函数 $g(v) = \inf_{x \in D} L(x, \lambda, v) = \inf_{x \in D} (x^T W x + \sum_{i=1}^n v_i (x_i^2 - 1))$

Part 1 对偶问题

对偶问题 (Dual Problem) $(D) \begin{cases} \max g(\lambda, v) \\ \text{s.t. } \lambda \geq 0 \end{cases}$ <p style="text-align: center;">d^*: 对偶问题的最优值</p>	原问题 (Primal Problem) $(P) \begin{cases} \min f(x) \\ \text{s.t. } f_i(x) \leq 0, i=1, \dots, m \\ Ax = b, x \geq 0 \end{cases}$ <p style="text-align: center;">p^*: 原问题的最优值</p>
---	--

$d^* \leq p^*$

例. $\min f_0(Ax+b)$ 拉格朗日函数: $L(x, \lambda) = f_0(Ax+b) - \lambda^T(Ax+b)$ 对偶函数: $g(\lambda, v) = \inf_x L(x, \lambda) = \inf_x f_0(Ax+b) - \lambda^T(Ax+b)$ 对偶问题: $\max g$

$$\begin{cases} \min f_0(y) \\ \text{s.t. } Ax+b=y \end{cases}$$

对偶问题: $\max g(v) = \inf_y \begin{cases} f_0(y) + v^T(Ax+b-y) \\ \text{s.t. } Ax+b=y \end{cases} = \inf_y f_0(y) + v^T(Ax+b) - \inf_y v^T(Ax+b), v \geq 0 \end{cases}$

例. (P1). $\begin{cases} \min c^T x \\ \text{s.t. } Ax=b, x \geq 0 \end{cases}$

对偶问题: $\begin{cases} \max -b^T v + c^T v - \lambda^T x \\ \text{s.t. } v^T A = 0 \end{cases}$

$g(\lambda, v) = \inf_x \begin{cases} -b^T v + (c+A^T v - \lambda)^T x \\ \text{s.t. } Ax=b \end{cases} = \begin{cases} -b^T v, c+A^T v - \lambda = 0 \\ -\infty, \text{otherwise} \end{cases}$

(D1) $\begin{cases} \max -b^T v \\ \text{s.t. } c+A^T v - \lambda = 0 \\ \lambda \geq 0 \end{cases}$ 对偶问题: $\begin{cases} \min b^T v \\ \text{s.t. } c+A^T v \geq 0 \end{cases}$

例. (P2): $\begin{cases} \min b^T x \\ \text{s.t. } A^T x + c \geq 0 \end{cases}$

拉格朗日函数: $L(x, \lambda) = b^T x + \lambda^T(-A^T x - c) = (b - A^T \lambda)^T x - c^T \lambda$

对偶函数: $g(\lambda) = \inf_x L(x, \lambda) = \begin{cases} -c^T \lambda, A\lambda = b \\ -\infty, A\lambda \neq b \end{cases}$

对偶问题: $\begin{cases} \max -c^T \lambda \\ \text{s.t. } A\lambda = b \\ \lambda \geq 0 \end{cases}$

Part 2 Duality

弱对偶性 (Weak Duality): $d^* \leq p^*$
 强对偶性 (Strong Duality): $d^* = p^*$
 对偶间隙 (Duality Gap): $p^* - d^*$
 D 的相对内部 (Relative Interior):
 $\text{Relint}(D) = \{x \in D | \exists r > 0, B(x, r) \cap \partial(D) \neq \emptyset\}$

Slater 条件 ($d^* = p^*$ 的充分条件)
 若有凸问题 $\min f(x)$
 s.t. $f_i(x) \leq 0, i=1, \dots, m$
 $Ax = b$
 其中 $f_i(x), i=0, \dots, m$ 为凸, 当存在 $x \in \text{Relint}(D)$, 使得 $f_i(x) < 0, i=1, \dots, m$
 和 $Ax = b$ 满足, 则 $p^* = d^*$

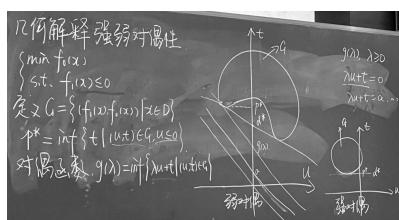
A Weaker Slater's Condition
 若上述凸问题是前 k 个不等式约束为仿射的, 则当存在
 $x \in \text{Relint}(D)$, 且 $f_i(x) \leq 0, i=1, \dots, k, f_i(x) < 0, i=k+1, \dots, m$ 和
 $Ax = b$, 则 $p^* = d^*$
 若线性规划问题存在可行解, 则必有 $p^* = d^*$

例. $\begin{cases} \min x^T x \\ \text{s.t. } Ax=b \end{cases}$ 对偶问题: $\max_{v \in S_{++}^n} v^T A^T A v - v^T v$
 B). QCQP
 $\begin{cases} \min \frac{1}{2} x^T P_0 x + q_0^T x + r_0 \\ \text{s.t. } \frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0, i=1, \dots, m \end{cases}$

拉格朗日函数: $L(x, \lambda) = \frac{1}{2} x^T P_0 x + q_0^T x + r_0 + \sum_{i=1}^m \lambda_i (\frac{1}{2} x^T P_i x + q_i^T x + r_i)$
 $= \frac{1}{2} x^T (P_0 + \sum_{i=1}^m \lambda_i P_i) x + q_0^T x + r_0 + \sum_{i=1}^m \lambda_i r_i$

至少有一个 $P_i, i=1, \dots, m$ 是正定的

对偶函数: $g(\lambda) = \inf_x L(x, \lambda) \stackrel{\lambda \geq 0}{=} -\frac{1}{2} \lambda^T P_0^{-1} \lambda + q_0^T \lambda + r_0$
 对偶问题: $\begin{cases} \max -\frac{1}{2} \lambda^T P_0^{-1} \lambda + q_0^T \lambda + r_0 \\ \text{s.t. } \lambda \geq 0 \end{cases}$
 若 $\exists x \in \text{Relint}(D) = \text{Relint}(R^n) = R^n$, 使得 $\frac{1}{2} x^T P_0 x + q_0^T x + r_0 < 0$, 则 $p^* = d^*$
 若 $q_0 = 0, r_i = 0, \forall i=0, \dots, m$, Slater 条件为 $\frac{1}{2} x^T P_0 x < 0, i=1, \dots, m$
 事实上, $d^* = p^*$



Part 1

鞍点

鞍点解释 (Saddle Point)

极大-极小不等式：若函数 $f(w, z)$ 定义在 $w \in S_w, z \in S_z$ 上，则有 $\sup_{z \in S_z} \inf_{w \in S_w} f(w, z) \leq \inf_{w \in S_w} \sup_{z \in S_z} f(w, z)$.

鞍点、李定行-函数 $L(x, \lambda)$ 定义在 $w \in S_w, z \in S_z$ 上，称其为 (\bar{w}, \bar{z}) 的鞍点。
 $\arg \max_{w \in S_w} \inf_{z \in S_z} f(w, z) = w^*$, $\inf_{w \in S_w} f(w, z^*)$ 为 z^* 的值。
 成：倘若 $\forall w \in S_w, \forall z \in S_z$, 有 $f(w, z) \leq f(\bar{w}, \bar{z}) \leq f(w, z^*)$.

原问题： $\min_x f_i(x)$
st. $f_j(x) \leq 0, j=1, \dots, m$

拉格朗日函数： $L(x, \lambda) = f(x) + \sum_{j=1}^m \lambda_j f_j(x)$

$$\sup_{\lambda \geq 0} \{L(x, \lambda)\} = \begin{cases} f_0(x), & f_0(x) \leq 0 \\ +\infty, & \text{else} \end{cases}$$

$$P^* = \min_x \{f_0(x) | f_j(x) \leq 0, j=1, \dots, m\} = \inf_{x \geq 0} \sup_{\lambda \geq 0} L(x, \lambda)$$

$$d^* = \sup_{\lambda \geq 0} \inf_x L(x, \lambda). \quad P^* = d^* \text{ 莱布尼茨}$$

鞍点定理：

若 $(\bar{x}, \bar{\lambda})$ 是在拉格朗日函数 $L(x, \lambda)$ 的鞍点 \Leftrightarrow 对偶问题且对偶问题成立
 证明： \Rightarrow 因为 $(\bar{x}, \bar{\lambda})$ 是 $L(x, \lambda)$ 的鞍点，则和对偶问题成立
 $\sup_{\lambda \geq 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \geq 0} L(x, \lambda)$ (对偶问题成立)
 且 $\bar{\lambda} = \arg \sup_{\lambda \geq 0} \inf_x L(x, \lambda)$. $\bar{x} = \inf_x \sup_{\lambda \geq 0} L(x, \lambda)$

\Leftarrow 因为 $(\bar{x}, \bar{\lambda})$ 为原问题和对偶问题的最优解，则有
 则 $f_i(\bar{x}) \leq 0, i=1, \dots, m$ ①
 由对偶问题成立知
 $f_0(\bar{x}) = g(\bar{\lambda}) = \inf_x (f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)) \leq f_0(\bar{x}) + \sum_{i=1}^m \lambda_i f_i(\bar{x})$
 因此 $\inf_x L(x, \bar{\lambda}) = \underline{L}(\bar{x}, \bar{\lambda}) = f_0(\bar{x})$. 从而 $\underline{L}(\bar{x}, \bar{\lambda}) \leq L(\bar{x}, \bar{\lambda})$
 再由 $\sup_{\lambda \geq 0} L(\bar{x}, \lambda) = \sup_{\lambda \geq 0} (f_0(\bar{x}) + \sum_{i=1}^m \lambda_i f_i(\bar{x})) \leq f_0(\bar{x}) = L(\bar{x}, \bar{\lambda})$
 可知 $L(\bar{x}, \bar{\lambda}) \leq \underline{L}(\bar{x}, \bar{\lambda})$

Part 2

KKT 条件

原问题： $\min_x f_0(x)$
st. $f_i(x) \leq 0, i=1, \dots, m$
 $h_i(x) = 0, i=1, \dots, p$

对偶问题： $g(\lambda) = \inf_x (f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \frac{1}{2} \sum_{i,j=1}^p \lambda_i \lambda_j h_i(x) h_j(x))$

对偶问题： (D) **st.** $\lambda \geq 0$

假设 $\left\{ \begin{array}{l} \text{对偶间隙} 0, d^* = 0 \\ \text{所有函数可微} \end{array} \right.$

记 x^*, λ^*, v^* 为原问题对偶问题的解，则以平行
 $f_i(x^*) \leq 0, i=1, \dots, m$
 $h_i(x^*) = 0, i=1, \dots, p$
 $\lambda^* \geq 0$
 由 $d^* = \inf_{\lambda \geq 0} f_0(x^*) = g(\lambda^*) = \inf_x (f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \frac{1}{2} \sum_{i,j=1}^p \lambda_i \lambda_j h_i(x) h_j(x))$
 $\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \frac{1}{2} \sum_{i,j=1}^p \lambda_i^* \lambda_j^* h_i(x^*) h_j(x^*) = f_0(x^*)$
 提 ① $\sum_{i=1}^m \lambda_i^* f_i(x^*) = 0$ 从而 $\sum_{i=1}^m \lambda_i^* f_i(x^*) = 0, i=1, \dots, m$
 则有 $\sum_{i=1}^m \lambda_i^* f_i(x^*) = 0 \Rightarrow \lambda^* = 0$
 $\left\{ \begin{array}{l} f_0(x^*) = 0 \\ \lambda^* = 0 \end{array} \right. \text{互补松弛条件 (Complementary Slackness)}$

② 因 $L(x^*, \lambda^*, v^*) = \underline{L}(x^*, \lambda^*, v^*)$, 由拉格朗日函数的弱对偶性知
 $\underline{L}(x^*, \lambda^*, v^*) \Big|_{\lambda^* \geq 0} = 0$ 即 $\nabla_{\lambda} \underline{L}(x^*, \lambda^*, v^*) = 0$ 稳定性条件 (Stability)

KKT 条件 原问题可行解条件 (I) x^* 为 x^* 的稳定点
 对偶问题可行解条件 (II) λ^* 为 λ^* 的稳定点
 补充松弛条件 (Karush-Kuhn-Tucker)
 和对偶条件

若原问题是凸问题，各函数分段线性，对偶间隙为零，则 KKT 条件
 为充要条件

证明：(充分性) 设 (x^*, λ^*) 满足 KKT 条件，即

$$\left\{ \begin{array}{l} f_i(x^*) \leq 0, i=1, \dots, m \\ h_i(x^*) = 0, i=1, \dots, p \\ \lambda^* \geq 0 \\ \sum_{i=1}^m \lambda_i^* f_i(x^*) = 0 \\ \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \lambda_i^* \nabla h_i(x^*) = 0 \end{array} \right.$$

由原问题凸，则 $L(x, \lambda, v) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \frac{1}{2} \sum_{i,j=1}^p \lambda_i \lambda_j h_i(x) h_j(x)$
 并且 KKT 条件和 $\underline{L}(x, \lambda, v) \leq L(x, \lambda, v)$ 令 $\lambda^* = 0$, 得 $\underline{L}(x, \lambda, v) = L(x, \lambda, v)$
 即 $\underline{L}(x, \lambda, v) = L(x, \lambda, v)$ 令 $\lambda^* = 0$, 得 $\underline{L}(x, \lambda, v) = L(x, \lambda, v)$
 又 $\underline{L}(x, \lambda, v) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \frac{1}{2} \sum_{i,j=1}^p \lambda_i \lambda_j h_i(x) h_j(x) = L(x, \lambda, v)$
 从而 $L(x, \lambda, v) = L(x, \lambda, v)$ 为非空闭合的凸集上的点，即 x^* 为原问题的解。



Part 1 KKT条件

例. 二次规划问题
 $\min \frac{1}{2} x^T P x + q^T x + r$, $P \in \mathbb{R}^{n \times n}$
 s.t. $Ax = b$

写出 KKT 条件: $Ax = b$

$$\begin{pmatrix} P & A \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ v^* \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}$$

例. 注水问题 (Water filling)
 $\min_{\substack{x \geq 0 \\ Ax = b}} \frac{1}{2} x^T L x$, $x \in \mathbb{R}^n$, $x \geq 0$, $Ax = b$

写出 KKT 条件: $x^* \geq 0$, $L x^* = 0$, $\lambda^* = 1$

$$\lambda^* \left(\frac{1}{2} x^T L x + \frac{1}{2} x^T A^T \lambda + b^T \lambda \right) = 0 \Rightarrow \lambda^* = \frac{1}{2} x^T L^{-1} A^T \lambda + b^T \lambda$$

$V^* \geq \frac{1}{2} (x_i + x_{i+1})$, $i = 1, \dots, n$

$$\lambda_i^* \left(V^* - \frac{1}{2} (x_i + x_{i+1}) \right) = 0, i = 1, \dots, n$$

1. 对于使得 $V^* \geq \frac{1}{2} (x_i + x_{i+1})$ 成立的 i , 有 $x_i^* = 0$

2. 对于使得 $V^* < \frac{1}{2} (x_i + x_{i+1})$ 成立的 i , 有 $x_i^* > 0$, 则 $V^* - \frac{1}{2} (x_i + x_{i+1}) = 0$, 即 $V^* = \frac{1}{2} (x_i + x_{i+1})$

于是 $x_i^* = \max \left\{ 0, \frac{1}{2} (x_i + x_{i+1}) \right\}$

KKT 条件与目标函数引数时最优性准则

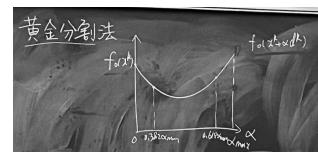
$$\begin{cases} \min f(x) \\ \text{s.t. } g_i(x) \geq 0, i=1, \dots, m \\ h_j(x) = 0, j=1, \dots, l \end{cases} \Leftrightarrow \begin{cases} x \geq 0 \\ \nabla f(x) = 0 \\ \nabla f_i(x) \geq 0, i=1, \dots, m \\ (\nabla f_i(x))^T \lambda = 0, i=1, \dots, m \\ h_j(x) = 0, j=1, \dots, l \end{cases}$$

KKT 条件: $x^* \geq 0$, $\nabla f(x^*) = 0$, $\lambda_i^* (\nabla f_i(x^*)) = 0$, $\nabla f_i(x^*) + (\lambda_i^*)^T \nabla h_j(x^*) = 0 \Leftrightarrow \lambda_i^* = \nabla f_i(x^*)$

Part 2 优化问题

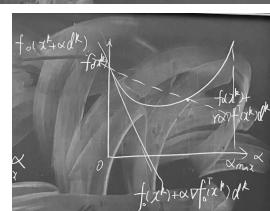
优化算法

a. 所有优化算法都是迭代算法
 在第 k 步, $x^{k+1} = x^k + \alpha^k d^k$; d^k 在第 k 步时的解
 α^k 是 k 步方向
 Δx^k
 $\alpha^k = \arg \min_{\alpha \geq 0} f(x^k + \alpha d^k)$ x^k 为 k 步



Armijo-Goldstein 法 (回溯法 backtracking)

核心思想: 1. 目标函数值有足够下降
 2. 搜索步长 α 不太小
 若 $f_0(x^k + \alpha^k d^k) > f_0(x^k) + \gamma \alpha^k \nabla f_0(x^k)^T d^k$, 则 $\alpha^k \leftarrow \beta \alpha^k$, 令 $\beta \in (0, 1)$



$\min f(x)$, 若 $f(x)$ 为凸数, KKT 条件为 $\nabla f(x) = 0$

(1): 当 $\nabla f(x) \leq 0$ 时 $x \rightarrow x^*$?
 $\nabla f(x) \leq f'(x)$

2. 假设 $f(x)$ 为凸且有强凸性
 强凸性: 存在 $m > 0$, $\forall x \in \text{dom}(f)$, $\nabla^2 f(x) \geq mI$

等价: $\nabla f(y) \in \text{dom}(f) \Rightarrow f(y) \geq f(x) + \frac{1}{2} m \|y - x\|^2$

当 $\nabla f(x) > 0$, $x \rightarrow x^*$?
 由强凸性知 $\nabla f(x) \geq f(x) + \frac{1}{2} m \|x - x^*\|^2$
 由强凸性 $\|x - x^*\| \geq \frac{\|\nabla f(x)\|}{m}$
 $\|x - x^*\| \geq \frac{\|\nabla f(x)\|}{m} \geq \frac{\|\nabla f(x)\|}{m} \geq \frac{\|\nabla f(x)\|}{m} \geq \frac{\|\nabla f(x)\|}{m}$
 再由 $f(x) > f(x^*)$ 知 $f(x) > f(x^*) - \frac{1}{2} m \|x - x^*\|^2 \geq f(x^*) + \frac{1}{2} m \|x - x^*\|^2$
 于是 $\|x - x^*\| \geq \frac{1}{m} \|x - x^*\| \geq \frac{1}{m} \|x - x^*\| \geq \frac{1}{m} \|x - x^*\|$
 $\|x - x^*\| \leq \frac{1}{m} \|x - x^*\|$

当 $\nabla f(x) = 0$ 时, $f(x) \rightarrow f(x^*)$?
 给定 x , $-f(x) + \nabla f(x)^T (y - x) + \frac{1}{2} m \|y - x\|^2$ 等于凸
 即 $\nabla f(x) + m(y - x) = 0$, 有 $y = x - \frac{\nabla f(x)}{m}$
 知上式最小值为 $f(x) - \frac{1}{2m} \|\nabla f(x)\|^2$
 假令强凸性有 $\forall x \in \text{dom}(f)$, $f(x) \geq f(x^*) + \frac{1}{2m} \|\nabla f(x)\|^2$
 于是 $f(x) \geq f(x^*) + \frac{1}{2m} \|\nabla f(x)\|^2$, 且 $f(x) \leq f(x^*) + \frac{1}{2m} \|\nabla f(x)\|^2$
 有 $\|f(x) - f(x^*)\| \leq \frac{1}{2m} \|\nabla f(x)\|^2$

Part 1 梯度下降法

$$\begin{aligned}
 p^* &= f(x^k) \geq f(x) - \frac{1}{2m} \|\nabla f(x)\|_2^2 \\
 \|f(x) - p^*\|_2 &\leq \frac{1}{2m} \|\nabla f(x)\|_2^2 \\
 \|x - x^k\|_2 &\leq \frac{1}{m} \|\nabla f(x)\|_2 \\
 \boxed{\text{由已知 } \exists M > 0. \forall x \in \text{dom}(f), \nabla^2 f(x) \leq M} \\
 \forall x, y \in \text{dom}(f), f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{1}{2} M \|y - x\|_2^2 \\
 p^* &\leq f(x) - \frac{1}{2M} \|\nabla f(x)\|_2^2
 \end{aligned}$$

梯度下降法: $d^k = -\nabla f(x^k)$

重复

1. $\alpha^k = \arg \min_{\alpha} f(x^k + \alpha d^k), 0 \leq \alpha \leq \alpha_{\max}$
2. $x^{k+1} = x^k + \alpha^k d^k$

直到停止准则

算法收敛性分析: 假设 $\forall x \in \text{dom}(f), M \geq \|\nabla^2 f(x)\|_2 \geq m$

记 $\tilde{f}(x) = f(x^k + \alpha d^k) = f(x^k + \nabla f(x^k)) \geq f(x^k)$

由假设可知, $f(x^{k+1}) \leq f(x^k) + \nabla f(x^k)^T (-\nabla f(x^k)) + \frac{M}{2} \|x^k - x^{k+1}\|_2^2$

于是 $\tilde{f}(x) \leq f(x^k) - \alpha \|\nabla f(x^k)\|_2^2 + \frac{M}{2} \alpha^2 \|\nabla f(x^k)\|_2^2$

$\forall x, y \in \text{dom}(f), f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{1}{2} M \|y - x\|_2^2$

精确线搜索:

式右边关于 α 为凸函数, 且 $\tilde{f}(x) - f(x^k) - \alpha \|\nabla f(x^k)\|_2^2 + M \alpha \|\nabla f(x^k)\|_2^2 = 0 \Rightarrow \alpha = \frac{1}{M}$

$\min_{\alpha} \tilde{f}(x) \leq f(x^k) - \frac{1}{M} \|\nabla f(x^k)\|_2^2 + \frac{1}{2m} \|\nabla f(x^k)\|_2^2$

于是 $f(x^{k+1}) \leq f(x^k) - \frac{1}{2m} \|\nabla f(x^k)\|_2^2$

$\uparrow^k = f(x^k) - f(x) - \frac{1}{2m} \|\nabla f(x^k)\|_2^2$

且 $\frac{1}{2m} \|\nabla f(x^k)\|_2^2 + f(x^{k+1}) - \uparrow^k \leq f(x^k) - \uparrow^k$

$- \frac{1}{2m} \|\nabla f(x^k)\|_2^2 + f(x^k) - \uparrow^k \leq 0$

$M(f(x^{k+1}) - p^*) + m(f(x^k) - p^*) \leq M(f(x^k) - p^*)$

则 $f(x^{k+1}) - p^* \leq (1 - \frac{m}{M})(f(x^k) - p^*)$

从而 $f(x^k) - p^* \leq (1 - \frac{m}{M})^k (f(x^0) - p^*)$

$\|f(x^{k+1}) - p^*\|_2 \leq C \|f(x^k) - p^*\|_2$

非精确线搜索: (回溯法: 若 $f(x^k + \alpha d^k) > f(x^k) + r \alpha \|\nabla f(x^k)\|_2^2$)

先说明当 $0 \leq \alpha \leq \frac{1}{M}$ 时, 回溯停止: $f(x) = f(x^{k+1}) = f(x^k + \alpha d^k) \leq f(x^k) + r \alpha \|\nabla f(x^k)\|_2^2$

因当 $0 \leq \alpha \leq \frac{1}{M}$ 时, $-\alpha + \frac{M}{2} \alpha^2 \leq \frac{\alpha}{2} (-\frac{\alpha}{2} + \frac{M}{2} \alpha^2 = \frac{\alpha}{2} (\alpha - 1) \leq 0)$

于是 $\tilde{f}(x) \leq f(x^k) - \alpha \|\nabla f(x^k)\|_2^2 + \frac{M}{2} \alpha^2 \|\nabla f(x^k)\|_2^2 \leq f(x^k) - \frac{\alpha}{2} \|\nabla f(x^k)\|_2^2 \leq f(x^k) - r \alpha \|\nabla f(x^k)\|_2^2$

因此终止时 $\alpha = \alpha_{\max}$ 或 $\alpha > \frac{1}{M}$ (若不然, $\alpha < \frac{1}{M}$, 上一步 $\alpha = \frac{1}{M}$ 时 $\alpha = \frac{1}{M}$)

$\therefore \tilde{f}(x^{k+1}) \leq f(x) - \min\left\{r\alpha_{\max}, \frac{r}{M}\right\} \|\nabla f(x)\|_2^2 \leq \min\left\{r\alpha_{\max}, \frac{r}{M}\right\} \|\nabla f(x)\|_2^2$

类似与精确线搜索的推导, 这里有

$\log(f(x^k) - p^*) - \log(f(x^{k+1}) - p^*) \leq (1 - \min\left\{2m\alpha_{\max}, \frac{2m}{M}\right\})(f(x^k) - p^*)$

$\frac{\|f(x^{k+1}) - p^*\|}{\|f(x^k) - p^*\|} = \frac{\|f(x^k) - p^*\|}{\|f(x^k) - p^*\|} = \frac{1}{C} \leq 1$

以 $f(x) = \frac{1}{2} x^T P x, P \in S_+^n, \nabla^2 f(x) = P$, M 为特征值, m 为特征值

$\frac{M}{m}$ 矩阵的条件数

$P = \begin{pmatrix} 100 & 0 \\ 0 & 1 \end{pmatrix}$

Part 1

无约束优化

最速下降法

$$\nabla f(x^k + v) \rightarrow \min_v \nabla f(x^k) + \nabla^2 f(x^k)v \rightarrow \min_v \|\nabla f(x^k) + \nabla^2 f(x^k)v\| \quad \text{令 } \|v\|=1 \Rightarrow v = d^k$$

1. ∞ 范数 $\|v\|_\infty = 1 \quad d^k = -\frac{\nabla f(x^k)}{\|\nabla f(x^k)\|_2}$

2. l_2 范数 $\|v\|_2 = 1 \quad d^k = \frac{\nabla f(x^k)}{\|\nabla f(x^k)\|_2}$

$\nabla f(x) = \begin{pmatrix} (\nabla f(x))_1 \\ \vdots \\ (\nabla f(x))_n \end{pmatrix}$

坐标轮换法 $d_j^k = \begin{cases} -1 & (\nabla f(x^k))_j > 0 \\ 1 & (\nabla f(x^k))_j < 0 \end{cases}$

坐标的绝对值 $\left| \frac{\partial f(x^k)}{\partial x_i} \right| > 0$

3. ∞ 范数 $\|v\|_\infty = 1 \quad d^k = \begin{cases} -1 & (\nabla f(x^k))_j > 0 \\ 1 & (\nabla f(x^k))_j < 0 \end{cases}$

变种:

坐标轮换法 $d^k = e_{\text{middle point}}$ C: 第 i 分量为 1, 其余分量为 0
步长 $\alpha \in [0.5, 1, \dots, X_{\max}]$

2. 次梯度 $\frac{\partial f(x)}{\partial x} = \begin{cases} \frac{\partial f(x)}{\partial x_i} & \text{左端点处任一点, 取某} \\ & \text{次梯度在 } [-1, 1] \end{cases}$

(3). $\min f(x) = \frac{1}{2} \|Ax-b\|_2^2 + \lambda \|x\|_1$

$$\frac{\partial f(x)}{\partial x} = A^T(Ax-b) + \lambda \frac{\partial \|x\|_1}{\partial x}$$

$$\left(\frac{\partial \|x\|_1}{\partial x} \right)_i = \begin{cases} 1 & , x_i > 0 \\ -1 & , x_i < 0 \\ \in [-1, 1] & , x_i = 0 \end{cases}$$

牛顿法

$$d^k = \arg \min_v \{f(x^k) + \nabla f(x^k)v + \frac{1}{2} v^T \nabla^2 f(x^k)v\}$$

$$\nabla^2 f(x^k) + \nabla^2 f(x^k)v = 0 \Rightarrow d^k = -(\nabla^2 f(x^k))^{-1} \nabla f(x^k)$$

对于一二阶可微的优化问题, 能够找到 $v > 0$,

① 在 $\|\nabla f(x)\|_2 \geq \eta$, 阻尼牛顿法

② 在 $\|\nabla f(x)\|_2 \leq \eta$, 二次收敛阶段 $\frac{f(x^k) - p}{\|f(x^k) - p\|^{1/(0.1)}}$

拟牛顿法 Quasi-Newton's Method

$$\nabla^2 f(x^k) d^k = -\nabla f(x^k)$$

左边用矩阵 B 代替 $\nabla^2 f(x^k)$, 满足 $Bd^k = -\nabla f(x^k)$

BFGS

Part 2

带约束的优化

有约束的优化问题

凸、等式约束 $\begin{cases} \min f(x) \\ \text{s.t. } Ax=b \end{cases}$

其 KKT 条件为 $Ax^*=b$.

$$\nabla f(x^*) + A^T v^* = 0$$

一、线性方程组

$$\begin{cases} \min \frac{1}{2} x^T P x + q^T x + r, \quad P \in S^n_+ \\ \text{s.t. } Ax=b \end{cases}$$

KKT: $\begin{cases} Ax^*=b \\ P x^* + q + A^T v^* = 0 \end{cases}$

整理 $\begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ v^* \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}$

二、非线性方程组

$$\begin{cases} \min f(x^k + d) \\ \text{s.t. } A(x^k + d) = b \Rightarrow Ad=0 \end{cases}$$

KKT 条件 $\begin{pmatrix} \nabla^2 f(x^k) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} d^k \\ v^* \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ 0 \end{pmatrix} \Rightarrow d^k = -\nabla f(x^k)$

$x^k = \arg \min_{x \geq 0} f(x^k + \alpha d^k)$

$x^{k+1} = x^k + \alpha x^k$

Part 1 拉格朗日乘子法

$$(P) \begin{cases} \min f(x) \\ \text{s.t. } Ax = b \end{cases} \quad \text{KKT 条件} \quad \begin{cases} Ax^* = b \\ \nabla f(x^*) + A^T v^* = 0 \end{cases} \quad \text{Lagrange Multiplier}$$

拉格朗日法 / 拉格朗日乘子法 (Lagrangian Method, Method of Multipliers)

$$\begin{aligned} x^{k+1} &= x^k - \alpha^k (\nabla f(x^k) + A^T v^k) \\ v^{k+1} &= v^k + \alpha^k (Ax^k - b) \end{aligned}$$

鞍点解释

$$L(x, v) = f(x) + v^T (Ax - b)$$

若 x^*, v^* 是 $L(x, v)$ 的鞍点, 那 x^*, v^* 同时是以下两个问题的最优解

$$\begin{cases} (x^*, v^*) = \arg \max_v \min_x L(x, v) \\ (x^*, v^*) = \arg \max_x \min_v L(x, v) \end{cases}$$

于是 $x^* = \arg \min_v L(x^*, v^*)$, $v^* = \arg \max_v L(x^*, v)$

对于 x^* 的求解, 使用梯度下降法, 负梯度为 $-\nabla f(x^k) - A^T v^k$, 且 $x^{k+1} = x^k - \alpha^k (-\nabla f(x^k) - A^T v^k)$
使用 v^k 逼近 v^*

对于 v^* 的求解, 关于 v 的梯度方向为 $Ax^k - b$ (在 x^* 固定时)

使用对偶代替 x^k 有 $v^{k+1} = v^k + \alpha^k (Ax^k - b)$

增广拉格朗日法, Augmented Lagrangian Method

增广拉格朗日函数: $L_c(x, v) = f(x) + v^T (Ax - b) + \frac{c}{2} \|Ax - b\|^2$

其为 $(P') \begin{cases} \min_x f(x) + \frac{c}{2} \|Ax - b\|^2 \\ \text{s.t. } Ax = b \end{cases}$

$P \Leftarrow P' = 1^\circ$ 最优解一样 2° 对偶问题的最优解一样

使用拉格朗日法求解 P'

$$\begin{cases} x^{k+1} = x^k - \alpha^k \nabla_x L_c(x^k, v^k) \\ v^{k+1} = v^k + \alpha^k (Ax^k - b) \end{cases}$$

若原问题 KKT 条件满足, $\nabla_x L(x, v) = 0, Ax = b$

若 P' 的 KKT 条件满足, $\nabla_x L_c(x^k, v^k) = 0, Ax^k = b$

在此基础上, 增广拉格朗日法进行改进:

$$\begin{cases} 1' x^{k+1} = \arg \min_x L_c(x, v^k) \\ 2' v^{k+1} = v^k + c(Ax^{k+1} - b) \end{cases}$$

好性质

1° 若 $v = v^*$, 则 $\forall c > 0$, $x^* = \arg \min_x L_c(x, v^*)$

2° 若 $c \rightarrow \infty$, 则 $\forall v$, $x^* = \arg \min_x L_c(x, v)$

例. $\min \frac{1}{2} x_1^2 + \frac{1}{2} x_2^2$ 最优解 $x^* = (1, 0)$
s.t. $x_1 = 1$

拉格朗日函数: $L(x, v) = \frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 + v(x_1 - 1)$

KKT 条件: $\begin{cases} x_1 = 1 \\ x_1 + v = 0 \\ x_2 = 0 \end{cases} \Rightarrow \begin{cases} x_1^* = 1 \\ x_2^* = 0 \\ v^* = -1 \end{cases}$

增广拉格朗日函数: $L_c(x, v) = \frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 + v(x_1 - 1) + \frac{c}{2} (x_1 - 1)^2$

对于性质 1°: $\arg \min_x L_c(x^k, v^k) = \arg \min_x \frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 + v(x_1 - 1) + \frac{c}{2} (x_1 - 1)^2$
由 $x_1 - 1 + c(x_1 - 1) = 0$ 得 $x_1 = 1$, $x_2 = 0$, $v = -1$

对于性质 2°: $\arg \min_x L_c(x, v) = \arg \min_x \frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 + v(x_1 - 1) + \frac{c}{2} (x_1 - 1)^2$
 $x_1 + v + c(x_1 - 1) = 0 \Rightarrow x_1 = \frac{c-v}{c+1}$ 当 $c \rightarrow \infty$, $x_1 \rightarrow 1$

增广拉格朗日法: $x^{k+1} = \begin{pmatrix} \frac{c-v^k}{c+1} \\ 0 \end{pmatrix}, v^{k+1} = v^k + c \left(\frac{c-v^k}{c+1} - 1 \right) = v^k - \frac{c}{c+1} (v^k + 1)$

又 $v^* = -1$, 于是

$$v^{k+1} - v^k = v^k - \frac{c}{c+1} (v^k + 1) = (v^k + 1) \left(1 - \frac{c}{c+1} \right) = \frac{1}{c+1} (v^k - v^*)$$

交替方向乘子法

$$\min f(x) + g(x) \Rightarrow \min f(x) + g(z)$$

$$\text{s.t. } x = z$$

增广拉格朗日函数: $L_c(x, z, v) = f(x) + g(z) + \frac{c}{2} \|x - z\|_2^2$

交替方向乘子法:

- 1) $\{x^{k+1}, z^{k+1}\} = \arg \min_{x, z} (f(x) + g(z)) + \frac{c}{2} \|x - z\|_2^2$
- 2) $v^{k+1} = v^k + c(x^{k+1} - z^{k+1})$

坐标轮换法求解 1)

- (a) $x^{k+1, t+1} = \arg \min_x f(x) + \frac{c}{2} \|x - z^{k+1, t}\|_2^2$
- (b) $z^{k+1, t+1} = \arg \min_z g(z) + \frac{c}{2} \|z - x^{k+1, t}\|_2^2$

实际上可以重复 (a)(b)(2), 交替方向乘子法
Alternating Direction Method

2023.10.31

Part 1

拉格朗日乘子法 2

$$\text{例} 1. \min \sum_{i=1}^n f_i(x_i)$$

等价变换 $\min \sum_{i=1}^n f_i(x_i)$

$$\text{s.t. } x_i = z, i=1, \dots, n$$

$$L_C = \sum_{i=1}^n f_i(x_i) + \frac{c}{2} \sum_{i=1}^n \|x_i - z\|_2^2$$

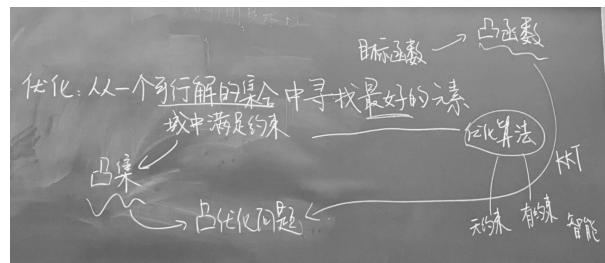
$$\Rightarrow \left\{ x_i^{k+1} \right\} = \arg \min_{\{x_i\}} \sum_{i=1}^n f_i(x_i) + \frac{c}{2} \sum_{i=1}^n \|x_i - z^k + \frac{v_i^k}{c}\|_2^2$$

$$\Leftrightarrow x_i^{k+1} = \arg \min_{x_i} f_i(x_i) + \frac{c}{2} \left\| x_i - z^k + \frac{v_i^k}{c} \right\|_2^2, \quad i=1, \dots, n$$

$$2) z^{k+1} = \arg \min_z \frac{c}{2} \sum_{i=1}^n \|z - x_i^{k+1} - \frac{v_i^k}{c}\|_2^2 \Leftrightarrow z^{k+1} = \frac{1}{n} \sum_{i=1}^n (x_i^{k+1} + \frac{v_i^k}{c})$$

$$3) v_i^{k+1} = v_i^k + c (x_i^{k+1} - z^{k+1}), \quad i=1, \dots, n$$

课程归纳



一、凸集

$\theta_1 x_1 + \dots + \theta_k x_k$ 为
 ① 仿射组合: $\theta_1 + \dots + \theta_k = 1$
 ② 凸组合: $\theta_1 + \dots + \theta_k = 1, \theta_i \geq 0$
 ③ 凸锥组合: $\theta_i \geq 0$

仿射集 / 凸集 / 凸锥, 若集合中任意元素的仿射 / 凸 / 凸锥组合仍在集中

仿射包 / 凸包 / 凸锥包

凸集例子

二、凸函数

四个定义:

- $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 凸 $\Leftrightarrow \text{dom}(f)$ 为凸且 $\forall x, y \in \text{dom}(f), \theta \in [0, 1]$,
 $\theta x + (1-\theta)y \in \text{dom}(f) \quad \text{且} \quad f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$
- $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 凸 $\Leftrightarrow \text{dom}(f)$ 为凸且 $\forall x \in \text{dom}(f), \forall t \in \mathbb{R}$,
 $g(t) = f(x+tv)$ 为凸, 其中 $\text{dom}(g) = \{t | x+tv \in \text{dom}(f)\}$
- (二阶条件)
 f 为仿射 $\Leftrightarrow \text{dom}(f)$ 为凸且 $\nabla^2 f(x)$ 半定

凸函数举例, 保凸运算

拟凸函数 α -下水平集

三、凸优化

$\min f_0(x)$ 凸

s.t. $f_i(x) \leq 0, i=1, \dots, m$ 凸

$h_i(x) = 0, i=1, \dots, p$ 仿射函数

线性规划

若 f_i, h_i 可微, 最优化条件:

可行解 $\Leftrightarrow \forall y \in X, L_f^*(y-x^*) \geq 0$ $\Leftrightarrow \lambda^* = d^*$

IV 对偶 (对于一般化问题)

$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$ $\Leftrightarrow x \in D, \lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$ \Leftrightarrow 原问题对偶问题

$g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu)$ \Leftrightarrow 对偶问题 $\max_{\lambda, \nu} g(\lambda, \nu)$ \Leftrightarrow 解释, 联立

$p^* = \inf_{x \in D} \sup_{\lambda \geq 0, \nu} L(x, \lambda, \nu) \rightarrow x^*, \lambda^*, \nu^*$ $(x^*, \lambda^*, \nu^*) = (x^*, \lambda^*, \nu^*)$ 满足

$d^* = \sup_{\lambda \geq 0, \nu} L(x^*, \lambda, \nu) \rightarrow x^*, \lambda^*, \nu^*$ $(x^*, \lambda^*, \nu^*) = (x^*, \lambda^*, \nu^*)$ 满足

鞍点定理

大KKT条件

$$\begin{cases} f_i(x^*) \leq 0, i=1, \dots, m \\ h_i(x^*) = 0, i=1, \dots, p \\ \lambda_i^* \geq 0, i=1, \dots, m \\ \lambda_i^* f_i(x^*) = 0, i=1, \dots, m \\ \lambda_i^* f_i(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) = 0 \end{cases}$$

五、优化算法

线搜索 (步长)

无约束 $\left\{ \begin{array}{l} \text{梯度下降法} \\ \text{最速下降法} \\ \text{坐标轮换法} \\ \text{牛顿法/拟牛顿法} \end{array} \right.$

有约束 $\left\{ \begin{array}{l} \text{牛顿/拟牛顿法} \\ \text{拉格朗日法} \\ \text{增广拉格朗日法} \\ \text{交替方向法} \end{array} \right.$

六、智能化
演化计算