1. 引言

优化/数学规划：从一个可行解的集合中，寻找出最优的元素。

minimize $\underset{\text{目标函数}}{f_0(x)}$ subject to $\underset{\text{不等式约束}}{f_i(x) \le b_i}$ $\quad i=1,\cdots,m$. $\qquad x=[x_1,\cdots,x_n]^T$

x^* optimal $\Longleftrightarrow$ $\forall z$, $z \in \{f_i(z) \le b_i,\ i=1,\cdots,m\}$ 可行域 feasible set $\quad f_0(z) \ge f_0(x^*)$.

e.g. 数据拟合问题 $y=ax^2+bx+c$ $\quad \min \sum_{k=1}^{n}[y_i-(ax_i^2+bx_i+c)]^2$ 先验已知 $a \ge 0$

线性二次调节器 LQR $\quad x_k = Ax_{k-1}+Bu_k \qquad \min_{\{u_k\}} J = \sum_{k=1}^{N}(x_k^T Q x_k + u_k^T R u_k)$

多用户能量控制问题 每个用户 u_i 耗费能量 p_i $\quad 0 \le p_i \le h_i$ 互相有干扰：α_{ij} 噪声：σ_i^2

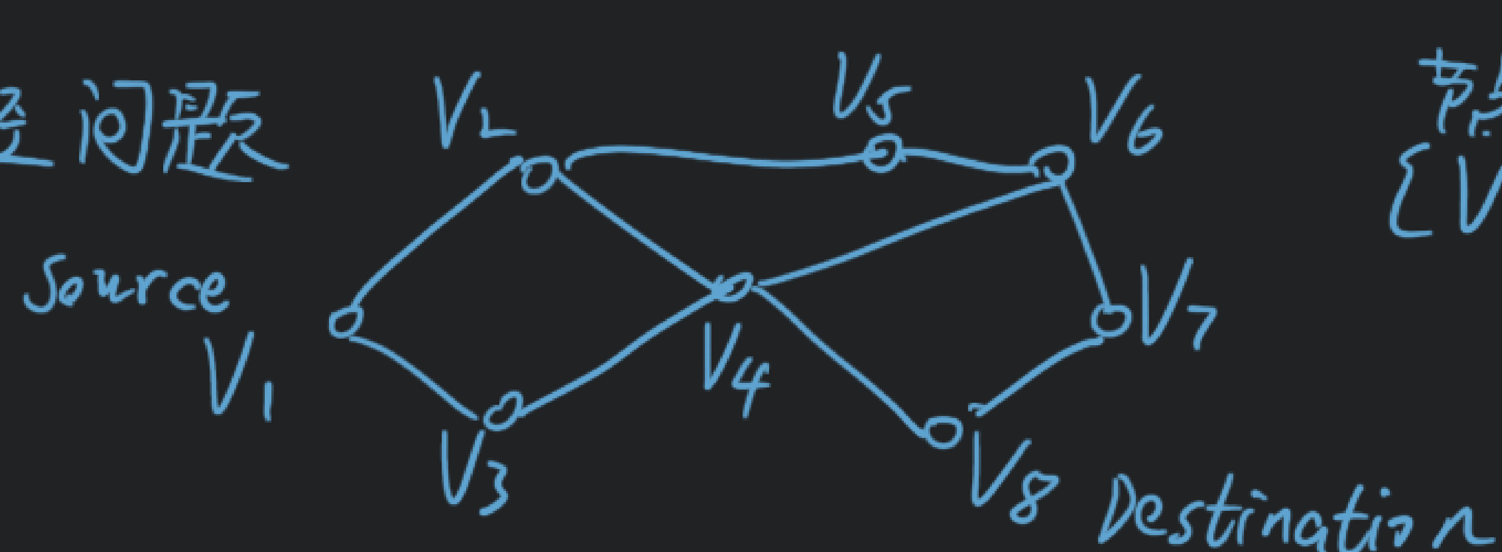
$$SINR_i = \frac{p_i}{\sigma_i^2 + \sum_{j\ne i} p_j \alpha_{ji}}$$ 流量 $f_i \sim \log(1+SINR_i)$ $\quad \max \sum_{i=1}^{n} f_i \quad s.t.\ 0 \le p_i \le b_i$.

图像处理 带噪声图像 $\Phi_0(x,y) \longrightarrow$ 恢复原图像 $\Phi(x,y)$ TV范数↓（分片光滑）

$$\|\Phi\|_{TV} = \sum_{x,y}\sqrt{(\Phi(x,y)-\Phi(x,y-1))^2+(\Phi(x,y)-\Phi(x-1,y))^2}$$

$$\min_{\Phi} \|\Phi\|_{TV} + \lambda\|\Phi-\Phi_0\|_F^2 \qquad (\text{TV-}L_2 \text{ 模型})$$

最短路径问题



节点 路径
$\{V,E\}$. 路径权重 w_{ij} 给定 求：源到目的地的最短路径.

$$\min \sum_{ij\in E} w_{ij}x_{ij} \quad s.t.\ x_{ij}=0 \text{ or } 1 \quad \sum_j^{\text{出}} x_{ij} - \sum_j^{\text{进}} x_{ji} = \begin{cases} 1 & i=S \\ -1 & i=D \\ 0 & \text{other.} \end{cases}$$

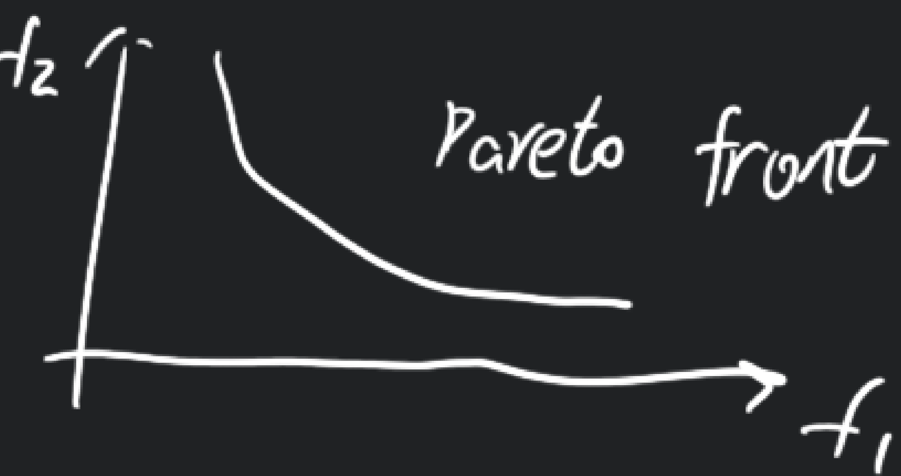
分类： 线性规划 $f_i(\alpha x+\beta y)=\alpha f_i(x)+\beta f_i(y)$ $\forall i\in\{0,\cdots,m\}$. 非线性规划

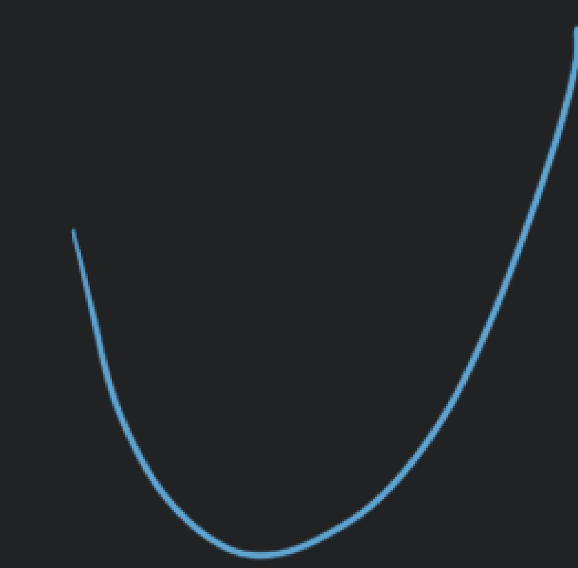
凸规划 $f_i(\alpha x+\beta y)\le \alpha f_i(x)+\beta f_i(y)$ $\forall i\in\{1,\cdots,m\}$.

光滑规划 f_i 在每一点可微. 凸规划一般更简单

连续/离散规划 针对可行域而言 连续规划一般更简单

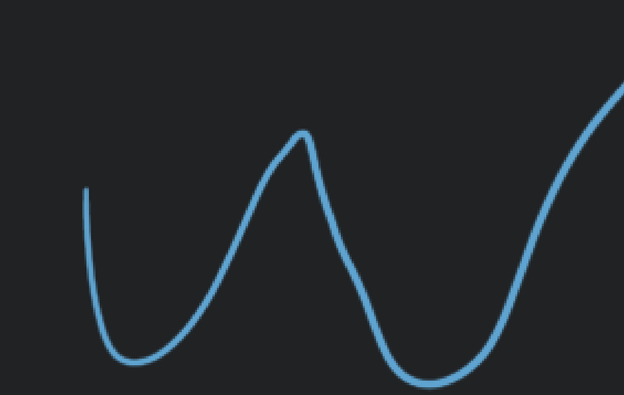
单目标/多目标 $\min \begin{matrix} f_1(x) \\ f_2(x) \end{matrix}$





凸函数. 无法找到不相邻的比较低的点.

线性规划 最优解只出现在顶点或边上.



非凸函数

主要内容 1) 凸集 凸函数

2) 凸优化

3) 若干算法.

不讨论目标函数的难易 (这些发展最少？)

2. 凸集

$x_1\neq x_2\in\mathbb{R}^n$ 直线 $\theta\in\mathbb{R}$ $y=\theta x_1+(1-\theta)x_2=x_2+\theta(x_1-x_2)$

线段 $\theta\in[0,1]$

仿射集 Affine Set 集合C，$\forall x_1,x_2\in C$，连接 x_1,x_2 两点的直线也在 C 内. ($\forall x_1,x_2\in C$, $\forall\theta\in\mathbb{R}$, $\theta x_1+(1-\theta)x_2\in C$)

设 $x_1,\cdots,x_k\in C$ $\theta_1,\cdots,\theta_k\in\mathbb{R}$, $\theta_1+\cdots+\theta_k=1$.

仿射组合：$\theta_1 x_1 + \cdots + \theta_k x_k$

仿射集（定义2） $\forall x_1, \cdots, x_k \in C$ 的仿射组合也在 C 内.（两定义等价）

与仿射集 C 相关的子空间 $\forall x_0 \in C$ $V = \{x - x_0 \mid x \in C\}$. V 也是仿射集 $\forall v_1, v_2 \in V$ $\forall \alpha, \beta \in \mathbb{R}$ $\Rightarrow \alpha v_1 + \beta v_2 \in V$.

不要求 $\alpha+\beta=1$，性质更好.

$\Leftarrow \alpha v_1 + \beta v_2 + x_0 \in C \Leftarrow \alpha(\overset{\in C}{v_1 + x_0}) + \beta(\overset{\in C}{v_1 + x_0}) + (1-\alpha-\beta)x_0 \in C$.

V 一定过原点. V 实际与 x_0 无关.

e.g. 线性方程组的解集是仿射集.

$C = \{x \mid Ax = b\}$. $A \in \mathbb{R}^{m \times n}$ $b \in \mathbb{R}^m$ $x \in \mathbb{R}^n$.

$\forall x_1, x_2 \in C$, $\forall \theta \in \mathbb{R}$. $A[\theta x_1 + (1-\theta)x_2] = b$. $\Rightarrow \theta x_1 + (1-\theta)x_2 \in C$.

$\forall x_0$, $V = \{x - x_0 \mid x \in C\}$. $= \{x - x_0 \mid Ax = b, Ax_0 = b\}$ 是 $Ax = 0$ 齐次方程的解集，A 的零空间.

任意集合 C 构造尽可能小的仿射集？

仿射包 $\operatorname{aff} C = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid \forall x_1, \cdots x_k \in C, \forall \theta_1, \cdots, \theta_k \in \mathbb{R}, \theta_1 + \cdots + \theta_k = 1\}$ C 的仿射组合构成集合

凸集 Convex Set 集合 C 是凸集 $\Leftrightarrow$ $\forall x_1, x_2 \in C$, $\forall \theta \in [0,1]$ $\theta x_1 + (1-\theta)x_2 \in C$ （任意两点线段在 C 内）

仿射集是凸集.

凸组合：$\forall x_1, \cdots, x_k$ $\forall \theta_1, \cdots, \theta_k \in [0,1]$ $\theta_1 + \cdots + \theta_k = 1$ $\theta_1 x_1 + \cdots + \theta_k x_k$ 称为 $x_1, \cdots, x_k$ 的凸组合

凸集第二定义：$\forall x_1, \cdots, x_k$ 的凸组合在 C 内

凸包 $\forall C \in \mathbb{R}^n$, $\operatorname{Conv} C$ = C 的凸组合构成的集合

锥 Cone　集合C是锥 $\Longleftrightarrow$ $\forall x\in C,\ \theta\geq 0 \Rightarrow \theta x\in C$　锥一定过原点.

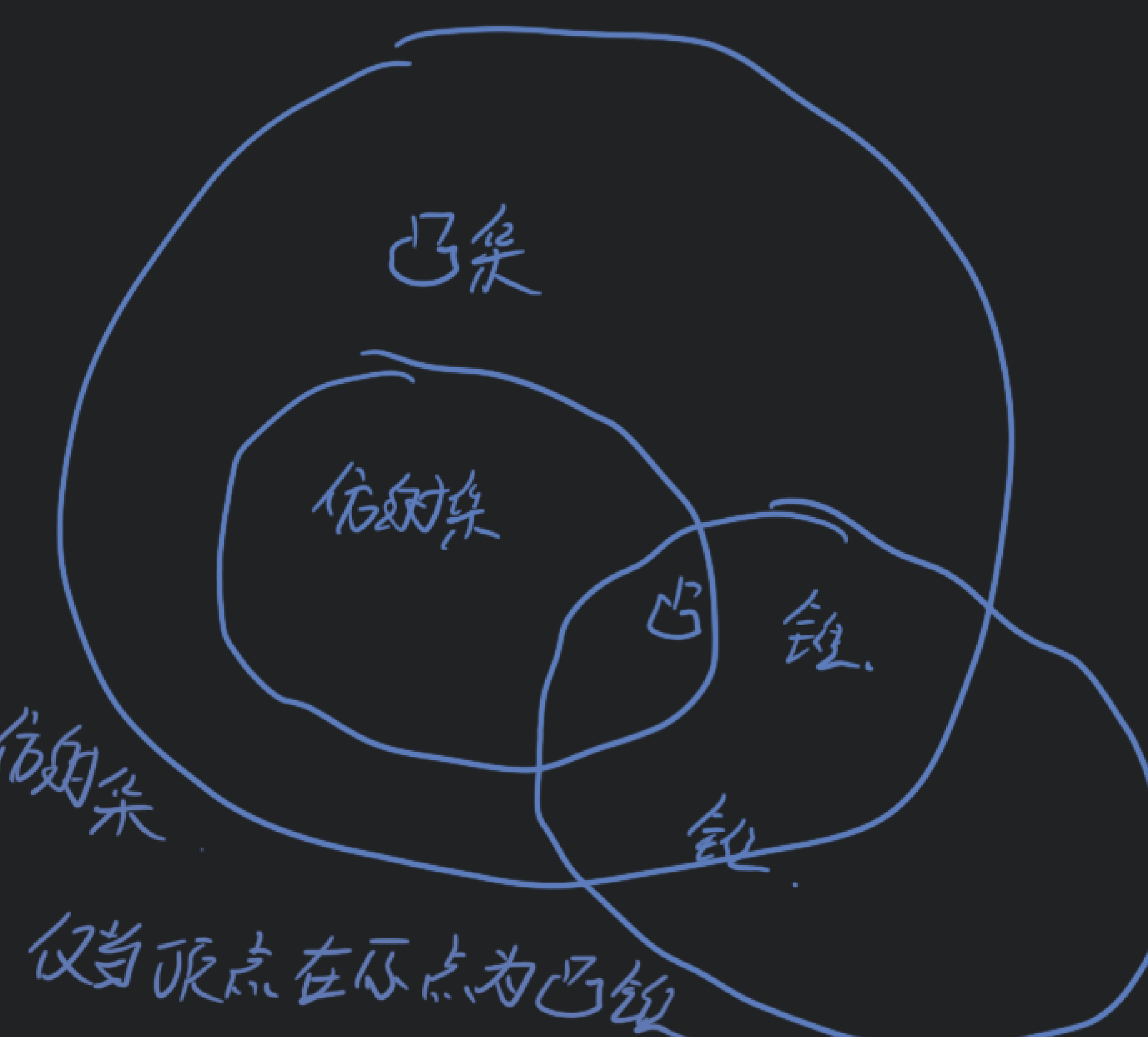
凸锥 Convex Cone $\Longleftrightarrow$ $\forall x_1,x_2\in C,\ \theta_1,\theta_2\geq 0 \Rightarrow \theta_1x_1+\theta_2x_2\in C$.　凸锥是凸集

凸锥组合　$\forall x_1,\cdots,x_k\in C$　$\forall\theta_1,\cdots,\theta_k\geq 0$　$\theta_1x_1+\cdots+\theta_kx_k$

凸锥包　C的凸锥组合构成的集合　(包含C的最小凸锥).

单点集合　$C=\{x\}$　一定是仿射集　一定是凸集　仅当$x=0$时是凸锥

空集, $\mathbb{R}^n$, $\mathbb{R}^n$的子空间　一定是仿射集, 凸集, 凸锥



仿射组合 → $\theta_1+\cdots+\theta_k=1$

凸组合

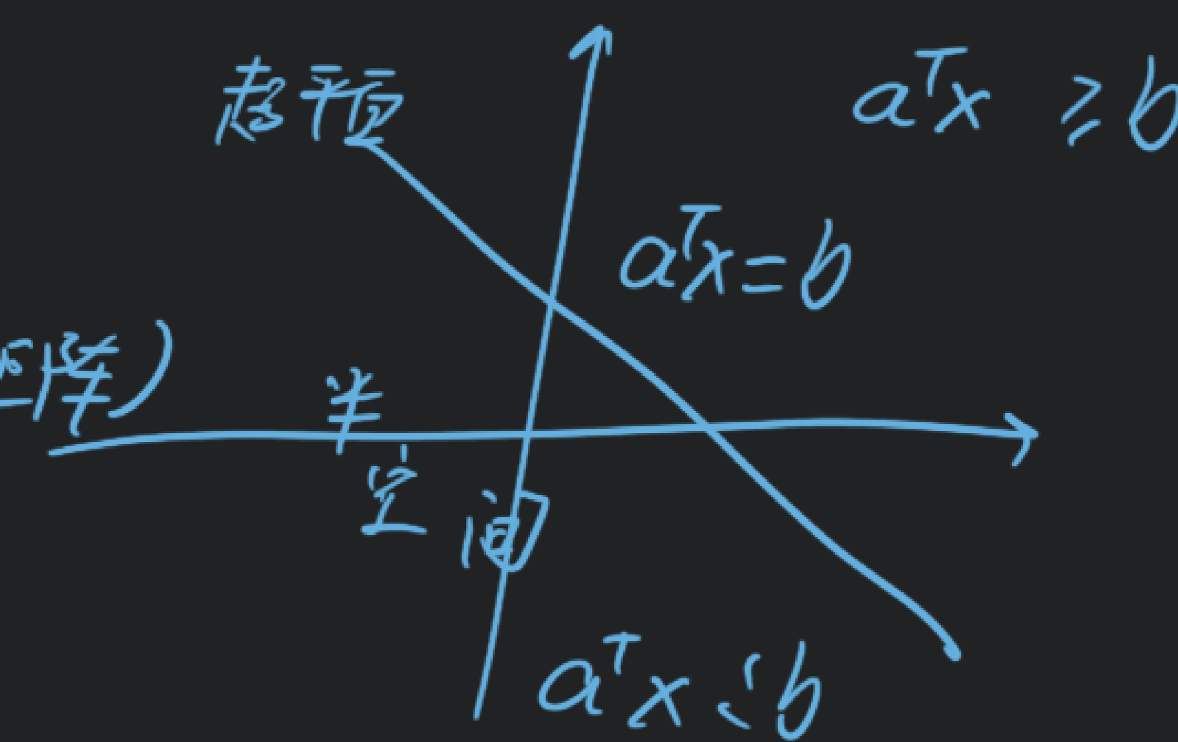
凸锥组合 → $\theta_1+\cdots+\theta_k\geq 0$.

任意线段是凸集, 仅当缩为一个点时为仿射集

任意射线是凸集, 仅当缩为一点为仿射集, 仅当顶点在原点为凸锥.

超平面 $\{x \mid a^Tx = b\}$ $x, a \in \mathbb{R}^n,\ b \in \mathbb{R}$. $a \neq 0$.

超平面分出两个半空间



椭球 $\mathcal{E}(x_c, P) = \{x \mid (x-x_c)^T P^{-1}(x-x_c) \leq 1\}$ 其中 $P \in S^n_{++}$ (对称正定矩阵)

奇异值 Singular Value $\sqrt{\mathrm{eig}(A^TA)}$ 椭球半轴长是奇异值

特征值，≥ 0.

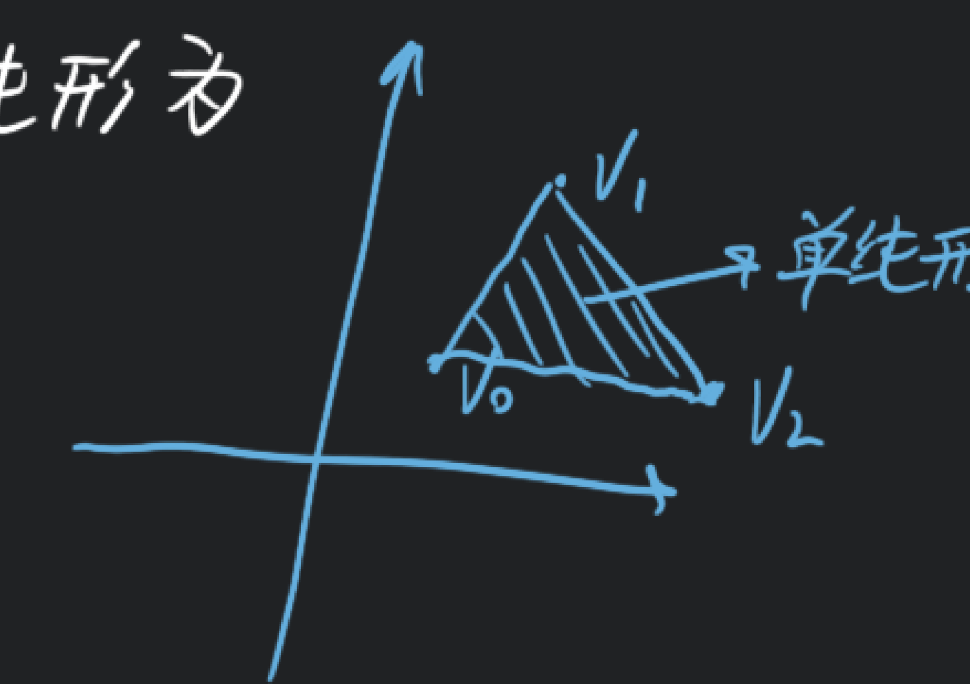
多面体 Polyhedron $P = \{x \mid a_j^T x \leq b_j \quad j=1,\cdots,m \quad a_j^T x = d_j \quad j=1,\cdots,p\}$. (一些半空间和超平面的交)

有界多面体. 多面体是凸集.

单纯形 Simplex $\forall V_0, \cdots, V_k \in \mathbb{R}^n$ $V_1 - V_0, \cdots, V_k - V_0$ 线性无关(要求 $k \leq n$)，则与上述点相关的单纯形为

$$C = \mathrm{Conv}\{V_0, \cdots, V_k\} = \{\theta_0 V_0 + \cdots + \theta_k V_k,\ \theta \succeq 0,\ 1^T\theta \succeq 1\}.$$

$\theta \succeq 0$: $\forall i \leq k,\ \theta_i \geq 0$.



二维空间的单纯形只能是线段或三角形

单纯形一定是多面体.

证：设 $x \in C \subseteq \mathbb{R}^n$ C为Simplex $\Leftrightarrow$ $x = \theta_0 V_0 + \cdots + \theta_k V_k$ $1^T\theta = 1,\ \theta \succeq 0$. $V_1 - V_0, \cdots, V_k - V_0$ 线性无关

定义 $[\theta_1, \cdots, \theta_k]^T = y$ $y \succeq 0$ $1^Ty = 1$ $[V_1 - V_0, \cdots, V_k - V_0] = B \subseteq \mathbb{R}^n$.

$x \in C \Leftrightarrow x = V_0 + By$. rank $B = k \leq n$ (列满秩) $\exists$ 非奇异矩阵 $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \in \mathbb{R}^{n\times n}$

$AB = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} B = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$

$$Ax = AV_0 + ABy \Leftrightarrow \begin{cases} A_1 x = A_1 V_0 + y \\ A_2 x = A_2 V_0 \end{cases} \Leftrightarrow \begin{cases} A_1 x \succeq A_1 V_0 \\ 1^T A_1 x \leq 1^T A_1 V_0 + 1 \\ A_2 x = A_2 V_0 \end{cases}$$

#

对称矩阵集合 $S^n=\{X\in\mathbb{R}^{n\times n}\mid X=X^T\}$.

对称半正定矩阵集合 $S_+^n=\{X\in S^n\mid X\succeq 0\}$. 所有特征值 ≥ 0

对称正定矩阵集合 $S_{++}^n=\{X\in S^n\mid X\succ 0\}$.

S_+^n 是凸锥（因此也是凸集）. S^n 是凸锥 S_{++}^n 不是凸锥

证: $\forall\theta_1,\theta_2\geq 0$ $\forall A,B\in S_+^n$, $\theta_1A+\theta_2B$ 对称, $\forall x\in\mathbb{R}^n$, $x^TAx\geq 0$ $x^TBx\geq 0$ $\Rightarrow x^T(\theta_1A+\theta_2B)x\geq 0$.

$\Rightarrow \theta_1A+\theta_2B\in S_+^n$.

交集 S_1,S_2 为凸集 $\Rightarrow S_1\cap S_2$ 为凸集.

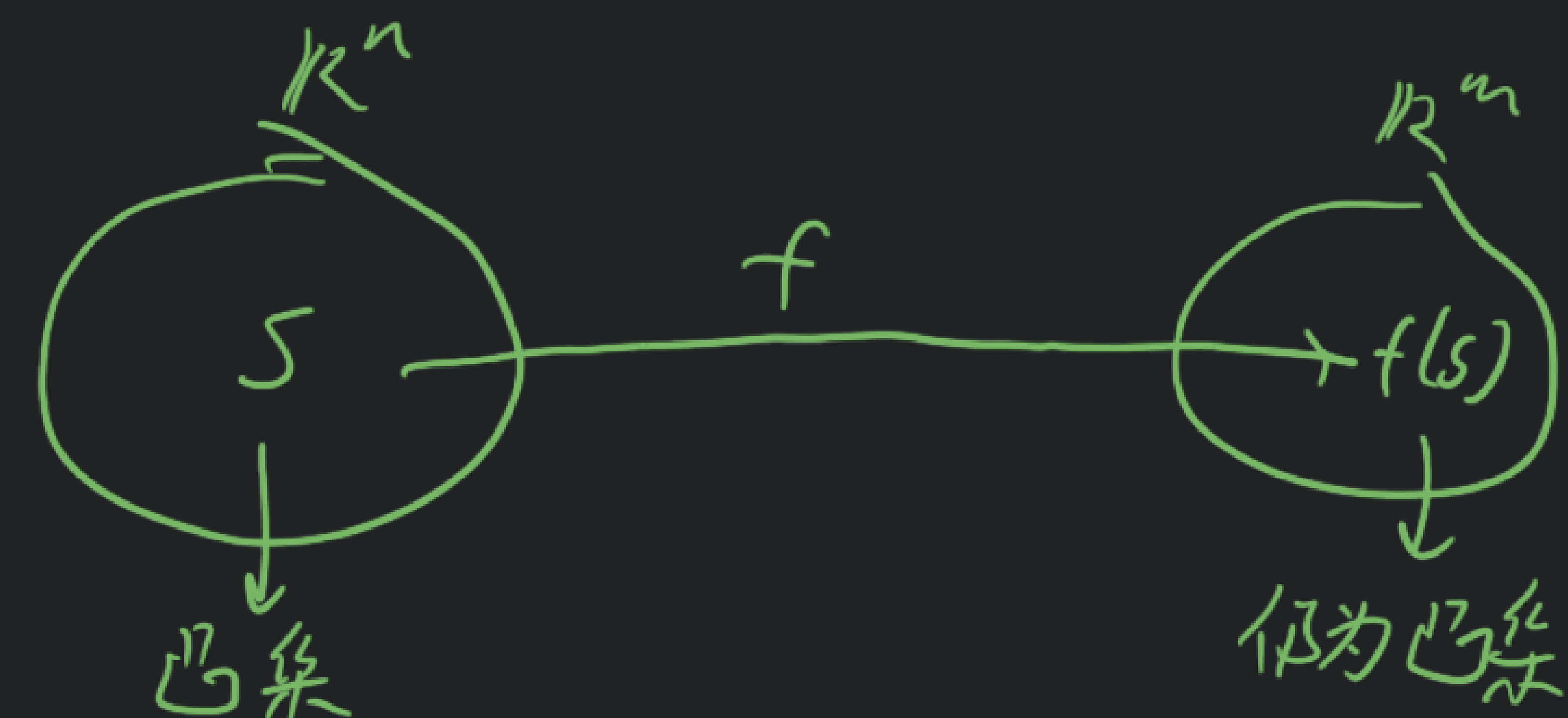
若 $\forall a\in A$ S_a 为凸集 $\Rightarrow \bigcap\limits_{a\in A} S_a$ 为凸集.

但并集不一定.

仿射函数 $f:\mathbb{R}^n\to\mathbb{R}^m$ 是仿射的 $\Leftrightarrow f(x)=Ax+B$ $A\in R^{m\times n}$ $B\in\mathbb{R}^m$.

仿射函数是保凸运算

逆映射也保凸



e.g. $\alpha S=\{\alpha x\mid x\in S\}$. 缩放和移位保持凸性

$S+a=\{x+a\mid x\in S\}$.

e.g. 两个凸集的和是凸集 $S_1+S_2=\{x+y \mid x\in S_1, y\in S_2\}$.

$S_1\times S_2 = \{(x,y) \mid x\in S_1, y\in S_2\}$ 是凸集 $\Rightarrow$ S_1+S_2 也是凸集.

e.g. 线性矩阵不等式 LMI $B, A_i\in S^n$, $x\in\mathbb{R}^n$. $A(x)=x_1A_1+\cdots+x_nA_n\preceq B$ (即 $(A(x)-B)$ 是半正定矩阵).

$\{x \mid A(x)\preceq B\}$ 为凸集.

prove: 定义仿射变换 $f(x)\triangleq B-A(x)$: $\mathbb{R}^n\to S^n$ S_+^n 为凸集. $\Rightarrow f^{-1}(S_+^n)=\{x \mid B-A(x)\succeq 0\}$ 也为凸集.

e.g. 椭球是球的仿射映射.

$\mathcal{E}=\{x \mid (x-x_c)^T P^{-1}(x-x_c)\le 1\}$ $P\in S_{++}^n$ $\{u \mid \|u\|^2\le 1\}$. $f(u)=P^{1/2}u+x_c$.

$\{f(u) \mid \|u\|^2\le 1\}=\{P^{1/2}u+x_c \mid \|u\|^2\le 1\}$. $x\triangleq P^{1/2}u+x_c$ 上式 $=\{x \mid \|P^{-1/2}(x-x_c)\|\le 1\}=\mathcal{E}$.

透视函数 $P:\mathbb{R}^n\times\mathbb{R}_{++}\longrightarrow\mathbb{R}^n$ $P(z,t)=\frac{z}{t}$ $z\in\mathbb{R}^n$ $t>0$.

透视函数是保凸运算

e.g. $\mathbb{R}^{n+1}$ 中线段 $x=(\tilde{x}, x_{n+1})$ $y=(\tilde{y}, y_{n+1})$ $\tilde{x},\tilde{y}\in\mathbb{R}^n$ $x_{n+1}, y_{n+1}>0$.

$0\le\theta\le 1$, 线段 $\theta x+(1-\theta)y$. $\xrightarrow{P}$ 仍为线段.

$$P(\theta x+(1-\theta)y)=\frac{\theta\tilde{x}+(1-\theta)\tilde{y}}{\theta x_{n+1}+(1-\theta)y_{n+1}}=\underbrace{\frac{\theta x_{n+1}}{\theta x_{n+1}+(1-\theta)y_{n+1}}}_{\triangleq\mu}\frac{\tilde{x}}{x_{n+1}}+\frac{(1-\theta)y_{n+1}}{\theta x_{n+1}+(1-\theta)y_{n+1}}\frac{\tilde{y}}{y_{n+1}}=\mu P(x)+(1-\mu)P(y)\quad 0\le\mu\le 1.$$

$\mu(\theta)$ 连续变化.

反透视映射 $P^{-1}(C)=\{(x,t)\in\mathbb{R}^{n+1} \mid \frac{x}{t}\in C, t>0\}$. 任意凸集的反透视映射也是凸集.

prove: $(x,t)\in P^{-1}(C)$ $(y,s)\in P^{-1}(C)$ $0\le\theta\le 1$ C是凸集，$\frac{x}{t}\in C$，$\frac{y}{s}\in C$.

$$\frac{\theta x+(1-\theta)y}{\theta t+(1-\theta)s} = \frac{\theta t}{\underbrace{\theta t+(1-\theta)s}_{\triangleq\mu}}\frac{x}{t} + \frac{(1-\theta)s}{\theta t+(1-\theta)s}\frac{y}{s} = \mu\frac{x}{t}+(1-\mu)\frac{y}{s}\in C$$

$$\Rightarrow \theta(x,t)+(1-\theta)(y,s)\in P^{-1}(C).$$

线性分数函数 $g:\mathbb{R}^n\longrightarrow\mathbb{R}^{m+1}$ 为仿射映射 $g(x)=\begin{bmatrix}A\\c^T\end{bmatrix}x+\begin{bmatrix}b\\d\end{bmatrix}$ $A\in\mathbb{R}^{m\times n}$ $b\in\mathbb{R}^m$ $c\in\mathbb{R}^n$ $d\in\mathbb{R}$

$P:\mathbb{R}^{m+1}\longrightarrow\mathbb{R}^m$ 为透视映射 线性分数函数 $f:\mathbb{R}^n\to\mathbb{R}^m\triangleq P\circ g$.

$$f(x)=\frac{Ax+b}{c^Tx+d} \qquad \underset{\text{定义域}}{\mathrm{dom}\,f} = \{x\mid c^Tx+d>0\}.$$

线性分数函数仍保凸

e.g. 两个随机变量的联合概率 $\longrightarrow$ 条件概率.

$u=\{1,\cdots,n\}$ $v=\{1,\cdots,m\}$. $P_{ij}=P(u=i,v=j)$ $f(ij)=P(u=i|v=j)$.

$$f_{ij}=\frac{P_{ij}}{\sum_{k=1}^{m}P(v=k)}$$

3. 凸函数

凸函数定义1　$f: \mathbb{R}^n \to \mathbb{R}$ 为凸函数 $\Longleftrightarrow$ $\operatorname{dom} f$ 为凸集，$\forall x, y \in \operatorname{dom} f$，$\forall \theta \in [0,1]$

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

严格凸：$\operatorname{dom} f$ 为凸集，$\forall x, y \in \operatorname{dom} f$，$x \neq y$，$\forall \theta \in (0,1)$　$f(\theta x + (1-\theta) y) < \theta f(x) + (1-\theta) f(y)$.

$-f$ 凸 $\Rightarrow$ f 凹　　$-f$ 严格凸 $\Rightarrow$ f 严格凹

凸函数定义2　$f: \mathbb{R}^n \to \mathbb{R}$ 为凸函数 $\Longleftrightarrow$ $\forall x \in \operatorname{dom} f$，$\forall v \in \mathbb{R}^n$，$g(t) \triangleq f(x+tv)$ 为凸函数，

其中 $\operatorname{dom} g = \{t \mid x+vt \in \operatorname{dom} f\}$.

（函数是凸的 $\Longleftrightarrow$ 在与其定义域相交的任何直线上是凸的）　高维凸 $\to$ 一维凸.

凸函数扩展

$f: C \to \mathbb{R}$ 为凸函数，$\operatorname{dom} f$（定义域）$= C \subseteq \mathbb{R}^n$　$\tilde{f} = \begin{cases} f(x) & x \in \operatorname{dom} f \\ +\infty & x \notin \operatorname{dom} f \end{cases}$　$\tilde{f}: \mathbb{R}^n \to \mathbb{R}$，$\operatorname{dom} \tilde{f} = \mathbb{R}^n$ 也是凸函数.

示性函数是凸函数

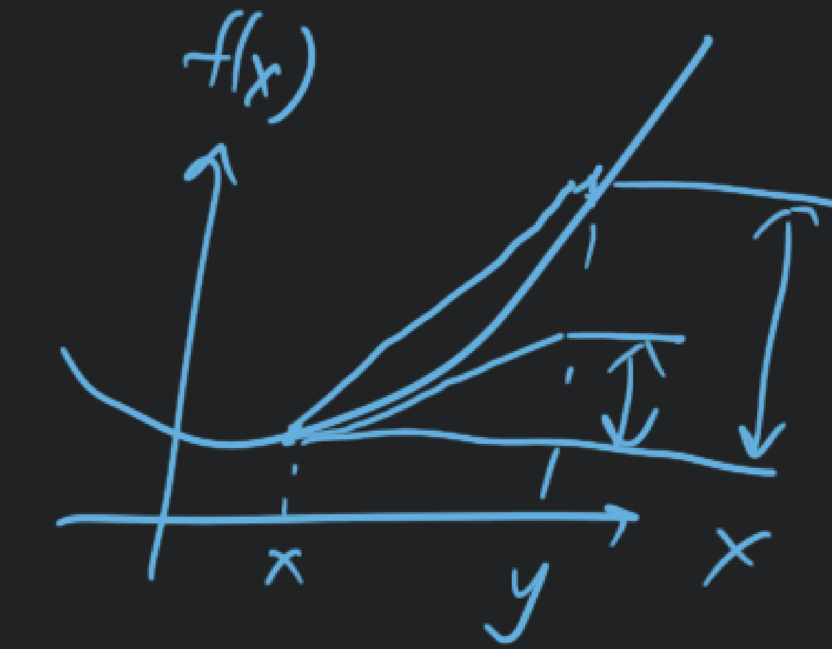
凸集 $C \subseteq \mathbb{R}^n$　$f_C(x) = \begin{cases} \text{无定义} & x \notin C \\ 0 & x \in C \end{cases}$ 是凸函数　$I_C(x) = \tilde{f}_C(x) = \begin{cases} \infty & x \notin C \\ 0 & x \in C \end{cases}$ 也是凸函数.

但 $J_C(x) = \begin{cases} 1 & x \notin C \\ 0 & x \in C \end{cases}$ 既不凸也不凹

一阶条件（凸函数定义3）

设 $f:\mathbb{R}^n\to\mathbb{R}$ 可微，即 ∇f 在 $\mathrm{dom}\, f$ 上均存在，（$\mathrm{dom}\, f$ 为开集）则 f 为凸函数 $\Longleftrightarrow$

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



若f(x)满足上述条件，$\exists x_0\ \nabla f(x_0)=0 \Rightarrow \forall y\in\mathrm{dom}\, f\quad f(y)\geq f(x_0) \Rightarrow \{x\in\mathrm{dom}\, f \mid \nabla f(x)=0\}$ 即最优解集.

证明一阶条件：　一维情况　$f:\mathbb{R}\to\mathbb{R}$ 为凸 $\Longleftrightarrow$ $\mathrm{dom}\, f$ 为凸，$f(y)\geq f(x)+f'(x)(y-x)$.　（假设 f 可导）

"⇒"　f 凸 $\Rightarrow \mathrm{dom}\, f$ 凸，$\forall t\in(0,1]\quad x+t(y-x)\in\mathrm{dom}\, f\quad f(x+t(y-x))\leq(1-t)f(x)+t f(y)$

$f(y)\geq f(x)+[f(x+t(y-x))-f(x)]/t$　　取 $t\to0^+$，$f(y)\geq f(x)+f'(x)(y-x)$.

"⇐"　设 $\forall x\neq y$，$x,y\in\mathrm{dom}\, f$　取 $\theta\in[0,1]$，$z\triangleq\theta x+(1-\theta)y\in\mathrm{dom}\, f$

$$\left\{\begin{array}{l} f(x)\geq f(z)+f'(z)(z-x)\\ f(y)\geq f(z)+f'(z)(y-z)\end{array}\right. \Rightarrow \theta f(x)+(1-\theta)f(y)\geq f(z)+f'(z)[\underbrace{\theta(z-x)+(1-\theta)(y-z)}_{=0}]$$

$$\Downarrow$$

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

高维情况

"⇒"　$\forall x,y\in\mathrm{dom}\, f$　　$g(t)\triangleq f(ty+(1-t)x)=f(x+t(y-x))$　　$g'(t)=\nabla f(ty+(1-t)x)^T(y-x)$.

按定义，$g(t)$ 为凸函数. $\Rightarrow g(t_1)\geq g(t_2)+g'(t_2)(t_1-t_2)\Rightarrow g(1)\geq g(0)+g'(0)$

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

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

$$f(ty+(1-t)x) \geq f(\tilde{t}y+(1-\tilde{t})x) + \nabla f(\tilde{t}y+(1-\tilde{t})x)^T \underbrace{[ty+(1-t)x-\tilde{t}y-(1-\tilde{t})x]}_{=(y-x)(t-\tilde{t})}.$$

$g(t) \triangleq f(ty+(1-t)x) \Rightarrow g(t)\geq g(\tilde{t})+g'(\tilde{t})(t-\tilde{t}) \xRightarrow{\text{定义2}}$ $f(x)$ 为凸函数.

二阶条件(定义4)

若 $f:\mathbb{R}^n\to\mathbb{R}$ 二阶可导，则 f 为凸函数 $\Longleftrightarrow$ $\mathrm{dom}f$ 为凸集，$\nabla^2 f(x)\succeq 0$ $\forall x\in\mathrm{dom}f$ (Hessian 矩阵半正定)

$\nabla^2 f(x)\succ 0 \Rightarrow$ 严格凸，但反过来不一定对. e.g. $f(x)=x^4$ $f''(x)=12x^2$ $f''(0)=0$. (但对二次函数成立).

e.g. 二次函数 $f:\mathbb{R}^n\to\mathbb{R}$ $\mathrm{dom}f=\mathbb{R}^n$ $f(x)=\frac{1}{2}x^TPx+q^Tx+r$ $P\in S^n$ $q\in\mathbb{R}^n$ $r\in\mathbb{R}$.

$\nabla^2 f(x)=P$ 半正定 $\Rightarrow$ 凸 半负定 $\Rightarrow$ 凹 正定 $\Rightarrow$ 严格凸 负定 $\Rightarrow$ 严格凹.

e.g. 仿射函数 $f(x)=Ax+b$ $\nabla^2 f=0$ 凸函数.

e.g. 负熵 $f(x)=x\log x$ 凸函数 (熵是凹的)

e.g. 范数 $\mathbb{R}^n$ 空间范数 $p:\mathbb{R}^n\to\mathbb{R}$ 满足 ① $p(ax)=|a|p(x)$ ② $p(x+y)\leq p(x)+p(y)$ ③ $p(x)=0\Leftrightarrow x=0$.

$\forall x,y\in\mathbb{R}^n$ $\forall 0\leq\theta\leq 1$ $p(\theta x+(1-\theta)y)\leq\theta p(x)+(1-\theta)p(y)$ $\Rightarrow$ $p(x)$ 是凸的.

e.g. 零范数 $\|x\|_0$ = x 中非零元素数目 零范数不是范数，也不是凸函数. $x\in\mathbb{R}$. $\|x\|_0=\begin{cases}0 & x=0\\ 1 & x\neq 0\end{cases}$.

e.g. 极大值函数 $f(x)=\max\{x_1,\cdots,x_n\}$ $x\in\mathbb{R}^n$.

$\forall x,y\in\mathbb{R}^n, \forall\theta\in[0,1]$ $f(\theta x+(1-\theta)y) = \max\{\theta x_i+(1-\theta)y_i, i=1,\cdots,n\} \le \theta\max\{x_i\} + (1-\theta)\max\{y_i\}$
$\Rightarrow$ 极大值函数是凸函数. $= \theta f(x)+(1-\theta)f(y)$.

解析逼近 log-sum-up $f(x) = \log(e^{x_1}+\cdots+e^{x_n})$ $x\in\mathbb{R}^n$. $\max\{x_1,\cdots x_n\} \le f(x) \le \max\{f_1+\cdots+f_n\}+\log n$.

$\frac{\partial f}{\partial x_i} = \frac{e^{x_i}}{\sum_l e^{x_l}}$ $\frac{\partial f}{\partial x_i\partial x_j} = \frac{\delta_{ij}e^{x_i}}{\sum_l e^{x_l}} - \frac{e^{x_i+x_j}}{(\sum_l e^{x_l})^2}$ $z \triangleq [e^{x_1},\cdots,e^{x_n}]^T \Rightarrow H = \frac{1}{(1^Tz)^2}\underbrace{[(1^Tz)\operatorname{diag}\{z\} - zz^T]}_{\triangleq K}$

$\forall v\in\mathbb{R}^n, v^TKv = (\sum_i z_i)(\sum_i v_i^2 z_i) - (\sum_i z_iv_i)^2$ $\underline{\underline{a_i = v_i\sqrt{z_i}, b_i=\sqrt{z_i}}}$ $|b|^2|a|^2 - |\vec{a}\cdot\vec{b}|^2 \ge 0. \Rightarrow f$ 也是凸函数

e.g. $f(x) = (x_1\cdots x_n)^{\frac{1}{n}}$ $x\in\mathbb{R}^n_{++}$ 几何平均是凹函数

e.g. 行列式对数 $f(x) = \log\det(x)$ $\operatorname{dom} f = S^n_{++}$

$n>1$ 时, $\forall z\in S^n_{++}$ $\forall t\in\mathbb{R}$, $\forall v\in\mathbb{R}^{n\times n}$ 且 $z+tv\in S^n_{++} \Rightarrow v\in S^n$. $g(t)\triangleq f(z+tv) = \log\det\{z^{1/2}(I+tz^{-1/2}vz^{-1/2})z^{1/2}\}$.

$g(t) = f(z)+\log\det\{I+tz^{-1/2}vz^{-1/2}\}$. λ_i: $z^{-1/2}vz^{1/2}$ 的特征值 $z^{-1/2}vz^{1/2} = Q\Lambda Q^T$.

$\det\{I+tz^{-1/2}vz^{-1/2}\} = \det\{QQ^T+tQ\Lambda Q^T\} = \det\{I+t\Lambda\} = \prod_{i=1}^n(1+t\lambda_i)$ $g(t) = f(z)+\sum_{i=1}^n\log(1+t\lambda_i)$

$g'(t) = \sum_i\frac{\lambda_i}{1+t\lambda_i}$ $g''(t) = \sum_i\frac{-\lambda_i^2}{(1+t\lambda_i)^2}\le 0$. $\Rightarrow g(t)$ 是凹函数 $\Rightarrow f(x)$ 是凹函数.

保持函数凸性的运算

① 非负加权和: $f_1\cdots f_n$ 为凸, $w_i\ge 0(\forall i)$ $\Rightarrow$ 则 $f=\sum_{i=1}^n w_if_i$ 为凸

连续拓展: 若 $f(x,y)$, 对 $\forall y\in A$, $f(x,y)$ 为凸, 设 $w(y)\ge 0$ $(\forall y\in A)$ $g(x) = \int_{y\in A} w(y)f(x,y)dy$ 为凸.

② 仿射映射 $f: \mathbb{R}^n \to \mathbb{R}$ 为凸 $A \in \mathbb{R}^{n\times m}$ $b \in \mathbb{R}^n$ $g(x) \triangleq f(Ax+b)$ $\text{dom}\, g = \{x \mid Ax+b \in \text{dom} f\}$ $\Rightarrow g(x)$ 也为凸

prove: 令 $x, y \in \text{dom}\, g$, $0 \le \theta \le 1$ $g(\theta x+(1-\theta)y) = f(\theta Ax + (1-\theta)Ay+b) = f(\theta(Ax+b) + (1-\theta)(Ay+b))$

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

③ 两个凸函数的极大值函数 f_1, f_2 为凸 $f(x) \triangleq \max\{f_1(x), f_2(x)\}$ $\text{dom} f = \text{dom} f_1 \cap \text{dom} f_2$. $\Rightarrow f(x)$ 也为凸.

prove: $\forall x, y \in \text{dom} f$, $0 \le \theta \le 1$. $f(\theta x+(1-\theta)y) = \max\{f_1(\theta x+(1-\theta)y), f_2(\theta x+(1-\theta)y)\}$.

$\le \max\{\theta f_1(x) + (1-\theta) f_1(y), \theta f_2(x) + (1-\theta) f_2(y)\}$.

$\le \theta \max\{f_1(x), f_2(x)\} + (1-\theta)\{f_1(y), f_2(y)\} = \theta f(x) + (1-\theta) f(y)$ #.

e.g. 向量中 r 个最大元素的和 $x \in \mathbb{R}^n$ $x_{[i]} \triangleq x$ 中第 i 大元素 $f(x) = \sum_{i=1}^{r} x_{[i]}$

$f(x) = \max\{x_{i_1} + \cdots + x_{i_r} \mid 1 \le i_1 \le \cdots \le i_r \le n\}$.

连续拓展: $f(x,y)$ $\forall y \in A$ 为凸函数 $g(x) \triangleq \overset{\text{上确界}}{\sup_{y\in A}} f(x,y)$ 为凸函数.

e.g. 实对称矩阵的最大特征值

$f(X) = \lambda_{max}(X)$ $\text{dom} f = S^m$ $Xy = \lambda y \Rightarrow y^T X y = \lambda |y|^2 \Rightarrow \lambda = \frac{y^T X y}{|y|^2}$ 约束 $|y| = 1$. $\Rightarrow \lambda = y^T X y$

$\lambda_{max}(X) = \sup\{y^T X y \mid |y| = 1\}$.

④ 函数复合. $g, h: \mathbb{R} \to \mathbb{R}$, $f = h \circ g$, $\text{dom} f = \text{dom}\, g = \text{dom}\, h = \mathbb{R}$, f, g, h 均二阶可微. $f(x) = h(g(x))$.

$f''(x) = h''(g(x))\, g'(x)^2 + h'(g(x))\, g''(x) \ge 0$

h 为凸，增 $(h' \geq 0)$，g 为凸 $\Rightarrow$ $f = h \circ g$ 为凸.

推广：$\begin{cases} \text{高维} \quad h: \mathbb{R} \to \mathbb{R} \quad g: \mathbb{R}^n \to \mathbb{R} \quad f = h \circ g: \mathbb{R}^n \to \mathbb{R} \\ \mathrm{dom}\, g \neq \mathbb{R}^n \quad \mathrm{dom}\, h \neq \mathbb{R} \\ h, g \text{ 均不二阶可微} \end{cases}$

$\Rightarrow$ h 凸，$\tilde{h}$ 增，g 凸 $\Rightarrow$ f 凸

h 凹，$\tilde{h}$ 增，g 凹 $\Rightarrow$ f 凹

$\xrightarrow{\text{取负}}$ h 凹，$\tilde{h}$ 减，g 凸 $\Rightarrow$ f 凹

h 凸，$\tilde{h}$ 减，g 凹 $\Rightarrow$ f 凸

证明：$\forall x, y \in \mathrm{dom} f \quad 0 \leq \theta \leq 1$. g 凸 $\Rightarrow$ $\mathrm{dom}\, g$ 凸. $\theta x + (1-\theta) y \in \mathrm{dom}\, g \quad \theta g(x) + (1-\theta) g(y) \geq g(\theta x + (1-\theta) y)$.

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

证明 $g(\theta x + (1-\theta) y) \in \mathrm{dom}\, h$ 反证：设 $g(\theta x + (1-\theta) y) \notin \mathrm{dom}\, h$ $\tilde{h}$ 不降 $+\infty = \tilde{h}(g(\theta x + (1-\theta) y)) \leq \tilde{h}(\theta g(x) + (1-\theta) g(y))$

$\Rightarrow f(\theta x + (1-\theta) y) = h(g(\theta x + (1-\theta) y)) \leq h(\theta g(x) + (1-\theta) g(y)) \leq \theta f(x) + (1-\theta) f(y)$ $\Rightarrow f$ 为凸函数

e.g. ① 若 g 为凸 $\Rightarrow \exp\{g(x)\}$ 为凸 ② 若 g 为凹，$g > 0$ $\log\{g(x)\}$ 为凹 ③ 若 g 为凹，$g > 0$，$\frac{1}{g(x)}$ 为凸

④ 若 g 为凸，$g \geq 0$，$p \geq 1$ $[g(x)]^p$ 为凸函数

⑤ 函数的透视 $\quad f: \mathbb{R}^n \to \mathbb{R} \quad g: \mathbb{R}^n \times R_{++} \to \mathbb{R} \quad g(x,t) = t f(\frac{x}{t}) \quad \mathrm{dom}\, g = \{(x,t) \mid t > 0, \frac{x}{t} \in \mathrm{dom} f\}$.

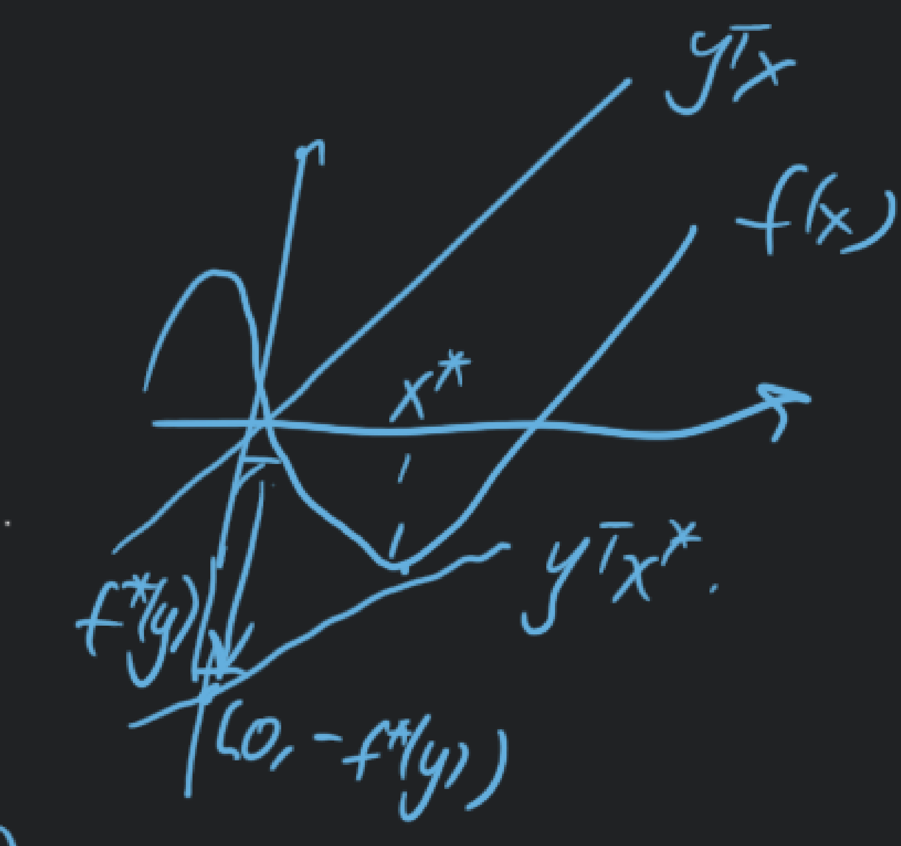
若 f 为凸 $\Rightarrow g(x,t)$ 也为凸. $\quad f$ 凹 $\Rightarrow g(x,t)$ 为凹

e.g. 欧几里德范数平方 $\quad f(x) = x^T x \quad \mathrm{dom} f = \mathbb{R}^n \quad f$ 的透视 $g(x,t) = t(\frac{x}{t})^T(\frac{x}{t}) = \frac{x^T x}{t}$ 为凸.

e.g. 负对数 $f(x) = -\log x \quad \mathrm{dom} f = R_{++}$ 为凸 $\quad g(x,t) = t(-\log \frac{x}{t}) = t \log \frac{t}{x} \quad \mathrm{dom}\, g = R_{++}^2 \ (x > 0, t > 0)$ 也为凸

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

$D_{KL}(u,v) \triangleq \sum_{i=1}^{n}\left(u_i \log\frac{u_i}{v_i} - u_i + v_i\right)$ 为凸，K-L散度.



⑥函数的共轭 (Conjugate)　$f:\mathbb{R}^n\to\mathbb{R} \Rightarrow f^*:\mathbb{R}^n\to\mathbb{R}$，$f^*(y) = \overset{\text{上确界}}{\sup\limits_{x\in dom f}} (y^Tx - f(x))$.

f(x)可微 ⇒ $f^*(y)$ 对应的x必是 $y=\nabla f(x)$ 的点、$\nabla_x[y^Tx-f(x)] = y-\nabla f = 0$. $\Rightarrow y=\nabla f(x)$.

函数的共轭一定为凸.（$y^Tx-f(x)$ 关于y是凸函数，据性质③得证）⇒ 函数共轭的共轭不一定等于其自身

若f凸，且f为闭函数 ⇒ $f^{**}=f$

e.g. $f(x)=ax+b$　$dom f=\mathbb{R}$　　$f^*(y)=\sup\limits_{x\in\mathbb{R}}(yx-(ax+b)) = \begin{cases} -b, & y=a \\ +\infty, & y\neq a \end{cases}$ 为凸

e.g. $f(x) = -\log x$　$dom f=\mathbb{R}_{++}$　　$f^*(y) = \sup\limits_{x>0}(yx+\log x) = \begin{cases} -1-\log(-y) & y<0 \\ +\infty & y\geq 0 \end{cases}$

e.g. $f(x)=\frac{1}{2}x^TQx$　$Q\in S_{++}^n$　$dom f=\mathbb{R}^n$　　$f^*(y) = \sup\limits_{x}(y^Tx-\frac{1}{2}x^TQx)$　　$\nabla_x \Rightarrow y-Qx=0$. $x=Q^{-1}y$

$\Rightarrow f^*(y) = \frac{1}{2}y^TQ^{-1}y$ 凸.

α-sublevel set（下水平集）：若 $f:\mathbb{R}^n\to\mathbb{R}$，$C_\alpha \triangleq \{x\in dom f \mid f(x)\leq\alpha\}$.

凸函数的下水平集都是凸集.

证明：$\forall x,y\in C_\alpha$　$x,y\in dom f$，$f(x)\leq\alpha$，$f(y)\leq\alpha$　f凸 $\Rightarrow \forall\theta\in[0,1]$　$f(\theta x+(1-\theta)y)\leq\theta f(x)+(1-\theta)f(y)\leq\alpha$.

$\Rightarrow \theta x+(1-\theta)y\in C_\alpha$.

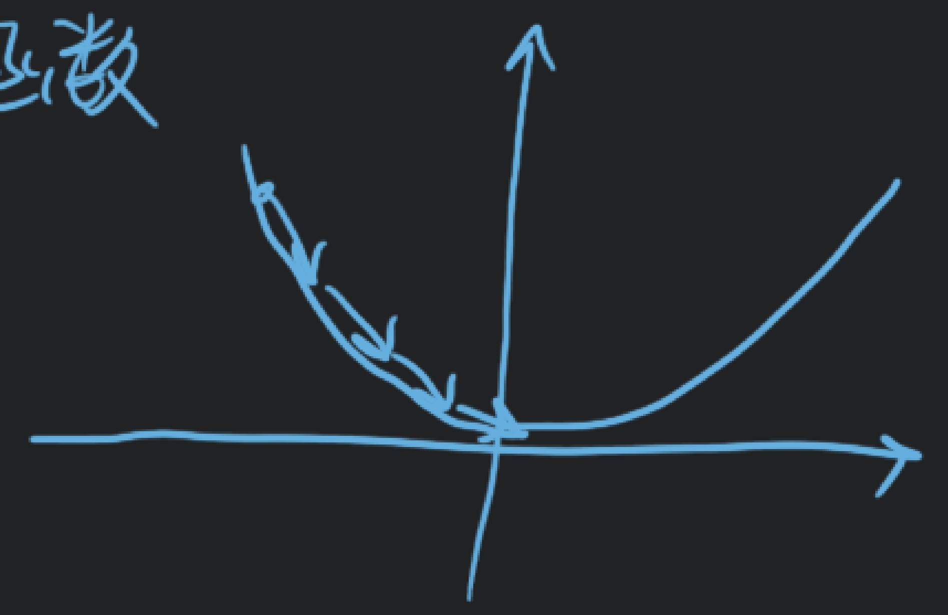
但反过来不一定对．函数的下水平集都是凸集，f不一定是凸函数　　e.g. $f(x) = -e^x$

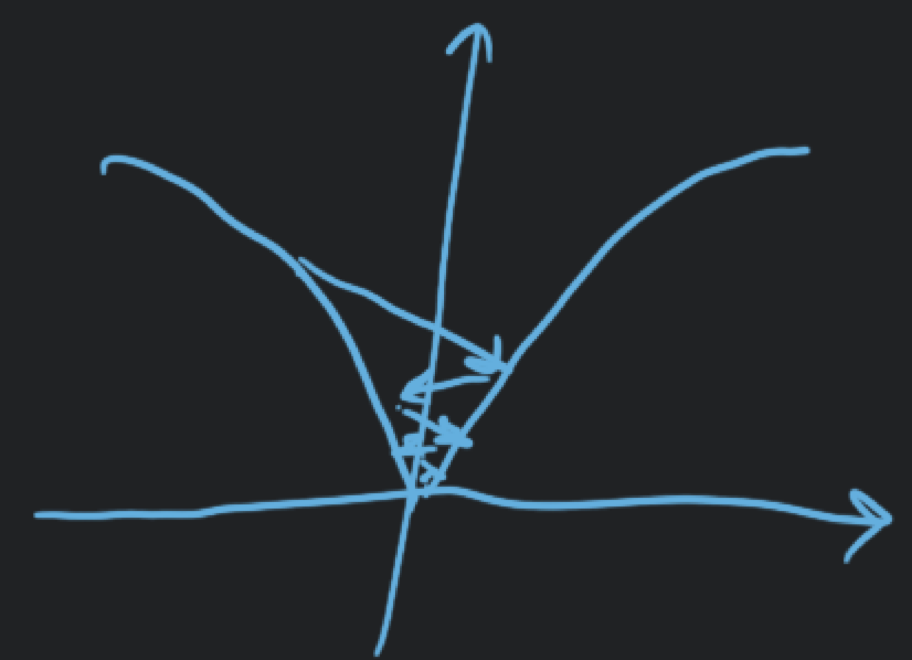
拟凸函数（Quasi Convex Function）（单模态函数，unimodel function）下水平集都是凸集的函数　拟凸函数不一定是凸函数．

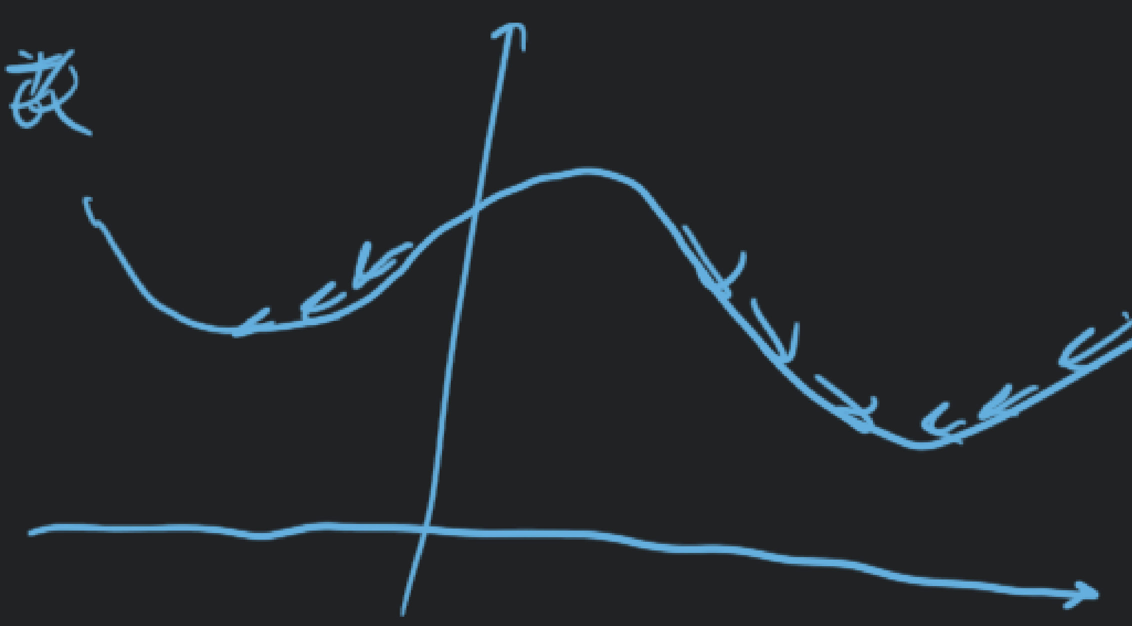
拟凹函数　$S_\alpha' = \{x \in \mathrm{dom} f \mid f(x) \geq \alpha\}$　上水平集都是凸集的函数．

拟线性函数　$S_\alpha'' = \{x \in \mathrm{dom} f \mid f(x) = \alpha\}$　都是凸集的函数．

梯度下降　凸函数　　单模态函数　　多模态函数







拟凸函数 定义2：$f: \mathbb{R}^n \to \mathbb{R}$ 拟凸 $\Leftrightarrow$ $\mathrm{dom} f$ 凸，$\forall x, y \in \mathrm{dom} f$，$\forall \theta \in [0,1]$，$\max\{f(x), f(y)\} \geq f(\theta x + (1-\theta) y)$．

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

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

$\{f(x) \leq \alpha\} \Rightarrow i = \lfloor \alpha \rfloor + 1, \cdots, n,\ x_i = 0$．即 $\mathbb{R}^n$ 中子空间，是凸集 $\Rightarrow f(x)$ 拟凸．

e.g. 线性分数函数　$f(x) = \dfrac{a^T x + b}{c^T x + d}$　　$\mathrm{dom} f = \{x \mid c^T x + d > 0\}$．

$S_\alpha = \{x \mid c^T x + d > 0,\ \dfrac{a^T x + b}{c^T x + d} \leq \alpha\}$ 是多面体，凸集 $\Rightarrow f(x)$ 拟凸．

e.g. $\min \|x\|_0$ s.t. $x \in C$．　$\|x\|_0$ 在高维时不是拟凸函数 $\Rightarrow$ 难以求解

近似：$\min \log(x^T x + 1)$ s.t. $x \in C$　　$\log(x^T x + 1)$ 是拟凸函数

可微拟凸函数一阶条件（定义3）　$f:\mathbb{R}^n\to\mathbb{R}$ 可微，f 拟凸 $\Longleftrightarrow$ $\mathrm{dom}f$ 为凸集，$\forall x,y\in\mathrm{dom}f$　$f(y)\le f(x)\Rightarrow$ $\nabla^T f(x)(y-x)\le 0$

prove：一维情况．"$\Rightarrow$"　$\forall x,y\in\mathrm{dom}f$　$0\le\theta\le1$　设 $f(y)\le f(x)$，$f(x)\ge f(\theta x+(1-\theta)y)\Longleftrightarrow$

$f(\theta x+(1-\theta)y)-f(\theta x+(1-\theta)x)\le0\Longleftrightarrow\dfrac{f(\theta x+(1-\theta)y)-f(\theta x+(1-\theta)x)}{(1-\theta)(y-x)}(1-\theta)(y-x)\le0$

$\xrightarrow{\theta\to1^-} f'(x)(y-x)\le0$.

"$\Leftarrow$"　$\forall x,y\in\mathrm{dom}f$，设 $f(y)\le f(x)\Rightarrow f'(x)(y-x)\le0$．　$\forall\theta\in(0,1)$　$z\triangleq\theta x+(1-\theta)y$.

先证：若 $f(z)\ge f(x)$，则 $f(z)=f(x)$

$f(z)\ge f(x)\ge f(y)\Rightarrow f'(z)(x-z)\le0$　　$f'(z)(y-z)\le0$　$\Rightarrow f'(z)(1-\theta)(x-y)\le0$　$f'(z)\theta(y-x)\le0$.

$\Rightarrow f'(z)=0\Rightarrow f(z)$是极大值　　若 $f(z)>f(x)$，其邻域 $f(z)\ge f(x)\Rightarrow f'(z)=0$，不断扩展 $\Rightarrow f(z)=f(x)$ 矛盾.

$\Rightarrow f(z)\le f(x)$　即证.

若 $\nabla f(x)=0$　凸函数：$\forall y$，$f(y)\ge f(x)$.　　拟凸函数：$0\le0$ 无意义．（e.g.　$f(x)=x^3$）

二阶条件（定义4）　$f:\mathbb{R}^n\to\mathbb{R}$ 可微，f 拟凸 $\Longleftrightarrow$ $\mathrm{dom}f$ 凸，$\forall x,y\in\mathrm{dom}f$，$y^T\nabla f(x)\ge0\Rightarrow y^T\nabla^2 f(x)y\ge0$.

$n=1$：　$yf'(x)\ge0\Rightarrow y^2f''(x)\ge0$　$\forall y\ne0$，$f'(x)=0\Rightarrow f''(x)\ge0$（高维：半正定）　$\nabla^2 f$ 仅需在 $\nabla f=0$ 的点半正定

对数凸/凹　　$f:\mathbb{R}^n\to\mathbb{R}$　对数凹：　若 $f(x)>0$，$\forall x\in\mathrm{dom}f$，且 $\log f$ 为凹函数.

对数凸：　$f(x)>0$，$\forall x\in\mathrm{dom}f$　且 $\log f$ 为凸函数

f 对数凸 $\Rightarrow$ f 凸　　$f=e^{\log f}$　　$\log f$ 凸，e^x 凸且增

f 凹，$f>0\Rightarrow f$ 对数凹　　　对数凸比凸更强，　对数凹比凹更弱.

4. 凸优化问题

一般优化问题描述： $\min\ f_0(x)$ s.t. $f_i(x)\le 0$ $i=1,\cdots,m$ $h_i(x)=0$ $i=1,\cdots,p$.

ε-最优解集： $X_\varepsilon=\{x \mid x\in X_f,\ f_0(x)\le \underbrace{p^*}_{\text{最优解 } \min f_0}+\varepsilon\}$

局部最优解 $\exists R>0$ s.t. $f_0(x)=\inf\{f_0(z) \mid f_i(z)\le 0\ \ i=1,\cdots,m\ \ h_i(z)=0\ \ i=1,\cdots,p\ \ \|z-x\|\le R\}$

若 $x\in X_f$，$f_i(x)=0$ 称 $f_i(x)\le 0$ 为活动约束， 若 $f_i(x)<0$，称 $f_i(x)\le 0$ 为不活动约束.

可行性优化问题： 令 $f_0(x)=\text{Const}$. $\Rightarrow \{x \mid f_i(x)\le 0,\ h_i(x)=0\}$.

凸优化问题（广义） 目标是凸集，可行域是凸集.

凸优化问题（狭义） $\min f_0(x)$ s.t. $f_i(x)\le 0$ $i=1,\cdots,m$ $a_i^Tx=b_i$ $i=1,\cdots,p$ 且 $f_0,f_1,\cdots,f_m$ 为凸函数

（目标和不等式约束是凸函数，等式约束是仿射函数） 可行域是凸集.

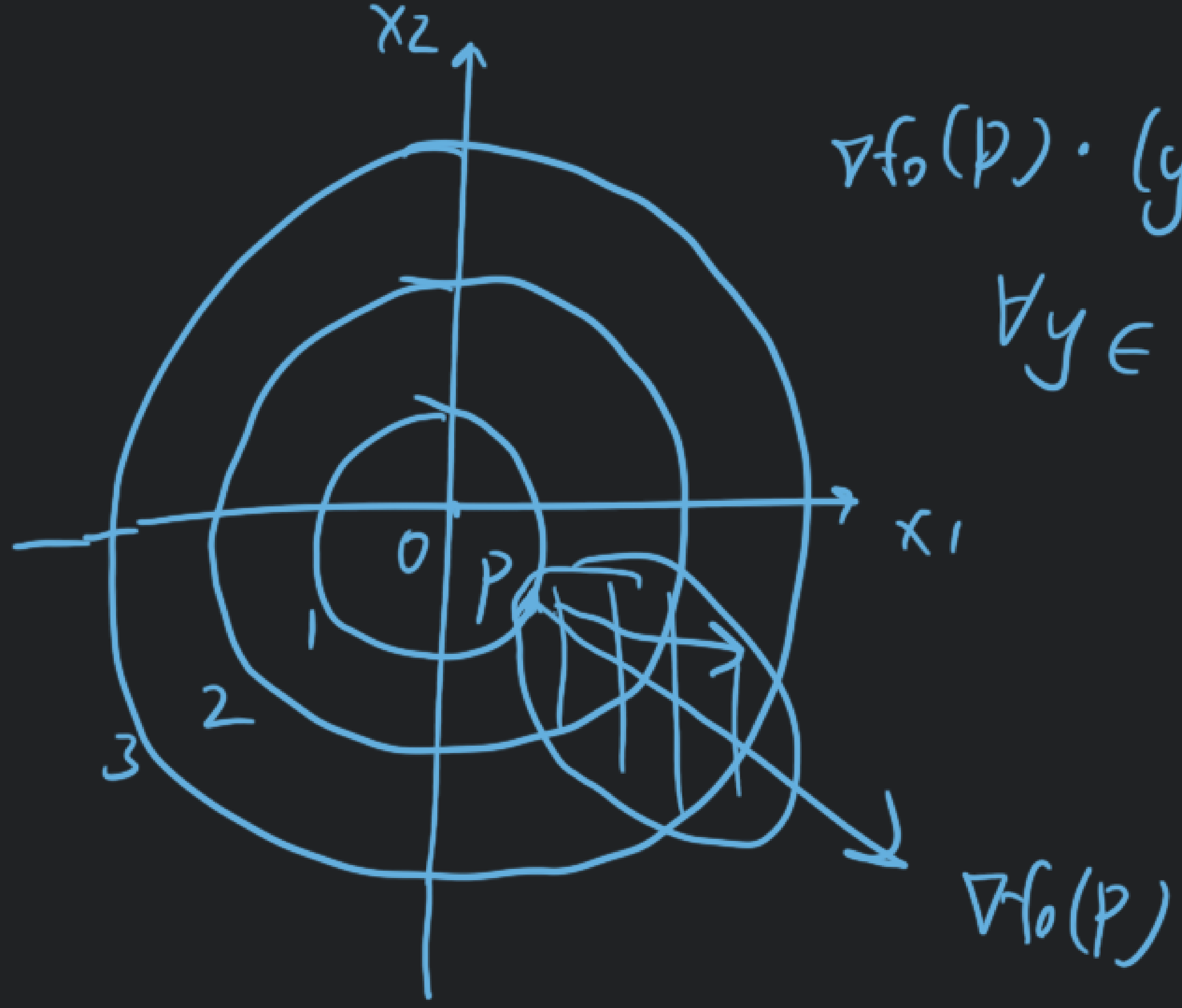
性质：局部最优 = 全局最优.

证明：设 x 局部最优但不全局最优 $\exists R>0$ $f_0(x)=\inf\{f_0(z) \mid z\in \text{可行域},\ \|x-z\|\le R\}$.

$\exists y$ 可行 $\|y-x\|>R$ $f_0(y)<f_0(x)$ $z=(1-\theta)x+\theta y$ $\theta=\dfrac{R}{2\|y-x\|}$ z 可行 且 $f_0(z)\le(1-\theta)f_0(x)+\theta f_0(y)<f_0(x)$.

$\|z-x\|=\theta\|x-y\|=\dfrac{R}{2}<R$ $f_0(y)<f_0(x)\le f_0(z)$ 但 $f_0(z)<f_0(x)$ 矛盾！

目标 f_0 可微： $x^*\in X_f$ 最优 $\Longleftrightarrow$ $\nabla f_0(x^*)^T(y-x^*)\ge 0$ $\forall y\in X_f$.



$\nabla f_0(p)\cdot(y-x_p)\geq 0$

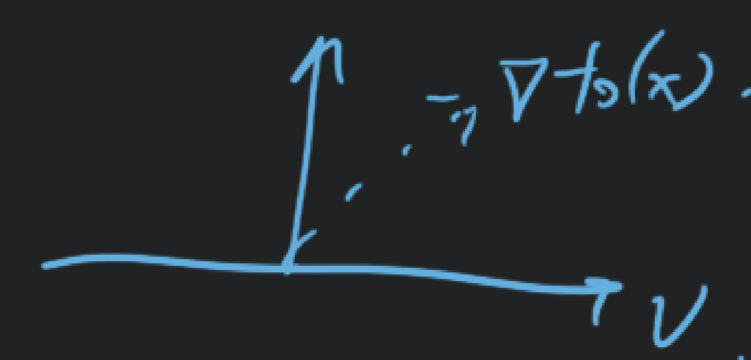
$\forall y\in X_f$.

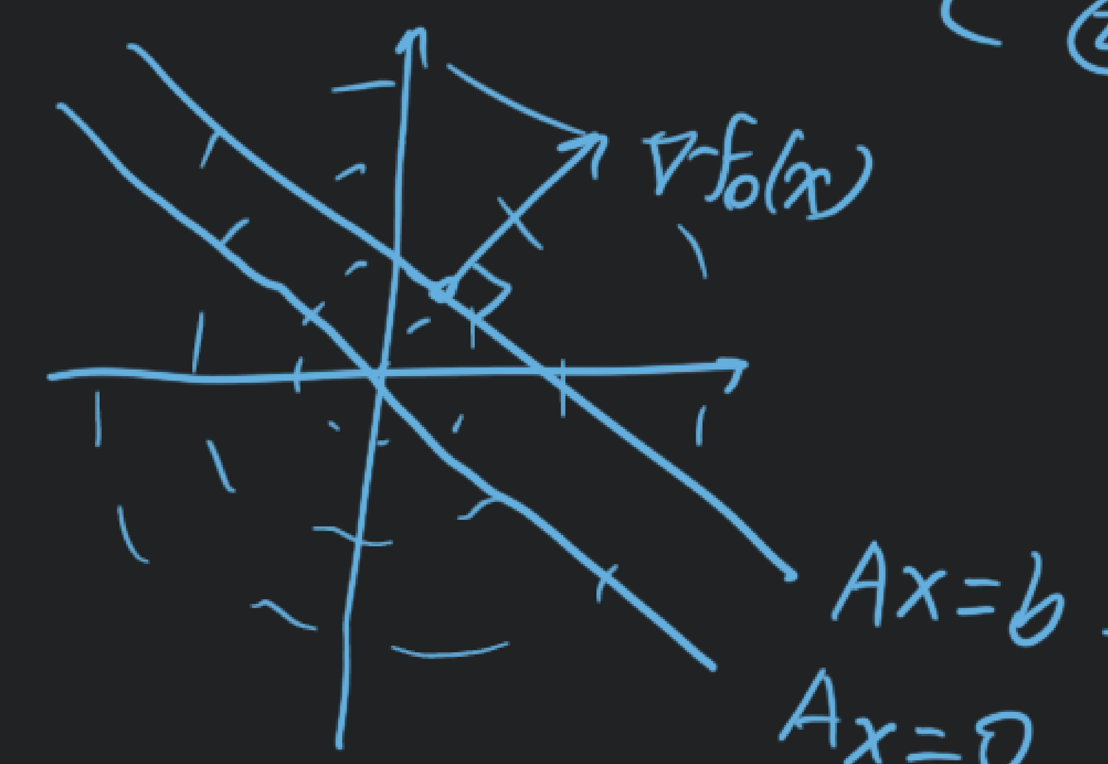
e.g. 约束仅为等式约束. $\min f_0(x)$ $\operatorname{dom} f_0=\mathbb{R}^n$ s.t. $Ax=b$.

若 $\exists x$, $Ax=b$ x最优 $\Longleftrightarrow$ $\forall y$, $Ay=b$ $\nabla f_0(x)^T(y-x)\geq 0$.

$y=x+v$ $v\in N(A)$ (A的零空间, 即 $Av=0$).

$\nabla f_0(x)^T v\geq 0$ $\forall v\in N(A)$. $\Longleftrightarrow$ $\begin{cases} ① \ v=0 & A\text{可逆} \quad x=A^{-1}b. \\ ② \ \nabla f_0(x)\text{与}N(A)\text{正交} \end{cases}$





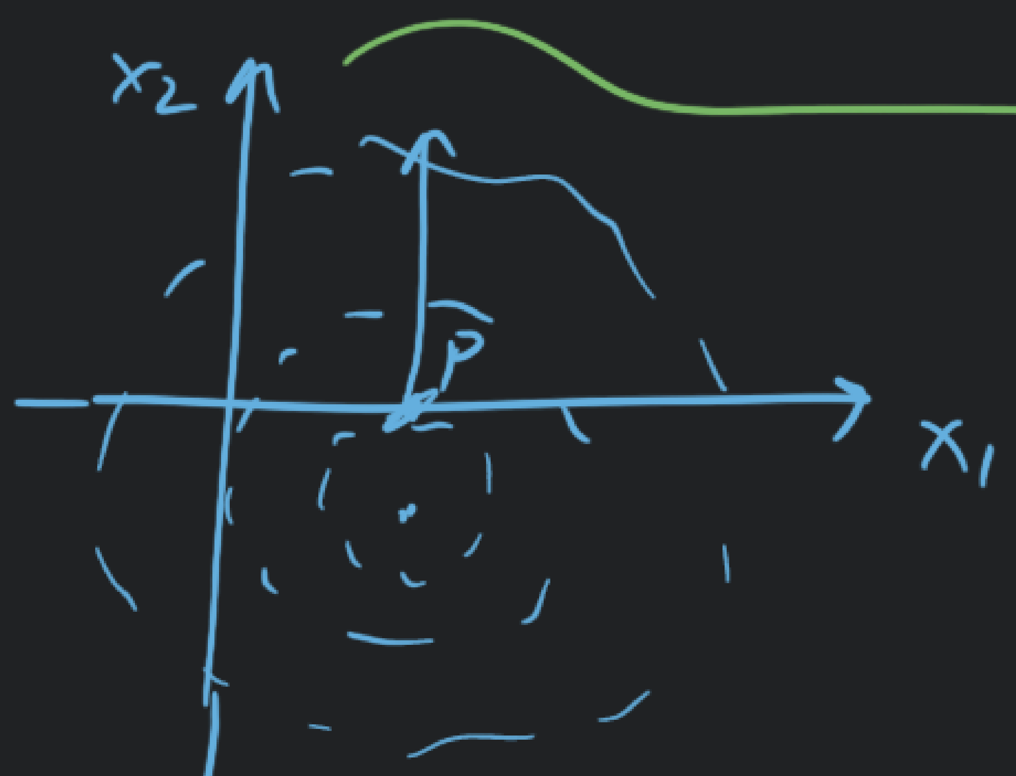
e.g. 约束仅为非负约束 $\min f_0(x)$ s.t. $x\geq 0$. (means $\forall i$, $x_i\geq 0$).

若 $\exists x\geq 0$, x最优 $\Longleftrightarrow$ $\forall y\geq 0$ $\nabla f_0(x)^T(y-x)\geq 0$.

① 若 $\nabla f_0(x)$有一项小于0, 可以取 $y-x=(0,\cdots,0,1,0,\cdots 0)$ $\Rightarrow$ $\nabla f_0(x)\geq 0$ ($\nabla f_0(x)$每一项均≥ 0).

② 取 $y=0$, $\nabla f_0(x)^T x\leq 0$ 又 $\nabla f_0(x)\geq 0$, $x\geq 0$. $\Rightarrow$ $\nabla f_0(x)^T x\geq 0$ $\Rightarrow$ $\nabla f_0(x)^T x=0$

$\Rightarrow$ $\forall i$, $x_i\geq 0$ $[\nabla f_0(x)]_i\geq 0$ $[\nabla f_0(x)]_i x_i=0$ (互补条件)



最优解P点. $x_1>0$ $x_2=0$

$[\nabla f_0(x)]_1=0$ $[\nabla f_0(x)]_2>0$.

显型的凸问题

① 线性规划 $\min\ c^Tx+d \quad s.t.\ Gx\le h \quad Ax=b \qquad c\in\mathbb{R}^n,\ d\in\mathbb{R} \quad G\in\mathbb{R}^{m\times n} \quad h\in\mathbb{R}^m$

目标和约束均为线性的. $A\in\mathbb{R}^{k\times n} \quad b\in\mathbb{R}^n$.

等价变换: $\min\ c^Tx+d \quad s.t.\ Gx+s=h \quad Ax=b \quad s\ge 0$.

令 $x=x^+-x^- \quad x^+,x^-\ge 0$ 等价问题 $\min\ c^Tx^+-c^Tx^-+d \quad s.t.\begin{cases} Gx^+-Gx^-+s=h \quad Ax^+-Ax^-=b \\ s,\ x^+,\ x^-\ge 0.\end{cases}$

e.g. 食谱问题. n种食物, m种营养, 第i种食物单位价格 c_i, 含第j种营养 a_{ij}

要求: 第j种营养至少有 b_j $(j=1,\cdots,m)$ 的前提下最小化总成本.

设食物量为 $x_1,\cdots,x_n$ $\quad x=[x_1,\cdots;x_n]^T. \quad \min\ c^Tx \quad s.t.\ A^Tx\ge b,\ x\ge 0.$

e.g. 线性分数规划 (P0) $\min\ f_0(x) \quad s.t.\ Gx\le h \quad Ax=b \qquad f_0(x)=\dfrac{c^Tx+d}{e^Tx+f} \quad \mathrm{dom}\ f_0=\{x\mid e^Tx+f>0\}.$

(P1) $\min\ c^Ty+dz \quad s.t.\ Gy-hz\le 0 \quad Ay-bz=0 \quad e^Ty+fz=1 \quad z\ge 0$

(P0)与(P1)等价. 证明: 1)若 x 在(P0)可行, 取 $y=\dfrac{x}{e^Tx+f} \qquad z=\dfrac{1}{e^Tx+f}$

$$\begin{cases} Gx\le h \\ Ax=b \\ e^Tx+f>0 \end{cases} \Rightarrow \begin{cases} Gy-hz=\dfrac{Gx-h}{e^Tx+f}\le 0 \qquad e^Ty+fz=\dfrac{e^Tx+f}{e^Tx+f}=1 \text{ 且 } \dfrac{c^Tx+d}{e^Tx+f}=c^Ty+dz. \\ Ay-bz=\dfrac{Ax-b}{e^Tx+f}=0 \qquad z>0. \end{cases}$$

2) 若 y, z 对 (P1) 可行，若 $z>0$，取 $x=\frac{y}{z}$. 同样可证 x 在 (P0) 可行，且 $f_0(x)=c^Ty+dz$.

若 $z=0$. 设 x_0 为 P0 的 1 个可行解，则 $x=x_0+ty$ 也对 P0 可行 $\forall t\geq 0$

$Gy\leq 0 \quad Ay=0 \quad e^Ty=1 \qquad Gx=Gx_0+tGy\leq h \qquad Ax=Ax_0+tAy=b \qquad e^Tx+f=e^Tx_0+f+te^Ty\geq 0$

且 $f_0(x)=f_0(x_0+ty)=\dfrac{c^Tx_0+c^Tty+d}{e^Tx_0+e^Tty+f}\xlongequal{t\to+\infty} c^Ty.$

② 二次规划 $\min \frac{1}{2}x^TPx+q^Tx+r \quad P\in S_+^n \quad$ s.t. $Gx\leq h \quad Ax=b.$

二次约束二次规划 (QCQP) $\min \frac{1}{2}x^TPx+q^Tx+r \quad$ s.t. $\begin{cases}\frac{1}{2}x^TP_ix+q_i^Tx+r_i\leq 0 \\ Ax=b\end{cases} \quad P\in S_+^n, \ P_i\in S_+^n$

e.g. 带噪声的测量系统 $\underset{\text{测量}}{b}=Ax+\underset{\text{噪声}}{e}$ $\quad \hat{x}=\underset{x}{\arg\min}\|b-Ax\|_2^2=\underset{x}{\arg\min}\ x^TA^TAx-2b^TAx+b^Tb$

无约束二次规划问题 若 A 可逆，$\hat{x}=(A^TA)^{-1}A^Tb.$

e.g. 若 x 稀疏为先验知识 $\hat{x}=\underset{x}{\arg\min}\|b-Ax\|_2^2+\lambda_0\|x\|_0 \simeq \arg\min\|b-Ax\|_2^2+\lambda_1\|x\|_1$

$x=x^+-x^- \qquad \hat{x}=\underset{x}{\arg\min}\|b-Ax^++Ax^-\|_2^2+\lambda_1 1^Tx^++\lambda_1 1^Tx^- \quad$ s.t. $x^+, x^-\geq 0.$

L1 范数规范的最小二乘

e.g. 若先验已知 x 中元素幅度类似 $\longrightarrow$ L2 范数规范的最小二乘.

$\hat{x}=\underset{x}{\arg\min}\|b-Ax\|_2^2+\lambda_2\|x\|_2^2 \qquad$ 岭回归

e.g. 投资组合问题　n种方案，收益率 $\bar{p} \sim N(\bar{p}, \Sigma)$　$\min X^T\Sigma X$　s.t. $\bar{p}^T X \geq r_{min}$　$1^T X = \overset{\text{预算}}{B}$　$X \geq 0$

半正定规划　$\min \mathrm{Tr}(CX)$　s.t. $\mathrm{Tr}(A_i X) = b_i$　$(i=1,\cdots,p)$　$X \succeq 0$　　$X \in S_+^n$, $C \in R^{n\times n}$　$A_i \in R^{n\times n}$　$b_i \in R$

特例：C, A_i, X 均为对角矩阵 $\Rightarrow$ $\min c^T x$　s.t. $A_i^T x = b_i$　$x \geq 0$　　$c, A_i \in \mathbb{R}^n$　$b_i \in \mathbb{R}$

$\min c^T x$　s.t. $x_1A_1 + \cdots + x_nA_n \preceq B$　　$x \in \mathbb{R}^n$　　$B, A_1, \cdots, A_n \in S^k$ (对称k维矩阵)　　$c \in \mathbb{R}^n$

e.g. $A(x) = A_0 + x_1A_1 + \cdots + x_nA_n$　$A_i \in \mathbb{R}^{p\times q}$　$(i=0,\cdots,n)$　　$x \in \mathbb{R}^n$　　谱范数 $\|A(x)\|_2 = A(x)$的最大奇异值

$\min \|A(x)\|_2$.　　利用 $\|A(x)\|_2 \leq \sqrt{s}$　$s \geq 0$ $\Longleftrightarrow$ $A^T(x)A(x) - sI \preceq 0$

等价问题：$\min \sqrt{s}$　s.t. $A^T(x)A(x) \preceq sI$, $s > 0$

$\Longleftrightarrow$ 凸问题：$\min t$　s.t. $A^T(x)A(x) - t^2 I \preceq 0$　$t \geq 0$

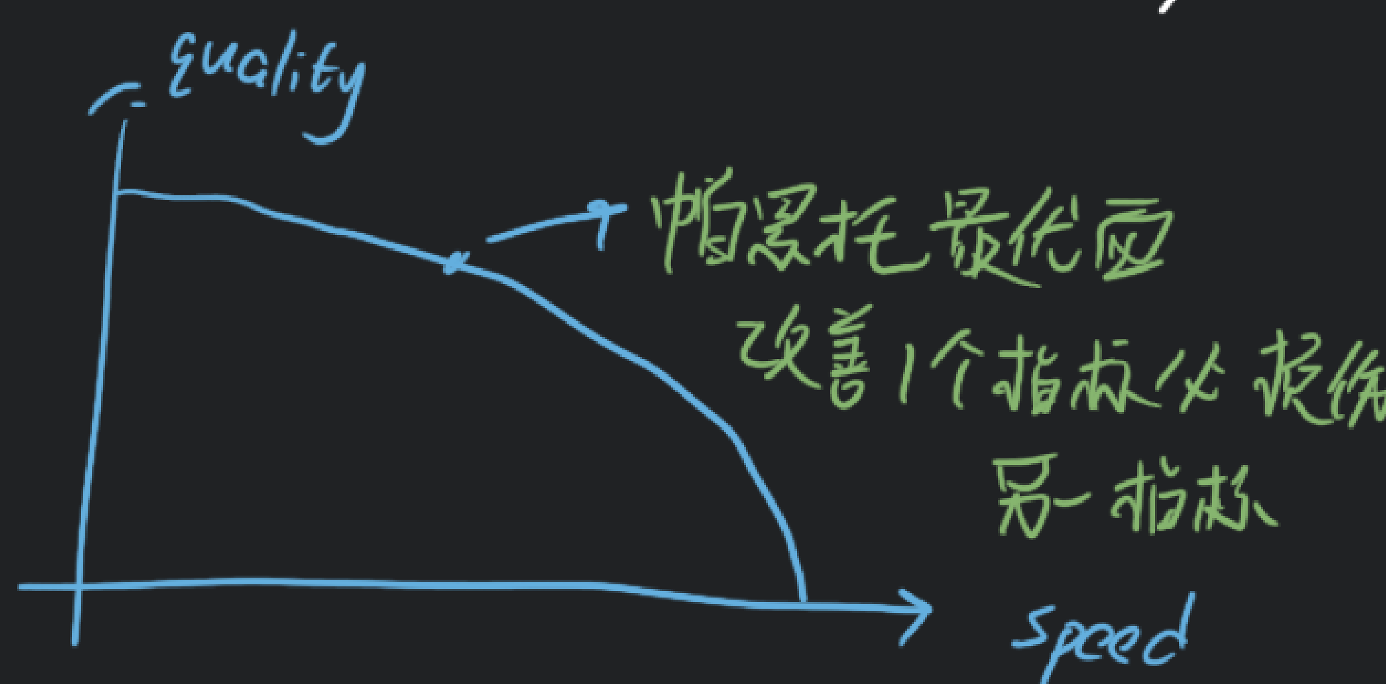
$\Longleftrightarrow$ $\min t$　s.t. $\begin{array}{c|cc} & p & q \\ \hline p & tI & A(x) \\ q & A(x)^T & tI \end{array} \succeq 0$　$t \geq 0$

$\Longleftrightarrow$ 半正定规划问题　$\min t$　s.t. $Y = \begin{bmatrix} tI & A(x) \\ A(x)^T & tI \end{bmatrix}$　$Y \succeq 0$　$t \geq 0$　(优化变量：t, Y, x)

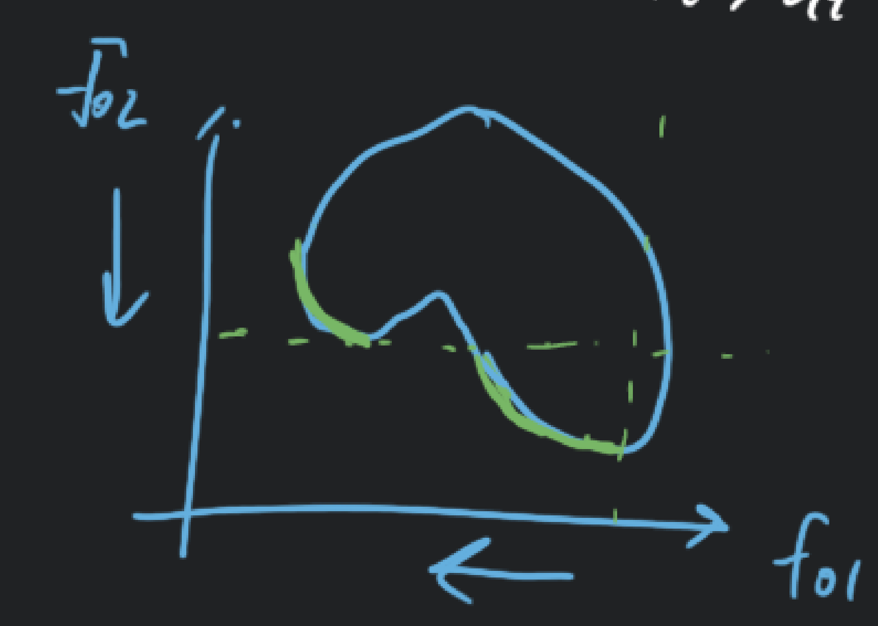
多目标优化问题　$\min f_0(x)$　s.t. $f_i(x) = 0$　$(i=1,\cdots,m)$　$h_i(x) = 0$　$(i=1,\cdots,p)$　　$f_0: \mathbb{R}^n \to \mathbb{R}^q$　　$f_i, h_i: \mathbb{R}^n \to \mathbb{R}$

e.g. min Risk, min −income　s.t. Resources

min −quality, min −speed



e.g.



若 $\{f_0(x)\}$ 在 $\mathbb{R}^q$ 中为凸集，$f_i(x)$ 为凸函数，$h_i(x)$ 为仿射函数，则可通过下述流程求得 Pareto 最优面

$$\min \sum_{i=1}^{q} \lambda_i [f_0(x)]_i \quad s.t.\ \ f_i(x)\le 0 \ \ (i=1,\cdots,m) \quad h_i(x)=0 \ \ (i=1,\cdots,p)$$

其中 $\lambda_i \geq 0$ 遍历 λ_i 可找到 Pareto 面所有点

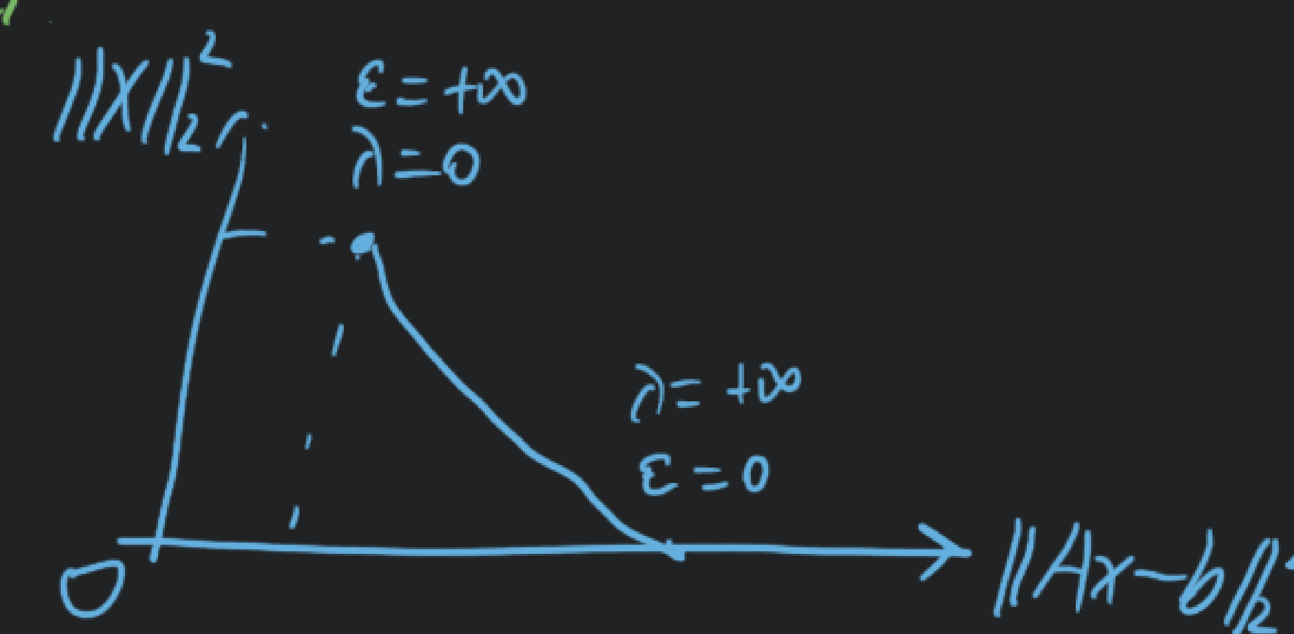
本质是对各目标取线性组合，但若 $\{f_0(x)\}$ 不是凸集，有些点不能通过组合求出

e.g. Ridge Regression $\quad b = Ax + e \quad A\in\mathbb{R}^{m\times n} \quad b\in\mathbb{R}^m \quad e\in\mathbb{R}^m \quad x\in\mathbb{R}^n$

$\min \|b-Ax\|_2^2 \quad \min \|x\|_2^2 \quad \{\|b-Ax\|_2^2, \|x\|_2^2\}$ 是凸集

$\Rightarrow \min \|b-Ax\|_2^2 + \lambda\|x\|_2^2$ 遍历 λ，可得到帕累托面所有点.

实践中另一方法：$\min \|b-Ax\|_2^2 \quad s.t.\ \|x\|_2^2 \leq \varepsilon$



5. 对偶性

$\min f_0(x) \quad s.t.\ f_i(x)\le 0 \ (i=1,\cdots,m) \quad h_i(x)=0 \ (i=1,\cdots,p)$. $\quad x\in\mathbb{R}^n$

域

$D = \left(\bigcap_{i=0}^{m} \mathrm{dom} f_i\right) \cap \left(\bigcap_{i=1}^{p} \mathrm{dom}\, h_i\right)$ $\quad p^*$ 最优值

拉格朗日函数 $\quad L(x,\lambda,v) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x) \quad \lambda\in\mathbb{R}^m \quad v\in\mathbb{R}^p$

拉格朗日对偶函数 $\quad g(\lambda,v) = \inf_{x\in D} L(x,\lambda,v)$ （是拉格朗日乘子的函数）

(1) 对偶函数总是凹函数（无论问题是否是凸的）

(2) $\forall \lambda \geq 0, \ \forall v, \quad g(\lambda,v) \leq p^*$ （对偶函数总给出最优值的1个下界）

设 x^* 是最优解，则必有 $\quad f_i(x^*) \leq 0 \ (i=1,\cdots,m) \quad h_i(x^*)=0 \ (i=1,\cdots,p)$

$\forall\lambda\geq 0, \forall v, \quad g(\lambda,v) = \inf_{x\in D} L(x,\lambda,v) \leq L(x^*,\lambda,v) \leq f_0(x^*) = p^*.$

e.g. $\min x^Tx$ s.t. $Ax=b$ $x\in\mathbb{R}^n$ $b\in\mathbb{R}^p$ $A\in\mathbb{R}^{p\times n}$

$L(x,v) = x^Tx + v^T(Ax-b) \Rightarrow g(v) = \inf_{x\in D} L(x,v) = \inf(x^Tx + v^TAx - v^Tb)$.

$2x^* + A^Tv = 0 \Rightarrow x^* = -\frac{A^Tv}{2}$ $g(v) = -\frac{1}{4}v^TAA^Tv - b^Tv$ 关于v是凹函数.

e.g. $\min c^Tx$ s.t. $Ax=b$ $x\geq 0$

$L(x,\lambda,v) = c^Tx - \lambda^Tx + v^T(Ax-b) = -b^Tv + (c+A^Tv-\lambda)^Tx$.

$\Rightarrow g(\lambda,v) = \inf_x L(x,\lambda,v) = \begin{cases} -b^Tv & A^Tv-\lambda+c=0 \\ -\infty & \text{else} \end{cases}$ 关于λ,v是凹函数.

e.g. $\min x^TWx$ s.t. $x_i=\pm 1$ $(i=1,\cdots,m)$ $x\in\mathbb{R}^m$ $W\in\mathbb{R}^{m\times m}$

$L(x,v) = x^TWx + \sum_{i=1}^n v_i(x_i^2-1) = x^T(W+\mathrm{Diag}(v))x - 1^Tv$.

$g(v) = \inf_x L(x,v) = \begin{cases} -1^Tv & W+\mathrm{Diag}(v)\geq 0 \\ -\infty & W+\mathrm{Diag}(v)<0 \end{cases}$ $\{v \mid W+\mathrm{Diag}(v)\geq 0\}$是凸集 $\Rightarrow g(v)$是凹函数.

(用定义证).

函数的共轭与对偶函数的联系 $f^*(y) = \sup_{x\in\mathrm{dom}f}(x^Ty - f(x))$.

e.g. $\min f(x)$ s.t. $x=0$ $L(x,v) = f(x) + v^Tx$ $\mathrm{dom}\,L = \mathrm{dom}f \times \mathbb{R}^n$

$g(v) = \inf_{x\in\mathrm{dom}f}(f(x)+v^Tx) = -\sup_{x\in\mathrm{dom}f}(-v^Tx - f(x)) = -f^*(-v)$.

e.g. $\min f_0(x)$ s.t. $Ax\leq b$ $Cx=d$. $L(x,\lambda,v) = f_0(x) + \lambda^T(Ax-b) + v^T(Cx-d) = f_0(x) + (\lambda^TA + v^TC)x - \lambda^Tb - v^Td$.

$\Rightarrow g(\lambda,\nu) = \inf_{x\in \mathrm{dom} f_0} L(x,\lambda,\nu) = -f_0^*\left(-(\lambda^TA+\nu^TC)^T\right) - \lambda^Tb - \nu^Td$.

对偶函数提供的 p^* 的最紧下界 (D) $\begin{cases} \max & g(\lambda,\nu) \triangleq d^* \\ \text{s.t.} & \lambda \geq 0 \end{cases}$ （拉格朗日对偶问题）

凸优化问题（无论原问题是否凸）

原问题 (P) $\begin{cases} \min & f_0(x) \triangleq p^* \\ \text{s.t.} & f_i(x) \leq 0 \quad (i=1,\cdots,m) \quad h_i(x)=0 \quad (i=1,\cdots,p). \end{cases}$

等价问题的对偶问题不一定等价.

(1) $d^* \leq p^*$ （弱对偶性）

(2) (D) $\longrightarrow \lambda^*, \nu^*$ 最优拉格朗日乘子.

e.g. $\min c^Tx$ s.t. $Ax=b, x\geq 0$ $\quad g(\lambda,\nu) = \begin{cases} -b^T\nu & A^T\nu-\lambda+c=0 \\ -\infty & \text{else} \end{cases}$

(D) $\Rightarrow \begin{matrix} \max & g(\lambda,\nu) \\ \text{s.t.} & \lambda\geq 0 \end{matrix} \iff \begin{matrix} \max & -b^T\nu \\ \text{s.t.} & \lambda \geq 0,\ A^T\nu-\lambda+c=0 \end{matrix} \iff \begin{matrix} \min & b^T\nu \\ \text{s.t.} & A^T\nu+c\geq 0 \end{matrix}$

e.g. $\min c^Tx$ s.t. $Ax\leq b \Rightarrow L(x,\lambda) = c^Tx+\lambda^T(Ax-b) = (c+A^T\lambda)^Tx - b^T\lambda$.

$\Rightarrow g(\lambda) = \inf_x L(x,\lambda) = \begin{cases} -b^T\lambda & c+A^T\lambda=0 \\ -\infty & \text{otherwise} \end{cases}$

(D) $\Rightarrow \begin{matrix} \max & g(\lambda). \\ \text{s.t.} & \lambda\geq 0 \end{matrix} \iff \begin{matrix} \min & b^T\lambda \\ \text{s.t.} & A^T\lambda+c=0 \\ & \lambda\geq 0 \end{matrix}$

原问题 变量 n 维 约束 p 维

对偶问题 变量 p 维 约束 n 维

上述两问题对偶的对偶等于其自身，但此性质一般不成立.（对偶问题一定可写成凸优化问题）

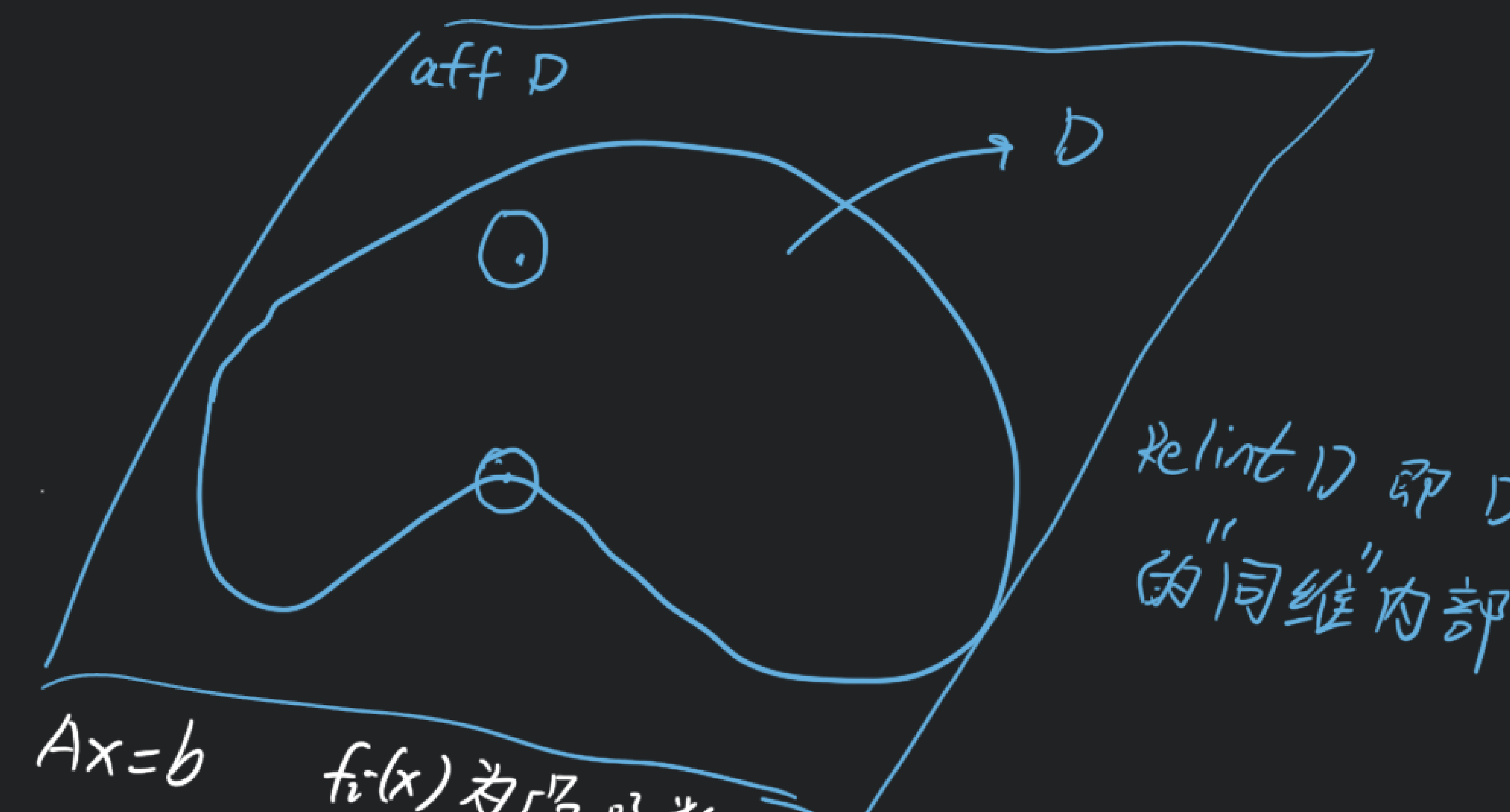
强对偶性： $d^* = p^*$.　（原问题和对偶问题最优解相等）

对偶间隙　$p^* - d^*$.

原问题　$\mathrm{relint}\ D = \{ x \in D \mid B(x,r) \cap \mathrm{aff}\ D \subseteq D,\ \exists r > 0 \}$

（上域；$B(x,r)$：以x为中心，r为半径的球；aff：仿射包）

$\left(D = \left(\bigcap_{i=0}^{m} \mathrm{dom} f_i\right) \cap \left(\bigcap_{i=1}^{p} h_i(x)\right) \right)$



Slater 条件　若有凸问题　$\min f_0(x)$　s.t. $f_i(x) \le 0$ $(i=1,\cdots,m)$　$Ax=b$　$f_i(x)$为凸函数 $(i=0,\cdots,m)$

且 $\exists x \in \mathrm{relint}\ D$ 且 $f_i(x) < 0$ $(i=1,\cdots,m)$　$Ax=b$ 满足时　（relint D 与严格可行域有交集）

$p^* = d^*$，强对偶性成立.　（不是必要条件）

弱 Slater 条件　若上述凸问题的不等式约束为仿射函数，只要 $\mathrm{dom} f_0$ 为开集，可行域非空，必有 $p^* = d^*$

e.g.　$\min\ x^T x$　s.t. $Ax = b$　　$p^* = d^*$.

e.g. QCQP　$\begin{cases} \min & \frac{1}{2} x^T P_0 x + q_0^T x + r_0 \\ \text{s.t.} & \frac{1}{2} x^T P_i x + q_i^T x + r_i \le 0 \quad (i=1,\cdots,m) \end{cases}$　$P_0 \in S_{++}^n$　$P_i \in S_+^n$

$$L(x,\lambda) = \frac{1}{2} x^T P_0 x + q_0^T x + r_0 + \sum_{i=1}^{m} \lambda_i \left(\frac{1}{2} x^T P_i x + q_i^T x + r_i \right)$$

$$g(\lambda) = \inf_x L(x,\lambda) = -\frac{1}{2} q^T(\lambda) P^{-1}(\lambda) q(\lambda) + r(\lambda) \quad (\lambda \ge 0 \text{ 时})$$

$P(\lambda) = P_0 + \sum_{i=1}^{m} \lambda_i P_i$　　$q(\lambda) = q_0 + \sum_{i=1}^{m} \lambda_i q_i$

$r(\lambda) = r_0 + \sum_{i=1}^{m} \lambda_i r_i$

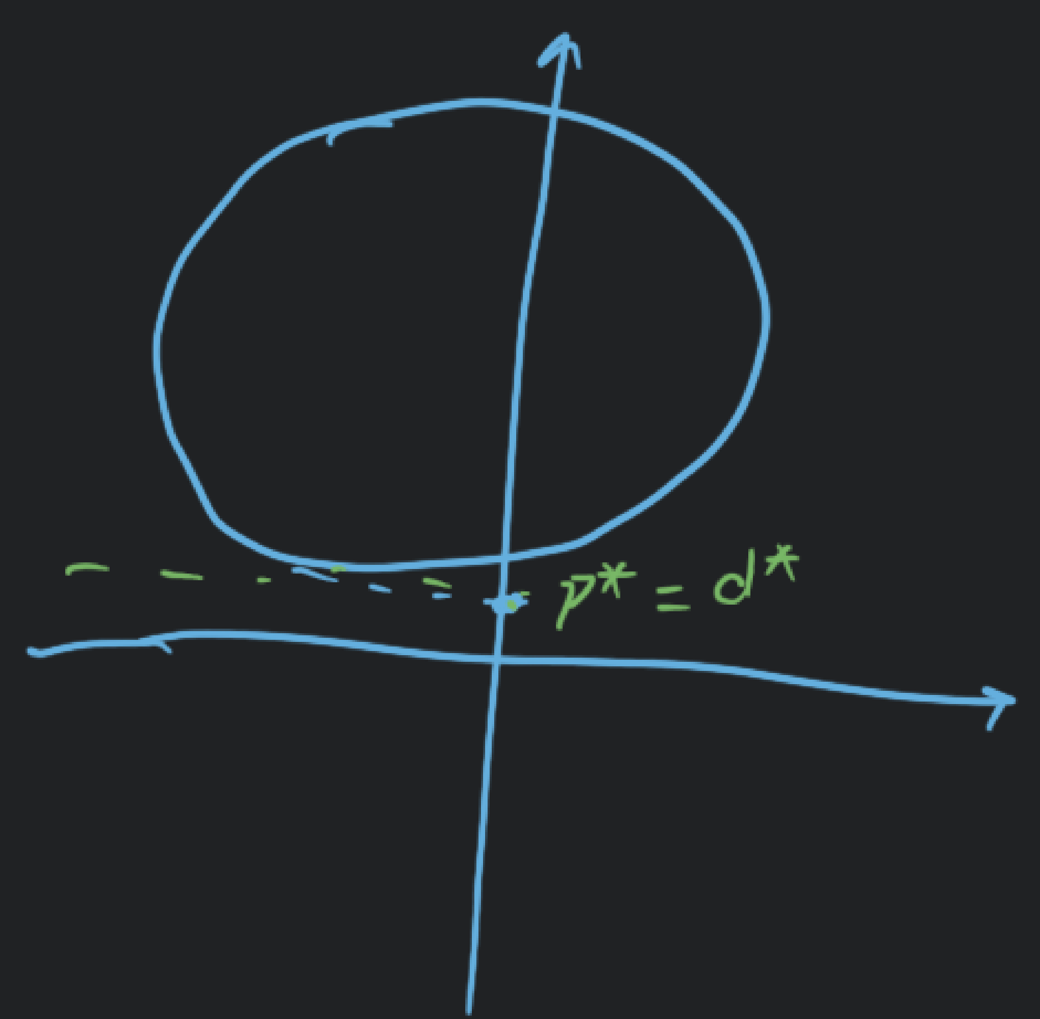
$\Rightarrow$ 对偶问题　$\begin{cases} \min & \frac{1}{2} q^T(\lambda) P^{-1}(\lambda) q(\lambda) - r(\lambda) \\ \text{s.t.} & \lambda \ge 0 \end{cases}$

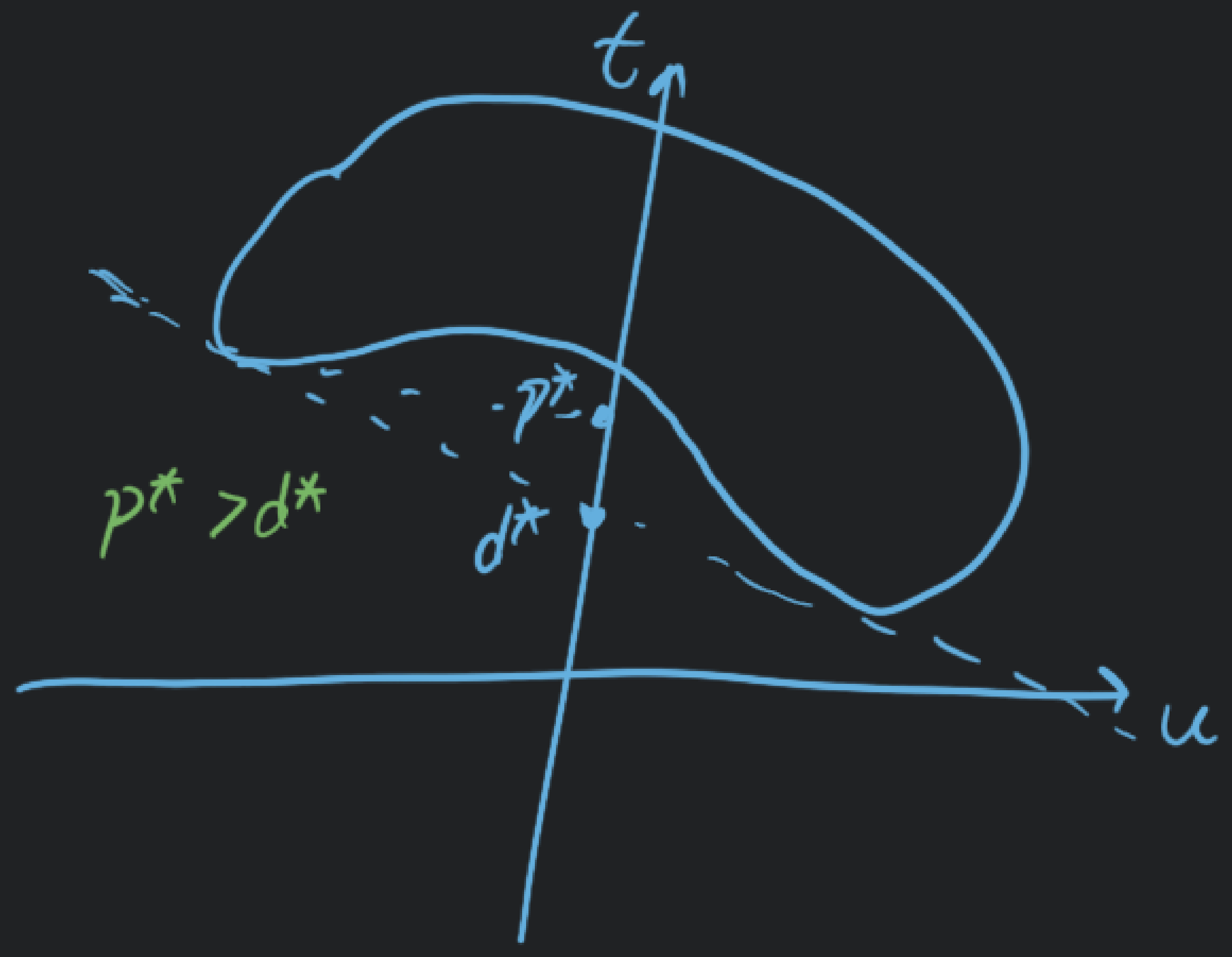
若 $\exists x\in\mathbb{R}^n$, $\frac{1}{2}x^TP_ix + q_i^Tx + r_i < 0$ $(i=1,\cdots,m)$ 则 $p^* = d^*$.

若 $\forall i=1,\cdots,m$ $q_i=0$, $r_i=0$ $\Rightarrow$ $p^*=d^*=0$ 但 Slater条件不满足 (Slater条件不必要).

几何解释 $\min\ f_0(x)$ s.t. $f_1(x)\le 0$ $\quad G \triangleq \{(f_1(x), f_0(x)) \mid x\in D\}$ $\quad p^* = \inf\{t \mid (u,t)\in G,\ u\le 0\}$.

$$g(\lambda) = \inf\{\lambda u + t \mid (u,t)\in G\}.$$





鞍点的解释 $L(x,\lambda)$ 强对偶性即:

$$\inf_{x\in D}\sup_{\lambda\ge 0} L(x,\lambda) = \sup_{\lambda\ge 0}\inf_{x\in D} L(x,\lambda)$$

$$d^* = \sup_{\lambda\ge 0}\inf_{x\in D} L(x,\lambda)$$

$$p^* = \inf_{x\in D}\sup_{\lambda\ge 0} L(x,\lambda)$$

解释: $\min\ f_0(x)$ s.t. $f_i(x)\le 0$ $i=1,\cdots,m$.

$$L(x,\lambda) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x).$$

$$\begin{cases} \forall i=1,\cdots,m\ \ f_i(x)\le 0 & \text{取 } \lambda=0 \\ \exists i\ \ f_i(x)>0 & \lambda_i=+\infty\ \ L=+\infty \end{cases} \xrightarrow[x\in D]{\inf} \begin{array}{l}\min\ f_0(x) \\ \text{s.t. } f_i(x)\le 0\end{array} \text{ 即原问题.}$$

一般情况 $f(w,z)$ $w\in S_w$ $z\in S_z$ f, S_w, S_z 均是一般的,

(min-max 不等式).

$$\sup_{z\in S_z}\inf_{w\in S_w} f(w,z) \le \inf_{w\in S_w}\sup_{z\in S_z} f(w,z)$$

(矮个子里找高个子 ≤ 高个子里找矮个子)

$(\tilde{w},\tilde{z})$称为f的鞍点，如果

$$\arg_{(\tilde{w},\tilde{z})}\{\sup_z\inf_w f(w,z)\} = \arg_{(\tilde{w},\tilde{z})}\{\inf_w\sup_z f(w,z)\}.$$

等价定义：$f(\tilde{w},z) \le f(\tilde{w},\tilde{z}) \le f(w,\tilde{z}) \quad \forall z\in S_z,\ \forall w\in S_w$ $\quad \inf\limits_{w}\sup\limits_{z} f(w,z) \ge \sup\limits_{z}\inf\limits_{w} f(w,z)$.

$\Rightarrow f(\tilde{w},\tilde{z}) = \sup\limits_{z} f(\tilde{w},z) \ge \inf\limits_{w}\sup\limits_{z} f(w,z) \quad f(\tilde{w},\tilde{z}) = \inf\limits_{w} f(w,\tilde{z}) \le \sup\limits_{z}\inf\limits_{w} f(w,z) \Rightarrow \inf\limits_{w}\sup\limits_{z} f(w,z) = \sup\limits_{z}\inf\limits_{w} f(w,z)$ 且在$(\tilde{w},\tilde{z})$取到.

鞍点定理：$(\tilde{x},\tilde{\lambda})$为$L(x,\lambda)$鞍点 $\Longleftrightarrow$ 强对偶存在，且 $\tilde{x},\tilde{\lambda}$ 为原问题与对偶问题最优解.

"$\Leftarrow$" $(\tilde{x},\tilde{\lambda})$可行 $f_i(\tilde{x}) \le 0 \ (i=1,\cdots,m)$ $\tilde{\lambda} \ge 0$. $f_0(\tilde{x}) = \inf\limits_{x}\{f_0(x) + \sum\limits_{i=1}^{m}\tilde{\lambda}_i f_i(x)\} \le f_0(\tilde{x}) + \sum\limits_{i=1}^{m}\tilde{\lambda}_i f_i(\tilde{x})$
$\le f_0(\tilde{x}) \Rightarrow$ 不等号均取等

$\Rightarrow \inf\limits_{x} L(x,\tilde{\lambda}) = L(\tilde{x},\tilde{\lambda}) \qquad \sup\limits_{\lambda\ge 0} L(\tilde{x},\lambda) = \sup\limits_{\lambda\ge 0}\{f_0(\tilde{x}) + \sum\limits_{i=1}^{m}\lambda_i f_i(\tilde{x})\} = f_0(\tilde{x}) = L(\tilde{x},\tilde{\lambda})$

$\Rightarrow \inf\limits_{x} L(x,\tilde{\lambda}) = \sup\limits_{\lambda\ge 0} L(\tilde{x},\lambda)$ 即$(\tilde{x},\tilde{\lambda})$为鞍点.

假设所有目标和约束可微，强对偶性成立. (x^*,λ^*,ν^*)为最优解 $\quad f_i(x^*) \le 0 \quad h_i(x^*) = 0 \quad \lambda^* \ge 0$

$$f_0(x^*) = \inf_x\{f_0(x) + \sum_{i=1}^{m}\lambda_i^* f_i(x) + \sum_{i=1}^{p}\nu_i^* h_i(x)\} \le f_0(x^*) + \sum_{i=1}^{m}\lambda_i^* f_i(x^*) + \sum_{i=1}^{p}\nu_i^* h_i(x^*) \le f_0(x^*)$$

$\Rightarrow$ 不等号均取等 $\quad \sum\limits_{i=1}^{m}\lambda_i^* f_i(x^*) = 0 \quad \lambda_i^* f_i(x^*) \le 0 \Rightarrow \forall i=1,\cdots,m \quad \lambda_i^* f_i(x^*) = 0 \Rightarrow \begin{cases} \lambda_i^* > 0 \Rightarrow f_i(x^*) = 0 \\ f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0 \end{cases}$

互补松弛条件

$\inf\limits_{x} L(x,\lambda^*,\nu^*) = L(x^*,\lambda^*,\nu^*) \Rightarrow \left.\dfrac{\partial L(x,\lambda^*,\nu^*)}{\partial x}\right|_{x=x^*} = 0$

$\Rightarrow \nabla f_0(x^*) + \sum\limits_{i=1}^{m}\lambda_i^* \nabla f_i(x^*) + \sum\limits_{i=1}^{p}\nu_i^* \nabla h_i(x^*) = 0$ (稳定性).

可行性：$f_i(x^*) \le 0 \ (i=1,\cdots,m) \quad h_i(x^*) = 0 \ (i=1,\cdots,p) \quad \lambda_i^* \ge 0 \ (i=1,\cdots,m)$.

以上条件统称 KKT条件. (是最优解必要条件，未必充分).

若原问题为凸问题，各函数可微，强对偶成立，则KKT条件为充要条件.

证明充分性：只需证 $g(\tilde\lambda,\tilde\nu)=f_0(\tilde x)$，即 $(\tilde x,\tilde\lambda,\tilde\nu)$ 是鞍点，由鞍点定理可知 $(\tilde x,\tilde\lambda,\tilde\nu)$ 最优.

$f_i(x^*)\le 0$ $\quad h_i(x^*)=0$ $\quad \lambda_i^*\ge 0$ $\quad \lambda_i f_i(x^*)=0$ $\quad L(x,\tilde\lambda,\tilde\nu)=f_0(x)+\sum_{i=1}^{m}\tilde\lambda_i f_i(x)+\sum_{i=1}^{p}\tilde\nu_i h_i(x)$ 是关于x的凸函数.

于是 $\frac{\partial L(x,\tilde\lambda,\tilde\nu)}{\partial x}\Big|_{x=\tilde x}=0$ 得到 $\tilde x$ 是最优解 $\Rightarrow g(\tilde\lambda,\tilde\nu)=\inf_x L(x,\tilde\lambda,\tilde\nu)=L(\tilde x,\tilde\lambda,\tilde\nu)=f_0(\tilde x)$ #

---

e.g. $\min \frac{1}{2}x^TPx+q^Tx+r$ $\quad P\in S_+^n$ $\quad$ s.t. $Ax=b$.

KKT: $Ax^*=b$ $\quad \frac{\partial}{\partial x}\left[\frac{1}{2}x^TPx+q^Tx+r+(Ax-b)^T\nu^*\right]\Big|_{x=x^*}=0 \Rightarrow P^Tx^*+q+\nu^{*T}A=0$

$$\Rightarrow \begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x^* \\ \nu^* \end{bmatrix}=\begin{bmatrix} -q \\ b \end{bmatrix}$$ 由此解出最优解 x^*.

e.g. water-filling 问题 $\quad \min_x -\sum_{i=1}^{n}\log(\alpha_i+x_i)$ $\quad$ s.t $x\succeq 0$ $\quad 1^Tx=1$ $\quad x\in\mathbb{R}^n$ $\quad \alpha\in\mathbb{R}^n$ $\quad \alpha\succeq 0$

KKT. $x^*\succeq 0$ $\quad 1^Tx^*=1$ $\quad \lambda^*\succeq 0$ $\quad \lambda_i^*x_i^*=0$ $(i=1,\cdots,n)$. $\quad -\frac{1}{\alpha_i+x_i^*}-\lambda_i^*+\nu^*=0$ $(i=1,\cdots,m)$

$\Rightarrow x_i^*\left(\nu^*-\frac{1}{\alpha_i+x_i^*}\right)=0$ $\quad \nu^*\ge\frac{1}{\alpha_i+x_i^*}$ $\quad$ 若 $\nu^*\ge\frac{1}{\alpha_i}\Rightarrow x_i^*=0$ $\quad$ 若 $\nu^*<\frac{1}{\alpha_i}\Rightarrow x_i^*>0,\ x_i^*=\frac{1}{\nu^*}-\alpha_i$.

注水算法,

$\frac{1}{\nu}\uparrow$ 直到 $\sum_{i=1}^{n}\mathrm{Relu}\left(\frac{1}{\nu}-\alpha_i\right)=1 \Rightarrow$ 解出 ν^*, x_i^*

e.g. $\min f_0(x)$ $\quad$ s.t. $x\succeq 0$ $\quad f_0(x)$ 凸函数 $\quad$ KKT: $x^*\succeq 0$ $\quad [\nabla f_0(x^*)]_i-\lambda_i^*=0$ $\quad \lambda_i^*\ge 0$ $\quad \lambda_i^*x_i^*=0$

$\Leftrightarrow \nabla f_0(x^*)\succeq 0$ $\quad x_i^*(\nabla f_0(x^*))_i=0$ $\quad x^*\succeq 0$

e.g. $\min f_0(Ax+b)$. 不能使用KKT. 变形: $\min f_0(y)$ s.t. $Ax+b=y$. $\Rightarrow L(x,y,\lambda) = f_0(y) + v^T(Ax+b-y)$.

$$\Rightarrow g(v) = \inf_{x,y} L(x,y,v) = \begin{cases} -f_0^*(v) + v^Tb & v^TA=0 \\ -\infty & v^TA\neq 0 \end{cases} \Rightarrow \text{共轭问题} \quad \max\ v^Tb - f_0^*(v) \quad \text{s.t. } v^TA=0$$

e.g. $\min \|Ax-b\| \Leftrightarrow \min \|y\|$ s.t. $y=Ax-b \Leftrightarrow \max -v^Tb - \|v\|^*$ s.t. $v^TA=0 \Leftrightarrow \max v^Tb - \|v\|^2$ s.t. $v^TA=0$

e.g. 带框约束的线性规划 $\min c^Tx$ s.t. $Ax=b$ $l\le x\le u$ $(l_i\le x_i\le u_i, \forall i)$.

$$\Rightarrow L(x,\lambda_1,\lambda_2,v) = c^Tx + v^T(Ax-b) + \lambda_1^T(l-x) + \lambda_2^T(x-u) = (c+A^Tv-\lambda_1+\lambda_2)^Tx - v^Tb + \lambda_1^Tl - \lambda_2^Tu$$

$$\Rightarrow g(\lambda_1,\lambda_2,v) = \begin{cases} -v^Tb + \lambda_1^Tl - \lambda_2^Tu & c+A^Tv-\lambda_1+\lambda_2=0 \\ -\infty & c+A^Tv-\lambda_1+\lambda_2\neq 0 \end{cases} \Rightarrow \begin{aligned} \max\ & -b^Tv + l^T\lambda_1 - u^T\lambda_2 \\ \text{s.t. } & c+A^Tv-\lambda_1+\lambda_2=0,\ \lambda_1\ge 0,\ \lambda_2\ge 0 \end{aligned}$$

等价问题: $\min f_0(x)$ s.t. $Ax=b$ 其中 $f_0(x) = \begin{cases} c^Tx & l\le x\le u \\ +\infty & \text{otherwise} \end{cases} \Rightarrow L(x,v) = f_0(x) + v^T(Ax-b)$

$$g(v) = \inf_x f_0(x) + v^TAx - v^Tb = \inf_{l\le x\le u} (A^Tv+c)^Tx - v^Tb \qquad x \triangleq x^+ - x^- \quad x^+, x^- \ge 0.$$

$$\Rightarrow g(v) = -v^Tb + l^T(A^Tv+c)^+ - u^T(A^Tv+c)^- \qquad \max g(v)$$

敏感性分析

原问题 $\begin{cases} \min & f_0(x) \\ \text{s.t.} & f_i(x)\le 0 \quad i=1,\cdots,m \\ & h_i(x)=0 \quad i=1,\cdots,p \end{cases}$

干扰问题 $\begin{cases} \min & f_0(x) \\ \text{s.t.} & f_i(x)\le u_i \quad i=1,\cdots,m \\ & h_i(x)=w_i \quad i=1,\cdots,p. \end{cases}$

最优解 $p^*(u,w)$

$p^*(0,0) = p^*$ 原问题最优解

性质1. 若原问题为凸问题，则 $p^*(u,w)$ 为 (u,w) 的凸函数.

prove: $p^*(u,w) = \inf_x \{ f_0(x) \mid f_i(x_i) \le u_i,\ h_i(x_i) = w_i \} = \inf_x g(x,u,w)$.

$g(x,u,w) \triangleq f_0(x)$ $\quad \mathrm{dom}\, g = (\mathrm{dom} f_0 \cap \{x \mid f_i(x_i) \le u_i,\ h_i(x_i) = w_i\}, \mathbb{R}^m, \mathbb{R}^p)$ 是凸集.

$\Rightarrow g(x,u,w)$ 是 (x,u,w) 的凸函数 $\quad p^*(u,w)$ 是 (u,w) 的凸函数.

$\forall (x,u_1,w_1), (x,u_2,w_2) \in \mathrm{dom}\, g$, $\forall \theta \in [0,1]$, $p^*(\theta u_1 + (1-\theta)u_2, \theta w_1 + (1-\theta) w_2) = \inf_x g(x, \theta u_1 + (1-\theta)u_2, \theta w_1 + (1-\theta) w_2)$

$= \inf_x \left[\theta g(x,u_1,w_1) + (1-\theta) g(x,u_2,w_2)\right] = \theta p^*(u_1,w_1) + (1-\theta) p^*(u_2,w_2)$. $\Rightarrow p^*(u,v)$ 是凸函数.

性质2. 若原问题为凸问题，强对偶，λ^*, v^* 为原问题对偶问题最优解. $\Longrightarrow$ $p^*(u,w) \ge p^*(0,0) - \lambda^{*T} u - v^{*T} w$

prove: 设 $\tilde{x}$ 为干扰问题最优解，$f_i(\tilde{x}) \le u_i$ $\quad h_i(\tilde{x}) = w_i$. $\quad p^*(0,0) = g(\lambda^*, v^*) \le f_0(\tilde{x}) + \sum_{i=1}^m \lambda_i^* f_i(\tilde{x}) + \sum_{i=1}^p v_i^* h_i(\tilde{x})$

$\Rightarrow p^*(0,0) \le f_0(\tilde{x}) + \lambda^{*T} u + v^{*T} w = p^*(u,w) + \lambda^{*T} u + v^{*T} w$. #

(1) 若 $\lambda_i^* \gg 1$，且加紧第 i 次不等式约束 $u_i < 0$ $\quad \Rightarrow p^*(u,w)$ 急剧增加.

(2) 若 $v^* \gg 1$，且 $w_i < 0$ $\quad$ 或 $v^* \ll -1$ 且 $w_i > 0$ $\quad \Rightarrow p^*(u,w)$ 急剧增加

性质3. (局部敏感性) 若原问题为凸问题，强对偶，且 $p^*(u,w)$ 在 $(u,w) = (0,0)$ 处可微

$$\lambda_i^* = -\frac{\partial p^*(0,0)}{\partial u_i} \qquad v_i^* = -\frac{\partial p^*(0,0)}{\partial w_i}$$

对偶与算法

e.g. Boolen LP 问题 $\quad \min c^T x \quad$ s.t. $Ax \le b \quad x_i \in \{0,1\}$ $\quad i = 1, \cdots, n$ $\quad$ 难以求解

$x_i \in \{0,1\} \Leftrightarrow x_i(x_i - 1) = 0$

$\Rightarrow L(x,\lambda,v) = c^Tx + \lambda^T(Ax-b) + \sum_{i=1}^n v_i x_i^2 - \sum_{i=1}^n v_i x_i = \sum_{i=1}^n v_i x_i^2 + (c+A^T\lambda - v)^T x - \lambda^T b.$

$\Rightarrow g(\lambda,v) = \inf_x L(x,\lambda,v) = \begin{cases} -\lambda^T b - \frac{1}{4}\sum_{i=1}^n (c_i + a_i^T\lambda - v_i)^2/v_i & v\geq 0 \\ -\infty & \text{otherwise} \end{cases}$ 其中 $A \triangleq (a_1 \cdots a_i \cdots a_n)$

$\min_{\lambda,v} \lambda^T b + \frac{1}{4}\sum_{i=1}^n (c_i + a_i^T\lambda - v_i)^2/v_i \quad s.t.\ \lambda\geq 0 \quad v\geq 0$

先求 $\min_v$，$\beta_i \triangleq c_i + a_i^T\lambda$ $\quad \min_{v_i} \frac{(\beta_i - v_i)^2}{4v_i} = \begin{cases} -\beta_i & \beta_i \leq 0 \\ 0 & \beta_i > 0 \end{cases} = \min\{\beta_i, 0\}.$

$\Leftrightarrow \min_\lambda \lambda^T b - \sum_{i=1}^n \min\{0, c_i + a_i^T\lambda\} \quad s.t.\ \lambda\geq 0$

$\Leftrightarrow \min_\lambda \lambda^T b - 1^T W \quad s.t.\ \lambda\geq 0 \quad w_i \leq 0 \quad w_i \leq a_i^T\lambda + c_i.$ → 拉格朗日松弛，弱对偶性.

Boolen LP 松弛问题 $\min c^Tx \quad s.t.\ Ax\leq b \quad 0\leq x_i\leq 1 \quad (i=1,\cdots,n).$

$L(x,u,v,w) = c^Tx + u^T(Ax-b) - v^Tx + w^T(x-1) = (c+A^Tu - v + w)^T x - b^Tu - 1^Tw.$

$\Rightarrow \min b^Tu + 1^Tw \quad s.t.\ c + A^Tu - v + w = 0 \quad u\geq 0 \quad v\geq 0 \quad w\geq 0.$

$\Leftrightarrow \min b^Tu + 1^Tw \quad s.t.\ c + A^Tu + w \geq 0 \quad u\geq 0 \quad w\geq 0$

二者等价

拉格朗日松弛与Boolen LP 松弛互为对偶.

e.g. 带等式约束的可微凸优化问题. $\min f_0(x) \ s.t.\ Ax-b=0 \xrightarrow{\text{罚函数}} \min f_0(x) + \frac{\alpha}{2}\|Ax-b\|_2^2 \quad \alpha > 0$

$\tilde{x} = \arg\min_x f_0(x) + \frac{\alpha}{2}\|Ax-b\|_2^2 \Rightarrow \nabla f_0(\tilde{x}) + \alpha A^T(A\tilde{x}-b) = 0 \Rightarrow \tilde{x} = \arg\min_x f_0(x) + \alpha(A\tilde{x}-b)^T(Ax-b)$

原问题 $L(x,v) = f_0(x) + v^T(Ax-b) \quad g(v) = \inf_x\{f_0(x) + v^T(Ax-b)\}$ 取 $v = \alpha(A\tilde{x}-b)$ 即罚函数问题

$\Rightarrow g(\alpha(A\tilde{x}-b)) = f_0(\tilde{x}) + \alpha\|A\tilde{x}-b\|_2^2$

$$\underbrace{f_0(x^*) = p^* = d^*}_{\text{原问题}} \geq g(\alpha(A\tilde{x}-b)) = \underbrace{f_0(\tilde{x}) + \alpha\|A\tilde{x}-b\|_2^2}_{\text{罚函数解}} \geq f_0(\tilde{x})$$

$$\begin{cases} \alpha = 0 & \arg\min f_0(x) \\ \alpha \to +\infty & \text{原问题} \quad f_0(x^*) = f_0(\tilde{x}) \end{cases}$$

罚函数相当于对原问题做了松弛.

e.g. 带线性不等式约束的可微凸优化问题.

$\min f_0(x) \quad s.t.\ Ax \geq b$ 罚问题: log-barrier $\min f_0(x) - \sum_{i=1}^{m} u \log(a_i^T x - b_i)$ 凸问题 $u \geq 0$

设 $\tilde{x}$ 为罚问题最优解. $\nabla f_0(\tilde{x}) - \sum_{i=1}^{m} u_i \dfrac{a_i}{a_i^T\tilde{x} - b_i} = 0 \Rightarrow \tilde{x} = \arg\min_x f_0(x) - \sum_{i=1}^{m} u \dfrac{a_i^T x - b_i}{a_i^T \tilde{x} - b_i}$.

原问题: $g(\lambda) = \inf_x \{f_0(x) + \sum_{i=1}^{m} \lambda_i (b_i - a_i^T x)\}$ 取 $\lambda_i = \dfrac{u}{a_i^T\tilde{x} - b_i}$ 即上述罚问题最优解.

内点法: 在约束范围内搜索, 且离约束边界越远越好.

原问题 ↗ 惩罚问题 → 惩罚因子

$\Updownarrow$ -- 对应

原问题 ↘ 对偶函数 → 拉格朗日乘子

6. 优化算法

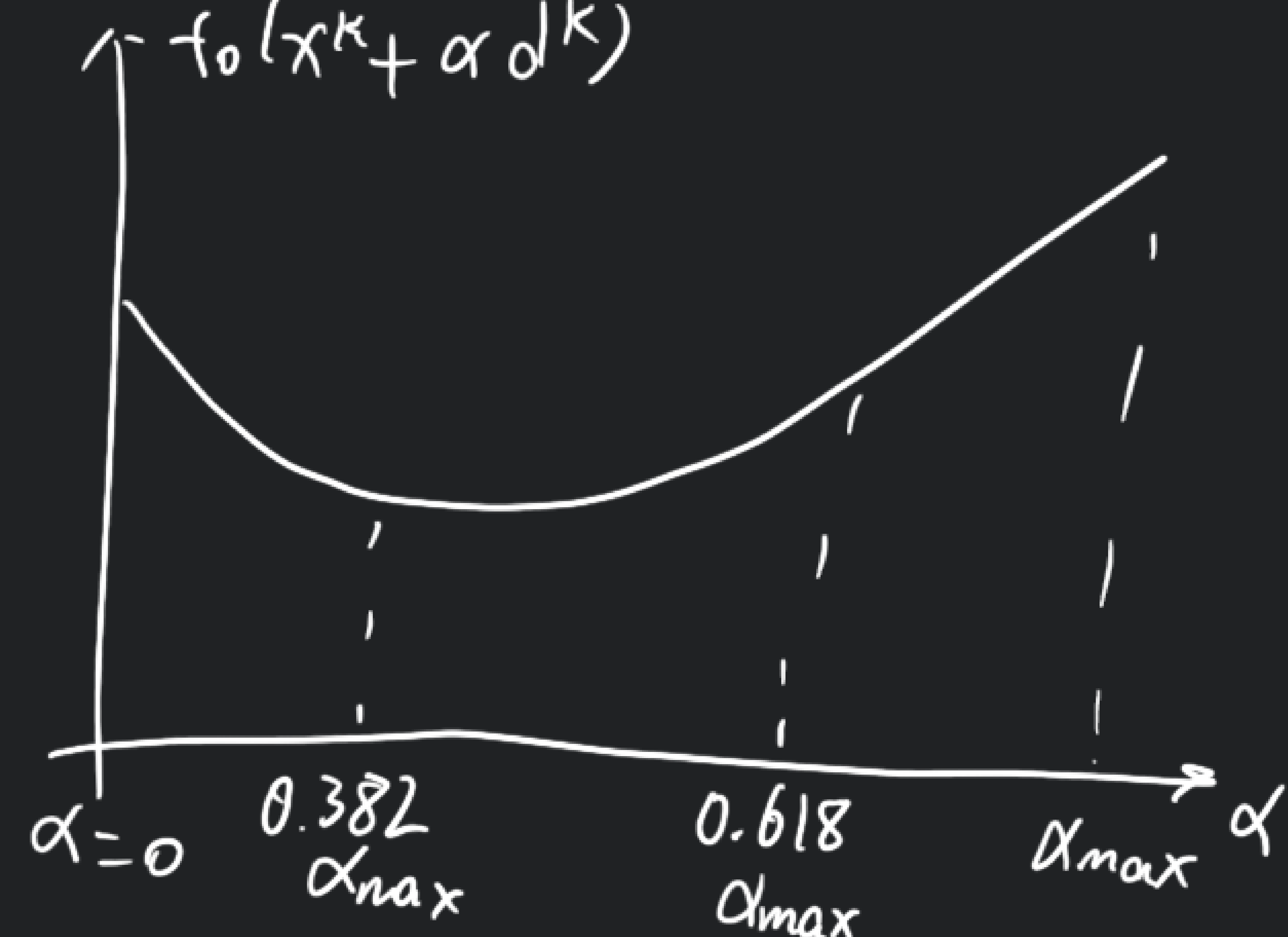
无约束优化问题

所有优化算法都是迭代算法 $x^{k+1} = x^k + \alpha^k d^k$ $\begin{cases} d^k: \text{方向} & \dim d^k = \dim x^k \\ \alpha^k: \text{步长} & \alpha^k \in R \end{cases}$

$\alpha^k = \arg\min_{\alpha \geq 0} f_0(x^k + \alpha d^k)$: 一维优化问题

原问题凸问题 ⇒ 步长问题也是凸问题. Line Search.

(1) 迭代算法　黄金分割法 (优选法)

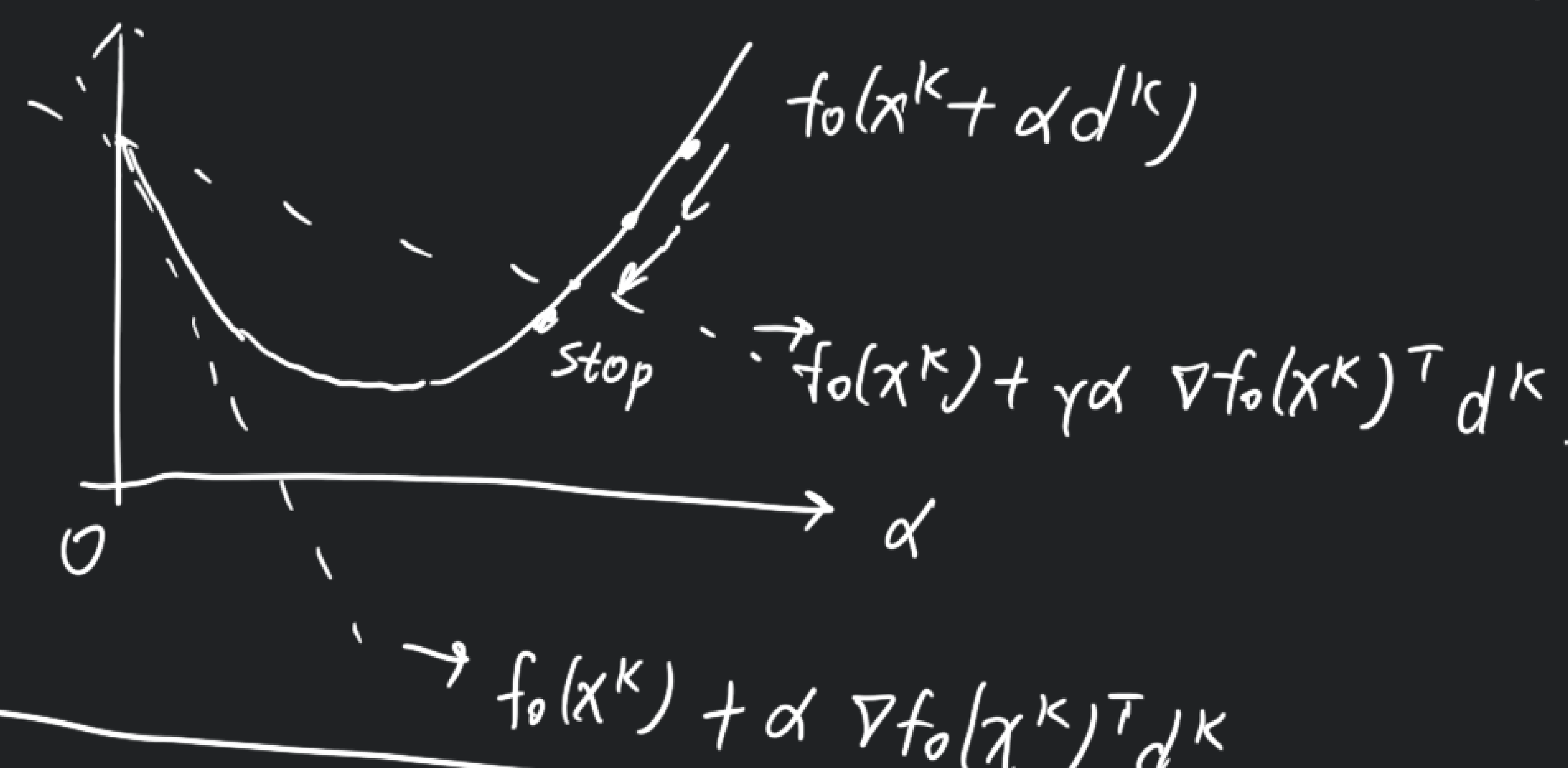


$f(0.382\alpha_{max}) > f(0.618\alpha_{max})$: 继续在 $[0.382\alpha_{max},\ \alpha_{max}]$ 迭代

$f(0.382\alpha_{max}) < f(0.618\alpha_{max})$: 继续在 $[0,\ 0.618\alpha_{max}]$ 迭代

(2) Armijo Rule

若 $f_0(x^k+\alpha d^k) > f(x^k) + \gamma\alpha \nabla f_0(x^k)^T d^k$　$0<\gamma<0.5$　则 $\alpha \Leftarrow \alpha\beta$　$0<\beta<1$　否则停止



可找到 α^* 的次优解.

(找到一个还可以的 α 就足够了)

强凸性　$\exists m>0$　$f(x)-\frac{1}{2}mx^Tx$ 是凸函数　　强凸 $\Rightarrow$ 凸

等价定义: (1) 若 $f(x)$ 可微　$\forall x,y\in \mathrm{dom}f$　$f(y)\geq f(x)+\nabla f(x)^T(y-x)+\frac{1}{2}m\|y-x\|_2^2$

$f(x)-\frac{1}{2}mx^Tx$ 凸 $\Leftrightarrow f(y)-\frac{1}{2}my^Ty \geq f(x)-\frac{1}{2}mx^Tx + \nabla f(x)^T(y-x) - mx^T(y-x)$.

$\Leftrightarrow f(y)\geq f(x)+\nabla f(x)^T(y-x)+\frac{1}{2}m(y^Ty-x^Tx-2x^Ty+2x^Tx) = f(x)+\nabla f(x)^T(y-x)+\frac{1}{2}m\|y-x\|_2^2$

(2) 若 $f(x)$ 二阶可微 $\exists m>0$ $\forall x\in \mathrm{dom} f$ $\nabla^2 f(x)\succeq mI$ (海森矩阵下界)

$\min f(x)$ $f(x)$ 凸, 可微 $\Rightarrow$ 解 $\nabla f(x)=0$ 迭代

$\nabla f(x)\to 0$ $f(x)\xrightarrow{?} f(x^*)$ 假设 $f(x)$ 强凸 $f(x)+\nabla f(x)^T(y-x)+\frac{1}{2}m\|y-x\|_2^2$ 是 y 的凸函数

最优 y : $\nabla f(x)+m(\tilde{y}-x)=0$ $\tilde{y}=x-\frac{\nabla f(x)}{m}$ $f(\tilde{y})\geq f(x)-\frac{1}{2m}\|\nabla f(x)\|_2^2$

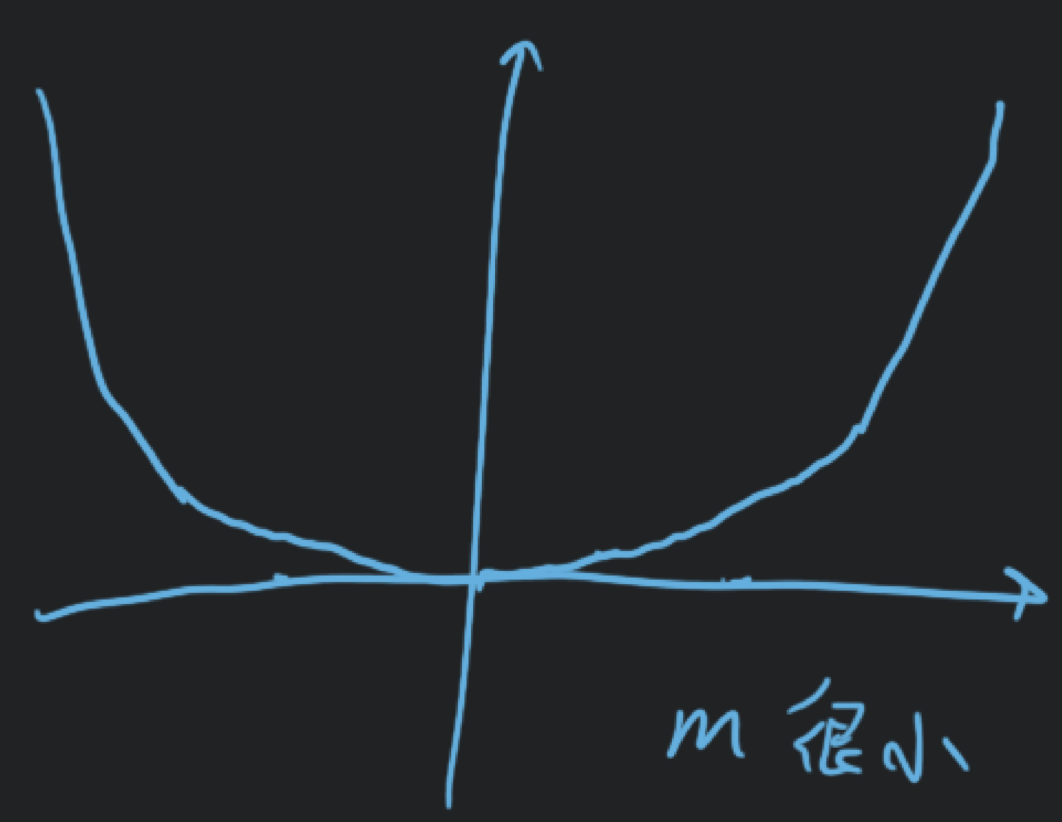
$\forall x,y$ $f(y)\geq f(x)-\frac{1}{2m}\|\nabla f(x)\|_2^2$ 取 $y=x^*=\underset{x}{\operatorname{argmin}}\, f(x)$ $\Rightarrow p^*\geq f(x)-\frac{1}{2m}\|\nabla f(x)\|_2^2$

$\Rightarrow \forall x$, $p^*\leq f(x)\leq p^*+\frac{1}{2m}\|\nabla f(x)\|_2^2$ $\|f(x)-p^*\|_2\leq\frac{1}{2m}\|\nabla f(x)\|_2^2$

$\nabla f(x)\to 0$, $x\xrightarrow{?} x^*$ $f(x)$ 强凸 取 $y=x^*$ $p^*\geq f(x)+\nabla f(x)^T(x^*-x)+\frac{m}{2}\|x^*-x\|_2^2$

$\langle a,b\rangle+\|a\|\|b\|\geq 0$ $\Rightarrow$ $f(x)\geq p^*\geq f(x)-\|\nabla f(x)\|_2\|x^*-x\|_2+\frac{m}{2}\|x^*-x\|_2^2$

$\Rightarrow -\|\nabla f(x)\|_2\|x^*-x\|_2+\frac{m}{2}\|x^*-x\|_2^2\leq 0$ $\|x^*-x\|_2\leq\frac{2}{m}\|\nabla f(x)\|_2$



$\|\nabla f(x)\|$ 给定, $m\downarrow$ 误差 $\uparrow$

$m\uparrow$ 误差 $\downarrow$

海森矩阵上界 $\exists M > 0 \quad \forall x \in \mathrm{dom} f,\ \nabla^2 f(x) \preceq MI$

性质 (1) $\forall x, y \in \mathrm{dom} f,\quad f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{M}{2} \|y-x\|_2^2$ （$f(x) - Mx^T x$ 是凹函数）

(2) $\|f(x) - p^*\|_2 \geqslant \frac{1}{2M} \|f(x)\|_2^2$

梯度下降法 $d^k = -\nabla f(x^k)$

Repeat $\alpha^k = \underset{0 \le \alpha \le \alpha_{max}}{\operatorname{argmin}}\ f(x^k + \alpha d^k)$

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

Until Convergence

分析算法收敛性 假设 $x \in \mathrm{dom} f \quad MI \succeq \nabla^2 f(x) \succeq mI$

1) 精确线搜索 $\tilde{f}(\alpha) = f(x^k + \alpha d^k) = f(x^k - \alpha \nabla f(x^k)) = f(x^{k+1})$

$\tilde{f}(\alpha) = f(x^{k+1}) \le f(x^k) + \nabla f(x^k)^T (-\alpha \nabla f(x^k)) + \frac{M}{2} \|-\alpha \nabla f(x^k)\|_2^2$

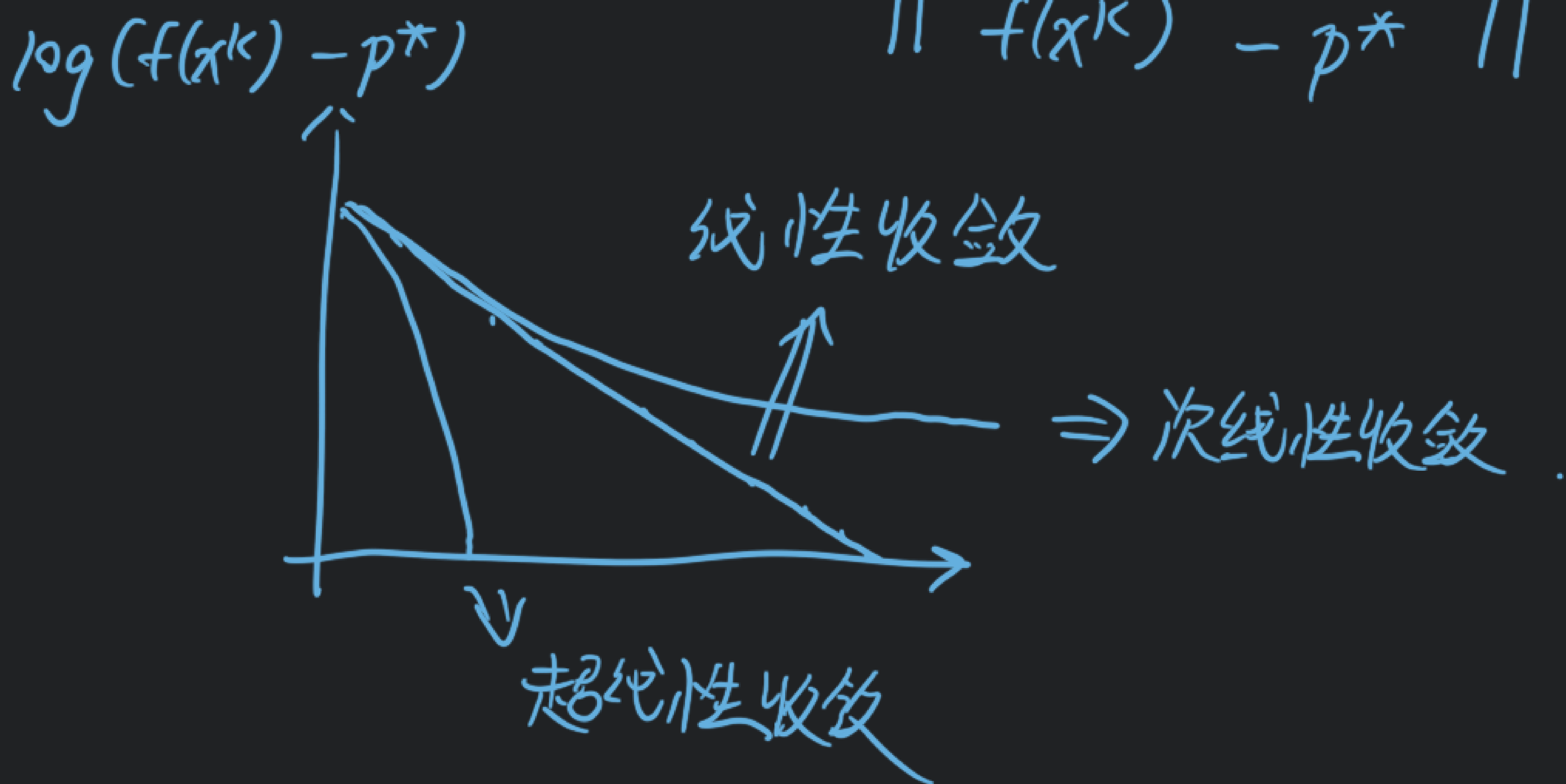
$\Leftrightarrow \quad \tilde{f}(\alpha) \le f(x^k) + \|\nabla f(x^k)\|_2^2 \left(\frac{M\alpha^2}{2} - \alpha\right)$

$\Rightarrow \quad \min_{\alpha} \tilde{f}(\alpha) \le f(x^k) - \frac{1}{2M} \|\nabla f(x^k)\|_2^2$

对精确线搜索， $\alpha = \underset{\alpha}{\operatorname{argmin}}\ \tilde{f}(\alpha)$. $\Rightarrow f(x^{k+1}) \le f(x^k) - \frac{1}{2M} \|\nabla f(x^k)\|_2^2$

又 $p^* \ge f(x^k) - \frac{1}{2m}\|\nabla f(x^k)\|_2^2 \Rightarrow f(x^{k+1}) - p^* \le (1-\frac{m}{M})(f(x^k)-p^*)$. $M>m \Rightarrow$ 收敛

给定允许误差 ε，何时 $\left\|\frac{f(x^{k+\tau})-p^*}{f(x^k)-p^*}\right\| \le \varepsilon$. $(1-\frac{m}{M})^\tau = \varepsilon \Rightarrow \tau = \frac{\log\varepsilon}{\log(1-\frac{m}{M})}$



梯度下降（精确线搜索）至少是线性收敛的.

2) 非精确线搜索 (Armijo Rule)

当 $0\le\alpha\le\frac{1}{M}$ 时，迭代必然停止　Armijo Rule: $\tilde f(\alpha) = f(x^{k+1}) \le f(x^k) + \gamma\alpha\nabla f(x^k)^T d^k$ 时接收 α. $(0<\gamma<\frac{1}{2})$

当 $0\le\alpha\le\frac{1}{M}$ 时，$(-\alpha+\frac{M\alpha^2}{2}) \le -\frac{\alpha}{2}$ 又 $\tilde f(\alpha) = f(x^{k+1}) \le f(x^k) - \alpha\|\nabla f(x^k)\|_2^2 + \frac{M\alpha^2}{2}\|\nabla f(x^k)\|_2^2$

$\Rightarrow \tilde f(\alpha) \le f(x^k) - \frac{\alpha}{2}\|\nabla f(x^k)\|_2^2 \le f(x^k) - \gamma\alpha\|\nabla f(x^k)\|_2^2 = f(x^k)+\gamma\alpha\nabla f(x^k)^T d^k$

$\Rightarrow \alpha_{inexect} = \alpha_{max}$ 或 $\ge \frac{\beta}{M}$. （要么是初始的 α，要么 $\ge \frac{\beta}{M}$，若 $\alpha_{inexect} < \frac{\beta}{M}$，上一步 $\frac{\alpha_{inexect}}{\beta} < \frac{1}{M}$ 已经停止）

$\tilde f(\alpha_{exect}) \le f(x^k) - \frac{1}{2M}\|\nabla f(x^k)\|_2^2$　非精确：$f(x^{k+1}) \le f(x) - \min\{\gamma\alpha_{max}, \frac{\gamma\beta}{M}\}\|\nabla f(x)\|_2^2$

$\Rightarrow \frac{f(x^{k+1})-p^*}{f(x^k)-p^*} \le 1-\min\{2m\gamma\alpha_{max}, \frac{2m\gamma\beta}{M}\}$. $0<\frac{2m\gamma\beta}{M}<1 \Rightarrow 0<\frac{f(x^{k+1})-p^*}{f(x^k)-p^*}<1 \Rightarrow$ Armijo Rule 梯度下降也是收敛的.

e.g. $f(x)=\frac{1}{2}x^TPx \quad P\in S_+^n \quad \nabla^2 f(x)=P \quad \lambda_{max}I\succeq P\succeq \lambda_{min}I \Rightarrow M=\lambda_{max} \quad m=\lambda_{min}$ → P的最大最小特征值.

$\frac{m}{M}=\frac{\lambda_{min}}{\lambda_{max}}\nearrow$，收敛越快 $\quad \frac{m}{M}\downarrow$（病态矩阵，e.g. $P=\begin{pmatrix}100 & \\ & 1\end{pmatrix}$），收敛越慢

$\frac{m}{M}\sim 1$，收敛快

$\frac{m}{M}\ll 1$，收敛慢

最速下降法 Steepest descent （未必是最快的）

$$\min f(x) \longrightarrow \min_v f(x^k)+\nabla f(x^k)^T v \quad s.t.\ \|v\|=1.$$

2-范数：$v^*=-\frac{\nabla f(x^k)}{\|\nabla f(x^k)\|_2}$ $\qquad$ 1-范数：$i\triangleq \arg\max_i |\nabla f(x^k)_i|$ 则 $v_i^*=-\mathrm{sgn}(\nabla f(x^k)_i)$ $\quad v_j^*=0\ (j\neq i)$.

∞-范数：$\|v\|_\infty=|\max v_i|$ $\qquad v_i^*=-\mathrm{sgn}(\nabla f(x^k)_i)\quad (i=1,\cdots,n)$.

变种：(1) 坐标轮换法 $\quad d^k=e_{mod(k,n)}$ $\quad e_i$：单位向量，第 i 个分量为 1，其余分量为 0. $\quad mod(k,n)=k$ 除以 n 的余数

$n=\dim x^k$.

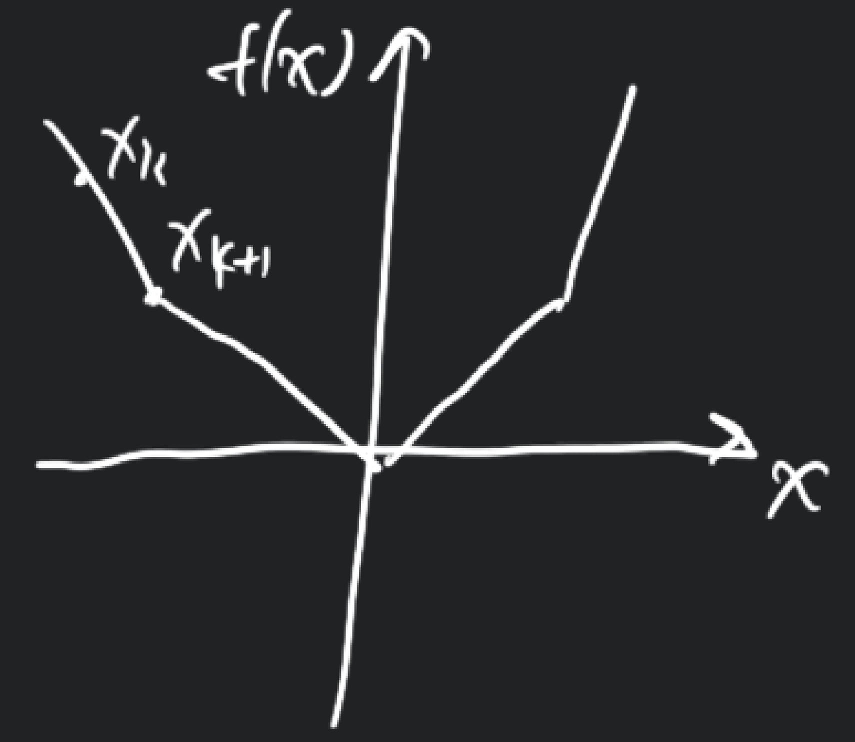
$\alpha\in[-\alpha_{max},\alpha_{max}]$ 中搜索.

(2) 分块坐标轮换法 $\quad \min f(x,y)\quad x\in\mathbb{R}^n\quad y\in\mathbb{R}^m$ $\quad$ 固定 y 优化 x → 固定 x 优化 y

$$x^{k+1}=\arg\min_x f(x,y^k) \rightarrow y^{k+1}=\arg\min_y f(x^k,y)$$

要求：$f(x,y)$ 应该是凸的.

(3) 若 $f(x)$ 在某些点不可微



$f(x_{k+1})$ 不可导　次梯度：在左极限和右极限范围任选一个

若 $0 \in$ 次梯度集合，即找到最优解.

牛顿法 Newton's Method

x^k　$d^k = \underset{v}{\operatorname{argmin}} \{ f(x^k + v) \mid \|v\| = 1 \} \simeq \underset{v}{\operatorname{argmin}} \{ f(x^k) + \nabla f(x^k)^T v + \frac{1}{2} v^T \nabla^2 f(x^k) v \mid \|v\| = 1 \}$

($\nabla f(x)$ 严格凸)

若 $\nabla^2 f(x) \geq mI \ (m > 0)$ 即 $\nabla^2 f(x)$ 正定，可去掉 $\|v\|=1$ 约束，得 $v^* = -(\nabla^2 f(x^k))^{-1} f(x^k)$ (牛顿方向).

牛顿法步骤：

Repeat　$d^k = -(\nabla^2 f(x^k))^{-1} \nabla f(x^k)$

$\alpha^k = \underset{0 \leq \alpha \leq \alpha_{max}}{\operatorname{argmin}} f(x^k + \alpha d^k)$

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

Until　$|\nabla f(x^k)^T (\nabla^2 f(x^k))^{-1} \nabla f(x^k)| \leq \varepsilon$

收敛性分析　$\exists \eta > 0$

1) 若 $\|\nabla f(x^k)\|_2 > \eta$　阻尼牛顿阶段 (距离 x^* 较远)

2) 若 $\|\nabla f(x^k)\|_2 < \eta$　二次收敛阶段　$\frac{f(x^{k+1}) - p^*}{(f(x^k) - p^*)^2} \sim \text{Const}$

拟牛顿法 Quasi-Newton Method

$$\nabla^2 f(x^k) d^k = -\nabla f(x^k) \xrightarrow{\text{替换}} B d^k = -\nabla f(x^k)$$ 计算 $\nabla^2 f(x^k)$ 和逆太麻烦

无约束优化算法 总结：
$\min f(x)$

- 梯度下降法
- 最速下降法
- (分块)坐标轮换法
- 次梯度法
- 牛顿法
- 拟牛顿法

有约束优化问题
线性等式约束

$$\begin{cases} \min f(x) \\ s.t.\ Ax=b \end{cases}$$

KKT条件 $Ax^* = b$

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

1) 线性方程组

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

KKT条件 $Ax^* = b$

$$Px^* + q + A^T v^* = 0$$

$$\Rightarrow \begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -q \\ 0 \end{bmatrix}$$

2) 非线性方程组

$$\min\ f(x^k+d)\quad \text{s.t.}\ A(x^k+d)=b \quad\approx\quad \min\ f(x^k)+\nabla f(x^k)^T d+\frac{1}{2}d^T\nabla^2 f(x^k)d\quad \text{s.t.}\ Ad=0$$

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

$$\begin{cases}\alpha^k=\arg\min\limits_{\alpha\geq 0} f(x^k+\alpha d^k)\\ x^{k+1}=x^k+\alpha^k d^k\end{cases}$$
x^{k+1} 也可行 $(Ax^{k+1}=b)$.

以上称为有约束的牛顿法

拉格朗日法

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

用梯度下降法优化拉格朗日函数

一般取 α^k 为递增且有上限的序列　　实践：收敛太慢

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

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

(x^*,v^*) 为鞍点 $\Rightarrow\begin{cases}x^*=\arg\min\limits_x L(x,v^*)\\ v^*=\arg\max\limits_v L(x^*,v)\end{cases}$

$$\Rightarrow x^{k+1}=x^k+\alpha^k(-\nabla f(x^k)-A^T v^*)\approx x^k+\alpha^k(-\nabla f(x^k)-A^T v^k)$$

$$\Rightarrow v^{k+1}=v^k+\alpha^k(Ax^*-b)\approx v^k+\alpha^k(Ax^k-b)$$

罚函数问题 $\min P(x,v) = \frac{1}{2}\|Ax-b\|_2^2 + \frac{1}{2}\|\nabla f(x) + A^T v\|_2^2$ $\nabla f(x)$ 非线性 $\Rightarrow$ 非凸问题

负梯度方向 $-\nabla P(x,v)|_{(x^k,v^k)} = -\begin{pmatrix} A^T(Ax^k-b) + \nabla^2 f(x^k)(\nabla f(x^k) + A^T v^k) \\ A(\nabla f(x^k) + A^T v^k) \end{pmatrix}$

可以证明：拉格朗日法下降方向与此负梯度方向夹角 $< 90°$（内积 >0）

$$(d^k)^T(-\nabla P(x^k,v^k)) = (\nabla f(x^k) + A^T v^k)^T \nabla^2 f(x^k) (\nabla f(x^k) + A^T v^k) > 0$$

需要：$\nabla^2 f(x^k) > 0$，$\nabla f(x^k) + A^T v^k \neq 0$（即 x^k, v^k 还未到最优解点）

增广拉格朗日法 $L_c(x,v) = f(x) + v^T(Ax-b) + \frac{c}{2}\|Ax-b\|_2^2$ （$\min f(x) + \frac{c}{2}\|Ax-b\|_2^2$ s.t. $Ax=b$ 的拉格朗日函数，与原问题等价，且对偶问题也等价）

$\begin{cases} \nabla_x L(x^*,v^*) = 0 \\ Ax^* = b \end{cases} \Leftrightarrow \begin{cases} \nabla_x L_c(x^*,v^*) = 0 \\ Ax^* = b \end{cases}$ 即两问题的对偶问题也等价.

拉格朗日法 $\Rightarrow \begin{cases} x^{k+1} = x^k - \alpha^k \nabla_x L_c(x^k,v^k) \\ v^{k+1} = v^k + \alpha^k(Ax^k-b) \end{cases}$ 更好的方法 $\begin{cases} x^{k+1} = \arg\min\limits_x L_c(x,v^k) \\ v^{k+1} = v^k + c(Ax^{k+1}-b) \end{cases}$

性质：1) 若 $v=v^*$，则 $\forall c>0$，$x^* = \arg\min\limits_x L_c(x,v^*)$

2) 若 $c\to+\infty$，则 $\forall v$，$x^* = \arg\min\limits_x L_c(x,v)$

e.g. $\min \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2$ s.t. $x_1 = 1$ $\Rightarrow L(x,v) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + v(x_1-1)$ KKT $\Rightarrow \begin{cases} x_1 = 1 \\ x_1 + v = 0 \\ x_2 = 0 \end{cases} \Rightarrow \begin{cases} x_1^* = 1 \\ x_2^* = 0 \\ v^* = -1 \end{cases}$

$L_c(x,v) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + v(x_1-1) + \frac{c}{2}(x_1-1)^2$

若 $v=v^*$，$\arg\min\limits_{x} L_c(x,v^*) = \frac{1}{2}x_1^2+\frac{1}{2}x_2^2-(x_1-1)+\frac{c}{2}(x_1-1)^2 \Rightarrow \begin{cases} x_1-1+c(x_1-1)=0 \\ x_2=0 \end{cases} \Rightarrow \begin{cases} x_1^*=1 \\ x_2^*=0 \end{cases}$

若 $c\to+\infty$，$\nabla_x L_c(x,v^*)=0 \Rightarrow \begin{cases} x_1+v+c(x_1-1)=0 \\ x_2=0 \end{cases} \to x_1^*=\frac{c-v}{c+1}\to 1$

迭代过程：
$$\begin{cases} x^{k+1}=\frac{c-v^k}{c+1} \\ v^{k+1}=v^k+c(x_1^{k+1}-1)=\frac{v^k-c}{c+1} \end{cases} \qquad \frac{v^{k+1}-v^*}{v^k-v^*}=\frac{1}{c+1}<1 \Rightarrow v^k\text{线性收敛到 } v^*$$

e.g. $\min\limits_{x} f(x)+g(x)$ f,g 为凸函数 $\Longleftrightarrow$ $\min f(x)+g(z)$ s.t. $x=z$

$\Rightarrow L_c(x,z,v)=f(x)+g(z)+v^T(x-z)+\frac{c}{2}\|x-z\|_2^2$

1) $\{x^{k+1},z^{k+1}\}=\arg\min\limits_{x,z} f(x)+g(z)+v_k^T(x-z)+\frac{c}{2}\|x-z\|_2^2$

2) $v^{k+1}=v^k+c(x^{k+1}-z^{k+1})$

子循环：$x^{k+1|t+1}=\arg\min\limits_{x} f(x)+\frac{c}{2}\|x-z^{k+1|t}+\frac{v^k}{c}\|_2^2$ $\quad z^{k+1|t+1}=\arg\min\limits_{z} g(z)+\frac{c}{2}\|z-x^{k+1|t+1}-\frac{v^k}{c}\|_2^2$.

交替方向乘子法 $\min \sum_{i=1}^{n} f_i(x) \Longleftrightarrow \min \sum_{i=1}^{n} f_i(x_i)$ s.t. $x_i=z$ $(i=1,\cdots,n)$

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

(1) $\{x_i^{k+1}\}=\arg\min\limits_{\{x_i\}} \sum_{i=1}^{n} f_i(x_i)+\frac{c}{2}\sum_{i=1}^{n}\|x_i-z^k+\frac{v_i}{c}\|_2^2 \Longleftrightarrow x_i^{k+1}=\arg\min\limits_{x_i} f_i(x_i)+\frac{c}{2}\|x_i-z^k+\frac{v_i}{c}\|_2^2$ $\forall i$.

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

(3) $v_i^{k+1}=v^k+c(x_i^{k+1}-z^{k+1})$ $\forall i$.

分布式计算

只需传 x_i, v_i, z.
不需要传函数形式.

7.总结

优化：从一个$\underbrace{\text{可行解集合}}_{\text{约束}}$$\overbrace{\text{寻找}}^{\text{优化算法}}$$\underbrace{\text{最好}}_{\text{目标函数}}$的元素

研究简单约束（凸集） 简单目标（凸函数）$\longrightarrow$简单问题（凸问题）

一. 凸集.

$\{x_1, \cdots, x_k\}$ $\{\theta_1, \cdots, \theta_k\}$. $\theta_1 x_1 + \cdots + \theta_k x_k$

$$\begin{cases} \theta_1 + \cdots + \theta_k = 1 & \text{仿射组合} \\ \theta_1 + \cdots + \theta_k = 1 \quad \theta_1, \cdots, \theta_k \geq 0 & \text{凸组合} \\ \theta_1, \cdots, \theta_k \geq 0 & \text{凸锥组合} \end{cases}$$

组合 $\longrightarrow$ 集

组合 $\longrightarrow$ 包

典型凸集：超平面，半空间，椭球，多面体，单纯形，S^n，S_+^n，S_{++}^n

保凸变换：交集，仿射，透视，反透视，线性分数.

二. 凸函数

等价定义

1. $\text{dom} f$ 凸，$\forall x, y \in \text{dom} f$，$0 \leq \theta \leq 1$，$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$.
2. (切西瓜) $\text{dom} f$ 凸，$\forall x \in \text{dom} f$，$v \in \mathbb{R}^n$ $g(t) = f(x + tv)$ 凸 其中 $\text{dom}\, g = \{t \mid x + tv \in \text{dom} f\}$
3. (一阶条件) $\text{dom} f$ 凸，f 可微，$\forall x, y \in \text{dom} f$ $f(y) \geq f(x) + \nabla f(x)^T (y - x)$
4. (二阶条件) $\text{dom} f$ 凸，f 二阶可微，$\forall x \in \text{dom} f$，$\nabla^2 f(x) \geq 0$

保凸运算：非负加权和，仿射，最大值，复合*，透视，共轭*（一定凸）

拟凸函数
$\begin{cases} 1.\ C_\alpha = \{x \in \mathrm{dom} f \mid f(x) \le \alpha\} \text{ 是凸集} \\ 2.\ \mathrm{dom} f \text{ 凸},\ \forall x, y \in \mathrm{dom} f,\ 0 \le \theta \le 1,\ \max\{f(x), f(y)\} \ge f(\theta x + (1-\theta) y) \\ 3.\ \text{一阶条件，二阶条件} \end{cases}$

对数凸 → 凸 → 拟凸　　凹 → 对数凹

三. 凸问题

定义：$\min f_0(x)$ s.t. $f_i(x) \le 0$ $(i = 1, \cdots, m)$　$Ax = b$　可行域是凸集

性质：全局最优 = 局域最优

可微情况最优解　x^* 最优 $\Longleftrightarrow \forall y \in X$，$\nabla f_0(x)^T (y - x^*) \ge 0$　（X → 可行集）

例：线性规划，二次规划，多目标优化（帕累托最优面）

四. 对偶性

$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$　$x \in D = \left(\bigcap_{i=0}^{m} \mathrm{dom} f_i\right) \cap \left(\bigcap_{i=1}^{p} \mathrm{dom} h_i\right)$　$\lambda \in \mathbb{R}^m$　$\nu \in \mathbb{R}^p$　$\lambda \ge 0$

对偶函数　$g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu)$　一定是凹函数

对偶问题　$\max g(\lambda, \nu)$ s.t. $\lambda_i \ge 0$ $(i = 1, \cdots, m)$　一定是凸问题

弱对偶性　$p^* \ge d^*$　　强对偶性　$p^* = d^*$（充分性：Slater 条件，一般的凸问题都满足）

鞍点解释　$p^* = \inf_{x} \sup_{\lambda \ge 0, \nu} L(x, \lambda, \nu) \longrightarrow (x^*, \lambda^*, \nu^*)$

$$d^* = \sup_{\lambda \geq 0, \nu} \inf_{x} L(x, \lambda, \nu) \longrightarrow \overset{\parallel}{(\tilde{x}^*, \tilde{\lambda}^*, \tilde{\nu}^*)} \longrightarrow \text{鞍点.}$$

鞍点定理：(x^*, λ^*) 是 $L(x, \lambda)$ 鞍点 $\Longleftrightarrow$ 原问题有强对偶性，且 x^* 和 λ^* 分别是原问题和对偶问题最优解.

KKT条件

$$\begin{cases} f_i(x^*) \leq 0 \quad (i=1, \cdots, m) \quad h_i(x^*) = 0 \quad \lambda_i^* \geq 0 \\ \lambda_i^* f_i(x^*) = 0 \quad (i=1, \cdots, m) \\ \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) = 0 \end{cases}$$

是强对偶凸问题的最优解的充要条件

强对偶非凸问题 最优解 必要条件

敏感性分析

五. 优化算法.

无约束
- 梯度下降
- 最速下降
- 坐标轮换
- 牛顿/拟牛顿

有约束
- 牛顿/拟牛顿
- 拉格朗日法
- 增广拉格朗日法
- 交替方向乘子法