凸优化视频课程

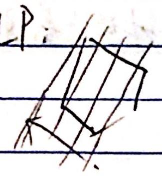
优化 = 数学规划 (Math Programing)

分类方法

线性规划 LP / 非线性

用单纯形法解决 LP

对LP:



是 feasible set

∧是 f_0 下降方向 (梯度反方向)

凸规划 / 非凸规划

所有 f_i 都是凸函数

||

f_0 凸, feasible set 是凸集

卜二于

凸规划一定可以在 多项式时间中得到求解 (但单纯形法不一定可以), 现在用'内点法'作为主要方法

证明: 从仿射集到仿射组合:

有仿射集 $x_1, x_2, x_3 \in C$, $\theta_1+\theta_2+\theta_3=1$

有 $\frac{\theta_1}{\theta_1+\theta_2}x_1 + \frac{\theta_2}{\theta_1+\theta_2}x_2$ 在仿集 (仿集定义)

又有 $(\theta_1+\theta_2)\left(\frac{\theta_1}{\theta_1+\theta_2}x_1 + \frac{\theta_2}{\theta_1+\theta_2}x_2\right) + \theta_3 x_3$ 也在C内

$= \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$ 也在C内

若= $x_2=-x_1$

有一类仿集，需求 $\theta_1+\theta_2=1$；

$V = C - x_0 = \{x - x_0 \mid x \in C\}$ $\quad \forall x_0 \in C$

C是一个 $\theta_1+\theta_2=1$ 的仿集

V是与C相关的子空间（只有C是仿集，才能它是子空间）

这个V就是 $\alpha+\beta\neq1$，但 $\alpha x_1+\beta x_2\in V$

（∀ 同时包括 $\alpha+\beta=1$，∴ V仿集

证：

$\forall v_1, v_2 \in V$，$\alpha+\beta\in R$

$\alpha v_1+\beta v_2 + x_0$ 认属于C，即 $\alpha v_1+\beta v_2$ 属于V

$$\alpha(\underbrace{v_1+x_0}_{\in C})+\beta(\underbrace{v_2+x_0}_{\in C})+(1-\alpha-\beta)\underbrace{x_0}_{\in C}$$

∵ $\alpha+\beta+(1-\alpha-\beta)=1$

∴ $\in \forall C$

∴ $\alpha v_1+\beta v_2 = C - x_0 \in V$

※ 从性质一般的仿集C得到性质更好的仿集V

V一定经过原点（因为从C上减掉了一个 x_0 点）

例2 线性方程的解集是仿集

证：$C=\{x \mid Ax=b\}$，

假设：$x_1\in C$，$x_2\in C$ $\quad \theta_1+\theta_2=1$

$A(\theta_1x_1+\theta_2x_2)$

$=\theta_1Ax_1+\theta_2Ax_2=(\theta_1+\theta_2)b=b \in C$

与 $C=\{x \mid Ax=b\}$ 相关的子空间

$V = C - x_0 \mid x$ $\quad x_0$是用它来

$= \{x - x_0 \mid A(x - x_0) = 0\}$ (即代入 $Ax_0 = b$)

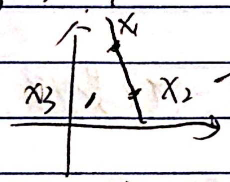
$= \{y \mid Ay = 0\}$ A的化零空间 一个

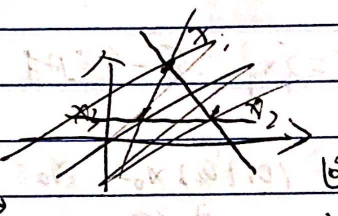
(反过来也对：任意仿集可理解为 Linear function 的解集)

例2 ∀集合C，构造一个最小仿射集 $\subset C$

用仿射包来构成 (affine ~~cone~~) 于C：$\{\theta_1 x_1 + \cdots \theta_k x_k \mid x_1, \dots x_k \in C, \theta_1 + \cdots \theta_k = 1\}$

假设 $C = \{x_1, x_2\}$





这条直线即C的仿包

~~这三条直线即 $\{x_1, x_2, x_3\}$ 的~~

当 $\{x_1, x_2, x_3\}$，其仿包是 $\mathbb{R}^2$

凸集 Convex

仿集一定是凸集

凸集合组合

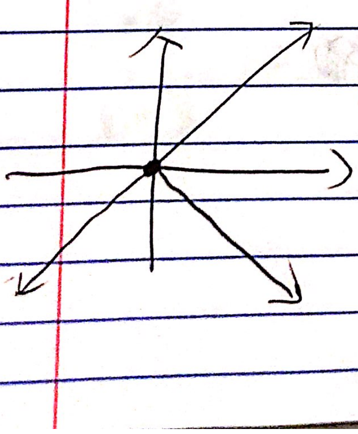
C为凸集 $\Longleftrightarrow$ ∀ $x_i \in C$ 的凸组合 $\in C$

凸包

(由C中任意个点的凸组合的全体 = 凸包)

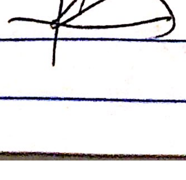
锥 cone　C是锥 $\Longleftrightarrow$ $\forall x \in C$, $\theta x \in C$ $(\theta \geq 0)$

凸锥 convex cone 一定是凸集：$(x_1, x_2 \in C,\ \theta_1, \theta_2 \geq 0,\ \theta_1 x_1 + \theta_2 x_2 \in C)$



三条射线，

三条射线是一个锥 (注意，锥必须过原点)

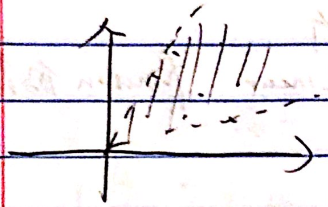


这是个锥 (= ∞ 个射线)

(也是凸锥)

4

How to judge convex cone combinations and convex cone hulls
阴影部分即为这集合的凸锥包



过原点的直线是凸锥

当一个集合$C=\{x_0\}$只有一个点时，

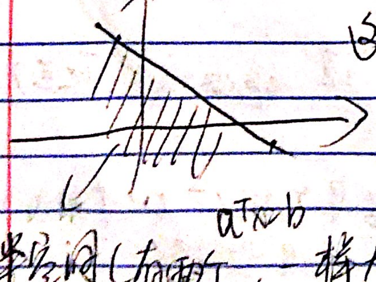
$\theta_1 x_0+\theta_2 x_0=(\theta_1+\theta_2)x_0=x_0$ ∴ C是仿射集、凸集。
$\theta_1 x_0\neq x_0$ ∴ 不是锥
(只有当x_0=原点时才是锥，且是凸锥)

当集合$C=\phi$，
C是仿射、凸、凸锥

几种重要的凸集

	仿射集	凸	凸锥
$\mathbb{R}^n$ 维空间	✓	✓	✓
n维空间的子空间	✓	✓	✓
任意直线	✓	✓	不一定(需过原点)
线段	X(只有一个点时)	✓	X(只有个原点)

超平面 hyperplane
$\{x \mid a^Tx=b\}$ $(a\neq \vec{0})$ 就被称为超平面



该直线即n=2时的超平面

半空间(右边的，一样大)

	仿	凸	凸锥
超平面	✓	✓	不一定(✓过原点 那就空间了)
半空间	X	✓	不一定(✓过原点)

球和椭球(欧式空间中)

$B = B(x_c, r) = \{x \mid \|x - x_c\|_2 \le r\}$

球	可能(r=0时)	✓	可能

用定义证球是凸

$\forall x_1, x_2 \in B$, 即 $\|x_1 - x_c\|_2 \le r$, $\|x_2 - x_c\|_2 \le r$

$\forall \theta \in [0,1]$

证 $\|\theta x_1 + (1-\theta) x_2 - x_c\|_2 \le r$

左边 $= \|\theta(x_1 - x_c) + (1-\theta)(x_2 - x_c)\|_2$

$\le \|\theta(x_1 - x_c)\| + \|(1-\theta)(x_2 - x_c)\|_2$ (三角不等式)

$= r$

∴ 得证

第6/55 —

椭球 加权的L2范数

$\mathcal{E}(x_c, P) = \{x \mid (x - x_c)^T P^{-1} (x - x_c) \le 1\}$, $P \in S_{++}^n$

对称的正定矩阵

⇓

~~eigenvalue~~ > 0

singular value

椭球的半轴长由P的奇异值决定

例. $x^T \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}^{-1} x \le 1$

singular $= 2$, 1

$\sin A = \sqrt{\text{eigenvalue}(A^T A)}$

奇异值和特征值不一定一样

多面体

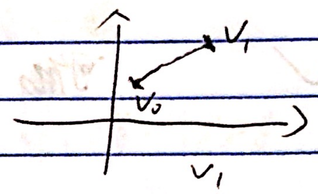
$P = \{x \mid a_j^T x \le b_j, \ c_i^T x = d_j\}$ (每个式子是一个半空间(超平面))

是凸集

单纯形：

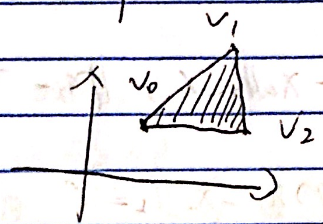
R^n 空间中选择 $v_0, \cdots v_k$ 共 $k+1$ 个点，$v_1-v_0, \cdots v_k-v_0$ 线性无关，
则与上述点相关的单纯形为
$C = \mathrm{Conv}\{v_0 \cdots v_k\} = \{\theta_0 v_0 + \cdots \theta_k v_k \mid \theta \ge 0, \mathbf{1}^T\theta = 1\}$
即基于这些点构造出的凸包

例：取 $k=$ ~~0~~ 1



v_1-v_0 本身线性无关

$k=2$



在 R^2 空间中，k 无法再变大

证 simplex 是多面体

思路：给一个 simplex，一定可以写成多面体

定义 $y = [\theta_1, \cdots \theta_k]^T$，有 $y \ge 0$，$\mathbf{1}^T y \le 1$ （因为 $y + \theta_0 = 1$）
$B = [v_1-v_0, \cdots v_k-v_0] \in R^{n\times k}$
将 x 写成 ~~x=~~ （x 是 simplex）
$x = v_0 + \theta_1(v_1-v_0) + \cdots \theta_k(v_k-v_0)$
$= v_0 + By$
$\because \mathrm{rank}(B) = k$ （因为 k 个线性无关）
B 通过线性变换可成为 $\begin{bmatrix} I_k \\ 0 \end{bmatrix} \begin{matrix} \to k\times k \\ \to (n-k)\times k \end{matrix}$
[$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ 非奇异，使 $AB = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$，~~则 B~~ $\begin{bmatrix} I_k \\ 0 \end{bmatrix}$]

$Ax = Av_0 + ABy$ （A 非奇异）

$\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} x = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} v_0 + \begin{bmatrix} A_1B \\ A_2B \end{bmatrix} y$

$\Rightarrow \begin{cases} A_1 x = A_1 v_0 + \overset{I}{A_1B}\, y \\ A_2 x = A_2 v_0 + \overset{0}{A_2B}\, y \end{cases}$

$$\Rightarrow \begin{cases} A_1 x \geq A_1 v_0 \\ 1 \cdot A_1 x < 1 + 1^T A v_0 \\ A_2 x = A_2 v_0 \end{cases} \quad (0 \leq y \leq 1)$$

即任意点由线性不等式和等式描述，即多面体 (30:40)

(一个无法想象的凸集)

$S^n = \{x \in R^{n\times n} \mid x = x^T\}$　　$S^n_+ = \{x \in R^{n\times n} \mid x = x^T, x \succeq 0\}$ 即所有特异值≥0

$S^n_{++} = \{x \in R^{n\times n} \mid x = x^T, x \succ 0\}$

~~都是凸锥~~

S^n_+ 是凸锥　证 $\forall \theta_1, \theta_2 \geq 0$，$\forall A, B \in S^n_+$，证 $\theta_1 A + \theta_2 B \in S^n_+$

易得 $\theta_1 A + \theta_2 B$ 也是对称

半正定(用性质 $\forall x \in R^n$，$x^T A x \geq 0$，则 $A \in S^n_+$

$x^T(\theta_1 A + \theta_2 B)x \geq 0$ 易得　∴ 凸锥

但 S^n_{++} 不是凸锥 (因为 $\theta_1 = 0$ $\theta_2 = 0$ 导到 0 不在 S^n_{++} 中)，但是凸集

7/55

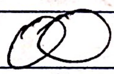
举例 $n=1$，$S^1_+ = x \geq 0$

$S^1_{++} = x > 0$ 不包含原点，即不是锥

$n=2$　$S^2_+ = \left\{\begin{bmatrix} x & y \\ y & z \end{bmatrix} \mid x \geq 0, z \geq 0, xz \geq y^2\right\}$

保凸运算

① 交集



② 仿射函数 $f(x) = Ax + b$　　即缩小与位移保持凸性

逆映射 也保凸 (因为还是线性)

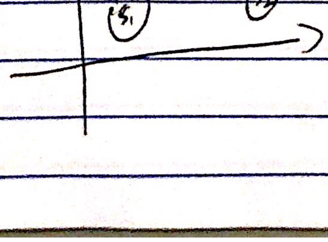
③ 两个凸集的合 (不是并集)

$S_1 + S_2 = \{x + y \mid x \in S_1, y \in S_2\}$

用仿映：$S_1 \times S_2 = \{(x, y) \mid x \in S_1, y \in S_2\}$ 也凸

$f(x, y) = x + y = (x, y) \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

看成一个变量 $\in S_1 \times S_2$，是凸集 $S_1 \times S_2$ 的仿映　∴ 保凸



二维的例子

8

例：线性矩阵不等式

$A(x) = x_1A_1 + \cdots x_nA_n \preceq B$

$B, x_i, A_i \in S^n$

$A(x) - B \preceq 0$，认为矩阵 $A(x) - B$ 非负定

解集是凸集

定义仿射变换 $f(x) = B - A(x)$，以高维矩阵空间变化到低维

S_+^n 为凸

$f^{-1}(S_+^n) = \{x \mid B - A(x) \succeq 0\}$

从低维映射到高维

f^{-1} 也是凸保凸 ∴ $\{x \mid B - A(x) \succeq 0\}$ 凸　得证

例2 椭球是球的仿射映射

$\varepsilon = \{x \mid (x - x_0)^T P^{-1} (x - x_0) \le 1\}$

$\{u \mid \|u\|_2 \le 1\}$ 单位球

$f(u) = P^{\frac{1}{2}} u + x_0$　　$((P^{\frac{1}{2}}) \cdot (P^{\frac{1}{2}}) = P)$

$\{f(u) \mid \|u\|_2 \le 1\}$

8/55　$= \{P^{\frac{1}{2}} u + x_0 \mid \|u\|_2 \le 1\}$

$x \triangleq P^{\frac{1}{2}} u + x_c$

$u = P^{-\frac{1}{2}}(x - x_c)$　(因为 P 是 S_{++})

$= \{x \mid \|P^{-\frac{1}{2}}(x - x_c)\|_2 \le 1\}$

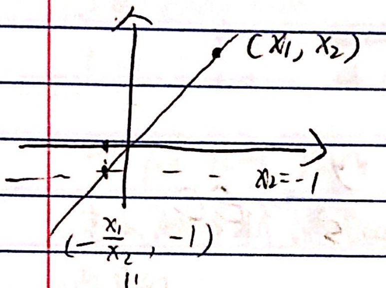
$= \{x \mid (x - x_c)^T P^{-1} (x - x_c) \le 1\}$

透视函数 (降维) → 后面一维实数空间

$P: R^{n+1} \to R^n$ dom $P = R^n \times R_{++}$

$P(z,t)=\frac{z}{t}$ $z\in R^n$, $t \in R_{++}$

除完之后最后的 t 没用,扔掉, $R^{n+1}\to R^n$



几何意义上的透视(小孔成象)

$(-P(x_1,x_2),-1)$

凸集透视后仍是凸集

例: 考虑 R^{n+1} 线段, $x=(\tilde{x}, x_{n+1})$, $y=(\tilde{y}, y_{n+1})$

$\theta x+(1-\theta)y$ $\theta\in[0,1]$

↓ 变换后仍凸, 且仍为线段

①先证凸集

$x \to P(x)$, $y\to P(y)$

$$P(\theta x+(1-\theta)y)=\frac{\theta\tilde{x}+(1-\theta)\tilde{y}}{\theta x_{n+1}+(1-\theta)y_{n+1}}$$

$$=\underbrace{\frac{\theta\cdot x_{n+1}}{\sim}}_{=\mu}\cdot\frac{\tilde{x}}{x_{n+1}}+\frac{(1-\theta)y_{n+1}}{\sim}\cdot\frac{\tilde{y}}{y_{n+1}} \quad (\mu\in(0,1)$$

$$=\mu\cdot P(x)+(1-\mu)\cdot P(y)$$

~~θ与μ一一对应~~

~~二、第二个~~ 证透视后仍线段

↓

证一一对应

10

透视函数的反函数

保凸

设 $C\in R^n$

定义 $P^{-1}(C)=\{(x,t)\in R^{n+1} \mid \frac{x}{t}\in C, t>0\}$

考虑 $(x,t),(y,s)\in P^{-1}(C)$

证 $\frac{\theta x+(1-\theta)y}{\theta t+(1-\theta)s}$ 仍属于 C

$$=\frac{\theta t}{\theta t+(1-\theta)s}\cdot\frac{x}{t}+\frac{(1-\theta)s}{\theta t+(1-\theta)s}\cdot\frac{y}{s}\in C$$

(因为 C 是凸集)

得证

线性分式函数

(把一个点先仿射,再透视)

$g(x)$ 是仿射映射

$g(x)=\begin{bmatrix}A\\c^T\end{bmatrix}x+\begin{bmatrix}b\\d\end{bmatrix}$

$P: R^{m+1}\to R^m$

$f: R^n\to R^m \triangleq P\circ g$ (先 g 再 P)

$f=\frac{Ax+b}{c^Tx+d}$ (要求 $c^Tx+d>0$)

保凸 (因为经过了两次凸变换)

[虽然看着非线性,但实际很好,凸]

例) 两个随机变量的联合分布 → 条件概率

$u\in\{1,\dots n\}$, $v\in\{1,\dots m\}$

$f_{ij}=P(u=i \mid v=j)$

$f_{ij}=\frac{P_{ij}}{\sum_k P_{kj}}$ 是一个线性分式

$P_{ij} = (0,\dots,0,\overset{i\text{th}}{1},\dots,0)\cdot\begin{pmatrix}P_{1j}\\ \vdots\\ P_{nj}\end{pmatrix}$

后者联合概率是凸：

9/55 缺失（凸函数的定义）

10/55 凸函数的扩展

$\tilde{f} = \begin{cases} f(x), & x\in dom\,f \\ +\infty & x\notin dom\,f \end{cases} \Rightarrow dom$ 扩 $\tilde{f} = R^n$

f 的 dom 为凸

定义的函数满足 $f(x)<\infty$

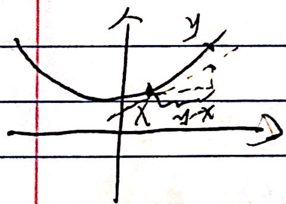
例：凸集的示性函数是凸函数　（当α为1时，只能是凸的）

$f_C(x) = \begin{cases} 常数 & x\notin C \\ 0 & x\in C \end{cases} \xrightarrow{扩} f_C(x) = \begin{cases} \infty & x\notin C \\ 0 & x\in C \end{cases}$ $(x\in R)$

一阶条件：（隐含：dom 是开集，否则在端点不可微）

当 f 可微时，对 $x,y\in dom\,f$，有

$f(y) \ge f(x) + \nabla f^T\cdot(y-x)$　（且 dom f 凸）



（定义是从 ∪ 的上面去理解）

根据一阶条件，当找到 x 使 $\nabla f(x)=0$，则 $f(y)\ge f(x)$，x 为 min 点

证一阶条件（部分）

①考虑一维：

有 $f(y) \ge f(x) + \nabla f(x)(y-x)$

证 ⇒

已知 f 为凸

∃ 找 $t\in(0,1)$　~~x+t~~ $x+t(y-x)\in dom\,f$

12

$f(x+t(y-x)) \le (1-t)f(x) + t f(y)$

$t f(y) \ge t f(x) + f(x+t(y-x)) - f(x)$

$f(y) \ge f(x) + \frac{f(x+t(y-x)) - f(x)}{t}$ → 一阶条件

~~取极限~~ $t \to 0^+$

$f(y) \ge f(x) + \nabla f(x)(y-x)$

∴得证

证⇐

设 $\forall\ x \ne y,\ x, y \in dom\ f,\ \theta \in [0,1]$

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

$\begin{cases} f(x) \ge f(z) + f'(z)(x-z) & \times\theta \\ f(y) \ge f(z) + f'(z)(y-z) & \times(1-\theta) \end{cases}$

$\theta f(x) + (1-\theta) f(y) \ge f(z) + \theta f'(z)x + (1-\theta)f'(z)y - f'(z)\cdot z$

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

$= f(z) + 0$

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

∴得证

11/55 (9中的一个定义)

$\forall\ x \in dom\ f,\ v \in dom\ f$ 一个方向

$g(t) = f(x + tv)$ 凸 ⇔ 凸

将高维降到低维

一阶条件 $n>1$ 的证明

$g(t) = f(ty + (1-t)x)$

$= f(x + t(y-x))$

$g(t) = \nabla f^T(ty + (1-t)x)(y-x)$

$\because g(t)$ 凸（用定~~性质~~义g中）

$\therefore g(t)$ 满足

$g(t_1) \geq g(t_2) + \nabla g(t_2)^T(t_1 - t_2)$ （一维已证过）

$t_1 = 1,\ t_2 = 0$

$g(1) \geq g(0) + \nabla g'(0)$

$g(1) = f(y),\quad g(0) = f(x)$

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

从一阶证凸 $\Leftarrow$

$\forall x, y \in dom\, f$

$ty + (1-t)x \in dom f,\quad \tilde{t}y + (1-\tilde{t})x \in dom f$

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

（即代入一阶条件）

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

$g(t) \geq g(\tilde{t}) + \nabla g(\tilde{t})(t - \tilde{t})$

（因为 $g(\tilde{t}) = \nabla f^T(\tilde{t}y + (1-\tilde{t})x)(y-x)$）

$\therefore g$ 是凸函数（定义）

根据 f 的高维定义

$\therefore f$ 是凸

二阶条件　　　（Hessain 矩阵半正定）

若f二阶可微　　$\nabla^2 f(x) \succeq 0$　（最实用的

$n=1$ 时　　即一阶偏导的值单调不减（即f有着像个碗）

14

若 $\nabla^2 f \succ 0$，严格凸

但严格凸 不需要 $\nabla^2 f \succ 0$

例：$f = x^4$ ∪

~~在 x=0时~~

$f' = 4x^3 \qquad f'' = 12x^2$

在 $x=0$ 时 $f'' = 0$

例：一个重要的凸函数

二次函数

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

凸性条件

求 $\nabla^2 f(x)$ $\quad \nabla^2 f(x) = P$

二次函数 $\quad \nabla^2 f \succ 0 \Leftrightarrow$ 严格凸

例：$f(x) = \frac{1}{x^2}$

$\nabla^2 f \qquad f'(x) = -2 \cdot x^{-3} \qquad f''(x) = 6x^{-4} = \frac{6}{x^4}$

但是 $\mathrm{dom} f = \{x \mid x \neq 0\}$，不是凸集

∴ 不是凸函数（但若单独 $x>0$ or $x<0$ 是凸）

例：$\nabla^2 f(x) = 0$ 矩阵 $\succeq 0$

12/55

$f(x) = e^{ax}$ 凸

$\nabla^2 f = a^2 e^{ax} \geq 0$

幂 $(x>0) \qquad f = x^a$

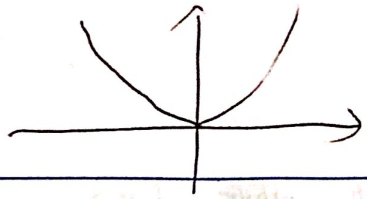
$f'(x) = a x^{a-1}$

$f'' = a \cdot (a-1) x^{a-2}$

$f'' \geq 0 \quad (a \geq 1, a \leq 0)$ 凸

$f'' < 0 \quad (0 < a \leq 1)$ 凹

也有可能既凸又凹



$|x|^3$

$x<0$

$(-x)^3$

$3(-x)^2\cdot(-1)$

$6(-x)$

绝对值的幂函数

$f=|x|^p \quad (x\in R)$

（f处处一定二阶可微）

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

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

~~p>1时 ∀f''>0，凸；但p=1时 f=|x|，也凸~~

p<1时，

$p>1$ （$p\neq1$，因为$p=1$不可导）

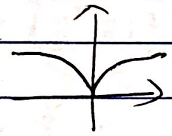
$p\geq2$时 $\nabla^2 f$ 有定义，易得

$p\in(1,2)$时，更复杂，但仍凸

$p=1$ V 也凸

（在$p<1$时无统一结论）

$p=\frac{1}{2}$



非凸也非凹

对数函数 $f=\lg x \quad (x\in R_+)$

凹函数

$f'=\frac{1}{x} \qquad f''=-\frac{1}{x^2}<0$

~~互熵~~

负熵 $f(x)=x\log x, \quad x\in R_+$

$f'=\log x+1 \qquad f''=\frac{1}{x}>0$ 严格凸

（$\log x$ 凹，x 凸，x 把 $\log x$ bāi 过来了）

极大化熵 = 极小化负熵（负熵凸，容易）

范数 $P(x)$

三性质 ① $P(ax)=|a|P(x)$

② $P(x+y)\leq P(x)+P(y)$

③ $P(x)=0 \Leftrightarrow x=0$

~~[illegible]~~ $\forall x,y\in R^n, \ \forall\, 0\leq\theta\leq1$

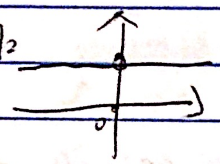
$P(\theta x+(1-\theta)y) \leq P(\theta x)+P((1-\theta)y)$ ②

$=\theta P(x)+(1-\theta)P(y)$ ①

零范数：$\|x\|_0$：非零元素总和 个数

~~女~~零范数非凸（因为不是真的范数）

例：



这个图就非凸

$P(ax) \neq |a|P(x)$

极大值函数

$f(x)=\max\{x_1,\cdots x_n\} \quad x\in R^n$

无法求导，用定义：

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

$=\max\{\theta x_1+(1-\theta)y_1, \cdots \theta x_n+(1-\theta)y_n\}$

$\le \max\{\theta x_1,\cdots \theta x_n\}+\max\{(1-\theta)y_1,\cdots (1-\theta)y_n\}$

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

∴凸

极小极大问题

$\min_x \max_y f(x,y)$ 也都是凸问题

做可导的逼近（解析逼近）

（log sum up）（softmax）

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

$= \cancel{n \log}$

$\max \le f(x) \le \max+\log n$

$\frac{\partial f}{\partial x_i}=\frac{e^{x_i}}{e^{x_i}+\cdots e^{x_n}}$

当 $i\neq j$ $\quad \frac{\partial^2 f}{\partial x_i \partial x_j}=\frac{-e^{x_i}e^{x_j}}{(e^{x_1}+\cdots e^{x_n})^2}<0$

$i=j$ $\quad \frac{\partial^2 f}{\partial x_i^2}=\frac{-e^{x_i}e^{x_i}+e^{x_i}(e^{x_1}+\cdots e^{x_n})}{(e^{x_1}+\cdots e^{x_n})^2} \quad \cancel{>0}$

定义 $z = \begin{bmatrix} e^{x_1} \\ \vdots \\ e^{x_n} \end{bmatrix}$

$\therefore \dfrac{\partial^2 f}{\partial x_i \partial x_j} = \dfrac{-e^{x_i}e^{x_j}}{\cancel{\|z\|_2^2}\ (\mathbb{1}^T z)^2}$

$\dfrac{\partial^2 f}{\partial x_i^2} = \dfrac{-e^{x_i}e^{x_i} + \cancel{\mathbb{1}^T z}\ e^{x_i}(e^{x_1} + \cdots e^{x_n})}{\cancel{\|z\|_2^2}\ (\mathbb{1}^T z)^2}$

B/55

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

$$= \underbrace{\frac{1}{(\mathbb{1}^T z)^2}}_{\text{一定}\geq} \underbrace{((\mathbb{1}^T z)\,\mathrm{diag}\{z\} - zz^T)}_{k}$$

$\begin{pmatrix} e^{x_1} & & \\ & \ddots & \\ & & e^{x_n} \end{pmatrix}$

半正定：对 $\forall\ v \in R^n$，$v^T P v \geq 0$ ~~⇔~~ 则 P 半正定

~~$v^T k v$~~ $v^T k v = (\cancel{\|z\|}\ \mathbb{1}^T z)\, v^T \mathrm{diag}\{z\} v - v^T z z^T v)$

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

定义 $a_i = v_i \sqrt{z_i}$ $\quad b_i = \sqrt{z_i}$

$$\boxed{\begin{aligned} &= (b^T b)\cdot(a^T a) - (a^T b)^2 \text{ 本身} \\ &\geq 0 \text{ (因为Cachy-Schwartz不等式)} \end{aligned}}$$

出现指数 → 三角不等

出现平方(和) → 柯西不等

$\therefore$ softmax 凸

18

几何平均

$f(x)=(x_1 x_2 \cdots x_n)^{\frac{1}{n}}$ $x\in R$ $x\in R^n_{++}$

凹函数

($-f$ 是凸)

例：行列式的对数

讲的真好

$f(X)=\log\det(X)$ $\quad dom\ f = S^n_{++}$ (此时 $\det(X)>0$)

此时 $\det(X)=\prod \lambda_i$

(把 x 想成标量，$f(x)=\log x$ 凹 ∴ $f(x)$ 多维可能是凹)

当 $n>1$ 时

用凹 f 的高维定义来做

$\forall\ Z\in S^n_{++}$ → 起点

$\forall t\in R$ → 距离

$\forall\ V\in R^{n\times n}$ → 方向

$Z+tV\in S^n_{++}=dom f$，且 $V\in S^n$

$g(t)=f(Z+tV)$

$=\log\det(Z+tV)$

$=\log\det\{Z^{\frac{1}{2}}(I+tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}})Z^{\frac{1}{2}}\}$

因为 $\det(AB)=\det(A)\det(B)$

$=\log\det Z^{\frac{1}{2}}+\log\det Z^{\frac{1}{2}}+\log\det(I+tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}})$

$=\log\det Z+\sum_i\log(1+\lambda_i)$ 的特征值

($tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}}$ 对称，进行分解为

$Q\Lambda Q^T$ $(QQ^T=I)$ $\quad\lambda_i\in(-1,1)$ 由定义域得

∴ $\det(I+tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}})$

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

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

$=\prod(1+\lambda_i)$

14/55 $\therefore g(t)$ 的凸性同原函数 $f(x)$ 的凸性

$g'(t)=\sum_i \frac{x_i}{1+tx_i}$ $\quad g''=\sum \frac{-x_i^2}{(1+tx_i)^2}\le 0$

$\therefore \forall g$ 凸 $\Rightarrow f$ 凸

保凸

非负的求加权和

$f_1 \cdots f_m$ 为凸，则 $f=\sum_i^m w_i f_i$ 为凸

证：看

① F 的定义域为 $\cap$，必凸

② 用定义得凸性

↓

变成积分

若 $f(x,y)$ 对任意 $y\in A$，$f(x,y)$ 均为凸（边际凸），~~则 $f(x,y)$ 为凸~~

则 (x,y) 为凸集 （$f(x,y)$ 不需要是凸 x ← 联合凸）

设 $w(y)\ge 0$, $\forall y\in A$

$g(x)=\int_{y\in A} w(y)\, f(x,y)\, dy$ 为凸 （对求和的扩展）

仿射映射

$Ax+b$ $\quad A\in R^{n\times m}$ $\quad b\in R^{mn}$

$g(x)=f(Ax+b)$ $\quad \mathrm{dom}\, g=\{x\mid Ax+b\in \mathrm{dom}\, f\}$

证：令 $x,y\in \mathrm{dom}\, f$

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

$\le \theta f(Ax+b)+(1-\theta)f(Ay+b)$

$=\theta g(x)+(1-\theta)g(y)$ $\quad\therefore$ 得证

20

~~若 $g(x) = Af(x)+b$~~,

~~$g(\theta x+(1-\theta)$~~

$f_i: R^n \to R$, $i=1,\cdots m$ 为凸　$A\in R^n$, $b\in R$

$g(x) = A^T[f_1(x)\cdots f_m(x)]^T + b$

$\boxed{\text{也不一定凸}}$ 因为实质是加权和:

如果A都非负则凸，否则不确定

两个函数的极大值函数

f_1, f_2 为凸，则 $\boxed{\max\{f_1(x), f_2(x)\}\text{ 凸}}$

定义域凸

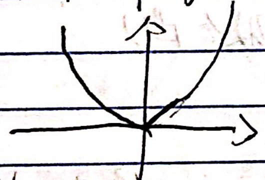
$g(\theta x+(1-\theta)y) = \max\{f_1(\theta x+(1-\theta)y), f_2(\theta x+1-\theta y)\}$

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

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

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

例1 $f(x) = \max\{x^2, x\}$



n个函数的极大值函数

例2 向量中r个最大元素的和

$x_{[i]}$ 为第i大的元素

$x_{[1]} \ge \cdots x_{[n]}$

$f(x) = \sum_{i=1}^{r} x_{[i]}$

$\|$　值意取r个元素

$f(x) = \max\{x_{i1} + \cdots x_{ir} \mid 1\le i_1, \cdots \le i_r \le n\}$

共有 C_n^r 种组合，其中最大的组合就是

仿射是凸

$x_{i1}+\cdots x_{ir}$ 任意一个组合可表示为 $(1,\cdots 0,1,\cdots)\cdot \delta$

$\therefore = \max\{x\ v_1 \delta, \cdots\ v_{C_n^r}\delta\} \Rightarrow \therefore$ 凸集

无限个凸函数的极大值

$g(x) = \cancel{\inf}\ \sup(f_1, \cdots f_n)$ 也凸

例：实对称阵的最大特征值

$\cancel{f_x =}\ f = \lambda_{max}(X)$, $\quad dom f = S^m$

$\delta y = \lambda y \Rightarrow y^T \delta y = y^T \lambda y \Rightarrow y^T \delta y = \lambda \cdot \|y\|_2^2$

$\therefore \cancel{y}\ \lambda = \frac{y^T \delta y}{\|y\|_2^2}$ （一个特向 y 对应一个特值 λ）

$\forall$ 将 y 缩放为 $\|y\|_2^2 = 1$

$\lambda = \cancel{\max_{\|y\|_2=1}}\ \sup_{\|y\|_2=1}\{y^T \delta y\}$

$\because y^T \delta y$ 对任意 y 凸（按定义，线性）

$\therefore f = \lambda_{max}(X)$ 凸

11/55 函数组合

（某些特殊）组合也保凸

$h: R^k \to R$，$\forall g: R^n \to R^k$

$f: h \circ g: R^n \to R$（先过 g 再过 h）

$f(x) = h(g(x))$ 定义域：$\begin{cases} x \in dom\, g(x) \\ g(x) \in dom\, h(x) \end{cases}$

先看特例，所有都一维，此时 $dom \in R$

判断 $f''(x) \geq 0$

$f' = h'(g(x)) \cdot g'(x) \qquad f'' = h''(g(x)) \cdot g'(x) \cdot g'(x) + g''(x) \cdot h'(g(x))$

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

22

① 假设 g 凸，则 $g''(x) \geq 0$，则 f 凸

$h'' \geq 0$ (凸)，$h'(\,) \geq 0$ ~~非减~~ 不降

∴ g 凸，h 凸且 h 不降

② g 凹，h 凸且 h 不增，则 f 凸

同理可得凹的条件

(一个放松的条件)

高维 $n, k \geq 1$

dom g, dom h, dom f $\neq R^n, R^k, R^n$ ← 解决了 dom 问题

h, g 的不二阶可微

与 1 维类似的推论

e.g. $h(x) = -\log x \rightarrow \tilde{h}(x) = \begin{cases} -\log x & x > 0 \\ \infty & x \leq 0 \end{cases}$

① h 为凸 $\tilde{h}$ 不降，g 凸，则 f 凸

h 凸 $\tilde{h}$ 不增，g 凹，则 f 凸

凹	不降	凹	凹
凹	不增	凸	凹

证明第一条

(不一定二阶可导)

$x, y \in \text{dom}\ f$，$g(x), g(y) \in \text{dom}\ h$

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

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

h 为凸，故 dom h 为凸

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

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

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

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

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

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

~~同为 $\tilde{h}$ (不降)~~

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

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

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

不能只用单调性，因为没有说 $g(\theta x+(1-\theta)y) \in \operatorname{dom} h$ 如 (1)(1)

证 $g(\theta x+(1-\theta)y) \in \operatorname{dom} h$

反证：若 $\notin \operatorname{dom} h$，则 $g(\theta x+(1-\theta)y) \in \operatorname{dom} \tilde{h}$

$\because \tilde{h}$ 不降

闭了不能性质 (扩展)

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

$\therefore$ 左边 $=\infty$，且右边 $\in \operatorname{dom} h$

$\therefore h(右) \ge \infty$ (因为 h 是凸)

即对 $\forall x,y$，都有 $h(右)=\infty$，则等于只有 $h(x)=\infty$，不是我们要的。

在其它情况下，矛盾 $\therefore g(\theta x+(1-\theta)y) \in \operatorname{dom} h$

$\Downarrow$

单调性

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

例：若 g 为凸，$\exp\{g(x)\}$ 为凸

$\because g$ 凸，$h=\exp()$ 凸且不降 $\therefore f=\exp(g(x))$ 凸

(这些条件在用于二阶条件不好用时，e.g. 不可微)

例：g 凹且 $g\ge 0$，$\log(g)$ 为凹

$\because g$ 凹 $\log$ 凹，且 $\tilde{\log}$ 不降

$\therefore \log g$ 凹

16/55 若 g 为凹，$g>0$，~~$\frac{1}{g}$ 为凸~~ 则 $\frac{1}{g}$ 为凸

$\because g$ 凹 $h=\frac{1}{z}$ 在 $g>0$ 时凸，扩展后不增

$\therefore f$ 凸

~~g凸，p常数 ≥ 0，$g^p(x)$ 为凸~~

~~$p=1$，g凸明显~~

~~例 $h(z)=z^p$~~

~~把z看作凸(~~

~~方便扩展保持单调性)~~

他乱说，下节课再讲此

函数的透视

(与透视函数不同之：$P: R^{n+1} \to R^n$ dom $P \in R^n \times R_{++}$

$P(z,t)=\frac{z}{t}$)

函数f: $R^n \to R$

其透视 g: $\cancel{R^{n+1}}\ R^n \times R_{++} \to R$

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

若f是凸，则g也凸，若f凹，则g凹

可表述为对(x,t) joint convex

例：对 l2 范数的平方

$f(x)=x^T x \quad x \in R^n$

$g(x,t) = t \cdot (\frac{x}{t})^T \cdot (\frac{x}{t}) = \frac{1}{t} \cdot x^T x = \frac{1}{t} \cdot f(x)$

例：负对数

$f(x) = -\lg x$ 凸

$g(x,t) = t \cdot (-\cancel{\lg x}\ \lg(\frac{x}{t}))$

$= t \cdot \lg(\frac{t}{x}) \quad (x,t) \in R^2_{++}$

(用f''条件可证g)

累加^n

将x,t扩展成 $u, v \in R^n_{++}$

$g(u,v) = \sum_i^n u_i \cdot \log\frac{u_i}{v_i}$

(~~因为~~ $= \sum g_i(u_i, v_i)$ 凸

这里可是非负加权和，因为定义域(u,v)是同一个

又凸又凹

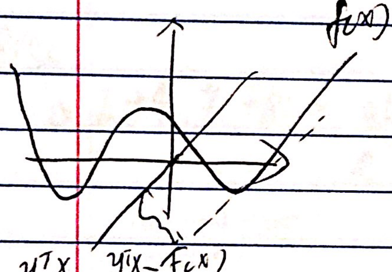
$D_{KL}(u,v) = \sum_{i=1}^{n} (u_i \log\frac{u_i}{v_i} - u_i + v_i) \to$ 凸

[KL-Divergence 的逻辑，D_{KL} 是 Bregman Divergence 特例，D_{breg}：
定义 $f: R \to R$ 为凸：$D_B(u,v) = f(u) - f(v) - \nabla f(v)(u-v)$
若 $f(u) = \sum_{i=1}^{n} u_i \log u_i - \sum u_i$，$D_B$ 变成 D_{KL}
但一般的 D_B 不一定凸（即使 f 必须凸）]

函数的共轭 Conjugate

$f: R^n \to R$，$f^*: R^n \to R$

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



（之所以不在左边找，因为 $f(x) > y^T x$，$y^T x - f(x) \to -\infty$）

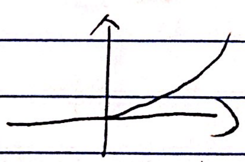
① $f(x)$ 若可微，则 ~~$f^*(x)$~~：$f^*(y)$ 对应的 x 必是 $f'(x) = y$

② ~~$f(x)$ 凸~~ 无论 $f(x)$ 是否凸，$f^*(x)$ 一定凸

（看 why f^* 凸，给定 x，f^* 是对 y 的 Linear function，
∵ 是 sup，即对一系列 LF 求极大值，∴ f^* 凸

17/55 例：若 g 为凸，$p \geq 1$，则 $g^p(x)$ 为凸

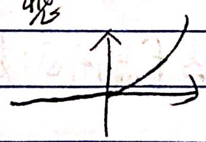
let $h(z) = z^p$ dom $h \in R_{++}$



$h(z)$ 的扩展无单调性

以 $h(z)$ 构造的函数，重新想

$h(z) = \begin{cases} z^p & x \geq 0 \\ 0 & x < 0 \end{cases}$



单调，凸

.26

函数的对偶(续)

① $f(x)=ax+b$

$f'(x)=y=a$

$\therefore f^*(y)=\begin{cases}\sup(-b)=-b & (y=a)\\ +\infty & (y\neq a)\end{cases}$

② $f(x)=-\log x \quad x\in R_{++}$

$f^*(y)=\sup(yx+\log x)$

$f'(x)=-\frac{1}{x}=y$

当 $y<0$ 时 $x=-\frac{1}{y}>0 \quad f^*(y)=\cancel{-1-\log(y)}$ $-1-\log(-y)$

当 $y\geq 0$ 时，$x=-\frac{1}{y}\notin R_{++} \quad \to +\infty$ (yx就变成∞了)

$\therefore f^*(y)=\begin{cases}-1-\log(-y) & y<0\\ +\infty & y\geq 0\end{cases}$

③ $f(x)=\frac{1}{2}x^TQx \qquad Q\in S^n_{++}$

$f^*(y)=\sup(yx-\frac{1}{2}x^TQx)$

$f'(x)=Qx=y \Rightarrow x=Q^{-1}y$

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

$=y^TQ^{-1}y-\frac{1}{2}y^TQ^{-1}y$

$=\frac{1}{2}y^TQ^{-1}y$

复数 $(a+bi)\to(a-bi)\to(a+bi)$

共轭的共轭是它自身

对f而言

$\because f^*$一定凸 $\therefore f^{**}=(f^*)^*$一定凸

$\alpha\because f$不一定凸 $\therefore f\neq f^{**}$

对于凸f，当f为闭函数，$f^{**}=f$

凸集与凸函数的关系

α-sublevel set

$C_\alpha = \{x \in \mathrm{dom} f \mid f(x) \le \alpha\}$

① 凸 f 的 $\forall \alpha$ set 都是凸集

证: $x, y \in C_\alpha$

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

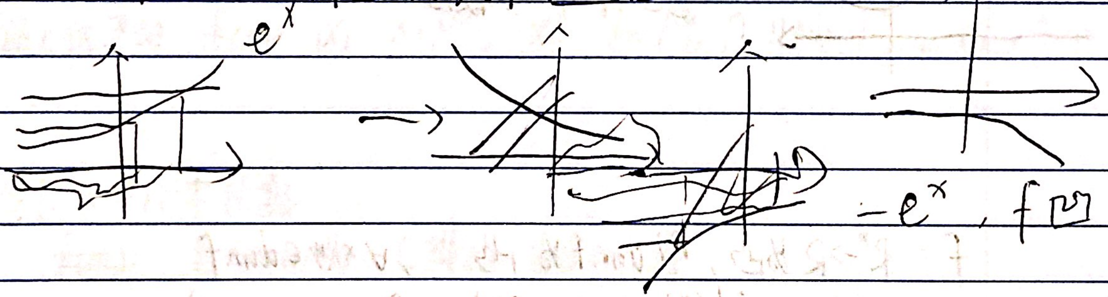
若 $f(\theta x+(1-\theta)y)$ 也 $\le \alpha$，则 $\theta x+(1-\theta)y \in C_\alpha$ 得证

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

∴ 得证

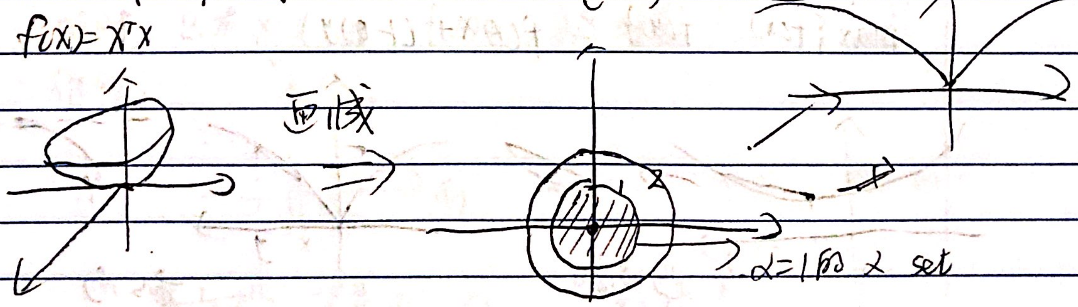
18/55

② 反过来，若 f 的 α-set 都凸集，则 f 不一定凸



高维: $f: \mathbb{R}^2 \to \mathbb{R}$，$x \in \mathbb{R}^2$，$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ 一定 不凸不凹

$f(x) = x^T x$



Quasi-Convex function 拟凸函数

① 不一定是凸函数

② 相对比较好优化的一类

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

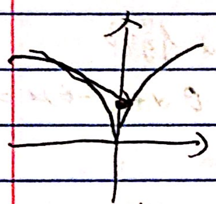
即所有的 α set 凸，即称为"拟凸"

$\le\alpha$ 凸图，"拟凸"

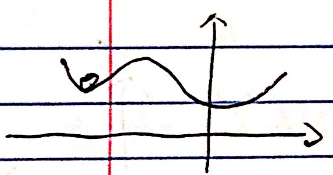
$=\alpha$ 凸 拟凸性

e^x 同时是拟凹/凸/线性
(因为 e^x 很像线)
凸 $\Rightarrow$ 拟凸 ，拟凸 $\nRightarrow$ 凸

拟凸也称'单模态'。不一定凸，但简单，因为很多凸优化算法可直接用在拟凸中。但理论上解释比较难

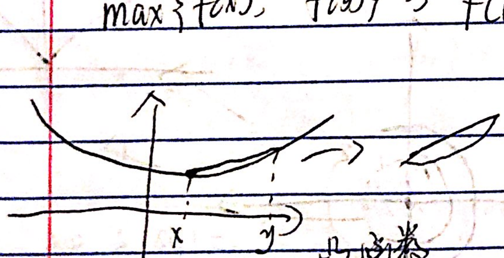


拟凸，仍然可降落到最低点

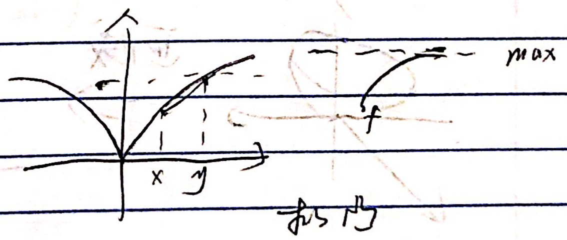


不是拟凸：会落停到 local-optimal
(这种叫'多模态')

$f: R^n \to R$ 为凸，则 dom f 为凸 ，$\forall x, y \in \text{dom} f$
$0 \leq \theta \leq 1$，$\theta f(x) + (1-\theta) f(y) \geq f(\theta x + (1-\theta) y)$
对于 f 是拟凸时（仍要求 dom f 凸）
$\max\{f(x), f(y)\} \geq f(\theta x + (1-\theta) y)$



凸函数



拟凸

向量的长度 $x \in R^n$
x 中最后一个非零 entry 的位置

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

向下取整 凹=3

照定义，看 $\{x \mid f(x) \le \alpha\}$，对所有 $i = \alpha+1, \cdots r$，$x_i = 0$

此 set 是个子空间（一些维度化成0）

∵ 子空间凸 ∴ f 凸

例1 线性分式函数

$f(x) = \frac{a^Tx+b}{c^Tx+d}$ f 包含凸集，但 f 不一定是凸函数，

但一定拟凸

$S_\alpha = \{x \mid \frac{a^Tx+b}{c^Tx+d} \le \alpha\}$

$\{x \mid a^Tx+b \le \alpha c^Tx + \alpha d,\ c^Tx+d>0\}$ 凸，因为是多面体

∴ 拟凸

19/55

证几何平均 $f(x) = (x_1 \cdots x_n)^{\frac{1}{n}}$，在 $x \in R^n_{++} \cap \|x\|_1 \le 1$（当成 Quiz）

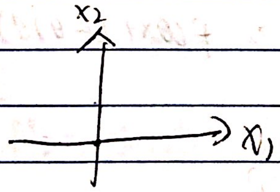
例2 向量的零范数

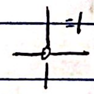
~~f(x)=~~ $f(x) = \|x\|_0$（非凸 由定义可得）

~~$f(x) \le \alpha$~~ α 为一维时，α>1时，x=

一维时，只有 x=0 才 f(x)=0，其它都是 1 ∴ 凸

二维时，





其它所有之 为 2

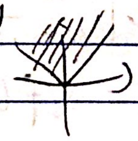
∴ ~~拟凸~~ 不是拟凸（十 非凸）

例. $\min \|x\|_0$，s.t. $x \in C$ 找一个最稀疏的 x

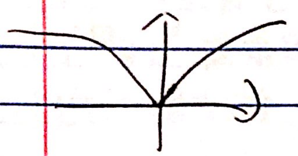
松弛解决近似问题

① 法一：处理成凸的问题（凸松弛）

$\min \|x\|_0 \Rightarrow \min \|x\|_1$（$\|x\|_1$ 凸）



各种近似 (n=1)



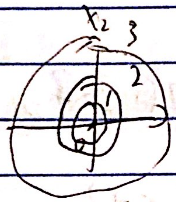
$\log(ax^2+1)$

$\min_{x \in C} \log(ax^2+1)$. 非凸，但拟凸

n=2 时

$\log(z^T z + 1)$

$x \in C$



拟凸

20/55

拟凸函数别的定义

可微拟凸的一阶条件

凸f: 若 dom f 为凸，$f(y) \ge f(x) + \nabla^T f(x)(y-x)$

↓

拟凸: dom f 为凸，$f(y) \le f(x) \Rightarrow \nabla^T f(x)(y-x) \le 0$

(存在一个推论关系)

证:

n=1

⇒:

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

设 $f(y) \le f(x)$, 则 $f(x) \ge f(\theta x + (1-\theta)y)$

凑一个一阶偏导

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

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

$\Rightarrow \frac{f(\theta x + (1-\theta)y) - f(\theta x + (1-\theta)x)}{(1-\theta)(y-x)}(1-\theta)(y-x) \le 0$

$\theta \to 1$, $\frac{\cdot}{\cdot} \to \nabla f(x)$ $\quad 1-\theta \to 0$

$\Rightarrow \nabla f(x) \cdot (y-x) \le 0$

$\Leftarrow$ $\forall x,y\in \mathrm{dom} f$，有 $f(y)\leq f(x) \Rightarrow \nabla^T f(x)(y-x)\leq 0$

$\max\{f(y), f(x)\} - f(\theta x+(1-\theta)y)$ （证 ≥ 0）

（设 $f(x)\geq f(y)$）

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

$= \frac{f(\theta x+(1-\theta)x) - f(\theta x+(1-\theta)y)}{(1-\theta)(x-y)} \cdot (1-\theta)(x-y)$

$= f'(x)(x-y) \geq 0$

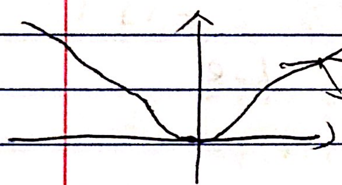
$\therefore \max\{\ \} \geq f(\theta x+(1-\theta)y)$

✗（这个证明有问题，因为已经 $\theta\to 1$，但是应该是 $\forall\ \theta\in[0,1]$，都成立

21/1/15

拟凸若 $\nabla f(x)=0$，则当 $f(y)\leq f(x)$ 时 $\Rightarrow 0\leq 0$ 没有意义

（这就是拟凸不如凸 f 的一点，一阶导 $=0$ 无意义）



拟凸

不一定有

这里 $f'=0$，但没有意义

二阶条件：

$\mathrm{dom} f$ 凸，且 $\Rightarrow y^T\nabla f(x)\geq 0 \Rightarrow y^T \nabla^2 f(x) y\geq 0$

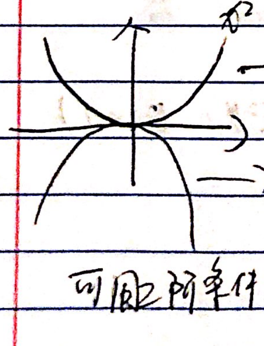
$n=1$ 时 $yf'(x)\geq 0 \Rightarrow y^2 f''(x)\geq 0$

$y=0$ 时 $0\geq 0 \Rightarrow 0\geq 0$

$y\neq 0, f'(x)=0 \Rightarrow \cancel{f''(x)\geq 0}\ f''(y)\geq 0$

极值点　　Hessain 半正定

（凸 f 要求每一点 $f''\geq 0$，拟凸只要求这些关键点）

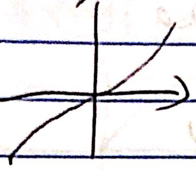


$\longrightarrow x=0\quad f'(x)=0\quad f''=2$

$\longrightarrow f'(0)=0,\ f''(0)=-2$（这个非凸）

可用二阶条件判断 f 不是拟凸

32



$f=x^3$ 非凸凹

$f'=3x^2=0 \quad x=0$

$f''=6x=0 \geq 0$

log convex / concave 对数凸/凹

$f: R^n \to R$ 为 ~~log convex~~ iff $\forall \log f$ 为凹
 logconcave.

同理有 log convex

(用来处理变量的乘积)

f 凸，log f 不一定凸：$f=x$ $\log x$ 为凹

若 log f 凸，f 凸

(用复合组合来证：令 $g=\log f$，求 e^g $h=e$

g 凸，h 凸，单调不降 → f 凸)

log f 凹，f 不一定凹 (因为 log 本身→凹)

f 凹，则 log f 凹

(用复合组，f 凹，log 凹 非降→凹)

(对数凸 > 凸；凹 > 对数凹

21/55 证拟凸的一阶条件

$\Leftarrow$

$x, y \in dom f$，$f(y) \leq f(x)$，$x \neq y$

$z=\theta x+(1-\theta)y$ (可扩展到 $x=y$，$\theta=0,1$)

$f(x)=\max\{f(x), f(y)\} \geq f(z)$

① $f(x) > f(z)$，无需做更证明

② $f(x)=f(z)$ }

③ $f(x)<f(z)$ } ⇒ $f(x) \leq f(z)$ 矛盾

证明 $f(x) \le f(z)$ 时 $f(x)$ 一定 $= f(z)$

② $\because$ $f(y) \le f(x)$, $f(z) \ge f(y)$ ($\ge$ max) 之为什么

$\therefore$

$f'(z)(y-z) \le 0$

$$\begin{cases} f(x) \le f(z) \\ f(y) \le f(z) \end{cases} \rightarrow \begin{aligned} f'(z)(x-z) \le 0 \\ f'(z)(y-z) \le 0 \end{aligned}$$

将 $z = \theta x + (1-\theta)y$ 代入

~~因为 f(y) ≤ f(x) < f(z)~~ $\Rightarrow f(y) \le f(z)$

$\theta \in (0,1)$

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

$$\therefore \begin{cases} f'(z)(x-y) \le 0 \\ f'(z)(x-y) \ge 0 \end{cases} \Rightarrow \cancel{x=y}\ f'(z)(x-y) = 0 \Rightarrow f'(z) = 0$$

z_1, z, x, y

$z_1 \to z_1$, 已知 $f(z_1) \ge f(x)$

$\therefore$ 同理又有 $f'(z_1) = 0$

$\therefore$ ~~所有~~ z 和 x 间 ~~存在~~ 不均匀的 $f' = 0$

$\because$ 连续

$\therefore f(z) = f(x)$

$\therefore$ 当 $f(x) \le f(z)$ 时 $f(x) = f(z)$

$\therefore f(x) \ge f(z)$ (综合①②③)

得 $\max(f(x), f(y)) \ge f(\theta x + (1-\theta)y)$

Convex Problem

① 目标函数为f
② 凸集约束 } 二次定义

一般的优化问题

$\min f_0$

s.t. $f_i(x) \le 0$ ← 不等式约束 } 没有时称为"无约束问题"

$h_i(x) = 0$

x 优化变量 f 目标函数

最优解：

若可行解集 $\neq \phi$ → $p^* = \inf\{f_0(x) \mid x \in X_f\}$

若 X_f 为空集 → 问题的设置或解答出错了

(因为无实际意义了)，写作 $p^* = +\infty$

若 x^* 可行，且 $p^* = f_0(x^*)$

称为 x^* 为最优解

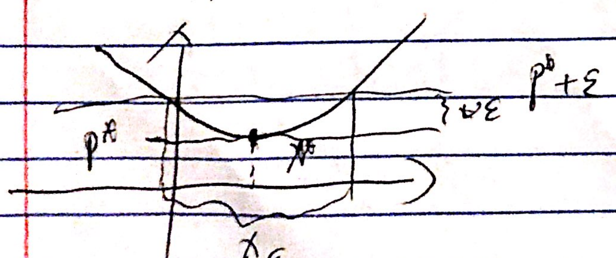
最优解的集合称为最优解集，可为 ϕ

ε 次优解集

Satisficing solution

一些对工程问题足够的解

$X_\varepsilon = \{x \in X_f \mid f_0(x) \le p^* + \varepsilon\}$



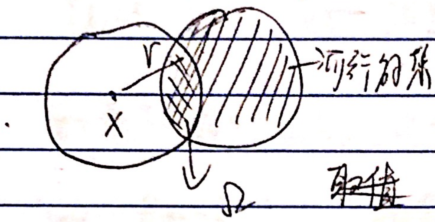
局部最优解

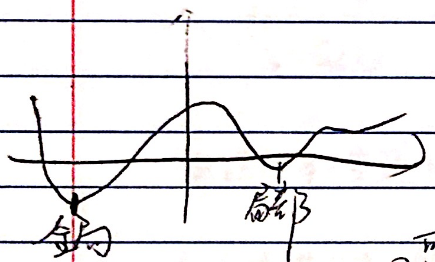
（有可能是非完全局解）

$f_0(x) = \inf\{f_0(z) \mid f_i(z) \le 0,\ h_i(z) = 0,\ \|z - x\| \le R\}$

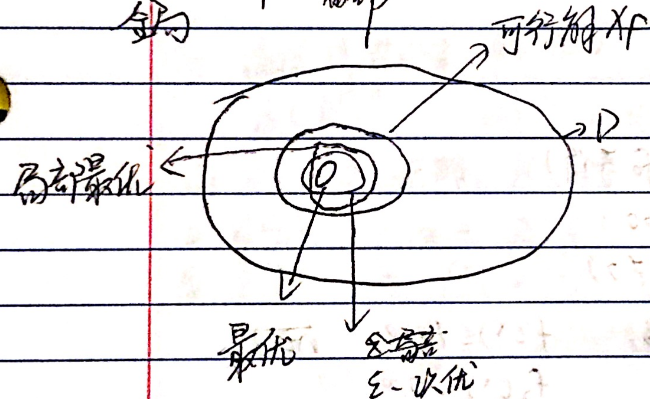
$\exists R > 0$ 使 x 局部最优

22/55





X_{loc} 局部最优解集



最优解 ⊂ 局部最优解

但 ε-次优 和局部最优 无确定关系

局部最优 ⊂ 可行集

为什么约束都是 '≤'，不是 '<'？

若 $x \in X_f$ 且 $f_i(x) = 0$，则 $f_i(x) \le 0$ 称为 活动约束 (active)（当约束起作用且在边缘）

→ $f_i(x) < 0$ 非活动约束 (inactive)

'<' 可以转化为 '≤' 的松弛形式

当我们把 "≤" → '<'，则本

s.t. $f_i(x) \le 0$ → x^* 时活动 等价 $f_i(x) < 0$
→ x^* 时活动 $f_i(x) < 0$ 不合理，因为实际上是在界限上取到的（不符合数学来源的实际问题）

e.g. s.t. money < 100 没有意义 → money ≤ 100 + 新的约束 |log(100 − money)| > 0

↓现实 money ≤ 100 ↘或 money ≤ 99.99

近似来说是 money ∈ [0, 100)

可行性优化问题 feasible problem

"关羽是最 x^*，磨化是 x_F"

find x $\xrightarrow[\text{化}]{\text{标准}}$ find min 0

s.t. $f_i(x) \le 0$ s.t. —

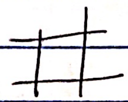
$h_i(x) = 0$ —

等价问题

例：Box Constraints

min $f_0(x)$

s.t. $l_i \le x_i \le u_i$



$\Updownarrow$

$l_i - x_i \le 0$

$x_i - u_i \le 0$

缩放 $\alpha_0 > 0$

例：min $\alpha_0 f_0(x)$ （与 min f_0 等价）

s.t. $\alpha_i f_i(x) \le 0$ $(\alpha_i > 0)$

$\beta_i h_i(x) = 0$ $(\beta_i \neq 0)$

现实意义：约束量级不一样：$f_1(x) \le 1000$，~~$h(x)$~~

$f_2(x) \le 0.001$，

缩放后 相当于normalize

构造新函数

例：ψ_0 R 单增

$\psi_1, \dots \psi_m$ $R \to R$ 条件 $\psi_i(u) \le 0 \Leftrightarrow u \le 0$

$Q_1, \dots Q_p$: $Q_i(u) = 0 \Leftrightarrow u = 0$

$\Downarrow$

min. $\psi_0(f_0(x))$ 等价于标准形式

s.t. $\psi_i(f_i(x)) \le 0$

$Q_i(h_i(x)) = 0$

例: min $\|Ax-b\|_2 \Leftrightarrow$ min $\|Ax-b\|_2^2$ (相当于 $\psi_0 = x^2$ (在 $x>0$ 时 ↗))

例: 消除等式约束

(约束越多，求解越复杂)

$\{h_i(x)=0\}$ 一组方程相当于 $\Leftrightarrow x = Q(z)$ 等价

($z \in R^k$, $Q: R^k \to R^n$)

如果能找到合适的 Q 和 z,

$$\begin{aligned} \min\ & f_0(x) \\ s.t.\ & f_i(x) \le 0 \\ & x = Q(z) \end{aligned} \xrightarrow[\text{代}x]{\text{用}z\text{替}} \begin{aligned} \min\ & f_0(Q(z)) \\ s.t.\ & f_i(Q(z)) \le 0 \end{aligned}$$

变成关于 z 的优化

$x^* = Q(z^*)$

但消除不一定保证更简单

例: 消除线性约束 $Ax-b=0$

$A \in R^{p\times n} \Rightarrow x = Q(z)$

① $Ax-b=0$ 无解，则 $x_F = \emptyset$

② A 可逆: $x = A^{-1}b$

③ 更一般 $x = Fz + x_0$ (这个容易理解) $z \in R^{n-r}$

A 的化零空间 $(Ax_0=b)$

~~Az~~ $AF=0$

23/55 当 $h_i(x) = a_i^T x_i = b_i$，称为"狭义"的凸问题

狭义与广义的区分

例: min $f_0(x) = x_1^2 + x_2^2$

s.t. $f_1(x)$ $\frac{x_1}{1+x_2^2} \le 0$ 狭义不是

$h_1(x)$ $(x_1+x_2)^2 = 0$

⇓ 变化

s.t. $x_1 \le 0 \to$ 凸

$x_1 + x_2 = 0 \to$ affine $\} \to$ 凸.

例: $\min\ f_0(Fz+x_0)$

s.t. $f_i(Fz+x_0)\le 0$　(消除affine 变为 $x=Fz+x_0$)

优点: ①降维

②减少约束维数

升维:

当 $f_i(x)$, $h_i(x)$ 都 affine

$\Downarrow$

引入 s. s.t. $s_i\le 0$　松弛变量 s

$f_i(x)-s_i=0$　(affine)

$h_i(x)=0$

Quasi-convex problem:

$f_i(x)$ 凸, $h_i(x)$ 仿射 (约束集为凸集),

f_0 为拟凸

max concave function = min convex function

凸问题的性质:

局部最优 = 全局最优

证: x 局部最优: $\exists R>0$　$f_0(x)=\inf\{f_0(z)\mid \|x-z\|\le R,\ x,z \text{可行}\}$

假设 x 不是全局最优,

$\exists y$ 可行 $f_0(y)<f_0(x)$, 因 x 局部最优

$\therefore \|y-x\|_2>R$

构造 $z=\theta y+(1-\theta)x$

$\theta=\frac{R}{2\|y-x\|}\in(0,\frac{1}{2})$

$\because$ 凸 $\therefore z$ 可行

$f_0(z)\le\theta f_0(y)+(1-\theta)f_0(x)$

(可以 z 在 $\|x-z\|\le R$ 中)

$\|z-x\|_2=\theta\|y-x\|_2=\frac{R}{2}$

$\therefore z\in\{\|x-z\|\le R\}$

$\therefore f_0(x) \le f_0(z)$ （$\because$ x 局部最优）

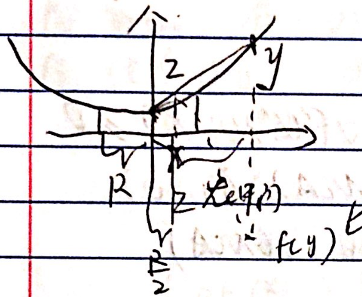
$\begin{cases} f_0(x) \le f_0(z) \\ f_0(z) \le \theta f_0(y) + (1-\theta) f_0(x) \end{cases}$

$f_0(y) < f_0(x)$

$\therefore f_0(y) < f_0(x) \le f_0(z)$

$\therefore \theta f_0(y) + (1-\theta) f_0(x) < f_0(z)$

矛盾 $\therefore$ x 一定全局最优



画出示图是说明 f(y) 一定比 f(x) 大

否设 f(y) < f(x)，则明显矛盾

② 目标函数 f_0 可微分时

（结合 f_0 可微：f_0 凸 $\Leftrightarrow$ $f_0(y) \ge f_0(x) + \nabla f_0(x)(y-x)$）

设可行域 $X_f = \{x \mid f_i(x) \le 0,\ h_i(x) = 0\} \cap \operatorname{dom} f_0$

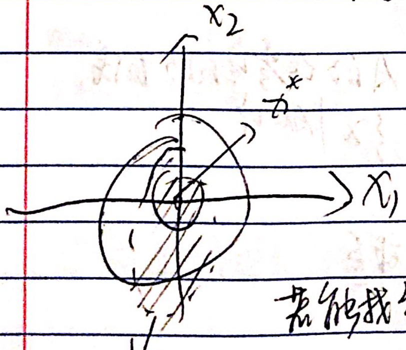
x^* 最优 $\Leftrightarrow$ $\nabla f_0^T(x^*)(y - x^*) \ge 0$

(?) 证明 有点没懂

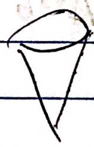
证明：设 $\operatorname{dom} f_0 = X_f$，若 x^* 使 $\nabla f_0^T(x^*)(y - x^*) \ge 0$ 对 $\forall y$ 成立，

则由于凸 f 性质 $f(y) \ge f(x) + \nabla f(x)(y-x)$

24/55



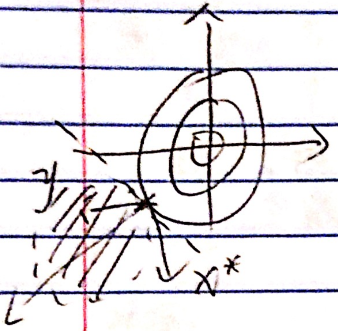
锥函数



若能找到 $f_0'(x) = 0$，则此点为 x^*（因为此时 $\nabla = 0$，$\nabla(y-x) \ge 0$）

新问题，加约束 $\because$，仍然 $f' \ge 0$，因为在 X_f

40



$\nabla f(x^*)(y-x) \geq 0$

梯度对话

例1：约束极值有 $h_i(x)=0$ 仿映

若∃ x^*, $Ax^*=b$. $x^* \Leftrightarrow \forall y$, $Ay=b$, $\nabla f^*(x^*)(y-x^*) \geq 0$

∵ $Ax=b$ $Ay=b$ ∴ $y=x+v$ ($v \in N(A)$, $Av=0$)

$\nabla f(x)(y-x) = \underline{\nabla f(x) \cdot v} \geq 0$ 对∀ $v \in N(A)$

这个对 ∇f 的要求很强.

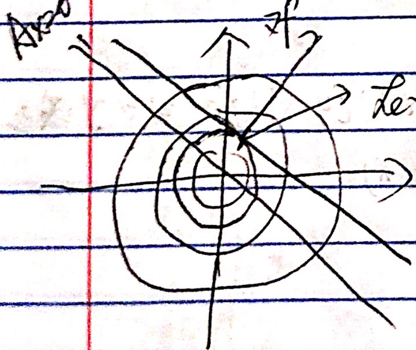
$n=2$

只有① 零子空间变为零空

或② ~~梯度与零空间~~ 与子空间正交

Case 1: $v=0 \Rightarrow y=x \Rightarrow Ax=b$ 有唯一解，A可逆 $x=A^{-1}B$ $x_f=\{A^{-1}B\}$

2: 正交 $\Rightarrow \nabla f \cdot v=0$ ∴ $\nabla f(x) \cdot v \geq 0$ 成立



在 x^* 切之，f最小

$Ax=b$ (A不可逆)，A的化零空间为直线 $\{x \mid Ax=0\}$

∇f 与 $Ax=0$ 之任意直线垂直

例：约束仅为非负约束

$\min \ f_0(x)$

$s.t. \ x \geq 0$

若∃ x ，x^* ⟺ $\forall y \geq 0$ $\nabla f(x)(y-x) \geq 0$

$\nabla f(x) \cdot y - \nabla f(x) \cdot x$

① 若 $\nabla f_0(x) \leq 0$ ，则 $\nabla f(x) \cdot y$ 会趋于无穷小，（因为 $y \geq 0$）则不再可能保证 $\nabla f(x) y - \nabla f(x) \cdot x \geq 0$ ∴ $\nabla f(x) \geq 0$

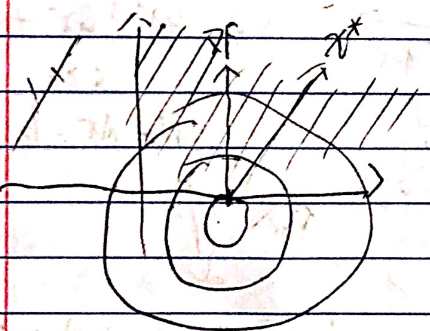
② 取 $y=0$ 有 $\nabla f(x) \cdot x \leq 0$

③ 结合①② ，有 ~~$\nabla f(x) = 0$~~ ~~又∵~~ $\nabla f(x) \geq 0$ ，$x \geq 0$ ∴ $\nabla f(x) \cdot x \geq 0$

∵② $\nabla f(x) \cdot x \leq 0$ ∴ ~~$\nabla f(x)=0$~~ $\nabla f(x) \cdot x = 0$

∴ $\begin{cases} x \geq 0 \\ \nabla f(x) \geq 0 \\ \nabla f_i(x) \cdot x_i = 0 \end{cases}$ 最优解的（充要）条件，称为互补条件（由一阶条件衍生出）

即 $\nabla f_i(x) = 0$ 或 $x_i = 0$



$x_1^* > 0$ $x_2^* = 0$ 互补

$\nabla_1 f(x^*) = 0$, $\nabla_2 f(x^*) > 0$

典型的凸问题

线性规划（目标和约束均为线性）

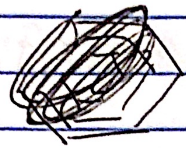
$\min \ c^T x + d$ $c \in R^n$ $d \in R$

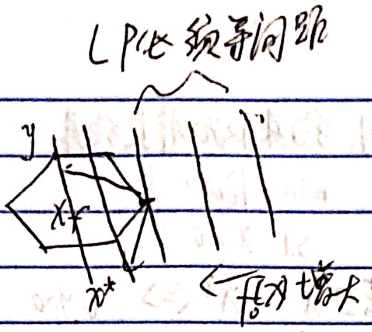
$s.t. \ Gx \leq h$ $G \in R^{m \times n}$ $h \in R^m$

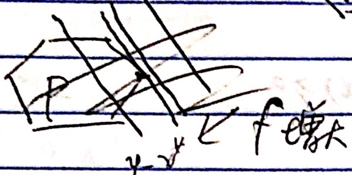
$Ax = b$ $A \in R^{k \times n}$ $b \in R^k$

由倒度问题发展来

25/35 X_f 是多面体







易得 $\langle y-x^*, \nabla f(x^*)\rangle \geq 0$

$\forall$LP，一定有一个 x^* 在顶点上（但可能 x^* 在整条边）

这个是LP的优势，即顶点，

等价形式

$\min\ c^Tx+d$ $\xrightarrow{\text{将x分成}}$ $\min\ c^Tx^+ - c^Tx^- + d$

s.t. $Gx+s \geq h$ 松弛变量 $\xleftrightarrow[x^+\text{和}x^-]{}$ s.t. $Gx^+ - Gx^- + s = h$

$s \geq 0$ $\qquad$ $s \geq 0$

$Ax=b$ $\qquad$ $Ax^+ - Ax^- = b$

($x=\begin{bmatrix}1\\2\end{bmatrix}$,
$x^+=\begin{bmatrix}1\\0\end{bmatrix}$
$x^-=\begin{bmatrix}0\\2\end{bmatrix}$)
$x=x^+-x^-$
$x^+, x^- \geq 0$

代入

优点：不等式约束是 ≥ 0

可规范写成 $\to$ 这个x包含 x^+, x^- 和s

$$\begin{cases} \min\ c^Tx \\ \text{s.t. } A^Tx=b \\ \quad x \geq 0 \end{cases}$$

例：食谱问题

m种营养元素 $b_1, \cdots b_m$

n种食物，单位营养为 $a_{1j}, \cdots a_{mj}$ $\quad j=1,\cdots n$

目标：单位食物价格 c_j，求总价最小

设食物量 $x_1, \cdots x_n$ $\qquad \min.\ \sum c_j x_j$

$\qquad$ s.t. $\sum_{j=1}^{n} a_{ij}\cdot x_j \geq b_i \quad (i=1,\dots m)$

$\qquad x_j \geq 0$

$$\Rightarrow \min \quad [c_1 \cdots c_n]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\text{s.t.} \quad \begin{bmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & & \\ a_{1n} & \cdots & a_{mn} \end{bmatrix}^T \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \geq \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \geq 0$$

(在现实中有时要求 $x_i = \text{int}$，则非凸)

例：线性分数规划 Linear fractional programming

是拟凸

P_0 min. $f_0(x)$ (一个线性分数函数 $\frac{c^Tx+d}{e^Tx+f}$ ← 非凸、拟凸 ⇄)

s.t. $Gx \leq h$

$Ax = b$

$\Updownarrow$ (前提是分式 ($f\neq 0$，否则无法转换)

P_1 min. $c^Ty + dz$

s.t. $Gy - hz \leq 0$ 等价

$Ay - bz = 0$

$e^Ty + fz = 1$

$z \geq 0$

证：

(1). 若 x 在 P_0 中可行 $y = \frac{x}{e^Tx+f}$，$z = \frac{1}{e^Tx+f}$ 构造的方式

$Gx \leq h$ $\frac{Gx-h}{e^Tx+f} \leq 0$ ✓ $z \geq 0$ ✓

$Ax = b$ $\frac{Ax-b}{e^Tx+f} = 0$ ✓

$e^Tx+f > 0$ (by def) $\frac{e^Tx+f}{e^Tx+f} = 1$ ✓

$$f_0(x) = c^Ty - hz$$

(2). 若 y, z 在 P_1 中可行，则 $x = \frac{y}{z}$ (设 ~~z>0~~ $z > 0$)

26/55 若 $z = 0$ 时，~~则构造 y,z 构造合乎条件~~ (神奇一。。)

无法构造 x，设 x_0 为 P_0 可行解一个，则

$x = x_0 + ty$ 也对 P_0 可行 ~~成立~~ ($\forall t \geq 0$)

∵此 $z = 0$ ∴ $Gy \leq 0$ $Ay = 0$ $e^Ty = 1$ (P_1 s.t.)

∴ 验证 $x = x_0 + ty$ 在 P_0 中

$$Gx = Gx_0 + Gty \leq h + \underbrace{(Gy)}_{\leq 0}\underbrace{(t)}_{\geq 0} \leq h$$

$$Ax = Ax_0 + tAy = \cancel{A} \; b + 0$$

$$e^T x + f = \underbrace{e^T x_0 + f}_{>0} + \overset{>0}{t} > 0$$

$\therefore x = x_0 + ty$ 也可行

设 $f_0(x) = f_0(y, z)$

P1 P1

$$f_0(x) = \frac{c^T x_0 + c^T t y + d}{e^T x_0 + e^T t y + f}$$ 找一个t 使 $= c^T y$

取 $t \to +\infty$

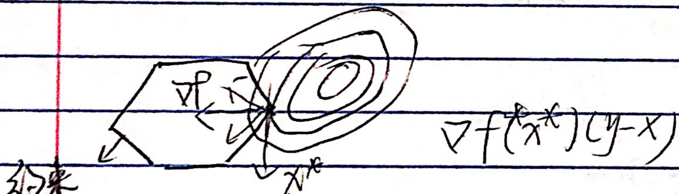
$$f_0(x) \to \frac{c^T t y}{e^T t y} \quad \because e^T y = 1 \quad \therefore = c^T y$$

二次规划 Quadratic P

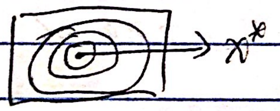
min: $\frac{1}{2} x^T P x + q^T x + r$, $P \in S_+^n$ 二次的

s.t. $Gx \le h$ } affine

$Ax = b$



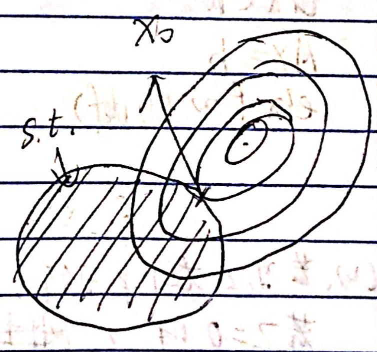
极值不一定在边上



二次约束二次规划 QCQP

(s.t. 为二次的, $Ax = b$)

$\frac{1}{2} x^T P_i x + q_i x + r_i \le 0$



例2 带噪声的测量系统

$b = Ax + e$ 信号恢复

$\hat{x} = \arg\min_x \|b - Ax\|_2$

$= \arg\min_x \|b - Ax\|_2^2$

$= \arg\min_x \left(x^T(A^TA)x - 2b^TAx + b^Tb \right)$

$A^TA \in S_+^n$

若A可逆 $\cancel{(A^{-1}A)^{-1}}$ $\hat{x} = (A^TA)^{-1}A^Tb$ $(\nabla f = 0)$

增加先验知识：

① $\hat{x}$ 稀疏 $\Rightarrow \hat{x} = \arg\min \|Ax-b\|^2 + \lambda_0 \|x\|_0 \rightarrow$ 尽可能稀疏

$\because \|x\|_0$ 非凸 $\therefore$ 近似成 $\|x\|_1$

$\hat{x} = \arg\min \|Ax-b\|^2 + \lambda_1 \|x\|_1 \leftarrow L_1$-regulized least square

（就是lasso）

☆ 用 $x = x^+ - x^-$ （常用于LP&QP）

$\hat{x} = \arg\min \|b - Ax^+ + Ax^-\|_2^2 + \underline{\lambda_1 \|x^+ - x^-\|_1}$

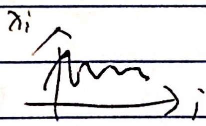
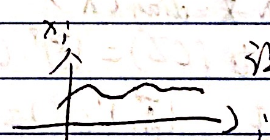
$= \lambda_1 \mathbf{1}^T x^+ + \lambda_1 \mathbf{1}^T x^-$ ☆ 可以

$+$ s.t. $x^+, x^- \geq 0$

例：L_2-regulized least square （岭回归）

先验：x_i 差别不大

$\hat{x} = \arg\min \|Ax-b\|^2 + \lambda_2 \|x\|_2^2$ QP

无 L_2 时  $\xrightarrow{\text{加上}L_2}$  让曲线更平

$\Updownarrow$ 存在 θ 使两个等价对于个给定 λ_2

例：

$\arg\min \|b - Ax\|_2^2$

s.t. $\|x\|_2^2 \leq \theta$ QCQP

27/55 例：投资组合问题 portfolio

n种投资，各投资 $x_1, \cdots x_n$ 万元，最终收入 $P_1x_1, \cdots P_nx_n$

$\max P_1x_1 + \cdots P_nx_n$

s.t. $\sum x_i \leq B$ 本金 budget $\rightarrow \sum x_i = B$ （现实意义）

$x_i \geq 0$ ~~房~~ 股 银行

假设三种投资 $P = [1.05, 1.05, 1]$ （考虑随机性，看 E）

$\Sigma = \begin{bmatrix} 1.2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ （ Var ）

4.6

ς 后改成 max

$$\min \ \vec{x}^T \Sigma \vec{x}$$ 一个的模型很重要

$$\text{s.t.} \ \bar{p}^T x \geq r_{min}$$

$$1^T x = B$$

$$x \geq 0$$

半定规划

$$\min. \ tr(CX)$$

$$\text{s.t.} \ tr(A_i X) = b_i$$

$$X \succeq 0$$

$$X \in S_+^n \quad C \in R^{n \cdot n}, \ A \in R^{n \cdot n} \ b_i \in R$$

是矩阵空间内的线性规划问题

(因为取 trace 之后是线性的)

例2 对角矩阵 C, A, X

$\vec{x} = diag(X)$

$$\min. \ tr(CX) = (\vec{C})^T \cdot (\vec{x})$$

$$\text{s.t.} \ (\vec{A_i})^T (\vec{x}) = b_i$$

$$diag(X) \geq 0$$

是关于 $\vec{x} = diag(X)$ 的线规 LP

例2 $\min \ C^T X$

$$\text{s.t.} \ x_1 A_1 + \cdots x_n A_n \preceq B$$ 半定约束

$X \in R^n$, $B, A_1, \cdots A_n \in S^k$, $C \in R^n$

是线性矩阵不等式，凸集

例2 $A(x) = A_0 + x_1 A_1 + \cdots x_n A_n$

$A_i \in R^{p \times q}$, $i = 0 \cdots n$, $x \in R^n$

谱范数 $\|A(x)\|_2 = A(x)$ 最大奇异值
找到个 x，最小化谱范数
$\min \|A(x)\|_2$
利用 $\|A(x)\|_2 \le \sqrt{s}$, $s>0 \iff A^T(x) \cdot A(x) - s \cdot I \preceq 0$

min ~~xx~~ t $\iff$ min s

~~$\|A(x)\|_2 \le \sqrt{s}$~~ $\underbrace{A^T(x)A(x) \le t^2 I}_{\text{关于 } x \text{ 二次的}}$ 对 x, s 联合的 s.t. ——

$\Updownarrow$ $\Downarrow$

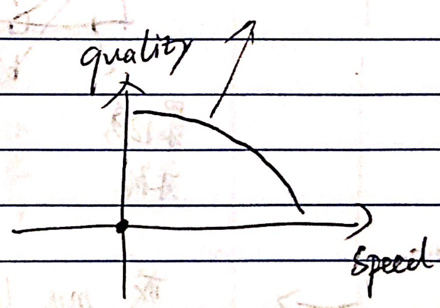
28/55 $\min t$
s.t. $\begin{bmatrix} tI & A(x) \\ A^T(x) & tI \end{bmatrix} \succeq 0$
$t \ge 0$

$\min t$
s.t. $Y = \begin{bmatrix} tI & A(x) \\ A^T(x) & tI \end{bmatrix}$
$Y \ge 0$
$t \ge 0$

多目标~~优化~~ 优化
$\min f_0(x)$ ($R^n \to \underline{\underline{R^q}}$, q个输出)
s.t. $f_i(x) \le 0$
$h_i(x) = 0$

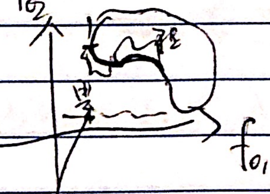
帕累托最优面

例 portofolio $\min$ Risk, $-$income
s.t. Resource
发展 $\min$ $-$speed, $-$quality
s.t. Resource



pareto: optimal value point
frontier 面

最优例



帕累托面不一定连续
Oracle solution: 让 f_{01} 和 f_{02} 都最小的点（但达不到）

一个特殊的多目标问题:

若 $\{f_0(x)\}$ 在 R^k 上凸, $f_i(x)$ 凸, $h_i(x)$ 仿射, 则可通过下述方法求Pareto面:

$$\min \sum_{i=1}^{q} \lambda_i f_{0i}(x) \quad (\lambda_i \geq 0)$$

非负加权

$$\text{s.t. } f_i(x) \leq 0$$
$$h_i(x) = 0$$

※ 通过 iterate λ_i 可找到所有 Pareto 点

常处理成单目标.

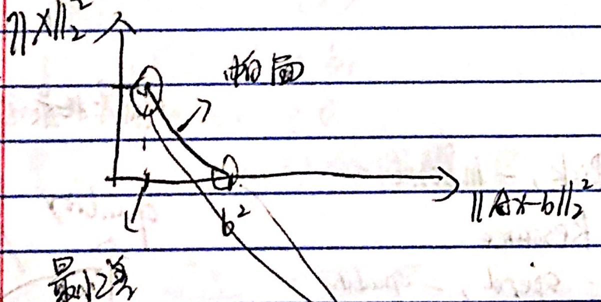
e.g. max quality
s.t. speed > 5

例: 岭回归

$b = Ax + e \qquad \min \|Ax-b\|^2 + \lambda_2 \|x\|_2^2$

↓ 多目标

$\min \|Ax-b\|_2^2, \quad \min \|x\|_2^2$



最小二乘
乘解

或 $\min \|Ax-b\|_2^2$

$\text{s.t. } \|x\|_2^2 \leq \varepsilon$

$\varepsilon = +\infty$

$\varepsilon = 0$

29/55 对偶性 Duality

凸分析的核心

$\min f_0(x)$ 　　不一定凸

s.t. $f_i(x) \le 0$

$h_i(x) = 0$

$x \in R$ 　 $D = \bigcap_{i=0}^{m} \mathrm{dom} f_i \cap \bigcap_{i=0}^{p} \mathrm{dom}\, h_i$

p^* optimal value

Lagrangian function

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

向量 ~~dim~~ $\lambda \in R^m$

$v \in R^p$

λ、v 拉格朗日乘子

关于 λ/v 都是线性的

↓ 构造

Lagrangian dual function

$g(\lambda, v) = \cancel{\inf_{x \in D} L}\ \inf_{x \in D} L(x, \lambda, v)$

① g 一定是凹函数

($\sup_{x \in D} L(x, \lambda, v)$ 必凸，因为关于 λ 和 v 是线性的)

↓ 极小

($\inf L($ 　 $)$ 关于 λ，v 线性，故凹函数)

② $\forall \lambda \ge 0$，$\forall v$，$g(\lambda, v) \le p^*$ 　 g是最优值的一个下界

证：设 x^*，则 $x^* \in X_f$，有 $f_i(x^*) \le 0$，$h_i(x^*) = 0$

当 $\forall \lambda \ge 0$，$\forall v$，有 $\sum \lambda_i f_i(x^*) + \sum v_i h_i(x^*) \le 0$

$$g(\lambda, v) \le L(x^*, \lambda, v) = f(x^*) + \left(\sum \lambda_i f_i(x^*) + \sum v_i h_i(x^*)\right)$$

$$\le f(x^*)$$

得证

极大化 $g(\ )$，就最逼近 p^*

∵ g凹, max 凹 = min 凸

又∵ $\lambda_i \geq 0$ 线性 ∴ max g 为凹问题

目标为得到尽可能接近的下界

例: min $x^T x$ s.t. $Ax=b$

$$L(x,\lambda,v) = x^T x + \cancel{\lambda^T(Ax-b)}\ v^T(Ax-b)$$

$$g(v) = \inf_{x\in D} \left(x^T x + v(Ax-b)\right) \quad D=R$$

$$= \inf\ x^T x + v^T A x - v^T b$$

用一阶导得 $\cancel{x}$ $2x + A^T v = 0$ $x = -\frac{1}{2}A^T v$ 代入得

$$g(v) = \inf \left(\frac{v^T A A^T v}{4} - \frac{v^T A A^T v}{2} - v^T b \right)$$

$$= -\frac{v^T A A^T v}{4} - b^T v \quad 凹$$

例: min $c^T x$

s.t. $Ax=b$

$x \geq 0$ (这里的是 $-x \leq 0$)

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

$$= (c^T - \lambda^T + v^T A)x - v^T b$$

$$g(\lambda, v) = \inf_x L(x,v,\lambda)$$

①如果系数为0, 则 $g = -v^T b$

②如 $\neq 0$, 则 $g = -\infty$

$$= \begin{cases} -b^T v & A^T v - \lambda + c = 0 \quad 一个超平面 \\ -\infty & \text{otherwise} \end{cases}$$

只需证 $-b^T v$ 时凹 就能得证 g 凹 (−∞ 是拓展)

∵ $-b^T v$ 线性

∴ 凸&凹

例: $\min\ x^TWx$ （W不需正定）

s.t. $x_i = \pm 1$ binary 约束

$\Updownarrow$

$x_i^2 - 1 = 0$

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

30/55 ~~$g(x,v)$~~ $g(v) = \inf\left(x^T(W + \mathrm{diag}(v))x - 1^Tv\right)$ 用凸定义得 dom 为凸

$$= \begin{cases} -1^Tv & (当\ W + \mathrm{diag}(v) \succeq 0) \\ -\infty & \text{otherwise} \end{cases}$$

~~-1^Tv 为凸~~

函数的共轭（与对偶的关系）

f^* 是 f $R^n \to R$ 的共轭

若 $f^*(y) = \sup\limits_{x \in \mathrm{dom} f} (y^Tx - f(x))$ （f^*为凸）

例: $\min f(x)$, s.t. $x = 0$ $\Rightarrow$ $\min = f(0)$

$L(x, v) = f(x) + vx$ $(x \in \{0\},\ v \in R^n)$

$$g(v) = \inf_{x \in \{0\}} (f(x) + vx) = -\sup_{x \in \mathrm{dom} f}(-f(x) - vx) = -\sup(-vx - f(x))$$
$$= f^*(-v)$$

$\therefore g(v) = f^*(-v)$

例: $\min f_0(x)$. s.t. $Ax \le b$, $cx = d$

$$L(x, \lambda, v) = f_0(x) + \lambda^T(Ax - b) + v^T(cx - d)$$
$$= f_0(x) + (\lambda^TA + v^Tc)x - \lambda^Tb - v^Td$$

$$g(\lambda, v) = \inf(L) = \inf\left(f_0(x) + (\lambda^TA + v^Tc)x\right) = -\sup\left(-(\lambda^TA + v^Tc)x - f_0(x)\right) \ldots$$
$$= -f^*\left(\underbrace{-(\lambda^TA + v^Tc)^T}_{要保证部分在\ \mathrm{dom} f^*\ 里}\right) - \lambda^Tb - v^Td$$

5.2

(D) $\begin{cases} \max & g(\lambda, \nu) \\ s.t. & \lambda \geq 0 \end{cases}$ ~~定义最优值为~~ d^* (duality problem)

(P) $\min \ f_0(x)$
$s.t. \ f_i(x) \leq 0$ primary problem
$h_i(x) = 0$

① $d^* \leq p^*$

② ~~d 是一个问题~~ λ^*, ν^* : optimal lagrange multiplier
也是对偶问题最优解

例: $\min \ c^T x$, s.t. $Ax = b$, $x \geq 0$

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

这一块无意义的对偶可丢掉

$\Downarrow$

$\max \ -b^T \nu$
s.t. $\lambda \geq 0$
$A^T \nu - \lambda + c = 0$ (若 $A^T \nu - \lambda + c \neq 0$ 时, 即无意义)

$\Downarrow$

$\boxed{\min b^T \nu}$ s.t. $A^T \nu + c \geq 0$

$\Downarrow$

注意 $\min b^T \nu$ 等价于 $\max -b^T \nu$, 但二者最优值并不一样
(是最优的 ν 是同一个)

例: $\min \ c^T x$ s.t. $Ax \leq b$

$\Rightarrow L(x, \lambda) = c^T x + \lambda (Ax - b)$
$= (c^T + \lambda A) x - \lambda b$

$g(\lambda) = \inf \ (c^T + \lambda A) x - \lambda b$
$= ~~\sup(-(c^T+\lambda A)x + \lambda~~$
$= ~~\inf(-\lambda b) +~~$

~~不太对~~

① $c^T+\lambda A=0$ $\quad g(\lambda)=-b^T\lambda$

② else $\quad g(\lambda)=-\infty$

D=> $\max\ -b^T\lambda$ $\qquad \min\ b^T\lambda$ $\quad$ D2

s.t. $\lambda\geq 0$, $c^T+\lambda A=0$ => s.t. $\lambda\geq 0$, $\cancel{c^T+\lambda A=0}$ c 很像

与一個的原问题，将 x 换成 λ => $\min\ c^T\lambda$

s.t. $A\lambda=b$ $\quad$ P1

$\lambda\geq 0$

本例 $\min\ c^Tx$ $\qquad$ 与一多例的对偶 $\min\ b^Tx$

s.t. $\cancel{Ax\leq b}$ $\qquad$ s.t. $Ax+c\geq 0$

P2 $\quad -Ax+b\geq 0$ $\quad$ P1

也很像

说明：P1 的对偶的对偶 是P1 自身 (对偶！得类推) (max 凹函数 凸问题)

(因 f^{**} 不一定 $=f$ 因为凸，所以 $D(D(P))$ 不一定 $=P$ 因为凸)

$D(D(P))=P$ 的等价条件 ①

以 LP 为例 $\cancel{\min\ c^Tx}$

P, P维约数 s.t. Ax ~~(P维约数)~~ 为n维

D, x 为P维，有 n维约束 <-> 对应

若 $P^*=D^*$ ②

54

31/55 $d^* = p^*$ 的条件 $\begin{cases} d^* \le p^* & \text{弱对偶} \\ d^* = p^* & \text{强对偶} \end{cases}$

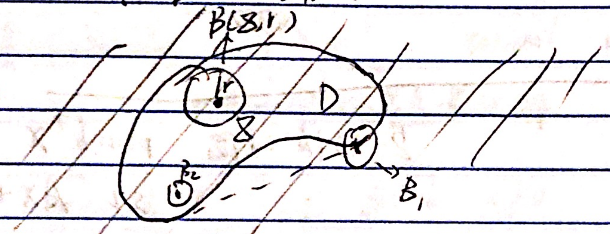
对偶间隙 $p^* - d^*$

定义：D的 relative interior := Relint D

Relint $D = \{x \in D \mid B(x, r) \cap \text{aff}\, D \subseteq D, \exists r > 0\}$

$\begin{cases} B(x,r) & \text{一个球} \\ \text{Aff } D & \text{D的仿射包} \end{cases}$

例：



R为仿射 D

$B \cap \text{Aff } D = B \subseteq D$

∴ 要求 δ 尤可在边缘 ∴ Relint D 即不包括边缘

Slater' 条件 $(p^* = d^*)$

充分但不必要

若有凸问题 min $f_0(x)$ (凸)

s.t. $f_i(x)$ (凸)

$Ax = b$

Slater 条件：~~不等式可行域~~

当 $\exists x \in \text{relint}(D)$，使 $f_i(x) < 0$，$Ax = b$， <，不是≤

此时 $p^* = d^*$

(一个不要紧的推论) 一般的凸问题都符合 slater 条件

(充分) 弱 slater 条件 $cx + d \le 0$ 半空间 ~~$cx+d=0$~~ 即 $x \in$ Relint D，且 $Ax \le b$，$cx = d$

若凸问题的不等式约束为仿射约束，此时只要 x_f ~~存在~~，

必有 $p^* = d^*$

~~[难以找到在这空间内部的点]~~

是所有D的∩，但须可行域！

$\text{relint } D = \text{relint}\{\text{dom} f_0 \cap (\cap_i \text{dom} f_i)\}$

在验slater条件

$\text{relint} D = \text{relint}\{\text{dom} f_0\}$（因为 $Ax+b\le 0$ $x\in R$）

若 $\text{dom} f_0 = R$，~~则一定~~ 则必有 $x_f \ne \phi$，则slater满足

若 $\text{dom} f_0$ 有内部，则 $\text{relint} D$ 一定 $\ne \phi$

例1：f_0 也仿射（LP）

若 $x_f \ne \phi$，则必 $d^* = p^*$

例2：$x^T x$ s.t. $Ax = b$ ∵ 若 $x_f \ne \phi$

∴ 必有 $p^* = d^*$

例：QCQP

$$\min \ \frac{1}{2}x^T P_0 x + q_0^T x + r_0 \quad (P_0 \in S_{++}^n)$$

$$\text{s.t.} \ \frac{1}{2}x^T P_i x + q_i x + r_i \le 0 \quad (P_i \in S_+^n)$$

~~$P_0 \in S_{++}^n$ $P_i \in$~~

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

$$g(\lambda) = \inf_{x\in D} L(x,\lambda)$$

$$= \inf \Big(\frac{1}{2} x^T \underbrace{(P_0 + \sum \lambda_i P_i)}_{P} x + \underbrace{(q_0^T + \sum \lambda_i q_i)^T}_{Q} x + \underbrace{(r_0 + \sum \lambda_i r_i)}_{R} \Big)$$

（$\lambda_i \ge 0$ 才是凸）

（$\lambda \ge 0$ 时）$= -\frac{1}{2} Q^T(\lambda) P^{-1}(\lambda) Q(\lambda) + R(\lambda)$ 是λ的凹函数

↓

验证slater条件

① $x\in D$ $x\in R$

② $\frac{1}{2}x^T P_i x + q_i x + r_i < 0$

(2-1) 若 $q_i, r_i = 0$，$\frac{1}{2}x^T P_i x$ 恒 ≥ 0

此时 s.t. $\frac{1}{2}x^T P_i x \le 0$，$x_f = \{\vec{0}\}$

不满足slater，但仍 $d^* = p^*$

$p^* = r_0$，$d^* = r_0$ $p^* = d^*$

56

(2-2) 若 $q_i \neq 0$, $r_i \neq 0$，找对称外

32/55 $p^* = d^*$ 的解释

几何解释：

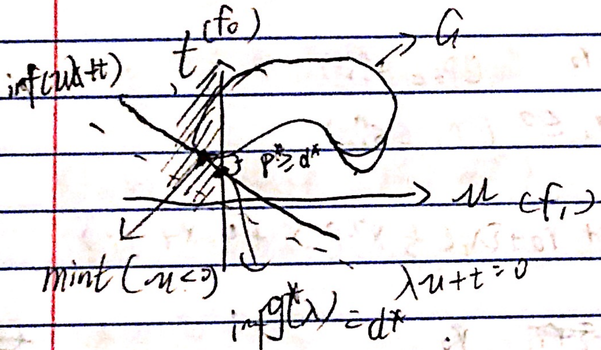
$\min f_0(x)$

s.t. $f_1(x) \le 0$

$G = \{(f_1(x), f_0(x)) \mid x \in D\}$

则 $p^* = \inf\{t \mid (u,t) \in G,\ u \le 0\}$

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



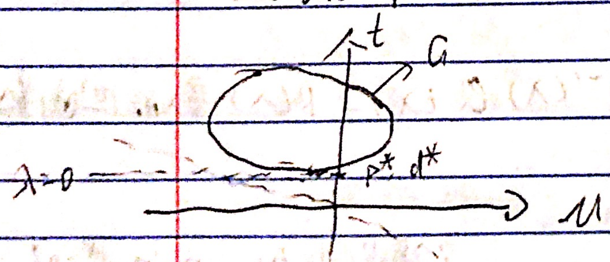
$g(\lambda)$ when $\lambda = 2$

$\inf(2u+t)$

$t = -2u$

当G很奇怪时，$p^* = d^*$ 很难成立。

当G规则时



$\lambda = 0$ 时 $p^* = d^*$

的解释

鞍点 saddle point

$$p^* = \left| L(x,\lambda),\quad \inf_{x\in D}\sup_{\lambda\geq 0} L(x,\lambda) \right| = \boxed{\sup_{\lambda\geq 0} \overbrace{\inf_{x\in D} L(x,\lambda)}^{g(\lambda)=}} = d^*$$

若成立，称此问题有鞍点，对应的 (λ^*, x^*) 为鞍点 L的

证左边 $= p^*$，即

$L = f_0(x) + \lambda f_i(x)$

$\sup L = f_0(x) + \sup \lambda_i f_i(x))$

① 若所有 $f_i(x) \leq 0$，则 $\sup = f_0(x)$

② 若存在 $f_i > 0$，$\sup L = f_0(x) + \infty = +\infty$

∵ $\inf(\sup L)$ ∴只能取①，即所有 s.t. 满足，$\inf(\sup ①)$

$= \inf((f_0(x))) = \inf f_0(x) = p^*$

多目标优化角度的解释

$\min\{f_0(x), \dots, f_m(x)\}$，是多目标，不是取最小的f

通过引入 $\{\lambda_i\} \Rightarrow \min(f_0(x) + \sum \lambda_i f_i(x))$ 即 Lagragian function

和对偶问题

step ① 给定 $\bar{\lambda}$，~~min(f0(x))~~ $\min_{x\in D} L(x,\lambda) \Rightarrow \bar{x}$ 为解

② 对λ求极大 $\max_{\lambda\geq 0} g(\lambda) \rightarrow \bar{\lambda}$ 为解

∴将 ~~$\bar{x}$~~ $\bar{\lambda}$ 代入 $L(x,\lambda)$，$\min L(x,\bar{\lambda})$ 的解与 $\min(f_0(x) + \sum \bar{\lambda}_i f_i(x))$

在 $d^* = p^*$ 时等价一致

经济学解释

$\min f_0(x)$

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

x是商品数量，f_0 是亏损（$-f_0$ 是利润）

$f_i(x)$ 是原材料i的约束

~~若原材料可买卖~~

58

解读举例

~~若~~原材料可交卖，~~交易价格为~~入；
则最终损失为 $\min_x \left(f_0(x) + \sum \lambda_i f_i(x)\right) = g(\lambda)$

剩下的材料；

去极小化总损失，
即可同时得到市场价入为何，故 minL 是入的函数
max g 表最大损失 $= d^*$
弱对偶：$d^* \le p^*$（当原材料可交易时，总损失一定 ≤ 不可交易时的总损失）

—o—

弱对偶说明市场经济比计划经济好 —o—
强对偶时 $d^* = p^*$ 只有分配不非常好时
x^* 是生产计划，λ^* 是市场价格调节，使市场全体效益最高（称为影子价格）

33/55 ⟶ Boyd 书的优化知识，另个人讲的，跳过了
34/55

35/55 影子价格 λ^* 的特点 市场价

~~若 $\lambda_i^* = 0$，则~~ ~~剩下的材料 $= f_i(x^*)$~~

x^* 时 对于 $f_i(x^*) < 0$ 剩下的材料 $\lambda_i^* = 0$（比时卖和不卖都一样）

对于 ~~$f_i(x^*)$~~ > 0
当 $\lambda_i > 0$（市场价），则 $f_i(x^*)$ 必无库存 $= 0$

那么在最优的价格下说明卖掉所有库存，即 $\lambda_i^* f_i(x^*) = 0$ 一定

~~稍~~ 稍后补充
$f(w, z)$ $w \in S_w$，$z \in S_z$（可用的函数）
必有 $\sup_z \inf_w f(w,z) \le \inf_w \sup_z f(w,z)$

minimax 不等式

想象 $f(w,z)$ 为 (w,z) 位置的人的身高

∴ sup(inf) 表示从一群矮子中找高的

inf(sup) 表示从高个子里找个最矮的

等式成立条件

① 等式成立 ⇔ 所对应的 w,z 也相等

称为鞍点 $(\tilde{w},\tilde{z})$

等价定义，$f(\tilde{w},z) \le f(\tilde{w},\tilde{z}) \le f(w,\tilde{z})$ 对 $\forall z,w$

$L(x,\lambda) = f_0(x) + \sum \lambda_i f_i(x)$ 只有 $f_i(x) \le 0$ 时，sup 才 $=0$；other $+\infty$ 要 $\lambda \ge 0$

$\sup_{\lambda \ge 0} \{L(x,\lambda)\} = f_0(x) + \sup\{\sum \lambda_i f_i(x)\}$

$= f_0(x)$ （在 $f_i(x) \le 0$ 情况下）

$p^* = \min_x \{f_0(x) \mid f_i(x) \le 0\}$

$= \inf_x \sup_{\lambda \ge 0} \{L(x,\lambda)\}$ 即 $p^* = \inf \cdot \sup$

$d^* = \max(\min L) = \sup \cdot \inf L$

必有 $p^* \ge d^*$

鞍点定理：

对只有不等式约束的优化问题，若 $(\tilde{x},\tilde{\lambda})$ 为 $L(x,\lambda)$ 的鞍点，

⇔ $\underbrace{\text{强对偶存在}}_{p^*=d^*}$，且 $\tilde{x},\tilde{\lambda}$ 为 d^* 与 p^* 的最优解

证明 ⇒ 都是基于鞍点定义

∵ 鞍 ∴ sup inf = inf sup ⇒ $d^* = p^*$

∵ 鞍 ∴ $(\tilde{x},\tilde{\lambda})$ 是鞍点左和右的最优解

$(\tilde{x},\tilde{\lambda}) = \arg \sup_{\lambda} \inf_{x} L(x,\lambda)$

↓

$\tilde{\lambda} = \arg_{\lambda} \left[\underbrace{\sup_{\lambda} \inf_{x} L(x,\lambda)}_{\text{对偶问题}} \right]$ 为对偶问题 $g(\lambda)$ 的最优解

② 视角右边 $(\tilde{x},\tilde{\lambda}) = \arg \inf \sup L(x,\lambda)$ → $\tilde{x}$ 是 primary 最优解

60

二证：

强对偶存在 $\Rightarrow$ ~~$d^* = p^*$~~ $f_0(x^*) = g(\lambda^*)$

$\because x^*, \lambda^*$ 分别为最优点

$\therefore x^*, \lambda^*$ 为可行解

$\therefore f_i(x^*) \le 0$，$\lambda^* \ge 0$

$$g(\lambda^*) = \inf_x \{f_0(x) + \sum \lambda_i^* f_i(x)\}$$

则任意 x，有

$$f_0(x^*) = g(\lambda^*) \overset{\ge g(\lambda^*)}{\le} f_0(x^*) + \sum \lambda_i^* \overset{\le 0}{f_i(x)} = L(x^*, \lambda^*)$$

$$= \le f_0(x^*) + 0$$

$\therefore$ 每一步都必须等号

36/55 因此生成的两个等式

$$g(\lambda^*) = L(x^*, \lambda^*) \Rightarrow \inf L(x, \lambda^*) = L(x^*, \lambda^*)$$

再考虑 $\sup L(x^*, \lambda)$

$$= \sup_{\lambda \ge 0} \{f_0(x^*) + \sum \lambda_i \overset{\le 0}{f_i(x^*)}\}$$

$$= \sup f_0(x^*) + 0$$

$$= L(x^*, \lambda^*)$$

$$\begin{cases} \inf L(x, \lambda^*) = L(x^*, \lambda^*) \\ \sup L(x^*, \lambda) = L(x^*, \lambda^*) \end{cases}$$

$\Rightarrow (x^*, \lambda^*)$ 是鞍点（由鞍点第一个定义）

例：$\min f_0(x)$ ← 第一定义

s.t. $f_i(x) \le 0$

$h_i(x) = 0$

对偶函数 $g(\lambda, v) = \inf_x \{f_0(x) + \sum \lambda_i f_i(x) + \sum v_i h_i(x)$

对偶问题： $\max\ g(\lambda, v)$

$\quad s.t.\ \lambda \geq 0$

(有)两个假设，① $d^* = p^*$ ——> 此时有许多好性质

② 所有 f_i, h_i 可微)

设此时已找到 (x^*, λ^*, v^*)

有 ③ $f_i(x^*) \leq 0 \quad i \geq 1$

④ $h_i(x^*) = 0$

⑤ $\lambda^* \geq 0$

$f_0(x^*) = g(\lambda^*, v^*) \qquad (\because p^* = d^*)$

$= \inf_x \{ f_0(x) + \sum \lambda_i^* f_i(x) + \sum v_i^* h_i(x) \} = \inf_x L(x, \lambda^*, v^*)$

$\leq f_0(x^*) + \underbrace{\sum \lambda_i^* f_i(x^*)}_{\leq 0} + \underbrace{\sum v_i^* h_i(x^*)}_{0} \quad \| \quad L(x^*, \lambda^*, v^*)$

变成等号 <— $\leq f_0(x^*)$

说明 x^* 是 inf 的全局最优

$\because$ 可微 $\therefore \nabla L(x^*) = 0$

$\because$ 等式成立

$\therefore \sum \lambda_i^* f_i(x^*) = 0$

$\because \forall\ \lambda_i^* f_i(x^*) \leq 0$

$\therefore \forall\ \lambda_i^* f_i(x^*) = 0$

① 若 $\lambda_i^* > 0 \Rightarrow f_i(x^*) = 0$ (价格>0 资源—> 无剩余

② 若 $f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0$ 　有剩余资源 —> 价格=0)

① 互补松弛条件 Complementary slackness

(含義，"≤"一个"≤"可由于另一个"="而 有在变成"<")

$\left.\frac{\partial L(x, \lambda^*, v^*)}{\partial x}\right|_{x^*} = 0$

② 稳定性 stationarity $\quad \nabla f_0(x^*) + \sum \lambda_i^* \nabla f_i(x^*) + \sum v_i^* \nabla h_i(x^*) = 0$

(λ^*, v^*) 是稳定的

共5个式子，可分为4类

一、原问题的可行性 ③④

二、对偶问题可行性 ⑤

三、互补松弛 ①

四、稳定性 ②

KKT条件

KKT条件，是最优解的必要条件

（充何时充分）

★ 若原问题为凸问题，各个函数可微，对偶间隙为0，
则KKT为充要条件

加上了凸问题的条件

证 凸问题时充分：

（充分性）设 $(\tilde{x}, \tilde{\lambda}, \tilde{v})$ 满足KKT条件，则必有 $(\tilde{x}, \tilde{\lambda}, \tilde{v})$ 为最优解

即证 $g(\tilde{\lambda}, \tilde{v}) = f_0(\tilde{x})$

（因为这个成立则 $(\tilde{x}, \tilde{\lambda}, \tilde{v})$ 为鞍点（极大 = 极小成立））

∵ 鞍点

∴ $g(\tilde{\lambda}, \tilde{v}) = f_0(\tilde{x})$ （间隙为0）

$(\tilde{x}, \tilde{\lambda}, \tilde{v})$ 为最优解

先用KKT的占个条件

$f_i(\tilde{x}) \le 0$

$h_i(\tilde{x}) = 0$

$\tilde{\lambda}_i \ge 0 \to L(x, \tilde{\lambda}, \tilde{v})$

$= f_0(x) + \sum \tilde{\lambda}_i f_i(x) + \sum \tilde{v}_i h_i(x)$

∵ 凸问题

∴ h_i 仿射函数，$\sum \tilde{\lambda}_i f_i(x)$ 凸，$f_0(x)$

∵ 非负加权

∴ L 凸函数

$\nabla L = 0$ 得最优点

$g(\tilde{\lambda}, \tilde{v}) = \inf\limits_{x \in D} L(x, \tilde{\lambda}, \tilde{v})$

$= L(\tilde{x}, \tilde{\lambda}, \tilde{v})$

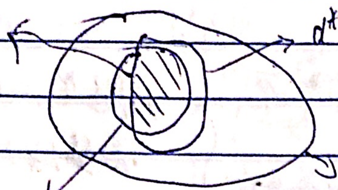
$= f_0(\tilde{x}) + \underbrace{\sum \tilde{\lambda}_i f_i(\tilde{x})}_{=0} + \overbrace{\sum \tilde{v}_i h_i(\tilde{x})}^{=0}$

互补松弛

$g(\tilde{x}, \tilde{v}) = f_0(\tilde{x})$ 也得证

（用了KKT中的互补松弛 和 对偶问题的可行性）

KKT的意义在于：求解KKT等价于求解原问题（在凸问题



例2 min $\frac{1}{2}x^TPx + q^Tx + r$ $\quad P \in S_+^n$

s.t. $Ax = b$

写出KKT：(1) $Ax^* = b$

(2) $\lambda_i \ge 0$ 不需要

(3) 互补也没了

(4) $\nabla f_0 = \cancel{P^Tx+q}$ $Px^* + q + A^Tv^* = 0$

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

（当 $Ax \ge b$ 时，很复杂，因为③）

Water-Filling 注水算法：

min $-\sum_{i}^{n} \log(\alpha_i + x_i)$ （x_i → 信道能量） 凸问题

s.t. $x \succeq 0$, $\mathbf{1}^T x = 1$

写出KKT：① $x^* \succeq 0$

② $\mathbf{1}^T x^* = 1$

③ $\lambda^* \succeq 0$ $\quad(-x^* \le 0)$

④ $\lambda_i^* x_i^* = 0$

⑤ $\cancel{\frac{1}{\alpha_i + x_i}}$ $\frac{-1}{\alpha_i + x_i^*} - \lambda_i^* + v^* = 0$

⑤ ⇒ $\lambda_i^* = v^* - \frac{1}{\alpha_i + x_i^*}$ 代入 ③，④

④ $x_i^*(v_i^* - \frac{1}{\alpha_i + x_i^*}) = 0$ (到这里，⑤ 就消掉了)

③ $v_i^* \geq \frac{1}{\alpha_i + x_i}$

case 1: 若 $v_i^* \geq \frac{1}{\alpha_i}$，

代入 ④ $x_i^* = 0$

case 2: 若 $v_i^* < \frac{1}{\alpha_i}$ ($\alpha_i < \frac{1}{v_i^*}$)

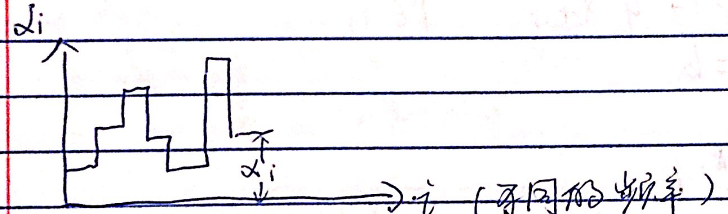
代入 ③

$\frac{1}{\alpha_i + x_i} < \frac{1}{\alpha_i} \Rightarrow x_i > 0$

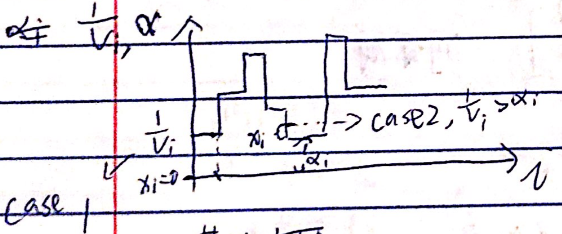
$\therefore v_i^* - \frac{1}{\alpha_i + x_i^*} = 0$

$x_i^* = \frac{1}{v_i^*} - \alpha_i$

到这里，③④ 消失成了一条



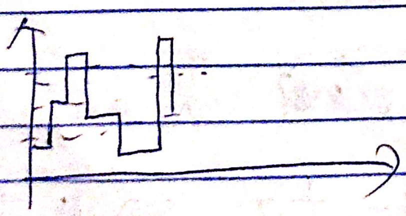
在case 1中



~~灌水条件~~

第1步灌水
① 要灌的水位比原来低，灌了白灌
② ——————————— 高，灌到 $\frac{1}{v}$

第2步灌水



是每个池子水的高度 ✓

$x_i \geq 0$

$\sum x_i = 1$

(灌水会使 x_i 相等)

$x_i = x_j$

KTT 与凸函数一阶可微

例： $\min f_0(x)$ 凸

s.t. $x \geq 0$

$\Rightarrow$ $x \geq 0$,

这是这个问题的最优解的条件 $\Leftarrow$ $\begin{cases} \nabla f_0(x) \geq 0 \\ x_i(\nabla f_0(x)) = 0 \end{cases}$ $\rightarrow$ ~~新最优性条件~~

(互补)

本质相同

和KKT的互补松弛

KTT条件：(充要)

$x^* \geq 0$, $\lambda^* \geq 0$, $\cancel{\lambda_i x^* \leq 0}$ $\lambda_i(-x_i^*) = 0$, $\underline{\nabla f_0(x^*) + (-\lambda^*) = 0}$

$\Downarrow$

$\lambda^* = \nabla f_0(x^*)$

一致

代入

$\nabla f_0(x_0^*) \geq 0$, $x_i \nabla f_0(x^*) = 0$

例： $f_0(Ax+b)$

$L(x) = f_0(Ax+b)$ $g(\ ) = \inf_x f_0(Ax+b)$ ($\max g$) 为对偶问题

$\downarrow$ 变形

$\min f_0(y)$

s.t. $Ax+b=y$

$$L(x, y, v) = f_0(y) + v(Ax - y + b)$$

$$= f_0(y) - v^T y + v^T Ax + v^T b$$

$$g(v) = \inf_{x,y} L = \inf_x \inf_y L \quad (\text{因为 } x \perp y)$$

$$\cancel{= \inf_x (-f_0^*(v)) + v^T Ax}$$

$$= -f_0^*(v) + \underset{\text{or } -\infty}{0} + v^T b$$

对偶问题变为 $\max(g(v)) = \max(-f_0^*(v) + v^T b)$

~~说明同个问题~~

※ 说明同一个问题变化后，对偶问题也会变