国家税务总局徐州市税务局稽查局

Convex Opt HW7

1、(a) 考察 $g(t) = f(X+tV)$，其中 $X+tV \in S^n_{++}$

由于还可以考虑 $g(t_0-t)+t)$，因此可以只考虑小临域上，之后延拓。的凹性. 令 $X^{-\frac{1}{2}} V X^{-\frac{1}{2}} = W$

$$g(t) = tr\left(X^{\frac{1}{2}}(tW+I) X^{\frac{1}{2}}\right)^{-1} = tr\left(X^{-\frac{1}{2}} (I+tW)^{-1} X^{-\frac{1}{2}}\right)$$

而 $\forall W$, $\exists \varepsilon$, s.t. $\forall t \in (-\varepsilon, \varepsilon)$, $\lim_{n\to\infty}(tW)^n = 0$ (可取 $t < \frac{1}{\|W\|_2}$)

则对于 $t \in (-\varepsilon, \varepsilon)$，有:

$$(I+tW)(I - tW + t^2W^2 - t^3W^3 \cdots) = I$$

故 $(I+tW)^{-1} = (I - tW + t^2W^2 \cdots)$，且易知该级数收敛于 $t\in(-\varepsilon,\varepsilon)$

(因为每一个矩阵元都以指数收敛)

故:

$g(t)$
$= tr(X^{-1}) - t\cdot tr(X^{-1}W) + t^2 \cdot tr(X^{-1}W^2) \cdots$ 故存在小收敛域.

且易证其收敛 (t前系数为指数级，$\lim_{n\to\infty} \frac{tr(X^{-1}W^n)}{tr(X^{-1}W^{n-1})} = \lambda_i$)

而考虑 $g(t)$ 在 t^2 前的系数 $tr(X^{-1}W^2) = tr(X^{-\frac{1}{2}} V^2 X^{-\frac{1}{2}})$

$X^{-\frac{1}{2}} V^2 X^{-\frac{1}{2}} = (X^{-\frac{1}{2}}V)(X^{-\frac{1}{2}}V)^T \geq 0$ $\therefore tr(X^{-1}W^2) \geq 0$

故 $g''(t)|_{t=0} \geq 0$, $\forall X, V$. 故平移后可得 g is convex 因此，$f(X)$ is convex.

(b) 令 $g(t) = |X+t\cdot V|^{\frac{1}{n}}$，$X+tV \in S^n_{++}$

则 $g(t) = |X|^{\frac{1}{n}} (|I + X^{-\frac{1}{2}} tV \cdot X^{-\frac{1}{2}}|)^{\frac{1}{n}}$

设 $X^{-\frac{1}{2}} V X^{-\frac{1}{2}}$ 的eigenvalue为 $\lambda_1, \cdots, \lambda_n$

有: $g(t) = |X|^{\frac{1}{n}} \cdot ((1+t\cdot\lambda_1)\cdots(1+t\cdot\lambda_n))^{\frac{1}{n}}$

由几何不等式，在0附近，$1+t\cdot\lambda_i > 0$

$\therefore g(t) \leq |X|^{\frac{1}{n}}(1+\frac{\lambda_1+\cdots+\lambda_n}{n} t)$

而可求得 $g'(0) = |X|^{\frac{1}{n}}(\frac{1}{n}(\lambda_1+\cdots+\lambda_n))$，故 $g(t) \leq g(0)+tg'(0)$，在0附近成立.

$\therefore g'(0) \leq 0$, $\forall X, V$. 故: $f(X)$ is concave.

由于g可平移，故: g is concave

2、 $B_f(x,y) = f(x) - f(y) - \langle \nabla f(y), -y+x \rangle$

$B_f(z,x) = f(z) - f(x) - \langle \nabla f(x), -x+z \rangle$

$B_f(z,y) = f(z) - f(y) - \langle \nabla f(y), -y+z \rangle$

$\therefore B_f(z,x) + B_f(x,y) - B_f(z,y)$

$= -\langle \nabla f(x), z-x \rangle - \langle \nabla f(y), z-y \rangle + \langle \nabla f(y), x-y \rangle$

$= \langle \nabla f(x) - \nabla f(y), x-z \rangle$

3、 $B_f(x,y) = f(x) - f(y) - \langle \nabla f(y), x-y \rangle$, where f is convex and differentiable:

则 $\frac{\partial B_f(x,y)}{\partial x} = \nabla f(x) - \nabla f(y)$, $\frac{\partial^2 B_f(x,y)}{\partial \vec{x} \partial \vec{x}} = \nabla^2 f(x) \geq 0$, for f is convex.

$\therefore B_f(x,y)$ is convex with respect to $\vec{x}$.

而取 $f(x) = e^x$

$B_f(x,y) = e^x - e^y - e^y(x-y) = e^x - e^y(1+x-y)$

取 $x=1$, $B_f(1,y) = e - e^y(2-y)$

可以验证 $-e^y(2-y)$ is not convex.

$\left(\frac{d^2}{dy^2}(-e^y(2-y)) = ye^y\right.$ 不定号).

4、 $\nabla \phi(\vec{x}) = \frac{\beta}{2}\|A\|_2^2 \nabla \|x-u\|_2^2 - \frac{\beta}{2}\nabla \|Ax-v\|^2$

而 $\nabla \|W\vec{x} - \vec{w}\|_2^2 = 2W^T(W\vec{x} - \vec{w})$

故 $\nabla \phi = \frac{\beta}{2}\|A\|_2^2 \cdot 2(\vec{x} - \vec{u}) - \frac{\beta}{2} 2A^T(A\vec{x} - v)$

$= \beta(\|A\|_2^2 - A^TA)\vec{x} - \beta\|A\|_2^2 u + \beta A^TAv$

因此 $B_\phi(x,x')$

$= \frac{\beta}{2}\|A\|_2^2\left(\|x-u\|^2 - \|x'-u\|^2\right) - \frac{\beta}{2}\left(\|Ax-v\|^2 - \|Ax'-v\|^2\right)$

$- (x-x')^T\left(\beta(\|A\|_2^2 - A^TA)\vec{x'} - \beta\|A\|_2^2 u + \beta A^TAv\right)$

<u>展开</u> $\frac{\beta}{2}\|A\|_2^2\left(|x|^2 - x^Tx'\right) - \frac{\beta}{2}\left(|Ax|^2 - x^TA^TAx\right)$, 与 u, v 无关.

5、否. 设 $f(x)=x^2$

$$B_f(1,2)=f(1)-f(2)-f'(2)\cdot(1-2)=-3+4=1$$

$$B_f(2,3)=f(2)-f(3)-f'(3)(2-3)=-5+6=1$$

$$B_f(1,3)=f(1)-f(3)-f'(3)(1-3)=-8+2\times6=4$$

$\therefore B_f(1,2)+B_f(2,3)<B_f(1,3)$

~~并~~非 triangular、

6、(1) 由 $\partial(f_1+f_2)=\partial f_1+\partial f_2$.

而 $\frac{1}{2}x_1^2$ 可导, $\therefore \partial f_1(\vec{x})=\{(x_1,0)^T\}$

对于 ∂f_2, 在 $x_2\neq0$ 处可导, $\partial f_2(\vec{x})=\mathrm{sgn}(x_2)$

在 $x_2=0$ 时, 若令 $|x_2'|\geq\langle g,\begin{bmatrix}x_1'\\x_2'\end{bmatrix}\rangle$, $\forall\begin{bmatrix}x_1'\\x_2'\end{bmatrix}\in\mathbb{R}^2$

则可知 $g=\{\begin{bmatrix}0\\x\end{bmatrix}, x\in[-1,1]\}$

$$\therefore \partial f=\begin{cases}\{\begin{bmatrix}x_1\\\mathrm{sgn}(x_2)\end{bmatrix}\}, & x_2\neq0\\ \{\begin{bmatrix}x_1\\x\end{bmatrix}, x\in[0,1]\} & x_2=0\end{cases}$$

(2) 对于 $\mathbb{R}/\{0,1\}$, $f(x)=\max(x,x^2)$ 可导, $\partial f=\begin{cases}\{2x\} & x\in(-\infty,0)\cup(1,+\infty)\\ \{1\} & x\in(0,1)\end{cases}$

对于 $x=0$, 需有: $f(x)\geq g\cdot x$, $\forall x$

故 $g\in[0,1]$

对于 $x=1$, 需有: $f(x)\geq1+g\cdot x$, $\forall x$

故 $g\in[1,2]$

$$\therefore \partial f=\begin{cases}\{2x\}, & x\in(-\infty,0)\cup(1,+\infty)\\ \{1\}, & x\in(0,1)\\ [0,1], & x=0\\ [1,2], & x=1\end{cases}$$

(3) Lemma: 对于 $\max\{f_i(x)\}$, 有: $\partial(\max f_i(x)) = \mathrm{conv}\left(\bigcup_{i\in I(x)} \partial f_i(x)\right)$

其中 $I(x)=\{i \mid f_i(x)=\max_j\{f_j(x)\}\}$ (写完这个 Lemma 才发现教材 174~175 页上有…)

Proof: $\forall g_i \in \partial f_i,\ i\in I(x)$

有: $\forall x',\ f_i(x')-f_i(x) \geq \langle g_i, x'-x\rangle$

而 $i\in I(x) \Leftrightarrow f_i(x)=f(x),\ f_i(x')\leq f(x')$

$\therefore f(x')-f(x) \geq \langle g_i, x'-x\rangle,\ \forall x',\ \forall g_i\in\partial f_i, i\in I(x)$

故 $\forall g\in \mathrm{conv}\left(\bigcup_{i\in I(x)}\partial f_i(x)\right)$, $f(x')-f(x) \geq \langle \sum\theta_i g_i, x'-x\rangle = \langle g, x'-x\rangle$ $\forall x'$

$\therefore \mathrm{conv}\left(\bigcup_{i\in I(x)}\partial f_i(x)\right) \subseteq \partial f$

若 $g\in\partial f$ 且 $g\notin \mathrm{conv}\left(\bigcup_{i\in I(x)}\partial f_i(x)\right)$

则由分离超平面, $\exists y$, s.t.

$\langle g, y\rangle > \langle \partial f_i(x), y\rangle$, $\forall i\in I(x)$

$\therefore \langle g,y\rangle > \max_{i\in I(x)} \langle y, \partial f_i(x)\rangle$, $\langle g,y\rangle > \partial f$

矛盾. 故得证.

在此，$\|x\|_\infty = \max_{i=1,\dots,n} |e_i^T x|$

① $\vec{x}=0$，则

$$\partial f = \operatorname*{conv}_{i=1,\dots,n} \{ a_i \cdot e_i,\ a_i \in [-1,1]$$

$$= \{ \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix},\ a_i \in [-1,1] \}$$

② $\vec{x} \neq 0$，假设共 r 个 x_i 有：$|x_i| = \|\vec{x}\|_\infty$

记为 $x_{a_1}, \dots, x_{a_r}$，为取到最大绝对值的元素

则 $\partial f = \operatorname*{conv}_{i=1,\dots,r} \{ e_{a_i} \cdot \operatorname{sgn}(x) \}$

$$= \{ \sum_{i=1}^{r} \theta_i e_{a_i} \cdot \operatorname{sgn}(x_{a_i}) \mid 其中 \sum \theta_i = 1,\ \theta_i \geq 0 \}$$

7. ① $g_1 \in \partial f(x_1)$，则 $\forall x$，$f(x) - f(x_1) \geq \langle g_1, x - x_1 \rangle$

取 $x = x_2$，$f(x_2) - f(x_1) \geq \langle g_1, x_2 - x_1 \rangle$ [1]

同理，$f(x_1) - f(x_2) \geq \langle g_2, x_1 - x_2 \rangle$ [2]

[1] + [2] 得：$\langle g_1 - g_2, x_2 - x_1 \rangle \leq 0$

即：$\langle g_1 - g_2, x_1 - x_2 \rangle \geq 0$

② 若 $f(x)$ 为 μ-strongly convex

则有：$\forall x$，$g_1 \in \partial f(x_1)$，$f(x) - f(x_1) \geq \langle g_1, x - x_1 \rangle + \frac{\mu^2}{2} \|x - x_1\|^2$

取 $x = x_2$

$f(x_2) - f(x_1) \geq \langle g_1, x_2 - x_1 \rangle + \frac{\mu^2}{2} \|x_2 - x_1\|^2$ [1]

同理，$f(x_1) - f(x_2) \geq \langle g_2, x_1 - x_2 \rangle + \frac{\mu^2}{2} \|x_1 - x_2\|^2$ [2]

[1] + [2] 得：$\langle g_1 - g_2, x_1 - x_2 \rangle \geq \mu \|x_1 - x_2\|^2$