国家税务总局徐州市税务局稽查局

Convex Opt HW 8

1、 $\partial\|x\|_1^2$， $\|x\|_1^2=(|x_1|+\cdots+|x_n|)^2$

不妨令 $x_{a_1},\cdots,x_{ar}$ 为 0，剩余不为 0

则记 $e_i^k=-1$ 或 1，有：$\|x\|_1^2=\max\limits_k(e_1^k x_1+\cdots+e_n^k x_n)^2$

而右侧可微，故由 Danskin rule，$I(k)=\{\vec{e}^k \mid e_i=\mathrm{sgn}(x_i)$ 或 $\forall i, e_i=-\mathrm{sgn}(x_i)$，$\forall i, i\neq a_j$ 而 for $i=a_j$，e_i 任意$\}$

而 $\partial(e_1^k x_1+\cdots+e_n^k x_n)^2$

$$=2(e_1^k x_1+\cdots e_n^k x_n)\begin{bmatrix}e_1^k\\ \vdots\\ e_n^k\end{bmatrix}$$ 对于 $k\in I(k)$，$\vec{e}^{k\,T}\vec{x}=\pm\|x\|_1$

$$\therefore \partial\|x\|_1^2=\mathrm{conv}\left\{2\|x\|_1\begin{bmatrix}e_1^k\\ \vdots\\ e_n^k\end{bmatrix}\right\}=\{2\|x\|_1\vec{e}\}, \text{其中 } \vec{e}_i=\begin{cases}\mathrm{sgn}(x_i), & x_i\neq 0\\ [-1,1], & x_i=0\end{cases}$$

2、 $\|x\|_M=\sqrt{x^TMx}$

设 $f(\vec{x})=\|x\|_2$．则 $\|x\|_M=f(M^{\frac{1}{2}}x)$，f is convex & differential

① $x\neq 0$， $\partial f(x)=\frac{1}{\|x\|_2}\vec{x}$ $\therefore \partial\|x\|_M=(M^{\frac{1}{2}})^T\frac{1}{\sqrt{x^TMx}}(M^{\frac{1}{2}}\vec{x})$

$$=\frac{1}{\|x\|_M}M\vec{x}$$

② $x=0$，此时 $\partial f(x)=\{g\in\mathbb{R}^n \mid \|g\|_2\leq 1\}$ $\therefore \partial\|x\|_m=\{Mx \mid x^TMx\leq 1\}$

$$=\{y \mid y^TM^{-1}y\leq 1\}$$

3、 $\max\vec{x}=\max\limits_i(e_i^T\vec{x})$

设 x_i，$i=a_1,\cdots,ar$ 处取最值，故由 danskin's rule，

$$\partial\max\vec{x}=\mathrm{conv}\{\vec{e}_i \mid i=a_1,\cdots,ar\}$$

$$=\left\{\vec{e} \;\middle|\; e_i=\begin{cases}0, & i\neq a_j\\ \theta_{a_j}, & i=a_j\end{cases}, \text{其中 } \theta_{a_j}\geq 0, \sum\theta_{a_j}=1\right\}$$

4、 $\|X\|_2 = \max_{\|a\|_2=1} \|Xa\|_2$

而 $\partial \|Xa\|_2 = \frac{1}{\|Xa\|_2} X a a^T$

且若 $\|Xa\|_2 = \max_{\|b\|_2=1} \|Xb\|_2$，则需有：设 $X = U \Sigma V^T$

$X\vec{a} = b \cdot \sigma_1$，$\|b\|_2=1$

$\Sigma = \begin{bmatrix} \underbrace{\sigma_1 \ \cdots \ \sigma_1}_{k个} & \cancel{\sigma_2} & \ddots \end{bmatrix}$，则需 $a \in I = \{ \|a\|_2 = 1, \cancel{a \in Null(X - \sigma_1 I)} \}$

故 $\partial \|X\|_2 = \underset{a \in I}{conv} \left\{ \frac{X a a^T}{\|X\|_2} \right\}$，而 for $a \in I$，$X\vec{a} = \sigma_1 \vec{b} = \|X\|_2 \vec{b}$

~~其中~~ $\|b\|_2 = 1$

对于满足 $\underset{(\|a\|=1)}{\|Xa\|_2} = \|X\|_2$ 的 a，由 SVD，有：（设等于 σ_1 的奇异值有 k 个）

$a \in span\{u_1 \cdots u_k\}$，$b \in span\{v_1, \cdots, v_k\}$，$Xa = b$，$\|a\|_2 = \|b\|_2 = 1$

u_i, v_i 为左右特征向量. 记 $[u_1, \cdots, u_k] = U$，$[v_1, \cdots, v_k] = V$

则以上可写作：$\partial \|X\|_2 = conv\{ U\alpha \cdot (V\alpha)^T \mid \|\alpha\|_2 = 1 \}$

由于 $conv\{\alpha\alpha^T \mid \|\alpha\|_2 = 1\} = \{W \mid W \succeq 0, trW = 1\}$

故可知：$\partial \|X\|_2 = \{ UWV^T \mid W \succeq 0, trW = 1 \}$，$X \neq 0$

同理，若 $X = 0$，$\partial \|X\|_2 = conv\{ uv^T \mid \|u\|_2 = \|v\|_2 = 1\} = \{ W \mid \|W\|_* \leq 1 \}$

5、 $f(x), g(x) > 0$. 则 $\forall \theta \in [0,1]$，$x, y \in J^\circ$. $f(\theta x + (1-\theta)y)\, g(\theta x + (1-\theta)y)$ 有：

$f(\theta x + (1-\theta)y)\, g(\theta x + (1-\theta)y) \geq (\theta f(x) + (1-\theta) f(y))(\theta g(x) + (1-\theta) g(y))$

$= \theta^2 f(x) g(x) + \theta(1-\theta)(f(x)g(y) + g(x)f(y)) + (1-\theta)^2 f(y) g(y)$

$\geq \theta^2 f(x) g(x) + \theta(1-\theta)(f(x)g(x) + f(y)g(y)) + (1-\theta)^2 f(y) g(y) = \theta f(x) g(x) + (1-\theta) f(y) g(y)$

（排序不等式）

故 $f(x)g(x)$ is convex.

① $f \partial g + g \partial f \subseteq \partial(fg)$

proof: $\forall a \in f\partial g + g \partial f$，$a = b + c$，$b \in f \partial g$，$c \in g \partial f$

则 $\frac{b}{f(x)} \in \partial g$，$\frac{c}{g(x)} \in \partial f$. 有：$f(x+y)\, g(x+y) \geq \left(f(x) + \left(\frac{c}{g(x)}\right)^T y\right) \cdot \left(g(x) + \left(\frac{b}{f(x)}\right)^T y\right)$

$\geq f(x) g(x) + (b+c)^T y$

$= f(x) g(x) + a^T y$

$\therefore$ $a \in \partial(fg)$，得证.

② $\partial(fg) \subseteq f\partial g + g\partial f$, proof:

取 $\alpha \in \partial(fg)$，有：$f(x+y)g(x+y) \geq f(x)g(x) + \alpha \cdot y,\ \forall y$.

由于 $\partial f, \partial g$ 为 convex & closed set，不妨设 $\partial f = [a, b]$，$\partial g = [c, d]$

则 $f\partial g + g\partial f = [cf(x) + ag(x),\ df(x) + bg(x)]$

在 x_0 处，考虑函数 $f(x_0)g(x) + g(x_0)f(x) \triangleq F(x)$

$$\partial F(x) = f(x_0)\partial g(x) + g(x_0)\partial f(x) = [cf(x_0) + ag(x_0),\ df(x_0) + bg(x_0)] = f\partial g + g\partial f$$

故若 $\alpha \notin f\partial g + g\partial f$，$\exists \varepsilon$, s.t. $f(x)g(x+y) + g(x)f(x+y) < f(x)g(x) + \alpha y,\ \forall y \in [0, \varepsilon)$

记 $f(x+y) - f(x) = \Delta f$，由于 f, g 在 J° 上连续，故有：$\forall \varepsilon > 0$，$\exists y'$, s.t. $\Delta f, \Delta g < \varepsilon$

故当 y' 足够小时，一定有：$f\Delta g + g\Delta f + \Delta f\Delta g < \alpha y'$，与 $\alpha \in \partial fg$ 矛盾.

综上，$\partial fg = f\partial g + g\partial f$

6. 由排序不等式：对于 $a_1 \leq a_2 \cdots \leq a_n$，$b_1 \leq \cdots \leq b_n$

$$\sum_{i=1}^{n} a_i b_i \geq \sum_{i=1}^{n} a_{c_i} b_{d_i},\quad c_i, d_i \text{为其它顺序}$$

故取 $a_n = \alpha_1, \cdots, a_{n+1-r} = \alpha_r, a_{n+r} = 0, \cdots$

$b_n = x_{[1]}, \cdots, b_1 = x_{[n]}$

有：$f(\vec{x}) = \max\limits_{\substack{k=1,\dots,n! \\ \text{为全部组合}}} \vec{a}^{k\,T} x$，而 $\vec{a}^{i\,T} x$ 为线性函数，故 convex

$\therefore f(\vec{x})$ 为一系列凸函数最大值

$\therefore f(x)$ is convex

7. 设 $g(\vec{y}) = \log(\sum_{i=1}^{n} e^{y_i})$，其为 convex

则 $\log(\sum_{i=1}^{m} e^{a_i^T x + b_i}) = g(Ax+b)$ 为线性变换，为 convex

在 dom f 上，$-\log(x)$ is convex & non increasing

$-g(Ax+b)$ is concave

故 $-\log(-g(Ax+b))$ is convex on dom f

8. $h(\vec{x}) = 1/(1+e^{-w^T x})$。又因为若 $F(x)$ 为凸，则 $F(\alpha^T \vec{y})$ 关于 $\vec{y}$ 也为凸.

设 $-\ln(h(\vec{x})) = g_1(w^T x)$，其中 $g_1(y) = -\ln(1/(1+e^{-y})) = \ln(1+e^{-y})$

$g_1''(y) = \frac{e^y}{(1+e^y)^2} > 0$ ∴ $g_1(y)$ is convex ∴ $-\ln(h(\vec{x})) = g_1(w^T x)$ is convex with respect to $\vec{w}$.

设 $-\ln(1-h(\vec{x})) = g_2(w^T x)$

$g_2(y) = -\ln(1-1/(1+e^{-y}))$，$g_2''(y) = \frac{e^y}{(1+e^y)^2} > 0$

∴ $g_2(y)$ is convex，∴ $-\ln(1-h(\vec{x})) = g_2(w^T x)$ is convex with respect to $\vec{w}$.

故：$H(w)$ 是以上两种凸函数的和

∴ $H(w)$ is convex.