优化作业5

1 (a) 设 $Z\in S_{++}^n$. $V\in S_n$ 取 $t$ s.t. $Z+tV\in S_{++}^n$

$$g(t)=f(Z+tV)=tr((Z+tV)^{-1})=tr\left[Z^{\frac12}(I+tZ^{-\frac12}VZ^{-\frac12})Z^{\frac12})^{-1}\right]$$

$$=tr(Z^{-\frac12}(I+tZ^{-\frac12}VZ^{-\frac12})^{-1}Z^{-\frac12})$$

对 $W=Z^{-\frac12}VZ^{-\frac12}$ 作特征值分解 $W=U\Lambda U^T$. $\Lambda$ 对角. $UU^T=I$.

则 $g(t)=tr\, Z^{-\frac12}(U(I+t\Lambda)U^T)^{-1}Z^{-\frac12})$

$$=tr((I+t\Lambda)^{-1}U^TZ^{-1}U)\overset{M=U^TZ^{-1}U>0}{=\!=\!=}tr((I+t\Lambda)^{-1}M)$$

$$=\sum_{i=1}^n\frac{m_{ii}}{1+t\lambda_i}\qquad m_{ii}\text{是}M\text{的对角元}.\ \lambda_i\text{是}W\text{的特征值}.$$

$$\therefore g'(t)=\sum_{i=1}^n\frac{-\lambda_i m_{ii}}{(1+t\lambda_i)^2}\qquad g''(t)=\sum_{i=1}^n\frac{2\lambda_i^2m_{ii}(1+t\lambda_i)}{(1+t\lambda_i)^4}=\sum_{i=1}^n\frac{2\lambda_i^2m_{ii}}{(1+t\lambda_i)^3}$$

$\because M=U^TZ^{-1}U,\ Z\in S_{++}^n\ \therefore M\in S_{++}^n\Rightarrow m_{ii}>0$

$(I+t\Lambda)=U(I+t\Lambda)U^T=I+tW=Z^{-\frac12}(Z+tV)Z^{-\frac12}\quad\because Z+tV\in S_{++}^n$

$\Rightarrow I+t\Lambda\in S_{++}^n\Rightarrow(I+t\Lambda)^{-1}\in S_{++}^n\Rightarrow 1+t\lambda_i>0.$

$\therefore g''(t)>0\ \forall t$ s.t. $Z+tV\in S_{++}^n\qquad\therefore f(X)=tr(X^{-1})$ 凸.

(b) $Z\in S_{++}^n$. $V\in S_n$. $Z+tV\in S_{++}^n$ $t\in R$.

$$g(t)=f(Z+tV)=(\det(Z+tV))^{\frac1n}=(\det(Z^{\frac12}(I+tZ^{-\frac12}VZ^{-\frac12})Z^{\frac12})^{\frac1n}$$

$$=\left[\det Z^{\frac12}\cdot\det(I+tZ^{-\frac12}VZ^{-\frac12})\det Z^{\frac12}\right]^{\frac1n}$$

$$=C\cdot\left[\det(I+t\Lambda)\right]^{\frac1n}$$

$C=(\det Z^{\frac12}\cdot\det Z^{\frac12}\cdot\det UU^T)^{\frac1n}$ 是常数.

其中 $Z^{-\frac12}VZ^{-\frac12}=U\Lambda U^T$ 是特征值分解.

$\therefore g(t)=C\left(\prod_{i=1}^n(1+t\lambda_i)\right)^{\frac1n}$ 同上题知 $1+t\lambda_i>0$

设 $s(t)=\left[\prod(1+t\lambda_i)\right]^{\frac1n}\quad\therefore s(t)>0$

$\ln s(t)=\frac1n\sum\ln(1+t\lambda_i)\qquad(\ln s(t))'=\frac1n\sum\frac{\lambda_i}{1+\lambda_i t}$

$\Rightarrow\frac{d\ln s(t)}{dt}=\frac{s'(t)}{s(t)}\Rightarrow s'(t)=s(t)(\ln s(t))'=\left[\prod(1+t\lambda_i)\right]^{\frac1n}\cdot\frac1n\sum\frac{\lambda_i}{1+t\lambda_i}$

$s''(t)=s'(t)(\ln s(t))'+s(t)\cdot(\ln s(t))''$

$$=s(t)\left[(\ln s(t))'\right]^2+s(t)\cdot\frac1n\sum\frac{-\lambda_i^2}{(1+t\lambda_i)^2}=s(t)\left(\frac{1}{n^2}\left(\sum\frac{\lambda_i}{1+\lambda_i t}\right)^2-\frac1n\sum\frac{\lambda_i^2}{(1+t\lambda_i)^2}\right)$$

由柯西不等式 $n\sum\frac{\lambda_i^2}{(1+t\lambda_i)^2} \ge \left(\sum\frac{|\lambda_i|}{1+\lambda_i t}\right)^2 \underset{\substack{\Downarrow \\ 1+t\lambda_i>0}}{\ge} \left(\sum\frac{\lambda_i}{1+t\lambda_i}\right)^2$

$\therefore\ s''(t)\le 0 \quad\Rightarrow\quad g''(t)\le 0 \quad\Rightarrow$ ~~g(x)~~ $f(x)$ 是凹函数.

2. $B_f(z,x)+B_f(x,y)-B_f(z,y)$

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

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

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

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

3. $B_f(x,y) = f(x)-f(y)-\langle\nabla f(y), x-y\rangle$.

$= f(x)+h(x) \qquad h(x) = -f(y)-\langle\nabla f(y), x-y\rangle$ 是仿射函数 $\Rightarrow$ 是凸的

又 $f(x)$ 凸 $\quad\therefore$ ~~即~~ $B_f(x,y)=f(x)+h(x)$ 关于 $x$ 是凸的

取 ~~f(x,y)=~~ $f(x) = -\sum_{i=1}^{n}\ln x_i \qquad$ 则 $d_f(x,y) = -\sum\left(\ln\frac{x_i}{y_i} - 1 + \frac{x_i}{y_i}\right)$

$\nabla_y^2 d_f(x,y) = -\sum\frac{1}{y_i^2} - \frac{2x_i}{y_i^3} = -\sum\frac{y_i-2x_i}{y_i^3} \qquad$ 取 $y_i > \frac{x_i}{2} > 0\ \forall i$

则 $\nabla_y^2 d_f(x,y) < 0 \quad\Rightarrow$ 关于 $y$ 不一定凸

4. 引理：

设 $\varphi(x) = \mu(x) + a + b^T x$

则 $B_\varphi(x,y) = B_\mu(x,y)$.

这是因为 $B_\varphi(x,y) = \varphi(x) - \varphi(y) - \langle \nabla\varphi(y), x-y \rangle$

$= \mu(x) + b^T x - \mu(y) - b^T y - \langle \nabla\mu(x) + b, x-y \rangle$

$= \mu(x) - \mu(y) - \langle \nabla\mu(x), x-y \rangle = B_\mu(x,y)$.

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

注意到含 $u$、$v$ 的均为常数或一次项. 由引理知 $B_\phi(x,y) = B_{\phi_0}(x,y)$

~~$B_{\phi(x)} \Rightarrow B$~~ ~~$B_{\phi(x)} \Rightarrow B$~~ 其中 $\phi_0(x) = \frac{\beta}{2}(\|A\|_2^2 x^T x - x^T A^T A x)$ 与 $u$、$v$ 无关

$\Rightarrow B_{\phi_0}(x,y)$ 与 $u$、$v$ 无关 $\Rightarrow B_\phi(x,y)$ 与 $u$、$v$ 无关.

5. ① $f(x) = \frac{1}{2}x_1^2 + |x_2|$

设 $x \in \mathbb{R}^n$

$$\partial f = \partial(\tfrac{1}{2}x_1^2) + \partial|x_2| = \begin{cases} \{g \mid g_1 = x_1,\ g_2 = 1,\ g_i = 0\ i \ge 3,\ g \in \mathbb{R}^n\} & x_2 > 1 \\ \{g \mid g_1 = x_1,\ g_2 = -1,\ g_i = 0,\ i \ge 3,\ g \in \mathbb{R}^n\} & x_2 < 1 \\ \{g \mid g_1 = x_1\ -1 \le g_2 \le 1,\ g_i = 0\ i \ge 3,\ g \in \mathbb{R}^n\} & x_2 = 0 \end{cases}$$

② $f(x) = \max(x, x^2)$.

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

(Danskin's)

③ $\|X\|_{2,1} = \sum_{i=1}^n \|X_i\|_2$. $X = (X_1, \dots, X_n)$. $X \in \mathbb{R}^{m\times n}$

$\partial\|X\|_{2,1} = \sum_{i=1}^n \partial\|X_i\|_2$.

$$\partial\|X_i\|_2 = \begin{cases} \{y \mid y = \frac{X_i}{\|X_i\|_2} \in \mathbb{R}^m\} & X_i \ne 0 \\ \{y \mid \|y\|_2 \le 1,\ y \in \mathbb{R}^m\} & X_i = 0 \end{cases}$$

$\therefore \partial\|X\|_{2,1} = \{G = (G_1, G_2, \dots, G_n) \mid G_i \in \partial\|X_i\|_2,\ G_i \in \mathbb{R}^{m\times n}\}$.

6. 对$\mu$强凸的$f$，有 $f(y)\ge f(x)+\langle g,y-x\rangle+\frac{\mu}{2}\|y-x\|^2 \quad \forall g\in\partial f(x)$

则 $f(x_1)\ge f(x_2)+\langle g_2,x_1-x_2\rangle+\frac{\mu}{2}\|x_1-x_2\|^2 \quad g_2\in\partial f(x_2)$

$f(x_2)\ge f(x_1)+\langle g_1,x_2-x_1\rangle+\frac{\mu}{2}\|x_1-x_2\|^2 \quad g_1\in\partial f(x_1)$

相加有 $0\ge\langle g_2-g_1,x_1-x_2\rangle+\mu\|x_1-x_2\|^2 \quad g_i\in\partial f(x_i)\ i=1,2$

也即 $\langle g_1-g_2,x_1-x_2\rangle\ge\mu\|x_1-x_2\|^2 \quad g_i\in\partial f(x_i)\ i=1,2$

将$\mu=0$代入即得$f$凸的时 $\langle g_1-g_2,x_1-x_2\rangle\ge 0 \quad \forall g_i\in\partial f(x_i)\ i=1,2$

7. 设 $g_j(x)=\sum_{i=1}^{j}x[i] \quad j=1,2\ldots r$ （$\alpha_i=w_i\ 1\le i\le r$ 用错字母了）

则 $f(x)=w_r g_r(x)+(w_{r-1}-w_r)g_{r-1}(x)+(w_{r-2}-w_{r-1})g_{r-2}(x)+\cdots+(w_1-w_2)g_1(x)$

$g_j(x)$ 是凸的 $1\le j\le r$，（课上的结论） $w_r\ge 0 \quad w_j-w_{j+1}\ge 0$

$\therefore f(x)$是凸函数的非负线性加和 $\Rightarrow f(x)$凸

8. $f_k(X)=\max\limits_{\substack{U\subseteq\mathbb{R}^n\\ \dim(U)=k}}\sum_{i=1}^{k}\|Xu_i\|_2$ 其中$u_1,u_2\ldots u_k$是$U$的标准正交基

则 $f_k(x)$是凸函数的最大值 $\Rightarrow f_k(x)$是凸函数

补充：对奇异值分解 $X=S\Sigma V^T \quad V=[v_1,v_2\ldots v_n]$是右奇异向量矩阵

取 $U=\mathrm{span}\{v_1,v_2\ldots v_k\}$ 有一组标准正交基 $v_1,v_2\ldots v_k$

~~有 $Xv_i=X(\sum c_jv_j)=\sum c_iXv_i=\sum c_i\sigma_iv_i=$~~

有 $Xv_i=\sigma_is_i \quad S=(s_1,s_2\ldots s_n)$是左奇异矩阵，$\|s_i\|_2=1$

$\therefore \sum_{i=1}^{k}\|Xv_i\|_2=\sum_{i=1}^{k}\|\sigma_is_i\|_2=\sum_{i=1}^{k}|\sigma_i|=\sum_{i=1}^{k}\sigma_i \quad \therefore \exists U$和$u_1,u_2\ldots u_k$, s.t. $\sum_{i=1}^{k}\|Xu_i\|=\sum_{i=1}^{k}\sigma_i$

其他的$U$和$u_i$均小于 $\sum_{i=1}^{k}\sigma_i$ 可由根据$\sigma_i$的序关系证得