HW2

方科晨

2024 年 3 月 16 日

Problem1.

(a) A 是凸集，B 不是凸集。$f''(x) = -2$ ，所以 f 是凹函数，因此 $\forall (x_1, y_1), (x_2, y_2) \in A, \forall 0 \le \lambda \le 1$ ，对于点 $(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2)$ 有 $1 = \lambda \cdot 1 + (1-\lambda) \cdot 1 \le \lambda x_1 + (1-\lambda)x_2 \le \lambda \cdot 3 + (1-\lambda) \cdot 3 = 3$ 且 $0 = \lambda \cdot 0 + (1-\lambda) \cdot 0 \le \lambda y_1 + (1-\lambda)y_2 \le \lambda f(x_1) + (1-\lambda) f(x_2) \le f(\lambda x_1 + (1-\lambda)x_2)$ ，因此 $(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \in A$ ，所以 A 是凸集。

$g''(x) = -2$,g 是凸函数,且不等式不取等,因此对于点 $(1, g(1)), (3, g(3)) \in B$ ，对于 $\forall 0 \le \lambda \le 1$ 都有 $1 \le \lambda \cdot 1 + (1-\lambda) \cdot 3 \le 3$ 且 $\lambda g(1) + (1-\lambda) g(3) > g(\lambda \cdot 1 + (1-\lambda) \cdot 3)$ ，因此 $(\lambda \cdot 1 + (1-\lambda) \cdot 3, \lambda g(1) + (1-\lambda) g(3)) \notin B$ ，故 B 不是凸集。

(b) g 是凸函数 f 不是，由 (1) 中求导可得。

Problem2.

$\Rightarrow$：若 Ω 是一个 convex cone，则 Ω 既是一个锥也是一个凸集。由于 Ω 是锥，所以 $\forall x \in \Omega, \forall \lambda > 0$ 都有 $\lambda x \in \Omega$ 。而且 $\forall x, y \in \Omega, \frac{1}{2}x, \frac{1}{2}y \in \Omega$ ，由于 $\frac{1}{2} + \frac{1}{2} = 1$ 且 Ω 为凸集，所以 $\frac{1}{2}x + \frac{1}{2}y \in \Omega$ ，故 $x + y = 2(\frac{1}{2}x + \frac{1}{2}y) \in \Omega$

$\Leftarrow$：由于 $\forall x \in \Omega, \forall \lambda > 0$ 都有 $\lambda x \in \Omega$ ，所以 Ω 是一个锥。其次 $\forall x, y \in \Omega, \forall 0 \le \lambda \le 1$ 。当 $\lambda = 0, \frac{1}{2}, 1$ 时是 trival 的。当 $0 \le \lambda < \frac{1}{2}$ 时，有 $\lambda x + (1-\lambda)y = \lambda(x+y) + (1-2\lambda)y$ ，其中由于 $x, y \in \Omega$ 故

$x+y \in \Omega$ 故 $\lambda(x+y) \in \Omega$ ，又 $1-2\lambda > 0$ 则 $(1-2\lambda)y \in \Omega$，两者相加可得 $\lambda(x+y)+(1-2\lambda)y \in \Omega$ 。当 $\frac{1}{2} < \lambda < 1$ 时同理可得。综上，Ω 是凸集，故 Ω 是 convex cone。

Problem3.

首先证明 $\partial f(\hat{x})$ 是凸集。考虑任意 $g_1, g_2 \in \partial f(\hat{x})$ 以及任意 $0 \le \alpha \le 1$ ，则有 $f(x) = \alpha f(x) + (1-\alpha) f(x) \ge \alpha(f(\hat{x}) + g_1^T(x-\hat{x})) + (1-\alpha)(f(\hat{x}) + g_2^T(x-\hat{x})) = (\alpha + (1-\alpha)) f(\hat{x}) + (\alpha g_1^T + (1-\alpha) g_2^T)(x-\hat{x}) = f(\hat{x}) + (\alpha g_1^T + (1-\alpha) g_2^T)(x-\hat{x})$ 。故有 $\alpha g_1^T + (1-\alpha) g_2^T \in \partial f(\hat{x})$ 。因此 $\partial f(\hat{x})$ 是凸集。

由于 $\partial f(\hat{x})$ 中的条件为等号，所以显然 $\partial f(\hat{x})$ 为闭集，任意 $\partial f(\hat{x})$ 中点列的极限仍然在 $\partial f(\hat{x})$ 中。

f 是凸函数，则 epi$f$ 是凸集，且 $(\hat{x}, f(\hat{x})) \in \partial \mathrm{epi} f$ 。由支撑超平面，可以找到 $(\boldsymbol{a}, b) \ne \mathbf{0}, (\boldsymbol{a}, b) \in \mathbb{R}^{n+1}$ 使得 $(\boldsymbol{a}, b)^T(x, f(x)) \ge (\boldsymbol{a}, b)^T(\hat{x}, f(\hat{x})), \forall x \in \mathbb{R}^n$ 。如果 $b = 0$ ，则式子变为 $\boldsymbol{a} \cdot x \ge \boldsymbol{a} \cdot \hat{x}, \forall x \in \mathbb{R}^n$ ，这显然不可能成立，故 $b \ne 0$ 。则原式两边同除以 b 就有 $(\frac{\boldsymbol{a}}{b}, 1)^T(x, f(x)) \ge (\frac{\boldsymbol{a}}{b}, 1)^T(\hat{x}, f(\hat{x})), \forall x \in \mathbb{R}^n$ ，由此可见 $(\frac{\boldsymbol{a}}{b}, 1) \in \partial f(\hat{x})$ ，因此 $\partial f(\hat{x})$ 非空。

综上，$\partial f(\hat{x})$ 为非空闭凸集。

Problem4.

考虑 $x, y \in \bar{C}$ ，则存在两个点列 $\{x_n\}_{n=1}^\infty \subset C, \{y_n\}_{n=1}^\infty \subset C$ 使得 $\lim_{n\to\infty} x_n = x, \lim_{n\to\infty} y_n = y$ 。则对于任意 $0 \le \alpha \le 1$ 。由于 C 为凸集，故有 $z_n = \alpha x_n + (1-\alpha) y_n \in C, \forall n \ge 1$ ，则 $z = \alpha x + (1-\alpha) y = \alpha \lim_{n\to\infty} x_n + (1-\alpha) \lim_{n\to\infty} y_n = \lim_{n\to\infty}(\alpha x_n + (1-\alpha) y_n) \in \bar{C}$ 。因此 $\bar{C}$ 为凸集。

Problem5.

(a) 对于任意的 $0 \le \alpha \le 1$ 和 $x, y \in \mathbb{R}^n$ 有 $h(\alpha x + (1-\alpha) y) = g(f(\alpha x +$

$(1-\alpha)y))$ 。由于 $f(\alpha x+(1-\alpha)y) \leq \alpha f(x)+(1-\alpha)f(y)$ 及 g 为单调非降函数，故 $h(\alpha x+(1-\alpha)y) \leq g(\alpha f(x)+(1-\alpha)f(y)) \leq \alpha g(f(x))+(1-\alpha)g(f(y)) = \alpha h(x)+(1-\alpha)h(y)$ ，所以 h 是凸函数。

(b) 令 $f,g:\mathbb{R}\to\mathbb{R}, f(x)=x^2-1, g(x)=x^2, h=g\circ f$ 。我们有 $h(0)=g(f(0))=g(-1)=1, h(1)=g(f(1))=g(0)=0, h(\frac{1}{2})=g(f(\frac{1}{2}))=g(-\frac{3}{4})=\frac{9}{16}$ 。可以发现并不满足 $h(\frac{1}{2}\cdot 0+\frac{1}{2}\cdot 1)\leq \frac{1}{2}h(0)+\frac{1}{2}h(1)$ ，故 h 不为凸函数。

Problem6.

由于 $\lambda_1>0$ ，则有 $-\frac{\lambda_2}{\lambda_1},\cdots,-\frac{\lambda_n}{\lambda_1},\frac{1}{\lambda_1}=\frac{\lambda_1+\cdots,\lambda_n}{\lambda_1}>0$ ，同时 $(-\frac{\lambda_2}{\lambda_1})+\cdots+(-\frac{\lambda_n}{\lambda_1})+(\frac{\lambda_1+\cdots+\lambda_n}{\lambda_1})=1$ ，由于 f 为凸函数，由 Jensen 不等式可得 $f(x_1)=f(\frac{\lambda_1}{\lambda_1})\leq -\frac{\lambda_2}{\lambda_1}f(x_2)+\cdots-\frac{\lambda_n}{\lambda_1}f(x_n)+\frac{1}{\lambda_1}f(\lambda_1x_1+\cdots+\lambda_nx_n)$ ，整理后即为要证的式子

Problem7.

可以求得 f 的 Hessian 矩阵为:

$$\nabla^2 f(x)=\{\frac{\partial^2 f}{\partial x_i\partial x_j}\}=\begin{pmatrix} \frac{\lambda_1e^{-x_1}}{1-e^{-x_1}}\frac{\lambda_1e^{-x_1}-1}{1-e^{-x_1}}f(x) & \frac{\lambda_1e^{-x_1}}{1-e^{-x_1}}\frac{\lambda_2e^{-x_2}}{1-e^{-x_2}}f(x) & \cdots & \frac{\lambda_1e^{-x_1}}{1-e^{-x_1}}\frac{\lambda_ne^{-x_n}}{1-e^{-x_n}}f(x) \\ \frac{\lambda_2e^{-x_2}}{1-e^{-x_2}}\frac{\lambda_1e^{-x_1}}{1-e^{-x_1}}f(x) & \frac{\lambda_2e^{-x_2}}{1-e^{-x_2}}\frac{\lambda_2e^{-x_2}-1}{1-e^{-x_2}}f(x) & \cdots & \frac{\lambda_2e^{-x_2}}{1-e^{-x_2}}\frac{\lambda_ne^{-x_n}}{1-e^{-x_n}}f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\lambda_ne^{-x_n}}{1-e^{-x_n}}\frac{\lambda_1e^{-x_1}}{1-e^{-x_1}}f(x) & \frac{\lambda_2e^{-x_2}}{1-e^{-x_2}}\frac{\lambda_ne^{-x_n}}{1-e^{-x_n}}f(x) & \cdots & \frac{\lambda_ne^{-x_n}}{1-e^{-x_n}}\frac{\lambda_ne^{-x_n}-1}{1-e^{-x_n}}f(x) \end{pmatrix}$$

不妨令 $y=(\hat{y}_1,\cdots,\hat{y}_n)=((1-e^{-x_1})y_1,\cdots,(1-e^{-x_n})y_n)\in\mathbb{R}^n$ ，其中 $1-e^{-x_i}>0,\forall i=1,\cdots,n$ 。再设 $c_i=\lambda_ie^{-x_i}$ ，因此 $\hat{y}^T\nabla^2 f(x)\hat{y}=f(x)\cdot(\sum_{i=1}^n c_i(c_i-1)y_i^2+2\cdot\sum_{i\neq j}c_ic_jy_iy_j)=f(x)\cdot(-\sum_{i=1}^n c_iy_i^2+(\sum_{i=1}^n c_iy_i)^2)(*)$

由 Cauchy-Schwarz 不等式以及 $f(x)>0,\forall x\in\mathbb{R}^n$ 以及题设 $\sum_{i=1}^n c_i\leq 1$,可得 $(*)\leq f(x)\cdot(-\sum_{i=1}^n c_iy_i^2+(\sum_{i=1}^n c_i)(\sum_{i=1}^n c_iy_i^2))\leq f(x)\cdot(-\sum_{i=1}^n c_iy_i^2+\sum_{i=1}^n c_iy_i^2)=0$ 。

综上 f 是凹函数。