最优化导论　第三次作业

1. $f(\boldsymbol{x})$ 为凸函数的充要条件是对任意的 $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{R}^n$, 一元函数 $\varphi(\alpha) = f(\boldsymbol{x} + \alpha\boldsymbol{y})$ 是关于 α 的凸函数.

证明　必要性. 设 $\lambda_1 \geqslant 0, \lambda_2 \geqslant 0$, 且 $\lambda_1 + \lambda_2 = 1$, 由 $\varphi(\alpha)$ 的定义和 $f(\boldsymbol{x})$ 的凸性，有

$$\begin{aligned}
\varphi\left(\lambda_1\alpha_1 + \lambda_2\alpha_2\right) &= f\left(\boldsymbol{x} + \left(\lambda_1\alpha_1 + \lambda_2\alpha_2\right)\boldsymbol{y}\right) \\
&= f\left(\lambda_1\boldsymbol{x} + \lambda_2\boldsymbol{x} + \left(\lambda_1\alpha_1 + \lambda_2\alpha_2\right)\boldsymbol{y}\right) \\
&= f\left(\lambda_1\left(\boldsymbol{x} + \alpha_1\boldsymbol{y}\right) + \lambda_2\left(\boldsymbol{x} + \alpha_2\boldsymbol{y}\right)\right) \\
&\leqslant \lambda_1 f\left(\boldsymbol{x} + \alpha_1\boldsymbol{y}\right) + \lambda_2 f\left(\boldsymbol{x} + \alpha_2\boldsymbol{y}\right) \\
&= \lambda_1\varphi\left(\alpha_1\right) + \lambda_2\varphi\left(\alpha_2\right)
\end{aligned}$$

由定义知 $\varphi(\alpha)$ 是凸函数.

充分性. 任取 $x, y \in \mathrm{R}^n$, 记 $x = x + 0(y - x), y = x + (y - x), \varphi(\alpha) = f(x + \alpha(y - x))$ ，则

$$\begin{aligned}
f\left(\lambda_1\boldsymbol{x} + \lambda_2\boldsymbol{y}\right) &= f\left(\lambda_1(\boldsymbol{x} + 0(\boldsymbol{y} - \boldsymbol{x})) + \lambda_2(\boldsymbol{x} + (\boldsymbol{y} - \boldsymbol{x}))\right) \\
&= f\left(\boldsymbol{x} + \left(\lambda_1 0 + \lambda_2 1\right)(\boldsymbol{y} - \boldsymbol{x})\right) \\
&= \varphi\left(\lambda_1 0 + \lambda_2 1\right) \\
&\leqslant \lambda_1\varphi(0) + \lambda_2\varphi(1) \\
&= \lambda_1 f(\boldsymbol{x}) + \lambda_2 f(\boldsymbol{y})
\end{aligned}$$

故知 $f(\boldsymbol{x})$ 是凸函数.

2. 利用凸函数二阶条件证明如下结论：ln-sum-exp 函数: $f(x) = \ln\sum_{k=1}^n \exp x_k$ 是凸函数;

证明　求海瑟矩阵, 为了方便记 $S = \sum_{k=1}^n \exp x_k$ 。则

$$\nabla^2 f = \frac{1}{S^2} \cdot \left[\begin{array}{cccc}
e^{x_1}\left(S - e^{x_1}\right) & -e^{x_1}e^{x_2} & \cdots & -e^{x_1}e^{x_n} \\
-e^{x_2}e^{x_1} & e^{x_2}\left(S - e^{x_2}\right) & \cdots & -e^{x_1}e^{x_n} \\
\vdots & \vdots & \ddots & \vdots \\
-e^{x_n}e^{x_1} & -e^{x_n}e^{x_2} & \cdots & e^{x_n}\left(S - e^{x_n}\right)
\end{array}\right]$$

现在只需证明 $\nabla^2 f$ 半正定. 对 $\forall y = \left(y_1, y_2, \cdots, y_n\right)^{\mathrm{T}} \in \mathbb{R}^n (y \neq 0)$ ，成立

$$y^{\mathrm{T}}\nabla^2 f y = \frac{1}{S^2}\left[\sum_{k=1}^n y_k^2 e^{x_k}\sum_{k=1}^n e^{x_k} - \left(\sum_{k=1}^n y_k e^{x_k}\right)^2\right]$$

利用柯西不等式可知 $y^{\mathrm{T}}\nabla^2 f y \geqslant 0$. 因此 $\nabla^2 f \succeq 0$.

3. 设 D 是 R^n 中的一个非空凸集，f 是定义在 D 上的凸函数，α 是一个实数, 则水平集 $D_\alpha=\{\boldsymbol{x} \mid \boldsymbol{x} \in D, f(\boldsymbol{x}) \leqslant \alpha\}$ 是凸集.

证明 证明对于任意的 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in D_\alpha$, 根据 D_α 定义, 有

$$f\left(\boldsymbol{x}^{(1)}\right) \leqslant \alpha, f\left(\boldsymbol{x}^{(2)}\right) \leqslant \alpha.$$

由于 D 为凸集, 因此对每个数 $\lambda \in[0,1]$, 必有

$$\lambda x^{(1)}+(1-\lambda) x^{(2)} \in D.$$

又由于 $f(x)$ 是 D 上的凸函数, 则有

$$\begin{aligned} f\left(\lambda x^{(1)}+(1-\lambda) x^{(2)}\right) & \leqslant \lambda f\left(x^{(1)}\right)+(1-\lambda) f\left(x^{(2)}\right) \\ & \leqslant \lambda \alpha+(1-\lambda) \alpha=\alpha \end{aligned}$$

因此 $\lambda x^{(1)}+(1-\lambda) x^{(2)} \in D_\alpha$, 故 D_α 为凸集.

4. 求下列函数的共轭函数：最大值函数: $f(x)=\max x_i$.

解 分情况讨论. 若 $\|y\|_1 \leqslant 1$ 且 $y \geqslant 0$, 则 y 与 x 内积时不会改变 x 分量的符号，且一定成立

$$y^{\mathrm{T}} x \leqslant \max_i x_i,$$

故此时 $f^*(y)=0$. 若不然, 则或有 $\|y\|_1>1$. 不妨设 $j=\arg\max x_i$ 且 $y_j=1+\delta$ ($\delta>0$ 且其他坐标取 0), 那么

$$\begin{aligned} f^*(y) & =\sup_{x \in \mathbb{R}^n}\left\{y^{\mathrm{T}} x-x_j\right\} \\ & =\sup_{x, \mathbb{R}}\left\{\delta x_j\right\} \\ & \rightarrow \infty . \end{aligned}$$

又或存在 i 使得 $y_i<0$, 类似可证 $f^*(y)$ 不存在. 综上, $f^*(y)=0$, 定义域是 $\{y \in \mathbb{R}^n \mid y \geqslant 0,\|y\|_1 \leqslant 1\}$.

5. (1) 证明有限个凸集的交集仍然是凸集.

(2) 设 $D_1=\{x \mid x_1+x_2 \leqslant 1, x_1 \geqslant 0\}, D_2=\{x \mid x_1-x_2 \geqslant 0, x_1 \leqslant 0\}$. 令 $D=D_1 \cup D_2$. 证明 D_1, D_2 均是凸集, 但 D 却不是凸的, 由此得出凸集的并集未必是凸集.

证明 (1) 设 $C=\bigcap_{i=1}^n C_i$, 其中 $C_i(i=1,2,\cdots,n)$ 均是凸集. 取 $x_1, x_2 \in C$, 则对于 $\forall i$ 都有 $x_1, x_2 \in C_i$. 取 $\mu_1, \mu_2 \in[0,1]$ 且 $\mu_1+\mu_2=1$, 则有 $\mu_1 x_1+\mu_2 x_2 \in C_i$. 因此有 $\mu_1 x_1+\mu_2 x_2 \in \bigcap_{i=1}^n C_i=C$, 即 C 为凸集. 说明有限个凸集的交集仍然是凸集.

(2) 证明 D_1 是凸集:

$$\forall x, y \in D_1, x=\left(x_1, x_2\right), y=\left(y_1, y_2\right), \lambda \in[0,1], \lambda x+(1-\lambda) y=\left(\lambda x_1+(1-\lambda) y_1, \lambda x_2+(1-\lambda) y_2\right),$$

$$\begin{aligned} \lambda x_1+(1-\lambda) y_1+\lambda x_2+(1-\lambda) y_2 & =\lambda\left(x_1+x_2\right)+(1-\lambda)\left(y_1+y_2\right) \\ & \leqslant \lambda+1-\lambda=1 \end{aligned}$$

$\lambda x_1+(1-\lambda) y_1 \geqslant 0 \Rightarrow \lambda x+(1-\lambda) y \in D_1 \Rightarrow D_1$ 为凸集。

证明 D_2 是凸集:

$$\forall x,y\in D_2, x=(x_1,x_2), y=(y_1,y_2), \lambda\in[0,1], \lambda x+(1-\lambda)y=(\lambda x_1+(1-\lambda)y_1, \lambda x_2+(1-\lambda)y_2),$$

$$\begin{aligned}\lambda x_1+(1-\lambda)y_1-\lambda x_2-(1-\lambda)y_2&=\lambda(x_1-x_2)+(1-\lambda)(y_1-y_2)\\&\geqslant 0\end{aligned}$$

$\lambda x_1+(1-\lambda)y_1\leqslant 0\Rightarrow \lambda x+(1-\lambda)y\in D_2\Rightarrow D_2$ 为凸集。

证明 D 不是凸集:

取 $x=(1,0)\in D, y=(0,-1)\in D, \lambda=\frac{1}{2}$, 则 $\lambda x+(1-\lambda)y=\left(\frac{1}{2},-\frac{1}{2}\right)\notin D\Rightarrow D$ 不是凸集.

6. 设 $c_i(x)(i=1,2,\cdots,m)$ 为 R^n 上的凹函数, $f(x)$ 是 R^n 上的凸函数, 证明函数

$$P(x)=f(x)-\mu\sum_{i=1}^{m}\log c_i(x)$$

在集合 $D=\{x\mid c_i(x)>0\}$ 上是凸函数. ($\mu>0$)

证明 易证 $\log c_i(x)$ 是凹函数, 所以 $\forall x,y\in D$, 则

$$\begin{aligned}P(\lambda x+(1-\lambda)y)&=f(\lambda x+(1-\lambda)y)-\mu\sum_{i=1}^{m}\log c_i(\lambda x+(1-\lambda)y)\\&\leqslant \lambda f(x)+(1-\lambda)f(y)-\mu\sum_{i=1}^{m}\log(\lambda c_i(x)+(1-\lambda)c_i(y))\\&\leqslant \lambda f(x)+(1-\lambda)f(y)-\lambda\mu\sum_{i=1}^{m}\log c_i(x)-(1-\lambda)\mu\sum_{i=1}^{m}\log c_i(y)\\&=\lambda P(x)+(1-\lambda)P(y)\end{aligned}$$

所以, 该函数在集合 $D=\{x\mid c_i(x)>0\}$ 上是凸函数.

课本习题: 《Convex Optimization》3.5, 3.6, 3.7, 3.8, 3.10, 3.11, 3.12, 3.16