作业一

最优化方法

截止时间：10 月 11 日 23:59（周三晚）

请在规定时间前提交到大夏学堂，超时得分将有折扣。
计算、证明题提交 pdf 电子版，编程题提交 Python 代码、结果及必要的解释。

1 阅读任务

略。

2 凸集

a. 证明若函数 f 二次可微，则 f 为凸函数的充要条件为：

- $\mathrm{dom}\, f$ 为凸集
- $\nabla^2 f(x) \succeq 0$, for all $x \in \mathrm{dom}\, f$

($\Rightarrow$): 若 f 为二次可微凸函数，则 $\mathrm{dom}\, f$ 为凸集，由 $f(x) \geq f(y) + \nabla f(y)^\top (x-y), \quad f(y) \geq f(x) + \nabla f(x)^\top (y-x)$ 相加，可得以下成立（事实上，反向也成立）：

$$(\nabla f(x) - \nabla f(y))^\top (x-y) \geq 0, \quad \forall x, y$$

令 $x_\tau = x + \tau s, \tau > 0$，因此

$$\begin{aligned} 0 &\leq \frac{1}{\tau}(\nabla f(x_\tau) - \nabla f(x))^\top (x_\tau - x) = \frac{1}{\tau}(\nabla f(x_\tau) - \nabla f(x))^\top s \\ &= \frac{1}{\tau}\int_0^\tau s^\top \nabla^2 f(x + \lambda s) s d\lambda \end{aligned}$$

令 $\tau \to 0$ 可得 $\nabla^2 f(x) \geq 0$.

($\Leftarrow$): 反之，二次可微函数有泰勒展开：

$$f(y) = f(x) + \nabla f(x)^\top (y-x) + \frac{1}{2}(y-x)^\top \nabla^2 f(\xi)(y-x) \geq f(x) + \nabla f(x)^\top (y-x)$$

即为凸函数的一阶等价条件.

b. 闭集与凸集

b.1. 证明多面体 $\{x : Ax \leq b\}, A \in \mathbb{R}^{mn}, b \in \mathbb{R}^m$ 为闭的凸集

假设多面体集合中的一列点 x_n，且有极限点 x^*，则有 $\lim_{n\to\infty} Ax_n = Ax^* \leq b$.
对于凸性，若 $x, y \in \{Ax \leq b\}$，对 $\theta \in [0,1]$，有 $A(\theta x + (1-\theta)y) = \theta Ax + (1-\theta)Ay \leq \theta b + (1-\theta)b = b$.

b.2. 举例：$\mathbb{R}^2$ 上闭集的凸包不一定是闭的

e.g., $\{(x,y): y \geq 1/(1+x^2)\}$ *or* $\{(x,y): x \in [0,1], y = 0 \text{ or } x = 1, y \geq 0\}$

b.3. 若 $A \in \mathbb{R}^{m\times n}, S \in \mathbb{R}^m$ 为凸集，证明集合 S 在 A 下的原像 $\{x \in \mathbb{R}^n : Ax \in S\}$ 是凸集

Let $B = \{x \in \mathbb{R}^n : Ax \in S\}$*, then for any* $x, y \in B$, $Ax, Ay \in S$*, thus* $A(\alpha x + (1-\alpha)y) \in S, \forall \alpha \in (0,1)$ *due to convexity of* S*, so* $\alpha x + (1-\alpha)y \in B$.

b.4. 若 $A \in \mathbb{R}^{m\times n}, S \in \mathbb{R}^n$ 为凸集，证明集合 S 在 A 下的像 $\{Ax : x \in S\}$ 是凸集

Let $B = \{Ax : x \in S\}$*, for* $x, y \in B$*, there exist* $x', y' \in S$ *such that* $Ax' = x, Ay' = y$*, and due to* $\alpha x' + (1-\alpha)y' \in S$, $\alpha x + (1-\alpha)y = A(\alpha x' + (1-\alpha)y') \in S, \forall \alpha \in (0,1)$.

b.5. 举例：存在 $A \in \mathbb{R}^{m\times n}$ 及闭凸集 $S \subseteq \mathbb{R}^n$，使得 $A(S)$ 不是闭集

For $\mathbb{R}^2$*, Let* $S = \{(x,y) : y \geq 1/x, x > 0\}$, $Ax = x_1$.

c. 多面体

c.1. 证明：若 $P \subseteq \mathbb{R}^n$ 为多面体，$A \in \mathbb{R}^{m\times n}$，则 $A(P)$ 为多面体，提示：可使用以下事实：

$$P \subseteq \mathbb{R}^{m+n} \text{ 为多面体} \Rightarrow \{x \in \mathbb{R}^n : (x,y) \in P \text{ for some } y \in \mathbb{R}^m\} \text{ 是多面体}$$

Let $B = \{(x,y) : x \in P, y = Ax\} = \{(x,y) : Mx \succeq 0, Lx = 0, Ax - y = 0\}$*, then* B *is a polytope. So* $\{y \in \mathbb{R}^m : (x,y) \in B \text{ for some } x \in \mathbb{R}^n\}$ *is a polytope.*

c.2. 证明若 $Q \subseteq \mathbb{R}^m$ 为多面体，$A \in \mathbb{R}^{m\times n}$，则 $A^{-1}(Q)$ 为多面体

提示（hint）：Let $B' = \{(x,y) : x \in P, y = Ax\} = \{(x,y) : Mx \succeq 0, Lx = 0, Ay - x = 0\}$.

3 凸函数

a. 证明熵函数：

$$f(x) = -\sum_{i=1}^n x_i \log(x_i), \quad \operatorname{dom} f = \{x \in \mathbb{R}^n_{++} : \sum_{i=1}^n x_i = 1\}$$

是严格凹的

提示（hint）：check the positive definiteness of Hessian matrix.

b. 若 f 为二次可微函数且 $\operatorname{dom} f$ 为凸集，证明 f 为凸函数的充要条件为：

$$(\nabla f(x) - \nabla f(y))^\top (x-y) \geq 0, \quad \forall x, y$$

这被称为梯度 ∇f 的单调性

只证明充分性。存在 $\lambda \in (0,1)$,

$$\begin{aligned} f(y) &= f(x) + \nabla f(x + \lambda(y-x))^\top (y-x) \\ &= f(x) + \nabla f(x)^\top (y-x) + (\nabla f(x+\lambda(y-x)) - \nabla f(x))^\top (y-x) \\ &\geq f(x) + \nabla f(x)^\top (y-x) \end{aligned}$$

c. 举例：严格凸函数并不一定能达到其极小值

e.g., $f(x) = 1/x, x \in \mathbb{R}_{++}$.

d. 函数 $f : \mathbb{R}^n \to \mathbb{R}$ 被称为强制的（coercive），如果当 $\|x\|_2 \to \infty$ 时，有 $f \to \infty$。强制函数的一个关键事实为其可以达到极小值。证明对于一个二次可微的强凸函数是强制的（coercive），并因此可达到极小值。

提示（hint）：$f(x) - \frac{m}{2}\|x\|^2 = f(x_0) + \nabla f(x_0)^\top (x - x_0) + \frac{1}{2}(x - x_0)^\top [\nabla^2 f(\tilde{x}) - \frac{m}{2} I](x - x_0)$，*根据强凸性，可知关于* $\nabla^2 f(\tilde{x}) - \frac{m}{2}$ *矩阵半正定，故* $\nabla^2 f(\tilde{x})$ *正定的，则* $f(x)$ *可看作关于* x *的二次函数，且二次矩阵正定，满足强制性条件* $\|x\|_2 \to \infty, f(x) \to \infty$。

e. 证明在有界多面体上的凸函数的最大值一定在其中一个顶点上。提示：已知一个有界的多面体可以被表示为其顶点的凸组合。

Denote $a_i, i = 1, \ldots, n$ *as vertices of the bounded polyhedron* P, *and*

$$P = \left\{ x \in \mathbb{R}^m : x = \sum_{i=1}^{n} \lambda_i a_i, \lambda_i \geq 0, \sum_{i=1}^{n} \lambda_i = 1 \right\}$$

Let f *be a convex function on* P, *Assume that* x_0 *is the point that* f *reaches its maximum value on* P. *Thus there exists* $\lambda_i \geq 0, i = 1, \ldots, n$, *and* $\sum_i \lambda_i = 1$ *such that* $\sum_{i=1}^{n} \lambda_i a_i = x_0$. *And*

$$f(x_0) = f(\sum_i \lambda_i a_i) \leq \sum_i \lambda_i f(a_i) \leq \max_i f(a_i) = f(x_0)$$

This implies that $\lambda_i > 0$ *iff (if and only if)* $f(a_i) = f(x_0)$.

4 实践：使用 CVXPY 解优化问题

见文件：homework1-cvxpy.ipynb，或在线 kaggle 文档：
`https://www.kaggle.com/code/mrsta1123/homework1-cvxpy`