CS257 Linear and Convex Optimization
Homework 3

Due: October 12, 2020

September 29, 2020

1. Prove that $\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$ are affinely independent if and only if

$$\sum_{i=0}^{m} c_i \boldsymbol{x}_i = \mathbf{0} \text{ and } \sum_{i=0}^{m} c_i = 0 \implies c_i = 0 \text{ for } i = 0, 1, \dots, m. \qquad (\star)$$

2. Given two probability distributions $\boldsymbol{x}, \boldsymbol{y} \in \Delta_{n-1}$, where Δ_{n-1} is the probability simplex, the **Kullback-Leibler (KL) divergence** between them is defined by

$$KL(\boldsymbol{x} \| \boldsymbol{y}) = \sum_{i=1}^{n} x_i \log \frac{x_i}{y_i}$$

Use the concavity of log to show that $KL(\boldsymbol{x} \| \boldsymbol{y}) \geq 0$. You can assume $\boldsymbol{x} > \mathbf{0}, \boldsymbol{y} > \mathbf{0}$.

3. Prove that the set of global minima of a convex function over a convex set is convex, i.e. if f is convex and $S \subset \operatorname{dom} f$ is convex, then the set M defined below is also convex

$$M = \{\boldsymbol{x}^* \in S : f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \forall \boldsymbol{x} \in S\}$$

4. Let f be convex. If $f(\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}) = \theta f(\boldsymbol{x}) + \bar{\theta} f(\boldsymbol{y})$ for some $\boldsymbol{x}, \boldsymbol{y}$ and $\theta = \theta_0 \in (0, 1)$, then it holds for the same $\boldsymbol{x}, \boldsymbol{y}$ and any $\theta \in [0, 1]$.

Hint: Assume $f(\theta_1 \boldsymbol{x} + \bar{\theta}_1 \boldsymbol{y}) < \theta_1 f(\boldsymbol{x}) + \bar{\theta}_1 f(\boldsymbol{y})$ for some θ_1. Without loss of generality, you may assume $\theta_1 \in (0, \theta_0)$; the case $\theta_1 \in (\theta_0, 1)$ is similar. Express $\theta_0 \boldsymbol{x} + \bar{\theta}_0 \boldsymbol{y}$ as a convex combination of $\theta_1 \boldsymbol{x} + \bar{\theta}_1 \boldsymbol{y}$ and $\boldsymbol{x}$. Then deduce a contradiction.

5. Suppose $f : S \subset \mathbb{R} \to \mathbb{R}$ is convex. Let $a, b \in S$ and $a < b$.

(a). Show

$$f(x) \leq \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b), \quad \forall x \in [a, b]$$

(b). Show

$$\frac{f(x) - f(a)}{x - a} \leq \frac{f(b) - f(a)}{b - a} \leq \frac{f(b) - f(x)}{b - x}, \quad \forall x \in (a, b)$$

What's the geometric interpretation of these inequalities? Draw a sketch to illustrate them.

(c). Suppose f is differentiable. Use (b) to show

$$f'(a) \leq \frac{f(b) - f(a)}{b - a} \leq f'(b)$$

What can you say about the monotonicity of the derivative f'?

(d). Suppose f is twice differentiable. Note $a < b$ are arbitrary points in S. Use (c) to show that $f''(x) \geq 0$.

6. Let $f : (a, b) \to \mathbb{R}$ be convex, where $-\infty \leq a < b \leq +\infty$. Let X be random variable taking values in (a, b). Suppose the expectations $\mathbb{E}X$ and $\mathbb{E}f(X)$ exist. Prove Jensen's inequality $f(\mathbb{E}X) \leq \mathbb{E}f(X)$ by completing the following steps.

(a). Let $\mu = \mathbb{E}X$. Show that there exists $\beta \in \mathbb{R}$ such that

$$f(x) \geq f(\mu) + \beta(x - \mu), \quad \forall x \in (a, b) \tag{$\star$}$$

Hint: You can take

$$\beta = \sup_{a<s<\mu} \frac{f(\mu) - f(s)}{\mu - s}.$$

Obviously $\beta > -\infty$. Use part (b) of Problem 5 to show that $\beta < +\infty$ and satisfies ($\star$) (consider $a < x < \mu$ and $\mu < x < b$ separately).

(b). Show that

$$f(X) \geq f(\mu) + \beta(X - \mu).$$

and conclude $\mathbb{E}f(X) \geq f(\mathbb{E}X)$ by taking expectation.

Remark. If f is differentiable, we can take $\beta = f'(\mu)$ by the first-order condition. Part (a) shows that ($\star$) holds without assuming differentiability. The number β used in the proof generalizes the concept of gradient (derivative) $f'(\mu)$. Any β satisfying ($\star$) is called a **subgradient** of f at μ. For example, any $\beta \in [-1, 1]$ is a subgradient of $f(x) = |x|$ at 0.

7. In this problem, we show that the so-called **midpoint convexity** and continuity implies convexity. Assume $f : \mathbb{R} \to \mathbb{R}$ is **midpoint convex**, i.e.

$$f\left(\frac{x_1 + x_2}{2}\right) \leq \frac{1}{2}\left[f(x_1) + f(x_2)\right], \quad \forall x_1, x_2 \in \mathbb{R}.$$

(a). Show that for $k \in \mathbb{N}$,

$$f\left(\frac{x_1 + x_2 + \cdots + x_{2^k}}{2^k}\right) \leq \frac{1}{2^k}\left[f(x_1) + f(x_2) + \cdots + f(x_{2^k})\right], \quad \forall x_1, x_2, \ldots, x_{2^k} \in \mathbb{R}.$$

Hint: Use induction on k.

(b). Show that for $n \in \mathbb{N}$,

$$f\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) \leq \frac{1}{n}\left[f(x_1) + f(x_2) + \cdots + f(x_n)\right], \quad \forall x_1, x_2, \ldots, x_n \in \mathbb{R}.$$

Hint: Assume $2^{k-1} < n \leq 2^k$. Let $x_i = \bar{x} = \frac{x_1 + \cdots + x_n}{n}$ for $i = n + 1, \ldots 2^k$ in part (a).

(c). Use (b) to show that for $p, q \in \mathbb{N}$,

$$f\left(\frac{p}{p+q}x_1 + \frac{q}{p+q}x_2\right) \le \frac{p}{p+q}f(x_1) + \frac{q}{p+q}f(x_2), \quad \forall x_1, x_2 \in \mathbb{R}.$$

Conclude that for rational $\theta \in \mathbb{Q} \cap (0,1)$,

$$f(\theta x_1 + \bar{\theta} x_2) \le \theta f(x_1) + \bar{\theta} f(x_2), \quad \forall x_1, x_2 \in \mathbb{R}.$$

(d). Assume f is also continuous. Show that for $\theta \in [0,1]$,

$$f(\theta x_1 + \bar{\theta} x_2) \le \theta f(x_1) + \bar{\theta} f(x_2), \quad \forall x_1, x_2 \in \mathbb{R}.$$

Hint: Let $\{\theta_n\}$ be a sequence of rational numbers in $\mathbb{Q} \cap (0,1)$ such that $\theta_n \to \theta$ as $n \to \infty$.