Introduction to Machine Learning
Fall 2022
University of Science and Technology of China

Lecturer: Jie Wang
Name: Yunqin Zhu

Homework 4
ID: PB20061372

Notice, to get the full credits, please present your solutions step by step.

Exercise 1: Convex Functions

1. (Optional) For each of the following functions, determine whether it is convex.
   (a) $f(x) = x^2 \log x$ on $\mathbb{R}_{++}$, where $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$.
   (b) $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2$.
   (c) $f(x_1, x_2) = \frac{x_1}{x_2}$ on $\mathbb{R}^2_{++}$, where $\mathbb{R}^2_{++} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1, x_2 > 0\}$.
   (d) $f(x_1, x_2) = \frac{x_1^2}{x_2}$ on $\mathbb{R} \times \mathbb{R}_{++}$.
   (e) $f(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha}$ on $\mathbb{R}^2_{++}$, where $0 \leq \alpha \leq 1$.

   Solution:
   (a) $f''(x) = 2\log x + 3 < 0$ for $x \in [0, e^{-\frac{3}{2}})$, so f is not convex.
   (b) $\nabla^2 f(x) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, which is indefinite, so f is not convex.
   (c) $\nabla^2 f(x) = \begin{pmatrix} 0 & -\frac{1}{x_2^2} \\ -\frac{1}{x_2^2} & \frac{2x_1}{x_2^3} \end{pmatrix}$, which is indefinite on $\mathbb{R}^2_{++}$, so f is not convex.
   (d) $\nabla^2 f(x) = \begin{pmatrix} \frac{2}{x_2} & -\frac{2x_1}{x_2^2} \\ -\frac{2x_1}{x_2^2} & \frac{2x_1^2}{x_2^3} \end{pmatrix} \succeq 0$ on $\mathbb{R} \times \mathbb{R}_{++}$, so f is convex.
   (e) $\nabla^2 f(x) = \begin{pmatrix} \alpha(\alpha-1)x_1^{\alpha-2}x_2^{1-\alpha} & \alpha(1-\alpha)x_1^{\alpha-1}x_2^{-\alpha} \\ \alpha(1-\alpha)x_1^{\alpha-1}x_2^{-\alpha} & \alpha(\alpha-1)x_1^{\alpha}x_2^{-\alpha-1} \end{pmatrix} \succeq 0$ iff. $\alpha \leq 0$ or $\alpha \geq 1$. ■

2. Please show that the following functions are convex.
   (a) $f(\mathbf{x}) = \log \sum_{i=1}^{n} e^{x_i}$ on $\mathbf{dom}\ f = \mathbb{R}^n$, where x_i denotes the i^{th} component of $\mathbf{x}$.
   (b) $f(\mathbf{x}) = \sum_{i=1}^{k} x_{[i]}$ on $\mathbf{dom}\ f = \mathbb{R}^n$, where $1 \leq k \leq n$ and $x_{[i]}$ denotes the i^{th} largest component of $\mathbf{x}$.
   (c) The extended-value extension of the indicator function of a convex set $C \subseteq \mathbb{R}^n$, i.e.,

$$\tilde{I}_C(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in C, \\ \infty, & \mathbf{x} \notin C. \end{cases}$$

   (d) The negative entropy, i.e.,

$$f(\mathbf{p}) = \sum_{i=1}^{n} p_i \log p_i$$

   on $\mathbf{dom}\ f = \{\mathbf{p} \in \mathbb{R}^n : 0 < p_i \leq 1, \sum_{i=1}^{n} p_i = 1\}$, where p_i denotes the i^{th} component of $\mathbf{p}$.

(e) The spectral norm, i.e.,

$$f(\mathbf{X}) = \|\mathbf{X}\|_2 = \sigma_{\max}(\mathbf{X})$$

on $\mathbf{dom}\ f = \mathbb{R}^{m\times n}$, where $\sigma_{\max}$ denotes the largest singular value of $\mathbf{X}$.

(f) $f(\mathbf{X}) = \operatorname{tr}(\mathbf{X}^{-1})$ on $\mathbf{dom}\ f = \mathbb{S}^n_{++}$, where $\mathbb{S}^n_{++}$ is the space of all $n \times n$ real positive definite matrices.

Solution:

(a) The Hessian of f is

$$\nabla^2 f(\mathbf{x}) = \frac{1}{(\sum_i e^{x_i})^2} \begin{pmatrix} e^{x_1} \sum\limits_{i\neq 1} e^{x_i} & -e^{x_1}e^{x_2} & \cdots & -e^{x_1}e^{x_n} \\ -e^{x_1}e^{x_2} & e^{x_2} \sum\limits_{i\neq 2} e^{x_i} & \cdots & -e^{x_2}e^{x_n} \\ \vdots & \vdots & \ddots & \vdots \\ -e^{x_1}e^{x_n} & -e^{x_2}e^{x_n} & \cdots & e^{x_n} \sum\limits_{i\neq n} e^{x_i} \end{pmatrix}.$$

$\nabla^2 f(\mathbf{x})$ is diagonally dominant with positive diagonal entries, so $\nabla^2 f(\mathbf{x}) \succeq 0$, and hence f is convex.

(b) We have

$$f(\mathbf{x}) = \max\left\{ \sum_{i\in I} x_i : I \subseteq \{1,\dots,n\}, |I| = k \right\}.$$

Given any I, the linear function $\sum_{i\in I} x_i$ is convex on $\mathbb{R}^n$. It follows that f is also convex.

(c) The epigraph of $\tilde{I}_C$ is

$$\mathbf{epi}\ \tilde{I}_C = C \times \mathbb{R}_+,$$

which is a convex set, so $\tilde{I}_C$ is convex.

(d) The Hessian of f is

$$\nabla^2 f(\mathbf{p}) = \begin{pmatrix} \frac{1}{p_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{p_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{p_n} \end{pmatrix} \succ 0,$$

and thus f is convex.

(e) We have

$$f(\mathbf{X}) = \sup\left\{ \langle \mathbf{u}, \mathbf{X}\mathbf{v} \rangle : \|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1, \mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n \right\}.$$

Given any $\mathbf{u}$ and $\mathbf{v}$, the linear function $\langle \mathbf{u}, \mathbf{X}\mathbf{v} \rangle$ is convex in $\mathbf{X}$, so f is convex.

(f) Let $\mathbf{X} \in \mathbb{S}^n_{++}$ and $\mathbf{Y} \in \mathbb{S}^n$. Then

$$(\mathbf{X} + t\mathbf{Y})^{-1} = (\mathbf{I} + t\mathbf{X}^{-1}\mathbf{Y})^{-1}\mathbf{X}^{-1} = \left(\sum_{k=0}^{\infty} (-1)^k t^k \left(\mathbf{X}^{-1}\mathbf{Y}\right)^k \right) \mathbf{X}^{-1}.$$

It follows that

$$f(\mathbf{X}+t\mathbf{Y}) = \operatorname{tr}\left((\mathbf{X}+t\mathbf{Y})^{-1}\right) = \sum_{k=0}^{\infty}(-1)^k t^k \operatorname{tr}\left(\left(\mathbf{X}^{-1}\mathbf{Y}\right)^k \mathbf{X}^{-1}\right).$$

Therefore, we have

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\mathbf{X}+t\mathbf{Y})\Big|_{t=0} = 2\operatorname{tr}\left(\left(\mathbf{X}^{-1}\mathbf{Y}\right)^2 \mathbf{X}^{-1}\right) \geq 0,$$

as $\mathbf{X}^{-1} \succ 0$ and $\left(\mathbf{X}^{-1}\mathbf{Y}\right)^2 \succeq 0$. By changing variables, we conclude that $\frac{\mathrm{d}^2}{\mathrm{d}t^2} f(\mathbf{X}+t\mathbf{Y}) \geq 0$ for any $t \in \{t : \mathbf{X}+t\mathbf{Y} \in \mathbf{dom}\, f\}$, which implies the convexity of f. ■

3. Please show that a continuously differentiable function f is strongly convex with parameter $\mu > 0$ if and only if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \frac{\mu}{2}\|\mathbf{y}-\mathbf{x}\|_2^2, \quad \forall\, \mathbf{x},\mathbf{y} \in \mathbb{R}^n.$$

Solution:
By definition, f is strongly convex with parameter $\mu > 0$ if and only if $g(\mathbf{x}) = f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|_2^2$ is convex. Since g is also continuously differentiable, we know g is convex if and only if $g(\mathbf{y}) \geq g(\mathbf{x}) + \langle \nabla g(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle$ for all $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$. Combining these two facts, f is strongly convex with parameter $\mu > 0$ if and only if

$$\begin{aligned}
f(\mathbf{y}) &\geq \frac{\mu}{2}\|\mathbf{y}\|_2^2 + f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|_2^2 + \langle \nabla f(\mathbf{x}) - \mu\mathbf{x}, \mathbf{y}-\mathbf{x}\rangle \\
&= f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \frac{\mu}{2}\langle \mathbf{y}+\mathbf{x}, \mathbf{y}-\mathbf{x}\rangle - \mu\langle \mathbf{x}, \mathbf{y}-\mathbf{x}\rangle \\
&= f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \frac{\mu}{2}\|\mathbf{y}-\mathbf{x}\|_2^2, \quad \forall\, \mathbf{x},\mathbf{y} \in \mathbb{R}^n
\end{aligned}$$

■

4. Suppose that f is twice continuously differentiable and strongly convex with parameter $\mu > 0$. Please show that $\mu \leq \lambda_{\min}(\nabla^2 f(\mathbf{x}))$ for any $\mathbf{x} \in \mathbb{R}^n$, where $\lambda_{\min}(\nabla^2 f(\mathbf{x}))$ is the smallest eigenvalue of $\nabla^2 f(\mathbf{x})$.

Solution:
By the definition of strong convexity, $f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|_2^2$ is convex and twice continuously differentiable, so we have $\nabla^2\left(f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|_2^2\right) = \nabla^2 f(\mathbf{x}) - \mu\mathbf{I} \succeq 0$, and hence $\lambda_{\min}(\nabla^2 f(\mathbf{x}) - \mu\mathbf{I}) \geq 0$, i.e. $\mu \leq \lambda_{\min}(\nabla^2 f(\mathbf{x}))$. ■

5. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, and the gradient of f is Lipschitz continuous, i.e.,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq L\|\mathbf{x}-\mathbf{y}\|_2, \quad \forall\, \mathbf{x},\mathbf{y} \in \mathbb{R}^n,$$

where $L > 0$ is the Lipschitz constant. Please show that $\lambda_{\max}(\nabla^2 f(\mathbf{x})) \leq L$ for any $\mathbf{x} \in \mathbb{R}^n$, where $\lambda_{\max}(\nabla^2 f(\mathbf{x}))$ is the largest eigenvalue of $\nabla^2 f(\mathbf{x})$.

Solution:
Suppose $\mathbf{x} \neq \mathbf{y}$ and let $\mathbf{y} \to \mathbf{x}$. Then

$$\limsup_{\mathbf{y}\to\mathbf{x}} \frac{\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2}{\|\mathbf{x}-\mathbf{y}\|_2} \leq L.$$

On the other hand, by the definition of differentiability,

$$\lim_{\mathbf{y}\to\mathbf{x}} \frac{\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}) - \nabla^2 f(\mathbf{x})^\top(\mathbf{x}-\mathbf{y})\|_2}{\|\mathbf{x}-\mathbf{y}\|_2} = 0.$$

Adding the two equations above and using the triangle inequality, we yield

$$\|\nabla^2 f(\mathbf{x})\|_2 = \limsup_{\mathbf{y}\to\mathbf{x}} \frac{\|\nabla^2 f(\mathbf{x})^\top(\mathbf{x}-\mathbf{y})\|_2}{\|\mathbf{x}-\mathbf{y}\|_2} \leq L + 0 = L,$$

i.e. $\lambda_{\max}(\nabla^2 f(\mathbf{x})) \leq L$ for any $\mathbf{x} \in \mathbb{R}^n$. ■

6. Consider the problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}), \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and convex, and $\mathbf{dom}\ f$ is closed.

(a) Please show that the α-sublevel set of f, i.e., $C_\alpha = \{\mathbf{x} \in \mathbf{dom}\ f : f(\mathbf{x}) \leq \alpha\}$ is closed.

(b) Please give an example to show that Problem (1) may be unsolvable even if f is strictly convex.

(c) Suppose that f can attain its minimum. Please show that the optimal set $\mathcal{C} = \{\mathbf{y} : f(\mathbf{y}) = \min_{\mathbf{x}} f(\mathbf{x})\}$ is closed and convex. Does this property still hold if $\mathbf{dom}\ f$ is not closed?

(d) Suppose that f is strongly convex with parameter $\mu > 0$. Please show that Problem (1) admits a unique solution.

Solution:

(a) If C_α has no limit point, then it is trivially closed. Otherwise, let $\{\mathbf{x}_k\} \subset C_\alpha$ be an arbitrary sequence that converges to some $\mathbf{x} \in \mathbf{dom}\ f$. Since $f(\mathbf{x}_k) \leq \alpha$ and f is continuous, we have $f(\mathbf{x}) = \lim_{k\to\infty} f(\mathbf{x}_k) \leq \alpha$. Thus $\mathbf{x} \in C_\alpha$, which implies that C_α is closed.

(b) $f(x) = e^{-x}$ is strictly convex on $\mathbb{R}$ but its infimum cannot be attained.

(c) $\mathcal{C} = \{\mathbf{y} : f(\mathbf{y}) \leq \min_{\mathbf{x}} f(\mathbf{x})\}$ is a nonempty sublevel set of f. According to Question 6(a), it is closed. For any $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{C}$ and $0 \leq \theta \leq 1$, we have $f(\theta\mathbf{y}_1 + (1-\theta)\mathbf{y}_2) \leq \theta f(\mathbf{y}_1) + (1-\theta) f(\mathbf{y}_2) \leq \min_{\mathbf{x}} f(\mathbf{x})$, so $\theta\mathbf{y}_1 + (1-\theta)\mathbf{y}_2 \in \mathcal{C}$, and hence $\mathcal{C}$ is convex.
This property does not hold if $\mathbf{dom}\ f$ is not closed. For example, $f(x) = 1$ on $(0, 1)$.

(d) Let $\mathbf{x}_0 \in \mathbf{dom}\ f$ and $\alpha_0 = f(\mathbf{x}_0)$. Then, for any $\mathbf{x} \in C_{\alpha_0}$, we have

$$f(\mathbf{x}_0) \geq f(\mathbf{x}) \geq f(\mathbf{x}_0) + \langle \nabla f(\mathbf{x}_0), \mathbf{x}-\mathbf{x}_0\rangle + \frac{\mu}{2}\|\mathbf{x}-\mathbf{x}_0\|_2^2.$$

Therefore,

$$\frac{\mu}{2}\|\mathbf{x}-\mathbf{x}_0\|_2^2 \le \langle \nabla f(\mathbf{x}_0), \mathbf{x}_0-\mathbf{x}\rangle \le \|\nabla f(\mathbf{x}_0)\|_2\|\mathbf{x}-\mathbf{x}_0\|_2,$$

which implies

$$\|\mathbf{x}-\mathbf{x}_0\|_2 \le \frac{2}{\mu}\|\nabla f(\mathbf{x}_0)\|_2.$$

That is, for any $\alpha \le \alpha_0$, the closed set C_α is bounded, and hence compact. Let $\{\alpha_k\}$ be a sequence converging to $\inf_{\mathbf{x}} f(\mathbf{x})$, where $\inf_{\mathbf{x}} f(\mathbf{x}) < \alpha_k \le \alpha_0$ and thus C_{α_k} is nonempty for all k. Let $\mathbf{x}_k \in C_{\alpha_k}$. Then, by Bolzano-Weierstrass theorem, there exists a subsequence $\{\mathbf{x}_{k_j}\}$ converging to some $\mathbf{x}^*$ such that $f(\mathbf{x}^*) = \inf_{\mathbf{x}} f(\mathbf{x})$.
Assume that there exists another $\mathbf{x}' \in \mathcal{C}$. By the convexity of $f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|_2^2$, we have

$$\begin{aligned} f\left(\frac{\mathbf{x}^*+\mathbf{x}'}{2}\right) &\le \frac{\mu}{2}\left\|\frac{\mathbf{x}^*+\mathbf{x}'}{2}\right\|_2^2 + \frac{1}{2}f(\mathbf{x}^*) - \frac{\mu}{4}\|\mathbf{x}^*\|_2^2 + \frac{1}{2}f(\mathbf{x}') - \frac{\mu}{4}\|\mathbf{x}'\|_2^2 \\ &= \min_{\mathbf{x}} f(\mathbf{x}) - \frac{\mu}{2}\left\|\frac{\mathbf{x}^*-\mathbf{x}'}{2}\right\|_2^2 < \min_{\mathbf{x}} f(\mathbf{x}), \end{aligned}$$

which is a contradiction. Therefore, $\mathbf{x}^*$ is the unique solution of Problem (1). ■

Exercise 2: Operations that Preserve Convexity

1. (a) Let $f : \mathbb{R}^m \to (-\infty, +\infty]$ be a given convex function, $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Please show that

$$F(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b}), \quad \mathbf{x} \in \mathbb{R}^n.$$

is convex.

(b) Let $f_i : \mathbb{R}^n \to (-\infty, +\infty]\,, i = 1, \ldots, m$, be given convex functions. Please show that

$$F(\mathbf{x}) = \sum_{i=1}^{m} w_i f_i(\mathbf{x})$$

is convex, where $w_i \geq 0$, $i = 1, \ldots, m$.

(c) Let $f_i : \mathbb{R}^n \to (-\infty, +\infty]$ be given convex functions for $i \in I$, where I is an arbitrary index set. Please show that the supremum

$$F(\mathbf{x}) = \sup_{i\in I} f_i(\mathbf{x})$$

is convex.

Solution:

(a) Restrict $\mathbf{dom}\ f$ to the image of $\mathbb{R}^m$ under the affine transformation $\mathbf{x} \mapsto \mathbf{A}\mathbf{x} + \mathbf{b}$. Then $\mathbf{dom}\ f$ is convex and the epigraph of f becomes $\{(\mathbf{A}\mathbf{x} + \mathbf{b}, y) : x \in \mathbb{R}^m, y \geq f(\mathbf{A}\mathbf{x} + \mathbf{b})\}$. Since $\mathbf{epi}\ F = \{(\mathbf{x}, y) : \mathbf{x} \in \mathbb{R}^m, y \geq f(\mathbf{A}\mathbf{x} + \mathbf{b})\}$, we have

$$\mathbf{epi}\ f = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{epi}\ F + \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix}.$$

Hence $\mathbf{epi}\ F$ is the inverse image of $\mathbf{epi}\ f$ under an affine transformation. As $\mathbf{epi}\ f$ is convex, $\mathbf{epi}\ F$ is also convex. Therefore, F is convex.

(b) Define $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m, \mathbf{x} \mapsto (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))^\top$. Then we have

$$\begin{pmatrix} \mathbf{x} \\ F(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{0} & \mathbf{w}^\top \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{f}(\mathbf{x}) \end{pmatrix},$$

and thus

$$\mathbf{epi}\ F = \begin{pmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{0} & \mathbf{w}^\top \end{pmatrix} \mathbf{epi}\ \mathbf{f},$$

where $\mathbf{w} = (w_1, \ldots, w_m)^\top$. Clearly, $\mathbf{epi}\ \mathbf{f}$ is convex, and its image under linear transformation is convex, so is $\mathbf{epi}\ F$. Therefore, F is convex.

(c) Because $F(\mathbf{x}) = \sup_{i\in I} f_i(\mathbf{x})$ if and only if $F(\mathbf{x}) \geq f_i(\mathbf{x})$ for all $i \in I$, we have

$$\mathbf{epi}\ F = \bigcap_{i\in I} \mathbf{epi}\ f_i.$$

The intersection of convex sets is convex, so $\mathbf{epi}\ F$ is convex. Therefore, F is convex. ■

2. (Optional) Let $\mathbf{A} \in \mathbb{R}^{n \times m}, \mathbf{x}_0 \in \mathbb{R}^n$. The restriction of $f : \mathbb{R}^n \to \mathbb{R}$ to the affine set $\{\mathbf{A}\mathbf{z} + \mathbf{x}_0 : \mathbf{z} \in \mathbb{R}^m\}$ is defined as the function $F : \mathbb{R}^m \to \mathbb{R}$ with

$$F(\mathbf{z}) = f(\mathbf{A}\mathbf{z} + \mathbf{x}_0)$$

on $\mathbf{dom}\ F = \{\mathbf{z} : \mathbf{A}\mathbf{z} + \mathbf{x}_0 \in \mathbf{dom}\ f\}$. Suppose f is twice differentiable with a convex domain.

   (a) Show that F is convex if and only if for all $\mathbf{z} \in \mathbf{dom}\ F$, we have

$$\mathbf{A}^\top \nabla^2 f(\mathbf{A}\mathbf{z} + \mathbf{x}_0)\mathbf{A} \succeq 0.$$

   (b) Suppose $\mathbf{B} \in \mathbb{R}^{p \times n}$ is a matrix whose nullspace is equal to the range of $\mathbf{A}$, i.e., $\mathbf{B}\mathbf{A} = \mathbf{O}$ and $\operatorname{rank}(\mathbf{B}) = n - \operatorname{rank}(\mathbf{A})$. Show that F is convex if for all $\mathbf{z} \in \mathbf{dom}\ F$, there exists a $\lambda \in \mathbb{R}$ such that

$$\nabla^2 f(\mathbf{A}\mathbf{z} + \mathbf{x}_0) + \lambda \mathbf{B}^\top \mathbf{B} \succeq 0.$$

   (**Hint:** you can use the result as follows. If $\mathbf{C} \in \mathbb{S}^n$ and $\mathbf{D} \in \mathbb{R}^{p \times n}$, then $\mathbf{x}^\top \mathbf{C}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathcal{N}(\mathbf{D})$ if there exists a λ such that $\mathbf{C} + \lambda \mathbf{D}^\top \mathbf{D} \succeq 0$.)

Solution:

   (a) F is convex if and only if for all $\mathbf{z} \in \mathbf{dom}\ F$, $\nabla^2 F(\mathbf{z}) = \mathbf{A}^\top \nabla^2 f(\mathbf{A}\mathbf{z} + \mathbf{x}_0)\mathbf{A} \succeq 0$.

   (b) F is convex if $\mathbf{A}^\top \nabla^2 f(\mathbf{A}\mathbf{z} + \mathbf{x}_0)\mathbf{A} \succeq 0$, or equivalently, if $\mathbf{x}^\top \nabla^2 f(\mathbf{A}\mathbf{z} + \mathbf{x}_0)\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathcal{C}(\mathbf{A})$. Because $\nabla^2 f(\mathbf{A}\mathbf{z} + \mathbf{x}_0) \in \mathbb{S}^n$ and $\mathcal{N}(\mathbf{B}) = \mathcal{C}(\mathbf{A})$, according to the hint, the result holds if there exists a λ such that $\nabla^2 f(\mathbf{A}\mathbf{z} + \mathbf{x}_0) + \lambda \mathbf{B}^\top \mathbf{B} \succeq 0$. ■

3. (Optional)

   (a) Consider the function $f(\mathbf{X}) = \lambda_{\max}(\mathbf{X})$, with $\mathbf{dom}\ f = \mathbb{S}^n$, where $\lambda_{\max}(\mathbf{X})$ is the largest eigenvalue of $\mathbf{X}$ and $\mathbb{S}^n$ is the set of $n \times n$ real symmetric matrices. Show that f is a convex function.

   (b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, with $\mathbf{dom}\ f = \mathbb{R}^n$. Show that it can be represented as the pointwise supremum of a family of affine functions, i.e.,

$$f(\mathbf{x}) = \sup\{g(\mathbf{x}) : g \text{ is affine}, g(\mathbf{z}) \leq f(\mathbf{z}) \text{ for all } \mathbf{z} \in \mathbb{R}^n\}.$$

Solution:

   (a) See Exercise 1 Question 2(e).

   (b) For any $g(\mathbf{x})$, we have $f(\mathbf{x}) \geq g(\mathbf{x})$, so $f(\mathbf{x}) \geq \sup\{g(\mathbf{x}) : g \text{ is affine}, g(\mathbf{z}) \leq f(\mathbf{z}) \text{ for all } \mathbf{z} \in \mathbb{R}^n\}$. Consider the epigraph of f, which is convex. Given any $(\mathbf{x}, f(\mathbf{x})) \in \mathbf{bd\ epi}\ f$, we can fine a supporting hyperplane to $\mathbf{epi}\ f$ at this point, i.e. there exists $\mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}$ such that

$$\left\langle \begin{pmatrix} \mathbf{a} \\ b \end{pmatrix}, \begin{pmatrix} \mathbf{x} \\ f(\mathbf{x}) \end{pmatrix} - \begin{pmatrix} \mathbf{z} \\ t \end{pmatrix} \right\rangle \leq 0$$

   for all $(\mathbf{z}, t) \in \mathbf{epi}\ f$. Letting $\mathbf{z} = \mathbf{x}$, we have $b(f(\mathbf{x}) - t) \leq 0$ for all $t \geq f(\mathbf{x})$, implying that $b \geq 0$. Clearly, $b \neq 0$, otherwise $\mathbf{a}^\top(\mathbf{x} - \mathbf{z}) \leq 0$ for all $\mathbf{z} \in \mathbb{R}^n$, which leads to

$\mathbf{a} = \mathbf{0}$. Therefore, letting $t = f(\mathbf{z})$, we have $f(\mathbf{x}) + b^{-1}\mathbf{a}^\top(\mathbf{x} - \mathbf{z}) \leq f(\mathbf{z})$, where the left hand side, as a function of $\mathbf{z}$, belongs to $\{g(\mathbf{x}) : g \text{ is affine}, g(\mathbf{z}) \leq f(\mathbf{z}) \text{ for all } \mathbf{z} \in \mathbb{R}^n\}$ and equals $f(\mathbf{x})$ at $\mathbf{z} = \mathbf{x}$. Therefore, $f(\mathbf{x}) \leq \sup\{g(\mathbf{x}) : g \text{ is affine}, g(\mathbf{z}) \leq f(\mathbf{z}) \text{ for all } \mathbf{z} \in \mathbb{R}^n\}$, which completes the proof. ■

4. Suppose that the training set is $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where $\mathbf{x}_i \in \mathbb{R}^d$ is the i^{th} data instance and $y_i \in \mathbb{R}$ is the corresponding label. Recall that Lasso is the regression problem:

$$\min_{\mathbf{w}} \frac{1}{2n}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{w}\|_1,$$

where $\mathbf{X} \in \mathbb{R}^{n\times d}$ with its i^{th} row being $\mathbf{x}_i^\top$, $\mathbf{w} \in \mathbb{R}^d$, and $\lambda > 0$ is the regularization parameter. Show that the objective function in the above problem is convex.

Solution:
The objective function is the weighted sum of $f(\mathbf{w}) = \|\mathbf{X}\mathbf{w}-\mathbf{y}\|_2^2$ and $g(\mathbf{w}) = \|\mathbf{w}\|_1$. We know that any p-norm of vectors is convex function, because $\|\theta\mathbf{x}+(1-\theta)\mathbf{y}\|_p \leq \theta\|\mathbf{x}\|_p+(1-\theta)\|\mathbf{y}\|_p$ for all $0 \leq \theta \leq 1$, using the triangle inequality. Therefore, $g(\mathbf{w})$ is convex. For $f(\mathbf{w})$, it is the composition of an affine function and l_2-norm, and according to Exercise 2 Question 1(a), it is convex. Finally, according to Exercise 2 Question 1(b), the objective function is convex. ■

Exercise 3: Subdifferentials

1. Calculation of subdifferentials.

   (a) Let $H \subset \mathbb{R}^n$ be a hyperplane. The extended-value extension of its indicator function I_H is

   $$\tilde{I}_H(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in H, \\ \infty, & \mathbf{x} \notin H. \end{cases}$$

   Find $\partial \tilde{I}_H(\mathbf{x})$.

   (b) Let $f(\mathbf{x}) = \exp \|\mathbf{x}\|_1$, $\mathbf{x} \in \mathbb{R}^n$. Find $\partial f(\mathbf{x})$.

   (c) Let $f(x) = \max\{0, x\}$, $x \in \mathbb{R}$. Find $\partial f(x)$.

   (d) For $\mathbf{x} \in \mathbb{R}^n$, let $x_{[i]}$ be the i^{th} largest component of $\mathbf{x}$. Find the subdifferential of

   $$f(\mathbf{x}) = \sum_{i=1}^{k} x_{[i]}.$$

   (e) Let $f(\mathbf{x}) = \max\limits_{1 \le i \le n} x_i$, $\mathbf{x} \in \mathbb{R}^n$. Find $\partial f(\mathbf{x})$.

   (f) Let $f(\mathbf{x}) = \|\mathbf{x}\|_\infty$, $\mathbf{x} \in \mathbb{R}^n$. Find $\partial f(\mathbf{x})$.

   (g) Let $f(X) = \max\limits_{1 \le i \le n} |\lambda_i|$, where $X \in \mathbb{S}^n$ and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of X. Find $\partial f(X)$.

   (h) (Optional) Let

   $$f(\mathbf{x}) = \left(\sum_{i=1}^{k} x_i^2\right)^{\frac{1}{2}} + \left(\sum_{i=k+1}^{n} x_i^2\right)^{\frac{1}{2}}, \quad \mathbf{x} \in \mathbb{R}^n,$$

   where $1 \le k \le n-1$. Find $\partial f(\mathbf{x})$.

   (i) (Optional) Let $f(\mathbf{X}) = \|\mathbf{X}\|_*$ be the trace norm of $\mathbf{X} \in \mathbb{R}^{m \times n}$. Find $\partial f(\mathbf{X})$.

   Solution:

   (a) Note that $\textbf{epi}\ \tilde{I}_H = H \times \mathbb{R}_+$ is a convex set, so $\tilde{I}_H(\mathbf{x})$ is convex. By definition, $\mathbf{g} \in \partial \tilde{I}_H(\mathbf{x})$ if and only if $\tilde{I}_H(\mathbf{y}) \ge \tilde{I}_H(\mathbf{x}) + \mathbf{g}^\top(\mathbf{y} - \mathbf{x})$ for all $\mathbf{y} \in \mathbb{R}^n$. If $\tilde{I}_H(\mathbf{x}) = \infty$, then $\partial \tilde{I}_H(\mathbf{x}) = \emptyset$. If $\tilde{I}_H(\mathbf{x}) = 0$, we only need to consider the case where $\tilde{I}_H(\mathbf{y}) = 0 < \infty$. That is,

   $$\langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle \le 0, \quad \forall \mathbf{y} \in H.$$

   Hence $\mathbf{x}$ is the projection of $\mathbf{g} + \mathbf{x}$ onto H, which holds if and only if $\mathbf{g} \perp H$. In conclusion,

   $$\partial \tilde{I}_H(\mathbf{x}) = \begin{cases} H^\perp, & \mathbf{x} \in H, \\ \emptyset, & \mathbf{x} \notin H. \end{cases}$$

   Here, we denote $H^\perp$ as the orthogonal complement of the subspace $H - \mathbf{x}, \forall \mathbf{x} \in H$.

(b) Note that
$$f(\mathbf{x}) = \max\left\{\exp\langle \mathbf{s}, \mathbf{x}\rangle : \mathbf{s} \in \mathbb{R}^n, s_i = \pm 1\right\}.$$
Given any $\mathbf{s}$, we see that $\exp\langle \mathbf{s}, \mathbf{x}\rangle$ is closed and convex in $\mathbf{x}$ because it is the composition of a closed convex function and an affine function. Moreover, the gradient of $\exp\langle \mathbf{s}, \mathbf{x}\rangle$ is $\mathbf{s}\exp\langle \mathbf{s}, \mathbf{x}\rangle$. Hence, the subdifferential of $f(\mathbf{x})$ is given by
$$\begin{aligned}
\partial f(\mathbf{x}) &= \mathbf{conv}\ \left\{\mathbf{s}\exp\langle \mathbf{s}, \mathbf{x}\rangle : \mathbf{s} \in \mathbb{R}^n, s_i = \pm 1, \langle \mathbf{s}, \mathbf{x}\rangle = \|\mathbf{x}\|_1\right\} \\
&= \left\{\mathbf{v} \in \mathbb{R}^n : \begin{cases} v_i = \exp\|\mathbf{x}\|_1, & \text{if } x_i > 0, \\ -\exp\|\mathbf{x}\|_1 \le v_i \le \exp\|\mathbf{x}\|_1, & \text{if } x_i = 0, \\ v_i = -\exp\|\mathbf{x}\|_1, & \text{if } x_i < 0 \end{cases}\right\}.
\end{aligned}$$

(c) Note that $f(x) = \max\{f_1(x), f_2(x)\}$, where both $f_1(x) = 0$ and $f_2(x) = x$ are closed and convex. Clearly, we have $\nabla f_1(x) = 0$ and $\nabla f_2(x) = 1$. Hence, the subdifferential of $f(x)$ is given by
$$\begin{aligned}
\partial f(x) &= \mathbf{conv}\left\{\nabla f_i(x) : f_i(x) = f(x)\right\} \\
&= \begin{cases} 0, & \text{if } x < 0, \\ [0,1], & \text{if } x = 0, \\ 1, & \text{if } x > 0. \end{cases}
\end{aligned}$$

(d) Note that
$$f(\mathbf{x}) = \max\left\{\sum_{i \in I} x_i : I \subseteq \{1, \dots, n\}, |I| = k\right\}.$$
Given any I, the linear function $\sum_{i\in I} x_i$ is closed and convex, so f is also closed and convex. Moreover, the gradient of $\sum_{i\in I} x_i$ is $\sum_{i\in I} \mathbf{e}_i$. Hence, the subdifferential of $f(\mathbf{x})$ is given by
$$\begin{aligned}
\partial f(\mathbf{x}) &= \mathbf{conv}\ \left\{\sum_{i \in I} \mathbf{e}_i : I \subseteq \{1, \dots, n\}, |I| = k, \sum_{i\in I} x_i = \sum_{i=1}^{k} x_{[i]}\right\} \\
&= \left\{\mathbf{v} \in \mathbb{R}^n : v_i = \begin{cases} 1, & \text{if } x_i > x_{[k]}, \\ \alpha_i \ge 0, & \text{if } x_i = x_{[k]}, \\ 0, & \text{if } x_i < x_{[k]}; \end{cases} \sum_{x_i = x_{[k]}} \alpha_i = 1\right\},
\end{aligned}$$

(e) Note that $f(\mathbf{x}) = \max_{1\le i\le n} f_i(\mathbf{x})$, where $f_i(\mathbf{x}) = x_i$ is closed and convex and $\nabla f_i(\mathbf{x}) = \mathbf{e}_i$. Hence, the subdifferential of $f(\mathbf{x})$ is given by
$$\begin{aligned}
\partial f(\mathbf{x}) &= \mathbf{conv}\ \left\{\mathbf{e}_i : x_i = x_{[1]}\right\} \\
&= \left\{\mathbf{v} \in \mathbb{R}^n : v_i = \begin{cases} \alpha_i \ge 0, & \text{if } x_i = x_{[1]}, \\ 0, & \text{if } x_i < x_{[1]}; \end{cases} \sum_{x_i = x_{[1]}} \alpha_i = 1\right\}.
\end{aligned}$$

(f) Note that $f(\mathbf{x}) = \max_{1\le i\le n} f_i(\mathbf{x})$, where $f_i(\mathbf{x}) = |x_i|$ is closed and convex and

$$\partial f_i(\mathbf{x}) = \begin{cases} \mathbf{e}_i, & \text{if } x_i > 0, \\ \mathbf{conv}\,\{\pm\mathbf{e}_i\}, & \text{if } x_i = 0, \\ -\mathbf{e}_i, & \text{if } x_i < 0. \end{cases}$$

Hence, the subdifferential of $f(\mathbf{x})$ is given by

$$\begin{aligned} \partial f(\mathbf{x}) &= \mathbf{conv}\,\{\partial f_i(\mathbf{x}) : |x_i| = \|\mathbf{x}\|_\infty\} \\ &= \begin{cases} \mathbf{conv}\,(\{\mathbf{e}_i : x_i = \|\mathbf{x}\|_\infty\} \cup \{-\mathbf{e}_i : x_i = -\|\mathbf{x}\|_\infty\}), & \text{if } \|\mathbf{x}\|_\infty > 0, \\ \mathbf{conv}\bigcup\{\mathbf{conv}\,\{\pm\mathbf{e}_i\} : x_i = 0\}, & \text{if } \|\mathbf{x}\|_\infty = 0 \end{cases} \\ &= \mathbf{conv}\,(\{\mathbf{e}_i : x_i = \|\mathbf{x}\|_\infty\} \cup \{-\mathbf{e}_i : x_i = -\|\mathbf{x}\|_\infty\}) \\ &= \left\{ \mathbf{v} \in \mathbb{R}^n : v_i = \begin{cases} \alpha_i \ge 0, & \text{if } x_i = \|\mathbf{x}\|_\infty \neq 0, \\ \alpha_i \le 0, & \text{if } x_i = -\|\mathbf{x}\|_\infty \neq 0, \\ \alpha_i, & \text{if } x_i = \|\mathbf{x}\|_\infty = 0, \\ 0, & \text{otherwise;} \end{cases} \sum_{|x_i| = \|\mathbf{x}\|_\infty} |\alpha_i| = 1 \right\}. \end{aligned}$$

(g) Note that $f(X) = \max\{\lambda_{\max}(X), -\lambda_{\min}(X)\} = \max\{\lambda_{\max}(X), \lambda_{\max}(-X)\}$. By Exercise 1 Question 2(e) and Exercise 2 Question 1(a), we see that both $\lambda_{\max}(X)$ and $\lambda_{\max}(-X)$ are convex, and clearly also closed, so is $f(X)$.

Suppose that $\lambda_{\max}(X), \lambda_{\min}(X)$ have multiplicity r, s, respectively. Let $U = (\mathbf{u}_1, \ldots, \mathbf{u}_r)$ and $V = (\mathbf{v}_1, \ldots, \mathbf{v}_s)$, where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the corresponding eigenvectors of $\lambda_{\max}(X)$ and $\lambda_{\min}(X)$, respectively. Then

$$\begin{aligned} \partial\lambda_{\max}(X) &= \left\{ U^r G (U^r)^\top : G \succeq 0, \operatorname{tr} G = 1 \right\}, \\ \partial\lambda_{\max}(-X) &= \left\{ -V^s G (V^s)^\top : G \succeq 0, \operatorname{tr} G = 1 \right\}, \end{aligned}$$

If $\lambda_{\max}(X) > -\lambda_{\min}(X)$, then $\partial f(X) = \partial\lambda_{\max}(X)$. If $\lambda_{\max}(X) < -\lambda_{\min}(X)$, then $\partial f(X) = \partial\lambda_{\max}(-X)$. If $\lambda_{\max}(X) = -\lambda_{\min}(X)$, then $\partial\lambda_{\max}(X) = -\partial\lambda_{\max}(-X)$ and hence $\partial f(X) = \mathbf{conv}\,\{\partial\lambda_{\max}(X), -\partial\lambda_{\max}(X)\}$. In conclusion,

$$\partial f(X) = \begin{cases} \left\{ U^r G (U^r)^\top : G \succeq 0, \operatorname{tr} G = 1 \right\}, & \text{if } \lambda_{\max}(X) > -\lambda_{\min}(X), \\ \left\{ -V^s G (V^s)^\top : G \succeq 0, \operatorname{tr} G = 1 \right\}, & \text{if } \lambda_{\max}(X) < -\lambda_{\min}(X), \\ \left\{ \alpha U^r G (U^r)^\top : G \succeq 0, \operatorname{tr} G = 1, |\alpha| \le 1 \right\}, & \text{if } \lambda_{\max}(X) = -\lambda_{\min}(X). \end{cases}$$

(h) Denote $\mathbf{x}_1 = (x_1, \ldots, x_k)$ and $\mathbf{x}_2 = (x_{k+1}, \ldots, x_n)$. Let $\mathbf{A}_1 = (\mathbf{e}_1, \ldots, \mathbf{e}_k, \mathbf{O}) \in \mathbb{R}^{n\times n}$ and $\mathbf{A}_2 = (\mathbf{O}, \mathbf{e}_{k+1}, \ldots, \mathbf{e}_n) \in \mathbb{R}^{n\times n}$. Then

$$f(\mathbf{x}) = \|\mathbf{x}_1\|_2 + \|\mathbf{x}_2\|_2 = \|\mathbf{A}_1\mathbf{x}\|_2 + \|\mathbf{A}_2\mathbf{x}\|_2.$$

By Exercise 2 Question 1(a), we see that both $\|\mathbf{A}_1\mathbf{x}\|_2$ and $\|\mathbf{A}_2\mathbf{x}\|_2$ is convex in $\mathbf{x}$, so is $f(\mathbf{x})$. Note that, by definition, the subdifferential of l_2-norm at $\mathbf{x} = \mathbf{0}$, denoted by $\mathbf{g}$, satisfies $\|\mathbf{x}\|_2 \geq \mathbf{g}^\top \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, or equivalently,

$$\partial \|\mathbf{x}\|_2\big|_{\mathbf{x}=\mathbf{0}} = \left\{ \mathbf{g} : \sup_{\|\mathbf{x}\|_2=1} \mathbf{g}^\top \mathbf{x} \leq 1 \right\}.$$

Hence, the subdifferential of $f(\mathbf{x})$ is given by

$$\begin{aligned} \partial f(X) &= \partial \|\mathbf{A}_1\mathbf{x}\|_2 + \partial \|\mathbf{A}_2\mathbf{x}\|_2 \\ &= \begin{cases} \left(\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|_2}, \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|_2}\right), & \text{if } \|\mathbf{x}_1\|_2 > 0, \|\mathbf{x}_2\|_2 > 0, \\ \left\{ \left(\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|_2}, \mathbf{h}\right) : \sup_{\|\mathbf{x}_2\|_2=1} \mathbf{h}^\top \mathbf{x}_2 \leq 1 \right\}, & \text{if } \|\mathbf{x}_1\|_2 > 0, \|\mathbf{x}_2\|_2 = 0 \\ \left\{ \left(\mathbf{g}, \frac{\mathbf{x}_2}{\|\mathbf{x}_2\|_2}\right) : \sup_{\|\mathbf{x}_1\|_2=1} \mathbf{g}^\top \mathbf{x}_1 \leq 1 \right\}, & \text{if } \|\mathbf{x}_1\|_2 = 0, \|\mathbf{x}_2\|_2 > 0, \\ \left\{ (\mathbf{g}, \mathbf{h}) : \sup_{\|\mathbf{x}_1\|_2=1} \mathbf{g}^\top \mathbf{x}_1 \leq 1, \sup_{\|\mathbf{x}_2\|_2=1} \mathbf{h}^\top \mathbf{x}_2 \leq 1 \right\}, & \text{if } \|\mathbf{x}_1\|_2 = 0, \|\mathbf{x}_2\|_2 = 0. \end{cases} \end{aligned}$$

(i) The trace norm is defined as

$$f(\mathbf{X}) = \|\mathbf{X}\|_* = \operatorname{tr}\left(\sqrt{\mathbf{X}^\top \mathbf{X}}\right) = \sum_{i=1}^n \sigma_i(\mathbf{X}).$$

The subdifferential of the trace norm is

$$\partial f(\mathbf{X}) = \left\{ \mathbf{U}\mathbf{V}^\top + \mathbf{W} : \|\mathbf{W}\|_2 \leq 1, \mathbf{U}^\top \mathbf{W} = \mathbf{W}\mathbf{V} = \mathbf{O} \right\}.$$

where $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top = \mathbf{X}$ is the singular value decomposition of $\mathbf{X}$.

To be proved.

■

Exercise 4: Problems from the Lecture Notes

1. **Mean Value Theorem in Vector Functions.** The mean value theorem is generally not holds in vector valued functions. In the proof of Theorem 5 in Lecture 06, we use some techniques to avoid applying mean value theorem directly on vector valued functions.

   Theorem 5 in Lecture 06. Suppose that f is twice continuously differentiable. Then, f is convex if and only if **dom** f is convex and $\nabla^2 f(\mathbf{x}) \succeq 0$.

   In the proof of necessity, we let $\mathbf{x}_t = \mathbf{x} + t\mathbf{s}, t > 0$.

   (a) We write down the following formula without proof in class,

   $$\begin{aligned} 0 \leq \frac{1}{t^2}\langle\nabla f\left(\mathbf{x}_t\right)-\nabla f(\mathbf{x}), \mathbf{x}_t-\mathbf{x}\rangle &= \frac{1}{t}\langle\nabla f\left(\mathbf{x}_t\right)-\nabla f(\mathbf{x}), \mathbf{s}\rangle \\ &= \frac{1}{t}\int_0^t\langle\nabla^2 f(\mathbf{x}+\tau\mathbf{s})\mathbf{s}, \mathbf{s}\rangle \mathrm{d}\tau. \end{aligned}$$

   Please show why the second equality holds.

   (b) By the mean value theorem, we can find an $\alpha \in (0, t)$ such that

   $$\int_0^t \left\langle\nabla^2 f(\mathbf{x}+\tau\mathbf{s})\mathbf{s}, \mathbf{s}\right\rangle \mathrm{d}\tau = t\left\langle\nabla^2 f(\mathbf{x}+\alpha\mathbf{s})\mathbf{s}, \mathbf{s}\right\rangle.$$

   Please explain how we use the mean value theorem in detail.

   Solution:

   (a) By Newton-Leibniz formula,

   $$\nabla f\left(\mathbf{x}_t\right) = \nabla f(\mathbf{x}) + \int_{\mathbf{x}}^{\mathbf{x}_t} \nabla^2 f(\mathbf{z})\mathrm{d}\mathbf{z} = \nabla f(\mathbf{x}) + \int_0^t \nabla^2 f(\mathbf{x}+\tau\mathbf{s})\mathrm{d}\tau.$$

   Since integral is linear and $\mathbf{s}$ is a constant vector, we have

   $$\langle\nabla f\left(\mathbf{x}_t\right)-\nabla f(\mathbf{x}), \mathbf{s}\rangle = \langle\int_0^t \nabla^2 f(\mathbf{x}+\tau\mathbf{s})\mathrm{d}\tau, \mathbf{s}\rangle = \int_0^t\langle\nabla^2 f(\mathbf{x}+\tau\mathbf{s})\mathbf{s}, \mathbf{s}\rangle\mathrm{d}\tau.$$

   (b) Let $g(\xi) = \int_0^\xi \left\langle\nabla^2 f(\mathbf{x}+\tau\mathbf{s})\mathbf{s}, \mathbf{s}\right\rangle \mathrm{d}\tau$. Then, g is a continuous function on $[0, t]$. By the mean value theorem, there exists an $\alpha \in (0, t)$ such that

   $$g(t) - g(0) = g'(\alpha)(t-0),$$

   or equivalently,

   $$\int_0^t \left\langle\nabla^2 f(\mathbf{x}+\tau\mathbf{s})\mathbf{s}, \mathbf{s}\right\rangle \mathrm{d}\tau - 0 = \left\langle\nabla^2 f(\mathbf{x}+\alpha\mathbf{s})\mathbf{s}, \mathbf{s}\right\rangle (t-0).$$

   This is the desired result. ■

2. **Log-determinant Function.** Recall Example 3 in Lecture06, the log-determinant function $f(X) = -\log\det X$ with $\mathbf{dom}\ f = \mathbb{S}^n_{++}$. Let $X_0 \in \mathbb{S}^n_{++}$ and $V \in \mathbb{S}^n$. We define

$$g(t) = f\left(X_0 + tV\right)$$

with $\mathbf{dom}\ g = \left\{t : X_0 + tV \in \mathbb{S}^n_{++}\right\}$.

   (a) Please show that $\mathbf{dom}\ g$ is nonempty.

   (b) (**Optional**) Please find $\mathbf{dom}\ g$.

   Solution:

   (a) $X_0 + 0V = X_0 \in \mathbf{dom}\ f = \mathbb{S}^n_{++}$, so $0 \in \mathbf{dom}\ g$.

   (b) Note that

$$X_0 + tV = X_0^{\frac{1}{2}}(\mathbf{I} + tX_0^{-\frac{1}{2}}VX_0^{-\frac{1}{2}})X_0^{\frac{1}{2}},$$

   which is positive definite if and only if $\mathbf{I} + tX_0^{-\frac{1}{2}}VX_0^{-\frac{1}{2}} \succ 0$. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $X_0^{-\frac{1}{2}}VX_0^{-\frac{1}{2}}$. Then, the eigenvalues of $\mathbf{I}+tX_0^{-\frac{1}{2}}VX_0^{-\frac{1}{2}}$ become $1+t\lambda_1, \ldots,$ $1 + t\lambda_n$, which are all positive if and only if $t \in \mathbf{dom}\ g$. That is,

$$\mathbf{dom}\ g = \begin{cases} (-\frac{1}{\lambda_{\max}}, \infty), & \text{if } \lambda_{\min} \geq 0, \lambda_{\max} > 0, \\ (-\frac{1}{\lambda_{\max}}, -\frac{1}{\lambda_{\min}}), & \text{if } \lambda_{\min} < 0 < \lambda_{\max}, \\ (-\infty, -\frac{1}{\lambda_{\min}}), & \text{if } \lambda_{\min} < 0, \lambda_{\max} \leq 0, \\ (-\infty, \infty), & \text{if } \lambda_{\min} = \lambda_{\max} = 0. \end{cases}$$

   where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and the largest eigenvalues of $X_0^{-\frac{1}{2}}VX_0^{-\frac{1}{2}}$, respectively. ■

3. **Subdifferential.** Recall the Example 5 in Lecture 07. Let $f : \mathbb{S}^n \to \mathbb{R}$ be defined by $f(X) = \lambda_{\max}(X)$. We want to Find $\partial f(X)$.

   By eigen-decomposition, a symmetric matrix can be written as $X = U\Lambda U^\top$, where $U^\top U = I$ and $\Lambda = \operatorname{diag}\left(\lambda_1, \ldots, \lambda_n\right)$ with $\lambda_1 \geq \cdots \geq \lambda_n$. Let $U = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$, i.e., $\mathbf{u}_i$ is the eigenvector corresponding to λ_i. We then write $f(X)$ as

$$f(X) = \max\{\langle \mathbf{s}, X\mathbf{s}\rangle : \|\mathbf{s}\| = 1\} = \max\{\langle \mathbf{s}\mathbf{s}^\top, X\rangle : \|\mathbf{s}\| = 1\}.$$

   Assume that $\lambda_{\max} = \lambda_1 = \cdots = \lambda_r$, where $1 \leq r \leq n$. Let $U^r = (\mathbf{u}_1, \ldots, \mathbf{u}_r)$,

$$S^* := \underset{\|\mathbf{s}\|=1}{\mathbf{argmax}}\, \langle \mathbf{s}\mathbf{s}^\top, X\rangle$$

   (a) Please show that

$$S^* = \{\mathbf{v} : \mathbf{v} \in \mathbf{span}\ U^r, \|\mathbf{v}\| = 1\} = \{\mathbf{v} : \mathbf{v} = U^r\mathbf{q}, \mathbf{q} \in \mathbb{R}^r, \|\mathbf{q}\| = 1\}.$$

   (b) Please find $\frac{\mathrm{d}}{\mathrm{d}X}\langle \mathbf{s}\mathbf{s}^\top, X\rangle$, then show that

$$\partial f(X) = \mathbf{conv}\left\{\mathbf{v}\mathbf{v}^\top : \mathbf{v} \in S^*\right\} = \left\{U^r G\left(U^r\right)^\top : G \succeq 0, \operatorname{tr} G = 1\right\}.$$

(c) Suppose $n = 3$, please find $\partial f(X)$ at $X = \operatorname{diag}(2, 4, 4)$ and $X = \operatorname{diag}(1, 2, 4)$.

Solution:

(a) Let $\mathbf{s} = \mathbf{v} + \mathbf{w}$, where $\mathbf{v} = U\mathbf{q} \in \mathbf{span}\ U$ and $\mathbf{w} \in (\mathbf{span}\ U)^\perp$. Then

$$\begin{aligned}\left\langle \mathbf{s}\mathbf{s}^\top, X \right\rangle &= \left\langle U^\top(\mathbf{v}+\mathbf{w})(U^\top(\mathbf{v}+\mathbf{w}))^\top, \Lambda \right\rangle \\ &= \left\langle U^\top \mathbf{v}\mathbf{v}^\top U, \Lambda \right\rangle \\ &= \left\langle \mathbf{q}\mathbf{q}^\top, \Lambda \right\rangle \\ &= \sum_{i=1}^n \lambda_i q_i^2 \le \lambda_1 \|\mathbf{q}\|^2 \le \lambda_1.\end{aligned}$$

The first inequality becomes equality if and only if $q_i = 0$ for $i > r$, i.e. $\mathbf{v} \in \mathbf{span}\ U^r$. The second inequality becomes equality if and only if $\|\mathbf{q}\| = \|\mathbf{v}\| = 1$, i.e. $\mathbf{s} = \mathbf{v}$. Thus, $S^* = \{\mathbf{v} : \mathbf{v} \in \mathbf{span}\ U^r, \|\mathbf{v}\| = 1\} = \{\mathbf{v} : \mathbf{v} = U^r\mathbf{q}, \mathbf{q} \in \mathbb{R}^r, \|\mathbf{q}\| = 1\}$.

(b) We have

$$\frac{\partial}{\partial x_{ij}} \langle \mathbf{s}\mathbf{s}^\top, X \rangle = s_i s_j \implies \mathrm{d}\langle \mathbf{s}\mathbf{s}^\top, X \rangle = \langle \mathbf{s}\mathbf{s}^\top, \mathrm{d}X \rangle \implies \frac{\mathrm{d}}{\mathrm{d}X} \langle \mathbf{s}\mathbf{s}^\top, X \rangle = \langle \mathbf{s}\mathbf{s}^\top, \cdot \rangle.$$

That is, $\nabla_X \langle \mathbf{s}\mathbf{s}^\top, X \rangle = \mathbf{s}\mathbf{s}^\top$. Clearly, $\langle \mathbf{s}\mathbf{s}^\top, X \rangle$ is closed and convex in X. Moreover, $\mathbf{s}$ belongs to a compact set and $\langle \mathbf{s}\mathbf{s}^\top, X \rangle$ is continuous in $\mathbf{s}$. Hence, the subdifferential of $f(X)$ is given by

$$\begin{aligned}\partial f(X) &= \mathbf{conv} \left\{ \nabla_X \langle \mathbf{s}\mathbf{s}^\top, X \rangle : \mathbf{s} = f(X) \right\} \\ &= \mathbf{conv} \left\{ \mathbf{v}\mathbf{v}^\top : \mathbf{v} \in S^* \right\} \\ &= \mathbf{conv} \left\{ U^r \mathbf{q}\mathbf{q}^\top (U^r)^\top : \mathbf{q} \in \mathbb{R}^r, \|\mathbf{q}\| = 1 \right\} \\ &= \mathbf{conv} \left\{ U^r Q (U^r)^\top : Q \succeq 0, \mathbf{rank}\,(Q) = 1, \operatorname{tr} Q = 1 \right\} \\ &= \left\{ U^r G (U^r)^\top : G \succeq 0, \operatorname{tr} G = 1 \right\}.\end{aligned}$$

To see that the fourth equality holds, note that $\mathbf{rank}\left(\mathbf{q}\mathbf{q}^\top\right) = \mathbf{rank}\left(\mathbf{q}^\top\mathbf{q}\right) = 1$, $\operatorname{tr}\left(\mathbf{q}\mathbf{q}^\top\right) = \operatorname{tr}\left(\mathbf{q}^\top\mathbf{q}\right) = 1$ and $(\mathbf{x}^\top\mathbf{q})^2 \ge 0, \forall \mathbf{x} \in \mathbb{R}^r$. Conversely, any positive semidefinite Q with a single nonzero eigenvalue $\lambda = 1$ and the corresponding eigenvector $\mathbf{q}$ can be decomposed as $Q = \lambda\mathbf{q}\mathbf{q}^\top = \mathbf{q}\mathbf{q}^\top$.

To see that the last equality holds, note that $\{G \in \mathbb{S}_+ : \operatorname{tr} G = 1\}$ is a convex set, so $\{G \in \mathbb{S}_+ : \operatorname{tr} G = 1\} \supset \mathbf{conv}\ \{Q \in \mathbb{S}_+ : \mathbf{rank}\,(Q) = 1, \operatorname{tr} Q = 1\}$. Conversely, any G can be decomposed as $\sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^\top$, where λ_i are the eigenvalues of G and $\mathbf{u}_i$ are the corresponding unit eigenvectors. Since $\|\mathbf{u}_i\| = 1$ and $\sum_{i=1}^n \lambda_i = 1$, we have $G \in \mathbf{conv}\ \{Q \in \mathbb{S}_+ : \mathbf{rank}\,(Q) = 1, \operatorname{tr} Q = 1\}$. Therefore,

$$\{G \in \mathbb{S}_+ : \operatorname{tr} G = 1\} = \mathbf{conv}\ \{Q \in \mathbb{S}_+ : \mathbf{rank}\,(Q) = 1, \operatorname{tr} Q = 1\}.$$

It is easy to see that, under an affine transformation, the image of the convex hull of a set is the same as the convex hull of the image of the set. This completes the proof of the last equality.

(c) i. For $X = \operatorname{diag}(2, 4, 4)$, we have

$$U^2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \ G = \begin{pmatrix} \frac{1}{2} + a & b \\ b & \frac{1}{2} - a \end{pmatrix}, \text{ where } a^2 + b^2 \le \frac{1}{4}.$$

Hence the subdifferential is

$$\partial f(X) = \left\{ \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} + a & b \\ 0 & b & \frac{1}{2} - a \end{pmatrix} : a^2 + b^2 \le \frac{1}{4} \right\}.$$

ii. For $X = \operatorname{diag}(1, 2, 4)$, we have

$$U^1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \ G = 1.$$

Hence the subdifferential is

$$\partial f(X) = \left\{ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}.$$

■