Introduction to Machine Learning
Fall 2023
University of Science and Technology of China

Lecturer: Jie Wang | Homework 4
Posted: Nov. 1, 2023 | Due: Nov. 8, 2023

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

Exercise 1: Convex Functions

1. Please show that the following functions are convex.
   (a) $f(\mathbf{x}) = \sum_{i=1}^{k} x_{[i]}$ on $\mathbf{dom}\ f = \mathbb{R}^n$, where $1 \leq k \leq n$ and $x_{[i]}$ denotes the i^{th} largest component of $\mathbf{x}$.
   (b) 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 $\mathbf{p}$.
   (c) 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 $\mathbf{X}$.

2. please show the following two equalities:
$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))^\top (\mathbf{y} - \mathbf{x}) \mathrm{d}t \tag{1}$$
$$\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}) = \int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) \mathrm{d}t \tag{2}$$
   (**Hint:** you may consider the function $g(t) = f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))$ and apply the fundamental theorem of calculus.)

3. (Optional) Please show that a continuously differentiable function f is strongly convex with parameter $\mu > 0$ if and only if
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{\mu}{2}\|\mathbf{y} - \mathbf{x}\|_2^2, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

4. (Optional) 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 $\nabla^2 f(\mathbf{x})$.

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 $\nabla^2 f(\mathbf{x})$.

6. Consider the problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}), \tag{3}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and convex, and **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 (3) 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 **dom** f is not closed?

(d) Suppose that f is strongly convex with parameter $\mu > 0$. Please show that Problem (3) admits a unique solution.

Solution: 1. (a) We know that function $f(\mathbf{x}) = \max\{f_1(\mathbf{x}, f_2(\mathbf{x}, \cdots, f_n(\mathbf{x},)\}$, where $f_1, f_2, \cdots, f_n$ are all convex functions, is a convex function.
$f(\mathbf{x}) = \sum_{i=1}^{k} x_{[i]} = \max\{\mathbf{x}_{i_1} + \mathbf{x}_{i_2} + \cdots + \mathbf{x}_{i_k} | 1 \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq n\}$. Actually the function is pointwise maximum of $\frac{n!}{r!(n-r)!}$ linear functions. And we know linear functions are convex. Thus $f(\mathbf{x}) = \sum_{i=1}^{k} x_{[i]}$ is convex.

(b) Consider the function $g : R_+ \to R, g(x) = x\log(x)$. We can easily find the second derivative of g: $g''(x) = \frac{1}{x} > 0$. It means g is convex function.
f is a linear combination of convex function g, and all coefficients are non-negative. Therefore f is convex function.

(c) We know that

$$f(\mathbf{X}) = \|\mathbf{X}\|_2 = \sigma_{\max}(\mathbf{X}) = \sqrt{\lambda_{\max}(\mathbf{X}^\top\mathbf{X})} = \max_{\|\mathbf{v}\|=1} \|\mathbf{X}v\|_2$$

Consider the function $g(\mathbf{X}) = \|\mathbf{X}\mathbf{v}\|_2$, $\forall \mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m\times n}$, $\forall\theta \in [0, 1]$, we have

$$g(\theta\mathbf{X} + (1-\theta)\mathbf{Y}) = \|\theta\mathbf{X} + (1-\theta)\mathbf{Y}\|_2 \leq \theta\|\mathbf{X}\|_2 + (1-\theta)\|\mathbf{Y}\|_2$$

g is convex function. f is maximum of several convex functions, thus $f(\mathbf{X}) = \|\mathbf{X}\|_2 = \sigma_{\max}(\mathbf{X})$ is convex function.

2. Applying the fundamental theorem of calculus, we can get

$$g(1) - g(0) = \int_0^1 g'(t)dt$$

Note that $g(1) = f(\mathbf{y}), g(0) = f(\mathbf{x})$, and $g'(t) = \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))^\top(\mathbf{y} - \mathbf{x})$. Thus we can derive that

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \nabla f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))^\top(\mathbf{y} - \mathbf{x})\mathrm{d}t$$

We also can get the second derivative $g''(t) = (\nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}))^\top(\mathbf{y} - \mathbf{x})$. And $g'(1) = \nabla f(\mathbf{y})^\top(\mathbf{y} - \mathbf{x})$, $g'(0) = \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x})$. Applying the fundamental theorem of calculus, we can get

$$\begin{aligned} g'(1) - g'(0) &= \int_0^1 g''(t)dt \\ (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}))^\top(\mathbf{y} - \mathbf{x}) &= \int_0^1 (\nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}))^\top(\mathbf{y} - \mathbf{x})dt \\ \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}) &= \int_0^1 \nabla^2 f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x})dt \end{aligned}$$

3.

4.

5. From the definition of the Hessian of a twice differentiable function f, we know that for any vector $\mathbf{v} \in \mathbb{R}^n$,

$$\begin{aligned} \nabla^2 f(\mathbf{x})\mathbf{v} &= \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} \\ \implies \|\nabla^2 f(\mathbf{x})\mathbf{v}\| &\le \lim_{h \to 0} \frac{\|f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})\|}{|h|} \\ &\le L \lim_{h \to 0} \frac{\|h\mathbf{v}\|}{|h|} \\ &= L\|\mathbf{v}\| \end{aligned}$$

Since this is true for any $\mathbf{v} \in \mathbf{R}^n$, it is also true for the eigenvectors for $\nabla^2 f(\mathbf{x})$. When $\mathbf{v}$ is eigenvectors,

$$\|\nabla^2 f(\mathbf{x})\mathbf{v}\| = |\lambda|\|\mathbf{v}\| \le L\|\mathbf{v}\|$$

It means $\lambda_{\max}(\nabla^2 f(\mathbf{x})) \le L$.

6. (a) Suppose that C_α is open. It means $\exists \mathbf{x}' \notin C_\alpha$ is a limit point of C_α. According to the definition of limit point, we know there exists a sequence $(\mathbf{x}_k) \subseteq C_\alpha$ that converges to $\mathbf{x}'$.
$\forall \mathbf{x}_k \in C_\alpha, \mathbf{x}_k \neq \mathbf{x}'$ and $\mathbf{x}_k \to \mathbf{x}'$. Because f is continuous, we have $f(\mathbf{x}_k) \to f(\mathbf{x}') > \alpha$. But from the definition of C_α, $f(\mathbf{x}_k) \le \alpha$. This is contradictory. Therefore, C_α is closed.

   (b) $f(x) = e^x$, **dom** $f = \mathbb{R}$.

   (c) Firstly, we show $\mathcal{C}$ is convex. $\forall \mathbf{y_1}, \mathbf{y_2} \in \mathcal{C}$, $\forall \theta \in [0, 1]$, we have

$$\begin{aligned} f(\theta\mathbf{y_1} + (1 - \theta)\mathbf{y_2}) &\le \theta f(\mathbf{y_1}) + (1 - \theta)f(\mathbf{y_2}) \\ &= \theta \min_{\mathbf{x}} f(\mathbf{x}) + (1 - \theta) \min_{\mathbf{x}} f(\mathbf{x}) \\ &= \min_{\mathbf{x}} f(\mathbf{x}) \end{aligned}$$

   Thus $\theta\mathbf{y_1} + (1 - \theta)\mathbf{y_2} \in \mathcal{C}$, it means $\mathcal{C}$ is a convex set.
   Next we show $\mathcal{C}$ is closed. Suppose $\mathcal{C}$ is open, there exists a limit point $\mathbf{y}' \notin \mathcal{C}$.

According to the definition of limit point, we know there is a sequence $(\mathbf{y_k}) \subseteq \mathcal{C}$ that converges to $\mathbf{y}'$. Because f is continuous, we have

$$\lim_{k\to\infty} f(\mathbf{y_k}) = f(\lim_{k\to\infty} \mathbf{y}_k) = f(\mathbf{y}') > \min_{\mathbf{x}} f(\mathbf{x})$$

But because $\mathbf{y}_k \in \mathcal{C}$, $f(\mathbf{y}_k) = \min_{\mathbf{x}} f(\mathbf{x})$. This is contradictory. Therefore $\mathcal{C}$ is closed. This property still hold if **dom** f is not closed.

(d) Suppose $\exists \mathbf{y_1}, \mathbf{y_2} \in \mathbf{dom}\ f$ and $\mathbf{y_1} \neq \mathbf{y_2}$, such that $f(\mathbf{y_1}) = f(\mathbf{y_2}) = \min f$. Because f is strongly convex with parameter $\mu > 0$, we have $\exists \theta = \frac{1}{2}$, such that

$$\begin{aligned} f(\frac{\mathbf{y_1}+\mathbf{y_2}}{2}) - \frac{\mu}{2}\|\frac{\mathbf{y_1}+\mathbf{y_2}}{2}\|_2^2 &\leq \frac{1}{2}(\min f - \frac{\mu}{2}\|\mathbf{y_1}\|) + \frac{1}{2}(\min f - \frac{\mu}{2}\|\mathbf{y_2}\|) \\ &= \min f - \frac{\mu}{2}(\frac{1}{2}\|\mathbf{y_1}\| + \frac{1}{2}\|\mathbf{y_2}\|) \\ \implies f(\frac{\mathbf{y_1}+\mathbf{y_2}}{2}) &\leq \min f + \frac{\mu}{4}(\frac{1}{2}\|\mathbf{y_1}+\mathbf{y_2}\|_2^2 - \|\mathbf{y_1}\|_2^2 - \|\mathbf{y_2}\|_2^2) \end{aligned}$$

Next step, we consider this term : $\frac{1}{2}\|\mathbf{y_1}+\mathbf{y_2}\|_2^2 - \|\mathbf{y_1}\|_2^2 - \|\mathbf{y_2}\|_2^2$.

$$\frac{1}{2}\|\mathbf{y_1}+\mathbf{y_2}\|_2^2 - \|\mathbf{y_1}\|_2^2 - \|\mathbf{y_2}\|_2^2 = \mathbf{y_1}^\top \mathbf{y_2} - \frac{1}{2}\|\mathbf{y_1}\|_2^2 - \frac{1}{2}\|\mathbf{y_2}\|_2^2 \geq 0$$

Note that the above equation only equals zero when $\mathbf{y_1} = \mathbf{y_2}$.We know $\mu > 0$, thus

$$f(\frac{\mathbf{y_1}+\mathbf{y_2}}{2}) < \min f$$

Obviously this contradicts our premise. Thus if f is strongly convex with parameter $\mu > 0$, Problem (3) admits a unique solution.

■

Exercise 2: Operations that Preserve Convexity

1. 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.

2. 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 $w_i \geq 0$, $i = 1, \ldots, m$.

3. 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:

1. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^m$, $\forall \theta \in [0, 1]$, we have

$$\begin{aligned} F(\theta\mathbf{x} + (1-\theta)\mathbf{y}) &= f(\mathbf{A}(\theta\mathbf{x} + (1-\theta)\mathbf{y}) + \mathbf{b}) \\ &= f(\theta(\mathbf{A}\mathbf{x} + \mathbf{b}) + (1-\theta)(\mathbf{A}\mathbf{y} + \mathbf{b})) \\ &\leq \theta f(\mathbf{A}\mathbf{x} + \mathbf{b}) + (1-\theta) f(\mathbf{A}\mathbf{y} + \mathbf{b}) \\ &= \theta F(\mathbf{x}) + (1-\theta)F(\mathbf{y}) \end{aligned}$$

   Therefore, F is convex.

2. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^m$, $\forall \theta \in [0, 1]$, we have

$$\begin{aligned} F(\theta\mathbf{x} + (1-\theta)\mathbf{y}) &= \sum_{i=1}^{m} w_i f_i(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \\ &\leq \sum_{i=1}^{m} w_i(\theta f_i(\mathbf{x}) + (1-\theta)f(\mathbf{y})) \\ &= \theta\sum_{i=1}^{m} w_i f_i(\mathbf{x}) + (1-\theta)\sum_{i=1}^{m} w_i f_i(\mathbf{y}) \\ &= \theta F(\mathbf{x}) + (1-\theta)F(\mathbf{y}) \end{aligned}$$

   Therefore, F is convex.

3. Obviously, from the definition of epigraph, we know **epi** $F = \bigcap_{i\in I}$ **epi** f_i. Because f_i is convex function, **epi** f_i is convex set. Therefore **epi** $F = \bigcap_{i\in I}$ **epi** f_i is also convex set. It means that F is convex function.

■

Exercise 3: Subdifferentials

Calculation of subdifferentials (you need to finish at least four of the problems).

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

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

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

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

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

5. 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 $\partial f(X)$(**Hint**: you can refer to Example 5 in Lec06).

Solution:

1. From the definition, we know $\forall \mathbf{g} \in \partial \tilde{I}_H(\mathbf{x}), \tilde{I}_H(\mathbf{y}) \ge \tilde{I}_H(\mathbf{x}) + < \mathbf{g}, \mathbf{y} - \mathbf{x} >$. To make the above formula true for all y it can be written as

$$0 \ge < \mathbf{g}, \mathbf{y} - \mathbf{x} >, \forall \mathbf{y} \in H$$

It means $\mathbf{g}^\top \mathbf{y} \le \mathbf{g}^\top \mathbf{x}$. Therefore $\partial \tilde{I}_H(\mathbf{x}) = \{\mathbf{g} \in \mathbb{R}^n : \mathbf{g}^\top \mathbf{y} \le \mathbf{g}^\top \mathbf{x}, \forall y \in H\}$

2. $f(\mathbf{x}) = \exp \|\mathbf{x}\|_1 = \max\{e^{\mathbf{s}^\top \mathbf{x}} : s \in \mathbb{R}^n, |s_i| = 1, i = 1, 2, \cdots, n\}$. Let $\phi(\mathbf{s}, \mathbf{x}) = e^{\mathbf{s}^\top \mathbf{x}}$ and $\Delta = \{s \in \mathbb{R}^n, |s_i| = 1, i = 1, 2, \cdots, n\}$, we have $\partial f(\mathbf{x}) = \textbf{conv}\,\{\partial \phi(\mathbf{s}, \mathbf{x}) : \phi(\mathbf{s}, \mathbf{x}) = f(\mathbf{x})\}$.
Because $\phi(\mathbf{s}, \mathbf{x})$ is continuous, we can get its subdifferentials easily

$$\partial \phi(\mathbf{s}, \mathbf{x}) = e^{\mathbf{s}^\top \mathbf{x}} \mathbf{s}^\top$$

Therefore,

$$\begin{aligned} \partial f(\mathbf{x}) &= \textbf{conv}\,\{e^{\mathbf{s}^\top \mathbf{x}} \mathbf{s}^\top : e^{\mathbf{s}^\top \mathbf{x}} = \exp \|\mathbf{x}\|_1\} \\ &= \{\mathbf{v} \in \mathbb{R}^n : v_i = \begin{cases} \exp(\mathbf{s}^\top \mathbf{x}), x_i > 0 \\ [-\exp(\mathbf{s}^\top \mathbf{x}), \exp(\mathbf{s}^\top \mathbf{x})], x_i = 0 \\ -\exp(\mathbf{s}^\top \mathbf{x}), x_i < 0 \end{cases} \} \end{aligned}$$

3. $f(\mathbf{x}) = \sum_{i=1}^{k} x_{[i]} = \max\{\mathbf{x}_{i_1} + \mathbf{x}_{i_2} + \cdots + \mathbf{x}_{i_k} | 1 \le i_1 \le i_2 \le \cdots \le i_k \le n\} = \max_{1 \le i \le \frac{n!}{k!(n-k)!}} f_i(\mathbf{x})$, where $f_i(\mathbf{x}) = \mathbf{x}_{i_1} + \mathbf{x}_{i_2} + \cdots + \mathbf{x}_{i_k}$. Because $f_i(\mathbf{x})$ is continuous, we can get its subdifferentials easily: $\partial f_i(\mathbf{x})$ is a n-dimensions vector, only k

dimensions is 1 and the rest of all is 0.

$$\partial f(\mathbf{x}) = \mathbf{conv}\left\{\partial f_i(\mathbf{x}) : f(\mathbf{x}) = f_i(\mathbf{x})\right\}$$

4. $f(\mathbf{x}) = \|\mathbf{x}\|_\infty = \max_{1\le i\le n}\{|x_i|\}$. Therefore,

$$\partial f(\mathbf{x}) = \mathbf{conv}\left\{\partial f_i(\mathbf{x}) : f(\mathbf{x}) = f_i(\mathbf{x})\right\},$$

where $f_i(\mathbf{x}) = |x_i|$. We know $\partial|x_i| = \begin{cases} 1, x_i > 0 \\ [-1,1], x_i = 0 \\ -1, x_i < 0 \end{cases}$

■