Homework

PB22010344 黄境

2024 年 11 月 8 日

Exercise 1. **Convex Functions**

Solution. 1.(a) $\forall\, \mathbf{x} = (x_1, x_2, \ldots, x_n)$, $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ and $\theta \in [0,1]$, let $\mathbf{z}_\theta = \theta\mathbf{x} + (1-\theta)\mathbf{y}$, we have

$$f(z_\theta) = \sum_{i=1}^{k} z_{[i]}.$$

Denote $z_{[i]}$ by z_{j_i}. Since $z_i = \theta x_i + (1-\theta)y_i$, we have

$$\begin{aligned}
\sum_{i=1}^{k} z_{[i]} = \sum_{i=1}^{k} z_{j_i} &= \sum_{i=1}^{k} \theta x_{j_i} + (1-\theta)y_{j_i} \\
&= \theta \sum_{i=1}^{k} x_{j_i} + (1-\theta)\sum_{i=1}^{k} y_{j_i} \\
&\le \theta \sum_{i=1}^{k} x_{[i]} + (1-\theta)\sum_{i=1}^{k} y_{[i]} \\
&= \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})
\end{aligned}$$

Therefore, f is convex.

(b) $\forall\, \mathbf{p}_1, \mathbf{p}_2 \in \mathbf{dom}\, f$ and $\theta \in [0,1]$, let $\mathbf{p}_\theta = \theta\mathbf{p}_1 + (1-\theta)\mathbf{p}_2$. Since

$$\nabla^2 x ln x = \frac{d^2 x \ln x}{dx^2} = \frac{1}{x} > 0,\, \forall\, x > 0,$$

we have $f(x) = xlnx$ is convex. Therefore,

$$\begin{aligned}
f(\mathbf{p}_\theta) &= \sum_{i=1}^{n} (\theta p_{1i} + (1-\theta)p_{2i}) \ln(\theta p_{1i} + (1-\theta)p_{2i}) \\
&\le \sum_{i=1}^{n} (\theta p_{1i} \ln p_{1i} + (1-\theta)p_{2i} \ln p_{2i}) \\
&= \theta \sum_{i=1}^{n} p_{1i} \ln p_{1i} + (1-\theta) \sum_{i=1}^{n} p_{2i} \ln p_{1i} \\
&= \theta f(\mathbf{p}_1) + (1-\theta) f(\mathbf{p}_2)
\end{aligned}$$

Therefore, f is convex.

(c) Since

$$\|\mathbf{X}\|_p = \sup_{\substack{\mathbf{a}\in\mathbb{R}^{n\times 1} \\ \|\mathbf{a}\|=1}} \|\mathbf{X}\mathbf{a}\|_p,$$

$\forall\, \mathbf{X}_1,\, \mathbf{X}_2 \in \textbf{dom}\ f$ and $\theta \in [0,1]$, we have

$$\begin{aligned}
f(\theta\mathbf{X}_1 + (1-\theta)\mathbf{X}_2) &= \sup_{\substack{\|\mathbf{a}\|=1 \\ \mathbf{a}\in\mathbb{R}^{n\times 1}}} \|(\theta\mathbf{X}_1 + (1-\theta)\mathbf{X}_2)\mathbf{a}\|_p \\
&\le \sup_{\substack{\|\mathbf{a}\|=1 \\ \mathbf{a}\in\mathbb{R}^{n\times 1}}} \|\theta\mathbf{X}_1\mathbf{a}\|_p + \sup_{\substack{\|\mathbf{a}\|=1 \\ \mathbf{a}\in\mathbb{R}^{n\times 1}}} \|(1-\theta)\mathbf{X}_2\mathbf{a}\|_p \\
&= \theta \sup_{\substack{\|\mathbf{a}\|=1 \\ \mathbf{a}\in\mathbb{R}^{n\times 1}}} \|\mathbf{X}_1\mathbf{a}\|_p + (1-\theta) \sup_{\substack{\|\mathbf{a}\|=1 \\ \mathbf{a}\in\mathbb{R}^{n\times 1}}} \|\mathbf{X}_2\mathbf{a}\|_p \\
&= \theta f(\mathbf{X}_1) + (1-\theta) f(\mathbf{X}_2)
\end{aligned}$$

Therefore, f is convex.

2. ($\Rightarrow$) From the definition of convex function, $\textbf{dom}\ f$ is convex. Therefore, $\forall\, t_1,\, t_2 \in \textbf{dom}\ g$ and $\theta \in [0,1]$, we have

$$\begin{aligned}
g(\theta t_1 + (1-\theta)t_2) &= f(\theta(\mathbf{x}_0 + t_1\mathbf{v}) + (1-\theta)(\mathbf{x}_0 + t_2\mathbf{v})) \\
&\le \theta f(\mathbf{x}_0 + t_1\mathbf{v}) + (1-\theta) f(\mathbf{x}_0 + t_2\mathbf{v}) \\
&= \theta g(t_1) + (1-\theta) g(t_2)
\end{aligned}$$

which implies that g is convex over its domain.

($\Leftarrow$) Prove by contradiction. Suppose that f is not convex, i.e. $\exists\, \mathbf{x}_1,\ \mathbf{x}_2 \in$ **dom** f and $\theta \in [0,1]$, s.t.

$$f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) > \theta f(\mathbf{x}_1) + (1-\theta)f(\mathbf{x}_2).$$

Since **dom** f is convex, the line segment $\overline{\mathbf{x}_1\mathbf{x}_2} \subset \mathbf{dom}\ f$, which implies that

$$g(t) = f(\mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1)),\ t \in [0,1]$$

is convex. Therefore, we have

$$\begin{aligned}
\theta f(\mathbf{x}_1) + (1-\theta)f(\mathbf{x}_2) &< f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \\
&= g(1-\theta) \\
&= g(0\cdot\theta + 1\cdot(1-\theta)) \\
&\leq \theta g(0) + (1-\theta)g(1) \\
&= \theta f(\mathbf{x}_1) + (1-\theta)f(\mathbf{x}_2)
\end{aligned}$$

leading to a contradiction. Therefore, f is convex.

5. Let $g(t) = \nabla f(\mathbf{x} + t\mathbf{r})$, where $\mathbf{r} \in \mathbb{R}^n$ is an arbitrary non-zero vector. Similar to exercise 1.2, we have

$$g'(t) = \nabla^2 f(\mathbf{x} + t\mathbf{r})\mathbf{r}.$$

Applying the fundamental theorem of calculus, we get

$$\nabla f(\mathbf{x} + \alpha\mathbf{r}) - \nabla f(\mathbf{x}) = g(\alpha) - g(0) = \int_0^\alpha \nabla^2 f(\mathbf{x} + t\mathbf{r})\mathbf{r}dt,$$

where $\alpha \geq 0$. Since the ∇f is Lipschitz continuous with the constant L, we obtain

$$\|\int_0^\alpha \nabla^2 f(\mathbf{x} + t\mathbf{r})\mathbf{r}dt\|_2 = \|\nabla f(\mathbf{x} + \alpha\mathbf{r}) - \nabla f(\mathbf{x})\|_2 \leq \alpha L\|\mathbf{r}\|_2$$

by taking the 2-norm on both sides, which is equals to

$$L \geq \frac{\| \int_0^\alpha \nabla^2 f(\mathbf{x} + t\mathbf{r})\mathbf{r} dt\|_2}{\alpha\|\mathbf{r}\|_2}.$$

Let $\alpha \to 0^+$, by the L'Hospital's rule, we have

$$L \geq \frac{\|\nabla^2 f(\mathbf{x})\mathbf{r}\|_2}{\|\mathbf{r}\|_2}, \forall\, \mathbf{x},\ \mathbf{r} \in \mathbb{R}^n,\ \mathbf{r} \neq \mathbf{0}.$$

Therefore, by the definition of 2-norm we have

$$L \geq \sup_{\mathbf{r}\neq\mathbf{0}} \frac{\|\nabla^2 f(\mathbf{x})\mathbf{r}\|_2}{\|\mathbf{r}\|_2} = \|\nabla^2 f(\mathbf{x})\|_2, \forall\, \mathbf{x}.$$

Suppose that λ_i is an eigenvalue of $\nabla^2 f(\mathbf{x})$, and $\mathbf{b}_i$ is a corresponded non-zero eigenvector. Let $\mathbf{B}_i = (\mathbf{b}_i, \mathbf{b}_i, \ldots, \mathbf{b}_i)^T \in \mathbb{R}^{n\times n}$, therefore

$$\lambda_i \mathbf{B}_i = \mathbf{A}\mathbf{B}_i,\ \|\mathbf{B}_i\|_2 > 0.$$

Thus, we have

$$\|\lambda_i \mathbf{B}_i\|_2 = |\lambda_i|\|\mathbf{B}_i\|_2 = \|\mathbf{A}\mathbf{B}_i\|_2 \leq \|\mathbf{A}\|_2\|\mathbf{B}_i\|_2 \Rightarrow |\lambda_i| \leq \|\mathbf{A}\|_2, \forall\, i.$$

Therefore,

$$L \geq \|\nabla^2 f(\mathbf{x})\|_2 \geq \max_i \lambda_i = \lambda_{max}(\nabla^2 f(\mathbf{x})), \forall\, \mathbf{x} \in \mathbb{R}^n.$$

6.(a) Suppose that $\{\mathbf{x}_n\}_{i=1}^\infty \subset C_\alpha$, $\mathbf{x}_n \to \mathbf{x}$. Since f is continuous, we have

$$f(\mathbf{x}) = \lim_{n\to\infty} f(\mathbf{x}_n) \leq \lim_{n\to\infty} \alpha = \alpha,$$

which implies that $\mathbf{x} \in C_\alpha$. Therefore C_α is closed.

(b) Let $f(x) = e^x$, $x \in \mathbb{R}$. Since $\nabla^2 f = e^x > 0$, f is strictly convex. Since f is strictly decreasing when $x \to -\infty$, the problem is unsolvable.

(c) Denote $\min_{\mathbf{x}} f(\mathbf{x})$ by m.

Suppose that $\mathbf{y}_1,\ \mathbf{y}_2 \in \mathcal{C},\ \lambda \in [0,1]$. Since f is convex, **dom** f is convex. Therefore, we have

$$\lambda \mathbf{y}_1 + (1-\lambda)\mathbf{y}_2 \in \textbf{dom } f,\ \forall\, \mathbf{y}_1,\ \mathbf{y}_2 \in \textbf{dom } f,\ \lambda \in [0,1].$$

Moreover,

$$m \le f(\lambda \mathbf{y}_1 + (1-\lambda)\mathbf{y}_2) \le \lambda f(\mathbf{y}_1) + (1-\lambda) f(\mathbf{y}_2) = \lambda m + (1-\lambda) m = m$$

$$\Rightarrow \lambda \mathbf{y}_1 + (1-\lambda)\mathbf{y}_2 \in \mathcal{C},\ \forall\, \mathbf{y}_1,\ \mathbf{y}_2 \in \mathcal{C},\ \lambda \in [0,1].$$

Therefore, $\mathcal{C}$ is convex.

Suppose that $\{\mathbf{x}_n\}_{i=1}^{n} \subset \mathcal{C}$, $\mathbf{x}_n \to \mathbf{x}$. Since f is continuous, we have

$$f(\mathbf{x}) = \lim_{n\to\infty} f(\mathbf{x}_n) = m.$$

Since **dom** f is closed, $\mathbf{x} \in \textbf{dom } f$. Therefore $\mathbf{x} \in \mathcal{C}$, which implies that $\mathcal{C}$ is closed.

If **dom** f is not closed, let $\textbf{dom } f = B_{\mathbf{0}}(1)$, $f = 0$. Therefore, f is convex and continuously differentiable, but $\mathcal{C} = B_{\mathbf{0}}(1)$ is not a closed set.

(d) From exercise 1.3, we have

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

Suppose that Problem(3) has two different solutions x_1, x_2, which implies that $f(\mathbf{x}_1) = f(\mathbf{x}_2)$. Therefore, we have

$$f(\mathbf{x}_2) \ge f(\mathbf{x}_1) + \nabla f^T(\mathbf{x}_1)(\mathbf{x}_2-\mathbf{x}_1) + \frac{\mu}{2}\|\mathbf{x}_2-\mathbf{x}_1\|_2^2$$
$$\Rightarrow \nabla f^T(\mathbf{x}_1)(\mathbf{x}_1-\mathbf{x}_2) \ge \frac{\mu}{2}\|\mathbf{x}_1-\mathbf{x}_2\|_2^2.$$

Since f is strongly convex, f is convex. Thus, $\mathbf{x}_1$ is a minimum point of $f \Rightarrow \mathbf{x}_1$ is a stationary point of $f \Rightarrow \nabla f(\mathbf{x}_1) = 0$, which implies that

$$0 = \nabla f^T(\mathbf{x}_1)(\mathbf{x}_1 - \mathbf{x}_2) \geq \frac{\mu}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 \geq 0 \Rightarrow \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 = 0 \Rightarrow \mathbf{x}_1 = \mathbf{x}_2,$$

contradicting to the assumption that $\mathbf{x}_1 \neq \mathbf{x}_2$. Therefore, Problem (3) admits a unique solution.

Exercise 2. **Operations that Preserve Convexity**

Solution. 1. For all $\mathbf{x}_1$, $\mathbf{x}_2 \in \mathbb{R}^n$ and $\lambda \in [0, 1]$, we have

$$\begin{aligned}
F(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) &= f(\mathbf{A}(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + \mathbf{b}) \\
&= f(\lambda(\mathbf{A}\mathbf{x}_1 + \mathbf{b}) + (1-\lambda)(\mathbf{A}\mathbf{x}_2 + \mathbf{b})) \\
&\leq \lambda f(\mathbf{A}\mathbf{x}_1 + \mathbf{b}) + (1-\lambda) f(\mathbf{A}\mathbf{x}_2 + \mathbf{b}) \\
&= \lambda F(\mathbf{x}_1) + (1-\lambda)F(\mathbf{x}_2)
\end{aligned}$$

Therefore, F is convex.

2. Similarly, we have

$$\begin{aligned}
F(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) &= \sum_{i=1}^{m} w_i f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \\
&\leq \sum_{i=1}^{m} w_i(\lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)) \\
&= \lambda \sum_{i=1}^{m} w_i f(\mathbf{x}_1) + (1-\lambda) \sum_{i=1}^{m} w_i f(\mathbf{x}_2) \\
&= \lambda F(\mathbf{x}_1) + (1-\lambda)F(\mathbf{x}_2)
\end{aligned}$$

Therefore, F is convex.

3. Since $f_i(\theta\mathbf{x}_1+(1-\theta)\mathbf{x}_2) \le \theta f_i(\mathbf{x}_1)+(1-\theta)f_i(\mathbf{x}_2)$, $\forall\, \mathbf{x}_1,\, \mathbf{x}_2 \in \mathbb{R}^n$, $\theta \in [0,1]$, $i \in I$, we have

$$\begin{aligned} F(\theta\mathbf{x}_1+(1-\theta)\mathbf{x}_2) &= \sup_{i\in I} f_i(\theta\mathbf{x}_1+(1-\theta)\mathbf{x}_2) \\ &\le \sup_{i\in I} \theta f_i(\mathbf{x}_1) + \sup_{i\in I}(1-\theta) f_i(\mathbf{x}_2) \\ &= \theta \sup_{i\in I} f_i(\mathbf{x}_1) + (1-\theta) \sup_{i\in I} f_i(\mathbf{x}_2) \\ &= \theta F(\mathbf{x}_1) + (1-\theta) F(\mathbf{x}_2) \end{aligned}$$

Therefore, F is convex.

Exercise 3. **Subdifferentials**

Solution. 1. Let $\mathbf{x}_0 \in \mathbf{H}$, we have

$$\tilde{I}_{\mathbf{H}}(\mathbf{y}) \ge \tilde{I}_{\mathbf{H}}(\mathbf{x}_0) + \langle \mathbf{g},\, \mathbf{y}-\mathbf{x}_0 \rangle = \mathbf{g}^T(\mathbf{y}-\mathbf{x}),\ \forall\, \mathbf{y} \in \mathbb{R}^n.$$

Since $\tilde{I}_{\mathbf{H}}(\mathbf{y}) = +\infty$, $\forall\, \mathbf{y} \notin \mathbf{H}$, we can assume that $\mathbf{y} \in \mathbf{H}$. Therefore, we have

$$0 \ge \mathbf{g}^T(\mathbf{y}-\mathbf{x}_0),\ \forall\, \mathbf{y} \in \mathbf{H}.$$

which implies that

$$\partial \tilde{I}_{\mathbf{H}}(\mathbf{x}) = \{\mathbf{g} \in \mathbb{R}^n | \mathbf{g}^T\mathbf{x} \ge \mathbf{g}^T\mathbf{y},\ \forall\, \mathbf{y} \in \mathbf{H}\}$$

2. Without loss of generality, we can assume that $x_i = 0$, $i = k+1, k+2, \ldots, n$. Since

$$\partial e^{\|\mathbf{x}\|_1} = e^{\|\mathbf{x}\|_1} \partial \|\mathbf{x}\|_1,$$

from Example 3 in Lec07, we have

$$\partial f(\mathbf{x}) = \left\{ \mathbf{v} \in \mathbb{R}^n : v_i = \begin{cases} sgn(x_i) e^{\|\mathbf{x}\|_1}, & \text{if } x_i \neq 0, \\ [-e^{\|\mathbf{x}\|_1}, e^{\|\mathbf{x}\|_1}], & \text{if } x_i = 0. \end{cases} \right\}$$

3. In fact, $f(\mathbf{x}) = \max_{i\in I} \langle \mathbf{p}_i,\, \mathbf{x}\rangle$, where

$$\{\mathbf{p}_i\}_{i\in I} = \{\sum_{j=1}^{k} e_{i_j} | i_j \text{ are different elements in } \{1,2,\ldots,n\}\}.$$

Let $f_i(\mathbf{x}) = \langle \mathbf{p}_i,\, \mathbf{x}\rangle$. Since $\nabla f_i(\mathbf{x}) = \mathbf{p}_i$, by Lemma 3 we have

$$\begin{aligned} \partial f(\mathbf{x}) &= \mathbf{conv}\,\{\mathbf{p}_i : i \in I, f_i(\mathbf{x}) = \langle \mathbf{p}_i,\, \mathbf{x}\rangle = f(\mathbf{x})\} \\ &= \{\mathbf{v} : \mathbf{v} \in \mathbb{R}^n_+,\, \|\mathbf{v}\|_1 = k,\, \langle \mathbf{v},\, \mathbf{x}\rangle\ = f(\mathbf{x})\} \end{aligned}$$

4. We have $f(\mathbf{x}) = \max_i |x_i| = \max_{1\le i\le n} |\langle \mathbf{e}_i,\, \mathbf{x}\rangle|$. Let $f_i(\mathbf{x}) = |\langle \mathbf{e}_i,\, \mathbf{x}\rangle|$, then $\partial f_i(\mathbf{x}) = \mathbf{e}_i \partial |x_i|$. Therefore, we have

$$\partial f(\mathbf{x}) = \mathbf{conv}\,\{\mathbf{e}_i \partial |x_i| : |x_i| = \max_j |x_j|\}.$$

Consider two cases:

(i) $\|\mathbf{x}\|_\infty = 0$, which implies that $\mathbf{x} = \mathbf{0}$. Therefore,

$$\partial f(\mathbf{x}) = \mathbf{conv}\,\{\mathbf{e}_i[-1,1] : 1 \le i \le n\} = \{\mathbf{v} \in \mathbb{R}^n : \sum_{i=1}^{n} |v_i| \le 1\}.$$

(ii) $\|\mathbf{x}\|_\infty > 0$. We have

$$\begin{aligned} \partial f(\mathbf{x}) &= \mathbf{conv}\,\{\mathbf{e}_i sgn(x_i) : |x_i| = \max_j |x_j|\} \\ &= \left\{ \mathbf{v} \in \mathbb{R}^n : v_i = \begin{cases} sgn(x_i)\theta_i, & \text{if } |x_i| = \max_j |x_j|, \\ 0, & \text{others.} \end{cases}, \sum_{i=1}^{n} \theta_i = 1 \right\} \end{aligned}$$

5. From Example 7 in Lec06, we have

$$f(\mathbf{X}) = \max_{\|\mathbf{s}\|=1} \langle \mathbf{s}, \mathbf{X}\mathbf{s}\rangle = \max_{\|\mathbf{s}\|=1} \left\langle \mathbf{s}\mathbf{s}^T, \mathbf{X} \right\rangle.$$

Let $f_{\mathbf{s}}(\mathbf{X}) = \left\langle \mathbf{s}\mathbf{s}^T, \mathbf{X} \right\rangle$ and $\Delta = \{\mathbf{s} : \|\mathbf{s}\| = 1\}$, we have $\nabla f_{\mathbf{s}}(\mathbf{X}) = \mathbf{s}\mathbf{s}^T$. Therefore,

$$\nabla f(\mathbf{X}) = \mathbf{conv}\left\{\mathbf{s}\mathbf{s}^T : \left\langle \mathbf{s}\mathbf{s}^T, \mathbf{X} \right\rangle = f(\mathbf{X})\right\}.$$

Assume that $\lambda_{max} = \lambda_1 = \cdots = \lambda_r$, $r \leq n$, we have

$$\mathbf{u}_i \in \mathop{\mathbf{argmax}}_{\|\mathbf{s}\|=1} \left\langle \mathbf{s}\mathbf{s}^T, \mathbf{X} \right\rangle, \, i = 1, 2, \ldots, r.$$

Let $\mathbf{U} = (u_1, u_2, \ldots, u_r)$, we have

$$\begin{aligned} \mathop{\mathbf{argmax}}_{\|\mathbf{s}\|\in\Delta} \left\langle \mathbf{s}\mathbf{s}^T, \mathbf{X} \right\rangle &= \{\mathbf{v} : \mathbf{v} \in \mathbf{span}\ \mathbf{U},\ \|\mathbf{v}\| = 1\} \\ &= \{\mathbf{v} : \mathbf{v} = \mathbf{U}\mathbf{q},\ \mathbf{q} \in \mathbf{R}^r,\ \|\mathbf{q}\| = 1\} \end{aligned}$$

By Lemma 4, we have

$$\begin{aligned} \nabla f(\mathbf{X}) &= \mathbf{conv}\left\{\mathbf{v}\mathbf{v}^T : \mathbf{v} \in \mathbf{span}\ \mathbf{U},\ \|\mathbf{v}\| = 1\right\} \\ &= \mathbf{conv}\left\{\mathbf{U}\mathbf{q}\mathbf{q}^T\mathbf{U}^T : \mathbf{q} \in \mathbf{R}^r,\ \|\mathbf{q}\| = 1\right\} \\ &= \left\{\mathbf{U}\mathbf{G}\mathbf{U}^T : \mathbf{G} \succeq 0,\ tr(\mathbf{G}) = 1\right\} \end{aligned}$$