Introductory Lectures on Optimization
Homework (2)

Student Name YangYichou
Student ID 3230105697

November 22, 2025

Excercise 1. Convex functions

Please show 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 $\mathbf{dom}\ f = \{p \in \mathbb{R}^n : 0 < p_i \leq 1, \sum_{i=1}^{n} p_i = 1\}$, 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 σ_{max} denotes the largest singular value of $\mathbf{X}$.

Proof of Excercise 1: Write down your solutions step by step here.
a.Let $z = \alpha x + (1-\alpha)y, \exists |I| = k, I \subset \{1, ..., n\}, s.t.$

$$\sum_{i=1}^{k} z_{[i]} = \sum_{i\in I} (\alpha x_i + (1-\alpha) y_i)$$

$\forall |J| = k, J \in \{1, ..., n\}, \sum_{i\in J} x_i \leq \sum_{i=1}^{k} x_{[i]} = f(x), \sum_{i\in J} y_i \leq \sum_{i=1}^{k} y_{[i]} = f(y)$
Let $I = J$,we have

$$\sum_{i=1}^{k} z_{[i]} = f(z) = f(\alpha x + (1-\alpha)y) = \alpha \sum_{i\in I} x_i + (1-\alpha) \sum_{i\in I} y_i \leq \alpha f(x) + (1-\alpha) f(y)$$

Therefore,f is convex.
b.
$$\frac{\partial f(p)}{\partial p_i} = \log p_i + 1, \frac{\partial^2 f}{\partial p_i^2} = \frac{1}{p_i}, \frac{\partial^2 f}{\partial p_i \partial p_j} = 0 (i \neq j)$$

$$\therefore \nabla^2 f(p) = diag(\frac{1}{p_1}, ..., \frac{1}{p_n}) \succeq 0$$

c.According the linear property and the triangle inequality of the matrix norm,we have

$$\begin{aligned}
&\|\alpha X+(1-\alpha)Y\|_2 \leq \|\alpha X\|_2+\|(1-\alpha)Y\|_2=\alpha\|X\|_2+(1-\alpha)\|Y\|_2\\
\Rightarrow &f(\alpha X+(1-\alpha)Y) \leq \alpha f(X)+(1-\alpha)f(Y)
\end{aligned}$$

Therefore,f is convex. □

Excercise 2. Operations that Preserve Convexity

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.

Proof of Excercise 2: Write down your solutions step by step here.

$$\begin{aligned}
&epi(f)=\{(x,t)|t \geq f_i(x), \forall i \in I, x \in \cap_{i\in I} dom f_i\}=\cap_{i\in I} epi(f_i)\\
&\because epi(f_i)\ are\ all\ closed\ and\ convex\\
&\therefore epi(F)\ is\ convex\\
&\therefore F(\mathbf{x})\ is\ convex
\end{aligned}$$

□

Excercise 3. Strong Convex Functions

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

Proof of Excercise 3: Write down your solution step by step here.
Apply the definition of strongly convex functions,we have

$$f(y) \geq f(x) + < \nabla f(x), y-x > + \frac{1}{2}\mu\|y-x\|^2 \tag{1}$$

Exchange the position of x and y,we have

$$f(x) \geq f(y) + < \nabla f(y), x-y > + \frac{1}{2}\mu\|x-y\|^2 \tag{2}$$

Add (1) and (2),we can easily get

$$< \nabla f(x) - \nabla f(y), x-y > \geq \mu\|x-y\|^2$$

Let $y = x + \tau s$,we can get

$$\begin{aligned}
&< \nabla f(x+\tau s) - \nabla f(x), \tau s > \geq \mu\|\tau s\|^2\\
\Rightarrow &\frac{< \nabla f(x+\tau s) - \nabla f(x), \tau s >}{\tau^2\|s\|^2} \geq \mu\\
\Rightarrow &\frac{< \nabla f(x+\tau s) - \nabla f(x), s >}{\tau\|s\|^2} \geq \mu
\end{aligned}$$

Set $\tau \to 0$,we have

$$\lim_{\tau\to 0} \frac{\nabla f(x+\tau s) - \nabla f(x)}{\tau} = \nabla^2 f(x)s$$
$$\Rightarrow \frac{< \nabla^2 f(x)s, s >}{\|s\|^2} \geq \mu$$
$$\Rightarrow < \nabla^2 f(x)s, s > \geq \mu\|s\|^2 = < \mu Ins, s >$$
$$\Rightarrow \nabla^2 f(x) \geq \mu In, \forall x$$
$$\Rightarrow \mu \leq \lambda_{min}(\nabla^2 f(x)), \forall x$$

□

Excercise 4. Subdifferentials

Calculation of subdifferentials

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

$$\widetilde{I}_H(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in H, \\ \infty, & else \end{cases} \tag{1}$$

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

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

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

$$f(\mathbf{x}) = \sum_{i=1}^{k} x_{[i]} \tag{2}$$

Proof of Excercise 4: Write down your solution step by step here.
a.**dom** $\tilde{I_H} = H$

$$\begin{aligned} \forall x \in H, g \in \partial \tilde{I_H}(x) &\iff \tilde{I_H}(y) \geq \tilde{I_H}(x) + < g, y - x >, \forall y \in \mathbb{R}^n \\ &\iff \tilde{I_H}(y) \geq < g, y - x > \end{aligned}$$

$$y \in H \Rightarrow < g, y - x > \leq 0$$
$$y \in \mathbb{R}^n \backslash H \Rightarrow \infty \geq < g, y - x >$$
$$\therefore \partial \tilde{I_H}(x) = \{ g \in \mathbb{R}^n | < g, y - x > \leq 0, \forall y \in H \}$$

b.$f = h \circ g, g(x) = \|x\|_1, h(x) = e^x$

$$\partial f = \partial(h \circ g)(x) = h'(g(x)) \cdot \partial g(x) = e^{\|x\|_1} \cdot \partial g(x)$$
$$\partial g(x) = \partial\|x\|_1 = \partial|x_1| \times ... \times \partial|x_n|, \partial|x_i| = \begin{cases} sgn(x_i), x_i \neq 0 \\ [-1, 1], x_i = 0 \end{cases}$$
$$\therefore \partial g(x) = \{ v \in \mathbb{R}^n | v_i = sgn(x_i), x_i \neq 0; |v_i| \leq 1, x_i = - \}$$
$$\partial f(x) = e^{\|x\|_1} \cdot \partial g(x) = \left\{ w \in \mathbb{R}^n | w_i = e^{\|x\|_1} \cdot sgn(x_i), x_i \neq 0; |w_i| \leq e^{\|x\|_1}, x_i = 0 \right\}$$

c. [Lemma] For supporting function $\sigma_P(x) = sup_{y \in P} < x, y >, \partial\sigma_P(x) = argmax_{y \in P} < x, y >$

$$f(x) = \sum_{i=1}^{k} x_{[i]} = sup_{y \in P} < x, y >= \sigma_P(x), P = \left\{ y \in [0,1]^n | \mathbf{1}^T y = k \right\}$$
$$\Rightarrow \partial f(x) = argmax_{y \in P} < x, y >$$

Assume $x_{[1]} > x_{[2]} > ... > x_{[r]} = x_{[r+1]} = ... = x_{[r+m]} > x_{[r+m+1]} > ..., \exists k, r \leq k < r+m$
1.$\forall i, i < r, g_i = 1$
2.$\forall i, i > r+m, g_i = 0$
3.For $i \in \{r, r+1, ..., r+m\}, \sum_{i=r}^{r+m} g_i = k - (r-1)$

$$\therefore \partial f(x) = \left\{ g \in [0,1]^n | g_i = 1, i < r; g_i = 0, i > r+m; \sum_{i=r}^{r+m} g_i = k - (r-1), i \in \{r, ..., r+m\} \right\}$$

□

Excercise 5. Supporting Hyperplane

We know that there exists supporting hyperplanes at the boundary point of a convex set.

Please solve the following questions.

a. Express the closed convex set $\{\mathbf{x} \in R_+^2 | x_1 x_2 \geq 1\}$ as an intersection of halfspaces.

b. Let C=$\{\mathbf{x} \in R^n | \ ||\mathbf{x}_\infty|| \leq 1\}$, the ∞-norm unit ball in $\mathbb{R}^n$, and let $\hat{\mathbf{x}}$ be a point in the boundary of C. Identify the supporting hyperplanes of C at $\hat{\mathbf{x}}$ explicitly. (The ∞-norm of a point $\mathbf{x} \in \mathbb{R}^n$ is defined as $max_{1 \leq i \leq n} |x_i|$.)

Proof of Excercise 5: Write down your solution step by step here.
The tangent of $x_1 x_2 = 1$ at $(t, \frac{1}{t})$ is

$$x_2 - \frac{1}{t} = -\frac{1}{t^2}(x_1 - t) \iff tx_2 + \frac{1}{t}x_1 \geq 2, \forall t > 0$$

Let

$$H = \cap_{t>0} \left\{ x \in \mathbb{R}^2 | x_1 > 0, x_2 > 0, tx_2 + \frac{1}{t}x_1 \geq 2 \right\}$$
$$C = \{\mathbf{x} \in R_+^2 | x_1 x_2 \geq 1\}$$

We show that $H = C$:

$$\forall x \in C, tx_2 + \frac{1}{t}x_1 \geq 2\sqrt{x_1 x_2} \geq 2 \Rightarrow x \in H.$$
$$\forall x \in H, set\ t = \sqrt{\frac{x_1}{x_2}}, tx_2 + \frac{1}{t}x_1 = 2\sqrt{x_1 x_2} \geq 2 \Rightarrow x_1 x_2 \geq 1, x \in C.$$
$$\therefore C = \cap_{t>0} \left\{ x_1 > 0, x_2 > 0, tx_2 + \frac{1}{t}x_1 \geq 2 \right\}$$

b.$\hat{x}$ at the boundary of $C \Rightarrow \|\hat{x}_\infty\| = 1$
Supporting function of C:

$$h_C(a) = sup_{x \in C} a^T x = sup_{|x_i| \leq 1} \sum_{i=1}^{n} a_i x_i = \|a\|_1, x_i = sgn(a_i)$$

Then we have the supporting hyperplanes at point $\hat{x}$:

$$a^T \hat{x} = \|a\|_1, \forall a \neq 0$$

$\square$

Excercise 6. Concave Function

Consider the following loss function for logistic regression:

$$l(\theta) = \sum_{i=1}^{N} \left\{ -\log[1 + \exp(-\theta^T u^{(i)})] - [1 - v^{(i)}]\theta^T u^{(i)} \right\}.$$

where $u^{(i)} \in R^n$ denotes the input variables, $v^{(i)} \in R$ denotes the output variable, a pair $(u^{(i)}, v^{(i)})$ is called a training example, and the dataset $\{(u^{(i)}, v^{(i)})\}, i = 1, ..., N$ is called a training set. Find the Hessian $\mathbf{H}$ for this function and show that l is a concave function.

Proof of Excercise 6: Write down your solution step by step here.

$$\begin{aligned}
l_i(\theta) &= -\log(1 + \exp(-\theta^T u^{(i)})) - (1 - v^{(i)}\theta^T u^{(i)} \\
\frac{\partial l_i}{\partial \theta} &= (\frac{1}{1 + \exp(\theta^T u^{(i)})} - (1 - v^{(i)}))u^{(i)} \\
\frac{\partial^2 l_i}{\partial \theta^2} &= -\frac{1}{(1 + \exp(\theta^T u^{(i)}))^2} \cdot u^{(i)} \exp(\theta^T u^{(i)})(u^{(i)})^T \\
\nabla^2 l(\theta) &= \sum_{i=1}^{n} -\frac{\exp(\theta^T u^{(i)})}{(1 + \exp(\theta^T u^{(i)}))^2} \cdot u^{(i)}(u^{(i)})^T
\end{aligned}$$

In logistic regression(where σ is the sigmoid function):

$$p_i = \sigma(\theta^T u^{(i)}) = \frac{1}{1 + \exp(-\theta^T u^{(i)})} = \frac{\exp(\theta^T u^{(i)})}{1 + \exp(\theta^T u^{(i)})}$$

$$\frac{1}{1 + \exp(\theta^T u^{(i)})} = \sigma(-\theta^T u^{(i)}) = 1 - p_i$$

$$\therefore \nabla^2 l(\theta) = \sum_{i=1}^{n} -p_i(1 - p_i)u^{(i)}(u^{(i)})^T = -X^T W X$$

$X \in \mathbb{R}^{N \times d}, X_{i\cdot} = (u^{(i)})^T, W = diag(p_i(1 - p_i))$
$\therefore \nabla^2 l(\theta) \preceq 0$
$\therefore l$ is concave. $\square$