CSE203B HW3 Solution

Enming Guo

January 2025

1 Assignment

I.1.1.a. (3 pts) The function

$$f(x) = \sqrt{x}, \quad dom\, f = \mathbb{R}_{++}$$

is concave. To prove $f(x)$ is concave, we analyze its second derivative $f''(x)$ and verify that $f''(x) \leq 0$ for all $x \in dom\, f$.

1. Compute the first derivative of $f(x)$:

$$f'(x) = \frac{d}{dx}\left(\sqrt{x}\right) = \frac{1}{2\sqrt{x}}.$$

2. Compute the second derivative of $f(x)$:

$$f''(x) = \frac{d}{dx}\left(\frac{1}{2\sqrt{x}}\right) = -\frac{1}{4x^{3/2}}.$$

Since $x > 0$, we have $-\frac{1}{4x^{3/2}} < 0$. The second derivative satisfies:

$$f''(x) \leq 0 \quad \text{for all } x \in \mathbb{R}_{++}.$$

3. Conclusion: Since $f''(x) \leq 0$ for all $x \in dom\, f$, the function $f(x) = \sqrt{x}$ is concave.

I.1.1.b. (3 pts) The function

$$f(x) = \sum_{i=1}^{n} e^{x_i}, \quad dom\, f = \mathbb{R}^n$$

is convex. To prove $f(x)$ is convex, we use the property that $f(x)$ is convex if and only if for all $x \in dom\, f$ and all directions v, the function $g(t) = f(x + tv)$ is convex on its domain $\{t \mid x + tv \in dom\, f\}$.

1. Define $g(t)$:

$$g(t) = f(x + tv) = \sum_{i=1}^{n} e^{x_i + tv_i}.$$

2. First derivative of $g(t)$:

$$g'(t) = \frac{d}{dt}\left(\sum_{i=1}^{n} e^{x_i + tv_i}\right) = \sum_{i=1}^{n} v_i e^{x_i + tv_i}.$$

3. Second derivative of $g(t)$:

$$g''(t) = \frac{d}{dt}\left(\sum_{i=1}^{n} v_i e^{x_i + tv_i}\right) = \sum_{i=1}^{n} v_i^2 e^{x_i + tv_i}.$$

Since $e^{x_i+tv_i} > 0$ and $v_i^2 \geq 0$, it follows that $g''(t) \geq 0$ for all t. Therefore, $g(t)$ is convex.

4. Conclusion: By the property of convex functions, $f(x) = \sum_{i=1}^{n} e^{x_i}$ is convex because $g(t)$ is convex for any $x \in \mathit{dom}\, f$ and any direction v.

I.1.1.c. (3 pts) $f(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha}, \quad 0 \leq \alpha \leq 1, \quad \mathit{dom}\, f = \mathbb{R}^2_{++}$.

The function f (x1, x2) is concave. To prove this, we calculate the Hessian matrix of $f(x_1, x_2)$ and verify that it is negative semidefinite.

1. Compute the first-order partial derivatives of $f(x_1, x_2)$:

$$\frac{\partial f}{\partial x_1} = \alpha x_1^{\alpha-1} x_2^{1-\alpha}, \quad \frac{\partial f}{\partial x_2} = (1-\alpha) x_1^{\alpha} x_2^{-\alpha}.$$

2. Compute the second-order partial derivatives to form the Hessian matrix:

$$\frac{\partial^2 f}{\partial x_1^2} = \alpha(\alpha - 1) x_1^{\alpha-2} x_2^{1-\alpha}, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = \alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha},$$

$$\frac{\partial^2 f}{\partial x_2^2} = (1-\alpha)(-\alpha) x_1^{\alpha} x_2^{-\alpha-1}.$$

The Hessian matrix is:

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} \alpha(\alpha-1)x_1^{\alpha-2}x_2^{1-\alpha} & \alpha(1-\alpha)x_1^{\alpha-1}x_2^{-\alpha} \\ \alpha(1-\alpha)x_1^{\alpha-1}x_2^{-\alpha} & \alpha(\alpha-1)x_1^{\alpha}x_2^{-\alpha-1} \end{pmatrix}.$$

$$= \alpha(\alpha-1)x_1^{\alpha-2}x_2^{-1-\alpha}\begin{pmatrix} x_2^2 & -x_1x_2 \\ -x_1x_2 & x_1^2 \end{pmatrix}.$$

$$v^T \nabla^2 f(x_1, x_2) v = \alpha(\alpha-1)x_1^{\alpha-2}x_2^{-1-\alpha}(x_2 - x_1)^2$$

Since $a(a-1) \leq 0$, $v^T \nabla^2 f(x_1, x_2) v \leq 0$

3. Conclusion: The determinant of the Hessian matrix $\det(\nabla^2 f(x_1, x_2)) \leq 0$. Hence, the Hessian matrix is negative semidefinite, and $f(x_1, x_2)$ is a concave function.

I.1.1.d. (3 pts) The function

$$f(x) = \max\{a_1^T x + b_1, \ldots, a_k^T x + b_k\}, \quad dom\, f = \mathbb{R}^n$$

is convex. To prove $f(x)$ is convex, we use the definition of convexity: $f(x)$ is convex if for any $\lambda \in [0, 1]$ and any $x_1, x_2 \in \mathbb{R}^n$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

1. By definition of $f(x)$, it represents the maximum of affine functions $a_i^T x + b_i$, where $i = 1, \ldots, k$.
2. Affine functions $a_i^T x + b_i$ are convex (since they are both linear and affine), and the maximum of convex functions is also convex. Formally:

$$f(x) = \max_{i \in \{1,\ldots,k\}} (a_i^T x + b_i).$$

For any $\lambda \in [0, 1]$ and $x_1, x_2 \in \mathbb{R}^n$,

$$f(\lambda x_1 + (1-\lambda)x_2) = \max_{i \in \{1,\ldots,k\}} \{a_i^T(\lambda x_1 + (1-\lambda)x_2) + b_i\}.$$

3. Expanding $a_i^T(\lambda x_1 + (1-\lambda)x_2) + b_i$:

$$a_i^T(\lambda x_1 + (1-\lambda)x_2) + b_i = \lambda(a_i^T x_1 + b_i) + (1-\lambda)(a_i^T x_2 + b_i).$$

4. Since the maximum function satisfies convexity:

$$\max_{i \in \{1,\ldots,k\}} \{\lambda(a_i^T x_1 + b_i) + (1-\lambda)(a_i^T x_2 + b_i)\} \leq \lambda \max_{i \in \{1,\ldots,k\}} (a_i^T x_1 + b_i) + (1-\lambda) \max_{i \in \{1,\ldots,k\}} (a_i^T x_2 + b_i).$$

5. Conclusion: By the above argument, $f(x)$ is convex because the maximum of affine functions is convex.

I.1.2. (7 pts) Consider the function:

$$f(x) = \left(\sum_{i=1}^n x_i^p\right)^{1/p}, \quad dom\, f = \mathbb{R}_{++}^n, \quad p < 1, p \neq 0.$$

Answer the following questions related to this function:

(a) Derive the Hessian matrix of $f(x)$. (3 pts)

(b) Is the function $f(x)$ convex or concave? Show your proof. (4 pts)

Solution

(a) The gradient of $f(x)$ is computed as:

$$\frac{\partial f}{\partial x_i} = x_i^{p-1} \cdot \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}-1}.$$

The second-order partial derivatives are:

$$\frac{\partial^2 f}{\partial x_i^2} = (p-1)\cdot x_i^{p-2}\cdot\left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}-1} + x_i^{p-1}\cdot(1-p)\cdot\left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}-2}\cdot x_i^{p-1}$$

$$= (p-1)\cdot x_i^{p-2}\cdot\left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}-2}\cdot\left(\left(\sum_{i=1}^{n} x_i^p\right) - x_i^p\right)$$

$$= (p-1)\cdot x_i^{p-2}\cdot f^{1-2p}(x)\cdot\left[f^p(x) - x_i^p\right].$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(i \neq j) = x_i^{p-1}\cdot\left(\frac{1}{p}-1\right)\cdot\left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}-2}\cdot p\cdot x_j^{p-1}$$

$$= (1-p)\cdot x_i^{p-1}\cdot x_j^{p-1}\cdot\left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}-2}$$

$$= (1-p)\cdot x_i^{p-1}\cdot x_j^{p-1}\cdot f^{1-2p}(x).$$

The Hessian matrix of $f(x)$ can be written as:

$$\nabla^2 f(x) = (p-1)\cdot x_i^{p-1}\cdot f^{1-2p}(x)\cdot A,$$

where A is a symmetric matrix defined as:

$$A = \begin{bmatrix} x_1^{p-2}\cdot(x_2^p+\cdots+x_n^p) & -x_2^{p-1}x_1^{p-1} & \cdots & -x_1^{p-1}x_n^{p-1} \\ -x_2^{p-1}x_1^{p-1} & x_2^{p-2}\cdot(x_1^p+x_3^p+\cdots+x_n^p) & \cdots & -x_2^{p-1}x_n^{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ -x_n^{p-1}x_1^{p-1} & -x_n^{p-1}x_2^{p-1} & \cdots & x_n^{p-2}\cdot(x_1^p+x_2^p+\cdots+x_{n-1}^p) \end{bmatrix}.$$

(b) To determine the concavity or convexity of $f(x)$, note the following:

The corresponding Hessian matrix is given by:

$$\nabla^2 f(x) = (p-1)\cdot x_i^{p-1}\cdot f^{1-2p}(x)\cdot A,$$

where:

$$A = \begin{bmatrix} x_1^{p-2}\cdot(x_2^p+\cdots+x_n^p) & -x_2^{p-1}x_1^{p-1} & \cdots & -x_n^{p-1}x_1^{p-1} \\ -x_2^{p-1}x_1^{p-1} & x_2^{p-2}\cdot(x_1^p+x_3^p+\cdots+x_n^p) & \cdots & -x_n^{p-1}x_2^{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ -x_n^{p-1}x_1^{p-1} & -x_n^{p-1}x_2^{p-1} & \cdots & x_n^{p-2}\cdot(x_1^p+x_2^p+\cdots+x_{n-1}^p) \end{bmatrix}.$$

$$v^TAv = A_{11}v_1^2+A_{12}v_1v_2+...+A_{1n}v_1v_n+A_{21}v_2v_1+A_{22}v_2^2+...+A_{2n}v_2v_n+...+A_{n1}v_nv_1+...+A_{nn}v_n^2$$

$$= \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left(x_i^{p-2} x_j^{p-2} (x_i - x_j)^2 \right) \geq 0$$

, which confirms that A is a positive semidefinite matrix.

Since $p < 1$, the prefactor

$$(p-1) \cdot x_i^{p-1} \cdot f^{1-2p}(x),$$

in $\nabla^2 f(x)$ is negative, and therefore $\nabla^2 f(x)$ is negative semidefinite. Thus, $f(x)$ is a **concave function** on its domain $\mathit{dom}\, f = \mathbb{R}^n_{++}$.

II.2.1 The conjugate function of $f(x) = 2x + 1$ is defined as:

$$f^*(y) = \sup_{x \in \mathbb{R}} (yx - (2x+1)) = \sup_{x \in \mathbb{R}} ((y-2)x - 1)\,.$$

To solve this, we classify into two cases:

- **Case 1:** $y = 2$

  When $y = 2$, the expression becomes:

  $$f^*(2) = \sup_{x \in \mathbb{R}} ((2-2)x - 1) = \sup_{x \in \mathbb{R}} (-1) = -1.$$

  Therefore, $f^*(2) = -1$.

- **Case 2:** $y \neq 2$

  When $y \neq 2$, the coefficient of x in $(y-2)x-1$ is $(y-2)$, which determines whether the supremum exists:

  - If $y > 2$, then $(y-2)x \to \infty$ as $x \to \infty$, so $f^*(y) = \infty$.
  - If $y < 2$, then $(y-2)x \to \infty$ as $x \to -\infty$, so $f^*(y) = \infty$.

  Therefore, $f^*(y) = \infty$ for all $y \neq 2$.

In conclusion:

$$f^*(y) = \begin{cases} -1, & \text{if } y = 2, \\ \infty, & \text{if } y \neq 2. \end{cases}$$

II.2.2. Let $f_2(x) = \frac{1}{3} x^\top Q x$, where $Q \in S^n_{++}$ (a symmetric positive definite matrix) and $x \in \mathbb{R}^n$.

The conjugate function of $f_2(x)$ is defined as:

$$f_2^*(y) = \sup_{x \in \mathbb{R}^n} \left(y^\top x - \frac{1}{3} x^\top Q x \right).$$

Solution:

Rewrite the expression to simplify:

$$y^\top x - \frac{1}{3} x^\top Q x = x^\top y - \frac{1}{3} x^\top Q x.$$

Combine terms into a quadratic form:

$$x^\top \left(y - \frac{1}{3} Qx \right).$$

The supremum is achieved by maximizing over x. Since $Q \in S_{++}^n$, the maximization can be solved by setting the gradient to 0:

$$\frac{\partial}{\partial x} \left(y^\top x - \frac{1}{3} x^\top Qx \right) = y - \frac{2}{3} Qx = 0.$$

Solve for x:

$$x = \frac{3}{2} Q^{-1} y.$$

Substitute $x = \frac{3}{2} Q^{-1} y$ back into the original function:

$$f_2^*(y) = y^\top \left(\frac{3}{2} Q^{-1} y \right) - \frac{1}{3} \left(\frac{3}{2} Q^{-1} y \right)^\top Q \left(\frac{3}{2} Q^{-1} y \right).$$

Simplify:

$$f_2^*(y) = \frac{3}{2} y^\top Q^{-1} y - \frac{1}{3} \left(\frac{9}{4} y^\top Q^{-1} Q Q^{-1} y \right).$$

Notice that $Q^{-1} Q = I$:

$$f_2^*(y) = \frac{3}{2} y^\top Q^{-1} y - \frac{1}{3} \left(\frac{9}{4} y^\top Q^{-1} y \right).$$

Simplify further:

$$f_2^*(y) = \frac{3}{2} y^\top Q^{-1} y - \frac{3}{4} y^\top Q^{-1} y.$$

Combine terms:

$$f_2^*(y) = \frac{3}{2} y^\top Q^{-1} y - \frac{3}{4} y^\top Q^{-1} y = \frac{3}{4} y^\top Q^{-1} y.$$

Final Result:

$$f_2^*(y) = \frac{3}{4} y^\top Q^{-1} y.$$

II.2.3. The conjugate function $f_3^*(y)$ is defined as:

$$f_3^*(y) = \sup_{x \in \mathbb{R}} \{ yx - f_3(x) \} = \sup_{x \in \mathbb{R}} \{ yx + \log(ax^2 + bx + c) \}.$$

Step 1: Analyze $f_3^*(y)$ for $y \geq 0$: If $y \geq 0$, the term $yx + \log(ax^2 + bx + c)$ is unbounded above as $x \to \infty$. Therefore:

$$f_3^*(y) = +\infty, \quad \text{for } y \geq 0.$$

Step 2: Analyze $f_3^*(y)$ for $y < 0$

If $y < 0$, the optimization problem reduces to finding the maximum of the function:

$$\Phi(x) = yx + \log(ax^2 + bx + c),$$

where x must satisfy $ax^2 + bx + c > 0$.

However, since $y < 0$, as $x \to -\infty$, we observe:

$$\Phi(x) \to +\infty.$$

Thus, there is no need to solve for the roots of the quadratic equation. We directly conclude:

$$f_3^*(y) = +\infty.$$

In other words, when $y < 0$, the function $\Phi(x)$ is unbounded above, meaning its supremum is $+\infty$.

II.2.4. The conjugate function $f_4^*(y)$ is defined as:

$$f_4^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^\top x - f_4(x) \right\},$$

where $f_4(x) = \sum_{i=1}^n e^{a_i x_i + b_i}$. Thus, the optimization problem becomes:

$$f_4^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^\top x - \sum_{i=1}^n e^{a_i x_i + b_i} \right\}.$$

Step 1: Analyze the Supremum We rewrite the function to maximize:

$$\Phi(x) = \sum_{i=1}^n \left(y_i x_i - e^{a_i x_i + b_i} \right).$$

The objective separates over i, so we can maximize each term independently. For each i, define:

$$\phi_i(x_i) = y_i x_i - e^{a_i x_i + b_i}.$$

Step 2: Compute the Critical Points To find the maximum of $\phi_i(x_i)$, take the derivative with respect to x_i:

$$\phi_i'(x_i) = y_i - a_i e^{a_i x_i + b_i}.$$

Set $\phi_i'(x_i) = 0$:

$$y_i = a_i e^{a_i x_i + b_i}.$$

Solve for x_i:

$$e^{a_i x_i + b_i} = \frac{y_i}{a_i}, \quad \text{which gives} \quad x_i = \frac{1}{a_i} \log\left(\frac{y_i}{a_i}\right) - \frac{b_i}{a_i}.$$

Step 3: Feasibility Condition The solution for x_i exists only if $y_i > 0$ (since $\log(\cdot)$ requires a positive argument). If $y_i = 0$, the supremum is 0 when x grows to negative infinity. If $y_i < 0$, the supremum becomes unbounded, so $f_4^*(y) = +\infty$ in those cases.

Step 4: Compute $f_4^*(y)$ When $y_i > 0$ for All i Substitute the critical point $x_i = \frac{1}{a_i} \log\left(\frac{y_i}{a_i}\right) - \frac{b_i}{a_i}$ into $\phi_i(x_i)$:

$$\phi_i(x_i) = y_i \left(\frac{1}{a_i} \log\left(\frac{y_i}{a_i}\right) - \frac{b_i}{a_i} \right) - \frac{y_i}{a_i}.$$

Simplify:

$$\phi_i(x_i) = \frac{y_i}{a_i} \log\left(\frac{y_i}{a_i}\right) - \frac{y_i b_i}{a_i} - \frac{y_i}{a_i}.$$

Thus, the conjugate function is:

$$f_4^*(y) = \begin{cases} \sum_{i=1}^n \left(\frac{y_i}{a_i} \log\left(\frac{y_i}{a_i}\right) - \frac{y_i b_i}{a_i} - \frac{y_i}{a_i} \right), & \text{if } y_i > 0 \text{ for all } i, \\ 0, & \text{if } y_i = 0 \\ +\infty, & \text{if } y_i < 0 \text{ for any } i. \end{cases}$$

II.2.5. The conjugate function $f_5^*(y)$ is defined as:

$$f_5^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^\top x - f_5(x) \right\},$$

where $f_5(x) = \log \sum_{i=1}^n \exp\left(\frac{x_i}{\gamma}\right)$. The optimization becomes:

$$f_5^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^\top x - \log \sum_{i=1}^n \exp\left(\frac{x_i}{\gamma}\right) \right\}.$$

Step 1: Reformulate the Problem Let:

$$\Phi(x) = y^\top x - \log \sum_{i=1}^n \exp\left(\frac{x_i}{\gamma}\right).$$

The first term is linear, while the second term involves a log-sum-exp function.

Step 2: Critical Point Condition To maximize $\Phi(x)$, compute its gradient:

$$\frac{\partial \Phi}{\partial x_i} = y_i - \frac{1}{\gamma} \frac{\exp\left(\frac{x_i}{\gamma}\right)}{\sum_{j=1}^n \exp\left(\frac{x_j}{\gamma}\right)}.$$

Setting $\frac{\partial \Phi}{\partial x_i} = 0$ gives:

$$y_i = \frac{1}{\gamma} \frac{\exp\left(\frac{x_i}{\gamma}\right)}{\sum_{j=1}^n \exp\left(\frac{x_j}{\gamma}\right)}.$$

Let:

$$p_i = \frac{\exp\left(\frac{x_i}{\gamma}\right)}{\sum_{j=1}^n \exp\left(\frac{x_j}{\gamma}\right)}.$$

Thus:

$$p_i = \gamma y_i, \quad \text{and} \quad \sum_{i=1}^n p_i = 1 \implies \sum_{i=1}^n \gamma y_i = 1.$$

Step 3: Feasibility of y The solution is valid if $y_i \geq 0$ and $\sum_{i=1}^n \gamma y_i = 1$. Otherwise, $f_5^*(y) = +\infty$.

Step 4: Compute $f_5^*(y)$ Substitute $p_i = \gamma y_i$ into the expression for x. Using the relationship:

$$x_i = \gamma \log\left(\frac{p_i}{\gamma}\right),$$

compute $y^\top x$:

$$y^\top x = \sum_{i=1}^n y_i \cdot \gamma \log\left(\frac{\gamma y_i}{\gamma}\right) = \sum_{i=1}^n \gamma y_i \log(y_i).$$

The conjugate function becomes:

$$f_5^*(y) = \sum_{i=1}^n \gamma y_i \log(y_i), \quad \text{if } \gamma \sum_{i=1}^n y_i = 1 \text{ and } y_i \geq 0.$$

Otherwise:

$$f_5^*(y) = +\infty.$$