Convex Optimization - AI2101
Assignment - IV

Bhuvan Chandra K
AI23BTECH11013

April 1, 2025

Question 3.16

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

Solution

(a) $f(x) = e^x - 1$ on $\mathbb{R}$ Calculating $f'(x)$:

$$f'(x) = e^x$$

Since $f'(x) > 0$ ($e^x > 0$) $\forall x \in \mathbb{R}$, $f(x)$ is strictly increasing on its domain ($x \in \mathbb{R}$).
To prove convexity, we calculate the second derivative:

$$f''(x) = e^x$$

Since $f''(x) > 0$ $\forall x \in \mathbb{R}$ (as $e^x > 0$ for all real values), the function is strictly convex on $\mathbb{R}$.
So, it is also quasi-convex.
Note: SInce this function is strictly increasing, its superlevel sets are convex . Hence the function is also quasi-convex.

$$\therefore \text{The function is convex,quasi-convex,quasi-concave only}$$

(b) $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2$
To determine the type of function for a multivariate function, we use the Hessian matrix:

$$\frac{\partial f}{\partial x_1} = x_2, \quad \frac{\partial f}{\partial x_2} = x_1$$

Second-order derivatives:

$$\frac{\partial^2 f}{\partial x_1^2} = 0, \quad \frac{\partial^2 f}{\partial x_2^2} = 0, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 1$$

Hessian matrix:

$$H = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

To determine if the Hessian is definite or indefinite, we calculate its eigenvalues.
The characteristic equation is:

$$\det(H - \lambda I) = \begin{vmatrix} -\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - 1 = 0$$

Solving for eigenvalues: $\lambda = +1, -1$.
Since the Hessian has both positive and negative eigenvalues, it is indefinite.
Therefore, $f(x_1, x_2)$ is neither concave nor convex.
The superlevel sets ($\{(x_1, x_2)|x_1x_2 \geq c\}$) form hyperbolic regions in $\mathbb{R}^2$, which are convex sets. Because the region is bounded by a hyperbola and if i take any two points from this region and draw a line , the line will lie in this region.
Hence, the function is quasi-concave.
The sublevel sets ($\{(x_1, x_2)|x_1x_2 \leq \alpha\}$) are hyperbolas in the positive quadrant (but not convex),Because if we take two points from opposite segemnts and draw a line between them it wont be entirely part of one of the regions.
So the function is not quasi-convex.

$$\therefore \text{The function is quasi concave only}$$

(c) $f(x_1, x_2) = 1/(x_1x_2)$ on $\mathbb{R}^{++}$
First-order partial derivatives:

$$\frac{\partial f}{\partial x_1} = -\frac{1}{x_1^2x_2}, \quad \frac{\partial f}{\partial x_2} = -\frac{1}{x_1x_2^2}$$

Second-order derivatives:

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{2}{x_1^3x_2}, \quad \frac{\partial^2 f}{\partial x_2^2} = \frac{2}{x_1x_2^3}, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{1}{x_1^2x_2^2}$$

Hessian matrix:

$$H = \begin{bmatrix} \frac{2}{x_1^3x_2} & \frac{1}{x_1^2x_2^2} \\ \frac{1}{x_1^2x_2^2} & \frac{2}{x_1x_2^3} \end{bmatrix}$$

To look at the nature of the Hessian matrix we find the trace and determinant

$$\text{Trace}(H) = \frac{2}{x_1^3x_2} + \frac{2}{x_1x_2^3}$$

Since we're in the domain $\mathbb{R}^2_{++}$ where $x_1 > 0$ and $x_2 > 0$, each term in the trace is positive.
Therefore, $\text{Trace}(H) > 0$.

$$\begin{aligned} \det(H) &= \frac{2}{x_1^3x_2} \cdot \frac{2}{x_1x_2^3} - \left(\frac{1}{x_1^2x_2^2}\right)^2 \\ &= \frac{4}{x_1^4x_2^4} - \frac{1}{x_1^4x_2^4} \\ &= \frac{3}{x_1^4x_2^4} \end{aligned}$$

Since $x_1 > 0$ and $x_2 > 0$ in our domain, we have:

$$\frac{3}{x_1^4x_2^4} > 0$$

So, $\det(H) > 0$.
SInce $\text{Trace}(H) > 0$ and $\det(H) > 0$ for all points in the domain $\mathbb{R}^2_{++}$, the Hessian matrix is positive definite everywhere in this domain.
So we can say that the function $f(x_1, x_2) = \frac{1}{x_1x_2}$ is strictly convex on $\mathbb{R}^2_{++}$.
Since the function is strictly convex, in turn it is quasiconvex.

Since Hessian is positive definite it cannot eb concave and the function is strictly decreasing for both variables so its superlevel sets wont be convex , so it is not quasi concave.

Therefore, the function is convex and quasiconvex only.

(d) $f(x_1, x_2) = x_1/x_2$ on $\mathbb{R}^2_+$
First-order partial derivatives:

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

Second-order derivatives:

$$\frac{\partial^2 f}{\partial x_1^2} = 0, \quad \frac{\partial^2 f}{\partial x_2^2} = \frac{2x_1}{x_2^3}, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = -\frac{1}{x_2^2}$$

The Hessian matrix:

$$H = \begin{bmatrix} 0 & -\frac{1}{x_2^2} \\ -\frac{1}{x_2^2} & \frac{2x_1}{x_2^3} \end{bmatrix}$$

To look at the nature of Haessian matrix we find the trace and determinant:

$$\text{Trace}(H) = 0 + \frac{2x_1}{x_2^3} = \frac{2x_1}{x_2^3}$$

$$\det(H) = 0 \cdot \frac{2x_1}{x_2^3} - \left(-\frac{1}{x_2^2}\right)^2 = -\frac{1}{x_2^4} < 0$$

Since the determinant is negative for all $x_2 > 0$, the Hessian is indefinite.
Thus, H is neither positive nor negative semi-definite, and the function $f(x_1, x_2)$ is neither convex nor concave.
Sublevel sets:

$$S = \{(x_1, x_2) | x_1/x_2 \leq \alpha\}$$

These represent halfspaces ($x_1 \leq \alpha x_2$), which are convex. Thus, $f(x_1, x_2)$ is quasi-convex.
Superlevel sets:

$$S = \{(x_1, x_2) | x_1/x_2 \geq \alpha\}$$

These represent halfspaces ($x_1 \geq \alpha x_2$), which are convex. Thus, $f(x_1, x_2)$ is quasi-concave.
Therefore, this function is both quasi-convex and quasi-concave.

$$\therefore \text{The function is quasi convex and quasi concave only}$$

(e) $f(x_1, x_2) = x_1^2/x_2$ on $\mathbb{R}_+ \times \mathbb{R}_+$ First-order partial derivatives:

$$\frac{\partial f}{\partial x_1} = \frac{2x_1}{x_2}, \quad \frac{\partial f}{\partial x_2} = -\frac{x_1^2}{x_2^2}$$

Second-order derivatives:

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{2}{x_2}, \quad \frac{\partial^2 f}{\partial x_2^2} = \frac{2x_1^2}{x_2^3}, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = -\frac{2x_1}{x_2^2}$$

The Hessian matrix:

$$H = \begin{bmatrix} \frac{2}{x_2} & -\frac{2x_1}{x_2^2} \\ -\frac{2x_1}{x_2^2} & \frac{2x_1^2}{x_2^3} \end{bmatrix}$$

To look at the nature of Hessian matrix we find the trace and determinant

$$\text{Trace}(H) = \frac{2}{x_2} + \frac{2x_1^2}{x_2^3} = \frac{2(x_2^2 + x_1^2)}{x_2^3} > 0 \quad \text{for } x_2 > 0$$

$$\det(H) = \frac{2}{x_2} \cdot \frac{2x_1^2}{x_2^3} - \left(-\frac{2x_1}{x_2^2}\right)^2 = \frac{4x_1^2 - 4x_1^2}{x_2^4} = 0$$

Since the trace is positive and the determinant is zero, the Hessian is positive semi-definite (not strictly positive definite).
Therefore, the function is convex, in turn it is also quasi convex.
The function isn't monotonic in x_1, so it is not concave.
The superlevel sets ($S = \{(x_1, x_2) \mid x_1^2/x_2 > a\}$) are not convex.
Hence, the function is not quasi-concave.

$$\therefore \text{The function is convex and in turn quasi-convex only.}$$

(f) $f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$, where $0 \leq \alpha \leq 1$ on $\mathbb{R}^{++}$
First-order partial derivatives:

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

Second-order partial derivatives:

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

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

The Hessian matrix:

$$H = \begin{bmatrix} \alpha(\alpha-1) x_1^{\alpha-2} x_2^{1-\alpha} & \alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha} \\ \alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha} & (1-\alpha)(-\alpha) x_1^\alpha x_2^{-\alpha-1} \end{bmatrix}$$

To look at the nature of Hessian matrix we find the trace and determinant

$$\text{Trace}(H) = \alpha(\alpha-1) x_1^{\alpha-2} x_2^{1-\alpha} + (1-\alpha)(-\alpha) x_1^\alpha x_2^{-\alpha-1}$$

Since $0 \leq \alpha \leq 1$, we have $\alpha - 1 \leq 0$, so $\alpha(\alpha-1) \leq 0$ and $(1-\alpha)(-\alpha) \leq 0$. Thus, $\text{Tr}(H) \leq 0$ for all $x_1, x_2 > 0$.

$$\begin{aligned} \det(H) &= \alpha(\alpha-1) x_1^{\alpha-2} x_2^{1-\alpha} \cdot (1-\alpha)(-\alpha) x_1^\alpha x_2^{-\alpha-1} - [\alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha}]^2 \\ &= \alpha^2 (1-\alpha)^2 x_1^{2\alpha-2} x_2^{-2\alpha} \cdot \left(\frac{\alpha-1}{x_1^2} \cdot \frac{-\alpha}{x_2} - 1\right) \\ &= \alpha^2 (1-\alpha)^2 x_1^{2\alpha-2} x_2^{-2\alpha} \cdot \left(\frac{-\alpha(\alpha-1)}{x_1^2 x_2} - 1\right) \end{aligned}$$

Since $0 \leq \alpha \leq 1$, both α and $1 - \alpha$ are non-negative, and $(\alpha - 1) \leq 0$. Thus, $-\alpha(\alpha-1) \geq 0$, and the term inside brackets $\left(\frac{-\alpha(\alpha-1)}{x_1^2 x_2} - 1\right) < 0$, which implies that $\det(H) \geq 0$ for all $x_1, x_2 > 0$.
With $\text{Tr}(H) \leq 0$ and $\det(H) \geq 0$, the Hessian is negative semi-definite for $0 \leq \alpha \leq 1$ on $\mathbb{R}^{++}$.
Since $H \leq 0$, the function is concave and hence quasi-concave. However, the function is not convex, and the sublevel sets are not convex. Thus, the function is not quasi-convex.

$$\therefore \text{The function is concave and hence quasi-concave only.}$$

Question 3.18

Adapt the proof of concavity of the log-determinant function in §3.1.5 to show the following.
(a) $f(X) = \text{tr}(X^{-1})$ is convex on $\text{dom} f = S_n^{++}$.
(b) $f(X) = (\det X)^{1/n}$ is concave on $\text{dom} f = S_n^{++}$.

Solution

(a) We need to show that $f(X) = \operatorname{tr}(X^{-1})$ is convex on the domain S_+,
i.e., for any two symmetric positive definite matrices X_1, X_2 and any $\theta \in [0, 1]$,

$$f(\theta X_1 + (1-\theta)X_2) \geq \theta f(X_1) + (1-\theta) f(X_2).$$

Let P be a positive definite matrix (symmetric), and Y be a symmetric matrix.
Define:

$$g(t) = \operatorname{tr}((P + tY)^{-1}).$$

If we show $g(t)$ is convex in t, then $f(X)$ is also convex.
From the matrix inversion identity:

$$(P + tY)^{-1} = P^{-1}(I + tP^{-1/2}YP^{-1/2})^{-1}.$$

So,

$$g(t) = \operatorname{tr}(P^{-1}(I + tP^{-1/2}YP^{-1/2})^{-1}).$$

Let:

$$P^{-\frac{1}{2}}YP^{-\frac{1}{2}} = Q\Lambda Q^T,$$

where Q is an orthogonal matrix ($QQ^T = I$) and Λ is a diagonal matrix containing the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.
Thus

$$g(t) = \operatorname{tr}(P^{-1}Q(I + t\Lambda)^{-1}Q^T).$$

Since the trace function is invariant under cyclic permutations:

$$g(t) = \sum_{i=1}^{n} \frac{1}{(I + t\lambda_i)}.$$

We can see that the function has the form:

$$h(\lambda_i) = (Q^T P^{-1} Q)_{ii}(1 + t\lambda_i)^{-1}.$$

We know that $h(\lambda_i)$ is convex for $t > 0$. So, $g(t)$, and hence $f(X)$, is a convex function.
Since each function is convex, the sum of convex functions is also convex. We conclude that $g(t)$ is convex in t.
Thus, $f(X) = \operatorname{tr}(X^{-1})$ is convex on S_+.

(b) To show: $f(X) = (\det X)^{1/n}$ is concave on S_{++}, the set of positive definite matrices.
Let:

$$g(t) = \det(P + tY)^{1/n}, \quad Z > 0,\ Z \text{ positive definite, and } Y \succeq 0.$$

If $g(t)$ is concave in t, then $f(X) = (\det X)^{1/n}$ is concave. Define:

$$P + tY = P^{1/2}\Big(I + tP^{-1/2}YP^{-1/2}\Big)P^{1/2}.$$

Using the determinant property:

$$\det AB = \det A \cdot \det B,$$

we have:

$$g(t) = \Big(\det(P)^{1/2}\det(I + tP^{-1/2}YP^{-1/2})\det(P)^{1/2}\Big)^{1/n}.$$

Thus:

$$g(t) = (\det(P))^{1/n}(\det(I + tP^{-1/2}YP^{-1/2}))^{1/n}.$$

We know that the determinant of a symmetric matrix is the product of its eigenvalues.
Let the eigenvalues of $P^{-1/2}YP^{-1/2}$ be $\lambda_1, \lambda_2, \dots, \lambda_n$.
Then:

$$\det(I + tP^{-1/2}YP^{-1/2}) = \prod_{i=1}^{n}(1 + t\lambda_i).$$

So:

$$g(t) = (\det(P))^{1/n} \prod_{i=1}^{n}(1 + t\lambda_i)^{1/n}.$$

The function:

$$\prod_{i=1}^{n}(1 + t\lambda_i)^{1/n}$$

is said to be concave due to the geometric-arithmetic mean inequality (GAM inequality).
Since $\prod_{i=1}^{n}(1 + t\lambda_i)^{1/n}$ is concave for $t \geq 0$, and each $(1 + t\lambda_i)$ is positive for small t, we conclude that $g(t)$ is concave.
Thus, $f(X) = (\det X)^{1/n}$ is concave on S_{++}.

Question 3.19

(a) Show that $f(x) = \sum_{i=1}^{r} \alpha_i x_{[i]}$ is a convex function of x, where $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r \geq 0$, and $x_{[i]}$ denotes the ith largest component of x. (You can use the fact that $f(x) = \sum_{i=1}^{k} x_{[i]}$ is convex on $\mathbb{R}^n$.)

(b) Let $T(x, \omega)$ denote the trigonometric polynomial

$$T(x, \omega) = x_1 + x_2 \cos \omega + x_3 \cos 2\omega + \cdots + x_n \cos(n-1)\omega.$$

Show that the function

$$f(x) = -\int_0^{2\pi} \log T(x, \omega)\, d\omega$$

is convex on $\{x \in \mathbb{R}^n \mid T(x, \omega) > 0, 0 \leq \omega \leq 2\pi\}$.

Solution

(a) Let:

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

where $d_1 \geq \alpha_2 \geq \cdots \geq \alpha_k \geq 0$, and $x_{[i]}$ is the i-th largest component of x. We know that:

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

is convex on $\mathbb{R}^n$. So, $f(x)$ as:

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

is convex for all $i = 1, 2, \dots, k$.
From the question:

$$f(x) = \alpha_1 x_{[1]} + \alpha_2 x_{[2]} + \cdots + \alpha_k x_{[k]},$$

where:

$$\alpha_1 x_{[1]} + \alpha_2 x_{[2]} = (\alpha_1 - \alpha_2)x_{[1]} + \alpha_2(x_{[1]} + x_{[2]}),$$

Expanding further:

$$f(x) = (\alpha_1 - \alpha_2)x_{[1]} + (\alpha_2 - \alpha_3)(x_{[1]} + x_{[2]}) + \cdots + (\alpha_k - \alpha_r)(x_{[1]} + x_{[2]} + \cdots + x_{[r-1]}) + \alpha_r(x_{[1]} + \cdots + x_{[r]}).$$

This is a linear combination of convex functions (coefficients are non-negative).
Thus, $x_{[1]} + \cdots + x_{[r]}$ is convex for $i = 1 \ldots r$. Therefore, $f(x)$ is a convex function of x.

(b) Let:

$$T(x, \omega) = x_1 + x_2 \cos \omega + \cdots + x_n \cos(n-1)\omega.$$

We need to show:

$$f(x) = -\int_0^{2\pi} \log(T(x, \omega))\, d\omega$$

is convex on $\{x \in \mathbb{R}^n : T(x, \omega) > 0,\ 0 \leq \omega \leq 2\pi\}$. Let:

$$g(x, \omega) = -\log(T(x, \omega)).$$

Since $-\log z$ is convex for $z > 0$, and:

$$T(x, \omega) = x_1 + x_2 \cos \omega + \cdots + x_n \cos(n-1)\omega$$

is linear in $x = (x_1, x_2, \ldots, x_n)$, it follows that $g(x, \omega) = -\log(T(x, \omega))$ is convex in x.
From the result that a function $g(x, t)$ is convex in x for every fixed t, then its integral over a fixed range (independent of x) is also convex in x.
Thus, the integral:

$$f(x) = -\int_0^{2\pi} g(x, \omega) d\omega = -\int_0^{2\pi} \log(T(x, \omega)) d\omega$$

is convex. Hence:

$$f(x) = -\int_0^{2\pi} \log(T(x, \omega)) d\omega$$

is convex.

Question 3.21

Show that the following functions $f : \mathbb{R}^n \to \mathbb{R}$ are convex.

(a) $f(x) = \max_{i=1,\ldots,k} \|A^{(i)}x - b^{(i)}\|$, where $A^{(i)} \in \mathbb{R}^{m \times n}$, $b^{(i)} \in \mathbb{R}^m$, and $\|\cdot\|$ is a norm on $\mathbb{R}^m$.

(b) $f(x) = \sum_{i=1}^r |x|_{[i]}$ on $\mathbb{R}^n$, where $|x|$ denotes the vector with $|x|_i = |x_i|$ (i.e., $|x|$ is the absolute value of x, componentwise), and $|x|_{[i]}$ is the ith largest component of $|x|$. In other words, $|x|_{[1]}, |x|_{[2]}, \ldots, |x|_{[n]}$ are the absolute values of the components of x, sorted in nonincreasing order.

Solution

(a) We need to show:

$$f(x) = \max_{i=1,\ldots,k} \|A^{(i)}x - b^{(i)}\|$$

is convex, where $A^{(i)} \in \mathbb{R}^{m \times n}, b^{(i)} \in \mathbb{R}^m$. For any $x_1, x_2 \in \mathbb{R}^n$, and $\lambda \in [0, 1]$,

$$f_i(\lambda x_1 + (1-\lambda)x_2) = \|A^{(i)}(\lambda x_1 + (1-\lambda)x_2) - b^{(i)}\|.$$

Expanding:

$$f_i(\lambda x_1 + (1-\lambda)x_2) = \|A^{(i)}(\lambda x_1 + (1-\lambda)x_2) - b^{(i)}\| = \|\lambda A^{(i)}x_1 + (1-\lambda)A^{(i)}x_2 - b^{(i)}\|.$$

From the convexity property of the norm:

$$\|\lambda u + (1-\lambda)v\| \leq \lambda\|u\| + (1-\lambda)\|v\|.$$

So:

$$\|A^{(i)}(\lambda x_1 + (1-\lambda)x_2) - b^{(i)}\| \leq \lambda\|A^{(i)}x_1 - b^{(i)}\| + (1-\lambda)\|A^{(i)}x_2 - b^{(i)}\|.$$

Hence:

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

So $f_i(x) = \|A^{(i)}x - b^{(i)}\|$ is convex for each $i = 1, 2, \ldots, k$.
Let $h_1(x), h_2(x), \ldots, h_k(x)$ be convex functions. Define:

$$g(x) = \max(g_1(x), g_2(x), \ldots, g_k(x)).$$

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

$$g(x_1) = g_p(x_1), \quad g(x_2) = g_q(x_2),$$

where p, q are the indices where the maximum is achieved for x_1, x_2, respectively.
Let r be the index where the maximum is achieved for:

$$g_r(\lambda x_1 + (1-\lambda)x_2).$$

Then:

$$g_r(\lambda x_1 + (1-\lambda)x_2) \geq g_p(\lambda x_1 + (1-\lambda)x_2).$$

By convexity:

$$g_r(\lambda x_1 + (1-\lambda)x_2) = g_r(\lambda x_1 + (1-\lambda)x_2) = g_p(\lambda x_1 + (1-\lambda)x_2).$$

Thus: $f(x)$ is convex.

$$g(\lambda x_1 + (1-\lambda)x_2) = g_r(\lambda x_1 + (1-\lambda)x_2).$$

Since $g_r(x)$ is convex:

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

Since p, q are maximum indices for x_1, x_2,

$$g_r(x_1) \leq g_p(x_1), \quad g_r(x_2) \leq g_q(x_2).$$

Thus:

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

So $g_r(x)$ is convex. From these two results, we can say:

$$f(x) = \max_{i=1,\ldots,k} \|A^{(i)}x - b^{(i)}\|$$

is a convex function.

(b) We need to show:

$$f(x) = \sum_{j=1}^{r} |x|_{[j]} \quad \text{on } \mathbb{R}^n,$$

where $x_{[j]}$ is the j-th largest component of x.
For any $x \in \mathbb{R}^n$, $f(x)$ is the maximum sum obtained by selecting r components from $|x|$.
So we are choosing the r-largest components.
Define:

$$f(x) = \max_{\substack{|C|=r \\ C \subseteq \{1,2,\ldots,n\}}} \sum_{i \in C} |x_i|.$$

Let $x, y \in \mathbb{R}^n$, $\lambda \in [0,1]$. We need to show:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

Let C_1 be the set of indices that results in the maximum for $f(\lambda x + (1-\lambda)y)$:

$$f(\lambda x + (1-\lambda)y) = \sum_{i \in C_1} |\lambda x_i + (1-\lambda)y_i|.$$

Since for $i \in C_1$,

$$|\lambda x_i + (1-\lambda)y_i| \leq \lambda |x_i| + (1-\lambda)|y_i|,$$

we have:

$$f(\lambda x + (1-\lambda)y) = \sum_{i \in C_1} |\lambda x_i + (1-\lambda)y_i| \leq \sum_{i \in C_1} (\lambda |x_i| + (1-\lambda)|y_i|).$$

Expanding further:

$$f(\lambda x + (1-\lambda)y) \leq \lambda \sum_{i \in C_1} |x_i| + (1-\lambda) \sum_{i \in C_1} |y_i|.$$

Let C_2, C_3 be the sets of indices for $f(x)$ and $f(y)$, respectively:

$$\sum_{i \in C_2} |x_i| = f(x), \quad \sum_{i \in C_3} |y_i| = f(y).$$

From the above:

$$f(\lambda x + (1-\lambda)y) \leq \lambda \sum_{i \in C_2} |x_i| + (1-\lambda) \sum_{i \in C_3} |y_i|.$$

Thus:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

Hence, $f(x)$ is a convex function.

Question 3.25

Let $p, q \in \mathbb{R}^n$ represent two probability distributions on $\{1, \ldots, n\}$ (so $p \geq 0$, $q \geq 0$, and $1^T p = 1^T q = 1$).
We define the maximum probability distance $d_{\text{mp}}(p, q)$ between p and q as the maximum difference in probability assigned by p and q, over all events:

$$d_{\text{mp}}(p,q) = \max\{|\operatorname{prob}(p,C) - \operatorname{prob}(q,C)| \mid C \subseteq \{1,\ldots,n\}\}.$$

Here, $\operatorname{prob}(p, C)$ is the probability of C under the distribution p, i.e.,

$$\operatorname{prob}(p,C) = \sum_{i \in C} p_i.$$

Find a simple expression for d_{mp}, involving $\|p - q\|_1 = \sum_{i=1}^n |p_i - q_i|$, and show that d_{mp} is a convex function on $\mathbb{R}^n \times \mathbb{R}^n$. (Its domain is $\{(p,q) \mid p, q \geq 0, 1^T p = 1^T q = 1\}$, but it has a natural extension to all of $\mathbb{R}^n \times \mathbb{R}^n$.)

Solution

Let $d_{mp}(p,q)$ denote the maximum difference in probability assigned by two distributions p and q over all possible events.

Events imply subsets of $\{1,2,\ldots,n\}$. For any subset $C \subseteq \{1,2,\ldots,n\}$, we have:

$$\text{prob}(p,C) = \sum_{i \in C} p_i,$$

and:

$$\text{prob}(q,C) = \sum_{i \in C} q_i.$$

Thus:

$$d_{mp}(p,q) = \max_C |\, \text{prob}(p,C) - \text{prob}(q,C)|,$$

where the maximum is taken over all possible subsets C. For any subset C,:

$$|\, \text{prob}(p,C) - \text{prob}(q,C)| = \Big| \sum_{i \in C} p_i - \sum_{i \in C} q_i \Big|.$$

Simplifying:

$$|\, \text{prob}(p,C) - \text{prob}(q,C)| = \Big| \sum_{i \in C} (p_i - q_i) \Big|.$$

Finding the subset C that maximizes this difference
To maximize:

$$\Big| \sum_{i \in C} (p_i - q_i) \Big|,$$

we should:
Include $i \in C$ when $p_i - q_i > 0$.
Exclude $i \in C$ when $p_i - q_i < 0$.
We can say:

$$d_{mp}(p,q) = \max_C \Big| \sum_{i=1}^{n} (p_i - q_i) \Big|,$$

where:

$$C = \sum_{i=1}^{n} \max(p_i - q_i, 0).$$

Define: $C_1 = \{i : p_i > q_i\}$, $C_2 = \{i : p_i < q_i\}$.
We start with:

$$d_{mp}(p,q) = \Big| \sum_{i=1}^{n} (p_i - q_i) \Big|, \quad \text{where } i \in C.$$

Expanding:

$$d_{mp}(p,q) = \sum_{i=1}^{n} (p_i - q_i), \quad \text{where } i \in C,$$

$$d_{mp}(p,q) = \sum_{i=1}^{n} \max(p_i - q_i, 0).$$

Let:

$$\sum_{i=1}^{n} (p_i - q_i) = 0, \quad \text{as } \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} q_i = 1.$$

So:

$$\sum_{i=1}^{n}(p_i - q_i), \quad \text{where } i \in C_1 = -\sum_{i=1}^{n}(p_i - q_i), \quad \text{where } i \in C_2$$

Thus:

$$d_{mp}(p,q) = \sum_{i=1}^{n} \max(p_i - q_i, 0).$$

The L1 norm is defined as:

$$\|p - q\|_1 = \sum_{i=1}^{n} |p_i - q_i|.$$

Expanding:

$$\|p - q\|_1 = \sum_{i=1}^{n}(p_i - q_i), \quad (\text{where } i \in C_1) + \sum_{i=1}^{n}(q_i - p_i), \quad (\text{where } i \in C_2.)$$

This simplifies to:

$$\|p - q\|_1 = 2\sum_{i=1}^{n} \max(p_i - q_i, 0), \quad [\text{since both sums are equal}].$$

Thus:

$$d_{mp}(p,q) = \frac{\|p - q\|_1}{2}.$$

We need to show:

$$d_{mp}(\lambda p_1 + (1-\lambda)p_2, \lambda q_1 + (1-\lambda)q_2) \leq \lambda d_{mp}(p_1, q_1) + (1-\lambda)d_{mp}(p_2, q_2).$$

Since:

$$d_{mp}(p,q) = \frac{\|p - q\|_1}{2},$$

we need to show:

$$\frac{1}{2}\|\lambda p_1 + (1-\lambda)p_2 - (\lambda q_1 + (1-\lambda)q_2)\|_1 \leq \frac{\lambda}{2}\|p_1 - q_1\|_1 + \frac{(1-\lambda)}{2}\|p_2 - q_2\|_1.$$

Since the L1 norm is convex:

$$\|\lambda x + (1-\lambda)y\|_1 \leq \lambda\|x\|_1 + (1-\lambda)\|y\|_1,$$

where x,y $\in \mathbb{R}^n$ and $\lambda \in [0,1]$
So,

$$\|\lambda(p_1 - q_1) + (1-\lambda)(p_2 - q_2)\|_1 \leq \lambda\|p_1 - q_1\|_1 + (1-\lambda)\|p_2 - q_2\|_1.$$

Thus:

$$d_{mp}(p,q) = \frac{\|p - q\|_1}{2},$$

is a convex function on $\mathbb{R}^n \times \mathbb{R}^n$.