Homework Assignment #2 Solutions

1. Before we begin, note that convexity (concavity) of a function implies its quasiconvexity (quasiconcavity). Hence, once we have shown convexity (concavity) of a function, quasiconvexity (quasiconcavity) follows immediately.

   For each f below, define $S_\alpha = \{ x \in \mathbf{dom}\, f \mid f(x) \leq \alpha \}$ (*i.e.*, the α-sublevel set) and $\bar{S}_\alpha = \{ x \in \mathbf{dom}\, f \mid f(x) \geq \alpha \}$ (*i.e.*, the α-superlevel set).

   (a) $f(x) = e^x - 1$ on $x \in \mathbf{R}_+$: f is obviously strictly convex, and therefore quasiconvex. But it is also quasiconcave, because the superlevel set

   $$\bar{S}_\alpha = \{ x \in \mathbf{R}_+ \mid e^x - 1 \geq \alpha \} = \mathbf{R}_+ \cap [\log(\alpha + 1), \infty)$$

   is convex.

   (b) $f(x) = \max_{i=1,\ldots,k} \|A^{(i)}x - b^{(i)}\|$ on $x \in \mathbf{R}^n$: We know that any norm is a convex function, and convexity is preserved under affine transformation of the domain, so the functions $\|A^{(i)}x - b^{(i)}\|$ are convex. The function f is the pointwise maximum of these, and so is convex (hence, also quasiconvex).

   Is the function quasiconcave? No. The set

   $$\bar{S}_\alpha = \bigcup_{i=1,\ldots,k} \{ x \mid \|A^{(i)}x - b^{(i)}\| \geq \alpha \}$$

   is just the union of the exteriors of several ellipsoids, which is not (in general) convex.

   (c) $f(X) = \mathbf{Tr}\, X^{-1}$, on $\{ X \mid X = X^T \succ 0 \}$: The first thing to do is guess whether it is convex or not (because if it is not, you can get yourself a lot of wasted time trying to show it is). You should check obvious things like the scalar case, where $f(x) = 1/x$, and is convex when $x > 0$.

   It turns out the function is convex, but it's not so easy to show. Here's one proof. We examine the behavior of $g(t) := f(X + tV)$, for some $V = V^T \neq 0$, around $t = 0$

   $$\begin{aligned} g(t) &= \mathbf{Tr}\,(X + tV)^{-1} = \mathbf{Tr}\, X^{-1/2} \left(I + tX^{-1/2} V X^{-1/2} \right)^{-1} X^{-1/2} \\ &= \mathbf{Tr}\, X^{-1/2} Q (I + t\Lambda)^{-1} Q^T X^{-1/2}, \end{aligned}$$

   where $Q \Lambda Q^T$ is the eigenvalue decomposition of $X^{-1/2} V X^{-1/2}$. If we define $y_1, \ldots, y_n$ to be the columns of $Y = X^{-1/2} Q$, then $g(t)$ becomes

   $$g(t) = \mathbf{Tr}\, Y (I + t\Lambda)^{-1} Y^T = \sum_{i=1}^{n} \frac{\|y_i\|^2}{1 + t\lambda_i}.$$

The individual terms $\|y_i\|^2/(1+t\lambda_i)$ are convex on the interval of interest (the neighborhood around $t=0$ for which $X+tV \succ 0$, so $1+t\lambda_i > 0$) because the second derivative

$$\frac{d^2}{dt^2}\left(\frac{\|y_i\|^2}{(1+t\lambda_i}\right) = \frac{2\|y_i\|^2\lambda_i^2}{(1+t\lambda_i)^3}$$

is positive because $1+t\lambda_i > 0$. Therefore, so is g and, in turn, so is f. As a result, f is *convex*, and *not* concave.

It is *not* quasiconcave: for a simple counterexample, consider $X_1 = \mathbf{diag}\,(1, 0.1)$, $X_2 = \mathbf{diag}\,(0.1, 1)$:

$$\mathbf{Tr}\left(\frac{X_1+X_2}{2}\right)^{-1} \not\geq \min\left\{\mathbf{Tr}\,X_1, \mathbf{Tr}\,X_2\right\}.$$

Here are a couple of additional convexity proofs for this function. The first examines the Hessian of g:

$$\begin{aligned} g'(t) &= -\mathbf{Tr}\,(X+tV)^{-1}V(X+tV)^{-1} \\ g''(t) &= 2\mathbf{Tr}\,(X+tV)^{-1}V(X+tV)^{-1}V(X+tV)^{-1} \\ \Longrightarrow \quad g''(0) &= 2\mathbf{Tr}\,X^{-1}VX^{-1}VX^{-1} \\ &= 2\left\|X^{-1/2}VX^{-1}\right\|_F^2 \end{aligned}$$

where $\|Z\|_F = \sqrt{\mathbf{Tr}\,Z^TZ}$ is the Frobenius norm of a matrix Z. Since, for any $V \neq 0$, $g''(0) > 0$, f must be convex.

Another proof uses the epigraph of f. This is the set

$$\mathbf{epi} f = \left\{(X,t) \mid t \in \mathbf{R},\ \ X = X^T \succ 0,\ \ \mathbf{Tr}\,X^{-1} \leq t\right\}.$$

Now, we have $\mathbf{Tr}\,X^{-1} \leq t$ if and only if there exist a matrix $Y \succeq X^{-1}$ with $\mathbf{Tr}\,Y \leq t$. In turn, from the Schur complement result of problem 2 (homework 1), the inequality $Y \succeq X^{-1}$, $X \succ 0$ is equivalent to

$$\begin{bmatrix} Y & I \\ I & X \end{bmatrix} \succeq 0,$$

which is a (convex) linear matrix inequality in X, Y. Therefore, $\mathbf{epi} f$ is the projection on the space of (X,t)-variables, of the convex set

$$S = \left\{(X,Y,t) \mid t \in \mathbf{R},\ \begin{bmatrix} Y & I \\ I & X \end{bmatrix} \succeq 0,\ \ \mathbf{Tr}\,Y \leq t\right\},$$

and therefore $\mathbf{epi} f$ is convex.

(d) $f(x) = \frac{a^Tx+b}{c^Tx+d}$, on $\{\, x \mid c^Tx + d > 0 \,\}$:

If we examine $g(t) = f(x+tv)$ for some direction $v \neq 0$ around $t=0$, we find that

$$g''(0) = (2/\tilde{d}^3)\langle c, v\rangle\langle \tilde{b}c - \tilde{d}a, v\rangle,$$

where $\tilde{b} = a^T x + b$ and $\tilde{d} = c^T x + d$. From this, it is easy to see that, in general, f is neither convex nor concave.

However, we have already been told that this function is quasilinear in x; to see this, note that

$$\begin{aligned} S_\alpha &= \{\, x \mid (a - \alpha c)^T x + (b - \alpha d) \leq 0, c^T x + d > 0 \,\} \text{ and} \\ \bar{S}_\alpha &= \{\, x \mid (a - \alpha c)^T x + (b - \alpha d) \geq 0, c^T x + d > 0 \,\} \end{aligned}$$

are both convex for any α.

(e) $f(x) = \sum_{i=1}^n x_i \log x_i - (\sum_{i=1}^n x_i) \log \sum_{i=1}^n x_i$ on $x_i > 0$, $i = 1, \ldots, n$: For some of you, this may resemble the familiar formula from information theory for the entropy of a probability distribution—especially if it is rewritten as follows:

$$f(x) = \sum_{i=1}^n x_i \log \frac{x_i}{\sum_{i=1}^n x_i}.$$

So, you may guess that this function is convex.

Unfortunately the hint given is of not much help, since it is wrong. I apologize for this ...

One way to prove convexity is to use by using the fact that the function $g(x,t) = \sum_{i=1}^n x_i \log(x_i/t)$ is convex on its domain $\mathbf{dom}\, g = \{(x,t) \mid x_i > 0,\ t > 0\}$. Therefore, it is also convex on the intersection of $\mathbf{dom}\, g$ and the hyperplane defined by the equality $t = \sum_i x_i$, where $g(x,t)$ coincides with $f(x)$.

Another method is to compute the Hessian of f. The Hessian of f is surprisingly simple:

$$D^2 f(x) = \begin{bmatrix} 1/x_1 & 0 & \cdots & 0 \\ 0 & 1/x_2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & 1/x_n \end{bmatrix} - \frac{1}{\sum_{i=1}^n x_i} \mathbf{1}\mathbf{1}^\mathbf{T}$$

where $\mathbf{1} \in \mathbf{R}^\mathbf{n}$ is the vector of all ones. We will show that this matrix is positive semidefinite which proves that f is convex. Define the diagonal matrix $X \in \mathbf{R}^{n \times n}$ as

$$X = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & x_n \end{bmatrix}$$

and therefore $D^2 f(x)$ can be re-written as

$$D^2 f(x) = X^{-1} - \frac{1}{\sum_{i=1}^n x_i} \mathbf{1}\mathbf{1}^\mathbf{T}.$$

Since X is diagonal and with positive diagonal elements ($x_i > 0$), we have $X = X^T \succ 0$. Therefore, by congruence, positive definiteness of $D^2 f(x)$ is equivalent to positive definiteness of $X^T D^2 f(x) X = X D^2 f(x) X$. But

$$X D^2 f(x) X = X - \frac{1}{\sum_{i=1}^n x_i} X \mathbf{1}\mathbf{1}^\mathbf{T}\mathbf{X}.$$

and

$$\frac{XD^2f(x)X}{\sum_{i=1}^n x_i} = \frac{X}{\sum_{i=1}^n x_i} - \left(\frac{X\mathbf{1}}{\sum_{i=1}^n x_i}\right)\left(\frac{\mathbf{1^TX}}{\sum_{i=1}^n x_i}\right).$$

Now let $p_i = x_i/\sum_{i=1}^n x_i$ for $i = 1, \ldots, n$. Clearly, these p_i's define a probability distribution because $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$. Therefore

$$\frac{X^TD^2f(x)X}{\sum_{i=1}^n x_i} = \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & p_n \end{bmatrix} - \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix}.$$

Consider $y^T\left[\frac{X^TD^2f(x)X}{\sum_{i=1}^n x_i}\right]y$ for $y \neq 0$:

$$\begin{aligned} y^T\left[\frac{X^TD^2f(x)X}{\sum_{i=1}^n x_i}\right]y &= \sum_{i=1}^n p_iy_i^2 - \left(\sum_{i=1}^n p_iy_i\right)\left(\sum_{i=1}^n p_iy_i\right) \\ &= \mathbf{E}y^2 - (\mathbf{E}y)^2 \\ &= \mathbf{var}(y) \\ &= \geq 0. \end{aligned}$$

where we have used the fact that the variance of any variable y is always non-negative. Therefore, $\frac{X^TD^2f(x)X}{\sum_{i=1}^n x_i} \succeq 0$ and it follows that $D^2f(x) \succeq 0$.

$f(\cdot)$ is not quasiconcave because the set

$$\bar{S}_\alpha = \left\{ x \in \mathbf{R}_+^n \;\middle|\; \sum_{i=1}^n x_i \log x_i - \sum_{i=1}^n x_i \log \sum_{i=1}^n x_i \geq \alpha \right\}$$

is not convex. Consider, for example, $n = 2$, $x = (1, 0.1)$, $y = (0.1, 1)$, and $\alpha = -0.5$; x and y are in $\bar{S}_{-0.5}$, but $\frac{1}{2}(x + y)$ is not.

(f) $f(x) = \sum_{i=1}^r |x|_{[i]}$ on $\mathbf{R}^n$, 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 $|x|$: We show that f can be written as the pointwise maximum of linear functions and is therefore convex. Consider the vectors $a_i \in \mathbf{R}^n$ with $n - r$ elements equal to zero, and r elements which each are either $+1$ or -1. Clearly, there are $2^rC_n^r$ of these vectors and

$$f(x) = \max_i a_i^Tx.$$

Therefore, it follows that f is convex and, as a result, also quasiconvex.

f is not concave or quasiconcave. For example, consider the simplest case with $r = 1$ and $n = 1$. Then $f(x) = |x|$ which is clearly neither concave nor quasiconcave.

2. What we have to show is that all sublevel sets are convex. Since in $\mathbf{dom}\, f$ we have $h(x) > 0$ we can write the sublevel set as

$$S(\alpha) = \{x | f(x) \leq \alpha\} = \{x | g(x) \leq \alpha h(x)\}.$$

Since $h(x)$ is concave, we have $-h(x)$ convex, hence for $\alpha > 0$ the function $g(x) - \alpha h(x)$ is convex. From this fact is follows that for any $\alpha > 0$ the set $g(x) - \alpha f(x) \leq 0$ (which equals $S(\alpha)$) is convex, so we conclude that $f(x)$ is quasiconvex.

3. (a) $\mathbf{E}x = \sum_{i=1}^{n} p_i a_i$ is a linear function in p. Therefore $\mathbf{E}x$ is convex, concave, quasiconvex and quasiconcave.

   (b) $\mathbf{Prob}(x \geq \alpha) = \sum_{i=i_\alpha}^{n} p_i$ where $i_\alpha = \min\{i | a_i \leq \alpha, i = 1, \ldots, n\}$. It is a linear function in p. Thus $\mathbf{Prob}(x \geq \alpha)$ is convex, concave, quasiconvex and quasiconcave. Moreover, it is a log-concave function because $\mathbf{Prob}(x \geq \alpha)$ is non-negative and concavity implies log-concavity.

   (c) $\mathbf{Prob}(\alpha \leq x \leq \beta) = \sum_{i=i_\alpha}^{i_\beta} p_i$, where $i_\alpha = \min\{i | a_i \leq \alpha, i = 1, \ldots, n\}$ and $i_\beta = \max\{i | a_i \geq \beta, i = 1, \ldots, n\}$. It is a linear function in p and hence convex, concave, quasiconvex and quasiconcave. $\mathbf{Prob}(\alpha \leq x \leq \beta)$ is also log-concave because it is non-negative and concave.

   (d) $p_i \log p_i$ is a convex function for each $i \in \{1, \ldots, n\}$, because its second derivative $1/p_i$ is positive on the probability simplex. Therefore, $\sum_i p_i \log p_i$ is convex (and hence quasiconvex) because the summation of convex functions is convex.

   (e) $\mathbf{var}x = \mathbf{E}x^2 - (\mathbf{E}x)^2$. Since $(\mathbf{E}x)^2$ is a quadratic function in p, it is convex and $-(\mathbf{E}x)^2$ is concave. $\mathbf{E}x^2$ is a linear function in p and is concave. Therefore, $\mathbf{var}x$ is concave in p because the summation of two concave functions is concave. By its concavity and non-negativity, $\mathbf{var}x$ is also quasiconcave and log-concave.

   (f) Recall that $\text{Quartile}(x) = \min\{\alpha | \mathbf{Prob}(x \leq \alpha) \geq 0.25\}$. We first show that the sublevel set $\{p | \text{Quartile}(x) \leq \alpha\}$ is convex: We have

$$\text{Quartile(x)} \leq \alpha \iff \mathbf{Prob}(x \leq a_j) \geq 0.25 \iff \sum_{i=1}^{j} p_i \geq 0.25,$$

   which is just a half-space, so the sublevel set is convex. The *superlevel set* $\{p | \text{Quartile}(x) \geq \alpha\}$ is also convex:

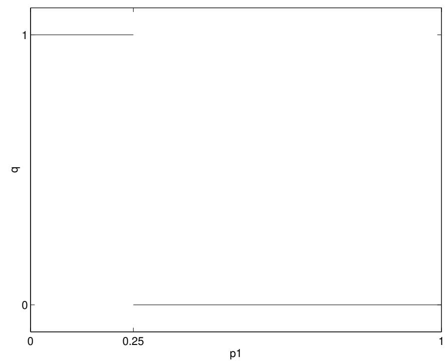
$$\text{Quartile(x)} \geq \alpha \iff \mathbf{Prob}(x \geq a_k) \geq 0.75 \iff \sum_{i=k}^{j} p_i \geq 0.75.$$

   Therefore $\text{Quartile}(x)$ is quasiconvex and quasiconcave (hence quasilinear).

   In general $\text{Quartile}(x)$ is neither convex nor concave. Consider the example with $n = 2$, $a_1 = 0$ and $a_2 = 1$, we have

$$\text{Quartile}(x) = \begin{cases} 0, & \text{for } p_1 + p_2 = 1,\ 0.25 \leq p_1 \leq 1 \\ 1, & \text{for } p_1 + p_2 = 1,\ 0 \leq p_1 < 0.25\ . \end{cases}$$

   Plot $\text{Quartile}(x)$ on the line segment $p_1 + p_2 = 1, p_1 \in [0, 1]$

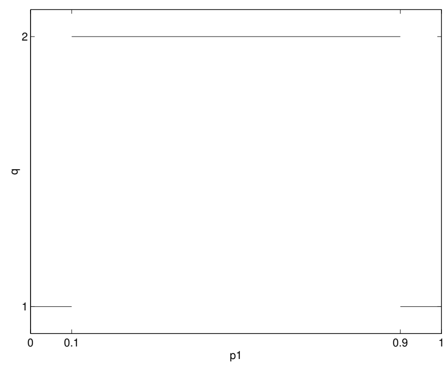


Clearly it is neither convex nor concave.

(g) Let $F(x)$ denote the function of the size of the smallest set $\mathcal{A} \subseteq \{a_1, \ldots, a_n\}$ with probability $\geq 90\%$. Clearly the range of $F(x)$ is $\{1, \ldots, n\}$. Consider the superlevel set $T_\alpha = \{\ p \mid F(x) \geq \alpha\ \}$ for $\alpha \in [1, n]$. If we define $M_i = \{\ p \mid f_i(p) = \sum_{j=1}^{i} p_{[j]} \leq 90\%\ \}$ ($p_{[i]}$ denotes the ith largest p_j for $j = 1, \ldots, n$), then we have $T_\alpha = M_{\lfloor \alpha \rfloor}$ where $\lfloor \alpha \rfloor$ denotes the smallest integer no greater than α. We have shown in the lecture that the function $f_i(p)$ is convex on $\mathbf{R}^n$, then it is also convex on the probability simplex because any convex restriction of a convex function is convex. Thus M_i is in fact sublevel sets of $f_i(p)$ for $i = 1, \ldots, n$ and is convex in p. Therefore the superlevel set T_α is convex and $F(x)$ is quasiconcave.

In general, $F(x)$ is not convex, concave nor quasiconvex. This can be shown by the example with $n = 2$, $a_1 = 0$ and $a_2 = 1$, we have

$$F(x) = \begin{cases} 1, & \text{for } p_1 + p_2 = 1,\ 0.9 \leq p_1 \leq 1 \\ 2, & \text{for } p_1 + p_2 = 1,\ 0.1 < p_1 < 0.9 \\ 1, & \text{for } p_1 + p_2 = 1,\ 0 \leq p_1 \leq 0.1\ . \end{cases}$$



$F(x)$ is not convex, concave nor quasiconvex.

The same example can also be used to show that $F(x)$ is neither log-convex nor log-concave in general.

4. *Matrix convexity.* All we need to show is that for any $X \succ 0$ ($X \in \mathbf{R}^{n \times n}$) and any vector $v \in \mathbf{R}^n$, $g(X) = v^T f(X) v$ is always positive and is convex in X. The former is easy because $X \succ 0$ implies that $X^{-1} \succ 0$ and therefore $g(X) = v^T f(X) v = v^T X^{-1} v > 0$.

One method to show g is convex is to prove that the epigraph of g is convex. We have

$$\mathbf{epi}\,(g) = \left\{(X, t) \mid v^T X^{-1} v \le t\right\}.$$

The condition $v^T X^{-1} v \le t$, $X \succ 0$ is equivalent to the linear matrix inequality

$$\begin{bmatrix} t & v^T \\ v & X \end{bmatrix} \succeq 0, \quad X \succ 0$$

and therefore

$$\mathbf{epi}\,(g) = \left\{(X, t) \mid \begin{bmatrix} t & v^T \\ v & X \end{bmatrix} \succeq 0,\ X \succ 0\right\}$$

is a convex set and g is a convex function.