DSA3102: Solutions to Tutorial Set 4
Assigned: 31/08/23
Never Due

1. BV Problem 3.1

   Solution:

   (a) This is just the definition of convexity with $\lambda = (b-x)/(b-a)$.

   (b) We obtain the first inequality by subtracting $f(a)$ from both sides of the inequality in part (a). The second inequality follows from subtracting $f(b)$. Geometrically, the inequalities mean that the slope of the line segment between $(a, f(a))$ and $(b, f(b))$ is larger than the slope of the segment between $(a, f(a))$ and $(x, f(x))$, and smaller than the slope of the segment between $(x, f(x))$ and $(b, f(b))$.

   (c) This follows from part (b) by taking the limit for $x \to a$ on both sides of the first inequality, and by taking the limit for $x \to b$ on both sides of the second inequality.

   (d) From part (c),
   $$\frac{f'(b) - f'(a)}{b-a} \geq 0$$
   and taking the limit as $b \to a$ shows that $f''(a) \geq 0$.

2. BV Problem 3.5

   Hint: Use the differentiability of f and the first-order condition for convexity.

   Solution: The function F is differentiable with
   $$\begin{aligned} F'(x) &= -\frac{1}{x^2}\int_0^x f(t)\,dt + \frac{f(x)}{x} \\ F''(x) &= \frac{2}{x^3}\int_0^x f(t)\,dt - \frac{2f(x)}{x^2} + \frac{f'(x)}{x} \\ &= \frac{2}{x^3}\int_0^x (f(t) - f(x) - f'(x)(t-x))\,dt \end{aligned}$$
   Convexity now follows from the fact that
   $$f(t) \geq f(x) + f'(x)(t-x)$$
   for all $x, t \in \mathbf{dom}\, f$, which implies that $F''(x) \geq 0$.

3. BV Problem 3.13

   Solution: The negative entropy is strictly convex and differentiable on $\mathbb{R}^n_{++}$ and so
   $$f(u) > f(v) + \nabla f(v)^T (u - v)$$

for all $u, v \in \mathbb{R}^n_{++}$ with $u \neq v$. Evaluating both sides of the inequality, we obtain

$$\begin{aligned}\sum_i u_i \log u_i &> \sum_i v_i \log v_i + \sum_i (\log v_i + 1)(u_i - v_i) \\ &= \sum_i u_i \log v_i + 1^T(u - v)\end{aligned}$$

Re-arranging this inequality gives the desired result.

4. BV Problem 3.16

Solution:

(a) $f(x) = e^x - 1$ with $\mathbf{dom}\, f = \mathbb{R}$. This function is strictly convex, and therefore quasiconvex. Also quasiconcave but not concave.

(b) $f(x_1, x_2) = x_1 x_2$ with $\mathbf{dom}\, f = \mathbb{R}^2_{++}$. The Hessian of f is

$$\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

which is neither positive semidefinite nor negative semidefinite. Therefore, f is neither convex nor concave. It is quasiconcave, since its superlevel sets

$$\{(x_1, x_2) \in \mathbb{R}^2_{++} | x_1 x_2 \geq 0\}$$

are convex. It is not quasiconvex.

(c) $f(x_1, x_2) = 1/(x_1 x_2)$ with $\mathbf{dom}\, f = \mathbb{R}^2_{++}$. The Hessian of f is

$$\nabla^2 f(x) = \frac{1}{x_1 x_2} \begin{bmatrix} 2/x_1^2 & 1/(x_1 x_2) \\ 1/(x_1 x_2) & 2/x_2^2 \end{bmatrix} \succeq 0$$

Therefore, f is convex and quasiconvex. It is not quasiconcave or concave.

(d) $f(x_1, x_2) = x_1/x_2$ with $\mathbf{dom}\, f = \mathbb{R}^2_{++}$. The Hessian of f is

$$\nabla^2 f(x) = \begin{bmatrix} 0 & -1/x_2^2 \\ -1/x_2^2 & 2x_1/x_2^3 \end{bmatrix}$$

which is not positive or negative semidefinite. Therefore, f is not convex or concave. It is quasiconvex and quasiconcave (i.e., quasilinear), since the sublevel and super- level sets are halfspaces.

(e) $f(x_1, x_2) = x_1^2/x_2$ with $\mathbf{dom}\, f = \mathbb{R}^2_{++}$. f is convex, as mentioned on page 72 and worked out in class. Therefore, f is convex and quasiconvex. It is not concave or quasiconcave (see the figure in the book).

(f) $f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$ where $0 \leq \alpha \leq 1$ with $\mathbf{dom}\, f = \mathbb{R}^2_{++}$. The Hessian of f is

$$\begin{aligned}\nabla^2 f(x) &= \begin{bmatrix} \alpha(\alpha-1)x_1^{\alpha-2}x_2^{1-\alpha} & \alpha(\alpha-1)x_1^{\alpha-1}x_2^{-\alpha} \\ \alpha(\alpha-1)x_1^{\alpha-1}x_2^{-\alpha} & (1-\alpha)(-\alpha)x_1^{\alpha}x_2^{-\alpha-1} \end{bmatrix} \\ &= \alpha(1-\alpha)x_1^\alpha x_2^{1-\alpha} \begin{bmatrix} -1/x_1^2 & 1/(x_1 x_2) \\ 1/(x_1 x_2) & -1/x_2^2 \end{bmatrix} \\ &= -\alpha(1-\alpha)x_1^\alpha x_2^{1-\alpha} \begin{bmatrix} 1/x_1 \\ -1/x_2 \end{bmatrix} \begin{bmatrix} 1/x_1 \\ -1/x_2 \end{bmatrix}^T\end{aligned}$$

Hence,

$$-\nabla^2 f(x) \succeq 0$$

We conclude that f is concave and quasiconcave. It is not convex or quasiconvex.

5. BV Problem 3.17

   Solution: The first derivatives of f are given by

$$\frac{\partial f}{\partial x_i} = \left(\sum_i x_i^p\right)^{(1-p)/p} x_i^{p-1} = \left(\frac{f(x)}{x_i}\right)^{1-p}$$

   The second derivatives are

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{1-p}{x_i}\left(\frac{f(x)}{x_i}\right)^{-p}\left(\frac{f(x)}{x_j}\right)^{1-p} = \frac{1-p}{f(x)}\left(\frac{f(x)^2}{x_i x_j}\right)^{1-p}$$

   for $i \neq j$ and

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{1-p}{f(x)}\left(\frac{f(x)^2}{x_i^2}\right)^{1-p} - \frac{1-p}{x_i}\left(\frac{f(x)}{x_i}\right)^{1-p}.$$

   We need to show that

$$y^T \nabla^2 f(x) y = \frac{1-p}{f(x)}\left[\left(\sum_i \frac{y_i f(x)^{1-p}}{x_i^{1-p}}\right)^2 - \sum_i \frac{y_i^2 f(x)^{2-p}}{x_i^{2-p}}\right] \leq 0$$

   This follows by applying the Cauchy-Schwarz inequality $a^T b \leq \|a\|\|b\|$ with

$$a_i = \left(\frac{f(x)}{x_i}\right)^{-p/2}, \qquad b_i = y_i\left(\frac{f(x)}{x_i}\right)^{1-p/2}$$

   and noting that $\sum_i a_i^2 = 1$.

6. BV Problem 3.18

   Solution:

   (a) Define $g(t) = f(Z + tV)$, where $Z \succ 0$ and $V \in \mathbf{S}^n$. We have

$$\begin{aligned} g(t) &= \mathbf{tr}((Z+tV)^{-1}) \\ &= \mathbf{tr}(Z^{-1}(I + tZ^{-1/2}VZ^{-1/2})^{-1}) \\ &= \mathbf{tr}(Z^{-1}Q(I+t\Lambda)^{-1}Q^T) \\ &= \mathbf{tr}(Q^T Z^{-1} Q (I+t\Lambda)^{-1}) \\ &= \sum_i (Q^T Z Q)_{ii}(1+t\lambda_i)^{-1} \end{aligned}$$

   where we used the eigenvalue decomposition $Z^{-1/2}VZ^{-1/2} = Q\Lambda Q^T$. In the last equality we express g as a positive weighted sum of convex functions $1/(1+t\lambda_i)$, hence it is convex.

   (b) Define $g(t) = f(Z + tV)$, where $Z \succ 0$ and $V \in \mathbf{S}^n$. We have

$$\begin{aligned} g(t) &= (\det(Z+tV))^{1/n} \\ &= (\det(Z)^{1/2}\det(I + tZ^{-1/2}VZ^{-1/2})\det(Z)^{1/2})^{1/n} \\ &= (\det Z)^{1/n}\left(\prod_i (1+t\lambda_i)\right)^{1/n} \end{aligned}$$

   where $\lambda_i, i = 1, \ldots, n$ are the eigenvalues of $Z^{-1/2}VZ^{-1/2}$. From the last equality, we see that g is a concave function of t on $\{t | Z + tV \succ 0\}$, since $\det Z > 0$ and the geometric mean $(\prod_i x_i)^{1/n}$ is concave on $\mathbb{R}^n_{++}$.

7. (Reverse Jensen's inequality) Suppose f is convex, $\lambda_1 > 0$, $\lambda_i \leq 0, i = 2, \ldots, n$, and $\sum_{i=1}^n \lambda_i = 1$ and let $x_1, \ldots, x_n \in \mathbf{dom}\, f$. Show that the inequality

$$f\left(\sum_{i=1}^n \lambda_i x_i\right) \geq \sum_{i=1}^n \lambda_i f(x_i)$$

always holds.

Hint: Draw a picture for the $n = 2$ case first. For the general case, express x_1 as a convex combination of $\lambda_1 x_1 + \ldots + \lambda_n x_n$ and $x_2, \ldots, x_n$, and use Jensen's inequality.

Solution: Let

$$x_1 = \mu_1 \left(\sum_{i=1}^n \lambda_i x_i\right) + \mu_2 x_2 + \ldots + \mu_n x_n$$

for some $\mu_i \geq 0$ and $\sum_{i=1}^n \mu_i = 1$. Then applying Jensen's inequality, we have

$$f\left(\mu_1 \left(\sum_{i=1}^n \lambda_i x_i\right) + \mu_2 x_2 + \ldots + \mu_n x_n\right) \leq \mu_1 f\left(\sum_{i=1}^n \lambda_i x_i\right) + \mu_2 f(x_2) + \ldots + \mu_n f(x_n)$$

Now set

$$\mu_1 = \frac{1}{\lambda_1}, \quad \mu_i = -\frac{\lambda_i}{\lambda_1}, \quad \forall i = 2, \ldots, n$$

so $\mu_i \geq 0$ for each $i = 1, \ldots, n$ and

$$\sum_{i=1}^n \mu_i = \frac{1}{\lambda_1} - \frac{\lambda_2}{\lambda_1} - \ldots - \frac{\lambda_n}{\lambda_1} = \frac{1}{\lambda_1}(1 - \lambda_2 - \ldots - \lambda_n) = 1$$

because we are given that $\sum_{i=1}^n \lambda_i = 1$. Plugging these choices of $\{\mu_i\}_{i=1}^n$ into Jensen's inequality, we obtain

$$f(x_1) \leq \frac{1}{\lambda_1} f\left(\sum_{i=1}^n \lambda_i x_i\right) - \frac{\lambda_2}{\lambda_1} f(x_2) - \ldots - \frac{\lambda_n}{\lambda_1} f(x_n)$$

which upon rearrangements, yields the desired reverse Jensen's inequality.

8. (Alternate Criteria for Convexity)

This problem is rather challenging (at least for me). It is a clever way of using continuity to replace the "for every $\theta \in [0,1]$" clause in the definition of convexity with a "there exists some $\theta \in [0,1]$" clause.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a *continuous* function with $C = \mathbf{dom}(f)$ being convex. Suppose that f is such that for all $x, y \in C$, <u>there exists</u> $\theta \in (0,1)$ such that

$$f((1-\theta)x + \theta y) \leq \theta f(x) + (1-\theta) f(y). \tag{1}$$

Prove that f is convex.

This result is attributed to Hardy, Littlewood, and Polya [HLP]. But they in turn attribute it to both Reisz [R] and to Jessen [J].

[HLP] G. H. Hardy, J. E. Littlewood, and G. Pólya. 1952. Inequalities, 2d. ed. Cambridge: Cambridge University Press.

[R] M. Riesz. 1927. Sure les maxima des formes bilinéaires et sur les fonctionelles linéaires. Acta Mathematica 49(3–4):465–497.

[J] Om Uligheder imellem Potensmiddelværdier, Mat. Tidsskr. B, 1931

Solution: Suppose, to the contrary, that f is continuous and satisfies (1), but is not convex. Then, there exists $x, y \in C$ and $\bar{\theta} \in (0,1)$ such that $f((1-\bar{\theta})x + \bar{\theta}y) > \bar{\theta}f(x) + (1-\bar{\theta})f(y)$. By continuity, the set $A = \{\theta : f((1-\theta)x + \theta y) > \theta f(x) + (1-\theta)f(y)\}$ is open and contains $\bar{\theta}$. However $0, 1 \notin A$, so $\bar{\theta}$ is contained in a maximal open interval included in A. That is, there exists α, β satisfying $0 \leq \alpha < \bar{\theta} < \beta \leq 1$ such that for all $\theta \in (\alpha, \beta)$, we have $f((1-\theta)x + \theta y) > \theta f(x) + (1-\theta)f(y)$ but $f((1-\alpha)x + \alpha y) = \alpha f(x) + (1-\alpha)f(y)$ and $f((1-\beta)x + \beta y) = \beta f(x) + (1-\beta)f(y)$. Now consider the points $x' = (1-\alpha)x + \alpha y$ and $y' = (1-\beta)x + \beta y$. By construction, for every $0 < \theta < 1$, the point $(1-\theta)x' + \theta y'$ strictly between x' and y' satisfies $f((1-\theta)x' + \theta y') > \theta f(x') + (1-\theta)f(y')$. However, this violates (1) applied to the points x' and y', a contradiction. Therefore f must be convex.