System Optimization Method (ECE 6233) October 7, 2020
NYU Tandon School of Engineering, Fall 2021
Instructor: *Prof. Quanyan Zhu*

Solution to Homework 2

1. (i) The proof is given by induction in k.

   - Base case: $k = 1$. $\theta_1 = 1$ and $1 \cdot x_1 \in C$ is true.
     $k = 2$. By definition of a convex set.
   - Inductive Step: Assume all length $k = n - 1 > 2$ combinations are contained in C. Take a length $k = n$ combination of points in C

     $$\hat{x} = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

     where $\theta_1, \ldots, \theta_n \in \mathbb{R}$ and $\theta_i > 0$, $\theta_1 + \cdots + \theta_n = 1$. We would like to show that $\hat{\mathbf{x}}$ is also in set C.
     By the inductive hypothesis, the following length $n - 1$ combination

     $$y = \frac{\theta_1}{\theta_1 + \cdots + \theta_{n-1}} x_1 + \frac{\theta_2}{\theta_1 + \cdots + \theta_{n-1}} x_2 + \cdots + \frac{\theta_{n-1}}{\theta_1 + \cdots + \theta_{n-1}} x_{n-1}$$

     is in C.Now, the point $\hat{x}$ can be written as

     $$\begin{aligned}\hat{x} =& (\theta_1 + \cdots + \theta_{n-1}) y + \theta_n x_n \\ =& (1 - \theta_n) y + \theta_n x_n\end{aligned}$$

     where $y \in C$ and $x_n \in C$. By definition, this point also belongs to the convex set C.
   - Conclusion: The statement is true.

(ii)

Solution. A set is convex if and only if its intersection with an arbitrary line $\{\hat{x}+tv \mid t \in \mathbf{R}\}$ is convex.

(a) We have

$$(\hat{x}+tv)^T A(\hat{x}+tv) + b^T(\hat{x}+tv) + c = \alpha t^2 + \beta t + \gamma$$

where

$$\alpha = v^T A v, \qquad \beta = b^T v + 2\hat{x}^T A v, \qquad \gamma = c + b^T\hat{x} + \hat{x}^T A\hat{x}.$$

The intersection of C with the line defined by $\hat{x}$ and v is the set

$$\{\hat{x}+tv \mid \alpha t^2 + \beta t + \gamma \le 0\},$$

which is convex if $\alpha \ge 0$. This is true for any v, if $v^T A v \ge 0$ for all v, *i.e.*, $A \succeq 0$. The converse does not hold; for example, take $A = -1$, $b = 0$, $c = -1$. Then $A \not\succeq 0$, but $C = \mathbf{R}$ is convex.

(iii)

Hyperbolic sets. Show that the *hyperbolic set* $\{x \in \mathbf{R}^2_+ \mid x_1x_2 \ge 1\}$ is convex. As a generalization, show that $\{x \in \mathbf{R}^n_+ \mid \prod_{i=1}^n x_i \ge 1\}$ is convex. *Hint.* If $a, b \ge 0$ and $0 \le \theta \le 1$, then $a^\theta b^{1-\theta} \le \theta a + (1-\theta)b$; see §3.1.9.

Solution.

(a) We prove the first part without using the hint. Consider a convex combination z of two points (x_1, x_2) and (y_1, y_2) in the set. If $x \succeq y$, then $z = \theta x + (1-\theta)y \succeq y$ and obviously $z_1z_2 \ge y_1y_2 \ge 1$. Similar proof if $y \succeq x$.
Suppose $y \not\succeq 0$ and $x \not\succeq y$, *i.e.*, $(y_1 - x_1)(y_2 - x_2) < 0$. Then

$$\begin{aligned}
&(\theta x_1 + (1-\theta)y_1)(\theta x_2 + (1-\theta)y_2) \\
&= \theta^2 x_1x_2 + (1-\theta)^2 y_1y_2 + \theta(1-\theta)x_1y_2 + \theta(1-\theta)x_2y_1 \\
&= \theta x_1x_2 + (1-\theta)y_1y_2 - \theta(1-\theta)(y_1 - x_1)(y_2 - x_2) \\
&\ge 1.
\end{aligned}$$

(b) Assume that $\prod_i x_i \ge 1$ and $\prod_i y_i \ge 1$. Using the inequality in the hint, we have

$$\prod_i(\theta x_i + (1-\theta)y_i) \ge \prod x_i^\theta y_i^{1-\theta} = (\prod_i x_i)^\theta (\prod_i y_i)^{1-\theta} \ge 1.$$

2. (i) Since function g is convex, for some $\lambda \in [0,1]$ we have

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y) \tag{1}$$

Function f is increasing, thus for LHS and RHS in (1), we have $f(LHS) \leq f(RHS)$.

$$f(g(\lambda x + (1-\lambda)y) \leq f(\lambda g(x) + (1-\lambda)g(y)) \tag{2}$$

Since function f is convex, the RHS of (2) also satisfies

$$f(\lambda g(x) + (1-\lambda)g(y)) \leq \lambda f(g(x)) + (1-\lambda)f(g(y)) \tag{3}$$

Combine (2) and (3) together, we have

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

for some $\lambda \in [0,1]$. Thus, $f(g(x))$ is a convex function.

(ii) For some $\lambda \in [0,1]$,

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \log\left(\sum_{i=1}^{n} e^{\lambda x_i + (1-\lambda)y_i}\right) \\ &= \log\left(\sum_{i=1}^{n} e^{\lambda x_i} \cdot e^{(1-\lambda)y_i}\right) \end{aligned}$$

According to Holder's inequality,

$$\sum_{k=1}^{n} |a_k b_k| \leq \left(\sum_{k=1}^{n} |a_k|^p\right)^{1/p} \left(\sum_{k=1}^{n} |b_k|^q\right)^{1/q}$$

where $1/p + 1/q = 1$ and $p, q > 1$.
Apply this inequality the expression inside log. Because e^x is always positive, we have

$$\left(\sum_{i=1}^{n} e^{\lambda x_i} \cdot e^{(1-\lambda)y_i}\right) \leq \left(\sum_{i=1}^{n} e^{\lambda x_i \cdot \frac{1}{\lambda}}\right)^{\lambda} \left(\sum_{i=1}^{n} e^{(1-\lambda)y_i \cdot \frac{1}{1-\lambda}}\right)^{1-\lambda}$$

Here we regard $1/\lambda$ as p and $1/(1-\lambda)$ as q. Take log on both sides won't change the inequality and $\log(AB) = \log(A) + \log(B)$. We obtain

$$
\begin{aligned}
f(\lambda x + (1-\lambda)y) =& \log\left(\sum_{i=1}^{n} e^{\lambda x_i} \cdot e^{(1-\lambda)y_i}\right) \\
\leq& \log\left(\sum_{i=1}^{n} e^{x_i}\right)^{\lambda} + \log\left(\sum_{i=1}^{n} e^{y_i}\right)^{1-\lambda} \\
=& \lambda \log\left(\sum_{i=1}^{n} e^{x_i}\right) + (1-\lambda)\log\left(\sum_{i=1}^{n} e^{y_i}\right) \\
=& \lambda f(x) + (1-\lambda) f(y)
\end{aligned}
$$

Thus, the log-exp function is convex.
Alternative proof:

Log-sum-exp. The Hessian of the log-sum-exp function is

$$\nabla^2 f(x) = \frac{1}{(\mathbf{1}^T z)^2}\left((\mathbf{1}^T z)\,\mathbf{diag}(z) - zz^T\right),$$

where $z = (e^{x_1}, \ldots, e^{x_n})$. To verify that $\nabla^2 f(x) \succeq 0$ we must show that for all v, $v^T \nabla^2 f(x) v \geq 0$, *i.e.*,

$$v^T \nabla^2 f(x) v = \frac{1}{(\mathbf{1}^T z)^2}\left(\left(\sum_{i=1}^{n} z_i\right)\left(\sum_{i=1}^{n} v_i^2 z_i\right) - \left(\sum_{i=1}^{n} v_i z_i\right)^2\right) \geq 0.$$

But this follows from the Cauchy-Schwarz inequality $(a^T a)(b^T b) \geq (a^T b)^2$ applied to the vectors with components $a_i = v_i\sqrt{z_i}$, $b_i = \sqrt{z_i}$.

(iii)

Solution. The first derivatives of f are given by

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

The second derivatives are

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

for $i \neq j$, and

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

We need to show that

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

This follows by applying the Cauchy-Schwarz inequality $a^T b \leq \|a\|_2 \|b\|_2$ 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 $\sum_i a_i^2 = 1$.

(iv) f is the composition of a norm, which is convex, and an affine function $Ax - b$.