# HW II Solutions

EE 588: Optimization for the information and data sciences

University of Southern California

Release Date: October 17, 2018

---

**3.15** *A family of concave utility functions.* For $0 < \alpha \leq 1$ let

$$u_\alpha(x) = \frac{x^\alpha - 1}{\alpha},$$

with $\mathbf{dom} u_\alpha = \mathbb{R}_+$. We also define $u_0(x) = \log x$ (with $\mathbf{dom} u_0 = \mathbb{R}_{++}$).

1. Show that for $x > 0$, $u_0(x) = \lim_{\alpha \to 0} u_\alpha(x)$.
2. Show that $u_\alpha$ are concave, monotone increasing, and all satisfy $u_\alpha(1) = 0$.

These functions are often used in economics to model the benefit or utility of some quantity of goods or money. Concavity of $u_\alpha$ means that the marginal utility (i.e., the increase in utility obtained for a fixed increase in the goods) decreases as the amount of goods increases. In other words, concavity models the effect of *satiation.*

**Solution**

1. In this limit, both the numerator and denominator go to zero, so we use l'Hopital's rule:

$$\lim_{\alpha \to 0} u_\alpha(x) = \lim_{\alpha \to 0} \frac{(d/d\alpha)(x^\alpha - 1)}{(d/d\alpha)\alpha} = \lim_{\alpha \to 0} \frac{x^\alpha \log x}{1} = \log x.$$

2. By inspection we have

$$u_\alpha(1) = \frac{1^\alpha - 1}{\alpha} = 0.$$

   The derivative is given by

$$u'_\alpha(x) = x^{\alpha - 1},$$

   which is positive for all $x$ (since $0 < \alpha < 1$), so these functions are increasing. To show concavity, we examine the second derivative:

$$u''_\alpha(x) = (\alpha - 1)x^{\alpha - 2}.$$

   Since this is negative for all $x$, we conclude that $u_\alpha$ is strictly concave.

■

- *Composition with an affine function.* Show that the following functions $f : \mathbb{R}^n \to \mathbb{R}$ are convex.

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

   **Solution.** $f$ is the composition of a norm, which is convex, and an affine function.

2. $f(x) = -\left(\det(A_0 + x_1A_1 + \cdots + x_nA_n)\right)^{1/m}$, on $\{x \mid A_0 + x_1A_1 + \cdots + x_nA_n \succ 0\}$, where $A_i \in \mathbb{S}^m$.

   **Solution.** $f$ is the composition of the convex function $h(X) = -(\det X)^{1/m}$ and an affine transformation. To see that $h$ is convex on $\mathbb{S}^m_{++}$, we restrict $h$ to a line and prove that $g(t) = -\det(Z + tV)^{1/m}$ is convex:

$$\begin{aligned} g(t) &= -(\det(Z + tV))^{1/m} \\ &= -(\det Z)^{1/m}(\det(I + tZ^{-1/2}VZ^{-1/2}))^{1/m} \\ &= -(\det Z)^{1/m}(\prod_{i=1}^{m}(1 + t\lambda_i))^{1/m} \end{aligned}$$

   where $\lambda_1, \ldots, \lambda_m$ denote the eigenvalues of $Z^{-1/2}VZ^{-1/2}$. We have expressed $g$ as the product of a negative constant and the geometric mean of $1 + t\lambda_i$, $i = 1, \ldots, m$. Therefore $g$ is convex. (See also exercise 3.18.)

3. $f(X) = \mathbf{tr}\,(A_0 + x_1A_1 + \cdots + x_nA_n)^{-1}$, on $\{x \mid A_0 + x_1A_1 + \cdots + x_nA_n \succ 0\}$, where $A_i \in \mathbb{S}^m$. (Use the fact that $\mathbf{tr}(X^{-1})$ is convex on $\mathbb{S}^m_{++}$; see exercise 3.18.)

   **Solution.** $f$ is the composition of $\mathbf{tr}X^{-1}$ and an affine transformation

$$x \mapsto A_0 + x_1A_1 + \cdots + x_nA_n.$$

**3.22** *Composition rules.* Show that the following functions are convex.

1. $f(x) = -\log(-\log(\sum_{i=1}^m e^{a_i^Tx+b_i}))$ on $\mathbf{dom}f = \{x \mid \sum_{i=1}^m e^{a_i^Tx+b_i} < 1\}$. You can use the fact that $\log(\sum_{i=1}^n e^{y_i})$ is convex.

   **Solution** $g(x) = \log(\sum_{i=1}^m e^{a_i^Tx+b_i})$ is convex (composition of the log-sum-exp function and an affine mapping), so $-g$ is concave. The function $h(y) = -\log y$ is convex and decreasing. Therefore $f(x) = h(-g(x))$ is convex. ■

2. $f(x, u, v) = -\sqrt{uv - x^Tx}$ on $\mathbf{dom}f = \{(x, u, v) \mid uv > x^Tx,\ u,\ v > 0\}$. Use the fact that $x^Tx/u$ is convex in $(x, u)$ for $u > 0$, and that $-\sqrt{x_1x_2}$ is convex on $\mathbb{R}^2_{++}$.

   **Solution** We can express $f$ as $f(x, u, v) = -\sqrt{u(v - x^Tx/u)}$. The function $h(x_1, x_2) = -\sqrt{x_1x_2}$ is convex on $\mathbb{R}^2_{++}$, and decreasing in each argument. The functions $g_1(u, v, x) = u$ and $g_2(u, v, x) = v - x^Tx/u$ are concave. Therefore $f(u, v, x) = h(g(u, v, x))$ is convex. ■

3. $f(x, u, v) = -\log(uv - x^Tx)$ on $\mathbf{dom}f = \{(x, u, v) \mid uv > x^Tx,\ u,\ v > 0\}$.

   **Solution** We can express $f$ as

$$f(x, u, v) = -\log u - \log(v - x^Tx/u).$$

   The first term is convex. The function $v - x^Tx/u$ is concave because $v$ is linear and $x^Tx/u$ is convex on $\{(x, u) \mid u > 0\}$. Therefore the second term in $f$ is convex: it is the composition of a convex decreasing function $-\log t$ and a concave function. ■

4. $f(x,t) = -(t^p - \|x\|_p^p)^{1/p}$ where $p > 1$ and $\mathbf{dom} f = \{(x,t) \mid t \geq \|x\|_p\}$. You can use the fact that $\|x\|_p^p/u^{p-1}$ is convex in $(x,u)$ for $u > 0$ (see exercise 3.23), and that $-x^{1/p}y^{1-1/p}$ is convex on $\mathbb{R}_+^2$ (see exercise 3.16).

   **Solution** We can express $f$ as

$$f(x,t) = -\left(t^{p-1}\left(t - \frac{\|x\|_p^p}{t^{p-1}}\right)\right)^{1/p} = -t^{1-1/p}\left(t - \frac{\|x\|_p^p}{t^{p-1}}\right)^{1/p}.$$

   This is the composition of $h(y_1, y_2) = -y_1^{1-1/p}y_2^{1/p}$ (convex and decreasing in each argument) and two concave functions

$$g_1(x,t) = t, \qquad g_1(x,t) = t - \frac{\|x\|_p^p}{t^{p-1}}.$$

   ■

5. $f(x,t) = -\log(t^p - \|x\|_p^p)$ where $p > 1$ and $\mathbf{dom} f = \{(x,t) \mid t > \|x\|_p\}$. You can use the fact that $\|x\|_p^p/u^{p-1}$ is convex in $(x,u)$ for $u > 0$ (see exercise 3.23).

   **Solution** Express $f$ as

$$\begin{aligned} f(x,t) &= -\log t^{p-1} - \log(t - \|x\|_p^p/t^{p-1}) \\ &= -(p-1)\log t - \log(t - \|x\|_p^p/t^{p-1}). \end{aligned}$$

   The first term is convex. The second term is the composition of a decreasing convex function and a concave function, and is also convex. ■

**3.24**

- *Some functions on the probability simplex.* Let $x$ be a real-valued random variable which takes values in $\{a_1, \ldots, a_n\}$ where $a_1 < a_2 < \cdots < a_n$, with $\mathbf{prob}(x = a_i) = p_i$, $i = 1, \ldots, n$. For each of the following functions of $p$ (on the probability simplex $\{p \in \mathbb{R}_+^n \mid \mathbf{1}^T p = 1\}$), determine if the function is convex, concave, quasiconvex, or quasiconcave.

  1. $\mathbb{E}x$.

     **Solution** $\mathbb{E}x = p_1a_1 + \cdots + p_na_n$ is linear, hence convex, concave, quasiconvex, and quasiconcave ■

  2. $\mathbf{prob}(x \geq \alpha)$.

     **Solution** Let $j = \min\{i \mid a_i \geq \alpha\}$. Then $\mathbf{prob}(x \geq \alpha) = \sum_{i=j}^n p_i$, This is a linear function of $p$, hence convex, concave, quasiconvex, and quasiconcave. ■

  3. $\mathbf{prob}(\alpha \leq x \leq \beta)$.

     **Solution** Let $j = \min\{i \mid a_i \geq \alpha\}$ and $k = \max\{i \mid a_i \leq \beta\}$. Then $\mathbf{prob}(\alpha \leq x \leq \beta) = \sum_{i=j}^k p_i$. This is a linear function of $p$, hence convex, concave, quasiconvex, and quasiconcave. ■

4. $\sum_{i=1}^n p_i \log p_i$, the negative entropy of the distribution.

**Solution** $p \log p$ is a convex function on $\mathbb{R}_+$ (assuming $0 \log 0 = 0$), so $\sum_i p_i \log p_i$ is convex (and hence quasiconvex).

The function is not concave or quasiconcave. Consider, for example, $n = 2$, $p^1 = (1, 0)$ and $p^2 = (0, 1)$. Both $p^1$ and $p^2$ have function value zero, but the convex combination $(0.5, 0.5)$ has function value $\log(1/2) < 0$. This shows that the superlevel sets are not convex. ■

5. $\mathbf{var} x = \mathbb{E}(x - \mathbb{E}x)^2$.

**Solution** We have

$$\mathbf{var} x = \mathbb{E}x^2 - (\mathbb{E}x)^2 = \sum_{i=1}^n p_i a_i^2 - (\sum_{i=1}^n p_i a_i)^2,$$

so $\mathbf{var} x$ is a concave quadratic function of $p$.

The function is not convex or quasiconvex. Consider the example with $n = 2$, $a_1 = 0$, $a_2 = 1$. Both $(p_1, p_2) = (1/4, 3/4)$ and $(p_1, p_2) = (3/4, 1/4)$ lie in the probability simplex and have $\mathbf{var} x = 3/16$, but the convex combination $(p_1, p_2) = (1/2, 1/2)$ has a variance $\mathbf{var} x = 1/4 > 3/16$. This shows that the sublevel sets are not convex. ■

6. $\mathbf{quartile}(x) = \inf\{\beta \mid \mathbf{prob}(x \le \beta) \ge 0.25\}$.

**Solution** The sublevel and the superlevel sets of $\mathbf{quartile}(x)$ are convex (see problem 2.15), so it is quasiconvex and quasiconcave.

$\mathbf{quartile}(x)$ is not continuous (it takes values in a discrete set $\{a_1, \ldots, a_n\}$), so it is not convex or concave. (A convex or a concave function is always continuous on the relative interior of its domain.) ■

7. The cardinality of the smallest set $\mathcal{A} \subseteq \{a_1, \ldots, a_n\}$ with probability $\ge 90\%$. (By cardinality we mean the number of elements in $\mathcal{A}$.)

**Solution** $f$ is integer-valued, so it can not be convex or concave. (A convex or a concave function is always continuous on the relative interior of its domain.)

$f$ is quasiconcave because its superlevel sets are convex. We have $f(p) \ge \alpha$ if and only if

$$\sum_{i=1}^k p_{[i]} < 0.9,$$

where $k = \max\{i = 1, \ldots, n \mid i < \alpha\}$ is the largest integer less than $\alpha$, and $p_{[i]}$ is the $i$th largest component of $p$. We know that $\sum_{i=1}^k p_{[i]}$ is a convex function of $p$, so the inequality $\sum_{i=1}^k p_{[i]} < 0.9$ defines a convex set.

In general, $f(p)$ is not quasiconvex. For example, we can take $n = 2$, $a_1 = 0$ and $a_2 = 1$, and $p^1 = (0.1, 0.9)$ and $p^2 = (0.9, 0.1)$. Then $f(p^1) = f(p^2) = 1$, but $f((p^1 + p^2)/2) = f(0.5, 0.5) = 2$. ■

8. The minimum width interval that contains 90% of the probability, i.e.,

$$\inf \{\beta - \alpha \mid \mathbf{prob}(\alpha \le x \le \beta) \ge 0.9\}.$$

**Solution** The minimum width interval that contains 90% of the probability must be of the form $[a_i, a_j]$ with $1 \leq i \leq j \leq n$, because

$$\mathbf{prob}(\alpha \leq x \leq \beta) = \sum_{k=i}^{j} p_k = \mathbf{prob}(a_i \leq x \leq a_j)$$

where $i = \min\{k \mid a_k \geq \alpha\}$, and $j = \max\{k \mid a_k \leq \beta\}$.
We show that the function is quasiconcave. We have $f(p) \geq \gamma$ if and only if all intervals of width less than $\gamma$ have a probability less than 90%,

$$\sum_{k=i}^{j} p_k < 0.9$$

for all $i, j$ that satisfy $a_j - a_i < \gamma$. This defines a convex set.
Since the function takes values on a finite set, it is not continuous and therefore neither convex nor concave. In addition it is not quasiconvex in general. Consider the example with $n = 2$, $a_1 = 0$, $a_2 = 1$, $p^1 = (0.95, 0.05)$ and $p^2 = (0.05, 0.95)$. Then $f(p^1) = 0$, $f(p^2) = 0$, but $f((p^1 + p^2)/2) = 1$. ■

**4.11** *Problems involving $\ell_1$- and $\ell_\infty$-norms.* Formulate the following problems as LPs. Explain in detail the relation between the optimal solution of each problem and the solution of its equivalent LP.

1. Minimize $\|Ax - b\|_\infty$ ($\ell_\infty$-norm approximation).
2. Minimize $\|Ax - b\|_1$ ($\ell_1$-norm approximation).
3. Minimize $\|Ax - b\|_1$ subject to $\|x\|_\infty \leq 1$.
4. Minimize $\|x\|_1$ subject to $\|Ax - b\|_\infty \leq 1$.
5. Minimize $\|Ax - b\|_1 + \|x\|_\infty$.

In each problem, $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$ are given. (See §**??** for more problems involving approximation and constrained approximation.)

**Solution**

1. Equivalent to the LP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & Ax - b \preceq t\mathbf{1} \\ & Ax - b \succeq -t\mathbf{1}. \end{array}$$

in the variables $x$, $t$. To see the equivalence, assume $x$ is fixed in this problem, and we optimize only over $t$. The constraints say that

$$-t \leq a_k^T x - b_k \leq t$$

for each $k$, i.e., $t \geq |a_k^T x - b_k|$, i.e.,

$$t \geq \max_k |a_k^T x - b_k| = \|Ax - b\|_\infty.$$

Clearly, if $x$ is fixed, the optimal value of the LP is $p^\star(x) = \|Ax - b\|_\infty$. Therefore optimizing over $t$ and $x$ simultaneously is equivalent to the original problem.

2. Equivalent to the LP

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T s \\ \text{subject to} & Ax - b \preceq s \\ & Ax - b \succeq -s. \end{array}$$

Assume $x$ is fixed in this problem, and we optimize only over $s$. The constraints say that

$$-s_k \leq a_k^T x - b_k \leq s_k$$

for each $k$, i.e., $s_k \geq |a_k^T x - b_k|$. The objective function of the LP is separable, so we achieve the optimum over $s$ by choosing

$$s_k = |a_k^T x - b_k|,$$

and obtain the optimal value $p^\star(x) = \|Ax - b\|_1$. Therefore optimizing over $t$ and $s$ simultaneously is equivalent to the original problem.

3. Equivalent to the LP

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T y \\ \text{subject to} & -y \preceq Ax - b \preceq y \\ & -\mathbf{1} \preceq x \preceq \mathbf{1}, \end{array}$$

with variables $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$.

4. Equivalent to the LP

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T y \\ \text{subject to} & -y \preceq x \preceq y \\ & -\mathbf{1} \preceq Ax - b \preceq \mathbf{1} \end{array}$$

with variables $x$ and $y$.

Another good solution is to write $x$ as the difference of two nonnegative vectors $x = x^+ - x^-$, and to express the problem as

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T x^+ + \mathbf{1}^T x^- \\ \text{subject to} & -\mathbf{1} \preceq Ax^+ - Ax^- - b \preceq \mathbf{1} \\ & x^+ \succeq 0, \quad x^- \succeq 0, \end{array}$$

with variables $x^+ \in \mathbb{R}^n$ and $x^- \in \mathbb{R}^n$.

5. Equivalent to

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T y + t \\ \text{subject to} & -y \preceq Ax - b \preceq y \\ & -t\mathbf{1} \preceq x \preceq t\mathbf{1}, \end{array}$$

with variables $x$, $y$, and $t$.

■

- Formulate the following optimization problems as semidefinite programs. The variable is $x \in \mathbb{R}^n$; $F(x)$ is defined as

$$F(x) = F_0 + x_1 F_1 + x_2 F_2 + \cdots + x_n F_n$$

with $F_i \in \mathcal{S}^m$. The domain of $f$ in each subproblem is $\mathbf{dom} f = \{x \in \mathbb{R}^n \mid F(x) \succ 0\}$.

1. Minimize $f(x) = c^T F(x)^{-1} c$ where $c \in \mathbb{R}^m$.
2. Minimize $f(x) = \max_{i=1,\ldots,K} c_i^T F(x)^{-1} c_i$ where $c_i \in \mathbb{R}^m$, $i = 1, \ldots, K$.
3. Minimize $f(x) = \sup_{\|c\|_2 \le 1} c^T F(x)^{-1} c$.
4. Minimize $f(x) = \mathbb{E}(c^T F(x)^{-1} c)$ where $c$ is a random vector with mean $\mathbb{E}c = \bar{c}$ and covariance $\mathbb{E}(c - \bar{c})(c - \bar{c})^T = S$.

**Solution.**

1.
$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \left[\begin{array}{cc} F(x) & c \\ c^T & t \end{array}\right] \succeq 0. \end{array}$$

2.
$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \left[\begin{array}{cc} F(x) & c_i \\ c_i^T & t \end{array}\right] \succeq 0, \quad i = 1, \ldots, K. \end{array}$$

3. $f(x) = \lambda_{\max}(F(x)^{-1})$, so $f(x) \le t$ if and only if $F(x)^{-1} \preceq tI$. Using a Schur complement we get
$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \left[\begin{array}{cc} F(x) & I \\ I & tI \end{array}\right] \succeq 0. \end{array}$$

4. $f(x) = \bar{c}^T F(x)^{-1} \bar{c} + (F(x)^{-1} S)$. If we factor $S$ as $S = \sum_{k=1}^m c_k c_k^T$ the problem is equivalent to
$$\text{minimize} \quad \bar{c}^T F(x)^{-1} \bar{c} + \textstyle\sum_{k=1}^m c_k^T F(x)^{-1} c_k,$$
which we can write as an SDP
$$\begin{array}{ll} \text{minimize} & t_0 + \sum_k t_k \\ \text{subject to} & \left[\begin{array}{cc} F(x) & \bar{c} \\ \bar{c}^T & t_0 \end{array}\right] \succeq 0 \\ & \left[\begin{array}{cc} F(x) & c_k \\ c_k^T & t_k \end{array}\right] \succeq 0, \quad k = 1, \ldots, m. \end{array}$$