AM 221: Advanced Optimization	Spring 2018
Kojin Oshiba	*Problem Set 3 — Due Wed, Feb. 14th at 23:59*

Instructions: All your solutions should be prepared in LaTeX and the PDF and .tex should be submitted to Canvas. Please submit all your files as ONE archive of filetype zip, tgz, or tar.gz. For each question, a well-written and correct answer will be selected a sample solution for the entire class to enjoy. If you prefer that we do not use your solutions, please indicate this clearly on the first page of your assignment.

The programming parts can be written in Python, Matlab, or Julia. If you strongly wish to use another language, please contact the instructor to ask for permission.

1. Unboundedness. Let us consider the polytope $P = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} \leq \mathbf{b}\}$ for some $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. In this problem, we are interested in the following linear program:

$$\max_{\mathbf{x}\in P} \mathbf{c}^\intercal \mathbf{x} \tag{LP}$$

We define the *recession cone* P^o associated with P by:

$$P^o \stackrel{\text{def}}{=} \{\mathbf{d} \in \mathbb{R}^n \mid \forall \mathbf{x} \in P, \forall\lambda \geq 0,\ \mathbf{x} + \lambda\mathbf{d} \in P\}$$

a. Show that $P^o = \{\mathbf{d} \in \mathbb{R}^n \mid A\mathbf{d} \leq 0\}$.

b. Show that P^o is a convex set.

c. Show that the linear program (LP) above is unbounded if and only if there exists $\mathbf{d} \in P^o$ such that $\mathbf{c}^\intercal\mathbf{d} > 0$.

a.

$$\begin{aligned}
&\forall \mathbf{x} \in P, \forall\lambda \geq 0, \mathbf{x} + \lambda\mathbf{d} \in P \\
\Leftrightarrow &\forall \mathbf{x} \in P, \forall\lambda \geq 0, A(\mathbf{x} + \lambda\mathbf{d}) \leq \mathbf{b} \ (\because \text{definition of P}) \\
\Leftrightarrow &\forall \mathbf{x} \in P, \forall\lambda > 0, A(\mathbf{x} + \lambda\mathbf{d}) \leq \mathbf{b} \ (\because \text{when } \lambda = 0,\ A\mathbf{x} \leq \mathbf{b} \text{ is satisfied by default beucase } \mathbf{x} \in P) \\
\Leftrightarrow &\forall \mathbf{x} \in P, \forall\lambda > 0, A\mathbf{d} \leq \frac{\mathbf{b} - A\mathbf{x}}{\lambda} \\
\Leftrightarrow &A\mathbf{d} \leq 0
\end{aligned}$$

The last equivalence comes from the fact that $\mathbf{b} - A\mathbf{x} \geq 0$, since $\mathbf{x} \in P$. So, for $A\mathbf{d} \leq \frac{\mathbf{b}-A\mathbf{x}}{\lambda}, \forall \mathbf{x} \in P, \forall\lambda > 0$, we need $A\mathbf{d} \leq 0$ because the lower bound of $\frac{\mathbf{b}-A\mathbf{x}}{\lambda}$ is 0.

b. Let $d_1 \in P_0$ and $d_2 \in P_0$. Then,

$$\begin{aligned}A(\lambda \mathbf{d}_1 + (1-\lambda)\mathbf{d}_2) &= \lambda A\mathbf{d}_1 + (1-\lambda)A\mathbf{d}_2 \\ &\leq \lambda 0 + (1-\lambda)0 \ (\because \ (a)) \\ &= 0\end{aligned}$$

c. Let (i) $max_{\mathbf{x}\in P}\mathbf{c}^T\mathbf{x}$ is unbounded, (ii) $\exists \mathbf{d} \in P^0$ s.t. $\mathbf{c}^T\mathbf{d} > 0$.

Proof of (i) $\Rightarrow$ (ii):

$$\begin{aligned}
&\quad max_{\mathbf{x}\in P}\mathbf{c}^T\mathbf{x} \text{ is unbounded}\\
&\Leftrightarrow min_{\mathbf{x}\in\mathbb{R}^n}\mathbf{c}^T\mathbf{x} \text{ s.t. } A\mathbf{x} \geq \mathbf{b} \text{ is unbounded}\\
&\Leftrightarrow \text{the dual of this LP } max_{\mathbf{y}\in\mathbb{R}^m}\mathbf{b}^T\mathbf{y} \text{ s.t. } A^T\mathbf{y} = c, y \geq 0 \text{ is infeasible } (\because \text{ weak duality})\\
&\Rightarrow \{\mathbf{y} \in \mathbb{R}^m : A^T\mathbf{y} = \mathbf{c}, y \geq 0\} = \phi\\
&\Rightarrow \ max_{\mathbf{y}\in\mathbb{R}^m}\mathbf{0}^T\mathbf{y} \text{ s.t. } A^T\mathbf{y} = c, y \geq 0 \text{ is infeasible}\\
&\Rightarrow \text{the dual of this LP } min_{\mathbf{x}\in\mathbb{R}^n}\mathbf{c}^T\mathbf{x} \text{ s.t. } A\mathbf{x} \geq \mathbf{0} \text{ is unbounded}\\
&\quad (\because \text{ weak duality. But since } \mathbf{x} = 0 \text{ is a feasible solution, it is unbounded})\\
&\Rightarrow \ max_{\mathbf{x}\in\mathbb{R}^n}\mathbf{c}^T\mathbf{x} \text{ s.t. } A\mathbf{x} \leq \mathbf{0} \text{ is unbounded}\\
&\Rightarrow \ min_{\mathbf{d}\in P^0}\mathbf{c}^T\mathbf{d} \text{ is unbounded}\\
&\Rightarrow \exists \mathbf{d} \in P^0 \text{ s.t. } \mathbf{c}^T\mathbf{d} > 0
\end{aligned}$$

Proof of (ii) $\Rightarrow$ (i):

$$\begin{aligned}
&\exists \mathbf{d} \in P^0 \text{ s.t. } \mathbf{c}^T\mathbf{d} > 0\\
&\Rightarrow \forall \mathbf{x} \in P, \forall \lambda \geq 0, \mathbf{x} + \lambda\mathbf{d} \in P \text{ and } \mathbf{c}^T\mathbf{d} > 0\\
&\Rightarrow \mathbf{c}^T(\mathbf{x} + \lambda\mathbf{d}) \text{ can be arbitrarily large by choosing } \lambda \text{ arbitrarily large.}\\
&\Rightarrow max_{\mathbf{x}\in P}\mathbf{c}^T\mathbf{x} \text{ is unbounded}
\end{aligned}$$

2. Linearly Separable Datasets. In linear classification, we are given a dataset $\mathcal{D} = \big\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\big\}$ with $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{-1, +1\}$. y_i is the *label* of data point i. Remember that we defined a dataset to be (strictly) linearly separable if and only if there exists $\mathbf{w} \in \mathbb{R}^d$ and $b \in \mathbb{R}$ such that:

$$y_i\big(\mathbf{w}^\intercal x_i - b\big) > 0, \quad 1 \leq i \leq n$$

a. Show that the condition of being linearly separable is equivalent to the following condition: there exists $\mathbf{w} \in \mathbb{R}^d$ and $b \in \mathbb{R}$ such that

$$y_i\big(\mathbf{w}^\intercal x_i - b\big) \geq 1, \quad 1 \leq i \leq n$$

b. Let us define $X^+ \stackrel{\text{def}}{=} \{\mathbf{x}_i \,|\, y_i = +1, \ 1 \leq i \leq n\}$ and $X^- \stackrel{\text{def}}{=} \{\mathbf{x}_i \,|\, y_i = -1, \ 1 \leq i \leq n\}$. Using Farkas' lemma, show that if $\mathcal{D}$ is not linearly separable then $C(X^+) \cap C(X^-) \neq \emptyset$. Remember from the second problem set that $C(X)$ denotes the convex hull of X.

a. Let $c = min\{b' \in \mathbb{R} : b + 1 < b'\}$

$$\begin{aligned}
&\exists \mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R} \text{ s.t. } y_i(\mathbf{w}^\intercal x_i - b) > 0, \quad 1 \leq i \leq n \\
\Leftrightarrow &\exists \mathbf{w} \in \mathbb{R}^d, c \in \mathbb{R} \text{ s.t. } y_i(\mathbf{w}^\intercal x_i - (c-1)) \geq 0, \quad 1 \leq i \leq n \\
\Leftrightarrow &\exists \mathbf{w} \in \mathbb{R}^d, c \in \mathbb{R} \text{ s.t. } y_i(\mathbf{w}^\intercal x_i - c) \geq 1, \quad 1 \leq i \leq n
\end{aligned}$$

b. We show that the contraposition is true. To show this, we use the Seperation Theorem I, referenced here. We first show that $C(X^+)$ is compact. $C(X^+)$ is closed by definition of a convex hull. Furthermore, since there are only finitely many points in X^+, $C(X^+)$ is also bounded. Hence, $C(X^+)$ is compact. Therefore, it follows from the Seperation Theorem I that there exist a nonzero vector v and real numbers $c_1 < c_2$ such that $\langle x, v \rangle > c_2$ and $\langle y, v \rangle < c_1$ for all $x \in C(X^+)$ and $y \in C(X^-)$.

3. Fractional Knapsack Problem. In the Fractional Knapsack Problem, there is a set of n items, each item i, $1 \leq i \leq n$ has a value $v_i \in \mathbb{R}_+$ (representing your "happiness" for owning this item) and a cost $c_i \in \mathbb{R}_+$. There is a budget constraint $b \in \mathbb{R}_+$ on the total amount of money you can spend and your goal is to buy the set of items of maximum value while not spending more than b. We assume that the items are infinitely divisible, meaning that you can buy a fraction x_i, $0 \leq x_i \leq 1$ of item i for a fraction $x_i c_i$ of its cost, but this will only give you a fraction $x_i v_i$ of its value.

Formally, we want to solve the following linear program:

$$\begin{aligned}
\max_{\mathbf{x} \in \mathbb{R}^n} &\sum_{i=1}^{n} v_i x_i \\
\text{s.t} &\sum_{i=1}^{n} c_i x_i \leq b \\
&x_i \geq 0, \ 1 \leq i \leq n \\
&x_i \leq 1, \ 1 \leq i \leq n
\end{aligned}$$

We will also assume that all items have positive value, $i.e$ $v_i > 0$ for all i (otherwise we can remove these items without changing the problem, since we will never want to buy them anyway).

a. Let us consider a feasible solution $\mathbf{x} \in \mathbb{R}^n$ of the Fractional Knapsack Problem. Show that $\mathbf{x} \in \mathbb{R}^n$ is optimal if and only if there exists $\mathbf{y} \in \mathbb{R}^n$ and $\xi \in \mathbb{R}$ such that:

$$\begin{gathered}
\xi \geq 0, \ \mathbf{y} \geq 0 \\
c_i \xi + y_i \geq v_i, \ 1 \leq i \leq n \\
y_i > 0 \Rightarrow x_i = 1, \ 1 \leq i \leq n \\
x_i > 0 \Rightarrow c_i \xi + y_i = v_i, \ 1 \leq i \leq n \\
\xi > 0 \Rightarrow \sum_{i=1}^{n} c_i x_i = b
\end{gathered}$$

We will now focus on characterizing an optimal solution $\mathbf{x} \in \mathbb{R}^n$. Let us consider one such solution $\mathbf{x} \in \mathbb{R}^n$. **Warning:** this problem is decomposed in many small questions, but most questions should only take a few lines to solve.

b. Show that when $\frac{v_i}{c_i} > \xi$ or $c_i = 0$ then $x_i = 1$.

c. Show that when $c_i \neq 0$ and $\frac{v_i}{c_i} < \xi$ then $x_i = 0$.

For simplicity, we will assume that the ratios $\frac{v_i}{c_i}$ are distinct across all items. In other words, for distinct indices $i \neq j$, we have that $\frac{v_i}{c_i} \neq \frac{v_j}{c_j}$. This is not a very restrictive assumption, and the analysis which follows can be adapted to accommodate ties.

d. Let us denote by I the set of indices such that $\frac{v_i}{c_i} > \xi$, and let us assume that there exists j such that $\frac{v_j}{c_j} = \xi$; if such a j exists, it is necessarily unique. Show that if $\sum_{i \in I} c_i + c_j \leq b$ then $x_j = 1$.

e. Assume that $\sum_{i \in I} c_i + c_j > b$ (j is the same index as in part d.), show that:

$$x_j = \frac{b - \sum_{i \in I} c_i}{c_j}$$

f. Combining parts a. to f. explain how to construct an optimal solution to the Fractional Knapsack Problem.

a. Let $A = \begin{pmatrix} -\mathbf{c}^T \\ I \\ -I \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} b \\ \underline{0} \\ \underline{-1} \end{pmatrix}$, $\mathbf{c}' = -\mathbf{v}^T$, $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

Then the primal LP can be written as:

$min\ \mathbf{c}'^T\mathbf{x}$, subject to $A\mathbf{x} \geq \mathbf{b}, \mathbf{x} \geq 0$.

Then, the dual LP can be written as:

$max\ \mathbf{b}^T\mathbf{y}'$ subject to $A^T\mathbf{y}' = \mathbf{c}', \mathbf{y}' \geq 0$.

Let $\mathbf{y}' = \begin{pmatrix} \xi \\ z_1 \\ \vdots \\ z_n \\ y_1 \\ \vdots \\ y_n \end{pmatrix}$.

Directly from the dual condition, we have

$$\mathbf{y} > 0, \xi > 0$$

For $1 \leq i \leq n$, we have from $A^T\mathbf{y}' = \mathbf{c}'$ that

$$\begin{aligned}
&- c_i\xi + z_i - y_i = -v_i \\
&\Leftrightarrow c_i\xi + y_i - z_i = v_i \\
&\Leftrightarrow c_i\xi + y_i \geq v_i
\end{aligned}$$

From both weak and strong duality, we have

$$-\mathbf{v}^T\mathbf{x} = \mathbf{y}^T A\mathbf{x} = \mathbf{b}^T\mathbf{y}$$

Hence,

$$\mathbf{y}^T A\mathbf{x} = \mathbf{b}^T\mathbf{y}$$

This means, if $\mathbf{y}' > 0$, $A\mathbf{x} = \mathbf{b}$. Thus,

if $\xi > 0$, $\sum_{i=1}^{n} c_i x_i = b$ and

if $y_i > 0$, $x_i = 1$ for $1 \leq i \leq n$.

Finally, we show the remaining $x_i > 0 \Rightarrow c_i\xi + y_i = v_i$. Let's use the contrapositive of the complementary slackness. First of all, we have complementary slackness from the strong duality since there exists x that is feasible in (P) and so it must be the case that there's a y feasible for (D) with $c^T x = b^T y$. Then, we have that $\forall i\ y_i' > 0 \Rightarrow a_i^T x = b_i^T$. The contrapositive of this is $\forall i\ a_i^T x > b_i^T \Rightarrow y_i' = 0$. Applying this for the a_i's corresponding to z's in y', we have $x_i > 0 \Rightarrow z_i = 0$ for all z_i's. Because we have $c_i\xi + y_i - z_i = v_i$ as we've seen, it follows that $c_i\xi + y_i = v_i$.

Since complementary slackness is an if and only if result, the proofs above will hold even in the converse case. Therefore, we proved the proposition.

b.

$$\begin{aligned}
&\frac{v_i}{c_i} > \xi \text{ or } c_i = 0 \\
\Rightarrow &y_i > 0 \\
&(\because c_i\xi + y_i \geq v_i. \text{ So when } c_i = 0,\ y_i \geq v_i > 0 \text{ and when } \frac{v_i}{c_i} > \xi,\ y_i > 0\ ) \\
\Rightarrow &x_i = 1
\end{aligned}$$

c.

$$\begin{aligned}
&c_i \neq 0, \frac{v_i}{c_i} < \xi \\
&\Rightarrow v_i < c_i\xi \text{ and } c_i \neq 0 \\
&\Rightarrow v_i < c_i\xi + y_i \text{ and } c_i \neq 0\ (\because\ y_i \geq 0) \\
&\Rightarrow c_i\xi + y_i \neq v_i \\
&\Rightarrow x_i = 0\ (\because \text{ contraposition of } x_i > 0 \Rightarrow c_i\xi + y_i = v_i)
\end{aligned}$$

d. Since we assume that there exists i such that $\xi = \frac{v_i}{c_i}$ and $v_i > 0$ always, it must be the case that $\xi > 0$. Hence, we assume that for the remaining of the problem. Using the results of b and c, we have,

$$\sum_{i=1}^{n} c_i x_i = \sum_{i \in I} c_i x_i + c_j x_j + \sum_{k \in K} c_k x_k = \sum_{i \in I} c_i + c_j x_j$$

On the other hand, we assumed that $\sum_{i\in I} c_i + c_j \leq b$. Moreover, we have $b = \sum_{i=1}^n c_i x_i$ from $\xi > 0$. So,

$$b = \sum_{i\in I} c_i + c_j x_j \leq \sum_{i\in I} c_i + c_j \leq b$$

Hence, $\sum_{i\in I} c_i + c_j x_j = b$ and so $x_j = 1$

e. We can again assume that $\xi > 0$. We have that if $\xi > 0$, $\sum_{i=1}^n c_i x_i = b$.

   Similarly to d, using the results from b and c, $\sum_{i\in I} c_i + c_j x_j = b$

   Hence, $x_j = \frac{b - \sum_{i\in I} c_i}{c_j}$.

f. First, sort items by $\frac{v_i}{c_i}$ for all $1 \geq i \geq n$ in descending order. Then, buy the entirety of each item, starting from the ones with high $\frac{v_i}{c_i}$, until you can't buy the entirety of items with lower $\frac{v_i}{c_i}$. Finally buy the fraction of the item with the next highest $\frac{v_i}{c_i}$ so as to use up the rest of the remaining budget.

4. Predicting wine quality. In this problem we will use Linear Programming to predict wine quality (as judged by oenologists) from chemical measurements. The dataset is available at `http://rasmuskyng.com/am221_spring18/psets/hw3/wines.csv`. In each line, the first 11 columns contain the results from various chemical tests performed on the wine, and the last column is the evaluation (a score between 0 and 10) of the wine.

For wine sample i, let us denote by $y_i \in \mathbb{R}$ its score and by $\mathbf{x}_i \in \mathbb{R}^{11}$ its chemical properties. We will construct a linear model to predict y_i as a function of $\mathbf{x}_i$, that is, we want to find $\mathbf{a} \in \mathbb{R}^{11}$ and $b \in \mathbb{R}$ such that:

$$y_i \simeq \mathbf{a}^\intercal \mathbf{x}_i + b$$

The quality of the model will be evaluated using the ℓ_1 norm, *i.e* we want to find a solution to this optimization problem:

$$\min_{\substack{\mathbf{a}\in\mathbb{R}^{11}\\ b\in\mathbb{R}}} \frac{1}{n}\sum_{i=1}^n |y_i - \mathbf{a}^\intercal \mathbf{x}_i - b|$$

a. Remember from class that the above problem is equivalent to the following linear program:

$$\begin{aligned} \min_{\substack{\mathbf{a}\in\mathbb{R}^{11}\\ b\in\mathbb{R}\\ \mathbf{z}\in\mathbb{R}^n}} \; & \frac{1}{n}\sum_{i=1}^n z_i \\ \text{s.t. } & z_i \geq y_i - \mathbf{a}^\intercal \mathbf{x}_i - b, 1 \leq i \leq n \\ & z_i \geq \mathbf{a}^\intercal \mathbf{x}_i + b - y_i, 1 \leq i \leq n \end{aligned}$$

   Explain how to rewrite this problem in matrix form:

$$\begin{aligned} \min_{\mathbf{d}\in\mathbb{R}^{12+n}} \; & \mathbf{c}^\intercal \mathbf{d} \\ \text{s.t. } & A\mathbf{d} \leq \mathbf{b} \end{aligned}$$

   In particular, give the dimensions and definitions of $\mathbf{c}$, A and $\mathbf{b}$.

b. Use an LP solver (we recommend using CVXOPT in Python, CVX in Matlab or JuMP & Clp in Julia) to solve the above problem and report your code as well as the optimal value of the problem. Note that the value of the problem is exactly the average absolute error of the linear model on the dataset. Does it seem to be within an acceptable range?

a. Let $A = \begin{pmatrix} \mathbf{x}_1^T & 1 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ \mathbf{x}_n^T & 1 & 0 & 0 & \dots & -1 \\ -\mathbf{x}_1^T & -1 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ \mathbf{x}_n^T & -1 & 0 & 0 & \dots & -1 \end{pmatrix}$.

A is a $2n \times 12 + n$ matrix, where the first 12 columns are $\mathbf{x}^T$ and $-\mathbf{x}^T$ stacked vertically, and the next column is n 1's and n -1's stacked vertically, and the last n columns are two $-I$ (I is an identity matrix) stacked vertically.

$\mathbf{b} = \begin{pmatrix} \mathbf{y} \\ -\mathbf{y} \end{pmatrix}$, which is a vector of length $2n$.

$\mathbf{c} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix}$, which is a vector of length $12 + n$ and the first 12 elements are 0, the rest are $\frac{1}{n}$.

$\mathbf{d} = \begin{pmatrix} \mathbf{a} \\ b \\ \mathbf{z} \end{pmatrix}$, which is a vector of length $12 + n$, where $\mathbf{a}, b, \mathbf{z}$ are stacked vertically. .

b. The full code is provided in The optimal cost of the primal is 0.49375, which seems to be reasonable.