AM 221: Advanced Optimization	Spring 2018
Dr. Rasmus Kyng	*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 $P = \{\mathbf{x} \in \mathbb{R}^n \,|\, A\mathbf{x} \leq \mathbf{b}\}$ 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:

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

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

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 $\mathbf{c}^\intercal \mathbf{d} > 0$.

2. Linearly Separable Datasets. In linear classification, we are given a dataset $\mathcal{D} = \big\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\big\}$ 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.

3. Fractional Knapsack Problem. In the Fractional Knapsack Problem, there is a set of n items, each item i, $1 \le i \le n$ 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, $0 \le x_i \le 1$ 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:

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

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:

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

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 $\sum_{i \in I} c_i + c_j \le b$ 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.

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?