EECS 127/227AT **Fall 2020**
Homework 5
Homework 5 is due on Gradescope by Friday 10/09 at 11.59 p.m.

1 A combinatorial problem formulated as a convex optimization problem.

In this question we will see how combinatorial optimization problems can also sometimes be solved via related convex optimization problems. We also saw this in a different context in problem 5 on Homework 3 when we related λ_2 to $\phi(G)$ for a graph. This problem does not have any connection with the earlier problem in Homework 3. The relevant sections of the textbooks are Sec. 8.3 of the textbook of Calafiore and El Ghaoui and Secs. 4.1 and 4.2 (except 4.2.5) of the textbook of Boyd and Vandenberghe.

We have to decide which items to sell among a collection of n items. The items have given selling prices $s_i > 0$, $i = 1, \ldots, n$, so selling the item i results in revenue s_i. Selling item i also implies a transaction cost $c_i > 0$. The total transaction cost is the sum of transaction costs of the items sold, plus a fixed amount, which we normalize to 1.

We are interested in the following decision problem: decide which items to sell, with the aim of maximizing the *margin*, i.e. the ratio of the total revenue to the total transaction cost.

Note that this is a combinatorial optimization problem, because the set of choices is a discrete set (here it is the set of all subsets of $\{1, \ldots, n\}$). In this question we will see that, nevertheless, it is possible to pose this combinatorial optimization problem as a convex optimization problem.

(a) Show that the original combinatorial optimization problem can be formulated as

$$\max_{x \in \{0,1\}^n} f(x)$$

where $f(x) := \frac{s^T x}{1 + c^T x}$, Here $s \in \mathbb{R}^n$ is the column vector of revenues and $c \in \mathbb{R}^n$ is the column vector of transaction costs associated to the individual items.

(b) Suppose you are starting a small shop in your hometown to sell some items, stocking it from the local supermarket, and deciding which items to stock based on the criterion of maximizing your margin. You would therefore like to solve the combinatorial optimization problem in the first part of this problem. According to Consumer Reports, a typical supermarket stocks about 45000 different items, so you can assume that $n = 45000$. The number of subsets of a set of size n is 2^n. Using the approximation $2^{10} \approx 10^3$, get an approximation for the integer k such that $2^{45000} \approx 10^k$. Given your result, do you think it'd be feasible to solve this problem via "brute force", that is, by checking the objective function value for every different $x \in \{0, 1\}^n$ to determine the optimum?

Remark: There are roughly 10^{80} atoms in the known universe.

(c) Show that the combinatorial optimization problem admits a convex formulation in the sense that the value of the convex optimization problem

$$\begin{aligned} \min_{t} \quad & t \\ \text{subject to :} \quad & t \geq \mathbb{1}^T (s - tc)_+, \\ & t \geq 0, \end{aligned}$$

is the same as the value of the original combinatorial optimization problem.

Here z_+, for a vector $z \in \mathbb{R}^n$, denotes the vector with components $\max(0, z_i)$, $i = 1, \ldots, n$. Also $\mathbb{1}$ denotes the all-ones column vector of length n.

Note: You should also justify why the problem formulated in this part of the question is a convex optimization problem.

Hint: For given $t \geq 0$, express the condition $f(x) \leq t$ for every $x \in \{0, 1\}^n$ in simple terms. Note that this is related to the idea of introducing slack variables, as in Sec. 8.3.4.4 of the textbook of Calafiore and El Ghaoui.

(d) Show that the inequality constraint $t \geq \mathbb{1}^T(s - ct)_+$ is active at the optimum in the convex optimization problem, namely, it will have to be satisfied with equality.

(e) How can you recover an optimal solution x^* for the original combinatorial optimization problem from an optimal value t^* for the above convex optimization problem?

2 Optimization over a Polytope

Consider an optimization problem with an affine objective with the feasible set being a convex polyhedron $\mathcal{P}$ in $\mathbb{R}^n$. Namely, consider the problem

$$\min_{x \in \mathcal{P}} \left(c^T x + b\right)$$

Recall that a polyhedron, by definition, is a subset of Euclidean space defined as the intersection of a finite number of half-spaces. To get the correct mental picture, recall that every affine subspace can be written as an intersection of half-spaces, so in general the affine hull of a polyhedron in $\mathbb{R}^n$ can have a lower dimension than n (the dimension of a polyhedron is, by definition, the dimension of its affine hull). Every polyhedron is a closed convex set. A polytope is, by definition, a bounded polyhdedron.

The relevant portions of the textbooks for this problem are Sec. 8.1 and Sec. 8.3 (up to Sec. 8.1.4) of the textbook of Calafiore and El Ghaoui.

(a) Prove that the optimal solution, if any exists, lies on the boundary of the feasible set.

Hint: In the case where the feasible set has a nonempty interior, you could try a proof by contradiction starting with the assumption that an optimal solution lies in the interior.

(b) Suppose now that $\mathcal{P}$ is a polytope.

i. Prove that the optimal value for the objective is attained at a vertex of the polytope,

Remark: Recall that a *vertex* of a polyhedron is, by definition, one of its extreme points. Here, an *extreme point* of a convex set $K \subseteq \mathbb{R}^n$ is any point $x \in K$ such that if we can write $x = \lambda x_1 + (1-\lambda)x_2$ for some $x_1, x_2 \in K$ and $\lambda \in (0,1)$, then we must have $x_1 = x_2$. Namely it is a point $x \in K$ that cannot be expressed in a nontrivial way as a convex combination of two distinct points of K. You can check that this way of defining the vertex of a polytope matches your intuitive geometric understanding of what a vertex of a polytope is.

You can check for yourself that every polyhedron has only finitely many vertices (possibly none). Further, you can check for yourself that every polytope is the convex hull of its vertices (this is not true for a polyhedron that is not a polytope). We will accept these as facts for this problem.

ii. Show that if the optimal value of the objective is attained at multiple vertices, then it is also achieved at any point in their convex hull.

(c) Consider the optimization problem:

$$\begin{aligned} \min\ & ax + by + c \\ \text{subject to}\ & x \geq 0, \\ & y \geq 0, \\ & x + 2y \leq 6, \\ & 5x + 2y \leq 10. \end{aligned}$$

The feasible set is the closed bounded set enclosed by the red lines as given in Fig. 1.

Find the optimal solutions for objectives (i) $-2x+3y+5$, (ii) $-x-2y+5$ and (iii) 5 by finding the values at the corners of the feasible set and then using the *cvxpy* package in this jupyter notebook.

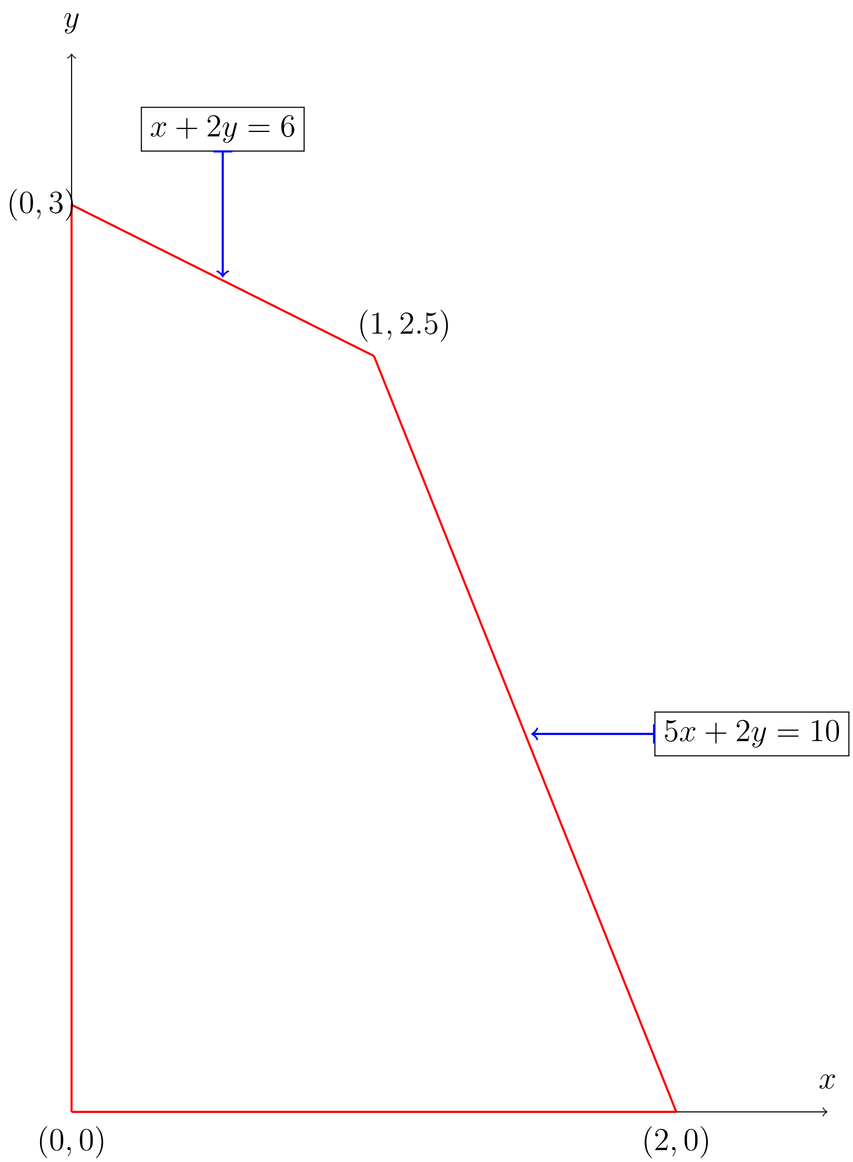

Figure 1: The feasible set for problem in 2(c) is the polytope bounded by the red lines.

Please refer to `https://www.cvxpy.org/tutorial/index.html` to gain some familiarity with `cvxpy` which we may use for additional problems later in the course and which is a widely used package for convex optimization.

3 Fenchel Conjugates

For a function $f : \mathbb{R}^n \to \mathbb{R}$, not necessarily a convex function, with a domain $\mathrm{dom}(f)$, which we assume to be nonempty, but not necessarily a convex set, we can define its *conjugate* (also called its *convex conjugate*, *Fenchel conjugate* or *Legendre-Fenchel conjugate*), $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ via the rule

$$f^*(z) := \sup_{x \in \mathrm{dom}(f)} \left(z^T x - f(x)\right).$$

Note that f^* is an extended real valued function and does not take the value $-\infty$. Also note that it is convenient to treat f also as an extended real valued function, taking the value ∞ outside $\mathrm{dom}(f)$, and with this viewpoint we can also write

$$f^*(z) = \sup_{x \in \mathbb{R}^n} \left(z^T x - f(x)\right). \tag{1}$$

Note that, as an extended real valued function, f also does not take the value $-\infty$.

The relevant portion of the textbooks for this problem is Sec. 3.3 of the textbook of Boyd and Vandenberghe.

(a) Suppose you are manufacturing and selling n products labeled $1, \ldots, n$. You can sell your products in the form of a basket containing x_i amount of product i, getting revenue $f(x)$. Let the cost of making product i be y_i per unit. Assuming that each product is infinitely divisible (i.e. not restricted to be in discrete units), show that the optimal amount of profit you can make is $(-f)^*(-y)$. Here the profit is the excess of revenue over costs.

(b) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a function with $\mathrm{dom}(f) = \mathbb{R}^n$ such that $f^* = f$.

Show that $f(x) = \frac{1}{2}\|x\|_2^2$.

(c) Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = e^x$, with $\mathrm{dom}(f) = \mathbb{R}$.

  i. Find the conjugate f^* of f.
  Note: Part of finding f^* is to find $\mathrm{dom}(f^*)$.

  ii. Find f^{**}.

(d) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the indicator function of the rectangle $[-1, 1] \times [-1, 1]$. Namely,

$$f(x_1, x_2) = \begin{cases} 0 & \text{if } -1 \leq x_1 \leq 1 \text{ and } -1 \leq x_2 \leq 1, \\ \infty & \text{otherwise.} \end{cases}$$

Note that $\mathrm{dom}(f) = [-1, 1] \times [-1, 1]$.

Find the conjugate f^*.

4 Perspective function

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function, with $\mathrm{dom}(f) = \mathbb{R}^n$.

The relevant portion of the textbooks for this problem is Sec. 8.2 of the textbook of Calafiore and El Ghaoui and Secs. 3.1 and 3.2 of the textbook of Boyd and Vandenberghe.

(a) Show that we can represent f as a pointwise maximum of affine functions, specifically that for all $x \in \mathbb{R}^n$ we have

$$f(x) = \max_{(a,b)\in\mathcal{A}} \left(a^T x + b\right), \tag{2}$$

where

$$\mathcal{A} := \left\{(a,b) \in \mathbb{R}^n \times \mathbb{R} \ : \ a = \nabla f(y), \ \ b = f(y) - y^T \nabla f(y), \ \ y \in \mathbb{R}^n\right\}.$$

(b) Given a function $f : \mathbb{R}^n \to \mathbb{R}$, recall that the *convex conjugate*, also known as the *Fenchel conjugate*, of f is another function $f^* : \mathbb{R}^n \to \mathbb{R}$ such that

$$f^*(y) = \sup_{x\in\mathrm{dom}(f)} (y^\top x - f(x)).$$

Suppose now that $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function, with $\mathrm{dom}(f) = \mathbb{R}^n$, as in the preceding part of the problem. Show that, for all $y \in \mathbb{R}^n$, we have

$$f^*(\nabla f(y)) = y^T \nabla f(y) - f(y).$$

Remark: The interpretation is that the maximization problem used to define the value of the conjugate function at $z = \nabla f(y)$ is achieved at y. This is an extremely important mental picture to have when thinking about the conjugate function. See Figure 2 for an illustration of what is going on in the construction of the convex conjugate of a function.

(c) We define the *perspective* of a function $f : \mathbb{R}^n \to \mathbb{R}$ with domain $\mathrm{dom}(f) \subseteq \mathbb{R}^n$ as the function $g : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ with domain $\{(x,t) \ : \ x/t \in \mathrm{dom}(f), t > 0\}$, with values

$$g(x,t) = \begin{cases} tf(x/t) & \text{if } x/t \in \mathrm{dom}(f), t > 0, \\ \infty & \text{otherwise.} \end{cases}$$

Assume as in the first part of this problem that f is a differentiable convex function with $\mathrm{dom}(f) = \mathbb{R}^n$. Prove that g is convex.

Hint: Use equation (2).

(d) Show that the function $h : \mathbb{R}^n \to \mathbb{R}$ with values

$$h(x) = \begin{cases} \dfrac{(a^T x + b)^2}{c^T x + d} & \text{if } c^T x + d > 0, \\ \infty & \text{otherwise,} \end{cases}$$

is convex.

Hint: Consider the perspective function of the function $f : \mathbb{R} \to \mathbb{R}$ with domain $\mathbb{R}$ given by $f(z) = z^2$.

Remark: The perspective of a function shows up in many applications including statistics and control. A prominent application is in computer vision. One can define a function $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$

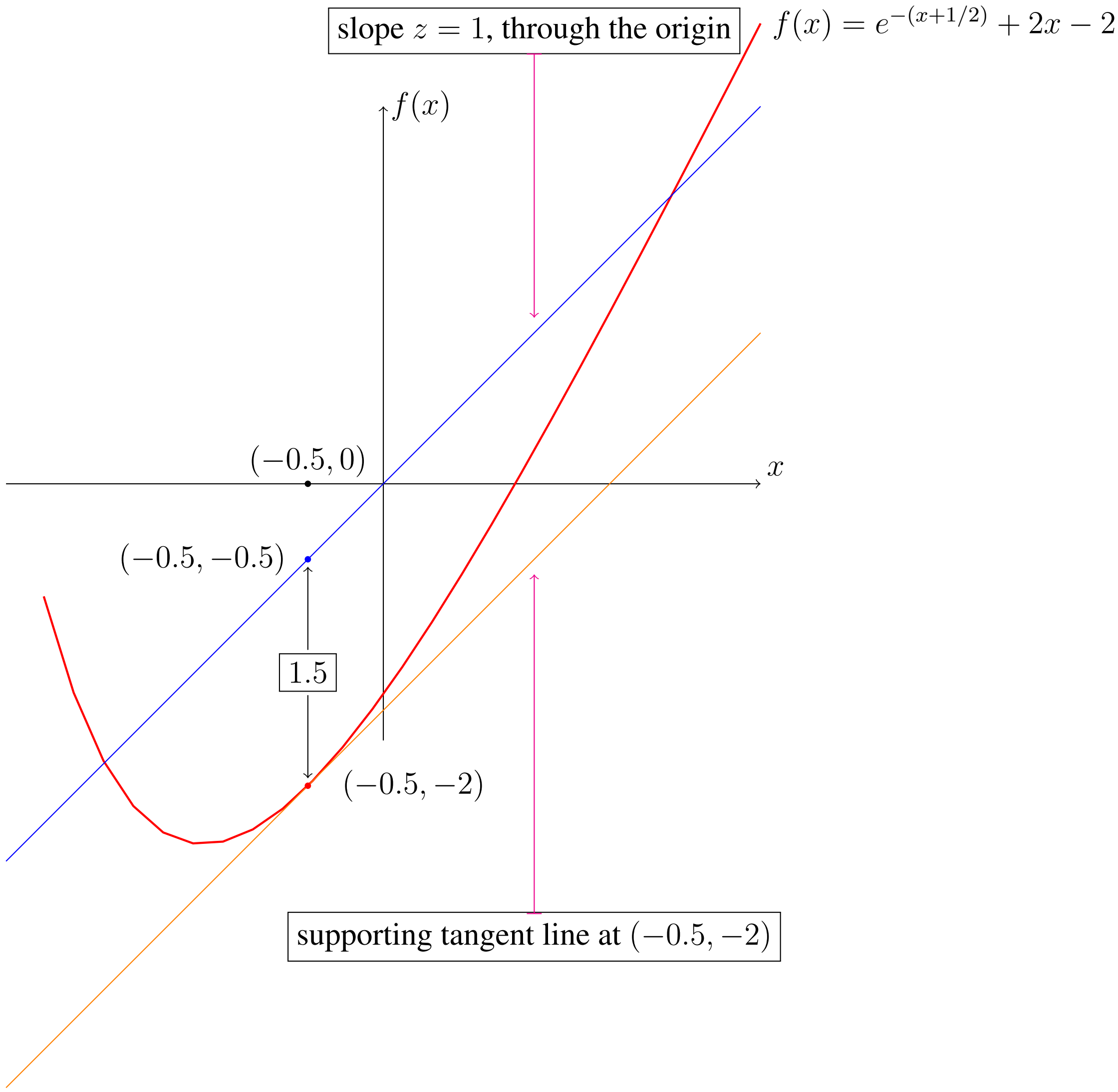

Figure 2: This figure illustrates the construction of the conjugate of a convex function f. The function given is $f(x) = e^{-(x+1/2)} + 2x - 2$, plotted in red. We want to evaluate $f^*(1)$. We consider the line of slope 1 through the origin, namely the graph of x, illustrated in blue. $x - f(x)$ is maximized at $x = -0.5$, where $f(-0.5) = -2$ and $-0.5 - f(-0.5) = 1.5$, as illustrated. Thus $f^*(1) = 1.5$. When the line of slope 1 through the origin is translated down by $f^*(1)$, we get the orange line, which is the supporting tangent line to the graph of f at $(-0.5, -2)$, as illustrated. If f were not differentiable at the abscissa of the point on its graph through which the orange line passes, the orange line would have been one of the subgradient lines to the graph of f at that point.

with domain $\mathbb{R}^n \times \mathbb{R}_{++}$ given by $P(z,t) = z/t$, called the *perspective function*. If one imagines looking at $\mathbb{R}^n \times \mathbb{R}_{++}$ from the origin in $\mathbb{R}^{n+1}$, i.e. drawing a ray from the origin through any given point $(z,t) \in \mathbb{R}^n \times \mathbb{R}_{++}$, the $\mathbb{R}^n$ coordinates of the intersection of that ray with the affine hyperplane $\{(z,1) \;:\; z \in \mathbb{R}^n\}$ is z/t, i.e. $P(z,t)$. It can be shown that the perspective function maps any convex subset of $\mathbb{R}^n \times \mathbb{R}_{++}$ to a convex subset of $\mathbb{R}^n$ (see Sec. 2.3.3 of the textbook of Boyd and Vandenberghe). Based on this, one can show that the perspective of any convex function $f : \mathbb{R}^n \to \mathbb{R}$ with domain $\mathrm{dom}(f)$, defined as the function $g : \mathbb{R}^{n+1} \to \mathbb{R}$ with

$$\mathrm{dom}(g) := \{(x,t) : x/t \in \mathrm{dom}(f), t > 0\},$$

is a convex function. This generalizes what was shown in part (c) of this problem. See Sec. 3.2.6 of the textbook of Boyd and Vandenberghe for more on the perspective function and on the perspective of a function.

5 Maximizing a sum of logarithms

Consider the following problem:

$$\begin{aligned} m^* := \max_{x \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} \alpha_i \log x_i \\ \text{s.t.} \quad & x \geq 0, \\ & \mathbb{1}^\top x = c, \end{aligned}$$

where $\mathbb{1}$ denotes the column vector of ones, $c > 0$ and $\alpha_i > 0$, $i = 1, \ldots, n$. In this exercise, we will determine a closed-form maximizer, and show that the optimal value of this problem is

$$m^* = \alpha \log(c/\alpha) + \sum_{i=1}^{n} \alpha_i \log \alpha_i,$$

where $\alpha := \sum_{i=1}^{n} \alpha_i$.

The relevant portion of the textbooks for this problem is the beginning of Sec. 8.5 and Secs. 8.5.1, 8.5.2 and 8.5.4 of the textbook of Calafiore and El Ghaoui and Secs. 5.1, 5.2.1 and 5.2.2 of the textbook of Boyd and Vandenberghe.

(a) First, formulate the problem as an equivalent minimization problem. Can you relax the equality constraint to an inequality while preserving the set of solutions?

(b) After relaxing the equality constraint to an inequality constraint, form the Lagrangian $\mathcal{L}(x, \mu)$ for this problem, where μ is the dual variable corresponding to the inequality constraint containing c. Do not dualize the constraint $x \geq 0$.

It is okay to not dualize the constraint $x \geq 0$ because we can treat it as implicit in defining the domain of the objective function. In fact, this sort of restricted dualization can be done in general simply by changing what we take to be the domain of the objective function, but it is particularly appropriate in this example, because the "natural" domain of the objective function in this problem, namely $\{x : x_i > 0 \text{ for all } i\}$, already satisfies the constraint $x \geq 0$. Recall from the discussion in Sec. 8.5. of the textbook of Calfiore and El Ghaoui that if there are m inequality constraints and q equality constraints which we dualize, then the Lagrangian is a function on $\mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^q$, where $\mathcal{D} \subset \mathbb{R}^n$ is the intersection of the domains of the objective and the constraint functions. In this problem we can take $\mathcal{D}$ to be $\{x : x_i > 0 \text{ for all } i\}$, and we are dualizing a single inequality constraint, i.e. we have $m = 1$ and $q = 0$ when compared to the general notation.

(c) Now derive the dual function $g(\mu)$ and solve the dual problem $d^* = \max_{\mu \geq 0} g(\mu)$. What is the optimal dual variable μ^*?

(d) Assume strong duality holds, so $p^* = d^*$. From the μ^* obtained in the previous part, how do we obtain the optimal primal variable x^*? And finally, what is the optimal objective function value p^*?