The University of Texas at Austin
Department of Electrical and Computer Engineering

EE381V-11: Large Scale Optimization — Fall 2012

Problem Set Five

Caramanis/Sanghavi Due: Thursday, October 18, 2012.

Reading Assignments

1. Reading: Boyd & Vandenberghe: Chapters 4 & 5.

Written Problems

1. For A an $m \times n$ matrix, and $b \in \mathbb{R}^m$, show that exactly one of the two following statements must hold:

   (i) There exists $x \geq 0$ such that $Ax = b$.

   (ii) There exists a vector s such that $s^\top A \geq 0$, and $s^\top b < 0$.

   (Hint: One direction is easy. For the other, think about separation arguments.)

2. **Compressed Sensing** Consider the following optimization problem

$$\begin{array}{rl} \min\limits_{x \in \mathbb{R}^n} & \|x\|_1 \\ s.t. & Ax = b \end{array}$$

   Write this as a linear program. Find its dual.

3. Problem 5.7 in the textbook, Boyd and Vandenberghe.

4. **Exponential Families** In this problem we investigate the natural motivation for an important class of distributions: exponential families. Let X be a discrete[1] random variable, with possible values $x \in \mathcal{X}$. Given a set of functions $\{\phi_k(x)\}$, the corresponding exponential family is all probability mass functions of the form

$$p(x) \;=\; \frac{1}{Z(\theta)} \exp\left(\sum_k \theta_k \phi_k(x) \right) \tag{1}$$

   where all $\theta_k \in \mathbb{R}$ and $Z(\theta)$ is a normalizing constant. Examples include bernoulli, exponential, gaussian, poisson etc.

   (a) Consider the entropy function $H(p) := -\sum_x p(x) \log p(x)$. As is well known, the higher the entropy of a random variable, the "more random" it is. Show that $H(\cdot)$ is a concave function of p.

[1] The same property holds for all random variables, but we will keep it discrete here for simplicity.

(b) Consider the following optimization problem, which maximizes entropy subject to moment constraints on certain functions:

$$\begin{aligned} \max_p \quad & H(p) \\ s.t. \quad & E_p[\phi_k(X)] = a_k \quad \text{for all } k \end{aligned}$$

where $E_p[\cdot]$ is the expectation when X has pmf p. Why is this a convex program ?

(c) Given a set of functions $\{\phi_k(x)\}$, show that the optimum of the convex program above is a pmf in the *corresponding* exponential family (1).

5. **Fast-Mixing Markov Chains** A doubly-stochastic matrix P is a symmetric matrix with non-negative entries such that every row and every column sums up to 1. Its leading eigenvalue is always 1, corresponding to the eigenvector $\mathbf{1}$. Consider the absolute values of all the other eigenvalues of P, say $\lambda_2(P) \geq \ldots \geq \lambda_n(P)$, and let $\mu(P) := \max_{i \neq 1} |\lambda_i(P)|$ denote the largest such absolute value[2].

   (a) Show that $\mu(P)$ is a convex function of P. *(Hint: $\mu(P) = \max\{\lambda_2(P), -\lambda_n(P)\}$).*

   (b) Write $\mu(P)$ as the spectral norm of P minus another matrix. (Recall: for symmetric matrices, spectral norm is the largest absolute value of an eignevalue.) *(Hint: all eigenvectors are orthogonal.)*

6. **Duality in graph theory**. Given a graph with edge weights $w_{ij} \geq 0$, the *max-weight matching* problem is: find the heaviest set of disjoint edges (i.e no two edges in the set share a node). The *min-weight vertex cover* problem is: put weights u_i on each vertex, so that *(a)* for every edge we have $w_{ij} \leq u_i + u_j$, and *(b)* the total node weights $\sum_i u_i$ is minimized. Show that these two problems are the duals of each other.

7. **Robust Optimization**. Recall the Robust Optimization framework we introduced in class. We have a linear program,

$$\begin{aligned} \min: \quad & c^\top x \\ \text{s.t.}: \quad & a_i^\top x \leq b, \quad \forall a_i \in \mathcal{U}_i, \quad i = 1, \ldots, m, \end{aligned}$$

   where $\mathcal{U}_i$ represents the uncertainty set. In class we considered polyhedral and ellipsoidal uncertainty sets. Now consider the following cardinality-constrained robust problem:[3] Each constraint, $a_i^\top x \leq b_i$, has some integer r_i of its entries that may deviate from some nominal value, while the remaining $(n - r_i)$ entries are known exactly. Thus we have:

$$\mathcal{U}_i = \{a = a_i^0 + \hat{a}_i \,:\, |\hat{a}_{ij}| \leq \Delta_{ij},\ |\text{supp}(\hat{a}_i)| \leq r_i\}.$$

   That is, $\hat{a}_i$ is non-zero on at most r_i entries. Here, $\hat{a}_{ij}$ is the j^{th} entry of the vector $\hat{a}_i$.

   Show that the robust linear program can be rewritten as a linear program. Note that you have a non-convex problem to deal with, because of the cardinality constraint. The final outcome, however, is just a linear program.

---

[2] $1 - \mu(P)$ governs the time it takes for a Markov chain, with probability transition matrix P, to converge to the unique stationary distribution $\frac{1}{n}\mathbf{1}$.

[3] As motivation, consider the following: if you are modeling measurements, it may make sense to assume that all entries may be off by some amount. But if you are modeling, say, faulty components, where something either fails or does not, it may make more sense to consider the case where at most some finite number of components fail, and the others operate perfectly.