Nonlinear Optimization (18799 B, PP)
IST-CMU PhD course
Spring 2017
Instructor: João Xavier (`jxavier@isr.ist.utl.pt`)
TA: Shanghang Zhang (`shzhang.pku@gmail.com`)

The homework is due June 9.

Homework 5

Problem A. (*Setting arc prices in a network*) For $c \in \mathbf{R}^m_+$, let

$$p_k^\star(c) := \inf\{c^T x \;:\; Ax = b_k, x \geq 0\} \tag{1}$$

for $k = 1, \ldots, K$. The matrix $A \in \mathbf{R}^{n \times m}$ and the vectors b_k, $k = 1, \ldots, K$, are given. Assume $p_k^\star(c)$ is finite for all k and c, *i.e.*, $p_k^\star(c) \in [0, \infty[$.

Let $q = (q(1), \ldots, q(K))$ be a probability mass function, i.e.,

$$q \in \Delta = \left\{x \in \mathbf{R}^K \;:\; x \geq 0, 1^T x = 1\right\}.$$

Formulate

$$\begin{array}{ll} \underset{c}{\text{minimize}} & \|c - \overline{c}\|_1 \\ \text{subject to} & \sum_{k=1}^K q(k)\, p_k^\star(c) \geq P \\ & c \geq 0, \end{array} \tag{2}$$

as a LP, QP, QCQP, SOCP or SDP. In (6), $\overline{c} \in \mathbf{R}^m_+$ and $P > 0$ are given constants.

We now give an interpretation for problem (6). You don't need the interpretation to solve the problem; it is just for your own curiosity (you can ignore the following motivation). We start by interpreting $p_k^\star(c)$ in (1): if A is the node-arc incidence matrix of a directed graph and b_k a vector with a 1 entry, a -1 entry and 0 elsewhere, we can interpret $p_k^\star(c)$ as the minimum cost of traveling from a given source to a given sink in the graph, in which c represents the arc prices. From this standpoint, we can interpret (6) as seeking the arc prices, closest to a nominal setting $\overline{c}$, that guarantee us an average profit greater or equal to P—here, $p_k^\star(c)$ is the money we get from a user k going from a certain source to a certain sink in the graph, and $q(k)$ is the probability of such an user showing up.

Formulate (6) as a LP, QP, SOCP or SDP.

Problem B. (*Dualizing a problem without constraints*) Consider the optimization problem

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}\|y - Ax\|^2 + \sum_{k=1}^K \rho_k \|G_k x\| \tag{3}$$

where y, A, G_k and $\rho_k > 0$ are given. (This problem structure arises in group sparse regularization applications.)

The goal of this problem is to dualize (3). There are no constraints in (3) to dualize, but we can consider the reformulation

$$\begin{array}{ll} \underset{x, z, w_1, \ldots, w_K}{\text{minimize}} & \frac{1}{2}\|y - z\|^2 + \sum_{k=1}^K \rho_k \|w_k\| \\ \text{subject to} & z = Ax \\ & w_k = G_k x, \quad k = 1, \ldots, K. \end{array} \tag{4}$$

Show that the dual of (4) can be written as

$$\begin{array}{ll} \underset{\lambda, \eta_1, \ldots, \eta_K}{\text{maximize}} & -\frac{1}{2}\|\lambda - y\|^2 + \frac{1}{2}\|y\|^2 \\ \text{subject to} & A^T\lambda = \sum_{k=1}^K G_k^T \eta_k, \quad k = 1, \ldots, K \\ & \|\eta_k\| \leq \rho_k, \quad k = 1, \ldots, K. \end{array}$$

Hint: the fact

$$\sup\left\{s^T x - \|x\| \, : x\right\} = \begin{cases} 0 & \text{, if } \|s\| \le 1 \\ +\infty & \text{, otherwise,} \end{cases}$$

may be helpful (if you use this fact, you must also prove it).

Problem C. (*Selecting penalization coefficients*) This problem applies the results of problem B. Consider again (3) reproduced here:

$$\underset{x}{\text{minimize}} \quad \tfrac{1}{2}\|y - Ax\|^2 + \textstyle\sum_{k=1}^{K} \rho_K \|G_k x\| \,. \tag{5}$$

In problem B, we assumed that the vector $y \in \mathbf{R}^m$, the matrices $A \in \mathbf{R}^{m\times n}$ and $G_k \in \mathbf{R}^{m_k \times n}$, and the penalization coefficients $\rho_k \ge 0$ were given when solving (5). In practice, the penalization coefficients ρ_k are often chosen through heuristics (sometimes, by trial-and-error until we "like" the solutions x that emerge from (5)). In this problem, we discuss a method to find reasonable values for the penalization coefficients (A and G_k are assumed known).

Suppose that for a given list of vectors $y_l^\star \in \mathbf{R}^m$, $l = 1, 2, \ldots, L$, we would like the corresponding solutions to be given vectors $x_l^\star \in \mathbf{R}^n$, $l = 1, 2, \ldots, L$. That is, when we take $y = y_l^\star$ in (5), the vector $x_l^\star$ should be a solution of (5).

Since there may be no choice of $\rho_1, \ldots, \rho_K$ that guarantees this property, we try instead to make each $x_l^\star$ as close to be a solution as possible. More precisely, for a fixed choice of penalization coefficients, we can assess how suboptimal is $x_l^\star$ for a given $y_l^\star$: we simply evaluate the optimality gap

$$\delta_k(\rho_1, \ldots, \rho_K) := \frac{1}{2}\|y_l^\star - Ax_l^\star\|^2 + \sum_{k=1}^{k} \rho_k \|G_k x_l^\star\| - p_l^\star(\rho_1, \ldots, \rho_K)$$

where $p_l^\star(\rho_1, \ldots, \rho_K)$ denotes the optimal value of (5) when $y = y_l^\star$. Our method consists in finding the penalization coefficients that minimize the sum of all optimality gaps:

$$\begin{array}{ll} \underset{\rho_1,\ldots,\rho_K}{\text{minimize}} & \sum_{l=1}^{L} \delta_l(\rho_1, \ldots, \rho_K) \\ \text{subject to} & \rho_k \ge 0, \quad k = 1, \ldots, K. \end{array} \tag{6}$$

Formulate (6) as a LP, QP, SOCP or SDP.