Convex Optimization

Problem set 2

Due Monday, March 29th

1. Consider the Quadratically Constrained Quadratic Program (QCQP):

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2}x'Hx + c'x \\
\text{s.t.} \quad & x'Q_ix + p_i'x + d_i \leq 0 \qquad i = 1..m \\
& A'x = b
\end{aligned}
\tag{1}
$$

where the minimization is w.r.t. $x \in \mathbb{R}^n$, and $H, Q_i \in S^n$, $c, p_i \in \mathbb{R}^n$, $d_i \in \mathbb{R}$, $A \in \mathbb{R}^{n \times p}$, $b \in \mathbb{R}^p$ are given.

   (a) What constraints must H and each Q_i satisfy for the problem to be convex?

   (b) Derive the dual of the problem.

   (c) When $Q_i = 0$ for all i, the problem is known simply as a "Quadratic Problem" (QP). By substituting $Q = 0$ in the general dual, verify the dual of a quadratic program is also a quadratic program.

   (d) Write down the QCQP as a semi-definite program (SDP), that is using only linear matrix inequality constraints and a linear objective.

2. In this problem we will investigate the problem of large margin linear classification. We are given data points $a_1, a_2, \ldots, a_m \in \mathbb{R}^d$ with binary labels $y_1, \ldots, y_m \in \pm 1$. We would like to find a (homogeneous) hyperplane separating between the positive and negative points with the largest possible margin. For a hyperplane with normal $w \in \mathbb{R}^d$, the distance between the hyperplane and a point a is $\langle w, a \rangle$—we would like all positive points to be on one side, all negative points on the other side, and to maximize the distance between the hyperplane the point closest to it. This corresponds to the following optimization problem:

$$
\begin{aligned}
\text{maximize}_{w \in \mathbb{R}^d, t \in \mathbb{R}} \quad & t \\
\text{s.t.} \quad & y_i \langle w, x_i \rangle \geq t \quad i = 1, \ldots, m \\
& \|w\| \leq 1
\end{aligned}
\tag{2}
$$

   (a) Is the problem (2) convex? Is it an LP? A QP? A QCQP? An SDP?

(b) We will use the change of variables $z = w/t$ to rewrite (2) as a quadratic program:

$$\begin{aligned} \text{minimize}_{z \in \mathbb{R}^d} \quad & \langle z, z \rangle \\ \text{s.t.} \quad & y_i \langle z, x_i \rangle \geq 1 \quad i = 1, \ldots, m \end{aligned} \tag{3}$$

State the relationship between the values of the problems (2) and (3).

(c) Does Slater's condition always hold for (3)? When does it hold?

(d) Use the KKT optimality conditions (without explicitly deriving the dual) in order to prove that the optimal z^* is always spanned by the data vectors a_i, i.e. can always be expressed as a linear combination of a_is.

(e) Use complimentary slackness to further argue that z^* can be written as a linear combination of only those points a_i which are exactly at the minimum distance from the optimal separating hyperplane.

3. Given data $a_1, \ldots, a_m \in \mathbb{R}^d$ with binary labels $y_1, \ldots, y_m \in \pm 1$, fitting a linear logistic regression model amounts to solving the following optimization problem:

$$\text{minimize}_{w \in \mathbb{R}^d} \quad \sum_{i=1}^{m} g(y_i \langle w, a_i \rangle) \tag{4}$$

where $g(\cdot)$ is the "logistic loss", and is given by $g(z) = \log(1 + e^{-z})$.

The dual of the above problem is not very interesting (what is it?). In order to be able to calculate a more meaningful dual, we will introduce new optimization variables $z \in \mathbb{R}^m$ and new equality constraints and rewrite (4) as:

$$\begin{aligned} \text{minimize} w \in \mathbb{R}^d, z \in \mathbb{R}^m \quad & \sum_{i=1}^{m} g(z_i) \\ \text{s.t.} \quad & z_i = y_i \langle w, a_i \rangle \end{aligned} \tag{5}$$

(a) Calculate the Fenchel Conjugate of the logistic loss function $g(\cdot)$.

(b) Use the equation for the conjugate of sums of independent functions to write down the conjugate of the objective function of (5).

(c) Use the Fenchel Conjugate to write down the dual of (5).

(d) Does strong duality always hold here? Explain.

Obtaining a meaningful dual can be useful in this case for obtaining a certificate for suboptimality in the form of a dual feasible point.

4. In this problem we will consider a different variant of the binary rating reconstruction problem we studied in class. Consider n "users" and m "movies", and a sparse set of ratings $y_{ij} \in \pm 1$ for $(i, j) \in S$, where S is a (small) subset of all user-movie pairs. We will again

want to find small-norm vectors $u_i \in \mathbb{R}^k$ and $v_j \in \mathbb{R}^k$ ($k > n, m$), associated with each user i and each user j, that explain the ratings in the sense that:

$$y_{ij}\langle u_i, v_j\rangle \geq 1$$

for each $(i, j) \in S$. However, this time we would like to minimize the maximum norm, i.e. optimize:

$$\begin{aligned} \text{minimize} \quad & \max(\max_i \|u_i\|, \max_j \|v_j\|) \\ \text{s.t.} \quad & y_{ij}\langle u_i, v_j\rangle \geq 1 \quad \forall (i, j) \in S \end{aligned} \tag{6}$$

(a) Express (6) as a semi-definite program.

(b) Derive the dual of the semi-definite program.

5. Problem 5.15 of Boyd and Vandenberghe.

Suggested review questions (please do not turn these in): 4.43 (try also deriving the dual of each one), 5.22, 5.41.