Convex Optimization

Problem set 3

Due Monday November 11th

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 expand the derivation of the dual to the logistic regression problem we did in class. Recall the logistic loss function is given by:

$$g(z) = \log(1 + e^{-z}) \tag{2}$$

You might want to plot $g(z)$ and see how $g(z)$ is close to zero when $z >> 0$ and increases roughly linearly when $z < 0$. It can thus be used to penalize values that we would like to be positive, and preferably away from zero.

   (a) In class we saw the expression of the Fenchel conjugate of $g(z)$. What is the value of $g^*(0)$ and $g^*(-1)$? What is the value of $g^*(p)$ for $p > 0$ or $p < -1$?

In a logistic regression model we would like to explain binary labels (responses) $y_1, \ldots, y_m$ using a linear function of input points (feature vectors, covariate vectors) $x_1, \ldots, x_m \in R^n$. In particular, we would like to find $w \in \mathbb{R}^n$ such that the sign of $w'x_i$ matches the label y_i, and we quantify this by minimizing $g(y_i w' x_i)$. Fitting a logistic regression model therefore corresponds to optimizing the following unconstrained convex optimization problem:

$$\text{minimize}_{w \in \mathbb{R}^n} \quad \sum_{i=1}^{m} g(y_i w' x_i) \tag{3}$$

In order to be able to derive a meaningful dual, e.g. in order to be able to obtain certificates of suboptimality, we instead rewrote (3) as:

$$\begin{aligned} \text{minimize}_{w \in \mathbb{R}^n, z \in \mathbb{R}^m} \quad & \sum_{i=1}^{m} g(z_i) \\ \text{s.t.} \quad & z_i = y_i w' x_i \qquad i = 1..m. \end{aligned} \tag{4}$$

(b) Use the dual of (4), seen in class, to write down the KKT conditions for a pair of primal and dual optimal solutions of (4). Explain how to use the KKT conditions to easily obtain a primal optimal solution if you are given a dual optimal solution.

(c) Consider adding a regulaization term, as is commonly done, to (3):

$$\text{minimize}_{w \in \mathbb{R}^n} \quad \sum_{i=1}^{m} g(y_i w' x_i) + \frac{\lambda}{2} \|w\|^2 \tag{5}$$

Modify (4) accordingly by adding a similar regularization term to its objective, and derive the dual of the resulting problem (Hint: first derive the Fenchel conjugate of the squared norm, possibly as a special case of the Fenchel conjugate of a quadratic).

(d) **[Optional]** An alternative loss function to the logistic loss is the hinge-loss (or Support Vector Machine loss) given by:

$$r(z) = [1 - z]_+ \tag{6}$$

Derive the Fenchel conjugate of $r(z)$, then replace the logistic loss $g(y_i w' x_i)$ with the hinge-loss $r(y_i w' x_i)$ in the regularized probelm (5), rewrite it using equality constraints and derive its dual.

An alternative way to obtain a dual is to represent the piecewise-linear hinge-loss $\xi_i = r(y_i w' x_i)$ using the two linear inequality constraints $\xi_i \geq 0$ and $\xi_i \geq 1 - y_i w' x_i$, resulting in:

$$\begin{aligned} \text{minimize}_{w \in \mathbb{R}^n, \xi \in \mathbb{R}^m} \quad & \sum_{i=1}^{m} \xi_i + \frac{\lambda}{2} \|w\|^2 \\ \text{s.t.} \quad & \xi_i \geq 0, \quad , \xi_i \geq 1 - y_i w' x_i \qquad i = 1..m. \end{aligned} \tag{7}$$

Derive the dual of (7) and compare it to the dual obtained above.

3. 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:

$$\begin{aligned} \text{minimize} \quad & \max(\max_i \|u_i\|, \max_j \|v_j\|) \\ \text{s.t.} \quad & y_{ij}\langle u_i, v_j \rangle \geq 1 \quad \forall (i,j) \in S \end{aligned} \tag{8}$$

The problem (8) can be reformulated as a semi-definite program:

$$\begin{aligned} \text{minimize} \quad & t \\ \text{s.t.} \quad & \left(\begin{smallmatrix} A & X \\ X' & B \end{smallmatrix} \right) \succeq 0 \\ & diag(A) \leq t \\ & diag(B) \leq t \\ & y_{ij} X_{ij} \geq 1 \quad \forall (i,j) \in S \end{aligned} \tag{9}$$

(a) Derive the dual of this semi-definite program.

(b) Write down, and simplify as much as you can, the KKT conditions for the problems.

Suggested review questions (please do not turn these in): 4.43 (try also deriving the dual of each one), 5.13, 5.22, 5.41.