

Management Sciences Topics: Convex Optimization

Homework 2: Due Feb 21 (11:59 pm)

(You can directly use any properties, theorems, examples or facts from the lectures.)

Problem 1: What are the KKT conditions and the dual problems of the following minimization problems?

- a. $\min_{x,y} c_1^\top x + c_2^\top y$ s.t. $Ax + By \leq d, x \geq 0$, where $c_1 \in \mathbb{R}^{d_1}, c_2 \in \mathbb{R}^{d_2}, A \in \mathbb{R}^{p \times d_1}, B \in \mathbb{R}^{p \times d_2}$ and $d \in \mathbb{R}^p$.

The dual problem:

$$\begin{aligned} \text{maximize} \quad & g(\lambda, \nu) = \inf_{x \in \mathcal{D}} c_1^\top x + c_2^\top y + \lambda_1(Ax + By - d) + \lambda_2(-x), \\ \text{subject to} \quad & \lambda \succeq 0. \end{aligned}$$

where $\mathcal{D} = \{(x, y) \mid Ax + By \leq d, x \geq 0\}$.

The KKT conditions:

$$\begin{aligned} A\tilde{x} + B\tilde{y} - d &\leq 0, \\ -\tilde{x} &\leq 0, \\ \tilde{\lambda}_i &\geq 0, \quad i = 1, 2 \\ \tilde{\lambda}_1(Ax + By - d) &= 0, \\ \tilde{\lambda}_2\tilde{x} &= 0, \\ \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \tilde{\lambda}_1 \begin{pmatrix} A \\ B \end{pmatrix} + \tilde{\lambda}_2 \begin{pmatrix} -1_x \\ 0 \end{pmatrix} &= 0. \end{aligned}$$

- b. $\min_{x \in \mathbb{R}^d} \frac{1}{2}x^\top Px + q^\top x$ s.t. $Ax \leq b$, where $P \in \mathbb{R}^{d \times d}, q \in \mathbb{R}^d, A \in \mathbb{R}^{p \times d}, b \in \mathbb{R}^p$ and P is a positive definite (and thus invertible) symmetric matrix.

The dual problem:

$$\begin{aligned} \text{maximize} \quad & g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \frac{1}{2}x^\top Px + q^\top x + \lambda(Ax - b) \\ \text{subject to} \quad & \lambda \succeq 0. \end{aligned}$$

where $\mathcal{D} = \{x \mid Ax \leq b\}$.

The KKT condition:

$$\begin{aligned} A\tilde{x} - b &\leq 0, \\ \tilde{\lambda} &\geq 0, \\ \tilde{\lambda}(A\tilde{x} - b) &= 0, \\ P\tilde{x} + q + \tilde{\lambda}A &= 0. \end{aligned}$$

Problem 2: Do the following problems satisfy Slater's condition? Does strong duality ($d^* = p^*$) hold for these problems? You only need to answer Yes or No for both questions. (Note that $d^* = p^*$ holds when they are both $+\infty$ or both $-\infty$.)

a. $\min_{x \in \mathbb{R}} \exp(-x)$ s.t. $x = 0$.

Satisfy Slater's condition, and strong duality holds.

b. $\min_{x \in \mathbb{R}} \exp(-x)$ s.t. $x^2 \leq 0$.

Does not satisfy Slater's condition, strong duality holds.

c. $\min_{x \in \mathbb{R}} x$ s.t. $x \leq -1, x \geq 1$.

Does not satisfy Slater's condition, strong duality holds.

Problem 3: Consider the following problem of projecting $z \in \mathbb{R}^d$ to a simplex:

$$\begin{aligned} \min_x \quad & \frac{1}{2} \|x - z\|_2^2 \\ \text{s.t.} \quad & \sum_{i=1}^d x_i = 1 \\ & x_i \geq 0, \quad i = 1, \dots, d. \end{aligned}$$

Answer the following questions:

- a. For $r \in \mathbb{R}$, let $x_i(r) := \max\{z_i - r, 0\}$ and $H(r) := \sum_{i=1}^d x_i(r)$. Show that there must be a r^* such that $H(r^*) = 1$. (Hint: Show that $H(r)$ is non-increasing in r and $H(r) \rightarrow 0$ as $r \rightarrow +\infty$ and $H(r) \rightarrow +\infty$ as $r \rightarrow -\infty$.)

Proof. First, for $r_1 < r_2$,

$$H(r_1) = \sum_{i=1}^d x_i(r_1) = \sum_{i=1}^r \max\{z_i - r_1, 0\} \geq \sum_{i=1}^r \max\{z_i - r_2, 0\} = H(r_2),$$

so $H(r)$ is nonincreasing. As $r \rightarrow +\infty$, $z_i - r < 0$ for some $r > \max\{z_i\}$. Hence $H(r) \rightarrow 0$ for $r \rightarrow +\infty$. On the other hand, when $r \rightarrow -\infty$, $z_i - r > -|z_i| - r \rightarrow +\infty$. Hence $H(r) \rightarrow +\infty$. Notice $H(r)$ is continuous in r , there is a r^* , s.t. $H(r^*) = 1$.

□

- b. Let r^* defined as in question a. Show that $x_i^* = x_i(r^*) = \max\{z_i - r^*, 0\}$ is the optimal solution. To do that, you need to find the corresponding dual variables which satisfy the KKT conditions together with x^* .

Proof. Consider the dual problem

$$\begin{aligned} \max \quad & \frac{1}{2} \|x - z\|_2^2 - \sum_{i=1}^d \lambda_i x_i + \nu \left(\sum_{i=1}^d x_i - 1 \right), \\ \text{s.t.} \quad & \lambda \succeq 0, \end{aligned}$$

and the KKT conditions

$$\begin{aligned} x_i &\geq 0, \quad i = 1, \dots, d, \\ \sum_{i=1}^d x_i &= 1, \\ \lambda_i &\geq 0, \quad i = 1, \dots, d, \\ \lambda_i x_i &= 0, \quad i = 1, \dots, d, \\ \begin{pmatrix} x_1 - z_1 \\ \vdots \\ x_d - z_d \end{pmatrix} - \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_d \end{pmatrix} + \nu \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} &= 0. \end{aligned}$$

Now we show the solution $x = x^*$ satisfies the KKT condition. Notice $\sum_{i=1}^d x_i^* = 1$, so there must be an index i , s.t. $x_i^* > 0$. Then for this x_i^* , $\lambda_i = 0$, so $\nu = z_i - x_i^* = r^*$. Now consider $x_j^* = 0$, then $\lambda_j = \nu - z_j = r^* - z_j \geq 0$. Hence (λ^*, ν^*, x^*) satisfies the KKT condition, and hence x^* is the optimal solution.

□

Problem 4: Implement the Proximal Gradient (PG) method to solve the following problems defined on the datasets from the LIBSVM library. You need to submit exactly two matlab files for each problem, one for the PG method without line search and one for the PG method with line search. You need to initialize both algorithms at $\mathbf{x} = \mathbf{0}$. In the PG method without line search, you need to manually tune η to obtain a reasonably good convergence speed. In the PG method with line search, you need to choose $\eta_0 = 1$, $\gamma_{\text{inc}} = 0.5$, and $\gamma_{\text{dec}} = 0.5$. Run the PG method for 200 iterations for each problem and plot the objective value at each iteration.

- a. Box-constrained linear regression: $\min_{\mathbf{x} \in \mathbb{R}^d} \frac{1}{2} \|A\mathbf{x} - b\|_2^2$ s.t. $-10 \leq x_i \leq 10$.
 (b, A) is from E2006-tfidf (the one we showed in the class).

b. ℓ_1 -ball-constrained logistic regression: $\min_{\mathbf{x} \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-b_i(\mathbf{a}_i^\top \mathbf{x})))$
s.t. $\|\mathbf{x}\|_1 \leq 10$. Here, $b_i \in \{1, -1\}$ is the class label. Dataset: rcv1.binary.
You can directly use the Matlab code by John Duchi (<https://stanford.edu/~jduchi/projects/DuchiShSiCh08.html>) for projections onto the ℓ_1 -ball.

The codes are in the “codes” subfolder, where the `box_constrained_linear_regression` is the code for problem (a), and `l1_constrained_logistic_regression` is for problem (b). The results are as follows:

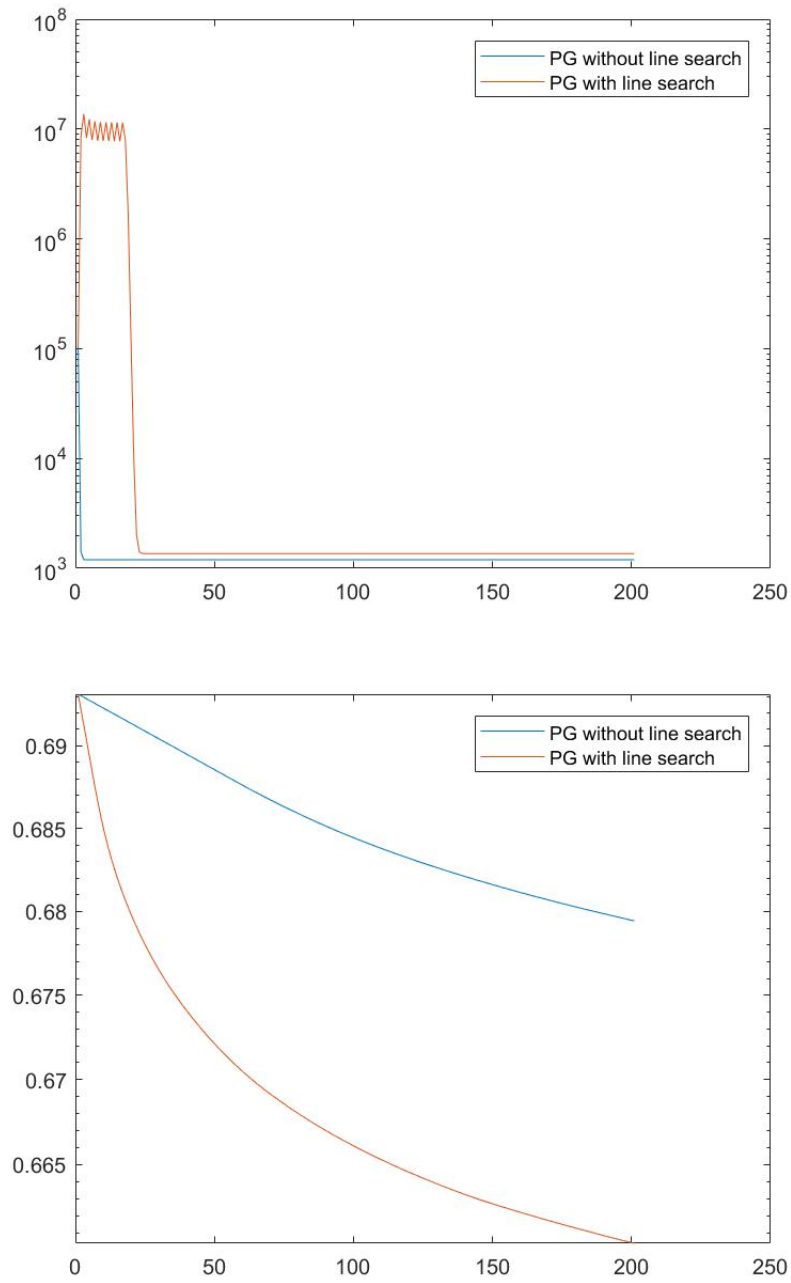


Figure 1: Results. The first one is for problem (a), while the second is for problem (b).