EECS 127/227AT Optimization Models in Engineering
Spring 2019

Homework 11

Release date: 11/21/19.
Due date: 12/05/19, 23:00 (11 pm). Please LaTeX or handwrite your homework solution and submit an electronic version.

Submission Format
Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned) as well as your IPython notebook saved as a PDF.

If you do not attach a PDF "printout" of your IPython notebook, you will not receive credit for problems that involve coding. Make sure that your results and your plots are visible. Assign the IPython printout to the correct problem(s) on Gradescope.

1. **Dual of a QP and differentiability**

   Consider a quadratic program of the form

   $$p^* = \max_x \ c^\top x - \frac{1}{2} x^\top Q x \ : \ Ax \le b,$$

   where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, and $Q \in \mathbb{S}^n_{++}$. We assume that the problem is feasible.

   (a) [A few lines]Form the dual of the QP. Show that strong duality holds.

   (b) [A few lines of justification] Show that p^*, considered as a function of c (resp. b), is convex (resp. concave).

   (c) [1-2 lines] [A few lines of justification] For any convex function $f : \mathbb{R}^n \to \mathbb{R}$, we say the vector $g \in \mathbb{R}^n$ is a subgradient of f at $x \in \textbf{dom } f$ if for all $z \in \textbf{dom } f$, $f(z) \ge f(x) + g^\top (z - x)$. The subgradient of f at x need not be unique. The set of all subgradients of f at x is known as the subdifferential and is denoted by $\partial f(x)$.

   Suppose f is the pointwise maximum of subdifferentiable (subdifferential exists at every point) convex functions $f_1, f_2, \cdots f_m$, so that $f(x) = \max_i f_i(x)$. Then, a subgradient of f at x is any vector g such that $g \in \partial f_k$ where k is any index such that $f_k(x) = f(x)$. In words, to find a subgradient of the maximum of functions at a point, we can choose one of the functions that achieves the maximum at that point, and choose any subgradient of that function at the point. This follows from the fact that $f(z) \ge f_k(z) \ge f_k(x) + g^\top (z - x) = f(x) + g^\top (z - x)$. (See section 8.2.3 in the course text for more details on subgradients).

   Explain how to form a subgradient of p^* considered as a function of c. You may assume subdifferentiability wherever it is helpful.

   (d) [1-2 lines] The convex function f is differentiable iff the subgradient at every point is unique. Is p^* differentiable with respect to c?

2. KKT conditions

Consider the optimization problem

$$\begin{aligned}&\min_{x\in\mathbb{R}^n} \sum_{i=1}^{n} \left(\tfrac{1}{2}d_i x_i^2 + r_i x_i\right)\\ &\text{s.t. } a^\top x = 1, \quad x_i \in [-1,1], \quad i = 1,\dots,n,\end{aligned}$$

where $|a_i| \geq 1$ and $d_i > 0$ for $i = 1,\dots,n$.

In this exercise, we will use the KKT conditions and/or the Lagrangian to find the dual of this optimization problem and develop some interesting results. Recall that the KKT conditions are:

- Primal feasibility: $x \in \mathcal{D}, f_i(x) \leq 0, i = 1,\dots,m$
- Dual feasibility: $\lambda \geq 0$
- Complementary slackness: $\lambda_i f_i(x) = 0, i = 1,\dots,m$
- Lagrangian stationarity: $x \in \arg\min \mathcal{L}(\cdot,\lambda)$

(a) [Rewritten program + 1-2 lines justification] Let $D = \operatorname{diag}(d_1,\dots,d_n)$, $r = (r_1,\dots,r_n)$. Show that the problem can be rewritten as a Quadratically Constrained Quadratic Program (QCQP) of the form

$$\begin{aligned}&\min_{x} \frac{1}{2}x^\top D x + r^\top x\\ &\text{s.t.} a^\top x = 1\\ &\quad x_i^2 \leq 1 \qquad\qquad i = 1,\dots,n.\end{aligned}$$

(b) [A few lines justification.] Recall that if

- The objective and all of the constraints are differentiable.
- Strong duality holds.
- The optimization problem is convex.

then the KKT optimality conditions are necessary and sufficient for any pair of (x^*,λ^*) to be optimal.

Verify that these three properties hold for this problem.

(c) [A few lines justification + KKT] Show that the Lagrangian is

$$\mathcal{L}(x,\lambda,\mu) = \frac{1}{2}x^\top (D+\Lambda)x + (r+\mu a)^\top x - \left(\mu + \frac{1}{2}\sum_{i=1}^{n}\lambda_i\right),$$

where $\Lambda = \operatorname{diag}(\lambda)$ and $\lambda = (\lambda_1,\dots,\lambda_n)$ are the dual variables corresponding to the inequality constraints, and μ is the dual variable corresponding to the equality constraint. Also, write down the KKT optimality conditions.

(d) [A few lines justification] Using the KKT condition that corresponds to the stationarity of the Lagrangian, write x^* in terms of λ^* and μ^*. In particular, show that

$$x_i^* = -\frac{r_i + \mu^* a_i}{d_i + \lambda_i^*}, \quad i = 1,\dots,n.$$

(e) [Algebraic justification] The dual problem amounts to maximizing the dual function $g(\lambda, \mu)$ w.r.t. μ and $\lambda \geq 0$. Show that the dual function can be written as

$$g(\lambda, \mu) = -\mu - \frac{1}{2} \sum_{i=1}^{n} \left[\frac{(r_i + \mu a_i)^2}{d_i + \lambda_i} + \lambda_i \right].$$

(f) [Algebraic expression + justification] Find the maximum of $g(\lambda, \mu)$ w.r.t. λ, for fixed μ. In particular, write down a closed-form expression for the corresponding optimal point $\lambda_i^*(\mu)$. *Do not forgot* the condition $\lambda_i^\star(\mu) \geq 0$.

(g) [Algebraic expression] Using your previous results, express the optimal primal point $x_i^*(\mu)$ as a function of the scalar dual variable μ only.

(h) [A few lines justification.]From previous parts, we found λ^* and x^* in terms of μ^*. Check that complementary slackness holds.

3. A matrix problem with strong duality

Consider the problem

$$p^* \doteq \min_{\Delta} \; c^\top (A+\Delta)^{-1} b \;:\; \|\Delta\| \leq 1,$$

where $A \in \mathbb{R}^{n \times n}$, with smallest singular value $\sigma_{\min}(A)$ strictly greater than one, and $b, c \in \mathbb{R}^n$. Here, $\|\cdot\|$ stands for the largest singular value norm. This problem arises in the study of equilibrium states of a dynamical system subject to perturbations.

(a) [A few lines justification] Show that the objective function is well-defined everywhere on the feasible set.

Hint: you can show that a square matrix is invertible if it has non singular values (or even values) equal to 0.

(b) [Several lines justification] Is the problem, as stated, convex? Give a proof or a counter-example.

(c) [Several lines justification] What is the abswer to the question in Part (b) if Δ is restricted to be symmetric, A is symmetric, and $b = c$? (We do not make such assumptions in the sequel.) *Hint:* Use Schur Complements. If you find the question hard, do not hesitate to work with some classmates.

(d) [A few lines justification] Show that the problem can be expressed as

$$\min_x \; c^\top x \;:\; \|Ax - b\|_2^2 \leq \|x\|_2^2.$$

(e) [A few lines justification] Given our assumption that $\sigma_{\min}(A) > 1$, describe the shape of the feasible set of the formulation in (c), in terms of the matrix $K \doteq A^\top A - I$, the vector $x_0 \doteq K^{-1} A^\top b$, and the scalar $\gamma \doteq x_0^\top K x_0 - b^\top b$. Explain why the above problem is convex.

(f) [Expression + several lines justification] Form a Lagrange dual to the problem. Does strong duality hold?

(g) [A few lines justification] Show that the optimal value can be written

$$p^* = c^\top (A^\top A - I)^{-1} A^\top b - \|(AA^\top - I)^{-1/2} b\|_2 \cdot \|(A^\top A - I)^{-1/2} c\|_2.$$

4. Ridge Regression Classifier Vs. SVM

In this problem, we explore Ridge Regression as a classifier, and compare it to SVM. Recall Ridge Regression solves the problem

$$\min_w \|Xw - y\|_2^2 + \lambda \|w\|_2^2,$$

where $X \in \mathbb{R}^{m,n}$, and $y \in \mathbb{R}^n$

(a) [A few lines explanation] Ridge Regression as is solves a regression problem. Given data $X \in \mathbb{R}^{m \times n}$ and labels $y \in \{0, 1\}^m$, explain how we might be able to train a Ridge Regression model and use it to classify a test point.

(b) [Notebook] Complete the accompanying Jupyter Notebook to compare Ridge Regression and SVM.