EECS 127/227AT Optimization Models in Engineering
Spring 2020 Homework 7 - PRACTICE

This homework is NEVER DUE. All problems are intended as practice for the midterm exam, and problems and solutions have been released simultaneously.

This version was compiled on 2020-03-11 16:31.

1. Optimizing over multiple variables

In this exercise, we consider several problems in which we optimize over two variables, $\vec{x} \in \mathbb{R}^n$ and $\vec{y} \in \mathbb{R}^m$, and a general (possibly nonconvex) objective function, $F_0(\vec{x}, \vec{y})$. Suppose also that $\vec{x}$ and $\vec{y}$ are constrained to different feasible sets $\mathcal{X}$ and $\mathcal{Y}$, respectively, which may or may not be convex.

(a) Show that

$$\min_{\vec{x}\in\mathcal{X}} \min_{\vec{y}\in\mathcal{Y}} \; F_0(\vec{x}, \vec{y}) = \min_{\vec{y}\in\mathcal{Y}} \min_{\vec{x}\in\mathcal{X}} \; F_0(\vec{x}, \vec{y}),$$

i.e., if we minimize over both $\vec{x}$ and $\vec{y}$, then we can exchange the minimization order without altering the optimal value.

Solution: We first consider the quantity $\min_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y})$, which can be viewed as a function of $\vec{x}$. We can write

$$\begin{aligned} F_0(\vec{x}, \vec{y}) &\geq \min_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y}) \\ &\geq \min_{\vec{x}\in\mathcal{X}} \min_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y}) \end{aligned}$$

where both lines follow from the definition of a minimum. The inequality above holds for every $\vec{x} \in \mathcal{X}$, so it holds for the value $\vec{x}$ that minimizes this quantity, i.e.,

$$\min_{\vec{x}\in\mathcal{X}} F_0(\vec{x}, \vec{y}) \geq \min_{\vec{x}\in\mathcal{X}} \min_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y}).$$

This inequality also holds for every $\vec{y} \in \mathcal{Y}$, so

$$\min_{\vec{y}\in\mathcal{Y}} \min_{\vec{x}\in\mathcal{X}} F_0(\vec{x}, \vec{y}) \geq \min_{\vec{x}\in\mathcal{X}} \min_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y}). \tag{1}$$

By symmetry, we can reverse our treatment of $\vec{x}$ and $\vec{y}$ and arrive at the reversed inequality

$$\min_{\vec{x}\in\mathcal{X}} \min_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y}) \geq \min_{\vec{y}\in\mathcal{Y}} \min_{\vec{x}\in\mathcal{X}} F_0(\vec{x}, \vec{y}). \tag{2}$$

Since both (1) and (2) must hold, the expressions must be equal, as desired.

(b) Show that $p^* \geq d^*$, where

$$\begin{aligned} p^* &\doteq \min_{\vec{x}\in\mathcal{X}} \max_{\vec{y}\in\mathcal{Y}} F_0(\vec{x}, \vec{y}) \\ d^* &\doteq \max_{\vec{y}\in\mathcal{Y}} \; \min_{\vec{x}\in\mathcal{X}} \; F_0(\vec{x}, \vec{y}). \end{aligned}$$

This statement is referred to as the *min-max theorem*.

Solution: By the definitions of minimization and maximization, we have that

$$L(\vec{y}) \doteq \min_{\vec{x}'} F_0(\vec{x}', \vec{y}) \leq F_0(\vec{x}, \vec{y}) \leq U(\vec{x}) \doteq \max_{\vec{y}'} F_0(\vec{x}, \vec{y}')$$

for every $\vec{x} \in \mathcal{X}$ and $\vec{y} \in \mathcal{Y}$, or more simply,

$$L(\vec{y}) \leq U(\vec{x}).$$

Since this inequality holds for all $\vec{x} \in \mathcal{X}$, it holds for the value of $\vec{x}$ that minimizes $U(\vec{x})$, and thus

$$p^* = \min_{\vec{x} \in \mathcal{X}} U(\vec{x}) \geq L(\vec{y}).$$

Similarly, since the above holds for all $\vec{y} \in \mathcal{Y}$, it holds for the value of $\vec{y}$ that maximizes $L(\vec{y})$, and thus

$$p^* \geq \max_{\vec{y} \in \mathcal{Y}} L(\vec{y}) = d^*$$

as desired.

2. **(Sp '19 Midterm 2 #7) Gradient descent algorithm**

Consider $g : \mathbb{R}^n \to \mathbb{R}$, $g(\vec{x}) = \frac{1}{2}\vec{x}^\top Q \vec{x} - \vec{x}^\top \vec{b}$, where Q is a symmetric positive definite matrix, i.e., $Q \in \mathbb{S}^n_{++}$.

(a) Write the update rule for the gradient descent algorithm

$$\vec{x}_{k+1} = \vec{x}_k - \eta \nabla g(\vec{x}_k),$$

where η is the step size of the algorithm, and bring it into the form

$$(\vec{x}_{k+1} - \vec{x}^*) = P_\eta(\vec{x}_k - \vec{x}^*),$$

where $P_\eta \in \mathbb{R}^{n \times n}$ is a matrix that depends on η. Find $\vec{x}^*$ and P_η in terms of Q, $\vec{b}$ and η. *Note:* $\vec{x}^*$ is a minimizer of g.

Solution: We have $\nabla g(\vec{x}) = Q\vec{x} - \vec{b}$ and

$$\vec{x}_{k+1} = \vec{x}_k - \eta(Q\vec{x}_k - \vec{b}) = \vec{x}_k - \eta Q(\vec{x}_k - Q^{-1}\vec{b}).$$

We can write

$$\vec{x}_{k+1} - Q^{-1}\vec{b} = \vec{x}_k - Q^{-1}\vec{b} - \eta Q(\vec{x}_k - Q^{-1}\vec{b}) = (I - \eta Q)(\vec{x}_k - Q^{-1}\vec{b}).$$

This shows that $\vec{x}^* = Q^{-1}\vec{b}$ and $P_\eta = I - \eta Q$.

(b) Write a condition on the step size η and the matrix Q that ensures convergence of $\vec{x}_k$ to $\vec{x}^*$ for every initialization of $\vec{x}_0$.

Solution: From part (a), we have

$$\vec{x}_k - \vec{x}^* = (I - \eta Q)^k (\vec{x}_0 - \vec{x}^*).$$

For every initialization $\vec{x}_0$, $(\vec{x}_k - \vec{x}^*)$ converges to zero if (and only if) all eigenvalues of $(I - \eta Q)$ is in $(-1, 1)$:

$$-1 < 1 - \eta\lambda < 1 \quad \text{for each eigenvalue } \lambda \text{ of } Q.$$

Since Q is positive definite, all of its eigenvalues are positive, and the right hand side of the inequality is satisfied for all $\eta > 0$. For the left hand side of the inequality, we need

$$-1 < 1 - n\lambda \ \forall \lambda_Q \iff \eta < \frac{2}{\lambda_{\max}(Q)}.$$

(c) Assume all eigenvalues of Q are distinct. Let η_m denote the largest stepsize that ensures convergence for all initializations $\vec{x}_0$, based on the condition computed in part (b). Does there exist an initialization $\vec{x}_0 \neq \vec{x}^*$ for which the algorithm converges to the minimum value of g for certain values of the step size η that are larger than η_m? Justify your answer.

Hint: The question asks if such initializations exist; not whether it is practical to find them.

Solution: From part (a), we have

$$\vec{x}_k - \vec{x}^* = (I - \eta Q)^k (\vec{x}_0 - \vec{x}^*).$$

If we want

$$(I - \eta Q)^k (\vec{x}_0 - \vec{x}^*) \to \vec{0} \quad \text{as } k \to \infty$$

for a specific initialization $\vec{x}_0$, the vector $(\vec{x}_0 - \vec{x}^*)$ must lie in the eigenspaces of $(I - \eta Q)$ corresponding to the eigenvalues in the range $(-1, 1)$. This explanation gets full credit.

For example, if $\frac{2}{\lambda_1} < \eta < \frac{2}{\lambda_2}$, where λ_1 and λ_2 are the largest two eigenvalues of Q, we have $(I - \eta Q)^k (\vec{x}_0 - \vec{x}^*) \to \vec{0}$ as long as $(\vec{x}_0 - \vec{x}^*)$ does not have any component in the eigenspace corresponding to the minimum eigenvalue of $(I - \eta Q)$.

3. Minimizing a sum of logarithms

Consider the following problem:

$$\begin{aligned} p^* = \max_{x \in \mathbb{R}^n} \quad & \textstyle\sum_{i=1}^n \alpha_i \ln x_i \\ \text{s.t.} \quad & x \geq 0, \quad \mathbf{1}^\top x = c, \end{aligned}$$

where $c > 0$ and $\alpha_i > 0$, $i = 1, \ldots, n$. Problems of this form arise, for instance, in maximum-likelihood estimation of the transition probabilities of a discrete-time Markov chain.

Determine in closed-form a minimizer, and show that the optimal objective value of this problem is

$$p^* = \alpha \ln(c/\alpha) + \sum_{i=1}^n \alpha_i \ln \alpha_i,$$

where $\alpha \doteq \sum_{i=1}^n \alpha_i$. We will show this in a series of steps.

(a) First, express the problem as a minimization problem. Then, can you relax the equality constraint to an inequality constraint while preserving the set of solutions?

Solution: Let us consider the equivalent problem

$$p^* = \min_{x \in \mathbb{R}^n} \quad \sum_{i=1}^n -\alpha_i \ln x_i \qquad \text{s.t.} \quad x \geq 0, \quad \mathbf{1}^\top x = c.$$

Since the objective is strictly decreasing over $x \geq 0$ and $\mathbf{1}^\top x$ is nondecreasing over $x \geq 0$, we can replace the equality constraint by an inequality one, thus we consider the problem

$$p^* = \min_{x \in \mathbb{R}^n} \quad \sum_{i=1}^n -\alpha_i \ln x_i \qquad \text{s.t.} \quad x \geq 0, \quad \mathbf{1}^\top x \leq c.$$

(b) After relaxing the equality constraint to an inequality constraint, form the Lagrangian $\mathcal{L}(x, \mu)$ for this problem, where μ is the dual variable corresponding to the inequality constraint containing c.

Solution: The partial Lagrangian for this problem is

$$\begin{aligned}
\mathcal{L}(x, \mu) &= \sum_{i=1}^n \alpha_i \ln 1/x_i + \mu(\mathbf{1}^\top x - c) \\
&= \sum_{i=1}^n (\alpha_i \ln 1/x_i + \mu x_i) - \mu c,
\end{aligned}$$

(c) Now derive the dual function $g(\mu)$ and solve the dual problem $d^* = \max_{\mu \geq 0} g(\mu)$. What is the optimal dual variable μ^*?

Solution: For the dual variable $\mu \geq 0$,

$$\begin{aligned}
g(\mu) &= \min_{x \geq 0} \mathcal{L}(x, \mu) = -\mu c + \sum_{i=1}^n \min_{x_i \geq 0} (\alpha_i \ln 1/x_i + \mu x_i) \\
&= -\mu c + \sum_{i=1}^n (\alpha_i \ln(\mu/\alpha_i) + \alpha_i) \\
&= -\mu c + \ln \mu \sum_{i=1}^n \alpha_i + \sum_{i=1}^n \alpha_i (1 - \ln \alpha_i).
\end{aligned}$$

The minimum with respect to $x_i \geq 0$ in the first expression is attained at the unique point $x_i = \alpha_i/\mu \geq 0$, which we obtain by verifying that the expression is convex with respect to x and setting the gradient to 0.

The dual is thus $d^* = \max_{\mu \geq 0} g(\mu)$.

Thus, the optimal dual variable is obtained as

$$\mu^* = \frac{\sum_{i=1}^n \alpha_i}{c} = \frac{\alpha}{c}.$$

(d) Assume strong duality holds, so $p^* = d^*$. (We will prove why this holds later). From the μ^* obtained in the previous part, how do we obtain the optimal primal variable x^*? And finally, what is the optimal objective function value p^*?

Solution: We obtain the optimal primal solution as

$$x_i^* = \frac{\alpha_i}{\mu^*} = \frac{c\alpha_i}{\alpha}, \quad i = 1, \ldots, n.$$

The expression for the optimal objective value follows by substituting this optimal solution back into the objective:

$$\begin{aligned} p^* &= \sum_{i=1}^{n} \alpha_i \ln\left(\frac{c\alpha_i}{\alpha}\right) \\ &= \sum_{i=1}^{n} \left(\alpha_i \ln\left(\frac{c}{\alpha}\right) + \alpha_i \ln \alpha_i\right) \\ &= \alpha \ln\left(\frac{c}{\alpha}\right) + \sum_{i=1}^{n} \alpha_i \ln \alpha_i. \end{aligned}$$