EECS 127/227AT Optimization Models in Engineering
Spring 2020 **Homework 13**

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

1. Multiple Choice

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Consider the following optimization problems:

$$\begin{aligned} p_1^* = \min_{t\in\mathbb{R},\vec{x}\in\mathbb{R}^n} \quad & t \\ \text{s.t. } & \|\vec{x}\|_2 = t, \\ & f(\vec{x}) \le 0, \end{aligned} \tag{1}$$

$$\begin{aligned} p_2^* = \min_{t\in\mathbb{R},\vec{x}\in\mathbb{R}^n} \quad & t \\ \text{s.t. } & \|\vec{x}\|_2 \le t, \\ & f(\vec{x}) \le 0. \end{aligned} \tag{2}$$

Write the statement labels (A, B, C) corresponding to statements that are true in the box given below. More than one statement might be true; and you will get credit for this problem only if you write the labels corresponding to all statements that are true and do not write a label corresponding to any statement that is false. No justification is required.

(A) Problem (1) as written is a convex problem.

(B) Problem (2) as written is a convex problem.

(C) We necessarily have $p_1^* = p_2^*$.

Solution: Statement with labels B and C are true. Problem (1) as written is not a convex problem since we have an equality constraint which is not affine.
Problem (2) as written is a convex problem since both inequality constraints are convex.
Further $p_1^* = p_2^*$ due to the following argument.
Since the second problem is a relaxation of the first we have $p_1^* \ge p_2^*$. Next suppose we have $(\vec{x}^*, t^*)$ as candidate solutions for (2) with $\|\vec{x}^*\|_2 = s < t^*$. Then we can decrease our objective value by using $(\vec{x}^*, s)$ which is feasible for both problems. Thus $p_1^* \le p_2^*$.

2. Linear algebra meets optimization Let wide matrix $A \in \mathbb{R}^{m\times n}$ $(m < n)$ be full row rank.

(a) Consider the ridge regression problem, where $\vec{b} \in \mathbb{R}^m, x \in \mathbb{R}^n$ and the constant $\lambda > 0$ is given:

$$\min_{\vec{x}} \quad ||A\vec{x} - \vec{b}||_2^2 + \lambda||\vec{x}||_2^2 \tag{3}$$

Since this is a convex problem and the objective function is differentiable, the optimum can be found by setting the gradient to zero. Use this to find the optimal solution $\vec{x}^*$.

Solution: Setting the gradient of the objective to 0 at optimum, we find that

$$2(A^\top A + \lambda I)\vec{x}^* - 2A^\top \vec{b} = 0$$

Since $A^\top A$ is PSD and λI is PD, it follows that $A^\top A + \lambda I$ is always invertible and hence

$$\vec{x}^* = (A^\top A + \lambda I)^{-1} A^\top \vec{b}$$

(b) Now we rewrite the problem in (3) by adding a constraint

$$\min_{\vec{z} = A\vec{x} - \vec{b}} ||\vec{z}||_2^2 + \lambda ||\vec{x}||_2^2. \tag{4}$$

Let the Lagrangian corresponding to this problem be $\mathcal{L}(\vec{x}, \vec{z}, \vec{\nu})$, where $\vec{\nu}$ is the dual variable corresponding to the equality constraint. Write out the dual function $g(\vec{\nu}) = \inf_{\vec{x}, \vec{z}} \mathcal{L}(\vec{x}, \vec{z}, \vec{\nu})$ explicitly. Solve the dual problem to get $\vec{\nu}^*$. Find the corresponding values of $\tilde{\vec{x}}, \tilde{\vec{z}}$ such that $g(\vec{\nu}^*) = \mathcal{L}(\tilde{\vec{x}}, \tilde{\vec{z}}, \vec{\nu}^*)$.

Solution: The dual problem is

$$\max_{\vec{\nu}} g(\vec{\nu})$$

where

$$\begin{aligned} g(\vec{\nu}) &= \min_{\vec{x}, \vec{z}} \mathcal{L}(\vec{x}, \vec{z}, \vec{\nu}) \\ &= \min_{\vec{x}, \vec{z}} \|\vec{z}\|^2 + \lambda \|\vec{x}\|^2 + \vec{\nu}^\top (\vec{z} - A\vec{x} + \vec{b}) \end{aligned}$$

First we minimize over $\vec{x}$. Setting the gradient to 0 we have that

$$2\lambda \vec{x}^* - A^\top \vec{\nu} = 0 \implies x^* = \frac{1}{2\lambda} A^\top \vec{\nu}$$

Setting the gradient to 0 for z we have that

$$2\vec{z}^* + \vec{\nu} = 0 \implies \vec{z}^* = -\frac{1}{2}\vec{\nu}$$

Plugging back in and simplifying the expression we have that

$$g(\vec{\nu}) = \vec{\nu}^\top \vec{b} - \frac{1}{4}\|\vec{\nu}\|_2^2 - \frac{1}{4\lambda}\|A^\top \vec{\nu}\|_2^2$$

Maximizing over $\vec{\nu}$ again amounts to setting the gradient to 0 at optimum. Hence we have

$$\vec{b} - \frac{1}{2}\vec{\nu}^* - \frac{1}{2\lambda} AA^\top \vec{\nu}^* = 0 \implies \vec{\nu}^* = 2(\frac{1}{\lambda} AA^\top + I)^{-1}\vec{b} = 2\lambda (AA^\top + \lambda I)^{-1}\vec{b}$$

It follows that

$$\begin{aligned} \tilde{\vec{x}} &= A^\top (AA^\top + \lambda I)^{-1}\vec{b} \\ \tilde{\vec{z}} &= A\tilde{\vec{x}} - \vec{b} \\ &= A(A^\top (AA^\top + \lambda I)^{-1}\vec{b}) - \vec{b}. \end{aligned}$$

(c) Show that for every $\lambda > 0$,

$$(A^\top A + \lambda I)^{-1} A^\top \vec{b} = A^\top (AA^\top + \lambda I)^{-1} \vec{b}.$$

Hint: One approach is to start by considering $\lambda A^\top + A^\top AA^\top$. Another approach is to use the SVD of A.

Solution:

Method 1: Let A have the thin SVD,

$$A = U\Sigma V^\top,$$

where $U \in \mathbb{R}^{m\times m}, \Sigma \in \mathbb{R}^{m\times m}, V^\top \in \mathbb{R}^{m\times n}$.

Using the SVD of A, $A^\top(AA^\top + \lambda I)^{-1}$ evaluates to,

$$\begin{aligned} A^\top(AA^\top + \lambda I_m)^{-1} &= V\Sigma U^\top(U\Sigma V^\top V\Sigma U^\top + \lambda I_m)^{-1} \\ &= V\Sigma U^\top(U\Sigma^2 U^\top + U\lambda I_m U^\top)^{-1} \\ &= V\Sigma U^\top(U(\Sigma^2 + \lambda I_m)^{-1}U^\top) \\ &= V\Sigma(\Sigma^2 + \lambda I_m)^{-1}U^\top. \end{aligned}$$

Note that $\Sigma^2 + \lambda I_m$ is invertible because A is full row rank and $\lambda > 0$.

Next we evaluate $(A^\top A + \lambda I)^{-1}A^\top$. We have,

$$\begin{aligned} (A^\top A + \lambda I_n)^{-1}A^\top &= (V\Sigma U^\top U\Sigma V^\top + \lambda I_n)^{-1}V\Sigma U^\top \\ &= (V\Sigma^2 V^\top + V\lambda I_m V^\top)^{-1}V\Sigma U^\top \\ &= (V(\Sigma^2 + \lambda I_m)V^\top)^{-1}V\Sigma U^\top \\ &= V(\Sigma^2 + \lambda I_m)^{-1}V^\top V\Sigma U^\top \\ &= V(\Sigma^2 + \lambda I_m)^{-1}\Sigma U^\top \\ &= V\Sigma(\Sigma^2 + \lambda I_m)^{-1}U^\top, \end{aligned}$$

where in the last equality we can interchange order of matrices since they are both diagonal.

Thus we have, $A^\top A + \lambda I)^{-1}A^\top = A^\top(AA^\top + \lambda I)^{-1}$,

which gives us,

$(A^\top A + \lambda I)^{-1}A^\top\vec{b} = A^\top(AA^\top + \lambda I)^{-1}\vec{b}$.

Method 2: Note

$$\lambda A^\top + A^\top AA^\top = A^\top(\lambda I + AA^\top) = (\lambda I + A^\top A)A^\top$$

Hence we have that

$$(\lambda I + A^\top A)^{-1}A^\top = A^\top(\lambda I + AA^\top)^{-1}$$

The result follows. Note that both $\lambda I + A^\top A$ and $\lambda I + AA^\top$ are invertible since they have strictly positive eigenvalues and are hence positive definite.

3. Best Approximation in the Uniform norm

Let $(x_1, y_1), \cdots, (x_n, y_n) \in \mathbb{R}^2$ be the given data points, and define vectors $\vec{x} = [x_1, \cdots, x_n]^\top$ and $\vec{y} = [y_1, \cdots, y_n]^\top$.

(a) We want to find $a, b \in \mathbb{R}$ that minimizes $\|a\vec{x} + b\vec{1} - \vec{y}\|_\infty$, where $\vec{1}$ is an n-dimensional vector of ones. Formulate this problem as an LP.

Solution: $\vec{x}$ and $\vec{y}$ are given and the decision variables are a and b, so the problem can be formulated as follows:

$$\begin{aligned}
&\min_{a,b} \left\|a\vec{x} + b\vec{1} - \vec{y}\right\|_\infty \\
&= \min_{a,b} \max_i |ax_i + b - y_i| \\
&= \min_{a,b,t} t \\
&\quad \text{s.t. } t \geq \pm(ax_i + b - y_i), \forall i
\end{aligned}$$

(b) Now we want to find $a, b \in \mathbb{R}$ that minimizes $\|a\vec{x} + b\vec{1} - \vec{y}\|_1$, where $\vec{1}$ is an n-dimensional vector of ones. Formulate this problem as an LP.

Solution: The problem can be formulated as follows:

$$\begin{aligned}
&\min_{a,b} \left\|a\vec{x} + b\vec{1} - \vec{y}\right\|_1 \\
&= \min_{a,b} \sum_i |ax_i + b - y_i| \\
&= \min_{a,b,t} \sum_i t_i \\
&\quad \text{s.t. } t_i \geq \pm(ax_i + b - y_i), \forall i
\end{aligned}$$

4. Newton's method

Given a symmetric positive definite matrix $Q \in \mathbb{S}^n_{++}$ and $\vec{b} \in \mathbb{R}^n$, consider the minimization of the function $f : \mathbb{R}^n \to \mathbb{R}$ defined as

$$f(\vec{x}) = \frac{1}{2}\vec{x}^\top Q\vec{x} - \vec{b}^\top \vec{x}.$$

Let $\vec{x}^*$ denote the point at which $f(\vec{x})$ is minimized, and define $\mathcal{B}(\vec{x}^*)$ as the ball centered at $\vec{x}^*$ with unit ℓ_2-norm:

$$\mathcal{B}(\vec{x}^*) = \{\vec{x} \in \mathbb{R}^n : \|\vec{x} - \vec{x}^*\|_2 \leq 1\}.$$

Assume we use Newton's method to minimize f:

$$\vec{x}_{k+1} = \vec{x}_k - (\nabla^2 f(\vec{x}_k))^{-1}\nabla f(\vec{x}_k),$$

where the initial point is $\vec{x}_0 \in \mathcal{B}(\vec{x}^*)$. For any $k \in \mathbb{N}$, find

$$\max_{\vec{x}_0 \in \mathcal{B}(\vec{x}^*)} \|\vec{x}_k - \vec{x}^*\|_2.$$

Solution: Note that $\nabla^2 f(\vec{x}_k) = Q$ for all $\vec{x}_k \in \mathbb{R}^n$. Therefore, the update rule is:

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

so we have

$$\vec{x}_k = \vec{x}^* \quad \forall k \geq 1.$$

As a result,

$$\max_{\vec{x}_0 \in \mathcal{B}(\vec{x}^*)} \|\vec{x}_k - \vec{x}^*\|_2 = \begin{cases} 1 & k = 0, \\ 0 & \forall k \geq 1. \end{cases}$$