# EECS 127/227AT Optimization Models in Engineering
Spring 2020 **Homework 5**

**This homework is due Friday, February 28, 2020 at 23:00 (11pm).**
**Self grades are due Friday, March 6, 2020 at 23:00 (11pm).**

This version was compiled on 2020-02-27 15:27.

**Submission Format:** Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned).

***Note****: While the problems on this homework can be completed in any order, they are most cohesive in the order in which they appear. Problems 1–3 all explore the same linear system $A\vec{x} = \vec{y}$ under different noise models; problem 1 examines the role of A's condition number in determining the system's sensitivity to different perturbations, and problems 2–3 show two different ways of using ridge regression to accommodate different models of noise. (You've already seen several ways of interpreting ridge regression in lecture; here are even more!) Problems 4–6 explore convexity and are unrelated to 1–3 or to each other.*

## 1. Condition number

In Lecture 8, we examined the sensitivity of solutions to linear system $A\vec{x} = \vec{y}$ (for nonsingular/invertible square matrix $A$) to perturbations in our measurements $\vec{y}$. Specifically, we showed that if we model measurement noise $\vec{\delta y}$ as a linear perturbation on $\vec{y}$, resulting in a linear perturbation $\vec{\delta x}$ on $\vec{x}$ — i.e., $A(\vec{x} + \vec{\delta x}) = \vec{y} + \vec{\delta y}$ — we can bound the magnitude of the solution perturbations $\vec{\delta x}$ as

$$\frac{\|\vec{\delta x}\|_2}{\|\vec{x}\|_2} \leq \kappa(A)\frac{\|\vec{\delta y}\|_2}{\|\vec{y}\|_2}, \tag{1}$$

where $\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$ is the condition number of $A$, or the ratio of $A$'s maximum and minimum singular values. In this problem, we will establish a similar bound for perturbations on $A$.

(a) Consider the linear system $A\vec{x} = \vec{y}$ above, where $A \in \mathbb{R}^{n\times n}$ is invertible (i.e., square and non-singular). Let $\delta A \in \mathbb{R}^{n\times n}$ denote a linear perturbation on matrix $A$ generating a corresponding linear perturbation $\vec{\delta x}$ in solution $\vec{x}$, i.e.,

$$(A + \delta A)(\vec{x} + \vec{\delta x}) = \vec{y}.$$

Show that

$$\frac{\|\vec{\delta x}\|_2}{\|\vec{x} + \vec{\delta x}\|_2} \leq \kappa(A)\frac{\|\delta A\|_2}{\|A\|_2}. \tag{2}$$

(b) (Practice: This subpart will not be graded, but it is helpful for understanding.) Note that Eq. 1 and 2 above bound two slightly different quantities: $\frac{\|\vec{\delta x}\|_2}{\|\vec{x}\|_2}$ and $\frac{\|\vec{\delta x}\|_2}{\|\vec{x}+\vec{\delta x}\|_2}$, respectively. In general, we wish to establish these bounds because we want to characterize the size of $\vec{\delta x}$ under different sizes of perturbation. Which of these two bounds better serves this purpose? *Hint:* Consider different relative values of $\vec{x}$ and $\vec{\delta x}$. What happens to the bounds when $\vec{\delta x} \gg \vec{x}$?

**2. Ridge regression for bounded output noise (1/2)**

We will first solve the ridge regression problem in the case where our output measurements $\vec{y}$ are noisy and we have some bounds on this noise, as well as some specific knowledge about data matrix $A$.

Let square matrix $A \in \mathbb{R}^{n\times n}$ have the singular value decomposition $A = U\Sigma V^\top$, and let its smallest singular value be $\sigma_{\min}(A) > 0$.

(a) Is $A$ invertible? If so, write the singular value decomposition of $A^{-1}$.

(b) Consider the linear equation $A\vec{x} = \vec{\tilde{y}}$, where $\vec{\tilde{y}} \in \mathbb{R}^n$ is a noisy measurement satisfying

$$\|\vec{\tilde{y}} - \vec{y}\|_2 \le r$$

for some vector $\vec{y} \in \mathbb{R}^n$ and $r > 0$. Let $\vec{x}^*(\vec{y})$ denote the solution of $A\vec{x} = \vec{y}$. Show that

$$\max_{\vec{\tilde{y}}:\|\vec{\tilde{y}}-\vec{y}\|_2\le r} \|\vec{x}^*(\vec{\tilde{y}}) - \vec{x}^*(\vec{y})\|_2 = \frac{r}{\sigma_{\min}(A)}$$

(c) What happens if the smallest singular value of $A$ is very close to zero? Why is this problematic for finding our solution vector $\vec{x}^*$?

(d) Now assume that we find optimal value $\vec{x}^*$ via ridge regression, i.e., we compute

$$\vec{x}^*_\lambda(\vec{\tilde{y}}) = \arg\min_{\vec{x}} \|A\vec{x} - \vec{\tilde{y}}\|_2^2 + \lambda\|\vec{x}\|_2^2$$

for some chosen value $\lambda \in \mathbb{R}$. Compute $\vec{x}^*_\lambda(\vec{\tilde{y}})$, our optimal solution vector (now parameterized by $\lambda$), by solving this optimization problem. You may use the solution from class/the last HW for this part.

(e) Show that for all $\lambda > 0$,

$$\max_{\vec{\tilde{y}}:\|\vec{\tilde{y}}-\vec{y}\|_2\le r} \|\vec{x}^*_\lambda(\vec{\tilde{y}}) - \vec{x}^*_\lambda(\vec{y})\|_2 \le \frac{r}{2\sqrt{\lambda}}.$$

How does the value of $\lambda$ affect the sensitivity of your solution $\vec{x}^*_\lambda(\vec{y})$ to noise in $\vec{y}$?

*Hint:* For every $\lambda > 0$, we have

$$\max_{\sigma>0} \frac{\sigma}{\sigma^2 + \lambda} = \frac{1}{2\sqrt{\lambda}}.$$

(You need not show this; this optimization can be solved by setting the derivative of the objective function to 0 and solving for $\sigma$.)

**3. Ridge regression for data matrix noise (2/2)**

Next, we will solve the ridge regression problem in the case where our data matrix $A$ is noisy and we know some properties of this noise.

Consider the standard least-squares problem

$$\min_{\vec{x}} \|A\vec{x} - \vec{y}\|_2^2,$$

in which the data matrix $A \in \mathbb{R}^{m\times n}$ is noisy. We model this noise by assuming that each row $\vec{a_i}^\top \in \mathbb{R}^n$ has the form $\vec{a_i} = \vec{\tilde{a}}_i + \vec{u_i}$, where the noise vector $\vec{u_i} \in \mathbb{R}^n$ has zero mean and covariance matrix $\sigma^2 I_n$, with $\sigma \in \mathbb{R}$ a measure of the size of the noise. Therefore, now the matrix $A$ is a function of the random variable $U = (\vec{u_1}, \ldots, \vec{u_m})$, which we denote by $A_U$. We will use $\hat{A}$ to denote the matrix with rows $\vec{a}_i^\top$, $i = 1, \dots, m$. To account for this noise, we replace the standard least squares formulation above with

$$\min_{\vec{x}} \mathbb{E}_U\{\|A_U\vec{x} - \vec{y}\|_2^2\},$$

where $\mathbb{E}_U$ denotes the expected value with respect to the random variable $U$. Show that this problem can be written as

$$\min_{\vec{x}} \|\hat{A}\vec{x} - \vec{y}\|_2^2 + \lambda\|\vec{x}\|_2^2,$$

where $\lambda \geq 0$ is some regularization parameter, which you will determine. In other words, show that regularized least-squares can be interpreted as a way to take into account uncertainties in the matrix $A$, in the expected value sense.
*Hint 1:* Compute the expected value of $((\vec{\hat{a}}_i + \vec{u}_i)^\top \vec{x} - y_i)^2$, for a specific row index $i$.
*Hint 2:* Trace trick: We can use the fact that trace of a scalar is equal to the scalar and write,

$$\begin{aligned}\mathbb{E}_U\{\vec{u}_i^\top B\vec{u}_i\} &= \mathbb{E}_U\{\operatorname{trace}(\vec{u}_i^\top B\vec{u}_i)\} \\ &= \mathbb{E}_U\{\operatorname{trace}(\vec{u}_i\vec{u}_i^\top B)\},\end{aligned}$$

for any matrix $B$.

**4. Visualizing rank-1 matrices**

In this problem, we explore the effect of rank constraints on the convexity of matrix sets.

First, consider the set of all $2 \times 2$ matrices with diagonal elements $(1, 2)$, which we can write explicitly as

$$S_{2\times2} = \left\{ \begin{bmatrix} 1 & x \\ y & 2 \end{bmatrix} \middle| x, y \in \mathbb{R} \right\}.$$

(a) Is set $S_{2\times2}$ convex? If so, provide a proof, and if not, provide a counterexample.

(b) Suppose we now wish to define $S_{2\times2}^{(1)} \subset S_{2\times2}$, the set of all rank-1 matrices in $S_{2\times2}$. Write out conditions on $x$ and $y$ (i.e. equation constraints that $x$ and $y$ must satisfy) to define $S_{2\times2}^{(1)}$ explicitly.

(c) Is set $S_{2\times2}^{(1)}$ convex? If so, provide a proof, and if not, provide a counterexample. Plot the $x$–$y$ curve described by constraints you found in the earlier part and observe its shape.
*Hint:* Any linear function applied to a convex set generates another convex set[1], and the function that maps set $S_{2\times2}^{(1)}$ to variables $(x, y)$ is linear.[2]

(d) In this class, we will sometimes pose optimization problems in which we optimize over sets of matrices. Since low-dimensional models are often easier to interpret, it would be nice to impose rank constraints on these solution matrices. Suppose we wish to solve the optimization problem

$$\min_{A \in S_{2\times2}^{(1)}} \|A\|_F^2$$

[1] You are asked to prove this in problem 6(a).

[2] We can show this directly from the definition of linearity: define function $f : S_{2\times2}^{(1)} \to \mathbb{R}^2$ that maps each set element $s$ to its corresponding off-diagonal values $(x, y)$. Then for any two elements $s_1, s_2 \in S_{2\times2}^{(1)}$ and $\alpha_1, \alpha_2 \in \mathbb{R}$, we have $f(\alpha_1 s_1 + \alpha_2 s_2) = \alpha_1 f(s_1) + \alpha_2 f(s_2)$.

which is equivalent to

$$\begin{aligned} \min_{A \in S_{2\times 2}} \quad & \|A\|_F^2 \\ \text{s.t.} \quad & \text{rk}(A) = 1. \end{aligned}$$

Is this optimization problem convex?

**5. Properties of convex functions**

In this exercise, we examine convexity and what it represents graphically.

(a) In what region between $[0, 2\pi]$ is $\sin(x)$ a convex function? In what region between $[0, 2\pi]$ is $\sin(x)$ a concave function? Give a region between $[0, 2\pi]$ where $\sin(x)$ is neither convex nor concave.

(b) Plot $\sin(x)$ between $[0, 2\pi]$. For each of the 3 intervals defined above in part (a), draw a chord to illustrate graphically on what regions the function is convex, concave, and neither convex nor concave.

(c) Show that for all $x \in [0, \frac{\pi}{2}]$,

$$\frac{2}{\pi}x \leq \sin x \leq x.$$

**6. Convexity**

(a) Show the conservation of convexity through affine transformation, i.e., prove that if $S \subseteq \mathbb{R}^n$ is convex, then the image of $S$ under an affine function $f$,

$$f(S) = \{f(\vec{x}) \mid \vec{x} \in S\},$$

is convex.

(b) Show that a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its epigraph, defined as $\text{epi}(f) = \{(\vec{x}, t) \mid \vec{x} \in \text{dom}(f), f(\vec{x}) \leq t\}$, is convex.

**7. Homework process**

Whom did you work with on this homework? List the names and SIDs of your group members.