EECS 127/227AT Optimization Models in Engineering

Spring 2019 **Midterm 1**

In class exam

PRINT your student ID: ________________________________

PRINT AND SIGN your name: ________________, ________________ ________________

(last name) (first name) (signature)

Name and SID of the person to your left: ________________________________

Name and SID of the person to your right: ________________________________

Name and SID of the person in front of you: ________________________________

Name and SID of the person behind you: ________________________________

1. (1 Point) What is one of your favorite things to do outside of school?

2. (1 Point) What is one of things you learned in 127/227AT that you enjoyed?

Do not turn this page until the proctor tells you to do so. You may work on the questions above.

PRINT your student ID: ____

Extra page for scratchwork.

Work on this page will NOT be graded.

PRINT your student ID: ____________________

3. (13 points) Singular value decomposition

The compact form of the singular value decomposition of a matrix $A \in \mathbb{R}^{3\times 3}$ is given as

$$A = \begin{bmatrix} \frac{2}{3} & \frac{1}{\sqrt{2}} \\ \frac{2}{3} & -\frac{1}{\sqrt{2}} \\ \frac{1}{3} & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \end{bmatrix}.$$

(a) (2 points) What is the rank of A? Justify.

(b) (3 points) What is the dimension of the column space (range) of A? Write a basis for the column space (range) of A.

(c) (4 points) What is the dimension of the null space of $A^\top$? Write a basis for the null space of $A^\top$.

PRINT your student ID: ____

(d) (4 points) Let $\mathcal{B}_2$ denote the unit-norm ball in ℓ_2 norm: $\mathcal{B}_2 = \{z \in \mathbb{R}^3 : \|z\|_2 \leq 1\}$. Compute the minimum value of $x^\top Ay$, where x and y are two vectors in $\mathcal{B}_2$; that is, find $\min_{x,y \in \mathcal{B}_2} x^\top Ay$.

PRINT your student ID: ____

4. (12 points) Symmetric and skew-symmetric matrices

A square matrix $A \in \mathbb{R}^{n \times n}$ is called skew-symmetric if all its diagonal elements are zero and $A_{ij} = -A_{ji}$ for all $i, j \in \{1, \ldots, n\}$. In other words, A is skew-symmetric if and only if $A^\top = -A$.

(a) (3 points) Let $A \in \mathbb{R}^{n \times n}$ be a skew-symmetric matrix, and let $B \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Show that $\langle A, B \rangle = 0$, where inner product of A and B is defined as

$$\langle A, B \rangle = \mathrm{Trace}(A^\top B).$$

Note that this implies the space of symmetric matrices is orthogonal to the space of skew-symmetric matrices.

(b) (2 points) The set of all matrices in $\mathbb{R}^{2 \times 2}$ forms a vector space. All elements in this space can be written as a linear combination of

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\},$$

and because these matrices are linearly independent, they provide a basis for $\mathbb{R}^{2 \times 2}$.

Similarly, the set of all skew-symmetric matrices in $\mathbb{R}^{2 \times 2}$ forms a vector space. Write a basis for the space of skew-symmetric matrices in $\mathbb{R}^{2 \times 2}$.

PRINT your student ID: ______________________________

(c) (4 points) Consider the matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{R}^{2\times 2}$. Find a symmetric matrix $A_{\text{sym}} \in \mathbb{R}^{2\times 2}$ and a skew-symmetric matrix $A_{\text{skew}} \in \mathbb{R}^{2\times 2}$ such that

$$A = A_{\text{sym}} + A_{\text{skew}}.$$

(d) (3 points) Consider the function $f : \mathbb{R}^2 \mapsto \mathbb{R}$, which is defined as $f(x) = \frac{1}{2} x^\top \begin{bmatrix} a & b \\ c & d \end{bmatrix} x$. Find the Hessian of the function. Show your calculations.

PRINT your student ID: ____________________

5. (12 points) Online Least Squares

Consider n sensors located in different areas of California to measure the temperature of the air (as a scalar).

Let $x_{i,t}$ denote the measurement from sensor i at time t, for for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$. Assume $T < n$.

We represent $x_t \in \mathbb{R}^n$ as a column vector of all measurements at time t. Thus,

$$x_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{n,t} \end{bmatrix}.$$

Let $X_t \in \mathbb{R}^{n \times t}$ denote the matrix with columns $x_1, x_2 ... x_t$. Thus we have,

$$X_t = \begin{bmatrix} x_1 & x_2 & \ldots & x_t \end{bmatrix}.$$

We additionally consider scalars $y_1, y_2, \ldots, y_n$ where $y_i \in \mathbb{R}$. Here y_i represents wind chill at the region corresponding to sensor i, as predicted by meteorological department with the help of weather satellites. Note that y_i does not depend on time t. Let $y \in \mathbb{R}^n$ denote the column vector containing the wind chill at all sensors. Thus,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Define $q_t \in \mathbb{R}^n$ iteratively as follows. First,

$$q_1 = \frac{x_1}{\|x_1\|_2}.$$

For $t = 2, 3, \ldots, T$:

$$s_t = x_t - \sum_{j=1}^{t-1} \langle x_t, q_j \rangle \, q_j$$

$$q_t = \frac{s_t}{\|s_t\|_2}.$$

Let $Q_t \in \mathbb{R}^{n \times t}$ denote the matrix with columns $q_1, q_2, \ldots, q_t$. Thus we have,

$$Q_t = \begin{bmatrix} q_1 & q_2 & \ldots & q_t \end{bmatrix}.$$

Assume that for all $t = 1, 2, \ldots, T$, the matrix X_t is full column rank, i.e the columns of X_t are linearly independent. Further for this problem, assume that the inner product of x and y, $\langle x, y \rangle$, is given by $x^\top y$.

For each time $i = 1, 2, \dots t$, we are interested in fitting a linear model to predict y from X_t, the sensor measurements up to time t. Consider the following two problems at time t:

$$w_t^* = \operatorname*{argmin}_w \|y - X_t w\|_2$$
$$v_t^* = \operatorname*{argmin}_v \|y - Q_t v\|_2 .$$

(a) (1 point) Write an expression for w_t^* in terms of X_t and y.

(b) (1 point) Write an expression for v_t^* in terms of Q_t and y.

(c) (3 points) Show that for each t,

$$X_t w_t^* = Q_t v_t^*.$$

Note that this implies that the matrices Q and X have the information for fitting a linear model for y.

PRINT your student ID: ____________________

(d) (3 points) For each t, show that we can express w_t^* in terms of X_t, Q_t and v_t^* as,

$$w_t^* = (X_t^\top X_t)^{-1} X_t^\top Q_t v_t^*.$$

Hint: It might be useful to start by justifying that y can be expressed as $y = X_t w_t^* + e_t$, where e_t is orthogonal to columns of X_t. Then use the fact that $X_t w_t^* = Q_t v_t^*$.

PRINT your student ID: ______________________________

(e) (4 points) Show that v_t^* can be obtained in terms of y and $q_i, i = 1, 2, \ldots, t$ as:

$$v_t^* = \begin{bmatrix} (v_t^*)_1 \\ (v_t^*)_2 \\ \vdots \\ (v_t^*)_t \end{bmatrix},$$

where,

$$(v_t^*)_i = q_i^\top y, \quad i = 1, 2, \ldots, t.$$

For $t = 2, 3, \ldots, T$, use the above result to obtain v_t^* using only v_{t-1}^*, q_t and y.

Print your student ID: ____

6. (9 points) PCA

Let $\{x_1, x_2, \ldots, x_m\}$ be a set of points in $\mathbb{R}^3$. Assume that their empirical mean $\hat{x}$ and empirical covariance matrix Σ are given as:

$$\hat{x} = \frac{1}{m}\sum_{i=1}^{m} x_i = 0,$$

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m} x_i x_i^\top = \begin{bmatrix} 0.8 & 0.6 & 0 \\ -0.36 & 0.48 & 0.8 \\ -0.48 & 0.64 & -0.6 \end{bmatrix} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 0.8 & -0.36 & -0.48 \\ 0.6 & 0.48 & 0.64 \\ 0 & 0.8 & -0.6 \end{bmatrix}.$$

Note that Σ is a symmetric matrix, and it is given with its singular value decomposition, which is equivalent to its eigen-decomposition.

(a) (3 points) Let $w \in \mathbb{R}^3$ be a vector. Let $\{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_m\}$ be projections of the points onto w. Write a vector w that maximizes the variance of the projected points $\{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_m\}$.

PRINT your student ID: ______________________________

(b) (6 points) Let $A \in \mathbb{R}^{3\times 3}$ be a matrix such that $A^\top A = \Sigma$, and consider the set $S \subset \mathbb{R}^3$ defined as $S = \{Au : u \in \mathbb{R}^3, \|u\|_2 \leq 1\}$. Assume that the points in S are projected onto the hyperplane $\mathcal{H}(w) = \{z \in \mathbb{R}^3 : w^\top z = 0\}$ for some $w \in \mathbb{R}^3$. Find the vector w for which the projection of S onto $\mathcal{H}(w)$ is a circular disc.

PRINT your student ID: ____

7. (12 points) Errors in the measurement apparatus

This question is about solving the following optimization problem:

$$Q^*_\lambda = \underset{Q}{\operatorname{argmin}} \|Qw - y\|_2^2 + \lambda \|X - Q\|_F^2 . \tag{1}$$

$w \in \mathbb{R}^m, y \in \mathbb{R}^n, X \in \mathbb{R}^{n \times m}, \lambda \in \mathbb{R}, \lambda > 0$ **are all known and constant in Eq.** (1)**.** You do not have to carefully read the rest of the setup to solve the question, but it might give you context.

We perform a series of experiments where we illuminate an object with patterned light and collect the reflected light after it was incident on the object, to help us understand the properties of the object. This is the key idea behind tomography.

The patterned light that is used for illumination is measured and recorded in the "measurement" matrix $X \in \mathbb{R}^{n \times m}$. The ith row of this matrix, $x_i^\top$, represents the illumination measurement for the ith experiment, with a total of n experiments. The intensity of the reflected light is measured as a scalar observation, y_i, for the ith experiment.

Thus we have n pairs $(x_i, y_i) \quad i = 1, 2, \ldots, n$, corresponding to the n experiments, where $x_i \in \mathbb{R}^m$ is a vector and $y_i \in \mathbb{R}$ is a scalar.

If our observations are accurate, we expect that (x_i, y_i) should satisfy the equation,

$$x_i^\top w = y_i, \quad i = 1, 2, \ldots, n.$$

where vector $w \in \mathbb{R}^m$ represents a **known** image.

To be precise, let $X \in \mathbb{R}^{n \times m}$ denote the matrix with rows as $x_i^\top$,

$$X = \begin{bmatrix} \leftarrow & x_1^\top & \rightarrow \\ \leftarrow & x_2^\top & \rightarrow \\ \leftarrow & \vdots & \rightarrow \\ \leftarrow & x_n^\top & \rightarrow \end{bmatrix}.$$

Let $y \in \mathbb{R}^n$ denote the column vector with entries y_i,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

We expect the following equation to hold:

$$Xw = y,$$

where w is a **known** vector.

Unfortunately, after completing the experiment we find that our apparatus made errors while measuring the illumination on the image, i.e. there there are small errors in the recorded values of X. However the observations, y_i, are accurate. We would like to recover the true value of the illumination, represented by $X_{true} = Q$. For this we use the two pieces of information that we have:

- $Qw \approx y$
- $Q \approx X$.

We mathematize this by writing the objective function Eq. (1).

PRINT your student ID: ____

(a) (3 points) Consider $Z \in \mathbb{R}^{m \times m}$ defined as $Z = ww^\top + \lambda I$. Show that Z is invertible.

PRINT your student ID: ____

(b) (8 points) Find an expression for Q_λ^* as defined in Eq. (1) in terms of λ, X, w and y. Assume that we can find the minimum value by setting gradient with respect to Q of the objective function to zero. Justify any algebraic manipulations you make.
Hint: The following identities might be useful:

$$\begin{aligned}
\|X\|_F^2 &= \mathrm{Trace}(X^\top X),\\
\nabla_Q \mathrm{Trace}(Q^\top Q B) &= QB^\top + QB, \quad \text{for B square},\\
\nabla_Q \mathrm{Trace}(AQ) &= A^\top.
\end{aligned}$$

PRINT your student ID: ______________________________

(c) (1 point) Find $\lim_{\lambda \to \infty} Q^*_\lambda$.

PRINT your student ID: ______________________________

Extra page for scratchwork.
Work on this page will NOT be graded.

PRINT your student ID: ____

Extra page for scratchwork.

Work on this page will NOT be graded.

PRINT your student ID: ____

Doodle page!
Draw us something if you want or give us suggestions, compliments, or complaints.
You can also use this page to report anything suspicious that you might have noticed.

EECS 127/227AT Optimization Models in Engineering

Spring 2019 — Midterm 1 Instructions

Read the following instructions before the exam.

There are 7 problems of varying numbers of points. You have 80 minutes for the exam. The problems are of varying difficulty, so pace yourself accordingly, do easier problems first, and avoid spending too much time on any one question until you have gotten all of the other points you can.

There are 20 pages on the exam, so there should be 10 sheets of paper in the exam. The exam is printed double-sided. Do not forget the problems on the back sides of the pages! Notify a proctor immediately if a page is missing. **Do not tear out or remove any of the pages. Do not remove the exam from the exam room.**

No collaboration is allowed, and do not attempt to cheat in any way. Cheating will not be tolerated.

Write your student ID on each page. If a page is found without a student ID, we are not responsible for identifying the student who wrote that page. You may consult ONE handwritten 8.5" $\times$ 11" note sheet (front and back). No phones, calculators, tablets, computers, other electronic devices, or scratch paper are allowed.

Please write your answers legibly in the boxed spaces provided on the exam. The space provided should be adequate.

In general, show all of your work in order to receive full credit.

Partial credit will be given for substantial progress on each problem.

If you need to use the restrooms during the exam, bring your student ID card, your phone, and your exam to a proctor. You can collect them once you return from the restrooms.

Good luck!

Do not turn this page until the proctor tells you to do so.