EECS 16A: Foundations of Signals, Dynamical Systems, and Information Processing Exam 3

Department of Electrical Engineering and Computer Sciences

University of California, Berkeley 01 May 2025

FIRST Name: __________ LAST Name: ____________ SID (All Digits): __________

- **(5 Points)** On *every* page, print legibly your name and ALL digits of your SID. For every page on which you do not write your name and SID, you forfeit a point, up to the maximum 5 points.
- **(10 Points) (Pledge of Academic Integrity)** Hand-copy, sign, and date the single-line text (which begins with *I have read, . . .*) of the Pledge of Academic Integrity on page 3 of this document. Your solutions will *not* be evaluated without this.
- **Urgent Contact with the Teaching Staff:** In case of an urgent matter, raise your hand.
- **This document consists of pages numbered 1 through 16.** Verify that your copy of the exam is free of anomalies, and contains all of the specified number of pages. If you find a defect in your copy, contact the teaching staff immediately.
- This exam is designed to be completed within 70 minutes. However, you may use up to 80 minutes total—*in one sitting*—to tackle the exam.

  The exam starts at 8:10 pm California time. Your allotted window begins with respect to this start time. Students who have official accommodations of 1.5× and 2× time windows have 120 and 160 minutes, respectively.
- **This exam is closed book.** You may not use or access, or cause to be used or accessed, any reference in print or electronic form at any time during the exam, except three double-sided 8.5"×11" sheets of handwritten, original notes having no appendage.

  Collaboration is not permitted.

  Computing, communication, and other electronic devices (except dedicated timekeepers) must be turned off.

  Scratch paper will be provided to you; ask for more if you run out. You may not use your own scratch paper.
- Please write neatly and legibly, because *if we can't read it, we can't evaluate it.*
- For each problem, limit your work to the space provided specifically for that problem. *No other work will be considered. For example, we will not evaluate scratch work. No exceptions.*
- Unless explicitly waived by the specific wording of a problem, you must explain your responses (and reasoning) succinctly, but clearly and convincingly.
- In some parts of a problem, we may ask you to establish a certain result—for example, "show this" or "prove that." Even if you're unable to establish the result that we ask of you, you may still take that result for granted—and use it in any subsequent part of the problem.

- If we ask you to provide a ”reasonably simple expression” for something, by default we expect your expression to be in closed form—one *not* involving a sum $\sum$ or an integral $\int$—*unless* we explicitly tell you otherwise.

- Noncompliance with these or other instructions from the teaching staff—*including, for example, commencing work prematurely, or continuing it beyond the allocated time window*—is a serious violation of the Code of Student Conduct.

Pledge of Academic Integrity

By my honor, I affirm that

(1) this document—which I have produced for the evaluation of my performance—reflects my original, bona fide work, and that I have neither provided to, nor received from, anyone excessive or unreasonable assistance that produces unfair advantage for me or for any of my peers;

(2) as a member of the UC Berkeley community, I have acted with honesty, integrity, respect for others, and professional responsibility—and in a manner consistent with the letter and intent of the campus Code of Student Conduct;

(3) I have not violated—nor aided or abetted anyone else to violate—the instructions for this exam given by the course staff, including, but not limited to, those on the cover page of this document; and

(4) More generally, I have not committed any act that violates—nor aided or abetted anyone else to violate—UC Berkeley, state, or Federal regulations, during this exam.

(10 Points) In the space below, hand-write the following sentence, verbatim. Then write your name in legible letters, sign, include your full SID, and date before submitting your work:

I have read, I understand, and I commit to adhere to the letter and spirit of the pledge above.

Full Name: ______________________ Signature: ______________________

Date: ______________________ Student ID: ______________________

E3.1 (35 Points) A special QR Decomposition in 2D

Consider the real 2×2 matrix $\mathbf{A}$ given by

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \text{where} \quad ad - bc > 0.$$

Derive the QR Decomposition of $\mathbf{A}$, and put it in the form below

$$\mathbf{A} = \underbrace{\frac{1}{\sqrt{a^2+c^2}} \begin{bmatrix} a & -c \\ c & a \end{bmatrix}}_{\mathbf{Q}} \underbrace{\frac{1}{\sqrt{a^2+c^2}} \begin{bmatrix} a^2+c^2 & ab+cd \\ 0 & ad-bc \end{bmatrix}}_{\mathbf{R}}.$$

Note that it is *not* sufficient for you to show that the matrix identified as $\mathbf{Q}$ above is orthogonal, and that the product of the two matrices labeled above as $\mathbf{Q}$ and $\mathbf{R}$ equals $\mathbf{A}$. That's why we've asked you to *derive* these matrices.

Hint: In $\mathbb{R}^2$, if you want to find a vector orthogonal to the vector $\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, an easy way is to flip the entries and negate one of them. For example, the $\begin{bmatrix} -\beta \\ \alpha \end{bmatrix}$ is orthogonal to $\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$.

It you do not use this hint, you're bound to drown in unwieldy expressions as you try to determine $\boldsymbol{q}_2$, the second column of $\mathbf{Q}$. So, tread carefully.

You may use all of the next page to continue your work.

E3.1 (Continued)

E3.2 (20 Points) Linear Regression

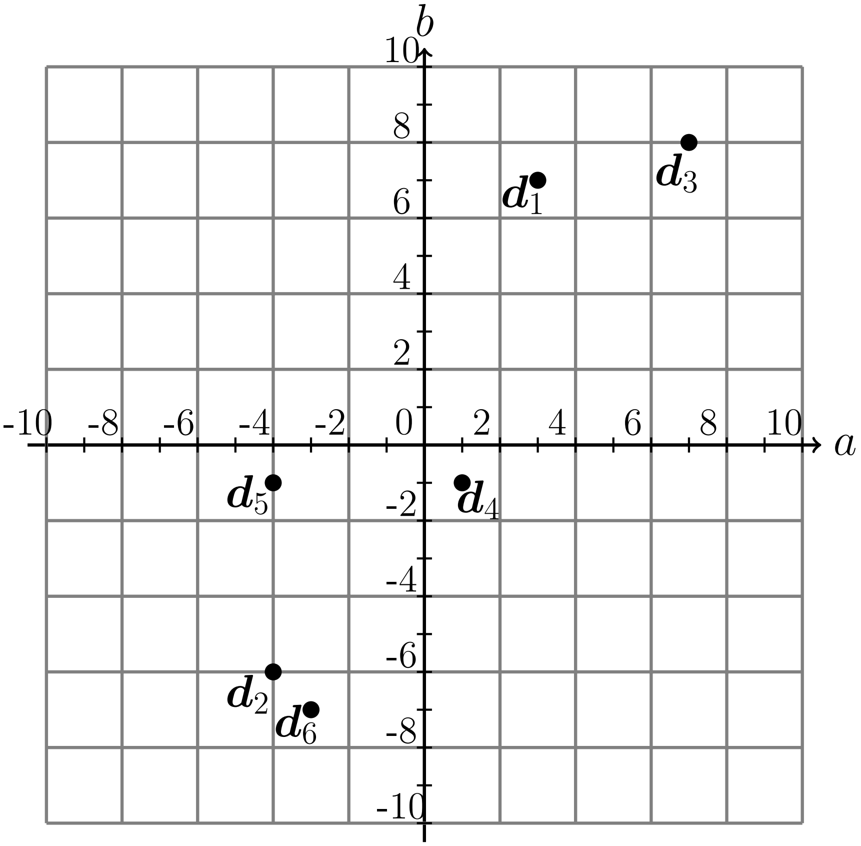
Consider a set of six mean-centered data points $\boldsymbol{d}_k$ given by the columns of the data matrix

$$\mathbf{D} = \begin{bmatrix} 3 & -4 & 7 & 1 & -4 & -3 \\ 7 & -6 & 8 & -1 & -1 & -7 \end{bmatrix}$$

These data points are shown in the figure below. Let each point $\boldsymbol{d}_k = \begin{bmatrix} a_k \\ b_k \end{bmatrix}$, where

$$\boldsymbol{a}^\mathsf{T} = \begin{bmatrix} 3 & -4 & 7 & 1 & -4 & -3 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{b}^\mathsf{T} = \begin{bmatrix} 7 & -6 & 8 & -1 & -1 & -7 \end{bmatrix}$$

denote the first and second rows of the data matrix $\mathbf{D}$, respectively.



We want to regress a line through the data—that is, determine a line that best fits the data in the least-squares sense. The data points are mean-centered already: $\frac{1}{6}\sum_{k=1}^{6} \boldsymbol{d}_k = \mathbf{0}$. Thus, the regression line that we seek crosses the origin—that is, $ax = b$. Notice that a and b here are scalars, because this is the equation that characterizes a straight line in the (a, b)-plane. The unknown slope of the line is denoted by x.

Set up an over-constrained set of six equations in one unknown—that is, $\boldsymbol{a}x = \boldsymbol{b}$—and determine the least-squares solution $\widehat{x}$. Note that $\boldsymbol{a}$ and $\boldsymbol{b}$ here are vectors (as they represent the collected data), and the slope x is the unknown scalar that we're after. Show your work *only* on the next page.

E3.2 (Continued)

E3.3 (55 Points) The Golden State Space

A linear time-invariant system is characterized by the following state-space representation:

$$\boldsymbol{q}[n+1] = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{q}[n] \qquad \text{State-Evolution Equation}$$

$$y[n] = \begin{bmatrix} 1 & 1 \end{bmatrix} \boldsymbol{q}[n] = \mathbf{1}^{\mathsf{T}} \boldsymbol{q}[n]. \qquad \text{Output Equation}$$

The state vector at time n is $\boldsymbol{q}[n] = \begin{bmatrix} q_1[n] \\ q_2[n] \end{bmatrix}$. The state-transition matrix is $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. The signal y, which is called the *output* of the system, is scalar-valued in this setup since it's defined by $y[n] = q_1[n] + q_2[n]$ for all n. The initial state is $\boldsymbol{q}[0] = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

(a) (10 Points) Compute $y[0]$, $y[1]$, $y[2]$, $y[3]$, and $y[4]$. What pattern do you observe?

Hint: Full credit will be given only if, as part of your work, you determine the state vectors $\boldsymbol{q}[1]$, $\boldsymbol{q}[2]$, $\boldsymbol{q}[3]$, and $\boldsymbol{q}[4]$.

E3.3 (Continued)

(b) (10 Points) Show that the eigenvalues of the state-transition matrix **A** are

$$\lambda_1 = \phi \triangleq \frac{1+\sqrt{5}}{2} \qquad \text{and} \qquad \lambda_2 = \widehat{\phi} \triangleq \frac{1-\sqrt{5}}{2},$$

where the symbol $\triangleq$ means ”*is defined as.*”

In one or more subsequent parts, it may or may not be useful for you to know that $\phi\,\widehat{\phi} = -1$, $\phi + \widehat{\phi} = 1$, and $\phi - \widehat{\phi} = \sqrt{5}$.

(c) (5 Points) Explain—in one short English sentence containing no math symbols, and without computing $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$—why the following assertion must be true:

> The eigenvectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ corresponding to
> λ_1 and λ_2, respectively, *must* be orthogonal.

E3.3 (Continued)

(d) (10 Points) Determine a vector in the null space of $\lambda\mathbf{I} - \mathbf{A}$ *for each* of the eigenvalues $\lambda_1 = \phi$ and $\lambda_2 = \widehat{\phi}$, and proceed to show that the following can denote their corresponding eigenvectors:

$$\boldsymbol{v}_1 = \begin{bmatrix} -1 \\ \widehat{\phi} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{v}_2 = \begin{bmatrix} -1 \\ \phi \end{bmatrix}.$$

(e) (5 Points) Let $\mathbf{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ \widehat{\phi} & \phi \end{bmatrix}$. Show that $\mathbf{V}^{-1} = \frac{1}{\sqrt{5}} \begin{bmatrix} -\phi & -1 \\ \widehat{\phi} & 1 \end{bmatrix}$.

E3.3 (Continued)

(f) (10 Points) Use the eigendecomposition $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ to derive a closed-form expression for $\boldsymbol{q}[n] = \mathbf{A}^n\boldsymbol{q}[0]$, in terms of powers of the eigenvalues. Then show that

$$y[n] = \frac{1}{\sqrt{5}}\Big(\phi^{n+1} - \widehat{\phi}^{n+1}\Big).$$

Hint: Determine $\mathbf{V}^{-1}\boldsymbol{q}[0]$ as an initial part of your work.

E3.3 (Continued)

(g) (5 Points) Note that $\phi = \dfrac{1+\sqrt{5}}{2} \approx 1.62 > 1$ is the dominant eigenvalue, whereas $\left|\widehat{\phi}\right| = \left|\dfrac{1-\sqrt{5}}{2}\right| \approx |-0.62| < 1$. Determine

$$\lim_{n\to\infty} \frac{y[n+1]}{y[n]}.$$

E3.4 (35 Points) Basic SVD

Determine the compact SVD of the following matrix.

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

E3.5 (40 Points) Using PCA to Analyze Bacteria

Aakarsh wants to determine how many species of bacteria live in the courtyard of Cory Hall. He collects a sample of cells—of size $n = 50,000$—and takes it back to the lab.

He then runs some experiments on the cells to construct a *gene-expression raw-data matrix* that tells him how much of each gene is present within each cell—a good indicator of the particular type of bacterium that the cell represents. Associated with each of the $n = 50,000$ cells are $m = 6$ gene-expression measurements.

Aakarsh is thus left with a giant $m \times n$ (i.e., a $6 \times 50,000$) gene-expression raw-data matrix $\mathbf{X}$, *each* of whose columns contains raw data for *one* collected cell. In particular, the $m \times n$ raw-data matrix can be written as $\mathbf{X} = \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix}$, where each column $\boldsymbol{x}_k \in \mathbb{R}^m$ contains raw gene-expression data for each of the m genes expressed by cell k. Here, $k \in \{1, \ldots, n\}$.

To get a sense of what the gene-expression data of a typical cell looks like, he averages the columns of the raw-data matrix $\mathbf{X}$ to create a mean vector $\boldsymbol{\mu}$ as follows:

$$\boldsymbol{\mu} = \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{x}_k.$$

Aakarsh is excited to apply Principal Component Analysis (PCA) to make sense of the data!

(a) (5 points) Provide at least one reason why PCA might be useful to analyze the data.

(b) (10 points) Show the preprocessing step that Aakarsh must perform on the matrix $\mathbf{X}$ to produce a matrix $\mathbf{A}$ that is ready for PCA. In particular, express the PCA-ready matrix $\mathbf{A}$ in terms of the raw-data matrix $\mathbf{X}$ and any other relevant parameters given in the problem stem above.

Aakarsh performs an SVD decomposition on his $6 \times 50,000$ PCA-ready gene-expression matrix $\mathbf{A}$, and discovers the following:

- That he needs *only two (2)* principal components to differentiate the various types of bacteria; and
- The $\mathbf{U} \in \mathbb{R}^{6\times 6}$ and $\mathbf{V} \in \mathbb{R}^{50000\times 50000}$ matrices in the SVD decomposition of $\mathbf{A}$ are as shown below (partially for $\mathbf{V}$ due to its size):

$$\mathbf{U} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \qquad \mathbf{V} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 & 0 & \cdots \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

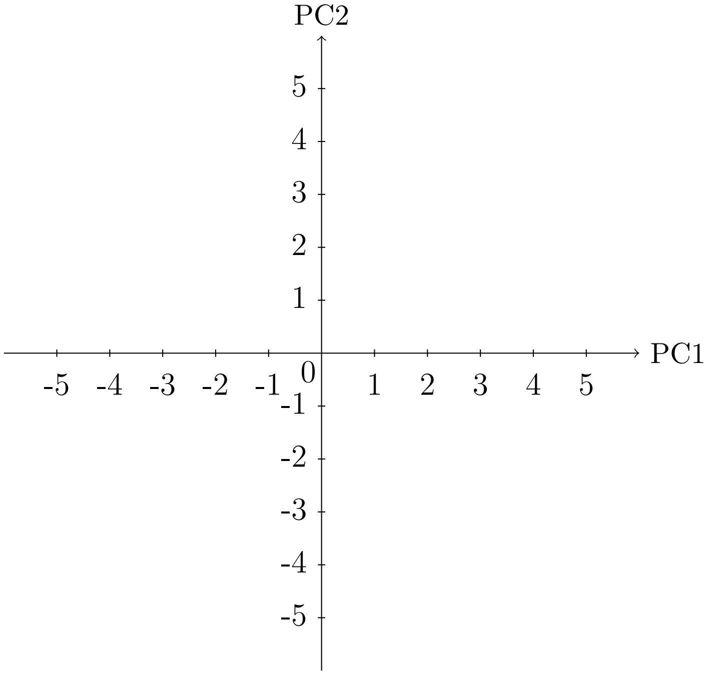
(c) (10 points) Describe qualitatively how the singular values decrease from σ_1 to σ_2 to σ_3, …to σ_6 to support Aakarsh's analysis.

E3.5 (Continued)

(d) (15 points) Aakarsh draws a random sample of six columns from the gene-expression matrix $\mathbf{A}$, to construct the matrix

$$\mathbf{S} = \begin{bmatrix} 1 & 4 & 2 & -3 & -4 & -2 \\ 3 & 4 & -4 & -5 & 3 & -5 \\ 4 & -7 & 3 & 2 & 0 & 6 \\ -4 & -3 & 3 & 4 & -5 & 5 \\ -2 & 1 & 0 & 3 & 4 & 5 \\ -3 & 0 & 1 & 2 & 3 & 4 \end{bmatrix}$$

On the axes below, plot the PCA coordinates of the random cells that were selected for $\mathbf{S}$. To do this, project the columns of $\mathbf{S}$ onto the first two principal components. Based on the distinct clusters you observe, how many different types of bacteria are present in the sample?


PC2
PC1