EE263 Final Exam, December 2025

- This is a 75-minute in-class exam.
- Write your answers in the provided blue book. Make sure to clearly label each page with the problem number you are working on.
- If we ask you to "show" something is true, we are expecting a concise, intuitive derivation *not a formal proof*.
- You may use any books or paper notes.
- You may not use any electronic resources. For example, you cannot use a laptop, tablet or phone, run Julia code, search online, or consult a large language model.
- You may not discuss the exam or course material with others until everyone has taken the exam.
- If you have a question, please ask the staff. We have done our best to make the exam unambiguous. So unless there is a mistake, we are unlikely to say much.
- Please box or otherwise highlight your solution on the page.
- The problems on this exam are split into parts. Each part has a solution under 1/2 page. Thus, if you find yourself producing a complicated and time-consuming solution, stop and reconsider your assumptions.
- Good luck!

1. Multiple choice (10 points). For each of the problems below, select exactly one correct answer. Throughout this question $A \in \mathbb{R}^{m\times n}$ and no assumption on the relationship between m and n is made unless explicitly stated.

a) Suppose $A \in \mathbb{R}^{m\times n}$ with $m > n$ and $\operatorname{rank}(A) = n$. Which statement about the system $Ax = y$ is correct?

(A) For every $y \in \mathbb{R}^m$, there is at least one solution and generally infinitely many.

(B) For every $y \in \mathbb{R}^m$, there is a unique solution.

(C) There exists some y for which $Ax = y$ has no solution, but whenever a solution exists, it is unique.

(D) There exist y_1, y_2 such that $Ax = y_1$ has no solutions and $Ax = y_2$ has infinitely many solutions.

b) Suppose $A \in \mathbb{R}^{m\times n}$ with $m > n$ and full column rank, and let

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

be a full QR factorization, where $Q_1 \in \mathbb{R}^{m\times n}$, $Q_2 \in \mathbb{R}^{m\times(m-n)}$ have orthonormal columns and $R_1 \in \mathbb{R}^{n\times n}$ is upper triangular and invertible. For a given $y \in \mathbb{R}^m$, the orthogonal projection of y onto $\operatorname{range}(A)$ is

(A) $A^\mathsf{T}(AA^\mathsf{T})^{-1}y$.

(B) $(A^\mathsf{T}A)^{-1}A^\mathsf{T}y$.

(C) $Q_1Q_1^\mathsf{T}y$.

(D) $Q_2Q_2^\mathsf{T}y$.

c) Let $A \in \mathbb{R}^{m\times n}$ have singular value decomposition

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^\mathsf{T},$$

where $r = \operatorname{rank}(A)$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Which expression equals the Frobenius norm $\|A\|_F$?

(A) $\displaystyle\sum_{i=1}^{r} \sigma_i$.

(B) $\displaystyle\sum_{i=1}^{r} |\sigma_i|$.

(C) $\displaystyle\sqrt{\sum_{i=1}^{r} \sigma_i^2}$.

(D) $\displaystyle\sqrt{\max_{1\leq i\leq r} \sigma_i^2}$.

d) Let $A \in \mathbb{R}^{m\times n}$ be arbitrary and let $A^\dagger$ denote its pseudo-inverse. For a given $y \in \mathbb{R}^m$, define $x^\star = A^\dagger y$. Which statement is always true?

(A) $x^\star$ is the unique solution of $Ax = y$ for every $y \in \mathbb{R}^m$.

(B) $x^\star$ minimizes $\|x\|_2$ subject to $Ax = y$ for every $y \in \mathbb{R}^m$.

(C) $x^\star$ minimizes $\|Ax - y\|_2$ over all x, and among all minimizers of $\|Ax - y\|_2$ has the smallest norm.

(D) $x^\star$ is in the range of A.

e) Let $A \in \mathbb{R}^{m\times n}$ and $y \in \mathbb{R}^m$. Consider the statement

$$z^\mathsf{T} y = 0 \quad \text{for every } z \in \operatorname{null}(A^\mathsf{T}). \qquad (*)$$

For which y does $(*)$ hold?

(A) $(*)$ holds if and only if $y \in \operatorname{range}(A)$.

(B) $(*)$ holds if and only if $y \in \operatorname{null}(A)$.

(C) $(*)$ holds if and only if $y \in \operatorname{range}(A^\mathsf{T})$.

(D) $(*)$ holds for every $y \in \mathbb{R}^m$.

f) Let $x \sim \mathcal{N}(\mu, \Sigma)$ in $\mathbb{R}^n$ with $\Sigma > 0$. For $\alpha > 0$ define the ellipsoid

$$\mathcal{E}_\alpha = \left\{ x \in \mathbb{R}^n \;\middle|\; (x-\mu)^\mathsf{T} \Sigma^{-1} (x-\mu) \le \alpha \right\}.$$

Which statement about $\mathcal{E}_\alpha$ is correct?

(A) The principal axes of $\mathcal{E}_\alpha$ are the eigenvectors of Σ, and the ith semi-axis length is $\sqrt{\alpha\,\lambda_i(\Sigma)}$.

(B) The principal axes of $\mathcal{E}_\alpha$ are the eigenvectors of Σ^{-1}, and the ith semi-axis length is $\sqrt{\alpha/\lambda_i(\Sigma)}$.

(C) $\mathcal{E}_\alpha$ is a ball of radius $\sqrt{\alpha}$ centered at μ for every positive definite Σ.

(D) Doubling α doubles every semi-axis length.

g) Let $A \in \mathbb{R}^{m\times n}$ have singular value decomposition $A = U\Sigma V^\mathsf{T}$ and pseudo-inverse $A^\dagger = V\Sigma^\dagger U^\mathsf{T}$. Which statement is always true?

(A) $AA^\dagger$ is the orthogonal projector onto $\operatorname{range}(A)$.

(B) $A^\dagger A$ is the orthogonal projector onto $\operatorname{null}(A)$.

(C) $AA^\dagger = I_m$ for every A.

(D) $A^\dagger A = I_n$ for every A.

h) Let L be the Laplacian of an undirected weighted graph with n nodes. Suppose L has exactly 3 linearly independent eigenvectors associated with eigenvalue 0. Which of the following statements is true?

(A) The graph is connected and has 3 self-loops.

(B) The graph has 3 connected components.

(C) The graph has 3 cycles.

(D) The graph has 3 nodes of maximum degree.

2. Warehouse location (10 points). We want to locate two warehouses in $\mathbb{R}^2$ to serve four cities. The city locations are described by two dimensional vectors

$$c_1, c_2, c_3, c_4 \in \mathbb{R}^2,$$

and warehouse k is placed at $x^{(k)} \in \mathbb{R}^2$, $k = 1, 2$. Cities 1 and 2 are served by warehouse 1; cities 3 and 4 are served by warehouse 2.

We choose a quadratic shipping cost

$$J_{\text{ship}}(x) = \|x^{(1)} - c_1\|_2^2 + \|x^{(1)} - c_2\|_2^2 + \|x^{(2)} - c_3\|_2^2 + \|x^{(2)} - c_4\|_2^2,$$

and we also penalize the warehouses being far apart

$$J_{\text{pen}}(x) = \|x^{(1)} - x^{(2)}\|_2^2.$$

For a given penalty parameter $\mu \geq 0$ we define the total cost

$$J_\mu(x) = J_{\text{ship}}(x) + \mu J_{\text{pen}}(x),$$

where

$$x = \begin{bmatrix} x^{(1)} \\ x^{(2)} \end{bmatrix} \in \mathbb{R}^4.$$

a) Find a matrix $A_{\text{ship}} \in \mathbb{R}^{8\times 4}$ and a vector $y \in \mathbb{R}^8$ such that

$$J_{\text{ship}}(x) = \|A_{\text{ship}}x - y\|_2^2.$$

b) Find a matrix $A_{\text{pen}} \in \mathbb{R}^{2\times 4}$ such that

$$J_{\text{pen}}(x) = \|A_{\text{pen}}x\|_2^2.$$

c) Show that for any $\mu \geq 0$ we can write

$$J_\mu(x) = \|\hat{A}x - \hat{y}\|_2^2$$

for some matrix $\hat{A}$ and vector $\hat{y}$ that you should write explicitly in terms of A_{ship}, A_{pen}, y, and μ.

d) Does the problem

$$\min_x J_\mu(x)$$

always have a unique minimizer for every $\mu \geq 0$? Justify your answer.

e) Give a short geometric interpretation of the optimal warehouse locations for $\mu = 0$. Consider the specific city locations

$$c_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad c_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \qquad c_3 = \begin{bmatrix} 0 \\ 4 \end{bmatrix}, \qquad c_4 = \begin{bmatrix} 2 \\ 4 \end{bmatrix},$$

and denote the optimal solution for a given μ by $x^\star(\mu)$. What is $x^\star(0)$?

f) Give a short geometric interpretation of the optimal warehouse locations for $\mu \to \infty$. What is $x^\star(\mu)$ for $\mu \to \infty$ for the specific city locations above?

g) For this problem we can express the solution as

$$x^\star(\mu) = \frac{1}{2} P \begin{bmatrix} I_2 & 0 \\ 0 & \frac{1}{1+\mu} I_2 \end{bmatrix} P^\mathsf{T} A_{\text{ship}}^\mathsf{T} y,$$

where P is some orthogonal matrix that is not a function of μ. Sketch the solutions $x^\star(\mu)$ for $\mu = 0$ and $\mu \to \infty$ and on the same diagram show where the solutions $x^\star(\mu)$ for $\mu \in (0, \infty)$ lie. Justify your answer using the provided expression for $x^\star(\mu)$.

3. Two variations on least-squares regression (10 points). In least-squares regression, we seek to predict a *target* y as a linear function of features $x_1, \ldots, x_p$: that is, $y \approx \beta_1 x_1 + \cdots + \beta_p x_p$. Given a dataset of $N > p$ samples $y^{(i)} \in \mathbb{R}$ and $x^{(i)} \in \mathbb{R}^p$, we define

$$y \in \mathbb{R}^N = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(N)} \end{bmatrix}, \quad X \in \mathbb{R}^{N\times p} = \begin{bmatrix} (x^{(1)})^\top \\ \vdots \\ (x^{(N)})^\top \end{bmatrix}.$$

a) Suppose X is not full rank. (Statisticians call this multicollinearity, but you do not need to know any statistics to solve this problem.) What can you say about the singular values of X? What about the least-squares solution $\beta = (X^\top X)^{-1} X^\top y$?

b) Explain how to find β for this problem such that $\hat{y} = X\beta$ minimizes the objective

$$J(\hat{y}) = \|\hat{y} - y\|^2.$$

The pseudoinverse $X^\dagger$ may be useful. Clearly explain why your method works when X is not full rank.

c) Now we consider a different situation. Given $X \in \mathbb{R}^{N\times p}, N > p$, you compute β using least-squares. Your friend (who has not taken ee263) proposes a shortcut. Instead of fitting the entire β vector using your method from part (b), he fits each β_j to solve the single-variable regression

$$\text{minimize} \sum_{i=1}^{N} (y_i - \beta_j X_{ij})^2$$

Then he arranges the β_j's into a new vector $\tilde{\beta}$. Which of the following is true?

i. Your friend's approach yields a LOWER J value than yours.

ii. Your friend's approach yields a HIGHER J value than yours.

iii. You and your friend have the SAME J value.

Briefly justify your answer.

d) You explain the SVD to your friend, and he decides to modify his approach by computing the SVD $X = U\Sigma V^\top$. Then, for each column u_j in U, he solves the single-variable regression problem from part (c) to obtain a coefficient β_j. Which of the following is true?

i. Your friend's new approach yields a LOWER J value than yours.

ii. Your friend's new approach yields a HIGHER J value than yours.

iii. You and your friend now have the SAME J value.

Briefly justify your answer.

Hint: It may help to think of the objective in terms of the columns of X: that is,

$$J(\hat{y}) = \|y - \hat{y}\|^2 = \|y - (\beta_1 x_1 + \cdots + \beta_p x_p)\|^2$$

where x_i is the ith column of X.

e) In part (d), which of the following is true?

i. Your friend's $\tilde{\beta}$ vector is DIFFERENT from your β.

ii. Your friend's $\tilde{\beta}$ vector is the SAME as your β.

Briefly justify your answer.

4. Regularized least squares and MMSE (10 points). Let $x \in \mathbb{R}^n$ be a random vector with prior distribution $x \sim \mathcal{N}(0, \tau^2 I)$. We observe $y \in \mathbb{R}^m$ according to the linear Gaussian model

$$y = Ax + w,$$

where $A \in \mathbb{R}^{m\times n}$ is known and $w \sim \mathcal{N}(0, \sigma^2 I)$ is Gaussian noise independent of x. We measure $y = y_{\text{meas}}$ and wish to estimate x.

a) Is the random vector $\begin{bmatrix} x \\ y \end{bmatrix}$ jointly Gaussian? Justify your answer using the properties of Gaussian vectors. Write down the covariance matrix for $\begin{bmatrix} x \\ y \end{bmatrix}$.

b) Write down the MMSE estimate of x given $y = y_{\text{meas}}$ in terms of A, τ, σ, and y_{meas}.

c) Now consider the regularized least-squares problem

$$\min_x \quad \|y_{\text{meas}} - Ax\|_2^2 + \lambda\|x\|_2^2$$

with $\lambda > 0$. State or derive the formula for the minimizer of this problem, x_{reg}, in terms of A, λ, and y_{meas}.

d) Show that x_{reg} is equal to the MMSE for appropriate choice of λ and express that λ in terms of τ and σ. You might find the following identity useful:

$$B(I + CB)^{-1} = (I + BC)^{-1}B.$$

e) Assume that $m \geq n$ and that A has full column rank with SVD

$$A = U\Sigma V^{\mathsf{T}}, \qquad \Sigma = \operatorname{diag}(s_1, \ldots, s_n), \;\; s_1 \geq \cdots \geq s_n > 0.$$

Express x_{mmse} in the right singular-vector basis as

$$x_{\text{mmse}} = \sum_{i=1}^{n} \beta_i v_i,$$

and write β_i as functions of s_i, σ, τ, and $u_i^{\mathsf{T}} y_{\text{meas}}$.

f) Let $x_{\text{ls}} = A^{\dagger} y_{\text{meas}}$ be the ordinary least-squares estimate. Show that

$$v_i^{\mathsf{T}} x_{\text{mmse}} = \gamma_i \, v_i^{\mathsf{T}} x_{\text{ls}}$$

for suitable factors γ_i that you should give explicitly. What values can γ_i take? Briefly interpret how γ_i depends on s_i, σ, and τ.

g) Consider the scalar case $n = m = 1$. Express the conditional mean-square error $\mathbb{E}\left((x - x_{\text{mmse}})^2 \mid y = y_{\text{meas}}\right)$ in terms of s, σ, τ, and y_{meas}. Give the limits of this error as $\sigma \to 0$ and as $\sigma \to \infty$, and briefly interpret the results.