Homework #1, EECS 553, F26. Due **Wed. Sep. 9**, by 11:00PM

Notes

- You must write the Engineering Honor Pledge on your exams in this class for them to be graded.
  To review the honor pledge, visit http://honorcode.engin.umich.edu
- Submit your solutions to https://gradescope.com.
- It is essential that you properly match your solution to each problem to the gradescope entry to earn credit.

Problems

The following problems are all designed to practice concepts from ECE 501 and 551 that will be used in this course. Several problems bridge topics of those courses.

1. [5] Honor code

True or false: According to the Engineering Honor Code, as described in the Honor Code Pamphlet on the Honor Council website.

(a) [1] It is the responsibility of faculty members to specify their policies in writing at the beginning of each semester. Students are responsible for understanding these policies and should consult the instructor if they are unclear.

(b) [1] Students who are not members of the College of Engineering are not bound by the policies of the Engineering Honor Code.

(c) [1] This exact wording appears in the pamphlet: *It is a violation of the Honor Code for students to submit, as their own, work that is not the result of their own labor and thoughts.*

(d) [1] During the first Honor Council hearing of a case, a professor presents the evidence.

(e) [1] Submitted work that includes material derived in any way from the efforts of another author, either by direct quotation or paraphrasing, should be fully and properly documented.

2. [7] Syllabus

Read the syllabus (linked on Canvas) and then answer these questions.

(a) [1] T/F: If you have a mandatory obligation on an exam date, you can tell the teaching staff one week before the exam.

(b) [1] T/F: One homework score will be dropped.

(c) [1] T/F: Late homework can be submitted up to 48 hours after the deadline.

(d) [1] T/F: The use of Great Lakes for the project is required.

(e) [1] T/F: You should email the professor with your questions about the course.

(f) [2] The GenAI section of the syllabus has a link to a concise document by Dr. M on advice about using AI tools. List *any one* of the "rationalizations" that you think some students in this course may be at risk of using.

3. [12] Probability and expectation

Let random variables X and Y be jointly distributed with distribution $\mathsf{p}(x, y)$. You can assume that they are jointly discrete so that $\mathsf{p}(x, y)$ is the probability mass function (pmf). Show the following results using fundamental properties of probability and random variables.

(a) [3] $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$, where $\mathbb{E}[X] = \sum_x x\, \mathsf{p}(x)$ denotes statistical expectation of X and $\mathbb{E}[X \mid Y] = \sum_x x\, \mathsf{p}(x \mid Y)$ denotes conditional expectation of X given Y.

(b) [3] $\mathbb{E}[\mathbb{I}_{\{X \in \mathcal{C}\}}] = \mathsf{P}\{X \in \mathcal{C}\}$, where $\mathbb{I}_{\{X \in \mathcal{C}\}}$ is the **indicator function** of an arbitrary set $\mathcal{C}$, *i.e.*, $\mathbb{I}_{\{X \in \mathcal{C}\}} = 1$ if $X \in \mathcal{C}$ and 0 otherwise.

(c) [3] If X and Y are **independent**, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

(d) [3] If X and Y take values in $\{0, 1\}$ and $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, then X and Y are independent.

4. [10] **Random vectors / covariance / PSD**
   (a) [4] The code line `v0 = mu0 .+ 1*randn(2,n0)` in this demo generates IID random vectors as columns of a matrix variable `v0`. Write the mathematical expression for the pdf of one of those random vectors $\boldsymbol{v}$.
   (b) [4] Determine analytically the covariance of the random vector $\boldsymbol{y} \triangleq \boldsymbol{A}\boldsymbol{v}$ where $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 3 & 0 \end{bmatrix}$.
   (c) [2] Is that covariance a positive definite matrix? Prove or disprove.

5. [6] **Covariance matrices for generative models**

   Let $\boldsymbol{x}$ denote a standard normal random vector with mean zero and covariance $\mathsf{Cov}\{\boldsymbol{x}\} = \boldsymbol{I}_2$.
   (a) [4] Find a matrix $\boldsymbol{A}$ such that the random vector defined by $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ has the covariance matrix $\mathsf{Cov}\{\boldsymbol{y}\} = \boldsymbol{\Sigma}_y = \begin{bmatrix} 1 & 3 \\ 3 & 25 \end{bmatrix}$.
   Numerical values with explanation are acceptable. (Try to solve it in a way that would work even for a much larger case.)
   (b) [2] Is your $\boldsymbol{A}$ unique? Discuss.

6. [12] **PCA-related matrix methods**
   (a) [3] Let $\boldsymbol{Q}$ denote a $d \times k$ semi-unitary matrix. Express concisely the pseudo-inverse of $\boldsymbol{I} - \boldsymbol{Q}\boldsymbol{Q}^\top$.
   (b) [3] Determine $\underset{\boldsymbol{\mu} \in \mathbb{R}^d}{\arg\min} \sum_{i=1}^{n} \|\boldsymbol{\mu} - \boldsymbol{x}_i\|_2^2$, for $\boldsymbol{x}_i \in \mathbb{R}^d$.
   (c) [3] Determine $\underset{\boldsymbol{\mu} \in \mathbb{R}^d}{\arg\min} \sum_{i=1}^{n} \left\|\boldsymbol{P}_{\boldsymbol{Q}}^{\perp}(\boldsymbol{\mu} - \boldsymbol{x}_i)\right\|_2^2$,
   where $\boldsymbol{P}_{\boldsymbol{Q}}^{\perp}$ denotes the projection onto the orthogonal complement of the range of $\boldsymbol{Q}$.
   (d) [3] Are the solutions to the 3 preceding parts unique?

7. [5] **PSD matrices**

   Use two different GenAI tools of your choice to prove that **covariance matrices** are PSD. Covariance matrices are matrices of the form $\mathbb{E}[(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])^\top]$, where $\boldsymbol{x} \in \mathbb{R}^d$ is a random vector.

   Include both transcripts, including your prompts, and write a few sentences to compare the two outputs. Did you notice any significant similarities or differences in the two outputs? Do you find that surprising? Why would you choose one over the other? Did you learn anything, or is there anything you especially like about the way they answered? How might you edit it if you were to include it in your own work?

8. [20] **Low-rank approximation**

Approximating a matrix by another matrix of lower rank is an important operation for dimensionality reduction, and low-rank adaptation (LoRA) methods are an active ML topic.

(a) [10] Write a JULIA or Python function `rank1` that computes a rank (at most) 1 approximation to its input matrix that is "best" in the sense of minimizing the Frobenius norm of the difference between the original matrix and its approximation. (The last character there is the digit 1 not the letter l.)
In all such assignments, you may use JULIA or Python. There will be code-related exam questions and you will be able to use either of these two languages.
We recommend using Colab, which supports both Python and JULIA without having to install anything locally.
To start a new JULIA notebook, use this Colab URL. Or open any Colab notebook, then select `Runtime` in the top menu, select `Change runtime type`, and select `Julia`.
Your submitted file should be named `rank1.jl` or `rank1.py` and should contain the following function:

```
"""
    B = rank1(A)

In:
* `A` `M x N` matrix

Out:
* `B` best rank-1 approximation to A in Frob. norm sense
"""
function rank1(A::AbstractMatrix)
```

Submit your solution to autograder.io
The comment style here is called a JULIA `docstring` and is the preferred way to document code. It supports markdown syntax. With this approach, you can type `?rank1` at the REPL or in a jupyter notebook for formatted documentation.
In Python, the docstring follows the function name instead of preceding it. Because JULIA is Prof. Fessler's preferred language, function prototypes will typically be provided in JULIA and you may convert them to Python for your own solutions. (BTW, Python fans may think of this as HW0 instead of HW1.)

(b) [5] Write code for at least two *test cases* that check the correctness of your `rank1` function. At least one test case should use a rank-1 matrix, and at least one test case should use an input matrix whose rank is larger than one, but for which you can still analytically determine the best rank-1 approximation to check your numerical solution. (In JULIA, the `isapprox` function is useful here.)
Unit testing is an fundamental part of modern software engineering. Several HW problems will require you to provide unit tests, both for encouraging best practices, and to aid your understanding of the connections between the math and the code.
Submit a screenshot of your test case code to gradescope, unless we post a Canvas message giving autograder instructions.

(c) [3] Is your code's best approximation unique? Discuss.

(d) [2] Does your function *always* return a matrix of rank 1? Explain why or why not.

Non-graded problems

The following problems also review important concepts from ECE 501 and/or ECE 551 but are essentially identical to homework problems assigned in a previous 501/551 term, so they will not be graded here. We will provide solutions for self check and future reference. Exam problems and future HW problems may use material from these problems.

9. [5] **Maximum likelihood estimation**

Let $x_1, \ldots, x_n$ be IID samples from the uniform distribution $U[0, \theta]$ where $\theta > 0$.
Express the maximum likelihood estimate $\hat{\theta} = \arg\max_\theta \mathsf{p}(\boldsymbol{x}; \theta)$ in terms of those samples.
Hint. Differentiation is unlikely to be helpful here.

10. [24] Prove each of the following statements, or provide a counterexample. Throughout assume that $f(\boldsymbol{x})$ is **convex** on $\mathbb{F}^N$.
Optional: in cases where the answer is False, try to find a small modification of the statement that would make it true.
These properties of convex functions are used extensively in optimization.
(a) [0] Optional. If $g(\boldsymbol{x})$ is convex, then $h(\boldsymbol{x}) \triangleq f(\boldsymbol{x}) + g(\boldsymbol{x})$ is convex. (sum)
(b) [0] Optional. If $g(\boldsymbol{x})$ is **strictly convex**, then $h(\boldsymbol{x}) \triangleq f(\boldsymbol{x}) + g(\boldsymbol{x})$ is strictly convex. (sum)
(c) [3] If $b \in \mathbb{R}$ and $\boldsymbol{v} \in \mathbb{F}^N$, then $f(\boldsymbol{x}) + \text{real}\{\boldsymbol{v}'\boldsymbol{x}\} + b$ is convex. (affine)
(d) [3] If $\alpha \in \mathbb{R}$ then $\alpha f(\boldsymbol{x})$ is convex. (scaling)
(e) [3] If $\boldsymbol{A} \in \mathbb{F}^{M \times N}$ then $g(\boldsymbol{y}) \triangleq f(\boldsymbol{A}'\boldsymbol{y})$ is convex on $\mathbb{F}^M$.
(f) [3] If g is strictly convex on $\mathbb{F}^N$ and if $\boldsymbol{A} \in \mathbb{F}^{N \times N}$, then $h(\boldsymbol{x}) \triangleq g(\boldsymbol{A}\boldsymbol{x})$ is strictly convex on $\mathbb{F}^N$.
(g) [3] Any **norm** $\|\cdot\|$ for $M \times N$ matrices is convex on $\mathbb{F}^{M \times N}$.
(h) [3] If $\psi : \mathbb{F} \to \mathbb{R}$ is convex, then the **additively separable** function $R(\boldsymbol{x}) \triangleq \mathbf{1}_N' \, \psi.(\boldsymbol{x})$ is convex on $\mathbb{F}^N$.
(i) [3] Both $\|\boldsymbol{x}\|_2^2$ and $\|\boldsymbol{x}\|_2$ are **strictly convex** on $\mathbb{F}^N$.
(j) [3] If $g : \mathbb{R} \mapsto \mathbb{R}$ is convex and non-decreasing, then $h(\boldsymbol{x}) = g(f(\boldsymbol{x}))$ is convex. (composition)

11. [0]
(a) Show that if $\boldsymbol{W}$ is a $N \times N$ **positive definite matrix**, then the weighted Euclidean norm $\|\boldsymbol{x}\|_{\boldsymbol{W}} = \sqrt{\boldsymbol{x}'\boldsymbol{W}\boldsymbol{x}}$ is a valid **norm** on $\mathbb{F}^N$.
(b) Now show the converse, *i.e.*, if $\|\boldsymbol{x}\|_{\boldsymbol{W}} = \sqrt{\boldsymbol{x}'\boldsymbol{W}\boldsymbol{x}}$ is a valid **norm** on $\mathbb{F}^N$, and $\boldsymbol{W}$ is (Hermitian) symmetric, then $\boldsymbol{W} \succ \mathbf{0}$.

12. [0] This problem examines import properties of **positive semidefinite** and **positive definite** matrices.

Recall that $\boldsymbol{B}'\boldsymbol{B} \succeq \mathbf{0}$ for any matrix $\boldsymbol{B}$.
(a) Show that $\boldsymbol{B}'\boldsymbol{B} \succ \mathbf{0}$ if $\boldsymbol{B}$ has **full column rank**.
(b) Show that $\boldsymbol{A} \succ \mathbf{0}$ implies $\boldsymbol{A}$ is invertible.
(c) Show $\boldsymbol{A} \succeq \mathbf{0}$ and $\boldsymbol{B} \succeq \mathbf{0} \implies \boldsymbol{A} + \boldsymbol{B} \succeq \mathbf{0}$.
(d) $\boldsymbol{A} \succ \mathbf{0}$ and $\boldsymbol{B} \succeq \mathbf{0} \implies \boldsymbol{A} + \boldsymbol{B} \succ \mathbf{0}$.
(e) Show $\boldsymbol{A}'\boldsymbol{A} + \boldsymbol{B}'\boldsymbol{B}$ is invertible if $\boldsymbol{B}$ has full column rank
(This property is relevant to **regularized least-squares** problems.)
(f) Show $\boldsymbol{A}'\boldsymbol{A} + \boldsymbol{B}'\boldsymbol{B}$ is invertible if $\mathcal{N}(\boldsymbol{A}) \cap \mathcal{N}(\boldsymbol{B}) = \{\mathbf{0}\}$, i.e., if $\boldsymbol{A}$ and $\boldsymbol{B}$ have disjoint **null spaces** other than $\mathbf{0}$.
(g) Show that $\boldsymbol{A} \succ \mathbf{0}, \ \boldsymbol{B} \succ \mathbf{0} \implies \boldsymbol{B}\boldsymbol{A}\boldsymbol{B} \succ \mathbf{0}$.