EECS 16B Designing Information Devices and Systems II
Spring 2021 UC Berkeley

Homework 11

This homework is due on Friday, April 9, 2021, at 11:00PM. Self-grades and HW Resubmission are due on Tuesday, April 13, 2021, at 11:00PM.

1. **Reading Lecture Notes**

   Staying up to date with lectures is an important part of the learning process in this course. Here are links to the notes that you need to read for this week: Note 12

   (a) Consider two vectors $\vec{x} \in \mathbb{R}^m$ and $\vec{y} \in \mathbb{R}^n$, what is the dimension of the matrix $\vec{x}\vec{y}^\top$ and what is the rank of it?

   (b) Consider a matrix $A \in \mathbb{R}^{m \times n}$ and the rank of A is r. Suppose its SVD is $A = U\Sigma V^\top$ where $U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$, and $V \in \mathbb{R}^{n \times n}$. Can you write A in terms of the singular values of A and outer products of the columns of U and V?

2. Proofs

In this problem we will review some of the important proofs we saw in the lecture as practice. Let's define $S = A^\top A$ where A is an arbitrary $m \times n$ matrix. $V = \begin{bmatrix} \vec{v}_1 \cdots \vec{v}_n \end{bmatrix}$ is the matrix of normalized eigenvectors of S with eigenvalues $\lambda_1, \cdots, \lambda_n$.

(a) Let $\vec{v_1}, \cdots, \vec{v_r}$ be those normalized eigenvectors that correspond to non-zero eigenvalues (i.e. $\|\vec{v}_i\| = 1$ and $\lambda_i \neq 0$ for $i = 1, \cdots, r$). Show that $\|A\vec{v}_i\|^2 = \lambda_i$.

(b) Following the assumptions in part (a), show that $A\vec{v}_i$ is orthogonal to $A\vec{v}_j$.

(c) Show that if $V \in \mathbb{R}^{n\times n}$ is an orthonormal square matrix, then $||\vec{x}||^2 = ||V\vec{x}||^2$ for all $\vec{x} \in \mathbb{R}^n$. ***Hint:*** *Write the norm as an inner product instead of trying to do this elementwise.*

3. SVD

(a) Consider the matrix

$$A = \begin{bmatrix} -1 & 1 & 5 \\ 3 & 1 & -1 \\ 2 & -1 & 4 \end{bmatrix}.$$

Observe that the columns of matrix A are mutually orthogonal with norms $\sqrt{14}$, $\sqrt{3}$, $\sqrt{42}$.

Verify numerically that columns $\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ **and** $\begin{bmatrix} 5 \\ -1 \\ 4 \end{bmatrix}$ **are orthogonal to each other.**

(b) **Write** $A = BD$**, where** B **is an orthonormal matrix and** D **is a diagonal matrix.** What is B? What is D?

(c) **Write out a singular value decomposition of** $A = U\Sigma V^{\top}$ **using the previous part.** Note the ordering of the singular values in Σ should be from the largest to smallest.

(HINT: From the previous part $A = BDI_{3\times 3}$*. Now, re-order to have eigenvalues in decreasing order.)*

4. The Moore-Penrose pseudoinverse for "wide" matrices

Say we have a set of linear equations given by $A\vec{x} = \vec{y}$. If A is invertible, we know that the solution for $\vec{x}$ is $\vec{x} = A^{-1}\vec{y}$. However, what if A is not a square matrix? In 16A, you saw how this problem could be approached for tall "standing up" matrices A where it really wasn't possible to find a solution that exactly matches all the measurements, using linear least-squares. The linear least-squares solution gives us a reasonable answer that asks for the "best" match in terms of reducing the norm of the error vector.

This problem deals with the other case — when the matrix A is wide — with more columns than rows. In this case, there are generally going to be lots of possible solutions — so which should we choose? Why? We will walk you through the **Moore-Penrose pseudoinverse** that generalizes the idea of the matrix inverse and is derived from the singular value decomposition.

This approach to finding solutions complements the OMP approach that you learned in 16A and that we used earlier in 16B in the context of outlier removal during system identification.

(a) Say you have the matrix

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}.$$

To find the Moore-Penrose pseudoinverse we start by calculating the SVD of A. That is to say, calculate U, Σ, V such that

$$A = U\Sigma V^\top$$

where U and V are orthonormal matrices.

Here we will give you that the decomposition of A is:

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

where:

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix}$$

$$V^\top = \begin{bmatrix} 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

It is a good idea to be able to calculate the SVD yourself as you may be asked to solve similar questions on your own in the exam. For this subpart, verify that this decomposition of A is correct. You may use a computer to do the matrix multiplication if you want, but it is better to verify by hand.

(b) Let us now think about what the SVD does. Consider a rank m matrix $A \in \mathbb{R}^{m\times n}$, with $n > m$. Let $U \in \mathbb{R}^{m\times m}, \Sigma \in \mathbb{R}^{m\times n}, V \in \mathbb{R}^{n\times n}$. Let us look at matrix A acting on some vector $\vec{x}$ to give the result $\vec{y}$. We have

$$A\vec{x} = U\Sigma V^\top \vec{x} = \vec{y}.$$

We can think of $V^\top \vec{x}$ as a rotation of the vector $\vec{x}$, then Σ as a scaling, and U as another rotation (multiplication by an orthonormal matrix does not change the norm of a vector, try to verify this for yourself). We will try to "reverse" these operations one at a time and then put them together to construct the Moore-Penrose pseudoinverse.

If U "rotates" the vector $\left(\Sigma V^\top\right)\vec{x}$, what operator can we derive that will undo the rotation?

(c) **Derive a matrix $\widetilde{\Sigma}$ that will "unscale", or undo the effect of Σ where it is possible to undo.** Recall that Σ has the same dimensions as A.

Hint: Consider

$$\widetilde{\Sigma} = \begin{bmatrix} \frac{1}{\sigma_0} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_1} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_{m-1}} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

Then compute $\widetilde{\Sigma}\Sigma$.

(d) **Derive an operator that would "unrotate" by $V^\top$.**

(e) **Try to use this idea of "unrotating" and "unscaling" to write out an "inverse", denoted as $A^\dagger$.** That is to say,

$$\vec{\hat{x}} = A^\dagger \vec{y}$$

where $\vec{\hat{x}}$ is the recovered $\vec{x}$. The reason why the word inverse is in quotes (or why this is called a pseudo-inverse) is because we're ignoring the "divisions" by zero and $\vec{\hat{x}}$ isn't exactly equal to $\vec{x}$.

(f) **Use $A^\dagger$ to solve for a vector $\vec{x}$ in the following system of equations.**

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \vec{x} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

(g) Now we will see why this matrix, $A^\dagger = V\widetilde{\Sigma}U^\top$, is a useful proxy for the matrix inverse in such circumstances. **Show that the solution given by the Moore-Penrose pseudoinverse satisfies the minimality property that if $\vec{\hat{x}} = A^\dagger \vec{y}$ is the pseudo-inverse solution to $A\vec{x} = \vec{y}$, then $\vec{\hat{x}}$ has no component in the nullspace of A.**

Hint: To show this recall that for any vector $\vec{x}$, the vector $V^{-1}\vec{x}$ represents the coordinates of $\vec{x}$ in the basis of the columns of V. Compute $V^{-1}\vec{x}$ and show that the last $(n-m)$ rows are all zero.

This minimality property is useful in many applications. You saw a control application in lecture. This is also used all the time in machine learning, where it is connected to the concept behind what is called ridge regression or weight shrinkage.

(h) Consider a generic wide matrix A. We know that A can be written using $A = U\Sigma V^\top$ where U and V each are the appropriate size and have orthonormal columns, while Σ is the appropriate size and is a diagonal matrix — all off-diagonal entries are zero. Further assume that the rows of A are linearly independent. **Prove that $A^\dagger = A^\top (AA^\top)^{-1}$.**

(HINT: Just substitute in $U\Sigma V^\top$ for A in the expression above and simplify using the properties you know about U, Σ, V. Remember the transpose of a product of matrices is the product of their transposes in reverse order: $(CD)^\top = D^\top C^\top$.)

5. Frobenius Norm

In this problem we will investigate the basic properties of the Frobenius norm.

Similar to how the norm of vector $\vec{x} \in \mathbb{R}^N$ is defined as $\|x\| = \sqrt{\sum_{i=1}^{N} x_i^2}$, the Frobenius norm of a matrix $A \in \mathbb{R}^{N\times N}$ is defined as

$$\|A\|_F = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} |A_{ij}|^2}.$$

A_{ij} is the entry in the i^{th} row and the j^{th} column. This is basically the norm that comes from treating a matrix like a big vector filled with numbers.

(a) With the above definitions, **show that for a 2×2 matrix A:**

$$\|A\|_F = \sqrt{\mathbf{Tr}\{A^\top A\}}.$$

Think about how this generalizes to $n \times n$ matrices. *Note:* The trace of a matrix is the sum of its diagonal entries. For example, let $A \in \mathbb{R}^{N\times N}$, then,

$$\text{Tr}\{A\} = \sum_{i=1}^{N} A_{ii}$$

(b) **Show that if U and V are square orthonormal matrices, then**

$$\|UA\|_F = \|AV\|_F = \|A\|_F\,.$$

(HINT: Use the trace interpretation from part (a).)

(c) **Use the SVD decomposition to show that $\|A\|_F = \sqrt{\sum_{i=1}^{N} \sigma_i^2}$, where $\sigma_1, \cdots, \sigma_N$ are the singular values of A.**

(HINT: The previous part might be quite useful.)

6. SVD from the other side

In lecture, we thought about the SVD for a wide matrix M with n rows and $m > n$ columns by looking at the big $m \times m$ symmetric matrix $M^T M$ and its eigenbasis. This question is about seeing what happens when we look at the small $n \times n$ symmetric matrix $Q = MM^T$ and its orthonormal eigenbasis $U = [\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_n]$ instead. Suppose we have sorted the eigenvalues so that the real eigenvalues $\widetilde{\lambda}_i$ are sorted in descending order where $\widetilde{\lambda}_i \vec{u}_i = Q\vec{u}_i$.

(a) **Show that $\widetilde{\lambda}_i \geq 0$.**
(HINT: You want to involve $\vec{u}_i^T \vec{u}_i$ somehow.)

(b) Suppose that we define $\vec{w}_i = \frac{M^T \vec{u}_i}{\sqrt{\widetilde{\lambda}_i}}$ for all i for which $\widetilde{\lambda}_i > 0$. Suppose that there are ℓ such eigenvalues. **Show that $W = [\vec{w}_1, \vec{w}_2, \cdots, \vec{w}_\ell]$ has orthonormal columns.**

7. Homework Process and Study Group

Citing sources and collaborators are an important part of life, including being a student!
We also want to understand what resources you find helpful and how much time homework is taking, so we can change things in the future if possible.

(a) **What sources (if any) did you use as you worked through the homework?**

(b) **If you worked with someone on this homework, who did you work with?**
List names and student ID's. (In case of homework party, you can also just describe the group.)

Contributors:

- Kourosh Hakhamaneshi.
- Elena Jia.
- Siddharth Iyer.
- Justin Yim.
- Anant Sahai.
- Gaoyue Zhou.
- Aditya Arun.