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?

Solution: Let $\vec{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}^\top$, then

$$\vec{x}\vec{y}^\top = \begin{bmatrix} | & | & & | \\ y_1\vec{x} & y_2\vec{x} & \dots & y_n\vec{x} \\ | & | & & | \end{bmatrix}$$

We can see that each column of $\vec{x}\vec{y}^\top$ is a multiple of $\vec{x}$, thus it has rank 1.

(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?

Solution: We have $A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top$ where $\vec{u}_i$ and $\vec{v}_i$ are the columns of U and V. Note that we only sum to r since A has rank r and hence it has r non-zero singular values $\sigma_1, ..., \sigma_r$. This is the outer product form of the SVD.

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 $\|A\vec{v_i}\|^2 = \lambda_i$.

Solution: We know that v_i is an eigenvector of S, so we can start with:

$$S\vec{v_i} = \lambda_i \vec{v_i}$$
$$\vec{v_i}^\top S\vec{v_i} = \lambda_i \vec{v_i}^\top \vec{v_i}.$$
$$\vec{v_i}^\top S\vec{v_i} = \lambda_i \|\vec{v_i}\|^2.$$

On the second step, we multiplied both sides with $\vec{v_i}^\top$. Since $\|\vec{v_i}\|^2 = 1$ the right hand side will simplify to λ_i. By substituting $S = A^\top A$:

$$\vec{v_i}^\top A^\top A\vec{v_i} = \lambda_i$$
$$(A\vec{v_i})^\top A\vec{v_i} = \lambda_i$$
$$\|A\vec{v_i}\|^2 = \lambda_i.$$

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

Solution: To show that two vectors are orthogonal we have to take the inner product and show that it is zero:

$$(A\vec{v_i})^\top A\vec{v_j} = \vec{v_i}^\top A^\top A\vec{v_j} = \vec{v_i}^\top S\vec{v_j}.$$

We know that $S\vec{v_j} = \lambda_j \vec{v_j}$. Therefore,

$$\vec{v_i}^\top S\vec{v_j} = \lambda_j \underbrace{\vec{v_i}^\top \vec{v_j}}_{=0(\text{since } \vec{v}_i \perp \vec{v}_j)} = 0.$$

(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.*

Solution:

$$||V\vec{x}||^2 =< V\vec{x}, V\vec{x} >= \vec{x}^\top V^\top V\vec{x} = \vec{x}^\top I_{n\times n}\vec{x} = ||\vec{x}||^2.$$

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.**

Solution: Taking the inner product of the two vectors, we have

$$\left\langle \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 \\ -1 \\ 4 \end{bmatrix} \right\rangle = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}^\top \begin{bmatrix} 5 \\ -1 \\ 4 \end{bmatrix} = 5 - 1 - 4 = 0.$$

So the two columns 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?

Solution: We compute the norm for each column and divide each column by its norm to obtain matrix B. Matrix D is formed by placing the norms on the diagonal.

$$B = \begin{bmatrix} -\frac{1}{\sqrt{14}} & \frac{1}{\sqrt{3}} & \frac{5}{\sqrt{42}} \\ \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{42}} \\ \frac{2}{\sqrt{14}} & -\frac{1}{\sqrt{3}} & \frac{4}{\sqrt{42}} \end{bmatrix}$$

$$D = \begin{bmatrix} \sqrt{14} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{42} \end{bmatrix}$$

(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.)

Solution:

$$A = BD = BDI = \begin{bmatrix} -\frac{1}{\sqrt{14}} & \frac{1}{\sqrt{3}} & \frac{5}{\sqrt{42}} \\ \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{42}} \\ \frac{2}{\sqrt{14}} & -\frac{1}{\sqrt{3}} & \frac{4}{\sqrt{42}} \end{bmatrix} \begin{bmatrix} \sqrt{14} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{42} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Reordering the singular values and corresponding left and right singular vectors, we have the SVD:

$$\begin{bmatrix} \frac{5}{\sqrt{42}} & -\frac{1}{\sqrt{14}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{42}} & \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{3}} \\ \frac{4}{\sqrt{42}} & \frac{2}{\sqrt{14}} & -\frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{42} & 0 & 0 \\ 0 & \sqrt{14} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

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.

Solution: You may give yourself full credit for this subpart if you showed the multiplication of the matrices out to get A, or stated that you did this using a computer. Though you did not have to do any work for this sub-part the following solutions will walk you through how to solve for the SVD:

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

.

$$AA^\top = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}.$$

Which has characteristic polynomial $\lambda^2 - 6\lambda + 8 = 0$, producing eigenvalues 4 and 2. Solving $Av = \lambda_i v$ produces eigenvectors $[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}]^\top$ and $[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^\top$ associated with eigenvalues 4 and 2 respectively. The singular values are the square roots of the eigenvalues of $AA^\top$, so

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

and

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

We can then solve for the $\vec{v}$ vectors using $A^\top \vec{u}_i = \sigma_i \vec{v}_i$, producing $\vec{v}_1 = [0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^\top$ and $\vec{v}_2 = [1, 0, 0]^\top$. The last $\vec{v}$ must be orthonormal to the other two, so we can pick $[0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^\top$.

The SVD is:

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

(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?

Solution: By orthonormality, we know that $U^\top U = UU^\top = I$. Therefore, $U^\top$ undoes 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 $\widetilde{\Sigma}\Sigma$.

Solution: If you observe the equation:

$$\Sigma\vec{x} = U^\top\vec{y} = \vec{\tilde{y}}, \tag{1}$$

you can see that $\sigma_i x_i = \tilde{y}_i$ for $i = 0, ..., m-1$, which means that to obtain x_i from y_i, we need to multiply y_i by $\frac{1}{\sigma_i}$. For any $i > m-1$, the information in x_i is lost by multiplying with 0. If the corresponding $\tilde{y}_i \neq 0$, there is no way of solving this equation. No solution exists, and we have to accept an approximate solution. If the corresponding $\tilde{y}_i = 0$, then any x_i would still work. Either way, it is reasonable to just say x_i is 0 in the case that $\sigma_i = 0$. That's why we can legitimately pad 0s in the bottom of $\widetilde{\Sigma}$ given below:

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

To recover $\vec{x}$, let's try to transform $\vec{y}$ with $\widetilde{\Sigma}$:

$$\begin{aligned} \widetilde{\Sigma}\vec{y} &= \widetilde{\Sigma}\Sigma\vec{x} \\ &= \begin{bmatrix} \frac{1}{\sigma_0} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_1} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_{m-1}} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \sigma_0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & 0 \\ 0 & 0 & 0 & \sigma_{m-1} & 0 & \dots & 0 \end{bmatrix} \vec{x} \\ &= \begin{bmatrix} \Sigma_1^{-1} \\ 0_{(n-m)\times m} \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0_{m\times(n-m)} \end{bmatrix} \vec{x} \\ &= \begin{bmatrix} I_{m\times m} & 0_{m\times(n-m)} \\ 0_{(n-m)\times m} & 0_{(n-m)\times(n-m)} \end{bmatrix} \vec{x} \\ &= \begin{bmatrix} x_1 \\ \vdots \\ x_m \\ 0 \\ \vdots \\ 0 \end{bmatrix} \end{aligned}$$

Therefore, we are able to recover all x_i for $1 \leq i \leq m$ with $\widetilde{\Sigma}$. The rest of the entries of $\vec{x}$ are lost when multiplied by the zeros in Σ and thus it's not possible to recover them.

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

Solution: By orthonormality, we know that $V^\top V = VV^\top = I$. Therefore, V undoes the rotation.

(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 $\vec{x}$.

Solution: You may give yourself full credit for this subpart if you write out $A^\dagger = V\widetilde{\Sigma}U^\top$.

To see how we get this, recall that $\vec{y} = A\vec{x} = U\Sigma V^\top \vec{x}$. We get $\vec{y}$ by first rotating $\vec{x}$ by $V^\top$, then scaling the resulting vector by Σ, and finally rotating by U. To undo the operations, we should first "unrotate" $\vec{y}$ by $U^\top$, then "unscale" it by $\widetilde{\Sigma}$, and finally "unrotate" it by V. Therefore, we have

$$\vec{\hat{x}} = V\widetilde{\Sigma}U^\top \vec{y}$$

which leads to the definition $A^\dagger = V\widetilde{\Sigma}U^\top$ to be the pseudo-inverse of A. Of course, nothing can possibly be done for the information that was destroyed by the nullspace of A — there is no way to recover any component of the true $\vec{x}$ that was in the nullspace of A. However, we can get back everything else.

$$\begin{aligned}
\vec{y} &= A\vec{x} = U\Sigma V^\top \vec{x} & \\
U^\top \vec{y} &= U^\top U\Sigma V^\top \vec{x} & \text{Unrotating by } U \\
U^\top \vec{y} &= I_{m\times m}\Sigma V^\top \vec{x} & \text{Unrotating by } U \\
U^\top \vec{y} &= \Sigma V^\top \vec{x} & \text{Unrotating by } U \\
\widetilde{\Sigma}U^\top \vec{y} &= \widetilde{\Sigma}\Sigma V^\top \vec{x} & \text{Unscaling by } \widetilde{\Sigma} \\
V\widetilde{\Sigma}U^\top \vec{y} &= V\widetilde{\Sigma}\Sigma V^\top \vec{x} & \text{Unrotating by } V
\end{aligned}$$

It's worth noting that $\vec{\hat{x}} = V\widetilde{\Sigma}U^\top \vec{y} = V\widetilde{\Sigma}\Sigma V^\top \vec{x}$ has no component in the null space of A. To see why this is the case, recall that $\{\vec{v}_{m+1}, ..., \vec{v_n}\}$ is a basis of the null space of A, and $V^\top \vec{x}$ is the coordinate of $\vec{x}$ in the V basis.

From previous parts, we know that

$$\widetilde{\Sigma}\Sigma V^\top \vec{x} = \begin{bmatrix} (V^\top \vec{x})_1 \\ \vdots \\ (V^\top \vec{x})_m \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $(V^\top \vec{x})_i$ is the ith entry of $V^\top \vec{x}$. This gives the coordinate of $\vec{x}$ in the basis of V with the components of $\vec{x}$ in the null space of A zeroed out.

Finally, we transform this vector back to the standard basis by V and this gives $\vec{\hat{x}} = V(\widetilde{\Sigma}\Sigma V^\top \vec{x})$. Therefore, $\vec{\hat{x}}$ has no component in the null space of A.

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

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

Solution: From the above, we have the solution given by:

$$\begin{aligned}\vec{x} = A^{\dagger}\vec{y} &= V\widetilde{\Sigma}U^{\top}\vec{y} \\ &= \begin{bmatrix} 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}\end{aligned}$$

Therefore, a reasonable solution to the system of equations is:

$$\vec{x} = \begin{bmatrix} 3 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$$

(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.

Solution:

Since $\vec{\hat{x}}$ is the pseudoinverse solution, we know that,

$$\vec{\hat{x}} = V\widetilde{\Sigma}U^{\top}\vec{y}$$

Let us write down what $\vec{\hat{x}}$ is with respect to the orthonormal basis formed by the columns of V. Let there be k non-zero singular values. The following expression comes from expanding the matrix multiplication.

$$\begin{aligned}V^{\top}\vec{\hat{x}} &= V^{\top}V\widetilde{\Sigma}U^{\top}\vec{y} \\ &= V^{\top}V\widetilde{\Sigma}U^{\top}\vec{y} \\ &= \widetilde{\Sigma}U^{\top}\vec{y} \\ &= \begin{bmatrix} \Sigma_1^{-1} \\ 0_{(n-m)\times m} \end{bmatrix} \begin{bmatrix} — & \vec{u_1}^{\top} & — \\ & \vdots & \\ — & \vec{u_m}^{\top} & — \end{bmatrix} \vec{y} \\ &= \begin{bmatrix} \Sigma_1^{-1}U^{\top}\vec{y} \\ 0_{(n-m)\times m} \end{bmatrix}\end{aligned}$$

Note that $V^\top \vec{x}$ is the coordinate of $\vec{x}$ in the V basis and recall that $\{\vec{v}_{m+1}, ..., \vec{v_n}\}$ is a basis of the null space of A. Since the last $n-m$ entries of $V^\top \vec{\hat{x}}$ are all zeros, this means that $\vec{\hat{x}}$ has no component in the direction of $\vec{v}_{m+1}, ..., \vec{v_n}$, i.e. $\vec{\hat{x}}$ has no component in the null space of A.

(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: $(CD)^\top = D^\top C^\top$.)

Solution:

We just substitute in to see what happens:

$$\begin{aligned}
A^\top(AA^\top)^{-1} &= (U\Sigma V^\top)^\top(U\Sigma V^\top(U\Sigma V^\top)^\top)^{-1} && (2)\\
&= V\Sigma^\top U^\top(U\Sigma V^\top V\Sigma^\top U^\top)^{-1} && (3)\\
&= V\Sigma^\top U^\top(U(\Sigma\Sigma^\top)U^\top)^{-1} && (4)\\
&= V\Sigma^\top U^\top U(\Sigma\Sigma^\top)^{-1}U^\top && (5)\\
&= V\Sigma^\top(\Sigma\Sigma^\top)^{-1}U^\top. && (6)
\end{aligned}$$

At this point, we are almost done in reaching $A^\dagger = V\widetilde{\Sigma}U^\top$. We have the leading V and the ending $U^\top$. All that we need to do is multiply out the diagonal matrices in the middle.

$$\Sigma^\top(\Sigma\Sigma^\top)^{-1} = (\begin{bmatrix} \sigma_0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & 0 \\ 0 & 0 & 0 & \sigma_{m-1} & 0 & \dots & 0 \end{bmatrix})^\top (\begin{bmatrix} \sigma_0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & 0 \\ 0 & 0 & 0 & \sigma_{m-1} & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \sigma_0 & 0 & \cdots & 0 \\ 0 & \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{m-1} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix})^{-1} \tag{7}$$

$$= (\begin{bmatrix} \sigma_0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & 0 \\ 0 & 0 & 0 & \sigma_{m-1} & 0 & \dots & 0 \end{bmatrix})^\top (\begin{bmatrix} \sigma_0^2 & 0 & \dots & 0 \\ 0 & \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma_{m-1}^2 \end{bmatrix})^{-1} \tag{8}$$

$$= \begin{bmatrix} \sigma_0 & 0 & \dots & 0 \\ 0 & \sigma_1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sigma_{m-1} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_0^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_1^2} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \frac{1}{\sigma_{m-1}^2} \end{bmatrix} \tag{9}$$

$$= \begin{bmatrix} \frac{1}{\sigma_0} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_1} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sigma_{m-1}} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} = \widetilde{\Sigma}. \tag{10}$$

This concludes the proof.

In case you were wondering, the alternative form $A^\top(AA^\top)^{-1}$ comes from first noticing that the columns of $A^\top$ are all orthogonal to the nullspace of A by definition, and so using them as the basis for the subspace in which we want to find the solution. The $(AA^\top)$ are the columns of where each of these basis elements ends up through A. Inverting this tells us how to get to where we want.

Anyway, it is interesting to step back and see that at this point, between 16A and 16B, you now know two *different* ways to solve problems in which there are fewer linear equations than you have unknowns. You learned OMP in 16A which proceeded in a greedy fashion and basically tried to minimize the number of variables that it set to anything other than zero. And now you have learned the Moore-Penrose Psuedoinverse that finds the solution that minimizes the Euclidean norm.

Both of these, as well as least-squares, are ways to manifest the philosophical principle of Occam's Razor algorithmically for learning. Occam's Razor says "Numquam ponenda est pluralitas sine necessitate" (translation: don't posit more than you need.) But there are two different ways to measure "more" — counting and weighing. Least squares (when we have more equations than variables) is on the path of weighing — the norm of the error is minimized. OMP iterates that to also follow the path of counting, where the number of nonzero variables corresponds to the things that are counted. The Moore-Penrose Psuedoinverse is fully in the path of weighing.

Both of these paths grow into major themes in machine learning generally, and both play a very important role in modern machine learning in particular. This is because in many contemporary approaches to machine learning, we try to learn models that have more parameters than we have data points.

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**:**

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

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,

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

Solution: This proof is for the general case of $n \times n$ matrices. You should give yourself full credit if you did this calculation only on the 2×2 case.

$$\begin{aligned}
\text{Tr}\{A^\top A\} &= \sum_{i=1}^N (A^\top A)_{ii} &(11)\\
&= \sum_{i=1}^N \left(\sum_{j=1}^N (A^\top)_{ij} A_{ji}\right) &(12)\\
&= \sum_{i=1}^N \left(\sum_{j=1}^N A_{ji} A_{ji}\right) &(13)\\
&= \sum_{i=1}^N \sum_{j=1}^N (A_{ji}^2) &(14)\\
&= \|A\|_F^2 &(15)
\end{aligned}$$

In the above solution, step 11 writes out the trace definition, step 12 expands the matrix multiplication on the diagonal indices (i.e. index (i,i) is the inner product of row i and column i), step 13 applies the definition of matrix transpose, and the last two steps collects the result into the definition of Frobenius norm.

(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).)

Solution: The direct path is just to compute using the trace formula:

$$\|UA\|_F = \sqrt{\text{Tr}\{(UA)^\top(UA)\}} = \sqrt{\text{Tr}\{A^\top U^\top UA\}} = \sqrt{\text{Tr}\{A^\top A\}} = \|A\|_F$$

Another path is to note that the Frobenius norm squared of a matrix is the sum of squared Euclidean norms of the columns of the matrix. Matrix multiplication UA proceeds to act on each column of A independently. None of those norms change since U is orthonormal, and so the Frobenius norm also doesn't change.

To show the second equality, we must note that $\left\|A^\top\right\|_F = \|A\|_F$, because we are just summing over the same numbers, just in a different order. Hence:

$$\begin{aligned}\|AV\|_F &= \left\|(AV)^\top\right\|_F \\ &= \left\|V^\top A^\top\right\|_F\end{aligned}$$

But the transpose of an orthonormal matrix is also orthonormal, hence this case reduces to the previous case.

(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.)

Solution:

$$\begin{aligned}\|A\|_F &= \left\|U\Sigma V^\top\right\|_F = \left\|\Sigma V^\top\right\|_F = \|\Sigma\|_F \\ &= \sqrt{\text{Tr}\{\Sigma^\top\Sigma\}} = \sqrt{\sum_{i=1}^N \sigma_i^2}\end{aligned}$$

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.)

Solution: We know that u_i is an eigenvector of Q, so we can start with:

$$\widetilde{\lambda}_i \vec{u}_i = Q\vec{u}_i$$
$$\widetilde{\lambda}_i \vec{u}_i^T \vec{u}_i = \vec{u}_i^T Q\vec{u}_i$$
$$\widetilde{\lambda}_i \|\vec{u}_i\|^2 = \vec{u}_i^T Q\vec{u}_i$$
$$\widetilde{\lambda}_i \|\vec{u}_i\|^2 = \vec{u}_i^T MM^T \vec{u}_i$$
$$\widetilde{\lambda}_i \|\vec{u}_i\|^2 = (M^T\vec{u}_i)^T (M^T\vec{u}_i)$$
$$\widetilde{\lambda}_i \|\vec{u}_i\|^2 = \left\|M^T\vec{u}_i\right\|^2$$

Since we know that any squared quantity is positive, the right side of the equation is positive. This means that $\widetilde{\lambda}_i$ must also be positive.

(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.**

Solution: To show orthonormality, we can compute the inner product $\vec{w}_i^T \vec{w}_j$ for all $i, j \in 1, 2, \ldots \ell$ and demonstrate that $\vec{w}_i^T \vec{w}_j = 1$ if $i = j$ and $\vec{w}_i^T \vec{w}_j = 0$ if $i \neq j$:

$$\vec{w}_i^T \vec{w}_j = \left(\frac{M^T\vec{u}_i}{\sqrt{\widetilde{\lambda}_i}}\right)^T \left(\frac{M^T\vec{u}_j}{\sqrt{\widetilde{\lambda}_j}}\right)$$
$$\vec{w}_i^T \vec{w}_j = \frac{\vec{u}_i^T MM^T \vec{u}_j}{\sqrt{\widetilde{\lambda}_i}\sqrt{\widetilde{\lambda}_j}}$$

If $i = j$ we saw that in the previous part $\widetilde{\lambda}_i \|\vec{u}_i\|^2 = \vec{u}_i^T MM^T \vec{u}_i$

$$\vec{w}_i^T \vec{w}_i = \frac{\widetilde{\lambda}_i \|\vec{u}_i\|^2}{\sqrt{\widetilde{\lambda}_i}\sqrt{\widetilde{\lambda}_i}}$$

Since we know U is orthonormal, $\|\vec{u}_i\| = 1$:

$$\vec{w}_i^T \vec{w}_i = \frac{\widetilde{\lambda}_i}{\widetilde{\lambda}_i} = 1$$

If $i \neq j$, we know that $\vec{u}_j$ is an eigenvector of MM^T, so the original equation becomes:

$$\vec{w}_i^T \vec{w}_j = \frac{\lambda_j \vec{u}_i^T \vec{u}_j}{\sqrt{\widetilde{\lambda}_i}\sqrt{\widetilde{\lambda}_j}}$$

Since U is orthonormal, $\vec{u}_i^T \vec{u}_j = 0$ and so $\vec{w}_i^T \vec{w}_j = 0$.

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.