Homework 12

This homework is due on Friday, November 18, 2022 at 11:59PM. Self-grades and HW Resubmissions are due the following Sunday, November 27, 2022 at 11:59PM.

1. **Min Norm Proofs**

Recall from lecture and the previous homework that we need to find a value of $\vec{x}_\star \in \mathbb{R}^n$ that best approximates

$$A\vec{x}_\star \approx \vec{y} \tag{1}$$

where $\vec{y} \in \mathbb{R}^m$. This is the typical problem of least squares, but sometimes we can have multiple values of $\vec{x}$ that approximate $A\vec{x} \approx \vec{y}$ equally well. To choose a unique solution, we pick the $\vec{x}_\star$ with minimum norm.

If A is rank $r = \text{rank}(A)$ and has SVD $A = U\Sigma V^\top$, we can write $U := \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}$, $V := \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$, and $\Sigma = \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$. From the previous homework, you determined that the optimal solution for $\vec{x}_\star$, given the requirements above, is

$$\vec{x}_\star = V \begin{bmatrix} \Sigma_r^{-1} U_r^\top \vec{y} \\ \vec{0}_{n-r} \end{bmatrix} \tag{2}$$

(a) The first property we will show is that $\vec{x}_\star \in \text{Col}\left(A^\top\right)$. **To do this, first prove that** $\text{Null}(A) \perp \text{Col}\left(A^\top\right)$**.** Use the fact that an SVD of $A^\top$ is $A^\top = V\Sigma U^\top$, and use Theorem 14 from Note 16. **Then, show that** $\dim \text{Null}(A) + \dim \text{Col}\left(A^\top\right) = n$**, and use this fact to argue that if a vector** $\vec{\ell} \perp \text{Null}(A)$ **(i.e., it is orthogonal to every vector in** $\text{Null}(A)$**), then** $\vec{\ell} \in \text{Col}\left(A^\top\right)$**.**

(HINT: When we are asked to show $\text{Null}(A) \perp \text{Col}\left(A^\top\right)$*, you need to argue that every vector in* $\text{Null}(A)$ *is orthogonal to every vector in* $\text{Col}\left(A^\top\right)$*.)*

Solution: From Theorem 14, we have that $\text{Col}\left(A^\top\right) = \text{Col}(V_r)$ and $\text{Null}(A) = \text{Col}(V_{n-r})$. Since the columns of V_r are orthogonal to the columns in V_{n-r}, we have that $\text{Col}(V_{n-r}) \perp \text{Col}(V_r)$ so $\text{Null}(A) \perp \text{Col}\left(A^\top\right)$. Since V is an orthonormal matrix, all the columns are linearly independent. Hence, $\dim \text{Col}(V_r) = r$ and $\dim \text{Col}(V_{n-r}) = n - r$. Thus, $\dim \text{Null}(A) + \dim \text{Col}\left(A^\top\right) = \dim \text{Col}(V_{n-r}) + \dim \text{Col}(V_r) = n - r + r = n$. From this, we know that $\text{Null}(A)$ and $\text{Col}\left(A^\top\right)$ together span $\mathbb{R}^n$, and they span distinct directions in $\mathbb{R}^n$ (i.e., there cannot be any vector in both $\text{Null}(A)$ and $\text{Col}\left(A^\top\right)$ simultaneously except $\vec{0}$). Thus, if we have a vector $\vec{\ell} \perp \text{Null}(A)$ (equivalently, $\vec{\ell} \notin \text{Null}(A)$), then $\vec{\ell}$ is in the remaining portion of $\mathbb{R}^n$ that happens to be spanned by $\text{Col}\left(A^\top\right)$.

(b) **Show that we can rewrite eq. (2) as**

$$\vec{x}_\star = V_r \Sigma_r^{-1} U_r^\top \vec{y} \tag{3}$$

Use this to show that $\vec{x}_\star \perp \text{Null}(A)$ **and hence** $\vec{x}_\star \in \text{Col}\left(A^\top\right)$**.**

(HINT: For the first part, write out $V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$ and perform block matrix multiplication.) (HINT: For the second part, write $\vec{x}_\star = V_r\vec{\alpha}$ where $\vec{\alpha} := \Sigma_r^{-1}U_r^\top\vec{y}$. What does this mean about $\vec{x}_\star$'s relationship with the columns of V_{n-r}?)

Solution: Following the hints, we can write

$$\vec{x}_\star = V\begin{bmatrix} \Sigma_r^{-1}U_r^\top\vec{y} \\ \vec{0}_{n-r} \end{bmatrix} \tag{4}$$

$$= \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}\begin{bmatrix} \Sigma_r^{-1}U_r^\top\vec{y} \\ \vec{0}_{n-r} \end{bmatrix} \tag{5}$$

$$= V_r\Sigma_r^{-1}U_r^\top\vec{y} + V_{n-r}\vec{0}_{n-r} \tag{6}$$

$$= V_r\Sigma_r^{-1}U_r^\top\vec{y} \tag{7}$$

For the second part of the problem, we can write $\vec{x} = V_r\vec{\alpha}$ where $\vec{\alpha} := \Sigma_r^{-1}U_r^\top\vec{y}$ as described in the hint. This means that $\vec{x}$ is orthogonal to the columns of V_{n-r} (since it is a linear combination of the columns of V_r), and hence, $\vec{x} \perp \text{Null}(A)$ so $\vec{x} \in \text{Col}(A^\top)$.

(c) Next, we will prove that, when $r = \text{rank}(A) = m$ (so A has to be a wide matrix), we have the following min norm solution:

$$\vec{x}_\star = A^\top\left(AA^\top\right)^{-1}\vec{y} \tag{8}$$

Using eq. (3), show that the above equation holds true. *(HINT: Use the compact SVD, namely $A = U_r\Sigma_rV_r^\top$.) (HINT: U_r should be a square, orthonormal matrix in this case. This is not necessarily the case for V_r, but remember that $V_r^\top V_r = I$.)*

Solution: Note that $U_r = U$ in this case since $r = m$ (so U_r has m columns). Let $A = U_r\Sigma_rV_r^\top$. Plugging this into the right hand side of eq. (8), we get

$$A^\top\left(AA^\top\right)^{-1}\vec{y} = \left(U_r\Sigma_rV_r^\top\right)^\top\left(U_r\Sigma_rV_r^\top\left(U_r\Sigma_rV_r^\top\right)^\top\right)^{-1}\vec{y} \tag{9}$$

$$= V_r\Sigma_rU_r^\top\left(U_r\Sigma_rV_r^\top V_r\Sigma_rU_r^\top\right)^{-1}\vec{y} \tag{10}$$

$$= V_r\Sigma_rU_r^\top\left(U_r\Sigma_r^2U_r^\top\right)^{-1}\vec{y} \tag{11}$$

$$= V_r\Sigma_rU_r^\top\left(U_r^\top\right)^{-1}\left(\Sigma_r^2\right)^{-1}(U_r)^{-1}\vec{y} \tag{12}$$

$$= V_r\Sigma_r\left(\Sigma_r^2\right)^{-1}(U_r)^{-1}\vec{y} \tag{13}$$

$$= V_r\Sigma_r\Sigma_r^{-2}U_r^\top\vec{y} \tag{14}$$

$$= V_r\Sigma_r^{-1}U_r^\top\vec{y} \tag{15}$$

which is exactly the right hand side of eq. (3).

2. Practical SVD System ID

Please answer all of the questions in the Jupyter notebook associated with this homework.

3. PCA Introduction

Let $X \in \mathbb{R}^{m \times n}$ be defined as $X := \begin{bmatrix} \vec{x}_1 & \cdots & \vec{x}_n \end{bmatrix}$ where each $\vec{x}_i \in \mathbb{R}^m$. Let X have an SVD $X = U\Sigma V^\top$. Now, let $U_\ell := \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_\ell \end{bmatrix}$ where $\vec{u}_i$ is the ith column of U. In other words, U_ℓ is the first ℓ columns of U. In this problem, we will go about showing that

$$U_\ell \in \underset{W \in \mathbb{R}^{m \times \ell}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\| \vec{x}_i - WW^\top \vec{x}_i \right\|^2 \tag{16}$$

where $W^\top W = I_\ell$ (i.e., it is a matrix with orthonormal columns). This is an important result for deriving PCA, as you will see in lecture.

(a) **First, show that**

$$\left\| \vec{x}_i - WW^\top \vec{x}_i \right\|^2 = \|\vec{x}_i\|^2 - \left\| W^\top \vec{x}_i \right\|^2 \tag{17}$$

(HINT: Expand the left hand side of the equation above using transposes. That is, use the fact that $\|\vec{v}\|^2 = \vec{v}^\top \vec{v}$.)

Solution: We have that

$$\begin{aligned}
\left\| \vec{x}_i - WW^\top \vec{x}_i \right\|^2 &= \left(\vec{x}_i - WW^\top \vec{x}_i \right)^\top \left(\vec{x}_i - WW^\top \vec{x}_i \right) && (18)\\
&= \left(\vec{x}_i^\top - \vec{x}_i^\top WW^\top \right) \left(\vec{x}_i - WW^\top \vec{x}_i \right) && (19)\\
&= \vec{x}_i^\top \vec{x}_i - \vec{x}_i^\top WW^\top \vec{x}_i - \vec{x}_i^\top WW^\top \vec{x}_i + \vec{x}_i^\top W \underbrace{W^\top W}_{I_\ell} W^\top \vec{x}_i && (20)\\
&= \|\vec{x}_i\|^2 - \left\| W^\top \vec{x}_i \right\|^2 - \left\| W^\top \vec{x}_i \right\|^2 + \left\| W^\top \vec{x}_i \right\|^2 && (21)\\
&= \|\vec{x}_i\|^2 - \left\| W^\top \vec{x}_i \right\|^2 && (22)
\end{aligned}$$

(b) Using the result from the previous part, we can simplify the original optimization problem to say

$$\underset{W \in \mathbb{R}^{m \times \ell}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\| \vec{x}_i - WW^\top \vec{x}_i \right\|^2 = \underset{W \in \mathbb{R}^{m \times \ell}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(\|\vec{x}_i\|^2 - \left\| W^\top \vec{x}_i \right\|^2 \right) \tag{23}$$

$$\underset{W \in \mathbb{R}^{m \times \ell}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(- \left\| W^\top \vec{x}_i \right\|^2 \right) \tag{24}$$

$$\underset{W \in \mathbb{R}^{m \times \ell}}{\operatorname{argmax}} \sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2 \tag{25}$$

where we get the second line from noticing that we cannot change $\vec{x}_i$, so we remove it from the optimization problem. Then, we pull out the negative to turn the minimization problem into a maximization problem. Now, let $W := \begin{bmatrix} \vec{w}_1 & \cdots & \vec{w}_\ell \end{bmatrix}$. **Show that**

$$\sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2 = \sum_{k=1}^{\ell} \vec{w}_k^\top \left(XX^\top \right) \vec{w}_k \tag{26}$$

You may use the fact that $\sum_{i=1}^{n} \vec{x}_i \vec{x}_i^\top = XX^\top$. *(HINT: Start by expanding out the norm squared expression as the sum of squares of the individual entries of $W^\top \vec{x}_i$.)*

Solution: We have that the kth element of $W^\top \vec{x}_i$ is $\vec{w}_k^\top \vec{x}_i$, so

$$\sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2 = \sum_{i=1}^{n} \sum_{k=1}^{\ell} \left(\vec{w}_k^\top \vec{x}_i \right)^2 \tag{27}$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{\ell} \left(\vec{w}_k^\top \vec{x}_i \right) \left(\vec{w}_k^\top \vec{x}_i \right) \tag{28}$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{\ell} \left(\vec{w}_k^\top \vec{x}_i \right) \left(\vec{x}_i^\top \vec{w}_k \right) \tag{29}$$

$$= \sum_{k=1}^{\ell} \vec{w}_k^\top \left(\sum_{i=1}^{n} \vec{x}_i \vec{x}_i^\top \right) \vec{w}_k \tag{30}$$

$$= \sum_{k=1}^{\ell} \vec{w}_k^\top \left(X X^\top \right) \vec{w}_k \tag{31}$$

(c) **Use the result of the previous part to show that**

$$\sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2 = \sum_{k=1}^{\ell} \vec{\tilde{w}}_k^\top \Sigma \Sigma^\top \vec{\tilde{w}}_k \tag{32}$$

where $\vec{\tilde{w}}_k = U^\top \vec{w}_k$. Then, argue that $\Sigma\Sigma^\top$ can be written as

$$\Sigma\Sigma^\top = \begin{bmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & & \\ & & \sigma_r^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \tag{33}$$

where $r = \text{rank}(X)$ *(HINT: Use the SVD of X to simplify the $XX^\top$ term from the previous part.)*

Solution: We have that $XX^\top = (U\Sigma V^\top)(U\Sigma V^\top)^\top = U\Sigma V^\top V \Sigma^\top U^\top = U\Sigma\Sigma^\top U^\top$. Plugging this in to eq. (31), we get

$$\sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2 = \sum_{k=1}^{\ell} \vec{w}_k^\top U \Sigma \Sigma^\top U^\top \vec{w}_k \tag{34}$$

$$= \sum_{k=1}^{\ell} \vec{\tilde{w}}_k^\top \Sigma \Sigma^\top \vec{\tilde{w}}_k \tag{35}$$

Since $\Sigma := \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$, we have that $\Sigma\Sigma^\top = \begin{bmatrix} \Sigma_r^2 & 0_{r\times(m-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(m-r)} \end{bmatrix}$ where $\Sigma_r^2 = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_r^2 \end{bmatrix}$.

(d) From the previous part, we have the following expression:

$$\sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2 = \sum_{k=1}^{\ell} \vec{\tilde{w}}_k^\top \begin{bmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & & \\ & & \sigma_r^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \vec{\tilde{w}}_k \tag{36}$$

One may show (via Cauchy-Schwarz) that

$$\sum_{k=1}^{\ell} \vec{\tilde{w}}_k^\top \begin{bmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & & \\ & & \sigma_r^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \vec{\tilde{w}}_k \leq \sum_{k=1}^{\ell} \sigma_k^2 \tag{37}$$

if $\vec{\tilde{w}}_k$ are required to be orthonormal (you are not required to show this). **Using this fact, find some specific values of $\vec{\tilde{w}}_i$ such that we attain eq. (37) with equality. Then, use this to show that U_ℓ maximizes $\sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2$ and hence is a solution to the original optimization problem.**

Solution: To obtain eq. (37) with equality, we can set $\vec{\tilde{w}}_k = \vec{e}_k$, which is the kth standard basis vector (i.e., a vector with 1 in the kth position and zeros everywhere else). Notice that

$$\begin{bmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & & \\ & & \sigma_r^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \vec{e}_k = \sigma_k^2 \tag{38}$$

so we obtain eq. (37) with equality. Since $\vec{\tilde{w}}_k = \vec{e}_k$ and $\vec{w}_k = U\vec{\tilde{w}}_k$, we have that $\vec{w}_k = \vec{u}_k$, which is the kth column of U. Hence,

$$W = \begin{bmatrix} \vec{w}_1 & \cdots & \vec{w}_\ell \end{bmatrix} = \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_\ell \end{bmatrix} = U_\ell \tag{39}$$

We can set $\vec{\tilde{w}}_1, \ldots, \vec{\tilde{w}}_\ell$ to be any permutation of the first ℓ standard basis vectors, but we choose this specific ordering so we end up with $W = U_\ell$. Since $W = U_\ell$ maximizes $\sum_{i=1}^{n} \left\| W^\top \vec{x}_i \right\|^2$, we have that it minimizes the original optimization problem, so $W = U_\ell$ is a solution.