Discussion 10A

1. **SVD Short Questions** Assume we have the compact form of the SVD of $A = U_1 S V_1^T = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T$.

   (a) Compute $AV_1V_1^T$

   Solutions: Recall that V_1 is an orthogonal matrix, so it has orthonormal columns, giving it the property $V_1^T V_1 = I$. Hence we can write:

   $$AV_1V_1^T = U_1SV_1^TV_1V_1^T = U_1SV_1^T = A$$

   (b) What is the subspace that spans the column space of A?

   Solutions: Given a vector $\vec{x}$, the column space of A is also the same as the space of all possible $A\vec{x}$.

   $$A\vec{x} = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T \vec{x}$$

   But, $\vec{v}_i^T \vec{x}$ is a scalar, hence,

   $$A\vec{x} = \sum_{i=1}^{r} (\sigma_i \vec{v}_i^T \vec{x}) \vec{u}_i$$

   From that decomposition, we can see that $A\vec{x}$ is a linear combination of $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r$. Hence the span of columns of A is the subspace spanned by the columns of U_1.

2. **Frobenius Norm** In this problem we will investigate the properties of the Frobenius norm.

   (a) The trace of a matrix is the sum of its diagonal entries. For example, let $Q \in \mathbb{R}^{N \times N}$, then,

   $$Tr\{Q\} = \sum_{i=1}^{N} Q_{ii}$$

   Much like the norm of a vector $\vec{x} \in \mathbb{R}^N$ is $\sqrt{\sum_{i=1}^{N} x_i^2}$, the Frobenius norm of a matrix Q is defined as,

   $$||Q||_F = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} |Q_{ij}|^2}$$

   Note that matrices have other types of norms as well. With the above definitions, show that,

   $$||A||_F = \sqrt{Tr\{A^T A\}}$$

Solutions:

$$
\begin{aligned}
Tr\{A^TA\} &= \sum_{i=1}^{N}(A^TA)_{ii} \\
&= \sum_{i=1}^{N}\left(\sum_{j=1}^{N}(A^T)_{ij}A_{ji}\right) \\
&= \sum_{i=1}^{N}\left(\sum_{j=1}^{N}A_{ji}A_{ji}\right) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}(A_{ji}^2) \\
&= ||A||_F^2
\end{aligned}
$$

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

$$||UA||_F = ||AV||_F = ||A||_F$$

Solutions:

$$||UA||_F = \sqrt{Tr\{(UA)^T(UA)\}} = \sqrt{Tr\{A^TU^TUA\}} = \sqrt{Tr\{A^TA\}} = ||A||_F$$

To show the second set of equality, we must note that $Tr\{A^TA\} = Tr\{AA^T\}$. Hence,

$$||AV||_F = \sqrt{Tr\{(AV)(AV)^T\}} = \sqrt{Tr\{AVV^TA^T\}} = \sqrt{Tr\{AA^T\}} = ||A||_F$$

(c) Show that $||A||_F = \sqrt{\sum_{i=1}^{N}\sigma_i^2}$

Solutions:

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

3)

Let $A \in \mathbb{R}^{M \times N}$ be a “fat” matrix, where M<N. A is full rank, with Rank{A}=M.

a). $A = U_1 S V_1^T$ Is the SVD of A. What are the sizes of U_1, S, V_1 ?

solution:

Since $\text{Rank}\{A\} = M$, $S \in \mathbb{R}^{M \times M}$

$$A = \underbrace{\boxed{\qquad\qquad\qquad\qquad}}_{N}\, M = M\,\underbrace{\boxed{\;U_1\;}}_{M}\, M\,\underbrace{\boxed{\;S\;}}_{M}\, M\,\underbrace{\boxed{\qquad V_1^T \qquad}}_{N}$$

so, $V_1 \in \mathbb{R}^{M \times N}$ $S \in \mathbb{R}^{M \times M}$ $U_1 \in \mathbb{R}^{M \times M}$

b) You are given the following equation, where $\vec{x}$ is unknown:

$$A\vec{x} = \vec{y} \qquad (1)$$

A is the same as above, and can represent some linear system. $\vec{y}$ is known and can represent a desired output of system A. We would like to design an input $\vec{x}$, which satisfies the above equality. Note, that since A is fat, we can not just compute an inverse. In fact, there are infinite number of solutions to Eq. 1.

We define a pseudo-inverse $A^{\dagger} = V_1 S^{-1} U_1^T$.

Show that $\hat{x} = A^{\dagger} \vec{y}$ is a solution to Eq. 1.

Solution:

U_1 is a square orthonormal matrix. Hence, $U_1^T U_1 = U_1 U_1^T = I_{M\times M}$

V_1 is tall, and orthonormal. Hence, $V_1^T V_1 = I_{M\times M}$ $\left(V_1 V_1^T \neq I_{N\times N} \;!!! \right)$

$$A\hat{x} = AA^{+}\vec{y} = U_1 S \underbrace{V_1^T V_1}_{=I_{M\times M}} S^{-1} U_1^T \vec{y} = U_1 \underbrace{S S^{-1}}_{=I_{M\times M}} U_1^T \vec{y} = \underbrace{U_1 U_1^T}_{=I_{M\times M}} \vec{y} = \vec{y}$$

c) Show that $\hat{x} + \tilde{x}$ is also a solution,

$$A(\hat{x} + \tilde{x}) = \vec{y}$$

only if $\tilde{x}$ is spanned by the null-space of V_1

Solution:

$$\vec{y} = A(\hat{x} + \tilde{x}) = \underbrace{A\hat{x}}_{y} + A\tilde{x} = \vec{y} + A\tilde{x} \Rightarrow \text{true only if } A\tilde{x} = 0$$

$$A\tilde{x} = U_1 S V_1^T \tilde{x} = 0$$

sine S has non-zero diagonals, this is true only if $V_1^T \tilde{x} = 0$

d) Show that when $\hat{x} = A^{\dagger}\vec{y}$, is a solution for Eq. 1. $\hat{x}$ has the minimum norm among all solutions that satisfy Eq. 1.

In other words: let $\vec{x} \mid A\vec{x} = y$. If $\vec{x} \neq \hat{x}$, then $\|\vec{x}\| > \|\hat{x}\|$.

Solution:

Let $A = U \Sigma V^T$ be the full SVD. $V = [V_1\ V_2]$

if $\vec{x} \neq \hat{x}$, then $\vec{x} = \hat{x} + \tilde{x}$

The norm does not change when multiplying by an orthonormal matrix. so,

$$\|\vec{x}\|^2 = \|VV^T\vec{x}\|^2 = \|VV^T(\hat{x} + \tilde{x})\|^2 =$$

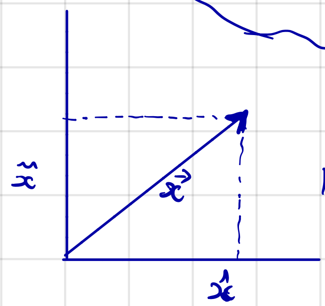
From part (C),

$$= \|V_1V_1^T\hat{x} + V_2V_2^T\tilde{x}\|^2$$

$V_1 // V_2$, so

$$= \|V_1U_1^T\hat{x}\|^2 + \|U_2V_2^T\tilde{x}\|^2 =$$

$$= \|\hat{x}\|^2 + \|\tilde{x}\|^2 > \|\hat{x}\|^2$$



$\|\vec{x}\|^2 = \|\hat{x}\|^2 + \|\tilde{x}\|^2$

e) From Midterm 1, Spring 16

Let $A = \begin{bmatrix} 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}^T$.

Find the vector $\vec{x}$ with the smallest norm, that satisfies,

$$A\vec{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

Solution:

$$\vec{x} = A^{\dagger} y = V_1 S^{-1} U_1^T$$

$$S = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} = S^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{bmatrix}$$

$$S^{-1} U_1^T = \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ -1 & 1 \end{bmatrix}$$

$$V_1 \cdot S^{-1} U_1^T = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \cdot \frac{1}{2} \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ -1 & 1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} \frac{4}{3} & -\frac{2}{3} \\ \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{bmatrix}$$

$$A^{\dagger} = \frac{1}{6} \begin{bmatrix} 2 & -1 \\ 2 & -1 \\ -1 & 2 \\ -1 & 2 \end{bmatrix} \quad \text{and} \quad \vec{x} = A^{\dagger} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 \\ 2 \\ -1 \\ -1 \end{bmatrix}$$

f) Now, let $A \in \mathbb{R}^{M \times N}$ be a tall full rank matrix, M>N. Given a set of equations,

$$A\vec{x} = \vec{y}$$

there is generally no solution that satisfies all the equations exactly. However, we know that the least squares solution x_{LS} minimizes the norm of the error $\|A\vec{x}_{LS} - \vec{y}\|$

In 16A we learned that the solution has a closed form:

$$\vec{x}_{LS} = (A^T A)^{-1} A^T \vec{y}$$

In that case, we can say that $(A^T A)^{-1} A^T$ is a pseudo-inverse of A.

Show that $(A^T A)^{-1} A^T = A^{+} = V_1 S^{-1} U_1^T$

Solution:

Note that A is tall, so,

$$\boxed{A} = \boxed{U_1}\,\boxed{S}\,\boxed{V_1^T}$$

Now $V_1 \in \mathbb{R}^{N \times N}$ is square and orthonormal. Also, $U_1 \in \mathbb{R}^{M \times N}$ is tall and orthonormal so $U_1^T U_1 = I_{N \times N}$

so,

$$(A^T A) = V_1 S \underbrace{U_1^T U_1}_{I_{N \times N}} S V_1^T = V_1 S^2 V_1^T$$

$$(A^T A)^{-1} = V_1 S^{-2} V_1^T$$

$$(A^T A)^{-1} A^T = V_1 S^{-2} \underbrace{V_1^T V_1}_{I_{N \times N}} S U_1^T = V_1 \underbrace{S^{-2} \cdot S}_{S^{-1}} \cdot U_1^T = \underline{\underline{V_1 S^{-1} U_1^T}}$$

$A^{+} = V_1 S^{-1} U_1^T$ is also called the

"Moore-Penrose Pseudo-Inverse"

The same equation using the SVD of A can be used for both tall and fat matrices.

When A is tall, $A^{+}y$ will be the least squares solution.

When A is fat, $A^{+}y$ will be the minimum norm solution.

Same-Same, but different!