(1) Minimum - Norm Solution.

Consider $A\vec{x} = \vec{b}$. $A \in \mathbb{R}^{m \times n}$. m [ ] n Rank(A)=m

If A is a wide matrix, ie. $m < n$

then $A\vec{x} = \vec{b}$ has an infinite # of solutions.

We want to find the minimum norm solution

We want to find the minimum norm solution, when A is full row rank.

ie. $\underset{st\ A\vec{x} = \vec{b}}{\operatorname{argmin}} \|\vec{x}\|$,

Claim: $\vec{x}^* = A^T (AA^T)^{-1} . \vec{b}$.

Proof: Consider any solution $\vec{x}$.

$\vec{x} = \vec{x}_n + \vec{y}$ by the Fundamental thm. of Lin. Alg.

where $\vec{x}_n \in N(A)$ and $\vec{y} \in R(A^T)$. Let $\vec{y} = A^T\vec{z}$

So the norm of $\|\vec{x}\|_2$ can be written as:

$$\|\vec{x}\|_2^2 = (\vec{x}_n + \vec{y})^T(\vec{x}_n + \vec{y})$$

$$= \|\vec{x}_n\|_2^2 + \|\vec{y}\|_2^2 + 2\underbrace{\langle \vec{x}_n, \vec{y} \rangle}_{=0 \text{ by FTLA}}$$

since $N(A) \perp R(A^T)$

Now $A\vec{x} = A\vec{x}_n + A\vec{y}$

$= A\vec{y}$

So any component of $\vec{x}$ that lies in N(A) does not change the value of $A\vec{x}$, however it does increase the norm. Hence to minimize the norm we must choose $\vec{x}_n = 0$,

ie. the min-norm solution belongs to $R(A^T)$.

So we have $A\vec{x} = A\vec{y} = AA^T\vec{z} = \vec{b}$

$\therefore \vec{z} = (AA^T)^{-1}\vec{b}$. $\Rightarrow$ $\vec{y} = A^T\vec{z} = A^T(AA^T)^{-1}\vec{b}$

$\therefore$ The optimal $\vec{x}^* = 0 + \vec{y} = A^T(AA^T)^{-1}\vec{b}$.

Now, to show that

$A^T(AA^T)^{-1}$ is the pseudoinverse of A, we

write:

$$A^T(AA^T)^{-1} = (U_r\Sigma_r V_r^T)^T(U_r\Sigma_r V_r^T(U_r\Sigma_r V_r^T)^T)^{-1}$$

$$= V_r\Sigma_r U_r^T(U_r\Sigma_r V_r^T V_r\Sigma_r U_r^T)^{-1}$$

(since $\underset{r\times n}{V_r^T}\underset{n\times r}{V_r} = I_r$) $$= V_r\Sigma_r U_r^T(U_r\Sigma_r^2 U_r^T)^{-1} \quad (*)$$

Now to compute $(U_r\Sigma_r^2 U_r^T)^{-1}$ recall that $r=m$, since A is full row rank. Hence:

$$U_r\Sigma_r^2 U_r^T = U_m\Sigma_m^2 U_m^T$$

<u>Now:</u> $$(U_m\Sigma_m^2 U_m^T)(U_m\Sigma_m^{-2}U_m^T) \quad (**)$$

$$= U_m\Sigma_m^2 \underbrace{U_m^T U_m}\cdot\Sigma_m^{-2}\cdot U_m^T$$

$$= U_m\Sigma_m^2\, I_m\, \Sigma_m^{-2}U_m^T$$

$$= U_m\cdot I_m U_m^T$$

$$= U_m U_m^T$$

$$= I_m.$$

Note: This last equality holds only because we can replace m in place of r, ie. m=r. In general $U_rU_r^T \neq I_r$ but $U_r^TU_r = I_r$.

Continuing from $(*)$

$$A^T(AA^T)^{-1} = V_r\Sigma_r U_r^T(U_r\Sigma_r^2 U_r^T)^{-1} \overset{\text{since } r=m}{=} V_m\Sigma_m U_m^T(U_m\Sigma_m^2 U_m^T)^{-1}$$

$$\overset{\text{by } **}{=} V_m\Sigma_m U_m^T U_m\Sigma_m^{-2}U_m^T = V_m\Sigma_m^{-1}U_m^T = A^+$$

for full row-rank.

2 Low-Rank Approximation

Thm: $A \in \mathbb{R}^{m \times n}$. $A = U\Sigma V^T$. $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_k \geq \sigma_{k+1} \cdots \geq \sigma_n > 0$

Let $A_k = \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^T$

Then:

1. $$A_k = \underset{\substack{B \in \mathbb{R}^{m \times n} \\ \text{Rank}(B) = k}}{\text{argmin}} \| A - B \|_2$$ Spectral / ~~2~~ 2-2 norm.

2. $$A_k = \underset{\substack{B \in \mathbb{R}^{m \times n} \\ \text{Rank}(B) = k}}{\text{argmin}} \| A - B \|_F$$ Frobenius norm.

$$A = \sum_{i=1}^{n} \sigma_i \vec{u}_i \vec{v}_i^T = U\Sigma V^T.$$

Proof of 1.

Consider $\| A - A_k \|_2 = \| \sum_{i=k+1}^{n} \sigma_i \vec{u}_i \vec{v}_i^T \|_2 = \sigma_{k+1}$

Since the spectral norm is equal to the max singular value, we have the last equality above. The max-singular value of

$$C = \sum_{i=k+1}^{n} \sigma_i \vec{u}_i \vec{v}_i^T \quad \text{is} \quad \sigma_{k+1}.$$

So to prove our theorem, we must show that for any other $B \neq A_k$ of rank-k, we have $\| A - B \|_2 \geq \sigma_{k+1}$.

To show this, first observe:

$$\| A - B \|_2^2 \geq \| (A - B)\vec{w} \|_2^2 \quad \text{for all } \| \vec{w} \|_2 = 1$$

This is true by the definition of the spectral norm.

So now, if we can choose a specific $\vec{w}$ that is helpful, we are done. ~~Note~~ Note, the RHS above is the norm of a vector, while the LHS is the norm of a matrix.

So how do we choose $\vec{w}$?

We want to remove B, so it would be nice if $\vec{w} \in N(B)$.

What else do we want? We want to compare the value of the norm to σ_{k+1}, the $(k+1)$th singular value.

So consider: $V_{k+1} = \begin{bmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_{k+1} \\ | & | & & | \end{bmatrix}$.

Since $\vec{v}_i$'s are orthogonal, $\text{Rank}(V_{k+1}) = \dim(\text{Range}(V_{k+1})) = k+1$.

Since $\text{Rank}(B) = k$, $\dim(N(B)) = n-k$.

Now, $N(B) \subseteq \mathbb{R}^n$ as is $\text{Range}(V_{k+1}) \subseteq \mathbb{R}^n$, both are subspaces of $\mathbb{R}^n$

$$(n-k) + (k+1) = n+1 > n.$$

Given the ambient space that all vectors live in is $\mathbb{R}^n$, there must be atleast one dimension of overlap between the subspaces

$N(B)$ and $\text{Range}(V_{k+1})$.

Choose $\vec{w}$ such that $\vec{w} \in N(B)$ and $\vec{w} \in \text{Range}(V_{k+1})$.

and $\|\vec{w}\|_2 = 1$.

$$\vec{w} \in \text{Range}(V_{k+1}) \Rightarrow \vec{w} = V \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Big\} \; k+1 \text{ non-zero entries.}$$

$$= \alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \cdots + \alpha_{k+1} \vec{v}_{k+1}.$$

$$= V \vec{\alpha}$$

Further $B\vec{w} = 0$.

Finally: $\|\vec{w}\|_2^2 = 1 \Rightarrow \sum_{i=1}^{k+1} \alpha_i^2 = 1$

Back to:

$\| A - B \|_2^2 \geq \| (A-B)\vec{w} \|_2^2$ (def^n of l_2 norm) $\|\vec{w}\| = 1$.

$= \| A\vec{w} \|_2^2$ $\vec{w} \in N(B)$.

$= \| U \Sigma V^T \cdot V \cdot \vec{\alpha} \|_2^2$ $\vec{w} = V \cdot \vec{\alpha}$

$= \| U \Sigma \vec{\alpha} \|_2^2$

$= \| \Sigma \vec{\alpha} \|_2^2$ U is orthonormal

$= \alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2 + \dots + \alpha_k^2 \sigma_k^2 + \alpha_{k+1}^2 \sigma_{k+1}^2$.

$\geq \alpha_1^2 \sigma_{k+1}^2 + \alpha_2^2 \sigma_{k+1}^2 + \dots + \alpha_{k+1}^2 \sigma_{k+1}^2$

$= \sigma_{k+1}^2 \left(\sum_{i=1}^{k+1} \alpha_i^2 \right)$

$= \sigma_{k+1}^2$. $\|\vec{w}\|_2^2 = \sum_{i=1}^{k+1} \alpha_i^2 = 1$.

$\Rightarrow \| A - B \|_2^2 \geq \sigma_{k+1}^2$ for all B.

$\blacksquare$