EECS 16B

Spring 2022

Lecture 22

4/7/2022

LECTURE 22: Applications of SVD:
- pseudoinverse
- least squares
- min energy solution

Recall: Given $A \in \mathbb{R}^{m \times n}$ with rank $= r$, SVD decomposes it as:

$$A = U \Sigma V^T \qquad \text{(full SVD)}$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices and

$$\Sigma = \begin{bmatrix} \Sigma_r & 0_{r \times (n-r)} \\ 0_{(m-r) \times r} & 0_{(m-r) \times (n-r)} \end{bmatrix}, \quad \Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}.$$

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are called the singular values.

If we partition U and V as $[\underset{m\times r}{U_r} \ \underset{m \times (m-r)}{U_{m-r}}]$, $[\underset{n \times r}{V_r} \ \underset{n \times (n-r)}{V_{n-r}}]$ then

$$U \Sigma V^T = U_r \Sigma_r V_r^T. \qquad \text{(compact form)}$$

U_{m-r}, V_{n-r} get canceled but contain useful information about A:

$$\mathrm{Col}(V_{n-r}) = \mathrm{Null}(A)$$
$$\mathrm{Col}(U_{m-r}) = \mathrm{Null}(A^T).$$

Also,

$$\mathrm{Col}(U_r) = \mathrm{Col}(A)$$
$$\mathrm{Col}(V_r) = \mathrm{Col}(A^T).$$

Suppose $m = n = r$ (i.e. A invertible), so

$$A = U \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} V^T.$$

What is A^{-1} in terms of U, Σ, V? $A^{-1} = V \Sigma^{-1} U^T$

because $A^{-1}A = V \Sigma^{-1} \underbrace{U^T \cdot U}_{=I} \Sigma V^T = V \underbrace{\Sigma^{-1} \Sigma}_{=I} V^T = V V^T = I.$

Thus, SVD makes inversion easy.
In addition, a "pseudo inverse" can be derived from SVD when an inverse doesn't exist.

<u>Def'n:</u> Given $A \in \mathbb{R}^{m \times n}$ with rank r and SVD

$$A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0_{(m-r)\times(n-r)} \end{bmatrix} V^T$$

the (Moore-Penrose) pseudo inverse is

$$A^+ = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0_{(n-r)\times(m-r)} \end{bmatrix} U^T$$

or, equivalently,

$$A^+ = V_r \Sigma_r^{-1} U_r^T \qquad \text{(compact form)}$$

$$= \begin{bmatrix} \vec{v}_1 \cdots \vec{v}_r \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_r} \end{bmatrix} \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_r^T \end{bmatrix}$$

$$= \sum_{i=1}^{r} \frac{1}{\sigma_i} \vec{v}_i \vec{u}_i^T \qquad \text{(outer-product form)}.$$

Example: $A = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ $\quad A^+ = \frac{1}{\sigma_1} \vec{v}_1 \cdot \vec{u}_1^T = \frac{1}{\sqrt{5}} \underbrace{\left(\frac{1}{\sqrt{5}} [1 \ 2] \right)}_{\vec{u}_1^T}$

$$= \underbrace{\sqrt{5}}_{\sigma_1} \underbrace{\left(\frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right)}_{\vec{u}_1} \underbrace{1}_{\vec{v}_1} \qquad = \begin{bmatrix} \frac{1}{5} & \frac{2}{5} \end{bmatrix}$$

<u>Note:</u> (i) $A^+ \in \mathbb{R}^{n \times m}$ when $A \in \mathbb{R}^{m \times n}$

(ii) applicable to any $A \neq 0$, not necessarily square nor full rank

(iii) if $m=n=r$ (A invertible) then

$$A^{+} = V\Sigma^{-1}U^{T} = A^{-1}$$

as shown above.

(iv) $$AA^{+} = \underbrace{U_r\Sigma_r V_r^T \cdot V_r}_{I_{r\times r}}\Sigma_r^{-1}U_r^T$$

$$= U_r\Sigma_r\Sigma_r^{-1}U_r^T = U_rU_r^T$$

(v) $$A^{+}A = V_r\Sigma_r^{-1}\underbrace{U_r^TU_r}_{=I}\Sigma_r V_r^T = V_rV_r^T$$

Recall: If $Q = [\vec{q}_1 \cdots \vec{q}_k]$ has orthonormal columns

$$Q^TQ = \begin{bmatrix} \vec{q}_1^T\vec{q}_1 & \vec{q}_1^T\vec{q}_2 & & \\ & \vec{q}_2^T\vec{q}_2 & & \\ & & \ddots & \end{bmatrix} = I_{k\times k}$$

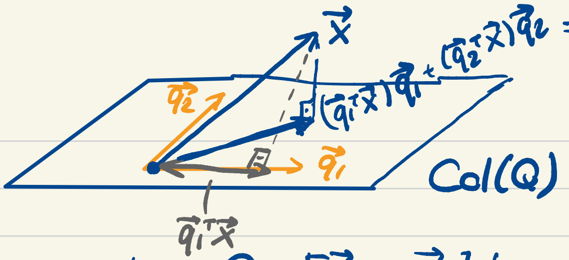
whether Q is square or not, but $QQ^T = I$ only when Q is square.

Example: $$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad Q^TQ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$QQ^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

What is an interpretation of QQ^T when Q is not square?

$$QQ^T\vec{x} = [\vec{q}_1 \cdots \vec{q}_k]\underbrace{\begin{bmatrix} \vec{q}_1^T \\ \vdots \\ \vec{q}_k^T \end{bmatrix}\vec{x}}_{\begin{bmatrix} \vec{q}_1^T\vec{x} \\ \vdots \\ \vec{q}_k^T\vec{x} \end{bmatrix}} = (\vec{q}_1^T\vec{x})\vec{q}_1 + \cdots + (\vec{q}_k^T\vec{x})\vec{q}_k$$



$=$ projection of $\vec{x}$ onto Col(Q) by orthonormality of $\vec{q}_1, \vec{q}_2 \dots$

Therefore, when $Q = [\vec{q}_1 \dots \vec{q}_k]$ has orthonormal columns

$QQ^T\vec{x}$ projects $\vec{x}$ onto Col(Q).

(iv) $\Rightarrow$ AA^+ is a projection onto $\text{Col}(U_r) = \text{col}(A)$

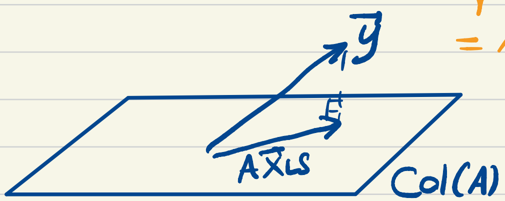
(v) $\Rightarrow$ A^+A " " " " $\text{Col}(V_r) = \text{col}(A^T)$

as shown in the last lecture

Least Squares with SVD:

Want to minimize $\|A\vec{x} - \vec{y}\|$ when $m > n$ (tall matrix).
Recall the minimizer $\vec{x}_{LS}$ is such that

$$A\vec{x}_{LS} = \underbrace{\text{projection of } \vec{y} \text{ onto Col}(A)}_{= AA^+\vec{y} \text{ from above}}.$$



Want $A\vec{x}_{LS} = AA^+\vec{y}$. $\boxed{\vec{x}_{LS} = A^+\vec{y}}$ does the job!

How does this compare to LS solution from before:

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

when A has full column rank $(r = n)$?

Answer: It's the same because $A^+ = (A^TA)^{-1}A^T$ --- (1)
when A has full column rank.

Example above: $A=\begin{bmatrix}1\\2\end{bmatrix}$ $\quad A^TA=5 \quad A^\dagger=\frac{1}{5}A^T=\frac{1}{5}[1\ 2]$

same as what we found above.

Let's show $(A^TA)^{-1}A^T=A^\dagger$ by substituting SVD:

$$A=U_r\Sigma_rV^T \longrightarrow V_r=V \text{ (square)}$$

because $r=n$

$$(A^TA)^{-1}A^T=(\underbrace{\underbrace{V\Sigma_rU_r^T}_{A^T}\cdot\underbrace{U_r\Sigma_rV^T}_{A}}_{I_{r\times r}})^{-1}V\Sigma_rU_r^T$$

$$\begin{bmatrix}\sigma_1^2 & & \\ & \ddots & \\ & & \sigma_r^2\end{bmatrix}$$

$$=\underbrace{(V\Sigma_r^2V^T)^{-1}}_{\underbrace{(V^T)^{-1}}_{V}\Sigma_r^{-2}\underbrace{V^{-1}}_{V^T}}V\Sigma_rU_r^T$$

$$\begin{bmatrix}\frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_r^2}\end{bmatrix}$$

$$=V\Sigma_r^{-2}\underbrace{V^TV}_{=I}\Sigma_rU_r^T=V\Sigma_r^{-1}U_r^T$$

$$=A^\dagger \text{ by def'n.}$$

Therefore (1) is proven.

Likewise, when A is wide and full row rank ($r=m$):

$$A^\dagger=A^T(AA^T)^{-1} \qquad --(2)$$

Can be shown similarly to derivation of (1), by substituting

$$A=U\Sigma_rV_r^T$$

where $U_r=U$ because $r=m$.

Minimal Norm Solution:

Now $m<n$ (wide) so, if $A\vec{x}=\vec{y}$ has a sol'n, it has infinitely many others. Want the one with least $\|\vec{x}\|$.

Substitute compact SVD for A in $A\vec{x}=\vec{y}$:

$$U_r \Sigma_r V_r^T \vec{x} = \vec{y}$$

Multiply from left by U_r^T:

$$\underbrace{\cancel{U_r^T U_r}}_{I} \Sigma_r V_r^T \vec{x} = U_r^T \vec{y}$$

$$\Rightarrow \quad V_r^T \vec{x} = \Sigma_r^{-1} U_r^T \vec{y} \qquad \text{--(3)}$$

Any $\vec{x}$ satisfying (3) solves $A\vec{x}=\vec{y}$. Minimize

$$\|\vec{x}\| \underset{\uparrow}{=} \|V^T\vec{x}\| = \left\| \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix} \vec{x} \right\| = \left\| \begin{bmatrix} V_r^T\vec{x} \\ V_{n-r}^T\vec{x} \end{bmatrix} \right\|$$

since V orthogonal

fixed by (3)

$V_r^T\vec{x}$ is fixed by (3). No constraint on $V_{n-r}^T\vec{x}$. Set it to zero to minimize the norm:

$$V_{n-r}^T \vec{x} = 0. \qquad \text{--(4)}$$

(3) ensures $A\vec{x}=\vec{y}$; (4) minimizes $\|\vec{x}\|$. Together:

$$\begin{bmatrix} V_r^T\vec{x} \\ V_{n-r}^T\vec{x} \end{bmatrix} = \begin{bmatrix} \Sigma_r^{-1} U_r^T \vec{y} \\ 0 \end{bmatrix}$$

i.e., $$V^T\vec{x} = \begin{bmatrix} \Sigma_r^{-1} U_r^T \vec{y} \\ 0 \end{bmatrix}.$$

Multiply from left by V:

$$\underbrace{VV^T}_{=I \text{ because } V \text{ is square and orthogonal}}\vec{x} = V \begin{bmatrix} \Sigma_r^{-1} U_r^T \vec{y} \\ 0 \end{bmatrix}$$

$$\Rightarrow \quad \vec{x} = \underbrace{\begin{bmatrix} V_r & V_{n-r} \end{bmatrix}}_{V} \begin{bmatrix} \Sigma_r^{-1} U_r^T \vec{y} \\ 0 \end{bmatrix}$$

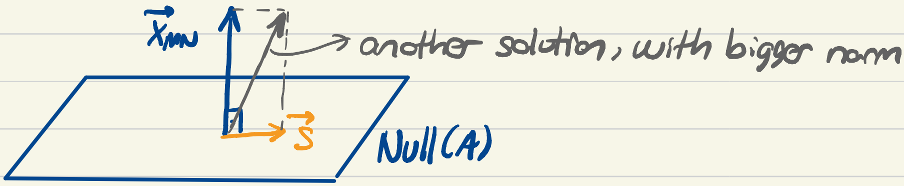
$$\Rightarrow \quad \underbrace{\vec{x}_{MN}}_{\text{"Minimum Norm"}} = V_r \Sigma_r^{-1} U_r^T \vec{y} = A^+ \vec{y} \text{ by def'n of } A^+$$

Thus, $\boxed{\vec{x}_{MN} = A^+ \vec{y}.}$

Note that (4) means: $\begin{bmatrix} \vec{v}_{r+1}^T \\ \vdots \\ \vec{v}_n^T \end{bmatrix} \vec{x} = 0 \Rightarrow \vec{x} \perp \vec{v}_{r+1}, \ldots \vec{v}_n$

Since $\vec{v}_{r+1} \cdots \vec{v}_n$ span Null(A), we can see that the minimum norm solution keeps $\vec{x}_{MN}$ orthogonal to the null space of A. This shows the "wisdom" of the minimum norm solution: any $\vec{x}$ of the form $\vec{x} = \vec{x}_{MN} + \vec{s}$, $\vec{s} \in \text{Null}(A)$ satisfies $A\vec{x} = A\vec{x}_{MN} + \underbrace{A\vec{s}}_{=0} = \vec{y}$

but $\vec{s}$ adds unnecessarily to the norm of $\vec{x}$.



When A is wide and full row rank, $A^+ = A^T(AA^T)^{-1}$

by (2). Thus: $\boxed{\vec{x}_{MN} = A^T(AA^T)^{-1}\vec{y}.}$