EECS 16B : Module 3 / Lecture 3

Announcements:

- 0.25 EC point for each lecture you attend for rest of the term.
- links.eecs16b.org/lecture-ec

Last time:

- SVD: $A = U \Sigma V^T$
  - "Full" SVD construction justification & algorithm
  - "Compact" SVD $A = U_r \Sigma_r V_r^T$

Today:

- Recap of SVD algorithm
- Outer-product SVD : examples
- Geometry of SVD
- Applications of SVD:
  - Pseudo-inverse
  - PCA (Principal Component Analysis) (time-permitting)

SVD procedure for $A = U\Sigma V^T$

① Compute $S = A^TA$

Compute e-vectors of S as $\mathbf{V}$ (orthonormal)

→ Use this to populate the V matrix for the SVD.

→ $\vec{v}_1, \vec{v}_2 \ldots \vec{v}$ correspond to positive e-vals

(rank of C) $\lambda_1 > \lambda_2 > \ldots > \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \ldots = \lambda_n = 0$

② Form $\boxed{\vec{u}_i = \dfrac{A\vec{v}_i}{\sqrt{\lambda_i}}}$ for $\lambda_i \neq 0$

$\sigma_i = \sqrt{\lambda_i}$

$$U = \begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_m \\ | & | & & | \end{bmatrix}$$

③ $$\Sigma = \begin{array}{c} \\ m \end{array}\left[\begin{array}{cccccc|c} \sigma_1 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_r & & & & 0_{m\times(n-m)} \\ & & & 0 & & & \\ & & & & \ddots & & \\ & & & & & 0 & \end{array}\right]$$
$\longleftarrow n \longrightarrow$

, where $\sigma_i = \sqrt{\lambda_i}$, $i = 1, 2 \ldots r$

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

SUMMARY OF SVD

Finding SVD for $A \in \mathbb{R}^{m\times n}$ (with rank $= r$) from evalues/evectors of:

$A^TA \in \mathbb{R}^{n\times n}$

Evalues of A^TA are real and nonnegative. r of them are strictly positive; the remaining $n-r$ are zero.

Step 1: Find orthogonal matrix V diagonalizing A^TA:

$$V^TA^TAV = \begin{bmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} \Big\} n-r$$

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$

Step 2: For each $i = 1, \dots, r$ pick ith column $\vec{v}_i$ of V (which is evector for A^TA for evalue λ_i). Let

$\sigma_i = \sqrt{\lambda_i}$, $\boxed{\vec{u}_i = \frac{1}{\sigma_i} A\vec{v}_i.}$

$AA^T \in \mathbb{R}^{m\times m}$

E-values of (AA^T) are: real, nonnegative evalues, r of which are strictly positive, remaining $m-r$ are zero.

Step 1: Find orthogonal matrix $U \in \mathbb{R}^{m\times m}$ diagonalizing AA^T:

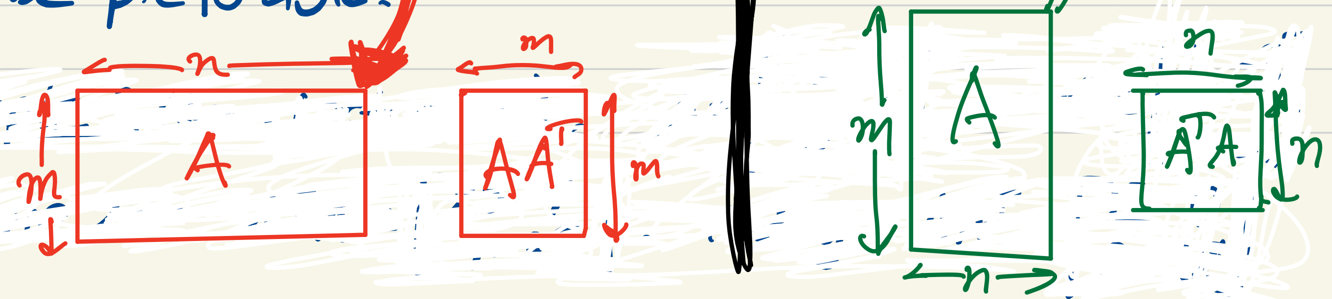
$$U^TAA^TU = \begin{bmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} \Big\} m-r$$

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$

Step 2: For each $i = 1 \dots r$ pick ith column $\vec{u}_i$ of U (which is evector of AA^T for evalue λ_i). Let

$\sigma_i = \sqrt{\lambda_i}$, $\boxed{\vec{v}_i = \frac{1}{\sigma_i} A^T\vec{u}_i.}$

Which procedure to use? Choose A^TA or AA^T based on which one looks simpler for finding evalues/evectors. If $m < n$, AA^T $(m\times m)$ is smaller than A^TA $(n\times n)$ and may be preferable.



Q) Is the SVD of a matrix unique?

A) No!

Ex.

$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ $\quad AA^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\quad U = I$ works for step 1

$\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\quad \vec{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\lambda_1 = \lambda_2 = 1$

$\Downarrow$

$\sigma_1 = \sigma_2 = 1$ $\quad \vec{v}_1 = A^T\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\quad \vec{v}_2 = A^T\vec{u}_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

Since $AA^T = I$, any other orthonormal $\vec{u}_1, \vec{u}_2$ will work.

$\vec{u}_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ $\quad \vec{u}_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$

$\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ above are a special case: $\theta = 0$

$\vec{v}_1 = A^T\vec{u}_1 = \begin{bmatrix} \cos\theta \\ -\sin\theta \end{bmatrix}$ $\quad \vec{v}_2 = A^T\vec{u}_2 = \begin{bmatrix} -\sin\theta \\ -\cos\theta \end{bmatrix}$. Conclusion:

Repeated evalues of A^TA or AA^T ($\lambda_1 = \lambda_2 = 1$ in this example) are another source of nonuniqueness in SVD.

Ex.:

$$A = \begin{bmatrix} 1 & 2 & 4 & 5 \\ 2 & 4 & 8 & 10 \end{bmatrix}_{2\times 4}$$

$$A = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}}_{U_{2\times 2}} \underbrace{\begin{bmatrix} 230 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\Sigma_{2\times 4}} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{46}} & \frac{2}{23} & 2\sqrt{\frac{2}{23}} & \frac{5}{\sqrt{46}} \\ \frac{-5}{\sqrt{26}} & 0 & 0 & \frac{1}{\sqrt{26}} \\ \frac{-2}{\sqrt{273}} & 0 & \sqrt{\frac{13}{21}} & \frac{-10}{\sqrt{273}} \\ \frac{-1}{\sqrt{483}} & \sqrt{\frac{21}{23}} & \frac{-4}{\sqrt{483}} & \frac{-5}{\sqrt{483}} \end{bmatrix}}_{V^T_{4\times 4}}$$

$\boxed{\text{("FULL" SVD)}}$

$$A = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}}_{U_r\,(2\times 1)} \underbrace{\begin{bmatrix} \sqrt{230} \end{bmatrix}}_{\Sigma_r\,(1\times 1)} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{46}} & \sqrt{\frac{2}{23}} & 2\sqrt{\frac{2}{23}} & \frac{5}{\sqrt{46}} \end{bmatrix}}_{V_r^T\,(1\times 4)}$$

$\boxed{\text{"COMPACT" SVD}}$

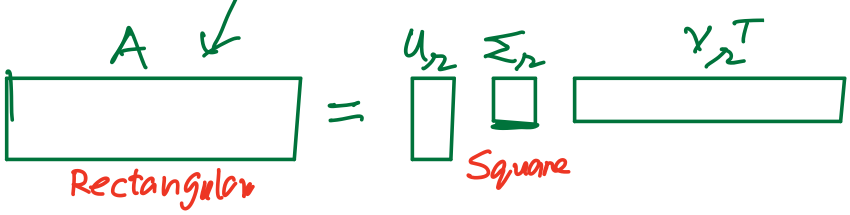
$$\underbrace{A}_{A_{(m\times n)}}\underbrace{\begin{bmatrix}\vec{v}_1 & \cdots & \vec{v}_r\end{bmatrix}}_{V_r\,(n\times r)} = \underbrace{\begin{bmatrix}\vec{u}_1 & \cdots & \vec{u}_r\end{bmatrix}}_{U_r\,(m\times r)}\underbrace{\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{bmatrix}}_{\Sigma_r\,(r\times r)}$$

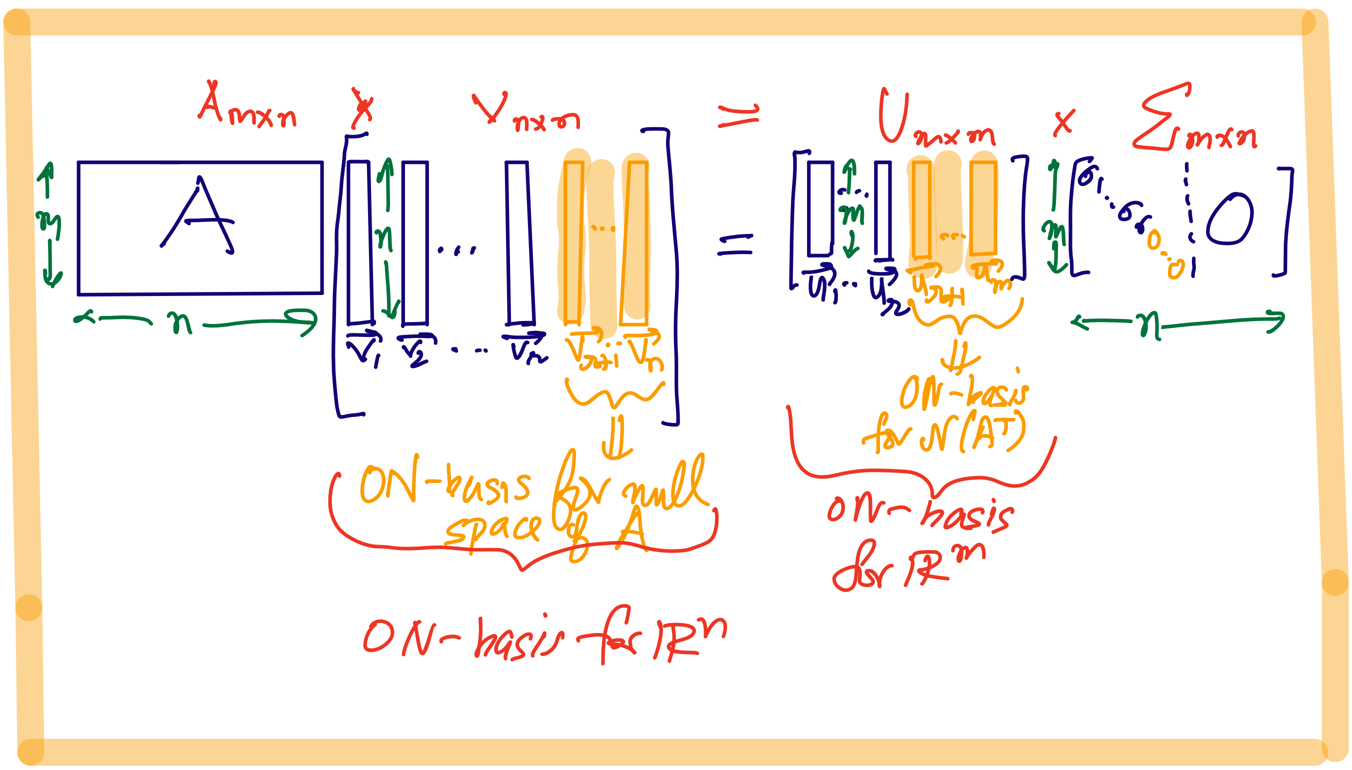
$$A V_r = U_r \Sigma_r$$

General "compact form" SVD:

$$\boxed{\underbrace{A}_{m\times n} = \underbrace{U_r}_{m\times r}\underbrace{\Sigma_r}_{r\times r}\underbrace{V_r^T}_{r\times n}}$$

$$\boxed{\begin{aligned} r &= \operatorname{rank}(A) \\ &\le \min(m,n)\end{aligned}}$$





REVIEW OF MATRIX MULTIPLICATION

Two ways to interpret matrix multiplication :

ex.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad ; \quad B = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

First method:

$$A \cdot B = \begin{bmatrix} \langle \text{Row } 1, \text{Col } 1 \rangle & \langle \text{Row } 1, \text{Col } 2 \rangle \\ \langle \text{Row } 2, \text{Col } 1 \rangle & \langle \text{Row } 2, \text{Col } 2 \rangle \end{bmatrix}$$

$$= \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix}$$

- "Inner Products" used to do matrix multiplication
- In our ex., 4 scalars corr. to 4 inner-products of Rows of A with Columns of B.

Alternate method:

- Take outer products of Columns of A with Rows of B.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad ; \quad B = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

$$AB = \underbrace{\begin{bmatrix} a \\ c \end{bmatrix} \begin{bmatrix} e & f \end{bmatrix}}_{C1(A) \times R1(B)} + \underbrace{\begin{bmatrix} b \\ d \end{bmatrix} \begin{bmatrix} g & h \end{bmatrix}}_{C2(A) \times R2(B)}$$

$$= \underbrace{\begin{bmatrix} ae & af \\ ce & cf \end{bmatrix}}_{\text{Rank-1}} + \underbrace{\begin{bmatrix} bg & bh \\ dg & dh \end{bmatrix}}_{\text{Rank-1}}$$

$$= \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix} \text{ as before!}$$

$$= (\text{Col. 1 of } A)(\text{Row 1 of } B) + (\text{Col. 2 of } A)(\text{Row 2 of } B)$$

- Can express AB as the sum of Rank-1 components
- Each Rank-1 computation is a matrix!

More generally,

$$\boxed{\sum_{i=1}^{r} \vec{u}_i \vec{v}_i^{\,T} = U_r V_r^T}$$

$$\begin{array}{ccccc} \left[\begin{array}{cccc} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_r \end{array}\right] & & \left[\begin{array}{c} \vec{v}_1^{\,T} \\ \vdots \\ \vec{v}_r^{\,T} \end{array}\right] & \begin{array}{c} R1 \\ R2 \\ Rr \end{array} \\ \begin{array}{cccc} C1 & C2 & & Cr \end{array} & & & \end{array}$$

Outer-Product SVD:

$C \in \mathbb{R}^{m\times n}$

$\text{Rank}(C) = r$

$$\underset{(m\times n)}{C} = U \Sigma V^T$$

$$= \left[\begin{array}{c|c} \underbrace{\vec{u}_1 \cdots \vec{u}_r}_{r} & \underbrace{\vec{u}_{r+1} \cdots \vec{u}_m}_{m-r} \end{array}\right] \left[\begin{array}{c|c|c} \underbrace{\begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{matrix}}_{r} & \underbrace{\begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix}}_{m-r} & \underbrace{0}_{n-m} \end{array}\right] \left[\begin{array}{c} \vec{v}_1^{\,T} \\ \vec{v}_2^{\,T} \\ \vdots \\ \vec{v}_n^{\,T} \end{array}\right]$$

$$= \left[\begin{array}{cc} \underbrace{\sigma_1\vec{u}_1 \quad \sigma_2\vec{u}_2 \cdots \sigma_r\vec{u}_r}_{r} & \underbrace{\vec{0} \quad \vec{0} \cdots \vec{0}}_{n-r} \end{array}\right] \left[\begin{array}{c} \vec{v}_1^{\,T} \\ \vec{v}_2^{\,T} \\ \vdots \\ \vec{v}_n^{\,T} \end{array}\right]$$

$$C = [\sigma_1 \vec{u}_1 \quad \sigma_2 \vec{u}_2 \cdots \sigma_r \vec{u}_r \quad \vec{0} \cdots \vec{0}] \begin{bmatrix} \vec{v}_1^T \\ \vec{v}_2^T \\ \vec{v}_r^T \\ \vec{v}_{r+1}^T \\ \vdots \\ \vec{v}_n^T \end{bmatrix}$$

$$= \sum_{i=1}^{r} \sigma_i \underbrace{\vec{u}_i}_{(m\times 1)} \underbrace{\vec{v}_i^T}_{(1\times n)}$$

r = rank of matrix = # of positive singular values.

$\boxed{\text{Outer-product form of SVD is the most efficient \& compact for representation.}}$

$$C = \underbrace{\begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_r \\ | & | & & | \end{bmatrix}}_{U_r} \underbrace{\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}}_{\Sigma_r} \underbrace{\begin{bmatrix} - & \vec{v}_1^T & - \\ - & \vec{v}_2^T & - \\ & \vdots & \\ - & \vec{v}_r^T & - \end{bmatrix}}_{V_r^T}$$

compact-SVD

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

(no unnecessary representation)

outer product SVD

Geometric interpretation of the SVD:

Note:

1) Multiplying a vector $\vec{x}$ by an orthogonal matrix Q does not change its length:

$$\|Q\vec{x}\| = \|\vec{x}\|$$

(Proof: $\|Q\vec{x}\|^2 = \langle Q\vec{x}, Q\vec{x}\rangle = \vec{x}^T \underbrace{Q^T Q}_{I} \vec{x} = \|\vec{x}\|^2$ □)

2) Multiplying a vector by $\Sigma_r = \begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{bmatrix}$ stretches the first entry by σ_1, second entry by σ_2, and so on.

Combining the observations above we can interpret multiplication of a vector $\vec{x}$ by $A = U\Sigma V^T$ as the composition of three operations:

i) $V^T\vec{x}$, which reorients $\vec{x}$ without changing its length;

ii) $\Sigma(V^T\vec{x})$, which stretches the vector $V^T\vec{x}$ along each axis with corresponding singular value;

iii) $U(\Sigma V^T\vec{x})$, which again reorients the resulting vector.

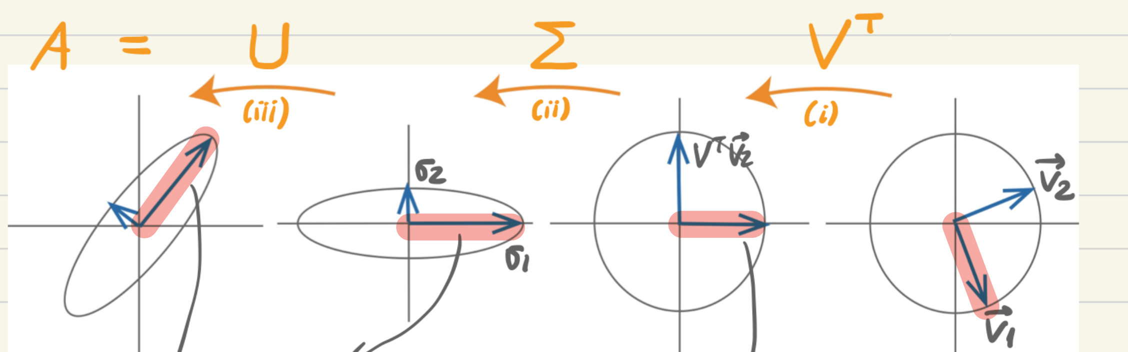


Illustration of multiplication $A\vec{x} = U\Sigma V^T\vec{x}$ when $\vec{x}$ is $\vec{v}_1$.

$$\begin{aligned} & U\Sigma V^T \vec{v}_1 \\ & = A\vec{v}_1 \end{aligned}$$

$$\Sigma V^T\vec{v}_1 = \begin{bmatrix}\sigma_1 & \\ & \sigma_2\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}\sigma_1\\0\end{bmatrix}$$

$$V^T\vec{v}_1 = \begin{bmatrix}\vec{v}_1^T\\ \vec{v}_2^T\end{bmatrix}\vec{v}_1 = \begin{bmatrix}\vec{v}_1^T\vec{v}_1\\ \vec{v}_2^T\vec{v}_1\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix}$$

"Applications" of the SVD:

- pseudo-inverse : Least-Squares & Min-norm solns:
- Principal Component Analysis:

$$\underset{m\times n}{A} = \underset{m\times m}{U}\,\underset{m\times n}{\Sigma}\,\underset{n\times n}{V^T} \quad \text{(full SVD)}$$

Suppose $m=n=r$ (square & Full-Rank)

$\Rightarrow$ A is invertible:

$$A = U \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix} V^T .$$

Q) What is A^{-1} in terms of U, Σ, V?

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

So, SVD makes inversion "easy."

If an inverse does not exist, then a "pseudoinverse" can be defined!

Defn: Given $A \in \mathbb{R}^{m\times n}$ with rank r, and SVD

$$\underset{m\times n}{A} = \underset{m\times m}{U} \cdot \begin{array}{c} \left[\begin{array}{c|c} \Sigma_r & 0_{r\times(n-r)} \\ \hline 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{array}\right] \begin{array}{l} \}\, r \\ \}\, m-r \end{array} \\ \underbrace{\hphantom{\Sigma_r 0_{(m-r)}}}_{r}\ \underbrace{\hphantom{0_{(m-r)\times(n-r)}}}_{(n-r)} \end{array} \underset{n\times n}{V^T},$$

the (Moore-Penrose) pseudoinverse of A is:

$\swarrow$ (full)

$$\boxed{A^{+}_{n\times m} = V_{n\times n} \begin{array}{c} \left[\begin{array}{c|c} \Sigma_r^{-1} & 0_{r\times(m-r)} \\ \hline \underbrace{0_{(n-r)\times r}}_{r} & \underbrace{0_{(n-r)\times(m-r)}}_{(m-r)} \end{array}\right] \begin{array}{l} \}r \\ \}n-r \end{array} \end{array} U^{T}_{m\times m}}$$

or equivalently,

$$\boxed{\underset{(n\times m)}{A^{+}} = \underset{(n\times r)}{V_r}\ \underset{(r\times r)}{\Sigma_r^{-1}}\ \underset{(r\times m)}{U_r^{T}}}$$

(compact form)

$$= \begin{bmatrix} \vec{v}_1 \cdots \vec{v}_r \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_r} \end{bmatrix} \begin{bmatrix} \vec{u}_1^{T} \\ \vdots \\ \vec{u}_r^{T} \end{bmatrix}$$

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

(outer-product form)

Ex. Given $A = [1 \quad 2]$, what is A^{+} (pseudo-inverse)?

$A = [1 \quad 2]$

<u>Soln:</u>

$$\underset{1\times 2}{A} = \underset{1\times 1}{U}\ \underset{1\times 2}{\Sigma}\ \underset{2\times 2}{V^{T}}$$

$$= \sigma_1 \vec{u}_1 \vec{v}_1^{T} = \sqrt{5} \cdot 1 \cdot \frac{1}{\sqrt{5}}[1 \quad 2]$$

$$\underset{2\times 1}{A^{+}} = \frac{1}{\sigma_1} \vec{v}_1 \vec{u}_1^{T} = \frac{1}{\sqrt{5}} \left(\frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right) \cdot 1$$

$$\boxed{A^{+} = \begin{bmatrix} 1/5 \\ 2/5 \end{bmatrix}}$$

__Remark__: If $Q_{n\times k} = [\vec{q}_1 \cdots \vec{q}_k]$ has orthonormal columns, then $Q^TQ = \begin{bmatrix}\vec{q}_1^{\,T}\\ \vdots \\ \vec{q}_k^{\,T}\end{bmatrix}[\vec{q}_1 \cdots \vec{q}_k]$

$$\underset{(k\times n)}{Q^T}\underset{(n\times k)}{Q} = \begin{bmatrix} \vec{q}_1^{\,T}\vec{q}_1 & \vec{q}_1^{\,T}\vec{q}_2 & \cdots & \vec{q}_1^{\,T}\vec{q}_k \\ & \vec{q}_2^{\,T}\vec{q}_2 & & \\ & & \ddots & \\ & & & \vec{q}_k^{\,T}\vec{q}_k \end{bmatrix} = I_{k\times k},$$

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

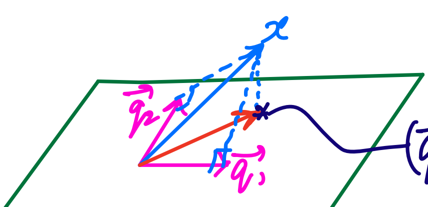
__Ex.__: $Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ $\quad Q^TQ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

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

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

$$QQ^T\vec{x} = [\vec{q}_1 \cdots \vec{q}_k]\begin{bmatrix}\vec{q}_1^{\,T}\\ \vdots \\ \vec{q}_k^{\,T}\end{bmatrix}\vec{x} = [\vec{q}_1 \cdots \vec{q}_k]\begin{bmatrix}\vec{q}_1^{\,T}\vec{x}\\ \vdots \\ \vec{q}_k^{\,T}\vec{x}\end{bmatrix}$$

$$= (\vec{q}_1^{\,T}\vec{x})\vec{q}_1 + \cdots + (\vec{q}_k^{\,T}\vec{x})\vec{q}_k$$



$(\vec{q}_1^{\,T}\vec{x})\vec{q}_1 + (\vec{q}_2^{\,T}\vec{x})\vec{q}_2 =$ Projection of $\vec{x}$ onto column space of Q by orthonormality of $\vec{q}_1, \vec{q}_2$ ..

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

- $AA^+ = U_r\Sigma_r \underbrace{V_r^T V_r}_{I_r} \Sigma_r^{-1} U_r^T = U_r U_r^T$

- $A^+A = V_r\Sigma_r^{-1} \underbrace{U_r^T U_r}_{I_r} \Sigma_r V_r^T = V_r V_r^T$

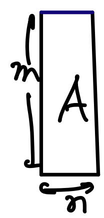
From •, AA^+ is a projection onto $Col(U_r) = Col(A)$
From •, AA^+ " " $Col(V_r) = Col(A^T)$

Pseudoinverse & Least Squares:

$\boxed{m>n}$

Least Squares w/ SVD:

Want to minimize $\|A\vec{x} - \vec{y}\|$ when $m > n \Rightarrow$



- If (A^TA) is invertible, then $\vec{x}_{LS} = (A^TA)^{-1}A^T\vec{y}$ (†)

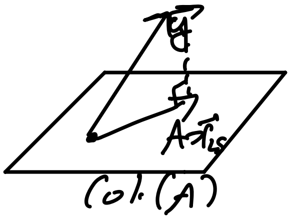
Q) what if (A^TA) is not invertible?

A) $\boxed{\vec{x}_{LS} = A^+\vec{y}}$ ⊗

In fact ⊗ is always valid, whether or not $(A^TA)^{-1}$ exists!

· Recall the minimizer $\vec{x}_{LS}$ is such that:

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



$$A x_{LS} = A A^{+} y \Rightarrow \boxed{\vec{x}_{LS} = A^{+}\vec{y}}$$

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

→ Let's verify $(\dagger)$ when $A^T A$ has full column rank $(r=n)$.

$$A_{m\times n} = U_{m\times n} \Sigma_{n\times n} V^T_{n\times n}$$

$$A = U_r \Sigma_r V_r^T \Rightarrow A = U_r \Sigma_r V^T \quad (V_r = V \text{ since } r = n)$$

$$A^T = V\Sigma_r U_r^T \Rightarrow A^T A = V\Sigma_r \underbrace{U_r^T U_r}_{I_r} \Sigma_r V^T$$

$$= V \Sigma_r^2 V^T$$

$$\Rightarrow (A^T A)^{-1} = V \Sigma_r^{-2} V^T$$

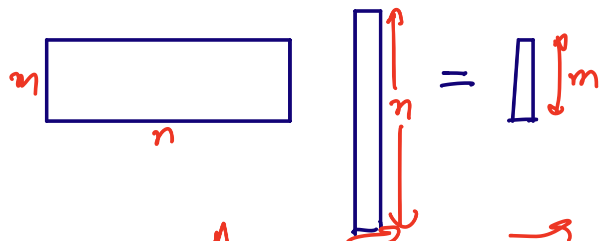
$$\Rightarrow (A^T A)^{-1} A^T = (V\Sigma_r^{-2} V^T)(V\Sigma_r U_r^T) = V\Sigma_r^{-1} U_r^T$$

$$= A^{+}$$

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

One has a similar story for pseudoinverse and minimum-norm (or minimum-energy) setting:

$\boxed{m < n}$



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

min. norm

$\vec{x}_{MN} = A^T (A A^T) \vec{y}$

Exercise: If $r = m$ (full row rank), then verify that $A^+ = A^T (AA^T)^{-1}$

Summary: If $A\vec{x} = \vec{y}$, where we have $m < n$ or $m > n$ (or $m = n$)

$\vec{x} = A^+ \vec{y}$ always works!

(POWER OF SVD!)