EECS 16B : Module 3 / Lecture 2

Announcements:

- 0.25 EC point for each lecture you attend for rest of the term.

links.eecs16b.org/lecture-ec

Last time:

- Minimum-Energy Control
- Recap of spectral theorem for symmetric matrices
- Singular Value Decomposition (SVD)

$$A = U \Sigma V^T$$

See Note 14

TODAY:

- SVD: $A = U \Sigma V^T$
  - "Full" SVD construction justification & algorithm
  - Examples
  - "Compact" SVD & "Outer-product" SVD

RECAP

Let's now recall our study of symmetric matrix S:

1. Symmetric matrix S can always be diagonalized.

2. The diagonalizing basis/matrix V is made up of the eigenvectors of S that are orthonormal.

3. All the eigenvalues of S are real.

$$V^T S V = \Lambda$$

V: orthonormal basis

Λ: diagonal

$$\Lambda = \left[\begin{array}{ccc|c} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_r & \\ \hline & & & 0 \cdots 0 \end{array}\right]$$

$$\Rightarrow S = V \Lambda V^T$$

$$S V = V \Lambda$$

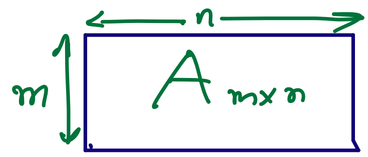
$$S \left[\begin{array}{cccc|ccc} | & | & & | & & & \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_r & \vec{v}_{r+1} & \cdots & \vec{v}_n \\ | & | & & | & & & \end{array}\right] = \left[\begin{array}{ccc|ccc} \vec{v}_1 & \cdots & \vec{v}_r & \vec{v}_{r+1} & \cdots & \vec{v}_n \end{array}\right] \left[\begin{array}{ccc|c} \lambda_1 & & & \\ & \ddots & & 0 \\ & & \lambda_r & \\ \hline & 0 & & 0 \cdots 0 \end{array}\right]$$

$\vec{v}_1, \ldots, \vec{v}_r$: correspond to non-zero e-vals

$\vec{v}_{r+1}, \ldots, \vec{v}_n$: correspond to zero e-vals

$$\left.\begin{array}{l} S \vec{v}_{r+1} = 0 \\ S \vec{v}_{r+2} = 0 \\ \vdots \\ S \vec{v}_n = 0 \end{array}\right\} \Rightarrow \vec{v}_{r+1}, \vec{v}_{r+2}, \ldots, \vec{v}_n$$

form an orthonormal basis for $\underbrace{\text{Null-space}(S)}_{\mathcal{N}(S)}$



Q) What is a "good" decomposition for a general (non-square) matrix A?

- We love an orthonormal basis
- We love diagonalization
- But we cannot rely on special structures like ~~symmetry~~ or even ~~square~~ matrices

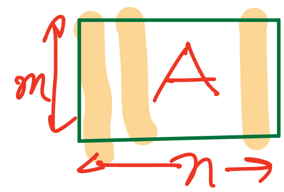
Let us see how to generalize the concept of EIGENVALUE and EIGENVECTOR for square matrices to a similar concept for rectangular matrices, while insisting that we have orthonormal bases.

Key insight: We need TWO orthonormal bases, one for the Column Space(A) & one for the Row-Space(A)

For square matrices

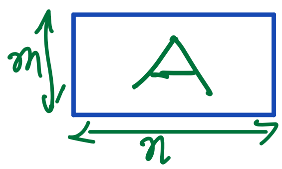
- $\cancel{A \vec{v}_i = \lambda_i \vec{v}_i}$; $(\lambda_i, \vec{v}_i)$ are an e-val/e-vector pair for A (e-vectors need not be ⊥)

How about $\boxed{A \vec{v}_i = \sigma_i \vec{u}_i}$



We now have 2 orthonormal bases:

- $\{\vec{u}_i\}$ for the col. space $(A) \in \mathbb{R}^m$ and
- $\{\vec{v}_i\}$ for the row space $(A) \in \mathbb{R}^n$



- Assume $m < n$
- $\text{Rank}(A) = m$

In matrix form:

$$\underbrace{A}_{A_{m\times n}} \underbrace{\begin{bmatrix} \vec{v}_1 \cdots \vec{v}_m \cdots \vec{v}_n \end{bmatrix}}_{V_{n\times n}} = \underbrace{\begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_m \end{bmatrix}}_{U_{m\times m}} \underbrace{\left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & 0 \\ 0 & & \ddots & & \\ & & & \sigma_m & \end{array}\right]}_{\Sigma_{m\times n}}$$

$$A V = U \Sigma$$

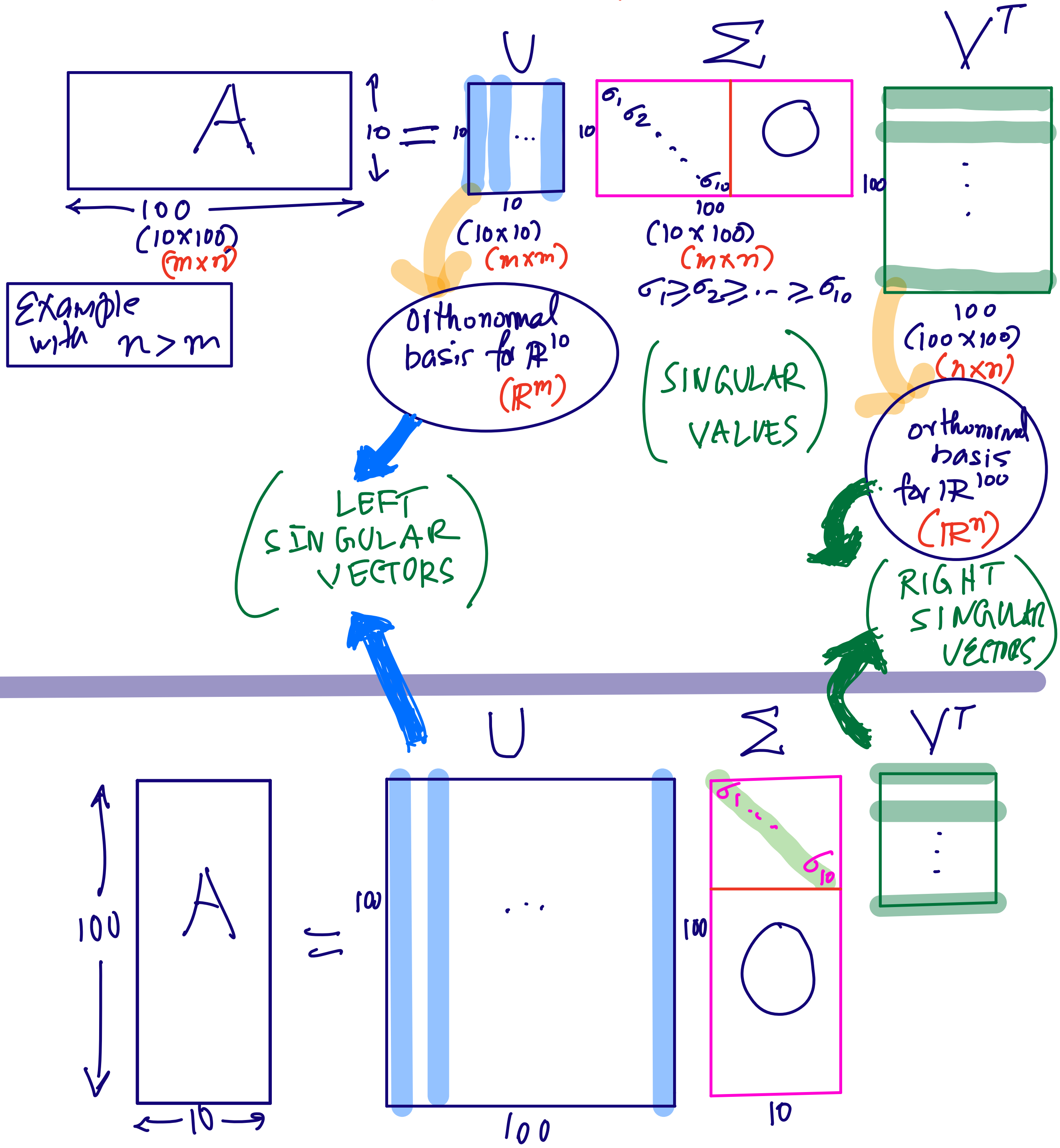
SINGULAR VALUE DECOMPOSITION (SVD)

$$\boxed{A = U \Sigma V^T}$$

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

Singular Value Decomposition (SVD)

$$A = U \Sigma V^T$$



$$AV = U\Sigma$$

$$\Rightarrow \boxed{A = \underbrace{U}_{\text{left singular vectors}}\ \underbrace{\Sigma}_{\text{singular values along diagonal}}\ \underbrace{V^T}_{\text{right singular vectors}}}$$

"Full" SVD of A

Ex. (recap)

$$A = \begin{bmatrix} 4 & 4 \\ 3 & 3 \end{bmatrix} = \underbrace{\frac{1}{5}\begin{bmatrix} 4 & 3 \\ 3 & -4 \end{bmatrix}}_{U}\ \underbrace{\begin{bmatrix} \sqrt{50} & \\ & 0 \end{bmatrix}}_{\Sigma}\ \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{V^T}$$

$\vec{u}_1 = \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix}$, $\vec{u}_2 = \begin{bmatrix} 3/5 \\ -4/5 \end{bmatrix}$

Two orthonormal bases,

$U = [\vec{u}_1\ \vec{u}_2]$ for $Col(A) \in \mathbb{R}^2$

and $V = [\vec{v}_1\ \vec{v}_2]$ for $Row(A) \in \mathbb{R}^2$

NOTE: $A = U\Sigma V^T$ is different from $A = Q\Lambda Q^{-1}$ (eigendecomposition) even for square matrix A!

WHY?

Consider A^TA

If we use SVD, $\underset{m\times n}{A} = \underset{m\times m}{U}\underset{m\times n}{\Sigma}\underset{n\times n}{V^T}$

then $A^TA = (U\Sigma V^T)^T U\Sigma V^T$

$= V\Sigma^T \underbrace{U^T U}_{I} \Sigma V^T$

$= V\Sigma^T\Sigma V^T = \boxed{\underbrace{V}_{n\times n}\underbrace{\Sigma^T\Sigma}_{\text{square } (n\times n)}\underbrace{V^T}_{(n\times n)}}$

Similarly, $AA^T = U\Sigma \underbrace{V^T V}_{I}\Sigma^T U^T$

$= \boxed{\underbrace{U}_{m\times m}\underbrace{\Sigma\Sigma^T}_{m\times m}\underbrace{U^T}_{m\times m}}$

Suggests that the key to understanding the SVD of $A = U\Sigma V^T$ is to study the square matrix A^TA !!!

Consider $C^T C$

Sorry, we will use C and A interchangeably!

Fact: $C^T C$ is symmetric

Proof: $(C^T C)^T = C^T C$ $\square$

Special property: eigenvalues of $S = C^T C$ are always real and non-negative.

Proof: $S\vec{v} = \lambda \vec{v}$ (Let $(\lambda, \vec{v})$ are an e-value/e-vector pair for S)

$C^T C \vec{v} = \lambda \vec{v}$ (We know that all e-values of $S = C^T C$ are real by the spectral theorem)

Left multiply by $\vec{v}^T$ $\Rightarrow$

$$\vec{v}^T C^T C \vec{v} = \vec{v}^T \lambda \vec{v}$$

$$(C\vec{v})^T C\vec{v} = \lambda \vec{v}^T \vec{v}$$

$$\|C\vec{v}\|^2 = \lambda \|\vec{v}\|^2$$

$$\lambda = \frac{\|C\vec{v}\|^2}{\|\vec{v}\|^2} \geq 0$$

$\square$

($C^T C$ is called a Positive Semi-Definite or PSD matrix)

$C^T C$ is a symmetric matrix having non-negative e-values. Order the e-values of $C^T C$ as:

$$\underbrace{\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_r}_{\text{positive eigenvalues}} > \underbrace{\lambda_{r+1} = \cdots = \lambda_n = 0}_{\text{zero e-values}}$$

$$V = \Big[\underbrace{\vec{v}_1 \; \vec{v}_2 \cdots \vec{v}_r}_{\substack{\text{e-vectors of} \\ C^TC \text{ corr. to} \\ \lambda_i > 0}} \; \underbrace{\vec{v}_{r+1} \cdots \vec{v}_n}_{\substack{\text{e-vectors} \\ \text{of } C^TC \text{ corr.} \\ \text{to } \lambda_i = 0}} \Big]$$

<u>Fact</u>: C and $C^T C$ have the same null space.

<u>Pf</u>: (16A) $N(A) = \{\vec{x} : C\vec{x} = 0\}$

- If $C\vec{x} = 0$, then $C^T C x = 0$ (Pf: obvious)
- If $C^T C\vec{x} = 0$, then $C\vec{x} = 0$

Proof of •

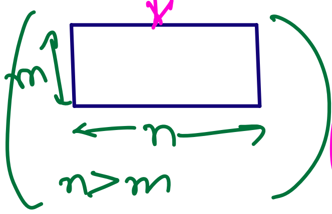
$$C^T C \vec{x} = 0$$

Left-multiply by $\vec{x}^T$:

$$x^T C^T C \vec{x} = 0$$

$$\|C\vec{x}\|^2 = 0 \Rightarrow C\vec{x} = 0 \qquad \square$$

$$C\,V = \left[\, C\vec{v}_1 \ldots C\vec{v}_m \;\middle|\; \underbrace{C\vec{v}_{m+1}, \ldots C\vec{v}_n}_{0} \,\right]$$



(assume full row rank m)

Null space of $C^T C$
$=$ Null space of C
(dim. $= n-m$)

Fact: $\{C\vec{v}_1, C\vec{v}_2, \ldots C\vec{v}_m\}$ form an orthogonal (but not ON) basis set for $\mathbb{R}^m$.

Why is this reassuring? Then, we would have:

$$C_{m\times n} \cdot \underbrace{\left[V_{col} \mid V_{null} \right]}_{n\times n} = \underbrace{\left[U \right]}_{m\times m} \underbrace{\left[\ddots \mid 0 \right]}_{m\times n}$$

and we would be done!

$$CV = \left[\, C\vec{v}_1 \cdots C\vec{v}_m \;\middle|\; C\vec{v}_{m+1} \cdots C\vec{v}_n \,\right]$$

m × n

(assume full rank $r = m < n$)

<u>FACT</u>: $\{C\vec{v}_1, C\vec{v}_2, \dots, C\vec{v}_m\}$ form an orthogonal basis for $\mathbb{R}^m$

why?

$$\{C\vec{v}_1, C\vec{v}_2, \dots, C\vec{v}_m\} = \{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_m\}$$

$$\langle C\vec{v}_i, C\vec{v}_j \rangle = (C\vec{v}_i)^T C\vec{v}_j$$

$$= \vec{v}_i^T \underbrace{\underbrace{C^T C}_{S}\, \vec{v}_j}_{\lambda_j \vec{v}_j}$$

$$= \lambda_j \underbrace{\langle \vec{v}_i, \vec{v}_j \rangle}_{0} = 0 \quad \text{if } i \neq j$$

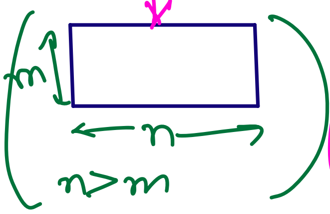
If we want orthonormality, we need to normalize $C\vec{v}_i$ to unit-length.

$$\langle C\vec{v}_i, C\vec{v}_i \rangle = \lambda_i \underbrace{\langle \vec{v}_i, \vec{v}_i \rangle}_{1}$$

$$\|C\vec{v}_i\|^2 = \lambda_i$$

$$\|C\vec{v}_i\| = \sqrt{\lambda_i} \quad \text{for } \lambda_i \neq 0$$

$$CV = \left[\, C\vec{v}_1 \dots C\vec{v}_m \;\middle|\; \underbrace{C\vec{v}_{m+1} \dots C\vec{v}_n}_{0} \,\right]$$



(Assume full row rank m)

Null space of C^TC = Null space of C (dim. $= n-m$)

Fact: $\{C\vec{v}_1, C\vec{v}_2, \dots C\vec{v}_m\}$ form an orthogonal basis for $\mathbb{R}^m$

Why? $\{C\vec{v}_1, C\vec{v}_2, \dots C\vec{v}_m\} = \{\vec{u}_1, \vec{u}_2, \dots \vec{u}_m\}$

$$\langle C\vec{v}_i, C\vec{v}_j \rangle = (C\vec{v}_i)^T C\vec{v}_j$$

$$= \vec{v}_i^T \underbrace{C^T C \vec{v}_j}_{\lambda_j \vec{v}_j} = \lambda_j \underbrace{\langle \vec{v}_i, \vec{v}_j \rangle}_{= 0 \text{ if } i \neq j}$$

If we want an ON-set, we need to normalize $C\vec{v}_i$ to unit length:

$$\langle C\vec{v}_i, C\vec{v}_i \rangle = \lambda_i \underbrace{\langle \vec{v}_i, \vec{v}_i \rangle}_{=1}$$

$$\|C\vec{v}_i\|^2 = \lambda_i \Rightarrow \|C\vec{v}_i\| = \sqrt{\lambda_i}$$

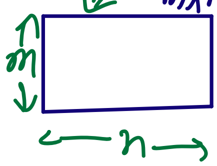
singular value

Fact: $\left\{ \vec{u}_i = \dfrac{C\vec{v}_i}{\sqrt{\lambda_i}} \right\}_{i=1}^{m}$ forms an ON basis for $\mathbb{R}^m$

corresponding to $\vec{v}_i$ for which $\lambda_i > 0$.

$\left(\vec{v}_{r+1}, \dots, \vec{v}_n \text{ corr. to } \lambda_i = 0 \right)$

$$C_{m\times n} V_{n\times n} = \left[\begin{array}{ccc|ccc} C\vec{v}_1 & \cdots & C\vec{v}_m & \multicolumn{3}{c}{\underbrace{C\vec{v}_{r+1} \;\cdots\; C\vec{v}_n}_{0}} \end{array}\right]$$



m, n

$$= \left[\begin{array}{ccc|c} \sqrt{\lambda_1}\,\vec{u}_1 & \cdots & \sqrt{\lambda_m}\,\vec{u}_m & 0 \end{array}\right]$$

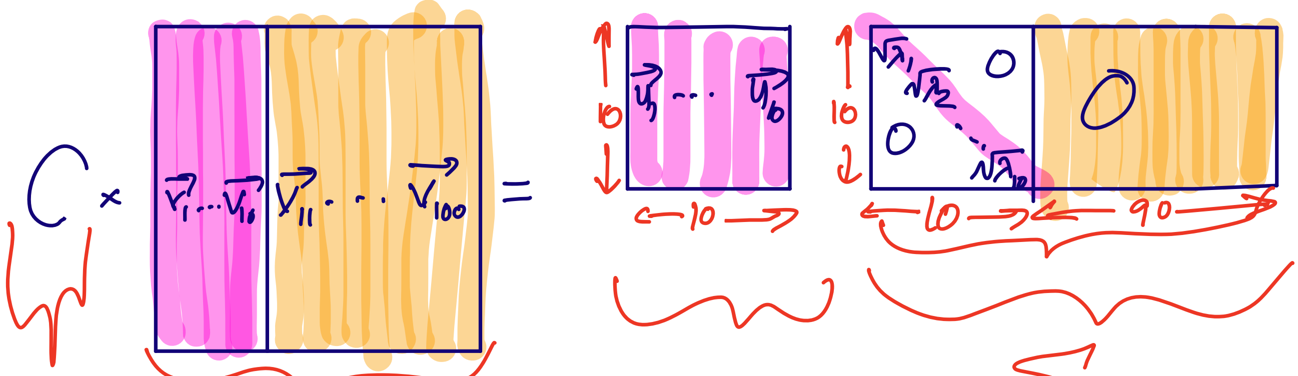
$$= \begin{array}{c} \\ m \\ \\ \end{array} \left[\begin{array}{ccccc} \vec{u}_1 & \vec{u}_2 & \cdots & \cdots & \vec{u}_m \\ & & & & \\ \end{array}\right]_{\leftarrow m \rightarrow} \begin{array}{c} \\ m \\ \\ \end{array} \left[\begin{array}{cccc|c} \sqrt{\lambda_1} & & & & \\ & \sqrt{\lambda_2} & & & 0 \\ & & \ddots & & \\ & & & \sqrt{\lambda_m} & \end{array}\right]_{\leftarrow n \rightarrow}$$

singular values

$$= U_{m\times m} \Sigma_{m\times n}$$

$$\Rightarrow CV = U\Sigma \Rightarrow C = U\Sigma V^T$$

SUMMARY:



$$C_{10\times100} \; V_{100\times100} = U_{10\times10} \; \Sigma_{10\times100}$$

$$C V = U \Sigma$$

or $$C = U \Sigma V^T$$

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

① Compute $S = C^T C$

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

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

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

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m > \lambda_{m+1} = \lambda_{m+2} = \cdots = \lambda_n = 0$

$\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$

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

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

③
$$\Sigma = \begin{array}{c} \\ m \end{array} \left[\begin{array}{cccc|c} \sqrt{\lambda_1} & & & & \\ & \sqrt{\lambda_2} & & & \\ & & \sqrt{\lambda_r} & & O_{m\times(n-m)} \\ & & & 0 \ \ 0 \ \ \sqrt{\lambda_m}\ 0 & \end{array}\right]$$
$$\longleftarrow n \longrightarrow$$

$C = U \Sigma V^T$ (SVD)

Q) What if C is not full rank?
e.g., if there are only r positive e-values of $C^T C$? ($r < m$, $r < n$)

Ex:

$$A = \begin{bmatrix} 4 & 4 \\ 3 & 3 \end{bmatrix}$$

$\boxed{\text{Note: We will use } A \text{ and } C \text{ interchangeably!}}$

1) $$A^T A = \begin{bmatrix} 4 & 3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 4 & 4 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 25 & 25 \\ 25 & 25 \end{bmatrix}$$

Eigenvalues of $A^T A$ are roots of $\det[A^T A - \lambda I] = 0$

$$\Rightarrow \begin{vmatrix} 25-\lambda & 25 \\ 25 & 25-\lambda \end{vmatrix} = 0 \Rightarrow (25-\lambda)^2 - 25^2 = 0$$

$$\Rightarrow (50-\lambda)(0-\lambda) = 0$$

$$\lambda = 0, 50$$

$\lambda_1 = 50$; $\vec{v_1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$\lambda_2 = 0$; $\vec{v_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$$A^T A = \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{V} \underbrace{\begin{bmatrix} 50 & \\ & 0 \end{bmatrix}}_{\Lambda} \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{V^T}$$

2) $$\vec{u_i} = \frac{A \vec{v_i}}{\sqrt{\lambda_i}} \quad \text{for } \lambda_i \neq 0$$

$$\vec{u}_1 = \frac{\begin{bmatrix} 4 & 4 \\ 3 & 3 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}}{\sqrt{50}} = \frac{1}{10}\begin{bmatrix} 8 \\ 6 \end{bmatrix} = \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix}$$

$\vec{u}_2 = ?$ $(\lambda_2 = 0)$

Ans: Use Gram-Schmidt to "complete" ON basis!

$$\vec{u}_2 = \begin{bmatrix} 3/5 \\ -4/5 \end{bmatrix}$$

$$\Rightarrow \quad U = \begin{bmatrix} 4/5 & 3/5 \\ 3/5 & -4/5 \end{bmatrix}$$

③ $$\Sigma = \begin{bmatrix} \sqrt{50} & 0 \\ 0 & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix} \sqrt{50} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

SVD for $A = \begin{bmatrix} 4 & 4 \\ 3 & 3 \end{bmatrix}$ is:

$$\underbrace{\begin{bmatrix} 4 & 4 \\ 3 & 3 \end{bmatrix}}_{A} \underbrace{=}_{=} \underbrace{\begin{bmatrix} 4/5 & 3/5 \\ 3/5 & -4/5 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sqrt{50} & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}}_{V^T}$$

Two orthonormal bases,

$U = [\vec{u}_1 \ \vec{u}_2]$ for $Col(A) \in \mathbb{R}^2$

and $V = [\vec{v}_1 \ \vec{v}_2]$ for $Row(A) \in \mathbb{R}^2$

"Compact form" SVD.

Ex.:

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

Rank(A) = ?

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

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

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

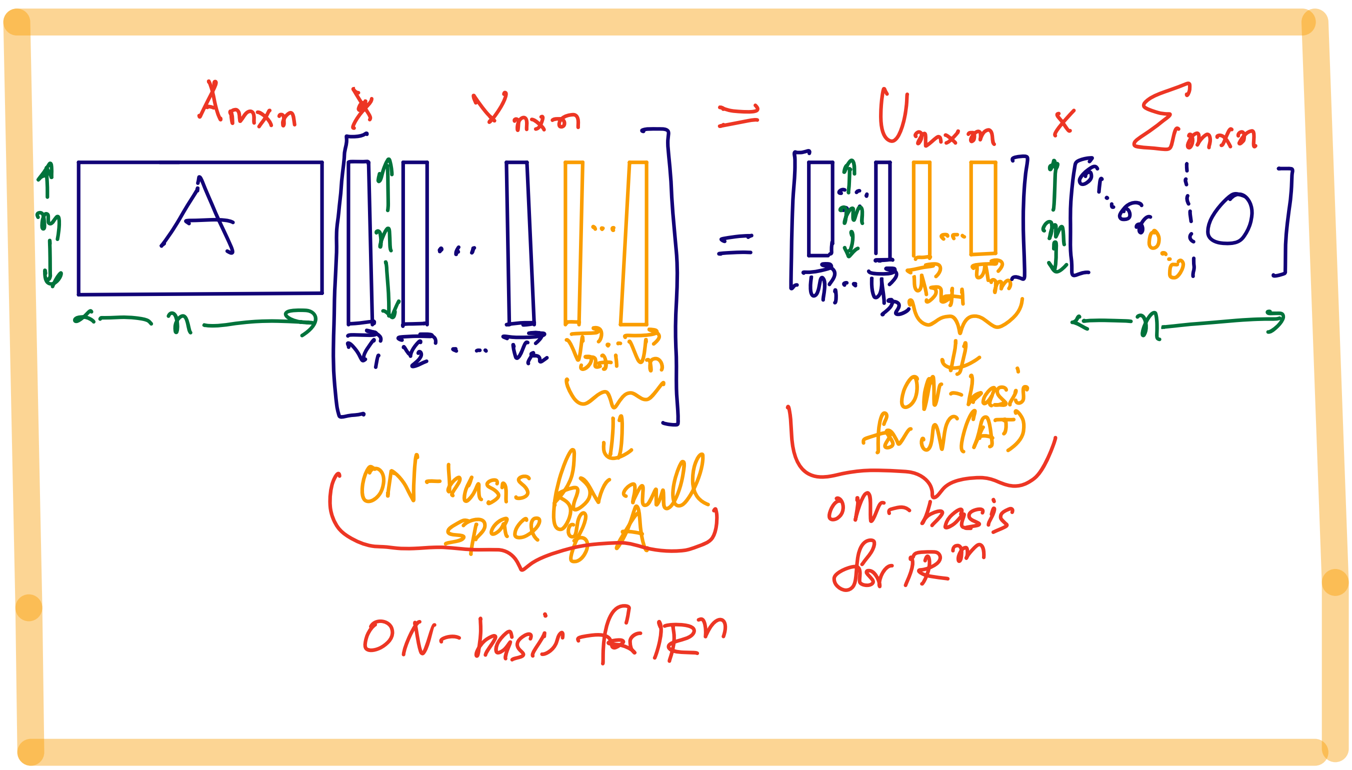
$$A V_r = U_r \Sigma_r$$

"Reduced" or "compact form" SVD:

$$\boxed{A = U_r \Sigma_r V_r^T}$$

$$\boxed{r = \text{rank}(A) \leq \min(m,n)}$$

Let's "expand" out V_r, U_r, Σ_r to "full" SVD:



Outer-Product Form

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

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

$$C = U \Sigma V^T$$

$$= \underbrace{\begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_m \\ | & | & & | \end{bmatrix}}_{m\times m} \underbrace{\left[\begin{array}{ccccc|c} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & O_{m\times(n-m)} \\ & & & 0 & & \\ & & & & \ddots \; 0 & \end{array}\right]}_{m\times n} \underbrace{\begin{bmatrix} - & \vec{v}_1^T & - \\ - & \vec{v}_2^T & - \\ & \vdots & \\ - & \vec{v}_n^T & - \end{bmatrix}}_{n\times n}$$

$$= \overbrace{\underbrace{\begin{bmatrix} | & | & & | \\ \sigma_1\vec{u}_1 & \sigma_2\vec{u}_2 & \cdots & \sigma_r\vec{u}_r \\ | & | & & | \end{bmatrix}}_{r \text{ non-zero columns}}}^{r} \overbrace{\underbrace{\begin{bmatrix} | & | & & | \\ \vec{0} & \vec{0} & \cdots & \vec{0} \\ | & | & & | \end{bmatrix}}_{(n-r) \text{ zero columns}}}^{(n-r)} \begin{bmatrix} - & \vec{v}_1^T & - \\ - & \vec{v}_2^T & - \\ & \vdots & \\ - & \vec{v}_n^T & - \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1\vec{u}_1 & \sigma_2\vec{u}_2 & \cdots & \sigma_r\vec{u}_r & \vec{0} & \cdots & \vec{0} \end{bmatrix} \begin{bmatrix} \vec{v}_1^T \\ \vec{v}_2^T \\ \vec{v}_r^T \\ \vdots \\ \vec{v}_n^T \end{bmatrix}$$

rank of matrix

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

r: rank of matrix

: # of positive singular values

Outer-product form of SVD is the most efficient & compact for representation.

SUMMARY

Finding an 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:

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

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

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

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

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

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:

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

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

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

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

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.

Example

$$A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \qquad A^T = \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}. \qquad AA^T = \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}$$

Example

$A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$ $\quad A^T = \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix}$

$A^TA = \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}$. $\quad AA^T = \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}$

Easier bc/ diagonal: choose $U = I$: $\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\vec{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\lambda_1 = 32$, $\lambda_2 = 18$ $\Rightarrow$ $\sigma_2 = \sqrt{\lambda_2} = 3\sqrt{2}$

$\Downarrow$

$\sigma_1 = \sqrt{\lambda_1} = 4\sqrt{2}$

$\boxed{\vec{v}_2 = \frac{1}{\sigma_2} A^T \vec{u}_2} = \frac{1}{3\sqrt{2}} \begin{bmatrix} -3 \\ 3 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

$\vec{v}_1 = \frac{1}{\sigma_1} A^T \vec{u}_1 = \frac{1}{4\sqrt{2}} \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

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

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

$\lambda_1=\lambda_2=1$

$\Downarrow$

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

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.