



EECS 16B

Designing Information Devices and Systems II

Lecture 23

Prof. Yi Ma

Department of Electrical Engineering and Computer Sciences, UC Berkeley,
yima@eecs.berkeley.edu

Outline

- Singular Value Decomposition (SVD)
 - Geometric Interpretation of SVD
- Applications of SVD: **unifying**
 - Matrix (Pseudo) Inverse
 - Least Squares
 - Minimum Norm Solution

$$y = Ax$$

$$\text{“}A^{-1}\text{”} \quad x = \text{“}A^{-1}\text{”} y$$

Singular Value Decomposition (SVD)

Given $\underline{A \in \mathbb{R}^{m \times n}}$ with $\text{rank}(A) = \underline{r}$, we like to decompose it into a special **matrix** form:

$$V = [\vec{v}_1, \dots, \vec{v}_n] \text{ orthonormal e.v.'s for } \underline{A^\top A} \quad \text{eigenvalues of } A^\top A \text{ (or } AA^\top\text{)} : \underline{\lambda_1 \geq \dots \geq \lambda_r > 0 \dots 0}$$

$$U = [\vec{u}_1, \dots, \vec{u}_m] \text{ orthonormal e.v.'s for } \underline{AA^\top} \quad \Sigma_r = \text{diag}\{\sigma_1 = \underline{\sqrt{\lambda_1}}, \dots, \sigma_r = \underline{\sqrt{\lambda_r}}\} > 0$$

Compact SVD: $\underline{A = U_r \Sigma_r V_r^\top = [\vec{u}_1, \vec{u}_2, \dots, \vec{u}_r]}$

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

$$\begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_r \end{bmatrix} \begin{bmatrix} \vec{v}_1^\top \\ \vec{v}_2^\top \\ \vdots \\ \vec{v}_r^\top \end{bmatrix}$$

Full SVD: $A = U \Sigma V^\top = [\vec{u}_1, \dots, \vec{u}_r, \vec{u}_{r+1}, \dots, \vec{u}_n]$

$$A : \mathbb{R}^n \rightarrow \mathbb{R}^m \quad \begin{matrix} U_r & U_{m-r} \\ \mathbb{R}^m & \end{matrix}$$

$$\begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 & \ddots & \vdots \\ 0 & 0 & \sigma_r & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & 0 & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \vec{v}_1^\top \\ \vdots \\ \vec{v}_{r+1}^\top \\ \vdots \\ \vec{v}_n^\top \end{bmatrix}$$

$V_r^\top \quad \mathbb{R}^m \quad V_{n-r}^\top$

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Geometric Interpretation of SVD

$$\vec{y} = A\vec{x} : A = U\Sigma V^\top = [U_r, U_{m-r}] \begin{bmatrix} \Sigma_r \\ \mathbf{0}_{(m-r) \times r} \end{bmatrix} \begin{bmatrix} \mathbf{0}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times (n-r)} \end{bmatrix} \begin{bmatrix} V_r^\top \\ V_{n-r}^\top \end{bmatrix}$$

The diagram illustrates the decomposition $A = U\Sigma V^\top$ as a sequence of operations:

- U:** A rectangular matrix representing a "rotation" (indicated by a curved arrow pointing left).
- Σ:** A diagonal matrix representing "scaling" (indicated by a curved arrow pointing right).
- V^T:** A rectangular matrix representing another "rotation" (indicated by a curved arrow pointing left).

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

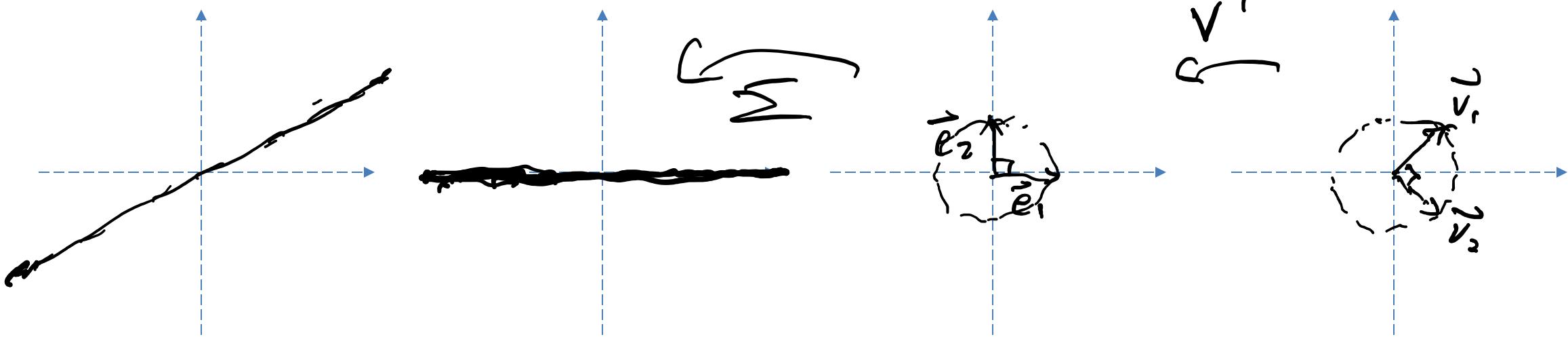
The diagram shows the components of the SVD in a coordinate system:

- U:** A transformation that rotates the original axes \vec{u}_1 and \vec{u}_2 into new axes \vec{e}_1 and \vec{e}_2 .
- Σ:** A scaling transformation that stretches the space along the \vec{e}_1 axis by a factor of σ_1 and the \vec{e}_2 axis by a factor of σ_2 .
- V^T:** A transformation that rotates the scaled axes \vec{v}_1 and \vec{v}_2 back into the original axes \vec{x} and \vec{y} .

Curved arrows indicate the flow of the decomposition process from left to right: "rotation" (red) for U, "scaling" (purple) for Σ, and "rotation" (blue) for V^T.

Geometric Interpretation of SVD (Example)

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



Algebraic Interpretation of SVD

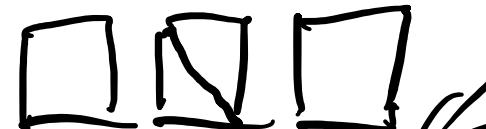
$$A = U\Sigma V^\top = [U_r, U_{m-r}] \begin{bmatrix} \Sigma_r & \mathbf{0}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{bmatrix} \begin{bmatrix} V_r^\top \\ V_{n-r}^\top \end{bmatrix} = \underbrace{U_r \Sigma_r V_r^\top}_{} \quad \downarrow$$

- ① $A V_{n-r} = 0 \Rightarrow \text{Null}(A) = \underline{\text{col}(V_{n-r})} \leftarrow$
- ② $U_{m-r}^\top A = 0 \Rightarrow \text{Null}(A^\top) = \underline{\text{col}(U_{m-r})}$
- ③ $\text{col}(U_r) = \text{col}(A) \perp \text{Null}(A^\top)$
- ④ $\text{col}(V_r) = \text{col}(A^\top) = \text{row}(A)$
 $\perp \text{col}(V_{n-r}) \perp \text{Null}(A)$

Applications of SVD: Matrix Inverse

Given $A \in \mathbb{R}^{m \times n}$ with $\underline{\text{rank}(A) = r = m = n}$: $A = U\Sigma V^\top$. What is its inverse?

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



$$A^{-1} = (U \Sigma V^\top)^{-1} = \underbrace{V}_{\Sigma^{-1}} \underbrace{\Sigma^{-1}}_{U^\top} \underbrace{U^\top}_{U}$$

$$\begin{bmatrix} 6 & 0 \\ 0 & \ddots & 6_r \end{bmatrix} \rightarrow \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \ddots & \frac{1}{6_r} \end{bmatrix}$$

$$AA^{-1} = U \Sigma \underbrace{V^\top}_{\Sigma^{-1}} \Sigma^{-1} U^\top = I$$

Applications of SVD: Matrix Pseudo Inverse

Definition: Given $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = r$ and SVD:

$$A = \underline{U\Sigma V^\top} = U \begin{bmatrix} \Sigma_r & \mathbf{0}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{bmatrix} V^\top \leftarrow \mathbb{C}^{m \times n}$$

its (Moore-Penrose) **pseudo inverse** is defined to be:

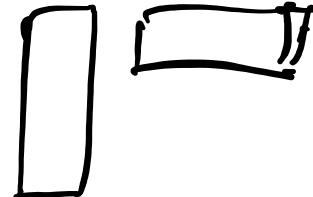
$$A^\dagger = \underline{V \begin{bmatrix} \Sigma_r^{-1} & \mathbf{0}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{bmatrix} U^T} = \underline{V_r \Sigma_r^{-1} U_r^\top} \mathbb{C}^{n \times m}$$

Example: $A = \begin{bmatrix} 4 & 4 \\ 3 & 3 \end{bmatrix}, \quad A^\dagger = ?$

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} 4 & 3 \\ -3 & 4 \end{bmatrix}$$

Applications of SVD: Matrix Pseudo Inverse

① $m = n = r \quad A^T = V \Sigma^{-1} U^T = A^{-1}$



② $r < \min\{m, n\}$

$$AA^T = \underbrace{V_r \Sigma_r V_r^T}_{U_r \Sigma_r^{-1} U_r^T} (U_r \Sigma_r^{-1} U_r^T) = \underbrace{U_r U_r^T}_{\Sigma_r^{-1}} = \sum_{i=1}^r \vec{u}_i \vec{u}_i^T$$

③ $A^T A = (\underbrace{V_r \Sigma_r^{-1} U_r^T}_{U_r \Sigma_r V_r^T}) U_r \Sigma_r V_r^T = \underbrace{U_r U_r^T}_{\Sigma_r} = \sum_{i=1}^r \vec{v}_i \vec{v}_i^T$

projection matr: x

Applications of SVD: Matrix Pseudo Inverse

Geometric interpretation of AA^\dagger or $A^\dagger A$.

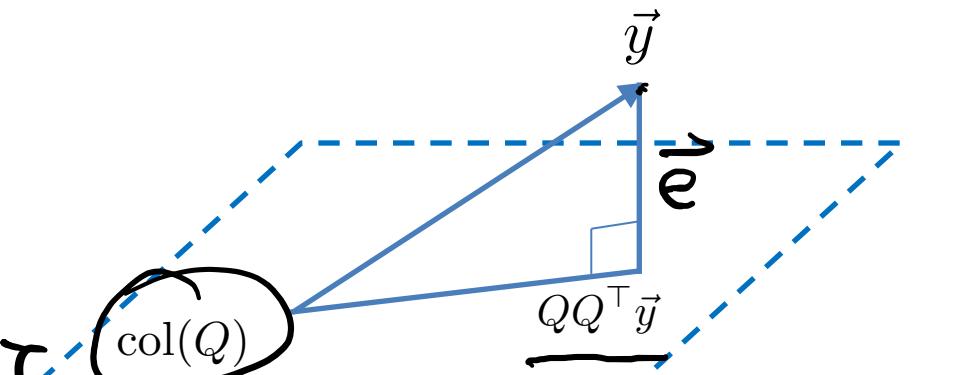
$$\underline{AA^\dagger = U_r V_r^T} \quad \underline{A^\dagger A = V_r V_r^T}$$

Q - orthogonal $Q = [\vec{q}_1, \vec{q}_2, \dots, \vec{q}_k]$

$\epsilon \in \mathbb{R}^{n \times k}$ $k \leq n$ $Q^T Q = I_{k \times k}$ $\underline{QQ^T = \sum_{i=1}^k \vec{q}_i \vec{q}_i^T}$ projection matrix

$$\vec{y} \in \mathbb{R}^n \quad (QQ^T)\vec{y} = \underline{Q(Q^T\vec{y})} \in \text{col}(Q)$$

$$\begin{aligned} \vec{e} &= \underline{\vec{y} - QQ^T\vec{y}} \quad \langle \vec{e}, QQ^T\vec{y} \rangle = \langle \vec{y} - QQ^T\vec{y}, QQ^T\vec{y} \rangle \\ &= \vec{y}^T QQ^T \vec{y} - \vec{y}^T QQ^T \vec{y} = 0. \end{aligned}$$



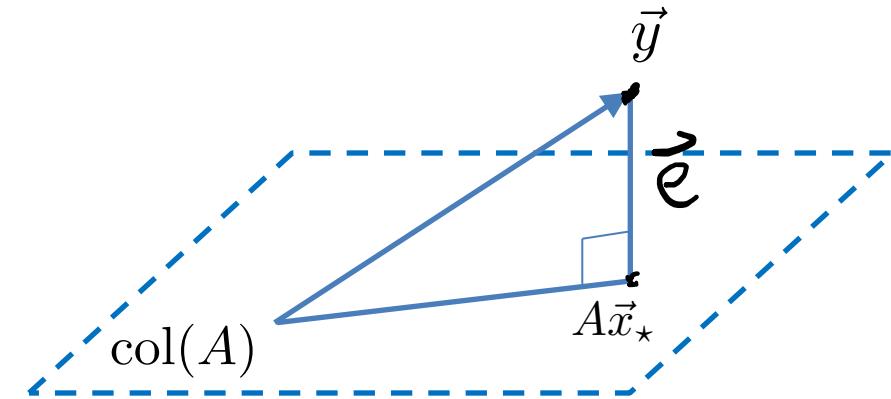
Applications of SVD: Least Squares



$$\underbrace{\min_{\vec{x}} \|\vec{y} - A\vec{x}\|_2^2, \text{ with } A \in \mathbb{R}^{m \times n} \text{ and } \text{rank}(A) = n : \vec{x}_* = \underbrace{(A^\top A)^{-1} A^\top \vec{y}}_{\text{the solution}}$$

an alternative way.

$A\vec{x}_*$ - orthogonal projection onto
 $\text{col}(A) = \text{col}(U_r)$



$AA^\dagger = U_r U_r^\top$ - proj. onto $\text{col}(U_r)$

$AA^\dagger \vec{y}$ - proj. of \vec{y} onto $\text{col}(A)$

$$A\vec{x}_* = AA^\dagger \vec{y} \Rightarrow \vec{x}_* = A^\dagger \vec{y}. \quad \square$$

Applications of SVD: Least Squares

Show: Given $A \in \mathbb{R}^{m \times n}$ and $\text{rank}(A) = n$: $\underbrace{A^\dagger = (A^\top A)^{-1} A^\top}_{?}$

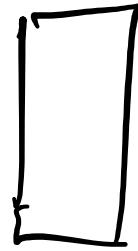
$$\underbrace{(A^\top A)^{-1}}_{?} A^\top \leftarrow A = U_r \Sigma_r V^\top \quad (r=n)$$

$$= (V \Sigma_r \underbrace{U_r^\top}_{U_r \Sigma_r V^\top})^{-1} V \Sigma_r U_r^\top$$

$$= (V \Sigma_r^2 V^\top)^{-1} V \Sigma_r U_r^\top$$

$$= V \underbrace{\Sigma_r^{-2} V^\top}_{V^\top} V \Sigma_r U_r^\top$$

$$= V \Sigma^{-1} U_r^\top = A^\dagger$$



$$\frac{1}{\sigma_i^2} \overset{\circ}{\sigma}_i = \frac{1}{\sigma_i}$$
$$\Sigma_r^{-2} \Sigma_r$$

Applications of SVD: Minimum Norm Solution

$$\min_{\vec{x}} \|\vec{x}\|_2^2 \text{ s.t. } \vec{y} = A\vec{x}, \text{ with } A \in \mathbb{R}^{m \times n} \text{ and } \text{rank}(A) = m : \vec{x}_* = A^\top (AA^\top)^{-1} \vec{y}$$

Show: $\vec{x}_* = A^\dagger \vec{y} (= A^\top (AA^\top)^{-1} \vec{y})$.

$m \boxed{}_n$ $\vec{y} = A(\vec{x} + \vec{s}) \quad \vec{s} \in \text{Nu}(A)$

$$\vec{y} = A\vec{x} = U_r \sum_r V_r^T \vec{x} \Rightarrow U_r^T \vec{y} = \sum_r V_r^T \vec{x}$$

$$\Rightarrow \sum_r^{-1} U_r^T \vec{y} = V_r^T \vec{x}$$

$$\|\vec{V}^T \vec{x}\|_2^2 = \|\vec{x}\|_2^2 = \left\| \begin{bmatrix} V_r^T \vec{x} \\ V_{n-r}^T \vec{x} \end{bmatrix} \right\|_2^2$$

fixed.

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

$$\vec{x}_* = [V_r, V_{n-r}] \begin{bmatrix} \Sigma_r^{-1} U_r^T \vec{y} \\ 0 \end{bmatrix} = \underbrace{V_r \Sigma_r^{-1} U_r^T \vec{y}}_{A^\dagger}$$

$$\vec{x}_* \in \text{col}(V_r) \perp \text{Null}(A)$$

$$\vec{y} = A \underbrace{(\vec{x}_* + \vec{s})}_{\vec{s} \in \text{Null}(A)} = \text{col}(V_{n-r})$$

Applications of SVD: Minimum Norm Solution

$$\vec{y} = A(\vec{x}_* + \vec{s})$$

Diagram illustrating the decomposition:

- \vec{x}_* is the minimum norm solution.
- \vec{s} is the residual vector.
- \vec{x}' is the total vector.
- y is the observed vector.
- V_r^T is the right singular vectors matrix.
- U_r is the left singular vectors matrix.

$$A = U_r \Sigma_r V_r^T$$

Diagram illustrating the SVD decomposition:

- y is the observed vector.
- \vec{z} is the latent variable.
- \vec{s} is the residual vector.
- V_r^T is the right singular vectors matrix.
- Σ_r is the diagonal matrix of singular values.
- U_r is the left singular vectors matrix.
- Proj. indicates the projection of \vec{z} onto the range of A .

~~lifting~~

$$\text{row}(A) = \text{col}(V_r)$$

Diagram illustrating the relationship between the rows of A and the columns of V_r :

- \vec{x}_* is the minimum norm solution.
- \vec{x}' is the total vector.
- \vec{x}_* lies in the column space of V_r .
- \vec{s} is the residual vector.
- $\text{null}(A)$ is the null space of A .