

# Mathematics for Machine Learning

## — Linear Algebra

### Singular Value Decomposition & Matrix Approximation

Joseph Chuang-Chieh Lin

Department of Computer Science & Engineering,  
National Taiwan Ocean University

Fall 2025

# Credits for the resource

- The slides are based on the textbooks:
  - *Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong: Mathematics for Machine Learning. Cambridge University Press. 2020.*
  - *Arnold J. Insel, Lawrence E. Spence, Stephen H. Friedberg: Linear Algebra, 4th Edition. Prentice Hall. 2013.*
  - *Howard Anton, Chris Rorres, Anton Kaul: Elementary Linear Algebra, 12th Edition. Wiley. 2019.*
- We could partially refer to the monograph:  
*Francesco Orabona: A Modern Introduction to Online Learning.*  
<https://arxiv.org/abs/1912.13213>

# Outline

- 1 Singular Value Decomposition (SVD)
  - Construction of the SVD
  - Example
- 2 Matrix Approximation

# Outline

## 1 Singular Value Decomposition (SVD)

- Construction of the SVD
- Example

## 2 Matrix Approximation

# Why Singular Value Decomposition?

- It can be applied to all matrices (not only to square matrices).
- It always exists.

# Illustration

$$\mathbf{A} \in \mathbb{R}^{m \times n}, \text{rank}(\mathbf{A}) = r \leq \min(m, n):$$

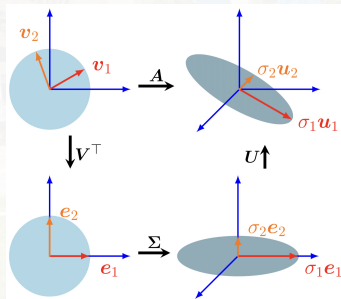
$$\begin{matrix} n \\ \boxed{\mathbf{A}} \\ m \end{matrix} = \begin{matrix} m \\ \boxed{\mathbf{U}} \\ m \end{matrix} \begin{matrix} n \\ \boxed{\Sigma} \\ m \end{matrix} \begin{matrix} n \\ \boxed{\mathbf{V}^T} \\ n \end{matrix}$$

Matrix  $\Sigma$  contains singular values  $\sigma_1, \sigma_2, \dots, \sigma_r$  on the diagonal and a zero block below.

- $\mathbf{U} \in \mathbb{R}^{m \times m}$  with orthogonal columns vectors  $\mathbf{u}_i$ ,  $i = 1, \dots, m$ .
- $\mathbf{V} \in \mathbb{R}^{n \times n}$  with orthogonal columns vectors  $\mathbf{v}_j$ ,  $j = 1, \dots, n$ .
- $\Sigma \in \mathbb{R}^{m \times n}$  with  $\Sigma_{ii} = \sigma_i \geq 0$  and  $\Sigma_{ij} = 0$  for  $i \neq j$ .
  - $\sigma_i$ : **singular values**;  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$ .
  - $\mathbf{u}_i$ : **left-singular vectors**;
  - $\mathbf{v}_j$ : **right-singular vectors**;

# Illustration & Example

$$\begin{aligned}
 \mathbf{A} &= \begin{bmatrix} 1 & -0.8 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\
 &= \begin{bmatrix} -0.79 & 0 & -0.62 \\ 0.38 & -0.78 & -0.49 \\ -0.48 & -0.62 & 0.62 \end{bmatrix} \begin{bmatrix} 1.62 & 0 \\ 0 & 1.0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.78 & 0.62 \\ -0.62 & -0.78 \end{bmatrix}
 \end{aligned}$$



# Exercise

## Exercise

Prove that for an  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{A}\mathbf{A}^\top$  and  $\mathbf{A}^\top\mathbf{A}$  have the same nonzero eigenvalues.



# Outline

- 1 Singular Value Decomposition (SVD)
  - Construction of the SVD
  - Example
- 2 Matrix Approximation

# SVD & Eigendecomposition

- Recall the eigendecomposition of a symmetric positive definite matrix

$$\mathbf{S} = \mathbf{S}^{\top} = \mathbf{P}\mathbf{D}\mathbf{P}^{\top}.$$

# SVD & Eigendecomposition

- Recall the eigendecomposition of a symmetric positive definite matrix

$$\mathbf{S} = \mathbf{S}^{\top} = \mathbf{P}\mathbf{D}\mathbf{P}^{\top}.$$

with the corresponding SVD

$$\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top}$$

so  $\mathbf{U} = \mathbf{P} = \mathbf{V}$ ,  $\mathbf{D} = \mathbf{\Sigma}$ .

# The Overall Idea

- Computing the SVD of  $\mathbf{A} \in \mathbb{R}^{m \times n} \Leftrightarrow$  Finding two sets of orthonormal bases  $U = (\mathbf{u}_1, \dots, \mathbf{u}_m)$  and  $V = (\mathbf{v}_1, \dots, \mathbf{v}_n)$  respectively.
- Images of  $\mathbf{A}\mathbf{v}_i$ 's form a set of orthogonal vectors.

# The first step: Constructing the right-singular vectors

- **Recall:** Eigenvectors of a *symmetric* matrix form an **orthonormal basis** (the Spectral theorem).
- Also, we can always construct a **symmetric, positive semidefinite** matrix  $\mathbf{A}^\top \mathbf{A} \in \mathbb{R}^{n \times n}$  from any matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ .
- Thus,

$$\mathbf{A}^\top \mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^\top = \mathbf{P} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \mathbf{P}^\top,$$

where  $\mathbf{P}$  is orthogonal and composed of orthonormal eigenbasis.

- ★  $\lambda_i \geq 0$  are the eigenvalues of  $\mathbf{A}^\top \mathbf{A}$ .

## The first step (2/2)

- Assume the SVD of  $\mathbf{A}$  exists.

$$\mathbf{A}^\top \mathbf{A} = (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top)^\top (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top) = \mathbf{V} \mathbf{\Sigma}^\top \mathbf{U}^\top \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$$

where  $\mathbf{U}$ ,  $\mathbf{V}$  are orthonormal matrices

## The first step (2/2)

- Assume the SVD of  $\mathbf{A}$  exists.

$$\mathbf{A}^\top \mathbf{A} = (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top)^\top (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top) = \mathbf{V} \mathbf{\Sigma}^\top \mathbf{U}^\top \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$$

where  $\mathbf{U}, \mathbf{V}$  are orthonormal matrices ( $\because \mathbf{U}^\top \mathbf{U} = \mathbf{I}$ ).

## The first step (2/2)

- Assume the SVD of  $\mathbf{A}$  exists.

$$\mathbf{A}^\top \mathbf{A} = (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top)^\top (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top) = \mathbf{V} \mathbf{\Sigma}^\top \mathbf{U}^\top \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$$

where  $\mathbf{U}, \mathbf{V}$  are orthonormal matrices ( $\because \mathbf{U}^\top \mathbf{U} = \mathbf{I}$ ). So,



## The first step (2/2)

- Assume the SVD of  $\mathbf{A}$  exists.

$$\mathbf{A}^\top \mathbf{A} = (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top)^\top (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top) = \mathbf{V} \mathbf{\Sigma}^\top \mathbf{U}^\top \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$$

where  $\mathbf{U}, \mathbf{V}$  are orthonormal matrices ( $\because \mathbf{U}^\top \mathbf{U} = \mathbf{I}$ ). So,

$$\mathbf{A}^\top \mathbf{A} = \mathbf{V} \mathbf{\Sigma}^\top \mathbf{\Sigma} \mathbf{V}^\top = \mathbf{V} \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix} \mathbf{V}^\top$$

- Hence, we identify  $\mathbf{V}^\top = \mathbf{P}^\top$  (right-singular vectors) and  $\sigma_i^2 = \lambda_i$ .

## The second step: Constructing the left-singular vectors

- Similarly, we can always construct a **symmetric, positive semidefinite** matrix  $\mathbf{A}\mathbf{A}^\top \in \mathbb{R}^{m \times m}$  from any matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ .
- Thus, by assuming the SVD of  $\mathbf{A}$  exists, we have

$$\begin{aligned}\mathbf{A}\mathbf{A}^\top &= (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top)(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top)^\top = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top\mathbf{V}\mathbf{\Sigma}^\top\mathbf{U}^\top \\ &= \mathbf{U} \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m^2 \end{bmatrix} \mathbf{U}^\top\end{aligned}$$

**Note:**  $\mathbf{A}\mathbf{A}^\top$  and  $\mathbf{A}^\top\mathbf{A}$  have the same nonzero eigenvalues.

## The second step: Constructing the left-singular vectors

- Similarly, we can always construct a **symmetric, positive semidefinite** matrix  $\mathbf{A}\mathbf{A}^\top \in \mathbb{R}^{m \times m}$  from any matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ .
- Thus, by assuming the SVD of  $\mathbf{A}$  exists, we have

$$\begin{aligned}\mathbf{A}\mathbf{A}^\top &= (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top)(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top)^\top = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top\mathbf{V}\mathbf{\Sigma}^\top\mathbf{U}^\top \\ &= \mathbf{U} \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m^2 \end{bmatrix} \mathbf{U}^\top\end{aligned}$$

**Note:**  $\mathbf{A}\mathbf{A}^\top$  and  $\mathbf{A}^\top\mathbf{A}$  have the same nonzero eigenvalues.

⇒ The nonzero entries of  $\mathbf{\Sigma}$  in the SVD for both steps must be the same.

## The last step: Link up all parts (1/2)

Images of the  $\mathbf{v}_i$  under  $\mathbf{A}$  must be orthogonal.

$$(\mathbf{A}\mathbf{v}_i)^\top (\mathbf{A}\mathbf{v}_j) = \mathbf{v}_i^\top (\mathbf{A}^\top \mathbf{A}) \mathbf{v}_j = \mathbf{v}_i^\top (\lambda_j \mathbf{v}_j) = \lambda_j \mathbf{v}_i^\top \mathbf{v}_j = 0.$$

(For  $m \geq r$ ) We observe that  $\{\mathbf{A}\mathbf{v}_1, \dots, \mathbf{A}\mathbf{v}_r\}$  is a basis of an  $r$ -dimensional subspace of  $\mathbb{R}^m$ .

## The last step: Link up all parts (1/2)

Images of the  $\mathbf{v}_i$  under  $\mathbf{A}$  must be orthogonal.

$$(\mathbf{A}\mathbf{v}_i)^\top (\mathbf{A}\mathbf{v}_j) = \mathbf{v}_i^\top (\mathbf{A}^\top \mathbf{A}) \mathbf{v}_j = \mathbf{v}_i^\top (\lambda_j \mathbf{v}_j) = \lambda_j \mathbf{v}_i^\top \mathbf{v}_j = 0.$$

(For  $m \geq r$ ) We observe that  $\{\mathbf{A}\mathbf{v}_1, \dots, \mathbf{A}\mathbf{v}_r\}$  is a basis of an  $r$ -dimensional subspace of  $\mathbb{R}^m$ .

- Normalize the images of these right-singular vectors:

$$\mathbf{u}_i := \frac{\mathbf{A}\mathbf{v}_i}{\|\mathbf{A}\mathbf{v}_i\|} = \frac{\mathbf{A}\mathbf{v}_i}{\sqrt{(\mathbf{A}\mathbf{v}_i)^\top (\mathbf{A}\mathbf{v}_i)}} = \frac{1}{\sqrt{\lambda_i}} \mathbf{A}\mathbf{v}_i = \frac{1}{\sigma_i} \mathbf{A}\mathbf{v}_i.$$

- That is,  $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$ , for  $i = 1, \dots, r$ .

## The last step: Link up all parts (2/2)

- Concatenate the  $\mathbf{v}_i$ 's as the columns of  $\mathbf{V}$ ;
- Concatenate the  $\mathbf{u}_i$ 's as the columns of  $\mathbf{U}$ ;

## The last step: Link up all parts (2/2)

- Concatenate the  $\mathbf{v}_i$ 's as the columns of  $\mathbf{V}$ ;
- Concatenate the  $\mathbf{u}_i$ 's as the columns of  $\mathbf{U}$ ;
- Then,

$$\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma} \in \mathbb{R}^{m \times n}.$$

## The last step: Link up all parts (2/2)

- Concatenate the  $\mathbf{v}_i$ 's as the columns of  $\mathbf{V}$ ;
- Concatenate the  $\mathbf{u}_i$ 's as the columns of  $\mathbf{U}$ ;
- Then,

$$\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma} \in \mathbb{R}^{m \times n}.$$

Thus,

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



## The last step: Link up all parts (2/2)

- Concatenate the  $\mathbf{v}_i$ 's as the columns of  $\mathbf{V}$ ;
- Concatenate the  $\mathbf{u}_i$ 's as the columns of  $\mathbf{U}$ ;
- Then,

$$\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma} \in \mathbb{R}^{m \times n}.$$

Thus,

$$\mathbf{A} = \mathbf{A}\mathbf{V}\mathbf{V}^{\top} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top}$$

### Exercise

Why do we have  $\mathbf{A} = \mathbf{A}\mathbf{V}\mathbf{V}^{\top}$ ?

# Resemblance to eigenvalues and eigenvectors

- $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$  (from  $\mathbf{A}^\top \mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$  &  $\mathbf{u}_i := \mathbf{A}\mathbf{v}_i / \|\mathbf{A}\mathbf{v}_i\|$ ).

# Resemblance to eigenvalues and eigenvectors

- $\mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i$  (from  $\mathbf{A}^\top\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{u}_i$  &  $\mathbf{u}_i := \mathbf{A}\mathbf{v}_i/\|\mathbf{A}\mathbf{v}_i\|$ ).  
 $\Rightarrow \mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i$ .
- $\mathbf{A}^\top\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}^\top\mathbf{A}\mathbf{v}_i = \frac{1}{\sigma_i}\lambda_i\mathbf{v}_i = \sigma_i\mathbf{v}_i$ .

## Resemblance to eigenvalues and eigenvectors

- $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$  (from  $\mathbf{A}^\top \mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{u}_i$  &  $\mathbf{u}_i := \mathbf{A}\mathbf{v}_i / \|\mathbf{A}\mathbf{v}_i\|$ ).

$$\Rightarrow \mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{A}\mathbf{v}_i.$$

- $\mathbf{A}^\top \mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{A}^\top \mathbf{A}\mathbf{v}_i = \frac{1}{\sigma_i} \lambda_i \mathbf{v}_i = \sigma_i \mathbf{v}_i.$

$$\text{Also, } \mathbf{A}\mathbf{A}^\top \mathbf{u}_i = \sigma_i \mathbf{A}\mathbf{v}_i = \sigma_i^2 \mathbf{u}_i = \lambda_i \mathbf{u}_i.$$

## Resemblance to eigenvalues and eigenvectors

- $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$  (from  $\mathbf{A}^\top \mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$  &  $\mathbf{u}_i := \mathbf{A}\mathbf{v}_i / \|\mathbf{A}\mathbf{v}_i\|$ ).  
 $\Rightarrow \mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{A}\mathbf{v}_i$ .

- $\mathbf{A}^\top \mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{A}^\top \mathbf{A}\mathbf{v}_i = \frac{1}{\sigma_i} \lambda_i \mathbf{v}_i = \sigma_i \mathbf{v}_i$ .

Also,  $\mathbf{A}\mathbf{A}^\top \mathbf{u}_i = \sigma_i \mathbf{A}\mathbf{v}_i = \sigma_i^2 \mathbf{u}_i = \lambda_i \mathbf{u}_i$ .

- Starting from the eigendecomposition of  $\mathbf{A}\mathbf{A}^\top$  we can still get  $\mathbf{u}_i$ .

# Outline

## 1 Singular Value Decomposition (SVD)

- Construction of the SVD
- Example

## 2 Matrix Approximation

## Example

Find the singular value decomposition of

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$$

## Example

Find the singular value decomposition of

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$$

Note:

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$



# SVD Example (step 1/2)

**Goal:** Find  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ .

Perform eigendecomposition of  $\mathbf{A}^\top \mathbf{A} = \mathbf{PDP}^\top$ :

# SVD Example (step 1/2)

**Goal:** Find  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ .

Perform eigendecomposition of  $\mathbf{A}^\top \mathbf{A} = \mathbf{PDP}^\top$ :

$$\mathbf{A}^\top \mathbf{A} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

# SVD Example (step 1/2)

**Goal:** Find  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ .

Perform eigendecomposition of  $\mathbf{A}^\top \mathbf{A} = \mathbf{PDP}^\top$ :

$$\mathbf{A}^\top \mathbf{A} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

So,

$$\mathbf{V} = \mathbf{P} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

where  $\sigma^2 = 6$ ,  $\sigma_2^2 = 1 \Rightarrow \sigma_1 = \sqrt{6}$ ,  $\sigma_2 = 1$ .

## SVD Example (step 2/2)

Left-singular vectors:

## SVD Example (step 2/2)

Left-singular vectors:

$$\mathbf{u}_1 = \frac{1}{\sigma_1} \mathbf{A} \mathbf{v}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 5/\sqrt{30} \\ -2/\sqrt{30} \\ 1/\sqrt{30} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix},$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2} \mathbf{A} \mathbf{v}_2 = \frac{1}{1} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}.$$

Then, we derive  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2] = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$ .

# Outline

- 1 Singular Value Decomposition (SVD)
  - Construction of the SVD
  - Example
- 2 Matrix Approximation

# Motivation

- Represent a matrix  $\mathbf{A}$  as a **sum of simpler low-rank matrices**  $\mathbf{A}_i$ .
- Cheaper than computing the full SVD.
- Rank-1 matrix  $\mathbf{A}_i \in \mathbb{R}^{m \times n}$ :

$$\mathbf{A}_i = \mathbf{u}_i \mathbf{v}_i^\top. \quad (\text{outer product})$$

# Motivation

- Represent a matrix  $\mathbf{A}$  as a **sum of simpler low-rank matrices**  $\mathbf{A}_i$ .
- Cheaper than computing the full SVD.
- Rank-1 matrix  $\mathbf{A}_i \in \mathbb{R}^{m \times n}$ :

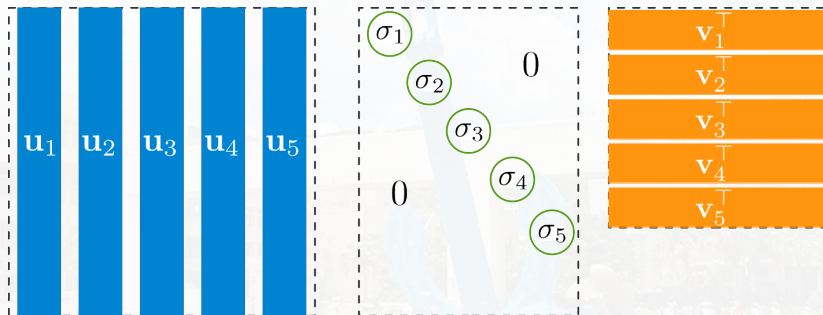
$$\mathbf{A}_i = \mathbf{u}_i \mathbf{v}_i^\top. \quad (\text{outer product})$$

- In fact, we can derive

$$\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^\top = \sum_{i=1}^r \sigma_i \mathbf{A}_i.$$

- Outer-product matrices  $\mathbf{A}_i$  weighted by the  $i$ th singular value  $\sigma_i$ .





- $\sigma_i$  is multiplied with  $\mathbf{u}_i$  and  $\mathbf{v}_i^\top$ .

# Rank- $k$ Approximation

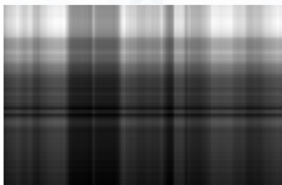
Up to an intermediate value  $k < r$  (assume that  $\sigma_i$ 's are sorted in decreasing order),

$$\hat{\mathbf{A}}(k) := \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^{\top} = \sum_{i=1}^k \sigma_i \mathbf{A}_i$$

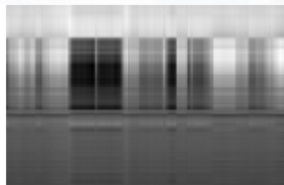
## Illustrating example



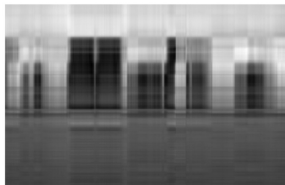
(a) Original image  $A$ .



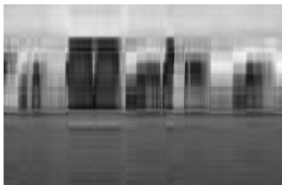
(b) Rank-1 approximation  $\hat{A}(1)$ .



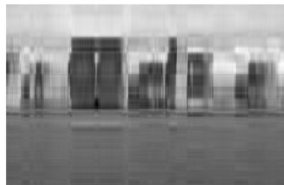
(c) Rank-2 approximation  $\hat{A}(2)$ .



(d) Rank-3 approximation  $\hat{A}(3)$ .



(e) Rank-4 approximation  $\hat{A}(4)$ .



(f) Rank-5 approximation  $\hat{A}(5)$ .

Measure the difference b/w  $\mathbf{A}$  and  $\hat{\mathbf{A}}$

## Spectral Norm

For  $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ , the spectral norm of  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is

$$\|\mathbf{A}\|_2 = \max_{\mathbf{x}} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2}$$

Measure the difference b/w  $\mathbf{A}$  and  $\hat{\mathbf{A}}$

## Spectral Norm

For  $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ , the spectral norm of  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is

$$\|\mathbf{A}\|_2 = \max_{\mathbf{x}} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2}$$

- Think about why we need to divide the norm  $\|\mathbf{x}\|_2$ .

## Theorem & Exercise

### Theorem (4.24)

The spectral norm of  $\mathbf{A}$  is its largest singular value  $\sigma_1$ .

# Eckart-Young Theorem

## Theorem [Eckart & Young 1936]

Consider  $\mathbf{A} \in \mathbb{R}^{m \times n}$  of rank  $r$  and let  $\mathbf{B} \in \mathbb{R}^{m \times n}$  be a matrix of rank  $k$ .

Then for any  $k \leq r$  with  $\hat{\mathbf{A}}(k) = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$ , it holds that

$$\hat{\mathbf{A}}(k) = \arg \min_{\text{rank}(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_2,$$

$$\|\mathbf{A} - \hat{\mathbf{A}}(k)\|_2 = \sigma_{k+1}.$$

Physical meaning:

- We can view the rank- $k$  approximation as a **projection** of the matrix  $\mathbf{A}$  onto a lower-dimensional space of rank-at-most- $k$  matrices.
- The approximation error: the next singular value (i.e.,  $\sigma_{k+1}$ )!

## Sketch of the Proof (1/2)

Note that

$$\mathbf{A} - \hat{\mathbf{A}}(k) = \sum_{i=k+1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^{\top}.$$

By Theorem 4.24, we have  $\|\mathbf{A} - \hat{\mathbf{A}}(k)\|_2 = \sigma_{k+1}$  (spectral norm).



## Sketch of the Proof (1/2)

Note that

$$\mathbf{A} - \hat{\mathbf{A}}(k) = \sum_{i=k+1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^{\top}.$$

By Theorem 4.24, we have  $\|\mathbf{A} - \hat{\mathbf{A}}(k)\|_2 = \sigma_{k+1}$  (spectral norm).

But why  $\hat{\mathbf{A}}$  is the *best* approximation in some sense?

## Sketch of the Proof (1/2)

Note that

$$\mathbf{A} - \hat{\mathbf{A}}(k) = \sum_{i=k+1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^{\top}.$$

By Theorem 4.24, we have  $\|\mathbf{A} - \hat{\mathbf{A}}(k)\|_2 = \sigma_{k+1}$  (spectral norm).

But why  $\hat{\mathbf{A}}$  is the *best* approximation in some sense?

Assume that  $r > k$  and there is another  $\mathbf{B}$  with  $\text{rank}(\mathbf{B}) \leq k$ , such that

$$\|\mathbf{A} - \mathbf{B}\|_2 < \|\mathbf{A} - \hat{\mathbf{A}}(k)\|_2.$$

## Sketch of the Proof (2/2)

Note that  $\dim(\text{null}(\mathbf{B})) \geq n - k$ .

## Sketch of the Proof (2/2)

Note that  $\dim(\text{null}(\mathbf{B})) \geq n - k$ .

There exists an  $\geq (n - k)$ -dimensional null space  $Z \subseteq \mathbb{R}^n$  such that

$$\mathbf{x} \in Z, \mathbf{B}\mathbf{x} = \mathbf{0}.$$

## Sketch of the Proof (2/2)

Note that  $\dim(\text{null}(\mathbf{B})) \geq n - k$ .

There exists an  $\geq (n - k)$ -dimensional null space  $Z \subseteq \mathbb{R}^n$  such that

$$\mathbf{x} \in Z, \mathbf{B}\mathbf{x} = \mathbf{0}.$$

$$\Rightarrow \|\mathbf{A}\mathbf{x}\|_2 = \|(\mathbf{A} - \mathbf{B})\mathbf{x}\|_2$$

## Sketch of the Proof (2/2)

Note that  $\dim(\text{null}(\mathbf{B})) \geq n - k$ .

There exists an  $\geq (n - k)$ -dimensional null space  $Z \subseteq \mathbb{R}^n$  such that

$$\mathbf{x} \in Z, \mathbf{B}\mathbf{x} = \mathbf{0}.$$

$$\Rightarrow \|\mathbf{A}\mathbf{x}\|_2 = \|(\mathbf{A} - \mathbf{B})\mathbf{x}\|_2 \leq \|\mathbf{A} - \mathbf{B}\|_2 \|\mathbf{x}\|_2$$

## Sketch of the Proof (2/2)

Note that  $\dim(\text{null}(\mathbf{B})) \geq n - k$ .

There exists an  $\geq (n - k)$ -dimensional null space  $Z \subseteq \mathbb{R}^n$  such that

$$\mathbf{x} \in Z, \mathbf{B}\mathbf{x} = \mathbf{0}.$$

$$\Rightarrow \|\mathbf{A}\mathbf{x}\|_2 = \|(\mathbf{A} - \mathbf{B})\mathbf{x}\|_2 \leq \|\mathbf{A} - \mathbf{B}\|_2 \|\mathbf{x}\|_2 < \sigma_{k+1} \|\mathbf{x}\|_2. \quad \cdots (\dagger)$$

## Sketch of the Proof (2/2)

Note that  $\dim(\text{null}(\mathbf{B})) \geq n - k$ .

There exists an  $\geq (n - k)$ -dimensional null space  $Z \subseteq \mathbb{R}^n$  such that

$$\mathbf{x} \in Z, \mathbf{B}\mathbf{x} = \mathbf{0}.$$

$$\Rightarrow \|\mathbf{A}\mathbf{x}\|_2 = \|(\mathbf{A} - \mathbf{B})\mathbf{x}\|_2 \leq \|\mathbf{A} - \mathbf{B}\|_2 \|\mathbf{x}\|_2 < \sigma_{k+1} \|\mathbf{x}\|_2. \quad \cdots (\dagger)$$

However, there exists a  $(k + 1)$ -dimensional subspace  $Y$  spanned by  $\mathbf{v}_1, \dots, \mathbf{v}_{k+1}$ .



## Sketch of the Proof (2/2)

Note that  $\dim(\text{null}(\mathbf{B})) \geq n - k$ .

There exists an  $\geq (n - k)$ -dimensional null space  $Z \subseteq \mathbb{R}^n$  such that

$$\mathbf{x} \in Z, \mathbf{B}\mathbf{x} = \mathbf{0}.$$

$$\Rightarrow \|\mathbf{A}\mathbf{x}\|_2 = \|(\mathbf{A} - \mathbf{B})\mathbf{x}\|_2 \leq \|\mathbf{A} - \mathbf{B}\|_2 \|\mathbf{x}\|_2 < \sigma_{k+1} \|\mathbf{x}\|_2. \quad \cdots (\dagger)$$

However, there exists a  $(k + 1)$ -dimensional subspace  $Y$  spanned by  $\mathbf{v}_1, \dots, \mathbf{v}_{k+1}$ .

Note that for  $\mathbf{x} \in Y$ ,  $\|\mathbf{A}\mathbf{x}\|_2 \geq \sigma_{k+1} \|\mathbf{x}\|_2. \quad \cdots (\ddagger) \quad (\text{Exercise!})$

## Sketch of the Proof (2/2)

Note that  $\dim(\text{null}(\mathbf{B})) \geq n - k$ .

There exists an  $\geq (n - k)$ -dimensional null space  $Z \subseteq \mathbb{R}^n$  such that

$$\mathbf{x} \in Z, \mathbf{B}\mathbf{x} = \mathbf{0}.$$

$$\Rightarrow \|\mathbf{A}\mathbf{x}\|_2 = \|(\mathbf{A} - \mathbf{B})\mathbf{x}\|_2 \leq \|\mathbf{A} - \mathbf{B}\|_2 \|\mathbf{x}\|_2 < \sigma_{k+1} \|\mathbf{x}\|_2. \quad \cdots (\dagger)$$

However, there exists a  $(k + 1)$ -dimensional subspace  $Y$  spanned by  $\mathbf{v}_1, \dots, \mathbf{v}_{k+1}$ .

Note that for  $\mathbf{x} \in Y$ ,  $\|\mathbf{A}\mathbf{x}\|_2 \geq \sigma_{k+1} \|\mathbf{x}\|_2. \quad \cdots (\ddagger) \quad (\text{Exercise!})$

But by the Dimension Theorem (rank-nullity theorem), there must be  $\mathbf{x} \in Y \cap Z. (\Rightarrow \times)$

# Sketch of the Proof of ( $\dagger$ )

- For any  $\mathbf{x} \in Y = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_{k+1}\}$ , write  $\mathbf{x} = \sum_{i=1}^{k+1} \alpha_i \mathbf{v}_i$ .
- $\mathbf{v}_i$ 's and  $\mathbf{u}_i$ 's are orthonormal  $\Rightarrow$

$$\|\mathbf{x}\|_2^2 = \sum_{i=1}^{k+1} \alpha_i^2, \quad \mathbf{Ax} = \sum_{i=1}^{k+1} \alpha_i \mathbf{Av}_i = \sum_{i=1}^{k+1} \alpha_i \sigma_i \mathbf{u}_i.$$

- Hence,

$$\|\mathbf{Ax}\|_2^2 = \left\| \sum_{i=1}^{k+1} \alpha_i \sigma_i \mathbf{u}_i \right\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 \alpha_i^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} \alpha_i^2 = \sigma_{k+1}^2 \|\mathbf{x}\|_2^2.$$

# Proof of Theorem 4.24

$$\|\mathbf{A}\|_2 = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} = \sup_{\mathbf{x} \neq \mathbf{0}} \left\| \mathbf{A} \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \right\|_2 = \sup_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2.$$

- Let  $\mathbf{B} := \mathbf{A}^\top \mathbf{A} \in \mathbb{R}^{n \times n}$ .
- Then  $\mathbf{B}$  is symmetric positive semidefinite and admits the eigen-decomposition  $\mathbf{A}^\top \mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^\top$ , for  $\mathbf{P} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$  containing orthonormal column vectors, and  $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$ ,  $\lambda_i \geq 0$ .

## Proof of Theorem 4.24 (Upper Bound)

Write  $\mathbf{x} = \sum_{i=1}^n a_i \mathbf{v}_i$ . Then

$$\begin{aligned}\|\mathbf{Ax}\|_2^2 &= \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} = \left( \sum_{i=1}^n a_i \mathbf{v}_i \right)^\top \left( \sum_{i=1}^n a_i \lambda_i \mathbf{v}_i \right) = \sum_{i=1}^n a_i^2 \lambda_i \\ &\leq \left( \max_{1 \leq i \leq n} \lambda_i \right) \sum_{i=1}^n a_i^2 = \left( \max_i \lambda_i \right) \|\mathbf{x}\|_2^2.\end{aligned}$$

Hence,

$$\|\mathbf{Ax}\|_2 \leq \sqrt{\max_i \lambda_i} \|\mathbf{x}\|_2 \quad \Rightarrow \quad \|\mathbf{A}\|_2 \leq \sqrt{\max_i \lambda_i}.$$

## Proof of Theorem 4.24 (Lower Bound and Final)

- Let  $\mathbf{v}_k$  be an eigenvector of  $\mathbf{A}^\top \mathbf{A}$  with  $\lambda_k = \max_i \lambda_i$  and  $\|\mathbf{v}_k\|_2 = 1$ . Then

$$\|\mathbf{A}\mathbf{v}_k\|_2^2 = \mathbf{v}_k^\top \mathbf{A}^\top \mathbf{A} \mathbf{v}_k = \lambda_k.$$

- Therefore  $\|\mathbf{A}\|_2 \geq \|\mathbf{A}\mathbf{v}_k\|_2 \geq \sqrt{\lambda_k} = \sqrt{\max_i \lambda_i}$ . Combining with the upper bound gives

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^\top \mathbf{A})} = \sigma_{\max}(\mathbf{A}) = \max_i \sigma_i,$$

where  $\sigma_i$  are the singular values of  $\mathbf{A}$  and  $\lambda_i = \sigma_i^2$ .

# Discussions