Linear Algebra

# Singular Value Decomposition

Department of Computer Engineering

Sharif University of Technology

Hamid R. Rabiee  rabiee@sharif.edu

Maryam Ramezani maryam.ramezani@sharif.edu

# Table of contents

# 01

# SVD Introduction

# SVD Introduction

- Generalization of the spectral decomposition that applies to all matrices, rather than just normal matrices.
- Applications:
  - Compute the size of a matrix (in a way that typically makes more sense than norm)
  - Provide a new geometric interpretation of linear transformations
  - Solve optimization problems
  - Construct an “almost inverse” for matrices that do not have an inverse.

# SVD

- Given any $m \times n$ matrix **A**, algorithm to find matrices **U**, **V**, and $\sum$ such that (**always exists**)
- $A = U\Sigma V^T$ $A = U\Sigma V^*$

  **$U$ is $m \times m$ and orthogonal (always real)**

  **$\sum$ is $m \times n$ and diagonal with non-negative (always real) called singular values.**

  **$V$ is $n \times n$ and orthogonal (always real)**
- Columns of U are eigenvectors of $AA^T$ (called the left singular vectors).
- Columns of V are eigenvectors of $A^T A$ (called the right singular vectors).
- The non-zero singular vectors are the positive square roots of non-zero eigenvalues of $AA^T$ or $A^T A$.

# SVD

- The $\Sigma_i$ are called the singular values of **A**
- If **A** is singular, some of the $\Sigma_i$ will be 0
- In general *rank*(**A**) = number of nonzero $\Sigma_i$
- SVD is mostly unique (up to permutation of singular values, or if some $\Sigma_i$ are equal)

# Columns of V are eigenvectors of $A^T A$

- Assume $A$ with singular value decomposition $A = U\Sigma V^T$. Let's take a look at the eigenpairs corresponding to $A^T A$:

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

Hence $A^T A = V\Sigma^2 V^T$

- Recall that columns of $V$ are all linear independent (orthogonal matrix), then from diagonalization ($B = XDX^{-1}$), we get:
  - The columns of $V$ are the eigenvectors of the matrix $A^T A$
  - The diagonal entries of $\Sigma^2$ are the eigenvalues of $A^T A$
- Let's call $\lambda$ the eigenvalues of $A^T A$, then $\sigma_i^2 = \lambda_i$

# Columns of U are eigenvectors of $AA^T$

- In a similar way,

$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T$$

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

Hence $AA^T = U\Sigma^2 U^T$

- Recall that columns of $U$ are all linear independent (orthogonal matrix), then from diagonalization ($B = XDX^{-1}$), we get:
  - The columns of $U$ are the eigenvectors of the matrix $AA^T$

# How can we compute an SVD of a matrix A?

1. Evaluate the n eigenvectors $v_i$ and eigenvalues $\lambda_i$ of $A^T A$
2. Make a matrix $V$ from the normalized vectors $v_i$. The columns are called "right singular vectors".

$$\mathrm{V} = \begin{pmatrix} \vdots & \cdots & \vdots \\ v_1 & \cdots & v_n \\ \vdots & \cdots & \vdots \end{pmatrix}$$

3. Make a diagonal matrix from the square roots of the eigenvalues.

$$\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \quad \sigma_i = \sqrt{\lambda_i} \quad \text{and } \sigma_1 \geq \sigma_2 \geq \cdots$$

4. Find $U: A = U\Sigma V^T \Rightarrow U\Sigma = AV \Rightarrow U = AV\Sigma^{-1}$. The columns are called "left singular values".

# Example

$$S = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} \to S^T S = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}, rank(S) = 1$$

$$\Delta(\lambda) = \lambda^2 - 18\lambda = 0 \Rightarrow \sigma_1 = \sqrt{18}, \sigma_2 = 0 \Rightarrow \boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{3\sqrt{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, v_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \Rightarrow \boldsymbol{V} = \frac{\mathbf{1}}{\sqrt{\mathbf{2}}} \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{-1} & \mathbf{1} \end{bmatrix}$$

$$Sv_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix} \Rightarrow u_1 = \frac{1}{\sigma_1} Sv_1 = \frac{1}{3\sqrt{2}} \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}$$

$$u_2 = \begin{bmatrix} 2/3 \\ 2/3 \\ 1/3 \end{bmatrix}, u_3 = \begin{bmatrix} -2/3 \\ 1/3 \\ 2/3 \end{bmatrix} \Rightarrow \boldsymbol{U} = \frac{\mathbf{1}}{\mathbf{3}} \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{-2} \\ \mathbf{-2} & \mathbf{2} & \mathbf{1} \\ \mathbf{2} & \mathbf{1} & \mathbf{2} \end{bmatrix}$$

$$S = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 2 & -2 \\ -2 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = U\Sigma V^T$$

02

# Reduced SVD

# SVD for Square Matrix

- The SVD is a factorization of a m x n matrix into

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

Where U is a m x m orthogonal matrix, $V^T$ is a n x n orthogonal matrix and $\Sigma$ is a m x n diagonal matrix.
For a square matrix (m=n):

$$A = \begin{pmatrix} \vdots & \cdots & \vdots \\ u_1 & \cdots & u_n \\ \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \begin{pmatrix} \cdots & v_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & v_n^T & \cdots \end{pmatrix}$$

$$A = \begin{pmatrix} \vdots & \cdots & \vdots \\ u_1 & \cdots & u_n \\ \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \begin{pmatrix} \vdots & \cdots & \vdots \\ v_1 & \cdots & v_n \\ \vdots & \cdots & \vdots \end{pmatrix}^T$$

# Reduced SVD

$$[Sv_1 \quad \dots \quad Sv_r \quad 0 \quad \dots \quad 0]_{m\times n} = [\sigma_1 u_1 \quad \dots \quad \sigma_r u_r \quad 0 \quad \dots \quad 0]_{m\times n}$$
$$[Sv_1 \quad \dots \quad Sv_r \quad Sv_{r+1} \quad \dots \quad Sv_n]_{m\times n} = [\sigma_1 u_1 \quad \dots \quad \sigma_r u_r \quad 0 \quad \dots \quad 0]_{m\times n}$$

$$S[v_1 \quad \dots \quad v_n] = [u_1 \quad \dots \quad u_m]\left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & & \vdots & 0 \\ 0 & \cdots & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right]$$

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

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

# Reduced SVD

- what happens when A is not a square matrix?
- n > m

$$A = U\Sigma V^T = \begin{pmatrix} \vdots & \cdots & \vdots \\ u_1 & \cdots & u_m \\ \vdots & \cdots & \vdots \end{pmatrix}_{m\times m} \begin{pmatrix} \sigma_1 & & & 0 & \\ & \ddots & & & \ddots \\ & & \sigma_m & & & 0 \end{pmatrix}_{m\times n} \begin{pmatrix} \cdots & v_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & v_m^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & v_n^T & \cdots \end{pmatrix}_{n\times n}$$

We can instead rewrite the above as:

$$A = U\Sigma_R V_R^T$$

where $V_R$ is n x m matrix and $\Sigma_R$ is a m x m matrix

**In general:**

$$A = U_R\Sigma_R V_R^T$$

Now U and V are not orthogonal. But their columns are orthonormal.

$U_R$ is a m x k matrix

$\Sigma_R$ is a k x k matrix

$V_R$ is a n x k matrix

k = min(m, n)

# Reduced SVD

- m > n

$$A = U\Sigma V^T = \begin{pmatrix} \vdots & \cdots & \vdots & \cdots & \vdots \\ u_1 & \cdots & u_n & \cdots & u_m \\ \vdots & \cdots & \vdots & \cdots & \vdots \end{pmatrix}_{m\times m} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}_{m\times n} \begin{pmatrix} \cdots & v_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & v_n^T & \cdots \end{pmatrix}_{n\times n}$$

We can instead rewrite the above as:

$$A = U\Sigma_R V_R^T$$

where $U_R$ is m x n matrix and $\Sigma_R$ is a n x n matrix

**Now U and V are not orthogonal. But their columns are orthonormal.**

# Reduced SVD

- Let's take a look at the product of $\Sigma^T\Sigma$ where $\Sigma$ has the singular values of a $A$, a m x n matrix.
  - **m > n:**

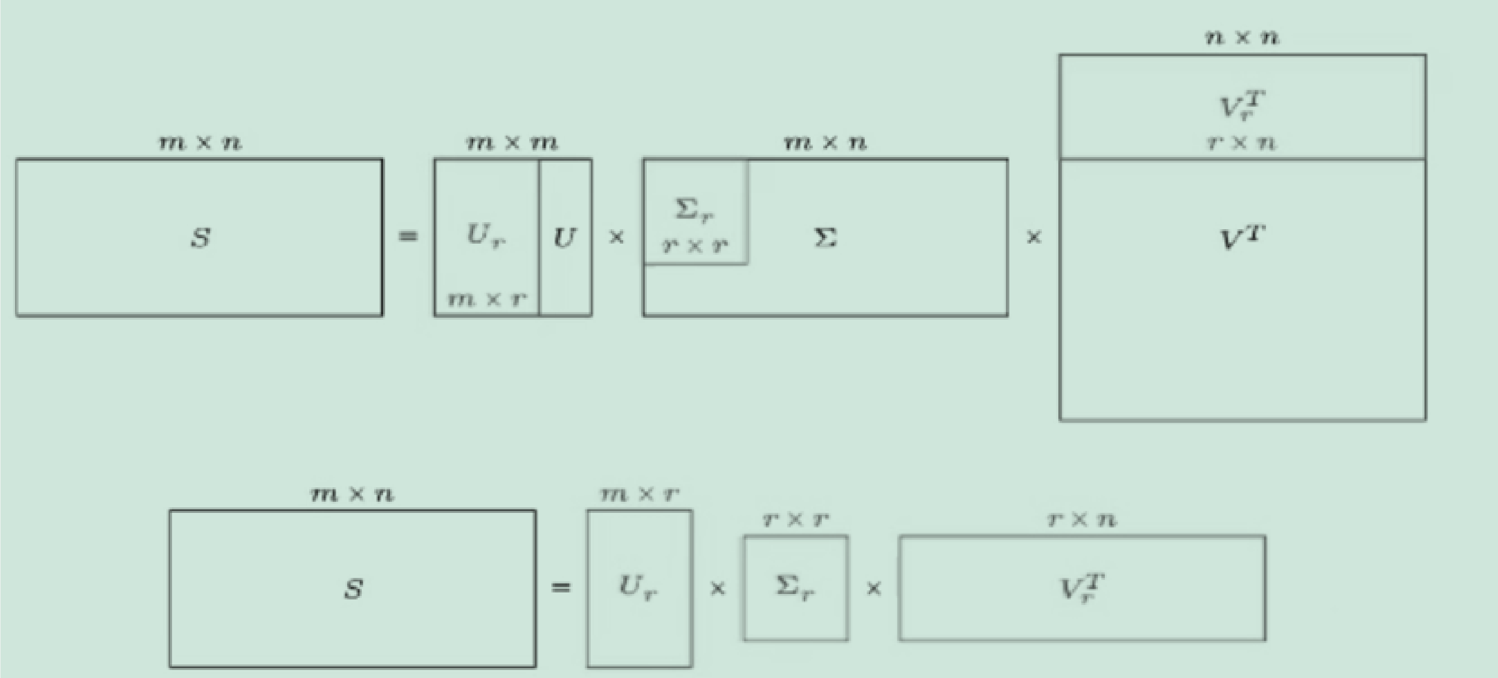
$$\Sigma^T\Sigma = \begin{pmatrix} \sigma_1 & & & 0 & & \\ & \ddots & & & \ddots & \\ & & \sigma_n & & & 0 \end{pmatrix}_{n\times m} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}_{m\times n} = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix}_{n\times n}$$

  - **n > m:**

$$\Sigma^T\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}_{n\times m} \begin{pmatrix} \sigma_1 & & & 0 & & \\ & \ddots & & & \ddots & \\ & & \sigma_m & & & 0 \end{pmatrix}_{m\times n} = \begin{pmatrix} \sigma_1^2 & & & 0 & & \\ & \ddots & & & \ddots & \\ & & \sigma_m^2 & & & 0 \\ 0 & & & 0 & & \\ & \ddots & & & \ddots & \\ & & 0 & & & 0 \end{pmatrix}_{n\times n}$$

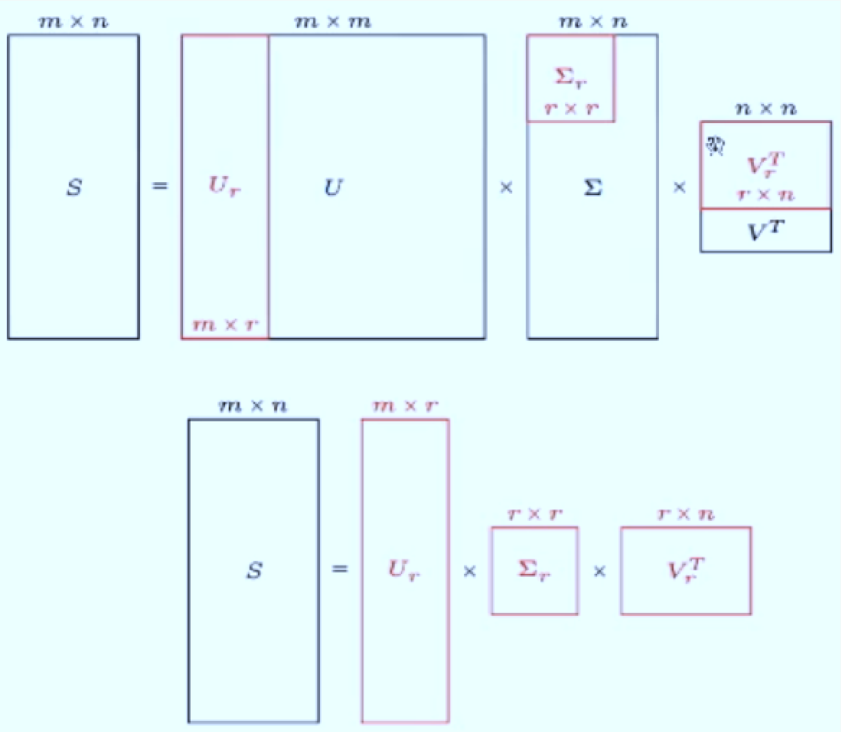
# Reduced SVD

- Wide Matrix



# Reduced SVD

- Tall Matrix



# SVD Comparison

| SVD | Diagonalization | Spectral Decomposition |
|---|---|---|
| applies to every single matrix (even rectangular ones). | only applies to matrices with a basis of eigenvectors | only applies to normal matrices |
| matrix $\sum$ in the middle of the SVD is diagonal (and even has real non-negative entries) | do not guarantee an entrywise non-negative matrix | do not guarantee an entrywise non-negative matrix |
| It requires two unitary matrices U and V | only required one invertible matrix | only required one unitary matrix |

# Lemma

- Unitary Freedom of PSD Decompositions

Suppose $B, C \in \mathcal{M}_{m,n}(\mathbb{F})$. The following are equivalent:

a. There exists a unitary matrix $U \in \mathcal{M}_m(\mathbb{F})$ such that $C = UB$,

b. $B^*B = C^*C$,

## Example

$$\begin{bmatrix} 3 & 2 \\ -2 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 1 \\ 3 & 2 & 1 \end{bmatrix}$$

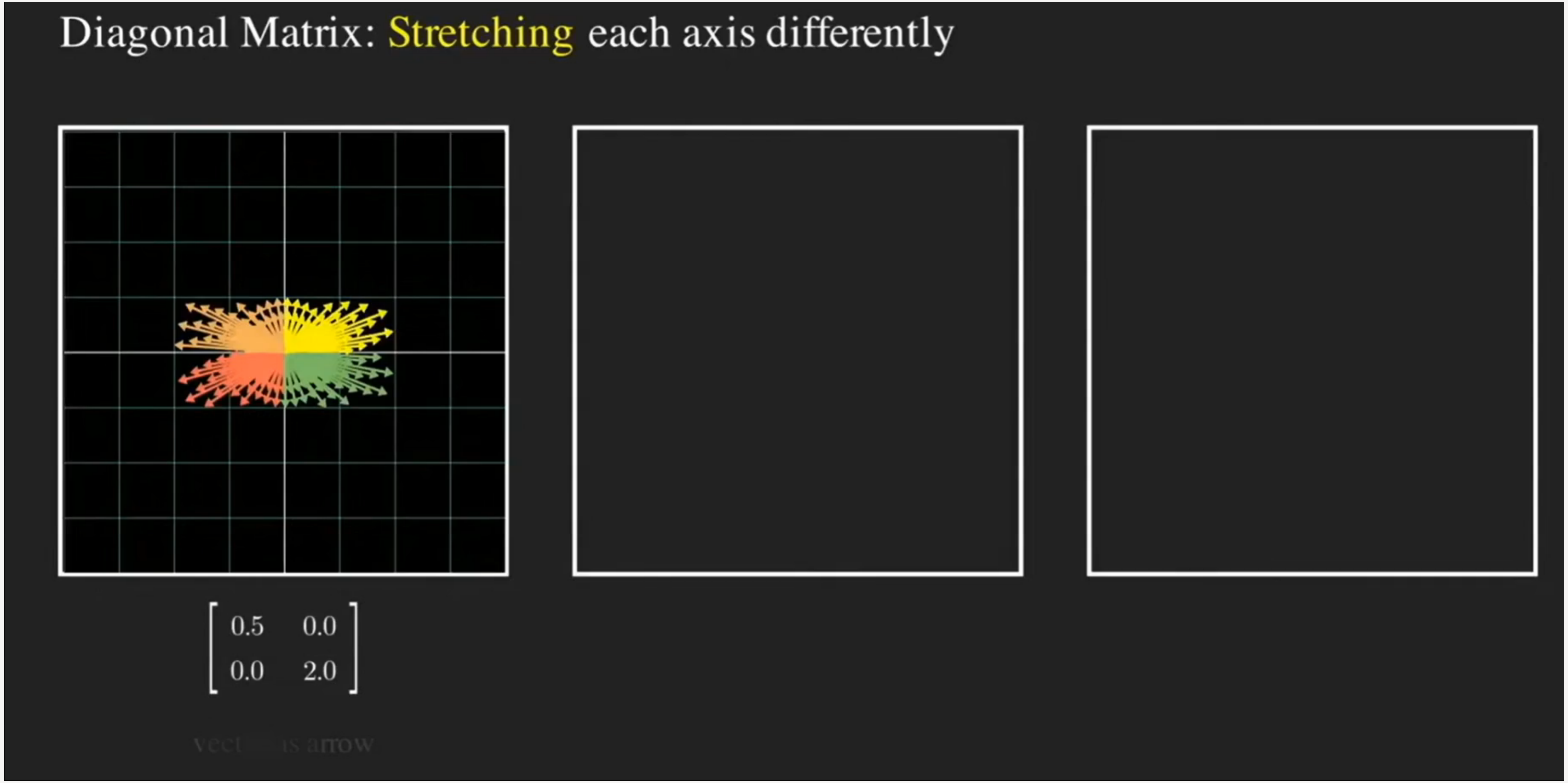
# SVD Proof

- If $m \neq n$ then $A^*A, AA^*$ have different sizes, but they still have essentially the same eigenvalues—whichever one is larger just has some extra 0 eigenvalues.
- The same is actually true of AB and BA for any A and B.

- Proof SVD in another view!!

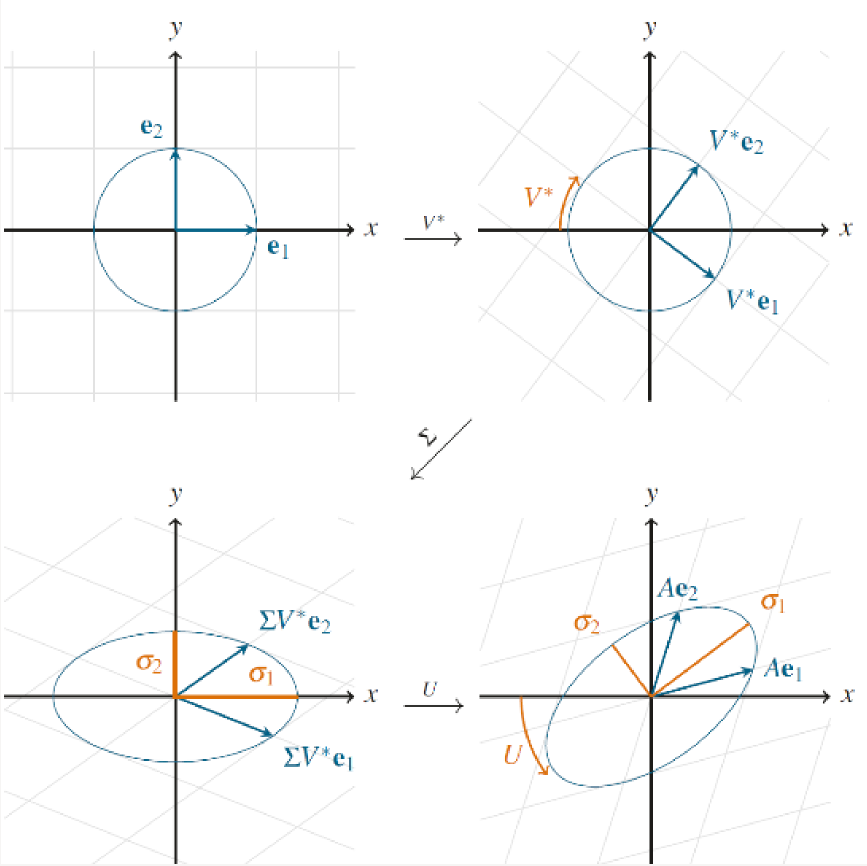
# 03

# SVD Intuition


Diagonal Matrix: Stretching each axis differently
0.5 0.0
0.0 2.0

$A = U\Sigma V^*$

The product of a matrix's singular values equals the absolute value of its determinant



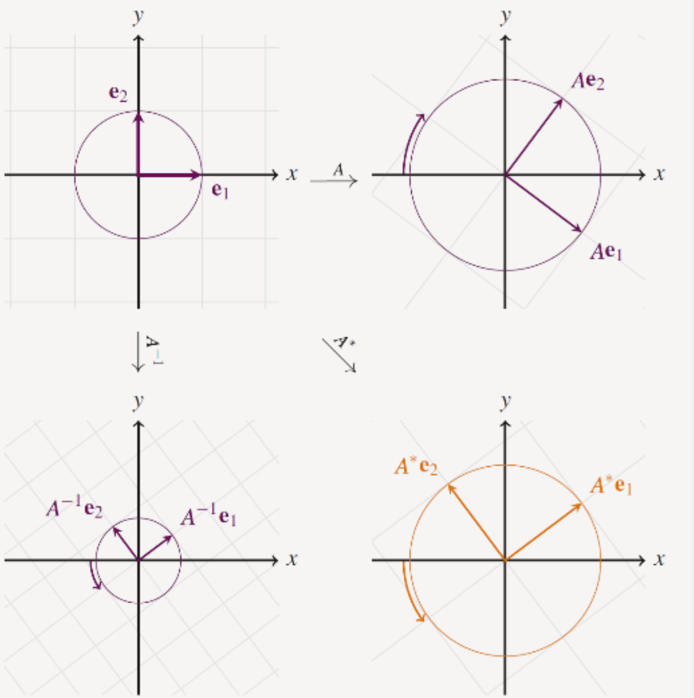

24

# A Geometric Interpretation

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

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

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



04

# SVD in Problem Solving

# Determining the rank of a matrix

- Suppose $A$ is a m x n rectangular matrix where m > n:

$$A = \begin{pmatrix} \vdots & \cdots & \vdots & \cdots & \vdots \\ u_1 & \cdots & u_n & \cdots & u_m \\ \vdots & \cdots & \vdots & \cdots & \vdots \end{pmatrix}_{m\times m} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}_{m\times n} \begin{pmatrix} \cdots & v_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & v_n^T & \cdots \end{pmatrix}_{n\times n}$$

$$A = \begin{pmatrix} \vdots & \cdots & \vdots \\ u_1 & \cdots & u_n \\ \vdots & \cdots & \vdots \end{pmatrix} \begin{pmatrix} \cdots & \sigma_1 v_1^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \sigma_n v_n^T & \cdots \end{pmatrix} = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots + \sigma_n u_n v_n^T$$

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T$$

$A_1 = \sigma_1 u_1 v_1^T$ what is rank$(A_1) = ?$

In general, rank$(A_k) = k$

# SVD and Rank

- Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A \in \mathcal{M}_{m,n}(\mathbb{F})$ has rank$(A) = r$. There exist orthonormal sets of vectors $\{u_j\}_{j=1}^{r} \subset \mathbb{F}^m$ and $\{v_j\}_{j=1}^{r} \subset \mathbb{F}^n$ such that

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^*,$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the non-zero singular values of A.

# Conclusion

- Let $A \in \mathcal{M}_{m,n}$ be a matrix with rank(A) = r and the singular value decomposition $A = U\Sigma V^T$, where

$$U = [u_1 \mid u_2 \mid \ \dots \mid u_m] \text{ and } V = [v_1 \mid v_2 \mid \ \dots \mid v_n]$$

Then

a. $\{u_1, u_2, \dots, u_r\}$ is an orthonormal basis of range($A$),
b. $\{u_{r+1}, u_{r+2}, \dots, u_m\}$ is an orthonormal basis of null($A^*$),
c. $\{v_1, v_2, \dots, v_r\}$ is an orthonormal basis of range($A^*$), and
d. $\{v_{r+1}, v_{r+2}, \dots, v_n\}$ is an orthonormal basis of null($A$)

# SVD and Matrix Similarity

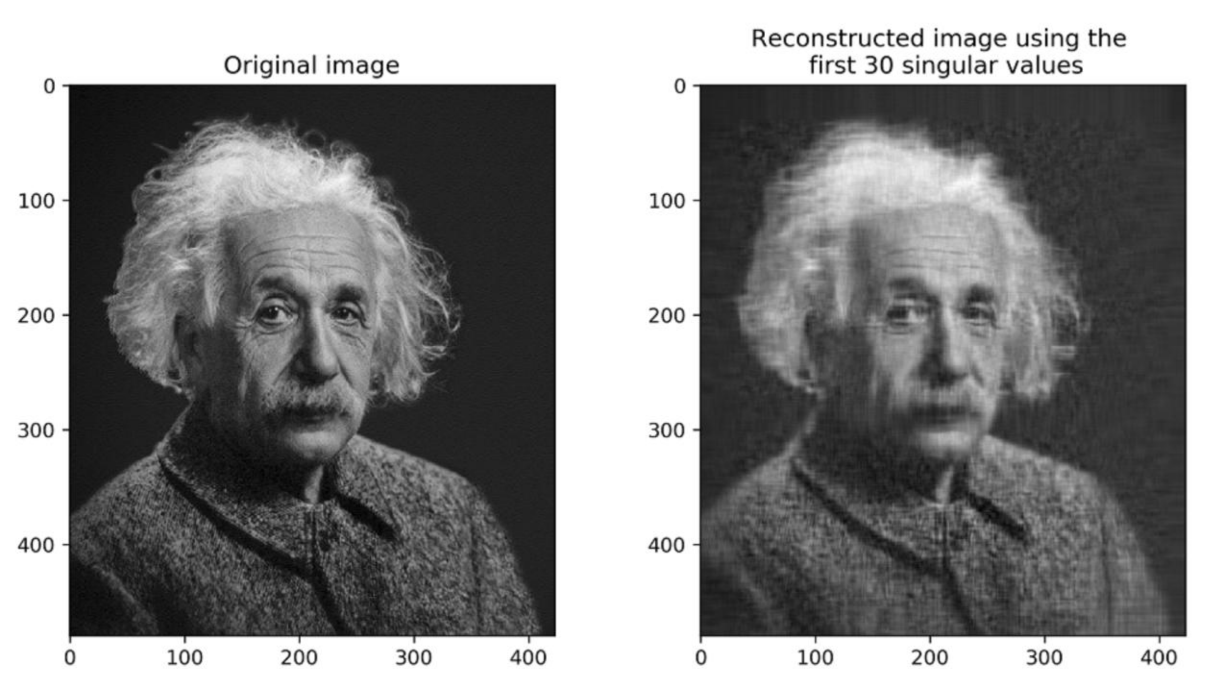
- Suppose you want to find best rank-*k* approximation to **A**
  - Answer: set all but the largest *k* singular values to zero
- Can form compact representation by eliminating columns of **U** and **V** corresponding to zeroed $\Sigma_i$

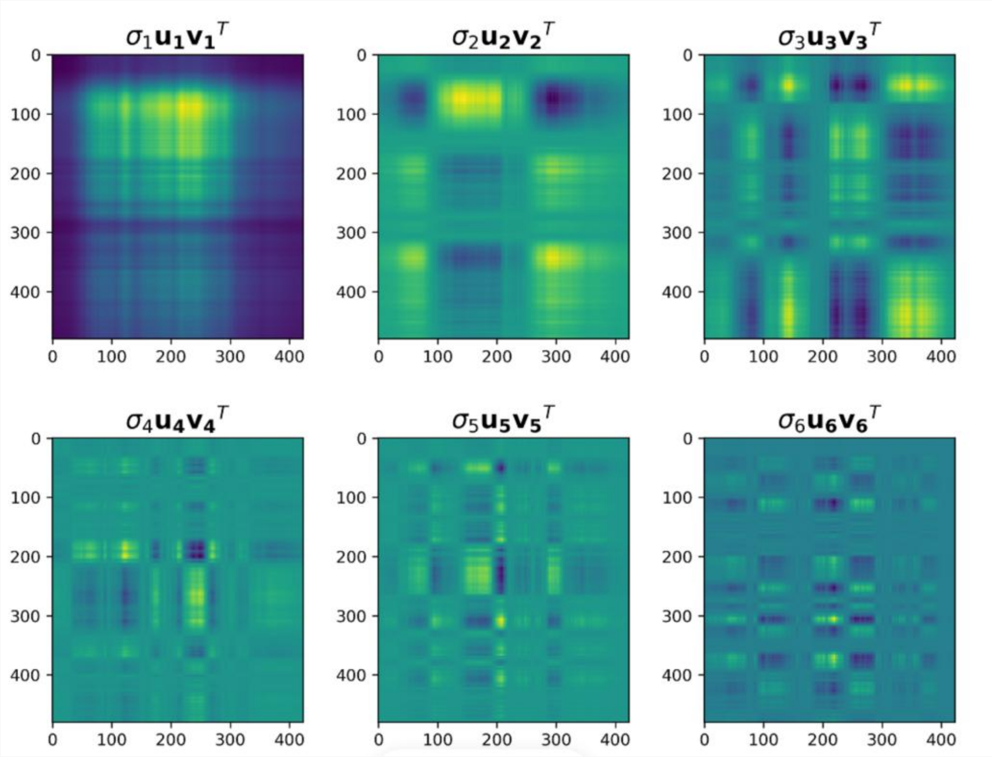
# SVD and PSD

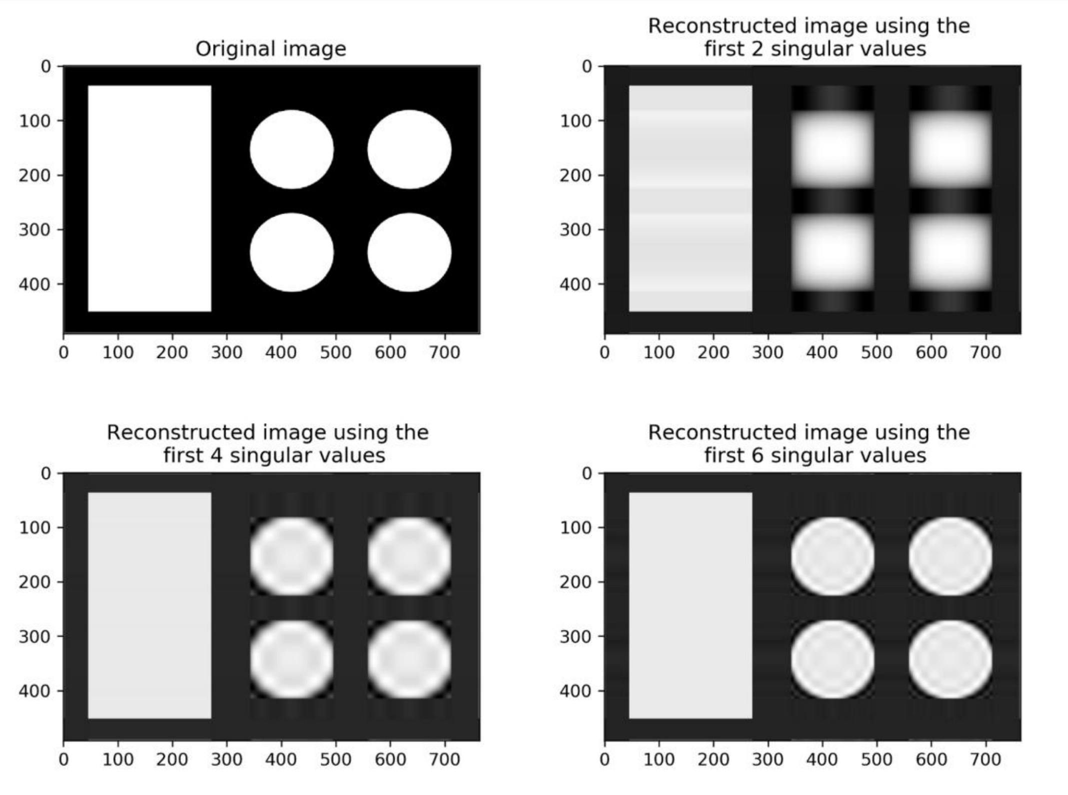
- If $A \in \mathcal{M}_n$ is positive semidefinite then its singular values equals its eigenvalues.

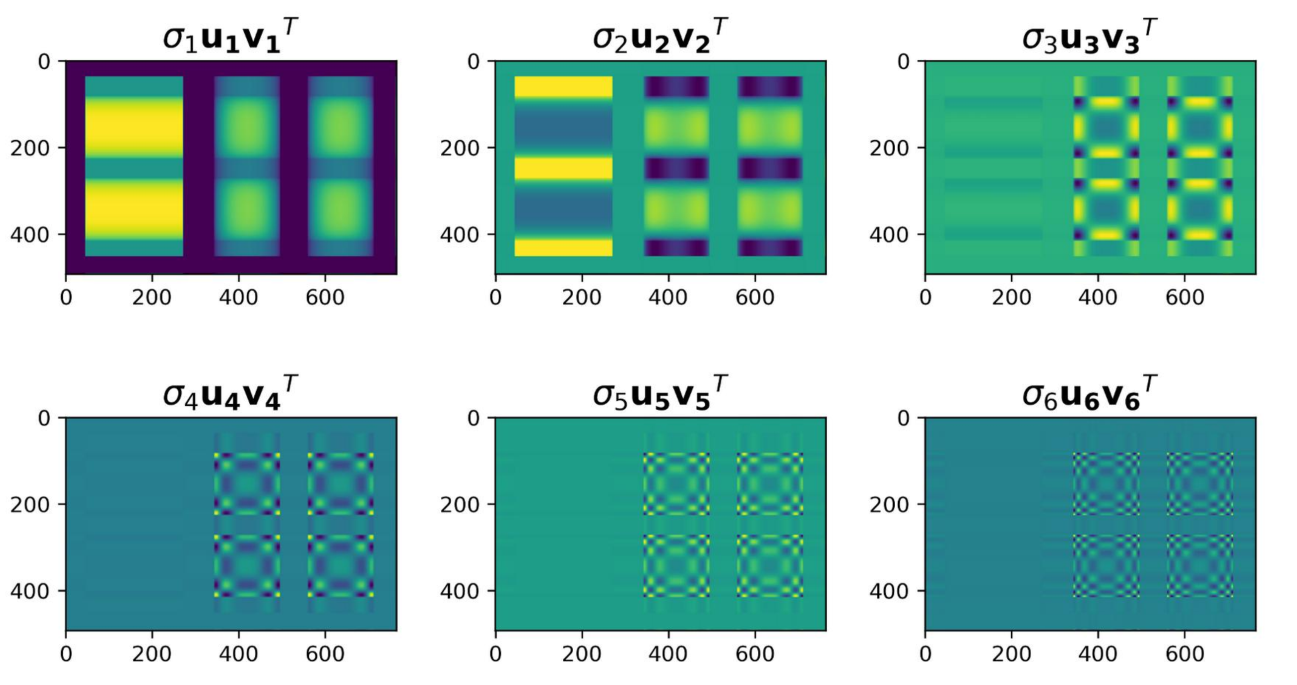
# 05

# Applications


Original image
Reconstructed image using the first 30 singular values
0
100
200
300
400


$\sigma_1 \mathbf{u_1} \mathbf{v_1}^T$
$\sigma_2 \mathbf{u_2} \mathbf{v_2}^T$
$\sigma_3 \mathbf{u_3} \mathbf{v_3}^T$
$\sigma_4 \mathbf{u_4} \mathbf{v_4}^T$
$\sigma_5 \mathbf{u_5} \mathbf{v_5}^T$
$\sigma_6 \mathbf{u_6} \mathbf{v_6}^T$
0
100
200
300
400


Original image
Reconstructed image using the first 2 singular values
Reconstructed image using the first 4 singular values
Reconstructed image using the first 6 singular values


$\sigma_1 \mathbf{u_1} \mathbf{v_1}^T$
$\sigma_2 \mathbf{u_2} \mathbf{v_2}^T$
$\sigma_3 \mathbf{u_3} \mathbf{v_3}^T$
$\sigma_4 \mathbf{u_4} \mathbf{v_4}^T$
$\sigma_5 \mathbf{u_5} \mathbf{v_5}^T$
$\sigma_6 \mathbf{u_6} \mathbf{v_6}^T$
0
200
400
600

# Low Rank Approximation of Image

06

# Pseudo Inverse

# SVD and Inverses

- Why is SVD so useful?
- $A^{-1} = V\Sigma^{-1}U^{-1} = V\Sigma^{-1}U^T$
  - Using fact that inverse = transpose for orthogonal matrices
  - Since $\Sigma$ is diagonal, $\Sigma^{-1}$ also diagonal with reciprocals of entries of $\Sigma$
- This fails when some $\Sigma_i$ are 0
  - It's *supposed* to fail – singular matrix
- Pseudoinverse: if $\Sigma_i = 0$, set $\frac{1}{\Sigma_i}$ to 0 (!)
  - "Closest" matrix to inverse
  - Defined for all (even non-square, singular, etc.) matrices
  - Equal to $(A^TA)^{-1}A^T$ if $A^TA$ invertible

# Pseudo Inverse

- Problem:

  if A is rank-deficient, $\Sigma$ is not invertible.
- How to fix it:

  Define the Pseudo Inverse
- Pseudo Inverse of a diagonal matrix:

$$(\Sigma^+)_i = \begin{cases} \dfrac{1}{\sigma_i}, & if\ \sigma_i \neq 0 \\ 0, & if\ \sigma_i = 0 \end{cases}$$

- Pseudo Inverse of a matrix A:

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

# Moore-Penrose inverse (Pseudo Inverse)

- If a matrix A has the singular value decomposition

$$A = UWV^T$$

then the pseudo-inverse or Moore-Penrose inverse of A is

$$A^+ = VW^{-1}U^T$$

### 1.3 Moore-Penrose Conditions

For a matrix $A^+$ to be the pseudoinverse of $A$, it must satisfy the following four **Moore-Penrose conditions:**

1. **(MP1)** $AA^+A = A$
2. **(MP2)** $A^+AA^+ = A^+$
3. **(MP3)** $(AA^+)^T = AA^+$
4. **(MP4)** $(A^+A)^T = A^+A$

These conditions ensure that $A^+$ behaves similarly to an inverse, even when $A$ is not invertible.

# Pseudo Inverse

$$A^+ = VW^{-1}U^T$$

If A is 'tall' (m > n) and has full rank

$$A^+ = (A^T A)^{-1} A^T$$ (it gives the least-squares solution $x_{lsq} = A^+ b$)

If A is 'short' (n > m) and has full rank

$A^+ = A^T(AA^T)^{-1}$ (it gives the least-norm solution $x_{l-n} = A^+ b$)

- In general, $x_{pinv} = A^+ b$ is the minimum-norm, least-square solution.

$$x^* = A^{-1}b = (UDV^T)^{-1}b,$$

- $(UDV^T)^{-1} = V^{-T}\, D^{-1}\, U^{-1}$

Moore–Penrose pseudoinverse

$$x^* = A^{-1}b = \boxed{VD^{-1}U^T}b$$

- Invert the diagonal entries in D that are nonzero, but leave the other diagonal entries alone as zeros.