

# Singular Value Decomposition and Pseudo-inverse

JrPhy

# Introduction

- SVD always appears in the last part of a Linear Algebra book, the reason is SVD includes the key ideas in Linear Algebra.
- Not only in teaching, but in many applications, like computer vision, statistic, etc.
- I'll prove this theorem, but it bases on many theorem, those won't be proved in this ppt.

- Definition:
- Let  $A$  in  $M_{m \times n}(\mathbb{C})$ . The unique scalars  $\sigma_1, \sigma_2, \dots, \sigma_n$  in SVD are called the singular values of  $A$ .

- Singular Value Decomposition Theorem:
- Let  $A$  is  $M_{m \times n}(\mathbb{C})$  and  $\text{rank}(A) = r$  with positive scalar singular values  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ .  $A$  factorization

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

- is called a singular value decomposition of  $A$ .  
Where  $U$  is  $M_{m \times m}(\mathbb{C})$  and  $V$  is  $M_{n \times n}(\mathbb{C})$  are unitary matrices and  $\Sigma$  is  $M_{m \times n}(\mathbb{C})$  and  $\Sigma_{ij} = \sigma_i \delta_{ij}$

# Prove the theorem of SVD

- $\because U, V$  are unitary,  $\therefore AV = USV^*V = U\Sigma$

So that there exists 2 orthonormal basis  $\beta = \{v_1, v_2, \dots, v_n\}$  and  $\gamma = \{u_1, u_2, \dots, u_m\}$  in finite-dimension inner product space  $V$  and  $U$  and a linear transform  $T: F^n \rightarrow F^m$  of rank  $r$ , and  $\Sigma$  can be written as

$$\Sigma_{ij} = \sigma_i \delta_{ij}, (1 \leq i \leq r), \Sigma_{ij} = 0, \text{ otherwise}$$

$$\text{Then } T(v_i) = \sigma_i u_i, 1 \leq i \leq r, T(v_i) = 0, i > r$$

# Prove the theorem of SVD

- So that we have

$$\begin{aligned}\langle u_i, u_j \rangle &= \left\langle \frac{1}{\sigma_i} T(v_i), \frac{1}{\sigma_j} T(v_j) \right\rangle = \frac{1}{\sigma_i \sigma_j} \langle T^* T(v_i), v_j \rangle \\ &= \frac{\sigma_i^2}{\sigma_i \sigma_j} \langle v_i, v_j \rangle = \delta_{ij}\end{aligned}$$

- So  $\sigma_{ij}^2$  is the eigenvalue of  $T^*T$  and  $v_i$  is the corresponding eigenvector.

# Procedure for calculating SVD

1. Compute  $A^*A$
2. Compute eigenvalue of  $A^*A$  its corresponding normalized eigenvector. The square root of eigenvalues are  $\sigma_i$  It forms  $\Sigma$  and  $V$  matrix
3. Compute  $AV/\sigma_i$  to get  $U$  matrix

Note that the size of is the same as  $A$ . If  $A$  is  $m$  by  $n$  matrix and  $m > n$ , then number of  $\sigma$  is  $n$ .

# Eigenvalues of $AA^*$

- By SVD theorem,  $A = U\Sigma V^*$ ,

$$A^* = (U\Sigma V^*)^* = V\Sigma^*U^* = V\Sigma U^* \rightarrow A^*U = V\Sigma$$

- Form the viewpoint of linear transformation,

$$\begin{aligned}\langle v_i, v_j \rangle &= \left\langle \frac{1}{\sigma_i} T^*(u_i), \frac{1}{\sigma_j} T^*(u_j) \right\rangle = \frac{1}{\sigma_i \sigma_j} \langle T^* T(u_i), u_j \rangle \\ &= \frac{\sigma_i^2}{\sigma_i \sigma_j} \langle u_i, u_j \rangle = \delta_{ij}\end{aligned}$$

- The eigenvalues of  $AA^*$  and  $A^*A$  are the same!

# Example

- Give a system of eq:  $\begin{cases} x_1 + x_2 - x_3 = 1 \\ x_1 + x_2 - x_3 = 2 \end{cases} \rightarrow \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$$AA^T = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \rightarrow \left\| \begin{bmatrix} 3-\lambda & 3 \\ 3 & 3-\lambda \end{bmatrix} \right\| = \lambda(6-\lambda) = 0$$

$$A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & -2 \\ 2 & 2 & -2 \\ -2 & -2 & 2 \end{bmatrix}$$

**$AA^T$  and  $A^T A$**   
has the same  
eigenvalue!

$$\rightarrow \left\| \begin{bmatrix} 2-\lambda & 2 & -2 \\ 2 & 2-\lambda & -2 \\ -2 & -2 & 2-\lambda \end{bmatrix} \right\| = \lambda^2(6-\lambda) = 0$$



# Example

- $\because A = U\Sigma V^T \rightarrow A^T = (U\Sigma V^T)^T = V\Sigma^T U^T$
- $AA^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma^2 U^T$
- $A^T A = V\Sigma^T U^T U\Sigma V^T = V^T \Sigma^2 V$
- But the corresponding eigenvector are  $U$  and  $V$  respectively,  $AA^T$  to get  $U$ ,  $A^T A$  to get  $V$ .
- It's easier that we use  $AV = \Sigma U$  or  $UA = \Sigma V^T$  to get its SVD. In this example,

# Example

$$AV = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix} \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & -\sqrt{3} & 1 \\ -\sqrt{2} & 0 & 2 \end{bmatrix} = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ \sqrt{3} & 0 & 0 \end{bmatrix}$$

- $U$  is  $M_{2 \times 2}(\mathbb{R})$ , so  $u_1 = \frac{1}{\sigma_1} Av_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} \\ \sqrt{3} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
- And  $(AV)_{21}$  is  $\frac{1}{\sqrt{2}}$  so  $u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \rightarrow U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

$$A = U\Sigma V^T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \end{pmatrix}$$

# Store $U\Sigma V^T$ elements

- In the example, we observe that

$$A = U\Sigma V^T = \begin{pmatrix} 1/\sqrt{2} & * \\ 1/\sqrt{2} & * \end{pmatrix} \begin{pmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ * & * & * \\ * & * & * \end{pmatrix}$$

- $u_2$ ,  $v_2$  and  $v_3$  has no effect about reconstructing  $A$ , so that in computation, it does not store full  $U$  and  $V$ , and  $\Sigma$  is the sparse matrix, which represents 0 is much more than other elements. It can save memory.

# The inverse matrix

- Matrix  $A$  is invertible  $\iff A$  is square  
 $\det(A) \neq 0$
- The definition of the a invertible is:
- If  $A$  is in  $M_{m \times m}(\mathbb{R})$  invertible, then there exists a matrix  $B$  in  $M_{m \times m}(\mathbb{R})$  such that  $AB = BA = I_m$
- $AI = A(AB) = A(BA) = A$
- And it uses to solve least-square problem.

# The Pseudo-inverse

- The Pseudo-inverse matrix is a generalized of a inverse matrix,
- If  $A$  is in  $M_{m \times n}(\mathbb{R})$ ,  $A$  is not invertible, but its Pseudo-inverse matrix exists.
- The Pseudo-inverse is useful for solving least-square problem, even if the matrix is not full rank.

# Definition of Pseudo-inverse

- Let  $A$  is in  $M_{m \times n}(\mathbb{R})$ , then the solution  $X$  is the Pseudo-inverse matrix of  $A$  satisfies the following eqs:
- $AXA = A, XAX = X, (AX)^* = XA, (XA)^* = AX$
- The Pseudo-inverse can be used to solve the least-square problem.
- If  $Ax = b$  and  $A$  is full rank  $\Leftrightarrow x_0 = (A^T A)^{-1} A^T b$
- If  $Ax = b$  and  $A$  is rank-deficient  $\Leftrightarrow x^+ = A^+ b$

# Pseudo-inverse by SVD

Theorem:

- $A$  is  $M_{m \times n}(\mathbb{C})$  and  $A = U\Sigma V^*$ , then  $A^+ = V\Sigma^+U^*$  (Pseudo-inverse of  $A$ ),  $\Sigma^+$  is the pseudo-inverse of  $\Sigma$ .
- $\Sigma_{ij} = \sigma_i \delta_{ij}$ ,  $(\Sigma_{ij})^+ = \Sigma_{ji} = \sigma_j^{-1} \delta_{ji}$

Note that  $AA^+$  and  $A^+A$  are not equal to identity matrix necessarily

# Least-square solution

Theorem:

- a. If  $A\mathbf{x} = \mathbf{b}$  is consistent, then  $\mathbf{z}$  is the unique solution to the system having minimum norm.
- b. If  $A\mathbf{x} = \mathbf{b}$  is inconsistent, then  $\mathbf{z}$  is the unique best approximation to a solution having minimum norm.

Part a has been proved in another [ppt](#), so here we prove part b.



# Least-square solution

*Proof:*

Suppose  $\mathbf{x}_0$  is the best-approximation solution of  $A\mathbf{x} = \mathbf{b}$ , then  $\|A\mathbf{x} - \mathbf{b}\| \geq \|A\mathbf{x}_0 - \mathbf{b}\|$

$$A\mathbf{x} - \mathbf{b} = A(\mathbf{x} - A^+\mathbf{b}) - (I - AA^+)(-\mathbf{b})$$

Since those vector forms a right triangle by Pythagorean theorem,

$$\begin{aligned}\|A\mathbf{x} - \mathbf{b}\|^2 &= \|A(\mathbf{x} - A^+\mathbf{b})\|^2 + \|(I - AA^+)(-\mathbf{b})\|^2 \\ &= \|A(\mathbf{x} - A^+\mathbf{b})\|^2 + \|(I - AA^+)(-\mathbf{b})\|^2 \\ &\geq \|A\mathbf{x}_0 - \mathbf{b}\|^2\end{aligned}$$

# Example

- Give a system of eq.:  $\begin{cases} x_1 + x_2 - x_3 = 1 \\ x_1 + x_2 - x_3 = 2 \end{cases} \rightarrow \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix} \quad A^+ = \frac{1}{6} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{bmatrix} \rightarrow z = A^+ b = \frac{1}{6} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

- We know the system of eq. has infinity solutions, so  $z$  is the “best-approximate” solution.