Notes on SVD

for complex vectors & complex matrices:

1. hermitian (星标) 指 conjugate transpose.

$$A^* = (\bar{A})^T = (\overline{A^T}) \in \mathbb{C}^{n\times m}$$

a. interchanges the row and columns

b. negates the imaginary part of complex #.

$$A = \begin{bmatrix} 3+4i & 1+9i \\ 5-10i & 1+7i \end{bmatrix}$$

$$\downarrow$$

$$B = A^* ;\ B = \begin{bmatrix} 3-4i & 5+10i \\ 1-9i & 1-7i \end{bmatrix}$$

2. Unitary matrix: 有着和 orthogonal 同步的特性.

$(U^* U = U U^* = I ;$

a. norm compu

b. symmetry

Typo LM p. 1629

$$U \in \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{m \times m}$$

Contents

Recap SVD.

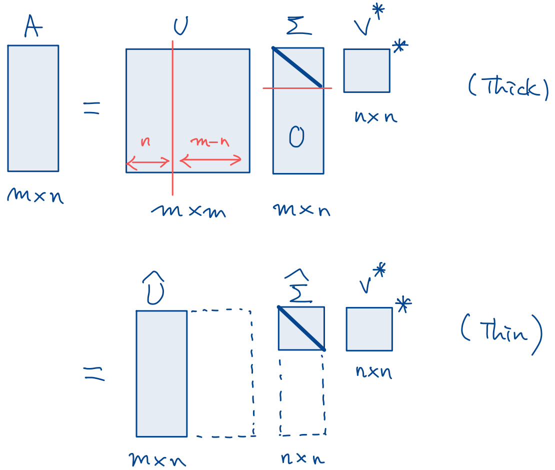
Let $A \in \mathbb{C}^{m\times n}$. Then

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

where

- $U \in \mathbb{C}^{m\times m}$ unitary
- $V \in \mathbb{C}^{n\times n}$ unitary
- $\Sigma \in \mathbb{R}^{m\times n}$ diagonal.

Illustration for $m > n$



Note: When $m \geq n$,

$$\Sigma = \left[\begin{array}{ccc} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \\ \hline & 0 & \end{array}\right], \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$$

Example

U unitary (orthogonal) $\rightarrow$ linearly indepent columns.

$$\begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2} \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Not an SVD

$$= \begin{bmatrix} -1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

an SVD

$$\sqrt{\left(-\tfrac{1}{\sqrt{2}}\right)^2 + 0^2 + \left(-\tfrac{1}{\sqrt{2}}\right)^2} = \sqrt{1} = 1$$

Properties of SVD

SVD and the 2-Norm

Note: σ_1 (largest singular value) is often called the principal singular value (of A).

Theorem 1

Let $A \in \mathbb{C}^{m \times n}$ have an SVD $A = U\Sigma V^*$. Then

1. $\|A\|_2 = \sigma_1$ *and* $\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}$.
2. *The rank of A is the number of nonzero singular values.*
3. *Let $r = \min\{m, n\}$. Then*

$$\kappa_2(A) = \|A\|_2 \|A^+\|_2 = \frac{\sigma_1}{\sigma_r}.$$

$$\Sigma = \begin{bmatrix} * & & & & \\ & * & & & \\ & & * & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}$$

$\sigma_1, \sigma_2, \sigma_3$ are nonzero

$\Rightarrow$ rank$(A) = 3$

the # of linearly independent columns of A.

Why is ① true?

$$\|A\|_p = \max_{\vec{x} \neq \vec{0}} \frac{\|A\vec{x}\|_p}{\|\vec{x}\|_p}$$

When $p=2$,

$$\|A\|_2 = \sqrt{\lambda_{max}(A^*A)}$$

largest eigenvalue

$$= \sqrt{\sigma_1^2} = \sigma_1$$

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

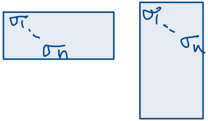
$$= (V^*)^* \underbrace{\Sigma^*}_{\Sigma^T} \underbrace{U^* U}_{I} \Sigma V^*$$

Σ^T (∵ Σ is real)

I (∵ U unitary)

$$= V \underbrace{\Sigma^T \Sigma} V^*$$

✓ if $m \geq n$



$$= \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix}$$

Connection to EVD

Let $A = U\Sigma V^* \in \mathbb{C}^{m\times n}$ and $B = A^*A$. Observe that

$B^* = (A^*A)^* = A^*(A^*)^* = A^*A = B.$

- $B \in \mathbb{C}^{n\times n}$ is a *hermitian matrix*[1], *i.e.*, $B^* = B$.
- B has an EVD:

I ∵ U is unitary

∵ V is unitary $(V^* = V^{-1})$

$$B = \underbrace{(V\Sigma^*U^*)}_{A^*}\,\underbrace{(U\Sigma V^*)}_{A} = V\Sigma^*\Sigma V^* = V\,\underbrace{(\Sigma^*\Sigma)}_{=\Sigma^T\Sigma}\,V^{-1}.$$

$\Sigma^T\Sigma$ ∵ Σ is real.

- The squares of singular values of A are eigenvalues of B.
- An EVD of $B = A^*A$ reveals the singular values and a set of right singular vectors of A.

$$\Sigma^T\Sigma = \begin{bmatrix} \sigma_1 & & 0 & \\ & \sigma_2 & & 0 \\ 0 & & \ddots & \\ & & & \sigma_n \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \ddots \\ & & & \sigma_n \\ & 0 & \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix}$$

[1] This is the $\mathbb{C}$-extension of real symmetric matrices.

(A is symmetric if $A^T = A$.)

To obtain left singular vectors of A:

Note

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

$$= (U\Sigma \underbrace{V^*)(V}_{= I} \Sigma^* U^*)$$

$$= U\underbrace{(\Sigma\Sigma^T)}_{\text{diagonal.}}U^{-1} \qquad (\text{EVD of } AA^*)$$

Connection to EVD (cont')

Theorem 2

The nonzero singular values of $A \in \mathbb{C}^{m \times n}$ are the square roots of the nonzero eigenvalues of A^*A or AA^*.

Reduction of Dimensions

Low-Rank Approximations

$\vec{u}_j \in \mathbb{C}^{m\times 1}$

$\vec{v}_j^{\,*} \in \mathbb{C}^{1\times n}$

Let $A \in \mathbb{C}^{m\times n}$ with $m \geqslant n$. Its thin SVD $A = \widehat{U}\widehat{\Sigma}V^*$ can be written as

$$A = \underbrace{\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}} \begin{bmatrix} \mathbf{v}_1^* \\ \vdots \\ \mathbf{v}_n^* \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1\mathbf{u}_1 & \cdots & \sigma_n\mathbf{u}_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^* \\ \vdots \\ \mathbf{v}_n^* \end{bmatrix} = \sum_{j=1}^{r} \sigma_j \underbrace{\mathbf{u}_j\mathbf{v}_j^*}_{\substack{\text{outer product} \\ (m\times n)}},$$

(outer product expansion)

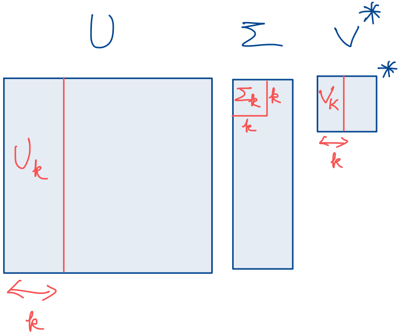
where r is the rank of A.

- Each outer product $\mathbf{u}_j\mathbf{v}_j^*$ is a rank-1 matrix.
- Since $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$, important contributions to A come from terms with small j.

Low-Rank Approximations (cont')

For $1 \leqslant k \leqslant r$, define

$$A_k = \sum_{j=1}^{k} \sigma_j \mathbf{u}_j \mathbf{v}_j^* = U_k \Sigma_k V_k^*,$$



where

- U_k is the first k columns of U;
- V_k is the first k columns of V;
- Σ_k is the upper-left $k \times k$ submatrix of Σ.

This is a rank-k approximation of A.

Best Rank-k Approximation

Theorem 3 (Eckart-Young)

Let $A \in \mathbb{C}^{m\times n}$. Suppose A has rank r and let $A = U\Sigma V^*$ be an SVD. Then

- $\|A - A_k\|_2 = \sigma_{k+1}$, *for* $k = 1, \ldots, r-1$.
- *For any matrix* B *with* $\operatorname{rank}(B) \leqslant k$, $\|A - B\|_2 \geqslant \sigma_{k+1}$. $= \|A - A_k\|_2$

first omitted singular value of A
in construction of A_k.

Appendix: Unitary Diagonalization and SVD

Unitary Diagonalization of Hermitian Matrices

The previous discussion is relevant to hermitian matrices constructed in a specific manner. For a generic hermitian matrix, we have the following result.

Theorem 4 (Spectral Decomposition)

Let $A \in \mathbb{C}^{n \times n}$ be hermitian. Then A has a unitary diagonalization

$$A = VDV^{-1},$$

where $V \in \mathbb{C}^{n \times n}$ is unitary and $D \in \mathbb{R}^{n \times n}$ is diagonal.

In words, a hermitian matrix (or symmetric matrix) has a complete set of orthonormal eigenvectors and all its eigenvalues are real.

Notes on Unitary Diagonalization and Normal Matrices

- A unitarily diagonalizable matrix $A = VDV^{-1}$ with $D \in \mathbb{C}^{n\times n}$, is called a **normal matrix**[2]. All hermitian matrices are normal.
- Let $A = VDV^{-1} \in \mathbb{C}^{n\times n}$ be normal. Since $\kappa_2(V) = 1$ (why?), Bauer-Fike implies that eigenvalues of A can be changed by no more than $\|\delta A\|_2$.

[2]Usual defintion: $A \in \mathbb{C}^{n\times n}$ is normal if $AA^* = A^*A$.

Unitary Diagonalization and SVD

Theorem 5

Let $A \in \mathbb{C}^{n\times n}$ be hermitian. Then the singular values of A are the absolute values of the eigenvalues of A.

Precisely, if $A = VDV^{-1}$ is a unitary diagonalization of A, then

$$A = \big(V\,\mathrm{sign}(D)\big)\,|D|\,V^*$$

is an SVD, where

$$\mathrm{sign}(D) = \begin{bmatrix} \mathrm{sign}(d_1) & & \\ & \ddots & \\ & & \mathrm{sign}(d_n) \end{bmatrix}, \qquad |D| = \begin{bmatrix} |d_1| & & \\ & \ddots & \\ & & |d_n| \end{bmatrix}.$$

When Do Unitary EVD and SVD Coincide?

Theorem 6

If $A = A^*$, then the following statements are equivalent:

1. *Any unitary EVD of A is also an SVD of A.*
2. *The eigenvalues of A are positive numbers.*
3. $\mathbf{x}^* A\mathbf{x} > 0$ *for all nonzero* $\mathbf{x} \in \mathbb{C}^n$. *(HPD)*

- The equivalence of 1 and 2 is immediate from Theorem 5
- The property in 3 is called the **hermitian positive definiteness**, *c.f.*, symmetric positive definiteness.
- The equivalence of 2 and 3 can be shown conveniently using **Rayleigh quotient**; see next slide.

Note: Rayleigh Quotient

Let $A \in \mathbb{R}^{n\times n}$ be fixed. The **Rayleigh quotient** is the map $R_A : \mathbb{R}^n \to \mathbb{R}$ given by

$$R_A(\mathbf{x}) = \frac{\mathbf{x}^{\mathrm{T}} A\mathbf{x}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}}.$$

- R_A maps an eigenvector of A into its associated eigenvalue, *i.e.*, if $A\mathbf{v} = \lambda\mathbf{v}$, then $R_A(\mathbf{v}) = \lambda$.
- If $A = A^{\mathrm{T}}$, then $\nabla R_A(\mathbf{v}) = \mathbf{0}$ for an eigenvector $\mathbf{v}$, and so

  $$R_A(\mathbf{v} + \epsilon\mathbf{z}) = R_A(\mathbf{v}) + 0 + O(\epsilon^2) = \lambda + O(\epsilon^2), \quad \text{as } \epsilon \to 0.$$

  The Rayleigh quotient is a quadratic approximation of an eigenvalue.