計算科学・量子計算における情報圧縮

Data Compression in Computational Science and Quantum Computing

2022.10.20

#3:特異値分解

Singular value decomposition

理学系研究科　大久保　毅

Graduate School of Science, **Tsuyoshi Okubo**

I put (URLs of) recordings of previous lectures on ITC-LMS.
You can also download lecture slide from ITC-LMS.

Today’s topic

1. Computational science, quantum computing, and data compression
2. Review of linear algebra
3. Singular value decomposition
4. Application of SVD and generalization to tensors
5. Entanglement of information and matrix product states
6. Application of MPS to eigenvalue problems
7. Tensor network representation
8. Data compression in tensor network
9. Tensor network renormalization
10. Quantum mechanics and quantum computation
11. Simulation of quantum computers
12. Quantum-classical hybrid algorithms and tensor network
13. Quantum error correction and tensor network

Outline

- Singular value decomposition (SVD)
  - Definition and properties
  - (Relation to quantum physics)
- Generalized inverse matrix
  - MoorPseudo inverse
  - Application to simultaneous linear equations
- Low-rank approximation by SVD
  - It is “optimal.”
  - (Relation to PCA)

Singular value decomposition

Diagonalization

Diagonalizaiton（対角化）：

$$A : N \times N$$

(Square matrix)

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{pmatrix}$$

$$A\vec{v} = \lambda\vec{v} \qquad P = (\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_N)$$

$$(\vec{u}^*)^t A = \lambda(\vec{u}^*)^t \qquad (P^{-1})^t = (\vec{u}_1^*, \vec{u}_2^*, \cdots, \vec{u}_N^*)$$

- Eigenvalue problems and diagonalizations are defined for a square matrix.
- Even if *A* is a square matrix, it may not be diagonalized.
  - Normal or Hermitian matrices are always diagonalized by a unitary matrix

(Normal matrix)

Normal matrix（正規行列）： $A^\dagger A = AA^\dagger$

We can always diagonalize it by a unitary matrix

$$U^\dagger = U^{-1}$$

as

$$U^\dagger A U = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{pmatrix} \qquad \lambda_i \in \mathbb{C}$$

Its eigenvalues could be complex.
(even if *A* is a real matrix)

Spectral decomposition

(For a normal matrix A,)

Spectral decomposition
(スペクトル分解)

$$A = U \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{pmatrix} U^\dagger$$

$$= \sum_{i=1}^{N} \lambda_i \underline{\vec{u}_i \vec{u}_i^\dagger}$$

$$\left(= \sum_{i=1}^{N} \lambda_i |u_i\rangle\langle u_i| \right)$$

Note:

$$\vec{u}_i \vec{u}_i^\dagger = \begin{pmatrix} u_1 u_1^* & u_1 u_2^* & \cdots & u_1 u_N^* \\ u_2 u_1^* & u_2 u_2^* & \cdots & u_2 u_N^* \\ \vdots & \vdots & \cdots & \vdots \\ u_N u_1^* & u_N u_2^* & \cdots & u_N u_N^* \end{pmatrix}$$

Projector:

$$P^2 = P$$

Matrix decomposition into a sum of **projectors** onto its eigensubspaces.

Generalization of spectral decomposition?

- Spectral decomposition is defined for a normal matrix.
  - Even if A is a square matrix, it may not be diagonalized (by unitary matrix)

- Is it possible to transform all square matrixes into diagonal forms by generalizing the spectral decomposition?
- Is it possible to generalize it to rectangular matrices?

Yes! One solution is **the singular value decomposition.**

Singular value decomposition (SVD)

Singular value decomposition (特異値分解)

$$A : M \times N$$

$$A_{ij} \in \mathbb{C}$$

$$A = \underline{U}\,\Sigma\,\underline{V}^\dagger$$

$U : M \times M$ **Unitary**

$V : N \times N$ **Unitary**

$$\Sigma = \begin{pmatrix} \Sigma_{r\times r} & 0_{r\times(N-r)} \\ 0_{(M-r)\times r} & 0_{(M-r)\times N-r} \end{pmatrix}$$

$$\Sigma_{r\times r} = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{pmatrix}$$

Diagonal matrix with non-negative real elements

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

Singular values

Properties of SVD 1

1. Any matrices can be decomposed as SVD: $A = U\Sigma V^\dagger$

$A : M \times N$ ➡ $A^\dagger A : N \times N$

* $A^\dagger A$ is a Hermitian matrix.

$$(A^\dagger A)^\dagger = A^\dagger A$$ ➡

It can be diagonalized by
a unitary matrix V .

$$V^\dagger (A^\dagger A) V = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$$

$$V = (\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_N)$$

$\vec{v}_i$: eigenvector

* $A^\dagger A$ is a positive semi-definite matrix.
(半正定値、準正定値)

$$\vec{x}^* \cdot (A^\dagger A \vec{x}) = \|A\vec{x}\|^2 \geq 0$$ ⬌

Its eigenvalues are
non-negative

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0$$

Properties of SVD 1

1. Any matrices can be decomposed as SVD: $A = U\Sigma V^\dagger$

$$V^\dagger (A^\dagger A) V = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$$
$$V = (\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_N)$$

$\Longleftrightarrow$

$$(A\vec{v}_i)^* \cdot (A\vec{v}_j) = \lambda_i \delta_{ij}$$
$$(\|A\vec{v}_i\|^2 = \lambda_i)$$

Suppose $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_N$

(There are r positive eigenvalues.)

Make new orthonormal basis $U = (\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_M)$ in $\mathbb{C}^M$

For $(i = 1, 2, \ldots, r)$ $\quad \sigma_i = \sqrt{\lambda_i}, \vec{u}_i = \dfrac{1}{\sigma_i} A\vec{v}$

For $(i = r+1, \ldots, M)$ Any orthonormal basis orthogonal to $\vec{u}_i \quad (i = 1, 2, \ldots, r)$

$$\vec{u}_i^* \cdot (A\vec{v}_j) = \sigma_i \delta_{ij} \quad (i = 1, \ldots, M; j = 1, \ldots, N)$$

(For simplicity, we set $\sigma_i = 0$ for $i > r$.)

Properties of SVD 1

1. Any matrices can be decomposed as SVD: $A = U\Sigma V^\dagger$

We can perform same "proof" by using $AA^\dagger$.

$U = (\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_M)$ is the unitary matrix which diagonalize $AA^\dagger$ as

$$U^\dagger (AA^\dagger) U = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_r, \underbrace{0, \ldots, 0}_{M-r}\}$$

In summary,

- A matrix A can be decomposed as SVD: $A = U\Sigma V^\dagger$
- Singular values are related to the eigenvalues of $A^\dagger A$ and $AA^\dagger$ as
$$\sigma_i = \sqrt{\lambda_i}\ .$$
- V and U are eigenvectors of $A^\dagger A$ and $AA^\dagger$,respectively.

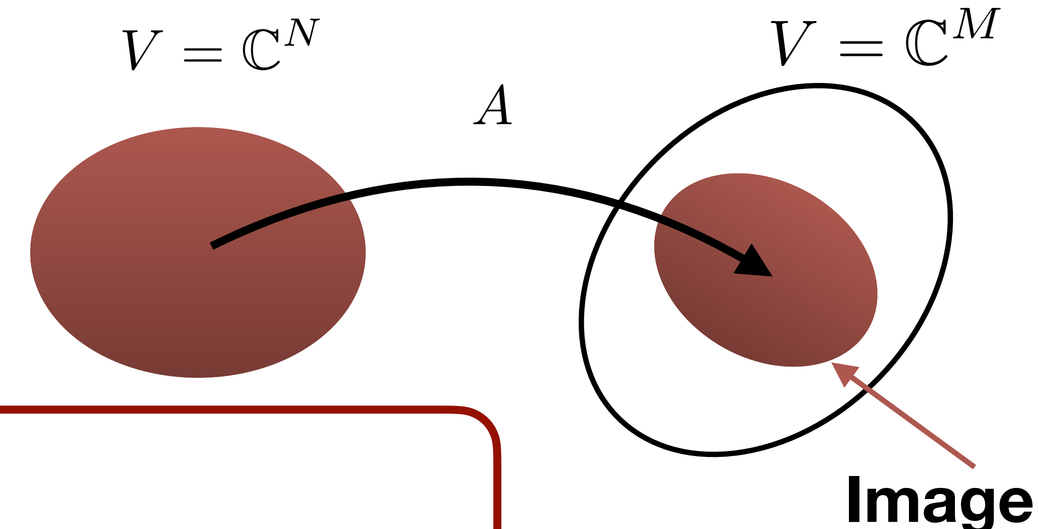
Properties of SVD 2

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

2. # of positive singular values is identical with the rank.

$$A : M \times N \Rightarrow A : \mathbb{C}^N \to \mathbb{C}^M$$

$$\mathrm{rank}(A) \equiv \dim(\mathrm{img}(A))$$



Remember

The orthonormal basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_N\}$ satisfies

$$(A\vec{v}_i)^* \cdot (A\vec{v}_j) = \lambda_i \delta_{ij}$$

Here, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_N$ and $\sigma_i = \sqrt{\lambda_i}$

$$\forall \vec{x} \in \mathbb{C}^N, \vec{x} = \sum_{i=1}^{N} C_i \vec{v}_i \Rightarrow A\vec{x} = \sum_{i=1}^{N} C_i (A\vec{v}_i) = \sum_{i=1}^{r} C_i (A\vec{v}_i)$$

$$\Rightarrow \mathrm{img}(A) = \mathrm{Span}\{A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_r\}$$

$\Rightarrow \dim(\mathrm{img}(A)) = r$ **= # of positive singular values**

Properties of SVD 3 (optional) $A = U\Sigma V^\dagger$

3. Singular vectors

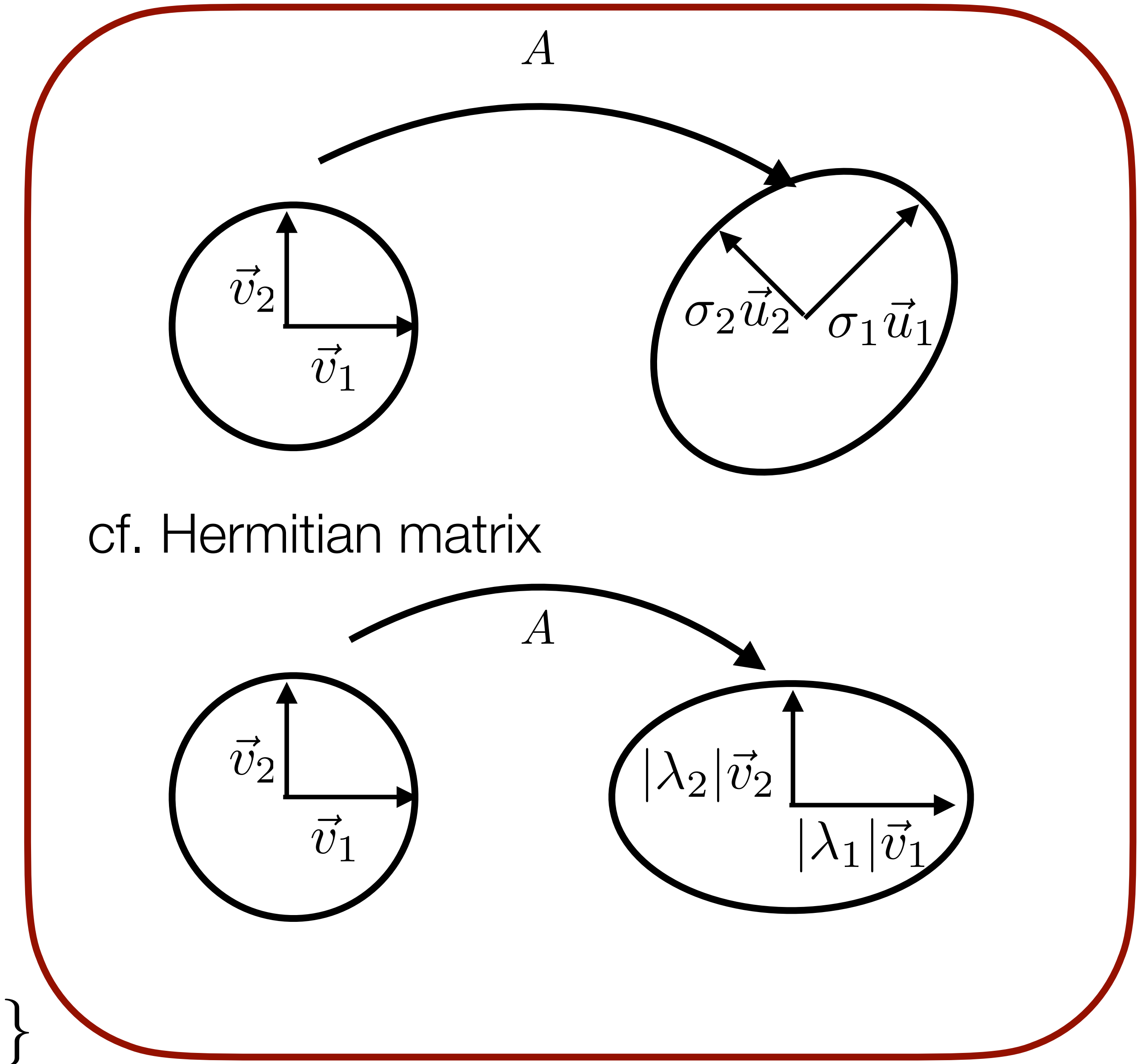
$A : M \times N$ $\quad U = (\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_M)$, $V = (\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_N)$

For $i = 1, 2, \ldots, r$

$$A\vec{v}_i = \sigma_i \vec{u}_i \ , \ A^\dagger \vec{u}_i = \sigma_i \vec{v}_i$$

$\vec{v}_i$: right singular vector

$\vec{u}_i$: left singular vector



Relation to image and kernel:

$$\mathrm{img}(A) = \mathrm{Span}\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r\}$$
$$\ker(A) = \mathrm{Span}\{\vec{v}_{r+1}, \vec{v}_{r+2}, \ldots, \vec{v}_N\}$$

$$\mathrm{img}(A^\dagger) = \mathrm{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r\}$$
$$\ker(A^\dagger) = \mathrm{Span}\{\vec{u}_{r+1}, \vec{u}_{r+2}, \ldots, \vec{u}_M\}$$

Properties of SVD 4 (optional) $A = U\Sigma V^\dagger$

4. Min-max theorem (Courant-Fischer theorem)

$A : N \times N$, Hermitian matrix

Suppose its eigenvalues are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$.

$$\lambda_k = \min_{\boldsymbol{S};\dim(\boldsymbol{S})\leq k-1} \max_{\vec{x}\in \boldsymbol{S}^\perp;\|\vec{x}\|=1} \vec{x}^* \cdot A\vec{x}$$

$$\boldsymbol{S}^\perp = \{\vec{x} : \vec{x}^* \cdot \vec{y} = 0, \vec{y} \in \boldsymbol{S}\}$$

Orthonormal complement（直交補空間）

We can prove this by considering vector subspace spanned by eigenvectors. (see references)

Intuitive examples:

Maximum appears for the eigenvector.

$$A\vec{u}_i = \lambda_i \vec{u}_i$$

$$\lambda_1 = \max_{\vec{x}\in \boldsymbol{C}^N;\|\vec{x}\|=1} \vec{x}^* \cdot (A\vec{x}) \quad \Rightarrow \quad \vec{x} = \vec{u}_1$$

$$\lambda_2 = \max_{\vec{x}\in \boldsymbol{C}^N;\vec{x}\perp\vec{u}_1,\|\vec{x}\|=1} \vec{x}^* \cdot (A\vec{x}) = \min_{\boldsymbol{S};\dim(\boldsymbol{S})\leq 1} \max_{\vec{x}\in \boldsymbol{S}^\perp;\|\vec{x}\|=1} \vec{x}^* \cdot (A\vec{x}) \quad \Rightarrow \quad \vec{x} = \vec{u}_2$$

Properties of SVD 4 (optional) $A = U\Sigma V^\dagger$

4. Min-max theorem (Courant-Fischer theorem)

$A : M \times N$

Suppose its singular values are $\sigma_1 \geq \sigma_2 \geq \cdots$

$$\sigma_k = \min_{\boldsymbol{S};\dim(\boldsymbol{S})\leq k-1} \max_{\vec{x}\in\boldsymbol{S}^\perp;\|\vec{x}\|=1} \|A\vec{x}\|$$

By setting k=1,

$$\sigma_1 = \max_{\vec{x}\in\boldsymbol{C}^N,\|\vec{x}\|=1} \|A\vec{x}\|$$

which means

$$\|A\vec{x}\| \leq \sigma_1 \|\vec{x}\|$$

for $\vec{x} \in \boldsymbol{C}^N$

We can easily prove this by using

$A^\dagger A$: Hermitian

$$A^\dagger A \vec{v}_i = \lambda_i$$

$$\sigma_i = \sqrt{\lambda_i}$$

Properties of SVD 5

$A = U \Sigma V^\dagger$

5. Singular values for multiplication and addition

$\sigma_i(A)$: singular value of matrix A

(for $i > \mathrm{rank}(A)$, we set $\sigma_i = 0$)

*Following properties can be proven by using min-max theorem.

Multiplication: $A : M \times L, B : L \times N$

$$\sigma_k(AB) \leq \sigma_1(A)\sigma_k(B) \qquad (k = 1, 2, \dots)$$

$$(\sigma_k(AB) \leq \sigma_k(A)\sigma_1(B))$$

$$\Rightarrow \quad \mathrm{rank}(AB) \leq \min(\mathrm{rank}(A), \mathrm{rank}(B))$$

We will use this in tensor networks.

Addition: $A, B : M \times N$

(optional)

$$\sigma_{k+j-1}(A+B) \leq \sigma_k(A) + \sigma_j(B) \qquad (k, j = 1, 2, \dots)$$

$$(\sigma_{k+j-1}(A+B) \leq \sigma_j(A) + \sigma_k(B))$$

$\Rightarrow$ If $\mathrm{rank}(B) \leq r$,

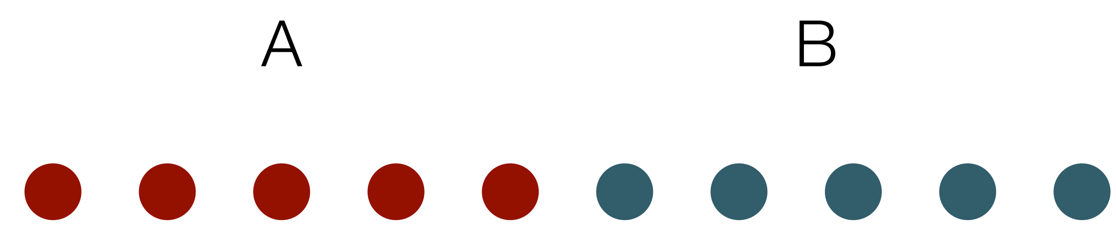
$$\sigma_{k+r}(A+B) \leq \sigma_k(A)$$

Relation to quantum physics

Quantum state $|\Psi\rangle = \sum_{\{i_1, i_2, \dots i_N\}} \Psi_{i_1 i_2 \dots i_N} |i_1 i_2 \dots i_N\rangle$

Schmidt decomposition

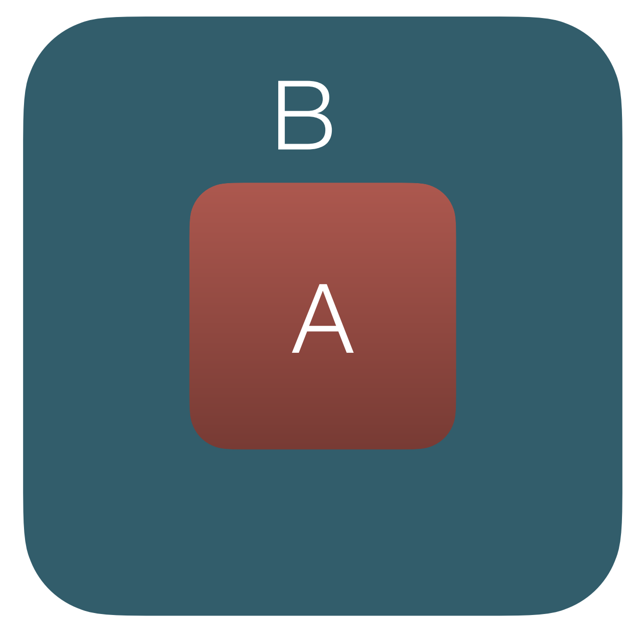
Divide a system into two parts, A and B:



General wave function can be represented by a superposition of orthonormal basis set.

$$|\Psi\rangle = \sum_{i,j} M_{i,j} |A_i\rangle \otimes |B_j\rangle = \sum_i \lambda_i |\alpha_i\rangle \otimes |\beta_i\rangle$$

$$\boxed{M_{i,j} \equiv \Psi_{\underbrace{(i_1, \dots)}_{\text{A}}, \underbrace{(\dots, i_N)}_{\text{B}}} \quad \begin{array}{l} |A_i\rangle = |i_1, i_2, \dots\rangle \\ |B_j\rangle = |\dots, i_{N-1}, i_N\rangle \end{array}}$$



Orthonormal basis: $\langle A_i | A_j \rangle = \langle B_i | B_j \rangle = \delta_{i,j}$,
$\langle \alpha_i | \alpha_j \rangle = \langle \beta_i | \beta_j \rangle = \delta_{i,j}$

Schmidt coefficient: $\lambda_i \geq 0$

Relation to quantum physics

$$|\Psi\rangle = \sum_{i,j} M_{i,j} |A_i\rangle \otimes |B_j\rangle$$

Singular values: $\lambda_m \geq 0$

SVD

$$M_{i,j} = \sum_m U_{i,m} \lambda_m V^\dagger_{m,j}$$

Singular vectors:
$$\sum_i U^\dagger_{m,i} U_{i,m'} = \delta_{m,m'}$$
$$\sum_j V^\dagger_{m,j} V_{i,m'} = \delta_{m,m'}$$

Relation to the Schmidt decomposition:

$$|\Psi\rangle = \sum_{i,j} M_{i,j} |A_i\rangle \otimes |B_j\rangle = \sum_m \lambda_m |\alpha_m\rangle \otimes |\beta_m\rangle$$

$$|\alpha_m\rangle = \sum_i U_{i,m} |A_i\rangle$$
$$|\beta_m\rangle = \sum_j V^\dagger_{m,j} |B_j\rangle$$

$$\Rightarrow \quad \langle \alpha_m | \alpha_{m'} \rangle = \langle \beta_m | \beta_{m'} \rangle = \delta_{m,m'}$$

SVD of the quantum state is directly related to the Schmidt decomposition.

We will revisit this topic on #5!

Generalized inverse matrix

Regular matrix and its inverse matrix

A square matrix A is a **regular matrix** （正則）if a matrix X satisfying

$$AX = XA = I$$

exists. The matrix X is called inverse matrix（逆行列） of A and it is written as $X = A^{-1}$.

Properties:

A^{-1} is unique.

$$(A^{-1})^{-1} = A$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

A is a regular matrix $\Longleftrightarrow$ $\mathrm{rank}(A) = N$

Can we consider an "inverse matrix" of a non-regular matrix (including a rectangular matrix) ?

Generalized inverse matrix

Generalized inverse matrix（一般化逆行列）：

For $A : M \times N$, a matrix $A^{-} : N \times M$ satisfying

$$AA^{-}A = A$$

is called generalized inverse matrix.

Properties:

- The generalized inverse matrix is not unique.
  - At least one generalized matrix exists for a given matrix.

- If A is a regular matrix, $A^{-} = A^{-1}$

→ A^{-} is a generalization of the inverse matrix.

Moore-Penrose pseudo inverse

Moore-Penrose pseudo inverse matrix（擬似逆行列）：

For $A : M \times N$, a matrix $A^+ : N \times M$ satisfying

(1) $AA^+A = A$ (2) $A^+AA^+ = A^+$

(3) $(AA^+)^\dagger = AA^+$ (4) $(A^+A)^\dagger = A^+A$

is called (Moore-Penrose) pseudo inverse matrix.

Relation to SVD

- Pseudo inverse is unique and calculated from SVD.

$$A = U\Sigma V^\dagger = U \begin{pmatrix} \Sigma_{r\times r} & 0_{r\times(N-r)} \\ 0_{(M-r)\times r} & 0_{(M-r)\times N-r} \end{pmatrix} V^\dagger$$

$$\Rightarrow \quad A^+ = V \begin{pmatrix} \Sigma^{-1}_{r\times r} & 0_{r\times(M-r)} \\ 0_{(N-r)\times r} & 0_{(N-r)\times M-r} \end{pmatrix} U^\dagger$$

$$A^+A = V \begin{pmatrix} \Sigma^{-1}_{r\times r} & 0_{r\times(M-r)} \\ 0_{(N-r)\times r} & 0_{(N-r)\times M-r} \end{pmatrix} U^\dagger U \begin{pmatrix} \Sigma_{r\times r} & 0_{r\times(N-r)} \\ 0_{(M-r)\times r} & 0_{(M-r)\times N-r} \end{pmatrix} V^\dagger$$

$$= \sum_{i=1}^{r} \vec{v}_i \vec{v}_i^\dagger \left(= \sum_{i=1}^{r} |v_i\rangle\langle v_i|\right)$$

A^+A is a projector onto img($A^\dagger$).
(AA^+ is a projector onto img(A).)

Simultaneous linear equation

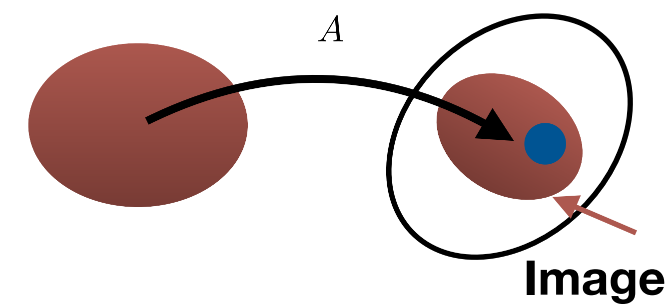
Simultaneous linear equation（連立一次方程式）

$$A\vec{x} = \vec{b} \qquad A : M \times N, \vec{x} \in \mathbb{C}^N, \vec{b} \in \mathbb{C}^M$$

$\vec{b}$: ●

Two situations:

(1) There are solutions. $\Longleftrightarrow$ $\vec{b} \in \mathrm{img}(A)$



(i) There is the unique solution.

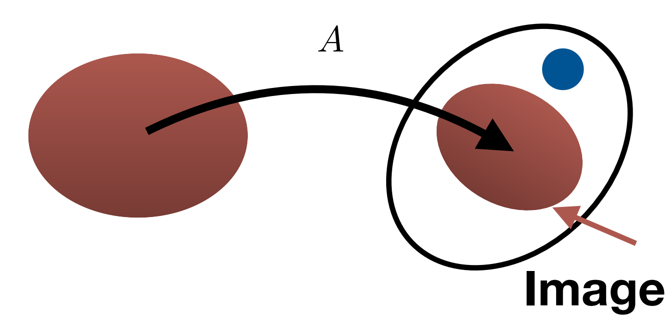
$$\mathrm{rank}(A) = N$$

(ii) There are infinite solutions (underdetermined).

$$\mathrm{rank}(A) < N$$

(We can add any vector $A\vec{y} = \vec{0}$.)

(2) There is no solution. $\Longleftrightarrow$ $\vec{b} \notin \mathrm{img}(A)$
(overdetermined)



Pseudo inverse and simultaneous linear equation

Simultaneous linear equation $A\vec{x} = \vec{b}$ $\quad A : M \times N, \vec{x} \in \mathbb{C}^N, \vec{b} \in \mathbb{C}^M$

(1) There are solutions. $\Leftrightarrow$ $\vec{b} \in \mathrm{img}(A)$

- A vector defined by the pseudo inverse as

$$\vec{x}' \equiv A^+\vec{b}$$

is one of the solutions.

Because $\vec{b} \in \mathrm{img}(A)$,there exists $\vec{v} : A\vec{v} = \vec{b}$.

$$\Rightarrow \quad A\vec{x}' = AA^+\vec{b} = AA^+A\vec{v} = A\vec{v} = \vec{b}$$

- $\vec{x}'$ has the smallest norm $||\vec{x}'||$ among the solutions.

$$||\vec{x}|| \geq ||A^+A\vec{x}|| = ||A^+\vec{b}|| = ||\vec{x}'||$$

$\because A^+A$ is a projector.

The pseudo inverse gives us the smallest norm solution.

Pseudo inverse and simultaneous linear equation

Simultaneous linear equation $A\vec{x} = \vec{b}$ $\quad A : M \times N, \vec{x} \in \mathbb{C}^N, \vec{b} \in \mathbb{C}^M$

(2) There is no solution. $\Longleftrightarrow$ $\vec{b} \notin \mathrm{img}(A)$

- A vector defined by the pseudo inverse as

$$\vec{x}' \equiv A^+\vec{b}$$

minimizes the "distance" $\|\vec{b} - A\vec{x}\|$.

$$\vec{y} = A\vec{c} \in \mathrm{img}(A), \vec{c} \in \mathbb{C}^N$$

$$\|\vec{y} - \vec{b}\|^2 = \|\underbrace{\vec{y} - AA^+\vec{b}}_{\in\, \mathrm{img}(A)} - \underbrace{(I - AA^+)\vec{b}}_{\in\, \mathrm{img}(A)^\perp}\|^2$$

$$= \|\vec{y} - AA^+\vec{b}\|^2 + \|\vec{b} - AA^+\vec{b}\|^2$$

$$\geq \|\vec{b} - AA^+\vec{b}\|^2 = \|\vec{b} - A\vec{x}'\|^2$$

The pseudo inverse gives us approximate "least square solution".

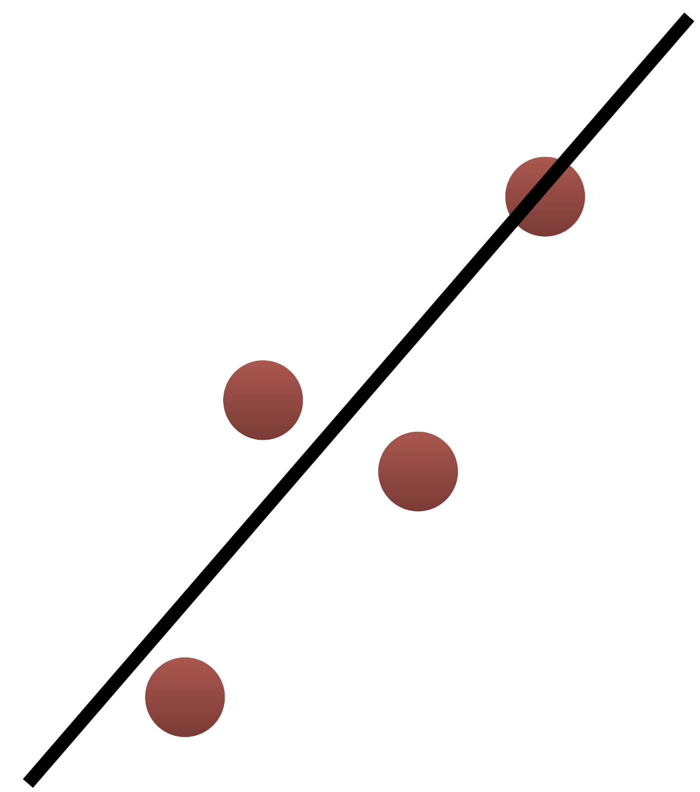
Example of Least square solution problem

Fitting of a line to data points

$$y = ax + b$$

Data poins:

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$$



$$ax_i + b = y_i$$

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$

$$A\vec{x} = \vec{b}$$

Solution by the pseudo inverse

Least square fitting（最小二乗法）

$$\begin{pmatrix} a \\ b \end{pmatrix} = A^+ \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$

Low-rank approximation

Amount of data in SVD representation

$$A : M \times N$$

$$A = U\Sigma V^\dagger = U \begin{pmatrix} \Sigma_{r\times r} & 0_{r\times(N-r)} \\ 0_{(M-r)\times r} & 0_{(M-r)\times N-r} \end{pmatrix} V^\dagger$$

neglect zero singular values →

$$= \bar{U}\Sigma_{r\times r}\bar{V}^\dagger$$

$$\bar{U} : M \times r, \bar{V}^\dagger : r \times N$$

$$U = (\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_M)$$
$$V = (\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_N)$$

↓

$$\bar{U} = (\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r)$$
$$\bar{V} = (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r)$$

If rank(A) is much smaller than M and N,

$$r \ll M, N$$

we can reduce the data to represent A.

(At this stage, no data loss)

Low-rank approximation

Low-rank approximation（低ランク近似）

Find an approximate matrix

$$A \simeq \tilde{A}$$

with lower rank:

$$\mathrm{rank}(A) > \mathrm{rank}(\tilde{A})$$

⇒ Through the low-rank approximation, we can reduce the amount of data.

An example of data compressions.

Low-rank approximation by SVD

Consider a matrix obtained by neglecting smaller singular values

$$A = \bar{U}\Sigma_{r\times r}\bar{V}^{\dagger} \quad \Rightarrow \quad \tilde{A} = \tilde{U}\Sigma_{k\times k}\tilde{V}^{\dagger} \qquad (k < r)$$

$$\boxed{\begin{aligned} \Sigma_{r\times r} &= \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r) \\ \bar{U} &= (\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r) \\ \bar{V} &= (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r) \end{aligned}}$$

$$\boxed{\begin{aligned} \Sigma_{k\times k} &= \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k) \\ \tilde{U} &= (\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k) \\ \tilde{V} &= (\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k) \end{aligned}}$$

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

$$\mathrm{rank}(A) = r$$

Keep the largest k singular values
(and corresponding singular vectors).

$$\mathrm{rank}(\tilde{A}) = k < r$$

This approximation is one of the low rank approximation.

* For this approximation, we need $\mathrm{O}(MNk)$ calculations for SVD of a $M\times N$ matrix.

Norm of matrices $\|A\|$

There are two popular norms:

(1) **Frobenius norm** （フロベニウス　ノルム）

$$\|A\|_F = \sqrt{\sum_{i,j} |A_{ij}|^2} = \sqrt{\mathrm{Tr}(A^\dagger A)}$$

*Trace（対角和）

$$\mathrm{Tr}(X) = \sum_i X_{ii}$$

(2) **Operator norm**（作用素ノルム）

$$\begin{aligned}\|A\|_O &= \inf\{c \geq 0; \|A\vec{x}\| \leq c\|\vec{x}\|\} \\ &= \sigma_1(A)\end{aligned}$$

*inf =infimum（下限）

*We define the norm for a vector as

$$\|\vec{x}\| = \sqrt{\sum_i |x_i|^2}$$

By using these norms, we define the distance between matrices:

$$\|A - \tilde{A}\|$$

Accuracy of low rank approximation by SVD

(A part of) Eckart-Young-Mirsky Theorem

C. Eckart, and G. Young, Psychometrika **1**, 211 (1936).
L. Mirsky, Q. J. Math. **11**, 50 (1960).

For $A : M \times N$

$$\min\{\|A - B\|_F : \mathrm{rank}(B) = k\} = \sqrt{\sum_{i=k+1}^{\min(N,M)} \sigma_i^2}$$

$$\min\{\|A - B\|_O : \mathrm{rank}(B) = k\} = \sigma_{k+1}$$

Because the k-rank approximation by SVD gives

$$\|A - \tilde{A}\|_F = \sqrt{\sum_{i=k+1}^{\min(N,M)} \sigma_i^2}, \qquad \|A - \tilde{A}\|_O = \sigma_{k+1}$$

it is an "optimal" approximation with rank k.

Short proof of the theorem: Frobenius norm (optional)

*This proof is based on
"システム制御のための数学（１）"太田快人 著

From the inequality of singular values for matrix addition (property 5),

for j=1,... ,min(M,N) $\quad (\mathrm{rank}(B) = k)$

$$\sigma_{j+k}(A) = \sigma_{j+k}((A-B)+B) \leq \sigma_j(A-B)$$

Property 5

By taking a square and summing up them

$$\sum_{i=k+1}^{\min(M,N)} \sigma_i^2(A) \leq \sum_{j=1}^{\min(M,N)} \sigma_j^2(A-B) = \|A-B\|_F^2$$

*Note $\sigma_j(A) = 0 \quad (j > \mathrm{rank}(A))$

Short proof of the theorem: operator norm (optional)

*This proof is based on
"システム制御のための数学（１）"太田快人 著

From the min-max theorem of singular values (property 4),

$$(\mathrm{rank}(B) = k)$$

$$\sigma_{k+1}(A) \leq \max_{\vec{x} \in \ker(B), \|\vec{x}\|=1} \|A\vec{x}\| = \max_{\vec{x} \in \ker(B), \|\vec{x}\|=1} \|(A-B)\vec{x}\|$$

Property4 with

$$\boldsymbol{S} = \mathrm{img}(B^\dagger)$$
$$\boldsymbol{S}^\perp = \ker(B)$$

$$B\vec{x} = 0 \quad (\vec{x} \in \ker(B))$$

$$\leq \max_{\|\vec{x}\|=1} \|(A-B)\vec{x}\| = \|A-B\|_O$$

Expand the vector space

Definition of the operator norm

Relation to **principal component analysis** （主成分分析）

Data set $\{X_{ij}\}$: X: $N \times M$ matrix

i = index for data, j = data type (coordinates, momentum, ...)

* Suppose the mean of data is 0: $\sum_i X_{ij} = 0$

Covariance matrix（共分散行列）: $C = X^T X$

Principal component analysis (PCA):

Data compression through the spectrum decomposition of C.

$C = V \Lambda V^T$ Λ: diagonal matrix, $\Lambda_{ii} = \lambda_i \geq 0$

$V = (\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_N)$

$\vec{v}_i$ corresponding to large λ_i contains important information.

By construction, λ and V are related to SVD of X!

$X = U \Sigma V^T$, $\sigma_i = \sqrt{\lambda_i}$

PCA can be regarded as the low-rank approximation of X.

Next week (Oct. 27)

Notice

- No classes on Nov. 3, Nov. 17, and Nov. 22
- Classes will be also held on Jan. 5 and Jan. 19

---

1. Computational science, quantum computing, and data compression
2. Review of linear algebra
3. Singular value decomposition
4. Application of SVD and generalization to tensors
5. Entanglement of information and matrix product states
6. Application of MPS to eigenvalue problems
7. Tensor network representation
8. Data compression in tensor network
9. Tensor network renormalization
10. Quantum mechanics and quantum computation
11. Simulation of quantum computers
12. Quantum-classical hybrid algorithms and tensor network
13. Quantum error correction and tensor network