SI231b: Matrix Computations

Lecture 19: Singular Value Decomposition

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology

ShanghaiTech University

Nov. 18, 2021

Main Results

Any matrix $\mathrm{A} \in \mathbb{R}^{m \times n}$ admits a singular value decomposition (SVD)

$$\mathrm{A} = \mathrm{U}\boldsymbol{\Sigma}\mathrm{V}^T,$$

where $\mathrm{U} \in \mathbb{R}^{m \times m}$ and $\mathrm{V} \in \mathbb{R}^{n \times n}$ are orthogonal, and $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ has $[\boldsymbol{\Sigma}]_{ij} = 0$ for all $i \neq j$ and $[\boldsymbol{\Sigma}]_{ii} = \sigma_i$ for all i, with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min\{m,n\}} \geq 0$.

- matrix 2-norm: $\|\mathrm{A}\|_2 = \sigma_1$
- let r be the number of nonzero σ_i's, partition $\mathrm{U} = [\ \mathrm{U}_1\ \mathrm{U}_2\ ]$, $\mathrm{V} = [\ \mathrm{V}_1\ \mathrm{V}_2\ ]$ with $\mathrm{U}_1 \in \mathbb{R}^{m \times r}$ and $\mathrm{V}_1 \in \mathbb{R}^{n \times r}$, and let $\tilde{\boldsymbol{\Sigma}} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$
  - rank(A) $= r$
  - pseudo-inverse: $\mathrm{A}^\dagger = \mathrm{V}_1 \tilde{\boldsymbol{\Sigma}}^{-1} \mathrm{U}_1^T$
  - LS solution: $\mathrm{x}_{\mathrm{LS}} = \mathrm{A}^\dagger \mathrm{y} + \boldsymbol{\eta}$ for any $\boldsymbol{\eta} \in \mathcal{R}(\mathrm{V}_2)$
  - orthogonal projection: $\mathrm{P}_\mathrm{A} = \mathrm{U}_1 \mathrm{U}_1^T$

Low-rank Approximation

Given $A \in \mathbb{R}^{m \times n}$ and $k \in \{1, \dots, \min\{m, n\}\}$, the problem

$$\min_{B \in \mathbb{R}^{m \times n},\ \ \mathrm{rank}(B) \leq k} \|A - B\|_2^2$$

has an optimal solution given by $B^\star = \sum_{i=1}^{k} \sigma_i u_i v_i^T$. Or equivalently, $B^\star$ gives the best rank k approximation of A while using the matrix 2-norm to optimize $\|A - B\|^2$.

Singular Value Decomposition

Theorem. Given any $\mathsf{A} \in \mathbb{R}^{m \times n}$, there exists a 3-tuple $(\mathsf{U}, \boldsymbol{\Sigma}, \mathsf{V})$ with $\mathsf{U} \in \mathbb{R}^{m \times m}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ and $\mathsf{V} \in \mathbb{R}^{n \times n}$ such that

$$\mathsf{A} = \mathsf{U} \boldsymbol{\Sigma} \mathsf{V}^T,$$

U and V are orthogonal, and $\boldsymbol{\Sigma}$ takes the form

$$[\boldsymbol{\Sigma}]_{ij} = \begin{cases} \sigma_i, & i = j \\ 0, & i \neq j \end{cases}, \qquad \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0, \; p = \min\{m, n\}.$$

- the above decomposition is called the singular value decomposition (SVD)
- σ_i is called the ith singular value
- u_i and v_i are called the ith left and right singular vectors, resp.
- the following notations may be used to denote singular values of a given A

$$\sigma_{\max}(\mathsf{A}) = \sigma_1(\mathsf{A}) \geq \sigma_2(\mathsf{A}) \geq \ldots \geq \sigma_p(\mathsf{A}) = \sigma_{\min}(\mathsf{A})$$

- partitioned form: let r be the number of nonzero singular values, and note $\sigma_1 \geq \ldots \sigma_r > 0$, $\sigma_{r+1} = \ldots = \sigma_p = 0$. Then,

$$\mathrm{A} = \begin{bmatrix} \mathrm{U}_1 & \mathrm{U}_2 \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathrm{V}_1^T \\ \mathrm{V}_2^T \end{bmatrix},$$

  where

  - $\tilde{\boldsymbol{\Sigma}} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$
  - $\mathrm{U}_1 = [\ \mathrm{u}_1, \ldots, \mathrm{u}_r\ ] \in \mathbb{R}^{m \times r}$, $\mathrm{U}_2 = [\ \mathrm{u}_{r+1}, \ldots, \mathrm{u}_m\ ] \in \mathbb{R}^{m \times (m-r)}$
  - $\mathrm{V}_1 = [\ \mathrm{v}_1, \ldots, \mathrm{v}_r\ ] \in \mathbb{R}^{n \times r}$, $\mathrm{V}_2 = [\ \mathrm{v}_{r+1}, \ldots, \mathrm{v}_n\ ] \in \mathbb{R}^{n \times (n-r)}$

- economic SVD: $\mathrm{A} = \mathrm{U}_1 \tilde{\boldsymbol{\Sigma}} \mathrm{V}_1^T$
- outer-product form: $\mathrm{A} = \sum_{i=1}^{r} \sigma_i \mathrm{u}_i \mathrm{v}_i^T$

SVD and Eigenvalue Decomposition

From the SVD $A = U\boldsymbol{\Sigma}V^T$, we see that

$$AA^T = UD_1U^T, \qquad D_1 = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2, \underbrace{0, \ldots, 0}_{m-p \text{ zeros}}) \qquad (*)$$

$$A^TA = VD_2V^T, \qquad D_2 = \boldsymbol{\Sigma}^T\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2, \underbrace{0, \ldots, 0}_{n-p \text{ zeros}}) \qquad (**)$$

Observations:

- $(*)$ and $(**)$ are the eigenvalue decompositions of AA^T and A^TA, resp.
- the left singular vector matrix U of A is the eigenvector matrix of AA^T
- the right singular vector matrix V of A is the eigenvector matrix of A^TA
- the squares of nonzero singular values of A, $\sigma_1^2, \ldots, \sigma_r^2$, are the nonzero eigenvalues of both AA^T and A^TA.

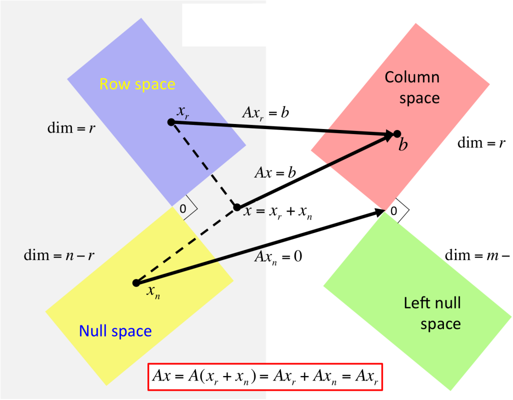

Figure 1: Four fundamental subspaces

In lecture 3, we have learnt that for $\mathrm{A} \in \mathbb{R}^{m\times n}$

- $\mathcal{R}(\mathrm{A}) \perp \mathcal{N}(\mathrm{A}^T)$, and $\mathcal{R}(\mathrm{A}) \oplus \mathcal{N}(\mathrm{A}^T) = \mathbb{R}^m$
- $\mathcal{R}(\mathrm{A}^T) \perp \mathcal{N}(\mathrm{A})$, and $\mathcal{R}(\mathrm{A}^T) \oplus \mathcal{N}(\mathrm{A}) = \mathbb{R}^n$

SVD and Four Fundamental Subspaces

Property: The following properties hold:

(a) $\mathcal{R}(\mathrm{A}) = \mathcal{R}(\mathrm{U}_1)$, $\mathcal{R}(\mathrm{A})^\perp = \mathcal{N}(\mathrm{A}^T) = \mathcal{R}(\mathrm{U}_2)$;

(b) $\mathcal{R}(\mathrm{A}^T) = \mathcal{R}(\mathrm{V}_1)$, $\mathcal{R}(\mathrm{A}^T)^\perp = \mathcal{N}(\mathrm{A}) = \mathcal{R}(\mathrm{V}_2)$;

(c) $\mathrm{rank}(\mathrm{A}) = r$ (the number of nonzero singular values).

Requires a proof.

Note:

- SVD can be used as a numerical tool to compute basis of $\mathcal{R}(\mathrm{A})$, $\mathcal{R}(\mathrm{A})^\perp$, $\mathcal{R}(\mathrm{A}^T)$, $\mathcal{N}(\mathrm{A})$
- we have previously learnt the following properties
  - $\mathrm{rank}(\mathrm{A}^T) = \mathrm{rank}(\mathrm{A})$
  - $\dim \mathcal{N}(\mathrm{A}) = n - \mathrm{rank}(\mathrm{A})$

  By SVD, the above properties are easily seen to be true.
- SVD is also used as a numerical tool to compute the rank of a matrix.

Matrix 2-Norm

Induced matrix p-norm from the vector p-norm

$$\|A\|_p = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \max_{\|x\|_p = 1} \|Ax\|_p$$

$p = 2$: matrix 2-norm or spectral norm

$$\|A\|_2 = \sigma_{\max}(A).$$

Proof:

- for any x with $\|x\|_2 \leq 1$,
$$\begin{aligned} \|Ax\|_2^2 &= \|U\boldsymbol{\Sigma}V^T x\|_2^2 = \|\boldsymbol{\Sigma}V^T x\|_2^2 \\ &\leq \sigma_1^2 \|V^T x\|_2^2 = \sigma_1^2 \|x\|_2^2 \leq \sigma_1^2 \end{aligned}$$
- $\|Ax\|_2 = \sigma_1$ if we choose $x = v_1$

Implication to linear transformation: let $y = Ax$ be a linear transformation maps x to y. Under the constraint $\|x\|_2 = 1$, the system output $\|y\|_2^2$ is maximized when x is chosen as the 1st right singular vector.

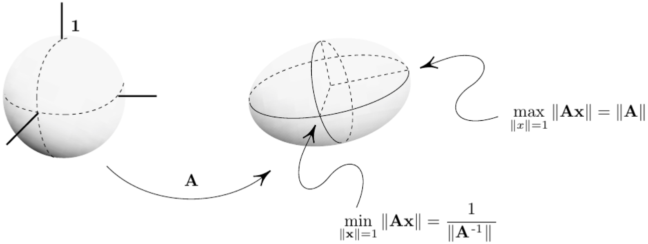

Figure 2: Linear transformation by nonsingular matrix A

When $A \in \mathbb{R}^{m \times n}$ is of full rank and $m \geq n$,

- $\|Ax\|_2 \geq \sigma_{\min}(A)\|x\|_2$ (hands-on exercise)
- can you use Figure 1 to help to understand?

Properties of Matrix 2-Norm

- $\|AB\|_2 \leq \|A\|_2 \|B\|_2$
  - in fact, $\|AB\|_p \leq \|A\|_p \|B\|_p$ for any $p \geq 1$
- $\|Ax\|_2 \leq \|A\|_2 \|x\|_2$
  - a special case of the 1st property
- $\|QAW\|_2 = \|A\|_2$ for any orthogonal Q, W
  - we also have $\|QAW\|_F = \|A\|_F$ for any orthogonal Q, W
- $\|A\|_2 \leq \|A\|_F \leq \sqrt{p}\|A\|_2$ (here $p = \min\{m, n\}$)
  - proof: $\|A\|_F = \|\boldsymbol{\Sigma}\|_F = \sqrt{\sum_{i=1}^{p} \sigma_i^2}$, and $\sigma_1^2 \leq \sum_{i=1}^{p} \sigma_i^2 \leq p\sigma_1^2$
- let A be square and nonsingular. Then, $\|A^{-1}\|_2 = 1/\sigma_{\min}(A)$

Schatten p-Norm

The function

$$f(\mathrm{A}) = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathrm{A})^p \right)^{1/p}, \qquad p \geq 1,$$

defines a matrix norm and is called the Schatten p-norm. Here $\sigma_i(\mathrm{A})$ ($i = 1,\ 2,\ \cdots,\ p$) are the singular values of A.

Nuclear norm:

$$\|\mathrm{A}\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathrm{A})$$

- a special case of the Schatten p-norm
- finds applications in rank approximation, e.g., for compressive sensing and matrix completion [Recht-Fazel-Parrilo'10]

1. B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Review, vol. 52, no. 3, pp. 471–501, 2010.

Readings

You are supposed to read

- Gene H. Golub and Charles F. Van Loan. *Matrix Computations*, Johns Hopkins University Press, 2013.

  Chapter 2.4.