SI231 Matrix Analysis and Computations
Singular Value Decomposition

Ziping Zhao

Spring Term 2022–2023

School of Information Science and Technology
ShanghaiTech University, Shanghai, China

http://si231.sist.shanghaitech.edu.cn

Singular Value Decomposition

- singular values, singular vectors, and singular value decomposition
- matrix norms
- linear systems
- least squares, pseudo-inverse, orthogonal projections
- low-rank matrix approximation
- variational characterizations for singular values
- singular value inequalities
- computations of the SVD

Main Results

- any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ admits a singular value decomposition

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

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

- matrix 2-norm: $\|\mathbf{A}\|_2 = \sigma_1$

- let r be the number of nonzero σ_i's, partition $\mathbf{U} = [\ \mathbf{U}_1\ \mathbf{U}_2\ ]$, $\mathbf{V} = [\ \mathbf{V}_1\ \mathbf{V}_2\ ]$, and let $\tilde{\mathbf{\Sigma}} = \mathrm{Diag}(\sigma_1, \ldots, \sigma_r)$
  - thin SVD: $\mathbf{A} = \mathbf{U}_1\tilde{\mathbf{\Sigma}}\mathbf{V}_1^T$
  - pseudo-inverse: $\mathbf{A}^\dagger = \mathbf{V}_1\tilde{\mathbf{\Sigma}}^{-1}\mathbf{U}_1^T$
  - linear system solution: $\mathbf{x} = \mathbf{A}^\dagger\mathbf{y} + \boldsymbol{\eta}$ for any $\boldsymbol{\eta} \in \mathcal{R}(\mathbf{V}_2)$ and $\mathbf{U}_2^T\mathbf{y} = \mathbf{0}$
  - least squares solution: $\mathbf{x} = \mathbf{A}^\dagger\mathbf{y} + \boldsymbol{\eta}$ for any $\boldsymbol{\eta} \in \mathcal{R}(\mathbf{V}_2)$
  - orthogonal projection: $\mathbf{P}_\mathbf{A} = \mathbf{U}_1\mathbf{U}_1^T$

Main Results

- low-rank matrix approximation: given $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $k \in \{1, \ldots, \min\{m, n\}\}$, the problem

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

  has a solution given by $\mathbf{B}^\star = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$

- in this lecture, we will deal with the real matrices–the complex case follows along the same lines

Singular Values, Singular Vectors, and Singular Value Decomposition

Theorem 1. Given any $\mathbf{A} \in \mathbb{R}^{m\times n}$, there exists a 3-tuple $(\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}) \in \mathbb{R}^{m\times m} \times \mathbb{R}^{m\times n} \times \mathbb{R}^{n\times n}$ such that

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

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

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

(requires a proof)

- the above decomposition is called the singular value decomposition (SVD)
- σ_i is called the ith singular value (always nonnegative even for complex matrices)
- $\mathbf{u}_i$ and $\mathbf{v}_i$ are called the ith left and right singular vectors, resp.

  $$\mathbf{u}_i^T\mathbf{A} = \sigma_i\mathbf{v}_i^T \Longleftarrow \mathbf{U}^T\mathbf{A} = \boldsymbol{\Sigma}\mathbf{V}^T \Longleftrightarrow \mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$$

  $$\Longleftrightarrow \mathbf{A}\mathbf{V} = \mathbf{U}\boldsymbol{\Sigma} \Longrightarrow \mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i \quad \text{for} \quad i = 1, \ldots, p$$

  $\mathbf{U}$ and $\mathbf{V}$ are called the left and right singular vector matrices, resp.
- the following notations may be used to denote singular values of a given $\mathbf{A}$

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

Singular Spectrum and Singular Subspace

- The set $\{\sigma_1, \ldots, \sigma_p\}$ is called the singular spectrum of $\mathbf{A}$
- A pair of k-dimensional subspaces $\mathcal{U}$ and $\mathcal{V}$ are called left and right singular subspaces of $\mathbf{A}$, resp., if

$$\mathbf{A}\mathbf{v} \in \mathcal{U} \quad \text{for all } \mathbf{v} \in \mathcal{V}$$

  and

$$\mathbf{A}^T\mathbf{u} \in \mathcal{V} \quad \text{for all } \mathbf{u} \in \mathcal{U}.$$

  We also write this as $\mathbf{A}\mathcal{V} \subset \mathcal{U}$ and $\mathbf{A}^T\mathcal{U} \subset \mathcal{V}$.

  – The simplest example is when $\mathcal{U}$ and $\mathcal{V}$ are spanned by a single pair of singular vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ of $\mathbf{A}$, respectively.

  – More generally, any pair of singular subspaces can be spanned by a subset of the singular vectors of $\mathbf{A}$, although the spanning vectors do not have to be singular vectors themselves.

Different Ways of Writing out SVD

- thin SVD: only compute the first $p = \min\{m, n\}$ columns of $\mathbf{U}$ and $\mathbf{V}$
  - for $m \leq n$: $(\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}) \in \mathbb{R}^{m\times m} \times \mathbb{R}^{m\times m} \times \mathbb{R}^{m\times n}$
  - for $n \leq m$: $(\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$
- partitioned form of the full SVD: let r be the number of nonzero singular values, and note $\sigma_1 \geq \ldots \geq \sigma_r > 0$, $\sigma_{r+1} = \ldots = \sigma_p = 0$. Then,

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

  where
  - $\tilde{\mathbf{\Sigma}} = \mathrm{Diag}(\sigma_1, \ldots, \sigma_r)$,
  - $\mathbf{U}_1 = [\ \mathbf{u}_1, \ldots, \mathbf{u}_r\ ] \in \mathbb{R}^{m\times r}$, $\mathbf{U}_2 = [\ \mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\ ] \in \mathbb{R}^{m\times(m-r)}$,
  - $\mathbf{V}_1 = [\ \mathbf{v}_1, \ldots, \mathbf{v}_r\ ] \in \mathbb{R}^{n\times r}$, $\mathbf{V}_2 = [\ \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\ ] \in \mathbb{R}^{n\times(n-r)}$.
- compact SVD: $\mathbf{A} = \mathbf{U}_1 \tilde{\mathbf{\Sigma}} \mathbf{V}_1^T$
  - both thin SVD and compact SVD are reduced SVD forms; in contrast, the one in Theorem 1 is also called full SVD
- outer-product form (i.e., dyadic decomposition): $\mathbf{A} = \sum_{i=1}^{p} \sigma_i \mathbf{u}_i \mathbf{v}_i^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$

SVD and Eigendecomposition

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

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{D}_1\mathbf{U}^T, \qquad \mathbf{D}_1 = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T = \mathrm{Diag}(\sigma_1^2, \ldots, \sigma_p^2, \underbrace{0, \ldots, 0}_{m-p \text{ zeros}}) \qquad (*)$$

$$\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{D}_2\mathbf{V}^T, \qquad \mathbf{D}_2 = \boldsymbol{\Sigma}^T\boldsymbol{\Sigma} = \mathrm{Diag}(\sigma_1^2, \ldots, \sigma_p^2, \underbrace{0, \ldots, 0}_{n-p \text{ zeros}}) \qquad (**)$$

Observations:

- $(*)$ and $(**)$ are the SVD's of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$, resp.
- $(*)$ and $(**)$ are the eigendecompositions of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$, resp.
- the left singular matrix $\mathbf{U}$ of $\mathbf{A}$ is the eigenvector matrix of $\mathbf{A}\mathbf{A}^T$
- the right singular matrix $\mathbf{V}$ of $\mathbf{A}$ is the eigenvector matrix of $\mathbf{A}^T\mathbf{A}$
- the squares of nonzero singular values of $\mathbf{A}$, $\sigma_1^2, \ldots, \sigma_r^2$, are the nonzero eigenvalues of both $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$.
- the relation between SVD and eigendec. can be used for analysis and computations

Insights of the Proof of SVD

- the proof of SVD is constructive
- to see the insights, consider the special case of square nonsingular $\mathbf{A}$
- $\mathbf{A}\mathbf{A}^T$ is PD, and denote its eigendecomposition by

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T, \qquad \text{with } \lambda_1 \geq \ldots \geq \lambda_n > 0.$$

- let $\boldsymbol{\Sigma} = \mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m})$, $\mathbf{V} = \mathbf{A}^T\mathbf{U}\boldsymbol{\Sigma}^{-1}$
- it can be verified that $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \mathbf{A}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$

- how to prove the SVD in the general case? (requires a proof)

Uniqueness of SVD

- the singular values σ_i's are uniquely determined and the nonzero singular values are the positive square roots of the eigenvalues of $\mathbf{A}\mathbf{A}^T$ or, equivalently, of $\mathbf{A}^T\mathbf{A}$

- the multiplicity of a singular value σ of $\mathbf{A}$ is the multiplicity of σ^2 as an eigenvalue of $\mathbf{A}\mathbf{A}^T$ or, equivalently, of $\mathbf{A}^T\mathbf{A}$

- a singular value σ of $\mathbf{A}$ is simple (algebraic multiplicity is 1) if σ^2 is a simple eigenvalue of $\mathbf{A}\mathbf{A}^T$ or, equivalently, of $\mathbf{A}^T\mathbf{A}$

uniqueness of SVD is higly related to the multiplicity of singular values and zero singular values of $\mathbf{A}$ and there are different kinds of characterizations; see Theorem 2.6.5 in **[Horn-Johnson'12]**.

Matrix Equivalences

Given $\mathbf{A}$, $\mathbf{B} \in \mathbb{R}^{m\times n}$, they are called equivalent if

$$\mathbf{B} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{P}$$

for some invertible $\mathbf{P}$ and $\mathbf{Q}$, which are called the equivalence transformations.

- Two matrices are equivalent if and only if they have the same rank.
- not be confused with matrix similarity, which is only defined for square matrices

Suppose $\mathbf{Q}$ and $\mathbf{P}$ are unitary matrices, i.e., $\mathbf{Q}^{-1} = \mathbf{Q}^H$ and $\mathbf{P}^{-1} = \mathbf{P}^H$. We say that $\mathbf{B}$ is unitarily (orthogonally) equivalent to $\mathbf{A}$, where $\mathbf{Q}$ and $\mathbf{P}$ are unitary (orthogonal) equivalence transformations.

- $\mathbf{B}$ has the same singular values as $\mathbf{A}$.
- If $\mathbf{u}$ and $\mathbf{v}$ are left and right singular vectors of $\mathbf{A}$, resp., so that $\mathbf{A}\mathbf{v} = \sigma\mathbf{u}$ and $\mathbf{A}^H\mathbf{u} = \sigma\mathbf{v}$, then $\mathbf{Q}^H\mathbf{u}$ and $\mathbf{P}^H\mathbf{v}$ are left and right singular vectors of $\mathbf{B}$, resp.

Properties of SVD

Property 1. The following properties hold:

(a) $\mathbf{A}^T = \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T$

(b) $\mathbf{A}$, $\mathbf{A}^*$, $\mathbf{A}^T$, and $\mathbf{A}^H$ have the same singular values

(c) $\mathbf{u}_i^T\mathbf{A}\mathbf{v}_i = \sigma_i$ for $i = 1, \ldots, p$, or, equivalently, in matrix form $\mathbf{U}^T\mathbf{A}\mathbf{V} = \mathbf{\Sigma}$

(d) $\mathrm{tr}(\mathbf{A}^T\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{A}^T) = \sum_{i=1}^p \sigma_i^2$

(e) let $\mathbf{A} \in \mathbb{R}^{n\times n}$, $|\det(\mathbf{A})| = |\det(\mathbf{\Sigma})| = \prod_{i=1}^n \sigma_i$

(f) $\mathrm{rank}(\mathbf{A}) < q$ (singular for square $\mathbf{A}$) if and only if 0 is one singular value of $\mathbf{A}$

(g) $\mathrm{rank}(\mathbf{A}) =$ number of nonzero singular values

(h) let $\mathbf{A} \in \mathbb{S}^n$, the singular values are the absolute values of eigenvalues of $\mathbf{A}$

(i) if $\mathbf{A}$ is invertible, $\mathbf{A}^{-1} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T$ (can be used to compute matrix inversion)

(j) for othogonal $\mathbf{P}$ and $\mathbf{Q}$, SVD of $\mathbf{P}\mathbf{A}\mathbf{Q}^T$ is given by $\tilde{\mathbf{U}}\mathbf{\Sigma}\tilde{\mathbf{V}}^T$ where $\tilde{\mathbf{U}} = \mathbf{P}\mathbf{U}$ and $\tilde{\mathbf{V}} = \mathbf{Q}\mathbf{V}$, i.e., singular values are orthogonally invariant (i.e., $\sigma_i(\mathbf{A}) = \sigma_i(\mathbf{P}\mathbf{A}\mathbf{Q}^T)$) but singular vectors not

Properties of SVD

Property 2. The following properties hold:

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

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

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

(Proof as a Quiz)

Note:

- in practice, SVD can be used a numerical tool for computing bases of $\mathcal{R}(\mathbf{A})$, $\mathcal{N}(\mathbf{A}^T)$, $\mathcal{R}(\mathbf{A}^T)$, $\mathcal{N}(\mathbf{A})$
- we have previously learnt the following properties
  - $\mathrm{rank}(\mathbf{A}^T) = \mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{A}\mathbf{A}^T) = \mathrm{rank}(\mathbf{A}^T\mathbf{A})$
  - $\mathrm{rank}(\mathbf{A}) + \mathrm{nullity}(\mathbf{A}) = n$

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