AMS526: Numerical Analysis I (Numerical Linear Algebra for Computational and Data Sciences)

Lecture 4: Singular Value Decomposition

Xiangmin Jiao

SUNY Stony Brook

Outline

1. Singular Value Decomposition (NLA§4-5)

Singular Value Decomposition (SVD)

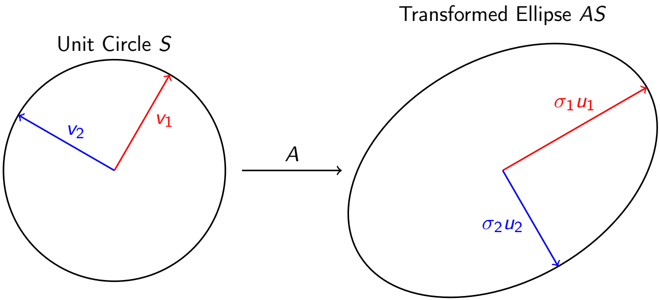
- Given $A \in \mathbb{R}^{m \times n}$, its *SVD* is

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

  where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal, and $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal
- If $A \in \mathbb{C}^{m \times n}$, then its SVD is $A = U \Sigma V^H$, where $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ are unitary, and $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal
- *Singular values* are diagonal entries of Σ, with entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$
- *Left singular vectors* of A are column vectors of U
- *Right singular vectors* of A are column vectors of V
- $Av_j = \sigma_j u_j$ for $1 \leq j \leq n$
- SVD plays a prominent role in data analysis and matrix analysis

Geometric Observation

- Given a unit hypersphere S in $\mathbb{R}^n$, AS denotes shape after transformation, which is a *hyperellipsoid* in $\mathbb{R}^m$
- Column vectors in V are the preimages of the principal semiaxes of the hyperellipsoid AS
- *Singular values* correspond to the principal semiaxes of hyperellipsoid
- *Left singular vectors* are parallel to principal semiaxes of AS
- *Right singular vectors* are preimages of principal semiaxes of AS



Two Different Types of SVD

- **Full SVD**: For $A \in \mathbb{R}^{m\times n}$, we have$U \in \mathbb{R}^{m\times m}$, $\Sigma \in \mathbb{R}^{m\times n}$, and $V \in \mathbb{R}^{n\times n}$ s.t.
$$A = U\Sigma V^T$$

Two Different Types of SVD

- **Full SVD**: For $A \in \mathbb{R}^{m \times n}$, we have$U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$, and $V \in \mathbb{R}^{n \times n}$ s.t.
$$A = U \Sigma V^T$$
- **Thin SVD** (**Reduced SVD**): Assuming $m \geq n$, we have $\hat{U} \in \mathbb{R}^{m \times n}$ and $\hat{\Sigma} \in \mathbb{R}^{n \times n}$ s.t.
$$A = \hat{U} \hat{\Sigma} V^T$$

Two Different Types of SVD

- **Full SVD**: For $A \in \mathbb{R}^{m \times n}$, we have$U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$, and $V \in \mathbb{R}^{n \times n}$ s.t.
$$A = U \Sigma V^T$$
- **Thin SVD (Reduced SVD)**: Assuming $m \geq n$, we have $\hat{U} \in \mathbb{R}^{m \times n}$ and $\hat{\Sigma} \in \mathbb{R}^{n \times n}$ s.t.
$$A = \hat{U} \hat{\Sigma} V^T$$
- Furthermore, notice that
$$A = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i v_i^T,$$
so we can keep only entries of U and V corresponding to nonzero σ_i.

Existence of SVD

Theorem

(Existence) Every matrix $A \in \mathbb{R}^{m\times n}$ has an SVD.

Proof: Let $\sigma_1 = \|A\|_2$. There exists $v_1 \in \mathbb{R}^n$ with $\|v_1\|_2 = 1$ and $\|Av_1\|_2 = \sigma_1$. Let U_1 and V_1 be orthogonal matrices whose first columns are $u_1 = Av_1/\sigma_1$ (or any unit-length vector if $\sigma_1 = 0$) and v_1, respectively. Note that (with block-matrix notation)

$$U_1^T A V_1 = S = \begin{bmatrix} \sigma_1 & \omega^T \\ 0 & B \end{bmatrix}. \qquad (1)$$

(Key remaining questions: What is ω, and how to deal with B?)
Furthermore, $\omega = 0$ because $\|S\|_2 = \sigma_1$, and

$$\left\| \begin{bmatrix} \sigma_1 & \omega^T \\ 0 & B \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \omega \end{bmatrix} \right\|_2 \geq \sigma_1^2 + \omega^T\omega = \sqrt{\sigma_1^2 + \omega^T\omega} \left\| \begin{bmatrix} \sigma_1 \\ \omega \end{bmatrix} \right\|_2,$$

implying that $\sigma_1 \geq \sqrt{\sigma_1^2 + \omega^T\omega}$ and $\omega = 0$.

Existence of SVD Cont'd

We then prove by induction using (1) from previous slide. If $m = 1$ or $n = 1$, then B is empty and we have $A = U_1 S V_1^T$. Otherwise, suppose $B = U_2 \Sigma_2 V_2^T$, and then

$$A = \underbrace{U_1 \begin{bmatrix} 1 & 0^T \\ 0 & U_2 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & 0^T \\ 0 & \Sigma_2 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1 & 0^T \\ 0 & V_2^T \end{bmatrix} V_1^T}_{V^T},$$

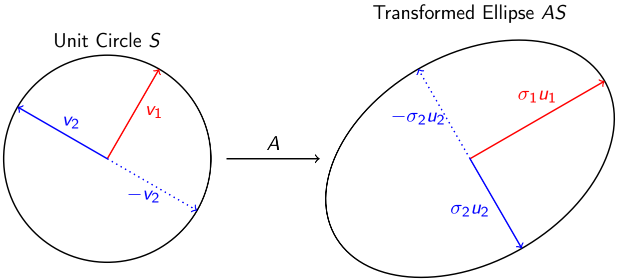
where U and V are orthogonal.

Uniqueness of SVD

Theorem

*(Uniqueness) The singular values $\{\sigma_j\}$ are uniquely determined. If A is square and the σ_j are distinct, the left and right singular vectors are uniquely determined **up to signs**.*

Geometric argument: If the lengths of semiaxes of a hyperellipsoid are distinct, then the orientations of semiaxes are determined up to signs.



Question: What are the signs of singular vectors if A is complex?

Uniqueness of SVD Cont'd

Algebraic argument: The proof can be done by induction. Consider the case where the σ_j are distinct. The 2-norm is unique, so σ_1 is unique. If v_1 is not unique up to sign, then the orthonormal bases of these vectors are right singular vectors of A, implying that σ_1 is not a simple singular value.

Once the first triplet σ_1, u_1, and v_1 are determined, the remainder of SVD follows from the subspace orthogonal to v_1. Because v_1 is unique up to sign, the orthogonal subspace is uniquely defined. The rest of the SVD can then be uniquely determined by induction.

Uniqueness of SVD Cont'd

Algebraic argument: The proof can be done by induction. Consider the case where the σ_j are distinct. The 2-norm is unique, so σ_1 is unique. If v_1 is not unique up to sign, then the orthonormal bases of these vectors are right singular vectors of A, implying that σ_1 is not a simple singular value.

Once the first triplet σ_1, u_1, and v_1 are determined, the remainder of SVD follows from the subspace orthogonal to v_1. Because v_1 is unique up to sign, the orthogonal subspace is uniquely defined. The rest of the SVD can then be uniquely determined by induction.

- Discussion: What if we change the sign of a singular vector?

Uniqueness of SVD Cont'd

Algebraic argument: The proof can be done by induction. Consider the case where the σ_j are distinct. The 2-norm is unique, so σ_1 is unique. If v_1 is not unique up to sign, then the orthonormal bases of these vectors are right singular vectors of A, implying that σ_1 is not a simple singular value.

Once the first triplet σ_1, u_1, and v_1 are determined, the remainder of SVD follows from the subspace orthogonal to v_1. Because v_1 is unique up to sign, the orthogonal subspace is uniquely defined. The rest of the SVD can then be uniquely determined by induction.

- Discussion: What if we change the sign of a singular vector?
- Discussion: What if σ_i is not distinct?

SVD vs. Eigenvalue Decomposition

- Eigenvalue decomposition of nondefective matrix A is $A = X \Lambda X^{-1}$

SVD vs. Eigenvalue Decomposition

- Eigenvalue decomposition of nondefective matrix A is $A = X\Lambda X^{-1}$
- Differences between SVD and eigenvalue decomposition
  - Not every matrix has eigenvalue decomposition, but every matrix has singular value decomposition
  - Eigenvalues may not always be real numbers, but singular values are always non-negative real numbers
  - Eigenvectors are not always orthogonal to each other (orthogonal for symmetric matrices), but left (or right) singular vectors are orthogonal to each other

SVD vs. Eigenvalue Decomposition

- Eigenvalue decomposition of nondefective matrix A is $A = X\Lambda X^{-1}$
- Differences between SVD and eigenvalue decomposition
  - Not every matrix has eigenvalue decomposition, but every matrix has singular value decomposition
  - Eigenvalues may not always be real numbers, but singular values are always non-negative real numbers
  - Eigenvectors are not always orthogonal to each other (orthogonal for symmetric matrices), but left (or right) singular vectors are orthogonal to each other
- Similarities
  - Singular values of A are square roots of eigenvalues of AA^T and A^TA, and their eigenvectors are left and right singular vectors, respectively
  - Singular values of Hermitian matrices are absolute values of eigenvalues, and there exist an SVD such that the eigenvectors are the singular vectors
  - This relationship can be used to compute singular values by hand

SVD vs. Eigenvalue Decomposition

- Eigenvalue decomposition of nondefective matrix A is $A = X\Lambda X^{-1}$
- Differences between SVD and eigenvalue decomposition
  - Not every matrix has eigenvalue decomposition, but every matrix has singular value decomposition
  - Eigenvalues may not always be real numbers, but singular values are always non-negative real numbers
  - Eigenvectors are not always orthogonal to each other (orthogonal for symmetric matrices), but left (or right) singular vectors are orthogonal to each other
- Similarities
  - Singular values of A are square roots of eigenvalues of AA^T and A^TA, and their eigenvectors are left and right singular vectors, respectively
  - Singular values of Hermitian matrices are absolute values of eigenvalues, and there exist an SVD such that the eigenvectors are the singular vectors
  - This relationship can be used to compute singular values by hand
- Discussion: Are the eigenvalues and eigenvectors of AA^T unique?

Matrix Properties via SVD

- Let r be number of nonzero singular values σ_i of $A \in \mathbb{R}^{m \times n}$
  - $\operatorname{rank}(A)$ is r
  - $\operatorname{range}(A) = \operatorname{span}\{u_1, u_2, \ldots, u_r\}$
  - $\operatorname{null}(A) = \operatorname{span}\{v_{r+1}, v_{r+2}, \ldots, v_n\}$
- 2-norm and Frobenius norm
  - $\|A\|_2 = \sigma_1$ and $\|A\|_F = \sqrt{\sum_i \sigma_i^2}$
- Determinant of matrix
  - For $A \in \mathbb{R}^{m \times m}$, $|\det(A)| = \prod_{i=1}^{m} \sigma_i$

Matrix Properties via SVD

- Let r be number of nonzero singular values σ_i of $A \in \mathbb{R}^{m\times n}$
  - rank(A) is r
  - range(A) $=$ span$\{u_1, u_2, \ldots, u_r\}$
  - null(A) $=$ span$\{v_{r+1}, v_{r+2}, \ldots, v_n\}$
- 2-norm and Frobenius norm
  - $\|A\|_2 = \sigma_1$ and $\|A\|_F = \sqrt{\sum_i \sigma_i^2}$
- Determinant of matrix
  - For $A \in \mathbb{R}^{m\times m}$, $|\det(A)| = \prod_{i=1}^{m} \sigma_i$
- However, SVD may not be the most efficient way in solving problems
- Algorithms for SVD are similar to those for eigenvalue decomposition and we will discuss them later