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

Lecture 4: More on SVD; Projectors

Xiangmin Jiao

SUNY Stony Brook

Outline

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 and are the preimages of the principal semiaxes of AS
- SVD plays prominent role in data analysis and matrix analysis

Two Different Types of SVD

- **Full SVD**: $U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$, and $V \in \mathbb{R}^{n \times n}$ is

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

Two Different Types of SVD

- **Full SVD**: $U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$, and $V \in \mathbb{R}^{n \times n}$ is

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

- **Thin SVD (Reduced SVD)**: $\hat{U} \in \mathbb{R}^{m \times n}$, $\hat{\Sigma} \in \mathbb{R}^{n \times n}$ (assume $m \geq n$)

$$A = \hat{U} \hat{\Sigma} V^T$$

Two Different Types of SVD

- **Full SVD**: $U \in \mathbb{R}^{m\times m}$, $\Sigma \in \mathbb{R}^{m\times n}$, and $V \in \mathbb{R}^{n\times n}$ is

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

- **Thin SVD (Reduced SVD)**: $\hat{U} \in \mathbb{R}^{m\times n}$, $\hat{\Sigma} \in \mathbb{R}^{n\times n}$ (assume $m \geq n$)

$$A = \hat{U}\hat{\Sigma}V^T$$

- 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

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

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). 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 hyperellipse are distinct, then the semiaxes themselves are determined by the geometry up to signs.

Uniqueness of SVD Cont'd

Algebraic argument: Based on 2-norm and prove by induction. Consider the case where the σ_j are distinct. The 2-norm is unique, so is σ_1. 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 σ_1, u_1, and v_1 are determined, the remainder of SVD is determined by the space orthogonal to v_1. Because v_1 is unique up to sign, the orthogonal subspace is uniquely defined. Then prove by induction.

Uniqueness of SVD Cont'd

Algebraic argument: Based on 2-norm and prove by induction. Consider the case where the σ_j are distinct. The 2-norm is unique, so is σ_1. 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 σ_1, u_1, and v_1 are determined, the remainder of SVD is determined by the space orthogonal to v_1. Because v_1 is unique up to sign, the orthogonal subspace is uniquely defined. Then prove by induction.

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

Uniqueness of SVD Cont'd

Algebraic argument: Based on 2-norm and prove by induction. Consider the case where the σ_j are distinct. The 2-norm is unique, so is σ_1. 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 σ_1, u_1, and v_1 are determined, the remainder of SVD is determined by the space orthogonal to v_1. Because v_1 is unique up to sign, the orthogonal subspace is uniquely defined. Then prove by induction.

- Question: What if we change the sign of a singular vector?
- Question: 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 eigenvectors are singular vectors (up to signs)
  - This relationship can be used to compute singular values by hand

Matrix Properties via SVD

- Let r be number of nonzero singular values of $A \in \mathbb{R}^{m \times n}$
  - $\text{rank}(A)$ is r
  - $\text{range}(A) = \text{span}\{u_1, u_2, \ldots, u_r\}$
  - $\text{null}(A) = \text{span}\{v_{r+1}, v_{r+2}, \ldots, v_n\}$
- 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 of $A \in \mathbb{R}^{m \times n}$
  - rank(A) is r
  - range(A) $= \text{span}\{u_1, u_2, \ldots, u_r\}$
  - null(A) $= \text{span}\{v_{r+1}, v_{r+2}, \ldots, v_n\}$
- 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

Outline

Projectors (Projection Matrices)

- A *projector* (aka *projection matrix*) satisfies $P^2 = P$. They are also said to be *idempotent*.
  - *Orthogonal* projector
  - *Oblique* projector
- An *orthogonal projector* is one that projects onto a subspace S_1 along a space S_2, where S_1 and S_2 are orthogonal.
  - $S_1 = \text{range}(P)$
  - $S_2 = \text{null}(P)$
- Example: $\begin{bmatrix} 0 & 0 \\ \alpha & 1 \end{bmatrix}$
  - is an oblique projector if $\alpha \neq 0$,
  - is orthogonal projector if $\alpha = 0$.

Orthogonal Projector

Theorem

A projector P is orthogonal if and only if $P = P^T$.

Note: An alternative definition of *orthogonal projection* is $P^2 = P$ and $P = P^T$, and it projects onto $S = \text{range}(P)$.

Proof.

"If" direction: If $P = P^T$, then $(Px)^T(I - P)y = x^T(P - P^2)y = 0$.
"Only if" direction: Use SVD. Suppose P projects onto S_1 along S_2 where $S_1 \perp S_2$, and S_1 has dimension n. Let $\{q_1, \dots, , q_n\}$ be orthonormal basis of S_1 and $\{q_{n+1}, \dots, q_m\}$ be a basis for S_2. Let Q be unitary matrix whose jth column is q_j, and we have $PQ = (q_1, q_2, \dots, q_n, 0, \dots, 0)$, so $Q^T PQ = \text{diag}(1, 1, \cdots, 1, 0, \cdots) = \Sigma$, and $P = Q\Sigma Q^T$. □

Orthogonal Projector

Theorem

A projector P is orthogonal if and only if $P = P^T$.

Note: An alternative definition of *orthogonal projection* is $P^2 = P$ and $P = P^T$, and it projects onto $S = \text{range}(P)$.

Proof.

"If" direction: If $P = P^T$, then $(Px)^T(I - P)y = x^T(P - P^2)y = 0$.
"Only if" direction: Use SVD. Suppose P projects onto S_1 along S_2 where $S_1 \perp S_2$, and S_1 has dimension n. Let $\{q_1, \ldots, , q_n\}$ be orthonormal basis of S_1 and $\{q_{n+1}, \ldots, q_m\}$ be a basis for S_2. Let Q be unitary matrix whose jth column is q_j, and we have $PQ = (q_1, q_2, \ldots, q_n, 0, \ldots, 0)$, so $Q^T PQ = \text{diag}(1, 1, \cdots, 1, 0, \cdots) = \Sigma$, and $P = Q\Sigma Q^T$. □

Question: Are orthogonal projectors orthogonal matrices?

Complementary Projectors

- Complementary projectors: P vs. $I - P$. We write $I - P$ as $P_{\perp}$
- What space does $I - P$ project?

Complementary Projectors

- Complementary projectors: P vs. $I - P$. We write $I - P$ as $P_\perp$
- What space does $I - P$ project?
  - Answer: null(P).
  - range$(I - P) \supseteq$ null(P) because $Pv = 0 \Rightarrow (I - P)v = v$.
  - range$(I - P) \subseteq$ null(P) because for any v $(I - P)v = v - Pv \in$ null(P).
- A projector separates $\mathbb{R}^m$ into two complementary subspace: range space and null space (i.e., range(P) + null$(P) = \mathbb{R}^m$ and range$(P) \cap$ null$(P) = \{0\}$ for projector $P \in \mathbb{R}^{m\times m}$)
- It projects onto range space along null space
  - In other words, $x = Px + r$, where $r \in$ null(P)
- Question: Are range space and null space of projector orthogonal to each other?

Uniqueness of Orthogonal Projector

- Orthogonal projector for a subspace is unique
- In other words, for $S \subseteq \mathbb{R}^n$ be a subspace, if P_1 and P_2 are each orthogonal projector onto S, then $P_1 = P_2$
- Proof: For any $z \in \mathbb{R}^n$,

$$
\begin{aligned}
\|(P_1 - P_2)z\|_2^2 &= z^T(P_1 - P_2)(P_1 - P_2)z \\
&= z^T P_1(P_1 - P_2)z - z^T P_2(P_1 - P_2)z \\
&= z^T P_1(I - P_2)z + z^T P_2(I - P_1)z \\
&= (P_1 z)^T(I - P_2)z + (P_2 z)^T(I - P_1)z \\
&= 0
\end{aligned}
$$

Therefore, $\|P_1 - P_2\|_2 = 0$, and $P_1 = P_2$.

Projections with Orthonormal Basis

- Given unit vector q, $P_q = qq^T$ and $P_{\perp q} = I - P_q$
- Given any matrix $Q \in \mathbb{R}^{m \times n}$ whose columns q_j are orthonormal, $P = QQ^T = \sum_j q_j q_j^T$ is orthogonal projector onto range(Q)
- SVD-related projections
  - Suppose $A = U\Sigma V^T \in \mathbb{R}^{m \times n}$ is SVD of A, and $r = \text{rank}(A)$
  - Partition U and V to

$$U = \Big[\underbrace{U_r}_{r} \big| \underbrace{\tilde{U}_r}_{m-r}\Big], \quad V = \Big[\underbrace{V_r}_{r} \big| \underbrace{\tilde{V}_r}_{n-r}\Big]$$

  - What do $U_r U_r^T$, $\tilde{U}_r \tilde{U}_r^T$, $V_r V_r^T$, and $\tilde{V}_r \tilde{V}_r^T$ project onto, respectively?

Projections with Orthonormal Basis

- Given unit vector q, $P_q = qq^T$ and $P_{\perp q} = I - P_q$
- Given any matrix $Q \in \mathbb{R}^{m \times n}$ whose columns q_j are orthonormal, $P = QQ^T = \sum_j q_j q_j^T$ is orthogonal projector onto range(Q)
- SVD-related projections
  - Suppose $A = U\Sigma V^T \in \mathbb{R}^{m \times n}$ is SVD of A, and $r = \text{rank}(A)$
  - Partition U and V to

$$U = \Big[\underset{r}{U_r} \,|\, \underset{m-r}{\tilde{U}_r}\Big], \quad V = \Big[\underset{r}{V_r} \,|\, \underset{n-r}{\tilde{V}_r}\Big]$$

  - What do $U_r U_r^T$, $\tilde{U}_r \tilde{U}_r^T$, $V_r V_r^T$, and $\tilde{V}_r \tilde{V}_r^T$ project onto, respectively?
    - ★ Answer: range(A), null(A^T), range(A^T), null(A)

Projection with Arbitrary Basis

- For arbitrary vector a, we write $P_a = \frac{aa^T}{a^T a}$ and $P_{\perp a} = I - P_a$
- Given any matrix $A \in \mathbb{R}^{m \times n}$ that **has full rank** and $m \geq n$. Let $A = U \Sigma V^T$ be its SVD

  $$P = UU^T = A(A^T A)^{-1} A^T$$

  is orthogonal projection onto range(P)
- $(A^T A)^{-1} A^T$ is called the *pseudo-inverse* of A, denoted as A^+
- Therefore,

  $$P = UU^T = AA^+$$

- In addition, $A^+ A = I$
- Note: If $m < n$, $A^+ = A^T (AA^T)^{-1}$, and $AA^+ = I$ and $A^+ A$ is orthogonal projection onto range(A^T)

Distance Between Subspaces

- Suppose S_1 and S_2 are subspaces of $\mathbb{R}^n$, $\dim(S_1) = \dim(S_2)$, and P_i is orthogonal projection onto S_i
- The *distance* between S_1 and S_2 is

$$\text{dist}(S_1, S_2) = \|P_1 - P_2\|_2$$

- Suppose $W = [\underset{k}{W_1}|\underset{n-k}{W_2}], Z = [\underset{k}{Z_1}|\underset{n-k}{Z_2}]$ are $n \times n$ orthogonal matrices. If $S_1 = \text{range}(W_1)$ and $S_2 = \text{range}(Z_1)$, then

$$\text{dist}(S_1, S_2) = \|W_1^T Z_2\|_2 = \|Z_1^T W_2\|_2$$

  (proof omitted here)
- In general, $0 \leq \text{dist}(S_1, S_2) \leq 1$
- If subspaces are lines or planes, $\text{dist}(S_1, S_2)$ is sine of angle between them