The singular value decomposition

Nicholas Hu · Last updated on 2025-01-09

Let $A \in \mathbb{C}^{m \times n}$. The **singular value decomposition (SVD)** is a factorization of A as $U \Sigma V^*$, 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 (rectangular) diagonal with *nonnegative* entries. [1] In other words, $A = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i v_i^*$, where u_i and v_i are the i^{th} columns of U and V and σ_i is the i^{th} diagonal entry of Σ. The vectors u_i and v_i are called **left** and **right singular vectors** of A and the scalars σ_i are called **singular values** of A; by convention, we arrange the singular values in decreasing order.

If an SVD of A has r nonzero singular values, then $\{u_i\}_{i=1}^r$ is an orthonormal basis of $\operatorname{im}(A)$ because $Av_i = \sigma_i u_i$ for all i. Hence r must be the rank of A and $\{u_i\}_{i=r+1}^m$ an orthonormal basis of $\ker(A^*)$; similarly, $\{v_i\}_{i=1}^r$ and $\{v_i\}_{i=r+1}^n$ are orthonormal bases of $\operatorname{im}(A^*)$ and $\ker(A)$.

Existence

Assume without loss of generality that $m \geq n$. Clearly, the matrix A^*A is (Hermitian) positive semidefinite, so by the spectral theorem, $A^*A = V \Lambda V^*$ for some unitary $V \in \mathbb{C}^{n \times n}$ and some diagonal $\Lambda \in \mathbb{R}^{n \times n}$ with diagonal entries $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. Set $\sigma_i = \sqrt{\lambda_i}$ for each i and $Av_i = \sigma_i u_i$ for each nonzero σ_i. If r is as above, $\hat{U} := [u_1 \quad \cdots \quad u_r] \in \mathbb{C}^{m \times r}$, and $\hat{\Sigma} := \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r \times r}$, then by construction

$$AV = \hat{U} \begin{bmatrix} \hat{\Sigma} & 0_{r \times (n-r)} \end{bmatrix}.$$

Moreover, $\langle u_i, u_j \rangle = \langle Av_i/\sigma_i, Av_j/\sigma_j \rangle = \langle \lambda_i v_i, v_j \rangle / \sigma_i \sigma_j = \delta_{ij}$, so $\{u_i\}_{i=1}^r$ is orthonormal. Extending this set to an orthonormal basis $\{u_i\}_{i=1}^m$, and defining $U = [u_1 \quad \cdots \quad u_m] \in \mathbb{C}^{m \times m}$ and

$$\Sigma = \begin{bmatrix} \hat{\Sigma} & 0_{r \times (n-r)} \\ 0_{(m-r) \times r} & 0_{(m-r) \times (n-r)} \end{bmatrix} \in \mathbb{R}^{m \times n},$$

we obtain $A = U \Sigma V^*$ as required. [2]

Although an SVD is not unique, this argument shows that the singular values are unique and that the singular vectors are unique up to complex signs if $m = n$ and the singular values are distinct, since we must have $A^*A = V(\Sigma^*\Sigma)V^*$.

Low-rank approximation

> **Eckart–Young theorem**
>
> Suppose that $U \Sigma V^*$ is an SVD of a matrix $A \in \mathbb{C}^{m \times n}$ with rank r. If $k \leq r$ and $A_k := \sum_{i=1}^k \sigma_i u_i v_i^*$, then $\|A - B\|_2 \geq \sigma_{k+1} = \|A - A_k\|_2$ for all $B \in \mathbb{C}^{m \times n}$ such that $\operatorname{rank}(B) \leq k$ (where $\sigma_{r+1} := 0$). In particular, $\|A\|_2 = \sigma_1$.

Proof. Suppose that $B \in \mathbb{C}^{m\times n}$ is such that $\operatorname{rank}(B) \le k$. Then $\dim(\ker(B)) \ge n-k$, so there exists a $v \in \ker(B) \cap \operatorname{span}\{v_i\}_{i=1}^{k+1}$ such that $\|v\|_2 = 1$. Hence $\|A-B\|_2^2 \ge \|(A-B)v\|_2^2 = \|Av\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 |\langle v, v_i\rangle|^2 \ge \sigma_{k+1}^2$. Similarly, if $v \in \mathbb{C}^n$ with $\|v\|_2 = 1$, then $\|(A-A_k)v\|_2^2 = \sum_{i=k+1}^{r} \sigma_i^2 |\langle v, v_i\rangle|^2 \le \sigma_{k+1}^2$, with equality if $v = v_{k+1}$. ∎

An analogous theorem holds for the Frobenius norm (which can be proven similarly). In fact, we have the following generalization.

> **Eckart–Young–Mirsky theorem**
>
> Suppose that $U\Sigma V^*$ is an SVD of a matrix $A \in \mathbb{C}^{m\times n}$ with rank r and let $\|\cdot\|$ be a *unitarily invariant* norm. If $k \le r$ and $A_k := \sum_{i=1}^{k} \sigma_i u_i v_i^*$, then $\|A-B\| \ge \|A-A_k\|$ for all $B \in \mathbb{C}^{m\times n}$ such that $\operatorname{rank}(B) \le k$.

Proof. We begin by proving *Weyl's inequality* for singular values:

$$\sigma_{i+j-1}(A+B) \le \sigma_i(A) + \sigma_j(B),$$

where $\sigma_i(\,\cdot\,)$ denotes the i^{th} singular value of a given matrix. Let $A_k := \sum_{i=1}^{k} \sigma_i(A) u_i v_i^*$. Then $\operatorname{rank}(A_{i-1} + B_{j-1}) \le (i-1) + (j-1) = i+j-2$, so by the Eckart–Young theorem, $\sigma_{i+j-1}(A+B) \le \|(A+B) - (A_{i-1} + B_{j-1})\|_2 \le \|A - A_{i-1}\|_2 + \|B - B_{j-1}\|_2 = \sigma_i(A) + \sigma_j(B)$.

Now suppose that $B \in \mathbb{C}^{m\times n}$ is such that $\operatorname{rank}(B) \le k$. By Weyl's inequality, $\sigma_{k+i}(A) \le \sigma_{k+1}(B) + \sigma_i(A-B) = \sigma_i(A-B)$ for all i (where $\sigma_{k+i}(\,\cdot\,) := 0$ if $k+i > \min\{m,n\}$), so there exist $\theta_i \in [0,1]$ such that $\sigma_{k+i}(A) = \theta_i \sigma_i(A-B)$. For $0 \le j \le \min\{m,n\}$, let $D_j^{\pm} \in \mathbb{R}^{m\times n}$ be the (rectangular) diagonal matrix with diagonal entries

$$(D_j^{\pm})_{ii} := \begin{cases} \sigma_i(A-B) & \text{if } i < j, \\ \pm\sigma_i(A-B) & \text{if } i = j, \\ \sigma_{k+i}(A) & \text{if } i > j. \end{cases}$$

Then $\|A - A_k\| = \|D_0^+\|$ and $\|A-B\| = \|D_{\min\{m,n\}}^+\|$ since $\|\cdot\|$ is unitarily invariant. Moreover, $\|D_j^+\| = \|D_j^-\|$ is increasing in j because $\|D_{j-1}^+\| = \|\frac{1+\theta_j}{2} D_j^+ + \frac{1-\theta_j}{2} D_j^-\| \le \frac{1+\theta_j}{2}\|D_j^+\| + \frac{1-\theta_j}{2}\|D_j^-\| = \|D_j^+\|$, so $\|A - A_k\| \le \|A-B\|$. ∎

---

1. If $A \in \mathbb{R}^{m\times n}$, an SVD is defined analogously; i.e., with U and V orthogonal. ↩

2. If we instead add $n-r$ rows of zeroes to $[\hat{\Sigma} \quad 0]$, forming a square matrix, the resulting decomposition is sometimes called the **thin SVD**; if we instead omit the last $n-r$ columns, what remains is sometimes called the **compact SVD**. ↩