The Singular Value Decomposition (SVD): Detailed Analysis

AI Assistant

November 18, 2025

I. Intuition: Geometric Interpretation

The Singular Value Decomposition (SVD) is a factorization of a complex $m \times n$ matrix $\mathbf{A}$ into three simple, interpretable linear transformations. Geometrically, the action of $\mathbf{A}$ on the unit sphere in the input space ($\mathbb{R}^n$) results in an ellipse (or ellipsoid) in the output space ($\mathbb{R}^m$).

- The singular values in $\mathbf{\Sigma}$ are the **lengths of the semi-axes** of this resulting ellipse. They measure the **"gain" or "amplification"** of the matrix along specific directions.
- The columns of $\mathbf{V}$ define the **input directions** (the principal axes of the original unit sphere) that get mapped to the axes of the ellipse.
- The columns of $\mathbf{U}$ define the **output directions** (the axes of the ellipse) where the transformation ends up.

Sequence of Transformation

The transformation can be visualized as a sequence of three operations:

1. **Rotation/Reflection ($\mathbf{V}^T$):** Orients the input space axes ($\mathbf{v}_i$) along the directions of maximum variance.
2. **Scaling ($\mathbf{\Sigma}$):** Stretches or shrinks these oriented axes by factors equal to the singular values (σ_i).
3. **Rotation/Reflection (U):** Orients the final resulting ellipse in the output space. The columns of $\mathbf{U}$ define the axes ($\mathbf{u}_i$) of this final ellipsoid.

II. The Mathematical Decomposition

The SVD of an $m \times n$ matrix $\mathbf{A}$ is given by:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{1}$$

A. Defining the Components

U: The $m \times m$ matrix of **Left Singular Vectors** ($\mathbf{u}_i$). It is an orthogonal matrix ($\mathbf{U}^T\mathbf{U} = \mathbf{I}_m$), meaning its column vectors form an orthonormal basis for $\mathbb{R}^m$.

V: The $n \times n$ matrix of **Right Singular Vectors** ($\mathbf{v}_i$). It is also an orthogonal matrix ($\mathbf{V}^T\mathbf{V} = \mathbf{I}_n$), and its column vectors form an orthonormal basis for $\mathbb{R}^n$.

$\mathbf{\Sigma}$: The $m \times n$ rectangular diagonal matrix of **Singular Values** (σ_i).

B. Derivation of Singular Values

The singular values are unique and always non-negative, ordered such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, where $r = \text{rank}(\mathbf{A})$. They are directly related to the eigenvalues of the square symmetric matrices $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$.

- The singular values σ_i are the square roots of the non-zero eigenvalues (λ_i) of $\mathbf{A}^T\mathbf{A}$: $\sigma_i = \sqrt{\lambda_i(\mathbf{A}^T\mathbf{A})}$.
- The right singular vectors ($\mathbf{v}_i$) are the corresponding eigenvectors of $\mathbf{A}^T\mathbf{A}$.
- The left singular vectors ($\mathbf{u}_i$) are the corresponding eigenvectors of $\mathbf{A}\mathbf{A}^T$.

III. Key Insights: The Four Fundamental Subspaces

The rank r defines the size of the non-trivial subspaces. The columns of $\mathbf{U}$ and $\mathbf{V}$ naturally span the four fundamental subspaces:

Subspace	**Basis (from SVD)**	**Dimension**
Column Space $\text{Col}(\mathbf{A}) \subset \mathbb{R}^m$	$\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$	r
Null Space $\text{Null}(\mathbf{A}) \subset \mathbb{R}^n$	$\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$	$n - r$
Row Space $\text{Col}(\mathbf{A}^T) \subset \mathbb{R}^n$	$\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$	r
Left Null Space $\text{Null}(\mathbf{A}^T) \subset \mathbb{R}^m$	$\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$	$m - r$

IV. Low-Rank Approximation and Data Compression

The full expansion of $\mathbf{A}$ is the sum of r rank-one matrices:

$$\mathbf{A} = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^T = \sum_{i=1}^{r} \sigma_i\mathbf{u}_i\mathbf{v}_i^T \tag{2}$$

The power of SVD lies in the **Eckart–Young Theorem**, which states that the best possible rank k approximation ($\mathbf{A}_k$) is achieved by keeping only the k largest singular values and their corresponding vectors:

$$\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i\mathbf{u}_i\mathbf{v}_i^T \quad \text{for } k < r \tag{3}$$

This property is critical for data analysis, image compression, and matrix denoising because the information is ordered by importance (magnitude of σ_i).