Matrix Decomposition : SVD, and Spectral Decomposition

Dr. Shailesh Sivan

June 11, 2025

Matrix Decomposition: Overview

- Matrix decomposition simplifies matrix operations and linear algebra problems.
- Decompositions factor a matrix into simpler components:
  - LU Decomposition (Lower-Upper)
  - QR Decomposition (Orthogonal-Triangular)
  - **SVD (Singular Value Decomposition)**
  - **Spectral (Eigenvalue) Decomposition**

Dataset as a Matrix

Let a dataset be represented as a matrix:

$$X \in \mathbb{R}^{n \times d} \quad \text{(n samples, d features)}$$

We compute the covariance matrix:

$$C = \frac{1}{n-1} X^T X$$

Then apply spectral decomposition:

$$C = Q \Lambda Q^T$$

- Q: matrix of eigenvectors
- Λ: diagonal matrix of eigenvalues

Significance of Eigenvectors and Eigenvalues

Eigenvectors

Indicate the directions (axes) along which the data varies the most — the principal directions.

Eigenvalues

Quantify the amount of variance (information) present along each eigenvector direction.

Use in PCA: Select top-k eigenvectors with highest eigenvalues for dimensionality reduction.

Geometric Interpretation

- Eigenvectors form a new coordinate system (orthogonal axes).
- Eigenvalues tell how "stretched" the data is along each axis.

$$X \to \text{rotate} \to \text{project} \to \text{reduce}$$

New representation:

$$Z = XQ_k$$

Where Q_k contains the top-k eigenvectors.

Summary Table

Concept	Significance
Eigenvectors	Principal axes in feature space (directions of variance)
Eigenvalues	Magnitude of variance in corresponding direction
PCA	Uses top-k eigenvectors to reduce dimensionality
Noise Filtering	Discarding low eigenvalue directions removes noise
Compression	Low-rank approximation using leading eigenvectors

Singular Value Decomposition (SVD) - Concept

$$\text{For any } A \in \mathbb{R}^{m \times n}, \quad A = U\Sigma V^T$$

- $U \in \mathbb{R}^{m \times m}$: orthogonal, columns = left singular vectors
- $V \in \mathbb{R}^{n \times n}$: orthogonal, columns = right singular vectors
- $\Sigma \in \mathbb{R}^{m \times n}$: diagonal matrix with singular values

Algorithm: Steps for SVD

Given $A \in \mathbb{R}^{m \times n}$:

1. Compute $A^T A \in \mathbb{R}^{n \times n}$
2. Find eigenvalues λ_i and eigenvectors v_i of $A^T A$
3. Singular values: $\sigma_i = \sqrt{\lambda_i}$
4. Right singular vectors: v_i (columns of V)
5. Left singular vectors: $u_i = \frac{1}{\sigma_i} A v_i$ (columns of U)
6. Construct:

$$U = [u_1, \ldots, u_r], \quad V = [v_1, \ldots, v_r], \quad \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r)$$

Example: SVD of a 2x2 Matrix

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \Rightarrow A^T A = \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix}$$

- Eigenvalues of $A^T A$: $\lambda_1 = 16, \lambda_2 = 4$
- $\sigma_1 = 4, \quad \sigma_2 = 2$
- Eigenvectors of $A^T A$ give V, then $U = \frac{1}{\sigma_i} A v_i$
- Final decomposition: $A = U \Sigma V^T$

Feature Reduction Using SVD

$$A = \begin{bmatrix} 2 & 4 & 1 & 3 \\ 1 & 3 & 2 & 5 \\ 4 & 2 & 3 & 1 \\ 3 & 5 & 4 & 2 \\ 5 & 1 & 2 & 4 \end{bmatrix}$$

- Size: 5×4 (5 samples, 4 features)
- Goal: Reduce to 5×2

Step 1: SVD Decomposition

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

Where:

$$U = \begin{bmatrix} -0.45 & 0.58 & -0.48 & 0.50 \\ -0.52 & 0.16 & 0.82 & -0.15 \\ -0.50 & -0.64 & -0.05 & -0.57 \\ -0.59 & 0.49 & -0.14 & -0.63 \\ -0.33 & -0.07 & -0.27 & 0.90 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 12.3 & 0 & 0 & 0 \\ 0 & 5.2 & 0 & 0 \\ 0 & 0 & 2.1 & 0 \\ 0 & 0 & 0 & 1.1 \end{bmatrix}$$

$$V = \begin{bmatrix} -0.52 & 0.43 & -0.57 & -0.47 \\ -0.50 & -0.58 & 0.27 & 0.58 \\ -0.47 & 0.47 & 0.72 & -0.22 \\ -0.51 & -0.49 & -0.28 & 0.64 \end{bmatrix}$$

Step 2: Keep Top-2 Components

We retain only top 2 singular values and vectors:

$U_2 =$ first 2 columns of U, $\quad \Sigma_2 =$ top-left $2{\times}2$, $\quad V_2 =$ first 2 columns of V

$$A_{\text{reduced}} = U_2 \Sigma_2 = A \cdot V_2$$

$$V_2 = \begin{bmatrix} -0.52 & 0.43 \\ -0.50 & -0.58 \\ -0.47 & 0.47 \\ -0.51 & -0.49 \end{bmatrix}$$

Step 3: Compute $A_{\text{reduced}} = A \cdot V_2$

$$A_{\text{reduced}} = \begin{bmatrix} 2 & 4 & 1 & 3 \\ 1 & 3 & 2 & 5 \\ 4 & 2 & 3 & 1 \\ 3 & 5 & 4 & 2 \\ 5 & 1 & 2 & 4 \end{bmatrix} \cdot \begin{bmatrix} -0.52 & 0.43 \\ -0.50 & -0.58 \\ -0.47 & 0.47 \\ -0.51 & -0.49 \end{bmatrix}$$

$$A_{\text{reduced}} \approx \begin{bmatrix} -5.94 & -2.67 \\ -5.65 & -2.62 \\ -5.16 & 1.09 \\ -7.45 & -1.04 \\ -5.84 & 0.17 \end{bmatrix}$$

Final Reduced Matrix A_{reduced}

$$A_{\text{reduced}} = \begin{bmatrix} -5.94 & -2.67 \\ -5.65 & -2.62 \\ -5.16 & 1.09 \\ -7.45 & -1.04 \\ -5.84 & 0.17 \end{bmatrix}$$

- New size: 5×2
- Each sample is now represented using only 2 features
- Dimensionality reduced while preserving structure

Spectral Decomposition: Concept

Applies to symmetric matrices: $A = A^T \in \mathbb{R}^{n \times n}$

$$A = Q \Lambda Q^T$$

- Q: orthogonal matrix of eigenvectors
- Λ: diagonal matrix of eigenvalues
- Basis for many ML techniques like PCA

Algorithm: Steps for Spectral Decomposition

Given symmetric matrix $A \in \mathbb{R}^{n \times n}$:

1. Compute eigenvalues $\lambda_1, \ldots, \lambda_n$
2. Compute eigenvectors $\mathbf{q}_1, \ldots, \mathbf{q}_n$
3. Normalize eigenvectors to unit length
4. Form:
$$Q = [\mathbf{q}_1 \cdots \mathbf{q}_n], \quad \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$$
5. Decomposition:
$$A = Q \Lambda Q^T$$

Example: Spectral Decomposition

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

- Eigenvalues: $\lambda_1 = 3, \lambda_2 = 1$
- Eigenvectors:

$$\mathbf{q}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{q}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

- Decomposition:

$$A = Q \Lambda Q^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^T$$

Original Symmetric Matrix A

$$A = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 1 & 3 & 0 & 1 \\ 2 & 0 & 5 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}$$

- $A \in \mathbb{R}^{4 \times 4}$, real symmetric
- Spectral decomposition is applicable

Spectral Decomposition

$$A = Q \Lambda Q^T$$

Where:

- Q: matrix of eigenvectors
- Λ: diagonal matrix of eigenvalues

Assume:

$$\Lambda = \begin{bmatrix} 8.1 & 0 & 0 & 0 \\ 0 & 3.2 & 0 & 0 \\ 0 & 0 & 2.1 & 0 \\ 0 & 0 & 0 & 0.6 \end{bmatrix} \quad Q = \begin{bmatrix} 0.5 & -0.4 & 0.6 & -0.5 \\ 0.3 & 0.8 & -0.2 & -0.4 \\ 0.7 & -0.2 & -0.6 & -0.3 \\ 0.4 & 0.3 & 0.5 & 0.7 \end{bmatrix}$$

Dimensionality Reduction (4D to 2D)

We choose the top 2 eigenvectors corresponding to the 2 largest eigenvalues:

$$Q_2 = \begin{bmatrix} 0.5 & -0.4 \\ 0.3 & 0.8 \\ 0.7 & -0.2 \\ 0.4 & 0.3 \end{bmatrix}$$

$$A_{\text{reduced}} = X \cdot Q_2$$

Let sample matrix X be:

$$X = \begin{bmatrix} 2 & 1 & 3 & 2 \\ 1 & 2 & 2 & 1 \\ 3 & 3 & 1 & 2 \end{bmatrix}$$

Then A_{reduced} is a 3×2 matrix.

Compute Reduced Matrix

$$A_{\text{reduced}} = \begin{bmatrix} 2 & 1 & 3 & 2 \\ 1 & 2 & 2 & 1 \\ 3 & 3 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0.5 & -0.4 \\ 0.3 & 0.8 \\ 0.7 & -0.2 \\ 0.4 & 0.3 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 3.9 & -0.4 \\ 3.0 & 0.6 \\ 3.8 & 1.0 \end{bmatrix}$$

- Each row is now a 2D representation of original 4D sample
- Major variance preserved by top eigenvectors

Summary

- Spectral decomposition works only for symmetric matrices
- Eigenvectors form a new feature space
- Dimensionality reduction achieved by projecting data on top eigenvectors
- Result: Fewer features with minimal information loss

Summary

Decomposition	**Input**	**Output**
SVD	Any matrix	$A = U\Sigma V^T$
Spectral	Symmetric matrix	$A = Q\Lambda Q^T$

Significance of Matrix Decompositions in ML and Data Science

- **Dimensionality Reduction:**
  - SVD and PCA reduce high-dimensional data to fewer components while preserving variance.
- **Noise Reduction and Denoising:**
  - Low-rank approximations from SVD remove noise in data.
- **Data Compression:**
  - Use top singular values to compress large datasets or images (e.g., JPEG compression).
- **Latent Semantic Analysis (LSA):**
  - SVD helps uncover latent structures in text data for topic modeling and document similarity.

Significance of Matrix Decompositions in ML and Data Science

- **Recommendation Systems:**
  - Matrix factorization (via SVD) is used in collaborative filtering (e.g., Netflix algorithm).
- **Eigenfaces in Facial Recognition:**
  - PCA is used to identify key patterns in face images for classification.
- **Feature Extraction and Engineering:**
  - Decompositions help extract dominant features from raw data.
- **Solving Linear Systems and Inversion:**
  - LU, QR, and SVD are used in efficient numerical solutions and stability.