SVD Decomposition

Shayan Shabani

February 2025

1 From Scratch: Singular Value Decomposition (SVD)

1.1 Mathematical Foundations

1.1.1 Definition

For any matrix $\mathbf{A} \in R^{m \times n}$, there exists a decomposition of the form:

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

Where: - $\mathbf{U} \in R^{m \times m}$ is an orthogonal matrix whose columns are the left singular vectors - $\mathbf{\Sigma} \in R^{m \times n}$ is a rectangular diagonal matrix containing the singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(m,n)} \geq 0$ - $\mathbf{V} \in R^{n \times n}$ is an orthogonal matrix whose columns are the right singular vectors

1.1.2 Derivation

The derivation of SVD begins by examining the squared matrices $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$. These matrices are symmetric and positive semi-definite, meaning they have non-negative eigenvalues and orthogonal eigenvectors.

Let's denote: - The eigenvectors of $\mathbf{A}^T\mathbf{A}$ as the columns of $\mathbf{V}$ - The eigenvectors of $\mathbf{A}\mathbf{A}^T$ as the columns of $\mathbf{U}$ - The eigenvalues of $\mathbf{A}^T\mathbf{A}$ (and $\mathbf{A}\mathbf{A}^T$) as λ_i

The singular values σ_i are defined as $\sigma_i = \sqrt{\lambda_i}$.

The connection between these becomes clear when we consider:

$$\mathbf{A}^T\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

For any eigenvector $\mathbf{v}_i$ with $\lambda_i > 0$, we can define:

$$\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i$$

Now we can verify that $\mathbf{u}_i$ is an eigenvector of $\mathbf{A}\mathbf{A}^T$:

$$\mathbf{A}\mathbf{A}^T\mathbf{u}_i = \mathbf{A}\mathbf{A}^T\frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i = \frac{1}{\sigma_i}\mathbf{A}(\mathbf{A}^T\mathbf{A})\mathbf{v}_i = \frac{\lambda_i}{\sigma_i}\mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i$$

1.1.3 Geometric Interpretation

SVD provides a powerful geometric interpretation of linear transformations:

1. $\mathbf{V}^T$ represents a rotation in the input space 2. $\mathbf{\Sigma}$ represents a scaling along the coordinate axes (stretching or shrinking) 3. $\mathbf{U}$ represents a rotation in the output space

This means that any linear transformation can be decomposed into a rotation, followed by a scaling, followed by another rotation. The singular values in $\mathbf{\Sigma}$ determine the amount of stretching or shrinking along each dimension.

Another interpretation is that SVD identifies the principal directions (orthogonal axes) in both the domain and range of the transformation, with the singular values indicating the "importance" or "strength" of each direction.

1.2 Complete Worked Example

: Computing the SVD

Let's compute the complete SVD decomposition of the following matrix:

$$\mathbf{A} = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$$

1.2.1 Step 1

: Compute $\mathbf{A}^T\mathbf{A}$ and its eigenvalues/eigenvectors

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 4 & 3 \\ 0 & -5 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix} = \begin{bmatrix} 25 & -15 \\ -15 & 25 \end{bmatrix}$$

To find the eigenvalues, we solve the characteristic equation:

$$\det(\mathbf{A}^T\mathbf{A} - \lambda\mathbf{I}) = 0$$

$$\det\left(\begin{bmatrix} 25-\lambda & -15 \\ -15 & 25-\lambda \end{bmatrix}\right) = 0$$

$$(25-\lambda)^2 - (-15)^2 = 0$$

$$(25-\lambda)^2 - 225 = 0$$

$$625 - 50\lambda + \lambda^2 - 225 = 0$$

$$\lambda^2 - 50\lambda + 400 = 0$$

Using the quadratic formula:

$$\lambda = \frac{50 \pm \sqrt{50^2 - 4 \times 1 \times 400}}{2 \times 1} = \frac{50 \pm \sqrt{2500 - 1600}}{2} = \frac{50 \pm \sqrt{900}}{2} = \frac{50 \pm 30}{2}$$

So, $\lambda_1 = 40$ and $\lambda_2 = 10$.

Now we find the eigenvectors for each eigenvalue:

For $\lambda_1 = 40$:

$$(\mathbf{A}^T\mathbf{A} - 40\mathbf{I})\mathbf{v}_1 = \mathbf{0}$$

$$\begin{bmatrix} 25-40 & -15 \\ -15 & 25-40 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -15 & -15 \\ -15 & -15 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This gives us $-15v_{11} - 15v_{12} = 0$, which simplifies to $v_{11} = -v_{12}$.
With the constraint that eigenvectors should have unit length, we get:

$$v_{11}^2 + v_{12}^2 = 1$$

Solving these equations:

$$v_{11} = -v_{12}$$

$$v_{11}^2 + v_{11}^2 = 1$$

$$2v_{11}^2 = 1$$

$$v_{11} = \pm\frac{1}{\sqrt{2}}$$

We'll choose $v_{11} = \frac{1}{\sqrt{2}}$ and $v_{12} = -\frac{1}{\sqrt{2}}$.

So the first eigenvector is $\mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$.

For $\lambda_2 = 10$: Using the same approach, we get $v_{21} = v_{22}$, which gives us $\mathbf{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

So our matrix $\mathbf{V}$ is:

$$\mathbf{V} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

1.2.2 Step 2: Calculate the singular values

The singular values are the square roots of the eigenvalues:

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{40} = 2\sqrt{10}$$

$$\sigma_2 = \sqrt{\lambda_2} = \sqrt{10}$$

So our matrix $\mathbf{\Sigma}$ is:

$$\mathbf{\Sigma} = \begin{bmatrix} 2\sqrt{10} & 0 \\ 0 & \sqrt{10} \end{bmatrix}$$

1.2.3 Step 3: Calculate the left singular vectors

For each $\sigma_i > 0$, we compute:

$$\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i$$

For $\mathbf{u}_1$:

$$\mathbf{u}_1 = \frac{1}{2\sqrt{10}}\mathbf{A}\mathbf{v}_1 = \frac{1}{2\sqrt{10}}\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\mathbf{u}_1 = \frac{1}{2\sqrt{10}}\begin{bmatrix} \frac{4}{\sqrt{2}} \\ \frac{3}{\sqrt{2}} + \frac{5}{\sqrt{2}} \end{bmatrix} = \frac{1}{2\sqrt{10}}\begin{bmatrix} \frac{4}{\sqrt{2}} \\ \frac{8}{\sqrt{2}} \end{bmatrix}$$

$$\mathbf{u}_1 = \frac{1}{2\sqrt{10}}\begin{bmatrix} \frac{4\sqrt{2}}{2} \\ \frac{8\sqrt{2}}{2} \end{bmatrix} = \frac{1}{2\sqrt{10}}\begin{bmatrix} 2\sqrt{2} \\ 4\sqrt{2} \end{bmatrix}$$

$$\mathbf{u}_1 = \begin{bmatrix} \frac{2\sqrt{2}}{2\sqrt{10}} \\ \frac{4\sqrt{2}}{2\sqrt{10}} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{10}} \\ \frac{2\sqrt{2}}{\sqrt{10}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$$

For $\mathbf{u}_2$:

$$\mathbf{u}_2 = \frac{1}{\sqrt{10}}\mathbf{A}\mathbf{v}_2 = \frac{1}{\sqrt{10}}\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sqrt{10}}\begin{bmatrix} \frac{4}{\sqrt{2}} \\ \frac{3}{\sqrt{2}} - \frac{5}{\sqrt{2}} \end{bmatrix} = \frac{1}{\sqrt{10}}\begin{bmatrix} \frac{4}{\sqrt{2}} \\ -\frac{2}{\sqrt{2}} \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sqrt{10}}\begin{bmatrix} \frac{4\sqrt{2}}{2} \\ -\frac{2\sqrt{2}}{2} \end{bmatrix} = \frac{1}{\sqrt{10}}\begin{bmatrix} 2\sqrt{2} \\ -\sqrt{2} \end{bmatrix}$$

$$\mathbf{u}_2 = \begin{bmatrix} \frac{2\sqrt{2}}{\sqrt{10}} \\ -\frac{\sqrt{2}}{\sqrt{10}} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{bmatrix}$$

We can verify that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal:

$$\mathbf{u}_1^T\mathbf{u}_2 = \frac{1}{\sqrt{5}} \cdot \frac{2}{\sqrt{5}} + \frac{2}{\sqrt{5}} \cdot (-\frac{1}{\sqrt{5}}) = \frac{2}{5} - \frac{2}{5} = 0$$

And they're both unit vectors:

$$\|\mathbf{u}_1\|^2 = \left(\frac{1}{\sqrt{5}}\right)^2 + \left(\frac{2}{\sqrt{5}}\right)^2 = \frac{1}{5} + \frac{4}{5} = 1$$

$$\|\mathbf{u}_2\|^2 = \left(\frac{2}{\sqrt{5}}\right)^2 + \left(-\frac{1}{\sqrt{5}}\right)^2 = \frac{4}{5} + \frac{1}{5} = 1$$

So our matrix $\mathbf{U}$ is:

$$\mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix}$$

1.2.4 Step 4: Verify the decomposition

We have:

$$\mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix}, \ \mathbf{\Sigma} = \begin{bmatrix} 2\sqrt{10} & 0 \\ 0 & \sqrt{10} \end{bmatrix}, \ \mathbf{V} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Let's compute $\mathbf{U\Sigma V}^T$ to see if we get $\mathbf{A}$:

$$\mathbf{U\Sigma V}^T = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 2\sqrt{10} & 0 \\ 0 & \sqrt{10} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

First, let's compute $\mathbf{U\Sigma}$:

$$\mathbf{U\Sigma} = \begin{bmatrix} \frac{1}{\sqrt{5}} \cdot 2\sqrt{10} & \frac{2}{\sqrt{5}} \cdot \sqrt{10} \\ \frac{2}{\sqrt{5}} \cdot 2\sqrt{10} & -\frac{1}{\sqrt{5}} \cdot \sqrt{10} \end{bmatrix} = \begin{bmatrix} \frac{2\sqrt{10}}{\sqrt{5}} & \frac{2\sqrt{10}}{\sqrt{5}} \\ \frac{4\sqrt{10}}{\sqrt{5}} & -\frac{\sqrt{10}}{\sqrt{5}} \end{bmatrix}$$

$$\mathbf{U\Sigma} = \begin{bmatrix} \frac{2\sqrt{2}}{\sqrt{1}} & \frac{2\sqrt{2}}{\sqrt{1}} \\ \frac{4\sqrt{2}}{\sqrt{1}} & -\frac{\sqrt{2}}{\sqrt{1}} \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} & 2\sqrt{2} \\ 4\sqrt{2} & -\sqrt{2} \end{bmatrix}$$

Now, let's compute $\mathbf{U\Sigma V}^T$:

$$\mathbf{U\Sigma V}^T = \begin{bmatrix} 2\sqrt{2} & 2\sqrt{2} \\ 4\sqrt{2} & -\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\mathbf{U\Sigma V}^T = \begin{bmatrix} 2\sqrt{2} \cdot \frac{1}{\sqrt{2}} + 2\sqrt{2} \cdot \frac{1}{\sqrt{2}} & 2\sqrt{2} \cdot (-\frac{1}{\sqrt{2}}) + 2\sqrt{2} \cdot \frac{1}{\sqrt{2}} \\ 4\sqrt{2} \cdot \frac{1}{\sqrt{2}} + (-\sqrt{2}) \cdot \frac{1}{\sqrt{2}} & 4\sqrt{2} \cdot (-\frac{1}{\sqrt{2}}) + (-\sqrt{2}) \cdot \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\mathbf{U\Sigma V}^T = \begin{bmatrix} 2+2 & -2+2 \\ 4-1 & -4-1 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix} = \mathbf{A}$$

The decomposition is verified!

1.3 Interpretation of the Example

We've successfully decomposed $\mathbf{A} = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$ into its SVD components:

$$\mathbf{A} = \mathbf{U\Sigma V}^T = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 2\sqrt{10} & 0 \\ 0 & \sqrt{10} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

The singular values $\sigma_1 = 2\sqrt{10} \approx 6.32$ and $\sigma_2 = \sqrt{10} \approx 3.16$ tell us about the "importance" of each dimension. The fact that $\sigma_1 > \sigma_2$ means that the transformation $\mathbf{A}$ stretches vectors more along the first principal direction (given by the first column of $\mathbf{V}$) than along the second.

The right singular vectors (columns of $\mathbf{V}$) show the principal directions in the input space. In our example, these directions are $\mathbf{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$, which are at 45° angles to the standard basis.

The left singular vectors (columns of $\mathbf{U}$) show the principal directions in the output space. These are $\mathbf{u}_1 = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{bmatrix}$.

The geometric interpretation of this transformation is: 1. A rotation by 45° in the input space (matrix $\mathbf{V}^T$) 2. A scaling by factors of $2\sqrt{10}$ and $\sqrt{10}$ along the new axes (matrix $\mathbf{\Sigma}$) 3. A rotation in the output space to align with the standard basis (matrix $\mathbf{U}$)

In the context of data analysis, if $\mathbf{A}$ were a data matrix, the SVD would tell us that the data has two principal components, with the first component about twice as significant as the second.

1.4 Connection to Applications

In a recommender system, if $\mathbf{A}$ were a user-item matrix: 1. The right singular vectors would represent latent item features 2. The left singular vectors would represent latent user preferences 3. The singular values would indicate the importance of each latent feature

For a social network adjacency matrix, the SVD would reveal: 1. Community structures (the singular vectors) 2. The relative strengths of these communities (the singular values) 3. How users are positioned within these communities (the components of the singular vectors)

The computational approach demonstrated here can be scaled to larger matrices using iterative methods such as the power iteration or Lanczos algorithm, which are much more efficient for large, sparse matrices commonly found in real-world applications.

2 Advanced SVD and Matrix Norms: Notes and Problems

2.1 Graduate Course in Social Networks Analysis

—

2.2 Part I: Vector and Matrix Norms

2.3 Theoretical Background

Vector norms quantify the "size" of vectors, while matrix norms measure the "magnitude" of matrices. These are critical in analyzing social network data, particularly when working with user-item interaction matrices or adjacency matrices.

2.3.1 Vector Norms

For a vector $\mathbf{x} \in R^n$:

1. **L_p Norm**: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ - L_1 Norm (Manhattan): $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$ - L_2 Norm (Euclidean): $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ - L_∞ Norm (Chebyshev): $\|\mathbf{x}\|_\infty = \max_i |x_i|$

2. **Properties**: - Positivity: $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$ - Homogeneity: $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ - Triangle Inequality: $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

2.3.2 Matrix Norms

For a matrix $\mathbf{A} \in R^{m\times n}$:

1. **Frobenius Norm**: $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\text{Tr}(\mathbf{A}^T\mathbf{A})}$
2. **Spectral Norm**: $\|\mathbf{A}\|_2 = \sigma_{\max}(\mathbf{A}) = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})}$
3. **Nuclear Norm**: $\|\mathbf{A}\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i(\mathbf{A})$
4. **Induced p-Norm**: $\|\mathbf{A}\|_p = \sup_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$

2.4 Applications in Social Network Analysis

- Centrality measures in networks (related to vector norms) - Network distances and similarities (matrix norms) - User similarity measures in recommender systems - Regularization in matrix factorization models

2.5 Problem 1

: Norm Properties in Social Networks

Problem: Consider a social network with 5 users and the following influence vector $\mathbf{v} = [0.8, 0.3, 0.5, 0.9, 0.1]$, representing each user's influence score.

a) Calculate $\|\mathbf{v}\|_1$, $\|\mathbf{v}\|_2$, and $\|\mathbf{v}\|_\infty$.

b) If we have two competing influence vectors $\mathbf{v}_1 = [0.8, 0.3, 0.5, 0.9, 0.1]$ and $\mathbf{v}_2 = [0.7, 0.7, 0.7, 0.4, 0.4]$, which vector would you consider more significant based on different norms? Interpret the meaning in a social network context.

Solution:

a) Calculating the norms: - $\|\mathbf{v}\|_1 = 0.8+0.3+0.5+0.9+0.1 = 2.6$ - $\|\mathbf{v}\|_2 = \sqrt{0.8^2+0.3^2+0.5^2+0.9^2+0.1^2} = \sqrt{1.8} \approx 1.34$ - $\|\mathbf{v}\|_\infty = \max(0.8, 0.3, 0.5, 0.9, 0.1) = 0.9$

b) Comparing $\mathbf{v}_1$ and $\mathbf{v}_2$: - $\|\mathbf{v}_1\|_1 = 2.6$ and $\|\mathbf{v}_2\|_1 = 2.9$ - $\|\mathbf{v}_1\|_2 \approx 1.34$ and $\|\mathbf{v}_2\|_2 \approx 1.39$ - $\|\mathbf{v}_1\|_\infty = 0.9$ and $\|\mathbf{v}_2\|_\infty = 0.7$

Interpretation: - $\mathbf{v}_2$ has larger L_1 and L_2 norms, indicating higher total and average influence across the network - $\mathbf{v}_1$ has a larger L_∞ norm, indicating the presence of a single highly influential user - In social networks, $\mathbf{v}_1$ represents a network with concentrated influence (potential influencers), while $\mathbf{v}_2$ represents more evenly distributed influence

2.6 Problem 2

: Matrix Norms in Interaction Data

Problem: Consider a user-item interaction matrix in a social recommendation system:

$$\mathbf{R} = \begin{bmatrix} 5 & 3 & 0 & 1 \\ 4 & 0 & 0 & 1 \\ 1 & 1 & 0 & 5 \\ 0 & 1 & 5 & 4 \end{bmatrix}$$

a) Calculate the Frobenius norm of this matrix.

b) If we know the singular values of $\mathbf{R}$ are $\sigma_1 = 8.82$, $\sigma_2 = 5.29$, $\sigma_3 = 2.24$, and $\sigma_4 = 0.89$, calculate the spectral norm and nuclear norm.

c) How do these different norms help us understand the complexity of the interaction data?

Solution:

a) Frobenius norm: $\|\mathbf{R}\|_F = \sqrt{5^2 + 3^2 + 0^2 + 1^2 + 4^2 + 0^2 + 0^2 + 1^2 + 1^2 + 1^2 + 0^2 + 5^2 + 0^2 + 1^2 + 5^2 + 4^2}$ $\|\mathbf{R}\|_F = \sqrt{25 + 9 + 0 + 1 + 16 + 0 + 0 + 1 + 1 + 1 + 0 + 25 + 0 + 1 + 25 + 16}$ $\|\mathbf{R}\|_F = \sqrt{121} = 11$

b) Using the singular values: - Spectral norm: $\|\mathbf{R}\|_2 = \sigma_1 = 8.82$ - Nuclear norm: $\|\mathbf{R}\|_* = \sigma_1 + \sigma_2 + \sigma_3 + \sigma_4 = 8.82 + 5.29 + 2.24 + 0.89 = 17.24$

c) Interpretation: - The Frobenius norm ($\|\mathbf{R}\|_F = 11$) measures the overall energy or magnitude of interactions in the system. - The spectral norm ($\|\mathbf{R}\|_2 = 8.82$) indicates the maximum amplification the matrix can apply to a unit vector, revealing the strength of the dominant pattern in the data. - The nuclear norm ($\|\mathbf{R}\|_* = 17.24$) is related to the "effective rank" of the matrix, suggesting how complex the underlying interaction patterns are. - The large gap between σ_1 and σ_4 indicates that a low-rank approximation may effectively capture the most important patterns in this recommendation data.

—

3 Part II

: Singular Value Decomposition (SVD)

3.1 Mathematical Definition

For any matrix $\mathbf{A} \in R^{m \times n}$, the Singular Value Decomposition is:

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

Where: - $\mathbf{U} \in R^{m \times m}$ is an orthogonal matrix whose columns are the left singular vectors - $\mathbf{\Sigma} \in R^{m \times n}$ is a diagonal matrix containing singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(m,n)} \geq 0$ - $\mathbf{V} \in R^{n \times n}$ is an orthogonal matrix whose columns are the right singular vectors

3.2 Key Properties

1. **Rank**: rank($\mathbf{A}$) = number of non-zero singular values 2. **Pseudo-inverse**: $\mathbf{A}^+ = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^T$ where $\mathbf{\Sigma}^+$ inverts non-zero singular values 3. **Low-rank approximation**: The best rank-k approximation to $\mathbf{A}$ is $\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ 4. **Eckart-Young Theorem**: $\|\mathbf{A} - \mathbf{A}_k\|_F = \sqrt{\sum_{i=k+1}^{\min(m,n)} \sigma_i^2}$ 5. **Nuclear norm**: $\|\mathbf{A}\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i$

3.3 Applications in Recommender Systems

1. **Matrix Factorization**: Using truncated SVD to factorize user-item matrices 2. **Latent Factor Models**: Interpreting $\mathbf{U}$ and $\mathbf{V}$ as latent user and item factors 3. **Collaborative Filtering**: Predicting missing entries in sparse interaction matrices 4. **Cold Start Problem**: Handling new users/items using SVD-based approaches

3.4 Problem 3

: SVD and Recommender Systems

Problem: Consider a user-item interaction matrix from a movie recommendation platform:

$$\mathbf{R} = \begin{bmatrix} 5 & 4 & ? & 4 & ? \\ ? & 5 & 4 & 3 & ? \\ 2 & ? & 1 & ? & 3 \\ 4 & 3 & ? & 3 & 5 \end{bmatrix}$$

where ? represents missing ratings.

a) Given the rank-2 decomposition from SVD:

$$\mathbf{U}_2 = \begin{bmatrix} 0.56 & 0.31 \\ 0.52 & 0.13 \\ 0.21 & -0.72 \\ 0.61 & 0.28 \end{bmatrix}, \mathbf{\Sigma}_2 = \begin{bmatrix} 9.82 & 0 \\ 0 & 3.94 \end{bmatrix}, \mathbf{V}_2^T = \begin{bmatrix} 0.45 & 0.38 & 0.29 & 0.32 & 0.42 \\ 0.31 & 0.24 & -0.61 & 0.18 & 0.56 \end{bmatrix}$$

Predict the missing rating for user 2 on item 5.

b) Calculate the Frobenius norm of the approximation error if we use this rank-2 approximation, given that the original matrix has singular values $\sigma_1 = 9.82$, $\sigma_2 = 3.94$, $\sigma_3 = 2.11$, $\sigma_4 = 0.87$.

Solution:

a) The rank-2 approximation is given by:

$$\hat{\mathbf{R}}_2 = \mathbf{U}_2\mathbf{\Sigma}_2\mathbf{V}_2^T$$

To predict user 2 (index 1) rating for item 5 (index 4), we compute:

$$\hat{r}_{1,4} = [\mathbf{U}_2]_{1,:}\mathbf{\Sigma}_2[\mathbf{V}_2^T]_{:,4}$$

$$\hat{r}_{1,4} = [0.52, 0.13] \begin{bmatrix} 9.82 & 0 \\ 0 & 3.94 \end{bmatrix} \begin{bmatrix} 0.42 \\ 0.56 \end{bmatrix}$$

$$\hat{r}_{1,4} = [5.11, 0.51] \begin{bmatrix} 0.42 \\ 0.56 \end{bmatrix} = 5.11 \times 0.42 + 0.51 \times 0.56 = 2.15 + 0.29 = 2.44$$

Therefore, the predicted rating for user 2 on item 5 is approximately 2.44, which would round to 2.5.

b) Using the Eckart-Young theorem, the Frobenius norm of the approximation error is:

$$\|\mathbf{R} - \hat{\mathbf{R}}_2\|_F = \sqrt{\sum_{i=3}^{4} \sigma_i^2} = \sqrt{2.11^2 + 0.87^2} = \sqrt{4.45 + 0.76} = \sqrt{5.21} \approx 2.28$$

This means that our rank-2 approximation has an average error of 2.28 across all elements of the matrix, which is relatively high for a 1-5 rating scale. We might consider using a rank-3 approximation for better accuracy.

3.5 Problem 4

: Advanced SVD Applications in Social Networks

Problem: In a social network with n users, the adjacency matrix $\mathbf{A}$ represents connections, where $A_{ij} = 1$ if users i and j are connected and 0 otherwise.

For a weighted network with the following adjacency matrix:

$$\mathbf{A} = \begin{bmatrix} 0 & 0.9 & 0.3 & 0.5 & 0.2 \\ 0.9 & 0 & 0.8 & 0.1 & 0 \\ 0.3 & 0.8 & 0 & 0.7 & 0.4 \\ 0.5 & 0.1 & 0.7 & 0 & 0.6 \\ 0.2 & 0 & 0.4 & 0.6 & 0 \end{bmatrix}$$

a) Given that SVD decomposes $\mathbf{A} = \mathbf{U\Sigma V}^T$ with singular values $\sigma_1 = 2.37$, $\sigma_2 = 0.92$, $\sigma_3 = 0.64$, $\sigma_4 = 0.43$, $\sigma_5 = 0.18$, how much of the network's variance is captured by the first two singular values?

b) In this social network context, interpret what the right singular vectors (columns of $\mathbf{V}$) represent.

c) If the first right singular vector is $\mathbf{v}_1 = [0.44, 0.43, 0.52, 0.47, 0.35]^T$, which users are most central in the dominant community structure?

Solution:

a) The total variance in the data is the sum of squared singular values:

Total variance $= \sum_{i=1}^{5} \sigma_i^2 = 2.37^2 + 0.92^2 + 0.64^2 + 0.43^2 + 0.18^2 = 5.62 + 0.85 + 0.41 + 0.18 + 0.03 = 7.09$

Variance captured by first two singular values $= \frac{\sigma_1^2+\sigma_2^2}{\sum_{i=1}^5 \sigma_i^2} = \frac{5.62+0.85}{7.09} = \frac{6.47}{7.09} \approx 0.912$ or 91.2

The first two singular values capture approximately 91.2

b) In a social network context, the right singular vectors (columns of $\mathbf{V}$) represent: - Community structures or clusters within the network - The coefficients in each vector indicate how strongly each user belongs to that particular community - The first right singular vector typically identifies the dominant community structure - Subsequent vectors identify increasingly refined community subdivisions - These can be interpreted as latent "social roles" or positions in the network

c) Looking at the first right singular vector $\mathbf{v}_1 = [0.44, 0.43, 0.52, 0.47, 0.35]^T$: - User 3 (index 2) has the highest coefficient (0.52), making them the most central in the dominant community - User 4 (index 3) has the second-highest coefficient (0.47) - User 1 (index 0) is third with 0.44

These users form the core of the main community structure in the network. User 3's high centrality means they are likely a key connector or influencer within the dominant community structure.

—

4 Part III

: Connecting SVD and Norms in Applications

4.1 Matrix Completion and Low-Rank Approximation

In many real-world applications like recommender systems and social network analysis, we often deal with incomplete data matrices. The mathematical formulation for matrix completion is:

$$\min_{\mathbf{X}} \operatorname{rank}(\mathbf{X}) \quad \text{subject to} \quad X_{ij} = M_{ij} \quad \forall (i,j) \in \Omega$$

where Ω is the set of observed entries. Since this problem is NP-hard, we use a convex relaxation:

$$\min_{\mathbf{X}} \|\mathbf{X}\|_* \quad \text{subject to} \quad X_{ij} = M_{ij} \quad \forall (i,j) \in \Omega$$

This is the foundation of many modern recommender systems.

4.2 Theoretical Connections

1. **Nuclear Norm as Convex Envelope**: The nuclear norm $\|\mathbf{X}\|_*$ is the convex envelope of rank(X) on the unit ball of matrices with spectral norm 1.

2. **Relationship to SVD**: The nuclear norm is the sum of singular values, which connects directly to the SVD.

3. **Regularization Techniques**: L_1 and L_2 regularization in matrix factorization models correspond to constraining different matrix norms.

4. **Recommender Systems Optimization**: Modern approaches often solve:

$$\min_{\mathbf{X}} \|\mathcal{P}_\Omega(\mathbf{X} - \mathbf{M})\|_F^2 + \lambda\|\mathbf{X}\|_*$$

where $\mathcal{P}_\Omega$ is a projection operator that keeps observed entries.

4.3 Problem 5

: Matrix Completion for Recommendation

Problem: Consider a partially observed user-item rating matrix:

$$\mathbf{M} = \begin{bmatrix} 5 & ? & 2 & ? \\ ? & 4 & ? & 1 \\ 3 & ? & ? & 4 \\ ? & 2 & 3 & ? \end{bmatrix}$$

a) Given that you want to complete this matrix using a nuclear norm minimization approach, write the formal optimization problem.

b) If we approximate the nuclear norm by decomposing $\mathbf{X} = \mathbf{U}\mathbf{V}^T$ where $\mathbf{U} \in R^{4\times k}$ and $\mathbf{V} \in R^{4\times k}$, what is the corresponding optimization problem for a rank-2 approximation?

c) If the resulting completed matrix has a nuclear norm of 15.3 and a Frobenius norm of 11.2, what does this indicate about the complexity of the recommendation patterns?

Solution:

a) The formal nuclear norm minimization problem is:

$$\min_{\mathbf{X}} \|\mathbf{X}\|_* \quad \text{subject to} \quad X_{ij} = M_{ij} \quad \forall (i,j) \in \Omega$$

where $\Omega = \{(1,1),(1,3),(2,2),(2,4),(3,1),(3,4),(4,2),(4,3)\}$ is the set of observed entries.

In practice, we often solve a relaxed version:

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{X} - \mathbf{M})\|_F^2 + \lambda\|\mathbf{X}\|_*$$

where $\lambda > 0$ is a regularization parameter.

b) For a rank-2 approximation using $\mathbf{X} = \mathbf{U}\mathbf{V}^T$, the optimization problem becomes:

$$\min_{\mathbf{U},\mathbf{V}} \frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{U}\mathbf{V}^T - \mathbf{M})\|_F^2 + \frac{\lambda}{2}(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2)$$

where $\mathbf{U} \in R^{4\times 2}$ and $\mathbf{V} \in R^{4\times 2}$.

This is the standard matrix factorization approach used in many recommender systems, where the columns of $\mathbf{U}$ represent latent user factors and the columns of $\mathbf{V}$ represent latent item factors.

c) The ratio of the nuclear norm to the Frobenius norm provides information about the effective rank:

$$\frac{\|\mathbf{X}\|_*}{\|\mathbf{X}\|_F} = \frac{15.3}{11.2} \approx 1.37$$

For a rank-1 matrix, this ratio would be exactly 1. The ratio of 1.37 suggests that while the matrix has a full rank of 4, its "effective rank" is relatively low (likely between 1 and 2).

This indicates that the recommendation patterns are relatively simple and can be well-approximated by a low-rank model. The user-item interactions likely follow a few dominant patterns, which is good news for a recommender system as it means we can effectively predict missing ratings with a simple model.

—

5 Conclusion

This class has covered advanced concepts in SVD and matrix norms with a focus on applications in social network analysis and recommender systems. The key takeaways are:

1. Vector and matrix norms provide important measures of magnitude in different contexts, with each norm highlighting different aspects of the data.

2. SVD offers a powerful framework for understanding the structure of matrices, enabling dimensionality reduction, noise filtering, and pattern discovery.

3. The connection between SVD and norms, particularly the nuclear norm, provides theoretical foundations for modern matrix completion techniques used in recommender systems.

4. In social network analysis, these tools help identify community structures, influential users, and latent interaction patterns.

The problems presented demonstrate how these mathematical concepts translate to practical applications in analyzing and predicting behavior in social networks and recommendation platforms.