PMAI LAB

Linear Algebra: Theory, Computation, and Modern Applications

Lecture 09 • Singular Value Decomposition,
PCA, and Data Compression

Course lens
Understand → Model → Compute → Apply → Code → Interpret

Audience
Undergraduate BS Mathematics students

Instructor
Dr. Hira Benish

SVD separates a matrix into meaningful directions, their strengths, and a change of coordinates.

1. Why Singular Value Decomposition?

Extend eigenvector geometry to rectangular, singular, and noisy matrices

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

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

The columns of $\mathbf{V}$ are right singular vectors, the columns of $\mathbf{U}$ are left singular vectors, and

$$\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots), \qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq 0.$$

Geometric interpretation

The transformation $\boldsymbol{x} \mapsto \mathbf{A}\boldsymbol{x}$ can be understood in three stages:

$$\boldsymbol{x} \xrightarrow{\mathbf{V}^{\mathsf{T}}} \text{new input coordinates} \xrightarrow{\Sigma} \text{axis-by-axis scaling} \xrightarrow{\mathbf{U}} \text{output orientation}.$$

Thus SVD separates orientation from stretching and compression.

Learning outcomes

By the end of this lecture, students should be able to:

- state full and reduced forms of the SVD;
- obtain singular values from $\mathbf{A}^{\mathsf{T}}\mathbf{A}$;
- interpret singular vectors and singular values geometrically;
- calculate a simple SVD and verify reconstruction;
- form a truncated SVD and quantify approximation error;
- explain rank, condition number, pseudoinverse, and PCA through SVD; and
- implement image compression and PCA in Python.

Central idea
SVD orders the information carried by a matrix. Large singular values identify strong structure; small singular values often represent weak structure, redundancy, or noise.

Where it appears

Image and audio compression, denoising, recommender systems, latent semantic analysis, PCA, inverse problems, numerical least squares, computer vision, signal processing, scientific computing, and machine learning.

2. Structure and Computation

Connect singular vectors to symmetric eigenvalue problems

The matrices $\mathbf{A}^\top\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$ are symmetric and positive semidefinite:

$$\mathbf{A}^\top\mathbf{A}\boldsymbol{v}_i = \sigma_i^2\boldsymbol{v}_i, \qquad \mathbf{A}\mathbf{A}^\top\boldsymbol{u}_i = \sigma_i^2\boldsymbol{u}_i.$$

Therefore,

$$\boxed{\sigma_i = \sqrt{\lambda_i(\mathbf{A}^\top\mathbf{A})}.}$$

For each nonzero singular value,

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

Full and reduced SVD

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

$$\mathbf{U} \in \mathbb{R}^{m\times m}, \quad \mathbf{\Sigma} \in \mathbb{R}^{m\times n}, \quad \mathbf{V} \in \mathbb{R}^{n\times n}.$$

If $\operatorname{rank}(\mathbf{A}) = r$, the reduced form is

$$\mathbf{A} = \mathbf{U}_r\mathbf{\Sigma}_r\mathbf{V}_r^\top,$$

where $\mathbf{U}_r \in \mathbb{R}^{m\times r}$, $\mathbf{\Sigma}_r \in \mathbb{R}^{r\times r}$, and $\mathbf{V}_r \in \mathbb{R}^{n\times r}$.

Outer-product form

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top.$$

Each term is a rank-one matrix. SVD builds $\mathbf{A}$ from independent information layers ordered by strength.

Immediate information from singular values

$$\operatorname{rank}(\mathbf{A}) = \#\{i : \sigma_i > 0\}, \qquad \|\mathbf{A}\|_2 = \sigma_1, \qquad \|\mathbf{A}\|_F^2 = \sum_i \sigma_i^2.$$

For full-rank problems, the 2-norm condition number is

$$\kappa_2(\mathbf{A}) = \frac{\sigma_{\max}}{\sigma_{\min}}.$$

Computation workflow
Step 1: Find eigenpairs of $\mathbf{A}^\top\mathbf{A}$. **Step 2:** Set $\sigma_i = \sqrt{\lambda_i}$. **Step 3:** obtain $\boldsymbol{u}_i = \mathbf{A}\boldsymbol{v}_i/\sigma_i$. **Step 4:** verify $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$.

3. Worked SVD Computation

Calculate a complete SVD for a small symmetric matrix

Let

$$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

First compute

$$\mathbf{A}^\mathsf{T}\mathbf{A} = \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix}.$$

Its eigenvalues are 16 and 4, so

$$\sigma_1 = 4, \qquad \sigma_2 = 2.$$

Corresponding orthonormal right singular vectors are

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \boldsymbol{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Find the left singular vectors

$$\boldsymbol{u}_1 = \frac{\mathbf{A}\boldsymbol{v}_1}{4} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \boldsymbol{u}_2 = \frac{\mathbf{A}\boldsymbol{v}_2}{2} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Thus

$$\mathbf{U} = \mathbf{V} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}.$$

Reconstruction

$$\mathbf{U}\Sigma\mathbf{V}^\mathsf{T} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

> **Interpretation**
> Along $(1,1)^\mathsf{T}$ the matrix stretches by 4; along $(1,-1)^\mathsf{T}$ it stretches by $2.$ Because this matrix is symmetric positive definite, its singular vectors agree with its eigenvectors.

Checks

$$\operatorname{rank}(\mathbf{A}) = 2, \qquad \|\mathbf{A}\|_F = \sqrt{4^2 + 2^2} = \sqrt{20}, \qquad \kappa_2(\mathbf{A}) = 4/2 = 2.$$

4. Low-Rank Approximation and Pseudoinverse

Keep the strongest information layers and control the approximation error

The rank-k truncated SVD is

$$\boxed{\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathsf{T}}}, \qquad k < r.$$

The Eckart–Young theorem states that $\mathbf{A}_k$ is the best rank-k approximation in both the spectral and Frobenius norms:

$$\|\mathbf{A} - \mathbf{A}_k\|_2 = \sigma_{k+1}, \qquad \|\mathbf{A} - \mathbf{A}_k\|_F = \sqrt{\sum_{i=k+1}^{r} \sigma_i^2}.$$

Rank-one approximation of the worked matrix

$$\mathbf{A}_1 = 4\boldsymbol{u}_1\boldsymbol{v}_1^{\mathsf{T}} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}.$$

Then

$$\|\mathbf{A} - \mathbf{A}_1\|_F = 2, \qquad \text{energy retained} = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} = \frac{16}{20} = 80\%.$$

Storage model

An $m \times n$ matrix stores mn numbers. A rank-k SVD representation stores approximately

$$k(m + n + 1)$$

numbers: mk for $\mathbf{U}_k$, k singular values, and nk for $\mathbf{V}_k$.

Moore–Penrose pseudoinverse

Invert only the nonzero singular values:

$$\mathbf{A}^{+} = \mathbf{V}\boldsymbol{\Sigma}^{+}\mathbf{U}^{\mathsf{T}}.$$

For the worked matrix,

$$\mathbf{A}^{+} = \mathbf{V} \begin{bmatrix} 1/4 & 0 \\ 0 & 1/2 \end{bmatrix} \mathbf{U}^{\mathsf{T}} = \begin{bmatrix} 3/8 & -1/8 \\ -1/8 & 3/8 \end{bmatrix}.$$

The pseudoinverse provides minimum-norm least-squares solutions:

$$\widehat{\boldsymbol{x}} = \mathbf{A}^{+}\boldsymbol{b}.$$

Numerical judgement
A very small σ_i makes inversion sensitive because $1/\sigma_i$ becomes large. Truncation or regularization can improve stability by suppressing weak directions.

5. Principal Component Analysis through SVD

Find directions of maximum variation in centred multivariable data

Suppose rows of $\mathbf{X} \in \mathbb{R}^{n\times p}$ are observations and columns are variables. First centre each column:

$$\mathbf{X}_c = \mathbf{X} - \text{column means.}$$

Then compute

$$\mathbf{X}_c = \mathbf{U}\Sigma\mathbf{V}^{\mathsf{T}}.$$

The columns of $\mathbf{V}$ are the principal directions, and the score matrix is

$$\mathbf{T} = \mathbf{X}_c\mathbf{V} = \mathbf{U}\Sigma.$$

The variance along component j is

$$\frac{\sigma_j^2}{n-1},$$

and its explained-variance ratio is

$$\frac{\sigma_j^2}{\sum_i \sigma_i^2}.$$

Small data illustration

For observations

$$\mathbf{X} = \begin{bmatrix} 2 & 1 \\ 3 & 2 \\ 4 & 2 \\ 5 & 4 \end{bmatrix}, \qquad \overline{\boldsymbol{x}} = \begin{pmatrix} 3.5 \\ 2.25 \end{pmatrix},$$

the centred matrix is

$$\mathbf{X}_c = \begin{bmatrix} -1.5 & -1.25 \\ -0.5 & -0.25 \\ 0.5 & -0.25 \\ 1.5 & 1.75 \end{bmatrix}.$$

Python gives

$$\sigma_1 \approx 3.0621, \qquad \sigma_2 \approx 0.6110.$$

Therefore,

$$\text{PC1 explained variance} \approx 96.17\%, \qquad \text{PC2} \approx 3.83\%.$$

The first principal direction is approximately

$$\boldsymbol{v}_1 = \begin{pmatrix} 0.7169 \\ 0.6972 \end{pmatrix}.$$

> **Interpretation**
> Both original variables increase together, so most variation lies near a single diagonal direction. Replacing two variables by the first component retains about 96% of the variation.

Important modelling choices

Centre before PCA. Standardize as well when variables use very different units or scales. PCA finds variance, not necessarily causality, class separation, or the most meaningful domain feature.

6. Application Case Study: Grayscale Image Compression

Use truncated SVD to preserve dominant image structure with fewer numbers

A grayscale image can be represented by a matrix whose entries are pixel intensities. Consider

$$\mathbf{A} = \begin{bmatrix} 9 & 9 & 2 & 2 \\ 9 & 9 & 2 & 2 \\ 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \end{bmatrix}.$$

The two large upper-left blocks and repeated rows create strong low-rank structure.

Singular spectrum

The nonzero singular values are approximately

$$\sigma_1 = 18.9443, \qquad \sigma_2 = 1.0557.$$

The matrix has rank 2, but the first component retains

$$\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \approx 0.9969 = 99.69\%$$

of its squared Frobenius energy.

Rank-one compressed image

$$\mathbf{A}_1 \approx \begin{bmatrix} 8.972 & 8.972 & 2.118 & 2.118 \\ 8.972 & 8.972 & 2.118 & 2.118 \\ 2.118 & 2.118 & 0.500 & 0.500 \\ 2.118 & 2.118 & 0.500 & 0.500 \end{bmatrix}.$$

Its reconstruction error is

$$\|\mathbf{A} - \mathbf{A}_1\|_F = \sigma_2 \approx 1.0557.$$

Compression count

The original stores $4 \times 4 = 16$ values. The rank-one representation stores

$$1(4 + 4 + 1) = 9$$

values, a reduction of

$$\frac{16 - 9}{16} \times 100\% = 43.75\%.$$

Application decision
Rank 1 is appropriate when a 99% energy threshold is acceptable. It greatly preserves the dominant bright and dark regions while reducing storage.

Responsible limitation
For tiny matrices, bookkeeping may reduce practical savings. Real image compression also stores metadata and quantized values; visual quality must be checked because energy retention does not capture every perceptual detail.

7. Python Laboratory

Compress a matrix, inspect the singular spectrum, and perform PCA

```
import numpy as np
import matplotlib.pyplot as plt

# SVD image-compression example
A = np.array([[9., 9., 2., 2.],
              [9., 9., 2., 2.],
              [2., 2., 1., 1.],
              [2., 2., 1., 1.]])

U, s, Vt = np.linalg.svd(A, full_matrices=False)
k = 1
A_k = (U[:, :k] * s[:k]) @ Vt[:k, :]

energy = np.cumsum(s**2) / np.sum(s**2)
error = np.linalg.norm(A - A_k, ord="fro")
stored_original = A.size
stored_svd = k * (A.shape[0] + A.shape[1] + 1)

print("Singular values:", np.round(s, 4))
print("Rank-1 approximation:\n", np.round(A_k, 3))
print("Energy retained:", round(float(energy[k - 1]), 6))
print("Frobenius error:", round(float(error), 4))
print("Stored values:", stored_original, "->", stored_svd)

# PCA example
X = np.array([[2., 1.], [3., 2.], [4., 2.], [5., 4.]])
X_centered = X - X.mean(axis=0)
Up, sp, Vtp = np.linalg.svd(X_centered, full_matrices=False)
scores = X_centered @ Vtp.T
explained_ratio = sp**2 / np.sum(sp**2)

print("PCA singular values:", np.round(sp, 4))
print("Principal directions:\n", np.round(Vtp.T, 4))
print("Explained-variance ratio:", np.round(explained_ratio, 4))
print("First-component scores:", np.round(scores[:, 0], 4))

fig, axes = plt.subplots(1, 3, figsize=(10, 3.2))
axes[0].imshow(A, cmap="gray", vmin=0, vmax=9)
axes[0].set_title("Original")
axes[1].imshow(A_k, cmap="gray", vmin=0, vmax=9)
axes[1].set_title("Rank-1 reconstruction")
axes[2].bar(np.arange(1, len(s) + 1), s, color="#1665D8")
axes[2].set_title("Singular spectrum")
axes[2].set_xlabel("Component")
for ax in axes[:2]:
    ax.axis("off")
plt.tight_layout()
plt.show()
```

Selected output

```
Singular values: [18.9443 1.0557 0. 0.]
Energy retained: 0.996904   Frobenius error: 1.0557
Stored values: 16 -> 9
PCA singular values: [3.0621 0.611 ]
Explained-variance ratio: [0.9617 0.0383]
```

Implementation note

`np.linalg.svd` returns $\mathbf{V}^\top$ as `Vt`. Multiplying `U[:, :k] * s[:k]` scales each retained column of $\mathbf{U}$ by its corresponding singular value.

8. Concept Check and Practice

Common misconceptions

Misconception	Why it fails	Correct practice
SVD is only for square matrices	It exists for every real rectangular matrix	Track the dimensions of $\mathbf{U}, \Sigma, \mathbf{V}$
The largest singular value is always enough	Important secondary structure may be lost	Select k using error, energy, and domain needs
PCA requires uncentred raw data	The mean can dominate the directions	Centre variables before decomposition
High energy guarantees perfect visual quality	Energy is a matrix norm criterion	Inspect reconstruction and task performance

Three concept-check MCQs

1. The singular values of $\mathbf{A}$ are square roots of eigenvalues of:

A. $\mathbf{A} + \mathbf{I}$

B. $\mathbf{A}^\mathsf{T}\mathbf{A}$

C. $\mathbf{A} - \mathbf{I}$

D. $\mathbf{A}^\mathsf{T} + \mathbf{A}$

2. The best rank-k approximation is obtained by:

A. retaining the smallest singular values

B. sorting rows

C. retaining the largest k singular values

D. taking a determinant

3. In PCA through SVD, principal directions are columns of:

A. $\mathbf{V}$

B. Σ

C. $\mathbf{A}^{-1}$

D. $\mathbf{I}$

MCQ answers: 1-B, 2-C, 3-A

Practice exercises

1. Find the singular values of $\begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}$.
2. Verify the SVD of $\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ given on page 4.
3. Find the rank of a matrix whose singular values are $7, 2, 0$.
4. Find κ_2 when $\sigma_{\text{max}} = 8$ and $\sigma_{\text{min}} = 0.5$.
5. Calculate the rank-one approximation of the worked matrix.
6. Compare storage for a 100×80 matrix and a rank-5 SVD representation.
7. If PCA singular values are $5, 2, 1$, find the variance ratio of PC1 and of the first two components.
8. Use Python to select the smallest rank retaining at least 99% energy for the image case study.

9. Scenario Challenge, Answers, and Takeaways

Scenario-based problem: recommendation compression

A streaming platform records the following four-user by three-category preference matrix:

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

Using Python, select the smallest rank retaining at least 80% of the squared singular-value energy. Report the approximation error, storage count, and practical interpretation.

Brief answers to practice

Exercise	Brief answer
1	$\sigma_1 = 6,\ \sigma_2 = 2$
2	$\mathbf{U} = \mathbf{V} = 2^{-1/2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $\mathbf{\Sigma} = \text{diag}(4, 2)$ reconstruct $\mathbf{A}$
3	Rank $= 2$ because two singular values are nonzero
4	$\kappa_2 = 8/0.5 = 16$
5	$\mathbf{A}_1 = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$
6	Original $= 8000$ values; rank-5 form $= 5(100 + 80 + 1) = 905$ values; reduction $\approx 88.69\%$
7	PC1 $= 25/30 = 83.33\%$; first two $= 29/30 = 96.67\%$
8	Rank 1 retains approximately 99.69% energy
Scenario	Singular values $\approx 10.5961, 5.0717, 0$; rank 1 retains 81.36%; error ≈ 5.0717; storage 8 instead of 12 values

Scenario interpretation

Rank 1 meets the stated threshold and compresses the matrix by 33.33%. It captures the dominant shared preference pattern, but it weakens the distinction between the third category and the first two; rank selection should therefore consider recommendation quality as well as energy.

Lecture summary

- SVD decomposes any real matrix into orthogonal directions and nonnegative scales.
- Singular values reveal rank, strength, conditioning, and approximation error.
- Truncated SVD gives the best low-rank approximation under standard matrix norms.
- PCA uses SVD of centred data to identify directions of maximum variance.
- Compression is a controlled trade-off among storage, error, and application quality.

Key terms: SVD • singular value • left singular vector • right singular vector • reduced SVD • truncated SVD • low rank • pseudoinverse • condition number • PCA • explained variance

Bridge to Lecture 10
The final lecture integrates linear systems, transformations, projections, regression, eigenpairs, and SVD through modern applications and a complete course review.

Suggested references: Gilbert Strang, Introduction to Linear Algebra; Gene H. Golub and Charles F. Van Loan, Matrix Computations; NumPy SVD documentation.