San José State University

Math 253: Mathematical Methods for Data Visualization

Lecture 5: Singular Value Decomposition (SVD)

Dr. Guangliang Chen

Outline

- Matrix SVD

Introduction

We have seen that symmetric matrices are always (orthogonally) diagonalizable.

That is, *for any symmetric matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$, *there exist an orthogonal matrix* $\mathbf{Q} = [\mathbf{q}_1 \ldots \mathbf{q}_n]$ *and a diagonal matrix* $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, *both real and square, such that*

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T.$$

We have pointed out that λ_i's are the eigenvalues of $\mathbf{A}$ and $\mathbf{q}_i$'s the corresponding eigenvectors (which are orthogonal to each other and have unit norm).

Thus, such a factorization is called the **eigendecomposition** of $\mathbf{A}$, also called the **spectral decomposition** of $\mathbf{A}$.

What about general rectangular matrices?

Existence of the SVD for general matrices

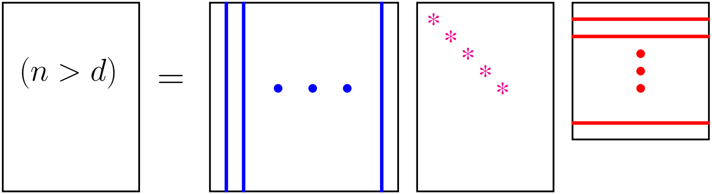
Theorem: For any matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$, there exist two orthogonal matrices $\mathbf{U} \in \mathbb{R}^{n\times n}, \mathbf{V} \in \mathbb{R}^{d\times d}$ and a nonnegative, "diagonal" matrix $\mathbf{\Sigma} \in \mathbb{R}^{n\times d}$ (of the same size as $\mathbf{X}$) such that

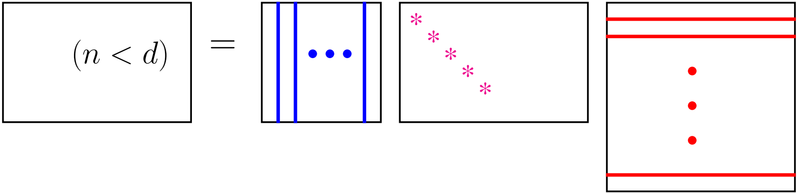
$$\mathbf{X}_{n\times d} = \mathbf{U}_{n\times n}\mathbf{\Sigma}_{n\times d}\mathbf{V}^T_{d\times d}.$$

Remark. This is called the *Singular Value Decomposition (SVD)* of $\mathbf{X}$:

- The diagonals of $\mathbf{\Sigma}$ are called the singular values of $\mathbf{X}$ (often sorted in decreasing order).
- The columns of $\mathbf{U}$ are called the left singular vectors of $\mathbf{X}$.
- The columns of $\mathbf{V}$ are called the right singular vectors of $\mathbf{X}$.

Singular Value Decomposition (SVD)





Connection to spectral decomposition of symmetric matrices

From the SVD of $\mathbf{X}$ we obtain that

$$\mathbf{X}\mathbf{X}^T = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \cdot \mathbf{V}\boldsymbol{\Sigma}^T\mathbf{U}^T = \mathbf{U}\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\right)\mathbf{U}^T$$
$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\boldsymbol{\Sigma}^T\mathbf{U}^T \cdot \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \mathbf{V}\left(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\right)\mathbf{V}^T$$

This shows that

- $\mathbf{U}$ is the eigenvectors matrix of $\mathbf{X}\mathbf{X}^T$;
- $\mathbf{V}$ is the eigenvectors matrix of $\mathbf{X}^T\mathbf{X}$;
- The eigenvalues of $\mathbf{X}\mathbf{X}^T, \mathbf{X}^T\mathbf{X}$ (which must be the same) are equal to the squared singular values of $\mathbf{X}$.

How to prove the SVD theorem

Given any matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$, the SVD can be thought of as solving a matrix equation for three unknown matrices (each with certain constraint):

$$\mathbf{X} = \underbrace{\mathbf{U}}_{\text{orthogonal}} \cdot \underbrace{\mathbf{\Sigma}}_{\text{diagonal}} \cdot \underbrace{\mathbf{V}^T}_{\text{orthogonal}}.$$

Suppose such solutions exist.

- From previous slide:

  $$\mathbf{X}^T\mathbf{X} = \mathbf{V}\left(\mathbf{\Sigma}^T\mathbf{\Sigma}\right)\mathbf{V}^T$$

  This tells us how to find $\mathbf{V}$ and $\mathbf{\Sigma}$ (which contain the eigenvectors and square roots of eigenvalues of $\mathbf{X}^T\mathbf{X}$, respectively).

- After we have found both $\mathbf{V}$ and $\mathbf{\Sigma}$, rewrite the matrix equation as

$$\mathbf{XV} = \mathbf{U\Sigma},$$

or in columns,

$$\mathbf{X}[\mathbf{v}_1 \ldots \mathbf{v}_r\, \mathbf{v}_{r+1} \ldots \mathbf{v}_d] = [\mathbf{u}_1 \ldots \mathbf{u}_r\, \mathbf{u}_{r+1} \ldots \mathbf{u}_n] \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \\ \end{bmatrix}.$$

By comparing columns, we obtain

$$\mathbf{X}\mathbf{v}_i = \begin{cases} \sigma_i \mathbf{u}_i, & 1 \le i \le r \text{ (\#nonzero singular values)} \\ \mathbf{0}, & r < i \le d \end{cases}$$

This tells us how to find the matrix $\mathbf{U}$: $\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{X}\mathbf{v}_i$ for $1 \le i \le r$.

A rigorous proof of the SVD theorem

Let $\mathbf{C} = \mathbf{X}^T\mathbf{X} \in \mathbb{R}^{d\times d}$. Then $\mathbf{C}$ is square, symmetric, and positive semidefinite.

Therefore, by the Spectral Theorem, $\mathbf{C} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ for an orthogonal $\mathbf{V} \in \mathbb{R}^{d\times d}$ and diagonal $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ with $\lambda_1 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_d$ (where $r = \mathrm{rank}(\mathbf{X}) \leq d$).

Let $\sigma_i = \sqrt{\lambda_i}$ and correspondingly form the matrix $\mathbf{\Sigma} \in \mathbb{R}^{n\times d}$:

$$\mathbf{\Sigma} = \begin{bmatrix} \mathrm{diag}(\sigma_1, \ldots, \sigma_r) & \mathbf{O}_{r\times(d-r)} \\ \mathbf{O}_{(n-r)\times r} & \mathbf{O}_{(n-r)\times(d-r)} \end{bmatrix}$$

Define also

$$\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{X}\mathbf{v}_i \in \mathbb{R}^n, \qquad \text{for each } 1 \leq i \leq r.$$

Then $\mathbf{u}_1, \ldots, \mathbf{u}_r$ are orthonormal vectors. To see this,

$$\begin{aligned}
\mathbf{u}_i^T \mathbf{u}_j &= \left(\frac{1}{\sigma_i}\mathbf{X}\mathbf{v}_i\right)^T \left(\frac{1}{\sigma_j}\mathbf{X}\mathbf{v}_j\right) = \frac{1}{\sigma_i \sigma_j}\mathbf{v}_i^T \underbrace{\mathbf{X}^T\mathbf{X}}_{=\mathbf{C}}\mathbf{v}_j \\
&= \frac{1}{\sigma_i \sigma_j}\mathbf{v}_i^T(\lambda_j \mathbf{v}_j) = \frac{\sigma_j}{\sigma_i}\mathbf{v}_i^T\mathbf{v}_j \qquad\qquad (\lambda_j = \sigma_j^2) \\
&= \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}
\end{aligned}$$

Choose $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n \in \mathbb{R}^n$ (through basis completion) such that

$$\mathbf{U} = [\mathbf{u}_1 \ldots \mathbf{u}_r \mathbf{u}_{r+1} \ldots \mathbf{u}_n] \in \mathbb{R}^{n \times n}$$

is an orthogonal matrix.

It remains to verify that $\mathbf{XV} = \mathbf{U\Sigma}$, i.e.,

$$\mathbf{X}[\mathbf{v}_1 \ldots \mathbf{v}_r \, \mathbf{v}_{r+1} \ldots \mathbf{v}_d] = [\mathbf{u}_1 \ldots \mathbf{u}_r \, \mathbf{u}_{r+1} \ldots \mathbf{u}_n] \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \\ & & & \end{bmatrix}.$$

Consider two cases:

- $1 \leq i \leq r$: $\mathbf{X}\mathbf{v}_i = \sigma_i \mathbf{u}_i$ by construction.
- $i > r$: $\mathbf{X}\mathbf{v}_i = \mathbf{0}$, which is due to $\mathbf{X}^T\mathbf{X}\mathbf{v}_i = \mathbf{C}\mathbf{v}_i = 0\mathbf{v}_i = \mathbf{0}$.

Consequently, we have obtained that $\mathbf{X} = \mathbf{U\Sigma V}^T$.

Example 0.1. Compute the SVD of

$$\mathbf{X} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Answer:

$$\mathbf{X} = \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix} \cdot \begin{pmatrix} \sqrt{3} & \\ & 1 \\ & \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^T$$

Geometric interpretation of SVD

Given any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, it defines a linear transformation:

$$f : \mathbb{R}^n \mapsto \mathbb{R}^m, \qquad \text{with} \quad f(\mathbf{x}) = \mathbf{A}\mathbf{x}.$$

The SVD of $\mathbf{A}$ indicates that the linear transformation f can be decomposed into a sequence of three operations:

$$\underbrace{\mathbf{A}\mathbf{x}}_{\text{full transformation}} = \underbrace{\mathbf{U}}_{\text{rotation}} \cdot \underbrace{\mathbf{\Sigma}}_{\text{rescaling}} \cdot \underbrace{\mathbf{V}^T\mathbf{x}}_{\text{rotation}}$$

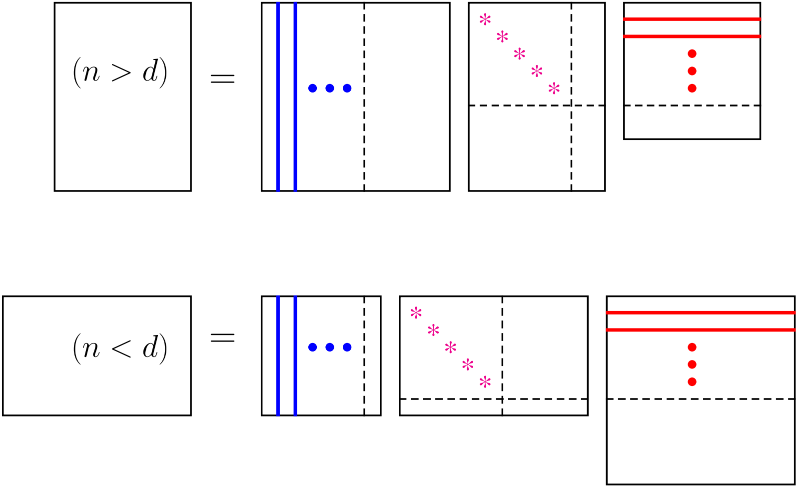
Different versions of SVD

- **Full SVD**: $\mathbf{X}_{n\times d} = \mathbf{U}_{n\times n}\mathbf{\Sigma}_{n\times d}\mathbf{V}^T_{d\times d}$
- **Compact SVD**: Suppose $\mathrm{rank}(\mathbf{X}) = r$. Define

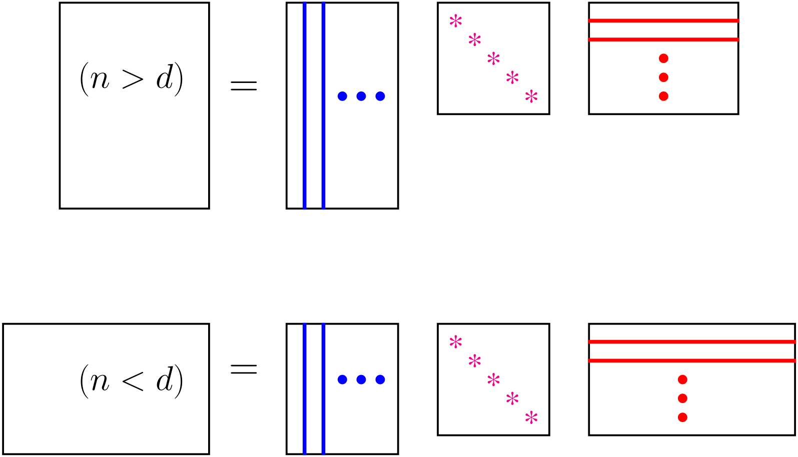
$$\mathbf{U}_r = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{R}^{n\times r}$$
$$\mathbf{V}_r = [\mathbf{v}_1, \ldots, \mathbf{v}_r] \in \mathbb{R}^{d\times r}$$
$$\mathbf{\Sigma}_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$$

Then

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


$(n > d)$
=
$(n < d)$
=

Singular Value Decomposition (SVD)



- **Rank-1 decomposition**:

$$\mathbf{X} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

This has the interpretation that $\mathbf{X}$ is a weighted sum of rank-one matrices, as for a square, symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T = \sum_{i=1}^{n} \lambda_i \mathbf{q}_i \mathbf{q}_i^T.$$

In sum, $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T$ where both $\mathbf{U}, \mathbf{V}$ have orthonormal columns and $\mathbf{\Sigma}$ is diagonal.

Remark. For any version of SVD, the form is not unique (this is mainly due to different choices of orthogonal basis for each eigenspace).

Remark. For any matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$ and integer $1 \le K \le r$, we define the truncated SVD of X with K terms as

$$\mathbf{X} \approx \sum_{i=1}^{K} \sigma_i \mathbf{u}_i \mathbf{v}_i^T = \mathbf{X}_K$$

where the singular values are assumed to be sorted from large to small (so $\sigma_1, \ldots, \sigma_K$ represent the largest K singular values).

Note that $\mathbf{X}_K$ has a rank of K and is not exactly equal to $\mathbf{X}$ (thus can be regarded as an approximation to $\mathbf{X}$).

Power method for numerical computing of SVD

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be a matrix whose SVD is to be computed: $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Consider $\mathbf{C} = \mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$. Then

$$\begin{aligned}
\mathbf{C} &= \mathbf{V}(\mathbf{\Sigma}^T\mathbf{\Sigma})\mathbf{V}^T = \sum \sigma_i^2 \mathbf{v}_i\mathbf{v}_i^T \\
\mathbf{C}^2 &= \mathbf{V}(\mathbf{\Sigma}^T\mathbf{\Sigma})^2\mathbf{V}^T = \sum \sigma_i^4 \mathbf{v}_i\mathbf{v}_i^T \\
&\vdots \\
\mathbf{C}^k &= \mathbf{V}(\mathbf{\Sigma}^T\mathbf{\Sigma})^k\mathbf{V}^T = \sum \sigma_i^{2k} \mathbf{v}_i\mathbf{v}_i^T
\end{aligned}$$

If $\sigma_1 > \sigma_2$, then the first term dominates, so $\mathbf{C}^k \to \sigma_1^{2k}\mathbf{v}_1\mathbf{v}_1^T$ as $k \to \infty$.

This means that a close estimate to $\mathbf{v}_1$ can be computed by simply taking the first column of $\mathbf{C}^k$ and normalizing it to a unit vector.

The previous method is very costly due to the matrix power part.

A better approach. Instead of computing $\mathbf{C}^k$, we select a random vector $\mathbf{x} \in \mathbb{R}^n$ and compute $\mathbf{C}^k\mathbf{x}$ through a sequence of matrix-vector multiplications (which are very efficient especially when one dimension of $\mathbf{A}$ is small, or $\mathbf{A}$ is sparse):

$$\mathbf{C}^k\mathbf{x} = \mathbf{A}^T\mathbf{A}\cdots\mathbf{A}^T\mathbf{A}\mathbf{x}$$

Write $\mathbf{x} = \sum c_i\mathbf{v}_i$ (since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form an orthonormal basis for $\mathbb{R}^n$). Then

$$\mathbf{C}^k\mathbf{x} \approx (\sigma_1^{2k}\mathbf{v}_1\mathbf{v}_1^T)\left(\sum c_i\mathbf{v}_i\right) = \sigma_1^{2k}c_1\mathbf{v}_1.$$

Normalizing the vector $\mathbf{C}^k\mathbf{x}$ for some large k then yields $\mathbf{v}_1$, the first right singular vector of $\mathbf{A}$.

MATLAB commands for computing matrix SVD

1. Full SVD

svd – Singular Value Decomposition.

[U,S,V] = svd(X) produces a diagonal matrix S, of the same dimension as X and with nonnegative diagonal elements in decreasing order, and orthogonal matrices U and V so that X = U*S*V^T.

s = svd(X) returns a vector containing the singular values.

2. Truncated SVD

svds – Find a few singular values and vectors.

S = svds(A,K) computes the K largest singular values of A.

[U,S,V] = svds(A,K) computes the singular vectors as well. If A is M-by-N and K singular values are computed, then U is M-by-K with orthonormal columns, S is K-by-K diagonal, and V is N-by-K with orthonormal columns.

In many applications, a truncated SVD is enough, and it is much easier to compute than the full SVD.