

# Singular value decomposition

Victor Kitov

[v.v.kitov@yandex.ru](mailto:v.v.kitov@yandex.ru)

# Table of Contents

- 1 Definition of SVD
- 2 Truncated SVD
- 3 Applications of SVD
- 4 Recommendation system with SVD

## SVD decomposition

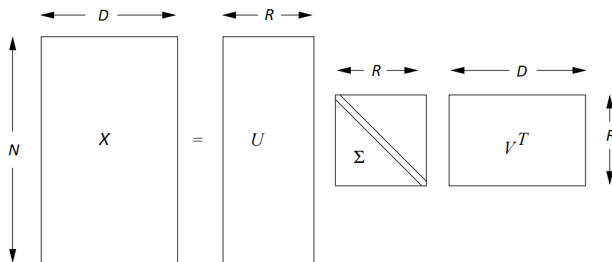
Every matrix  $X \in \mathbb{R}^{N \times D}$ ,  $\text{rank } X = R$ , can be decomposed into the product of three matrices:

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

where

- $U \in \mathbb{R}^{N \times R}$ ,  $\Sigma \in \mathbb{R}^{R \times R}$ ,  $V^T \in \mathbb{R}^{R \times D}$
- $\Sigma = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_R\}$ ,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_R \geq 0$
- $U^T U = I$ ,  $V^T V = I$ , where  $I \in \mathbb{R}^{R \times R}$  is identity matrix.

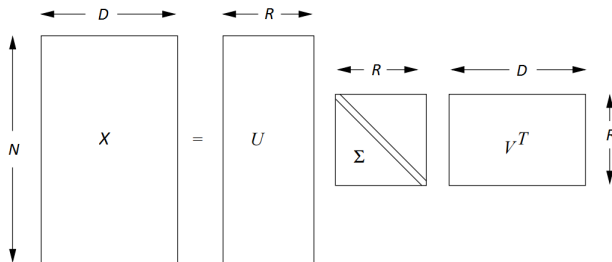
# Interpretation of SVD



For  $X_{ij}$  let  $i$  denote objects and  $j$  denote properties.

- Columns of  $U$  - orthonormal basis of columns of  $X$
- Rows of  $V^T$  - orthonormal basis of rows of  $X$
- $\Sigma$  - scaling.
- Efficient representations of low-rank matrix!

# Interpretation of SVD



For  $X_{ij}$  let  $i$  denote objects and  $j$  denote properties.

- Rows of  $U$  are normalized coordinates of rows in  $V^T$
- $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_R\}$  shows the importances of each column of  $U$  and respectively each row from  $V^T$ .
- rows of  $V^T$  are principal components

## SVD: existence & uniqueness

### Theorem 1

*For any matrix  $X \in \mathbb{R}^{N \times D}$  SVD decomposition exists.*

### Theorem 2

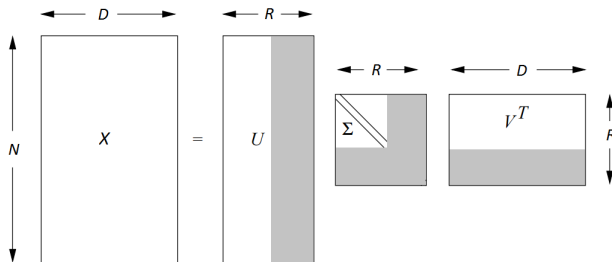
*SVD decomposition is unique up to sign if  $X^T X \in \mathbb{R}^{D \times D} \Leftrightarrow$  has a set of  $D$  unique eigenvalues.*

*Unique up to sign* means that we can always simultaneously change signs of  $u_i$  and  $v_i^T$  for  $\forall i = 1, 2, \dots, R$ .

# Table of Contents

- 1 Definition of SVD
- 2 Truncated SVD**
- 3 Applications of SVD
- 4 Recommendation system with SVD

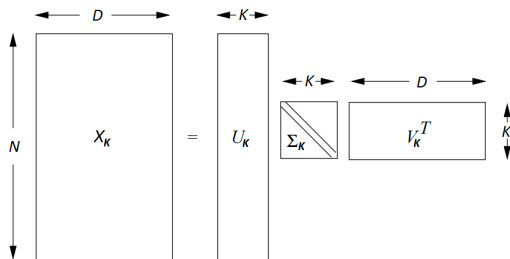
## Truncated SVD decomposition



$$\Sigma = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K, \sigma_{K+1}, \dots, \sigma_R\} \longrightarrow \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K, 0, 0, \dots, 0\} = \Sigma_K$$



# Truncated SVD decomposition



Simplification to rank  $K \leq R$ :

$$X_K = U_K \Sigma_K V_K$$

$$\Sigma = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K, \sigma_{K+1}, \dots, \sigma_R\} \longrightarrow \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_K\} = \Sigma_K$$

$$U = [u_1, u_2, \dots, u_K, u_{K+1}, \dots, u_R] \longrightarrow [u_1, u_2, \dots, u_K] = U_K$$

$$V = [v_1, v_2, \dots, v_K, v_{K+1}, \dots, v_R] \longrightarrow [v_1, v_2, \dots, v_K] = V_K$$

- Now rows of  $U$  give reduced representation of rows of  $X$ .

## Properties of truncated SVD decomposition

### Frobenius norm of matrix

$$\|X\|_F^2 = \sum_{n=1}^N \sum_{d=1}^D x_{nd}^2$$

- For matrix  $X$  and its approximation  $\hat{X}$  we can measure

$$\text{approximation error} = \|\hat{X} - X\|_F^2$$

### Theorem 3

Suppose  $X \in \mathbb{R}^{N \times D}$ , is approximated with  $\hat{X}_K = U_K \Sigma_K V_K$ . Then:

- 1 rank  $X_K = K$ .
- 2  $X_K = \arg \min_{B: \text{rank } B \leq K} \|X - B\|_F^2$

## Which K to choose for approximation?

## Theorem 4

For any matrix  $X$  and its singular value decomposition  $A = U\Sigma V^T$ ,  $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_R\}$ :

$$\|X\|_F^2 = \sum_{i=1}^R \sigma_i^2$$

- Suppose  $X = U\Sigma V^T$ ,  $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_R\}$
- Approximation  $\hat{X}_K = U\Sigma_K V^T$ ,  $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_K, 0, 0, \dots, 0\}$ .
- Then error of approximation  $E_K = X - \hat{X}_K = U\tilde{\Sigma}V^T$ , where  $\tilde{\Sigma} = \text{diag}\{0, 0, \dots, 0, \sigma_{K+1}, \dots, \sigma_R\}$

## Which $K$ to choose for approximation?

Select  $K$  giving relative error below some threshold  $t$ :

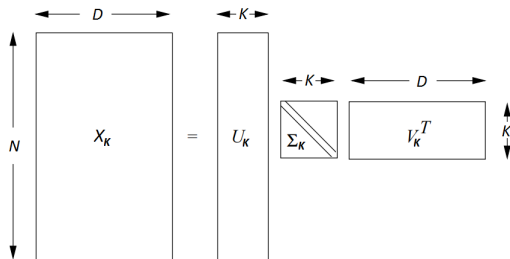
$$K = \arg \min_K \left\{ \frac{\|E_K\|_F^2}{\|X\|_F^2} = \frac{\sum_{i=K+1}^R \sigma_i^2}{\sum_{i=1}^R \sigma_i^2} < t \right\}$$

We used theorem 4 for calculation of Frobenius matrix norm.

# Table of Contents

- 1 Definition of SVD
- 2 Truncated SVD
- 3 Applications of SVD**
- 4 Recommendation system with SVD

# Dimensionality reduction



- rows of  $U$  give truncated representation of rows of  $X$ .
- $x_n \in \mathbb{R}^D \longrightarrow u_n \in \mathbb{R}^K$

## Memory efficiency

Storage costs of  $X \in \mathbb{R}^{N \times D}$ , assuming  $N \geq D$  and each element taking 1 byte:

Memory storage costs

representation of $X$	memory requirements
original $X$	?
fully SVD decomposed	?
truncated SVD to rank $K$	?

## Performance efficiency

- Multiplication  $Xq$ 
  - $X$  - normalized documents representation
  - $q$  - normalized search query

representation of $X$	$Xq$ complexity
original $X$	?
truncated SVD to rank $K$	?



## Finding similar objects and similar features

- Similar objects have coappearing features.
- Similar features coappear in objects.
- Example: text analysis.
  - LSA gives compact representations, invariant to synonyms
  - can compare documents
  - can compare words

# Table of Contents

- 1 Definition of SVD
- 2 Truncated SVD
- 3 Applications of SVD
- 4 Recommendation system with SVD

# Example

	Terminator	Gladiator	Rambo	Titanic	Love story	A walk to remember
Andrew	4	5	5	0	0	0
John	4	4	5	0	0	0
Matthew	5	5	4	0	0	0
Anna	0	0	0	5	5	5
Maria	0	0	0	5	5	4
Jessika	0	0	0	4	5	4

## Example

$$U = \begin{pmatrix} 0. & 0.6 & -0.3 & 0. & 0. & -0.8 \\ 0. & 0.5 & -0.5 & 0. & 0. & 0.6 \\ 0. & 0.6 & 0.8 & 0. & 0. & 0.2 \\ 0.6 & 0. & 0. & -0.8 & -0.2 & 0. \\ 0.6 & 0. & 0. & 0.2 & 0.8 & 0. \\ 0.5 & 0. & 0. & 0.6 & -0.6 & 0. \end{pmatrix}$$

$$\Sigma = \text{diag}\{(14. \quad 13.7 \quad 1.2 \quad 0.6 \quad 0.6 \quad 0.5)\}$$

$$V^T = \begin{pmatrix} 0. & 0. & 0. & 0.6 & 0.6 & 0.5 \\ 0.5 & 0.6 & 0.6 & 0. & 0. & 0. \\ 0.5 & 0.3 & -0.8 & 0. & 0. & 0. \\ 0. & 0. & 0. & -0.2 & 0.8 & -0.6 \\ -0. & -0. & -0. & 0.8 & -0.2 & -0.6 \\ 0.6 & -0.8 & 0.2 & 0. & 0. & 0. \end{pmatrix}$$

## Example (excluded insignificant concepts)

$$U_2 = \begin{pmatrix} 0. & 0.6 \\ 0. & 0.5 \\ 0. & 0.6 \\ 0.6 & 0. \\ 0.6 & 0. \\ 0.5 & 0. \end{pmatrix}$$

$$\Sigma_2 = \text{diag}\{(14. \quad 13.7)\}$$

$$V_2^T = \begin{pmatrix} 0. & 0. & 0. & 0.6 & 0.6 & 0.5 \\ 0.5 & 0.6 & 0.6 & 0. & 0. & 0. \end{pmatrix}$$

Concepts may be

- patterns among movies (along  $j$ ) - action movie / romantic movie
- patterns among people (along  $i$ ) - boys / girls

**Dimensionality reduction case:** patterns along  $j$  axis.

# Applications

- Example: new movie rating by new person

$$x = (5 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0)$$

- **Dimensionality reduction:** map  $x$  into concept space:

$$y = V_2^T x = (0 \quad 2.7)$$

- **Recommendation system:** map  $y$  back to original movies space:

$$\hat{x} = yV_2^T = (1.5 \quad 1.6 \quad 1.6 \quad 0 \quad 0 \quad 0)$$

## Summary

- SVD decomposition  $X = U\Sigma V^T$ ,  $U^T U = I$ ,  $V^T V = I$ ,  $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_R\}$  exists  $\forall X$ .
- Reduced SVD decomposition of order  $K$  solves:

$$X_K = \arg \min_{B: \text{rank } B \leq K} \|X - B\|_F^2$$

- SVD (reduced SVD) extracts structure of large matrices with small (close to small) rank
  - gives intuitive representation
  - efficient representations
  - fast matrix multiplications
- Helpful in recommendation engines
  - pitfall: treats 0 (no vote) as real vote.