MATH 637: Mathematical Techniques in Data Science
The singular value decomposition

Dominique Guillot

Departments of Mathematical Sciences
University of Delaware

April 29 & May 1, 2020

Theorem. Let $A \in \mathbb{R}^{m \times n}$. Then we can factor

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

where

1. $U \in \mathbb{R}^{m \times m}$ is an orthogonal matrix ($UU^T = U^T U = I$).
2. $\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with non-negative diagonal.
3. $V \in \mathbb{R}^{n \times n}$ is an orthogonal matrix ($VV^T = V^T V = I$).

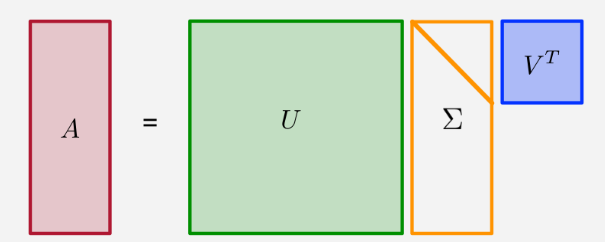


Figure from J.M. Phillips, *Mathematical Foundations for Data Analysis*.

Changing bases.

- Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ are bases of $\mathbb{R}^n$.

Changing bases.

- Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ are bases of $\mathbb{R}^n$.
- Change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$ is $P = P_{\mathcal{B} \to \mathcal{C}}$. Its columns

are the vectors of $\mathcal{B}$ expressed in the basis $\mathcal{C}$.

Changing bases.

- Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ are bases of $\mathbb{R}^n$.
- Change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$ is $P = P_{\mathcal{B}\to\mathcal{C}}$. Its columns are the vectors of $\mathcal{B}$ expressed in the basis $\mathcal{C}$.
- If $v \in \mathbb{R}^n$ has coordinates $(v_1, \ldots, v_n)$ in basis $\mathcal{B}$, meaning

$$v = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n,$$

then v has coordinates $Pv = (w_1, \ldots, w_n)$ in basis $\mathcal{C}$, meaning

$$v = w_1\mathbf{c}_1 + w_2\mathbf{c}_2 + \cdots + w_n\mathbf{c}_n.$$

Changing bases.

- Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ are bases of $\mathbb{R}^n$.
- Change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$ is $P = P_{\mathcal{B}\to\mathcal{C}}$. Its columns are the vectors of $\mathcal{B}$ expressed in the basis $\mathcal{C}$.
- If $v \in \mathbb{R}^n$ has coordinates $(v_1, \ldots, v_n)$ in basis $\mathcal{B}$, meaning

$$v = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n,$$

then v has coordinates $Pv = (w_1, \ldots, w_n)$ in basis $\mathcal{C}$, meaning

$$v = w_1\mathbf{c}_1 + w_2\mathbf{c}_2 + \cdots + w_n\mathbf{c}_n.$$

- Remark: $P_{\mathcal{C}\to\mathcal{B}} = P^{-1}_{\mathcal{B}\to\mathcal{C}}$.

Changing bases.

- Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ are bases of $\mathbb{R}^n$.
- Change of basis matrix from $\mathcal{B}$ to $\mathcal{C}$ is $P = P_{\mathcal{B}\to\mathcal{C}}$. Its columns are the vectors of $\mathcal{B}$ expressed in the basis $\mathcal{C}$.
- If $v \in \mathbb{R}^n$ has coordinates $(v_1, \ldots, v_n)$ in basis $\mathcal{B}$, meaning

$$v = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n,$$

then v has coordinates $Pv = (w_1, \ldots, w_n)$ in basis $\mathcal{C}$, meaning

$$v = w_1\mathbf{c}_1 + w_2\mathbf{c}_2 + \cdots + w_n\mathbf{c}_n.$$

- Remark: $P_{\mathcal{C}\to\mathcal{B}} = P_{\mathcal{B}\to\mathcal{C}}^{-1}$.
- We think of a matrix $A \in \mathbb{R}^{m\times n}$ as a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ written with respect to bases $\mathcal{C}$ and $\mathcal{D}$ of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively:

$$A = \begin{pmatrix} | & | & & | \\ A\mathbf{c_1} & A\mathbf{c_2} & \ldots & A\mathbf{c_n} \\ | & | & & | \end{pmatrix}$$

Columns of A = images of the vectors $\mathbf{c_i}$, expressed in the basis $\mathcal{D}$.

Change of basis for a matrix

- Suppose $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator from $V = \mathbb{R}^n$ to $W = \mathbb{R}^m$.

Change of basis for a matrix

- Suppose $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator from $V = \mathbb{R}^n$ to $W = \mathbb{R}^m$.
- We write it with respect to bases $\mathcal{C}$ and $\mathcal{D}$.

Change of basis for a matrix

- Suppose $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator from $V = \mathbb{R}^n$ to $W = \mathbb{R}^m$.
- We write it with respect to bases $\mathcal{C}$ and $\mathcal{D}$.
- We would like to re-write it using bases $\mathcal{B}$ and $\mathcal{E}$ instead.

Change of basis for a matrix

- Suppose $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator from $V = \mathbb{R}^n$ to $W = \mathbb{R}^m$.
- We write it with respect to bases $\mathcal{C}$ and $\mathcal{D}$.
- We would like to re-write it using bases $\mathcal{B}$ and $\mathcal{E}$ instead.

$$\begin{tikzcd}
V_{\mathcal{C}} \arrow[r, "A"] & W_{\mathcal{D}} \\
V_{\mathcal{B}} \arrow[u, "S = P_{\mathcal{B}\to\mathcal{C}}"] \arrow[r, "T^{-1}AS"] & W_{\mathcal{E}} \arrow[u, "T = P_{\mathcal{E}\to\mathcal{D}}"']
\end{tikzcd}$$

The matrix A becomes $T^{-1}AS$ in the new bases, where $S = P_{\mathcal{B}\to\mathcal{C}}$ and $T = P_{\mathcal{E}\to\mathcal{D}}$.

Diagonalization

Special case: $A : \mathbb{R}^n \to \mathbb{R}^n$, $\mathcal{C}$ = canonical basis.

$$
\begin{tikzcd}
V_{\mathcal{C}} \arrow[r, "A"] & V_{\mathcal{C}} \\
V_{\mathcal{B}} \arrow[u, "P = P_{\mathcal{B}\to\mathcal{C}}"] \arrow[r, "P^{-1}AP"] & V_{\mathcal{B}} \arrow[u, "P = P_{\mathcal{B}\to\mathcal{C}}"']
\end{tikzcd}
$$

Definition. A matrix A is *diagonalizable* if $P^{-1}AP = D$ for some invertible matrix P and some diagonal matrix D.

Special case: $A : \mathbb{R}^n \to \mathbb{R}^n$, $\mathcal{C}$ = canonical basis.

$$\begin{tikzcd} V_{\mathcal{C}} \arrow[r, "A"] & V_{\mathcal{C}} \\ V_{\mathcal{B}} \arrow[u, "P = P_{\mathcal{B}\to\mathcal{C}}"] \arrow[r, "P^{-1}AP"] & V_{\mathcal{B}} \arrow[u, "P = P_{\mathcal{B}\to\mathcal{C}}"'] \end{tikzcd}$$

Definition. A matrix A is *diagonalizable* if $P^{-1}AP = D$ for some invertible matrix P and some diagonal matrix D.

- In that case, $\mathcal{B}$ consists of eigenvectors of A associated to eigenvalues d_i.

Special case: $A : \mathbb{R}^n \to \mathbb{R}^n$, $\mathcal{C}$ = canonical basis.

$$\begin{array}{ccc} V_{\mathcal{C}} & \xrightarrow{\quad A \quad} & V_{\mathcal{C}} \\ P = P_{\mathcal{B}\to\mathcal{C}} \Big\uparrow & & \Big\uparrow P = P_{\mathcal{B}\to\mathcal{C}} \\ V_{\mathcal{B}} & \xrightarrow{P^{-1}AP} & V_{\mathcal{B}} \end{array}$$

Definition. A matrix A is *diagonalizable* if $P^{-1}AP = D$ for some invertible matrix P and some diagonal matrix D.

- In that case, $\mathcal{B}$ consists of eigenvectors of A associated to eigenvalues d_i.

Theorem. A matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if $\mathbb{R}^n$ has a basis consisting of eigenvectors of A.

Special case: $A : \mathbb{R}^n \to \mathbb{R}^n$, $\mathcal{C}$ = canonical basis.

$$\begin{tikzcd}
V_{\mathcal{C}} \arrow[r, "A"] & V_{\mathcal{C}} \\
V_{\mathcal{B}} \arrow[u, "P = P_{\mathcal{B}\to\mathcal{C}}"] \arrow[r, "P^{-1}AP"] & V_{\mathcal{B}} \arrow[u, "P = P_{\mathcal{B}\to\mathcal{C}}"']
\end{tikzcd}$$

Definition. A matrix A is *diagonalizable* if $P^{-1}AP = D$ for some invertible matrix P and some diagonal matrix D.

- In that case, $\mathcal{B}$ consists of eigenvectors of A associated to eigenvalues d_i.

Theorem. A matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if $\mathbb{R}^n$ has a basis consisting of eigenvectors of A.

- The columns of P are the eigenvectors of A expressed in the canonical basis of $\mathbb{R}^n$.

Special case: symmetric matrix

Theorem. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix ($A = A^T$). Then there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

- Hence every symmetric matrix is diagonalizable.

Special case: symmetric matrix

Theorem. Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix ($A = A^T$). Then there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

- Hence every symmetric matrix is diagonalizable.
- The columns of P contain the (orthonormal) eigenvectors of A.
- Hence $PP^T = P^T P = I$ (P is an orthogonal matrix) and

Special case: symmetric matrix

Theorem. Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix ($A = A^T$). Then there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

- Hence every symmetric matrix is diagonalizable.
- The columns of P contain the (orthonormal) eigenvectors of A.
- Hence $PP^T = P^T P = I$ (P is an orthogonal matrix) and

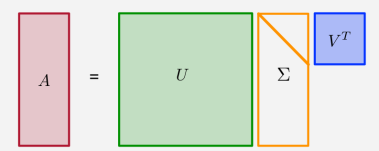
$$A = PDP^T.$$

Special case: symmetric matrix

Theorem. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix ($A = A^T$). Then there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

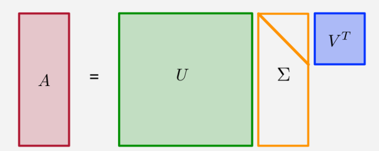
- Hence every symmetric matrix is diagonalizable.
- The columns of P contain the (orthonormal) eigenvectors of A.
- Hence $PP^T = P^TP = I$ (P is an orthogonal matrix) and
$$A = PDP^T.$$
- In general, a given matrix is **not** diagonalizable.

Back to SVD



- The SVD diagonalizes $A \in \mathbb{R}^{m \times n}$ with respect to two different orthonormal bases of $\mathbb{R}^n$ and $\mathbb{R}^m$.

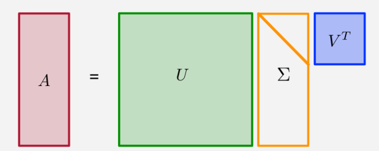
Back to SVD



- The SVD diagonalizes $A \in \mathbb{R}^{m\times n}$ with respect to two different orthonormal bases of $\mathbb{R}^n$ and $\mathbb{R}^m$.
- We have $U^T AV = \Sigma = T^{-1}AS$.

$$\begin{tikzcd}
\mathbb{R}^n_{\mathcal{C}} \arrow[r, "A"] & \mathbb{R}^m_{\mathcal{D}} \\
\mathbb{R}^n_{\mathcal{B}} \arrow[u, "V = P_{\mathcal{B}\to\mathcal{C}}"] \arrow[r, "U^TAV = \Sigma"] & \mathbb{R}^m_{\mathcal{E}} \arrow[u, "U = P_{\mathcal{E}\to\mathcal{D}}"']
\end{tikzcd}$$

Back to SVD

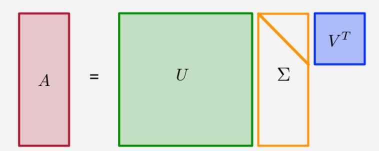


- The SVD diagonalizes $A \in \mathbb{R}^{m\times n}$ with respect to two different orthonormal bases of $\mathbb{R}^n$ and $\mathbb{R}^m$.
- We have $U^T AV = \Sigma = T^{-1}AS$.

$$
\begin{tikzcd}
\mathbb{R}^n_{\mathcal{C}} \arrow[r, "A"] & \mathbb{R}^m_{\mathcal{D}} \\
\mathbb{R}^n_{\mathcal{B}} \arrow[u, "V = P_{\mathcal{B}\to\mathcal{C}}"] \arrow[r, "U^T AV = \Sigma"] & \mathbb{R}^m_{\mathcal{E}} \arrow[u, "U = P_{\mathcal{E}\to\mathcal{D}}"']
\end{tikzcd}
$$

- Works for any matrix (even rectangular ones).

Back to SVD



- The SVD diagonalizes $A \in \mathbb{R}^{m \times n}$ with respect to two different orthonormal bases of $\mathbb{R}^n$ and $\mathbb{R}^m$.
- We have $U^T AV = \Sigma = T^{-1}AS$.

$$\begin{tikzcd}
\mathbb{R}^n_{\mathcal{C}} \arrow[r, "A"] & \mathbb{R}^m_{\mathcal{D}} \\
\mathbb{R}^n_{\mathcal{B}} \arrow[u, "V = P_{\mathcal{B}\to\mathcal{C}}"] \arrow[r, "U^T AV = \Sigma"] & \mathbb{R}^m_{\mathcal{E}} \arrow[u, "U = P_{\mathcal{E}\to\mathcal{D}}"']
\end{tikzcd}$$

- Works for any matrix (even rectangular ones).
- Columns of U are the *left singular vectors* of A
- Columns of V are the *right singular vectors* of A.
- Diagonal elements of Σ are the *singular values* of A.

Properties of the SVD

Write

$$\Sigma = \begin{pmatrix} D \\ \mathbf{0} \end{pmatrix}$$

Write

$$\Sigma = \begin{pmatrix} D \\ \mathbf{0} \end{pmatrix}$$

Observe that

- $AA^T = U\Sigma V^T V \Sigma^T U^T = U\Sigma\Sigma^T U^T = U \begin{pmatrix} D^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} U^T$.
- $A^T A = V\Sigma^T U^T U \Sigma V^T = V\Sigma^T\Sigma V^T = VD^2V^T$.

Properties of the SVD

Write

$$\Sigma = \begin{pmatrix} D \\ \mathbf{0} \end{pmatrix}$$

Observe that

- $AA^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T = U \begin{pmatrix} D^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} U^T.$
- $A^T A = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T\Sigma V^T = VD^2V^T.$
- These are eigendecompositions (diagonalizations) of the psd matrices AA^T and $A^T A$.

Properties of the SVD

Write

$$\Sigma = \begin{pmatrix} D \\ \mathbf{0} \end{pmatrix}$$

Observe that

- $AA^T = U\Sigma V^T V \Sigma^T U^T = U\Sigma\Sigma^T U^T = U \begin{pmatrix} D^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} U^T$.
- $A^T A = V\Sigma^T U^T U \Sigma V^T = V\Sigma^T \Sigma V^T = V D^2 V^T$.
- These are eigendecompositions (diagonalizations) of the psd matrices AA^T and $A^T A$.

Consequence:

- Columns of V are the eigenvectors of $A^T A$.
- Columns of U are the eigenvectors of AA^T.
- The (non-zero) singular values of A are the square roots of the (non-zero) eigenvalues of $A^T A$ or AA^T.

Application 1: Best rank k approximation

1. **Best rank k approximation.**

- Let $X \in \mathbb{R}^{n \times p}$ be a usual "data matrix".

Application 1: Best rank k approximation

1. Best rank k approximation.

- Let $X \in \mathbb{R}^{n\times p}$ be a usual "data matrix".
- Corresponding svd: $X = U\Sigma V^T = \sum_{i=1}^{\min(n,p)} \sigma_i u_i v_i^T$, with

$\sigma_1 \geq \sigma_2 \geq \dots$

1. Best rank k approximation.

- Let $X \in \mathbb{R}^{n\times p}$ be a usual "data matrix".
- Corresponding svd: $X = U\Sigma V^T = \sum_{i=1}^{\min(n,p)} \sigma_i u_i v_i^T$, with

$\sigma_1 \geq \sigma_2 \geq \dots$

Theorem. For $1 \leq k \leq \min(n,p)$, define:

$$X_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T.$$

1. Best rank k approximation.

- Let $X \in \mathbb{R}^{n \times p}$ be a usual "data matrix".
- Corresponding svd: $X = U\Sigma V^T = \sum_{i=1}^{\min(n,p)} \sigma_i u_i v_i^T$, with

$\sigma_1 \geq \sigma_2 \geq \dots$

Theorem. For $1 \leq k \leq \min(n,p)$, define:

$$X_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T.$$

Then
$$\min_{\substack{B \in \mathbb{R}^{n \times p} \\ \text{rank}(B) \leq k}} \|X - B\|_F^2 = \|X - X_k\|^2 = \sum_{j=k+1}^{p} \sigma_j^2$$

$$\min_{\substack{B \in \mathbb{R}^{n \times p} \\ \text{rank}(B) \leq k}} \|X - B\|_2 = \|X - X_k\|_2 = \sigma_{k+1}.$$

1. Best rank k approximation.

- Let $X \in \mathbb{R}^{n\times p}$ be a usual "data matrix".
- Corresponding svd: $X = U\Sigma V^T = \sum_{i=1}^{\min(n,p)} \sigma_i u_i v_i^T$, with $\sigma_1 \geq \sigma_2 \geq \dots$

Theorem. For $1 \leq k \leq \min(n,p)$, define:

$$X_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T.$$

Then

$$\min_{\substack{B\in\mathbb{R}^{n\times p}\\ \mathrm{rank}(B)\leq k}} \|X - B\|_F^2 = \|X - X_k\|^2 = \sum_{j=k+1}^{p} \sigma_j^2$$

$$\min_{\substack{B\in\mathbb{R}^{n\times p}\\ \mathrm{rank}(B)\leq k}} \|X - B\|_2 = \|X - X_k\|_2 = \sigma_{k+1}.$$

- The truncated SVD provides the *best rank k approximation* to X.
- Applications: data compression, data recovery, etc.

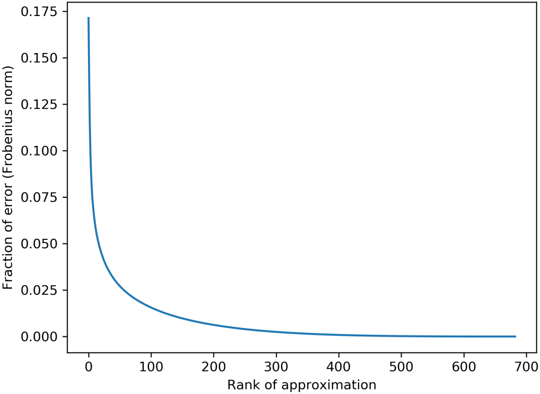
Example

Compressing the following image using the svd:

- Original image $X \in \mathbb{R}^{683\times1024}$.
- $X = U\Sigma V^T$.
- Approximate X by X_k.

Example (cont.)

We examine $\sum_{i=k+1}^{683} \sigma_i^2 / \sum_{i=1}^{683} \sigma_i^2$.



Example (cont.)

- Best rank 10 approximation:

Example (cont.)

- Best rank 50 approximation:

Example (cont.)

- Best rank 100 approximation:

- Best rank 200 approximation:

Example (cont.)

- Best rank 300 approximation:

- Best rank 400 approximation:

- Best rank 500 approximation:

Example (cont.)

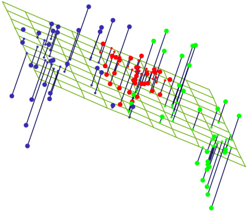
- Best rank 600 approximation:

Example (cont.)

- Full image (rank 683):

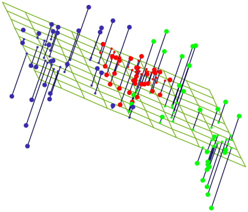
2. Projecting data on low dimensional subspace and PCA.

- The rows $x_1, \dots, x_n \in \mathbb{R}^p$ of X are observations of a p-dimensional vector.
- We would like to find a (lower dimensional) subspace of $\mathbb{R}^p$ approximating the data well



2. Projecting data on low dimensional subspace and PCA.

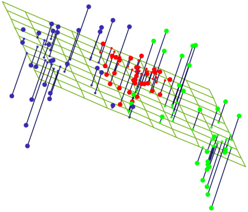
- The rows $x_1, \ldots, x_n \in \mathbb{R}^p$ of X are observations of a p-dimensional vector.
- We would like to find a (lower dimensional) subspace of $\mathbb{R}^p$ approximating the data well



- Assume the data is *centered*, i.e.,each column of X has mean 0.

2. Projecting data on low dimensional subspace and PCA.

- The rows $x_1, \dots, x_n \in \mathbb{R}^p$ of X are observations of a p-dimensional vector.
- We would like to find a (lower dimensional) subspace of $\mathbb{R}^p$ approximating the data well



- Assume the data is *centered*, i.e.,each column of X has mean 0.
- For a given $1 \leq k \leq p$, we want to solve:

$$\min_{\substack{F \text{ subspace of } \mathbb{R}^p \\ \dim F = k}} \sum_{i=1}^{n} \|x_i - \pi_F(x_i)\|_2^2,$$

where $\pi_F(x)$ denotes the projection of x onto F.

Application 2 (cont.)

Theorem. Let $v_1, \ldots, v_p$ denote the right singular vectors of X associated to $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$. The optimal k dimensional subspace solving the previous problem is $\mathrm{span}(v_1, v_2, \ldots, v_k)$.

Application 2 (cont.)

Theorem. Let $v_1, \ldots, v_p$ denote the right singular vectors of X associated to $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$. The optimal k dimensional subspace solving the previous problem is $\mathrm{span}(v_1, v_2, \ldots, v_k)$.

(Sketch of proof) For any vector $x \in \mathbb{R}^p$, we have

$$\|x\|_2^2 = \|\pi_F(x)\|_2^2 + \|\pi_{F^\perp}(x)\|_2^2.$$

Let $f_1, \ldots, f_k$ be an orthonormal basis of F. Then

$$\min_{\substack{F \text{ subspace of } \mathbb{R}^p \\ \dim F = k}} \sum_{i=1}^n \|x_i - \pi_F(x_i)\|_2^2 = \min_{\substack{F \text{ subspace of } \mathbb{R}^p \\ \dim F = k}} \sum_{i=1}^n \|\pi_{F^\perp}(x_i)\|_2^2$$

$$= \max_{\substack{F \text{ subspace of } \mathbb{R}^p \\ \dim F = k}} \sum_{i=1}^n \|\pi_F(x_i)\|_2^2 = \max_{f_1,\ldots,f_k \text{ orthonormal}} \sum_{j=1}^k \|X f_j\|_2^2$$

$$= \max_{f_1,\ldots,f_k \text{ orthonormal}} \sum_{j=1}^k f_j^T X^T X f_j$$

Using the min-max theorem for Rayleigh quotients, one can show that this is maximized when $\{f_1, \ldots, f_k\} = \{v_1, \ldots, v_k\}$.

3. Recommender system




$$
\begin{array}{l}
\text{Alice} \\
\text{Bob} \\
\text{Carol} \\
\text{Dave} \\
\\
\end{array}
\begin{pmatrix}
1 & 9 & 10 & 5 & \cdots \\
1 & 6 & 7 & 8 & \cdots \\
7 & 3 & 9 & 6 & \cdots \\
6 & 2 & 8 & 9 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
$$

- X_{ij} = ranking from person i of movie j.

Recommender system (cont.)

Idea. Try to explain why user i liked movie j as follows:

- Each movie is a combination of some unknown independent "basic features" (e.g. action, explosions, nature, romance, etc.)
- Each features has a degree of importance (weights).
- Each person likes a given combination of the basic features.

Recommender system (cont.)

Idea. Try to explain why user i liked movie j as follows:

- Each movie is a combination of some unknown independent "basic features" (e.g. action, explosions, nature, romance, etc.)
- Each features has a degree of importance (weights).
- Each person likes a given combination of the basic features.

- Write $X_{n\times p} = U_{n\times n}\Sigma_{n\times p}V_{p\times p}^T$.
- Then

Importance of feature k

$$x_{ij} = \sum_{k=1}^{p} u_{ik}\,\sigma_k\,v_{jk}$$

Amount of feature k in movie j

Ranking of Movie j from User i

How much User i likes feature k

Idea. Try to explain why user i liked movie j as follows:

- Each movie is a combination of some unknown independent "basic features" (e.g. action, explosions, nature, romance, etc.)
- Each features has a degree of importance (weights).
- Each person likes a given combination of the basic features.

- Write $X_{n\times p} = U_{n\times n}\Sigma_{n\times p}V^T_{p\times p}$.
- Then

$$\underbrace{x_{ij}}_{\text{Ranking of Movie j from User i}} = \sum_{k=1}^{p} \underbrace{u_{ik}}_{\text{How much User i likes feature k}} \overbrace{\sigma_k}^{\text{Importance of feature k}} \underbrace{v_{jk}}_{\text{Amount of feature k in movie j}}$$

- Can try to label each learned "feature" (or genre) from data (e.g. "Critically acclaimed western with a romantic component") .

Idea. Try to explain why user i liked movie j as follows:

- Each movie is a combination of some unknown independent "basic features" (e.g. action, explosions, nature, romance, etc.)
- Each features has a degree of importance (weights).
- Each person likes a given combination of the basic features.

- Write $X_{n\times p} = U_{n\times n}\Sigma_{n\times p}V^T_{p\times p}$.
- Then

Importance of
feature k

$$x_{ij} = \sum_{k=1}^{p} u_{ik}\,\sigma_k\,v_{jk}$$

Amount of feature k
in movie j

Ranking of Movie j
from User i

How much User i
likes feature k

- Can try to label each learned "feature" (or genre) from data (e.g. "Critically acclaimed western with a romantic component") .
- Other matrix factorization are possible (e.g. Non-negative matrix factorization).

Application 4: Text summarization

4. Text summarization.

Give a text document, find a few sentences summarizing it.

4. Text summarization.

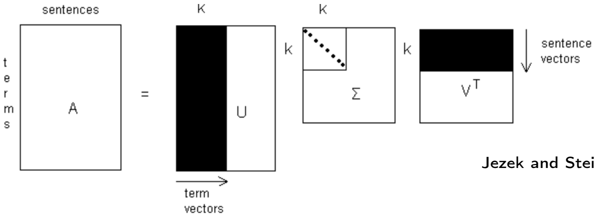
Give a text document, find a few sentences summarizing it.

- Define a *sentence matrix* $A = (a_{ij})$.
- $a_{ij} =$ frequency of word i in sentence j.

4. Text summarization.

Give a text document, find a few sentences summarizing it.

- Define a *sentence matrix* $A = (a_{ij})$.
- $a_{ij} =$ frequency of word i in sentence j.



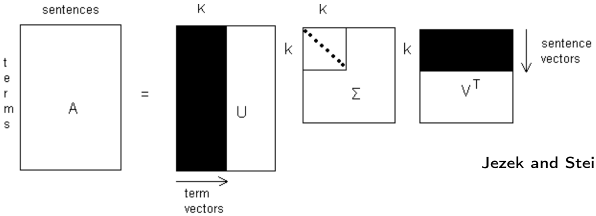
Jezek and Steinberger, 2004.

Application 4: Text summarization

4. Text summarization.

Give a text document, find a few sentences summarizing it.

- Define a *sentence matrix* $A = (a_{ij})$.
- $a_{ij} =$ frequency of word i in sentence j.



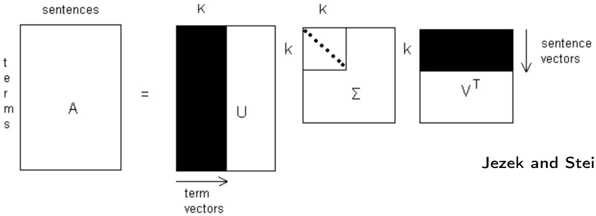
Jezek and Steinberger, 2004.

- The SVD provides statistics to guide the choice of the best sentences.

4. Text summarization.

Give a text document, find a few sentences summarizing it.

- Define a *sentence matrix* $A = (a_{ij})$.
- a_{ij} = frequency of word i in sentence j.



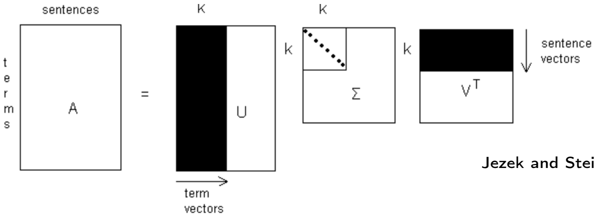
Jezek and Steinberger, 2004.

- The SVD provides statistics to guide the choice of the best sentences.
- Recall: the first k columns of V (rows of V^T) span the k-dim subspace that best approximates the columns of A (the sentences).

4. Text summarization.

Give a text document, find a few sentences summarizing it.

- Define a *sentence matrix* $A = (a_{ij})$.
- a_{ij} = frequency of word i in sentence j.



Jezek and Steinberger, 2004.

- The SVD provides statistics to guide the choice of the best sentences.
- Recall: the first k columns of V (rows of V^T) span the k-dim subspace that best approximates the columns of A (the sentences).
- E.g., Gong and Liu (2001). From each row of the V^T matrix, the sentence with the highest score is selected.