Mathematics Methods for Computer Science

Instructor: Xubo Yang

SJTU-SE DALAB

Lecture

Singular Value Decomposition

Understanding the Geometry of $A \in \mathbb{R}^{m \times n}$

Critical points of the ratio:

$$R(\vec{v}) = \frac{||A\vec{v}||_2}{||\vec{v}||_2}$$

Understanding the Geometry of $A \in \mathbb{R}^{m \times n}$

Critical points of the ratio:

$$R(\vec{v}) = \frac{||A\vec{v}||_2}{||\vec{v}||_2}$$

- $R(\alpha\vec{v}) = R(\vec{v}) \Rightarrow$ take $||\vec{v}||_2 = 1$
- $R(\vec{v}) \geq 0 \Rightarrow$ study $R^2(\vec{v})$ instead

Once Again...

Critical points satisfy $A^\top A\vec{v}_i = \lambda_i \vec{v}_i$.

Once Again...

Critical points satisfy $A^\top A\vec{v}_i = \lambda_i \vec{v}_i$.

Properties:

- $\lambda_i \geq 0 \; \forall i$
- Basis is full and orthonormal

Geometric Question

What about A instead of $A^\top A$?

Geometric Question

What about A instead of $A^\top A$?

Object of study: $\vec{u}_i \equiv A\hat{v}_i$

Observation

Lemma

Either $\vec{u}_i = \vec{0}$ or $\vec{u}_i$ is an eigenvector of $AA^\top$ with $||\vec{u}_i||_2 = \sqrt{\lambda_i}||\hat{v}_i||_2 = \sqrt{\lambda_i}$.

Lemma

Simpler proof than in book (top p. 132):

$$A^\top A\hat{v}_i = \lambda_i \hat{v}_i$$
$$AA^\top(A\hat{v}_i) = \lambda_i A\hat{v}_i$$
$$AA^\top \vec{u}_i = \lambda_i \vec{u}_i$$

Length of $\vec{u}_i = A\hat{v}_i$ follows from

$$\|\vec{u}_i\|_2^2 = \|A\hat{v}_i\|_2^2 = \hat{v}_i^\top A^\top A\hat{v}_i = \lambda_i \hat{v}_i^\top \hat{v}_i = \lambda_i$$

Corresponding Eigenvalues

$$
\begin{aligned}
k &= \text{ number of } \lambda_i > 0 \\
A^\top A \hat{v}_i &= \lambda_i \hat{v}_i \\
A A^\top \hat{u}_i &= \lambda_i \hat{u}_i \\
\bar{U} \in \mathbb{R}^{n \times k} &= \text{ matrix of unit } \hat{u}_i \text{ 's} \\
\bar{V} \in \mathbb{R}^{m \times k} &= \text{ matrix of unit } \hat{v}_i \text{ 's}
\end{aligned}
$$

Observation

Simpler lemma + proof than book (bottom p.132):

Lemma

$\hat{u}_i^\top A \hat{v}_j = \sqrt{\lambda_i}\delta_{ij}$

$$
\begin{aligned}
\bar{\Sigma} &\equiv diag(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_k}) \\
&= diag(\sigma_i, \cdots, \sigma_k)(\sigma_i \text{ are singular values})
\end{aligned}
$$

Observation

Simpler lemma + proof than book (bottom p.132):

Lemma

$\hat{u}_i^\top A \hat{v}_j = \sqrt{\lambda_i}\delta_{ij}$

$$\begin{aligned}\bar{\Sigma} &\equiv diag(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_k}) \\ &= diag(\sigma_i, \cdots, \sigma_k)(\sigma_i \text{ are singular values})\end{aligned}$$

Corollary

$\bar{U}^\top A \bar{V} = \bar{\Sigma}$

Fat SVD: Completing the Basis

Add $\hat{v}_i$ with $A^\top A\vec{\hat{v}}_i = \vec{0}$ and $\hat{u}_i$ with $AA^\top \hat{u}_i = \vec{0}$

Fat SVD: Completing the Basis

Add $\hat{v}_i$ with $A^\top A\vec{\hat{v}}_i = \vec{0}$ and $\hat{u}_i$ with $AA^\top \hat{u}_i = \vec{0}$

$$\bar{U} \in \mathbb{R}^{m\times k}, \bar{V} \in \mathbb{R}^{n\times k} \mapsto$$
$$U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$$

Fat SVD: Completing the Basis

Add $\hat{v}_i$ with $A^\top A\vec{\hat{v}}_i = \vec{0}$ and $\hat{u}_i$ with $AA^\top \hat{u}_i = \vec{0}$

$$\bar{U} \in \mathbb{R}^{m\times k}, \bar{V} \in \mathbb{R}^{n\times k} \mapsto$$
$$U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$$
$$\Sigma_{ij} \equiv \begin{cases} \sqrt{\lambda_i} & i = j \text{ and } i \leq k \\ 0 & \text{otherwise} \end{cases}$$

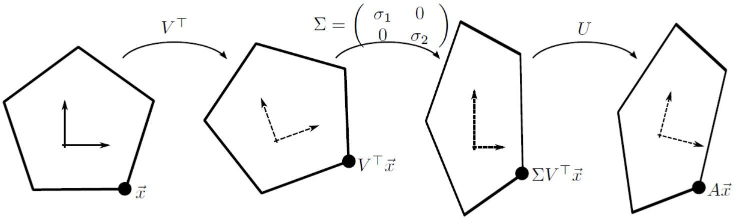
Singular Value Decomposition

$$A = U\Sigma V^\top$$

Geometry of Linear Transformations

$$A = U\Sigma V^\top$$

1. Rotate $(V^\top)$
2. Scale (Σ)
3. Rotate (U)



SVD Vocabulary

$$A = U\Sigma V^\top$$

- **Left singular vectors:**
  Columns of U; span $col\ A$
- **Right singular vectors:**
  Columns of V; span $row\ A$
- **Singular values:**
  Diagonal σ_i of Σ; sort $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$

Computing SVD: Simple Strategy

1. Columns of V are eigenvectors of $A^\top A$
2. $AV = U\Sigma \Rightarrow$ columns of U corresponding to nonzero singular values are normalized columns of AV
3. Remaining columns of U satisfy $AA^\top \vec{u}_i = \vec{0}$.

Computing SVD: Simple Strategy

1. Columns of V are eigenvectors of $A^\top A$
2. $AV = U\Sigma \Rightarrow$ columns of U corresponding to nonzero singular values are normalized columns of AV
3. Remaining columns of U satisfy $AA^\top \vec{u}_i = \vec{0}$.

$\exists$ **more specialized methods!**

Solving Linear Systems with $A = U\Sigma V^\top$

$$
\begin{aligned}
A\vec{x} &= \vec{b} \\
\implies U\Sigma V^\top \vec{x} &= \vec{b} \\
\implies \vec{x} &= V\Sigma^{-1}U^\top \vec{b}
\end{aligned}
$$

Solving Linear Systems with $A = U\Sigma V^\top$

$$
\begin{aligned}
A\vec{x} &= \vec{b} \\
\implies U\Sigma V^\top \vec{x} &= \vec{b} \\
\implies \vec{x} &= V\Sigma^{-1}U^\top \vec{b}
\end{aligned}
$$

What is Σ^{-1} ?

Uniting Short/Tall Matrices

$$\text{minimize } ||\vec{x}||_2^2$$
$$\text{such that } A^\top A\vec{x} = A^\top \vec{b}$$

Simplification

$$A^\top A = V\Sigma^\top \Sigma V^\top$$

Simplification

$$A^\top A = V\Sigma^\top \Sigma V^\top$$

$$A^\top A\vec{x} = A^\top \vec{b} \Leftrightarrow \Sigma^\top \Sigma \vec{y} = \Sigma^\top \vec{d}$$

$$\vec{y} \equiv V^\top \vec{x}$$

$$\vec{d} \equiv U^\top \vec{b}$$

Resulting Optimization

$$\text{minimize } ||\vec{y}||_2^2$$
$$\text{such that } \Sigma^\top \Sigma \vec{y} = \Sigma^\top \vec{d}$$

Solution

$$\Sigma^+_{ij} \equiv \begin{cases} 1/\sigma_i & i = j, \sigma_i \neq 0, \text{ and } i \leq k \\ 0 & \text{otherwise} \end{cases}$$

$$\implies \vec{y} = \Sigma^+ \vec{d}$$

$$\implies \vec{x} = V\Sigma^+ U^\top \vec{b}$$

Pseudoinverse

$$A^+ = V\Sigma^+ U^\top$$

Pseudoinverse Properties

- A **square** and **invertible** $\Rightarrow A^+ = A^{-1}$
- A **overdetermined** $\Rightarrow A^+\vec{b}$ gives least-squares
- A **underdetermined** $\Rightarrow A^+\vec{b}$ gives least-squares solution to $A\vec{x} \approx \vec{b}$ with least (Euclidean) norm

Alternative Form

$$A = U\Sigma V^\top \Longrightarrow A = \sum_{i=1}^{\ell} \sigma_i \vec{u}_i \vec{v}_i^\top$$
$$\ell \equiv \min\{m, n\}$$

Outer Product

$$\vec{u} \otimes \vec{v} \equiv \vec{u}\vec{v}^\top$$

Computing $A\vec{x}$

$$A\vec{x} = \sum_i \sigma_i(\vec{v}_i \cdot \vec{x})\vec{u}_i$$

Computing $A\vec{x}$

$$A\vec{x} = \sum_i \sigma_i (\vec{v}_i \cdot \vec{x}) \vec{u}_i$$

Trick:

Ignore small σ_i.

Computing $A^+\vec{x}$

$$A^+ = \sum_{\sigma_i \neq 0} \frac{\vec{v}_i \vec{u}_i^\top}{\sigma_i}$$

Trick:

Ignore large σ_i.

Even Better Trick

Do not compute large (small) σ_i at all!

Eckart-Young Theorem

Theorem

Suppose $\tilde{A}$ is obtained from $A = U\Sigma V^\top$ by truncating all but the k largest singular values σ_i of A to zero. Then, $\tilde{A}$ minimizes both $||A - \tilde{A}||_{Fro}$ and $||A - \tilde{A}||_2$ subject to the constraint that the column space of $\tilde{A}$ has at most dimension k.

Matrix Norm Expressions

$$||A||_{Fro}^2 = \sum \sigma_i^2$$

$$||A||_2 = max\{\sigma_i\}$$

$$cond\ A = \sigma_{max}/\sigma_{min}$$

Revisiting Tikhonov Regularization

Regularized least-squares problem:

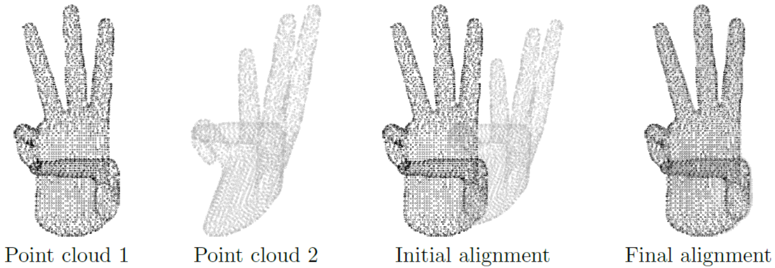
$$(A^\top A + \alpha I)\vec{x} = A^\top \vec{b}.$$

Perform SVD analysis.

What does α do to the singular values?

Example: Vandermonde matrix, V

Rigid Alignment



Variational Formulation

Given $\vec{x}_{1i} \mapsto \vec{x}_{2i}$

$$\min_{R^\top R = I_{3\times 3}, \vec{t} \in \mathbb{R}^3} \sum_i \|R\vec{x}_{1i} + \vec{t} - \vec{x}_{2i}\|_2^2$$

Variational Formulation

Given $\vec{x}_{1i} \mapsto \vec{x}_{2i}$

$$\min_{R^\top R = I_{3\times 3}, \vec{t} \in \mathbb{R}^3} \sum_i ||R\vec{x}_{1i} + \vec{t} - \vec{x}_{2i}||_2^2$$

Alternate:

1. Minimize with respect to $\vec{t}$: Least-squares
2. Minimize with respect to R: SVD

Procrustes via SVD

$$\min_{R^\top R = I_{3\times 3}} ||RX_1 - X_2^t||_{Fro}^2$$

Procrustes via SVD

$$\min_{R^\top R = I_{3\times 3}} \|RX_1 - X_2^t\|_{Fro}^2$$

Orthogonal Procrustes Theorem

The orthogonal matrix R minimizing $\|RX - Y\|^2$ is given by $UV^\top$, where SVD is applied to factor $YX^\top = U\Sigma V^\top$.

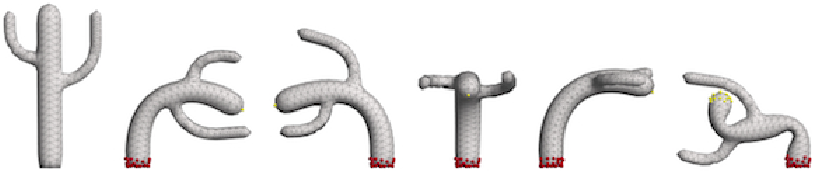
Application: As-Rigid-As-Possible

As-Rigid-As-Possible Surface Modeling

Olga Sorkine and Marc Alexa

Eurographics/ACM SIGGRAPH Symposium on Geometry Processing 2007.

http://www.youtube.com/watch?v=ltX-qUjbkdc



Related: Polar Decomposition

$$F = Ru$$

Special case:

- F is square real-valued matrix;
- R is best rotation matrix approximation;
- U is right symmetric PSD stretch matrix.
- Proof by SVD.

Recall: Statistics Problem

Given: Collection of data points $\vec{x}_i$

- Age
- Weight
- Blood pressure
- Heart rate

Find: Correlations between different dimensions

Simplest Model

One-dimensional subspace

$$\vec{x}_i \approx c_i \vec{v}$$

More General Statement

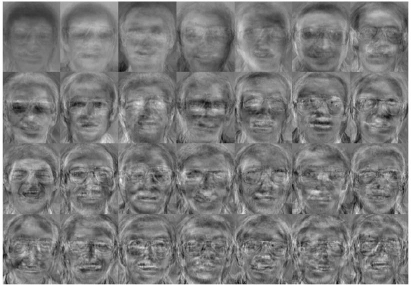
Principal Component Analysis

The matrix $C \in \mathbb{R}^{n\times d}$ minimizing $||X - CC^\top X||_{Fro}$ subject to $C^\top C = I_{d\times d}$ is given by the first d columns of U, for $X = U\Sigma V^\top$.

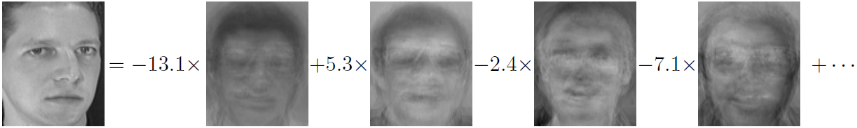
Proved in textbook.

Application: Eigenfaces

(a) Input faces



(b) Eigenfaces



(c) Projection