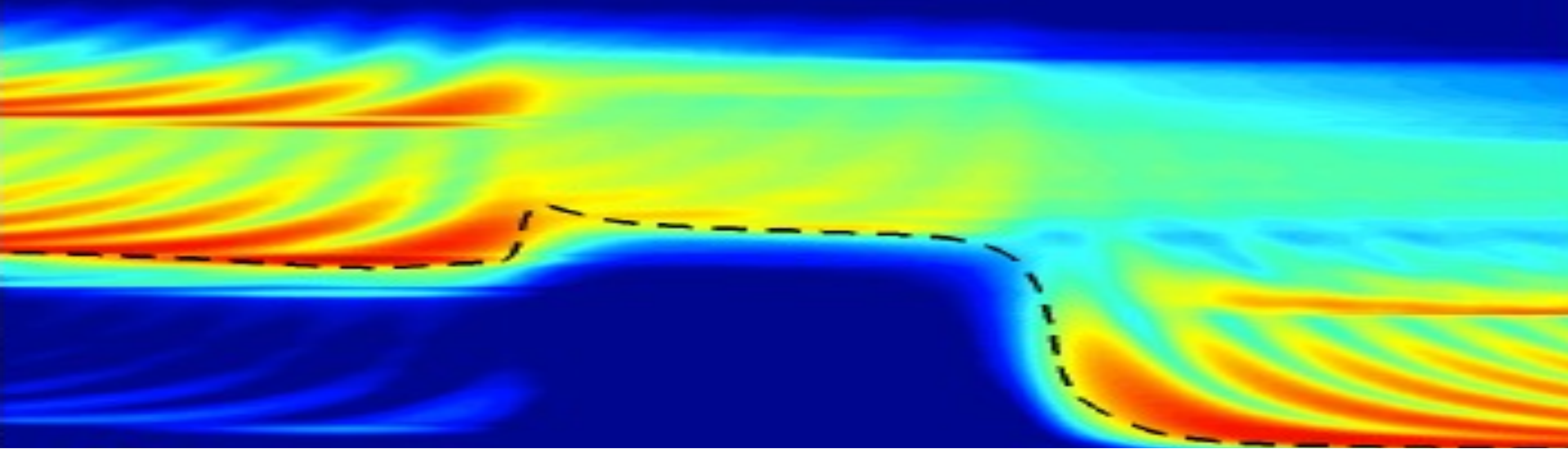


EECS 16B
Designing Information Devices and Systems II
Lecture 21

Prof. Yi Ma
Department of Electrical Engineering and Computer Sciences, UC Berkeley,
yima@eecs.berkeley.edu

Outline

- Motivations for Singular Value Decomposition
  - Least Squares and Minimum Norm Solution
  - Identifying Low-dim Linear Subspace
- Singular Value Decomposition (SVD)
  - Algorithm
  - Example
  - Theorem (with proof)

Least-Squares vs Minimum-Norm Solutions

Moore-Penrose pseudo inverse of $A \in \mathbb{R}^{m\times n}$: $\vec{y} = A\vec{x}$, $\vec{x} = A^{\dagger}\vec{y}$.

$y = Ax$

$m \geq n$ and $\mathrm{rank}(A) = n$: $A^{\dagger} = \underline{(A^{\top}A)^{-1}A^{\top}}$

system id

$y = \boxed{A}\,\underline{x} + e$

$m \leq n$ and $\mathrm{rank}(A) = m$: $A^{\dagger} = A^{\top}\underline{(AA^{\top})}^{-1}$

control

$y = \boxed{A}\,\underline{x}$

$\min \|x\|_2^2$ s.t. $y = Ax$.

$A \in \mathbb{R}^{m\times n}$ not full column or row rank?

$\underline{y = Ax}$

$\boxed{A}_{m\times n} = \boxed{W}\,\boxed{F}$

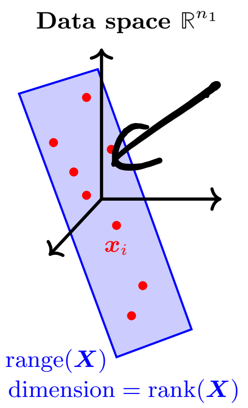
Low-Dim Structures in High-Dim Data

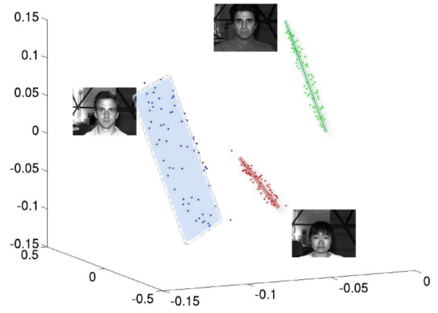
Principal Component Analysis: Finding one linear subspace

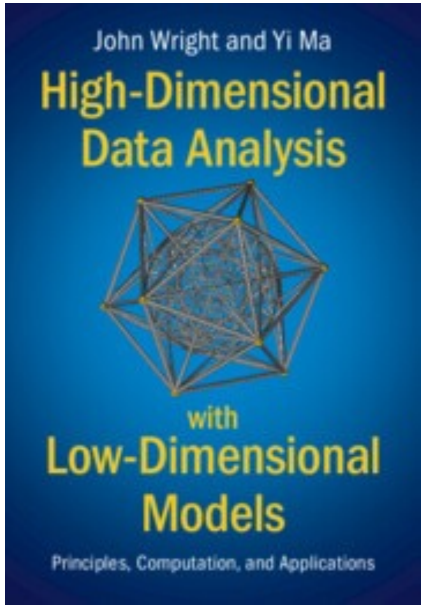
Compressive Sensing: Finding multiple low-dim linear structures

- Solving under-determined systems of linear equations
- Low-rank matrix approximation or recovery

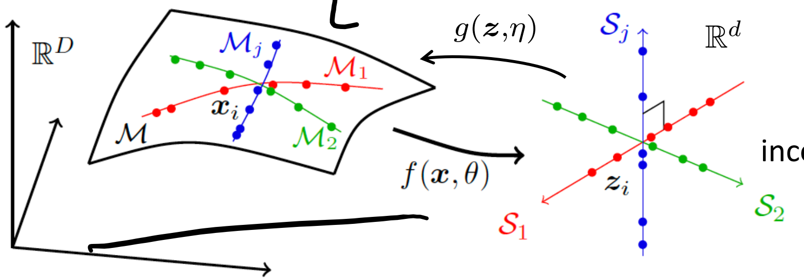
Deep Learning: Finding non-linear low-dimensional structures







A mixture of
low-dim submanifolds



A mixture of
incoherent subspaces

EECS 208: [Computational Principles for High-Dimensional Data Analysis]

(from SVD/PCA, to Generalized PCA, Robust PCA, Nonlinear PCA, and to Deep Networks…)

$\langle \vec{u}, \vec{v} \rangle = \vec{u}^T \vec{v}$

Identifying a Low-dim Linear Subspace

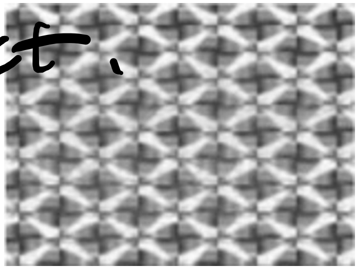
Given $X = [\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n] \in \mathbb{R}^{m\times n}$, find a low rank L: $\min_L \|X - L\|_F^2, \ \text{s.t.} \ \text{rank}(L) \le r.$

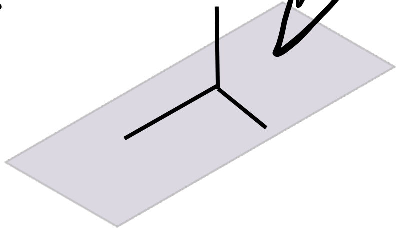
$\text{rank}(X) = 1$

$\tilde{x}_i \propto \vec{u}$ — normalized

$\vec{x}_i = v_i \vec{u}$

outer product.





$X = \vec{u}\,[v_1, v_2, \dots, v_n] = \vec{u}\cdot\vec{v}^T$

m $\vec{v}^T$ n

$X = \alpha \cdot \vec{u} \cdot \vec{v}^T$ — $\vec{u}, \vec{v}$ normalized

$X = [\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n] \in \mathbb{R}^{m\times n}$

$\text{rank}(X) = 2, 3, \dots, r.$

orthogonal

$\vec{u}_1, \vec{u}_2, \dots, \vec{u}_r$ linearly ind

$\vec{v}_1, \vec{v}_2, \dots, \vec{v}_r$ linearly ind.

$X = \alpha_1 \vec{u}_1 \vec{v}_1^T + \alpha_2 \vec{u}_2 \vec{v}_2 + \dots + \alpha_r \vec{u}_r \vec{v}_r^T$

atoms

$\alpha_i \neq 0$.

$\alpha_i > 0$?

Singular Value Decomposition (SVD)

Given $A \in \mathbb{R}^{m \times n}$ with $\mathrm{rank}(A) = r$, we like to decompose it into a special outer-product form:

$$\underline{A} = \underline{\sigma_1 \vec{u}_1 \vec{v}_1^\top + \sigma_2 \vec{u}_2 \vec{v}_2^\top + \cdots + \sigma_r \vec{u}_r \vec{v}_r^\top}$$

$\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r$ orthonormal in $\mathbb{R}^m$

$\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r$ orthonormal in $\mathbb{R}^n$

$\underline{\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0}$

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix} = \underline{\begin{bmatrix} 1 \\ 2 \end{bmatrix}} \underline{[1 \; 1 \; 1]}$$

$$= \underset{6}{\sqrt{15}} \cdot \underbrace{\left(\frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right)}_{\vec{u}} \underbrace{\left(\frac{1}{\sqrt{3}} [1, \; 1, \; 1] \right)}_{\vec{v}^\top}$$

Singular Value Decomposition (SVD)

Given $A \in \mathbb{R}^{m\times n}$ with $\mathrm{rank}(A) = r$, we like to decompose it into a special matrix form:

$$A = \underbrace{[\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r]}_{U_{m\times r}} \underbrace{\begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_r \end{bmatrix}}_{\Sigma_{r\times r}} \underbrace{\begin{bmatrix} \vec{v}_1^\top \\ \vec{v}_2^\top \\ \vdots \\ \vec{v}_r^\top \end{bmatrix}}_{V^T_{r\times n}} = U\Sigma V^T$$

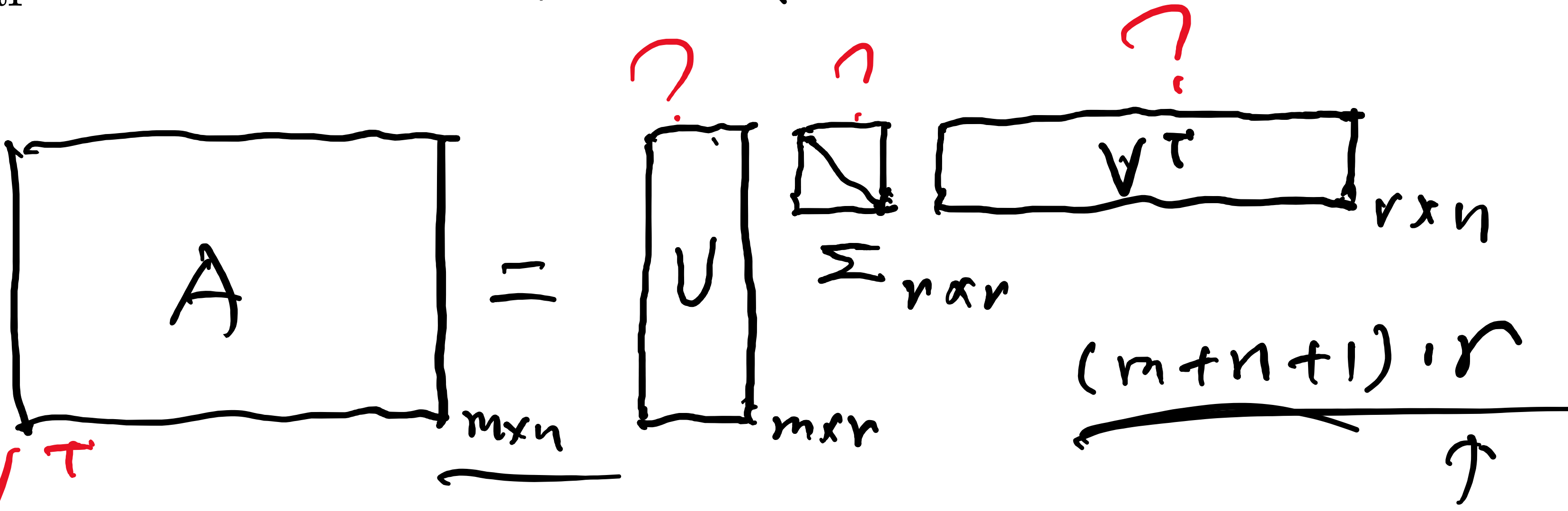
$U^T U = I$

$U = [\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r]$ orthogonal

$V = [\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r]$ orthogonal

$\Sigma = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_r\} > 0$

$$A^T A = (U\Sigma V^T)^T (U\Sigma V^T)$$
$$= V\Sigma \underline{U^T U} \Sigma V^T$$
$$= V \underline{\Sigma^2} V^T = V \underline{\Lambda} V^T$$



Singular Value Decomposition (SVD)

Claim: $A^\top A \in \mathbb{R}^{n\times n}$ all eigenvalues are non-negative and eigenvectors are orthogonal.

Proof: A^TA real symmetric

$$A^TA = V_n \Lambda V_n^T$$

$$= \sum_{i=1}^{n} \vec{v}_i \lambda_i \vec{v}_i^T$$

$$(A^TA)\vec{v}_i = \lambda_i \vec{v}_i \qquad \lambda_i \neq 0$$

$$\vec{v}_i^T (A^TA)\vec{v}_i = \lambda_i \vec{v}_i^T \vec{v}_i$$

$$V_n^T (A^TA) V_n = \begin{bmatrix} \begin{matrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{matrix} & 0 \\ 0 & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \end{bmatrix}$$

orthogonal

Λ

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots \geq \lambda_r > 0 \;?$$

$$\lambda_i = (\vec{v}_i^T A^T)(A\vec{v}_i)$$

$$= \|A\vec{v}_i\|_2^2 > 0.$$

$\square$

Singular Value Decomposition (SVD)

Claim: $\underline{\operatorname{rank}(A^\top A)} = \underline{\operatorname{rank}(A)} = r$ hence r eigenvalues of $A^\top A$ are positive.

Proof: ① $\underline{Null(A) \subseteq Null(A^TA)}$

$A\vec{v} = 0 \Rightarrow A^T \underline{Av} = 0$

② $Null(A) \supseteq Null(A^TA)$

$\underline{A^T} \underline{A\vec{v}} = 0 \Longrightarrow \vec{v}^T A^T A \vec{v} = 0$

$\Longrightarrow \|A\vec{v}\|_2^2 = 0$

$\Longrightarrow A\vec{v} = 0$.

$\mathbb{R}^n$

$\underline{A^TA_{n\times n}}$

$\underline{A_{m\times n}}$

$rank(A) = r$

$Null(A) = n - r$

$\underline{rank(A^TA) = r}$.

A^TA r positive eigen.

AA^T

Singular Value Decomposition (Algorithm)

Algorithm: given $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = r$, start with $A^\top A \in \mathbb{R}^{n\times n}$:

If $A = U\Sigma V^T$?

$A^TA = V\Lambda V^T = V\Sigma^2 V^T$

$\Sigma = \sqrt{\Lambda}$, $\sigma_i = \sqrt{\lambda_i}$, $i=1,\dots,r$

$AV = U\Sigma V^T V$

$AV = U\Sigma$ $\quad U = AV\Sigma^{-1}$

$A\vec{v}_i = \vec{u}_i\sigma_i$

$\vec{u}_i = \frac{1}{\sigma_i}A\vec{v}_i$

step 1: $A^TA = V_n\Lambda V_n^T$

$$V_n^T(A^TA)V_n = \begin{bmatrix} \lambda_1 & & & & 0 \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \lambda_r & \\ 0 & & & & 0 \ddots 0 \end{bmatrix}$$

$\lambda_1 \geq \lambda_2 \geq \dots \lambda_r > 0$

$V = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_r]$

step 2: $\sigma_i = \sqrt{\lambda_i}$, $i=1,\dots,r$.

$\vec{u}_i = \frac{1}{\sigma_i}A\vec{v}_i$, $i=1,\dots,r$.

$U = \{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_r]$?

Singular Value Decomposition (Example)

$A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$ $\vec{v}$ rank $=2$ $\qquad A^TA = \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}$

① $\lambda I - A^TA = \begin{bmatrix} \lambda-25 & -7 \\ -7 & \lambda-25 \end{bmatrix} = (\lambda-25)^2 - 7^2 = \lambda - 25 \pm 7 = 0$

$\underline{\lambda_1 = 32}, \underline{\lambda_2 = 18}$

> 0

$\lambda_1 I - A^TA = \begin{bmatrix} 7 & -7 \\ -7 & 7 \end{bmatrix} \quad \vec{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$\lambda_2 I - A^TA = \begin{bmatrix} -7 & -7 \\ -7 & -7 \end{bmatrix} \quad \vec{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

V

② $\sigma_1 = \sqrt{\lambda_1} = 4\sqrt{2} \qquad \sigma_2 = \sqrt{\lambda_2} = 3\sqrt{2} \qquad \vec{u}_2 = \frac{1}{\sigma_2} A\vec{v}_2$

$\vec{u}_1 = \frac{1}{\sigma_1} A\vec{v}_1 = \frac{1}{4\sqrt{2}}\frac{1}{\sqrt{2}}\begin{bmatrix} 8 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{3\sqrt{2}}\frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ -6 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

Singular Value Decomposition (Example)

$\vec{u}_1 \perp \vec{u}_2$

$$A = \underbrace{4\sqrt{2}}_{\sigma_1} \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\vec{u}_1} \underbrace{\frac{1}{\sqrt{2}}[1, 1]}_{\vec{v}_1} + 3\sqrt{2} \underbrace{\begin{bmatrix} 0 \\ -1 \end{bmatrix}}_{\vec{u}_2} \underbrace{\frac{1}{\sqrt{2}}[1, -1]}_{\vec{v}_2}$$

$$= \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$$

<u>atomic</u>

$\sigma \vec{u} \vec{v}^T = \sigma (-\vec{u})(-v^T)$

Singular Value Decomposition (Theorem)

Theorem: given $A \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(A) = r$, let $A^\top A = \sum_{i=1}^{r} \lambda_i \vec{v}_i \vec{v}_i^\top$ and $\sigma_i = \sqrt{\lambda_i}$,

$\vec{u}_i = \frac{1}{\sigma_i} A\vec{v}_i \in \mathbb{R}^m$, $i = 1, \ldots, r$. Then we have $U = [\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r]$ orthogonal, and

$$A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top = U \Sigma V^\top$$

$$\Sigma = \operatorname{diag}\{\sigma_1, \ldots, \sigma_r\} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_r \end{bmatrix}$$

Proof: (1) $\vec{u}_i \perp \vec{u}_j \quad i \neq j$

$$\vec{u}_i^T \vec{u}_j = \left(\frac{1}{\sigma_i} A \vec{v}_i\right)^T \left(\frac{1}{\sigma_j} A \vec{v}_j\right) = \frac{1}{\sigma_i \sigma_j} v_i^T \underbrace{A^T A \vec{v}_j}_{\lambda_j \vec{v}_j}$$

$$= \frac{1}{\sigma_i \sigma_j} v_i^T \lambda_j \vec{v}_j = 0 \quad i \neq j$$

$$= \frac{\lambda_i}{\sigma_i^2} \quad i = j$$

$$= 1$$

$\boxed{\Sigma}$

Singular Value Decomposition (Theorem)

Theorem: given $A \in \mathbb{R}^{m\times n}$ with $\mathrm{rank}(A) = r$, let $A^\top A = \sum_{i=1}^{r} \lambda_i \vec{v}_i \vec{v}_i^\top$ and $\sigma_i = \sqrt{\lambda_i}$, $\vec{u}_i = \frac{1}{\sigma_i} A\vec{v}_i \in \mathbb{R}^m,\ i = 1, \ldots, r$. Then we have $U = [\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r]$ $\mathrm{orthogonal}$, and

$$A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top = U\Sigma V^\top$$

Proof: