Thm (Review): A matrix $A \in \mathbb{R}^n$ is symmetric

$$\iff A = Q\,D\,Q^T = Q \begin{pmatrix} \ddots & & \\ & & \\ & & \end{pmatrix} Q^T, \text{ with orthogonal } Q.$$

Thm: Let matrix $A \in \mathbb{R}^{n \times n}$ be symmetric, then

$$M \stackrel{def}{=} \mathbf{max}_{\|\mathbf{x}\|=1} \mathbf{x}^T A \mathbf{x} \quad \text{is the largest eigenvalue of } A,$$

$$m \stackrel{def}{=} \mathbf{min}_{\|\mathbf{x}\|=1} \mathbf{x}^T A \mathbf{x} \quad \text{is the least eigenvalue of } A.$$

Proof: Write $A = Q\,D\,Q^T$, with orthogonal matrix $Q \in \mathbb{R}^{n \times n}$ and diagonal matrix $D = \mathbf{diag}\,(\lambda_1, \cdots, \lambda_n)$ with eigenvalues.

Thm: Let matrix $A \in \mathbb{R}^{n \times n}$ be symmetric, then

$$M \stackrel{def}{=} \mathbf{max}_{\|\mathbf{x}\|=1} \mathbf{x}^T A \mathbf{x} \quad \text{is the largest eigenvalue of } A,$$

$$m \stackrel{def}{=} \mathbf{min}_{\|\mathbf{x}\|=1} \mathbf{x}^T A \mathbf{x} \quad \text{is the least eigenvalue of } A.$$

Proof: Write $A = Q\,D\,Q^T$, with orthogonal matrix $Q \in \mathbb{R}^{n \times n}$ and diagonal matrix $D = \mathbf{diag}(\lambda_1, \cdots, \lambda_n)$ with eigenvalues.

- Define change of variable $\mathbf{y} = Q^T \mathbf{x}$. Then $\|\mathbf{y}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$,

$$\text{and} \quad M = \mathbf{max}_{\|\mathbf{y}\|=1} \mathbf{y}^T D \mathbf{y}, \quad m = \mathbf{min}_{\|\mathbf{y}\|=1} \mathbf{y}^T D \mathbf{y}.$$

Proof: Write $A = Q\,D\,Q^T$, $D = \mathbf{diag}\,(\lambda_1, \cdots, \lambda_n)$ with eigenvalues,

$$\text{and} \quad M = \mathbf{max}_{\|\mathbf{y}\|=1}\, \mathbf{y}^T\, D\, \mathbf{y}, \quad m = \mathbf{min}_{\|\mathbf{y}\|=1}\, \mathbf{y}^T\, D\, \mathbf{y}, \quad \text{for} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

$$\mathbf{y}^T\, D\, \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}^T \mathbf{diag}\,(\lambda_1, \cdots, \lambda_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \lambda_1\, y_1^2 + \cdots + \lambda_n\, y_n^2.$$

Let $\lambda_{\mathbf{max}} = \mathbf{max}\,\{\lambda_1, \cdots, \lambda_n\} = \lambda_{\ell_1}$, $\lambda_{\mathbf{min}} = \mathbf{min}\,\{\lambda_1, \cdots, \lambda_n\} = \lambda_{\ell_2}$, then

$$\lambda_{\mathbf{min}}\,\left(y_1^2 + \cdots + y_n^2\right) \le \lambda_1\, y_1^2 + \cdots + \lambda_n\, y_n^2 \le \lambda_{\mathbf{max}}\,\left(y_1^2 + \cdots + y_n^2\right),$$

$$\text{or,} \quad \lambda_{\mathbf{min}}\,\|\mathbf{y}\|^2 \le \mathbf{y}^T\, D\, \mathbf{y} \le \lambda_{\mathbf{max}}\,\|\mathbf{y}\|^2.$$

$$\text{So for all } \|\mathbf{y}\| = 1, \quad \lambda_{\mathbf{min}} \le m \le \mathbf{y}^T\, D\, \mathbf{y} \le M \le \lambda_{\mathbf{max}}.$$

Proof: For $D = \mathbf{diag}(\lambda_1, \cdots, \lambda_n)$ with eigenvalues,

$$M = \mathbf{max}_{\|\mathbf{y}\|=1} \mathbf{y}^T D \mathbf{y}, \quad m = \mathbf{min}_{\|\mathbf{y}\|=1} \mathbf{y}^T D \mathbf{y}, \text{ for } \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

$$\mathbf{y}^T D \mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2.$$

Let $\lambda_{\mathbf{max}} = \mathbf{max}\{\lambda_1, \cdots, \lambda_n\} = \lambda_{\ell_1}$, $\lambda_{\mathbf{min}} = \mathbf{min}\{\lambda_1, \cdots, \lambda_n\} = \lambda_{\ell_2}$, then

$$\text{for all } \|\mathbf{y}\| = 1, \quad \lambda_{\mathbf{min}} \leq m \leq \mathbf{y}^T D \mathbf{y} \leq M \leq \lambda_{\mathbf{max}}.$$

- Let $\mathbf{e}_j$ be the j^{th} column of the identity.
  - Choose $\mathbf{y} = \mathbf{e}_{\ell_1}$, then $M \geq \mathbf{y}^T D \mathbf{y} = \lambda_{\mathbf{max}}$.
  - Choose $\mathbf{y} = \mathbf{e}_{\ell_2}$, then $m \leq \mathbf{y}^T D \mathbf{y} = \lambda_{\mathbf{min}}$.
- Therefore $M = \lambda_{\mathbf{max}}$, $m = \lambda_{\mathbf{min}}$.

Thm: Let matrix $A \in \mathbb{R}^{n\times n}$ be symmetric, then

$$M \overset{def}{=} \mathbf{max}_{\|\mathbf{x}\|=1}\, \mathbf{x}^T A\, \mathbf{x} \quad \text{is the largest eigenvalue of } A,$$
$$m \overset{def}{=} \mathbf{min}_{\|\mathbf{x}\|=1}\, \mathbf{x}^T A\, \mathbf{x} \quad \text{is the least eigenvalue of } A.$$

EXAMPLE: Matrix $A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix} \in \mathcal{R}^{3\times 3}$ is symmetric with eigenvalues $\lambda_1 = 6,\ \lambda_2 = 3,\ \lambda_3 = 1$ and unit eigenvectors

$$\mathbf{u}_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},\ \mathbf{u}_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix},\ \mathbf{u}_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}.$$

Therefore

$$M = \mathbf{u}_1^T A\, \mathbf{u}_1 = 6, \quad m = \mathbf{u}_3^T A\, \mathbf{u}_3 = 1.$$

Thm (Review with Proof): A matrix $A \in \mathbb{R}^n$ is symmetric

$$\iff A = Q\,D\,Q^T = Q \begin{pmatrix} \ddots & & \\ & & \\ & & \end{pmatrix} Q^T, \text{ with orthogonal } Q.$$

Thm (REVIEW WITH PROOF): A matrix $A \in \mathbb{R}^n$ is symmetric

$$\iff A = Q\,D\,Q^T = Q \begin{pmatrix} \diagdown & & \\ & & \\ & & \end{pmatrix} Q^T, \text{ with orthogonal } Q.$$

Proof: Define $\lambda = \max_{\|\mathbf{x}\|=1} \mathbf{x}^T A\,\mathbf{x}$. Then λ is an eigenvalue,

$$A\,\mathbf{u} = \lambda\,\mathbf{u}, \quad \text{with UNIT eigenvector} \quad \mathbf{u}.$$

- Extend $\mathbf{u}$ into an orthonormal basis for $\mathbb{R}^n$:
  $\mathbf{u}, \mathbf{u}_2, \cdots, \mathbf{u}_n$ are unit and mutually orthogonal vectors,
- $$U \stackrel{def}{=} (\mathbf{u}, \mathbf{u}_2, \cdots, \mathbf{u}_n) \stackrel{def}{=} \left(\mathbf{u}, \widehat{U}\right) \in \mathbb{R}^{n\times n} \quad \text{is orthogonal.}$$

$$U^T A\,U = \begin{pmatrix} \lambda & \mathbf{0}^T \\ \mathbf{0} & \widehat{U}^T \left(A\,\widehat{U}\right) \end{pmatrix}, \quad \widehat{U}^T \left(A\,\widehat{U}\right) \text{ is symmetric.}$$

- Repeat same procedure on $\widehat{U}^T \left(A\,\widehat{U}\right)$.

Let matrix $A \in \mathbb{R}^n$ be symmetric.

Thm : Let λ_1 be largest eigenvalue of A with unit eigenvector $\mathbf{u}_1$.

Then, $\widehat{M} \stackrel{def}{=} \mathbf{max}_{\|\mathbf{x}\|=1,\, \mathbf{x}^T \mathbf{u}_1=0}\, \mathbf{x}^T A\, \mathbf{x}$ is SECOND largest eigenvalue of A.

Let matrix $A \in \mathbb{R}^n$ be symmetric.

Thm : Let λ_1 be largest eigenvalue of A with unit eigenvector $\mathbf{u}_1$.

Then, $\widehat{M} \stackrel{def}{=} \mathbf{max}_{\|\mathbf{x}\|=1,\, \mathbf{x}^T \mathbf{u}_1=0}\, \mathbf{x}^T A\, \mathbf{x}$ is SECOND largest eigenvalue of A.

Proof: Extend $\mathbf{u}_1$ into an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$,

- $U \stackrel{def}{=} (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n) \stackrel{def}{=} \left(\mathbf{u}_1, \widehat{U}\right) \in \mathbb{R}^{n\times n}$ is orthogonal.

$$U^T A\, U = \begin{pmatrix} \lambda_1 & \mathbf{0}^T \\ \mathbf{0} & \widehat{U}^T \left(A\, \widehat{U}\right) \end{pmatrix}, \quad \widehat{U}^T \left(A\, \widehat{U}\right) \text{ is symmetric.}$$

- But $\mathbf{u}_2, \cdots, \mathbf{u}_n$ is orthonormal basis for $(\mathbf{Span}\,\{\mathbf{u}_1\})^{\perp}$.

$$\text{Thus,} \quad \|\mathbf{x}\| = 1,\ \mathbf{x}^T \mathbf{u}_1 = 0 \iff \mathbf{x} = \widehat{U}\, \mathbf{y},\ \ \|\mathbf{y}\| = 1.$$

$$\Longrightarrow \quad \widehat{M} = \mathbf{max}_{\|\mathbf{y}\|=1}\, \mathbf{y}^T \left(\widehat{U}^T A\, \widehat{U}\right) \mathbf{y} = \text{ largest eigenvalue of } \widehat{U}^T A\, \widehat{U},$$

and therefore SECOND largest eigenvalue of A.

Let matrix $A \in \mathcal{R}^{n \times n}$ be symmetric.

Thm : Let $\lambda_1, \cdots, \lambda_{k-1}$ be largest $k-1$ eigenvalues of A with unit eigenvectors $\mathbf{u}_1, \cdots, \mathbf{u}_{k-1}$. Then,

$$\widehat{M} \stackrel{def}{=} \mathbf{max}_{\|\mathbf{x}\|=1,\, \mathbf{x}^T \mathbf{u}_1=0, \cdots, \mathbf{x}^T \mathbf{u}_{k-1}=0} \mathbf{x}^T A \mathbf{x} \quad \text{is } k^{th} \text{ largest eigenvalue of } A.$$

Let matrix $A \in \mathcal{R}^{n \times n}$ be symmetric.

Thm : Let $\lambda_1, \cdots, \lambda_{k-1}$ be largest $k-1$ eigenvalues of A with unit eigenvectors $\mathbf{u}_1, \cdots, \mathbf{u}_{k-1}$. Then,

$$\widehat{M} \stackrel{def}{=} \mathbf{max}_{\|\mathbf{x}\|=1,\, \mathbf{x}^T \mathbf{u}_1=0, \cdots, \mathbf{x}^T \mathbf{u}_{k-1}=0}\, \mathbf{x}^T A\, \mathbf{x} \quad \text{is } k^{th} \text{ largest eigenvalue of } A.$$

Proof: Extend $\mathbf{u}_1, \cdots, \mathbf{u}_{k-1}$ into an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$,

- $U \stackrel{def}{=} (\mathbf{u}_1, \cdots, \mathbf{u}_{k-1}, \mathbf{u}_k, \cdots, \mathbf{u}_n) \stackrel{def}{=} \left(\widehat{U}_1, \widehat{U}_2\right) \in \mathcal{R}^{n \times n}$ orthogonal.

$$U^T A\, U = \begin{pmatrix} D_1 & \mathbf{0}^T \\ \mathbf{0} & \widehat{U}_2^T \left(A\, \widehat{U}_2\right) \end{pmatrix}, \quad \text{with } D_1 \stackrel{def}{=} \mathbf{diag}\,(\lambda_1, \cdots, \lambda_{k-1})$$

- $\mathbf{u}_k, \cdots, \mathbf{u}_n$ is orthonormal basis for $(\mathbf{Span}\,\{\mathbf{u}_1, \cdots, \mathbf{u}_{k-1}\})^{\perp}$.

Thus, $\quad \|\mathbf{x}\| = 1,\ \mathbf{x}^T \mathbf{u}_1 = 0, \cdots, \mathbf{x}^T \mathbf{u}_{k-1} = 0 \iff \mathbf{x} = \widehat{U}_2\, \mathbf{y},\ \ \|\mathbf{y}\| = 1.$

$$\Longrightarrow \quad \widehat{M} = \mathbf{max}_{\|\mathbf{y}\|=1}\, \mathbf{y}^T \left(\widehat{U}_2^T A\, \widehat{U}_2\right) \mathbf{y} = \text{ largest eigenvalue of } \widehat{U}_2^T A\, \widehat{U}_2,$$

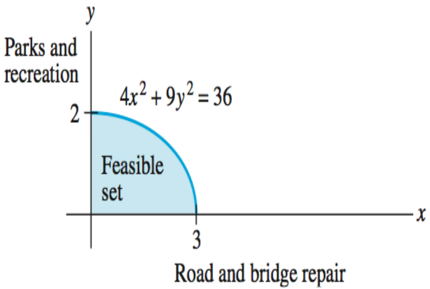
and therefore k^{th} largest eigenvalue of A.

Example: Public repair works planning (I)

public roads/bridges:	x	(hundred miles)
public recreation areas:	y	(hundred acres)
cost:	$4\,x^2 + 9\,y^2$	
Available Resources:	36	
Utility (effectiveness):	$x\,y$	

SOLUTION: Choose x and y to maximize the utility

$$\mathbf{max}_{x,y\geq 0,\ 4\,x^2+9\,y^2\leq 36}\ x\,y$$



Example: Public repair works planning (II)

SOLUTION: Choose x and y to maximize the utility

$$\mathbf{max}_{x,y\geq 0,\ 4\,x^2+9\,y^2\leq 36}\, x\, y$$

Define $\mathbf{x} = \frac{1}{6}\begin{pmatrix} 2\,x \\ 3\,y \end{pmatrix}$ and $A = \begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix}$. Problem becomes

$$\mathbf{max}_{\|\mathbf{x}\|\leq 1}\, \mathbf{x}^T A\, \mathbf{x} = \mathbf{max}_{\|\mathbf{x}\|=1}\, \mathbf{x}^T A\, \mathbf{x}.$$

Maximum is largest eigenvalue, 3, of matrix A:

$$\begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = 3\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix},$$

optimal solution

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{3}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} \end{pmatrix}.$$

Example: Maximum length of linear transform

Let $A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$. Solve problem

$$\mathbf{max}_{\|\mathbf{x}\|=1} \; \|A\,\mathbf{x}\| \,.$$

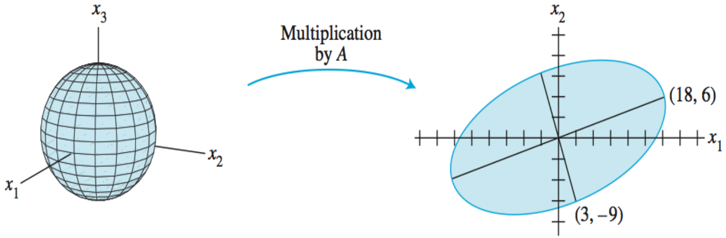


FIGURE 1 A transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$.

Let $A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$. Solve problem

$$\max_{\|\mathbf{x}\|=1} \|A\,\mathbf{x}\|\,.$$

SOLUTION: Re-write problem as

$$\max_{\|\mathbf{x}\|=1} \|A\,\mathbf{x}\| = \sqrt{\max_{\|\mathbf{x}\|=1} \|A\,\mathbf{x}\|^2} = \sqrt{\max_{\|\mathbf{x}\|=1} \ \mathbf{x}^T \left(A^T A\right) \mathbf{x}}.$$

$\Longrightarrow$ Maximum is largest eigenvalue of matrix $A^T A$:

$$A^T A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}^T \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}.$$

The eigenvalues of matrix $A^T A$ are $\lambda_1 = 360, \lambda_2 = 90, \lambda_3 = 0$,

with unit eigenvector $\mathbf{v}_1 = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$, $A^T A\,\mathbf{v}_1 = \lambda_1\,\mathbf{v}_1$, $\|A\,\mathbf{v}_1\| = \sqrt{360}$.

optimal solution $\mathbf{x} = \mathbf{v}_1$ and optimal value $\|A\,\mathbf{v}_1\| = \sqrt{360}$.

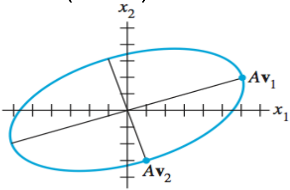
$$A^T A = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}, \quad \text{with eigenvalues } \lambda_1 = 360, \lambda_2 = 90, \lambda_3 = 0.$$

$$\mathbf{v}_1 = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \quad \mathbf{v}_2 = \frac{1}{3}\begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix}, \quad A^T A\,\mathbf{v}_1 = \lambda_1 \mathbf{v}_1, \quad A^T A\,\mathbf{v}_2 = \lambda_2 \mathbf{v}_2.$$

$$A\,\mathbf{v}_1 = \frac{1}{3}\begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = \sqrt{360}\,\mathbf{u}_1, \quad \mathbf{u}_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ 1 \end{pmatrix},$$

$$A\,\mathbf{v}_2 = \frac{1}{3}\begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}\begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix} = \sqrt{90}\,\mathbf{u}_2, \quad \mathbf{u}_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ -3 \end{pmatrix}.$$

$\mathbf{u}_1, \mathbf{u}_2$ orthonormal.



§7.4 Singular Value Decomposition

Let $A \in \mathbb{R}^{m \times n}$. The SVD of A is

$$A = U\,S\,V^T = U \begin{pmatrix} \diagdown & \\ & \end{pmatrix} V^T,$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices,

$$U^T\,U = I_m, \quad V^T\,V = I_n, \quad \text{and}$$

$$S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m \times n} \text{ is diagonal with non-negative entries.}$$

Rank and linear independence

- **Def:** The RANK of $A \in \mathbb{R}^{m\times n}$, denoted **Rank**$(A)$, is the number of linearly independent COLUMNS in A.
- **Thm:** $\mathbf{Rank}(A) = \mathbf{Rank}(A^T A)$
- Eigenvalues of $A^T A$ are real and non-negative.

Proof: For any $\mathbf{v}$ in $\mathbb{R}^n$:

$$\begin{aligned} A\mathbf{v} &= \mathbf{0}, \quad \longrightarrow \quad A^T A\mathbf{v} = \mathbf{0}, \\ A^T A\mathbf{v} &= \mathbf{0}, \quad \longrightarrow \quad \|A\mathbf{v}\|_2^2 = \mathbf{v}^T A^T A\mathbf{v} = 0, \quad \longrightarrow \quad A\mathbf{v} = \mathbf{0}. \end{aligned}$$

therefore $A\mathbf{v} = \mathbf{0} \Longleftrightarrow A^T A\mathbf{v} = \mathbf{0}, \implies \mathrm{Nul}\ A = \mathrm{Nul}\ A^T A.$

$$\mathbf{Rank}(A) = n - \mathbf{dim}\,(\mathrm{Nul}\ A) = n - \mathbf{dim}\left(\mathrm{Nul}\ A^T A\right) = \mathbf{Rank}(A^T A).$$

Let λ be eigenvalue of $A^T A$ and $\mathbf{u} \in \mathbb{R}^n$ unit eigenvector:

$$A^T A\mathbf{u} = \lambda\mathbf{u}, \implies \lambda = \mathbf{u}^T(\lambda\mathbf{u}) = \mathbf{u}^T\left(A^T A\mathbf{u}\right) = \|A\mathbf{u}\|_2^2 \geq 0.$$

Constructing the SVD for $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

- Let $A^T A = V\, D\, V^T$ be eigendecomposition, with

$$D = \textbf{diag}\left(\lambda_1, \lambda_2, \cdots, \lambda_n\right) \quad \text{with} \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

be eigenvalues and $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ be orthogonal.

$$\text{So} \quad A^T A\, \mathbf{v}_j = \lambda_j\, \mathbf{v}_j \quad \text{for} \quad j = 1, \cdots, n.$$

Constructing the SVD for $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

- Let $A^T A = V D V^T$ be eigendecomposition, with

$$D = \mathbf{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \quad \text{with} \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

be eigenvalues and $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ be orthogonal.

$$\text{So} \quad A^T A \mathbf{v}_j = \lambda_j \mathbf{v}_j \quad \text{for} \quad j = 1, \cdots, n.$$

- Define $\sigma_j = \sqrt{\lambda_j}$ for $j = 1, \cdots, n$. Let k be such that $\sigma_k > 0$ and $\sigma_{k+1} = 0$.

Constructing the SVD for $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

- Let $A^T A = V D V^T$ be eigendecomposition, with

  $$D = \mathbf{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \quad \text{with} \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

  be eigenvalues and $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ be orthogonal.

  $$\text{So} \quad A^T A \mathbf{v}_j = \lambda_j \mathbf{v}_j \quad \text{for} \quad j = 1, \cdots, n.$$

- Define $\sigma_j = \sqrt{\lambda_j}$ for $j = 1, \cdots, n$. Let k be such that $\sigma_k > 0$ and $\sigma_{k+1} = 0$.
- For $j = 1, \cdots, k$, define $\mathbf{u}_j = \frac{1}{\sigma_j} A \mathbf{v}_j$.
  - $\mathbf{u}_j$ is unit vector: $\mathbf{u}_j^T \mathbf{u}_j = \frac{1}{s_j^2} \mathbf{v}_j^T \left(A^T A \mathbf{v}_j\right) = \frac{\lambda_j}{\sigma_j^2} = 1$.

Constructing the SVD for $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

- Let $A^T A = V D V^T$ be eigendecomposition, with

  $$D = \textbf{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \quad \text{with} \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

  be eigenvalues and $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ be orthogonal.

  $$\text{So} \quad A^T A \mathbf{v}_j = \lambda_j \mathbf{v}_j \quad \text{for} \quad j = 1, \cdots, n.$$

- Define $\sigma_j = \sqrt{\lambda_j}$ for $j = 1, \cdots, n$. Let k be such that $\sigma_k > 0$ and $\sigma_{k+1} = 0$.
- For $j = 1, \cdots, k$, define $\mathbf{u}_j = \frac{1}{\sigma_j} A \mathbf{v}_j$.
  - $\mathbf{u}_j$ is unit vector: $\mathbf{u}_j^T \mathbf{u}_j = \frac{1}{s_j^2} \mathbf{v}_j^T \left(A^T A \mathbf{v}_j\right) = \frac{\lambda_j}{\sigma_j^2} = 1.$

  $(\mathbf{u}_1, \cdots, \mathbf{u}_k)$ column orthogonal: $\mathbf{u}_i^T \mathbf{u}_j = \frac{1}{\sigma_i \sigma_j} \mathbf{v}_i^T \left(A^T A \mathbf{v}_j\right) = 0, \ i \neq j.$

Constructing the SVD for $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

- Let $A^T A = V D V^T$ be eigendecomposition, with

  $$D = \mathbf{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \quad \text{with} \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

  be eigenvalues and $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ be orthogonal.

  $$\text{So} \quad A^T A \mathbf{v}_j = \lambda_j \mathbf{v}_j \quad \text{for} \quad j = 1, \cdots, n.$$

- Define $\sigma_j = \sqrt{\lambda_j}$ for $j = 1, \cdots, n$. Let k be such that $\sigma_k > 0$ and $\sigma_{k+1} = 0$.
- For $j = 1, \cdots, k$, define $\mathbf{u}_j = \frac{1}{\sigma_j} A \mathbf{v}_j$.
  - $\mathbf{u}_j$ is unit vector: $\mathbf{u}_j^T \mathbf{u}_j = \frac{1}{s_j^2} \mathbf{v}_j^T \left(A^T A \mathbf{v}_j \right) = \frac{\lambda_j}{\sigma_j^2} = 1$.

  $(\mathbf{u}_1, \cdots, \mathbf{u}_k)$ column orthogonal: $\mathbf{u}_i^T \mathbf{u}_j = \frac{1}{\sigma_i \sigma_j} \mathbf{v}_i^T \left(A^T A \mathbf{v}_j \right) = 0, \; i \neq j$.

- Choose $U = (\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{u}_{k+1}, \cdots \mathbf{u}_m) \in \mathbb{R}^{m \times m}$ orthogonal.

Then $A V = U S$ with $S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m \times n}$.

Constructing the SVD for $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

- Let $A^T A = V D V^T$ be eigendecomposition, with

  $$D = \mathbf{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \quad \text{with} \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

  be eigenvalues and $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ be orthogonal.

  $$\text{So} \quad A^T A \mathbf{v}_j = \lambda_j \mathbf{v}_j \quad \text{for} \quad j = 1, \cdots, n.$$

- Define $\sigma_j = \sqrt{\lambda_j}$ for $j = 1, \cdots, n$. Let k be such that $\sigma_k > 0$ and $\sigma_{k+1} = 0$.
- For $j = 1, \cdots, k$, define $\mathbf{u}_j = \frac{1}{\sigma_j} A \mathbf{v}_j$.
  - $\mathbf{u}_j$ is unit vector: $\mathbf{u}_j^T \mathbf{u}_j = \frac{1}{s_j^2} \mathbf{v}_j^T \left(A^T A \mathbf{v}_j \right) = \frac{\lambda_j}{\sigma_j^2} = 1.$

  $(\mathbf{u}_1, \cdots, \mathbf{u}_k)$ column orthogonal: $\mathbf{u}_i^T \mathbf{u}_j = \frac{1}{\sigma_i \sigma_j} \mathbf{v}_i^T \left(A^T A \mathbf{v}_j \right) = 0, \; i \neq j.$

- Choose $U = (\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{u}_{k+1}, \cdots \mathbf{u}_m) \in \mathbb{R}^{m \times m}$ orthogonal.

Then $A V = U S$ with $S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m \times n}$. So $A = U S V^T$.

EX: $A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$. Eigenvalues of $A^T A$: $\lambda_1 = 360, \lambda_2 = 90, \lambda_3 = 0$,

unit eigenvectors $\mathbf{v}_1 = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \ \mathbf{v}_2 = \frac{1}{3}\begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix}, \ \mathbf{v}_3 = \frac{1}{3}\begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}$.

$$A\,\mathbf{v}_1 \quad = \quad \sqrt{360}\,\mathbf{u}_1, \quad A\,\mathbf{v}_2 = \sqrt{90}\,\mathbf{u}_2, \quad , A\,\mathbf{v}_3 = \mathbf{0}.$$

where $\mathbf{u}_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ 1 \end{pmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ -3 \end{pmatrix}$.

Putting together

$$A\,(\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3) \quad = \quad \left(\sqrt{360}\,\mathbf{u}_1,\,\sqrt{90}\,\mathbf{u}_2,\,\mathbf{0}\right) = (\mathbf{u}_1,\,\mathbf{u}_2)\begin{pmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{pmatrix}.$$

Therefore $A = (\mathbf{u}_1,\,\mathbf{u}_2)\begin{pmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{pmatrix}(\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3)^T$.

Thm. 9: Let $A = U\,S\,V^T$ with

$$U = (\mathbf{u}_1, \cdots, \mathbf{u}_m)\,, \quad V = (\mathbf{v}_1, \cdots, \mathbf{v}_n)\,\text{orthogonal},$$

$$S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \text{ with } \sigma_1 \geq \cdots \geq \sigma_k > \sigma_{k+1} = \cdots = \sigma_n = 0.$$

Then **Rank** $(A) = k$.

Thm. 9: Let $A = U\,S\,V^T$ with

$$U = (\mathbf{u}_1, \cdots, \mathbf{u}_m)\,, \quad V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \,\text{orthogonal},$$

$$S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \text{ with } \sigma_1 \geq \cdots \geq \sigma_k > \sigma_{k+1} = \cdots = \sigma_n = 0.$$

Then **Rank** $(A) = k$.

PROOF: By definition, **Rank** $(A) =$ **dim** $(\text{Col} \;\; A)$, with

$$\text{Col} \;\; A = \{A\,\mathbf{x} \mid \mathbf{x} \in \mathcal{R}^n\}\,.$$

- $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are orthonormal basis for $\mathcal{R}^n$. Thus

$$\text{Col} \;\; A = \{A\,(c_1\,\mathbf{v}_1 + \cdots + c_n\,\mathbf{v}_n) \mid c_1, \;\cdots\;, \; c_n \in \mathcal{R}\}\,.$$

- Now $A\,(c_1\,\mathbf{v}_1 + \cdots + c_n\,\mathbf{v}_n) = \sigma_1\,c_1\,\mathbf{u}_1 + \cdots + \sigma_k\,c_k\,\mathbf{u}_k$. Thus,

$$\text{Col} \;\; A = \{\sigma_1\,c_1\,\mathbf{u}_1 + \cdots + \sigma_k\,c_k\,\mathbf{u}_k \mid c_1, \;\cdots\;, \; c_k \in \mathcal{R}\} = \textbf{Span}\,\{\mathbf{u}_1, \;\cdots\;, \; \mathbf{u}_k\}\,.$$

$$\textbf{Rank}\,(A) = \textbf{dim}\,(\textbf{Span}\,\{\mathbf{u}_1, \;\cdots\;, \;\mathbf{u}_k\}) = k. \quad \square$$

The Invertible Matrix Theorem

Let $A = U\,S\,V^T$ be the SVD of $A \in \mathcal{R}^{n\times n}$ with

$$S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix}, \; \sigma_1 \geq \cdots \geq \sigma_n \geq 0. \text{ Then}$$

$$\textbf{Rank}\,(A) = n \iff \sigma_n > 0.$$

Solving Least Squares Problem with SVD (I)

Let the SVD of $A = U\,S\,V^T \in \mathbb{R}^{m\times n}$, with $m \geq n$, where

$U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \in \mathbb{R}^{m\times m}$ and $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n\times n}$

be orthogonal; and $S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m\times n}$ be diagonal

with $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_n \geq 0$.

Solving Least Squares Problem with SVD (I)

Let the SVD of $A = U\,S\,V^T \in \mathbb{R}^{m\times n}$, with $m \geq n$, where

$U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \in \mathbb{R}^{m\times m}$ and $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n\times n}$

be orthogonal; and $S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m\times n}$ be diagonal

with $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_n \geq 0$.

The Least Squares Problem (LS) is

$$\mathbf{min}_{\mathbf{x}\in\mathbb{R}^n} \|A\,\mathbf{x} - \mathbf{b}\|_2 \quad \text{for a given } \mathbf{b} \in \mathbb{R}^m$$

and the LS solution satisfies $A^T A\,\mathbf{x} = A^T\,\mathbf{b}$.

$$A = U\,S\,V^T = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)^T .$$

$$A = U\,S\,V^T = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)^T.$$

$$A^T A\mathbf{x} = A^T\mathbf{b} \Leftrightarrow VS^TS\left(V^T\mathbf{x}\right) = VS^T\left(U^T\mathbf{b}\right) \Leftrightarrow S^TS\left(V^T\mathbf{x}\right) = S^T\left(U^T\mathbf{b}\right),$$

which is $\sigma_j^2 \left(V^T\,\mathbf{x}\right)_j = \sigma_j \left(U^T\,\mathbf{b}\right)_j$, for $j = 1, \cdots, n.$ (ℓ)

Define $\sigma_j^\dagger = \begin{cases} \sigma_j^{-1}, & \text{if } \sigma_j > 0, \\ 0, & \text{otherwise.} \end{cases}$

Equation (ℓ) solves to $\left(V^T\,\mathbf{x}\right)_j = \sigma_j^\dagger \left(U^T\,\mathbf{b}\right)_j$, for $j = 1, \cdots, n.$

Therefore $V^T\,\mathbf{x} = \begin{pmatrix} \sigma_1^\dagger & & & \mathbf{0} \\ & \ddots & & \\ & & \sigma_n^\dagger & \mathbf{0} \end{pmatrix} \left(U^T\,\mathbf{b}\right) \stackrel{def}{=} S^\dagger \left(U^T\,\mathbf{b}\right),$

and

$$A = U\,S\,V^T = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)^T.$$

$$A^T A\mathbf{x} = A^T\mathbf{b} \Leftrightarrow VS^TS\left(V^T\mathbf{x}\right) = VS^T\left(U^T\mathbf{b}\right) \Leftrightarrow S^TS\left(V^T\mathbf{x}\right) = S^T\left(U^T\mathbf{b}\right),$$

which is $\sigma_j^2 \left(V^T\mathbf{x}\right)_j = \sigma_j \left(U^T\mathbf{b}\right)_j$, for $j = 1, \cdots, n$. (ℓ)

Define $\sigma_j^\dagger = \begin{cases} \sigma_j^{-1}, & \text{if } \sigma_j > 0, \\ 0, & \text{otherwise.} \end{cases}$

Equation (ℓ) solves to $\left(V^T\mathbf{x}\right)_j = \sigma_j^\dagger \left(U^T\mathbf{b}\right)_j$, for $j = 1, \cdots, n$.

Therefore $V^T\mathbf{x} = \begin{pmatrix} \sigma_1^\dagger & & & \mathbf{0} \\ & \ddots & & \\ & & \sigma_n^\dagger & \mathbf{0} \end{pmatrix} \left(U^T\mathbf{b}\right) \stackrel{def}{=} S^\dagger \left(U^T\mathbf{b}\right),$

and $\mathbf{x} = \left(V\,S^\dagger\,U^T\right)\mathbf{b}$.

$$A \quad = \quad U\,S\,V^T = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)^T .$$

Least Squares Solution

$$\mathbf{x} = \left(V\,S^\dagger\,U^T \right)\,\mathbf{b},$$

$$\text{with} \quad S^\dagger = \begin{pmatrix} \sigma_1^\dagger & & & \mathbf{0} \\ & \ddots & & \vdots \\ & & \sigma_n^\dagger & \mathbf{0} \end{pmatrix}.$$

DEFINITION: $A^\dagger = V\,S^\dagger\,U^T =$ is PSEUDO-INVERSE of A.

§7.5 Applications

Table: EXAMPLE Class Grades

Student	Midterm #1	Midterm #2	Final	Homework	Quizzes
Alice	50	83	97	64	77
Ben	47	87	60	0	0
Cindy	91	95	95	90	99
Eric	85	100	88	87	91
Fiona	89	99	86	76	65
Gloria	70	76	67	78	77
Henry	100	80	90	91	83

$$\textbf{matrix of observations} \quad \mathbf{X} = \begin{pmatrix} 50 & 47 & 91 & 85 & 89 & 70 & 100 \\ 83 & 87 & 95 & 100 & 99 & 76 & 80 \\ 97 & 60 & 95 & 88 & 86 & 67 & 90 \\ 64 & 0 & 90 & 87 & 76 & 78 & 91 \\ 77 & 0 & 99 & 91 & 65 & 77 & 83 \end{pmatrix}.$$

$$\begin{matrix} \Uparrow & \Uparrow & & & & & \Uparrow \\ \mathbf{x}_1 & \mathbf{x}_2 & & \cdots & & & \mathbf{x}_7 \end{matrix}$$

- 5 **variables**: Midterm #1, Midterm #2, Final, Homework, Quizzes,
- 7 **samples**: 7 students.

Example Class Grades

matrix of observations $\mathbf{X} = \begin{pmatrix} 50 & 47 & 91 & 85 & 89 & 70 & 100 \\ 83 & 87 & 95 & 100 & 99 & 76 & 80 \\ 97 & 60 & 95 & 88 & 86 & 67 & 90 \\ 64 & 0 & 90 & 87 & 76 & 78 & 91 \\ 77 & 0 & 99 & 91 & 65 & 77 & 83 \end{pmatrix}.$

$$\Uparrow \quad \Uparrow \qquad\qquad \Uparrow$$
$$\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_7$$

sample mean $\mathbf{m} = \frac{1}{7}(\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_7) = \frac{1}{7}\begin{pmatrix} 532 \\ 620 \\ 583 \\ 486 \\ 492 \end{pmatrix}$

mean-deviation form $\mathbf{B} \stackrel{def}{=} (\mathbf{x}_1 - \mathbf{m}, \cdots, \mathbf{x}_7 - \mathbf{m})$

$$= \frac{1}{7}\begin{pmatrix} -182 & -203 & 105 & 63 & 91 & -42 & 168 \\ -39 & -11 & 45 & 80 & 73 & -88 & -60 \\ 96 & -163 & 82 & 33 & 19 & -114 & 47 \\ -38 & -486 & 144 & 123 & 46 & 60 & 151 \\ 47 & -492 & 201 & 145 & -37 & 47 & 89 \end{pmatrix}.$$

EXAMPLE Class Grades

matrix of observations $\mathbf{X} = \begin{pmatrix} 50 & 47 & 91 & 85 & 89 & 70 & 100 \\ 83 & 87 & 95 & 100 & 99 & 76 & 80 \\ 97 & 60 & 95 & 88 & 86 & 67 & 90 \\ 64 & 0 & 90 & 87 & 76 & 78 & 91 \\ 77 & 0 & 99 & 91 & 65 & 77 & 83 \end{pmatrix}.$

$\Uparrow \quad \Uparrow \qquad\qquad \Uparrow$

$\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_7$

sample mean $\mathbf{m} = \frac{1}{7}(\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_7) = \frac{1}{7}\begin{pmatrix} 532 \\ 620 \\ 583 \\ 486 \\ 492 \end{pmatrix}$

mean-deviation form $\mathbf{B} \stackrel{def}{=} (\mathbf{x}_1 - \mathbf{m}, \cdots, \mathbf{x}_7 - \mathbf{m}),$

sample covariance matrix $\mathbf{S} \stackrel{def}{=} \frac{1}{6}\mathbf{B}\,\mathbf{B}^T.$

$$\boxed{\textbf{sample mean}} \quad \mathbf{m} \quad = \quad \frac{1}{N}\left(\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_N\right).$$

$$\boxed{\textbf{mean-deviation form}} \; \mathbf{B} \quad \stackrel{def}{=} \quad \left(\mathbf{x}_1 - \mathbf{m}, \cdots, \mathbf{x}_N - \mathbf{m}\right),$$

$$\boxed{\textbf{sample covariance matrix}} \; \mathbf{S} \quad \stackrel{def}{=} \quad \frac{1}{N-1}\mathbf{B}\,\mathbf{B}^T.$$

$$\mathbf{S}_{i,j} = \begin{cases} \text{covariance of } x_i \text{ and } x_j, & \text{if } i \neq j, \\ \text{variance of } x_j, & \text{if } i = j. \end{cases}$$

variables x_i and x_j are **uncorrelated** if $\mathbf{S}_{i,j} = 0$.

Principal Component Analysis

In the setting of p variables $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$, *determine a change of variables*

$$\mathbf{x} = P\,\mathbf{y} = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}$$

so that the new variables $y_1, y_2, \cdots, y_p$ *are uncorrelated and in order of decreasing variance.*

EXAMPLE Class Grades

$$\boxed{\textbf{sample mean}}\quad \mathbf{m} = \frac{1}{7}(\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_7)\,.$$

$$\boxed{\textbf{mean-deviation form}}\ \mathbf{B} \overset{def}{=} (\mathbf{x}_1 - \mathbf{m}, \cdots, \mathbf{x}_7 - \mathbf{m}) = U\Sigma V^T,$$

$$\boxed{\textbf{change of variables}}\quad \mathbf{x} \overset{def}{=} U\,\mathbf{y},$$

$$\boxed{\textbf{new covariance matrix}}\quad \mathbf{S}^{new} = \frac{1}{6}U^T\,\mathbf{B}\,\mathbf{B}^T\,U = \Sigma\,\Sigma^T$$

$$= \begin{pmatrix} 2432 & 0 & 0 & 0 & 0 \\ 0 & 225 & 0 & 0 & 0 \\ 0 & 0 & 120 & 0 & 0 \\ 0 & 0 & 0 & 44 & 0 \\ 0 & 0 & 0 & 0 & 9 \end{pmatrix} \approx \begin{pmatrix} 2432 & 0 & 0 & 0 & 0 \\ 0 & 225 & 0 & 0 & 0 \\ 0 & 0 & 120 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\boxed{\textbf{Approximate change of variables}}\quad \mathbf{x} \overset{def}{\approx} U \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ 0 \\ 0 \end{pmatrix}.$$

y_1, y_2, y_3 leading principal components.

Eckart-Young Theorem: For $A \in \mathbb{R}^{m\times n}$ with $m \geq n$

Let the SVD of $A = U\,S\,V^T$, where $U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \in \mathbb{R}^{m\times m}$,

$V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n\times n}$ orthogonal; and

$$S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m\times n} \text{ diagonal with } \sigma_1 \geq \sigma_2 \cdots \geq \sigma_n \geq 0.$$

Eckart-Young Theorem: For $A \in \mathbb{R}^{m \times n}$ with $m \geq n$

Let the SVD of $A = U\,S\,V^T$, where $U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \in \mathbb{R}^{m \times m}$,

$V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ orthogonal; and

$$S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m \times n} \text{ diagonal with } \sigma_1 \geq \sigma_2 \cdots \geq \sigma_n \geq 0.$$

For any $1 \leq k \leq n$, the rank-k TRUNCATED SVD of A is

$$A_k \stackrel{def}{=} (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)^T.$$

Eckart-Young Theorem: For $A \in \mathbb{R}^{m\times n}$ with $m \ge n$

Let the SVD of $A = U\,S\,V^T$, where $U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \in \mathbb{R}^{m\times m}$, $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n\times n}$ orthogonal; and

$$S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m\times n} \text{ diagonal with } \sigma_1 \ge \sigma_2 \cdots \ge \sigma_n \ge 0.$$

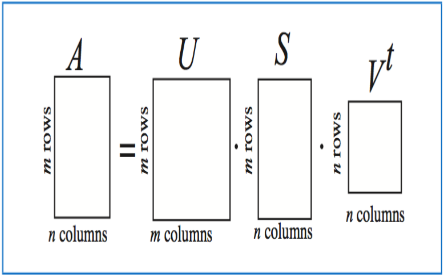
For any $1 \le k \le n$, the rank-k TRUNCATED SVD of A is

$$A_k \stackrel{def}{=} (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)^T. \text{ Then}$$

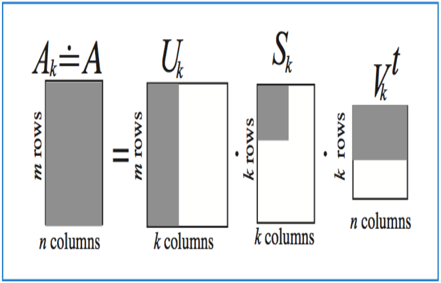
$$\mathbf{min}_{B\in\mathbb{R}^{m\times n},\, \mathbf{Rank}(B)\le k} \|A - B\|_F = \|A - A_k\|_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_n^2},$$

where $\|X\|_F \stackrel{def}{=} \sqrt{x_{11}^2 + \cdots + x_{1n}^2 + \cdots + x_{m1}^2 + \cdots + x_{mn}^2}$.

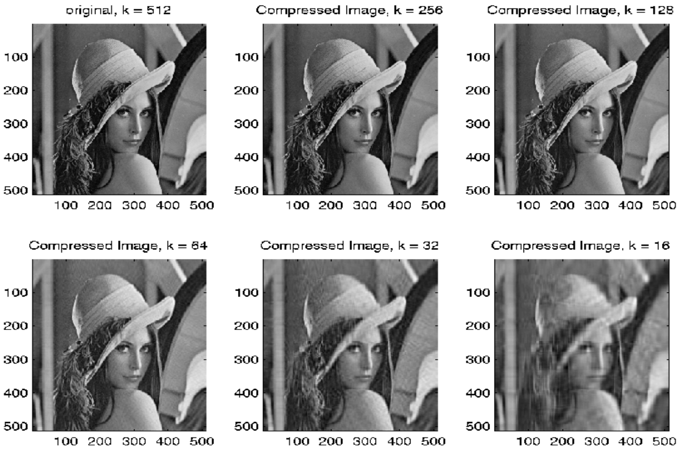
- SVD



- TRUNCATED SVD



Compressing Lena with TRUNCATED SVD



Hilbert Matrix $H = \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n-1} & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{n-1} & \frac{1}{n} & \cdots & \frac{1}{2\,n-3} & \frac{1}{2\,n-2} \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2\,n-2} & \frac{1}{2\,n-1} \end{pmatrix} = H^T$

- Eigendecomposition $H = U\,S\,U^T$ is SVD of H, where $U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n) \in \mathbb{R}^{n \times n}$ is orthogonal;
$$S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \in \mathbb{R}^{n \times n} \text{ diagonal, } \sigma_1 \geq \sigma_2 \cdots \geq \sigma_n > 0.$$

Hilbert Matrix $H = \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n-1} & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{n-1} & \frac{1}{n} & \cdots & \frac{1}{2n-3} & \frac{1}{2n-2} \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2n-2} & \frac{1}{2n-1} \end{pmatrix} = H^T$

- Eigendecomposition $H = U\,S\,U^T$ is SVD of H, where $U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n) \in \mathbb{R}^{n\times n}$ is orthogonal;
  $$S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \in \mathbb{R}^{n\times n} \text{ diagonal, } \sigma_1 \geq \sigma_2 \cdots \geq \sigma_n > 0.$$