Lecture 35: Quick review from previous lecture

- Suppose that a symmetric matrix A has real eigenvalues

$$\lambda_1 \geq \cdots \geq \lambda_n.$$

Then

$$\lambda_1 = \max\{\overbrace{\langle A\mathbf{x}, \mathbf{x}\rangle}^{x^TAx} : \|\mathbf{x}\|_2 = 1\}, \qquad \lambda_n = \min\{\overbrace{\langle A\mathbf{x}, \mathbf{x}\rangle}^{x^TAx} : \|\mathbf{x}\|_2 = 1\}.$$

The **maximal value** is achieved when $\mathbf{x} = \pm\mathbf{u}_1$, the unit eigenvector associated with the largest eigenvalue λ_1.

The **minimal value** is achieved when $\mathbf{x} = \pm\mathbf{u}_n$, the unit eigenvector associated with the smallest eigenvalue λ_n.

- A is complete $\Rightarrow$ $A = VDV^{-1}$

$$= \underbrace{[v_1 \cdots v_n]}_{\text{eigenvectors}} \underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}}_{\lambda_i:\ \text{eigenvalues}} [v_1 \cdots v_n]^{-1}$$

Today we will

- discuss Section 8.7 Singular Values

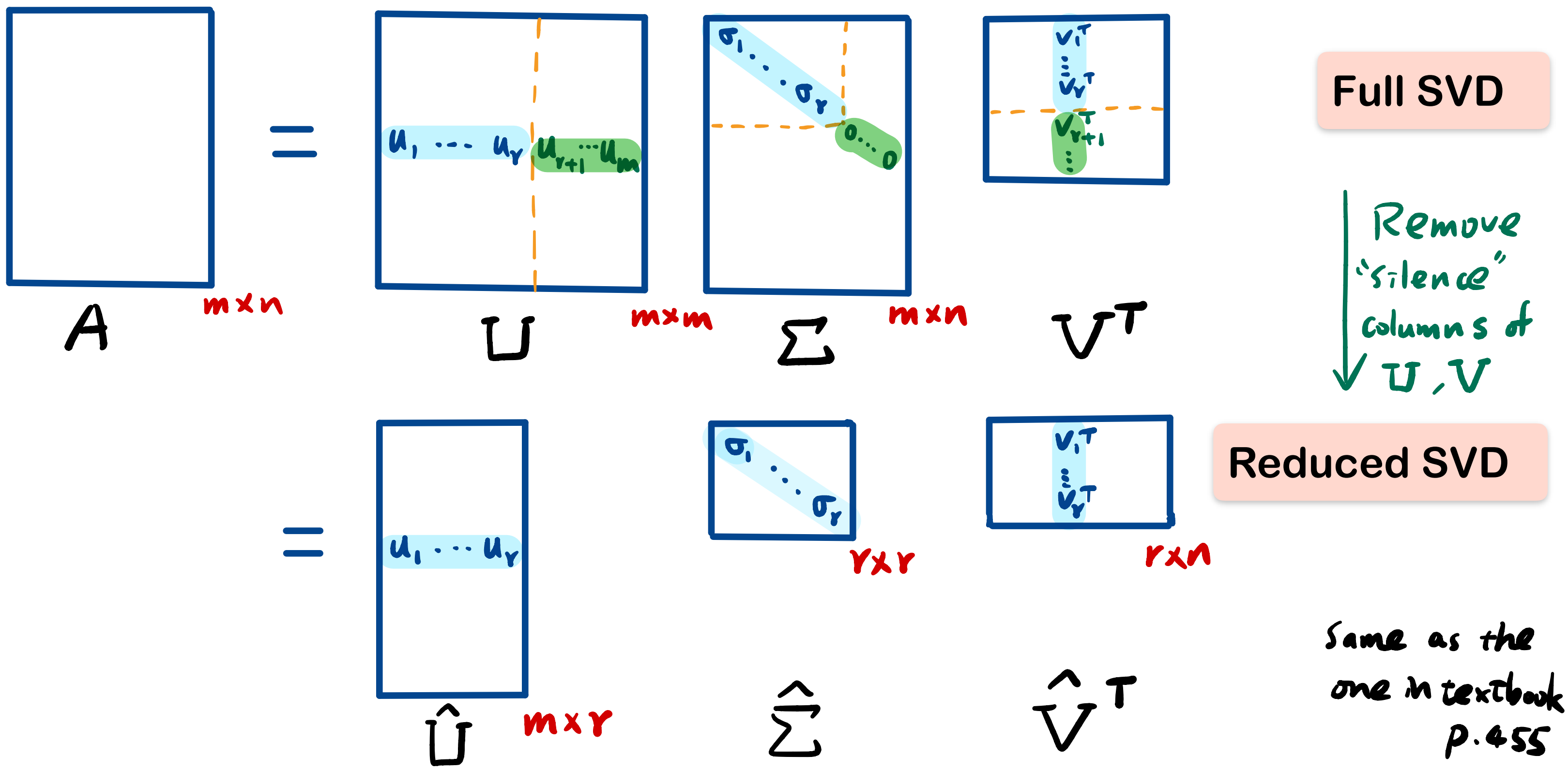
- Lecture will be recorded -

- $A = A^T \Rightarrow A = QD\overbrace{Q^{-1}}^{=Q^T}$, where Q is orthogonal

$Q = [u_1 \cdots u_n]$, u_i is orthonormal.

A is mxn matrix with rank(A)=r

$m \geq n$: rank$(A) = r$.



$\lambda_j \neq 0$: $\quad v_j$ in $\ker(A^TA - \lambda_j I)$, $1 \leq j \leq r$ (right singular vectors)

$\lambda_j = 0$: $\quad v_{r+1}, \ldots, v_n$ in $\ker(A^TA)$

$\sigma_j = \sqrt{\lambda_j}$ (singular values)

$u_j = \dfrac{Av_j}{\sigma_j}$, $1 \leq j \leq r$ (left singular vectors)

$u_{r+1}, \ldots, u_m \in \operatorname{coker} A$

NOTE : $\{v_1, \ldots, v_n\}$ is orthonormal basis for $\mathbb{R}^n$;
$\{u_1, \ldots, u_m\}$ " $\mathbb{R}^m$.

Definition: The square roots of the eigenvalues λ_j of A^TA are called the singular values $\sigma_1, \sigma_2, \cdots, \sigma_n$ of an $m \times n$ matrix A.

Thus, we have shown that

Full SVD for a matrix:
Let A be an $m \times n$ matrix of rank r with the positive singular values

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \qquad (\sigma_{r+1} = \ldots = \sigma_n = 0),$$

and let Σ be the $m \times n$ matrix defined by

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sigma_r & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \cdots & 0 & \cdots & 0 \end{bmatrix}_{m \times n}$$

Then there exist an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

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

Example 1. Find both full and reduced SVD for $A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix}_{2\times 3}$. $: \mathbb{R}^3 \to \mathbb{R}^2$ $\{v_1, v_2, v_3\}$ $\{u_1, u_2\}$

Solution: $A^TA = \begin{bmatrix} 2 & 2 & -2 \\ 2 & 2 & -2 \\ -2 & -2 & 2 \end{bmatrix}$ has eigenvalues $6, 0, 0$

singular values $\sqrt{6}, 0, 0$.
$\sigma_1 \quad \sigma_2 \quad \sigma_3$

① $\lambda = 6$, Find v_1 in $\ker(A^TA - 6I)$.

$A^TA - 6I = \begin{pmatrix} -4 & 2 & -2 \\ 2 & -4 & -2 \\ -2 & -2 & -4 \end{pmatrix}$. Then $v_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$

normalized.

② $\lambda = 0$. Find orthonormal v_2, v_3 in $\ker(A^TA)$. $A^TA \to \begin{pmatrix} 2 & 2 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $\begin{pmatrix} -y+z \\ y \\ z \end{pmatrix} \to \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

$v_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$, $v_3 = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$. are orthonormal. G-S make them orthogonal.

③ Find orthonormal basis of $\operatorname{img} A$: $u_1 = \frac{Av_1}{\sigma_1} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

④ Find orthonormal basis of $\operatorname{coker} A$: $u_2 \perp u_1$; $u_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

[Example continue]

Full SVD: $A = U\Sigma V^T = [u_1 \ u_2]\begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{2\times 3}\begin{bmatrix} v_1^T \\ v_2^T \\ v_3^T \end{bmatrix}$

$= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{2\times 3}\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \end{bmatrix}$

Reduced SVD: $A = \hat{U}\hat{\Sigma}\hat{V}^T = [u_1]\,[\sqrt{6}]_{1\times 1}\,[v_1^T]$

$= \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}[\sqrt{6}]\,[1/\sqrt{3} \ \ 1/\sqrt{3} \ \ -1/\sqrt{3}]$ #

We show Fact 2 now.

Fact 2:

(1) $\lambda_1, \ldots, \lambda_r > 0$.

(2) $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is orthonormal basis for coimg A and $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ is orthonormal basis for ker A.

(3) $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ are mutually orthogonal and, moreover,

$$\left\{\frac{A\mathbf{v}_1}{\sqrt{\lambda_1}}, \ldots, \frac{A\mathbf{v}_r}{\sqrt{\lambda_r}}\right\} \quad \text{is orthonormal basis for img } A$$

(1) We knew $\lambda_j \neq 0$, $1 \le j \le r$. To show $\lambda_j > 0$.
$A^TAv_j = \lambda_j v_j \Rightarrow v_j^T A^TA v_j = \lambda_j v_j^T v_j = \lambda_j$ (since $v_j^Tv_j = \|v_j\|^2 = 1$)
Since $v_j^T A^T A v_j = \|Av_j\|^2 > 0$, one has $\lambda_j > 0$.

(2) Since $\ker(A^TA) = \ker(A)$ by Fact 1, and $v_{r+1}, \ldots, v_n$ in $\ker(A^TA)$, we thus have $\{v_{r+1}, \ldots, v_n\}$ is $\ker A$.
Since $\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_n\}$ is orthonormal basis of $\mathbb{R}^n$ and coimg $A = (\ker A)^\perp$ it implies $\{v_1, \ldots, v_r\}$ is orthonormal basis for coimg A.

(3) $\langle Av_i, Av_j\rangle = (Av_i)^TAv_j = v_i^T A^TAv_j = v_i^T \lambda_j v_j = \begin{cases} \lambda_j, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$
when $i=j$, $\|Av_i\|^2 = \langle Av_i, Av_i\rangle = \lambda_i \Rightarrow \|Av_i\| = \sqrt{\lambda_i}$ #

Fact 3: Let $A \in M_{m\times n}$ ($m \times n$ real matrices). Then the following are true.

1. The nonzero eigenvalues of A^TA and AA^T are the same.
2. A and A^T have the same nonzero singular values.

(1) We have SVD of A: $A = U\Sigma V^T$, $\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}_{m\times n}$, $\operatorname{rank}(A) = r$.

$$A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = (V^T)^T\Sigma^T \underbrace{U^TU}_{I}\Sigma V^T$$

$$= V\Sigma^T\Sigma V^T$$

$$= V \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}_{n\times m} \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}_{m\times n} V^T$$

$$= V \begin{bmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}_{n\times n} V^T \quad (V^T = V^{-1})$$

A^TA has eigenvalues $\sigma_1^2 \ldots \sigma_r^2, 0, \ldots 0$,

$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T = U \begin{bmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}_{m\times m} U^T, \quad U^T = U^{-1}.$$

AA^T has eigenvalues $\sigma_1^2, \ldots \sigma_r^2, 0, \ldots, 0$,

(2) singular value of A : square root of eigenvalues of A^TA.

" of A^T : " of AA^T.

Fact 4: If $A = A^T$ is a symmetric matrix, then its singular values, σ_i, are the absolute values of the eigenvalues of A. (that is, $\sigma_i = |\lambda_i|$)

Suppose $Av = \lambda v$.

$$A^TAv = A^T(\lambda v) \overset{A=A^T}{=} \lambda A v = \lambda^2 v$$

So singular value of A is $\sqrt{\lambda^2} = |\lambda|$. #

§ Data fitting: Least squares problems

The problem here is: How do we "almost" solve a system $A\mathbf{x} = \mathbf{b}$?

For example, an experimenter collects data by taking measurements $\{(t_i, b_i)\}$:

$$b_1, b_2, \ldots, b_m \qquad \text{at times } \ t_1, t_2, \ldots, t_m, \ \text{respectively.}$$

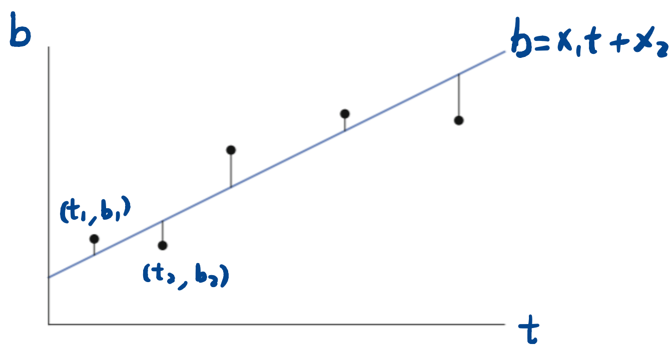


Figure 1: Least squares approximation of data by a straight line.

Suppose we use a linear model $\boxed{b = x_1 t + x_2}$ (x_1, x_2 to be determined) to make a prediction so that the line $b = x_1 t + x_2$ **best fits the data collected**.

One way is to minimize the error

$$E \overset{def}{=} \sum_{i=1}^{m} (\overbrace{b_i}^{\text{real data}} - \underbrace{(x_1 t_i + x_2)}_{\text{predicted data}})^2 = \|A\mathbf{x} - \mathbf{b}\|_2^2.$$

Here

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \quad A = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

→ to be solved

Unfortunately, the system $A\mathbf{x} = \mathbf{b}$ can NOT be solved in many cases.

In this scenario, we can try to find a $\mathbf{x}^*$ that

$$\text{minimizes the error } \|A\mathbf{x} - \mathbf{b}\|.$$

That means

$$\|A\mathbf{x}^* - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\| \qquad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

Poll Question 1: Suppose that the matrix A can be factored in the form

$$A = VDV^{-1}$$

where D is diagonal and V is nonsingular. Then every column vector of V are eigenvectors of A.

A) Yes

B) No