Lecture 35: Quick review from previous lecture

- Suppose that a symmetric matrix A has real eigenvalues

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

  Then

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

  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.

Today we will discuss Singular Value Decomposition.

- Lecture will be recorded -

- Information for Final Exam and Course Grade has been posted on Canvas, see "Announcements".

Definition: The square roots of the eigenvalues 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

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r,$$

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

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

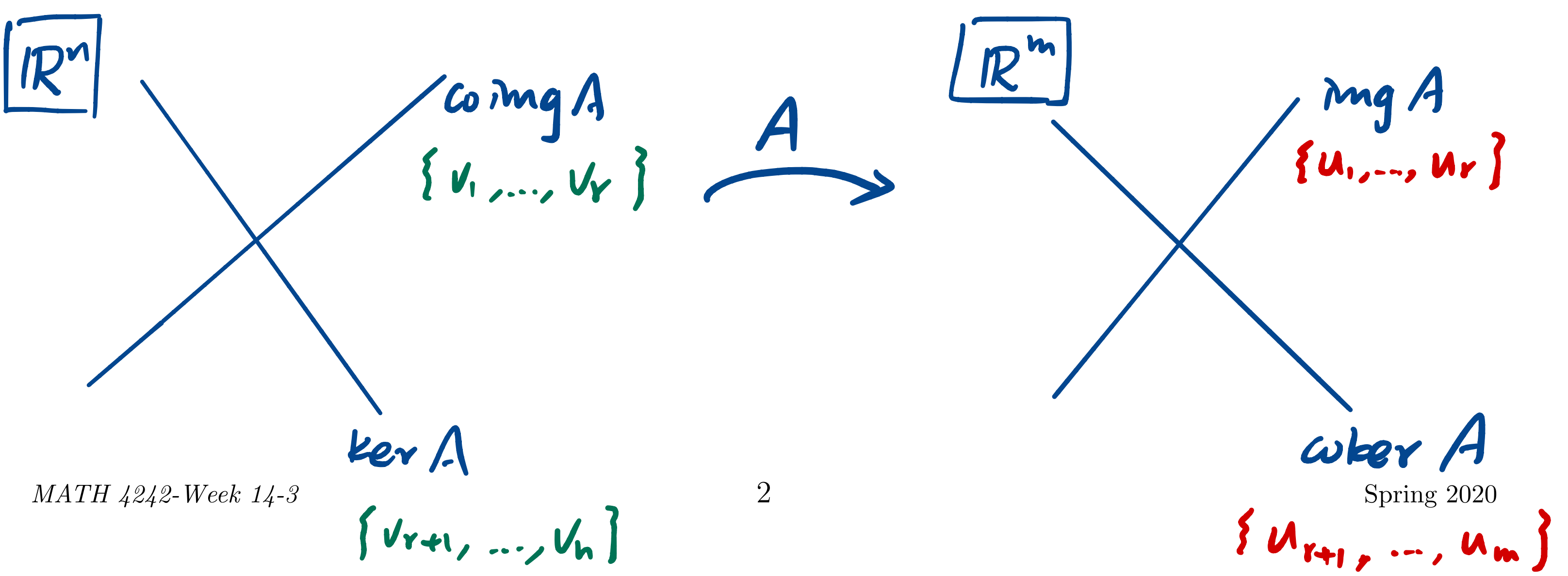
Then there exist an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

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

$(A^TA)v_j = \lambda_j v_j$, $1 \le j \le r$; $u_j = \dfrac{Av_j}{\sqrt{\lambda_j}} = \dfrac{Av_j}{\sigma_j}$, $1 \le j \le r$

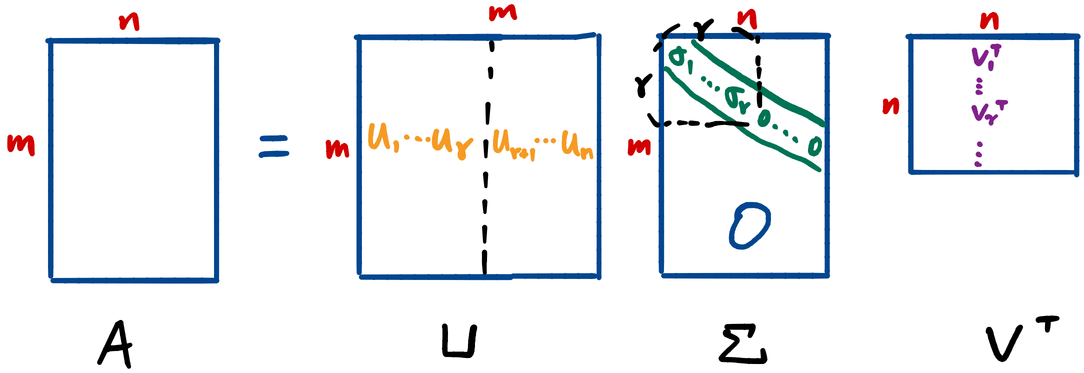
$(A^TA)v_j = 0$, $r+1 \le j \le n$;

$u_{r+1}, \dots, u_m$ ONB for coker A.

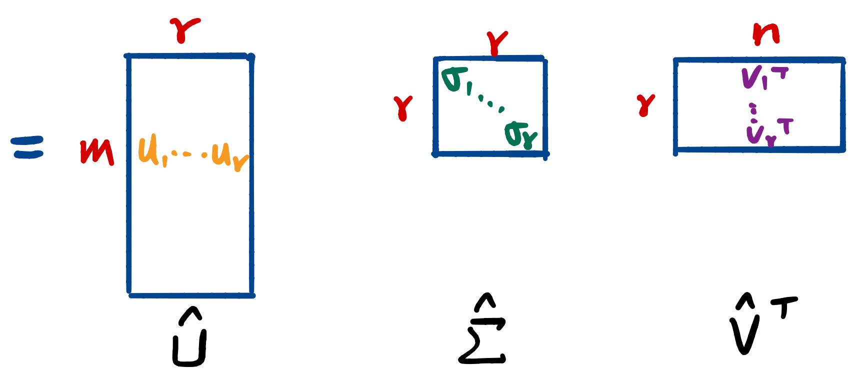


[SVD] $A\underbrace{[v_1\ v_2 \cdots v_r\ v_{r+1} \cdots v_n]}_{V} = \underbrace{[u_1\ u_2 \cdots u_r\ u_{r+1} \cdots u_m]}_{U}\underbrace{\begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ & & & & 0 & \\ & & & & & \ddots \ 0 \\ & & & 0 & & \end{bmatrix}}_{\Sigma}$

If $m \geq n$.



(Full SVD)



(Reduced SVD)

same as the one in p.455 textbook

Example. Find full and reduced SVD for $A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix}$.

Solution: $A^TA = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 2 & -2 \\ 2 & 2 & -2 \\ -2 & -2 & 2 \end{pmatrix}$ has eigenvalues $\lambda = 6, 0, 0$.

$\underline{\lambda = 6}$. $v_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$; $\underline{\lambda = 0}$. $v_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$, $v_3 = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$.

$$u_1 = \frac{Av_1}{\sqrt{6}} = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix}\frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Finding $u_2 \perp u_1$, ($u_2 \in \operatorname{coker} A$), we have

$$u_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

[Example continue]

(Full SVD)

Then $\underset{3\times 2}{A} = U \Sigma V^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \underset{3\times 2}{\begin{bmatrix} \boxed{\sqrt{6}} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}} \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ -1/\sqrt{3} & 0 & 2/\sqrt{6} \end{bmatrix}^T.$

Reduced SVD:

$$A = \underset{\hat{U}}{\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}} \underset{\hat{\Sigma}}{\begin{bmatrix} \sqrt{6} \end{bmatrix}} \underset{\hat{V}^T}{\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \end{bmatrix}} . \#$$

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

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

---

1. SVD of A (rank $A = r$) : $A = U \Sigma V^T$.

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

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

$\Sigma^T \Sigma$ and $\Sigma \Sigma^T$ have the same nonzero eigenvalues $\sigma_1^2, \ldots, \sigma_r^2$.

Since $A^T A$ is similar to $\Sigma^T \Sigma$, $A^T A$ has nonzero eigenvalues $\sigma_1^2, \ldots, \sigma_r^2$. Similarly, AA^T also have nonzero eigenvalues $\sigma_1^2, \ldots, \sigma_r^2$ since AA^T is similar to $\Sigma\Sigma^T$. #

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

1. If $A = A^T$, then the singular values of A are the absolute values of the eigenvalues of A.

$Av = \lambda v, \; v \neq 0.$

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

So, λ^2 is eigenvalue of $A^T A$. it leads to

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

Likewise, we also have

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

1. $\|A\|_2 = \sigma_1$
2. $\|A\|_F = \sqrt{\sigma_1^2 + \ldots + \sigma_r^2}$.

The proof is similar to the one for symmetric matrix we showed in Lecture 33. Thus, we skip the proof here.

§ Least square solutions - Pseudo inverse

How do we “almost” solve a system?

For instance, we consider an experimenter collects data by taking measurements

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

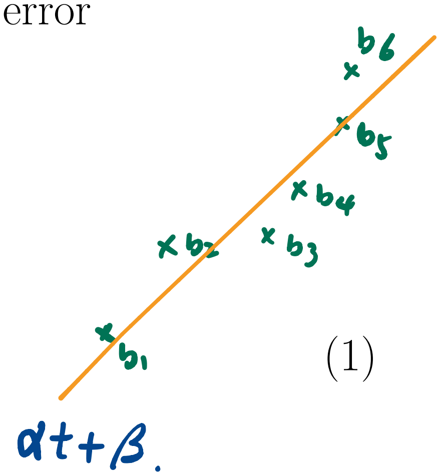
Suppose that the data

$$(b_1, t_1), (b_2, t_2), \ldots, (b_m, t_m)$$

are plotted in the plane.

Suppose there exists a “linear relationship” between b and t, say $b = \alpha t + \beta$. We want to find the constants α, β so that the line $b = \alpha t + \beta$ represents the best possible fit to the data collected. One way is to minimize the error

$$E \stackrel{def}{=} \sum_{i=1}^{m} \Big(\underbrace{b_i}_{collected\ data} - \underbrace{(\alpha t_i + \beta)}_{linear\ relationship} \Big)^2,$$



which can be written as

$$\|Ax - b\|^2, \tag{1}$$

where

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

Since we cannot solve $Ax = b$ exactly in many cases, our goal here is to find a x that minimizes $\|Ax - b\|$. We will develop a general method for finding a vector x^* that **minimizes the error** E, that is,

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

Definition: Suppose that $A \in M_{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$. **The least squares problem** is to find $\mathbf{x} \in \mathbb{R}^n$ for which that $\|A\mathbf{x} - \mathbf{b}\|$ is minimized.

A vector $\mathbf{x}$ that minimizes $\|A\mathbf{x} - \mathbf{b}\|$ is called **the least squares solution**.

Definition: The **pseudoinverse** of a nonzero $m \times n$ matrix A with SVD $A = U\Sigma V^T$ is the $n \times m$ matrix

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

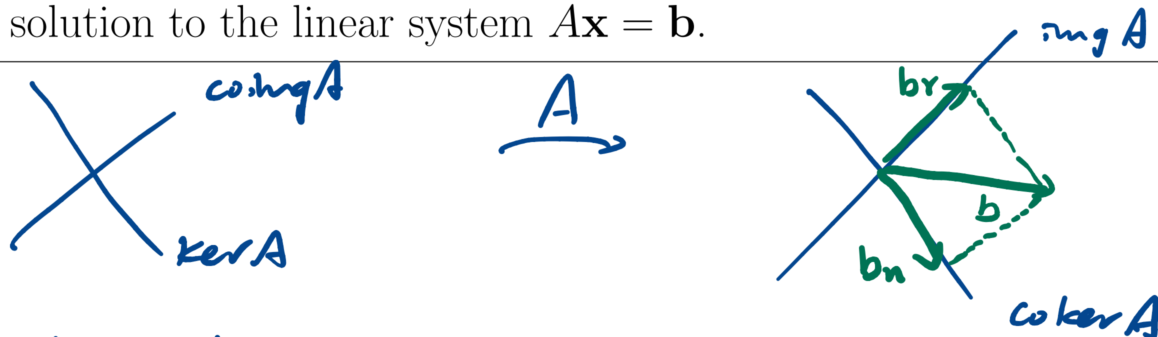
$(m \geq n)$

A has rank r with singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$.

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

$$\Sigma^+ = \begin{bmatrix} 1/\sigma_1 & & & & & \\ & \ddots & & & 0 & \\ & & 1/\sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}_{n\times m} ; \quad \Sigma^+_{ij} = \begin{cases} \frac{1}{\sigma_i}, & \text{if } i=j \leq r, \\ 0, & \text{otherwise.} \end{cases}$$

Fact: Suppose that $A \in M_{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$. Let $\mathbf{x}^* = A^+\mathbf{b}$. Then $\mathbf{x}^*$ is the least squares solution to the linear system $A\mathbf{x} = \mathbf{b}$.



$$b = \underset{\text{img}A}{\underset{\cap}{b_r}} + \underset{\text{coker}A}{\underset{\cap}{b_n}}$$

- b_r : orthogonal projection of b onto img A,
- b_n : " " coker A.

$$\|Ax - b\|_2^2 = \|\overbrace{Ax - b_r}^{\text{img}A} - \overbrace{b_n}^{\text{coker}A}\|_2^2 \underset{(Ax-b_r)\perp b_n.}{\underset{\uparrow}{=}} \|Ax - b_r\|^2 + \|b_n\|_2^2$$

$$\min_x \|Ax - b\|_2^2 = \|b_n\|_2^2.$$

The least squares solution satisfies

$$\boxed{Ax = b_r}$$

<u>check</u> : $x^* = A^+b$ satisfies <u>$Ax = b_r$</u>.

[Continue] $A^T A x = A^T b_r = A^T(b - b_n)$

$= A^T b.$ ($A^T b_n = 0$ since $b_n \in \operatorname{coker} A$)

Remark: The equation $A^T A\mathbf{x} - A^T\mathbf{b} = 0$ is called the **normal equation**.

Fact: Suppose that $A \in M_{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$. Let $\mathbf{x}^* = A^+\mathbf{b}$. If A has n linearly independent columns($\underbrace{\ker A = \{\mathbf{0}\}}_{\operatorname{rank} A = n}$), then

$$\mathbf{x}^* = A^+\mathbf{b} = (A^T A)^{-1}A^T\mathbf{b},$$

where

$$A^+ = (A^T A)^{-1}A^T.$$

Thus, $(A^TA)^{-1}$ exists. x^* satisfies $A^TAx = A^Tb$.

so, $x^* = (A^TA)^{-1}A^Tb$.

Example. Consider the linear system

$$\begin{cases} x + y - z = 1, \\ x + y - z = 1. \end{cases}$$

Find the best approximation to a solution having minimum norm.

To be continued!