Lecture 36: Quick review from previous lecture

- **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

$$A = U\Sigma V^T. \quad \text{(Full SVD)}$$

$Ax=b$

- 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**.

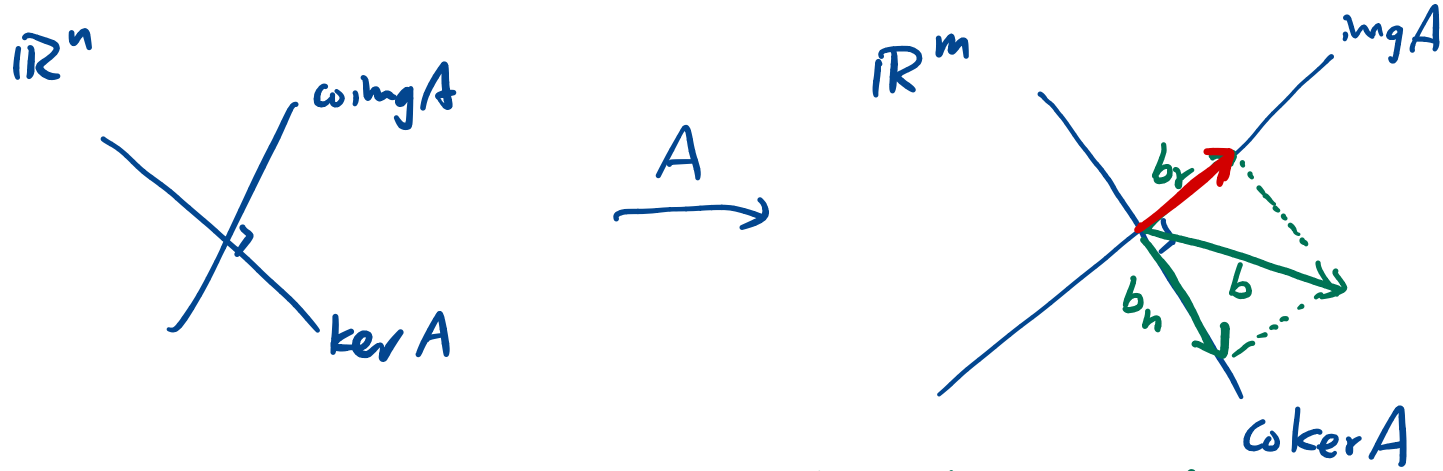
Today we will discuss Singular Value Decomposition.

- Lecture will be recorded -

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

Let's recap. Even if $A\mathbf{x} = \mathbf{b}$ (Data) has no exact solution, we can always find a least squares solution $\mathbf{x}^* = A^+\mathbf{b}$ that minimizes the error $\|A\mathbf{x} - \mathbf{b}\|^2$.
Recall that the **pseudoinverse** of a nonzero $m \times n$ matrix A with SVD $A = U\Sigma V^T$ is the $n \times m$ matrix $\boxed{A^+ = V\Sigma^+U^T}$.

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}_{m\times n}, \quad \Sigma^+ = \begin{pmatrix} \frac{1}{\sigma_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_r} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}_{n\times m}, \text{ pseudo inverse of } \Sigma.$$



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

Recalling: the least squares solution must satisfies the equation $\boxed{Ax = b_r}$.

Check: $x^* = A^+ b$ satisfies $\boxed{Ax = b_r}$.

[To see this] $x^* = A^+b = V\Sigma^+U^Tb = V\begin{bmatrix} \frac{1}{\sigma_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_r} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}\begin{bmatrix} u_1^T \\ \vdots \\ u_r^T \\ \vdots \\ u_n^T \end{bmatrix} b$

$$= V\begin{bmatrix} \frac{1}{\sigma_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_r} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}_{m\times n}\begin{bmatrix} \langle b, u_1\rangle \\ \vdots \\ \langle b, u_r\rangle \\ \vdots \\ \langle b, u_n\rangle \end{bmatrix}_{n\times 1}$$

$$Ax^* = (U\Sigma \underbrace{V^T)(V}_{I_n}\begin{bmatrix} \frac{1}{\sigma_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_r} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}\begin{bmatrix} \langle b, u_1\rangle \\ \vdots \\ \langle b, u_r\rangle \\ \vdots \\ \langle b, u_n\rangle \end{bmatrix}) = U\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}\begin{bmatrix} \langle b, u_1\rangle \\ \vdots \\ \langle b, u_r\rangle \\ \vdots \\ \langle b, u_n\rangle \end{bmatrix}$$

$= \langle b, u_1\rangle u_1 + \cdots + \langle b, u_r\rangle u_r$

$\| \; b_r$ *

Remark: The least squares solution $\mathbf{x}^*$ to the system $A\mathbf{x} = \mathbf{b}$ satisfies the **normal equation** $A^T A\mathbf{x} - A^T\mathbf{b} = 0$.

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 $A\mathbf{x} = \mathbf{b}$.

If A has n linearly independent columns($\ker A = \{\mathbf{0}\}$), then

$\to A^TA$ is invertible

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

where

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

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.

We write $A\vec{x} = b$, where $A = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix}$, $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

In lecture 35, we have found SVD of A is

$$A = \underbrace{\begin{pmatrix} \frac{1}{\sqrt2} & \frac{1}{\sqrt2} \\ \frac{1}{\sqrt2} & \frac{-1}{\sqrt2} \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} \sqrt6 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} \frac{1}{\sqrt3} & \frac{1}{\sqrt2} & \frac{1}{\sqrt6} \\ \frac{1}{\sqrt3} & \frac{-1}{\sqrt2} & \frac{1}{\sqrt6} \\ \frac{-1}{\sqrt3} & 0 & \frac{2}{\sqrt6} \end{pmatrix}^T}_{V^T}.$$

So, pseudo inverse of A is

$$A^+ = V \underbrace{\begin{bmatrix} \frac{1}{\sqrt6} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma^+} U^T = \frac{1}{6}\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{pmatrix}.$$

So, $x^* = A^+b = A^+\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$. *

§ Low rank approximations to a matrix.

Suppose we want to approximate a matrix $A = A_{m\times n}$ with rank r by a matrix $B = B_{m\times n}$ with rank $k < r$. That is, we want such B to minimize $\|A - B\|$.

Recall a matrix A with rank r has full SVD as follows:

$$A = U\Sigma V^T = [u_1 \cdots u_m]\begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}_{m\times n}\begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}, \quad \sigma_1 \geq \cdots \geq \sigma_r > 0.$$

If we take $A_k = [u_1 \cdots u_m]\underbrace{\begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_k & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}}_{\Sigma_k}\begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}$, $\sigma_{k+1} = \cdots = \sigma_r = 0$,

then the matrix A_k has rank $\underline{k}$. #

The best rank k approximation to A is

$$B = U\Sigma_k V^T.$$

$$= U\begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_k & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}V^T, \quad k < n.$$

Fact: This matrix B minimizes the distance to A as measured by Frobenius norm and operator norm:

$$\|A - B\|_F, \quad \|A - B\|_2.$$

Recall: $\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$; $\|A\|_2 = \sigma_1$ (largest singular value of A).

$$A - B = U\Sigma V^T - U\Sigma_k V^T = U\underline{(\Sigma - \Sigma_k)}V^T$$

$$= U\begin{array}{cc} & k \\ k & \begin{pmatrix} 0 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & & & \\ & & & \sigma_{k+1} & & & \\ & & & & \ddots & & \\ & & & & & \sigma_r & \\ & & & & & & 0 \\ & & & & & & & \ddots \\ & & & & & & & & 0 \end{pmatrix}\end{array} V^T$$

So, the largest singular value of $A - B$ is σ_{k+1}, thus $\|A - B\|_2 = \sigma_{k+1}$. Moreover,

$$\|A - B\|_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}. \quad \#$$

Example. Find the best rank 1 approximation to $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Find SVD of A. $A^TA = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. $0 = \det(A^TA - \lambda I)$

$\lambda = 1, 3$.

$\underline{\lambda = 3}$: $A^TA - 3I$. $v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$. $V = [v_1 \; v_2]$.

$\underline{\lambda = 1}$: $A^TA - I$. $v_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix}$.

$u_1 = \dfrac{Av_1}{\sqrt{3}} = \dfrac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$; $u_2 = \dfrac{Av_2}{\sqrt{1}} = \dfrac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$

Finding u_3 that is orthogonal to u_1, u_2, so

$$u_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

Thus, full SVD of A is

$$A = U \Sigma V^T = [u_1 \; u_2 \; u_3] \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} [v_1 \, v_2]^T.$$

⌜Reduced SVD of A is

$$\underset{3\times 2}{A} = \hat{U} \hat{\Sigma} \hat{V}^T = \underset{3\times 2}{[u_1 \; u_2]} \underset{2\times 2}{\begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}} \underset{2\times 2}{[v_1 \, v_2]^T}. \lrcorner$$

So, $B = U \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} V^T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. ✗

$\underline{\text{Exercise}}$: $\|A - B\|_F = \underline{\qquad}$