# 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

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

  and let $\Sigma$ 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

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

---

Today we will

- Continue Section 8.7 Singular Values

**- Lecture will be recorded -**

---

- HW 12 due today at 6pm

## § 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 $\{(t_i, \boxed{b_i})\}$:

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

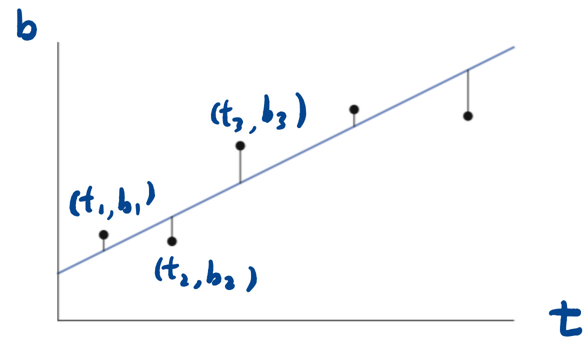


Figure 1: Least squares approximation of data by a straight line.

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

$$\|\vec{x}\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$$

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

definition $E :=$

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 solve

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

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

That means

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

**Definition:** Suppose that $A \in M_{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$.

(1) **The least squares problem** is to find $\mathbf{x}^* \in \mathbb{R}^n$ for which that the error $\|A\mathbf{x} - \mathbf{b}\|$ is minimized.

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

The least square solution $\mathbf{x}^*$ will satisfy this equation

$$\boxed{A^T A\mathbf{x} - A^T\mathbf{b} = 0,}$$ which is called the **normal equation**.

Solve x: $A^TAx = A^Tb$.

**Fact 5:** Suppose that $A \in M_{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$. If $A$ has $n$ linearly independent columns ($\mathrm{rank}(A) = n$), then **the least square solution** is

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

A has n l. indep. columns, i.e, rank(A) = n.

By Fact 1, rank$(A^TA)$ = rank(A) = n.

$A^TA$ is $n\times n$ matrix with rank = n $\Rightarrow$ $A^TA$ is invertible.

$\Rightarrow (A^TA)^{-1}$ exists.

$A^TAx = A^Tb \Rightarrow x^* = (A^TA)^{-1}A^Tb$. #

**Example 2.** Return to our experiment, suppose the collected data $(t_i, b_i)$ are $(1, 2), (2, 3), (3, 5), (4, 7)$. Then

$t_1$ $b_1$ $t_2$ $b_2$ $t_3$ $b_3$ $t_4$, $b_4$

$$\mathbf{b} = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \end{bmatrix}, \quad A = \begin{array}{cc} t & 1 \\ \end{array} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}. \qquad \boxed{Ax = b}$$

Find the least squares solution $x^*$.

A has 2 l. indep columns. By Fact 5,

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

$$A^TA = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}$$

$$(A^TA)^{-1} = \frac{1}{20}\begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix}$$

So, $$x^* = \frac{1}{20}\begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix} = \underline{\begin{pmatrix} 1.7 \\ 0 \end{pmatrix}}$$ #

This means linear model $b = x_1 t + x_2 = \underline{\underline{1.7t}}$

The error $E$ is $\|Ax^* - b\|_2^2$

$$= \left\| \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix}\begin{pmatrix} 1.7 \\ 0 \end{pmatrix} - \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix} \right\|_2^2$$

$$= \left\| \begin{pmatrix} 0.3 \\ 0.4 \\ 0.1 \\ 0.2 \end{pmatrix} \right\|_2^2 = (0.3)^2 + (0.4)^2 + 0.1^2 + 0.2^2$$

$$= \underline{0.3}$$ #

**Remark:** The method above may also be applied to different models. For example, if we consider a polynomial model $\boxed{b = x_1t^2 + x_2t + x_3}$ $(x_1, x_2, x_3$ to be determined). Then Here

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \quad A = \begin{bmatrix} t_1^2 & t_1 & 1 \\ t_2^2 & t_2 & 1 \\ \vdots & \vdots & \vdots \\ t_m^2 & t_m & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

§ In general, when $\mathsf{rank}(A) = r \leq n$, we use the "pseudoinverse" to find the least square solution.

**Definition:** Let $A$ be an $m \times n$ matrix with $\mathsf{rank}(A) = r$ and nonzero singular values $\sigma_1 \geq \ldots \geq \sigma_r$ and SVD

$\sigma_{r+1} = \cdots = \sigma_n = 0$

$$A = U\Sigma V^T. = U \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \\ \hline & & 0 & & \end{bmatrix}_{m\times n} V^T \quad \text{if } m \geq n$$

The **pseudoinverse** of $A$ is the $n \times m$ matrix

$$\underset{n\times m}{A^+} = V\Sigma^+U^T. = V \left[\begin{array}{cccccc|c} 1/\sigma_1 & & & & & & \\ & \ddots & & & & & \\ & & 1/\sigma_r & & & & 0 \\ & & & 0 & & & \\ & & & & \ddots & & \\ & & & & & 0 & \end{array}\right]_{n\times m} U^T$$

Here

$$\Sigma\Sigma^+ = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}_{m\times m}$$

$$\Sigma^+_{ij} = \begin{cases} \frac{1}{\sigma_i} & \text{if } i = j \leq r \\ 0 & \text{otherwise} \end{cases}$$

$$\underset{n\times m}{\Sigma^+}\,\underset{m\times n}{\Sigma} = \left[\begin{array}{ccc|c} 1/\sigma_1 & & & \\ & \ddots & & 0 \\ & & 1/\sigma_r \; 0 \cdots 0 & \end{array}\right] \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \cdots 0 \\ \hline & & 0 & \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}_{n\times n}$$

**Fact 6:** Suppose that $A \in M_{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then $\mathbf{x}^* = A^+\mathbf{b}$ is the **least squares solution** to the linear system $A\mathbf{x} = \mathbf{b}$.

Check : $x^* = A^+ b$ satisfies the "Normal equation":

$$(A^TA)\,x = A^Tb.$$

① $(A^TA)\,x^* = (U\Sigma V^T)^T(U\Sigma V^T)\,x^*$

$$= V\Sigma^T \underbrace{U^TU}_{I}\Sigma V^T x^*$$

$$= V\Sigma^T\Sigma V^T x^* = V \begin{bmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}_{n\times n} V^T \underbrace{x^*}_{V\Sigma^+U^Tb}$$

$$= V \begin{bmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} \Sigma^+U^Tb.$$

$$= V \underbrace{\left[\begin{array}{cccccc|c} \sigma_1 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_r & & & & 0 \\ & & & 0 & & & \\ & & & & \ddots & & \\ & & & & & 0 & \end{array}\right]}_{\Sigma^T} U^Tb.$$

② $A^Tb = V\Sigma^TU^Tb.$

Then ① = ②. #

**Fact 7:** Suppose that $A \in M_{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$ with rank$(A) = n$. Then

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

which gives

$$\mathbf{x}^* = \overbrace{A^+\mathbf{b}}^{(\text{rank}A\le n)\ \text{Fact 6}} = \overbrace{(A^TA)^{-1}A^T}^{\text{Fact 5}\ (\text{rank}A=n)}\mathbf{b}.$$

Exercise: $A^+ = (A^TA)^{-1}A^T$ if rank$(A) = n$.

**Example ~~2~~ 3.** Consider the linear system

$$\begin{cases} x+y-z=1, \\ x+y-z=0. \end{cases}$$

Find the best approximation to a solution having minimum norm, that is, find a least squares solution to this system.

$b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. $A = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix}$. $\boxed{A\begin{pmatrix} x \\ y \\ z \end{pmatrix} = b.}$

Previously, we have found in Ex 1.

$$A = U\Sigma V^T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}\begin{pmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}_{2\times 3}\underbrace{\begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \end{pmatrix}}_{V^T}.$$

Pseudo inverse

$$A^+ = V\Sigma^+U^T = V\begin{pmatrix} 1/\sqrt{6} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}_{3\times 2} U^T$$

The least squares solution is

$$x^* = A^+b = \frac{1}{6}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

Error: $\|Ax^* - b\|_2^2 = \|A\begin{pmatrix} 1/6 \\ 1/6 \\ -1/6 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix}\|_2^2 = \frac{1}{2}$. #

$\text{EX}: D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}. \quad \|D\|_2 = 3$

## § Revisit Matrix norm.

## Frobenius norm and Natural Matrix norm.

Let's consider $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \ldots + x_n^2}$.

$Tr(B) = b_{11} + \cdots + b_{nn}$ if $B = (b_{ij})$

The **natural matrix norm** of $A$ is

$$\|A\|_2 = \max\{\|A\mathbf{u}\|_2 : \|\mathbf{u}\|_2 = 1\}.$$

The **Frobenius norm** of a matrix $A = (a_{ij})$ is defined by

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2} = \sqrt{tr(A^T A)}$$

**Fact 8:** Let $A \in M_{m \times n}$ and $Q$ is an orthogonal matrix. Then

$$\|QA\|_2 = \|A\|_2, \qquad \|QA\|_F = \|A\|_F.$$

[To see this]

To be continued!