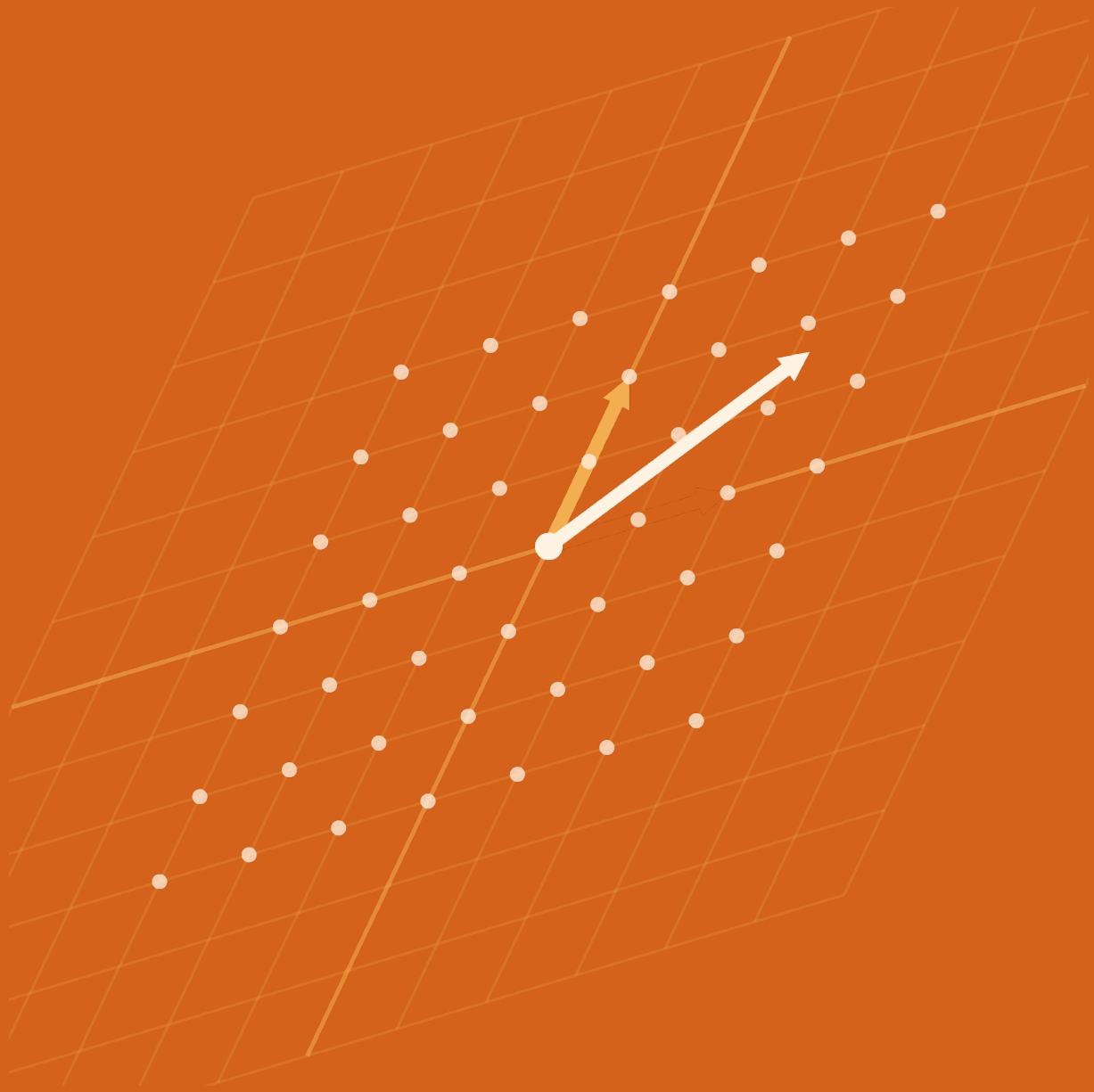
LINEAR ALGEBRA FOR AIML AND 6G

Least Squares



Lecture 21

Prof. Aditya K. Jagannatham
IIT Kanpur

Outline

1. **Least squares problem**
   Overdetermined systems and the normal equations
2. **Geometric intuition**
   The error is orthogonal to the column space
3. **Worked example**
   Fitting a 4 × 2 least squares problem

01

Least squares problem

Overdetermined systems and the normal equations



Tall matrix

- Let $\mathbf{A}$ be an $m \times n$ matrix with $m > n$
- Such a matrix $\mathbf{A}$ is known as a **tall matrix**

Overdetermined system

- Consider the **system of equations**

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

- Since $m > n$, this is termed as an overdetermined system
  - Such a system typically has **no solution**

Least squares problem

- We therefore find the solution which **minimizes** the error $\bar{\mathbf{e}} = \bar{\mathbf{b}} - \mathbf{A}\bar{\mathbf{x}}$.
- This can be formulated as the problem

$$\hat{\mathbf{x}} = \arg\min_{\bar{\mathbf{x}}} \left\| \bar{\mathbf{b}} - \mathbf{A}\bar{\mathbf{x}} \right\|^2$$

  - This is termed the **least squares** problem

Least squares solution

- The **least squares** solution can be found as follows
- Firstly, we expand the cost function as

$$\left\|\bar{\mathbf{b}} - \mathbf{A}\bar{\mathbf{x}}\right\|^2 = \left(\bar{\mathbf{b}} - \mathbf{A}\bar{\mathbf{x}}\right)^T \left(\bar{\mathbf{b}} - \mathbf{A}\bar{\mathbf{x}}\right)$$

$$= \left(\bar{\mathbf{b}}^T - \bar{\mathbf{x}}^T\mathbf{A}^T\right)\left(\bar{\mathbf{b}} - \mathbf{A}\bar{\mathbf{x}}\right)$$

$$= \bar{\mathbf{b}}^T\bar{\mathbf{b}} - \bar{\mathbf{x}}^T\mathbf{A}^T\bar{\mathbf{b}} - \bar{\mathbf{b}}^T\mathbf{A}\bar{\mathbf{x}} + \bar{\mathbf{x}}^T\mathbf{A}^T\mathbf{A}\bar{\mathbf{x}}$$

Least squares solution

- To find the **least squares solution**, we set

$$\frac{\partial}{\partial \bar{\mathbf{x}}} \left\| \bar{\mathbf{b}} - \mathbf{A}\bar{\mathbf{x}} \right\|^2 = \mathbf{0}$$

Least squares solution

- This implies that

$$\frac{\partial}{\partial \bar{\mathbf{x}}} \|\bar{\mathbf{b}} - \mathbf{A}\bar{\mathbf{x}}\|^2$$

$$= \frac{\partial}{\partial \bar{\mathbf{x}}} \left(\bar{\mathbf{b}}^T\bar{\mathbf{b}} - \bar{\mathbf{x}}^T\mathbf{A}^T\bar{\mathbf{b}} - \bar{\mathbf{b}}^T\mathbf{A}\bar{\mathbf{x}} + \bar{\mathbf{x}}^T\mathbf{A}^T\mathbf{A}\bar{\mathbf{x}}\right)$$

$$= \mathbf{0} - \mathbf{A}^T\bar{\mathbf{b}} - \left(\bar{\mathbf{b}}^T\mathbf{A}\right)^T + 2\mathbf{A}^T\mathbf{A}\bar{\mathbf{x}}$$

$$= 2\mathbf{A}^T\mathbf{A}\bar{\mathbf{x}} - 2\mathbf{A}^T\bar{\mathbf{b}} = \mathbf{0}$$

Least squares solution

- Therefore, the **least squares solution** is determined as

$$
\begin{aligned}
2\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} - 2\mathbf{A}^T\bar{\mathbf{b}} &= \mathbf{0} \\
\Rightarrow \mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} &= \mathbf{A}^T\bar{\mathbf{b}} \\
\Rightarrow \hat{\mathbf{x}} &= \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\bar{\mathbf{b}}
\end{aligned}
$$

02

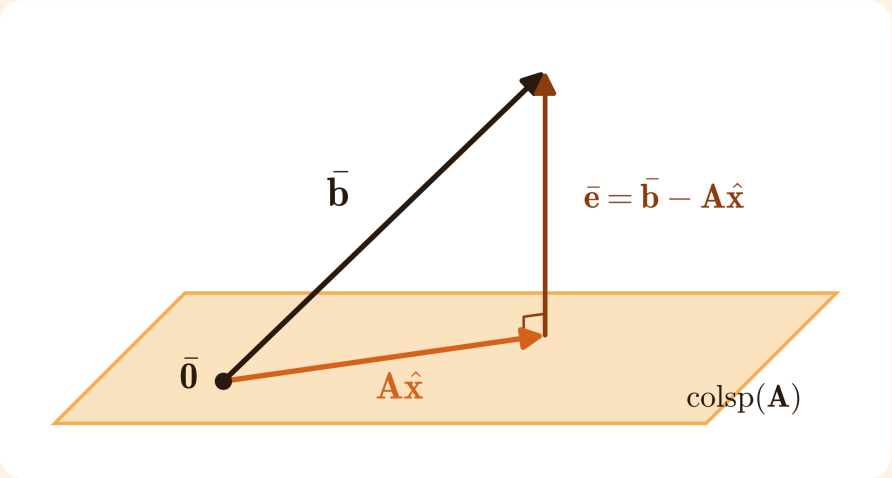
Geometric intuition



The error is orthogonal to the column space

Least squares intuition

- We now explore an **intuitive** solution to least squares
- Note that $\mathbf{A}\hat{\mathbf{x}}$ lies in the **column space** of $\mathbf{A}$
- Also $\bar{\mathbf{e}} = \bar{\mathbf{b}} - \mathbf{A}\hat{\mathbf{x}}$ is the error of approximation of $\bar{\mathbf{b}}$
- The error is minimum when it is **perpendicular** to $\text{colsp}(\mathbf{A})$



Least squares intuition

- Therefore, the minimum error should be **perpendicular** to each column of $\mathbf{A}$

$$\bar{\mathbf{a}}_1^T \bar{\mathbf{e}} = 0$$
$$\bar{\mathbf{a}}_2^T \bar{\mathbf{e}} = 0$$
$$\vdots$$
$$\bar{\mathbf{a}}_n^T \bar{\mathbf{e}} = 0$$

Least squares intuition

- This can be combined as

$$\begin{bmatrix} \bar{\mathbf{a}}_1^T \\ \bar{\mathbf{a}}_2^T \\ \vdots \\ \bar{\mathbf{a}}_n^T \end{bmatrix} \bar{\mathbf{e}} = \mathbf{0} \Rightarrow \mathbf{A}^T \bar{\mathbf{e}} = \mathbf{0}$$

Least squares intuition

- Therefore,

$$\mathbf{A}^T\left(\bar{\mathbf{b}} - \mathbf{A}\hat{\mathbf{x}}\right) = \mathbf{0}$$

$$\Rightarrow \mathbf{A}^T\bar{\mathbf{b}} - \mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{0}$$

$$\Rightarrow \mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\bar{\mathbf{b}}$$

$$\Rightarrow \hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\bar{\mathbf{b}}$$

03

Worked example

Fitting a 4 × 2 least squares problem



Least squares – example

- Solve the **LS problem** below and calculate $\hat{\mathbf{x}}$

$$\min \left\| \underbrace{\begin{bmatrix} 2 \\ 0 \\ 3 \\ -1 \end{bmatrix}}_{\bar{\mathbf{b}}} - \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}}_{\mathbf{A}} \bar{\mathbf{x}} \right\|^2$$

Least squares – example

- We start as follows

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}$$

Least squares – example

- Next

$$(\mathbf{A}^T\mathbf{A})^{-1} = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}^{-1}$$

$$= \frac{1}{20}\begin{bmatrix} 14 & -6 \\ -6 & 4 \end{bmatrix} = \begin{bmatrix} \frac{7}{10} & -\frac{3}{10} \\ -\frac{3}{10} & \frac{1}{5} \end{bmatrix}$$

Least squares – example

- Subsequently

$$(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \frac{1}{20}\begin{bmatrix}14 & -6\\ -6 & 4\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\\ 0 & 1 & 2 & 3\end{bmatrix}$$

$$= \frac{1}{20}\begin{bmatrix}14 & 8 & 2 & -4\\ -6 & -2 & 2 & 6\end{bmatrix}$$

Least squares – example

- Finally

$$\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\bar{\mathbf{b}} = \frac{1}{20}\begin{bmatrix} 14 & 8 & 2 & -4 \\ -6 & -2 & 2 & 6 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \\ 3 \\ -1 \end{bmatrix}$$

$$= \frac{1}{20}\begin{bmatrix} 38 \\ -12 \end{bmatrix} = \begin{bmatrix} \frac{19}{10} \\ -\frac{3}{5} \end{bmatrix}$$

Summary

- A tall $\mathbf{A}$ $(m > n)$ gives an **overdetermined** system $\bar{\mathbf{b}} = \mathbf{A}\bar{\mathbf{x}}$, usually with no exact solution
- **Least squares** minimizes the error $\|\bar{\mathbf{b}} - \mathbf{A}\bar{\mathbf{x}}\|^2$
- Setting the gradient to zero gives $\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\bar{\mathbf{b}}$
- The solution is $\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\bar{\mathbf{b}}$
- Geometrically, the error $\bar{\mathbf{b}} - \mathbf{A}\hat{\mathbf{x}}$ is orthogonal to the column space of $\mathbf{A}$



Thank You!

Questions?