EE 205003 Session 18

Che Lin

Institute of Communications Engineering

Department of Electrical Engineering

Least Square Approximation

Issue : It often happens that $A\mathbf{x} = \mathbf{b}$ has no solution
($m > n, \mathbf{C}(A)$ only spans a small part of $\mathbb{R}^m$.
If $\mathbf{b} \notin \mathbf{C}(A)$, no solution)

Q: Do we stop here ?

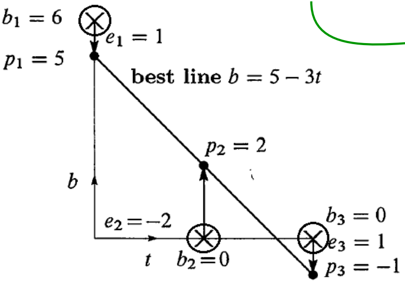
No! measurement includes noise.
Instead, we try to find the "best solution"

To repeat : we cannot always get error $\mathbf{e} = \mathbf{b} - A\mathbf{x}$ down to zero
When $\mathbf{e} = \mathbf{0}$, $\mathbf{x}$ is exact solution to $A\mathbf{x} = \mathbf{b}$
When $\mathbf{e}$ is as small as possible,
$\widehat{\mathbf{x}}$ is the least square solution, or "best solution"
($\mathbf{p} = A\widehat{\mathbf{x}}$ is the projection of $\mathbf{b}$ onto $\mathbf{C}(A)$.
To find the "best solution", we solve $A^\intercal A\widehat{\mathbf{x}} = A^\intercal\mathbf{b}$.)

Ch 4.3 Least Squares Approximations

Ex : Fitting a line (linear regression)

Find the <u>closest line</u> through $(0, 6)$, $(1, 0)$, $(2, 0)$



errors = vertical distances to line

$(b = C + Dt)$

If $(0, 6)$ on the line,
$C + D \cdot 0 = 6$

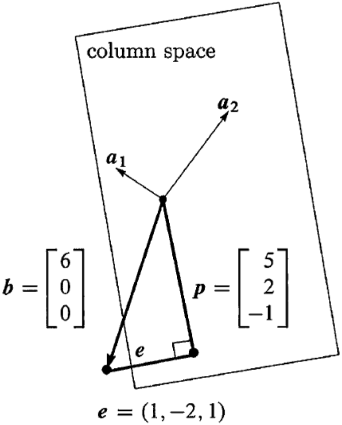
If $(1, 0)$ on the line,
$C + D \cdot 1 = 0$

If $(2, 0)$ on the line,
$C + D \cdot 2 = 0$

3 equations, 2 unknowns : $A\mathbf{x} = \mathbf{b}$

$$\Rightarrow A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} C \\ D \end{bmatrix} \mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

(not in $\mathbf{C}(A)$, no solution)



By Geometry

Every $\mathbf{Ax}$ lies on the plane $\mathbf{C}(A)$
want to find the point closest to $\mathbf{b}$
$\Rightarrow$ The nearest point is projection

$$P = A\widehat{\mathbf{x}}$$

Normal equation : $A^\intercal A\widehat{\mathbf{x}} = A^\intercal \mathbf{b}$
(same as Ex3 in session 17, we already computed $\widehat{\mathbf{x}} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$)
$\Rightarrow b = 5 - 3t$ is the "best" line
(linear regression works if no outlier)

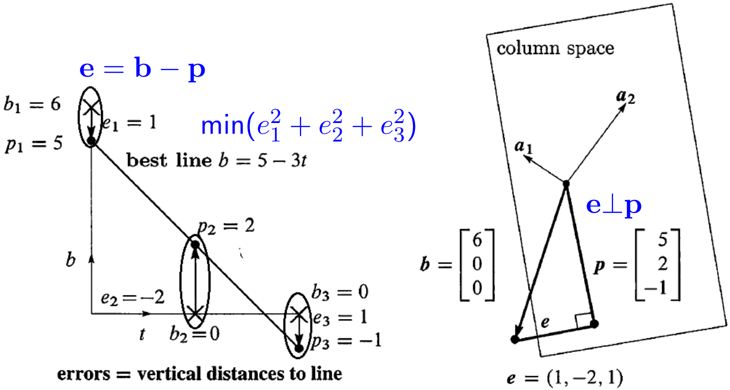


Figure 4.6: **Best line and projection: Two pictures, same problem.** The line has heights $\boldsymbol{p} = (5, 2, -1)$ with errors $\boldsymbol{e} = (1, -2, 1)$. The equations $A^{\mathrm{T}} A \widehat{\boldsymbol{x}} = A^{\mathrm{T}} \boldsymbol{b}$ give $\widehat{\boldsymbol{x}} = (5, -3)$. The best line is $b = 5 - 3t$ and the projection is $\boldsymbol{p} = 5\boldsymbol{a}_1 - 3\boldsymbol{a}_2$.

By Algebra

Every $\mathbf{b} = \underset{\in \mathbf{C}(A)}{\mathbf{p}} + \underset{\in \mathbf{N}(A^\intercal)}{\mathbf{e}}$

orthogonal components

$A\mathbf{x} = \mathbf{b} = \mathbf{p} + \mathbf{e}$ (impossible) $A\widehat{\mathbf{x}} = \mathbf{p}$ (solvable) (by removing $\mathbf{e}$)

For all $\mathbf{x}$,

$$\|A\mathbf{x} - \mathbf{b}\|^2 = \|A\mathbf{x} - \mathbf{p} - \mathbf{e}\|^2 = \|\underbrace{A\mathbf{x} - \mathbf{p}}_{\in \mathbf{C}(A)}\|^2 + \|\underbrace{\mathbf{e}}_{\in \mathbf{N}(A^\intercal)}\|^2$$

$\widehat{\mathbf{x}}$ makes $\|A\mathbf{x} - \mathbf{p}\|^2 = 0$

this leaves the smallest possible error $\mathbf{e}$

$\boxed{\text{Fact}}$ The least square sol. $\widehat{\mathbf{x}}$ makes

$E = \|A\mathbf{x} - \mathbf{b}\|^2$ as small as possible

By Calculus

$$E = \left\|A\mathbf{x} - \mathbf{b}\right\|^2 = (C + D \cdot 0 - 6)^2 + (C + D \cdot 1 - 0)^2 + (C + D \cdot 2 - 0)^2$$

$$\frac{\partial E}{\partial C} = \cancel{2}(C + D \cdot 0 - 6)^2 + \cancel{2}(C + D \cdot 1)^2 + \cancel{2}(C + D \cdot 2)^2 = 0$$

$$\frac{\partial E}{\partial D} = \cancel{2}(C + D \cdot 0 - 6)^2(0) + \cancel{2}(C + D \cdot 1)^2(1) + \cancel{2}(C + D \cdot 2)^2(2) = 0$$

$$\Rightarrow \quad \begin{matrix} 3C + 3D = 6 \\ 3C + 5D = 0 \end{matrix} \qquad \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \text{ is } A^\intercal A$$

(same as $A^\intercal A\mathbf{x} = A^\intercal\mathbf{b}$) $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$

$\boxed{\text{Fact}}$ The partial derivatives of $\left\|A\mathbf{x}-\mathbf{b}\right\|^2$ are zero when $A^\intercal A\widehat{\mathbf{x}} = A^\intercal\mathbf{b}$

Recall :

$b = 5 - 3t$ is the "best" line

$$\begin{aligned} t &= 0, P_1 = 5 - 0 = 5 \\ t &= 1, P_2 = 5 - 3 = 2 \\ t &= 2, P_3 = 5 - 6 = -1 \end{aligned} \quad \Rightarrow \mathbf{p} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

$$\Rightarrow \mathbf{e} = \mathbf{b} - \mathbf{p} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \quad \mathbf{e}\perp\mathbf{p}$$ ($\perp$ col.s of A)

The Big Picture

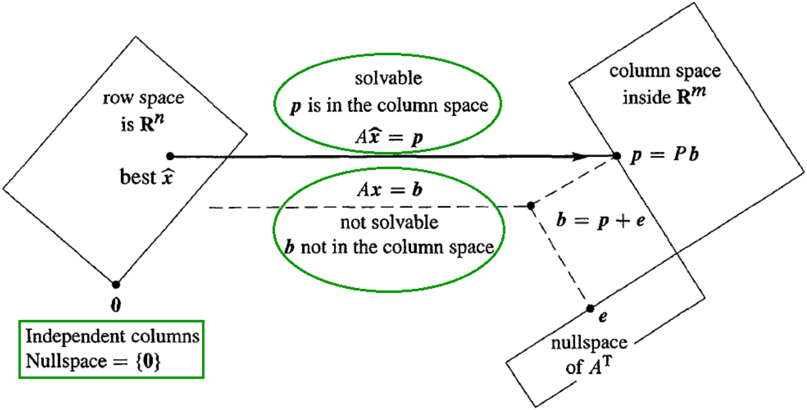


Figure 4.7: The projection $p = A\hat{x}$ is closest to b, so $\hat{x}$ minimizes $E = \|b - Ax\|^2$.

Recall :

If A has independent col.s, then $AA^\intercal$ invertible

$\Rightarrow$ we can solve for least square solution $\widehat{\mathbf{x}}$

$\Rightarrow$ we can use linear regression to find approximate solution to unsolvable $A\mathbf{x} = \mathbf{b}$

(col.s of A is guaranteed to be independent if they are orthonormal)

(Topic for next session)