EE203001 SES # 18

Least square approx.

Issue: It often happens that

$A\underline{x} = \underline{b}$ has no sol.

($m > n$, $C(A)$ only spans a small part of R^m. If $\underline{b} \notin C(A)$, no sol.)

Q: Do we stop here?

No! measurement includes noise

Instead, we try to find the "best sol."

To repeat: we cannot always get error $\underline{e} = \underline{b} - A\underline{x}$ down to zero

When $\underline{e} = \underline{0}$, $\underline{x}$ is exact sol. to $A\underline{x} = \underline{b}$

When $\underline{e}$ is as small as possible, $\hat{x}$ is the least square sol. or "best sol."

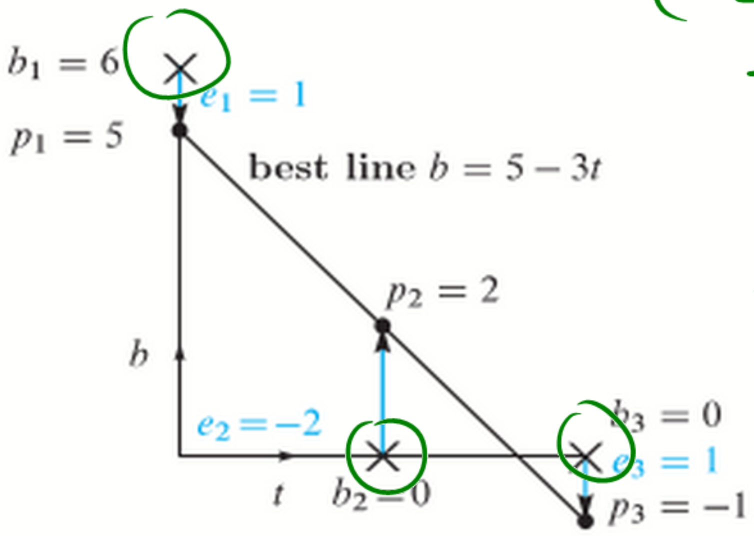
($\underline{p} = A\underline{\hat{x}}$ is the projection of $\underline{b}$ onto $C(A)$. To find the "best sol.", we solve $A^TA\underline{\hat{x}} = A^T\underline{b}$)

Ex: Fitting a line (linear regression)

Find the closest line through $(0,6)$ $(1,0)$, $(2,0)$

$\Updownarrow$

$(b = C + Dt)$



errors = vertical distances to line

If $(0,6)$ on the line

$C + D \cdot 0 = 6$

If $(1,0)$ ... ...

$C + D \cdot 1 = 0$

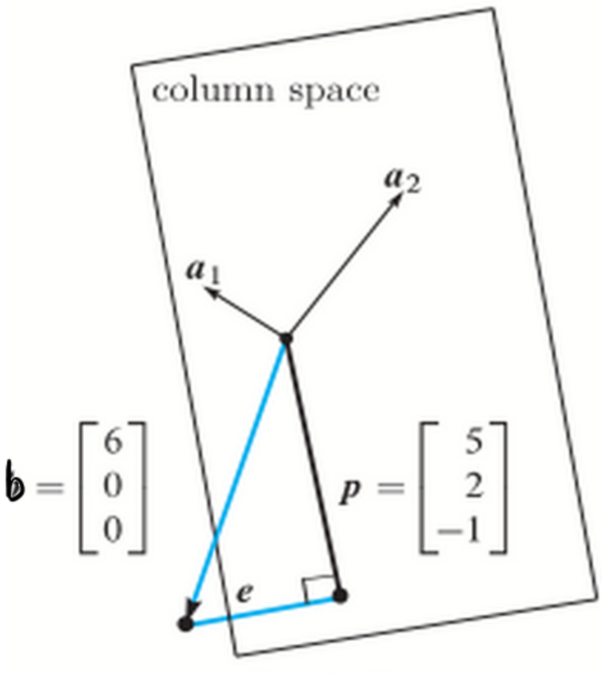
If $(2,0)$ ... ...

$C + D \cdot 2 = 0$

3 equs, 2 unknowns; $A\underline{x} = \underline{b}$

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

(not in $C(A)$, $\Leftrightarrow$ no sol.)



$e = (1, -2, 1)$

<u>By geometry</u>

Every $A\underline{x}$ lies on the plane $C(A)$. Want to find the point closest to $\underline{b}$ $\Rightarrow$ The nearest point is projection

$\underline{p} = A\hat{\underline{x}}$

Normal eqn :

$$A^T A \underline{\hat{x}} = A^T \underline{b}$$

(same as Ex 3 in SES-17, we already computed $\hat{x} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$)

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

(linear regression works if no outlier)

$\underline{e} = \underline{b} - \underline{p}$ $\qquad$ $\underline{e} \perp \underline{p}$

$\min(e_1^2 + e_2^2 + e_3^2)$

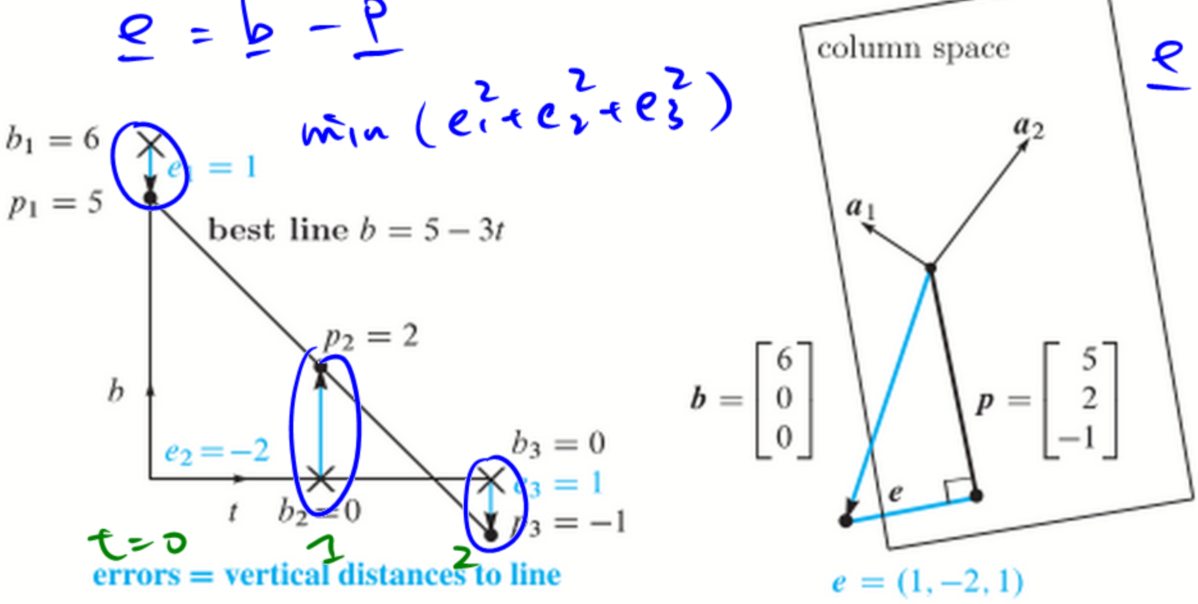


Figure 28: **Best line and projection: Two pictures, same problem.** The line has heights $\boldsymbol{p} = (5, 2, -1)$ with errors $\boldsymbol{e} = (1, -2, 1)$. The equations $A^T A \hat{\boldsymbol{x}} = A^T \boldsymbol{b}$ give $\hat{\boldsymbol{x}} = (5, -3)$. The best line is $b = 5 - 3t$ and the projection is $\boldsymbol{p} = 5\boldsymbol{a}_1 - 3\boldsymbol{a}_2$.

By algebra

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

orthogonal complements

$A\underline{x} = \underline{b} = \underline{p} + \underline{e}$ (impossible)

$A\underline{\hat{x}} = \underline{p}$ (solvable) (by removing $\underline{e}$)

For any $\underline{x}$,

$$\|A\underline{x} - \underline{b}\|^2 = \|A\underline{x} - \underline{p} - \underline{e}\|^2$$

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

$\hat{x}$ makes $\|A\hat{\underline{x}} - \underline{p}\|^2 = 0$

this leaves the smallest possible error

$\underline{e}$

$\boxed{\text{Fact}}$ The least square sol. $\hat{x}$ makes $E = \|A\underline{x} - b\|^2$ as small as possible

By calculus

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

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

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

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

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

$\boxed{\text{Fact}}$ The partial derivatives of $\|A\underline{x}-b\|^2$ are zero when $A^TA\hat{\underline{x}} = A^T\underline{b}$

Recall:

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

$$\left.\begin{array}{ll} t=0, & p_1 = 5-0 = 5 \\ t=1, & p_2 = 5-3 = 2 \\ t=2, & p_3 = 5-6 = -1 \end{array}\right. \Rightarrow \underline{p} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

$$\Rightarrow \underline{e} = \underline{b} - \underline{p} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \quad \underline{e} \perp \underline{p}$$

($\perp$ col.s of A)

The big picture

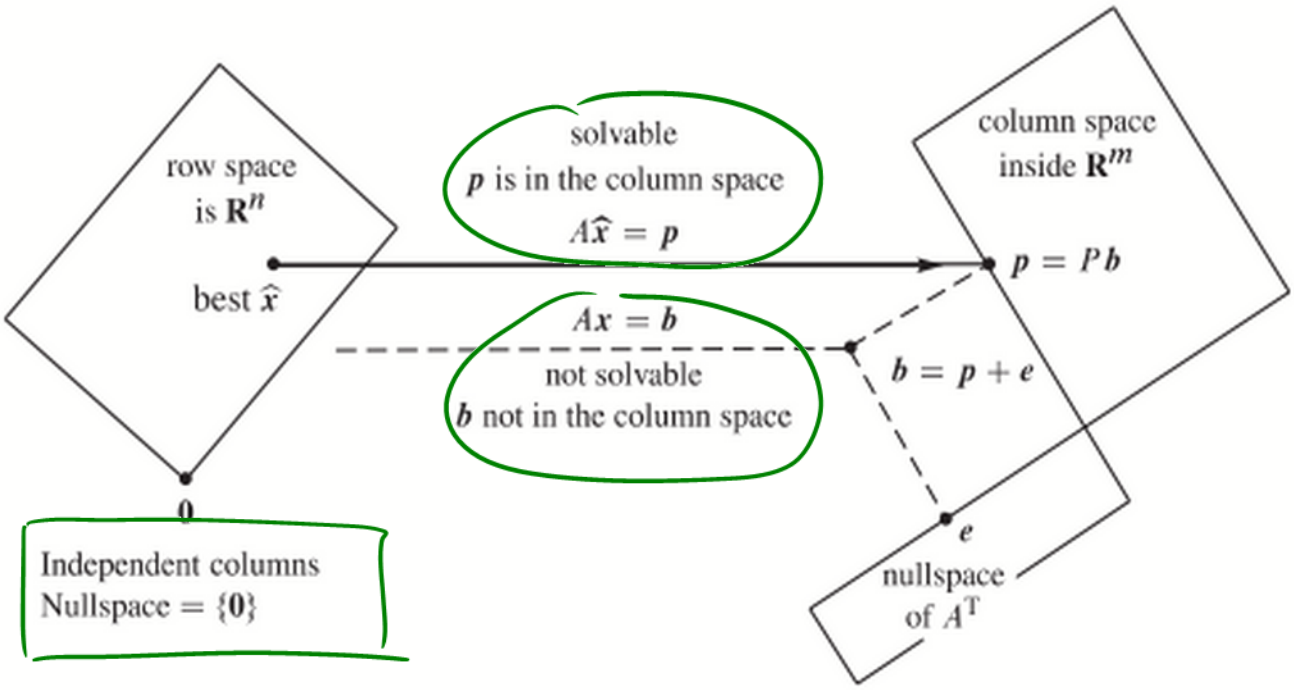


Figure 29: The projection $p = A\hat{x}$ is closest to b, so $\hat{x}$ minimizes $E = \|b - Ax\|^2$.

Recall:

If A has indep. col.s, then $A^T A$ invertible

$\Rightarrow$ we can solve for least square sol. $\underline{\hat{x}}$

$\Rightarrow$ we can use linear regression to find approx. sol. to unsolvable $A\underline{x} = \underline{b}$

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

(Topic for next session)