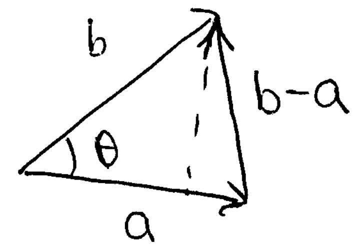
§ 3.3 Projections & Least squares —

- Projection onto a line — (Law of Cosines) $\mathbb{R}^n$

$$\|b-a\|^2 = \|a\|^2 + \|b\|^2 - 2\|a\|\|b\|\cos\theta$$

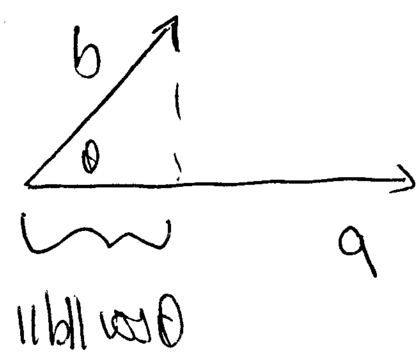


$$(a-b)\cdot(a-b) = \|a\|^2 - 2a\cdot b + \|b\|^2$$

$$\Rightarrow \quad 2ab = 2\|a\|\|b\|\cos\theta$$

$$\boxed{\cos\theta = \frac{a\cdot b}{\|a\|\|b\|}}$$

- $$\mathrm{Proj}_a b = \underbrace{(\|b\|\cos\theta)}_{\text{length}} \underbrace{\frac{a}{\|a\|}}_{\text{direction}}$$



$$\mathrm{Proj}_a b = \frac{a\cdot b}{\|a\|^2}\, a = a\,\frac{a^T\cdot b}{\|a\|^2} = \underbrace{\frac{a\cdot a^T}{\|a\|^2}}_{P_{n\times n}}\, b$$

$$\underbrace{\underset{n\times 1}{\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}} \underset{1\times n}{[a_1 \cdots a_n]}}\ \underset{n\times 1}{\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}}$$

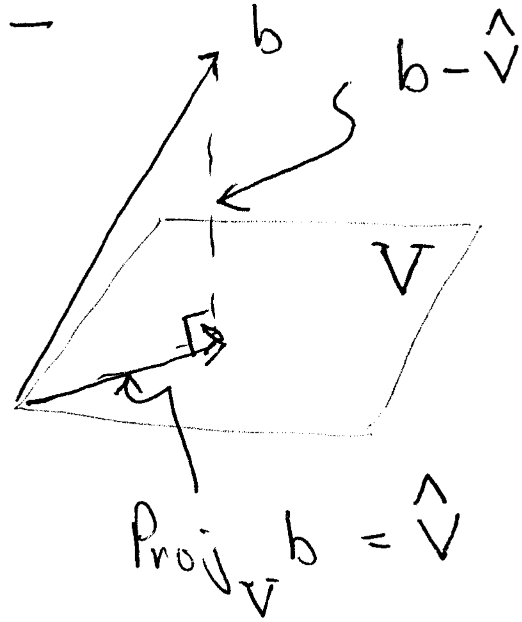
$$\underset{n\times n}{a\cdot a^T} = \begin{bmatrix} a_1 \begin{matrix}|\\a\\|\end{matrix}, a_2 \begin{matrix}|\\a\\|\end{matrix}, \ldots, a_n \begin{matrix}|\\a\\|\end{matrix} \end{bmatrix} = \begin{bmatrix} a_1(-\,a\,-) \\ a_2(-\,a\,-) \\ \vdots \\ a_n(-\,a\,-) \end{bmatrix}$$

Projection onto a subspace $V \subseteq \mathbb{R}^m$

$V = \text{Span}\{c_1, \dots, c_n\} \subseteq \mathbb{R}^m \quad (m > n)$

- Form the matrix of columns –

$$A = \begin{bmatrix} | & & | \\ c_1 & \cdots & c_n \\ | & & | \end{bmatrix}$$



- Find $\in V$ closest to b

Any $v \in V = \text{Span}\{c_1, \dots, c_n\}$ is given by

$$v = Ax \qquad \text{some } x \in \mathbb{R}^n$$

$$= \sum_{i=1}^{n} x_i \begin{bmatrix} | \\ c_i \\ | \end{bmatrix}$$

Thus: as that $(b - \hat{v}) \perp V$

or

$$\boxed{b - A\hat{x} \perp V}$$

But: $\quad b - A\hat{x} \perp \bar{V}$

$$\Leftrightarrow \quad [b - A\hat{x}]^T \begin{bmatrix} | \\ c_i \\ | \end{bmatrix} = 0 \qquad i = 1, \dots, n$$

$$\Leftrightarrow \quad [b - A\hat{x}]^{\perp} \cdot \begin{bmatrix} | & & | \\ c_1 & \cdots & c_n \\ | & & | \end{bmatrix} = 0$$

$$\Leftrightarrow \quad [b - A\hat{x}]^{\perp} A = 0$$

$$\Leftrightarrow \quad A^T [b - A\hat{x}] = 0$$

$$\Leftrightarrow \quad A^T b - A^T A \hat{x} = 0$$

$(n \times m)(m \times 1) \quad (n \times m)(m \times n) \quad (n \times 1)$

$$\Leftrightarrow \quad \boxed{A^T A \hat{x} = A^T b}$$

- **Thm** If $\{c_1, \ldots, c_n\}$ is a basis for V, then A^TA has an inverse -

I.e. $m > n$ so

$$A^TA = \underset{n \times m}{\begin{bmatrix} & A^T & \end{bmatrix}} \underset{m \times n}{\begin{bmatrix} \\ A \\ \\ \end{bmatrix}} = \underset{n \times n}{\begin{bmatrix} A^TA \end{bmatrix}}$$

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

$$(A^TA)\hat{x} = A^Tb$$

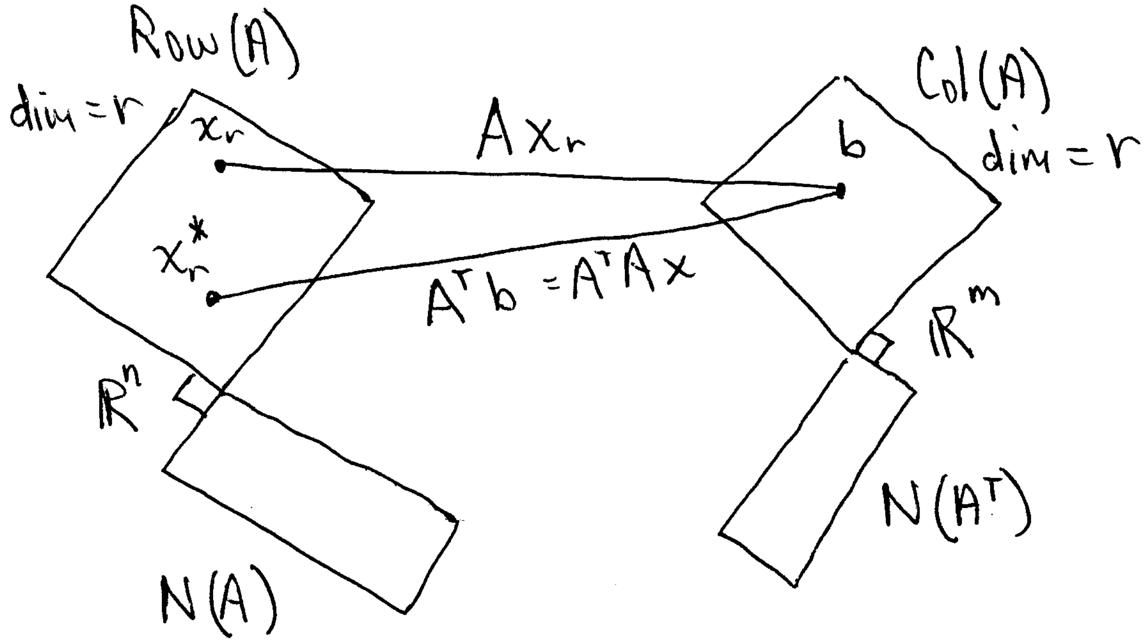
$$\hat{x} = (A^TA)^{-1}A^Tb$$

$$V = A\hat{x} = \underbrace{A(A^TA)^{-1}A^T}_{\text{"The } m \times m \text{ matrix that projects } b \text{ onto } V\text{"}} b$$

Pf of Thm: It suffices to show

Lemma: The rank of (A^TA) = rank of A

Picture:



"Idea $A^TA : \text{Row}(A) \longrightarrow \text{Row}(A)$ 1-1 and onto

so Span {rows of A} = span {rows of A^TA}

$\Rightarrow$ r = rank A = rank (A^TA)"

For an alternative proof...

It suffices to prove

$$N(A) = N(A^TA) \qquad (*)$$

I.e.

$$r = n - \dim(N(A))$$

$$\operatorname{rank}(A^TA) = n - \dim(N(A^TA))$$

For $(*)$:

If $x \in N(A)$ then $Ax = 0 \Rightarrow A^TAx = A^T0 = 0$

$\therefore x \in N(A^TA) \Rightarrow N(A) \subseteq N(A^TA)$

If $x \notin N(A)$ then

$$Ax \in \operatorname{Col}(A) \quad Ax \neq 0$$

But $A^T : \operatorname{Col}(A) \to \operatorname{Row}(A) = \operatorname{Col}(A^T)$ is 1-1 onto

$$\therefore A^T \underbrace{(Ax)}_{\text{nonzero}} = A^TAx \neq 0$$

So A^TA has no elements of the nullspace except $N(A)$

start Friday

Corollary: If $m > n$, then A^TA is invertible iff the colms of A are indept

Pf. $\operatorname{rank}(A^TA) = \operatorname{rank}(A)$. If colms of A are indept, then $\operatorname{rank}(A) = n$. Thus

$\underset{(n\times m)(m\times n)}{A^T A}$ is an $n\times n$ matrix of rank $n \Rightarrow$ invertible

Summary: Among all vectors $v \in V = \operatorname{Col}(A)$

$$Pb = A(A^TA)^{-1}A^Tb$$

minimizes the distance $\|b - Ax\|$ (or $\|b-Ax\|^2$)

$$\|b - Ax\| = \sqrt{(b_1-(Ax)_1)^2 + \cdots + (b_n - (Ax)_n)^2}$$

$$\|b - Ax\|^2 = \sum_{i=1}^{n} (b-Ax)_i^2$$

We Say: "Pb is the least squares best estimate for b in V"

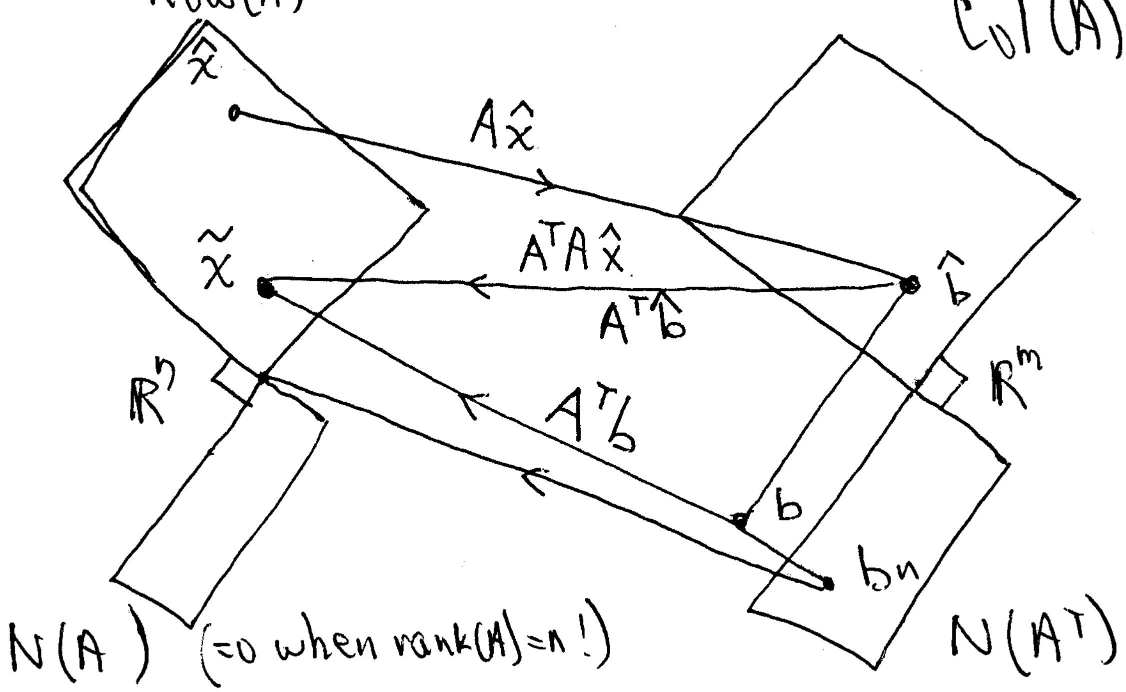
Note : $A\,(A^TA)^{-1}A^T b = Pb = \hat{b}$

$$\underset{(m\times n)}{\begin{bmatrix} A \end{bmatrix}}\ \underset{(n\times n)}{\left[(A^TA)^{-1}\right]}\ \underset{n\times m}{\left[\quad A^T \quad\right]}\ \underset{(m\times 1)}{\begin{bmatrix} b \end{bmatrix}} = \underset{(m\times 1)}{\begin{bmatrix} \hat{b} \end{bmatrix}}$$

$$\mathrm{Col}(A) \ni \hat{b} \xleftarrow{A} \mathrm{Row}(A) \xleftarrow{(A^TA)^{-1}} \mathrm{Row}(A) \xleftarrow{A^T} b \in \mathbb{R}^m$$

Picture:

$\hat{x} = (A^TA)^{-1}\tilde{x}$



"P takes any $b \in \mathbb{R}^m$ to $\hat{b} \in \mathrm{Col}(A)$ closest to b"

"P takes $\hat{b} \in \mathrm{Col}(A)$ to $\hat{b}$"

■ Projection Matrices : P has following Properties

① $P^2 = P$ (P is a projection)

② $P^T = P$ (P is an orthogonal projection)

check ① : $P^2 = A\underbrace{\cancel{(A^TA)^{-1}}\,\cancel{A^TA}}_{I}(A^TA)^{-1}A^T = P$ ✓

check ② : $P^T = (A(A^TA)^{-1}A^T)^T = A\,[A^TA]^{T^{-1}}A^T = A(A^TA)^{-1}A^T = P$ ✓

Theorem : Every matrix that satisfies ① & ② is the "least squares" projection onto Col(P)

Pf. If P is projection then ① & ② hold.

Now assume ① & ② hold:

<u>Claim</u>: $\|Pb - b\|$ is minimized among all $v \in \mathrm{Col}(P)$. Need $Pb - b \perp \mathrm{Col}(P)$, or equiv

$$(Pb-b)^T P = 0$$

$$\Leftrightarrow \quad P^T (Pb-b) = 0$$

$$\Leftrightarrow \quad P(Pb-b) = P^2 b - Pb = Pb - Pb = 0 \quad \checkmark$$

<u>Summary</u>: "Start with colms of A, $\min \|b - Ax\|$

$$\Rightarrow \quad \hat{x} = (A^TA)^{-1}A^Tb$$

$$\hat{b} = A\hat{x} = \underbrace{A(A^TA)^{-1}A^T}_{P} b$$

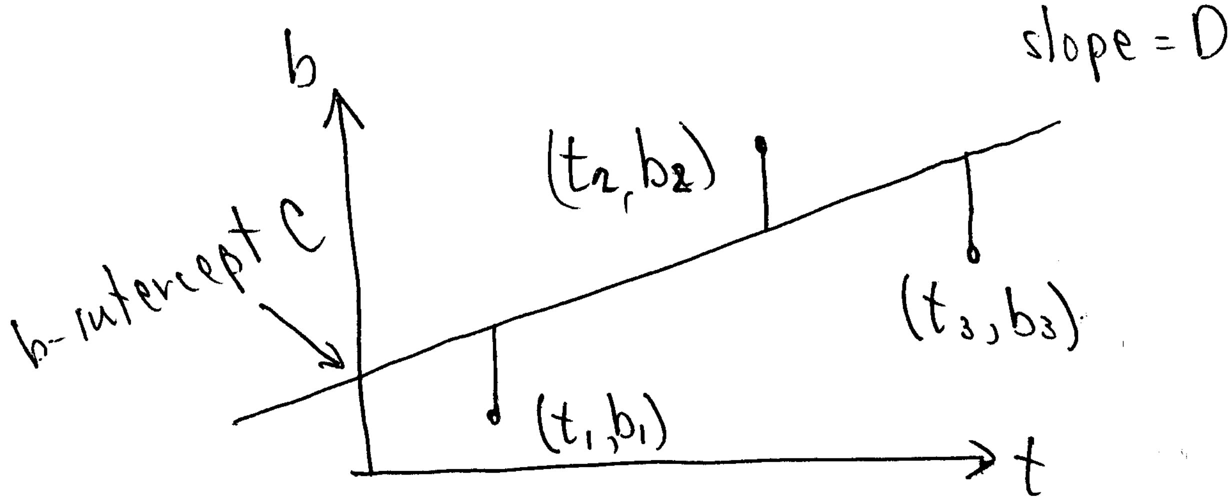
assume a basis for V, $m \geq n$

Now colms of P span V also, and the only properties needed to show its a projection are $P^2 = P$ & $P^T = P$!

▫ Application : (Fundamental) Least squares fit of a line...

- Consider the line $b = C + Dt$

graph



- Say we get measurements (t_i, b_i) $i = 1, \dots, m$

Q: what line fits the data "best"

Least squares "minimize sum of squares of vertical distances"

(Measured value at t_i) $= b_i$
(Actual value at t_i) $= C + Dt_i$

Minimize : $\sum_{i=1}^{m} \| b_i - (C + Dt_i) \|^2$

Minimize : $\sum_{i=1}^{m} \| b_i - (C+Dt_i) \|^2$

$$\Leftrightarrow \quad \Big\| \underbrace{b}_{m\times 1} - \underbrace{\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}}_{A_{m\times 2}} \underbrace{\begin{bmatrix} C \\ D \end{bmatrix}}_{x = \begin{bmatrix} C \\ D \end{bmatrix}} \Big\|^2$$

$$b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}, \quad x = \begin{bmatrix} C \\ D \end{bmatrix}$$

$\Leftrightarrow$ Minimize $\| b - Ax \|^2$

Soln : $\boxed{\hat{x} = (A^T A)^{-1} A^T b}$ (least squre formula)

Formula For $\hat{x} = \begin{pmatrix} C \\ D \end{pmatrix}$:

$$\hat{x} = (A^TA)^{-1}A^Tb$$

$$\Leftrightarrow \quad A^TA\hat{x} = A^Tb$$

$$A^TA = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \sum_{i=1}^m t_i \\ \sum_{i=1}^m t_i & \sum_{i=1}^m t_i^2 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m b_i \\ \sum_{i=1}^m t_i b_i \end{bmatrix}$$

$$\therefore \quad A^TA\hat{x} = A^Tb \quad \Leftrightarrow$$

$$\begin{bmatrix} m & \sum_{i=1}^m t_i \\ \sum_{i=1}^m t_i & \sum_{i=1}^m t_i^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m b_i \\ \sum_{i=1}^m t_i b_i \end{bmatrix}$$

2x2 matrix problem !

Ex: What line fits the data best in least squares sense?

Data: $(-1,1)$, $(1,1)$, $(2,3)$

Soln: Look for $b = C + Dt$

$$b = \text{vector of outputs} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

Minimize: $\sum (b_i - (C + Dt_i))^2$.

$$\Leftrightarrow \left\| b - \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} \right\|^2 \qquad A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\| b - A x \|^2$$

$$\hat{x} = \qquad A^T b = A^T A \hat{x} \qquad \hat{x} = (A^T A)^{-1} A^T b$$

$$(A^TA) = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

Want: $\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix} \Rightarrow \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 9 \\ 4 \end{bmatrix} \frac{1}{7}$

$$b = \frac{9}{7} + \frac{4}{7}t$$

is line of best fit —