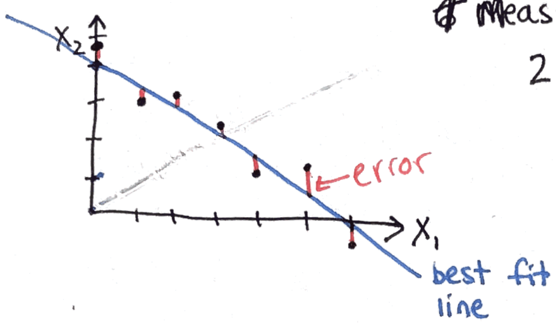
M3 Lecture 4 $\boxed{\text{Least Squares}}$ cont'd.

- algorithim for finding a solution when no unique sol'n exists
  - e.g. case of imperfect measuring
  - OR: how to average measurements that are different

find best estimate in 'least squares' sense
↳ minimize $(\text{error})^2$

Example: Fit to a line:



Measurements of (a_i, b_i): but only need to solve for 2 unknowns!

$$a_i x_1 + x_2 = b_i$$

Mtx form:
$$\begin{bmatrix} a_1 & 1 \\ a_2 & 1 \\ \vdots & \\ a_n & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

What happens if I try GE?
"inconsistent"
↳ $\boxed{\text{No Solution}}$
So then what? Find best estimate (min error)

How to find eqn of line ↑ that "most closely" matches data?

↳ Project data onto allowed line? (but don't know line...)

↳ Minimize "error" (difference b/w data & line)

↳ LS: minimize error squared cause |abs| is annoying...

Define Error: $\boxed{\vec{e} = A\vec{x} - \vec{b}}$ *if data not inconsistent, then error → 0

vector error has direction, but we only care about magnitude!

So minimize $\|\vec{e}\|^2$:

$$\underbrace{\min_{\vec{x}}}_{\text{find the } \vec{x}} \overbrace{\|\vec{e}\|^2}^{\text{that minimizes this}} = \min_{\vec{x}} \left[\| A\vec{x} - \vec{b} \|^2 \right]$$

$$\begin{bmatrix} a_1 & 1 \\ a_2 & 1 \\ \vdots & \vdots \\ a_n & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}}_{\vec{e}}$$

← ERROR term is now explicit
but it error can be anything → $\boxed{\infty \text{ sol'ns}}$
↳ Find the one with min error^2!
How? error ⊥ subspace (A)

So error vector $\vec{e}$ should be ~~the~~ orthogonal to all the columns of A matrix!

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}^T \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = 0 \qquad \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}^T \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = 0 \quad \rightarrow \quad (A^T)\vec{e} = 0$$

$$A^T(A\vec{x} - \vec{b}) = 0$$

$$A^T A\vec{x} - A^T\vec{b} = 0$$

$$A^T A\vec{x} = A^T\vec{b}$$

*iff A^TA invertible

Least-squares solution "~~pseudo inverse~~" $\boxed{\hat{\vec{x}} = (A^TA)^{-1}A^T\vec{b}}$ *optimal sol'n in least squares sense

Consider $\hat{b} = A\vec{x} = A\left[(A^TA)^{-1}A^T\vec{b}\right]$

orthogonal proj. of $\vec{b}$: $\hat{b} = \boxed{A \cdot (A^TA)^{-1}A^T}\,\vec{b}$

Projection Matrix! (projects $\vec{b}$ onto subspace of A)

What are dimensions?

$$\underbrace{\underset{m\times n}{A}\ \underbrace{(\underset{n\times m}{A^T}\ \underset{m\times n}{A})^{-1}}\ \underset{n\times m}{A^T}}_{\substack{n\times m \\ m\times m\ \checkmark}}$$

What happens if cols of A are orthogonal?

OVERDET → tall, skinny A. $\hat{b}$ from proj. on cols + adding up proj. only if orthog.

$A = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \\ | & | & & | \end{bmatrix}$ orthogonal

orthog. → inner prod. zero

$\langle \vec{a}_1, \vec{a}_j \rangle \underset{j\neq 1}{=} 0$ or generally $\langle \vec{a}_i, \vec{a}_j \rangle = 0$ if $i \neq 0$

Let's compute $A^TA = \begin{bmatrix} - & \vec{a}_1^T & - \\ - & \vec{a}_2^T & - \\ & \vdots & \\ - & \vec{a}_n^T & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \vec{a}_1^T\vec{a}_1 & \cancel{\vec{a}_1^T\vec{a}_2} \to 0 & \cdots & \cancel{0} \to 0 \\ \cancel{\vec{a}_2^T\vec{a}_1} \to 0 & \vec{a}_2^T\vec{a}_2 & \cdots & 0 \\ \cancel{\vec{a}_3^T\vec{a}_1} \to 0 & \cancel{\vec{a}_3^T\vec{a}_2} \to 0 & \vec{a}_3^T\vec{a}_3 & \cdots 0 \\ \vdots & & & \ddots \\ 0 & & & \vec{a}_n^T\vec{a}_n \end{bmatrix}$

only diag. are non-zero if orthog. cols!

$(A^TA)^{-1} = \begin{bmatrix} 1/\|\vec{a}_1\|^2 & & & 0 \\ & 1/\|\vec{a}_2\|^2 & & \\ & & \ddots & \\ 0 & & & 1/\|\vec{a}_n\|^2 \end{bmatrix} \leftarrow A^TA = \begin{bmatrix} \|\vec{a}_1\|^2 & & & 0 \\ & \|\vec{a}_2\|^2 & & \\ & & \ddots & \\ 0 & & & \|\vec{a}_n\|^2 \end{bmatrix}$

diagonal matrix

↳ inverse is easy!!

Now we want to find $\hat{b} = A(A^TA)^{-1}A^T\vec{b}$

$$= A \begin{bmatrix} 1/\|\vec{a}_1\|^2 & & \\ & 1/\|\vec{a}_2\|^2 & \\ & & \ddots\ 1/\|\vec{a}_n\|^2 \end{bmatrix} \underbrace{\begin{bmatrix} - & \vec{a}_1^T & - \\ - & \vec{a}_2^T & - \\ & \vdots & \\ - & \vec{a}_n^T & - \end{bmatrix} \vec{b}}_{\text{mtx-vector mult in "row view"}}$$

$$= A \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \begin{bmatrix} \vec{a}_1^T\vec{b} \\ \vec{a}_2^T\vec{b} \\ \vdots \\ \vec{a}_n^T\vec{b} \end{bmatrix}$$

another mtx-vec mult.

$$= A \begin{bmatrix} \vec{a}_1^T\vec{b}/\|\vec{a}_1\|^2 \\ \vdots \\ \vec{a}_n^T\vec{b}/\|\vec{a}_n\|^2 \end{bmatrix}$$

looks a bit like proj....!

$$\hat{\vec{b}} = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \vec{a}_1^T\vec{b}/\|\vec{a}_1\|^2 \\ \vec{a}_2^T\vec{b}/\|\vec{a}_2\|^2 \\ \vdots \\ \vec{a}_n^T\vec{b}/\|\vec{a}_n\|^2 \end{bmatrix} = \underbrace{\frac{\vec{a}_1^T\vec{b}}{\|\vec{a}_1\|^2}}_{\text{scalar}} \vec{a}_1 + \frac{\vec{a}_2^T\vec{b}}{\|\vec{a}_2\|^2}\vec{a}_2 + \dots + \frac{\vec{a}_n^T\vec{b}}{\|\vec{a}_n\|^2}\vec{a}_n$$

$m \times 1$ $m \times 1$

project onto each col & add up results

*only works if cols orthogonal

OVERDETERMINED

$m > n$

$A\vec{x} = \vec{b}$: $m \times n$, $n \times 1$, $m \times 1$ → Least squares $\boxed{\hat{x} = (A^TA)^{-1}A^T\vec{b}}$

more eqns than unknowns

UNDERDETERMINED

$m < n$

less eqns than unknowns

$A\vec{x} = \vec{b}$: $m \times n$, $n \times 1$, $m \times 1$ → ? Infinite solutions - no unique sol'n

What to do?

*assume A has full row rank

Ex. $\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [1] \rightarrow x_1 + x_2 = 1$

can be many combos. → Pick one! $\boxed{\text{Which one?}}$ 'smallest'? maybe

A popular choice is minimum norm solution.

↳ What's that?

Constrained optimization: $\min_{\vec{x}} \|\vec{x}\|^2$ such that $A\vec{x} = \vec{b}$ → $(x_1 = \frac{1}{2}, x_2 = \frac{1}{2})$

Ex. to solve, use $\min_{\vec{x}, \lambda} \|\vec{x}\|^2 + \boxed{\vec{\lambda}^T(b - A\vec{x})}$ extra new term!

$\vec{\lambda}$ scalars: Lagrange multipliers

take $\frac{\partial}{\partial \vec{x}}\left(\vec{x}^T\vec{x} + \vec{\lambda}^T(\vec{b} - A\vec{x})\right) = 0$

$$2\vec{x}^T - \vec{\lambda}^T A = 0$$

Left-mult. by A: $A2\vec{x} - AA^T\lambda = 0A$

$$\vec{\lambda} = (AA^T)^{-1}\, 2A\vec{x}$$

Then, differentiate w.r.t $\vec{\lambda}$ and set $= 0$:

$$A\vec{x} = \vec{b}$$

$$\vec{\lambda} = (AA^T)^{-1}\, 2\vec{b}$$

Minimum norm solution to $Ax = b$: $\boxed{\vec{x} = A^T(AA^T)^{-1}\vec{b}}$