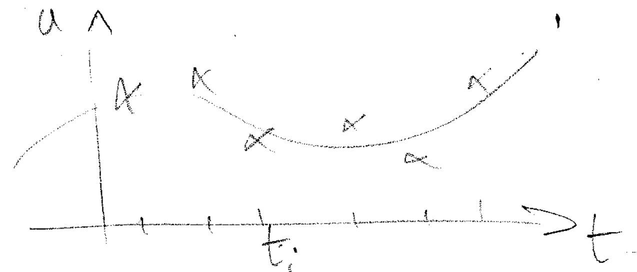
02/08/18 Least-squares regression.

$J(x) = \| \underbrace{Ax}_{\to \text{ on } \mathcal{R}(x)} - b \|_2^2$ with $\|b\|_2^2 = \sum b_i^2$

LSM. $\min_x J(x)$. $A \in \mathbb{R}^{m \times n}$.

ex.



$$u_i = \sum_{\alpha=0}^{n-1} c_\alpha t_i^\alpha \qquad i = 1, \dots, m$$

$$\begin{array}{c} \\ m \\ \end{array} \begin{bmatrix} 1 & t_1 & \cdots & t_1^d \\ \vdots & & & \\ 1 & t_m & \cdots & t_m^d \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_{n-1} \end{bmatrix} = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} \qquad Ax = b$$

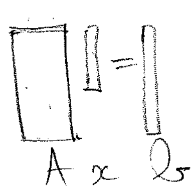
n

interpolation ($m = n$)
or approximation ($m > n$)

Case I. $m = n$, A full rank

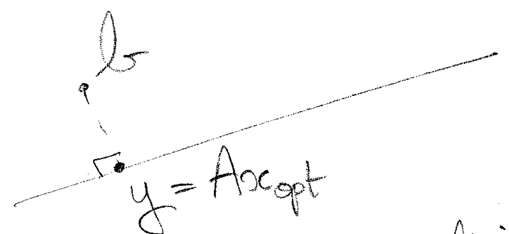
$x = A^{-1} b$. $\min J = 0$.

Case II. $m > n$ A full col-rank
$\Rightarrow$ overdetermined



$\to$ solution to $Ax = b$ only if $b \in \operatorname{Ran} A$

(Why full col rank? Having a nullspace $\Rightarrow$ minimizer cannot be unique)



$\operatorname{Ran} A = \{Ax : x \in \mathbb{R}^n\}.$

$$\min \|Ax - b\| \to y = P_{\operatorname{Ran} A}\, b$$

$$b - y \perp \operatorname{Ran} A$$

$\langle x, y\rangle = x^T y$

$$\langle b - y, Ax\rangle = 0 \qquad \forall x \in \mathbb{R}^n$$

$$\langle A^*(b - y), x\rangle = 0 \qquad \forall x \in \mathbb{R}^n$$

transpose conjugate if complex

$$A^*(b - y) = 0 \qquad y = Ax_{opt}$$

$$\Rightarrow \underbrace{A^*A}_{\text{invertible?}}\, x_{opt} = A^* b \qquad \text{normal equations}$$

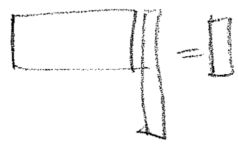
$$\left[\begin{array}{l} A^*Ax = 0 \;\Rightarrow\; x^*A^*Ax = 0 = \|Ax\|^2 \\ \qquad\qquad\;\Rightarrow\; Ax = 0 \\ \qquad\qquad\;\Rightarrow\; x = 0 \quad \text{because } A \text{ has full col-rank} \\ \text{so } \operatorname{Null}(A^*A) = \{0\}, \text{ yes, invertible.} \end{array}\right]$$

$$\boxed{x_{opt} = (A^*A)^{-1} A^* b} \qquad \text{(OLS)}$$

$$y = Ax_{opt} = \boxed{A(A^*A)^{-1}A^*}\, b$$

$$= \boxed{P_{\operatorname{Ran} A}}\, b$$

Case III $m < n$. A full row-rank

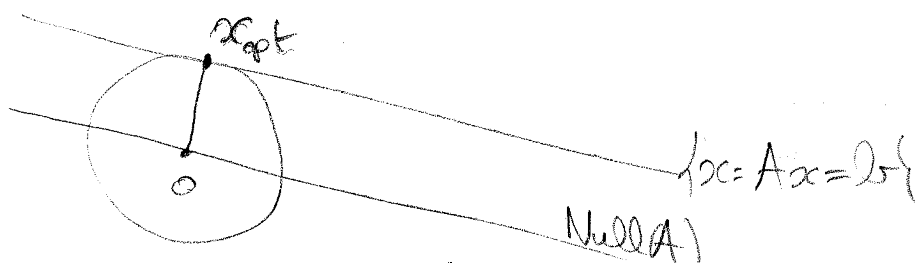


$\Rightarrow$ underdetermined

many solutions to $Ax=b$.
(restrict to any m cols).

(Why full row rank? May not have a solution to $Ax=b$ otherwise, $\operatorname{Ran} A \not\subseteq \mathbb{R}^m$).

ULS: $\min_x \|x\|_2$ s.t. $Ax=b$.



$x_{opt} \perp \{x : Ax=b\}$.

HW

Vectors normal to $Ax=b$:
$\operatorname{Span}(\text{rows}(A)) = \operatorname{Ran}(A^*)$

$x_{opt} \in \operatorname{Ran} A^*$
$x_{opt} = A^* v$ for some $\underline{v}$.

and $A x_{opt} = b \Rightarrow A A^* v = b$.
$v = (AA^*)^{-1} b$ full row rank (same as before)

normal eq

$\Rightarrow \boxed{x_{opt} = A^*(AA^*)^{-1} b}$ (ULS,
A)

orthon. cols

diag, pos.

(thin) SVD: $A = U \Sigma V^*$ — orthon rows.

Case I
tall & thin

$A = U \Sigma V^*$

unitary
$V^*V = VV^* = I$

Isometry:
$U^*U = I$
$UU^* = P_{\mathrm{Ran}A} = P_{\mathrm{Ran}U}$

$$A^*A\, x_{opt} = A^* b.$$

$$\cancel{V}\cancel{\Sigma}\cancel{U^*U}\Sigma V^* x_{opt} = \cancel{V}\cancel{\Sigma} U^* b$$

$$\boxed{x_{opt} = V\Sigma^{-1}U^* b}$$

$$= A^+ b = (A^*A)^{-1}A^*b$$

pseudoinverse of A.

$A^+A = I$
$AA^+ = P_{\mathrm{Ran}A}$

Case II
short & wide

$A = U\Sigma V^*$

unitary
$U^*U = UU^* = I$

Isometry
$V^*V = I$
$VV^* = P_{\mathrm{Ran}A^*} = P_{\mathrm{Ran}V}$

$$\begin{cases} x_{opt} = A^* v \\ AA^* v = b \end{cases}$$

$$U\Sigma \cancel{V^*V}\Sigma U^* v = b$$

$$v = U\Sigma^{-2}U^* b$$

$$x_{opt} = V\Sigma \cancel{U^*U}\Sigma^{-2}U^* b$$

$$\boxed{x_{opt} = V\Sigma^{-1}U^* b}$$

$$= A^+ b = A^*(AA^*)^{-1} b$$

$AA^+ = I$
$A^+A = P_{\mathrm{Ran}A^*}$

Rmk: formula still holds in degenerate (not full rank) cases → adequate thin SVD.

("Stability" often used in the context of algorithms instead).

Conditioning. $y = Ax_0 + e$
$\Rightarrow x = A^+ y$.

ex polyn. interp. is ill conditioned in general

perturbation e in data space
$\Rightarrow$ how large of a perturbation $x - x_0$ in model space? (assume overdetermined, otherwise x could be far from x_0)

$$\frac{\|x-x_0\|}{\|x_0\|} \Big/ \frac{\|e\|}{\|Ax_0\|}$$

$$x = A^+y = A^+(Ax_0+e) = x_0 + A^+e$$

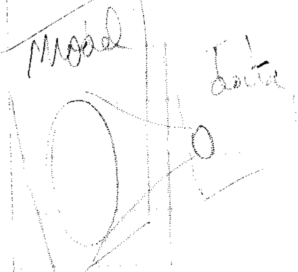
$$\|x-x_0\| = \|A^+e\| \le \|A^+\|\,\|e\|$$

$$\|A\| = \sup \frac{\|Ax\|}{\|x\|}$$

$$(\cdot) \le \|A^+\| \frac{\|Ax_0\|}{\|x_0\|}$$

$$\le \boxed{\|A^+\|\,\|A\| = K(A)}$$

Condition number of A.



The inverse problem is well-posed when $\frac{\|\Delta \text{model}\|}{\|\text{model}\|} \Big/ \frac{\|\Delta \text{data}\|}{\|\text{data}\|}$ is small, otherwise it is ill-posed. (poor conditioning)

Thm

Rmk. $\|A\| = \sigma_{max}(A)$, $\|A^+\| = \sigma_{min}(A)$

$\sigma_1(A)$ $\qquad$ $\sigma_{min(m,n)}(A)$

so $K(A) \ge 1$.