RECALL: $\boxed{\min_{\theta \in \mathbb{R}^d} f(\theta)}$

$\nabla f(\theta_*)$ & f CONVEX $\Rightarrow$ θ_* IS GLOB. OPT.

$\nabla^2 f(\theta_*) \succ 0$ (SOSC) $\Rightarrow$ STABILITY UNDER PERTURBATIONS, WELL-POSED PROBLEM

PART II: GENERAL ESTIMATION METHODS

CH 4: LINEAR LEAST SQUARES (LLS)

4.1

DATA: $y(1), y(2), \ldots, y(N) \in \mathbb{R}$ (NOISY) — DEPENDENT VARS

$\phi(1), \phi(2), \ldots, \phi(N) \in \mathbb{R}^d$ (NOT NOISY) — EXPLANATORY VARS, INDEPENDENT VARS. (REGRESSION VECTORS)

QUESTION: GIVEN A NEW ϕ, WHAT COULD BE y?

MODEL: $\boxed{y(k) = \phi(k)^T \cdot \theta + \varepsilon(k)}$

IDEA (LLS): FIND θ THAT MINIMIZES "PREDICTION" ERRORS $(y(k) - \phi(k)^T\theta)$

$$\boxed{\min_{\theta \in \mathbb{R}^d} \sum_{k=1}^{N} (y(k) - \phi(k)^T\theta)^2}$$

$$Y_N := \begin{bmatrix} y(1) \\ \vdots \\ \vdots \\ y(N) \end{bmatrix} \in \mathbb{R}^N \qquad \Phi_N := \begin{bmatrix} \phi(1)^T \\ \vdots \\ \vdots \\ \phi(N)^T \end{bmatrix} \in \mathbb{R}^{N \times d}$$

$$\sum_{i=1}^{N} \left(y(i) - \phi(i)^T \theta\right)^2 = \| Y_N - \Phi_N \theta \|_2^2$$

WHERE

$$\|x\|_2^2 := \sum_{k=1}^{n} x_k^2$$

SQUARE OF "EUCLIDEAN NORM" $\|x\|_2 = \sqrt{\sum x_i^2}$

$$\begin{bmatrix} y(1) - \phi(1)^T \theta \\ \vdots \\ y(N) - \phi(N)^T \theta \end{bmatrix} = Y_N - \Phi_N \cdot \theta$$

$$\boxed{\min_{\theta \in \mathbb{R}^d} f(\theta)} \qquad f(\theta) = \| Y_N - \Phi_N \theta \|_2^2$$

$S^* = \arg\min_{\theta} f(\theta)$, S^* : SET OF MINIMIZERS. IF WE ARE LUCKY, $S^* = \{\theta^*\}$, AND WE WRITE

$$\hat{\theta}_{LS} := \theta^* = \arg\min_{\theta} f(\theta)$$

4.2 SOLUTION OF LLS

$f(\theta) = \|y - \Phi\theta\|_2^2$ IS CONVEX (OMIT "N" FOR NOT. SIMPL.)

$= (y-\Phi\theta)^T(y-\Phi\theta) = \theta^T(\Phi^T\Phi)\theta + y^Ty - 2\,\theta^T\Phi^Ty$

NOTE: $y^T\Phi\theta = \theta^T\Phi^Ty$

$\nabla f(\theta) = 2\cdot\Phi^T\Phi\cdot\theta - 2\,\Phi^Ty \overset{!}{=} 0 \iff \Phi^T\Phi\,\theta = \Phi^Ty$

$S^* = \{\theta\in\mathbb{R}^d \mid \Phi^T\Phi\theta = \Phi^Ty\}$ IF LUCKY, $\Phi^T\Phi$ IS INVERTIBLE, $S^* = \{\theta^*\}$ WITH

$$\boxed{\theta^* = \underline{(\Phi^T\Phi)^{-1}\Phi^T}\cdot y}$$

DEF: PSEUDOINVERSE OF $\Phi\in\mathbb{R}^{N\times d}$ (IF $\Phi^T\Phi$ INV.):

$$\boxed{\Phi^+ := (\Phi^T\Phi)^{-1}\cdot\Phi^T}$$

"pinv" IN MATLAB

$$\boxed{\theta^* = \Phi^+\cdot y}$$

WHAT DOES $\Phi^T\Phi$ INV. MEAN? (LUCKY CASE) $\iff$ WELL-POSED

$\iff$ STABLE AGAINST PERTURBATIONS

CLEARLY, $\Phi^T\Phi \succeq 0$ FOR ANY Φ.

$\Phi^T\Phi \succ 0 \iff$ LUCKY CASE $\iff z^T\Phi^T\Phi\cdot z > 0$ FOR $\forall z\neq 0$

$\iff \Phi z \neq 0$ FOR ANY $z\neq 0$

$\iff$ COLUMNS OF Φ ARE LINEARLY INDEP.

EX: 1 MODEL: $y(k) = \theta + \varepsilon(k) = \underbrace{1}_{\phi(k)} \cdot \theta + \varepsilon(k)$

$$Y_N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}, \quad \Phi_N = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \qquad \Phi_N^T \Phi_N = N$$

$$\theta^* = \underline{(\Phi_N^T \Phi_N)^{-1}} \Phi_N^T \cdot Y_N = \frac{1}{N} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} = \frac{1}{N} \sum_{k=1}^{N} y(k)$$

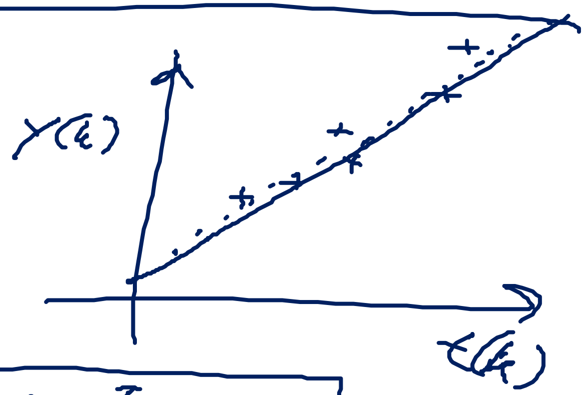
AVERAGE!

EX: (FITTING A LINE) MODEL

$$\Phi_N = \begin{bmatrix} 1 & t(1) \\ \vdots & \vdots \\ 1 & t(N) \end{bmatrix}$$

$$y(k) = 1 \cdot \theta_1 + t(k) \cdot \theta_2 + \varepsilon(k) \qquad \phi(k) = \begin{bmatrix} 1 \\ t(k) \end{bmatrix}$$

$$= \phi(k)^T \theta + \varepsilon(k)$$



$$\Phi_N^T \Phi_N = \begin{bmatrix} N & \sum t(k) \\ \sum t(k) & \sum t(k)^2 \end{bmatrix} \implies \theta^* = \Phi_N^+ Y_N \longrightarrow \text{MODEL } \boxed{y = \begin{bmatrix} 1 \\ t \end{bmatrix}^T \cdot \theta^*}$$

LINE IN (t, y) SPACE

ALSO FITTING HIGHER ORDER POLYNOMIALS IS LLS, E.G. $\boxed{y(k) = 1 \cdot \theta_1 + t(k)\,\theta_2 + t(k)^2 \cdot \theta_3 + \varepsilon(k)}$

NEED: <u>LINEAR IN PARAMETERS (LIP)</u>

4.3 WEIGHTED LS

$$f(\theta) = \sum_{k=1}^{N} \frac{\left(y(k) - \phi(k)^T \cdot \theta\right)^2}{\sigma_\varepsilon^2(k)}$$

"DIAGONAL WEIGHTING"
WITH NOISE VARIANCES $\sigma_\varepsilon^2(k) > 0$
(OR ESTIMATES OF THEM)
WITH $W = \begin{bmatrix} \frac{1}{\sigma_\varepsilon^2(1)} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_\varepsilon^2(N)} \end{bmatrix}$
(SPECIAL CASE)

GENERAL WEIGHTED LS WITH ANY $W \succ 0$

$$f(\theta) = (y_N - \Phi_N \cdot \theta)^T \, W \cdot (y_N - \Phi_N \cdot \theta)$$

$$= \|y_N - \Phi_N \theta\|_W^2$$

WITH $\|x\|_W^2 := x^T W x$

$\|x\|_W := \sqrt{x^T W x}$

$$= y_N^T W y_N - 2\theta^T \Phi_N^T W \cdot y_N + \theta^T \Phi_N^T W \Phi_N \theta$$

$$\nabla f(\theta) = -2\Phi_N^T W y + 2\Phi_N^T W \Phi_N \theta \iff \theta^* = (\Phi_N^T W \Phi_N)^{-1} \Phi_N^T W y$$

ONE WAY TO COMPUTE SOLUTION

$$f(\theta) = \underbrace{(y_N - \Phi_N \theta)^T W^{\frac{1}{2}T}} \; \underbrace{W^{\frac{1}{2}} (y_N - \Phi_N \theta)} = \| \underline{\underline{W^{\frac{1}{2}} y_N}} - \underline{\underline{W^{\frac{1}{2}} \Phi_N}} \theta \|_2^2$$

USE $W = W^{\frac{1}{2}} \cdot W^{\frac{1}{2}}$
SYMMETRIC MATRIX SQUARE ROOT
(EXISTS & IS UNIQUE FOR $W \succ 0$)

$$\theta^* = (W^{\frac{1}{2}} \Phi_N)^+ W^{\frac{1}{2}} y_N$$

USING "PINV"
(SECOND WAY)

SECOND WAY : (FOR DIAGONAL WEIGHTING)

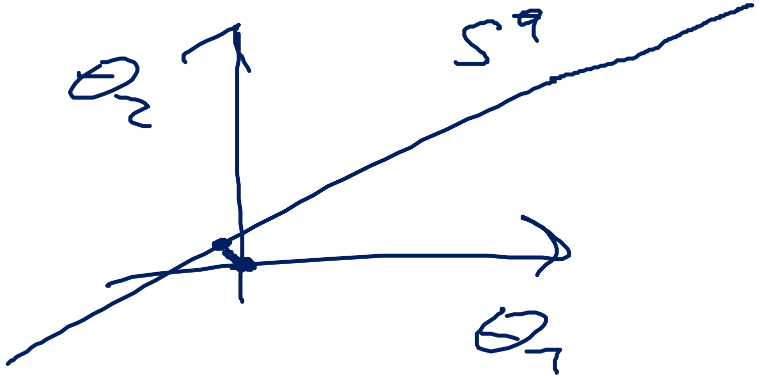
$$y(k) \longrightarrow \tilde{y}(k) = \frac{y(k)}{G_\varepsilon(k)} \qquad \tilde{Y}_N = W^{\frac{1}{2}} Y_N$$

$$\phi(k) \longrightarrow \tilde{\phi}(k) = \frac{\phi(k)}{G_\varepsilon(k)} = \frac{1}{G_\varepsilon(k)} \cdot \phi(k) \qquad \tilde{\Phi}_N = W^{\frac{1}{2}} \Phi_N$$

4.4 ILL-POSED LEAST SQUARES AND THE MOORE-PENROSE PSEUDO INVERSE (MPPI)

IF $\Phi_N^T \Phi$ NOT INVERTIBLE

$S^* = \{\theta \mid \Phi_N^T \Phi_N \cdot \theta = \Phi_N^T Y\}$ IS <u>NOT</u> A POINT, BUT AN AFFINE SET



WHAT TO DO?

1) TAKE LESS θ (REDUCE d) $\longrightarrow$ BEST WAY

2) CHOOSE (AUTOMATICALLY) ONE $\theta \in S^*$, E.G. MINIMAL NORM SOLUTION

$$\boxed{\min_{\theta} \frac{1}{2}\|\theta\|_2^2 \quad \text{s.t.} \quad \theta \in S^*}$$

SOLUTION IS GIVEN BY "$\mathrm{pinv}(\Phi_N) \cdot Y_N$" WITH MPPI

DEF: MPPI GIVEN ~~$\Phi \in \mathbb{R}^{m \times n}$~~ $\Phi \in \mathbb{R}^{m \times n}$ WITH SVD

$\Phi = U S V^T$. THEN MPPI Φ^+ IS DEFINED BY

$$\boxed{\Phi^+ := V S^+ U^T}$$

WITH

$$S^+ := \left[\begin{array}{ccc|c} \sigma_1^{-1} & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_r^{-1} \; 0 \cdots 0 & \\ & & & 0 \end{array}\right]$$

NOTE: IN "LUCKY CASE", $\Phi^+ = (\Phi^T \Phi)^{-1} \Phi^T$

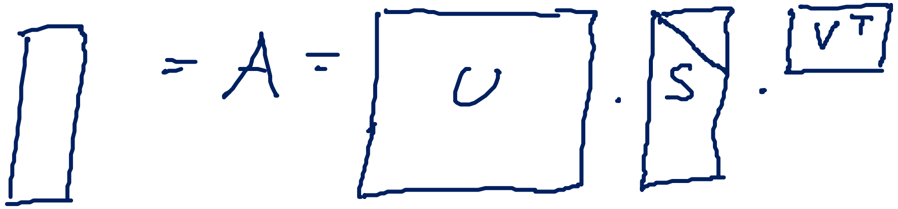
DEF: Singular Value Decomposition (SVD) FOR ANY $A \in \mathbb{R}^{m \times n}$

EXIST $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$, ORTHOGONAL $U^T U = I$, $V^T V = I$

AND S DIAGONAL, WITH

REAL, NON~~ZERO~~ NEGATIVE ELEMENTS, $S \in \mathbb{R}^{m \times n}$ SUCH THAT $\boxed{A = U \cdot S \cdot V^T}$

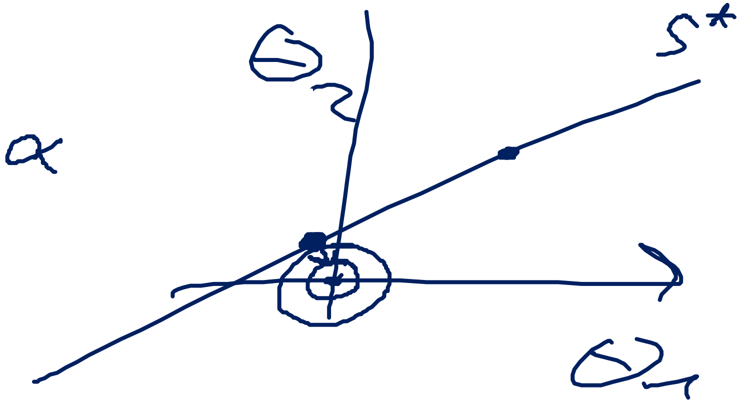
$$S = \left[\begin{array}{ccc} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \; 0 \cdots 0 \\ \hline & 0 & \end{array}\right]$$



$\sigma_1, .., \sigma_r$: SINGULAR VALUES $r \le \min(m, n)$ IS RANK OF A

$\sigma_1 \ge \sigma_2 .. \ge \sigma_r$

4.4.1 REGULARIZATION



$$\theta^*(\alpha) = \arg\min_{\theta} \frac{1}{2}\|y - \Phi\theta\|_2^2 + \frac{1}{2}\|\theta\|_2^2 \cdot \alpha$$

UNIQUE (LUCKY CASE) FOR ANY $\alpha > 0$

$$\Phi^T\Phi\,\theta^* - \Phi^T y + \alpha\cdot\theta^* = 0$$

$$\Leftrightarrow (\Phi^T\Phi + \alpha I)\,\theta^* = \Phi^T y \Leftrightarrow \theta^*(\alpha) = \underbrace{(\Phi^T\Phi + \alpha I)^{-1}\Phi^T}\cdot y$$

$$\boxed{\lim_{\alpha \to 0} (\Phi^T\Phi + \alpha I)^{-1}\Phi^T = \Phi^+}$$

TAKE HOME MESSAGE: $\boxed{\mathrm{pinv}(A) = A^+}$ ALWAYS GIVES AN ANSWER, MOORE-PENROSE PSEUDO INVERSE

IF $A^TA \succ 0$: $A^+ = (A^TA)^{-1}A^T$

IF A INV : $A^+ = A^{-1}$