0 MODEL:

$$H(k) = \theta_1 + T_{out}(k)\cdot\underline{\theta_2} + L_{out}(k)\cdot\theta_3 + T_{in}(k)\,\underline{\theta_4} + D_{sleep}(k)\cdot\theta_5 + \left(T_{in}(k)+T_{out}(k)\right)\underline{\theta_6} + C_{drink}(k)\cdot\theta_7 + T_{in}(k)^2\cdot\theta_8 + \varepsilon(k) = \phi(k)^T\theta + \varepsilon(k)$$

THIS WILL LEAD TO ILL-POSED ESTIMATION PROBLEM. WHY?

$$\Phi_N = \begin{bmatrix} \phi(1)^T \\ \vdots \\ \\ \end{bmatrix} = \begin{bmatrix} 1 & T_{out}(1) & L_{out}(1) & T_{in}(1) & D_{sleep}(1) & (T_{in}(1)+T_{out}(1)) & C_{drink}(1) & T_{in}(1)^2 \\ 1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & & & & & & & \end{bmatrix}$$

$\Phi_N^T\Phi_N > 0$

$\theta_2, \theta_4, \theta_6$ CANNOT BE ESTIMATED SIMULTANEOUSLY!

① DROP ONE OF THEM (BEST)

② USE PSEUDO-INVERSE

$$Y_N = \begin{bmatrix} H(1) \\ \vdots \end{bmatrix}$$

$$\boxed{\theta = \Phi_N^{+}\cdot Y_N}$$

4.5 STATISTICAL ANALYSIS OF WEIGHTED LS

ASSUMPTIONS: 1) $\boxed{y(k) = \phi(k)^T \theta + \varepsilon(k)}$ WITH θ_0 TRUE BUT UNKNOWN θ

2) $\phi(1), \phi(2), \ldots$ ARE DETERMINISTIC

3) $\varepsilon(1), \ldots$ HAVE ZERO MEAN

[3b) OFTEN, $\varepsilon(1), \varepsilon(2), \ldots$ INDEPENDENT, OR EVEN
[3c) IND. IDENT. DISTRIB (i.i.d.)]

$$\boxed{\hat{\theta}_{WLS} = \left(\Phi_N^T W \Phi_N\right)^{-1} \Phi_N^T W y_N}$$ WITH $W \succeq 0$

QUESTIONS: a) BIASED OR NOT ?

b) PERFORMANCE DEPENDING ON W ? (COVARIANCE OF $\hat{\theta}_{WLS}$)

c) HOW GOOD IS OUR FIT ?

4.5.1 EXPECTATION OF WLS

$$\boxed{y_N = \Phi_N \cdot \theta_0 + \varepsilon_N}$$

REMARK: $\boxed{\text{EXPECTATION}}$ = MEAN (OF RANDOM)

$\boxed{\text{AVERAGE}}$ = SAMPLE MEAN (OF A SEQUENCE)

$$\mathbb{E}\{\hat{\theta}_{WLS}\} = \mathbb{E}\{(\Phi_N^T W \Phi_N)^{-1}\Phi_N^T W \; y_N\} = \underbrace{(\Phi_N^T W \Phi_N)^{-1}\Phi_N^T W}_{=:A} \; \mathbb{E}\{y_N\}$$

$$= A \cdot \mathbb{E}\{\underbrace{\Phi_N \cdot \theta_0}_{\text{DETERMINISTIC}} + \varepsilon_N\} = A \, \Phi_N \theta_0 + A \cdot \underbrace{\mathbb{E}\{\varepsilon_N\}}_{=0}$$

$$= (\Phi_N^T W \Phi_N)^{-1}(\Phi_N^T W \Phi_N)\theta_0 = \theta_0 \qquad \underline{\underline{\text{UNBIASED}}} \; !$$

4.5.2 COVARIANCE OF $\hat{\theta}_{WLS}$

$$\mathrm{cov}(\hat{\theta}_{WLS}) = \mathbb{E}\{(\hat{\theta}_{WLS} - \theta_0)(\hat{\theta}_{WLS} - \theta_0)^T\}$$

$$= \mathbb{E}\{A \cdot \varepsilon_N \cdot \varepsilon_N^T \cdot A^T\} = A \cdot \underbrace{\mathbb{E}\{\varepsilon_N \cdot \varepsilon_N^T\}}_{=\mathrm{cov}(\varepsilon_N) = \Sigma_{\varepsilon_N}} \cdot A^T$$

$$= A \cdot \Sigma_{\varepsilon_N} \cdot A^T$$

$$\rightarrow \; = (\Phi_N^T W \Phi_N)^{-1} \underline{\Phi_N^T W \; \Sigma_{\varepsilon_N} \; W \Phi_N} \; \underline{(\Phi_N^T W \Phi_N)}^{-1} \succeq (\Phi_N^T \Sigma_{\varepsilon_N}^{-1} \Phi_N)^{-1}$$

OBSERVATIONS: RESCALING W TO $\alpha \cdot W$ WILL NOT CHANGE $\mathrm{cov}(\hat{\theta}_{WLS})$

$$\hat{\theta}_{WLS} - \theta_0 = A \cdot \varepsilon_N$$

$$= (\Phi_N^T W \Phi_N)^{-1} \Phi_N^T W \cdot \varepsilon_N$$

OPTIMAL CHOICE: $W = \Sigma_{\varepsilon_N}^{-1}$

$$\boxed{\mathrm{cov}(\hat{\theta}_{WLS}) = (\Phi_N^T \Sigma_{\varepsilon_N}^{-1} \Phi_N)^{-1}}$$

FOR i.i.d. NOISE, $W = I$

UNWEIGHTED LS IS OPTIMAL

EX:

$$\varepsilon(1) \sim \mathcal{N}(0, \sigma_1^2)$$

$$\varepsilon(2) \sim \mathcal{N}(0, \sigma_2^2)$$

independent

$$N = 2$$

$$\Sigma_{\varepsilon N} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

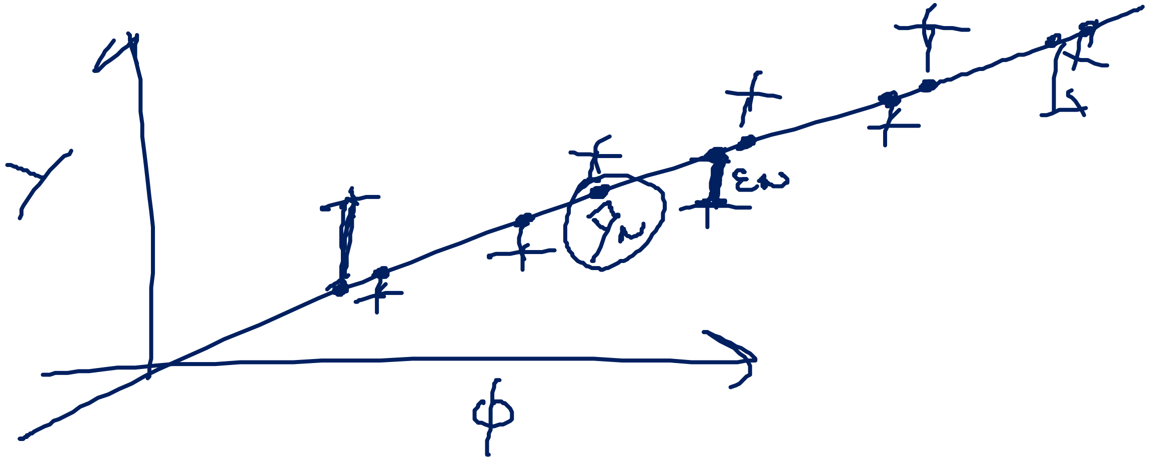
$$W^{OPT} = \Sigma_{\varepsilon N}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix}$$

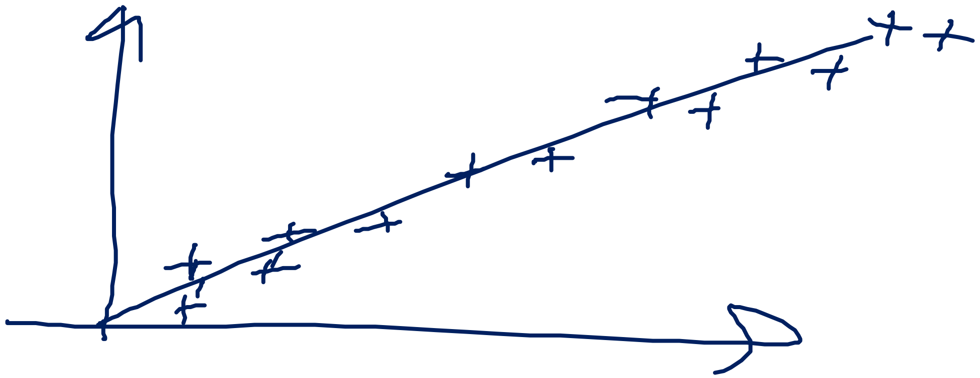
$$(Y_N - \Phi_N \theta)^T W (Y_N - \Phi_N \theta)$$

$$= \sum_{k=1}^{2} (y(k) - \varphi(k)^T \theta) \frac{1}{\sigma_k^2} (y(k) - \varphi(k)^T \theta)$$

$$\Sigma_{\varepsilon N} = \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

$$\Sigma_{\varepsilon N}^{-1} = \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$





4.6 MEASURE "GOODNESS OF FIT" USING R-SQUARED

INTUITIVELY: $\| Y_N - \underbrace{\Phi_N \hat{\theta}}_{\hat{Y}_N} \|_2^2$ SMALL $\longrightarrow$ GOOD

$\epsilon_N = Y_N - \Phi_N \hat{\theta}$

$Y_N - \hat{Y}$

$$\boxed{R^2 = 1 - \frac{\| Y_N - \Phi_N \hat{\theta} \|_2^2}{\| Y_N \|_2^2}}$$

BEST CASE $R^2 = 1$

WORST CASE $R^2 = 0$

INTERPRETATION USING $\hat{\theta}$ IS OPTIMAL

$\hat{\theta}$ SATISFIES $\Phi_N^T \Phi_N \hat{\theta} = \Phi_N^T Y_N$

$\Longleftrightarrow \Phi_N^T (\Phi_N \hat{\theta} - Y_N) = 0$

$\Longleftrightarrow \Phi_N^T \epsilon_N = 0 \Longrightarrow \underbrace{\hat{\theta}^T \Phi_N^T}_{\hat{Y}_N} \epsilon_N = 0 \Longleftrightarrow \hat{Y}_N^T \cdot \epsilon_N = 0$

$Y_N = \hat{Y}_N + \epsilon_N$

$\boxed{\hat{Y}_N \perp \epsilon_N}$

ORTHOGONAL

$\| Y_N \|_2^2 = \| \hat{Y}_N \|_2^2 + \| \epsilon_N \|_2^2$

$$R^2 = 1 - \frac{\|\varepsilon_N\|_2^2}{\|Y_N\|_2^2}$$

$$= \frac{\|Y_N\|_2^2 - \|\varepsilon_N\|_2^2}{\|Y_N\|_2^2} = \frac{\|\hat{Y}_N\|_2^2}{\|Y_N\|_2^2}$$

$$\boxed{\|\hat{Y}_N\|_2^2 + \|\varepsilon_N\|_2^2 = \|Y_N\|_2^2}$$

"HOW MUCH OF THE DATA CAN BE EXPLAINED BY THE MODEL"

IN PRACTICE, FIRST ~~WE~~ SUBTRACT AVERAGE FROM ALL MEASUREMENTS, I.E.

$$\boxed{\sum_{k=1}^{N} y(k) = 0}$$

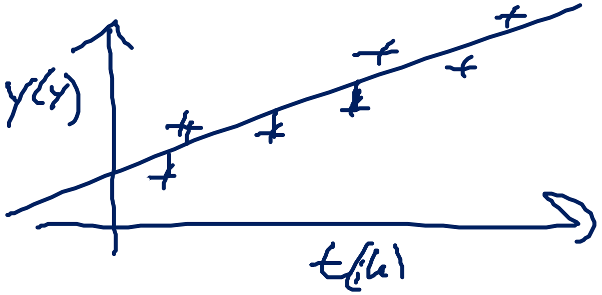
4.7 ESTIMATING COVARIANCE WITH SINGLE EXPERIMENT (FOR i.i.d. NOISE)

$$\mathrm{cov}(\hat{\theta}_{LS}) = (\Phi_N^T \Sigma_{\varepsilon_N}^{-1} \Phi_N)^{-1} = \sigma_\varepsilon^2 (\Phi_N^T \Phi_N)^{-1} \qquad \Sigma_{\varepsilon_N} = I \cdot \sigma_\varepsilon^2$$

PROBLEM: DO NOT KNOW σ_ε^2. IDEA: ESTIMATE σ_ε^2 FROM THE FIT

$$\boxed{\hat{\Sigma}_{\hat{\theta}_{LS}} := \hat{\sigma}_\varepsilon^2 \cdot (\Phi_N^T \Phi_N)^{-1}} \quad \boxed{\hat{\sigma}_\varepsilon^2 = \frac{1}{N-d} \sum_{k=1}^{N} (y(k) - \varphi(k)^T \hat{\theta}_{LS})^2} \qquad \theta \in \mathbb{R}^d$$

$$\boxed{\hat{\Sigma}_{\hat{\theta}_{LS}} = \frac{\|y_N - \Phi_N \hat{\theta}_{LS}\|_2^2}{N-d}\,\underline{\underline{(\Phi_N^T \Phi_N)^{-1}}}}$$



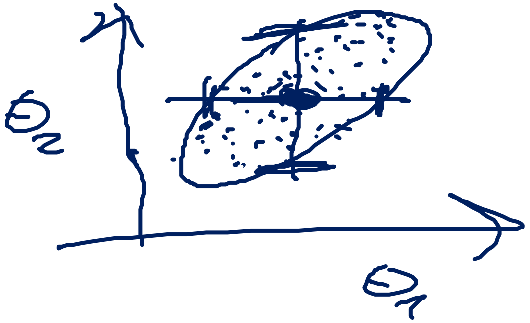
$$\phi(k)^T = \begin{bmatrix} 1 & t(k) \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

$$y(k) = \phi(k)^T \theta + \varepsilon(k)$$
$$= \theta_1 + t(k) \cdot \theta_2 + \varepsilon(k)$$

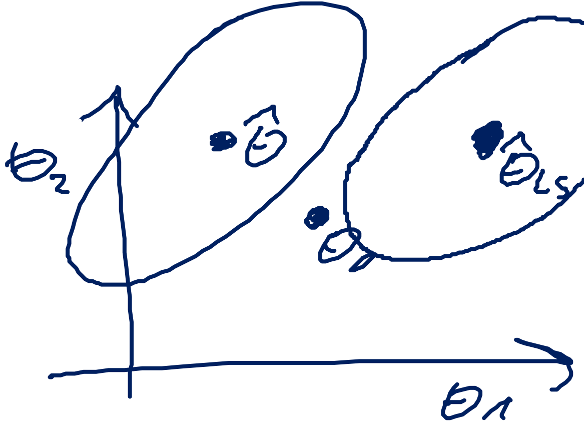
$$\Phi_N = \begin{bmatrix} 1 & t(1) \\ \vdots & t(2) \\ \vdots & \vdots \end{bmatrix}$$

$$\Phi_N^T \Phi_N = \begin{bmatrix} N & \sum t(k) \\ \sum t(k) & \sum t(k)^2 \end{bmatrix}$$

$$(\Phi_N^T \Phi_N)^{-1} = \begin{bmatrix} x & x \\ x & x \end{bmatrix}$$



$$\boxed{\hat{\Sigma}_{\hat{\theta}_{LS}} = \sigma_\varepsilon^2 \cdot \begin{bmatrix} x & x \\ x & x \end{bmatrix}}$$



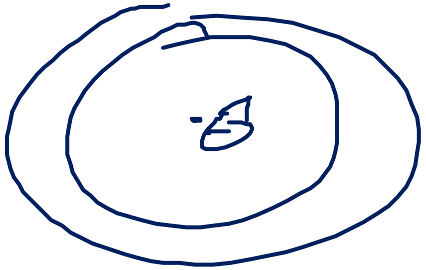
$$\boxed{P(\theta_0 \in \mathcal{E}_X(\hat\theta_{LS}))} = F(x, d)$$

$$\mathcal{E}_X(\hat\theta_{LS}) = \{\theta \in \mathbb{R}^d \mid \|\theta - \hat\theta_{LS}\|^2_{\Sigma_\theta^{-1}} \le \cancel{1}x\}$$

"CONFIDENCE ELLIPSOID"

(χ^2-DISTRIBUTION)

$\|\theta - \hat\theta_{LS}\|_2^2 \le 1$



$\|\theta - \hat\theta\|^2_{\Sigma_\theta^{-1}} \le 1$