CH 9 THE CRAMER RAO INEQUALITY

RECALL WEIGHTED LEAST SQUARES

$$\boxed{y(k) = \theta^T \varphi(k) + \varepsilon(k)} \qquad k = 1, \dots, N$$

$$\boxed{\mathbb{E}\{\varepsilon \varepsilon^T\} = \Sigma}$$

$$\hat{\theta}_{ML} = \arg\min_{\theta} \frac{1}{2}(y - \Phi\theta)^T \Sigma^{-1} (y - \Phi\theta) \qquad \Phi = \begin{pmatrix} \varphi(1)^T \\ \vdots \\ \varphi(N)^T \end{pmatrix}$$

$$= \underbrace{J^+ \Sigma^{-\frac{1}{2}}}\, y \qquad J = \Sigma^{-\frac{1}{2}} \Phi$$

$$J^+ = (J^T J)^{-1} J^T = \underline{(\Phi^T \Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} \Phi)^{-1} \Phi^T \Sigma^{-\frac{1}{2}}}$$

$\mathbb{E}\{\hat{\theta}_{ML}\} = \theta_0 \Leftrightarrow$ UNBIASED

$$\mathbb{E}\{(\hat{\theta}_{ML} - \theta_0)(\hat{\theta}_{ML} - \theta_0)^T\} = \operatorname{cov} \hat{\theta} = J^+ \underbrace{\Sigma^{-\frac{1}{2}} \, \Sigma \, \Sigma^{-\frac{1}{2}}} J^{+T} = J^+ J^{+T}$$

$$\operatorname{Cov}\hat{\theta} = (\Phi^T \Sigma^{-1} \Phi)^{-1} \cancel{(\Phi^T \underbrace{\Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}}}_{\Sigma^{-1}} \Phi)} \cancel{(\Phi^T \Sigma^{-1} \Phi)}^{-1}$$

$$= \underline{\underline{(\Phi^T \Sigma^{-1} \Phi)^{-1}}}$$

$$\underbrace{\frac{1}{2} (y - \Phi\theta)^T \Sigma^{-1} (y - \Phi\theta)}_{=: g(\theta)} = \frac{1}{2} \| \Sigma^{-\frac{1}{2}} (y - \Phi\theta) \|_2^2$$

$$= \frac{1}{2} \| \Sigma^{-\frac{1}{2}} y - \Sigma^{-\frac{1}{2}} \Phi \theta \|_2^2$$

$$= \frac{1}{2} \| \underline{\underline{\Sigma^{-\frac{1}{2}} y}} - J\theta \|_2^2$$

$$\arg\min g(\theta) = J^+ \Sigma^{-\frac{1}{2}} y$$

$$\frac{\partial^2 g}{\partial \theta^2} = \Phi^T \Sigma^{-1} \Phi$$

JUST A COINCIDENCE?

QUESTION: GIVEN

a) ANY ESTIMATOR $\hat{\theta}(y)$
(THAT IS UNBIASED)

b) A PDF $f(\theta, y)$
(AND TRUE PARAMETER θ_0)

WHAT IS THE COVARIANCE OF $\hat{\theta}$?

PARTIAL ANSWER (CRAMER-RAO-LOWER BOUND)

$$\boxed{\mathrm{Cov}\,\hat{\theta} \succeq M^{-1}} \iff \mathrm{Cov}\,\hat{\theta} - M^{-1} \succeq 0$$

WHERE M IS THE "FISHER INFORMATION MATRIX"

$$\boxed{M = \mathbb{E}\left\{ \frac{d^2}{d\theta^2} \left(-\log f(\theta, y) \right) \right\} \Big|_{\theta = \theta_0}}$$

[CF. LJUNG A7]

"NO ESTIMATOR CAN HAVE LOWER COVARIANCE THAN M^{-1}"

EXAMPLE: GAUSSIAN NOISE, LINEAR MODEL, independent

$$\Sigma = \begin{bmatrix} \sigma(1)^2 & & 0 \\ & \ddots & \\ 0 & & \sigma(N)^2 \end{bmatrix} \qquad y = \begin{pmatrix} y(1) \\ \vdots \\ y(N) \end{pmatrix}$$

$$f(\theta, y) = \prod_{k=1}^{N} \text{CONST.} \exp\left(-\frac{1}{2}\left(\frac{y(k)-\theta^T\varphi(k)}{\sigma(k)}\right)^2\right)$$

$$-\log f(\theta, y) = \frac{1}{2}\sum_{k=1}^{N}\left(\frac{y(k)-\theta^T\varphi(k)}{\sigma(k)}\right)^2 + \text{CONST}$$

$$= \frac{1}{2}(y-\Phi\theta)^T\Sigma^{-1}(y-\Phi\theta) + \text{CONST}$$

$$= \frac{1}{2}\theta^T \underbrace{\Phi^T\Sigma^{-1}\Phi}\,\theta + \text{LINEAR} + \text{CONSTANT}$$

$$\frac{\partial^2(-\log f(\theta, y))}{\partial\theta^2} = \Phi^T\Sigma^{-1}\Phi = M$$

"FISHER INFORMATION MATRIX"

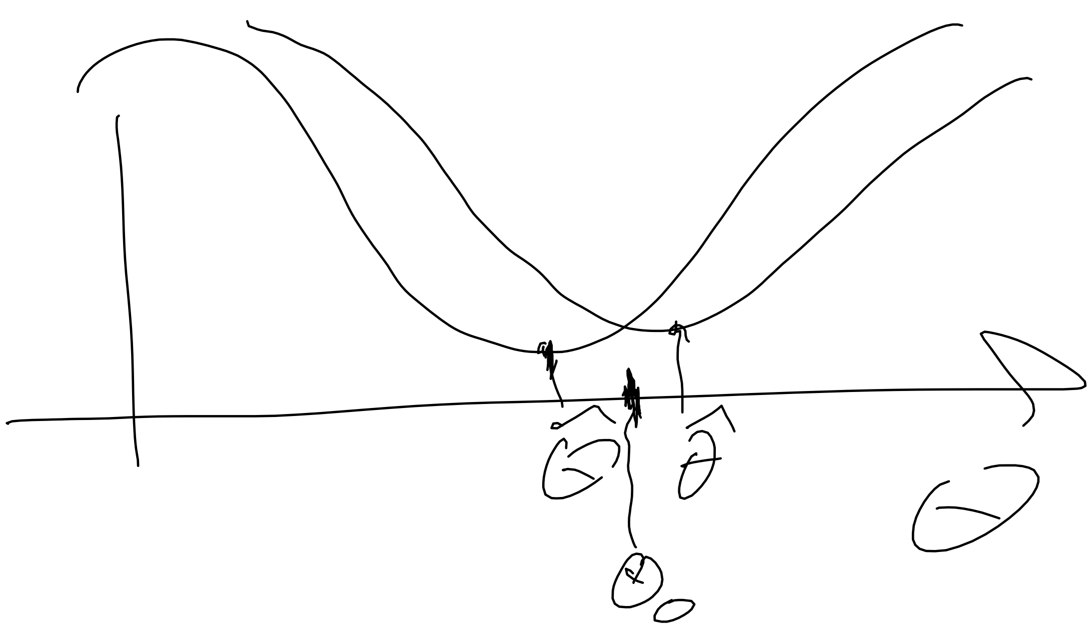
CONCLUSION (FROM EXAMPLE)

NO ESTIMATOR $\hat{\theta}(y)$ CAN BE BETTER THAN MAX. LIKELIHOOD ESTIMATOR

(BECAUSE

$$\mathrm{Cov}\,\hat{\theta} \succeq M^{-1} = \mathrm{Cov}\,\hat{\theta}_{ML} = (\Phi^T \Sigma^{-1} \Phi)^{-1}$$

FOR NONLINEAR LEAST SQUARES



HESSIAN MATRIX OF NEG-LOG-LIKELIHOOD IS ESTIMATE FOR FISHER-INFORMATION MATRIX