Advanced Signal Processing

Minimum Variance Unbiased Estimation (MVU)

Danilo Mandic

room 813, ext: 46271

Department of Electrical and Electronic Engineering
Imperial College London, UK

d.mandic@imperial.ac.uk, URL: www.commsp.ee.ic.ac.uk/∼mandic

Objectives

- Learn the concept of **minumum variance unbiased (MVU)** estimation
- Investigate how the accuracy of an estimator depends upon the relationship between the unknown parameter(s) and the PDF of noise
- Study the requirements for the design of an efficient estimator
- Analyse the Cramer–Rao Lower Bound (CRLB) for the scalar case
- Extension to the Cramer–Rao Lower Bound (CRLB) for the vector case
- Sinusoid parameter estimation, linear models
- Dependence on data length (motivation for 'sufficient statistics')
- Examples:
  - ⊛ DC level in WGN (frequency estimation in power, bioengineering)
  - ⊛ Parameters of a sinusoid (scalar case, vector case)
  - ⊛ A new view of Fourier analysis
  - ⊛ System identification

What is the Cramer–Rao Lower Bound (CRLB)

☞ The CRLB is a lower bound on the variance **of any unbiased estimator**.

In other words, if $\hat{\theta}$ is an unbiased estimator of θ, then

$$\sigma^2_{\hat{\theta}} \geq CRLB_{\hat{\theta}}(\theta) \qquad \text{or} \qquad \sigma^2_{\hat{\theta}} \geq \sqrt{CRLB_{\hat{\theta}}(\theta)}$$

Therefore, the CRLB is a benchmark which tells us the best we can ever expect to be able to achieve with an unbiased estimator.

The CRLB is a must–check quantitative bound for:

- Feasibility studies (sensor relevance, if we met problem specifications)
- Assessment of quality of any derived estimator (we can only do as good or worse than CRLB)
- It can sometimes provide the form of MVU (we just read it out from CRLB theorem)
- It may be used to demonstrate the importance of physical/signal parameters to the estimation problem (e.g. optimum freq. for power)

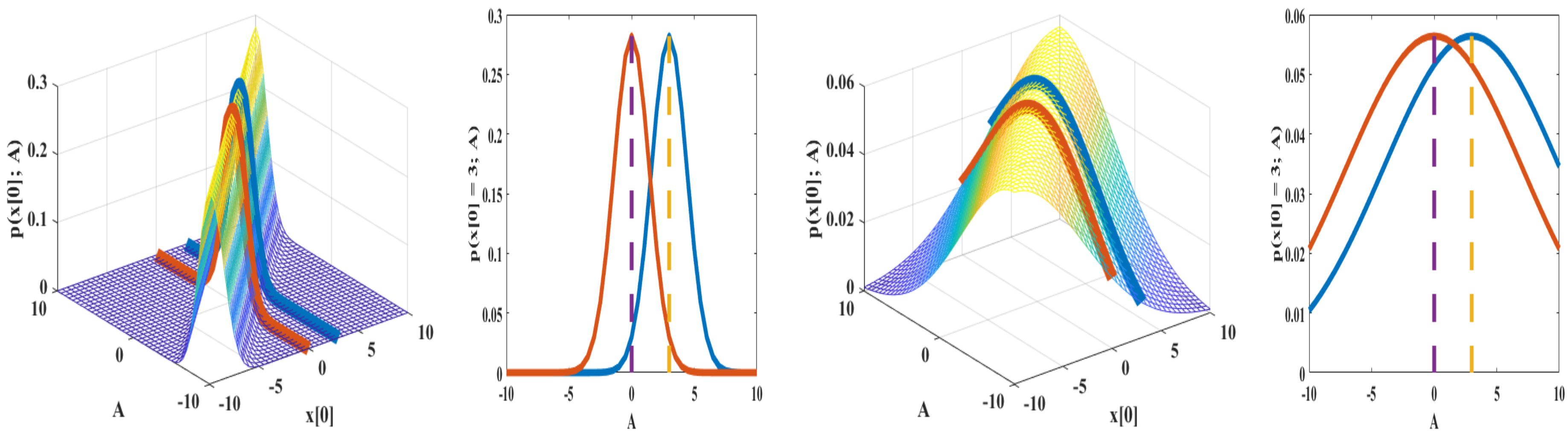
The need for the “parametrised” pdf, $p(x[0];\theta)$

Interpretation of $p(\mathbf{x};\theta)$ $\looparrowright$ a function of θ for fixed observed data $\mathbf{x}$

Q: What determines how well we estimate the unknown θ from the observed data $\mathbf{x}$?

A: Since the data $\mathbf{x}$ is a random process which depends on θ, it is the **parametrised pdf** which describes that dependence, denoted by $p(\mathbf{x};\theta)$

☞ Clearly, if $p(\mathbf{x};\theta)$ depends strongly/weakly on θ, then this implies that we should be able to estimate θ well/poorly.



Left: Strong dependence on θ **Right:** Weak dependence on θ

☞ The mean of the parametrised pdf (red & blue slices) depends on the observed point $x[0]$.

Example 1: Consider a single observation $x[0] = A + w[0]$, where $w[0] \sim \mathcal{N}(0, \sigma^2)$

The simplest estimator of the DC level A in white noise $w[0] \sim \mathcal{N}(0, \sigma^2)$ is

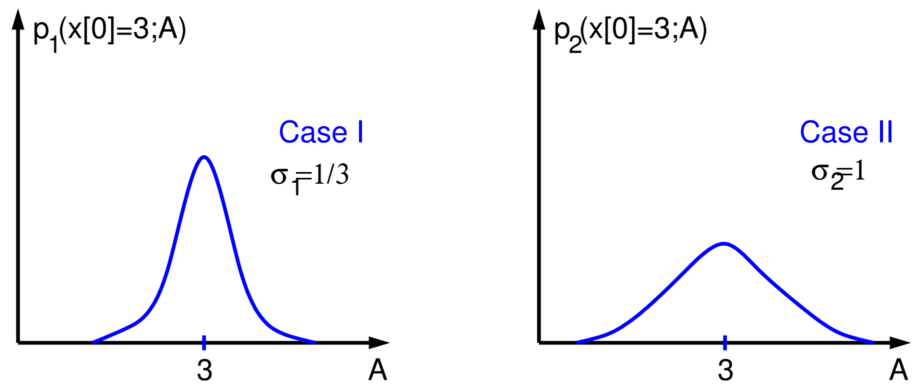
$\hat{A} = x[0] \quad \Rightarrow \quad$ **estimator $\hat{A}$ is unbiased, with the variance of σ^2**

To show that the estimator accuracy improves as σ^2 decreases:

◦ Consider

$$p_i(x[0]; A) = \frac{1}{\sqrt{2\pi\sigma_i^2}} exp\left[-\frac{1}{2\sigma_i^2}(x[0] - A)^2\right]$$

for $\underbrace{x[0] = 3}_{fundamental\ step,\ we\ are\ fixing\ the\ data\ value}$ and $i = 1, 2$ with $\sigma_1 = \frac{1}{3}$ and $\sigma_2 = 1$



Clearly, as $\sigma_1 < \sigma_2$, the DC level A is estimated more accurately with $p_1(x[0]; A)$

Likely candidates for values of $A \in 3 \pm 3\sigma \;\Rightarrow\;$ therefore [2, 4] for σ_1 and [0, 6] for σ_2.

Likelihood function

When the PDF is viewed as a function of an unknown parameter **(with the dataset $\{x\} = x[0], x[1], \ldots$ fixed)** it is termed the **"likelihood function"**.

- The **"sharpness"** of the likelihood function determines the accuracy at which the unknown parameter may be estimated.
- Sharpness is measured by the **"curvature"** $\looparrowright$ a negative of the second derivative of the logarithm of the likelihood function **at its peak**.

Example 2: Estimation based on one sample of a DC level in WGN

$$\ln p(x[0]; A) = -\ln \sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}\big(x[0] - A\big)^2$$

then
$$\frac{\partial \ln p(x[0]; A)}{\partial A} = \frac{1}{\sigma^2}\big(x[0] - A\big)$$

and the curvature
$$-\frac{\partial^2 \ln p(x[0]; A)}{\partial A^2} = \frac{1}{\sigma^2}$$

Therefore, as expected, the curvature increases as σ^2 decreases.

☞ **Curvature** $\nearrow$ $\Rrightarrow$ **PDF concentration** $\nearrow$ $\Rrightarrow$ **Accuracy** $\nearrow$

Likelihood function: Curvature

Since we know that the variance of the estimator equals σ^2, then

$$var(\hat{A}) = \frac{1}{-\frac{\partial^2 \ln p(x[0];A)}{\partial A^2}}$$

and **the variance decreases as the curvature increases.**

Generally, the **second derivative does depend upon one data point,** $x[0]$, and hence a **more appropriate measure of curvature is the statistical measure** (average over many random $x[0]$)

$$-E\left[\frac{\partial^2 \ln p(x[0];A)}{\partial A^2}\right]_{A=\text{true value}}$$

which measures the average curvature of the log-likelihood function

Recall: The likelihood function is a random variable due to $x[0]$

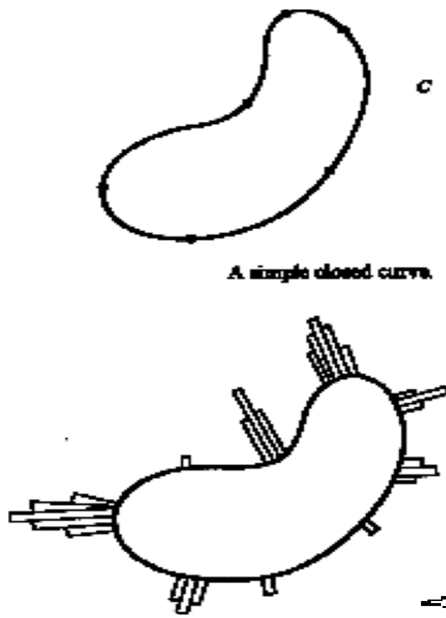
Recall: The Mean Square Error $\looparrowright$ MSE $=$ Bias^2 $+$ variance

☞ It makes perfect sense to look for a minimum variance unbiased (MVU) solution

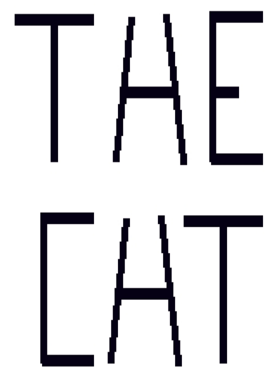
Link with human perception

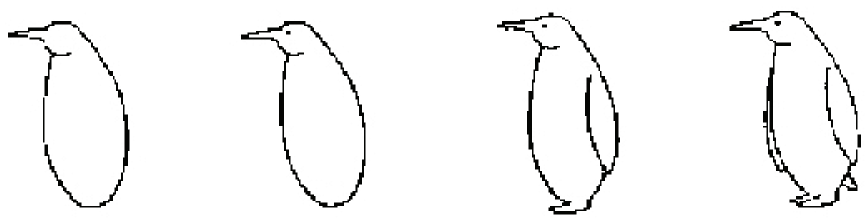
In the 50s, phsychologist Fred Attneave recorded eye dwellings on objects

Example 3a): The drawing of a bean (top) and the histogram of eye dwellings (bottom)



Example 3b): Read the words below ... now read letter by letter ... are you still sure?





Example 3c): Is the drawing on the left still a penguin?

So, what is the sufficient information to 'estimate' an object?

The Cramer–Rao Lower Bound (CRLB)

Theorem: [CRLB] **Assumption:** The PDF $p(\mathbf{x};\theta)$ satisfies the "**regularity**" condition

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0, \qquad \forall \theta$$

where the expectation is taken with respect to $p(\mathbf{x};\theta)$.

Then, the variance of any **unbiased** estimator, $\hat{\theta}$, must satisfy

$$var(\hat{\theta}) \geq \frac{1}{-E\left\{\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right\}}$$

where the derivative is evaluated at the true value of θ.

CRLB for a scalar parameter, continued

Moreover, an unbiased estimator may be found that attains the bound for all θ, if and only if for some functions g and $\mathcal{I}$

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = \mathcal{I}(\theta)\big(g(\mathbf{x}) - \theta\big)$$

This estimator is the **minimum variance unbiased (MVU) estimator**, for which

$$\hat{\theta} = g(\mathbf{x})$$

and its minimum variance

$$\frac{1}{\mathcal{I}(\theta)}$$

—— end of CRLB theorem ——

Remark: Since the variance $var(\hat{\theta}) \geq \frac{1}{-E\left\{\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right\}}$, the evaluation of the "curvature term" gives

$$E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right] = \int \frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} p(\mathbf{x};\theta) d\mathbf{x}$$

Obviously, in general the bound depends on the parameter θ and the data length

Example 4: Physical relevance of CRLB

☞ Point 3 from Slide 3: "CRLB can sometimes provide the form of MVU"

Shall we therefore compare the form of regularity condition with Example 3

Regularity condition: $$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = \mathcal{I}(\theta)\big(g(\mathbf{x}) - \theta\big)$$

inverse of the minimum achievable variance ↑ ↑ form of the optimum est.

Compare with what we have derived for $x[0] = A + w[0]$ (Slide 6)

↓ $g(x[0])$

$$\frac{\partial \ln p(x[0];A)}{\partial A} = \frac{1}{\sigma^2}\big(x[0] - A\big)$$

inverse of the Fisher information ↑ ↑ the unknown parameter

☞ By inspection, the optimum estimate is $\hat{A} = g([x[0]) = x[0]$

☞ From the CRLB theorem the optimum variance of this estimator: $\frac{1}{\mathcal{I}(\theta)} = \sigma^2$

Therefore: Good estimator ⇛ variance ↘ and curvature ↗

Poor estimator ⇛ variance ↗ and curvature ↘ (see Slide 5)

Example 5: Estimation of a DC level in WGN

Consider the estimation of a DC level in WGN, assume N observations

$$x[n] = \underbrace{A}_{unknown\ DC\ level} + \underbrace{w[n]}_{noise\ with\ known\ pdf} \qquad n = 0, 1, 2, \ldots, N-1$$

where $w[n] \sim \mathcal{N}(0, \sigma^2)$.

Determine the CRLB for the unknown DC level A, starting from $(\theta = A)$

$$p(\mathbf{x}; \theta) = p(\mathbf{x}; A) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2\sigma^2}\left(x[n] - A\right)^2\right]$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n] - A\right)^2\right]$$

☞ Estimation of a DC level is very useful, e.g. in the time-frequency plane a sinusoid of frequency f is represented by a straight line ☺

Example 5: DC level in WGN $\looparrowright$ continued

Upon taking the first derivative, we have

$$\begin{aligned}\frac{\partial \ln p(\mathbf{x};A)}{\partial A} &= \frac{\partial}{\partial A}\left[-\ln\left[2\pi\sigma^2\right]^{N/2} - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right] \\ &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right) = \frac{N}{\sigma^2}(\bar{x}-A)\end{aligned}$$

where $\bar{x}$ is the sample mean.

CRLB connection: $\bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x(n) = g(\mathbf{x})$, $\quad var(\hat{A}) = \frac{1}{\mathcal{I}(A)} = \frac{\sigma^2}{N}$

Upon differ. again $\downarrow$ does not depend on $\mathbf{x}$, so no $E\{\cdot\}$

$$\frac{\partial^2 \ln p(\mathbf{x};A)}{\partial A^2} = -\frac{N}{\sigma^2}$$

Therefore $var(\hat{A}) = \frac{\sigma^2}{N} =$ CRLB, which implies that **the sample mean estimator attains the Cramer-Rao LB and must, therefore, be an MVU estimator in WGN.**

Example 5: DC level in WGN, spelling out previous slide

Upon taking the first derivative, we have

$$\frac{\partial \ln p(\mathbf{x};A)}{\partial A} = \frac{\partial}{\partial A}\left[-\ln\left[2\pi\sigma^2\right]^{N/2} - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right]$$

$$= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right) = \frac{N}{\sigma^2}(\bar{x}-A) \overset{\text{CRLB Th}}{=} \mathcal{I}(A)\big(g(\mathbf{x})-A\big)$$

$\mathcal{I}(A)\uparrow \quad \uparrow g(\mathbf{x})$

CRLB connection: $\bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x(n) = g(\mathbf{x})$, $\quad var(\hat{A}) = \frac{1}{\mathcal{I}(A)} = \frac{\sigma^2}{N}$

Upon differentiating again $\quad \downarrow$ does not depend on $\mathbf{x}$, so no $E\{\cdot\}$

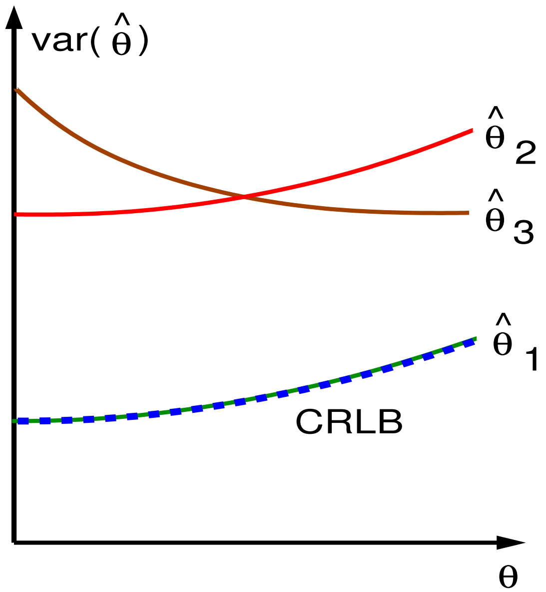
$$\frac{\partial^2 \ln p(\mathbf{x};A)}{\partial A^2} = -\frac{N}{\sigma^2}$$

Therefore $var(\hat{A}) = \frac{\sigma^2}{N} =$ CRLB, which implies that **the sample mean estimator attains the Cramer-Rao LB and must, therefore, be an MVU estimator in WGN.** (for any other estimator, $\tilde{A}$, $var(\tilde{A}) \geq \sigma^2/N$)

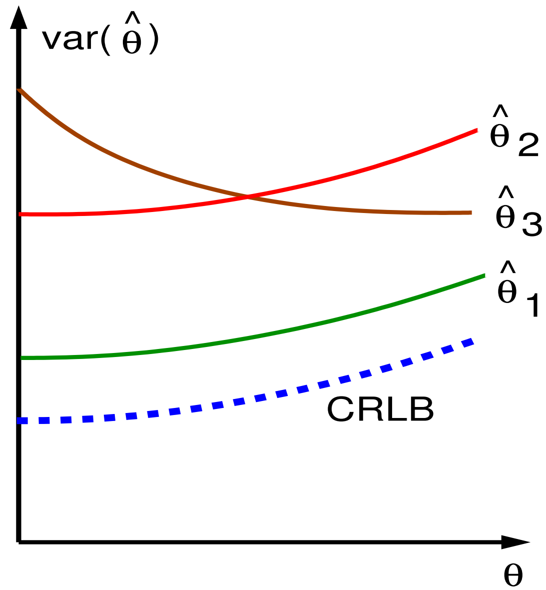
Efficient estimator ↬ concept

Def: An estimator which is unbiased and attains the CRLB is said to be efficient. In other words, an estimator is efficient if:

- ◦ It is an Minimum Variance Unbiased (MVU) estimator, and
- ◦ It efficiently uses the data.



$\hat{\theta}_1$ is efficient and MVU, $\hat{\theta}_2, \hat{\theta}_3$ are not



$\hat{\theta}_1$ may be MVU but is not efficient

☞ Not all estimators (phase est.) & not all MVU estimators are efficient

Fisher information

The term
$$\mathcal{I}(\theta) = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right]$$
in the CRLB theorem is referred to as the **Fisher information.**

Intuitively:
the more information available $\rightsquigarrow$ the lower the bound $\rightsquigarrow$ less variance

☞ **Essential properties of an information measure:**

- ☞ Non–negative
- ☞ Additive for independent observations

☞ General CRLB for **arbitrary signals** in WGN (see the next slide)
$$var(\hat{\theta}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1}\left(\frac{\partial s[n;\theta]}{\partial\theta}\right)^2}$$

Accurate estimators: Signals change rapidly with the parameter changes. Therefore $\frac{\partial s[n;\theta]}{\partial\theta}$ above acts as a "sensitivity" term. (see Appendix)

General case: Arbitrary signal in noise

Consider a deterministic signal $s[n;\theta]$ observed in WGN, $w \sim \mathcal{N}(0, \sigma^2)$

$$x[n] = s[n;\theta] + w[n], \qquad n = 0, 1, \ldots, N-1$$

Then, the PDF for $\mathbf{x}$ parametrised by $\boldsymbol{\theta}$ has the form

$$p(\mathbf{x};\theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-s[n;\theta]\right)^2}$$

and so

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-s[n;\theta]\right)\frac{\partial\, s[n;\theta]}{\partial\theta}$$

$$\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\Big[\big(\underbrace{x[n]-s[n;\theta]}_{E\{x[n]\}=s[n;\theta]}\big)\frac{\partial^2 s[n;\theta]}{\partial\theta^2} - \big(\frac{\partial s[n;\theta]}{\partial\theta}\big)^2\Big]$$

Therefore, the Fisher information

$$\mathcal{I}(\theta) = -E\big[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\big] = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(\frac{\partial s[n;\theta]}{\partial\theta}\right)^2$$

Example 6: Sinusoidal frequency estimation

(the CRLB depends both on the unknown parameter f_0 and the data length N)

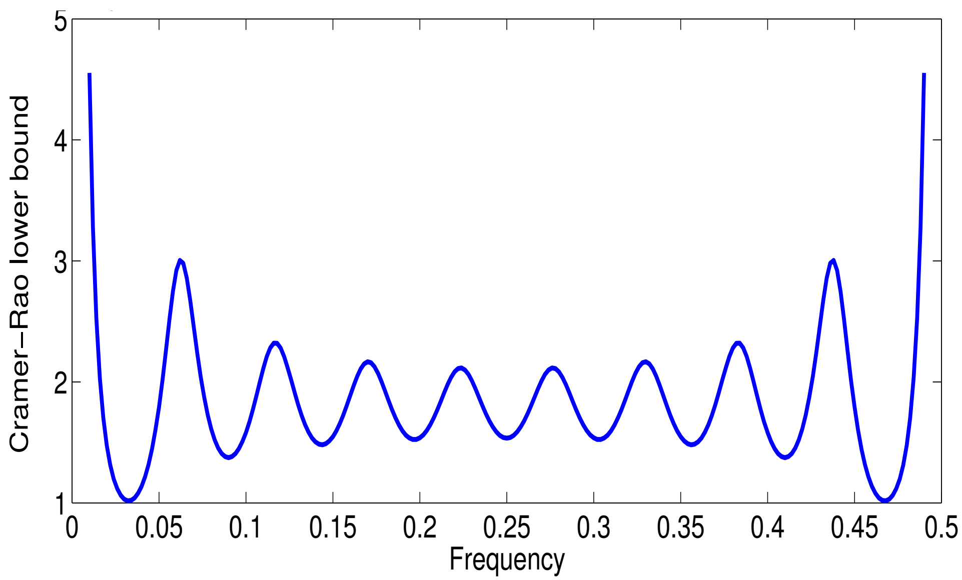
Consider a general sinewave in noise: $x[n] = A\cos\left(2\pi f_0 n + \Phi\right) + w[n]$

If only the frequency f_0 is unknown, then (for normalised frequency)

$$s[n; f_0] = \underbrace{A}_{known} \cos(2\pi f_0 n + \underbrace{\Phi}_{known}), \qquad 0 < f_0 < \frac{1}{2}$$

and

$$var(\hat{f}_0) \geq \frac{\sigma^2}{A^2 \sum_{n=0}^{N-1} \left[2\pi n \ \sin(2\pi f_0 n + \Phi)\right]^2}$$



Note the preferred frequencies, e.g.

$f \approx 0.03$, and that

for $f_0 \rightarrow \{0, 1/2\}$ the CRLB $\rightarrow \infty$

Paramet.: N $=$ 10, Φ $=$ 0, SNR$=A^2/\sigma^2 = 1$

Extension to a vector parameter

we now have the Fisher Information Matrix $\mathcal{I}$, s.t. $[\mathcal{I}(\boldsymbol{\theta})]_{ij} = -E\left[\frac{\partial^2 ln\ p(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right]$

Formulation: Estimate a vector parameter $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_p]^T$

- Recall that an unbiased estimator $\hat{\boldsymbol{\theta}}$ is efficient (and therefore an MVU estimator) when it satisfies the conditions of the CRLB

- It is assumed that the PDF $p(\mathbf{x};\boldsymbol{\theta})$ satisfies the **regularity conditions**

$$E\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right] = \mathbf{0}, \quad \forall\boldsymbol{\theta}$$

- Then, the covariance matrix of any unbiased estimator $\hat{\boldsymbol{\theta}}$ satisfies

  $\mathbf{C}_{\hat{\boldsymbol{\theta}}} - \mathcal{I}^{-1}(\boldsymbol{\theta}) \geq \mathbf{0}$ (symbol $\geq \mathbf{0}$ means that $\mathbf{C}_{\hat{\boldsymbol{\theta}}}$ is positive semidefinite)

- The Fisher Information Matrix is given by $[\mathcal{I}(\boldsymbol{\theta})]_{ij} = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right]$

☞ An unbiased estimator $\hat{\boldsymbol{\theta}} = \mathbf{g}(\mathbf{x})$ exists that satisfies the bound $\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathcal{I}^{-1}(\boldsymbol{\theta})$ if and only if

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \mathcal{I}(\boldsymbol{\theta})\big(\mathbf{g}(\mathbf{x}) - \boldsymbol{\theta}\big)$$

Extension to a vector parameter: Fisher information

Some observations:

- Elements of the Information Matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$ are given by

$$[\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})]_{ij} = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right]$$

  where the derivatives are evaluated at the true values of the parameter vector.

- The CRLB theorem provides a powerful tool for finding MVU estimators for a vector parameter.

☞ **MVU estimators for linear models are found with the Cramer–Rao Lower Bound (CRLB) theorem.**

Example 7: Sinusoid parameter estimation $\looparrowright$ vector case

Consider again a general sinewave

$$s[n] = A\cos\left(2\pi f_0 n + \Phi\right)$$

where A, f_0 and Φ are all unknown. Then, the data model becomes

$$x[n] = A\ cos(2\pi f_0 n + \Phi) + w[n] \qquad n = 0, 1, \ldots, N-1$$

where $A > 0$, $0 < f_0 < 1/2$, and $w[n] \sim \mathcal{N}(0, \sigma^2)$.

Task: Determine CRLB for the parameter vector $\boldsymbol{\theta} = [A, \quad f_0, \quad \Phi]^T$.

Solution: The elements of the Fisher Information Matrix become

$$\frac{\partial^2 \ln p(\mathbf{x}; A, f_0, \Phi)}{\partial A^2} \downarrow \qquad \downarrow \frac{\partial^2 \ln p(\mathbf{x}; A, f_0, \Phi)}{\partial A \partial f_0} \qquad \swarrow \frac{\partial^2 \ln p(\mathbf{x}; A, f_0, \Phi)}{\partial A \partial \Phi}$$

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\begin{bmatrix} N/2 & 0 & 0 \\ 0 & 2A^2\pi^2\sum_{n=0}^{N-1} n^2 & \pi A \sum_{n=0}^{N-1} n \\ 0 & \pi A \sum_{n=0}^{N-1} n & \frac{NA^2}{2} \end{bmatrix}$$

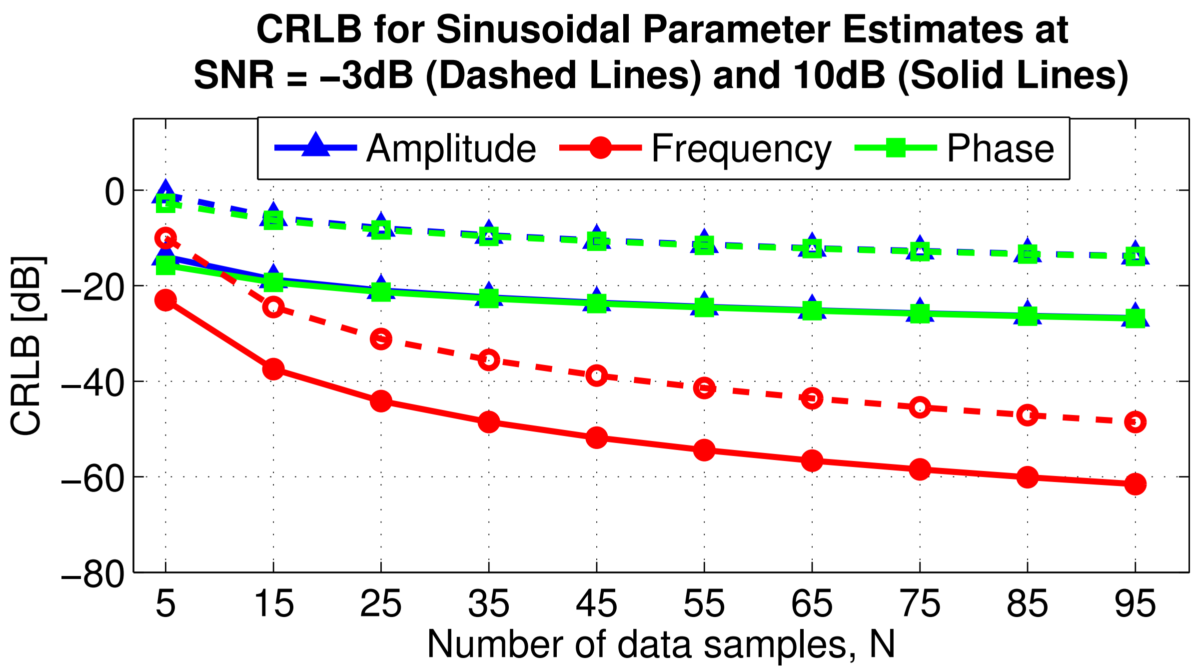
$$\frac{\partial^2 \ln p(\mathbf{x}; A, f_0, \Phi)}{\partial \Phi \partial A} \uparrow \qquad \uparrow \frac{\partial^2 \ln p(\mathbf{x}; A, f_0, \Phi)}{\partial \Phi \partial f_0} \qquad \nwarrow \frac{\partial^2 \ln p(\mathbf{x}; A, f_0, \Phi)}{\partial \Phi^2}$$

Example 7: Sinusoid parameter estimation $\looparrowright$ continued

since $C_{\hat{\theta}} = \mathcal{I}^{-1}(\theta)$ (slide 15) make an inverse of the FIM

After inversion of $\mathcal{I}(\boldsymbol{\theta})$, its diagonal components are ($\eta = \frac{A^2}{2\sigma^2}$ is SNR):

$$var(\hat{A}) \geq \frac{2\sigma^2}{N} \qquad var(\hat{f}_0) \geq \frac{12}{(2\pi)^2\eta N(N^2-1)} \qquad var(\hat{\Phi}) \geq \frac{2(2N-1)}{\eta N(N+1)}$$



☞ the variance of the estimated parameters of a sinusoid behaves $\propto 1/\eta$ and $\propto 1/N^3$, thus **exhibiting strong sensitivity to data length**

Linear models

Generally, it is difficult to determine the MVU estimator.

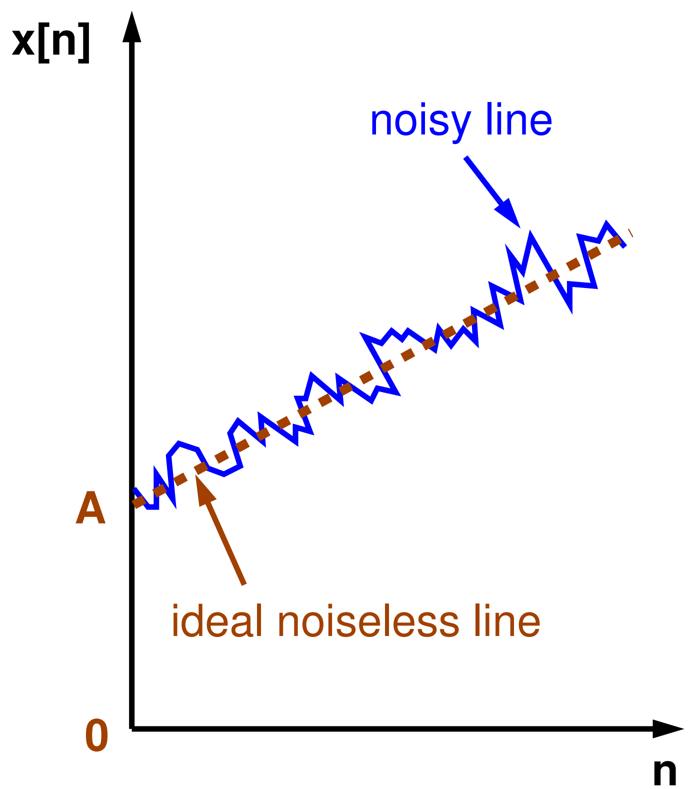
- In signal processing, however, a **linear data model** can often be employed $\Rightarrow$ straightforward to determine the MVU estimator.

Example 8: Linear model of a straight line in noise

$$x[n] = A + Bn + w[n]$$
$$n = 0, 1, \ldots, N-1$$

where

- $w[n] \sim \mathcal{N}(0, \sigma^2)$,
- B - slope and
- A - intercept.



Linear models: Compact notation (Example 8 contd.)

This data model can be written more compactly in the matrix notation as

known $\downarrow$ $\swarrow$ known pdf

$$\underline{x} = H\underline{\theta} + \underline{w} \qquad \text{or} \qquad \mathbf{x} = \mathbf{H}\,\boldsymbol{\theta} + \mathbf{w}$$

where observed $\nearrow$ $\uparrow$ unknown

$$\mathbf{x} = \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix} = \big[x[0], x[1], \ldots, x[N-1]\big]^T \qquad \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix}$$

and

$$\begin{aligned} \boldsymbol{\theta} &= [A \quad B]^T \\ \mathbf{w} &= [w[0], w[1], \ldots, w[N-1]]^T \\ \mathbf{w} &\sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}) \\ \mathbf{I} &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \operatorname{diag}(1, 1, \ldots, 1) \end{aligned}$$

Linear models: Fisher information matrix

NB: $p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})}{2\sigma^2}}$

☞ The CRLB theorem can be used to obtain the MVU estimator for $\boldsymbol{\theta}$

The MVU estimator, $\hat{\boldsymbol{\theta}} = g(\mathbf{x})$, will satisfy

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})\big(g(\mathbf{x}) - \boldsymbol{\theta}\big)$$

where $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$ is the **Fisher information matrix**, whose elements are

$$[\boldsymbol{\mathcal{I}}]_{ij} = -E\Big[\frac{\partial^2 \ \ln \ p(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\Big]$$

Applying the Linear Model

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}}\left[-\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\big(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}\big)^T\big(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}\big)\right]$$

$$= -\frac{1}{2\sigma^2}\frac{\partial}{\partial \boldsymbol{\theta}}\Big[N\sigma^2\ln(2\pi\sigma^2) + \mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}\Big]$$

Note that only the quadratic term in θ involves the matrix $\mathbf{H}$

Linear models: Some useful matrix/vector derivatives

the derivation is given in the Lecture Supplement

Use the identities (remember that both $\mathbf{b}^T\boldsymbol{\theta}$ and $\boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}$ are scalars)

$$\frac{\partial \mathbf{b}^T\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{b} \quad \looparrowright \quad \frac{\partial \mathbf{x}^T\mathbf{H}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = (\mathbf{x}^T\mathbf{H})^T = \mathbf{H}^T\mathbf{x}$$

$$\frac{\partial \boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta} \quad \looparrowright \quad \frac{\partial \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$$

(which you should prove for yourself), that is, follow the rules of vector/matrix differentiation.

As a rule of thumb, watch for the position of the $(\cdot)^T$ operator

Then,

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\sigma^2}\left[\mathbf{H}^T\mathbf{x} - \mathbf{H}^T\mathbf{H}\boldsymbol{\theta}\right]$$

Linear models: Cramer-Rao lower bound

Find the MVU estimator: $\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})\big(g\left(\mathbf{x}\right) - \boldsymbol{\theta}\big)$

Similarly to the vector CRLB, $\looparrowright$ recall that $(\mathbf{H}^T\mathbf{H})^T = \mathbf{H}^T\mathbf{H}$,

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) = -\frac{\partial^T}{\partial \boldsymbol{\theta}}\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \frac{1}{\sigma^2}\mathbf{H}^T\mathbf{H}$$

Therefore

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \underbrace{\frac{1}{\sigma^2}\mathbf{H}^T\mathbf{H}}_{\mathcal{I}(\boldsymbol{\theta})}\Big[\underbrace{\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}}_{\mathbf{g}(\mathbf{x})} - \boldsymbol{\theta}\Big]$$

By inspection, the **linear estimator** is then given by

$$\hat{\boldsymbol{\theta}} = g\left(\mathbf{x}\right) = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}$$

provided $\left(\mathbf{H}^T\mathbf{H}\right)^{-1}$ is invertible (it is, as $\mathbf{H}$ is full rank, with orthogonal rows and columns).

The covariance matrix of $\hat{\boldsymbol{\theta}}$ now becomes $\quad \mathbf{C}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{\mathcal{I}}^{-1}\left(\boldsymbol{\theta}\right) = \sigma^2\left(\mathbf{H}^T\mathbf{H}\right)^{-1}$

CRLB for linear models – continued

- **The MVU estimator for the linear model is efficient $\looparrowright$ it attains the CRLB**

- The columns of $\mathbf{H}$ must be **linearly independent** for $\left(\mathbf{H}^T\mathbf{H}\right)$ to be invertible

Theorem: (Minimum Variance Unbiased Estimator for the Linear Model)
If the observed data can be modelled as

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

where

$\mathbf{x}$	is an N×1 "vector of observed data"
$\mathbf{H}$	is an N×p "observation (measurement) matrix" of rank p
$\boldsymbol{\theta}$	is a p×1 unknown "parameter vector"
$\mathbf{w}$	is an N×1 additive "noise vector" $\sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$

CRLB for linear models: Theorem, continued

Then, the MVU estimator is given by

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}$$

for which the covariance matrix

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2\left(\mathbf{H}^T\mathbf{H}\right)^{-1}$$

Note that the statistical performance of $\hat{\boldsymbol{\theta}}$ is now completely described because $\hat{\boldsymbol{\theta}}$ is **linear transformation** of a Gaussian vector $\mathbf{x}$, i.e.

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}\left(\boldsymbol{\theta}, \sigma^2\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\right)$$

End of Theorem □

Example 9: Fourier analysis

The data model is given by $(n = 0, 1, \ldots, N-1,\ w[n] \sim \mathcal{N}(0, \sigma^2))$

$$x[n] = \sum_{k=1}^{M} a_k\ cos\left(\frac{2\pi kn}{N}\right) + \sum_{k=1}^{M} b_k\ sin\left(\frac{2\pi kn}{N}\right) + w[n]$$

where the Fourier coefficients, a_k and b_k, that is the amplitudes of the cosine and sine terms, are to be estimated.

∘ Frequencies are **harmonically related**, i.e. $f_1 = \frac{1}{N}$, and $f_k = \frac{k}{N}$.

∘ Then, the parameter vector is $\boldsymbol{\theta} = [a_1, a_2, \ldots, a_M, b_1, b_2, \ldots, b_M]^T$

and the observation matrix **H** ($N \times \underbrace{2M}_{p}$–dimensional) takes the form

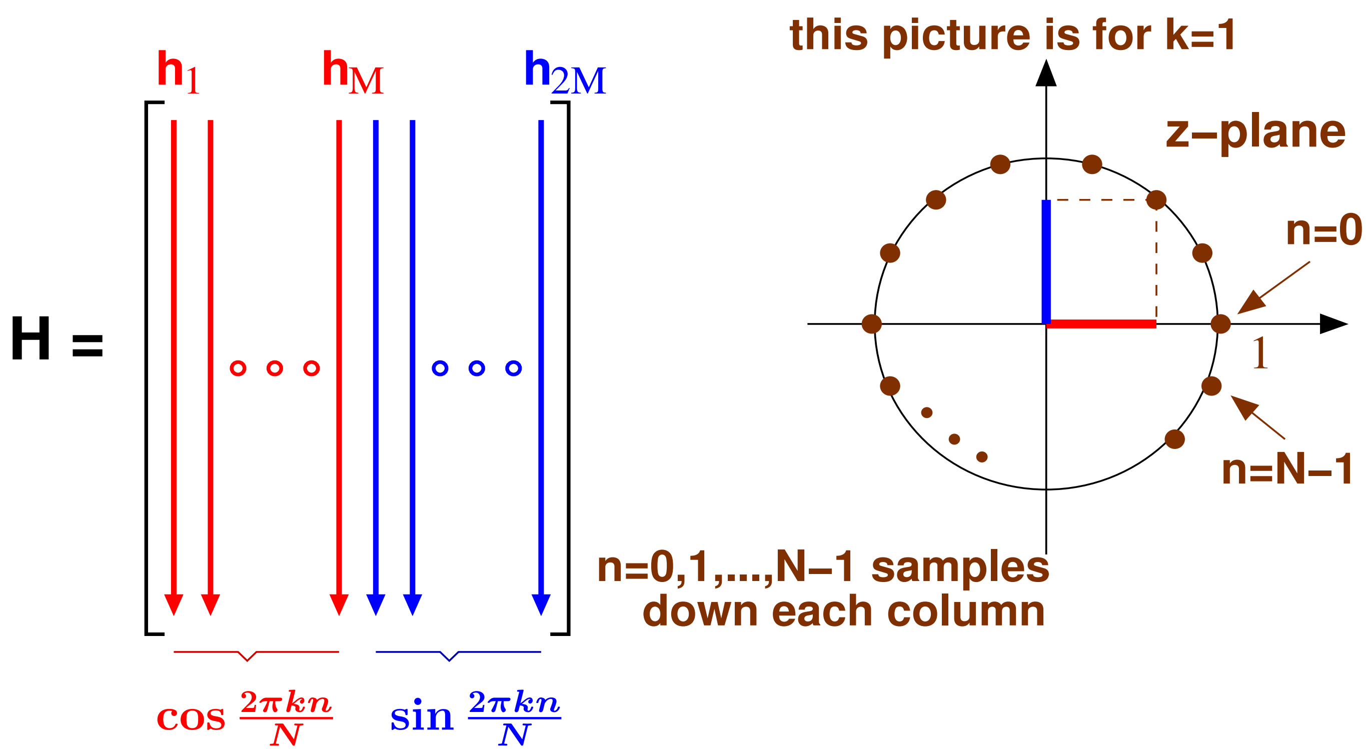
$$\mathbf{H} = \begin{bmatrix} 1 & \ldots & 1 & 0 & \ldots & 0 \\ \cos\frac{2\pi}{N} & \ldots & \cos\frac{2\pi M}{N} & \sin\frac{2\pi}{N} & \ldots & \sin\frac{2\pi M}{N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \cos\frac{2\pi(N-1)}{N} & \ldots & \cos\frac{2\pi M(N-1)}{N} & \sin\frac{2\pi(N-1)}{N} & \ldots & \sin\frac{2\pi M(N-1)}{N} \end{bmatrix}_{N\times 2M}$$

Example 9: Fourier analysis, geometric view

Data model: $$x[n] = \sum_{k=1}^{M} a_k \; cos\left(\frac{2\pi k n}{N}\right) + \sum_{k=1}^{M} b_k \; sin\left(\frac{2\pi k n}{N}\right) + w[n]$$

↑ parameters to estimate ↑ WGN ↑

Parameters: $\boldsymbol{\theta} = [a_1, a_1, \ldots, a_M, b_1, b_2, \ldots, b_M]^T]$ (Fourier coeffs.)



Example 9: Fourier analysis $\hookrightarrow$ continued

For $\mathbf{H}$ not to be under-determined, it has to satisfy $N > p \Rightarrow M < \frac{N}{2}$,

☞ For **mathematical convenience**, the columns of $\mathbf{H}$ should be **orthogonal**.

This is because the **columns of $\mathbf{H}$ form a basis of a new representation space**, and is obvious if we rewrite the measurement matrix in the form

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{2M}]$$

where $\underline{h}_i = \mathbf{h}_i$ is the i-th column of $\mathbf{H}$.

Then, for a large enough number of data points, N, due to the properties of products of sines and cosines of different frequencies

$$\mathbf{h}_i^T \mathbf{h}_j = 0 \qquad for \quad i \neq j$$

In other words, $\mathbf{h}_i \perp \mathbf{h}_j$, that is, the columns of matrix $\mathbf{H}$ are orthogonal

Example 9: Fourier analysis $\looparrowright$ contd. contd.

The **orthogonality of the columns of $\mathbf{H}$** (for large N) follows from the discrete Fourier transform (DFT) rationale

$$\sum_{n=0}^{N-1} \cos\left(\frac{2\pi i n}{N}\right)\cos\left(\frac{2\pi j n}{N}\right) = \frac{N}{2}\delta_{ij} \qquad \text{as power of a sinusoid is } A^2/2$$

$$\sum_{n=0}^{N-1} \sin\left(\frac{2\pi i n}{N}\right)\sin\left(\frac{2\pi j n}{N}\right) = \frac{N}{2}\delta_{ij}$$

$$\sum_{n=0}^{N-1} \cos\left(\frac{2\pi i n}{N}\right)\sin\left(\frac{2\pi j n}{N}\right) = 0 \qquad \forall i,j, \quad \text{s.t. } i,j = 1,2,\ldots,M < \frac{N}{2}$$

where

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

In other words: (i) $\cos i\alpha \perp \sin j\alpha,\ \forall i,j$, (ii) $\cos i\alpha \perp \cos j\alpha,\ \forall i \neq j$, (iii) $\sin i\alpha \perp \sin j\alpha,\ \forall i \neq j$

Example 9: Fourier analysis $\rightarrow$ observation matrix

Therefore (orthogonality)

$$\mathbf{H}^T\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^T \\ \vdots \\ \mathbf{h}_{2M}^T \end{bmatrix} [\mathbf{h}_1, \ldots, \mathbf{h}_{2M}] = \begin{bmatrix} \frac{N}{2} & 0 & \ldots & 0 \\ 0 & \frac{N}{2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{N}{2} \end{bmatrix} = \frac{N}{2}\mathbf{I}$$

and the MVU estimator of the Fourier coefficients is given by

$$\hat{\boldsymbol{\theta}} = (\underbrace{\mathbf{H}^T\mathbf{H}}_{=\frac{N}{2}\mathbf{I}})^{-1}\mathbf{H}^T\mathbf{x} \qquad \Rightarrow \qquad \hat{\boldsymbol{\theta}}_{MVU} = \frac{2}{N}\mathbf{H}^T\mathbf{x}$$

$$\hat{\boldsymbol{\theta}} = \frac{2}{N}\mathbf{H}^T\mathbf{x} = \frac{2}{N}\begin{bmatrix} \mathbf{h}_1^T \\ \vdots \\ \mathbf{h}_{2M}^T \end{bmatrix}\mathbf{x} = \begin{bmatrix} \frac{2}{N}\mathbf{h}_1^T\mathbf{x} \\ \vdots \\ \frac{2}{N}\mathbf{h}_{2M}^T\mathbf{x} \end{bmatrix} = \begin{bmatrix} \frac{2}{N}\sum_{n=0}^{N-1} x[n]\cos\left(\frac{2\pi\times 1\times n}{N}\right) \\ \vdots \\ \frac{2}{N}\sum_{n=0}^{N-1} x[n]\sin\left(\frac{2\pi\times 2M\times n}{N}\right) \end{bmatrix}$$

☞ Fourier coefficients of a "signal + WGN" are MVU estimates of the Fourier coefficients of the noise–free signal.

Example 9: Finally ↬ Fourier coefficients (Fourier coefficients of "signal + AWGN" are MVU estimates of Fourier coeff. of noise–free signal)

Therefore, the Fourier analysis is a linear MVU estimator, given by

$$\hat{a}_k = \frac{2}{N}\sum_{n=0}^{N-1} x[n]\, cos\left(\frac{2\pi kn}{N}\right)$$

$$\hat{b}_k = \frac{2}{N}\sum_{n=0}^{N-1} x[n]\, sin\left(\frac{2\pi kn}{N}\right)$$

where the a_k and b_k are **discrete Fourier transform coefficients.**

From CRLB for Linear Model, the covariance matrix of this estimator is

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2\left(\mathbf{H}^T\mathbf{H}\right)^{-1} = \frac{2\sigma^2}{N}\mathbf{I}$$

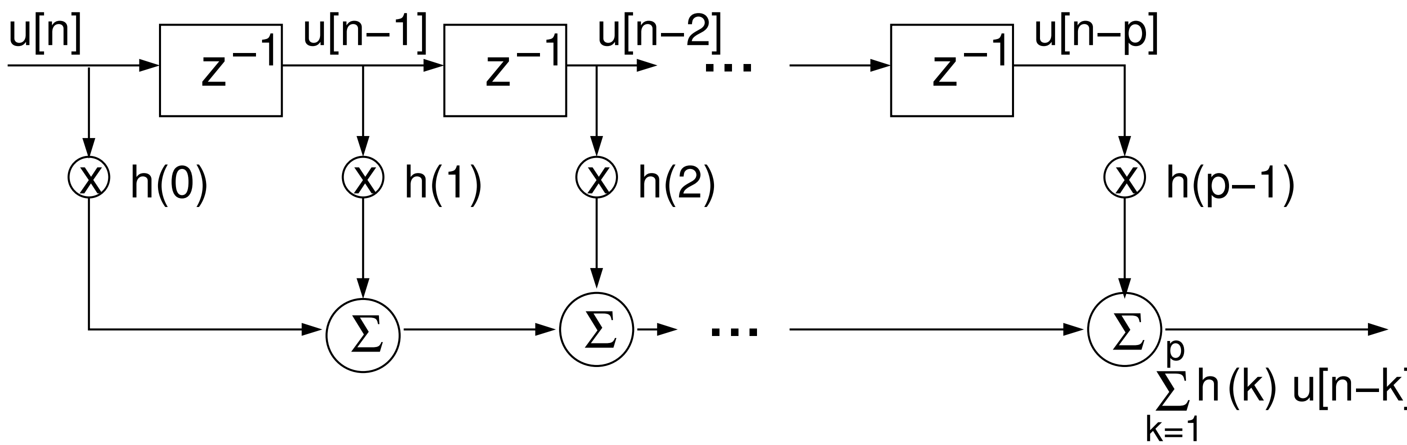
i) Note that, as $\hat{\boldsymbol{\theta}}$ is a Gaussian random variable and the covariance matrix is diagonal, the amplitude estimates are statistically independent;

ii) The orthogonality of the columns of $\mathbf{H}$ is fundamental in the computation of the MVU estimator (invertible parsimonious basis);

iii) The measurement matrix $\mathbf{H}$ is desired to be a **tall matrix with orthogonal columns.**

Example 10: System Identification (SYS ID)

Aim: To identify the model of a system (filter coefficients $\{h\}$) from input/output data. Assume an FIR filter system model given below



- The input $u[n]$ "probes" the system, then the output of the FIR filter is given by the convolution $x[n] = \sum_{k=0}^{p-1} h(k)x[n-k]$
- We wish to estimate the filter coefficients $[h(0), \ldots, h(p-1)]^T$
- In practice, the output is corrupted by additive WGN

Example 10: SYS ID $\looparrowright$ data model in noise $w \sim \mathcal{N}(0, \sigma^2)$

Data model

$$x[n] = \sum_{k=0}^{p-1} h(k)u[n-k] + w[n] \qquad n = 0, 1, \ldots, N-1$$

Equivalently, in the matrix–vector form

$$\underbrace{\begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}}_{\text{obs. vec. } \mathbf{x}} = \underbrace{\begin{bmatrix} u[0] & 0 & \ldots & 0 \\ u[1] & u[0] & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u[N-1] & u[N-2] & \ldots & u[N-p] \end{bmatrix}}_{\text{measurement matrix } \mathbf{H}} \underbrace{\begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(p-1) \end{bmatrix}}_{\text{coeff. vec. } \boldsymbol{\theta}} + \underbrace{\begin{bmatrix} w[0] \\ w[1] \\ \vdots \\ w[N-1] \end{bmatrix}}_{\text{noise vec. } \mathbf{w}}$$

that is

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w} \qquad \text{where} \qquad \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

The MVU estimator

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x} \qquad \text{with} \qquad \mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2 \left(\mathbf{H}^T\mathbf{H}\right)^{-1}$$

This representation also lends itslef to state-space modelling

Example 10: SYS ID $\looparrowright$ more about $\mathbf{H}$

Now, $\mathbf{H}^T\mathbf{H}$ becomes a symmetric Toeplitz autocorrelation matrix, given by

$$\mathbf{H}^T\mathbf{H} = N \begin{bmatrix} r_{uu}(0) & r_{uu}(1) & \dots & r_{uu}(p-1) \\ r_{uu}(1) & r_{uu}(0) & \dots & r_{uu}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{uu}(p-1) & r_{uu}(p-2) & \dots & r_{uu}(0) \end{bmatrix}$$

where

$$r_{uu}(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} u[n]u[n+k]$$

For $\mathbf{H}^T\mathbf{H}$ to be diagonal, we must have $r_{uu}(k) = 0$ for $k \neq 0$, which holds for a pseudorandom (PRN) input sequence.

Finally, when $\quad \mathbf{H}^T\mathbf{H} = N\, r_{uu}(0)\mathbf{I}$

$$\text{then} \qquad var\big(\hat{h}(i)\big) = \frac{\sigma^2}{N\, r_{uu}(0)}, \qquad i = 0, 1, \dots, p-1$$

Example 10: SYS ID $\looparrowright$ MVU estimator

For a PRN sequence, the MVU estimator becomes

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}$$

Then

$$\hat{h}(k) = \frac{1}{N\, r_{uu}(0)}\sum_{n=0}^{N-1} u[n-k]x[n]$$

and

$$\frac{r_{ux}(k)}{r_{uu}(0)} = \frac{\frac{1}{N}\sum_{n=0}^{N-1-k} u[n]x[n+k]}{r_{uu}(0)}$$

$$k = 0, 1, \ldots, p-1$$

Thus, the MVU estimator is the ratio of the input-output cross-correlation to the input autocorrelation.

☞ **Compare with the Wiener filter in Lecture 7.**

Theorem: The MVU Estimator for a General Linear Model (GLM)

i) Data model

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \underbrace{\mathbf{s}}_{\text{known signal}} + \mathbf{w}$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$$

ii) Then, the MVU estimator

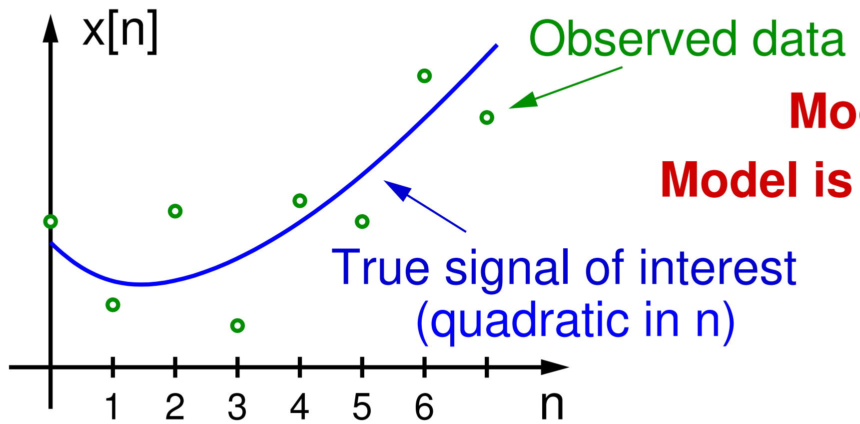
$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\mathbf{x} - \mathbf{s})$$

iii) with covariance matrix

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = (\mathbf{H}^T\mathbf{C}\mathbf{H})^{-1}$$

Example 11: The concept of "linear in the parameters" models

☞ Recall that the notion "linear" in the term "Linear Models" **does not arise from fitting straight lines to data!**



Model is quadratic in time "n"

Model is "linear in the parameters!"

Observations: $$x[n] = \underbrace{\theta_0 + \theta_1 n + \theta_2 n^2}_{\text{linear in parameters } \theta} + w[n] \quad \Rightarrow \quad \mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

where $$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} \qquad \mathbf{H} = \begin{bmatrix} 1 & 0 & 0^2 \\ 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \\ \vdots & \vdots & \vdots \\ N & N & N^2 \end{bmatrix}$$

What to remember about MVU estimators

- **An estimator is a random variable** and as such its performance can only be described statistically by its PDF
- The use of computer simulations for assessing the performance of an estimator is **rarely conclusive**
- Unbiased estimators **tend to have symmetric PDFs** centred about the true value of θ
- The minimum mean square error (MMSE) criterion is natural to search for optimal estimators, but it most often leads to unrealisable estimators (those that cannot be written solely as a function of data)
- Since *MSE*= *Bias*^{2} + *variance*, any criterion that depends on bias should be abandoned $\looparrowright$ we need to consider alternative approaches
- **Remedy:** Constrain the bias to zero and find an estimator which minimises the variance $\looparrowright$ the minimum variance unbiased (MVU) estim.
- Minimising the variance of an unbiased estimator also has the effect of concentrating the PDF of the estimation error, $\hat{\theta} - \theta$, about zero $\looparrowright$ **estimation error less likely to be large**

A few things about CRLB to remember

Even if the MVU estimator exists, there is no "turn of the crank" procedure to find it.

The CRLB sets a lower bound on the variance of any unbiased estimator!

This can be extremely useful in several ways:

- If we find an estimator that achieves the CRLB $\Rightarrow$ we known we have found an MVU estimator
- The CRLB can provide a benchmark against which we can compare the performance of any unbiased estimator
- The CRLB enables us to rule out impossible estimators. **It is physically impossible to find an unbiased estimator that beats the CRLB**
- We may require the estimator to be linear, which is not necessarily a severe restriction, as shown in the example on the estimation of Fourier coefficients

Some "rule of thumb" practical hints with CRLB

1. Start from the log–likelihood parametrised PDF function, which depends on the unknown parameter θ, that is, $\ln p(\mathbf{x};\theta)$

2. Fix $\mathbf{x}$ and take 2nd partial derivative of the log-likelhood function, that is, $\partial^2 \ln p(\mathbf{x};\theta)/\partial\theta^2$

3. If the result still depends on $\mathbf{x}$, then fix the θ and take the expected value with respect to $\mathbf{x}$.
   Otherwise, this step is not needed.

4. Should the result still depend on θ, then evaluate at every specific value of θ

5. For the CRLB, perform the reciprocal and negate

For some problems, an efficient estimator may not exist, for example the estimation of sinusoidal phase.

Appendix: An alternative from of CRLB (via sensitivity of $p(\mathbf{x};\theta)$ to θ)

Sometimes, it is easier to find CRLB as

$$var(\hat{\theta}) \geq \frac{1}{E\left\{\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right]^2\right\}} \qquad cf. \text{ the original } \quad var(\hat{\theta}) \geq \frac{1}{-E\left\{\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right\}}$$

Motivation: Sensitivity analysis, ease of interpretation

For an increment in θ, i.e. $\theta \rightarrow \theta + \Delta\theta \quad \Rightarrow \quad p(\mathbf{x};\theta) \rightarrow p(\mathbf{x};\theta+\Delta\theta)$

Then, the sensitivity of $p(\mathbf{x};\theta)$ to that change is

$$\tilde{S}_\theta^p(\mathbf{x}) = \frac{\left[\frac{\Delta p(\mathbf{x};\theta)}{p(\mathbf{x};\theta)}\right]}{\left[\frac{\Delta\theta}{\theta}\right]} = \frac{\%\text{ change in } \ p(\mathbf{x};\theta)}{\%\text{ change in } \ \theta} = \left[\frac{\Delta p(\mathbf{x};\theta)}{\Delta\theta}\right]\left[\frac{\theta}{p(\mathbf{x};\theta)}\right]$$

For $\Delta\theta \rightarrow 0 \quad S_\theta^p(\mathbf{x}) = \lim_{\Delta\theta\rightarrow 0} \tilde{S}_\theta^p(\mathbf{x}) = \left[\frac{\partial p(\mathbf{x};\theta)}{\partial\theta}\right]\left[\frac{\theta}{p(\mathbf{x};\theta)}\right] = \theta\frac{\partial\, ln\, p(\mathbf{x}'\theta)}{\partial\theta}$

(recall the derivative rules of a $\log$ function, $\frac{\partial\, ln\, f(x)}{\partial x} = \frac{1}{f(x)}\frac{\partial f(x)}{\partial x}$)

Appendix: An alternative from of CRLB (contd.) (via sensitivity of $p(\mathbf{x};\theta)$ to θ)

Therefore (Gardner, IEEE Transactions on Information Theory, July 1979)

$$\frac{var(\hat{\theta}}{\theta^2} = \frac{1}{\theta^2\, E\left\{\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right]^2\right\}} = \frac{1}{\theta^2\, E\left\{\left[S_\theta^p(\mathbf{x})\right]^2\right\}}$$

Interpretation: This is an inverse mean square sensitivity of $p(\mathbf{x};\theta)$ to θ.

- Modelling and estimation are obviously intertwined
- Unknown parameters may have a physical interpretation, such as e.g. direction in beamforming, delay in radar, ...
- Otherwise, parameters may be part of an imposed model, such as e.g. the fixed sine-cosine bases in Fourier analysis

Notes:

○

Notes:

◦