Chapter 5

- **Estimation Theory and Applications**

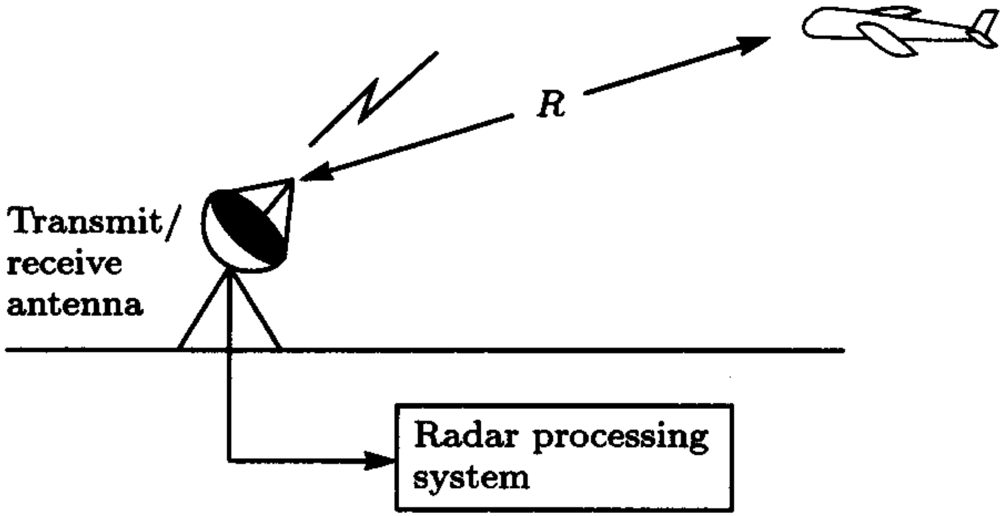
References:

- S.M.Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory,* Prentice Hall, 1993

Estimation Theory and Applications

Application Areas

1. Radar

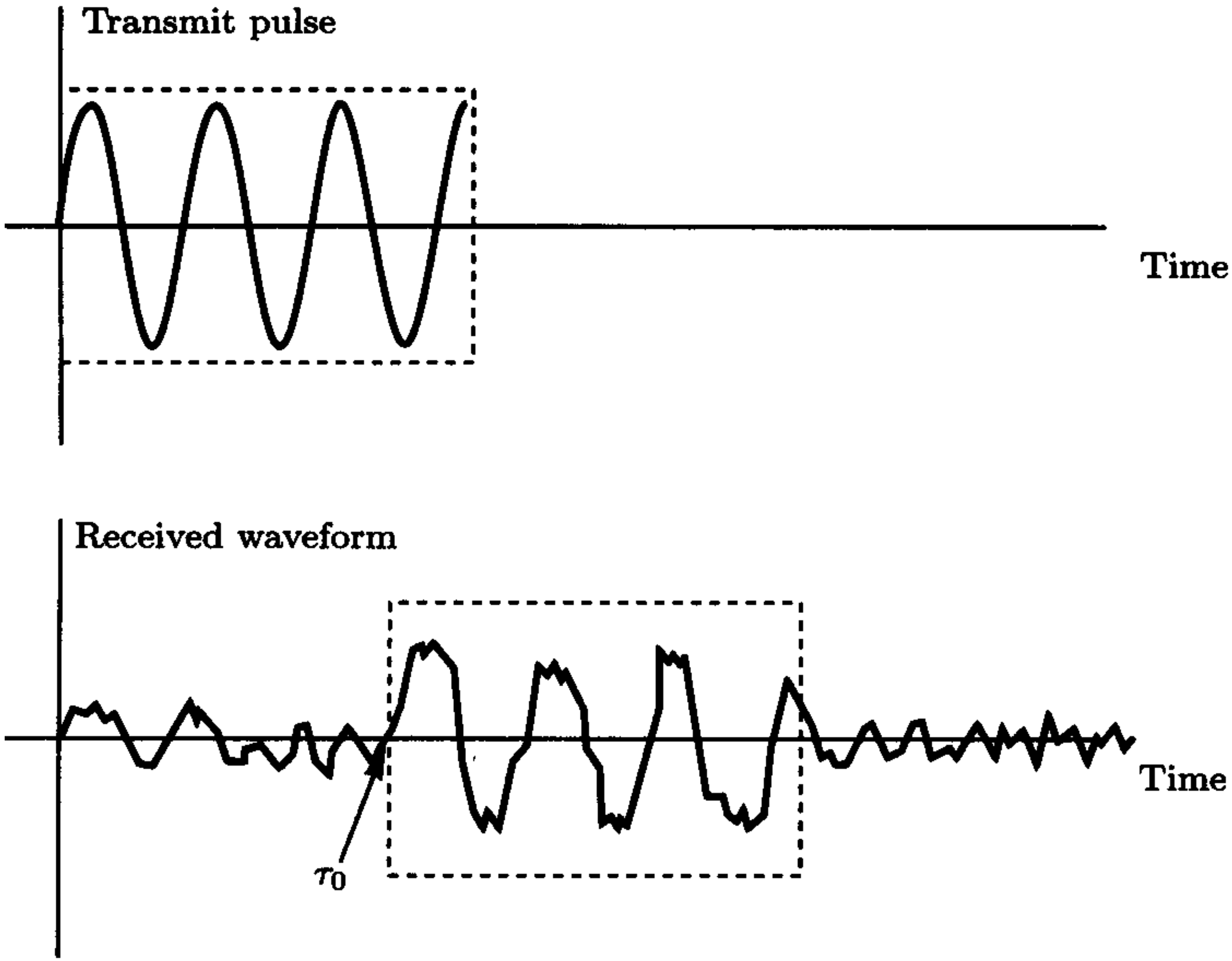


Radar system transmits an electromagnetic pulse $s(n)$. It is reflected by an aircraft, causing an echo $r(n)$ to be received after τ_0 seconds:

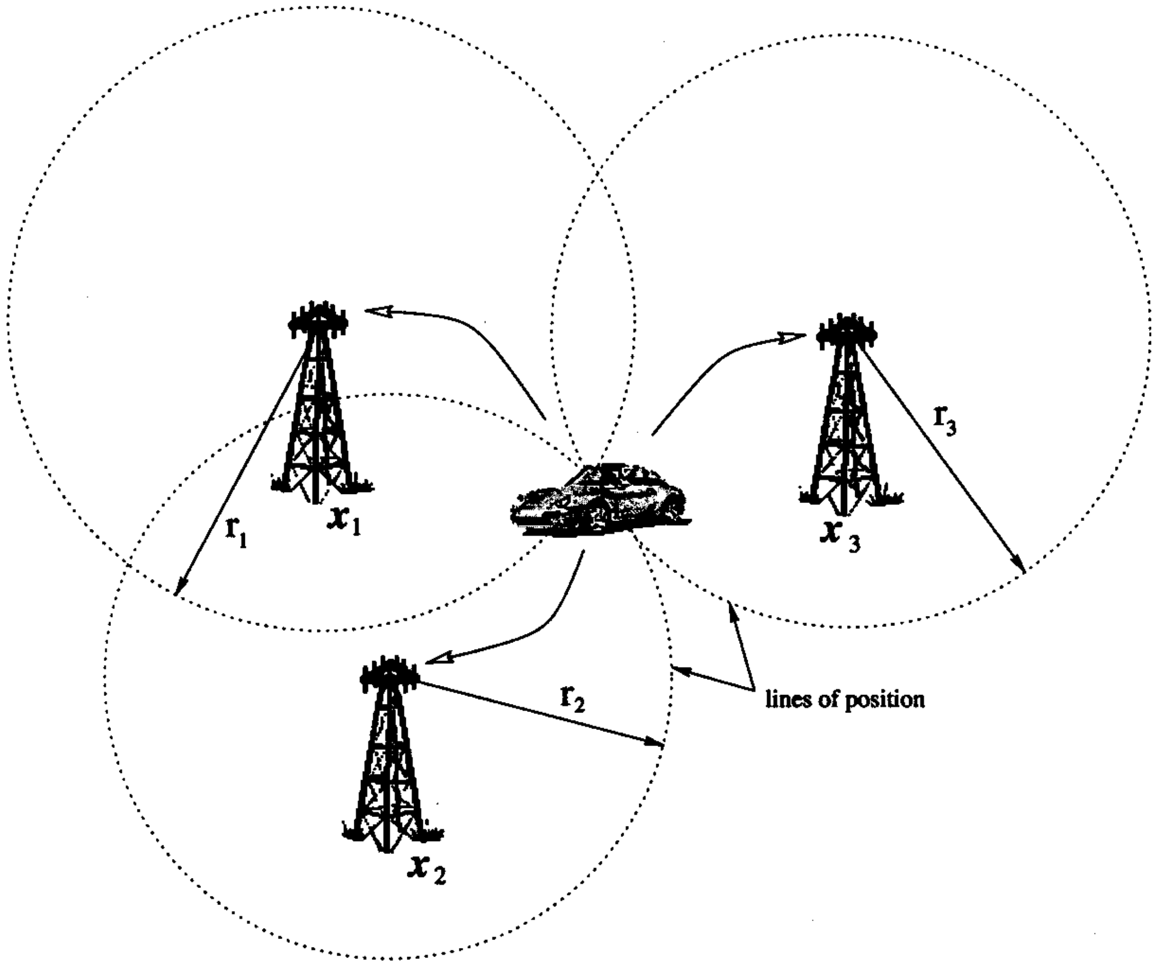
$$r(n) = \alpha s(n - \tau_0) + w(n)$$

where the range R of the aircraft is related to the time delay by

$$\tau_0 = 2R / c$$


Transmit pulse
Time
Received waveform
Time
τ_0

2. Mobile Communications



The position of the mobile terminal can be estimated using the time-of-arrival measurements received at the base stations.

3. Speech Processing

Recognition of human speech by a machine is a difficult task because our voice changes from time to time.

Given a human voice, the estimation problem is to determine the speech as close as possible.

4. Image Processing

Estimation of the position and orientation of an object from a camera image is useful when using a robot to pick it up, e.g., bomb-disposal

5. Biomedical Engineering

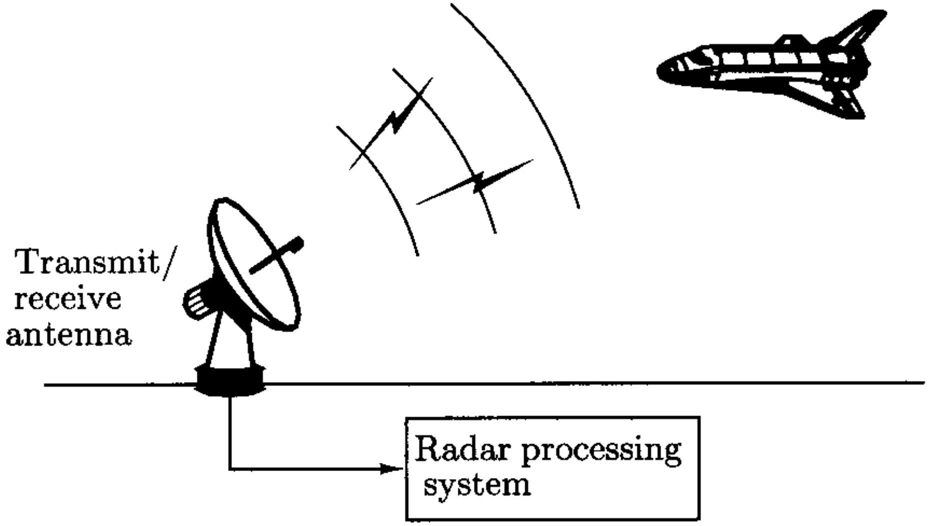
Estimation the heart rate of a fetus and the difficulty is that the measurements are corrupted by the mother's heart beat as well.

6. Seismology

Estimation of the underground distance of an oil deposit based on sound reflection due to the different densities of oil and rock layers.

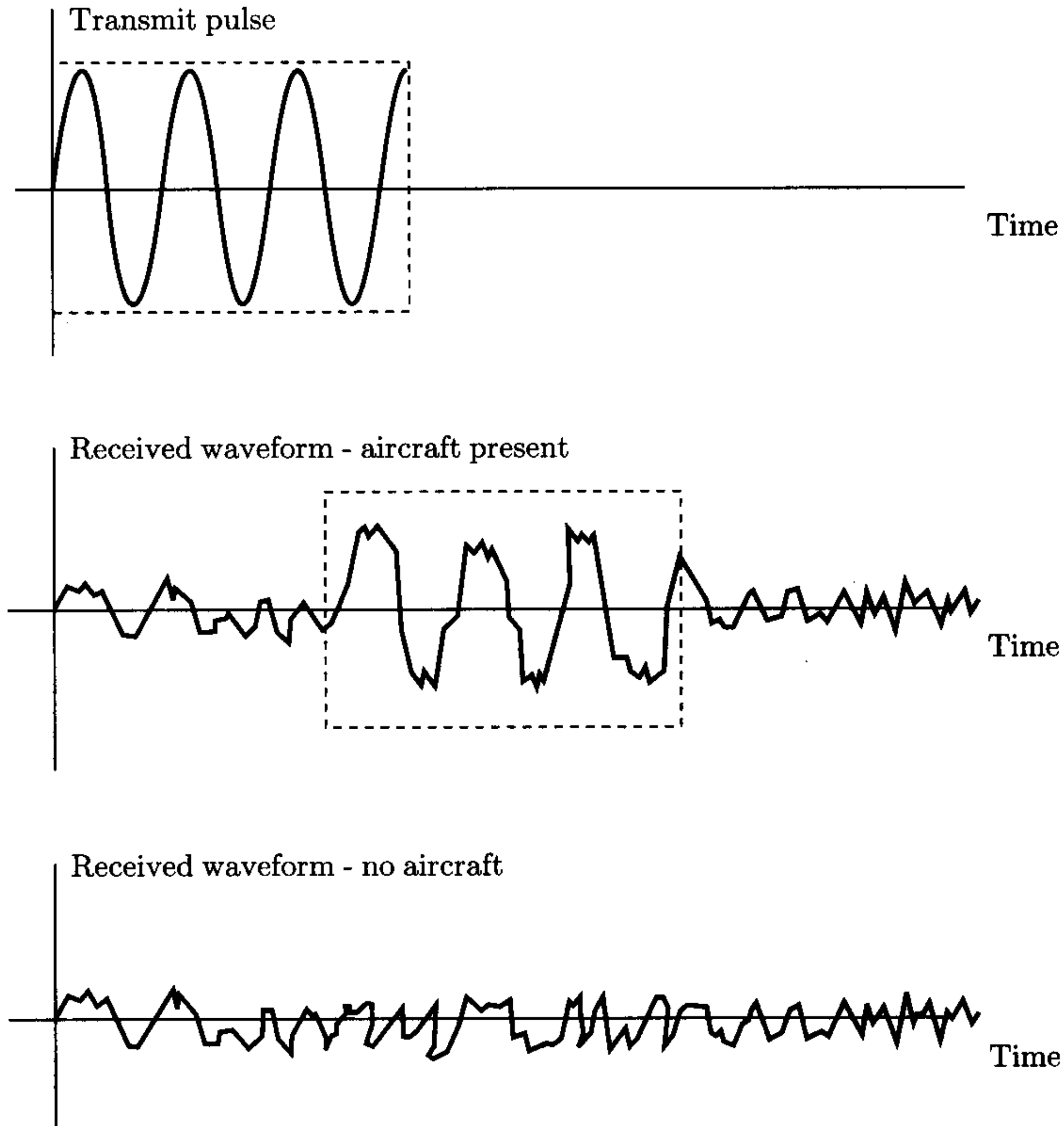
Differences from Detection

1. Radar



Radar system transmits an electromagnetic pulse $s(n)$. After some time, it receives a signal $r(n)$. The detection problem is to decide whether $r(n)$ is

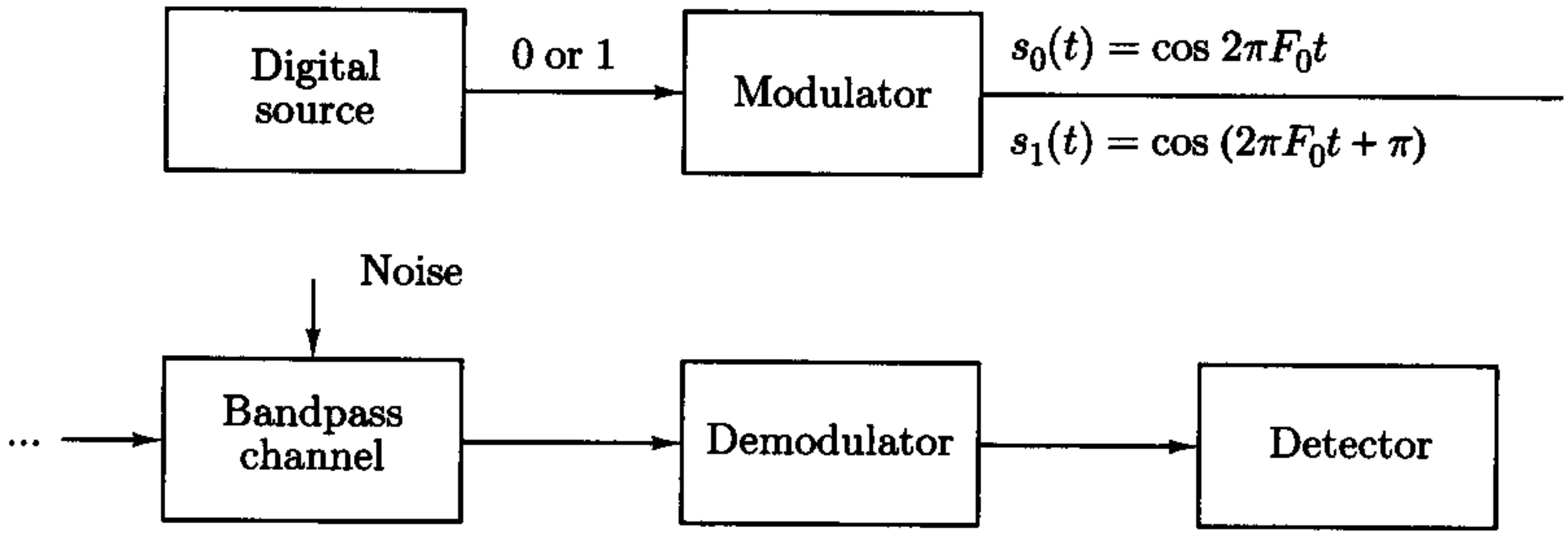
echo from an object or it is not an echo

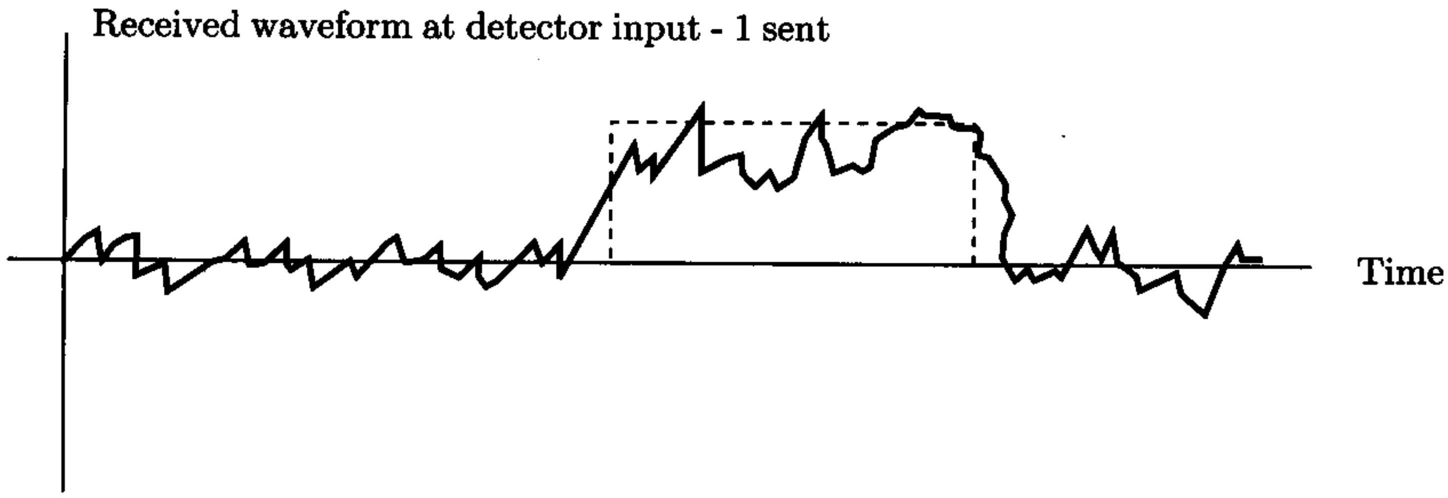

Transmit pulse
Time
Received waveform - aircraft present
Time
Received waveform - no aircraft
Time

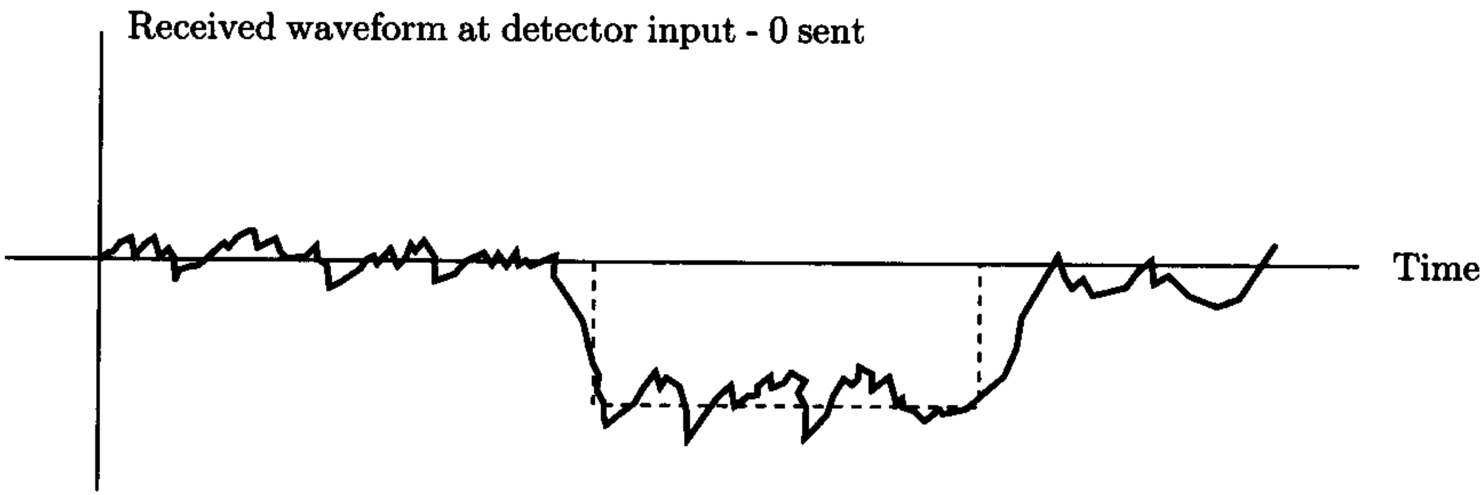
2. Communications

In wired or wireless communications, we need to know the information sent from the transmitter to the receiver.

e.g., for binary phase shift keying (BPSK) signals, it consists of only two symbols, “0” or “1”. The detection problem is to decide whether it is “0” or “1”.

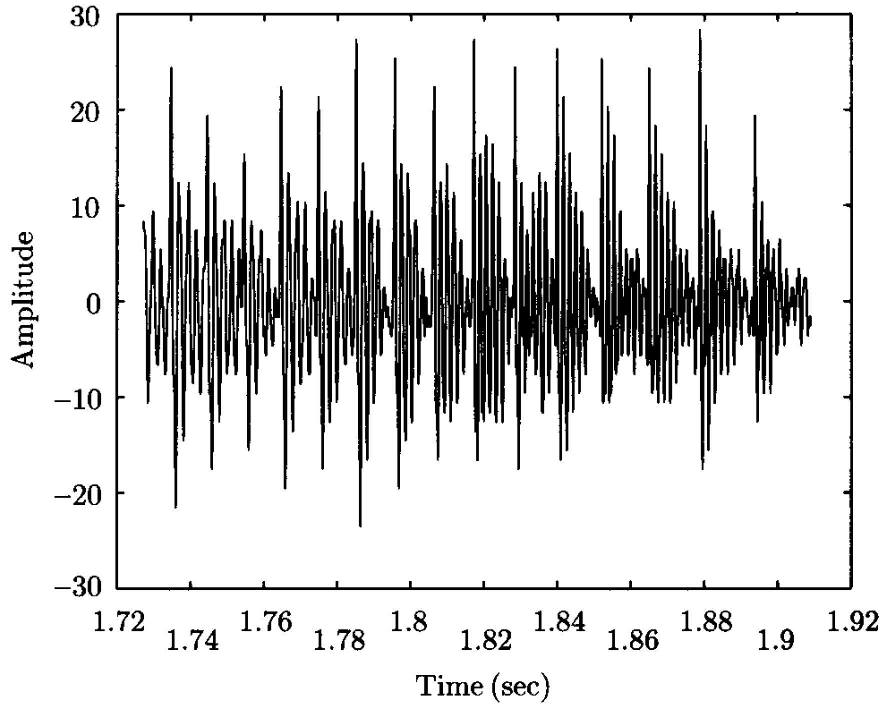



Received waveform at detector input - 1 sent
Time


Received waveform at detector input - 0 sent
Time

3. Speech Processing

Given a human speech signal, the detection problem is decide what is the spoken word from a set of predefined words, e.g., “0”, “1”,…, “9”



Waveform of “0”

Another example is voice authentication: given a voice and it is indicated that the voice is from George Bush, we need to decide it’s Bush or not.

4. Image Processing

Fingerprint authentication: given a fingerprint image and his owner says he is “A”, we need to verify if it is true or not

Other biometric examples include face authentication, iris authentication, etc.

5. Biomedical Engineering

X光造影驗乳癌 準確性低

每百宗陽性結果 86至98人「虛驚」

【本報記者關天燕報道】遲婚、遲生育、飲食西化，都是婦女罹患乳癌的誘因。根據香港癌病資料統計中心的數字顯示，目前每24名婦女中，便有一人患上乳癌，縱使患者不斷增加，但醫學界至今仍未找出預防乳癌的最佳方法，就連早前提倡的乳房X光造影檢查，也被評定準確性甚低。

香港大學醫學院社會醫學系助理教授梁卓偉指出，1973至99年間，乳癌發病率增加了37%，80歲或以上的婦女佔最大增長；30至49歲的發病率也有相當的增長；但50至79歲婦女的平均風險卻較低。研究亦發現1900年與1960年出生的婦女，發病率有兩倍的差異。

無理據證明 檢查可預防乳癌

是次研究亦探討患病婦女有否作定期檢查，研究結論顯示，無論患病增長率最多及最少組別人士，發病率亦相若，因此看不到有理據證明乳房X光造影檢查的服務，可以預防乳癌。

梁卓偉指出，全球性包括本港的研究得出，每100名接受乳房X光造影檢查後得出陽性結果的婦女中，有86至98名婦女化驗結果最終是假陽性。

由於陽性結果要進一步檢查，例如細針管抽吸術，甚至動手術切除活細胞組織，此舉會有一成機會產生傷口感染，亦對婦女造成不必要擔憂。

50歲內未收經婦女 準確度更低

香港癌病資料統計中心主管傅惠霖指出，雖然乳房X光造影是唯一能預早驗出乳癌的方法，但不是減低乳癌死亡率的最好辦法，尤其是50歲以下還未「收經」的婦女，由於乳房密度高，X光照不通透，準確性更低，假陽性及假陰性結果出現的機會率亦高。

港大醫學院社會醫學系主任林大慶表示，本港社會及生活模式轉變，包括遲婚、低出生率、遲生育、較豐裕及西化的生活習慣，加上運動量少及體重上升，都是患上乳癌主因。

目前本港沒有全民乳房X光造影檢查，但新加坡即將會推行，由於華人患乳癌的風險比西方女性低，所以醫學界都質疑新加坡的做法。

另外，傅惠霖稱30至60歲婦女，應提高乳癌意識及警覺性，經常自我檢查乳房及多了解身體狀況。

經乳房造影檢查後，若發現結果為陽性，需進一步接受細針管抽吸術檢查，但會引致併發症。圖為一女士接受檢查的情況。（資料圖片）

港大醫學院社會醫學系助理教授梁卓偉稱，研究證實乳房X光造影檢查準確性低，因此沒有理據擴展乳房X光造影檢查服務。（車耀開攝）

本港診斷乳癌方法
- 乳房X光造影
- 乳房超音波
- 紅外綫檢查
- 細針管抽吸術
- 切除活細胞組織化驗
- 自我檢查

資料來源：港大醫學院及各醫療組織

17 Jan. 2003, Hong Kong Economics Times

e.g., given some X-ray slides, the detection problem is to determine if she has breast cancer or not

6. Seismology

To detect if there is oil or there is no oil at a region

What is Estimation?

Extract or estimate some parameters from the observed signals, e.g.,

- Use a voltmeter to measure a DC signal

$$x[n] = A + w[n], \qquad n = 0,1,\cdots,N-1$$

Given $x[n]$, we need to find the DC value, A

$\Rightarrow$ the parameter is the observed signal

- Estimate the amplitude, frequency and phase of a sinusoid in noise

$$x[n] = \alpha\cos(\omega n + \phi) + w[n], \qquad n = 0,1,\cdots,N-1$$

Given $x[n]$, we need to find α, ω and ϕ

$\Rightarrow$ the parameters are not directly observed in the received signal

- Estimate the value of resistance R from a set of voltage and current readings:

$$V[n] = V_{\text{actual}}[n] + w_1[n], \qquad I[n] = I_{\text{actual}}[n] + w_2[n], \quad n = 0,1,\cdots,N-1$$

Given N pairs of $(V[n], I[n])$, we need to estimate the resistance R, ideally, $R = V / I$

$\Rightarrow$ the parameter is not directly observed in the received signals

- Estimate the position of the mobile terminal using time-of-arrival measurements:

$$r[n] = \frac{\sqrt{(x_s - x_n)^2 - (y_s - y_n)^2}}{c} + w[n], \qquad n = 0,1,\cdots,N-1$$

Given $r[n]$, we need to find the mobile position (x_s, y_s) where c is the signal propagation speed and (x_n, y_n) represent the known position of the nth base station

$\Rightarrow$ the parameters are not directly observed in the received signals

Types of Parameter Estimation

- Linear or non-linear

  Linear: DC value, amplitude of the sine wave
  Non-linear: Frequency of the sine wave, mobile position

- Single parameter or multiple parameters

  Single: DC value; scalar
  Multiple: Amplitude, frequency and phase of sinusoid; vector

- Constrained or unconstrained

  Constrained: Use other available information & knowledge, e.g., from the N pairs of $(V[n], I[n])$, we draw a line which best fits the data points and the estimate of the resistance is given by the slope of the line. We can add a constraint that the line should cross the origin (0,0)
  Unconstrained: No further information & knowledge is available

- Parameter is unknown deterministic or random

Unknown deterministic:	constant but unknown (classical) DC value is an unknown constant
Random :	random variable with prior knowledge of PDF (Bayesian) If we have prior knowledge that the DC value is bounded by $-A_0$ and A_0 with a particular PDF $\Rightarrow$ better estimate

- Parameter is stationary or changing

Stationary :	Unknown deterministic for whole observation period, time-of-arrivals of a static source
Changing :	Unknown deterministic at different time instants, time-of-arrivals of a moving source

Performance Measures for Classical Parameter Estimation

Accuracy:

- Is the estimator biased or unbiased?

e.g., $x[n] = A + w[n], \qquad n = 0,1,\cdots,N-1$

where $w[n]$ is a zero-mean random noise with variance σ_w^2

Proposed estimators:

$$\hat{A}_1 = x[0]$$

$$\hat{A}_2 = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$$

$$\hat{A}_3 = \frac{1}{N-1}\sum_{n=0}^{N-1} x[n]$$

$$\hat{A}_4 = \sqrt[N]{\prod_{n=0}^{N-1} x[n]} = \sqrt[N]{x[0]\cdot x[1]\cdots x[N-1]}$$

Biased : $E\{\hat{A}\} \neq A$

Unbiased : $E\{\hat{A}\} = A$

Asymptotically unbiased : $E\{\hat{A}\} = A$ only if $N \to \infty$

Taking the expected values for $\hat{A}_1$, $\hat{A}_2$ and $\hat{A}_3$, we have

$$E\{\hat{A}_1\} = E\{x[0]\} = E\{A\} + E\{w[0]\} = A + 0 = A$$

$$E\{\hat{A}_2\} = E\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right\} = E\left\{\frac{1}{N}\sum_{n=0}^{N-1} A\right\} + E\left\{\frac{1}{N}\sum_{n=0}^{N-1} w[n]\right\}$$

$$= \frac{1}{N}\sum_{n=0}^{N-1} A + \frac{1}{N}\sum_{n=0}^{N-1} E\{w[n]\} = \frac{1}{N}\cdot N \cdot A + \frac{1}{N}\sum_{n=0}^{N-1} 0 = A$$

$$E\{\hat{A}_3\} = \frac{N}{N-1}\cdot A = \frac{1}{1-1/N}\cdot A$$

Q. State the biasedness of $\hat{A}_1$, $\hat{A}_2$ and $\hat{A}_3$.

For $\hat{A}_4$, it is difficult to analyze the biasedness. However, for $w[n] = 0$:

$$\sqrt[N]{x[0] \cdot x[1] \cdots x[N-1]} = \sqrt[N]{A \cdot A \cdots A} = \sqrt[N]{A^N} = A$$

- What is the value of the mean square error or variance?

  They correspond to the fluctuation of the estimate in the second order:

$$\text{MSE} = E\{(\hat{A} - A)^2\} \qquad (5.1)$$

$$\text{var} = E\{(\hat{A} - E\{\hat{A}\})^2\} \qquad (5.2)$$

:

  If the estimator is unbiased, then MSE = var

In general,

$$\begin{aligned}\text{MSE} &= E\{(\hat{A}-A)^2\} = E\{(\hat{A}-E\{\hat{A}\}+E\{\hat{A}\}-A)^2\} \\ &= E\{(\hat{A}-E\{\hat{A}\})^2\} + E\{(E\{\hat{A}\}-A)^2\} + 2E\{(\hat{A}-E\{\hat{A}\})(E\{\hat{A}\}-A)\} \\ &= \text{var} + (E\{\hat{A}\}-A)^2 + 2(E\{\hat{A}\}-E\{\hat{A}\})(E\{\hat{A}\}-A) \\ &= \text{var} + (\text{bias})^2\end{aligned} \tag{5.3}$$

$$E\{(\hat{A}_1-A)^2\} = E\{(x[0]-A)^2\} = E\{(A+w[0]-A)^2\} = E\{w^2[0]\} = \sigma_w^2$$

$$E\{(\hat{A}_2-A)^2\} = E\left\{\left(\frac{1}{N}\sum_{n=0}^{N-1}x[n]-A\right)^2\right\} = \frac{1}{N}E\left\{\sum_{n=0}^{N-1}w^2[n]\right\} = \frac{\sigma_w^2}{N}$$

$$E\{(\hat{A}_3-A)^2\} = E\left\{\left(\frac{1}{N-1}\sum_{n=0}^{N-1}x[n]-A\right)^2\right\} = \left(\frac{A}{N-1}\right)^2 + \frac{\sigma_w^2}{N-1}$$

An optimum estimator should give estimates which are

- Unbiased
- Minimum variance (MSE as well)

Q. How do we know the estimator has the minimum variance?

Cramer-Rao Lower Bound (CRLB)

Performance bound in terms of minimum achievable variance provided by any unbiased estimators

Use for classical parameter estimation

Require knowledge of the noise PDF and the PDF must have closed form

More easier to determine than other variance bounds

Let the parameters to be estimated be $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_P]^T$, the CRLB for θ_i in Gaussian noise is stated as follows

$$\mathrm{CRLB}(\theta_i) = [\mathbf{J}(\boldsymbol{\theta})]_{i,i} = \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{i,i} \tag{5.4}$$

where

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} -E\left\{\dfrac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_1^2}\right\} & -E\left\{\dfrac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2}\right\} & \cdots & -E\left\{\dfrac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_P}\right\} \\ -E\left\{\dfrac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_2 \partial \theta_1}\right\} & -E\left\{\dfrac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_2^2}\right\} & & \\ \vdots & & \ddots & \\ -E\left\{\dfrac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_P \partial \theta_1}\right\} & & & -E\left\{\dfrac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_P^2}\right\} \end{bmatrix} \tag{5.5}$$

$p(\mathbf{x};\boldsymbol{\theta})$ represents PDF of $\mathbf{x}=[x[0],x[1],\cdots,x[N-1]]^T$ and it is parameterized by the unknown parameter vector $\boldsymbol{\theta}$

Note that

- $\mathbf{I}(\boldsymbol{\theta})$ is known as Fisher information matrix

- $[\mathbf{J}]_{i,i}$ is the (i,i) element of $\mathbf{J}$

  e.g., $\mathbf{J}=\begin{bmatrix}1 & 2\\ 2 & 3\end{bmatrix} \Rightarrow [\mathbf{J}]_{2,2}=3$

- $E\left\{\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_i\,\partial\theta_j}\right\}=E\left\{\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_j\,\partial\theta_i}\right\}$

Review of Gaussian (Normal) Distribution

The Gaussian PDF for a scalar random variable x is defined as

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 \right) \tag{5.6}$$

We can write $x \sim N(\mu, \sigma)$

The Gaussian PDF for a random vector $\mathbf{x}$ of size N is defined as

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} \det^{1/2}(\mathbf{C})} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \cdot \mathbf{C}^{-1} \cdot (\mathbf{x}-\boldsymbol{\mu}) \right) \tag{5.7}$$

We can write $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{C})$

The covariance matrix $\mathbf{C}$ has the form of

$$\mathbf{C} = E\{(\mathbf{x} - \boldsymbol{\mu}) \cdot (\mathbf{x} - \boldsymbol{\mu})^T\}$$

$$= \begin{bmatrix} E\{(x[0]-\mu_0)^2\} & \cdots & E\{(x[0]-\mu_0)(x[N-1]-\mu_{N-1})\} \\ E\{(x[0]-\mu_0)(x[1]-\mu_1)\} & \ddots & \vdots \\ \vdots & & \\ E\{(x[0]-\mu_0)(x[N-1]-\mu_{N-1})\} & \cdots & E\{(x[N-1]-\mu_{N-1})^2\} \end{bmatrix} \tag{5.8}$$

where

$$\mathbf{x} = [x[0], x[1], \cdots, x[N-1]]^T$$

$$\boldsymbol{\mu} = E\{\mathbf{x}\} = [\mu_0, \mu_1, \cdots, \mu_{N-1}]^T$$

If $\mathbf{x}$ is a zero-mean white vector and all vector elements have variance σ^2

$$\mathbf{C} = E\{(\mathbf{x}-\boldsymbol{\mu})\cdot(\mathbf{x}-\boldsymbol{\mu})^T\} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \cdot \mathbf{I}_N$$

The Gaussian PDF for the random vector $\mathbf{x}$ can be simplified as

$$p(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right) \qquad (5.9)$$

with the use of

$$\mathbf{C}^{-1} = \sigma^{-2} \cdot \mathbf{I}_N$$
$$\det(\mathbf{C}) = (\sigma^2)^N = \sigma^{2N}$$

Example 5.1

Determine the PDF of

$$x[0] = A + w[0]$$

and

$$x[n] = A + w[n], \qquad n = 0,1,\cdots,N-1$$

where $\{w(n)\}$ is a white Gaussian process with known variance σ_w^2 and A is a constant

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

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

Example 5.2

Find the CRLB for estimating A based on single measurement:

$$x[0] = A + w[0]$$

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

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

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

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

As a result,

$$E\left\{\frac{\partial^2 \ln(p(x[0];A))}{\partial A^2}\right\} = -\frac{1}{\sigma_w^2}$$

$$\mathbf{I}(A) = I(A) = \frac{1}{\sigma_w^2}$$

$$\Rightarrow J(A) = \sigma_w^2$$

$$\Rightarrow \mathrm{CRLB}(A) = \sigma_w^2$$

This means the best we can do is to achieve estimator variance = σ_w^2 or

$$\mathrm{var}(\hat{A}) \geq \sigma_w^2$$

where $\hat{A}$ is any unbiased estimator for estimating A

We also observe that a simple unbiased estimator

$$\hat{A}_1 = x[0]$$

achieves the CRLB:

$$E\{(\hat{A}_1 - A)^2\} = E\{(x[0]-A)^2\} = E\{(A+w[0]-A)^2\} = E\{w^2[0]\} = \sigma_w^2$$

<u>Example 5.3</u>

Find the CRLB for estimating A based on N measurements:

$$x[n] = A + w[n], \qquad n = 0,1,\cdots,N-1$$

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

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

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

$$\Rightarrow \frac{\partial \ln(p(\mathbf{x};A))}{\partial A}=-\frac{1}{2\sigma_w^2}\cdot 2\cdot\sum_{n=0}^{N-1}(x[n]-A)\cdot -1=\frac{\sum_{n=0}^{N-1}(x[n]-A)}{\sigma_w^2}$$

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

$$\Rightarrow E\left\{\frac{\partial^2 \ln(p(\mathbf{x};A))}{\partial A^2}\right\}=-\frac{N}{\sigma_w^2}$$

As a result,

$$\mathbf{I}(A) = I(A) = \frac{N}{\sigma_w^2}$$

$$\Rightarrow J(A) = \frac{\sigma_w^2}{N}$$

$$\Rightarrow \mathrm{CRLB}(A) = \frac{\sigma_w^2}{N}$$

This means the best we can do is to achieve estimator variance = σ_w^2 / N or

$$\mathrm{var}(\hat{A}) \geq \frac{\sigma_w^2}{N}$$

where $\hat{A}$ is any unbiased estimator for estimating A

We also observe that a simple unbiased estimator

$$\hat{A}_1 = x[0]$$

does not achieve the CRLB

$$E\{(\hat{A}_1 - A)^2\} = E\{(x[0] - A)^2\} = E\{(A + w[0] - A)^2\} = E\{w^2[0]\} = \sigma_w^2$$

On the other hand, the sample mean estimator

$$\hat{A}_2 = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

achieve the CRLB

$$E\{(\hat{A}_2 - A)^2\} = E\left\{\left(\frac{1}{N} \sum_{n=0}^{N-1} x[n] - A\right)^2\right\} = \frac{1}{N} E\left\{\sum_{n=0}^{N-1} w^2[n]\right\} = \frac{\sigma_w^2}{N}$$

$\Rightarrow$ sample mean is the optimum estimator for white Gaussian noise

<u>Example 5.4</u>

Find the CRLB for A and σ_w^2 given $\{x[n]\}$:

$$x[n] = A + w[n], \qquad n = 0,1,\cdots,N-1$$

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma_w^2)^{N/2}} \exp\left(-\frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} (x[n]-A)^2 \right), \qquad \boldsymbol{\theta} = [A, \sigma_w^2]$$

$$\Rightarrow \ln(p(\mathbf{x};\boldsymbol{\theta})) = -\ln((2\pi\sigma_w^2)^{N/2}) - \frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} (x[n]-A)^2$$

$$= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_w^2) - \frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} (x[n]-A)^2$$

$$\Rightarrow \frac{\partial \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial A} = -\frac{1}{2\sigma_w^2} \cdot 2 \cdot \sum_{n=0}^{N-1} (x[n]-A) \cdot -1 = \frac{\sum_{n=0}^{N-1} (x[n]-A)}{\sigma_w^2}$$

$$\Rightarrow \frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial A^2} = -\frac{N}{\sigma_w^2}$$

$$\Rightarrow E\left\{\frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial A^2}\right\} = -\frac{N}{\sigma_w^2}$$

$$\Rightarrow \frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial A \partial \sigma_w^2} = -\frac{\sum_{n=0}^{N-1}(x[n]-A)}{\sigma_w^4} = -\frac{\sum_{n=0}^{N-1}(w[n])}{\sigma_w^4}$$

$$\Rightarrow E\left\{\frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial A \partial \sigma_w^2}\right\} = -\frac{\sum_{n=0}^{N-1}(E\{w[n]\})}{\sigma_w^4} = 0$$

$$\ln(p(\mathbf{x};\boldsymbol{\theta})) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_w^2) - \frac{1}{2\sigma_w^2}\sum_{n=0}^{N-1}(x[n]-A)^2$$

$$\Rightarrow \frac{\partial \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial \sigma_w^2} = -\frac{N}{2\sigma_w^2} + \frac{1}{2\sigma_w^4}\sum_{n=0}^{N-1}(x[n]-A)^2$$

$$\Rightarrow \frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial (\sigma_w^2)^2} = \frac{N}{2\sigma_w^4} - \frac{1}{\sigma_w^6}\sum_{n=0}^{N-1}(w[n])^2$$

$$\Rightarrow E\left\{\frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial (\sigma_w^2)^2}\right\} = \frac{N}{2\sigma_w^4} - \frac{1}{\sigma_w^6}\cdot N\sigma_w^2 = -\frac{N}{2\sigma_w^4}$$

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{N}{\sigma_w^2} & 0 \\ 0 & \frac{N}{2\sigma_w^4} \end{bmatrix}$$

$$\mathbf{J}(\boldsymbol{\theta}) = \mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\sigma_w^2}{N} & 0 \\ 0 & \frac{2\sigma_w^4}{N} \end{bmatrix}$$

$$\Rightarrow \mathrm{CRLB}(A) = \frac{\sigma_w^2}{N}$$

$$\Rightarrow \mathrm{CRLB}(\sigma_w^2) = \frac{2\sigma_w^4}{N}$$

$\Rightarrow$ the CRLBs for unknown and known noise power are identical

Q. The CRLB is not affected by knowledge of noise power. Why?

Q. Can you suggest a method to estimate σ_w^2 ?

Example 5.5

Find the CRLB for phase of a sinusoid in white Gaussian noise:

$$x[n] = A\cos(\omega_0 n + \phi) + w[n], \qquad n = 0,1,\cdots,N-1$$

where A and ω_0 are assumed known

The PDF is

$$p(\mathbf{x};\phi) = \frac{1}{(2\pi\sigma_w^2)^{N/2}} \exp\left(-\frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} (x[n] - A\cos(\omega_0 n + \phi))^2 \right)$$

$$\Rightarrow \ln(p(\mathbf{x};\phi)) = -\ln((2\pi\sigma_w^2)^{N/2}) - \frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} (x[n] - A\cos(\omega_0 n + \phi))^2$$

$$\frac{\partial \ln(p(\mathbf{x};\phi))}{\partial \phi} = -\frac{1}{2\sigma_w^2}\sum_{n=0}^{N-1} 2(x[n] - A\cos(\omega_0 n + \phi)) \cdot -A \cdot -\sin(\omega_0 n + \phi)$$

$$= -\frac{A}{\sigma_w^2}\sum_{n=0}^{N-1}\left[x[n]\sin(\omega_0 n + \phi) - \frac{A}{2}\sin(2\omega_0 n + 2\phi)\right]$$

$$\frac{\partial^2 \ln(p(\mathbf{x};\phi))}{\partial \phi^2} = -\frac{A}{\sigma_w^2}\sum_{n=0}^{N-1}\left[x[n]\cos(\omega_0 n + \phi) - A\cos(2\omega_0 n + 2\phi)\right]$$

$$E\left\{\frac{\partial^2 \ln(p(\mathbf{x};\phi))}{\partial \phi^2}\right\} = -\frac{A}{\sigma_w^2}\sum_{n=0}^{N-1}\left[A\cos(\omega_0 n + \phi) \cdot \cos(\omega_0 n + \phi) - A\cos(2\omega_0 n + 2\phi)\right]$$

$$= -\frac{A^2}{\sigma_w^2}\sum_{n=0}^{N-1}\left[\cos^2(\omega_0 n + \phi) - \cos(2\omega_0 n + 2\phi)\right]$$

$$= -\frac{A^2}{\sigma_w^2}\sum_{n=0}^{N-1}\left[\frac{1}{2} + \frac{1}{2}\cos(2\omega_0 n + 2\phi) - \cos(2\omega_0 n + 2\phi)\right]$$

$$E\left\{\frac{\partial^2 \ln(p(\mathbf{x};\phi))}{\partial\phi^2}\right\} = -\frac{A^2}{\sigma_w^2}\cdot\frac{N}{2} + \frac{A^2}{2\sigma_w^2}\sum_{n=0}^{N-1}\cos(2\omega_0 n + 2\phi)$$

$$= -\frac{NA^2}{2\sigma_w^2} + \frac{A^2}{2\sigma_w^2}\sum_{n=0}^{N-1}\cos(2\omega_0 n + 2\phi)$$

As a result,

$$\text{CRLB}(\phi) = \left[\frac{NA^2}{2\sigma_w^2} - \frac{A^2}{2\sigma_w^2}\sum_{n=0}^{N-1}\cos(2\omega_0 n + 2\phi)\right]^{-1} = \frac{2\sigma_w^2}{NA^2}\left[1 - \frac{1}{N}\sum_{n=0}^{N-1}\cos(2\omega_0 n + 2\phi)\right]^{-1}$$

If $N >> 1$,
$$\frac{1}{N}\sum_{n=0}^{N-1}\cos(2\omega_0 n + 2\phi) \approx 0$$

then

$$\text{CRLB}(\phi) \approx \frac{2\sigma_w^2}{NA^2}$$

Example 5.6

Find the CRLB for A, ω_0 and ϕ for

$$x[n] = A\cos(\omega_0 n + \phi) + w[n], \qquad n = 0,1,\cdots,N-1, \qquad N >> 1$$

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma_w^2)^{N/2}} \exp\left(-\frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} (x[n] - A\cos(\omega_0 n + \phi))^2 \right), \qquad \boldsymbol{\theta} = [A, \omega_0, \phi]$$

$$\Rightarrow \ln(p(\mathbf{x};\boldsymbol{\theta})) = -\ln((2\pi\sigma_w^2)^{N/2}) - \frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} (x[n] - A\cos(\omega_0 n + \phi))^2$$

$$\Rightarrow \frac{\partial \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial A} = -\frac{1}{2\sigma_w^2} \cdot 2 \cdot \sum_{n=0}^{N-1} (x[n] - A\cos(\omega_0 n + \phi)) \cdot -\cos(\omega_0 n + \phi)$$

$$= \frac{1}{\sigma_w^2} \sum_{n=0}^{N-1} (x[n]\cos(\omega_0 n + \phi) - A\cos^2(\omega_0 n + \phi))$$

$$\frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial A^2} = -\frac{1}{\sigma_w^2}\sum_{n=0}^{N-1}\cos^2(\omega_0 n + \phi) = -\frac{1}{\sigma_w^2}\sum_{n=0}^{N-1}\left[\frac{1}{2}+\frac{1}{2}\cos(2\omega_0 n + 2\phi)\right]$$

$$\approx -\frac{N}{2\sigma_w^2}$$

$$E\left\{\frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial A^2}\right\} \approx -\frac{N}{2\sigma_w^2}$$

Similarly,

$$E\left\{\frac{\partial^2 \ln(p(\mathrm{x};\theta))}{\partial A \partial \omega_0}\right\} = \frac{A}{2\sigma_w^2}\sum_{n=0}^{N-1} n\sin(2\omega_0 n + 2\phi) \approx 0$$

$$E\left\{\frac{\partial^2 \ln(p(\mathrm{x};\theta))}{\partial A \partial \phi}\right\} = \frac{A}{2\sigma_w^2}\sum_{n=0}^{N-1} \sin(2\omega_0 n + 2\phi) \approx 0$$

$$E\left\{\frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial \omega_0{}^2}\right\} = -\frac{A^2}{\sigma_w^2}\sum_{n=0}^{N-1} n^2\left(\frac{1}{2}-\frac{1}{2}\cos(2\omega_0 n + 2\phi)\right) \approx -\frac{A^2}{2\sigma_w^2}\sum_{n=0}^{N-1} n^2$$

$$E\left\{\frac{\partial^2 \ln(p(\mathrm{x};\theta))}{\partial \omega_0 \partial \phi}\right\} = -\frac{A^2}{\sigma_w^2}\sum_{n=0}^{N-1} n \sin^2(\omega_0 n + \phi) \approx -\frac{A^2}{2\sigma_w^2}\sum_{n=0}^{N-1} n$$

$$E\left\{\frac{\partial^2 \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial \phi^2}\right\} = -\frac{A^2}{\sigma_w^2}\sum_{n=0}^{N-1} \sin^2(\omega_0 n + \phi) \approx -\frac{NA^2}{2\sigma_w^2}$$

$$\mathbf{I}(\boldsymbol{\theta}) \approx \frac{1}{\sigma_w^2}\begin{bmatrix} \frac{N}{2} & 0 & 0 \\ 0 & \frac{A^2}{2}\sum_{n=0}^{N-1} n^2 & \frac{A^2}{2}\sum_{n=0}^{N-1} n \\ 0 & \frac{A^2}{2}\sum_{n=0}^{N-1} n & \frac{NA^2}{2} \end{bmatrix}$$

After matrix inversion, we have

$$\mathrm{CRLB}(A) \approx \frac{2\sigma_w^2}{N}$$

$$\mathrm{CRLB}(\omega_0) \approx \frac{12}{\mathrm{SNR} \cdot N(N^2 - 1)}, \qquad \mathrm{SNR} = \frac{A^2}{2\sigma_w^2}$$

$$\mathrm{CRLB}(\phi) \approx \frac{2(2N-1)}{\mathrm{SNR} \cdot N(N+1)}$$

Note that

$$\mathrm{CRLB}(\phi) \approx \frac{2(2N-1)}{\mathrm{SNR} \cdot N(N+1)} \approx \frac{4}{\mathrm{SNR} \cdot N} > \frac{1}{\mathrm{SNR} \cdot N} = \frac{2\sigma_w^2}{NA}$$

$\Rightarrow$ In general, the CRLB increases as the number of parameters to be estimated increases

$\Rightarrow$ CRLB decreases as the number of samples increases

Parameter Transformation in CRLB

Find the CRLB for $\boldsymbol{\alpha} = g(\boldsymbol{\theta})$ where $g()$ is a function

e.g., $x[n] = A + w[n], \qquad n = 0,1,\cdots,N-1$

What is the CRLB for A^2?

The CRLB for parameter transformation of $\alpha = g(\theta)$ is given by

$$\mathrm{CRLB}(\alpha) = \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{-E\left\{\frac{\partial^2 \ln(p(\mathbf{x};\theta))}{\partial \theta^2}\right\}} \tag{5.10}$$

For nonlinear function, "=" is replaced by "≈" and it is true only for large N

Example 5.7

Find the CRLB for the power of the DC value, i.e., A^2:

$$x[n] = A + w[n], \qquad n = 0,1,\cdots,N-1$$

$$\alpha = g(A) = A^2$$

$$\Rightarrow \frac{\partial g(A)}{\partial A} = 2A \Rightarrow \left(\frac{\partial g(A)}{\partial A}\right)^2 = 4A^2$$

From Example 5.3, we have

$$-E\left\{\frac{\partial^2 \ln(p(\mathbf{x};A))}{\partial A^2}\right\} = \frac{N}{\sigma_w^2}$$

As a result,

$$\mathrm{CRLB}(A^2) \approx 4A^2 \cdot \frac{\sigma_w^2}{N} = \frac{4A^2\sigma_w^2}{N}, \qquad N >> 1$$

Example 5.8

Find the CRLB for $\alpha = c_1 + c_2 A$ from

$$x[n] = A + w[n], \qquad n = 0,1,\cdots,N-1$$

$$\alpha = g(A) = c_1 + c_2 A$$

$$\Rightarrow \frac{\partial g(A)}{\partial A} = c_2 \Rightarrow \left(\frac{\partial g(A)}{\partial A}\right)^2 = c_2^2$$

As a result,

$$\text{CRLB}(\alpha) = c_2^2 \cdot \text{CRLB}(A) = c_2^2 \cdot \frac{\sigma_w^2}{N}$$

$$= \frac{c_2^2 \sigma_w^2}{N}$$

Maximum Likelihood Estimation

Parameter estimation is achieved via maximizing the likelihood function

Optimum realizable approach and can give performance close to CRLB

Use for classical parameter estimation

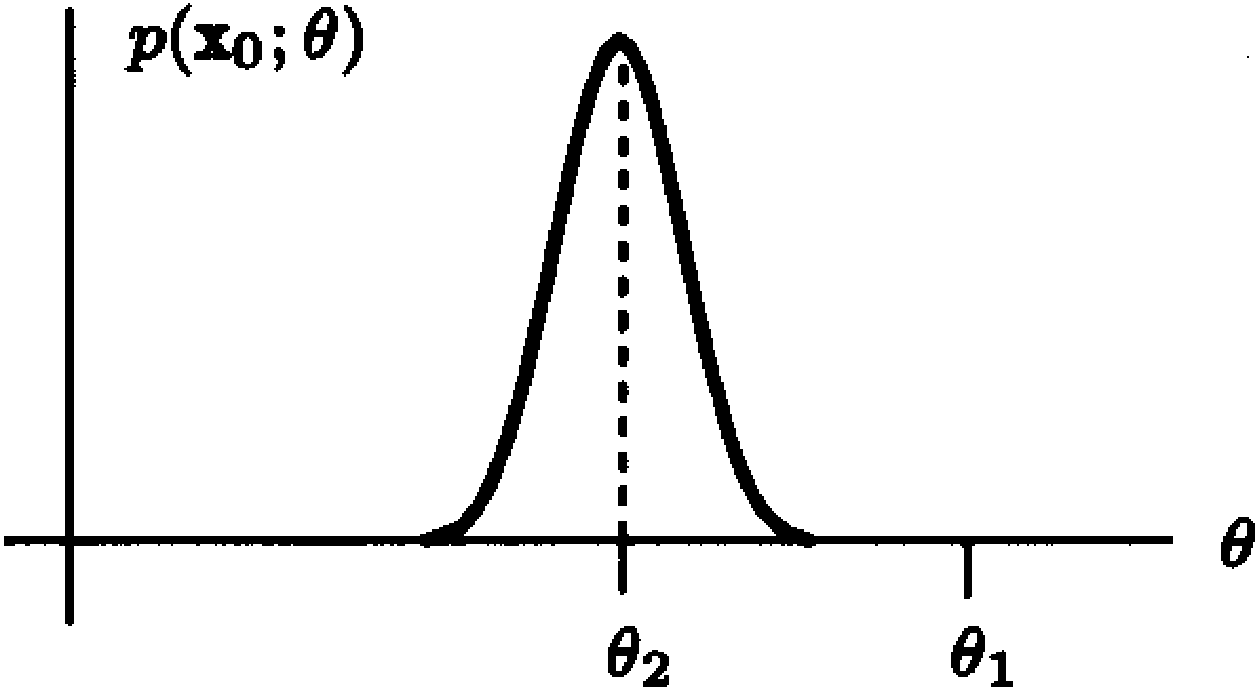
Require knowledge of the noise PDF and the PDF must have closed form

Generally computationally demanding

Let $p(\mathbf{x};\boldsymbol{\theta})$ be the PDF of the observed vector $\mathbf{x}$ parameterized by the parameter vector $\boldsymbol{\theta}$. The maximum likelihood (ML) estimate is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\mathbf{x};\boldsymbol{\theta}) \tag{5.11}$$

e.g., given $p(\mathbf{x} = \mathbf{x_0}; \theta)$ where $\mathbf{x_0}$ is the observed data, as below



Q. What is the most possible value of θ?

Example 5.9

Given

$$x[n] = A + w[n], \qquad n = 0,1,\cdots,N-1$$

where A is an unknown constant and $w[n]$ is a white Gaussian noise with known variance σ_w^2. Find the ML estimate of A.

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

Since $\arg\max_{\boldsymbol{\theta}} p(\mathbf{x};\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}} \{\ln(p(\mathbf{x};\boldsymbol{\theta}))\}$, taking log for $p(\mathbf{x};A)$ gives

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

Differentiate with respect to A yields

$$\frac{\partial \ln(p(\mathbf{x};A))}{\partial A} = -\frac{1}{2\sigma_w^2} \cdot 2 \cdot \sum_{n=0}^{N-1} (x[n]-A) \cdot -1 = \frac{\sum_{n=0}^{N-1} (x[n]-A)}{\sigma_w^2}$$

$\hat{A} = \arg\max_{A} \{\ln(p(\mathbf{x};A))\}$ is determined from

$$\frac{\sum_{n=0}^{N-1} (x[n]-\hat{A})}{\sigma_w^2} = 0 \Rightarrow \sum_{n=0}^{N-1} (x[n]-\hat{A}) = 0 \Rightarrow \hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

Note that

- ML estimate is identical to the sample mean
- Attain CRLB

Q. How about if σ_w^2 is unknown?

Example 5.10

Find the ML estimate for phase of a sinusoid in white Gaussian noise:

$$x[n] = A\cos(\omega_0 n + \phi) + w[n], \qquad n = 0,1,\cdots,N-1$$

where A and ω_0 are assumed known

The PDF is

$$p(\mathbf{x};\phi) = \frac{1}{(2\pi\sigma_w^2)^{N/2}} \exp\left(-\frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} (x[n] - A\cos(\omega_0 n + \phi))^2 \right)$$

$$\Rightarrow \ln(p(\mathbf{x};\phi)) = -\ln((2\pi\sigma_w^2)^{N/2}) - \frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} (x[n] - A\cos(\omega_0 n + \phi))^2$$

It is obvious that the maximum of $p(\mathbf{x};\phi)$ or $\ln(p(\mathbf{x};\phi))$ corresponds to the minimum of

$$\frac{1}{2\sigma_w^2}\sum_{n=0}^{N-1}(x[n]-A\cos(\omega_0 n+\phi))^2 \quad \text{or} \quad \sum_{n=0}^{N-1}(x[n]-A\cos(\omega_0 n+\phi))^2$$

Differentiating with respect to ϕ and then set the result to zero:

$$\sum_{n=0}^{N-1} 2(x[n]-A\cos(\omega_0 n+\phi))\cdot -A\cdot -\sin(\omega_0 n+\phi)$$

$$= A\sum_{n=0}^{N-1}\left[x[n]\sin(\omega_0 n+\phi)-\frac{A}{2}\sin(2\omega_0 n+2\phi)\right]=0$$

$$\Rightarrow \qquad \sum_{n=0}^{N-1} x[n]\sin(\omega_0 n+\hat{\phi}) = \frac{A}{2}\sum_{n=0}^{N-1}\sin(2\omega_0 n+2\hat{\phi})$$

The ML estimate for ϕ is determined from the root of the above equation

Q. Any ideas to solve the nonlinear equation?

Approximate ML (AML) solution may exist and it depends on the structure of the ML expression. For example, there exists an AML solution for ϕ

$$\sum_{n=0}^{N-1} x[n]\sin(\omega_0 n + \hat{\phi}) = \frac{A}{2}\sum_{n=0}^{N-1} \sin(2\omega_0 n + 2\hat{\phi})$$

$$\Rightarrow \frac{1}{N}\sum_{n=0}^{N-1} x[n]\sin(\omega_0 n + \hat{\phi}) = \frac{A}{2}\cdot\frac{1}{N}\sum_{n=0}^{N-1} \sin(2\omega_0 n + 2\hat{\phi}) \approx \frac{A}{2}\cdot 0 = 0, \qquad N >> 1$$

The AML solution is obtained from

$$\sum_{n=0}^{N-1} x[n]\sin(\omega_0 n + \hat{\phi}) = 0$$

$$\Rightarrow \sum_{n=0}^{N-1} x[n]\sin(\omega_0 n)\cos(\hat{\phi}) + \sum_{n=0}^{N-1} x[n]\cos(\omega_0 n)\sin(\hat{\phi}) = 0$$

$$\Rightarrow \cos(\hat{\phi})\cdot\sum_{n=0}^{N-1} x[n]\sin(\omega_0 n) = -\sin(\hat{\phi})\cdot\sum_{n=0}^{N-1} x[n]\cos(\omega_0 n)$$

$$\hat{\phi} = -\tan^{-1}\left(\frac{\sum_{n=0}^{N-1} x[n]\sin(\omega_0 n)}{\sum_{n=0}^{N-1} x[n]\cos(\omega_0 n)}\right)$$

In fact, the AML solution is reasonable:

$$\hat{\phi} = -\tan^{-1}\left(\frac{\sum_{n=0}^{N-1} (A\cos(\omega_0 n + \phi) + w[n])\sin(\omega_0 n)}{\sum_{n=0}^{N-1} (A\cos(\omega_0 n + \phi) + w[n])\cos(\omega_0 n)}\right)$$

$$\approx -\tan^{-1}\left(\frac{-\frac{NA}{2}\sin(\phi) + \sum_{n=0}^{N-1} w[n]\sin(\omega_0 n)}{\frac{NA}{2}\cos(\phi) + \sum_{n=0}^{N-1} w[n]\cos(\omega_0 n)}\right), \qquad N >> 1$$

$$= \tan^{-1}\left(\frac{\sin(\phi) - \frac{2}{NA}\sum_{n=0}^{N-1} w[n]\sin(\omega_0 n)}{\cos(\phi) + \frac{2}{NA}\sum_{n=0}^{N-1} w[n]\cos(\omega_0 n)}\right)$$

For parameter transformation, if there is a one-to-one relationship between $\alpha = g(\theta)$ and θ, the ML estimate for α is simply:

$$\hat{\alpha} = g(\hat{\theta}) \tag{5.12}$$

<u>Example 5.11</u>

Given N samples of a white Gaussian process $w[n]$, $n = 0,1,\cdots,N-1$, with unknown variance σ^2. Determine the power of $w[n]$ in dB.

The power in dB is related to σ^2 by

$$P = 10\log_{10}(\sigma^2)$$

which is a one-to-one relationship. To find the ML estimate for P, we first find the ML estimate for σ^2

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

$$\Rightarrow \ln(p(\mathbf{w};\sigma^2)) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n]$$

Differentiating the log-likelihood function w.r.t. to σ^2:

$$\frac{\partial \ln(p(\mathbf{w};\sigma^2))}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{n=0}^{N-1} x^2[n]$$

Setting the resultant expression to zero:

$$\frac{N}{2\hat{\sigma}^2} = \frac{1}{2\hat{\sigma}^4} \sum_{n=0}^{N-1} x^2[n] \Rightarrow \hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n]$$

As a result,

$$\hat{P} = 10\log_{10}(\hat{\sigma}^2) = 10\log_{10}\left(\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \right)$$

Example 5.12

Given

$$x[n] = A + w[n], \qquad n = 0,1,\cdots,N-1$$

where A is an unknown constant and $w[n]$ is a white Gaussian noise with unknown variance σ^2. Find the ML estimates of A and σ^2.

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n]-A)^2 \right), \qquad \boldsymbol{\theta} = [A \quad \sigma^2]^T$$

$$\Rightarrow \quad \frac{\partial \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n]-A)$$

$$\frac{\partial \ln(p(\mathbf{x};\boldsymbol{\theta}))}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{n=0}^{N-1} (x[n]-A)^2$$

Solving the first equation:

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \bar{x}$$

Putting $A = \hat{A} = \bar{x}$ in the second equation:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2$$

Numerical Computation of ML Solution

When the ML solution is not of closed form, it can be computed by

- Grid search
- Numerical methods: Newton-Raphson, Golden section, bisection, etc

Example 5.13

From Example 5.10, the ML solution of ϕ is determined from

$$\sum_{n=0}^{N-1} x[n]\sin(\omega_0 n + \hat{\phi}) = \frac{A}{2}\sum_{n=0}^{N-1} \sin(2\omega_0 n + 2\hat{\phi})$$

Suggest methods to find $\hat{\phi}$

Approach 1: Grid search

Let

$$g(\phi) = \sum_{n=0}^{N-1} x[n]\sin(\omega_0 n + \phi) - \frac{A}{2}\sum_{n=0}^{N-1} \sin(2\omega_0 n + 2\phi)$$

It is obvious that

$$\hat{\phi} = \text{root of } g(\phi)$$

The idea of grid search is simple:

- Search for all possible values of $\hat{\phi}$ or a given range of $\hat{\phi}$ to find root
- Values are discrete $\Rightarrow$ tradeoff between resolution & computation

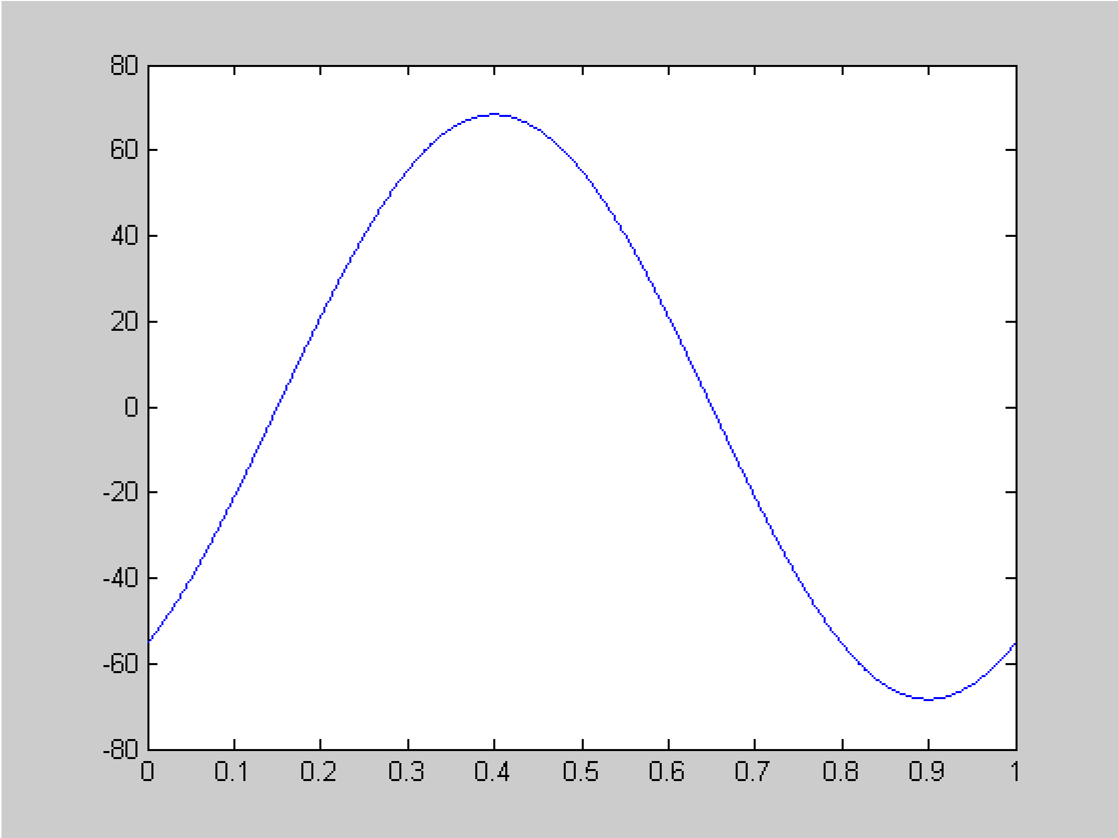
e.g., Range for $\hat{\phi}$: any values in $[0,2\pi)$

Discrete points : 1000 $\Rightarrow$ resolution is $2\pi/1000$

MATLAB source code:

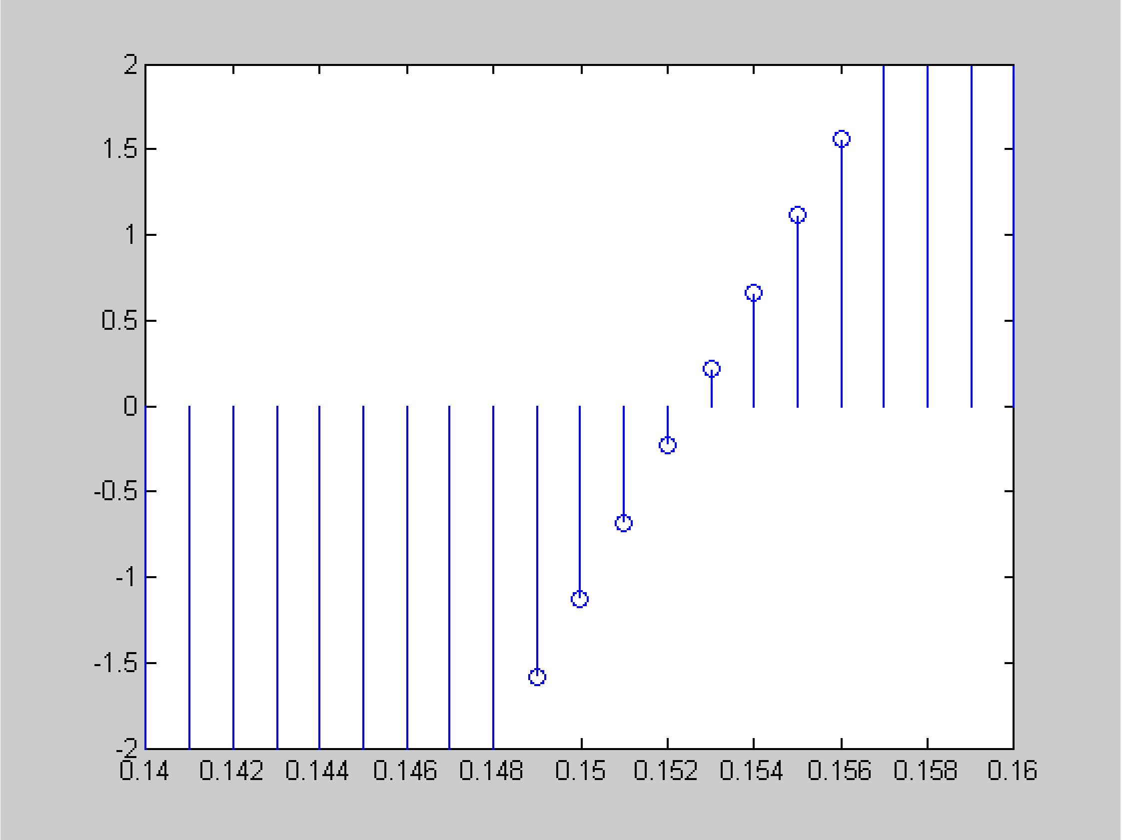
```
N=100;
n=[0:N-1];
w = 0.2*pi;
A = sqrt(2);
p = 0.3*pi;
np = 0.1;
q = sqrt(np).*randn(1,N);
x = A.*cos(w.*n+p)+q;
for j=1:1000
   pe = j/1000*2*pi;
   s1 =sin(w.*n+pe);
   s2 =sin(2.*w.*n+2.*pe);
   g(j) = x*s1'-A/2*sum(s2);
end
```

```
pe = [1:1000]/1000;
plot(pe,g)
```



Note: x-axis is $\phi/(2\pi)$

```
stem(pe,g)
axis([0.14 0.16 -2 2])
```

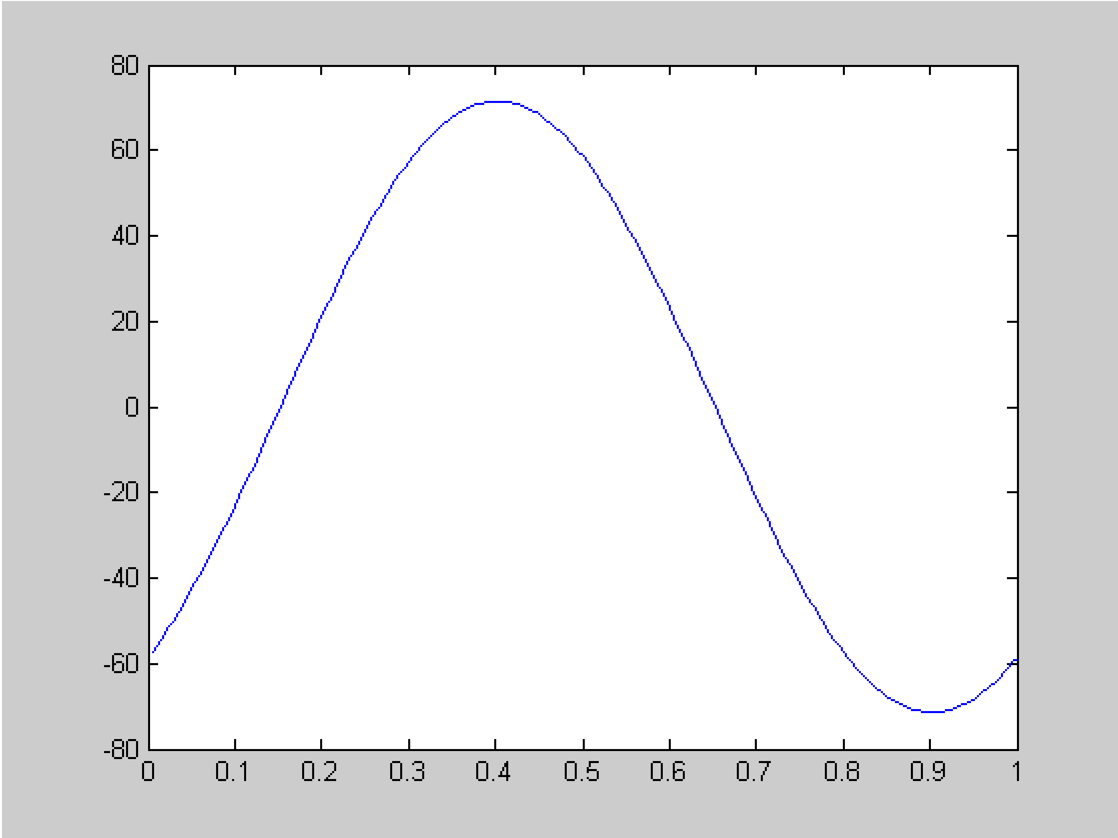


$$g(0.152 \cdot 2\pi) = -0.2324, \; g(0.153 \cdot 2\pi) = 0.2168$$

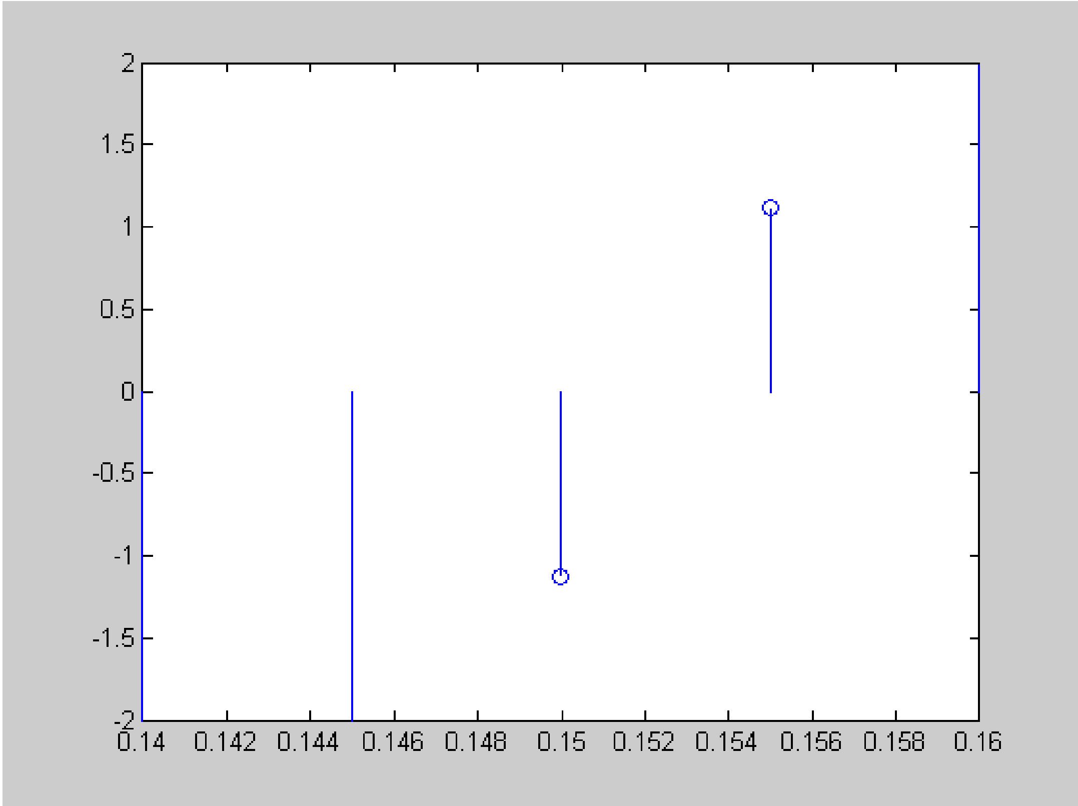
$$\Rightarrow \qquad \hat{\phi} = 0.153 \cdot 2\pi = 0.306\pi \; (\pm 0.001\pi)$$

For a smaller resolution, say 200 discrete points:

```
clear pe;
clear s1;
clear s2;
clear g;
for j=1:200
    pe = j/200*2*pi;
    s1 =sin(w.*n+pe);
    s2 =sin(2.*w.*n+2.*pe);
    g(j) = x*s1'-A/2*sum(s2);
end
pe = [1:200]/200;
plot(pe,g)
```



```
stem(pe,g)
axis([0.14 0.16 -2 2])
```



$$g(0.150 \cdot 2\pi) = -1.1306,\ g(0.155 \cdot 2\pi) = 1.1150$$

$\Rightarrow$ $$\hat{\phi} = 0.155 \cdot 2\pi = 0.310\pi\ (\pm 0.005\pi)$$

$\Rightarrow$ Accuracy increases as number of grids increases

Approach 2: Newton/Raphson iterative procedure

With initial guess $\hat{\phi}_0$, the root of $g(\phi)$ can be determined from

$$\hat{\phi}_{k+1} = \hat{\phi}_k - \frac{g(\hat{\phi}_k)}{\left.\frac{dg(\phi)}{d\phi}\right|_{\phi=\hat{\phi}_k}} = \frac{g(\hat{\phi}_k)}{g'(\hat{\phi}_k)} \tag{5.13}$$

$$g(\phi) = \sum_{n=0}^{N-1} x[n]\sin(\omega_0 n + \phi) - \frac{A}{2}\sum_{n=0}^{N-1} \sin(2\omega_0 n + 2\phi)$$

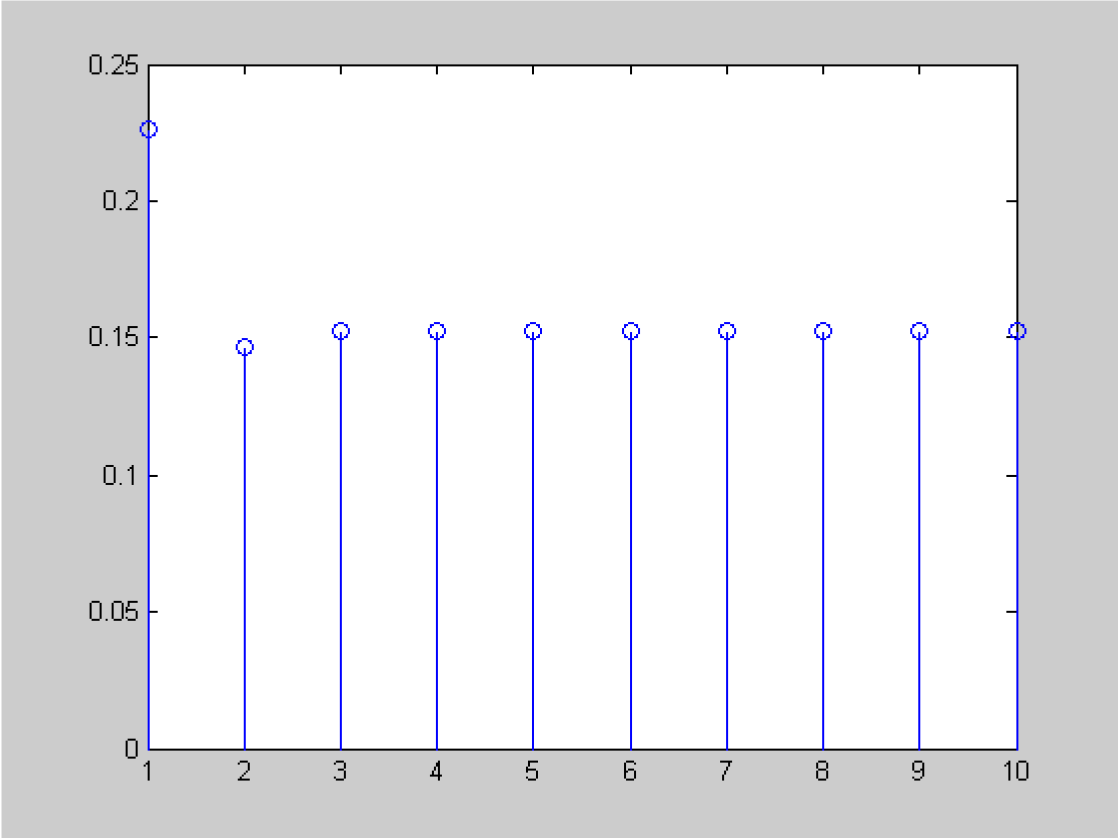
$$g'(\phi) = \sum_{n=0}^{N-1} x[n]\cos(\omega_0 n + \phi) - \frac{A}{2}\sum_{n=0}^{N-1} \cos(2\omega_0 n + 2\phi)\cdot 2$$

$$= \sum_{n=0}^{N-1} x[n]\cos(\omega_0 n + \phi) - A\sum_{n=0}^{N-1} \cos(2\omega_0 n + 2\phi)$$

with

$$\hat{\phi}_0 = 0$$

```
p1 = 0;
for k=1:10
    s1 =sin(w.*n+p1);
    s2 =sin(2.*w.*n+2.*p1);
    c1 =cos(w.*n+p1);
    c2 =cos(2.*w.*n+2.*p1);
    g = x*s1'-A/2*sum(s2);
    g1 = x*c1'-A*sum(c2);
    p1 = p1 - g/g1;
    p1_vector(k) = p1;
end
stem(p1_vector/(2*pi))
```



Newton/Raphson method converges at ~ 3rd iteration

$$\hat{\phi} = 0.1525 \cdot 2\pi = 0.305\pi$$

Q. Can you comment on the grid search & Newton/Raphson method?

ML Estimation for General Linear Model

The general linear data model is given by

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w} \tag{5.14}$$

where

$\mathbf{x}$ is the observed vector of size N
$\mathbf{w}$ is Gaussian noise vector with known covariance matrix $\mathbf{C}$
$\mathbf{H}$ is known matrix of size $N \times p$
$\boldsymbol{\theta}$ is parameter vector of size p

Based on (5.7), the PDF of $\mathbf{x}$ parameterized by $\boldsymbol{\theta}$ is

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}\det^{1/2}(\mathbf{C})}\exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T \cdot \mathbf{C}^{-1} \cdot (\mathbf{x}-\mathbf{H}\boldsymbol{\theta})\right) \tag{5.15}$$

Since $\mathbf{C}$ is not a function of $\boldsymbol{\theta}$, the ML solution is equivalent to

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \{J(\boldsymbol{\theta})\} \quad \text{where} \quad J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \cdot \mathbf{C}^{-1} \cdot (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

Differentiating $J(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and then set the result to zero:

$$-2\mathbf{H}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{x} + 2\mathbf{H}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{H}\hat{\boldsymbol{\theta}} = 0$$

$$\Rightarrow \mathbf{H}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{x} = \mathbf{H}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{H} \cdot \hat{\boldsymbol{\theta}}$$

As a result, the ML solution for linear model is

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \cdot \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} \qquad (5.16)$$

For white noise:

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T (\sigma_w^2 \cdot \mathbf{I})^{-1} \mathbf{H})^{-1} \cdot \mathbf{H}^T (\sigma_w^2 \cdot \mathbf{I})^{-1} \mathbf{x} = (\mathbf{H}^T \mathbf{H})^{-1} \cdot \mathbf{H}^T \mathbf{x} \qquad (5.17)$$

Example 5.14

Given N pair of (x, y) where x is error-free but y is subject to error:

$$y[n] = m \cdot x[n] + c + w[n] \qquad , n = 0,1,\cdots, N-1$$

where $\mathbf{w}$ is white Gaussian noise vector with known covariance matrix $\mathbf{C}$

Find the ML estimates for m and c

$$y[n] = m \cdot x[n] + c + w[n]$$

$$\Rightarrow y[n] = [x[n] \quad 1] \cdot \begin{bmatrix} m \\ c \end{bmatrix} + w[n] = [x[n] \quad 1] \cdot \boldsymbol{\theta} + w[n], \qquad \boldsymbol{\theta} = [m \quad c]^T$$

$$\Rightarrow \qquad \begin{aligned} y[0] &= [x[0] \quad 1] \cdot \boldsymbol{\theta} + w[0] \\ y[1] &= [x[1] \quad 1] \cdot \boldsymbol{\theta} + w[1] \\ \cdots & \qquad \cdots \qquad \cdots \\ y[N-1] &= [x[N-1] \quad 1] \cdot \boldsymbol{\theta} + w[N-1] \end{aligned}$$

Writing in matrix form:

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

where

$$\mathbf{y} = [y[0], y[1], \cdots, y[N-1]]^T$$

$$\mathbf{H} = \begin{bmatrix} x[0] & 1 \\ x[1] & 1 \\ \vdots & \vdots \\ x[N-1] & 1 \end{bmatrix}$$

Applying (5.16) gives

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{m} \\ \hat{c} \end{bmatrix} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \cdot \mathbf{H}^T \mathbf{C}^{-1} \mathbf{y}$$

Example 5.15

Find the ML estimates of A, ω_0 and ϕ for

$$x[n] = A\cos(\omega_0 n + \phi) + w[n], \qquad n = 0,1,\cdots,N-1, \qquad N >> 1$$

where $w[n]$ is a white Gaussian noise with variance σ_w^2

Recall from Example 5.6:

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma_w^2)^{N/2}} \exp\left(-\frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} (x[n] - A\cos(\omega_0 n + \phi))^2 \right), \qquad \boldsymbol{\theta} = [A, \omega_0, \phi]$$

The ML solution for $\boldsymbol{\theta}$ can be found by minimizing

$$J(A, \omega_0, \phi) = \sum_{n=0}^{N-1} (x[n] - A\cos(\omega_0 n + \phi))^2$$

This can be achieved by using a 3-D grid search or Netwon/Raphson method but it is computationally complex

Another simpler solution is as follows

$$J(A,\omega_0,\phi) = \sum_{n=0}^{N-1} (x[n] - A\cos(\omega_0 n + \phi))^2$$

$$= \sum_{n=0}^{N-1} (x[n] - A\cos(\phi)\cos(\omega_0 n) + A\sin(\phi)\sin(\omega_0 n))^2$$

Since A and ϕ are not quadratic in $J(A,\omega_0,\phi)$, the first step is to use parameter transformation:

$$\begin{aligned} \alpha_1 &= A\cos(\phi) \\ \alpha_2 &= -A\sin(\phi) \end{aligned} \quad \Rightarrow \quad \begin{aligned} A &= \sqrt{\alpha_1^2 + \alpha_2^2} \\ \phi &= \tan^{-1}\left(\frac{-\alpha_2}{\alpha_1}\right) \end{aligned}$$

Let

$$\mathbf{c} = [1 \quad \cos(\omega_0) \cdots \cos(\omega_0(N-1))]^T$$

$$\mathbf{s} = [0 \quad \sin(\omega_0) \cdots \sin(\omega_0(N-1))]^T$$

We have

$$J(\alpha_1, \alpha_2, \omega_0) = (\mathbf{x} - \alpha_1 \mathbf{c} - \alpha_2 \mathbf{s})^T (\mathbf{x} - \alpha_1 \mathbf{c} - \alpha_2 \mathbf{s})$$

$$= (\mathbf{x} - \mathbf{H}\boldsymbol{\alpha})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\alpha}), \qquad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \quad \mathbf{H} = [\mathbf{c} \quad \mathbf{s}]$$

Applying (5.17) gives

$$\hat{\boldsymbol{\alpha}} = (\mathbf{H}^T \mathbf{H})^{-1} \cdot \mathbf{H}^T \mathbf{x}$$

Substituting back to $J(\alpha_1, \alpha_2, \omega_0)$:

$$
\begin{aligned}
J(\omega_0) &= (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\alpha}})^T (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\alpha}}) \\
&= (\mathbf{x} - \mathbf{H} \cdot (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x})^T (\mathbf{x} - \mathbf{H} \cdot (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}) \\
&= \left((\mathbf{I} - \mathbf{H} \cdot (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \cdot \mathbf{x}\right)^T \left((\mathbf{I} - \mathbf{H} \cdot (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \cdot \mathbf{x}\right) \\
&= \mathbf{x}^T \cdot (\mathbf{I} - \mathbf{H} \cdot (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T)^T (\mathbf{I} - \mathbf{H} \cdot (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \cdot \mathbf{x} \\
&= \mathbf{x}^T \cdot (\mathbf{I} - \mathbf{H} \cdot (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \cdot \mathbf{x} \\
&= \mathbf{x}^T \cdot \mathbf{x} - \mathbf{x}^T \cdot \mathbf{H} \cdot (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \cdot \mathbf{x}
\end{aligned}
$$

Minimizing $J(\omega_0)$ is identical to maximizing

$$
\mathbf{x}^T \cdot \mathbf{H} \cdot (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \cdot \mathbf{x}
$$

or

$$
\hat{\omega}_0 = \arg\max_{\omega_0} \left\{ \mathbf{x}^T \cdot \mathbf{H} \cdot (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \cdot \mathbf{x} \right\}
$$

$\Rightarrow$ 3-D search is reduced to a 1-D search

After determining $\hat{\omega}_0$, $\hat{\boldsymbol{\alpha}}$ can be obtained as well

For sufficiently large N:

$$\mathbf{x}^T \cdot \mathbf{H} \cdot (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \cdot \mathbf{x} = \begin{bmatrix} \mathbf{c}^T \mathbf{x} & \mathbf{s}^T \mathbf{x} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{c}^T \mathbf{c} & \mathbf{c}^T \mathbf{s} \\ \mathbf{s}^T \mathbf{c} & \mathbf{s}^T \mathbf{s} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}$$

$$\approx \begin{bmatrix} \mathbf{c}^T \mathbf{x} & \mathbf{s}^T \mathbf{x} \end{bmatrix} \cdot \begin{bmatrix} N/2 & 0 \\ 0 & N/2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}$$

$$= \frac{2}{N} \left(\left(\mathbf{c}^T \mathbf{x} \right)^2 + \left(\mathbf{s}^T \mathbf{x} \right)^2 \right)$$

$$= \frac{2}{N} \left| \sum_{n=0}^{N-1} x[n] \exp(-j\omega_0 n) \right|^2$$

$$\Rightarrow \hat{\omega}_0 = \arg\max_{\omega_0} \left\{ \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] \exp(-j\omega_0 n) \right|^2 \right\} \Rightarrow \text{periodogram maximum}$$

Least Squares Methods

Parameter estimation is achieved via minimizing a least squares (LS) cost function

Generally not optimum but computationally simple

Use for classical parameter estimation

No knowledge of the noise PDF is required

Can be considered as a generalization of LS filtering

Variants of LS Methods

1. Standard LS

Consider the general linear data model:

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

where

$\mathbf{x}$ is the observed vector of size N
$\mathbf{w}$ is zero-mean noise vector with unknown covariance matrix
$\mathbf{H}$ is known matrix of size $N \times p$
$\boldsymbol{\theta}$ is parameter vector of size p

The LS solution is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right\} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} \qquad (5.18)$$

which is equal to (5.17)

$\Rightarrow$ LS solution is optimum if covariance matrix of $\mathbf{w}$ is $\mathbf{C} = \sigma_w^2 \cdot \mathbf{I}$ and $\mathbf{w}$ is Gaussian distributed

Define

$$\mathbf{e} = \mathbf{x} - \mathbf{H}\boldsymbol{\theta}$$

where

$$\mathbf{e} = [e(0) \quad e(1) \quad \cdots \quad e(N-1)]^T$$

(5.18) is equivalent to

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \sum_{k=0}^{N-1} e^2(k) \right\} \tag{5.19}$$

which is similar to LS filtering

Q. Any differences between (5.19) and LS filtering?

Example 5.16

Given

$$x[n] = A + w[n], \qquad n = 0,1,\cdots,N-1$$

where A is an unknown constant and $w[n]$ is a zero-mean noise

Find the LS solution of A

Using (5.19),

$$\hat{A} = \arg\min_{A} \left\{ \sum_{n=0}^{N-1} \left(x[n] - A \right)^2 \right\}$$

Differentiating $\sum_{n=0}^{N-1} \left(x[n] - A \right)^2$ with respect to A and set the result to 0:

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

On the other hand, writing $\{x[n]\}$ in matrix form:

$$\mathbf{x} = \mathbf{H}A + \mathbf{w}$$

where

$$\mathbf{H} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Using (5.18),

$$\hat{A} = \left([1 \;\; 1 \;\; \cdots \;\; 1] \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right)^{-1} [1 \;\; 1 \;\; \cdots \;\; 1] \cdot \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix} = N^{-1} \cdot \sum_{n=0}^{N-1} x[n]$$

Both (5.18) and (5.19) give the same answer and the LS solution is optimum if the noise is white Gaussian

Example 5.17

Consider the LS filtering problem again. Given

$$d[n] = \underline{X}^T[n] \cdot \underline{W} + q[n], \qquad n = 0,1,\cdots,N-1$$

where

$d[n]$ is desired response

$\underline{X}[n] = [x[n] \quad x[n-1] \quad \cdots x[n-L+1]]^T$ is the input signal vector

$\underline{W} = [w_0 \quad w_1 \quad \cdots \quad w_{L-1}]^T$ is the unknown filter weight vector

$q[n]$ is zero-mean noise

Writing in matrix form:

$$\mathbf{d} = \mathbf{H} \cdot \mathbf{W} + \mathbf{q}, \qquad \mathbf{W} = \underline{W}$$

Using (5.18):

$$\hat{\mathbf{W}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{d}$$

where

$$\mathbf{H} = \begin{bmatrix} \underline{X}^T(0) \\ \underline{X}^T(1) \\ \vdots \\ \underline{X}^T(N-1) \end{bmatrix} = \begin{bmatrix} x[0] & 0 & \cdots & 0 \\ x[1] & x[0] & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ x[N-1] & x[N-2] & \cdots & x[N-L] \end{bmatrix}$$

with $x[-1] = x[-2] = \cdots = 0$

Note that

$$\underline{R}_{xx} = \mathbf{H}^T \mathbf{H}$$

$$\underline{R}_{dx} = \mathbf{H}^T \mathbf{d}$$

where $\underline{R}_{xx}$ is not the original version but not modified version of (3.6)

Example 5.18

Find the LS estimate of A for

$$x[n] = A\cos(\omega_0 n + \phi) + w[n], \qquad n = 0,1,\cdots,N-1, \qquad N >> 1$$

where ω_0 and ϕ are known constants while $w[n]$ is zero-mean noise

Using (5.19),

$$\hat{A} = \arg\min_{A}\left\{\sum_{n=0}^{N-1}\left(x[n] - A\cos(\omega_0 n + \phi)\right)^2\right\}$$

Differentiate $\sum_{n=0}^{N-1}\left(x[n] - A\cos(\omega_0 n + \phi)\right)^2$ with respect to A & set result to 0:

$$2\sum_{n=0}^{N-1}\left(x[n] - A\cos(\omega_0 n + \phi)\right)\cdot -\cos(\omega_0 n + \phi) = 0$$

$$\Rightarrow \sum_{n=0}^{N-1} x[n]\cos(\omega_0 n + \phi) = A\sum_{n=0}^{N-1}\cos^2(\omega_0 n + \phi)$$

The LS solution is then

$$\hat{A} = \frac{\sum_{n=0}^{N-1} x[n]\cos(\omega_0 n + \phi)}{\sum_{n=0}^{N-1} \cos^2(\omega_0 n + \phi)}$$

2. Weighted LS

Use a general form of LS via a symmetric weighting matrix $\mathbf{W}$

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{W} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right\} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{x} \qquad (5.20)$$

such that

$$\mathbf{W} = \mathbf{W}^T$$

Due to the presence of $\mathbf{W}$, it is generally difficult to write the cost function $(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{W} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$ in scalar form as in (5.19)

Rationale of using $\mathbf{W}$: put larger weights on data with smaller errors

put smaller weights on data with larger errors

When $\mathbf{W}=\mathbf{C}^{-1}$ where $\mathbf{C}$ is covariance matrix of the noise vector:

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x} \qquad (5.21)$$

which is equal to the ML solution and is optimum for Gaussian noise

Example 5.19

Given two noisy measurements of A:

$$x_1 = A + w_1 \quad \text{and} \quad x_2 = A + w_2$$

where w_1 and w_2 are zero-mean uncorrelated noises with known variances σ_1^2 and σ_2^2. Determine the optimum weighted LS solution

Use

$$\mathbf{W} = \mathbf{C}^{-1} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1} = \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix}$$

Grouping x_1 and x_2 into matrix form:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot A + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

or

$$\mathbf{x} = \mathbf{H} \cdot A + \mathbf{w}$$

Using (5.21)

$$\hat{A} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} = \left(\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

As a result,

$$\hat{A} = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)^{-1} \left(\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2} \right) = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \cdot x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \cdot x_2$$

Note that

- If $\sigma_2^2 > \sigma_1^2$, a larger weight is placed on x_1 and vice versa
- If $\sigma_2^2 = \sigma_1^2$, the solution is equal to the standard sample mean
- The solution will be more complicated if w_1 and w_2 are correlated
- Exact values for σ_1^2 and σ_2^2 are not necessary, only ratio is needed

  Define $\lambda = \sigma_1^2 / \sigma_2^2$, we have

$$\hat{A} = \frac{1}{1+\lambda} \cdot x_1 + \frac{\lambda}{1+\lambda} \cdot x_2$$

3. Nonlinear LS

The LS cost function cannot be represented as a linear model as in

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

In general, it is more complex to solve, e.g.,

The LS estimates for A, ω_0 and ϕ can be found by minimizing

$$\sum_{n=0}^{N-1} (x[n] - A\cos(\omega_0 n + \phi))^2$$

whose solution is not straightforward as seen in Example 5.15

Grid search and numerical methods are used to find the minimum

4. Constrained LS

The linear LS cost function is minimized subject to constraints:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right\} \quad \text{subject to } \mathbf{S} \qquad (5.22)$$

where $\mathbf{S}$ is a set of equalities/inequalities in terms of $\boldsymbol{\theta}$

Generally it can be solved by linear/nonlinear programming, but simpler solution exists for linear and quadratic constraint equations, e.g.,

Linear constraint equation: $\theta_1 + \theta_2 + \theta_3 = 10$

Quadratic constraint equation: $\theta_1^2 + \theta_2^2 + \theta_3^2 = 100$

Other types of constraints: $\theta_1 > \theta_2 > \theta_3 > 10$

$$\theta_1 + \theta_2^2 + \theta_3^3 \geq 100$$

Consider the constraints $\mathbf{S}$ is

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$$

which contains r linear equations. The constrained LS problem for linear model is

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right\} \quad \text{subject to } \mathbf{A}\boldsymbol{\theta} = \mathbf{b} \tag{5.23}$$

The technique of Lagrangian multipliers can solve (5.23) as follows

Define the Lagrangian

$$J_c = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) + \boldsymbol{\lambda}^T (\mathbf{A}\boldsymbol{\theta} - \mathbf{b}) \tag{5.24}$$

where $\boldsymbol{\lambda}$ is a r-length vector of Lagrangian multipliers

The procedure is first solve $\boldsymbol{\lambda}$ then $\boldsymbol{\theta}$

Expanding (5.24):

$$J_c = \mathbf{x}^T\mathbf{x} - 2\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{x} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} + \boldsymbol{\lambda}^T\mathbf{A}\boldsymbol{\theta} - \boldsymbol{\lambda}^T\mathbf{b}$$

Differentiate J_c with respect to $\boldsymbol{\theta}$:

$$\frac{\partial J_c}{\partial \boldsymbol{\theta}} = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} + \mathbf{A}^T\boldsymbol{\lambda}$$

Set the result to zero:

$$-2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\hat{\boldsymbol{\theta}}_c + \mathbf{A}^T\boldsymbol{\lambda} = \mathbf{0}$$

$$\Rightarrow \hat{\boldsymbol{\theta}}_c = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} - \frac{1}{2}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\boldsymbol{\lambda} = \hat{\boldsymbol{\theta}} - \frac{1}{2}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\boldsymbol{\lambda}$$

where $\hat{\boldsymbol{\theta}}$ is the LS solution. Put $\hat{\boldsymbol{\theta}}_c$ into $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$:

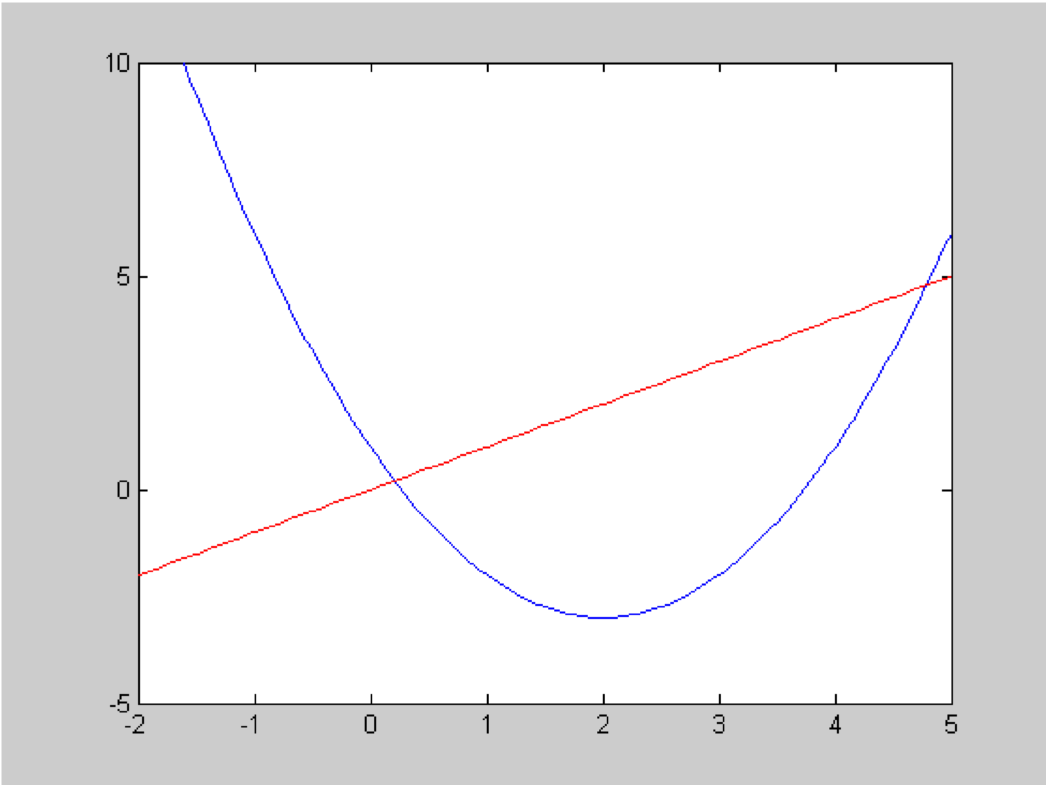
$$\mathbf{A}\hat{\boldsymbol{\theta}}_c = \mathbf{A}\hat{\boldsymbol{\theta}} - \frac{1}{2}\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T\boldsymbol{\lambda} = \mathbf{b} \Rightarrow \frac{\boldsymbol{\lambda}}{2} = (\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T)^{-1}(\mathbf{A}\hat{\boldsymbol{\theta}} - \mathbf{b})$$

Put $\boldsymbol{\lambda}$ back to $\hat{\boldsymbol{\theta}}_c$:

$$\hat{\boldsymbol{\theta}}_c = \hat{\boldsymbol{\theta}} - (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T(\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T)^{-1}(\mathbf{A}\hat{\boldsymbol{\theta}} - \mathbf{b})$$

Idea of constrained LS can be illustrated by finding minimum value of y:

$$y = x^2 - 3x + 2 \qquad \text{subject to } x - y = 1$$



5. Total LS

Motivation: Noises at both $\mathbf{x}$ and $\mathbf{H}\boldsymbol{\theta}$:

$$\mathbf{x} + \mathbf{w_1} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w_2} \tag{5.25}$$

where $\mathbf{w_1}$ and $\mathbf{w_2}$ are zero-mean noise vectors

A typical example is LS filtering in the presence of both input noise and output noise. The noisy input is

$$x(k) = s(k) + n_i(k), \qquad n = 0,1,\cdots,N-1$$

and the noisy output is

$$r(k) = s(k) \otimes h(k) + n_o(k), \qquad n = 0,1,\cdots,N-1$$

The parameters to be estimated are $\{h(k)\}$ given $x(k)$ and $y(k)$

Another example is in frequency estimation using linear prediction:

For a single sinusoid $s(k) = A\cos(\omega k + \phi)$, it is true that

$$s(k) = 2\cos(\omega) s(k-1) - s(k-2)$$

$s(k)$ is perfectly predicted by $s(k-1)$ and $s(k-1)$:

$$s(k) = a_0 s(k-1) + a_1 s(k-2)$$

It is desirable to obtain $a_0 = 2\cos(\omega)$ and $a_1 = -1$ in estimation process

In the presence of noise, the observed signal is

$$x(k) = s(k) + w(k), \qquad n = 0,1,\cdots, N-1$$

The linear prediction model is now

$$x(k) = a_0 x(k-1) + a_1 x(k-2), \qquad n = 0,1,\cdots,N-1$$

$$\begin{aligned} x(2) &= a_0 x(1) + a_1 x(0) \\ x(3) &= a_0 x(2) + a_1 x(1) \\ \cdots \quad & \cdots \quad \cdots \\ x(N-1) &= a_0 x(N-2) + a_1 x(N-3) \end{aligned} \Rightarrow \begin{bmatrix} x(2) \\ x(3) \\ \vdots \\ x(N-1) \end{bmatrix} = \begin{bmatrix} x(1) & x(0) \\ x(2) & x(1) \\ \vdots & \vdots \\ x(N-2) & x(N-2) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$

$$\begin{bmatrix} s(2) \\ s(3) \\ \vdots \\ s(N-1) \end{bmatrix} + \begin{bmatrix} w(2) \\ w(3) \\ \vdots \\ w(N-1) \end{bmatrix} = \begin{bmatrix} s(1) & s(0) \\ s(2) & s(1) \\ \vdots & \vdots \\ s(N-2) & s(N-2) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} + \begin{bmatrix} w(1) & w(0) \\ w(2) & w(1) \\ \vdots & \vdots \\ w(N-2) & w(N-2) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$

6. Mixed LS

A combination of LS, weighted LS, nonlinear LS, constrained LS and/or total LS

Examples: weighted LS with constraints, total LS with constraints, etc.

Questions for Discussion

1. Suppose you have N pairs of (x_i, y_i), $i = 1,2,\cdots,N$ and you need to fit them into the model of $y = ax$. Assuming that only $\{y_i\}$ contain zero-mean noise, determine the least squares estimate for a.

   (Hint: ithe relationship between x_i and y_i is

$$y_i = ax_i + n_i, \qquad i = 1,2,\cdots,N$$

   where $\{n_i\}$ are the noise in $\{y_i\}$.)

2. Use least squares to estimate the line $y = ax$ in Q.1 but now only $\{x_i\}$ contain zero-mean noise.

3. In a radar system, the received signal is

$$r(n) = \alpha s(n - \tau_0) + w(n)$$

where the range R of an object is related to the time delay by

$$\tau_0 = 2R / c$$

Suppose we get an unbiased estimate of τ_0, say, $\hat{\tau}_0$, and its variance is $\mathrm{var}(\hat{\tau}_0)$. Determine the corresponding range variance $\mathrm{var}(\hat{R})$, where $\hat{R}$ is the estimate of R.

If $\mathrm{var}(\hat{\tau}_0) = (0.1 \mu s)^2$ and $c = 3 \times 10^8 ms^{-1}$, what is the value of $\mathrm{var}(\hat{R})$?