Chapter 9:
Estimation Theory

- Fundamentals
- Minimum Variance Unbiased (MVU) Estimators
- Maximum Likelihood (ML) Estimation
- Spectral Estimation

References

[1] Simon Haykin, *Adaptive Filter Theory*, Prentice Hall, 1996 (3rd Ed.), 2001 (4th Ed.).

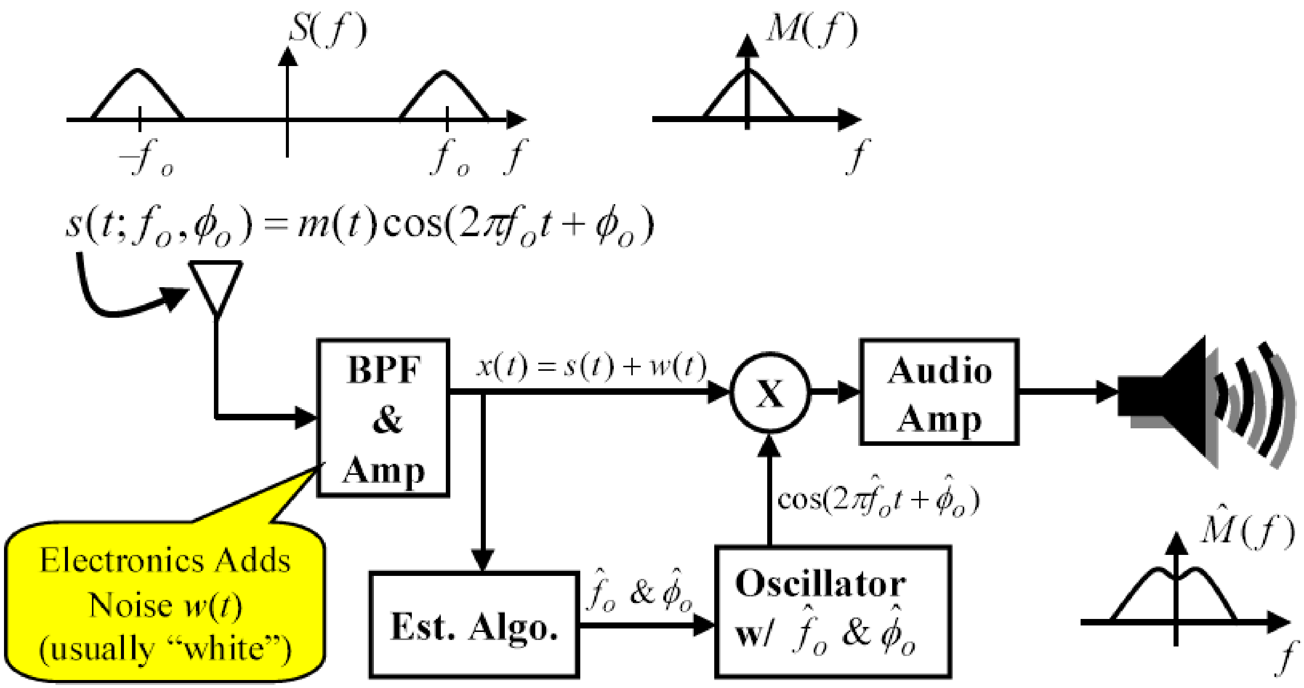
[2] Steven M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.

[3] Alan V. Oppenheim, Ronald W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, 1989.

[4] Athanasios Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1991 (3rd Ed.), 2001 (4th Ed.).

9. Fundamentals (1)

- **An example estimation problem**: **DSB Rx**



Goal: Given $x(t) = s(t; f_o, \phi_o) + w(t)$

Find Estimates

(that are optimal in some sense)

Describe with Probability Model: PDF & Correlation

9. Fundamentals (2)

- **Discrete-time estimation problem**: working with samples of the observed signal (signal plus noise):

$$x[n] = s[n; f_0, \phi_0] + w[n]$$

Each time "observing" $x[n]$ it contains same $s[n]$ but different "realization" of noise $w[n]$, so the estimate is different each time.

So, $\hat{f}_0$ & $\hat{\phi}_0$ **are RVs.**

Our job: Given finite data set $x[0]$, $x[1]$, …, $x[N\text{-}1]$, find estimator functions that map data into estimates:

$$\hat{f}_0 = g_1\left(x[0], x[1], \cdots x[N\text{-}1]\right) = g_1(\mathbf{x})$$

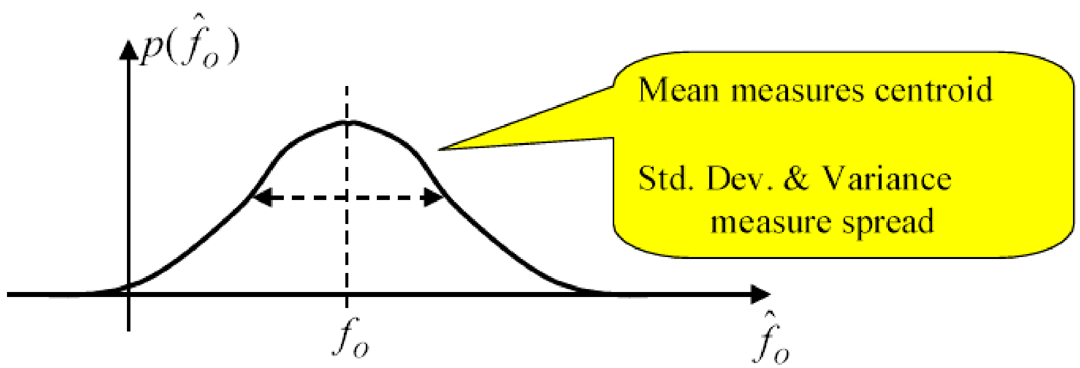
$$\hat{\phi}_0 = g_2\left(x[0], x[1], \cdots x[N\text{-}1]\right) = g_2(\mathbf{x})$$

These are RVs, need to describe with **probability model.**

9. Fundamentals (3)

PDF of estimate: Because estimates are RVs we describe them with a PDF depending on:

- structure of $s[n]$
- probability model of $w[n]$
- form of est. function $g(\mathbf{x})$



Desire: $E\{\hat{f}_o\} = f_o$

$$\sigma^2_{\hat{f}_o} = E\left\{\left(\hat{f}_o - E\{\hat{f}_o\}\right)^2\right\} = \text{small}$$

9. Fundamentals (4)

- **General mathematical statement of estimation problem:**

For measured data $\mathbf{x} = [x[0], x[1], \ldots, x[N\text{-}1]]$, unknown parameter $\boldsymbol{\theta} = [\theta_1, \theta_2 \ldots \theta_p]$

$\boldsymbol{\theta}$ is not random,

$\mathbf{x}$ is an N-dimensional random data vector.

Q: What captures all the statistical information needed for an estimation problem ?

A: Need the N-dimensional PDF of the data, parameterized by $\boldsymbol{\theta}$: $p(\mathbf{x}; \boldsymbol{\theta})$.

We will use $p(\mathbf{x}; \boldsymbol{\theta})$ to find $g(x)$. In practice, not given PDF ! Choose a suitable **model**:

- Captures essence of reality
- Leads to tractable answer

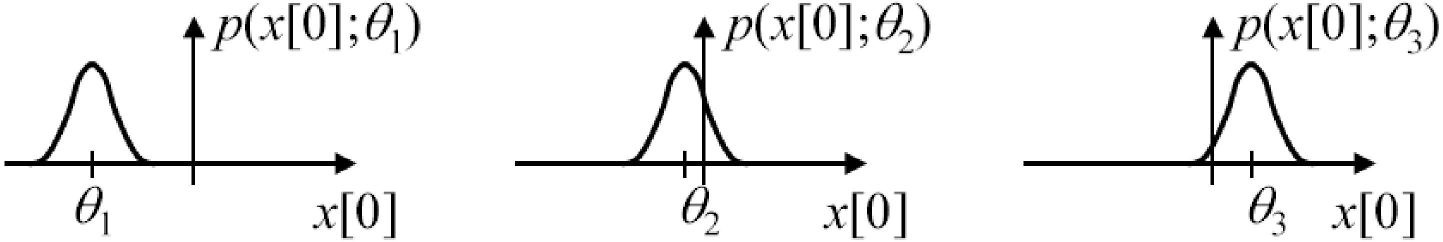
9. Fundamentals (5)

<u>Example</u>: Estimating a DC level in zero mean AWGN

Consider a single data point is observed: $x[0] = \theta + w[0]$, $w[0]$ is $N(0, \sigma^2)$. Then, $\theta + w[0]$ is $N(\theta, \sigma^2)$.

So, the needed parameterized PDF is: $p(x[0];\theta)$ which is Gaussian with mean of θ.

So, in this case the parameterization changes the data PDF mean:



9. Fundamentals (6)

- **<u>Typical assumptions for noise model</u>:**
  - **Whiteness** and **Gaussian** are always easiest to analyze
    - ✓ Usually assumed unless you have reason to believe otherwise
    - ✓ Whiteness is usually first assumption removed
    - ✓ Gaussian is less often removed due to the validity of Central Limit Theorem
  - **Zero mean** is a nearly universal assumption
    - ✓ Most practical cases have zero mean
    - ✓ But if not… $w[n] = w_{zm}[n] + \mu$
  - **Variance** of noise doesn't always have to be known to make an estimate
    - ✓ But, must know to assess expected "goodness" of the estimate
    - ✓ Usually perform "goodness" analysis as a function of noise variance (or SNR). Noise variance sets the SNR level of the problem.

9. Fundamentals (7)

❑ **<u>Classical vs. Bayesian estimation approaches</u>:**

- If we view **θ** (parameter to estimate) as **non-random** → **Classical Estimation**, provides no way to include a priori information about **θ** (Minimum Variance, Maximum Likelihood, Least Squares).

- If we view **θ** (parameter to estimate) as **random** → **Bayesian Estimation**, allows use of some a priori PDF on **θ**

  (MMSE, MAP, Wiener filter, Kalman Filter).

9. Fundamentals (8)

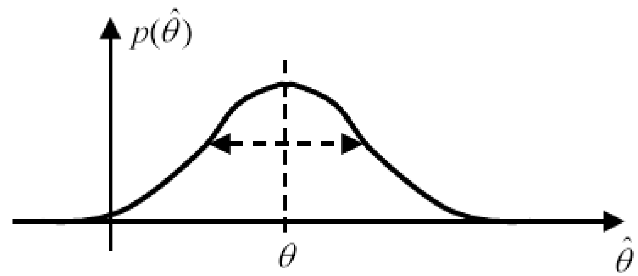
❑ Assessing estimator performance

Can only do this when the value of θ is known: Theoretical analysis, simulations, field tests, etc.

Recall that the estimate $\hat{\theta} = g(\mathbf{x})$ is a random variable.

Thus it has a PDF of its own and that PDF completely displays the quality of the estimate.

Illustrate with 1-D parameter case



Often just capture quality through <u>mean</u> and <u>variance</u> of $\hat{\theta} = g(\mathbf{x})$

Desire: $m_{\hat{\theta}} = E\{\hat{\theta}\} = \theta$

$$\sigma_{\hat{\theta}}^2 = E\left\{\left(\hat{\theta} - E\{\hat{\theta}\}\right)^2\right\} = \text{small}$$

If this is true: say estimate is "unbiased"

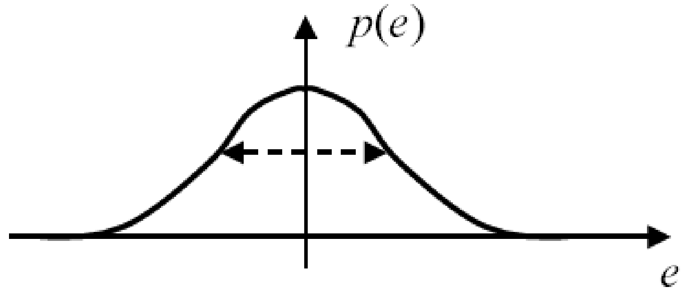
9. Fundamentals (9)

Equivalent view of assessing performance

Define estimation error: $\hat{\theta} = e + \theta$ or $e = \hat{\theta} - \theta$

RV RV not RV

Completely describe estimator quality with error PDF: $p(e)$



Desire:

$$m_e = E\{e\} = 0$$

$$\sigma_e^2 = E\left\{\left(e - E\{e\}\right)^2\right\} = \text{small}$$

If this is true: say estimate is "unbiased"

9. Fundamentals (10)

<u>Example</u>: DC level in AWGN

<u>Model</u>: $x[n] = A + w[n]$, $n = 0, 1, \ldots, N$-1

$w[n]$: Gaussian, zero mean, variance σ^2, white (uncorrelated sample-to-sample)

<u>PDF of an individual data sample</u>:

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

Uncorrelated Gaussian RVs are independent, so <u>joint PDF</u> is the product of the individual PDFs:

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

9. Fundamentals (11)

Each data sample has the same <u>mean</u> (A), which is the thing we are trying to estimate, we can imagine trying to estimate by finding the <u>sample mean:</u>

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

Let's analyze the quality of this estimator:

- Is it unbiased ?

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

Unbiased!

- Can we get a small variance ?

$$\operatorname{var}\left\{\hat{A}\right\} = \operatorname{var}\left\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right\} = \frac{1}{N^2}\sum_{n=0}^{N-1} \operatorname{var}\left\{x[n]\right\} = \frac{1}{N^2}\sum_{n=0}^{N-1}\sigma^2 = \frac{\sigma^2}{N}$$

Due to Indep.
(white & Gauss.)

$\Rightarrow$ Can make var small by increasing N!

9. Fundamentals (12)

❑ **<u>Theoretical analysis vs. simulations</u>**

- Ideally we'd like to be always be able to theoretically analyze the problem to find the bias and variance of the estimator.

  Theoretical results show <u>how performance depends on the problem specifications</u>.

- But sometimes we make use of simulations:
  - ✓ To verify that our theoretical analysis is correct.
  - ✓ Sometimes can't find theoretical results.

9. Fundamentals (13)

- **<u>Goal = Find "Optimal" estimators</u>**
  - There are several different definitions or criteria for optimality!
  - Most logical: **Minimum MSE** (Mean-Square-Error):

$$mse\left\{\hat{\theta}\right\} = E\left\{\left(\hat{\theta}-\theta\right)^2\right\} = E\left\{\left[\left(\hat{\theta}-E\left\{\hat{\theta}\right\}\right)+\left(\underbrace{E\left\{\hat{\theta}\right\}-\theta}_{b(\theta)}\right)\right]^2\right\}$$

$$= E\left\{\left[\hat{\theta}-E\left\{\hat{\theta}\right\}\right]^2\right\}+2b(\theta)\underbrace{E\left\{\hat{\theta}-E\left\{\hat{\theta}\right\}\right\}}_{=0}+b^2(\theta) = \mathrm{var}\left\{\hat{\theta}\right\}+b^2(\theta)$$

where: $b(\theta)$ is the **bias of estimate**.

9. Minimum Variance Unbiased (MVU) Estimation (1)

- **Basic idea of MVU:** Out of all unbiased estimates, find the one with the lowest variance.

- **Unbiased Estimators:** An estimator is unbiased if

$$E\left\{\hat{\theta}\right\} = \theta \qquad \text{for all} \quad \boldsymbol{\theta}$$

Example: Estimate DC in white uniform noise

$$x[n] = A + w[n], \quad n = 0, \ldots, N\text{-}1$$

Unbiased estimator:

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

same as before $E\left\{\hat{A}\right\} = A$ regardless of A value.

9. MVU Estimation (2)

- **<u>Minimum variance criterion</u>:** Constrain bias to be zero find the estimator that minimizes variance. Such an estimator is termed **Minimum Variance Unbiased** (**MVU**) estimator.

  Note that MSE of an unbiased estimator is the variance. So, MVU could also be called "Minimum MSE Unbiased Est."

- **<u>Existence of MVU Estimator</u>:**

  Sometimes there is no MVU Estimator, it can happen 2 ways:

  - There may be no unbiased estimators
  - None of the above unbiased estimators has a uniformly minimum variance

9. MVU Estimation (3)

- **Finding the MVU estimator:** Even if MVU exists: may not be able to find it.

  Three approaches to finding the MVU Estimators:

  - **Determine Cramer-Rao Lower Bound (CRLB)** and see if some estimator satisfies it. (Note: MVU can exist but not achieve the CRLB)
  - Apply Rao-Blackwell-Lechman-Scheffe Theorem: Rare in practice!
  - Restrict to Linear Unbiased & find MVLU: Only gives true MVU if problem is linear.

9. MVU Estimation (4)

- **<u>Vector Parameter</u>:** When we wish to estimate multiple parameters we group them into a vector:

$$\boldsymbol{\theta} = \left[\theta_1, \theta_2, \cdots \theta_p\right]^T$$

Then an estimator is notated as:

$$\hat{\boldsymbol{\theta}} = \left[\hat{\theta}_1, \hat{\theta}_2, \cdots \hat{\theta}_p\right]^T$$

Unbiased requirement becomes: $E\left\{\hat{\boldsymbol{\theta}}\right\} = \boldsymbol{\theta}$

Minimum Variance requirement becomes:

For each i…

$$var\left\{\hat{\boldsymbol{\theta}}\right\} = \boldsymbol{\theta} : \textbf{min over all unbiased estimates}$$

9. Cramer-Rao Lower Bound (CRLB) (1)

- **What is the Cramer-Rao Lower Bound:** CRLB is a lower bound on the variance of any unbiased estimator.

  If $\hat{\theta}$ is an unbiased estimator of θ, then

  $$\sigma^2_{\hat{\theta}}(\theta) \geq CRLB_{\hat{\theta}}(\theta) \Rightarrow \sigma_{\hat{\theta}}(\theta) \geq \sqrt{CRLB_{\hat{\theta}}(\theta)}$$

  The CRLB tells us the best we can ever expect to be able to do (with an unbiased estimator)

9. Cramer-Rao Lower Bound (CRLB) (2)

- **<u>Some uses of the CRLB</u>:**
  - Feasibility studies (e.g. Sensor usefulness, etc.)
    - Can we meet our specifications?
  - Judgment of proposed estimators
    - Estimators that don't achieve CRLB are looked down upon in the technical literature.
  - Can sometimes provide form for MVU estimate.
  - Demonstrates importance of physical and/or signal parameters to the estimation problem.

    e.g. We'll see that a signal's BW determines delay estimate accuracy $\Rightarrow$ Radars should use wide BW signals.

9. Cramer-Rao Lower Bound (CRLB) (3)

- **<u>Estimation accuracy consideration</u>:** What determines how well you can estimate θ ?

  Recall: Data vector is **x:** samples from a random process that depends on $\theta \Rightarrow$ the PDF describes that dependence: $p(\mathbf{x};\theta)$.

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

  $\Rightarrow$ Should look at $p(\mathbf{x};\theta)$ as a function of θ for fixed value of observed data **x**.

9. Cramer-Rao Lower Bound (CRLB) (4)

Example: PDF dependence for DC level in noise

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

where $w[0] \sim N(0, \sigma^2)$. Then the parameter-dependent PDF of the data point $x[0]$ is:

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

Define: **Likelihood Function (LF)**

The LF = the PDF $p(\mathbf{x}; \theta)$, but as a function of parameter θ with the data vector $\mathbf{x}$ fixed.

We will also often need the **Log Likelihood Function (LLF):**

$$\text{LLF} = \ln\{\text{LF}\} = \ln\{ p(\mathbf{x}; \theta)\}$$

9. Cramer-Rao Lower Bound (CRLB) (5)

- **<u>LF characteristics that affect accuracy</u>:** Intuitively, "sharpness" of the LF sets accuracy.

  Sharpness is measured using <u>curvature:</u>

$$\left.\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right|_{\substack{\mathbf{x}\text{: given data} \\ \theta\text{:true value}}}$$

Curvature ↑ ⇒ PDF concentration ↑ ⇒ Accuracy ↑

But this is for a particular set of data we want "in general". So, average over random vector to give the <u>average curvature</u>

$$-E\left\{\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right\}\Bigg|_{\theta\text{:true value}}$$

9. Cramer-Rao Lower Bound (CRLB) (6)

- **<u>Theorem</u>: CRLB for scalar parameter**

  Assume "regularity" condition is met:

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

  Then

$$\sigma_{\hat{\theta}}^2(\theta)\geq \underbrace{\frac{1}{-E\left\{\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right\}\Bigg|_{\theta\text{:true value}}}}_{CRLB}$$

  where

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

9. Cramer-Rao Lower Bound (CRLB) (7)

- **<u>Steps to find the CRLB</u>:**

  1. Write log 1ikelihood function as a function of θ:

$$\ln p(\mathbf{x};\, \theta)$$

  2. Fix $\mathbf{x}$ and take 2nd partial of LLF:

$$\partial^2 \ln p(\mathbf{x};\, \theta)/\partial\theta^2$$

  3. If result still depends on $\mathbf{x}$:
     - Fix θ and take expected value with respect to $\mathbf{x}$
     - Otherwise skip this step

  4. Result may still depend on θ: Evaluate at each specific value of θ desired.

  5. Negate and form reciprocal

9. Cramer-Rao Lower Bound (CRLB) (8)

<u>Example</u>: CRLB for DC in AWGN

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

where $w[n] \sim N(0, \sigma^2)$ and white.

Need likelihood function:

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

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

9. Cramer-Rao Lower Bound (CRLB) (9)

Now take ln to get LLF:

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

Now take first partial with respect to A:

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

Take partial again:

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

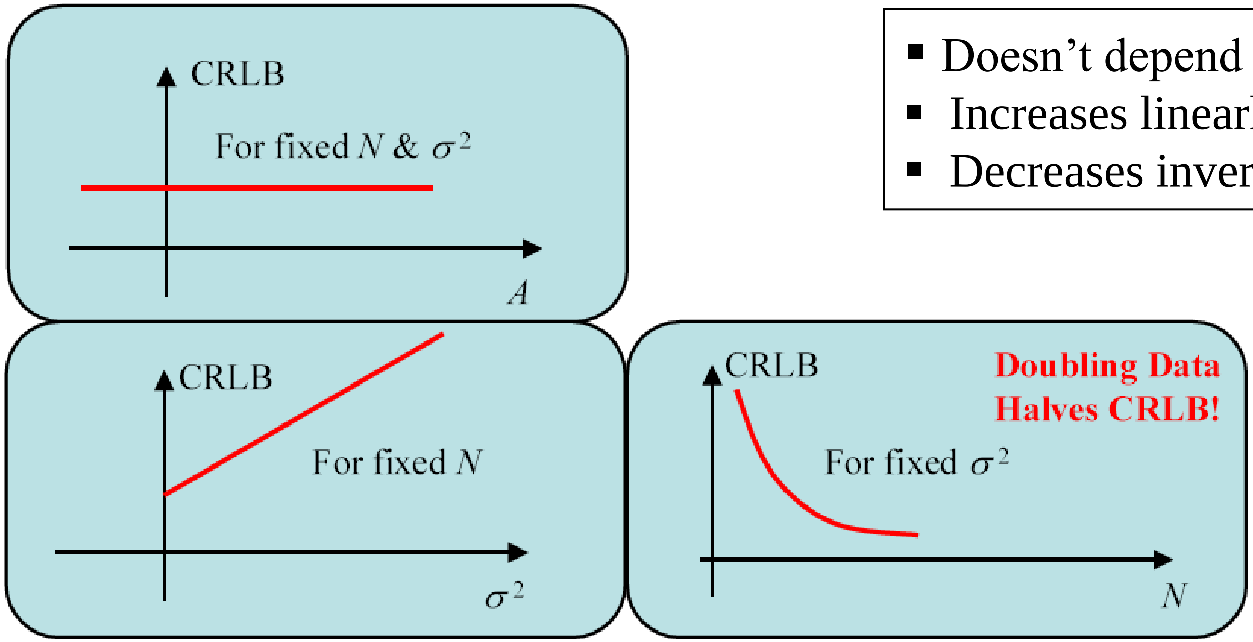
(Doesn’t depend on **x** so we don’t need to do $E\{.\}$)

9. Cramer-Rao Lower Bound (CRLB) (10)

Since the result doesn't depend on **x** or A, all we do is negate and form reciprocal to get CRLB:

$$CRLB = \frac{1}{-E\left\{\frac{\partial^2}{\partial A^2}\ln p(\mathbf{x};A)\right\}\Big|_{\theta=\text{true value}}} = \frac{\sigma^2}{N} \qquad \Rightarrow var\left\{\hat{A}\right\} \geq \frac{\sigma^2}{N}$$

- Doesn't depend on A
- Increases linearly with σ^2
- Decreases inversely with N



9. Cramer-Rao Lower Bound (CRLB) (11)

❑ <u>Continuation of theorem on CRLB</u>

There exists an unbiased estimator that <u>attains</u> the CRLB iff:

$$\frac{\partial}{\partial \theta} \ln p(\mathbf{x};\theta) = I(\theta)\left[g(\mathbf{x}) - \theta\right] \qquad (**)$$

for some functions $I(\theta)$ and $g(\mathbf{x})$.

Furthermore, the estimator that achieves the CRLB is then given by:

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

$$var\left\{\hat{\theta}\right\} = \frac{1}{I(\theta)} = CRLB$$

Since no unbiased estimator can do better, this is the <u>MVU estimate</u> !

<u>**This gives a possible way to find the MVU**</u>:

- Compute $\partial \ln p(\mathbf{x};\theta)/\partial\theta$ (need to anyway)
- Check to see if it can be put in form like (**)
- If so…then $g(\mathbf{x})$ is the MVU estimator.

9. Cramer-Rao Lower Bound (CRLB) (12)

Example: Find MVU estimate for DC level in AWGN (see example slide 27)

We found in (*) that: $$\frac{\partial}{\partial A}\ln p(\mathbf{x};A) = \frac{N}{\sigma^2}\left(\bar{x} - A\right)$$

has the form of $I(A)\left[g(\mathbf{x}) - A\right]$

where $$I(A) = \frac{N}{\sigma^2} \Rightarrow var\left\{\hat{A}\right\} = \frac{1}{I(A)} = \frac{\sigma^2}{N} = CRLB$$

and $$\hat{\theta} = g(\mathbf{x}) = \bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$$

So, for the DC Level in AWGN: the **sample mean** is the MVU Estimate !

9. Cramer-Rao Lower Bound (CRLB) (13)

- **Definition: Efficient estimator**

An estimator that is:

- unbiased and
- attains the CRLB

is said to be an **efficient estimator**

Notes:

- Not all estimators are efficient (see next example: Phase estimate)
- Not even all MVU estimators are efficient

9. Cramer-Rao Lower Bound (CRLB) (14)

Example: CRLB for Phase Estimation

This is related to the DSB carrier estimation problem (slide 3), except here, we have a pure sinusoid and we only wish to estimate only its phase.

Signal model: $$x[n] = \underbrace{A\cos\left(2\pi f_0 n + \phi_0\right)}_{s[n;\phi_0]} + w[n]$$

where $w[n]$ is AWGN with zero mean and variance σ^2.

Signal-to-Noise Ratio (SNR):

Signal Power = $A^2/2$, Noise Power = $\sigma^2 \Rightarrow \quad SNR = \dfrac{A^2}{2\sigma^2}$

Assumptions:

- $0 < f_0 < 1/2$ (f_0 is in cycles/sample)
- A and f_0 are known (we'll remove this assumption later)

9. Cramer-Rao Lower Bound (CRLB) (15)

<u>Problem</u>: Find the CRLB for estimating the phase.

We need the PDF:

$$p(\mathbf{x};\phi) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[\frac{-\sum_{n=0}^{N-1}\left(x[n] - A\cos\left(2\pi f_0 n + \phi_0\right)\right)^2}{2\pi\sigma^2} \right]$$

Now taking the log gets rid of the exponential, then taking partial derivative gives (see [2] for details):

$$\frac{\partial \ln p(\mathbf{x};\phi)}{\partial \phi} = \frac{-A}{\sigma^2}\sum_{n=0}^{N-1}\left(x[n]\sin\left(2\pi f_0 n + \phi_0\right) - \frac{A}{2}\sin\left(4\pi f_0 n + 2\phi_0\right)\right)^2$$

Taking partial derivative again:

$$\frac{\partial^2 \ln p(\mathbf{x};\phi)}{\partial \phi^2} = \frac{-A}{\sigma^2}\sum_{n=0}^{N-1}\left(x[n]\cos\left(2\pi f_0 n + \phi_0\right) - A\cos\left(4\pi f_0 n + 2\phi_0\right)\right)$$

9. Cramer-Rao Lower Bound (CRLB) (16)

Taking the expected value:

$$-E\left\{\frac{\partial^2 \ln p(\mathbf{x};\phi)}{\partial \phi^2}\right\} = E\left\{\frac{A}{\sigma^2}\sum_{n=0}^{N-1}\Big(x[n]\cos\left(2\pi f_0 n+\phi_0\right)-A\cos\left(4\pi f_0 n+2\phi_0\right)\Big)\right\}$$

$$= \frac{A}{\sigma^2}\sum_{n=0}^{N-1}\Big(E\left\{x[n]\right\}\cos\left(2\pi f_0 n+\phi_0\right)-A\cos\left(4\pi f_0 n+2\phi_0\right)\Big)$$

So, plug that in, get a $\cos^2$ term, use trigonometric identity, and get

$$-E\left\{\frac{\partial^2 \ln p(\mathbf{x};\phi)}{\partial \phi^2}\right\} = \frac{A}{\sigma^2}\left[\sum_{n=0}^{N-1}1-\sum_{n=0}^{N-1}\cos\left(4\pi f_0 n+2\phi_0\right)\right] \approx \frac{NA^2}{2\sigma^2} = N\times SNR$$

9. Cramer-Rao Lower Bound (CRLB) (17)

- **CRLB for signal in AWGN:**

When we have the case that our data is "signal + AWGN" then we get a simple form for the CRLB:

Signal model: $x[n] = s[n;\theta] + w[n]$, $n = 0, 1, 2, \dots, N\text{-}1$

where: $w[n]$ is white, Gaussian and zero mean.

First write the likelihood function:

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

Differentiate Log of likelihood function twice to get:

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

9. Cramer-Rao Lower Bound (CRLB) (18)

Since this equation depends on random $x[n]$, so we must take $E\{\}$:

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

Therefore, CRLB for signal in AWGN:

$$var\{\hat{\theta}\} \geq \frac{\sigma^2}{\sum_{n=0}^{N-1}\left[\frac{\partial s[n;\theta]}{\partial\theta}\right]^2}$$

- **Transformation of parameters:** There is a parameter θ with known $CRLB_\theta$, but now we are interested in estimating some other parameter α that is a function of θ: $\alpha = g(\theta)$. What is $CRLB_\alpha$?

$$var\{\alpha\} \geq CRLB_\alpha = \left(\frac{\partial g(\theta)}{\partial\theta}\right)^2 CRLB_\theta$$

9. Cramer-Rao Lower Bound (CRLB) (19)

- **CRLB for vector parameter:**

Vector parameter: $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_p]^T$

Its estimate: $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_p]^T$

Assume that estimate is unbiased: $E\left\{\hat{\boldsymbol{\theta}}\right\} = \boldsymbol{\theta}$

For a scalar parameter we looked at its variance, but for a vector parameter we look at its covariance matrix:

$$var\left\{\hat{\boldsymbol{\theta}}\right\} = E\left\{\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right]\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right]^T\right\} = \mathbf{C}_{\hat{\boldsymbol{\theta}}}$$

Example: For $\boldsymbol{\theta} = [x\ y\ z]^T$ then

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \begin{bmatrix} var(\hat{x}) & cov(\hat{x}, \hat{y}) & cov(\hat{x}, \hat{z}) \\ cov(\hat{y}, \hat{x}) & var(\hat{y}) & cov(\hat{y}, \hat{z}) \\ cov(\hat{z}, \hat{x}) & cov(\hat{z}, \hat{y}) & var(\hat{z}) \end{bmatrix}$$

9. Cramer-Rao Lower Bound (CRLB) (20)

- **<u>Fisher Information Matrix</u>**: For the vector parameter case, Fisher Info becomes the Fisher Info Matrix (FIM) $\mathbf{I}(\boldsymbol{\theta})$, whose *mn*th element is given by:

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{mn} = -E\left\{\frac{\partial^2 \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_n\partial\theta_n}\right\}, m, n = 1, 2, \cdots, p$$

  The CRLB Matrix: Then, under the same kind of regularity conditions, the CRLB matrix is the inverse of the FIM: $CRLB = \mathbf{I}^{-1}(\boldsymbol{\theta})$

  It means: $\sigma^2_{\hat{\theta}_n} = \left[\mathbf{C}_{\hat{\boldsymbol{\theta}}}\right]_{nn} \geq \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{nn}$

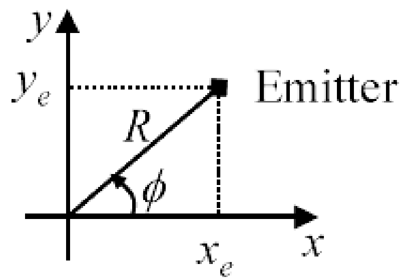
$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \begin{bmatrix} \mathrm{var}(\hat{x}) & \mathrm{cov}(\hat{x},\hat{y}) & \mathrm{cov}(\hat{x},\hat{z}) \\ \mathrm{cov}(\hat{y},\hat{x}) & \mathrm{var}(\hat{y}) & \mathrm{cov}(\hat{y},\hat{z}) \\ \mathrm{cov}(\hat{z},\hat{x}) & \mathrm{cov}(\hat{z},\hat{y}) & \mathrm{var}(\hat{z}) \end{bmatrix} \qquad \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \mathbf{I}^{-1}(\boldsymbol{\theta})$$

9. Cramer-Rao Lower Bound (CRLB) (21)

❑ **<u>Vector Transformations</u>:** Just like for the scalar case $\alpha = g(\theta)$, if you know $CRLB_\theta$ you can find $CRLB_\alpha$

$$\mathbf{CRLB}_{\boldsymbol{\alpha}} = \underbrace{\left[\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]}_{Jacobian \quad Matrix} \underbrace{\mathbf{I}^{-1}(\boldsymbol{\theta})}_{CRLB_\theta} \left[\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]^T$$

<u>Example</u>: Can estimate Range (R) and Bearing (ϕ) directly, but might really want to estimate emitter location (x_e, y_e)



9. Cramer-Rao Lower Bound (CRLB) (22)

Direct Parameters

$$\boldsymbol{\theta} = \begin{bmatrix} R \\ \phi \end{bmatrix}$$

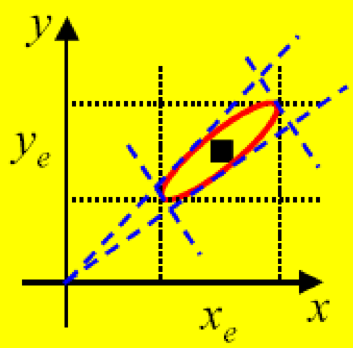
$$\boldsymbol{\alpha} = \begin{bmatrix} x_e \\ y_e \end{bmatrix} = g(\boldsymbol{\theta}) = \begin{bmatrix} R\cos\phi \\ R\sin\phi \end{bmatrix}$$

Mapped Parameters

Jacobian Matrix

$$\frac{\partial g(\theta)}{\partial \theta} = \begin{bmatrix} \dfrac{\partial R\cos\phi}{\partial R} & \dfrac{\partial R\cos\phi}{\partial \phi} \\ \dfrac{\partial R\sin\phi}{\partial R} & \dfrac{\partial R\cos\phi}{\partial \phi} \end{bmatrix}$$

$$= \begin{bmatrix} \cos\phi & -R\sin\phi \\ \sin\phi & R\cos\phi \end{bmatrix}$$



$$\boxed{\mathbf{CRLB}_{\boldsymbol{\alpha}} = \left[\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] \mathbf{CRLB}_{\boldsymbol{\theta}} \left[\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]^T}$$

9. Cramer-Rao Lower Bound (CRLB) (23)

- **CRLB for general Gaussian case:**
  - On slide 36 we saw the CRLB for "signal + AWGN", where signal is deterministic with scalar deterministic parameter. The PDF's parameter-dependence showed up only in the mean of the PDF.
  - Now generalize to the case where: $\mathbf{x} \sim N(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta}))$

    Data is still Gaussian, but

    Parameter-dependence not restricted to mean,

    Noise not restricted to white, covariance matrix not necessarily diagonal.
  - One way to get this case: "signal + AGN", where signal is random Gaussian signal with vector deterministic parameter, AGN is non-white noise. For this case the FIM is given by:

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i}\right]^T \mathbf{C}^{-1}(\boldsymbol{\theta}) \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j}\right] + \frac{1}{2} trace\left[\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_j}\right]$$

9. Cramer-Rao Lower Bound (CRLB) (24)

- **CRLB Examples:**

  1. Range Estimation
     – sonar, radar, robotics, emitter location.

  2. Sinusoidal Parameter Estimation (Amp., Frequency, Phase)
     – sonar, radar, communication receivers, etc.

  3. Bearing Estimation
     – sonar, radar, emitter location.

  4. Autoregressive Parameter Estimation
     – speech processing.

  We'll now apply the CRLB theory to several examples of practical signal processing problems.

9. Cramer-Rao Lower Bound (CRLB) (25)

Example 1: **Range estimation problem**

Transmit Pulse: $s(t)$ nonzero over $t \in [0, T_s]$

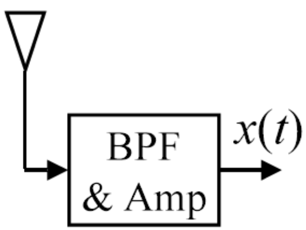
Receive Reflection: $s(t - \tau_o)$

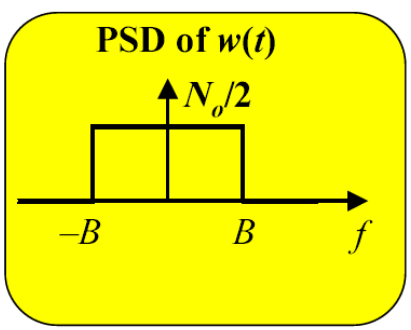
Measure Time Delay: τ_o

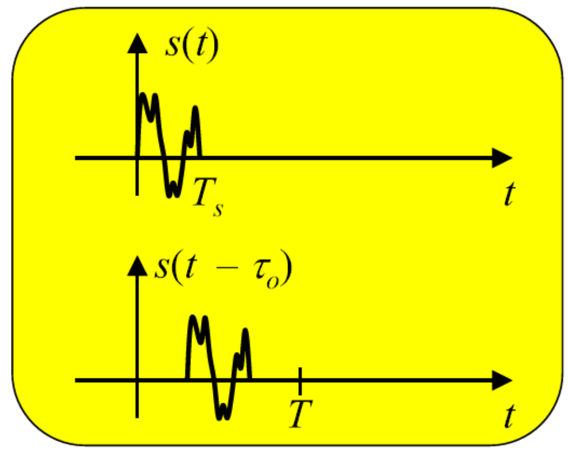
C-T Signal Model

$$x(t) = \underbrace{s(t - \tau_o)}_{s(t;\tau_o)} + w(t) \qquad 0 \le t \le T = T_s + \tau_{o,\max}$$

Bandlimited White Gaussian

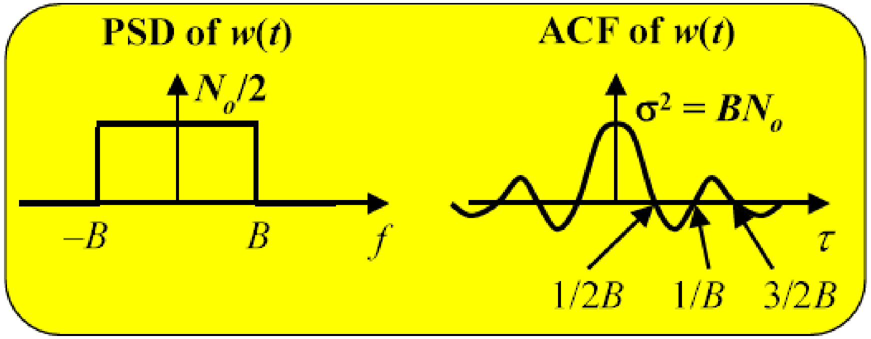






9. Cramer-Rao Lower Bound (CRLB) (26)

Range estimation D-T (Descrete-Time) Signal Model:



Sample Every $\Delta = 1/2B$ sec

$w[n] = w(n\Delta)$

DT White Gaussian Noise Var $\sigma^2 = BN_o$

$$x[n] = \underbrace{s(n\Delta - \tau_o)}_{s[n;\tau_o]} + w[n] \qquad n = 0, 1, \ldots, N-1$$

$s[n;\tau_o]$… has M non-zero samples starting at n_o

$$\boxed{x[n] = \begin{cases} w[n] & 0 \le n \le n_o - 1 \\ s(n\Delta - \tau_o) + w[n] & n_o \le n \le n_o + M - 1 \\ w[n] & n_o + M \le n \le N - 1 \end{cases}}$$

9. Cramer-Rao Lower Bound (CRLB) (27)

Range Estimation CRLB: Applying CRLB result for "signal + WGN:

Plug in… and keep non-zero terms

$$\mathrm{var}(\hat{\tau}_o) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1}\left(\frac{\partial s[n;\tau_o]}{\partial \tau_o}\right)^2} = \frac{\sigma^2}{\sum_{n=n_o}^{n_o+M-1}\left(\frac{\partial s(n\Delta-\tau_o)}{\partial \tau_o}\right)^2}$$

$$= \frac{\sigma^2}{\sum_{n=n_o}^{n_o+M-1}\left(\left.\frac{\partial s(t)}{\partial t}\right|_{t=n\Delta-\tau_o}\right)^2} = \frac{\sigma^2}{\sum_{n=0}^{M-1}\left(\left.\frac{\partial s(t)}{\partial t}\right|_{t=n\Delta}\right)^2}$$

Exploit Calculus!!!

Use approximation: $\tau_o = \Delta\, n_o$
Then do change of variables!!

9. Cramer-Rao Lower Bound (CRLB) (28)

Assuming sample spacing is small, approximate sum by integral:

$$\operatorname{var}(\hat{\tau}_o) \geq \frac{\sigma^2}{\frac{1}{\Delta}\int_0^{T_s}\left(\frac{\partial s(t)}{\partial t}\right)^2 dt} = \frac{N_o/2}{\int_0^{T_s}\left(\frac{\partial s(t)}{\partial t}\right)^2 dt} = \frac{1}{\frac{E_s}{N_o/2}\frac{\int_0^{T_s}\left(\frac{\partial s(t)}{\partial t}\right)^2 dt}{E_s}}$$

$$E_s = \int_0^{T_s} s^2(t)dt$$

FT Theorem & Parseval

$$\operatorname{var}(\hat{\tau}_o) \geq \frac{1}{\frac{E_s}{N_o/2}\frac{\int_0^{T_s}(2\pi f)^2|S(f)|^2 df}{E_s}}$$

$$= \frac{1}{\frac{E_s}{N_o/2}\frac{\int_{-\infty}^{\infty}(2\pi f)^2|S(f)|^2 df}{\int_{-\infty}^{\infty}|S(f)|^2 dt}}$$

Parseval

A type of "SNR"

Define a BW measure:

$$B_{rms} = \sqrt{\frac{\int_{-\infty}^{\infty}(2\pi f)^2|S(f)|^2 df}{\int_{-\infty}^{\infty}|S(f)|^2 dt}}$$

B_{rms} is "RMS BW" (Hz)

9. Cramer-Rao Lower Bound (CRLB) (29)

Therefore, the CRLB on the delay can be written as

$$var\left\{\hat{\tau}_0\right\} \geq \frac{1}{SNR \times B_{rms}^2}, (\sec^2)$$

To obtain the CRLB on the range, using $R = c\tau_0/2$, then

$$var\left\{\hat{R}\right\} \geq CRLB_{\hat{R}} = \left(\frac{\partial R}{\partial \tau_0}\right)^2 CRLB_{\hat{\tau}_0} = \frac{c^2/4}{SNR \times B_{rms}^2}, (m^2)$$

CRLB is inversely proportional to:

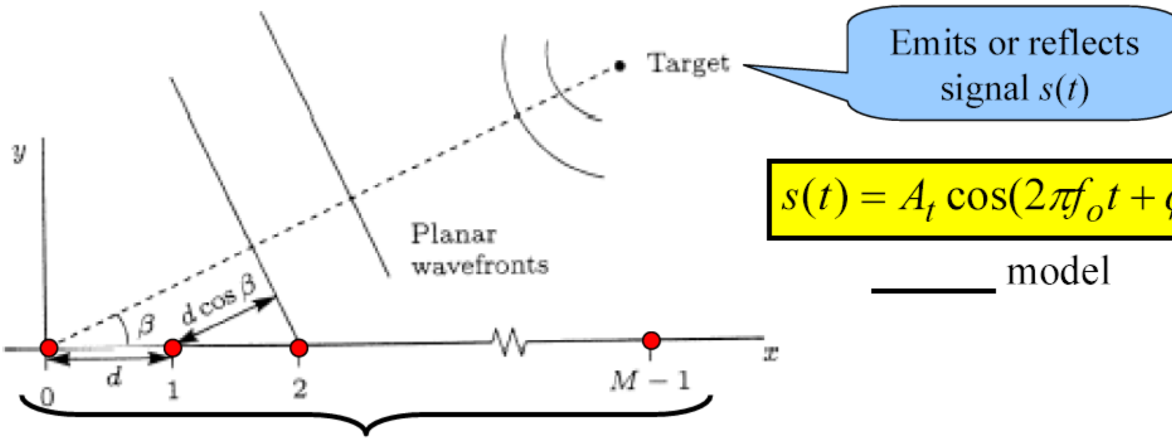
- SNR Measure
- RMS BW Measure.

Therefore, the CRLB tells us:

- Choose signal with large B_{rms}
- Ensure that *SNR* is large.
- Better on nearby/large targets.
- Which is better?
  - – Double transmitted energy?
  - – Double RMS bandwidth?

9. Cramer-Rao Lower Bound (CRLB) (30)

<u>Example 2</u>: **Bearing estimation CRLB problem**



$$\boxed{s(t) = A_t \cos(2\pi f_o t + \phi)}$$

______ model

Uniformly spaced linear array with *M* sensors:
- **Sensor Spacing of *d* meters**
- **Bearing angle to target β radians**

<u>Propagation Time to n^{th} Sensor</u>: $t_n = t_0 - n\dfrac{d}{c}\cos\beta \qquad n = 0, 1, \ldots, M-1$

<u>Signal at n^{th} Sensor</u>:

$$\boxed{\begin{aligned} s_n(t) &= \alpha s(t - t_n) \\ &= A\cos\left(2\pi f_o\left(t - t_0 + n\frac{d}{c}\cos\beta\right) + \phi\right) \end{aligned}}$$

9. Cramer-Rao Lower Bound (CRLB) (31)

<u>Bearing estimation snapshot of sensor signals:</u> Now instead of sampling each sensor at lots of time instants, we just grab one "snapshot" of all M sensors at a single instant t_s

$$s_n(t_s) = A\cos\left(2\pi f_o\left(t_s - t_0 + n\frac{d}{c}\cos\beta\right) + \phi\right)$$

$$= A\cos\left(\underbrace{\left(\underbrace{\left(\frac{2\pi f_o}{c}\cos\beta\right)}_{\omega_s} d\right)}_{\Omega_s} n + \tilde{\phi}\right) = A\cos\left(\Omega_s n + \tilde{\phi}\right)$$

Spatial sinusoid w/ spatial frequency Ω_s

For sinusoidal transmitted signal… Bearing Est. reduces to Frequency Est. And… we already know its FIM & CRLB!!!

<u>Spatial Frequencies</u>:
- ω_s is in rad/meter
- Ω_s is in rad/sensor

9. Cramer-Rao Lower Bound (CRLB) (32)

Bearing estimation data and parameters: Each sample in the snapshot is corrupted by a noise sample, and these M samples make the data vector:

$$\mathbf{x} = [x[0], x[1], \ldots, x[M\text{-}1]\ ]$$

$$\boxed{x[n] = s_n(t_s) + w[n] = A\cos\left(\Omega_s n + \tilde{\phi}\right) + w[n]}$$

Each $w[n]$ is a noise sample that comes from a different sensor so…
Model as uncorrelated Gaussian RVs (same as white temporal noise)
Assume each sensor has same noise variance σ^2

So… the parameters to consider are: $\boxed{\boldsymbol{\theta} = [A \quad \Omega_s \quad \hat{\phi}]^T}$

which get transformed to: $\boxed{\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta}) = \begin{bmatrix} A \\ \beta \\ \hat{\phi} \end{bmatrix} = \begin{bmatrix} A \\ \arccos\left(\dfrac{c\Omega_s}{2\pi f_o d}\right) \\ \hat{\phi} \end{bmatrix}}$

Parameter of interest!

9. Cramer-Rao Lower Bound (CRLB) (33)

Bearing estimation CRLB result: Using the FIM for the sinusoidal parameter proble, together with the transformation of parameters result (see p. 59, [2]):

$$\operatorname{var}(\hat{\beta}) \geq \frac{12}{(2\pi)^2 SNR \times M \dfrac{M+1}{M-1}\left(\dfrac{L}{\lambda}\right)^2 \sin^2(\beta)} \quad (\text{rad}^2)$$

Define: $L_r = L/\lambda$
Array Length "in wavelengths"

L = Array physical length in meters
M = Number of array elements
$\lambda = c/f_o$ Wavelength in meters (per cycle)

- **Bearing Accuracy:**
  - **Decreases as $1/SNR$**
  - **Decreases as $1/M$**
  - **Decreases as $\boxed{1/L_r^2}$**
  - **Depends on actual bearing β**
    - **Best at $\beta = \pi/2$ ("Broadside")**
    - **Impossible at $\beta = 0$! ("Endfire")**

Low-frequency (i.e., long wavelength) signals need very large physical lengths to achieve good accuracy

9. Cramer-Rao Lower Bound (CRLB) (34)

Example 3: **Sinusoid estimation CRLB problem**

Given discrete-time signal samples of a sinusoid in noise, estimate its amplitude, frequency, and phase.

$$\boxed{x[n] = A\cos(\Omega_o n + \phi) + w[n] \qquad n = 0, 1, \ldots, N-1}$$

Ω_o is DT frequency in cycles/sample: $0 < \Omega_o < \pi$

DT White Gaussian Noise
Zero Mean & Variance of σ2

Multiple parameters, so parameter vector: $\boldsymbol{\theta} = \begin{bmatrix} A & \Omega_0 & \phi \end{bmatrix}^T$

Recall that SNR of sinusoid in noise is:

$$SNR = \frac{P_s}{P_n} = \frac{A^2}{2\sigma^2}$$

9. Cramer-Rao Lower Bound (CRLB) (35)

Sinusoid estimation CRLB approach:

- Find Fisher Info Matrix
- Invert to get CRLB matrix
- Look at diagonal elements to get bounds on parameter variances

Recall: Result for FIM for general Gaussian case specialized to signal in AWGN case:

$$\boxed{\begin{aligned}[\mathbf{I}(\boldsymbol{\theta})]_{ij} &= \frac{1}{\sigma^2}\left(\frac{\partial \mathbf{s}_{\boldsymbol{\theta}}}{\partial \theta_i}\right)\left(\frac{\partial \mathbf{s}_{\boldsymbol{\theta}}}{\partial \theta_j}\right)^T \\ &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\frac{\partial s[n;\boldsymbol{\theta}]}{\partial \theta_i}\frac{\partial s[n;\boldsymbol{\theta}]}{\partial \theta_j}\end{aligned}}$$

9. Cramer-Rao Lower Bound (CRLB) (36)

Sinusoid estimation Fisher info elements: Taking the partial derivatives and using approximations given in [2] (valid when Ω_0 is not near 0 or π) :

$$[\mathbf{I}(\boldsymbol{\theta})]_{11} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\cos^2(\Omega_o n + \phi) = \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(1 + \cos(2\Omega_o n + 2\phi)\right) \approx \frac{N}{2\sigma^2}$$

$$[\mathbf{I}(\boldsymbol{\theta})]_{12} = [\mathbf{I}(\boldsymbol{\theta})]_{21} = \frac{-1}{\sigma^2}\sum_{n=0}^{N-1} An\cos(\Omega_o n + \phi)\sin(\Omega_o n + \phi) = \frac{-A}{2\sigma^2}\sum_{n=0}^{N-1} n\sin(2\Omega_o n + 2\phi) \approx 0$$

$$[\mathbf{I}(\boldsymbol{\theta})]_{13} = [\mathbf{I}(\boldsymbol{\theta})]_{31} = \frac{-1}{\sigma^2}\sum_{n=0}^{N-1} A\cos(\Omega_o n + \phi)\sin(\Omega_o n + \phi) = \frac{-A}{2\sigma^2}\sum_{n=0}^{N-1}\sin(2\Omega_o n + 2\phi) \approx 0$$

$$[\mathbf{I}(\boldsymbol{\theta})]_{22} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1} A^2(n)^2\sin^2(\Omega_o n + \phi) = \frac{A^2}{2\sigma^2}\sum_{n=0}^{N-1} n^2\left(1 - \cos(2\Omega_o n + 2\phi)\right) \approx \frac{A^2}{2\sigma^2}\sum_{n=0}^{N-1} n^2$$

$$[\mathbf{I}(\boldsymbol{\theta})]_{23} = [\mathbf{I}(\boldsymbol{\theta})]_{32} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1} A^2 n\sin^2(\Omega_o n + \phi) \approx \frac{A^2}{2\sigma^2}\sum_{n=0}^{N-1} n$$

$$[\mathbf{I}(\boldsymbol{\theta})]_{33} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1} A^2\sin^2(\Omega_o n + \phi) \approx \frac{NA^2}{2\sigma^2}$$

9. Cramer-Rao Lower Bound (CRLB) (37)

Fisher Info Matrix then is:

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

9. Cramer-Rao Lower Bound (CRLB) (38)

Sinusoid estimation CRLBs: Inverting the FIM by hand gives the CRLB matrix and then extracting the diagonal elements gives the three bounds:

$$\operatorname{var}(\hat{A}) \geq \frac{2\sigma^2}{N} \quad (\text{volts}^2)$$

$$\operatorname{var}(\hat{\Omega}_o) \geq \frac{12}{SNR \times N(N^2-1)} \quad ((\text{rad/sample})^2)$$

To convert to Hz²
multiply by $(F_s/2\pi)^2$

$$\operatorname{var}(\hat{\phi}) \geq \frac{2(2N-1)}{SNR \times N(N+1)} \approx \frac{4}{SNR \times N} \quad (\text{rad}^2)$$

- **Amp. Accuracy: Decreases as $1/N$, Depends on Noise Variance (not SNR)**
- **Freq. Accuracy: Decreases as $1/N^3$, Decreases as as $1/SNR$**
- **Phase Accuracy: Decreases as $1/N$, Decreases as as $1/SNR$**

9. Linear Models (1)

❑ General Linear Model

Recall signal + WGN case: $x[n] = s[n; \boldsymbol{\theta}] + w[n]$

In general, it is dependent on $\boldsymbol{\theta}$: $\mathbf{x} = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{w}$

Now we consider a special case: Linear "Observations":

$$\mathbf{s}(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\theta} + \mathbf{b}$$

$N\times 1$

known "observation matrix" $(N\times p)$

$p\times 1$

known "offset" $(p\times 1)$

The General Linear Model:

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

Data Vector

Known & Full Rank

To Be Estimated

Known

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

zero-mean, Gaussian, $\mathbf{C}$ is pos. def.

Note: "Gaussian" is part of the "Linear Model"

9. Linear Models (2)

We must assume **H** is full rank:

If **H** is not full rank, then for any **s**: $\exists\ \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ such that $\mathbf{s} = \mathbf{H}\,\boldsymbol{\theta}_1 = \mathbf{H}\,\boldsymbol{\theta}_2$

- **Importance of the Linear Model**:

  There are several reasons:

  1. Some applications admit this model
  2. Nonlinear models can sometimes be linearized
  3. Finding optimal estimator is easy (see next slide)

9. Linear Models (3)

❑ MVUE for Linear Model

Theorem: The MVUE for the General Linear Model and its covariance (i.e. its accuracy performance) are given by:

$$\boxed{\hat{\boldsymbol{\theta}}_{MVU} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{b})}$$

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

and achieves the CRLB.

Proof: We'll do this for the **b** = 0 case but it can easily be done for the more general case.

First we have that $\mathbf{x} \sim N(\mathbf{H}\boldsymbol{\theta}, \mathbf{C})$ because:

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

$$\text{cov}\{\mathbf{x}\} = E\{(\mathbf{x} - \mathbf{H}\theta)\,(\mathbf{x} - \mathbf{H}\theta)^{\mathrm{T}}\} = E\{\mathbf{w}\,\mathbf{w}^{\mathrm{T}}\} = \mathbf{C}$$

9. Linear Models (4)

Recalling CRLB theorem, look at the partial of LLF:

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

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

$$= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\theta}}\left[\underbrace{\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}} - \underbrace{2\mathbf{x}^T\mathbf{C}^{-1}\mathbf{H}\boldsymbol{\theta}} + \underbrace{\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\boldsymbol{\theta}}\right]$$

Constant w.r.t. $\boldsymbol{\theta}$

Linear w.r.t. $\boldsymbol{\theta}$

Quadratic w.r.t. $\boldsymbol{\theta}$
(Note: $\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}$ is symmetric)

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

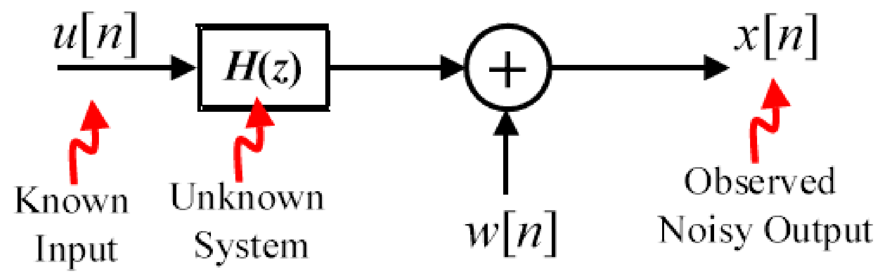
Pull out
$\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}$

$$= \underbrace{\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}}_{\mathbf{I}(\boldsymbol{\theta})}\left[\underbrace{\left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}}_{g(\boldsymbol{\theta})=\hat{\boldsymbol{\theta}}} - \boldsymbol{\theta}\right]$$

The "CRLB Theorem" says that if we have this form we have found the MVU <u>and</u> it achieves the CRLB of $\mathbf{I}^{-1}(\boldsymbol{\theta})$!!

9. Linear Models (5)

<u>Example</u>: **System Identification**



Goal: Determine a model for the system

In many applications: assume that the system is FIR (length p)

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

unknown, but here we'll assume known

Measured

Estimation Parameters

Known Input
Assume $u[n] = 0$, $n < 0$

AWGN

9. Linear Models (6)

Write FIR convolution in matrix form:

$$
\mathbf{x} = \underbrace{\begin{bmatrix}
u[0] & 0 & 0 & \cdots & \cdots & 0 \\
u[1] & u[0] & 0 & \cdots & \cdots & 0 \\
u[2] & u[1] & u[0] & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & & \ddots & \ddots & \ddots & 0 \\
\vdots & & & & \ddots & u[0] \\
\vdots & & & & & u[1] \\
\vdots & & & & & \vdots \\
u[N-1] & \cdots & \cdots & \cdots & \cdots & u[N-p]
\end{bmatrix}}_{\mathbf{H}\,(N x p)}
\underbrace{\begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[p-1] \end{bmatrix}}_{\boldsymbol{\theta}} + \mathbf{w}
$$

Measured Data

Known Input Signal Matrix

Estimate This

The Theorem for the Linear Model says: $\boxed{\hat{\boldsymbol{\theta}}_{MVU} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}}$

and achieves the CRLB. $\boxed{\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2\left(\mathbf{H}^T\mathbf{H}\right)^{-1}}$

9. Best Linear Unbiased Estimate (BLUE) (1)

- **<u>Motivation for BLUE</u>:**

  Except for Linear Model case, the optimal MVU estimator might:

  1. not even exist
  2. be difficult or impossible to find

  $\Rightarrow$ Resort to a **sub-optimal estimate**.

  BLUE is one such **sub-optimal estimate**.

  <u>Idea for BLUE</u>:

  1. Restrict estimate to be linear in data **x**
  2. Restrict estimate to be unbiased
  3. Find the best one (i.e. with minimum variance)

  <u>Advantage of BLUE</u>: Needs only mean and covariance of PDF.

  <u>Disadvantages of BLUE</u>:

  1. Sub-optimal (in general)
  2. Sometimes totally inappropriate

9. Best Linear Unbiased Estimate (BLUE) (2)

- **Definition of BLUE (scalar case)**

  Observed Data: $\mathbf{x} = [x[0]\ x[1]\ .\ .\ .\ x[N-1]\ ]^T$

  PDF: $p(x; \boldsymbol{\theta})$ depends on unknown $\boldsymbol{\theta}$

  BLUE constrained to be linear in data:

$$\boxed{\hat{\theta}_{BLU} = \sum_{n=0}^{N-1} a_n x[n] = \mathbf{a}^T \mathbf{x}}$$

  Choose **a**'s to give:

  1. unbiased estimator
  2. then minimize variance

- **Finding The BLUE (scalar case)**

  1. Constrain to be linear: $\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n]$
  2. Constrain to be unbiased: $E\left\{\hat{\theta}\right\} = \theta \Rightarrow \sum_{n=0}^{N-1} a_n E\left\{x[n]\right\} = \theta$

9. Best Linear Unbiased Estimate (BLUE) (3)

Finding BLUE for **scalar linear observations**:

Consider scalar-parameter linear observation:

$$x[n] = \theta s[n] + w[n] \;\Rightarrow E\{x[n]\} = \theta s[n]$$

where $s[n]$'s are known. Then for the unbiased condition we need:

$$E\left\{\hat{\theta}\right\} = \theta \sum_{n=0}^{N-1} a_n s[n] = \theta$$

$$\Downarrow$$

$$\boxed{\begin{cases} \mathbf{a}^T\mathbf{s} = 1 \\ \hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] \end{cases}}$$

Now, given that these constraints are met, we need to minimize the variance. Given that **C** is the covariance matrix of **x** we have (p. 136, [2]):

$$var\left\{\mathbf{a}^T\mathbf{x}\right\} = \mathbf{a}^T\mathbf{C}\mathbf{a}$$

9. Best Linear Unbiased Estimate (BLUE) (4)

Goal: minimize $\mathbf{a}^T\,\mathbf{C}\mathbf{a}$ subject to $\mathbf{a}^T\,\mathbf{s} = 1 \Rightarrow$ Constrained optimization.

From Appendix 6A, pp. 151-152, [2]: Use Lagrangian Multipliers: Minimize

$$J = \mathbf{a}^T\,\mathbf{C}\mathbf{a} + \lambda(\mathbf{a}^T\,\mathbf{s} - 1)$$

$$\text{Set}: \quad \frac{\partial J}{\partial a} = 0 \quad \Rightarrow \quad \underbrace{\boxed{\mathbf{a} = -\frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{s}}}_{\mathbf{a}^T\mathbf{s}=1}$$

$$\Rightarrow \quad \mathbf{a}^T\mathbf{s} = -\frac{\lambda}{2}\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = 1 \quad \Rightarrow \quad \boxed{-\frac{\lambda}{2} = \frac{1}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}}$$

$$\boxed{\mathbf{a} = \frac{\mathbf{C}^{-1}\mathbf{s}}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}}$$

$$\Rightarrow \quad \boxed{\hat{\theta}_{BLUE} = \mathbf{a}^T\mathbf{x} = \frac{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{x}}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}} \qquad \boxed{\mathrm{var}(\hat{\theta}) = \frac{1}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}}$$

9. Best Linear Unbiased Estimate (BLUE) (5)

- **Applicability of BLUE**

  We just derived the BLUE under the following:

  1. Linear observations but with no constraint on the noise PDF.
  2. No knowledge of the noise PDF other than its mean and covariance.
  3. Noise does not need to be Gaussian.

  See Examples 6.1 and 6.2, pp. 137-139, [2].

9. Best Linear Unbiased Estimate (BLUE) (6)

- **Vector Parameter Case: Gauss-Markov Theorem (Linear model without assumption of Gaussian noise)**

  If data can be modeled as having linear observations in noise:

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

Known Matrix

Known Mean & Cov (PDF is otherwise arbitrary & unknown)

Then the BLUE is (pp. 139-141, [2]):

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

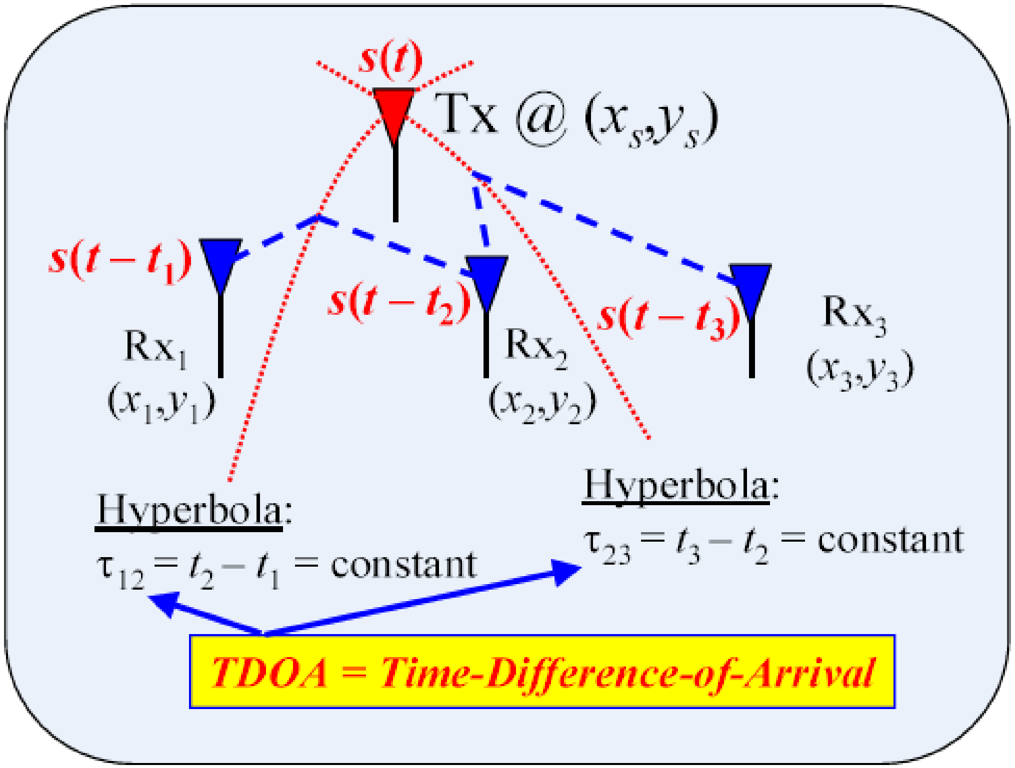
and its covariance is:

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

Note: If noise *is* Gaussian then BLUE *is* MVUE

9. Best Linear Unbiased Estimate (BLUE) (7)

Example: **TDOA-based emitter location**



Assume that the i^{th} Rx can measure its TOA: t_i

Then, from the set of TOAs, compute TDOAs

Then, from the set of TDOAs, estimate location (x_s, y_s)

9. Best Linear Unbiased Estimate (BLUE) (8)

TOA measurement model: Assume measurements (with errors) of TOAs at N receivers (only 3 shown above):

$$t_0 \,,\, t_1 \,,\, \ldots \,,\, t_{N\text{-}1}$$

TOA measurement model:

T_o = Time the signal emitted, R_i = Range from Tx to Rx$_i$

c = Speed of propagation (for EM: $c = 3\times10^8$ m/s)

$$\boxed{t_i = T_o + R_i/c + \varepsilon_i \qquad i = 0, 1, \ldots, N\text{-}1}$$

Measurement Noise $\Rightarrow$ zero-mean, variance σ^2, independent (but **PDF unknown**)
(variance determined from estimator used to estimate t_i's)

Now, we use $R_i = \sqrt{\left[(x_s - x_i)^2 + (y_s - y_i)^2\right]}$ then

$$\boxed{t_i = f(x_s, y_s) = T_o + \frac{1}{c}\sqrt{(x_s - x_i)^2 + (y_s - y_i)^2} + \varepsilon_i}$$

Nonlinear Model

9. Best Linear Unbiased Estimate (BLUE) (9)

Linearization of TOA model: So, we linearize the model so that we can apply BLUE: Assume some rough estimate is available (x_n, y_n)

$$x_s = x_n + \delta x_s \qquad y_s = y_n + \delta y_s$$

know estimate know estimate

$$\Rightarrow \quad \theta = [\delta x \ \ \delta y]^T$$

Now use truncated Taylor series to linearize R_i (x_n, y_n):

$$R_i \approx R_{n_i} + \underbrace{\frac{x_n - x_i}{R_{n_i}}}_{\triangleq A_i} \delta x_s + \underbrace{\frac{y_n - y_i}{R_{n_i}}}_{\triangleq B_i} \delta y_s$$

Known

Apply to TOA: $\tilde{t}_i = t_i - \frac{R_{n_i}}{c} = \boxed{T_o} + \frac{A_i}{c}\boxed{\delta x_s} + \frac{B_i}{c}\boxed{\delta y_s} + \varepsilon_i$

known known known

Three unknown parameters to estimate: T_o, δy_s, δy_s

9. Best Linear Unbiased Estimate (BLUE) (10)

Conversion to TDOA model:

N–1 TDOAs rather than N TOAs

TDOAs: $\tau_i = \tilde{t}_i - \tilde{t}_{i-1}, \quad i = 1, 2, \ldots, N-1$

$$= \underbrace{\frac{A_i - A_{i-1}}{c}}_{\text{known}} \delta x_s + \underbrace{\frac{B_i - B_{i-1}}{c}}_{\text{known}} \delta y_s + \underbrace{\varepsilon_i - \varepsilon_{i-1}}_{\text{correlated noise}}$$

In matrix form: $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$

$$\mathbf{x} = \begin{bmatrix} \tau_1 & \tau_2 & \cdots & \tau_{N-1} \end{bmatrix}^T \qquad \boldsymbol{\theta} = \begin{bmatrix} \delta x_s & \delta y_s \end{bmatrix}^T$$

$$\mathbf{H} = \frac{1}{c} \begin{bmatrix} (A_1 - A_0) & \vdots & (B_1 - B_0) \\ (A_2 - A_1) & \vdots & (B_2 - B_1) \\ \vdots & \vdots & \vdots \\ (A_{N-1} - A_{N-2}) & \vdots & (B_{N-1} - B_{N-2}) \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} \varepsilon_1 - \varepsilon_0 \\ \varepsilon_2 - \varepsilon_1 \\ \vdots \\ \varepsilon_{N-1} - \varepsilon_{N-2} \end{bmatrix} = \mathbf{A}\boldsymbol{\varepsilon}$$

$$\mathbf{C_w} = \operatorname{cov}\{\mathbf{w}\} = \sigma^2 \mathbf{A}\mathbf{A}^T$$

See [2] for structure of matrix **A**

9. Best Linear Unbiased Estimate (BLUE) (11)

<u>Apply BLUE to TDOA linearized model:</u>

$$\hat{\theta}_{BLUE} = \left(\mathbf{H}^T\mathbf{C}_\mathbf{w}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}_\mathbf{w}^{-1}\mathbf{x}$$

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

Dependence on σ^2 cancels out!!!

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

$$= \sigma^2\left(\mathbf{H}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{H}\right)^{-1}$$

Describes how large the location error is

9. Maximum Likelihood Estimate (MLE) (1)

- **Motivation for MLE:**

  Problems:

  1. MVUE often does not exist or can't be found (See example 7.1 in [2])
  2. BLUE may not be applicable ($\mathbf{x} \neq \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$)

  Solution: If the PDF is **known**, then **MLE can always** be used. This makes the MLE one of the **most popular practical methods**. It will be "**optimal**" for large enough data size.

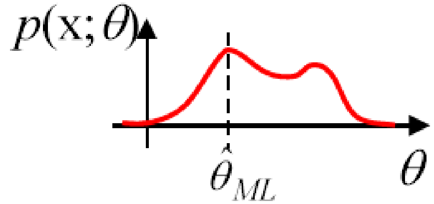
  Disadvantages:

  1. Not optimal for small data size
  2. Can be computationally complex - may require numerical methods.

9. Maximum Likelihood Estimate (MLE) (2)

❑ **Definition of the MLE:**

$\hat{\theta}_{ML}$ is the value of θ that maximizes the "Likelihood Function" $p(x; \theta)$ for the specific measured data x



$\hat{\theta}_{ML}$ maximizes the likelihood function

Note: Because ln(z) is a monotonically increasing function, $\hat{\theta}_{ML}$ maximizes the log likelihood function $\ln\{p(x; \theta)\}$.

General analytical procedure to find the MLE:

1. Find log-likelihood function: $\ln p(x;\theta)$
2. Differentiate w.r.t θ and set to 0: $\partial \ln p(x;\theta)/\partial\theta = 0$
3. Solve for θ value that satisfies the equation

9. Maximum Likelihood Estimate (MLE) (3)

Example: **DC level in WGN** (Example of MLE when MVUE non-existent)

$$x[n] = \underbrace{A}_{A>0} + \underbrace{w[n]}_{\substack{\text{WGN} \\ \sim N(0,A)}} \qquad \Rightarrow x[n] \sim N(A,A)$$

Likelihood Function: $$\underbrace{p(\mathbf{x};A) = \frac{1}{(2\pi A)^{\frac{N}{2}}} \exp\left[-\frac{1}{2A}\sum_{n=0}^{N-1}(x[n]-A)^2\right]}_{\text{To take } ln \text{ of this… use log properties:}}$$

Take $\partial/\partial A$, set $= 0$, and change A to $\hat{A}$

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

Expand this:

$$-\frac{N}{2\hat{A}} + \boxed{\frac{1}{\hat{A}}\sum x[n]} - \cancel{\frac{1}{\hat{A}}N\hat{A}} + \frac{1}{2\hat{A}}\sum x^2[n] - \boxed{\frac{1}{2\hat{A}^2}2\hat{A}\sum x[n]} + \cancel{\frac{\hat{A}^2 N}{2\hat{A}^2}} = 0$$

Cancel

9. Maximum Likelihood Estimate (MLE) (4)

Manipulate to get: $\hat{A}^2 + \hat{A} - \frac{1}{N}\sum_{n=0}^{N-1} x^2[n] = 0$

Solve quadratic equation to get MLE:

$$\hat{A}_{ML} = -\frac{1}{2} + \sqrt{\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] + \frac{1}{4}}$$

Can show this estimator biased (see bottom of p. 160)

But it is asymptotically unbiased…

Use the "Law of Large Numbers":

Sample Mean $\rightarrow$ True Mean

$$\frac{1}{N}\sum_{n=0}^{N-1} x^2[n] \xrightarrow[as\ N\to\infty]{} E\{x^2[n]\}$$

So can use this to show:

$$E\{\hat{A}_{ML}\} \rightarrow E\left\{-\frac{1}{2} + \sqrt{E\{x^2[n]\} + \frac{1}{4}}\right\} = -\frac{1}{2} + \sqrt{\underbrace{E\{x^2[n]\}}_{=A^2+A} + \frac{1}{4}} = A$$

$$\mathrm{var}(\hat{A}) \rightarrow \frac{A^2}{N\left(A + \frac{1}{2}\right)} = CRLB$$

(p. 161)

Asymptotically…Unbiased & Efficient

9. Maximum Likelihood Estimate (MLE) (5)

❑ **Properties of the MLE:**

The MLE is asymptotically:

1. unbiased
2. efficient (i.e. achieves CRLB)
3. Gaussian PDF

Also, if a truly efficient estimator exists, then the ML procedure finds it !

The asymptotic properties are captured in Theorem 7.1:

If $p(x;\theta)$ satisfies some "regularity" conditions, then the MLE is asymptotically distributed according to

$$\boxed{\hat{\theta}_{ML} \overset{a}{\sim} N(\theta, I^{-1}(\theta))}$$

where $\mathbf{I}(\theta)$ = Fisher Information Matrix

9. Maximum Likelihood Estimate (MLE) (6)

<u>Size of N to achieve asymptotic</u>:

This Theorem only states what happens asymptotically. When N is small there is no guarantee how the MLE behaves.

How large must N be to achieve the asymptotic properties?

$\Rightarrow$ In practice: use "Monte Carlo Simulations" (see Appendix 7, [2]) to answer this.

9. Maximum Likelihood Estimate (MLE) (7)

- **<u>Monte Carlo Simulations:</u>** A methodology for doing computer simulations to evaluate performance of any estimation method (not only for the MLE).

  Illustrate Monte Carlo Simulation for deterministic signal $s[n; \theta]$ in AWGN:

  - <u>Data Collection</u>:

    1. Select a particular true parameter value, θ_{true}
       - you are often interested in doing this for a variety of values of θ so you would run one MC simulation for each θ value of interest
    2. Generate signal having true θ: $s[n;\theta_t]$ (call it s in matlab)
    3. Generate WGN having unit variance

       w = randn (size(s));
    4. Form measured data: x = s + sigma*w;
       - choose σ to get the desired SNR
       - usually want to run at many SNR values

         → do one MC simulation for each SNR value

9. Maximum Likelihood Estimate (MLE) (8)

5. Compute estimate from data **x**
6. Repeat steps 3-5 *M* times

   - (call *M* "# of MC runs" or just "# of runs")
7. Store all M estimates in a vector EST (assumes scalar θ)

- <u>Statistical Evaluation:</u>

  1. Compute bias
  2. Compute error RMS
  3. Compute the error Variance
  4. Plot Histogram or Scatter Plot
     (if desired)

$$\boxed{b = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_i - \theta_{true}\right)}$$

$$\boxed{RMS = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_i - \theta_t\right)^2}}$$

$$\boxed{VAR = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_i - \left(\frac{1}{M}\sum_{i=1}^{M}\hat{\theta}_i\right)\right)^2}$$

9. Maximum Likelihood Estimate (MLE) (9)

- Now explore (via plots): How Bias, RMS, and VAR vary with θ value, SNR value, N value, etc.
  - Is $b \approx 0$?
  - Is RMS $\approx$ (CRLB)$^{1/2}$?

9. Maximum Likelihood Estimate (MLE) (10)

- **Phase estimation for a sinusoid:**

  Signal model: $x[n] = A\cos(2\pi f_o n + \varphi) + w[n]$, $n = 0, 1, \ldots, N-1$

  A and f_o known, φ unknown.

  White $\sim N(0,\sigma^2)$

Recall CRLB:

$$\mathrm{var}(\hat{\phi}) \geq \frac{2\sigma^2}{NA^2} = \frac{1}{N \cdot SNR}$$

For this problem… all methods for finding the MVUE will fail!!
$\Rightarrow$ So… try MLE!!

9. Maximum Likelihood Estimate (MLE) (11)

So first we write the **likelihood function**:

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

GOAL: Find ϕ that maximizes this

… equivalent to minimizing this

End up in same place if we maximize LLF

So, minimize: $J(\phi) \triangleq \sum_{n=0}^{N-1} \left[x[n] - A\cos\left(2\pi f_o n + \phi\right)\right]^2$ Setting $\frac{\partial J(\phi)}{\partial \phi} = 0$ gives

$$\sum_{n=0}^{N-1} x[n]\sin\left(2\pi f_o n + \hat{\phi}\right) = A \underbrace{\sum_{n=0}^{N-1} \sin\left(2\pi f_o n + \hat{\phi}\right)\cos\left(2\pi f_o n + \phi\right)}_{\approx 0}$$

sin and cos are $\perp$ when summed over full cycles

So… MLE Phase Estimate satisfies:

$$\boxed{\sum_{n=0}^{N-1} x[n]\sin\left(2\pi f_o n + \hat{\phi}\right) = 0}$$

Interpret via inner product or correlation

9. Maximum Likelihood Estimate (MLE) (12)

Now…using a Trig Identity and then re-arranging gives:

$$\cos(\hat{\phi})\left[\sum_n x[n]\sin(2\pi f_o n)\right] = -\sin(\hat{\phi})\left[\sum_n x[n]\cos(2\pi f_o n)\right]$$

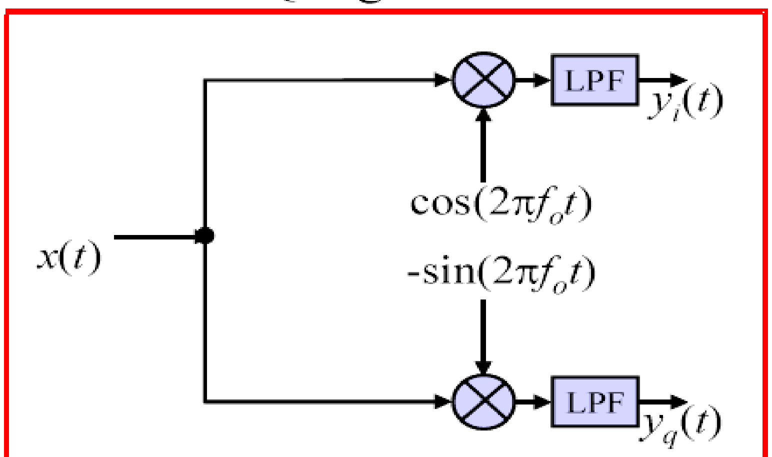
Or…

$$\boxed{\hat{\phi}_{ML} = -\tan^{-1}\left[\frac{\sum_n x[n]\sin(2\pi f_o n)}{\sum_n x[n]\cos(2\pi f_o n)}\right]}$$

Recall: This is the *approximate* MLE

Don't need to know A or σ^2 but do need to know f_o

Recall: I-Q Signal Generation



The "sums" in the above equation play the role of the LPF's in the figure (why?)
Thus, ML phase estimator can be viewed as: atan of ratio of Q/I

Monte Carlo results for ML phase estimation: See figures 7.3 & 7.4 in [2].

9. MLE Examples (1)

We'll now apply the MLE theory to several examples of practical signal processing problems. These are the same examples for the CRLB.

1. **Range estimation**

 –sonar, radar, robotics, emitter location

2. **Sinusoidal parameter estimation** (amplitude, frequency, phase)

–sonar, radar, communication receivers (recall DSB example), etc.

3. **Bearing estimation**

–sonar, radar, emitter location.

4. **Autoregressive parameter estimation**

–speech processing, econometrics.

9. MLE Examples (2)

❑ Example 1: Range estimation

Transmit Pulse: $s(t)$ nonzero over $t \in [0, T_s]$

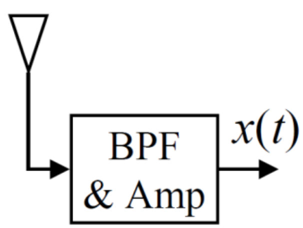
Receive Reflection: $s(t - \tau_o)$

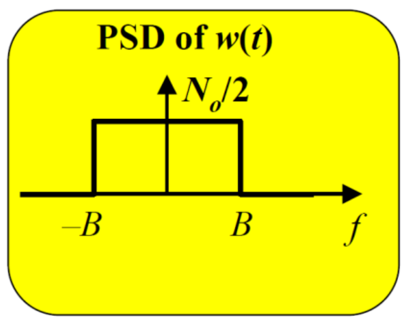
Measure Time Delay: τ_o

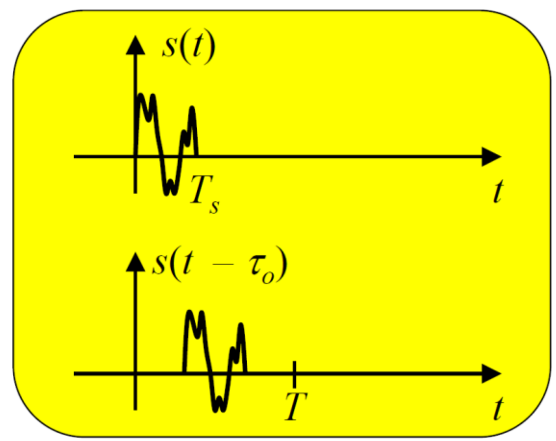
C-T Signal Model

$$x(t) = \underbrace{s(t-\tau_o)}_{s(t;\tau_o)} + w(t) \qquad 0 \le t \le T = T_s + \tau_{o,\max}$$

Bandlimited White Gaussian

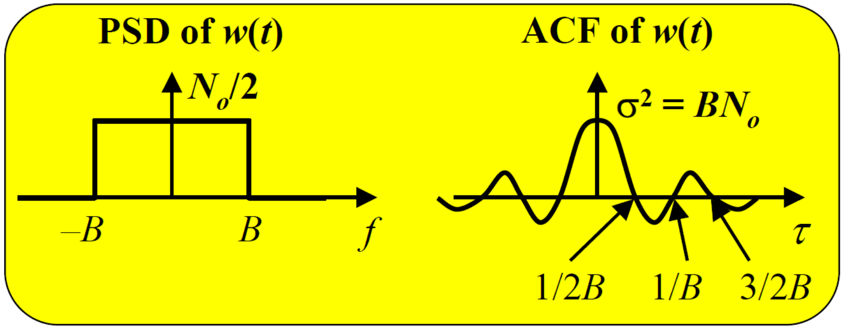






9. MLE Examples (3)

Range estimation D-T signal model:



Sample Every $\Delta = 1/2B$ sec

$$w[n] = w(n\Delta)$$

DT White Gaussian Noise Var $\sigma^2 = BN_o$

$$x[n] = \underbrace{s[n-n_o]}_{s[n;n_o]\ldots \text{ has } M \text{ non-zero samples starting at } n_o} + w[n] \qquad n = 0, 1, \ldots, N-1$$

$n_o \approx \tau_o/\Delta$

$$x[n] = \begin{cases} w[n] & 0 \le n \le n_o - 1 \\ s[n-n_o] + w[n] & n_o \le n \le n_o + M - 1 \\ w[n] & n_o + M \le n \le N - 1 \end{cases}$$

9. MLE Examples (4)

Range estimation likelihood function:

White and Gaussian $\Rightarrow$ Independent $\Rightarrow$ Product of PDFs

3 different PDFs – one for each subinterval

$$p(\mathbf{x};n_o) = \overbrace{\left[\prod_{n=0}^{n_o-1} C\exp\left[-\frac{x^2[n]}{2\sigma^2}\right]\right]}^{\#1} \bullet \overbrace{\left[\prod_{n=n_o}^{n_o+M-1} C\exp\left[-\frac{(x[n]-s[n-n_o])^2}{2\sigma^2}\right]\right]}^{\#2} \bullet \overbrace{\left[\prod_{n=n_o+M}^{N-1} C\exp\left[-\frac{x^2[n]}{2\sigma^2}\right]\right]}^{\#3}$$

Expand to get an $x^2[n]$ term… group it with the other $x^2[n]$ term

$$C = \frac{1}{\sqrt{2\pi\sigma^2}}$$

$$p(\mathbf{x};n_o) = \underbrace{C^N \exp\left[-\frac{\sum_{n=0}^{N-1} x^2[n]}{2\sigma^2}\right]}_{\text{Does not depend on } n_o} \bullet \exp\left[-\frac{1}{2\sigma^2}\underbrace{\sum_{n=n_o}^{n_o+M-1}\left(-2x[n]s[n-n_0]+s^2[n-n_o]\right)}_{\text{must minimize this or maximize its negative over values of } n_o}\right]$$

Does not depend on n_o

must <u>min</u>imize this or <u>max</u>imize its negative over values of n_o

9. MLE Examples (5)

So maximize this: $2\sum_{n=n_o}^{n_o+M-1} x[n]s[n-n_0] + \sum_{n=n_o}^{n_o+M-1} s^2[n-n_o]$

Because $s[n-n_o]=0$ outside summation range… so can extend it!

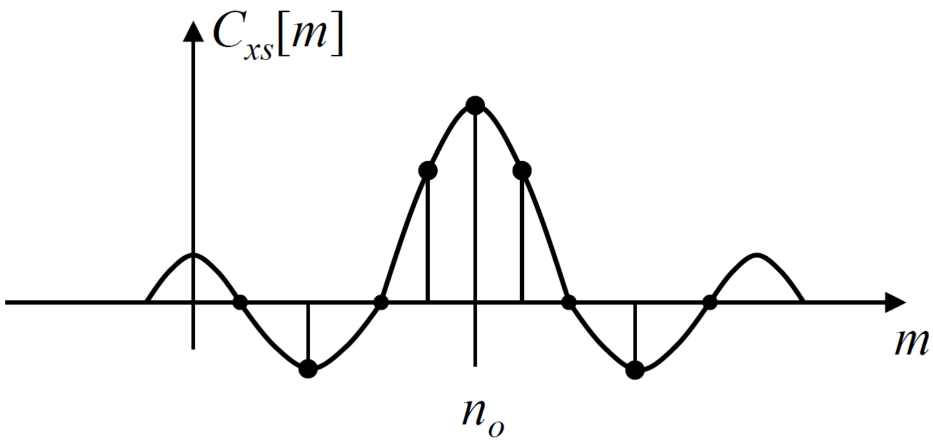
Doesn't depend on n_o! …Summand moves with the limits as n_o changes.

So maximize this: $\sum_{n=0}^{N-1} x[n]s[n-n_0]$

So…. MLE Implementation is based on Cross-correlation:
"Correlate" Received signal $x[n]$ with transmitted signal $s[n]$

$$\hat{n}_o = \underset{0\le m\le N-M}{\arg\max} \{C_{xs}[m]\} \qquad C_{xs}[m] = \sum_{n=0}^{N-1} x[n]s[n-m],$$

9. MLE Examples (6)



$$\boxed{C_{xs}[m] = \sum_{n=0}^{N-1} x[n]s[n-m],}$$

- Think of this as an inner product for each *m*.
- Compare data $x[n]$ to all possible delays of signal $s[n]$
  - pick n_o to make them **most similar**.

9. MLE Examples (7)

Example 2: Sinusoid parameter estimation

Given D-T signal samples of a sinusoid in noise ⇒ Estimate its amplitude, frequency, and phase.

$$\boxed{x[n] = A\cos(\Omega_o n + \phi) + w[n] \qquad n = 0, 1, \ldots, N-1}$$

Ω_o is DT frequency in cycles/sample: $0 < \Omega_o < \pi$

DT White Gaussian Noise
Zero Mean & Variance of σ^2

Multiple parameters… so parameter vector: $\boldsymbol{\theta} = [A \quad \Omega_o \quad \phi]^T$

The **likelihood function** is:

$$p(\mathbf{x};\boldsymbol{\theta}) = C^N \exp\left[-\frac{1}{2\sigma^2}\underbrace{\sum_{n=0}^{N-1}(x[n] - A\cos(\Omega_o n + \phi))^2}_{\triangleq J(A,\Omega_o,\phi)}\right]$$

For MLE: Minimize This

9. MLE Examples (8)

To make things easier, define an equivalent parameter set:

$$[\alpha_1 \quad \alpha_2 \quad \Omega_o\,]^T \qquad \alpha_1 = A\cos(\phi) \quad \alpha_2 = -A\sin(\phi)$$

$$\alpha = [\alpha_1 \quad \alpha_2]^T$$

Then $\quad J'(\alpha_1, \alpha_2, \Omega_o) = J(A, \Omega_o, \phi)$

Define

$$\mathbf{c}(\Omega_o) = [1 \quad \cos(\Omega_o) \quad \cos(\Omega_o 2) \quad \ldots \quad \cos(\Omega_o(N\text{-}1))]^T$$

$$\mathbf{s}(\Omega_o) = [0 \quad \sin(\Omega_o) \quad \sin(\Omega_o 2) \quad \ldots \quad \sin(\Omega_o(N\text{-}1))]^T$$

and $\quad \mathbf{H}(\Omega_o) = [\mathbf{c}(\Omega_o) \quad \mathbf{s}(\Omega_o)] \quad$ an $N \times 2$ matrix.

Then

$$J'(\alpha_1, \alpha_2, \Omega_o) = [\mathbf{x} - \mathbf{H}\,(\Omega_o)\,\alpha]^T\,[\mathbf{x} - \mathbf{H}\,(\Omega_o)\,\alpha]$$

This looks like the linear model case, except for Ω_o dependence of **$\mathbf{H}(\Omega_o)$**

9. MLE Examples (9)

Thus, for any fixed Ω_o value, the optimal $\boldsymbol{\alpha}$ estimate is

$$\hat{\boldsymbol{\alpha}} = \left[\mathbf{H}^T(\Omega_o)\mathbf{H}(\Omega_o)\right]^{-1}\mathbf{H}^T(\Omega_o)\mathbf{x}$$

Then plug that into $J'(\alpha_1, \alpha_2, \Omega_o)$:

$$\begin{aligned}
J'(\hat{\alpha}_1, \hat{\alpha}_2, \Omega_o) &= \left[\mathbf{x} - \mathbf{H}(\Omega_o)\hat{\boldsymbol{\alpha}}\right]^T\left[\mathbf{x} - \mathbf{H}(\Omega_o)\hat{\boldsymbol{\alpha}}\right] \\
&= \left[\mathbf{x}^T - \hat{\boldsymbol{\alpha}}^T\mathbf{H}^T(\Omega_o)\right]\left[\mathbf{x} - \mathbf{H}(\Omega_o)\hat{\boldsymbol{\alpha}}\right] \\
&= \mathbf{x}^T \underbrace{\left[\mathbf{I} - \mathbf{H}(\Omega_o)\left[\mathbf{H}^T(\Omega_o)\mathbf{H}(\Omega_o)\right]^{-1}\mathbf{H}^T(\Omega_o)\right]^2}_{=\mathbf{I} - \mathbf{H}(\Omega_o)\left[\mathbf{H}^T(\Omega_o)\mathbf{H}(\Omega_o)\right]^{-1}\mathbf{H}^T(\Omega_o)} \mathbf{x} \\
&= \mathbf{x}^T\mathbf{x} - \underbrace{\mathbf{x}^T\mathbf{H}(\Omega_o)\left[\mathbf{H}^T(\Omega_o)\mathbf{H}(\Omega_o)\right]^{-1}\mathbf{H}^T(\Omega_o)\mathbf{x}}_{\text{minimize w.r.t. } \Omega_o}
\end{aligned}$$

9. MLE Examples (10)

Sinusoid parameter estimation MLE procedure:

Step 1: Minimize

$$\underbrace{\mathbf{x}^T \mathbf{H}(\Omega_o)\left[\mathbf{H}^T(\Omega_o)\mathbf{H}(\Omega_o)\right]^{-1}\mathbf{H}^T(\Omega_o)\mathbf{x}}_{\text{minimize w.r.t. } \Omega_o}$$

over Ω_o to find:

$$\hat{\Omega}_o = \arg\min_{0\le\Omega_o\le\pi}\left\{\mathbf{x}^T \mathbf{H}(\Omega_o)\left[\mathbf{H}^T(\Omega_o)\mathbf{H}(\Omega_o)\right]^{-1}\mathbf{H}^T(\Omega_o)\mathbf{x}\right\}$$

Step 2: Use result of Step 1 to get

$$\hat{\boldsymbol{\alpha}} = \left[\mathbf{H}^T(\hat{\Omega}_o)\mathbf{H}(\hat{\Omega}_o)\right]^{-1}\mathbf{H}^T(\hat{\Omega}_o)\mathbf{x}$$

Step 3: Convert Step 2 result by solving

$$\begin{aligned}\hat{\alpha}_1 &= \hat{A}\cos(\hat{\phi}) \\ \hat{\alpha}_2 &= -\hat{A}\sin(\hat{\phi})\end{aligned} \qquad \text{for} \quad \hat{A} \quad \& \quad \hat{\phi}$$

9. MLE Examples (11)

Approximation:

First we look at a specific structure:

$$\mathbf{x}^T\mathbf{H}(\Omega_o)\left[\mathbf{H}^T(\Omega_o)\mathbf{H}(\Omega_o)\right]^{-1}\mathbf{H}^T(\Omega_o)\mathbf{x} =$$

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

If Ω_o is not near 0 or π, then approximately

$$\approx \begin{bmatrix} \frac{N}{2} & 0 \\ 0 & \frac{N}{2} \end{bmatrix}^{-1}$$

9. MLE Examples (12)

Then, Step 1 becomes:

$$\hat{\Omega}_o = \arg\min_{0 \le \Omega_o \le \pi} \left\{ \frac{2}{N} \left| \sum_{n=0}^{N-1} x[n] \exp(-j\Omega_o n) \right|^2 \right\} = \arg\min_{0 \le \Omega \le \pi} \left\{ \left| X(\Omega) \right|^2 \right\}$$

DTFT of Data $x[n]$

and Steps 2 & 3 become:

$$\hat{A} = \frac{2}{N} \left| X(\hat{\Omega}_o) \right|$$

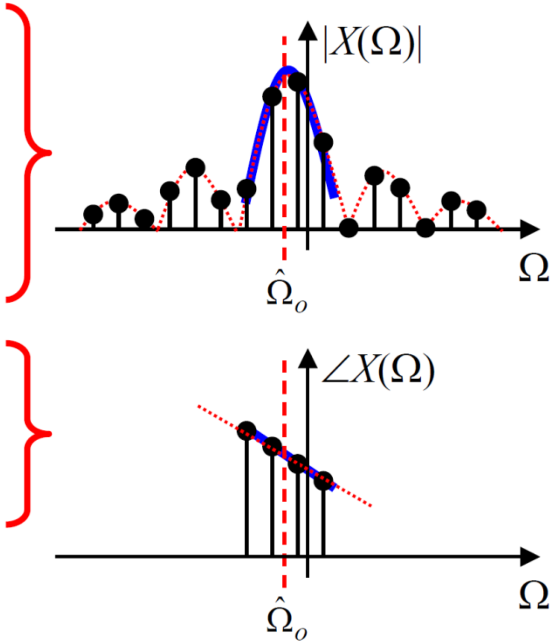
$$\hat{\phi} = \angle X(\hat{\Omega}_o)$$

9. MLE Examples (13)

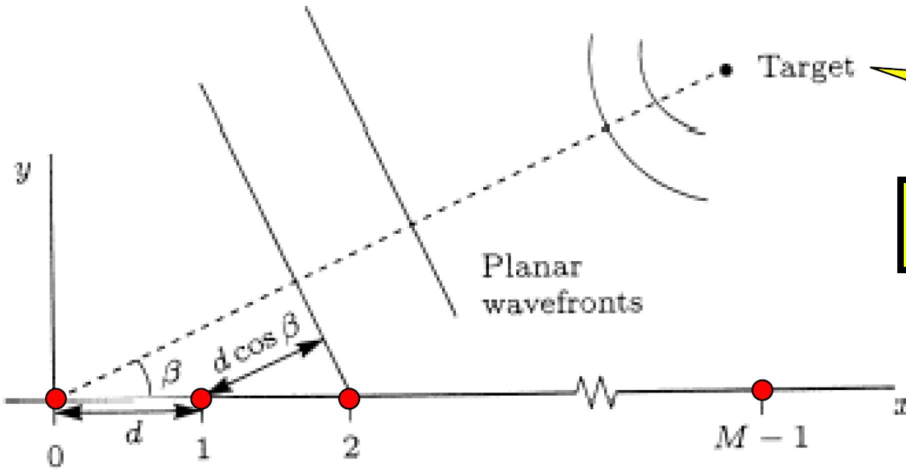
The processing is implemented as follows:

Given the data: $x[n]$, $n = 0, 1, 2, \ldots, N-1$

1. **Compute the DFT** $X[m]$, $m = 0, 1, 2, \ldots, M-1$ of the data
    - Zero-pad to length $M = 4N$ to ensure dense grid of frequency points.
    - Use the FFT algorithm for computational efficiency.
2. **Find location of peak**
    - Use quadratic interpolation of $|X[m]|$
3. **Find height at peak**
    - Use quadratic interpolation of $|X[m]|$
4. **Find angle at peak**
    - Use linear interpolation of $\angle X[m]$



❑ Example 3: Bearing estimation MLE



Emits or reflects signal $s(t)$

$$s(t) = A_t \cos(2\pi f_o t + \phi)$$

Simple model

Grab one snapshot of all M sensors at a single instant t_s:

$$x[n] = s_n(t_s) + w[n] = A\cos\left(\Omega_s n + \tilde{\phi}\right) + w[n]$$

Solution: Same as Sinusoidal estimation (example 2). So, compute DFT and find location of peak.

9. Spectral Estimation (1)

- **Definition of PSD (Power Spectral Density)**

Given a WSS discrete-time random process $x[n]$, the PSD is defined by:

$$S_x(\omega) = \lim_{M\to\infty} E\left\{ \frac{1}{2M+1} \left| \sum_{n=-M}^{M} x[n]e^{-j\omega n} \right|^2 \right\} \quad (1)$$

Recall the **Wiener-Khinchine theorem**:

$$S_x(\omega) = \mathcal{F}\{r_x[k]\}$$

$$= \sum_{k=-\infty}^{\infty} r_x[k]e^{-j\omega k} \quad (2)$$

$$\boxed{\underbrace{r_x[k] = E\{x[n]x^*[n+k]\}}_{\text{ACF of RP } x[n]}}$$

9. Spectral Estimation (2)

Problem of PSD estimation:

- Both (1) and (2) involve ensemble averaging but **in practice we get only one realization from the ensemble.**
- Both (1) and (2) use a Fourier transform of infinite length but **in practice we get only a finite number of samples.**

  (Note: a finite number of samples allows only a finite number of autocorrelation function - ACF values)

(1) & (2) motivate two approaches to PSD estimation:

1. Compute the DFT of the signal and then do some form of averaging.
2. Compute and estimate of the ACF using some form of averaging and then compute the DFT.

Both of them are called **classical non-parametric approaches** –they strive to do the best with the available data without making any assumptions other than that the underlying process is WSS.

9. Spectral Estimation (3)

Modern parametric approach:

There is a so-called modern approach to PSD estimation that tries to deal with the issue of having only a finite number of samples:

- Assume a recursive model for the ACF.
- Allows recursive extension of ACF using the known values.

Example model:

$$r_x[k] = -a_1 r_x[k-1] - a_2 r_x[k-2] - \cdots - a_p r_x[k-p], \quad k \geq p+1$$

We'll see that for this approach all we'll need to do is estimate the model parameters $\{a_i\}$ and then use them to get an estimate of the PSD. Thus, this approach is called **parametric**.

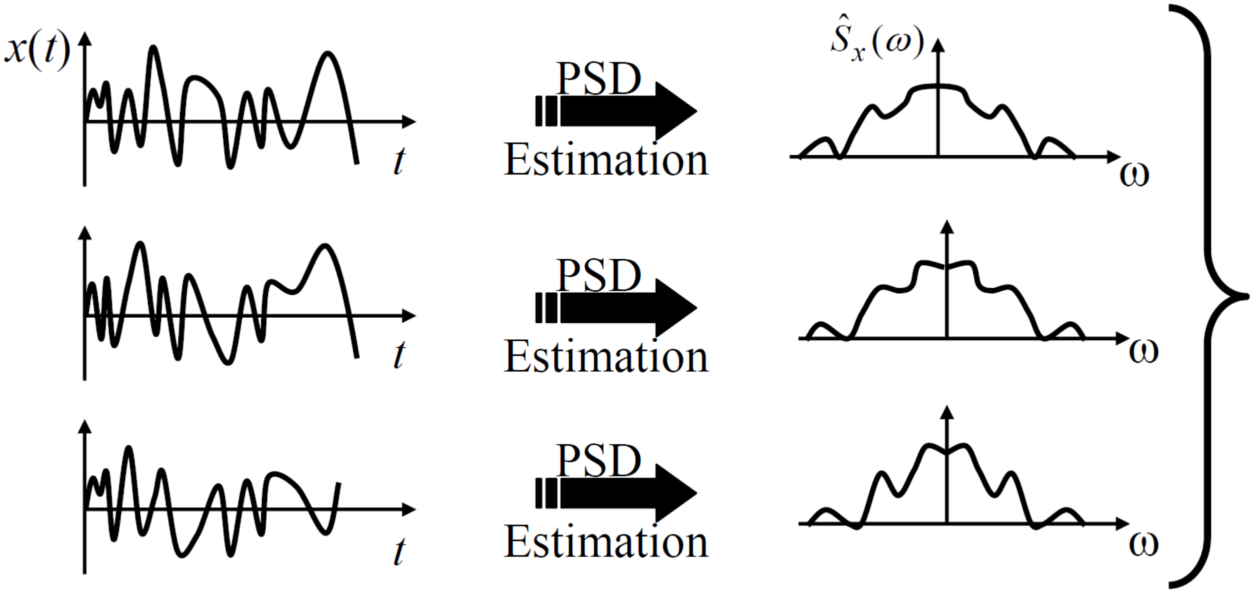
9. Spectral Estimation (4)

Review of statistics

What are we doing in PSD estimation?

Given: Finite number of samples from one realization.

Get: Something that "resembles" the PSD of the process.



Each signal realization gives a different PSD estimate realization

Each PSD estimate is a realization of a random process.

9. Spectral Estimation (5)

Thus, must view PSD estimate as a random process.

Need to characterize its mean and variance:

- Want mean of PSD estimate = true PSD
- Want variance of PSD estimate = small

To make things easier to discuss, we use a slightly different estimation problem to illustrate the ideas, consider the process:

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

Constant

AWGN, zero-mean, σ^2

Given a finite set of data samples $x[0], \ldots, x[N\text{-}1]$, estimate A.

Reasonable estimate (sample mean) is:

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

For each realization of $x[n]$ we get a different value for the estimate of A.

9. Spectral Estimation (6)

We want two things for the estimate:

1. We want our estimate to be "correct on average": $\boxed{E\{\hat{A}\} = A}$

   If this is true, we say the estimate is **unbiased**.

   If it is not true then we say the estimate is **biased**.

   If it is not true, but

$$\boxed{\lim_{N\to\infty} E\{\hat{A}\} = A}$$

   we say that the estimate is **asymptotically unbiased**.

2. We want small fluctuations from estimate to estimate:

$$\mathrm{var}\{\hat{A}\} = small$$

   Also, we'd like

$$\mathrm{var}\{\hat{A}\} \to 0 \quad as \quad N \to \infty$$

9. Spectral Estimation (7)

Can capture both mean and variance of an estimate by using Mean-Square-Error (MSE):

$$MSE\{\hat{A}\} = \mathrm{var}\{\hat{A}\} + B^2\{\hat{A}\}$$

$$where \quad B\{\hat{A}\} = A - E\{\hat{A}\}$$

Usual goal of estimation: Minimize MSE

- Minimize Bias
- Minimize Variance.

For PSD Estimation want:

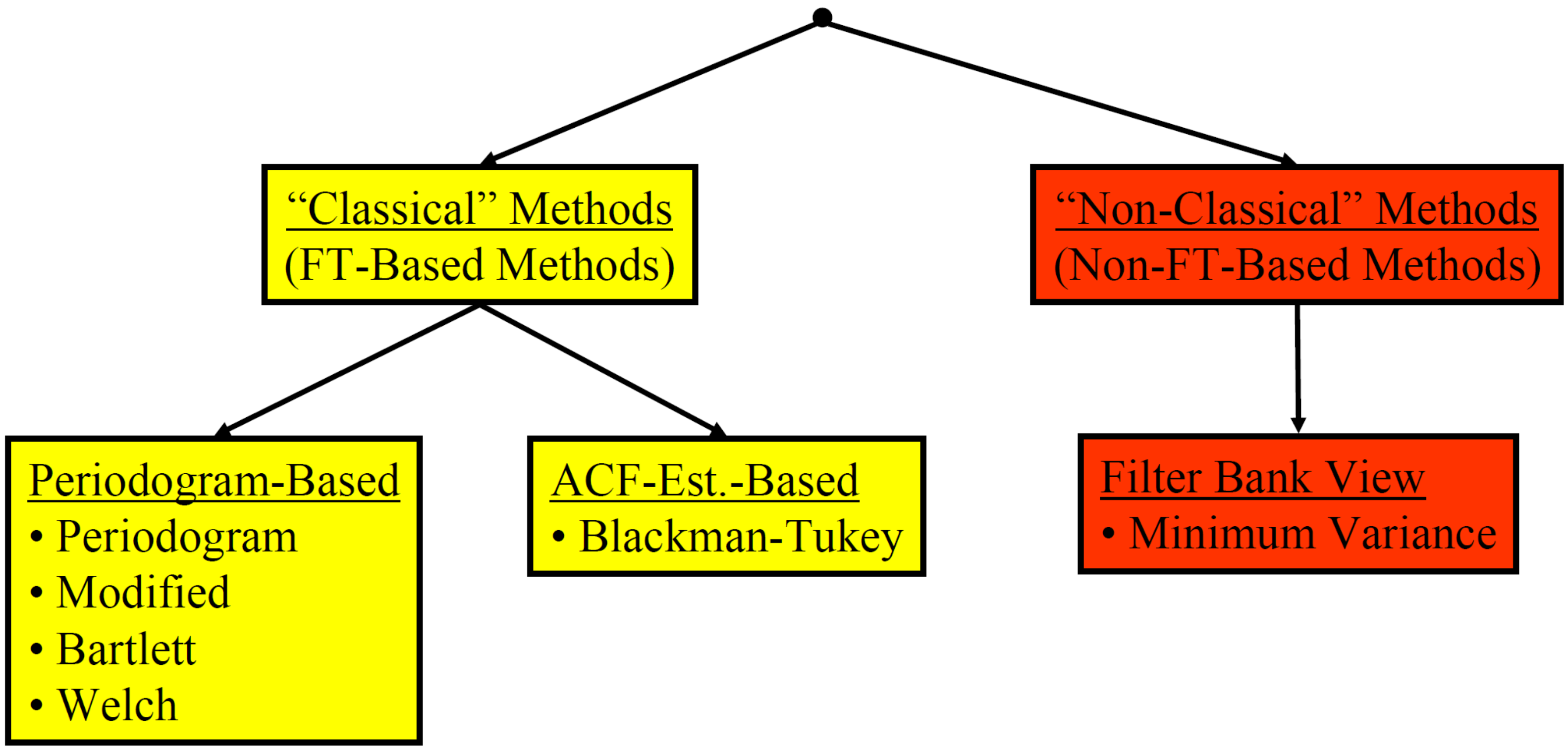
$$E\{\hat{S}_x(\omega)\} = S_x(\omega)$$

$$\mathrm{var}\{\hat{S}_x(\omega)\} = small$$

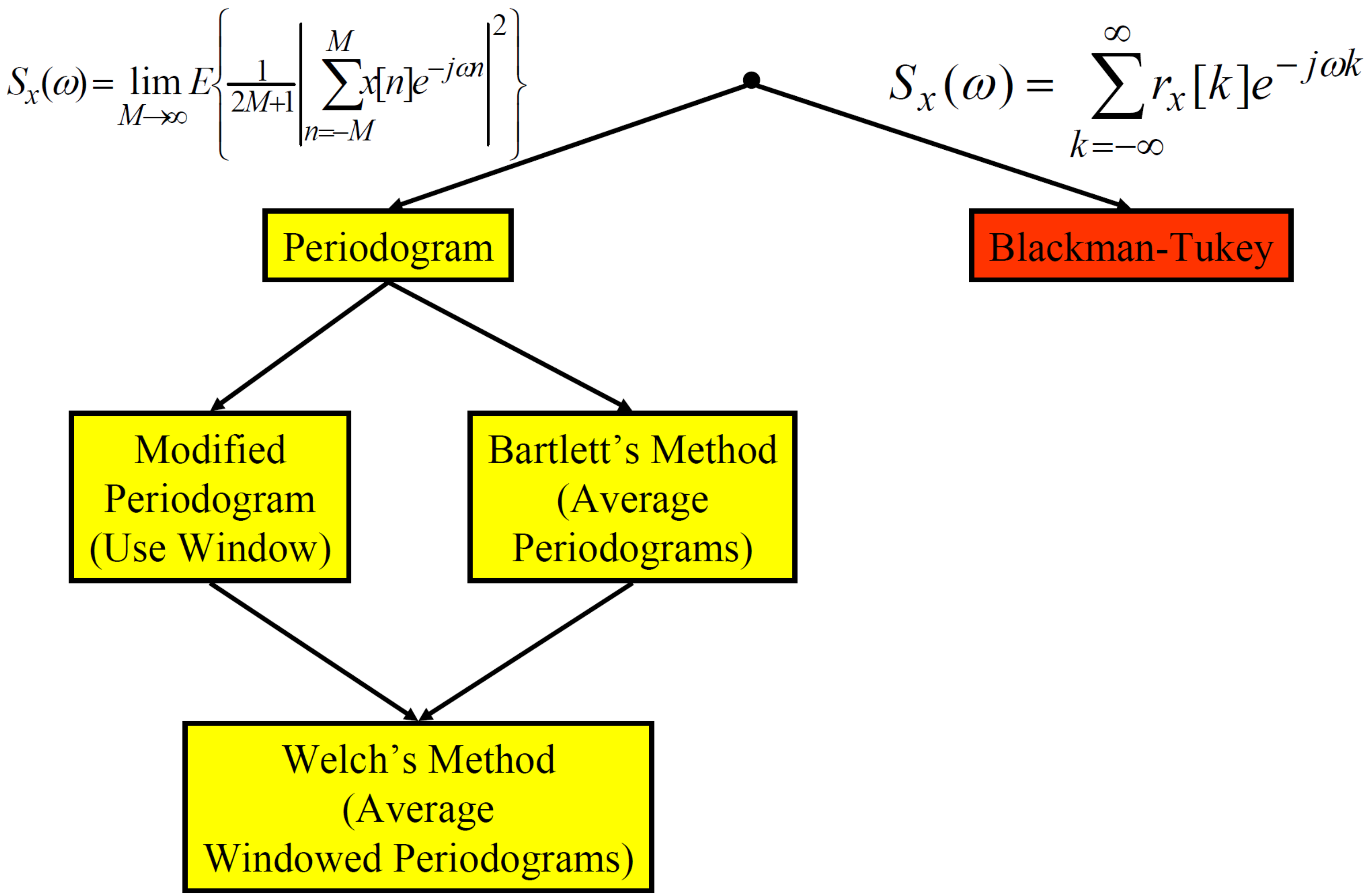
Non-Parametric Spectral Estimation

- Periodogram
- Windowed Periodogram
- Averaged Periodogram
- Windowed & Averaged Periodogram
- Blackman-Tukey Method
- Minimum Variance Method

9. Family of Non-Parametric Methods



9. Family of Classical Methods



9. Periodogram (1)

- **Periodogram –Definition**

Based on:

$$S_x(\omega) = \lim_{M\to\infty} E\left\{ \frac{1}{2M+1} \left| \sum_{n=-M}^{M} x[n]e^{-j\omega n} \right|^2 \right\}$$

In practice, we have one set of finite-duration data. Two practical problems:

1. Can’t do the expected value
2. Can’t do the limit.

The periodogram is a method that ignores them both:

$$\boxed{\hat{S}_{PER}(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n]e^{-j\omega n} \right|^2}$$

In practice, we compute this using the DFT (possibly using zero-padding) –which computes the DTFT at discrete frequency points (DFT bins).

9. Periodogram (2)

Periodogram – Computation

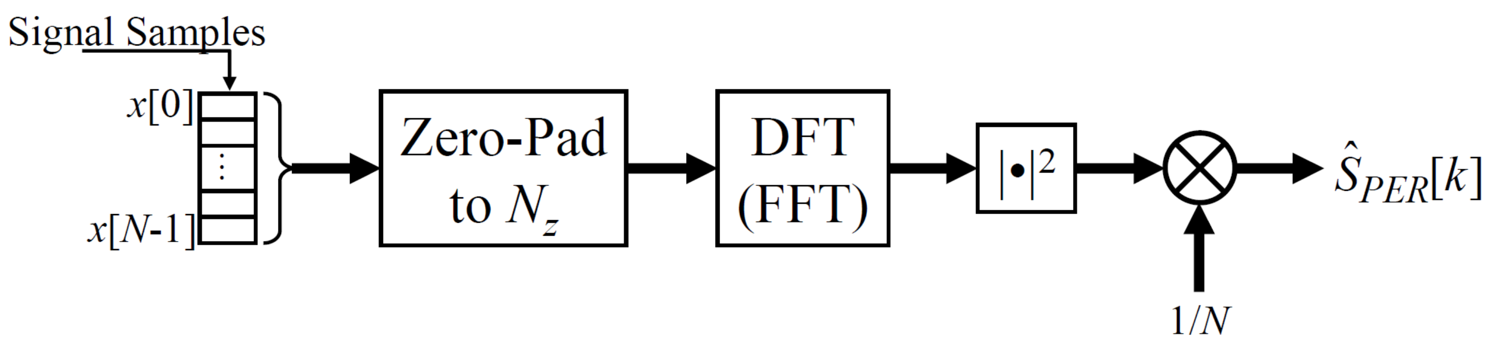
In practice, we compute this using the DFT (FFT) (usually using zero-padding) – which computes the DTFT at discrete frequency points (DFT bins):

$$\boxed{\hat{S}_{PER}[k] = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N_z}\right|^2}$$

$$\omega_k = 2\pi k / N_z$$

N = number of signal samples

N_z = DFT size – after zero-padding



9. Periodogram (3)

- **Periodogram –viewed as Filter Bank**

  Although we always implement the periodogram using the DFT, it is helpful to interpret it as a filter bank.

  Define the impulse response of an FIR filter as:

$$h_i[n] = \begin{cases} \dfrac{1}{N} e^{jn\omega_i}, & 0 \le n < N \\ 0, & otherwise \end{cases}$$

  Frequency Response of this filter is:

$$H(\omega) = \sum_{n=0}^{N-1} h_i[n] e^{-jn\omega}$$

$$= e^{-jn(\omega-\omega_i)(N-1)/2} \frac{\sin[N(\omega-\omega_i)/2]}{N\sin[(\omega-\omega_i)/2]}$$

9. Periodogram (4)

Now the output of the i^{th} filter is:

$$y_i[n] = x[n] * h_i[n] = \sum_{k=n-N+1}^{n} x[k] h_i[n-k]$$

$$= \frac{1}{N} \sum_{k=n-N+1}^{n} x[k] e^{j(n-k)\omega_i}$$

Now one estimate of the power at the output of this filter i$|y_i[n]|^2$ for any value of n. Choosing $n= N$-1 gives the periodogram:

$$\left|y_i[N-1]\right|^2 = \left|\frac{1}{N}\sum_{k=0}^{N-1} x[k] e^{j(N-1-k)\omega_i}\right|^2 = \underbrace{\left|e^{j(N-1)\omega_i}\right|^2}_{=1} \left|\frac{1}{N}\sum_{k=0}^{N-1} x[k] e^{j-k\omega_i}\right|^2$$

$$= \left|\frac{1}{N}\sum_{k=0}^{N-1} x[k] e^{j-k\omega_i}\right|^2 = N\hat{S}_{PER}(\omega)$$

9. Periodogram (5)

Periodogram –Performance

For a good PSD estimate we'd like to have (at the very least):

$$\lim_{N\to\infty} E\{\hat{S}_x(\omega)\} = S_x(\omega) \qquad \text{: Asymptotic unbiased}$$

$$\lim_{N\to\infty} \operatorname{var}\{\hat{S}_x(\omega)\} = 0$$

Actually, we would prefer it to be unbiased even for finite *N*.

Does the periodogram have these characteristics?

9. Periodogram (6)

Property #1: The periodogram is **biased**.

Property #2: But, the periodogram is **asymptotically unbiased**.

9. Periodogram (7)

Proof: Taking the expected value of the periodogram gives

$$E\left\{\hat{S}_{PER}(\omega)\right\} = \frac{1}{N} E\left\{\left|\sum_{n=0}^{N-1} x[n] e^{-j\omega n}\right|^2\right\}$$

$$= \frac{1}{N} E\left\{\left[\sum_{n=0}^{N-1} x[n] e^{-j\omega n}\right]\left[\sum_{m=0}^{N-1} x^*[m] e^{j\omega m}\right]\right\}$$

$$= \frac{1}{N} \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} r_x[n-m] e^{-j\omega(n-m)}$$

Sum on diagonals $r_x[n\text{-}m]$ is constant on each diagonal

$$= \sum_{k=-(N-1)}^{N-1} \underbrace{\left(1-\frac{|k|}{N}\right)}_{w_B[k]} r_x[k] e^{-j\omega k}$$

Bartlett (Triangle) window

$$= \frac{1}{2\pi} S_x(\omega) \underset{circ}{*} W_B(\omega) \neq S_x(\omega)$$

9. Periodogram (8)

This shows that the periodogram is **biased**. **The bias comes from the smoothing effect of Bartlett window.**

(Smoothing also reduces the resolution of sharp spectral features)

But, as $N\to\infty$, the Bartlett kernel tends to a delta function in the frequency domain, or –equivalently –in the time domain the Bartlett window tends 1:

$$\lim_{N\to\infty} E\left\{\hat{S}_{PER}(\omega)\right\} = \lim_{N\to\infty} \sum_{k=-(N-1)}^{N-1} \underbrace{\left(1-\frac{|k|}{N}\right)}_{\to 1} r_x[k]e^{-j\omega k}$$

$$= \sum_{k=-\infty}^{\infty} r_x[k]e^{-j\omega k} = S_x(\omega)$$

Thus, the periodogram is asymptotically unbiased.

9. Periodogram (9)

Property #3: The variance of the periodogram does not (in general) tend to zero as $N \to \infty$.

Proof: Difficult to prove for general case, so this is proved under the assumption: complex-valued white Gaussian process with zero mean and variance σ^2.

Under this assumption, the true PSD and ACF are:

$$S(\omega) = \sigma^2, \quad \forall \omega \quad \& \quad r_x[k] = \sigma^2 \delta[k]$$

The variance of the periodogram is what we want to analyze and is given by:

$$\operatorname{var}\left\{\hat{S}_{PER}(\omega)\right\} = E\left\{\hat{S}^2_{PER}(\omega)\right\} - \left[E\left\{\hat{S}_{PER}(\omega)\right\}\right]^2$$

Bias term

9. Periodogram (10)

So from our previous analysis of bias (and our assumptions on the process) we know that the second term is:

$$\left[E\left\{\hat{S}_{PER}(\omega)\right\}\right]^2 = \left[\sum_{k=-(N-1)}^{N-1}\left(1-\frac{|k|}{N}\right)r_x[k]e^{-j\omega k}\right]^2$$

$$= \left[\sum_{k=-(N-1)}^{N-1}\left(1-\frac{|k|}{N}\right)\left[\sigma^2\delta[k]\right]e^{-j\omega k}\right]^2 = \sigma^4\left[\left(1-\frac{|k|}{N}\right)e^{-j\omega k}\right]_{k=0}^2$$

$$= \sigma^4$$

So the variance of periodogram is now:

$$\mathrm{var}\left\{\hat{S}_{PER}(\omega)\right\} = E\left\{\hat{S}_{PER}^2(\omega)\right\} - \sigma^4$$

9. Periodogram (11)

As a means of looking at this first term we consider:

$$E\left\{\hat{S}_{PER}(\omega_1)\hat{S}_{PER}(\omega_2)\right\} = E\left\{\left[\frac{1}{N}\underbrace{\left|\sum_{l=0}^{N-1} x[l]e^{-j\omega_1 l}\right|^2}_{\left[\sum_{k=0}^{N-1} x[k]e^{-j\omega_1 k}\right]\left[\sum_{l=0}^{N-1} x[l]e^{-j\omega_1 l}\right]^*}\right]\left[\frac{1}{N}\underbrace{\left|\sum_{n=0}^{N-1} x[n]e^{-j\omega_2 n}\right|^2}_{\left[\sum_{m=0}^{N-1} x[m]e^{-j\omega_2 m}\right]\left[\sum_{n=0}^{N-1} x[n]e^{-j\omega_2 n}\right]^*}\right]\right\}$$

$$E\left\{\hat{S}_{PER}(\omega_1)\hat{S}_{PER}(\omega_2)\right\}$$
$$= \frac{1}{N^2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} E\left\{x[k]x^*[l]x[m]x^*[n]\right\}\exp\{-j[(k-l)\omega_1 + (m-n)\omega_2]\}$$

9. Periodogram (12)

Since we assumed the process is Gaussian we can use a standard result for complex joinly Gaussian RVs:

$$E\left\{x[k]x^*[l]x[m]x^*[n]\right\} = E\left\{x[k]x^*[l]\right\}E\left\{x[m]x^*[n]\right\} + E\left\{x[k]x^*[n]\right\}E\left\{x[m]x^*[l]\right\}$$

Using this result together with the assumption of whiteness:

$$E\left\{x[l]x[k]x[n]x[m]\right\}$$
$$= \sigma^4\left[\delta[k-l]\delta[m-n] + \delta[k-n]\delta[m-l]\right]$$

9. Periodogram (13)

Therefore,

$$E\left\{\hat{S}_{PER}(\omega_1)\hat{S}_{PER}(\omega_2)\right\}$$

$$= \frac{\sigma^4}{N^2}\left[\sum_{l=0}^{N-1}\sum_{k=0}^{N-1}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \delta[k-l]\delta[m-n]\exp\{-j[\omega_1(k-l)+\omega_2(m-n)]\}\right.$$

$$\left. + \sum_{l=0}^{N-1}\sum_{k=0}^{N-1}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\delta[k-n]\delta[m-l]\exp\{-j[\omega_1(k-l)+\omega_2(m-n)]\}\right]$$

$$= \frac{\sigma^4}{N^2}\left[\sum_{l=0}^{N-1}\sum_{n=0}^{N-1}1 + \sum_{l=0}^{N-1}\sum_{k=0}^{N-1}\exp\{-j[\omega_1(k-l)-\omega_2(k-l)]\}\right]$$

$$= \frac{\sigma^4}{N^2}\left[\sum_{l=0}^{N-1}\sum_{n=0}^{N-1}1 + \sum_{l=0}^{N-1}\sum_{k=0}^{N-1}\exp\{-j[(\omega_1-\omega_2)(k-l)]\}\right]$$

9. Periodogram (14)

$$E\left\{\hat{S}_{PER}(\omega_1)\hat{S}_{PER}(\omega_2)\right\}$$

$$= \frac{\sigma^4}{N^2}\left[N^2 + N\sum_{k=-(N-1)}^{N-1}\underbrace{\left(1-\frac{|k|}{N}\right)}_{w_B[k]}\exp\{-j[(\omega_1-\omega_2)k]\}\right]$$

Now the Fourier transform of the Bartlett window is:

$$\mathfrak{F}\{w_B[k]\} = \left(\frac{\sin(N\omega/2)}{\sin(\omega/2)}\right)^2$$

So using it in the above result gives:

$$E\left\{\hat{S}_{PER}(\omega_1)\hat{S}_{PER}(\omega_2)\right\} = \sigma^4\left[1+\left(\frac{\sin[N(\omega_1-\omega_2)/2]}{N\sin[(\omega_1-\omega_2)/2]}\right)^2\right]$$

9. Periodogram (15)

To find the first term in the variance expression of interest, we must set $\omega = \omega_1 = \omega_2$ in the above expression to get:

$$E\left\{\hat{S}^2_{PER}(\omega)\right\} = 2\sigma^4$$

Now using this in the expression for variance gives:

$$\begin{aligned}\operatorname{var}\left\{\hat{S}_{PER}(\omega)\right\} &= E\left\{\hat{S}^2_{PER}(\omega)\right\} - \sigma^4 \\ &= 2\sigma^4 - \sigma^4\end{aligned}$$

$$\boxed{\operatorname{var}\left\{\hat{S}_{PER}(\omega)\right\} = \sigma^4}$$

which does not go to zero as $N \rightarrow \infty$

9. Periodogram (16)

Property #4: Increasing N leads to rapidly fluctuating periodograms (even where the true PSD is smooth).

Proof: Use the previous results, the covariance of the periodogram is given by

$$\operatorname{cov}\left\{\hat{S}_{PER}(\omega_1)\hat{S}_{PER}(\omega_2)\right\} = E\left\{\hat{S}_{PER}(\omega_1)\hat{S}_{PER}(\omega_2)\right\} - E\left\{\hat{S}_{PER}(\omega_1)\right\}E\left\{\hat{S}_{PER}(\omega_2)\right\}$$

$$= \sigma^4\left[1 + \left(\frac{\sin[N(\omega_1-\omega_2)/2]}{N\sin[(\omega_1-\omega_2)/2]}\right)^2\right] - \sigma^4$$

$$= \sigma^4\left(\frac{\sin[N(\omega_1-\omega_2)/2]}{N\sin[(\omega_1-\omega_2)/2]}\right)^2$$

Covariance is a measure of how correlated two RVs are. Thus, $\operatorname{cov}(X,Y) = 0$ indicates that there is a high probability that X & Y will be very unalike.

9. Periodogram (17)

Now, the equation above indicates there are (ω_1, ω_2) pairs for which the covariance of the periodogram is zero.

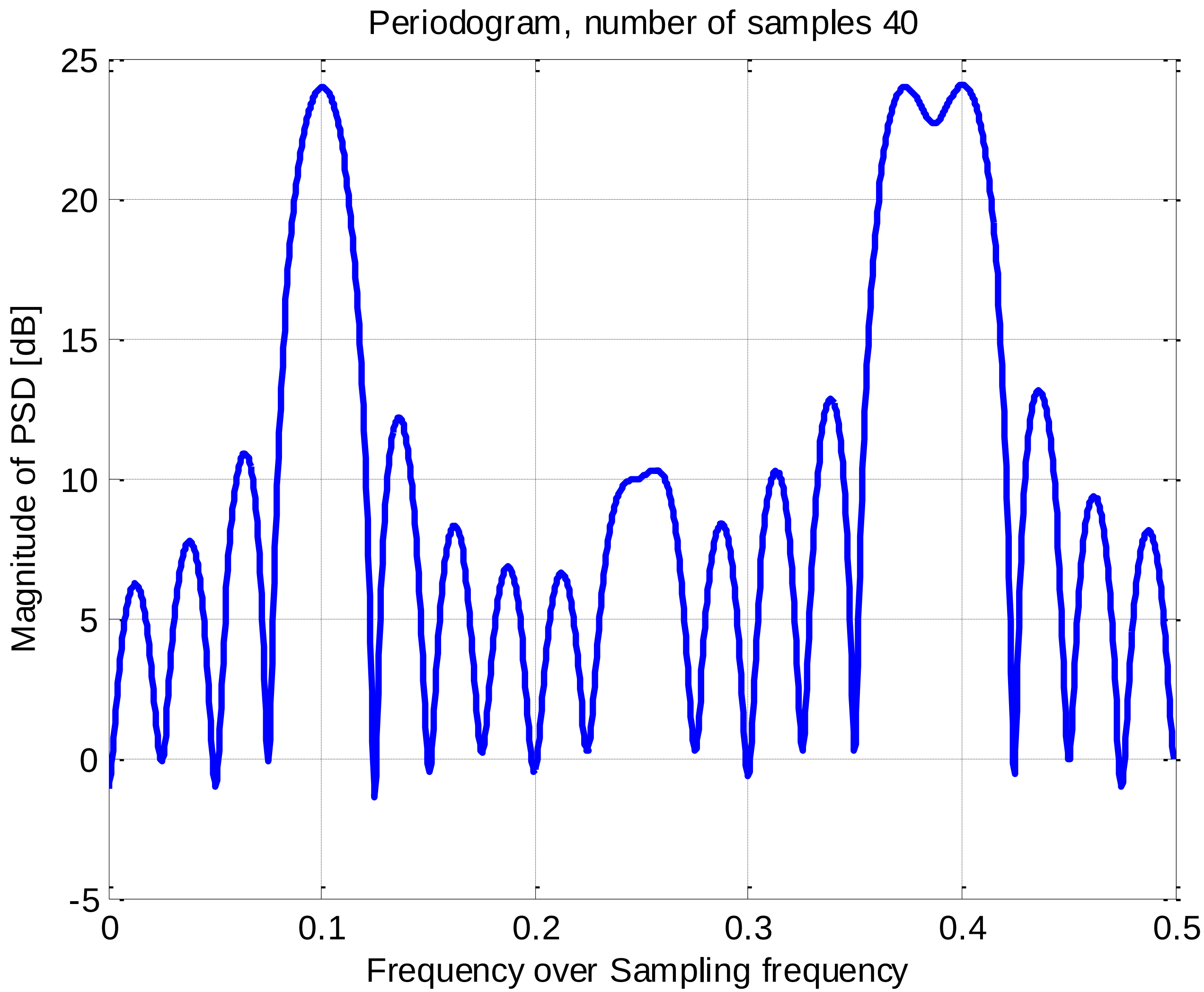
$\Rightarrow$ **Periodogram fluctuates rapidly from frequency-to-frequency.**

9. Periodogram (18)

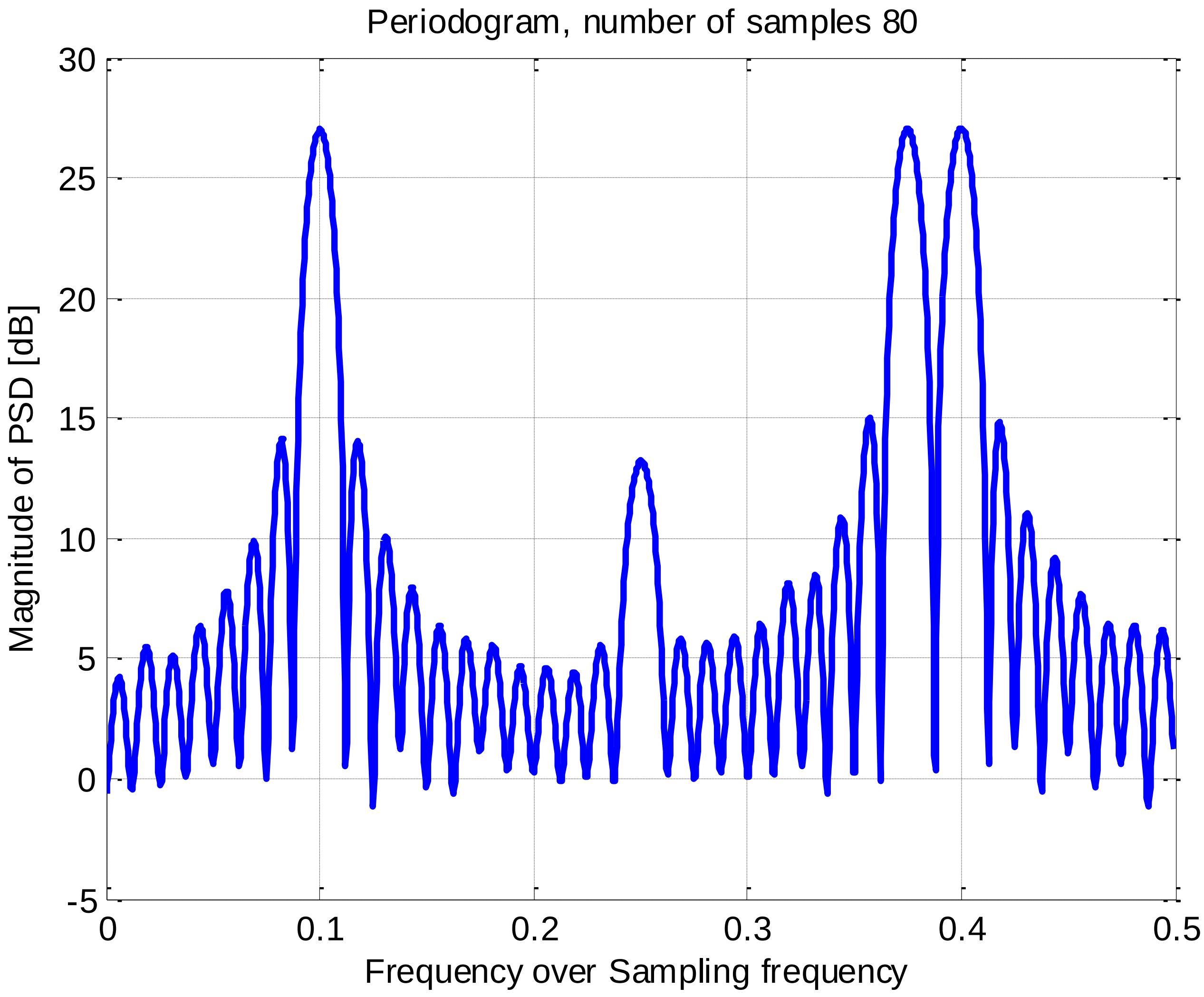
Example:

Periodogram of a sum of sinusoids in white Gaussian noise. The sinusoids are with frequencies of $0.1f_s$, $0.25f_s$, $0.375f_s$ and $0.4f_s$, and with amplitudes of 5, 1, 5 and 5, respectively. The f_s is sampling frequency. The second sinusoid is with smaller amplitude than others while the third sinusoid is very close to the fourth sinusoid. Number of signal samples is 40, DFT size is 2048. The white Gaussian noise is with variance of 1.

9. Periodogram (19)



9. Periodogram (20)



9. Practical Classical Methods (1)

- **<u>Main problems of the periodogram</u>**
  - Biased estimate
  - Variance does not decrease with increasing N
  - Rapid fluctuations

All of these arise due to the fact that the periodogram ignores:

- The expected value – (It includes no averaging)
- The limit operation – (It applies a rectangular window)

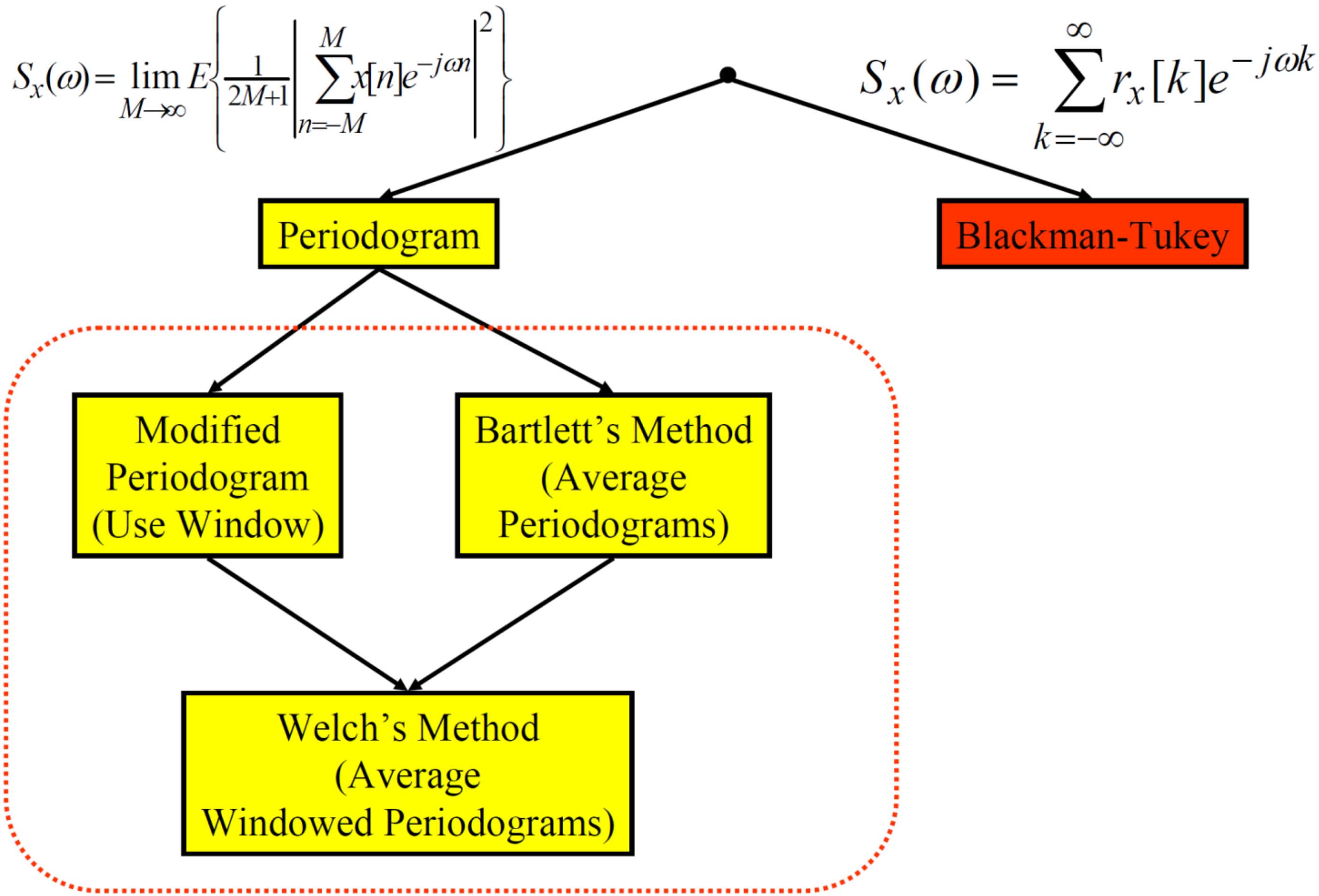
in the PSD definition.

Several classical methods for partially fixing these have been proposed.

9. Practical Classical Methods (2)

- **<u>Modifications based on periodogram view</u>**

  Recall: Family of classical methods:



9. Practical Classical Methods (3)

- <u>**Modified periodogram –windowed**</u>

The **modified periodogram** uses a **non-rectangular window** and therefore has to be scaled to account for the loss of power due to the window. This scaling is required to make the modified periodogram asymptotically unbiased:

$$\boxed{\hat{S}_{MP}(\omega) = \frac{1}{NU}\left|\sum_{n=0}^{N-1} x[n]w[n]e^{-j\omega n}\right|^2}$$

where the scaling factor is

$$U = \frac{1}{N}\sum_{n=0}^{N-1}\left|w[n]\right|^2$$

Note: $U = 1$ for a rectangular window.

As in the ordinary periodogram, the DFT/FFT is used for computation and zero-padding is usually used.

9. Practical Classical Methods (4)

The modified periodogram:

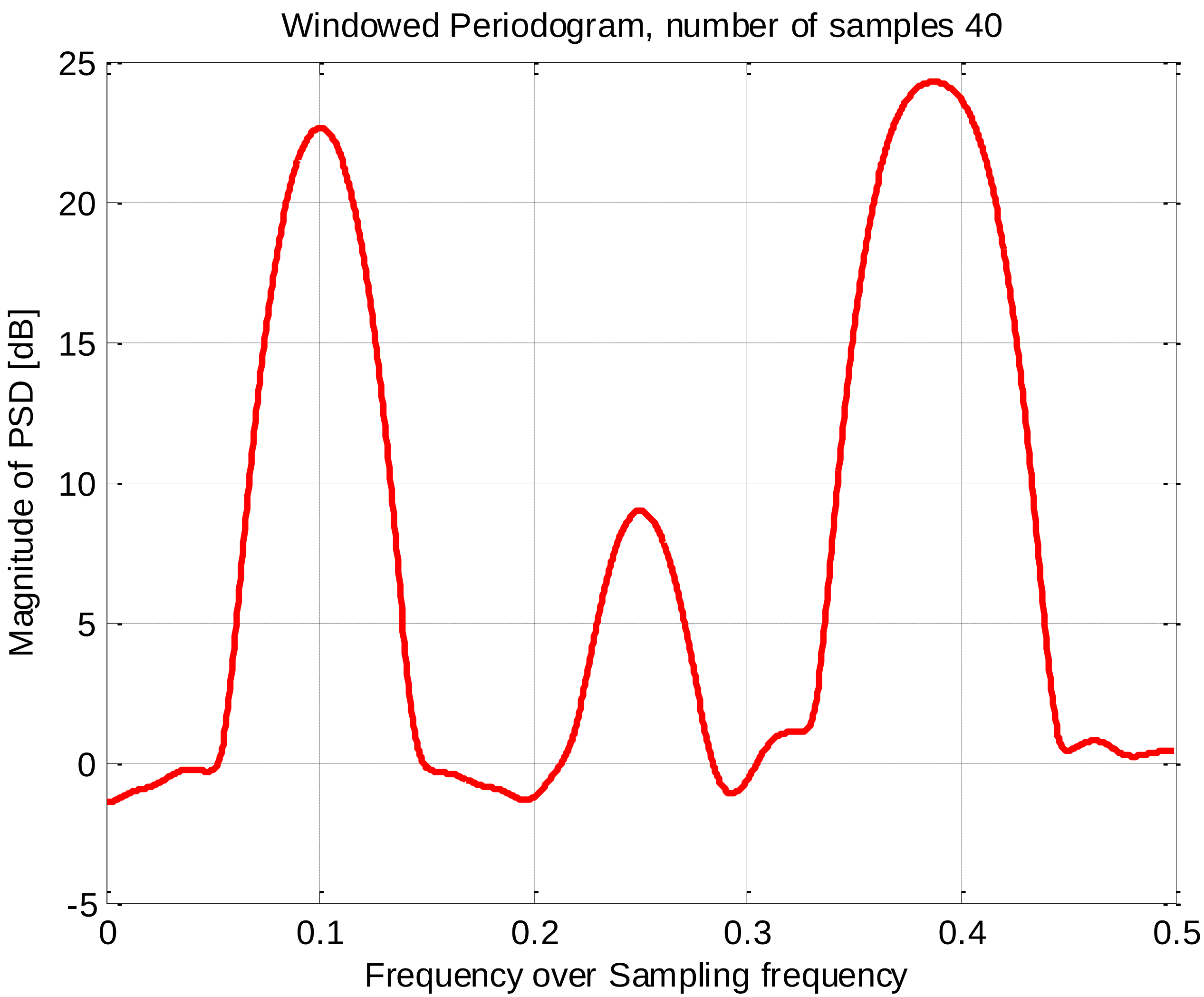
- Has reduced bias but is still biased.
- Is asymptotically unbiased.
- Has variance that roughly equals that of the periodogram.

9. Practical Classical Methods (5)

<u>Example</u>:

Modified (windowed) periodogram using Hamming window. The sinusoids in white Gaussian noise are with frequencies of $0.1f_s$, $0.25f_s$, $0.375f_s$ and $0.4f_s$, and with amplitudes of 5, 1, 5 and 5, respectively. The f_s is sampling frequency. The second sinusoid is with smaller amplitude than others while the third sinusoid is very close to the fourth sinusoid. Number of signal samples is 40, DFT size is 2048. The white Gaussian noise is with variance of 1.

9. Practical Classical Methods (6)



9. Practical Classical Methods (7)

- **Bartlett's method: Averaged periodogram**

  One of the main flaws in the periodogram is the lack of averaging.

  This lack of averaging is what leads to the non-decreasing variance as well as the rapid fluctuations of the periodogram.

  Now, in practice we have only one realization.

  We hope that the process is ergodic (A process is ergodic if time averaging of any realization is equivalent to ensemble averaging).

9. Practical Classical Methods (8)

- **<u>Bartlett's method –definition</u>**

  The signal data of length N is chopped into K non-overlapping blocks of length L (the length L is a design choice); $N = KL$:

  Block definition:

$$x_i[n] = x[n + iL] \quad n = 0, 1, \cdots, L-1$$
$$i = 0, 1, \cdots, K-1$$

$$\boxed{\begin{aligned}\hat{S}_B(\omega) &= \frac{1}{K}\sum_{i=0}^{K-1}\frac{1}{L}\left|\sum_{n=0}^{L-1} x_i[n] e^{-j\omega n}\right|^2 \\ &= \frac{1}{N}\sum_{i=0}^{K-1}\left|\sum_{n=0}^{L-1} x_i[n] e^{-j\omega n}\right|^2\end{aligned}}$$

9. Practical Classical Methods (9)

Variance improvement

The intent of averaging here is to improve the variance. To see how lets first just look at a simple related example:

Let X_i, i= 0, 1, ..., K-1 be a sequence of independent, identically distributed RVs each having zero-mean and variance σ^2. What is the variance of the data analysis average of them?

$$\bar{X} = \frac{1}{K}\sum_{i=0}^{K-1} X_i$$

$$\begin{aligned}\operatorname{var}\{\bar{X}\} &= E\{\bar{X}^2\} - \big[\underbrace{E\{\bar{X}\}}_{=0}\big]^2 \\ &= \frac{1}{K^2}\sum_{i=0}^{K-1}\sum_{k=0}^{K-1} E\{X_i X_k\} \\ &= \frac{1}{K^2}\sum_{i=0}^{K-1}\sum_{k=0}^{K-1} \sigma^2 \delta[i-k] \\ &= \frac{K\sigma^2}{K^2} = \frac{\sigma^2}{K}\end{aligned}$$

9. Practical Classical Methods (10)

- **<u>Welch's method: Averaged windowed periodogram</u>**

  We've seen:

  - Windowing helps the bias
  - Averaging helps the variance

  Welch: Do both and use overlapped blocks.

  Block definition: $x_i[n] = x[n + iD] \quad n = 0, 1, \cdots, L-1$

  $$i = 0, 1, \cdots, K-1$$

  The amount of overlap is $L\text{–}D$ points:

  $D = L$: No overlap

  $D = L/2$: 50% overlap (most common)

  $D = 3L/4$: 25% overlap

$$\boxed{\hat{S}_W(\omega) = \frac{1}{KLU} \sum_{i=0}^{K-1} \left| \sum_{n=0}^{L-1} x_i[n] w[n] e^{-j\omega n} \right|^2}$$

Implement using DFT/FFT and zero-padding

9. Practical Classical Methods (11)

Variance has been shown to be for 50% overlap:

$$\mathrm{var}\left\{\hat{S}_W(\omega)\right\} \approx \frac{9}{16}\frac{L}{N}S_x^2(\omega)$$

Compared to Bartlett's method (no overlap) for the same N and L:

$$\mathrm{var}\left\{\hat{S}_W(\omega)\right\} \approx \frac{9}{16}\mathrm{var}\left\{\hat{S}_B(\omega)\right\}$$

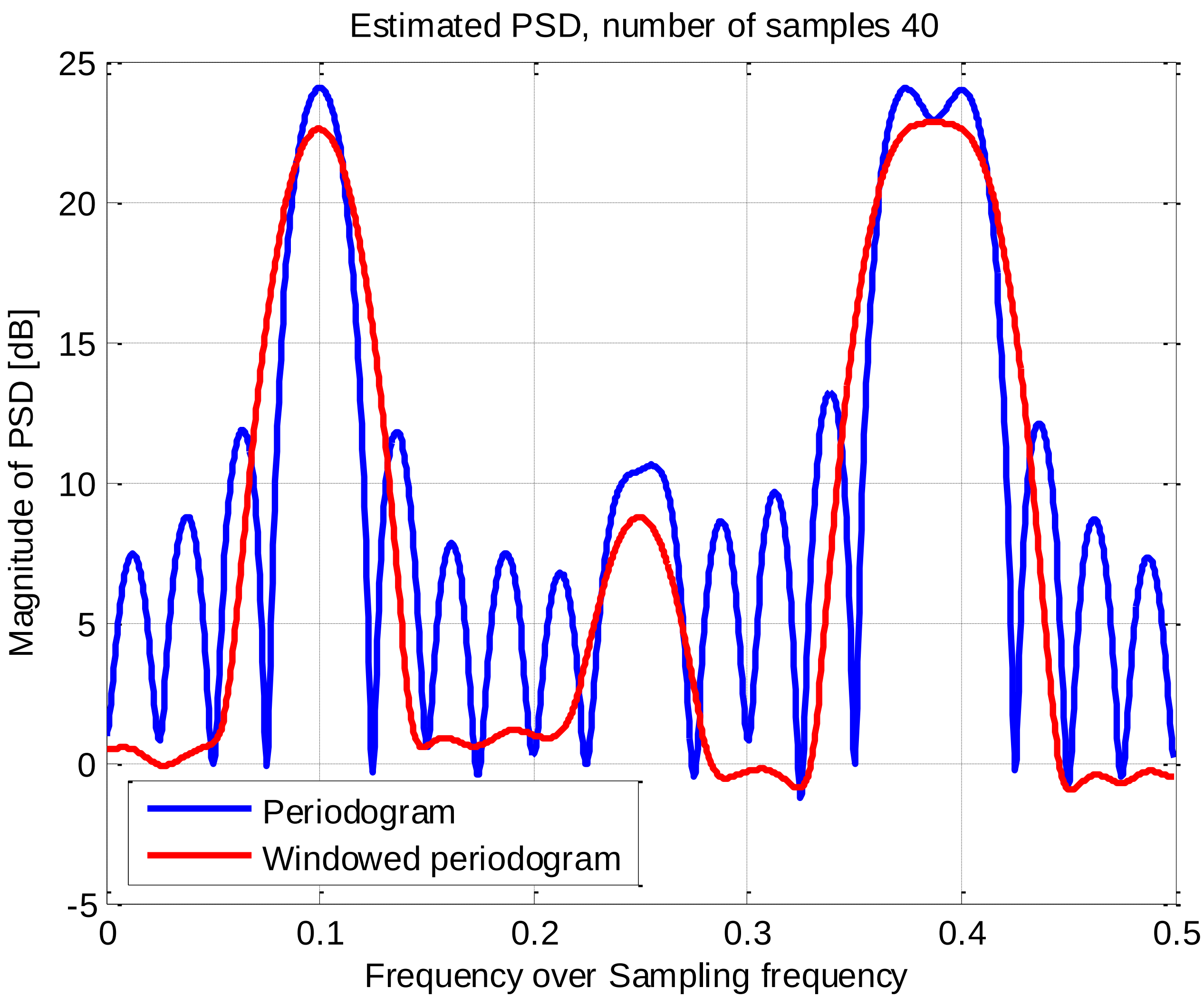
almost a 50% reduction.

9. Practical Classical Methods (12)

Windowing is useful for signals containing sinusoids + noise

- Sidelobes of a stronger sinusoid may mask the main lobe of a nearby weak sinusoid.
- We multiply $x[n]$ by window $w[n]$ before computing periodogram
- Weaker sinusoid becomes more visible, but main lobe of each sinusoid broadens: two close peaks may merge into one.

9. Practical Classical Methods (13)



9. Practical Classical Methods (14)

Commonly used windows:

Name	$w[k]$	Fourier transform
Rectangular	1	$W_R(f) = \dfrac{\sin\ \pi f(2M+1)}{\sin\ \pi f}$
Bartlett	$1 - \dfrac{\lvert k\rvert}{M}$	$\dfrac{1}{M}\left(\dfrac{\sin\ \pi fM}{\sin\ \pi f}\right)^2$
Hanning	$0.5 + 0.5\cos\dfrac{\pi k}{M}$	$0.25\, W_R\left(f - \frac{1}{2M}\right) + 0.5\, W_R(f) + 0.25\, W_R\left(f + \frac{1}{2M}\right)$
Hamming	$0.54 + 0.46\cos\dfrac{\pi k}{M}$	$0.23\, W_R\left(f - \frac{1}{2M}\right) + 0.54\, W_R(f) + 0.23\, W_R\left(f + \frac{1}{2M}\right)$

$w[k] = 0$ for $\lvert k\rvert > M$

9. Practical Classical Methods (15)

- **Modifications based on ACF view: Blackman-Tukey method**

  Periodogram's biggest problem is a variance that does not decrease with increasing *N*.

  The methods we've seen dealt with this by averaging.

  There is another way to combat this. To see how, we need to write the periodogram differently –motivated by the Wiener-Khinchine theorem:

$$S_x(\omega) = \sum_{k=-\infty}^{\infty} r_x[k] e^{-j\omega k}$$

$$\text{ACF}: \; r_x[k] = E\left\{x[n]x^*[n+k]\right\}$$

9. Practical Classical Methods (16)

The periodogram can be written as:

$$\hat{S}_{PER}(\omega)=\frac{1}{N}\left|\sum_{n=0}^{N-1}x[n]e^{-j\omega n}\right|^2=\frac{1}{N}\left[\sum_{m=0}^{N-1}x[m]e^{-j\omega m}\right]\left[\sum_{n=0}^{N-1}x[n]e^{-j\omega n}\right]^*$$

$$=\frac{1}{N}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}x[m]x^*[n]e^{-j\omega(m-n)}$$

Let $m=n+k \Rightarrow k=m-n \Rightarrow k\in[-(N-1),(N-1)]$

$$=\frac{1}{N}\sum_{k=-(N-1)}^{N-1}\sum_{n=0}^{N-1}x[n+k]x^*[n]e^{-j\omega k}$$

Note: $x[n+k]=0$ for $n+k>N-1$ i.e. for $n>N-k-1$

$$=\sum_{k=-(N-1)}^{N-1}\underbrace{\left[\frac{1}{N}\sum_{n=0}^{N-|k|-1}x[n+k]x^*[n]\right]}_{=\hat{r}_b[k]}e^{-j\omega k}$$

9. Practical Classical Methods (17)

Thus, the periodogram can be written as:

$$\hat{S}_{PER}(\omega) = \sum_{k=-(N-1)}^{N-1} \hat{r}_b[k] e^{-j\omega k}$$

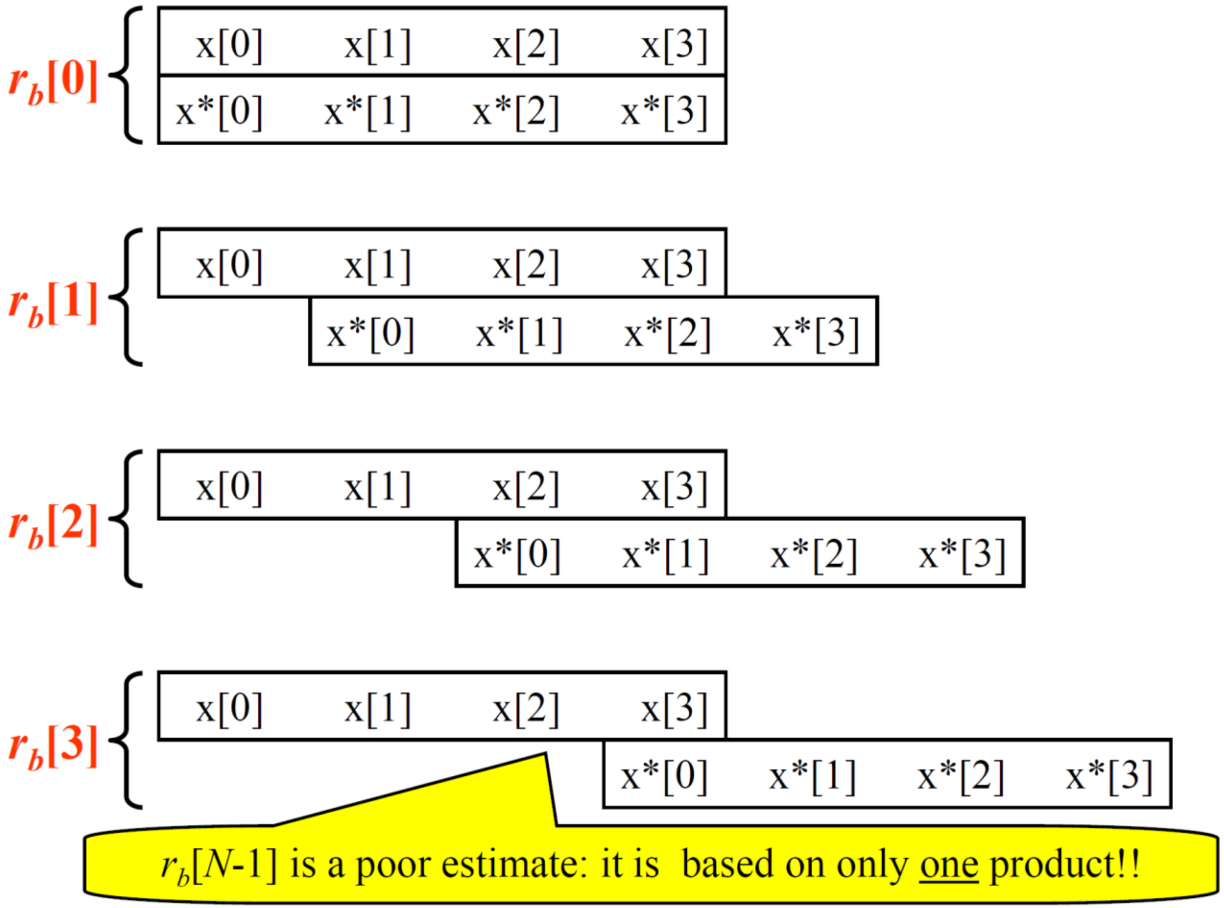
i.e, as a DTFT of an **estimated ACF** given by

$$\hat{r}_b[k] = \frac{1}{N} \sum_{n=0}^{N-|k|-1} x[n+k] x^*[n]$$

It is the poor quality of this ACF estimate that gives rise to the periodogram's poor quality.

9. Practical Classical Methods (18)

Example of ACF estimation:



9. Practical Classical Methods (19)

- **<u>Blackman-Tukey method –defined:</u>**

  For the biased ACF estimate, the estimated ACF lags for large $|k|$ values are unreliable.

  De-emphasize these unreliable lags by **applying a window to the biased ACF estimate**. This is the **Blackman-Tukey method**:

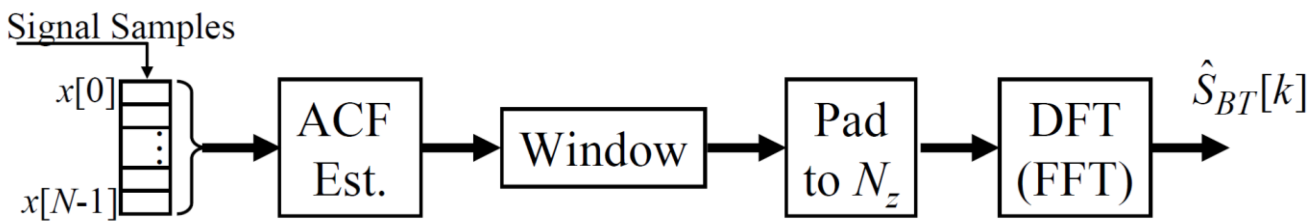
$$\hat{S}_{BT}(\omega) = \sum_{k=-M}^{M} w[k]\hat{r}_b[k]e^{-j\omega k}$$

  with $M < N - 1$.

  Since windows taper off to zero at their edges this causes the poor-quality estimates at large $|k|$ values to have less impact on the PSD estimate.

9. Practical Classical Methods (20)

In practice we compute this using the DFT(FFT) (usually using zero-padding) –which computes the DTFT at discrete frequency points (DFT bins):



It is useful to explore a frequency domain interpretation of it. By using the multiplication-convolution theorem for DTFT we have:

$$\hat{S}_{BT}(\omega) = \sum_{k=-M}^{M} w[k]\hat{r}_b[k]e^{-j\omega k}$$

(Product in Time Domain) ⇔ (Convolution in Frequency Domain)

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega-\xi)\hat{S}_{PER}(\xi)d\xi = \frac{1}{2\pi}W(\omega) \underset{circ}{*} \hat{S}_{PER}(\omega)$$

The BT estimate is a **smoothed version** of the periodogram.

9. Practical Classical Methods (21)

BT performance: Bias:

$$E\left\{\hat{S}_{BT}(\omega)\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega-\xi)\ \underbrace{E\left\{\hat{S}_{PER}(\xi)\right\}}_{\approx S_x(\xi)\text{ for large }N}\ d\xi$$

Since Asymp. Unbiased

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega-\xi) S_x(\xi) d\xi$$

$$\boxed{E\left\{\hat{S}_{BT}(\omega)\right\} \approx \frac{1}{2\pi} W(\omega) \underset{circ}{*} S_x(\omega)}$$

9. Practical Classical Methods (22)

BT performance: Variance:

$$\mathrm{var}\left\{\hat{S}_{BT}(\omega)\right\} = E\left\{\left[\hat{S}_{BT}(\omega) - E\left\{\hat{S}_{BT}(\omega)\right\}\right]^2\right\}$$

$$= \frac{1}{4\pi^2} E\left\{\left[\int_{-\pi}^{\pi} W(\omega-\xi)\hat{S}_{PER}(\xi)d\xi - \int_{-\pi}^{\pi} W(\omega-\xi)E\left\{\hat{S}_{PER}(\xi)\right\}d\xi\right]^2\right\}$$

$$= \frac{1}{4\pi^2} E\left\{\left[\int_{-\pi}^{\pi} W(\omega-\xi)\left[\hat{S}_{PER}(\xi) - E\left\{\hat{S}_{PER}(\xi)\right\}\right]d\xi\right]^2\right\}$$

$$= \frac{1}{4\pi^2} E\left\{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} W(\omega-\xi)W(\omega-\lambda)\left[\hat{S}_{PER}(\xi) - E\left\{\hat{S}_{PER}(\xi)\right\}\right]\left[\hat{S}_{PER}(\lambda) - E\left\{\hat{S}_{PER}(\lambda)\right\}\right]d\xi d\lambda\right\}$$

9. Practical Classical Methods (23)

$$= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} W(\omega-\xi) W(\omega-\lambda) \underbrace{E\left\{\left[\hat{S}_{PER}(\xi) - E\left\{\hat{S}_{PER}(\xi)\right\}\right]\left[\hat{S}_{PER}(\lambda) - E\left\{\hat{S}_{PER}(\lambda)\right\}\right]\right\}}_{=\operatorname{cov}\left\{\hat{S}_{PER}(\xi), \hat{S}_{PER}(\lambda)\right\}} d\xi d\lambda$$

$$= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} W(\omega-\xi) W(\omega-\lambda) \operatorname{cov}\left\{\hat{S}_{PER}(\xi), \hat{S}_{PER}(\lambda)\right\} d\xi d\lambda$$

Now, further approximation must be done do get a meaningful result. If N is large enough, the sin-over-sin kernel will be enough like a delta function (with area $2\pi/N$) to treat it approximately as one:

$$\operatorname{var}\left\{\hat{S}_{BT}(\omega)\right\} \approx \frac{1}{2\pi N} \int_{-\pi}^{\pi} W^2(\omega-\lambda) S_x^2(\lambda) d\lambda$$

Now, further approximation: assume that the true PSD is fairly constant over any interval of width = to mainlobe of $W(\omega)$:

$$\operatorname{var}\left\{\hat{S}_{BT}(\omega)\right\} \approx \frac{1}{2\pi N} S_x^2(\omega) \int_{-\pi}^{\pi} W^2(\omega-\lambda) d\lambda = S_x^2(\omega) \frac{1}{N} \sum_{k=-M}^{M} w^2[k]$$

Paserval theorem

9. Practical Classical Methods (24)

$$E\left\{\hat{S}_{BT}(\omega)\right\} \approx \frac{1}{2\pi} W(\omega) \underset{circ}{*} S_x(\omega)$$ Bias

$$\operatorname{var}\left\{\hat{S}_{BT}(\omega)\right\} \approx S_x^2(\omega)\left[\frac{1}{N}\sum_{k=-M}^{M} w^2[k]\right]$$ Variance

Basic Tradeoff between Bias and Variance:

- Need large M to get small bias
  - – In order to get narrow mainlobe and low sidelobes.
- Need $M << N$ to get low variance
  - – In order to reduce the bracketed term in variance equation.

<u>Recommended</u>: $M < N/5$

Performance measures

- **Variability** (small v is desirable):

$$v = \frac{\operatorname{var}\{\hat{S}(\omega)\}}{E^2\{\hat{S}(\omega)\}}$$

- **Resolution:** One of the important measures of goodness for spectral analysis is resolution –the **ability to see two closely-spaced sinusoids**.

  The width of the mainlobe of the window's kernel impacts this ability. There are many ways to measure resolution, one way is:

$$\Delta\omega = 6\,\text{dB Width of Mainlobe}$$

  Small $\Delta\omega$ is desirable. Two things impact mainlobe width:

  1. Window length: $\Delta\omega \downarrow$ as Length $\uparrow$
  2. Window shape (e.g. Hanning, Hamming, etc.)

  There is a tradeoff between $\Delta\omega$ and sidelobe level.

9. Practical Classical Methods (26)

- **<u>Overall Figure of Merit</u>**

  It is helpful to have a single measure by which to compare methods. This is done using the following **Figure of Merit**:

$$\mathcal{M} = \nu \times \Delta\omega$$

  Since ν and $\Delta\omega$ are both required to be as small as possible, we also want the figure of merit to be **as small as possible**.

9. Practical Classical Methods (27)

❑ Performance comparison of classical methods

	Variability ν	Resolution $\Delta\omega$	Merit $\mathcal{M}$
Periodogram	1	$0.89\frac{2\pi}{N}$	$0.89\frac{2\pi}{N}$
Bartlett	$\frac{L}{N}$	$0.89\frac{2\pi}{L}$	$0.89\frac{2\pi}{N}$
Welch (50% Overlap & Bartlett Window)	$\frac{9}{16}\frac{L}{N}$	$1.28\frac{2\pi}{L}$	$0.72\frac{2\pi}{N}$
Blackman-Tukey	$\frac{2}{3}\frac{M}{N}$	$0.64\frac{2\pi}{M}$	$0.43\frac{2\pi}{N}$

Other Windows Can be Used

=

Trade-Off Between Variance & Resolution

Res ~ 1/(DFT Size)

Merit ~ $1/N$

9. Practical Classical Methods (28)

- **<u>Complexity comparison of classical methods</u>**

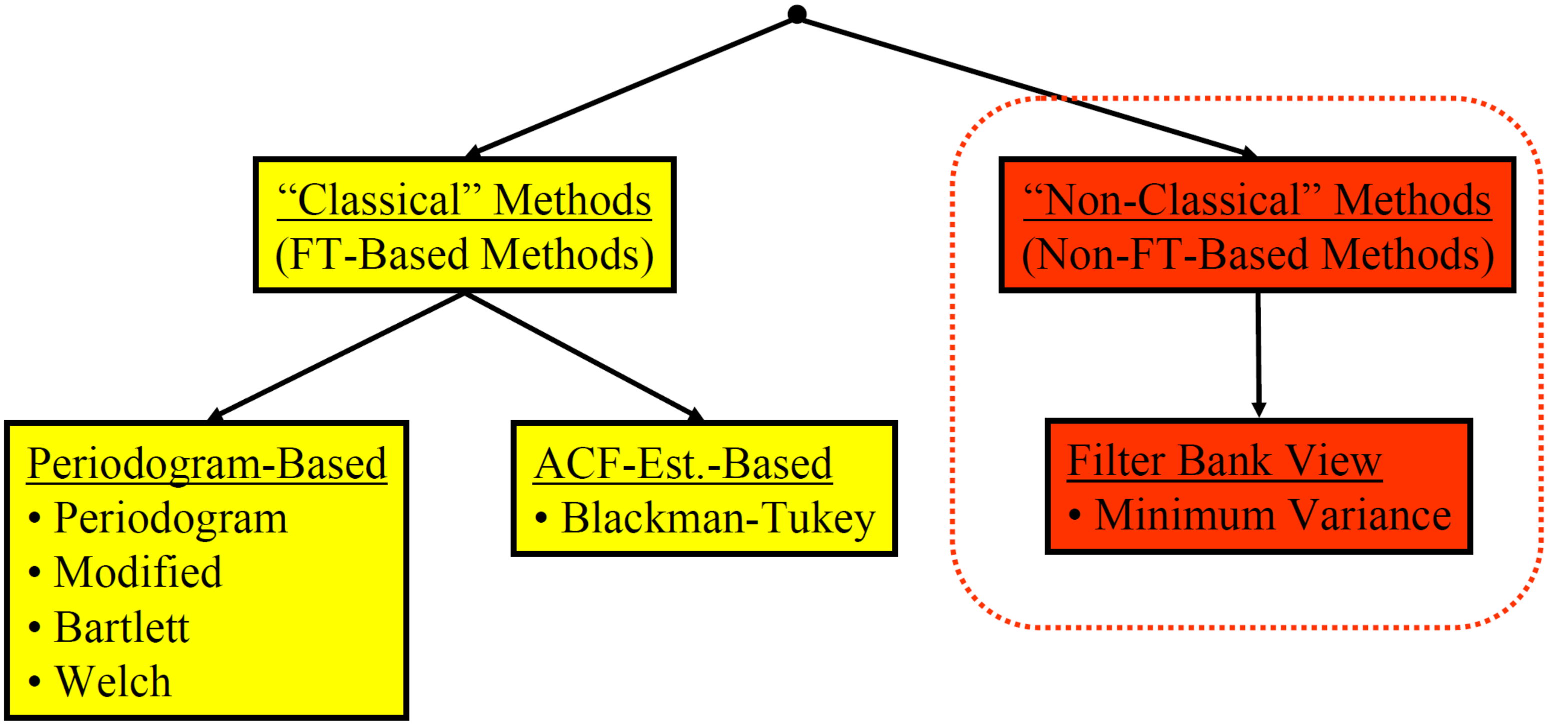
Welch and BT methods are the most commonly used ones. But counting the number of complex multiplies needed for each one, it is easy to see that: Welch requires a bit more computation then BT.

But, for BT, none of the ACF lags can be estimated until ALL of the data is obtained –therefore no computing can be done until all the data is obtained. For Welch, DFT’s can be started as soon as each block arrives.

Welch might have a real-time advantage.

9. Minimum Variance Method (MVSE) (1)

Family of Non-Parametric Methods



9. Minimum Variance Method (2)

The Minimum Variance Spectral Estimation (MVSE) method has two other names:

- Maximum Likelihood Method (MLM)
- Capon's Method

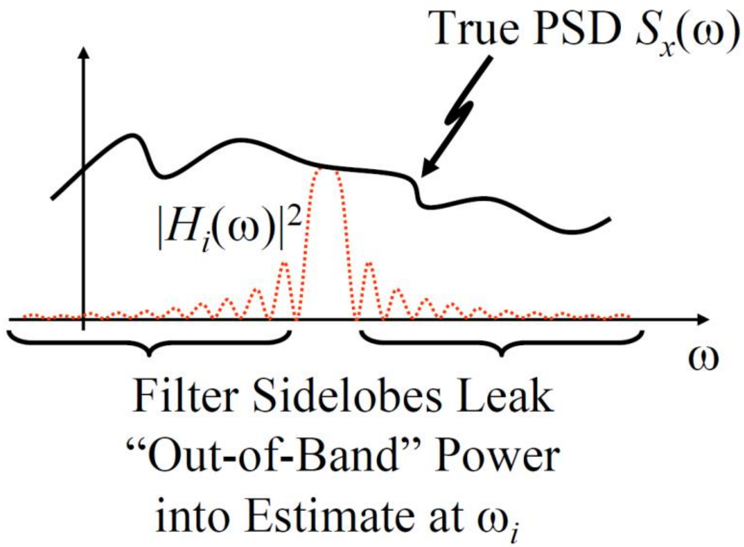
Note: The names MVSE and MLM are actually misnomers –this method:

- does NOT minimize the variance of the estimate
- does NOT maximize the "likelihood function"

9. Minimum Variance Method (3)

- **Recall: Filter bank view of periodogram**

  The problem is leakage from nearby frequencies:



9. Minimum Variance Method (4)

❑ Goal for MVSE Method

Figure out a way to design each filter bank channel response to minimize the leakage –this is thus a **data-dependent design**.

Collect Data ⇒ Design filters for filter bank

Want to design filters to **minimize the sidelobes while keeping the mainlobe height at 1**:

Design goals:

1. Want $H_i(\omega_i) = 1$ to let through the desired $S_x(\omega_i)$

2. Minimize total output power in the filter:

$$\rho_i = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|H_i(\omega)\right|^2 S_x(\omega)d\omega$$

This is equivalent to minimizing the sidelobe contribution even though the integral includes the desired ω_i

9. Minimum Variance Method (5)

- **<u>Useable form for ρ</u>**

  The frequency response of filter $h_i[n]$ is:

$$H_i(\omega) = \sum_{n=0}^{p-1} h_i[n] e^{-j\omega n} \qquad p = \text{Filter length, } p < N$$

  Using this in the expression for ρ gives:

$$\rho_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\sum_{k=0}^{p-1} h_i[k] e^{-j\omega k} \right] \left[\sum_{l=0}^{p-1} h_i^*[l] e^{j\omega l} \right] S_x(\omega) d\omega$$

$$= \sum_{k=0}^{p-1} \sum_{l=0}^{p-1} h_i[k] h_i^*[l] \underbrace{\frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) e^{j\omega(l-k)} d\omega}_{= r_x[l-k]}$$

$$= \sum_{k=0}^{p-1} \sum_{l=0}^{p-1} h_i[k] h_i^*[l] r_x[l-k]$$

Recognize as vector-matrix-vector multiplication, $\mathbf{R}_x$: correlation matrix

$$= \mathbf{h}_i^H \mathbf{R}_x \mathbf{h}_i$$

9. Minimum Variance Method (6)

For each i, minimize this: $\rho_i = \mathbf{h}_i^H \mathbf{R}_x \mathbf{h}_i$

Under this constraint:

$$H_i(\omega_i) = 1 \quad \Rightarrow \quad \mathbf{h}_i^H \mathbf{e}_i = 1$$

where

$$\mathbf{e}_i = [1 \quad e^{j\omega_i} \quad e^{j2\omega_i} \quad \cdots \quad e^{j(p-1)\omega_i}]^T$$

Most common way to do **constrained optimization** is using the **Lagrange Multiplier method**:

$$J = \mathbf{h}_i^H \mathbf{R}_x \mathbf{h}_i - \lambda\left(\mathbf{h}_i^H \mathbf{e}_i - 1\right)$$

Find $\mathbf{h}_i$ and λ to minimize J, in general set:

$$\frac{\partial J}{\partial \mathbf{h}_i} = \mathbf{0}^T \quad \& \quad \frac{\partial J}{\partial \lambda} = 0$$

But often an easier way is to do these two steps:

1. Do the partial w.r.t. $\mathbf{h}_i$ and solve for $\mathbf{h}_i$

2.Then choose λ to ensure solution meets the constraint

9. Minimum Variance Method (7)

Gradient:

If $g(\mathbf{x})$ is a scalar-valued function of a real-valued vector $\mathbf{x}$, then the gradient of $g(\mathbf{x})$ is defined as:

$$\nabla_{\mathbf{x}}(g) = \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial g(\mathbf{x})}{\partial x_1} & \frac{\partial g(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g(\mathbf{x})}{\partial x_N} \end{bmatrix}$$

Gradient here is nothing more than: the vector whose elements are the partials w.r.t. each element of $\mathbf{x}$.

(Note: There are similar definitions when $g(\mathbf{x})$ is vector-valued.)

<u>Example</u>: $g(\mathbf{x}) = \mathbf{c}^T\mathbf{x} = \sum_{i=1}^{N} c_i x_i = c_1x_1 + c_2x_2 + \cdots + c_n x_N$

$$\begin{aligned} \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} &= \begin{bmatrix} \frac{\partial g(\mathbf{x})}{\partial x_1} & \frac{\partial g(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g(\mathbf{x})}{\partial x_N} \end{bmatrix} \\ &= \begin{bmatrix} c_1 & c_2 & \cdots & c_N \end{bmatrix} \\ &= \mathbf{c}^T \end{aligned}$$

9. Minimum Variance Method (8)

Step 1: For our scalar-valued function J, we get:

$$\frac{\partial J}{\partial \mathbf{h}_i} = \mathbf{h}_{i,o}^H \mathbf{R}_x - \lambda \mathbf{e}_i^H \overset{set}{=} \mathbf{0}^T$$

Solve this for $\mathbf{h}_{i,o}$: $\mathbf{h}_{i,o}^H = \lambda \mathbf{e}_i^H \mathbf{R}_x^{-1}$

Step 2: Choose λ to make this solution satisfy constraint:

$$\mathbf{h}_{i,o}^H \mathbf{e}_i = 1 \quad \Rightarrow \quad \left(\lambda \mathbf{e}_i^H \mathbf{R}_x^{-1}\right)\mathbf{e}_i = 1 \quad \Rightarrow \quad \lambda = \frac{1}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i}$$

Now use this λ in optimal $\mathbf{h}_{i,o}$:

$$\boxed{\mathbf{h}_{i,o}^H = \frac{\mathbf{e}_i^H \mathbf{R}_x^{-1}}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i}}$$

This gives the optimal filter for estimating the power at the frequency ω_i

9. Minimum Variance Method (9)

In principle–we need to solve $\mathbf{h}_{i,o}$ for each frequency ω_i at which we wish to get a PSD estimate. Then we would compute the output power at each filter and that would be our PSD estimate (based on an equation for the output power: ρ_i)

The estimate of the power at ω_i is nothing more than the minimized value of ρ_i:

$$\rho_{i,o} = \mathbf{h}_{i,o}^H \mathbf{R}_x \mathbf{h}_{i,o}$$

$$= \left[\frac{\mathbf{e}_i^H \mathbf{R}_x^{-1}}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i}\right] \mathbf{R}_x \left[\frac{\mathbf{e}_i^H \mathbf{R}_x^{-1}}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i}\right]^H = \left[\frac{\mathbf{e}_i^H \mathbf{R}_x^{-1}}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i}\right] \mathbf{R}_x \left[\frac{\mathbf{R}_x^{-1} \mathbf{e}_i}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i}\right]$$

$$= \frac{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{R}_x \mathbf{R}_x^{-1} \mathbf{e}_i}{(\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i)(\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i)} = \frac{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i}{(\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i)(\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i)}$$

Thus, the **estimated power** at frequency ω_i is:

$$\hat{\sigma}_x^2(\omega_i) = \frac{1}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i}$$

9. Minimum Variance Method (10)

To get the power spectral density (PSD) we need to divide by the filter's bandwidth –for a filter of length *p* the bandwidth is approximately 1/*p* so our **MVSE PSD estimate** is:

$$\hat{S}_{MV}(\omega_i) = \frac{p}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i}$$

This estimate requires the ACF in matrix form. So, we need an estimate of the ACF in matrix form.

9. Minimum Variance Method (11)

❑ **Estimating ACF in matrix form**:

$$\hat{\mathbf{R}}_x = \begin{bmatrix} \hat{r}_x[0] & \hat{r}_x[1] & \cdots & \hat{r}_x[p-1] \\ \hat{r}_x[-1] & \hat{r}_x[0] & \ddots & \\ \vdots & \ddots & \ddots & \hat{r}_x[1] \\ \hat{r}_x[-p+1] & \cdots & \hat{r}_x[-1] & \hat{r}_x[0] \end{bmatrix}$$

where

$$\hat{r}_x[k] = \frac{1}{N} \sum_{n=0}^{N-|k|-1} x[n+k]x^*[m]$$

9. Minimum Variance Method (12)

- **<u>MVSE –Comments:</u>**

  Implementation of MVSE:

  - Generally done directly on the data matrix **X** for efficiency.
  - Even with that, it is more complex than classical methods.

  Performance of MVSE:

  - Provides better resolution than classical methods.
  - Mostly used when spiky spectra are expected

    (Although, the AR methods are usually better in that case)

  If required resolution can be met with classical methods, then use them. If not –consider either MVSE or Parametric methods.

Parametric Spectral Estimation

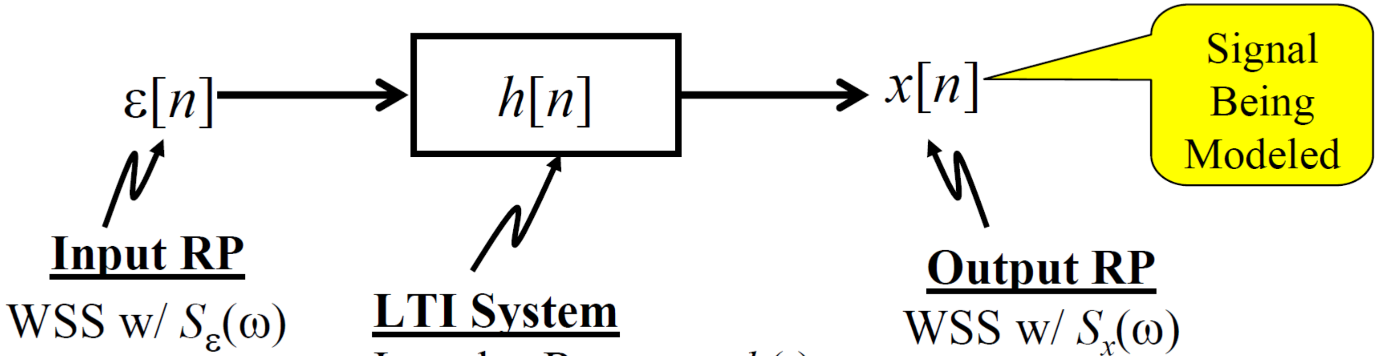
9. Parametric Spectral Estimation (1)

Time series models = Parametric models

Time series = Discrete-time random signal

9. Parametric Spectral Estimation (2)

❑ Motivation for time series models



Impulse Response $h(t)$

Frequency Response $H(\omega) = \mathfrak{F}\{h(t)\}$

$$\boxed{S_x(\omega) = |H(\omega)|^2 S_\varepsilon(\omega)}$$

If the input $\varepsilon[n]$ is white with power σ^2 then: $S_x(\omega) = |H(\omega)|^2 \sigma^2$

Then the shape of output PSD is completely set by $H(\omega)$.

Thus, under this model, **knowing the LTI system's transfer function (or frequency response) tells everything about the PSD**.

9. Parametric Spectral Estimation (3)

The **transfer function** of an LTI system is completely determined by a set of parameters $\{b_k\}$ and $\{a_k\}$:

$$H(z) = \frac{B(z)}{A(z)} = \frac{1+\sum_{k=1}^{q} b_k z^{-k}}{1+\sum_{k=1}^{p} a_k z^{-k}}$$

If we can assure that the random processes can be **modeled** as the output of a LTI system driven by white noise, then

Estimating Parameters = Estimating PSD

9. Parametric Spectral Estimation (4)

- **Parametric PSD models**

  The most general parametric PSD model is then:

$$S_x(\omega) = \sigma^2 \frac{\left|1 + \sum_{k=1}^{q} b_k e^{-j\omega k}\right|^2}{\left|1 + \sum_{k=1}^{p} a_k e^{-j\omega k}\right|^2}$$

Model parameters: $\sigma^2, \quad \{a_k\}_{k=1}^{p}, \quad \{b_k\}_{k=1}^{q}$

The output of the LTI system gives a time-domain model for the process:

$$x[n] = -\sum_{k=1}^{p} a_k x[n-k] + \sum_{k=0}^{q} b_k \varepsilon[n-k]$$

$$(b_0 = 1)$$

9. Parametric Spectral Estimation (5)

There are three special cases that are considered for these models:

- Autoregressive (AR)
- Moving Average (MA)
- Autoregressive Moving Average (ARMA)

9. Parametric Spectral Estimation (6)

❑ <u>Autoregressive (AR) PSD models</u>

If the LTI system's model is constrained to have only poles, then:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{k=1}^{p} a_k z^{-k}}$$

$$\underbrace{x[n] = -\sum_{k=1}^{p} a_k x[n-k] + \varepsilon[n]}_{\text{Output depends regressively on itself}} \quad (b_0 = 1)$$

Order of the model is p: called **AR(p) model**

$$\boxed{S_{AR}(\omega) = \frac{\sigma^2}{\left|1 + \sum_{k=1}^{p} a_k e^{-j\omega k}\right|^2}}$$

Poles give rise to PSD spikes.

9. Parametric Spectral Estimation (7)

Moving Average (MA) PSD models

If the LTI system's model is constrained to have only zeros, then:

$$H(z) = B(z) = 1 + \sum_{k=1}^{q} b_k z^{-k}$$

$$\underbrace{x[n] = -\sum_{k=0}^{q} b_k \varepsilon[n-k],}_{\text{Output is an average of values inside a moving window}} \quad b_0 = 1$$

Order of the model is q: called **MA(q) model.**

$$\boxed{S_{MA}(\omega) = \sigma^2 \left| 1 + \sum_{k=1}^{q} b_k e^{-j\omega k} \right|^2}$$

Zeros give rise to PSD nulls.

9. Parametric Spectral Estimation (8)

- **Autoregressive Moving Average (ARMA)**

If the LTI system's model is allowed to have Poles & Zeros, then:

$$H(z) = \frac{B(z)}{A(z)} = \frac{1 + \sum_{k=1}^{q} b_k z^{-k}}{1 + \sum_{k=1}^{p} a_k z^{-k}}$$

$$x[n] = -\sum_{k=1}^{p} a_k x[n-k] + \sum_{k=0}^{q} b_k \varepsilon[n-k]$$

$$(b_0 = 1)$$

Order of the model is p,q : called **ARMA(p,q) model**

$$S_x(\omega) = \sigma^2 \frac{\left|1 + \sum_{k=1}^{q} b_k e^{-j\omega k}\right|^2}{\left|1 + \sum_{k=1}^{p} a_k e^{-j\omega k}\right|^2}$$

Poles & Zeros give rise to PSD spikes and nulls.

9. Parametric Spectral Estimation (9)

- **ACF model of a process**

  So far we've seen relationships between:

  - PSD model
  - Time-domain model

  These models impart a corresponding model to the ACF:

  Let the process obey an ARMA(p,q) model:

$$x[n] = -\sum_{k=1}^{p} a_k x[n-k] + \sum_{k=0}^{q} b_k \varepsilon[n-k]$$

  To get ACF: multiply both sides of this by $x[n\text{-}k]$ and take $E\{\,\}$:

$$E\{x[n]x[n-k]\} = -\sum_{l=1}^{p} a_l E\{x[n-l]x[n-k]\} + \sum_{l=0}^{q} b_l E\{\varepsilon[n-l]x[n-k]\}$$

$$\Rightarrow \quad r_x[k] = -\sum_{l=1}^{p} a_l r_x[k-l] + \sum_{l=0}^{q} b_l r_{x\varepsilon}[k-l]$$

9. Parametric Spectral Estimation (10)

To evaluate this –write $x[n]$ as output of filter with input $\varepsilon[n]$:

$$
\begin{aligned}
r_{x\varepsilon}[k] &= E\{x[n]\varepsilon[n+k]\} \\
&= E\left\{\varepsilon[n+k]\sum_{l=-\infty}^{\infty} h[n-l]\varepsilon[l]\right\} \\
&= \sum_{l=-\infty}^{\infty} h[n-l]E\{\varepsilon[n+k]\varepsilon[l]\} \\
&= \sum_{l=-\infty}^{\infty} h[n-l]\sigma^2\delta[n+k-l] \\
&= h[-k]\sigma^2
\end{aligned}
$$

We have assumed a causal filter for a model:

$$r_{x\varepsilon}[k] = 0 \quad k > 0$$

9. Parametric Spectral Estimation (11)

Using this result gives the **Yule-Walker equations for ARMA:**

$$r_x[k] = \begin{cases} -\sum_{l=1}^{p} a_l r_x[k-l] + \sigma^2 \sum_{l=0}^{q} b_{l+k} h[l] & k = 0, 1, \ldots, q \\ -\sum_{l=1}^{p} a_l r_x[k-l] & k \geq q+1 \end{cases} \qquad \text{(}\mathbf{ARMA}\text{)}$$

These equations are the key to estimating the model parameters.

We now look at simplifications of these for the AR and MA cases.

9. Parametric Spectral Estimation (12)

- **<u>ACF model for an AR process</u>**

Specializing to the AR case, we set q= 0 and get:

$$r_x[k] = \begin{cases} -\sum_{l=1}^{p} a_l r_x[k-l] + \sigma^2 h[0] & k = 0 \\ -\sum_{l=1}^{p} a_l r_x[k-l] & k \geq 1 \end{cases}$$

Now, we see that

$$\underbrace{h[0] = \lim_{z\to\infty} H(z)}_{\text{Initial value theorem for Z-Transform}} = \lim_{z\to\infty} \frac{1 + \sum_{k=1}^{q} b_k z^{-k}}{1 + \sum_{k=1}^{p} a_k z^{-k}} = 1$$

Initial value theorem
for Z-Transform

Yule-Walker equations (AR):

$$r_x[k] = \begin{cases} -\sum_{l=1}^{p} a_l r_x[k-l] + \sigma^2 & k = 0 \\ -\sum_{l=1}^{p} a_l r_x[k-l] & k \geq 1 \end{cases} \qquad \textbf{(AR)}$$

If we look at $k = 0, 1, \ldots p$ for these AR Yule-Walker equations, we get $p+1$ simultaneous equations that can be solved for the $p+1$ model parameters of $\{a_i\}_{i=1,\ldots,p}$ and σ^2:

9. Parametric Spectral Estimation (14)

Yule-Walker Equations (AR)

$$\begin{bmatrix} r_x[0] & r_x[1] & \cdots & r_x[p-1] \\ r_x[1] & r_x[0] & \cdots & \vdots \\ \vdots & \cdots & \ddots & r_x[1] \\ r_x[p-1] & \cdots & r_x[1] & r_x[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} r_x[1] \\ r_x[2] \\ \vdots \\ r_x[p] \end{bmatrix}$$

$$\sigma^2 = r_x[0] + \sum_{l=1}^{p} a_l r_x[l]$$

If we know the $p \times p$ autocorrelation matrix, then we can solve these equations for the model parameters.

9. Parametric Spectral Estimation (15)

ACF model for an MA process

Specializing to the MA case, we set $p = 0$ and get:

$$r_x[k] = \begin{cases} \sigma^2 \sum_{l=0}^{q-k} b_{l+k} h[l] & k = 0, 1, \ldots, q \\ 0 & k \geq q+1 \end{cases}$$

But, for the MA case the system is a FIR filter and we have

$$h[k] = \begin{cases} b_k, & k = 0, 1, \cdots, q \\ 0, & otherwise \end{cases}$$

Yule-Walker equations (MA):

$$r_x[k] = \begin{cases} \sigma^2 \sum_{l=0}^{q-|k|} b_{l+k} b_l & |k| = 0, 1, \ldots, q \\ 0 & |k| \geq q+1 \end{cases} \qquad \textbf{(MA)}$$

9. Parametric Spectral Estimation (16)

❑ Parametric PSD estimation

As mentioned above, the idea here is to find a good estimate of the model parameters and then use those to get an estimate of the PSD. The basic idea holds regardless if it is ARMA, AR, or MA. We consider only the AR case.

Here is the general AR method: Given data $\{x[n], 0 \leq n \leq N\text{-}1\}$

1. Estimate the $p \times p$ autocorrelation matrix from the data:

$$\{x[n], 0 \leq n \leq N-1\} \quad \Rightarrow \quad \{\hat{r}[k], 0 \leq k \leq p\}$$

2. Solve the AR Yule-Walker equations for the AR model:

9. Parametric Spectral Estimation (17)

$$\begin{bmatrix} \hat{r}_x[0] & \hat{r}_x[1] & \cdots & \hat{r}_x[p-1] \\ \hat{r}_x[1] & \hat{r}_x[0] & \cdots & \vdots \\ \vdots & \cdots & \ddots & \hat{r}_x[1] \\ \hat{r}_x[p-1] & \cdots & \hat{r}_x[1] & \hat{r}_x[0] \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix} = - \begin{bmatrix} \hat{r}_x[1] \\ \hat{r}_x[2] \\ \vdots \\ \hat{r}_x[p] \end{bmatrix}$$

$$\hat{\sigma}^2 = \hat{r}_x[0] + \sum_{l=1}^{p} \hat{a}_l \hat{r}_x[l]$$

3. Compute the PSD estimate from the model

$$\boxed{\hat{S}_{AR}(\omega) = \frac{\hat{\sigma}^2}{\left| 1 + \sum_{k=1}^{p} \hat{a}_k e^{-j\omega k} \right|^2}}$$

9. Parametric Spectral Estimation (18)

Two common methods (but there are many others):

- **Autocorrelation method**

  Estimate the ACF using:

$$\hat{r}_x[k] = \frac{1}{N} \sum_{i=0}^{N-1-k} x[i]x[i+|k|], \quad 0 \le k \le p$$

- **Covariance method**

  Estimate using:

$$\hat{c}_{jk} = \frac{1}{N-p} \sum_{n=p}^{N-1} x[n-j]x[n-k], \quad 0 \le j,k \le p$$

9. Parametric Spectral Estimation (19)

Solve using

$$\begin{bmatrix} \hat{c}_{11} & \hat{c}_{12} & \cdots & \hat{c}_{1p} \\ \hat{c}_{21} & \hat{c}_{22} & \cdots & \hat{c}_{2p} \\ \vdots & \cdots & \ddots & \vdots \\ \hat{c}_{p1} & \hat{c}_{p2} & \cdots & \hat{c}_{pp} \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix} = - \begin{bmatrix} \hat{c}_{10} \\ \hat{c}_{20} \\ \vdots \\ \hat{c}_{p0} \end{bmatrix}$$

$$\hat{\sigma}^2 = \hat{c}_{00} + \sum_{l=1}^{p} \hat{a}_l \hat{c}_{0l}$$

9. Parametric Spectral Estimation (20)

- **<u>Least Squares method & linear prediction</u>**

  There is another method that is often used that comes at the problem from a little different direction.

  Recall: The above idea was based on the Yule-Walker equations, which are in terms of the ACF (which is unknown in practice). Thus we need to estimate the ACF to use this view.

  Least Squares provides a different way to estimate the AR parameters.

  Recall: The output of an AR model is given by

$$\varepsilon[n] \longrightarrow \boxed{\frac{1}{1+\sum_{k=1}^{p} a_k z^{-k}}} \longrightarrow x[n] = -\sum_{k=1}^{p} a_k x[n-k] + \varepsilon[n]$$

9. Parametric Spectral Estimation (21)

If we re-arrange this output equation we get:

$$x[n] - \underbrace{\left[-\sum_{k=1}^{p} a_k x[n-k] \right]}_{\hat{x}[n]} = \varepsilon[n]$$

Prediction Error

Prediction of $x[n]$

There are lots of applications where linear prediction is used: Data compression, target tracking, noise cancellation etc.

Goal: Find a set of prediction coefficients $\{a_k\}$ such that the sum of squares (least squares) of the prediction error is minimized.

$$\text{minimize} \quad V = \frac{1}{N\text{-}p} \sum_{n=p}^{N-1} \varepsilon^2[n]$$

9. Parametric Spectral Estimation (22)

To choose the $\{a_k\}$ to minimize V we differentiate and set = 0

$$\frac{\partial V}{\partial a_l} = \frac{2}{N-p}\sum_{n=p}^{N-1}\frac{\partial \varepsilon[n]}{\partial a_l}\varepsilon[n] = \frac{2}{N-p}\sum_{n=p}^{N-1}x[n-l]\varepsilon[n]$$

Now we use:

$$\varepsilon[n] = x[n] - \underbrace{\left[-\sum_{k=1}^{p}a_k x[n-k]\right]}_{\hat{x}[n]} = \sum_{k=0}^{p}a_k x[n-k]; \quad a_0 = 1$$

$$\frac{\partial V}{\partial a_l} = \frac{2}{N-p}\sum_{n=p}^{N-1}x[n-l]\left[\sum_{k=0}^{p}a_k x[n-k]\right]$$

$$= \frac{2}{N-p}\sum_{k=0}^{p}a_k\sum_{n=p}^{N-1}x[n-l]x[n-k] = 0; \quad 1 \le l \le p$$

9. Parametric Spectral Estimation (23)

So to solve the LS linear prediction problem we need:

$$\frac{2}{N-p}\sum_{k=0}^{p} a_k \sum_{n=p}^{N-1} x[n-l]x[n-k] = 0; \quad 1 \le l \le p \qquad (*)$$

Define:

1. Matrix $\mathbf{\Gamma}$ with elements λ_{lk}
2. Vector $\boldsymbol{\lambda}$ with elements λ_{l0}
3. Vector $\mathbf{a}$ with elements $a_1, \ldots, a_p$

where

$$\lambda_{lk} = \frac{1}{N-p}\sum_{n=p}^{N-1} x[n-l]x[n-k]; \quad 1 \le l, k \le p$$

Then (*) can be written as (exploiting that $a_0 = 1$):

$$\mathbf{\Gamma a} + \boldsymbol{\lambda} = \mathbf{0} \quad \text{or} \quad \mathbf{a} = -\mathbf{\Gamma}^{-1}\boldsymbol{\lambda}$$

Modern Spectral Estimation

High Resolution Spectral Estimation Based on Subspace Eigen-Analysis

- ❑ Pisarenko harmonic decomposition
- ❑ Multiple Signal Classification (MUSIC) spectral estimation

9. Eigenvector-Based Methods

- A very common problem in spectral estimation is concerned with the extraction of uncorrelated sinusoids from noise.
- The so-called eigen-decomposition methods are amongst the best at high SNR.
- They are used also extensively in array signal processing, e.g. for the estimation of the Direction of Arrival (DoA).

  (Measurements of spatial frequencies is equivalent to direction finding)
- These techniques can **resolve frequencies that are closely spaced** and hence are often referred to as **super-resolution** methods

9. Mathematical Background (1)

- **<u>The signal model</u>**: Assume that the signal is given by

$$x[n] = \sum_{i=1}^{p} a_i e^{j(n\omega_i + \phi_i)}$$

Then

$$r_{xx}[k] = E\{x[n]x[n-k]\} = \sum_{i=1}^{p} p_i e^{jn\omega_i}$$

where

$$p_i = a_i^2$$

9. Mathematical Background (2)

Set

$$\mathbf{x}[n] = \begin{bmatrix} x[n] \\ x[n+1] \\ . \\ . \\ x[n+M-2] \\ x[n+M-1] \end{bmatrix}$$

the autocorrelation matrix:

$$\mathbf{R}_{xx} = \begin{bmatrix} r_{xx}[0] & r_{xx}[-1] & . & r_{xx}[-(M-1)] \\ r_{xx}[1] & r_{xx}[0] & . & r_{xx}[-(M-2)] \\ . & . & . & . \\ r_{xx}[M-1] & . & . & r_{xx}[0] \end{bmatrix}$$

9. Mathematical Background (3)

Then we can write the autocorrelation matrix as

$$\mathbf{R}_{xx} = \sum_{k=1}^{p} p_k s_k s_k^H$$

where

$$\mathbf{s}_i = \begin{bmatrix} 1 \\ e^{j\omega_1} \\ e^{j2\omega_1} \\ \vdots \\ e^{j(M-2)\omega_1} \\ e^{j(M-1)\omega_1} \end{bmatrix}$$

9. Mathematical Background (4)

And hence

$$\mathbf{R}_{xx} = \mathbf{S}\mathbf{P}\mathbf{S}^{H}$$

where

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & . & . & \mathbf{s}_p \end{bmatrix}$$

and

$$\mathbf{P} = diag(\begin{bmatrix} p_1, & p_2, & . & . & p_p \end{bmatrix})$$

The vector space

$$S = span\{\mathbf{s}_1, \quad \mathbf{s}_2, \quad . \quad . \quad \mathbf{s}_p\}$$

is the **signal subspace** of $\{x[n]\}$

If $M > p$, then $\mathbf{R}_{xx}$ has rank p.

9. Mathematical Background (5)

Let the eigenvalues of $\mathbf{R}_{xx}$ be

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_M$$

corresponding to the normalized eigenvectors:

$$\mathbf{u}_1, \quad \mathbf{u}_2, \quad . \quad . \quad \mathbf{u}_M$$

Then

$$\mathbf{R}_{xx}\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

and

$$\mathbf{R}_{xx} = \sum_{i=1}^{M} \lambda_i \mathbf{u}_i \mathbf{u}_i^H$$

9. Mathematical Background (6)

Since $\mathbf{R}_{xx}$ is of rank p then

$$\lambda_{p+1} = \lambda_{p+2} = \lambda_{p+3} = \ldots = \lambda_M = 0$$

and hence

$$\mathbf{R}_{xx} = \sum_{i=1}^{p} \lambda_i \mathbf{u}_i \mathbf{u}_i^H$$

The eigenvectors

$$\mathbf{u}_1, \quad \mathbf{u}_2, \quad . \quad . \quad \mathbf{u}_p$$

are the **principal eigenvectors** of $\mathbf{R}_{xx}$.

An important result is **the principal eigenvectors span the signal subspace**. Thus given a sequence of observations we can determine the autocorrelation matrix and its eigenvectors.

Knowing the first *p* eigenvectors we can determine the space in which the signals reside even though at this point we do not know their frequencies.

9. Mathematical Background (7)

- **<u>The noise model</u>**: We assume that we observe signal contaminated additively, by a stationary, zero mean, white noise, independent of it

$$y[n] = x[n] + w[n]$$

Then

$$r_{yy}[k] = r_{xx}[k] + \sigma_w^2 \delta[k]$$

From the above we have

$$r_{yy}[k] = \sum_{i=1}^{p} p_i e^{jk\omega_i} + \sigma_w^2 \delta[k]$$

Moreover with

$$\mathbf{y}[n] = \begin{bmatrix} y[n] \\ y[n+1] \\ . \\ . \\ y[n+M-2] \\ y[n+M-1] \end{bmatrix} \qquad \mathbf{w}[n] = \begin{bmatrix} w[n] \\ w[n+1] \\ . \\ . \\ w[n+M-2] \\ w[n+M-1] \end{bmatrix}$$

9. Mathematical Background (8)

We have

$$\mathbf{R}_{yy} = \mathbf{R}_{xx} + \sigma_w^2 \mathbf{I}$$

Clearly $\mathbf{R}_{yy}$ is of rank M , i.e full rank.

Let its eigenvalues be

$$\mu_1 \geq \mu_2 \geq \mu_3 \geq \ldots \geq \mu_M$$

where the first p are

$$\mu_i = \lambda_i + \sigma_w^2$$

While the rest are all equal to the variance of noise. Thus we can write:

$$\mathbf{R}_{yy} = \sum_{i=1}^{p} (\lambda_i + \sigma_w^2) \mathbf{u}_i \mathbf{u}_i^H + \sum_{i=p+1}^{M} \sigma_w^2 \mathbf{u}_i \mathbf{u}_i^H$$

The space

$$N = span\{\mathbf{u}_{p+1}, \quad \mathbf{u}_{p+2}, \quad . \quad . \quad \mathbf{u}_M \}$$

is called the **noise subspace**.

<u>Important result</u>: Any vector in the signal subspace is orthogonal to the noise subspace.

9. Pisarenko Harmonic Decomposition (1)

The Pisarenko harmonic decomposition exploits the orthogonality of two subspaces directly.

Let the number of sinusoids (modes) p be known.

Set $M = p + 1$, so that the noise is spanned a single vector $\mathbf{u}_M$

Note that $\mathbf{u}_M$ must be orthogonal to all the signal subspace vectors:

$$\mathbf{s}_i^H \mathbf{u}_M = 0 \qquad i = 1, 2, \ldots, p$$

with

$$\mathbf{u}_M = \begin{bmatrix} u_{M,0} & u_{M,1} & . & . & u_{M,M-1} \end{bmatrix}^T$$

We have

$$\sum_{k=0}^{M-1} u_{M,k} e^{-jk\omega_i} = 0$$

9. Pisarenko Harmonic Decomposition (2)

The last equation is a polynomial in $e^{-j\omega_i}$ and hence its M -1 = p which all lie on the unit circle correspond to the frequencies of the sinuoidal signal.

The amplitudes are obtained from the autocorrelation relationships of the observations as given earlier.

The noise strength is given from the last eigenvalue of the same autocorrelation matrix.

9. Multiple Signal Classification (MUSIC) (1)

- **Multiple Signal Classification** relies on the same principle of orthogonality.

  Let

  $$\mathbf{s}^T(\omega) = \begin{bmatrix} 1 & e^{j\omega} & e^{j2\omega} & . & e^{j(M-1)\omega} \end{bmatrix}$$

  $$\mathbf{s}^H(\omega)\mathbf{x} = 0$$

  Now define the function:

  $$M(\omega) = \sum_{k=p+1}^{M} \left| \mathbf{s}^H(\omega)\mathbf{u}_k \right|^2$$

  Clearly $M(\omega_i) = 0$. Hence its reciprocal is infinite.Thus the reciprocal of the above function exhibits **peaks at the input frequencies**.

  The signal strengths can be computed as in the Pisarenko Harmonic Decomposition.

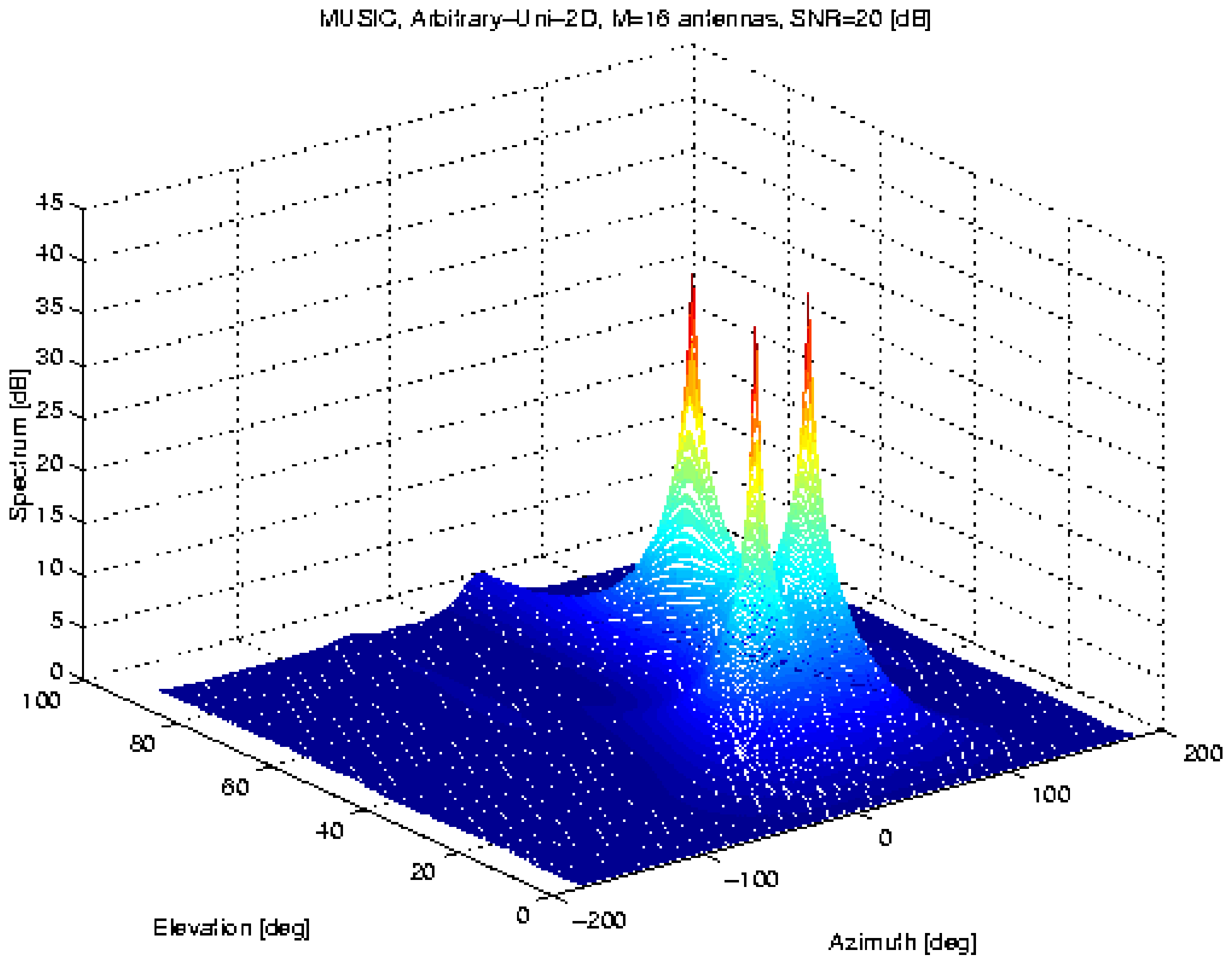
9. Multiple Signal Classification (MUSIC) (2)

- The quantity below is known as the **MUSIC spectrum**.

$$P(\omega) = \frac{1}{M(\omega)} = \frac{1}{\sum_{k=p+1}^{M} \left| \mathbf{s}^H(\omega)\mathbf{u}_k \right|^2}$$

9. Multiple Signal Classification (MUSIC) (3)

MUSIC spectrum example:



9. Multiple Signal Classification (MUSIC) (4)