ESTIMATION OF DETERMINISTIC PARAMETERS: METHOD OF MOMENTS AND MAXIMUM LIKELIHOOD ESTIMATORS

Fulvio GINI
Dipartimento di Ingegneria dell'Informazione
University of Pisa
Via G. Caruso 16, I-56122, Pisa, Italy
fulvio.gini@unipi.it

Taxonomy of the Estimation Problems

Deterministic parameters:

- probability of an event $p = \Pr\{A\}$
- moments of a distribution (mean value, variance, skewness, kurtosis, *etc*.)
- parameters of a distribution
- parameters of a signal (amplitude, phase, frequency, time of arrival, *etc*.)
- moments of a random vector (mean vector, correlation matrix, *etc*.)
- moments of a random signal (mean value function, ACF, PSD, *etc*.)

ACF=Autocorrelation Function, PSD=Power Spectral Density

Deterministic parameters:

- Sample estimates (application of the Law of large numbers)
- Method of Moments (MM)
- Maximum Likelihood (ML) method
- Least Squares (LS): Linear LS or Non Linear LS (NLLS)

Estimation of Deterministic Parameters

The Estimation Problem: Optimal Estimator

- **Question**: Does it exist an estimator that is the most efficient? i.e. an estimator that has the lowest MSE?

the **Minimum Mean Square Error (MMSE) Estimator**

- For example, assume we want to estimate the mean value η of a r.v. *X* given a vector **X** of independent observations, i.e. the $\{X_i\}$ are *N* IID r.v.'s: Does it exist an estimator that is more efficient than the **Sample Mean**?

- Let us suppose that if it exists it has the following structure:

$$\hat{\eta} = \frac{\alpha}{N}\sum_{i=1}^{N} X_i = \alpha \bar{X} \quad \Rightarrow \quad \text{if } \alpha = 1 \;\rightarrow\; \hat{\eta} = \bar{X} \equiv \text{Sample Mean}$$

$$\text{where} \quad \bar{X} \triangleq \frac{1}{N}\sum_{i=1}^{N} X_i \text{ is the Sample Mean}$$

- What is the value of α that minimizes the mean square error (MSE)?

- What is the value of α that minimizes the mean square error (MSE)?

$$\bar{X} \triangleq \frac{1}{N}\sum_{i=1}^{N} X_i \quad \Rightarrow \quad E\{\bar{X}\} = \eta, \quad \operatorname{var}\{\bar{X}\} = \frac{\sigma^2}{N}$$

$$MSE(\alpha) = E\left\{(\hat{\eta} - \eta)^2\right\} = E\left\{\left(\alpha\bar{X} - \eta\right)^2\right\} = \alpha^2 E\{\bar{X}^2\} - 2\alpha\eta E\{\bar{X}\} + \eta^2$$

$$= \alpha^2\left(\eta^2 + \frac{\sigma^2}{N}\right) - 2\alpha\eta^2 + \eta^2$$

$$\frac{dMSE(\alpha)}{d\alpha} = 2\alpha\left(\eta^2 + \frac{\sigma^2}{N}\right) - 2\eta^2$$

$$\min_{\alpha} MSE(\alpha) \quad \Leftrightarrow \quad \frac{dMSE(\alpha)}{d\alpha} = 0$$

The Estimation Problem: Optimal Estimator

$$\min_{\alpha} MSE(\alpha) \quad \Leftrightarrow \quad \frac{dMSE(\alpha)}{d\alpha} = 0$$

$$\frac{dMSE(\alpha)}{d\alpha} = 2\alpha\left(\eta^2 + \frac{\sigma^2}{N}\right) - 2\eta^2 = 0 \quad \Rightarrow \quad \alpha = \frac{\eta^2}{\eta^2 + \frac{\sigma^2}{N}} = \frac{1}{1 + \frac{\sigma^2}{N\eta^2}}$$

$$\alpha_{MMSE} = \frac{1}{1 + \frac{1}{N\gamma}}$$

where $\gamma \triangleq \dfrac{\eta^2}{\sigma^2}$ is a sort of Signal to Noise Ratio (SNR)

$\rightarrow$ Note that: $0 \leq \alpha_{MMSE} \leq 1$ and $\lim\limits_{N \text{ or } \gamma \to \infty} \alpha_{MMSE} = 1$

In practice if $N \gg 1$ or $\gamma \gg 1 \quad \rightarrow \quad \alpha_{MMSE} \cong 1$

The Estimation Problem: Optimal Estimator

$$\hat{\eta}_{MMSE} = \alpha_{MMSE} \cdot \frac{1}{N}\sum_{i=1}^{N} X_i \quad \rightarrow \quad b\{\hat{\eta}_{MMSE}\} = E\{\eta - \hat{\eta}_{MMSE}\} = \eta - \alpha_{MMSE}\eta$$

$$b\{\hat{\eta}_{MMSE}\} = (1-\alpha_{MMSE})\eta = \left(1 - \frac{1}{1+\frac{1}{N\gamma}}\right)\eta = \frac{1}{1+N\gamma}\eta \neq 0$$

$$MMSE = MSE(\alpha_{MMSE}) = \frac{\sigma^2}{N} \cdot \frac{1}{1+\frac{1}{N\gamma}} = \alpha_{MMSE} \cdot \frac{\sigma^2}{N} \leq \frac{\sigma^2}{N}$$

- The problem is that the structure of this optimal (MMSE) estimator depends on the unknown parameter η we want to estimate, so it cannot be implemented, i.e. it is not feasible.
- In other words, it does not exist an estimator that is optimal (minimum MSE) whatever is the value of the unknown parameter we want to estimate.

The Estimation Problem: Optimal Estimator

- The results we have just derived tell us that the **Sample Mean** is non optimal but asymptotically optimal (i.e. for $N \to \infty$ or $\gamma \to \infty$), in fact for $N >> 1$ or $\gamma >> 1$ the unfeasible MMSE estimator and the Sample Mean coincide.

- Since an MMSE estimator for the estimation of deterministic parameters does not exist, we should look for other methods (in addition to the **Sample Estimators** that we have already introduced, which are suitable for some specific estimation problems).

- These other methods should provide feasible estimators, that even if sub-optimal, have good performance, i.e. as close as possible to the optimal one.

- One of the simplest and most widely adopted methods is the **Method of Moments** (**MM** or sometime **MoM**).

- We start assuming a specific statistical model, i.e. a specific pdf, for our discrete observed data. The observed data should depend on the parameter we want to estimate, so as a consequence also their pdf: $\mathbf{X}(\theta) \to f_{\mathbf{X}}(\mathbf{x};\theta)$.

The Method of Moments

The Method of Moments (MM)

- Once we assume a specific pdf model (Gaussian, Laplace, Exponential, Weibull, Chauchy, *t*-student, *etc*.), the model is completely specified, except for a number of unknown deterministic parameters θ, that must be derived in such a way to match at the "best" the theoretical model to the observed data → it is a "*data fitting*" problem.

- **Assumption**: the *N* observed data $\{X_i\}$ are independent realizations of a r.v. *X*.

- Let us calculate the relationship between the unknown parameter θ and one of the moments of the r.v. *X*, e.g. the mean value:

$$m_1 \triangleq E\{X\} = \int_{-\infty}^{+\infty} x f_X(x;\theta)dx \quad \Rightarrow \quad m_1 = h_1(\theta)$$

- If this relationship can be inverted, we can express the unknown parameter θ as a function of the mean value of *X*:

$$\theta = h_1^{-1}(m_1)$$

The Method of Moments (MM)

- Since the mean value m_1 is unknown a priori, we cannot derive the true value of θ from this relationship. However, we can replace the unknown mean value with its estimate, e.g. obtained by using the **Sample Mean** estimator :

$$\theta = h_1^{-1}(m_1) \quad \Rightarrow \quad \hat{\theta}_{MM} = h_1^{-1}(\hat{m}_1) \qquad \text{where} \qquad \hat{m}_1 \triangleq \frac{1}{N}\sum_{i=1}^{N} X_i$$

- The estimator we find in this way is called the **Method of Moments (MM)** estimator. To analyze the estimation performance, we have to derive the estimator **bias** and **MSE**.

- As concerning the **efficiency** of the estimator, we cannot say anything. However, since the Sample Mean is a **consistent** estimator, also the MM estimate of θ is **consistent** (under mild conditions on the function h_1):

$$\lim_{N\to\infty} \hat{\theta}_{MM} = \lim_{N\to\infty} h_1^{-1}(\hat{m}_1) = h_1^{-1}\left(\lim_{N\to\infty} \hat{m}_1\right) = h_1^{-1}\left(m_1\right) = \theta$$

The Method of Moments (MM)

- The **Method of Moments** is fairly simple and yields **consistent** estimators (under very weak assumptions), though these estimators are often **biased**.

- In fact, the unbiasdness of the Sample Mean does not imply that the MM estimator of θ is also unbiased:

$$E\left\{\hat{\theta}_{MM}\right\} = E\left\{h_1^{-1}(\hat{m}_1)\right\} \neq h_1^{-1}\left(E\{\hat{m}_1\}\right) = h_1^{-1}\left(m_1\right) = \theta$$

- This is because $h_1(.)$ and its inverse are in general non linear transformations.

The Method of Moments (MM)

- The **Method of Moments** can be easily extended to the case of multiple unknown parmaters:

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_P \end{bmatrix}^T$$

- To estimate *P* unknown parameters we need at least *P* equations. We can obtain these *P* equations by calculating *P* moments of the pdf of *X*.
- It is a good practice to calculate the mean and the first *P*-1 central moments:

$$\begin{cases} m_1 = E\{X_i\} = h_1(\boldsymbol{\theta}) \\ \mu_2 = E\{(X_i - m_1)^2\} = h_2(\boldsymbol{\theta}) \\ \quad \vdots \\ \mu_P = E\{(X_i - m_1)^P\} = h_P(\boldsymbol{\theta}) \end{cases} \Rightarrow \quad \boldsymbol{\mu} = \mathbf{h}(\boldsymbol{\theta})$$

- System of *P* equations (generally nonlinear) with *P* unknowns.

The Method of Moments (MM)

- Solving the system of equations with respect to vector θ, we get the inverse relationship between the moments and the unknown parameter vector:

$$\boldsymbol{\theta} = \mathbf{h}^{-1}(\boldsymbol{\mu})$$

- Since the moments are not a priori known, we should replace their values with their estimates, obtained by using **Sample Estimators**:

$$\hat{m}_1 \triangleq \frac{1}{N}\sum_{i=1}^{N} X_i, \qquad \hat{\mu}_k \triangleq \frac{1}{N}\sum_{i=1}^{N} \left(X_i - \hat{m}_1\right)^k, \quad k = 2, 3, \cdots, P$$

$$\hat{\boldsymbol{\mu}} = \left[\hat{m}_1 \quad \hat{\mu}_2 \quad \hat{\mu}_3 \quad \cdots \quad \hat{\mu}_P\right]^T$$

$$\Rightarrow \qquad \hat{\boldsymbol{\theta}}_{MM} = \mathbf{h}^{-1}(\hat{\boldsymbol{\mu}})$$

- The estimators obtained by the Method of Moments are **consistent** (under the very weak assumption that the central moments of X exist and are finite up to the order $2P$).

The Method of Moments (MM)

- **Example**: Estimate the parameter η of an Exponential pdf, given *N* independent realizations of the r.v. *X*, assumed to be exponentially distributed.

$$\underset{N\times 1}{\mathbf{X}} = \begin{bmatrix} X_1 & X_2 & \cdots & X_N \end{bmatrix}^T, \quad \{X_i\}_{i=1}^{N} \quad \text{IID}$$

$$f_{X_i}(x_i) = \eta e^{-\eta x_i} u(x_i) \quad \rightarrow \quad m_1 \triangleq E\{X_i\} = \frac{1}{\eta}, \quad \mu_2 \triangleq \mathrm{var}\{X_i\} = \frac{1}{\eta^2}.$$

- The distribution is monoparametric (it depends on one parameter only). Hence, to estimate η we need only one moment. We can use for example the mean value or the variance:

$$m_1 = E\{X_i\} = \frac{1}{\eta} \quad \rightarrow \quad \eta = \frac{1}{m_1} \quad \rightarrow \quad \hat{\eta} = \frac{1}{\hat{m}_1}$$

$$\mu_2 = \mathrm{var}\{X_i\} = \frac{1}{\eta^2} \quad \rightarrow \quad \eta = \frac{1}{\sqrt{\mu_2}} \quad \rightarrow \quad \tilde{\eta} = \frac{1}{\sqrt{\hat{\mu}_2}}$$

The Method of Moments (MM)

- Which one of the two should we use?

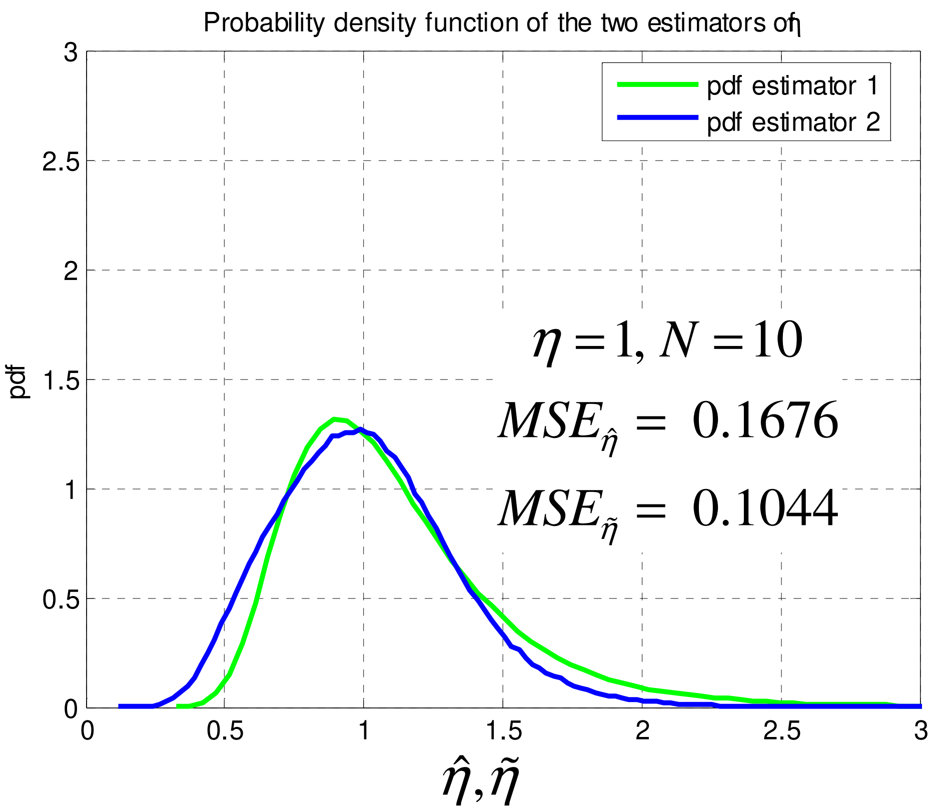
$$\hat{\eta} = \frac{1}{\hat{m}_1} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} X_i}$$

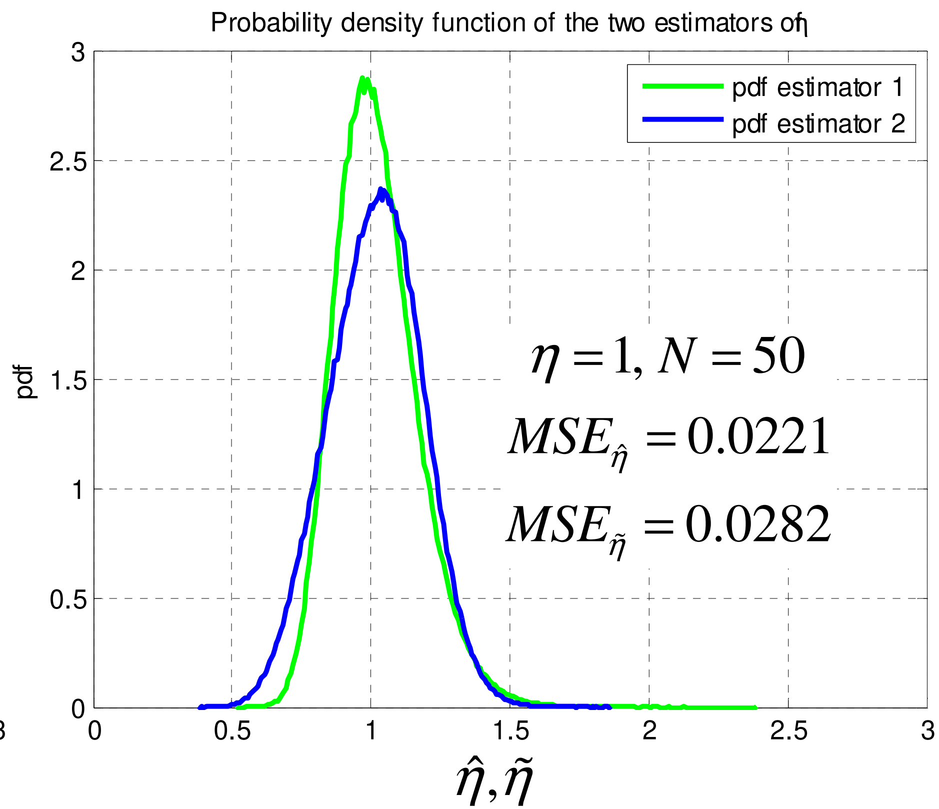
$$\tilde{\eta} = \frac{1}{\sqrt{\hat{\mu}_2}} = \frac{1}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{m}_1\right)^2}}$$

- The first estimator is simpler to implement (lower **computational complexity**).
- Let us compare them in terms of **estimation accuracy**.
- It is difficult to obtain the pdf's of the two estimators in closed-form. We derived the **normalized histogram** of the two estimators by generating a large number *M* of realizations of the estimates (the normalized histogram represents a non-parametric method to estimate the pdf of a r.v.).

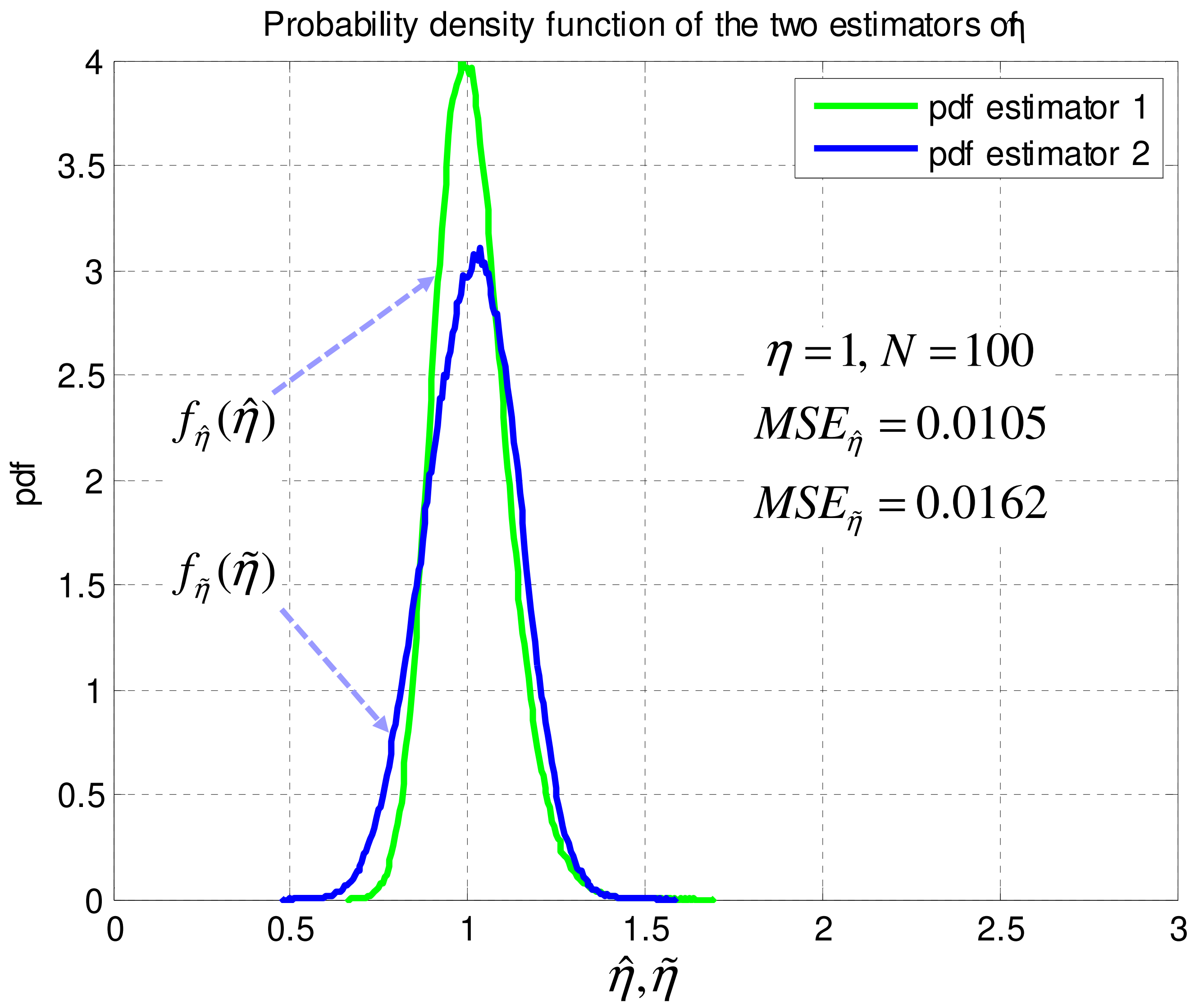
The Method of Moments (MM)

- The pdf's have been obtained by generating M=10^6 Monte Carlo runs.
- For N=10, estimator #2 (the one based on the variance) has MSE lower than that of estimator #1. However, for larger N the estimator #1 is always better (it is the one based on the Sample Mean, i.e. the moment of lowest order).





The Method of Moments (MM)



The Method of Moments (MM)

- **Example**: Estimate the two parameters *a* e *b* of a uniform pdf, given *N* independent realizations of the r.v. *X*, assumed to be uniformly distributed between *a* and *b*.

$$\underset{N\times 1}{\mathbf{X}} = \begin{bmatrix} X_1 & X_2 & \cdots & X_N \end{bmatrix}^T, \quad \{X_i\}_{i=1}^{N} \quad \text{IID}, \quad f_{X_i}(x_i) = \frac{1}{b-a} rect\left(\frac{x_i - \dfrac{a+b}{2}}{b-a} \right)$$

$$\rightarrow \quad m_1 = E\{X_i\} = \frac{a+b}{2}, \qquad \mu_2 = \text{var}\{X_i\} = \frac{(b-a)^2}{12}$$

- The distribution is biparametric, so we need at least two equations, i.e. we need to estimate at least two moments.
- We can obtain two equations by calculating the relationship between the mean value and the variance and the two parameters *a* and *b*.

The Method of Moments (MM)

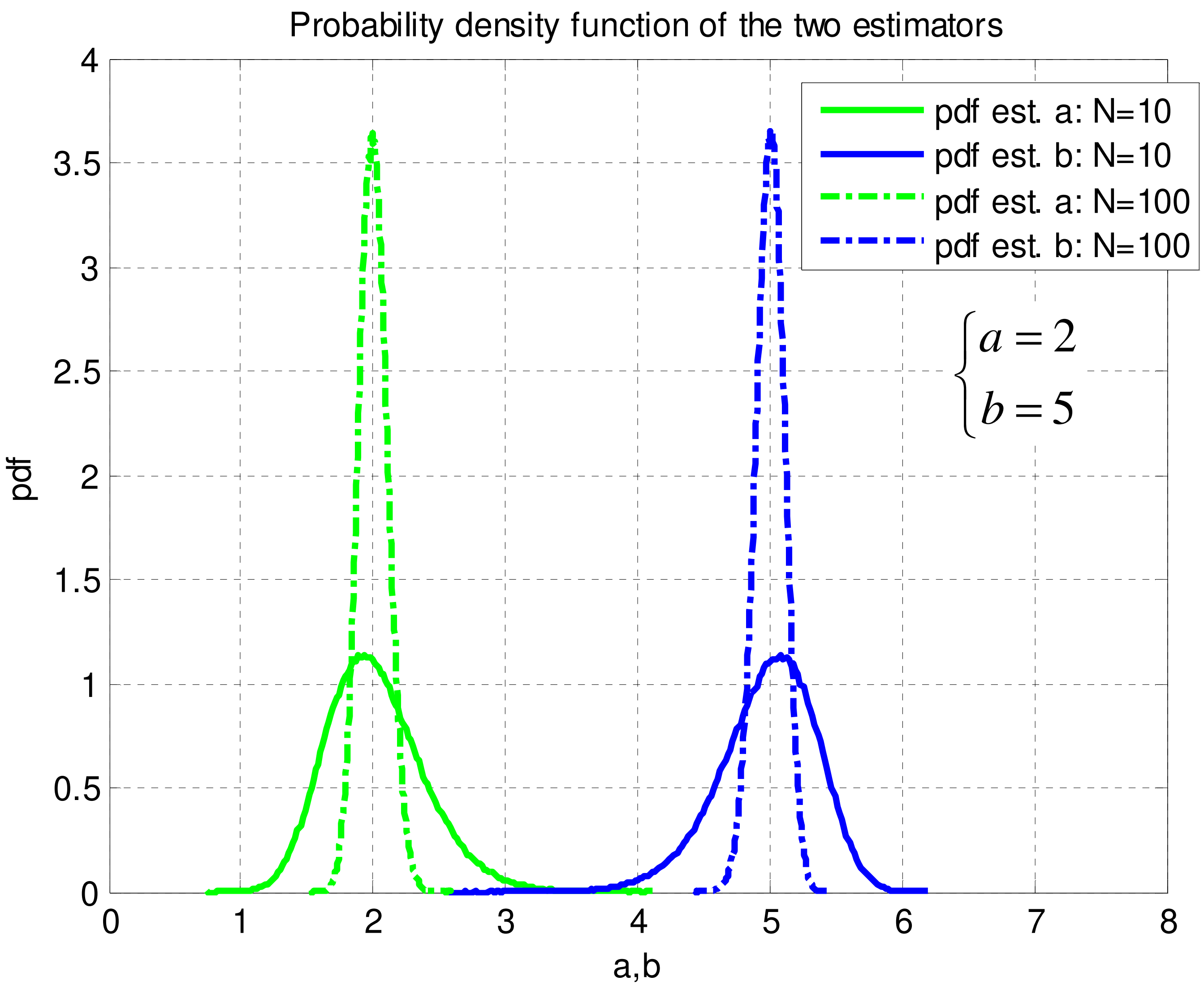
- Once we have the two equations, we have to solve them with respect to parameters *a* e *b*:

$$\begin{cases} m_1 = E\{X_i\} = \dfrac{a+b}{2} \\ \mu_2 = \mathrm{var}\{X_i\} = \dfrac{(b-a)^2}{12} \end{cases} \Rightarrow \begin{cases} b+a = 2m_1 \\ b-a = 2\sqrt{3\mu_2} \end{cases} \Rightarrow \begin{cases} a = m_1 - \sqrt{3\mu_2} \\ b = m_1 + \sqrt{3\mu_2} \end{cases}$$

- A this point, to obtain the MM estimate of parameters *a* and *b,* we just have to replace in the above equations the true values with their **Sample Estimates**:

$$\begin{cases} \hat{a}_{MM} = \hat{m}_1 - \sqrt{3\hat{\mu}_2} = \dfrac{1}{N}\sum_{i=1}^{N} X_i - \sqrt{3\dfrac{1}{N}\sum_{i=1}^{N}(X_i - \hat{m}_1)^2} \\ \hat{b}_{MM} = \hat{m}_1 + \sqrt{3\hat{\mu}_2} = \dfrac{1}{N}\sum_{i=1}^{N} X_i + \sqrt{3\dfrac{1}{N}\sum_{i=1}^{N}(X_i - \hat{m}_1)^2} \end{cases}$$

The Method of Moments (MM)



The Method of Moments (MM)

- **Example**: Estimate the two parameters η e σ² of a Gaussian pdf, given *N* independent realizations of the r.v. *X*, assumed to be Gaussian distributed.

$$\underset{N\times 1}{\mathbf{X}} = \begin{bmatrix} X_1 & X_2 & \cdots & X_N \end{bmatrix}^T, \quad \{X_i\}_{i=1}^{N} \quad \text{IID}$$

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\eta)^2}{2\sigma^2}} \quad \rightarrow \quad m_1 \triangleq E\{X_i\} = \eta, \quad \mu_2 \triangleq \mathrm{var}\{X_i\} = \sigma^2$$

- In this particular case, the two unknown parameters η e σ² coincide with the first two moments of the distribution, mean and variance. Hence, there are no equations to be solved. The **Sample Mean** and **Sample Variance** directly provides the MM estimates of parameters η e σ²:

$$\hat{\eta}_{MM} = \hat{m}_1 = \frac{1}{N}\sum_{i=1}^{N} X_i, \qquad \hat{\sigma}^2_{MM} = \hat{\mu}_2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{m}_1\right)^2$$

The Method of Moments (MM)

- The performance of the **Sample Mean** and **Sample Variance** have already been derived under the assumption that the *N* observed data are IID:

$$\hat{\eta} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$E\{\hat{\eta}\} = E\left\{\frac{1}{N}\sum_{i=1}^{N} X_i\right\} = \frac{1}{N}\sum_{i=1}^{N} E\{X_i\} = \frac{1}{N}\sum_{i=1}^{N}\eta = \eta$$

$$\mathrm{var}\{\hat{\eta}\} = \mathrm{var}\left\{\frac{1}{N}\sum_{i=1}^{N} X_i\right\} = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{var}\{X_i\} = \frac{\sigma^2}{N}$$

- The **Sample Mean** estimator is **unbiased** and **consistent**.
- Moreover, since in this case the data are Gaussian distributed and the estimator is **linear**, the Sample Mean estimate is also **Gaussian** distributed.

The Method of Moments (MM)

- The Gaussianity of the **Sample Mean** estimator is asymptotically (N>>1) valid in any case, even if the data are non Gaussian, thanks to the **Central-Limit Theorem**:

$$\hat{\eta} \overset{a}{\in} \mathcal{N}\left(\eta, \frac{\sigma^2}{N}\right)$$

- As concerning the **Sample Variance** estimator:

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{\eta}\right)^2$$

$$E\left\{\hat{\sigma}^2\right\} = E\left\{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{\eta}\right)^2\right\} = \left(1 - \frac{1}{N}\right)\sigma^2 \quad \rightarrow \quad b\left\{\hat{\sigma}^2\right\} = E\left\{\sigma^2 - \hat{\sigma}^2\right\} = \frac{\sigma^2}{N}$$

- The estimator is **biased** but **asymptotically unbiased**.

The Method of Moments (MM)

- **Assumption**: the observed data $\{X_i\}$ are IID.
- **Mean value** of the Sample Variance:

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{\eta}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}X_i^2 - 2\frac{1}{N}\sum_{i=1}^{N}X_i\hat{\eta} + \frac{1}{N}\sum_{i=1}^{N}\hat{\eta}^2 = \frac{1}{N}\sum_{i=1}^{N}X_i^2 - \hat{\eta}^2$$

$$E\left\{\hat{\sigma}^2\right\} = E\left\{\frac{1}{N}\sum_{i=1}^{N}X_i^2 - \hat{\eta}^2\right\} = E\left\{\frac{1}{N}\sum_{i=1}^{N}X_i^2\right\} - E\left\{\hat{\eta}^2\right\}$$

$$= \frac{1}{N}\sum_{i=1}^{N}E\left\{X_i^2\right\} - \left(\mathrm{var}\left\{\hat{\eta}\right\} + \left(E\left\{\hat{\eta}\right\}\right)^2\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(\sigma^2 + \eta^2\right) - \left(\frac{\sigma^2}{N} + \eta^2\right) = \sigma^2 - \frac{\sigma^2}{N} = \left(1 - \frac{1}{N}\right)\sigma^2$$

The Method of Moments (MM)

- As concerning the **variance** of the Sample Variance for IID data:

$$\operatorname{var}\left\{\hat{\sigma}^2\right\} = \frac{\mu_X(4) - \sigma^4}{N} - \frac{2\left(\mu_X(4) - 2\sigma^4\right)}{N^2} + \frac{\mu_X(4) - 3\sigma^4}{N^3}$$

$$\text{where} \quad \mu_X(4) \triangleq E\left\{\left(X_i - \eta\right)^4\right\}$$

- If the data are IID Gaussian distributed:

$$\mu_X(4) = 3\sigma^4 \quad \rightarrow \quad \operatorname{var}\left\{\hat{\sigma}^2\right\} = \frac{2\sigma^4}{N} - \frac{2\sigma^4}{N^2} = \frac{2\sigma^4}{N}\left(1 - \frac{1}{N}\right)$$

The Method of Moments (MM)

- When the data are IID Gaussian, we can also derive the exact pdf of the estimator **Sample Variance**, using the following known result:

$$\text{Given} \quad S \triangleq \sum_{i=1}^{N} \left(\frac{X_i - \hat{\eta}}{\sigma} \right)^2, \quad \text{where} \quad \hat{\eta} \triangleq \frac{1}{N} \sum_{k=1}^{N} X_k$$

if the $\{X_i\}$ are IID Gaussian, then S is a **central Chi-Squared** distributed r.v. with N-1 degrees of freedom. Its **mean** value and **variance** are given by:

$$E\{S\} = N - 1, \qquad \text{var}\{S\} = 2(N-1)$$

$$E\{\hat{\sigma}^2\} = E\left\{ \frac{\sigma^2}{N} S \right\} = \frac{\sigma^2}{N} E\{S\} = \frac{N-1}{N}\sigma^2 = \left(1 - \frac{1}{N}\right)\sigma^2$$

The Method of Moments (MM)

$$\operatorname{var}\left\{\hat{\sigma}^2\right\} = \operatorname{var}\left\{\frac{\sigma^2}{N}S\right\} = \frac{\sigma^4}{N^2}\operatorname{var}\left\{S\right\} = \frac{2(N-1)}{N^2}\sigma^4 = \frac{2\sigma^4}{N}\left(1-\frac{1}{N}\right)$$

$$MSE\left\{\hat{\sigma}^2\right\} = b^2\left\{\hat{\sigma}^2\right\} + \operatorname{var}\left\{\hat{\sigma}^2\right\} = \left(\frac{\sigma^2}{N}\right)^2 + \frac{2(N-1)}{N^2}\sigma^4 = \frac{2\sigma^4}{N}\left(1-\frac{1}{2N}\right)$$

- For large sample size (N>>1): $MSE\left\{\hat{\sigma}^2\right\} \cong \dfrac{2\sigma^4}{N}$

- For the Central-Limit Theorem, we have that the Sample Variance is asymptotically (N>>1) Gaussian distributed with the mean and variance derived above:

$$\hat{\sigma}^2 \overset{a}{\in} \mathcal{N}\left(\sigma^2, \frac{2\sigma^4}{N}\right)$$

The Method of Moments (MM)

- The exact pdf of the Sample Variance estimator is immediately obtained from the pdf of *S*, which is known in the scientific literature:

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{\eta}\right)^2 = \frac{\sigma^2}{N}S \quad \rightarrow \quad f_{\hat{\sigma}^2}(\hat{\sigma}^2) = \frac{N}{\sigma^2} f_S\left(\frac{N}{\sigma^2}\hat{\sigma}^2\right)$$

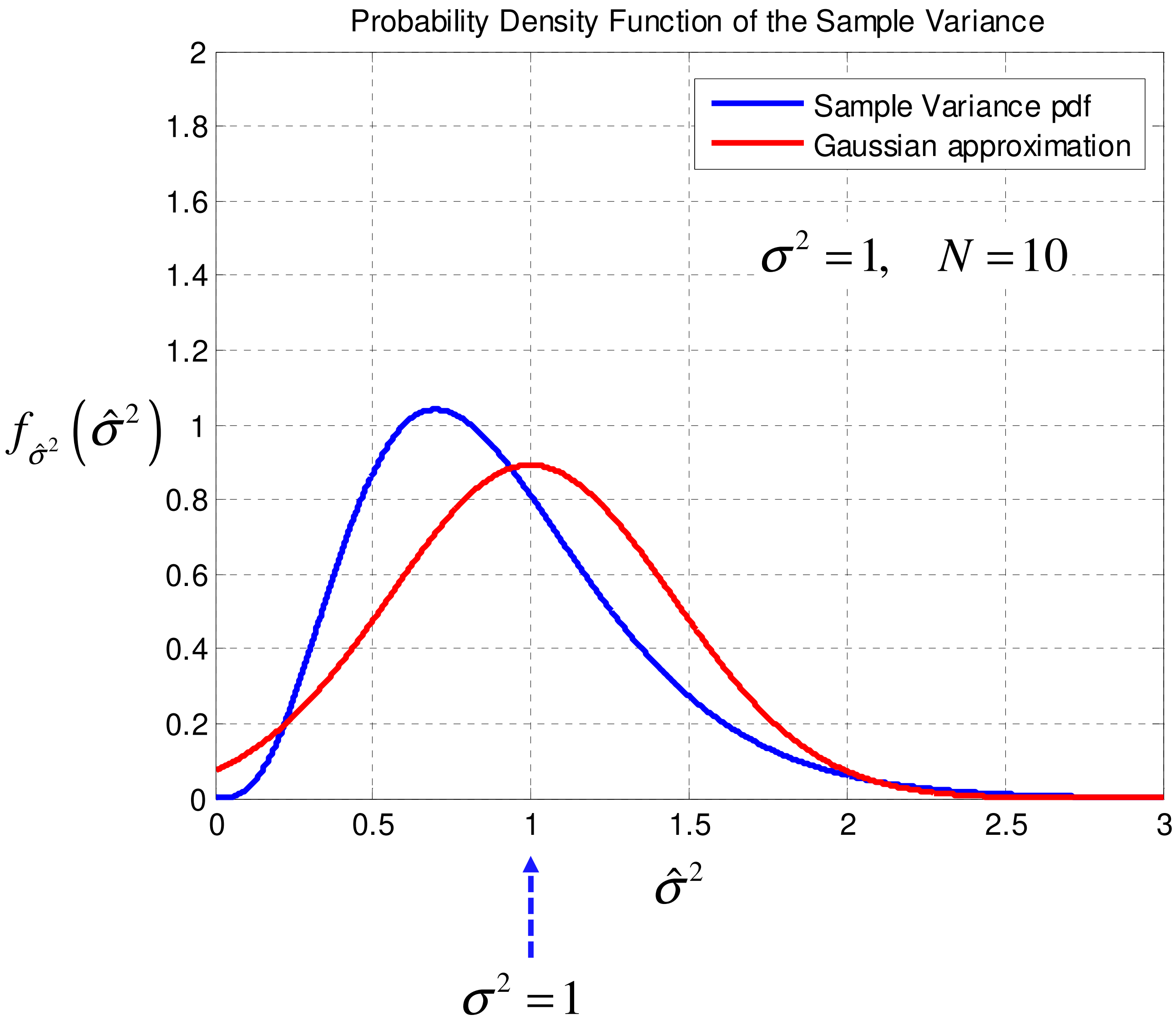
- The pdf of a a **central Chi-Squared** r.v. *S* with *N*-1 degrees of freedom is:

$$f_S\left(s\right) = \frac{1}{2^{(N-1)/2}\Gamma((N-1)/2)} s^{(N-1)/2-1} e^{-s/2} u(s)$$

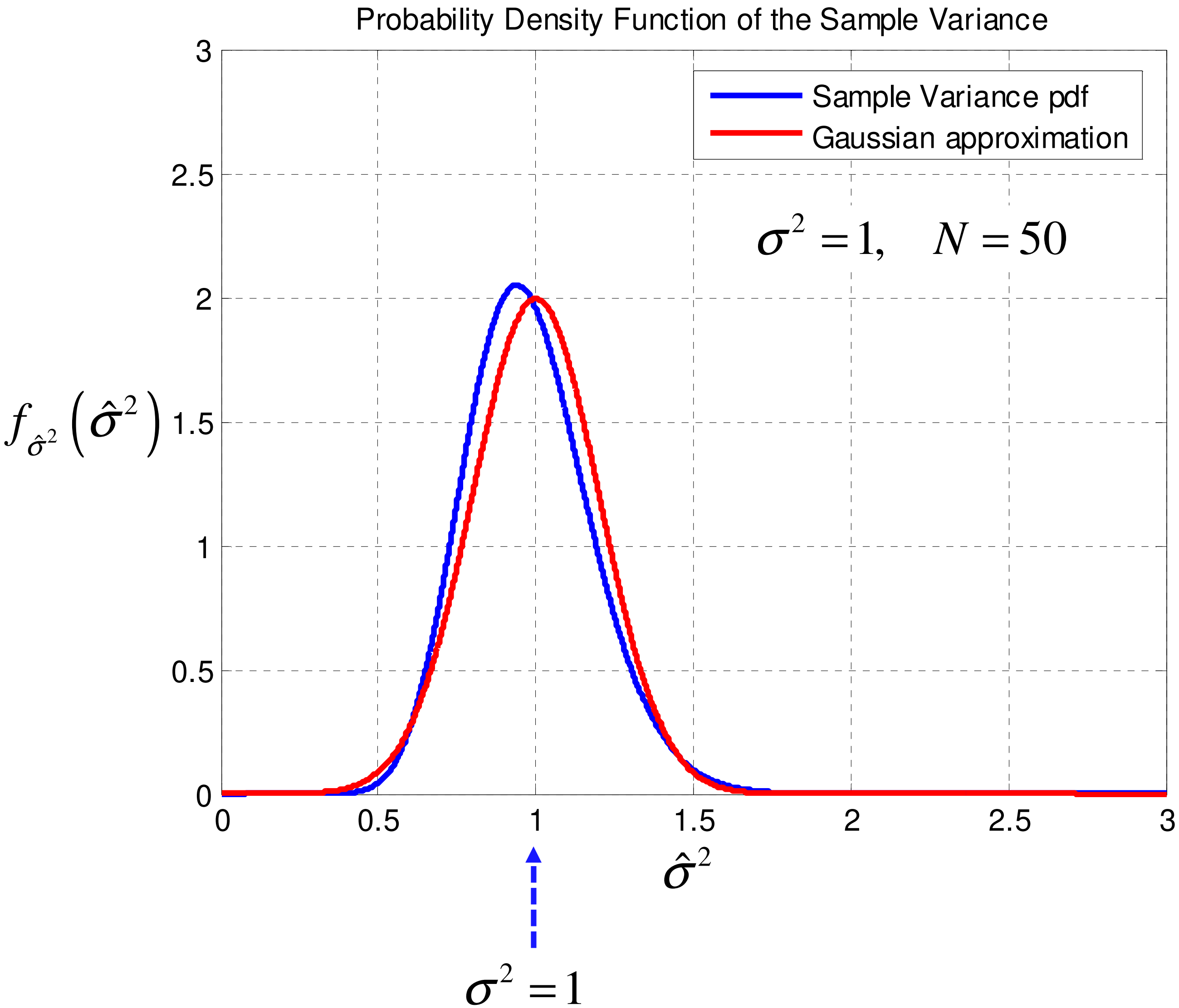
$$\text{where} \;\; \Gamma\left(\frac{k}{2}\right) = \sqrt{\pi}\,\frac{(k-2)!!}{2^{(k-1)/2}} \quad \text{per} \;\; k{=}1,3,5,7,\cdots$$

$$f_{\hat{\sigma}^2}\left(\hat{\sigma}^2\right) = \frac{1}{\Gamma((N-1)/2)}\left(\frac{N}{2\sigma^2}\right)^{(N-1)/2}\left(\hat{\sigma}^2\right)^{(N-1)/2-1} e^{-\frac{N}{2\sigma^2}\hat{\sigma}^2} u(\hat{\sigma}^2)$$

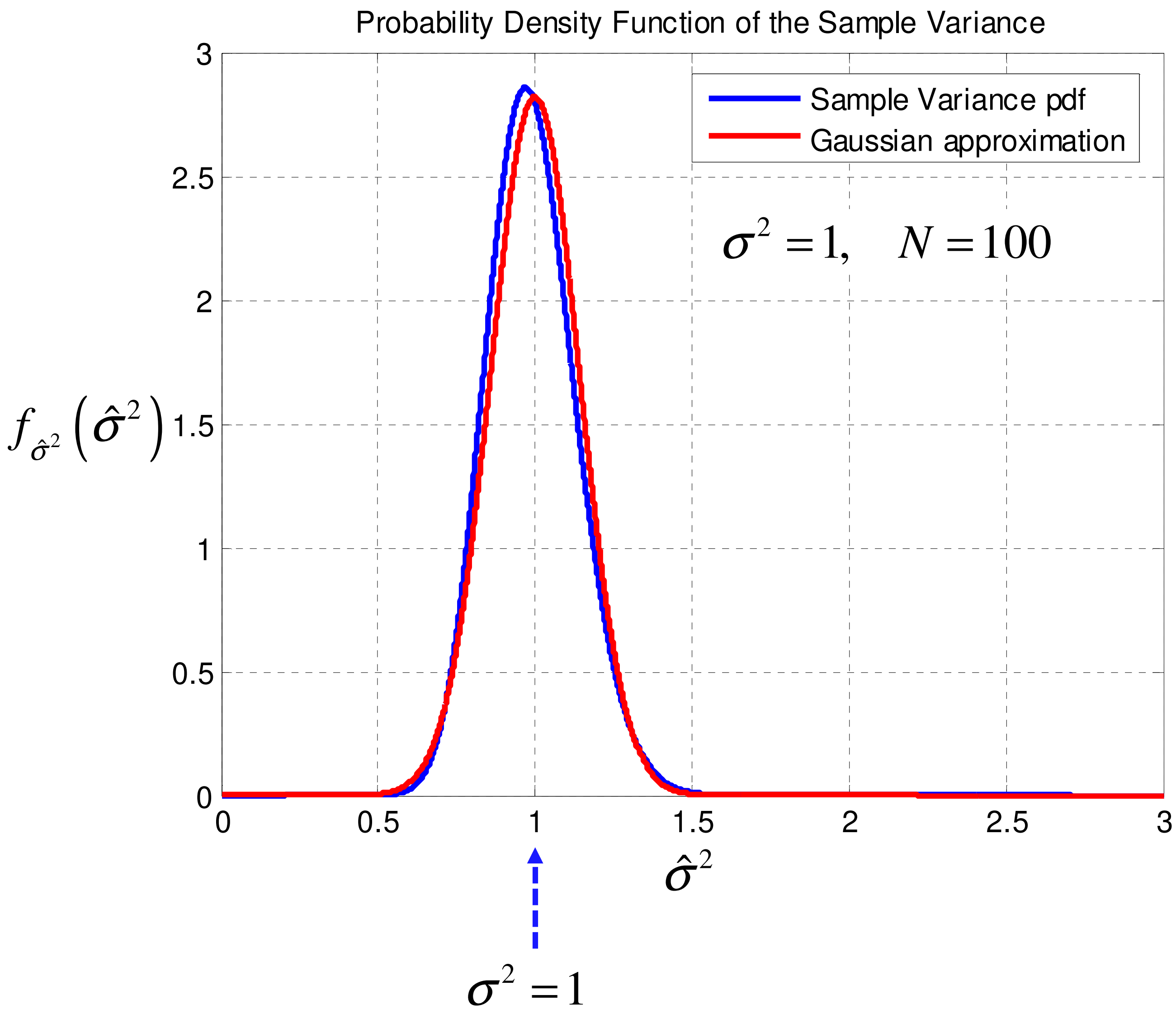
The Method of Moments (MM)



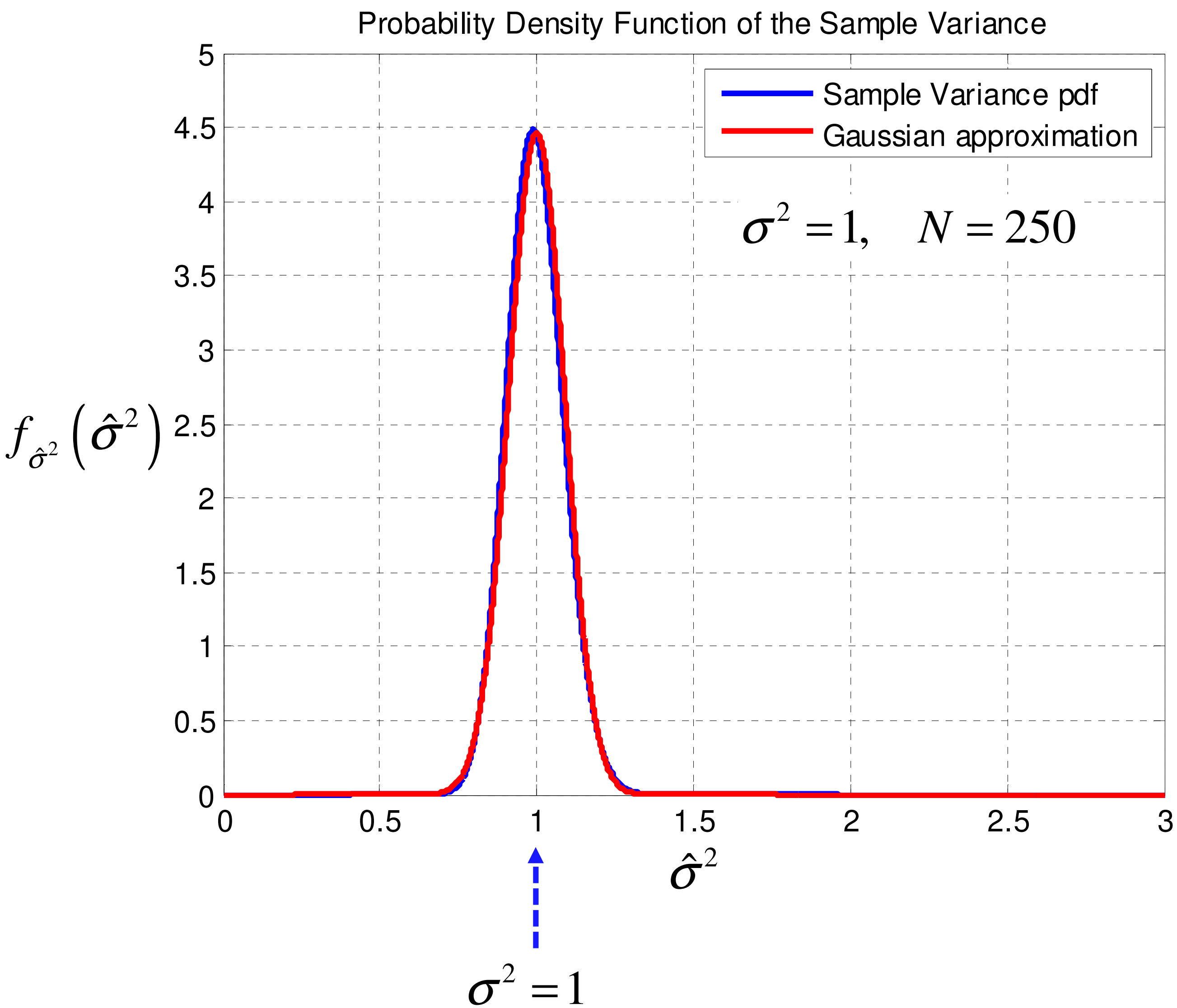
The Method of Moments (MM)



The Method of Moments (MM)

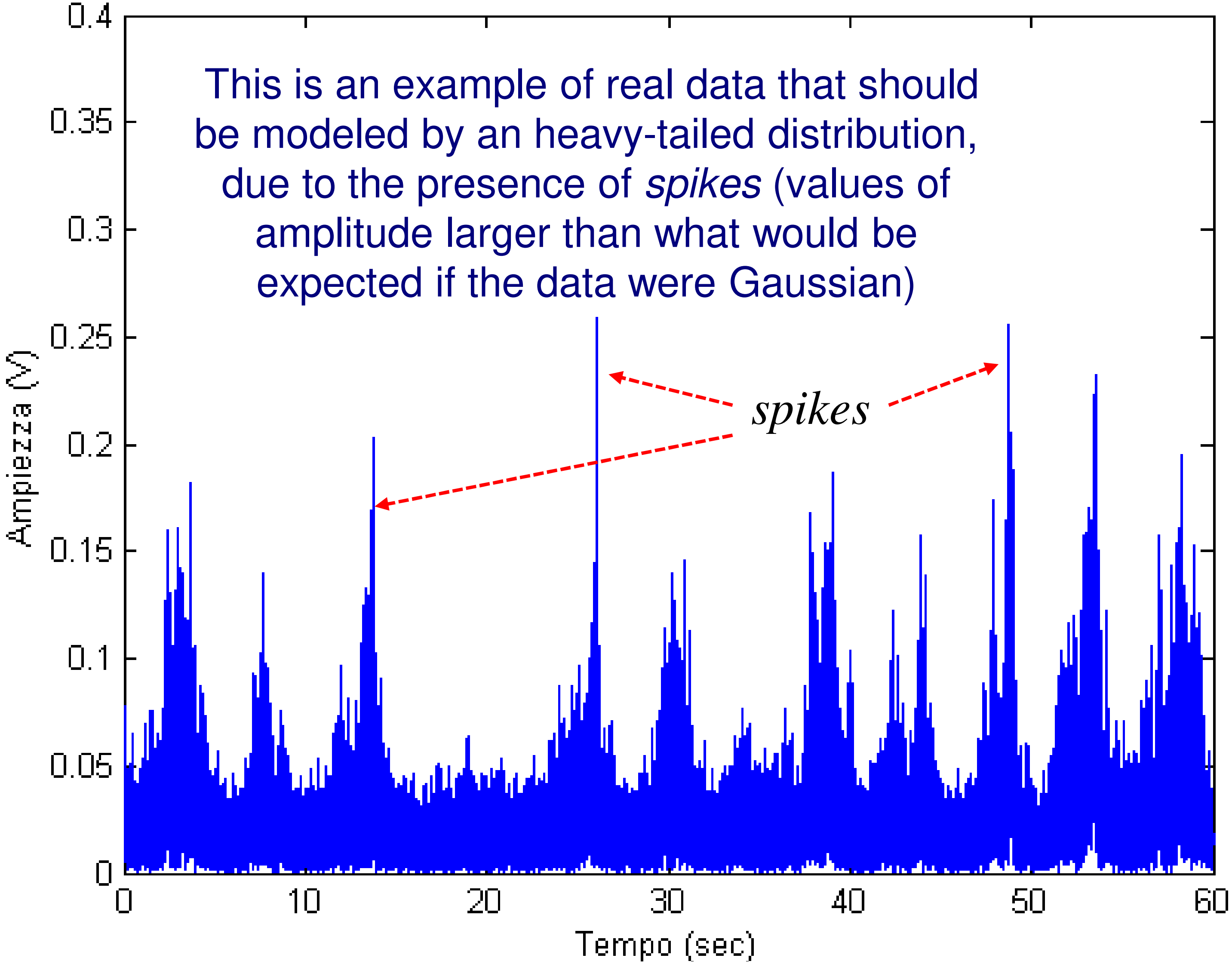


The Method of Moments (MM)



The Method of Moments (MM)

$$\text{Clutter amplitude:} \quad R = \left| X_I + jX_Q \right|$$



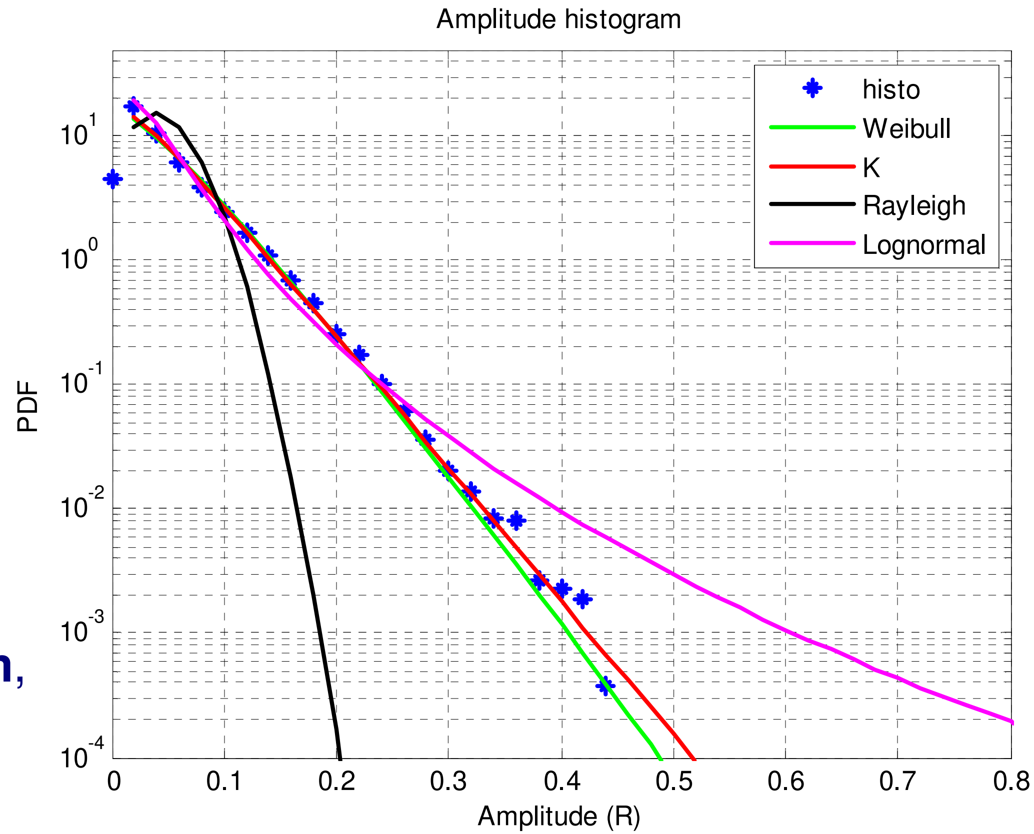
The Method of Moments (MM)

- **Real high-resolution sea clutter data.**

Clutter amplitude : $R = \left| X_I + jX_Q \right|$

- We want to find the model that "best" matches the observed data.

- We hypothesize different parametric models, estimate their parameters from the data, and then plot the various pdf's calculated by using the estimated values of the unknown parameters.



- Finally, we plot the data **histogram**, that represents a non model-based estimator of the data pdf.

- Finally, we select the one which provides the best fit.

$$f_R(r;\boldsymbol{\theta}) \quad \rightarrow \quad \hat{\boldsymbol{\theta}}_{MM} \quad \rightarrow \quad \hat{f}_R(r) = f_R\left(r;\hat{\boldsymbol{\theta}}_{MM}\right)$$

The Method of Moments (MM)

- The **Method of Moments (MM)** is fairly simple and yields **consistent** estimators (under very weak assumptions), though these estimators are often **biased**.

- In some respects, when estimating parameters of a known family of pdf's, this method was superseded by **Fisher's method** of **Maximum Likelihood (ML)**, because ML estimators have higher probability of being close to the quantities to be estimated and are more often unbiased.

- However, in some cases the likelihood equations may be intractable without computers and the ML estimators cannot be found in closed-form, whereas the MM estimators can be quickly and easily calculated by hand, provided we are able to solve the equation $\mu=\mathbf{h}(\theta)$.

The Method of Moments (MM)

- MM estimates may be used as the first approximation to the solutions of the likelihood equations, and successive improved approximations may then be found by the Newton–Raphson method. In this way the MM estimates can help in finding the ML estimates.

- In some cases, infrequent with large samples but not so infrequent with small samples, the MM estimates are outside of the parameter space. This problem never arises in the ML estimators.

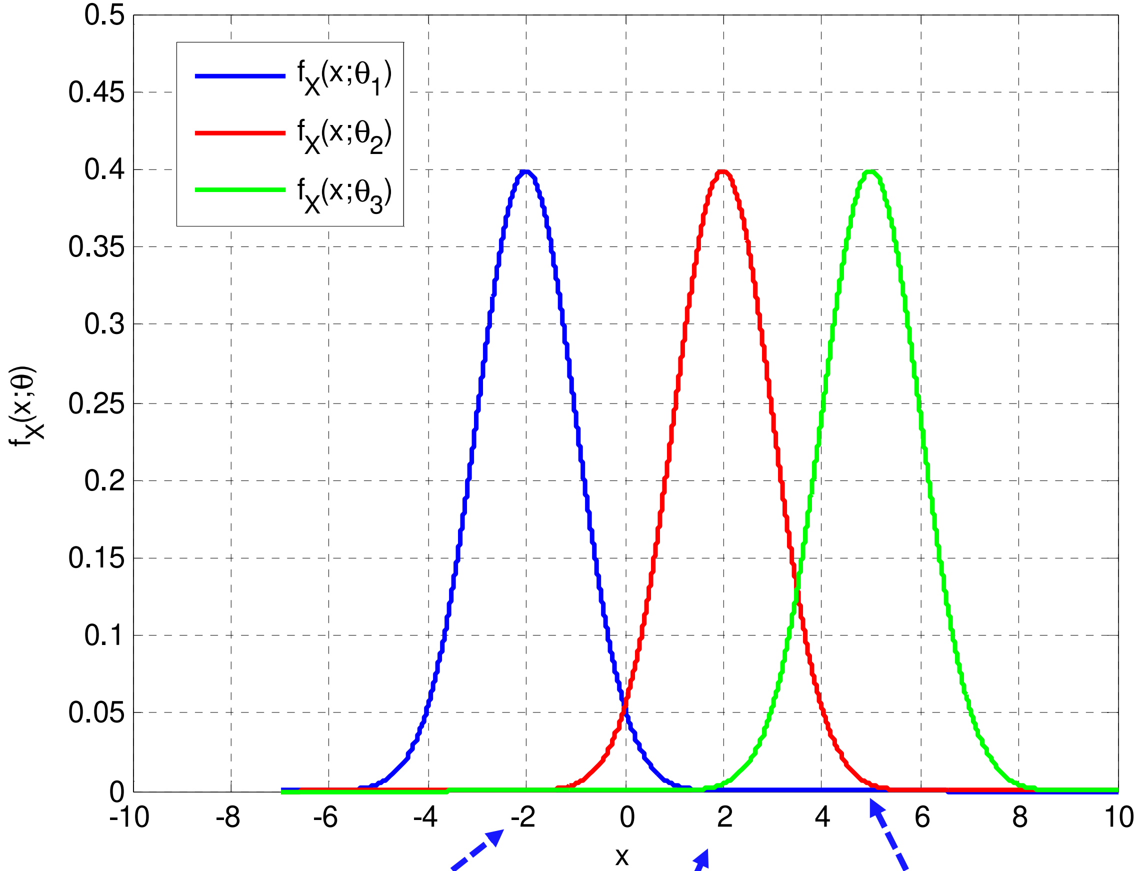
- MM estimates by the method of moments are not necessarily **sufficient statistics**, i.e., they sometimes fail to take into account all relevant information in the sample.

- Let us now introduce the **Maximum Likelihood (ML) method**.

The Maximum Likelihood Estimate

The Maximum Likelihood (ML) Estimate

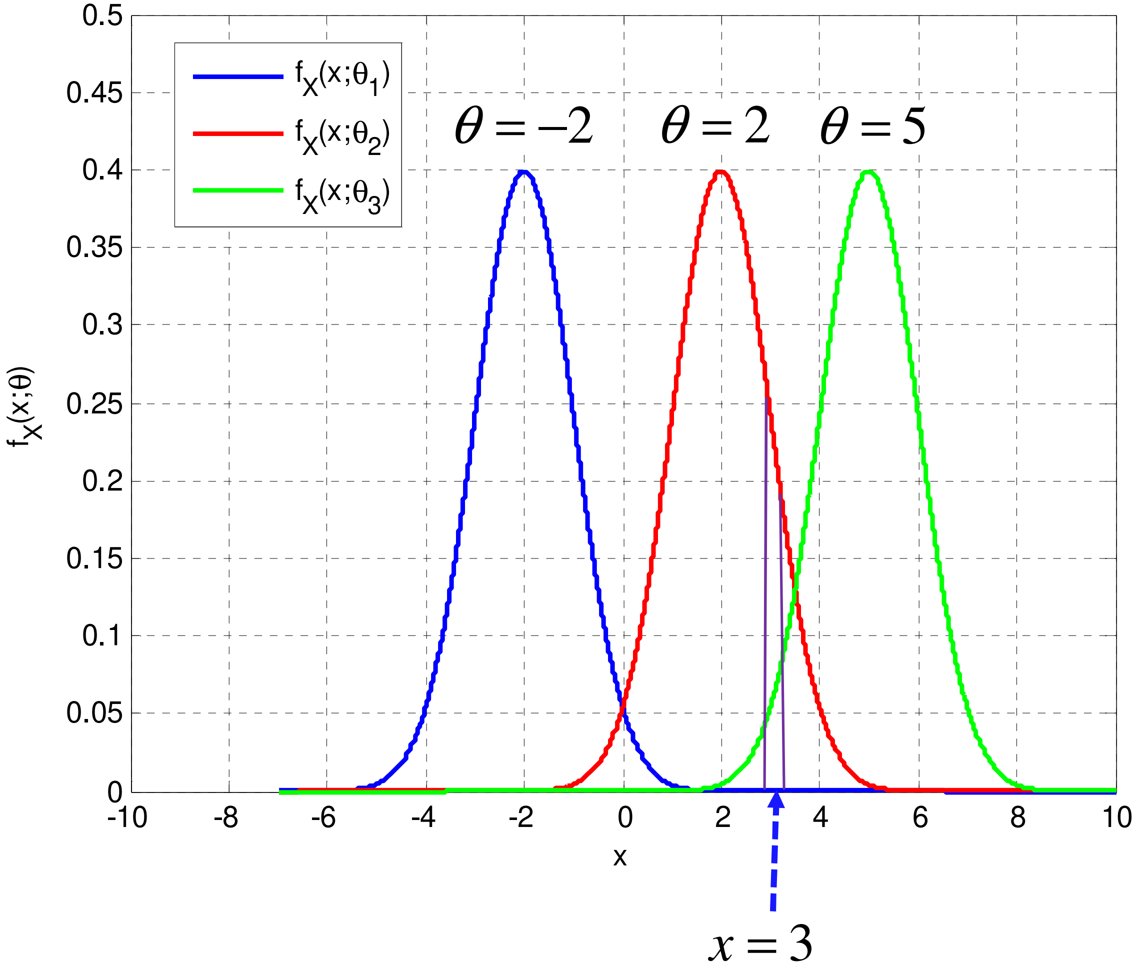
- *Example:* θ is the mean value of X and can assume only three different values.



$$\theta = \theta_1 = -2, \qquad \theta = \theta_2 = 2, \qquad \theta = \theta_3 = 5$$

The Maximum Likelihood (ML) Estimate

- Based on the observation of $X=x$, *e.g.* $x=3$, we want to estimate θ, i.e. decide which one of the three possible values is the true one.



The Maximum Likelihood (ML) Estimate

- Calculate now the probability that the r.v. *X* assumes a value that belongs to the interval centered around the observed value *x* of infinitesimal length Δ: $X \in [x-\Delta/2,\ x+\Delta/2]$, for each of the three values of θ:

$$f_X(x;\theta) \quad \rightarrow \quad P_i \triangleq \int_{x-\Delta/2}^{x+\Delta/2} f_X(\alpha;\theta_i)\,d\alpha = f_X(x;\theta_i)\Delta$$

- The idea of the ML method is to choose the value of θ which maximizes the probability P_i of observing a value of $X \in [x-\Delta/2,\ x+\Delta/2]$:

$$\hat{\theta} = \underset{\theta \in \{\theta_1,\theta_2,\theta_3\}}{\arg\max}\, P_i = \underset{\theta \in \{\theta_1,\theta_2,\theta_3\}}{\arg\max}\, f_X(x;\theta)\Delta = \underset{\theta \in \{\theta_1,\theta_2,\theta_3\}}{\arg\max}\, f_X(x;\theta)$$

- In the previous example, where $x=3$, the ML criterion brings us to select $\theta=2$ as the most likely value, i.e. the value that maximizes the likelihood of observing that specific value of x. In this case the ML estimate of θ would be 2.

The Maximum Likelihood (ML) Estimate

- Hence, the ML estimate is obtained by looking for the value of θ which maximizes the pdf of the observed data, given the observed data *X=x:*

$$\hat{\theta}_{ML} = \arg\max_{\theta} f_X(x;\theta)$$

- In general, the parameter θ is continuous, i.e. $\theta \in \Re$.

- The generalization to the case of *N* observed data is immediate:

$$\hat{\theta}_{ML} = \arg\max_{\theta} f_{\mathbf{X}}(\mathbf{x};\theta)$$

- It is worth observing that we should look at the pdf no more as a function of the data vector **x**, but as a function of the unknown parameter θ.

The Maximum Likelihood (ML) Estimate

- When seen as a function of the unknown parameter θ, the pdf of the observed data is called **Likelihood Function (LF)**.

- Sometimes, we prefer to operate on the natural logarithm of the likelihood function, the **Log-Likelihood Function (LLF)**:

$$L(\theta) \triangleq f_{\mathbf{X}}(\mathbf{x};\theta) \qquad\qquad \ln L(\theta) \triangleq \ln f_{\mathbf{X}}(\mathbf{x};\theta)$$

- The estimator that is obtained by maximizing the likelihood function is called the **Maximum Likelihood (ML) estimator**:

$$\hat{\theta}_{ML} = \arg\max_{\theta} f_{\mathbf{X}}(\mathbf{x};\theta) = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} \ln L(\theta)$$

- Clearly, the estimate is a function of the observed data **x**, even if not explicitly written.

The Maximum Likelihood (ML) Estimate

- In general, any time that the **Likelihood Function (LF)** is continuous and differentiable in the definition interval of the parameter θ to be estimated and the maximum is not on the border of the definition interval, the ML estimate can be obtained by solving the following equation (the **likelihood equation**):

$$\left.\frac{\partial L(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_{ML}} = 0, \qquad \text{or} \qquad \left.\frac{\partial \ln L(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_{ML}} = 0$$

- When we have to estimate **multiple parameters**, e.g. P parameters, we must derive the LF with respect to all the parameters and then set the derivative equal to zero, so obtaining a system of P equations in P unknowns:

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_P \end{bmatrix}^T \qquad \rightarrow \qquad \left.\frac{\partial \ln L(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ML}} = 0, \qquad i = 1, 2, \ , P$$

The Maximum Likelihood (ML) Estimate

- In general, we cannot claim that the ML estimate is the optimal one, i.e. the one with the minimum MSE, since, as we have seen, an optimal estimator does not exist (for deterministic parameters). However, the **ML estimate** enjoys the following **properties**:
- asymptotically unbiased;
- consistent;
- asymptotically efficient (the one with the lowest possible MSE);
- asymptotically Gaussian distributed;
- There is no guarantee that the ML estimator is efficient, but if an efficient estimator exists, it is the ML estimator.
- The concept of **efficiency** is related to the definition of the so-called **Cramér-Rao lower bound** (**CRB** or **CRLB**), that we will introduce in the following.

The Maximum Likelihood (ML) Estimate

- **Example**: Derive the ML estimate of the parameters η and σ^2 of a Gaussian r.v. *X*, given *N* independent realizations of *X*:

$$\underset{N\times 1}{\mathbf{X}} = \begin{bmatrix} X_1 & X_2 & \cdots & X_N \end{bmatrix}^T, \quad \{X_i\}_{i=1}^{N} \quad \text{IID}$$

$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\eta)^2}{2\sigma^2}} \quad \rightarrow \quad m_1 = E\{X_i\} = \eta, \quad \mu_2 = \text{var}\{X_i\} = \sigma^2.$$

$$f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2) = \prod_{i=1}^{N} f_{X_i}(x_i;\eta,\sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\eta)^2}{2\sigma^2}} = \left(2\pi\sigma^2\right)^{-\frac{N}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\eta)^2}$$

$$\ln L(\eta,\sigma^2) = \ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2) = -\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(x_i - \eta\right)^2$$

- **Case study 1**: η unknown and σ^2 known (>0).

$$\frac{d\ln L(\eta,\sigma^2)}{d\eta} = \frac{d\ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{d\eta} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\eta) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{N}x_i - N\eta\right)$$

$$\frac{d\ln L(\eta,\sigma^2)}{d\eta} = 0 \quad \rightarrow \quad \sum_{i=1}^{N}x_i - N\eta = 0 \quad \rightarrow \quad \eta = \frac{1}{N}\sum_{i=1}^{N}x_i$$

$$\hat{\eta}_{ML} = \frac{1}{N}\sum_{i=1}^{N}X_i$$

- The ML estimate of the parameter η (the mean value) of a Gaussian r.v. *X,* when the variance σ^2 is a priori known, coincides with the **Sample Mean**, that is also the estimator we obtained by the **Method of Moments**.

- **Caso study 2**: η known and σ² unknown (>0).

$$\frac{d\ln L(\eta,\sigma^2)}{d\sigma^2}=\frac{d\ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{d\sigma^2}=-\frac{N}{2}\cdot\frac{1}{\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{N}\left(x_i-\eta\right)^2$$

$$\frac{d\ln L(\eta,\sigma^2)}{d\sigma^2}=0 \quad\rightarrow\quad -\frac{N}{2}\cdot\frac{1}{\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{N}\left(x_i-\eta\right)^2=0 \quad\rightarrow\quad \sigma^2=\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\eta\right)^2$$

$$\hat{\sigma}^2_{ML}=\frac{1}{N}\sum_{i=1}^{N}\left(X_i-\eta\right)^2$$

- The ML estimate of the variance σ² when the mean η is a priori known is unbiased:

$$E\left\{\hat{\sigma}^2_{ML}\right\}=E\left\{\frac{1}{N}\sum_{i=1}^{N}\left(X_i-\eta\right)^2\right\}=\frac{1}{N}\sum_{i=1}^{N}E\left\{\left(X_i-\eta\right)^2\right\}=\sigma^2$$

The Maximum Likelihood (ML) Estimate

$$MSE\left\{\hat{\sigma}_{ML}^2\right\} = \mathrm{var}\left\{\hat{\sigma}_{ML}^2\right\} + \left(b\left\{\hat{\sigma}_{ML}^2\right\}\right)^2 = \mathrm{var}\left\{\hat{\sigma}_{ML}^2\right\}$$

$$= \mathrm{var}\left\{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \eta\right)^2\right\} = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{var}\left\{\left(X_i - \eta\right)^2\right\}$$

$$= \frac{1}{N}\mathrm{var}\left\{\left(X_i - \eta\right)^2\right\}$$

$$= \frac{1}{N}\left[E\left\{\left(X_i - \eta\right)^4\right\} - \left(E\left\{\left(X_i - \eta\right)^2\right\}\right)^2\right]$$

$$= \frac{\mu_X(4) - \left(\sigma^2\right)^2}{N} = \frac{3\sigma^4 - \left(\sigma^2\right)^2}{N} = \frac{2\sigma^4}{N}$$

- It is worth observing that this value coincides with the asymptotical (N>>1) value of the MSE of the Sample Variance estimator (obtained under the assumption that also the mean is unknown).

The Maximum Likelihood (ML) Estimate

- **Caso study 3**: η e σ^2 both unknowns ($\sigma^2>0$).

$$\begin{cases} \dfrac{d \ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{d\eta} = \dfrac{1}{\sigma^2}\left(\displaystyle\sum_{i=1}^{N} x_i - N\eta\right) = 0 \\ \dfrac{d \ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{d\sigma^2} = -\dfrac{N}{2}\cdot\dfrac{1}{\sigma^2} + \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{N}\left(x_i-\eta\right)^2 = 0 \end{cases}$$

- We solve the first equation w.r.t. η and we insert it in the second equation:

$$\begin{cases} \eta = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} x_i \\ \sigma^2 = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\left(x_i - \dfrac{1}{N}\displaystyle\sum_{k=1}^{N} x_k\right)^2 \end{cases}$$

The Maximum Likelihood (ML) Estimate

- The joint **ML estimators of mean and variance of a Gaussian distribution** are given by the **Sample Mean** and the **Sample Variance**, previously derived, also via the Method of Moments:

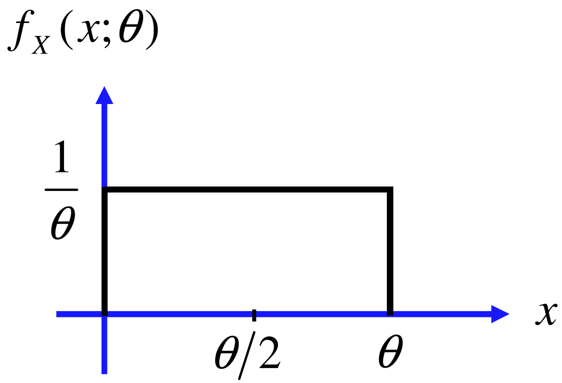
$$\hat{\eta}_{ML} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$\hat{\sigma}^2_{ML} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \frac{1}{N}\sum_{k=1}^{N} X_k \right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{\eta}_{ML} \right)^2$$

$$MSE\left\{\hat{\eta}_{ML}\right\} = \frac{\sigma^2}{N}, \qquad MSE\left\{\hat{\sigma}^2_{ML}\right\} = \frac{2\sigma^4}{N}\left(1 - \frac{1}{2N}\right) \overset{N \gg 1}{\cong} \frac{2\sigma^4}{N}$$

The Maximum Likelihood (ML) Estimate

- ***Example***: We have a random variable *X* having uniform distribution between 0 and θ, where $\theta>0$. We want to derive the ML and MM estimators of the parameter θ, given *N* independent realizations of *X* and then to compare their performance.



$$f_X(x;\theta)=\frac{1}{\theta}rect\left(\frac{x-\theta/2}{\theta}\right)=\begin{cases}1/\theta, & 0\le x\le\theta\\ 0, & otherwise\end{cases}$$

- Let us derive first the **MM estimator**.
- The first step is to calculate the mean of X:

$$m_1=E\{X\}=\frac{\theta}{2}\quad \text{(the mean value is the symmetry point of the pdf)}$$

The Maximum Likelihood (ML) Estimate

- By inverting this relationship we get:

$$\theta = 2m_1 \quad \Rightarrow \quad \hat{\theta}_{MM}(\mathbf{x}) = 2\hat{m}_1 = \frac{2}{N}\sum_{i=1}^{N} X_i$$

- Calculation of the **bias**:

$$E\left\{\hat{\theta}_{MM}(\mathbf{x})\right\} = E\left\{\frac{2}{N}\sum_{i=1}^{N} X_i\right\} = \frac{2}{N}\sum_{i=1}^{N} E\{X_i\} = \frac{2}{N}\sum_{i=1}^{N}\frac{\theta}{2} = \theta$$

- The MM estimator is **unbiased.** Hence, the MSE coincides with the variance:

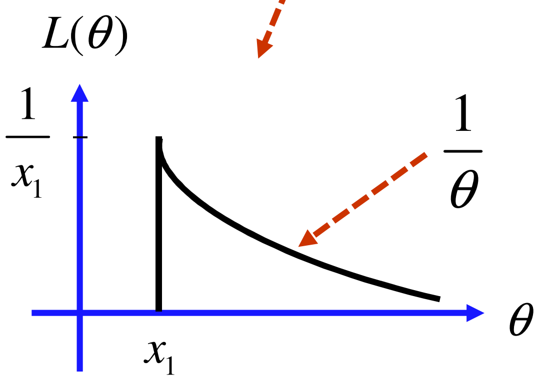
$$MSE\left\{\hat{\theta}_{MM}(\mathbf{x})\right\} = \mathrm{var}\left\{\hat{\theta}_{MM}(\mathbf{x})\right\} = \mathrm{var}\left\{\frac{2}{N}\sum_{i=1}^{N} X_i\right\} = \frac{4}{N^2}\sum_{i=1}^{N}\mathrm{var}\{X_i\}$$

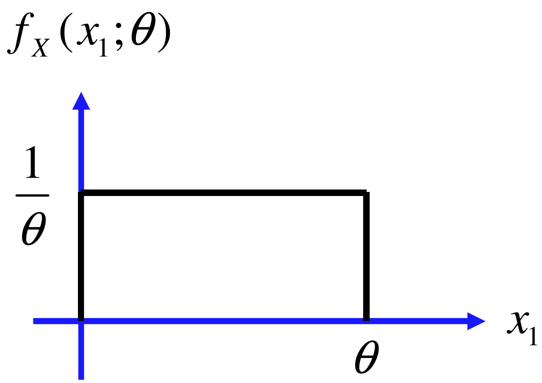
$$= \frac{4}{N^2}\cdot N\cdot\frac{\theta^2}{12} = \frac{\theta^2}{3N}$$

The Maximum Likelihood (ML) Estimate

- Let us now derive the **ML estimator**. The first step is to understand which is the behavior of the likelihood function (LF).
- Let us start from the simplest case of *N*=1.

$$L(\theta) = f_X(x_1;\theta) = \begin{cases} 1/\theta, & 0 \le x_1 \le \theta \\ 0, & otherwise \end{cases}$$





The Maximum Likelihood (ML) Estimate

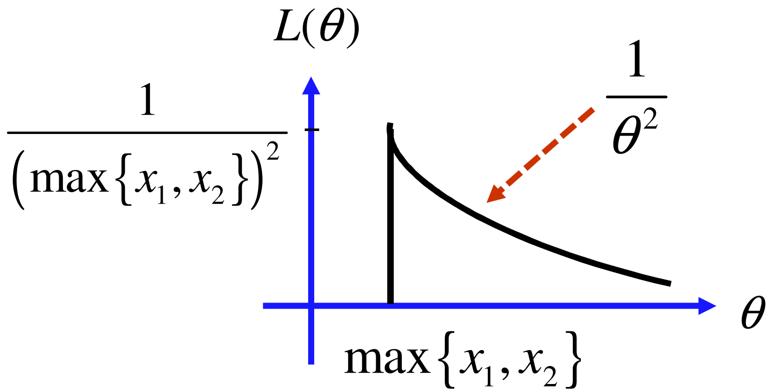
- For *N*=2, exploiting the independence of the observed data:

$$L(\theta) = f_{\mathbf{X}}(\mathbf{x};\theta) = \prod_{i=1}^{N} f_X(x_i;\theta)$$

$$\stackrel{N=2}{=} f_X(x_1;\theta) f_X(x_2;\theta) = \frac{1}{\theta} rect\left(\frac{x_1 - \theta/2}{\theta}\right) \cdot \frac{1}{\theta} rect\left(\frac{x_2 - \theta/2}{\theta}\right)$$

$$= \begin{cases} 1/\theta^2, & 0 \le x_1 \le \theta, 0 \le x_2 \le \theta \\ 0, & otherwise \end{cases}$$

$$= \begin{cases} 1/\theta^2, & 0 \le \max\{x_1, x_2\} \le \theta \\ 0, & otherwise \end{cases}$$



The Maximum Likelihood (ML) Estimate

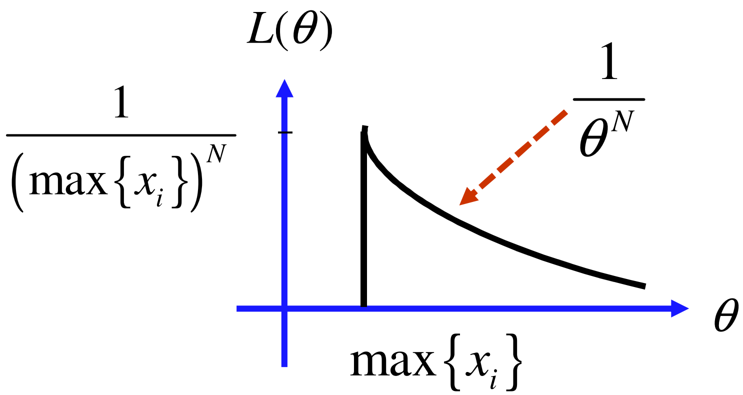
- For a generic *N*:

$$L(\theta) = f_{\mathbf{X}}(\mathbf{x};\theta) = \prod_{i=1}^{N} f_X(x_i;\theta)$$

$$= \prod_{i=1}^{N}\left[\frac{1}{\theta} rect\left(\frac{x_i - \theta/2}{\theta}\right)\right] = \frac{1}{\theta^N}\prod_{i=1}^{N} rect\left(\frac{x_i - \theta/2}{\theta}\right)$$

$$= \begin{cases} 1/\theta^N, & 0 \le x_1 \le \theta, 0 \le x_2 \le \theta, \cdots, 0 \le x_N \le \theta \\ 0, & otherwise \end{cases}$$

$$= \begin{cases} 1/\theta^N, & 0 \le \max\{x_1, x_2, \cdots, x_N\} \le \theta \\ 0, & otherwise \end{cases}$$



$$\max\{x_i\} \triangleq \max\{x_1, x_2, \cdots, x_N\}$$

The Maximum Likelihood (ML) Estimate

- In this case, the likelihood function is not continuous and differentiable everywhere in definition interval of the parameter θ, that is $\theta \in (0,+\infty)$, in particular, as we will see, it is not not continuous and differentiable in the maximum. Hence, the ML estimate cannot be obtained by solving the **likelihood equation**, but it must be obtained by looking directly for the value of θ for which $L(\theta)$ is maximum.

- In our case, by inspection of the previous figures we immediately obtain:

$$\hat{\theta}_{ML} = \arg\max_{\theta} L(\theta) = \max\{x_1, x_2, \cdots, x_N\}$$

- ML estimation requires ordering of the data (***Ordered Statistics, OS***).

- For performance analysis of the ML estimator, we have to derive the **bias** and the **MSE**.

- To this purpose, we need to derive the **pdf** of the estimator and then its **mean** value and **variance**.

The Maximum Likelihood (ML) Estimate

- Let us derive the pdf of the ML estimator by applying the method of the **cumulative distribution function (cdf)**:

$$X_M \triangleq \hat{\theta}_{ML} = \max\{X_1, X_2, \cdots, X_N\}$$

$$f_{X_M}(x_M) = \frac{dF_{X_M}(x_M)}{dx_M}, \quad \text{where } F_{X_M}(x_M) \triangleq \Pr\{X_M \le x_M\}$$

- The maximum is lower than x_M only if all the observed data are lower than x_M:

$$\begin{aligned} F_{X_M}(x_M) &\triangleq \Pr\{X_M \le x_M\} = \Pr\left\{\max\{X_1, X_2, \cdots, X_N\} \le x_M\right\} \\ &= \Pr\{X_1 \le x_M, X_2 \le x_M, \cdots, X_N \le x_M\} \\ &= \Pr\{X_1 \le x_M\}\Pr\{X_2 \le x_M\}\cdots\Pr\{X_N \le x_M\} \end{aligned}$$

- where we exploited the assumption that the observed data are independent.

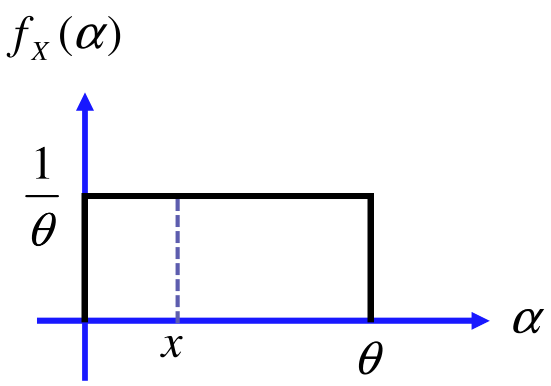
The Maximum Likelihood (ML) Estimate

- If we now exploit also the assumption that the data are identically distributed, i.e. they have the same pdf and so the same cdf:

$$F_{X_M}(x_M) = \Pr\{X_1 \le x_M\}\Pr\{X_2 \le x_M\}\cdots\Pr\{X_N \le x_M\}$$
$$= \prod_{i=1}^{N}\Pr\{X_i \le x_M\} = \prod_{i=1}^{N} F_X(x_M) = \left[F_X(x_M)\right]^N$$

- where the cdf of X is given by:

$$F_X(x) = \int_{-\infty}^{x} f_X(\alpha)d\alpha = \frac{1}{\theta}\int_{-\infty}^{x} rect\left(\frac{\alpha-\theta/2}{\theta}\right)d\alpha$$
$$= \begin{cases} 0, & x < 0 \\ \dfrac{x}{\theta}, & 0 \le x \le \theta \\ 1, & x > \theta \end{cases}$$



The Maximum Likelihood (ML) Estimate

- Hence, the cdf of the ML estimate is given by:

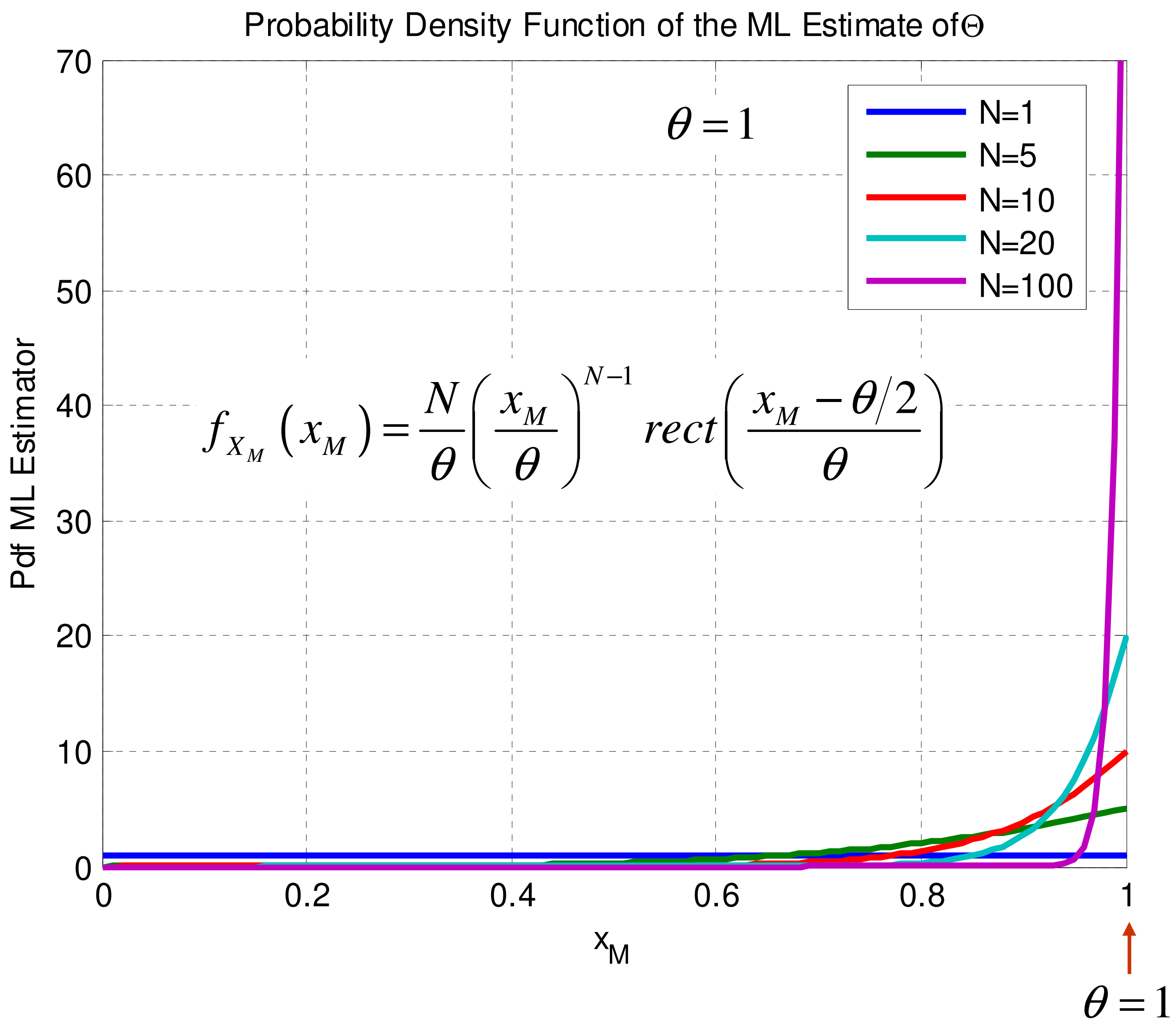
$$F_{X_M}(x_M) = \left[F_X(x_M)\right]^N = \begin{cases} 0, & x_M < 0 \\ \left(\dfrac{x_M}{\theta}\right)^N, & 0 \le x_M \le \theta \\ 1, & x_M > \theta \end{cases}$$

- Finally, the pdf of the ML estimate is obtained by calculating the derivative:

$$f_{X_M}(x_M) = \frac{dF_{X_M}(x_M)}{dx_M} = \frac{d\left[F_X(x_M)\right]^N}{dx_M} = N\left[F_X(x_M)\right]^{N-1} f_X(x_M)$$

$$= \begin{cases} \dfrac{N}{\theta}\left(\dfrac{x_M}{\theta}\right)^{N-1}, & 0 \le x_M \le \theta \\ 0 & otherwise \end{cases} = \frac{N}{\theta}\left(\frac{x_M}{\theta}\right)^{N-1} rect\left(\frac{x_M - \theta/2}{\theta}\right)$$

The Maximum Likelihood (ML) Estimate



The Maximum Likelihood (ML) Estimate

- We can now calculate **mean** and **variance** of the ML estimate:

$$\text{pdf of } \hat{\theta}_{ML}: \quad f_{X_M}(x_M) = \frac{N}{\theta}\left(\frac{x_M}{\theta}\right)^{N-1} rect\left(\frac{x_M - \theta/2}{\theta}\right)$$

$$E\left\{\hat{\theta}_{ML}\right\} = \int_{-\infty}^{+\infty} x_M f_{X_M}(x_M)dx_M = \frac{N}{\theta^N}\int_0^{\theta}(x_M)^N dx_M = \frac{N}{\theta^N}\left[\frac{(x_M)^{N+1}}{N+1}\right]\Bigg|_{x_M=0}^{\theta} = \frac{N\theta}{N+1}$$

$$E\left\{\hat{\theta}_{ML}^2\right\} = \int_{-\infty}^{+\infty} x_M^2 f_{X_M}(x_M)dx_M = \frac{N}{\theta^N}\int_0^{\theta}(x_M)^{N+1} dx_M = \frac{N}{\theta^N}\left[\frac{(x_M)^{N+2}}{N+2}\right]\Bigg|_{x_M=0}^{\theta} = \frac{N\theta^2}{N+2}$$

$$\text{var}\left\{\hat{\theta}_{ML}\right\} = E\left\{\hat{\theta}_{ML}^2\right\} - \left(E\left\{\hat{\theta}_{ML}\right\}\right)^2 = \frac{N\theta^2}{N+2} - \left(\frac{N\theta}{N+1}\right)^2 = \frac{N\theta^2}{(N+1)^2(N+2)}$$

The Maximum Likelihood (ML) Estimate

- Finally, we derive **bias** and **MSE** of the ML estimate:

$$b\left\{\hat{\theta}_{ML}\right\} = E\left\{\theta - \hat{\theta}_{ML}\right\} = \theta - E\left\{\hat{\theta}_{ML}\right\} = \theta - \frac{N\theta}{N+1} = \frac{\theta}{N+1} \overset{N\to\infty}{\longrightarrow} 0$$

- The estimator is **biased,** but **asymptotically unbiased**.

$$\begin{aligned} MSE\left\{\hat{\theta}_{ML}(\mathbf{x})\right\} &= \left(b\left\{\hat{\theta}_{ML}\right\}\right)^2 + \mathrm{var}\left\{\hat{\theta}_{ML}\right\} \\ &= \left(\frac{\theta}{N+1}\right)^2 + \frac{N\theta^2}{(N+1)^2(N+2)} \\ &= \frac{2\theta^2}{(N+1)(N+2)} \overset{N\to\infty}{\longrightarrow} 0 \end{aligned}$$

- The estimator is **consistent**.

The Maximum Likelihood (ML) Estimate

- **Performance comparison**:

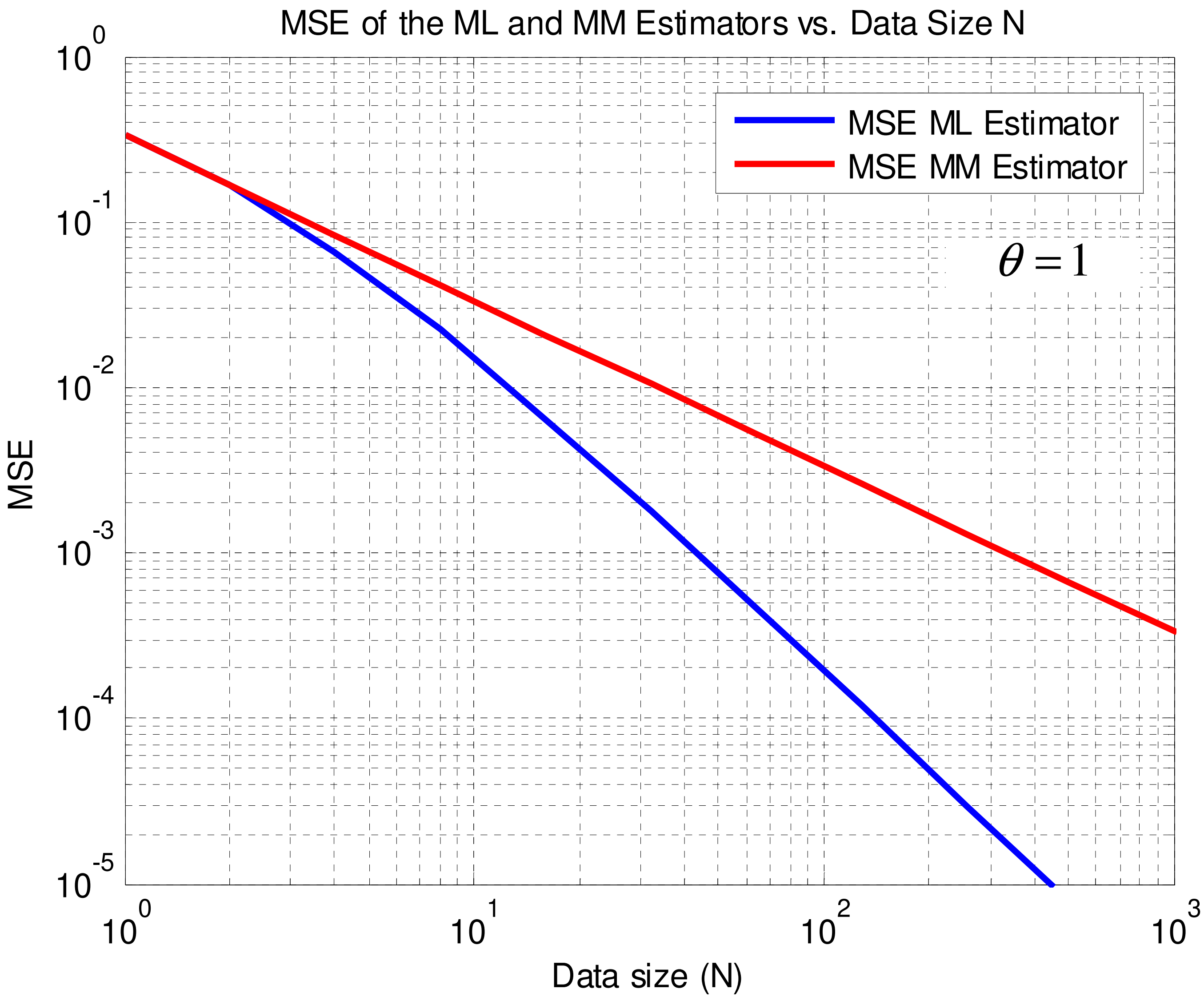
- The MSE of the ML estimator goes to zero with the law $1/N^2$. Instead, the MSE of the MM estimator goes to zero with the law $1/N$. Hence, the ML estimator is much more efficient.

- The two MSE's coincide only for $N=1$.

$$MSE\left\{\hat{\theta}_{ML}(\mathbf{x})\right\} = \frac{2\theta^2}{(N+1)(N+2)} \overset{N \gg 1}{\cong} \frac{2\theta^2}{N^2}$$

$$MSE\left\{\hat{\theta}_{MM}(\mathbf{x})\right\} = \frac{\theta^2}{3N}$$

The Maximum Likelihood (ML) Estimate



The Maximum Likelihood (ML) Estimate

- **Comparison in terms of computational complexity**:

$$\hat{\theta}_{MM}(\mathbf{x}) = \frac{2}{N}\sum_{i=1}^{N} X_i, \qquad \hat{\theta}_{ML}(\mathbf{x}) = \max\{X_i\}$$

- The ML estimator is non-linear, it requires calculation of the maximum of a set of N data (data ordering).

- The MM estimator is linear, hence much more efficient from the point of view of the **computational complexity**.

- The choice between the two depends on the constraints imposed by the specific application: if estimation accuracy is the most important issue, then we should use the ML estimate; if instead it is more important to keep low the computational complexity, we should use the MM estimator.

The Maximum Likelihood (ML) Estimate

- ***Example***: In a digital communication channel the probability that a symbol is received correctly is p. Assuming that we can observe the result of N independent trials of the experiment "transmission of a symbol" and check if that symbol has been received correctly or not for any trial, find the **ML estimator** of p.

- Thanks to the fact that the trials are run independently and in the same condition, the random variable K={*Number of times the symbol has been correctly received*} is a discrete Binomial random variable with parameters N and p:

$$K \in B(p, N) \quad \text{"Number of successes"}$$

$$p_K(k) = \Pr\{K = k\} = \binom{N}{k} p^k (1-p)^{N-k}, \quad 0 < p < 1, \quad 0 \le k \le N$$

The Maximum Likelihood (ML) Estimate

- Where:

$$\binom{N}{k} \triangleq \frac{N!}{(N-k)!k!}, \qquad E\{K\} = Np, \quad \operatorname{var}\{K\} = Np(1-p)$$

- To derive the ML estimate of *p* we have to find the value of *p* which maximizes the likelihood function.
- For a discrete random variable, the likelihood function is the joint **mass probability function (mpf)**, seen as a function of the unknown parameter, in this case *p*:

$$L(p) = p_K(k) = \binom{N}{k} p^k (1-p)^{N-k}, \quad 0 < p < 1, \quad 0 \le k \le N$$

The Maximum Likelihood (ML) Estimate

- Since $L(p)$ is continuous and differentiable w.r.t. p, we derive the ML estimate by solving the **likelihood function**:

$$\left.\frac{dL(p)}{dp}\right|_{p=\hat{p}_{ML}} = 0$$

$$\begin{aligned}
\frac{dL(p)}{dp} &= \frac{d}{dp}\left\{\binom{N}{k} p^{k}\left(1-p\right)^{N-k}\right\} \\
&= \binom{N}{k} k p^{k-1}\left(1-p\right)^{N-k} - \binom{N}{k}(N-k)p^{k}\left(1-p\right)^{N-k-1} \\
&= \binom{N}{k} p^{k-1}\left(1-p\right)^{N-k-1}\left[k(1-p)-(N-k)p\right]
\end{aligned}$$

The Maximum Likelihood (ML) Estimate

- Since the definition interval for p is $0<p<1$, we can exclude the trivial solutions $p=0$ and $p=1$, so we have:

$$k(1-p)-(N-k)p = k - Np = 0 \quad \Rightarrow \quad p = \frac{k}{N}$$

$$\hat{p}_{ML} = \frac{k}{N}$$

- Hence, the estimator "*frequency of occurrence*" that we already derived (see the *Law of Large Numbers*) is also the ML estimator. As a consequence, it has all the good properties of the ML estimators.

The Maximum Likelihood (ML) Estimate

- **Bias** and **MSE**:

$$E\{\hat{p}_{ML}\} = E\left\{\frac{K}{N}\right\} = \frac{E\{K\}}{N} = \frac{Np}{N} = p \quad \rightarrow \quad b\{\hat{p}_{ML}\} = 0$$

- The estimator is **unbiased**.

$$\mathrm{var}\{\hat{p}_{ML}\} = \mathrm{var}\left\{\frac{K}{N}\right\} = \frac{\mathrm{var}\{K\}}{N^2} = \frac{Np(1-p)}{N^2} = \frac{p(1-p)}{N}$$

$$MSE\{\hat{p}_{ML}\} = \mathrm{var}\{\hat{p}_{ML}\} = \frac{p(1-p)}{N} \quad \Rightarrow \quad \lim_{N\to\infty} MSE\{\hat{p}_{ML}\} = 0$$

- The estimator is **consistent**.

The Cramér-Rao Lower Bound

The Cramér-Rao Lower Bound (CRB)

- One of the most important results in estimation theory establishes that the **Mean Square Error (MSE)** of any **unbiased** estimator is always greater or equal to a value that depends on the data pdf and on how this pdf depends on the unknown parameter θ:

$$\text{If } \; E\left\{\hat{\theta}(\mathbf{x})\right\} = \theta \quad i.e. \text{ the estimator is unbiased}$$

$$MSE\left\{\hat{\theta}(\mathbf{x})\right\} = \operatorname{var}\left\{\hat{\theta}(\mathbf{x})\right\} \geq CRB(\theta) = \frac{1}{-E\left\{\dfrac{\partial^2 \ln f_{\mathbf{x}}(\mathbf{x};\theta)}{\partial \theta^2}\right\}}$$

- This result is known as the **Cramér-Rao lower bound (CRB)** or the **information inequality**, even if Fisher had already derived it 23 years before Cramér and Rao. The denominator of the CRB is called **Fisher Information**.

The Cramér-Rao Lower Bound (CRB)

$$MSE\left\{\hat{\theta}(\mathbf{x})\right\} = \mathrm{var}\left\{\hat{\theta}(\mathbf{x})\right\} \geq CRB(\theta) = \frac{1}{-E\left\{\dfrac{\partial^2 \ln f_{\mathbf{x}}(\mathbf{x};\theta)}{\partial \theta^2}\right\}}$$

- This results is valid only if the 2nd order derivative of the log-likelihood function exists and is absolutely integrable.
- There is also a CRB for biased estimators, but it is much less useful, since it depends on the bias expression of the particular estimator.
- An unbiased estimator which achieves this lower bound is said to be **efficient**. Such a solution achieves the lowest possible MSE among all unbiased methods, and it is therefore the **Minimum Variance Unbiased (MVU)** estimator.
- However, in some cases, no unbiased estimator exists which achieves the bound, i.e. an efficient estimator does not exist.

The Cramér-Rao Lower Bound (CRB)

- For an **efficient**, i.e. **Minimum Variance Unbiased (MVU)** estimator:

$$MSE\left\{\hat{\theta}(\mathbf{x})\right\} = \mathrm{var}\left\{\hat{\theta}(\mathbf{x})\right\} = CRB(\theta) = \frac{1}{-E\left\{\dfrac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\theta)}{\partial \theta^2}\right\}}$$

- It can be proved that an efficient unbiased estimator exists **if and only if** the derivative of the log-likelihood function can be factored as follows:

$$\frac{\partial \ln f_{\mathbf{X}}(\mathbf{x};\theta)}{\partial \theta} = I(\theta)\left[g(\mathbf{x}) - \theta\right] \qquad \rightarrow \quad \hat{\theta}(\mathbf{x}) = g(\mathbf{x}) \text{ is efficent}$$

- where $I(\theta)$ is the **Fisher information**, a quantity that can depend on the unknown parameter θ but not on the observed data vector **x**, whereas g(**x**) is a function of **x** but not of θ. g(**x**) is an unbiased efficient estimator of θ.

The Cramér-Rao Lower Bound (CRB)

- The Fisher information $I(\theta)$ is the reciprocal of the CRB(θ), in fact:

$$\frac{1}{CRB(\theta)} = -E\left\{\frac{\partial^2 \ln f_{\mathbf{x}}(\mathbf{x};\theta)}{\partial \theta^2}\right\} = -E\left\{\frac{\partial}{\partial\theta}\left(\frac{\partial \ln f_{\mathbf{x}}(\mathbf{x};\theta)}{\partial\theta}\right)\right\}$$

$$= -E\left\{\frac{\partial}{\partial\theta}\left(I(\theta)[g(\mathbf{x})-\theta]\right)\right\}$$

$$= -\frac{\partial I(\theta)}{\partial\theta}E\{[g(\mathbf{x})-\theta]\} - E\left\{I(\theta)\frac{\partial[g(\mathbf{x})-\theta]}{\partial\theta}\right\}$$

$$= -I(\theta)E\left\{\frac{\partial[g(\mathbf{x})-\theta]}{\partial\theta}\right\} = I(\theta), \qquad \text{since} \qquad E\{g(\mathbf{x})-\theta\}=0$$

$$\rightarrow \quad MSE\{\hat{\theta}(\mathbf{x})\} = MSE\{g(\mathbf{x})\} = CRB(\theta) = \frac{1}{I(\theta)}$$

The Cramér-Rao Lower Bound (CRB)

- From the relationship: $\dfrac{\partial \ln f_{\mathbf{x}}(\mathbf{x};\theta)}{\partial \theta} = I(\theta)[g(\mathbf{x})-\theta]$

we also obtain that <u>if an efficient unbiased estimator exists, this estimator is the ML estimator</u>:

$$\left.\frac{\partial \ln f_{\mathbf{x}}(\mathbf{x};\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_{ML}} = 0 \qquad \Rightarrow \qquad I(\theta)[g(\mathbf{x})-\theta]\big|_{\theta=\hat{\theta}_{ML}} = 0$$

$$\Rightarrow \qquad [g(\mathbf{x})-\theta]\big|_{\theta=\hat{\theta}_{ML}} = 0 \qquad \Rightarrow \qquad \hat{\theta}_{ML} = g(\mathbf{x})$$

- In summary, if the derivative of the log-likelihood function can be factored as above, then $g(\mathbf{x})$ is the ML estimator of θ, it is unbiased and efficient, and its MSE is given by the CRB, that is the reciprocal if the Fisher information:

$$MSE\left\{\hat{\theta}(\mathbf{x})\right\} = MSE\{g(\mathbf{x})\} = CRB(\theta) = \frac{1}{I(\theta)}$$

The Cramér-Rao Lower Bound (CRB)

- It can be proved that if the derivative of the log-likelihood function exists and is absolutely integrable, the ML estimator is **consistent**, **asymptotically Gaussian** and **asymptotically efficient**.

- **Example**: Estimate the parameters η e σ^2 of a Gaussian pdf, given *N* independent realizations of the r.v. *X*, assumed to be Gaussian distributed.

- **Caso study 1**: η unknown and σ^2 known (>0).

$$\frac{d \ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{d\eta} = \frac{1}{\sigma^2}\left(\sum_{i=1}^{N} x_i - N\eta\right) = \underbrace{\frac{N}{\sigma^2}}_{I(\eta)}\left(\underbrace{\frac{1}{N}\sum_{i=1}^{N} x_i}_{g(\mathbf{x})} - \eta\right)$$

The Cramér-Rao Lower Bound (CRB)

$$\frac{d \ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{d\eta} = \frac{1}{\sigma^2}\left(\sum_{i=1}^{N} x_i - N\eta\right) = \underbrace{\frac{N}{\sigma^2}}_{I(\eta)}\left(\underbrace{\frac{1}{N}\sum_{i=1}^{N} x_i}_{g(\mathbf{x})} - \eta\right)$$

$$\hat{\eta}_{ML} = g(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$MSE\left\{\hat{\eta}_{ML}\right\} = \mathrm{var}\left\{\hat{\eta}_{ML}\right\} = CRB(\eta) = \frac{1}{I(\eta)} = \frac{\sigma^2}{N}$$

- The sample mean is the ML estimator of the mean of a Gaussian r.v., it is unbiased and efficient (remember that it is not the MMSE estimator).

The Cramér-Rao Lower Bound (CRB)

- **Caso study 2**: η known and σ^2 unknown (>0).

$$\frac{d\ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{d\sigma^2} = -\frac{N}{2}\cdot\frac{1}{\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{N}(x_i-\eta)^2$$

$$= \underbrace{\frac{N}{2\sigma^4}}_{I(\sigma^2)}\left(\underbrace{\frac{1}{N}\sum_{i=1}^{N}(x_i-\eta)^2}_{g(\mathbf{x})}-\sigma^2\right)$$

$$\hat{\sigma}^2_{ML} = g(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}(X_i-\eta)^2$$

$$MSE\left\{\hat{\sigma}^2_{ML}\right\} = \mathrm{var}\left\{\hat{\sigma}^2_{ML}\right\} = CRB\left(\sigma^2\right) = \frac{1}{I\left(\sigma^2\right)} = \frac{2\sigma^4}{N}$$

The Vector Cramér-Rao Lower Bound (CRB)

- The CRB can also be defined in the case of multiple unknown parameters.
- In this case the CRB is expressed in vector form:

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_P \end{bmatrix}^T, \quad \hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \cdots & \hat{\theta}_P \end{bmatrix}^T$$

For any unbiased estimator of $\hat{\boldsymbol{\theta}}$, i.e. such that $E\left\{\hat{\boldsymbol{\theta}}\right\} = \boldsymbol{\theta}$, we have:

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = E\left\{\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^T\right\} \geq CRB\left(\boldsymbol{\theta}\right) = \mathbf{I}^{-1}\left(\boldsymbol{\theta}\right)$$

$$\left[\mathbf{C}_{\hat{\boldsymbol{\theta}}}\right]_{i,j} = E\left\{\left(\hat{\theta}_i - \theta_i\right)\left(\hat{\theta}_j - \theta_j\right)\right\} = \operatorname{cov}\left\{\hat{\theta}_i, \hat{\theta}_j\right\}$$

- $\mathbf{I}(\theta)$ is the *PxP* **Fisher Information Matrix (FIM)**. It is a positive semidefinite simmetric matrix.

The Vector Cramér-Rao Lower Bound (CRB)

- The **Fisher Information Matrix (FIM)** is defined as follows:

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{i,j} \triangleq -E\left\{\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right\}, \qquad i, j = 1, 2, \cdots, P$$

- The previous matrix inequality should be interpreted as follows:

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \geq \mathbf{I}^{-1}(\boldsymbol{\theta}) \qquad \rightarrow \qquad \mathbf{C}_{\hat{\boldsymbol{\theta}}} - \mathbf{I}^{-1}(\boldsymbol{\theta}) \text{ is a semidefinite positive matrix}$$

- As concerning the MSE on the various unknown parameters, it holds true that:

$$MSE\left\{\hat{\theta}_i\right\} = \mathrm{var}\left\{\hat{\theta}_i\right\} \geq CRB(\theta_i) = \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{i,i}, \quad i = 1, 2, \cdots, P$$

The Cramér-Rao Lower Bound (CRB)

- In this case, an efficient unbiased estimator of the parameter vector θ exists **if and only if** the gradient of the log-likelihood function can be factored as follows:

$$\nabla_{\boldsymbol{\theta}} \ln L(\boldsymbol{\theta}) = \mathbf{I}(\boldsymbol{\theta})\left[\mathbf{g}(\mathbf{x}) - \boldsymbol{\theta}\right]$$

$$\begin{cases} \hat{\boldsymbol{\theta}}_{ML} = \mathbf{g}(\mathbf{x}) \\ \mathbf{I}(\boldsymbol{\theta}) = FIM \end{cases}$$

$$E\left\{\hat{\boldsymbol{\theta}}_{ML}\right\} = \boldsymbol{\theta}, \quad MSE\left\{\hat{\boldsymbol{\theta}}_{ML}\right\} = CRB(\boldsymbol{\theta}) = \mathbf{I}^{-1}(\boldsymbol{\theta})$$

- $\mathbf{I}(\theta)$ is the **Fisher information matrix (FIM)**, which can depend on the unknown parameter vector θ but not on the observed data vector **x**, whereas **g**(**x**) is a function of **x** but not of θ.

- **g**(**x**) is the ML estimator of θ, unbiased, consistent and efficient.

The Vector Cramér-Rao Lower Bound (CRB)

- **Example**: Estimate the two parameters η e σ^2 of a Gaussian pdf, given *N* independent realizations of the r.v. *X*, assumed to be Gaussian distributed.

- **Caso study 3**: η and σ^2 both unknown (σ^2 >0): $\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} \eta \\ \sigma^2 \end{bmatrix}$

- We had already derived: $\begin{cases} \dfrac{d \ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{d\eta} = \dfrac{1}{\sigma^2}\left(\displaystyle\sum_{i=1}^{N} x_i - N\eta\right) \\ \dfrac{d \ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{d\sigma^2} = -\dfrac{N}{2}\cdot\dfrac{1}{\sigma^2} + \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{N}(x_i-\eta)^2 \end{cases}$

$$\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{\partial\eta^2} = \frac{\partial}{\partial\eta}\left[\frac{1}{\sigma^2}\left(\sum_{i=1}^{N} x_i - N\eta\right)\right] = -\frac{N}{\sigma^2}$$

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{1,1} = -E\left\{\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_1^2}\right\} = -E\left\{-\frac{N}{\sigma^2}\right\} = \frac{N}{\sigma^2}$$

$$\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{\left(\partial\sigma^2\right)^2} = \frac{\partial}{\partial\sigma^2}\left[-\frac{N}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}(x_i-\eta)^2\right]$$

$$= \frac{N}{2}\cdot\frac{1}{\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{N}(x_i-\eta)^2$$

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{2,2} = -E\left\{\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_2^2}\right\} = -E\left\{\frac{N}{2}\cdot\frac{1}{\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{N}(X_i-\eta)^2\right\}$$

$$= -\frac{N}{2\sigma^4} + \frac{N\sigma^2}{\sigma^6} = \frac{N}{2\sigma^4}$$

The Vector Cramér-Rao Lower Bound (CRB)

$$\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{\partial\eta\partial\sigma^2} = \frac{\partial}{\partial\sigma^2}\left[\frac{1}{\sigma^2}\left(\sum_{i=1}^{N} x_i - N\eta\right)\right] = -\frac{1}{\sigma^4}\left(\sum_{i=1}^{N} x_i - N\eta\right)$$

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{1,2} = -E\left\{\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2}\right\} = -E\left\{-\frac{1}{\sigma^4}\left(\sum_{i=1}^{N} X_i - N\eta\right)\right\}$$

$$= \frac{N}{\sigma^4} E\left\{\frac{1}{N}\sum_{i=1}^{N} X_i - \eta\right\} = 0$$

$$\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\eta,\sigma^2)}{\partial\sigma^2\partial\eta} = \frac{\partial}{\partial\eta}\left[-\frac{N}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}(x_i-\eta)^2\right] = -\frac{1}{\sigma^4}\sum_{i=1}^{N}(x_i-\eta)$$

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{2,1} = -E\left\{\frac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_2\partial\theta_1}\right\} = -E\left\{-\frac{1}{\sigma^4}\sum_{i=1}^{N}(X_i-\eta)\right\} = 0$$

The Vector Cramér-Rao Lower Bound (CRB)

- Note that $\left[\mathbf{I}(\boldsymbol{\theta})\right]_{1,2} = \left[\mathbf{I}(\boldsymbol{\theta})\right]_{2,1}$,
- In fact the FIM is always simmetric: $\left[\mathbf{I}(\boldsymbol{\theta})\right]_{i,j} = \left[\mathbf{I}(\boldsymbol{\theta})\right]_{j,i} \quad \rightarrow \quad \mathbf{I}(\boldsymbol{\theta}) = \mathbf{I}^T(\boldsymbol{\theta})$
- In this example we have:

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{N}{\sigma^2} & 0 \\ 0 & \dfrac{N}{2\sigma^4} \end{bmatrix} \quad \Rightarrow \quad CRB(\boldsymbol{\theta}) = \mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{\sigma^2}{N} & 0 \\ 0 & \dfrac{2\sigma^4}{N} \end{bmatrix}$$

$$CRB(\eta) = \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{1,1} = \frac{\sigma^2}{N}, \quad CRB\left(\sigma^2\right) = \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{2,2} = \frac{2\sigma^4}{N}$$

- The fact that the FIM is diagonal means that the CRB in the vector case (both the parameters are unknown) is the same as the CRB in scalar case (only one parameter is unknown).

- In general, the FIM is not diagonal, e.g. in the case of two unknown parameters we have:

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} -E\left\{\dfrac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_1^2}\right\} & -E\left\{\dfrac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2}\right\} \\ -E\left\{\dfrac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_2\partial\theta_1}\right\} & -E\left\{\dfrac{\partial^2 \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_2^2}\right\} \end{bmatrix} = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}$$

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbf{I}^T(\boldsymbol{\theta}), \quad \text{because} \quad I_{21} = I_{12}$$

The Vector Cramér-Rao Lower Bound (CRB)

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}^{-1} = \frac{1}{\Delta}\begin{bmatrix} I_{22} & -I_{12} \\ -I_{21} & I_{11} \end{bmatrix}$$

where $\Delta = I_{11}I_{22} - I_{21}I_{12}$ is the determinant of $\mathbf{I}(\boldsymbol{\theta})$

$$CRB(\theta_1) = \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{1,1} = \frac{I_{22}}{I_{11}I_{22} - I_{21}I_{12}} = \frac{1}{I_{11} - \frac{(I_{21})^2}{I_{22}}} \geq \frac{1}{I_{11}}$$

$$CRB(\theta_2) = \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{2,2} = \frac{I_{11}}{I_{11}I_{22} - I_{21}I_{12}} = \frac{1}{I_{22} - \frac{(I_{21})^2}{I_{11}}} \geq \frac{1}{I_{22}}$$

Equality holds iff $I_{21} = I_{12} = 0$

The Vector Cramér-Rao Lower Bound (CRB)

- In general, it holds true that:

$$\left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{i,i} \geq \frac{1}{\left[\mathbf{I}(\boldsymbol{\theta})\right]_{i,i}}$$

$\left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{i,i}$ is the CRB on θ_i in the vector case (all parameters unknown)

$\dfrac{1}{\left[\mathbf{I}(\boldsymbol{\theta})\right]_{i,i}}$ is the CRB on θ_i in the scalar case (only θ_i unknown)

- which means that the asymptotic performance of the ML estimator of a certain parameter, e.g. θ_i, gets worse when the data model depends on additional unknown parameters that are mutually coupled with it.

The Vector Cramér-Rao Lower Bound (CRB)

- This result cannot be extended to biased estimators:

$$\eta \text{ known} \quad \rightarrow \quad \hat{\sigma}^2_{ML} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \eta\right)^2 \quad \rightarrow \quad MSE\left\{\hat{\sigma}^2_{ML}\right\} = CRB(\sigma^2) = \frac{2\sigma^4}{N}$$

$$\eta \text{ unknown} \quad \rightarrow \quad \hat{\sigma}^2_{ML} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{\eta}_{ML}\right)^2 \quad \text{where} \quad \hat{\eta}_{ML} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$\text{In the 2nd case } \hat{\sigma}^2_{ML} \text{ is biased and } \quad MSE\left\{\hat{\sigma}^2_{ML}\right\} = \frac{2\sigma^4}{N}\left(1 - \frac{1}{2N}\right) < \frac{2\sigma^4}{N} = CRB(\sigma^2)$$

- The estimator is biased, so its MSE does not have to be greater than the CRB. Actually, the MSE is lower than the CRB! In this (rare) case the estimator is called "**super-efficient**".

The Vector Cramér-Rao Lower Bound (CRB)

- *Summary*: Estimation of the **variance** of a Gaussian distribution based on *N* IID data, when the mean is also unknown:

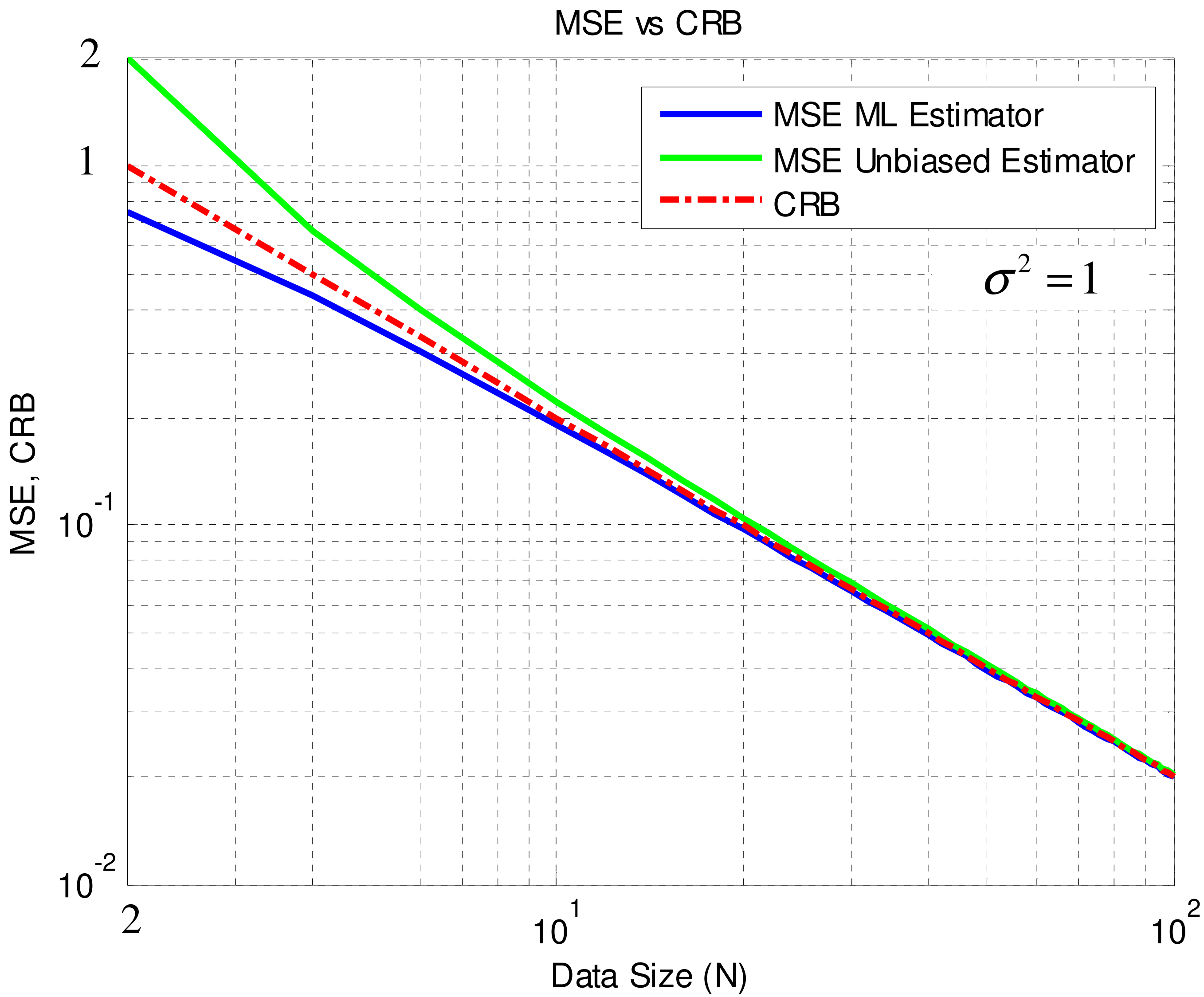
$$\hat{\eta}_{ML} = \frac{1}{N}\sum_{i=1}^{N} X_i, \qquad \hat{\sigma}^2_{ML} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{\eta}_{ML}\right)^2, \qquad \hat{\sigma}^2_{UB} = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \hat{\eta}_{ML}\right)^2$$

$$CRB\left(\sigma^2\right) = \frac{2\sigma^4}{N}$$

$$MSE\left\{\hat{\sigma}^2_{ML}\right\} = \frac{2\sigma^4}{N}\left(1 - \frac{1}{2N}\right) < CRB\left(\sigma^2\right) \quad \rightarrow \quad \text{super-efficient}$$

$$MSE\left\{\hat{\sigma}^2_{UB}\right\} = \left(\frac{N}{N-1}\right)^2 \operatorname{var}\left\{\hat{\sigma}^2_{ML}\right\} = \frac{2}{N-1}\sigma^4 = \frac{2\sigma^4}{N}\left(1 + \frac{1}{N-1}\right) > CRB\left(\sigma^2\right)$$

The Vector Cramér-Rao Lower Bound (CRB)



ML Estimate of the Parameters of a Signal: Amplitude

Estimate of the Amplitude of a Signal

- **Example**: We observe over the time interval $[0,T)$ a signal of known shape but unknown amplitude A, embedded in Additive White Gaussian Noise (AWGN):

$$X(t) = A \cdot s(t) + W(t), \quad t \in [0,T)$$

$$E_s \triangleq \int_0^T s^2(t)dt < +\infty$$

$$W(t)\ AWGN: \quad S_W(f) = \frac{N_0}{2}$$

- Without loss of generality, let us assume that $X(t)$, $s(t)$, $W(t)$ and A are real-valued.

Baseband rectangular pulse

Narrowband sinusoidal pulse

$$\text{Ex. 1}: \quad s(t) = rect\left(\frac{t - T/2}{T}\right)$$

$$\text{Ex. 2}: \quad s(t) = \cos(2\pi f_0 t + \theta)\, rect\left(\frac{t - T/2}{T}\right), \quad \text{with } f_0 \text{ and } \theta \text{ known}, \quad f_0 T \gg 1$$

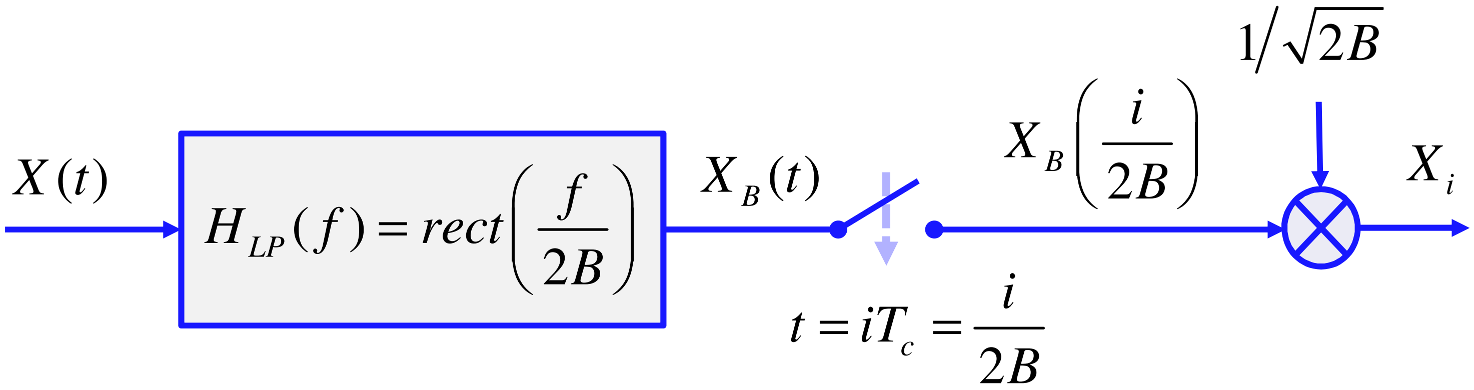
Estimate of the Amplitude of a Signal

- The first step is to transform the observed data in discrete format. To this purpose we have to choose the value of the essential bandwidth B at level α of the useful signal $s(t)$, in such a way that $N=2BT>>1$, so that $\alpha<<1$.

- Once we have chosen B, we can expand the signal in a base of *sinc* functions:

$$\boldsymbol{\Psi} \equiv \left\{ \psi_i(t) = \sqrt{2B} \cdot \operatorname{sinc}\left(2B\left(t - \frac{i}{2B} \right) \right) \right\}_{i=0}^{2BT-1}$$

- i.e. we filter the continuous-time signal with a low pass filter of bandwidth B (*anti-aliasing filter*) and we sample the filtered signal $X_B(t)$ at intervals $T_c=1/2B$:



Estimate of the Amplitude of a Signal

- Since the bandwidth *B* has been selected so that *2BT*>>1 and α<<1, the useful signal passes undistorted through the anti-aliasing filter *s*(*t*):

$$X_B(t) = X(t) \otimes h_{LP}(t) = A \cdot s(t) \otimes h_{LP}(t) + W(t) \otimes h_{LP}(t) \cong A \cdot s(t) + W_B(t)$$

- In other words, the spectrum of the useful signal is not affected by the filter:

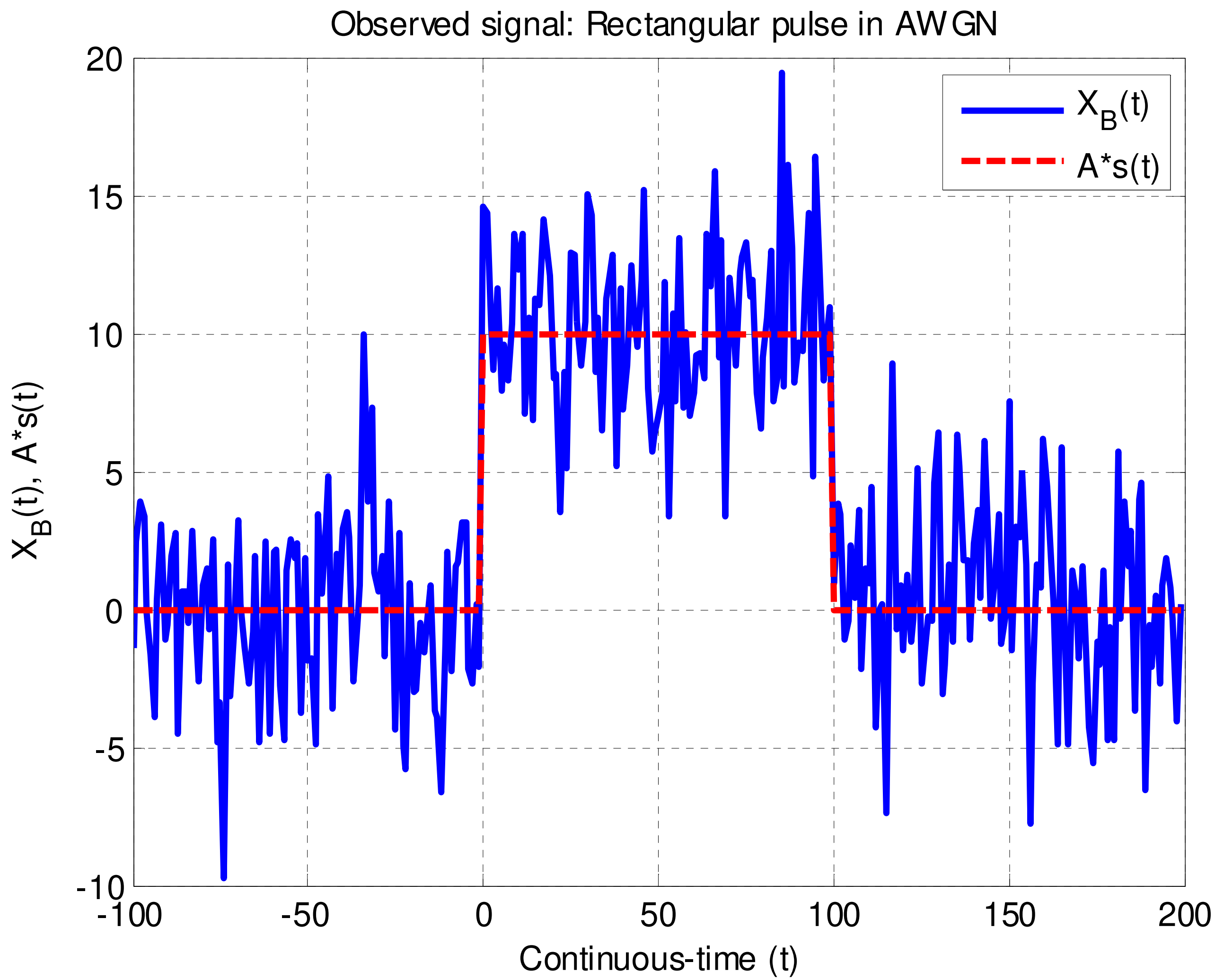
$$S(f)H_{LP}(f) = S(f) \cdot rect\left(\frac{f}{2B}\right) \cong S(f) \qquad \text{if} \quad N = 2BT \gg 1$$

- Power Spectral Density (PSD) and power of the filtered (bandlimited) noise:

$$S_{W_B}(f) = S_W(f)\left|H_{LP}(f)\right|^2 = \frac{N_0}{2} rect\left(\frac{f}{2B}\right) \quad \rightarrow \quad \sigma_{W_B}^2 = \int_{-\infty}^{+\infty} S_{W_B}(f)df = N_0 B$$

Estimate of the Amplitude of a Signal

Signal to Noise Power Ratio for each sample: $SNR_{in\,dB} = 10 Log_{10}\left(\dfrac{A^2}{\sigma^2_{W_B}}\right)$



Power of the bandlimited noise $W_B(t)$.

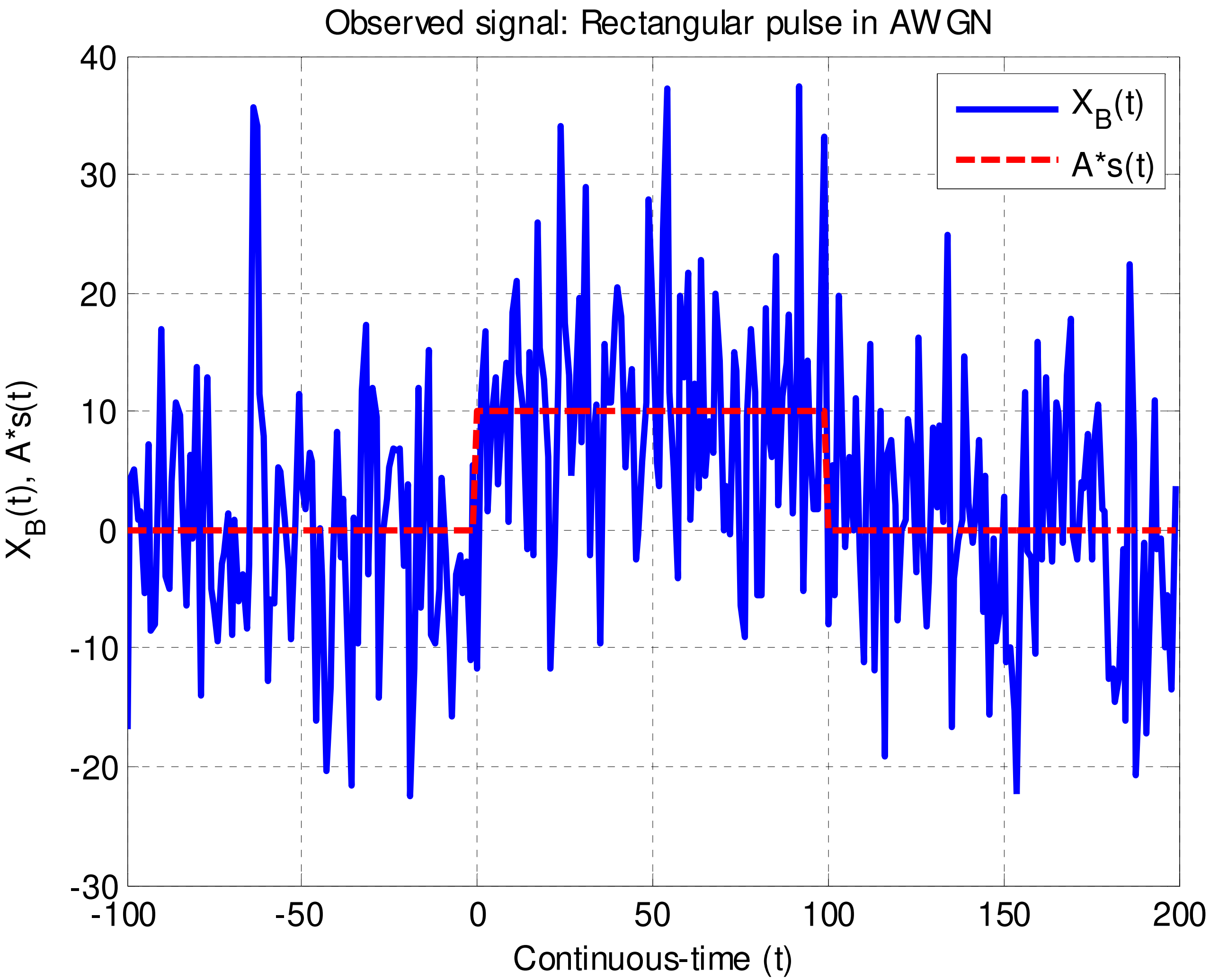
$$\sigma^2_{W_B} = \frac{N_0}{2} \cdot 2B = N_0 B$$

$$SNR_{in\,dB} = 10\,dB$$

$$A = 10$$

$$T = 100$$

Estimate of the Amplitude of a Signal



$$SNR_{in\,dB} = 0\,dB$$

$$A = 10$$

$$T = 100$$

Estimate of the Amplitude of a Signal

- The Autocorrelation Function (ACF) and the Power Spectral Density (PSD) of a process are related by the Fourier transform (**Wiener-Khintchine theorem**):

$$R_{W_B}(\tau) \overset{FT}{\Leftrightarrow} S_{W_B}(f)$$

$$S_{W_B}(f) = \frac{N_0}{2} rect\left(\frac{f}{2B}\right) \quad \rightarrow \quad R_{W_B}(\tau) \triangleq E\left\{W_B(t)W_B(t+\tau)\right\} = N_0 B \cdot \mathrm{sinc}\left(2B\tau\right)$$

- The image of the observed (noisy) signal is:

$$\mathbf{X} \triangleq \left[\frac{1}{\sqrt{2B}} X_B(0) \quad \frac{1}{\sqrt{2B}} X_B\left(\frac{1}{2B}\right) \quad \cdots \quad \frac{1}{\sqrt{2B}} X_B\left(\frac{N-1}{2B}\right)\right]^T$$

$$X_i \triangleq \frac{1}{\sqrt{2B}} X_B\left(\frac{i}{2B}\right) = \frac{A}{\sqrt{2B}} s\left(\frac{i}{2B}\right) + \frac{1}{\sqrt{2B}} W_B\left(\frac{i}{2B}\right) = A \cdot s_i + W_i, \quad i = 0, 1, \cdots, N-1$$

Estimate of the Amplitude of a Signal

$$\mathbf{X} = A \cdot \mathbf{s} + \mathbf{W} \quad \text{image of } X(t) \text{ for } t \in [0, T)$$

$$s_i \triangleq \frac{1}{\sqrt{2B}} s\left(\frac{i}{2B}\right), \quad W_i \triangleq \frac{1}{\sqrt{2B}} W_B\left(\frac{i}{2B}\right), \quad i = 0, 1, \cdots, N-1$$

$$\{W_i\}_{i=0}^{N-1} \text{ are IID, and so } \{X_i\}_{i=0}^{N-1} \text{ are mutually independent}$$

- The filtered noise is a Gaussian process, since it is obtained by filtering with a linear filter a Gaussian process. It is also zero mean, since the input noise is a zero mean process:

$$E\{W_i\} = E\left\{\frac{1}{\sqrt{2B}} W_B\left(\frac{i}{2B}\right)\right\} = \frac{1}{\sqrt{2B}} E\left\{W_B\left(\frac{i}{2B}\right)\right\} = \frac{1}{\sqrt{2B}} H_{LP}(0) \cdot E\{W(t)\} = 0$$

Estimate of the Amplitude of a Signal

- The samples of a Gaussian process are Gaussian random variables, so if they are **uncorrelated** they are also **independent**:

$$
\begin{aligned}
\operatorname{cov}\{W_i, W_j\} &= E\{W_i W_j\} = E\left\{\frac{1}{\sqrt{2B}} W_B\left(\frac{i}{2B}\right)\frac{1}{\sqrt{2B}} W_B\left(\frac{j}{2B}\right)\right\} \\
&= \frac{1}{2B} E\left\{W_B\left(\frac{i}{2B}\right) W_B\left(\frac{j}{2B}\right)\right\} = \frac{1}{2B} R_{W_B}\left(\frac{j-i}{2B}\right) \\
&= \frac{1}{2B} N_0 B \cdot \operatorname{sinc}\left(2B \cdot \frac{j-i}{2B}\right) = \frac{N_0}{2} \operatorname{sinc}(j-i) \\
&= \begin{cases} \dfrac{N_0}{2}, & i = j \\ 0, & i \neq j \end{cases}
\end{aligned}
$$

- Hence, the filtered noise samples $\{W_i\}$ are zero-mean, Gaussian distributed, independent, with variance $N_0/2$.

Estimate of the Amplitude of a Signal

$$\mathbf{X} = A \cdot \mathbf{s} + \mathbf{W} \qquad \text{image of } X(t) \text{ for } t \in [0,T)$$

$$s_i \triangleq \frac{1}{\sqrt{2B}} s\left(\frac{i}{2B}\right), \quad W_i \triangleq \frac{1}{\sqrt{2B}} W_B\left(\frac{i}{2B}\right), \quad i = 0,1,\cdots,N-1$$

$$X_i \in \mathcal{N}\left(A \cdot s_i, \sigma_W^2\right), \qquad \{X_i\}_{i=0}^{N-1} \text{ mutually independent}$$

$$\sigma_W^2 = \operatorname{var}\{W_i\} = \frac{1}{2B} \operatorname{var}\left\{W_B\left(\frac{i}{2B}\right)\right\} = \frac{1}{2B}\left(2B \cdot \frac{N_0}{2}\right) = \frac{N_0}{2}$$

$$f_{\mathbf{X}}(\mathbf{x}; A) = \prod_{i=0}^{N-1} f_{X_i}(x_i; A) = \prod_{i=0}^{N-1} \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(x_i - A \cdot s_i)^2}{N_0}} = \left(\pi N_0\right)^{-N/2} e^{-\frac{1}{N_0}\sum_{i=0}^{N-1}(x_i - A \cdot s_i)^2}$$

Estimate of the Amplitude of a Signal

$$\ln L(A) \triangleq \ln f_{\mathbf{X}}(\mathbf{x};A) = \sum_{i=0}^{N-1} \ln f_{X_i}(x_i;A) = \kappa - \frac{1}{N_0}\sum_{i=0}^{N-1}\left(x_i - A\cdot s_i\right)^2$$

$$\frac{d\ln L(A)}{dA} = \frac{2}{N_0}\sum_{i=0}^{N-1} s_i\left(x_i - A\cdot s_i\right) = \frac{2}{N_0}\left(\sum_{i=0}^{N-1} s_i x_i - A\sum_{i=0}^{N-1} s_i^2\right)$$

$$= \frac{2}{N_0}\left(\sum_{i=0}^{N-1} s_i x_i - AE_s\right) = \frac{2E_s}{N_0}\left(\frac{1}{E_s}\sum_{i=0}^{N-1} s_i x_i - A\right)$$

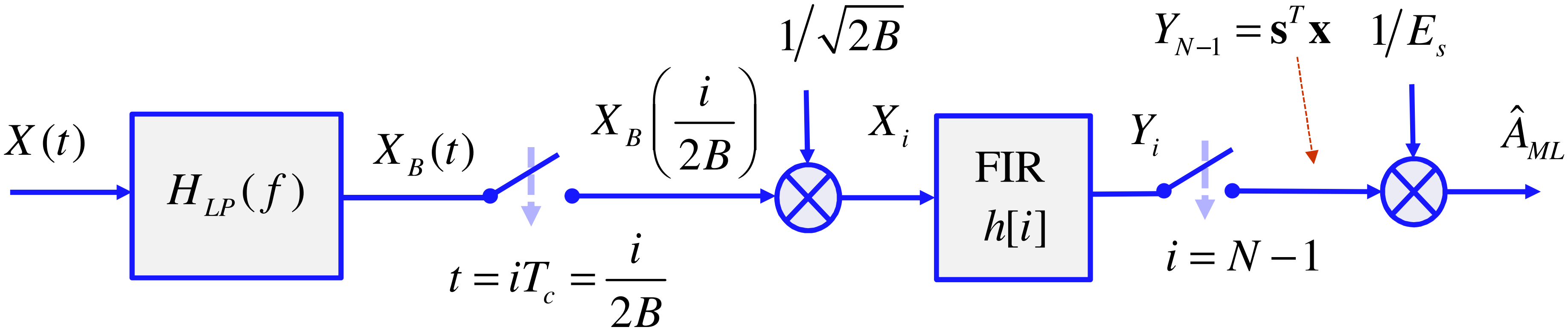
where $\sum_{i=0}^{N-1} s_i^2 = \|\mathbf{s}\|^2 = (\mathbf{s},\mathbf{s}) \overset{\text{if } N\gg 1}{=} (s(t),s(t)) = \int_0^T s^2(t)dt = E_s$

$$\hat{A}_{ML} = \frac{1}{E_s}\sum_{i=0}^{N-1} s_i x_i = \frac{1}{E_s}\mathbf{s}^T\mathbf{x}, \qquad MSE\left\{\hat{A}_{ML}\right\} = CRB(A) = \frac{1}{I(A)} = \frac{N_0}{2E_s}$$

Estimate of the Amplitude of a Signal

- The ML estimate of *A* is obtained by discrete-time FIR filtering the received data vector **x**.

- The FIR filter is called **Matched Filter (MF)** since its impulse response **h** is equal (i.e. matched) to the useful signal vector **s**, that is a priori known, since the signal shape is known.



$$\hat{A}_{ML} = \frac{1}{E_s}\sum_{i=0}^{N-1} s_i x_i = \frac{1}{E_s}\mathbf{s}^T\mathbf{x}$$

Matched Filter (MF): discrete-time FIR(N-1) filter with impulse response **h**=**s**, i.e. to be precise:

$h[i]=[\mathbf{h}]_{N-i}=s[N-1-i]$, $i=0,1,2, ,N-1$

Estimate of the Amplitude of a Signal

- The output of the MF is the cross-energy between the useful signal and the observed signal. It can be proved that if *N*>>1, this cross-energy can be calculated by processing either the discrete-time data or the continuous-time data, we get the same result.

$$\sum_{i=0}^{N-1} s_i x_i = \mathbf{s}^T \mathbf{x} = (\mathbf{x}, \mathbf{s}) \overset{\text{if } N \gg 1}{=} (X(t), s(t)) = \int_0^T X(t) s(t) dt = E_{sx} \quad \text{cross-energy}$$

$$\hat{A}_{ML} = \frac{1}{E_s} \sum_{i=0}^{N-1} s_i x_i = \frac{1}{E_s} \mathbf{s}^T \mathbf{x} = \frac{E_{sx}}{E_s} = \frac{1}{E_s} \int_0^T X(t) s(t) dt = \frac{1}{\sqrt{E_s}} \int_0^T X(t) \frac{1}{\sqrt{E_s}} s(t) dt$$

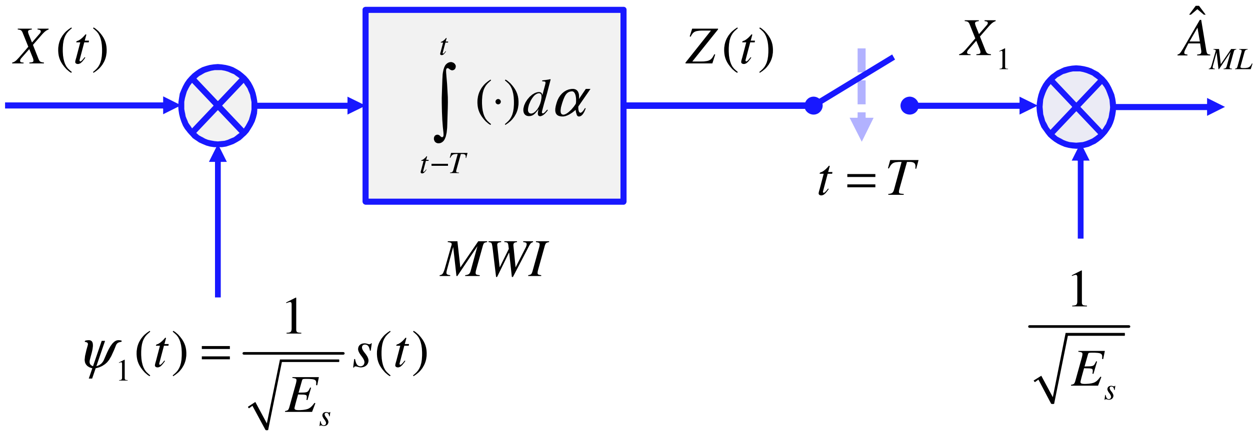
$$= \frac{1}{\sqrt{E_s}} \int_0^T X(t) \psi_1(t) dt = \frac{1}{\sqrt{E_s}} (X(t), \psi_1(t)) = \frac{1}{\sqrt{E_s}} X_1$$

where $\psi_1(t) \triangleq \frac{1}{\sqrt{E_s}} s(t)$ has unit energy, i.e. unit norm

Estimate of the Amplitude of a Signal

- Analog implementation of the ML estimator through a **correlator**, i.e. a moving window integrator (MWI):

$$X_1 \triangleq \int_0^T X(t)\psi_1(t)dt = \left[\int_{t-T}^{t} X(\alpha)\psi_1(\alpha)d\alpha \right]\Bigg|_{t=T}$$



$$\text{Ex. 1:} \quad s(t) = rect\left(\frac{t-T/2}{T}\right) \quad \rightarrow \quad E_s = T$$

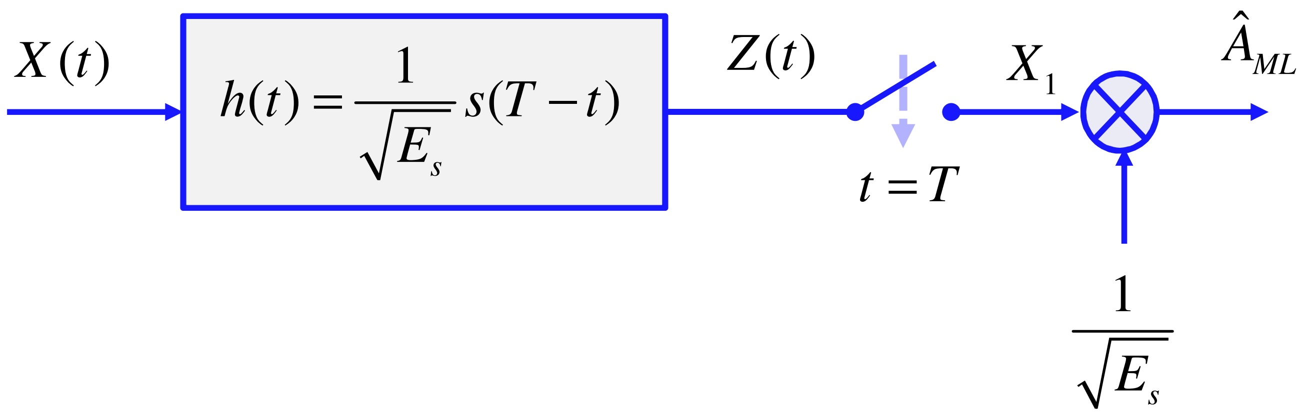
Estimate of the Amplitude of a Signal

- Analog implementation of the ML estimator through an analog **Matched Filter**:

$$Z(t) = X(t) \otimes h(t) = \int_0^T X(\alpha) h(t-\alpha) d\alpha = \int_0^T X(\alpha) \psi_1(T - t + \alpha) d\alpha$$

where $h(t) \triangleq \psi_1(T-t) = \dfrac{1}{\sqrt{E_s}} s(T-t)$ [analog Matched Filter (MF)]

$$X_1 \triangleq \int_0^T X(t)\psi_1(t) dt = \int_0^T X(\alpha)\psi_1(\alpha) d\alpha = [\underbrace{X(t) \otimes \psi_1(T-t)}_{Z(t)}]\Big|_{t=T} = Z(T)$$



Estimate of the Amplitude of a Signal

- The **matched filter** can be implemented using digital or analog technology.
- We can implement a **matched filter** since we *a-priori* know the signal shape (but not the amplitude).
- Matched filtering means calculate the inner product between the received vector $\mathbf{x}$ and the useful signal vector $\mathbf{s} \rightarrow \mathbf{s}^T\mathbf{x}$ (if the signal vector $\mathbf{s}$ is real).
- This tells us that <u>matched filtering is equivalent to the orthogonal projection of the received data vector **x** onto the 1D subspace generated by the signal vector **s**</u>.
- In other words, the only (principal) component we need to calculate is the one onto the direction of **s.** It is as if we calculate the **KLDT** where the first vector of the discrete-time basis is **s** (properly normalized) and we retain only the first component that is the one along **s**, the others being irrelevant.

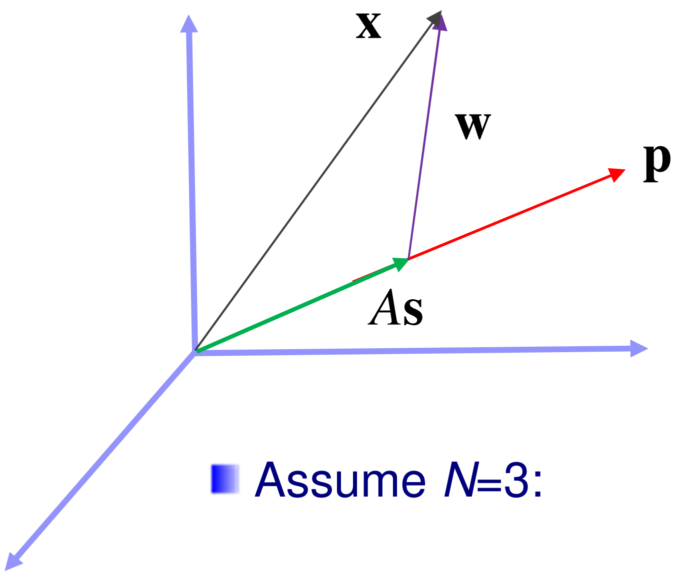
Geometrical Interpretation

$$\mathbf{x} = A\mathbf{s} + \mathbf{w} = A\sqrt{E_s} \cdot \frac{1}{\sqrt{E_s}}\mathbf{s} + \mathbf{w} = A\sqrt{E_s}\,\mathbf{p} + \mathbf{w}$$

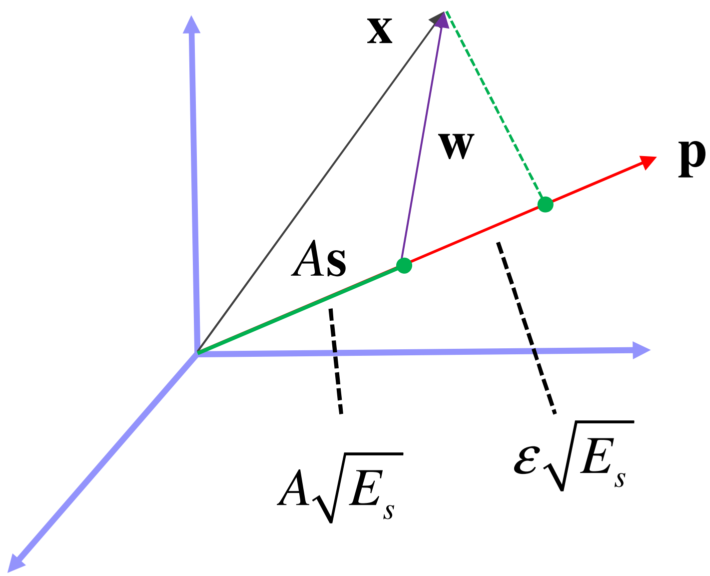
where $\mathbf{p} \triangleq \frac{1}{\sqrt{E_s}}\mathbf{s}$ so that $\|\mathbf{p}\|^2 = \frac{\|\mathbf{s}\|^2}{E_s} = 1$

$$\hat{A}_{ML} = \frac{1}{E_s}\mathbf{s}^T\mathbf{x} = \frac{1}{\sqrt{E_s}}\mathbf{p}^T\mathbf{x} = \frac{1}{\sqrt{E_s}}\left(A\sqrt{E_s}\,\mathbf{p}^T\mathbf{p} + \mathbf{p}^T\mathbf{w}\right) = \frac{1}{\sqrt{E_s}}\left(A\sqrt{E_s} + \varepsilon\sqrt{E_s}\right) = A + \varepsilon$$

where $\varepsilon \triangleq \left(\mathbf{p}^T\mathbf{w}\right)\big/\sqrt{E_s}$



- Assume N=3:



Estimate of the Amplitude of a Signal

- <u>**SNR at the output of the Matched Filter**</u>
- The ratio between the output and the input SNR is called **Processing Gain (PG)**.
- Let us calculate the input SNR (at the input of the digital Matched Filter, i.e. after the anti-aliasing filter):

Signal to Noise Power Ratio at the input of the digital Matched Filter:

$$SNR_{in} = \frac{E\left\{\left\|A\mathbf{s}\right\|^2\right\}}{E\left\{\left\|\mathbf{W}\right\|^2\right\}} = \frac{A^2\left\|\mathbf{s}\right\|^2}{E\left\{\sum_{i=1}^{N} W_i^2\right\}} = \frac{A^2\left\|\mathbf{s}\right\|^2}{\sum_{i=1}^{N} E\left\{W_i^2\right\}} = \frac{A^2 E_s}{N \cdot N_0/2}$$

Estimate of the Amplitude of a Signal

- The output of the Matched Filter (scaled by E_s) at discrete-time $i=N-1$ provides the ML estimate of A:

$$\hat{A}_{ML} = \frac{1}{E_s}\mathbf{s}^T\mathbf{x} = \frac{1}{E_s}\mathbf{s}^T(A\cdot\mathbf{s}+\mathbf{W}) = A\frac{1}{E_s}\mathbf{s}^T\mathbf{s} + \frac{1}{E_s}\mathbf{s}^T\mathbf{W} = \underbrace{A}_{\substack{\text{useful signal}\\ \text{component}}} + \underbrace{\frac{1}{E_s}\mathbf{s}^T\mathbf{W}}_{\text{Estimation error}}$$

$$= A + \varepsilon$$

$$MSE = E\{\varepsilon^2\} = E\left\{\left(\frac{1}{E_s}\mathbf{s}^T\mathbf{W}\right)^2\right\} = \frac{1}{E_s^2}E\left\{\mathbf{s}^T\mathbf{W}\left(\mathbf{s}^T\mathbf{W}\right)^T\right\} = \frac{1}{E_s^2}E\left\{\mathbf{s}^T\mathbf{W}\mathbf{W}^T\mathbf{s}\right\}$$

$$= \frac{1}{E_s^2}\mathbf{s}^T E\left\{\mathbf{W}\mathbf{W}^T\right\}\mathbf{s} = \frac{1}{E_s^2}\mathbf{s}^T\mathbf{C}_W\mathbf{s} = \frac{1}{E_s^2}\mathbf{s}^T\frac{N_0}{2}\mathbf{I}\mathbf{s} = \frac{N_0}{2E_s^2}\mathbf{s}^T\mathbf{s} = \frac{N_0}{2E_s}$$

$$\text{where } \mathbf{C}_W \triangleq E\left\{\mathbf{W}\mathbf{W}^T\right\} = \frac{N_0}{2}\mathbf{I}, \quad \text{since } \operatorname{cov}\left\{W_i, W_j\right\} = \begin{cases} \dfrac{N_0}{2}, & i=j \\ 0, & i\neq j \end{cases}$$

Estimate of the Amplitude of a Signal

- The Signal to Noise power Ratio (SNR) at the output of the digital Matched Filter (scaled by E_s) at discrete-time $i=N$-1 is given by:

$$SNR_{out} = \frac{A^2}{E\left\{\varepsilon^2\right\}} = \frac{A^2}{N_0/(2E_s)} = \frac{A^2 E_s}{N_0/2}$$

$$SNR_{out} = \frac{A^2 E_s}{N_0/2} = N \cdot \frac{A^2 E_s}{N \cdot N_0/2} = N \cdot SNR_{in}$$

$$PG \triangleq \frac{SNR_{out}}{SNR_{in}} = N$$

- The SNR at the output of the **Matched Filter** improves by a factor *N*. Hence, the **Processing Gain** of the MF is given by the number *N* of processed samples.

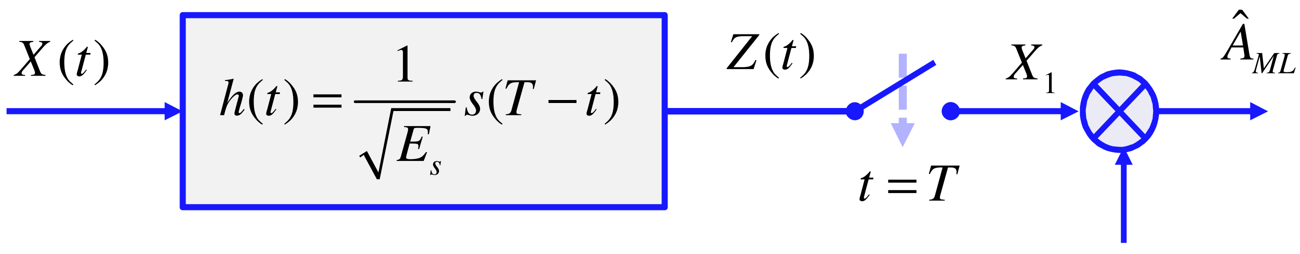
Estimate of the Amplitude of a Signal

- In the case of rectangular pulse, *s*(*t*) is constant over the observation interval, so the amplitude *A* also represents the mean of the observed data:

$$s(t) = rect\left(\frac{t - T/2}{T}\right) \quad \rightarrow \quad E_s = T$$

$$\begin{aligned}
\hat{A}_{ML} &= \frac{1}{E_s}\sum_{i=0}^{N-1} s_i x_i \\
&= \frac{1}{T}\sum_{i=0}^{N-1} \frac{1}{\sqrt{2B}} s\left(\frac{i}{2B}\right) \frac{1}{\sqrt{2B}} X\left(\frac{i}{2B}\right) \\
&= \frac{1}{2BT}\sum_{i=0}^{N-1} X\left(\frac{i}{2B}\right) \\
&= \frac{1}{N}\sum_{i=0}^{N-1} X\left(\frac{i}{2B}\right) \quad \leftarrow \quad \text{Sample Mean of the observed data}
\end{aligned}$$

Estimate of the Amplitude of a Signal



$$s(t)=rect\left(\frac{t-T/2}{T}\right) \quad \rightarrow \quad E_s=T$$

$$Z(t)=X(t)\otimes h(t)=\left[A\cdot s(t)+W_B(t)\right]\otimes h(t)$$

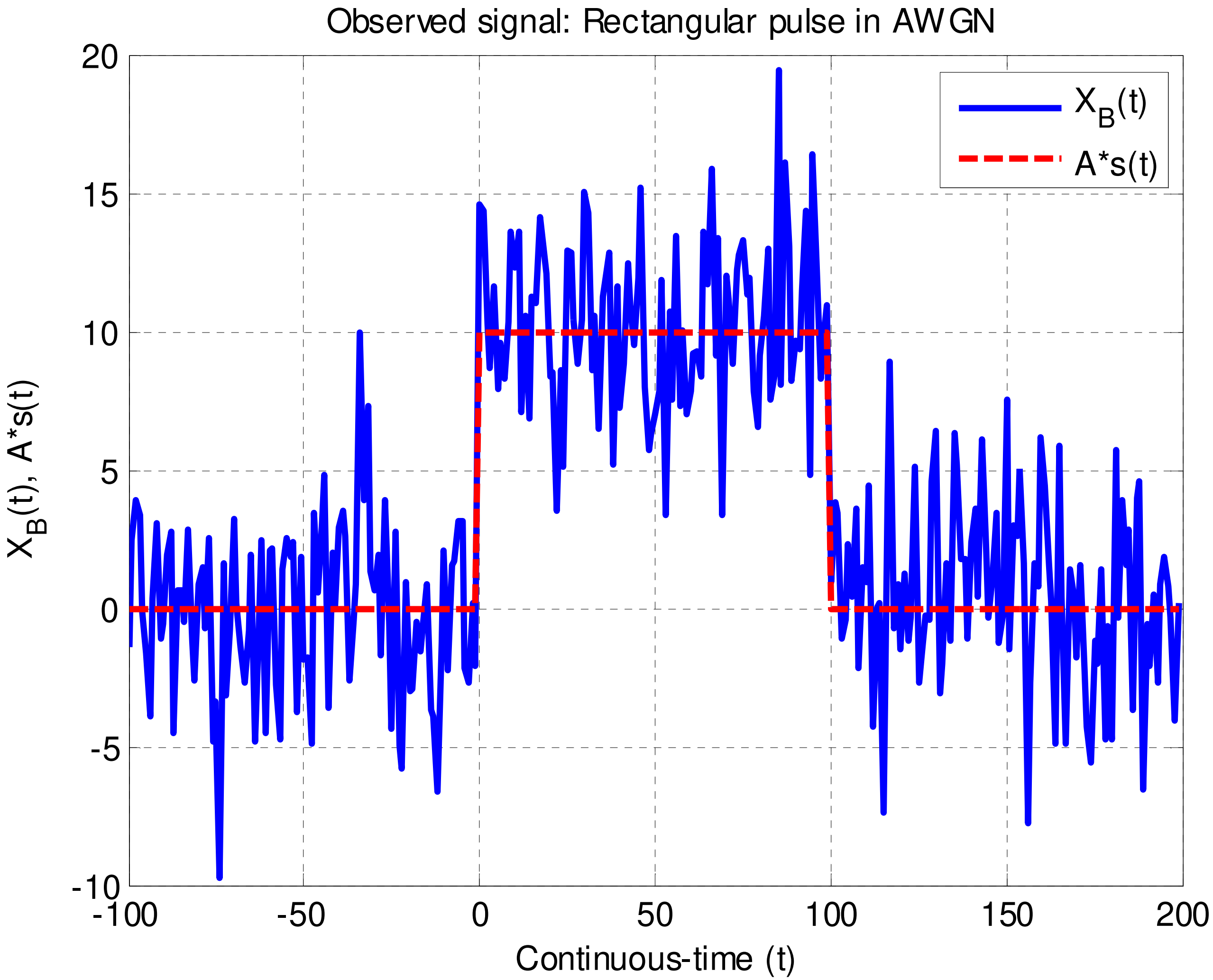
$$=A\cdot s(t)\otimes h(t)+W_B(t)\otimes h(t)$$

$$=Z_s(t)+Z_{W_B}(t)$$

$$\hat{A}_{ML}=\frac{1}{\sqrt{E_s}}X_1=\frac{1}{\sqrt{T}}Z(T)$$

Estimate of the Amplitude of a Signal

Signal to Noise Power Ratio for each sample: $SNR_{in\,dB} = 10Log_{10}\left(\dfrac{A^2}{\sigma_{W_B}^2}\right)$



Power of the bandlimited noise $W_B(t)$.

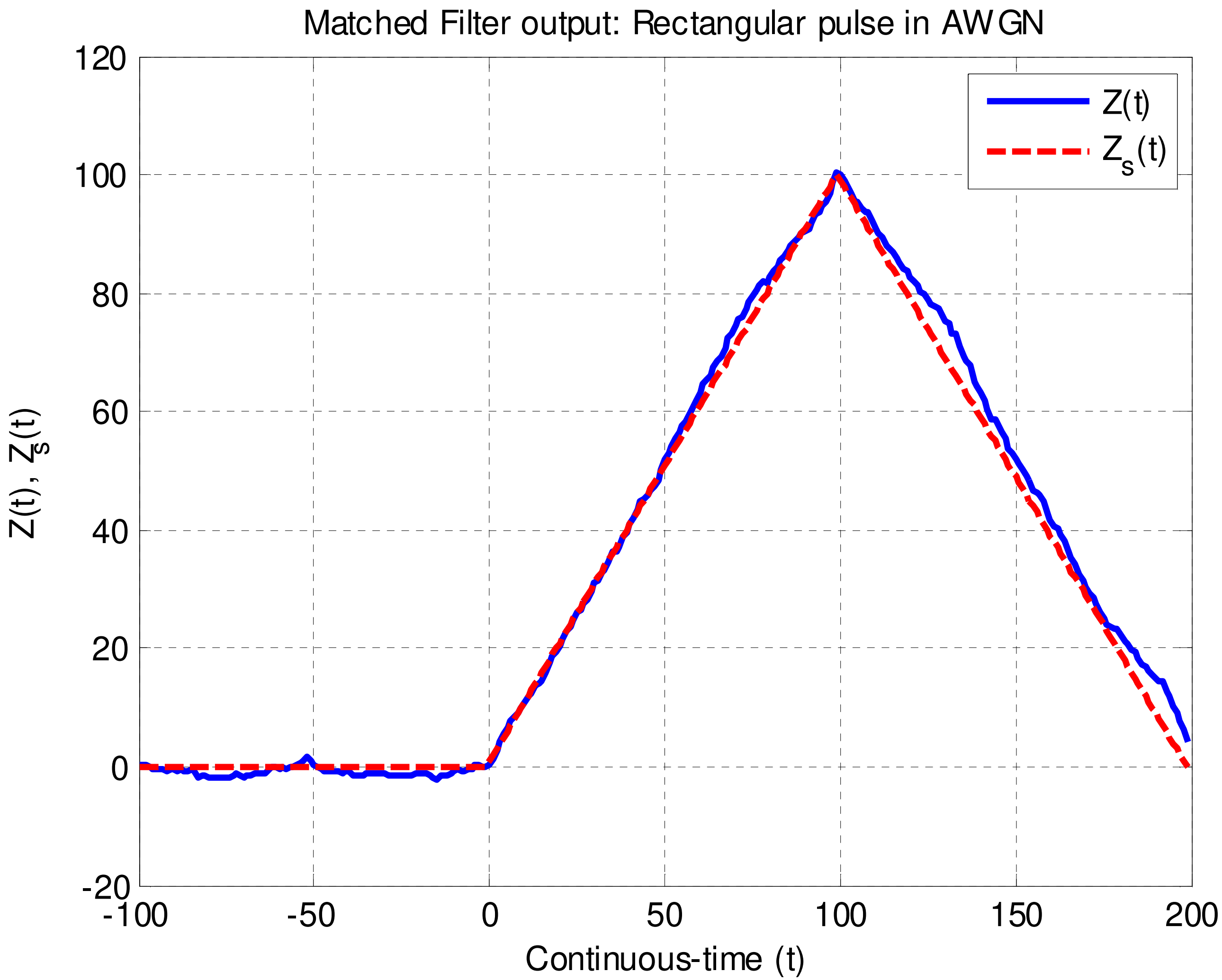
$$\sigma_{W_B}^2 = \frac{N_0}{2} \cdot 2B = N_0 B$$

$$SNR_{in\,dB} = 10\,dB$$

$$A = 10$$

$$T = 100$$

Estimate of the Amplitude of a Signal



$$SNR_{in\,dB} = 10\,dB$$

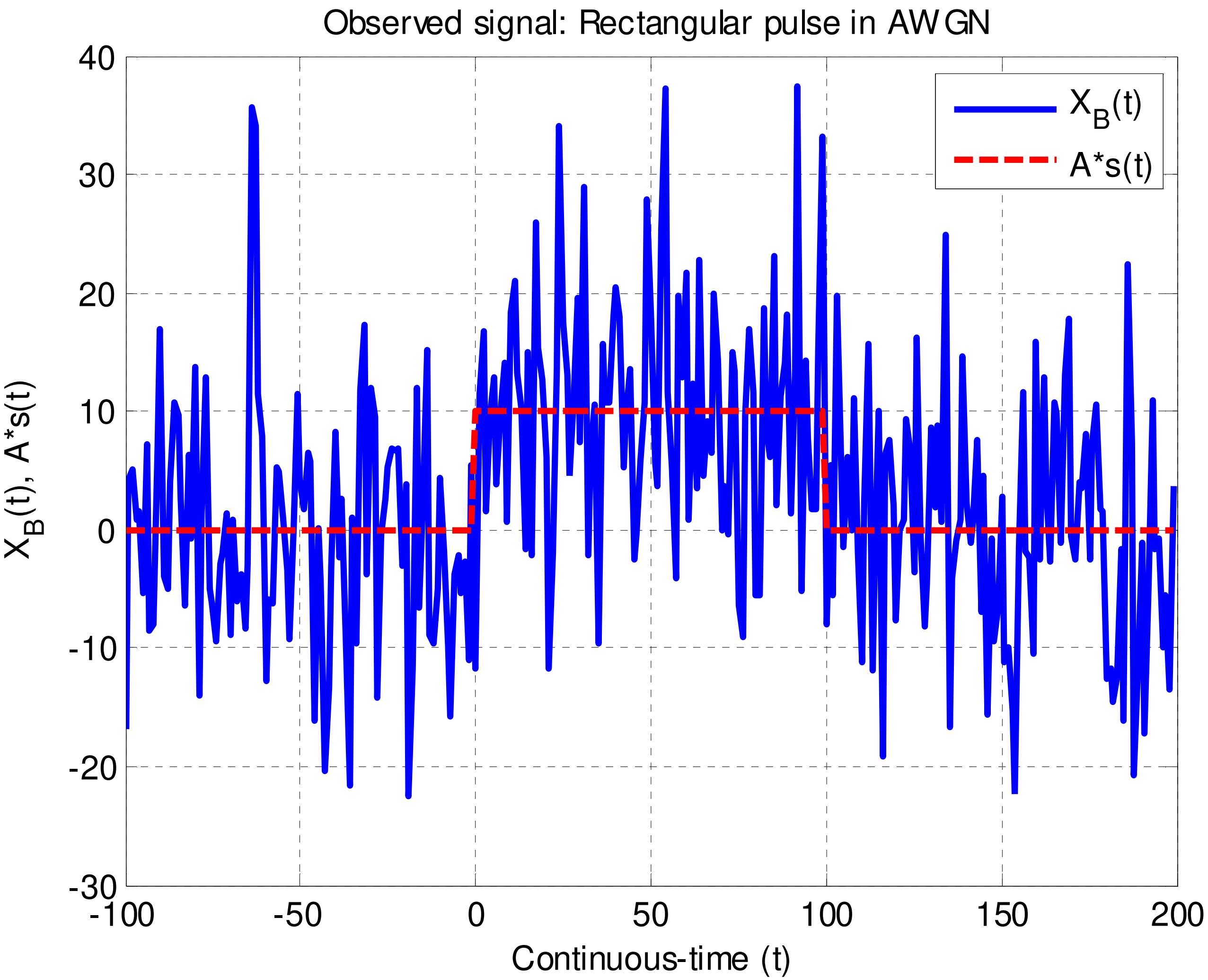
$$A = 10$$

$$T = 100$$

$$\hat{A}_{ML} = \frac{1}{\sqrt{E_s}} X_1 = \frac{1}{\sqrt{T}} Z(T) = \frac{Z(T)}{10}$$

$$A = \frac{Z_s(T)}{10}$$

Estimate of the Amplitude of a Signal

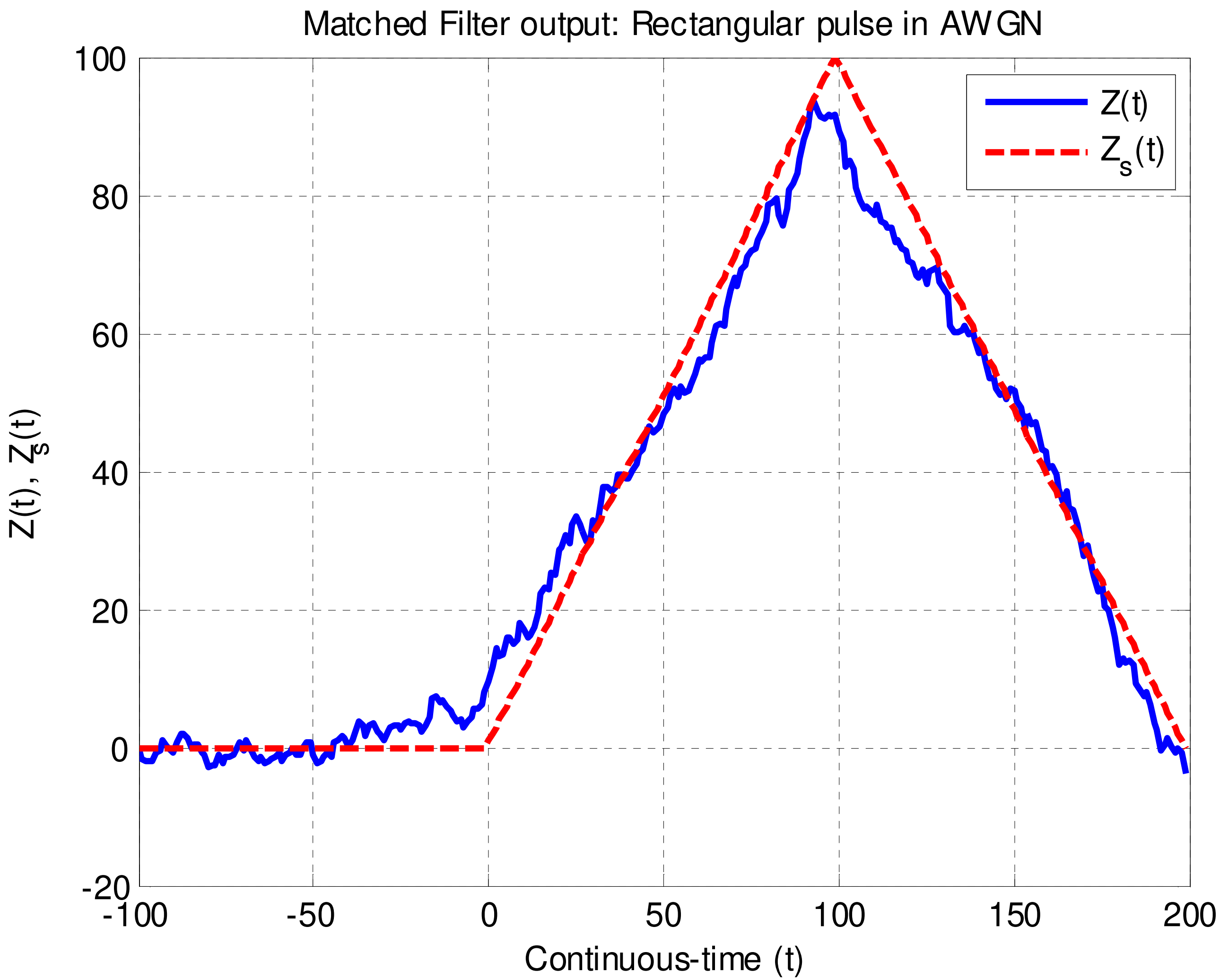


$$SNR_{in\,dB} = 0\,dB$$

$$A = 10$$

$$T = 100$$

Estimate of the Amplitude of a Signal



$$SNR_{in\,dB} = 0\,dB$$

$$A = 10$$

$$T = 100$$

$$\hat{A}_{ML} = \frac{1}{\sqrt{E_s}} X_1$$

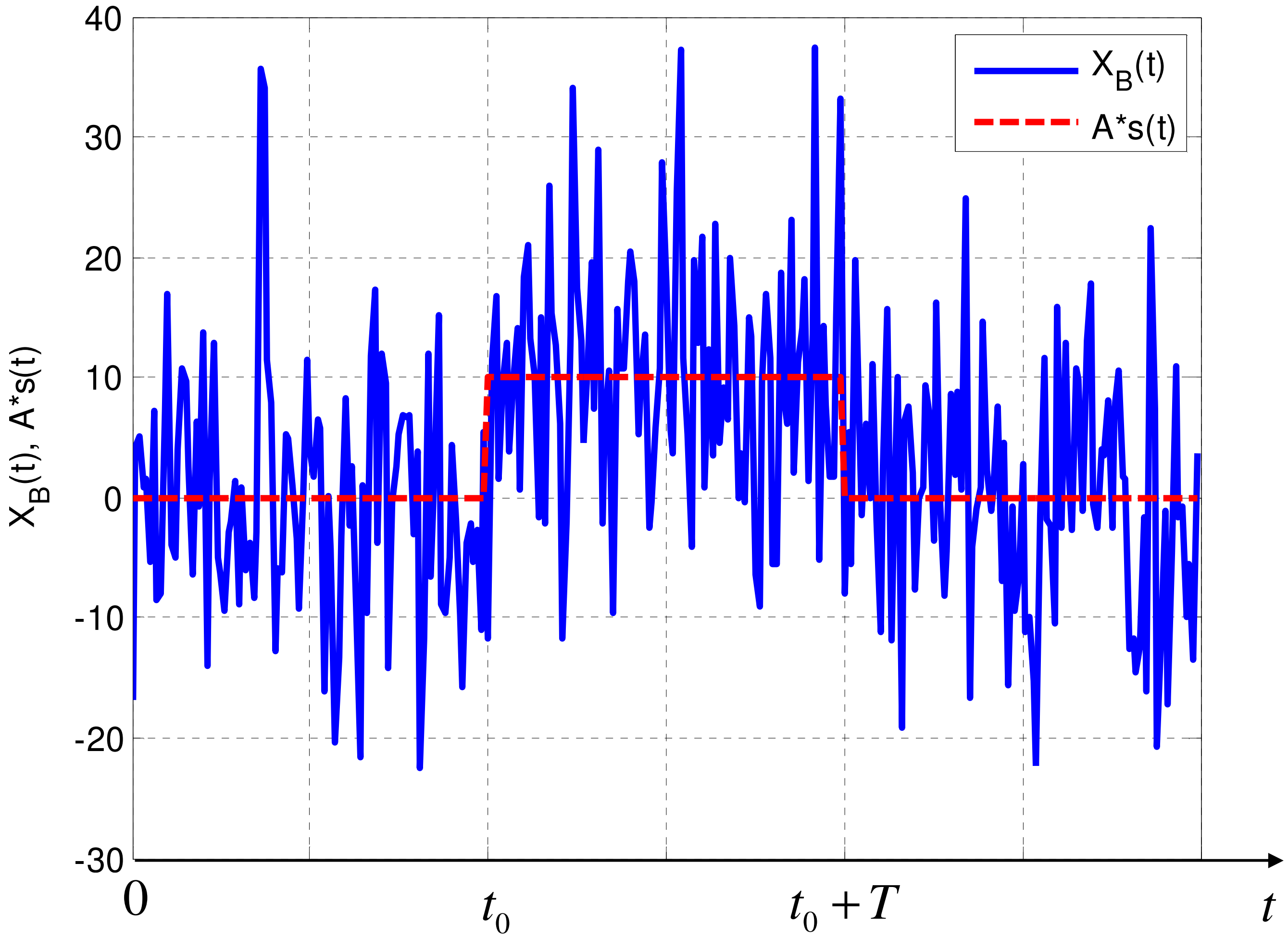
$$= \frac{1}{\sqrt{T}} Z(T)$$

$$= \frac{Z(T)}{10}$$

$$A = \frac{Z_s(T)}{10}$$

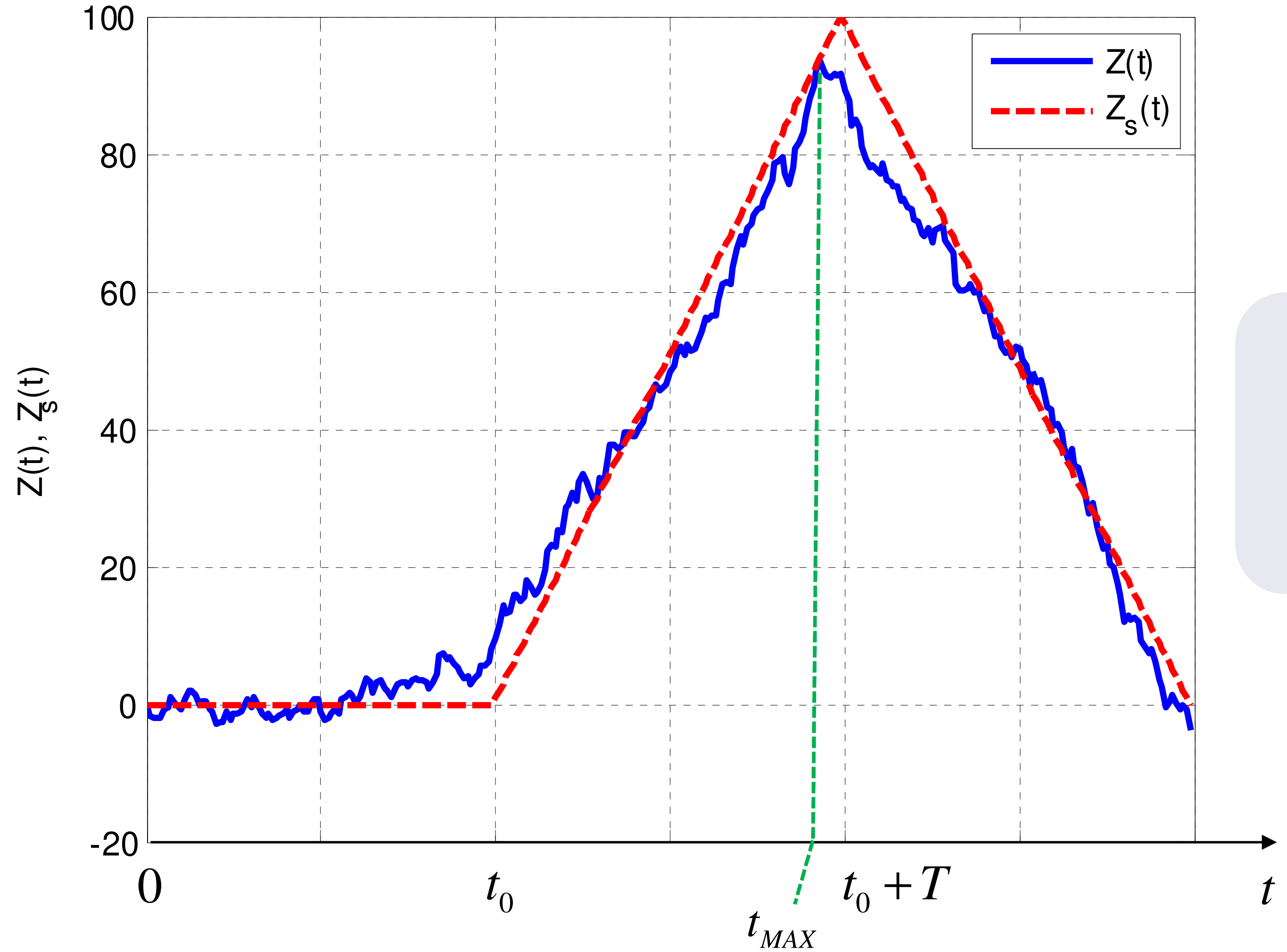
Estimate of the Time of Arrival (Range)

- If the Time of Arrival (ToA) t_0 is unknown, it should also be estimated.



Estimate of the Time of Arrival (Range)

- The ToA can be estimated from the time the output of the MF (or the correlator) is maximum.



$$t_{MAX} = t_0 + T + \varepsilon_t$$

$$\hat{t}_0 = t_{MAX} - T$$

ML Estimate of the Parameters of a Signal: Amplitude and Phase

Estimate of the Parameters of a Signal

- We observe a sinusoidal pulse *s*(*t*) with constant envelope over the time interval [0,*T*). The carrier frequency f_0 is perfectly known and <u>we want to estimate the unknown signal amplitude *A* and phase θ</u>. The signal is received corrupted by Additive White Gaussian Noise (AWGN):

$$X(t) = s(t) + W(t) = A \cdot p(t) + W(t), \quad t \in [0,T)$$

$$E_s \triangleq \int_0^T A^2 p^2(t)dt = A^2 E_p = \frac{A^2 T}{2} < +\infty$$

$$W(t)\ AWGN: \quad S_W(f) = \frac{N_0}{2}$$

$$p(t) = \cos\left(2\pi f_0 t + \theta\right) rect\left(\frac{t - T/2}{T}\right), \quad \text{with } f_0 \text{ known,} \quad f_0 T \gg 1$$

Estimate of the Parameters of a Signal

- The energy of the signal *p*(*t*) is:

$$E_p = \int_0^T p^2(t)dt = \int_0^T \cos^2\left(2\pi f_0 t + \theta\right)dt = \int_0^T \left[\frac{1}{2} + \frac{1}{2}\cos\left(4\pi f_0 t + 2\theta\right)\right]dt$$

$$= \frac{1}{2}\int_0^T dt + \frac{1}{2}\int_0^T \cos\left(4\pi f_0 t + 2\theta\right)dt = \frac{T}{2}\left(1 + \frac{1}{T}\int_0^T \cos\left(4\pi f_0 t + 2\theta\right)dt\right)$$

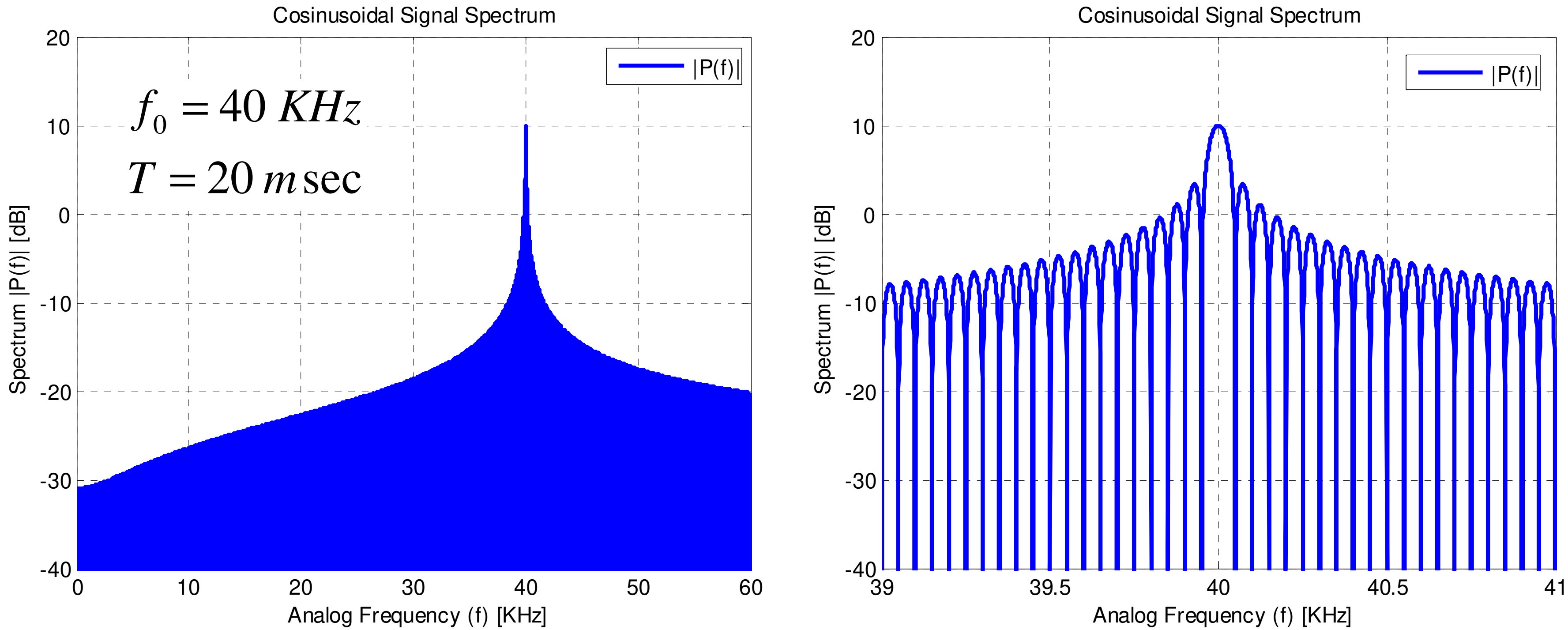
$$\cong \frac{T}{2}$$

$$\text{In fact:} \quad \frac{1}{T}\int_0^T \cos\left(4\pi f_0 t + 2\theta\right)dt \ll 1 \quad \text{if } f_0 T \gg 1$$

- The spectrum of the signal *p*(*t*) is given by its the Fourier Transform (FT):

$$P(f) = FT\left\{p(t)\right\} = \frac{T}{2} sinc\left((f + f_0)T\right)e^{-j\theta}e^{-j\pi(f+f_0)T} + \frac{T}{2} sinc\left((f - f_0)T\right)e^{j\theta}e^{-j\pi(f-f_0)T}$$

Estimate of the Parameters of a Signal



- The FT of the useful signal $s(t)=Ap(t)$ calculated at the known central frequency f_0:

$$A \cdot P(f_0) = \frac{AT}{2} sinc\left(2 f_0 T\right) e^{-j\theta} e^{-j2\pi f_0 T} + \frac{AT}{2} e^{j\theta} \cong \frac{AT}{2} e^{j\theta} \quad \text{if} \quad f_0 T \gg 1$$

- Note that the amplitude and phase of $A \cdot P(f_0)$ contains the information on the unknown parameters A and θ.

Estimate of the Parameters of a Signal

- The FT of the signal of interest $s(t)=Ap(t)$ calculated at the known frequency f_0:

$$S(f_0) = FT\left\{s(t)\right\}\Big|_{f=f_0} = A \cdot P(f_0) \cong \frac{AT}{2} e^{j\theta} \quad \text{if} \quad f_0 T \gg 1$$

- Hence, the amplitude and phase of $A \cdot P(f_0)$ contains the information on the unknown parameters A and θ.

- <u>In the absence of noise</u>, we can retrieve the true value of the amplitude and the phase from the FT of received signal $X(t)=s(t)$ as follows:

$$X(f_0) = S(f_0) = FT\left\{s(t)\right\}\Big|_{f=f_0} = A \cdot P(f_0) \cong \frac{AT}{2} e^{j\theta} \quad \text{if} \quad f_0 T \gg 1$$

$$\Rightarrow \quad \begin{cases} A = \dfrac{2}{T}\left|S(f_0)\right| = \dfrac{2}{T}\left|X(f_0)\right| \\ \theta = \angle S(f_0) = \angle X(f_0) \end{cases} \qquad \text{where } X(f_0) = FT\left\{X(t)\right\}\Big|_{f=f_0}$$

Estimate of the Parameters of a Signal

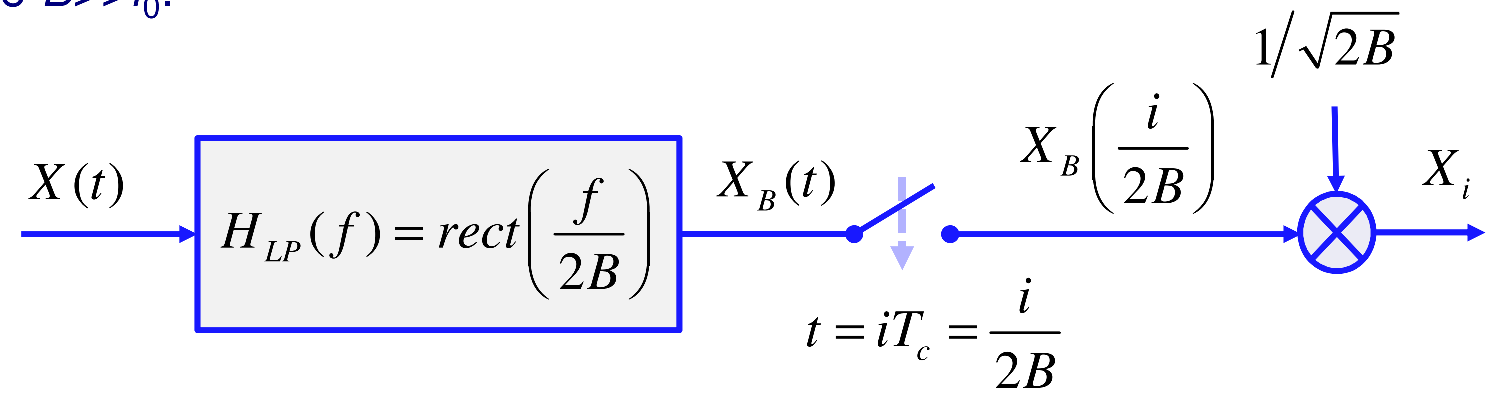
- When the noise is present, the FT of the received signal $X(t)$ represents an estimate of the FT of the signal of interest $s(t)$. Hence, we can use the previous relationships to estimate the amplitude and the phase of $s(t)$ from the FT of received (noisy) signal $X(t)$:

$$\begin{cases} \hat{A}_{MM} = \dfrac{2}{T}\left|\hat{S}(f_0)\right| = \dfrac{2}{T}\left|X(f_0)\right| \\ \hat{\theta}_{MM} = \angle\hat{S}(f_0) = \angle X(f_0) \end{cases} \qquad \text{where } X(f_0) = FT\{X(t)\}\Big|_{f=f_0}$$

- The estimators derived in this way are a sort of **Method of Moment estimators**, since the estimate is obtained by exploiting the relationship between the unknown parameters and a “moment” of the data, i.e. the data Fourier Transform at the frequency f_0 (which is the frequency where the signal spectrum is maximum).

Estimate of the Parameters of a Signal

- As in the previous case, the received signal is converted in discrete form by *anti-aliasing* filtering and sampling the filtered signal $X_B(t)$ at intervals T_c=1/2B, where B>>f_0.



$$X_i \triangleq \frac{1}{\sqrt{2B}} X_B\left(\frac{i}{2B}\right) = \frac{A}{\sqrt{2B}} p\left(\frac{i}{2B}\right) + \frac{1}{\sqrt{2B}} W_B\left(\frac{i}{2B}\right) = A \cdot p_i(\theta) + W_i, \quad i = 0, \cdots, N-1$$

$$p_i(\theta) \triangleq \frac{1}{\sqrt{2B}} p\left(\frac{i}{2B}\right) = \frac{1}{\sqrt{2B}} \cos\left(2\pi f_0 \frac{i}{2B} + \theta\right), \quad W_i \triangleq \frac{1}{\sqrt{2B}} W_B\left(\frac{i}{2B}\right)$$

$$X_i \in \mathcal{N}\left(A \cdot p_i(\theta), \sigma_W^2\right), \qquad \{X_i\}_{i=0}^{N-1} \text{ are mutually independent}$$

Estimate of the Parameters of a Signal

$\mathbf{X}$ image of $X(t)$ for $t \in [0,T)$

$$\mathbf{X} = \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{bmatrix} = A \cdot \mathbf{p}(\theta) + \mathbf{W} = \begin{bmatrix} \frac{A}{\sqrt{2B}}\cos(\theta) \\ \frac{A}{\sqrt{2B}}\cos\left(2\pi f_0 \frac{1}{2B} + \theta\right) \\ \vdots \\ \frac{A}{\sqrt{2B}}\cos\left(2\pi f_0 \frac{N-1}{2B} + \theta\right) \end{bmatrix} + \begin{bmatrix} W_0 \\ W_1 \\ \vdots \\ W_{N-1} \end{bmatrix}$$

$$W_i \in \mathcal{N}\left(0, \sigma_W^2\right), \text{ where } \sigma_W^2 = \operatorname{var}\{W_i\} = \frac{N_0}{2} \quad \rightarrow \quad X_i \in \mathcal{N}\left(A \cdot p_i(\theta), \sigma_W^2\right)$$

$$p_i(\theta) = \frac{1}{\sqrt{2B}}\cos\left(2\pi f_0 \frac{i}{2B} + \theta\right) = \frac{1}{\sqrt{2B}}\cos\left(2\pi F_0 i + \theta\right)$$

$$F_0 \triangleq \frac{f_0}{2B} \quad \text{Digital Frequency - Nyquist Sampling} \quad \rightarrow \quad 0 < F_0 < \frac{1}{2}$$

Estimate of the Parameters of a Signal

- The input Signal to Noise power Ratio (SNR_{in}):

$$SNR_{in} = \frac{\left\| A \cdot \mathbf{p}(\theta) \right\|^2}{E\left\{ \left\| \mathbf{W} \right\|^2 \right\}} = \frac{A^2 \left\| \mathbf{p}(\theta) \right\|^2}{E\left\{ \left\| \mathbf{W} \right\|^2 \right\}} = \frac{A^2}{2B} \cdot \frac{N}{2} \cdot \frac{2}{N_0 N} = \frac{A^2}{2 N_0 B}$$

$$\left\| \mathbf{p}(\theta) \right\|^2 = \sum_{i=0}^{N-1} \left(p_i(\theta) \right)^2 = \frac{N}{2B} \cdot \underbrace{\frac{1}{N} \sum_{i=0}^{N-1} \cos^2\left(2\pi F_0 i + \theta \right)}_{\cong 1/2 \ \text{ se } N \gg 1} = \frac{N}{2B} \cdot \frac{1}{2}$$

$$E\left\{ \left\| \mathbf{W} \right\|^2 \right\} = E\left\{ \sum_{i=0}^{N-1} W_i^2 \right\} = \sum_{i=0}^{N-1} E\left\{ W_i^2 \right\} = N \sigma_W^2 = N \cdot \frac{N_0}{2}$$

Estimate of the Parameters of a Signal

- Joint pdf of the observed data:

$$f_{\mathbf{X}}(\mathbf{x};A,\theta)=\prod_{i=0}^{N-1} f_{X_i}(x_i;A,\theta)=\prod_{i=0}^{N-1}\frac{1}{\sqrt{\pi N_0}}e^{-\frac{1}{N_0}\left(x_i-\frac{A}{\sqrt{2B}}\cos(2\pi F_0 i+\theta)\right)^2}$$

$$=\left(\pi N_0\right)^{-N/2}e^{-\frac{1}{N_0}\sum_{i=0}^{N-1}\left(x_i-\frac{A}{\sqrt{2B}}\cos(2\pi F_0 i+\theta)\right)^2}$$

- The log-likelihood function (log-LF):

$$\ln L(A,\theta)=\ln f_{\mathbf{X}}(\mathbf{x};A,\theta)=-\frac{N}{2}\ln\left(\pi N_0\right)-\frac{1}{N_0}\sum_{i=0}^{N-1}\left(x_i-\frac{A}{\sqrt{2B}}\cos\left(2\pi F_0 i+\theta\right)\right)^2$$

$$=-\frac{N}{2}\ln\left(\pi N_0\right)-\frac{1}{N_0}\sum_{i=0}^{N-1}x_i^2-\frac{A^2}{N_0 2B}\sum_{i=0}^{N-1}\cos^2\left(2\pi F_0 i+\theta\right)+\frac{2A}{N_0\sqrt{2B}}\sum_{i=0}^{N-1}x_i\cos\left(2\pi F_0 i+\theta\right)$$

$$\cong\kappa-\frac{A^2}{N_0 2B}\cdot\frac{N}{2}+\frac{2A}{N_0\sqrt{2B}}\sum_{i=0}^{N-1}x_i\cos\left(2\pi F_0 i+\theta\right)$$

Estimate of the Parameters of a Signal

In fact, if $N \gg 1$:

$$\frac{1}{N}\sum_{i=0}^{N-1}\cos^2\left(2\pi F_0 i+\theta\right)=\frac{1}{N}\sum_{i=0}^{N-1}\left[\frac{1}{2}+\frac{1}{2}\cos\left(4\pi F_0 i+2\theta\right)\right]$$

$$=\frac{1}{2}\left[1+\underbrace{\frac{1}{N}\sum_{i=0}^{N-1}\cos\left(4\pi F_0 i+2\theta\right)}_{\cong 0 \text{ if } N\gg 1}\right]\cong\frac{1}{2}$$

We can also prove that:

$$\frac{1}{N}\sum_{i=0}^{N-1}\sin^2\left(2\pi F_0 i+\theta\right)\cong\frac{1}{2}$$

$$\frac{1}{N}\sum_{i=0}^{N-1}\cos\left(2\pi F_0 i+\theta\right)\sin\left(2\pi F_0 i+\theta\right)=\frac{1}{2}\cdot\frac{1}{N}\sum_{i=0}^{N-1}\sin\left(4\pi F_0 i+2\theta\right)\cong 0$$

Estimate of the Parameters of a Signal

- The **joint ML estimators** of A and θ are obtained by jointly solving the two **likelihood equations**, which are obtained by deriving the log-LF with respect to both the unknown parameters:

$$\begin{cases} \dfrac{\partial \ln L(A,\theta)}{\partial A} = 0 \\ \dfrac{\partial \ln L(A,\theta)}{\partial \theta} = 0 \end{cases}$$

$$\frac{\partial \ln L(A,\theta)}{\partial A} = \frac{\partial}{\partial A}\left[-\frac{A^2}{N_0 2B} \cdot \frac{N}{2} + \frac{2A}{N_0\sqrt{2B}} \sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \theta\right) \right]$$

$$= -\frac{AN}{N_0 2B} + \frac{2}{N_0\sqrt{2B}} \sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \theta\right) = 0$$

$$\rightarrow \quad A = \sqrt{2B} \cdot \frac{2}{N} \sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \theta\right)$$

Estimate of the Parameters of a Signal

$$\frac{\partial \ln L(A,\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left[-\frac{A^2}{N_0 2B}\cdot\frac{N}{2} + \frac{2A}{N_0\sqrt{2B}}\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \theta\right)\right]$$

$$= -\frac{2A}{N_0\sqrt{2B}}\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i + \theta\right) = 0$$

$$\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i + \theta\right) = \sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right)\cos\left(\theta\right) + \sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right)\sin\left(\theta\right) = 0$$

$$\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right)\sin\left(\theta\right) = -\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right)\cos\left(\theta\right)$$

$$\frac{\sin\left(\theta\right)}{\cos\left(\theta\right)} = -\frac{\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right)}{\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right)} \quad \rightarrow \quad \theta = -\arctan\left(\frac{\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right)}{\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right)}\right)$$

Estimate of the Parameters of a Signal

- Hence, the ML estimator of θ is given by:

$$\hat{\theta}_{ML} = -\arctan\left(\frac{\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right)}{\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right)}\right) = -\arctan\left(\frac{\mathbf{s}^T\mathbf{x}}{\mathbf{c}^T\mathbf{x}}\right)$$

where we defined the two orthonormal (orthogonal and unit norm vectors):

$$\mathbf{c} \triangleq \begin{bmatrix} c_0 & c_1 & \cdots & c_{N-1} \end{bmatrix}, \text{ where } c_i \triangleq \sqrt{\frac{2}{N}}\cos\left(2\pi F_0 i\right), \quad \left\|\mathbf{c}\right\|^2 = \mathbf{c}^T\mathbf{c} = 1$$

$$\mathbf{s} \triangleq \begin{bmatrix} s_0 & s_1 & \cdots & s_{N-1} \end{bmatrix}, \text{ where } s_i \triangleq \sqrt{\frac{2}{N}}\sin\left(2\pi F_0 i\right), \quad \left\|\mathbf{s}\right\|^2 = \mathbf{s}^T\mathbf{s} = 1$$

$$\mathbf{s}^T\mathbf{c} = \mathbf{c}^T\mathbf{s} = 0 \quad \rightarrow \quad \mathbf{s} \perp \mathbf{c}$$

Estimate of the Parameters of a Signal

- Now, we should insert the ML estimate of θ in the previous expression of A:

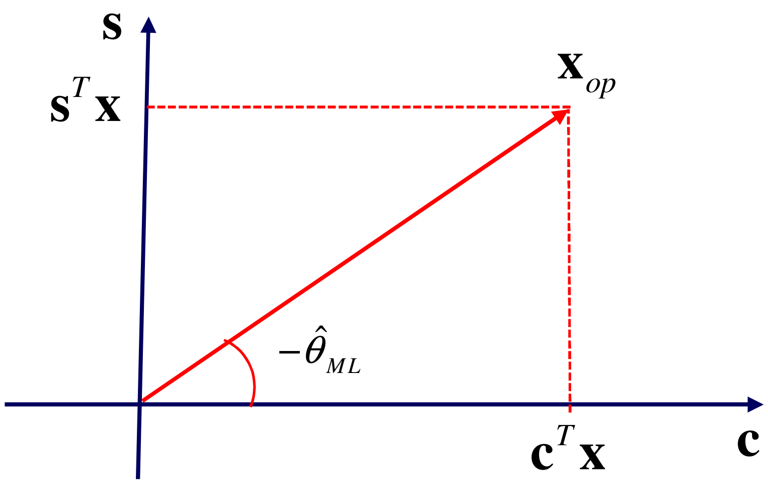
$$\hat{A}_{ML} = \sqrt{2B} \cdot \frac{2}{N} \sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \hat{\theta}_{ML}\right)$$

$$= \sqrt{2B} \cdot \frac{2}{N} \sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right) \cos\left(\hat{\theta}_{ML}\right) - \sqrt{2B} \cdot \frac{2}{N} \sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right) \sin\left(\hat{\theta}_{ML}\right)$$

$$= \sqrt{2B} \sqrt{\frac{2}{N}} \left(\mathbf{c}^T \mathbf{x}\right) \cos\left(\hat{\theta}_{ML}\right) - \sqrt{2B} \sqrt{\frac{2}{N}} \left(\mathbf{s}^T \mathbf{x}\right) \sin\left(\hat{\theta}_{ML}\right)$$

$$\overset{N=2BT}{=} \sqrt{\frac{2}{T}} \left(\mathbf{c}^T \mathbf{x}\right) \cos\left(\hat{\theta}_{ML}\right) - \sqrt{\frac{2}{T}} \left(\mathbf{s}^T \mathbf{x}\right) \sin\left(\hat{\theta}_{ML}\right)$$

Estimate of the Parameters of a Signal



- where $\mathbf{x}_{op}$ is the orthogonal projection of the received N-dim vector $\mathbf{x}$ onto the plane generated by the orthonormal vectors **c** and **s** (which is the **signal subspace**):

$$
\begin{cases}
\cos\left(\hat{\theta}_{ML}\right) = \cos\left(-\hat{\theta}_{ML}\right) = \dfrac{\mathbf{c}^T\mathbf{x}}{\sqrt{\left(\mathbf{c}^T\mathbf{x}\right)^2 + \left(\mathbf{s}^T\mathbf{x}\right)^2}} \\
-\sin\left(\hat{\theta}_{ML}\right) = \sin\left(-\hat{\theta}_{ML}\right) = \dfrac{\mathbf{s}^T\mathbf{x}}{\sqrt{\left(\mathbf{c}^T\mathbf{x}\right)^2 + \left(\mathbf{s}^T\mathbf{x}\right)^2}}
\end{cases}
$$

$$
\boxed{\begin{aligned} \mathbf{c}^T\mathbf{x} &= \mathbf{c}^T\mathbf{x}_{op} \\ \mathbf{s}^T\mathbf{x} &= \mathbf{s}^T\mathbf{x}_{op} \end{aligned}}
$$

Estimate of the Parameters of a Signal

- The ML estimator of *A* is immediately derived:

$$
\begin{aligned}
\hat{A}_{ML} &= \sqrt{\frac{2}{T}}\left(\mathbf{c}^T\mathbf{x}\right)\cos\left(\hat{\theta}_{ML}\right) - \sqrt{\frac{2}{T}}\left(\mathbf{s}^T\mathbf{x}\right)\sin\left(\hat{\theta}_{ML}\right) \\
&= \sqrt{\frac{2}{T}}\left[\left(\mathbf{c}^T\mathbf{x}\right)\cos\left(\hat{\theta}_{ML}\right) + \left(\mathbf{s}^T\mathbf{x}\right)\sin\left(-\hat{\theta}_{ML}\right)\right] \\
&= \sqrt{\frac{2}{T}}\left[\left(\mathbf{c}^T\mathbf{x}\right)\frac{\mathbf{c}^T\mathbf{x}}{\sqrt{\left(\mathbf{c}^T\mathbf{x}\right)^2 + \left(\mathbf{s}^T\mathbf{x}\right)^2}} + \left(\mathbf{s}^T\mathbf{x}\right)\frac{\mathbf{s}^T\mathbf{x}}{\sqrt{\left(\mathbf{c}^T\mathbf{x}\right)^2 + \left(\mathbf{s}^T\mathbf{x}\right)^2}}\right] \\
&= \sqrt{\frac{2}{T}}\left[\frac{\left(\mathbf{c}^T\mathbf{x}\right)^2}{\sqrt{\left(\mathbf{c}^T\mathbf{x}\right)^2 + \left(\mathbf{s}^T\mathbf{x}\right)^2}} + \frac{\left(\mathbf{s}^T\mathbf{x}\right)^2}{\sqrt{\left(\mathbf{c}^T\mathbf{x}\right)^2 + \left(\mathbf{s}^T\mathbf{x}\right)^2}}\right] \\
&= \sqrt{\frac{2}{T}}\sqrt{\left(\mathbf{c}^T\mathbf{x}\right)^2 + \left(\mathbf{s}^T\mathbf{x}\right)^2}
\end{aligned}
$$

Estimate of the Parameters of a Signal

- In summary, the **joint ML estimators** of *A* and θ are given by:

$$\hat{A}_{ML} = \sqrt{\frac{2}{T}}\sqrt{\left(\mathbf{c}^T\mathbf{x}\right)^2 + \left(\mathbf{s}^T\mathbf{x}\right)^2}$$

$$= \sqrt{\frac{2}{T}}\sqrt{\left(\sqrt{\frac{2}{N}}\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right)\right)^2 + \left(\sqrt{\frac{2}{N}}\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right)\right)^2}$$

$$\hat{\theta}_{ML} = -\arctan\left(\frac{\mathbf{s}^T\mathbf{x}}{\mathbf{c}^T\mathbf{x}}\right) = -\arctan\left(\frac{\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right)}{\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right)}\right)$$

- Note that we are able to implement them only if the frequency F_0 is known.

Estimate of the Parameters of a Signal

$$\mathbf{X} = A \cdot \mathbf{p}(\theta) + \mathbf{W} \quad \text{image of } X(t) \text{ for } t \in [0, T)$$

$$p_i(\theta) = \frac{1}{\sqrt{2B}} \cos\left(2\pi F_0 i + \theta\right)$$

$$= \frac{1}{\sqrt{2B}} \cos\left(2\pi F_0 i\right) \cos\left(\theta\right) - \frac{1}{\sqrt{2B}} \sin\left(2\pi F_0 i\right) \sin\left(\theta\right)$$

$$= \frac{1}{\sqrt{2B}} \sqrt{\frac{N}{2}} \cos\left(\theta\right) \sqrt{\frac{2}{N}} \cos\left(2\pi F_0 i\right) - \frac{1}{\sqrt{2B}} \sqrt{\frac{N}{2}} \sin\left(\theta\right) \sqrt{\frac{2}{N}} \sin\left(2\pi F_0 i\right)$$

$$\overset{N=2BT}{=} \sqrt{\frac{T}{2}} \cos\left(\theta\right) \sqrt{\frac{2}{N}} \cos\left(2\pi F_0 i\right) - \sqrt{\frac{T}{2}} \sin\left(\theta\right) \sqrt{\frac{2}{N}} \sin\left(2\pi F_0 i\right)$$

Estimate of the Parameters of a Signal

$$\mathbf{X} = A \cdot \mathbf{p}(\theta) + \mathbf{W}$$

$$= A\sqrt{\frac{T}{2}}\cos(\theta)\cdot\mathbf{c} - A\sqrt{\frac{T}{2}}\sin(\theta)\cdot\mathbf{s} + \mathbf{W}$$

$$= \alpha_0\mathbf{c} + \alpha_1\mathbf{s} + \mathbf{W}$$

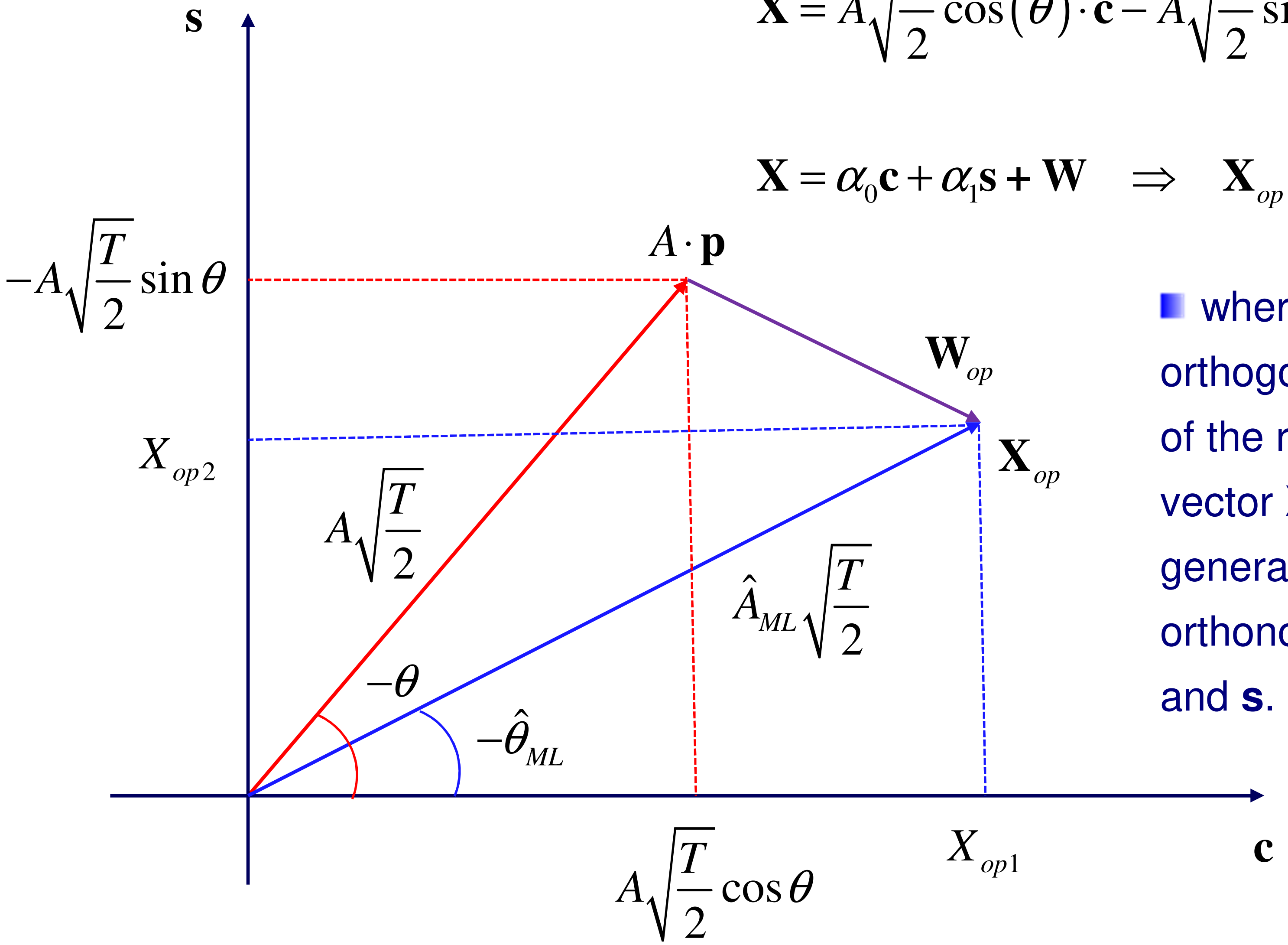
- The useful signal $A\mathbf{p}(\theta)$ is the linear combination of the two orthonormal vectors **c** and **s**.
- Hence, the useful signal belongs to the 2D subspace generated by vectors **c** and **s**. That's why the ML estimator projects the received vector signal **X** onto **c** and **s.**
- α_0 and α_1 are the two principal components, the other *N*-2 components are irrelevant (thanks also to the fact that the noise samples are IID).

Estimate of the Parameters of a Signal

- **Geometrical interpretation of the ML estimators:**

$$\mathbf{X} = A\sqrt{\frac{T}{2}}\cos(\theta)\cdot\mathbf{c} - A\sqrt{\frac{T}{2}}\sin(\theta)\cdot\mathbf{s} + \mathbf{W}$$

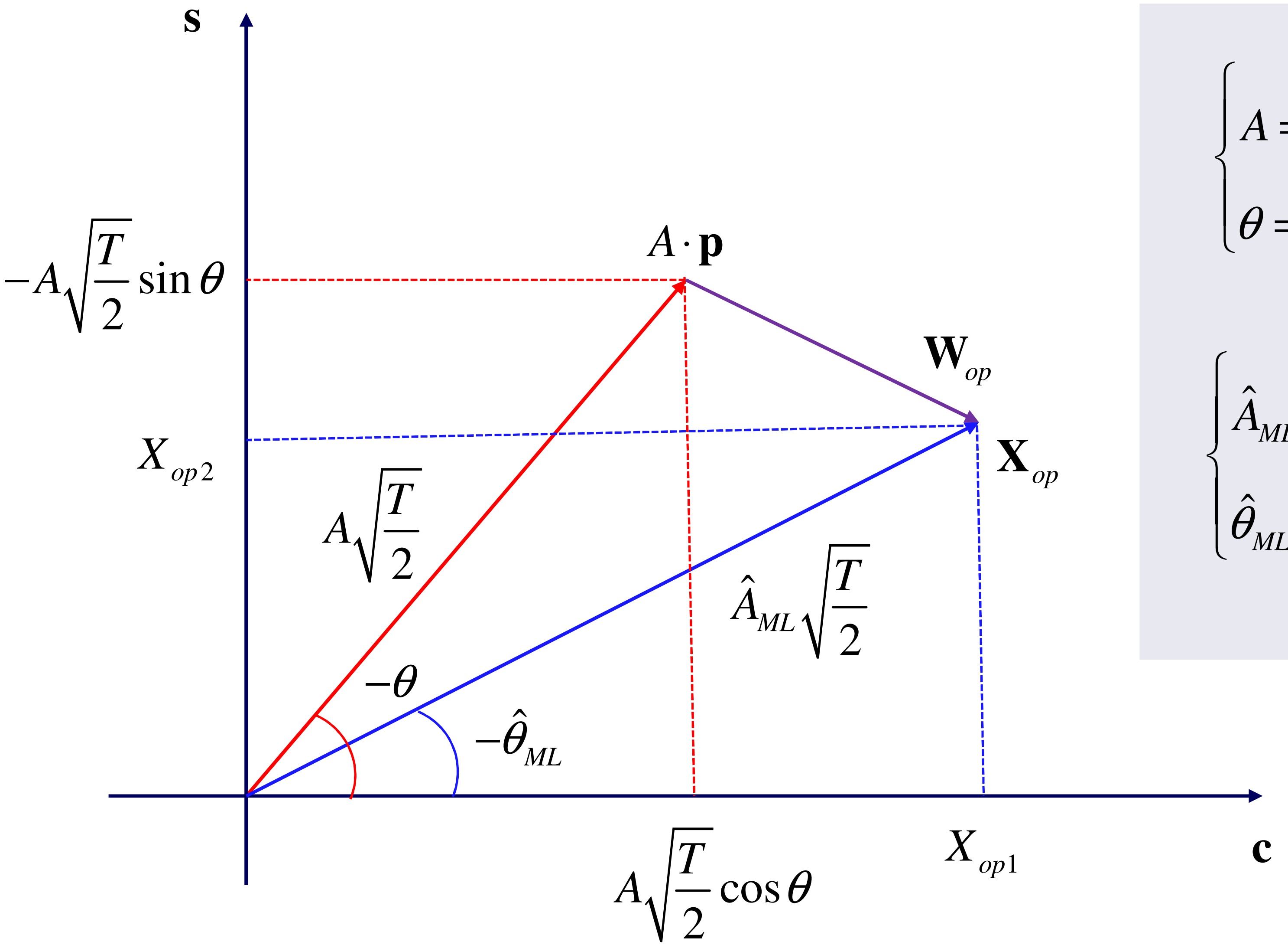
$$\mathbf{X} = \alpha_0\mathbf{c} + \alpha_1\mathbf{s} + \mathbf{W} \quad \Rightarrow \quad \mathbf{X}_{op} = \alpha_0\mathbf{c} + \alpha_1\mathbf{s} + \mathbf{W}_{op}$$



- where $\mathbf{X}_{op}$ is the orthogonal projection of the received *N*-dim vector **X** onto the plane generated by the orthonormal vectors **c** and **s**.

Estimate of the Parameters of a Signal

- **Geometrical interpretation of the ML estimators:**



$$\begin{cases} A = \sqrt{\dfrac{2}{T}}\,\|A\cdot\mathbf{p}\| \\ \theta = -\measuredangle(A\cdot\mathbf{p}) \end{cases}$$

$$\Downarrow$$

$$\begin{cases} \hat{A}_{ML} = \sqrt{\dfrac{2}{T}}\,\|\mathbf{X}_{op}\| \\ \hat{\theta}_{ML} = -\measuredangle\mathbf{X}_{op} \end{cases}$$

Estimate of the Parameters of a Signal

- Another interpretation of the ML estimators as the DTFT of the data vector **x**:

$$Y \triangleq \mathbf{c}^T\mathbf{x} - j\mathbf{s}^T\mathbf{x}$$

$$= \sqrt{\frac{2}{N}}\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right) - j\sqrt{\frac{2}{N}}\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right)$$

$$= \sqrt{\frac{2}{N}}\sum_{i=0}^{N-1} x_i \left[\cos\left(2\pi F_0 i\right) - j\sin\left(2\pi F_0 i\right)\right]$$

$$= \sqrt{\frac{2}{N}}\sum_{i=0}^{N-1} x_i e^{-j2\pi F_0 i} = \sqrt{\frac{2}{N}} \cdot DTFT\left\{\mathbf{x}\right\}\Big|_{F=F_0}$$

$$\hat{A}_{ML} = \sqrt{\frac{2}{T}}\sqrt{\left(\mathbf{c}^T\mathbf{x}\right)^2 + \left(\mathbf{s}^T\mathbf{x}\right)^2} = \sqrt{\frac{2}{T}}\left|\mathbf{c}^T\mathbf{x} - j\mathbf{s}^T\mathbf{x}\right|$$

$$= \sqrt{\frac{2}{T}}\left|Y\right| = \sqrt{\frac{2}{T}}\sqrt{\frac{2}{N}}\left|DTFT\left\{\mathbf{x}\right\}\Big|_{F=F_0}\right|$$

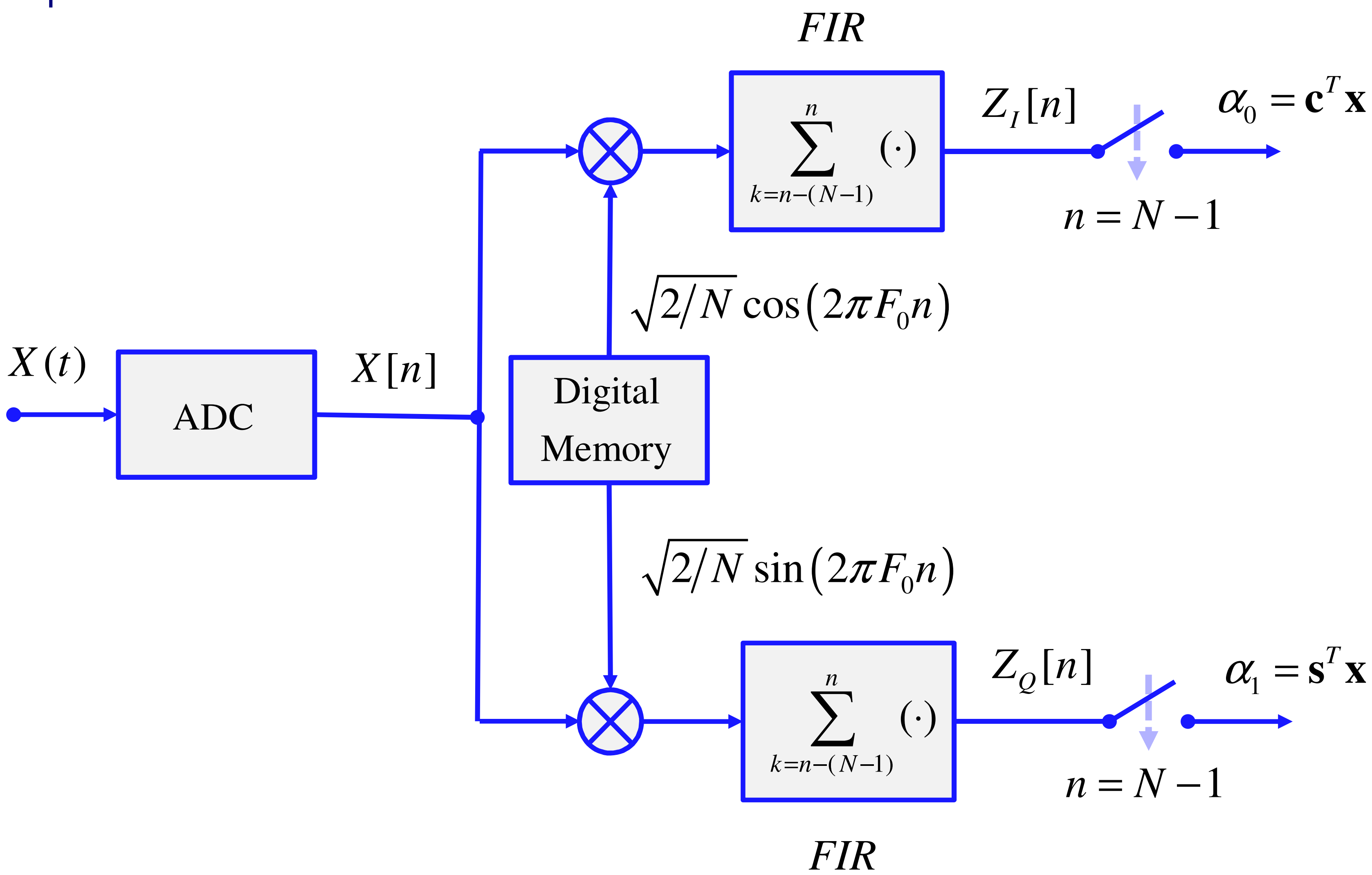
$$\hat{\theta}_{ML} = -\arctan\left(\frac{\mathbf{s}^T\mathbf{x}}{\mathbf{c}^T\mathbf{x}}\right) = \arctan\left(\frac{-\mathbf{s}^T\mathbf{x}}{\mathbf{c}^T\mathbf{x}}\right)$$

$$= \measuredangle Y = \measuredangle DTFT\left\{\mathbf{x}\right\}\Big|_{F=F_0}$$

- Hence, the ML estimators coincide with the Method of Moments (MM) estimators reported at p. 125.

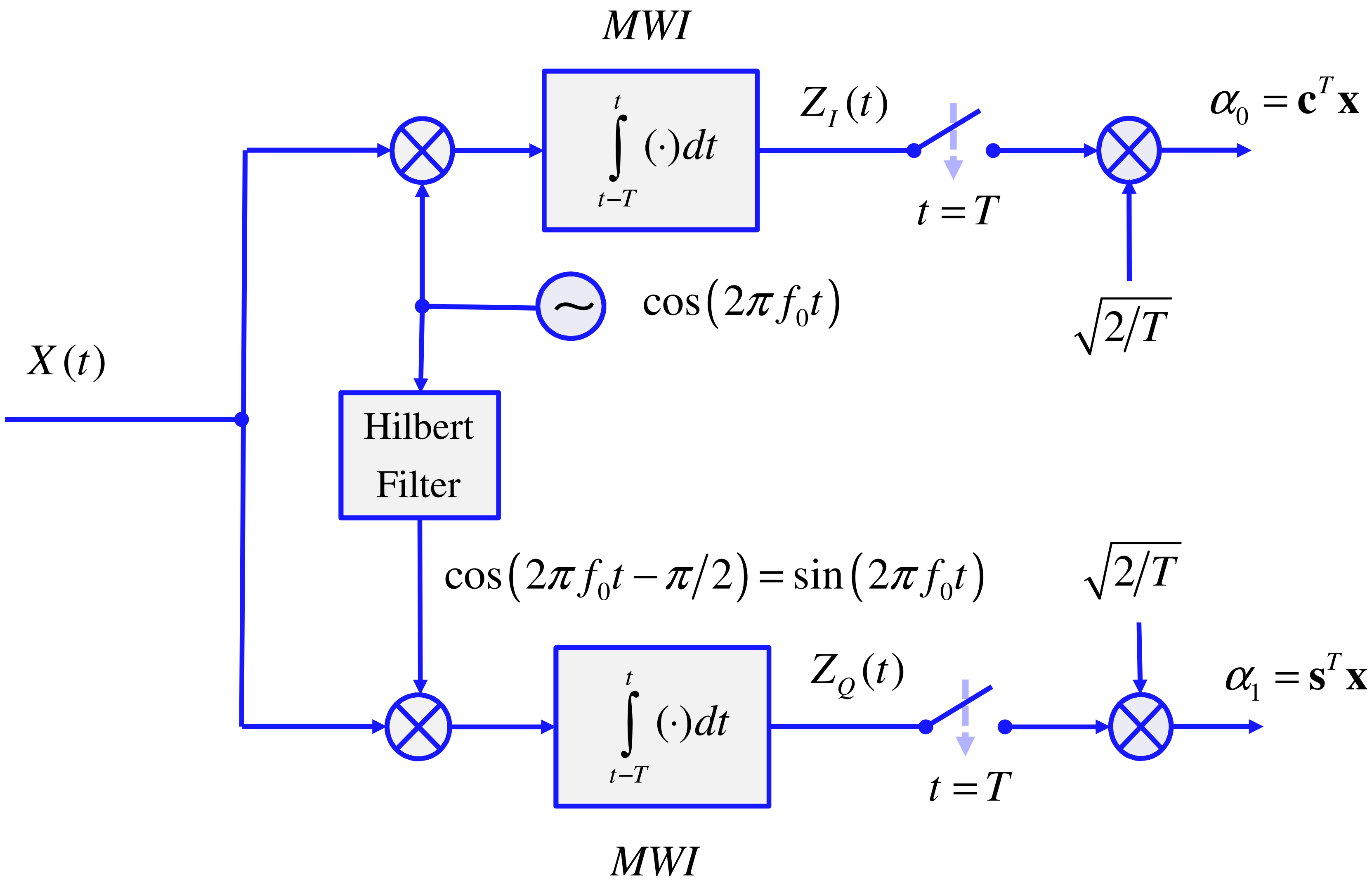
Estimate of the Parameters of a Signal

- The block diagram to implement the estimators is exactly the same as that at p. 63 of Ch. 1.



Estimate of the Parameters of a Signal

- Alternative approach: **correlator** structure that uses an analog **moving window integrator (MWI)**:



- **Performance of the ML estimators**:

$$\hat{A}_{ML} = \sqrt{\frac{2}{T}}\sqrt{\left(\mathbf{c}^T\mathbf{X}\right)^2 + \left(\mathbf{s}^T\mathbf{X}\right)^2} = \sqrt{\left(\sqrt{\frac{2}{T}}\mathbf{c}^T\mathbf{X}\right)^2 + \left(\sqrt{\frac{2}{T}}\mathbf{s}^T\mathbf{X}\right)^2}$$

$$\hat{\theta}_{ML} = -\arctan\left(\frac{\mathbf{s}^T\mathbf{X}}{\mathbf{c}^T\mathbf{X}}\right) = -\arctan\left(\frac{\sqrt{\frac{2}{T}}\mathbf{s}^T\mathbf{X}}{\sqrt{\frac{2}{T}}\mathbf{c}^T\mathbf{X}}\right)$$

- **X** is a Gaussian distributed random vector, so $\mathbf{c}^T\mathbf{X}$ and $\mathbf{s}^T\mathbf{X}$ are two jointly Gaussian r.v.'s:

$$\mathbf{X} = A \cdot \mathbf{p}(\theta) + \mathbf{W} \in \mathcal{N}\left(A \cdot \mathbf{p}(\theta), \sigma_W^2 \mathbf{I}\right)$$

Estimate of the Parameters of a Signal

- Let us derive the joint pdf of $\mathbf{c}^T\mathbf{X}$ and $\mathbf{s}^T\mathbf{X}$:

$$\mathbf{X} = A \cdot \mathbf{p}(\theta) + \mathbf{W} = A\sqrt{\frac{T}{2}}\cos\theta \cdot \mathbf{c} - A\sqrt{\frac{T}{2}}\sin\theta \cdot \mathbf{s} + \mathbf{W}$$

$$\begin{cases} \mathbf{c}^T\mathbf{X} = A\sqrt{\frac{T}{2}}\cos\theta \cdot \mathbf{c}^T\mathbf{c} - A\sqrt{\frac{T}{2}}\sin\theta \cdot \mathbf{c}^T\mathbf{s} + \mathbf{c}^T\mathbf{W} = A\sqrt{\frac{T}{2}}\cos\theta + \mathbf{c}^T\mathbf{W} \\ \mathbf{s}^T\mathbf{X} = A\sqrt{\frac{T}{2}}\cos\theta \cdot \mathbf{s}^T\mathbf{c} - A\sqrt{\frac{T}{2}}\sin\theta \cdot \mathbf{s}^T\mathbf{s} + \mathbf{s}^T\mathbf{W} = -A\sqrt{\frac{T}{2}}\sin\theta + \mathbf{s}^T\mathbf{W} \end{cases}$$

$$E\left\{\mathbf{c}^T\mathbf{X}\right\} = A\sqrt{\frac{T}{2}}\cos\theta, \quad E\left\{\mathbf{s}^T\mathbf{X}\right\} = -A\sqrt{\frac{T}{2}}\sin\theta$$

Estimate of the Parameters of a Signal

$$\mathrm{var}\left\{\mathbf{c}^T\mathbf{X}\right\} = \mathrm{var}\left\{\mathbf{c}^T\mathbf{W}\right\} = E\left\{\mathbf{c}^T\mathbf{W}\left(\mathbf{c}^T\mathbf{W}\right)^T\right\} = E\left\{\mathbf{c}^T\mathbf{W}\mathbf{W}^T\mathbf{c}\right\}$$

$$= \mathbf{c}^T E\left\{\mathbf{W}\mathbf{W}^T\right\}\mathbf{c} = \mathbf{c}^T\left(\sigma_W^2\mathbf{I}\right)\mathbf{c} = \sigma_W^2\mathbf{c}^T\mathbf{c} = \sigma_W^2 = \frac{N_0}{2}$$

$$\mathrm{var}\left\{\mathbf{s}^T\mathbf{X}\right\} = \mathrm{var}\left\{\mathbf{c}^T\mathbf{X}\right\} = \frac{N_0}{2}$$

$$\mathrm{cov}\left\{\mathbf{c}^T\mathbf{X},\mathbf{s}^T\mathbf{X}\right\} = E\left\{\mathbf{c}^T\mathbf{W}\left(\mathbf{s}^T\mathbf{W}\right)^T\right\} = \mathbf{c}^T E\left\{\mathbf{W}\mathbf{W}^T\right\}\mathbf{s} = \mathbf{c}^T\left(\sigma_W^2\mathbf{I}\right)\mathbf{s}$$

$$= \sigma_W^2\mathbf{c}^T\mathbf{s} = 0$$

- Hence, $\mathbf{c}^T\mathbf{X}$ and $\mathbf{s}^T\mathbf{X}$ are uncorrelated, and since they are jointly Gaussian, they are also independent.

Estimate of the Parameters of a Signal

- In summary:

$$\mathbf{c}^T\mathbf{X} \in \mathcal{N}\left(A\sqrt{\frac{T}{2}}\cos\theta, \frac{N_0}{2}\right), \quad \mathbf{s}^T\mathbf{X} \in \mathcal{N}\left(-A\sqrt{\frac{T}{2}}\sin\theta, \frac{N_0}{2}\right),$$

mutually independent

- The ML estimates of A and θ are obtained by applying the **Fundamental Theorem of transformation of random variables** to this transformation of two two jointly Gaussian r.v.'s:
- In practice, it is the well-known transformation from Cartesian to Polar coordinates.

Estimate of the Parameters of a Signal

- Skipping the details of the derivation, the ML estimate of *A* is **Rice** distributed:

$$f_{\hat{A}_{ML}}(a) = \frac{a}{N_0/T}\exp\left(-\frac{a^2+A^2}{2N_0/T}\right)I_0\left(\frac{aA}{N_0/T}\right)u(a)$$

- where $I_0(.)$ is the modified Bessel function of first kind of zero order.
- It is a biparametric pdf which depends on the parameters *A* and N_0/T.
- When ν=*A*=0, we get the Rayleigh distribution, since $I_0(0)=1$:

$$f_{\hat{A}_{ML}}(a) = \frac{a}{N_0/T}\exp\left(-\frac{a^2}{2N_0/T}\right)u(a)$$

- and the pdf of the ML estimate of the phase is uniformly distributed in $[-\pi, \pi)$.

Estimate of the Parameters of a Signal

$$f_{\hat{A}_{ML}}(a) = \frac{a}{N_0/T} \exp\left(-\frac{a^2 + A^2}{2N_0/T} \right) I_0\left(\frac{aA}{N_0/T} \right) u(a)$$

- It can be shown that: $\mathrm{var}\left\{ \hat{A}_{ML} \right\} \propto \frac{N_0}{T}$

- When $A^2T/(2N_0>1$, that means high output Signal-to-Noise power Ratio (SNR_{out}), the Rice pdf tends to a Gaussian pdf.

$$\lim_{T\to\infty} \frac{N_0}{T} = 0 \quad \Rightarrow \quad \lim_{T\to\infty} f_{\hat{A}_{ML}}(a) = \delta(a - A) \quad \text{so the estimator is consistent}$$

$$\text{where} \quad SNR_{out} \triangleq \frac{A^2T}{2N_0}$$

Estimate of the Parameters of a Signal

- The **input Signal to Noise power Ratio** (*SNR_{in}*) is the *SNR* after the digital conversion, i.e. calculated for the image vector ***X***:

$$\mathbf{X} = A \cdot \mathbf{p}(\theta) + \mathbf{W} = A\sqrt{\frac{T}{2}}\cos(\theta)\cdot\mathbf{c} - A\sqrt{\frac{T}{2}}\sin(\theta)\cdot\mathbf{s} + \mathbf{W}$$

$$SNR_{in} = \frac{\left\| A\sqrt{\frac{T}{2}}\cos(\theta)\cdot\mathbf{c} - A\sqrt{\frac{T}{2}}\sin(\theta)\cdot\mathbf{s} \right\|^2}{E\left\{\|\mathbf{W}\|^2\right\}} = \frac{A^2T}{2}\cdot\frac{1}{N\left(N_0/2\right)} = \frac{A^2T}{2N_0}\cdot\frac{2}{N}$$

$$= \frac{A^2T}{2N_0}\cdot\frac{2}{2BT} = \frac{A^2}{2N_0B}$$

$$\left\| A\sqrt{\frac{T}{2}}\cos(\theta)\cdot\mathbf{c} - A\sqrt{\frac{T}{2}}\sin(\theta)\cdot\mathbf{s} \right\|^2 = \left(A\sqrt{\frac{T}{2}}\cos(\theta)\right)^2 + \left(-A\sqrt{\frac{T}{2}}\sin(\theta)\right)^2 = \frac{A^2T}{2}$$

Estimate of the Parameters of a Signal

- The **output Signal to Noise power Ratio** (*SNR*$_{out}$) is the *SNR* after the two matched filters that implement the orthogonal projection of ***X*** onto the plane generated by the vectors **c** and **s**, i.e. onto the signal subspace.

$$\mathbf{X}_{op} = A \cdot \mathbf{p}(\theta) + \mathbf{W}_{op} = A\sqrt{\frac{T}{2}}\cos(\theta)\cdot\mathbf{c} - A\sqrt{\frac{T}{2}}\sin(\theta)\cdot\mathbf{s} + \mathbf{W}_{op}$$

$$SNR_{out} = \frac{\left\| A\sqrt{\frac{T}{2}}\cos(\theta)\cdot\mathbf{c} - A\sqrt{\frac{T}{2}}\sin(\theta)\cdot\mathbf{s} \right\|^2}{E\left\{\left\|\mathbf{W}_{op}\right\|^2\right\}} = \frac{A^2T}{2}\cdot\frac{1}{2\left(N_0/2\right)} = \frac{A^2T}{2N_0}$$

$$E\left\{\left\|\mathbf{W}_{op}\right\|^2\right\} = E\left\{\left(\mathbf{c}^T\mathbf{W}\right)^2 + \left(\mathbf{s}^T\mathbf{W}\right)^2\right\} = 2\cdot\frac{N_0}{2}$$

Estimate of the Parameters of a Signal

- The **Processing Gain (PG)** is immediately obtained:

$$SNR_{in} = \frac{A^2}{2BN_0}, \quad SNR_{out} = \frac{A^2 T}{2N_0} \quad \rightarrow \quad PG = \frac{SNR_{out}}{SNR_{in}} = BT = \frac{N}{2}$$

- Note that the PG is one half the PG in the previous case, when the signal shape was perfectly known, i.e. known θ.
- This should not be a surprise since now the Signal of Interest (SOI) belongs to a 2D subspace instead of a 1D subspace, so there is more uncertainty about the SOI.

Estimate of the Parameters of a Signal

- The **Rice** distribution with parameters (A,N_0/T) has moments of even order that can be expressed as a polynomials in A and N_0/T, e.g. the 2nd order moment:

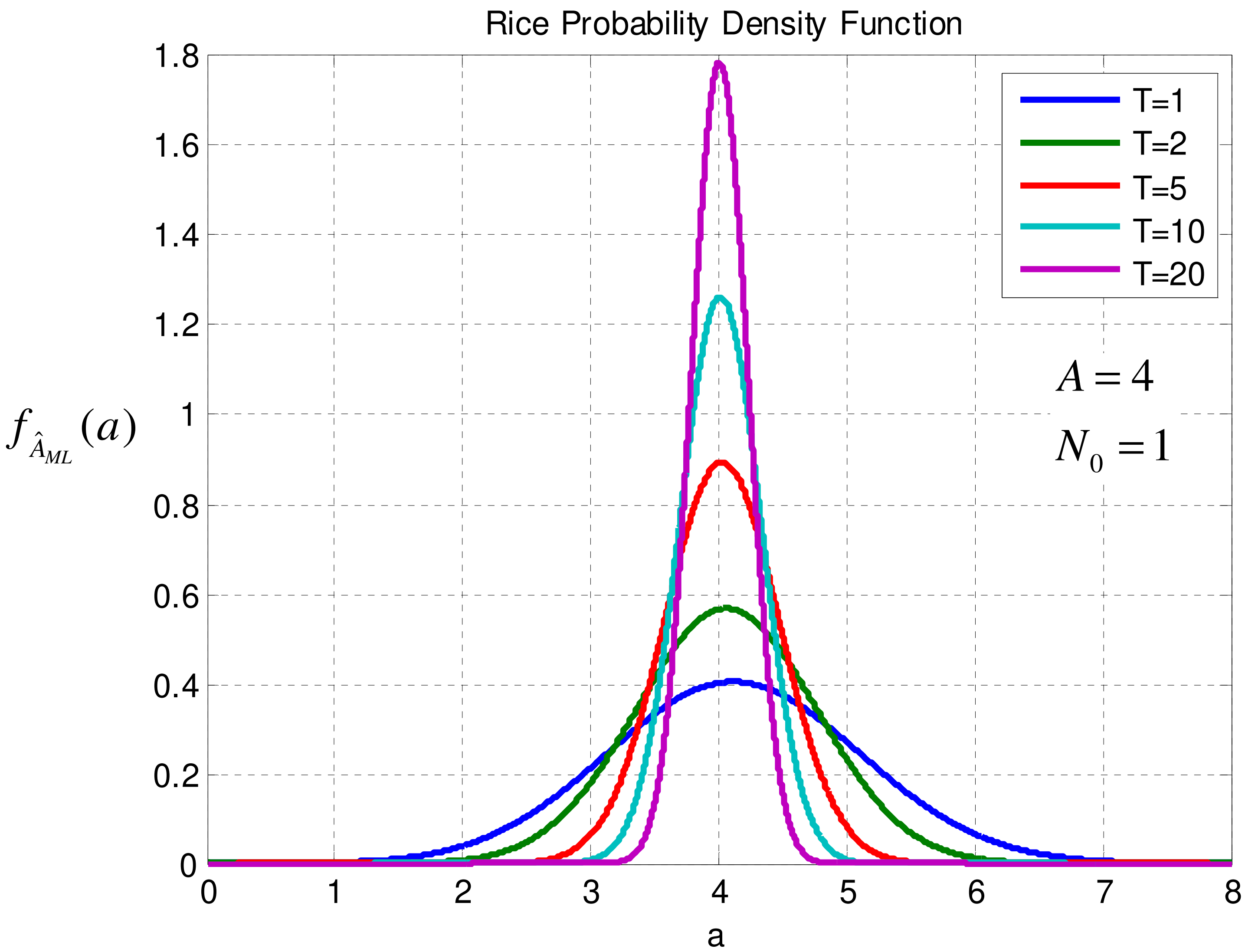
$$E\left\{\hat{A}_{ML}^2\right\} = A^2 + \frac{2N_0}{T} = A^2\left(1 + \frac{1}{SNR_{out}}\right), \quad \text{where} \quad SNR_{out} \triangleq \frac{A^2 T}{2N_0} = \frac{N}{2} SNR_{in}$$

- The odd order moments cannot be expressed in closed form; they can be expressed as a function of the generalized Laguerre polynomials, e.g. the mean:

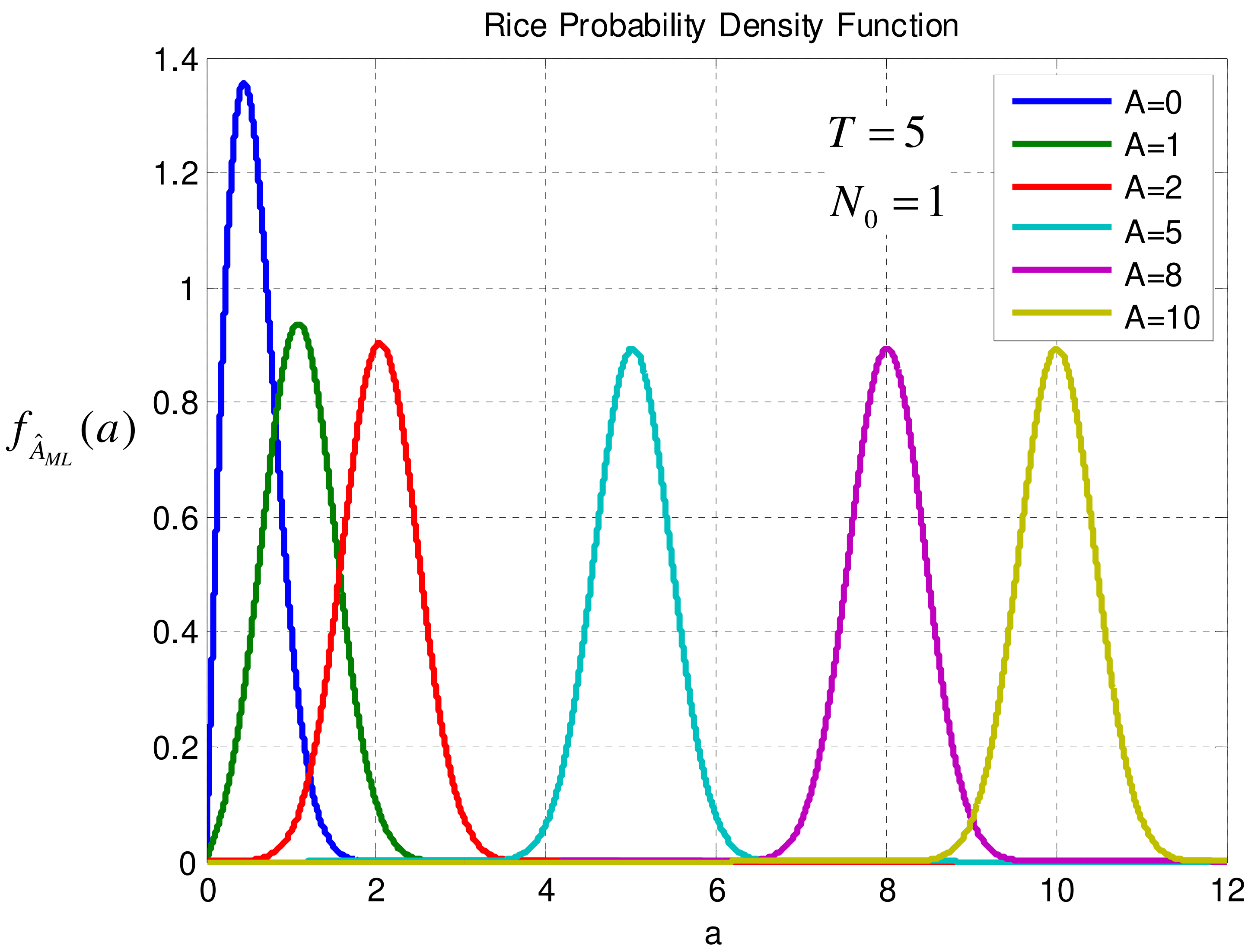
$$E\left\{\hat{A}_{ML}\right\} = \sqrt{\frac{\pi}{2}} \sigma L_{1/2}\left(-\frac{\nu^2}{2\sigma^2}\right) \neq A$$

- Hence, it is not possible to calculate bias and MSE in closed form.

Estimate of the Parameters of a Signal



Estimate of the Parameters of a Signal



- **Cramér-Rao lower Bound (CRB)** on amplitude and phase and comparison of the MSE (derived by Monte Carlo simulation) with the CRB:

$$\mathbf{C}_{\boldsymbol{\theta}} = E\left\{\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^T\right\} = E\left\{\begin{bmatrix}\hat{A}-A \\ \hat{\theta}-\theta\end{bmatrix}\begin{bmatrix}\hat{A}-A \\ \hat{\theta}-\theta\end{bmatrix}^T\right\}$$

$$= \begin{bmatrix} E\left\{(\hat{A}-A)^2\right\} & E\left\{(\hat{A}-A)(\hat{\theta}-\theta)\right\} \\ E\left\{(\hat{A}-A)(\hat{\theta}-\theta)\right\} & E\left\{(\hat{\theta}-\theta)^2\right\} \end{bmatrix}$$

$$\mathbf{C}_{\boldsymbol{\theta}} \geq CRB(A,\theta) = \mathbf{I}^{-1}(A,\theta) = \begin{bmatrix} -E\left\{\dfrac{\partial^2 \ln L(A,\theta)}{\partial A^2}\right\} & -E\left\{\dfrac{\partial^2 \ln L(A,\theta)}{\partial A \partial \theta}\right\} \\ -E\left\{\dfrac{\partial^2 \ln L(A,\theta)}{\partial \theta \partial A}\right\} & -E\left\{\dfrac{\partial^2 \ln L(A,\theta)}{\partial \theta^2}\right\} \end{bmatrix}^{-1}$$

Estimate of the Parameters of a Signal

- We have already derived the first order derivatives of the log-likelihood function with respect to the unknown parameters *A* and θ:

$$\begin{cases} \dfrac{\partial \ln L(A,\theta)}{\partial A} = -\dfrac{AN}{N_0 2B} + \dfrac{2}{N_0\sqrt{2B}} \displaystyle\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \theta\right) \\ \\ \dfrac{\partial \ln L(A,\theta)}{\partial \theta} = -\dfrac{2A}{N_0\sqrt{2B}} \displaystyle\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i + \theta\right) \end{cases}$$

- The 2nd order derivatives and the **Fisher Information Matrix (FIM)** elements are:

$$\frac{\partial^2 \ln L(A,\theta)}{\partial A^2} = -\frac{N}{N_0 2B} = -\frac{T}{N_0} \quad \rightarrow \quad I_{11} = -E\left\{\frac{\partial^2 \ln L(A,\theta)}{\partial A^2}\right\} = \frac{T}{N_0}$$

$$\frac{\partial^2 \ln L(A,\theta)}{\partial A \partial \theta} = \frac{\partial^2 \ln L(A,\theta)}{\partial \theta \partial A} = -\frac{2}{N_0\sqrt{2B}} \sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i + \theta\right)$$

$$= -\frac{2}{N_0\sqrt{2B}} \left[\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right)\cos\left(\theta\right) + \sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right)\sin\left(\theta\right) \right]$$

$$= -\frac{2}{N_0\sqrt{2B}} \sqrt{\frac{N}{2}} \left[\mathbf{s}^T \mathbf{X} \cos\left(\theta\right) + \mathbf{c}^T \mathbf{X} \sin\left(\theta\right) \right]$$

$$I_{21} = I_{12} = -E\left\{ \frac{\partial^2 \ln L(A,\theta)}{\partial A \partial \theta} \right\} = \frac{2}{N_0\sqrt{2B}} \sqrt{\frac{N}{2}} \left[E\left\{\mathbf{s}^T \mathbf{X}\right\} \cos\left(\theta\right) + E\left\{\mathbf{c}^T \mathbf{X}\right\} \sin\left(\theta\right) \right]$$

$$= \frac{2}{N_0\sqrt{2B}} \sqrt{\frac{N}{2}} \left[-A\sqrt{\frac{T}{2}} \sin\left(\theta\right)\cos\left(\theta\right) + A\sqrt{\frac{T}{2}} \cos\left(\theta\right)\sin\left(\theta\right) \right] = 0$$

where we used the results: $E\left\{\mathbf{c}^T \mathbf{X}\right\} = A\sqrt{\frac{T}{2}}\cos\left(\theta\right), \quad E\left\{\mathbf{s}^T \mathbf{X}\right\} = -A\sqrt{\frac{T}{2}}\sin\left(\theta\right)$

$$\frac{\partial^2 \ln L(A,\theta)}{\partial \theta^2} = -\frac{2A}{N_0\sqrt{2B}}\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \theta\right)$$

$$= -\frac{2A}{N_0\sqrt{2B}}\left[\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right)\cos\left(\theta\right) - \sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right)\sin\left(\theta\right)\right]$$

$$= -\frac{2A}{N_0\sqrt{2B}}\sqrt{\frac{N}{2}}\left[\mathbf{c}^T\mathbf{X}\cos\left(\theta\right) - \mathbf{s}^T\mathbf{X}\sin\left(\theta\right)\right]$$

$$I_{22} = -E\left\{\frac{\partial^2 \ln L(A,\theta)}{\partial \theta^2}\right\} = \frac{2A}{N_0\sqrt{2B}}\sqrt{\frac{N}{2}}\left[E\left\{\mathbf{c}^T\mathbf{X}\right\}\cos\left(\theta\right) - E\left\{\mathbf{s}^T\mathbf{X}\right\}\sin\left(\theta\right)\right]$$

$$= \frac{2A}{N_0\sqrt{2B}}\sqrt{\frac{N}{2}}\left[A\sqrt{\frac{T}{2}}\cos\left(\theta\right)\cos\left(\theta\right) + A\sqrt{\frac{T}{2}}\sin\left(\theta\right)\sin\left(\theta\right)\right]$$

$$= \frac{2A}{N_0\sqrt{2B}}\sqrt{\frac{N}{2}}A\sqrt{\frac{T}{2}} = \frac{2A^2}{N_0\sqrt{2B}}\sqrt{\frac{2BT}{2}}\sqrt{\frac{T}{2}} = \frac{A^2T}{N_0}$$

Estimate of the Parameters of a Signal

- Hence, the FIM is diagonal:

$$\mathbf{I}(A,\theta)=\begin{bmatrix} \dfrac{T}{N_0} & 0 \\ 0 & \dfrac{TA^2}{N_0} \end{bmatrix} \Rightarrow \mathbf{I}^{-1}(A,\theta)=\begin{bmatrix} \dfrac{N_0}{T} & 0 \\ 0 & \dfrac{N_0}{TA^2} \end{bmatrix}$$

- The CRB's on the estimation of amplitude and phase are:

$$\mathrm{var}\left\{\hat{A}\right\}\geq CRB(A)=\frac{N_0}{T}$$

$$\mathrm{var}\left\{\hat{\theta}\right\}\geq CRB(\theta)=\frac{N_0}{A^2T}=\frac{N_0}{2E_s}$$

$$E_s=A^2E_p=\frac{A^2T}{2}$$

Estimate of the Parameters of a Signal

- The CRB's can be expressed in terms of SNR_{in} *or* SNR_{out}:

$$SNR_{in} = \frac{A^2}{2BN_0}, \qquad SNR_{out} = \frac{A^2 T}{2N_0} = \frac{N}{2} SNR_{in}$$

- The two CRBs are given by:

$$\begin{cases} \operatorname{var}\left\{\hat{A}\right\} \geq CRB(A) = \dfrac{N_0}{T} = \dfrac{A^2}{2 \cdot SNR_{out}} = \dfrac{A^2}{N \cdot SNR_{in}} \\ \\ \operatorname{var}\left\{\hat{\theta}\right\} \geq CRB(\theta) = \dfrac{N_0}{A^2 T} = \dfrac{N_0}{2E_s} = \dfrac{1}{2 \cdot SNR_{out}} = \dfrac{1}{N \cdot SNR_{in}} \end{cases}$$

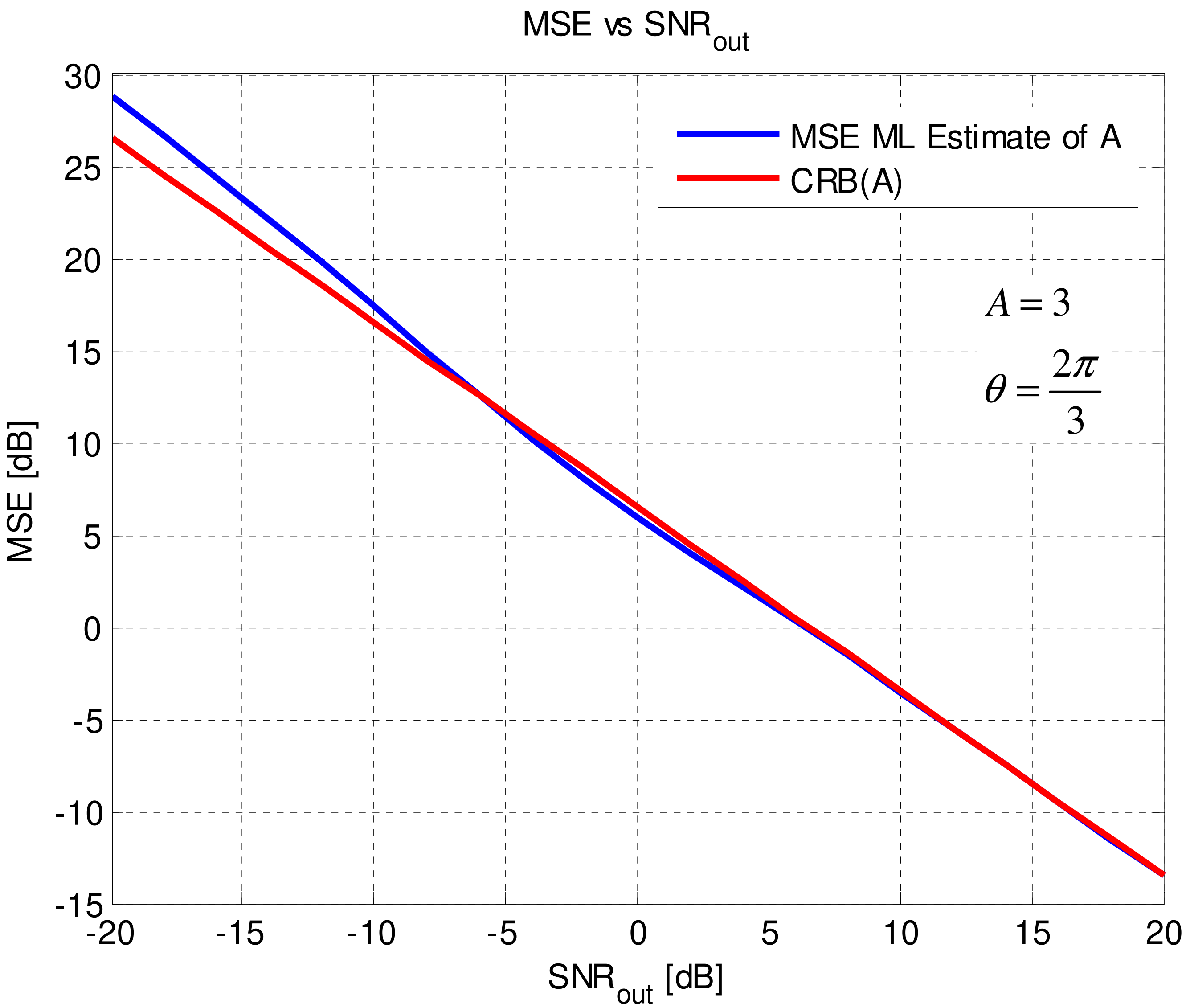
Estimate of the Parameters of a Signal

- Due to the asymptotic properties of all ML estimators (under some mild conditions), we have that when the observation interval T goes to infinity, i.e. $T \to \infty$, or equivalently $SNR_{out} \to \infty$, we have:

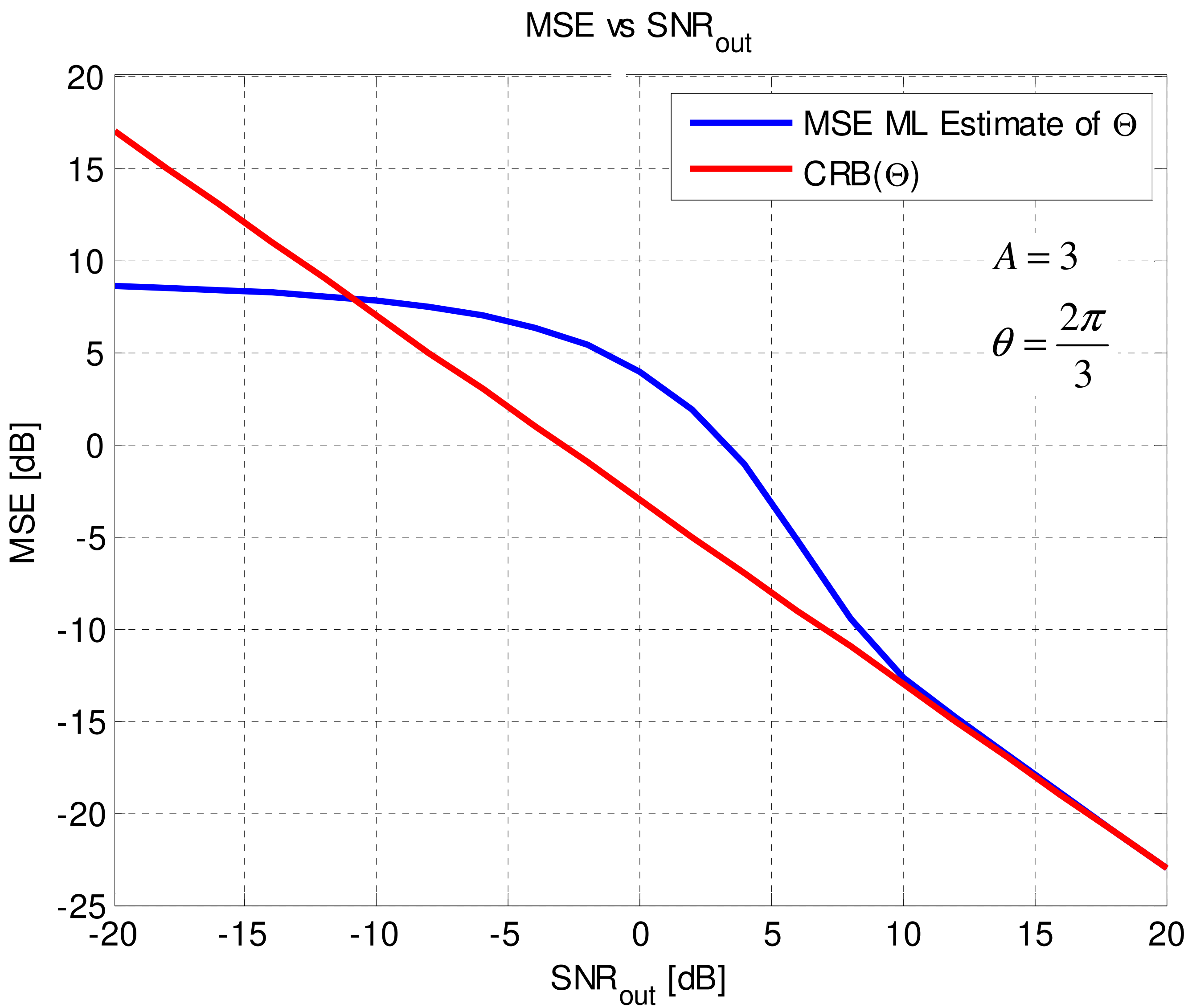
$$\begin{cases} \hat{A}_{ML} \overset{a}{\in} \mathcal{N}\left(A, \dfrac{N_0}{T}\right) \equiv \mathcal{N}\left(A, \dfrac{A^2}{2 \cdot SNR_{out}}\right) \\ \\ \hat{\theta}_{ML} \overset{a}{\in} \mathcal{N}\left(\theta, \dfrac{N_0}{A^2 T}\right) \equiv \mathcal{N}\left(\theta, \dfrac{1}{2 \cdot SNR_{out}}\right) \end{cases}$$

- i.e. the estimators are asymptotically efficient, asymptotically unbiased and consistent, since MSE$\to 0$ when $T \to \infty$ (i.e. $SNR_{out} \to \infty$).

Estimate of the Parameters of a Signal



Estimate of the Parameters of a Signal



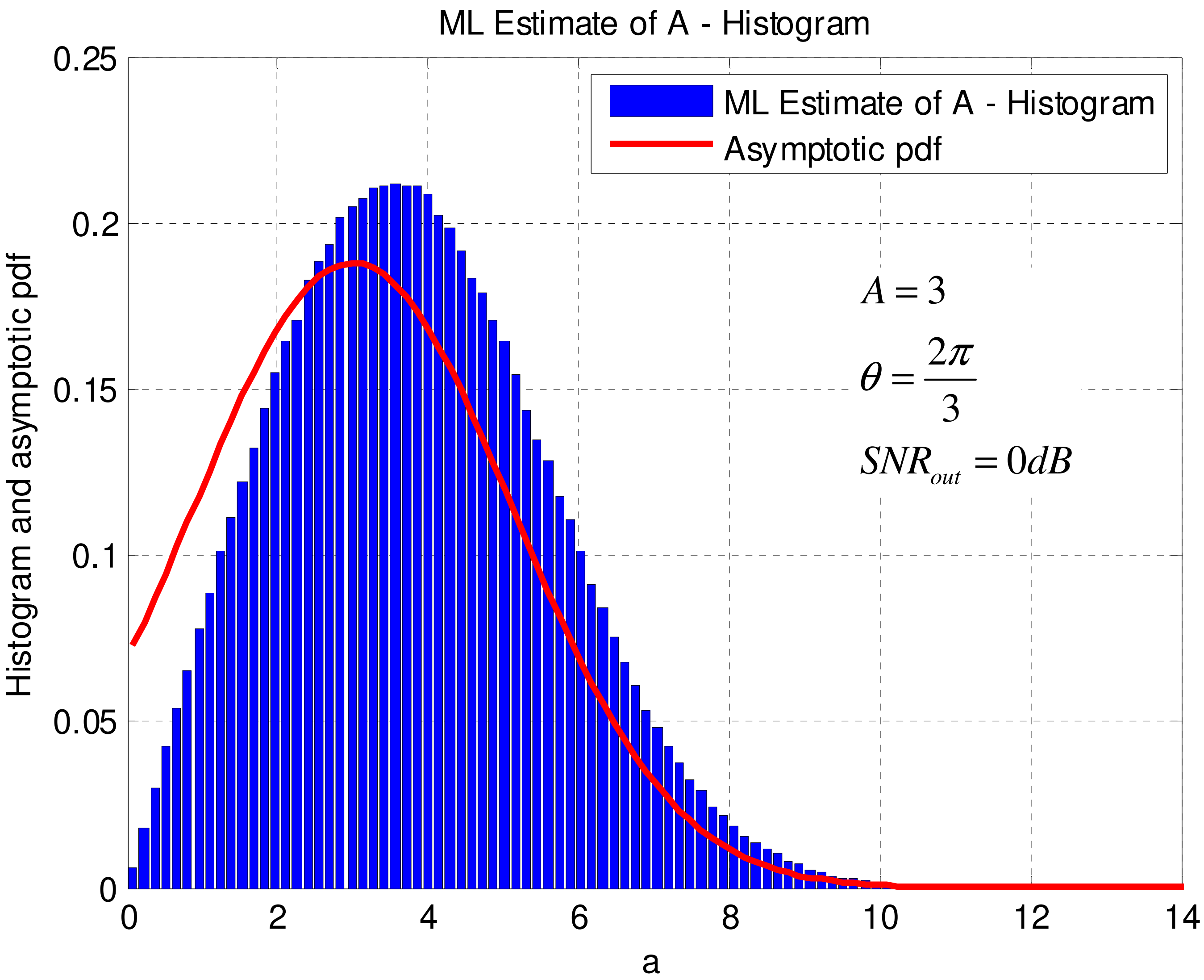
Estimate of the Parameters of a Signal

- Note that the plot is in dB on both scales, that's why the CRB curves are lines with slope -1:

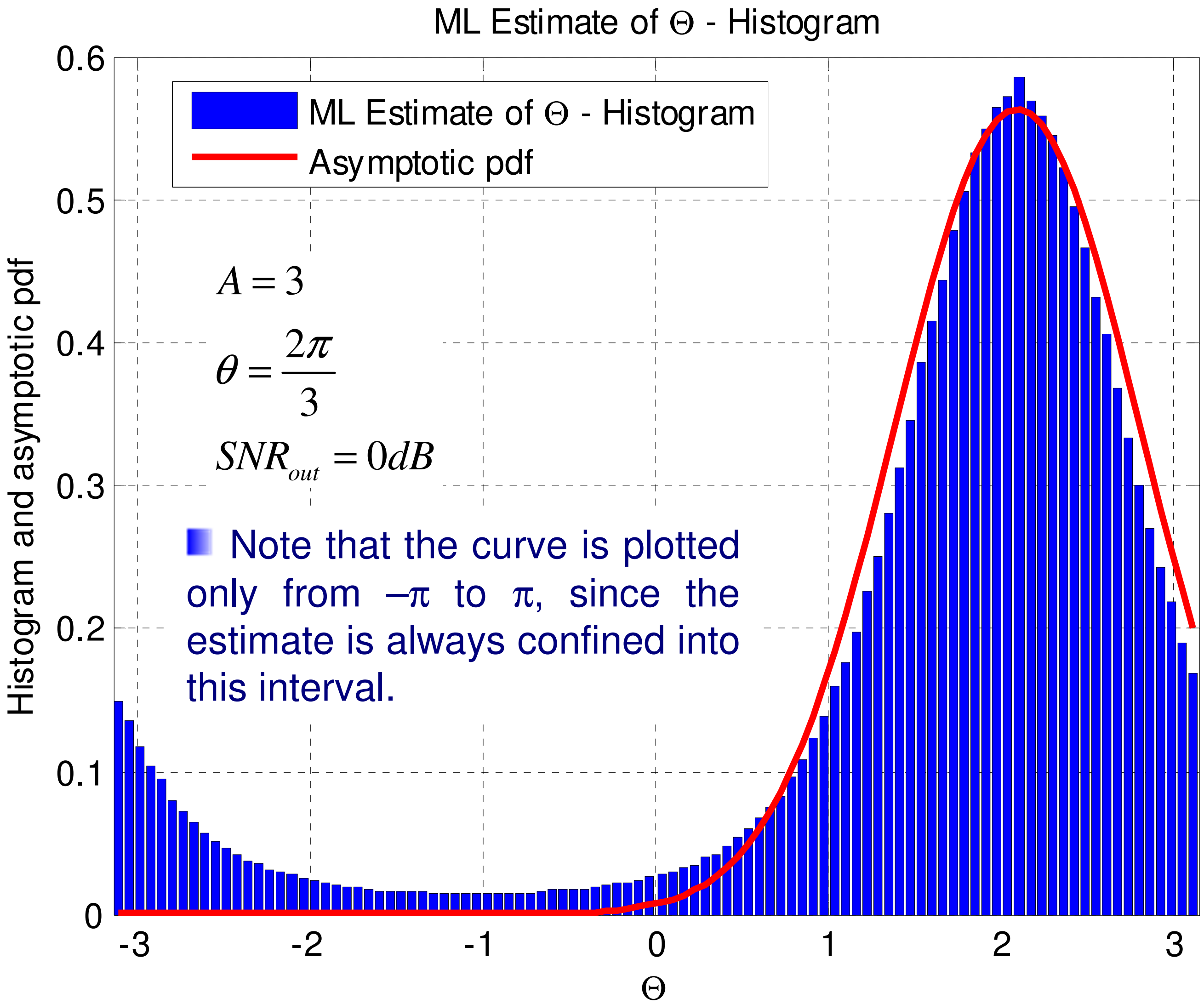
$$CRB(A)_{dB} = 10Log_{10}CRB(A) = 10Log_{10}\left(\frac{A^2}{2SNR_{out}}\right)$$
$$= 20Log_{10}\left(A\right) - 10Log_{10}\left(2\right) - 10Log_{10}\left(SNR_{out}\right)$$
$$= 20Log_{10}\left(A\right) - 10Log_{10}\left(2\right) - SNR_{out\,dB}$$

$$CRB(\theta)_{dB} = 10Log_{10}CRB(\theta) = 10Log_{10}\left(\frac{1}{2SNR_{out}}\right)$$
$$= -10Log_{10}\left(2\right) - 10Log_{10}\left(SNR_{out}\right)$$
$$= -10Log_{10}\left(2\right) - SNR_{out\,dB}$$

Estimate of the Parameters of a Signal

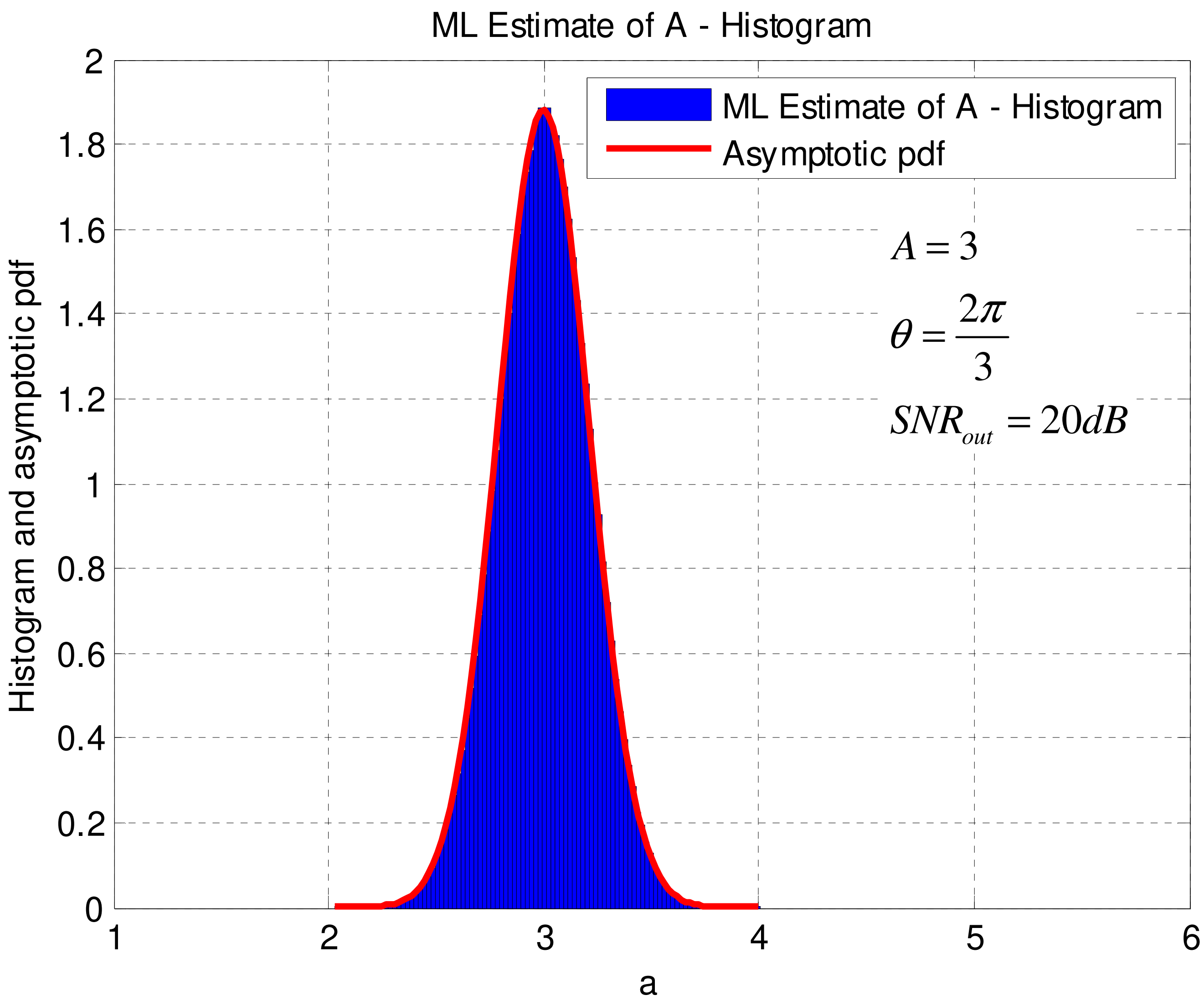


Estimate of the Parameters of a Signal

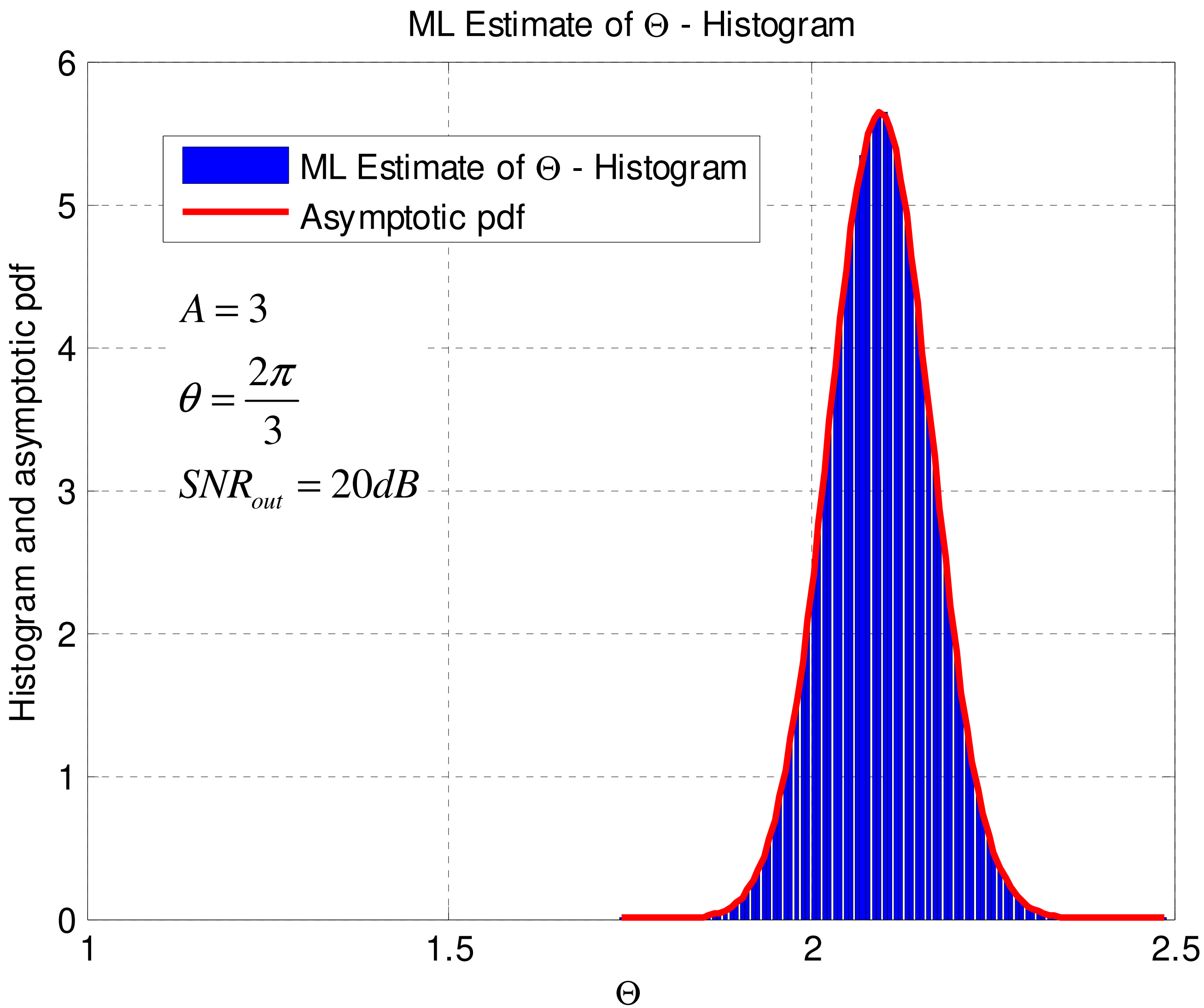


- Note that the curve is plotted only from $-\pi$ to π, since the estimate is always confined into this interval.

Estimate of the Parameters of a Signal



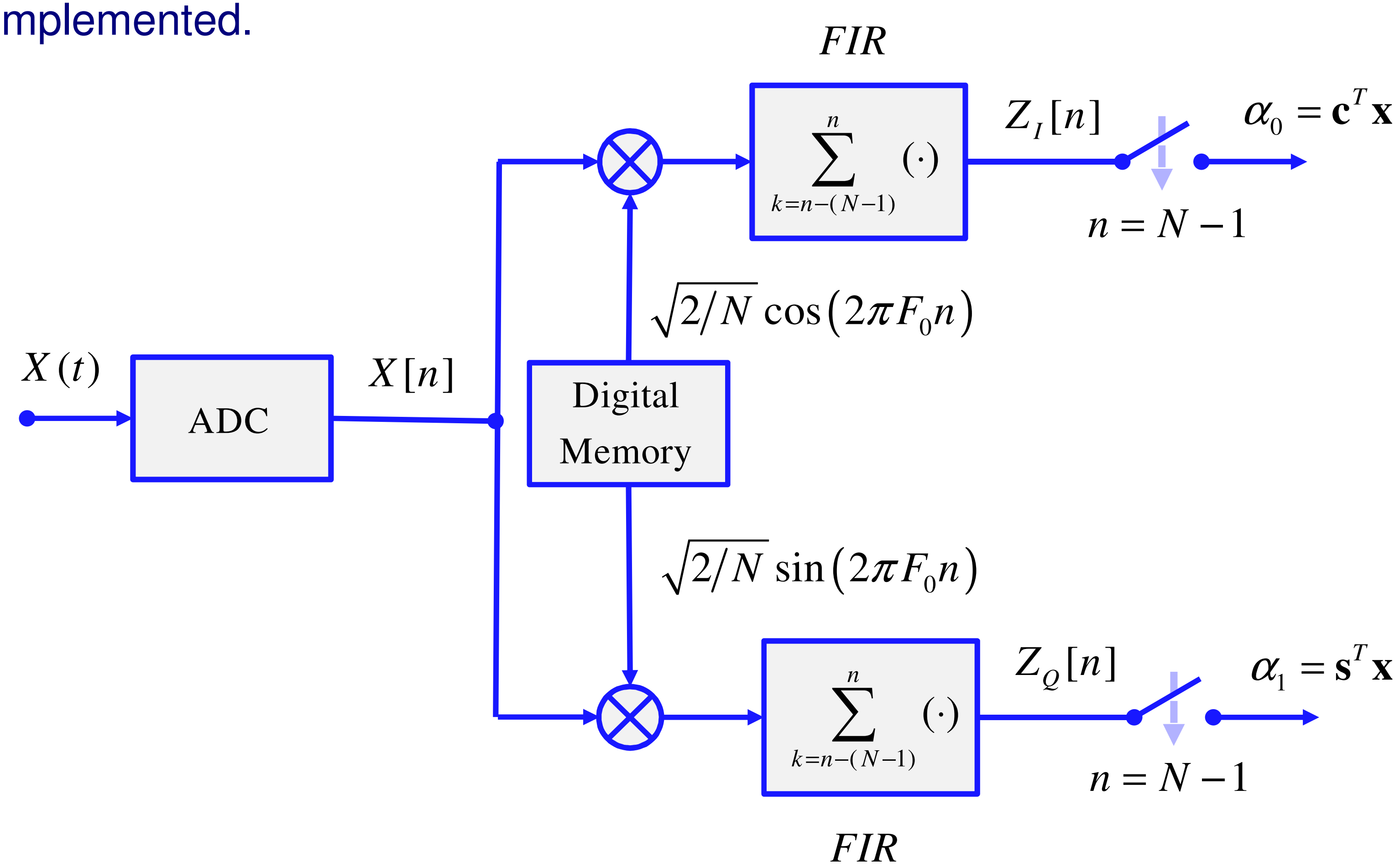
Estimate of the Parameters of a Signal



ML Estimate of the Parameters of a Signal: Amplitude, Phase, and Frequency

Estimate of the Parameters of a Signal

- Consider now the case when all the signal parameters (amplitude, phase and frequency) are unknown.
- If the frequency F_0 is also unknown, the previously derived algorithm cannot be implemented.



Estimate of the Parameters of a Signal

- The observed data vector **X** is exactly the same as before:

$$\mathbf{X} \;\; \text{image of} \;\; X(t) \;\; \text{for} \;\; t \in [0,T)$$

$$\mathbf{X} = \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{bmatrix} = A \cdot \mathbf{p}(\theta) + \mathbf{W} = \begin{bmatrix} \dfrac{A}{\sqrt{2B}}\cos(\theta) \\ \dfrac{A}{\sqrt{2B}}\cos(2\pi F_0 + \theta) \\ \vdots \\ \dfrac{A}{\sqrt{2B}}\cos(2\pi F_0 (N-1) + \theta) \end{bmatrix} + \begin{bmatrix} W_0 \\ W_1 \\ \vdots \\ W_{N-1} \end{bmatrix}$$

$$\{W_i\}_{i=0}^{N-1} \;\; \text{IID}, \;\; W_i \in \mathcal{N}\left(0, \sigma_W^2\right), \quad \sigma_W^2 = \text{var}\{W_i\} = \frac{N_0}{2}, \quad X_i \in \mathcal{N}\left(A \cdot p_i(\theta), \sigma_W^2\right)$$

$$F_0 \triangleq \frac{f_0}{2B} \;\; \text{Digital Frequency - Nyquist Sampling} \quad \rightarrow \quad 0 < F_0 < \frac{1}{2}$$

Estimate of the Parameters of a Signal

- The **log-likelihood function (log-LF)** is also the same as before:

$$\ln L(A,\theta,F_0) = \ln f_{\mathbf{X}}(\mathbf{x};A,\theta,F_0) = \kappa - \frac{A^2}{N_0 2B}\cdot\frac{N}{2} + \frac{2A}{N_0\sqrt{2B}}\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \theta\right)$$

- The joint ML estimators of A, θ, and F_0 are obtained by jointly solving the likelihood equations, which are obtained by deriving the log-LF with respect to the three unknown parameters:

$$\begin{cases} \dfrac{\partial \ln L(A,\theta,F_0)}{\partial A} = 0 \\ \dfrac{\partial \ln L(A,\theta,F_0)}{\partial \theta} = 0 \\ \dfrac{\partial \ln L(A,\theta,F_0)}{\partial F_0} = 0 \end{cases}$$

- We have already solved the first two equations and derived the ML estimators of amplitude and phase when the signal frequency F_0 is known.

- Note that the ML estimators of amplitude and phase are functions of F_0:

$$\begin{cases} \hat{A}_{ML}(F_0) = \sqrt{\dfrac{2}{T}} \sqrt{\left(\sqrt{\dfrac{2}{N}} \displaystyle\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right) \right)^2 + \left(\sqrt{\dfrac{2}{N}} \displaystyle\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right) \right)^2} \\ \hat{\theta}_{ML}(F_0) = -\arctan\left(\dfrac{\sqrt{\dfrac{2}{N}} \displaystyle\sum_{i=0}^{N-1} x_i \sin\left(2\pi F_0 i\right)}{\sqrt{\dfrac{2}{N}} \displaystyle\sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i\right)} \right) \end{cases}$$

- They are basically obtained by the modulo and phase of the data vector DTFT calculated at the digital frequency F_0:

$$\begin{cases} \hat{A}_{ML}(F_0) = \sqrt{\dfrac{2}{T}} \sqrt{\dfrac{2}{N}} \left| DTFT\{\mathbf{x}\}\big|_{F=F_0} \right| \\ \hat{\theta}_{ML}(F_0) = \measuredangle\, DTFT\{\mathbf{x}\}\big|_{F=F_0} \end{cases}$$

Estimate of the Parameters of a Signal

- We can insert the ML estimators of A and θ in the log-LF and look for the maximum of the **"compressed" log-LF** with respect to the last unknown F_0 :

$$C(F_0) \triangleq \ln L\left(\hat{A}_{ML}(F_0), \hat{\theta}_{ML}(F_0), F_0\right)$$

$$= \kappa - \frac{\left(\hat{A}_{ML}(F_0)\right)^2}{N_0 2B} \cdot \frac{N}{2} + \frac{2\hat{A}_{ML}(F_0)}{N_0\sqrt{2B}} \sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \hat{\theta}_{ML}(F_0)\right)$$

$$\Downarrow$$

$$\hat{F}_{0ML} = \arg\max_{F_0} C(F_0) = \arg\max_{F_0} \ln L\left(\hat{A}_{ML}(F_0), \hat{\theta}_{ML}(F_0), F_0\right)$$

$$= \arg\max_{F_0} \left[-\frac{\left(\hat{A}_{ML}(F_0)\right)^2}{N_0 2B} \cdot \frac{N}{2} + \frac{2\hat{A}_{ML}(F_0)}{N_0\sqrt{2B}} \sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \hat{\theta}_{ML}(F_0)\right) \right]$$

- If we now recall that (see the derivative w.r.t. *A* at page 131):

$$\hat{A}_{ML} = \sqrt{2B} \cdot \frac{2}{N} \sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \hat{\theta}_{ML}\right) \quad \rightarrow \quad \sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \hat{\theta}_{ML}\right) = \frac{N}{2\sqrt{2B}} \hat{A}_{ML}$$

$$\hat{F}_{0ML} = \arg\max_{F_0} \left[-\frac{\left(\hat{A}_{ML}(F_0)\right)^2}{N_0 2B} \cdot \frac{N}{2} + \frac{2\hat{A}_{ML}(F_0)}{N_0\sqrt{2B}} \sum_{i=0}^{N-1} x_i \cos\left(2\pi F_0 i + \hat{\theta}_{ML}(F_0)\right) \right]$$

$$= \arg\max_{F_0} \left[-\frac{\left(\hat{A}_{ML}(F_0)\right)^2}{N_0 2B} \cdot \frac{N}{2} + \frac{2\hat{A}_{ML}(F_0)}{N_0\sqrt{2B}} \frac{N}{2\sqrt{2B}} \hat{A}_{ML}(F_0) \right]$$

$$= \arg\max_{F_0} \left[\frac{N}{2} \cdot \frac{\left(\hat{A}_{ML}(F_0)\right)^2}{N_0 2B} \right] = \arg\max_{F_0} \left(\hat{A}_{ML}(F_0)\right)^2$$

Estimate of the Parameters of a Signal

- Recalling that the ML estimate of the amplitude is the modulo of the observed data vector DTFT:

$$\hat{A}_{ML}(F_0) = \sqrt{\frac{2}{T}}\sqrt{\frac{2}{N}}\left| DTFT\{\mathbf{x}\}\Big|_{F=F_0} \right| = \sqrt{\frac{2}{T}}\sqrt{\frac{2}{N}}\left| \sum_{i=0}^{N-1} x_i e^{-j2\pi F_0 i} \right|$$

$$\hat{F}_{0ML} = \arg\max_{F_0} \left(\hat{A}_{ML}(F_0) \right)^2 = \arg\max_{F_0} \underbrace{\frac{1}{N}\left| \sum_{i=0}^{N-1} x_i e^{-j2\pi F_0 i} \right|^2}_{Periodogram}$$

- Hence, F_0 is estimated as the location of the maximum of the observed data **Periodogram**. Once we have an estimate of F_0, A and θ are estimated as:

$$\begin{cases} \hat{A}_{ML} = \sqrt{\frac{2}{T}}\sqrt{\frac{2}{N}}\left| DTFT\{\mathbf{x}\}\Big|_{F=\hat{F}_{0ML}} \right| \\ \hat{\theta}_{ML} = \measuredangle DTFT\{\mathbf{x}\}\Big|_{F=\hat{F}_{0ML}} \end{cases}$$

Estimate of the Parameters of a Signal

- The **Periodogram** is calculated via the efficient Fast Fourier Transform (FFT) with **zero-padding** to N_{zp} points:

$$\hat{F}_{0ML} = \arg\max_{F_0} \underbrace{\frac{1}{N}\left|\sum_{i=0}^{N_{zp}-1} x_i e^{-j2\pi F_0 i}\right|^2}_{Periodogram} = \arg\max_{F_0} \frac{1}{N}\left|FFT\{\mathbf{x}\}\right|^2$$

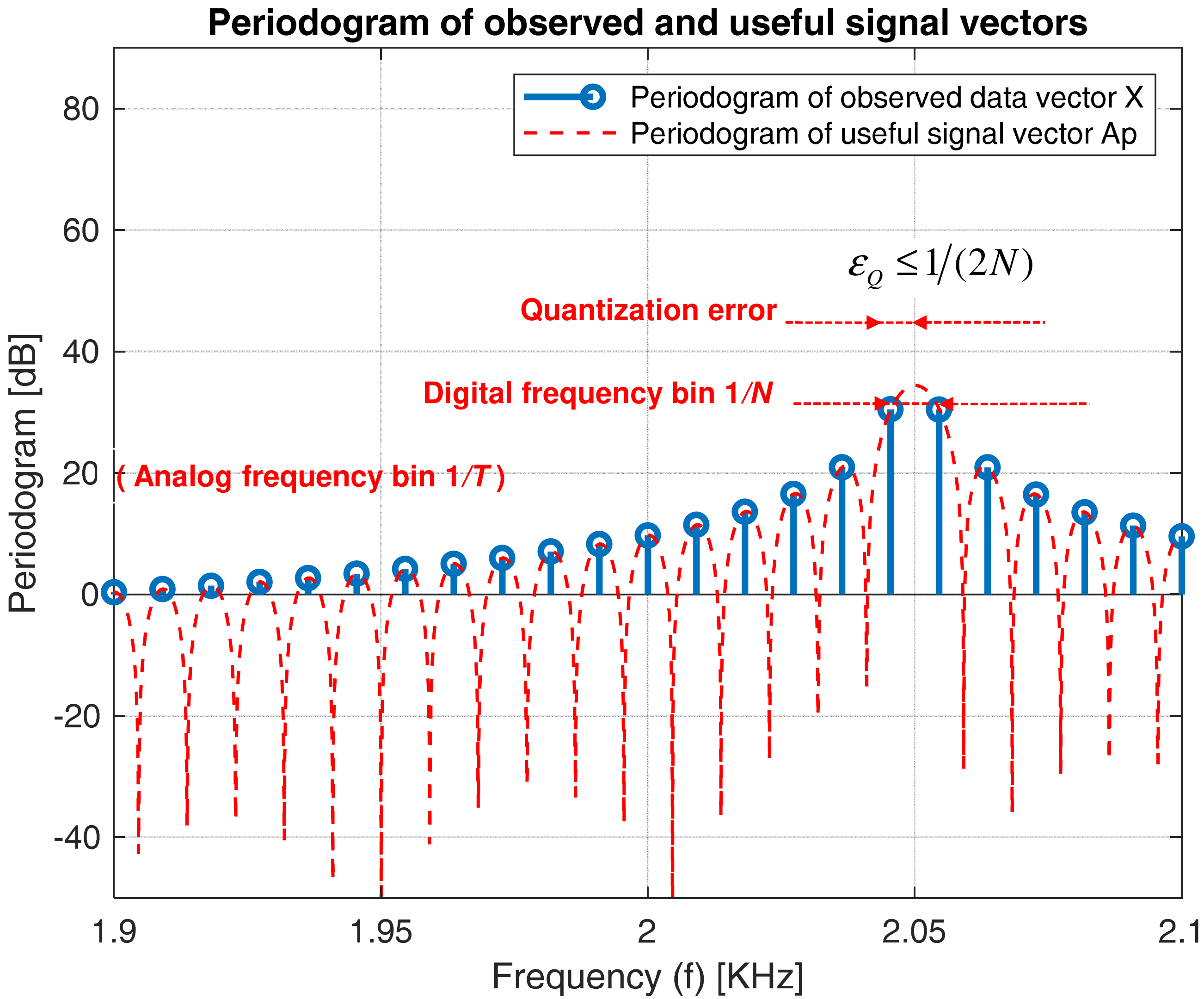
$$\text{where } x_i = 0 \text{ for } N \le i \le N_{zp} - 1$$

- The FFT is of order N_{zp}, so the maximum is obtained with an accuracy (in the absence of noise) of $1/N_{zp} \rightarrow$ there is a quantization error. In the following we will investigate how the estimation accuracy is affected by the value of N_{zp}.

- We will also observe the typical "**threshold effect**" for low SNR: below a threshold value of *SNR* the performance degrades very fast (we have the so-called "*catastrophic errors*"). This is a typical behavior for any non linear estimator.

Estimate of the Parameters of a Signal

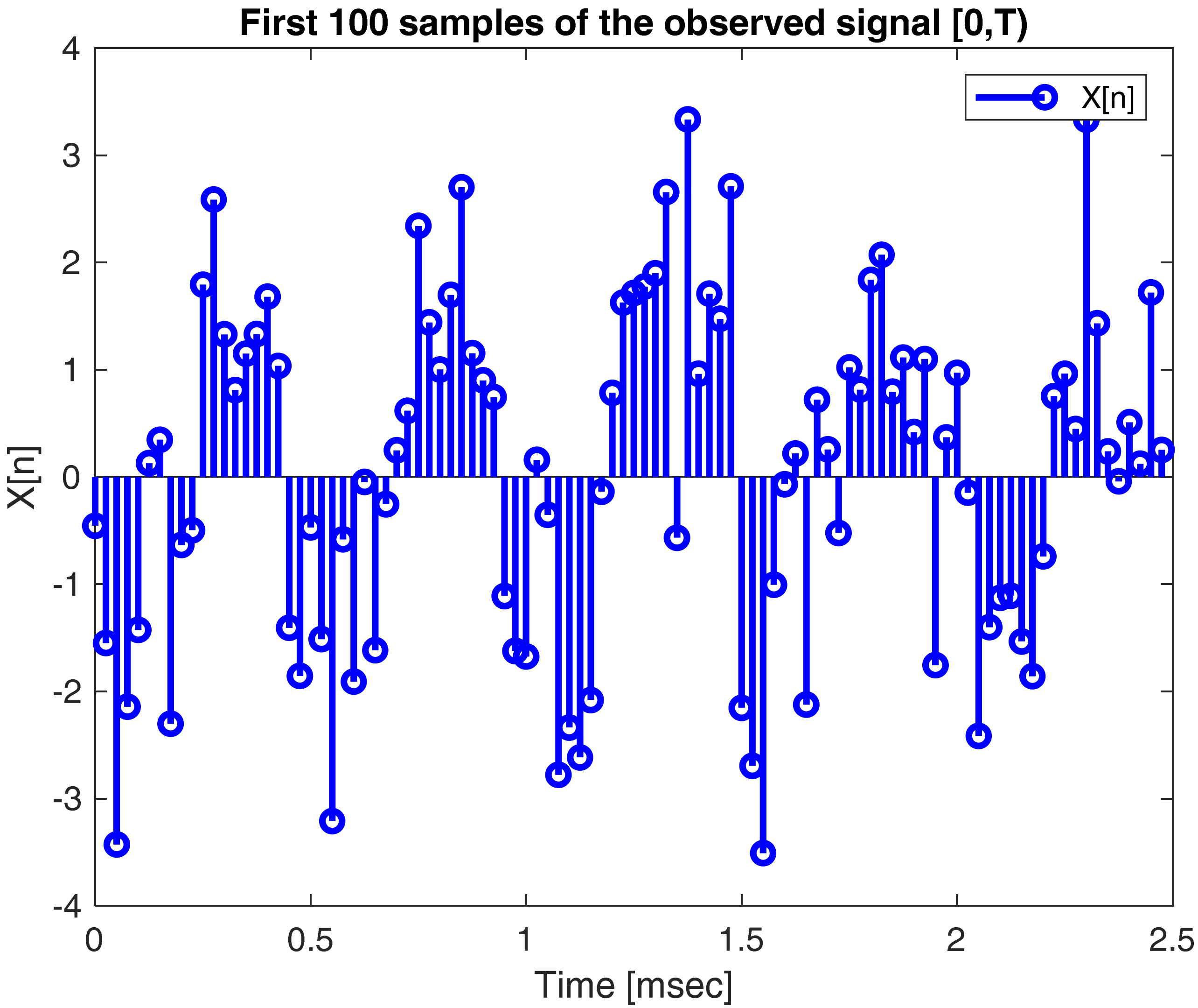
$$f_k = 2B \cdot F_k = 2B \cdot k/N = k/T\,, \quad k = 0,1,2,\cdots,N/2$$



$$A = 10$$
$$\theta = 2\pi/3$$
$$f_0 = 2.05\ KHz$$
$$SNR_{in} = 100\,dB$$
$$T = 105\ m\sec$$
$$B = 20\ KHz$$
$$N = 2010$$

- The quantization error is reduced by using the **zero-padding** method, so that the frequency bin of the FFT is $1/N_{zp}$ instead of $1/N$.

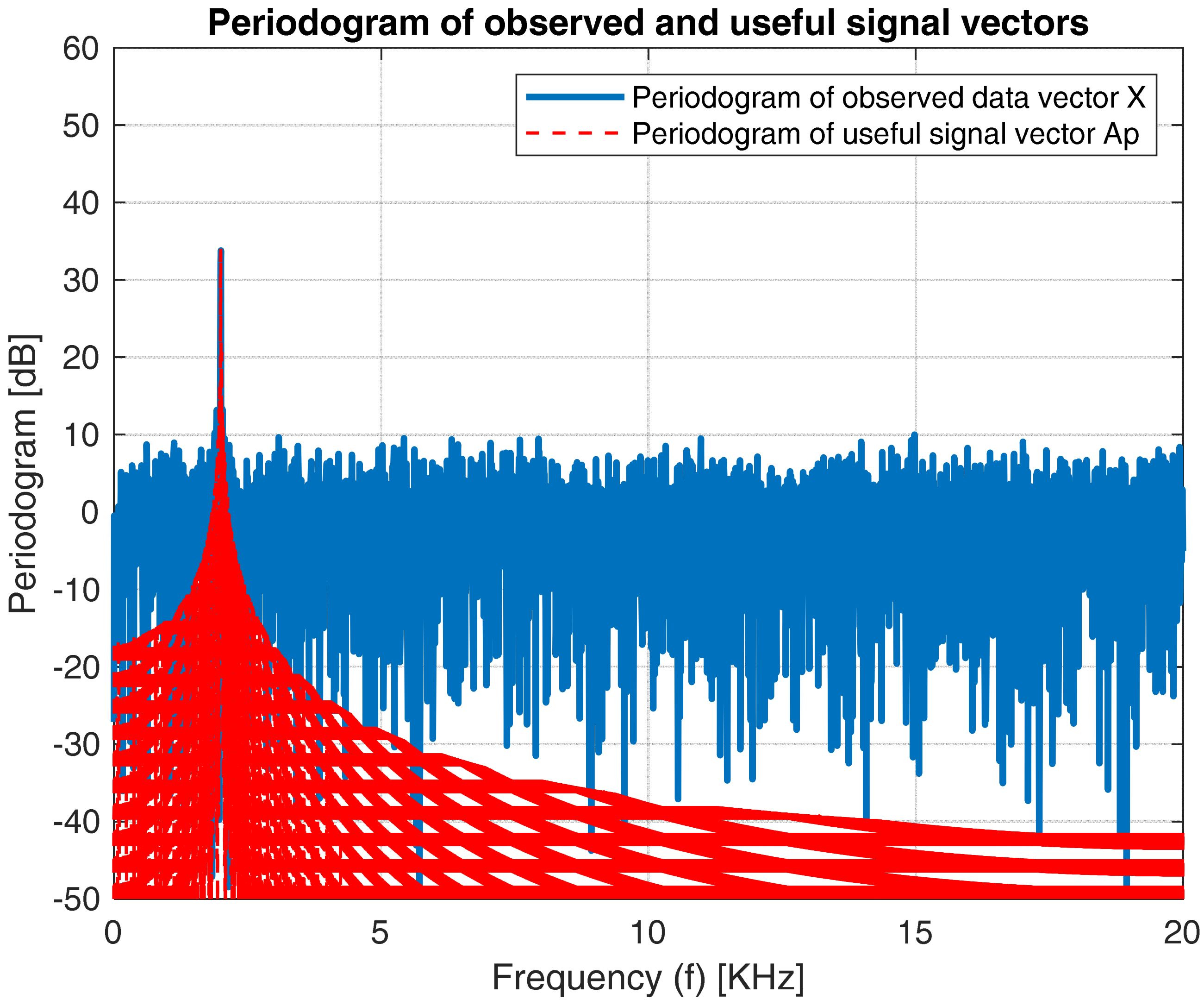
Estimate of the Parameters of a Signal



$$A = 10$$
$$\theta = 2\pi/3$$
$$f_0 = 2\ KHz$$
$$SNR_{in} = 0\,dB$$
$$T = 100\ m\sec$$
$$B = 20\ KHz$$
$$N = 4 \cdot 10^3$$
$$N_{zp} = 2^{19}$$

- We use the **zero-padding** method to interpolate in the frequency domain and reduce the quantization error.

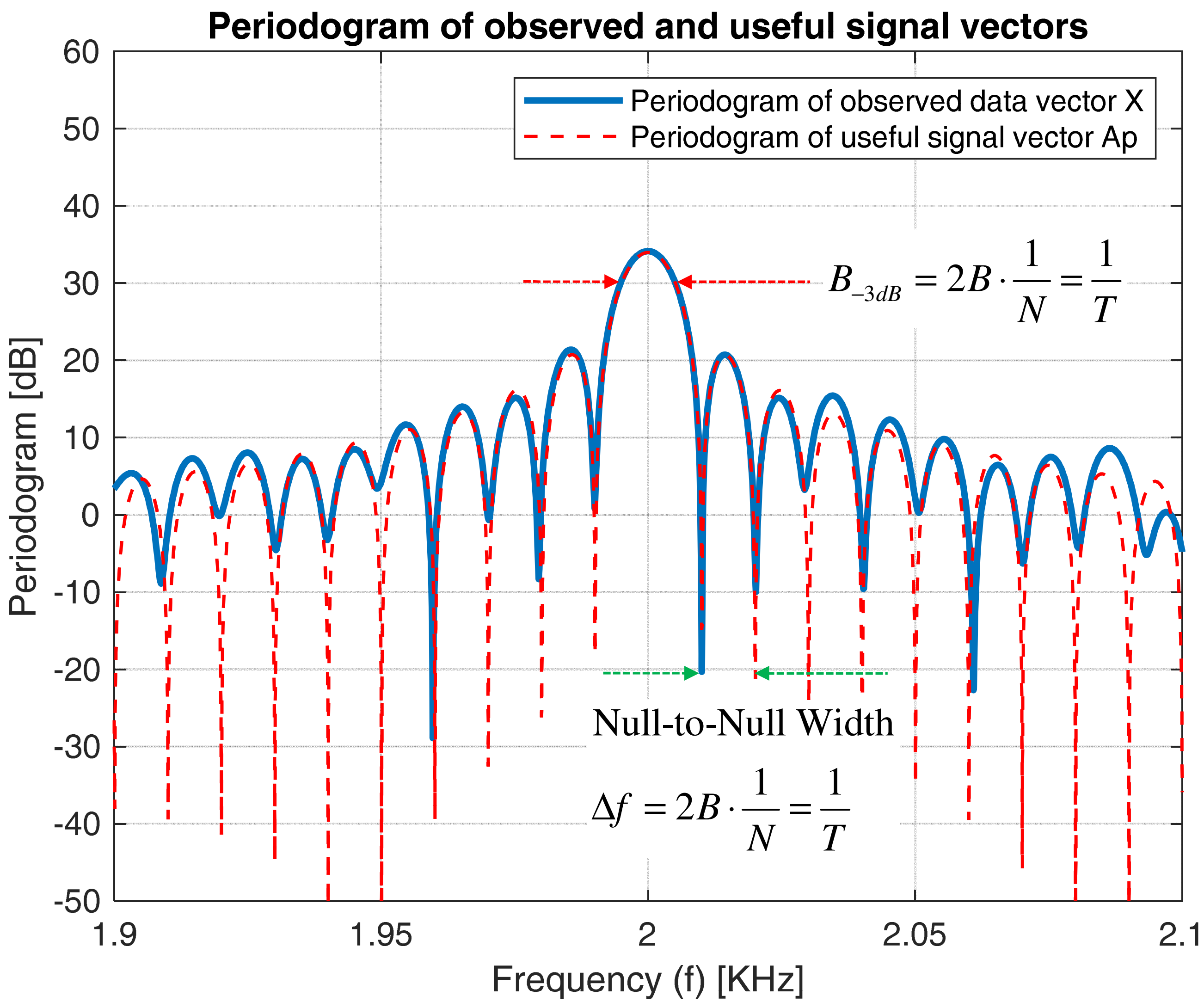
Estimate of the Parameters of a Signal



$$A = 10$$
$$\theta = 2\pi/3$$
$$f_0 = 2\ KHz$$
$$SNR_{in} = 0\,dB$$
$$T = 100\ m\sec$$
$$B = 20\ KHz$$
$$N = 4 \cdot 10^3$$
$$N_{zp} = 2^{19}$$

- By using the zero-padding, the frequency bin of the FFT is $1/N_{zp}$ instead of $1/N$.

Estimate of the Parameters of a Signal



$A = 10$

$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$SNR_{in} = 0\,dB$

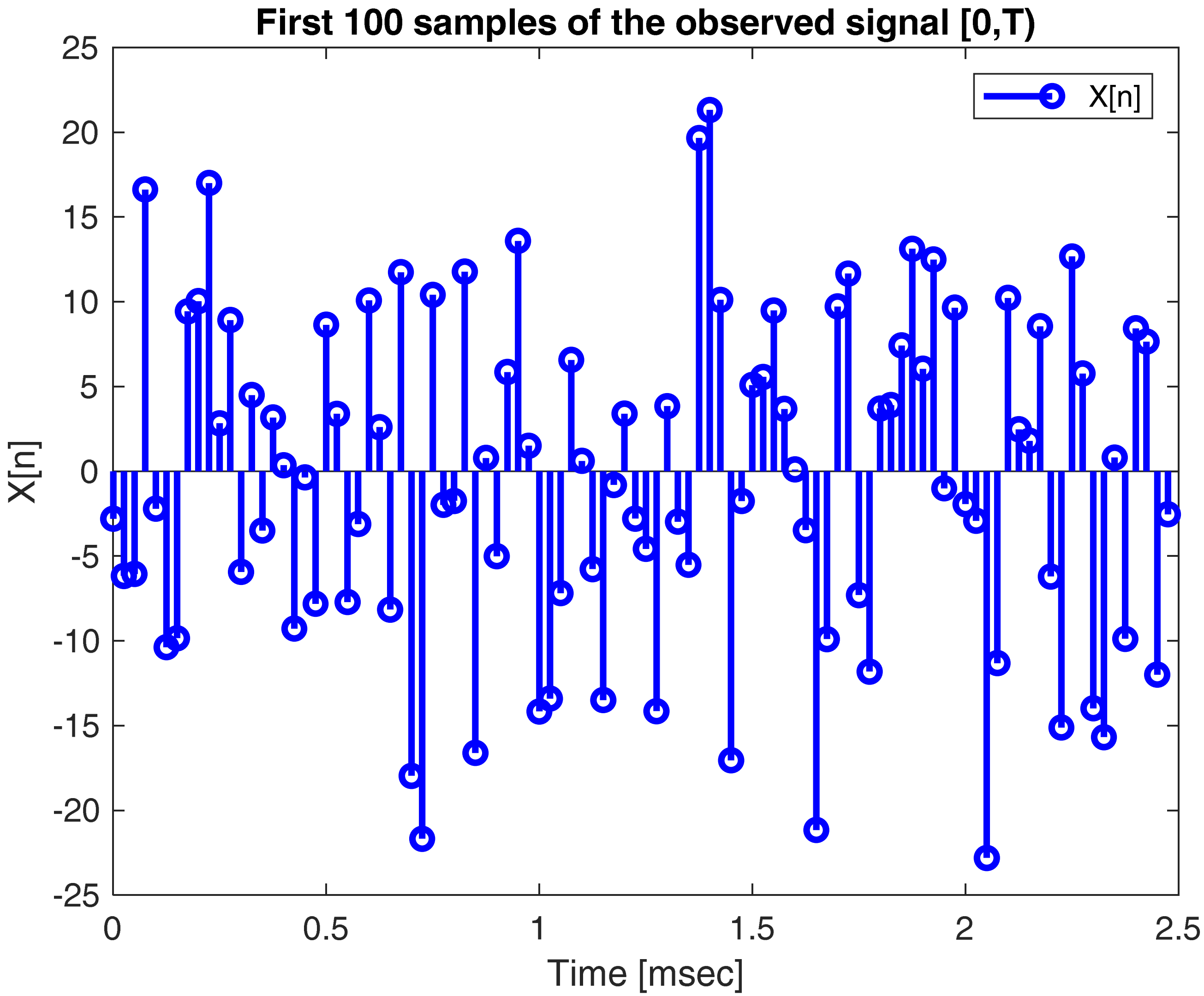
$T = 100\,m\sec$

$B = 20\ KHz$

$N = 4 \cdot 10^3$

$N_{zp} = 2^{19}$

Estimate of the Parameters of a Signal



$A = 10$

$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$SNR_{in} = -20\,dB$

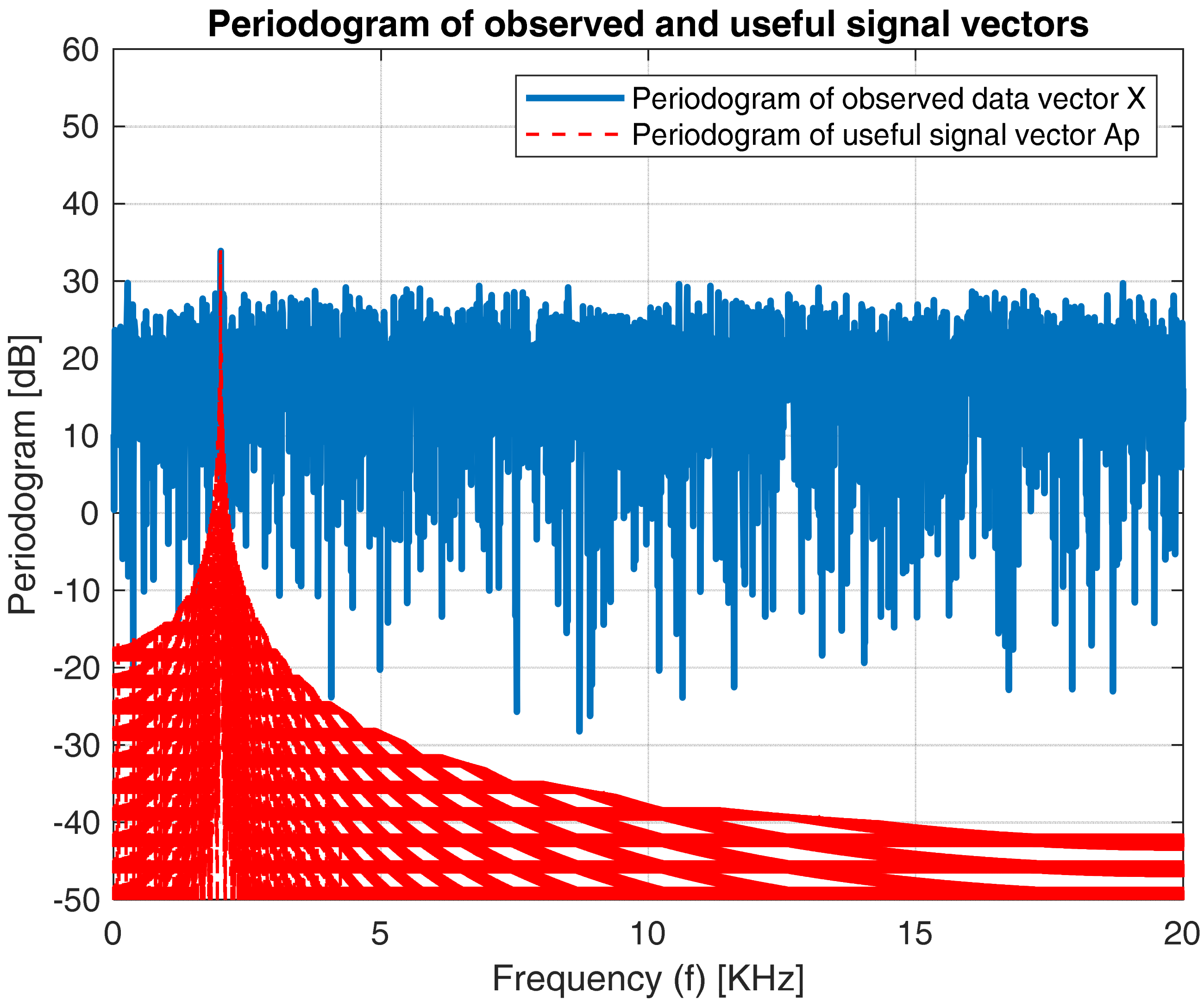
$T = 100\ m\sec$

$B = 20\ KHz$

$N = 4 \cdot 10^3$

$N_{zp} = 2^{19}$

Estimate of the Parameters of a Signal



$A = 10$

$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$SNR_{in} = -20\,dB$

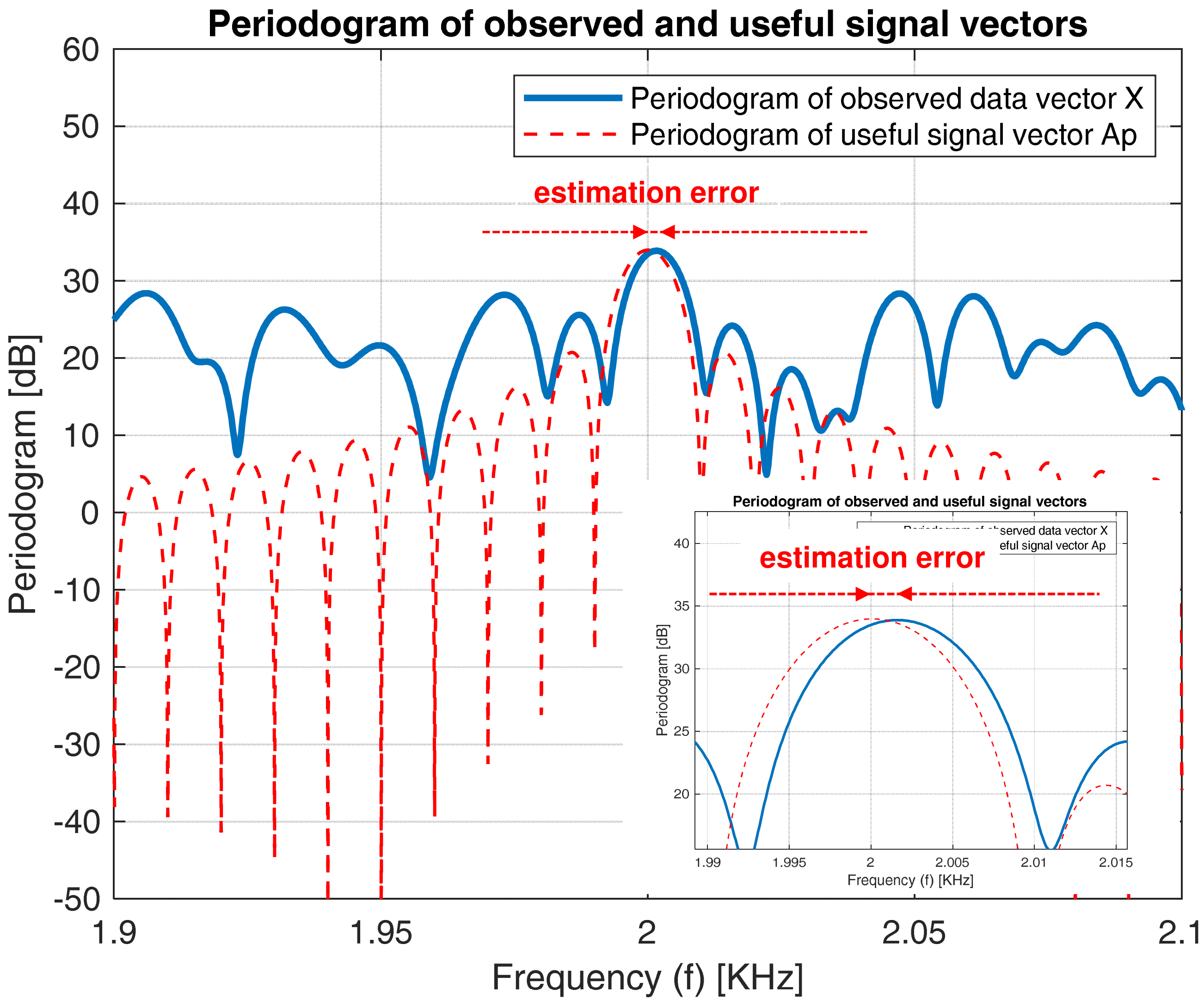
$T = 100\ m\sec$

$B = 20\ KHz$

$N = 2000$

$N_{zp} = 2^{19}$

Estimate of the Parameters of a Signal



$A = 10$

$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$SNR_{in} = -20\,dB$

$T = 100\ m\sec$

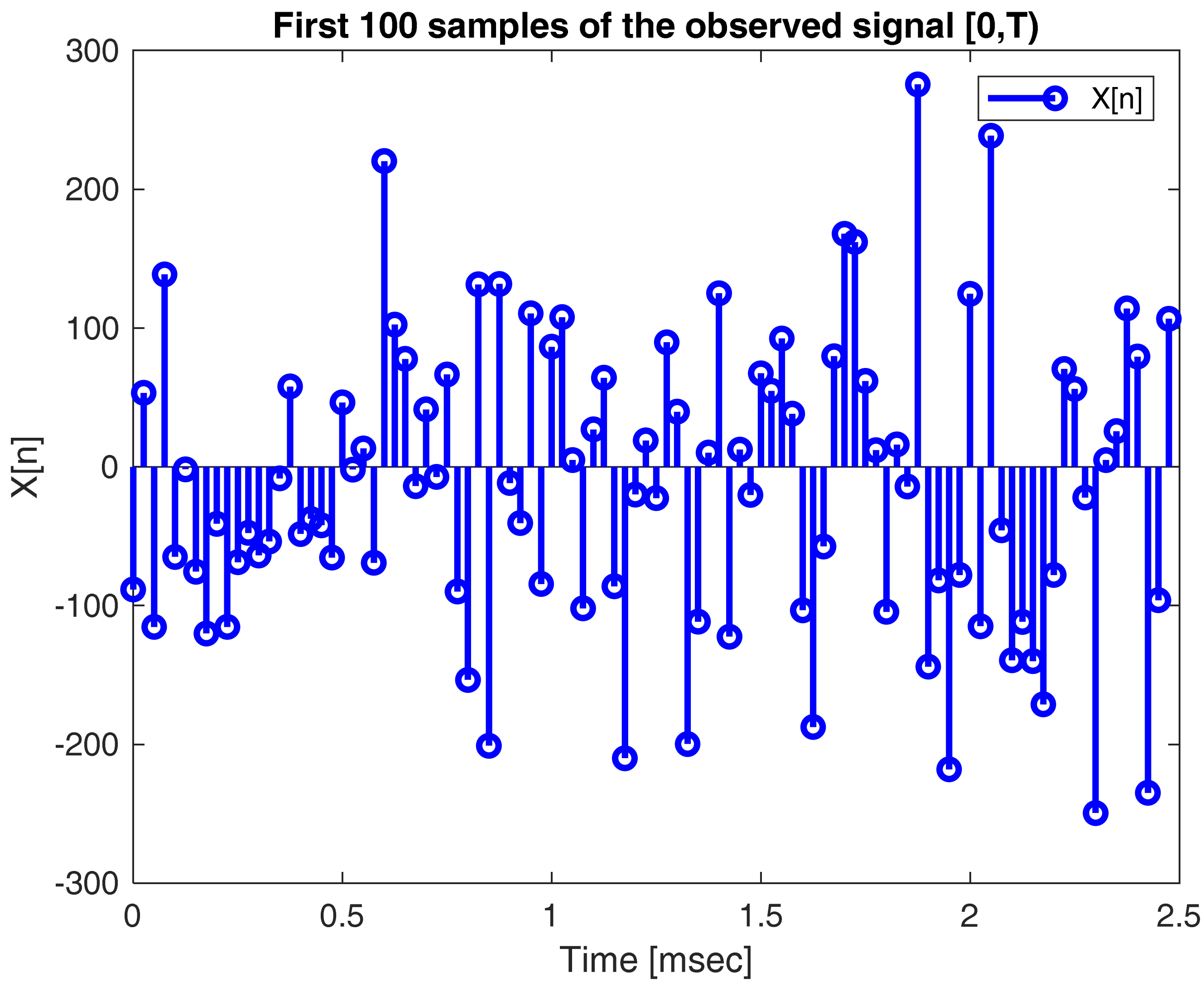
$B = 20\ KHz$

$N = 4 \cdot 10^3$

$N_{zp} = 2^{19}$

- For small SNR_{in} the estimation error is mainly due to noise, the quantization error is negligible.

Estimate of the Parameters of a Signal



$$A = 10$$

$$\theta = 2\pi/3$$

$$f_0 = 2\ KHz$$

$$SNR_{in} = -40\, dB$$

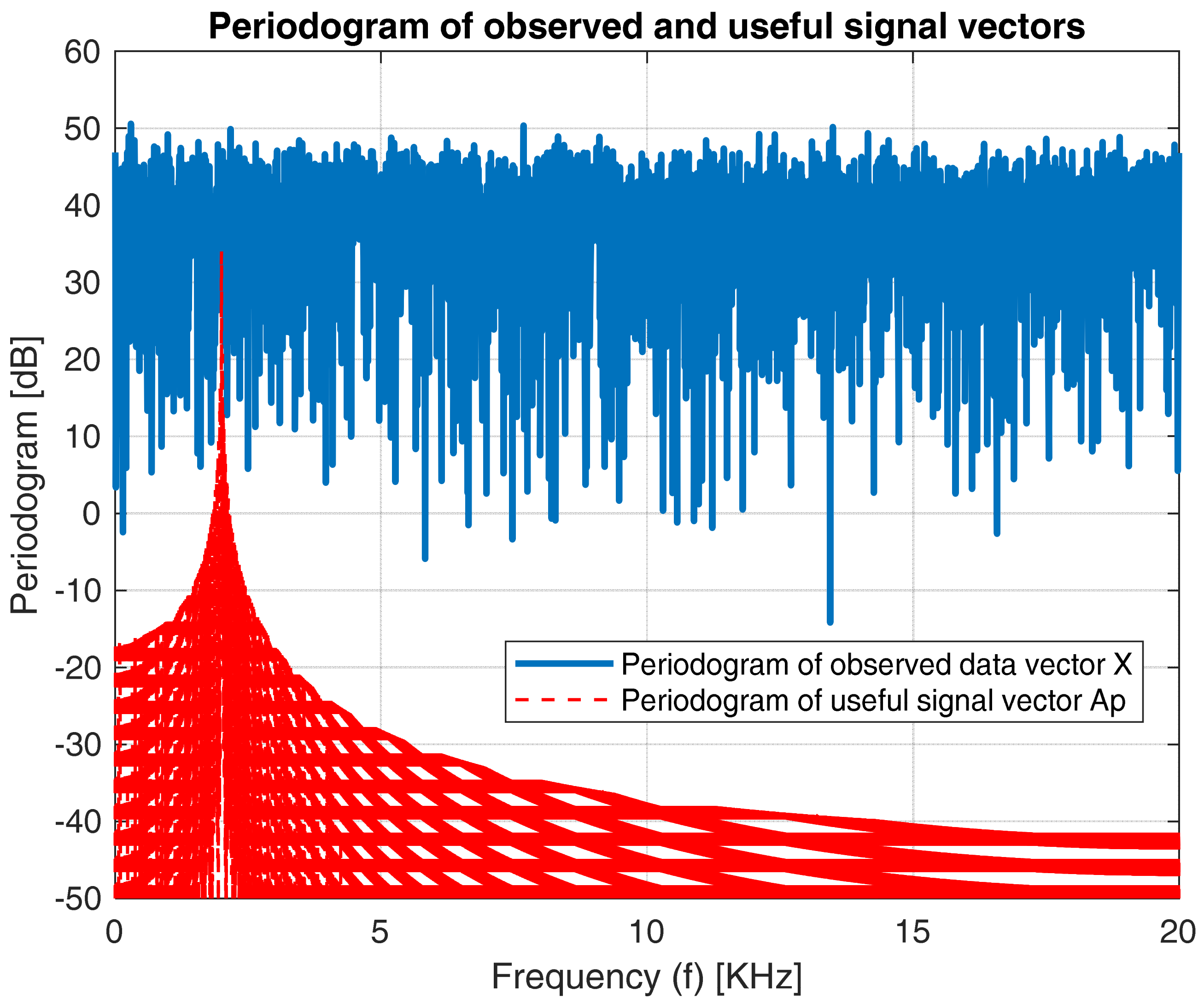
$$T = 100\ m\sec$$

$$B = 20\ KHz$$

$$N = 4 \cdot 10^3$$

$$N_{zp} = 2^{19}$$

Estimate of the Parameters of a Signal



$$A = 10$$

$$\theta = 2\pi/3$$

$$f_0 = 2\ KHz$$

$$SNR_{in} = -40\,dB$$

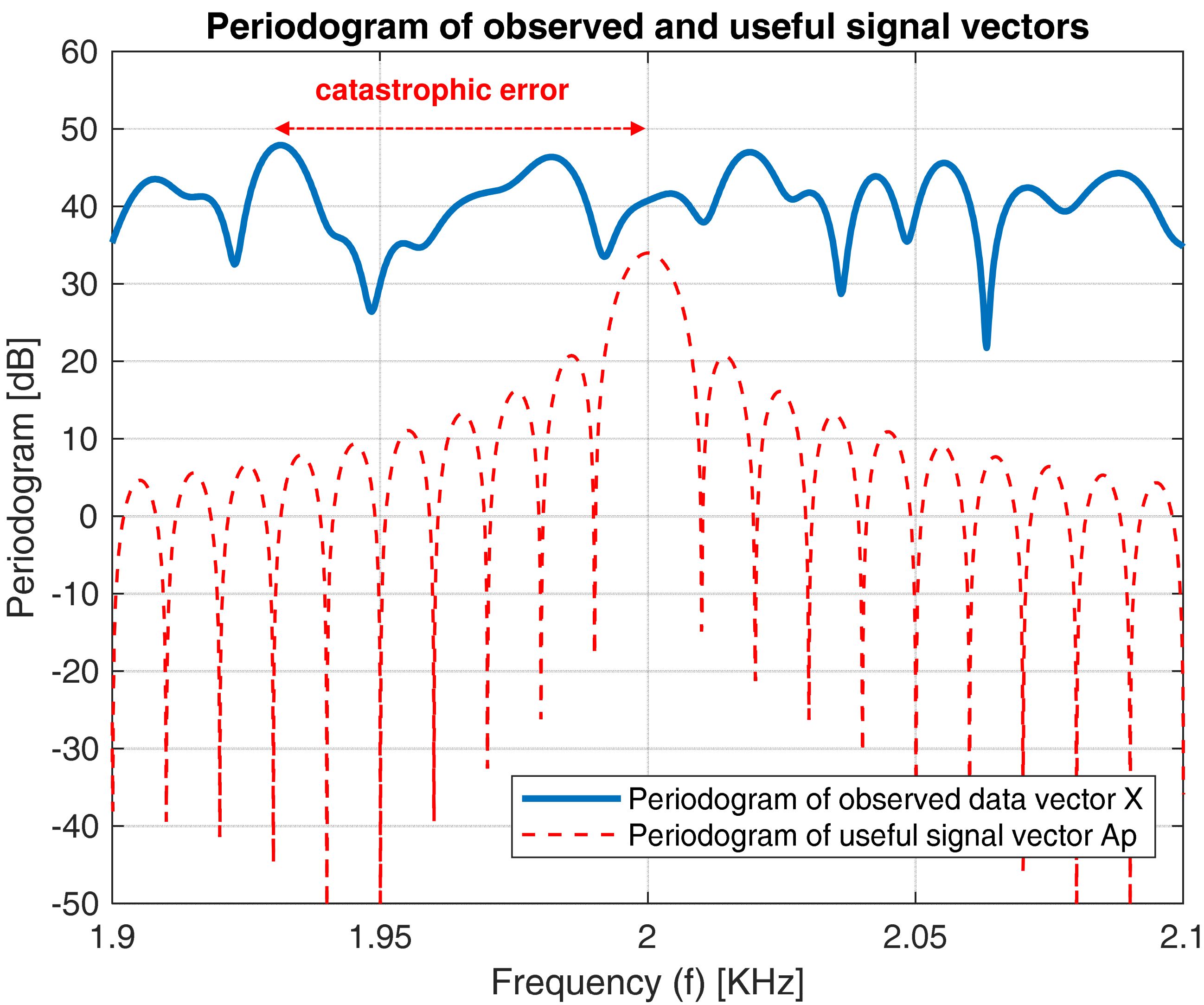
$$T = 100\ m\sec$$

$$B = 20\ KHz$$

$$N = 4 \cdot 10^3$$

$$N_{zp} = 2^{19}$$

Estimate of the Parameters of a Signal



$A = 10$

$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$SNR_{in} = -40\,dB$

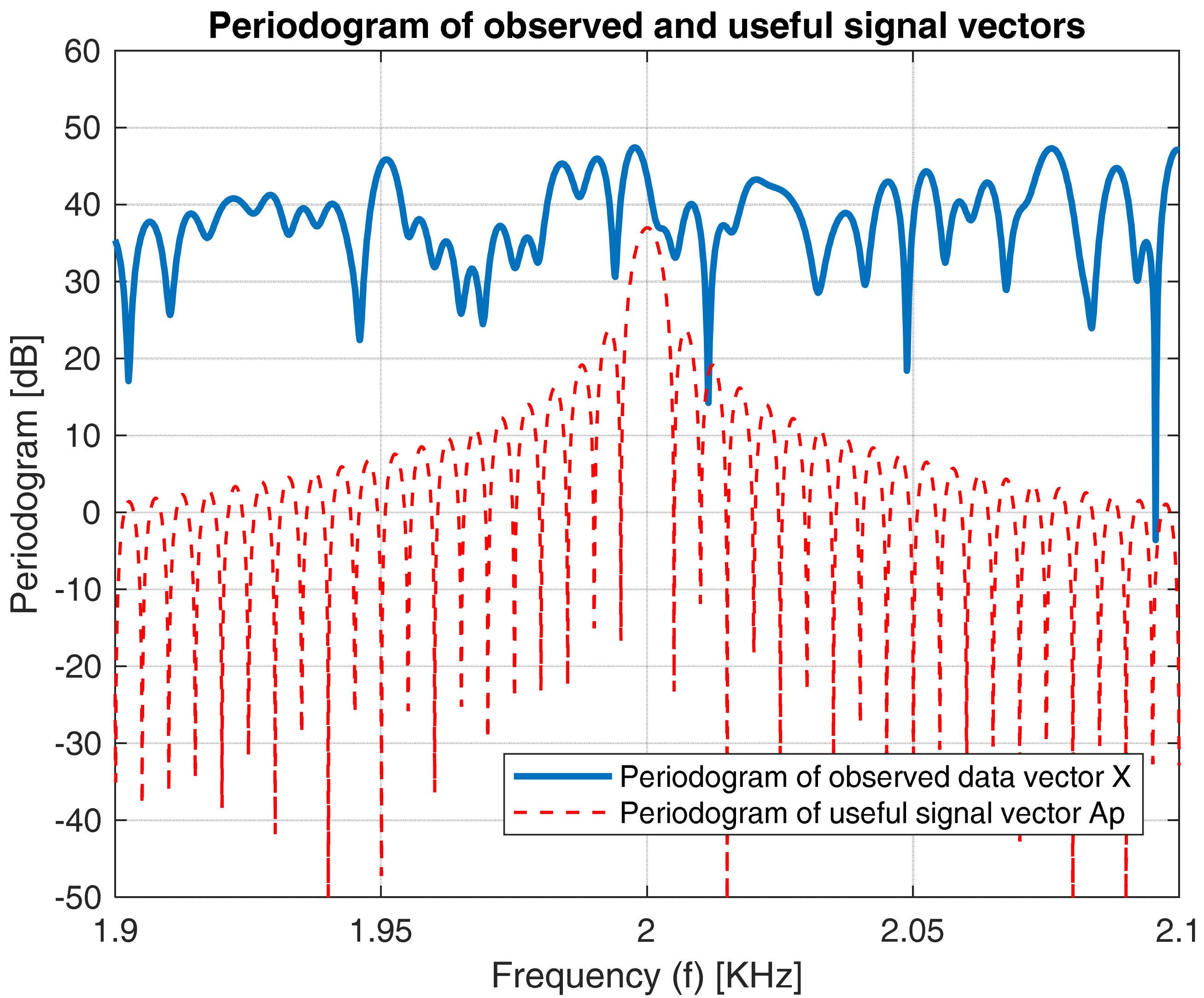
$T = 100\ m\sec$

$B = 20\ KHz$

$N = 4 \cdot 10^3$

$N_{zp} = 2^{19}$

Estimate of the Parameters of a Signal



$A = 10$

$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$SNR_{in} = -40\,dB$

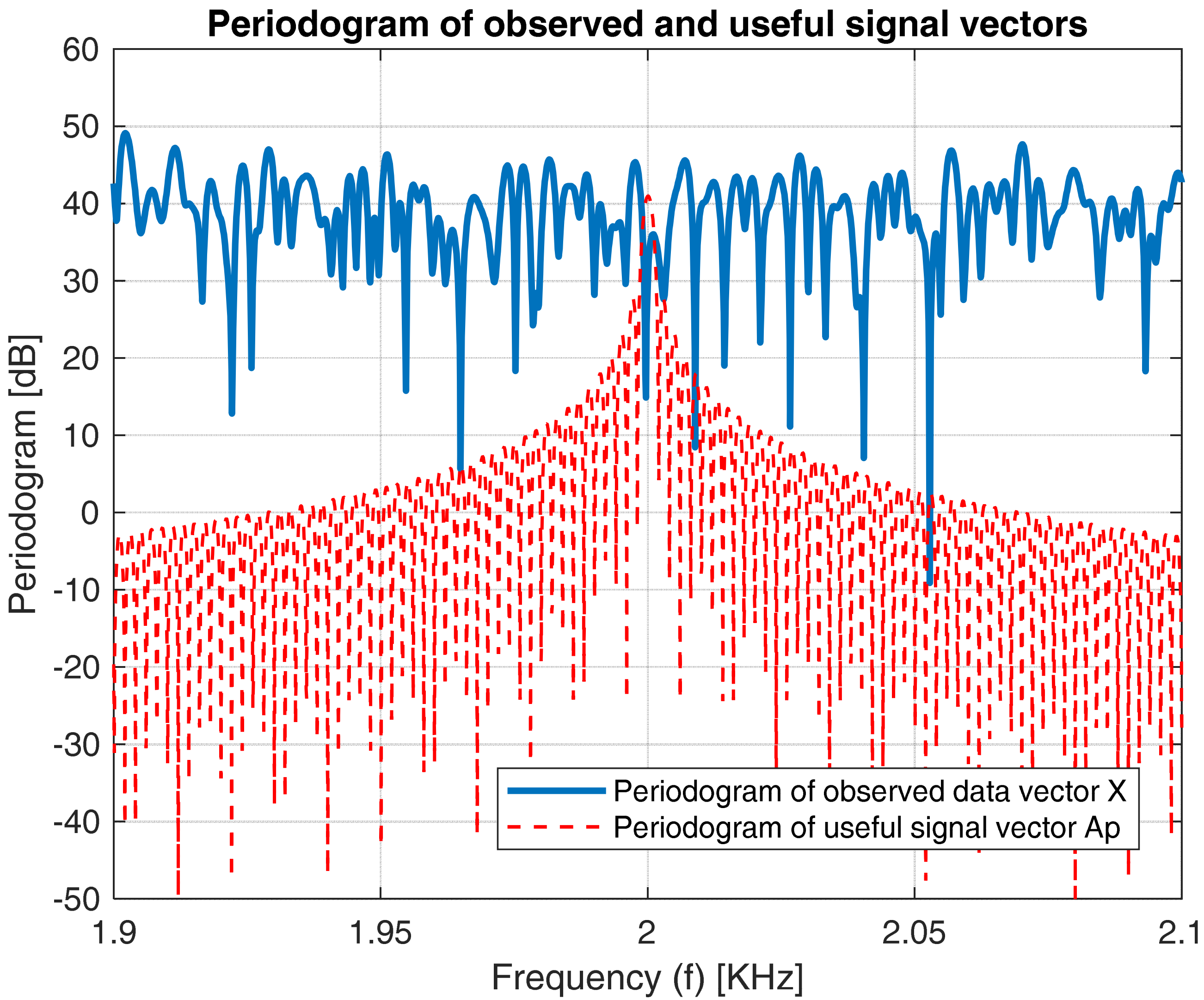
$T = 200\ m\sec$

$B = 20\ KHz$

$N = 8 \cdot 10^3$

$N_{zp} = 2^{19}$

Estimate of the Parameters of a Signal



$$A = 10$$

$$\theta = 2\pi/3$$

$$f_0 = 2\ KHz$$

$$SNR_{in} = -40\, dB$$

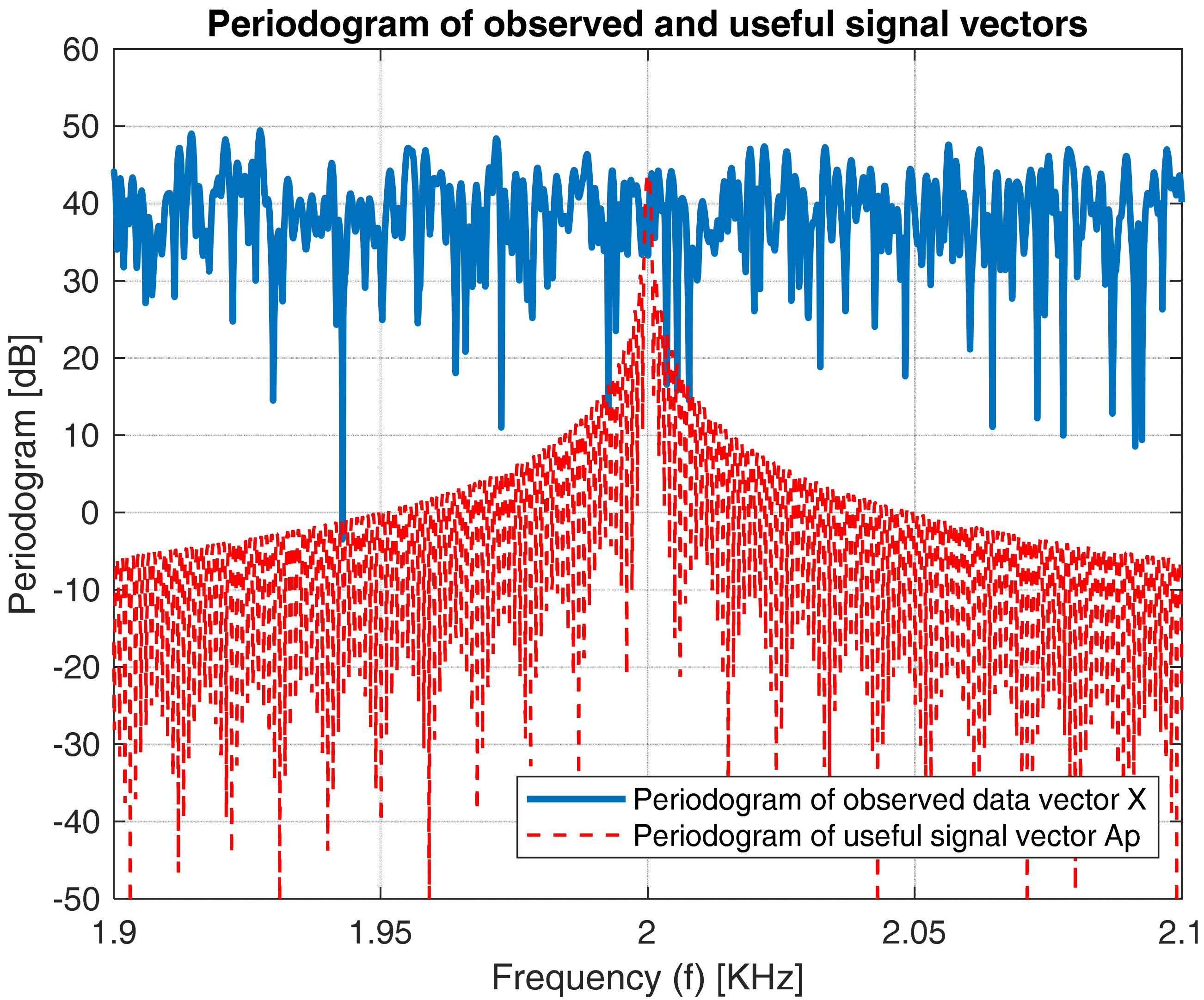
$$T = 500\ m\sec$$

$$B = 20\ KHz$$

$$N = 2 \cdot 10^4$$

$$N_{zp} = 2^{19}$$

Estimate of the Parameters of a Signal



$A = 10$

$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$SNR_{in} = -40\,dB$

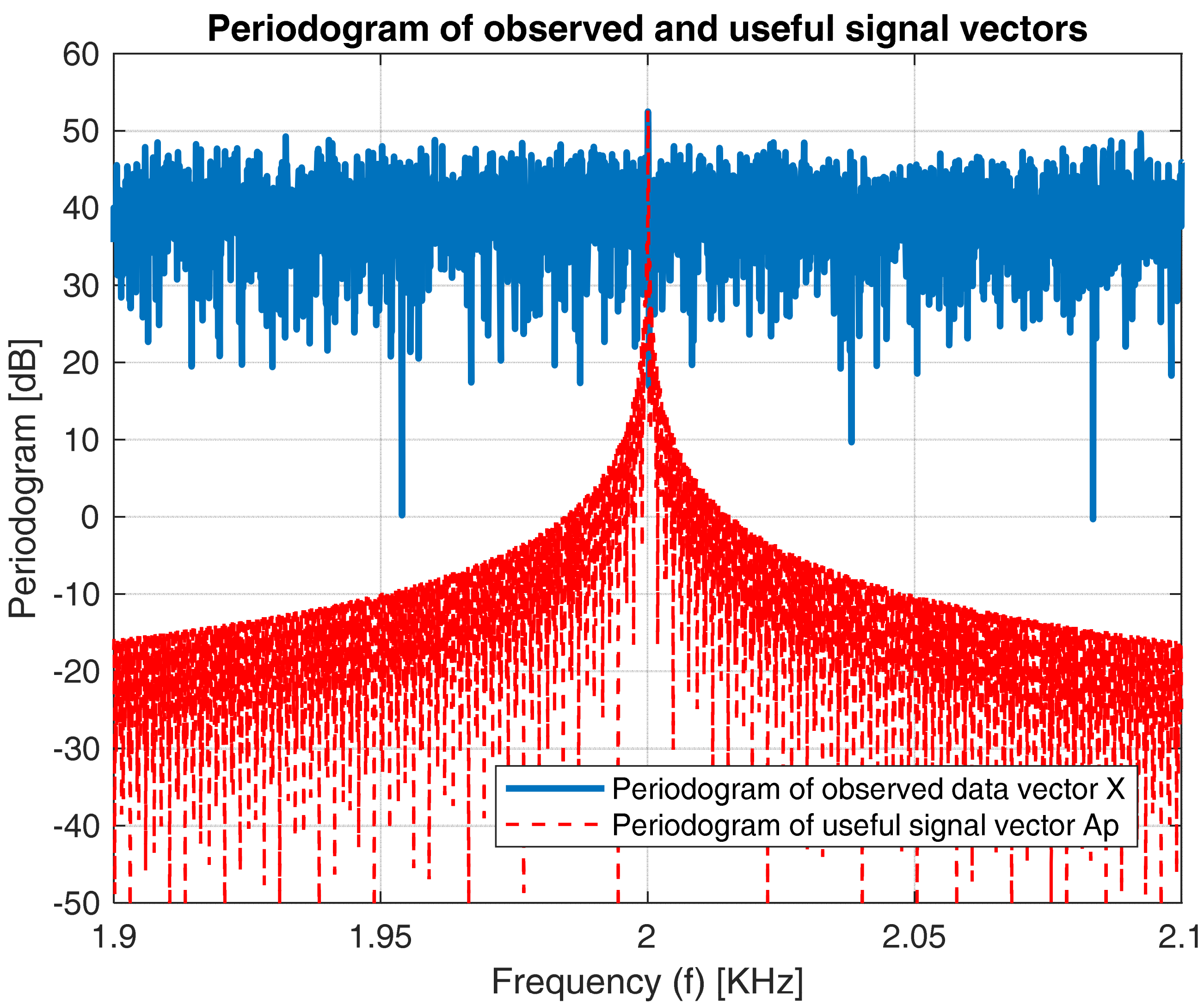
$T = 10^3\ m\sec$

$B = 20\ KHz$

$N = 4 \cdot 10^4$

$N_{zp} = 2^{19}$

Estimate of the Parameters of a Signal



$A = 10$

$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$SNR_{in} = -40\, dB$

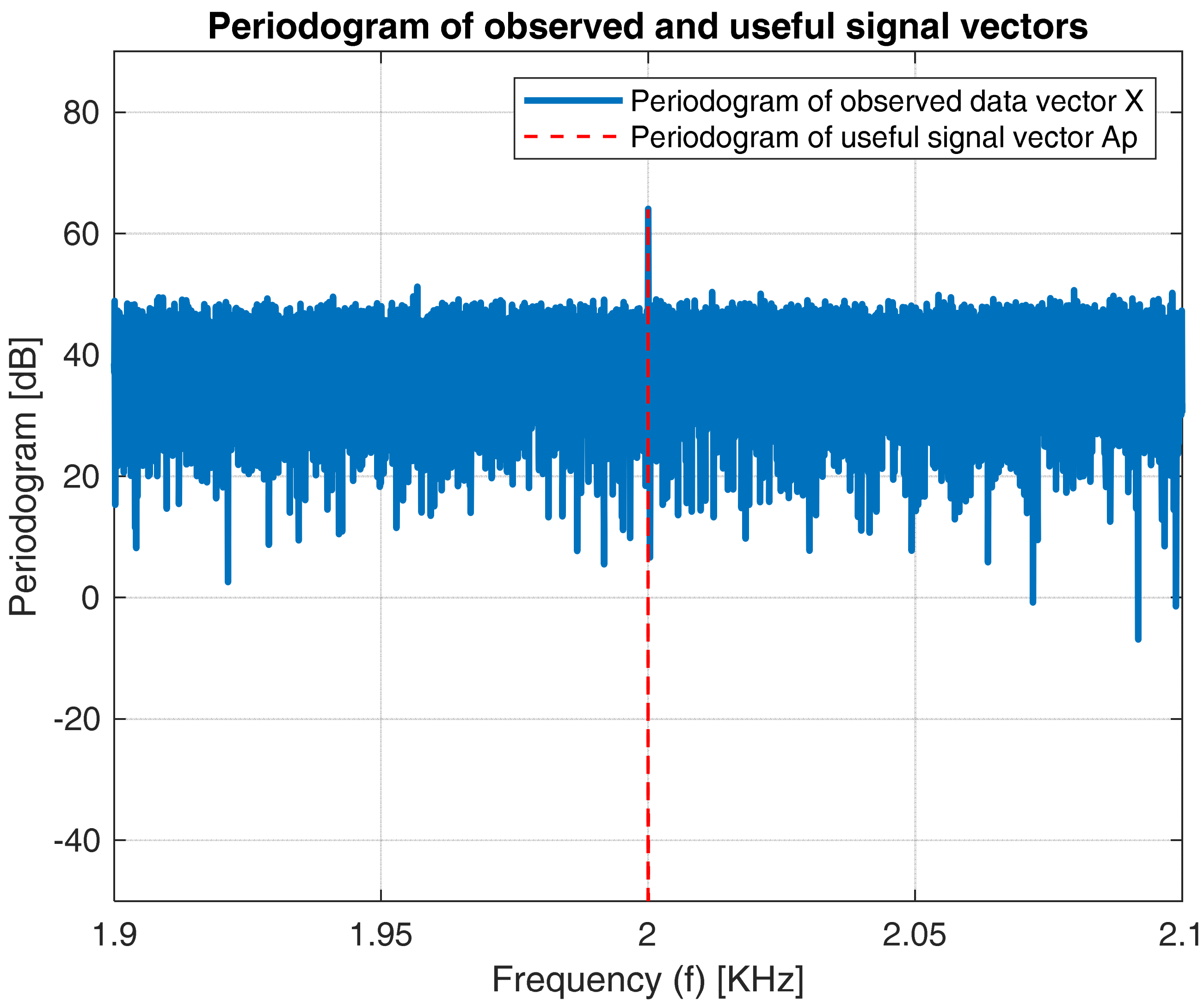
$T = 10^4\ m\sec$

$B = 20\ KHz$

$N = 4 \cdot 10^5$

$N_{zp} = 2^{19}$

Estimate of the Parameters of a Signal



$$A = 10$$

$$\theta = 2\pi/3$$

$$f_0 = 2\ KHz$$

$$SNR_{in} = -40\,dB$$

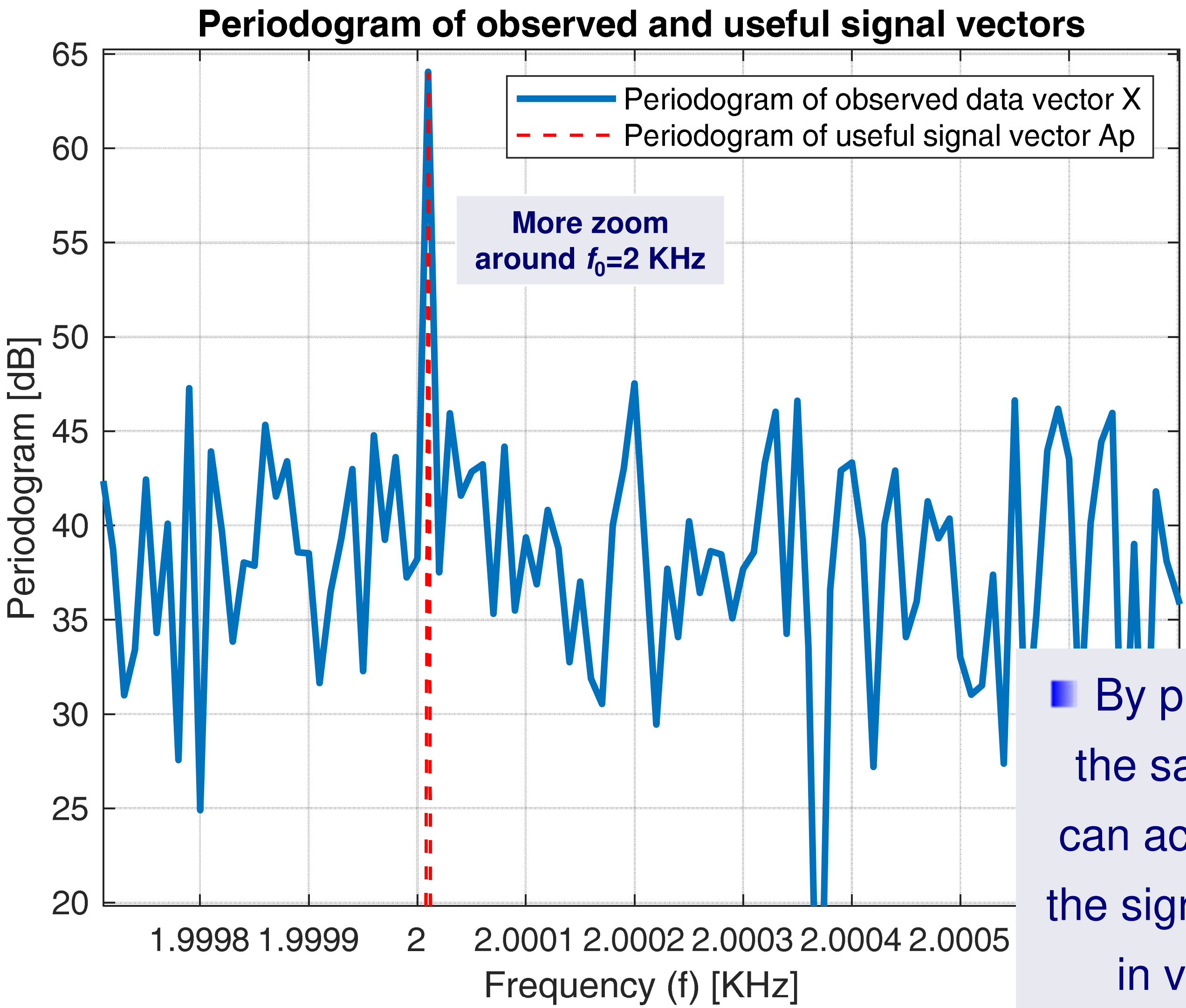
$$T = 10^5\ m\sec$$

$$B = 20\ KHz$$

$$N = 4 \cdot 10^6$$

- For such a large value of *N*, we do not use zero-padding.

Estimate of the Parameters of a Signal



$A = 10$

$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$SNR_{in} = -40\,dB$

$T = 10^5\ m\sec$

$B = 20\ KHz$

$N = 4 \cdot 10^6$

- By properly increasing the sample size *N*, we can accurately estimate the signal frequency also in very low SNR_{in}.

Estimate of the Parameters of a Signal

- **Cramér-Rao lower bounds (CRBs)** for the joint estimate of the parameters of a deterministic signal $s(t;\theta)$ observed in $t \in [0, T)$ in the presence of AWGN.

- After sampling, the data vector is: $\mathbf{X} = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{W}$

where θ is the vector of unknown parameters to be estimated.

- If the noise **W** is AWGN: $\mathbf{X} \in \mathcal{N}\left(\mathbf{s}(\boldsymbol{\theta}), \frac{N_0}{2}\mathbf{I}\right)$

- The log-likelihood function (log-LF):

$$\ln L(\boldsymbol{\theta}) = \ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) = -\frac{N}{2}\ln\left(\pi N_0\right) - \frac{1}{N_0}\sum_{i=0}^{N-1}\left(x_i - s_i(\boldsymbol{\theta})\right)^2$$

$$\text{where } \sum_{i=0}^{N-1}\left(x_i - s_i(\boldsymbol{\theta})\right)^2 = \left\|\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})\right\|^2$$

- Let's calculate now the elements of the **Fisher Information Matrix (FIM)**:

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{i,j} = -E\left\{\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right\}, \quad i, j = 1, 2, \cdots \quad , P$$

where $\boldsymbol{\theta} \triangleq [\theta_1 \quad \theta_2 \quad \cdots \quad \theta_P]^T$

$$\ln L(\boldsymbol{\theta}) = -\frac{N}{2}\ln\left(\pi N_0\right) - \frac{1}{N_0}\sum_{k=0}^{N-1}\left(x_k - s_k(\boldsymbol{\theta})\right)^2$$

$$\frac{\partial \ln L(\boldsymbol{\theta})}{\partial\theta_i} = \frac{2}{N_0}\sum_{k=0}^{N-1}\frac{\partial s_k(\boldsymbol{\theta})}{\partial\theta_i}\left(x_k - s_k(\boldsymbol{\theta})\right)$$

$$\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} = -\frac{2}{N_0}\sum_{k=0}^{N-1}\frac{\partial s_k(\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial s_k(\boldsymbol{\theta})}{\partial\theta_j} + \frac{2}{N_0}\sum_{k=0}^{N-1}\frac{\partial^2 s_k(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\left(x_k - s_k(\boldsymbol{\theta})\right)$$

$$\text{since} \quad x_k \in \mathcal{N}\left(s_k(\boldsymbol{\theta}), \frac{N_0}{2}\right) \quad \Rightarrow \quad E\left\{x_k - s_k(\boldsymbol{\theta})\right\} = 0$$

- Hence, we get the following general relationship:

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{i,j} = -E\left\{\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right\} = \frac{2}{N_0}\sum_{k=0}^{N-1}\frac{\partial s_k(\boldsymbol{\theta})}{\partial\theta_i}\cdot\frac{\partial s_k(\boldsymbol{\theta})}{\partial\theta_j}$$

- We now apply this formula to derive the **Cramér-Rao lower bounds (CRBs)** for the joint estimate of amplitude, phase, and frequency of a narrowband cosinusoidal signal observed in $t \in [0, T)$ in the presence of AWGN.

- In our case, the vector of unknown parameters to be estimated is:

$$\boldsymbol{\theta} \triangleq \left[A \; \theta \; F_0 \right]^T$$

and the SOI vector is given by:

$$\mathbf{s}(\boldsymbol{\theta}) = A \cdot \mathbf{p}(\theta, F_0) = \frac{A}{\sqrt{2B}} \left[\cos(\theta) \quad \cos(2\pi F_0 + \theta) \quad \cdots \quad \cos(2\pi F_0 (N-1) + \theta) \right]^T$$

$$s_k(\boldsymbol{\theta}) = \frac{A}{\sqrt{2B}} \cos(2\pi F_0 k + \theta) \quad \Rightarrow \quad \begin{cases} \dfrac{\partial s_k(\boldsymbol{\theta})}{\partial A} = \dfrac{1}{\sqrt{2B}} \cos(2\pi F_0 k + \theta) \\ \dfrac{\partial s_k(\boldsymbol{\theta})}{\partial \theta} = -\dfrac{A}{\sqrt{2B}} \sin(2\pi F_0 k + \theta) \\ \dfrac{\partial s_k(\boldsymbol{\theta})}{\partial F_0} = -\dfrac{2\pi k A}{\sqrt{2B}} \sin(2\pi F_0 k + \theta) \end{cases}$$

- The FIM elements are:

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{1,1} = \frac{2}{N_0}\sum_{k=0}^{N-1}\left[\frac{1}{\sqrt{2B}}\cos\left(2\pi F_0 k+\theta\right)\right]^2 = \frac{2}{N_0}\cdot\frac{N}{2B}\cdot\frac{1}{N}\sum_{k=0}^{N-1}\cos^2\left(2\pi F_0 k+\theta\right)$$

$$= \frac{2}{N_0}\cdot\frac{N}{2B}\cdot\frac{1}{2} = \frac{N}{2N_0 B} = \frac{T}{N_0} = \frac{SNR_{in}\cdot N}{A^2}$$

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{2,2} = \frac{2}{N_0}\sum_{k=0}^{N-1}\left[-\frac{A}{\sqrt{2B}}\sin\left(2\pi F_0 k+\theta\right)\right]^2 = \frac{2}{N_0}\cdot\frac{A^2 N}{2B}\cdot\frac{1}{N}\sum_{k=0}^{N-1}\sin^2\left(2\pi F_0 k+\theta\right)$$

$$= \frac{A^2 N}{2N_0 B} = \frac{A^2 T}{N_0} = SNR_{in}\cdot N$$

- where: $SNR_{in} = \dfrac{A^2}{2N_0 B} = \dfrac{A^2 T}{N_0 N}$

$$
\begin{aligned}
\left[\mathbf{I}(\boldsymbol{\theta})\right]_{3,3} &= \frac{2}{N_0}\sum_{k=0}^{N-1}\left[-\frac{2\pi k A}{\sqrt{2B}}\sin\left(2\pi F_0 k+\theta\right)\right]^2 \\
&= \frac{2}{N_0}\cdot\frac{4\pi^2 A^2}{2B}\sum_{k=0}^{N-1}k^2\sin^2\left(2\pi F_0 k+\theta\right) \\
&= \frac{4\pi^2 A^2}{N_0 B}\sum_{k=0}^{N-1}k^2\left[\frac{1}{2}-\frac{1}{2}\cos\left(4\pi F_0 k+2\theta\right)\right] \\
&= \frac{2\pi^2 A^2}{N_0 B}\sum_{k=0}^{N-1}k^2-\frac{2\pi^2 A^2}{N_0 B}\sum_{k=0}^{N-1}k^2\cos\left(4\pi F_0 k+2\theta\right) \\
&\cong \frac{2\pi^2 A^2}{N_0 B}\sum_{k=0}^{N-1}k^2=\frac{2\pi^2 A^2}{N_0 B}\cdot\frac{N(N-1)(2N-1)}{6} \\
&= \frac{2\pi^2 A^2 T}{3N_0}(N-1)(2N-1)
\end{aligned}
$$

- where we used the fact that for N>>1: $\frac{1}{N}\sum_{k=0}^{N-1}k^2\cos\left(4\pi F_0 k+2\theta\right)\cong 0$

Estimate of the Parameters of a Signal

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{1,2} = \left[\mathbf{I}(\boldsymbol{\theta})\right]_{2,1} = \frac{2}{N_0}\sum_{k=0}^{N-1}\left[\frac{1}{\sqrt{2B}}\cos\left(2\pi F_0 k + \theta\right)\right]\left[-\frac{A}{\sqrt{2B}}\sin\left(2\pi F_0 k + \theta\right)\right]$$

$$= -\frac{A}{N_0 B}\sum_{k=0}^{N-1}\cos\left(2\pi F_0 k + \theta\right)\sin\left(2\pi F_0 k + \theta\right)$$

$$= -\frac{A}{2N_0 B}\sum_{k=0}^{N-1}\left[\sin\left(4\pi F_0 k + 2\theta\right) + \sin\left(0\right)\right]$$

$$= -\frac{AN}{2N_0 B}\cdot\frac{1}{N}\sum_{k=0}^{N-1}\sin\left(4\pi F_0 k + 2\theta\right) \cong 0$$

- where we used the fact that for N>>1: $\frac{1}{N}\sum_{k=0}^{N-1}\sin\left(4\pi F_0 k + 2\theta\right) \cong 0$

Estimate of the Parameters of a Signal

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{1,3} = \left[\mathbf{I}(\boldsymbol{\theta})\right]_{3,1} = \frac{2}{N_0}\sum_{k=0}^{N-1}\left[\frac{1}{\sqrt{2B}}\cos\left(2\pi F_0 k + \theta\right)\right]\left[-\frac{2\pi k A}{\sqrt{2B}}\sin\left(2\pi F_0 k + \theta\right)\right]$$

$$= -\frac{2\pi A}{N_0 B}\sum_{k=0}^{N-1} k\cos\left(2\pi F_0 k + \theta\right)\sin\left(2\pi F_0 k + \theta\right)$$

$$= -\frac{2\pi A}{2 N_0 B}\sum_{k=0}^{N-1} k\left[\sin\left(4\pi F_0 k + 2\theta\right) + \sin(0)\right]$$

$$= -\frac{\pi A N}{N_0 B}\cdot\frac{1}{N}\sum_{k=0}^{N-1} k\sin\left(4\pi F_0 k + 2\theta\right) \cong 0$$

- where we used the fact that for N>>1: $\frac{1}{N}\sum_{k=0}^{N-1} k\sin\left(4\pi F_0 k + 2\theta\right) \cong 0$

Estimate of the Parameters of a Signal

$$\left[\mathbf{I}(\boldsymbol{\theta})\right]_{2,3} = \left[\mathbf{I}(\boldsymbol{\theta})\right]_{3,2} = \frac{2}{N_0}\sum_{k=0}^{N-1}\left[-\frac{A}{\sqrt{2B}}\sin\left(2\pi F_0 k+\theta\right)\right]\left[-\frac{2\pi k A}{\sqrt{2B}}\sin\left(2\pi F_0 k+\theta\right)\right]$$

$$= \frac{2\pi A^2}{N_0 B}\sum_{k=0}^{N-1} k\sin^2\left(2\pi F_0 k+\theta\right)$$

$$= \frac{2\pi A^2}{2N_0 B}\sum_{k=0}^{N-1} k\left[1-\cos\left(4\pi F_0 k+2\theta\right)\right]$$

$$= \frac{\pi A^2}{N_0 B}\sum_{k=0}^{N-1} k - \frac{\pi A N}{N_0 B}\cdot\frac{1}{N}\sum_{k=0}^{N-1} k\cos\left(4\pi F_0 k+2\theta\right)$$

$$\cong \frac{\pi A^2}{N_0 B}\sum_{k=0}^{N-1} k = \frac{\pi A^2}{N_0 B}\cdot\frac{N(N-1)}{2} = \frac{\pi A^2 T(N-1)}{N_0}$$

- where we used the fact that for N>>1: $\frac{1}{N}\sum_{k=0}^{N-1} k\cos\left(4\pi F_0 k+2\theta\right)\cong 0$

Estimate of the Parameters of a Signal

- The FIM is finally obtained:

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{T}{N_0} & 0 & 0 \\ 0 & \dfrac{A^2T}{N_0} & \dfrac{\pi A^2 T(N-1)}{N_0} \\ 0 & \dfrac{\pi A^2 T(N-1)}{N_0} & \dfrac{2\pi^2 A^2 T(N-1)(2N-1)}{3N_0} \end{bmatrix}$$

$$= \frac{A^2T}{N_0} \begin{bmatrix} \dfrac{1}{A^2} & 0 & 0 \\ 0 & 1 & \pi(N-1) \\ 0 & \pi(N-1) & \dfrac{2\pi^2 (N-1)(2N-1)}{3} \end{bmatrix}$$

- The inverse of the FIM:

$$CRB(\boldsymbol{\theta}) = \mathbf{I}^{-1}(\boldsymbol{\theta}) = \frac{N_0}{A^2T}\begin{bmatrix} \frac{1}{A^2} & 0 & 0 \\ 0 & 1 & \pi(N-1) \\ 0 & \pi(N-1) & \frac{2\pi^2(N-1)(2N-1)}{3} \end{bmatrix}^{-1}$$

$$= \frac{N_0}{A^2T}\begin{bmatrix} A^2 & 0 & 0 \\ 0 & \frac{2(2N-1)}{N+1} & -\frac{3}{\pi(N+1)} \\ 0 & -\frac{3}{\pi(N+1)} & \frac{3}{\pi^2(N^2-1)} \end{bmatrix}$$

- We note that the estimate of the amplitude is decoupled from the estimate of phase and frequency. This is true asymptotically (large *N*), since the ML estimators are asymptotically efficient, but not always efficient.

Estimate of the Parameters of a Signal

$$SNR_{in} = \frac{A^2}{2N_0 B}$$

- The CRB's on the three parameters are given by:

$$CRB(A) = \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{1,1} = \frac{N_0}{T} = \frac{A^2}{N \cdot SNR_{in}}$$

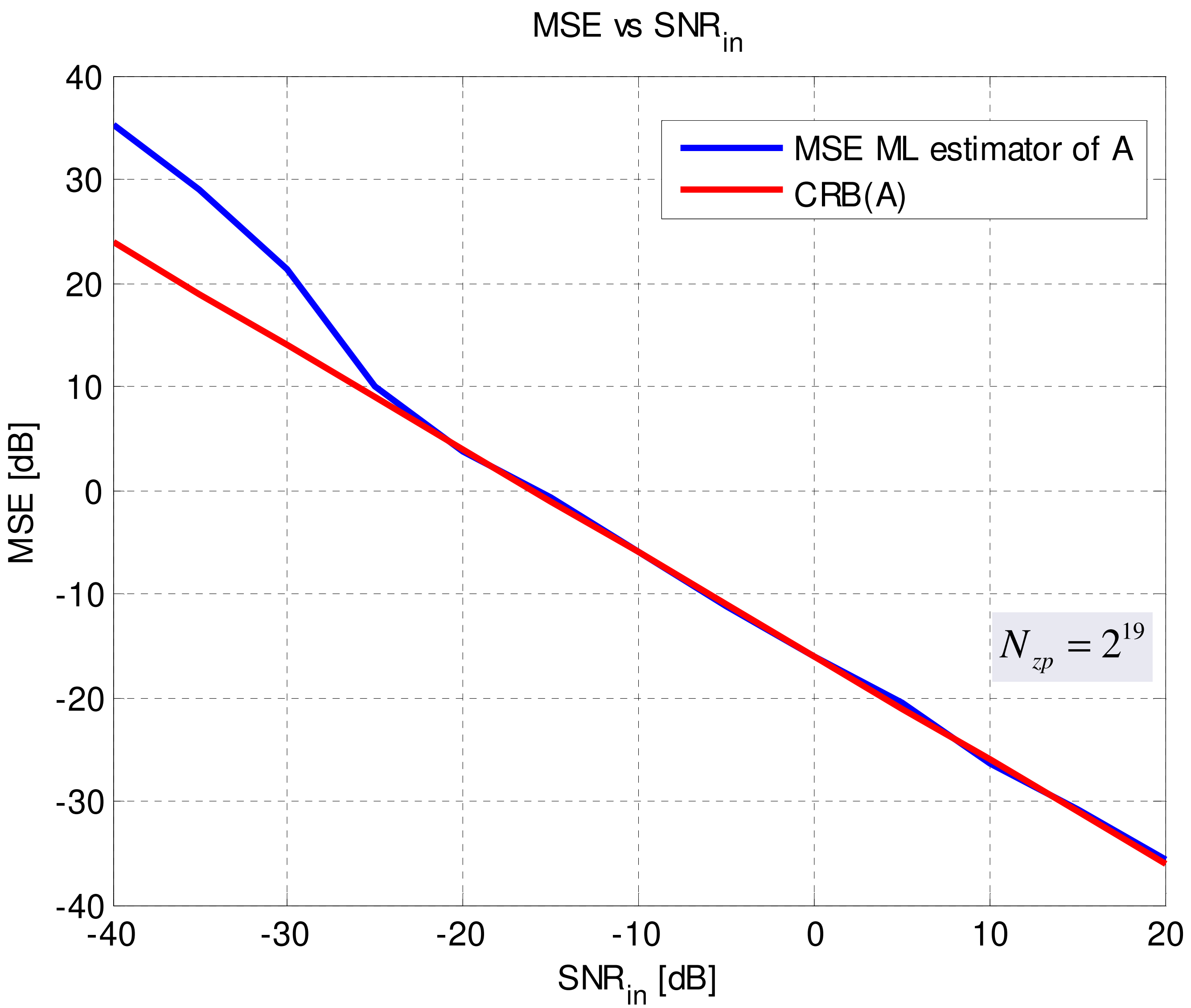
$$CRB(\theta) = \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{2,2} = \frac{N_0}{A^2 T} \cdot \frac{2(2N-1)}{N+1} = \frac{2(2N-1)}{N(N+1)SNR_{in}}$$

$$CRB(F_0) = \left[\mathbf{I}^{-1}(\boldsymbol{\theta})\right]_{3,3} = \frac{N_0}{A^2 T} \cdot \frac{3}{\pi^2 (N^2-1)} = \frac{3}{\pi^2 N(N^2-1)SNR_{in}}$$

- The CRB's on the analog frequency $f_0 = F_0 / T_c = (2B) F_0$ is easily obtained:

$$CRB(f_0) = (2B)^2 \cdot CRB(F_0) = \frac{N_0}{A^2 T} \cdot \frac{12B^2}{\pi^2 (N^2-1)} = \frac{12B^2}{\pi^2 N(N^2-1)SNR_{in}}$$

Estimate of the Parameters of a Signal



$A = 10$

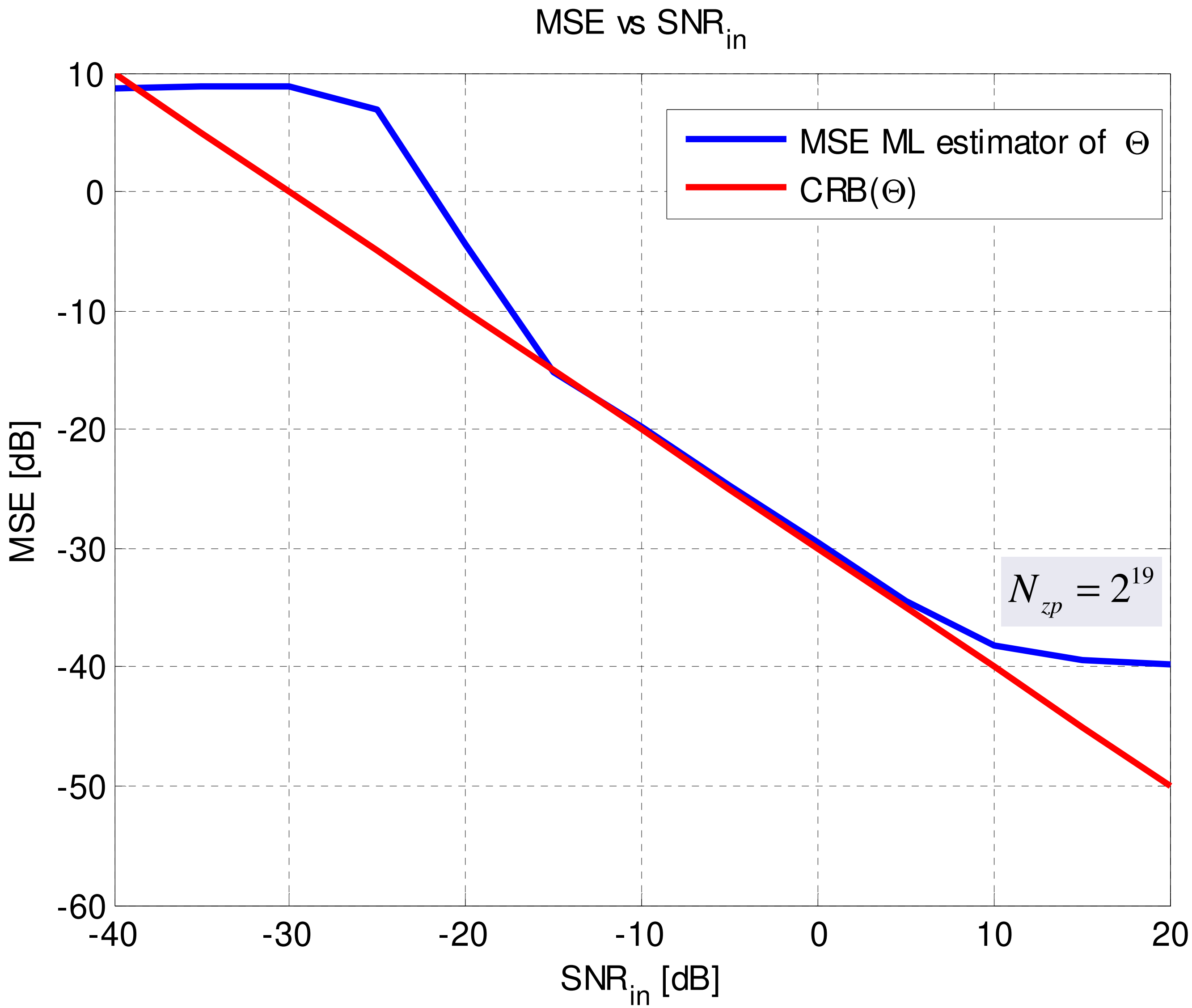
$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$T = 100\ m\sec$

$B = 20\ KHz$

Estimate of the Parameters of a Signal



$$A = 10$$

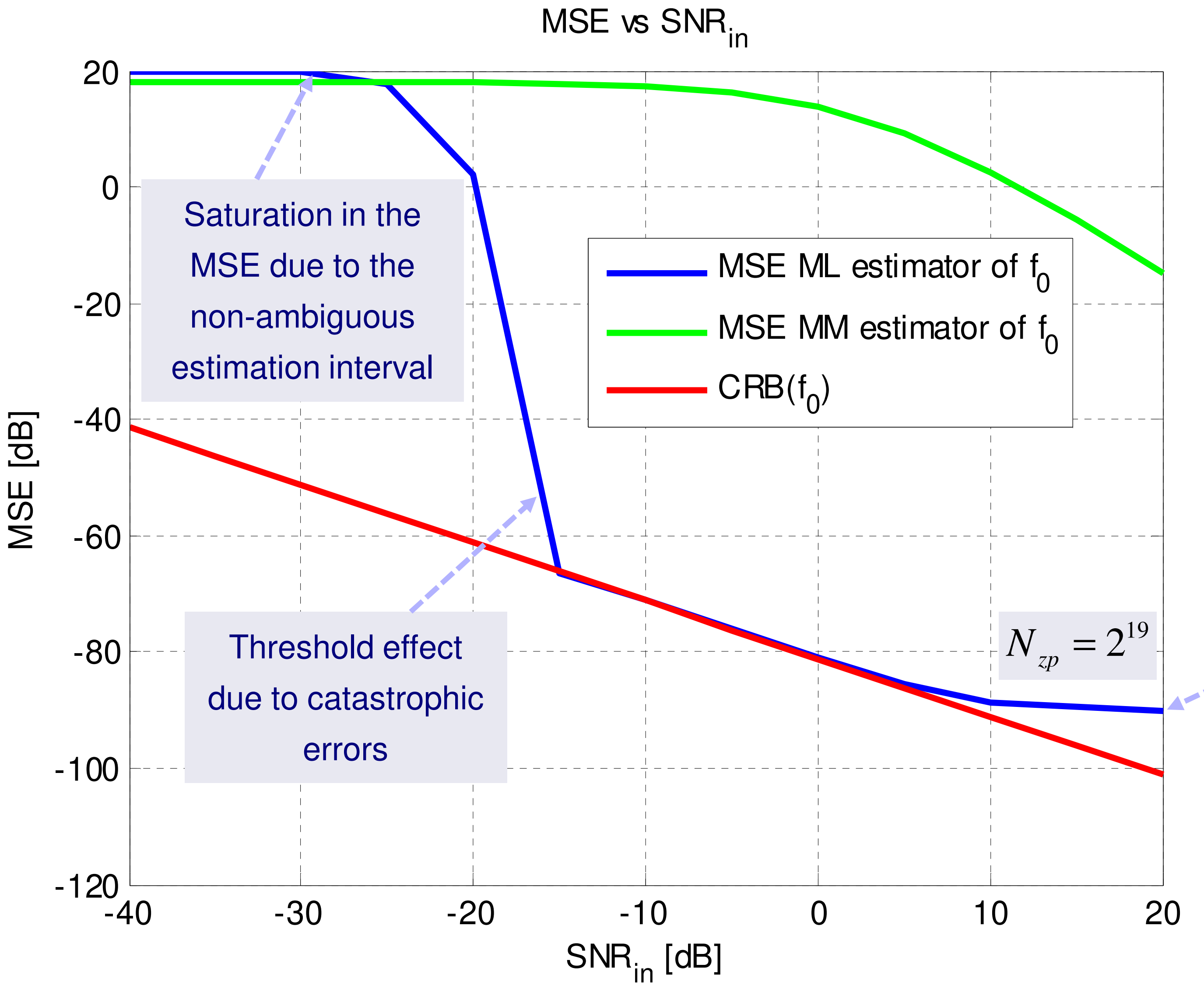
$$\theta = 2\pi/3$$

$$f_0 = 2\ KHz$$

$$T = 100\ m\sec$$

$$B = 20\ KHz$$

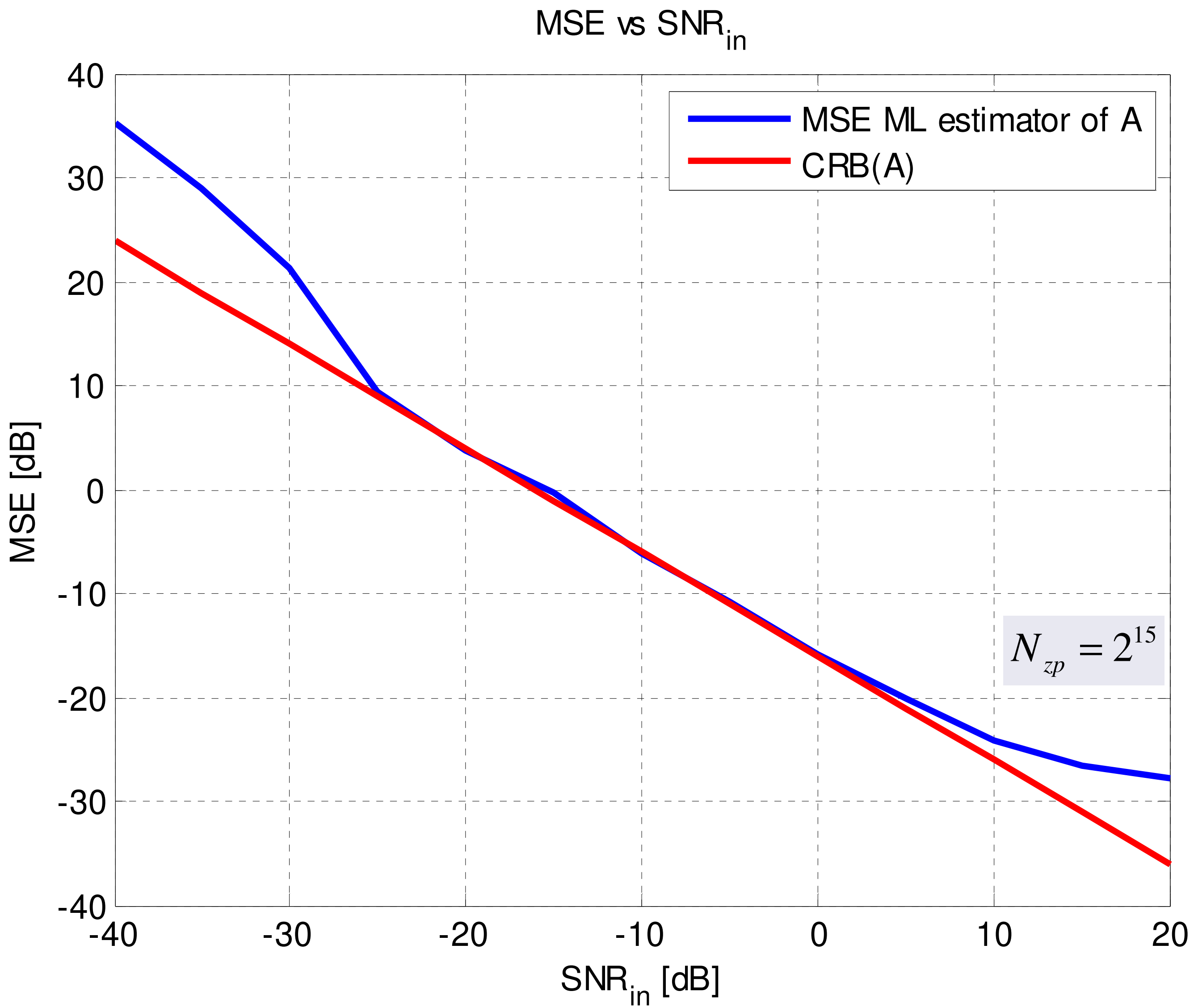
Estimate of the Parameters of a Signal



$$A = 10$$
$$\theta = 2\pi/3$$
$$f_0 = 2\ KHz$$
$$T = 100\ m\sec$$
$$B = 20\ KHz$$

Floor in the MSE due to the quantization error, that does not depend on the SNR.

Estimate of the Parameters of a Signal



$A = 10$

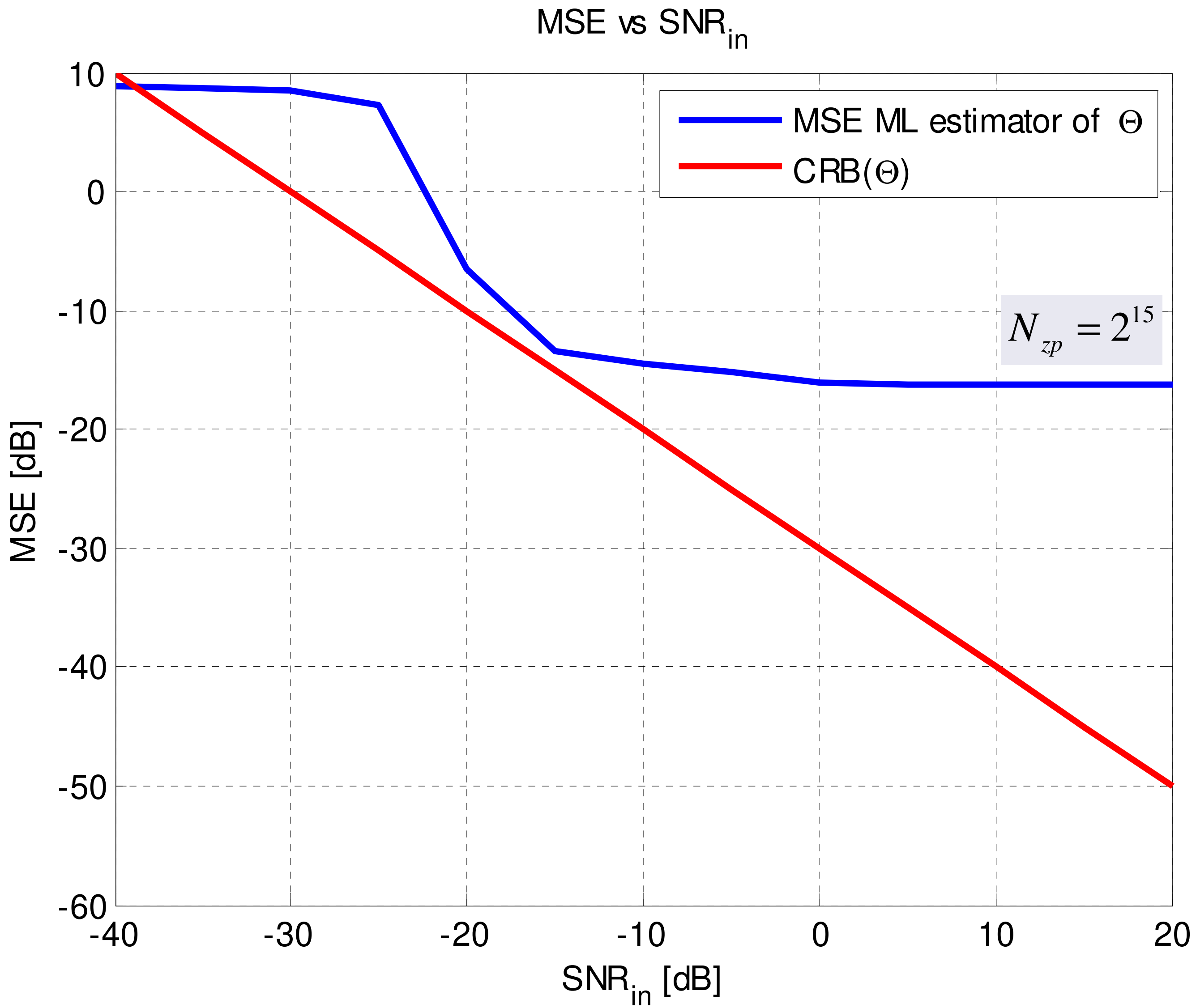
$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$T = 100\ m\sec$

$B = 20\ KHz$

Estimate of the Parameters of a Signal



$A = 10$

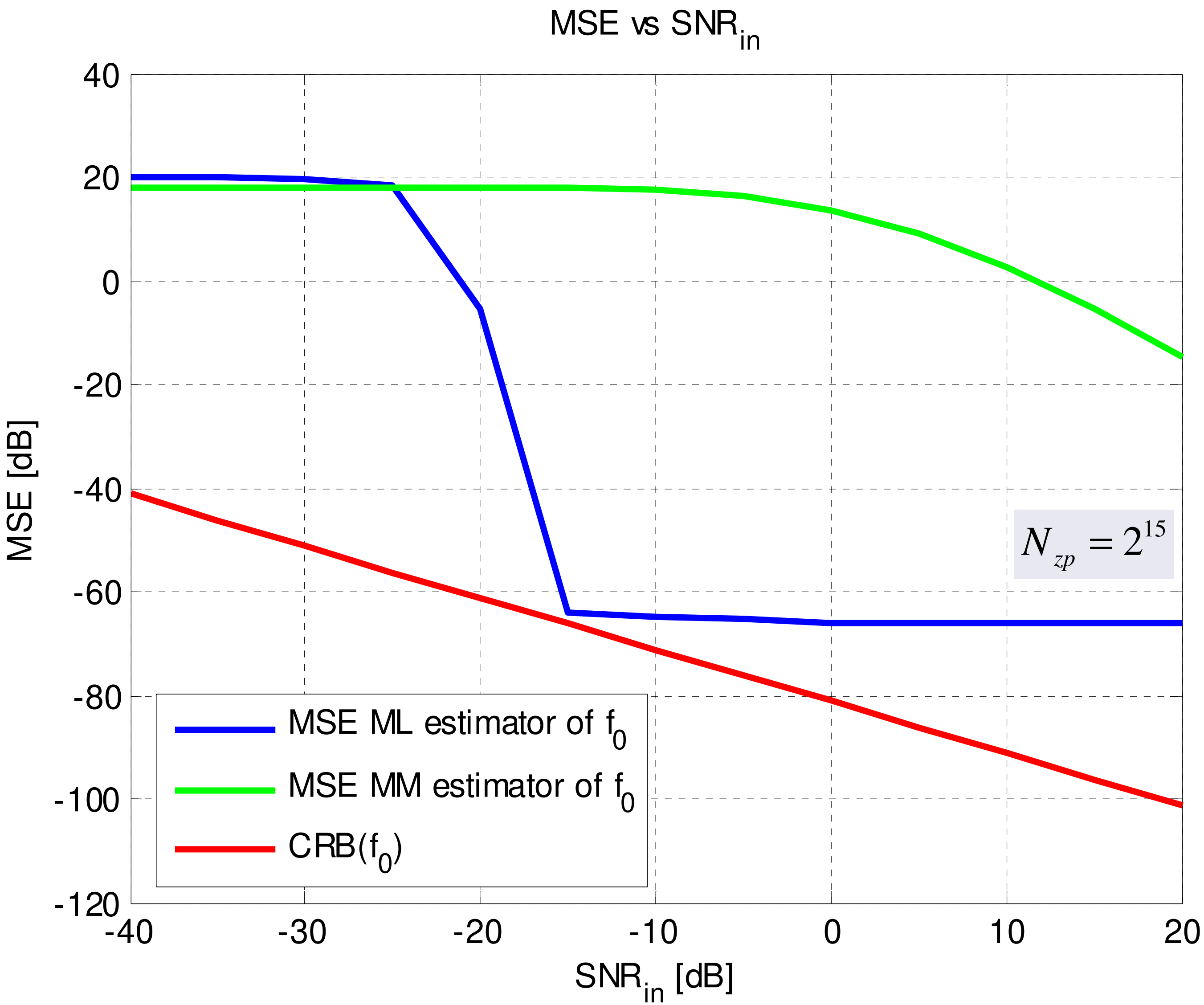
$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$T = 100\ m\sec$

$B = 20\ KHz$

Estimate of the Parameters of a Signal



$A = 10$

$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$T = 100\ m\sec$

$B = 20\ KHz$

Estimate of the Parameters of a Signal

- The ML estimator of frequency f_0 has high computational complexity, since it requires calculation of FFT of high order and then a search for the location of the maximum.
- An alternative estimator can be derived by the **Method of Moments (MM)**.
- To calculate the moments of the observed data, let us assume that the initial phase θ is uniformly distributed in [-π,π).
- The mean value does not contain any information on f_0:

$$\eta_X[k] \triangleq E\{X_k\} = E\left\{\frac{1}{\sqrt{2B}} X_B\left(\frac{k}{2B}\right)\right\}$$

$$= \frac{A}{\sqrt{2B}} E\{\cos(2\pi F_0 k + \theta)\} + \frac{1}{\sqrt{2B}} E\left\{W_B\left(\frac{k}{2B}\right)\right\}$$

$$= \frac{A}{\sqrt{2B}} \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(2\pi F_0 k + \theta)\, d\theta + \frac{1}{\sqrt{2B}} E\left\{W_B\left(\frac{k}{2B}\right)\right\} = 0$$

Estimate of the Parameters of a Signal

- Let us calculate the observed data Autocorrelation Function (ACF):

$$
\begin{aligned}
R_X[k,k+m] &\triangleq E\{X_k X_{k+m}\} = \frac{1}{2B} E\left\{ X_B\left(\frac{k}{2B}\right) X_B\left(\frac{k+m}{2B}\right)\right\} \\
&= \frac{A^2}{2B}\cdot\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos\left(2\pi F_0 k+\theta\right)\cos\left(2\pi F_0 (k+m)+\theta\right)d\theta \\
&\quad + \frac{1}{2B} E\left\{ W_B\left(\frac{k}{2B}\right) W_B\left(\frac{k+m}{2B}\right)\right\} \\
&= \frac{A^2}{2B}\cdot\frac{1}{2}\cos\left(2\pi F_0 m\right) + \frac{1}{2B}\cdot R_{W_B}\left(\frac{m}{2B}\right) \\
&= \frac{A^2}{4B}\cos\left(2\pi F_0 m\right) + \frac{1}{2B}\cdot N_0 B\cdot \mathrm{sinc}\left(2B\left(\frac{m}{2B}\right)\right) \\
&= \frac{A^2}{4B}\cos\left(2\pi F_0 m\right) + \frac{N_0}{2}\cdot\delta[m]
\end{aligned}
$$

Estimate of the Parameters of a Signal

- The ACF provides the information we are looking for:

$$R_X[m] = \frac{A^2}{4B}\cos(2\pi F_0 m) + \frac{N_0}{2}\cdot\delta[m]$$

- In fact, if we calculate the ACF for lag 0 and 1 we get:

$$R_X[0] = \frac{A^2}{4B} + \frac{N_0}{2} = \frac{A^2}{4B}\left(1 + \frac{2N_0 B}{A^2}\right) = \frac{A^2}{4B}\left(1 + \frac{1}{SNR_{in}}\right), \quad R_X[1] = \frac{A^2}{4B}\cos(2\pi F_0)$$

- If SNR_{in}>>1 we obtain:

$$R_X[0] \cong \frac{A^2}{4B}, \quad \frac{R_X[1]}{R_X[0]} \cong \cos(2\pi F_0) \quad \Rightarrow \quad F_0 \cong \frac{1}{2\pi}\arccos\left(\frac{R_X[1]}{R_X[0]}\right)$$

- So we get an estimate of the frequency through the estimation of the first two values of the observed data ACF.

Estimate of the Parameters of a Signal

- In summary, the **MM estimator** of f_0 is given by:

$$\hat{f}_{0MM} = 2B \cdot \hat{F}_{0MM} = \frac{B}{\pi} \arccos\left(\frac{\hat{R}_X[1]}{\hat{R}_X[0]} \right)$$

- The two values of the ACF are estimated by using the **sample estimator**:

$$R_X[m] = E\{X_k X_{k+m}\} \quad \rightarrow \quad \hat{R}_X[m] = \frac{1}{N-m} \sum_{k=0}^{N-m-1} X_k X_{k+m}$$

where the available data are: $\{X_k\}_{k=0}^{N-1}$

- The sample estimator of the ACF is **unbiased**:

$$E\{\hat{R}_X[m]\} = \frac{1}{N-m} \sum_{k=0}^{N-m-1} E\{X_k X_{k+m}\} = \frac{1}{N-m} \sum_{k=0}^{N-m-1} R_X[m] = R_X[m]$$

Estimate of the Parameters of a Signal

- Hence, the final expression of the **MM estimator** of f_0 is:

$$\hat{f}_{0MM} = 2B \cdot \hat{F}_{0MM} = \frac{B}{\pi} \arccos\left(\frac{\hat{R}_X[1]}{\hat{R}_X[0]} \right) = \frac{B}{\pi} \arccos\left(\frac{\frac{1}{N-1} \sum_{k=0}^{N-2} X_k X_{k+1}}{\frac{1}{N} \sum_{k=0}^{N-1} X_k^2} \right)$$

- The estimator is **consistent**, provided that the variance of $X_k X_{k+m}$ is finite:

$$\lim_{N\to\infty} \hat{R}_X[m] = \lim_{N\to\infty} \frac{1}{N-m} \sum_{k=0}^{N-m-1} X_k X_{k+m} = E\left\{ X_k X_{k+m} \right\} = R_X[m]$$

- so we have:

$$\lim_{N\to\infty} \hat{f}_{0MM} = \frac{B}{\pi} \arccos\left(\frac{R_X[1]}{R_X[0]} \right) \cong f_0 \quad \text{(under the assumption that } SNR_{in} \gg 1\text{)}$$

Estimate of the Parameters of a Signal

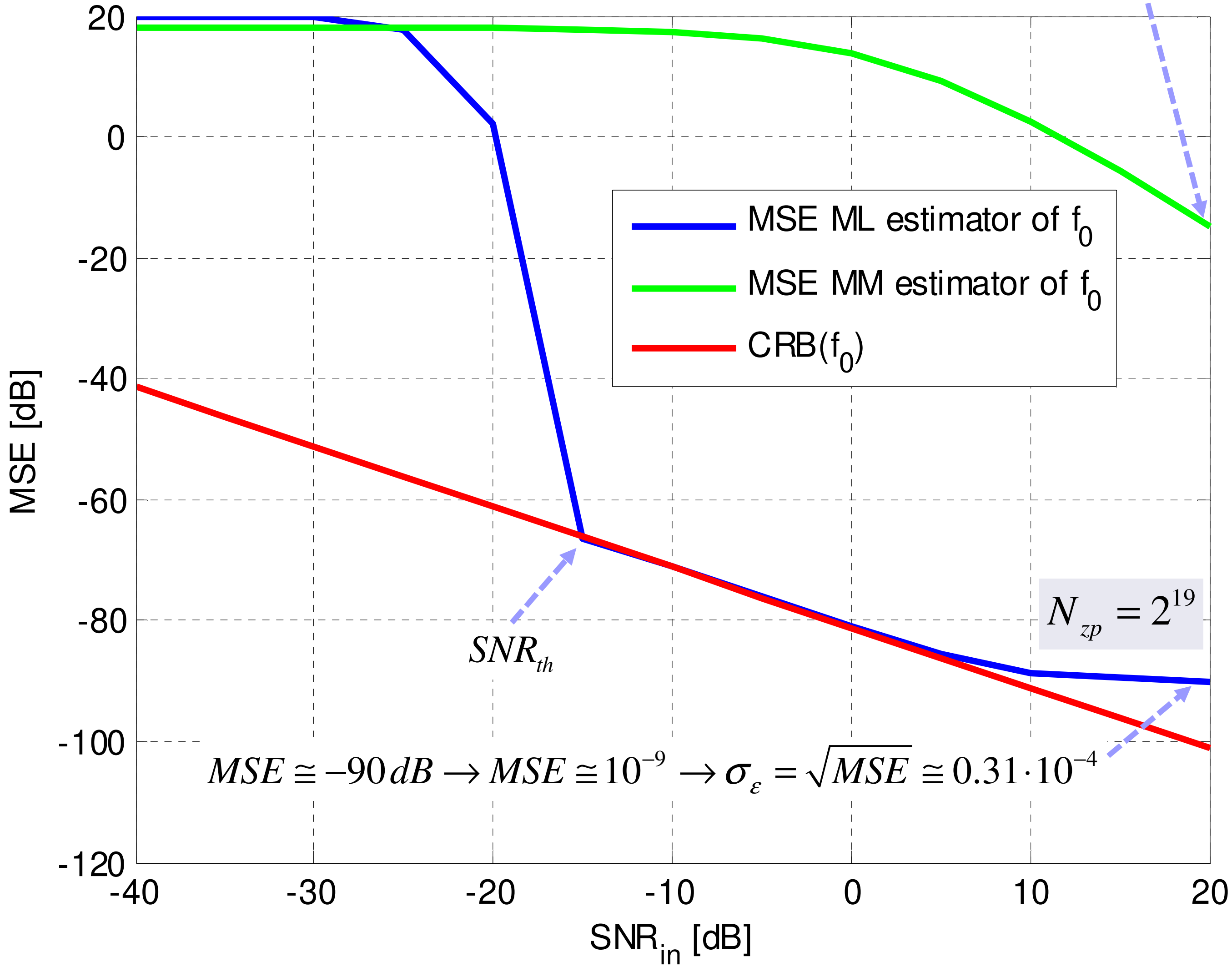
$$\hat{f}_{0MM} = 2B \cdot \hat{F}_{0MM} = \frac{B}{\pi} \arccos\left(\frac{\hat{R}_X[1]}{\hat{R}_X[0]} \right) = \frac{B}{\pi} \arccos\left(\frac{\frac{1}{N-1} \sum_{k=0}^{N-2} X_k X_{k+1}}{\frac{1}{N} \sum_{k=0}^{N-1} X_k^2} \right)$$

- It is also worth observing that the non-ambiguous interval for the *arccos* is π, and not 2π. Hence, the estimate of f_0 is non-ambiguous of $B=F_c/2$ rather than $2B=F_c$, as it is for the Periodogram.

- From the next figures we derive that the MM estimator is not efficient. In fact, its MSE is far from the CRB. However, the **computational complexity** is much lower than that of the ML estimator, so in some applications it can be preferred, thanks to its reduced computational complexity. This typically happens in applications where the processing must be done in real-time.

Estimate of the Parameters of a Signal

$$MSE \cong -18\,dB \rightarrow MSE \cong 0.0158 \rightarrow \sigma_\varepsilon = \sqrt{MSE} \cong 0.1257$$



$$MSE \cong -90\,dB \rightarrow MSE \cong 10^{-9} \rightarrow \sigma_\varepsilon = \sqrt{MSE} \cong 0.31 \cdot 10^{-4}$$

$A = 10$

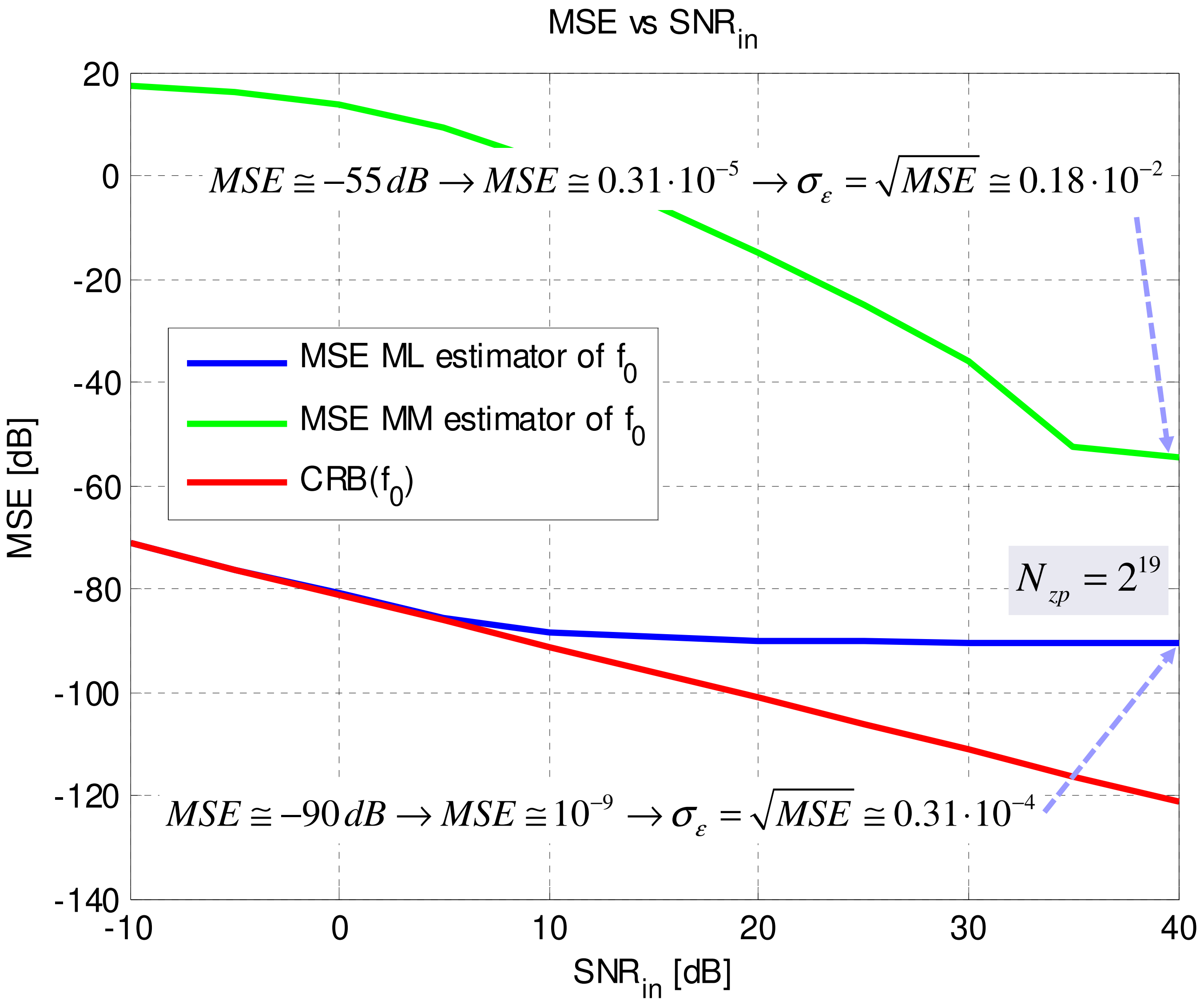
$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$T = 100\ m\sec$

$B = 20\ KHz$

Estimate of the Parameters of a Signal



$A = 10$

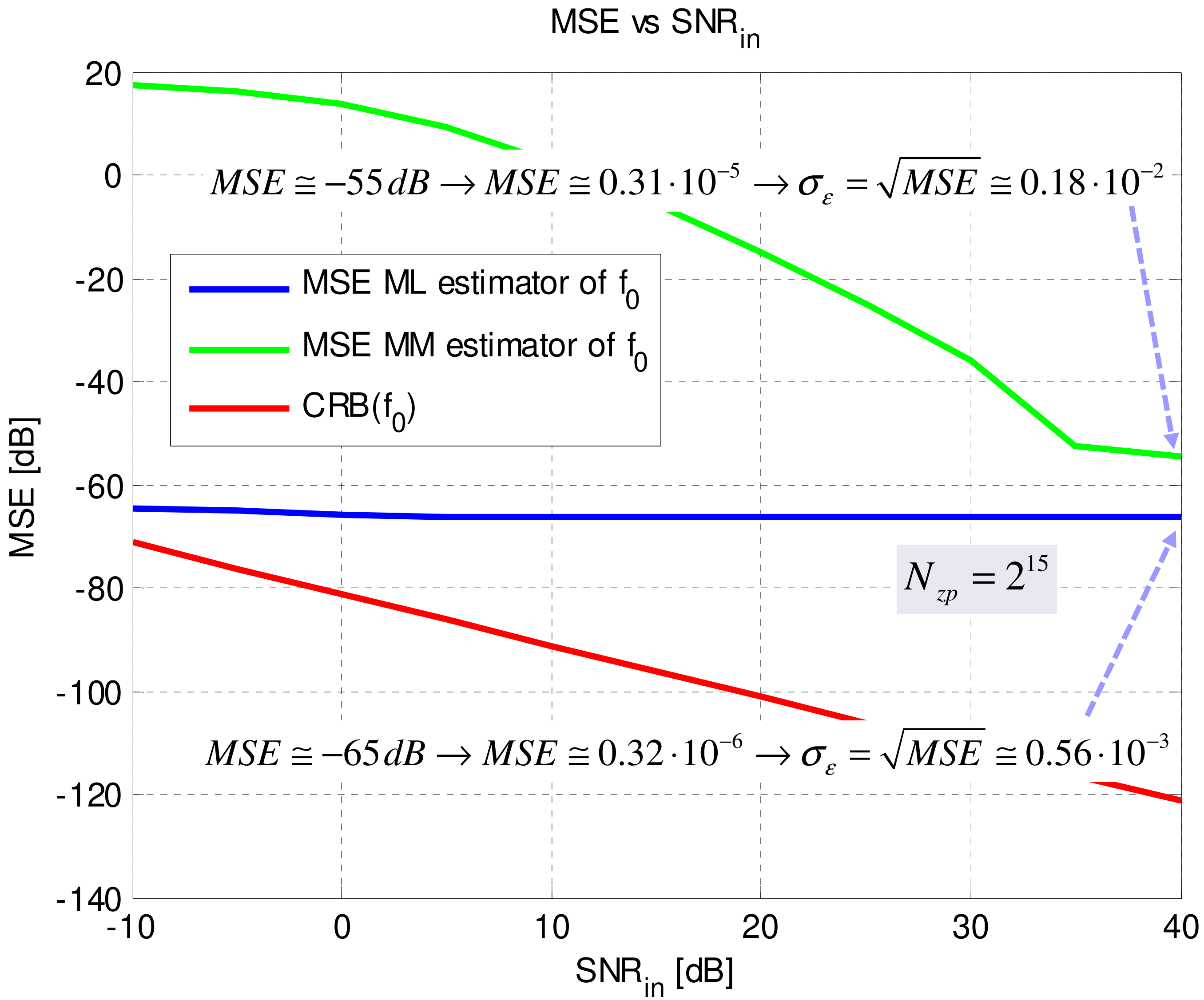
$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$T = 100\ m\sec$

$B = 20\ KHz$

Estimate of the Parameters of a Signal



$A = 10$

$\theta = 2\pi/3$

$f_0 = 2\ KHz$

$T = 100\ m\sec$

$B = 20\ KHz$

Estimate of the Parameters of a Signal

- **Summary on the estimation of the frequency of a cosinusoidal signal**:
- **ML estimate**: efficient if $SNR>SNR_{th}$; the algorithm is not in closed-form; the estimate requires the search for the maximum of the **Periodogram**; computational complexity: medium-high; non-ambiguous estimation interval: $2B=F_c$.
- **MM estimate**: need for $SNR_{in}>>1$, not efficient, the algorithm is in closed-form; low computational complexity, to be preferred when we need a fast estimate; non-ambiguous estimation interval: $B=F_c/2$.
- The estimation accuracy of the MM estimator can be improved by exploiting the sample estimates of the ACF at more lags.
- There are other advanced methods for frequency estimation not investigated here: Pisarenko, MUSIC, ESPRIT, etc.

ML Estimate of the Parameters of a Multi-component Signal

Multicomponent Signal in AWGN

$$X(t) = s(t) + W(t), \quad t \in [0,T)$$

$$s(t) = A_1 \cos\left(2\pi f_1 t + \theta_1\right) + A_2 \cos\left(2\pi f_2 t + \theta_2\right), \quad \text{with } f_1 T \gg 1 \text{ and } f_2 T \gg 1$$

$$W(t)\ AWGN: \quad S_W(f) = \frac{N_0}{2}$$

- The observation interval is of finite duration *T*. We can formally take this into account by introducing a *window* *q*(*t*) in the data model:

$$X(t) = \left[s(t) + W(t)\right] q(t) = s_q(t) + W_q(t), \text{ where } q(t) \triangleq rect\left(\frac{t - T/2}{T}\right)$$

$$s_q(t) \triangleq s(t)q(t) = \left[A_1 \cos\left(2\pi f_1 t + \theta_1\right) + A_2 \cos\left(2\pi f_2 t + \theta_2\right)\right] rect\left(\frac{t - T/2}{T}\right)$$

Multicomponent Signal in AWGN

$$X(t) = s_q(t) + W_q(t), \quad \text{where} \quad q(t) \triangleq rect\left(\frac{t - T/2}{T}\right)$$

$$s_q(t) \triangleq s(t)q(t) = \left[A_1 \cos\left(2\pi f_1 t + \theta_1\right) + A_2 \cos\left(2\pi f_2 t + \theta_2\right)\right] rect\left(\frac{t - T/2}{T}\right)$$

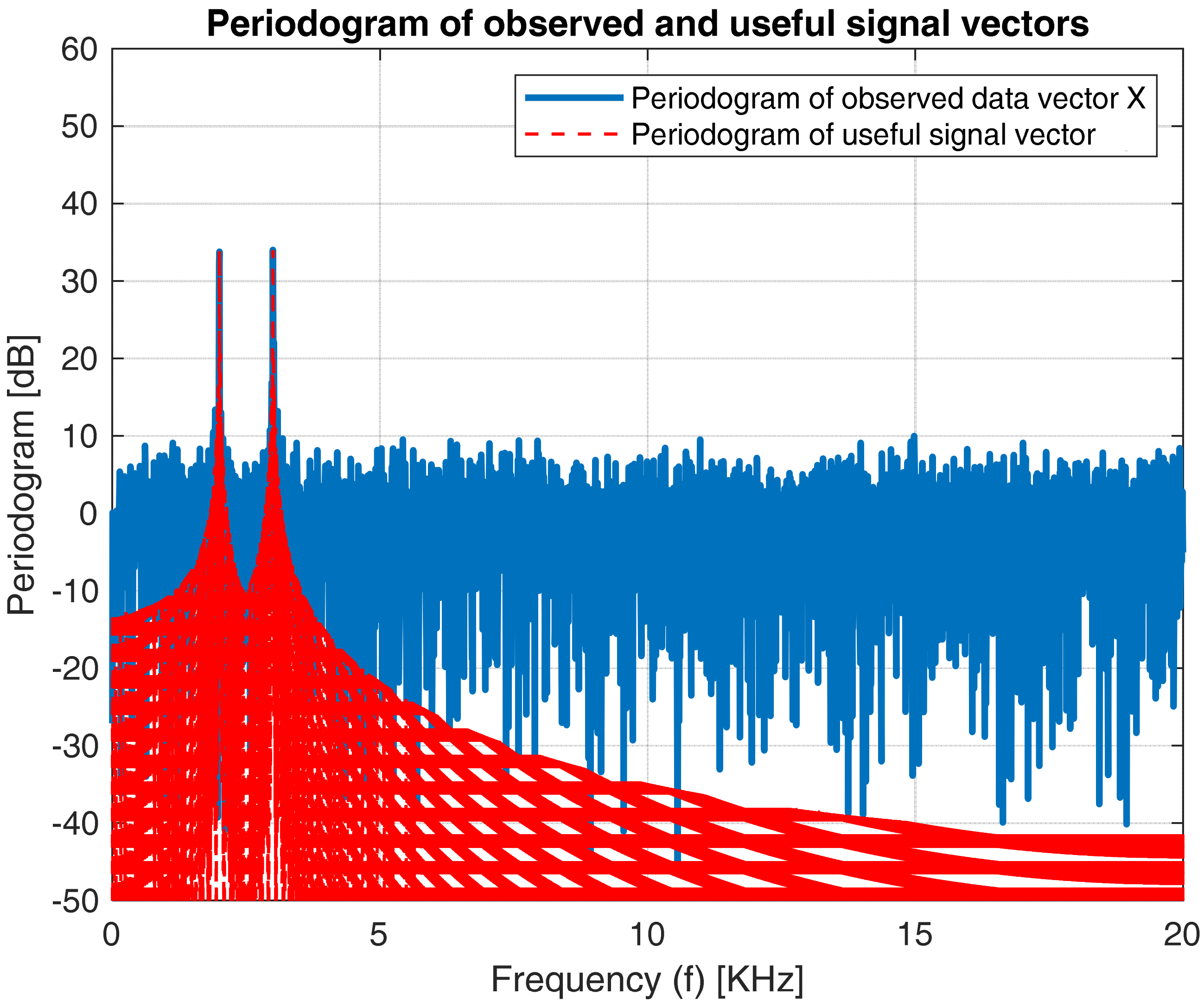
$$S_q(f) = S(f) \otimes Q(f) = S(f) \otimes \left[T sinc\left(fT\right) e^{-j\pi f T}\right]$$

$$f > 0:$$

$$S_q(f) \simeq \frac{TA_1}{2} sinc\left((f - f_1)T\right) e^{j\theta_1} e^{-j\pi(f - f_1)T} + \frac{TA_2}{2} sinc\left((f - f_2)T\right) e^{j\theta_2} e^{-j\pi(f - f_2)T}$$

- **ML estimate of f_1 and f_2**: It can be shown that if $\Delta f = |f_2 - f_1| > 1/T$ the ML estimate of the two frequencies is obtained as the locations of the two main peaks in the Periodogram.

Multicomponent Signal in AWGN



$$A_1 = 10,\ A_2 = 10$$

$$\theta_1 = 2\pi/3,\ \theta_2 = \pi/3$$

$$f_1 = 2\ KHz$$

$$f_2 = 3\ KHz$$

$$\Delta f \gg 1/T$$

$$SNR_{1in} = 0\, dB$$

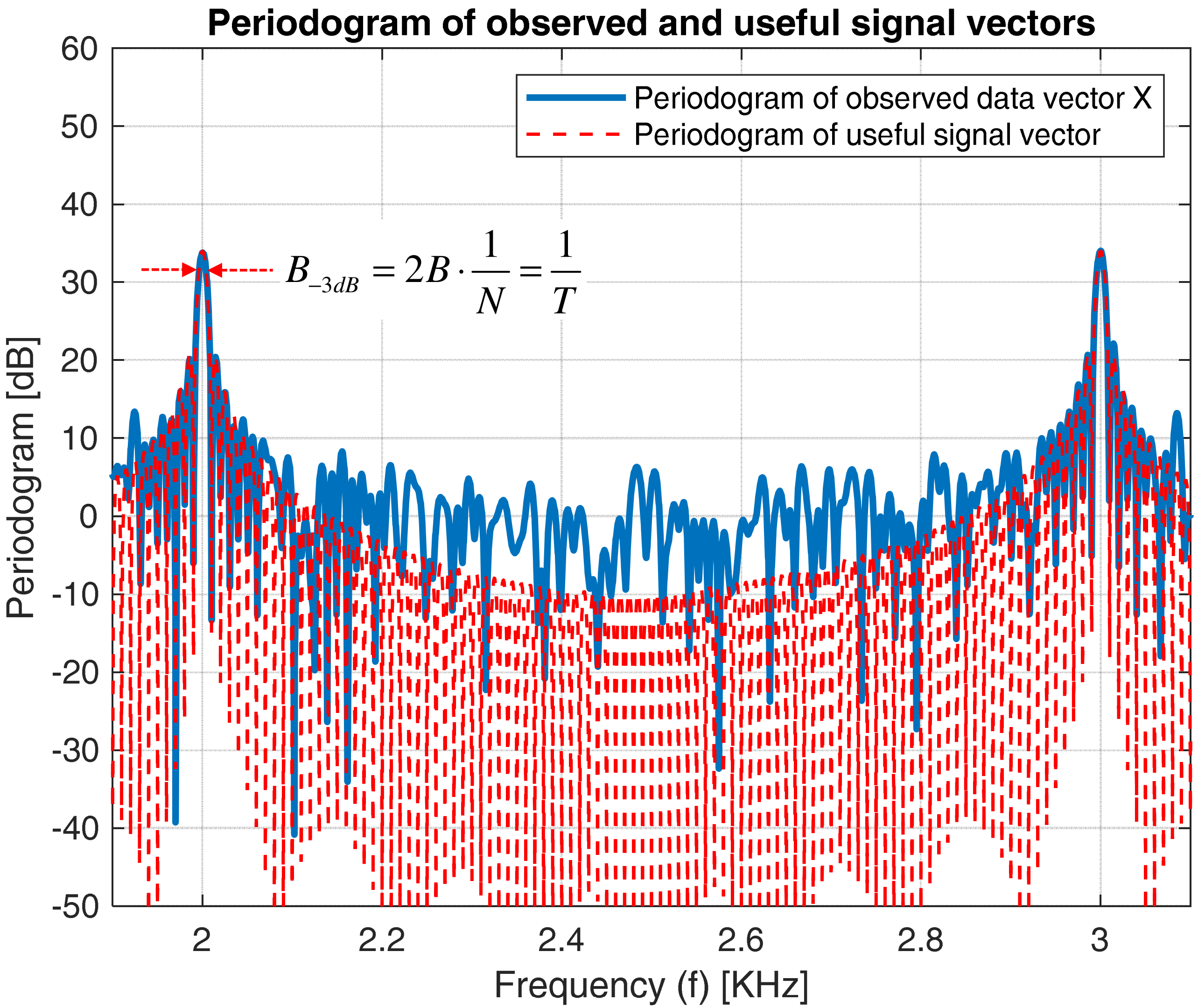
$$T = 100\ m\sec$$

$$B = 20\ KHz$$

$$N = 4 \cdot 10^3$$

$$N_{zp} = 2^{19}$$

Multicomponent Signal in AWGN



$A_1 = 10,\ A_2 = 10$

$\theta_1 = 2\pi/3,\ \theta_2 = \pi/3$

$f_1 = 2\ KHz$

$f_2 = 3\ KHz$

$\Delta f \gg 1/T$

$SNR_{1in} = 0\,dB$

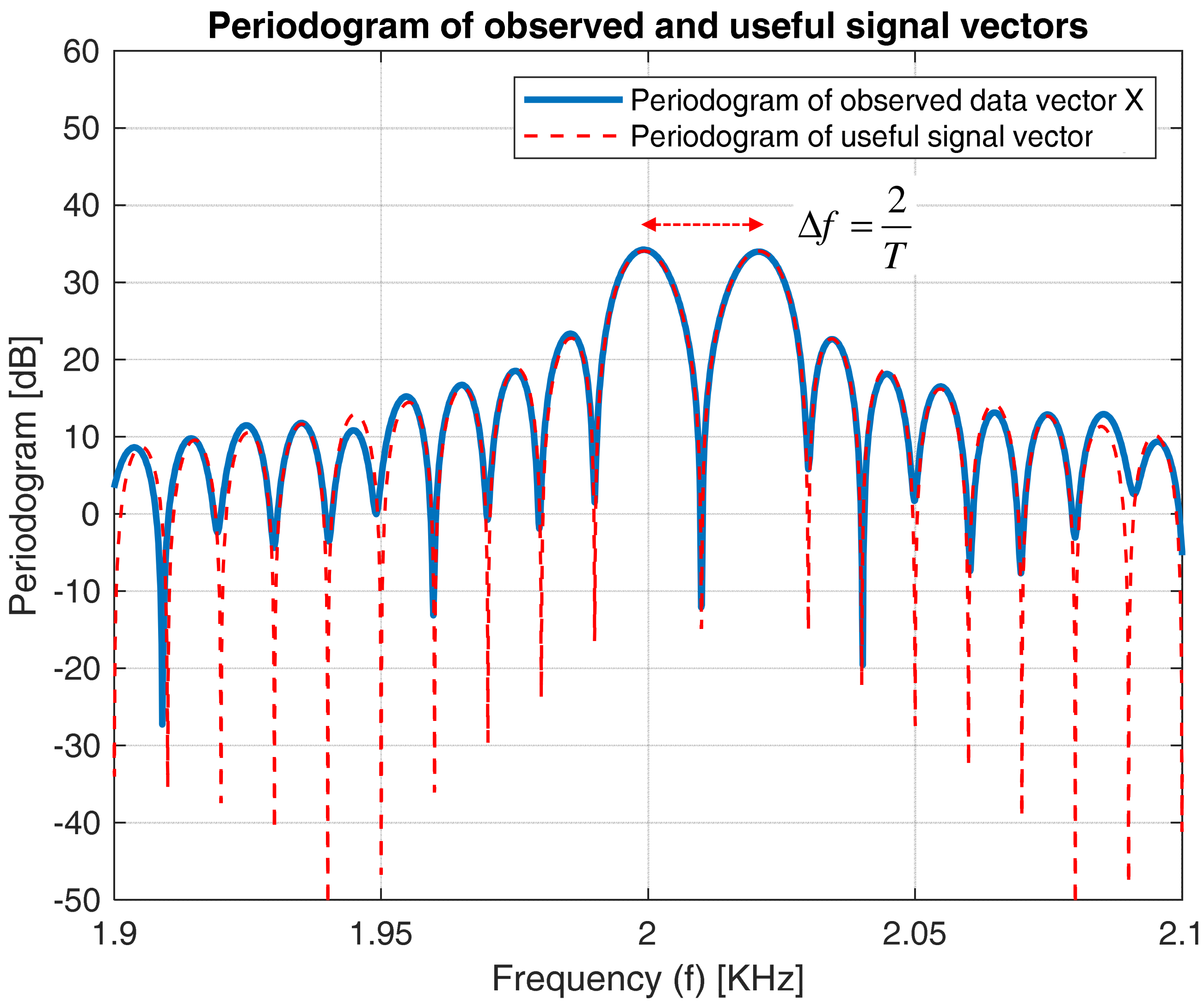
$T = 100\ m\sec$

$B = 20\ KHz$

$N = 4 \cdot 10^3$

$N_{zp} = 2^{19}$

Multicomponent Signal in AWGN



$A_1 = 10,\ A_2 = 10$

$\theta_1 = 2\pi/3,\ \theta_2 = \pi/3$

$f_1 = 2\ KHz$

$f_2 = 2 + 2/T\ KHz$

$\Delta f > 1/T$

$SNR_{1in} = 0\,dB$

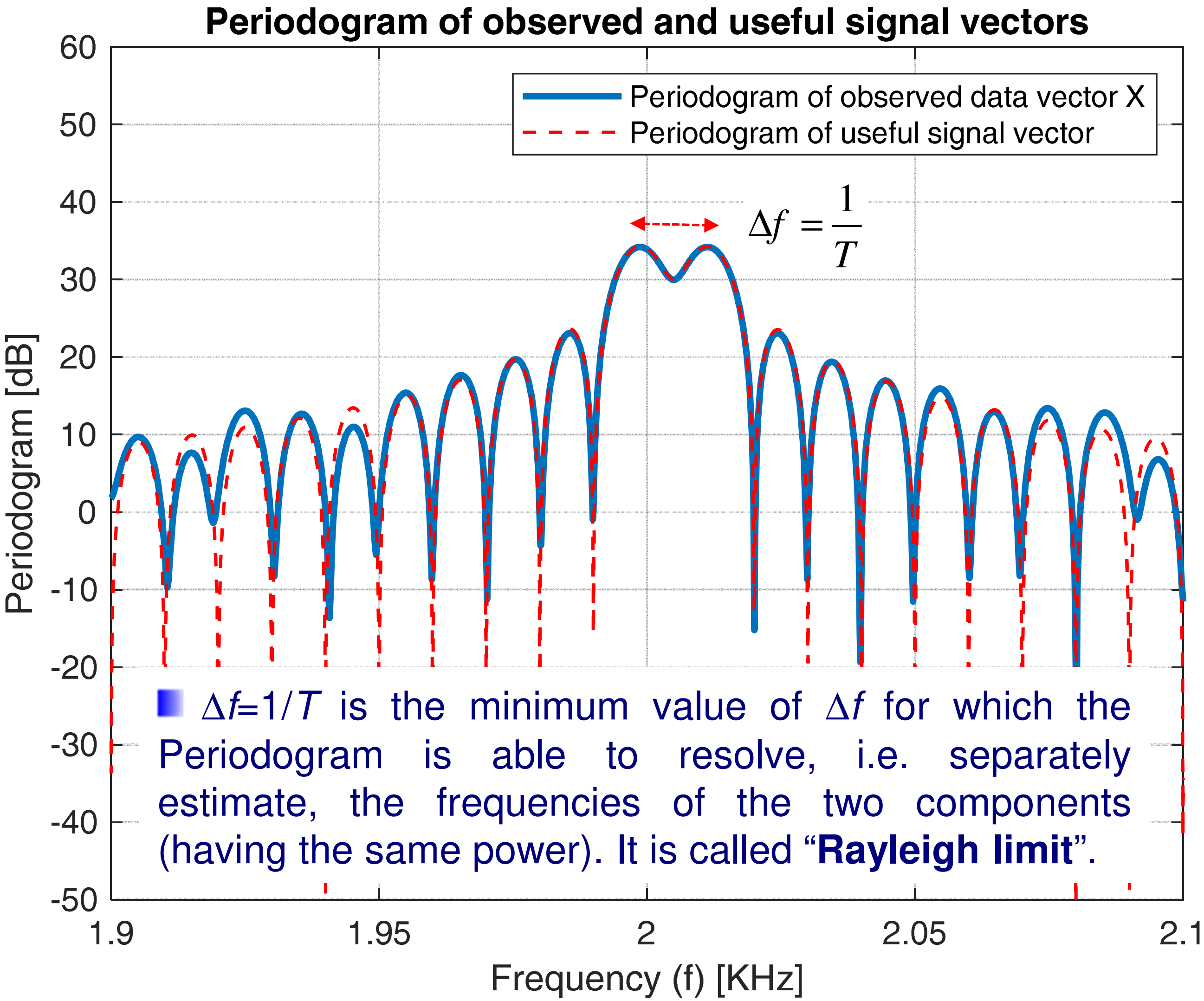
$T = 100\ m\sec$

$B = 20\ KHz$

$N = 4 \cdot 10^3$

$N_{zp} = 2^{19}$

Multicomponent Signal in AWGN



- $\Delta f = 1/T$ is the minimum value of Δf for which the Periodogram is able to resolve, i.e. separately estimate, the frequencies of the two components (having the same power). It is called "**Rayleigh limit**".

$A_1 = 10,\ A_2 = 10$

$\theta_1 = 2\pi/3,\ \theta_2 = \pi/3$

$f_1 = 2\ KHz$

$f_2 = 2 + 1/T\ \ KHz$

$\Delta f = 1/T$

$SNR_{1in} = 0\,dB$

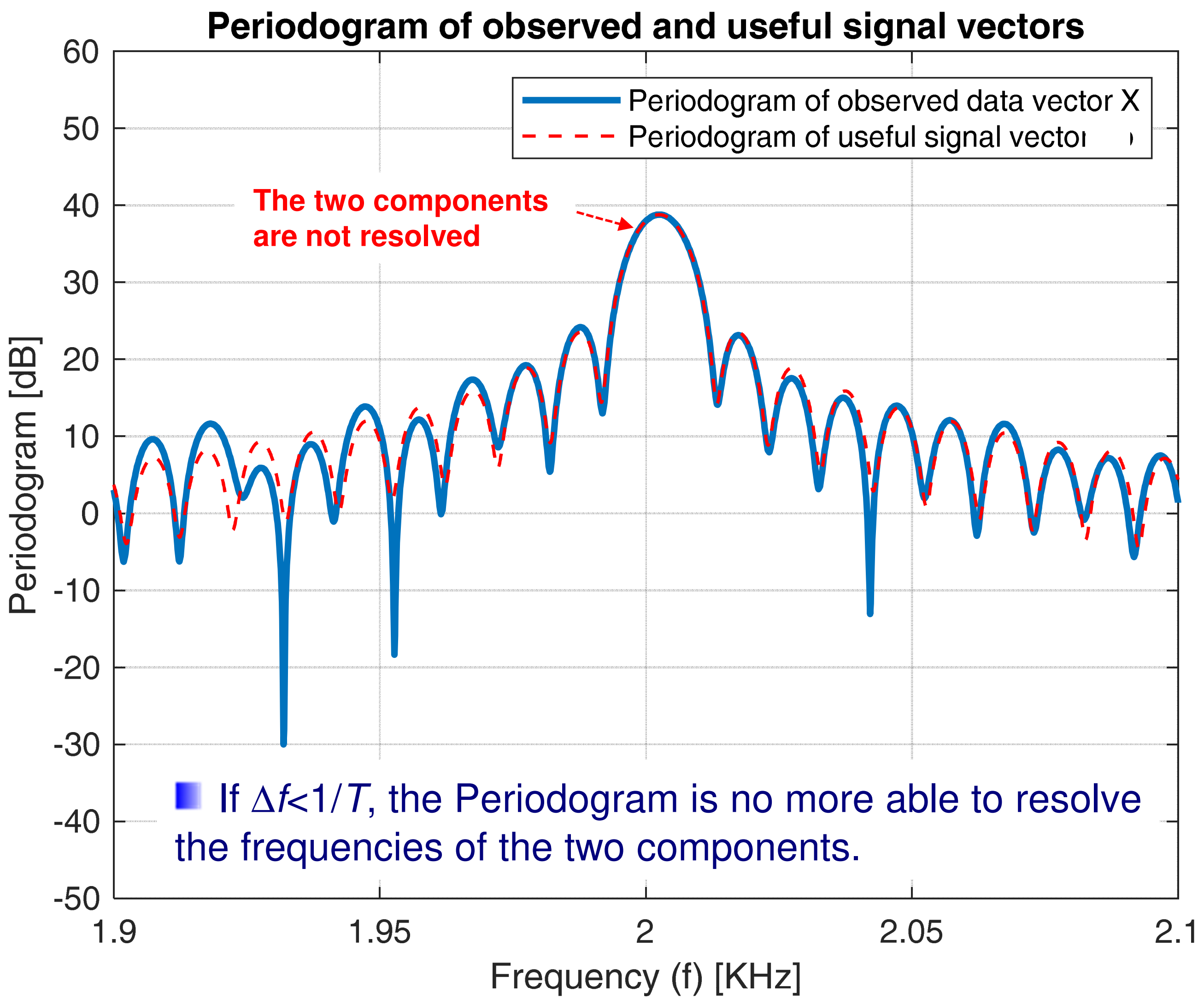
$T = 100\ m\sec$

$B = 20\ KHz$

$N = 4 \cdot 10^3$

$N_{zp} = 2^{19}$

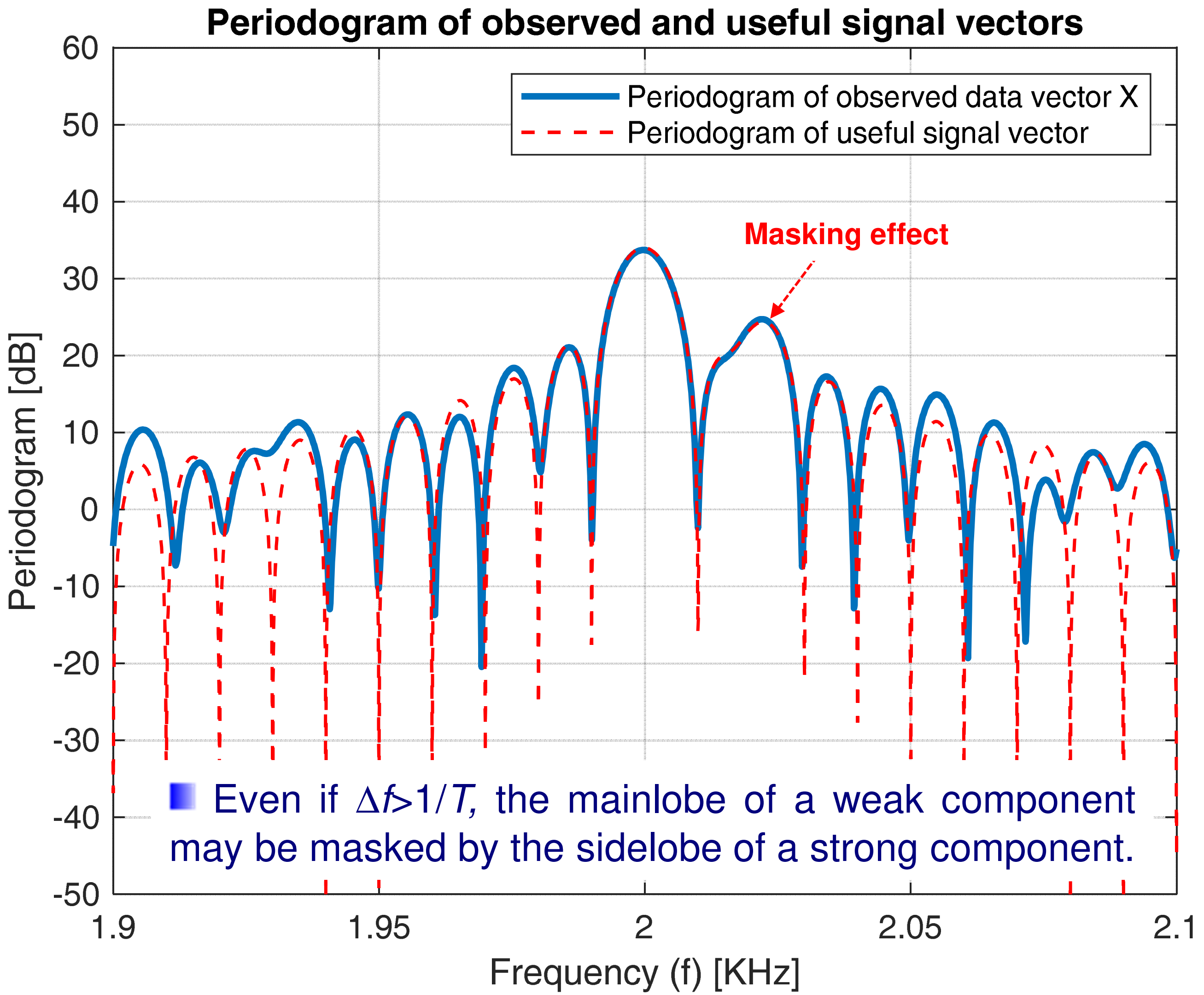
Multicomponent Signal in AWGN



- If $\Delta f < 1/T$, the Periodogram is no more able to resolve the frequencies of the two components.

$$A_1 = 10,\ A_2 = 10$$
$$\theta_1 = 2\pi/3,\ \theta_2 = \pi/3$$
$$f_1 = 2\ KHz$$
$$f_2 = 2 + 1/2T\ \ KHz$$
$$\Delta f = 1/2T < 1/T$$
$$SNR_{1in} = 0\,dB$$
$$T = 100\ m\sec$$
$$B = 20\ KHz$$
$$N = 4 \cdot 10^3$$
$$N_{zp} = 2^{19}$$

Multicomponent Signal in AWGN



$A_1 = 10,\ A_2 = 3$

$\theta_1 = 2\pi/3,\ \theta_2 = \pi/3$

$f_1 = 2\ KHz$

$f_2 = 2 + 2/T\ KHz$

$\Delta f = 2/T > 1/T$

$SNR_{1in} = 0\,dB$

$T = 100\,m\sec$

$B = 20\ KHz$

$N = 4 \cdot 10^3$

$N_{zp} = 2^{19}$

Multicomponent Signal in AWGN

- **Resolution of the Periodogram**: $\Delta f = 1/T$ (**Reyleigh limit**)
- To resolve multiple components that are separated by less than $1/T$ (in the digital frequency domain the Rayleigh limit is $1/N$) we need to use "**superresolution methods**", e.g. Nonlinear Least Squares (NLLS), Pisarenko, MUSIC, ESPRIT, Yule-Walker, Burg, RELAX, etc.
- Once we have estimated the frequencies of the various components, we can estimate amplitudes and phases in the usual way, from the amplitude and phase of the Fourier Transform (FT) of the received signal at the estimated frequencies.

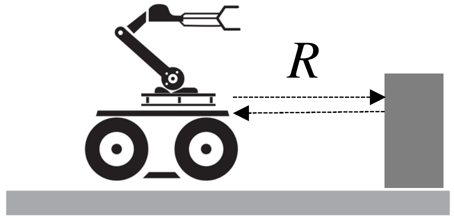
Range/Time-Delay (TD) Estimation

Range/Time Delay Estimation

Given $s(t) \neq 0, \forall t \in [0, T_S)$ and $X(t) = s(t - t_0) + W(t), \quad t \in [0, T)$

Find $t_0 = 2R/c$, where c is the speed of propagation and R is the range

$W(t)\ AWGN: \quad S_W(f) = \dfrac{N_0}{2}$



- The observation interval is of finite duration T, long enough to contain the delayed version of the signal: $T \geq T_s + t_0$.
- Given T, $t_{max} = \max\{t_0\} = T - T_s$ and as a consequence $R_{max} = \max\{R\} = c(T - T_s)/2$.
- Hence, given the R_{max} we want, we select T such that: $T = T_s + t_{max} = T_s + 2R_{max}/c$.
- As usual, we set B such that $2BT >> 1$ and we filter the signal with a LTI low-pass filter (the anti-aliasing filter) with bandwidth B and then we sample at the Nyquist rate $T_c = 1/2B$.

Range/Time Delay Estimation

$$X[n] = \frac{1}{\sqrt{2B}} X(t)\Big|_{t=\frac{n}{2B}} = \frac{1}{\sqrt{2B}} s\left(\frac{n}{2B} - t_0\right) + \frac{1}{\sqrt{2B}} W\left(\frac{n}{2B}\right) = s[n-n_0] + W[n]$$

where $n_0 = \left\lceil \frac{t_0}{1/2B} \right\rceil$ is the delay in samples, $N_S = \left\lceil \frac{T_S}{1/2B} \right\rceil$ is the signal length in samples

$N = \left\lceil \frac{T}{1/2B} \right\rceil$ is the observation interval in samples, i.e. the data size

- Discrete-time data model:

$$X[n] = \begin{cases} s[n-n_0] + W[n], & n \in [n_0, n_0 + N_S - 1] \\ W[n], & n \in [0, n_0 - 1] \cup [n_0 + N_S, N - 1] \end{cases}$$

$$\{W[n]\}_{n=0}^{N-1} \in \mathcal{N}\left(0, \sigma_W^2\right), \text{ IID}, \quad \sigma_W^2 = \frac{N_0}{2}$$

Range/Time Delay Estimation

$$\{X[n]\}_{n=0}^{N-1} \text{ mutually independent}, \quad X_n \triangleq X[n],\ s_n(n_0) \triangleq s[n-n_0]$$

$$f_{\mathbf{X}}(\mathbf{x};n_0) = \prod_{n=0}^{N-1} f_{X_n}(x_n;n_0) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(x_n - s_n(n_0))^2}{N_0}} = (\pi N_0)^{-N/2} e^{-\frac{1}{N_0}\sum_{n=0}^{N-1}(x_n - s_n(n_0))^2}$$

$$\ln L(n_0) \triangleq \ln f_{\mathbf{X}}(\mathbf{x};n_0) = \sum_{n=0}^{N-1} \ln f_{X_n}(x_n;n_0) = \kappa - \frac{1}{N_0}\sum_{n=0}^{N-1}(x_n - s_n(n_0))^2$$

$$\hat{n}_{0ML} = \arg\max_{n_0} \ln L(n_0) = \arg\min_{n_0} \sum_{n=0}^{N-1}(x_n - s_n(n_0))^2 = \arg\min_{n_0} \|\mathbf{x} - \mathbf{s}(n_0)\|^2 = \hat{n}_{0NLLS}$$

$$\sum_{n=0}^{N-1}(x_n - s_n(n_0))^2 = \sum_{n=0}^{N-1} x_n^2 + \sum_{n=0}^{N-1} s_n^2(n_0) - 2\sum_{n=0}^{N-1} x_n s_n(n_0) = E_X + E_S - 2E_{XS}(n_0)$$

Range/Time Delay Estimation

$$\hat{n}_{0ML} = \arg\min_{n_0} \sum_{n=0}^{N-1} \left(x_n - s_n(n_0) \right)^2 = \arg\min_{n_0} \left(E_X + E_S - 2E_{XS}(n_0) \right) = \arg\max_{n_0} E_{XS}(n_0)$$

- The ML estimate of n_0 is the value that maximazes the mutual energy between the received data $X[n]$ and $s[n-n_0]$.

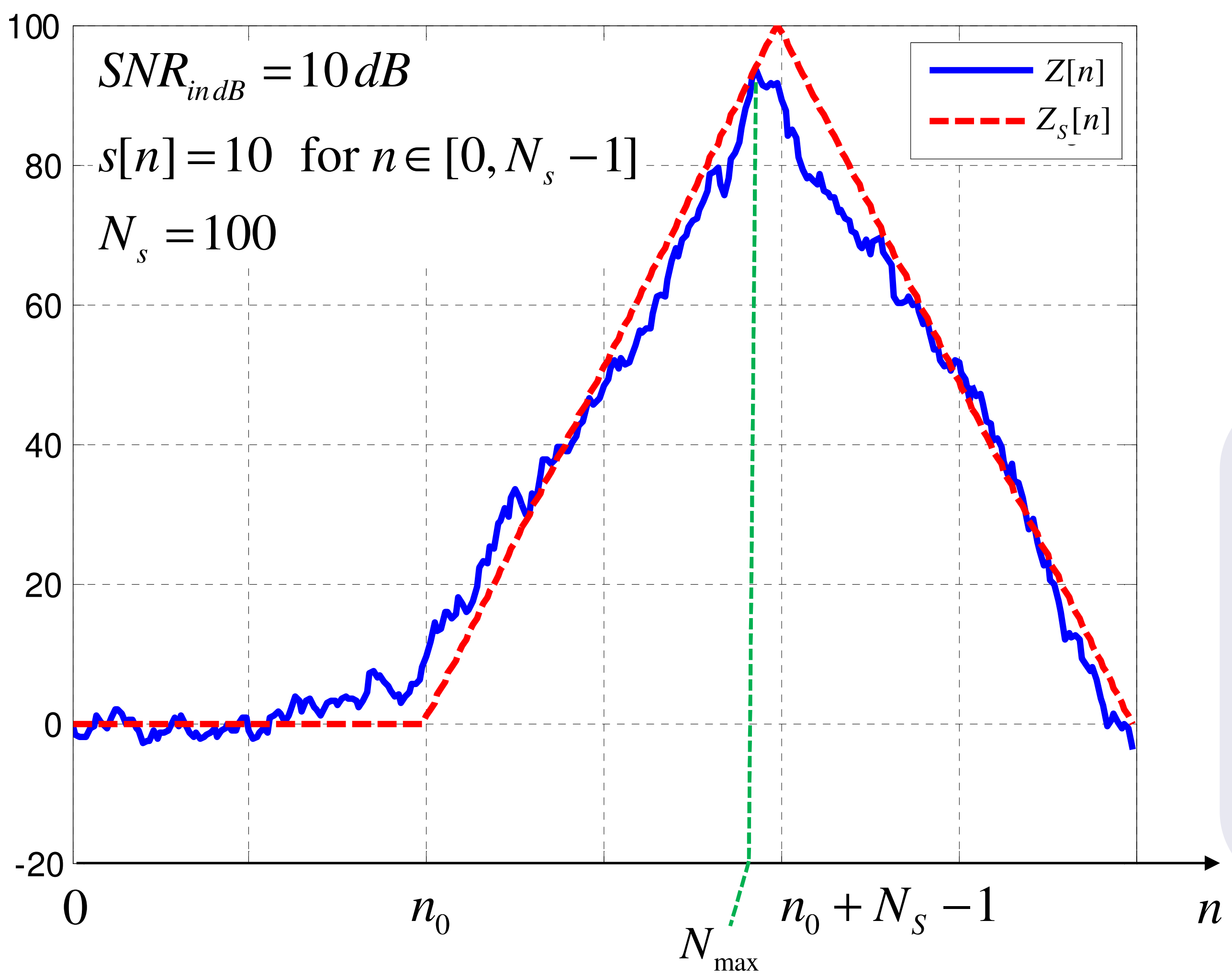
$$E_{XS}(n_0) = \sum_{n=0}^{N-1} x_n s_n(n_0) \text{ is an inner product, i.e. the matched filter output}$$

- We should implement a bank of LTI filters matched to $s[n-n_0]$ for all possible values of n_0 and select the value for which the MF output is maximal.
- In practice, instead of implementing a bank of filters matched to $s[n-n_0]$ for all possible values of n_0, this can be obtained by looking for the time N_{max} when the output of the filter matched to $s[n]$ has the peak:

$$\hat{n}_{0ML} = \arg\max_{n_0} E_{XS}(n_0) = N_{\max} - (N_S - 1)$$

Range/Time Delay Estimation

- For example, if $s[n]$ is a rectangular pulse of length N_S, i.e. from 0 to N_S-1:



$$h[n] = s[N_S - 1 - n]$$

$$Z[n] = h[n] \otimes X[n]$$

$$Z_S[n] = h[n] \otimes s[n - n_0]$$

$$N_{\max} = n_0 + N_S - 1 + \varepsilon$$

$$\hat{n}_{0ML} = N_{\max} - (N_S - 1)$$

$$\hat{R}_{ML} = \frac{c}{2}\left(\frac{\hat{n}_{0ML}}{2B}\right)$$

Range/Time Delay Estimation

- CRB on the estimate of the Time Delay (TD) t_0 and on the range R:

$$\ln L(n_0) \triangleq \ln f_{\mathbf{X}}(\mathbf{x};n_0) = \sum_{n=0}^{N-1} \ln f_{X_n}(x_n;n_0) = \kappa - \frac{1}{N_0}\sum_{n=0}^{N-1}\left(x_n - s_n(n_0)\right)^2$$

$$= \kappa - \frac{1}{N_0}\left\|\mathbf{x} - \mathbf{s}(n_0)\right\|^2 = \kappa - \frac{1}{N_0}\int_0^T \left(x(t) - s(t-t_0)\right)^2 dt \triangleq \ln L(t_0)$$

$$\frac{d\ln L(t_0)}{dt_0} = \frac{2}{N_0}\int_0^T \left(x(t) - s(t-t_0)\right)\frac{ds(t-t_0)}{dt_0}dt$$

$$\frac{d^2\ln L(t_0)}{dt_0^2} = \frac{2}{N_0}\int_0^T \left(x(t) - s(t-t_0)\right)\frac{d^2 s(t-t_0)}{dt_0^2}dt - \frac{2}{N_0}\int_0^T \left(\frac{ds(t-t_0)}{dt_0}\right)^2 dt$$

Range/Time Delay Estimation

$$I(t_0) = -E\left\{\frac{d^2 \ln L(t_0)}{dt_0^2}\right\} = \frac{2}{N_0}\int_0^T \left(\frac{ds(t-t_0)}{dt_0}\right)^2 dt$$

$$MSE\left\{\hat{t}_{0ML}\right\} \geq CRB(t_0) = \frac{1}{I(t_0)} = \frac{N_0/2}{\int_0^T \left(\frac{ds(t-t_0)}{dt_0}\right)^2 dt} = \frac{N_0/2}{\int_{t_0}^{t_0+T_S} \left(\frac{ds(t-t_0)}{dt_0}\right)^2 dt}$$

$$= \frac{N_0/2}{\int_0^{T_S} \left(\frac{ds(t)}{dt}\right)^2 dt} = \frac{N_0/2}{\int_{-\infty}^{+\infty} \left|j2\pi f \cdot S(f)\right|^2 df} = \frac{N_0/2}{E_S \overline{\beta}^2}$$

$$\overline{\beta}^2 \triangleq \frac{\int_{-\infty}^{+\infty} \left|2\pi f \cdot S(f)\right|^2 df}{\int_{-\infty}^{+\infty} \left|S(f)\right|^2 df} = \frac{\int_{-\infty}^{+\infty} (2\pi f)^2 \left|S(f)\right|^2 df}{E_S}$$

Mean Square Bandwidth (MSB)

Range/Time Delay Estimation

$$n_0 = \left\lceil \frac{t_0}{1/2B} \right\rceil, \quad R = \frac{c}{2} t_0, \quad SNR \triangleq \frac{E_S}{\sigma_W^2} = \frac{E_S}{N_0/2}, \quad \hat{R}_{ML} = \frac{c}{2} \hat{t}_{0ML} = \frac{c}{2} \left(\frac{\hat{n}_{0ML}}{2B} \right)$$

$$MSE\left\{ \hat{t}_{0ML} \right\} \geq CRB(t_0) = \frac{N_0/2}{E_S \overline{\beta}^2} = \frac{1}{SNR \cdot \overline{\beta}^2}$$

$$MSE\left\{ \hat{R}_{ML} \right\} \geq CRB(R) = \frac{c^2}{4} \cdot CRB(t_0) = \frac{c^2}{4SNR \cdot \overline{\beta}^2}$$

- We observe that the **range estimation accuracy** improves with the **bandwidth** of the transmitted signal, and obviously also with the *SNR*.
- This CRB does not take the **quantization error** (due to the temporal sampling) into account; in other words, due to the fact that $t_0/(1/2B)$ in general is not an integer number.

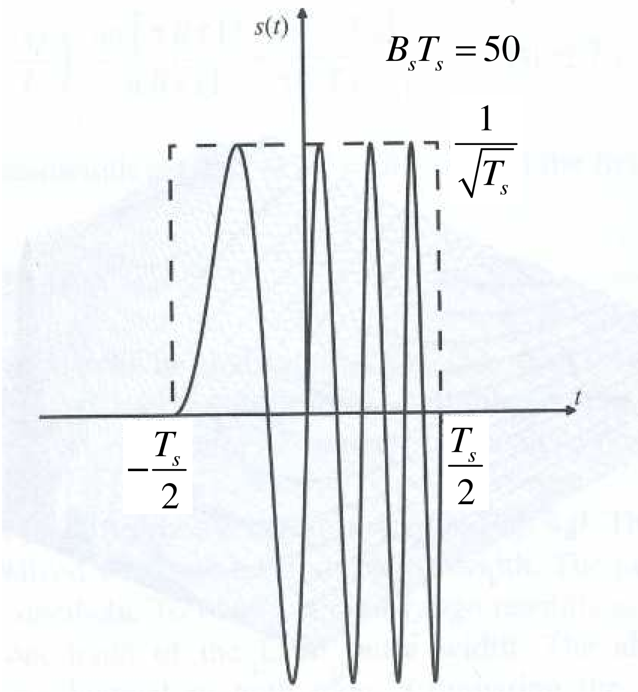
Range/Time Delay Estimation

- There are many methods to increase the **bandwidth** without reducing the pulse length T_s (since reducing T_s would imply reducing E_s).
- The most widely adopted method is to use a **Linearly Frequency Modulated (LFM)** waveform.
- The waveform is commonly referred to as a "**chirp**" waveform because a similar pulse in the audible frequency range produces a chirping sound.
- The basic idea is to sweep the frequency band B_s linearly during the pulse length T_s, as shown in the next page.
- A narrowband **linear-FM (LFM)** pulse of energy E_s is given by:

$$s(t) = \sqrt{\frac{2E_s}{T_s}} \cdot \cos\left(2\pi f_0 t + \pi k t^2\right) \cdot rect\left(\frac{t}{T_s}\right), \quad \text{where} \quad k = \pm \frac{B_s}{T_s}$$

- If k is positive, the pulse is an *upchirp*, if negative it is a *downchirp*.

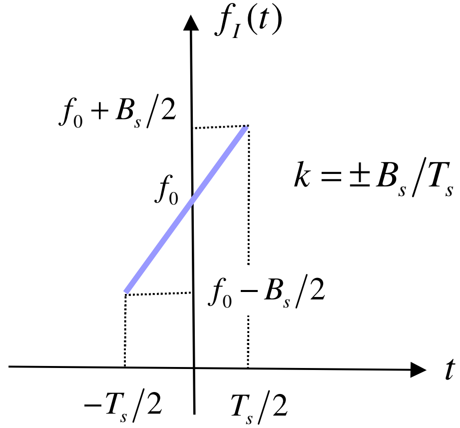
Range/Time Delay Estimation



$$s(t) = \sqrt{\frac{2E_s}{T_s}} \cos\left(2\pi f_0 t + \pi k t^2\right) \cdot rect\left(\frac{t}{T_s}\right)$$

The **instantaneous frequency** $f_I(t)$ is obtained by differentiating the argument $\phi(t)$ of the cosine:

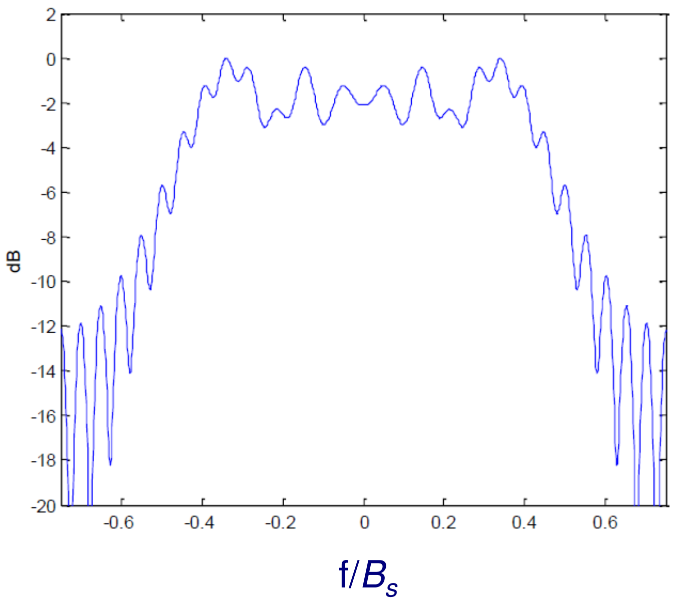
$$f_I(t) \triangleq \frac{1}{2\pi} \cdot \frac{d\phi(t)}{dt} = \frac{1}{2\pi} \cdot \frac{d\left(2\pi f_0 t + \pi k t^2\right)}{dt} = f_0 + kt$$

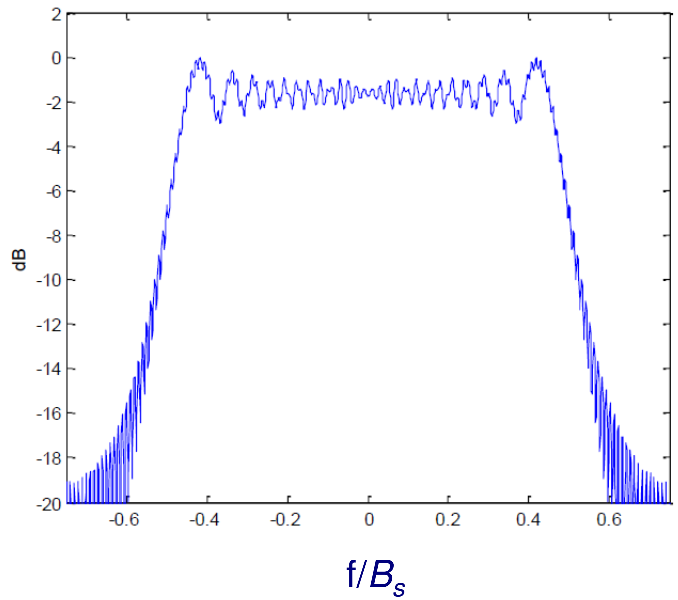


- The instantaneous frequency is a linear function of time.

Range/Time Delay Estimation

- For low $B_s T_s$, the spectrum is relatively poorly defined.
- As the $B_s T_s$ product increases, the spectrum takes on more rectangular shape (around the central frequency f_0). This is intuitively reasonable, because the sweep is linear, the waveform spreads its energy uniformly across the spectrum.





Spectrum of an LFM waveform, *a*) $B_s T_s$=20, *b*) $B_s T_s$=100.

Signal Estimation in Gaussian/Laplace Noise

Gaussian vs Laplace noise

- **Problem**: We observe N samples of a realization of the discrete-time random process $X[n]=A+W[n]$, for $n=0,1,2,\ldots,N-1$, where the signal of interest A is modelled as an unknown deterministic parameter and $\{W[n]\}$ are independent and identically distributed (IID) zero mean random variables with symmetric distribution, modelling the additive noise. The noise power is also unknown.

$$\underset{N\times 1}{\mathbf{X}} = \left[X[0] \quad X[1] \quad \cdots \quad X[N-1] \right]^T, \quad \left\{ X_i \right\}_{i=0}^{N-1} \quad \text{IID}$$

$$m_1 = E\left\{ X_i \right\} = median\left\{ X_i \right\} = A, \qquad \mu_2 = \operatorname{var}\left\{ X_i \right\} = noise\ power$$

- When the noise samples are modelled as zero mean **Gaussian** r.v.s with variance σ^2, we know that the **Maximum Likelihood (ML)** estimators of A (the mean of $X[n]$) and σ^2 (the variance of $X[n]$) are given by the **sample mean** and the **sample variance** of the observed data, and they coincide with the **Method of Moments (MM)** estimators.

Gaussian vs Laplace noise

$$\underset{N\times 1}{\mathbf{X}} = \left[X[0] \quad X[1] \quad \cdots \quad X[N-1] \right]^T, \quad \left\{ X_n \right\}_{n=0}^{N-1} \quad \text{IID}$$

$$\text{Gaussian noise}: \quad f_{X_n}(x_n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-A)^2}{2\sigma^2}}, \quad n = 0,1,\cdots,N-1$$

$$\rightarrow \quad m_1 = E\left\{X_n\right\} = median\left\{X_n\right\} = A, \quad \mu_2 = \text{var}\left\{X_n\right\} = \sigma^2$$

$$\begin{cases} A = m_1 = E\left\{X[n]\right\} \\ \sigma^2 = \mu_2 = E\left\{(X[n]-A)^2\right\} \end{cases} \Rightarrow \begin{cases} \hat{A}_{ML} = \hat{A}_{MM} = \hat{m}_1 = \dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1} X[n] \\ \hat{\sigma}^2_{ML} = \hat{\sigma}^2_{MM} = \hat{\mu}_2 = \dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1}\left(X[n]-\hat{m}_1\right)^2 \end{cases}$$

Gaussian vs Laplace noise

- Let us now investigate the case where the noise is **non Gaussian** and should be modelled by an heavy-tailed distribution, e.g. the **Laplace** distribution:

$$\text{Laplace noise}: \quad f_{X_n}(x_n) = \frac{1}{2\eta} e^{-\frac{|x_n - A|}{\eta}}, \quad n = 0, 1, \cdots, N-1$$

$$\rightarrow \quad m_1 = E\{X_n\} = median\{X_n\} = A, \qquad \mu_2 = \mathrm{var}\{X_n\} = 2\eta^2$$

- The **Method of Moments (MM)** estimators are the same as in the Gaussian case, i.e. they are the **sample mean** and the **sample variance**:

$$\begin{cases} \hat{A}_{MM} = \hat{m}_1 = \dfrac{1}{N} \displaystyle\sum_{n=0}^{N-1} X[n] \\ 2\hat{\eta}_{MM}^2 = \hat{\mu}_2 = \dfrac{1}{N} \displaystyle\sum_{n=0}^{N-1} \left(X[n] - \hat{m}_1 \right)^2 \quad \Rightarrow \quad \hat{\eta}_{MM} = \sqrt{\dfrac{1}{2N} \displaystyle\sum_{n=0}^{N-1} \left(X[n] - \hat{m}_1 \right)^2} \end{cases}$$

Gaussian vs Laplace noise

- However, remember that the **sample** estimators, the **MM** estimators**,** and the **ML** estimators coincide only if the additive noise is white **Gaussian** distributed.
- The **bias** and the **MSE** of the **MM estimator** of A in Laplace noise are given by:

$$E\left\{\hat{A}_{MM}\right\} = \frac{1}{N}\sum_{n=0}^{N-1} E\{X[n]\} = \frac{1}{N}\sum_{n=0}^{N-1} A = A \quad \rightarrow \quad \textit{the MM estimator is unbiased}$$

$$MSE\left\{\hat{A}_{MM}\right\} = \mathrm{var}\left\{\hat{A}_{MM}\right\} = \mathrm{var}\left\{\frac{1}{N}\sum_{n=0}^{N-1} X[n]\right\} = \frac{1}{N^2}\mathrm{var}\left\{\sum_{n=0}^{N-1} X[n]\right\}$$

$$= \frac{1}{N^2}\sum_{n=0}^{N-1}\mathrm{var}\{X[n]\} = \frac{1}{N^2}\sum_{n=0}^{N-1} E\left\{W^2[n]\right\} = \frac{2\eta^2}{N}$$

$$\rightarrow \quad \textit{the MM estimator is consistent}$$

- The bias and the MSE of the MM estimator of η cannot be derived analytically.

Gaussian vs Laplace noise

- Let us now derive the **ML estimators** of *A* and η when the noise is **Laplace** distributed:

$$\text{Laplace noise}: \quad f_{X_n}(x_n) = \frac{1}{2\eta} e^{-\frac{|x_n - A|}{\eta}}, \quad n = 0, 1, \cdots, N-1, \qquad \text{IID}$$

$$f_{\mathbf{X}}(\mathbf{x}; A, \eta) = \prod_{n=0}^{N-1} f_X(x_n; A, \eta) = \left(\frac{1}{2\eta}\right)^N \exp\left(-\frac{1}{\eta}\sum_{n=0}^{N-1}|x_n - A|\right)$$

$$\ln LF(A, \eta) = \ln f_{\mathbf{X}}(\mathbf{x}; A, \eta) = -N\ln(2) - N\ln(\eta) - \frac{1}{\eta}\sum_{n=0}^{N-1}|x_n - A|$$

- The ML estimator is derived by solving the **likelihood equation**.

Gaussian vs Laplace noise

- $$\frac{d\ln LF(A,\eta)}{dA} = -\frac{1}{\eta}\cdot\frac{d}{dA}\left(\sum_{n=0}^{N-1}\left|x_n - A\right|\right) = \frac{1}{\eta}\sum_{n=0}^{N-1} sign\left(x_n - A\right) = 0$$

$$\sum_{n=0}^{N-1} sign\left(x_n - A\right) = 0$$

- The value of *A* such that the ML equation is satisfied is given by the **Sample Median** of the data, i.e. we order the data vector ***X***, e.g. with increasing order, and we select the median value. Assuming that *N* is an odd integer number, the median value is the sample at position (*N*-1)/2:

$$\hat{A}_{ML} = median\{\mathbf{X}\} = Y\left[\frac{N-1}{2}\right] \qquad Sample\ Median$$

- where ***Y*** is the ordered data vector such that *Y*[0]≤*Y*[1]≤ … ≤ *Y*[*N*-2]≤*Y*[*N*-1].

Gaussian vs Laplace noise

- The ML estimator of *A* in Laplace noise does not require knowledge of the parameter η, i.e. they are decoupled; it is **nonlinear** and requires an ordering of the data vector, so the computational complexity is higher than that of the MM estimator, which is linear.
- Now, let us derive the ML estimator of the parameter η, which provides the ML estimate of the noise power: $\hat{\mu}_{2ML} = 2\hat{\eta}_{ML}^2$

$$\frac{d \ln LF(A,\eta)}{d\eta} = -\frac{N}{\eta} + \frac{1}{\eta^2}\sum_{n=0}^{N-1}\left|x_n - A\right| = 0$$

$$\eta = \frac{1}{N}\sum_{n=0}^{N-1}\left|x_n - A\right| \quad \rightarrow \quad \hat{\eta}_{ML} = \frac{1}{N}\sum_{n=0}^{N-1}\left|x_n - \hat{A}_{ML}\right|$$

- which is the so-called **Mean Absolute Deviation (MAD) from the Median**.

Gaussian vs Laplace noise

- To derive analytically the **bias** and the **MSE** of the **ML estimator** of *A* and η in **Laplace noise** is a very difficult task.
- We can derive bias and MSE of the ML estimators by Monte Carlo simulation.
- Let us derive the **Cramér-Rao lower bounds (CRB)** for *A* and η, which provide the asymptotic MSE of the ML estimators:

$$\frac{d^2 \ln LF(A,\eta)}{dA^2} = \frac{d}{dA}\left(\frac{d \ln LF(A,\eta)}{dA}\right) = \frac{1}{\eta}\sum_{n=0}^{N-1}\frac{d}{dA}\left(sign\left(x_n - A\right)\right) = -\frac{2}{\eta}\sum_{n=0}^{N-1}\delta\left(x_n - A\right)$$

$$I_{11} = -E\left\{\frac{d^2 \ln LF(A,\eta)}{dA^2}\right\} = -E\left\{-\frac{2}{\eta}\sum_{n=0}^{N-1}\delta\left(x_n - A\right)\right\} = \frac{2}{\eta}\sum_{n=0}^{N-1}\left[\int_{-\infty}^{+\infty}\delta\left(x_n - A\right) f_X(x_n)dx_n\right]$$

$$= \frac{2}{\eta}\sum_{n=0}^{N-1} f_X(A) = \frac{2N}{\eta} f_X(A) = \frac{2N}{\eta}\cdot\frac{1}{2\eta} = \frac{N}{\eta^2}$$

$$\frac{d^2 \ln LF(A,\eta)}{d\eta^2} = \frac{d}{d\eta}\left(\frac{d \ln LF(A,\eta)}{d\eta}\right) = \frac{d}{d\eta}\left(-\frac{N}{\eta} + \frac{1}{\eta^2}\sum_{n=0}^{N-1}|x_n - A|\right)$$

$$= \frac{N}{\eta^2} - \frac{2}{\eta^3}\sum_{n=0}^{N-1}|x_n - A|$$

$$I_{22} = -E\left\{\frac{d^2 \ln LF(A,\eta)}{d\eta^2}\right\} = -\frac{N}{\eta^2} + \frac{2}{\eta^3}\sum_{n=0}^{N-1} E\{|x_n - A|\}$$

- We know that if X~*Laplace*(A,η) then $Y=|X-A|$~*Exp*$(1/\eta)$, i.e.

$$f_X(x) = \frac{1}{2\eta} e^{-\frac{|x-A|}{\eta}} \overset{Y=|X-A|}{\Rightarrow} f_Y(y) = \frac{1}{\eta} e^{-\frac{y}{\eta}} u(y), \quad E\{Y\} = \eta$$

$$I_{22} = -\frac{N}{\eta^2} + \frac{2}{\eta^3}\sum_{n=0}^{N-1} E\{|x_n - A|\} = -\frac{N}{\eta^2} + \frac{2N\eta}{\eta^3} = \frac{N}{\eta^2}$$

Gaussian vs Laplace noise

$$\frac{d^2 \ln LF(A,\eta)}{dAd\eta} = \frac{d}{dA}\left(\frac{d \ln LF(A,\eta)}{d\eta}\right) = \frac{d}{dA}\left(-\frac{N}{\eta} + \frac{1}{\eta^2}\sum_{n=0}^{N-1}|x_n - A|\right)$$

$$= -\frac{1}{\eta^2}\sum_{n=0}^{N-1} sign\{x_n - A\}$$

$$I_{12} = -E\left\{\frac{d^2 \ln LF(A,\eta)}{dAd\eta}\right\} = \frac{1}{\eta^2}\sum_{n=0}^{N-1} E\{sign\{x_n - A\}\}$$

$$= \frac{1}{\eta^2}\sum_{n=0}^{N-1}\left[1 \cdot \Pr\{x_n > A\} + (-1) \cdot \Pr\{x_n < A\}\right]$$

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

- Hence, FIM is diagonal and the two parameters are decoupled.

Gaussian vs Laplace noise

- The CRBs on the estimate of parameters *A* and η are given by:

$$CRB(A) = \frac{1}{I_{11}} = \frac{\eta^2}{N}, \qquad CRB(\eta) = \frac{1}{I_{22}} = \frac{\eta^2}{N}$$

- Remember that:

$$MSE\left\{\hat{A}_{MM}\right\} = \frac{2\eta^2}{N} = 2 \cdot CRB(A)$$

- so the MM estimator is not efficient, it is not even asymptotically efficient; whereas we know that the ML estimator is always asymptotically efficient. Hence, asymptotically (N>>1) we have

$$MSE\left\{\hat{A}_{ML}\right\} \cong CRB(A) = \frac{\eta^2}{N} = \frac{MSE\left\{\hat{A}_{MM}\right\}}{2}$$

Gaussian vs Laplace noise

- **Summary on the estimate of *A***

- **Gaussian noise**: *Sample estimator = MM estimator = ML estimator*

- Mean value = Median = A → we can use both the Sample Mean and the Sample Median to estimate A (the Sample Median is not the ML estimate, it is more robust to the presence of outliers, but it has higher computational complexity and higher MSE).

- **Laplace noise**: *Sample estimator = MM estimator ≠ ML estimator*

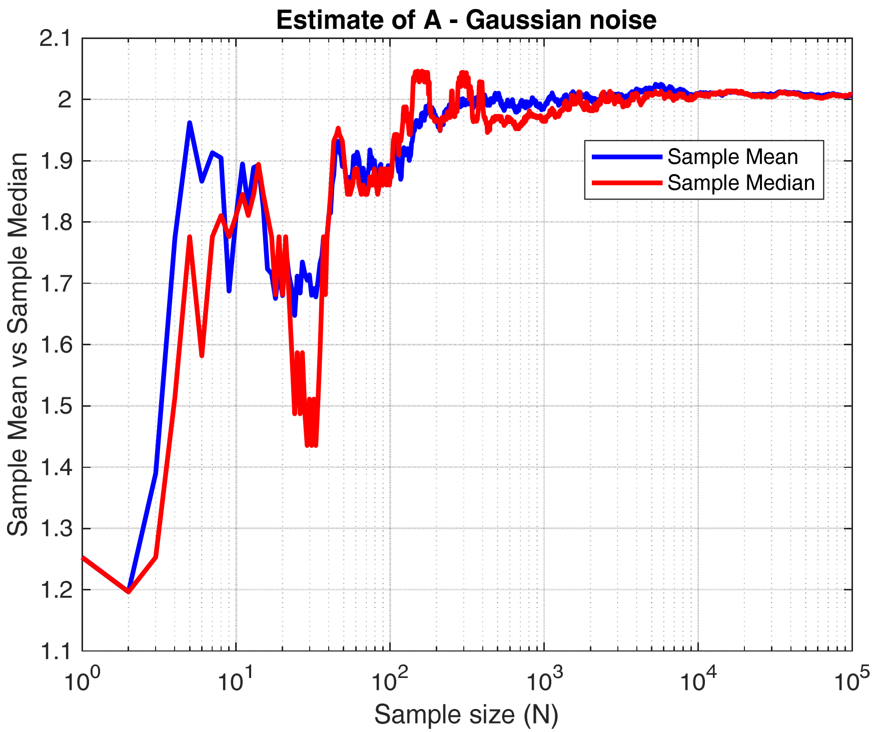
- Mean value = Median = A → we can use both the Sample Mean and the Sample Median to estimate A (the Sample Median is the ML estimate, it has higher computational complexity but lower MSE; the Sample Mean is the MM estimate but not the ML estimate and it has higher MSE than the ML).

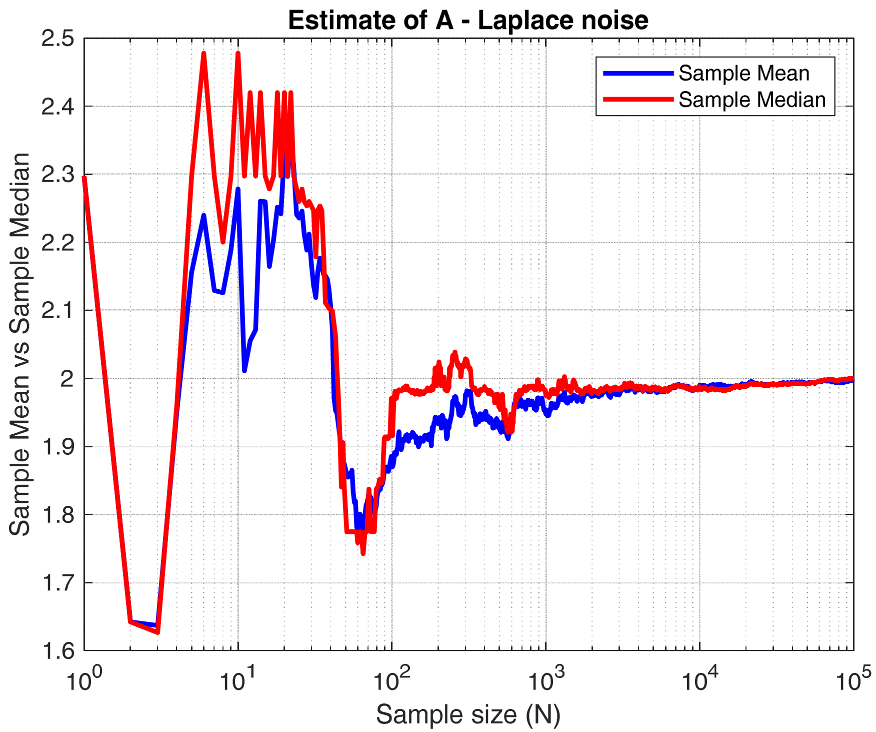
Gaussian vs Laplace noise

Estimate of *A* – Sample Mean vs Sample Median

$$G: A = m_1 = 2, \sigma^2 = \mu_2 = 1$$

$$L: A = m_1 = 2, \eta = 1/\sqrt{2}\ \left[\mu_2 = 2\eta^2 = 1\right]$$



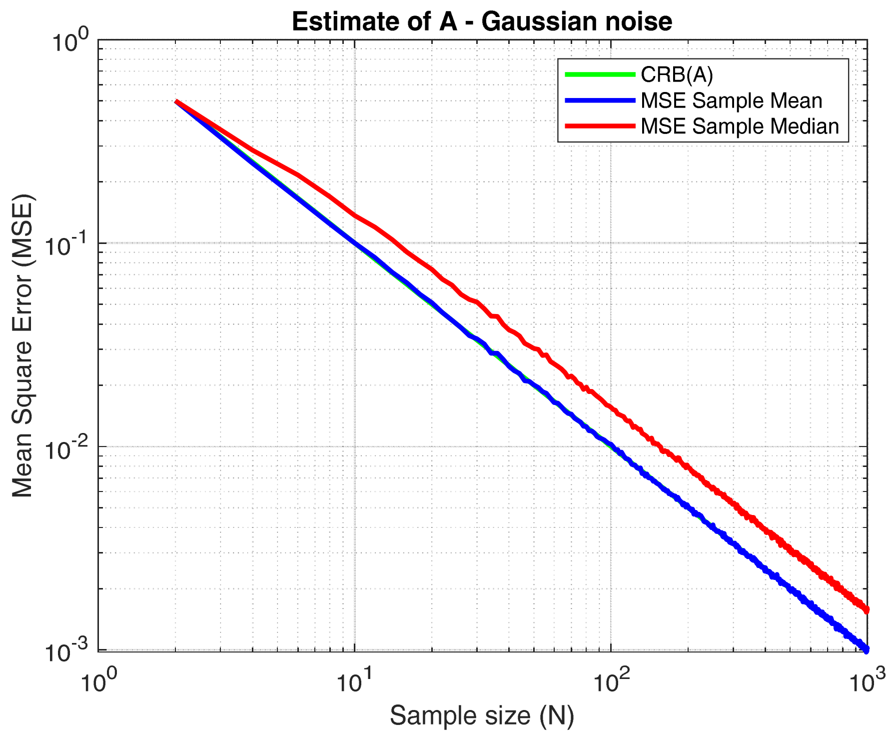


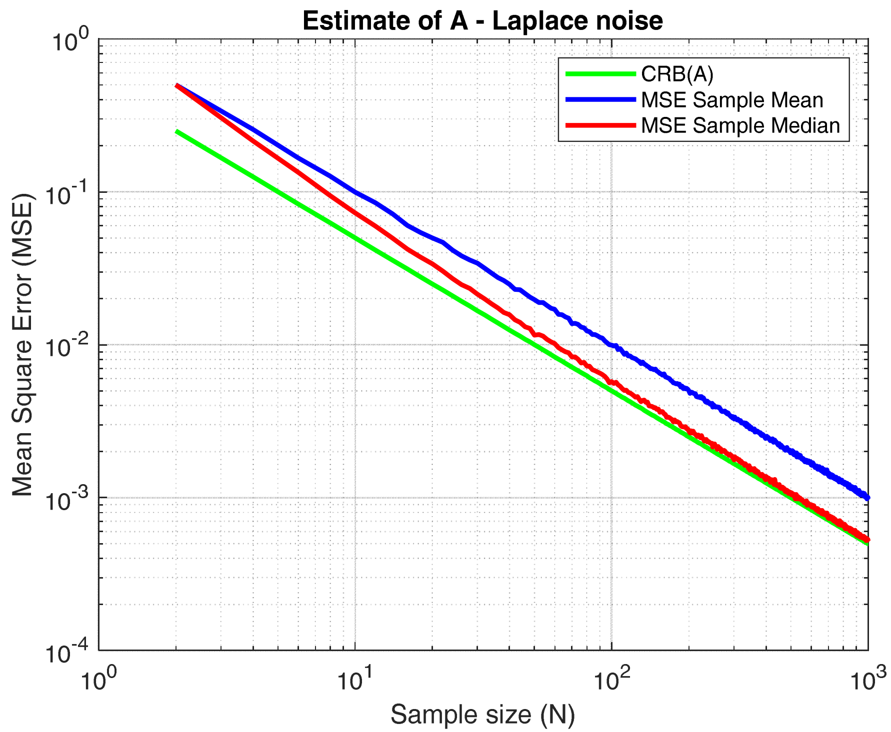
Gaussian vs Laplace noise

Estimate of *A* – Sample Mean vs Sample Median

$$G: A = m_1 = 2, \sigma^2 = \mu_2 = 1$$

$$L: A = m_1 = 2, \eta = 1/\sqrt{2} \; \left[\mu_2 = 2\eta^2 = 1\right]$$





Gaussian vs Laplace noise

Estimate of η - Mean Absolute Deviation (MAD) from the Median

$$L: A = m_1 = 2, \eta = 1/\sqrt{2} \quad \left[\mu_2 = 2\eta^2 = 1\right] \qquad L: A = m_1 = 2, \eta = 10/\sqrt{2} \quad \left[\mu_2 = 2\eta^2 = 100\right]$$