# MP-208
# Optimal Filtering with Aerospace Applications
## Chapter 3: Parameter Estimation

Prof. Dr. Davi Antônio dos Santos
Instituto Tecnológico de Aeronáutica
www.professordavisantos.com

São José dos Campos - SP
2023

# Contents

## Introduction. . .

# Introduction

**Motivation:**

In general, we are interested in two applications of parameter estimation techniques:

- System Identification – estimation of model parameters.
- Sensor Calibration – estimation of sensor parameters.

# Introduction

**Approaches:**

There are two approaches to parameter estimation:

- **Classical Approach:** The parameter to be estimated is modeled as an unknown deterministic constant.
- **Bayesian Approach:** The parameter to be estimated is modeled as a realization of a random variable. In this case, one is supposed to have *a priori* probabilistic information about the parameter.

# Introduction

**General Problem:**

Consider a set of measures $\mathbf{y}_{1:N}$, with $\mathbf{y}_i$ modeled by

$$\mathbf{y}_i = \mathbf{h}_i\left(\theta, \mathbf{v}_i\right), \quad i = 1, 2, ..., N \tag{1}$$

where $\mathbf{h}_i : \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}^m$ is a known function and $\mathbf{v}_i \in \mathbb{R}^n$ is an error vector. The vector $\boldsymbol{\theta} \in \mathbb{R}^p$ contains the unknown parameters that we want to estimate.

In general, the estimator $\hat{\boldsymbol{\theta}} \in \mathbb{R}^p$ of $\boldsymbol{\theta}$ from $\mathbf{y}_{1:N}$ has the form

$$\hat{\boldsymbol{\theta}} = \mathbf{g}\left(\mathbf{y}_{1:N}\right) \tag{2}$$

where $\mathbf{g}$ is a function obtained according to an optimality criterion.

# Introduction

**Criteria:**

The usual criteria for parameter estimation are the following:

- Least Squares (Classical Approach)
- Maximum Likelihood (Classical Approach)
- Maximum *a Posteriori* Probability (Bayesian Approach)
- Minimum Mean Square Error (Bayesian Approach)

# Least Squares. . .

# Least Squares

**Problem Definition:**

Consider a set of measures $\mathbf{y}_{1:N}$, with $\mathbf{y}_i \in \mathbb{R}^m$ modeled by

$$\mathbf{y}_i = \mathbf{h}_i(\boldsymbol{\theta}) + \mathbf{v}_i, \quad i = 1, 2, ..., N \tag{3}$$

where $\mathbf{h}_i : \mathbb{R}^p \rightarrow \mathbb{R}^m$ is a known function and $\mathbf{v}_i \in \mathbb{R}^m$ is an additive error; $\boldsymbol{\theta} \in \mathbb{R}^p$ is the parameter vector.

Consider that $\boldsymbol{\theta}$ is an unknown constant $\rightarrow$ Classical Approach

# Least Squares

The least-squares estimator (LS) $\hat{\boldsymbol{\theta}}_N \in \mathbb{R}^p$ of $\boldsymbol{\theta}$ from $\mathbf{y}_{1:N}$ is given by

$$\hat{\boldsymbol{\theta}}_N = \arg\min_{\boldsymbol{\theta}} J_N(\boldsymbol{\theta}) \tag{4}$$

where

$$J_N(\boldsymbol{\theta}) \triangleq \sum_{i=1}^{N} \left(\mathbf{y}_i - \mathbf{h}_i\left(\boldsymbol{\theta}\right)\right)^{\mathrm{T}} \mathbf{W}_i \left(\mathbf{y}_i - \mathbf{h}_i\left(\boldsymbol{\theta}\right)\right) \tag{5}$$

and $\mathbf{W}_i \in \mathbb{R}^{m \times m}$ is a weighting matrix.

## Explicit Solution for the Linear Model:

Consider that (3) is a linear model in the form

$$\mathbf{y}_i = \mathbf{H}_i\theta + \mathbf{v}_i, \quad i = 1, 2, ..., N \tag{6}$$

In this case, the LS estimator defined in (4) is given explicitly by

$$\hat{\theta}_N = \left(\sum_{i=1}^{N} \mathbf{H}_i^{\mathrm{T}}\mathbf{W}_i\mathbf{H}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{H}_i^{\mathrm{T}}\mathbf{W}_i\mathbf{y}_i \tag{7}$$

**Remark:** Note that we have not established any particular property for the measurement error $\mathbf{v}_i$, $i = 1, ..., N$.

## Maximum Likelihood. . .

# Maximum Likelihood

**Problem Definition:**

Consider a set of measures $\mathbf{y}_{1:N}$, with $\mathbf{y}_i \in \mathbb{R}^m$ modeled by

$$\mathbf{y}_i = \mathbf{h}_i(\boldsymbol{\theta}) + \mathbf{v}_i, \quad i = 1, 2, ..., N \tag{8}$$

where $\mathbf{h}_i : \mathbb{R}^p \to \mathbb{R}^m$ is a known function and $\mathbf{v}_i \in \mathbb{R}^m$ is an additive error; $\boldsymbol{\theta} \in \mathbb{R}^p$ is the parameter vector.

Consider that $\boldsymbol{\theta}$ is an unknown constant $\to$ Classical Approach

# Maximum Likelihood

The maximum likelihood (ML) estimator $\hat{\boldsymbol{\theta}}_N \in \mathbb{R}^p$ of $\boldsymbol{\theta}$ from $\mathbf{y}_{1:N}$ is

$$\hat{\boldsymbol{\theta}}_N = \arg\max_{\boldsymbol{\theta}} \; \Lambda_N(\boldsymbol{\theta}) \tag{9}$$

where $\Lambda_N(\boldsymbol{\theta}) \in \mathbb{R}$ is the likelihood function, which is defined as the joint pdf of $\mathbf{Y}_{1:N}$ given the parameter $\boldsymbol{\theta}$, *i.e.*,

$$\Lambda_N(\boldsymbol{\theta}) \triangleq f_{\mathbf{Y}_{1:N}}(\mathbf{y}_{1:N}; \boldsymbol{\theta}) \tag{10}$$

# Maximum Likelihood

**Explicit Solution for the Linear Gaussian Model:**

Consider that (8) is a linear Gaussian model in the form

$$\mathbf{y}_i = \mathbf{H}_i \theta + \mathbf{v}_i, \quad i = 1, 2, ..., N \tag{11}$$

where $\mathbf{v}_{1:N}$ is a realization of an uncorrelated random sequence $\mathbf{V}_{1:N}$, with marginal distribution $\mathbf{V}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$, and $\mathbf{H}_i \in \mathbb{R}^{m \times p}$ is a known matrix.

In this case, the ML estimator defined in (9) is given explicitly by

$$\hat{\theta}_N = \left( \sum_{i=1}^{N} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \mathbf{H}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \mathbf{y}_i \tag{12}$$

**Remark:** Note that by choosing $\mathbf{W}_i = \mathbf{R}_i^{-1}$, the LS estimator given in (7) coincides with the ML estimator given in (12).

# Maximum Likelihood

**Properties:**

Consider the estimation error RV $\tilde{\boldsymbol{\Theta}}_N \triangleq \hat{\boldsymbol{\Theta}}_N - \boldsymbol{\theta}$. The ML estimator given in (12) has the following properties:

1. Bias:

$$E\left(\tilde{\boldsymbol{\Theta}}_N\right) = \mathbf{0}$$

   In this case, we say that estimator (12) is unbiased.

2. Covariance:

$$E\left(\tilde{\boldsymbol{\Theta}}_N \tilde{\boldsymbol{\Theta}}_N^{\mathrm{T}}\right) = \left(\sum_{i=1}^{N} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \mathbf{H}_i\right)^{-1}$$

   Note that by making $N \rightarrow \infty$, the above expression goes to zero, *i.e.*, $\hat{\boldsymbol{\Theta}}_N \rightarrow \boldsymbol{\theta}$ in the mean square (ms) sense. In this case, we say that the ML estimator is consistent.

# Maximum a Posteriori Probability. . .

# Maximum a Posteriori Probability

## Problem Definition:

Consider a set of measures $\mathbf{y}_{1:N}$, with $\mathbf{y}_i \in \mathbb{R}^m$ modeled by

$$\mathbf{y}_i = \mathbf{h}_i(\boldsymbol{\theta}) + \mathbf{v}_i, \quad i = 1, 2, ..., N \tag{13}$$

where $\mathbf{h}_i : \mathbb{R}^p \rightarrow \mathbb{R}^m$ is a known function and $\mathbf{v}_i \in \mathbb{R}^m$ is an additive error modeled as a realization of a random vector $\mathbf{V}_i$; $\boldsymbol{\theta} \in \mathbb{R}^p$ is the parameter vector.

Consider that $\boldsymbol{\theta}$ is a realization of a RV $\boldsymbol{\Theta}$ with known pdf $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \rightarrow$ Bayesian Approach

# Maximum a Posteriori Probability

The Maximum *a Posteriori* Probability (MAP) estimator $\hat{\boldsymbol{\theta}}_N \in \mathbb{R}^p$ of $\boldsymbol{\theta}$ from $\mathbf{y}_{1:N}$ is given by

$$\hat{\boldsymbol{\theta}}_N = \arg\max_{\boldsymbol{\theta}} \; f_{\boldsymbol{\Theta}|\mathbf{Y}_{1:N}}(\boldsymbol{\theta}|\mathbf{y}_{1:N}) \tag{14}$$

where $f_{\boldsymbol{\Theta}|\mathbf{Y}_{1:N}}(\boldsymbol{\theta}|\mathbf{y}_{1:N})$ is the *a posteriori* pdf given by the Bayes Theorem:

$$f_{\boldsymbol{\Theta}|\mathbf{Y}_{1:N}}(\boldsymbol{\theta}|\mathbf{y}_{1:N}) = \frac{f_{\mathbf{Y}_{1:N}|\boldsymbol{\Theta}}(\mathbf{y}_{1:N}|\boldsymbol{\theta}) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{f_{\mathbf{Y}_{1:N}}(\mathbf{y}_{1:N})} \tag{15}$$

where $f_{\mathbf{Y}_{1:N}|\boldsymbol{\Theta}}(\mathbf{y}_{1:N}|\boldsymbol{\theta})$ is the likelihood function of $\mathbf{Y}_{1:N}$ given $\{\boldsymbol{\Theta} = \boldsymbol{\theta}\}$, $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ is the *a priori* pdf of $\boldsymbol{\Theta}$, and $f_{\mathbf{Y}_{1:N}}(\mathbf{y}_{1:N})$ is a normalizing factor.

# Maximum a Posteriori Probability

## Explicit Solution for the Linear Gaussian Model:

Consider that (13) is a linear Gaussian model in the form

$$\mathbf{y}_i = \mathbf{H}_i\boldsymbol{\theta} + \mathbf{v}_i, \quad i = 1, 2, ..., N \tag{16}$$

where $\boldsymbol{\theta}$ is the realization of a RV $\boldsymbol{\Theta} \sim \mathcal{N}(\mathbf{m}_\Theta, \mathbf{P}_\Theta)$, $\mathbf{v}_{1:N}$ is the realization of an uncorrelated random sequence $\mathbf{V}_{1:N}$ with marginal distribution $\mathbf{V}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$, and $\mathbf{H}_i \in \mathbb{R}^{m \times p}$ is a known matrix.

In this case, the MAP estimator (14) is given explicitly by

$$\hat{\boldsymbol{\theta}}_N = \mathbf{P}_N \mathbf{P}_\Theta^{-1} \mathbf{m}_\Theta + \mathbf{P}_N \sum_{i=1}^{N} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \mathbf{y}_i \tag{17}$$

with

$$\mathbf{P}_N \triangleq \left( \sum_{i=1}^{N} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \mathbf{H}_i + \mathbf{P}_\Theta^{-1} \right)^{-1} \in \mathbb{R}^{p \times p} \tag{18}$$

# Maximum a Posteriori Probability

**Properties:**

Consider the estimation error RV $\tilde{\mathbf{\Theta}}_N \triangleq \hat{\mathbf{\Theta}}_N - \mathbf{\Theta}$. The MAP estimator given in (17) has the following properties:

1. Bias:

$$E\left(\tilde{\mathbf{\Theta}}_N\right) = \mathbf{0}$$

   In this case, estimator (17) is said to be unbiased.

2. Covariance: Define $\check{\mathbf{\Theta}}_N \triangleq \hat{\mathbf{\Theta}}_N - E(\hat{\mathbf{\Theta}}_N)$. The covariance of $\hat{\mathbf{\Theta}}_N$ is

$$E\left(\check{\mathbf{\Theta}}_N \check{\mathbf{\Theta}}_N^{\mathrm{T}}\right) = \mathbf{P}_N \sum_{i=1}^{N} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \mathbf{P}_{Y_i} \mathbf{R}_i^{-1} \mathbf{H}_i \mathbf{P}_N$$

   where $\mathbf{P}_{Y_i} = \mathbf{H}_i \mathbf{P}_{\Theta} \mathbf{H}_i^{\mathrm{T}} + \mathbf{R}_i$.

3 Mean Square Error (MSE):

$$E\left(\tilde{\boldsymbol{\Theta}}_N \tilde{\boldsymbol{\Theta}}_N^{\mathrm{T}}\right) = \mathbf{P}_N \mathbf{P}_{\Theta}^{-1} \mathbf{m}_{\Theta} \mathbf{m}_{\Theta}^{\mathrm{T}} \mathbf{P}_{\Theta}^{-1} \mathbf{P}_N + \mathbf{P}_1 \left(\mathbf{P}_{\Theta} + \mathbf{m}_{\Theta} \mathbf{m}_{\Theta}^{\mathrm{T}}\right) \mathbf{P}_1 + \mathbf{P}_N \mathbf{P}_{\Theta}^{-1} \mathbf{m}_{\Theta} \mathbf{m}_{\Theta}^{\mathrm{T}} \mathbf{P}_1 + \mathbf{P}_1 \mathbf{m}_{\Theta} \mathbf{m}_{\Theta}^{\mathrm{T}} \mathbf{P}_{\Theta}^{-1} \mathbf{P}_N + \mathbf{P}_N \sum_{i=1}^{N} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \mathbf{H}_i \mathbf{P}_N$$

where

$$\mathbf{P}_1 \triangleq \left(\mathbf{P}_N \sum_{i=1}^{N} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \mathbf{H}_i - \mathbf{I}_p\right)$$

Note that estimator (17) is consistent, since the MSE converges to $\mathbf{0}$ as $N \to \infty$, which is equivalent to say that $\hat{\boldsymbol{\Theta}}_N \to \boldsymbol{\Theta}$ (ms).

# Minimum Mean Square Error . . .

# Minimum Mean Square Error

**Problem Definition:**

Consider a set of measures $\mathbf{y}_{1:N}$, with $\mathbf{y}_i \in \mathbb{R}^m$ modeled by

$$\mathbf{y}_i = \mathbf{h}_i(\boldsymbol{\theta}) + \mathbf{v}_i, \quad i = 1, 2, ..., N \tag{19}$$

where $\mathbf{h}_i : \mathbb{R}^p \to \mathbb{R}^m$ is a known function, and $\mathbf{v}_i \in \mathbb{R}^m$ is an additive error modeled as a realization of a RV; $\boldsymbol{\theta} \in \mathbb{R}^p$ is the parameter vector.

Consider that $\boldsymbol{\theta}$ is a realization of a RV $\boldsymbol{\Theta}$ with known pdf $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \to$ Bayesian Approach

# Minimum Mean Square Error

The Minimum Mean Square Error estimator (MMSE) $\hat{\boldsymbol{\theta}}_N \in \mathbb{R}^p$ of $\boldsymbol{\theta}$ from $\mathbf{y}_{1:N}$ is given by

$$\hat{\boldsymbol{\theta}}_N = \arg\min_{\bar{\boldsymbol{\theta}}} \; E\left(\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\Theta}\right)^{\mathrm{T}} \left(\bar{\boldsymbol{\theta}} - \boldsymbol{\Theta}\right) |\mathbf{Y}_{1:N}\right) \tag{20}$$

# Minimum Mean Square Error

**General Solution:**

We can show that, for any measurement model (19), the general solution to problem (20) is given by the conditional mean:

$$\hat{\boldsymbol{\theta}}_N = E\left(\boldsymbol{\Theta}|\mathbf{Y}_{1:N}\right) \tag{21}$$

which is calculated by means of an *a posteriori* pdf given by the Bayes Theorem:

$$f_{\boldsymbol{\Theta}|\mathbf{Y}_{1:N}}(\boldsymbol{\theta}|\mathbf{y}_{1:N}) = \frac{f_{\mathbf{Y}_{1:N}|\boldsymbol{\Theta}}(\mathbf{y}_{1:N}|\boldsymbol{\theta})f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{f_{\mathbf{Y}_{1:N}}(\mathbf{y}_{1:N})} \tag{22}$$

where $f_{\mathbf{Y}_{1:N}|\boldsymbol{\Theta}}(\mathbf{y}_{1:N}|\boldsymbol{\theta})$ is the likelihood function of $\mathbf{Y}_{1:N}$ given $\{\boldsymbol{\Theta} = \boldsymbol{\theta}\}$, $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ is the *a priori* pdf of $\boldsymbol{\Theta}$, and $f_{\mathbf{Y}_{1:N}}(\mathbf{y}_{1:N})$ is a normalizing factor.

# Minimum Mean Square Error

## Explicit Solution for the Linear Gaussian Model:

Consider that (19) is a linear Gaussian model in the form

$$\mathbf{y}_i = \mathbf{H}_i \theta + \mathbf{v}_i, \quad i = 1, 2, ..., N \tag{23}$$

where $\boldsymbol{\theta}$ is a realization of a RV $\boldsymbol{\Theta} \sim \mathcal{N}(\mathbf{m}_\Theta, \mathbf{P}_\Theta)$, $\mathbf{v}_{1:N}$ is a realization of an uncorrelated random sequence $\mathbf{V}_{1:N}$, with marginal distribution $\mathbf{V}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$, and $\mathbf{H}_i \in \mathbb{R}^{m \times p}$ is a known matrix.

In this case, the MMSE estimator (21) is given explicitly by

$$\hat{\boldsymbol{\theta}}_N = \mathbf{P}_N \mathbf{P}_\Theta^{-1} \mathbf{m}_\Theta + \mathbf{P}_N \sum_{i=1}^{N} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \mathbf{y}_i \tag{24}$$

where $\mathbf{P}_N$ is the matrix defined in (18).

# Minimum Mean Square Error

**Properties:**

The MMSE estimator (24) is identical to the MAP estimator (17). This is due to the Gaussianity of $\boldsymbol{\Theta}$ conditioned on $\mathbf{Y}_{1:N}$, *i.e.*,

$$f_{\boldsymbol{\Theta}|\mathbf{Y}_{1:N}}\left(\boldsymbol{\theta}|\mathbf{y}_{1:N}\right) = \mathcal{N}\left(\mathbf{m}_{\Theta|Y}, \mathbf{P}_{\Theta|Y}\right) \tag{25}$$

with

$$\mathbf{m}_{\Theta|Y} = \hat{\boldsymbol{\theta}}_N \tag{26}$$

$$\mathbf{P}_{\Theta|Y} = \mathbf{P}_N \tag{27}$$

Therefore, these estiamators have the same properties.

# Cramér-Rao Lower Bound...

# Cramér-Rao Lower Bound

**For Deterministic Parameters:**

In this case, the Cramér-Rao Lower Bound (CRLB) says that the covariance (or MSE) of an unbiased estimator is lower limited according to:

$$E\left(\left(\hat{\boldsymbol{\Theta}}_N - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\Theta}}_N - \boldsymbol{\theta}\right)^{\mathrm{T}}\right) \geq \mathbf{J}^{-1} \tag{28}$$

where $\mathbf{J}$ is the Fisher information matrix, which is defined as

$$\begin{aligned} \mathbf{J} &\triangleq -E\left(\nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}^{\mathrm{T}} \ln \Lambda_N(\boldsymbol{\theta})\right) & (29)\\ &= E\left(\left(\nabla_{\boldsymbol{\theta}} \ln \Lambda_N(\boldsymbol{\theta})\right)\left(\nabla_{\boldsymbol{\theta}} \ln \Lambda_N(\boldsymbol{\theta})\right)^{\mathrm{T}}\right) & (30) \end{aligned}$$

where $\Lambda_N(\boldsymbol{\theta})$ is the likelihood function (defined in *slide* 11).

# Cramér-Rao Lower Bound

**For Random Parameters:**

In this case, the CRLB has the same form of (28)–(30), however:

- instead of the deterministic vector $\boldsymbol{\theta}$, it considers the RV $\boldsymbol{\Theta}$ and
- the likelihood function is the following conditional pdf:

$$\Lambda_N(\boldsymbol{\Theta}) = f_{\mathbf{Y}_{1:N}|\boldsymbol{\Theta}}(\mathbf{Y}_{1:N}|\boldsymbol{\Theta})$$

**Remark:**

Note that in the deterministic case, the expectations in (29)–(30) are taken along $\mathbf{Y}_{1:N}$. In the random case, these expectations are taken along $\mathbf{Y}_{1:N}$ and $\boldsymbol{\Theta}$.

# Cramér-Rao Lower Bound

## Linear Gaussian Model with Deterministic Parameters:

Consider the linear Gaussian model

$$\mathbf{y}_i = \mathbf{H}_i \boldsymbol{\theta} + \mathbf{v}_i \ \in \mathbb{R}^m, \quad i = 1, 2, ..., N \tag{31}$$

where $\boldsymbol{\theta}$ is an unknown deterministic vector, $\mathbf{v}_{1:N}$ is a realization of an uncorrelated random sequence $\mathbf{V}_{1:N}$, with marginal distribution $\mathbf{V}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$, and $\mathbf{H}_i \in \mathbb{R}^{m \times p}$ is a known matrix.

In this case, the Fisher information matrix is given by

$$\mathbf{J} = \sum_{i=1}^{N} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \mathbf{H}_i \tag{32}$$

Note that $\mathbf{J}$ is equal to the inverse of the covariance of the ML estimator. Because the covariance of the ML estimator reaches its lower bound, we say that it is efficient.

References. . .

# Reference

Bar-Shalom, Y.; Li, X.R.; Kirubarajan, T. **Estimation with Applications to Tracking and Navigation**. New York: John Wiley & Sons, 2001.