- **Lecture 1:**
  - Course organization
  - Motivations
  - Introductive Examples
  - Review of Probability
- **Lecture 2:**
  - Estimation theory
  - Exercises
- **Lecture 3:**
  - Linear estimators
  - Exercise
- **Lecture 4:**
  - Linear estimator: Kalman Filter
  - Exercises & Matlab class
- **Lecture 5:**
  - Nonlinear estimation: EKF and UKF
  - Exercises & Matlab class
- **Lecture 6:**
  - Nonlinear estimation: PF and RBPF
  - Exercises & Matlab Class

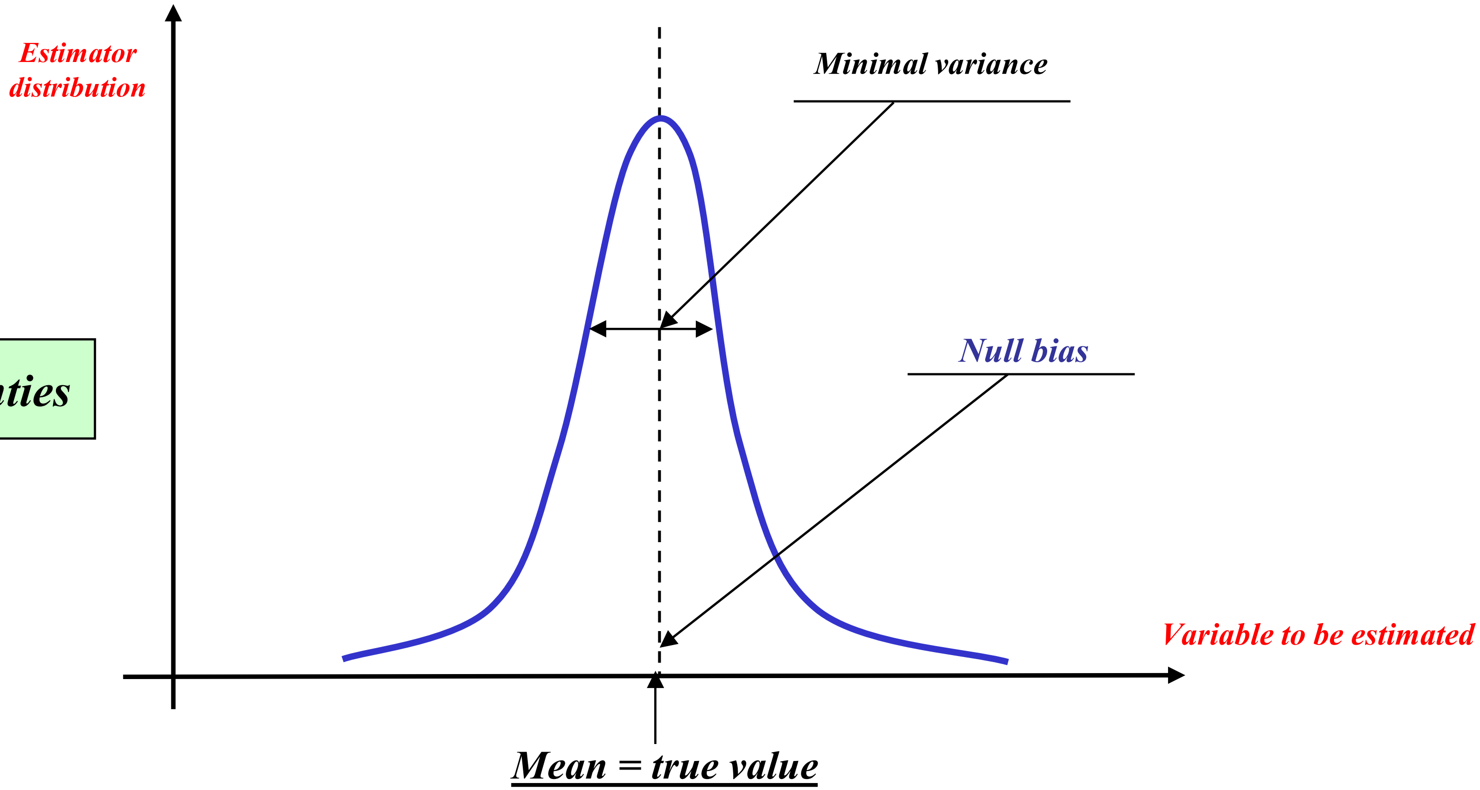
- **Definitions:**
  - Delayed estimation (offline): It has all the necessary information, and the treatment takes place in one operation
  - Real time estimation (online): We improves at every time sample the last estimator obtained according to new observations
  - Linear estimation: linear dependence between observations and parameters, e.g. the least squares estimator
  - Nonlinear estimation: one can't generally estimate the parameters by a simple combination of observations.
  - Point estimation: use of sample data to calculate a single value. it is the application of a point estimator to the data.
  - Confidence interval estimation: provides results for an interval associated with a risk or confidence level.

- **Estimation techniques:** We present in this section several estimators:
  - $\underline{\theta}$ is a **deterministic** variable/vector:
    - Minimum Variance Unbiased Estimator (MVU)
    - Linear Minimum Variance Unbiased Estimator (LMVU)
    - Maximum Likelihood Estimator (ML)
    - Least Squares Estimator (LS)
  - $\underline{\theta}$ is a **random** variable/vector: Bayesian philosophy
    - Bayesian Mean Square Estimator (BMS)
    - Linear Bayesian Mean Square Estimator (LBMS)
    - Maximum A Posteriori estimator (MAP)

✷ Minimum Variance Unbiased Estimator (MVU):

- Consider the problem of estimating a parameter $\underline{\theta}$: a possible criterion is to search among the unbiased estimators whose variance is minimal.

Variance = Uncertainties



Intuitive criterion: *We seek to minimize the fluctuations (variance) of the estimator around its true value.*

Minimum Variance Unbiased Estimator (MVU):

- Approach:. *Constrain the bias to zero and choose an estimator that minimizes the variance:*

$$\begin{cases} \textit{Unbiased} \Rightarrow E(\hat{\theta}) = \theta \quad \textit{for all } \theta \\ \textbf{Minimum Variance} \Rightarrow Var(\hat{\theta}) \text{ is minimal} \quad \textit{for all } \theta \end{cases} \Rightarrow \boxed{\underset{\hat{\theta}}{\operatorname{argmin}}\, J = E\left((\hat{\theta} - \theta)^2\right)}$$

This estimator is generally difficult to find

- Theorem Rao-Blackwell-Lehmann-Sheffe

*If $\underline{\hat{\theta}}_1(\mathbf{z})$ **is** an unbiased estimator of $\underline{\theta}$ and if $T(z)$ **is a sufficient statistic** then the estimator*

$$\underline{\hat{\theta}}(\mathbf{z}) = E\left(\underline{\hat{\theta}}_1(\mathbf{z}) / T(\mathbf{z})\right)$$

*is an unbiased estimator which does not depend on $\underline{\theta}$ and with a variance that is less or equal to the variance of $\underline{\hat{\theta}}_1(\mathbf{z})$. If now the statistic is complete (only one unbiased statistic) then $\underline{\hat{\theta}}(\mathbf{z})$ is a Minimum Variance Unbiased **(MVU)** estimator.*

Mechbal N.

✷ Linear Minimum Variance Unbiased Estimator (LMVU):

- Linear Gaussian model: consider the *special case* of **linear model** with **Gaussian** noises:

$$\mathbf{z} = \Phi\boldsymbol{\theta} + \omega \qquad \textbf{\textit{and}} \qquad \omega \sim N(0, \boldsymbol{\Gamma})$$

- For this special case, we will show in the sequel, that the MVU exists and using the Cramer-Rao lower bound, its expression is given by :

$$\boxed{\hat{\theta} = (\Phi^T\Gamma^{-1}\Phi)^{-1}\Phi^T\Gamma^{-1}\mathbf{z}}$$

Linear → ***LMVU***

- Properties (Exercise): The estimator is efficient (its variance is equal to CRLB) and has a Gaussian distribution:

$$\begin{cases} E(\hat{\theta}) = \theta \\ Var(\hat{\theta}) = (\Phi^T\Gamma^{-1}\Phi)^{-1} \end{cases}$$

Mechbal N.

Linear Minimum Variance Unbiased Estimator (LMVU):

- In practice, the MVU does not always exist or cannot be found (e.g., *we do not always know the PDF*). We then seek a suboptimal estimator which its variance will be sufficient to solve the given problem. A usual approach is to consider that the estimator is *linear with respect to the measures*.
- Then the LMVU has the following form:

$$\hat{\theta}(\mathbf{z}) = \sum_{k=0}^{N-1} a_k \mathbf{z}(k) = a^T \mathbf{z}$$

- and its expression is (*see the proof on the next slide*)

$$\boxed{\hat{\theta} = (\Phi^T \Gamma^{-1} \Phi)^{-1} \Phi^T \Gamma^{-1} \mathbf{z}}$$

- In the case of estimating the parameters of a linear model, the LMVU equals the MVU if the noise is Gaussian.
  - Hence, to compute the LMVU, we do not need the complete PDF, we *only need to know the mean* (up to scaling) and the variance Γ of **z**.

Linear Minimum Variance Unbiased Estimator (LMVU):

- **Proof:** We assume a linear model with null mean noise of unknown distribution but known variance, Γ.

$$E(\mathbf{z}) = \Phi\theta$$

- Unbiased estimation:

$$E(\hat{\theta}) = \theta \Rightarrow \begin{cases} E(\hat{\theta}) = E(a^T\mathbf{z}) = a^T E(\mathbf{z}) = a^T\Phi\theta = \theta \\ \qquad \Rightarrow a^T\Phi = 1 \end{cases}$$

- Minimum variance

$$\min\left(\operatorname{var}(\hat{\theta})\right) \Rightarrow \begin{cases} \operatorname{var}(\hat{\theta}) = E\left(\left(\hat{\theta} - E(\hat{\theta})\right)^2\right) = E\left(\left(a^T(\mathbf{z} - E(\mathbf{z}))\right)^2\right) \\ \qquad = a^T E\left((\mathbf{z} - E(\mathbf{z}))(\mathbf{z} - E(\mathbf{z}))^T\right)a = a^T\Gamma a \end{cases}$$

- **Problem**: we have to solve

$$\boxed{\min_a(a^T\Gamma a) \ \textbf{subject to} \ a^T\Phi = 1}$$

Mechbal N.

Linear Minimum Variance Unbiased Estimator (LMVU):

- We have to solve

$$\min_{a} \left(a^T \Gamma a\right) \quad \text{subject to} \quad a^T \Phi = 1$$

- It is a classical *optimization problem with constraint*. Using the method of the *Lagrange multipliers*, we should optimize the function

$$J = a^T \Gamma a + \lambda \left(a^T \Phi - 1\right)$$

- Calculating the gradient with respect to ***a*** and setting it to zero we get:

$$\frac{\partial J}{\partial a} = 0 \Rightarrow 2\Gamma a + \lambda \Phi = 0 \Rightarrow a = -\frac{1}{2} \lambda \Gamma^{-1} \Phi$$

- Recalling the constraint, we obtained the Lagrange multiplier λ as

$$a^T \Phi = 1 \Rightarrow -\frac{1}{2} \Phi^T \Gamma^{-1} \Phi \lambda = 1 \Rightarrow \lambda = -2\left(\Phi^T \Gamma^{-1} \Phi\right)^{-1}$$

- The optimal solution and estimator are then given by

$$a_{opt} = \Gamma^{-1} \Phi \left(\Phi^T \Gamma^{-1} \Phi\right)^{-1} \Longrightarrow \boxed{\hat{\theta} = \left(\Phi^T \Gamma^{-1} \Phi\right)^{-1} \Phi^T \Gamma^{-1} \mathbf{z}}$$

Maximum Likelihood Estimator (ML):

- Consider a vector $\mathbf{z}$ of known measurement. Then its density *probability function depends on* $\underline{\theta}$, *i.e.*, $p(\mathbf{z}/\underline{\theta})$ is parameterized on $\underline{\theta}$ and it is called **the Likelihood** function *since it tells us how likely it is to observe a certain* $\mathbf{z}$.
- As $\underline{\theta}$ is deterministic, we can seek for the value of $\underline{\theta}$ that maximize this *pdf*. the result is the ML estimator.
- **Definition**: *Let* $\mathbf{z}$ *be set of measures data, the maximum likelihood estimator (Likelihood) is defined as the value of* $\underline{\theta}$ *that maximizes the probability density, i.e.,*

$$\boxed{\hat{\underline{\theta}}_{mv} = \arg\max_{\hat{\underline{\theta}}} p_{\underline{z}/\underline{\theta}}(\mathbf{z}/\underline{\theta})}$$

- **Properties:**
  - The MLE is generally easy to derive.
  - Asymptotically, the ML is efficient and has the same mean and variance as the MVU
  - If an efficient estimator exists, then the ML procedure will produce it.
- **Procedure:** For the estimation we need:
  1. First, the definition of a function of $\underline{\theta}$, called the likelihood function describing the probability of obtaining the observed values of $\mathbf{z}$.
  2. Then the maximization of this function for $\underline{\theta}$.

Maximum Likelihood Estimator (ML):

- **Calculation**: To calculate the maximum, we use, for example, a *gradient method* or, in the case of scalar parameter, a simple derivative, *i.e.*:

$$\hat{\underline{\theta}}_{mv} \mapsto \frac{\partial p_{\underline{z},\underline{\theta}}(z/\underline{\theta})}{\partial \underline{\theta}}$$

- It is usually more convenient to operate on the *logarithmic* transformation of pdf, expressed then as a sum of functions of θ rather than a product of functions:

$$\hat{\underline{\theta}}_{mv} = \arg\max_{\hat{\underline{\theta}}} \log\left(p_{\underline{z},\underline{\theta}}(z/\theta)\right) = \arg\max_{\hat{\underline{\theta}}} L(z/\theta)$$

- $L(\boldsymbol{z}/\theta)$ represents the *Log* function or the log-likelihood of the pdf that we shall find it maximum with respect to θ:

$$\hat{\underline{\theta}}_{mv} \mapsto \frac{\partial L(z/\theta)}{\partial \underline{\theta}} = 0$$

Mechbal N.

Maximum Likelihood Estimator (ML)

- Linear Gaussian model: consider the following linear Gaussian model:

$$\underline{\mathbf{z}} = \Phi\underline{\theta} + \underline{\omega}, \qquad \omega \sim N(0,\Gamma) \qquad \dim(\underline{\mathbf{z}}) = N$$

- The pdf of the Gaussian noise vector is given by

$$p_{\omega}(\underline{\omega}) = (2\pi)^{\frac{-N}{2}} \det(\Gamma)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\mathbf{x}^T\Gamma^{-1}\mathbf{x}\right]$$

- But as the pdf of $\underline{\mathbf{z}}$ is parameterized by θ then the likelihood function is given by:

$$p_z(\underline{\mathbf{z}}/\underline{\theta}) = p_\omega(\mathbf{z} - \Phi\underline{\theta}) = (2\pi)^{\frac{-N}{2}} \det(\Gamma)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\underline{\mathbf{z}} - \Phi\underline{\theta})^T\Gamma^{-1}(\underline{\mathbf{z}} - \Phi\underline{\theta})\right]$$

- We can then show (**Exercise**) that the ML estimator is given by

$$\underline{\hat{\theta}}_{mv} = (\Phi^T\Gamma^{-1}\Phi)^{-1}\Phi^T\Gamma^{-1}\underline{\mathbf{z}}$$

- For the *linear Gaussian model, the ML is equivalent to the MVU*

Least Squares Estimator:

- Up to this point, the determination of the estimators involved **explicit probabilistic** knowledge of the measure (pdf, variance, ..). But it **is not always possible** to dispose of this knowledge that requires a statistical study of disturbances acting on the system.
- In this context, the approach by the method of least squares allows to give a solution to the problem of estimation.
- The main advantage of this method is its simplicity of implementation. Its main drawback is that, since there is no knowledge of the probabilistic nature of the disturbances, we can not analytically determine the performance of such an estimator. Therefore, **no conclusion in terms of statistical optimality can be given.**
- **No probabilistic assumptions required**
- **The performance highly depends on the noise**

Least Squares Estimator:

- The method of least squares (TM) was discovered in 1795 by KF **Gauss** who studied the motion of planets and comets. This method of determining the orbital parameters from noisy measurements, has been the subject of many developments.
- This method provides a solution in finding parameters that minimize a **quadratic expression** of the *gap between* the desired solution and the resulting solution with these parameters. It used to compare experimental data, usually corrupted by errors, to a mathematical model intended to describe the data.
- The mathematical model is generally a family of functions of one or more variables $\boldsymbol{x}$, indexed by one or more parameters unknown $\boldsymbol{\theta}$. The least squares method allows to select among these functions, the one that best reproduced the experimental data. In this case, we talk about **adjustment** by the method of least squares.
- If the parameters θ have a physical meaning, adjustment procedure also gives an indirect estimate of the value of these parameters: **Parametric Identification**
- **Note also, that the quadratic criterion is the energy of the error or its stochastic variance.**

Mechbal N.

Least Squares Estimator:

$\underline{z} = [z_1, z_2, \ldots, z_N]^T$

- Suppose that we have N measurements $\underline{z}$ and let $f(x, \underline{\theta})$ be the function modeling the data (obtained, for e.g., from physical law) then:
  - The method consists of a prescription (initially empirical) which is: the "best" function $f(x, \theta)$ that describes the data is the one that minimizes the quadratic sum of the deviations of measurements to the predictions of $f(x, \underline{\theta})$.
  - The optimal parameters $\underline{\theta}$ in the sense of "least squares" are those that minimize the quantity:

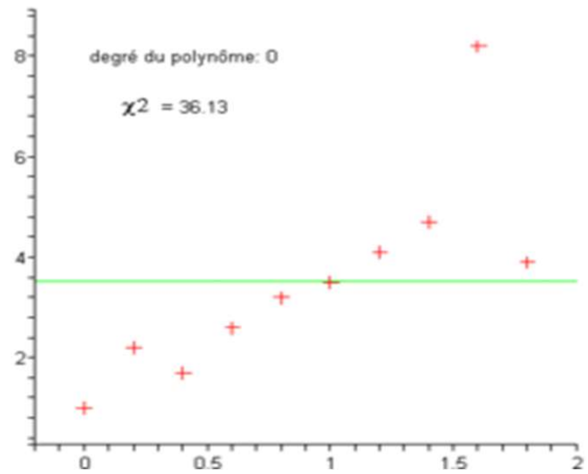
$$\boxed{J = \sum_{i=1}^{N} \left(z_i - f(x_i, \underline{\theta})\right)^2} \Rightarrow \hat{\theta} = \min_{\theta} J(\underline{\theta}) \Leftrightarrow \frac{\partial J(\underline{\theta})}{\partial \underline{\theta}} = 0$$

  - We can also weight this criterion, by introducing **(if available)** an estimate of the standard deviation σ_i of each measurement z_i (the error that affects each z_i), it is used to "weigh" the contribution of measurement to criterion:

A measure will have greater weight as the error will be small

$$J = \chi^2 = \sum_{i=1}^{N} \left(\frac{z_i - f(x_i, \underline{\theta})}{\sigma_i}\right)^2$$

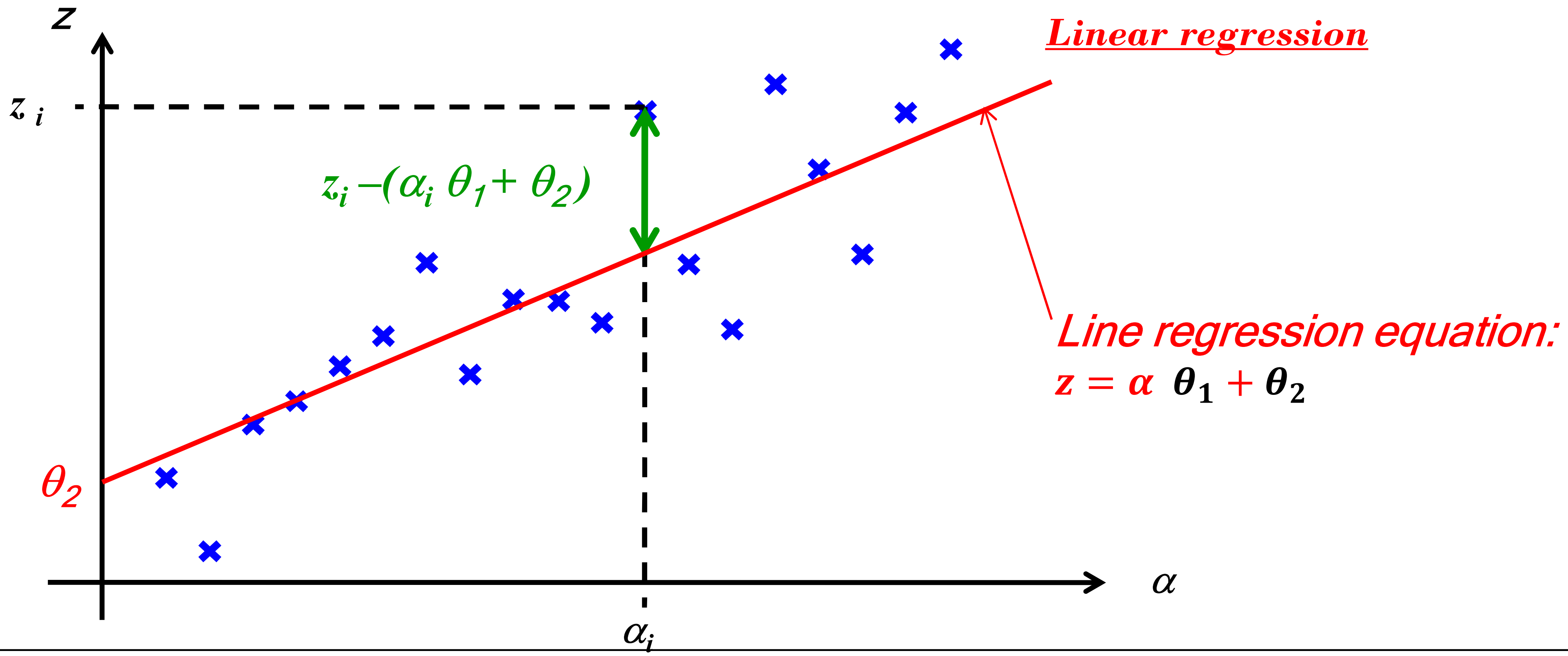
Gaussian case: Khi 2 law



Least Squares Estimator:

- <u>Linear regression:</u> $Case\ of\ \underline{\theta} = [\theta_1 \ \ \theta_2]^T \in \mathbb{R}^2$ and $z_i = \alpha_i \theta_1 + \theta_2 \in \mathbb{R}$

Criterion: $$J(\underline{\theta}) = \sum_{i=1}^{N} [z_i - (\alpha_i \theta_1 + \theta_2)]^2$$

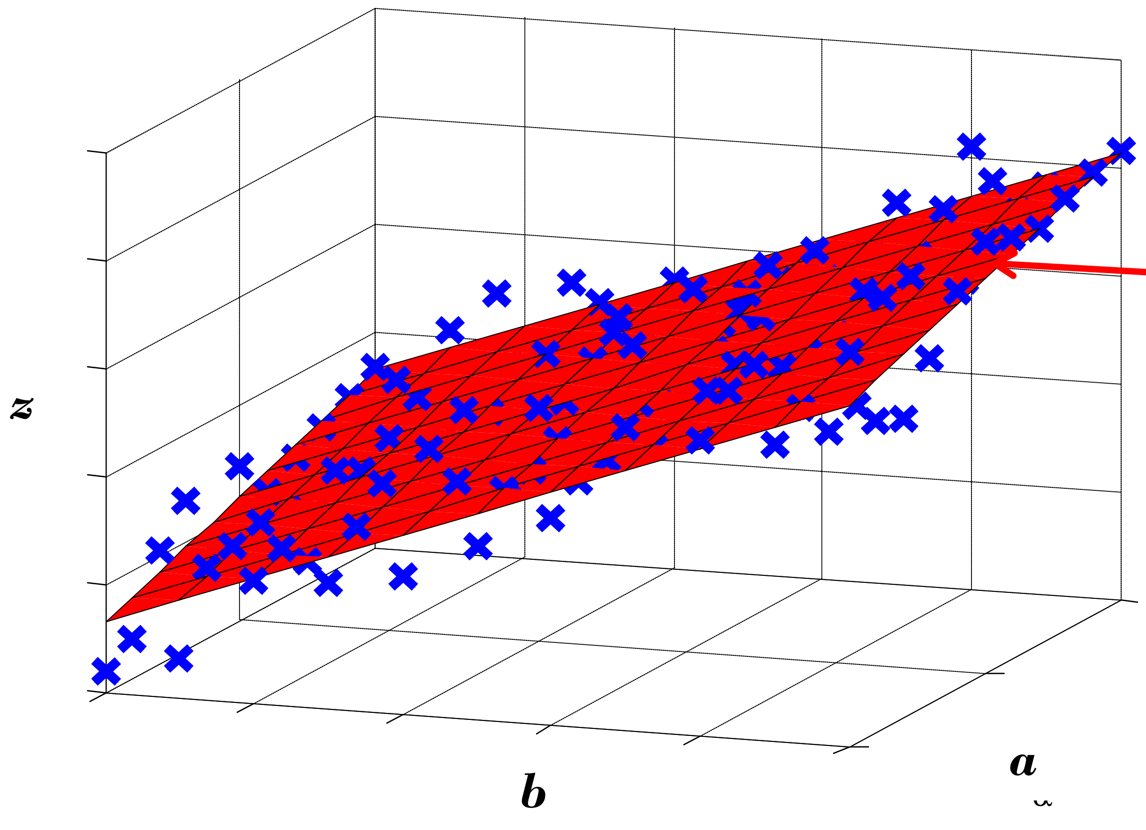


Least Squares Estimator:

- **Linear regression:** *Case of* $\underline{\theta} = [\theta_1 \ \ \theta_2 \ \ \theta_3]^T \in \mathbb{R}^3$ *and* $z_i = a_i\,\theta_1 + b_i\,\theta_2 + \theta_3$

Criterion:

$$J(\underline{\theta}) = \sum_{i=1}^{N} [z_i - (\alpha_i\theta_1 + \beta_i\theta_2 + \theta_3)]^2$$



Linear regression

Linear Regression surface:
$$z = \theta_1\, a + \theta_2\, b + \theta_3$$

✷ Least Squares Estimator:

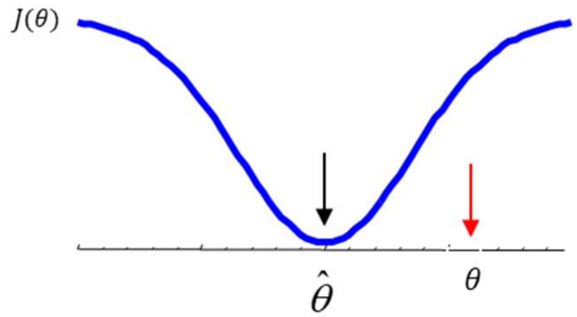
✦ Direct Solution

Derivative: Let A a matrix and $\underline{b}$ *et* $\underline{\theta}$ vectors, then we have

$$\frac{\partial(\underline{b}^T\underline{\theta})}{\partial\underline{\theta}} = \underline{b} \quad et \quad \frac{\partial(\underline{\theta}^T A\,\underline{\theta})}{\partial\underline{\theta}} = 2A\,\underline{\theta}$$

$$\hat{\underline{\theta}}_{|N} = \arg\min_{x} J(\underline{\theta}) = \arg\min_{x} \left[\left(\underline{z}^N - \Phi\underline{\theta}\right)^T \left(\underline{z}^N - \Phi\underline{\theta}\right) \right]$$

$$= \arg\min_{x} \left[\underline{\theta}^T \Phi^T \Phi \underline{\theta} - 2(\Phi^T \underline{z}^N)^T \underline{\theta} + (\underline{z}^N)^T \underline{z}^N \right]$$



Calculate $\hat{\underline{\theta}}_{|N}$ such as: $\left.\frac{\partial J(\underline{\theta})}{\partial\underline{\theta}}\right|_{\underline{\theta}=\hat{\underline{\theta}}_{|N}} = 2(\Phi^T\Phi)\hat{\underline{\theta}}_{|N} - 2(\Phi^T\underline{z}^N) = 0$

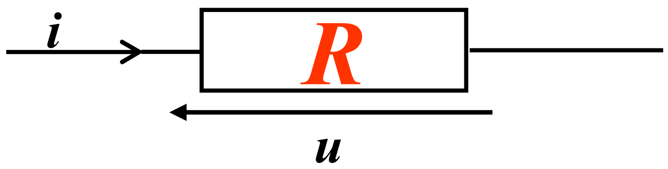
We deduce:

$$\boxed{\hat{\underline{\theta}}_{|N} = (\Phi^T\Phi)^{-1}\,\Phi^T\underline{z}^N = \mathbf{\Phi}^{\dagger}\underline{z}^N}$$

where $\Phi^{\dagger} = (\Phi^T\Phi)^{-1}\Phi^T$ **is the pseudo-inverse of matrix** Φ

$\hat{\underline{\theta}}_{|N}$ is a minimum of $J(\underline{\theta})$ as we also verify that: $\left.\frac{\partial^2 J(\underline{\theta})}{\partial\underline{\theta}^2}\right|_{x=\hat{\underline{\theta}}_{|N}} = 2(\Phi^T\Phi) > 0$

Example : Estimation of the value R of a resistance



- ***Formulation of LS problem:***

$$i(k) = i_o + b_i(k) \qquad u(k) = u_o + b_u(k)$$

- As the measurements are corrupted by noise, we define the following error

$$u(k) - Ri(k) = e(k)$$

The LS criterion to be minimized is then:

$$J_R = \sum_{k=1}^{N} \left(e(k)\right)^2 = \sum_{k=1}^{N} \left(u(k) - Ri(k)\right)^2$$

- ***The direct LS solution leads to***

$$\boxed{\begin{aligned}\hat{\theta} = \hat{R} &= \left(\Phi^T \Phi\right)^{-1} \Phi^T y(k) \\ &= \left(\sum_{k=1}^{N} \left(i^2(k)\right)\right)^{-1} \sum_{k=1}^{N} \left(i(k)u(k)\right)\end{aligned}} \Rightarrow \boxed{\hat{\theta} = \hat{R}_1 = \frac{\sum_{k=1}^{N} u(k)i(k)}{\sum_{k=1}^{N} i^2(k)}}$$

It is the estimator of Example 3 of Chapter 1

Least Squares Estimator:

Remarks:

$$\underline{\hat{\theta}}_{|N} = (\Phi^T \Phi)^{-1} \Phi^T \underline{z}^N$$

- For the linear model and considering unit covariance matrix then the LS estimator corresponds to the LMVU when the noise is white, and to the MVU when the noise is Gaussian and white

- To use this direct analytical solution, it is necessary that the matrix $(\Phi^T\Phi)$ has full rank in order to be invertible.

Problem: when **N is large**, Φ is of large dimension and invert $(\Phi^T\Phi)$ can be longer. In this case, we prefer the:

Recursive approach

$$z_i(k) = a_1 x_{1i}(k) + a_2 x_{2i}(k) + \ldots + a_N x_{Ni}(k) + \omega_i(k) = \Phi\underline{\theta} + \omega_i(k)$$

$$\Phi(k) = \underbrace{\begin{bmatrix} x_1 & \cdots & x_{1N} \\ \vdots & \vdots & \vdots \\ x_k & \cdots & x_{kN} \end{bmatrix}}_{Data} \quad \underline{\theta} = \underbrace{\begin{bmatrix} a_1 & \cdots & a_k \\ \vdots & \vdots & \vdots \\ a_{1N} & \cdots & a_{kN} \end{bmatrix}}_{Parameters}^T \quad and \quad \underline{z} = \underbrace{\begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_{N-1} \end{bmatrix}}_{Mesurement}$$

Mechbal N.

Least Squares Estimator:

- **Recursive algorithm**: This is a different organization of computations in order to avoid inverting matrix of large dimension. In the end, the value of the estimate determined by this approach is the same as in the direct calculation.
- Solution:
  - Considerer the value of $\underline{\hat{\theta}}(k-1)$ and $\underline{\hat{\theta}}(k)$ obtained successively from $k-1$ and k observations. Then from the optimal solution

$$\underline{\hat{\theta}} = (\Phi^T \Phi)^{-1} \Phi^T \underline{z}$$

  - one can rewrite this expression, as:

with $\underline{\hat{\theta}}(k) = P(k)Q(k)$ *and* $\underline{\hat{\theta}}(k-1) = P(k-1)Q(k-1)$

$$P(k) = (\Phi^T\Phi)^{-1} = \left[\sum_{j=1}^{k} \underline{\boldsymbol{\varphi}}(j)\underline{\boldsymbol{\varphi}}_i^T(j)\right]^{-1} \quad and \quad Q(k) = \Phi^T\underline{z}(k) = \sum_{j=1}^{k} \underline{\boldsymbol{\varphi}}_i(j)\underline{z}(k) \quad i = 1, \dots, N$$

$\underline{\boldsymbol{\varphi}}_i^T$ is a row matrix of Φ

Least Squares Estimator:

- **Recursive algorithm**:

$$P(k) = \left[\sum_{j=1}^{k} \underline{\varphi}_i(j)\underline{\varphi}_i^T(j)\right]^{-1} \Longrightarrow P(k-1)^{-1} = \left[\sum_{j=1}^{k-1} \underline{\varphi}_i(j)\underline{\varphi}_i^T(j)\right] \;\; \textit{and} \;\; Q(k-1) = \sum_{j=1}^{k-1} \underline{\varphi}_i(j)\underline{z}(k)$$

- We rewrite this two matrices by separating the two sample times $(k-1)$ and k:

$$P(k) = \left[P^{-1}(k-1) + \underline{\varphi}(k)\underline{\varphi}^T(k)\right]^{-1} \;\; \textit{and} \;\; Q(k) = Q(k-1) + \underline{\varphi}(k)\underline{z}(k)$$

- By using now the matrix inversion lemma,

$$\left[A^{-1} + BC^{-1}D\right]^{-1} = A - AB[C + DAB]^{-1}DA \quad \Rightarrow \quad \begin{cases} A = P(k-1), B = \underline{\varphi}(k) \\ C = I, \quad D = \underline{\phi}^T(k) \end{cases}$$

$$P(k) = P(k-1) - P(k-1)\varphi(k)\left(1 + \varphi^T(k)P(k-1)\varphi(k)\right)^{-1}\varphi^T(k)P(k-1)$$

- Which leads to

$$P(k) = P(k-1) - K(k)\varphi^T(k)P(k-1) \;\; \text{with} \;\; \boxed{K(k) = P(k-1)\underline{\varphi}(k)\left(1 + \underline{\varphi}^T(k)P(k-1)\underline{\varphi}(k)\right)^{-1}}$$

- The recursive estimator is then given by:

$$\boxed{\underline{\hat{\theta}}(k) = \underline{\hat{\theta}}(k-1) + \boldsymbol{K(k)}\left(z(k) - \underline{\varphi}^T(k)\underline{\hat{\theta}}(k-1)\right)}$$

$K(k)$ is the gain estimator to calculate

Mechbal N.

✷ Least Squares Estimator:

Initialization with k_0 measurements:

$$P(k_0) = \left(\Phi^T\Phi\right)^{-1} = \sum_{i=1}^{k_0} \underline{\varphi}_i \underline{\varphi}_i^T \qquad \underline{\hat{\theta}}_{k_0} = (\Phi^T\Phi)^{-1}\ \Phi^T \underline{z}^{k_0}$$

$$\underline{z}^{k_0} = \underbrace{\begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_{k_0} \end{bmatrix}}_{Mesurement}$$

Direct analytical solution

$$\downarrow$$

$$\boxed{k = k_0 + 1}$$

$$\downarrow$$

$$K(k) = P(k-1)\underline{\varphi}_k \left(1 + \underline{\varphi}_k^T P(k-1)\ \underline{\varphi}_k\right)^{-1}$$

$$\underline{\hat{\theta}}_k = \underline{\hat{\theta}}_{k-1} + K(k)\left(\underline{z}_k - \underline{\varphi}_k^T\ \underline{\hat{\theta}}_{k-1}\right)$$

$$P(k) = P(k-1) - K(k)\ \underline{\varphi}_k^T\ P(k-1)$$

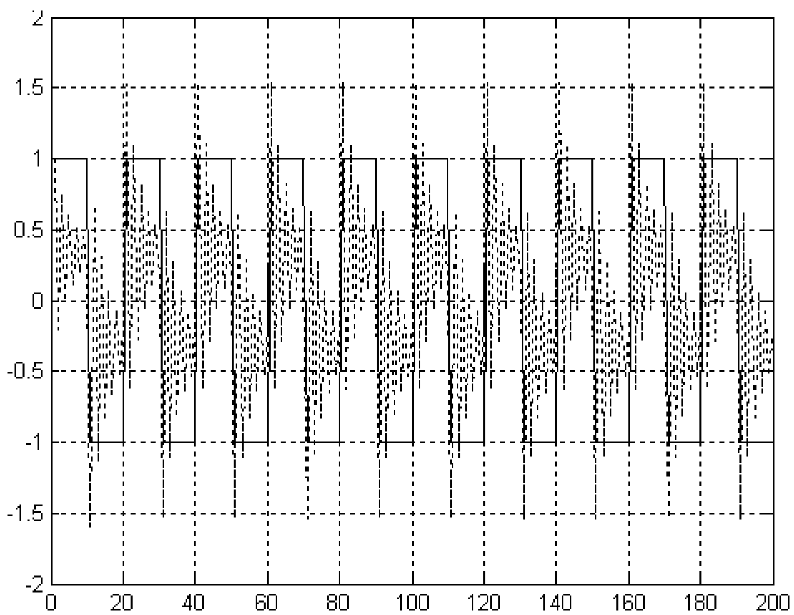
$$\downarrow$$

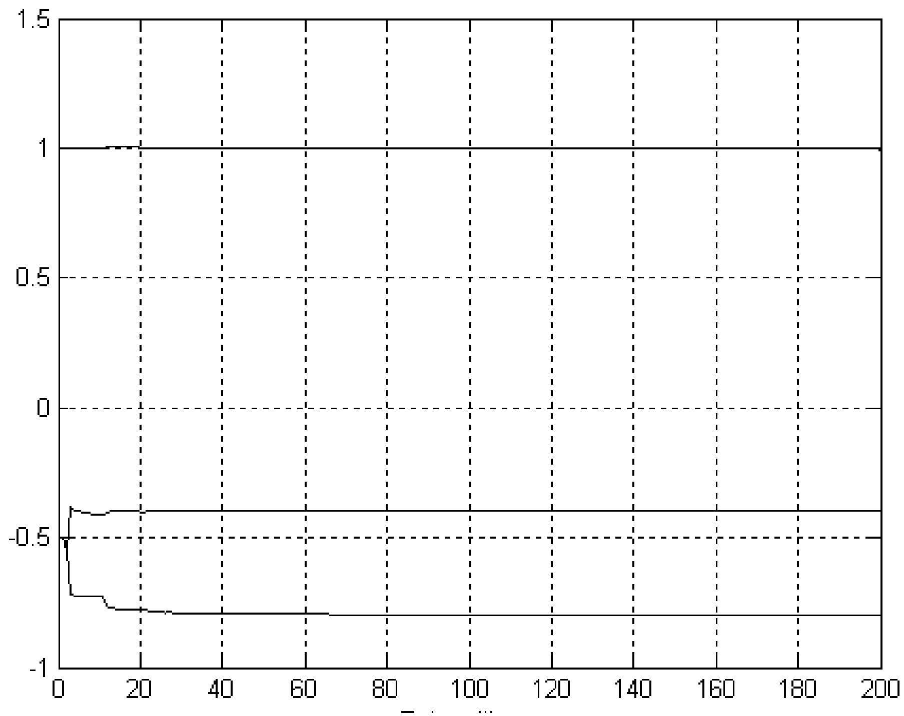
$$\boxed{k = k + 1}$$

Mechbal N.

Simple example: Recursive Least square (RLS) method

- Function : $$J = \sum_{i=1}^{N} \left(z_i - f(x_i, \theta)\right)^2$$
- The system to be identified is given by: $$G(z) = \frac{1 - 0.4z^{-1}}{1 - 0.8z^{-1}}$$
- The parameters to be estimated are : -0.4, -0.8 and 1
- Input/output and parameter estimation

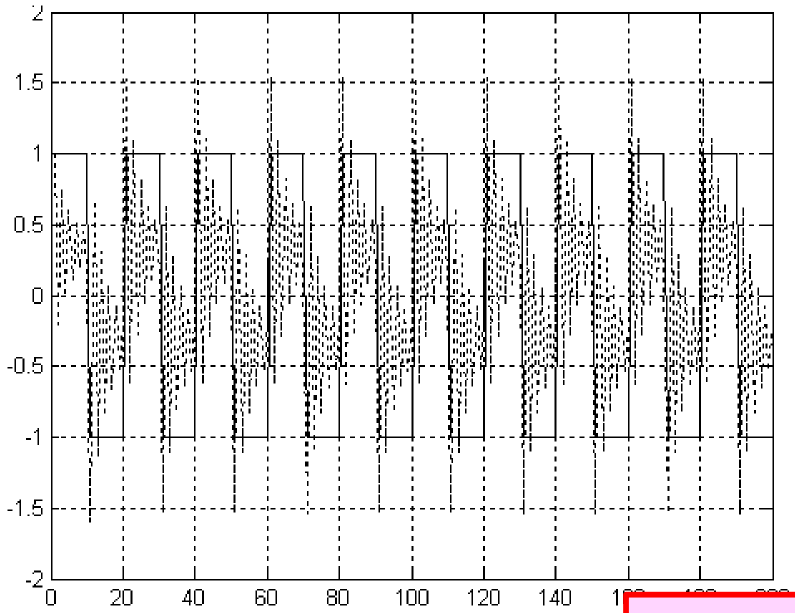


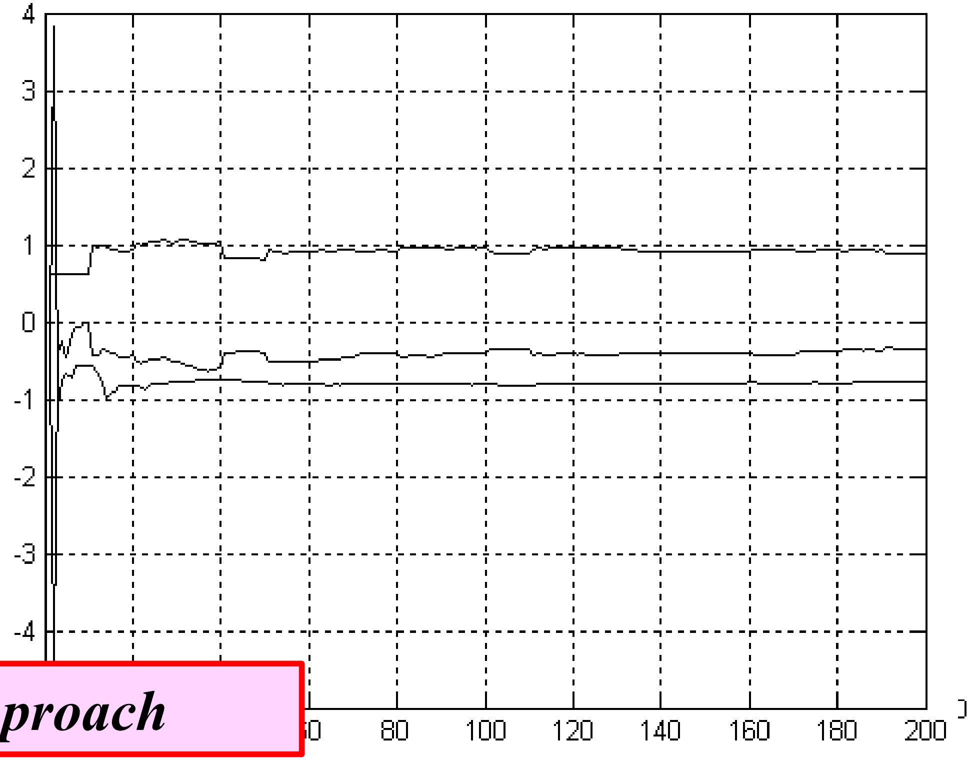


✷ Simple example: Recursive Least square (RLS) method

- Function : $$J=\sum_{i=1}^{N}\left(z_i - f(x_i,\theta)\right)^2$$
- The system to be identified is given by: $$G(z)=\frac{1-0.4z^{-1}}{1-0.8z^{-1}}$$
- The parameters to be estimated are : -0.4, -0.8 and 1

Case of noisy data





Stochastic approach

Mechbal N.

Least Squares Estimator:

- **Stochastic interpretation**: The least squares method was presented in the deterministic case without worrying on the statistical properties of variables or noises.
- Consider now that the measurement noise is **white Gaussian** with zero mean and variance R, such as:

$$E\{\underline{\omega}\} = 0 \qquad E\{\underline{\omega\omega}^T\} = \boldsymbol{R} \qquad E\{\underline{\omega x}^T\} = 0$$

- Statistically, we can always characterize the estimate by calculating the mean and variance of the estimation error:

$$\underline{\tilde{\theta}}(k) = \underline{\theta}(k) - \underline{\hat{\theta}}(k) = \underline{\theta}(k) - \left(\Phi^T\Phi\right)^{-1}\Phi^T\underline{y}(k)$$
$$= \underline{\theta}(k) - \left(\Phi^T\Phi\right)^{-1}\Phi^T(\Phi\underline{\theta}(k) - \underline{\omega}(k)) = \left(\Phi^T\Phi\right)^{-1}\Phi^T\underline{\omega}(k)$$

$$\Rightarrow \boxed{E\{\underline{\tilde{\theta}}(k)\} = \left(\Phi^T\Phi\right)^{-1}\Phi^T E\{\underline{\omega}(k)\} = 0}$$

$$\tilde{P}(k) = E\{\underline{\tilde{\theta}}(k)\underline{\tilde{\theta}}^T(k)\} = \left(\Phi^T\Phi\right)^{-1}\Phi^T E\{\underline{\omega}(k)\underline{\omega}^T(k)\}\Phi\left(\Phi^T\Phi\right)^{-1} \Rightarrow \boxed{\tilde{P}(k) = \boldsymbol{R}(\Phi^T\Phi)^{-1} = \boldsymbol{R}P(k)}$$

- Thus, the matrix that occurs at each step of calculation is, *up to a coefficient*, the **covariance matrix** of the estimation error. *It provides the accuracy of the estimate*.

Mechbal N.

The Bayesian philosophy

- In the Bayesian framework the parameter θ is viewed as a **random variable** and we must estimate its particular realization.
- We introduce an error function called "**innovation**" defined by:

$$\underline{\tilde{\theta}} = \underline{\theta} - \underline{\hat{\theta}}(z)$$

- Its is a function of two vectors of random variables. As now the error is random in nature, we can draw conclusion on the accuracy of the estimate, only by using **mean values**.
- The Bayesian approach generalizes this argument by introducing a *cost function, C,* and by finding the estimator that minimizes the following *Bayesian risk*, *R*, defined by

$$R = E\left(C(\underline{\tilde{\theta}})\right)$$

- An example of cost function is the Mean Square criterion:

$$C(\underline{\tilde{\theta}}) = \underline{\tilde{\theta}}^T \underline{\tilde{\theta}}$$

In this case we talk about Bayesian Mean Square estimator

Bayesian Mean Square Estimator (BMS)

- As now θ ***is a random*** vector, we will use *prior knowledge* about it. Hence, we will consider that θ has a *pdf* noted $p(\theta)$.
- Considering a linear model, we will also use the joint density $p(z, \theta)$
- The BMS is based on the minimization of the following risk:

$$R = E\left(\underline{\tilde{\theta}}^T \underline{\tilde{\theta}}\right) = E\left((\theta - \hat{\theta}(z))^T (\theta - \hat{\theta}(z))\right)$$

- Using the density functions this risk is written

$$R = \int_z \int_\theta \left((\theta - \hat{\theta})^T (\theta - \hat{\theta})\right) p(z, \theta) dz\, d\theta$$

- Which, according to Bayes (*conditional probability*), could also be rewritten as

$$R = \int_z p(z) \left(\int_\theta \left((\theta - \hat{\theta}(z))^T (\theta - \hat{\theta}(z))\right) p(\theta/z) d\theta \right) dz$$

- But as the density $p(z)$ is a positive function, the minimization of $\boldsymbol{R}$ is equivalent to minimize the integral

$$I(\hat{\theta}) = \int_\theta \left((\theta - \hat{\theta})^T (\theta - \hat{\theta})\right) p(\theta/z) d\theta$$

Bayesian Mean Square Estimator (BMS)

- Recall the integral

$$I(\hat{\theta}) = \int_{\theta} \left((\theta - \hat{\theta})^T (\theta - \hat{\theta})\right) p(\theta/z) d\theta$$

- To minimize this integral we calculate the derivative

$$\frac{\partial I(\hat{\theta})}{\partial \hat{\theta}} = 0 \Rightarrow \int_{\theta} \left(2(\theta - \hat{\theta}(z))\right) p(\theta/z) d\theta = 0 \qquad \text{with } \int_{\theta} p(\theta/z) d\theta = 1$$

$$\Rightarrow \int_{\theta} \theta p(\theta/z) d\theta = \hat{\theta}(z) \int_{\theta} p(\theta/z) d\theta \;\Rightarrow\; \hat{\theta}(z) = \int_{\theta} \theta p(\theta/z) d\theta = \boldsymbol{E(\theta/z)}$$

- Therefore, the **optimal BMS estimator** is equal to the conditional density function of the parameter given the measurement vector, *i.e.*,

$$\boxed{\hat{\theta}(z) = E(\theta/z)}$$

- The **drawback** of this estimator is that it needs *an* ***explicit knowledge*** of this conditional density which is difficult to obtain in practice, except *in the Gaussian case*.

- A way to get around the problem is to assume that the estimator is linear with respect to the measures. this approach leads to an estimator called: Linear Bayesian Mean Square estimator (LBMS)

Linear Bayesian Mean Square estimator (LBMS)

- As for the LMVU, we now constrain the LBM estimator to have the following form:

$$\hat{\theta}(\mathbf{z}) = \sum_{k=0}^{N-1} a_k \mathbf{z}(k) = a^T \mathbf{z} \quad with \quad \mathbf{z} = \Phi \underline{\theta} + \omega$$

- The Bayesian risk is the written (*for simplicity, we suppose here the scalar case*)

$$R = E\left((a^T z - \theta)^2\right) = a^T E(zz^T)a - 2a^T E(z\theta) + E(\theta^2)$$
$$= a^T \Gamma_{zz} a - 2a^T Cov(z,\theta) + \mathrm{var}(\theta)$$

- By setting the derivative to zero

$$\frac{\partial R}{\partial a} = 0 \quad \Rightarrow \quad a_{opt} = \Gamma_{zz}^{-1} Cov(z,\theta)$$

- The estimator is then given by

$$\boxed{\hat{\theta} = \left(Cov_{z\theta}^T \Gamma_{zz}^{-1}\right) z}$$

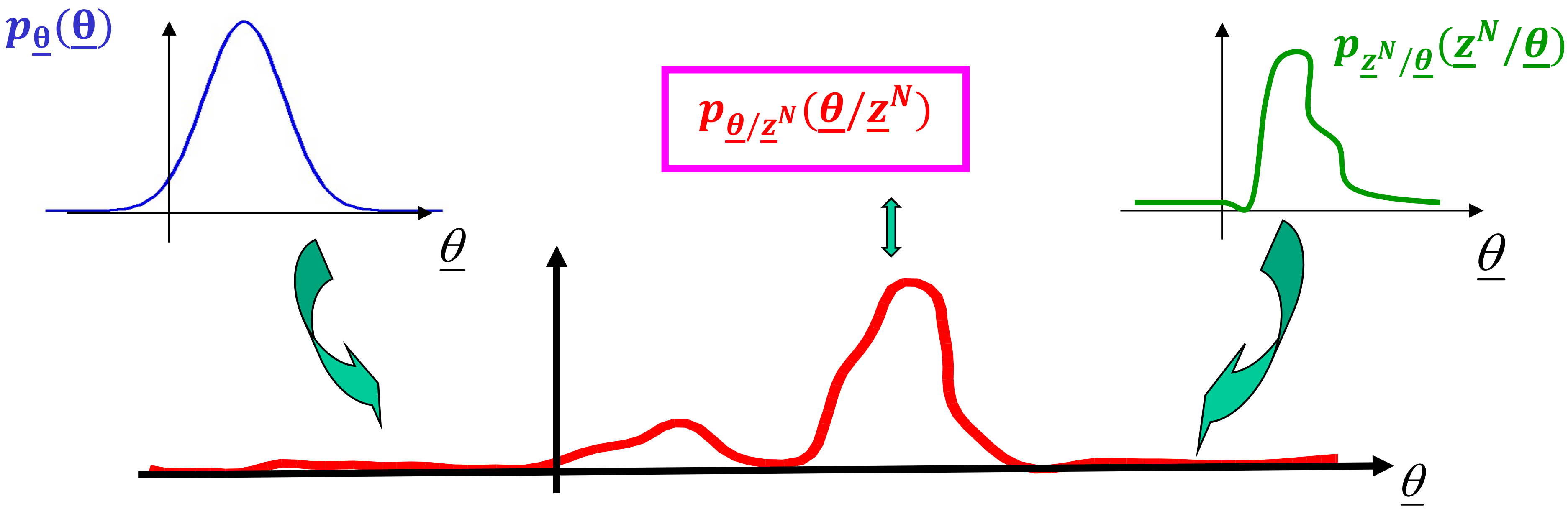
solution

$$\boxed{\hat{\theta} = \sigma^2 \Phi^T \left(\sigma^2 \Phi \Phi^T + \Gamma_\omega\right)^{-1} z}$$

- **Exercise**: *Drive the expression of the LBMS in the scalar case and when θ is assumed to have zero mean and variance* σ^2 (use matrix inversion lemma).

Maximum A Posteriori estimator (MAP)

- Elements of the problem: we assume that we have:
  - The ***a priori*** density function of $\underline{\theta}$: $p(\underline{\theta})$
  - The pdf of measure vector conditioned on θ, the ***likelihood***: $p(\boldsymbol{z}/\underline{\theta})$
- **Objective:** *Merge these two sources of information to obtain a better estimate*:



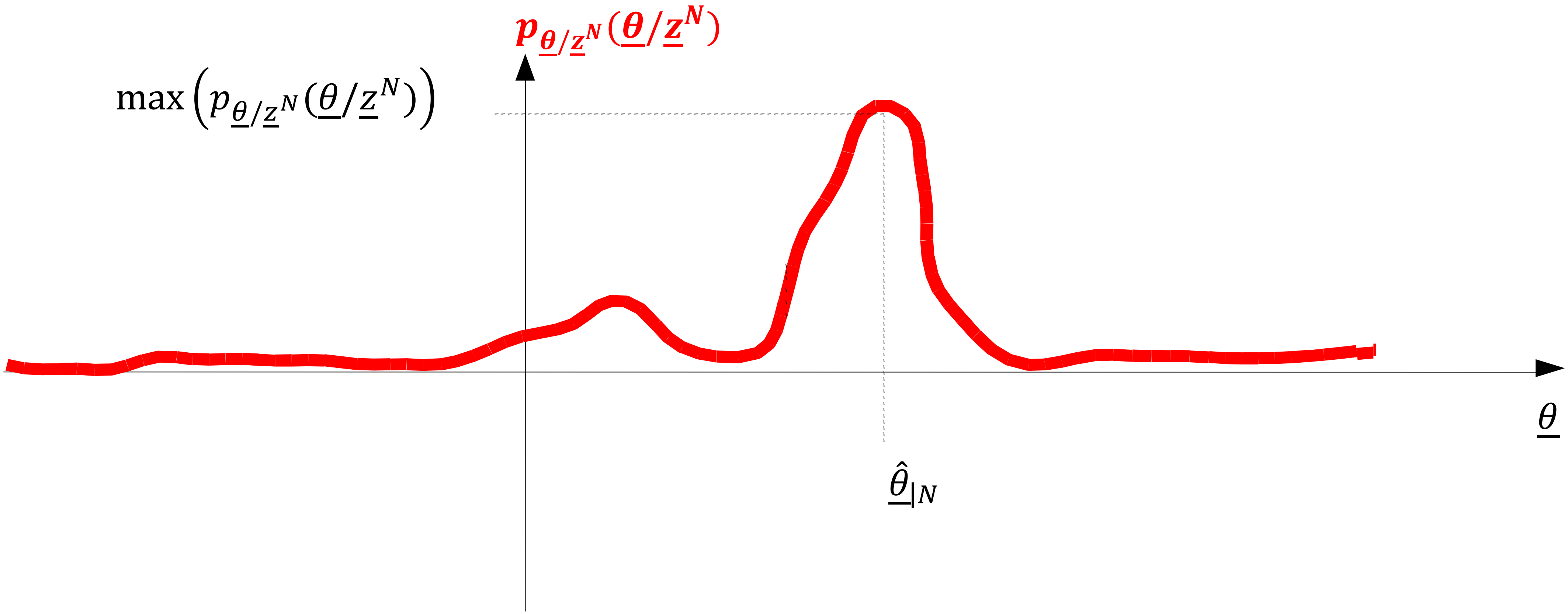
- The best information is the probability density of θ conditioned on the measurements: The ***a posteriori conditional density*** $p_{\underline{\theta}/\underline{z}^N}(\underline{\theta}/\underline{z}^N)$

Mechbal N.

Maximum A Posteriori estimator (MAP)

Definition

$$\underline{\hat{\theta}}_{|N} = \operatorname{argmax}_{\underline{\theta}}\left(p_{\underline{\theta}/\underline{z}^N}(\underline{\theta}/\underline{z}^N)\right)$$



Mechbal N.

Maximum A Posteriori estimator (MAP)

- Determination: we use available data

Bayes rules $\Rightarrow p_{\underline{\theta}/\underline{z}^N}(\underline{\theta}/\underline{z}^N) = \dfrac{p_{\underline{z}^N/\underline{\theta}}(\underline{z}^N/\underline{\theta})p_{\underline{\theta}}(\underline{\theta})}{p_{\underline{z}^N}(\underline{z}^N)}$

Marginal density $\Rightarrow p_{\underline{z}^N}(\underline{z}^N) = \displaystyle\int_{-\infty}^{+\infty} p_{\underline{z}^N,\underline{\theta}}(\underline{z}^N,\underline{\theta})\,\mathrm{d}\theta$

$$\boxed{p_{\underline{\theta}/\underline{z}^N}(\underline{\theta}/\underline{z}^N) = \frac{p_{\underline{z}^N/\underline{\theta}}(\underline{z}^N/\underline{\theta})p_{\underline{\theta}}(\underline{\theta})}{\int_{-\infty}^{+\infty} p_{\underline{z}^N/\underline{\theta}}(\underline{z}^N/\underline{\theta})p_{\underline{\theta}}(\underline{\theta})\,\mathrm{d}\theta}}$$

Mechbal N.

Maximum A Posteriori estimator (MAP)

- Determination: $\hat{\underline{\theta}}_{|N} = \arg\max_{\underline{\theta}} \left(p_{\underline{\theta}|\underline{z}^N}(\underline{\theta}|\underline{z}^N) \right)$

$$p_{\underline{\theta}/\underline{z}^N}(\underline{\theta}/\underline{z}^N) = \frac{p_{\underline{z}^N/\underline{\theta}}(\underline{z}^N/\underline{\theta})p_{\underline{\theta}}(\underline{\theta})}{\int_{-\infty}^{+\infty} p_{\underline{z}^N/\underline{\theta}}(\underline{z}^N/\underline{\theta})p_{\underline{\theta}}(\underline{\theta})\,\mathrm{d}\theta} = \frac{p_{\underline{z}^N/\underline{\theta}}(\underline{z}^N/\underline{\theta})p_{\underline{\theta}}(\underline{\theta})}{p_{\underline{z}^N}(\underline{z}^N)}$$

However,

$$p_{\underline{z}^N}(\underline{z}^N) = \int_{-\infty}^{+\infty} p_{\underline{z}^N,\underline{\theta}}(\underline{z}^N,\underline{\theta})\,\mathrm{d}\theta$$

Joint pdf

doesn't depend on θ.

Therefore:

$$\boxed{\hat{\underline{\theta}}_{|N} = \mathrm{argmax}_{\underline{\theta}} \left(p_{\underline{z}^N/\underline{\theta}}(\underline{z}^N/\underline{\theta})p_{\underline{\theta}}(\underline{\theta}) \right)}$$

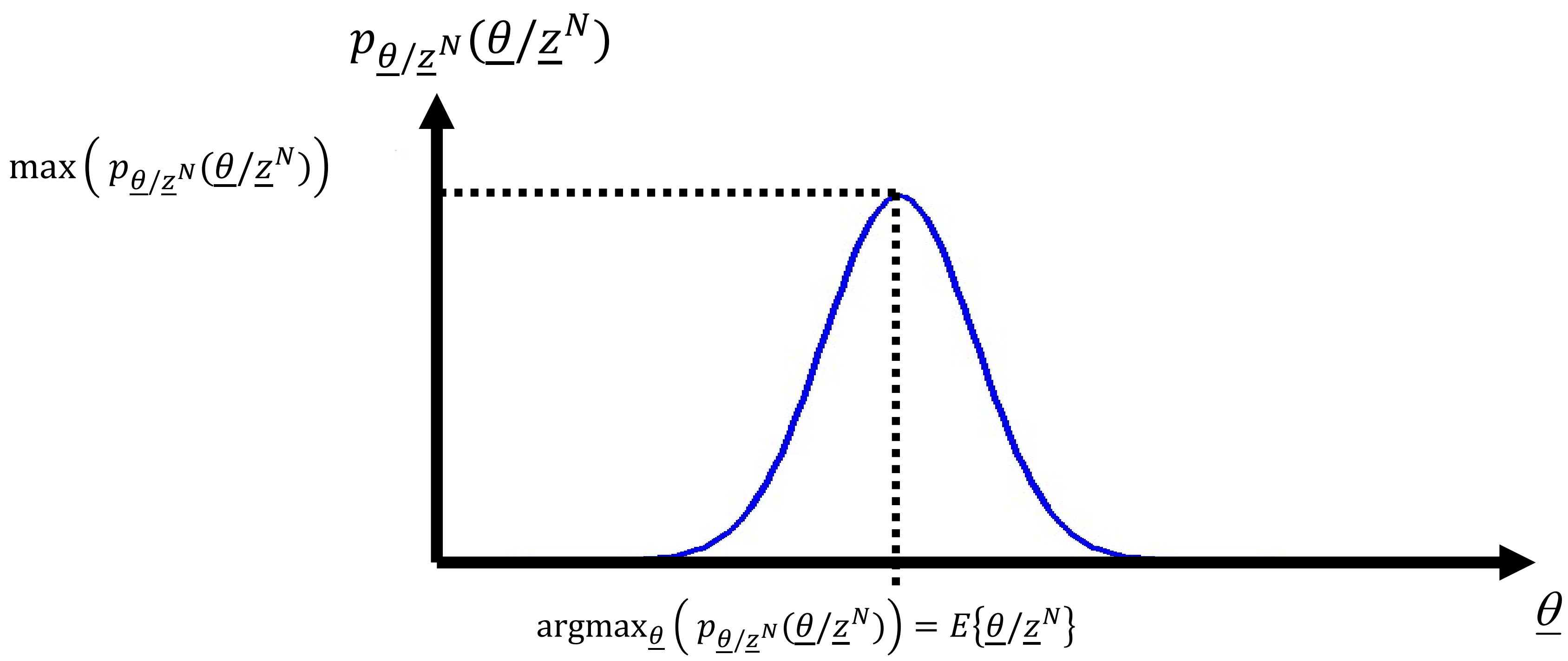
Equivalently : $\hat{\underline{\theta}}_{|N} = \mathrm{argmax}_{\underline{\theta}} \left(\ln\left[p_{\underline{z}^N/\underline{\theta}}(\underline{z}^N/\underline{\theta})p_{\underline{\theta}}(\underline{\theta}) \right] \right)$

Maximum A Posteriori estimator (MAP)

- Determination: Case of a symmetric and unimodal probability density (only a single maximum). The conditional *a posteriori* density is then given by:

$$\mathrm{argmax}_{\underline{\theta}}\left(p_{\underline{\theta}/\underline{z}^N}(\underline{\theta}/\underline{z}^N)\right) = E\{\underline{\theta}/\underline{z}^N\}$$

- For example, the **Gaussian** probability density



Maximum A Posteriori estimator (MAP)

- Determination: Unimodal density (Gaussian)

$$\hat{\underline{\theta}}_{|N} = \arg\max_{\underline{\theta}} \left(p_{\underline{\theta}|\underline{z}^N}(\underline{\theta}|\underline{z}^N) \right) = E\left\{ \underline{\theta}|\underline{z}^N \right\}$$

$$E\left\{ \underline{\theta}|\underline{z}^N \right\} = \int_{-\infty}^{+\infty} \underline{\theta} \quad p_{\underline{\theta}|\underline{z}^N}(\underline{\theta}|\underline{z}^N) \mathrm{d}\underline{\theta}$$

From definition of the expectation

$$p_{\underline{\theta}|\underline{z}^N}(\underline{\theta}|\underline{z}^N) = \frac{p_{\underline{z}^N|\underline{\theta}}(\underline{z}^N|\underline{\theta})\, p_{\underline{\theta}}(\underline{\theta})}{\int_{-\infty}^{+\infty} p_{\underline{z}^N|\underline{\theta}}(\underline{z}^N|\underline{\theta})\, p_{\underline{\theta}}(\underline{\theta})\, \mathrm{d}\theta}$$

$$\boxed{\hat{\underline{\theta}}_{|N} = E\left\{ \underline{\theta}|\underline{z}^N \right\} = \frac{\int_{-\infty}^{+\infty} \underline{\theta} \quad p_{\underline{z}^N|\underline{\theta}}(\underline{z}^N|\underline{\theta})\, p_{\underline{\theta}}(\underline{\theta}) \quad \mathrm{d}\underline{\theta}}{\int_{-\infty}^{+\infty} p_{\underline{z}^N|\underline{\theta}}(\underline{z}^N|\underline{\theta})\, p_{\underline{\theta}}(\underline{\theta})\, \mathrm{d}\theta}}$$

Mechbal N.

Maximum A Posteriori estimator (MAP)

- **Properties**: in case of **Gaussian density** the MAP has the following properties:
  - It is an unbiased estimator: $B\{\hat{\underline{\theta}}_{|N}\} = E\{\hat{\underline{\theta}}_{|N}\} - E\{\underline{\theta}\} = 0$
  - It is of minimal covariance
  - We can show that it is equivalent to the **LBMS estimator**

- In the case **of non Gaussian** world

$$\hat{\underline{\theta}}_{|N} = E\{\underline{\theta} | \underline{z}^N\} = \frac{\int_{-\infty}^{+\infty} \underline{\theta} \; p_{\underline{z}^N|\underline{\theta}}(\underline{z}^N | \underline{\theta}) \, p_{\underline{\theta}}(\underline{\theta}) \; \mathrm{d}\underline{\theta}}{\int_{-\infty}^{+\infty} p_{\underline{z}^N|\underline{\theta}}(\underline{z}^N | \underline{\theta}) \, p_{\underline{\theta}}(\underline{\theta}) \, \mathrm{d}\theta}$$

Bayesian estimation → *calculus of integral* → *Complicated If not Gaussian* → ***MCMC approches*** *Part 3 of this course*

✷ Estimated from a Linear Gaussian static model: The problem

- Problem formulation: Consider the problem of estimating linear static Gaussian models. It is a system where the relationship between phenomena involves no dynamic effect and errors. Uncertainties acting on the process are assumed to be Gaussian.

$$\underline{z} = H\underline{x} + \underline{v} \qquad \textit{with} \qquad \dim(\underline{x}) = n \quad et \quad \dim(\underline{z}) = m$$

where $\underline{x}$ is a random Gaussian vector with mean, m_x and covariance Q. $\underline{v}$ is also a random Gaussian vector having mean 0 and covariance R.

- The goal is to estimate the variable $\underline{x}$ given noisy measurements $\underline{z}$. In fact, we want to elaborate a *Linear Bayesian Mean Square Estimator* and hence calculating the following conditional pdf:

$$f_{\underline{x}/\underline{z}}(\underline{x}/\underline{z})$$

- The estimator will be then

$$\boxed{\hat{\underline{x}} = E\left\{\underline{x}/\underline{z}\right\}}$$

Mechbal N.

Estimated from a Linear Gaussian static model: Background

- Statistic characteristics of two Gaussian variables: Let x and z be two Gaussian random vectors of dimensions n and m respectively and with m_x, P_{xx} and m_z, P_{zz} as respective *mean and covariance*. Their joint distribution is written:

$$\underline{y} = \begin{pmatrix} \underline{x} \\ \underline{z} \end{pmatrix} \qquad f_{\underline{x},\underline{z}}(\underline{x},\underline{z}) = \left[(2\pi)^{\frac{n+m}{2}} \det\begin{pmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{pmatrix}\right]^{-1} \exp\left[-\frac{1}{2}\begin{pmatrix} \underline{x}-\underline{m}_x \\ \underline{z}-\underline{m}_z \end{pmatrix}^T \begin{pmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{pmatrix}^{-1} \begin{pmatrix} \underline{x}-\underline{m}_x \\ \underline{z}-\underline{m}_z \end{pmatrix}\right]$$

- Using the conditional density definition, we have

$$f_{\underline{x}/\underline{z}}(\underline{x}/\underline{z}) = f_{\underline{x},\underline{z}}(\underline{x},\underline{z})/f_{\underline{z}}(\underline{z})$$

- The calculus leads then to

$$f_{\underline{x}/\underline{z}}(\underline{x}/\underline{z}) = \frac{1}{(2\pi)^{1/2}\left(P_{\underline{x}/\underline{z}}\right)^{1/2}} \exp\left[-\frac{1}{2}\left(\underline{x}-\underline{m}_{x/z}\right)^T\left(P_{\underline{x}/\underline{z}}\right)^{-1}\left(\underline{x}-\underline{m}_{x/z}\right)\right]$$

with

$$\boxed{\begin{aligned} \underline{m}_{x/z} &= \underline{m}_x + P_{xz}(P_{zz})^{-1}(\underline{z}-\underline{m}_z) \\ P_{\underline{x}/\underline{z}} &= P_{xx} - P_{xz}(P_{zz})^{-1}P_{zx} \end{aligned}}$$

$$\underline{m}_y = \begin{pmatrix} \underline{m}_x \\ \underline{m}_z \end{pmatrix} \qquad P_{yy} = \begin{pmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{pmatrix}$$

- The conditional estimator is then given by

$$\boxed{\hat{x} = E\{\underline{x}/\underline{z}\} = \underline{m}_{x/z} = \underline{m}_x + P_{xz}(P_{zz})^{-1}(\underline{z}-\underline{m}_z)}$$

Estimated from a Linear Gaussian static model: The solution

- *LBMS estimator*: we recall that the considered model is

$$\underline{z} = H\,\underline{x} + \underline{v}$$

  - We define the following augmented vector

$$\underline{u} = \begin{bmatrix} \underline{x} \\ \underline{v} \end{bmatrix}$$

  - This vector has the following statistic characteristics (x et v are independent):

$$\underline{m}_u = \begin{bmatrix} \underline{m}_x \\ 0 \end{bmatrix} \;,\; P_{uu} = \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix}$$

  - As any *linear operation* of Gaussian random variables gives Gaussian random variables, we can write:

$$\underline{y} = \begin{pmatrix} \underline{x} \\ \underline{z} \end{pmatrix} = \begin{pmatrix} \underline{x} \\ H\underline{x} + \underline{v} \end{pmatrix} = \begin{pmatrix} I & 0 \\ H & I \end{pmatrix} \begin{pmatrix} \underline{x} \\ \underline{v} \end{pmatrix} = \begin{pmatrix} I & 0 \\ H & I \end{pmatrix} \underline{u}$$

with

$$\underline{m}_y = \begin{pmatrix} \underline{m}_x \\ \underline{m}_z \end{pmatrix} = \begin{pmatrix} I & 0 \\ H & 0 \end{pmatrix} \underline{m}_u = \begin{pmatrix} I & 0 \\ H & 0 \end{pmatrix} \begin{pmatrix} \underline{m}_x \\ 0 \end{pmatrix} = \begin{pmatrix} \underline{m}_x \\ H\underline{m}_x \end{pmatrix}$$

$$P_{yy} = \begin{pmatrix} I & 0 \\ H & I \end{pmatrix} \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix} \begin{pmatrix} I & H^T \\ 0 & I \end{pmatrix} = \begin{pmatrix} Q & QH^T \\ HQ & HQH^T + R \end{pmatrix} = \begin{pmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{pmatrix}$$

Estimated from a Gaussian static model

- In the case of *two random vectors* we have already seen that

$$E\{\underline{x}/\underline{z}\} = \underline{m}_x + P_{xz}(P_{zz})^{-1}(\underline{z} - \underline{m}_z)$$

$$\hat{P} = P_{\underline{x}/\underline{z}} = P_{xx} - P_{xz}(P_{zz})^{-1}P_{zx}$$

- Therefore the estimator is given by

$$\underline{\hat{x}} = E\{\underline{x}/\underline{z}\} = \underline{m}_x + QH^T(HQH^T + R)^{-1}(\underline{z} - H\underline{m}_x)$$

with the uncertainties

$$\hat{P} = P_{\underline{x}/\underline{z}} = Q - QH^T(HQH^T + R)^{-1}HQ$$

- if we introduce now a gain *K* we can rewrite these expressions to have

$$\boxed{\begin{aligned} K &= QH^T(HQH^T + R)^{-1} \\ \underline{\hat{x}} &= \underline{m}_x + K(\underline{z} - H\underline{m}_x) \\ \hat{P} &= Q - KHQ \end{aligned}}$$

⇨ ***It is similar to a Kalman filter structure***

- It is a recursive calculation: *the current estimate is the sum of the previous estimate plus a correction term weighted by a gain*

- **Lecture 1:**
  - Course organization
  - Motivations
  - Introductive Examples
  - Review of Probability
- **Lecture 2:**
  - Estimation theory
  - Exercises
- **Lecture 3:**
  - Linear estimator
  - Exercises
- **Lecture 4:**
  - Linear estimator: Kalman Filter
  - Exercises & Matlab class
- **Lecture 5:**
  - Nonlinear estimation: EKF and UKF
  - Exercises & Matlab class
- **Lecture 6:**
  - Nonlinear estimation: PF and RBPF
  - Exercises & Matlab class