Advanced Control Systems
Detection, Estimation, and Filtering

Graduate Course on the
MEng PhD Program
Spring 2012/2013

Chapter 5
Best Linear Unbiased Estimators

Instructor:
Prof. Paulo Jorge Oliveira
p.oliveira@dem.ist.utl.pt or pjcro @ isr.ist.utl.pt
Phone: 21 8419511 or 21 8418053 (3511 or 2053 inside IST)

DEM
DEPARTAMENTO
DE ENGENHARIA MECÂNICA
TÉCNICO LISBOA

Syllabus:

Classical Estimation Theory

…

Chap. 4 - ***Linear Models in the Presence of Stochastic Signals*** [1 week]

Stationary and transient analysis; White Gaussian noise and linear systems; Examples;

Chap. 5 - ***Best Linear Unbiased Estimators*** [1 week]

Definition of BLUE estimators; White Gaussian noise and bandlimited systems; Examples; Generalized MVU estimation;

Chap. 6 - ***Maximum Likelihood Estimation*** [1 week]

The maximum likelihood estimator; Properties of the ML estimators; Solution for ML estimation; Examples; Monte-Carlo methods;

continues…

An alternative strategy:

FACT:

It occurs that the MVU estimator, if it exists, can not be found.

e.g. the PDF for the data is not known, the user would not like to assume a model for the PDF, or the problem can be mathematically untreatable.

An alternative strategy can be pursued is to study the class of

Best Linear Unbiased Estimators

Only suboptimal performance can be achieved.

The performance degradation, relative to the MVU estimator, is unknown but the resulting performance can be enough for the problem at hand.

DEM
DEPARTAMENTO
DE ENGENHARIA MECÂNICA
TÉCNICO LISBOA

BLUE structure:

The Best Linear Unbiased Estimator consists of restrict the estimator to be a linear function of the data, i.e.

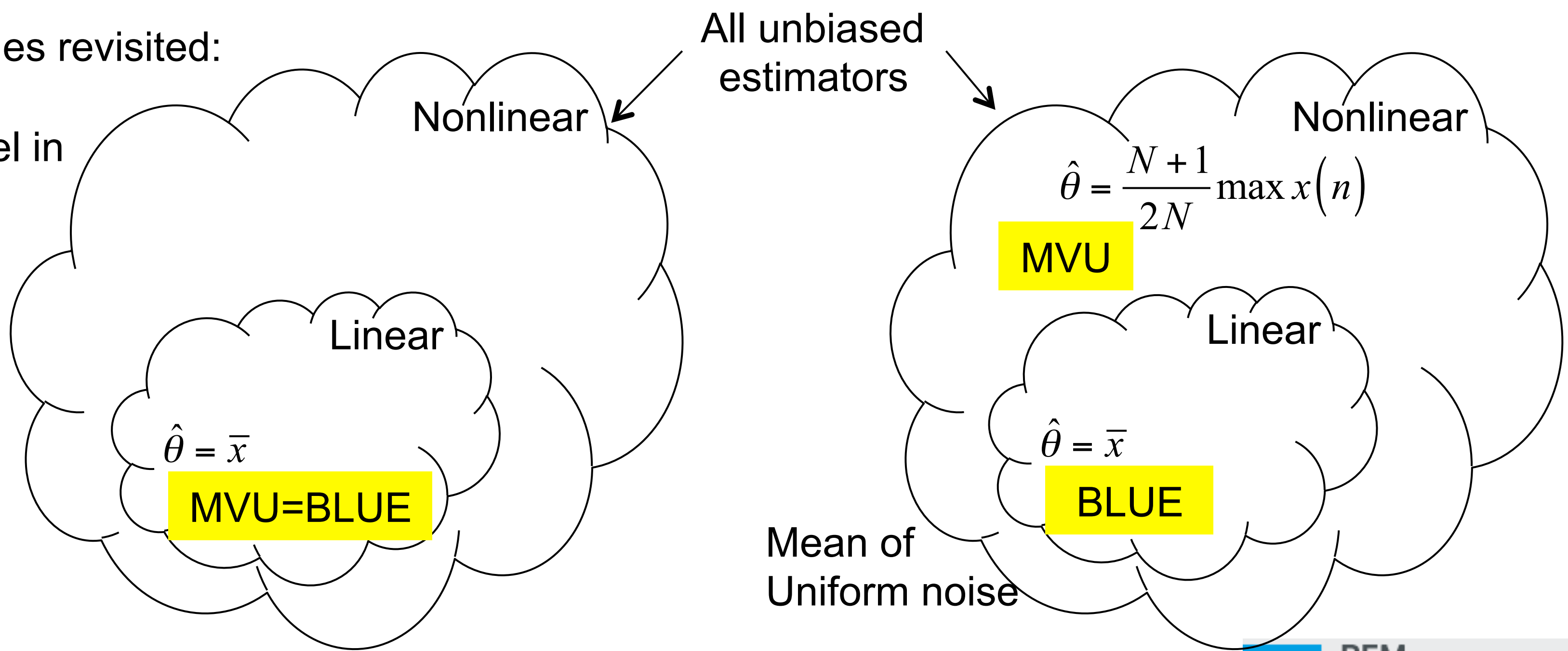
$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n]$$

where the a_n's are constants to be determined.

Optimality in general is lost.

Examples revisited:

DC level in WGN



Finding the BLUE:

To find the BLUE we constrain the estimator

- to be linear $$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n]$$
- to be unbiased $$E[\hat{\theta}] = \sum_{n=0}^{N-1} a_n E[x[n]] = \theta \qquad (6.2)$$
- to minimize its variance $$\text{var}(\hat{\theta}) = E\left[\left(\sum_{n=0}^{N-1} a_n x[n] - E\left[\sum_{n=0}^{N-1} a_n x[n]\right]\right)^2\right]$$

Defining ***a***$=[a_0\ a_1\ \ldots\ a_{N-1}]^T$ and ***x***$=[x_0\ x_1\ \ldots\ x_{N-1}]^T$ this last expression can be simplified:

$$\text{var}(\hat{\theta}) = E\left[\left(\mathbf{a}^T\mathbf{x} - \mathbf{a}^T E[\mathbf{x}]\right)^2\right] = E\left[\left(\mathbf{a}^T\left(\mathbf{x} - E[\mathbf{x}]\right)\right)^2\right] =$$

$$= E\left[\mathbf{a}^T\left(\mathbf{x} - E[\mathbf{x}]\right)\left(\mathbf{x} - E[\mathbf{x}]\right)^T\mathbf{a}\right] = \mathbf{a}^T C\mathbf{a}. \qquad (6.3)$$

The problem of finding the BLUE can be stated as, for $\mathbf{a} \in R^N$

$$\begin{array}{ll} \min & \mathbf{a}^T C\mathbf{a} \\ \text{subject to} & \mathbf{a}^T E[\mathbf{x}] = \theta \end{array}$$

DEM
DEPARTAMENTO DE ENGENHARIA MECÂNICA
TÉCNICO LISBOA

Finding the BLUE:

Given the scalar parameter θ, the expected value of the samples can be assumed as

$E[x[n]]=s[n]\theta,$

where $s[n]$ is known.

$$\sum_{n=0}^{N-1} a_n E\left[x\left[n\right]\right] = \sum_{n=0}^{N-1} a_n s\left[n\right]\theta = \theta, \qquad \text{(from 6.2)}$$

Thus the previous problem can be stated as

$$\min_{\mathbf{a} \in R^N} \quad \mathbf{a}^T \mathbf{C}\mathbf{a} \qquad \text{s.t. } \mathbf{a}^T \mathbf{s} = 1$$

The method of Lagrangian multipliers can be used to solve this problem. Define the Lagrangian function as

$$J\left(\mathbf{a},\lambda\right) = \mathbf{a}^T \mathbf{C}\mathbf{a} + \lambda\left(\mathbf{a}^T \mathbf{s} - 1\right), \qquad \lambda \in R$$

The gradient of J relative to **a** is

$$\frac{\partial J\left(\mathbf{a},\lambda\right)}{\partial \mathbf{a}} = 2\mathbf{C}\mathbf{a} + \lambda \mathbf{s} = 0$$

Finding the BLUE:

Solving for **a** produces

$$\mathbf{a} = -\frac{\lambda}{2}\mathbf{C}^{-1}\mathbf{s}$$

Using the constrain as

$$\mathbf{a}^T\mathbf{s} = -\frac{\lambda}{2}\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} = 1, \quad \text{or} \quad \lambda = -\frac{2}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}} \quad .$$

Finally, the solution is

$$\mathbf{a}_{opt} = -\frac{\mathbf{C}^{-1}\mathbf{s}}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}, \qquad \text{with a variance} \quad \mathrm{var}\left(\hat{\theta}\right) = \mathbf{a}_{opt}^T\mathbf{C}\mathbf{a}_{opt} = \frac{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{C}\mathbf{C}^{-1}\mathbf{s}}{\left(\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}\right)^2} = \frac{1}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}.$$

Taking into account that *E[**x**]=**s**θ,* finally the estimator

$$\hat{\theta} = \mathbf{a}_{opt}^T\mathbf{x} = \frac{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{x}}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}, \qquad \text{Its expected value is} \quad E\left(\hat{\theta}\right) = \frac{\mathbf{s}^T\mathbf{C}^{-1}E\left[\mathbf{x}\right]}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}} = \frac{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}\theta}{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}} = \theta!$$

Thus it is unbiased, as required.

DEM
DEPARTAMENTO
DE ENGENHARIA MECÂNICA
TÉCNICO LISBOA

Example:

Example (DC level in white ~~Gaussian~~ noise revisited):

Model of signal: $x\left[n\right] = A + w\left[n\right], \qquad n = 0,\dots,N-1$

Where $w[n]$ is zero mean white noise with variance σ^2 (and an unspecified PDF), the problem is to estimate A.

Because $E[x[n]]=A$, we have $\mathbf{s}=\mathbf{1}$, where $\mathbf{1}=[1\ 1\ \dots\ 1]^T$.

The BLUE is $\hat{A} = \dfrac{\mathbf{1}^T \frac{1}{\sigma^2}\mathbf{I}\mathbf{x}}{\mathbf{1}^T \frac{1}{\sigma^2}\mathbf{I}\mathbf{1}} = \dfrac{1}{N}\sum_{N=0}^{N-1} x\left[n\right] = \bar{x}$, and the variance is $\text{var}\left(\hat{A}\right) = \dfrac{1}{\mathbf{1}^T \frac{1}{\sigma^2}\mathbf{I}\mathbf{1}} = \dfrac{\sigma^2}{N}.$

Hence the sample mean is the BLUE independent of the PDF of data. It is the MVU estimator for the Gaussian case.

And in general: is it optimal?...

Example:

Example (DC level UNCORRELATED noise):

Model of signal: $x[n] = A + w[n], \qquad n = 0,\dots,N-1$

Where $w[n]$ is zero mean uncorrelated noise with $var(w[n])=\sigma^2$. Once again, the problem is to estimate A.

We have again $\boldsymbol{s=1}$, and $\boldsymbol{C}=diag(\sigma_0^2\ \sigma_1^2 \dots \sigma_{N-1}^2)$, and $\boldsymbol{C}^{-1}=diag(1/\sigma_0^2\ 1/\sigma_1^2 \dots 1/\sigma_{N-1}^2)$..

The BLUE is

$$\hat{A} = \frac{\mathbf{1}^T \mathbf{C}^{-1}\mathbf{x}}{\mathbf{1}^T \mathbf{C}^{-1}\mathbf{1}} = \frac{\sum_{N=0}^{N-1} \frac{x[n]}{\sigma_n^2}}{\sum_{N=0}^{N-1} \frac{1}{\sigma_n^2}}$$

and the variance is

$$\operatorname{var}(\hat{A}) = \frac{1}{\sum_{N=0}^{N-1} \frac{1}{\sigma_n^2}}$$

The BLUE weights those samples more heavily with smallest variances, in an attempt to equalize the noise contribution from each sample…

Is it optimal? In what cases?...

DEM
DEPARTAMENTO
DE ENGENHARIA MECÂNICA
TÉCNICO LISBOA

Extending BLUE to a Vector Parameter:

To find the BLUE for a $p \times 1$ vector parameter, we constrain the estimator

- to be linear $\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} x[n] \qquad i = 1,2,\ldots,p \quad \text{or} \quad \hat{\theta} = \mathbf{A}\mathbf{x}, \qquad \mathbf{A} \in R^{p \times N}$
- to be unbiased $E\left[\hat{\boldsymbol{\theta}}\right] = \mathbf{A}E\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}$
- to minimize its variance $\operatorname{var}\left(\hat{\theta}_i\right) = E\left[\left(\sum_{n=0}^{N-1} a_{in} x[n] - E\left[\sum_{n=0}^{N-1} ai_n x[n]\right]\right)^2\right]$

The problem of finding the BLUE can be stated as, for

$$\min \qquad \operatorname{var}\left(\hat{\theta}_i\right) = \mathbf{a}_i^T C \mathbf{a}_i$$

$$\text{subject to} \qquad \mathbf{a}_i^T E[\mathbf{x}] = \theta$$

where

$$E[\mathbf{x}] = \mathbf{H}\boldsymbol{\theta}.$$

DEM
DEPARTAMENTO
DE ENGENHARIA MECÂNICA
TÉCNICO LISBOA

Extension to non-white Gaussian noise:

Theorem 6.1(Gauss-Markov Theorem) *– If the data observed are of the general linear model form*

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

where **H** *is a known N x p observation matrix (with N>p) and rank p,* **x** *is a N x 1 vector of observations,* **θ** *is a p x 1 vector of parameters to be estimated, and* **w** *is an N x 1 noise vector with zero mean and covariance* ***C*** *(for an arbitrary PDF), then the BLUE is*

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x} \qquad \textit{(1bis)}$$

and the covariance matrix of estimate is

$$C_{\hat{\theta}} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}. \qquad \textit{(2bis)}$$

Bibliography:

Further reading

- Thomas Kailath, ***Linear Systems,*** Prentice Hall, 1980.
- Thomas Kailath, Ali Sayed, and Babak Hassibi, ***Linear Estimation,*** Prentice Hall, 2000.
- Harry L. Van Trees, ***Detection, Estimation, and Modulation Theory, Parts I to IV,*** John Wiley, 2001.
- J. Bibby, H. Toutenburg, ***Prediction and Improved Estimation in Linear Models,*** John Wiley, 1977.
- C.Rao, ***Linear Statistical Inference and Its Applications,*** John Wiley, 1973.