Lecture 4b: Linear MMSE Estimation

Optimal State Estimation

Guoquan (Paul) Huang

University of Delaware / Zhejiang University

June 20, 2018

MMSE Estimator

Consider $\mathbf{x}$ and $\mathbf{z}$ are jointly Gaussian distributed:

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{z} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{z}} \end{bmatrix}, \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xz} \\ \mathbf{C}_{zx} & \mathbf{C}_{zz} \end{bmatrix} \right) \quad (1)$$

The estimate of the r.v. $\mathbf{x}$ in terms of $\mathbf{z}$ according to the minimum mean square error (MMSE) criterion is the *conditional* mean of $\mathbf{x}$ given $\mathbf{z}$:

$$\hat{\mathbf{x}} \triangleq E[\mathbf{x}|\mathbf{z}] = \bar{\mathbf{x}} + \mathbf{C}_{xz}\mathbf{C}_{zz}^{-1}(\mathbf{z} - \bar{\mathbf{z}}) \quad (2)$$

$$\mathbf{C}_{xx|z} \triangleq E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T|\mathbf{z}] = \mathbf{C}_{xx} - \mathbf{C}_{xz}\mathbf{C}_{zz}^{-1}\mathbf{C}_{zx} \quad (3)$$

- The MMSE estimate – the conditional mean – of a Gaussian r.v. in terms of another Gaussian r.v. (the measurement) is a linear combination of:
  - The prior (unconditional) mean of the variable to be estimated
  - The difference btw. the measurement and its prior mean
- The conditional covariance of one Gaussian r.v. given another Gaussian r.v. (the measurement) is *independent* of the measurement.

Linear MMSE Estimator

- MMSE estimate of a random variable in terms of another random variable is the *conditional mean.*
- However, in many problems the distributed info needed for the evaluation of the conditional mean is not available. Even if it were available, the evaluation could be prohibitively complicated.
- *Linear MMSE estimation* is such that
  - The estimate is unbiased
  - The estimation error is uncorrelated from the measurements, that is, they are *orthogonal*, which is so called **principle of orthogonality**

Linear MMSE Estimation: Formulation

- To estimate a scalar parameter x (which is modeled as the realization of a random variable) based on the data $\mathbf{z} = [z(0) \ \cdots \ z(N-1)]^T$. We do not assume any specific form for the joint pdf $p(\mathbf{z}, x)$ but only the first two moments.

- The *linear minimum mean square error (LMMSE) estimator*:

$$\min_{\hat{x}} \ E[(x - \hat{x})^2] \tag{4}$$

$$s.t. \ \ \hat{x} = \sum_{n=0}^{N-1} a_i z(n) + a_N = \mathbf{a}^T \mathbf{z} + a_N \tag{5}$$

- Note that the estimator will be suboptimal (as we restrict to the class of linear (actually affine) estimators) unless the MMSE estimator happens to be linear, *in analogy to BLUE.*

- Note that LMMSE relies on the correlation between random variables, and a parameter uncorrelated with the data cannot be linearly estimated.

Linear MMSE Estimation: Derivation

- We now drive the optimal coefficients for the LMMSE (4)-(5). We first find a_N:

$$\frac{\partial}{\partial a_N} E\left[\left(x - \mathbf{a}^T\mathbf{z} + a_N\right)^2\right] = -2E\left[x - \mathbf{a}^T\mathbf{z} - a_N\right] = 0 \tag{6}$$

$$\Rightarrow \quad a_N = E[x] - \mathbf{a}^T E[\mathbf{z}] \tag{7}$$

- The cost function (4) becomes:

$$E\left[\left(x - \mathbf{a}^T\mathbf{z} + E[x] - \mathbf{a}^T E[\mathbf{z}]\right)^2\right] \overset{how?}{=} \mathbf{a}^T\mathbf{C}_{zz}\mathbf{a} - \mathbf{a}^T\mathbf{C}_{zx} - \mathbf{C}_{xz}\mathbf{a} + \mathbf{C}_{xx} \tag{8}$$

where $\mathbf{C}$ denotes the covariance matrix.

- We then find the other coefficients:

$$\frac{\partial E(\cdot)}{\partial \mathbf{a}} = 2\mathbf{C}_{zz}\mathbf{a} - 2\mathbf{C}_{zx} = \mathbf{0} \quad \Rightarrow \quad \mathbf{a} = \mathbf{C}_{zz}^{-1}\mathbf{C}_{zx} \tag{9}$$

- Finally, the LMMSE estimator:

$$\hat{x} = E[x] + \mathbf{C}_{xz}^T\mathbf{C}_{zz}^{-1}(\mathbf{z} - E[\mathbf{z}]) \tag{10}$$

Linear MMSE Estimation: Example

- Consider the measurement model (DC level in WGN):

$$z(n) = A + w(n) \quad n = 0, 1, \cdots, N-1 \tag{11}$$

where $A \sim \mathcal{U}[-A_0, A_0]$, $w(n)$ is WGN with variance σ^2, and A and $w(n)$ are independent. We wish to estimate A.

- The MMSE estimator cannot be obtained in closed form due to the integration (uniform distribution).

- We consider the LMMSE estimator, which is given by:

$$\hat{A} = \sigma_A^2 \mathbf{1}^T \left(\sigma_A^2 \mathbf{1}\mathbf{1}^T + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{z} \tag{12}$$

where $\sigma_A^2 = E[A^2]$.

Principle of Orthogonality

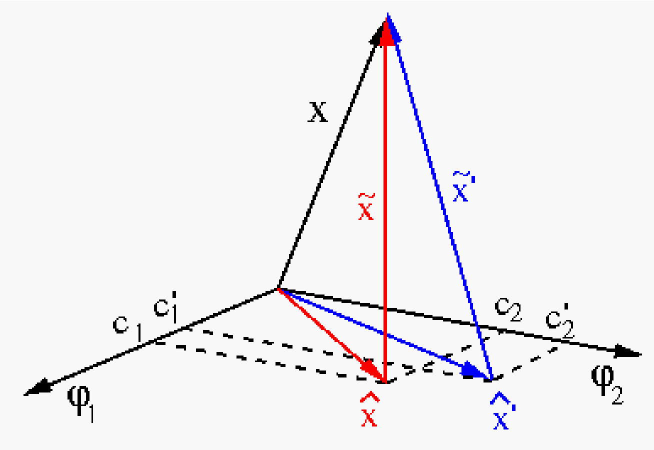

Figure 1: Orthogonal projection of r.v. x into the subspace spanned by z_i. In order to have the minimum error, it has to be orthogonal to the measurements.

Illustration with LMMSE Estimation

- Suppose a set of zero-mean random variables (measurements) $\mathbf{z}_i$ ($i = 1, \cdots, n$). Two vectors are orthogonal $\mathbf{z}_i \perp \mathbf{z}_j$, if and only if their inner product is zero, i.e.,

$$< \mathbf{z}_i, \mathbf{z}_j >= E[\mathbf{z}_i^T \mathbf{z}_j] = 0 \tag{13}$$

which is equivalent to these *zero-mean* r.v. being *uncorrelated.*

- The LMMSE estimator of a *zero-mean* random variable x is given by $\hat{x} = \sum_{i=1}^{n} a_i z_i$, and its norm of the estimation error $\tilde{x} \triangleq x - \hat{x}$ is minimized [see (4)-(5)]:

$$\min_{\{\beta_i\}_{i=1}^{n}} ||\tilde{x}||^2 = E[(x - \hat{x})^2] = E\left[(x - \sum_{i=1}^{n} a_i z_i)^2\right] \tag{14}$$

$$\Rightarrow -\frac{1}{2}\frac{\partial ||\tilde{x}||^2}{\partial a_j} = E\left[(x - \sum_{i=1}^{n} a_i z_i) z_j\right] = E[\tilde{x} z_j] =< \tilde{x}, z_j >= 0 \tag{15}$$

$$\Rightarrow \tilde{x} \perp z_j \ , \quad \forall j \qquad \text{it is orthogonal!} \tag{16}$$

The Vector LMMSE Estimator

- The best linear estimator that minimizes the scalar MSE criterion:

$$\min_{\mathbf{A},\mathbf{b}} J \triangleq E[(\mathbf{x}-\hat{\mathbf{x}})^T(\mathbf{x}-\hat{\mathbf{x}})] \tag{17}$$

$$= E[(\mathbf{x}-(\mathbf{A}\mathbf{z}+\mathbf{b}))^T(\mathbf{x}-(\mathbf{A}\mathbf{z}+\mathbf{b}))] \tag{18}$$

- The linear MMSE estimator is such that the estimation error is zero-mean and orthogonal to the observation:

$$E[\tilde{\mathbf{x}}] = \bar{\mathbf{x}} - (\mathbf{A}\mathbf{z}+\mathbf{b}) = 0 \tag{19}$$

$$\Rightarrow \ \mathbf{b} = \bar{\mathbf{x}} - \mathbf{A}\bar{\mathbf{z}} \tag{20}$$

$$E[\tilde{\mathbf{x}}\mathbf{z}^T] = E[(\mathbf{x}-\bar{\mathbf{x}}-\mathbf{A}(\mathbf{z}-\bar{\mathbf{z}}))\mathbf{z}^T] = \mathbf{P}_{xz} - \mathbf{A}\mathbf{P}_{zz} = 0 \tag{21}$$

$$\Rightarrow \ \mathbf{A} = \mathbf{P}_{xz}\mathbf{P}_{zz}^{-1} \tag{22}$$

- Therefore, the linear MMSE estimator for the multidimensional case:

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{P}_{xz}\mathbf{P}_{zz}^{-1}(\mathbf{z}-\bar{\mathbf{z}}) \tag{23}$$

$$E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] = \mathbf{P}_{xx} - \mathbf{P}_{xz}\mathbf{P}_{zz}^{-1}\mathbf{P}_{zx} =: \mathbf{P}_{xx|z} \tag{24}$$

which are the *fundamental equations of linear estimation.*

Sequential LMMSE Estimation

- Consider the example of DC level in WGN with Gaussian prior pdf:

$$z(n) = A + w(n) \quad n = 0, 1, \cdots, N-1 \tag{25}$$

$$p(A) = \mathcal{N}(0, \sigma_A^2), \quad w(n) \sim \mathcal{N}(0, \sigma^2) \tag{26}$$

- The LMMSE estimator based on $\{z(0), \cdots, z(N-1)\}$ can be found as:

$$\hat{A}(N-1) = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N} \frac{\sum_{n=0}^{N-1} z(n)}{N} = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N}\bar{z} \tag{27}$$

$$var(\hat{A}(N-1)) = \frac{\sigma_A^2 \sigma^2}{N\sigma_A^2 + \sigma^2} \tag{28}$$

- To update the estimator recursively as $z(N)$ becomes available:

$$\hat{A}(N) = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/(N+1)} \frac{\sum_{n=0}^{N} z(n)}{N+1} \tag{29}$$

$$= \hat{A}(N-1) + \underbrace{\frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2}}_{K(N)}(z(N) - \hat{A}(N-1)) \tag{30}$$

Sequential LMMSE Estimation (cont.)

- The gain can be computed as:

$$K(N) = \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} = \frac{var(\hat{A}(N-1))}{var(\hat{A}(N-1)) + \sigma^2} \tag{31}$$

- To update the minimum MSE (variance):

$$var(\hat{A}(N)) = \frac{\sigma_A^2\sigma^2}{(N+1)\sigma_A^2 + \sigma^2} = (1 - K(N))var(\hat{A}(N-1)) \tag{32}$$

- Summarize the sequential LMMSE estimator:
  - **Estimator Update:**

$$\hat{A}(N) = \hat{A}(N-1) + K(N)(z(N) - \hat{A}(N-1)) \tag{33}$$

$$K(N) = \frac{var(\hat{A}(N-1))}{var(\hat{A}(N-1)) + \sigma^2} \tag{34}$$

  - **Variance (Minimum MSE) Update:**

$$var(\hat{A}(N)) = (1 - K(N))var(\hat{A}(N-1)) \tag{35}$$