

430.714

Estimation Theory

Prof. Songhwai Oh
ECE, SNU

Linear Bayesian Estimators

Linear MMSE (LMMSE) or Linear Least Square Error (LLSE) Estimation

- Optimal Bayesian estimators are often too difficult to determine in closed form.
- They are too computationally intensive.
- Restrict estimators to be linear -> Linear MMSE estimator

LMMSE estimator is a linear estimator

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

which minimizes the Bayesian MSE

$$Bmse(\hat{\theta}) = E[(\theta - \hat{\theta})^2]$$

Clearly, the LMMSE estimator is suboptimal unless the MMSE estimator is linear.

Example: LMMSE is not always feasible

$x[0] \sim \mathcal{N}(0, \sigma^2)$ Parameter: $\theta = x^2[0]$. Estimate θ using $x[0]$.

Perfect Estimator: $\hat{\theta} = x^2[0]$ (Bayesian MSE is zero).

- Let us consider a linear estimator: $\hat{\theta} = a_0 x[0] + a_1$
- LMMSE estimator can be founded by minimizing

$$\begin{aligned} B_{\text{mse}}(\hat{\theta}) &= \mathbb{E} [(\theta - \hat{\theta})^2] \\ &= \mathbb{E} [(\theta - a_0 x[0] - a_1)^2] \end{aligned}$$

$$\frac{\partial}{\partial a_0} B_{\text{mse}}(\hat{\theta}) = 0 \implies \mathbb{E} [(\theta - a_0 x[0] - a_1)x[0]] = 0 \implies a_0 \mathbb{E}(x^2[0]) + a_1 \cancel{\mathbb{E}(x[0])} = \cancel{\mathbb{E}(\theta x[0])}$$

$$\frac{\partial}{\partial a_1} B_{\text{mse}}(\hat{\theta}) = 0 \implies \mathbb{E} [(\theta - a_0 x[0] - a_1)] = 0 \implies a_0 \cancel{\mathbb{E}(x[0])} + a_1 = \mathbb{E}(\theta)$$

$$\begin{aligned} a_0 &= 0 \\ a_1 &= \mathbb{E}(\theta) = \mathbb{E}(x^2[0]) = \sigma^2 \end{aligned} \implies \text{LMMSE Estimator: } \hat{\theta} = \sigma^2$$

$$B_{\text{mse}}(\hat{\theta}) = 2\sigma^4$$

No good!

LMMSE Estimator

$$\text{Bmse}(\hat{\theta}) = E[(\theta - \hat{\theta})^2]$$

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

$$\frac{\partial}{\partial a_N} E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] - a_N \right)^2 \right] = -2E \left[\theta - \sum_{n=0}^{N-1} a_n x[n] - a_N \right]$$

$$a_N = E(\theta) - \sum_{n=0}^{N-1} a_n E(x[n])$$

$$\text{Bmse}(\hat{\theta}) = E \left\{ \left[\sum_{n=0}^{N-1} a_n (x[n] - E(x[n])) - (\theta - E(\theta)) \right]^2 \right\}$$

$$\begin{aligned} \text{Bmse}(\hat{\theta}) &= E \left\{ [\mathbf{a}^T (\mathbf{x} - E(\mathbf{x})) - (\theta - E(\theta))]^2 \right\} \\ &= E [\mathbf{a}^T (\mathbf{x} - E(\mathbf{x})) (\mathbf{x} - E(\mathbf{x}))^T \mathbf{a}] - E [\mathbf{a}^T (\mathbf{x} - E(\mathbf{x})) (\theta - E(\theta))] \\ &\quad - E [(\theta - E(\theta)) (\mathbf{x} - E(\mathbf{x}))^T \mathbf{a}] + E [(\theta - E(\theta))^2] \\ &= \mathbf{a}^T \mathbf{C}_{xx} \mathbf{a} - \mathbf{a}^T \mathbf{C}_{x\theta} - \mathbf{C}_{\theta x} \mathbf{a} + C_{\theta\theta} \end{aligned}$$

$$\frac{\partial \text{Bmse}(\hat{\theta})}{\partial \mathbf{a}} = 2\mathbf{C}_{xx} \mathbf{a} - 2\mathbf{C}_{x\theta}$$

$$\mathbf{a} = \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}$$

$$\begin{aligned} \hat{\theta} &= \mathbf{a}^T \mathbf{x} + a_N \\ &= \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} \mathbf{x} + E(\theta) - \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} E(\mathbf{x}) \end{aligned}$$

LMMSE Estimator:

$$\hat{\theta} = E(\theta) + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x}))$$

LMMSE Estimator

$$\text{Bmse}(\hat{\theta}) = \mathbf{a}^T \mathbf{C}_{xx} \mathbf{a} - \mathbf{a}^T \mathbf{C}_{x\theta} - \mathbf{C}_{\theta x} \mathbf{a} + C_{\theta\theta}$$

$$\mathbf{a} = \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}$$

$$\begin{aligned} \text{Bmse}(\hat{\theta}) &= \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{xx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} - \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} \\ &\quad - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} + C_{\theta\theta} \\ &= \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} - 2 \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} + C_{\theta\theta} \end{aligned}$$

$$\boxed{\text{Bmse}(\hat{\theta}) = C_{\theta\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}} \quad ; \text{ Minimum Bayes MSE}$$

LMMSE Estimator:

$$\hat{\theta} = E(\theta) + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x}))$$

$$\text{Bmse}(\hat{\theta}) = C_{\theta\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}$$

cf. MMSE estimator for a multivariate Gaussian r.v.

When $\mathbb{E}(\theta) = 0$ and $\mathbb{E}(\mathbf{x}) = 0$, $\hat{\theta} = \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{x}$.

Example

$$x[n] = A + w[n], \quad n = 0, 1, \dots, N-1, \quad \text{where } A \sim \mathcal{U}[-A_0, A_0], w[n] \sim \mathcal{N}(0, \sigma^2)$$

MMSE estimator cannot be obtained in closed form.

LMMSE estimator:

$$\begin{aligned} E(\mathbf{x}) &= \mathbf{0} & \mathbf{C}_{xx} &= E(\mathbf{x}\mathbf{x}^T) & \mathbf{C}_{\theta x} &= E(A\mathbf{x}^T) \\ \sigma_A^2 &= E(A^2) & &= E[(A\mathbf{1} + \mathbf{w})(A\mathbf{1} + \mathbf{w})^T] & &= E[A(A\mathbf{1} + \mathbf{w})^T] \\ & & &= E(A^2)\mathbf{1}\mathbf{1}^T + \sigma^2\mathbf{I} & &= E(A^2)\mathbf{1}^T \end{aligned}$$

$$\begin{aligned} \hat{A} &= \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{x} \\ &= \sigma_A^2 \mathbf{1}^T (\sigma_A^2 \mathbf{1}\mathbf{1}^T + \sigma^2 \mathbf{I})^{-1} \mathbf{x} \end{aligned} \quad \boxed{(\mathbf{A} + \mathbf{u}\mathbf{u}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1} \mathbf{u}\mathbf{u}^T \mathbf{A}^{-1}}{1 + \mathbf{u}^T \mathbf{A}^{-1} \mathbf{u}}}$$

$$\hat{A} = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \bar{x} = \frac{\frac{A_0^2}{3}}{\frac{A_0^2}{3} + \frac{\sigma^2}{N}} \bar{x} \quad \sigma_A^2 = E(A^2) = (2A_0)^2/12 = A_0^2/3$$

- We don't care about the distribution of A.
- We only need to know its mean and variance to compute the LMMSE estimator.
- It is required that A and w are **uncorrelated** (as opposed to the independence assumption).
- But the LMMSE estimator is suboptimal.
- MMSE estimator:

$$\begin{aligned} \hat{A} &= E(A|\mathbf{x}) \\ &= \int_{-\infty}^{\infty} A p(A|\mathbf{x}) dA = \frac{\int_{-A_0}^{A_0} A \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}}(A - \bar{x})^2\right] dA}{\int_{-A_0}^{A_0} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}}(A - \bar{x})^2\right] dA}. \end{aligned}$$

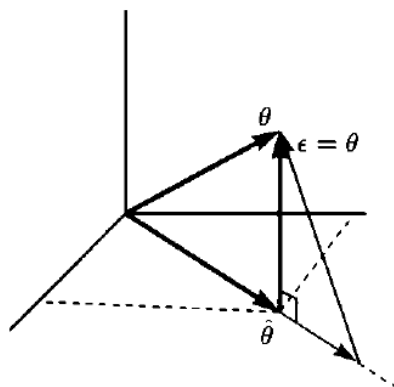
Estimation

Geometrical Interpretations

- Vector space of random variables (assume **mean zero** vectors).
- Inner product between X and Y is $E(XY)$.
- Random variables X and Y are orthogonal if $E(XY) = 0$ (i.e., uncorrelated).

LMMSE estimator $\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n]$ minimizes $E[(\theta - \hat{\theta})^2] = E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] \right)^2 \right]$

$$= \left\| \theta - \sum_{n=0}^{N-1} a_n x[n] \right\|^2.$$



Orthogonality principle (a.k.a. projection theorem):

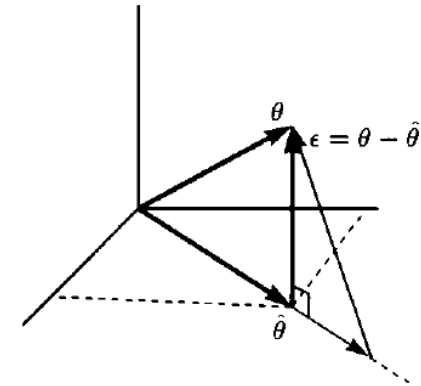
- $\hat{\theta}$ is in the subspace $\mathcal{S}$ spanned by $\{x[0], x[1], \dots, x[N-1]\}$.
- Error vector $\epsilon = \theta - \hat{\theta}$ is minimized if ϵ is orthogonal to the subspace $\mathcal{S}$.
- $\epsilon \perp \mathcal{S}$
- $\epsilon \perp x[0], x[1], \dots, x[N-1]$
- $\mathbb{E}[(\theta - \hat{\theta})x[n]] = 0$ for $n = 0, 1, \dots, N-1$

Geometrical Interpretations

Orthogonality: $\mathbb{E}[(\theta - \hat{\theta})x[n]] = 0$ for $n = 0, 1, \dots, N-1$

$$\mathbb{E} \left[\left(\theta - \sum_{m=0}^{N-1} a_m x[m] \right) x[n] \right] = 0 \quad n = 0, 1, \dots, N-1$$

$$\sum_{m=0}^{N-1} a_m \mathbb{E}(x[m]x[n]) = \mathbb{E}(\theta x[n]) \quad n = 0, 1, \dots, N-1$$



$$\begin{bmatrix} E(x^2[0]) & E(x[0]x[1]) & \dots & E(x[0]x[N-1]) \\ E(x[1]x[0]) & E(x^2[1]) & \dots & E(x[1]x[N-1]) \\ \vdots & \vdots & \ddots & \vdots \\ E(x[N-1]x[0]) & E(x[N-1]x[1]) & \dots & E(x^2[N-1]) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{bmatrix} = \begin{bmatrix} E(\theta x[0]) \\ E(\theta x[1]) \\ \vdots \\ E(\theta x[N-1]) \end{bmatrix}$$

$$\mathbf{C}_{xx} \mathbf{a} = \mathbf{C}_{x\theta}$$

$$\mathbf{a} = \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}$$

$$\hat{\theta} = \mathbf{a}^T \mathbf{x}$$

$$= \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} \mathbf{x} = \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{x}$$

$$\begin{aligned} \text{Bmse}(\hat{\theta}) &= E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] \right) \left(\theta - \sum_{m=0}^{N-1} a_m x[m] \right) \right] \\ &= E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] \right) \theta \right] - E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] \right) \sum_{m=0}^{N-1} a_m x[m] \right] \\ &= E(\theta^2) - \sum_{n=0}^{N-1} a_n E(x[n]\theta) - \sum_{m=0}^{N-1} a_m E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] \right) x[m] \right] \end{aligned}$$

Orthogonality principle

$$\begin{aligned} \text{Bmse}(\hat{\theta}) &= C_{\theta\theta} - \mathbf{a}^T \mathbf{C}_{x\theta} \\ &= C_{\theta\theta} - \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} \\ &= C_{\theta\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} \end{aligned}$$

Vector LMMSE Estimator

$$\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in}x[n] + a_{iN} \quad \text{for } i = 1, 2, \dots, p \quad \text{minimizes the Bayes MSE for each } i.$$

$$\hat{\theta}_i = E(\theta_i) + \mathbf{C}_{\theta_i x} \mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x})) \quad i = 1, 2, \dots, p$$

$$\text{Bmse}(\hat{\theta}_i) = C_{\theta_i \theta_i} - \mathbf{C}_{\theta_i x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x \theta_i} \quad i = 1, 2, \dots, p.$$

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \begin{bmatrix} E(\theta_1) \\ E(\theta_2) \\ \vdots \\ E(\theta_p) \end{bmatrix} + \begin{bmatrix} \mathbf{C}_{\theta_1 x} \mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x})) \\ \mathbf{C}_{\theta_2 x} \mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x})) \\ \vdots \\ \mathbf{C}_{\theta_p x} \mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x})) \end{bmatrix} \\ &= \begin{bmatrix} E(\theta_1) \\ E(\theta_2) \\ \vdots \\ E(\theta_p) \end{bmatrix} + \begin{bmatrix} \mathbf{C}_{\theta_1 x} \\ \mathbf{C}_{\theta_2 x} \\ \vdots \\ \mathbf{C}_{\theta_p x} \end{bmatrix} \mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x})) \\ &= E(\boldsymbol{\theta}) + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x})) \end{aligned}$$

$$\begin{aligned} \mathbf{M}_{\hat{\boldsymbol{\theta}}} &= E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T] \\ &= \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} \end{aligned}$$

$$\text{Bmse}(\hat{\theta}_i) = [\mathbf{M}_{\hat{\boldsymbol{\theta}}}]_{ii}$$

Properties of the LMMSE Estimator

1. Commutes over linear (affine) transformations

$$\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\theta} + \mathbf{b} \longrightarrow \hat{\boldsymbol{\alpha}} = \mathbf{A}\hat{\boldsymbol{\theta}} + \mathbf{b}$$

2. LMMSE estimator of a sum of unknown parameters is the sum of the individual estimators

$$\boldsymbol{\alpha} = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2 \longrightarrow \hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\theta}}_1 + \hat{\boldsymbol{\theta}}_2$$

$$\hat{\boldsymbol{\theta}}_1 = E(\boldsymbol{\theta}_1) + \mathbf{C}_{\boldsymbol{\theta}_1 x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x}))$$

$$\hat{\boldsymbol{\theta}}_2 = E(\boldsymbol{\theta}_2) + \mathbf{C}_{\boldsymbol{\theta}_2 x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})).$$

3. LMMSE estimator = MMSE estimator for linear Gaussian models, for which the MMSE estimator is linear in data

Theorem 12.1 (Bayesian Gauss-Markov Theorem) *If the data are described by the Bayesian linear model form*

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w} \quad (12.25)$$

where $\mathbf{x}$ is an $N \times 1$ data vector, $\mathbf{H}$ is a known $N \times p$ observation matrix, $\boldsymbol{\theta}$ is a $p \times 1$ random vector of parameters whose realization is to be estimated and has mean $E(\boldsymbol{\theta})$ and covariance matrix $\mathbf{C}_{\theta\theta}$, and $\mathbf{w}$ is an $N \times 1$ random vector with zero mean and covariance matrix $\mathbf{C}_w$ and is uncorrelated with $\boldsymbol{\theta}$ (the joint PDF $p(\mathbf{w}, \boldsymbol{\theta})$ is otherwise arbitrary), then the LMMSE estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}) + \mathbf{C}_{\theta\theta}\mathbf{H}^T(\mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_w)^{-1}(\mathbf{x} - \mathbf{H}E(\boldsymbol{\theta})) \quad (12.26)$$

$$= E(\boldsymbol{\theta}) + (\mathbf{C}_{\theta\theta}^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}_w^{-1}(\mathbf{x} - \mathbf{H}E(\boldsymbol{\theta})). \quad (12.27)$$

The performance of the estimator is measured by the error $\boldsymbol{\epsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ whose mean is zero and whose covariance matrix is

$$\begin{aligned} \mathbf{C}_{\boldsymbol{\epsilon}} &= E_{\mathbf{x}, \boldsymbol{\theta}}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) \\ &= \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta\theta}\mathbf{H}^T(\mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_w)^{-1}\mathbf{H}\mathbf{C}_{\theta\theta} \end{aligned} \quad (12.28)$$

$$= (\mathbf{C}_{\theta\theta}^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H})^{-1}. \quad (12.29)$$

The error covariance matrix is also the minimum MSE matrix $\mathbf{M}_{\hat{\boldsymbol{\theta}}}$ whose diagonal elements yield the minimum Bayesian MSE

$$\begin{aligned} [\mathbf{M}_{\hat{\boldsymbol{\theta}}}]_{ii} &= [\mathbf{C}_{\boldsymbol{\epsilon}}]_{ii} \\ &= \text{Bmse}(\hat{\theta}_i). \end{aligned} \quad (12.30)$$