ELEG 5633 Detection and Estimation
Minimum Mean Squared Error Estimation

Jingxian Wu

Department of Electrical Engineering
University of Arkansas

Outline

- Minimum Mean Square Error (MMSE) Estimator
- Performance Metrics
- Linear Minimum Mean Square Error (LMMSE) Estimators
- Orthogonality Principle

MMSE Estimator (Scalar Case)

- Mean square error (MSE)

$$\mathbb{E}[(\theta - \hat{\theta})^2 | X = x]$$

- MMSE Estimator: find $\hat{\theta}$ to minimize the MSE

$$\text{minimize}_{\hat{\theta}} \quad \mathbb{E}[(\theta - \hat{\theta})^2 | X = x]$$

- Solution:
  - $\mathbb{E}[(\theta - \hat{\theta})^2 | X = x] = \int_{-\infty}^{\infty} (\theta - \hat{\theta}(x))^2 p(\theta | x) d\theta$
  - Taking derivative with respect to $\hat{\theta}$,

$$-2 \int_{-\infty}^{\infty} \theta p(\theta | x) d\theta + 2\hat{\theta}(x) \int_{-\infty}^{\infty} p(\theta | x) d\theta = 0$$

  - Thus

$$\hat{\theta}_{\text{MMSE}}(x) = \int_{-\infty}^{\infty} \theta p(\theta | x) d\theta = \mathbb{E}[\theta | X = x]$$

$\hat{\theta}_{\text{MMSE}}$ is the posterior mean of θ

Example

Consider N conditionally i.i.d observations generated according to $X_i = A + W_i$, where $W_i \sim \mathcal{N}(0, \sigma^2)$, A is a random parameter uniformly distributed on $[-A_0, A_0]$. A and W_i are independent. What is the MMSE estimator of A?

Solution: Page 314, [M. Kay Volumn 1]. Let $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$, $\sigma_0 = \frac{\sigma^2}{N}$

$$
\begin{aligned}
p(A|\mathbf{x}) &= \frac{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i - A)^2\right]}{\int_{-A_0}^{A_0}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i - A)^2\right] dA} \\
&= \frac{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N} X_i^2\right]\exp\left[-\frac{1}{2\sigma^2}(-2AN\bar{X} + NA^2)\right]}{\int_{-A_0}^{A_0}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N} X_i^2\right]\exp\left[-\frac{1}{2\sigma^2}(-2AN\bar{X} + NA^2)\right] dA} \\
&= \frac{\exp\left[-\frac{1}{2\sigma_0^2}(-2A\bar{X} + A^2)\right]}{\int_{-A_0}^{A_0}\exp\left[-\frac{1}{2\sigma_0^2}(-2A\bar{X} + A^2)\right] dA}
\end{aligned}
$$

Solution (Cont'd)

$$
\begin{aligned}
p(A|\mathbf{x}) &= \frac{\exp\left[-\frac{1}{2\sigma_0^2}(\bar{X}^2 - 2A\bar{X} + A^2)\right]}{\int_{-A_0}^{A_0} \exp\left[-\frac{1}{2\sigma_0^2}(\bar{X}^2 - 2A\bar{X} + A^2)\right] dA} \\
&= \frac{\frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2\sigma_0^2}(A - \bar{X})^2\right]}{\int_{-A_0}^{A_0} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2\sigma_0^2}(A - \bar{X})^2\right] dA} \\
&= \frac{1}{c}\frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2\sigma_0^2}(A - \bar{X})^2\right], \quad -A_0 \le A \le A_0
\end{aligned}
$$

where

$$
c = \int_{-A_0}^{A_0} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2\sigma_0^2}(A - \bar{X})^2\right] dA = Q\left(\frac{A_0 - \bar{X}}{\sigma_0}\right) - Q\left(-\frac{A_0 + \bar{X}}{\sigma_0}\right)
$$

Solution (Cont'd)

The MMSE estimate is

$$\begin{aligned}
\hat{A}_{\text{MMSE}} &= \mathbb{E}[A|\mathbf{x}] = \int_{-A_0}^{A_0} Ap(A|\mathbf{x})dA \\
&= \frac{1}{c}\int_{-A_0}^{A_0} \frac{A}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2\sigma_0^2}(A-\bar{X})^2\right] dA \\
&= \frac{1}{c}\int_{-A_0}^{A_0} \frac{A-\bar{X}}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2\sigma_0^2}(A-\bar{X})^2\right] dA + \bar{X} \\
&= \frac{1}{c}\int_{-A_0}^{A_0} \frac{\sigma_0}{\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_0^2}(A-\bar{X})^2\right] d\frac{(A-\bar{X})^2}{2\sigma_0^2} + \bar{X} \\
&= \frac{\sigma_0}{c\sqrt{2\pi}}\left[\exp\left(\frac{\bar{X}+A_0}{2\sigma_0^2}\right) - \exp\left(\frac{\bar{X}-A_0}{2\sigma_0^2}\right)\right] + \bar{X}
\end{aligned}$$

Vector MMSE

- $\boldsymbol{\theta} \in \mathbb{R}^p$ is a vector: $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_p]^T$
- $\ell(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2$
- MSE

$$\mathbb{E}[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2 | X = x]$$

- MMSE

$$\text{minimize}_{\hat{\boldsymbol{\theta}}} \quad \mathbb{E}[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2 | X = x]$$

- Solution

$$\hat{\boldsymbol{\theta}}_{\text{MMSE}}(x) = \int_{-\infty}^{\infty} \boldsymbol{\theta} p(\boldsymbol{\theta}|x) d\theta = \mathbb{E}[\boldsymbol{\theta}|X = x]$$

  i.e., $\hat{\theta}_i = \mathbb{E}[\theta_i | X = x] = \int \theta_i p(\theta_i | x) d\theta_i, \quad i = 1, 2, \ldots, p$

- This is the MMSE estimator that minimizes $\mathbb{E}[(\hat{\theta}_i - \theta_i)^2]$.
- The vector MMSE estimator $\mathbb{E}[\boldsymbol{\theta}|X = x]$ minimizes the MSE for each component of the vector parameter!

Example

If $\boldsymbol{\theta}$ and $\mathbf{x}$ are jointly Gaussian so that

$$\begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{x} \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$$

where

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_\theta \\ \boldsymbol{\mu}_x \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} \mathbf{C}_{\theta\theta} & \mathbf{C}_{\theta x} \\ \mathbf{C}_{x\theta} & \mathbf{C}_{xx} \end{bmatrix}.$$

Find the MMSE estimator of $\boldsymbol{\theta}$ based on $\mathbf{x}$.

Solution: Review: X, Y are joint Gaussian random vectors. Conditional pdf of Y given $X = x$ is still Gaussian, with

$$\mu_{Y|X}(x) = \mu_Y + \Sigma_{YX}\Sigma_X^{-1}(x - \mu_x)$$
$$\Sigma_{Y|X}(x) = \Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{YX}^T$$

$$\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}} = \mathbb{E}[\boldsymbol{\theta}|\mathbf{x}] = \boldsymbol{\mu}_{\theta|x}(\mathbf{x}) = \boldsymbol{\mu}_\theta + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$$

The solution is linear in $\mathbf{x}$!

Example

Consider N conditionally i.i.d observations generated according to $X_i = A + W_i$, where $W_i \sim \mathcal{N}(0, \sigma^2)$, A is a random and $A \sim \mathcal{N}(0, \sigma_A^2)$.$A$ and W_i are independent. What is the MMSE estimator of A?

(Example 10.1 on Page 317, [M. Kay Volumn 1].)

Solution:

- Let $\mathbf{X} = [X_1, \cdots, X_N]^T$, then $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_x, \mathbf{C}_x)$

$$\boldsymbol{\mu}_x = \mathbf{0}, \qquad \mathbf{C}_x = \sigma_A^2 \mathbf{P}_N + \sigma^2 \mathbf{I}_N$$

  where $\mathbf{P}_N$ is a size $N \times N$ all-one matrix.

- Based on the relationship $(a\mathbf{P}_N + b\mathbf{I}_N)^{-1} = -\frac{a}{b(Na+b)}\mathbf{P}_N + \frac{1}{b}\mathbf{I}_N$, we have

$$\mathbf{C}_x^{-1} = \frac{1}{\sigma^2}\mathbf{I}_N - \frac{\sigma_A^2}{\sigma^2(N\sigma_A^2 + \sigma^2)}\mathbf{P}_N$$

Solution: (Cont'd)

- $$\mathbf{c}_{Ax} = [\sigma_A^2, \cdots, \sigma_A^2] = \sigma_A^2 \mathbf{1}_N^T$$

  where $\mathbf{1}_N$ is a length-N all-one column vector.
- The MMSE estimate is then

$$\begin{aligned}
\hat{A}_{\mathsf{MMSE}} &= \boldsymbol{\mu}_\theta + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x) = \mathbf{c}_{Ax}\mathbf{C}_x^{-1}\mathbf{x} \\
&= \left(\gamma_0 \mathbf{1}_N^T - \gamma_0 \frac{N\gamma_0}{N\gamma_0 + 1}\mathbf{1}^T\right)\mathbf{x} \\
&= (\gamma_0 - \gamma_0 \frac{N\gamma_0}{N\gamma_0 + 1}) \sum_{i=1}^{N} x_i \\
&= \frac{\gamma_0}{\gamma_0 + \frac{1}{N}}\bar{X} = \frac{\sigma_A^2}{\sigma_A^2 + \frac{1}{N}\sigma^2}\bar{X} = \hat{A}_{\mathsf{MAP}}
\end{aligned}$$

where $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ and $\gamma_0 = \frac{\sigma_A^2}{\sigma^2}$

Performance Metrics of Estimators

- Estimation error: $e(\theta, x) := \hat{\theta}(x) - \theta \Rightarrow$ A random variable!

- Bias b:

$$\text{bias}(\hat{\theta}) := \mathbb{E}_{\theta,x}[e] = \mathbb{E}_x[\hat{\theta}(x)] - \mathbb{E}_\theta[\theta]$$

  - If $\text{bias}(\hat{\theta}) = 0$, $\mathbb{E}_x[\hat{\theta}(x)] = \mathbb{E}_\theta[\theta]$, $\hat{\theta}$ is an unbiased estimator of θ.
  - If $\text{bias}(\hat{\theta}) \neq 0$, $\mathbb{E}_x[\hat{\theta}(x)] \neq \mathbb{E}_\theta[\theta]$, $\hat{\theta}$ is an biased estimator of θ.

- Error Covariance Matrix:

$$\mathbf{\Sigma}_e = \mathbb{E}[(\mathbf{e} - \mathbf{b})(\mathbf{e} - \mathbf{b})^T]$$

$$\mathbb{E}[\mathbf{e}\mathbf{e}^T] = \mathbf{\Sigma}_e + \mathbf{b}\mathbf{b}^T$$

Performance Metrics of Estimators (Cont'd)

- Bayesian Mean Square Error (BMSE)

$$\begin{aligned}\text{BMSE}(\hat{\theta}) &= \mathbb{E}_{\theta,x}\|\theta - \hat{\theta}(x)\|_2^2 \\ &= \mathbb{E}[\mathbf{e}^T\mathbf{e}] = \text{tr}(\mathbb{E}[\mathbf{e}\mathbf{e}^T]) \\ &= \text{tr}(\mathbf{\Sigma}_e) + \text{tr}[\mathbf{b}\mathbf{b}^T]\end{aligned}$$

Performance of Scalar MMSE

- $e(x,\theta) = \hat{\theta}_{\text{MMSE}}(x) - \theta = \mathbb{E}[\theta|X=x] - \theta$
- bias $b = \mathbb{E}_{x,\theta}[e(x,\theta)] = \mathbb{E}_x[\mathbb{E}[\theta|X=x] - \mathbb{E}_\theta[\theta] = \mathbb{E}_\theta[\theta] - \mathbb{E}_\theta[\theta] = 0$
  MMSE is an unbiased estimator!
- MMSE minimizes $\mathbb{E}_\theta[\|\hat{\theta}-\theta\|^2|X=x]$ for all x, thus it minimizes the BMSE

$$\mathbb{E}[\|\hat{\theta}-\theta\|^2] = \mathbb{E}_x\left[\mathbb{E}_\theta[\|\hat{\theta}-\theta\|^2|X=x]\right]$$

  Bayesian MSE is defined as $\mathbb{E}[\|\hat{\theta}-\theta\|^2]$, thus MMSE minimizes BMSE.
- Variance of e: var$(e) = \mathbb{E}[(\hat{\theta}_{\text{MMSE}}(x) - \theta)^2]$ =minimum BMSE

Performance of Vector MMSE

- $\mathbf{e}(x,\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}}_{\mathsf{MMSE}}(x) - \boldsymbol{\theta} = \mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\theta}|X=x] - \boldsymbol{\theta}$
- bias $\mathbf{b} = \mathbb{E}_{x,\boldsymbol{\theta}}[e(x,\boldsymbol{\theta})] = \mathbb{E}_{x,\boldsymbol{\theta}}[\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\theta}|X=x] - \boldsymbol{\theta}] = \mathbb{E}[\boldsymbol{\theta}] - \mathbb{E}[\boldsymbol{\theta}] = 0$
  Vector MMSE is unbiased!
- Error Covariance Matrix:

$$\boldsymbol{\Sigma}_e = \mathbb{E}[(\mathbf{e}-\mathbf{b})(\mathbf{e}-\mathbf{b})^T] = \mathbb{E}[\mathbf{e}\mathbf{e}^T]$$

- Bayesian MSE $\mathbb{E}[\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2]$, thus Vector MMSE minimizes BMSE.
- The corresponding BMSE$= \mathbb{E}[\mathbf{e}^T\mathbf{e}] = \mathrm{tr}(\boldsymbol{\Sigma}_e) = \sum_i \mathbb{E}[(\hat{\theta}_{i\mathsf{MMSE}}(x) - \theta_i)^2]$

Example

If $\boldsymbol{\theta}$ and $\mathbf{x}$ are jointly Gaussian so that $[\boldsymbol{\theta}, \mathbf{x}]^T \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$, where $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_\theta \\ \boldsymbol{\mu}_x \end{bmatrix}$ and $\mathbf{C} = \begin{bmatrix} \mathbf{C}_{\theta\theta} & \mathbf{C}_{\theta x} \\ \mathbf{C}_{x\theta} & \mathbf{C}_{xx} \end{bmatrix}$. Find the error covariance matrix of the MMSE estimator.

Solutions: $\hat{\boldsymbol{\theta}}_{\text{MMSE}} = \mathbb{E}[\boldsymbol{\theta}|\mathbf{x}] = \boldsymbol{\mu}_\theta + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$

$$\begin{aligned}
\boldsymbol{\Sigma}_e &= \mathbb{E}[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T] \\
&= \mathbb{E}[(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)][(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)]^T \\
&= \mathbf{C}_\theta - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} \\
&= \mathbf{C}_\theta - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}
\end{aligned}$$

Example

Consider N conditionally i.i.d observations generated according to $X_i = A + W_i$, where $W_i \sim \mathcal{N}(0, \sigma^2)$, A is a random and $A \sim \mathcal{N}(0, \sigma_A^2)$.$A$ and W_i are independent. What is the error variance of the MMSE estimator of A?

Solutions:

- $\sigma_e^2 = \mathbf{C}_\theta - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta}$
- $\mathbf{C}_\theta = \sigma_A^2$, $\mathbf{C}_{\theta x} = \sigma_A^2 \mathbf{1}_N^T$, $\mathbf{C}_x^{-1} = \frac{1}{\sigma^2}\mathbf{I}_N - \frac{\sigma_A^2}{\sigma^2(N\sigma_A^2+\sigma^2)}\mathbf{P}_N$
- $$\begin{aligned}\sigma_e^2 &= \sigma_A^2 - \sigma_A^2 \mathbf{1}_N^T \left(\frac{1}{\sigma^2}\mathbf{I}_N - \frac{\sigma_A^2}{\sigma^2(N\sigma_A^2+\sigma^2)}\mathbf{P}_N\right)\sigma_A^2 \mathbf{1}_N \\ &= \frac{\sigma_A^2}{N\gamma_0 + 1} = \frac{\sigma_A^2\sigma^2}{N\sigma_A^2 + \sigma^2}\end{aligned}$$

Example

Assume

$$x_n = a\cos(2\pi f_0 n) + b\sin(2\pi f_0 n) + w_n, n = 0, 1, \ldots, N-1$$

where $f_0 = 1/N$, and $w[n]$ is WGN with variance σ^2. It is desired to estimate $\boldsymbol{\theta} = [a, b]^T$, under the assumption that $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \sigma_\theta^2 \mathbf{I})$, and θ is independent of $w[n]$. Find the MMSE estimator of $\boldsymbol{\theta}$.

Note $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where $\mathbf{H} = \begin{bmatrix} 1 & 0 \\ \cos 2\pi/N & \sin 2\pi/N \\ \vdots & \vdots \\ \cos 2\pi(N-1)/N & \sin 2\pi(N-1)/N \end{bmatrix}$ Find the MMSE estimator and error covariance matrix.

Solutions:

$\mathbf{x}$ and $\boldsymbol{\theta}$ are jointly Gaussian distributed

$$\hat{\boldsymbol{\theta}}_{\text{MMSE}} = \mathbb{E}[\boldsymbol{\theta}|\mathbf{x}] = \boldsymbol{\mu}_\theta + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$$

Solutions: (Cont'd))

- $\boldsymbol{\mu}_x = \boldsymbol{\mu}_\theta = \mathbf{0}$
- $$\begin{aligned}\mathbf{C}_{xx} &= \mathbb{E}[(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)^T] = \mathbb{E}[(\mathbf{H}\boldsymbol{\theta}+\mathbf{w})(\mathbf{H}\boldsymbol{\theta}+\mathbf{w})^T] \\ &= \mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \mathbf{C}_w\end{aligned}$$
- $\mathbf{C}_{\theta x} = \mathbb{E}[(\boldsymbol{\theta}-\boldsymbol{\mu}_\theta)(\mathbf{x}-\boldsymbol{\mu}_x)^T] = \mathbb{E}[\boldsymbol{\theta}(\mathbf{H}\boldsymbol{\theta}+\mathbf{w})^T] = \mathbf{C}_\theta\mathbf{H}^T$
- $\hat{\boldsymbol{\theta}}_{\text{MMSE}} = \mathbf{C}_\theta\mathbf{H}^T(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T+\mathbf{C}_w)^{-1}\mathbf{x} = \mathbf{H}^T\left(\mathbf{H}\mathbf{H}^T + \frac{\sigma^2}{\sigma_\theta^2}\mathbf{I}_N\right)^{-1}\mathbf{x}$
- Error covariance matrix

$$\begin{aligned}\mathbf{C}_e &= \mathbb{E}\left[(\boldsymbol{\theta}-\mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{x})(\boldsymbol{\theta}-\mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{x})^T\right] \\ &= \mathbf{C}_\theta - \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\theta} \\ &= \mathbf{C}_\theta - \mathbf{C}_\theta\mathbf{H}^T(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T+\mathbf{C}_w)^{-1}\mathbf{H}\mathbf{C}_\theta\end{aligned}$$

An important lemma:
$\mathbf{C}_\theta - \mathbf{C}_\theta\mathbf{H}^T(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T+\mathbf{C}_w)^{-1}\mathbf{H}\mathbf{C}_\theta = (\mathbf{C}_\theta^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H})^{-1}$

Linear Minimum MSE (LMMSE) Estimator

- Optimal Bayesian estimators are difficult to determine in closed form, and computationally intensive to implement.
  - Multidimensional integration for MMSE estimator;
  - Multidimensional maximization for MAP estimator

- Under jointly Gaussian assumption, estimators can be easily found.
  - The MMSE estimator is linear in the observation $\mathbf{x}$.
  - The MAP estimator is the same as the MMSE estimator.

$$\hat{\boldsymbol{\theta}}_{\text{MMSE}} = \hat{\boldsymbol{\theta}}_{\text{MAP}} = \boldsymbol{\mu}_\theta + \mathbf{C}_{\theta x}\mathbf{C}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$$

Linear Minimum MSE (LMMSE) Estimator

For non-Gaussian case, we want to retain the MMSE criterion, but constrain the estimator to be linear.

- Assume $\boldsymbol{\theta} \in \mathbb{R}^n$, $\mathbf{x} \in \mathbb{R}^k$, $\mathbb{E}[\boldsymbol{\theta}] = \mathbf{0}$, $\mathbb{E}[\mathbf{x}] = \mathbf{0}$.
- Linear estimator: $\hat{\boldsymbol{\theta}} = \mathbf{A}^T\mathbf{x}, \quad \mathbf{A} \in \mathbb{R}^{k \times n}$

$$\begin{aligned} \text{minimize}_{\mathbf{A}} \quad & \mathbb{E}_{\theta,x}[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_2^2] \\ \text{subject to} \quad & \hat{\boldsymbol{\theta}} = \mathbf{A}^T\mathbf{x} \end{aligned}$$

- For Gaussian case, the LMMSE estimator is the same is MMSE estimator.
- For non-Gaussian case, the LMMSE estimator is sub-optimum compared to the MMSE estimator (which is generally non-linear)

LMMSE Estimator

- Let $\hat{\boldsymbol{\theta}} = \mathbf{A}^T\mathbf{x}$, we want to find the $\mathbf{A}$ to minimize BMSE $\mathbb{E}_{\theta,x}[\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2]$.
Solutions:

$$\begin{aligned}
\text{BMSE}(\mathbf{A}) &= \mathbb{E}[\|\boldsymbol{\theta} - \mathbf{A}^T\mathbf{x}\|^2] \\
&= \mathbb{E}\Big[\text{tr}\Big((\boldsymbol{\theta} - \mathbf{A}^T\mathbf{x})(\boldsymbol{\theta} - \mathbf{A}^T\mathbf{x})^T\Big)\Big] \\
&= \text{tr}\Big(\mathbb{E}\Big[(\boldsymbol{\theta} - A^T\mathbf{x})(\boldsymbol{\theta} - A^T\mathbf{x})^T\Big]\Big) \\
&= \text{tr}\big(\Sigma_{\theta\theta} - \mathbf{A}^T\Sigma_{x\theta} - \Sigma_{\theta x}\mathbf{A} + \mathbf{A}^T\Sigma_{xx}A\big)
\end{aligned}$$

  - Set $\frac{\partial \text{BMSE}(\mathbf{A})}{\partial \mathbf{A}} = 0 \Rightarrow -2\Sigma_{x\theta} + 2\Sigma_{xx}\hat{\mathbf{A}} = \mathbf{0}$

$$\begin{aligned}
\hat{\mathbf{A}} &= \Sigma_{xx}^{-1}\Sigma_{x\theta} \\
\hat{\boldsymbol{\theta}}_{\text{LMMSE}} &= \Sigma_{\theta x}\Sigma_{xx}^{-1}\mathbf{x}
\end{aligned}$$

- $\Sigma_{\theta x}\Sigma_{xx}^{-1}$ is often called the Wiener Filter.

LMMSE Estimator

- If $\mathbb{E}[\boldsymbol{\theta}] = \boldsymbol{\mu}_\theta$, $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}_x$.
- Let the estimator be in the form $\hat{\boldsymbol{\theta}} = A^T(\mathbf{x} - \boldsymbol{\mu}_x) + \mathbf{d}$.
- Then,

$$\hat{\boldsymbol{\theta}}_{\mathsf{LMMSE}} = \boldsymbol{\mu}_\theta + \Sigma_{\theta x}\Sigma_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$$

- It is identical to the Gaussian MMSE estimator.

$$\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}} = \boldsymbol{\mu}_\theta + \Sigma_{\theta x}\Sigma_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$$

- LMMSE is unbiased.

Error Covariance Matrix and BMSE

- Error Covariance Matrix

$$\begin{aligned}
\boldsymbol{\Sigma}_e &= \mathbb{E}[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T] \\
&= \mathbb{E}[(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) - \Sigma_{\theta x}\Sigma_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)][(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) - \Sigma_{\theta x}\Sigma_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)]^T \\
&= \Sigma_\theta - \Sigma_{\theta x}\Sigma_{xx}^{-1}\Sigma_{x\theta} - \Sigma_{\theta x}\Sigma_{xx}^{-1}\Sigma_{x\theta} + \Sigma_{\theta x}\Sigma_{xx}^{-1}\Sigma_{xx}\Sigma_{xx}^{-1}\Sigma_{x\theta} \\
&= \Sigma_\theta - \Sigma_{\theta x}\Sigma_{xx}^{-1}\Sigma_{x\theta}
\end{aligned}$$

- BMSE

$$\begin{aligned}
\text{BMSE}(\hat{\mathbf{A}}) &= \mathbb{E}[\|\boldsymbol{\theta} - \Sigma_{\theta x}\Sigma_{xx}^{-1}\mathbf{x}\|^2] \\
&= \text{tr}(\Sigma_{\theta\theta} - \Sigma_{\theta x}\Sigma_{xx}^{-1}\Sigma_{x\theta})
\end{aligned}$$

- This is identical to the BMSE for Gaussian MMSE estimator as well.
- In general (non-Gaussian case), LMMSE estimator gives a higher BMSE than MMSE estimator.

Example

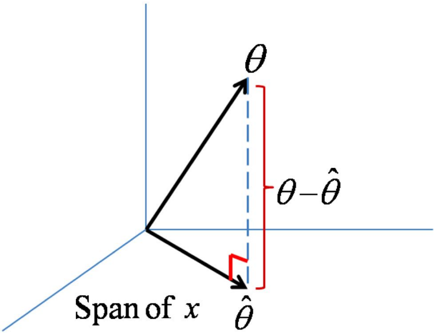
Let $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \sigma_\theta^2\mathbf{I})$, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ are independent. Find the LMMSE of $\boldsymbol{\theta}$.

Orthogonality Principle

Let $\hat{\boldsymbol{\theta}} = \Sigma_{\theta x}\Sigma_{xx}^{-1}\mathbf{x}$ be the LMMSE estimator. Then

$$\begin{aligned}\mathbb{E}[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\mathbf{x}^T] &= \Sigma_{\theta x} - \Sigma_{\theta x}\Sigma_{xx}^{-1}\Sigma_{xx} \\ &= \mathbf{0}\end{aligned}$$

In other words, the error $\mathbf{e} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ is orthogonal to the data $\mathbf{x}$.
Each $\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i$ is orthogonal to every component in $\mathbf{x}$.



Another Way to Derive LMMSE Estimator

- LMMSE Estimator can be derived based on the orthogonality principle.
- Consider any linear estimator of the form $\hat{\boldsymbol{\theta}} = \mathbf{B}^T\mathbf{x}$. If we impose the orthogonality condition

$$\mathbf{0} = \mathbb{E}[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\mathbf{x}^T] = \Sigma_{\theta x} - \mathbf{B}^T\Sigma_{xx}$$

- Then, $\mathbf{B}^T = \Sigma_{\theta x}\Sigma_{xx}^{-1} = \mathbf{A}^T$
- Error Covariance Matrix

$$\begin{aligned}\Sigma_e &= \mathbb{E}[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T] = \mathbb{E}[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\boldsymbol{\theta}^T] + \mathbb{E}[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\mathbf{x}^T\mathbf{A}^T] \\ &= \mathbb{E}[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\boldsymbol{\theta}^T] = \Sigma_\theta - \mathbf{A}^T\Sigma_{x\theta} \\ &= \Sigma_\theta - \Sigma_{\theta x}\Sigma_{xx}^{-1}\Sigma_{x\theta}\end{aligned}$$

Example

Suppose we model our detected signal as $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, where $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\theta} \in \mathbb{R}^k$, $\mathbf{H}_{n\times k}$ is a known linear transformation, and $\mathbf{w}$ is a noise process. Furthermore assume that $\mathbb{E}[\mathbf{w}] = \mathbf{0}$, $\mathbb{E}[\mathbf{w}\mathbf{w}^T] = \sigma_w^2 \mathbf{I}_{n\times n}$, $\mathbb{E}[\boldsymbol{\theta}] = \mathbf{0}$, $\mathbb{E}[\boldsymbol{\theta}\boldsymbol{\theta}^T] = \sigma_\theta^2 \mathbf{I}_{k\times k}$. In addition, assume we know that the parameter and the noise process are uncorrelated, i.e., $\mathbb{E}[\boldsymbol{\theta}\mathbf{w}^T] = \mathbb{E}[\mathbf{w}\boldsymbol{\theta}^T] = 0$. What is the LMMSE estimator?