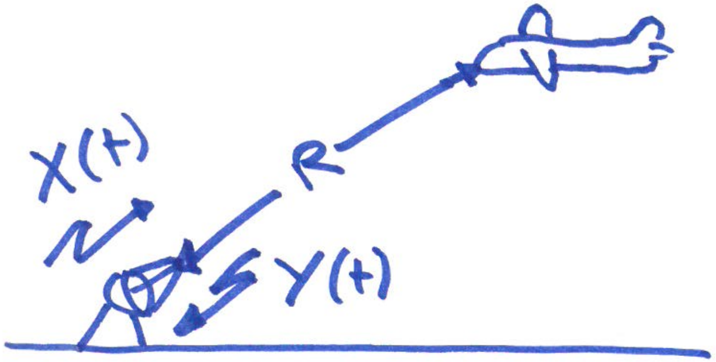
Estimation Problem

Given a set of data $\underline{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$, estimate the value of an unknown quantity θ.

Definition: an estimator is a function $\hat{\theta} = g(\underline{X})$ used to guess the value of an unknown entity θ.

Example: Range estimation in radar

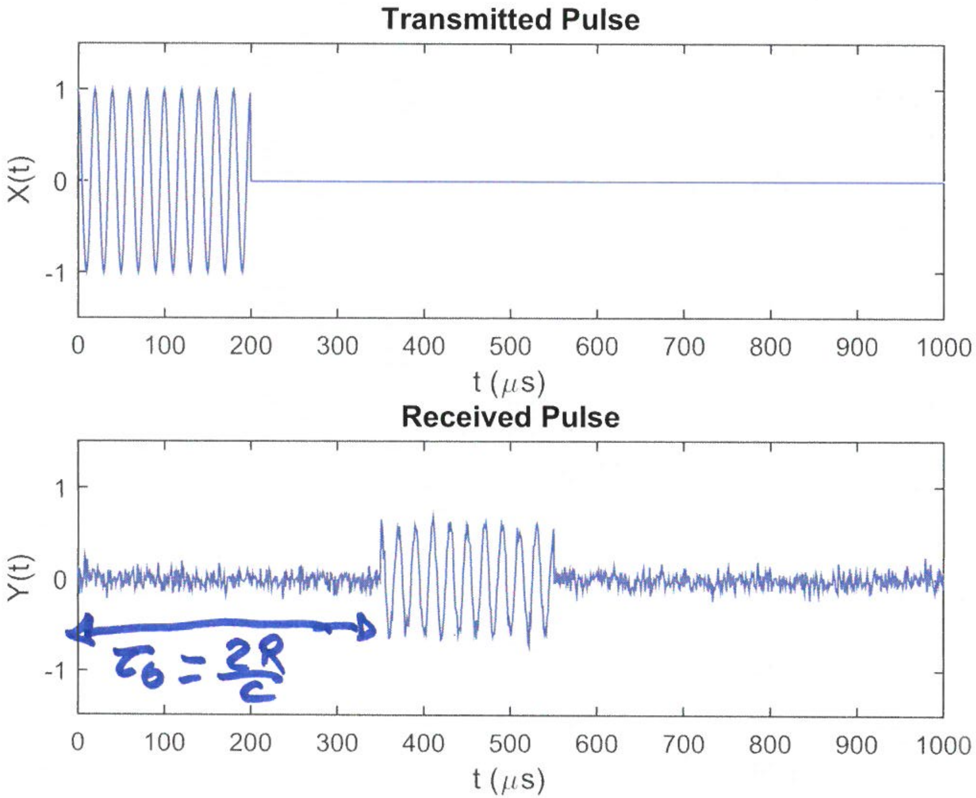


Radar transmits $X(t)$

$Y(t)$ will reach the radar with a delay $\tau_0 = \frac{2R}{c}$.

$$Y(t) = a \cdot X(t - \tau_0) + W(t)$$

$$Y_n = Y(nT_s) = a \cdot X(nT_s - \tau_0) + W(nT_s)$$
$$= a \cdot X\left(nT_s - \frac{2R}{c}\right) + W(nT_s)$$



Unknown: R (range)

Data: Y_n, $n = 1, 2, \ldots, N$

Estimator: $\hat{R} = g(Y_1, Y_2, \ldots, Y_N)$

Estimation error, bias, and mean squared error (MSE)

- Estimation error: $\hat{\theta} - \theta$

- Bias (mean error): $E[\hat{\theta} - \theta]$

  $E[\hat{\theta} - \theta] = 0 \Rightarrow E[\hat{\theta}] = E[\theta] \Rightarrow \hat{\theta}$ is unbiased.

- Mean Squared Error (MSE): $E[(\theta - \hat{\theta})^2]$

Example: Estimation of DC voltage from noisy measurements

- We measure an unknown DC voltage with a noisy voltmeter.
  The measurements are modeled as:

$$X_n = \theta + W_n \,, \quad n = 1, 2, \ldots, N.$$

where

θ: unknown DC voltage

$\{W_n\} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$: noise samples

- Estimator 1: $\tilde{\theta} = X_1.$

- Estimator 2: $\hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} X_i$

$$X_n = \theta + W_n, \quad n = 1, 2, .., N$$
$$\{W_n\} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

- Estimator 1: $\tilde{\theta} = X_1$

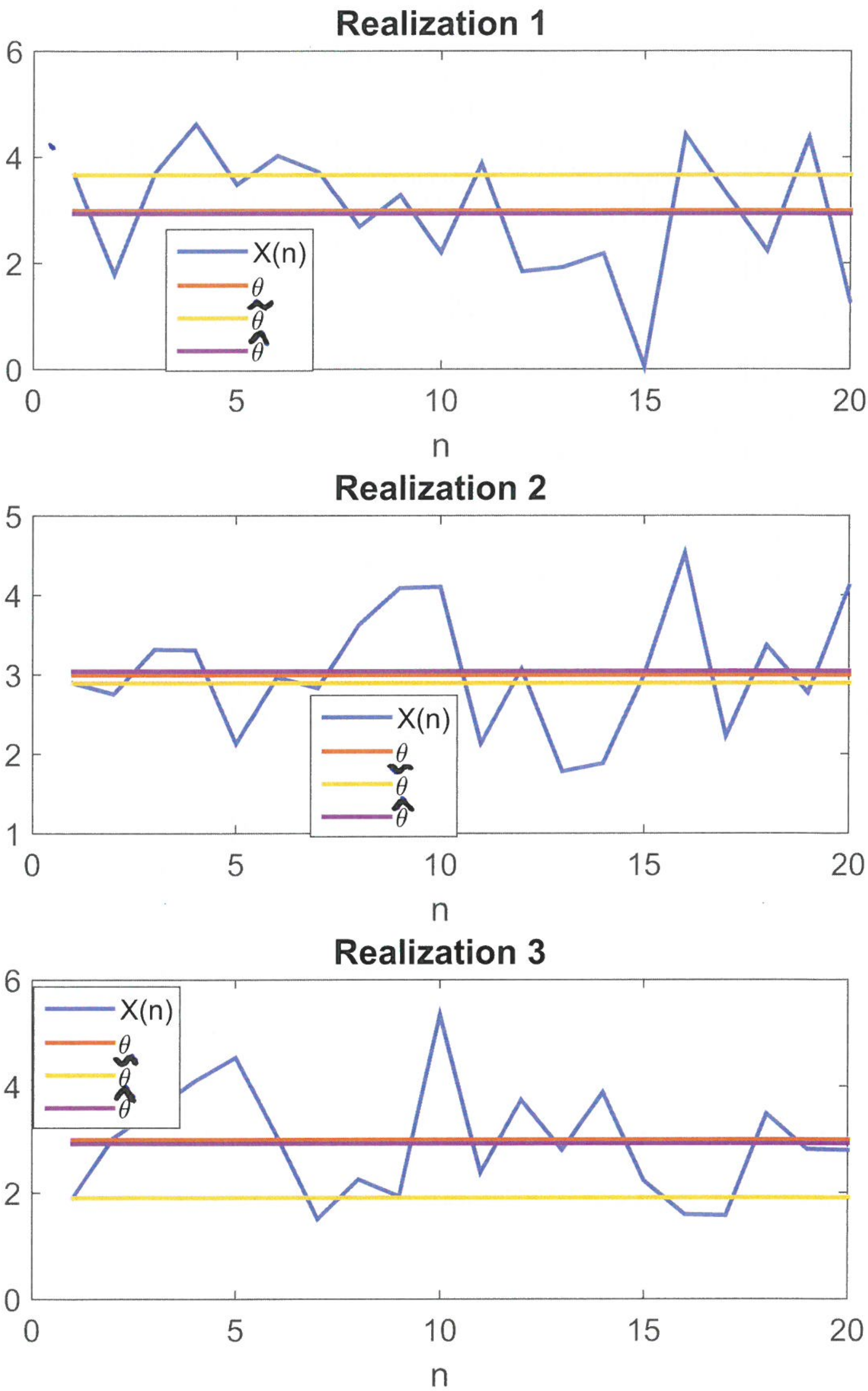
Bias: $E[\tilde{\theta} - \theta] = E[X_1 - \theta] = E[\theta + W_1 - \theta]$
$= E[W_1] = 0.$

MSE: $E[(\tilde{\theta} - \theta)^2] = E[(X_1 - \theta)^2] = E[(\cancel{\theta} + W_1 - \cancel{\theta})^2] = E[W_1^2] = \sigma^2$

- Estimator 2: $\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} X_i$

Bias: $E[\hat{\theta} - \theta] = E\left[\frac{1}{N}\sum_i X_i - \theta\right] = E\left[\frac{1}{N}\sum_i(\theta + W_i) - \theta\right]$
$= E\left[\frac{1}{N}\underbrace{\sum_i \theta}_{N\theta} + \frac{1}{N}\sum_i W_i - \theta\right] = E\left[\theta - \theta + \frac{1}{N}\sum_i W_i\right]$
$= \frac{1}{N}\sum_i E[W_i] = 0.$

MSE: $E[(\hat{\theta} - \theta)^2] = E\left[\left(\frac{1}{N}\sum X_i - \theta\right)^2\right] = E\left[\left(\frac{1}{N}\sum_i(\theta + W_i) - \theta\right)^2\right]$
$= E\left[\left(\frac{1}{N}\sum_i W_i + \cancel{\theta} - \cancel{\theta}\right)^2\right] = E\left[\left(\frac{1}{N}\sum W_i\right)^2\right] = \frac{1}{N^2} E\left[\left(\sum W_i\right)^2\right]$
$= \frac{1}{N^2}\mathrm{Var}\left(\sum_i W_i\right) = \frac{1}{N^2}\sum_i \mathrm{Var}(W_i) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$



Realization 1
X(n)
θ
$\tilde{\theta}$
$\hat{\theta}$
n
Realization 2
X(n)
θ
$\tilde{\theta}$
$\hat{\theta}$
n
Realization 3
X(n)
θ
$\tilde{\theta}$
$\hat{\theta}$
n

The Minimum MSE (MMSE) Estimator

- The MMSE is a Bayesian estimator:
  - θ is a random variable with pdf $p(\theta)$.
  - θ and X have joint pdf $\underbrace{p(\theta, X)}$.

<u>Definition</u>: The MMSE estimator of θ is the function $\hat{\theta} = g(X)$ that minimizes the MSE $E[(\theta - \hat{\theta})^2]$.

$$\boxed{\hat{\theta} = E[\theta | X]}$$

$$E[\theta | X] = \int \theta \underbrace{p(\theta | X)} d\theta = g(X)$$

Properties of the MMSE estimator

- The MMSE estimator is unbiased.

$$E[\hat{\theta}-\theta] = E\big[E[\theta|X]-\theta\big] = \underbrace{E\big[E[\theta|X]\big]}_{E[\theta]} - E[\theta] = 0.$$

- It has the lowest MSE among <u>all</u> estimators.

- It fulfills the "orthogonality principle": $E\big[(\theta-\hat{\theta})\cdot h(X)\big]=0.$

$$E\Big[E\big[(\theta-\hat{\theta})h(X)\,|X\big]\Big] = E\Big[\underbrace{E\big[(\theta-\hat{\theta})|X\big]}_{=0}\cdot h(X)\Big] = 0.$$

$$E\big[(\theta-\hat{\theta})\cdot X\big] = 0.$$

The Linear MMSE (LMMSE) Estimator

Definition: The LMMSE is the estimator of the form

$$\boxed{\hat{\theta} = h_0 + \sum_{n=1}^{N} h_n X_n = h_0 + h^T X}$$

with coefficients $h_0, h_1, \ldots, h_N$ that minimize $E[(\theta - \hat{\theta})^2]$.

The coefficients that minimize the MSE are:

$$\boxed{h_0 = E[\theta] - h^T E[X]}, \quad \boxed{h = \underline{\underline{C}}_{XX}^{-1} C_{X\theta}}$$

$$\hat{\theta} = E[\theta] - C_{\theta X} C_{XX}^{-1} E[X] + C_{\theta X} C_{XX}^{-1} X$$

$$= E[\theta] + C_{\theta X} C_{XX}^{-1} (X - E[X])$$

Properties of the LMMSE Estimator

$\boxed{\hat{\theta} = h_0 + h^T X}$ $\boxed{h_0 = E[\theta] - h^T E[X]}$ $\boxed{h = C_{XX}^{-1} C_{X\theta}}$

- Only first- and second-order statistical modeling of the problem is needed.
- Computational complexity is fixed (and low): N products and additions.
- The LMMSE estimator is unbiased.

$$E[\hat{\theta}] = E[h_0] + E[h^T X] = E\big[E[\theta] - h^T E[X]\big] + h^T E[X] = E[\theta] - \cancel{h^T E[X]} + \cancel{h^T E[X]} = E[\theta]$$

- It minimizes the MSE among all affine estimators.
- It fulfills the orthogonality principle for affine functions of the data X.

$$f(X) = f_0 + f^T X \Rightarrow E\big[(\theta - \hat{\theta})\, f(X)\big] = 0$$

- Its MSE is known upfront:

$$\boxed{E[(\theta - \hat{\theta})^2] = \mathrm{Var}(\theta) - C_{\theta X} C_{XX}^{-1} C_{X\theta}}$$

LMMSE for Multiple Variables ("Vector LMMSE")

- We estimate $\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_K \end{bmatrix}$ from $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$ with an estimator $\hat{\theta}$ of affine form.

- The objective is to minimize the total MSE:

$$MSE_{\hat{\theta}} = \sum_{k=1}^{K} MSE_{\hat{\theta}_k} = \sum_{k=1}^{K} E\left[(\theta_k - \hat{\theta}_k)^2\right]$$

- The optimal estimator is then: $\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_K \end{bmatrix} = E[\theta] + C_{\theta X} C_{XX}^{-1} (X - E[X])$

- $MSE_{\hat{\theta}} = \sum_{k=1}^{K} MSE_{\hat{\theta}_k} = \operatorname{trace}\left\{ C_{\theta\theta} - C_{\theta X} C_{XX}^{-1} C_{X\theta} \right\}$

Conditional Expectation: $E[\theta|x] = \int \theta \, \underbrace{p(\theta|x)} \, d\theta$

$$p(\theta|x) = \frac{p(\theta, x)}{p(x)} = \frac{p(\theta, x)}{\int p(\theta, x)\, d\theta}$$

$$E[h(x,\theta)] = E\big[E[h(x,\theta)|x]\big] = E\big[E[h(x,\theta)|\theta]\big]$$

$$E[h(x,\theta)] = \iint h(x,\theta) \underbrace{p(x,\theta)}_{p(\theta|x)\cdot p(x) = p(x|\theta)\cdot p(\theta)} dx\, d\theta$$

$$= \iint h(x,\theta)\, p(\theta|x)\, p(x)\, d\theta\, dx = \int \underbrace{\left(\int h(x,\theta)\, p(\theta|x)\, d\theta\right)}_{E[h(x,\theta)|x]} p(x)\, dx$$

$$E\big[E[h(x,\theta)|x]\big]$$

Derivation of the MMSE Estimator

- Find $\hat{\theta} = g(X)$ such that $E[(\theta-\hat{\theta})^2]$ is minimized.

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

$\hat{\theta}$ that minimizes $E[(\theta-\hat{\theta})^2|X]$ for all $X \Rightarrow$ It will minimize $E[(\theta-\hat{\theta})^2]$.

$$\frac{\partial E[(\theta-\hat{\theta})^2|X]}{\partial\hat{\theta}} = E\left[\frac{\partial(\theta-\hat{\theta})^2}{\partial\hat{\theta}}\Big|X\right] = E[2(\theta-\hat{\theta})(-1)|X]$$

$$= -2E[(\theta-\hat{\theta})|X] = 0 \Rightarrow \boxed{E[(\theta-\hat{\theta})|X] = 0}$$

$$E[(\theta)|X] - \underbrace{E[\hat{\theta}|X]}_{\hat{\theta}} = 0 \rightarrow \boxed{\hat{\theta} = E[\theta|X]}$$

$$\boxed{E[\theta - \hat{\theta} \mid X] = 0}$$

- Bias of the MMSE

$$E[(\hat{\theta} - \theta)] = E\Big[\underbrace{E[(\hat{\theta} - \theta) \mid X]}_{=0}\Big] = 0.$$

- Orthogonality Principle

$$E[(\theta - \hat{\theta}) \cdot h(X)] = E\Big[E[(\theta - \hat{\theta}) h(X) \mid X]\Big]$$

$$= E\Big[\underbrace{E[(\theta - \hat{\theta}) \mid X]}_{\shortparallel\ 0} \cdot h(X)\Big] = 0. //$$

Derivation of the LMMSE

- Find $\hat{\theta} = h_0 + h^T X$ such that $E[(\theta-\hat{\theta})^2]$ is minimized.

$\Rightarrow$ Find h_0, h^T such that $E[(\theta-\hat{\theta})^2]$ is minimized.

$$\frac{\partial E[(\theta-\hat{\theta})^2]}{\partial h_0} = 0 \, ; \quad \frac{\partial E[(\theta-\hat{\theta})^2]}{\partial h^T} = 0 \, ;$$

Derivation of h_0

$$\frac{\partial E\left[E[(\theta-\hat{\theta})^2 | X]\right]}{\partial h_0} = E\left[\frac{\partial E[(\theta-\hat{\theta})^2 | X]}{\partial h_0}\right]$$

$$\frac{\partial E[(\theta-\hat{\theta})^2 | X]}{\partial h_0} = \underbrace{\frac{\partial E[(\theta-\hat{\theta})^2 | X]}{\partial \hat{\theta}}}_{-2E[(\theta-\hat{\theta})|X]} \underbrace{\frac{\partial \hat{\theta}}{\partial h_0}}_{1} = -2E[(\theta-\hat{\theta})|X]$$

$$E\left[-2E[(\theta)|X] - 2\hat{\theta}\right] = -2E[\theta] - 2E[h_0 + h^T X] = 0 \Rightarrow h_0 = E[\theta] - h^T E[$$

$$\hat{\theta} = h_0 + h^T X = E[\theta] - h^T E[X] + h' X = \underline{E[\theta] + h'(X - E[X])}$$

- Derivation of h^T

$$\frac{\partial E[(\theta-\hat{\theta})^2 | X]}{\partial h^T} = \underbrace{\frac{\partial [(\theta-\hat{\theta})^2 | X]}{\partial \hat{\theta}}}_{-2E[(\theta-\hat{\theta})|X]} \underbrace{\frac{\partial \hat{\theta}}{\partial h^T}}_{X} = -2E[(\theta-\hat{\theta}) \cdot X | X]$$

$$E[-\cancel{2}E[(\theta-\hat{\theta})|X]] = 0 \Rightarrow \boxed{E[(\theta-\hat{\theta}) \cdot X] = 0}$$

$$E[\underbrace{(\theta - E[\theta] - h^T(X - E[X]))}_{Y \rightarrow Y \text{ is zero mean}} \cdot X] = E[XY^T] = Cov(X, Y) + \cancel{E[X]E[Y]^T}$$

$$= E[\underbrace{(X - E[X])(\theta - E[\theta])^T}_{C_{X\theta}}] - \underbrace{E[(X - E[X])(X - E[X])^T](h^T)^T}_{C_{XX} h} = 0$$

$$C_{X\theta} - C_{XX} h = 0 \Rightarrow \boxed{h = C_{XX}^{-1} C_{X\theta}}$$

Orthogonality principle for the LMMSE Estimator

- Recall from the derivation of h^T: $E[(\theta-\hat{\theta})\cdot X] = 0$

Orthogonality principle: $E[(\theta-\hat{\theta})\cdot f(X)] = 0$ for $\boxed{f(X) = f_0 + f^T X}$

$$E[(\theta-\hat{\theta})\cdot(f_0 + f^T X)] = \underbrace{E[(\theta-\hat{\theta}) f_0]}_{0 \,\Rightarrow\, \hat{\theta} \text{ is unbiased}} + E[(\theta-\hat{\theta}) f^T X]$$

$$= f^T \underbrace{E[(\theta-\hat{\theta})\cdot X]}_{0} = 0$$