Week 7 : Session 2

9. LMMSE Estimation

LMMSE stands for

- Linear MMSE
- Linear Minimum Mean Square Error.

Recall, MMSE is given as

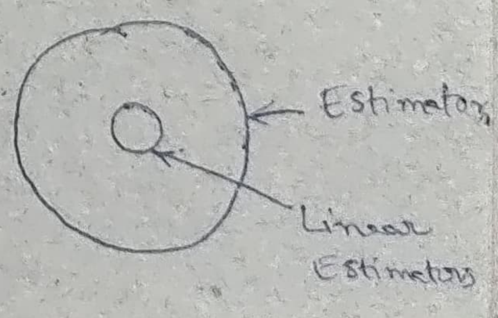
$$\min E\left\{ \| \hat{h} - \bar{h} \|^2 \right\} \longleftarrow \text{Cost function for MMSE estimate}$$

Conditional Expectation

$$\hat{h} = E\left\{ \bar{h} \mid \bar{y} \right\} \longleftarrow \text{General Expression for MMSE estimate}$$

However, this is extremely challenging to determine. (i) First we have to derive the conditional PDF $f_{\bar{h}|\bar{y}}(\bar{h} \mid \bar{y})$ and then take the expected value of the same, which is fairly difficult to evaluate, especially when $\bar{h}, \bar{y}$ are NOT Gaussian. (ii) $\bar{h}, \bar{y}$ is arbitrarily distributed. Hence we settle for the best LINEAR estimator. (ignoring the non-linear estimators)

LMMSE $\rightarrow$ Best Linear Estimator that has minimum Mean Square Error!



Thus, the general expression for Linear Estimator is given as,

$$\hat{h} = \underbrace{C\,\bar{y}}_{\text{Linear Transformation}}$$

$\hat{h}$ — $M \times 1$

C — $M \times N$ matrix.

$\bar{y}$ — $N \times 1$

Which is easier to determine but suboptimal.

Now, how to determine C that yields the lowest MSE?

LMMSE

The cost function to optimize is

Estimator is constrained to be LINEAR.

$$\underbrace{\min \; \underbrace{E\left\{ \underbrace{\| \underbrace{C\bar{y}}_{\text{Linear}} - \bar{h} \|^2}_{\text{square Error}} \right\}}_{\text{Mean Square Error}}}_{\text{Linear Minimum Mean Square Error}}$$

This can be simplified as follows:

$$\| C\bar{y} - \bar{h} \|^2 = (C\bar{y} - \bar{h})^T (C\bar{y} - \bar{h})$$

$$= Tr\left\{ (C\bar{y} - \bar{h})^T (C\bar{y} - \bar{h}) \right\}$$

$$= Tr\left\{ (C\bar{y} - \bar{h}) (C\bar{y} - \bar{h})^T \right\}$$

$$= Tr\left\{ (C\bar{y} - \bar{h}) (\bar{y}^T C^T - \bar{h}^T) \right\}$$

$$\|\bar{x}\|^2 = \bar{x}^T \bar{x}$$

$$= Tr\left\{ \bar{x}^T \bar{x} \right\}$$

$$= Tr\left\{ \bar{x} \bar{x}^T \right\}$$

$$= Tr\left\{ \begin{bmatrix} x_1^2 & & \\ & x_i^2 & \\ & & x_n^2 \end{bmatrix} \right\}$$

$$= x_1^2 + x_2^2 + \dots + x_n^2$$

$$= Tr\left\{ C\bar{y}\bar{y}^T C^T - \bar{h}\bar{y}^T C^T - C\bar{y}\bar{h}^T + \bar{h}\bar{h}^T \right\}$$

$$E\left\{ \| C\bar{y} - \bar{h} \|^2 \right\} = E\left\{ Tr\left\{ C\bar{y}\bar{y}^T C^T - \bar{h}\bar{y}^T C^T - C\bar{y}\bar{h}^T + \bar{h}\bar{h}^T \right\} \right\}$$

Interchanging Tr and E,

$$= Tr\left\{ E\left\{ C\bar{y}\bar{y}^T C^T - \bar{h}\bar{y}^T C^T - C\bar{y}\bar{h}^T + \bar{h}\bar{h}^T \right\} \right\}$$

$$= Tr\left\{ C\,E[\bar{y}\bar{y}^T]C^T - E[\bar{h}\bar{y}^T]C^T - C\,E\{\bar{y}\bar{h}^T\} + E\{\bar{h}\bar{h}^T\} \right\}$$

$$= Tr\left\{ C R_{yy} C^T - R_{hy} C^T - C R_{yh} + R_{hh} \right\}$$

This is the Mean square Error of the Linear Estimator $\hat{h} = C\bar{y}$. Find the matrix C that minimizes this.

Further simplification,

$$MSE = Tr\left\{ \underbrace{(C R_{yy} - R_{hy}) R_{yy}^{-1} (C R_{yy} - R_{hy})^T}_{\text{① Function of C}} + \underbrace{R_{hh} - R_{hy} R_{yy}^{-1} R_{yh}}_{\text{② Independent of C}} \right\}$$

How?

① ⇒ R_{yy} is Positive Semidefinite (PSD).

⇒ $\bar{x}^T R_{yy}^{-1} \bar{x} \geq 0$

⇒ Trace of the first term ① is always ≥ 0 with minimum of 0.

Therefore, Minimization reduces to

$$\min\ Tr\left\{ (C R_{yy} - R_{hy}) R_{yy}^{-1} (C R_{yy} - R_{hy})^T + R_{hh} - R_{hy} R_{yy}^{-1} R_{yh} \right\}$$

$$\Rightarrow \underbrace{\min\ Tr\left\{ (C R_{yy} - R_{hy}) R_{yy}^{-1} (C R_{yy} - R_{hy})^T \right\}}_{\text{depends on C, min. value = 0}} + \underbrace{Tr\left\{ R_{hh} - R_{hy} R_{yy}^{-1} R_{yh} \right\}}_{\text{does not depends on C}}$$

So, to achieve minimum MSE, set the first term to zero.

$\bar{x}^T R_{yy}^{-1} \bar{x} = 0$, only if $\bar{x} = 0$, coz minimum value $= 0$

$$MSE = Tr\left\{ \underbrace{\left(C R_{yy} - R_{xy}\right) R_{yy}^{-1} \left(C R_{yy} - R_{xy}\right)^T}_{\text{Function of } C} + \underbrace{R_{xx} - R_{xy} R_{yy}^{-1} R_{yx}}_{\text{Independent of } C} \right\}$$

$$= Tr\left\{ \left(C \cancel{R_{yy} R_{yy}^{-1}} - R_{xy} R_{yy}^{-1}\right)\left(R_{yy}^T C^T - R_{xy}^T\right) + R_{xx} - R_{xy} R_{yy}^{-1} R_{yx} \right\}$$

$R_{yy} R_{yy}^{-1} = I$

$(AB)^T = B^T A^T$

$(A+B)^T = A^T + B^T$

$$= Tr\left\{ \left(C - R_{xy} R_{yy}^{-1}\right)\left(R_{yy}^T C^T - R_{xy}^T\right) + R_{xx} - R_{xy} R_{yy}^{-1} R_{yx} \right\}$$

$$= Tr\left\{ C R_{yy}^T C^T - R_{xy} \cancel{R_{yy}^{-1} R_{yy}^T} C^T - C R_{xy}^T + \cancel{R_{xy} R_{yy}^{-1} R_{xy}^T} + R_{xx} - \cancel{R_{xy} R_{yy}^{-1} R_{yx}} \right\}$$

$R_{yy}^T = R_{yy}$

$R_{xy}^T = R_{yx}$

$$= Tr\left\{ C R_{yy}^T C^T - R_{xy} C^T - C R_{yx} + R_{xx} \right\}$$

Therefore, minimum occurs for

$$C R_{yy} - R_{hy} = 0$$

$$\Rightarrow C R_{yy} = R_{hy}$$

$$\Rightarrow \boxed{C = R_{hy} R_{yy}^{-1}}$$

This is the C for which minimum MSE is achieved.

Therefore, the LMMSE estimate is

$$\hat{h} = \underset{M \times N}{C} \bar{y}$$

$$\boxed{\hat{h} = R_{hy} R_{yy}^{-1} \bar{y}}$$

The corresponding MSE is given by

$$E\left\{ \| \hat{h} - \bar{h} \|^2 \right\} = Tr\left\{ R_{hh} - R_{hy} R_{yy}^{-1} R_{yh} \right\}$$

We have, LMMSE estimate

$$\boxed{\hat{h} = C\bar{y} = R_{hy} R_{yy}^{-1} \bar{y}}$$

where, $\bar{h}, \bar{y}$ are arbitrarily distributed

This seems exactly similar to MMSE !!! when $\bar{h}, \bar{y}$ are Gaussian.

In other words, "MMSE and LMMSE are identical when $\bar{h}, \bar{y}$ are Gaussian".

- MMSE – The best amongst all the Estimators
- LMMSE – The best amongst the Linear Estimators.

TO Summarize ...

$\hat{h}(\bar{y}) = R_{hy} R_{yy}^{-1} \bar{y}$	MMSE	LMMSE
$\bar{y}, \bar{h}$ are Jointly Gaussian	YES	YES
$\bar{y}, \bar{h}$ are Arbitrary PDF	NO	YES

Consider now the MISO channel estimation model.

$$\underset{N \times 1}{\bar{y}} = \underset{N \times M}{X} \underset{M \times 1}{\bar{h}} + \underset{N \times 1}{\bar{v}}$$

where, there are M Transmit Antennas and 1 Receive Antennas.

The LMMSE estimate is given as

$$\hat{h} = R_{hy} R_{yy}^{-1} \bar{y} = \left(X^T X + \frac{1}{SNR} I\right)^{-1} X^T \bar{y}.$$

Note that, this is valid even when $\bar{h}, \bar{y}$ are NOT jointly Gaussian, but zero mean. (i.e.) $E[\bar{h}] = E\{\bar{y}\} = 0$.

The Error covariance of LMMSE is given as

$$R_{hh} - R_{hy} R_{yy}^{-1} R_{yh} = \sigma^2 \left(X^T X + \frac{1}{SNR} I\right)^{-1}$$

The MSE is given as

$$MSE = Tr\left\{\text{Error covariance}\right\}$$

$$= \sigma^2 \, Tr\left\{\left(X^T X + \frac{1}{SNR} I\right)^{-1}\right\}$$

Week 7 : Session 3.

LMMSE Interpretation

LMMSE Interpretation is a very intuitive explanation of the LMMSE. Let us now take a deeper look at LMMSE.

The cost function to optimize is

$$\min \; E\left\{ \| C\bar{y} - \bar{h} \|^2 \right\}$$

LMMSE estimate is

$$\hat{h} = C\bar{y} = R_{hy} R_{yy}^{-1} \bar{y}$$

where, $\bar{h}, \bar{y}$ are arbitrarily distributed and zero mean. $E\{\bar{h}\} = E\{\bar{y}\} = 0$.

We first derive the LMMSE estimate for non-zero mean parameter / observation.

(i) $E\{\bar{h}\} = \bar{\mu}_h \leftarrow$ M×1 vector

$E\{\bar{y}\} = \bar{\mu}_y \leftarrow$ N×1 vector

Now, we make the zero-mean quantities.

(ii) $E\{\bar{h} - \bar{\mu}_h\} = 0$

$E\{\bar{y} - \bar{\mu}_y\} = 0$

Therefore, the LMMSE estimate for non-zero mean is given as follows.

$$\hat{h} - \bar{\mu}_h = R_{hy} R_{yy}^{-1} (\bar{y} - \bar{\mu}_y)$$

$$\Rightarrow \underbrace{\hat{h} = R_{hy} R_{yy}^{-1} (\bar{y} - \bar{\mu}_y) + \bar{\mu}_h}_{\text{LMMSE estimate for arbitrarily distributed } \bar{h}, \bar{y} \text{ with non-zero mean}}$$

$\bar{\mu}_y, \bar{\mu}_h$ are Prior information

For LMMSE Non-Zero mean, note that

The covariance Matrix is

$$R_{yy} = E\left\{(\bar{y}-\bar{\mu}_y)(\bar{y}-\bar{\mu}_y)^T\right\}$$

The Cross covariance Matrix is

$$R_{hy} = E\left\{(\bar{h}-\bar{\mu}_h)(\bar{y}-\bar{\mu}_y)^T\right\}$$

Therefore, the LMMSE for the Linear MISO channel estimation model $\bar{y} = X\bar{h} + \bar{v}$, with non-zero mean is

$$\underbrace{\hat{h} = \left(X^TX + \frac{1}{SNR}I\right)^{-1} X^T(\bar{y}-\bar{\mu}_y) + \bar{\mu}_h}_{\text{(aka) Linear Parameter estimation model}}$$

Further, $\bar{\mu}_y$ is derived as follows.

We have, $\bar{y} = X\bar{h} + \bar{v}$

Since noise is zero-mean, (i.e) $E\{\bar{v}\} = 0$,

$$\bar{\mu}_y = E\{\bar{y}\} = E\{X\bar{h} + \bar{v}\}$$

$$= E\{X\bar{h}\} + \cancelto{0}{E\{\bar{v}\}}$$

$$= X\bar{\mu}_h$$

Therefore,

$$\boxed{\hat{h} = \left(X^TX + \frac{1}{SNR}I\right)^{-1} X^T(\bar{y} - X\bar{\mu}_h) + \bar{\mu}_h}$$

For simplicity, let us consider the SISO Model.

$$y(1) = h\,x(1) + v(1)$$
$$y(2) = h\,x(2) + v(2)$$
$$\vdots$$
$$y(N) = h\,x(N) + v(N)$$

The different quantities are

$$\bar{x} = \begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(N) \end{bmatrix}, \quad \bar{y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix}$$

Pilot vector   Output vector

Therefore,

$$\bar{y} = \bar{x}\,h + \bar{v}$$

Since noise is zero mean, (i) $E\{\bar{v}\} = 0$

Mean of the Fading channel coefficient, (ii) $E\{h\} = \mu_h$

Mean of $\bar{y}$,

$$\bar{\mu}_y = E\{\bar{y}\} = E\{\bar{x}h + \bar{v}\} = E\{\bar{x}h\} + \cancel{E\{\bar{v}\}}^{0}$$
$$= \bar{x}\,E\{h\}$$
$$= \bar{x}\,\mu_h$$

$$\Rightarrow \bar{\mu}_y = \bar{x}\,\mu_h$$

Therefore, the LMMSE for the linear SISO channel estimation model is

M×1 Identity matrix (M = 1)

$$\hat{h} = \left(\bar{x}^T\bar{x} + \frac{1}{SNR}\mathbb{1}\right)^{-1}\bar{x}^T\left(\bar{y} - \bar{x}\mu_h\right) + \mu_h$$

$$= \frac{\bar{x}^T\left(\bar{y} - \bar{x}\mu_h\right)}{\|\bar{x}\|^2 + \frac{\sigma^2}{\sigma_h^2}} + \mu_h$$

$$\Rightarrow \hat{h} = \frac{\frac{\bar{x}^T}{\sigma^2}\left(\bar{y} - \bar{x}\mu_h\right)}{\frac{\|\bar{x}\|^2}{\sigma^2} + \frac{1}{\sigma_h^2}} + \mu_h$$

$$= \frac{\frac{\bar{x}^T\bar{y}}{\sigma^2} - \frac{\bar{x}^T\bar{x}}{\sigma^2}\mu_h}{\frac{\|\bar{x}\|^2}{\sigma^2} + \frac{1}{\sigma_h^2}} + \mu_h$$

$$= \frac{\frac{\bar{x}^T\bar{y}}{\sigma^2} - \cancel{\frac{\|\bar{x}\|^2}{\sigma^2}\mu_h} + \cancel{\frac{\|\bar{x}\|^2}{\sigma^2}}\mu_h + \frac{\mu_h}{\sigma_h^2}}{\frac{\|\bar{x}\|^2}{\sigma^2} + \frac{1}{\sigma_h^2}}$$

$$= \frac{\frac{\bar{x}^T\bar{y}}{\sigma^2} + \frac{\mu_h}{\sigma_h^2}}{\frac{\|\bar{x}\|^2}{\sigma^2} + \frac{1}{\sigma_h^2}}$$

ML Estimate

Prior

$$= \frac{\frac{\bar{x}^T\bar{y}/\|\bar{x}\|^2}{\sigma^2/\|\bar{x}\|^2} + \frac{\mu_h}{\sigma_h^2}}{\frac{1}{\sigma^2/\|\bar{x}\|^2} + \frac{1}{\sigma_h^2}}$$

Variance of Prior

MSE of ML Estimate

$$= \frac{\frac{\text{ML Est}}{\text{ML MSE}} + \frac{\text{Prior}}{\text{Prior Var.}}}{\frac{1}{\text{ML MSE}} + \frac{1}{\text{Prior Var.}}}$$

Linear combination of ML / Prior.

$$= \frac{W_1(\text{ML}) + W_2(\text{Prior})}{W_1 + W_2}$$

Weighted Linear combination

weights
$$\begin{cases} W_1 = \dfrac{1}{\sigma^2/\|\bar{x}\|^2} \propto \text{Reliability!} \\[2ex] W_2 = \dfrac{1}{\sigma_h^2} \propto \text{Reliability!} \end{cases}$$

The LMMSE estimate is performing a weighted linear combination of the ML and the Prior, with the weights given by the inverse of the MSE and the variance.

- Weights, $\propto \dfrac{1}{\text{MSE / variance}}$

$\Rightarrow$ Higher MSE / variance, Lower is the Weight

When ML MSE $\rightarrow \infty$

(i) $\sigma^2 \rightarrow \infty$

(very noisy observation)

$$\hat{h} = \frac{\cancelto{0}{\frac{\text{ML Est}}{\text{ML MSE}}} + \frac{\text{Prior}}{\text{Prior var.}}}{\cancelto{0}{\frac{1}{\text{ML MSE}}} + \frac{1}{\text{Prior var}}} = \frac{\frac{\text{Prior}}{\text{Prior var}}}{\frac{1}{\text{Prior var}}}$$

$$= \frac{\frac{\mu_h}{\sigma_h^2}}{\frac{1}{\sigma_h^2}}$$

$$\boxed{\hat{h} = \mu_h}$$ ← Prior information

When Prior var $\rightarrow \infty$

(i) $\sigma_h^2 \rightarrow \infty$

(Prior information is unreliable)

(ii) non-informative prior.

$$\hat{h} = \frac{\frac{\text{ML Est}}{\text{ML MSE}} + \cancelto{0}{\frac{\text{Prior}}{\text{Prior var.}}}}{\frac{1}{\text{ML MSE}} + \cancelto{0}{\frac{1}{\text{Prior var}}}} = \frac{\frac{\text{ML Est}}{\text{ML MSE}}}{\frac{1}{\text{ML MSE}}}$$

$$\boxed{\hat{h} = \frac{\bar{x}^T \bar{y}}{\|\bar{x}\|^2}}$$ ← ML Estimate