MM5. Minimum Mean Squared Error Estimation

Reading page: Chapt 7, pp.377-397

- Explain MM4 exercise
- 5.1 Linear minimum mean squared error estimators
- 5.2 (Nonlinear) minimum mean squared error estimators

What have we talked through MM4?

- 4.1 Binary detection of discrete-time signals
- 4.2 Binary detection of continuous-time signals
- 4.3 M-ary detection

Maximum "a posteriori" (MAP) Rule

- MAP decision rule:

$$f(y\mid H_1)P(H_1) \underset{H_0}{\overset{H_1}{\gtrless}} f(y\mid H_0)P(H_0)$$

- Prior distribution $P(H_i)$, i=1,2

- Likelihood ratio $L(y)$:

$$L(y)=\frac{f(y\mid H_1)}{f(y\mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)}{P(H_1)}$$

$$l(y)=\ln(L(y))$$

$$l(y)=\ln\left(\frac{f(y\mid H_1)}{f(y\mid H_0)}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \ln\left(\frac{P(H_0)}{P(H_1)}\right)$$

Bayes' Decision Rule

- Average cost:

$$\overline{C} = C_{00}P[D=H_0|H_0]P[H_0]+C_{10}P[D=H_1|H_0]P[H_0]+ \\ +C_{01}P[D=H_0|H_1]P[H_1]+C_{11}P[D=H_1|H_1]P[H_1]$$

- Bayes' decision rule: minimize the average cost

$$L(y)=\frac{f(y\mid H_1)}{f(y\mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)(C_{10}-C_{00})}{P(H_1)(C_{01}-C_{11})}$$

4.1 Binary Discrete: Decision Rules

- Decision rule:

$$l(\mathbf{y})=\frac{1}{\sigma^2}\left[\mathbf{y}^T(\mathbf{s_1}-\mathbf{s_0})+\frac{1}{2}\left(\|\mathbf{s_0}\|^2-\|\mathbf{s}_1\|^2\right)\right] \underset{H_0}{\overset{H_1}{\gtrless}} \ln(\gamma)$$

$$\mathbf{y}^T(\mathbf{s_1}-\mathbf{s_0}) \underset{H_0}{\overset{H_1}{\gtrless}} \sigma^2\ln(\gamma)+\frac{1}{2}\left(E_{\mathbf{s}_1}-E_{\mathbf{s}_0}\right) \qquad E_{\mathbf{s}_i}=\|\mathbf{s}_i\|^2=\sum_{n=0}^{N-1}{s_{in}}^2$$

$$\sum_{n=0}^{N-1} y_n(s_{1n}-s_{0n}) \underset{H_0}{\overset{H_1}{\gtrless}} \sigma^2\ln(\gamma)+\frac{1}{2}\left(E_{\mathbf{s}_1}-E_{\mathbf{s}_0}\right)$$

- MAP decision rule:

$$\mathbf{y}^T(\mathbf{s_1}-\mathbf{s_0}) \underset{H_0}{\overset{H_1}{\gtrless}} \sigma^2\ln\left(\frac{P(H_0)}{P(H_1)}\right)+\frac{1}{2}\left(E_{\mathbf{s}_1}-E_{\mathbf{s}_0}\right) \qquad E_{\mathbf{s}_i}=\|\mathbf{s}_i\|^2=\sum_{n=0}^{N-1}{s_{in}}^2$$

4.2 Binary Continuous: Decision Rules

- Time-limited but possibly bandwidth unlimited finite-energy signals
- Decision rule:

$$\int_{-T_s}^{T_e} y(t)(s_1(t)-s_0(t))dt \underset{H_0}{\overset{H_1}{\gtrless}} \frac{N_0}{2}\ln(\gamma)+\frac{1}{2}\left(E_{\mathbf{s}_1}-E_{\mathbf{s}_0}\right) \qquad E_{\mathbf{s}_i}=\|\mathbf{s}_i\|^2=\int_{T_s}^{T_e}(s_i(t))^2dt$$

$s_i(t)$

- MAP decision rule:

$$\int_{T_s}^{T_e} y(t)(s_1(t)-s_0(t))dt \underset{H_0}{\overset{H_1}{\gtrless}} \frac{N_0}{2}\ln\left(\frac{P(H_0)}{P(H_1)}\right)+\frac{1}{2}\left(E_{\mathbf{s}_1}-E_{\mathbf{s}_0}\right) \qquad E_{\mathbf{s}_i}=\|\mathbf{s}_i\|^2=\int_{T_s}^{T_e}(s_i(t))^2dt$$

4.3 MAP for M-ary Decision

- MAP decision rule:

$$select \quad H_i \quad if \quad P(H_i \mid y) \geq P(H_j \mid y) \quad for \quad any \quad j = 0,1,\cdots,M-1$$

or

$$select \quad H_i \quad if \quad \frac{f(y \mid H_i)}{f(y \mid H_j)} \geq \frac{P(H_j)}{P(H_i)} \quad for \quad any \quad j = 0,1,\cdots,M-1$$

- MAP decision rule for time-limited discrete-time signals:

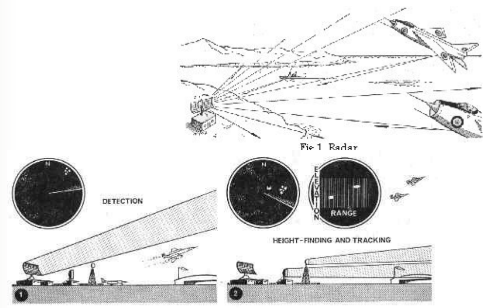
$$select \quad H_i \quad if \quad \mathbf{y}^T\mathbf{s}_i + \sigma^2 \ln(P(H_i)) - \frac{1}{2}E_{\mathbf{s}_i} \geq \mathbf{y}^T\mathbf{s}_j + \sigma^2 \ln(P(H_j)) - \frac{1}{2}E_{\mathbf{s}_j} \quad for \quad any \quad j = 0,1,\cdots,M-1$$

- Further with uniform"a priori" pdf, i.e., $P(H_0)=P(H_1)= \ldots =P(H_{M-1})=1/M$

$$select \quad H_i \quad if \quad \mathbf{y}^T\mathbf{s}_i - \frac{1}{2}E_{\mathbf{s}_i} \geq \mathbf{y}^T\mathbf{s}_j - \frac{1}{2}E_{\mathbf{s}_j} \quad for \quad any \quad j = 0,1,\cdots,M-1$$

Motivation – Signal Estimation

- Target detection and tracking – Radar system



MM5. Minimum Mean Squared Error Estimation

- **5.1 Linear minimum mean squared error estimators**
- 5.2 (Nonlinear) minimum mean squared error estimator

5.1 Linear Minimum Mean Squared Error (LMMSE) Estimators

LMMSE Problem formulation

- A random sequence X(1), ... X(M) whose realization can be observed
- A random variable Y which has to be estimated
- Seek a linear estimator as:

$$\hat{Y} = h_0 + \sum_{m=1}^{M} h_m X(m)$$

- By miminizing the mean squared error(MSE):

$$\min_{h_m, m=0,1,\ldots M} E\{(Y-\hat{Y})^2\}$$

5.1.0 Simple Estimation Cases

- Estimating a random variable with a constant

The mean of the random variable minimizes the MSE

$$E\{(Y-a)^2\} = E\{[(Y-\mu_Y)+(\mu_Y-a)]^2\}$$
$$= \sigma_{YY} + (\mu_Y - a)^2 + 2(\mu_Y - a)E\{(Y-\mu_Y)\}$$
$$= \sigma_{YY} + (\mu_Y - a)^2$$

- Estimating with one observation

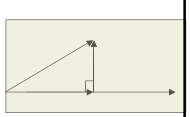
$$\hat{Y} = h_0 + h_1 X \quad \text{the LMMSE solution is}$$

$$h_1 = \frac{\sigma_{XY}}{\sigma_{XX}} \qquad h_0 = \mu_Y - h_1\mu_X = \mu_Y - \frac{\sigma_{XY}}{\sigma_{XX}}\mu_X$$

the minimized residual is

$$E\{(Y-\hat{Y})^2\} = \sigma_{YY} - \frac{\sigma_{XY}{}^2}{\sigma_{XX}} \quad \text{another expression:}$$

$$E\{(Y-\hat{Y})^2\} = \sigma_{YY}(1-\rho_{XY}{}^2), \quad where \quad \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}, \quad \sigma_X = \sqrt{\sigma_{XX}}, \sigma_Y = \sqrt{\sigma_{YY}}$$

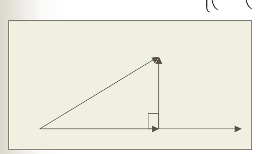


5.1.1 Orthogonality Principle

- A necessary condition for a linear estimator denoted by $h=[h_0, \ldots h_M]^T$ to be the solution of the LMMSE is that

$$E\{Y-\hat{Y}\} = E\left\{Y - \left(h_0 + \sum_{m=1}^{M} h_m X(m)\right)\right\} = 0 \qquad E\{Y\} = E\{\hat{Y}\}$$

$$E\left\{(Y-\hat{Y})X(j)\right\} = E\left\{\left(Y - \left(h_0 + \sum_{m=1}^{M} h_m X(m)\right)\right)X(j)\right\} = 0, \quad j = 1,\ldots,M$$



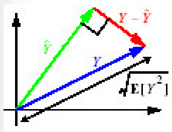
- Orthogonality Principle

5.1.2 Orthogonality Consequences

- Pythagoras Theorem:

$$E\{Y-\hat{Y}\}=0$$
$$E\{(Y-\hat{Y})X(j)\}=0, \quad j=1,\ldots,M$$

$$E\{(Y-\hat{Y})\hat{Y}\}=0$$
$$E\{(Y-\hat{Y})^2\}=E\{Y^2\}-E\{\hat{Y}^2\}=E\{(Y-\hat{Y})Y\}$$



Geometrical interpretation

5.1.3 LMMSE Soultion

$$\mathbf{h}=[h_1 \ \cdots \ h_M]^T \qquad \mu_{\mathbf{X}}=[\mu_{X(1)} \ \cdots \ \mu_{X(M)}]^T$$

$$\sum_{\mathbf{X}Y}=\begin{bmatrix}\sum_{X(1)Y}\\ \vdots \\ \sum_{X(M)Y}\end{bmatrix}, \qquad \sum_{\mathbf{XX}}=\begin{bmatrix}\sum_{X(1)X(1)} & \cdots & \sum_{X(1)X(M)}\\ \vdots & \ddots & \vdots\\ \sum_{X(M)X(1)} & \cdots & \sum_{X(M)X(M)}\end{bmatrix},$$

$$\mu_Y=h_0+\sum_{m=1}^{M}h_m\mu_{X(m)}=h_0+\mathbf{h}^T\mu_{\mathbf{X}}$$
$$\sum_{\mathbf{X}Y}=\sum_{\mathbf{XX}}\mathbf{h}$$

$$\mathbf{h}=\left(\sum_{\mathbf{XX}}\right)^{-1}\sum_{\mathbf{X}Y}$$ ← Under the invertable assumption

$$h_0=\mu_Y-\mathbf{h}^T\mu_{\mathbf{X}}=\mu_Y-\left(\sum_{\mathbf{X}Y}\right)^T\left(\sum_{\mathbf{XX}}\right)^{-1}\mu_{\mathbf{X}}$$

5.1.4 LMMSE Residual

- The solution for LMMSE results the minimal residual as

$$E\{(Y-\hat{Y})^2\}=E\{Y^2\}-E\{\hat{Y}^2\}=E\{(Y-\hat{Y})Y\}$$
$$E\{(Y-\hat{Y})^2\}=\sigma_Y{}^2-\mathbf{h}^T\sum_{\mathbf{X}Y}$$

How do we get this?

5.1.5 LMMSE Summary

- A random sequence X(1), ... X(M) whose realization can be observed
- A random variable Y which has to be estimated
- Seek a linear estimator as: $\hat{Y}=h_0+\sum_{m=1}^{M}h_mX(m)$
- By solving an optimal problem: $\min_{h_m,m=0,1,\ldots M} E\{(Y-\hat{Y})^2\}$
- The solution following the orthogonality is

$$\mathbf{h}=\left(\sum_{\mathbf{XX}}\right)^{-1}\sum_{\mathbf{X}Y}$$
$$h_0=\mu_Y-\mathbf{h}^T\mu_{\mathbf{X}}=\mu_Y-\left(\sum_{\mathbf{X}Y}\right)^T\left(\sum_{\mathbf{XX}}\right)^{-1}\mu_{\mathbf{X}}$$

5.1.6 Example

- ***Example: Linear prediction of a WSS process***

Let $Y(n)$ denote a WSS process with

- zero mean, i.e $\mathbf{E}[Y(n)]=0$,
- autocorrelation function $\mathbf{E}[Y(n)Y(n+k)]=R_{YY}(k)$

We seek the LMMSEE for the present value of $Y(n)$ based on the M past observations $Y(n-1),\ldots,Y(n-M)$ of the process. Hence,

- $Y = Y(n)$
- $X(m)=Y(n-m)$, $m=1,\ldots,M$, i.e.

$$\boldsymbol{X}=[Y(n-1),\ldots,Y(n-M)]^T$$

Because $\mu_Y=0$ and $\mu_X=0$, it follows from $(3.4b)$ that

$$h_0=0$$

Computation of Σ_{XY} and Σ_{XX}:

5.1.6 Example (Cont'd)

$$-\Sigma_{XY}=[\mathbf{E}[Y(n-1)Y(n)],\ldots,\mathbf{E}[Y(n-M)Y(n)]]^T$$
$$=[R_{YY}(1),\ldots,R_{YY}(M)]^T$$

$$-\Sigma_{XX}=$$

$$=\begin{bmatrix}\mathbf{E}[Y(n-1)^2] & \mathbf{E}[Y(n-1)Y(n-2)] & \ldots & \mathbf{E}[Y(n-1)Y(n-M)]\\ \mathbf{E}[Y(n-2)Y(n-1)] & \mathbf{E}[Y(n-2)^2] & \ldots & \mathbf{E}[Y(n-2)Y(n-M)]\\ \ldots & \ldots & \ldots & \ldots\\ \mathbf{E}[Y(n-M)Y(n-1)] & \mathbf{E}[Y(n-M)Y(n-2)] & \ldots & \mathbf{E}[Y(n-M)^2]\end{bmatrix}$$

$$=\begin{bmatrix}R_{YY}(0) & R_{YY}(1) & R_{YY}(2) & \ldots & R_{YY}(M-1)\\ R_{YY}(1) & R_{YY}(0) & R_{YY}(1) & \ldots & R_{YY}(M-2)\\ R_{YY}(2) & R_{YY}(1) & R_{YY}(0) & \ldots & R_{YY}(M-3)\\ \ldots & \ldots & \ldots & \ldots & \ldots\\ R_{YY}(M-1) & R_{YY}(M-2) & R_{YY}(M-3) & \ldots & R_{YY}(0)\end{bmatrix}$$

MM5. Minimum Mean Squared Error Estimation

- 5.1 Linear minimum mean squared error estimators
- **5.2 (Nonlinear) minimum mean squared error estimator**

5.2 Minimum Mean Squared Error Estimators (MMSEE)

- **MMSEE:** The solution for MMSEE of Y based on the observation of X(1),....,X(M) is:

$$\tilde{Y}(X(1),\dots,X(M)) = E\{Y \mid X(1),\dots,X(M)\}$$

- Which reaches $\min E\{(Y-\tilde{Y})^2\}$
- Specially, if X(1)=x(1),....,X(M)=x(M) is observaed, then

$$\tilde{Y}(x(1),\dots,x(M)) = E\{Y \mid x(1),\dots,x(M)\} = \int y p(y \mid x(1),\dots,x(M))dy$$

5.2.1 Remarks of MMSEE

- **Conditional expectation:**
  - Let U and V denote two random variables.
  - E{V|U} is a random variable, E{E{V|U}}=E{V}
- **Proof of MMSEE solution**

$$E\{(Y-\tilde{Y})^2\} = E_{\mathbf{X}}\{E_{Y|\mathbf{X}}\{(Y-\tilde{Y})^2 \mid X(1),\dots,X(M)\}\}$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} E_{Y|\mathbf{X}}\{(Y-\tilde{Y})^2 \mid X(1)=x_1,\dots,X(M)=x_M\} f_{X(1),\dots X(M)}(x_1,\dots,x_M)dx_1,\dots,dx_M$$

$$E_{Y|\mathbf{X}}\{(Y-\bar{Y})^2 \mid X(1)=x_1,\dots,X(M)=x_M\} = E\{(Y-\tilde{Y})^2 \mid X(1)=x_1,\dots,X(M)=x_M\}$$
$$+ E\{(\tilde{Y}-\bar{Y})^2 \mid X(1)=x_1,\dots,X(M)=x_M\} + 2\underbrace{E\{(Y-\tilde{Y})(\tilde{Y}-\bar{Y}) \mid X(1)=x_1,\dots,X(M)=x_M\}}_{=0}$$

5.2.2 Example - I

- Example 7.7 on page 394 of the tetxbook

$$f_{S|X}(s \mid x) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(s-x-\frac{x^2}{2})^2\right\} \quad \leftarrow Gaussian$$

$$(a) \quad MSE-solution: \tilde{S} = E\{S \mid X\} = X + \frac{X^2}{2} \quad (mean)$$

$$MSE-residual: \quad E\{(S-\tilde{S})^2\} = 1 \qquad (conditional\ variance)$$

$$(b)\ Assume \quad F_X(x) = e^{-x}, x \ge 0 \qquad then \quad E\{X\}=1, \quad E\{X^2\}=2, \quad E\{X^3\}=6,$$

$$Linear\ MMSE\ estimator: \quad \hat{S} = h_0 + h_1 X$$

where

$$h_1 = \frac{\sum_{SX}}{\sum_{XX}} = \frac{E\{(S-\mu_S)(X-\mu_X)\}}{E\{(X-\mu_X)^2\}} = \frac{E\{SX\}-\mu_S\mu_X}{E\{X^2\}-\mu_X{}^2} = \frac{E_X\{X(X+\frac{X^2}{2})\}-2}{2-1} = 3$$

$$h_0 = \mu_S - h_1\mu_X = 2-3 = -1$$

$$LMSE\ residual: \quad E\{(S-\hat{S})^2\} = 2$$

5.2.3 Example–II

Example: Multivariate Gaussian variables:

$[Y, X(1), \dots, X(M)]^T \sim \mathcal{N}(\mu, \Sigma)$ with

- $\mu = [\mu_Y, \mu_{X(1)}, \dots, \mu_{X(M)}]^T$
- $\Sigma = \begin{bmatrix} \sigma_Y^2 & (\Sigma_{XY})^T \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix}$

From Equation (6.22) in [Shanmugan] it follows that

$$\widehat{Y} = \mathbf{E}[Y|X] = \mu_Y + (\Sigma_{XY})^T(\Sigma_{XX})^{-1}(X-\mu_X)$$

Bivariate case: $M = 1, X(1) = X$

- $\Sigma_{XX} = \sigma_X^2$,
- $\Sigma_{XY} = \rho\sigma_X\sigma_Y$, where $\rho = \frac{\Sigma_{XY}}{\sigma_X\sigma_Y}$ is the correlation coefficient of Y and X.

In this case,

$$\widehat{Y} = \mathbf{E}[Y|X] = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(X-\mu_X) = \underbrace{\left(\mu_Y - \frac{\rho\sigma_Y}{\sigma_X}\mu_X\right)}_{h_0} + \underbrace{\left(\frac{\rho\sigma_Y}{\sigma_X}\right)}_{h_1}X$$

We can observe that $\widehat{Y}$ is linear, i.e. is the LMMSEE $\hat{Y} = \widehat{Y}$ in the bivariate case. This is also true in the general multivariate Gaussian case. In fact,

$\hat{Y} = \widehat{Y}$ if, and only if, $[Y, X(1), \dots, X(M)]^T$ is a Gaussian random vector.

5.2.4 **Appendix** Properties of the Multivariate Gaissian Distribution (p.50-51)

- PDF: $f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m \sum_{\mathbf{XX}}}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})^T \sum_{\mathbf{XX}}{}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})\right\}$

$$Partition\ \mathbf{X}\ as \quad \mathbf{X} = \begin{bmatrix}\mathbf{X}_1 \\ \mathbf{X}_2\end{bmatrix} \quad where\ \mathbf{X}_1 = \begin{bmatrix}X_1 \\ \vdots \\ X_k\end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix}X_k \\ \vdots \\ X_m\end{bmatrix},$$

$$\mu_{\mathbf{X}} = \begin{bmatrix}\mu_{\mathbf{X}_1} \\ \mu_{\mathbf{X}_2}\end{bmatrix}, \quad \sum_{\mathbf{XX}} = \begin{bmatrix}\sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22}\end{bmatrix}$$

- Uncorrelatedness implies independence
- The random vector **Y=AX** has the Gaussian pdf with

$$\mu_{\mathbf{Y}} = \mathbf{A}\mu_{\mathbf{X}} \quad \sum_{\mathbf{Y}} = \mathbf{A}\sum_{\mathbf{X}}\mathbf{A}^T$$

- Conditional pdf:

$$\mu_{\mathbf{X}_1|\mathbf{X}_2} = E\{\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2\} = \mu_{\mathbf{X}_1} + \sum_{12}\sum_{22}{}^{-1}(\mathbf{x}_2 - \mu_{\mathbf{X}_2}), \quad \sum_{\mathbf{X}_1|\mathbf{X}_2} = \sum_{11} - \sum_{12}\sum_{22}{}^{-1}\sum_{21}$$