Linear Estimation

In Bayesian estimation, we saw a special problem of finding a linear estimator

\[ \hat{\theta}(Y) = AY + b \]

such that \( \hat{\theta}(Y) \) minimizes the MSE.

What we want to do now is to take a closer look at this linear estimator.

To make our notation consistent to Kalman-Bucy Filtering, we consider the following estimator

\[ X_t = \sum_{n=a}^{b} h_{t,n} Y_n + c_t \quad (i) \]

where \( \{ h_{t,n} \}_{n=a}^{b} \) is a sequence of coefficients, and \( c_t \) is a scalar. The index \( t \) can be thought of as the \( t \)th component of a vector.

Remark: When \( a = -\infty \), or \( b = +\infty \), or both, we define the relation in (i) as

\[ \lim_{n \to -\infty} \mathbb{E}\left( (X_t - \sum_{n=m}^{b} h_{t,n} Y_n - c_t)^2 \right) = 0. \]
Now, let us consider the MSE minimization problem:

$$\min_{\hat{X}_t \in H^b_a} E\left[ (\hat{X}_t - X_t)^2 \right]. \tag{2}$$

Here, we defined

$$H^b_a = \text{set of all linear estimators based on} \{Y_a, \ldots, Y_b\}.$$

**Proposition 1.**

$\hat{X}_t \in H^b_a$ solves (2) if and only if

$$E\left[ (\hat{X}_t - X_t)Z \right] = 0 \tag{3}$$

for all $Z \in H^b_a$.

**Proof:** ($\Rightarrow$) Suppose $\hat{X}_t \in H^b_a$ satisfies

$$E\left[ (\hat{X}_t - X_t)^2 \right] = E\left[ \left( \hat{X}_t - \hat{X}_t + \hat{X}_t - X_t \right)^2 \right]$$

$$= E\left[ (\hat{X}_t - \hat{X}_t)^2 \right] + 2 E\left[ (\hat{X}_t - \hat{X}_t)(\hat{X}_t - X_t) \right]$$

$$+ E\left[ (\hat{X}_t - X_t)^2 \right].$$

Clearly, if $\hat{X}_t \in H^b_a$ and $\hat{X}_t \in H^b_a$, then

$$\hat{X}_t - \hat{X}_t \in H^b_a.$$
By (3), we have
\[ \mathbb{E}\left[ (\hat{X}_t - X_t) \left( \frac{\tilde{X}_t - \hat{X}_t}{\mathbb{E}[Z]} \right) \right] = 0, \]

So,
\[ \mathbb{E}\left[ (X_t - \tilde{X}_t)^2 \right] = \mathbb{E}\left[ (\hat{X}_t - \tilde{X}_t)^2 \right] + \mathbb{E}\left[ (\hat{X}_t - X_t)^2 \right] \]
\[ \geq \mathbb{E}\left[ (\hat{X}_t - X_t)^2 \right]. \]

Since \( \tilde{X}_t \) is chosen arbitrarily, \( \hat{X}_t \) is the minimizer of (2).

Suppose \( \tilde{X}_t \in \mathcal{H}_a^b \) and \( \tilde{Z} \in \mathcal{H}_a^b \) such that \( \mathbb{E}\left[ (X_t - \tilde{X}_t)^2 \right] \neq 0 \). We seek to find a contradiction.

Define
\[ \hat{X}_t = X_t + \frac{\mathbb{E}\left[ (X_t - \tilde{X}_t) \tilde{Z} \right]}{\mathbb{E}[\tilde{Z}^2]} \tilde{Z}. \]

(assume \( \mathbb{E}[\tilde{Z}^2] > 0 \)). Then
\[ \mathbb{E}\left[ (X_t - \hat{X}_t)^2 \right] = \mathbb{E}\left[ (X_t - \tilde{X}_t)^2 \right] - 2 \frac{\mathbb{E}\left[ (X_t - \tilde{X}_t) \mathbb{E}(X_t \tilde{Z}) \right]}{\mathbb{E}[\tilde{Z}^2]} \tilde{Z}^2 \]
\[ + \frac{\mathbb{E}\left[ (X_t - \tilde{X}_t)^2 \right] \mathbb{E}[\tilde{Z}^2]}{\mathbb{E}[\tilde{Z}^2]^2} \tilde{Z}^2 \]
\[ = \mathbb{E}\left[ (X_t - \hat{X}_t)^2 \right] - \mathbb{E}\left[ (X_t - \tilde{X}_t)^2 \right] \frac{\mathbb{E}[(X_t - \tilde{X}_t)^2]}{\mathbb{E}[\tilde{Z}^2]} \]
\[ < \mathbb{E}\left[ (X_t - \tilde{X}_t)^2 \right]. \]

Thus \( \hat{X}_t \) is a better estimator. Contradiction.
Proposition 1 provides an important geometric interpretation known as the orthogonality principle. Orthogonality principles says that

\[ \hat{x}_t \text{ is the MMSE estimator if and only if } \hat{x}_t - X_t \text{ is orthogonal to every linear function of } Y_{ta}. \]

**Example**

Let \( x \) and \( y \) be two vectors in \( \mathbb{R}^2 \). Suppose that we would like to approximate \( x \) by \( \alpha y \) for some \( \alpha \) such that the estimation error \( \alpha y - x \) is minimized.

![Diagram showing vector \( x \), \( \alpha y \), and estimation error \( x - \alpha y \). The estimation error is orthogonal to any linear function of \( y \).](image-url)

estimation error is orthogonal to any linear function of \( y \).
**Wiener-Hopf Equation**

**Proposition 2.**

\[ \hat{X}_t \text{ solves (2) if and only if} \]
\[ \begin{align*}
\mathbb{E}[\hat{X}_t] &= \mathbb{E}[X_t] \\
\text{and } \mathbb{E}[ (X_t - \hat{X}_t) Y_e ] &= 0 \text{ for all } a \leq e \leq b.
\end{align*} \tag{4} \]

**Proof:** \((\Rightarrow)\) if \(\hat{X}_t\) solves (2), then by Proposition 1 we have \(\mathbb{E}[ (\hat{X}_t - X_t) Z ] = 0\).

Put \(Z = 1\) yields \(\mathbb{E}[\hat{X}_t] = \mathbb{E}[X_t]\).

Put \(Z = Y_e\) yields \(\mathbb{E}[ (\hat{X}_t - X_t) Y_e ] = 0\).

\((\Leftarrow)\) if \(\hat{X}_t\) satisfies (4), then let \(Z = \sum_{n=a}^{b} h_{t,n} Y_n + c_t\).

\[ \mathbb{E}[ (X_t - \hat{X}_t) Z ] = \sum_{n=a}^{b} h_{t,n} \mathbb{E}[ (X_t - \hat{X}_t) Y_n ] + \mathbb{E}[ X_t - \hat{X}_t ] c_t \]
\[ = 0. \tag*{\square} \]

Using Proposition 2, we obtain the equation

\[ \mathbb{E}[\hat{X}_t] = \mathbb{E}[X_t] \]

\[ \Rightarrow \mathbb{E}\left[ \sum_{n=a}^{b} h_{t,n} Y_n + c_t \right] = \mathbb{E}[X_t] \]

\[ \Rightarrow c_t = \mathbb{E}[X_t] - \sum_{n=a}^{b} h_{t,n} \mathbb{E}[Y_n]. \tag{5} \]
Moreover,
\[ \mathbb{E} \left[ (X_t - \mathbb{E}(X_t)) Y_e \right] = 0 \]
\[ \Rightarrow \mathbb{E} \left[ (X_t - \sum_{n=a}^{b} h_{t,n} Y_n - c_t) Y_e \right] = 0 \quad (6) \]
Substitute (5) into (6) yields
\[ \mathbb{E} \left[ \left( (X_t - \mathbb{E}(X_t)) - \sum_{n=a}^{b} h_{t,n} (Y_n - \mathbb{E}(Y_n)) \right) Y_e \right] = 0 \]
\[ \Rightarrow \mathbb{E} \left[ (X_t - \mathbb{E}(X_t)) Y_e \right] = \sum_{n=a}^{b} h_{t,n} \mathbb{E} \left[ (Y_n - \mathbb{E}(Y_n)) Y_e \right] \]
\[ \Rightarrow \quad \text{Cov}(X_t, Y_e) = \sum_{n=a}^{b} h_{t,n} \text{Cov}(Y_n, Y_e) \]
\[ = \quad \text{Cxy}(t, e) = \sum_{n=a}^{b} h_{t,n} \text{Cxy}(n, e) \quad (7) \]

- (7) is called the Wiener-Hopf equation.
- (7) and (5) are the necessary and sufficient conditions for \( \{h_{t,n}\}_{n=a}^{b} \) and \( c_t \) to be the optimal filter.
- (7) can be written as
\[
\begin{bmatrix}
C_{xy}(t,a) \\
\vdots \\
C_{xy}(t,b)
\end{bmatrix}
= \text{Cov}(X) \begin{bmatrix}
\text{Cov}(a,a) & \text{Cov}(b,a) \\
\vdots & \vdots \\
\text{Cov}(a,b) & \text{Cov}(b,b)
\end{bmatrix}
\begin{bmatrix}
h_{t,a} \\
\vdots \\
h_{t,b}
\end{bmatrix}
\]