Linear Estimation

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

$$\hat{\theta}(\underline{Y}) = A\underline{Y} + b$$

such that $\hat{\theta}(\underline{Y})$ 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

$$\hat{X}_t = \sum_{n=a}^{b} h_{t,n} Y_n + c_t \qquad \text{(i)}$$

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 tth component of a vector.

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

$$\lim_{m \to -\infty} \mathbb{E}\left[\left(\hat{X}_t - \sum_{n=m}^{b} h_{t,n} Y_n - c_t\right)^2\right] = 0.$$

Now, let us consider the MSE minimization problem:

$$\min_{\hat{X}_t \in \mathcal{H}_a^b} \mathbb{E}\left[(\hat{X}_t - X_t)^2\right]. \qquad (2)$$

Here, we defined

$\mathcal{H}_a^b \overset{\text{def}}{=}$ set of all linear estimators based on $\{Y_a, \ldots, Y_b\}$.

<u>Proposition 1.</u>

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

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

for all $Z \in \mathcal{H}_a^b$.

Pf: ($\Rightarrow$) Suppose $\hat{X}_t \in \mathcal{H}_a^b$ ~~solves~~ satisfies (3). Then, let $\tilde{X}_t$ be any other estimator in $\mathcal{H}_a^b$. Then,

$$\begin{aligned}\mathbb{E}\left[(\tilde{X}_t - X_t)^2\right] &= \mathbb{E}\left[(\tilde{X}_t - \hat{X}_t + \hat{X}_t - X_t)^2\right] \\ &= \mathbb{E}\left[(\tilde{X}_t - \hat{X}_t)^2\right] + 2\mathbb{E}\left[(\tilde{X}_t - \hat{X}_t)(\hat{X}_t - X_t)\right] \\ &\quad + \mathbb{E}\left[(\hat{X}_t - X_t)^2\right].\end{aligned}$$

Clearly, if $\hat{X}_t \in \mathcal{H}_a^b$ and $\tilde{X}_t \in \mathcal{H}_a^b$, then

$$\tilde{X}_t - \hat{X}_t \in \mathcal{H}_a^b.$$

By (3), we have

$$\mathbb{E}\Big[(\hat{X}_t - X_t)\underbrace{(\tilde{X}_t - \hat{X}_t)}_{Z}\Big] = 0,$$

So,

$$\mathbb{E}\big[(X_t - \tilde{X}_t)^2\big] = \mathbb{E}\big[(\tilde{X}_t - \hat{X}_t)^2\big] + \mathbb{E}\big[(\hat{X}_t - X_t)^2\big]$$
$$\geq \mathbb{E}\big[(\hat{X}_t - X_t)^2\big].$$

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

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

Define

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

(assume $\mathbb{E}(Z^2) > 0$). Then

$$\mathbb{E}\big[(X_t - \hat{X}_t)^2\big] = \mathbb{E}\Big[(X_t - \tilde{X}_t)^2 - 2\,\mathbb{E}\Big[\frac{(X_t - \tilde{X}_t)\mathbb{E}[(X_t - \tilde{X}_t)Z]}{\mathbb{E}[Z^2]} \cdot Z\Big] + \frac{\mathbb{E}\big[(X_t - \tilde{X}_t)Z\big]^2 \mathbb{E}[Z^2]}{\mathbb{E}[Z^2]^2}$$

$$= \mathbb{E}\big[(X_t - \tilde{X}_t)^2\big] - \frac{\mathbb{E}\big[(X_t - \tilde{X}_t)Z\big]^2}{\mathbb{E}[Z^2]}$$

$$< \mathbb{E}\big[(X_t - \tilde{X}_t)^2\big].$$

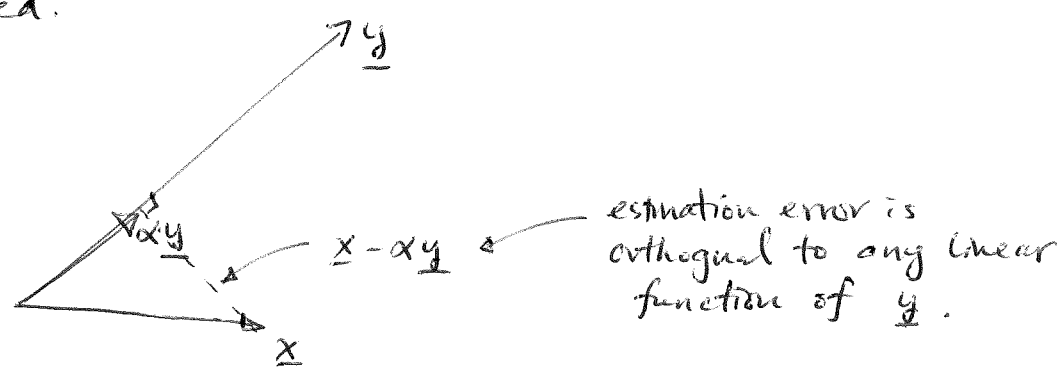
Thus $\hat{X}_t$ is a better estimator. Contradiction.

Proposition 1 provides an important geometric interpretation known as the orthogonality principle. Orthogonality principles says that

$\hat{X}_t$ is the MMSE estimator if and only if the estimation error $\hat{X}_t - X_t$ is orthogonal to every linear function of the observations Y_a^b.

<u>Example</u>

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



Wiener-Hopf Equation

Proposition 2.

$\hat{X}_t$ solves (2) if and only if

$$\left.\begin{aligned} &\mathbb{E}[\hat{X}_t] = \mathbb{E}[X_t] \\ \text{and } &\mathbb{E}[(X_t - \hat{X}_t) Y_\ell] = 0 \quad \text{for all } a \le \ell \le b. \end{aligned}\right\} \text{—(4)}$$

Pf: ($\Rightarrow$) 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 $Z = Y_\ell$ yields $\mathbb{E}[(\hat{X}_t - X_t) Y_\ell] = 0$.

($\Leftarrow$) if $\hat{X}_t$ satisfies (4), then

let $$Z = \sum_{n=a}^{b} h_{t,n} Y_n + c_t.$$

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

$$= 0. \qquad \blacksquare$$

Using Proposition 2, we obtain the equation

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

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

$$\Rightarrow c_t = \mathbb{E}[X_t] - \sum_{n=a}^{b} h_{t,n} \mathbb{E}[Y_n]. \qquad \text{—— (5)}$$

Moreover,

$$\mathbb{E}\left[(X_t - \hat{X}_t) Y_\ell\right] = 0$$

$$\Rightarrow \mathbb{E}\left[\left(X_t - \sum_{n=a}^{b} h_{t,n} Y_n - c_t\right) Y_\ell\right] = 0 \qquad \text{——— (6)}$$

Substitute (5) into (6) yields

$$\mathbb{E}\left[\left(\left(X_t - \mathbb{E}[X_t]\right) - \sum_{n=a}^{b} h_{t,n}\left(Y_n - \mathbb{E}[Y_n]\right)\right) Y_\ell\right] = 0$$

$$\Rightarrow \mathbb{E}\left[\left(X_t - \mathbb{E}[X_t]\right) Y_\ell\right] = \sum_{n=a}^{b} h_{t,n}\, \mathbb{E}\left[\left(Y_n - \mathbb{E}[Y_n]\right) Y_\ell\right]$$

$$\Rightarrow \mathrm{Cov}(X_t, Y_\ell) = \sum_{n=a}^{b} h_{t,n}\, \mathrm{Cov}(Y_n, Y_\ell)$$

$$\equiv \quad C_{XY}(t,\ell) = \sum_{n=a}^{b} h_{t,n}\, C_{YY}(n,\ell) \,. \qquad \text{——— (7)}$$

- (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

$$\begin{bmatrix} C_{XY}(t,a) \\ \vdots \\ C_{XY}(t,b) \end{bmatrix} = \begin{bmatrix} \mathrm{Cov}(a,a) & \cdots & \mathrm{Cov}(b,a) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(a,b) & \cdots & \mathrm{Cov}(b,b) \end{bmatrix} \begin{bmatrix} h_{ta} \\ \vdots \\ h_{tb} \end{bmatrix}$$