ECE 6010
Lecture 9 – Linear Minimum Mean-Square Error Filtering

Background

Recall that for random variable X and Y with finite variance, the MSE $E[(X - h(Y))^2]$ is minimized by $h(Y) = E[X|Y]$. That is, the best estimate of X using a measured value of Y is to find the conditional average of X. One aspect of this estimate is that:

The error is orthogonal to the data.

More precisely, the error $X - E[X|Y]$ is orthogonal to Y *and* to every function of Y:

$$E[(X - E[X|Y])g(Y)] = 0$$

for all measurable functions g. We will assume that $E[g^2(Y)] < \infty$.

We want to show that h minimizes $E[(X-h(Y))^2]$ if and only if $E[(X-h(Y))g(Y)] = 0$ (orthogonality), for all measurable g such that $E[g^2(Y)] < \infty$.

$$E[(X - E[X|Y])g(Y)] = E[E[(X - E[X|Y])|Y]g(Y)] = E[(E[X|Y] - E[X|Y])g(Y)] = 0.$$

Conversely, suppose for some g, $E[(X - h(Y))g(Y)] \neq 0$. Consider the estimate

$$\hat{h}(Y) = h(Y) + \alpha g(Y),$$

where

$$\alpha = \frac{E[(X - h(Y))g(Y)]}{E[g^2(Y)]}.$$

Then

$$E[(X - \hat{h}(Y))^2] = E[(X - h(Y))^2] - \frac{(E[(X - h(Y))g(Y)])^2}{E[g^2(Y)]} < E[(X - h(Y))^2].$$

Suppose now we are given two random processes $\{X_t\}$ and $\{Y_t\}$ that are statistically related (that is, not independent). Suppose, to begin, that $T = \mathbb{R}$. Suppose we observe Y over the interval $[a, b]$, and based on the information gained we want to estimate X_t for some fixed t as a function of $\{Y_t, a \leq t \leq b\}$. That is, we form

$$\hat{X}_t = f(\{Y_\tau, a \leq \tau \leq b\})$$

for some functional f mapping the function to real numbers.

If $t < b$: We say that the operation of the function is *smoothing*.

If $t = b$: We way that the operation of the function is *filtering*.

If $t > b$: We way that the operation of the function is *prediction*.

The error in the estimate is $X_t - \hat{X}_t$. The mean-squared error is $E[(X_t - \hat{X}_t)^2]$.

Fact (built on our previous intuition): The MSE $E[(X_t - \hat{X}_t)^2]$ is minimized by the conditional expectation

$$\hat{X}(t) = E[X_t|Y_\tau, a \leq \tau \leq b].$$

Furthermore, the orthogonality principle applies: $X_t - E[X_t|Y_\tau, a \leq \tau \leq b]$ is orthogonal to every function of $\{Y_\tau, a \leq \tau \leq b\}$.

While we know the theoretical result, it is difficult in general to compute the desired conditional expectation.

Definition 1 Suppose $\{Y_t\}$ is second order. Let $\mathcal{H}_y$ be the set of all random variables of the form $\sum_{i=1}^n a_i Y_{t_i} + c$ for $n \in \mathbb{Z}$ and $a_i, c \in \mathbb{R}$ and $t_i \in [a, b]$. □

Note that $\mathcal{H}_y$ may include infinite sequences, so we assume mean-square limits. The set $\mathcal{H}_y$ contains mean-square derivatives, mean-square integrals, and other linear transformations of $\{Y_t, t \in [a,b]\}$. (The set $\mathcal{H}_y$ is the Hilbert space generated by the linear span of Y_t.)

Let's now solve

$$\min_{\hat{X}_t \in \mathcal{H}_y} E[(X_t - \hat{X}_t)^2]. \qquad ((*))$$

A couple important properties:

- If $E[X_t^2] < \infty$ then $\hat{X}_t \in \mathcal{H}_y$ solves (*) if and only if $E[(X_t - \hat{X}_t)Z] = 0$ for all $Z \in \mathcal{H}_y$. That is, the error is orthogonal to all elements of $\mathcal{H}_y$.

  Proof "If": Suppose $\hat{X}_t \in \mathcal{H}_y$ satisfies $E[(X_t - \hat{X}_t)Z] = 0$ for all $Z \in \mathcal{H}_y$. Let X_t^* be an element of $\mathcal{H}_y$.

  $$\begin{aligned} E[(X_t - X_t^*)^2] &= E[(X_t - \hat{X}_t + \hat{X}_t - X_t^*)^2] \\ &= E[(X_t - \hat{X}_t)^2] + 2E[(X_t - \hat{X}_t)\underbrace{(\hat{X}_t - X_t^*)}_{\in \mathcal{H}_t}] + E[(\hat{X}_t - X_t^*)^2] \\ &= E[(X_t - \hat{X}_t)^2] + E[(\hat{X}_t - X_t^*)^2] \\ &\geq E[(X_t - \hat{X}_t)^2]. \end{aligned}$$

  So the orthogonality condition is sufficient for achieving MMSE.

  "Only if": Suppose $\hat{X}_t \in \mathcal{H}_y$, and there is an element $Z \in \mathcal{H}_y$ such that $E[(X_t - \hat{X}_t)Z] \neq 0$. We will show that there would then be a better estimate: Let

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

  Then

  $$E[(X_t - X_t*)^2] = E[(X_t - \hat{X}_t)^2] - \frac{(E[(X_t - \hat{X}_t)Z])^2}{E[Z^2]} < E[(X_t - X_t^*)^2].$$

  So $\hat{X}_t$ cannot be the MMSE estimator, which implies the necessity of the orthogonality condition.
  □

- $E[(X_t - \hat{X}_t)Z] = 0$ for all $Z \in \mathcal{H}_y$ if and only if $E[\hat{X}_t] = E[X_t]$ and $E[(X_t - \hat{X}_t)Y_\tau] = 0$ for all $\tau \in [a,b]$.

  This is a restatement of orthogonality, but for a restricted space.

  Proof "Only if" (necessity): Want to show that $E[(X_t - \hat{X}_t)Z] = 0$ only if $E[(X_t - \hat{X}_t)] = 0$ and $E[(X_t - \hat{X}_t)Y_\tau] = 0$ for all $\tau \in [a,b]$. But this comes by definition, since $1 \in \mathcal{H}_y$ and $Y_\tau \in \mathcal{H}_y$ for each $\tau \in [a,b]$.

  "If": (sufficiency) Suppose $Z \in \mathcal{H}_y$ and $E[X_t - \hat{X}_t] = 0$ and $E[(X_t - \hat{X}_t)Y_\tau] = 0$ for all $\tau \in [a,b]$. (That is, the error is orthogonal to each $Y\tau$.)

  Then for

  $$Z = \lim_{n\to\infty} (MS)(\sum_{i=1}^{n} a_i Y_{t_i} + c)$$

  we have

  $$E[(X_t - \hat{X}_t)Z] = \lim_{n\to\infty} E[(X_t - \hat{X}_t)(\sum_{i=1}^{n} a_i Y_{t_i} + c)],$$

where the limit may be interchanged because X_t is assumed to be second order,

$$E[(X_t - \hat{X}_t)Z] = \lim_{n\to\infty} \sum_{i=1}^{n} a_i E[(X_t - \hat{X}_t)Y_{t_i}] + cE[(X_t - \hat{X}_t)] = 0.$$

□

Suppose we further restrict $\hat{X}_t$ to be of the form

$$\hat{X}_t = \int_a^b h(t,\tau) Y_\tau d\tau + c_t.$$

That is, $\hat{X}_t$ is the output of a *linear* filter driven by Y_t. Note that $\hat{X}_t \in \mathcal{H}_y$. By property 2, we must have (1)

$$E[X_t] = E[\hat{X}_t] = \int_a^b h(t,\tau)\mu_Y(\tau) d\tau + c_t$$

so that

$$c_t = \mu_x(t) - \int_a^b h(t,\tau)\mu_y(\tau) d\tau$$

and (2):

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

for $\tau \in [a,b]$. That is,

$$R_{XY}(t,\tau) = \int_a^b h(t,\sigma) R_Y(\sigma,\tau) d\sigma + c_t \mu_y(\tau)$$

This gives us two equations in the unknowns c_t and h. We can eliminate c_t:

$$R_{XY}(t,\tau) = \int_a^b h(t,\sigma)(R_Y(\sigma,\tau) - \mu_y(\tau)\mu_y(\sigma)) d\sigma + \mu_x(t)\mu_x(\tau)$$

or

$$(R_{XY}(t,\tau) - \mu_x(t)\mu_y(\tau)) = \int_a^b h(t,\sigma) C_Y(\sigma,\tau)\, d\sigma$$

or

$$C_{XY}(t,\tau) = \int_a^b h(t,\sigma) C_Y(\sigma,\tau)\, d\sigma, \qquad \tau \in [a,b].$$

The optimal h is that which solves this integral equation.

Since we are dealing with covariances, the means have been eliminated. It is frequently assumed that X_t and Y_t have zero means. In this case, the covariances are equal to the correlations, and we can write

$$\boxed{R_{XY}(t,\tau) = \int_a^b h(t,\sigma) R_Y(\sigma,\tau)\, d\sigma.}$$

This equation is called the **Wiener-Hopf** equation.

An integral equation of this form is called a **Fredholm equation**. The theory on the existence of solutions Fredholm integral equations is well-known. In practice, solutions are usually numerical.

The solution h is sometimes called a **Wiener filter**.

Example 1 A Non-Causal Wiener filter. Suppose $a = -\infty$ and $b = \infty$. Suppose that $\{X_t\}$ and $\{Y_t\}$ are individually and jointly W.S.S. Then the Wiener-Hopf equation becomes

$$R_{XY}(t-\tau) = \int_{-\infty}^{\infty} h(t,\sigma) R_Y(\sigma - \tau)\, d\sigma$$

for $-\infty < \tau < \infty$. Let $\nu = \sigma - \tau$. Then

$$R_{XY}(t-\tau) = \int_{-\infty}^{\infty} h(t, \nu + \tau) R_Y(\nu)\, d\nu$$

Let $s = t - \tau$:

$$R_{XY}(s) = \int_{-\infty}^{\infty} h(t, \nu + t - s) R_Y(\nu)\, d\nu.$$

Observe that the left-hand side is independent of t. Thus, if there is a solution, there must be a solution which is independent of t. This means that there is a **time-invariant** solution; we will call it h_0. Then, by a particular choice of the form of h_0 we can write

$$R_{XY}(s) = \int_{-\infty}^{\infty} h_0(s - \nu) R_Y(\nu)\, d\nu.$$

That is,

$$R_{XY} = h_0 * R_Y.$$

How to solve for h_0? Easiest way is to use Fourier transforms:

$$S_{XY}(\omega) = H_0(\omega) S_Y(\omega)$$

so

$$\boxed{H_0(\omega) = \frac{S_{XY}(\omega)}{S_Y(\omega)}.}$$

□

The filter in this case is called a **Non-Causal Wiener Filter**.

Example 2 Suppose

$$Y_t = S_t + N_t$$

where S_t is some signal (random process) of interest and N_t is some "noise" process. Assume that S_t and N_t are independent, and individually and jointly W.S.S. Also assume that they are zero mean. Let

$$X_t = S_{t+\lambda}.$$

Given $\{Y_t\}$, we want to estimate $X_{t+\lambda}$. If $\lambda = 0$, this is a *filtering* problem. If $\lambda > 0$, this is a *prediction* problem.If $\lambda < 0$, this is a *smoothing* problem.

We find

$$S_Y(\omega) = S_S(\omega) + S_N(\omega) + 2\,\mathrm{Re}[S_{SN}(\omega)]$$
$$S_{XY}(\omega) = e^{i\omega\lambda} S_{SY}(\omega) = e^{i\omega\lambda}[S_S(\omega) + S_{SN}(\omega)]$$

We obtain the transfer function

$$H_0(\omega) = \frac{S_{XY}(\omega)}{S_Y(\omega)} = \frac{S_S(\omega) + S_{SN}(\omega)}{S_S(\omega) + S_N(\omega) + 2\Re[S_{SN}(\omega)]} e^{i\omega\lambda}.$$

If signal and noise are orthogonal,

$$H_0(\omega) = \frac{S_{XY}(\omega)}{S_Y(\omega)} = \frac{S_S(\omega)}{S_S(\omega) + S_N(\omega)} e^{i\omega\lambda}.$$

Let us look at the amplitude gain part:

$$\frac{S_S}{S_S + S_N} = \frac{S_S/S_N}{S_S/S_N + 1} \approx \begin{cases} 1 & S_S/S_N \gg 1 \\ 0 & S_S/S_N \ll 1 \end{cases}.$$

□

It can be shown that the residual error for the noncausal Wiener filter is

$$MMSE = \frac{1}{2\pi}\int_{-\infty}^{\infty}[S_X(\omega) - \frac{|S_{XY}(\omega)|^2}{S_Y(\omega)}]d\omega.$$

This can be seen as follows:

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

By orthogonality, the last term is 0, which implies that $E[X_t\hat{X}_t] = E[\hat{X}_t^2]$. We thus obtain

$$E[(X_t - \hat{X}_t)^2] = E[X_t^2] - E[\hat{X}_t X_t] = E[X_t^2] - E[\hat{X}_t^2] = \frac{1}{2\pi}\int_{-\infty}^{\infty}[S_X(\omega) - S_{\hat{X}}(\omega)]\,d\omega$$

The MMSE is sometimes written as

$$MMSE = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega)[1 - |\rho_{XY}(\omega)|^2]\,d\omega$$

where

$$\rho_{XY}(\omega) = \frac{S_{XY}(\omega)}{\sqrt{S_X(\omega)S_Y(\omega)}}.$$

Example 3 For the signal + noise problem, we have

$$MMSE = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{S_S(\omega)S_N(\omega)}{S_S(\omega) + S_N(\omega)}\,d\omega.$$

□

Example 4 Let us now do the signal + noise problem for a particular signal source. Suppose

$$S_S(\omega) = \frac{A^2}{\alpha^2 + \omega^2}$$

and

$$S_N(\omega) = \frac{N_0}{2}.$$

(white noise). Then

$$\begin{aligned}
H_0(\omega) &= \frac{S_S(\omega)}{S_S(\omega) + S_N(\omega)}e^{i\omega\lambda} \\
&= \frac{A^2/(\alpha^2+\omega^2)}{A^2/(\alpha^2+\omega^2) + N_0/2}e^{i\omega\lambda} \\
&= \frac{2A^2}{2A^2 + N_0(\alpha^2+\omega^2)}e^{i\omega\lambda} \\
&= \frac{2A^2}{N_0}\frac{1}{\omega^2 + \alpha^2 + 2A^2/N_0}e^{i\omega\lambda} \\
&= \frac{2A^2}{N_0}\frac{1}{\omega^2+\beta^2}e^{i\omega\lambda} \\
&= \frac{A^2}{\beta N_0}\frac{2\beta}{\omega^2+\beta^2}e^{i\omega\lambda}
\end{aligned}$$

Then

$$h_0(t) = \frac{A^2}{\beta N_0}e^{-\beta|t+\lambda|}.$$

This is not a causal filter. Plot for various values of λ. □

Causal Wiener Filtering

The examples we have seen so far have produced *noncausal* filters, i.e., practically nonimplementable in many cases. We will see what can be done now to make causal filters.

Let us take $a = -\infty$, with $\{X_t\}$ and $\{Y_t\}$ jointly and individually W.S.S., and take $b = t$ (filtering). Furthermore, assume that the filter is time-invariant and *causal*. Then the Wiener-Hopf equations can be written

$$\boxed{R_{XY}(s) = \int_{-\infty}^{t} h(s-\nu)R_Y(\nu)\,d\nu.}$$

The question is, how can this be solved? Because the limit does not proceed to ∞, we can't use conventional transform techniques.

Here are some facts to help. Suppose $S_Y(\omega)$ satisfies the following condition:

$$\int_{-\infty}^{\infty} \frac{\log|S_Y(\omega)|}{1+\omega^2}\,d\omega < \infty$$

(This is known as the Paley-Wiener condition.) Then it turns out that we can write

$$S_Y(\omega) = S_Y^+(\omega)S_Y^(\omega)$$

where

$$|S_Y^+(\omega)|^2 = |S_Y^-(\omega)|^2 = S_Y(\omega)$$

and $\mathcal{F}^{-1}S_Y^+(\omega)$ is zero for negative times (that is, it is causal) and $\mathcal{F}^{-1}S_Y^-(\omega)$ is zero for positive times (that is, it is anticausal). Moreover,

$$\frac{1}{S_Y^+(\omega)}$$

is also causal and

$$\frac{1}{S_Y^-(\omega)}$$

is also anticausal. The proof in general is rather difficult (we will skip it, but give some examples). This factorization is known as the **spectral factorization**.