Lecture 12, Nov. 24, 2010

Linear Estimation

Assume that we have two random variables X_n an Y_n and we have observed Y_n for $a \leq n \leq b$, i.e., we have Y_a^b and we would like to estimate X_t at some time t. We saw before that the MMSE estimate is given as:

$$\hat{X}_t = E[X_t \mid Y_a^b].$$

We saw, in the previous lecture, that when the underlying system has a linear structure and the variables are Gaussian, Kalman-Bucy filter is the optimum estimator. Kalman-Bucy filter gives an efficient (i.e., low complexity) solution to the estimation problem.

In the general case, however, the complexity grows to an impossible to tackle level as the number of observations grow. Also, Computing the expectation $E[X_t | Y_a^b]$ requires the knowledge of the joint density of

$X_t, Y_a, \ldots, Y_b$. This is difficult and most often impossible to have.

A simplifying assumption is to confine our search of the estimator to a smaller class of estimators. For example, confining our attention to the linear class, leads us to find the <u>optimum</u> <u>linear</u> <u>estimator</u>.

One example of the use of linear estimation is LPC (linear Prediction Coding) used for voice coding in cell phones and more recently in VOIP and other applications.

In linear estimation case, one tries to minimize the MSE over the estimators of the form:

$$\hat{X}_t = \sum_{n=a}^{b} h_{t,n} Y_n + c_t$$

where $h_{t,a}, \ldots, h_{t,b}$ and c_t are scalars.

The estimation is optimized by choice of $h_{t,a}, \dots, h_{t,b}$ such that

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

is minimized.

note that, as before, if $t > b$, we are dealing with a <u>prediction</u> problem. When $t = b$, we have <u>filtering</u> problem and when $a < t < b$ we have <u>smoothing</u> problem.

Minimizing

$$E[(X_t - \hat{X}_t)^2] = E\Big[\Big(X_t - \sum_{n=a}^{b} h_{t,n} Y_n - c_t\Big)^2\Big]$$

with respect to $\{h_{t,n}\}_a^b$ results in finding a set of colfficient and an excitation c_n. The colfficients $h_{t,a}, \dots, h_{t,b}$ define a filter whose output is $\hat{X}_t$ when excited with c_t. Finding $\{h_{t,n}\}_a^b$ is called analysis and generating $\hat{X}_t$ is the synthesis operation. For example, in the case of speech compression (say in cellular phones), samples of speech over a 20 ms interval are used as Y_a^b and a set of coefficients are generated (calculated) such that the MSE is minimized. These coefficients plus the resulting error (which is small and can be represented with very few bits) are sent to the receiver. At the receiver a filter is formed with the received coefficients and the error (the excitation) is input to it. This results in generation of $\hat{X}_t$.

<u>Orthogonality Principle</u>:

To minimize

$$E[(X_t - \hat{X}_t)^2] = E[(X_t - \sum_{n=a}^{b} h_{t,n} Y_n - c_t)^2]$$

we take its derivates with respect to $h_{t,a} \cdots h_{t,b}$ and equate them to zero.

$$\frac{\partial}{\partial h_{t,\ell}} E[(X_t - \hat{X}_t)^2] = -2E[(X_t - \sum_{n=a}^{b} h_{t,n} Y_n - c_t) Y_\ell] =$$

So the solution is found from

$$E[(X_t - \hat{X}_t) Y_\ell] = 0 \quad \text{for all } a \leq \ell \leq b$$

Taking derivative with respect to c_t results in

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

or

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

Hence:

$$c_t = E\{X_t\} - \sum_{n=a}^{b} h_{t,n} E[Y_n]$$

Now, we have

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

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

or, equivalently:

Wiener-Hopf equation

$$\boxed{C_{XY}(t,\ell) = \sum_{n=a}^{b} h_{t,n} C_Y(n,\ell)} \quad a \leq \ell \leq b$$

we can write this in vector form:

$$\underline{\sigma}_{XY}(t) = \Sigma_Y \underline{h}_t \, .$$

where

$$\underline{\sigma}_{XY}(t) = [C_{XY}(t,a), \cdots C_{XY}(t,b)]^T$$

$\underline{h}_t = (h_{t,a}, \dots, h_{t,b})^T$

and Σ_y is the covariance matrix of the vector $(Y_a, \dots, Y_b)^T$.

Assuming that Σ_y is positive definite:

$$\underline{h}_t = \Sigma_y^{-1} \underline{\sigma}_{xy}(t)$$

One step prediction:

Assume that we observe $Y_0, Y_1, \dots Y_t$ and we would like to estimate Y_{t+1}, i.e., $X_t = Y_{t+1}$. Then

$$
\begin{aligned}
C_{XY}(t,\ell) &= \mathrm{Cov}(X_t, Y_\ell) \\
&= \mathrm{Cov}(Y_{t+1}, Y_\ell) \\
&= C_Y(t+1, \ell)
\end{aligned}
$$

assume that the process is wide sense stationary (w.s.s.). Then

$$C_{XY}(t,\ell) = C_Y(t+1-\ell)$$

We also have

$$C_Y(n,\ell) = C_Y(n-\ell).$$

The vector equation for finding $h_{t,0}, \dots h_{t,t}$ is then:

$$\begin{bmatrix} C_y(t+1) \\ C_y(t) \\ \vdots \\ C_y(1) \end{bmatrix} = \begin{bmatrix} C_y(0) & C_y(1) & \cdots & C_y(t) \\ C_y(1) & C_y(0) & C_y(1) \cdots & C_y(t-1) \\ \vdots & \vdots & \vdots & C_y(1) \\ C_y(t) & \cdots & C_y(1) & C_y(0) \end{bmatrix} \begin{bmatrix} h_{t,0} \\ h_{t,1} \\ \vdots \\ h_{t,t} \end{bmatrix}$$

These are called <u>Yule-Walker</u> equations. Note that in this case the covariance matrix has a particular structure (the entries on each diagonal are the same). This type of matrix is called a Toeplitz matrix. Toeplitz matrices are easier to invert. They can be inverted in $O(N^2)$ operations as opposed to $O(N^3)$ in general.

The problem can be solved recursively using Levingson algorithm (see the text: page 230).

Wiener-Kolmogrov Filtering:

In addition to assuming stationarity (as was the case of Levinson algorithm), here we assume that the number of observations is infinite and non-causal. That is, we assume that our estimate of X_t is based on the observation of Y_n from $n=-\infty$ to $+\infty$. That is:

$$\hat{X}_t = \sum_{n=-\infty}^{\infty} h_{t,n} Y_n$$

The Wiener-Hopf equations for this problem is:

$$C_{XY}(t,l) = \sum_{n=-\infty}^{\infty} h_{t,n} C_Y(n,l) \qquad -\infty < l < \infty$$

Using the stationarity assumption:

$$C_{XY}(t-l) = \sum_{n=-\infty}^{\infty} h_{t,n} C_Y(n-l) \qquad -\infty < l < \infty$$

Let $\tau = (t-l)$ to get

$$C_{XY}(\tau) = \sum_{n=-\infty}^{\infty} h_{t,n} C_Y(n+\tau-t) \qquad -\infty < \tau < \infty$$

We change variable in the sum by substituting $\alpha = t - n$ to get

$$C_{xy}(\tau) = \sum_{\alpha=-\infty}^{\infty} h_{t,t-\alpha} C_y(\alpha - \tau) \qquad -\infty < \tau < \infty$$

Note that t only appears in the coefficients not in the covariances. So, if the Wiener-Hopf equation has a solution in this case, it can be chosen independent of t.

So, we can replace $h_{t,t-\alpha} = h_{\alpha,0} = h_\alpha$ to get:

$$C_{xy}(\tau) = \sum_{\alpha=-\infty}^{\infty} h_\alpha C_y(\tau - \alpha)$$

This is a convolution of sequences $\{h_n\}_{n=-\infty}^{\infty}$ and $\{C_y(n)\}_{n=-\infty}^{\infty}$. It can be converted into an algebraic equation by taking the discrete-time Fourier Transform to the two sides to get

$$\Phi_{xy}(\omega) = H(\omega) \Phi_y(\omega) \qquad -\pi \leq \omega \leq \pi$$

where

$$H(\omega) = \sum_{n=-\infty}^{\infty} h_n e^{-j\omega n} \qquad -\pi \leq \omega \leq \pi$$

is the transfer function of the filter $\{h_n\}_{n=-\infty}^{\infty}$

$$\Phi_{xy}(\omega) = \sum_{n=-\infty}^{\infty} C_{xy}(n) e^{-j\omega n} \qquad -\pi \leq \omega \leq \pi$$

is the cross power spectral density of $\{X_n\}_{n=-\infty}^{\infty}$ and $\{Y_n\}_{-\infty}^{\infty}$,

and

$$\Phi_y(\omega) = \sum_{n=-\infty}^{\infty} C_y(n) e^{-j\omega n} \qquad -\pi \leq \omega \leq \pi.$$

is the power spectral density of $\{Y_n\}_{n=-\infty}^{\infty}$

From the above form of Wiener-Hopf equation, we find the optimum estimator as:

$$H(\omega) = \frac{\Phi_{xy}(\omega)}{\Phi_y(\omega)} \qquad -\pi \leq \omega \leq \pi$$

from which we can find the filter coefficients using Inverse Fourier Transform

$$h_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_{xy}(\omega)}{\Phi_y(\omega)} e^{j\omega n} \, d\omega$$

The MMSE is given by

$$MMSE = E\{(X_t - \hat{X}_t)^2\}$$

$$= E\{(X_t - \hat{X}_t)X_t\} - \underbrace{E\{(X_t - \hat{X}_t)\hat{X}_t\}}_{\to 0}$$

$$= E\{(X_t - \hat{X}_t)X_t\}$$

$$= E\{X_t^2\} - E\{\hat{X}_t X_t\}$$

$$E\{\hat{X}_t X_t\} = E\{(\sum_{n=-\infty}^{\infty} h_{t-n} y_n) X_t\}$$

$$= \sum_{n=-\infty}^{\infty} h_{t-n} E\{Y_n X_t\}$$

$$= \sum_{n=-\infty}^{\infty} h_{t-n} C_{XY}(t-n) = \sum_{\alpha=-\infty}^{\infty} h_\alpha C_{XY}(\alpha)$$

Note that $\sum_{\alpha=-\infty}^{\infty} h_\alpha C_{XY}(\alpha) = \sum_{\alpha=-\infty}^{\infty} h_\alpha C_{XY}(\tau+\alpha)\Big|_{\tau=0}$

That is, it is the zeroth term of the convolution of $\{h_n\}_{n=-\infty}^{\infty}$ and $\{C_{XY}(-n)\}_{n=-\infty}^{\infty}$

So:

$$E\{\hat{X}_t X_t\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} H(\omega)\,\bar{\Phi}_{XY}(\omega)\,d\omega$$

where $\bar{\Phi}_{XY}(\omega) = DFT\{C_{XY}(-n)\}$, i.e.,

$$\bar{\Phi}_{XY}(\omega) = \sum_{n=-\infty}^{\infty} C_{XY}(-n)\,e^{-jn\omega} \qquad -\pi \le \omega \le \pi$$

Set $\alpha = -n$, we get

$$\bar{\Phi}_{xy}(\omega) = \sum_{\alpha=-\infty}^{\infty} C_{xy}(\alpha) e^{j\omega\alpha} = \Phi_{xy}^{*}(\omega)$$

So:

$$E\{\hat{X}_t X_t\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_{xy}(\omega)}{\Phi_y(\omega)} \Phi_{xy}^{*}(\omega)\, d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|\Phi_{xy}(\omega)|^2}{\Phi_y(\omega)} d\omega .$$

The first term in MMSE, i.e., $E\{X_t^2\}$ is:

$$E\{X_t^2\} = C_X(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_X(\omega)\, d\omega$$

So:

$$MMSE = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\Phi_X(\omega) - \frac{|\Phi_{xy}(\omega)|^2}{\Phi_y(\omega)} \right] d\omega$$

or

$$\boxed{MMSE = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[1 - \frac{|\Phi_{xy}(\omega)|^2}{\Phi_X(\omega)\Phi_y(\omega)} \right] \Phi_X(\omega)\, d\omega}$$

Signal Estimation in Additive Noise

Let the observation be

$$Y_n = S_n + N_n \qquad n \in \mathbb{Z}$$

where the sequences $\{S_n\}_{n=-\infty}^{\infty}$ and $\{N_n\}_{n=-\infty}^{\infty}$ are uncorrelated.

Assume that we would like to estimate

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

if $\lambda > 0$ we have prediction problem.

if $\lambda = 0$ we have filtering problem.

and

if $\lambda < 0$ we have Smoothing problem.

We have:

$$\Phi_Y(\omega) = \Phi_S(\omega) + \Phi_N(\omega) \qquad -\pi \leq \omega \leq \pi$$

and

$$\Phi_{XY}(\omega) = e^{j\omega\lambda}\,\Phi_S(\omega) \qquad -\pi \leq \omega \leq \pi$$

and

$$\Phi_X(\omega) = \Phi_S(\omega) \qquad -\pi \leq \omega \leq \pi$$

So:

$$H(\omega) = \frac{e^{j\omega\lambda}\,\Phi_S(\omega)}{\Phi_S(\omega) + \Phi_N(\omega)} \qquad -\pi \le \omega \le \pi$$

This consists of $e^{j\omega\lambda}$ that accounts for a shift of λ to the data sequence to align it with the time of estimation. The magnitude is

$$|H(\omega)| = \frac{\Phi_S(\omega)/\Phi_N(\omega)}{\Phi_S(\omega)/\Phi_N(\omega) + 1} \qquad -\pi \le \omega \le \pi$$

which varies between 0 and 1 as $\frac{\Phi_S(\omega)}{\Phi_N(\omega)}$ varies between 0 and infinity.

The MMSE is:

$$MMSE = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left[\Phi_X(\omega) - \frac{|\Phi_{XY}(\omega)|^2}{\Phi_Y(\omega)}\right] d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left[\Phi_S(\omega) - \frac{\Phi_S^2(\omega)}{\Phi_S(\omega) + \Phi_N(\omega)}\right] d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Phi_S(\omega)\Phi_N(\omega)}{\Phi_S(\omega) + \Phi_N(\omega)}\, d\omega$$