LECTURE 25 : LEVINSON FILTERING

Levinson filtering is concerned with one-step prediction of a random sequence whose second order statistics are stationary in time. The goal is to find the best linear estimator of Y_{t+1} given the observations $\{Y_0,Y_1,...,Y_t\}$.

Definition 0.1. Observation sequence $\{Y_n\}$,where $n \in (-\infty, \infty)$ is "wide sense stationary" process (w.s.s), if the mean $\mu = \mathbb{E}\{Y_n\}$ is constant, and the covariance function,

$$C_y(\tau) = C_y(t, t+\tau) = \mathbb{E}\{Y_t Y_{t+\tau}\},\ \forall\ \tau \in \mathbf{Z},$$

depends only on the time difference τ.

Let, $X_t = Y_{t+1}$ (refer lecture 24 for the notation), the cross-covariance function of $X_n = Y_{n+1}$ and Y_n, where $n \in (-\infty, \infty)$ is,

$$\begin{aligned} C_{X,Y}(t,l) &= Cov(X_t, Y_l) \\ &= Cov(Y_{t+1}, Y_l) \\ &= C_y(t+1-l) \end{aligned}$$

The Weiner-Hopf equation i.e $\sigma_{XY} = \Sigma_y h_t$ becomes,

$$\begin{bmatrix} C_y(t+1) \\ \vdots \\ C_y(1) \end{bmatrix} = \begin{bmatrix} C_y(0) & \dots & C_y(t) \\ \vdots & \ddots & \vdots \\ C_y(t) & \dots & C_y(0) \end{bmatrix} \quad \begin{bmatrix} h_{t,0} \\ \vdots \\ h_{t,t} \end{bmatrix}, \tag{1}$$

a set of equations also known as Yule Walker equations.

$$\mathbf{\Sigma}_Y^{0:t} = \begin{bmatrix} C_y(0) & \dots & C_y(t) \\ \vdots & \ddots & \vdots \\ C_y(t) & \dots & C_y(0) \end{bmatrix},$$

where $\Sigma_Y^{0:t}$ is a Toeplitz matrix , meaning its enries are constant along the diagonals. Also, a $t \times t$ Toeplitz matrix can be inverted in number of operations that is of the order of t^2, as against t^3 for a unstructured matrix.

0.1 Levinson-Durbin algorithm

In this section, we discuss the algorithm to compute $h_{t,0}, \ldots, h_{t,t}$ in number of operations of the order of t^2. Let $\hat{Y}_{t+1} = -\sum_{n=0}^{t} a_{t+1,t+1-n} Y_n$, then Yule-Walker equations become,

$$-\begin{bmatrix} C_y(t+1) \\ \vdots \\ C_y(1) \end{bmatrix} = \begin{bmatrix} C_y(0) & \ldots & C_y(t) \\ \vdots & \ddots & \vdots \\ C_y(t) & \ldots & C_y(0) \end{bmatrix} \begin{bmatrix} a_{t+1,t+1} \\ \vdots \\ a_{t+1,1} \end{bmatrix} \tag{2}$$

Let $e_{t+1} := \hat{Y}_{t+1} - Y_{t+1}$, and let $\epsilon_{t+1} = \mathbb{E}\{e_{t+1}^2\}$,

$$\begin{aligned} \epsilon_{t+1} &= \mathbb{E}\{e_{t+1}(-\sum_{n=0}^{t} a_{t+1,t+1-n} Y_n - Y_{t+1})\}, \\ &= \mathbb{E}\{-e_{t+1} Y_{t+1}\}, \\ &= -\mathbb{E}\{(-\sum_{n=0}^{t} a_{t+1,t+1-n} Y_n - Y_{t+1}) Y_{t+1}\}, \\ &= C_Y(0) + \sum_{n=0}^{t} C_y(t+1-n) a_{t+1,t+1-n}. \end{aligned}$$

Now transform eq. (1) by, (a) making LHS=$\vec{0}$,

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} C_Y(0) & \ldots & C_Y(t) & C_Y(t+1) \\ \vdots & \ddots & \vdots & \vdots \\ c_Y(t) & \ldots & C_Y(0) & C_Y(1) \end{bmatrix} \begin{bmatrix} a_{t+1,t+1} \\ \vdots \\ a_{t+1,1} \\ 1 \end{bmatrix}, \tag{3}$$

and (b) add equation for ϵ_{t+1},

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \epsilon_{t+1} \end{bmatrix} = \begin{bmatrix} C_Y(0) & \ldots & C_Y(t+1) \\ \vdots & \ddots & \vdots \\ c_Y(t) & \ldots & C_Y(1) \\ C_Y(t+1) & \ldots & \ldots C_y(0) \end{bmatrix} \begin{bmatrix} a_{t+1,t+1} \\ \vdots \\ a_{t+1,1} \\ 1 \end{bmatrix}. \tag{4}$$

Now, consider normal equations for prediction at order $(t+1)$. Let's add an extra equation

$$\begin{bmatrix} \alpha_{t+1} \\ 0 \\ \vdots \\ 0 \\ \epsilon_{t+1} \end{bmatrix} = \begin{bmatrix} C_Y(0) & \ldots & C_Y(t+1) & C_y(t+2) \\ C_Y(1) & C_Y(0) & \ldots & C_Y(t+1) \\ \vdots & \vdots & \ddots & \vdots \\ C_Y(t+1) & C_Y(t+1) & \ldots & C_Y(0) \end{bmatrix} \begin{bmatrix} 0 \\ a_{t+1,t+1} \\ \vdots \\ a_{t+1,1} \\ 1 \end{bmatrix}, \tag{5}$$

where $\alpha_{t+1} := C_y(1)a_{t+1,t+1} + \ldots + C_y(t+1)a_{t+1,1} + C_y(t+2)$. Transforming eqn. (5) by reversing rows and then columns,

$$\begin{bmatrix} \epsilon_{t+1} \\ 0 \\ \vdots \\ 0 \\ \alpha_{t+1} \end{bmatrix} = \begin{bmatrix} C_Y(0) & \ldots & C_Y(t+1) & C_y(t+2) \\ C_Y(1) & C_Y(0) & \ldots & C_Y(t+1) \\ \vdots & \vdots & \ddots & \vdots \\ C_Y(t+2) & C_Y(t+1) & \ldots & C_Y(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_{t+1,1} \\ \vdots \\ a_{t+1,t+1} \\ 0 \end{bmatrix} \tag{6}$$

Now multiplying eqn. (6) with K_{t+1} and then adding with eqn. (5), where $K_{t+1} = -\frac{\alpha_{t+1}}{\epsilon_{t+1}}$, we get

$$\begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ \epsilon_{t+1} - \frac{\alpha_{t+1}^2}{\epsilon_{t+1}} \end{bmatrix} = \begin{bmatrix} C_Y(0) & \ldots & C_Y(t+1) & C_y(t+2) \\ C_Y(1) & C_Y(0) & \ldots & C_Y(t+1) \\ \vdots & \vdots & \ddots & \vdots \\ C_Y(t+2) & C_Y(t+1) & \ldots & C_Y(0) \end{bmatrix} \begin{bmatrix} b_{t+2,t+2} \\ \vdots \\ \vdots \\ b_{t+2,1} \end{bmatrix} \tag{7}$$

Comparing the form of eqn (7) and (4), by inspection, we get,

$$\epsilon_{t+2} = \epsilon_{t+1} - \frac{\alpha_{t+1}^2}{\epsilon_{t+1}} = \epsilon_{t+1} - \frac{(K_{t+1}\epsilon_{t+1})^2}{\epsilon_{t+1}} = \epsilon_{t+1}(1 - K_{t+1}^2). \tag{8}$$

For order of t operations, we have, $a_{t+2,t+2} = K_{t+1}$, and $a_{t+2,t+2-n} = a_{t+1,t+2-n} + K_{t+1}a_{t+1-n}$ $1 \leq n \leq t+1$.

Example 0.2. Illustration

Time	0	1	2
Observation	ϕ	Y_0	Y_0, Y_1
To Predict	Y_0	Y_1	Y_2
Operations			
a	$\hat{Y}_0 = 0$	$a_{1,1} = K_0$	$a_{2,2} = K_1$
b	$\epsilon_0 = C_y(0)$	$\epsilon_1 = (1 - K_0^2)\epsilon_0$	$a_{2,1} = a_{1,1} + k_1 a_{1,1}$
c	$k_0 = -\frac{C_y(1)}{C_y(0)}$	$k_1 = -\frac{[C_Y(2)+C_Y(1)a_{1,1}]}{\epsilon_1}$	$\epsilon_2 = (1 - K_1^2)\epsilon_1$
d			$k_2 = -\frac{[C_Y(3)+C_Y(2)a_{2,1}+C_Y(1)a_{2,2}]}{\epsilon_2}$

0.2 Pseudocode: Levinson-Durbin

Initialize $\epsilon_0 = C_y(0)$ and $K_0 = -\frac{C_Y(1)}{C_Y(0)}$

for $t = 1, 2, \ldots$ **do**

$\quad a_{t,t} = K_{t-1}$

$\quad a_{t,t-n} = a_{t-1,t-n} + K_{t-1}a_{t-1-n}$, where $1 \leq n \leq t-1$

$\quad \epsilon_t = \epsilon_{t-1}(1 - k_{t-1}^2)$

$\quad k_t = -\frac{C_y(t+1) + \sum_{n=0}^{t-1} C_y(n+1)a_{t,t-n}}{\epsilon_t}$

end for

Remark 1. Interpretation of K_t:
By Orthogonality condition 'error and observation are orthogonal', i.e., $(Y_t - \hat{Y}_t) \quad \coprod \quad (Y_0, Y_1, \ldots .Y_{t-1})$. But the error $Y_t - \hat{Y}_t$ need not be orthogonal to Y_{-1}, and hence,

$$
\begin{aligned}
E[(Y_t - \hat{Y}_t)Y_{-1}] &= E[(Y_t + \sum_{n=0}^{t-1} a_{t,t-n}Y_n)Y_{-1}], \\
&= C_Y(t+1) + \sum_{n=0}^{t-1} a_{t,t-n}C_Y(n+1), \\
&= \alpha_t.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
K_t &= -\frac{\alpha_t}{\epsilon_t}, \\
&= \frac{E[(\hat{Y}_t - Y_t)Y_{-1}]}{E[(\hat{Y}_t - Y_t)^2]}, \\
&= \frac{Cov[\hat{Y}_t - Y_t, Y_{-1}]}{E[(\hat{Y}_t - Y_t)^2]},
\end{aligned}
$$

where K_t are called PARCOR (Partial Correlation) coefficients.

Remark 2. Mean square error (MSE) at time t is monotone, $\epsilon_0 \geq \epsilon_1 \geq \epsilon_2 \geq \ldots$

Remark 3. Suppose for all $m > M$, we have $\epsilon_m = \epsilon_{m+1} = \ldots$. This is called "stalling". When stalling occurs, the prediction errors become orthogonal.