Chapter 9

Linear Estimation

In previous chapters we saw that the optimum estimation, in a MMSE sense, of an unknown signal X_t, given the observations $Y_{a:b} = \{Y_a, \ldots, Y_b\}$ of a related quantity, is given by $\widehat{X}_t = \mathrm{E}\left[X_t | Y_{a:b}\right]$. This estimate is not, in general, a linear function of the data and its computation needs the knowledge of the joint distribution of $\{X_t, Y_{a:b}\}$. Only when this joint distribution is Gaussian and when X_t is a related to $Y_{a:b}$ by a linear relation, this optimal estimate is a linear function of the data. Even in this case, its computation needs the inversion of the covariance matrix of the data $\mathbf{\Sigma}_Y$ whose dimensions increase with the number of data.

One way to circumvent these drawbacks is, from the first step, to constraint the estimate to be a linear function of the data. Doing so, as we will see below, we do not need anymore the joint distribution of $\{X_t, Y_{a:b}\}$ but only its second order statistics. Furthermore, we will see that, in this case, we can develop real time or on-line algorithms with lower complexity and lower cost, if we assume data to be stationary.

9.1 Introduction

Assume that we want to obtain an estimate $\widehat{X}_t$ of a quantity X_t which is a linear (or more precisely an affine) function of the data $Y_{a:b} = \{Y_a, \ldots, Y_b\}$, *i.e.*

$$\widehat{X}_t = \sum_{n=a}^{b} h_{t,n} Y_n + c_t \tag{9.1}$$

where, in general, a can be either $-\infty$ or finite and b can also be either finite or ∞. When a and b are finite the meaning of the summation is clear. For the cases where $a = -\infty$ or $b = \infty$, these and all the following summations have to be understood in the MMSE sense, for example for the case $a = -\infty$

$$\lim_{m \mapsto -\infty} \mathrm{E}\left[\left(\sum_{n=m}^{b} h_{t,n} Y_n + c_t - \widehat{X}_t\right)^2\right] = 0 \tag{9.2}$$

In these cases we need also to assume that

$$\mathrm{E}\left[X_n^2\right] < \infty \quad \text{and} \quad \mathrm{E}\left[Y_n^2\right] < \infty.$$

The following propositions resume all we need for developing linear estimation theory.

Proposition 1 *Assume $\widehat{X}_t \in \mathcal{H}_a^b$ where $\mathcal{H}_a^b$ is the Hilbert space generated by the affine transform (9.1). Then $E\left[\widehat{X}_t^2\right] < \infty$ and if Z is a random variable satisfying $E\left[Z^2\right] < \infty$, then*

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

Proposition 2 (Orthogonality principle) *$\widehat{X}_t \in \mathcal{H}_a^b$ solves*

$$\min_{\widehat{X}_t \in \mathcal{H}_a^b} E\left[(\widehat{X}_t - X_t)^2\right] \tag{9.3}$$

if and only if

$$E\left[(\widehat{X}_t - X_t)\,Z\right] = 0 \quad \forall Z \in \mathcal{H}_a^b. \tag{9.4}$$

In other words, $\widehat{X}_t$ is a MMSE linear estimate of X_t given $Y_{a:b}$ if and only if the estimation error $(\widehat{X}_t - X_t)$ is orthogonal to every linear function of the observation $Y_{a:b}$.

Considering the particular cases of $Z = 1$ and $Z = Y_l,\ a \le l \le b$ we can rewrite this proposition in the following way

Proposition 3 *$\widehat{X}_t \in \mathcal{H}_a^b$ solves (9.3) if and only if*

$$E\left[\widehat{X}_t\right] = E[X_t] \tag{9.5}$$

and

$$E\left[(\widehat{X}_t - X_t)\,Y_l\right] = 0 \quad \forall a \le l \le b. \tag{9.6}$$

Now replacing (9.1) in (9.6) we obtain

$$\mathrm{E}\left[(X_t - \widehat{X}_t)\,Y_l\right] = \mathrm{E}\left[(X_t - \sum_{n=a}^{b} h_{t,n} Y_n - c_t)\,Y_l\right] = 0 \quad \forall a \le l \le b. \tag{9.7}$$

To go further in details more easily and without any loss of generality, we assume $\mathrm{E}\left[Y_l\right] = 0, \forall a \le l \le b$. Then, since $\mathrm{E}\left[X_t\right] = c_t$, the previous equation becomes

$$\mathrm{Cov}\left\{X_t, Y_l\right\} = \sum_{n=a}^{b} h_{t,n} \mathrm{Cov}\left\{Y_n, Y_l\right\} \quad \forall a \le l \le b. \tag{9.8}$$

which is known as the *Wiener-Hopft* equation.

Writing this in a matrix form we have

$$\boldsymbol{\sigma}_{XY}(t) = \boldsymbol{\Sigma}_Y \boldsymbol{h}_t \tag{9.9}$$

where

$$\begin{aligned}
\boldsymbol{\sigma}_{XY}(t) &\stackrel{\text{def}}{=} \left[\mathrm{Cov}\left\{X_t, Y_a\right\}, \dots, \mathrm{Cov}\left\{X_t, Y_b\right\}\right]^t \\
\boldsymbol{\Sigma}_{XY}(t) &\stackrel{\text{def}}{=} \left[\mathrm{Cov}\left\{Y_n, Y_l\right\}\right] \\
\boldsymbol{h}_t &\stackrel{\text{def}}{=} \left[h_{t,a}, \dots, h_{t,b}\right]^t
\end{aligned}$$

So, theoretically we have

$$\boldsymbol{h}_t = \boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\sigma}_{XY}(t) \tag{9.10}$$

The main difficulty is however the computation of $\boldsymbol{\Sigma}_Y^{-1}$. Note that this matrix is symmetric and positive definite. So, theoretically, it is not singular. However, its inversion cost increases exponentially with the number of data. In the following we will see how the stationary assumption will help to reduce this cost.

9.2 One step prediction

Consider the case where $a = 0$, $b = t$ and $X_t = Y_{t+1}$ and assume that Y_l is wide sense stationary, *i.e.*, $\mathrm{E}\left[Y_l\right] = 0$ and $\mathrm{Cov}\left\{Y_l, Y_m\right\} = C_Y(l-m)$. Then we have

$$\mathrm{Cov}\left\{X_t, Y_l\right\} = \mathrm{Cov}\left\{Y_{t+1}, Y_l\right\} = C_Y(t+1-l) \tag{9.11}$$

and the Wiener-Hopft equation becomes

$$\begin{pmatrix} C_Y(t+1) \\ C_Y(t) \\ \vdots \\ \\ \vdots \\ C_Y(1) \end{pmatrix} = \begin{pmatrix} C_Y(0) & C_Y(1) & & & C_Y(t) \\ C_Y(1) & \ddots & \ddots & & \\ & \ddots & & & \\ & & & & \\ & & & & C_Y(1) \\ C_Y(t) & & & C_Y(1) & C_Y(0) \end{pmatrix} \begin{pmatrix} h_{t,0} \\ h_{t,1} \\ \vdots \\ \\ \vdots h_{t,t} \end{pmatrix} \tag{9.12}$$

called *Yule-Walker* equation.

Note that the w.s.s. hypothesis of the data Y_n leads to a covariance matrix which is Toeplitz. Unlike the general case, the cost of the inversion of this matrix is only $O(n^2)$ against $O(n^3)$ for the general case, where n is the number of data. Thus, in any linear MMSE estimation problem, the w.s.s. assumption can reduce the complexity of the computation of the coefficients by a factor equal to the number of the data.

In the following, we will see that, we can still go further and use the specific structure of the Yule-Walker equation to keep on reducing this cost.

9.3 Levinson algorithm

Levinson algorithm uses the special structure of the Yule-Walker equation for the one step prediction problem where the left hand side vector of this equation is equal to the last column of the covariance matrix shifted by one time unit.

Rewriting this equation

$$\widehat{Y}_{t+1} = \sum_{n=0}^{t} h_{t,n} Y_n = -\sum_{n=0}^{t} a_{t+1,t+1-n} Y_n \tag{9.13}$$

the coefficients $a_{t,1}, \ldots, a_{t,t}$ can be updated recursively in t through the Levinson algorithm :

$$\begin{aligned} a_{t+1,k} &= a_{t,k} - k_t\, a_{t,t+1-k}, \quad k = 1, \ldots, t \\ a_{t+1,t+1} &= -k_t \end{aligned}$$

where k_t itself, is generated recursively with $\epsilon_t \stackrel{\text{def}}{=} \mathrm{E}\left[(Y_t - \widehat{Y}_t)^2\right]$ via

$$k_t = \frac{1}{\epsilon_t}[C_Y(t+1) + \sum_{k=1}^{t} a_{t,k}\, C_Y(t+1-k)]$$

$$\epsilon_{t+1} \;=\; (1-k_t^2)\epsilon_t$$

with the initialization $k_0 = \frac{C_Y(1)}{C_Y(0)}$ and $\epsilon_0 = C_Y(0)$.

The coefficients a_k are called *reflection coefficients* or still *partial correlation coefficients* (PARCOR).

9.4 Vector observation case

The linear estimation can be extended to the case where both the observation sequence and the quantity to be estimated are vectors. This extension is straight forward and we have:

$$\underline{\widehat{X}}_t = \sum_{n=a}^{b} \boldsymbol{H}_{t,n} \underline{Y}_n + \underline{c}_t \tag{9.14}$$

where $\boldsymbol{H}_{t,n}$ is a sequence of matrices. When a or b are infinite, the summations have the MSE sense. For example when $a = -\infty$, we have

$$\lim_{m \mapsto -\infty} \mathrm{E}\left[\left\|\sum_{n=m}^{b} \boldsymbol{H}_{t,n}\underline{Y}_n + \underline{c}_t - \underline{\widehat{X}}_t\right\|^2\right] = 0 \tag{9.15}$$

where $\|\boldsymbol{x}\| \stackrel{\text{def}}{=} \boldsymbol{x}^t\boldsymbol{x}$.

The orthogonality principle becomes:

Proposition 4 (Orthogonality principle) $\underline{\widehat{X}}_t \in \mathcal{H}_a^b$ *solves*

$$\min_{\underline{\widehat{X}}_t \in \mathcal{H}_a^b} E\left[\left\|\underline{\widehat{X}}_t - \underline{X}_t\right\|^2\right] \tag{9.16}$$

if and only if

$$E\left[(\underline{\widehat{X}}_t - \underline{X}_t)^t\, \underline{Z}\right] = 0 \quad \forall Z \in \mathcal{H}_a^b. \tag{9.17}$$

Writing this last equation for $\underline{Z} = \underline{1}$ and for $\underline{Z} = \underline{Y}_l^t$ we obtain

$$\begin{aligned} \mathrm{E}\,[\underline{X}_t] &= \mathrm{E}\left[\underline{\widehat{X}}_t\right] \\ \mathrm{E}\left[(\underline{X}_t - \underline{\widehat{X}}_t)\,\underline{Y}_l^t\right] &= 0, \quad \forall a \le l \le b. \end{aligned}$$

Using these relations we obtain

$$\mathrm{E}\left[(\underline{X}_t - \underline{\widehat{X}}_t)\,\underline{Y}_l^t\right] = \mathrm{E}\left[\left(\underline{X}_t - \sum_{n=a}^{b} \boldsymbol{H}_{t,n}\underline{Y}_n - \underline{c}_t\right)\,\underline{Y}_l^t\right] = [\boldsymbol{0}], \quad \forall a \le l \le b. \tag{9.18}$$

where $[\mathbf{0}]$ means a matrix whose all elements are equal to zero. The Wiener-Hopft equation becomes:

$$\boldsymbol{C}_{XY}(t,l) = \sum_{n=a}^{b} \boldsymbol{H}_{t,n}\boldsymbol{C}_{Y}(n,l), \quad \forall a \leq l \leq b$$

where $\boldsymbol{C}_{XY}(t,l) \stackrel{\text{def}}{=} \text{Cov}\{\underline{X}_t, \underline{Y}_l\}$ is the cross-covariance of $\{\underline{X}_n\}_{n=-\infty}^{\infty}$ and $\{\underline{Y}_l\}_{l=-\infty}^{\infty}$ and $\boldsymbol{C}_Y(n,l) \stackrel{\text{def}}{=} \text{Cov}\{\underline{Y}_n, \underline{Y}_l\}$ is the auto-covariance of $\{\underline{Y}_n\}_{n=-\infty}^{\infty}$. Note that $\boldsymbol{C}_{XY}(t,l)$ and $\boldsymbol{C}_Y(n,l)$ are $(m \times k)$ and $(k \times k)$ matrices respectively, where k and m are respectively the dimensions of the vectors $\underline{Y}_n$ and $\underline{X}_t$.

9.5 Wiener-Kolmogorov filtering

We assume here that Y_n is wide sense stationarity (w.s.s.) and that there is an infinite number of observations.

Two cases are of interest: Non causal where $(a = -\infty, b = t)$ and Causal where $(a = -\infty, b = \infty)$.

9.5.1 Non causal Wiener-Kolmogorov

Without losing any generality, we assume $\text{E}[Y_n] = \text{E}[X_n] = 0$. Then we have

$$\widehat{X}_t = \sum_{n=-\infty}^{\infty} h_{t,n} Y_n \tag{9.19}$$

The Wiener-Hopft equation becomes

$$C_{XY}(t-l) = \sum_{n=-\infty}^{\infty} h_{t,n} C_Y(n-l) \tag{9.20}$$

Changing variable $\tau = t - l$ we obtain

$$C_{XY}(\tau) = \sum_{n=-\infty}^{\infty} h_{t,n} C_Y(n+\tau-t) \tag{9.21}$$

Changing now the summation variable $n = t - \alpha$ we obtain

$$C_{XY}(\tau) = \sum_{\alpha=-\infty}^{\infty} h_{t,t-\alpha} C_Y(\tau-\alpha) \tag{9.22}$$

In this summation t appears only in $h_{t,t-\alpha}$. This means that we can choose it to be independent of t, *i.e.* if this equation has a solution, it can be chosen to be time-invariant with coefficients $h_{t,t-\alpha} = h_{0,0-\alpha} \stackrel{\text{def}}{=} h_\alpha$. Then we have

$$C_{XY}(\tau) = \sum_{\alpha=-\infty}^{\infty} h_\alpha C_Y(\tau-\alpha) \tag{9.23}$$

which is a convolution equation. Using then the following DFTs

$$\begin{aligned} H(\omega) &\stackrel{\text{def}}{=} \sum_{n=-\infty}^{\infty} h_n \exp\left[-j\omega n\right], \quad -\pi < \omega < \pi \\ S_{XY}(\omega) &\stackrel{\text{def}}{=} \sum_{n=-\infty}^{\infty} C_{XY}(n) \exp\left[-j\omega n\right] \quad -\pi < \omega < \pi \\ S_Y(\omega) &\stackrel{\text{def}}{=} \sum_{n=-\infty}^{\infty} C_Y(n) \exp\left[-j\omega n\right] \quad -\pi < \omega < \pi \end{aligned}$$

we have

$$S_{XY}(\omega) = H(\omega)\, S_Y(\omega)$$

and finally

$$H(\omega) = \frac{S_{XY}(\omega)}{S_Y(\omega)} \tag{9.24}$$

The coefficients h_n can then be obtained by inverse FT

$$h_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{S_{XY}(\omega)}{S_Y(\omega)} \exp\left[j\omega n\right] \mathrm{d}\omega, \quad n \in \mathbf{Z} \tag{9.25}$$

It is interesting to see that we have

$$\begin{aligned} \mathrm{E}\left[\widehat{X}_t X_t\right] &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{S_{XY}(\omega)}{S_Y(\omega)} \mathrm{d}\omega \\ \mathrm{E}\left[X_t^2\right] &= C_X(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_X(\omega)\, \mathrm{d}\omega \\ \mathrm{E}\left[(\widehat{X}_t - X_t)^2\right] &= \frac{1}{2\pi} \int_{-\pi}^{\pi} S_X(\omega) - \frac{S_{XY}(\omega)}{S_Y(\omega)} \mathrm{d}\omega \end{aligned}$$

This last equation can be written

$$MMSE = \mathrm{E}\left[(\widehat{X}_t - X_t)^2\right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[1 - \frac{|S_{XY}(\omega)|^2}{S_X(\omega) S_Y(\omega)}\right] S_X(\omega)\, \mathrm{d}\omega \tag{9.26}$$

Noting that we have $|S_{XY}(\omega)|^2 \leq S_X(\omega)\, S_Y(\omega)$, with equality if $\{X_t\}_{t=-\infty}^{\infty}$ and $\{Y_n\}_{n=-\infty}^{\infty}$ are perfectly correlated, we can conclude that the MMSE ranges from $\mathrm{E}\left[X_t^2\right]$ to zero as the relationship between $\{X_t\}_{t=-\infty}^{\infty}$ and $\{Y_n\}_{n=-\infty}^{\infty}$ ranges from independence to perfect correlation.

Example 1 (Noise filtering) *Consider the model*

$$Y_n = S_n + N_n, \quad n \in \boldsymbol{Z}$$

where S_n and N_n are assumed uncorrelated, zero mean and w.s.s. Suppose that we want to estimate $X_t = S_{t+\lambda}$ for some integer λ. The problem represents filtering, prediction *and smoothing respectively when $\lambda = 0$, $\lambda > 0$ and when $\lambda < 0$. To obtain the necessary equations, it is straightforward to show*

$$\begin{aligned} S_Y(\omega) &= S_X(\omega) + S_N(\omega) \\ S_{XY}(\omega) &= \exp[j\omega\lambda]\ S_X(\omega) \\ S_X(\omega) &= S_S(\omega) \end{aligned}$$

So, the transfer function of the optimum non causal filter is

$$H(\omega) = \frac{\exp[j\omega\lambda]\ S_S(\omega)}{S_S(\omega) + S_N(\omega)}$$

Example 2 (Deconvolution) *Consider the model*

$$Y_n = \sum_{k=0}^{p} h_k S_{n-k} + N_n, \quad n \in \boldsymbol{Z}$$

where S_n and N_n are assumed uncorrelated, zero mean and w.s.s. Suppose that we want to estimate $X_t = S_{t+\lambda}$ for some integer λ. Again here, it is straightforward to show

$$\begin{aligned} S_Y(\omega) &= |H(\omega)|^2 S_X(\omega) + S_N(\omega) \\ S_{XY}(\omega) &= \exp[j\omega\lambda]\ H^*(\omega)\, S_X(\omega) \\ S_X(\omega) &= S_S(\omega) \end{aligned}$$

where

$$H(\omega) = \sum_{n=0}^{p} h_n \exp[-jn\omega], -\pi < \omega < \pi$$

So, the transfer function of the optimum non causal filter is

$$H(\omega) = \frac{\exp[j\omega\lambda]\ H(\omega)^* S_S(\omega)}{|H(\omega)|^2 S_S(\omega) + S_N(\omega)}$$

For $\lambda = 0$ we have

$$H(\omega) = \frac{H(\omega)^* S_S(\omega)}{|H(\omega)|^2 S_S(\omega) + S_N(\omega)} = \frac{1}{H(\omega)} \frac{|H(\omega)|^2}{|H(\omega)|^2 + \frac{S_N(\omega)}{S_S(\omega)}}$$

9.6 Causal Wiener-Kolmogorov

To develop the causal Wiener-Kolmogorov filtering, let note by $\widetilde{X}_t$ the non causal Wiener-Kolmogorov solution and by $\widehat{X}_t$ the causal one, *i.e.*

$$\begin{aligned} \widehat{X}_t &= \sum_{n=-\infty}^{\infty} h_{t-n} Y_n \\ \widetilde{X}_t &= \sum_{n=-\infty}^{\infty} \tilde{h}_{t-n} Y_n \end{aligned}$$

Note also that $\mathcal{H}_{-\infty}^{t}$ is a subset of $\mathcal{H}_{-\infty}^{\infty}$. So, if the solution to the noncausal Wiener-Kolmogorov problem happens to be causal, it also solves the causal Wiener-Kolmogorov problem. But unfortunately, this is not the case excepted very special cases. However, there is surely a relation between these two solutions.

To obtain this relation we start by writing

$$(X_t - \widehat{X}_t) = (\widetilde{X}_t - \widehat{X}_t) + (X_t - \widetilde{X}_t)$$

So, for any $Z \in \mathcal{H}_{-\infty}^{t}$ we have

$$\mathrm{E}\left[(X_t - \widehat{X}_t)Z\right] = \mathrm{E}\left[(\widetilde{X}_t - \widehat{X}_t)Z\right] + \mathrm{E}\left[(X_t - \widetilde{X}_t)Z\right]$$

The left hand term $\mathrm{E}\left[(X_t - \widehat{X}_t)Z\right]$ is zero due to the orthogonality principle applied to $\widehat{X}_t$. The second right hand term $\mathrm{E}\left[(X_t - \widetilde{X}_t)Z\right]$ is zero due to the orthogonality principle applied to $\widetilde{X}_t$. So we have

$$\mathrm{E}\left[(\widetilde{X}_t - \widehat{X}_t)Z\right] = 0, \quad \forall Z \in \mathcal{H}_{-\infty}^{t}$$

which means that $\widehat{X}_t$ is the MMSE estimate of $\widetilde{X}_t$ among all estimates in $\mathcal{H}_{-\infty}^{t}$. In other words, $\widehat{X}_t$ which is the projection of X_t on $\mathcal{H}_{-\infty}^{t}$ can be obtained by first projecting X_t on $\mathcal{H}_{-\infty}^{\infty}$ to get $\widetilde{X}_t$ and then projecting $\widetilde{X}_t$ onto $\mathcal{H}_{-\infty}^{t}$.

Now, let define

$$\bar{X}_t \stackrel{\text{def}}{=} \sum_{n=-\infty}^{t} \tilde{h}_{t-n} Y_n$$

and consider the error

$$\widetilde{X}_t - \bar{X}_t = \sum_{n=-\infty}^{\infty} \tilde{h}_{t-n} Y_n - \sum_{n=-\infty}^{t} \tilde{h}_{t-n} Y_n = \sum_{n=t+1}^{\infty} \tilde{h}_{t-n} Y_n$$

If this error could be orthogonal to Y_m for all $m \leq t$, we could consider $\bar{X}_t$ as the projection of $\widetilde{X}_t$ on $\mathcal{H}_{-\infty}^{t}$. But this is not the case in general. This could be the case if $\{Y_n\}_{n=-\infty}^{\infty}$ was a sequence of uncorrelated random variables, because in that case we would have

$$\begin{aligned}
\mathrm{E}\left[(\widetilde{X}_t - \bar{X}_t)Y_m\right] &= \mathrm{E}\left[\left(\sum_{n=t+1}^{\infty} \tilde{h}_{t-n} Y_n\right) Y_m\right] \\
&= \sum_{n=t+1}^{\infty} \tilde{h}_{t-n} \mathrm{E}\left[Y_n, Y_m\right] \\
&= \sigma^2 \sum_{n=t+1}^{\infty} \tilde{h}_{t-n} \delta_{n,m} = 0, \quad \forall\, m \leq t
\end{aligned}$$

where $\delta_{n,m}$ is the Kronecker delta ($\delta_{n,m} = 1$ if $n = m$ and $\delta_{n,m} = 0$ if $n \neq m$ and where $\sigma^2 = \mathrm{E}\left[Y_n^2\right]$

From the above discussion we see that, if we transform first the data $\{Y_n\}_{n=-\infty}^{\infty}$ into an equivalent w.s.s. and uncorrelated sequence $\{Z_n\}_{n=-\infty}^{\infty}$, then the causal Wiener-Kolmogorov filter coefficients can be obtained by simple truncation on the non causal Wiener-Kolmogorov filter, *i.e.*

$$\widehat{X}_t = \sum_{n=-\infty}^{t} \widehat{h}_{t-n} Z_n$$

where

$$\widehat{h}_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{XZ}(\omega) \exp\left[jn\omega\right] \mathrm{d}\omega = C_{XZ}(n), \quad n \geq 0$$

where $S_{XZ}(\omega)$ and $C_{XZ}(n)$ are, respectively, the cross spectrum and the cross covariance of the sequences $\{X_n\}_{n=-\infty}^{\infty}$ and $\{Z_n\}_{n=-\infty}^{\infty}$.

Now the aim is to obtain $\{Z_n\}_{n=-\infty}^{\infty}$ from $\{Y_n\}_{n=-\infty}^{\infty}$. The analyse of the one step prediction problem in previous section in this chapter gives us the solution. If we note by $\widehat{Y}_{t|t-1}$ the one step prediction of Y_t from the data $\{Y_n\}_{n=-\infty}^{t-1}$ and by $\sigma_t^2 = \mathrm{E}\left[(Y_t - \widehat{Y}_{t|t-1})\right]$, then define

$$Z_n = \frac{Y_n - \widehat{Y}_{n|n-1}}{\sigma_n}, \quad n \in \mathbf{Z}$$

we can verify that $\mathrm{E}[Z_n] = 0$, $\mathrm{E}[Z_n^2] = 1$ and $\mathrm{Cov}\{Z_n, Z_m\} = 0$. So, Z_n has all the necessary properties that we need. Now, still we have to show that $\{Z_n\}_{n=-\infty}^{\infty}$ is equivalent to $\{Y_n\}_{n=-\infty}^{\infty}$ for the purpose of MMSE. To do this we need the result of the following theorem.

Theorem 1 (Spectral Factorization) *Assume Y_n has a power spectral density satisfying the Paley-Wiener condition*

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln S_Y(\omega)\, d\omega > -\infty.$$

Then

$$S_Y(\omega) = S_Y^-(\omega)\, S_Y^+(\omega), \quad -\pi < \omega < \pi \tag{9.27}$$

where

$$S_Y(\omega) = |S_Y^-(\omega)|^2 = |S_Y^+(\omega)|^2 \tag{9.28}$$

and

$$\begin{aligned}
\frac{1}{2\pi} \int_{-\pi}^{\pi} S_Y^-(\omega) \exp\left[jn\omega\right] d\omega &= 0 \quad n < 0 \\
\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{S_Y^-(\omega)} \exp\left[jn\omega\right] d\omega &= 0 \quad n < 0 \\
\frac{1}{2\pi} \int_{-\pi}^{\pi} S_Y^+(\omega) \exp\left[jn\omega\right] d\omega &= 0 \quad n > 0 \\
\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{S_Y^+(\omega)} \exp\left[jn\omega\right] d\omega &= 0 \quad n > 0
\end{aligned}$$

Now consider the time invariant filter with $H(\omega) = \frac{1}{S_Y^+(\omega)}$. This filter, by definition, is causal. The output of this filter with the wide sense stationnary input sequence Y_n will be another wide sense stationnary sequence Z_n with

$$S_Z(\omega) = \left|\frac{1}{S_Y^-(\omega)}\right|^2 S_Y(\omega) = 1, \quad -\pi < \omega < \pi \tag{9.29}$$

Since $S_Z(\omega) = 1$ corresponds to a white sequence, the filter is called a *whitenning filter*.

Note also that the input sequence Y_n can also be obtained causally from the output Z_n by

$$Y_t = \sum_{n=-\infty}^{\infty} f_{t-n} Z_n, \quad t \in \mathbf{Z} \tag{9.30}$$

where

$$f_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_Y^+(\omega) \exp[jn\omega]\, \mathrm{d}\omega, \quad n \geq 0$$

Now consider the λ-step prediction of Y_n from Z_n:

$$\begin{aligned} Y_{t+\lambda} &= \sum_{n=-\infty}^{t+\lambda} f_{t+\lambda-n} Z_n \\ &= \sum_{n=-\infty}^{t} f_{t+\lambda-n} Z_n + \sum_{n=t+1}^{t} f_{t+\lambda-n} Z_n \end{aligned}$$

But Z_n is white, so $Z_{n+1}, \ldots, Z_{t+\lambda}$ are orthogonal to $\{Z_n\}_{n=-\infty}^{t}$. So the best estimate $\widehat{Y}_{t+\lambda}$ of $Y_{t+\lambda}$ from $\{Y_n\}_{n=-\infty}^{t}$ is

$$\widehat{Y}_{t+\lambda} = \sum_{n=-\infty}^{t} f_{t+\lambda-n} Z_n$$

Combining the whitening filter and this prediction we obtain

$$\cdots, Y_{t-2}, Y_{t-1}, Y_t \longrightarrow \boxed{\frac{1}{S_Y^+(\omega)}} \longrightarrow Z_n \longrightarrow \boxed{\Big[\exp[j\omega\lambda]\, S_Y^+(\omega)\Big]_+} \longrightarrow \widehat{Y}_{t+\lambda}$$

where the operator $[H(\omega)]_+$ is defined as

$$[H(\omega)]_+ = \sum_{n=0}^{\infty} h_n \exp[-j\omega n]$$

where

$$h_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} H(\omega) \exp[jn\omega]\, \mathrm{d}\omega$$

Now we obtain a procedure for causally prewhitenning a stationary sequence of observations. Thus the solution of the general causal Wiener-Kolmogorov problem follows immediately. Assuming that $\{Y_n\}_{n=-\infty}^{\infty}$ satisfies the Paley-Winner condition, it can be transformed equivalently into $\{Z_n\}_{n=-\infty}^{t}$ by passing it through the causal filter $1/S_Y^+(\omega)$

$$Y_n \longrightarrow \boxed{\frac{1}{S_Y^+(\omega)}} \longrightarrow Z_n$$

Then we need to find the cross spectrum $S_{XZ}(\omega)$ and pass $\{Z_n\}_{n=-\infty}^{t}$ through the causal filter $[S_{XZ}(\omega)]_+$ to obtain the required result.

$$Z_n \longrightarrow \boxed{S_{XZ}^+(\omega)} \longrightarrow \widehat{Y}_t$$

Knowing that

$$S_{XZ}(\omega) = \frac{S_{XY}(\omega)}{[S_Y^+(\omega)]^*} = \frac{S_{XY}(\omega)}{S_Y^-(\omega)}$$

we obtain

$$Y_n \longrightarrow \boxed{\frac{1}{S_Y^+(\omega)}} \longrightarrow Z_n \longrightarrow \boxed{\left[\frac{S_{XY}(\omega)}{S_Y^-(\omega)}\right]_+} \longrightarrow \widehat{Y}_t$$

It is interesting to compare this filter with the non causal Wiener-Kolmogorov filter

$$\frac{S_{XY}(\omega)}{S_Y(\omega)} = \frac{1}{S_Y^+(\omega)}\left[\frac{S_{XY}(\omega)}{S_Y^-(\omega)}\right]$$

The following diagrams summarize this comparison:

$$Y_n \longrightarrow \boxed{\frac{1}{S_Y^+(\omega)}} \longrightarrow \boxed{\left[\frac{S_{XY}(\omega)}{S_Y^-(\omega)}\right]_+} \longrightarrow \widehat{Y}_{t+\lambda} \qquad Y_n \longrightarrow \boxed{\frac{1}{S_Y^+(\omega)}} \longrightarrow \boxed{\frac{S_{XY}(\omega)}{S_Y^-(\omega)}} \longrightarrow \widehat{Y}_{t+\lambda}$$

Causal Non Causal

9.7 Rational spectra

Definition 1 (Rational spectra) *$S_y(\omega)$ is said rational if it can be written as the ratio of two real trigonometric polynomials*

$$S_y(\omega) = \frac{n_0 + 2\sum_{k=1}^{p} n_k \cos k\omega}{d_0 + 2\sum_{k=1}^{m} d_k \cos k\omega}, \quad n_k, d_k \in \mathbb{R} \tag{9.31}$$

Using the relation $\cos k\omega = \exp[jk\omega] + \exp[-jk\omega]$ we deduce

$$S_y(\omega) = \frac{N(\exp[jk\omega])}{D(\exp[jk\omega])}.$$

Noting $z = \exp[jk\omega]$ we have

$$\begin{aligned} N(z) &= \sum_{k=-p}^{p} n_{|k|} z^{-k} \\ D(z) &= \sum_{k=-p}^{p} d_{|k|} z^{-k} \end{aligned}$$

$z^p N(z)$ is a $(2p)$-th order polynomial. So we can write

$$N(z) = n_p\, z^{-p} \prod_{k=1}^{2p} (z - z_k)$$

Since $N(z) = N(1/z)$ the roots z_k are reciprocal pairs. If we order them in such a way that $|z_1| > |z_2| > \cdots > |z_{2p}|$ we have

$$|z_{2p}| = \frac{1}{|z_1|}, \quad |z_{2p-1}| = \frac{1}{|z_2|}, \cdots, |z_{p+1}| = \frac{1}{|z_p|}.$$

We deduce that all the roots are outside or over the unit circle. Due to the reciprocity of the roots we can write

$$N(z) = B(z)\, B(1/Z) \tag{9.32}$$

where

$$B(z) = \sqrt{(-1)^p np / \prod_{k=1}^{p} z_k} \prod_{k=1}^{p} (z^{-1} - z_k) \tag{9.33}$$

So $B(z)$ is a polynomial of degree p and can be extended as

$$B(z) = \sum_{k=0}^{p} b_k\, z^{-k} \tag{9.34}$$

Similarly, we can do exactly the same analysis for the denominator $D(z)$:

$$D(z) = A(z)\, A(1/Z) \tag{9.35}$$

where

$$A(z) = \sum_{k=0}^{m} a_k\, z^{-k} \tag{9.36}$$

Putting these together we have

$$S_y(\omega) = \frac{B\left(\exp\left[jk\omega\right]\right)\, B\left(\exp\left[-jk\omega\right]\right)}{A\left(\exp\left[jk\omega\right]\right)\, A\left(\exp\left[-jk\omega\right]\right)} = \frac{B\left(\exp\left[jk\omega\right]\right)}{A\left(\exp\left[jk\omega\right]\right)} \frac{B\left(\exp\left[-jk\omega\right]\right)}{A\left(\exp\left[-jk\omega\right]\right)}$$

Assuming that none of the roots of B or A is on the unit circle $|z| = 1$ we have

$$\begin{aligned} S_y^+(\omega) &= \frac{B\left(\exp\left[jk\omega\right]\right)}{A\left(\exp\left[jk\omega\right]\right)} \\ S_y^-(\omega) &= \frac{B\left(\exp\left[-jk\omega\right]\right)}{A\left(\exp\left[-jk\omega\right]\right)} \end{aligned}$$

Now consider the whitenning filter of the last section and assume that the power spectrum of the data $S_Y(\omega)$ is a rational fraction.

$$Y_n \longrightarrow \boxed{\frac{1}{S_Y^+(\omega)}} \longrightarrow Z_n \longrightarrow Y_n \longrightarrow \boxed{\frac{A(z)}{B(z)} = \frac{\sum_{k=0}^{m} a_k\, z^{-k}}{\sum_{k=0}^{p} b_k\, z^{-k}}} \longrightarrow Z_n$$

Then, we can see easily that we have

$$\sum_{k=0}^{m} a_k\, Y_{n-k} = \sum_{k=0}^{p} b_k\, Z_{n-k} \tag{9.37}$$

We can rewrite this equation in two other equivalent forms

$$\begin{aligned} b_0 z_n &= -\sum_{k=1}^{p} b_k\, Z_{n-k} + \sum_{k=0}^{m} a_k\, Y_{n-k} \\ a_0 Y_n &= -\sum_{k=1}^{m} a_k\, Y_{n-k} + \sum_{k=0}^{p} b_k\, Z_{n-k} \end{aligned}$$

Autoregressive, moving Average (ARMA) sequence of order (m, p). For $p = 0$, we have an Autoregressive (AR) and for $m = 0$ we have a moving average (MA) sequence.

Example 3 (Wide-Sense Markov sequences) *A simple and useful model for the correlation structure of a stationary random sequence is the so-called* wide-sense Markov model*:*

$$C_Y(n) = \sigma^2 r^{|n|}, \quad n \in \boldsymbol{Z} \tag{9.38}$$

where $|r| < 1$. *The power spectrum of such a sequence is*

$$S_Y(\omega) = \frac{\sigma^2(1 - r^2)}{1 - 2r\cos(\omega) + r^2}$$

which is a rational fraction, and we can see easily that we can write it

$$S_Y(\omega) = \frac{\sigma^2(1 - r^2)}{(1 - r\exp[-j\omega])\,(1 - r\exp[+j\omega])} = \frac{1}{A(\exp[-j\omega])\,A(\exp[+j\omega])} = \frac{1}{A(Z)\,A(z^{-1})}$$

where

$$A(z) = a_0 + a_1 z^{-1}$$

with $a_0 = \sqrt{\sigma^2(1 - r^2)}$, $a_1 = -r\, a_0$.

We can conclude here that a wide-sense Markov sequence with the covariance structure (9.38) is an AR(1) sequence.

Example 4 (Prediction of a Wide-Sense Markov sequences) *Consider now the prediction problem where we wish to predict* $Y_{n+\lambda}$ *from the sequence* $Y_{k}{}_{k=-\infty}^{n}$. *Using the relations we obtained in the last section, this can be done through a causal filter whose transfer function is*

$$Y_n \longrightarrow \boxed{H(\omega) = \frac{1}{S_Y^+(\omega)} \left[\frac{S_{XY}(\omega)}{S_Y^-(\omega)}\right]_+} \longrightarrow \widehat{Y}_{t+\lambda}$$

Here we have

$$H(\omega) = A(\exp[j\omega]) \left[\frac{\exp[j\omega\lambda]}{A(\exp[j\omega])}\right]_+$$

Using the following geometric series relations

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1 - x}, \quad \textit{and} \quad \sum_{k=1}^{\infty} x^k = \frac{x}{1 - x}, \quad |x| < 1$$

we obtain easily

$$\frac{1}{A(z)} = \frac{1}{a_0} \sum_{k=0}^{\infty} r^n \, z^{-1}$$

and

$$\begin{aligned}
\left[\frac{\exp\,[j\omega\lambda]}{A(\exp\,[j\omega])}\right]_+ &= \left[\frac{1}{a_0} \sum_{n=0}^{\infty} r^n \, \exp\,[-j\omega(n-\lambda)]\right]_+ \\
&= \frac{1}{a_0} \sum_{n=\lambda}^{\infty} r^n \, \exp\,[-j\omega(n-\lambda)] \\
&= \frac{1}{a_0} \sum_{l=0}^{\infty} r^{l+\lambda} \, \exp\,[-jl\omega] \\
&= \frac{r^\lambda}{A(\exp\,[j\omega])}
\end{aligned}$$

Finally, we obtain

$$H(\omega) = A(\exp\,[j\omega]) \, \frac{r^\lambda}{A(\exp\,[j\omega])} = r^\lambda$$

which is a pure pure gain

$$\widehat{Y}_{t+\lambda} = r^\lambda \, Y_t$$

It is also easy to show that, in this case we have

$$MSE = E\left[(\widehat{Y}_{t+\lambda} - Y_{t+\lambda})^2\right] = \sigma^2(1 - r^{2\lambda})$$

which means that the prediction error increases monotonically from $\sigma^2(1 - r^{2\lambda})$ *to* σ^2 *as* λ *increases from 1 to* ∞.