Statistical and Adaptive Signal Processing [Manolakis, Ingle Kogon], 2005.

(6.4.17) in the frequency domain by using (6.4.20). Indeed, we have

$$P_o = r_y(0) - \frac{1}{2\pi}\int_{-\pi}^{\pi} H_o(e^{j\omega})R_{yx}^*(e^{j\omega})\,d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi}[R_y(e^{j\omega}) - H_o(e^{j\omega})R_{yx}^*(e^{j\omega})]\,d\omega \tag{6.4.22}$$

where $H_o(e^{j\omega})$ is the frequency response of the optimum filter. The above equation holds for any filter, FIR or IIR, as long as we use the proper limits to compute the summation in (6.4.19).

We will now obtain a formula for the MMSE that holds only for IIR filters whose impulse response extends from $-\infty$ to ∞. In this case, (6.4.16) is a convolution equation that holds for $-\infty < m < \infty$. Using the convolution theorem of the Fourier transform, we obtain

$$H_o(e^{j\omega}) = \frac{R_{yx}(e^{j\omega})}{R_x(e^{j\omega})} \tag{6.4.23}$$

which, we again stress, holds for noncausal IIR filters *only*. Substituting into (6.4.22), we obtain

$$P_o = \frac{1}{2\pi}\int_{-\pi}^{\pi}[1 - \frac{|R_{yx}(e^{j\omega})|^2}{R_y(e^{j\omega})R_x(e^{j\omega})}]R_y(e^{j\omega})\,d\omega$$

or

$$P_o = \frac{1}{2\pi}\int_{-\pi}^{\pi}[1 - \mathcal{G}_{yx}(e^{j\omega})]R_y(e^{j\omega})\,d\omega \tag{6.4.24}$$

where $\mathcal{G}_{yx}(e^{j\omega})$ is the coherence function between $x(n)$ and $y(n)$.

This important equation indicates that the performance of the optimum filter depends on the coherence between the input and desired response processes. As we recall from Section 5.4, the coherence is a measure of both the noise disturbing the observations and the relative linearity between $x(n)$ and $y(n)$. The optimum filter can reduce the MMSE at a certain band only if there is significant coherence, that is, $\mathcal{G}_{yx}(e^{j\omega}) \simeq 1$. Thus, the optimum filter $H_o(z)$ constitutes the best, in the MMSE sense, linear relationship between the stochastic processes $x(n)$ and $y(n)$. These interpretations apply to causal IIR and FIR optimum filters, even if (6.4.23) and (6.4.24) only hold approximately in these cases (see Section 6.6).

6.5 LINEAR PREDICTION

Linear prediction plays a prominent role in many theoretical, computational, and practical areas of signal processing and deals with the problem of estimating or predicting the value $x(n)$ of a signal at the time instant $n = n_0$, by using a set of other samples from the *same* signal. Although linear prediction is a subject useful in itself, its importance in signal processing is also due, as we will see later, to its use in the development of fast algorithms for optimum filtering and its relation to all-pole signal modeling.

6.5.1 Linear Signal Estimation

Suppose that we are given a set of values $x(n), x(n-1), \ldots, x(n-M)$ of a stochastic process and we wish to estimate the value of $x(n-i)$, using a linear combination of the remaining samples. The resulting estimate and the corresponding estimation error are given by

$$\hat{x}(n-i) \triangleq -\sum_{\substack{k=0\\k\neq i}}^{M} c_k^*(n)x(n-k) \tag{6.5.1}$$

and

$$e^{(i)}(n) \triangleq x(n-i) - \hat{x}(n-i) = \sum_{k=0}^{M} c_k^*(n)x(n-k) \quad \text{with } c_i(n) \triangleq 1 \tag{6.5.2}$$

where $c_k(n)$ are the coefficients of the estimator as a function of discrete-time index n. The process is illustrated in Figure 6.16.

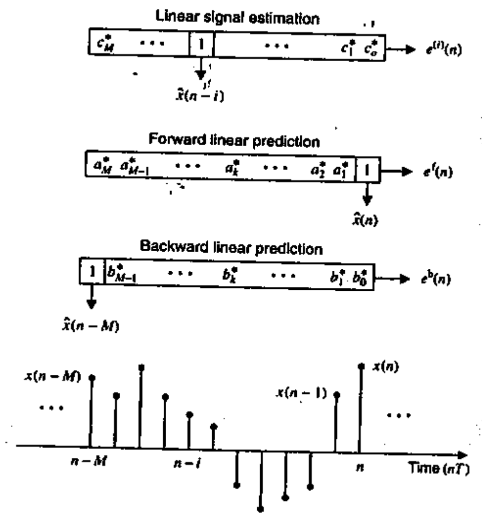

FIGURE 6.16
Illustration showing the samples, estimates, and errors used in linear signal estimation, forward linear prediction, and backward linear prediction.

To determine the MMSE signal estimator, we partition (6.5.2) as

$$\begin{aligned} e^{(i)}(n) &= \sum_{k=0}^{i-1} c_k^*(n)x(n-k) + x(n-i) + \sum_{k=i+1}^{M} c_k^*(n)x(n-k) \\ &\triangleq \mathbf{c}_1^H(n)\mathbf{x}_1(n) + x(n-i) + \mathbf{c}_2^H(n)\mathbf{x}_2(n) \\ &\triangleq [\bar{\mathbf{c}}^{(i)}(n)]^H \bar{\mathbf{x}}(n) \end{aligned} \tag{6.5.3}$$

where the partitions of the coefficient and data vectors, around the ith component, are easily defined from the context. To obtain the normal equations and the MMSE for the optimum *linear signal estimator*, we note that

$$\text{Desired response} = x(n-i) \qquad \text{data vector} = \begin{bmatrix} \mathbf{x}_1(n) \\ \mathbf{x}_2(n) \end{bmatrix}$$

Using (6.4.6) and (6.4.9) or the orthogonality principle, we have

$$\begin{bmatrix} \mathbf{R}_{11}(n) & \mathbf{R}_{12}(n) \\ \mathbf{R}_{12}^T(n) & \mathbf{R}_{22}(n) \end{bmatrix} \begin{bmatrix} \mathbf{c}_1(n) \\ \mathbf{c}_2(n) \end{bmatrix} = - \begin{bmatrix} \mathbf{r}_1(n) \\ \mathbf{r}_2(n) \end{bmatrix} \tag{6.5.4}$$

or more compactly[†]

$$\mathbf{R}^{(i)}(n)\mathbf{c}_o^{(i)}(n) = -\mathbf{d}^{(i)}(n) \tag{6.5.5}$$

and

$$P_o^{(i)}(n) = P_x(n-i) + \mathbf{r}_1^H(n)\mathbf{c}_1(n) + \mathbf{r}_2^H(n)\mathbf{c}_2(n) \tag{6.5.6}$$

where for $j, k = 1, 2$

$$\mathbf{R}_{jk}(n) \triangleq E\{\mathbf{x}_j(n)\mathbf{x}_k^H(n)\} \tag{6.5.7}$$

$$\mathbf{r}_j(n) \triangleq E\{\mathbf{x}_j(n)x^*(n-i)\} \tag{6.5.8}$$

$$P_x(n) = E\{|x(n)|^2\} \tag{6.5.9}$$

For various reasons, to be seen later, we will combine (6.5.4) and (6.5.6) into a single equation. To this end, we note that the correlation matrix of the extended vector

$$\bar{\mathbf{x}}(n) = \begin{bmatrix} \mathbf{x}_1(n) \\ x(n-i) \\ \mathbf{x}_2(n) \end{bmatrix} \tag{6.5.10}$$

can be partitioned as

$$\bar{\mathbf{R}}(n) = E\{\bar{\mathbf{x}}(n)\bar{\mathbf{x}}^H(n)\} = \begin{bmatrix} \mathbf{R}_{11}(n) & \mathbf{r}_1(n) & \mathbf{R}_{12}(n) \\ \mathbf{r}_1^H(n) & P_x(n-i) & \mathbf{r}_2^H(n) \\ \mathbf{R}_{12}^H(n) & \mathbf{r}_2(n) & \mathbf{R}_{22}(n) \end{bmatrix} \tag{6.5.11}$$

with respect to its ith row and ith column. Using (6.5.4), (6.5.6), and (6.5.11), we obtain

$$\bar{\mathbf{R}}(n)\bar{\mathbf{c}}_o^{(i)}(n) = \begin{bmatrix} \mathbf{0} \\ P_o^{(i)}(n) \\ \mathbf{0} \end{bmatrix} \leftarrow i\text{th row} \tag{6.5.12}$$

which completely determines the linear signal estimator $\mathbf{c}^{(i)}(n)$ and the MMSE $P_o^{(i)}(n)$.

If $M = 2L$ and $i = L$, we have a *symmetric linear smoother* $\bar{\mathbf{c}}(n)$ that produces an estimate of the middle sample by using the L past and the L future samples. The above formulation suggests an easy procedure for the computation of the linear signal estimator for any value of i, which is outlined in Table 6.3 and implemented by the function `[ci,Pi]=olsigest(R,i)`. We next discuss two types of linear signal estimation that are of special interest and have their own dedicated notation.

6.5.2 Forward Linear Prediction

One-step *forward linear prediction* (*FLP*) involves the estimation or prediction of the value $x(n)$ of a stochastic process by using a linear combination of the past samples $x(n-1), \ldots, x(n-M)$ (see Figure 6.16). We should stress that in signal processing applications

[†] The minus sign on the right-hand side of the normal equations is the result of arbitrarily setting the coefficient $c_i(n) \triangleq 1$.

TABLE 6.3
Steps for the computation of optimum signal estimators.

1. Determine the matrix $\bar{\mathbf{R}}(n)$ of the extended data vector $\bar{\mathbf{x}}(n)$.
2. Create the $M \times M$ submatrix $\mathbf{R}^{(i)}(n)$ of $\bar{\mathbf{R}}(n)$ by removing its ith row and its ith column.
3. Create the $M \times 1$ vector $\mathbf{d}^{(i)}(n)$ by extracting the ith column $\bar{\mathbf{d}}^{(i)}(n)$ of $\bar{\mathbf{R}}(n)$ and removing its ith element.
4. Solve the linear system $\mathbf{R}^{(i)}(n)\mathbf{c}_o^{(i)}(n) = -\mathbf{d}^{(i)}(n)$ to obtain $\mathbf{c}_o^{(i)}(n)$.
5. Compute the MMSE $P_o^{(i)}(n) = [\bar{\mathbf{d}}^{(i)}(n)]^H \bar{\mathbf{c}}_o^{(i)}(n)$.

of linear prediction, what is important is the ability to obtain a good estimate of a sample, pretending that it is unknown, instead of forecasting the future. Thus, the term *prediction* is used more with signal estimation than forecasting in mind. The forward predictor is a linear signal estimator with $i = 0$ and is denoted by

$$e^{\mathrm{f}}(n) \triangleq x(n) + \sum_{k=1}^{M} a_k^*(n)x(n-k) \tag{6.5.13}$$

$$= x(n) + \mathbf{a}^H(n)\mathbf{x}(n-1)$$

where

$$\mathbf{a}(n) \triangleq [a_1(n)\, a_2(n) \cdots a_M(n)]^T \tag{6.5.14}$$

is known as the *forward linear predictor* and $a_k(n)$ with $a_0(n) \triangleq 1$ as the FLP error filter. To obtain the normal equations and the MMSE for the optimum FLP, we note that for $i = 0$, (6.5.11) can be written as

$$\bar{\mathbf{R}}(n) = \begin{bmatrix} P_x(n) & \mathbf{r}^{\mathrm{f}H}(n) \\ \mathbf{r}^{\mathrm{f}}(n) & \mathbf{R}(n-1) \end{bmatrix} \tag{6.5.15}$$

where

$$\mathbf{R}(n) = E\{\mathbf{x}(n)\mathbf{x}^H(n)\} \tag{6.5.16}$$

and

$$\mathbf{r}^{\mathrm{f}}(n) = E\{\mathbf{x}(n-1)x^*(n)\} \tag{6.5.17}$$

Therefore, (6.5.5) and (6.5.6) give

$$\mathbf{R}(n-1)\mathbf{a}_o(n) = -\mathbf{r}^{\mathrm{f}}(n) \tag{6.5.18}$$

and

$$P_o^{\mathrm{f}}(n) = P_x(n) + \mathbf{r}^{\mathrm{f}H}(n)\mathbf{a}_o(n) \tag{6.5.19}$$

or

$$\bar{\mathbf{R}}(n)\begin{bmatrix} 1 \\ \mathbf{a}_o(n) \end{bmatrix} = \begin{bmatrix} P_o^{\mathrm{f}}(n) \\ \mathbf{0} \end{bmatrix} \tag{6.5.20}$$

which completely specifies the FLP parameters.

6.5.3 Backward Linear Prediction

In this case, we want to estimate the sample $x(n-M)$ in terms of the future samples $x(n), x(n-1), \ldots, x(n-M+1)$ (see Figure 6.16). The term *backward linear prediction* (*BLP*) is not accurate but is used since it is an established convention. A more appropriate name might be *postdiction* or *hindsight*. The BLP is basically a linear signal estimator with $i = M$ and is denoted by

$$e^{\mathrm{b}}(n) \triangleq \sum_{k=0}^{M-1} b_k^*(n)x(n-k) + x(n-M) \tag{6.5.21}$$

$$= \mathbf{b}^H(n)\mathbf{x}(n) + x(n-M)$$

where

$$\mathbf{b}(n) \triangleq [b_0(n)\, b_1(n) \cdots b_{M-1}(n)]^T \tag{6.5.22}$$

is the BLP and $b_k(n)$ with $b_M(n) \triangleq 1$ is the *backward prediction error filter (BPEF)*. For $i = M$, (6.5.11) gives

$$\bar{\mathbf{R}}(n) = \begin{bmatrix} \mathbf{R}(n) & \mathbf{r}^b(n) \\ \mathbf{r}^{bH}(n) & P_x(n-M) \end{bmatrix} \quad (6.5.23)$$

where

$$\mathbf{r}^b(n) \triangleq E\{\mathbf{x}(n)x^*(n-M)\} \quad (6.5.24)$$

The optimum backward linear predictor is specified by

$$\mathbf{R}(n)\mathbf{b}_o(n) = -\mathbf{r}^b(n) \quad (6.5.25)$$

and the MMSE is

$$P_o^b(n) = P_x(n-M) + \mathbf{r}^{bH}(n)\mathbf{b}_o(n) \quad (6.5.26)$$

and can be put in a single equation as

$$\bar{\mathbf{R}}(n)\begin{bmatrix} \mathbf{b}_o(n) \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ P_o^b(n) \end{bmatrix} \quad (6.5.27)$$

In Table 6.4, we summarize the definitions and design equations for optimum FIR filtering and prediction. Using the entries in this table, we can easily obtain the normal equations and the MMSE for the FLP and BLP from those of the optimum filter.

TABLE 6.4
Summary of the design equations for optimum FIR filtering and prediction.

	Optimum filter	FLP	BLP
Input data vector	$\mathbf{x}(n)$	$\mathbf{x}(n-1)$	$\mathbf{x}(n)$
Desired response	$y(n)$	$x(n)$	$x(n-M)$
Coefficient vector	$\mathbf{h}(n)$	$\mathbf{a}(n)$	$\mathbf{b}(n)$
Estimation error	$e(n) = y(n) - \mathbf{c}^H(n)\mathbf{x}(n)$	$e^f(n) = x(n) + \mathbf{a}^H(n)\mathbf{x}(n-1)$	$e^b(n) = x(n-M) + \mathbf{b}^H(n)\mathbf{x}(n)$
Normal equations	$\mathbf{R}(n)\mathbf{c}_o(n) = \mathbf{d}(n)$	$\mathbf{R}(n-1)\mathbf{a}_o(n) = -\mathbf{r}^f(n)$	$\mathbf{R}(n)\mathbf{b}_o(n) = -\mathbf{r}^b(n)$
MMSE	$P_o^c(n) = P_y(n) - \mathbf{c}_o^H(n)\mathbf{d}(n)$	$P_o^f(n) = P_x(n) + \mathbf{a}^H(n)\mathbf{r}_o^f(n)$	$P_o^b(n) = P_x(n-M) + \mathbf{b}^H(n)\mathbf{r}_o^b(n)$
Required moments	$\mathbf{R}(n) = E\{\mathbf{x}(n)\mathbf{x}^H(n)\}$ $\mathbf{d}(n) = E\{\mathbf{x}(n)y^*(n)\}$	$\mathbf{r}^f(n) = E\{\mathbf{x}(n-1)x^*(n)\}$	$\mathbf{r}^b(n) = E\{\mathbf{x}(n)x^*(n-M)\}$
Stationary processes	$\mathbf{R}\mathbf{c}_o = \mathbf{d}$, $\mathbf{R}$ is Toeplitz	$\mathbf{R}\mathbf{a}_o = -\mathbf{r}^*$	$\mathbf{R}\mathbf{b}_o = -\mathbf{J}\mathbf{r} \Rightarrow \mathbf{b}_o = \mathbf{J}\mathbf{a}_o^*$

6.5.4 Stationary Processes

If the process $x(n)$ is stationary, then the correlation matrix $\bar{\mathbf{R}}(n)$ does *not* depend on the time n and it is Toeplitz

$$\bar{\mathbf{R}} = \begin{bmatrix} r(0) & r(1) & \cdots & r(M) \\ r^*(1) & r(0) & \cdots & r(M-1) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(M) & r^*(M-1) & \cdots & r(0) \end{bmatrix} \quad (6.5.28)$$

Therefore, all the resulting linear MMSE signal estimators are *time-invariant*. If we define the correlation vector

$$\mathbf{r} \triangleq [r(1)\, r(2) \cdots r(M)]^T \quad (6.5.29)$$

where $r(l) = E\{x(n)x^*(n-l)\}$, we can easily see that the cross-correlation vectors for the FLP and the BLP are

$$\mathbf{r}^f = E\{\mathbf{x}(n-1)x^*(n)\} = \mathbf{r}^* \quad (6.5.30)$$

and

$$\mathbf{r}^b = E\{\mathbf{x}(n)x^*(n-M)\} = \mathbf{J}\mathbf{r} \quad (6.5.31)$$

where

$$\mathbf{J} = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{bmatrix}, \qquad \mathbf{J}^H\mathbf{J} = \mathbf{J}\mathbf{J}^H = \mathbf{I} \quad (6.5.32)$$

is the exchange matrix that simply reverses the order of the vector elements. Therefore,

$$\mathbf{R}\mathbf{a}_o = -\mathbf{r}^* \quad (6.5.33)$$

$$P_o^f = r(0) + \mathbf{r}^H\mathbf{a}_o \quad (6.5.34)$$

$$\mathbf{R}\mathbf{b}_o = -\mathbf{J}\mathbf{r} \quad (6.5.35)$$

$$P_o^b = r(0) + \mathbf{r}^H\mathbf{J}\mathbf{b}_o \quad (6.5.36)$$

where the Toeplitz matrix $\mathbf{R}$ is obtained from $\bar{\mathbf{R}}$ by deleting the last column and row. Using the centrosymmetry property of symmetric Toeplitz matrices

$$\mathbf{R}\mathbf{J} = \mathbf{J}\mathbf{R}^* \quad (6.5.37)$$

and (6.5.33), we have

$$\mathbf{J}\mathbf{R}^*\mathbf{a}_o^* = -\mathbf{J}\mathbf{r} \quad \text{or} \quad \mathbf{R}\mathbf{J}\mathbf{a}_o^* = -\mathbf{J}\mathbf{r} \quad (6.5.38)$$

Comparing the last equation with (6.5.35), we have

$$\mathbf{b}_o = \mathbf{J}\mathbf{a}_o^* \quad (6.5.39)$$

that is, the BLP coefficient vector is the reverse of the conjugated FLP coefficient vector. Furthermore, from (6.5.34), (6.5.36), and (6.5.39), we have

$$P_o \triangleq P_o^f = P_o^b \quad (6.5.40)$$

that is, the forward and backward prediction error powers are equal.

This remarkable symmetry between the MMSE forward and backward linear predictors holds for stationary processes but disappears for nonstationary processes. Also, we do not have such a symmetry if a criterion other than the MMSE is used and the process to be predicted is non-Gaussian (Weiss 1975; Lawrence 1991).

EXAMPLE 6.5.1. To illustrate the basic ideas in FLP, BLP, and linear smoothing, we consider the second-order estimators for stationary processes.

The augmented equations for the first-order FLP are ($r(0)$ is always real)

$$\begin{bmatrix} r(0) & r(1) \\ r^*(1) & r(0) \end{bmatrix}\begin{bmatrix} a_0^{(1)} \\ a_1^{(1)} \end{bmatrix} = \begin{bmatrix} P_1^f \\ 0 \end{bmatrix}$$

and they can be solved by using Cramer's rule. Indeed, we obtain

$$a_0^{(1)} = \frac{\det\begin{bmatrix} P_1^f & r(1) \\ 0 & r(0) \end{bmatrix}}{\det \mathbf{R}_2} = \frac{r(0)P_1^f}{\det \mathbf{R}_2} = 1 \Rightarrow P_1^f = \frac{\det \mathbf{R}_2}{\det \mathbf{R}_1} = \frac{r^2(0) - |r(1)|^2}{r(0)}$$

and
$$a_1^{(1)} = \frac{\det\begin{bmatrix} r(0) & P_1^f \\ r^*(1) & 0 \end{bmatrix}}{\det \mathbf{R}_2} = \frac{-P_1^f r^*(1)}{\det \mathbf{R}_2} = -\frac{r^*(1)}{r(0)}$$

for the MMSE and the FLP. For the second-order case we have

$$\begin{bmatrix} r(0) & r(1) & r(2) \\ r^*(1) & r(0) & r(1) \\ r^*(2) & r^*(1) & r(0) \end{bmatrix} \begin{bmatrix} a_0^{(2)} \\ a_1^{(2)} \\ a_2^{(2)} \end{bmatrix} = \begin{bmatrix} P_2^f \\ 0 \\ 0 \end{bmatrix}$$

whose solution is

$$a_0^{(2)} = \frac{P_2^f \det \mathbf{R}_2}{\det \mathbf{R}_3} = 1 \Rightarrow P_2^f = \frac{\det \mathbf{R}_3}{\det \mathbf{R}_2}$$

$$a_1^{(2)} = \frac{-P_2^f \det\begin{bmatrix} r^*(1) & r(1) \\ r^*(2) & r(0) \end{bmatrix}}{\det \mathbf{R}_3} = \frac{-\det\begin{bmatrix} r^*(1) & r(1) \\ r^*(2) & r(0) \end{bmatrix}}{\det \mathbf{R}_2} = \frac{r(1)r^*(2) - r(0)r^*(1)}{r^2(0) - |r(1)|^2}$$

and

$$a_2^{(2)} = \frac{P_2^f \det\begin{bmatrix} r^*(1) & r(0) \\ r^*(2) & r^*(1) \end{bmatrix}}{\det \mathbf{R}_3} = \frac{\det\begin{bmatrix} r^*(1) & r(0) \\ r^*(2) & r^*(1) \end{bmatrix}}{\det \mathbf{R}_2} = \frac{[r^*(1)]^2 - r(0)r^*(2)}{r^2(0) - |r(1)|^2}$$

Similarly, for the BLP

$$\begin{bmatrix} r(0) & r(1) \\ r^*(1) & r(0) \end{bmatrix} \begin{bmatrix} b_0^{(1)} \\ b_1^{(1)} \end{bmatrix} = \begin{bmatrix} 0 \\ P_1^b \end{bmatrix}$$

where $b_1^{(1)} = 1$, we obtain

$$P_1^b = \frac{\det \mathbf{R}_2}{\det \mathbf{R}_1} \qquad \text{and} \qquad b_0^{(1)} = -\frac{r(1)}{r(0)}$$

$$\begin{bmatrix} r(0) & r(1) & r(2) \\ r^*(1) & r(0) & r(1) \\ r^*(2) & r^*(1) & r(0) \end{bmatrix} \begin{bmatrix} b_0^{(2)} \\ b_1^{(2)} \\ b_2^{(2)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ P_2^b \end{bmatrix}$$

$$P_2^b = \frac{\det \mathbf{R}_3}{\det \mathbf{R}_2} \qquad b_1^{(2)} = \frac{r^*(1)r(2) - r(0)r(1)}{r^2(0) - |r(1)|^2} \qquad b_0^{(2)} = \frac{r^2(1) - r(0)r(2)}{r^2(0) - |r(1)|^2}$$

We note that
$$P_1^f = P_1^b \qquad a_1^{(1)} = b_0^{(1)*}$$

and
$$P_2^f = P_2^b \qquad a_1^{(2)} = b_1^{(2)*} \qquad a_2^{(2)} = b_0^{(2)*}$$

which is a result of the stationarity of $x(n)$ or equivalently of the Toeplitz structure of $\mathbf{R}_m$.

For the linear signal estimator, we have

$$\begin{bmatrix} r(0) & r(1) & r(2) \\ r^*(1) & r(0) & r(1) \\ r^*(2) & r^*(1) & r(0) \end{bmatrix} \begin{bmatrix} c_0^{(2)} \\ c_1^{(2)} \\ c_2^{(2)} \end{bmatrix} = \begin{bmatrix} 0 \\ P_2 \\ 0 \end{bmatrix}$$

with $c_1^{(2)} = 1$. Using Cramer's rule, we obtain

$$P_2 = \frac{\det \mathbf{R}_3}{\det \mathbf{R}_3^{(2)}}$$

$$c_0^{(2)} = \frac{-P_2 \det\begin{bmatrix} r(1) & r(2) \\ r^*(1) & r(0) \end{bmatrix}}{\det \mathbf{R}_3} = -\frac{\det\begin{bmatrix} r(1) & r(2) \\ r^*(1) & r(0) \end{bmatrix}}{\det \mathbf{R}_3^{(2)}} = \frac{r^*(1)r(2) - r(0)r(1)}{r^2(0) - |r(1)|^2}$$

$$c_2^{(2)} = \frac{-P_2 \det\begin{bmatrix} r(0) & r(1) \\ r^*(2) & r^*(1) \end{bmatrix}}{\det \mathbf{R}_3} = -\frac{\det\begin{bmatrix} r(0) & r(1) \\ r^*(2) & r^*(1) \end{bmatrix}}{\det \mathbf{R}_3^{(2)}} = \frac{r(1)r^*(2) - r(0)r^*(1)}{r^2(0) - |r(1)|^2}$$

from which we see that $c_0^{(2)} = c_2^{(2)*}$; that is, we have a linear phase estimator.

6.5.5 Properties

Linear signal estimators and predictors have some interesting properties that we discuss next.

PROPERTY 6.5.1. If the process $x(n)$ is stationary, then the symmetric, linear smoother has linear phase.

Proof. Using the centrosymmetry property $\bar{\mathbf{R}}\mathbf{J} = \mathbf{J}\bar{\mathbf{R}}^*$ and (6.5.12) for $M = 2L$, $i = L$, we obtain

$$\bar{\mathbf{c}} = \mathbf{J}\bar{\mathbf{c}}^* \tag{6.5.41}$$

that is, the symmetric, linear smoother has even symmetry and, therefore, has linear phase (see Problem 6.12).

PROPERTY 6.5.2. If the process $x(n)$ is stationary, the forward prediction error filter (PEF) $1, a_1, a_2, \ldots, a_M$ is minimum-phase and the backward PEF $b_0, b_1, \ldots, b_{M-1}, 1$ is maximum-phase.

Proof. The system function of the Mth-order forward PEF can be factored as

$$A(z) = 1 + \sum_{k=1}^{M} a_k^* z^{-k} = G(z)(1 - qz^{-1})$$

where q is a zero of $A(z)$ and

$$G(z) = 1 + \sum_{k=1}^{M-1} g_k z^{-k}$$

is an $(M-1)$st-order filter. The filter $A(z)$ can be implemented as the cascade connection of the filters $G(z)$ and $1 - qz^{-1}$ (see Figure 6.17). The output $s(n)$ of $G(z)$ is

$$s(n) = x(n) + g_1 x(n-1) + \cdots + g_{M-1} x(n-M+1)$$

and it is easy to see that

$$E\{s(n-1)e^{f*}(n)\} = 0 \tag{6.5.42}$$

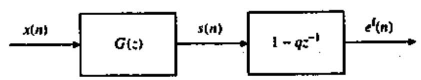

FIGURE 6.17
The prediction error filter with one zero factored out.

because $E\{x(n-k)e^{f*}(n)\} = 0$ for $1 \le k \le M$. Since the output of the second filter can be expressed as

$$e^f(n) = s(n) - qs(n-1)$$

we have

$$E\{s(n-1)e^{f*}(n)\} = E\{s(n-1)s^*(n)\} - q^*E\{s(n-1)s^*(n-1)\} = 0$$

which implies that

$$q = \frac{r_s(-1)}{r_s(0)} \Rightarrow |q| \le 1$$

because q is equal to the normalized autocorrelation of $s(n)$. If the process $x(n)$ is not predictable, that is, $E\{|e^f(n)|^2\} \ne 0$, we have

$$\begin{aligned} E\{|e^f(n)|^2\} &= E\{e^f(n)[s^*(n) - q^*s^*(n-1)]\} \\ &= E\{e^f(n)s^*(n)\} \qquad \text{due to (6.5.42)} \\ &= E\{[s(n) - qs(n-1)]s^*(n)\} \\ &= r_s(0)(1 - |q|^2) \ne 0 \end{aligned}$$

which implies that

$$|q| < 1$$

that is, the zero q of the forward PEF filter is strictly inside the unit circle. Repeating this process, we can show that all zeros of $A(z)$ are inside the unit circle; that is, $A(z)$ is minimum-phase. This proof was presented in Vaidyanathan et al. (1996). The property $\mathbf{b} = \mathbf{J}\mathbf{a}^*$ is equivalent to

$$B(z) = z^{-M}A^*\left(\frac{1}{z^*}\right)$$

which implies that $B(z)$ is a maximum-phase filter (see Section 2.4).

PROPERTY 6.5.3. The forward and backward prediction error filters can be expressed in terms of the eigenvalues $\bar{\lambda}_i$ and the eigenvectors $\bar{\mathbf{q}}_i$ of the correlation matrix $\bar{\mathbf{R}}(n)$ as follows

$$\begin{bmatrix} 1 \\ \mathbf{a}_o(n) \end{bmatrix} = P_o^f(n) \sum_{i=1}^{M+1} \frac{1}{\bar{\lambda}_i} \bar{\mathbf{q}}_i \bar{q}_{i,1}^* \qquad (6.5.43)$$

and

$$\begin{bmatrix} \mathbf{b}_o(n) \\ 1 \end{bmatrix} = P_o^b(n) \sum_{i=1}^{M+1} \frac{1}{\bar{\lambda}_i} \bar{\mathbf{q}}_i \bar{q}_{i,M+1}^* \qquad (6.5.44)$$

where $\bar{q}_{i,1}$ and $\bar{q}_{i,M+1}$ are the first and last components of $\bar{\mathbf{q}}_i$. The first equation of (6.5.43) and the last equation in (6.5.44) can be solved to provide the MMSEs $P_o^f(n)$ and $P_o^b(n)$, respectively.

Proof. See Problem 6.13.

PROPERTY 6.5.4. Let $\bar{\mathbf{R}}^{-1}(n)$ be the inverse of the correlation matrix $\bar{\mathbf{R}}(n)$. Then, the inverse of the ith element of the ith column of $\bar{\mathbf{R}}^{-1}(n)$ is equal to the MMSE $P^{(i)}(n)$, and the ith column normalized by the ith element is equal to $\mathbf{c}^{(i)}(n)$.

Proof. See Problem 6.14.

PROPERTY 6.5.5. The MMSE prediction errors can be expressed as

$$P_o^f(n) = \frac{\det \bar{\mathbf{R}}(n)}{\det \mathbf{R}(n-1)} \qquad P_o^b(n) = \frac{\det \bar{\mathbf{R}}(n)}{\det \mathbf{R}(n)} \qquad (6.5.45)$$

Proof. Problem 6.17.

The previous concepts are illustrated in the following example.

EXAMPLE 6.5.1. A random sequence $x(n)$ is generated by passing the white Gaussian noise process $w(n) \sim \text{WN}(0, 1)$ through the filter

$$x(n) = w(n) + \tfrac{1}{2}w(n-1)$$

Determine the second-order FLP, BLP, and symmetric linear signal smoother.

Solution. The complex power spectrum is

$$R(z) = H(z)H(z^{-1}) = (1 + \tfrac{1}{2}z^{-1})(1 + \tfrac{1}{2}z) = \tfrac{1}{2}z + \tfrac{5}{4} + \tfrac{1}{2}z^{-1}$$

Therefore, the autocorrelation sequence is equal to $r(0) = \frac{5}{4}$, $r(\pm 1) = \frac{1}{2}$, $r(l) = 0$ for $|l| \ge 2$. Since the power spectrum $R(e^{j\omega}) = \frac{5}{4} + \cos\omega > 0$ for all ω, the autocorrelation matrix is positive definite. The same is true of any principal submatrix. To determine the second-order linear signal estimators, we start with the matrix

$$\bar{\mathbf{R}} = \begin{bmatrix} \frac{5}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{5}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{5}{4} \end{bmatrix}$$

and follow the procedure outlined in Section 6.5.1 or use the formulas in Table 6.3. The results are

Forward linear prediction ($i = 0$):	$\{a_k\} \to \{1, -0.476, 0.190\}$	$P_o^f = 1.0119$
Symmetric linear smoothing ($i = 1$):	$\{c_k\} \to \{-0.4, 1, -0.4\}$	$P_o^s = 0.8500$
Backward linear prediction ($i = 2$):	$\{b_k\} \to \{0.190, -0.476, 1\}$	$P_o^b = 1.0119$

The inverse of the correlation matrix $\bar{\mathbf{R}}$ is

$$\bar{\mathbf{R}}^{-1} = \begin{bmatrix} 0.9882 & -0.4706 & 0.1882 \\ -0.4706 & 1.1765 & -0.4706 \\ 0.1882 & -0.4706 & 0.9882 \end{bmatrix}$$

and we see that dividing the first, second, and third columns by 0.9882, 1.1765, and 0.9882 provides the forward PEF, the symmetric linear smoothing filter, and the backward PEF, respectively. The inverses of the diagonal elements provide the MMSEs P_o^f, P_o^s, and P_o^b. The reader can easily see, by computing the zeros of the corresponding system functions, that the FLP is minimum-phase, the BLP is maximum-phase, and the symmetric linear smoother is mixed-phase. It is interesting to note that the smoother performs better than either of the predictors.

6.6 OPTIMUM INFINITE IMPULSE RESPONSE FILTERS

So far we have dealt with optimum FIR filters and predictors for nonstationary and stationary processes. In this section, we consider the design of optimum IIR filters for stationary stochastic processes. For nonstationary processes, the theory becomes very complicated. The Wiener-Hopf equations for optimum IIR filters are the same for FIR filters; only the limits in the convolution summation and the range of values for which the normal equations hold are different. Both are determined by the limits of summation in the filter convolution equation. We can easily see from (6.4.16) and (6.4.17), or by applying the orthogonality principle (6.2.41), that the optimum IIR filter

$$\hat{y}(n) = \sum_k h_o(k)x(n-k) \qquad (6.6.1)$$

is specified by the Wiener-Hopf equations

$$\sum_k h_o(k)r_x(m-k) = r_{yx}(m) \qquad (6.6.2)$$