

Linear Prediction

7. Linear Prediction

J. Benesty, J. Chen, Y. Huang

Linear prediction plays a fundamental role in all aspects of speech. Its use seems natural and obvious in this context since for a speech signal the value of its current sample can be well modeled as a linear combination of its past values. In this chapter, we attempt to present the most important ideas on linear prediction. We derive the principal results, widely recognized by speech experts, in a very intuitive way without sacrificing mathematical rigor.

7.1	Fundamentals	111
7.2	Forward Linear Prediction	112
7.3	Backward Linear Prediction	113
7.4	Levinson–Durbin Algorithm	114
7.5	Lattice Predictor	116
7.6	Spectral Representation	117
7.7	Linear Interpolation	118
7.8	Line Spectrum Pair Representation	119
7.9	Multichannel Linear Prediction	120
7.10	Conclusions	123
	References	123

7.1 Fundamentals

Linear prediction (LP) is a fundamental tool in many diverse areas such as adaptive filtering, system identification, economics, geophysics, spectral estimation, and speech. Recently, a nice history of LP in the context of speech coding was written by *Atal* [7.1]. Readers are invited to consult this reference for more information about this topic and how it has evolved.

Linear prediction is widely used in speech applications (recognition, compression, modeling, etc.) [7.2,3]. This is due to the fact that the speech production process is well modeled with LP. Indeed, it is well recognized that a speech signal can be written in the following form [7.4,5],

$$x(k) = \sum_{l=1}^L a_l x(k-l) + Gu(k), \quad (7.1)$$

where k is the time index, L represents the number of coefficients in the model (the order of the predictor), a_l , $l = 1, \dots, L$, are defined as the linear prediction coefficients, G is the gain of the system, and $u(k)$ is the excitation signal, which can be either a quasiperiodic train of impulses or a random noise source (also a combination of both signals for voiced fricatives such as ‘v’, ‘z’, and ‘zh’). The periodic source produces voiced sounds such as vowels and nasals, and the noise

source produces unvoiced or fricated sounds such as the fricatives. The parameters, a_l , determine the spectral characteristics of the particular sound for each of the two types of excitation and are widely used directly in many speech coding schemes and automatic speech recognition systems [7.4].

Equation (7.1) can be rewritten in the frequency domain, by using the z -transform. If $H(z)$ is the transfer function of the system, we have:

$$\begin{aligned} H(z) &= \frac{G}{1 - \sum_{l=1}^L a_l z^{-l}} \\ &= \frac{G}{A(z)}, \end{aligned} \quad (7.2)$$

which is an all-pole transfer function. This filter [$H(z)$] is a good model of the human vocal tract [7.2]. Our main concern is to determine the predictor coefficients, a_l , $l = 1, 2, \dots, L$, and to study the properties of the filter $A(z)$.

The applications of LP are numerous. Before addressing the estimation of LP coefficients, we give some examples to show the importance of LP. In many aspects of speech processing (noise reduction, speech separation, speech dereverberation, speech coding, etc.), it is of great interest to compare the closeness of the spectral envelope of two speech signals (the desired and

the processed ones) [7.6, 7]. One way of doing this is through comparing their LP coefficients. Consider the two speech signals $x(k)$ (desired) and $\hat{x}(k)$ (processed). Without entering too much into the details, one possible measure to evaluate the closeness of these two signals is the Itakura distance:

$$\text{ID}_{x\hat{x}} = \ln \frac{E_x}{E_{\hat{x}}}, \quad (7.3)$$

where E_x and $E_{\hat{x}}$ are the prediction-error powers of the signals $x(k)$ and $\hat{x}(k)$, respectively (see the following sections for more details). Note that the Itakura distance is not symmetric, i. e.,

$$\text{ID}_{x\hat{x}} \neq \text{ID}_{\hat{x}x}, \quad (7.4)$$

therefore, it is not a distance metric. However, asymmetry is usually not a problem for applications such as speech quality evaluation.

A more-powerful distance was proposed by Itakura and Saito in their formulation of linear prediction as an approximate maximum-likelihood estimation [7.8]. This distance between the two signals $x(k)$ and $\hat{x}(k)$ is defined as,

$$\text{ISD}_{x\hat{x}} = \frac{E_{\hat{x}}}{E_x} - \ln \frac{E_{\hat{x}}}{E_x} - 1. \quad (7.5)$$

7.2 Forward Linear Prediction

Consider a stationary random signal $x(k)$. The objective of the forward linear prediction is to predict the value of the sample $x(k)$ from its past values, i. e., $x(k-1)$, $x(k-2)$, etc. We define the forward prediction error as [7.10, 11],

$$\begin{aligned} e_{f,L}(k) &= x(k) - \hat{x}(k) \\ &= x(k) - \sum_{l=1}^L a_{L,l} x(k-l) \\ &= x(k) - \mathbf{a}_L^T \mathbf{x}(k-1), \end{aligned} \quad (7.6)$$

where the superscript ‘T’ denotes transposition, $\hat{x}(k)$ is the predicted sample,

$$\mathbf{a}_L = [a_{L,1} \ a_{L,2} \ \cdots \ a_{L,L}]^T$$

is the forward predictor of length L , and

$$\mathbf{x}(k-1) = [x(k-1) \ x(k-2) \ \cdots \ x(k-L)]^T$$

is a vector containing the L most recent samples starting with and including $x(k-1)$.

Like the Itakura distance, this measure is not symmetric either; therefore, it is not a true metric.

The Itakura–Saito distance has many interesting properties. It has been shown that this measure is highly correlated with subjective quality judgements [7.6]. For example, a recent report on speech codec evaluation reveals that, if the Itakura–Saito measure between two speech signals is less than 0.5, the difference in their mean opinion score would be less than 1.6 [7.9]. Many other reported experiments also confirmed that when the Itakura–Saito distance between two speech signals is below 0.1, they would be perceived nearly identically by human ears. As a result, the Itakura–Saito distance, which is based on LP, is often used as an objective measure of speech quality. It is probably the most widely used measure of similarity between speech signals.

The two previous examples of the vocal-tract filter and the speech quality measure clearly show the importance of LP in speech applications.

In this chapter, we study the theory of linear prediction and derive the most important LP techniques that are often encountered in many speech applications. We assume here that all signals of interest are real, stationary, and zero mean.

We would like to find the optimal Wiener predictor. For that, we seek to minimize the mean-square error (MSE):

$$J_f(\mathbf{a}_L) = E\{e_{f,L}^2(k)\}, \quad (7.7)$$

where $E\{\cdot\}$ denotes mathematical expectation. Taking the gradient of $J_f(\mathbf{a}_L)$ with respect to \mathbf{a}_L and equating to $\mathbf{0}_{L \times 1}$ (a vector of length L containing only zeroes), we easily find the Wiener–Hopf equations:

$$\mathbf{R}_L \mathbf{a}_{o,L} = \mathbf{r}_{f,L}, \quad (7.8)$$

where the subscript ‘o’ in $\mathbf{a}_{o,L}$ stands for optimal,

$$\begin{aligned} \mathbf{R}_L &= E\{\mathbf{x}(k-1)\mathbf{x}^T(k-1)\} \\ &= E\{\mathbf{x}(k)\mathbf{x}^T(k)\} \\ &= \begin{pmatrix} r(0) & r(1) & \cdots & r(L-1) \\ r(1) & r(0) & \cdots & r(L-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(L-1) & r(L-2) & \cdots & r(0) \end{pmatrix} \end{aligned} \quad (7.9)$$

is the correlation matrix, and

$$\begin{aligned} \mathbf{r}_{f,L} &= E\{\mathbf{x}(k-1)x(k)\} \\ &= [r(1) \ r(2) \ \cdots \ r(L)]^T \end{aligned} \quad (7.10)$$

is the correlation vector. The matrix \mathbf{R}_L has a Toeplitz structure (i. e., all the entries along the diagonals are the same); assuming that it is nonsingular, we deduce the optimal forward predictor:

$$\mathbf{a}_{o,L} = \mathbf{R}_L^{-1} \mathbf{r}_{f,L}. \quad (7.11)$$

Expanding $e_{f,L}^2(k)$ in (7.7) and using (7.8) shows that the minimum mean-square error (MMSE),

$$\begin{aligned} J_{f,\min} &= J_f(\mathbf{a}_{o,L}) \\ &= r(0) - \mathbf{r}_{f,L}^T \mathbf{a}_{o,L} = E_{f,L}. \end{aligned} \quad (7.12)$$

This is also called the forward prediction-error power.

Define the augmented correlation matrix:

$$\mathbf{R}_{L+1} = \begin{pmatrix} r(0) & \mathbf{r}_{f,L}^T \\ \mathbf{r}_{f,L} & \mathbf{R}_L \end{pmatrix}, \quad (7.13)$$

equations (7.8) and (7.12) may be combined in a convenient way:

$$\mathbf{R}_{L+1} \begin{pmatrix} 1 \\ -\mathbf{a}_{o,L} \end{pmatrix} = \begin{pmatrix} E_{f,L} \\ \mathbf{0}_{L \times 1} \end{pmatrix}. \quad (7.14)$$

We refer to (7.14) as the augmented Wiener–Hopf equations of a forward predictor of order L . From (7.13) we derive that,

$$\det(\mathbf{R}_{L+1}) = E_{f,L} \det(\mathbf{R}_L), \quad (7.15)$$

7.3 Backward Linear Prediction

The aim of the backward linear prediction is to predict the value of the sample $x(k-L)$ from its future values, i. e., $x(k)$, $x(k-1)$, \dots , $x(k-L+1)$. We define the backward prediction error as,

$$\begin{aligned} e_{b,L}(k) &= x(k-L) - \hat{x}(k-L) \\ &= x(k-L) - \sum_{l=1}^L b_{L,l} x(k-l+1) \\ &= x(k-L) - \mathbf{b}_L^T \mathbf{x}(k), \end{aligned} \quad (7.21)$$

where $\hat{x}(k-L)$ is the predicted sample,

$$\mathbf{b}_L = [b_{L,1} \ b_{L,2} \ \cdots \ b_{L,L}]^T$$

is the backward predictor of order L , and

$$\mathbf{x}(k) = [x(k) \ x(k-1) \ \cdots \ x(k-L+1)]^T.$$

The minimization of the MSE,

$$J_b(\mathbf{b}_L) = E\{e_{b,L}^2(k)\}, \quad (7.22)$$

where ‘det’ stands for determinant.

Let us now write the forward prediction errors for the optimal predictors of orders L and $L-i$:

$$e_{f,o,L}(k) = x(k) - \sum_{l=1}^L a_{o,L,l} x(k-l), \quad (7.16)$$

$$e_{f,o,L-i}(k) = x(k) - \sum_{l=1}^{L-i} a_{o,L-i,l} x(k-l). \quad (7.17)$$

From the principle of orthogonality [7.11], we know that:

$$E\{e_{f,o,L}(k)\mathbf{x}(k-1)\} = \mathbf{0}_{L \times 1}. \quad (7.18)$$

For $1 \leq i \leq L$, we can verify by using (7.18), that:

$$E\{e_{f,o,L}(k)e_{f,o,L-i}(k-i)\} = 0. \quad (7.19)$$

As a result,

$$\begin{aligned} \lim_{L \rightarrow \infty} E\{e_{f,o,L}(k)e_{f,o,L-i}(k-i)\} \\ = E\{e_{f,o}(k)e_{f,o}(k-i)\} = 0. \end{aligned} \quad (7.20)$$

This indicates that the signal $e_{f,o}(k)$ is a white noise. So the optimal forward predictor has this important property of being able to whiten a stationary random process, provided that the order of the predictor is high enough.

leads to the Wiener–Hopf equations:

$$\mathbf{R}_L \mathbf{b}_{o,L} = \mathbf{r}_{b,L}, \quad (7.23)$$

where

$$\begin{aligned} \mathbf{r}_{b,L} &= E\{\mathbf{x}(k)x(k-L)\} \\ &= [r(L) \ r(L-1) \ \cdots \ r(1)]^T. \end{aligned} \quad (7.24)$$

Therefore, the optimal backward predictor is:

$$\mathbf{b}_{o,L} = \mathbf{R}_L^{-1} \mathbf{r}_{b,L}. \quad (7.25)$$

The MMSE for backward prediction,

$$\begin{aligned} J_{b,\min} &= J_b(\mathbf{b}_{o,L}) \\ &= r(0) - \mathbf{r}_{b,L}^T \mathbf{b}_{o,L} = E_{b,L}, \end{aligned} \quad (7.26)$$

is also called the backward prediction-error power.

Define the augmented correlation matrix:

$$\mathbf{R}_{L+1} = \begin{pmatrix} \mathbf{R}_L & \mathbf{r}_{b,L} \\ \mathbf{r}_{b,L}^T & r(0) \end{pmatrix}, \quad (7.27)$$

equations (7.23) and (7.26) may be combined in a convenient way:

$$\mathbf{R}_{L+1} \begin{pmatrix} -\mathbf{b}_{o,L} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{L \times 1} \\ E_{b,L} \end{pmatrix}. \quad (7.28)$$

We refer to this expression as the augmented Wiener-Hopf equations of a backward predictor of order L .

One important property of backward prediction is that the error signals of different orders with the optimal predictors are uncorrelated, i.e., $E\{e_{b,o,i}(k)e_{b,o,l}(k)\} = 0, i \neq l, i, l = 0, 1, \dots, L-1$. To prove this, let us rewrite the error signal in vector form:

$$\mathbf{e}_{b,o}(k) = \mathbf{L}\mathbf{x}(k), \quad (7.29)$$

where

$$\mathbf{e}_{b,o}(k) = [e_{b,o,0}(k) \ e_{b,o,1}(k) \ \dots \ e_{b,o,L-1}(k)]^T \quad (7.30)$$

and

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\mathbf{b}_{o,1}^T & 1 & 0 & \dots & 0 \\ -\mathbf{b}_{o,2}^T & & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\mathbf{b}_{o,L-1}^T & & & & 1 \end{pmatrix} \quad (7.31)$$

is a lower triangular matrix with 1s along its main diagonal. The covariance matrix corresponding to the vector signal $\mathbf{e}_{b,o}(k)$ is:

7.4 Levinson–Durbin Algorithm

The Levinson–Durbin algorithm is an efficient way to solve the Wiener–Hopf equations for the forward and backward prediction coefficients. This efficient method can be derived thanks to the Toeplitz structure of the correlation matrix \mathbf{R}_L . This algorithm was first invented by *Levinson* [7.13] and independently reformulated at a later date by *Durbin* [7.14, 15]. *Burg* gave a more-elegant presentation [7.16]. Before describing this algorithm, we first need to show some important relations between the forward and backward predictors.

We define the co-identity matrix as:

$$\mathbf{J}_L = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

$$E\{\mathbf{e}_{b,o}(k)\mathbf{e}_{b,o}^T(k)\} = \mathbf{L}\mathbf{R}_L\mathbf{L}^T. \quad (7.32)$$

By definition, the previous matrix is symmetric. The matrix product $\mathbf{R}_L\mathbf{L}^T$ is a lower triangular matrix because of (7.28) and the main diagonal contains the backward prediction-error powers $E_{b,l}$ ($0 \leq l \leq L-1$). Since \mathbf{L} is also a lower triangular matrix, the product between the two matrices \mathbf{L} and $\mathbf{R}_L\mathbf{L}^T$ should have the same structure and, since it has to be symmetric, the only possibility is that this resulting matrix is diagonal:

$$E\{\mathbf{e}_{b,o}(k)\mathbf{e}_{b,o}^T(k)\} = \text{diag}[E_{b,0}, E_{b,1}, \dots, E_{b,L-1}], \quad (7.33)$$

and hence the prediction errors are uncorrelated.

Furthermore,

$$\mathbf{L}\mathbf{R}_L\mathbf{L}^T = \text{diag}[E_{b,0}, E_{b,1}, \dots, E_{b,L-1}], \quad (7.34)$$

taking the inverse of the previous equation,

$$\mathbf{L}^{-T}\mathbf{R}_L^{-1}\mathbf{L}^{-1} = \text{diag}[E_{b,0}^{-1}, E_{b,1}^{-1}, \dots, E_{b,L-1}^{-1}], \quad (7.35)$$

we finally get:

$$\mathbf{R}_L^{-1} = \mathbf{L}^T \text{diag}[E_{b,0}^{-1}, E_{b,1}^{-1}, \dots, E_{b,L-1}^{-1}]\mathbf{L}. \quad (7.36)$$

Expression (7.36) defines the Cholesky factorization of the inverse matrix \mathbf{R}_L^{-1} [7.10, 12].

We can easily check that:

$$\mathbf{R}_L\mathbf{J}_L = \mathbf{J}_L\mathbf{R}_L. \quad (7.37)$$

The matrix \mathbf{R}_L is said to be persymmetric. We also have, $\mathbf{r}_{f,L} = \mathbf{J}_L\mathbf{r}_{b,L}$. If we left-multiply both sides of the Wiener–Hopf equations (7.23) by \mathbf{J}_L , we get:

$$\begin{aligned} \mathbf{J}_L\mathbf{R}_L\mathbf{b}_{o,L} &= \mathbf{J}_L\mathbf{r}_{b,L} = \mathbf{r}_{f,L} \\ &= \mathbf{R}_L\mathbf{J}_L\mathbf{b}_{o,L} = \mathbf{R}_L\mathbf{a}_{o,L}, \end{aligned} \quad (7.38)$$

and, assuming that \mathbf{R}_L is nonsingular, we see that:

$$\mathbf{a}_{o,L} = \mathbf{J}_L\mathbf{b}_{o,L}. \quad (7.39)$$

Furthermore,

$$\begin{aligned}
 E_{b,L} &= r(0) - \mathbf{r}_{b,L}^T \mathbf{b}_{o,L} \\
 &= r(0) - \mathbf{r}_{b,L}^T \mathbf{J}_L \mathbf{J}_L \mathbf{b}_{o,L} \\
 &= r(0) - \mathbf{r}_{f,L}^T \mathbf{a}_{o,L} \\
 &= E_{f,L} = E_L .
 \end{aligned} \tag{7.40}$$

Therefore, for a stationary process, the forward and backward prediction-error powers are equal and the coefficients of the optimal forward predictor are the same as those of the optimal backward predictor, but in a reverse order.

The Levinson–Durbin algorithm is based on recursions of the orders of the prediction equations. Consider the following expression,

$$\begin{pmatrix} \mathbf{R}_L & \mathbf{r}_{b,L} \\ \mathbf{r}_{b,L}^T & r(0) \end{pmatrix} \begin{pmatrix} 1 \\ -\mathbf{a}_{o,L-1} \\ 0 \end{pmatrix} = \begin{pmatrix} E_{L-1} \\ \mathbf{0}_{(L-1) \times 1} \\ K_L \end{pmatrix}, \tag{7.41}$$

where

$$\begin{aligned}
 K_L &= r(L) - \mathbf{a}_{o,L-1}^T \mathbf{r}_{b,L-1} \\
 &= r(L) - \mathbf{a}_{o,L-1}^T \mathbf{J}_{L-1} \mathbf{r}_{f,L-1}.
 \end{aligned} \tag{7.42}$$

We define the reflection coefficient as,

$$\kappa_L = \frac{K_L}{E_{L-1}}. \tag{7.43}$$

From backward linear prediction, we have:

$$\begin{pmatrix} r(0) & \mathbf{r}_{f,L}^T \\ \mathbf{r}_{f,L} & \mathbf{R}_L \end{pmatrix} \begin{pmatrix} 0 \\ -\mathbf{b}_{o,L-1} \\ 1 \end{pmatrix} = \begin{pmatrix} K_L \\ \mathbf{0}_{(L-1) \times 1} \\ E_{L-1} \end{pmatrix}. \tag{7.44}$$

Multiplying both sides of the previous equation by κ_L , we get,

$$\mathbf{R}_{L+1} \begin{pmatrix} 0 \\ -\kappa_L \mathbf{b}_{o,L-1} \\ \kappa_L \end{pmatrix} = \begin{pmatrix} \kappa_L^2 E_{L-1} \\ \mathbf{0}_{(L-1) \times 1} \\ K_L \end{pmatrix}. \tag{7.45}$$

If we now subtract (7.45) from (7.41), we obtain,

$$\mathbf{R}_{L+1} \begin{pmatrix} 1 \\ \kappa_L \mathbf{b}_{o,L-1} - \mathbf{a}_{o,L-1} \\ -\kappa_L \end{pmatrix} = \begin{pmatrix} E_{L-1}(1 - \kappa_L^2) \\ \mathbf{0}_{L \times 1} \end{pmatrix}. \tag{7.46}$$

Assuming that \mathbf{R}_{L+1} is nonsingular and identifying (7.46) with (7.14), we can deduce the recursive equa-

tions:

$$\mathbf{a}_{o,L} = \begin{pmatrix} \mathbf{a}_{o,L-1} \\ 0 \end{pmatrix} - \kappa_L \begin{pmatrix} \mathbf{b}_{o,L-1} \\ -1 \end{pmatrix}, \tag{7.47}$$

$$E_L = E_{L-1}(1 - \kappa_L^2), \tag{7.48}$$

$$\mathbf{a}_{o,L,L} = \kappa_L. \tag{7.49}$$

Iterating on the prediction-error power given in (7.48), we find that,

$$E_L = r(0) \prod_{l=1}^L (1 - \kappa_l^2), \tag{7.50}$$

and since $E_L \geq 0$, this implies that,

$$|\kappa_l| \leq 1, \quad \forall l \geq 1. \tag{7.51}$$

Also, from (7.48) we see that we have,

$$0 \leq E_l \leq E_{l-1}, \quad \forall l \geq 1, \tag{7.52}$$

so, as the order of the predictors increases, the prediction-error power decreases.

Table 7.1 summarizes the Levinson–Durbin algorithm, whose arithmetic complexity is proportional to L^2 . This algorithm is much more efficient than standard methods such as the Gauss elimination technique, whose complexity is on the order of L^3 . The saving in number of operations to find the optimal Wiener predictor can be very important, especially when L is large. The other advantage of the Levinson–Durbin algorithm is that it gives the predictors of all orders and the algorithm can be stopped if the prediction-error power is under a threshold, which can be very useful in practice when the choice of the predictor order is not easy to get in advance. A slightly more-efficient approach, called the split Levinson algorithm, can be found in [7.17]. This algorithm requires roughly half the number of multiplications and the same number of additions as the classical Levinson–Durbin algorithm. Even more-efficient algorithms have been proposed (see, for example, [7.18]) but they are numerically unstable, which is not acceptable in most speech applications.

Table 7.1 Levinson–Durbin algorithm

Initialization: $E_0 = r(0)$

For $1 \leq l \leq L$

$$\kappa_l = \frac{1}{E_{l-1}} [r(l) - \mathbf{a}_{o,l-1}^T \mathbf{J}_{l-1} \mathbf{r}_{f,l-1}]$$

$$\mathbf{a}_{o,l} = \begin{pmatrix} \mathbf{a}_{o,l-1} \\ 0 \end{pmatrix} - \kappa_l \mathbf{J}_l \begin{pmatrix} -1 \\ \mathbf{a}_{o,l-1} \end{pmatrix}$$

$$E_l = E_{l-1}(1 - \kappa_l^2)$$

7.5 Lattice Predictor

In this section, we will show that the order-recursive structure of the forward and backward prediction errors has the form of a ladder, which is called a lattice predictor.

Inserting (7.47) into the forward prediction error for the optimal predictor of order L ,

$$e_{f,o,L}(k) = x(k) - \mathbf{a}_{o,L}^T \mathbf{x}(k-1), \quad (7.53)$$

we obtain,

$$e_{f,o,L}(k) = e_{f,o,L-1}(k) - \kappa_L \left(-\mathbf{b}_{o,L-1}^T \mathbf{1} \right) \mathbf{x}(k-1). \quad (7.54)$$

The second term (without the reflection coefficient) on the right-hand side of (7.54) is the backward prediction error, at time $k-1$, for the optimal predictor of order $L-1$. Therefore, (7.54) can be rewritten

$$e_{f,o,L}(k) = e_{f,o,L-1}(k) - \kappa_L e_{b,o,L-1}(k-1). \quad (7.55)$$

If we insert (7.47) again into the backward prediction error for the optimal predictor of order L ,

$$\begin{aligned} e_{b,o,L}(k) &= x(k-L) - \mathbf{b}_{o,L}^T \mathbf{x}(k) \\ &= x(k-L) - \mathbf{a}_{o,L}^T \mathbf{J}_L \mathbf{x}(k), \end{aligned} \quad (7.56)$$

we get,

$$e_{b,o,L}(k) = e_{b,o,L-1}(k-1) - \kappa_L e_{f,o,L-1}(k). \quad (7.57)$$

If we put (7.55) and (7.57) into a matrix form, we have,

$$\begin{aligned} \begin{pmatrix} e_{f,o,L}(k) \\ e_{b,o,L}(k) \end{pmatrix} &= \begin{pmatrix} 1 & -\kappa_L \\ -\kappa_L & 1 \end{pmatrix} \begin{pmatrix} e_{f,o,L-1}(k) \\ e_{b,o,L-1}(k-1) \end{pmatrix} \\ &= \prod_{l=1}^L \begin{pmatrix} 1 & -\kappa_l \\ -\kappa_l & 1 \end{pmatrix} \begin{pmatrix} x(k) \\ x(k-1) \end{pmatrix}, \end{aligned} \quad (7.58)$$

where we have taken for initial conditions (order 0), $e_{f,o,0}(k) = x(k)$ and $e_{b,o,0}(k-1) = x(k-1)$. Figure 7.1

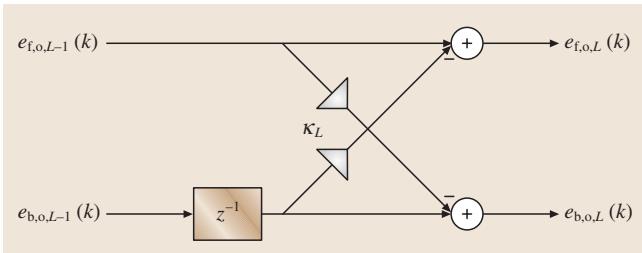


Fig. 7.1 Stage l of a lattice predictor

depicts the l -th stage of a lattice predictor. For the whole lattice predictor, L of these stages are needed and are connected in cascade, one to each other, starting from order 0 to order L .

Now let us compute the variance of $e_{b,o,L}(k)$ from (7.57),

$$\begin{aligned} E\{e_{b,o,L}^2(k)\} &= E_L \\ &= E_{L-1} + \kappa_L^2 E_{L-1} \\ &\quad - 2\kappa_L E\{e_{f,o,L-1}(k)e_{b,o,L-1}(k-1)\} \\ &= E_{L-1}(1 - \kappa_L^2). \end{aligned} \quad (7.59)$$

Developing the previous expression, we obtain,

$$\begin{aligned} \kappa_L &= \frac{E\{e_{f,o,L-1}(k)e_{b,o,L-1}(k-1)\}}{E_{L-1}} \\ &= \frac{E\{e_{f,o,L-1}(k)e_{b,o,L-1}(k-1)\}}{\sqrt{E\{e_{f,o,L-1}^2(k)\}E\{e_{b,o,L-1}^2(k-1)\}}}. \end{aligned} \quad (7.60)$$

We see from (7.60) that the reflection coefficients are also the normalized cross-correlation coefficients between the forward and backward prediction errors, which is why they are also often called partial correlation (PARCOR) coefficients [7.3, 19, 20]. These coefficients are linked to the zeroes of the forward prediction-error FIR filter of order L , whose transfer function is

$$A_{o,L}(z) = 1 - \sum_{l=1}^L a_{o,L,l} z^{-l} = \prod_{l=1}^L (1 - z_{o,l} z^{-1}), \quad (7.61)$$

where $z_{o,l}$ are the roots of $A_{o,L}(z)$. Since $\kappa_L = a_{o,L,L}$, we have,

$$\kappa_L = (-1)^{L+1} \prod_{l=1}^L z_{o,l}. \quad (7.62)$$

The filter $A_{o,L}(z)$ can be shown to be minimum phase, i. e., $|z_{o,l}| \leq 1, \forall l$. As a result [because of the relation (7.39)], the filter $B_{o,L}(z)$ corresponding to the backward predictor is maximum phase. We will now show this very important property that the forward predictor is minimum phase. As far as we know, this simple and elegant proof was first shown by M. Mohan Sondhi but was never been published. A similar proof can be found in [7.21] and [7.22].

To avoid cumbersome notation, redefine the coefficients $w_l = -a_{o,L,l}$, with $w_0 = 1$, so that the polynomial

becomes,

$$A_{o,L}(z) = \sum_{l=0}^L w_l z^{-l}. \quad (7.63)$$

Also, define the vector,

$$\mathbf{w} = [w_0 \ w_1 \ \cdots \ w_L]^T.$$

We know that,

$$\mathbf{R}_{L+1} \mathbf{w} = \begin{pmatrix} E_L \\ \mathbf{0}_{L \times 1} \end{pmatrix}. \quad (7.64)$$

If λ is a root of the polynomial, it follows that,

$$A_{o,L}(z) = (1 - \lambda z^{-1}) \sum_{l=0}^{L-1} g_l z^{-l}, \text{ with } g_0 = 1. \quad (7.65)$$

(Note that since λ can be complex, the coefficients g_l are, in general, complex.) Thus the vector \mathbf{w} can be written

$$\mathbf{w} = \mathbf{g} - \lambda \tilde{\mathbf{g}}, \quad (7.66)$$

where

$$\mathbf{g} = [1 \ g_1 \ g_2 \ \cdots \ g_{L-1} \ 0]^T = [\mathbf{g}^T \ 0]^T,$$

$$\tilde{\mathbf{g}} = [0 \ 1 \ g_1 \ g_2 \ \cdots \ g_{L-1}]^T = [0 \ \mathbf{g}^T]^T.$$

Substituting (7.66) in (7.64), we obtain,

$$\mathbf{R}_{L+1} \mathbf{g} = \lambda \mathbf{R}_{L+1} \tilde{\mathbf{g}} + \begin{pmatrix} E_L \\ \mathbf{0}_{L \times 1} \end{pmatrix}. \quad (7.67)$$

Now, premultiplying by $\tilde{\mathbf{g}}^H$ (where the superscript H denotes conjugate transpose) gives,

$$\tilde{\mathbf{g}}^H \mathbf{R}_{L+1} \mathbf{g} = \lambda \tilde{\mathbf{g}}^H \mathbf{R}_{L+1} \tilde{\mathbf{g}}. \quad (7.68)$$

7.6 Spectral Representation

It is important to understand the link between the spectrum of a speech signal and its prediction coefficients. Let us again take the speech model given in Sect. 7.1,

$$x(k) = \sum_{l=1}^L a_l x(k-l) + Gu(k), \quad (7.74)$$

where we now assume that $u(k)$ is a white random signal with variance $\sigma_u^2 = 1$. Since $x(k)$ is the output of the filter $H(z)$ (see Sect. 7.1), whose input is $u(k)$, its spectrum is [7.11],

$$S_x(\omega) = |H(e^{j\omega})|^2 S_u(\omega), \quad (7.75)$$

Thus,

$$\left| \tilde{\mathbf{g}}^H \mathbf{R}_{L+1} \mathbf{g} \right|^2 = |\lambda|^2 (\tilde{\mathbf{g}}^H \mathbf{R}_{L+1} \tilde{\mathbf{g}})^2. \quad (7.69)$$

Using the Schwartz inequality,

$$\left| \tilde{\mathbf{g}}^H \mathbf{R}_{L+1} \mathbf{g} \right|^2 \leq (\tilde{\mathbf{g}}^H \mathbf{R}_{L+1} \tilde{\mathbf{g}}) (\mathbf{g}^H \mathbf{R}_{L+1} \mathbf{g}). \quad (7.70)$$

However,

$$\begin{aligned} \tilde{\mathbf{g}}^H \mathbf{R}_{L+1} \tilde{\mathbf{g}} &= \begin{pmatrix} 0 & \mathbf{g}'^H \end{pmatrix} \begin{pmatrix} r^{(0)} & \mathbf{r}_{f,L}^T \\ \mathbf{r}_{f,L} & \mathbf{R}_L \end{pmatrix} \begin{pmatrix} 0 \\ \mathbf{g}' \end{pmatrix} \\ &= \mathbf{g}'^H \mathbf{R}_L \mathbf{g}', \end{aligned} \quad (7.71)$$

Similarly,

$$\begin{aligned} \mathbf{g}^H \mathbf{R}_{L+1} \mathbf{g} &= \begin{pmatrix} \mathbf{g}'^H & 0 \end{pmatrix} \begin{pmatrix} \mathbf{R}_L & \mathbf{r}_{b,L} \\ \mathbf{r}_{b,L}^T & r^{(0)} \end{pmatrix} \begin{pmatrix} \mathbf{g}' \\ 0 \end{pmatrix} \\ &= \mathbf{g}'^H \mathbf{R}_L \mathbf{g}'. \end{aligned} \quad (7.72)$$

Therefore, $\tilde{\mathbf{g}}^H \mathbf{R}_{L+1} \tilde{\mathbf{g}} = \mathbf{g}'^H \mathbf{R}_L \mathbf{g}'$, and the Schwartz inequality becomes,

$$\left| \tilde{\mathbf{g}}^H \mathbf{R}_{L+1} \mathbf{g} \right|^2 \leq (\tilde{\mathbf{g}}^H \mathbf{R}_{L+1} \tilde{\mathbf{g}})^2. \quad (7.73)$$

From (7.69) we see that $|\lambda|^2 \leq 1$. This completes the proof.

This property allows one easily to ensure that the all-pole system in (7.2) is stable (when the correlation matrix is positive definite) by simply imposing the constraint that the PARCOR coefficients are less than 1 in magnitude. As a result, in speech communication, transmitting PARCOR coefficients is more advantageous than directly transmitting linear prediction coefficients.

where ω is the angular frequency, $H(e^{j\omega})$ is the frequency response of the filter $H(z)$, and $S_u(\omega)$ is the spectrum of u . We have $S_u(\omega) = 1$ (u is white). Using (7.2), we deduce the spectrum of x ,

$$\begin{aligned} S_x(\omega) &= \frac{G^2}{|A(e^{j\omega})|^2} \\ &= \frac{G^2}{\left| 1 - \sum_{l=1}^L a_l e^{-jl\omega} \right|^2}. \end{aligned} \quad (7.76)$$

Therefore, the spectrum of a speech signal can be modeled by the frequency response of an all-pole filter,

whose elements are the prediction coefficients [7.23–25].

Consider the prediction error signal,

$$e_L(k) = x(k) - \mathbf{a}_L^T \mathbf{x}(k-1). \quad (7.77)$$

Taking the z -transform of (7.77) and setting $z = e^{i\omega}$, we obtain:

$$S_{x,L}(\omega) = \frac{|E_L(e^{i\omega})|^2}{|A_L(e^{i\omega})|^2}, \quad (7.78)$$

where $|E_L(e^{i\omega})|^2$ is the spectrum of $e_L(k)$. From Sect. 7.2, we know that, for a large order L , linear prediction tends to whiten the signal, so the power spectrum

$|E_L(e^{i\omega})|^2$ of the error signal, $e_L(k)$, will tend to be flat. Hence,

$$\lim_{L \rightarrow \infty} |E_L(e^{i\omega})|^2 = G^2. \quad (7.79)$$

As a result,

$$\lim_{L \rightarrow \infty} S_{x,L}(\omega) = \frac{G^2}{|1 - \sum_{l=1}^{\infty} a_l e^{-il\omega}|^2}. \quad (7.80)$$

This confirms that (7.76) can be a very good approximation of the spectrum of a speech signal, as long as the order of the predictor is large enough.

7.7 Linear Interpolation

Linear interpolation can be seen as a straightforward generalization of forward and backward linear predictions. Indeed, in linear interpolation, we try to predict the value of the sample $x(k-i)$ from its past and future values [7.26, 27]. We define the interpolation error as

$$\begin{aligned} e_i(k) &= x(k-i) - \hat{x}(k-i) \\ &= x(k-i) - \sum_{l=0, l \neq i}^L c_{i,l} x(k-l) \\ &= \mathbf{c}_i^T \mathbf{x}_{L+1}(k), \quad i = 0, 1, \dots, L, \end{aligned} \quad (7.81)$$

where $\hat{x}(k-i)$ is the interpolated sample,

$$\mathbf{c}_i = [-c_{i,0} \quad -c_{i,1} \quad \dots \quad c_{i,i} \quad \dots \quad -c_{i,L}]^T$$

is a vector of length $L+1$ containing the interpolation coefficients, with $c_{i,i} = 1$, and

$$\mathbf{x}_{L+1}(k) = [x(k) \quad x(k-1) \quad \dots \quad x(k-L)]^T.$$

The special cases $i = 0$ and $i = L$ are the forward and backward prediction errors, respectively.

To find the optimal Wiener interpolator, we need to minimize the cost function,

$$\begin{aligned} J_i(\mathbf{c}_i) &= E\{e_i^2(k)\} \\ &= \mathbf{c}_i^T \mathbf{R}_{L+1} \mathbf{c}_i, \end{aligned} \quad (7.82)$$

subject to the constraint

$$\mathbf{c}_i^T \mathbf{v}_i = c_{i,i} = 1, \quad (7.83)$$

where

$$\mathbf{v}_i = [0 \quad 0 \quad \dots \quad 0 \quad 1 \quad 0 \quad \dots \quad 0]^T$$

is a vector of length $L+1$ with its i -th component equal to one and all others equal to zero. By using a Lagrange multiplier, it is easy to see that the solution to this optimization problem is

$$\mathbf{R}_{L+1} \mathbf{c}_{0,i} = E_i \mathbf{v}_i, \quad (7.84)$$

where

$$\begin{aligned} E_i &= \mathbf{c}_{0,i}^T \mathbf{R}_{L+1} \mathbf{c}_{0,i} \\ &= \frac{1}{\mathbf{v}_i^T \mathbf{R}_{L+1}^{-1} \mathbf{v}_i} \end{aligned} \quad (7.85)$$

is the interpolation-error power.

From (7.84) we find,

$$\frac{\mathbf{c}_{0,i}}{E_i} = \mathbf{R}_{L+1}^{-1} \mathbf{v}_i, \quad (7.86)$$

hence the i -th column of \mathbf{R}_{L+1}^{-1} is $\mathbf{c}_{0,i}/E_i$. We can now see that \mathbf{R}_{L+1}^{-1} can be factorized as follows [7.28]:

$$\begin{aligned} \mathbf{R}_{L+1}^{-1} &= \begin{pmatrix} 1 & -c_{0,1,0} & \dots & -c_{0,L,0} \\ -c_{0,0,1} & 1 & \dots & -c_{0,L,1} \\ \vdots & \vdots & \ddots & \vdots \\ -c_{0,0,L} & -c_{0,1,L} & \dots & 1 \end{pmatrix} \\ &\quad \cdot \begin{pmatrix} 1/E_0 & 0 & \dots & 0 \\ 0 & 1/E_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/E_{L-1} \end{pmatrix} \\ &= \mathbf{C}_0^T \mathbf{D}_e^{-1}. \end{aligned} \quad (7.87)$$

Furthermore, since \mathbf{R}_{L+1}^{-1} is a symmetric matrix, (7.87) can be written as,

$$\mathbf{R}_{L+1}^{-1} = \begin{pmatrix} 1/E_0 & 0 & \cdots & 0 \\ 0 & 1/E_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/E_{L-1} \end{pmatrix} \cdot \begin{pmatrix} 1 & -c_{0,0,1} & \cdots & -c_{0,0,L} \\ -c_{0,1,0} & 1 & \cdots & -c_{0,1,L} \\ \vdots & \vdots & \ddots & \vdots \\ -c_{0,L,0} & -c_{0,L,1} & \cdots & 1 \end{pmatrix} = \mathbf{D}_e^{-1} \mathbf{C}_0. \quad (7.88)$$

Therefore, we deduce that,

$$\frac{c_{0,i,l}}{E_i} = \frac{c_{0,l,i}}{E_l}, \quad i, l = 0, 1, \dots, L. \quad (7.89)$$

The first and last columns of \mathbf{R}_{L+1}^{-1} contain, respectively, the normalized forward and backward predictors and all the columns between contain the normalized interpolators.

We are now going to show how the condition number of the correlation matrix depends on the interpolators. The condition number of the matrix \mathbf{R}_{L+1} is defined as [7.29]:

$$\chi(\mathbf{R}_{L+1}) = \|\mathbf{R}_{L+1}\| \|\mathbf{R}_{L+1}^{-1}\|, \quad (7.90)$$

where $\|\cdot\|$ can be any matrix norm. Note that $\chi(\mathbf{R})$ depends on the underlying norm. Let us compute $\chi(\mathbf{R}_{L+1})$ using the Frobenius norm:

$$\begin{aligned} \|\mathbf{R}_{L+1}\|_F &= [\text{tr}(\mathbf{R}_{L+1}^T \mathbf{R}_{L+1})]^{1/2} \\ &= [\text{tr}(\mathbf{R}_{L+1}^2)]^{1/2} \end{aligned} \quad (7.91)$$

7.8 Line Spectrum Pair Representation

Line spectrum pair (LSP) representation, first introduced by *Itakura* [7.33], is a more-robust way to represent the coefficients of linear predictive models. The LSP polynomials have some very interesting properties shown in [7.34].

A polynomial $P(z)$ of order L is said to be symmetric if

$$P(z) = z^{-L} P(z^{-1}) \quad (7.97)$$

and a polynomial $Q(z)$ is antisymmetric if

$$Q(z) = -z^{-L} Q(z^{-1}). \quad (7.98)$$

and

$$\|\mathbf{R}_{L+1}^{-1}\|_F = [\text{tr}(\mathbf{R}_{L+1}^{-2})]^{1/2}. \quad (7.92)$$

From (7.86), we have,

$$\frac{\mathbf{c}_{0,i}^T \mathbf{c}_{0,i}}{E_i^2} = \mathbf{v}_i^T \mathbf{R}_{L+1}^{-2} \mathbf{v}_i, \quad (7.93)$$

which implies that,

$$\begin{aligned} \sum_{i=0}^L \frac{\mathbf{c}_{0,i}^T \mathbf{c}_{0,i}}{E_i^2} &= \sum_{i=0}^L \mathbf{v}_i^T \mathbf{R}_{L+1}^{-2} \mathbf{v}_i \\ &= \text{tr}(\mathbf{R}_{L+1}^{-2}). \end{aligned} \quad (7.94)$$

Also, we can easily check that,

$$\text{tr}(\mathbf{R}_{L+1}^2) = (L+1)r^2(0) + 2 \sum_{l=1}^L (L+1-l)r^2(l). \quad (7.95)$$

Therefore, the square of the condition number of the correlation matrix associated with the Frobenius norm is

$$\begin{aligned} \chi_F^2(\mathbf{R}_{L+1}) &= \left[(L+1)r^2(0) + 2 \sum_{l=1}^L (L+1-l)r^2(l) \right] \\ &\quad \times \sum_{i=0}^L \frac{\mathbf{c}_{0,i}^T \mathbf{c}_{0,i}}{E_i^2}. \end{aligned} \quad (7.96)$$

Some other interesting relations between the forward predictors and the condition number can be found in [7.30].

To conclude this section, we would like to let readers know that several algorithms exist to compute the optimal predictors efficiently, see for example, [7.31] and [7.32]. All these algorithms are based on Levinson–Durbin recursions.

Let

$$A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_L z^{-L} \quad (7.99)$$

be the optimal polynomial predictor of order L . It is well known that in speech compression the coefficients of this polynomial are inappropriate for quantization because of their relatively large dynamic range and also because, as stated earlier, quantization can change a stable LPC filter into an unstable one [7.35]. From (7.99), we can construct two artificial $(L+1)$ -th-order (symmetric and antisymmetric) polynomials by setting the $(L+1)$ -th

reflection coefficient, κ_{L+1} , to be $+1$ and -1 . These two cases correspond, respectively, to an entirely closed or to an entirely open end at the last section of an acoustic tube of $L + 1$ piecewise-uniform sections [7.35],

$$P(z) = A(z) + z^{-L}A(z^{-1}), \quad (7.100)$$

$$Q(z) = A(z) - z^{-L}A(z^{-1}). \quad (7.101)$$

The polynomial $A(z)$ can be easily reconstructed from $P(z)$ and $Q(z)$ by

$$A(z) = \frac{1}{2}[P(z) + Q(z)]. \quad (7.102)$$

It was proved in [7.36] and [7.34] that the LSP polynomials, $P(z)$ and $Q(z)$, have the following important properties:

- all zeros of LSP polynomials are on the unit circle,
- the zeros of $P(z)$ and $Q(z)$ are interlaced, and
- the minimum-phase property of $A(z)$ can be easily preserved if the first two properties are intact after quantization.

Now, define the two prediction error signals:

$$e^+(k) = x(n) - \frac{1}{2}[\mathbf{x}(n-1) + \mathbf{x}(n)]^T \mathbf{a}^+, \quad (7.103)$$

$$e^-(k) = x(n) - \frac{1}{2}[\mathbf{x}(n-1) - \mathbf{x}(n)]^T \mathbf{a}^-. \quad (7.104)$$

It is shown in [7.37] and [7.38] that the LSP polynomials, whose trivial zeroes have been removed, are equivalent

to the two optimal Wiener predictors \mathbf{a}_0^+ and \mathbf{a}_0^- . This is easy to see if we rewrite, $e^+(k)$ for example as

$$\begin{aligned} e^+(k) &= \begin{bmatrix} 1 & -\frac{\mathbf{a}^{+T}}{2} \end{bmatrix} \mathbf{x}_{L+1}(n) \\ &\quad + \begin{bmatrix} -\frac{\mathbf{a}^{+T}}{2} & 0 \end{bmatrix} \mathbf{x}_{L+1}(n) \\ &= \mathbf{g}^T \mathbf{I}_d^T \mathbf{x}_{L+1}(n), \end{aligned} \quad (7.105)$$

where

$$\mathbf{I}_d = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \quad (7.106)$$

and

$$\mathbf{g} = \begin{pmatrix} 1 \\ -\frac{\mathbf{a}^+}{2} \end{pmatrix}. \quad (7.107)$$

By minimizing the MSE, $E\{e^{+2}(k)\}$, with respect to \mathbf{g} , with the constraint $\mathbf{g}^T \mathbf{v}_1 = 1$, one can find the most important results. For readers who are interested in more details on the properties of LSP polynomials, we recommend the paper by *Bäckström* and *Magi* [7.39].

7.9 Multichannel Linear Prediction

Multichannel linear prediction can be very useful in stereo or multichannel speech compression. In an increasing number of speech or audio applications, we have at least two channels available, which are often highly correlated with each other. Therefore, it makes sense to take this interchannel correlation into account in order to obtain more-efficient compression schemes. Multichannel linear prediction is the best way to do this.

Let

$$\chi(k) = \begin{bmatrix} x_1(k) & x_2(k) & \cdots & x_M(k) \end{bmatrix}^T$$

be a real, zero-mean, stationary M -channel time series. We define the multichannel forward prediction error vector as,

$$\begin{aligned} e_{f,L}(k) &= \chi(k) - \hat{\chi}(k) \\ &= \chi(k) - \sum_{l=1}^L \mathbf{A}_{L,l} \chi(k-l) \\ &= \chi(k) - \mathbf{A}_L^T \mathbf{x}(k-1), \end{aligned} \quad (7.108)$$

where

$$\mathbf{A}_L = [\mathbf{A}_{L,1} \quad \mathbf{A}_{L,2} \quad \cdots \quad \mathbf{A}_{L,L}]^T$$

is the forward predictor matrix of size $ML \times M$, each one of the square matrices $\mathbf{A}_{L,l}$ is of size $M \times M$, and

$$\mathbf{x}(k-1) = [\chi^T(k-1) \quad \chi^T(k-2) \quad \cdots \quad \chi^T(k-L)]^T$$

is a vector of length ML . (For convenience, some of the notation used in this section is the same as that in the previous sections.)

To derive the optimal Wiener forward predictors, we need to minimize the MSE,

$$J_f(\mathbf{A}_L) = E\{\mathbf{e}_{f,L}^T(k)\mathbf{e}_{f,L}(k)\}. \quad (7.109)$$

We find the multichannel Wiener–Hopf equations:

$$\mathbf{R}_L \mathbf{A}_{o,L} = \mathbf{R}_f(1/L), \quad (7.110)$$

where

$$\begin{aligned} \mathbf{R}_L &= E\{\mathbf{x}(k-1)\mathbf{x}^T(k-1)\} \\ &= E\{\mathbf{x}(k)\mathbf{x}^T(k)\} \\ &= \begin{pmatrix} \mathbf{R}(0) & \mathbf{R}(1) & \cdots & \mathbf{R}(L-1) \\ \mathbf{R}^T(1) & \mathbf{R}(0) & \cdots & \mathbf{R}(L-2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}^T(L-1) & \mathbf{R}^T(L-2) & \cdots & \mathbf{R}(0) \end{pmatrix} \end{aligned} \quad (7.111)$$

is the block-Toeplitz covariance matrix of size $ML \times ML$,

$$\mathbf{R}(l) = E\{\chi(k)\chi^T(k-l)\}, \quad l = 0, 1, \dots, L-1,$$

$$\mathbf{R}(-l) = E\{\chi(k-l)\chi^T(k)\} = \mathbf{R}^T(l),$$

and

$$\begin{aligned} \mathbf{R}_f(1/L) &= [\mathbf{R}(1) \ \mathbf{R}(2) \ \cdots \ \mathbf{R}(L)]^T \\ &= E\{\mathbf{x}(k-1)\chi^T(k)\} \end{aligned}$$

is the intercorrelation matrix of size $ML \times M$.

Using the augmented block-Toeplitz covariance matrix of size $(ML+M) \times (ML+M)$:

$$\mathbf{R}_{L+1} = \begin{pmatrix} \mathbf{R}(0) & \mathbf{R}_f^T(1/L) \\ \mathbf{R}_f(1/L) & \mathbf{R}_L \end{pmatrix}, \quad (7.113)$$

we deduce the augmented multichannel Wiener–Hopf equations:

$$\mathbf{R}_{L+1} \begin{pmatrix} \mathbf{I}_{M \times M} \\ -\mathbf{A}_{o,L} \end{pmatrix} = \begin{pmatrix} \mathbf{E}_{f,L} \\ \mathbf{0}_{ML \times M} \end{pmatrix}, \quad (7.114)$$

where $\mathbf{I}_{M \times M}$ is the identity matrix of size $M \times M$ and

$$\begin{aligned} \mathbf{E}_{f,L} &= E\{\mathbf{e}_{f,o,L}(k)\mathbf{e}_{f,o,L}^T(k)\} \\ &= \mathbf{R}(0) - \mathbf{R}_f^T(1/L)\mathbf{A}_{o,L} \end{aligned} \quad (7.115)$$

is the forward error covariance matrix of size $M \times M$, with

$$\mathbf{e}_{f,o,L}(k) = \chi(k) - \mathbf{A}_{o,L}^T \mathbf{x}(k-1). \quad (7.116)$$

We will proceed with the same philosophy to derive important equations for the multichannel backward prediction. We define the multichannel backward prediction error vector as

$$\begin{aligned} \mathbf{e}_{b,L}(k) &= \chi(k-L) - \hat{\chi}(k-L) \\ &= \chi(k-L) - \sum_{l=1}^L \mathbf{B}_{L,l} \chi(k-l+1) \\ &= \chi(k-L) - \mathbf{B}_L^T \mathbf{x}(k), \end{aligned} \quad (7.117)$$

where

$$\mathbf{B}_L = [\mathbf{B}_{L,1} \ \mathbf{B}_{L,2} \ \cdots \ \mathbf{B}_{L,L}]^T$$

is the backward predictor matrix of size $ML \times M$ with each one of the square submatrices $\mathbf{B}_{L,l}$ being of size $M \times M$.

The minimization of the MSE,

$$J_b(\mathbf{B}_L) = E\{\mathbf{e}_{b,L}^T(k)\mathbf{e}_{b,L}(k)\}, \quad (7.118)$$

leads to the multichannel Wiener–Hopf equations for the backward prediction:

$$\mathbf{R}_L \mathbf{B}_{o,L} = \mathbf{R}_b(1/L), \quad (7.119)$$

where

$$\begin{aligned} \mathbf{R}_b(1/L) &= E\{\mathbf{x}(k)\chi^T(k-L)\} \\ &= [\mathbf{R}^T(L) \ \mathbf{R}^T(L-1) \ \cdots \ \mathbf{R}^T(1)]^T. \end{aligned} \quad (7.120)$$

By using the augmented block-Toeplitz covariance matrix:

$$\mathbf{R}_{L+1} = \begin{pmatrix} \mathbf{R}_L & \mathbf{R}_b(1/L) \\ \mathbf{R}_b^T(1/L) & \mathbf{R}(0) \end{pmatrix}, \quad (7.121)$$

we find the augmented multichannel Wiener–Hopf equations:

$$\mathbf{R}_{L+1} \begin{pmatrix} -\mathbf{B}_{o,L} \\ \mathbf{I}_{M \times M} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{ML \times M} \\ \mathbf{E}_{b,L} \end{pmatrix}, \quad (7.122)$$

where

$$\begin{aligned} \mathbf{E}_{b,L} &= E\{\mathbf{e}_{b,o,L}(k)\mathbf{e}_{b,o,L}^T(k)\} \\ &= \mathbf{R}(0) - \mathbf{R}_b^T(1/L)\mathbf{B}_{o,L} \end{aligned} \quad (7.123)$$

is the backward error covariance matrix of size $M \times M$, with

$$\mathbf{e}_{b,o,L}(k) = \chi(k-L) - \mathbf{B}_{o,L}^T \mathbf{x}(k). \quad (7.124)$$

To solve the multichannel Wiener–Hopf equations efficiently, we need to derive some important relations [7.40]. Consider the following system,

$$\begin{aligned} & \begin{pmatrix} \mathbf{R}_L & \mathbf{R}_b(1/L) \\ \mathbf{R}_b^T(1/L) & \mathbf{R}(0) \end{pmatrix} \begin{pmatrix} \mathbf{I}_{M \times M} \\ -\mathbf{A}_{o,L-1} \\ \mathbf{0}_{M \times M} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{E}_{f,L-1} \\ \mathbf{0}_{(ML-M) \times M} \\ \mathbf{K}_{f,L} \end{pmatrix}, \end{aligned} \quad (7.125)$$

where

$$\mathbf{K}_{f,L} = \mathbf{R}^T(L) - \mathbf{R}_b^T(1/L-1)\mathbf{A}_{o,L-1}. \quad (7.126)$$

Consider the other system,

$$\begin{aligned} & \begin{pmatrix} \mathbf{R}(0) & \mathbf{R}_f^T(1/L) \\ \mathbf{R}_f(1/L) & \mathbf{R}_L \end{pmatrix} \begin{pmatrix} \mathbf{0}_{M \times M} \\ -\mathbf{B}_{o,L-1} \\ \mathbf{I}_{M \times M} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{K}_{b,L} \\ \mathbf{0}_{(ML-M) \times M} \\ \mathbf{E}_{b,L-1} \end{pmatrix}, \end{aligned} \quad (7.127)$$

where

$$\mathbf{K}_{b,L} = \mathbf{R}(L) - \mathbf{R}_f^T(1/L-1)\mathbf{B}_{o,L-1}. \quad (7.128)$$

If we post-multiply both sides of (7.127) by $\mathbf{E}_{b,L-1}^{-1}\mathbf{K}_{f,L}$, we get:

$$\begin{aligned} & \mathbf{R}_{L+1} \begin{pmatrix} \mathbf{0}_{M \times M} \\ -\mathbf{B}_{o,L-1} \\ \mathbf{I}_{M \times M} \end{pmatrix} \mathbf{E}_{b,L-1}^{-1} \mathbf{K}_{f,L} \\ &= \begin{pmatrix} \mathbf{K}_{b,L} \mathbf{E}_{b,L-1}^{-1} \mathbf{K}_{f,L} \\ \mathbf{0}_{(ML-M) \times M} \\ \mathbf{K}_{f,L} \end{pmatrix}. \end{aligned} \quad (7.129)$$

Subtracting (7.129) from (7.125) and identifying the resulting system with the augmented multichannel Wiener–Hopf equations for forward prediction [eq. (7.114)], we deduce the two recursions:

$$\mathbf{E}_{f,L} = \mathbf{E}_{f,L-1} - \mathbf{K}_{b,L} \mathbf{E}_{b,L-1}^{-1} \mathbf{K}_{f,L}, \quad (7.130)$$

$$\begin{aligned} \mathbf{A}_{o,L} &= \begin{pmatrix} \mathbf{A}_{o,L-1} \\ \mathbf{0}_{M \times M} \end{pmatrix} \\ &\quad - \begin{pmatrix} \mathbf{B}_{o,L-1} \\ -\mathbf{I}_{M \times M} \end{pmatrix} \mathbf{E}_{b,L-1}^{-1} \mathbf{K}_{f,L}. \end{aligned} \quad (7.131)$$

Similarly, if we post-multiply both sides of (7.125) by $\mathbf{E}_{f,L-1}^{-1}\mathbf{K}_{b,L}$, we obtain:

$$\begin{aligned} & \mathbf{R}_{L+1} \begin{pmatrix} \mathbf{I}_{M \times M} \\ -\mathbf{A}_{o,L-1} \\ \mathbf{0}_{M \times M} \end{pmatrix} \mathbf{E}_{f,L-1}^{-1} \mathbf{K}_{b,L} \\ &= \begin{pmatrix} \mathbf{K}_{b,L} \\ \mathbf{0}_{(ML-M) \times M} \\ \mathbf{K}_{f,L} \mathbf{E}_{f,L-1}^{-1} \mathbf{K}_{b,L} \end{pmatrix}. \end{aligned} \quad (7.132)$$

Subtracting (7.132) from (7.127) and identifying the resulting system with the augmented multichannel Wiener–Hopf equations for backward prediction (7.122), we deduce the two recursions:

$$\mathbf{E}_{b,L} = \mathbf{E}_{b,L-1} - \mathbf{K}_{f,L} \mathbf{E}_{f,L-1}^{-1} \mathbf{K}_{b,L}, \quad (7.133)$$

$$\begin{aligned} \mathbf{B}_{o,L} &= \begin{pmatrix} \mathbf{0}_{M \times M} \\ \mathbf{B}_{o,L-1} \end{pmatrix} \\ &\quad - \begin{pmatrix} -\mathbf{I}_{M \times M} \\ \mathbf{A}_{o,L-1} \end{pmatrix} \mathbf{E}_{f,L-1}^{-1} \mathbf{K}_{b,L}. \end{aligned} \quad (7.134)$$

Relations (7.130), (7.131), (7.133), and (7.134) were independently discovered by *Whittle* [7.41] and *Wiggins* and *Robinson* [7.42].

Another important relation needs to be found. Indeed, using (7.116) and (7.124), we can easily verify,

$$\mathbf{E}\{\mathbf{e}_{f,o,L-1}(k)\mathbf{e}_{b,o,L-1}^T(k-1)\} = \mathbf{K}_{b,L}, \quad (7.135)$$

$$\mathbf{E}\{\mathbf{e}_{b,o,L-1}(k-1)\mathbf{e}_{f,o,L-1}^T(k)\} = \mathbf{K}_{f,L}, \quad (7.136)$$

which implies that,

$$\mathbf{K}_{b,L} = \mathbf{K}_{f,L}^T. \quad (7.137)$$

Table 7.2 summarizes the Levinson–Wiggins–Robinson algorithm [7.42–44], which is a generalization of the Levinson–Durbin algorithm to the multichannel case.

Table 7.2 Levinson–Wiggins–Robinson algorithm

Initialization: $\mathbf{E}_{f,0} = \mathbf{E}_{b,0} = \mathbf{R}(0)$
For $1 \leq l \leq L$
 $\mathbf{K}_{b,l} = \mathbf{R}(l) - \mathbf{R}_f^T(1:l-1)\mathbf{B}_{o,l-1}$
 $\mathbf{A}_{o,l} = \begin{bmatrix} \mathbf{A}_{o,l-1} \\ \mathbf{0}_{M \times M} \end{bmatrix} - \begin{bmatrix} \mathbf{B}_{o,l-1} \\ -\mathbf{I}_{M \times M} \end{bmatrix} \mathbf{E}_{b,l-1}^{-1} \mathbf{K}_{b,l}^T$
 $\mathbf{B}_{o,l} = \begin{bmatrix} \mathbf{0}_{M \times M} \\ \mathbf{B}_{o,l-1} \end{bmatrix} - \begin{bmatrix} -\mathbf{I}_{M \times M} \\ \mathbf{A}_{o,l-1} \end{bmatrix} \mathbf{E}_{f,l-1}^{-1} \mathbf{K}_{b,l}$
 $\mathbf{E}_{f,l} = \mathbf{E}_{f,l-1} - \mathbf{K}_{b,l} \mathbf{E}_{b,l-1}^{-1} \mathbf{K}_{b,l}^T$
 $\mathbf{E}_{b,l} = \mathbf{E}_{b,l-1} - \mathbf{K}_{b,l}^T \mathbf{E}_{f,l-1}^{-1} \mathbf{K}_{b,l}$

7.10 Conclusions

In this chapter, we have tried to present the most important results in linear prediction for speech. We have explained the principle of forward linear prediction and have shown that the optimal prediction error signal tends to be a white signal. We have extended the principle of forward linear prediction to backward linear prediction and derived the Cholesky factorization of the inverse correlation matrix. We have developed the classical Levinson–Durbin algorithm, which is a very efficient way to solve the Wiener–Hopf equations for the forward

and backward prediction coefficients. We have explained the idea behind the lattice predictor. We have shown how the spectrum of a speech signal can easily be estimated thanks to the prediction coefficients. We have given some notions of linear interpolation and have demonstrated how the condition number of the correlation matrix is related to the optimal interpolators. We have also presented some notions of line spectrum pair polynomials. Finally, in the last section, we have generalized some of these ideas to the multichannel case.

References

- 7.1 B.S. Atal: The history of linear prediction, *IEEE Signal Proc. Mag.* **23**, 154–161 (2006)
- 7.2 L.R. Rabiner, R.W. Schaffer: *Digital Processing of Speech Signals* (Prentice–Hall, Englewood Cliffs 1976)
- 7.3 J.D. Markel, A.H. Gray Jr.: *Linear Prediction of Speech* (Springer, New York 1976)
- 7.4 J. Makhoul: Linear prediction: a tutorial review, *Proc. IEEE* **63**, 561–580 (1975)
- 7.5 J.W. Picone: Signal modeling techniques in speech recognition, *Proc. IEEE* **81**, 1215–1247 (1993)
- 7.6 S.R. Quackenbush, T.P. Barnwell, M.A. Clements: *Objective Measures of Speech Quality* (Prentice–Hall, Englewood Cliffs 1988)
- 7.7 Y. Huang, J. Benesty, J. Chen: *Acoustic MIMO Signal Processing* (Springer, New York 2006)
- 7.8 F. Itakura, S. Saito: A statistical method for estimation of speech spectral density and formant frequencies, *Electron. Commun. Jpn.* **53(A)**, 36–43 (1970)
- 7.9 G. Chen, S.N. Koh, I.Y. Soon: Enhanced Itakura measure incorporating masking properties of human auditory system, *Signal Process.* **83**, 1445–1456 (2003)
- 7.10 M.G. Bellanger: *Adaptive Digital Filters and Signal analysis* (Marcel Dekker, New York 1987)
- 7.11 S. Haykin: *Adaptive Filter Theory*, 4th edn. (Prentice–Hall, Upper Saddle River 2002)
- 7.12 C.W. Therrien: On the relation between triangular matrix decomposition and linear prediction, *Proc. IEEE* **71**, 1459–1460 (1983)
- 7.13 N. Levinson: The Wiener RMS (root mean square) error criterion in filter design and prediction, *J. Math. Phys.* **25(4)**, 261–278 (1947), Also Appendix B, in *N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series* (MIT, Cambridge 1949)
- 7.14 J. Durbin: Efficient estimation of parameters in moving–average models, *Biometrika* **46**, 306–316 (1959), Parts 1 and 2
- 7.15 J. Durbin: The fitting of time–series models, *Rev. Inst. Int. Stat.* **28(3)**, 233–243 (1960)
- 7.16 J.P. Burg: *Maximum Entropy Spectral Analysis Ph.D. Dissertation* (Stanford Univ., Stanford 1975)
- 7.17 P. Delsarte, Y.V. Genin: The split Levinson algorithm, *IEEE Trans. Acoust. Speech* **ASSP-34**, 470–478 (1986)
- 7.18 R. Kumar: A fast algorithm for solving a Toeplitz system of equations, *IEEE Trans. Acoust. Speech* **ASSP-33**, 254–265 (1985)
- 7.19 J. Makhoul: Stable and efficient lattice methods for linear prediction, *IEEE Trans. Acoust. Speech* **ASSP-25**, 423–428 (1977)
- 7.20 F. Itakura, S. Saito: Digital filtering techniques for speech analysis and synthesis, *Proc. 7th Int. Conf. Acoust.* **25(C-1)**, 261–264 (1971)
- 7.21 S.W. Lang, J.H. McClellan: A simple proof of stability for all–pole linear prediction models, *Proc. IEEE* **67**, 860–861 (1979)
- 7.22 M.H. Hayes: *Statistical Digital Signal Processing and Modeling* (Wiley, New York 1996)
- 7.23 J. Makhoul: Spectral analysis of speech by linear prediction, *IEEE Trans. Acoust. Speech* **AU-21(3)**, 140–148 (1973)
- 7.24 S.M. Kay, S.L. Marple Jr.: Spectrum analysis – a modern perspective, *Proc. IEEE* **69**, 1380–1419 (1981)
- 7.25 J. M. Cadzow: Spectral estimation: an overdetermined rational model equation approach, *Proc. IEEE* **70**, 907–939 (1982)
- 7.26 S. Kay: Some results in linear interpolation theory, *IEEE Trans. Acoust. Speech* **ASSP-31**, 746–749 (1983)
- 7.27 B. Picinbono, J.–M. Kerilis: Some properties of prediction and interpolation errors, *IEEE Trans. Acoust. Speech* **ASSP-36**, 525–531 (1988)
- 7.28 J. Benesty, T. Gaensler: New insights into the RLS algorithm, *EURASIP J. Appl. Si. Pr.* **2004**, 331–339 (2004)
- 7.29 G.H. Golub, C.F. Van Loan: *Matrix Computations* (Johns Hopkins Univ. Press, Baltimore 1996)

- 7.30 J. Benesty, T. Gaensler: Computation of the condition number of a non-singular symmetric Toeplitz matrix with the Levinson-Durbin algorithm, *IEEE Trans. Signal Process.* **54**, 2362–2364 (2006)
- 7.31 C.K. Coursey, J.A. Stuller: Linear interpolation lattice, *IEEE Trans. Signal Process.* **39**, 965–967 (1991)
- 7.32 M.R.K. Khansari, A. Leon-Garcia: A fast algorithm for optimal linear interpolation, *IEEE Trans. Signal Process.* **41**, 2934–2937 (1993)
- 7.33 F. Itakura: Line spectrum representation of linear predictive coefficients of speech signal, *J. Acoust. Soc. Am.* **57**(1), 35 (1975)
- 7.34 F.K. Soong, B.-H. Juang: Line spectrum pair (LSP) and speech data compression, *Proc. ICASSP* (1984) pp.1.10.1–1.10.4
- 7.35 F.K. Soong, B.-H. Juang: Optimal quantization of LSP parameters, *IEEE Trans. Speech Audio Process.* **1**, 15–24 (1993)
- 7.36 S. Sagayama: Stability condition of LSP speech synthesis digital filter, *Proc. Acoust. Soc. Jpn.* (1982) pp.153–154, (in Japanese)
- 7.37 B. Kleijn, T. Bäckström, P. Alku: On line spectral frequencies, *IEEE Signal Proc. Lett.* **10**, 75–77 (2003)
- 7.38 T. Bäckström, P. Alku, T. Paatero, B. Kleijn: A time-domain interpretation for the LSP decomposition, *IEEE Trans. Speech Audio Process.* **12**, 554–560 (2004)
- 7.39 T. Bäckström, C. Magi: Properties of line spectrum pair polynomials – A review, *Elsevier Signal Process.* **86**, 3286–3298 (2006)
- 7.40 T. Kailath: A view of three decades of linear filtering theory, *IEEE Trans. Inf. Theory* **IT-20**, 146–181 (1974)
- 7.41 P. Whittle: On the fitting of multivariate autoregressions and the approximate canonical factorization of a spectral density matrix, *Biometrika* **50**, 129–134 (1963)
- 7.42 R.A. Wiggins, E.A. Robinson: Recursive solution to the multichannel filtering problem, *J. Geophys. Res.* **70**, 1885–1891 (1965)
- 7.43 O.N. Strand: Multichannel complex maximum entropy (autoregressive) spectral analysis, *IEEE T. Automat. Contr.* **AC-22**, 634–640 (1977)
- 7.44 P. Delsarte, Y.V. Genin: Multichannel singular predictor polynomials, *IEEE Trans. Circuits Syst.* **35**, 190–200 (1988)