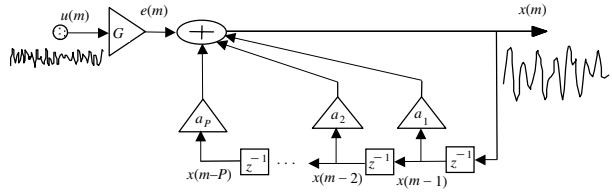
Advanced Digital Signal Processing and Noise Reduction, Second Edition.
Saeed V. Vaseghi
Copyright © 2000 John Wiley & Sons Ltd
ISBNs: 0-471-62692-9 (Hardback): 0-470-84162-1 (Electronic)

8



LINEAR PREDICTION MODELS

Linear prediction modelling is used in a diverse area of applications, such as data forecasting, speech coding, video coding, speech recognition, model-based spectral analysis, model-based interpolation, signal restoration, and impulse/step event detection. In the statistical literature, linear prediction models are often referred to as autoregressive (AR) processes. In this chapter, we introduce the theory of linear prediction modelling and consider efficient methods for the computation of predictor coefficients. We study the forward, backward and lattice predictors, and consider various methods for the formulation and calculation of predictor coefficients, including the least square error and maximum a posteriori methods. For the modelling of signals with a quasi-periodic structure, such as voiced speech, an extended linear predictor that simultaneously utilizes the short and long-term correlation structures is introduced. We study sub-band linear predictors that are particularly useful for sub-band processing of noisy signals. Finally, the application of linear prediction in enhancement of noisy speech is considered. Further applications of linear prediction models in this book are in Chapter 11 on the interpolation of a sequence of lost samples, and in Chapters 12 and 13 on the detection and removal of impulsive noise and transient noise pulses.

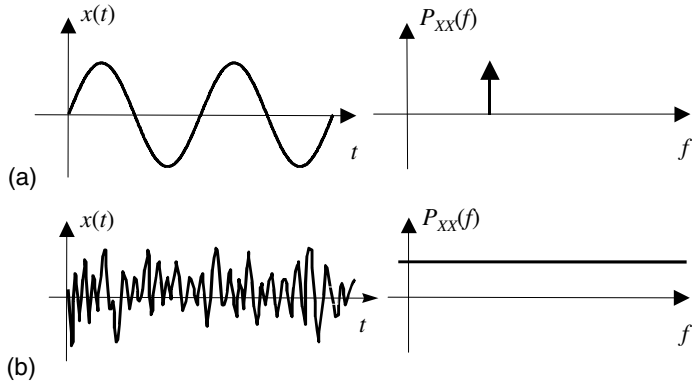


Figure 8.1 The concentration or spread of power in frequency indicates the predictable or random character of a signal: (a) a predictable signal; (b) a random signal.

8.1 Linear Prediction Coding

The success with which a signal can be predicted from its past samples depends on the autocorrelation function, or equivalently the bandwidth and the power spectrum, of the signal. As illustrated in Figure 8.1, in the time domain, a predictable signal has a smooth and correlated fluctuation, and in the frequency domain, the energy of a predictable signal is concentrated in narrow band/s of frequencies. In contrast, the energy of an unpredictable signal, such as a white noise, is spread over a wide band of frequencies.

For a signal to have a capacity to convey information it must have a degree of randomness. Most signals, such as speech, music and video signals, are partially predictable and partially random. These signals can be modelled as the output of a filter excited by an uncorrelated input. The random input models the unpredictable part of the signal, whereas the filter models the predictable structure of the signal. The aim of linear prediction is to model the mechanism that introduces the correlation in a signal.

Linear prediction models are extensively used in speech processing, in low bit-rate speech coders, speech enhancement and speech recognition. Speech is generated by inhaling air and then exhaling it through the glottis and the vocal tract. The noise-like air, from the lung, is modulated and shaped by the vibrations of the glottal cords and the resonance of the vocal tract. Figure 8.2 illustrates a source-filter model of speech. The source models the lung, and emits a random input excitation signal which is filtered by a pitch filter.

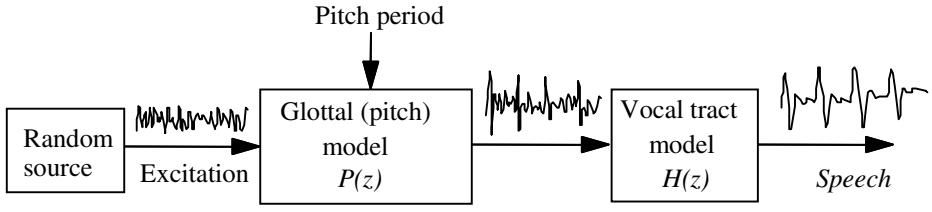


Figure 8.2 A source–filter model of speech production.

The pitch filter models the vibrations of the glottal cords, and generates a sequence of quasi-periodic excitation pulses for voiced sounds as shown in Figure 8.2. The pitch filter model is also termed the "long-term predictor" since it models the correlation of each sample with the samples a pitch period away. The main source of correlation and power in speech is the vocal tract. The vocal tract is modelled by a linear predictor model, which is also termed the "short-term predictor", because it models the correlation of each sample with the few preceding samples. In this section, we study the short-term linear prediction model. In Section 8.3, the predictor model is extended to include long-term pitch period correlations.

A linear predictor model forecasts the amplitude of a signal at time m, $x(m)$, using a linearly weighted combination of P past samples [$x(m-1)$, $x(m-2)$, ..., $x(m-P)$] as

$$\hat{x}(m) = \sum_{k=1}^{P} a_k x(m-k) \tag{8.1}$$

where the integer variable m is the discrete time index, $\hat{x}(m)$ is the prediction of $x(m)$, and a_k are the predictor coefficients. A block-diagram implementation of the predictor of Equation (8.1) is illustrated in Figure 8.3.

The prediction error $e(m)$, defined as the difference between the actual sample value $x(m)$ and its predicted value $\hat{x}(m)$, is given by

$$\begin{aligned} e(m) &= x(m) - \hat{x}(m) \\ &= x(m) - \sum_{k=1}^{P} a_k x(m-k) \end{aligned} \tag{8.2}$$

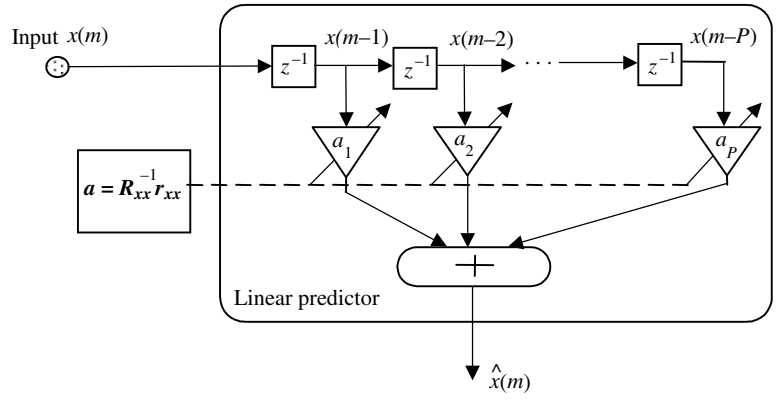


Figure 8.3 Block-diagram illustration of a linear predictor.

For information-bearing signals, the prediction error $e(m)$ may be regarded as the information, or the innovation, content of the sample $x(m)$. From Equation (8.2) a signal generated, or modelled, by a linear predictor can be described by the following feedback equation

$$x(m) = \sum_{k=1}^{P} a_k x(m-k) + e(m) \tag{8.3}$$

Figure 8.4 illustrates a linear predictor model of a signal $x(m)$. In this model, the random input excitation (i.e. the prediction error) is $e(m)=Gu(m)$, where $u(m)$ is a zero-mean, unit-variance random signal, and G, a gain term, is the square root of the variance of $e(m)$:

$$G=\left(\mathcal{E}\left[e^2(m)\right]\right)^{1/2} \tag{8.4}$$

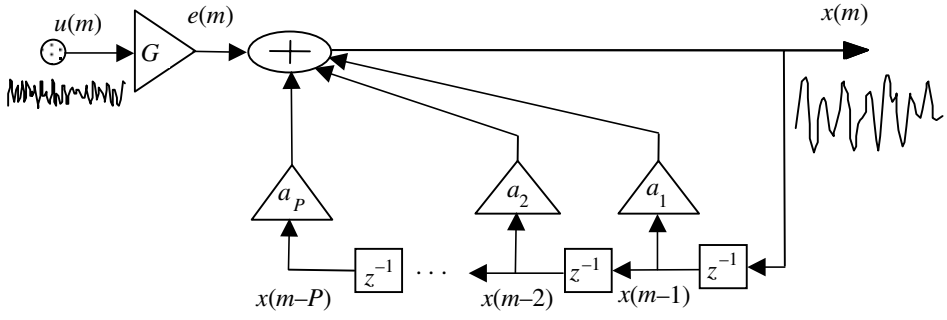


Figure 8.4 Illustration of a signal generated by a linear predictive model.

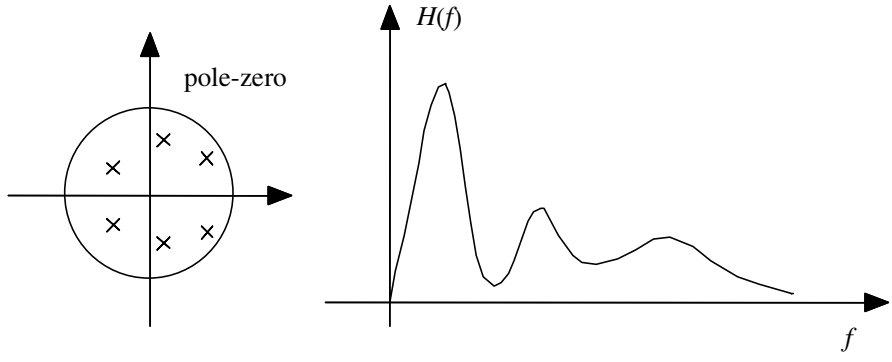


Figure 8.5 The pole–zero position and frequency response of a linear predictor.

where $\mathcal{E}[\cdot]$ is an averaging, or expectation, operator. Taking the z-transform of Equation (8.3) shows that the linear prediction model is an all-pole digital filter with z-transfer function

$$H(z)=\frac{X(z)}{U(z)}=\frac{G}{1-\sum_{k=1}^{P} a_k z^{-k}} \tag{8.5}$$

In general, a linear predictor of order P has $P/2$ complex pole pairs, and can model up to $P/2$ resonance of the signal spectrum as illustrated in Figure 8.5. Spectral analysis using linear prediction models is discussed in Chapter 9.

8.1.1 Least Mean Square Error Predictor

The "best" predictor coefficients are normally obtained by minimising a mean square error criterion defined as

$$\begin{aligned}\mathcal{E}[e^2(m)]&=\mathcal{E}\left[\left(x(m)-\sum_{k=1}^{P} a_k x(m-k)\right)^2\right]\\&=\mathcal{E}[x^2(m)]-2\sum_{k=1}^{P} a_k\,\mathcal{E}[x(m)x(m-k)]+\sum_{k=1}^{P} a_k \sum_{j=1}^{P} a_j\,\mathcal{E}[x(m-k)x(m-j)]\\&=r_{xx}(0)-2\boldsymbol{r}_{\boldsymbol{xx}}^{\mathrm{T}}\boldsymbol{a}+\boldsymbol{a}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{xx}}\boldsymbol{a}\end{aligned} \tag{8.6}$$

where $\boldsymbol{R}_{xx} = \mathcal{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}]$ is the autocorrelation matrix of the input vector $\boldsymbol{x}^{\mathrm{T}}=[x(m-1), x(m-2), \ldots, x(m-P)]$, $\boldsymbol{r}_{xx}=\mathcal{E}[x(m)\boldsymbol{x}]$ is the autocorrelation vector and $\boldsymbol{a}^{\mathrm{T}}=[a_1, a_2, \ldots, a_P]$ is the predictor coefficient vector. From Equation (8.6), the gradient of the mean square prediction error with respect to the predictor coefficient vector $\boldsymbol{a}$ is given by

$$\frac{\partial}{\partial \boldsymbol{a}}\mathcal{E}[e^2(m)] = -2\boldsymbol{r}_{xx}^{\mathrm{T}} + 2\boldsymbol{a}^{\mathrm{T}}\boldsymbol{R}_{xx} \tag{8.7}$$

where the gradient vector is defined as

$$\frac{\partial}{\partial \boldsymbol{a}} = \left(\frac{\partial}{\partial a_1}, \frac{\partial}{\partial a_2}, \ldots, \frac{\partial}{\partial a_{\mathrm{P}}}\right)^{\mathrm{T}} \tag{8.8}$$

The least mean square error solution, obtained by setting Equation (8.7) to zero, is given by

$$\boldsymbol{R}_{xx}\,\boldsymbol{a} = \boldsymbol{r}_{xx} \tag{8.9}$$

From Equation (8.9) the predictor coefficient vector is given by

$$\boldsymbol{a} = \boldsymbol{R}_{xx}^{-1}\boldsymbol{r}_{xx} \tag{8.10}$$

Equation (8.10) may also be written in an expanded form as

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \end{pmatrix} = \begin{pmatrix} r_{xx}(0) & r_{xx}(1) & r_{xx}(2) & \cdots & r_{xx}(P-1) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(P-2) \\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(P-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{xx}(P-1) & r_{xx}(P-2) & r_{xx}(P-3) & \cdots & r_{xx}(0) \end{pmatrix}^{-1} \begin{pmatrix} r_{xx}(1) \\ r_{xx}(2) \\ r_{xx}(3) \\ \vdots \\ r_{xx}(P) \end{pmatrix} \tag{8.11}$$

An alternative formulation of the least square error problem is as follows. For a signal block of N samples $[x(0), \ldots, x(N-1)]$, we can write a set of N linear prediction error equations as

$$\begin{pmatrix} e(0) \\ e(1) \\ e(2) \\ \vdots \\ e(N-1) \end{pmatrix} = \begin{pmatrix} x(0) \\ x(1) \\ x(2) \\ \vdots \\ x(N-1) \end{pmatrix} - \begin{pmatrix} x(-1) & x(-2) & x(-3) & \cdots & x(-P) \\ x(0) & x(-1) & x(-2) & \cdots & x(1-P) \\ x(1) & x(0) & x(-1) & \cdots & x(2-P) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x(N-2) & x(N-3) & x(N-4) & \cdots & x(N-P-1) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \end{pmatrix} \tag{8.12}$$

where $\boldsymbol{x}^{\mathrm{T}} = [x(-1), ..., x(-P)]$ is the initial vector. In a compact vector/matrix notation Equation (8.12) can be written as

$$\boldsymbol{e} = \boldsymbol{x} - \boldsymbol{X}\boldsymbol{a} \tag{8.13}$$

Using Equation (8.13), the sum of squared prediction errors over a block of N samples can be expressed as

$$\boldsymbol{e}^{\mathrm{T}}\boldsymbol{e} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - 2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{a} - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{a} \tag{8.14}$$

The least squared error predictor is obtained by setting the derivative of Equation (8.14) with respect to the parameter vector $\boldsymbol{a}$ to zero:

$$\frac{\partial \boldsymbol{e}^{\mathrm{T}}\boldsymbol{e}}{\partial \boldsymbol{a}} = -2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{X} - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = 0 \tag{8.15}$$

From Equation (8.15), the least square error predictor is given by

$$\boldsymbol{a} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{x}\right) \tag{8.16}$$

A comparison of Equations (8.11) and (8.16) shows that in Equation (8.16) the autocorrelation matrix and vector of Equation (8.11) are replaced by the time-averaged estimates as

$$\hat{r}_{xx}(m) = \frac{1}{N}\sum_{k=0}^{N-1} x(k)x(k-m) \tag{8.17}$$

Equations (8.11) and (8.16) may be solved efficiently by utilising the regular Toeplitz structure of the correlation matrix $\boldsymbol{R}_{xx}$. In a Toeplitz matrix,

all the elements on a left–right diagonal are equal. The correlation matrix is also cross-diagonal symmetric. Note that altogether there are only P+1 unique elements [$r_{xx}(0)$, $r_{xx}(1)$, . . . , $r_{xx}(P)$] in the correlation matrix and the cross-correlation vector. An efficient method for solution of Equation (8.10) is the Levinson–Durbin algorithm, introduced in Section 8.2.2.

8.1.2 The Inverse Filter: Spectral Whitening

The all-pole linear predictor model, in Figure 8.4, shapes the spectrum of the input signal by transforming an uncorrelated excitation signal $u(m)$ to a correlated output signal $x(m)$. In the frequency domain the input–output relation of the all-pole filter of Figure 8.6 is given by

$$X(f)=\frac{G\,U(f)}{A(f)}=\frac{E(f)}{1-\sum_{k=1}^{P} a_k\, e^{-\mathrm{j}2\pi fk}} \tag{8.18}$$

where $X(f)$, $E(f)$ and $U(f)$ are the spectra of $x(m)$, $e(m)$ and $u(m)$ respectively, G is the input gain factor, and $A(f)$ is the frequency response of the inverse predictor. As the excitation signal $e(m)$ is assumed to have a flat spectrum, it follows that the shape of the signal spectrum $X(f)$ is due to the frequency response $1/A(f)$ of the all-pole predictor model. The inverse linear predictor,

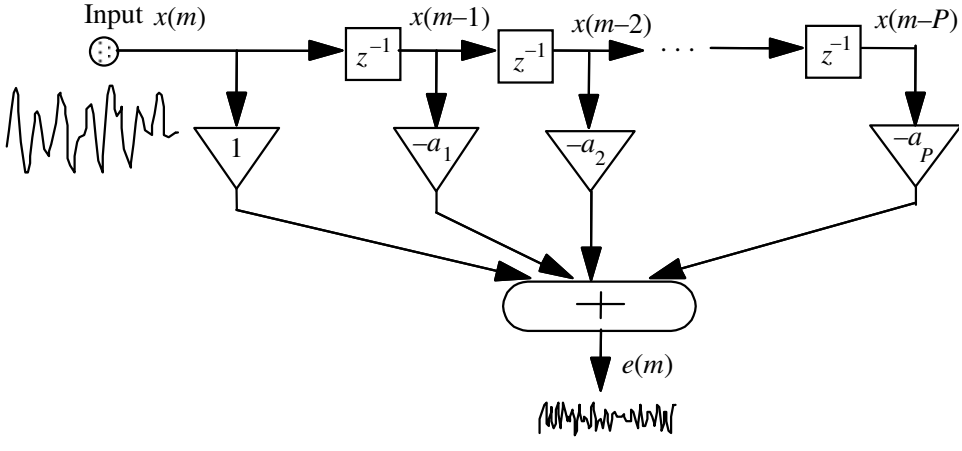


Figure 8.6 Illustration of the inverse (or whitening) filter.

as the name implies, transforms a correlated signal $x(m)$ back to an uncorrelated flat-spectrum signal $e(m)$. The inverse filter, also known as the prediction error filter, is an all-zero finite impulse response filter defined as

$$\begin{aligned} e(m) &= x(m) - \hat{x}(m) \\ &= x(m) - \sum_{k=1}^{P} a_k x(m-k) \\ &= (\boldsymbol{a}^{\text{inv}})^{\text{T}} \boldsymbol{x} \end{aligned} \tag{8.19}$$

where the inverse filter $(\boldsymbol{a}^{\text{inv}})^{\text{T}} = [1, -a_1, \ldots, -a_P] = [1, -\boldsymbol{a}]$, and $\boldsymbol{x}^{\text{T}} = [x(m), \ldots, x(m-P)]$. The z-transfer function of the inverse predictor model is given by

$$A(z) = 1 - \sum_{k=1}^{P} a_k \, z^{-k} \tag{8.20}$$

A linear predictor model is an all-pole filter, where the poles model the resonance of the signal spectrum. The inverse of an all-pole filter is an all-zero filter, with the zeros situated at the same positions in the pole–zero plot as the poles of the all-pole filter, as illustrated in Figure 8.7. Consequently, the zeros of the inverse filter introduce anti-resonances that cancel out the resonances of the poles of the predictor. The inverse filter has the effect of flattening the spectrum of the input signal, and is also known as a spectral whitening, or decorrelation, filter.

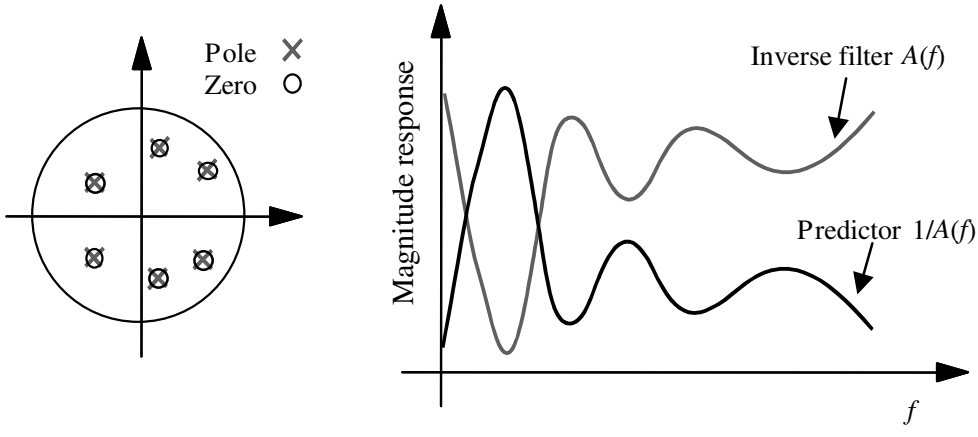


Figure 8.7 Illustration of the pole-zero diagram, and the frequency responses of an all-pole predictor and its all-zero inverse filter.

8.1.3 The Prediction Error Signal

The prediction error signal is in general composed of three components:

(a) the input signal, also called the excitation signal;
(b) the errors due to the modelling inaccuracies;
(c) the noise.

The mean square prediction error becomes zero only if the following three conditions are satisfied: (a) the signal is deterministic, (b) the signal is correctly modelled by a predictor of order P, and (c) the signal is noise-free. For example, a mixture of $P/2$ sine waves can be modelled by a predictor of order P, with zero prediction error. However, in practice, the prediction error is nonzero because information bearing signals are random, often only approximately modelled by a linear system, and usually observed in noise. The least mean square prediction error, obtained from substitution of Equation (8.9) in Equation (8.6), is

$$E^{(P)} = \mathcal{E}[e^2(m)] = r_{xx}(0) - \sum_{k=1}^{P} a_k r_{xx}(k) \tag{8.21}$$

where $E^{(P)}$ denotes the prediction error for a predictor of order P. The prediction error decreases, initially rapidly and then slowly, with increasing predictor order up to the correct model order. For the correct model order, the signal $e(m)$ is an uncorrelated zero-mean random process with an autocorrelation function defined as

$$\mathcal{E}[e(m)e(m-k)] = \begin{cases} \sigma_e^2 = G^2 & \text{if } m = k \\ 0 & \text{if } m \neq k \end{cases} \tag{8.22}$$

where σ_e^2 is the variance of $e(m)$.

8.2 Forward, Backward and Lattice Predictors

The forward predictor model of Equation (8.1) predicts a sample $x(m)$ from a linear combination of P *past* samples $x(m-1)$, $x(m-2)$, . . .,$x(m-P)$.

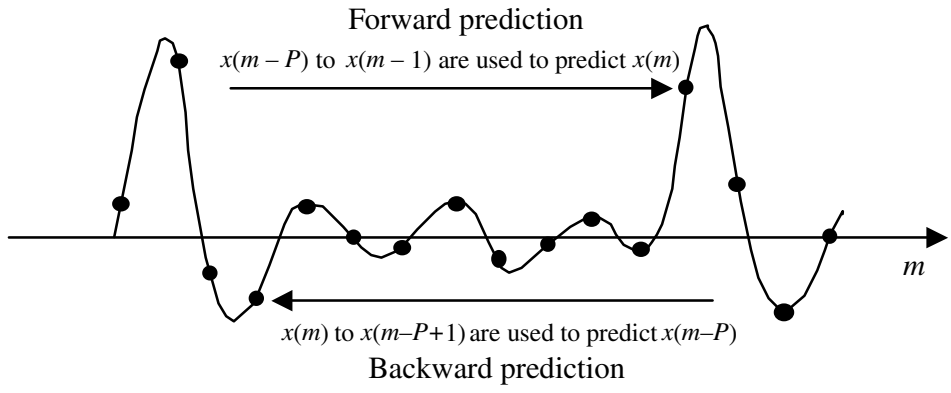


Figure 8.8 Illustration of forward and backward predictors.

Similarly, as shown in Figure 8.8, we can define a backward predictor, that predicts a sample $x(m-P)$ from P *future* samples $x(m-P+1), \ldots, x(m)$ as

$$\hat{x}(m-P)=\sum_{k=1}^{P} c_k\, x(m-k+1) \tag{8.23}$$

The backward prediction error is defined as the difference between the actual sample and its predicted value:

$$\begin{aligned} b(m) &= x(m-P)-\hat{x}(m-P) \\ &= x(m-P)-\sum_{k=1}^{P} c_k\, x(m-k+1) \end{aligned} \tag{8.24}$$

From Equation (8.24), a signal generated by a backward predictor is given by

$$x(m-P)=\sum_{k=1}^{P} c_k\, x(m-k+1)+b(m) \tag{8.25}$$

The coefficients of the least square error backward predictor, obtained in a similar method to that of the forward predictor in Section 8.1.1, are given by

$$\begin{pmatrix} r_{xx}(0) & r_{xx}(1) & r_{xx}(2) & \dots & r_{xx}(P-1) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(1) & \dots & r_{xx}(P-2) \\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0) & \dots & r_{xx}(P-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{xx}(P-1) & r_{xx}(P-2) & r_{xx}(P-3) & \dots & r_{xx}(0) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_P \end{pmatrix} = \begin{pmatrix} r_{xx}(P) \\ r_{xx}(P-1) \\ r_{xx}(P-2) \\ \vdots \\ r_{xx}(1) \end{pmatrix} \tag{8.26}$$

Note that the main difference between Equations (8.26) and (8.11) is that the correlation vector on the right-hand side of the backward predictor, Equation (8.26) is upside-down compared with the forward predictor, Equation (8.11). Since the correlation matrix is Toeplitz and symmetric, Equation (8.11) for the forward predictor may be rearranged and rewritten in the following form:

$$\begin{pmatrix} r_{xx}(0) & r_{xx}(1) & r_{xx}(2) & \dots & r_{xx}(P-1) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(1) & \dots & r_{xx}(P-2) \\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0) & \dots & r_{xx}(P-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{xx}(P-1) & r_{xx}(P-2) & r_{xx}(P-3) & \dots & r_{xx}(0) \end{pmatrix} \begin{pmatrix} a_P \\ a_{P-1} \\ a_{P-2} \\ \vdots \\ a_1 \end{pmatrix} = \begin{pmatrix} r_{xx}(P) \\ r_{xx}(P-1) \\ r_{xx}(P-2) \\ \vdots \\ r_{xx}(1) \end{pmatrix} \tag{8.27}$$

A comparison of Equations (8.27) and (8.26) shows that the coefficients of the backward predictor are the time-reversed versions of those of the forward predictor

$$\boldsymbol{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_P \end{pmatrix} = \begin{pmatrix} a_P \\ a_{P-1} \\ a_{P-2} \\ \vdots \\ a_1 \end{pmatrix} = \boldsymbol{a}^{\mathrm{B}} \tag{8.28}$$

where the vector $\boldsymbol{a}^{\mathrm{B}}$ is the reversed version of the vector $\boldsymbol{a}$. The relation between the backward and forward predictors is employed in the Levinson–Durbin algorithm to derive an efficient method for calculation of the predictor coefficients as described in Section 8.2.2.

8.2.1 Augmented Equations for Forward and Backward Predictors

The inverse forward predictor coefficient vector is $[1, -a_1, ..., -a_P]=[1, -\boldsymbol{a}^{\mathrm{T}}]$. Equations (8.11) and (8.21) may be combined to yield a matrix equation for the inverse forward predictor coefficients:

$$\begin{pmatrix} r(0) & \boldsymbol{r}_{\boldsymbol{xx}}^{\mathrm{T}} \\ \boldsymbol{r}_{\boldsymbol{xx}} & \boldsymbol{R}_{\boldsymbol{xx}} \end{pmatrix} \begin{pmatrix} 1 \\ -\boldsymbol{a} \end{pmatrix} = \begin{pmatrix} E^{(P)} \\ \mathbf{0} \end{pmatrix} \tag{8.29}$$

Equation (8.29) is called the augmented forward predictor equation. Similarly, for the inverse backward predictor, we can define an augmented backward predictor equation as

$$\begin{pmatrix} \boldsymbol{R}_{\boldsymbol{xx}} & \boldsymbol{r}_{\boldsymbol{xx}}^{\mathrm{B}} \\ \boldsymbol{r}_{\boldsymbol{xx}}^{\mathrm{BT}} & r(0) \end{pmatrix} \begin{pmatrix} -\boldsymbol{a}^{\mathrm{B}} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ E^{(P)} \end{pmatrix} \tag{8.30}$$

where $\boldsymbol{r}_{\boldsymbol{xx}}^{\mathrm{T}} = [r_{xx}(1), \cdots, r_{xx}(P)]$ and $\boldsymbol{r}_{\boldsymbol{xx}}^{\mathrm{BT}} = [r_{xx}(P), \ldots, r_{xx}(1)]$. Note that the superscript BT denotes backward and transposed. The augmented forward and backward matrix Equations (8.29) and (8.30) are used to derive an order-update solution for the linear predictor coefficients as follows.

8.2.2 Levinson–Durbin Recursive Solution

The Levinson–Durbin algorithm is a recursive order-update method for calculation of linear predictor coefficients. A forward-prediction error filter of order i can be described in terms of the forward and backward prediction error filters of order $i-1$ as

$$\begin{pmatrix} 1 \\ -a_1^{(i)} \\ \vdots \\ -a_{i-1}^{(i)} \\ -a_i^{(i)} \end{pmatrix} = \begin{pmatrix} 1 \\ -a_1^{(i-1)} \\ \vdots \\ -a_{i-1}^{(i-1)} \\ 0 \end{pmatrix} + k_i \begin{pmatrix} 0 \\ -a_{i-1}^{(i-1)} \\ \vdots \\ -a_1^{(i-1)} \\ 1 \end{pmatrix} \tag{8.31}$$

or in a more compact vector notation as

$$\begin{pmatrix} 1 \\ -\boldsymbol{a}^{(i)} \end{pmatrix} = \begin{pmatrix} 1 \\ -\boldsymbol{a}^{(i-1)} \\ 0 \end{pmatrix} + k_i \begin{pmatrix} 0 \\ -\boldsymbol{a}^{(i-1)\mathrm{B}} \\ 1 \end{pmatrix} \tag{8.32}$$

where k_i is called the reflection coefficient. The proof of Equation (8.32) and the derivation of the value of the reflection coefficient for k_i follows shortly. Similarly, a backward prediction error filter of order i is described in terms of the forward and backward prediction error filters of order i–1 as

$$\begin{pmatrix} -\boldsymbol{a}^{(i)\mathrm{B}} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -\boldsymbol{a}^{(i-1)\mathrm{B}} \\ 1 \end{pmatrix} + k_i \begin{pmatrix} 1 \\ -\boldsymbol{a}^{(i-1)} \\ 0 \end{pmatrix} \tag{8.33}$$

To prove the order-update Equation (8.32) (or alternatively Equation (8.33)), we multiply both sides of the equation by the $(i+1)\times(i+1)$ augmented matrix $\boldsymbol{R}_{\boldsymbol{xx}}^{(i+1)}$ and use the equality

$$\boldsymbol{R}_{\boldsymbol{xx}}^{(i+1)} = \begin{pmatrix} \boldsymbol{R}_{\boldsymbol{xx}}^{(i)} & \boldsymbol{r}_{\boldsymbol{xx}}^{(i)\mathrm{B}} \\ \boldsymbol{r}_{\boldsymbol{xx}}^{(i)\mathrm{BT}} & r_{xx}(0) \end{pmatrix} = \begin{pmatrix} r_{xx}(0) & \boldsymbol{r}_{\boldsymbol{xx}}^{(i)\mathrm{T}} \\ \boldsymbol{r}_{\boldsymbol{xx}}^{(i)} & \boldsymbol{R}_{\boldsymbol{xx}}^{(i)} \end{pmatrix} \tag{8.34}$$

to obtain

$$\begin{pmatrix} \boldsymbol{R}_{\boldsymbol{xx}}^{(i)} & \boldsymbol{r}_{\boldsymbol{xx}}^{(i)\mathrm{B}} \\ \boldsymbol{r}_{\boldsymbol{xx}}^{(i)\mathrm{BT}} & r_{xx}(0) \end{pmatrix} \begin{pmatrix} 1 \\ -\boldsymbol{a}^{(i)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{R}_{\boldsymbol{xx}}^{(i)} & \boldsymbol{r}_{\boldsymbol{xx}}^{(i)\mathrm{B}} \\ \boldsymbol{r}_{\boldsymbol{xx}}^{(i)\mathrm{BT}} & r_{xx}(0) \end{pmatrix} \begin{pmatrix} 1 \\ -\boldsymbol{a}^{(i-1)} \\ 0 \end{pmatrix} + k_i \begin{pmatrix} r_{xx}(0) & \boldsymbol{r}_{\boldsymbol{xx}}^{(i)\mathrm{T}} \\ \boldsymbol{r}_{\boldsymbol{xx}}^{(i)} & \boldsymbol{R}_{\boldsymbol{xx}}^{(i)} \end{pmatrix} \begin{pmatrix} 0 \\ -\boldsymbol{a}^{(i-1)\mathrm{B}} \\ 1 \end{pmatrix} \tag{8.35}$$

where in Equation (8.34) and Equation (8.35) $\boldsymbol{r}_{\boldsymbol{xx}}^{(i)\mathrm{T}} = [r_{xx}(1), \cdots, r_{xx}(i)]$, and $\boldsymbol{r}_{\boldsymbol{xx}}^{(i)B\mathrm{T}} = [r_{xx}(i), \cdots, r_{xx}(1)]$ is the reversed version of $\boldsymbol{r}_{\boldsymbol{xx}}^{(i)\mathrm{T}}$. Matrix–vector multiplication of both sides of Equation (8.35) and the use of Equations (8.29) and (8.30) yields

$$\begin{pmatrix} E^{(i)} \\ \mathbf{0}^{(i)} \end{pmatrix} = \begin{pmatrix} E^{(i-1)} \\ \mathbf{0}^{(i-1)} \\ \Delta^{(i-1)} \end{pmatrix} + k_i \begin{pmatrix} \Delta^{(i-1)} \\ \mathbf{0}^{(i-1)} \\ E^{(i-1)} \end{pmatrix} \tag{8.36}$$

where

$$\begin{aligned} \Delta^{(i-1)} &= \begin{bmatrix} 1 & -\boldsymbol{a}^{(i-1)} \end{bmatrix}^{\mathrm{T}} \boldsymbol{r}_{xx}^{(i)\mathrm{B}} \\ &= r_{xx}(i) - \sum_{k=1}^{i-1} a_k^{(i-1)} r_{xx}(i-k) \end{aligned} \tag{8.37}$$

If Equation (8.36) is true, it follows that Equation (8.32) must also be true. The conditions for Equation (8.36) to be true are

$$E^{(i)} = E^{(i-1)} + k_i \Delta^{(i-1)} \tag{8.38}$$

and

$$0 = \Delta^{(i-1)} + k_i E^{(i-1)} \tag{8.39}$$

From (8.39),

$$k_i = -\frac{\Delta^{(i-1)}}{E^{(i-1)}} \tag{8.40}$$

Substitution of $\Delta^{(i-1)}$ from Equation (8.40) into Equation (8.38) yields

$$\begin{aligned} E^{(i)} &= E^{(i-1)}(1-k_i^2) \\ &= E^{(0)} \prod_{j=1}^{i} (1-k_j^2) \end{aligned} \tag{8.41}$$

Note that it can be shown that $\Delta^{(i)}$ is the cross-correlation of the forward and backward prediction errors:

$$\Delta^{(i-1)} = \mathcal{E}[b^{(i-1)}(m-1)e^{(i-1)}(m)] \tag{8.42}$$

The parameter $\Delta^{(i-1)}$ is known as the partial correlation.

Durbin's algorithm

Equations (8.43)–(8.48) are solved recursively for i=1, . . ., P. The Durbin algorithm starts with a predictor of order zero for which $E^{(0)}=r_{xx}(0)$. The algorithm then computes the coefficients of a predictor of order i, using the coefficients of a predictor of order $i-1$. In the process of solving for the coefficients of a predictor of order P, the solutions for the predictor coefficients of all orders less than P are also obtained:

$$E^{(0)} = r_{xx}(0) \tag{8.43}$$

For i =1, ..., P

$$\Delta^{(i-1)} = r_{xx}(i) - \sum_{k=1}^{i-1} a_k^{(i-1)} r_{xx}(i-k) \tag{8.44}$$

$$k_i = -\frac{\Delta^{(i-1)}}{E^{(i-1)}} \tag{8.45}$$

$$a_i^{(i)} = k_i \tag{8.46}$$

$$a_j^{(i)} = a_j^{(i-1)} - k_i\, a_{i-j}^{(i-1)} \qquad 1 \le j \le i-1 \tag{8.47}$$

$$E^{(i)} = (1-k_i^2)E^{(i-1)} \tag{8.48}$$

8.2.3 Lattice Predictors

The lattice structure, shown in Figure 8.9, is a cascade connection of similar units, with each unit specified by a single parameter k_i, known as the *reflection* coefficient. A major attraction of a lattice structure is its modular form and the relative ease with which the model order can be extended. A further advantage is that, for a stable model, the magnitude of k_i is bounded by unity ($|k_i|<1$), and therefore it is relatively easy to check a lattice structure for stability. The lattice structure is derived from the forward and backward prediction errors as follows. An order-update recursive equation can be obtained for the forward prediction error by multiplying both sides of Equation (8.32) by the input vector [$x(m)$, $x(m-1)$, . . . , $x(m-i)$]:

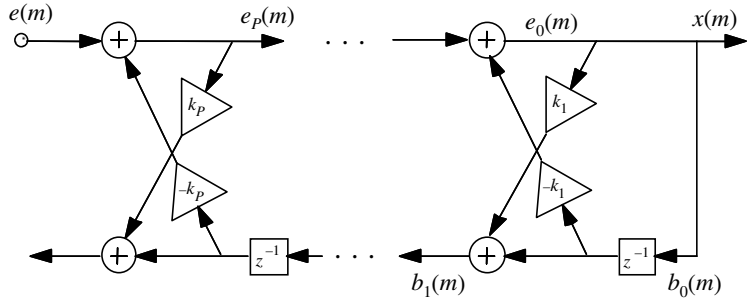
$$e^{(i)}(m) = e^{(i-1)}(m) - k_i\, b^{(i-1)}(m-1) \tag{8.49}$$

Similarly, we can obtain an order-update recursive equation for the backward prediction error by multiplying both sides of Equation (8.33) by the input vector $[x(m–i), x(m–i+1), \ldots, x(m)]$ as

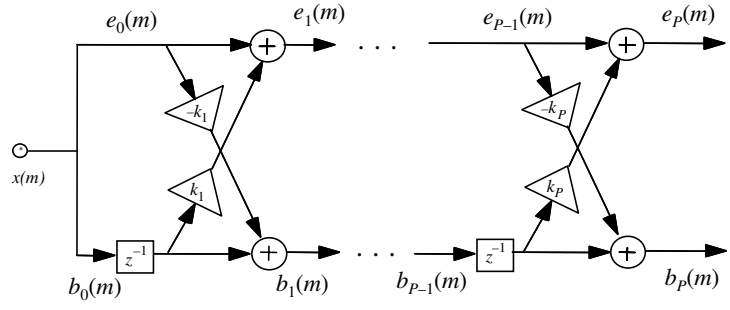
$$b^{(i)}(m)=b^{(i-1)}(m-1)-k_i e^{(i-1)}(m) \tag{8.50}$$

Equations (8.49) and (8.50) are interrelated and may be implemented by a lattice network as shown in Figure 8.8. Minimisation of the squared forward prediction error of Equation (8.49) over N samples yields

$$k_i=\frac{\sum_{m=0}^{N-1} e^{(i-1)}(m)b^{(i-1)}(m-1)}{\sum_{m=0}^{N-1}\left(e^{(i-1)}(m)\right)^2} \tag{8.51}$$



(a)



(b)

Figure 8.9 Configuration of (a) a lattice predictor and (b) the inverse lattice predictor.

Note that a similar relation for k_i can be obtained through minimisation of the squared backward prediction error of Equation (8.50) over N samples. The reflection coefficients are also known as the normalised partial correlation (PARCOR) coefficients.

8.2.4 Alternative Formulations of Least Square Error Prediction

The methods described above for derivation of the predictor coefficients are based on minimisation of either the forward or the backward prediction error. In this section, we consider alternative methods based on the minimisation of the sum of the forward and backward prediction errors.

Burg's Method Burg's method is based on minimisation of the sum of the forward and backward squared prediction errors. The squared error function is defined as

$$E_{fb}^{(i)} = \sum_{m=0}^{N-1} \left\{ \left[e^{(i)}(m) \right]^2 + \left[b^{(i)}(m) \right]^2 \right\} \tag{8.52}$$

Substitution of Equations (8.49) and (8.50) in Equation (8.52) yields

$$E_{fb}^{(i)} = \sum_{m=0}^{N-1} \left\{ \left[e^{(i-1)}(m) - k_i b^{(i-1)}(m-1) \right]^2 + \left[b^{(i-1)}(m-1) - k_i e^{(i-1)}(m) \right]^2 \right\} \tag{8.53}$$

Minimisation of $E_{fb}^{(i)}$ with respect to the reflection coefficients k_i yields

$$k_i = \frac{2\sum_{m=0}^{N-1} e^{(i-1)}(m) b^{(i-1)}(m-1)}{\sum_{m=0}^{N-1} \left\{ \left[e^{(i-1)}(m) \right]^2 + \left[b^{(i-1)}(m-1) \right]^2 \right\}} \tag{8.54}$$

Simultaneous Minimisation of the Backward and Forward Prediction Errors From Equation (8.28) we have that the backward predictor coefficient vector is the reversed version of the forward predictor coefficient vector. Hence a predictor of order P can be obtained through simultaneous minimisation of the sum of the squared backward and forward prediction errors defined by the following equation:

$$\begin{aligned} E_{fb}^{(P)} &= \sum_{m=0}^{N-1}\left\{\left[e^{(P)}(m)\right]^2+\left[b^{(P)}(m)\right]^2\right\} \\ &= \sum_{m=0}^{N-1}\left\{\left[x(m)-\sum_{k=1}^{P}a_k x(m-k)\right]^2+\left[x(m-P)-\sum_{k=1}^{P}a_k x(m-P+k)\right]^2\right\} \\ &= (\boldsymbol{x}-\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{X}\boldsymbol{a})+\left(\boldsymbol{x}^{\mathrm{B}}-\boldsymbol{X}^{\mathrm{B}}\boldsymbol{a}\right)^{\mathrm{T}}\left(\boldsymbol{x}^{\mathrm{B}}-\boldsymbol{X}^{\mathrm{B}}\boldsymbol{a}\right) \end{aligned} \tag{8.55}$$

where $\boldsymbol{X}$ and $\boldsymbol{x}$ are the signal matrix and vector defined by Equations (8.12) and (8.13), and similarly $\boldsymbol{X}^{\mathrm{B}}$ and $\boldsymbol{x}^{\mathrm{B}}$ are the signal matrix and vector for the backward predictor. Using an approach similar to that used in derivation of Equation (8.16), the minimisation of the mean squared error function of Equation (8.54) yields

$$\boldsymbol{a}=\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}+\boldsymbol{X}^{\mathrm{BT}}\boldsymbol{X}^{\mathrm{B}}\right)^{-1}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{X}^{\mathrm{BT}}\boldsymbol{x}^{\mathrm{B}}\right) \tag{8.56}$$

Note that for an ergodic signal as the signal length N increases Equation (8.56) converges to the so-called normal Equation (8.10).

8.2.5 Predictor Model Order Selection

One procedure for the determination of the correct model order is to increment the model order, and monitor the differential change in the error power, until the change levels off. The incremental change in error power with the increasing model order from i–1 to i is defined as

$$\Delta E^{(i)}=E^{(i-1)}-E^{(i)} \tag{8.57}$$

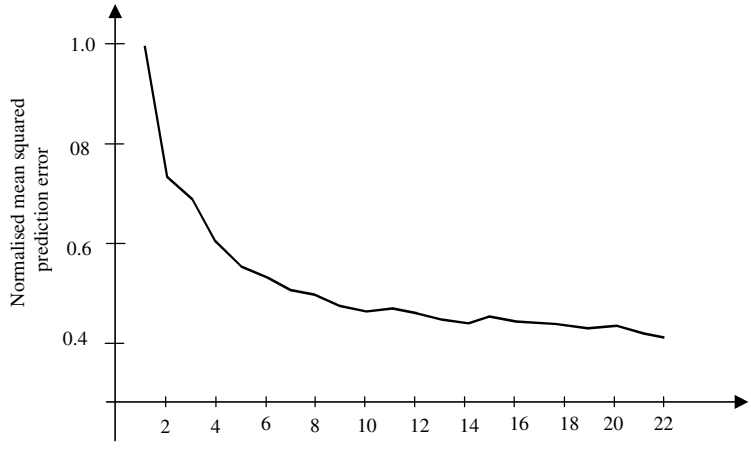


Figure 8.10 Illustration of the decrease in the normalised mean squared prediction error with the increasing predictor length for a speech signal.

Figure 8.10 illustrates the decrease in the normalised mean square prediction error with the increasing predictor length for a speech signal. The order P beyond which the decrease in the error power $\Delta E^{(P)}$ becomes less than a threshold is taken as the model order.

In linear prediction two coefficients are required for modelling each spectral peak of the signal spectrum. For example, the modelling of a signal with K dominant resonances in the spectrum needs $P=2K$ coefficients. Hence a procedure for model selection is to examine the power spectrum of the signal process, and to set the model order to twice the number of significant spectral peaks in the spectrum.

When the model order is less than the correct order, the signal is under-modelled. In this case the prediction error is not well decorrelated and will be more than the optimal minimum. A further consequence of under-modelling is a decrease in the spectral resolution of the model: adjacent spectral peaks of the signal could be merged and appear as a single spectral peak when the model order is too small. When the model order is larger than the correct order, the signal is over-modelled. An over-modelled problem can result in an ill-conditioned matrix equation, unreliable numerical solutions and the appearance of spurious spectral peaks in the model.

8.3 Short-Term and Long-Term Predictors

For quasi-periodic signals, such as voiced speech, there are two types of correlation structures that can be utilised for a more accurate prediction, these are:

(a) the short-term correlation, which is the correlation of each sample with the P immediate past samples: $x(m-1), \ldots, x(m-P)$;
(b) the long-term correlation, which is the correlation of a sample $x(m)$ with say $2Q+1$ similar samples a pitch period T away: $x(m-T+Q), \ldots, x(m-T-Q)$.

Figure 8.11 is an illustration of the short-term relation of a sample with the P immediate past samples and its long-term relation with the samples a pitch period away. The short-term correlation of a signal may be modelled by the linear prediction Equation (8.3). The remaining correlation, in the prediction error signal $e(m)$, is called the long-term correlation. The long-term correlation in the prediction error signal may be modelled by a pitch predictor defined as

$$\hat{e}(m) = \sum_{k=-Q}^{Q} p_k \, e(m-T-k) \tag{8.58}$$

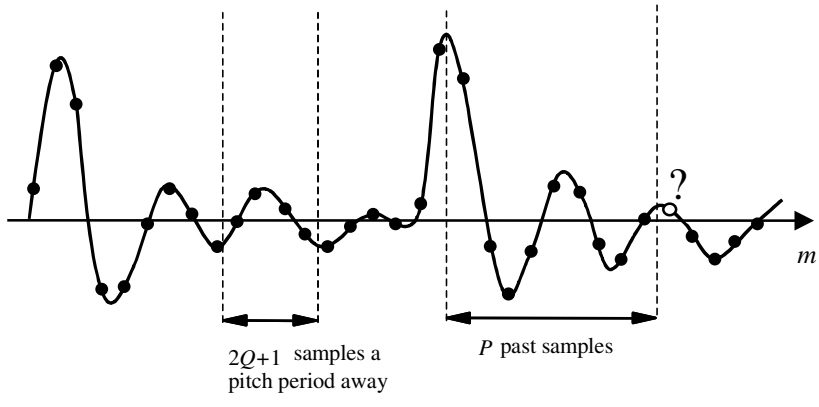


Figure 8.11 Illustration of the short-term relation of a sample with the P immediate past samples and the long-term relation with the samples a pitch period away.

where p_k are the coefficients of a long-term predictor of order $2Q+1$. The pitch period T can be obtained from the autocorrelation function of $x(m)$ or that of $e(m)$: it is the first non-zero time lag where the autocorrelation function attains a maximum. Assuming that the long-term correlation is correctly modelled, the prediction error of the long-term filter is a completely random signal with a white spectrum, and is given by

$$\begin{aligned}\varepsilon(m) &= e(m) - \hat{e}(m) \\ &= e(m) - \sum_{k=-Q}^{Q} p_k\, e(m-T-k)\end{aligned} \tag{8.59}$$

Minimisation of $\mathcal{E}[e^2(m)]$ results in the following solution for the pitch predictor:

$$\begin{pmatrix} p_{-Q} \\ p_{-Q+1} \\ \vdots \\ p_{Q-1} \\ p_Q \end{pmatrix} = \begin{pmatrix} r_{xx}(0) & r_{xx}(1) & r_{xx}(2) & \cdots & r_{xx}(2Q) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(2Q-1) \\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(2Q-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{xx}(2Q) & r_{xx}(2Q-1) & r_{xx}(2Q-2) & \cdots & r_{xx}(0) \end{pmatrix}^{-1} \begin{pmatrix} r_{xx}(T-Q) \\ r_{xx}(T-Q+1) \\ \vdots \\ r_{xx}(T+Q-1) \\ r_{xx}(T+Q) \end{pmatrix} \tag{8.60}$$

An alternative to the separate, cascade, modelling of the short- and long-term correlations is to combine the short- and long-term predictors into a single model described as

$$x(m) = \underbrace{\sum_{k=1}^{P} a_k\, x(m-k)}_{\text{short term prediction}} + \underbrace{\sum_{k=-Q}^{Q} p_k\, x(m-k-T)}_{\text{long term prediction}} + \varepsilon(m) \tag{8.61}$$

In Equation (8.61), each sample is expressed as a linear combination of P immediate past samples and $2Q+1$ samples a pitch period away. Minimisation of $\mathcal{E}[e^2(m)]$ results in the following solution for the pitch predictor:

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \\ p_{-Q} \\ p_{-Q+1} \\ \vdots \\ p_{+Q} \end{pmatrix} = \begin{pmatrix} r(0) & r(1) & \dots & r(P-1) & r(T+Q-1) & r(T+Q) & \dots & r(T-Q-1) \\ r(1) & r(0) & \dots & r(P-2) & r(T+Q-2) & r(T+Q-1) & \dots & r(T+Q-2) \\ r(2) & r(1) & \dots & r(P-3) & r(T+Q-3) & r(T+Q-2) & \dots & r(T+Q-3) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r(P-1) & r(P-2) & \dots & r(0) & r(T+Q-P) & r(T+Q-P+1) & \dots & r(T+Q-P) \\ r(T+Q-1) & r(T+Q-2) & \dots & r(T+Q-P) & r(0) & r(1) & \dots & r(2Q) \\ r(T+Q) & r(T+Q-1) & \dots & r(T+Q-P+1) & r(1) & r(0) & \dots & r(2Q-1) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r(T-Q-1) & r(T-Q-2) & \cdots & r(T-Q-P) & r(2Q) & r(2Q-1) & \dots & r(0) \end{pmatrix}^{-1} \begin{pmatrix} r(1) \\ r(2) \\ r(3) \\ \vdots \\ r(P) \\ r(T+Q) \\ r(T+Q-1) \\ \vdots \\ r(T-Q) \end{pmatrix} \tag{8.62}$$

In Equation (8.62), for simplicity the subscript xx of $r_{xx}(k)$ has been omitted. In Chapter 10, the predictor model of Equation (8.61) is used for interpolation of a sequence of missing samples.

8.4 MAP Estimation of Predictor Coefficients

The posterior probability density function of a predictor coefficient vector $\boldsymbol{a}$, given a signal $\boldsymbol{x}$ and the initial samples $\boldsymbol{x}_\mathrm{I}$, can be expressed, using Bayes' rule, as

$$f_{\boldsymbol{A}|\boldsymbol{X},\boldsymbol{X}_\mathrm{I}}(\boldsymbol{a} \mid \boldsymbol{x},\boldsymbol{x}_\mathrm{I}) = \frac{f_{\boldsymbol{X}|\boldsymbol{A},\boldsymbol{X}_I}(\boldsymbol{x} \mid \boldsymbol{a},\boldsymbol{x}_\mathrm{I}) f_{\boldsymbol{A}|\boldsymbol{X}_\mathrm{I}}(\boldsymbol{a} \mid \boldsymbol{x}_\mathrm{I})}{f_{\boldsymbol{X}|\boldsymbol{X}_I}(\boldsymbol{x} \mid \boldsymbol{x}_\mathrm{I})} \tag{8.63}$$

In Equation (8.63), the pdfs are conditioned on P initial signal samples $\boldsymbol{x}_\mathrm{I}=[x(-P),\ x(-P+1),\ \ldots,\ x(-1)]$. Note that for a given set of samples $[\boldsymbol{x},\ \boldsymbol{x}_\mathrm{I}]$, $f_{\boldsymbol{X}|\boldsymbol{X}_\mathrm{I}}(\boldsymbol{x} \mid \boldsymbol{x}_\mathrm{I})$ is a constant, and it is reasonable to assume that $f_{\boldsymbol{A}|\boldsymbol{X}_I}(\boldsymbol{a} \mid \boldsymbol{x}_\mathrm{I}) = f_{\boldsymbol{A}}(\boldsymbol{a})$.

8.4.1 Probability Density Function of Predictor Output

The pdf $f_{\boldsymbol{X}|\boldsymbol{A},\boldsymbol{X}_\mathrm{I}}(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{x}_\mathrm{I})$ of the signal $\boldsymbol{x}$, given the predictor coefficient vector $\boldsymbol{a}$ and the initial samples $\boldsymbol{x}_\mathrm{I}$, is equal to the pdf of the input signal $\boldsymbol{e}$:

$$f_{\boldsymbol{X}|\boldsymbol{A},\boldsymbol{X}_\mathrm{I}}(\boldsymbol{x} \mid \boldsymbol{a},\boldsymbol{x}_\mathrm{I}) = f_{\boldsymbol{E}}(\boldsymbol{x} - \boldsymbol{X}\boldsymbol{a}) \tag{8.64}$$

where the input signal vector is given by

$$\boldsymbol{e} = -\boldsymbol{X}\boldsymbol{a} \tag{8.65}$$

and $f_E(\boldsymbol{e})$ is the pdf of $\boldsymbol{e}$. Equation (8.64) can be expanded as

$$\begin{pmatrix} e(0) \\ e(1) \\ e(2) \\ \vdots \\ e(N-1) \end{pmatrix} = \begin{pmatrix} x(0) \\ x(1) \\ x(2) \\ \vdots \\ x(N-1) \end{pmatrix} - \begin{pmatrix} x(-1) & x(-2) & x(-3) & \cdots & x(-P) \\ x(0) & x(-1) & x(-2) & \cdots & x(1-P) \\ x(1) & x(0) & x(-1) & \cdots & x(2-P) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x(N-2) & x(N-3) & x(N-4) & \cdots & x(N-P-1) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \end{pmatrix} \tag{8.66}$$

Assuming that the input excitation signal $e(m)$ is a zero-mean, uncorrelated, Gaussian process with a variance of σ_e^2, the likelihood function in Equation (8.64) becomes

$$\begin{aligned} f_{\boldsymbol{X}|\boldsymbol{A},\boldsymbol{X}_I}(\boldsymbol{x} \mid \boldsymbol{a}, \boldsymbol{x}_I) &= f_{\boldsymbol{E}}(\boldsymbol{x} - \boldsymbol{X}\boldsymbol{a}) \\ &= \frac{1}{\left(2\pi\sigma_e^2\right)^{N/2}} \exp\left(\frac{1}{2\sigma_e^2} (\boldsymbol{x} - \boldsymbol{X}\boldsymbol{a})^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{X}\boldsymbol{a}) \right) \end{aligned} \tag{8.67}$$

An alternative form of Equation (8.67) can be obtained by rewriting Equation (8.66) in the following form:

$$\begin{pmatrix} e_0 \\ e_1 \\ e_3 \\ e_4 \\ \vdots \\ e_{N-1} \end{pmatrix} = \begin{pmatrix} -a_P & \cdots & -a_2 & -a_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -a_P & \cdots & -a_2 & -a_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -a_P & \cdots & -a_2 & -a_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -a_P & \cdots & -a_2 & -a_1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & -a_P & \cdots & -a_2 & -a_1 & 1 \end{pmatrix} \begin{pmatrix} x_{-P} \\ x_{-P+1} \\ x_{-P+2} \\ x_{-P+3} \\ \vdots \\ x_{N-1} \end{pmatrix} \tag{8.68}$$

In a compact notation Equation (8.68) can be written as

$$\boldsymbol{e} = \boldsymbol{A}\boldsymbol{x} \tag{8.69}$$

Using Equation (8.69), and assuming that the excitation signal $e(m)$ is a zero mean, uncorrelated process with variance σ_e^2, the likelihood function of Equation (8.67) can be written as

$$f_{\boldsymbol{X}|\boldsymbol{A},\boldsymbol{X}_I}(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{x}_{\mathrm{I}})=\frac{1}{\left(2\pi\sigma_e^2\right)^{N/2}}\exp\left(-\frac{1}{2\sigma_e^2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\right) \tag{8.70}$$

8.4.2 Using the Prior pdf of the Predictor Coefficients

The prior pdf of the predictor coefficient vector is assumed to have a Gaussian distribution with a mean vector $\boldsymbol{\mu}_{\boldsymbol{a}}$ and a covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{aa}}$:

$$f_{\boldsymbol{A}}(\boldsymbol{a})=\frac{1}{(2\pi)^{P/2}|\boldsymbol{\Sigma}_{\boldsymbol{aa}}|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{a}-\boldsymbol{\mu}_{\boldsymbol{a}})^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{aa}}^{-1}(\boldsymbol{a}-\boldsymbol{\mu}_{\boldsymbol{a}})\right] \tag{8.71}$$

Substituting Equations (8.67) and (8.71) in Equation (8.63), the posterior pdf of the predictor coefficient vector $f_{\boldsymbol{A}|\boldsymbol{X},\boldsymbol{X}_{\mathrm{I}}}(\boldsymbol{a}\mid\boldsymbol{x},\boldsymbol{x}_{\mathrm{I}})$ can be expressed as

$$\begin{aligned}f_{\boldsymbol{A}|\boldsymbol{X},\boldsymbol{X}_I}(\boldsymbol{a}\mid\boldsymbol{x},\boldsymbol{x}_{\mathrm{I}})=&\frac{1}{f_{\boldsymbol{X}|\boldsymbol{X}_{\mathrm{I}}}(\boldsymbol{x}\,|\boldsymbol{x}_{\mathrm{I}})}\frac{1}{(2\pi)^{(N+P)/2}\sigma_e^N|\boldsymbol{\Sigma}_{\boldsymbol{aa}}|^{1/2}}\\&\times\exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_e^2}(\boldsymbol{x}-\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{X}\boldsymbol{a})+(\boldsymbol{a}-\boldsymbol{\mu}_{\boldsymbol{a}})^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{aa}}^{-1}(\boldsymbol{a}-\boldsymbol{\mu}_{\boldsymbol{a}})\right]\right\}\end{aligned} \tag{8.72}$$

The maximum a posteriori estimate is obtained by maximising the log-likelihood function:

$$\frac{\partial}{\partial\boldsymbol{a}}\left[\ln f_{\boldsymbol{A}|\boldsymbol{X},\boldsymbol{X}_I}(\boldsymbol{a}\mid\boldsymbol{x},\boldsymbol{x}_I)\right]=\frac{\partial}{\partial\boldsymbol{a}}\left[\frac{1}{\sigma_e^2}(\boldsymbol{x}-\boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{X}\boldsymbol{a})+(\boldsymbol{a}-\boldsymbol{\mu}_{\boldsymbol{a}})^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{aa}}^{-1}(\boldsymbol{a}-\boldsymbol{\mu}_{\boldsymbol{a}})\right]=0 \tag{8.73}$$

This yields

$$\hat{\boldsymbol{a}}^{MAP}=\left(\boldsymbol{\Sigma}_{\boldsymbol{aa}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}+\sigma_e^2\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{aa}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{x}+\sigma_e^2\left(\boldsymbol{\Sigma}_{\boldsymbol{aa}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}+\sigma_e^2\mathbf{I}\right)^{-1}\boldsymbol{\mu}_{\boldsymbol{a}} \tag{8.74}$$

Note that as the Gaussian prior tends to a uniform prior, the determinant covariance matrix $\boldsymbol{\Sigma}_{aa}$ of the Gaussian prior increases, and the MAP solution tends to the least square error solution:

$$\hat{\boldsymbol{a}}^{LS} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{x}\right) \tag{8.75}$$

Similarly as the observation length N increases the signal matrix $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ becomes more significant than $\boldsymbol{\Sigma}_{aa}$ and again the MAP solution tends to a least squared error solution.

8.5 Sub-Band Linear Prediction Model

In a P^{th} order linear prediction model, the P predictor coefficients model the signal spectrum over its full spectral bandwidth. The distribution of the LP parameters (or equivalently the poles of the LP model) over the signal bandwidth depends on the signal correlation and spectral structure. Generally, the parameters redistribute themselves over the spectrum to minimize the mean square prediction error criterion. An alternative to a conventional LP model is to divide the input signal into a number of sub-bands and to model the signal within each sub-band with a linear prediction model as shown in Figure 8.12. The advantages of using a sub-band LP model are as follows:

(1) Sub-band linear prediction allows the designer to allocate a specific number of model parameters to a given sub-band. Different numbers of parameters can be allocated to different bands.
(2) The solution of a full-band linear predictor equation, i.e. Equation (8.10) or (8.16), requires the inversion of a relatively large correlation matrix, whereas the solution of the sub-band LP models require the inversion of a number of relatively small correlation matrices with better numerical stability properties. For example, a predictor of order 18 requires the inversion of an 18×18 matrix, whereas three sub-band predictors of order 6 require the inversion of three 6×6 matrices.
(3) Sub-band linear prediction is useful for applications such as noise reduction where a sub-band approach can offer more flexibility and better performance.

In sub-band linear prediction, the signal $x(m)$ is passed through a bank of N band-pass filters, and is split into N sub-band signals $x_k(m)$, k=1, …,N. The k^{th} sub-band signal is modelled using a low-order linear prediction model as

$$x_k(m)=\sum_{i=1}^{P_k} a_k(i)x_k(m-i)+g_k e_k(m) \tag{8.76}$$

where $[a_k, g_k]$ are the coefficients and the gain of the predictor model for the k^{th} sub-band. The choice of the model order P_k depends on the width of the sub-band and on the signal correlation structure within each sub-band. The power spectrum of the input excitation of an ideal LP model for the k^{th} sub-band signal can be expressed as

$$P_{EE}(f,k)=\begin{cases}1 & f_{k,start}<f<f_{k,end}\\ 0 & \text{otherwise}\end{cases} \tag{8.77}$$

where $f_{k,start}$, $f_{k,end}$ are the start and end frequencies of the k^{th} sub-band signal. The autocorrelation function of the excitation function in each sub-band is a sinc function given by

$$r_{ee}(m)=B_k \operatorname{sinc}[m(B_k-f_{k0})/2] \tag{8.78}$$

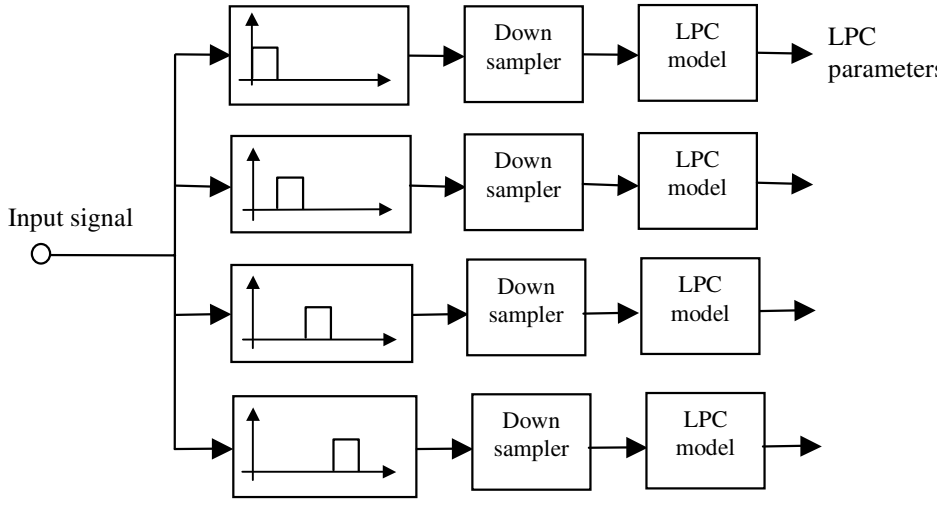


Figure 8.12 Configuration of a sub-band linear prediction model.

where B_k and f_{k0} are the bandwidth and the centre frequency of the k^{th} sub-band respectively. To ensure that each sub-band LP parameters only model the signal within that sub-band, the sub-band signals are down-sampled as shown in Figure 8.12.

8.6 Signal Restoration Using Linear Prediction Models

Linear prediction models are extensively used in speech and audio signal restoration. For a noisy signal, linear prediction analysis models the combined spectra of the signal and the noise processes. For example, the frequency spectrum of a linear prediction model of speech, observed in additive white noise, would be flatter than the spectrum of the noise-free speech, owing to the influence of the flat spectrum of white noise. In this section we consider the estimation of the coefficients of a predictor model from noisy observations, and the use of linear prediction models in signal restoration. The noisy signal $y(m)$ is modelled as

$$\begin{aligned} y(m) &= x(m) + n(m) \\ &= \sum_{k=1}^{P} a_k x(m-k) + e(m) + n(m) \end{aligned} \tag{8.79}$$

where the signal $x(m)$ is modelled by a linear prediction model with coefficients a_k and random input $e(m)$, and it is assumed that the noise $n(m)$ is additive. The least square error predictor model of the noisy signal $y(m)$ is given by

$$\boldsymbol{R}_{yy}\hat{\boldsymbol{a}} = \boldsymbol{r}_{yy} \tag{8.80}$$

where $\boldsymbol{R}_{yy}$ and $\boldsymbol{r}_{yy}$ are the autocorrelation matrix and vector of the noisy signal $y(m)$. For an additive noise model, Equation (8.80) can be written as

$$(\boldsymbol{R}_{xx} + \boldsymbol{R}_{nn})(\boldsymbol{a} + \tilde{\boldsymbol{a}}) = (\boldsymbol{r}_{xx} + \boldsymbol{r}_{nn}) \tag{8.81}$$

where $\tilde{\boldsymbol{a}}$ is the error in the predictor coefficients vector due to the noise. A simple method for removing the effects of noise is to subtract an estimate of the autocorrelation of the noise from that of the noisy signal. The drawback

of this approach is that, owing to random variations of noise, correlation subtraction can cause numerical instability in Equation (8.80) and result in spurious solutions. In the following, we formulate the p.d.f. of the noisy signal and describe an iterative signal-restoration/parameter-estimation procedure developed by Lee and Oppenheim.

From Bayes' rule, the MAP estimate of the predictor coefficient vector $\boldsymbol{a}$, given an observation signal vector $\boldsymbol{y}$=[$y(0)$, $y(1)$, ..., $y(N-1)$], and the initial samples vector $\boldsymbol{x}_I$ is

$$f_{\boldsymbol{A}|\boldsymbol{Y},\boldsymbol{X}_I}(\boldsymbol{a} \mid \boldsymbol{y},\boldsymbol{x}_I)=\frac{f_{\boldsymbol{Y}|\boldsymbol{A},\boldsymbol{X}_I}(\boldsymbol{y} \mid \boldsymbol{a},\boldsymbol{x}_I) f_{\boldsymbol{A},\boldsymbol{X}_I}(\boldsymbol{a},\boldsymbol{x}_I)}{f_{\boldsymbol{Y},\boldsymbol{X}_I}(\boldsymbol{y},\boldsymbol{x}_I)} \tag{8.82}$$

Now consider the variance of the signal $\boldsymbol{y}$ in the argument of the term $f_{\boldsymbol{Y}|\boldsymbol{A},\boldsymbol{X}_I}(\boldsymbol{y} \mid \boldsymbol{a},\boldsymbol{x}_I)$ in Equation (8.82). The innovation of $y(m)$ can be defined as

$$\begin{aligned}\varepsilon(m) &= y(m)-\sum_{k=1}^{P} a_k\, y(m-k) \\ &= e(m)+n(m)-\sum_{k=1}^{P} a_k\, n(m-k)\end{aligned} \tag{8.83}$$

The variance of $y(m)$, given the previous P samples and the coefficient vector $\boldsymbol{a}$, is the variance of the innovation signal $\varepsilon(m)$, given by

$$\mathrm{Var}[\,y(m)\,|\,y(m-1),\ldots,y(m-P),\boldsymbol{a}\,]=\sigma_\varepsilon^2+\sigma_e^2+\sigma_n^2-\sigma_n^2\sum_{k=1}^{P} a_k^2 \tag{8.84}$$

where σ_e^2 and σ_n^2 are the variance of the excitation signal and the noise respectively. From Equation (8.84), the variance of $y(m)$ is a function of the coefficient vector $\boldsymbol{a}$. Consequently, maximisation of $f_{\boldsymbol{Y}|\boldsymbol{A},\boldsymbol{X}_I}(\boldsymbol{y}|\boldsymbol{a},\boldsymbol{x}_I)$ with respect to the vector $\boldsymbol{a}$ is a non-linear and non-trivial exercise.

Lim and Oppenheim proposed the following iterative process in which an estimate $\hat{\boldsymbol{a}}$ of the predictor coefficient vector is used to make an estimate $\hat{\boldsymbol{x}}$ of the signal vector, and the signal estimate $\hat{\boldsymbol{x}}$ is then used to improve the estimate of the parameter vector $\hat{\boldsymbol{a}}$, and the process is iterated until

convergence. The posterior pdf of the noise-free signal $\boldsymbol{x}$ given the noisy signal $\boldsymbol{y}$ and an estimate of the parameter vector $\hat{\boldsymbol{a}}$ is given by

$$f_{\boldsymbol{X}|\boldsymbol{A},\boldsymbol{Y}}(\boldsymbol{x}|\hat{\boldsymbol{a}},\boldsymbol{y}) = \frac{f_{\boldsymbol{Y}|\boldsymbol{A},\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{a}},\boldsymbol{x})\, f_{\boldsymbol{X}|\boldsymbol{A}}(\boldsymbol{x}|\hat{\boldsymbol{a}})}{f_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}|\hat{\boldsymbol{a}})} \tag{8.85}$$

Consider the likelihood term $f_{\boldsymbol{Y}|\boldsymbol{A},\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{a}},\boldsymbol{x})$. Since the noise is additive, we have

$$\begin{aligned} f_{\boldsymbol{Y}|\boldsymbol{A},\boldsymbol{X}}(\boldsymbol{y}\,|\,\hat{\boldsymbol{a}},\boldsymbol{x}) &= f_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{x}) \\ &= \frac{1}{\left(2\pi\sigma_n^2\right)^{N/2}} \exp\left[-\frac{1}{2\sigma_n^2}(\boldsymbol{y}-\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x})\right] \end{aligned} \tag{8.86}$$

Assuming that the input of the predictor model is a zero-mean Gaussian process with variance σ_e^2, the pdf of the signal $\boldsymbol{x}$ given an estimate of the predictor coefficient vector $\boldsymbol{a}$ is

$$\begin{aligned} f_{\boldsymbol{Y}|\boldsymbol{A},\boldsymbol{X}}(\boldsymbol{x}\,|\,\hat{\boldsymbol{a}}) &= \frac{1}{\left(2\pi\sigma_e^2\right)^{N/2}} \exp\left(-\frac{1}{2\sigma_e^2}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{e}\right) \\ &= \frac{1}{\left(2\pi\sigma_e^2\right)^{N/2}} \exp\left(-\frac{1}{2\sigma_e^2}\boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{A}}^{\mathrm{T}}\hat{\boldsymbol{A}}\boldsymbol{x}\right) \end{aligned} \tag{8.87}$$

where $\boldsymbol{e} = \hat{\boldsymbol{A}}\boldsymbol{x}$ as in Equation (8.69). Substitution of Equations (8.86) and (8.87) in Equation (8.85) yields

$$f_{\boldsymbol{X}|\boldsymbol{A},\boldsymbol{Y}}(\boldsymbol{x}\,|\,\hat{\boldsymbol{a}},\boldsymbol{y}) = \frac{1}{f_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}\,|\,\hat{\boldsymbol{a}})} \frac{1}{(2\pi\sigma_n\sigma_e)^N} \exp\left[-\frac{1}{2\sigma_n^2}(\boldsymbol{y}-\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}) - \frac{1}{2\sigma_e^2}\boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{A}}^{\mathrm{T}}\hat{\boldsymbol{A}}\boldsymbol{x}\right] \tag{8.88}$$

In Equation (8.88), for a given signal $\boldsymbol{y}$ and coefficient vector $\hat{\boldsymbol{a}}$, $f_{\boldsymbol{Y}|\boldsymbol{A}}(\boldsymbol{y}|\hat{\boldsymbol{a}})$ is a constant. From Equation (8.88), the ML signal estimate is obtained by maximising the log-likelihood function as

$$\frac{\partial}{\partial \boldsymbol{a}}\left(\ln f_{\boldsymbol{X}|\boldsymbol{A},\boldsymbol{Y}}\left(\boldsymbol{x}\,|\,\hat{\boldsymbol{a}},\boldsymbol{y}\right)\right)=\frac{\partial}{\partial \boldsymbol{x}}\left(-\frac{1}{2\sigma_e^2}\boldsymbol{x}^{\mathrm{T}}\hat{\boldsymbol{A}}^{\mathrm{T}}\hat{\boldsymbol{A}}\boldsymbol{x}-\frac{1}{2\sigma_n^2}(\boldsymbol{y}-\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x})\right)=\boldsymbol{0} \tag{8.89}$$

which gives

$$\hat{\boldsymbol{x}}=\sigma_e^2\left(\sigma_n^2\hat{\boldsymbol{A}}^{\mathrm{T}}\hat{\boldsymbol{A}}+\sigma_e^2\mathbf{I}\right)^{-1}\boldsymbol{y} \tag{8.90}$$

The signal estimate of Equation (8.90) can be used to obtain an updated estimate of the predictor parameter. Assuming that the signal is a zero mean Gaussian process, the estimate of the predictor parameter vector $\boldsymbol{a}$ is given by

$$\hat{\boldsymbol{a}}(\hat{\boldsymbol{x}})=\left(\hat{\boldsymbol{X}}^{\mathrm{T}}\hat{\boldsymbol{X}}\right)^{-1}\left(\hat{\boldsymbol{X}}^{\mathrm{T}}\hat{\boldsymbol{x}}\right) \tag{8.91}$$

Equations (8.90) and (8.91) form the basis for an iterative signal restoration/parameter estimation method.

8.6.1 Frequency-Domain Signal Restoration Using Prediction Models

The following algorithm is a frequency-domain implementation of the linear prediction model-based restoration of a signal observed in additive white noise.

Initialisation: Set the initial signal estimate to noisy signal $\hat{\boldsymbol{x}}_0=\boldsymbol{y}$,
For iterations $i = 0, 1, ...$

Step 1 Estimate the predictor parameter vector $\hat{\boldsymbol{a}}_i$:

$$\hat{\boldsymbol{a}}_i(\hat{\boldsymbol{x}}_i)=\left(\hat{\boldsymbol{X}}_i^{\mathrm{T}}\hat{\boldsymbol{X}}_i\right)^{-1}\left(\hat{\boldsymbol{X}}_i^{\mathrm{T}}\hat{\boldsymbol{x}}_i\right) \tag{8.92}$$

Step 2 Calculate an estimate of the model gain G using the Parseval's theorem:

$$\frac{1}{N}\sum_{f=0}^{N-1}\frac{\hat{G}^2}{\left|1-\sum_{k=1}^{P}\hat{a}_{k,i}e^{-j2\pi fk/N}\right|^2}=\sum_{m=0}^{N-1}y^2(m)-N\hat{\sigma}_n^2 \tag{8.93}$$

where $\hat{a}_{k,i}$ are the coefficient estimates at iteration i, and $N\hat{\sigma}_n^2$ is the energy of white noise over N samples.

Step 3 Calculate an estimate of the power spectrum of speech model:

$$\hat{P}_{X_iX_i}(f)=\frac{\hat{G}^2}{\left|1-\sum_{k=1}^{P}\hat{a}_{k,i}e^{-j2\pi fk/N}\right|^2} \tag{8.94}$$

Step 4 Calculate the Wiener filter frequency response:

$$\hat{W}_i(f)=\frac{\hat{P}_{X_iX_i}(f)}{\hat{P}_{X_iX_i}(f)+\hat{P}_{N_iN_i}(f)} \tag{8.95}$$

where $\hat{P}_{N_iN_i}(f)=\hat{\sigma}_n^2$ is an estimate of the noise power spectrum.

Step 5 Filter the magnitude spectrum of the noisy speech as

$$\hat{X}_{i+1}(f)=\hat{W}_i(f)Y(f) \tag{8.96}$$

Restore the time domain signal $\hat{\boldsymbol{x}}_{i+1}$ by combining $\hat{X}_{i+1}(f)$ with the phase of noisy signal and the complex signal to time domain.

Step 6 Goto step 1 and repeat until convergence, or for a specified number of iterations.

Figure 8.13 illustrates a block diagram configuration of a Wiener filter using a linear prediction estimate of the signal spectrum. Figure 8.14 illustrates the result of an iterative restoration of the spectrum of a noisy speech signal.

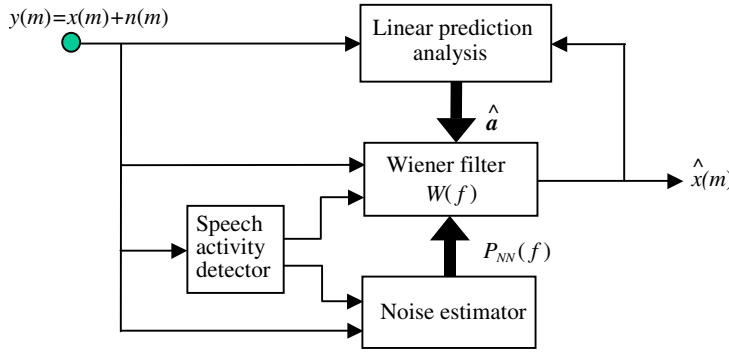


Figure 8.13 Iterative signal restoration based on linear prediction model of speech.

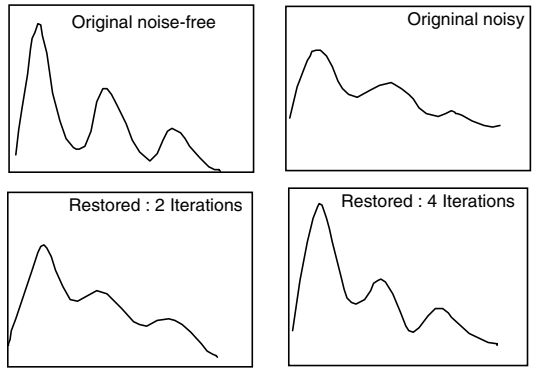


Figure 8.14 Illustration of restoration of a noisy signal with iterative linear prediction based method.

8.6.2 Implementation of Sub-Band Linear Prediction Wiener Filters

Assuming that the noise is additive, the noisy signal in each sub-band is modelled as

$$y_k(m) = x_k(m) + n_k(m) \tag{8.97}$$

The Wiener filter in the frequency domain can be expressed in terms of the power spectra, or in terms of LP model frequency responses, of the signal and noise process as

$$W_k(f) = \frac{P_{X,k}(f)}{P_{Y,k}(f)} = \frac{g_{X,k}^2}{\left|A_{X,k}(f)\right|^2} \frac{\left|A_{Y,k}(f)\right|^2}{g_{Y,k}^2} \tag{8.98}$$

where $P_{X,k}(f)$ and $P_{Y,k}(f)$ are the power spectra of the clean signal and the noisy signal for the k^{th} subband respectively. From Equation (8.98) the square-root Wiener filter is given by

$$W_k^{1/2}(f) = \frac{g_{X,k}}{\left|A_{X,k}(f)\right|} \frac{\left|A_{Y,k}(f)\right|}{g_{Y,k}} \tag{8.99}$$

The linear prediction Wiener filter of Equation (8.99) can be implemented in the time domain with a cascade of a linear predictor of the clean signal, followed by an inverse predictor filter of the noisy signal as expressed by the following relations (see Figure 8.15):

$$z_k(m) = \sum_{i=1}^{P} a_{Xk}(i) z_k(m-i) + \frac{g_X}{g_Y} y_k(m) \tag{8.100}$$

$$\hat{x}_k(m) = \sum_{i=0}^{P} a_{Yk}(i) z_k(m-i) \tag{8.101}$$

where $\hat{x}_k(m)$ is the restored estimate of $x_k(m)$ the clean speech signal and $z_k(m)$ is an intermediate signal.

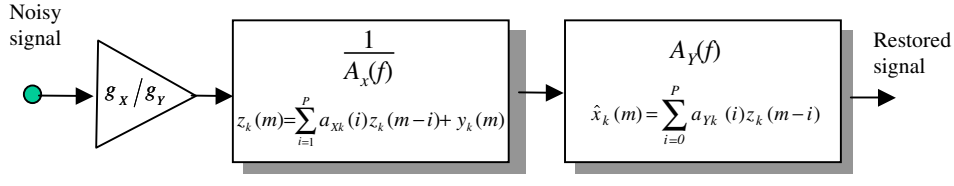


Figure 8.15 A cascade implementation of the LP squared-root Wiener filter.

8.7 Summary

Linear prediction models are used in a wide range of signal processing applications from low-bit-rate speech coding to model-based spectral analysis. We began this chapter with an introduction to linear prediction theory, and considered different methods of formulation of the prediction problem and derivations of the predictor coefficients. The main attraction of the linear prediction method is the closed-form solution of the predictor coefficients, and the availability of a number of efficient and relatively robust methods for solving the prediction equation such as the Levinson–Durbin method. In Section 8.2, we considered the forward, backward and lattice predictors. Although the direct-form implementation of the linear predictor is the most convenient method, for many applications, such as transmission of the predictor coefficients in speech coding, it is advantageous to use the lattice form of the predictor. This is because the lattice form can be conveniently checked for stability, and furthermore a perturbation of the parameter of any section of the lattice structure has a limited and more localised effect. In Section 8.3, we considered a modified form of linear prediction that models the short-term and long-term correlations of the signal. This method can be used for the modelling of signals with a quasi-periodic structure such as voiced speech. In Section 8.4, we considered MAP estimation and the use of a prior pdf for derivation of the predictor coefficients. In Section 8.5, the sub-band linear prediction method was formulated. Finally in Section 8.6, a linear prediction model was applied to the restoration of a signal observed in additive noise.

Bibliography

AKAIKE H. (1970) Statistical Predictor Identification, Annals of the Institute of Statistical Mathematics. **22**, pp. 203–217.

AKAIKE H. (1974) A New Look at Statistical Model Identification, IEEE Trans. on Automatic Control, **AC-19**, pp. 716–723, Dec.

ANDERSON O.D. (1976) Time Series Analysis and Forecasting, The Box-Jenkins Approach. Butterworth, London.

AYRE A.J. (1972) Probability and Evidence Columbia University Press.

BOX G.E.P and JENKINS G.M. (1976) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, California.

BURG J.P. (1975) Maximum Entropy Spectral Analysis. P.h.D. thesis, Stanford University, Stanford, California.

COHEN J. and Cohen P. (1975) Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Halsted, New York.

DRAPER N.R. and Smith H. (1981) Applied Regression Analysis, 2nd Ed. Wiley, New York.

DURBIN J. (1959) Efficient Estimation of Parameters in Moving Average Models. Biometrica, **46**, pp. 306–317.

DURBIN J. (1960) The Fitting of Time Series Models. Rev. Int. Stat. Inst., **28**, pp. 233–244.

FULLER W.A. (1976) Introduction to Statistical Time Series. Wiley, New York.

HANSEN J.H. and CLEMENTS M.A. (1987). Iterative Speech Enhancement with Spectral Constrains. IEEE Proc. Int. Conf. on Acoustics, Speech and Signal Processing ICASSP-87, **1**, pp. 189–192, Dallas, April.

HANSEN J.H. and CLEMENTS M.A. (1988). Constrained Iterative Speech Enhancement with Application to Automatic Speech Recognition. IEEE Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP-88, **1**, pp. 561–564, New York, April.

HOCKING R.R. (1996): The Analysis of Linear Models. Wiley.

KOBATAKE H., INARI J. and KAKUTA S. (1978) Linear prediction Coding of Speech Signals in a High Ambient Noise Environment. IEEE Proc. Int. Conf. on Acoustics, Speech and Signal Processing, pp. 472–475, April.

LIM J.S. and OPPENHEIM A.V. (1978) All-Pole Modelling of Degraded Speech. IEEE Trans. Acoustics, Speech and Signal Processing, **ASSP-26**, **3**, pp. 197-210, June.

LIM J.S. and OPPENHEIM A.V. (1979) Enhancement and Bandwidth Compression of Noisy Speech, Proc. IEEE, **67**, pp. 1586-1604.

MAKOUL J.(1975) Linear Prediction: A Tutorial review. Proceedings of the IEEE, **63**, pp. 561-580.

MARKEL J.D. and GRAY A.H. (1976) Linear Prediction of Speech. Springer Verlag, New York.

RABINER L.R. and SCHAFER R.W. (1976) Digital Processing of Speech Signals. Prentice-Hall, Englewood Cliffs, NJ.

TONG H. (1975) Autoregressive Model Fitting with Noisy Data by Akaike's Information Criterion. IEEE Trans. Information Theory, **IT-23**, pp. 409–48.

STOCKHAM T.G., CANNON T.M. and INGEBRETSEN R.B. (1975) Blind Deconvolution Through Digital Signal Processing. IEEE Proc. **63**, **4**, pp. 678–692.