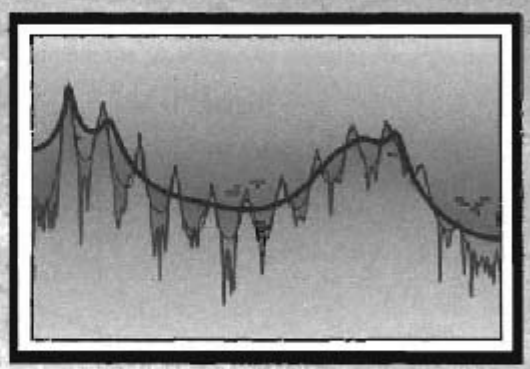
11 Parametric Signal Modeling



11.0 INTRODUCTION

Throughout this text, we have found it convenient to use several different representations of signals and systems. For example, the representation of a discrete-time signal as a sequence of scaled impulses was used in Eq. (2.5) of Chapter 2 to develop the convolution sum for LTI systems. The representation as a linear combination of sinusoidal and complex exponential signals led to the Fourier series, the Fourier transform, and the frequency domain characterization of signals and LTI systems. Although these representations are particularly useful because of their generality, they are not always the most efficient representation for signals with a known structure.

This chapter introduces another powerful approach to signal representation called *parametric signal modeling*. In this approach, a signal is represented by a mathematical model that has a predefined structure involving a limited number of parameters. A given signal $s[n]$ is represented by choosing the specific set of parameters that results in the model output $\hat{s}[n]$ being as close as possible in some prescribed sense to the given signal. A common example is to model the signal as the output of a discrete-time linear system as shown in Figure 11.1. Such models, which are comprised of the input signal $v[n]$ and the system function $H(z)$ of the linear system, become useful with the addition of constraints that make it possible to solve for the parameters of $H(z)$ given the signal

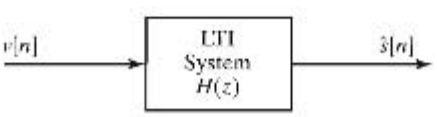


Figure 11.1 Linear system model for a signal $s[n]$.

to be represented. For example, if the input $v[n]$ is specified, and the system function is assumed to be a rational function of the form

$$H(z) = \frac{\sum_{k=0}^{q} b_k z^{-k}}{1 - \sum_{k=1}^{p} a_k z^{-k}}, \tag{11.1}$$

then the signal is modeled by the values of the a_ks and b_ks or equivalently, by the poles and zeros of $H(z)$, along with knowledge of the input. The input signal $v[n]$ is generally assumed to be a unit impulse $\delta[n]$ for deterministic signals, or white noise if the signal $s[n]$ is viewed as a random signal. When the model is appropriately chosen, it is possible to represent a large number of signal samples by a relatively small set of parameters.

Parametric signal modeling has a wide range of applications, including data compression, spectrum analysis, signal prediction, deconvolution, filter design, system identification, signal detection, and signal classification. In data compression, for example, it is the set of model parameters that is transmitted or stored, and the receiver then uses the model with those parameters to regenerate the signal. In filter design, the model parameters are chosen to best approximate, in some sense, the desired frequency response, or equivalently, the desired impulse response, and the model with these parameters then corresponds to the designed filter. The two key elements for success in all of the applications are an appropriate choice of model and an accurate estimate of the parameters for the model.

11.1 ALL-POLE MODELING OF SIGNALS

The model represented by Eq. (11.1) in general has both poles and zeros. While there are a variety of techniques for determining the full set of numerator and denominator coefficients in Eq. (11.1), the most successful and most widely used have concentrated on restricting q to be zero, in which case, $H(z)$ in Figure 11.1 has the form

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \frac{G}{A(z)}, \tag{11.2}$$

where we have replaced the parameter b_0 by the parameter G to emphasize its role as an overall gain factor. Such models are aptly termed "all-pole" models.[1] By its very nature, it would appear that an all-pole model would be appropriate only for modeling signals of infinite duration. While this may be true in a theoretical sense, this choice for the system function of the model works well for signals found in many applications, and as we will show, the parameters can be computed in a straightforward manner from finite-duration segments of the given signal.

[1] Detailed discussion of this case and the general pole/zero case are given in Kay (1988), Thierrien (1992), Hayes (1996) and Stoica and Moses (2005).

The input and output of the all-pole system in Eq. (11.2) satisfy the linear constant-coefficient difference equation

$$\hat{s}[n] = \sum_{k=1}^{p} a_k \hat{s}[n-k] + Gv[n], \tag{11.3}$$

which indicates that the model output at time n is comprised of a linear combination of past samples plus a scaled input sample. As we will see, this structure suggests that the all-pole model is equivalent to the assumption that the signal can be approximated as a linear combination of (or equivalently, is linearly predictable from) its previous values. Consequently, this method for modeling a signal is often also referred to as *linear predictive analysis* or *linear prediction*.[2]

11.1.1 Least-Squares Approximation

The goal in all-pole modeling is to choose the input $v[n]$ and the parameters G, and $a_1, \ldots, a_p$ in Eq. (11.3) such that $\hat{s}[n]$ is a close approximation in some sense to $s[n]$, the signal to be modeled. If, as is usually the case, $v[n]$ is specified in advance (e.g., $v[n] = \delta[n]$), a direct approach to determining the best values for the parameters might be to minimize the total energy in the error signal $e_{se}[n] = (s[n] - \hat{s}[n])$, thereby obtaining a least-squares approximation to $s[n]$. Specifically, for deterministic signals, the model parameters might be chosen to minimize the total squared error

$$\sum_{n=-\infty}^{\infty} (s[n] - \hat{s}[n])^2 = \sum_{n=-\infty}^{\infty} \left(s[n] - \sum_{k=1}^{p} a_k \hat{s}[n-k] - Gv[n] \right)^2. \tag{11.4}$$

In principle, the a_ks minimizing this squared error can be found by differentiating the expression in Eq. (11.4) with respect to each parameter, setting that derivative to zero, and solving the resulting equations. However, this results in a nonlinear system of equations, the solution of which is computationally difficult, in general. Although this least-squares problem is too difficult for most practical applications, the basic least-squares principle can be applied to slightly different formulations with considerable success.

11.1.2 Least-Squares Inverse Model

A formulation based on inverse filtering provides a relatively straightforward and tractable solution for the parameter values in the all-pole model. In any approach to approximation, it is recognized at the outset that the model output will in most cases not be exactly equal to the signal to be modeled. The inverse filtering approach is based on the recognition that if the given signal $s[n]$ is in fact the output of the filter $H(z)$ in the model of Figure 11.1 then with $s[n]$ as the input to the inverse of $H(z)$, the output will be $v[n]$. Consequently, as depicted in Figure 11.2 and with $H(z)$ assumed to be an all-pole system as specified in Eq. (11.2), the inverse filter, whose system function

$$A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k}, \tag{11.5}$$

[2]When used in the context of speech processing, linear predictive analysis is often referred to as *linear predictive coding* (LPC). (See Rabiner and Schafer, 1978 and Quatieri, 2002.)

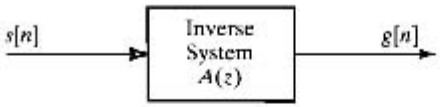


Figure 11.2 Inverse filter formulation for all-pole signal modeling.

is sought so that its output $g[n]$ would be equal to the scaled input $Gv[n]$. In this formulation, then, we choose the parameters of the inverse filter (and therefore implicitly the parameters of the model system) to minimize the mean-squared error between $g[n]$ and $Gv[n]$. As we will see, this leads to a set of well-behaved linear equations.

From Figure 11.2 and Eq. (11.5) it follows that $g[n]$ and $s[n]$ satisfy the difference equation

$$g[n] = s[n] - \sum_{k=1}^{p} a_k s[n-k]. \tag{11.6}$$

The modeling error $\hat{e}[n]$ is now defined as

$$\hat{e}[n] = g[n] - Gv[n] = s[n] - \sum_{k=1}^{p} a_k s[n-k] - Gv[n]. \tag{11.7}$$

If $v[n]$ is an impulse, then, for $n > 0$, the error $\hat{e}[n]$ corresponds to the error between $s[n]$ and the linear prediction of $s[n]$ using the model parameters. Thus, it is convenient to also express Eq. (11.7) as

$$\hat{e}[n] = e[n] - Gv[n], \tag{11.8}$$

where $e[n]$ is the prediction error given by

$$e[n] = s[n] - \sum_{k=1}^{p} a_k s[n-k]. \tag{11.9}$$

For a signal that exactly fits the all-pole model of Eq. (11.3), the modeling error $\hat{e}[n]$ will be zero, and the prediction error $e[n]$ will be the scaled input, i.e.,

$$e[n] = Gv[n]. \tag{11.10}$$

This formulation in terms of inverse filtering leads to considerable simplification, since $v[n]$ is assumed known and $e[n]$ can be computed from $s[n]$ using Eq. (11.9). The parameter values a_k are then chosen to minimize

$$\mathcal{E} = \left\langle |\hat{e}[n]|^2 \right\rangle, \tag{11.11}$$

where the notation $\langle \cdot \rangle$ denotes a summing operation for finite energy deterministic signals and an ensemble averaging operation for random signals. Minimizing $\mathcal{E}$ in Eq. (11.11) results in an inverse filter that minimizes the total energy in the modeling error in the case of deterministic signals or the mean-squared value of the modeling error in the case of random signals. For convenience, we will often refer to $\langle \cdot \rangle$ as the averaging operator where its interpretation as a sum or as an ensemble average should be clear from the context. Again, note that in solving for the parameters a_k specifying the inverse system of Figure 11.2, the all-pole system is implicitly specified, as well.

To find the optimal parameter values, we substitute Eq. (11.8) into Eq. (11.11) to obtain

$$\mathcal{E} = \left\langle (e[n] - Gv[n])^2 \right\rangle. \tag{11.12}$$

or equivalently,

$$\mathcal{E} = \left\langle e^2[n] \right\rangle + G^2 \left\langle v^2[n] \right\rangle - 2G \left\langle v[n]e[n] \right\rangle. \tag{11.13}$$

To find the parameters that minimize $\mathcal{E}$, we differentiate Eq. (11.12) with respect to the i^{th} filter coefficient a_i and set the derivative equal to zero, leading to the set of equations

$$\frac{\partial \mathcal{E}}{\partial a_i} = \frac{\partial}{\partial a_i}\left[\left\langle e^2[n] \right\rangle - 2G \left\langle v[n]s[n-i] \right\rangle\right] = 0, \quad i = 1, 2, \ldots, p. \tag{11.14}$$

where we have assumed that G is independent of a_i and, of course, so is $v[n]$, and consequently that

$$\frac{\partial}{\partial a_i}\left[G^2 \left\langle v^2[n] \right\rangle\right] = 0. \tag{11.15}$$

For models that will be of interest to us, $v[n]$ will be an impulse if $s[n]$ is a causal finite-energy signal and white noise if $s[n]$ is a wide-sense stationary random process. With $v[n]$ an impulse and $s[n]$ zero for $n < 0$, the product $v[n]s[n-i] = 0$ for $i = 1, 2, \ldots p$. With $v[n]$ as white noise,

$$\langle v[n]s[n-i] \rangle = 0, \qquad i = 1, 2, \ldots p, \tag{11.16}$$

since for any value of n, the input of a causal system with white-noise input is uncorrelated with the output values prior to time n. Thus, for both cases, Eq. (11.14) reduces to

$$\frac{\partial \mathcal{E}}{\partial a_i} = \frac{\partial}{\partial a_i}\left\langle e^2[n] \right\rangle = 0 \quad i = 1, 2, \ldots, p \tag{11.17}$$

In other words, choosing the coefficients to minimize the average squared modeling error $\langle \hat{e}^2[n] \rangle$ is equivalent to minimizing the average squared prediction error $\langle e^2[n] \rangle$. Expanding Eq. (11.17) and invoking the linearity of the averaging operator, we obtain from Eq. (11.17) the equations

$$\langle s[n]s[n-i] \rangle - \sum_{k=1}^{p} a_k \langle s[n-k]s[n-i] \rangle = 0, \quad i = 1, \ldots, p. \tag{11.18}$$

Defining

$$\phi_{ss}[i, k] = \langle s[n-i]s[n-k] \rangle, \tag{11.19}$$

Eqs. (11.18) can be rewritten more compactly as

$$\sum_{k=1}^{p} a_k \phi_{ss}[i, k] = \phi_{ss}[i, 0], \quad i = 1, 2, \ldots, p. \tag{11.20}$$

Equations (11.20) comprise a system of p linear equations in p unknowns. Computation of the parameters of the model can be achieved by solving the set of linear equations for the parameters a_k for $k = 1, 2, \ldots, p$, using known values for $\phi_{ss}[i, k]$ for $i = 1, 2, \ldots, p$ and $k = 0, 1, \ldots, p$ or first computing them from $s[n]$.

11.1.3 Linear Prediction Formulation of All-Pole Modeling

As suggested earlier, an alternative and useful interpretation of all-pole signal modeling stems from the interpretation of Eq. (11.3) as a linear prediction of the output in terms of past values, with the prediction error $e[n]$ being the scaled input $Gv[n]$, i.e.,

$$e[n] = s[n] - \sum_{k=1}^{p} a_k s[n-k] = Gv[n]. \tag{11.21}$$

As indicated by Eq. (11.17), minimizing the inverse modeling error $\mathcal{E}$ in Eq. (11.11) is equivalent to minimizing the averaged prediction error $\langle e^2[n] \rangle$. If the signal $s[n]$ were produced by the model system, and if $v[n]$ is an impulse, and if $s[n]$ truly fits the all-pole model, then the signal at any $n > 0$ is linearly predictable from past values, i.e., the prediction error is zero. If $v[n]$ is white noise, then the prediction error is white.

The interpretation in terms of prediction is depicted in Figure 11.3, where the transfer function of the prediction filter $P(z)$ is

$$P(z) = \sum_{k=1}^{p} a_k z^{-k}. \tag{11.22}$$

This system is referred to as the p^{th}-order *linear predictor* for the signal $s[n]$. Its output is

$$\tilde{s}[n] = \sum_{k=1}^{p} a_k s[n-k], \tag{11.23}$$

and as Figure 11.3 shows, the prediction error signal is $e[n] = s[n] - \tilde{s}[n]$. The sequence $e[n]$ represents the amount by which the linear predictor fails to exactly predict the signal $s[n]$. For this reason, $e[n]$ is also sometimes called the *prediction error residual* or simply the *residual*. With this point of view, the coefficients a_k are called the *prediction coefficients*. As is also shown in Figure 11.3, the prediction error filter is related to the linear predictor by

$$A(z) = 1 - P(z) = 1 - \sum_{k=1}^{p} a_k z^{-k}. \tag{11.24}$$

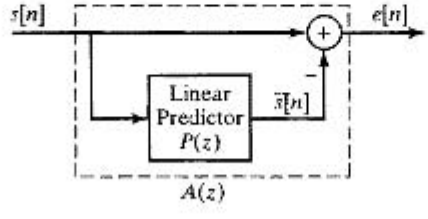


Figure 11.3 Linear prediction formulation for all-pole signal modeling.

11.2 DETERMINISTIC AND RANDOM SIGNAL MODELS

To use the optimum inverse filter or equivalently the optimum linear predictor as a basis for parametric signal modeling, it is necessary to be more specific about the assumed input $v[n]$ and about the method of computing the averaging operator $\langle\cdot\rangle$. To this end, we consider separately the case of deterministic signals and the case of random signals. In both cases, we will use averaging operations that assume knowledge of the signal to be modeled over all time $-\infty < n < \infty$. In Section 11.3, we discuss some of the practical considerations when only a finite-length segment of the signal $s[n]$ is available.

11.2.1 All-Pole Modeling of Finite-Energy Deterministic Signals

In this section, we assume an all-pole model that is causal and stable and also that both the input $v[n]$ and the signal $s[n]$ to be modeled are zero for $n < 0$. We further assume that $s[n]$ has finite energy and is known for all $n \geq 0$. We choose the operator $\langle\cdot\rangle$ in Eq. (11.11) as the total energy in the modeling error sequence $\hat{e}[n]$, i.e.,

$$\mathcal{E} = \left\langle |\hat{e}[n]|^2 \right\rangle = \sum_{n=-\infty}^{\infty} |\hat{e}[n]|^2. \tag{11.25}$$

With this definition of the averaging operator, $\phi_{ss}[i, k]$ in Eq. (11.19) is given by

$$\phi_{ss}[i, k] = \sum_{n=-\infty}^{\infty} s[n-i]s[n-k], \tag{11.26}$$

and equivalently,

$$\phi_{ss}[i, k] = \sum_{n=-\infty}^{\infty} s[n]s[n-(i-k)]. \tag{11.27}$$

The coefficients $\phi_{ss}[i, k]$ in Eq. (11.20) are now

$$\phi_{ss}[i, k] = r_{ss}[i-k], \tag{11.28}$$

where for real signals $s[n]$, $r_{ss}[m]$ is the deterministic autocorrelation function

$$r_{ss}[m] = \sum_{n=-\infty}^{\infty} s[n+m]s[n] = \sum_{n=-\infty}^{\infty} s[n]s[n-m]. \tag{11.29}$$

Therefore, Eq. (11.20) takes the form

$$\sum_{k=1}^{p} a_k r_{ss}[i-k] = r_{ss}[i] \quad i = 1, 2, \ldots, p. \tag{11.30}$$

These equations are called the *autocorrelation normal equations* and also the *Yule–Walker equations*. They provide a basis for computing the parameters $a_1, \ldots, a_p$ from the autocorrelation function of the signal. In Section 11.2.5, we discuss an approach to choosing the gain factor G.

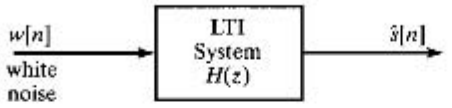


Figure 11.4 Linear system model for a random signal $s[n]$.

11.2.2 Modeling of Random Signals

For all-pole modeling of zero-mean, wide-sense stationary, random signals, we assume that the input to the all-pole model is zero-mean, unit-variance, white noise as indicated in Figure 11.4. The difference equation for this system is

$$\hat{s}[n] = \sum_{k=1}^{p} a_k \hat{s}[n-k] + Gw[n], \tag{11.31}$$

where the input has autocorrelation function $E\{w[n+m]w[n]\} = \delta[m]$, zero mean ($E\{w[n]\} = 0$), and unit average power ($E\{(w[n])^2\} = \delta[0] = 1$), with $E\{\cdot\}$ representing the expectation or probability average operator.[3]

The resulting model for analysis is the same as that depicted in Figure 11.2, but the desired output $g[n]$ changes. In the case of random signals, we want to make $g[n]$ as much like a white-noise signal as possible, rather than the unit sample sequence that was desired in the deterministic case. For this reason, the optimal inverse filter for random signals is often referred to as a *whitening filter*.

We also choose the operator $\langle\cdot\rangle$ in Eq. (11.11) as an appropriate one for random signals, specifically the mean-squared value or equivalently the average power. Then Eq. (11.11) becomes

$$\mathcal{E} = E\{(\hat{e}[n])^2\}. \tag{11.32}$$

If $s[n]$ is assumed to be a sample function of a stationary random process, then $\phi_{ss}[i,k]$ in Eq. (11.19) would be the autocorrelation function

$$\phi_{ss}[i,k] = E\{s[n-i]s[n-k]\} = r_{ss}[i-k]. \tag{11.33}$$

The system coefficients can be found as before from Eq. (11.20). Thus, the system coefficients satisfy a set of equations of the same form as Eq. (11.30), i.e.,

$$\sum_{k=1}^{p} a_k r_{ss}[i-k] = r_{ss}[i], \quad i = 1, 2, \ldots, p. \tag{11.34}$$

Therefore, modeling random signals again results in the Yule–Walker equations, with the autocorrelation function in this case being defined by the probabilistic average

$$r_{ss}[m] = E\{s[n+m]s[n]\} = E\{s[n]s[n-m]\}. \tag{11.35}$$

[3]Computation of $E\{\cdot\}$ requires knowledge of the probability densities. In the case of stationary random signals, only one density is required. In the case of ergodic random processes, a single infinite time average could be used. In practical applications, however, such averages must be approximated by estimates obtained from finite time averages.

11.2.3 Minimum Mean-Squared Error

For modeling of either deterministic signals (Section 11.2.1) or random signals (Section 11.2.2) the minimum value of the prediction error $e[n]$ in Figure 11.3 can be expressed in terms of the corresponding correlation values in Eq. (11.20) to find the optimum predictor coefficients. To see this, we write $\mathcal{E}$ as

$$\mathcal{E} = \left\langle \left(s[n] - \sum_{k=1}^{p} a_k s[n-k] \right)^2 \right\rangle. \tag{11.36}$$

As outlined in more detail in Problem 11.2, if Eq. (11.36) is expanded, and Eq. (11.20) is substituted into the result, it follows that in general,

$$\mathcal{E} = \phi_{ss}[0,0] - \sum_{k=1}^{p} a_k \phi_{ss}[0,k]. \tag{11.37}$$

Equation (11.37) is true for any appropriate choice of the averaging operator. In particular, for averaging definitions for which $\phi_{ss}[i,k] = r_{ss}[i-k]$, Eq. (11.37) becomes

$$\mathcal{E} = r_{ss}[0] - \sum_{k=1}^{p} a_k r_{ss}[k]. \tag{11.38}$$

11.2.4 Autocorrelation Matching Property

An important and useful property of the all-pole model resulting from the solution of Eq. (11.30) for deterministic signals and Eq. (11.34) for random signals is referred to as the autocorrelation matching property (Makhoul, 1973). Equations (11.30) and (11.34) represent a set of p equations to be solved for the model parameters a_k for $k = 1, \ldots, p$. The coefficients in these equations on both the left- and right-hand sides of the equations are comprised of the $(p+1)$ correlation values $r_{ss}[m], m = 0, 1, \ldots, p$, where the correlation function is appropriately defined, depending on whether the signal to be modeled is deterministic or random.

The basis for verifying the autocorrelation matching property is to observe that the signal $\hat{s}[n]$ obviously fits the model when the model system $H(z)$ in Figure 11.1 is specified as the all-pole system in Eq. (11.2). If we were to consider again applying all-pole modeling to $\hat{s}[n]$, we would of course again obtain Eqs. (11.30) or (11.34), but this time, with $r_{\hat{s}\hat{s}}[m]$ in place of $r_{ss}[m]$. The solution must again be the same parameter values $a_k, k = 1, 2, \ldots, p$, since $\hat{s}[n]$ fits the model, and this solution will result if

$$r_{ss}[m] = c r_{\hat{s}\hat{s}}[m] \quad 0 \le m \le p, \tag{11.39}$$

where c is any constant. The fact that the equality in Eq. (11.39) is required follows from the form of the recursive solution of the Yule–Walker equations as developed in Section 11.6. In words, the autocorrelation normal equations require that for the lags $|m| = 0, 1, \ldots, p$ the autocorrelation functions of the model output and the signal being modeled are proportional.

11.2.5 Determination of the Gain Parameter G

With the approach that we have taken, determination of the optimal choice for the coefficients a_k of the model does not depend on the system gain G. From the perspective of the inverse filtering formulation in Figure 11.2, one possibility is to choose G so that $\langle(\tilde{s}[n])^2\rangle = \langle(s[n])^2\rangle$. For finite-energy deterministic signals, this corresponds to matching the total energy in the model output to the total energy in the signal that is being modeled. For random signals, it is the average power that is matched. In both cases, this corresponds to choosing G, so that $r_{\tilde{s}\tilde{s}}[0] = r_{ss}[0]$. With this choice, the proportionality factor c in Eq. (11.39) is unity.

Example 11.1 1st-Order System

Figure 11.5 shows two signals, both of which are outputs of a 1st-order system with system function

$$H(z) = \frac{1}{1 - \alpha z^{-1}}. \tag{11.40}$$

The signal $s_d[n] = h[n] = \alpha^n u[n]$ is the output when the input is a unit impulse $\delta[n]$, while the signal $s_r[n]$ is the output when the input to the system is a zero mean, unit variance white-noise sequence. Both signals extend over the range $-\infty < n < \infty$, as suggested by Figure 11.5.

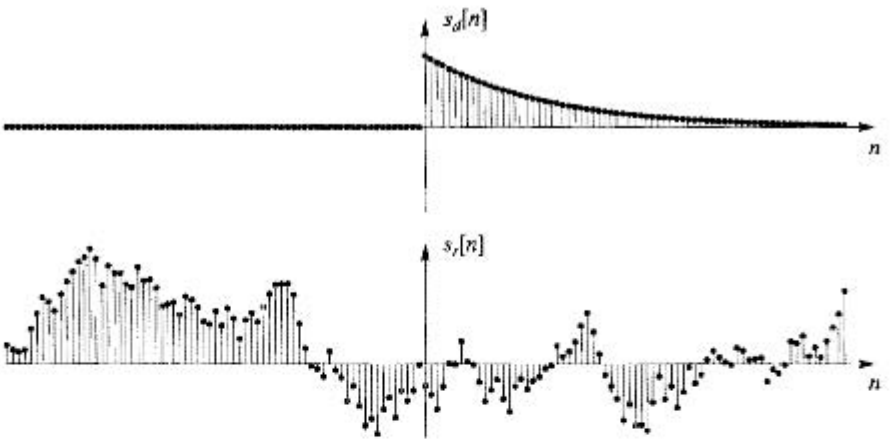


Figure 11.5 Examples of deterministic and random outputs of a 1st-order all-pole system.

The autocorrelation function for the signal $s_d[n]$ is

$$r_{s_d s_d}[m] = r_{hh}[m] = \sum_{n=0}^{\infty} \alpha^{n+m}\alpha^n = \frac{\alpha^{|m|}}{1 - \alpha^2}, \tag{11.41}$$

the autocorrelation function of $s_r[n]$ is also given by Eq. (11.41) since $s_r[n]$ is the response of the system to white noise, for which the autocorrelation function is a unit impulse.

Since both signals were generated with a 1st-order all-pole system, a 1st-order all-pole model will be an exact fit. In the deterministic case, the output of the optimum

inverse filter will be a unit impulse, and in the random signal case, the output of the optimum inverse filter will be a zero-mean white-noise sequence with unit average power. To show that the optimum inverse filter will be exact, note that for a 1st-order model, Eqs. (11.30) or (11.34) reduce to

$$r_{s_d s_d}[0]a_1 = r_{s_d s_d}[1], \tag{11.42}$$

so from Eq. (11.41), it follows that the optimum predictor coefficient for both the deterministic and the random signal is

$$a_1 = \frac{r_{s_d s_d}[1]}{r_{s_d s_d}[0]} = \frac{\dfrac{\alpha}{1-\alpha^2}}{\dfrac{1}{1-\alpha^2}} = \alpha. \tag{11.43}$$

From Eq. (11.38), the minimum mean-squared error is

$$\mathcal{E} = \frac{1}{1-\alpha^2} - a_1 \frac{\alpha}{1-\alpha^2} = \frac{1-\alpha^2}{1-\alpha^2} = 1, \tag{11.44}$$

which is the size of the unit impulse in the deterministic case and the average power of the white-noise sequence in the random case.

As mentioned earlier, and as is clear in this example, when the signal is generated by an all-pole system excited by either an impulse or white noise, all-pole modeling can determine the parameters of the all-pole system exactly. This requires prior knowledge of the model order p and the autocorrelation function. This was possible to obtain for this example, because a closed-form expression was available for the infinite sum required to compute the autocorrelation function. In a practical setting, it is generally necessary to estimate the autocorrelation function from a finite-length segment of the given signal. Problem 11.14 considers the effect of finite autocorrelation estimates (to be discussed next) for the deterministic signal $s_d[n]$ of this section.

11.3 ESTIMATION OF THE CORRELATION FUNCTIONS

To use the results of Sections 11.1 and 11.2 for modeling of either deterministic or random signals, we require *apriori* knowledge of the correlation functions $\phi_{ss}[i, k]$ that are needed to form the system equations satisfied by the coefficients a_k, or we must estimate these from the given signal. Furthermore, we may want to apply block processing or short-time analysis techniques to represent the time-varying properties of a nonstationary signal, such as speech. In this section, we will discuss two distinct approaches to the computation of the correlation estimates for practical application of the concepts of parametric signal modeling. These two approaches have come to be known as the *autocorrelation method* and the *covariance method*.

11.3.1 The Autocorrelation Method

Suppose that we have available a set of $M+1$ signal samples $s[n]$ for $0 \leq n \leq M$, and we wish to compute the coefficients for an all-pole model. In the autocorrelation method, it is assumed that the signal ranges over $-\infty < n < \infty$, with the signal samples taken to

be zero for all n outside the interval $0 \leq n \leq M$, even if they have been extracted from a longer sequence. This, of course, imposes a limit to the exactness that can be expected of the model, since the IIR impulse response of an all-pole model will be used to model the finite-length segment of $s[n]$.

Although the prediction error sequence need not be computed explicitly to solve for the filter coefficients, it is nevertheless informative to consider its computation in some detail. The impulse response of the prediction error filter is, by the definition of $A(z)$, in Eq. (11.24),

$$h_A[n] = \delta[n] - \sum_{k=1}^{p} a_k \delta[n-k]. \tag{11.45}$$

It can be seen that since the signal $s[n]$ has finite length $M+1$ and $h_A[n]$, the impulse response of the prediction filter $A[z]$, has length $p+1$, the prediction error sequence $e[n] = h_A[n] * s[n]$ will always be identically zero outside the interval $0 \leq n \leq M+p$. Figure 11.6 shows an example of the prediction error signal for a linear predictor with $p = 5$. In the upper plot, $h_A[n-m]$ the (time-reversed and shifted) impulse response of the prediction error filter, is shown as a function of m for three different values of n. The dark lines with square dots depict $h_A[n-m]$, and the lighter lines with round dots show the sequence $s[m]$ for $0 \leq m \leq 30$. On the left side is $h_A[0-m]$, which shows that the first nonzero prediction error sample is $e[0] = s[0]$. This, of course, is consistent with Eq. (11.9). On the extreme right is $h_A[M+p-m]$, which shows that the last nonzero error sample is $e[M+p] = -a_p s[M]$. The second plot in Figure 11.6 shows the error signal $e[n]$ for $0 \leq n \leq M+p$. From the point of view of linear prediction, it follows that the first p samples (dark lines and dots) are predicted from samples that are assumed to be zero. Similarly, the samples of the input for $n \geq M+1$ are assumed to be zero to obtain a finite-length signal. The linear predictor attempts to predict the zero samples

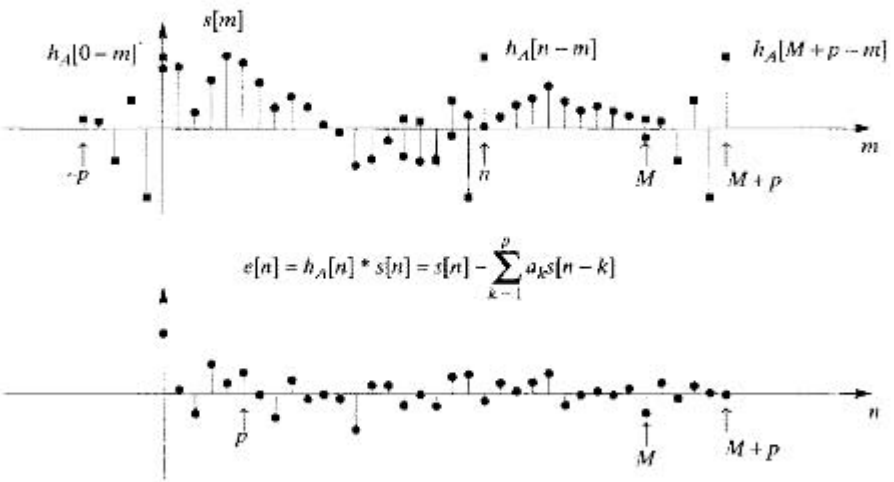


Figure 11.6 Illustration (for $p = 5$) of computation of prediction error for the autocorrelation method. (Square dots denote samples of $h_A[n-m]$ and light round dots denote samples of $s[m]$ for the upper plot and $e[n]$ for the lower plot.)

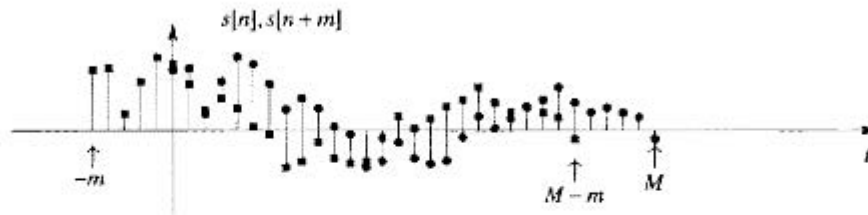


Figure 11.7 Illustration of computation of the autocorrelation function for a finite-length sequence. (Square dots denote samples of $s[n+m]$, and light round dots denote samples of $s[n]$.)

in the interval $M+1 \leq n \leq M+p$ from prior samples that are nonzero and part of the original signal. Indeed, if $s[0] \neq 0$ and $s[M] \neq 0$, then it will be true that both $e[0] = s[0]$ and $e[M+p] = -a_p s[M]$ will be nonzero. That is, the prediction error (total-squared error $\mathcal{E}$) can never be exactly zero if the signal is defined to be zero outside the interval $0 \leq n \leq M$. Furthermore, the total-squared prediction error for a p^{th}-order predictor would be

$$\mathcal{E}^{(p)} = \left\langle e[n]^2 \right\rangle = \sum_{n=-\infty}^{\infty} e[n]^2 = \sum_{n=0}^{M+p} e[n]^2, \tag{11.46}$$

i.e., the limits of summation can be infinite for convenience, but practically speaking, they are finite.

When the signal is assumed to be identically zero outside the interval $0 \leq n \leq M$, the correlation function $\phi_{ss}[i,k]$ reduces to the autocorrelation function $r_{ss}[m]$ where the values needed in Eq. (11.30) are for $m = |i-k|$. Figure 11.7 shows the shifted sequences used in computing $r_{ss}[m]$ with $s[n]$ denoted by round dots and $s[n+m]$ by square dots. Note that for a finite-length signal, the product $s[n]s[n+m]$ is nonzero only over the interval $0 \leq n \leq M-m$ when $m \geq 0$. Since r_{ss} is an even function, i.e., $r_{ss}[-m] = r_{ss}[m] = r_{ss}[|m|]$ it follows that the autocorrelation values needed for the Yule- Walker equations can be computed as,

$$r_{ss}[|m|] = \sum_{n=-\infty}^{\infty} s[n]s[n+|m|] = \sum_{n=0}^{M-|m|} s[n]s[n+|m|]. \tag{11.47}$$

For the finite-length sequence $s[n]$, Eq. (11.47) has all the necessary properties of an autocorrelation function and $r_{ss}[m] = 0$ for $m > M$. But of course $r_{ss}[m]$ is not the same as the autocorrelation function of the infinite length signal from which the segment was extracted.

Equation (11.47) can be used to compute estimates of the autocorrelation function for either deterministic or random signals.[4] Often, the finite-length input signal is extracted from a longer sequence of samples. This is the case, for example, in applications to speech processing, where voiced segments (e.g., vowel sounds) of speech are treated as deterministic and unvoiced segments (fricative sounds) are treated as

[4] In the context of random signals, it was shown in Section 10.6 that Eq. (11.47) is a biased estimate of the autocorrelation function. When $p \ll M$ as is often the case, this statistical bias is generally negligible.

random signals.[5] According to the previous discussion, the first p and last p samples of the prediction error can be large due to the attempt to predict nonzero samples from zero samples and to predict zero samples from nonzero samples. Since this can bias the estimation of the predictor coefficients, a signal-tapering window, such as a Hamming window is generally applied to the signal before computation of the autocorrelation function.

11.3.2 The Covariance Method

An alternative choice for the averaging operator for the prediction error for a p^{th}-order predictor is

$$\mathcal{E}_{\text{cov}}^{(p)} = \left\langle (e[n])^2 \right\rangle = \sum_{n=p}^{M} (e[n])^2. \tag{11.48}$$

As in the autocorrelation method, the averaging is over a finite interval ($p \leq n \leq M$), but the difference is that the signal to be modeled is known over the larger interval $0 \leq n \leq M$. The total-squared prediction error only includes values of $e[n]$ that can be computed from samples within the interval $0 \leq n \leq M$. Consequently, the averaging takes place over a shorter interval $p \leq n \leq M$. This is significant, since it relieves the inconsistency between the all-pole model and the finite-length signal.[6] In this case, we only seek to match the signal over a finite interval rather than over all n as in the autocorrelation method. The upper plot in Figure 11.8 shows the same signal $s[m]$ as

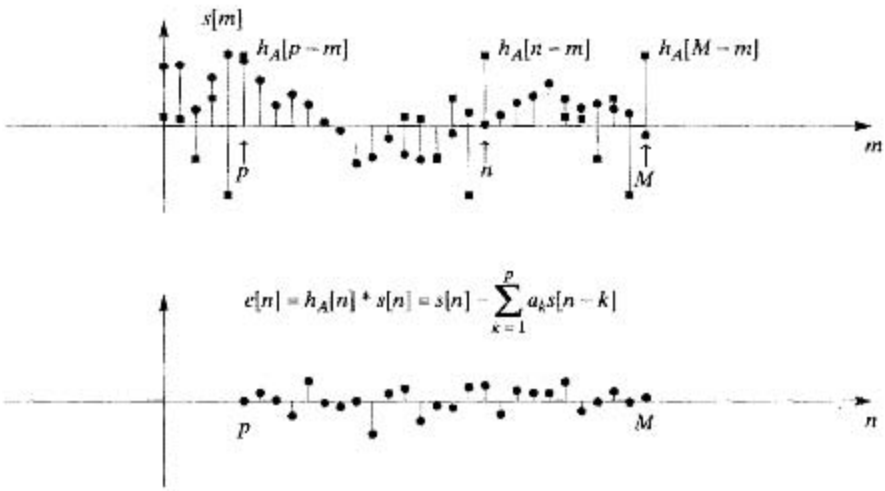


Figure 11.8 Illustration (for $p = 5$) of computation of prediction error for the covariance method. (In upper plot, square dots denote samples of $h_A[n-m]$, and light round dots denote samples of $s[m]$.)

[5] In both cases, the deterministic autocorrelation function in Eq. (11.47) is used as an estimate.

[6] The definitions of total-squared prediction error in Eqs. (11.48) and (11.46) are distinctly different, so we use the subscript $_{\text{cov}}$ to distinguish them.

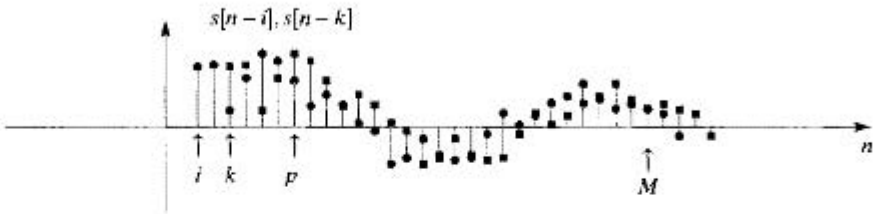


Figure 11.9 Illustration of computation of covariance function for a finite-length sequence. (Square dots denote samples of $s[n-k]$ and light round dots denote samples of $s[n-i]$.)

in the upper part of Figure 11.6, but in this case, the prediction error is only computed over the interval $p \le n \le M$ as needed in Eq. (11.48). As shown by the prediction error filter impulse responses $h_A[n-m]$ in the upper plot, there are no end effects when the prediction error is computed in this way, since all the signal samples needed to compute the prediction error are available. Because of this, it is possible for the prediction error to be precisely zero over the entire interval $p \le n \le M$, if the signal from which the finite length segment was extracted was generated as the output of an all-pole system. Seen another way, if $s[n]$ is the output of an all-pole system with an input that is zero for $n > 0$, then as seen from Eqs. (11.9) and (11.10) the prediction error will be zero for $n > 0$.

The covariance function inherits the same definition of the averaging operator, i.e.,

$$\phi_{ss}[i,k] = \sum_{n=p}^{M} s[n-i]s[n-k]. \tag{11.49}$$

The shifted sequences $s[n-i]$ (light lines and round dots) and $s[n-k]$ (dark lines and square dots) are shown in Figure 11.9. This figure shows that since we need $\phi_{ss}[i,k]$ only for $i = 0, 1, \ldots, p$ and $k = 1, 2, \ldots, p$, the segment $s[n]$ for $0 \le n \le M$ contains all the samples that are needed to compute $\phi_{ss}[i,k]$ in Eq. (11.49).

11.3.3 Comparison of Methods

The autocorrelation and covariance methods have many similarities, but there are many important differences in the methods and the resulting all-pole models. In this section, we summarize some of the differences that we have already demonstrated and call attention to some others.

Prediction Error

Both the averaged prediction error $\langle e^2[n] \rangle$ and averaged modeling error $\langle \hat{e}^2[n] \rangle$ are nonnegative and nonincreasing with increasing model order p. In the autocorrelation method based on estimates obtained from finite-length signals, the averaged modeling or prediction error will never be zero, because the autocorrelation values will not be exact. Furthermore, the minimum value of the prediction error even with an exact model is $Gv[n]$ as indicated by Eq. (11.10). In the covariance method, the prediction error for $n > 0$ can be exactly zero if the original signal was generated by an all-pole model. This will be demonstrated in Example 11.2.

Equations for Predictor Coefficients

In both methods, the predictor coefficients that minimize the averaged prediction error satisfy a general set of linear equations expressed in matrix form as $\mathbf{\Phi a} = \boldsymbol{\psi}$. The coefficients of the all-pole model are obtained by inverting the matrix $\mathbf{\Phi}$; i.e., $\boldsymbol{a} = \mathbf{\Phi}^{-1}\boldsymbol{\psi}$. In the covariance method, the elements $\phi_{ss}[i, k]$ of the matrix $\mathbf{\Phi}$ are computed using Eq. (11.49). In the autocorrelation method, the covariance values become autocorrelation values, i.e., $\phi_{ss}[i, k] = r_{ss}[|i - k|]$ and are computed using Eq. (11.47). In both cases, the matrix $\mathbf{\Phi}$ is symmetric and positive-definite, but in the autocorrelation method, the matrix $\mathbf{\Phi}$ is also a Toeplitz matrix. This implies numerous special properties of the solution, and it implies that the solution of the equations can be done more efficiently than would be true in general. In Section 11.6, we will explore some of these implications for the autocorrelation method.

Stability of the Model System

The prediction error filter has a system function $A(z)$ that is a polynomial in z^{-1}. Therefore, it can be represented in terms of its zeros as

$$A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} = \prod_{k=1}^{p} (1 - z_k z^{-1}). \tag{11.50}$$

In the autocorrelation method, the zeros of the prediction error filter $A(z)$ are guaranteed to lie strictly within the unit circle of the z plane; i.e., $|z_k| < 1$. This means that the poles of the causal system function $H(z) = G/A(z)$ of the model lie inside the unit circle, which implies that the model system is stable. A simple proof of this assertion is given by Lang and McClellan (1979) and McClellan (1988). Problem 11.10 develops a proof that depends on the lattice filter interpretation of the prediction error system to be discussed in Section 11.7.1. In the covariance method as we have formulated it, no such guarantee can be given.

11.4 MODEL ORDER

An important issue in parametric signal modeling is the model order p, the choice of which has a major impact on the accuracy of the model. A common approach to choosing p is to examine the averaged prediction error (often referred to as the residual) from the optimum p^{th}-order model. Let $a_k^{(p)}$ be the parameters for the optimal p^{th}-order predictor found using Eq. (11.30). The prediction error energy for the p^{th}-order model using the autocorrelation method is[7]

$$\mathcal{E}^{(p)} = \sum_{n=-\infty}^{\infty} \left(s[n] - \sum_{k=1}^{p} a_k^{(p)} s[n-k] \right)^2. \tag{11.51}$$

For the zero$^{\text{th}}$-order predictor, ($p = 0$), there are no delay terms in Eq. (11.51), i.e., the "predictor" is just the identity system so $e[n] = s[n]$. Consequently, for $p = 0$,

$$\mathcal{E}^{(0)} = \sum_{n=-\infty}^{\infty} s^2[n] = r_{ss}[0]. \tag{11.52}$$

[7]Recall that $\mathcal{E}_{\text{cov}}^{(p)}$ denotes the total-squared prediction error for the covariance method, while we use $\mathcal{E}^{(p)}$ with no subscript to denote the total-squared prediction error for the autocorrelation method.

Plotting the normalized mean-squared prediction error $\mathcal{V}^{(p)} = \mathcal{E}^{(p)}/\mathcal{E}^{(0)}$ as a function of p shows how increasing p changes this error energy. In the autocorrelation method, we showed that the averaged prediction error can never be precisely zero, even if the signal $s[n]$ was generated by an all-pole system, and the model order is the same as the order of the generating system. In the covariance method, however, if the all-pole model is a perfect model for the signal $s[n]$, $\mathcal{E}_{\text{cov}}^{(p)}$ will become identically zero at the correct choice of p, since the averaged prediction error only considers values for $p \leq n \leq M$. Even if $s[n]$ is not perfectly modeled by an all-pole system, there is often a value of p above which increasing p has little or no effect on either $\mathcal{V}^{(p)}$ or $\mathcal{V}_{\text{cov}}^{(p)} = \mathcal{E}_{\text{cov}}^{(p)}/\mathcal{E}_{\text{cov}}^{(0)}$. This threshold is an efficient choice of model order for representing the signal as an all-pole model.

Example 11.2 Model Order Selection

To demonstrate the effect of model order, consider a signal $s[n]$ generated by exciting a 10^{th}-order system

$$H(z) = \frac{0.6}{(1 - 1.03z^{-1} + 0.79z^{-2} - 1.34z^{-3} + 0.78z^{-4} - 0.92z^{-5} + 1.22z^{-6} - 0.43z^{-7} + 0.6z^{-8} - 0.29z^{-9} - 0.23z^{-10})} \tag{11.53}$$

with an impulse $v[n] = \delta[n]$. The samples of $s[n]$ for $0 \leq n \leq 30$ are shown as the sequence in the upper plots in Figures 11.6 and 11.8. This signal was used as the signal to be modeled by an all-pole model with both the autocorrelation method and the covariance method. Using the 31 samples of $s[n]$, the appropriate autocorrelation and covariance values were computed and the predictor coefficients computed by solving Eqs. (11.30) and (11.34) respectively. The normalized mean-squared prediction errors are plotted in Figure 11.10. Note that in both the autocorrelation and covariance methods the normalized error decreases abruptly at $p = 1$ in both plots, then decreasing more slowly as p increases. At $p = 10$, the covariance method gives zero error, while the autocorrelation method gives a nonzero averaged error for $p \geq 10$. This is consistent with our discussion of the prediction error in Section 11.3.

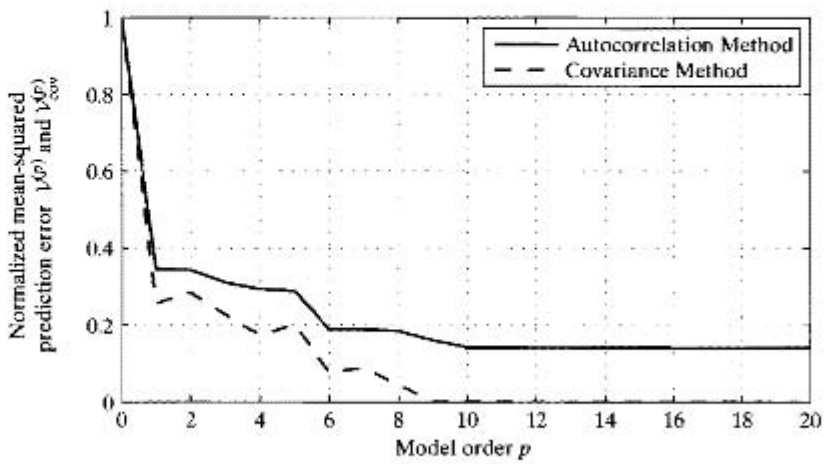


Figure 11.10 Normalized mean-squared prediction error $\mathcal{V}^{(p)}$ as a function of model order p in Example 11.2.

While Example 11.2 is an ideal simulation, the general nature of the dependence of averaged prediction error as a function of p is typical of what happens when all-pole modeling is applied to sampled signals. The graph of $\mathcal{V}^{(p)}$ as a function of p tends to flatten out at some point, and that value of p is often selected as the value to be used in the model. In applications such as speech analysis, it is possible to choose the model order based on physical models for the production of the signal to be modeled. (See Rabiner and Schafer, 1978.)

11.5 ALL-POLE SPECTRUM ANALYSIS

All-pole signal modeling provides a method of obtaining high-resolution estimates of a signal's spectrum from truncated or windowed data. The use of parametric signal modeling in spectrum analysis is based on the fact that if the data fits the model, then a finite segment of the data can be used to determine the model parameters and, consequently, also its spectrum. Specifically, in the deterministic case

$$|\hat{S}(e^{j\omega})|^2 = |H(e^{j\omega})|^2 |V(e^{j\omega})|^2 = |H(e^{j\omega})|^2 \tag{11.54}$$

since $|V(e^{j\omega})|^2 = 1$ for a unit impulse excitation to the model system. Likewise, for random signals the power spectrum of the output of the model is

$$P_{\hat{s}\hat{s}}(e^{j\omega}) = |H(e^{j\omega})|^2 P_{ww}(e^{j\omega}) = |H(e^{j\omega})|^2, \tag{11.55}$$

since $P_{ww}(e^{j\omega}) = 1$ for the white-noise input. Thus, we can obtain an estimate of the spectrum of a signal $s[n]$ by computing an all-pole model for the signal and then computing the magnitude-squared of the frequency response of the model system. For both the deterministic and random cases, the spectrum estimate takes the form

$$\text{Spectrum estimate} = |H(e^{j\omega})|^2 = \left| \frac{G}{1 - \sum_{k=1}^{p} a_k e^{-j\omega k}} \right|^2. \tag{11.56}$$

To obtain an understanding of the nature of the spectrum estimate in Eq. (11.56) for the deterministic case, it is useful to recall that the DTFT of the finite-length signal $s[n]$ is

$$S(e^{j\omega}) = \sum_{n=0}^{M} s[n] e^{-j\omega n}. \tag{11.57}$$

Furthermore, note that

$$r_{ss}[m] = \sum_{n=0}^{M-|m|} s[n+m]s[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} |S(e^{j\omega})|^2 e^{j\omega m} d\omega, \tag{11.58}$$

where, due to the finite length of $s[n]$, $r_{ss}[m] = 0$ for $|m| > M$. The values of $r_{ss}[m]$ for $m = 0, 1, 2, \ldots, p$ are used in the computation of the all-pole model using the autocorrelation method. Thus, it is reasonable to suppose that there is a relationship

between the Fourier spectrum of the signal, $|S(e^{j\omega})|^2$, and the all-pole model spectrum, $|\hat{S}(e^{j\omega})|^2 = |H(e^{j\omega})|^2$.

One approach to illuminating this relationship is to obtain an expression for the averaged prediction error in terms of the DTFT of the signal $s[n]$. Recall that the prediction error is $e[n] = h_A[n] * s[n]$, where $h_A[n]$ is the impulse response of the prediction error filter. From Parseval's Theorem, the averaged prediction error is

$$\mathcal{E} = \sum_{n=0}^{M+p} (e[n])^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |S(e^{j\omega})|^2 |A(e^{j\omega})|^2 d\omega. \tag{11.59}$$

where $S(e^{j\omega})$ is the DTFT of $s[n]$ as given by Eq. (11.57). Since $H(z) = G/A(z)$, Eq. (11.59) can be expressed in terms of $H(e^{j\omega})$ as

$$\mathcal{E} = \frac{G^2}{2\pi} \int_{-\pi}^{\pi} \frac{|S(e^{j\omega})|^2}{|H(e^{j\omega})|^2} d\omega. \tag{11.60}$$

Since the integrand in Eq. (11.60) is positive, and $|H(e^{j\omega})|^2 > 0$ for $-\pi < \omega \leq \pi$, it therefore follows from Eq. (11.60) that minimizing $\mathcal{E}$ is equivalent to minimizing the ratio of the energy spectrum of the signal $s[n]$ to the magnitude-squared of the frequency response of the linear system in the all-pole model. The implication of this is that the all-pole model spectrum will attempt to match the energy spectrum of the signal more closely at frequencies where the signal spectrum is large, since frequencies where $|S(e^{j\omega})|^2 > |H(e^{j\omega})|^2$ contribute more to the mean-squared error than frequencies where the opposite is true. Thus, the all-pole model spectrum estimate favors a good fit around the peaks of the signal spectrum. This will be illustrated by the discussion in Section 11.5.1. Similar analysis and reasoning also applies to the case in which $s[n]$ is random.

11.5.1 All-Pole Analysis of Speech Signals

All-pole modeling is widely used in speech processing both for speech coding, where the term linear predictive coding (LPC) is often used, and for spectrum analysis. (See Atal and Hanauer, 1971, Makhoul, 1975, Rabiner and Schafer, 1978, and Quatieri, 2002.) To illustrate many of the ideas discussed in this chapter, we discuss in some detail the use of all-pole modeling for spectrum analysis of speech signals. This method is typically applied in a time-dependent manner by periodically selecting short segments of the speech signal for analysis in much the same way as is done in time-dependent Fourier analysis as discussed in Section 10.3. Since the time-dependent Fourier transform is essentially a sequence of DTFTs of finite-length segments, the above discussion of the relationship between the DTFT and the all-pole spectrum characterizes the relationship between time-dependent Fourier analysis and time-dependent all-pole model spectrum analysis, as well.

Figure 11.11 shows a 201-point Hamming-windowed segment of a speech signal $s[n]$ in the top panel and the corresponding autocorrelation function $r_{ss}[m]$ below. During this time interval, the speech signal is voiced (vocal cords vibrating), as evidenced by the periodic nature of the signal. This periodicity is reflected in the autocorrelation function as the peak at about 27 samples (27/8 = 3.375 ms for 8 kHz sampling rate) and integer multiples thereof.

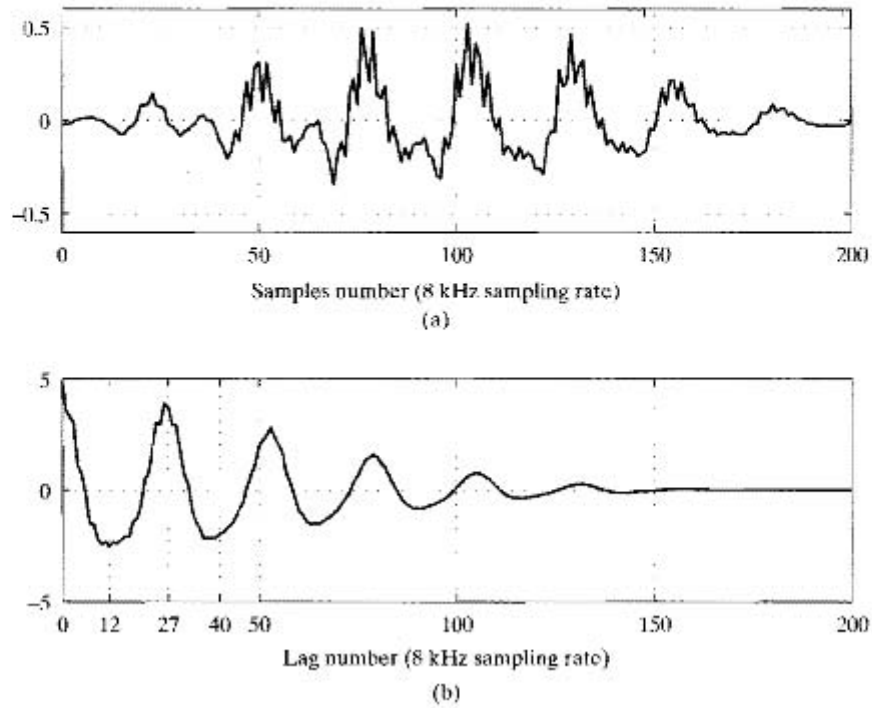


Figure 11.11 (a) Windowed voiced speech waveform. (b) Corresponding autocorrelation function (samples connected by straight lines).

When applying all-pole modeling to voiced speech, it is useful to think of the signal as being deterministic, but with an excitation function that is a periodic train of impulses. This accounts for the periodic nature of the autocorrelation function when several periods of the signal are included in the window as in Figure 11.11(a).

Figure 11.12 shows a comparison of the DTFT of the signal in Figure 11.11(a) with spectra computed from all-pole modeling with two different model orders and using the autocorrelation function in Figure 11.11(b). Note that the DTFT of $s[n]$ shows peaks at multiples of the fundamental frequency $F_0 = 8\,\text{kHz}/27 = 296\,\text{Hz}$, as well as many other less prominent peaks and dips that can be attributed to the windowing effects discussed in Section 10.2.1. If the first 13 samples of $r_{ss}[m]$ in Figure 11.11(b) are used to compute an all-pole model spectrum ($p = 12$), the result is the smooth curve shown with the heavy line in Figure 11.12(a). With the filter order as 12 and the fundamental period of 27 samples, this spectrum estimate in effect ignores the spectral structure owing to the periodicity of the signal and produces a much smoother spectrum estimate. If 41 values of $r_{ss}[m]$ are used, however, we obtain the spectrum plotted with the thin line. Since the period of the signal is 27, a value of $p = 40$ includes the periodicity peak in the autocorrelation function and thus, the all-pole spectrum tends to represent much of the fine detail in the DTFT spectrum. Note that both cases support our assertion above that the all-pole model spectrum estimate tends to favor good representation at the peaks of the DTFT spectrum.

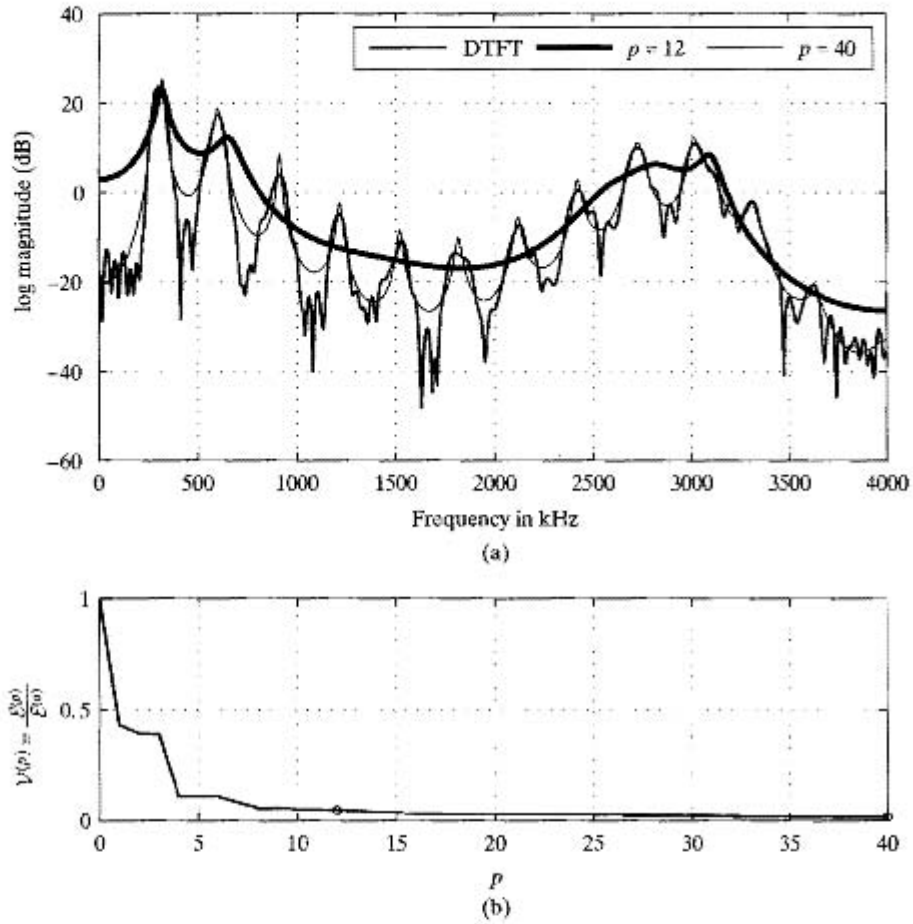


Figure 11.12 (a) Comparison of DTFT and all-pole model spectra for voiced speech segment in Figure 11.11(a). (b) Normalized prediction error as a function of p.

This example illustrates that the choice of the model order p controls the degree of smoothing of the DTFT spectrum. Figure 11.12(b) shows that as p increases, the mean-squared prediction error decreases quickly and then levels off, as in our previous example. Recall that in Sections 11.2.4 and 11.2.5, we argued that the all-pole model with appropriately chosen gain results in a match between the autocorrelation functions of the signal and the all-pole model up to p correlation lags as in Eq. (11.39). This implies that as p increases, the all-pole model spectrum will approach the DTFT spectrum, and when $p \to \infty$, it follows that $r_{hh}[m] = r_{ss}[m]$ for all m, and therefore, $|H(e^{j\omega})|^2 = |S(e^{j\omega})|^2$. However, this does not mean that $H(e^{j\omega}) = S(e^{j\omega})$ because $H(z)$ is an IIR system, and $S(z)$ is the z-transform of a finite-length sequence. Also note that as $p \to \infty$, the averaged prediction error does not approach zero, even though $|H(e^{j\omega})|^2 \to |S(e^{j\omega})|^2$. As we have discussed, this occurs because the total error in Eq. (11.11) is the prediction error $\hat{e}[n]$ minus $Gv[n]$. Said differently, the linear predictor must always predict the first nonzero sample from the zero-valued samples that precede it.

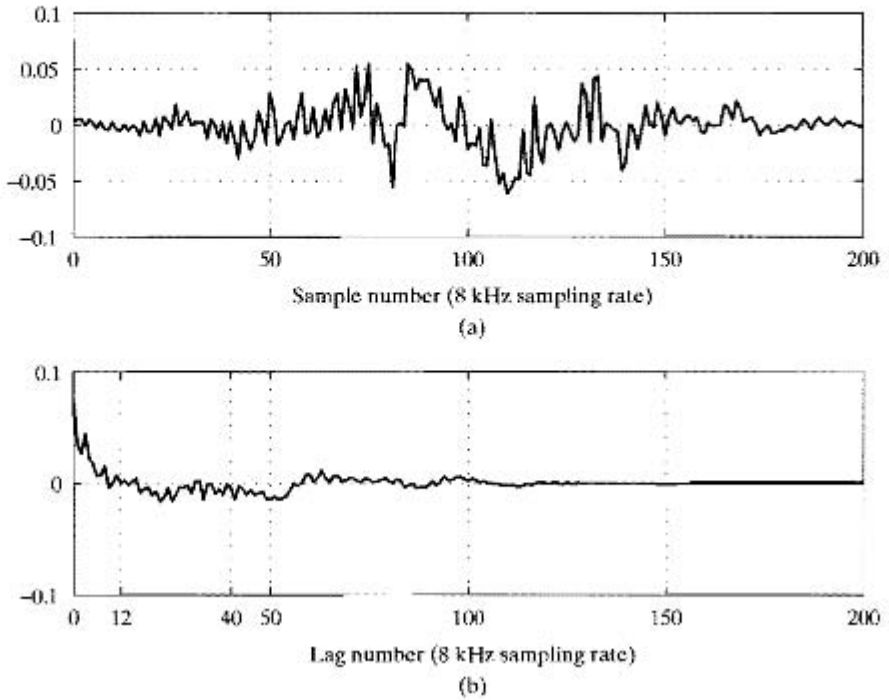


Figure 11.13 (a) Windowed unvoiced speech waveform. (b) Corresponding autocorrelation function (samples connected by straight lines).

The other main class of speech sounds is comprised of the unvoiced sounds such as fricatives. These sounds are produced by creating random turbulent air flow in the vocal tract; therefore, they are best modeled in terms of an all-pole system excited by white noise. Figure 11.13 shows an example of a 201-point Hamming-windowed segment of unvoiced speech and its corresponding autocorrelation function. Note that the autocorrelation function shows no indication of periodicity in either the signal waveform or the autocorrelation function. A comparison of the DTFT of the signal in Figure 11.13(a) with two all-pole model spectra computed from the autocorrelation function in Figure 11.13(b) is shown in Figure 11.14(a). From the point of view of spectrum analysis of random signals, the magnitude-squared of the DTFT is a periodogram. Thus, it contains a component that is randomly varying with frequency. Again, by choice of the model order, the periodogram can be smoothed to any desired degree.

11.5.2 Pole Locations

In speech processing, the poles of the all-pole model have a close relationship to the resonance frequencies of the vocal tract, thus, it is often useful to factor the polynomial $A(z)$ to obtain its zeros for representation as in Eq. (11.50). As discussed in Section 11.3.3, the zeros z_k of the prediction error filter are the poles of the all-pole model system function. It is the poles of the system function that are responsible for the peaks in the spectrum estimates discussed in Section 11.5.1. The closer a pole is to the unit circle, the more peaked is the spectrum for frequencies close to the angle of the pole.

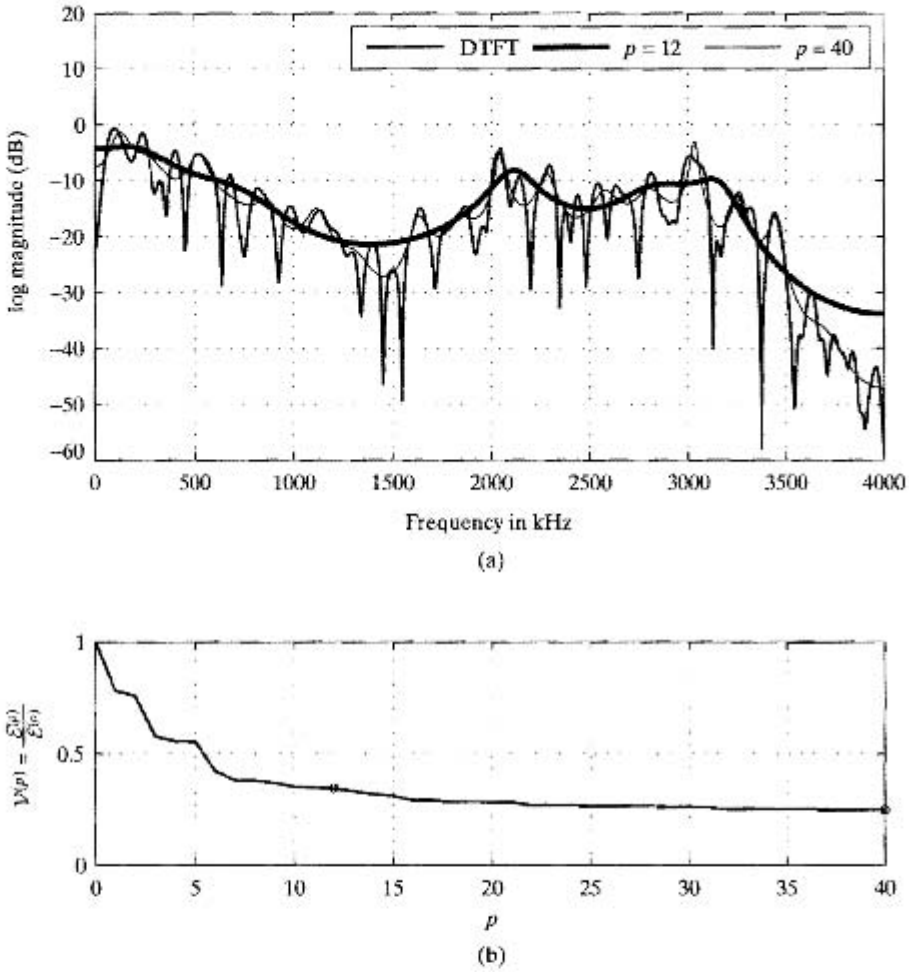


Figure 11.14 (a) Comparison of DTFT and all-pole model spectra for unvoiced speech segment in Figure 11.13(a). (b) Normalized prediction error as a function of p.

Figure 11.15 shows the zeros of the prediction error system function $A(z)$ (poles of the model system) for the two spectrum estimates in Figure 11.12(a). For $p = 12$, the zeros of $A(z)$ are denoted by the open circles. Five complex conjugate pairs of zeros are close to the unit circle, and their manifestations as poles are clearly evident in heavy line curve of Figure 11.12(a). For the case $p = 40$, the zeros of $A(z)$ are denoted by the large filled dots. Observe that most of the zeros are close to the unit circle, and they are more or less evenly distributed around the unit circle. This produces the peaks in the model spectrum that are spaced approximately at multiples of the normalized radian frequency corresponding to the fundamental frequency of the speech signal; i.e., at angles $2\pi(296 \text{ Hz})/8 \text{ kHz}$.

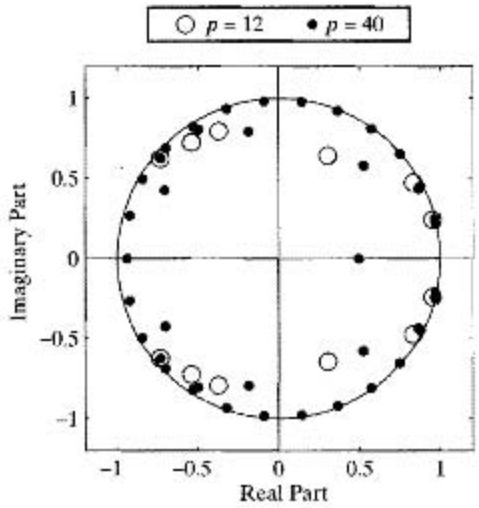


Figure 11.15 Zeros of prediction error filters (poles of model systems) used to obtain the spectrum estimates in Figure 11.12.

11.5.3 All-Pole Modeling of Sinusoidal Signals

As another important example, we consider the use of the poles of an all-pole model to estimate frequencies of sinusoidal signals. To see why this is possible, consider the sum of two sinusoids

$$s[n] = [A_1 \cos(\omega_1 n + \theta_1) + A_2 \cos(\omega_2 n + \theta_2)]\, u[n]. \tag{11.61}$$

The z-transform of $s[n]$ has the form

$$S(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3}}{(1 - e^{j\omega_1} z^{-1})(1 - e^{-j\omega_1} z^{-1})(1 - e^{j\omega_2} z^{-1})(1 - e^{-j\omega_2} z^{-1})}. \tag{11.62}$$

That is, the sum of two sinusoids can be represented as the impulse response of an LTI system whose system function has both poles and zeros. The numerator polynomial would be a somewhat complicated function of the amplitudes, frequencies, and phase shifts. What is important for our discussion is that the numerator is a 3rd-order polynomial and the denominator is a 4th-order polynomial, the roots of which are all on the unit circle at angles equal to $\pm\omega_1$ and $\pm\omega_2$. The difference equation describing this system with impulse excitation has the form

$$s[n] - \sum_{k=1}^{4} a_k s[n-k] = \sum_{k=0}^{3} b_k \delta[n-k] \tag{11.63}$$

where the coefficients a_k would result from multiplying the denominator factors. Note that

$$s[n] - \sum_{k=1}^{4} a_k s[n-k] = 0 \quad \text{for } n \geq 4, \tag{11.64}$$

which suggests that the signal $s[n]$ can be predicted with no error by a 4th-order predictor except at the very beginning ($0 \leq n \leq 3$). The coefficients of the denominator can be

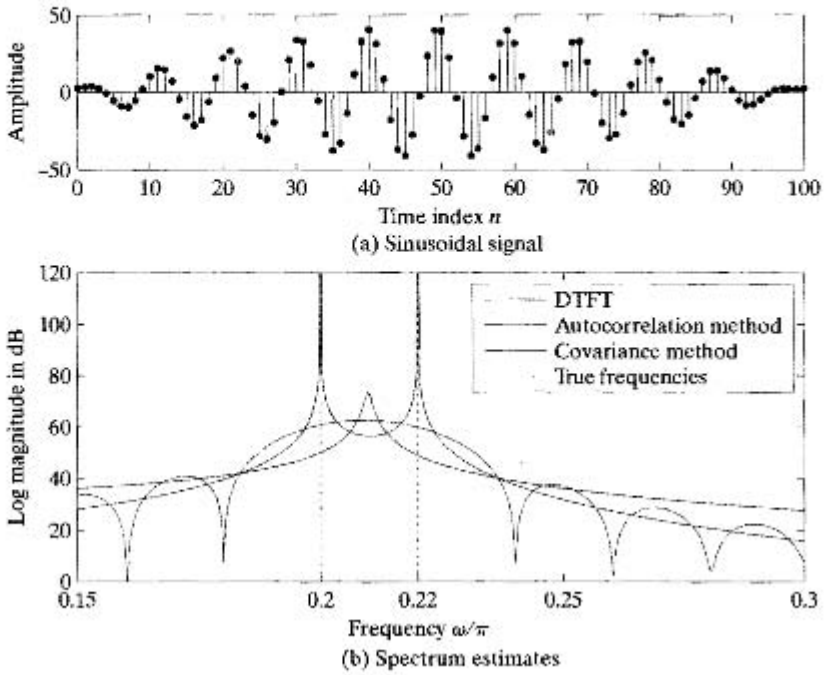


Figure 11.16 Spectrum estimation for a sinusoidal signal.

estimated from the signal by applying the covariance method to a short segment of the signal selected so as not to include the first four samples. In the ideal case for which Eq. (11.61) accurately represents the signal (e.g., high SNR), the roots of the resulting polynomial provide good estimates of the frequencies of the component sinusoids.

Figure 11.16(a) shows a plot of 101 samples of the signal[8]

$$s[n] = 20\cos(0.2\pi n - 0.1\pi) + 22\cos(0.22\pi n + 0.9\pi). \tag{11.65}$$

Because the two frequencies are close together, it is necessary to use a large number of samples to resolve the two frequencies by Fourier analysis. However, since the signal fits the all-pole model perfectly, the covariance method can be used to obtain very accurate estimates of the frequencies from very short segments of the signal. This is illustrated in Figure 11.16(b).

The DTFT of the 101 samples (with rectangular window) shows no indication that there are two distinct sinusoid frequencies around $\omega = 0.21\pi$. Recall that the main lobe width for an $(M+1)$-point rectangular window is $\Delta\omega = 4\pi/(M+1)$. Consequently, a 101-point rectangular window can clearly resolve two frequencies only if they are no closer than about $.04\pi$ rad/s. Correspondingly, the DTFT does not show two spectral peaks.

Similarly, use of the autocorrelation method results in the spectrum estimate shown by the heavy line. This estimate also contains only one spectral peak. The predic-

[8]The tapering of the segment of the signal in Figure 11.16(a) is not a result of windowing. It is caused by the "beating" of the two cosines of nearly the same frequency. The period of the beat frequency (difference between 0.22π and 0.2π) is 100 samples.

tion error polynomial (in factored form) obtained with the autocorrelation method is

$$A_e(z) = (1 - 0.998e^{j0.21\pi}z^{-1})(1 - 0.998e^{-j0.21\pi}z^{-1}) \cdot (1 - 0.426z^{-1})(1 - 0.1165z^{-1}) \tag{11.66}$$

The two real poles contribute no peaks, and the complex poles are close to the unit circle, but at $\pm 0.21\pi$, which is halfway between the two frequencies. Thus, the windowing inherent in the autocorrelation method causes the resulting model to lock onto the average frequency 0.21π.

On the other hand, the factored prediction error polynomial obtained with the covariance method is (with rounding of the magnitudes and angles) given by

$$A_e(z) = (1 - e^{j0.2\pi}z^{-1})(1 - e^{-j0.2\pi}z^{-1}) \cdot (1 - e^{j0.22\pi}z^{-1})(1 - e^{-j0.22\pi}z^{-1}). \tag{11.67}$$

In this case, the angles of the zeros are almost exactly equal to the frequencies of the two sinusoids. Also shown in Figure 11.16(b) is the frequency response of the model, i.e.,

$$|H_{\text{cov}}(e^{j\omega})|^2 = \frac{1}{|A_{\text{cov}}(e^{j\omega})|^2}. \tag{11.68}$$

plotted in dB. In this case, the prediction error is very close to zero, which, if used to estimate the gain of the all-pole model, would lead to an indeterminant estimate. Therefore, the gain is arbitrarily set to one, which leads to a plot of Eq. (11.68) on a similar scale to the other estimates. Since the poles are almost exactly on the unit circle, the magnitude spectrum becomes exceedingly large at the pole frequencies. Note that the roots of the prediction error polynomial give an accurate estimate of the frequencies. This method, of course, does not provide accurate information about the amplitudes and phases of the sinusoidal components.

11.6 SOLUTION OF THE AUTOCORRELATION NORMAL EQUATIONS

In both the autocorrelation and covariance methods of computing the correlation values, the predictor coefficients that minimize the mean-squared inverse filter error and equivalently the prediction error satisfy a set of linear equations of the general form:

$$\begin{bmatrix} \phi_{ss}[1,1] & \phi_{ss}[1,2] & \phi_{ss}[1,3] & \cdots & \phi_{ss}[1,p] \\ \phi_{ss}[2,1] & \phi_{ss}[2,2] & \phi_{ss}[2,3] & \cdots & \phi_{ss}[2,p] \\ \phi_{ss}[3,1] & \phi_{ss}[3,2] & \phi_{ss}[3,3] & \cdots & \phi_{ss}[3,p] \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \phi_{ss}[p,1] & \phi_{ss}[p,2] & \phi_{ss}[p,3] & \cdots & \phi_{ss}[p,p] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \phi_{ss}[1,0] \\ \phi_{ss}[2,0] \\ \phi_{ss}[3,0] \\ \vdots \\ \phi_{ss}[p,0] \end{bmatrix}. \tag{11.69}$$

In matrix notation, these linear equations have the representation

$$\boldsymbol{\Phi a} = \boldsymbol{\psi}. \tag{11.70}$$

Since $\phi[i, k] = \phi[k, i]$, in both the autocorrelation and covariance methods, the matrix $\boldsymbol{\Phi}$ is symmetric, and, because it arises in a least-squares problem, it is also positive-definite, which guarantees that it is invertible. In general, this leads to efficient solution methods, such as the Cholesky decomposition (see Press, et al., 2007), that are based on matrix factorization and applicable when $\boldsymbol{\Phi}$ is symmetric and positive definite. However, in the specific case of the autocorrelation method or any method for which $\phi_{ss}[i, k] = r_{ss}[|i - k|]$, Eqs. (11.69) become the autocorrelation normal equations (also referred to as the Yule–Walker equations).

$$\begin{bmatrix} r_{ss}[0] & r_{ss}[1] & r_{ss}[2] & \cdots & r_{ss}[p-1] \\ r_{ss}[1] & r_{ss}[0] & r_{ss}[1] & \cdots & r_{ss}[p-2] \\ r_{ss}[2] & r_{ss}[1] & r_{ss}[0] & \cdots & r_{ss}[p-3] \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ r_{ss}[p-1] & r_{ss}[p-2] & r_{ss}[p-3] & \cdots & r_{ss}[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r_{ss}[1] \\ r_{ss}[2] \\ r_{ss}[3] \\ \vdots \\ r_{ss}[p] \end{bmatrix}. \tag{11.71}$$

In this case, in addition to the matrix $\boldsymbol{\Phi}$ being symmetric and positive-definite, it is also a Toeplitz matrix, i.e., all the elements on each subdiagonal are equal. This property leads to an efficient algorithm, referred to as the Levinson–Durbin recursion, for solving the equations.

11.6.1 The Levinson–Durbin Recursion

The Levinson–Durbin algorithm for computing the predictor coefficients that minimize the total-squared prediction error results from the high degree of symmetry in the matrix $\boldsymbol{\Phi}$ and furthermore, as Eq. (11.71) confirms, the elements of the right-hand side vector $\boldsymbol{\psi}$ are primarily the same values that populate the matrix $\boldsymbol{\Phi}$. Equations (L–D.1) to (L–D.6) in Figure 11.17 define the computations. A derivation of these equations is given in Section 11.6.2, but before developing the details of the derivation, it is helpful to simply examine the steps of the algorithm.

(L–D.1) This step initializes the mean-squared prediction error to be the energy of the signal. That is, a zero$^{\text{th}}$-order predictor (no predictor) yields no reduction in prediction error energy, since the prediction error $e[n]$ is identical to the signal $s[n]$.

The next line in Figure 11.17 states that steps (L–D.2) through (L–D.5) are repeated p times, with each repetition of those steps increasing the order of the predictor by one. In other words, the algorithm computes a predictor of order i from the predictor of order $i - 1$ starting with $i - 1 = 0$.

(L–D.2) This step computes a quantity k_i. The sequence of parameters $k_i, i = 1, 2, \ldots, p$ which we refer to as the k-parameters, plays a key role in generating the next set of predictor coefficients.[9]

[9]For reasons to be discussed in Section 11.7, the k-parameters are also called *PARCOR* (for PARtial CORrelation) *coefficients* or also, *reflection coefficients*.

Levinson–Durbin Algorithm

$$\mathcal{E}^{(0)} = r_{ss}[0] \qquad \text{(L–D.1)}$$

for $i = 1, 2, \ldots, p$

$$k_i = \left(r_{ss}[i] - \sum_{j=1}^{i-1} a_j^{(i-1)} r_{ss}[i-j] \right) / \mathcal{E}^{(i-1)} \qquad \text{(L–D.2)}$$

$$a_i^{(i)} = k_i \qquad \text{(L–D.3)}$$

if $i > 1$ then for $j = 1, 2, \ldots, i-1$

$$a_j^{(i)} = a_j^{(i-1)} - k_i a_{i-j}^{(i-1)} \qquad \text{(L–D.4)}$$

end

$$\mathcal{E}^{(i)} = (1 - k_i^2)\mathcal{E}^{(i-1)} \qquad \text{(L–D.5)}$$

end

$$a_j = a_j^{(p)} \quad j = 1, 2, \ldots, p \qquad \text{(L–D.6)}$$

Figure 11.17 Equations defining the Levinson–Durbin algorithm.

(L–D.3) This equation states that $a_i^{(i)}$, the i^{th} coefficient of the i^{th}-order predictor, is equal to k_i.

(L–D.4) In this equation, k_i is used to compute the remaining coefficients of the i^{th}-order predictor as a combination of the coefficients of the predictor of order $(i-1)$ with those same coefficients in reverse order.

(L–D.5) This equation updates the prediction error for the i^{th}-order predictor.

(L–D.6) This is the final step where the p^{th}-order predictor is defined to be the result after p iterations of the algorithm.

The Levinson–Durbin algorithm is valuable because it is an efficient method of solution of the autocorrelation normal equations and also for the insight that it provides about the properties of linear prediction and all-pole models. For example, from Eq. (L–D.5), it can be shown that the averaged prediction error for a p^{th}-order predictor is the product of the prediction errors for all lower-order predictors, from which it follows that $0 < \mathcal{E}^{(i)} \leq \mathcal{E}^{(i-1)} < \mathcal{E}^{(p)}$ and

$$\mathcal{E}^{(p)} = \mathcal{E}^{(0)} \prod_{i=1}^{p} (1 - k_i^2) = r_{ss}[0] \prod_{i=1}^{p} (1 - k_i^2). \qquad (11.72)$$

Since $\mathcal{E}^{(i)} > 0$, it must be true that $-1 < k_i < 1$ for $i = 1, 2, \ldots, p$. That is, the k-parameters are strictly less than one in magnitude.

11.6.2 Derivation of the Levinson–Durbin Algorithm

From Eq. (11.30), the optimum predictor coefficients satisfy the set of equations

$$r_{ss}[i] - \sum_{k=1}^{p} a_k r_{ss}[i-k] = 0 \qquad i = 1, 2, \ldots, p, \qquad (11.73a)$$

and the minimum mean-squared prediction error is given by

$$r_{ss}[0] - \sum_{k=1}^{p} a_k r_{ss}[k] = \mathcal{E}^{(p)}. \tag{11.73b}$$

Since Eq. (11.73b) contains the same correlation values as in Eq. (11.73a), it is possible to take them together and write a new set of $p + 1$ equations that are satisfied by the p unknown predictor coefficients and the corresponding unknown mean-squared prediction error $\mathcal{E}^{(p)}$. These equations have the matrix form

$$\begin{bmatrix} r_{ss}[0] & r_{ss}[1] & r_{ss}[2] & \cdots & r_{ss}[p] \\ r_{ss}[1] & r_{ss}[0] & r_{ss}[1] & \cdots & r_{ss}[p-1] \\ r_{ss}[2] & r_{ss}[1] & r_{ss}[0] & \cdots & r_{ss}[p-2] \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ r_{ss}[p] & r_{ss}[p-1] & r_{ss}[p-2] & \cdots & r_{ss}[0] \end{bmatrix} \begin{bmatrix} 1 \\ -a_1^{(p)} \\ -a_2^{(p)} \\ \vdots \\ -a_p^{(p)} \end{bmatrix} = \begin{bmatrix} \mathcal{E}^{(p)} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{11.74}$$

It is this set of equations that can be solved recursively by the Levinson–Durbin algorithm. This is done by successively incorporating a new correlation value at each iteration and solving for the next higher-order predictor in terms of the new correlation value and the previously found predictor.

For any order i, the set of equations in Eq. (11.74) can be represented in matrix notation as

$$\mathbf{R}^{(i)}\mathbf{a}^{(i)} = \mathbf{e}^{(i)}. \tag{11.75}$$

We wish to show how the i^{th} solution can be derived from the $(i-1)^{\text{st}}$ solution. In other words, given $\mathbf{a}^{(i-1)}$, the solution to $\mathbf{R}^{(i-1)}\mathbf{a}^{(i-1)} = \mathbf{e}^{(i-1)}$, we wish to derive the solution to $\mathbf{R}^{(i)}\mathbf{a}^{(i)} = \mathbf{e}^{(i)}$.

First, write the equations $\mathbf{R}^{(i-1)}\mathbf{a}^{(i-1)} = \mathbf{e}^{(i-1)}$ in expanded form as

$$\begin{bmatrix} r_{ss}[0] & r_{ss}[1] & r_{ss}[2] & \cdots & r_{ss}[i-1] \\ r_{ss}[1] & r_{ss}[0] & r_{ss}[1] & \cdots & r_{ss}[i-2] \\ r_{ss}[2] & r_{ss}[1] & r_{ss}[0] & \cdots & r_{ss}[i-3] \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ r_{ss}[i-1] & r_{ss}[i-2] & r_{ss}[i-3] & \cdots & r_{ss}[0] \end{bmatrix} \begin{bmatrix} 1 \\ -a_1^{(i-1)} \\ -a_2^{(i-1)} \\ \vdots \\ -a_{i-1}^{(i-1)} \end{bmatrix} = \begin{bmatrix} \mathcal{E}^{(i-1)} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{11.76}$$

Then append a 0 to the vector $\mathbf{a}^{(i-1)}$ and multiply by the matrix $\mathbf{R}^{(i)}$ to obtain

$$\begin{bmatrix} r_{ss}[0] & r_{ss}[1] & r_{ss}[2] & \cdots & r_{ss}[i] \\ r_{ss}[1] & r_{ss}[0] & r_{ss}[1] & \cdots & r_{ss}[i-1] \\ r_{ss}[2] & r_{ss}[1] & r_{ss}[0] & \cdots & r_{ss}[i-2] \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ r_{ss}[i-1] & r_{ss}[i-2] & r_{ss}[i-3] & \cdots & r_{ss}[1] \\ r_{ss}[i] & r_{ss}[i-1] & r_{ss}[i-2] & \cdots & r_{ss}[0] \end{bmatrix} \begin{bmatrix} 1 \\ -a_1^{(i-1)} \\ -a_2^{(i-1)} \\ \vdots \\ -a_{i-1}^{(i-1)} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathcal{E}^{(i-1)} \\ 0 \\ 0 \\ \vdots \\ 0 \\ \gamma^{(i-1)} \end{bmatrix}. \tag{11.77}$$

where, to satisfy Eq. (11.77),

$$\gamma^{(i-1)} = r_{ss}[i] - \sum_{j=1}^{i-1} a_j^{(i-1)} r_{ss}[i-j]. \tag{11.78}$$

It is in Eq. (11.78) that the new autocorrelation value $r_{ss}[i]$ is introduced. However, Eq. (11.77) is not yet in the desired form $\mathbf{R}^{(i)}\mathbf{a}^{(i)} = \mathbf{e}^{(i)}$. The key step in the derivation is to recognize that due to the special symmetry of the Toeplitz matrix $\mathbf{R}^{(i)}$, the equations can be written in reverse order (first equation last and last equation first, and so on) and the matrix for the resulting set of equations is still $\mathbf{R}^{(i)}$; i.e.,

$$\begin{bmatrix} r_{ss}[0] & r_{ss}[1] & r_{ss}[2] & \cdots & r_{ss}[i] \\ r_{ss}[1] & r_{ss}[0] & r_{ss}[1] & \cdots & r_{ss}[i-1] \\ r_{ss}[2] & r_{ss}[1] & r_{ss}[0] & \cdots & r_{ss}[i-2] \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ r_{ss}[i-1] & r_{ss}[i-2] & r_{ss}[i-3] & \cdots & r_{ss}[1] \\ r_{ss}[i] & r_{ss}[i-1] & r_{ss}[i-2] & \cdots & r_{ss}[0] \end{bmatrix} \begin{bmatrix} 0 \\ -a_{i-1}^{(i-1)} \\ -a_{i-2}^{(i-1)} \\ \vdots \\ -a_1^{(i-1)} \\ 1 \end{bmatrix} = \begin{bmatrix} \gamma^{(i-1)} \\ 0 \\ 0 \\ \vdots \\ 0 \\ \mathcal{E}^{(i-1)} \end{bmatrix}. \tag{11.79}$$

Now Eq. (11.77) is combined with Eq. (11.79) according to

$$\mathbf{R}^{(i)}\left[\begin{bmatrix} 1 \\ -a_1^{(i-1)} \\ -a_2^{(i-1)} \\ \vdots \\ -a_{i-1}^{(i-1)} \\ 0 \end{bmatrix} - k_i \begin{bmatrix} 0 \\ -a_{i-1}^{(i-1)} \\ -a_{i-2}^{(i-1)} \\ \vdots \\ -a_1^{(i-1)} \\ 1 \end{bmatrix}\right] = \left[\begin{bmatrix} \mathcal{E}^{(i-1)} \\ 0 \\ 0 \\ \vdots \\ 0 \\ \gamma^{(i-1)} \end{bmatrix} - k_i \begin{bmatrix} \gamma^{(i-1)} \\ 0 \\ 0 \\ \vdots \\ 0 \\ \mathcal{E}^{(i-1)} \end{bmatrix}\right]. \tag{11.80}$$

Equation (11.80) is now approaching the desired form $\mathbf{R}^{(i)}\mathbf{a}^{(i)} = \mathbf{e}^{(i)}$. All that remains is to choose $\gamma^{(i-1)}$, so that the right hand vector has only a single nonzero entry. This requires that

$$k_i = \frac{\gamma^{(i-1)}}{\mathcal{E}^{(i-1)}} = \frac{r_{ss}[i] - \sum_{j=1}^{i-1} a_j^{(i-1)} r_{ss}[i-j]}{\mathcal{E}^{(i-1)}}, \tag{11.81}$$

which ensures cancelation of the last element of the right hand side vector, and causes the first element to be

$$\mathcal{E}^{(i)} = \mathcal{E}^{(i-1)} - k_i \gamma^{(i-1)} = \mathcal{E}^{(i-1)}(1 - k_i^2). \tag{11.82}$$

With this choice of $\gamma^{(i-1)}$, it follows that the vector of i^{th}-order prediction coefficients is

$$\begin{bmatrix} 1 \\ -a_1^{(i)} \\ -a_2^{(i)} \\ \vdots \\ -a_{i-1}^{(i)} \\ -a_i^{(i)} \end{bmatrix} = \begin{bmatrix} 1 \\ -a_1^{(i-1)} \\ -a_2^{(i-1)} \\ \vdots \\ -a_{i-1}^{(i-1)} \\ 0 \end{bmatrix} - k_i \begin{bmatrix} 0 \\ -a_{i-1}^{(i-1)} \\ -a_{i-2}^{(i-1)} \\ \vdots \\ -a_1^{(i-1)} \\ 1 \end{bmatrix} \tag{11.83}$$

From Eq. (11.83), we can write the set of equations for updating the coefficients as

$$a_j^{(i)} = a_j^{(i-1)} - k_i a_{i-j}^{(i-1)} \qquad j = 1, 2, \ldots, i-1, \tag{11.84a}$$

and

$$a_i^{(i)} = k_i. \tag{11.84b}$$

Equations (11.81), (11.84b), (11.84a), and (11.82) are the key equations of the Levinson–Durbin algorithm. They correspond to Eqs. (L–D.2), (L–D.3), (L–D.4), and (L–D.5) in Figure 11.17, which shows how they are used order-recursively to compute the optimum prediction coefficients as well as the corresponding mean-squared prediction errors and coefficients k_i for all linear predictors up to order p.

11.7 LATTICE FILTERS

Among the many interesting and useful concepts that emerge from the Levinson–Durbin algorithm is its interpretation in terms of the lattice structures introduced in Section 6.6. There, we showed that any FIR filter with system function of the form

$$A(z) = 1 - \sum_{k=1}^{M} \alpha_k z^{-k} \tag{11.85}$$

can be implemented by a lattice structure as depicted in Figure 6.37. Furthermore, we showed that the coefficients of the FIR system function are related to the k-parameters of a corresponding lattice filter by a recursion given in Figure 6.38, which is repeated for convenience in the bottom half of Figure 11.18. By reversing the steps in the k-to-α algorithm, we obtained an algorithm given in Figure 6.39 for computing the k-parameters from the coefficients $\alpha_j, j = 1, 2, \ldots, M$. Thus, there is a unique relationship between the coefficients of the direct form representation and the lattice representation of an FIR filter.

In this chapter, we have shown that a p^{th}-order prediction error filter is an FIR filter with system function

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

whose coefficients can be computed from the autocorrelation function of a signal through a process that we have called the Levinson–Durbin algorithm. A by-product of the Levinson–Durbin computation is a set of parameters that we have also denoted k_i and called the k-parameters. A comparison of the two algorithms in Figure 11.18 shows that their steps are identical except for one important detail. In the algorithm derived in Chapter 6, we started with the lattice filter with known coefficients k_i and derived the recursion for obtaining the coefficients of the corresponding direct form FIR filter. In the Levinson–Durbin algorithm, we begin with the autocorrelation function of a signal and compute the k-parameters recursively as an intermediate result in computing the coefficients of the FIR prediction error filter. Since both algorithms give a unique result after p iterations, and since there is a unique relationship between the k-parameters and the coefficients of an FIR filter, it follows that if $M = p$ and $a_j = \alpha_j$ for $j = 1, 2, \ldots, p$, the k-parameters produced by the Levinson–Durbin algorithm must be the k-parameters of a lattice filter implementation of the FIR prediction error filter $A^{(p)}(z)$.

Levinson–Durbin Algorithm

$\mathcal{E}^{(0)} = r_{ss}[0]$
for $i = 1, 2, \ldots, p$
$$k_i = \left(r_{ss}[i] - \sum_{j=1}^{i-1} a_j^{(i-1)} r_{ss}[i-j] \right) / \mathcal{E}^{(i-1)} \quad \text{Eq. (11.81)}$$
$a_i^{(i)} = k_i$ Eq. (11.84b)
if $i > 1$ then for $j = 1, 2, \ldots, i-1$
$a_j^{(i)} = a_j^{(i-1)} - k_i a_{i-j}^{(i-1)}$ Eq. (11.84a)
end
$\mathcal{E}^{(i)} = (1 - k_i^2)\mathcal{E}^{(i-1)}$ Eq. (11.82)
end
$a_j = a_j^{(p)} \quad j = 1, 2, \ldots, p$

Lattice k-to-α Algorithm

Given $k_1, k_2, \ldots, k_M$
for $i = 1, 2, \ldots, M$
$\alpha_i^{(i)} = k_i$ Eq. (6.66b)
if $i > 1$ then for $j = 1, 2, \ldots, i-1$
$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}$ Eq. (6.66a)
end
end
$\alpha_j = \alpha_j^{(M)} \quad j = 1, 2, \ldots, M$ Eq. (6.68b)

Figure 11.18 Comparison of the Levinson–Durbin algorithm and the algorithm for converting from k-parameters of a lattice structure to the FIR impulse response coefficients in Eq. (11.85).

11.7.1 Prediction Error Lattice Network

To explore the lattice filter interpretation further, suppose that we have an i^{th}-order prediction error system function

$$A^{(i)}(z) = 1 - \sum_{k=1}^{i} a_k^{(i)} z^{-k}. \tag{11.86}$$

The z-transform representation of the prediction error[10] would be

$$E^{(i)}(z) = A^{(i)}(z)S(z), \tag{11.87}$$

[10]The z-transform equations are used assuming that the z-transforms of $e[n]$ and $s[n]$ exist. Although this would not be true for random signals, the relationships between the variables remain in effect for the system. The z-transform notation facilitates the development of these relationships.

and the time-domain difference equation for this FIR filter is

$$e^{(i)}[n] = s[n] - \sum_{k=1}^{i} a_k^{(i)} s[n-k]. \tag{11.88}$$

The sequence $e^{(i)}[n]$ is given the more specific name *forward prediction error* because it is the error in predicting $s[n]$ from i *previous* samples.

The source of the lattice filter interpretation is Eqs. (11.84a) and (11.84b), which, if substituted into Eq. (11.86), yield the following relation between $A^{(i)}(z)$ and $A^{(i-1)}(z)$:

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}). \tag{11.89}$$

This is not a surprising result if we consider the matrix representation of the polynomial $A^{(i)}(z)$ in Eq. (11.83).[11] Now, if Eq. (11.89) is substituted for $A^{(i)}(z)$ in Eq. (11.87), the result is

$$E^{(i)}(z) = A^{(i-1)}(z)S(z) - k_i z^{-i} A^{(i-1)}(z^{-1})S(z). \tag{11.90}$$

The first term in Eq. (11.90) is $E^{(i-1)}(z)$, i.e., the prediction error for an $(i-1)^{\text{st}}$-order filter. The second term has a similar interpretation, if we define

$$\tilde{E}^{(i)}(z) = z^{-i} A^{(i)}(z^{-1})S(z) = B^{(i)}(z)S(z), \tag{11.91}$$

where we have defined $B^{(i)}(z)$ as

$$B^{(i)}(z) = z^{-i} A^{(i)}(z^{-1}) \tag{11.92}$$

The time-domain interpretation of Eq. (11.91) is

$$\tilde{e}^{(i)}[n] = s[n-i] - \sum_{k=1}^{i} a_k^{(i)} s[n-i+k]. \tag{11.93}$$

The sequence $\tilde{e}^{(i)}[n]$ is called the *backward prediction error*, since Eq. (11.93) suggests that $s[n-i]$ is "predicted" (using coefficients $a_k^{(i)}$) from the i samples that *follow* sample $n-i$.

With these definitions, it follows from Eq. (11.90) that

$$E^{(i)}(z) = E^{(i-1)}(z) - k_i z^{-1} \tilde{E}^{(i-1)}(z). \tag{11.94}$$

and hence,

$$e^{(i)}[n] = e^{(i-1)}[n] - k_i \tilde{e}^{(i-1)}[n-1]. \tag{11.95}$$

By substituting Eq. (11.89) into Eq. (11.91), we obtain

$$\tilde{E}^{(i)}(z) = z^{-1} \tilde{E}^{(i-1)}(z) - k_i E^{(i-1)}(z), \tag{11.96}$$

which, in the time domain, corresponds to

$$\tilde{e}^{(i)}[n] = \tilde{e}^{(i-1)}[n-1] - k_i e^{(i-1)}[n]. \tag{11.97}$$

[11] The algebraic manipulations to derive this result are suggested as an exercise in Problem 11.21.

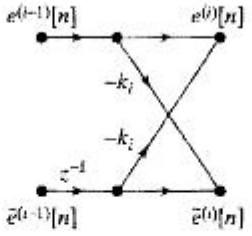


Figure 11.19 Signal flow graph of prediction error computation.

The difference equations in Eq. (11.95) and Eq. (11.97) express the i^{th}-order forward and backward prediction errors in terms of k_i and the $(i-1)^{\text{st}}$-order forward and backward prediction errors. This pair of difference equations is represented by the flow graph of Figure 11.19. Therefore, Figure 11.19 represents a pair of difference equations that embody one iteration of the Levinson–Durbin recursion. As in the Levinson–Durbin recursion, we start with a zero$^{\text{th}}$-order predictor for which

$$e^{(0)}[n] = \tilde{e}^{(0)}[n] = s[n]. \tag{11.98}$$

With $e^{(0)}[n] = s[n]$ and $\tilde{e}^{(0)}[n] = s[n]$ as inputs to a first stage as depicted in Figure 11.19 with k_1 as coefficient, we obtain $e^{(1)}[n]$ and $\tilde{e}^{(1)}[n]$ as outputs. These are the required inputs for stage 2. We can use p successive stages of the structure in Figure 11.19 to build up a system whose output will be the desired p^{th}-order prediction error signal $e[n] = e^{(p)}[n]$. Such a system, as depicted in Figure 11.20, is identical to the lattice network in Figure 6.37 of Section 6.6.[12] In summary, Figure 11.20 is a signal flow graph representation of the equations

$$e^{(0)}[n] = \tilde{e}^{(0)}[n] = s[n] \tag{11.99a}$$

$$e^{(i)}[n] = e^{(i-1)}[n] - k_i\tilde{e}^{(i-1)}[n-1] \quad i = 1, 2, \ldots, p \tag{11.99b}$$

$$\tilde{e}^{(i)}[n] = \tilde{e}^{(i-1)}[n-1] - k_i e^{(i-1)}[n] \quad i = 1, 2, \ldots, p \tag{11.99c}$$

$$e[n] = e^{(p)}[n], \tag{11.99d}$$

where, if the coefficients k_i are determined by the Levinson–Durbin recursion, the variables $e^{(i)}[n]$ and $\tilde{e}^{(i)}[n]$ are the forward and backward prediction errors for the i^{th}-order optimum predictor.

11.7.2 All-Pole Model Lattice Network

In Section 6.6.2, we showed that the lattice network of Figure 6.42 is an implementation of the all-pole system function $H(z) = 1/A(z)$, where $A(z)$ is the system function of an FIR system; i.e., $H(z)$ is the exact inverse of $A(z)$, and in the present context, it is the system function of the all-pole model with $G = 1$. In this section, we review the all-pole lattice structure in terms of the notation of forward and backward prediction error.

[12]Note that in Figure 6.37 the node variables were denoted $a^{(i)}[n]$ and $b^{(i)}[n]$ instead of $e^{(i)}[n]$ and $\tilde{e}^{(i)}[n]$, respectively.

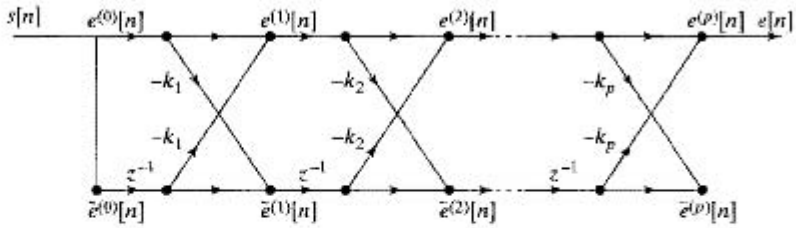


Figure 11.20 Signal flow graph of lattice network implementation of p^{th}-order prediction error computation.

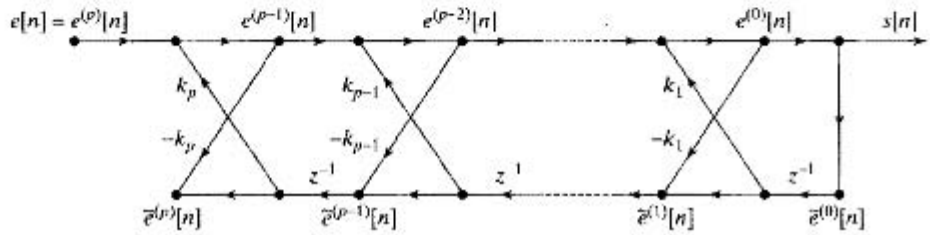


Figure 11.21 All-pole lattice system.

If we replace the node variable labels $a^{(i)}[n]$ and $b^{(i)}[n]$ in Figure 6.42 with the corresponding $e^{(i)}[n]$ and $\bar{e}^{(i)}[n]$ we obtain the flow graph of Figure 11.21, which represents the set of equations

$$e^{(p)}[n] = e[n] \tag{11.100a}$$

$$e^{(i-1)}[n] = e^{(i)}[n] + k_i \bar{e}^{(i-1)}[n-1] \quad i = p, p-1, \ldots, 1 \tag{11.100b}$$

$$\bar{e}^{(i)}[n] = \bar{e}^{(i-1)}[n-1] - k_i e^{(i-1)}[n] \quad i = p, p-1, \ldots, 1 \tag{11.100c}$$

$$s[n] = e^{(0)}[n] = \bar{e}^{(0)}[n]. \tag{11.100d}$$

As we discussed in Section 6.6.2, any stable all-pole system can be implemented by a lattice structure such as Figure 11.21. For such systems, the guarantee of stability inherent in the condition $|k_i| < 1$ is particularly important. Even though the lattice structure requires twice the number of multiplications per output sample as the direct form, it may be the preferred implementation when coefficients must be coarsely quantized. The frequency response of the direct form is exceedingly sensitive to quantization of the coefficients. Furthermore, we have seen that high-order direct form IIR systems can become unstable owing to quantization of their coefficients. This is not the case for the lattice network, as long as the condition $|k_i| < 1$ is maintained for the quantized k-parameters. Furthermore, the frequency response of the lattice network is relatively insensitive to quantization of the k-parameters.

11.7.3 Direct Computation of the *k*-Parameters

The structure of the flow graph in Figure 11.20 is a direct consequence of the Levinson–Durbin recursion, and the parameters k_i, $i = 1, 2, \ldots, p$ can be obtained from the autocorrelation values $r_{ss}[m]$, $m = 0, 1, \ldots, p$ through iterations of the algorithm of Figure 11.17. From our discussion so far, the k_i parameters have been an ancillary consequence of computing the predictor parameters. However, Itakura and Saito (1968, 1970), showed that the k_i parameters can be computed directly from the forward and backward prediction errors in Figure 11.20. And because of the iterative structure as a cascade of the stages in Figure 11.19, the k_i parameters can be computed sequentially from signals available from previous stages of the lattice. The direct computation of the parameter k_i is achieved with the following equation:

$$k_i^P = \frac{\sum_{n=-\infty}^{\infty} e^{(i-1)}[n]\bar{e}^{(i-1)}[n-1]}{\left\{\sum_{n=-\infty}^{\infty}(e^{(i-1)}[n])^2 \sum_{n=-\infty}^{\infty}(\bar{e}^{(i-1)}[n-1])^2\right\}^{1/2}}. \tag{11.101}$$

Observe that Eq. (11.101) is in the form of the energy-normalized cross-correlation between the forward and backward prediction errors at the output of the i^{th} stage. For this reason k_i^P computed using Eq. (11.101) is called a PARCOR coefficient, or more precisely *PARtial CORrelation coefficient*. Figure 11.20 has the interpretation that the correlation in $s[n]$ represented by the autocorrelation function $r_{ss}[m]$ is removed step-by-step by the lattice filter. For a more detailed discussion of the concept of partial correlation, see Stoica and Moses (2005) or Markel and Gray (1976).

Equation (11.101) for computing k_i^P is the geometric mean between a value k_i^f that minimizes the mean-squared forward prediction error and a value k_i^b that minimizes the mean-squared backward prediction error. The derivation of this result is considered in Problem 11.28. Note that we have shown the limits on the sums as infinite simply to emphasize that *all* error samples are involved in the sum. To be more specific, all the sums in Eq. (11.101) could start at $n = 0$ and end at $n = M + i$, since this is the range over which the error signal output of both the forward and backward i^{th}-order predictors would be nonzero. This is the same assumption that was made in setting up the autocorrelation method for finite-length sequences. Indeed, Problem 11.29 outlines a proof that k_i^P computed by Eq. (11.101) gives identically the same result as k_i computed by Eq. (11.81) or Eq. (L–D.2) in Figure 11.17. Therefore, Eq. (11.101) can be substituted for Eq. (L–D.2) in Figure 11.17, and the resulting set of prediction coefficients will be identical to those computed from the autocorrelation function.

To use Eq. (11.101), it is necessary to actually compute the forward and backward prediction errors by employing the computations of Figure 11.19. In summary, the following steps result in computation of the PARCOR coefficients k_i^P for $i = 1, 2, \ldots, p$:

PARCOR.0 Initialize with $e^{(0)}[n] = \bar{e}^{(0)}[n] = s[n]$ for $0 \le n \le M$.

For $i = 1, 2, \ldots, p$ repeat the following steps.

PARCOR.1 Compute $e^{(i)}[n]$ and $\bar{e}^{(i-1)}[n]$ using Eq. (11.99b) and Eq. (11.99c) respectively for $0 \leq n \leq M+i$. Save these two sequences as input for the next stage.

PARCOR.2 Compute k_i^P using Eq. (11.101).

Another approach to computing the coefficients in Figure 11.20 was introduced by Burg, 1975, who formulated the all-pole modeling problem in terms of the maximum entropy principle. He proposed to use the structure of Figure 11.20, which embodies the Levinson–Durbin algorithm, with coefficients k_i^B that minimize the sum of the mean-squared forward and backward prediction errors at the output of each stage. The result is given by the equation

$$k_i^B = \frac{2\sum_{n=i}^{N} e^{(i-1)}[n]\bar{e}^{(i-1)}[n-1]}{\sum_{n=i}^{N}(e^{(i-1)}[n])^2 + \sum_{n=i}^{N}(\bar{e}^{(i-1)}[n-1])^2} \tag{11.102}$$

The procedure for using this equation to obtain the sequence k_i^B, $i = 1, 2, \ldots, p$ is the same as the PARCOR method. In statement PARCOR.2, k_i^P is simply replaced by k_i^B from Eq. (11.102). In this case, the averaging operator is the same as in the covariance method, which means that very short segments of $s[n]$ can be used, while maintaining high spectral resolution.

Even though the Burg method uses a covariance-type analysis, the condition $|k_i^B| < 1$ holds, implying that the all-pole model implemented by the lattice filter will be stable. (See Problem 11.30.) Just as in the case of the PARCOR method, Eq. (11.102) can be substituted for Eq. (L–D.2) in Figure 11.17 to compute the prediction coefficients. While the resulting coefficients will differ from those obtained from the autocorrelation function or from Eq. (11.101), the resulting all-pole model will still be stable. The derivation of Eq. (11.102) is the subject of Problem 11.30.

11.8 SUMMARY

This chapter provides an introduction to parametric signal modeling. We have emphasized all-pole models, but many of the concepts discussed apply to more general techniques involving rational system functions. We have shown that the parameters of an all-pole model can be computed by a two-step process. The first step is the computation of correlation values from a finite-length signal. The second step is solving a set of linear equations, where the correlation values comprise the coefficients. We showed that the solutions obtained depend on how the correlation values are computed, and we showed that if the correlation values are true autocorrelation values, a particularly useful algorithm, called the Levinson–Durbin algorithm, can be derived for the solution of the equations. Furthermore, the structure of the Levinson–Durbin algorithm was shown to illuminate many useful properties of the all-pole model. The subject of parametric signal modeling has a rich history, a voluminous literature, and abundant applications, all of which make it a subject worthy of further advanced study.

Problems

Basic Problems

11.1. $s[n]$ is a finite-energy signal known for all n. $\phi_{ss}[i, k]$ is defined as

$$\phi_{ss}[i, k] = \sum_{n=-\infty}^{\infty} s[n - i]s[n - k].$$

Show that $\phi_{ss}[i, k]$ can be expressed as a function of $|i - k|$.

11.2. In general, the mean-squared prediction error is defined in Eq. (11.36) as

$$\mathcal{E} = \left\langle \left(s[n] - \sum_{k=1}^{p} a_k s[n - k] \right)^2 \right\rangle. \tag{P11.2-1}$$

(a) Expand Eq. (P11.2-1) and use the fact that $\langle s[n - i]s[n - k] \rangle = \phi_{ss}[i, k] = \phi_{ss}[k, i]$ to show that

$$\mathcal{E} = \phi_{ss}[0, 0] - 2\sum_{k=1}^{p} a_k \phi_{ss}[0, k] + \sum_{i=1}^{p} a_i \sum_{k=1}^{p} a_k \phi_{ss}[i, k] \tag{P11.2-2}$$

(b) Show that for the optimum predictor coefficients, which satisfy Eqs. (11.20), Eq. (P11.2-2) becomes

$$\mathcal{E} = \phi_{ss}[0, 0] - \sum_{k=1}^{p} a_k \phi_{ss}[0, k]. \tag{P11.2-3}$$

11.3. The impulse response of a causal all-pole model of the form of Figure 11.1 and Eq. (11.3) with system parameters G and $\{a_k\}$ satisfies the difference equation

$$h[n] = \sum_{k=1}^{p} a_k h[n - k] + G\delta[n] \tag{P11.3-1}$$

(a) The autocorrelation function of the impulse response of the system is

$$r_{hh}[m] = \sum_{n=-\infty}^{\infty} h[n]h[n + m]$$

By substituting Eq. (P11.3-1) into the equation for $r_{hh}[-m]$, and using the fact that $r_{hh}[-m] = r_{hh}[m]$ show that

$$\sum_{k=1}^{p} a_k r_{hh}[|m - k|] = r_{hh}[m], \quad m = 1, 2, \ldots, p \tag{P11.3-2}$$

(b) Using the same approach as in (a), now show that

$$r_{hh}[0] - \sum_{k=1}^{p} a_k r_{hh}[k] = G^2. \tag{P11.3-3}$$

11.4. Consider a signal $x[n] = s[n] + w[n]$, where $s[n]$ satisfies the difference equation

$$s[n] = 0.8s[n-1] + v[n].$$

$v[n]$ is a zero-mean white-noise sequence with variance $\sigma_v^2 = 0.49$ and $w[n]$ is a zero-mean white-noise sequence with variance $\sigma_w^2 = 1$. The processes $v[n]$ and $w[n]$ are uncorrelated. Determine the autocorrelation sequences $\phi_{ss}[m]$ and $\phi_{xx}[m]$.

11.5. The inverse filter approach to all-pole modeling of a deterministic signal $s[n]$ is discussed in Section 11.1.2 and depicted in Fig. 11.2. The system function of the inverse filter is given in Eq. (11.5).

(a) Based on this approach, determine the coefficients a_1 and a_2 of the best all-pole model for $s[n] = \delta[n] + \delta[n-2]$ with $p = 2$.

(b) Again, based on this approach, determine the coefficients a_1, a_2 and a_3 of the best all-pole model for $s[n] = \delta[n] + \delta[n-2]$ with $p = 3$.

11.6. Suppose that you have computed the parameters G and a_k, $k = 1, 2, \ldots, p$ of the all-pole model

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}.$$

Explain how you might use the DFT to evaluate the all-pole spectrum estimate $|H(e^{j\omega_k})|$ at N frequencies $\omega_k = 2\pi k/N$ for $k = 0, 1, \ldots, N-1$.

11.7. Consider a desired causal impulse response $h_d[n]$ that we wish to approximate by a system having impulse response $h[n]$ and system function

$$H(z) = \frac{b}{1 - az^{-1}}.$$

Our optimality criterion is to minimize the error function given by

$$\mathcal{E} = \sum_{n=0}^{\infty} (h_d[n] - h[n])^2.$$

(a) Suppose a is given, and we wish to determine the unknown parameter b which minimizes $\mathcal{E}$. Assume that $|a| < 1$. Does this result in a nonlinear set of equations? If so, show why. If not, determine b.

(b) Suppose b is given, and we wish to determine the unknown parameter a which minimizes $\mathcal{E}$. Is this a nonlinear problem? If so, show why. If not, determine a.

11.8. Assume that $s[n]$ is a finite-length (windowed) sequence that is zero *outside* the interval $0 \le n \le M-1$. The p^{th}-order *backward* linear prediction error sequence for this signal is defined as

$$\bar{e}[n] = s[n] - \sum_{k=1}^{p} \beta_k s[n+k]$$

That is, $s[n]$ is "predicted" from the p samples that *follow* sample n. The mean-squared backward prediction error is defined as

$$\bar{\mathcal{E}} = \sum_{m=-\infty}^{\infty} (\bar{e}[m])^2 = \sum_{m=-\infty}^{\infty} \left(s[m] - \sum_{k=1}^{p} \beta_k s[m+k] \right)^2$$

where the infinite limits indicate that the sum is over all nonzero values of $(\tilde{e}[m])^2$ as in the autocorrelation method used in "forward prediction."

(a) The prediction error sequence $\tilde{e}[n]$ is zero outside a finite interval $N_1 \leq n \leq N_2$. Determine N_1 and N_2.

(b) Following the approach used in this chapter to derive the forward linear predictor, derive the set of normal equations that are satisfied by the β_ks that minimize the mean-squared prediction error $\tilde{\mathcal{E}}$. Give your final answer in a concise, well-defined form in terms of autocorrelation values.

(c) Based on the result in part (b), describe how the backward predictor coefficients $\{\beta_k\}$ related to the forward predictor coefficients $\{\alpha_k\}$?

Advanced Problems

11.9. Consider a signal $s[n]$ that we model as the impulse response of an all-pole system of order p. Denote the system function of the p^{th}-order all-pole model as $H^{(p)}(z)$ and the corresponding impulse response as $h^{(p)}[n]$. Denote the inverse of $H^{(p)}(z)$ as $H_{\text{inv}}^{(p)}(z) = 1/H^{(p)}(z)$. The corresponding impulse response is $h_{\text{inv}}^{(p)}[n]$. The inverse filter, characterized by $h_{\text{inv}}^{(p)}[n]$, is chosen to minimize the total squared error $\mathcal{E}^{(p)}$ given by

$$\mathcal{E}^{(p)} = \sum_{n=-\infty}^{\infty} \left[\delta[n] - g^{(p)}[n]\right]^2,$$

where $g^{(p)}[n]$ is the output of the filter $H_{\text{inv}}^{(p)}(z)$ when the input is $s[n]$.

(a) Figure P11.9 depicts a signal flow graph of the lattice filter implementation of $H_{\text{inv}}^{(4)}(z)$. Determine $h_{\text{inv}}^{(4)}[1]$, the impulse response at $n = 1$.

(b) Suppose we now wish to model the signal $s[n]$ as the impulse response of a 2^{nd}-order all-pole filter. Draw a signal flow graph of the lattice filter implementation of $H_{\text{inv}}^{(2)}(z)$.

(c) Determine the system function $H^{(2)}(z)$ of the 2^{nd}-order all-pole filter.

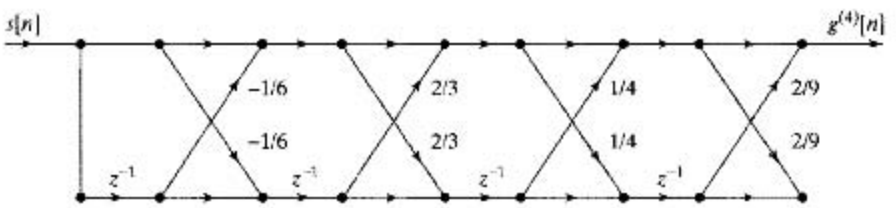


Figure P11.9

11.10. Consider an i^{th}-order predictor with prediction error system function

$$A^{(i)}(z) = 1 - \sum_{j=1}^{i} a_j^{(i)} z^{-j} = \prod_{j=1}^{i} (1 - z_j^{(i)} z^{-1}) \qquad \text{(P11.10-1)}$$

From the Levinson–Durbin recursion, it follows that $a_i^{(i)} = k_i$. Use this fact with Eq. (P11.10-1) to show that if $|k_i| \geq 1$, it must be true that $|z_j^{(i)}| \geq 1$ for some j. That

is, show that the condition $|k_i| < 1$ is a *necessary* condition for $A^{(p)}(z)$ to have all its zeros strictly *inside* the unit circle.

11.11. Consider an LTI system with system function $H(z) = h_0 + h_1 z^{-1}$. The signal $y[n]$ is the output of this system due to an input that is white noise with zero mean and unit variance.

(a) What is the autocorrelation function $r_{yy}[m]$ of the output signal $y[n]$?

(b) The 2nd-order forward prediction error is defined as

$$e[n] = y[n] - a_1 y[n-1] - a_2 y[n-2].$$

Without using the Yule–Walker equations directly, find a_1 and a_2, such that the variance of $e[n]$ is minimized.

(c) The backward prediction filter for $y[n]$ is defined as

$$\tilde{e}[n] = y[n] - b_1 y[n+1] - b_2 y[n+2].$$

Find b_1 and b_2 such that the variance of $\tilde{e}[n]$ is minimized. Compare these coefficients to those determined in part (b).

11.12. **(a)** The autocorrelation function, $r_{yy}[m]$ of a zero-mean wide-sense stationary random process $y[n]$ is given. In terms of $r_{yy}[m]$, write the Yule–Walker equations that result from modeling the random process as the response to a white noise sequence of a 3rd-order all-pole model with system function

$$H(z) = \frac{A}{1 - az^{-1} - bz^{-3}}.$$

(b) A random process $v[n]$ is the output of the system shown in Figure P11.12-1, where $x[n]$ and $z[n]$ are independent, unit variance, zero mean, white noise signals, and $h[n] = \delta[n-1] + \frac{1}{2}\delta[n-2]$. Find $r_{vv}[m]$, the autocorrelation of $v[n]$.

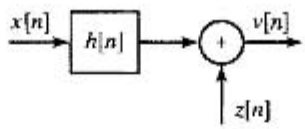


Figure P11.12-1

(c) Random process $y_1[n]$ is the output of the system shown in Figure P11.12-2, where $x[n]$ and $z[n]$ are independent, unit variance, zero-mean, white noise signals, and

$$H_1(z) = \frac{1}{1 - az^{-1} - bz^{-3}}.$$

The same a and b as found in part (a) are used for all-pole modeling of $y_1[n]$. The inverse modeling error, $w_1[n]$, is the output of the system shown in Figure P11.12-3. Is $w_1[n]$ white? Is $w_1[n]$ zero mean? Explain.

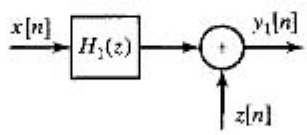


Figure P11.12-2

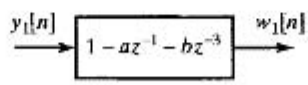


Figure P11.12-3

(d) What is the variance of $w_1[n]$?

11.13. We have observed the first six samples of a causal signal $s[n]$ given by $s[0] = 4$, $s[1] = 8$, $s[2] = 4$, $s[3] = 2$, $s[4] = 1$, and $s[5] = 0.5$. For the first parts of this problem, we will model the signal using a stable, causal, minimum-phase, two-pole system having impulse response $\hat{s}[n]$ and system function

$$H(z) = \frac{G}{1 - a_1 z^{-1} - a_2 z^{-2}}.$$

The approach is to minimize the modeling error $\mathcal{E}$ given by

$$\mathcal{E} = \min_{a_1, a_2, A} \sum_{n=0}^{5} (g[n] - G\delta[n])^2,$$

where $g[n]$ is the response of the inverse system to $s[n]$, and the inverse system has system function

$$A(z) = 1 - a_1 z^{-1} - a_2 z^{-2}.$$

(a) Write $g[n] - G\delta[n]$ for $0 \leq n \leq 5$.
(b) Based on your work in part (a), write the linear equations for the desired parameters a_1, a_2, and G.
(c) What is G?
(d) For this $s[n]$, without solving the linear equations in part (b), discuss whether you expect that the modeling error $\mathcal{E}$ will be zero.

For the rest of this problem, we will model the signal using a different stable, causal, minimum-phase system having impulse response $\hat{s}_2[n]$ and system function

$$H_2(z) = \frac{b_0 + b_1 z^{-1}}{1 - a z^{-1}}.$$

The modeling error to be minimized in this case is $\mathcal{E}_2$ given by

$$\mathcal{E}_2 = \min_{a, b_0, b_1} \sum_{n=0}^{5} (g[n] - r[n])^2,$$

where $g[n]$ is the response of the inverse system to $s[n]$, and the inverse system now has system function

$$A(z) = 1 - a z^{-1}.$$

Furthermore, $r[n]$ is the impulse response of a system with system function

$$B(z) = b_0 + b_1 z^{-1}.$$

(e) For this model, write $g[n] - r[n]$ for $0 \leq n \leq 5$.
(f) Calculate the parameter values a, b_0, and b_1 that minimize the modeling error.
(g) Calculate the modeling error $\mathcal{E}_2$ in part (f).

11.14. In Example 11.1, we considered the sequence $s_d[n] = \alpha^n u[n]$, which is the impulse response of a 1st-order all-pole system having system function

$$H(z) = \frac{1}{1 - \alpha z^{-1}}.$$

In this problem we consider the estimation of the parameters of an all-pole model for the signal $s_d[n]$ known only over the interval $0 \leq n \leq M$.

(a) First, consider the estimation of a 1st-order model by the autocorrelation method. To begin, show that the autocorrelation function of the finite-length sequence $s[n] = s_d[n](u[n] - u[n-M-1]) = \alpha^n(u[n] - u[n-M-1])$ is

$$r_{ss}[m] = \alpha^{|m|}\frac{1-\alpha^{2(M-|m|+1)}}{1-\alpha^2}. \qquad \text{(P11.14-1)}$$

(b) Use the autocorrelation function determined in (a) in Eq. (11.34), and solve for the coefficient a_1 of the 1st-order predictor.

(c) You should find that the result obtained in (b) is not the exact value (i.e., $a_1 \neq \alpha$) as obtained in Example 11.1, when the autocorrelation function was computed using the infinite sequence. Show, however, that $a_1 \to \alpha$ for $M \to \infty$.

(d) Use the results of (a) and (b) in Eq. (11.38) to determine the minimum mean-squared prediction error for this example. Show that for $M \to \infty$ the error approaches the minimum mean-squared error found in Example 11.1 for the exact autocorrelation function.

(e) Now, consider the covariance method for estimating the correlation function. Show that for $p = 1$, $\phi_{ss}[i,k]$ in Eq. (11.49) is given by

$$\phi_{ss}[i,k] = \alpha^{2-i-k}\frac{1-\alpha^{2M}}{1-\alpha^2} \qquad 0 \le (i,k) \le 1. \qquad \text{(P11.14-2)}$$

(f) Use the result of (e) in Eq. (11.20) to solve for the coefficient of the optimum 1st-order predictor. Compare your result to the result in (b) and to the result in Example 11.1.

(g) Use the results of (e) and (f) in Eq. (11.37) to find the minimum mean-squared prediction error. Compare your result to the result in (d) and to the result in Example 11.1.

11.15. Consider the signal

$$s[n] = 3\left(\frac{1}{2}\right)^n u[n] + 4\left(-\frac{2}{3}\right)^n u[n].$$

(a) We want to use a causal, 2nd-order all-pole model, i.e., a model of the form

$$H(z) = \frac{A}{1 - a_1 z^{-1} - a_2 z^{-2}},$$

to optimally represent the signal $s[n]$, in the least-square error sense. Find a_1, a_2, and A.

(b) Now, suppose we want to use a causal, 3rd-order all-pole model, i.e., a model of the form

$$H(z) = \frac{B}{1 - b_1 z^{-1} - b_2 z^{-2} - b_3 z^{-3}},$$

to optimally represent the signal $s[n]$, in the least-square error sense. Find, b_1, b_2, b_3, and B.

11.16. Consider the signal

$$s[n] = 2\left(\frac{1}{3}\right)^n u[n] + 3\left(-\frac{1}{2}\right)^n u[n]. \qquad \text{(P11.16-1)}$$

We wish to model this signal using a 2nd-order ($p = 2$) all-pole model or, equivalently, using 2nd-order linear prediction.

For this problem, since we are given an analytical expression for $s[n]$ and $s[n]$ is the impulse response of an all-pole filter, we can obtain the linear prediction coefficients directly from

the z-transform of $s[n]$. (You are asked to do this in part (a).) In practical situations, we are typically given data, i.e., a set of signal values, and not an analytical expression. In this case, even when the signal to be modeled is the impulse response of an all-pole filter, we need to perform some computation on the data, using methods such as those discussed in Section 11.3, to determine the linear prediction coefficients.

There are also situations in which an analytical expression is available for the signal, but the signal is not the impulse response of an all-pole filter, and we would like to model it as such. In this case, we again need to carry out computations such as those discussed in Section 11.3.

(a) For $s[n]$ as given in Eq. (P11.16-1), determine the linear prediction coefficients a_1, a_2 directly from the z-transform of $s[n]$.

(b) Write the normal equations for $p = 2$ to obtain equations for a_1, a_2 in terms of $r_{ss}[m]$.

(c) Determine the values of $r_{ss}[0]$, $r_{ss}[1]$, and $r_{ss}[2]$ for the signal $s[n]$ given in Eq. (P11.16-1).

(d) Solve the system of equations from part (a) using the values you found in part (b) to obtain values for the a_ks.

(e) Are the values of a_k from part (c) what you would expect for this signal? Justify your answer clearly.

(f) Suppose you wish to model the signal now with $p = 3$. Write the normal equations for this case.

(g) Find the value of $r_{ss}[3]$.

(h) Solve for the values of a_k when $p = 3$.

(i) Are the values of a_k found in part (h) what you would expect given $s[n]$? Justify your answer clearly.

(j) Would the values of a_1, a_2 you found in (h) change if we model the signal with $p = 4$?

11.17. $x[n]$ and $y[n]$ are sample sequences of jointly wide-sense stationary, zero-mean random processes. The following information is known about the autocorrelation function $\phi_{xx}[m]$ and cross correlation $\phi_{yx}[m]$:

$$\phi_{xx}[m] = \begin{cases} 0 & m \text{ odd} \\ \frac{1}{2^{|m|}} & m \text{ even} \end{cases}$$

$$\phi_{yx}[-1] = 2 \quad \phi_{yx}[0] = 3 \quad \phi_{yx}[1] = 8 \quad \phi_{yx}[2] = -3$$

$$\phi_{yx}[3] = 2 \quad \phi_{yx}[4] = -0.75$$

(a) The linear estimate of y given x is denoted $\hat{y}_x$. It is designed to minimize

$$\mathcal{E} = E\left(|\, y[n] - \hat{y}_x[n] \,|^2 \right), \tag{P11.17-1}$$

where the $\hat{y}_x[n]$ is formed by processing $x[n]$ with an FIR filter whose impulse response $h[n]$ is of length 3 and is given by

$$h[n] = h_0\delta[n] + h_1\delta[n-1] + h_2\delta[n-2].$$

Determine h_0, h_1, and h_2 to minimize $\mathcal{E}$.

(b) In this part, $\hat{y}_x$, the linear estimate of y given x, is again designed to minimize $\mathcal{E}$ in Eq. (P11.17-1), but with different assumptions on the structure of the linear filter. Here the estimate is formed by processing $x[n]$ with an FIR filter whose impulse response $g[n]$ is of length 2 and is given by

$$g[n] = g_1\delta[n-1] + g_2\delta[n-2].$$

Determine the g_1 and g_2 to minimize $\mathcal{E}$.

(c) The signal, $x[n]$ can be modeled as the output from a two-pole filter $H(z)$ whose input is $w[n]$, a wide-sense stationary, zero-mean, unit-variance white-noise signal.

$$H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2}}$$

Determine a_1 and a_2 based on the least-squares inverse model in Section 11.1.2.

(d) We want to implement the system shown in Figure P11.17 where the coefficients a_i are from all-pole modeling in part (c) and the coefficients h_i are the values of the impulse response of the linear estimator in part (a). Draw an implementation that minimizes the total cost of delays, where the cost of each individual delay is weighted linearly by its clock rate.

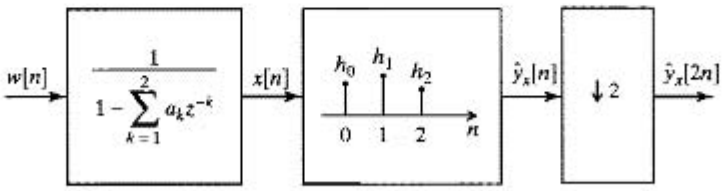


Figure P11.17

(e) Let $\mathcal{E}_a$ be the cost in part (a) and let $\mathcal{E}_b$ be the cost in part (b), where each $\mathcal{E}$ is defined as in Eq. (P11.17-1). Is $\mathcal{E}_a$ larger than, equal to, or smaller than $\mathcal{E}_b$, or is there not enough information to compare them?

(f) Calculate $\mathcal{E}_a$ and $\mathcal{E}_b$ when $\phi_{yy}[0] = 88$. (*Hint*: The optimum FIR filters calculated in parts (a) and (b) are such that $E[\hat{y}_x[n](y[n] - \hat{y}_x[n])] = 0$.)

11.18. A discrete-time communication channel with impulse response $h[n]$ is to be compensated for with an LTI system with impulse response $h_c[n]$ as indicated in Figure P11.18. The channel $h[n]$ is known to be a one-sample delay, i.e.,

$$h[n] = \delta[n - 1].$$

The compensator $h_c[n]$ is an N-point causal FIR filter, i.e.,

$$H_c(z) = \sum_{k=0}^{N-1} a_k z^{-k}.$$

The compensator $h_c[n]$ is designed to invert (or compensate for) the channel. Specifically, $h_c[n]$ is designed so that with $s[n] = \delta[n]$, $\hat{s}[n]$ is as "close" as possible to an impulse; i.e., $h_c[n]$ is designed so that the error

$$\mathcal{E} = \sum_{n=-\infty}^{\infty} |\hat{s}[n] - \delta[n]|^2$$

is minimized. Find the optimal compensator of length N, i.e., determine $a_0, a_1, \ldots, a_{N-1}$ to minimize $\mathcal{E}$.

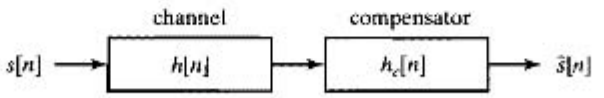


Figure P11.18

11.19. A speech signal was sampled with a sampling rate of 8 kHz. A 300-sample segment was selected from a vowel sound and multiplied by a Hamming window as shown in Figure P11.19. From this signal a set of linear predictors

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

with orders ranging from $i = 1$ to $i = 11$ was computed using the autocorrelation method. This set of predictors is shown in Table 11.1 below in a form suggestive of the Levinson–Durbin recursion.

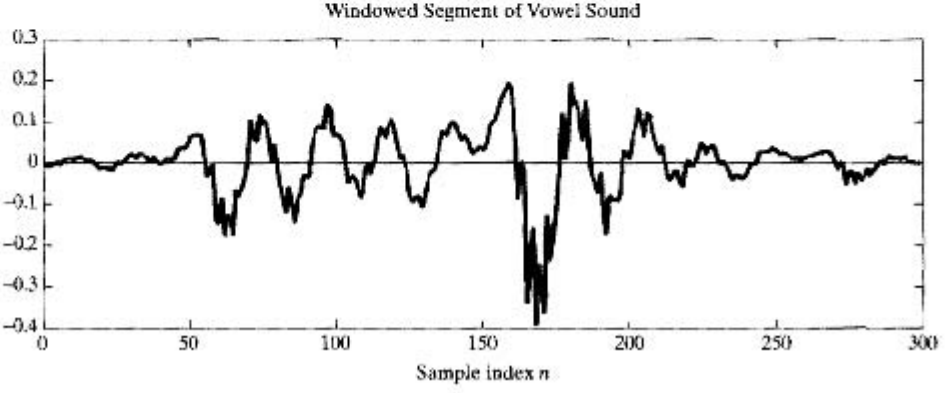


Figure P11.19

(a) Determine the z-transform $A^{(4)}(z)$ of the 4^{th}-order prediction error filter. Draw and label the flow graph of the direct form implementation of this system.

(b) Determine the set of k-parameters $\{k_1, k_2, k_3, k_4\}$ for the 4^{th}-order prediction error lattice filter. Draw and label the flow graph of the lattice implementation of this system.

(c) The minimum mean-squared prediction error for the 2^{nd}-order predictor is $E^{(2)} = 0.5803$. What is the minimum mean-squared prediction error for the 3^{rd}-order predictor? What is the total energy of the signal $s[n]$? What is the value of the autocorrelation function $r_{ss}[1]$?

TABLE 11.1 PREDICTION COEFFICIENTS FOR A SET OF LINEAR PREDICTORS

i	$a_1^{(i)}$	$a_2^{(i)}$	$a_3^{(i)}$	$a_4^{(i)}$	$a_5^{(i)}$	$a_6^{(i)}$	$a_7^{(i)}$	$a_8^{(i)}$	$a_9^{(i)}$	$a_{10}^{(i)}$	$a_{11}^{(i)}$
1	0.8328										
2	0.7459	0.1044									
3	0.7273	−0.0289	0.1786								
4	0.8047	−0.0414	0.4940	−0.4337							
5	0.7623	0.0069	0.4899	−0.3550	−0.0978						
6	0.6889	−0.2595	0.8576	−0.3498	0.4743	−0.7505					
7	0.6839	−0.2563	0.8553	−0.3440	0.4726	−0.7459	−0.0067				
8	0.6834	−0.3095	0.8890	−0.3685	0.5336	−0.7642	0.0421	−0.0713			
9	0.7234	−0.3331	1.3173	−0.6676	0.7402	−1.2624	0.2155	−0.4544	0.5605		
10	0.6493	−0.2730	1.2888	−0.5007	0.6423	−1.1741	0.0413	−0.4103	0.4648	0.1323	
11	0.6444	−0.2902	1.3040	−0.5022	0.6859	−1.1980	0.0599	−0.4582	0.4749	0.1081	0.0371

(d) The minimum mean-squared prediction errors for these predictors form a sequence $\{E^{(0)}, E^{(1)}, E^{(2)}, \ldots, E^{(11)}\}$. It can be shown that this sequence decreases abruptly in going from $i = 0$ to $i = 1$ and then decreases slowly for several orders and then makes a sharp decrease. At what order i would you expect this to occur?

(e) Sketch carefully the prediction error sequence $e^{(11)}[n]$ for the given input $s[n]$ in Figure P11.19. Show as much detail as possible.

(f) The system function of the 11^{th}-order all-pole model is

$$H(z) = \frac{G}{A^{(11)}(z)} = \frac{G}{1 - \sum_{k=1}^{11} a_k^{(11)} z^{-k}} = \frac{G}{\prod_{i=1}^{11}(1 - z_i z^{-1})}.$$

The following are five of the roots of the 11^{th}-order prediction error filter $A^{(11)}(z)$.

i	$\lvert z_i \rvert$	$\angle z_i$ (rad)
1	0.2567	2.0677
2	0.9681	1.4402
3	0.9850	0.2750
4	0.8647	2.0036
5	0.9590	2.4162

State briefly in words where the other six zeros of $A^{(11)}(z)$ are located. Be as precise as possible.

(g) Use information given in the table and in part (c) of this problem to determine the gain parameter G for the 11^{th}-order all-pole model.

(h) Carefully sketch and label a plot of the frequency response of the 11^{th}-order all-pole model filter for analog frequencies $0 \leq F \leq 4$ kHz.

11.20. Spectrum analysis is often applied to signals comprised of sinusoids. Sinusoidal signals are particularly interesting, because they share properties with both deterministic and random signals. On the one hand, we can describe them in terms of a simple equation. On the other hand, they have infinite energy, so we often characterize them in terms of their average power, just as with random signals. This problem explores some theoretical issues in modeling sinusoidal signals from the point of view of random signals.

We can consider sinusoidal signals as stationary random signals by assuming that the signal model is $s[n] = A\cos(\omega_0 n + \theta)$ for $-\infty < n < \infty$, where both A and θ can be considered to be random variables. In this model, the signal is considered to be an ensemble of sinusoids described by underlying probability laws for A and θ. For simplicity, assume that A is a constant, and θ is a random variable that is uniformly distributed over $0 \leq \theta < 2\pi$.

(a) Show that the autocorrelation function for such a signal is

$$r_{ss}[m] = E\{s[n+m]s[n]\} = \frac{A^2}{2}\cos(\omega_0 m). \qquad \text{(P11.20-1)}$$

(b) Using Eq. (11.34), write the set of equations that is satisfied by the coefficients of a 2^{nd}-order linear predictor for this signal.

(c) Solve the equations in (b) for the optimum predictor coefficients. Your answer should be a function of ω_0.

(d) Factor the polynomial $A(z) = 1 - a_1 z^{-1} - a_2 z^{-2}$ describing the prediction error filter.

(e) Use Eq. (11.37) to determine an expression for the minimum mean-squared prediction error. Your answer should confirm why random sinusoidal signals are called "predictable" and/or "deterministic."

11.21. Using Eqs. (11.84a) and (11.84b) from the Levinson–Durbin recursion, derive the relation between the i^{th} and $i-1^{\text{st}}$ prediction error filters given in Eq. (11.89).

11.22. In this problem, we consider the construction of lattice filters to implement the inverse filter for the signal

$$s[n] = 2\left(\frac{1}{3}\right)^n u[n] + 3\left(-\frac{1}{2}\right)^n u[n].$$

(a) Find the values of the k-parameters k_1 and k_2 for the 2^{nd}-order case (i.e., $p = 2$).
(b) Draw the signal flow graph of a lattice filter implementation of the inverse filter, i.e., the filter that outputs $y[n] = A\delta[n]$ (a scaled impulse) when the input $x[n] = s[n]$.
(c) Verify that the signal flow graph you drew in part (b) has the correct impulse response by showing that the z-transform of this inverse filter is indeed proportional to the inverse of $S(z)$.
(d) Draw the signal flow graph for a lattice filter that implements an all-pole system such that when the input is $x[n] = \delta[n]$, the output is the signal $s[n]$ given above.
(e) Derive the system function of the signal flow graph you drew in part (d) and demonstrate that its impulse response $h[n]$ satisfies $h[n] = s[n]$.

11.23. Consider the signal

$$s[n] = \alpha\left(\frac{2}{3}\right)^n u[n] + \beta\left(\frac{1}{4}\right)^n u[n]$$

where α and β are constants. We wish to linearly predict $s[n]$ from its past p values using the relationship

$$\hat{s}[n] = \sum_{k=1}^{p} a_k s[n-k]$$

where the coefficients a_k are constants. The coefficients a_k are chosen to minimize the prediction error

$$\mathcal{E} = \sum_{n=-\infty}^{\infty} (s[n] - \hat{s}[n])^2.$$

(a) With $r_{ss}[m]$ denoting the autocorrelation function of $s[n]$, write the equations for the case $p = 2$ the solution to which will result in a_1, a_2.
(b) Determine a pair of values for α and β such that when $p = 2$, the solution to the normal equations is $a_1 = \frac{11}{12}$ and $a_2 = -\frac{1}{6}$. Is your answer unique? Explain.
(c) If $\alpha = 8$ and $\beta = -3$, determine k-parameter k_3, resulting from using the Levinson recursion to solve the normal equations for $p = 3$. Is that different from k_3 when solving for $p = 4$?

11.24. Consider the following Yule–Walker equations: $\mathbf{\Gamma}_p \boldsymbol{a}_p = \boldsymbol{\gamma}_p$, where:

$$\boldsymbol{a}_p = \begin{bmatrix} a_1^p \\ \vdots \\ a_p^p \end{bmatrix} \qquad \boldsymbol{\gamma}_p = \begin{bmatrix} \phi[1] \\ \vdots \\ \phi[p] \end{bmatrix}$$

and

$$\mathbf{\Gamma}_p = \begin{bmatrix} \phi[0] & \cdots & \phi[p-1] \\ \vdots & \ddots & \vdots \\ \phi[p-1] & \cdots & \phi[0] \end{bmatrix} \qquad \text{(a Toeplitz matrix)}$$

The Levinson–Durbin algorithm provides the following recursive solution for the normal equation $\mathbf{\Gamma}_{p+1}\,\boldsymbol{a}_{p+1} = \boldsymbol{\gamma}_{p+1}$:

$$a_{p+1}^{p+1} = \frac{\phi[p+1] - \left(\boldsymbol{\gamma}_p^b\right)^T \boldsymbol{a}_p}{\phi[0] - \left(\boldsymbol{\gamma}_p\right)^T \boldsymbol{a}_p} \qquad a_m^{p+1} = a_m^p - a_{p+1}^{p+1} \cdot a_{p-m+1}^p \quad m = 1, \ldots, p$$

where $\boldsymbol{\gamma}_p^b$ is the backward version of $\boldsymbol{\gamma}_p$: $\boldsymbol{\gamma}_p^b = [\phi[p] \ldots \phi[1]]^T$, and $a_1^1 = \frac{\phi[1]}{\phi[0]}$. Note that for vectors, the model order is shown in the subscript; but for scalars, the model order is shown in the superscript.

Now consider the following normal equation: $\mathbf{\Gamma}_p\,\boldsymbol{b}_p = \boldsymbol{c}_p$, where

$$\boldsymbol{b}_p = \begin{bmatrix} b_1^p \\ \vdots \\ b_p^p \end{bmatrix} \qquad \boldsymbol{c}_p = \begin{bmatrix} c[1] \\ \vdots \\ c[p] \end{bmatrix}$$

Show that the recursive solution for $\mathbf{\Gamma}_{p+1}\,\boldsymbol{b}_{p+1} = \boldsymbol{c}_{p+1}$ is:

$$b_{p+1}^{p+1} = \frac{c[p+1] - \left(\boldsymbol{\gamma}_p^b\right)^T \boldsymbol{b}_p}{\phi[0] - \left(\boldsymbol{\gamma}_p\right)^T \boldsymbol{a}_p} \qquad b_m^{p+1} = b_m^p - b_{p+1}^{p+1} \cdot a_{p-m+1}^p \quad m = 1, \ldots, p$$

where $b_1^1 = \frac{c[1]}{\phi[0]}$.

(*Note*: You may find it helpful to note that $\boldsymbol{a}_p^b = \mathbf{\Gamma}_p^{-1}\boldsymbol{\gamma}_p^b$.)

11.25. Consider a colored wide-sense stationary random signal $s[n]$ that we desire to whiten using the system in Figure P11.25-1: In designing the optimal whitening filter for a given order p, we pick the coefficient $a_k^{(p)}$, $k = 1, \ldots, p$ that satisfy the autocorrelation normal equations given by Eq. (11.34), where $r_{ss}[m]$ is the autocorrelation of $s[n]$.

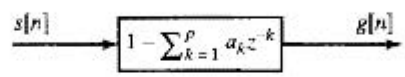


Figure P11.25-1

It is known that the optimal 2nd-order whitening filter for $s[n]$ is $H_2(z) = 1 + \frac{1}{4}z^{-1} - \frac{1}{8}z^{-2}$, (i.e., $a_1^{(2)} = -\frac{1}{4}$, $a_2^{(2)} = \frac{1}{8}$), which we implement in the 2nd-order lattice structure in Figure P11.25-2. We would also like to use a 4th-order system, with transfer function

$$H_4(z) = 1 - \sum_{k=1}^{4} a_k^{(4)} z^{-k}.$$

We implement this system with the lattice structure in Figure P11.25-3. Determine which, if any of $H_4(z)$, k_1, k_2, k_3, k_4 can be exactly determined from the information given above. Explain why you cannot determine the remaining, if any, parameters.

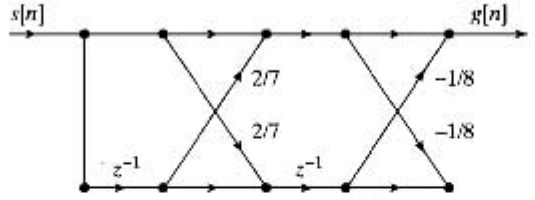


Figure P11.25-2 Lattice structure for 2nd-order system

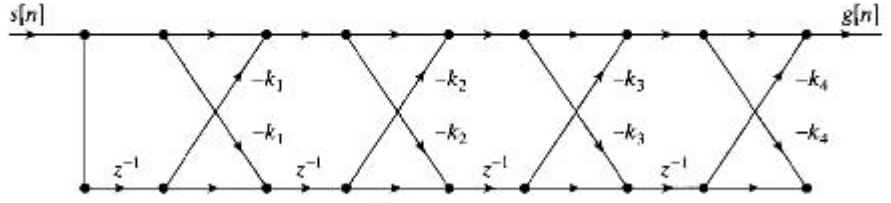


Figure P11.25-3 Lattice structure for 4th-order system

Extension Problems

11.26. Consider a stable all-pole model with system function

$$H(z) = \frac{G}{1 - \sum_{m=1}^{p} a_m z^{-m}} = \frac{G}{A(z)}.$$

Assume that g is positive.

In this problem, we will show that a set of $(p+1)$ samples of the magnitude-squared of $H(z)$ on the unit circle; i.e.,

$$C[k] = |H(e^{j\pi k/p})|^2 \qquad k = 0, 1, \ldots, p,$$

is sufficient to represent the system. Specifically, given $C[k], k = 0, 1, \ldots, p$, show that the parameters G and $a_m, m = 0, 1, \ldots, p$ can be determined.

(a) Consider the z-transform

$$Q(z) = \frac{1}{H(z)H(z^{-1})} = \frac{A(z)A(z^{-1})}{G^2},$$

which corresponds to a sequence $q[n]$. Determine the relationship between $q[n]$ and $h_A[n]$, the impulse response of the prediction error filter whose system function is $A(z)$. Over what range of n will $q[n]$ be nonzero?

(b) Design a procedure based on the DFT for determining $q[n]$ from the given magnitude-squared samples $C[k]$.

(c) Assuming that the sequence $q[n]$ as determined in (b) is known, state a procedure for determining $A(z)$ and G.

11.27. The general IIR lattice system in Figure 11.21 is restricted to all-pole systems. However, both poles and zeros can be implemented by the system of Figure P11.27-1 (Gray and Markel, 1973, 1975). Each of the sections in Figure P11.27-1 is described by the flow graph of Figure P11.27-2. In other words, Figure 11.21 is embedded in Figure P11.27-1 with the output formed as a linear combination of the backward prediction error sequences.

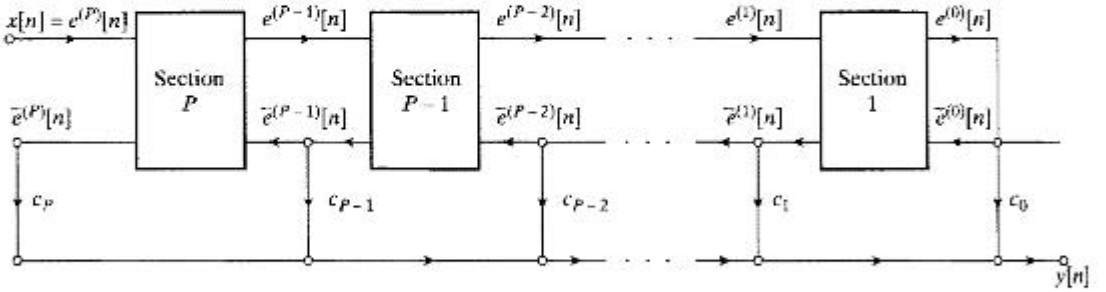


Figure P11.27-1

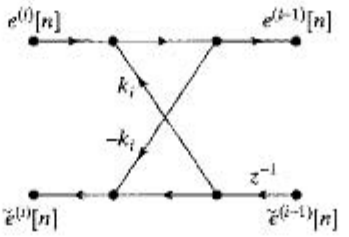


Figure P11.27-2

(a) Show that the system function between the input $X(z) = E^{(p)}(z)$ and $\tilde{E}^{(i)}(z)$ is

$$\tilde{H}^{(i)}(z) = \frac{\tilde{E}^{(i)}(z)}{X(z)} = \frac{z^{-1}A^{(i)}(z^{-1})}{A^{(p)}(z)}. \tag{P11.27-1}$$

(b) Show that $\tilde{H}^{(p)}(z)$ is an all-pass system. (This result is not needed for the rest of the problem.)

(c) The overall system function from $X(z)$ to $Y(z)$ is

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{i=0}^{p} \frac{c_i z^{-1} A^{(i)}(z^{-1})}{A^{(p)}(z)} = \frac{Q(z)}{A^{(p)}(z)}. \tag{P11.27-2}$$

Show that the numerator $Q(z)$ in Eq. (P11.27-2) is a p^{th}-order polynomial of the form

$$Q(z) = \sum_{m=0}^{p} q_m z^{-m} \tag{P11.27-3}$$

where the coefficients c_m in Figure P11.27 are given by the equation

$$c_m = q_m + \sum_{i=m+1}^{p} c_i a_{i-m}^{(i)} \quad m = p, p-1, \ldots, 1, 0. \tag{P11.27-4}$$

(d) Give a procedure for computing all the parameters needed to implement a system function such as Eq. (P11.27-2) using the lattice structure of Figure P11.27.

(c) Using the procedure described in (c), draw the complete flow graph of the lattice implementation of the system

$$H(z) = \frac{1 + 3z^{-1} + 3z^{-2} + z^{-3}}{1 - 0.9z^{-1} + 0.64z^{-2} - 0.576z^{-3}}. \qquad \text{(P11.27-5)}$$

11.28. In Section 11.7.3, the k-parameters are computed by Eqs. (11.101). Using the relations $e^{(i)}[n] = e^{(i-1)}[n] - k_i \tilde{e}^{(i-1)}[n-1]$ and $\tilde{e}^{(i)}[n] = \tilde{e}^{(i-1)}[n-1] - k_i e^{(i-1)}[n]$, show that

$$k_i^P = \sqrt{k_i^f k_i^b},$$

where k_i^f is the value of k_i that minimizes the mean-squared forward prediction error

$$\mathcal{E}^{(i)} = \sum_{n=-\infty}^{\infty} (e^{(i)}[n])^2,$$

and k_i^b is the value of k_i that minimizes the mean-squared backward prediction error

$$\tilde{\mathcal{E}}^{(i)} = \sum_{n=-\infty}^{\infty} (\tilde{e}^{(i)}[n])^2.$$

11.29. Substitute Eq. (11.88) and Eq. (11.93) into Eq. (11.101) to show that

$$k_i^P = \frac{\displaystyle\sum_{n=-\infty}^{\infty} e^{(i-1)}[n]\tilde{e}^{(i-1)}[n-1]}{\left\{\displaystyle\sum_{n=-\infty}^{\infty} (e^{(i-1)}[n])^2 \sum_{n=-\infty}^{\infty} (\tilde{e}^{(i-1)}[n-1])^2\right\}^{1/2}}$$

$$= \frac{r_{ss}[i] - \displaystyle\sum_{j=1}^{i-1} a_j^{(i-1)} r_{ss}[i-j]}{\mathcal{E}^{(i-1)}} = k_i.$$

11.30. As discussed in Section 11.7.3, Burg (1975) proposed computing the k parameters so as to minimize the sum of the forward and backward prediction errors at the i^{th} stage of the lattice filter; i.e.,

$$\mathcal{B}^{(i)} = \sum_{n=i}^{M} \left[(e^{(i)}[n])^2 + (\tilde{e}^{(i)}[n])^2\right] \qquad \text{(P11.30-1)}$$

where the sum is over the fixed interval $i \le n \le M$.

(a) Substitute the lattice filter signals $e^{(i)}[n] = e^{(i-1)}[n] - k_i \tilde{e}^{(i-1)}[n-1]$ and $\tilde{e}^{(i)}[n] = \tilde{e}^{(i-1)}[n-1] - k_i e^{(i-1)}[n]$ into (P11.30-1) and show that the value of k_i that minimizes $\mathcal{B}^{(i)}$ is

$$k_i^B = \frac{2\displaystyle\sum_{n=i}^{M} e^{(i-1)}[n]\tilde{e}^{(i-1)}[n-1]}{\left\{\displaystyle\sum_{n=i}^{M} (e^{(i-1)}[n])^2 + \sum_{n=i}^{M} (\tilde{e}^{(i-1)}[n-1])^2\right\}}. \qquad \text{(P11.30-2)}$$

(b) Prove that $-1 < k_i^B < 1$.

Hint: Consider the expression $\displaystyle\sum_{n=i}^{M} (x[n] \pm y[n])^2 > 0$ where $x[n]$ and $y[n]$ are two distinct sequences.

(c) Given a set of Burg coefficients k_i^B, $i = 1, 2, \ldots, p$, how would you obtain the coefficients of the corresponding prediction error filter $A^{(p)}(z)$?