

# SPECTRUM ESTIMATION

---

# 8

## 8.1 INTRODUCTION

In this chapter we consider the problem of estimating the power spectral density of a wide-sense stationary random process. As discussed in Chapter 3, the power spectrum is the Fourier transform of the autocorrelation sequence. Therefore, estimating the power spectrum is equivalent to estimating the autocorrelation. For an autocorrelation ergodic process, recall that

$$\lim_{N \rightarrow \infty} \left\{ \frac{1}{2N+1} \sum_{n=-N}^N x(n+k)x^*(n) \right\} = r_x(k) \quad (8.1)$$

Thus, if  $x(n)$  is known for all  $n$ , estimating the power spectrum is straightforward, in theory, since all that must be done is to determine the autocorrelation sequence  $r_x(k)$  using Eq. (8.1), and then compute its Fourier transform. However, there are two difficulties with this approach that make spectrum estimation both an interesting and a challenging problem. First, the amount of data that one has to work with is never unlimited and, in many cases, it may be very small. Such a limitation may be an inherent characteristic of the data collection process as is the case, for example, in the analysis of seismic data from an earthquake in which the signal persists for only a short period of time. It is also possible, however, that a limited data set is imposed by the requirement that the spectral characteristics of the process remain constant over the duration of the data record. In speech, for example, a stationarity requirement will restrict the length of time over which the signal may be assumed to be approximately stationary to a few milliseconds or less. The second difficulty is that the data is often corrupted by noise or contaminated with an interfering signal. Thus, spectrum estimation is a problem that involves estimating  $P_x(e^{j\omega})$  from a finite number of noisy measurements of  $x(n)$ . In some applications, however, estimating the power spectrum may be facilitated by having prior knowledge about how the process is generated. It may be known, for example, that  $x(n)$  is an autoregressive process or that it consists of one or more sinusoids in noise. This type of information may then allow one to parametrically estimate the power spectrum or, perhaps, to extrapolate the data or its autocorrelation in order to improve the performance of a spectrum estimation algorithm.

Spectrum estimation is a problem that is important in a variety of different fields and applications. It was shown in Chapter 7, for example, that the frequency response of a

noncausal Wiener smoothing filter is

$$H(e^{j\omega}) = \frac{P_d(e^{j\omega})}{P_d(e^{j\omega}) + P_v(e^{j\omega})}$$

where  $P_d(e^{j\omega})$  is the power spectrum of  $d(n)$ , the desired output of the Wiener filter, and  $P_v(e^{j\omega})$  is the power spectrum of the noise,  $v(n)$ . Therefore, before a Wiener smoothing filter can be designed and implemented, the power spectrum of both  $d(n)$  and  $v(n)$  must be determined. Since these power spectral densities are not generally known a priori, one is faced with the problem of estimating them from measurements. Another application in which spectrum estimation plays an important role is signal detection and tracking. Suppose, for example, that a sonar array is placed on the ocean floor to listen for the narrow-band acoustic signals that are generated by the rotating machinery or propellers of a ship. Once a narrow-band signal is detected, the problem of interest is to estimate its center frequency in order to determine the ships direction or velocity. Since these narrow-band signals are typically recorded in a very noisy environment, signal detection and frequency estimation are nontrivial problems that require robust, high-resolution spectrum estimation techniques. Other applications of spectrum estimation include harmonic analysis and prediction, time series extrapolation and interpolation, spectral smoothing, bandwidth compression, beam-forming and direction finding [25,34].

The approaches for spectrum estimation may be generally categorized into one of two classes. The first includes the *classical* or *nonparametric* methods that begin by estimating the autocorrelation sequence  $r_x(k)$  from a given set of data. The power spectrum is then estimated by Fourier transforming the estimated autocorrelation sequence. The second class includes the nonclassical or parametric approaches, which are based on using a model for the process in order to estimate the power spectrum. For example, if it is known that  $x(n)$  is a  $p$ th-order autoregressive process, then measured values of  $x(n)$  may be used to estimate the parameters of the all-pole model,  $a_p(k)$ , and these estimated model parameters,  $\hat{a}_p(k)$ , may then, in turn, be used to estimate the power spectrum as follows:

$$\hat{P}_x(e^{j\omega}) = \frac{1}{\left| \sum_{k=0}^p \hat{a}_p(k) e^{-jk\omega} \right|^2}$$

We begin this chapter with the classical or nonparametric spectrum estimation techniques. These methods, which are described in Section 8.2, include the periodogram, the modified periodogram, Bartlett's method, Welch's method, and the Blackman-Tukey method. The minimum variance method is then considered in Section 8.3. This technique involves the design of a narrow-band filter bank to generate a set of narrow-band random processes. The power spectrum at the center frequency of each bandpass filter is then estimated by measuring the power in the narrow-band process and dividing by the filter bandwidth. Next, in Section 8.4, the maximum entropy method (MEM) is presented, and it is shown that MEM is equivalent to spectrum estimation using an all-pole model. Then, in Section 8.5, we look at power spectrum estimation techniques that are based on a parametric model for the data. These models include moving average (MA), autoregressive (AR), and autoregressive moving average (ARMA). Next, in Section 8.6, we consider frequency estimation algorithms for harmonic processes that consist of a sum of sinusoids or complex exponentials in noise. These methods, sometimes referred to as noise sub-

space methods, include the Pisarenko harmonic decomposition, MUSIC, the eigenvector method, and the minimum norm algorithm. Finally, in Section 8.7, we look at principal components frequency estimation. This approach, which also assumes that the process is harmonic, forms a low-rank approximation to the autocorrelation matrix, which is then incorporated into a spectrum estimation algorithm such as the minimum variance method or MEM.

## **8.2 NONPARAMETRIC METHODS**

In this section, we consider nonparametric techniques of spectrum estimation. These methods are based on the idea of estimating the autocorrelation sequence of a random process from a set of measured data, and then taking the Fourier transform to obtain an estimate of the power spectrum. We begin with the *periodogram*, a nonparametric method first introduced by Schuster in 1898 in his study of periodicities in sunspot numbers [47,49]. As we will see, although the periodogram is easy to compute, it is limited in its ability to produce an accurate estimate of the power spectrum, particularly for short data records. We will then examine a number of modifications to the periodogram that have been proposed to improve its statistical properties. These include the modified periodogram, Bartlett's method, Welch's method, and the Blackman-Tukey method.

### **8.2.1 The Periodogram**

The power spectrum of a wide-sense stationary random process is the Fourier transform of the autocorrelation sequence,

$$P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_x(k)e^{-jk\omega}$$

Therefore, spectrum estimation is, in some sense, an autocorrelation estimation problem. For an *autocorrelation ergodic* process and an unlimited amount of data, the autocorrelation sequence may, in theory, be determined using the time-average

$$r_x(k) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x(n+k)x^*(n) \quad (8.2)$$

However, if  $x(n)$  is only measured over a finite interval, say  $n = 0, 1, \dots, N-1$ , then the autocorrelation sequence must be estimated using, for example, Eq. (8.2) with a finite sum,

$$\hat{r}_x(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n+k)x^*(n) \quad (8.3)$$

In order to ensure that the values of  $x(n)$  that fall outside the interval  $[0, N-1]$  are excluded from the sum, Eq. (8.3) will be rewritten as follows:

$$\hat{r}_x(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n+k)x^*(n) ; \quad k = 0, 1, \dots, N-1 \quad (8.4)$$

with the values of  $\hat{r}_x(k)$  for  $k < 0$  defined using conjugate symmetry,  $\hat{r}_x(-k) = \hat{r}_x^*(k)$ , and with  $\hat{r}_x(k)$  set equal to zero for  $|k| \geq N$ . Taking the discrete-time Fourier transform of

$\hat{r}_x(k)$  leads to an estimate of the power spectrum known as the *periodogram*,

$$\hat{P}_{per}(e^{j\omega}) = \sum_{k=-N+1}^{N-1} \hat{r}_x(k) e^{-jk\omega} \quad (8.5)$$

Although defined in terms of the estimated autocorrelation sequence  $\hat{r}_x(k)$ , it will be more convenient to express the periodogram directly in terms of the process  $x(n)$ . This may be done as follows. Let  $x_N(n)$  be the finite length signal of length  $N$  that is equal to  $x(n)$  over the interval  $[0, N - 1]$ , and is zero otherwise,

$$x_N(n) = \begin{cases} x(n) & ; \quad 0 \leq n < N \\ 0 & ; \quad \text{otherwise} \end{cases} \quad (8.6)$$

Thus,  $x_N(n)$  is the product of  $x(n)$  with a rectangular window  $w_R(n)$ ,

$$x_N(n) = w_R(n)x(n) \quad (8.7)$$

In terms of  $x_N(n)$ , the estimated autocorrelation sequence may be written as follows:

$$\hat{r}_x(k) = \frac{1}{N} \sum_{n=-\infty}^{\infty} x_N(n+k)x_N^*(n) = \frac{1}{N} x_N(k) * x_N^*(-k) \quad (8.8)$$

Taking the Fourier transform and using the convolution theorem, the periodogram becomes

$$\hat{P}_{per}(e^{j\omega}) = \frac{1}{N} X_N(e^{j\omega}) X_N^*(e^{j\omega}) = \frac{1}{N} |X_N(e^{j\omega})|^2 \quad (8.9)$$

where  $X_N(e^{j\omega})$  is the discrete-time Fourier transform of the  $N$ -point data sequence  $x_N(n)$ ,

$$X_N(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x_N(n) e^{-jn\omega} = \sum_{n=0}^{N-1} x(n) e^{-jn\omega} \quad (8.10)$$

Thus, the periodogram is proportional to the squared magnitude of the DTFT of  $x_N(n)$ , and may be easily computed using a DFT as follows:

$$x_N(n) \xrightarrow{\text{DFT}} X_N(k) \longrightarrow \frac{1}{N} |X_N(k)|^2 = \hat{P}_{per}(e^{j2\pi k/N})$$

A MATLAB program for the periodogram is given in Fig. 8.1.

### The Periodogram

```
function Px = periodogram(x,n1,n2)
%
x = x(:);
if nargin == 1
    n1 = 1; n2 = length(x); end;
Px = abs(fft(x(n1:n2),1024)).^2/(n2-n1+1);
Px(1)=Px(2);
end;
```

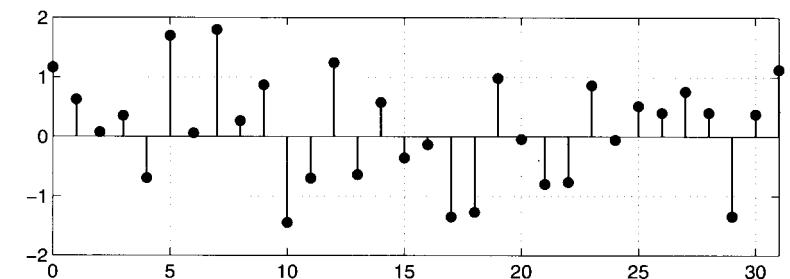
**Figure 8.1** A MATLAB program for computing the periodogram of  $x(n)$ .

**Example 8.2.1 Periodogram of White Noise**

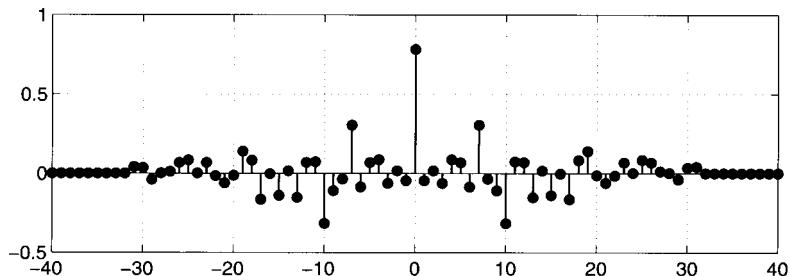
If  $x(n)$  is white noise with a variance of  $\sigma_x^2$ , then  $r_x(k) = \sigma_x^2\delta(k)$  and the power spectrum is a constant,

$$P_x(e^{j\omega}) = \sigma_x^2$$

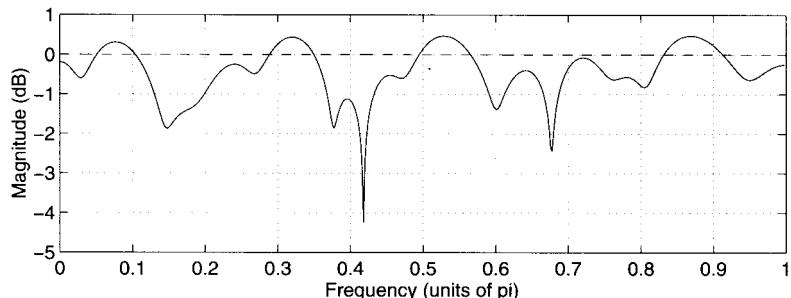
Shown in Fig. 8.2a is a sample realization of unit variance white noise of length  $N = 32$ . The autocorrelation sequence that is estimated using Eq. (8.3) is shown in Fig. 8.2b. Note that



(a)



(b)



(c)

**Figure 8.2** (a) A sample realization of unit variance white noise of length  $N = 32$ . (b) The estimated autocorrelation sequence. (c) The periodogram along with the true power spectrum,  $P_x(e^{j\omega}) = 1$ , which is indicated by the dotted line.

although  $\hat{r}_x(k)$  is zero for  $|k| \geq 32$ , it is nonzero for all other values of  $k$ . The periodogram, which is the Fourier transform of  $\hat{r}_x(k)$ , is shown in Fig. 8.2c along with the true power spectrum. Note that although the periodogram is approximately equal to  $P_x(e^{j\omega})$ , on the average, we see that there is a considerable amount of variation in  $\hat{P}_{per}(e^{j\omega})$  as  $\omega$  varies. As we will see, such variations are not uncommon with the periodogram.

As we now show, the periodogram has an interesting interpretation in terms of filter banks. Let  $h_i(n)$  be an FIR filter of length  $N$  that is defined as follows:

$$h_i(n) = \frac{1}{N} e^{jn\omega_i} w_R(n) = \begin{cases} \frac{1}{N} e^{jn\omega_i} & ; \quad 0 \leq n < N \\ 0 & ; \quad \text{otherwise} \end{cases} \quad (8.11)$$

The frequency response of this filter is

$$H_i(e^{j\omega}) = \sum_{n=0}^{N-1} h_i(n) e^{-jn\omega} = e^{-j(\omega-\omega_i)(N-1)/2} \frac{\sin[N(\omega-\omega_i)/2]}{N \sin[(\omega-\omega_i)/2]} \quad (8.12)$$

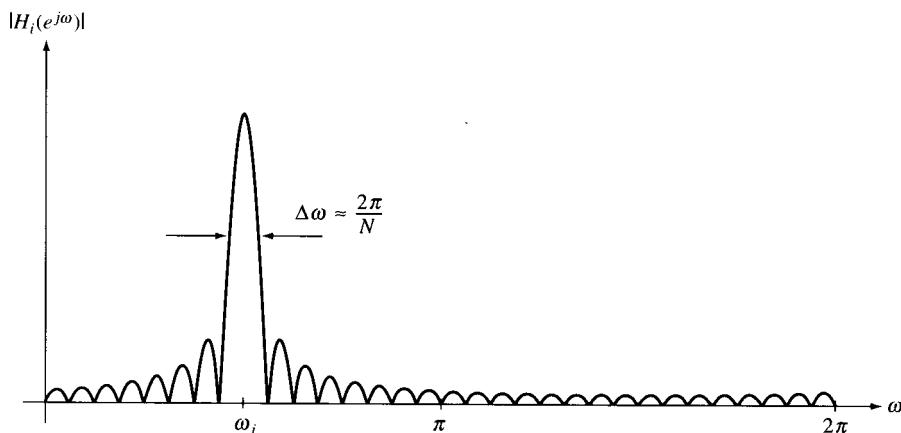
which, as illustrated in Fig. 8.3, is a bandpass filter with a center frequency  $\omega_i$ , and a bandwidth that is approximately equal to  $\Delta\omega = 2\pi/N$ . If a WSS random process  $x(n)$  is filtered with  $h_i(n)$ , then the output process is

$$y_i(n) = x(n) * h_i(n) = \sum_{k=n-N+1}^n x(k) h_i(n-k) = \frac{1}{N} \sum_{k=n-N+1}^n x(k) e^{j(n-k)\omega_i} \quad (8.13)$$

Since  $|H_i(e^{j\omega})|_{\omega=\omega_i} = 1$ , then the power spectrum of  $x(n)$  and  $y(n)$  are equal at frequency  $\omega_i$ ,

$$P_x(e^{j\omega_i}) = P_y(e^{j\omega_i})$$

Furthermore, if the bandwidth of the filter is small enough so that the power spectrum of  $x(n)$  may be assumed to be approximately constant over the passband of the filter, then the



**Figure 8.3** The magnitude of the frequency response of the bandpass filter used in the filter bank interpretation of the periodogram.

power in  $y_i(n)$  will be approximately<sup>1</sup>

$$E\{|y_i(n)|^2\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x(e^{j\omega}) |H_i(e^{j\omega})|^2 d\omega \approx \frac{\Delta\omega}{2\pi} P_x(e^{j\omega_i}) = \frac{1}{N} P_x(e^{j\omega_i})$$

and, therefore,

$$P_x(e^{j\omega_i}) \approx N E\{|y_i(n)|^2\} \quad (8.14)$$

Thus, if we are able to estimate the power in  $y_i(n)$ , then the power spectrum at frequency  $\omega_i$  may be estimated as follows:

$$\hat{P}_x(e^{j\omega_i}) = N \hat{E}\{|y_i(n)|^2\} \quad (8.15)$$

One simple yet very crude way to estimate the power is to use a one-point sample average,

$$\hat{E}\{|y_i(n)|^2\} = |y_i(N-1)|^2$$

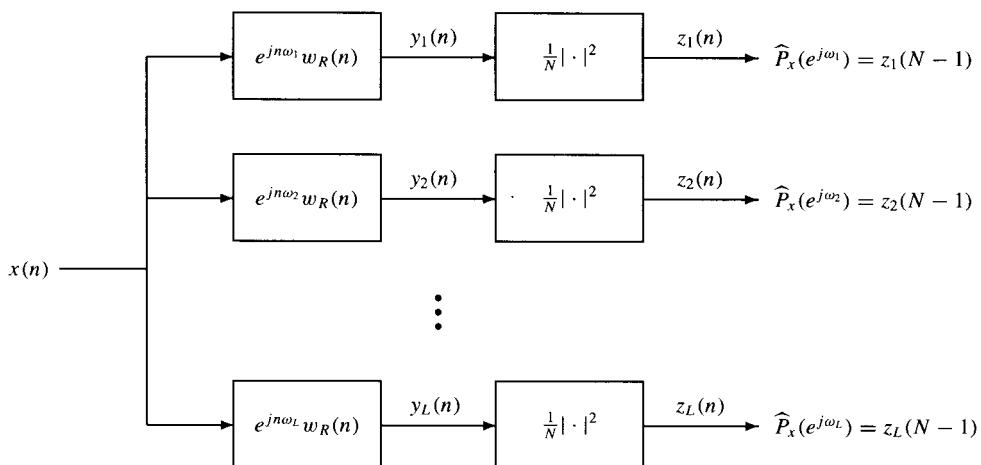
From Eq. (8.13) we see that this is equivalent to

$$|y_i(N-1)|^2 = \frac{1}{N^2} \left| \sum_{k=0}^{N-1} x(k) e^{-jk\omega_i} \right|^2 \quad (8.16)$$

Therefore,

$$\hat{P}_x(e^{j\omega_i}) = N |y_i(N-1)|^2 = \frac{1}{N} \left| \sum_{k=0}^{N-1} x(k) e^{-jk\omega_i} \right|^2 \quad (8.17)$$

which is equivalent to the periodogram. Thus, the periodogram may be viewed as the estimate of the power spectrum that is formed using a filter bank of bandpass filters as illustrated in Fig. 8.4, with  $\hat{P}_{per}(e^{j\omega_i})$  being derived from a one-point sample average of the power in the filtered process  $y_i(n)$ . Of course, the periodogram has such a filter bank “built into it” so that it is not necessary to implement the filter bank. However, in Section 8.3



**Figure 8.4** A filter bank interpretation of the periodogram.

<sup>1</sup>See the discussion leading up to Eq. (3.96) on p. 103.

we will consider a generalization of this filter bank idea that allows the filters to be *data dependent* and to have a frequency response that varies with the center frequency  $\omega_i$ .

### 8.2.2 Performance of the Periodogram

In the previous section, it was shown that the periodogram is proportional to the squared magnitude of the DTFT of the finite length sequence  $x_N(n)$ . Therefore, from a computational point of view, the periodogram is simple to evaluate. In this section, we look at the performance of the periodogram. Ideally, as the length of the data record increases, the periodogram should converge to the power spectrum of the process,  $P_x(e^{j\omega})$ . However, we must be careful when discussing the convergence of the periodogram to  $P_x(e^{j\omega})$ . Specifically, since  $\hat{P}_{per}(e^{j\omega})$  is a function of the random variables  $x(0), \dots, x(N)$ , it is necessary to consider convergence in a statistical sense. Therefore, in this section, we will look at mean-square convergence of the periodogram [33,40], i.e., we will be interested in whether or not

$$\lim_{N \rightarrow \infty} E \left\{ \left[ \hat{P}_{per}(e^{j\omega}) - P_x(e^{j\omega}) \right]^2 \right\} = 0$$

In order for the periodogram to be mean-square convergent, it is necessary that it be asymptotically unbiased

$$\lim_{N \rightarrow \infty} E \left\{ \hat{P}_{per}(e^{j\omega}) \right\} = P_x(e^{j\omega}) \quad (8.18)$$

and have a variance that goes to zero as the data record length  $N$  goes to infinity,

$$\lim_{N \rightarrow \infty} \text{Var} \left\{ \hat{P}_{per}(e^{j\omega}) \right\} = 0 \quad (8.19)$$

In other words,  $\hat{P}_{per}(e^{j\omega})$  must be a *consistent* estimate of the power spectrum (see Section 3.2.8). First, we consider the bias of the periodogram.

**Periodogram Bias.** To compute the bias of the periodogram, we begin by finding the expected value of  $\hat{r}_x(k)$ . From Eq. (8.4) it follows that the expected value of  $\hat{r}_x(k)$  for  $k = 0, 1, \dots, N-1$  is

$$E \left\{ \hat{r}_x(k) \right\} = \frac{1}{N} \sum_{n=0}^{N-1-k} E \left\{ x(n+k)x^*(n) \right\} = \frac{1}{N} \sum_{n=0}^{N-1-k} r_x(k) = \frac{N-k}{N} r_x(k)$$

and, for  $k \geq N$ , the expected value is zero. Using the conjugate symmetry of  $\hat{r}_x(k)$  we have

$$E \left\{ \hat{r}_x(k) \right\} = w_B(k) r_x(k) \quad (8.20)$$

where

$$w_B(k) = \begin{cases} \frac{N-|k|}{N} & ; \quad |k| \leq N \\ 0 & ; \quad |k| > N \end{cases} \quad (8.21)$$

is a Bartlett (triangular) window.<sup>2</sup> Therefore,  $\hat{r}_x(k)$  is a biased estimate of the autocorrelation.

<sup>2</sup>Since the Bartlett window is applied to the autocorrelation sequence,  $w_B(k)$  is referred to as a *lag window*. This is in contrast to a *data window* that is applied to  $x(n)$ .

Using Eq. (8.20) it follows that the expected value of the periodogram is

$$\begin{aligned} E \left\{ \hat{P}_{per}(e^{j\omega}) \right\} &= E \left\{ \sum_{k=-N+1}^{N-1} \hat{r}_x(k) e^{-jk\omega} \right\} = \sum_{k=-N+1}^{N-1} E \left\{ \hat{r}_x(k) \right\} e^{-jk\omega} \\ &= \sum_{k=-\infty}^{\infty} r_x(k) w_B(k) e^{-jk\omega} \end{aligned} \quad (8.22)$$

Since  $E \left\{ \hat{P}_{per}(e^{j\omega}) \right\}$  is the Fourier transform of the product  $r_x(k)w_B(k)$ , using the frequency convolution theorem we have

$$E \left\{ \hat{P}_{per}(e^{j\omega}) \right\} = \frac{1}{2\pi} P_x(e^{j\omega}) * W_B(e^{j\omega}) \quad (8.23)$$

where  $W_B(e^{j\omega})$  is the Fourier transform of the Bartlett window,  $w_B(k)$ ,

$$W_B(e^{j\omega}) = \frac{1}{N} \left[ \frac{\sin(N\omega/2)}{\sin(\omega/2)} \right]^2 \quad (8.24)$$

Thus, the expected value of the periodogram is the convolution of the power spectrum  $P_x(e^{j\omega})$  with the Fourier transform of a Bartlett window and, therefore, the periodogram is a biased estimate. However, since  $W_B(e^{j\omega})$  converges to an impulse as  $N$  goes to infinity, the periodogram is *asymptotically unbiased*

$$\lim_{N \rightarrow \infty} E \left\{ \hat{P}_{per}(e^{j\omega}) \right\} = P_x(e^{j\omega}) \quad (8.25)$$

To illustrate the effect of the lag window,  $w_B(k)$ , on the expected value of the periodogram, consider a random process consisting of a random phase sinusoid in white noise

$$x(n) = A \sin(n\omega + \phi) + v(n)$$

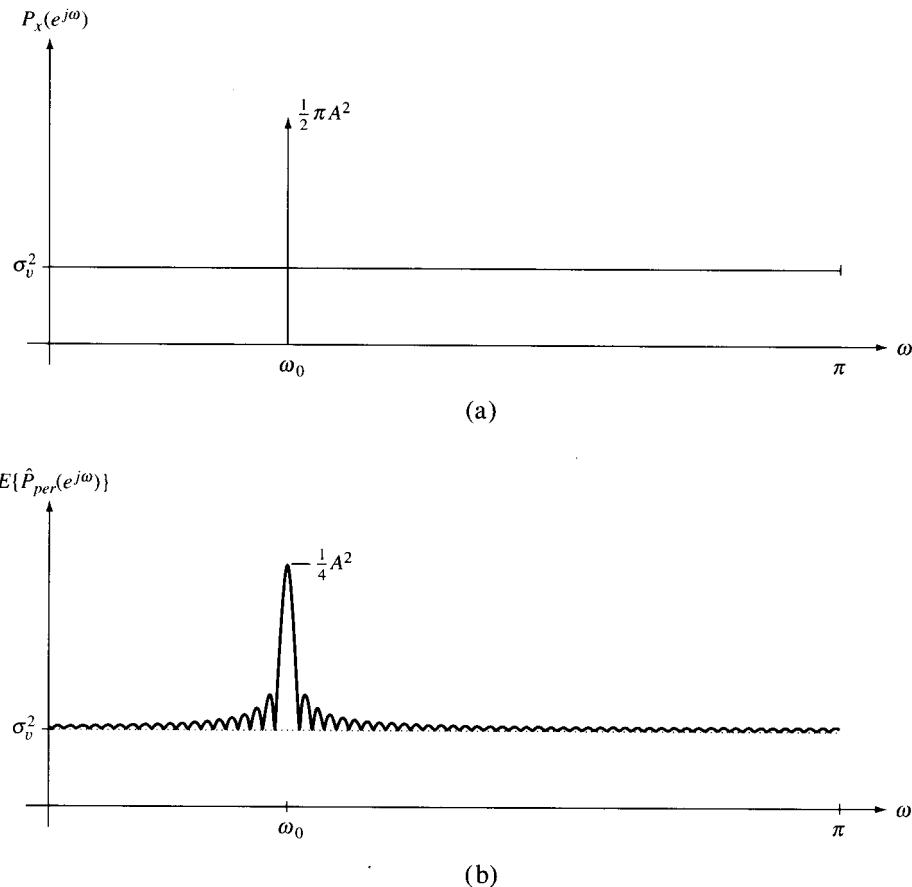
where  $\phi$  is a random variable that is uniformly distributed over the interval  $[-\pi, \pi]$ , and  $v(n)$  is white noise with a variance  $\sigma_v^2$ . The power spectrum of  $x(n)$  is

$$P_x(e^{j\omega}) = \sigma_v^2 + \frac{1}{2}\pi A^2 [u_0(\omega - \omega_0) + u_0(\omega + \omega_0)]$$

Therefore, it follows from Eq. (8.23) that the expected value of the periodogram is

$$\begin{aligned} E \left\{ \hat{P}_{per}(e^{j\omega}) \right\} &= \frac{1}{2\pi} P_x(e^{j\omega}) * W_B(e^{j\omega}) \\ &= \sigma_v^2 + \frac{1}{4} A^2 [W_B(e^{j(\omega-\omega_0)}) + W_B(e^{j(\omega+\omega_0)})] \end{aligned} \quad (8.26)$$

The power spectrum  $P_x(e^{j\omega})$  and the expected value of the periodogram are shown in Fig. 8.5 for  $N = 64$ . There are two effects that should be noted in this example. First, is the spectral smoothing that is produced by  $W_B(e^{j\omega})$ , which leads to a spreading of the power in the sinusoid over a band of frequencies that has a bandwidth of approximately  $4\pi/N$ . The second effect is the power leakage through the sidelobes of the window, which creates secondary spectral peaks at frequencies  $\omega_k \approx \omega_0 \pm \frac{2\pi}{N} k$ . As we will see in Section 8.2.3, it is possible for these sidelobes to mask low-level narrowband components.



**Figure 8.5** (a) The power spectrum of a single sinusoid in white noise and (b) the expected value of the periodogram.

---

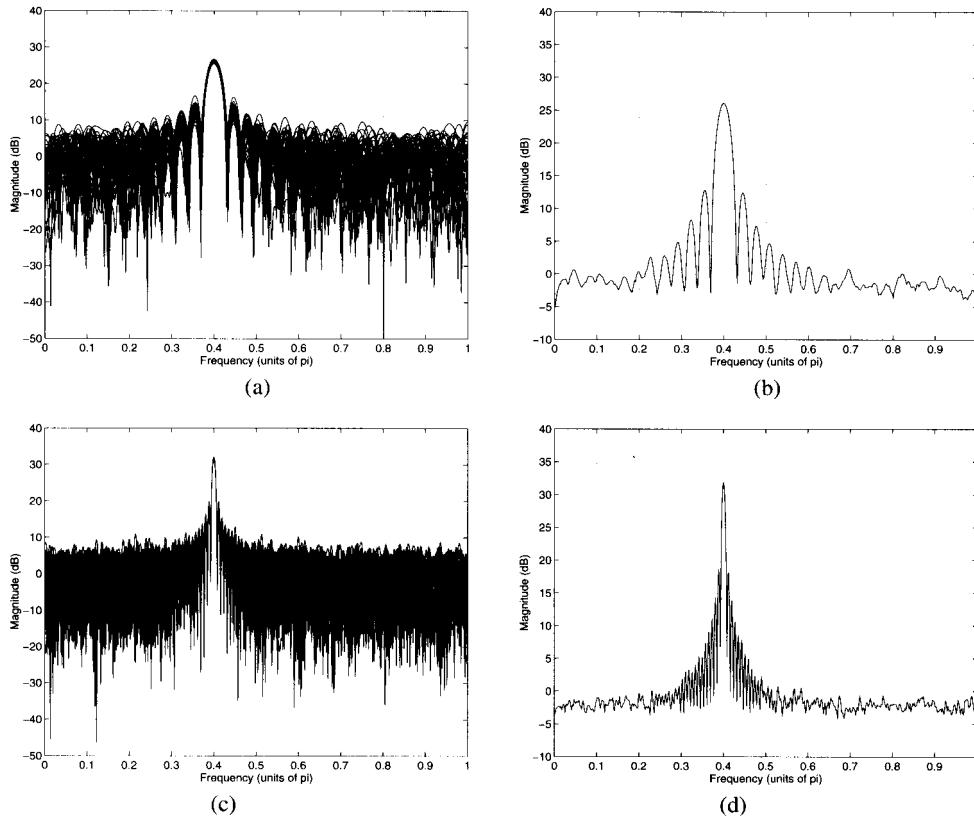
### Example 8.2.2 Periodogram of a Sinusoid in Noise

Let  $x(n)$  be a wide-sense stationary process consisting of a random phase sinusoid in unit variance white noise

$$x(n) = A \sin(n\omega_0 + \phi) + v(n)$$

With  $A = 5$ ,  $\omega_0 = 0.4\pi$ , and  $N = 64$ , fifty different realizations of this process were generated and the periodogram of each was computed. Shown in Fig. 8.6a is an overlay of the 50 periodograms. Note that although each periodogram has a peak at approximately  $\omega = 0.4\pi$ , there is a considerable amount of variation from one periodogram to the next. In Fig. 8.6b the average of all 50 periodograms is shown. This average is approximately equal to the expected value given in Eq. (8.26). By increasing the number of data values to  $N = 256$ , we obtain the periodograms shown in Fig. 8.6c and d. Note that with the additional data, the power in the sinusoid is spread out over a much narrower band of frequencies.

---



**Figure 8.6** The periodogram of a sinusoid in white noise. (a) Overlay plot of 50 periodograms using  $N = 64$  data values and (b) the periodogram average. (c) Overlay plot of 50 periodograms using  $N = 256$  data values and (d) the periodogram average.

In addition to biasing the periodogram, the smoothing that is introduced by the Bartlett window also limits the ability of the periodogram to resolve closely-spaced narrowband components in  $x(n)$ . Consider, for example, a random process consisting of two sinusoids in white noise

$$x(n) = A_1 \sin(n\omega_1 + \phi_1) + A_2 \sin(n\omega_2 + \phi_2) + v(n)$$

where  $\phi_1$  and  $\phi_2$  are uncorrelated uniformly distributed random variables and where  $v(n)$  is white noise with a variance of  $\sigma_v^2$ . The power spectrum of  $x(n)$  is

$$P_x(e^{j\omega}) = \sigma_v^2 + \frac{1}{2}\pi A_1^2 [u_0(\omega - \omega_1) + u_0(\omega + \omega_1)] + \frac{1}{2}\pi A_2^2 [u_0(\omega - \omega_2) + u_0(\omega + \omega_2)]$$

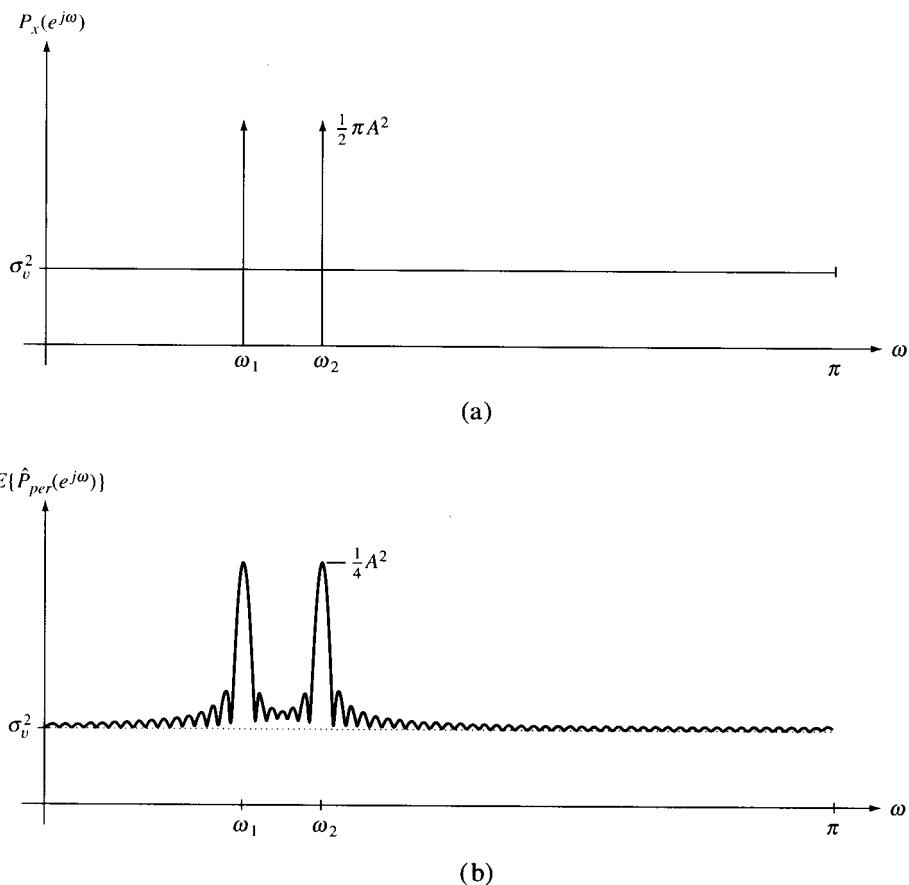
and the expected value of the periodogram is

$$\begin{aligned} E \left\{ \hat{P}_{per}(e^{j\omega}) \right\} &= \frac{1}{2\pi} P_x(e^{j\omega}) * W_B(e^{j\omega}) \\ &= \sigma_v^2 + \frac{1}{4} A_1^2 [W_B(e^{j(\omega-\omega_1)}) + W_B(e^{j(\omega+\omega_1)})] \\ &\quad + \frac{1}{4} A_2^2 [W_B(e^{j(\omega-\omega_2)}) + W_B(e^{j(\omega+\omega_2)})] \end{aligned} \quad (8.27)$$

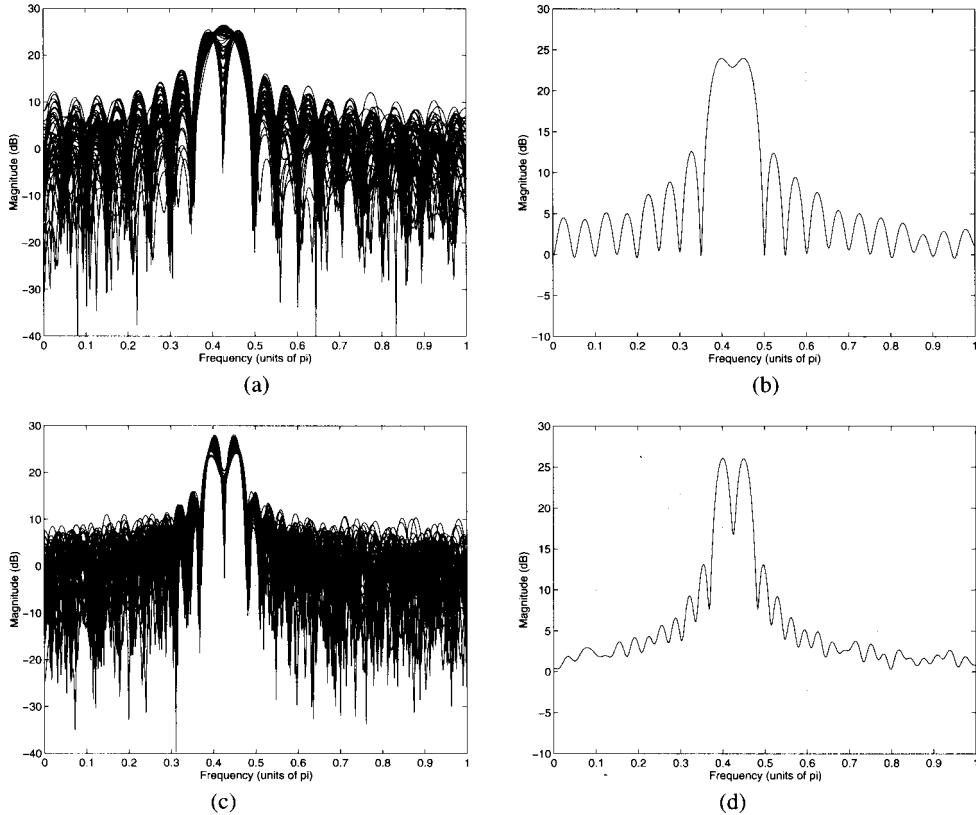
which is shown in Fig. 8.7 for  $A_1 = A_2$  and  $N = 64$ . Since the width of the main lobe of  $W_B(e^{j\omega})$  increases as the data record length decreases, for a given data record length,  $N$ , there is a limit on how closely two sinusoids or two narrowband processes may be located before they can no longer be resolved. One way to define this resolution limit is to set  $\Delta\omega$  equal to the width of the main lobe of the spectral window,  $W_B(e^{j\omega})$ , at its “half-power” or 6 dB point. For the Bartlett window,  $\Delta\omega = 0.89(2\pi/N)$ , which means that the resolution of the periodogram is

$$\text{Res} [\hat{P}_{per}(e^{j\omega})] = 0.89 \frac{2\pi}{N} \quad (8.28)$$

It is important to note, however, that Eq. (8.28) is nothing more than a *rule of thumb* that should only be used as a guideline in determining the amount of data that is necessary for a given resolution. There is nothing sacred, for example, with the proportionality constant  $0.89(2\pi)$ . On the other hand, what is important is the fact that the resolution is inversely proportional to the amount of data,  $N$ . Nevertheless, although the definition of  $\Delta\omega$  given in Eq. (8.28) is somewhat arbitrary, one generally finds that it is difficult to resolve details in the spectrum that are much finer than this.



**Figure 8.7** (a) The power spectrum of two sinusoids in white noise and (b) the expected value of the periodogram.



**Figure 8.8** The periodogram of two sinusoids in white noise with  $\omega_1 = 0.4\pi$  and  $\omega_2 = 0.45\pi$ . (a) Overlay plot of 50 periodograms using  $N = 40$  data values and (b) the ensemble average. (c) Overlay plot of 50 periodograms using  $N = 64$  data values and (d) the ensemble average.

### Example 8.2.3 Periodogram Resolution

Let  $x(n)$  be a random process consisting of two equal amplitude sinusoids in unit variance white noise

$$x(n) = A \sin(n\omega_1 + \phi_1) + A \sin(n\omega_2 + \phi_2) + v(n)$$

where  $\omega_1 = 0.4\pi$ ,  $\omega_2 = 0.45\pi$ , and  $A = 5$ . Using Eq. (8.28), with  $\Delta\omega = 0.05\pi$  we see that in order to resolve the two narrowband components, a data record length on the order of  $N = 36$  is required. Using  $N = 40$ , an overlay of 50 periodograms is shown in Fig. 8.8a. It is clear from this figure that it is not always possible to resolve the two sinusoidal components when  $N = 40$ . The average of the 50 periodograms, shown in 8.8b, illustrates the overlap of the two Bartlett windows. However, as shown in Figures 8.8c and d, if  $N = 64$  then the two sinusoids are clearly resolved.

**Variance of the Periodogram.** We have seen that the periodogram is an asymptotically unbiased estimate of the power spectrum. In order for it to be a consistent estimate, it is necessary that the variance go to zero as  $N \rightarrow \infty$ . Unfortunately, it is difficult to evaluate the variance of the periodogram for an arbitrary process  $x(n)$  since the variance depends

on the fourth-order moments of the process. However, as we show next, the variance may be evaluated in the special case of white Gaussian noise.

Let  $x(n)$  be a Gaussian white noise process with variance  $\sigma_x^2$ . Using Eq. (8.9) the periodogram may be expressed as follows:

$$\begin{aligned}\hat{P}_{per}(e^{j\omega}) &= \frac{1}{N} \left| \sum_{k=0}^{N-1} x(k)e^{-jk\omega} \right|^2 = \frac{1}{N} \left\{ \sum_{k=0}^{N-1} x(k)e^{-jk\omega} \right\} \left\{ \sum_{l=0}^{N-1} x^*(l)e^{jl\omega} \right\} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} x(k)x^*(l)e^{-j(k-l)\omega}\end{aligned}\quad (8.29)$$

Therefore, the second-order moment of the periodogram is

$$\begin{aligned}E \left\{ \hat{P}_{per}(e^{j\omega_1}) \hat{P}_{per}(e^{j\omega_2}) \right\} &= \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E \{x(k)x^*(l)x(m)x^*(n)\} e^{-j(k-l)\omega_1} e^{-j(m-n)\omega_2}\end{aligned}\quad (8.30)$$

which depends on the fourth-order moments of  $x(n)$ . Since  $x(n)$  is Gaussian, we may use the moment factoring theorem to simplify these moments [17,44]. For complex Gaussian random variables, the moment factoring theorem is<sup>3</sup>

$$\begin{aligned}E \{x(k)x^*(l)x(m)x^*(n)\} &= E \{x(k)x^*(l)\} E \{x(m)x^*(n)\} \\ &\quad + E \{x(k)x^*(n)\} E \{x(m)x^*(l)\}\end{aligned}\quad (8.31)$$

Substituting Eq. (8.31) into Eq. (8.30), the second-order moment of the periodogram becomes a sum of two terms. The first term contains products of  $E \{x(k)x^*(l)\}$  with  $E \{x(m)x^*(n)\}$ . For white noise, these terms are equal to  $\sigma_x^4$  when  $k = l$  and  $m = n$ , and they are equal to zero otherwise. Thus, the first term simplifies to

$$\frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{m=0}^{N-1} \sigma_x^4 = \sigma_x^4 \quad (8.32)$$

The second term, on the other hand, contains products of  $E \{x(k)x^*(n)\}$  with  $E \{x(m)x^*(l)\}$ . Again, for white noise, these terms are equal to  $\sigma_x^4$  when  $k = n$  and  $l = m$ , and they are equal to zero otherwise. Therefore, the second term becomes

$$\begin{aligned}\frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \sigma_x^4 e^{-j(k-l)\omega_1} e^{j(k-l)\omega_2} &= \frac{\sigma_x^4}{N^2} \sum_{k=0}^{N-1} e^{-jk(\omega_1-\omega_2)} \sum_{l=0}^{N-1} e^{jl(\omega_1-\omega_2)} \\ &= \frac{\sigma_x^4}{N^2} \left[ \frac{1 - e^{-jN(\omega_1-\omega_2)}}{1 - e^{-j(\omega_1-\omega_2)}} \right] \left[ \frac{1 - e^{jN(\omega_1-\omega_2)}}{1 - e^{j(\omega_1-\omega_2)}} \right] \\ &= \sigma_x^4 \left[ \frac{\sin N(\omega_1 - \omega_2)/2}{N \sin(\omega_1 - \omega_2)/2} \right]^2\end{aligned}\quad (8.33)$$

<sup>3</sup>Note that this is different from the moment factoring theorem for real Gaussian random variables, which contains three terms instead of two.

Combining Eq. (8.32) and Eq. (8.33) it follows that

$$E \left\{ \hat{P}_{per}(e^{j\omega_1}) \hat{P}_{per}(e^{j\omega_2}) \right\} = \sigma_x^4 \left\{ 1 + \left[ \frac{\sin N(\omega_1 - \omega_2)/2}{N \sin(\omega_1 - \omega_2)/2} \right]^2 \right\} \quad (8.34)$$

Since

$$\begin{aligned} \text{Cov} \left\{ \hat{P}_{per}(e^{j\omega_1}) \hat{P}_{per}(e^{j\omega_2}) \right\} &= E \left\{ \hat{P}_{per}(e^{j\omega_1}) \hat{P}_{per}(e^{j\omega_2}) \right\} \\ &\quad - E \left\{ \hat{P}_{per}(e^{j\omega_1}) \right\} E \left\{ \hat{P}_{per}(e^{j\omega_2}) \right\} \end{aligned}$$

and  $E \left\{ \hat{P}_{per}(e^{j\omega}) \right\} = \sigma_x^2$ , then the covariance of the periodogram is

$$\text{Cov} \left\{ \hat{P}_{per}(e^{j\omega_1}) \hat{P}_{per}(e^{j\omega_2}) \right\} = \sigma_x^4 \left[ \frac{\sin N(\omega_1 - \omega_2)/2}{N \sin(\omega_1 - \omega_2)/2} \right]^2 \quad (8.35)$$

Finally, setting  $\omega_1 = \omega_2$  we have, for the variance,

$$\text{Var} \left\{ \hat{P}_{per}(e^{j\omega}) \right\} = \sigma_x^4 \quad (8.36)$$

Thus, the variance does not go to zero as  $N \rightarrow \infty$ , and the periodogram is *not a consistent estimate* of the power spectrum. In fact, since  $P_x(e^{j\omega}) = \sigma_x^2$  then the variance of the periodogram of white Gaussian noise is proportional to the square of the power spectrum,

$$\text{Var} \left\{ \hat{P}_{per}(e^{j\omega}) \right\} = P_x^2(e^{j\omega}) \quad (8.37)$$

#### **Example 8.2.4 Periodogram of White Noise**

Let  $x(n)$  be white Gaussian noise with

$$P_x(e^{j\omega}) = 1$$

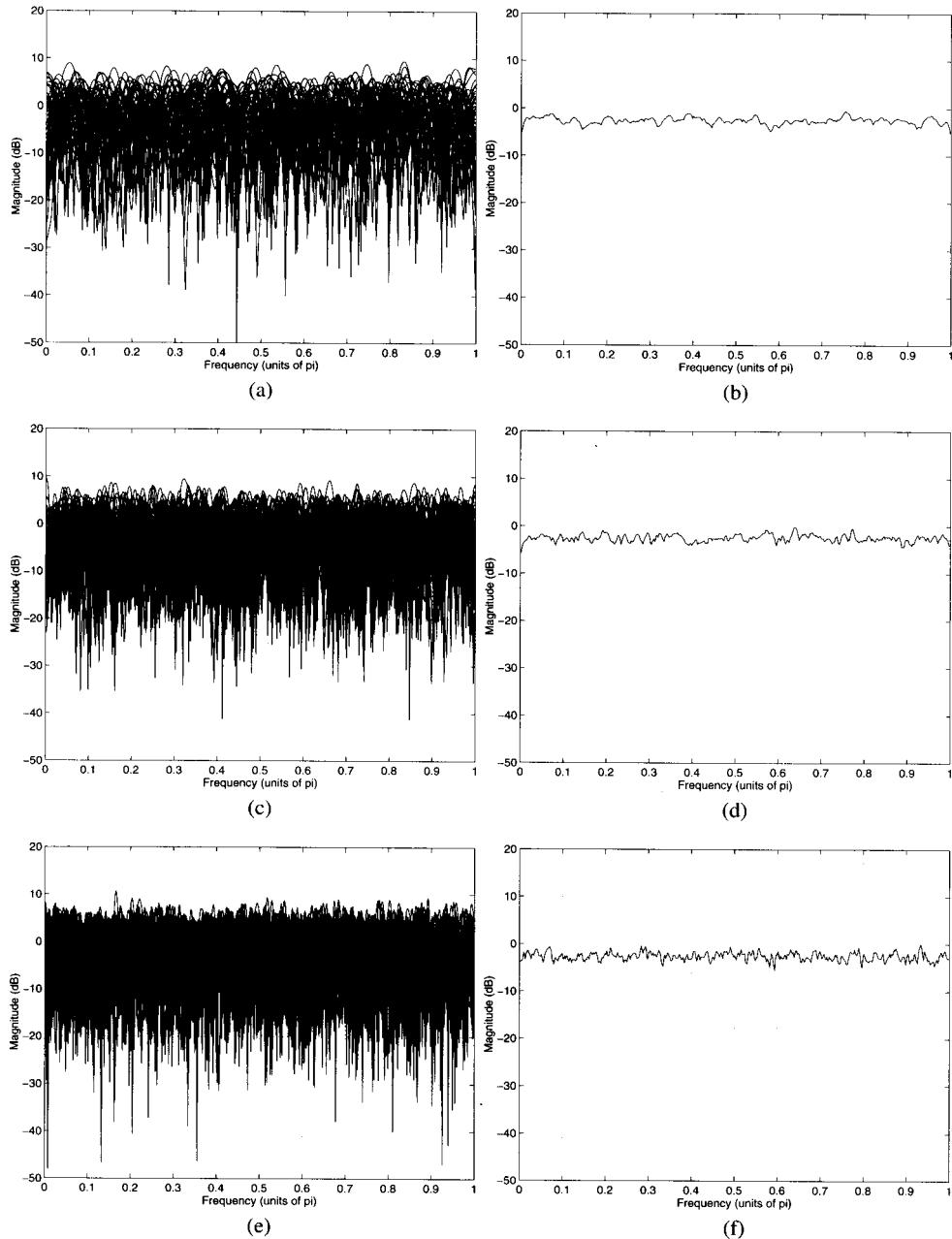
From Eq. (8.23) it follows that the expected value of the periodogram is equal to one,

$$E \left\{ \hat{P}_{per}(e^{j\omega}) \right\} = 1$$

and from Eq. (8.36) it follows that the variance is also equal to one,

$$\text{Var} \left\{ \hat{P}_{per}(e^{j\omega}) \right\} = 1$$

Thus, although the periodogram is unbiased, the variance is equal to a constant that is independent of the data record length,  $N$ . In Fig. 8.9a, c, and e are overlay plots of 50 periodograms of white noise that were generated using data records of length  $N = 64$ , 128, and 256, respectively. In Fig. 8.9b, d, and f, on the other hand, are the periodogram averages. What we observe is that, although the average value of the periodogram is approximately equal to  $\sigma_x^2 = 1$ , the variance does not decrease as the amount of data increases.



**Figure 8.9** The periodogram of unit variance white Gaussian noise. (a) Overlay plot of 50 periodograms with  $N = 64$  data values and (b) the periodogram average. (c) Overlay plot of 50 periodograms with  $N = 128$  data values and (d) the periodogram average. (e) Overlay plot of 50 periodograms with  $N = 256$  data values and (f) the periodogram average.

The analysis given above for the variance of the periodogram assumes that  $x(n)$  is white Gaussian noise. Although the statistical analysis of a nonwhite Gaussian process is much

**Table 8.1 Properties of the Periodogram**

$\hat{P}_{per}(e^{j\omega}) = \frac{1}{N} \left  \sum_{n=0}^{N-1} x(n)e^{-jn\omega} \right ^2$
<i>Bias</i>
$E \{ \hat{P}_{per}(e^{j\omega}) \} = \frac{1}{2\pi} P_x(e^{j\omega}) * W_B(e^{j\omega})$
<i>Resolution</i>
$\Delta\omega = 0.89 \frac{2\pi}{N}$
<i>Variance</i>
$\text{Var} \{ \hat{P}_{per}(e^{j\omega}) \} \approx P_x^2(e^{j\omega})$

### 8.2.3 The Modified Periodogram

In Section 8.2.1 we saw that the periodogram is proportional to the squared magnitude of the Fourier transform of the windowed signal  $x_N(n) = x(n)w_R(n)$ ,

$$\hat{P}_{per}(e^{j\omega}) = \frac{1}{N} |X_N(e^{j\omega})|^2 = \frac{1}{N} \left| \sum_{n=-\infty}^{\infty} x(n)w_R(n)e^{-jn\omega} \right|^2 \quad (8.45)$$

Instead of applying a rectangular window to  $x(n)$ , Eq. (8.45) suggests the possibility of using other data windows. Would there be any benefit, for example, in replacing the rectangular window with a triangular (Bartlett) window? To answer this question, let us examine the effect of the data window on the bias of the periodogram.

Using Eq. (8.45), the expected value of the periodogram is

$$\begin{aligned} E \{ \hat{P}_{per}(e^{j\omega}) \} &= \frac{1}{N} E \left\{ \left[ \sum_{n=-\infty}^{\infty} x(n)w_R(n)e^{-jn\omega} \right] \left[ \sum_{m=-\infty}^{\infty} x(m)w_R(m)e^{-jm\omega} \right]^* \right\} \\ &= \frac{1}{N} E \left\{ \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(n)x^*(m)w_R(m)w_R(n)e^{-j(n-m)\omega} \right\} \\ &= \frac{1}{N} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} r_x(n-m)w_R(m)w_R(n)e^{-j(n-m)\omega} \end{aligned} \quad (8.46)$$

With the change of variables,  $k = n - m$ , Eq. (8.46) becomes

$$\begin{aligned} E \{ \hat{P}_{per}(e^{j\omega}) \} &= \frac{1}{N} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} r_x(k)w_R(n)w_R(n-k)e^{-jk\omega} \\ &= \frac{1}{N} \sum_{k=-\infty}^{\infty} r_x(k) \left[ \sum_{n=-\infty}^{\infty} w_R(n)w_R(n-k) \right] e^{-jk\omega} \\ &= \frac{1}{N} \sum_{k=-\infty}^{\infty} r_x(k)w_B(k)e^{-jk\omega} \end{aligned} \quad (8.47)$$

more difficult, we may derive an approximate expression for the variance as follows. Recall that a random process  $x(n)$  with power spectrum  $P_x(e^{j\omega})$  may be generated by filtering unit variance white noise  $v(n)$  with a linear shift-invariant filter  $h(n)$  that has a frequency response  $H(e^{j\omega})$  with

$$|H(e^{j\omega})|^2 = P_x(e^{j\omega}) \quad (8.38)$$

As defined in Eq. (8.7), if  $x_N(n)$  and  $v_N(n)$  are the sequences of length  $N$  that are formed by windowing  $x(n)$  and  $v(n)$ , respectively, then the periodograms of these processes are

$$\hat{P}_{per}^{(x)}(e^{j\omega}) = \frac{1}{N} |X_N(e^{j\omega})|^2 \quad (8.39)$$

$$\hat{P}_{per}^{(v)}(e^{j\omega}) = \frac{1}{N} |V_N(e^{j\omega})|^2 \quad (8.40)$$

Although  $x_N(n)$  is not equal to the convolution of  $v_N(n)$  with  $h(n)$ , if  $N$  is large compared to the length of  $h(n)$  so that the transient effects are small, then

$$x_N(n) \approx h(n) * v_N(n)$$

Since

$$|X_N(e^{j\omega})|^2 \approx |H(e^{j\omega})|^2 |V_N(e^{j\omega})|^2 = P_x(e^{j\omega}) |V_N(e^{j\omega})|^2 \quad (8.41)$$

substituting Eqs. (8.39) and (8.40) into Eq. (8.41) we have

$$\hat{P}_{per}^{(x)}(e^{j\omega}) \approx P_x(e^{j\omega}) \hat{P}_{per}^{(v)}(e^{j\omega})$$

Therefore,

$$\text{Var} \left\{ \hat{P}_{per}^{(x)}(e^{j\omega}) \right\} \approx P_x^2(e^{j\omega}) \text{Var} \left\{ \hat{P}_{per}^{(v)}(e^{j\omega}) \right\}$$

and, since the variance of the periodogram of  $v(n)$  is equal to one, then we have

$$\text{Var} \left\{ \hat{P}_{per}^{(x)}(e^{j\omega}) \right\} \approx P_x^2(e^{j\omega}) \quad (8.42)$$

Thus, assuming that  $N$  is large, the variance of the periodogram of a Gaussian random process is proportional to the square of its power spectrum.

The second-order moment and covariance of the periodogram may be similarly generalized for nonwhite Gaussian noise. For the second-order moment, Eq. (8.34) becomes

$$E \left\{ \hat{P}_{per}(e^{j\omega_1}) \hat{P}_{per}(e^{j\omega_2}) \right\} \approx P_x(e^{j\omega_1}) P_x(e^{j\omega_2}) \left\{ 1 + \left[ \frac{\sin N(\omega_1 - \omega_2)/2}{N \sin(\omega_1 - \omega_2)/2} \right]^2 \right\} \quad (8.43)$$

and, for the covariance, Eq. (8.35) becomes

$$\text{Cov} \left\{ \hat{P}_{per}(e^{j\omega_1}) \hat{P}_{per}(e^{j\omega_2}) \right\} \approx P_x(e^{j\omega_1}) P_x(e^{j\omega_2}) \left[ \frac{\sin N(\omega_1 - \omega_2)/2}{N \sin(\omega_1 - \omega_2)/2} \right]^2 \quad (8.44)$$

Note that for large  $N$ , the term in brackets is approximately equal to zero provided  $\omega_1 - \omega_2 \gg 2\pi/N$ , which implies that there is little correlation between one frequency and another. The properties of the periodogram are summarized in Table 8.1.

where

$$w_B(k) = w_R(k) * w_R(-k) = \sum_{n=-\infty}^{\infty} w_R(n)w_R(n-k)$$

is a Bartlett window. Using the frequency convolution theorem, it follows that the expected value of the periodogram is

$$E \left\{ \hat{P}_{per}(e^{j\omega}) \right\} = \frac{1}{2\pi N} P_x(e^{j\omega}) * |W_R(e^{j\omega})|^2 \quad (8.48)$$

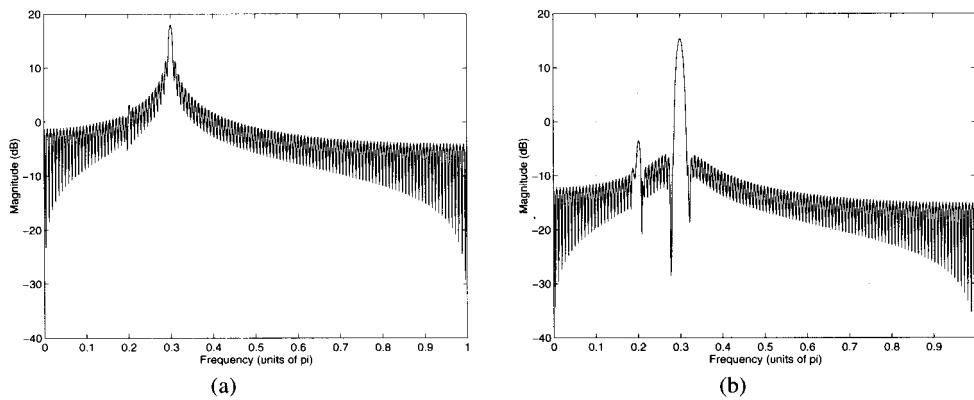
where

$$W_R(e^{j\omega}) = \frac{\sin(N\omega/2)}{\sin(\omega/2)} e^{-j(N-1)\omega/2}$$

is the Fourier transform of the rectangular data window,  $w_R(n)$ .<sup>4</sup> Therefore, the amount of smoothing in the periodogram is determined by the window that is applied to the data. Although a rectangular window has a narrow main lobe compared to other windows and, therefore, produces the least amount of spectral smoothing, it has relatively large sidelobes that may lead to masking of weak narrowband components. Consider, for example, a random process consisting of two sinusoids in white noise

$$x(n) = 0.1 \sin(n\omega_1 + \phi_1) + \sin(n\omega_2 + \phi_2) + v(n)$$

With  $\omega_1 = 0.2\pi$ ,  $\omega_2 = 0.3\pi$ , and  $N = 128$ , the expected value of the periodogram is shown in Fig. 8.10a. What we observe is that the sinusoid at frequency  $\omega_1$  is almost completely masked by the sidelobes of the window at frequency  $\omega_2$ . However, if the rectangular window is replaced with a Hamming window, then the sinusoid at frequency  $\omega_1$  is clearly visible as illustrated in Fig. 8.10b. This is due to the smaller sidelobes of a Hamming window, which are down about 30 dB compared to the sidelobes of a rectangular window. On the



**Figure 8.10** Spectral analysis of two sinusoids in white noise with sinusoidal frequencies of  $\omega_1 = .2\pi$  and  $\omega_2 = .3\pi$  and a data record length of  $N = 128$  points. (a) The expected value of the periodogram. (b) The expected value of the modified periodogram using a Hamming data window.

<sup>4</sup>Note the equivalence of Eq. (8.48) and Eq. (8.23).

other hand, this reduction in the sidelobe amplitude comes at the expense of an increase in the width of the mainlobe which, in turn, affects the resolution. The periodogram of a process that is windowed with a general window  $w(n)$  is called a *modified periodogram* and is given by

$$\hat{P}_M(e^{j\omega}) = \frac{1}{NU} \left| \sum_{n=-\infty}^{\infty} x(n)w(n)e^{-jnw} \right|^2 \quad (8.49)$$

where  $N$  is the length of the window and

$$U = \frac{1}{N} \sum_{n=0}^{N-1} |w(n)|^2 \quad (8.50)$$

is a constant that, as we will see, is defined so that  $\hat{P}_M(e^{j\omega})$  will be asymptotically unbiased. A MATLAB program for the modified periodogram is given in Fig. 8.11.

Let us now evaluate the performance of the modified periodogram. It follows from the derivation of Eq. (8.48) that the expected value of  $\hat{P}_M(e^{j\omega})$  is

$$E \left\{ \hat{P}_M(e^{j\omega}) \right\} = \frac{1}{2\pi NU} P_x(e^{j\omega}) * |W(e^{j\omega})|^2 \quad (8.51)$$

where  $W(e^{j\omega})$  is the Fourier transform of the data window. Note that with

$$U = \frac{1}{N} \sum_{n=0}^{N-1} |w(n)|^2 = \frac{1}{2\pi N} \int_{-\pi}^{\pi} |W(e^{j\omega})|^2 d\omega \quad (8.52)$$

### The Modified Periodogram

```

function Px = mper(x,win,n1,n2)
%
x = x(:);
if nargin == 2
    n1 = 1; n2 = length(x); end;
N = n2 - n1 + 1;
w = ones(N,1);
if (win == 2) w = hamming(N);
elseif (win == 3) w = hanning(N);
elseif (win == 4) w = bartlett(N);
elseif (win == 5) w = blackman(N);
end;
xw = x(n1:n2).*w/norm(w);
Px = N*periodogram(xw);
end;

```

**Figure 8.11** A MATLAB program for computing the modified periodogram of  $x(n)$ .

then

$$\frac{1}{2\pi NU} \int_{-\pi}^{\pi} |W(e^{j\omega})|^2 d\omega = 1$$

and, with an appropriate window,  $|W(e^{j\omega})|^2/NU$  will converge to an impulse of unit area as  $N \rightarrow \infty$ , and the modified periodogram will be asymptotically unbiased. (Note that if  $w(n)$  is a rectangular window, then  $U = 1$  and the modified periodogram reduces to the periodogram).

Since the modified periodogram is simply the periodogram of a windowed data sequence, the variance of  $\hat{P}_M(e^{j\omega})$  will be approximately the same as that for the periodogram, i.e.,

$$\text{Var} \left\{ \hat{P}_M(e^{j\omega}) \right\} \approx P_x^2(e^{j\omega}) \quad (8.53)$$

Therefore, the modified periodogram is not a consistent estimate of the power spectrum and the data window offers no benefit in terms of reducing the variance. What the window does provide, however, is a trade-off between spectral resolution (main lobe width) and spectral masking (sidelobe amplitude). For example, with the resolution of  $\hat{P}_M(e^{j\omega})$  defined to be the 3 dB bandwidth of the data window,<sup>5</sup>

$$\text{Res} \left[ \hat{P}_M(e^{j\omega}) \right] = (\Delta\omega)_{3\text{dB}} \quad (8.54)$$

we see from Table 8.2 that a 43 dB reduction in the sidelobe amplitude that comes from using a Hamming data window instead of a rectangular window results in a reduction in the spectral resolution of about 50%. The properties of the modified periodogram are summarized in Table 8.3. An extensive list of windows and window characteristics may be found in [14].

**Table 8.2 Properties of a Few Commonly Used Windows. Each Window is Assumed to be of Length N.**

Window	Sidelobe Level (dB)	3 dB BW $(\Delta\omega)_{3\text{dB}}$
Rectangular	-13	$0.89(2\pi/N)$
Bartlett	-27	$1.28(2\pi/N)$
Hanning	-32	$1.44(2\pi/N)$
Hamming	-43	$1.30(2\pi/N)$
Blackman	-58	$1.68(2\pi/N)$

<sup>5</sup>Note that this is consistent with the definition of resolution given in Eq. (8.28) for the periodogram. Specifically, although the periodogram resolution is defined to be the 6 dB bandwidth of  $W_B(e^{j\omega})$ , since  $W_B(e^{j\omega}) = |W_R(e^{j\omega})|^2$ , this is equivalent to the 3 dB bandwidth of  $W_R(e^{j\omega})$ .

**Table 8.3 Properties of the Modified Periodogram**

$\hat{P}_M(e^{j\omega}) = \frac{1}{NU} \left  \sum_{n=-\infty}^{\infty} w(n)x(n)e^{-jnw} \right ^2$	
	$U = \frac{1}{N} \sum_{n=0}^{N-1}  w(n) ^2$
<i>Bias</i>	
	$E \{ \hat{P}_M(e^{j\omega}) \} = \frac{1}{2\pi NU} P_x(e^{j\omega}) *  W(e^{j\omega}) ^2$
<i>Resolution</i>	Window dependent
<i>Variance</i>	
	$\text{Var} \{ \hat{P}_M(e^{j\omega}) \} \approx P_x^2(e^{j\omega})$

### 8.2.4 Bartlett's Method: Periodogram Averaging

In this section, we look at Bartlett's method of periodogram averaging, which, unlike either the periodogram or the modified periodogram, produces a consistent estimate of the power spectrum [5]. The motivation for this method comes from the observation that the expected value of the periodogram converges to  $P_x(e^{j\omega})$  as the data record length  $N$  goes to infinity,

$$\lim_{N \rightarrow \infty} E \{ \hat{P}_{per}(e^{j\omega}) \} = P_x(e^{j\omega}) \quad (8.55)$$

Therefore, if we can find a consistent estimate of the mean,  $E \{ \hat{P}_{per}(e^{j\omega}) \}$ , then this estimate will be a consistent estimate of  $P_x(e^{j\omega})$ .

In our discussion of the sample mean in Section 3.2.8, we saw how averaging a set of uncorrelated measurements of a random variable  $x$  yields a consistent estimate of the mean,  $E \{ x \}$ . This suggests that we consider estimating the power spectrum of a random process by periodogram averaging. Thus, let  $x_i(n)$  for  $i = 1, 2, \dots, K$  be  $K$  uncorrelated realizations of a random process  $x(n)$  over the interval  $0 \leq n < L$ . With  $\hat{P}_{per}^{(i)}(e^{j\omega})$  the periodogram of  $x_i(n)$ ,

$$\hat{P}_{per}^{(i)}(e^{j\omega}) = \frac{1}{L} \left| \sum_{n=0}^{L-1} x_i(n) e^{-jnw} \right|^2 ; \quad i = 1, 2, \dots, K \quad (8.56)$$

the average of these periodograms is

$$\hat{P}_x(e^{j\omega}) = \frac{1}{K} \sum_{i=1}^K \hat{P}_{per}^{(i)}(e^{j\omega}) \quad (8.57)$$

Evaluating the expected value of  $\hat{P}_x(e^{j\omega})$  we have

$$E \{ \hat{P}_x(e^{j\omega}) \} = E \{ \hat{P}_{per}^{(i)}(e^{j\omega}) \} = \frac{1}{2\pi} P_x(e^{j\omega}) * W_B(e^{j\omega}) \quad (8.58)$$

where  $W_B(e^{j\omega})$  is the Fourier transform of a Bartlett window,  $w_B(k)$ , that extends from  $-L$

to  $L$ . Therefore, as with the periodogram,  $\hat{P}_x(e^{j\omega})$  is asymptotically unbiased. In addition, with our assumption that the data records are uncorrelated, it follows that the variance of  $\hat{P}_x(e^{j\omega})$  is

$$\text{Var} \left\{ \hat{P}_x(e^{j\omega}) \right\} = \frac{1}{K} \text{Var} \left\{ \hat{P}_{per}^{(i)}(e^{j\omega}) \right\} \approx \frac{1}{K} P_x^2(e^{j\omega}) \quad (8.59)$$

which goes to zero as  $K$  goes to infinity. Therefore,  $\hat{P}_x(e^{j\omega})$  is a consistent estimate of the power spectrum provided that both  $K$  and  $L$  are allowed to go to infinity. However, the difficulty with this approach is that uncorrelated realizations of a process are generally not available. Instead, one typically only has a single realization of length  $N$ . Therefore, Bartlett proposed that  $x(n)$  be partitioned into  $K$  nonoverlapping sequences of length  $L$  where  $N = KL$  as illustrated in Fig. 8.12. The Bartlett estimate is then computed as in Eq. (8.56) and Eq. (8.57) with

$$\begin{aligned} x_i(n) &= x(n + iL) & n = 0, 1, \dots, L-1 \\ i &= 0, 1, \dots, K-1 \end{aligned}$$

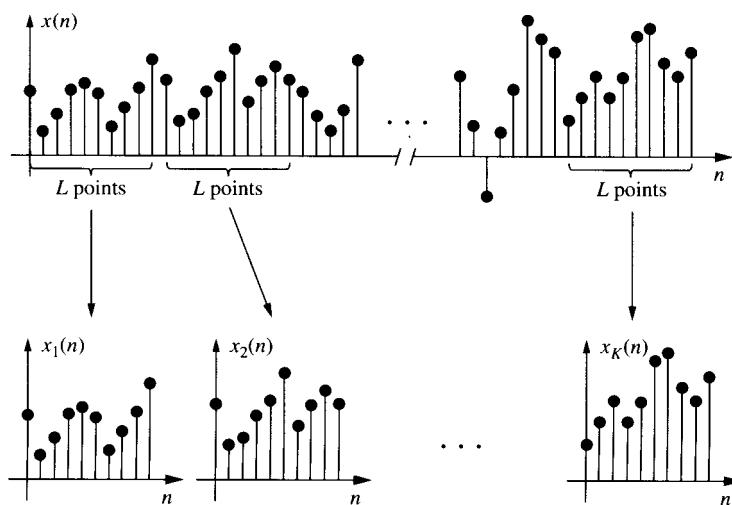
Thus, the Bartlett estimate is

$$\boxed{\hat{P}_B(e^{j\omega}) = \frac{1}{N} \sum_{i=0}^{K-1} \left| \sum_{n=0}^{L-1} x(n + iL) e^{-jn\omega} \right|^2} \quad (8.60)$$

A MATLAB program to compute the Bartlett estimate is given in Fig. 8.13.

Based on our analysis of the periodogram and the modified periodogram, we may easily evaluate the performance of Bartlett's method as follows. First, as in Eq. (8.58), the expected value of Bartlett's estimate is

$$\boxed{E \left\{ \hat{P}_B(e^{j\omega}) \right\} = \frac{1}{2\pi} P_x(e^{j\omega}) * W_B(e^{j\omega})} \quad (8.61)$$



**Figure 8.12** Partitioning  $x(n)$  into nonoverlapping subsequences.

**Bartlett's Method**

```

function Px = bart(x,nsect)
%
L = floor(length(x)/nsect);
Px = 0;
n1 = 1;
for i=1:nsect
    Px = Px + periodogram(x(n1:n1+L-1))/nsect;
    n1 = n1 + L;
end;

```

**Figure 8.13** A MATLAB program for estimating the power spectrum using Bartlett's method of averaging periodograms. Note that this m-file calls periodogram.m.

Therefore,  $\hat{P}_B(e^{j\omega})$  is asymptotically unbiased. Second, since the periodograms used in  $\hat{P}_B(e^{j\omega})$  are computed using sequences of length  $L$ , then the resolution is

$$\text{Res} \left[ \hat{P}_B(e^{j\omega}) \right] = 0.89 \frac{2\pi}{L} = 0.89K \frac{2\pi}{N} \quad (8.62)$$

which is  $K$  times larger (worse) than the periodogram. Finally, since the sequences  $x_i(n)$  are generally correlated with one another, unless  $x(n)$  is white noise, then the variance reduction will not be as large as that given in Eq. (8.59). However, the variance will be inversely proportional to  $K$  and, assuming that the data sequences are approximately uncorrelated, for large  $N$  the variance is approximately

$$\text{Var} \left\{ \hat{P}_B(e^{j\omega}) \right\} \approx \frac{1}{K} \text{Var} \left\{ \hat{P}_{per}^{(i)}(e^{j\omega}) \right\} \approx \frac{1}{K} P_x^2(e^{j\omega}) \quad (8.63)$$

Thus, if both  $K$  and  $L$  are allowed to go to infinity as  $N \rightarrow \infty$ , then  $\hat{P}_B(e^{j\omega})$  will be a consistent estimate of the power spectrum. In addition, for a given value of  $N$ , Bartlett's method allows one to trade a reduction in spectral resolution for a reduction in variance by simply changing the values of  $K$  and  $L$ . Table 8.4 summarizes the properties of Bartlett's method.

**Table 8.4 Properties of Bartlett's Method**

$\hat{P}_B(e^{j\omega}) = \frac{1}{N} \sum_{i=0}^{K-1} \left  \sum_{n=0}^{L-1} x(n+iL)e^{-jn\omega} \right ^2$ <p><i>Bias</i></p> $E \left\{ \hat{P}_B(e^{j\omega}) \right\} = \frac{1}{2\pi} P_x(e^{j\omega}) * W_B(e^{j\omega})$ <p><i>Resolution</i></p> $\Delta\omega = 0.89K \frac{2\pi}{N}$ <p><i>Variance</i></p> $\text{Var} \left\{ \hat{P}_B(e^{j\omega}) \right\} \approx \frac{1}{K} P_x^2(e^{j\omega})$
--

---

**Example 8.2.5 Bartlett's Method**

In Example 8.2.4, we considered using the periodogram to estimate the power spectrum of white noise. What we observed was that the variance of the estimate did not decrease as we increased the length  $N$  of the data sequence. Shown in Fig. 8.14a are the periodograms ( $K = 1$ ) of 50 different unit variance white noise sequences of length  $N = 512$  and in Fig. 8.14b is the ensemble average. The Bartlett estimates for these sequences with  $K = 4$  sections of length  $L = 128$  are shown in Fig. 8.14c with the average given in Fig. 8.14d. Similarly, the Bartlett estimates of these 50 processes using  $K = 16$  sections of length  $L = 32$  are shown in Fig. 8.14e and the average in Fig. 8.14f. What we observe in these examples is that the variance of the estimate decreases in proportion to the number of sections  $K$ .

As another example, let  $x(n)$  be a process consisting of two sinusoids in unit variance white noise,

$$x(n) = A \sin(n\omega_1 + \phi_1) + \sin(n\omega_2 + \phi_2) + v(n)$$

where  $\omega_1 = 0.2\pi$ ,  $\omega_2 = 0.25\pi$ , and  $A = \sqrt{10}$ . With  $N = 512$ , an overlay plot of 50 periodograms is shown in Fig. 8.15a and the ensemble average of these periodograms is given in Fig. 8.15b. Using Bartlett's method with  $K = 4$  and  $K = 16$  sections, the overlay plots and ensemble averages are shown in Fig. 8.15c-f. Note that although the variance of the estimate decreases with  $K$ , there is a corresponding decrease in the resolution as evidenced by the broadening of the spectral peaks.

---

### 8.2.5 Welch's Method: Averaging Modified Periodograms

In 1967, Welch proposed two modifications to Bartlett's method [59]. The first is to allow the sequences  $x_i(n)$  to overlap, and the second is to allow a data window  $w(n)$  to be applied to each sequence, thereby producing a set of *modified periodograms* that are to be averaged. Assuming that successive sequences are offset by  $D$  points and that each sequence is  $L$  points long then the  $i$ th sequence is given by

$$x_i(n) = x(n + iD) ; n = 0, 1, \dots, L - 1$$

Thus, the amount of overlap between  $x_i(n)$  and  $x_{i+1}(n)$  is  $L - D$  points, and if  $K$  sequences cover the entire  $N$  data points, then

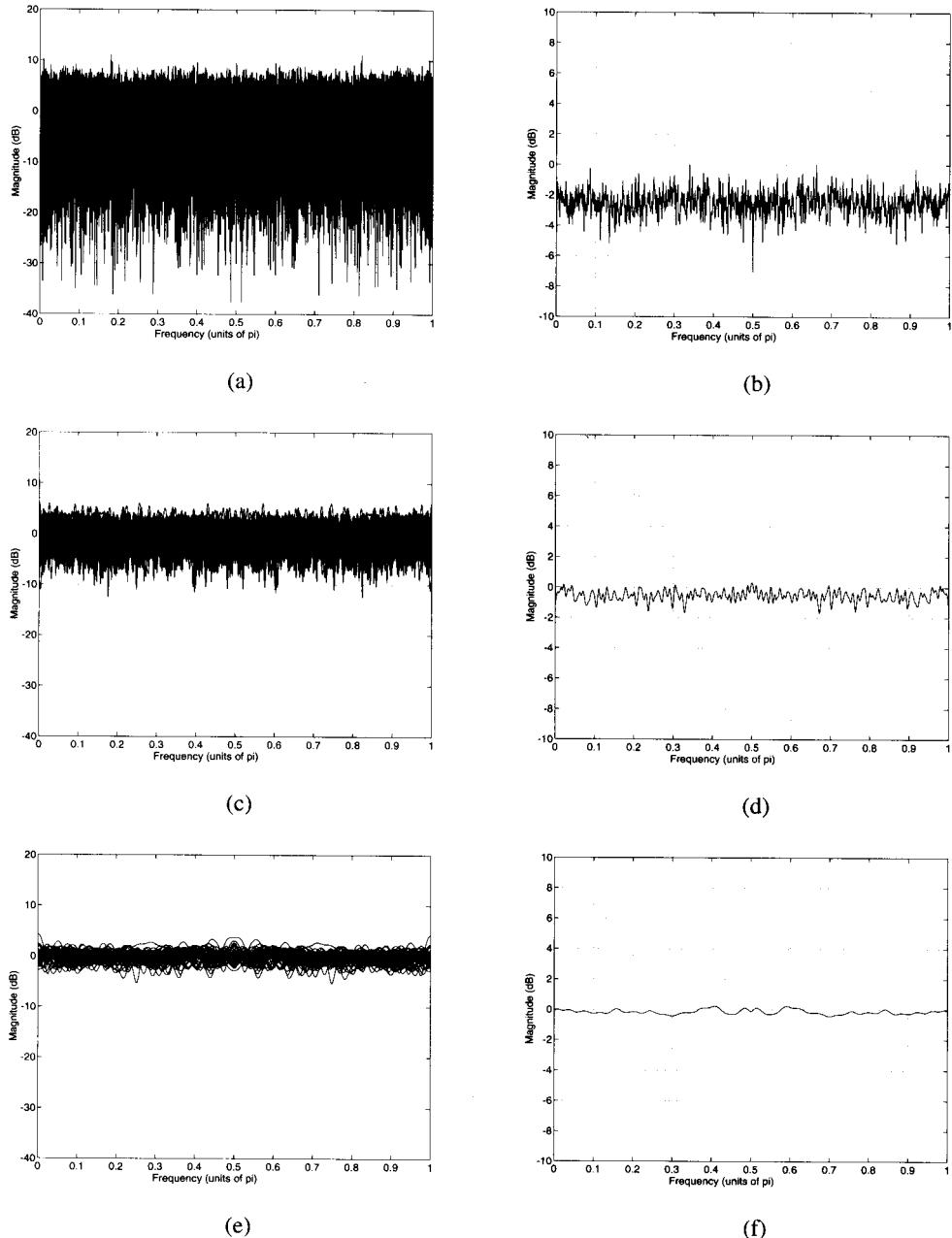
$$N = L + D(K - 1).$$

For example, with no overlap ( $D = L$ ) we have  $K = N/L$  sections of length  $L$  as in Bartlett's method. On the other hand, if the sequences are allowed to overlap by 50% ( $D = L/2$ ) then we may form

$$K = 2 \frac{N}{L} - 1$$

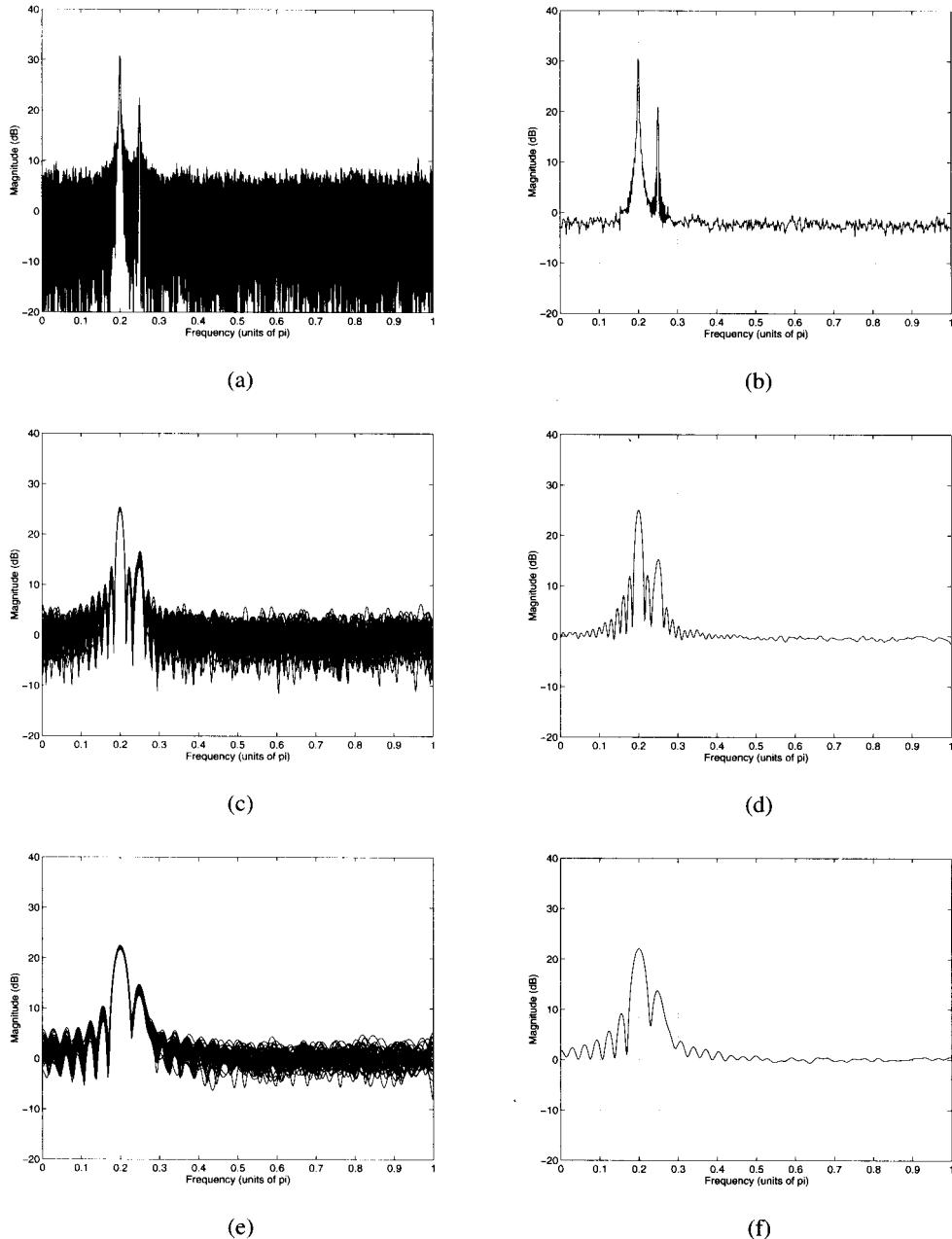
sections of length  $L$ , thus maintaining the same resolution (section length) as Bartlett's method while doubling the number of modified periodograms that are averaged, thereby reducing the variance. However, with a 50% overlap we could also form

$$K = \frac{N}{L} - 1$$



**Figure 8.14** Spectrum estimation of unit variance white Gaussian noise. (a) Overlay plot of 50 periodograms with  $N = 512$  and (b) the ensemble average. (c) Overlay plot of 50 Bartlett estimates with  $K = 4$  and  $L = 128$  and (d) the ensemble average. (e) Overlay plot of 50 Bartlett estimates with  $K = 8$  and  $L = 64$  and (f) the ensemble average.

sequences of length  $2L$ , thus increasing the resolution while maintaining the same variance as Bartlett's method. Therefore, by allowing the sequences to overlap it is possible to increase the number and/or length of the sequences that are averaged, thereby trading a



**Figure 8.15** Spectrum estimation of two sinusoids in white noise using  $N = 512$  data values. (a) Overlay plot of 50 periodograms and (b) the ensemble average. (c) Overlay plot of 50 Bartlett estimates with  $K = 4$  and  $L = 128$  and (d) the ensemble average. (e) Overlay plot of 50 Bartlett estimates with  $K = 8$  and  $L = 64$  and (f) the ensemble average.

reduction in variance for a reduction in resolution. A MATLAB program for estimating the power spectrum using Welch's method is given in Fig. 8.16.

Let us now evaluate the performance of Welch's method. Note that the estimate produced

**Welch's Method**

```

function Px = welch(x,L,over,win)
%
if (over >= 1) | (over < 0)
    error('Overlap is invalid'), end
n1 = 1;
n0 = (1-over)*L;
nsect=1+floor((length(x)-L)/(n0));
Px=0;
for i=1:nsect
    Px = Px + mper(x,win,n1,n1+L-1)/nsect;
    n1 = n1 + n0;
end;

```

**Figure 8.16** A MATLAB program for estimating the power spectrum using Welch's method of averaging modified periodograms. Note that this m-file calls mper.m.

with Welch's method may be written explicitly in terms of  $x(n)$  as follows

$$\hat{P}_W(e^{j\omega}) = \frac{1}{KLU} \sum_{i=0}^{K-1} \left| \sum_{n=0}^{L-1} w(n)x(n+iD)e^{-jnw} \right|^2 \quad (8.64)$$

or, more succinctly, in terms of modified periodograms as

$$\hat{P}_W(e^{j\omega}) = \frac{1}{K} \sum_{i=0}^{K-1} \hat{P}_M^{(i)}(e^{j\omega}) \quad (8.65)$$

Therefore, the expected value of Welch's estimate is

$$E\{\hat{P}_W(e^{j\omega})\} = E\{\hat{P}_M(e^{j\omega})\} = \frac{1}{2\pi LU} P_x(e^{j\omega}) * |W(e^{j\omega})|^2 \quad (8.66)$$

where  $W(e^{j\omega})$  is the Fourier transform of the  $L$ -point data window,  $w(n)$ , used in Eq. (8.64) to form the modified periodograms. Thus, as with each of the previous periodogram-based methods, Welch's method is an asymptotically unbiased estimate of the power spectrum. The resolution, however, depends on the data window. As with the modified periodogram, the resolution is defined to be the 3 dB bandwidth of the data window (see Table 8.2). The variance, on the other hand, is more difficult to compute since, with overlapping sequences, the modified periodograms cannot be assumed to be uncorrelated. Nevertheless, it has been shown that, with a Bartlett window and a 50% overlap, the variance is approximately [59]

$$\text{Var}\{\hat{P}_W(e^{j\omega})\} \approx \frac{9}{8K} P_x^2(e^{j\omega}) \quad (8.67)$$

Comparing Eqs. (8.67) and (8.63) we see that, for a given number of sections  $K$ , the variance of the estimate using Welch's method is larger than that for Bartlett's method by a factor of 9/8. However, for a fixed amount of data,  $N$ , and a given resolution (sequence length  $L$ ), with a 50% overlap twice as many sections may be averaged in Welch's method. Expressing

**Table 8.5 Properties of Welch's Method**

$\hat{P}_W(e^{j\omega}) = \frac{1}{KLU} \sum_{i=0}^{K-1} \left  \sum_{n=0}^{L-1} w(n)x(n+iD)e^{-jnw} \right ^2$
$U = \frac{1}{L} \sum_{n=0}^{L-1}  w(n) ^2$
<i>Bias</i>
$E \{ \hat{P}_W(e^{j\omega}) \} = \frac{1}{2\pi LU} P_x(e^{j\omega}) *  W(e^{j\omega}) ^2$
<i>Resolution</i> Window dependent
<i>Variance</i> †
$\text{Var} \{ \hat{P}_W(e^{j\omega}) \} \approx \frac{9}{16} \frac{L}{N} P_x^2(e^{j\omega})$

† Assuming 50% overlap and a Bartlett window.

the variance in terms of  $L$  and  $N$ , we have (assuming a 50% overlap)

$$\boxed{\text{Var} \{ \hat{P}_W(e^{j\omega}) \} \approx \frac{9}{16} \frac{L}{N} P_x^2(e^{j\omega})} \quad (8.68)$$

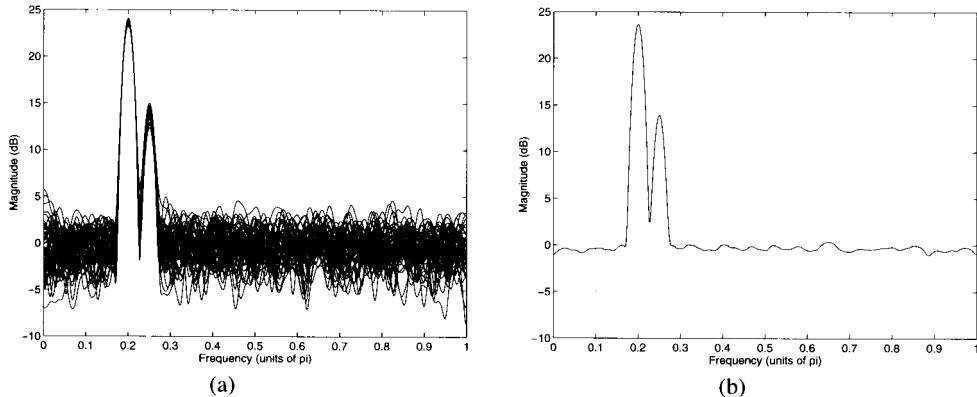
Since  $N/L$  is the number of sections that are used in Bartlett's method, it follows from Eq. (8.63) that

$$\text{Var} \{ \hat{P}_W(e^{j\omega}) \} \approx \frac{9}{16} \text{Var} \{ \hat{P}_B(e^{j\omega}) \}$$

Although it is possible to average more sequences for a given amount of data by increasing the amount of overlap, the computational requirements increase in proportion with  $K$ . In addition, since an increase in the amount of overlap increases the correlation between the sequences  $x_i(n)$ , there are diminishing returns when increasing  $K$  for a fixed  $N$ . Therefore the amount of overlap is typically either 50% or 75%. The properties of Welch's method are summarized in Table 8.5.

#### **Example 8.2.6 Welch's Method**

Consider the process defined in Example 8.2.5 consisting of two sinusoids in unit variance white noise. Using Welch's method with  $N = 512$ , a section length  $L = 128$ , a 50% overlap (7 sections), and a Hamming window, an overlay plot of the spectrum estimates for 50 different realizations of the process are shown in Fig. 8.17a and the ensemble average is shown in Fig. 8.17b. Comparing these estimates with those shown in Fig. 8.15e and f of Example 8.2.5, we see that, since the number of sections used in both examples are about the same (7 versus 8), then the variance of the two estimates are approximately the same. In addition, although the width of the main lobe of the Hamming window used in Welch's method is 1.46 times the width of the rectangular window used in Bartlett's method, the resolution is about the same. The reason for this is due to the fact that the 50% overlap that



**Figure 8.17** (a) An overlay plot of 50 estimates of the spectrum of two sinusoids in noise using Welch's method with  $N = 512$ , a section length of  $L = .128$ , 50% overlap (7 sections), and a Hamming window. (b) The average of the estimates in (a).

is used in Welch's method allows for the section length to be twice the length of that used in Bartlett's method. So what do we gain with Welch's method? We gain a reduction in the amount of spectral leakage that takes place through the sidelobes of the data window.

### 8.2.6 Blackman-Tukey Method: Periodogram Smoothing

The methods of Bartlett and Welch are designed to reduce the variance of the periodogram by averaging periodograms and modified periodograms, respectively. Another method for decreasing the statistical variability of the periodogram is periodogram smoothing, often referred to as the Blackman-Tukey method after the pioneering work of Blackman and Tukey in spectrum analysis [6]. To see how periodogram smoothing may reduce the variance of the periodogram, recall that the periodogram is computed by taking the Fourier transform of a consistent estimate of the autocorrelation sequence. However, for any finite data record of length  $N$ , the variance of  $\hat{r}_x(k)$  will be large for values of  $k$  that are close to  $N$ . For example, note that the estimate of  $r_x(k)$  at lag  $k = N - 1$  is

$$\hat{r}_x(N - 1) = \frac{1}{N} x(N - 1)x(0)$$

Since there is little averaging that goes into the formation of the estimates of  $r_x(k)$  for  $|k| \approx N$ , no matter how large  $N$  becomes, these estimates will always be unreliable. Consequently, the only way to reduce the variance of the periodogram is to reduce the variance of these estimates or to reduce the contribution that they make to the periodogram. In the methods of Bartlett and Welch, the variance of the periodogram is decreased by reducing the variance of the autocorrelation estimate by averaging. In the Blackman-Tukey method, the variance of the periodogram is reduced by applying a window to  $\hat{r}_x(k)$  in order to decrease the contribution of the unreliable estimates to the periodogram. Specifically, the Blackman-Tukey spectrum estimate is

$$\hat{P}_{BT}(e^{j\omega}) = \sum_{k=-M}^M \hat{r}_x(k)w(k)e^{-jk\omega} \quad (8.69)$$

where  $w(k)$  is a *lag window* that is applied to the autocorrelation estimate. For example, if  $w(k)$  is a rectangular window extending from  $-M$  to  $M$  with  $M < N - 1$ , then the estimates of  $r_x(k)$  having the largest variance are set to zero and, consequently, the power spectrum estimate will have a smaller variance. What is traded for this reduction in variance, however, is a reduction in resolution since a smaller number of autocorrelation estimates are used to form the estimate of the power spectrum.

Using the frequency convolution theorem, the Blackman-Tukey spectrum may be written in the frequency domain as follows:

$$\hat{P}_{BT}(e^{j\omega}) = \frac{1}{2\pi} \hat{P}_{per}(e^{j\omega}) * W(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{P}_{per}(e^{ju}) W(e^{j(\omega-u)}) du \quad (8.70)$$

Therefore, the Blackman-Tukey estimate smoothes the periodogram by convolving with the Fourier transform of the autocorrelation window,  $W(e^{j\omega})$ . Although there is considerable flexibility in the choice of the window that may be used,  $w(k)$  should be conjugate symmetric so that  $W(e^{jw})$  is real-valued, and the window should have a nonnegative Fourier transform,  $W(e^{j\omega}) \geq 0$ , so that  $\hat{P}_{BT}(e^{j\omega})$  is guaranteed to be nonnegative. A MATLAB program for generating the Blackman-Tukey spectrum estimate is given in Fig. 8.18.

To analyze the performance of the Blackman-Tukey method, we will evaluate the bias and the variance (the resolution is window-dependent). The bias may be computed by taking the expected value of Eq. (8.70) as follows:

$$E\{\hat{P}_{BT}(e^{j\omega})\} = \frac{1}{2\pi} E\{\hat{P}_{per}(e^{j\omega})\} * W(e^{j\omega}) \quad (8.71)$$

#### **Blackman-Tukey Method**

```
function Px = per_smooth(x,win,M,n1,n2)
%
x = x(:);
if nargin == 3
    n1 = 1; n2 = length(x); end;
R = covar(x(n1:n2),M);
r = [fliplr(R(1,2:M)),R(1,1),R(1,2:M)];
M = 2*M-1;
w = ones(M,1);
if (win == 2) w = hamming(M);
elseif (win == 3) w = hanning(M);
elseif (win == 4) w = bartlett(M);
elseif (win == 5) w = blackman(M);
end;
r = r'*w;
Px = abs(fft(r,1024));
Px(1)=Px(2);
end;
```

**Figure 8.18** A MATLAB program for estimating the power spectrum using the Blackman-Tukey method (periodogram smoothing).

Substituting Eq. (8.23) for the expected value of the periodogram we have

$$E\{\hat{P}_{BT}(e^{j\omega})\} = \frac{1}{2\pi} P_x(e^{j\omega}) * W_B(e^{j\omega}) * W(e^{j\omega}) \quad (8.72)$$

or, equivalently,

$$E\{\hat{P}_{BT}(e^{j\omega})\} = \sum_{k=-M}^M r_x(k)w_B(k)w(k)e^{-jk\omega} \quad (8.73)$$

If we let  $w_{BT}(k) = w_B(k)w(k)$  be the combined window that is applied to the autocorrelation sequence  $r_x(k)$ , using the frequency convolution theorem we have

$$E\{\hat{P}_{BT}(e^{j\omega})\} = \frac{1}{2\pi} P_x(e^{j\omega}) * W_{BT}(e^{j\omega}) \quad (8.74)$$

If we assume that  $M \ll N$  so that  $w_B(k)w(k) \approx w(k)$ , then we have

$$E\{\hat{P}_{BT}(e^{j\omega})\} \approx \frac{1}{2\pi} P_x(e^{j\omega}) * W(e^{j\omega}) \quad (8.75)$$

where  $W(e^{j\omega})$  is the Fourier transform of the lag window  $w(k)$ .

Evaluating the variance of the Blackman-Tukey spectrum estimate requires a bit more work.<sup>6</sup> Since

$$\text{Var}\{\hat{P}_{BT}(e^{j\omega})\} = E\{\hat{P}_{BT}^2(e^{j\omega})\} - E^2\{\hat{P}_{BT}(e^{j\omega})\} \quad (8.76)$$

we begin by finding the mean-square value  $E\{\hat{P}_{BT}^2(e^{j\omega})\}$ . From Eq. (8.70) we have

$$\hat{P}_{BT}^2(e^{j\omega}) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \hat{P}_{per}(e^{ju}) \hat{P}_{per}(e^{jv}) W(e^{j(\omega-u)}) W(e^{j(\omega-v)}) du dv$$

Therefore, the mean-square value is

$$E\{\hat{P}_{BT}^2(e^{j\omega})\} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} E\{\hat{P}_{per}(e^{ju}) \hat{P}_{per}(e^{jv})\} W(e^{j(\omega-u)}) W(e^{j(\omega-v)}) du dv$$

Using the approximation given in Eq. (8.43) for  $E\{\hat{P}_{per}(e^{ju}) \hat{P}_{per}(e^{jv})\}$  leads to an expression for the mean-square value that contains two terms. The first term is

$$\begin{aligned} & \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} P_x(e^{ju}) P_x(e^{jv}) W(e^{j(\omega-u)}) W(e^{j(\omega-v)}) du dv \\ &= \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x(e^{ju}) W(e^{j(\omega-u)}) du \right]^2 = E^2\{\hat{P}_{BT}(e^{j\omega})\} \end{aligned} \quad (8.77)$$

which is cancelled by the second term in Eq. (8.76). Therefore, the variance is

$$\begin{aligned} \text{Var}\{\hat{P}_{BT}(e^{j\omega})\} &= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} P_x(e^{ju}) P_x(e^{jv}) \\ &\quad \times \left[ \frac{\sin N(u-v)/2}{N \sin(u-v)/2} \right]^2 W(e^{j(\omega-u)}) W(e^{j(\omega-v)}) du dv \end{aligned} \quad (8.78)$$

<sup>6</sup>The reader may jump to the final result given in Eq. (8.79) without any loss in continuity.

Since

$$W_B(e^{j\omega}) = \frac{1}{N} \left[ \frac{\sin(N\omega/2)}{\sin(\omega/2)} \right]^2$$

is the discrete-time Fourier transform of a Bartlett window, then  $w_B(k)$  approaches a constant as  $N \rightarrow \infty$  and  $W_B(e^{j\omega})$  converges to an impulse. Therefore, if  $N$  is large then the term in brackets may be approximated by an impulse of area  $2\pi/N$ ,

$$\left[ \frac{\sin N(u-v)/2}{N \sin(u-v)/2} \right]^2 \approx \frac{2\pi}{N} u_0(u-v)$$

Thus, for large  $N$ , the variance of the Blackman-Tukey estimate is approximately

$$\text{Var} \left\{ \hat{P}_{BT}(e^{j\omega}) \right\} \approx \frac{1}{2\pi N} \int_{-\pi}^{\pi} P_x^2(e^{ju}) W^2(e^{j(\omega-u)}) du$$

If  $M$  is large enough so that we may assume that  $P_x(e^{j\omega})$  is constant across the main lobe of  $W(e^{j(\omega-u)})$ , then  $P_x^2(e^{j\omega})$  may be pulled out of the integral,

$$\text{Var} \left\{ \hat{P}_{BT}(e^{j\omega}) \right\} \approx \frac{1}{2\pi N} P_x^2(e^{j\omega}) \int_{-\pi}^{\pi} W^2(e^{j(\omega-u)}) du$$

Finally, using Parseval's theorem we have

$$\text{Var} \left\{ \hat{P}_{BT}(e^{j\omega}) \right\} \approx P_x^2(e^{j\omega}) \frac{1}{N} \sum_{k=-M}^M w^2(k) \quad (8.79)$$

provided  $N \gg M \gg 1$ . Thus, from Eqs. (8.75) and (8.79) we again see the trade-off between bias and variance. For a small bias,  $M$  should be large in order to minimize the width of the main lobe of  $W(e^{j\omega})$  whereas  $M$  should be small in order to minimize the sum in Eq. (8.79). Generally, it is recommended that  $M$  have a maximum value of  $M = N/5$  [26]. The properties of the Blackman-Tukey method are summarized in Table 8.6.

**Table 8.6 Properties of the Blackman-Tukey Method**

---

$\hat{P}_{BT}(e^{j\omega}) = \sum_{k=-M}^M \hat{r}_x(k) w(k) e^{-jk\omega}$
<i>Bias</i>
$E \left\{ \hat{P}_{BT}(e^{j\omega}) \right\} \approx \frac{1}{2\pi} P_x(e^{j\omega}) * W(e^{j\omega})$
<i>Resolution</i>
Window dependent
<i>Variance</i>
$\text{Var} \left\{ \hat{P}_{BT}(e^{j\omega}) \right\} \approx P_x^2(e^{j\omega}) \frac{1}{N} \sum_{k=-M}^M w^2(k)$

---

### 8.2.7 Performance Comparisons

An important issue in the selection of a spectrum estimation technique is the performance of the estimator. In the previous sections, we have seen that, in comparing one nonparametric method to another, there is a trade-off between resolution and variance. This section summarizes the performance of each nonparametric technique in terms of two criteria. The first is the *variability* of the estimate,

$$\mathcal{V} = \frac{\text{Var} \left\{ \hat{P}_x(e^{j\omega}) \right\}}{E^2 \left\{ \hat{P}_x(e^{j\omega}) \right\}}$$

which is a *normalized variance*. The second measure is an overall *figure of merit* that is defined as the product of the variability and the resolution,

$$\mathcal{M} = \mathcal{V} \Delta\omega$$

This figure of merit should be as small as possible. However, as we will soon discover, this figure of merit is approximately the same for all of the nonparametric methods that we have considered.

**Periodogram.** In Section 8.2.2 it was shown that the periodogram is asymptotically unbiased and that, for large  $N$ , the variance is approximately equal to  $P_x^2(e^{j\omega})$ . Asymptotically, therefore, the variability of the periodogram is equal to one

$$\mathcal{V}_{per} = \frac{P_x^2(e^{j\omega})}{P_x^2(e^{j\omega})} = 1$$

Thus, since the resolution of the periodogram is

$$\Delta\omega = 0.89 \frac{2\pi}{N}$$

then the overall figure of merit is

$$\mathcal{M}_{per} = 0.89 \frac{2\pi}{N}$$

which is inversely proportional to the data record length,  $N$ .

**Bartlett's Method.** In Bartlett's method, a reduction in variance is achieved by averaging periodograms. With  $N = KL$ , if  $N$  is large then the variance is approximately

$$\text{Var} \left\{ \hat{P}_B(e^{j\omega}) \right\} \approx \frac{1}{K} P_x^2(e^{j\omega})$$

and the variability is

$$\mathcal{V}_B = \frac{1}{K} \frac{P_x^2(e^{j\omega})}{P_x^2(e^{j\omega})} = \frac{1}{K}$$

Since the resolution is  $\Delta\omega = 0.89(2\pi K/N)$  then the figure of merit is

$$\mathcal{M}_B = 0.89 \frac{2\pi}{N}$$

which is the same as the figure of merit for the periodogram.

**Welch's Method.** The statistical properties of Welch's method depend on the amount of overlap that is used and on the type of data window. With a 50% overlap and a Bartlett window, the variability for large  $N$  is approximately

$$\mathcal{V}_W = \frac{9}{8} \frac{1}{K} = \frac{9}{16} \frac{L}{N}$$

Since the 3 dB bandwidth of a Bartlett window of length  $L$  is  $1.28(2\pi/L)$  the resolution is

$$\Delta\omega = 1.28 \frac{2\pi}{L}$$

and the figure of merit becomes

$$\mathcal{M}_W = 0.72 \frac{2\pi}{N}$$

**Blackman-Tukey Method.** Since the variance and resolution of the Blackman-Tukey method depend on the window that is used, suppose  $w(k)$  is a Bartlett window of length  $2M$  that extends from  $k = -M$  to  $k = M$ . Assuming that  $N \gg M \gg 1$ , the variance of the Blackman-Tukey estimate is approximately

$$\text{Var} \left\{ \hat{P}_{BT}(e^{j\omega}) \right\} \approx P_x^2(e^{j\omega}) \frac{1}{N} \sum_{k=-M}^M \left( 1 - \frac{|k|}{M} \right)^2 \approx P_x^2(e^{j\omega}) \frac{2M}{3N}$$

(here we have used the series given in Table 2.3 on p. 16 to evaluate the sum). Therefore, the variability is

$$\mathcal{V}_{BT} = \frac{2M}{3N}$$

Since the 3 dB bandwidth of a Bartlett window of length  $2M$  is equal to  $1.28(2\pi/2M)$ , then the resolution is

$$\Delta\omega = 0.64 \frac{2\pi}{M}$$

and the figure of merit becomes

$$\mathcal{M}_{BT} = 0.43 \frac{2\pi}{N}$$

which is slightly smaller than the figure of merit for Welch's method.

**Summary.** Table 8.7 provides a summary of the performance measures presented above for the periodogram-based spectrum estimation techniques discussed in this section. What is apparent from this table is that each technique has a figure of merit that is approximately the same, and that these figures of merit are inversely proportional to the length of the data sequence,  $N$ . Therefore, although each method differs in its resolution and variance, the overall performance is fundamentally limited by the amount of data that is available. In the following sections, we will look at some entirely different approaches to spectrum estimation in the hope of being able to find a high-resolution spectrum estimate with a small variance that works well on short data records.

**Table 8.7 Performance Measures for the Nonparametric Methods of Spectrum Estimation**

	Variability $\mathcal{V}$	Resolution $\Delta\omega$	Figure of Merit $\mathcal{M}$
Periodogram	1	$0.89 \frac{2\pi}{N}$	$0.89 \frac{2\pi}{N}$
Bartlett	$\frac{1}{K}$	$0.89K \frac{2\pi}{N}$	$0.89 \frac{2\pi}{N}$
Welch <sup>†</sup>	$\frac{9}{8} \frac{1}{K}$	$1.28 \frac{2\pi}{L}$	$0.72 \frac{2\pi}{N}$
Blackman-Tukey	$\frac{2}{3} \frac{M}{N}$	$0.64 \frac{2\pi}{M}$	$0.43 \frac{2\pi}{N}$

<sup>†</sup> 50% overlap and a Bartlett window.

### 8.3 MINIMUM VARIANCE SPECTRUM ESTIMATION

Up to this point, we have been considering nonparametric techniques for estimating the power spectrum of a random process. Relying on the DTFT of an estimated autocorrelation sequence, the performance of these methods is limited by the length of the data record. In this section, we develop the Minimum Variance (MV) method of spectrum estimation, which is an adaptation of the Maximum Likelihood Method (MLM) developed by Capon for the analysis of two-dimensional power spectral densities [8]. In the MV method, the power spectrum is estimated by filtering a process with a bank of narrowband bandpass filters. The motivation for this approach may be seen by looking, once again, at the effect of filtering a WSS random process with a narrowband bandpass filter. Therefore, let  $x(n)$  be a zero mean wide-sense stationary random process with a power spectrum  $P_x(e^{j\omega})$  and let  $g_i(n)$  be an ideal bandpass filter with a bandwidth  $\Delta$  and center frequency  $\omega_i$ ,

$$|G_i(e^{j\omega})| = \begin{cases} 1 & ; \quad |\omega - \omega_i| < \Delta/2 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

If  $x(n)$  is filtered with  $g_i(n)$ , then the power spectrum of the output process,  $y_i(n)$ , is

$$P_i(e^{j\omega}) = P_x(e^{j\omega})|G_i(e^{j\omega})|^2$$

and the power in  $y_i(n)$  is

$$\begin{aligned} E\{|y_i(n)|^2\} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} P_i(e^{j\omega}) d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x(e^{j\omega}) |G_i(e^{j\omega})|^2 d\omega \\ &= \frac{1}{2\pi} \int_{\omega_i - \Delta/2}^{\omega_i + \Delta/2} P_x(e^{j\omega}) d\omega \end{aligned} \quad (8.80)$$

If  $\Delta$  is small enough so that  $P_x(e^{j\omega})$  is approximately constant over the passband of the filter, then the power in  $y_i(n)$  is approximately

$$E\{|y_i(n)|^2\} \approx P_x(e^{j\omega_i}) \frac{\Delta}{2\pi} \quad (8.81)$$

Therefore, it is possible to estimate the power spectral density of  $x(n)$  at frequency  $\omega = \omega_i$  from the filtered process by estimating the power in  $y_i(n)$  and dividing by the normalized filter bandwidth,  $\Delta/2\pi$ ,

$$\hat{P}_x(e^{j\omega_i}) = \frac{E\{|y_i(n)|^2\}}{\Delta/2\pi} \quad (8.82)$$

As we saw in Section 8.2.1, the periodogram produces an estimate of the power spectrum in a similar fashion. Specifically,  $x(n)$  is filtered with a bank of bandpass filters,  $h_i(n)$ , where

$$|H_i(e^{j\omega})| = \frac{\sin[N(\omega - \omega_i)/2]}{N \sin[(\omega - \omega_i)/2]} \quad (8.83)$$

and the power in each of the filtered signals is estimated using a one-point sample average,

$$\hat{E}\{|y_i(n)|^2\} = |y_i(N-1)|^2$$

The periodogram is then formed by dividing this power estimate by the filter bandwidth,  $\Delta = 2\pi/N$ . Since each filter in the filter bank for the periodogram is the same, differing only in the center frequency, these filters are *data independent*. As a result, when a random process contains a significant amount of power in frequency bands within the sidelobes of the bandpass filter, leakage through the sidelobes will lead to significant distortion in the power estimates. Therefore, a better approach would be to allow each filter in the filter bank to be *data adaptive* so that each filter may be designed to be “optimum” in the sense of rejecting as much out-of-band signal power as possible. The minimum variance spectrum estimation technique described in this section is based on this idea and involves the following steps:

1. Design a bank of bandpass filters  $g_i(n)$  with center frequency  $\omega_i$  so that each filter rejects the maximum amount of out-of-band power while passing the component at frequency  $\omega_i$  with no distortion.
2. Filter  $x(n)$  with each filter in the filter bank and estimate the power in each output process  $y_i(n)$ .
3. Set  $\hat{P}_x(e^{j\omega_i})$  equal to the power estimated in step (2) divided by the filter bandwidth.

To derive the minimum variance spectrum estimate, we begin with the design of the bandpass filter bank.

Suppose that we would like to estimate the power spectral density of  $x(n)$  at frequency  $\omega_i$ . Let  $g_i(n)$  be a complex-valued FIR bandpass filter of order  $p$ . To ensure that the filter does not alter the power in the input process at frequency  $\omega_i$ ,  $G_i(e^{j\omega})$  will be constrained to have a gain of one at  $\omega = \omega_i$ ,

$$G_i(e^{j\omega_i}) = \sum_{n=0}^p g_i(n)e^{-jn\omega_i} = 1 \quad (8.84)$$

Let  $\mathbf{g}_i$  be the vector of filter coefficients  $g_i(n)$ ,

$$\mathbf{g}_i = [g_i(0), g_i(1), \dots, g_i(p)]^T$$

and let  $\mathbf{e}_i$  be the vector of complex exponentials  $e^{jk\omega_i}$ ,

$$\mathbf{e}_i = [1, e^{j\omega_i}, \dots, e^{jp\omega_i}]^T$$

The constraint on the frequency response given in Eq. (8.84) may be written in vector form as follows<sup>7</sup>

$$\mathbf{g}_i^H \mathbf{e}_i = \mathbf{e}_i^H \mathbf{g}_i = 1 \quad (8.85)$$

Now, in order for the power spectrum of  $x(n)$  at frequency  $\omega_i$  to be measured as accurately as possible, the bandpass filter should reject as much out-of-band power as possible. Therefore, the criterion that will be used to design the bandpass filter will be to minimize the power in the output process subject to the linear constraint given in Eq. (8.85). Since the power in  $y_i(n)$  may be expressed in terms of the autocorrelation matrix  $\mathbf{R}_x$  as follows (see Eq. (3.90) on p. 101)

$$E\{|y_i(n)|^2\} = \mathbf{g}_i^H \mathbf{R}_x \mathbf{g}_i \quad (8.86)$$

the filter design problem becomes one of minimizing Eq. (8.86) subject to the linear constraint given in Eq. (8.85). As we saw in Section 2.3.10, the solution to this problem is

$$\mathbf{g}_i = \frac{\mathbf{R}_x^{-1} \mathbf{e}_i}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i} \quad (8.87)$$

where the minimum value of  $E\{|y_i(n)|^2\}$  is equal to

$$\min_{\mathbf{g}_i} E\{|y_i(n)|^2\} = \frac{1}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i} \quad (8.88)$$

Thus, Eq. (8.87) defines the optimum filter for estimating the power in  $x(n)$  at frequency  $\omega_i$ , and Eq. (8.88) gives the power in  $y_i(n)$ , which is used as the estimate,  $\hat{\sigma}_x^2(\omega_i)$ , of the power in  $x(n)$  at frequency  $\omega_i$ . Note that although these equations were derived for a specific frequency  $\omega_i$ , since this frequency was arbitrary, then these equations are valid for all  $\omega$ . Thus, the optimum filter for estimating the power in  $x(n)$  at frequency  $\omega$  is

$$\boxed{\mathbf{g} = \frac{\mathbf{R}_x^{-1} \mathbf{e}}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}}} \quad (8.89)$$

and the power estimate is

$$\boxed{\hat{\sigma}_x^2(\omega) = \frac{1}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}}} \quad (8.90)$$

where  $\mathbf{e} = [1, e^{j\omega}, \dots, e^{jp\omega}]^T$ .

### Example 8.3.1 White Noise

Let us consider using Eq. (8.90) to estimate the power in white noise that has a variance of  $\sigma_x^2$ . Since the autocorrelation matrix is  $\mathbf{R}_x = \sigma_x^2 \mathbf{I}$ , then the minimum variance bandpass filter is

$$\mathbf{g} = \frac{\mathbf{R}_x^{-1} \mathbf{e}}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}} = \frac{\sigma_x^{-2} \mathbf{e}}{\sigma_x^{-2} \mathbf{e}^H \mathbf{e}} = \frac{1}{p+1} \mathbf{e}$$

<sup>7</sup>The first equality follows from the fact that  $\mathbf{g}_i^H \mathbf{e}_i$  is constrained to be equal to one and, therefore, is real-valued.

which is the same, to within a constant, as the filter in Eq. (8.11) that appears in the filter bank interpretation of the periodogram. From Eq. (8.90) it follows that the estimate of the power in  $x(n)$  at frequency  $\omega$  is

$$\hat{\sigma}_x^2(\omega) = \frac{1}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}} = \frac{1}{p+1} \sigma_x^2$$

which is independent of  $\omega$ . Therefore, the distribution of power in  $x(n)$  is a constant. In addition, note that the estimated power decreases as the filter order,  $p$ , increases. This is a result of the fact that, as the order of the bandpass filter increases, the bandwidth decreases and less power is allowed to pass through the filter. Since the power spectral density of white noise is constant, the total power over a frequency band of width  $\Delta\omega$  is  $\sigma_x^2 \Delta\omega$ , which goes to zero as  $\Delta\omega$  goes to zero. As we will see, what must be done in order to obtain an estimate of the power spectrum is to divide the power estimate by the bandwidth of the filter.

Having designed the bandpass filter bank and estimated the distribution of power in  $x(n)$  as a function of frequency, we may now estimate the power spectrum by dividing the power estimate by the bandwidth of the bandpass filter. Although there are several different criteria that may be used to define bandwidth, the simplest is to use the value of  $\Delta$  that produces the correct power spectral density for white noise. Since the minimum variance estimate of the power in white noise is  $E\{|y_i(n)|^2\} = \sigma_x^2/(p+1)$ , it follows from Eq. (8.82) that the spectrum estimate is

$$\hat{P}_x(e^{j\omega_i}) = \frac{E\{|y_i(n)|^2\}}{\Delta/2\pi} = \frac{\sigma_x^2}{p+1} \frac{2\pi}{\Delta} \quad (8.91)$$

Therefore, if we set

$$\Delta = \frac{2\pi}{p+1} \quad (8.92)$$

then  $\hat{P}_x(e^{j\omega}) = \sigma_x^2$ . Using Eq. (8.92) as the bandwidth of the filter  $g(n)$ , the power spectrum estimate becomes, in general,

$$\hat{P}_{MV}(e^{j\omega}) = \frac{p+1}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}}$$

(8.93)

which is the *minimum variance spectrum estimate*. Note that  $\hat{P}_{MV}(e^{j\omega})$  is defined in terms of the autocorrelation matrix  $\mathbf{R}_x$ , of  $x(n)$ . If, as is normally the case, the autocorrelation matrix is unknown, then  $\mathbf{R}_x$  may be replaced with an estimated autocorrelation matrix,  $\hat{\mathbf{R}}_x$ . A MATLAB program for computing the MV spectrum estimate is given in Fig. 8.19.

From a computational point of view, the minimum variance spectrum estimate requires the inversion of the autocorrelation matrix  $\mathbf{R}_x$  (or  $\hat{\mathbf{R}}_x$ ). Since  $\mathbf{R}_x$  is Toeplitz, the inverse may be found using either the Levinson recursion or the Cholesky decomposition. Once the inverse has been found, the quadratic form  $\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}$  must be evaluated, which may be done

**The Minimum Variance Method**

```

function Px = minvar(x,p)
%
x = x(:);
R = covar(x,p);
[v,d]=eig(R);
U = diag(inv(abs(d)+eps));
V = abs(fft(v,1024)).^2;
Px = 10*log10(p)-10*log10(V*U);
end;

```

**Figure 8.19** A MATLAB program for estimating the power spectrum using the minimum variance method.

efficiently as follows. Let  $v_x(k, l)$  denote the  $(k, l)$ th entry in the inverse of  $\mathbf{R}_x$

$$\mathbf{R}_x^{-1} = \begin{bmatrix} v_x(0, 0) & v_x(0, 1) & \cdots & v_x(0, p) \\ v_x(1, 0) & v_x(1, 1) & \cdots & v_x(1, p) \\ \vdots & \vdots & & \vdots \\ v_x(p, 0) & v_x(p, 1) & \cdots & v_x(p, p) \end{bmatrix}$$

The quadratic form is

$$\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e} = \sum_{k=0}^p \sum_{l=0}^p e^{-jk\omega} v_x(k, l) e^{jl\omega} = \sum_{k=0}^p \sum_{l=0}^p v_x(k, l) e^{j(l-k)\omega} \quad (8.94)$$

which may be written as

$$\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e} = \sum_{n=-p}^p \left[ \sum_{k=\max(0,n)}^{\min(p,p+n)} v_x(k, k-n) \right] e^{-jn\omega}$$

Note that the expression in brackets,

$$q(n) = \sum_{k=\max(0,n)}^{\min(p,p+n)} v_x(k, k-n)$$

is the sequence that is formed by summing the terms in the inverse matrix along the diagonals, i.e.,  $q(0)$  is the sum along the main diagonal,  $q(1)$  is the sum along the first diagonal below the main diagonal,  $q(-1)$  is the sum along the first diagonal above the main diagonal, and so on. Therefore,

$$\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e} = \sum_{n=-p}^p q(n) e^{-jn\omega}$$

which is the discrete-time Fourier transform of the sequence  $q(n)$ . Since  $\mathbf{R}_x^{-1}$  is Hermitian, then  $q(n) = q^*(-n)$  and we also have

$$\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e} = q(0) + 2 \sum_{n=1}^p \operatorname{Re} \left\{ q(n) e^{-jn\omega} \right\}$$

or, if  $r_x(k)$  is real,

$$\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e} = q(0) + 2 \sum_{n=1}^p q(n) \cos(n\omega)$$

In the following example, we consider the minimum variance estimate of an AR(1) process.

---

**Example 8.3.2** *The MV Estimate of an AR(1) Process*

Let  $x(n)$  be an AR(1) random process with autocorrelation

$$r_x(k) = \frac{1}{1 - \alpha^2} \alpha^{|k|}$$

where  $|\alpha| < 1$ . The power spectrum of  $x(n)$  is

$$P_x(e^{j\omega}) = \frac{1}{|1 - \alpha e^{-j\omega}|^2} = \frac{1}{1 + \alpha^2 - 2\alpha \cos \omega}$$

Given  $r_x(k)$  for  $|k| \leq p$ , the  $p$ th-order minimum variance spectrum estimate is

$$\hat{P}_{MV}(e^{j\omega}) = \frac{p+1}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}}$$

where  $\mathbf{R}_x = \frac{1}{1 - \alpha^2} \text{Toep}\{1, \alpha, \dots, \alpha^p\}$ . From Example 5.2.11 (p. 261) it follows that the inverse of  $\mathbf{R}_x$  is

$$\mathbf{R}_x^{-1} = \begin{bmatrix} 1 & -\alpha & 0 & \cdots & 0 & 0 \\ -\alpha & 1 + \alpha^2 & -\alpha & \cdots & 0 & 0 \\ 0 & -\alpha & 1 + \alpha^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 + \alpha^2 & -\alpha \\ 0 & 0 & 0 & \cdots & -\alpha & 1 \end{bmatrix}$$

Therefore,

$$q(0) = [2 + (p-1)(1 + \alpha^2)]$$

and

$$q(1) = -\alpha p$$

so the minimum variance estimate is

$$\hat{P}_{MV}(e^{j\omega}) = \frac{(p+1)}{2 + (p-1)(1 + \alpha^2) - 2\alpha p \cos \omega}$$

Note that  $\hat{P}_{MV}(e^{j\omega})$  converges to  $P_x(e^{j\omega})$  as  $p \rightarrow \infty$ .

---

In a number of applications it is important to be able to estimate the frequency of a sinusoid or a complex exponential in noise. The following example considers the use of the minimum variance method for frequency estimation.

---

**Example 8.3.3** *The MV Estimate of a Complex Exponential in Noise*

Let  $x(n)$  be a random phase complex exponential in white noise

$$x(n) = A_1 e^{jn\omega_1} + w(n)$$

where  $A_1 = |A_1|e^{j\phi}$  with  $\phi$  a random variable that is uniformly distributed over the interval  $[-\pi, \pi]$ . If the variance of  $w(n)$  is  $\sigma_w^2$ , then the autocorrelation of  $x(n)$  is

$$r_x(k) = P_1 e^{jk\omega_1} + \sigma_w^2 \delta(k)$$

where  $P_1 = |A_1|^2$ . Therefore, we may write the autocorrelation matrix as follows:

$$\mathbf{R}_x = P_1 \mathbf{e}_1 \mathbf{e}_1^H + \sigma_w^2 \mathbf{I}$$

where  $\mathbf{e}_1 = [1, e^{j\omega_1}, \dots, e^{jp\omega_1}]^T$ . Using Woodbury's identity (see p. 29) it follows that the inverse of  $\mathbf{R}_x$  is

$$\mathbf{R}_x^{-1} = \frac{1}{\sigma_w^2} \mathbf{I} - \frac{\frac{1}{P_1} \mathbf{e}_1 \mathbf{e}_1^H}{1 + \frac{P_1}{\sigma_w^2} \mathbf{e}_1^H \mathbf{e}_1} = \frac{1}{\sigma_w^2} \left[ \mathbf{I} - \frac{P_1}{\sigma_w^2 + (p+1)P_1} \mathbf{e}_1 \mathbf{e}_1^H \right]$$

(also see Example 2.3.9 on p. 45). Thus,

$$\begin{aligned} \hat{P}_{MV}(e^{j\omega}) &= \frac{p+1}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}} = \frac{p+1}{\frac{1}{\sigma_w^2} \mathbf{e}^H \left[ \mathbf{I} - \frac{P_1}{\sigma_w^2 + (p+1)P_1} \mathbf{e}_1 \mathbf{e}_1^H \right] \mathbf{e}} \\ &= \frac{\sigma_w^2}{1 - \frac{P_1/(p+1)}{\sigma_w^2 + (p+1)P_1} |\mathbf{e}^H \mathbf{e}_1|^2} \end{aligned} \quad (8.95)$$

Since

$$\mathbf{e}^H \mathbf{e}_1 = \sum_{k=0}^p e^{-jk\omega} e^{jk\omega_1} = \sum_{k=0}^p e^{-jk(\omega-\omega_1)} = W_R(e^{j(\omega-\omega_1)})$$

where  $W_R(e^{j\omega})$  is the discrete-time Fourier transform of a rectangular window that extends from  $k = 0$  to  $k = p$ , then the minimum variance estimate is

$$\hat{P}_{MV}(e^{j\omega}) = \frac{\sigma_w^2}{1 - \frac{P_1/(p+1)}{\sigma_w^2 + (p+1)P_1} |W_R(e^{j(\omega-\omega_1)})|^2}$$

From this expression we see that the minimum variance estimate attains its maximum at  $\omega = \omega_1$  with

$$\hat{P}_{MV}(e^{j\omega}) \Big|_{\omega=\omega_1} = \frac{\sigma_w^2}{1 - \frac{(p+1)P_1}{\sigma_w^2 + (p+1)P_1}} = \sigma_w^2 + (p+1)P_1$$

Therefore, the estimate of the power in  $x(n)$  at frequency  $\omega = \omega_1$  is

$$\hat{\sigma}_x^2(\omega_1) = \frac{1}{p+1} \hat{P}_{MV}(e^{j\omega_1}) = \frac{\sigma_w^2}{p+1} + P_1$$

Thus, if the signal-to-noise ratio is large,  $P_1 \gg \sigma_w^2$ , then the minimum variance estimate of the power at  $\omega = \omega_1$  becomes

$$\hat{\sigma}_x^2(\omega_1) = P_1$$

Furthermore, if  $p \gg 1$  and  $\omega \neq \omega_1$ , then  $W_R(e^{j(\omega-\omega_1)}) \approx 0$  and the minimum variance estimate of the power spectrum becomes

$$\hat{P}_{MV}(e^{j\omega}) \approx \sigma_w^2$$

One issue that we have not yet addressed concerns the question of how to select the filter order,  $p$ . Clearly, the larger the order of the filter the better the filter will be in rejecting out-of-band power. Therefore, with this in mind, the order of the filter should be as large as possible. From a practical point of view, however, there is a limit on how large the filter order may be. Specifically, note that for a  $p$ th-order filter the MV spectrum estimate requires the evaluation of the inverse of a  $(p+1) \times (p+1)$  autocorrelation matrix  $\mathbf{R}_x$ . However, in order to be able to invert this matrix, it is necessary that  $r_x(k)$  be known or estimated for  $k = 0, 1, \dots, p$ . Therefore, for a fixed data record length,  $N$ , since it is only possible to estimate  $r_x(k)$  for  $k = 0, 1, \dots, N-1$ , the filter order is limited to  $p \leq N$ . In practice, however, the filter order will generally be much smaller than this due to the large variance of the autocorrelation estimates for values of  $k$  that are close to  $N$ .

In addition to the approach described above, there are other forms of minimum variance spectrum estimates that differ either in the way that the filter is designed or in the way that the power in the output signal is computed [12,31,54]. For example, in the Data Adaptive Spectrum Estimation (DASE) algorithm [12],  $G_i(e^{j\omega})$  is constrained to have unit energy within a passband of width  $\Delta$  that is centered at frequency  $\omega_i$ ,

$$\frac{1}{\Delta} \int_{\omega_i - \Delta/2}^{\omega_i + \Delta/2} |G_i(e^{j\omega})|^2 d\omega = 1$$

Solving the constrained minimization problem in this case leads to a generalized eigenvalue problem.

## **8.4 THE MAXIMUM ENTROPY METHOD**

One of the limitations with the classical approach to spectrum estimation is that, for a data record of length  $N$ , the autocorrelation sequence can only be estimated for lags  $|k| < N$ . As a result,  $\hat{r}_x(k)$  is set to zero for  $|k| \geq N$ . Since many signals of interest have autocorrelations that are nonzero for  $|k| \geq N$ , this windowing may significantly limit the resolution and accuracy of the estimated spectrum. This is particularly true, for example, in the case of narrowband processes that have autocorrelations that decay slowly with  $k$ . Since the classical methods effectively extrapolate the autocorrelation sequence with zeros, if it were possible to perform a more accurate extrapolation of the autocorrelation sequence, then the effects of the window could be mitigated and a more accurate estimate of the spectrum could be found. A difficult question to answer, however, is the following: How should this extrapolation be performed? The maximum entropy method (MEM) that is developed in this section suggests one possible way to perform this extrapolation.

Given the autocorrelation  $r_x(k)$  of a WSS process for lags  $|k| \leq p$ , the problem that we wish to address, illustrated in Fig. 8.20, is how to extrapolate  $r_x(k)$  for  $|k| > p$ . Denoting the extrapolated values by  $r_e(k)$ , it is clear that some constraints should be placed on  $r_e(k)$ . For example, if

$$P_x(e^{j\omega}) = \sum_{k=-p}^p r_x(k)e^{-jk\omega} + \sum_{|k|>p} r_e(k)e^{-jk\omega} \quad (8.96)$$

then  $P_x(e^{j\omega})$  should correspond to a valid power spectrum, i.e.,  $P_x(e^{j\omega})$  should be real-valued and nonnegative for all  $\omega$ . In general, however, only constraining  $P_x(e^{j\omega})$  to be real and nonnegative is not sufficient to guarantee a unique extrapolation. Therefore, some additional constraints must be imposed on the set of allowable extrapolations.<sup>8</sup> One such constraint, proposed by Burg [7], is to perform the extrapolation in such a way so as to maximize the entropy of the process.<sup>9</sup> Since entropy is a measure of randomness or uncertainty, a maximum entropy extrapolation is equivalent to finding the sequence of autocorrelations,  $r_e(k)$ , that make  $x(n)$  as *white* (random) as possible. In some sense, such an extrapolation places as few constraints as possible or the least amount of structure on  $x(n)$ . In terms of the power spectrum, this corresponds to the constraint that  $P_x(e^{j\omega})$  be “as flat as possible” (see Fig. 8.21).

For a Gaussian random process with power spectrum  $P_x(e^{j\omega})$ , the entropy is [11]

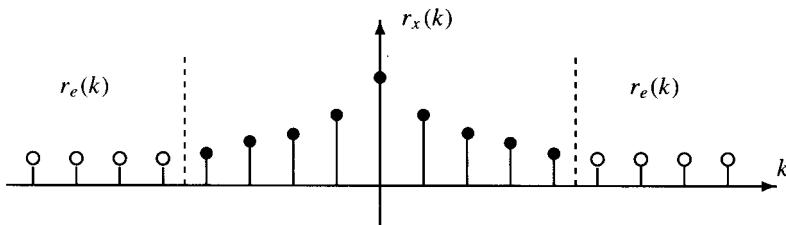
$$H(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln P_x(e^{j\omega}) d\omega \quad (8.97)$$

Therefore, for Gaussian processes with a given partial autocorrelation sequence,  $r_x(k)$  for  $|k| \leq p$ , the maximum entropy power spectrum is the one that maximizes Eq. (8.97) subject to the constraint that the inverse discrete-time Fourier transform of  $P_x(e^{j\omega})$  equals the given set of autocorrelations for  $|k| \leq p$ ,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} P_x(e^{j\omega}) e^{jk\omega} d\omega = r_x(k) \quad ; \quad |k| \leq p \quad (8.98)$$

The values of  $r_e(k)$  that maximize the entropy may be found by setting the derivative of  $H(x)$  with respect to  $r_e^*(k)$  equal to zero as follows

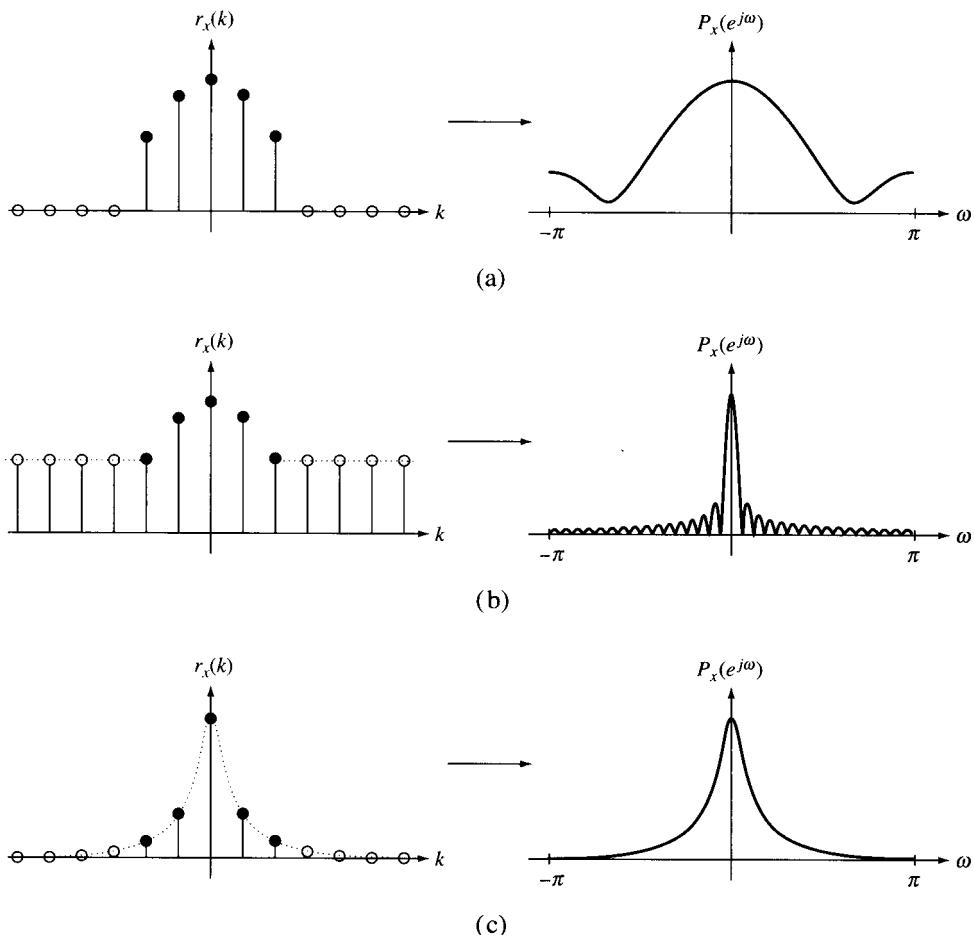
$$\frac{\partial H(x)}{\partial r_e^*(k)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{P_x(e^{j\omega})} \frac{\partial P_x(e^{j\omega})}{\partial r_e^*(k)} d\omega = 0 \quad ; \quad |k| > p \quad (8.99)$$



**Figure 8.20** Extrapolating the autocorrelation sequence.

<sup>8</sup>It is assumed that the given partial autocorrelation sequence is *extendible* so that there is at least one extrapolation that produces a valid power spectrum. See Section 5.2.8 for a discussion of the conditions under which an autocorrelation sequence is extendible.

<sup>9</sup>Entropy has its origins in information theory and is a subject that is well outside the scope of this book. For a treatment of entropy the reader may consult [4, 11, 40].



**Figure 8.21** Different extrapolations of a partial autocorrelation sequence and the corresponding power spectral densities. The problem is to find the extrapolation that produces a spectrum that is as flat as possible.

From Eq. (8.96) we see that<sup>10</sup>

$$\frac{\partial P_x(e^{j\omega})}{\partial r_e^*(k)} = e^{jk\omega}$$

which, when substituted into Eq. (8.99), yields

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{P_x(e^{j\omega})} e^{jk\omega} d\omega = 0 \quad ; \quad |k| > p \quad (8.100)$$

Defining  $Q_x(e^{j\omega}) = 1/P_x(e^{j\omega})$ , Eq. (8.100) states that the inverse discrete-time Fourier transform of  $Q_x(e^{j\omega})$  is a finite-length sequence that is equal to zero for  $|k| > p$ ,

$$q_x(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} Q_x(e^{j\omega}) e^{jk\omega} d\omega = 0 \quad ; \quad |k| > p \quad (8.101)$$

<sup>10</sup>Recall that  $r_x(k)$  is conjugate symmetric so  $r_x(-k) = r_x^*(k)$ .

Therefore,

$$Q_x(e^{j\omega}) = \frac{1}{P_x(e^{j\omega})} = \sum_{k=-p}^p q_x(k) e^{-jk\omega}$$

and it follows that the maximum entropy power spectrum for a Gaussian process, which we will denote by  $\hat{P}_{mem}(e^{j\omega})$ , is an all-pole power spectrum,

$$\hat{P}_{mem}(e^{j\omega}) = \frac{1}{\sum_{k=-p}^p q_x(k) e^{-jk\omega}} \quad (8.102)$$

Using the spectral factorization theorem, it follows that Eq. (8.102) may be expressed as

$$\hat{P}_{mem}(e^{j\omega}) = \frac{|b(0)|^2}{A_p(e^{j\omega}) A_p^*(e^{j\omega})} = \frac{|b(0)|^2}{\left| 1 + \sum_{k=1}^p a_p(k) e^{-jk\omega} \right|^2}$$

Alternatively, in terms of the vectors  $\mathbf{a}_p = [1, a_p(1), \dots, a_p(p)]^T$  and  $\mathbf{e} = [1, e^{j\omega}, \dots, e^{jp\omega}]^T$ , the MEM spectrum may be written as

$$\hat{P}_{mem}(e^{j\omega}) = \frac{|b(0)|^2}{|\mathbf{e}^H \mathbf{a}_p|^2} \quad (8.103)$$

Having determined the form of the MEM spectrum, all that remains is to find the coefficients  $a_p(k)$  and  $b(0)$ . Due to the constraint given in Eq. (8.98), these coefficients must be chosen in such a way that the inverse discrete-time Fourier transform of  $\hat{P}_{mem}(e^{j\omega})$  produces an autocorrelation sequence that matches the given values of  $r_x(k)$  for  $|k| \leq p$ . As we saw in Section 5.2.3, if the coefficients  $a_p(k)$  are the solution to the autocorrelation normal equations

$$\begin{bmatrix} r_x(0) & r_x^*(1) & \cdots & r_x^*(p) \\ r_x(1) & r_x(0) & \cdots & r_x^*(p-1) \\ \vdots & \vdots & & \vdots \\ r_x(p) & r_x(p-1) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_p(1) \\ \vdots \\ a_p(p) \end{bmatrix} = \epsilon_p \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (8.104)$$

and if

$$|b(0)|^2 = r_x(0) + \sum_{k=1}^p a_p(k) r_x^*(k) = \epsilon_p \quad (8.105)$$

then the autocorrelation matching constraint given in Eq. (8.98) will be satisfied. Thus, the MEM spectrum is

$$\hat{P}_{mem}(e^{j\omega}) = \frac{\epsilon_p}{|\mathbf{e}^H \mathbf{a}_p|^2} \quad (8.106)$$

where  $\mathbf{a}_p$  is the solution to Eq. (8.104). In summary, given a sequence of autocorrelations,  $r_x(k)$  for  $k = 0, 1, \dots, p$ , the MEM spectrum is computed as follows. First, the autocorrelation normal equations (8.104) are solved for the all-pole coefficients  $a_p(k)$  and  $\epsilon_p$ . Then, the MEM spectrum is formed by incorporating these parameters into Eq. (8.106). Note that since  $\hat{P}_{mem}(e^{j\omega})$  is an all-pole power spectrum, then  $r_x(k)$  satisfies the Yule-Walker

***The Maximum Entropy Method***

```

function Px = mem(x,p)
%
[a,e] = acm(x,p);
Px = 20*(log10(e)-log10(abs(fft(a,1024)))); 
end;

```

**Figure 8.22** A MATLAB program for estimating the power spectrum using the maximum entropy method. Note that this m-file calls acm.m.

equations

$$r_x(l) = - \sum_{k=1}^p a_p(k) r_x(k-l) \quad \text{for } l > 0 \quad (8.107)$$

Therefore, the maximum entropy method extrapolates the autocorrelation sequence according to this recursion. A MATLAB program for computing the MEM spectrum is given in Fig. 8.22.

The properties of the maximum entropy method have been studied extensively and, as a spectrum analysis tool, the maximum entropy method is subject to different interpretations [10,29]. It may be argued, for example, that in the absence of any information or constraints on a process  $x(n)$ , given a set of autocorrelation values,  $r_x(0), \dots, r_x(p)$ , the best way to estimate the power spectrum is to Fourier Transform the autocorrelation sequence formed from the given values along with an extrapolation that imposes the least amount of structure on the data, i.e., perform a maximum entropy extrapolation. This would seem to be preferable to an extrapolation that somewhat arbitrarily sets  $r_x(k) = 0$  for  $|k| > p$  as in the classical approach. On the other hand, it may also be argued that since the maximum entropy extrapolation imposes an all-pole model on the data, unless the process is known to be consistent with this model, then the estimated spectrum may not be very accurate. Whether or not the MEM estimate is “better” than the minimum variance method or an estimate produced using a classical approach depends critically on what type of process is being analyzed and on how closely the process may be modeled as an autoregressive process.

---

**Example 8.4.1 The MEM Estimate of a Complex Exponential in Noise**

In Example 8.3.3, we found the minimum variance estimate of a complex exponential in white noise

$$x(n) = A_1 e^{jn\omega_1} + w(n)$$

where  $A_1 = |A_1|e^{j\phi}$  with  $\phi$  being a random variable that is uniformly distributed over the interval  $[-\pi, \pi]$  and with the variance of  $w(n)$  being equal to  $\sigma_w^2$ . To find the  $p$ th-order MEM spectrum, we must solve the autocorrelation normal equations (8.104) for the AR coefficients  $a_p(k)$ . The  $(p+1) \times (p+1)$  autocorrelation matrix for  $x(n)$  is

$$\mathbf{R}_x = P_1 \mathbf{e}_1 \mathbf{e}_1^H + \sigma_w^2 \mathbf{I}$$

where  $\mathbf{e}_1 = [1, e^{j\omega_1}, \dots, e^{jp\omega_1}]^T$  and  $P_1 = |A_1|^2$ . As we saw in Example 8.3.3, the inverse

of  $\mathbf{R}_x$  is

$$\mathbf{R}_x^{-1} = \frac{1}{\sigma_w^2} \left[ \mathbf{I} - \frac{P_1}{\sigma_w^2 + (p+1)P_1} \mathbf{e}_1 \mathbf{e}_1^H \right]$$

Therefore,

$$\mathbf{a}_p = \epsilon_p \mathbf{R}_x^{-1} \mathbf{u}_1 = \frac{\epsilon_p}{\sigma_w^2} \left[ \mathbf{u}_1 - \frac{P_1}{\sigma_w^2 + (p+1)P_1} \mathbf{e}_1 \right] \quad (8.108)$$

where  $\mathbf{u}_1 = [1, 0, \dots, 0]^T$ , and the MEM spectrum is

$$\begin{aligned} \hat{P}_{mem}(e^{j\omega}) &= \frac{\epsilon_p}{|\mathbf{e}^H \mathbf{a}_p|^2} = \frac{\epsilon_p}{\left( \frac{\epsilon_p}{\sigma_w^2} \right)^2 \left| \mathbf{e}^H \left( \mathbf{u}_1 - \frac{P_1}{\sigma_w^2 + (p+1)P_1} \mathbf{e}_1 \right) \right|^2} \\ &= \frac{\epsilon_p}{\left( \frac{\epsilon_p}{\sigma_w^2} \right)^2 \left| 1 - \frac{P_1}{\sigma_w^2 + (p+1)P_1} \mathbf{e}^H \mathbf{e}_1 \right|^2} \end{aligned}$$

Since

$$\mathbf{e}^H \mathbf{e}_1 = \sum_{k=0}^p e^{-jk(\omega - \omega_1)} = W_R(e^{j(\omega - \omega_1)})$$

where  $W_R(e^{j\omega})$  is the discrete-time Fourier transform of a rectangular window, then the MEM spectrum becomes

$$\hat{P}_{mem}(e^{j\omega}) = \frac{\epsilon_p}{\left( \frac{\epsilon_p}{\sigma_w^2} \right)^2 \left| 1 - \frac{P_1}{\sigma_w^2 + (p+1)P_1} W_R(e^{j(\omega - \omega_1)}) \right|^2}$$

To determine the value of  $\epsilon_p$ , note that since  $a_p(0) = 1$ , then it follows from Eq. (8.108) that

$$a_p(0) = \frac{\epsilon_p}{\sigma_w^2} \left[ 1 - \frac{P_1}{\sigma_w^2 + (p+1)P_1} \right] = 1$$

Solving for  $\epsilon_p$  we have

$$\epsilon_p = \sigma_w^2 \left[ 1 + \frac{P_1}{\sigma_w^2 + pP_1} \right]$$

and the MEM spectrum becomes

$$\hat{P}_{mem}(e^{j\omega}) = \frac{\sigma_w^2 \left[ 1 - \frac{P_1}{\sigma_w^2 + (p+1)P_1} \right]}{\left| 1 - \frac{P_1}{\sigma_w^2 + (p+1)P_1} W_R(e^{j(\omega - \omega_1)}) \right|^2}$$

As with the minimum variance method, the peak of the MEM spectrum occurs at frequency  $\omega = \omega_1$ , and

$$\hat{P}_{mem}(e^{j\omega}) \Big|_{\omega=\omega_1} = \frac{\sigma_w^2 \left[ 1 - \frac{P_1}{\sigma_w^2 + (p+1)P_1} \right]}{\left| 1 - \frac{(p+1)P_1}{\sigma_w^2 + (p+1)P_1} \right|^2} = \frac{1}{\sigma_w^2} [\sigma_w^2 + pP_1][\sigma_w^2 + (p+1)P_1]$$

If the signal-to-noise ratio is large,  $P_1 \gg \sigma_w^2$ , then the MEM spectrum estimate at  $\omega = \omega_1$  is approximately

$$\hat{P}_{mem}(e^{j\omega}) \Big|_{\omega=\omega_1} \approx p^2 \frac{P_1^2}{\sigma_w^2}$$

Thus, the peak in the MEM spectrum is proportional to the *square* of the power in the complex exponential.

There is an interesting relationship that exists between the MEM and MV spectrum estimates. This relationship states that the  $p$ th-order MV estimate is the harmonic mean of the MEM estimates of orders  $k = 0, 1, \dots, p$ . To derive this relationship, we will use the recursion given in Eq. (5.114) for computing the inverse of a Toeplitz matrix. This recursion provides the following expression for the inverse of the  $(p+1) \times (p+1)$  Toeplitz matrix  $\mathbf{R}_p$ ,

$$\mathbf{R}_p^{-1} = \left[ \begin{array}{c|cccc} 0 & 0 & \cdots & 0 \\ \hline 0 & & & & \\ \vdots & & \mathbf{R}_{p-1}^{-1} & & \\ 0 & & & & \end{array} \right] + \frac{1}{\epsilon_p} \mathbf{a}_p \mathbf{a}_p^H \quad (8.109)$$

For a WSS process with an autocorrelation matrix  $\mathbf{R}_p$ , the  $p$ th-order MV estimate is

$$\hat{P}_{MV}(e^{j\omega}) = \frac{p+1}{\mathbf{e}^H \mathbf{R}_p^{-1} \mathbf{e}} \quad (8.110)$$

Multiplying  $\mathbf{R}_p^{-1}$  on the left by  $\mathbf{e}^H$  and on the right by  $\mathbf{e}$  we have

$$\mathbf{e}^H \mathbf{R}_p^{-1} \mathbf{e} = \mathbf{e}^H \left[ \begin{array}{c|cccc} 0 & 0 & \cdots & 0 \\ \hline 0 & & & & \\ \vdots & & \mathbf{R}_{p-1}^{-1} & & \\ 0 & & & & \end{array} \right] \mathbf{e} + \frac{1}{\epsilon_p} \mathbf{e}^H \mathbf{a}_p \mathbf{a}_p^H \mathbf{e} \quad (8.111)$$

or,<sup>11</sup>

$$\mathbf{e}^H \mathbf{R}_p^{-1} \mathbf{e} = \mathbf{e}^H \mathbf{R}_{p-1}^{-1} \mathbf{e} + \frac{1}{\epsilon_p} |\mathbf{e}^H \mathbf{a}_p|^2 \quad (8.112)$$

Note that the first two terms in Eq. (8.112) are proportional to the inverse of the minimum variance estimates of order  $p$  and  $p-1$ , respectively, and the last term is the reciprocal of the  $p$ th-order MEM spectrum. Therefore, we have

$$\frac{p+1}{\hat{P}_{MV}^{(p)}(e^{j\omega})} = \frac{p}{\hat{P}_{MV}^{(p-1)}(e^{j\omega})} + \frac{1}{\hat{P}_{mem}^{(p)}(e^{j\omega})}$$

which is a recursion for the  $p$ th-order MV estimate in terms of the  $(p-1)$ st-order MV estimate and the  $p$ th-order MEM spectrum. Solving this recursion for  $\hat{P}_{MV}^{(p)}(e^{j\omega})$  we find

$$\frac{1}{\hat{P}_{MV}^{(p)}(e^{j\omega})} = \frac{1}{p+1} \sum_{k=0}^p \frac{1}{\hat{P}_{mem}^{(k)}(e^{j\omega})} \quad (8.113)$$

<sup>11</sup>Note that, in order to keep the notation as simple as possible, we are using  $\mathbf{e}$  to denote a vector of complex exponentials  $e^{jk\omega}$  of varying lengths. Thus, the number of elements in  $\mathbf{e}$  is determined by the size of the vector or matrix that it is multiplied by.

Therefore, the MV estimate is the harmonic mean of the MEM spectra from the low order to the high order estimates. As a result of this smoothing, for WSS processes consisting of narrowband components in noise, the MEM spectrum generally provides a higher resolution spectrum estimate than the minimum variance method.

## 8.5 PARAMETRIC METHODS

One of the limitations of the nonparametric methods of spectrum estimation presented in Sections 8.2 and 8.3 is that they are not designed to incorporate information that may be available about the process into the estimation procedure. In some applications this may be an important limitation, particularly when some knowledge is available about how the data samples are generated. In speech processing, for example, an acoustic tube model for the vocal tract imposes an autoregressive model on the speech waveform [43]. As a result, for intervals of time over which the speech waveform may be assumed to be approximately stationary, the spectrum is of the form,

$$P_x(e^{j\omega}) = \frac{|b(0)|^2}{\left|1 + \sum_{k=1}^p a_p(k)e^{-jk\omega}\right|^2}$$

However, with a nonparametric approach such as the periodogram, the power spectrum has a form that is consistent with a moving average process,

$$\hat{P}_{per}(e^{j\omega}) = |X_N(e^{j\omega})|^2 = \sum_{k=-N}^N \hat{r}_x(k)e^{-jk\omega}$$

Therefore, one would hope that if it were possible to incorporate a model for the process directly into the spectrum estimation algorithm, then a more accurate and higher resolution estimate could be found. This may be easily done using a parametric approach to spectrum estimation. With a parametric approach, the first step is to select an appropriate model for the process. This selection may be based on a priori knowledge about how the process is generated or, perhaps, on experimental results indicating that a particular model “works well.” Models that are commonly used include autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), and harmonic (complex exponentials in noise). Once a model has been selected, the next step is to estimate the model parameters from the given data. The final step is to estimate the power spectrum by incorporating the estimated parameters into the parametric form for the spectrum. For example, with an autoregressive moving average model, with estimates  $\hat{b}_q(k)$  and  $\hat{a}_p(k)$  of the model parameters, the spectrum estimate would be

$$\hat{P}_x(e^{j\omega}) = \frac{\left|\sum_{k=0}^q \hat{b}_q(k)e^{-jk\omega}\right|^2}{\left|1 + \sum_{k=1}^p \hat{a}_p(k)e^{-jk\omega}\right|^2}$$

Although it is possible to significantly improve the resolution of the spectrum estimate with a parametric method, it is important to realize that, unless the model that is used

is appropriate for the process that is being analyzed, inaccurate or misleading estimates may be obtained. Consider, for example, the two estimates shown in Fig. 8.23a that were formed from  $N = 64$  values of a process consisting of two sinusoids in unit variance white noise

$$x(n) = 5 \sin(0.45\pi n + \phi_1) + 5 \sin(0.55\pi n + \phi_2) + w(n)$$

The spectrum estimate shown by the solid line was derived assuming an all-pole model for  $x(n)$  whereas the estimate shown by the dashed line was formed using the Blackman-Tukey method. Clearly, the estimate produced with the model-based approach provides much better resolution than the Blackman-Tukey method. Fig. 8.23b, on the other hand, shows the spectrum estimates that are formed using the same two approaches for the MA(2) process

$$x(n) = w(n) - w(n - 2)$$

where  $w(n)$  is unit variance white noise. The power spectrum of  $x(n)$  is

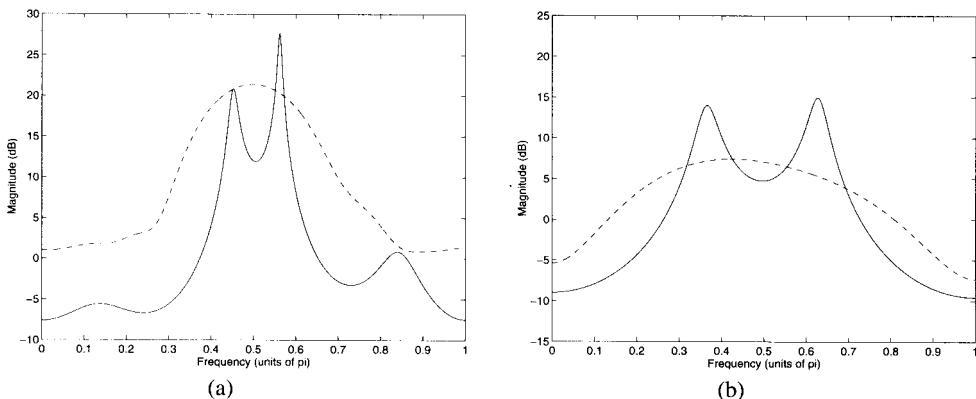
$$P_x(e^{j\omega}) = 2 - 2 \cos 2\omega$$

In this case, the model-based approach inaccurately represents the underlying spectrum, indicating that the process may contain two narrowband components. The Blackman-Tukey method, on the other hand, which makes no assumptions about the process, produces a more accurate estimate of the power spectrum.

In this section, we consider power spectrum estimation using AR, MA, and ARMA models. The problem of frequency estimation for processes consisting of complex exponentials in noise will be considered in Section 8.6.

### 8.5.1 Autoregressive Spectrum Estimation

An autoregressive process,  $x(n)$ , may be represented as the output of an all-pole filter that is driven by unit variance white noise. The power spectrum of a  $p$ th-order autoregressive



**Figure 8.23** Illustration of how an inappropriate model may lead to an inaccurate spectrum estimate. Using a spectrum estimation technique that assumes an all-pole model for the process (solid line) and the Blackman-Tukey method (dashed line), the spectrum estimates are shown for (a) a process consisting of two sinusoids in noise and (b) a second-order moving average process.

process is

$$P_x(e^{j\omega}) = \frac{|b(0)|^2}{\left|1 + \sum_{k=1}^p a_p(k)e^{-jk\omega}\right|^2} \quad (8.114)$$

Therefore, if  $b(0)$  and  $a_p(k)$  can be estimated from the data, then an estimate of the power spectrum may be formed using

$$\hat{P}_{AR}(e^{j\omega}) = \frac{|\hat{b}(0)|^2}{\left|1 + \sum_{k=1}^p \hat{a}_p(k)e^{-jk\omega}\right|^2} \quad (8.115)$$

Clearly, the accuracy of  $\hat{P}_{AR}(e^{j\omega})$  will depend on how accurately the model parameters may be estimated and, even more importantly, on whether or not an autoregressive model is consistent with the way in which the data is generated. If, for example, Eq. (8.115) is applied to a moving average process, then one should expect the estimate to be poor.

Since autoregressive spectrum estimation requires that an all-pole model be found for the process, there is a variety of techniques that may be used to estimate the all-pole parameters. However, once the all-pole parameters have been estimated, each method generates an estimate of the power spectrum in exactly the same way, i.e., using Eq. (8.115). In the following subsections, we briefly review the AR modeling techniques and describe some of the properties of these techniques as they apply to spectrum estimation.

**The Autocorrelation Method.** In the autocorrelation method of all-pole modeling, the AR coefficients  $a_p(k)$  are found by solving the autocorrelation normal equations

$$\begin{bmatrix} r_x(0) & r_x^*(1) & r_x^*(2) & \cdots & r_x^*(p) \\ r_x(1) & r_x(0) & r_x^*(1) & \cdots & r_x^*(p-1) \\ r_x(2) & r_x(1) & r_x(0) & \cdots & r_x^*(p-2) \\ \vdots & \vdots & \vdots & & \vdots \\ r_x(p) & r_x(p-1) & r_x(p-2) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \\ \vdots \\ a_p(p) \end{bmatrix} = \epsilon_p \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (8.116)$$

where

$$r_x(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n+k)x^*(n) \quad ; \quad k = 0, 1, \dots, p \quad (8.117)$$

(for simplicity we are suppressing the hats over the autocorrelations  $r_x(k)$  and all-pole parameters  $a_p(k)$ ). Solving Eq. (8.116) for the coefficients  $a_p(k)$ , setting

$$|b(0)|^2 = \epsilon_p = r_x(0) + \sum_{k=1}^p a_p(k)r_x^*(k) \quad (8.118)$$

and incorporating these parameters into Eq. (8.115) produces an estimate of the power spectrum, which is sometimes referred to as the *Yule-Walker method*.<sup>12</sup> Note that the Yule-Walker method is equivalent to the maximum entropy method. In fact, the only difference between the two methods lies in the assumptions that are made about the process  $x(n)$ .

<sup>12</sup>This name comes from the fact that the autocorrelation normal equations are equivalent to the Yule-Walker equations for autoregressive processes (see Section 3.6.2).

Specifically, with the Yule-Walker method it is assumed that  $x(n)$  is an autoregressive process, whereas with the maximum entropy method it is assumed that  $x(n)$  is Gaussian.

Since the autocorrelation matrix  $\mathbf{R}_x$  in the autocorrelation normal equations is Toeplitz, the Levinson-Durbin recursion may be used to solve these equations for  $a_p(k)$ . Furthermore, if  $\mathbf{R}_x > 0$ , then the roots of  $A_p(z)$  will lie inside the unit circle. However, because the autocorrelation method effectively applies a rectangular window to the data when estimating the autocorrelation sequence using Eq. (8.117), the data is effectively extrapolated with zeros. Therefore, the autocorrelation method generally produces a lower resolution estimate than the approaches that do not window the data, such as the covariance method and Burg's method. Consequently, for short data records the autocorrelation method is not generally used.

An artifact that may be observed with the autocorrelation method is *spectral line splitting*. This artifact involves the splitting of a single spectral peak into two separate and distinct peaks. Typically, spectral line splitting occurs when  $x(n)$  is *overmodeled*, i.e., when  $p$  is too large. An example of spectral line splitting is shown in Fig. 8.24 for an AR(2) process that is generated according to the difference equation

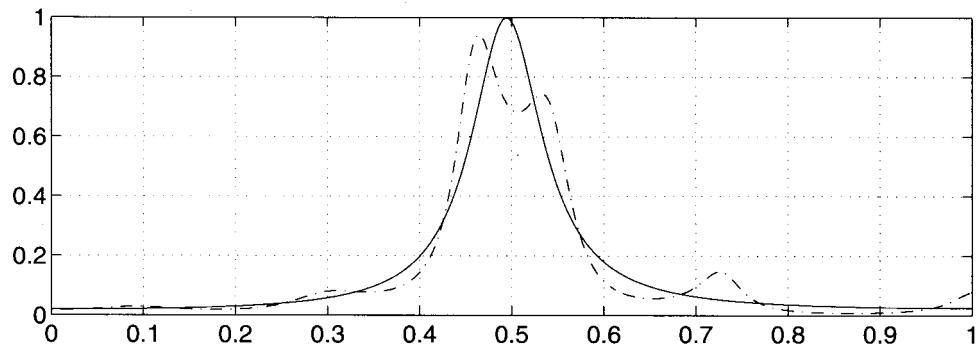
$$x(n) = -0.9x(n-2) + w(n)$$

where  $w(n)$  is unit variance white noise. Using model orders of  $p = 4$  and  $p = 12$  and a data record of length  $N = 64$ , spectrum estimates are shown for the autocorrelation method. What we observe is that, although the true spectrum has a single spectral peak at  $\omega = \pi/2$ , when  $p = 12$  this peak is split into two peaks.<sup>13</sup>

Since the autocorrelation estimate in Eq. (8.117) is biased, a variation of the autocorrelation method is to use the unbiased estimate

$$\hat{r}_x(k) = \frac{1}{N-k} \sum_{n=0}^{N-1-k} x(n+k)x^*(n) \quad ; \quad k = 0, 1, \dots, p \quad (8.119)$$

In this case, however, the autocorrelation matrix is not guaranteed to be positive definite and, as a result, the variance of the spectrum estimate tends to become large when  $\hat{\mathbf{R}}_x$



**Figure 8.24** Spectral line splitting of an AR(2) process. Two all-pole spectrum estimates were computed using the autocorrelation method with orders  $p = 4$  (solid line) and  $p = 12$  (dash-dot line).

<sup>13</sup>Note that for a process such as this, spectral line splitting will not *always* be observed. Whether or not line splitting occurs depends on the specific white noise process that generates  $x(n)$ .

is ill-conditioned or singular [26,37]. Therefore, the biased estimate of  $r_x(k)$  is generally preferred over the unbiased estimate.

**The Covariance Method.** Another approach for estimating the AR parameters is the covariance method. The covariance method requires finding the solution to the set of linear equations,

$$\begin{bmatrix} r_x(1, 1) & r_x(2, 1) & \cdots & r_x(p, 1) \\ r_x(1, 2) & r_x(2, 2) & \cdots & r_x(p, 2) \\ \vdots & \vdots & & \vdots \\ r_x(1, p) & r_x(2, p) & \cdots & r_x(p, p) \end{bmatrix} \begin{bmatrix} a_p(1) \\ a_p(2) \\ \vdots \\ a_p(p) \end{bmatrix} = - \begin{bmatrix} r_x(0, 1) \\ r_x(0, 2) \\ \vdots \\ r_x(0, p) \end{bmatrix} \quad (8.120)$$

where

$$r_x(k, l) = \sum_{n=p}^{N-1} x(n-l)x^*(n-k) \quad (8.121)$$

Unlike the linear equations in the autocorrelation method, these equations are not Toeplitz. However, the advantage of the covariance method over the autocorrelation method is that no windowing of the data is required in the formation of the autocorrelation estimates,  $r_x(k, l)$ . Therefore, for short data records the covariance method generally produces higher resolution spectrum estimates than the autocorrelation method. However, as the data record length increases and becomes large compared to the model order,  $N \gg p$ , the effect of the data window becomes small and the difference between the two approaches becomes negligible.

**The Modified Covariance Method.** The modified covariance method is similar to the covariance method in that no window is applied to the data. However, instead of finding the autoregressive model that minimizes the sum of the squares of the forward prediction error, the modified covariance method minimizes the sum of the squares of the forward and backward prediction errors.<sup>14</sup> As a result, the autoregressive parameters in the modified covariance method are found by solving a set of linear equations of the form given in Eq. (8.120) with

$$r_x(k, l) = \sum_{n=p}^{N-1} [x(n-l)x^*(n-k) + x(n-p+l)x^*(n-p+k)]$$

replacing the estimate in Eq. (8.121). As with the covariance method, the autocorrelation matrix is not Toeplitz.

In contrast to other AR spectrum estimation techniques, the modified covariance method appears to give statistically stable spectrum estimates with high resolution [37,51]. Furthermore, in the spectral analysis of sinusoids in white noise, although the modified covariance method is characterized by a shifting of the peaks from their true locations due to additive noise, this shifting appears to be less pronounced than with other autoregressive estimation techniques [53]. In addition, the peak locations tend to be less sensitive to phase [9,57]. Finally, unlike the previous methods, it appears that the modified covariance method is not subject to spectral line splitting [27].

<sup>14</sup>The modified covariance method has also been referred to as the *Forward-Backward Method* [37] and as the *Least Squares Method* [57].

**The Burg Algorithm.** As with the modified covariance algorithm, the Burg algorithm finds a set of all-pole model parameters that minimizes the sum of the squares of the forward and backward prediction errors. However, in order to assure that the model is stable, this minimization is performed sequentially with respect to the reflection coefficients. Although less accurate than the modified covariance method, since the Burg algorithm does not apply a window to the data, the estimates of the autoregressive parameters are more accurate than those obtained with the autocorrelation method. In the analysis of sinusoids in noise, the Burg algorithm is subject to spectral line splitting and the peak locations are highly dependent upon the phases of the sinusoids [27,35].

**Examples.** It is difficult to provide a complete set of examples to illustrate the properties of the AR spectrum estimation techniques for different types of processes and to compare them to other spectrum estimation algorithms. It would be interesting, for example, to compare the effectiveness of each approach in estimating the spectra of narrowband and wideband autoregressive processes of various orders using different data record lengths. It would also be interesting to look at what happens when these techniques are applied to other types of processes such as MA processes, ARMA processes, and harmonic processes. Rather than attempting to cover all of the possibilities, in the following we only consider the use of AR spectrum estimation techniques to analyze a short data record that is derived from a fourth-order narrowband autoregressive process. Additional experiments with the AR techniques are left to the computer exercises.

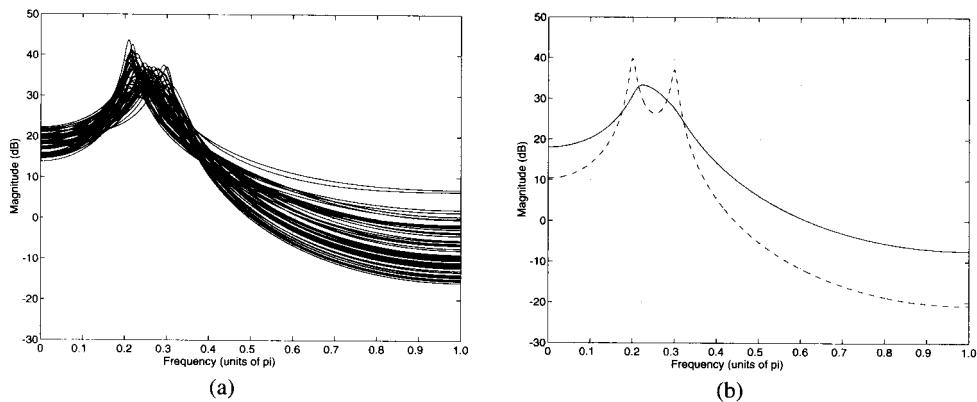
Consider the AR(4) process that is generated by the difference equation

$$\begin{aligned} x(n) = & 2.7377x(n-1) - 3.7476x(n-2) + 2.6293x(n-3) \\ & - 0.9224x(n-4) + w(n) \end{aligned} \quad (8.122)$$

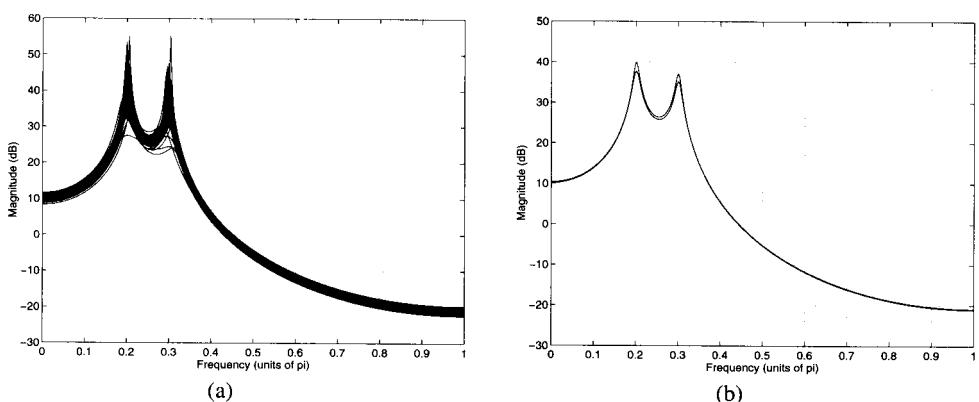
where  $w(n)$  is unit variance white Gaussian noise. The filter that generates  $x(n)$  has a pair of complex poles at  $z = 0.98e^{\pm j(0.2\pi)}$  and a pair of complex poles at  $z = 0.98e^{\pm j(0.3\pi)}$ . Using data records of length  $N = 128$ , an ensemble of 50 spectrum estimates were computed using the Yule-Walker method, the covariance method, the modified covariance method, and Burg's method. Shown in part *a* of Figs. 8.25 to 8.28 are overlay plots of these estimates and in part *b* the ensemble average is plotted along with the true power spectrum. What we observe from these plots is that, for this narrowband process, all of the estimates, except the Yule-Walker method, appear to be unbiased and to have a comparable variance. The Yule-Walker method, on the other hand, is unable to resolve the spectral peaks and has a larger variance.

**Selecting the Model Order.** A question that remains to be answered in the use of an AR spectrum estimation method is how to select the model order  $p$  of the AR process. If the model order that is used is too small, then the resulting spectrum will be smoothed and will have poor resolution. If, on the other hand, the model order is too large, then the spectrum may contain spurious peaks and, as illustrated in Fig. 8.24, may lead to spectral line splitting. Therefore, it would be useful to have a criterion that indicates the appropriate model order to use for a given set of data. One approach would be to increase the model order until the modeling error is minimized. The difficulty with this, however, is that the error is a monotonically nonincreasing function of the model order  $p$ . This problem may be overcome by incorporating a penalty function that increases with the model order  $p$ . Several criteria have been proposed that include a penalty term that increases *linearly* with  $p$

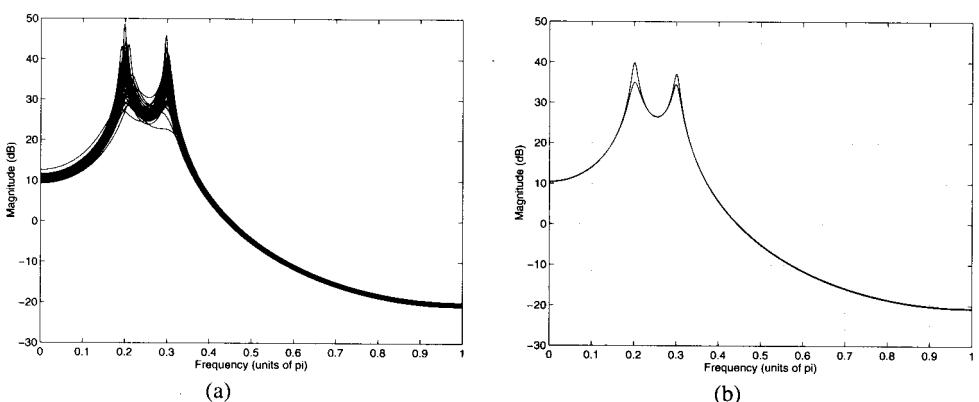
$$C(p) = N \log \mathcal{E}_p + f(N)p \quad (8.123)$$



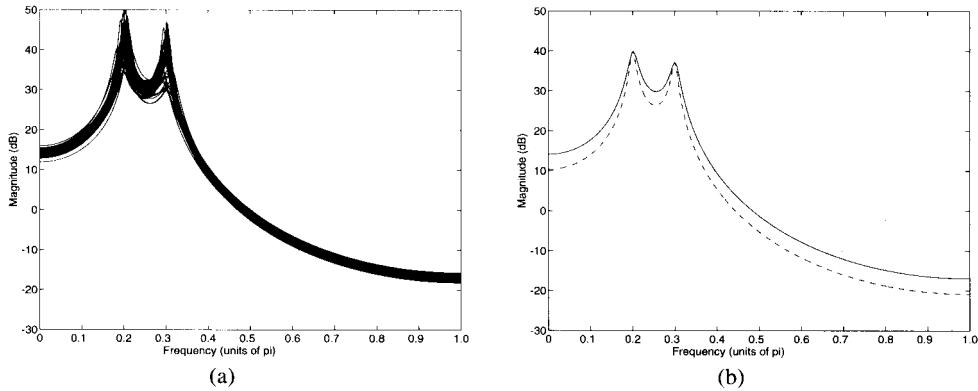
**Figure 8.25** Spectrum estimation of a fourth-order autoregressive process using the Yule-Walker method. (a) Overlay plot of 50 spectrum estimates. (b) The average of the estimates in (a) with the true power spectrum indicated by the dashed line.



**Figure 8.26** Spectrum estimation of a fourth-order autoregressive process using the covariance method. (a) Overlay plot of 50 spectrum estimates. (b) The average of the estimates in (a) with the true power spectrum indicated by the dashed line.



**Figure 8.27** Spectrum estimation of a fourth-order autoregressive process using the modified covariance method. (a) Overlay plot of 50 spectrum estimates. (b) The average of the estimates in (a) with the true power spectrum indicated by the dashed line.



**Figure 8.28** Spectrum estimation of a fourth-order autoregressive process using Burg's method. (a) Overlay plot of 50 spectrum estimates. (b) The average of the estimates in (a) with the true power spectrum indicated by the dashed line.

Here,  $\mathcal{E}_p$  is the modeling error,  $N$  is the data record length, and  $f(N)$  is a constant that may depend upon  $N$ . The idea, then, is to select the value of  $p$  that minimizes  $C(p)$ . Two criteria that are of this form are the Akaike Information Criterion [1,2]

$$\text{AIC}(p) = N \log \mathcal{E}_p + 2p \quad (8.124)$$

and the minimum description length proposed by Rissanen [45],

$$\text{MDL}(p) = N \log \mathcal{E}_p + (\log N)p \quad (8.125)$$

The AIC was derived by minimizing an information theoretic function and includes the penalty  $2p$  for any extra AR coefficients that do not significantly reduce the prediction error. It has been observed that the AIC gives an estimate for the order  $p$  that is too small when applied to nonautoregressive processes and that it tends to overestimate the order as  $N$  increases [23,24,55]. The MDL, on the other hand, contains the penalty term  $(\log N)p$ , which increases with the data record length  $N$  and the model order  $p$ . It has been shown that the MDL is a consistent model-order estimator in the sense that it converges to the true order as the number of observations,  $N$ , increases [45,58]. Two other model order selection criteria that are often used are Akaike's Final Prediction Error [2],

$$\text{FPE}(p) = \mathcal{E}_p \frac{N + p + 1}{N - p - 1} \quad (8.126)$$

and Parzen's Criterion Autoregressive Transfer function [39]

$$\text{CAT}(p) = \left[ \frac{1}{N} \sum_{j=1}^p \frac{N-j}{N\mathcal{E}_j} \right] - \frac{N-p}{N\mathcal{E}_p} \quad (8.127)$$

The success of each of these criteria in estimating model orders has been mixed. In fact, for short data sequences, none of the criteria tend to work particularly well [56]. Therefore, these criteria should only be used as "indicators" of the model order. It should also be pointed out that since each of these criteria depends upon the prediction error,  $\mathcal{E}_p$ , the model order will depend on the modeling technique that is used, e.g. the predicted model order may not be the same when using the autocorrelation method and the Burg algorithm. Table 8.8 summarizes these model order criteria.

**Table 8.8 Model Order Selection Criteria**

$AIC(p) = N \log \mathcal{E}_p + 2p$
$MDL(p) = N \log \mathcal{E}_p + (\log N)p$
$FPE(p) = \mathcal{E}_p \frac{N + p + 1}{N - p - 1}$
$CAT(p) = \left[ \frac{1}{N} \sum_{j=1}^p \frac{N - j}{N \mathcal{E}_j} \right] - \frac{N - p}{N \mathcal{E}_p}$

### 8.5.2 Moving Average Spectrum Estimation

A moving average process may be generated by filtering unit variance white noise,  $w(n)$ , with an FIR filter as follows:

$$x(n) = \sum_{k=0}^q b_q(k)w(n-k) \quad (8.128)$$

As discussed in Section 3.6.3, the relationship between the power spectrum of a moving average process and the coefficients  $b_q(k)$  is

$$P_x(e^{j\omega}) = \left| \sum_{k=0}^q b_q(k)e^{-jk\omega} \right|^2 \quad (8.129)$$

Equivalently, the power spectrum may be written in terms of the autocorrelation sequence  $r_x(k)$  as

$$P_x(e^{j\omega}) = \sum_{k=-q}^q r_x(k)e^{-jk\omega} \quad (8.130)$$

where  $r_x(k)$  is related to the filter coefficients  $b_q(k)$  through the Yule-Walker equations (see Section 3.6.3)

$$r_x(k) = \sum_{l=0}^{q-k} b_q(l+k)b_q^*(l) \quad ; \quad k = 0, 1, \dots, q \quad (8.131)$$

with  $r_x(-k) = r_x^*(k)$  and  $r_x(k) = 0$  for  $|k| > q$ .

With a moving average model, the spectrum may be estimated in one of two ways. The first approach is to take advantage of the fact that the autocorrelation sequence of a moving average process is finite in length. Specifically, since  $r_x(k) = 0$  for  $|k| > q$ , then a natural estimate to use is

$$\hat{P}_{MA}(e^{j\omega}) = \sum_{k=-q}^q \hat{r}_x(k)e^{-jk\omega} \quad (8.132)$$

where  $\hat{r}_x(k)$  is a suitable estimate of the autocorrelation sequence. Note that although  $\hat{P}_{MA}(e^{j\omega})$  is equivalent to the Blackman-Tukey estimate using a rectangular window, there

is a subtle difference in the assumptions that are behind these two estimates. In particular, since Eq. (8.132) assumes that  $x(n)$  is a moving average process of order  $q$ , then the true autocorrelation sequence is zero for  $|k| > q$ . Thus, if an unbiased estimate of the autocorrelation sequence is used for  $|k| \leq q$ , then

$$E\{\hat{P}_{MA}(e^{j\omega})\} = P_x(e^{j\omega})$$

i.e.,  $\hat{P}_{MA}(e^{j\omega})$  is unbiased. The Blackman-Tukey method, on the other hand, makes no assumptions about  $x(n)$  and may be applied to any type of process. Therefore, due to the windowing of the autocorrelation sequence, unless  $x(n)$  is a moving average process, the Blackman-Tukey spectrum will be biased.

The second approach is to estimate the moving average parameters,  $b_q(k)$ , from  $x(n)$  and then substitute these estimates into Eq. (8.129) as follows:

$$\hat{P}_{MA}(e^{j\omega}) = \left| \sum_{k=0}^q \hat{b}_q(k) e^{-jk\omega} \right|^2 \quad (8.133)$$

For example,  $b_q(k)$  may be estimated using the two-stage approach developed by Durbin (see Section 4.7.3).

As with autoregressive spectrum estimation, it is useful to have a criterion for estimating the order of the MA model that should be used for a given process  $x(n)$ . A discussion of some of these estimation methods may be found in [25].

**Examples.** As we did with the autoregressive methods, here we present only a few examples of MA spectrum estimation rather than attempting to be complete in covering all of the interesting possibilities. Specifically, we consider a fourth-order moving average process that is generated by the difference equation

$$x(n) = w(n) - 1.5857w(n-1) + 1.9208w(n-2) \quad (8.134)$$

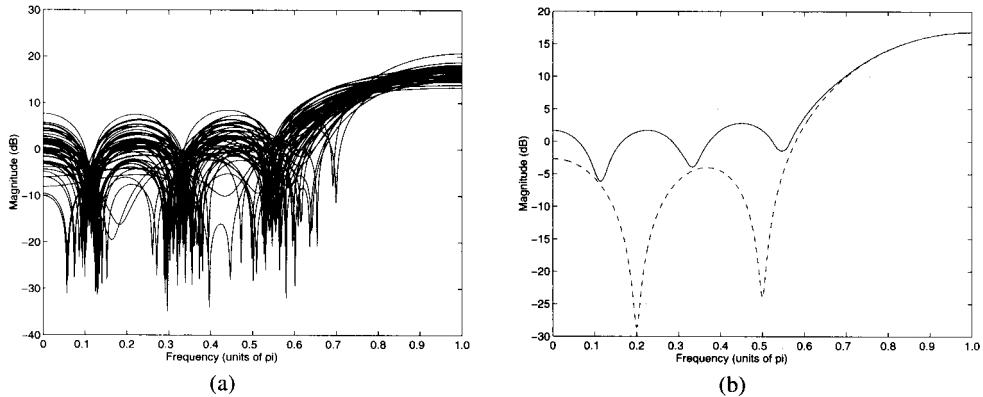
$$- 1.5229w(n-3) + 0.9224w(n-4)$$

where  $w(n)$  is unit variance white Gaussian noise. The filter that generates  $x(n)$  has a pair of complex zeroes at  $z = 0.98e^{\pm j(0.2\pi)}$  and a pair of complex zeroes at  $z = 0.98e^{\pm j(0.5\pi)}$ . Using data records of length  $N = 128$ , an ensemble of 50 spectrum estimates were computed using the Blackman-Tukey method with a rectangular window that extends from  $k = -4$  to  $k = 4$ . Shown in Fig. 8.29a is an overlay plot of these estimates and in b is the ensemble average along with the true power spectrum. These plots are repeated in Fig. (8.30) using Durbin's method with  $q = 4$  and  $p = 32$  (recall that  $p$  is the order of the all-pole model that is used to model  $x(n)$  prior to estimating the moving average coefficients).

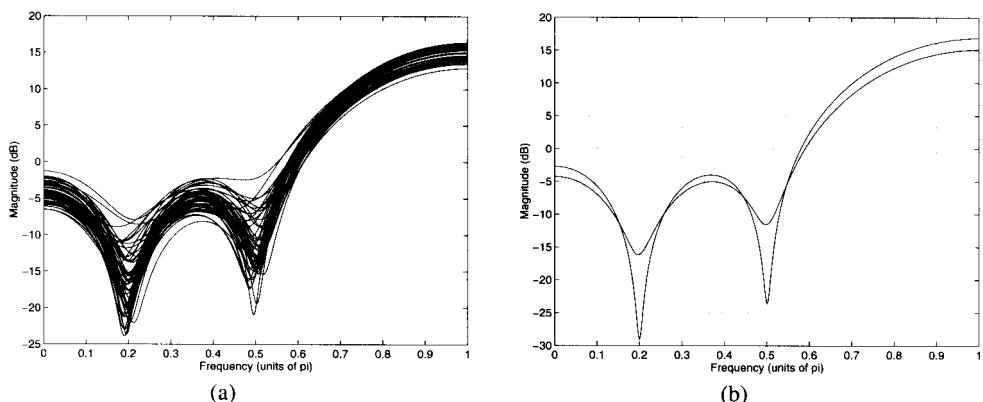
### 8.5.3 Autoregressive Moving Average Spectrum Estimation

An autoregressive moving average process has a power spectrum of the form

$$P_x(e^{j\omega}) = \frac{\left| \sum_{k=0}^q b_q(k) e^{-jk\omega} \right|^2}{\left| 1 + \sum_{k=1}^p a_p(k) e^{-jk\omega} \right|^2} \quad (8.135)$$



**Figure 8.29** Spectrum estimation of a fourth-order moving average process using the Blackman-Tukey method. (a) Overlay plot of 50 spectrum estimates. (b) The average of the estimates in (a) with the true power spectrum indicated by the dashed line.



**Figure 8.30** Spectrum estimation of a fourth-order moving average process using Durbin's method. (a) Overlay plot of 50 spectrum estimates. (b) The average of the estimates in (a) with the true power spectrum indicated by the dashed line.

which may be generated by filtering unit variance white noise with a filter having both poles and zeros,

$$H(z) = \frac{B_q(z)}{A_p(z)} = \frac{\sum_{k=0}^q b_q(k)z^{-k}}{1 + \sum_{k=1}^p a_p(k)z^{-k}}$$

Following the approach used for AR( $p$ ) and MA( $q$ ) spectrum estimation, the spectrum of an ARMA( $p, q$ ) process may be estimated from Eq. (8.135) using estimates of the model parameters

$$\hat{P}_{ARMA}(e^{j\omega}) = \frac{\left| \sum_{k=0}^q \hat{b}_q(k)e^{-jk\omega} \right|^2}{\left| 1 + \sum_{k=1}^p \hat{a}_p(k)e^{-jk\omega} \right|^2} \quad (8.136)$$

As discussed in Section 4.7.1, the AR model parameters may be estimated from the modified Yule-Walker equations either directly or by using a least squares approach. Once the coefficients  $\hat{a}_p(k)$  have been estimated, a moving average modeling technique such as Durbin's method may be used to estimate the moving average parameters  $b_q(k)$ . Examples may be found in [25].

## **8.6 FREQUENCY ESTIMATION**

In the previous section, we considered the problem of estimating the power spectrum of a WSS random process that could be modeled as the output of a linear shift-invariant filter that is driven by white noise. Another model of importance is one in which  $x(n)$  is a sum of complex exponentials in white noise,

$$x(n) = \sum_{i=1}^p A_i e^{jn\omega_i} + w(n) \quad (8.137)$$

It is assumed that the amplitudes  $A_i$  are complex,

$$A_i = |A_i|e^{j\phi_i}$$

with  $\phi_i$  uncorrelated random variables that are uniformly distributed over the interval  $[-\pi, \pi]$ . Although the frequencies and magnitudes of the complex exponentials,  $\omega_i$  and  $|A_i|$ , respectively, are not random, they are assumed to be unknown. Thus, the power spectrum of  $x(n)$  consists of a set of  $p$  impulses of area  $2\pi|A_i|$  at frequency  $\omega_i$  for  $i = 1, 2, \dots, p$ , plus the power spectrum of the additive noise  $w(n)$ . Signals of this form are found in a number of applications such as sonar signal processing and speech processing. Typically, the complex exponentials are the “information bearing” part of the signal and it is the estimation of the frequencies and amplitudes that is of interest rather than the estimation of the power spectrum itself. In the case of sonar signals, for example, the frequencies  $\omega_i$  may represent bearing or velocity information, whereas for speech signals they would correspond to the formant frequencies [42]. Although it is possible to estimate the frequencies of the complex exponentials from the peaks of the spectrum that is estimated using any of the techniques discussed in the previous sections, this approach would not fully exploit the assumed parametric form of the process. Therefore, in this section we consider *frequency estimation* algorithms that take into account the known properties of the process. The methods that we will be considering are based on an eigendecomposition of the autocorrelation matrix into two subspaces, a *signal* subspace and a *noise* subspace. Once these subspaces have been determined, a frequency estimation function is then used to extract estimates of the frequencies. Before looking at these *subspace methods*, we begin with a discussion of the eigendecomposition of the autocorrelation matrix.

### **8.6.1 Eigendecomposition of the Autocorrelation Matrix**

In order to motivate the use of an eigendecomposition of the autocorrelation matrix as an approach that may be used for frequency estimation, consider the first-order process

$$x(n) = A_1 e^{jn\omega_1} + w(n)$$

that consists of a single complex exponential in white noise. The amplitude of the complex exponential is  $A_1 = |A_1|e^{j\phi_1}$  where  $\phi_1$  is a uniformly distributed random variable, and

$w(n)$  is white noise that has a variance of  $\sigma_w^2$ . As shown in Section 3.6.4, the autocorrelation sequence of  $x(n)$  is

$$r_x(k) = P_1 e^{j k \omega_1} + \sigma_w^2 \delta(k)$$

where  $P_1 = |A_1|^2$  is the *power* in the complex exponential. Therefore, the  $M \times M$  autocorrelation matrix for  $x(n)$  is a sum of an autocorrelation matrix due to the signal,  $\mathbf{R}_s$ , and an autocorrelation matrix due to the noise,  $\mathbf{R}_n$ ,

$$\boxed{\mathbf{R}_x = \mathbf{R}_s + \mathbf{R}_n} \quad (8.138)$$

where the signal autocorrelation matrix is

$$\mathbf{R}_s = P_1 \begin{bmatrix} 1 & e^{-j\omega_1} & e^{-j2\omega_1} & \dots & e^{-j(M-1)\omega_1} \\ e^{j\omega_1} & 1 & e^{-j\omega_1} & \dots & e^{-j(M-2)\omega_1} \\ e^{j2\omega_1} & e^{j\omega_1} & 1 & \dots & e^{-j(M-3)\omega_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{j(M-1)\omega_1} & e^{j(M-2)\omega_1} & e^{j(M-3)\omega_1} & \dots & 1 \end{bmatrix} \quad (8.139)$$

and has a rank of one, and the noise autocorrelation matrix is diagonal,

$$\mathbf{R}_n = \sigma_w^2 \mathbf{I}$$

and has full rank. Note that if we define

$$\mathbf{e}_1 = [1, e^{j\omega_1}, e^{j2\omega_1}, \dots, e^{j(M-1)\omega_1}]^T \quad (8.140)$$

then  $\mathbf{R}_s$  may be written in terms of  $\mathbf{e}_1$  as follows:

$$\mathbf{R}_s = P_1 \mathbf{e}_1 \mathbf{e}_1^H$$

Since the rank of  $\mathbf{R}_s$  is equal to one, then  $\mathbf{R}_s$  has only one nonzero eigenvalue. With

$$\mathbf{R}_s \mathbf{e}_1 = P_1 (\mathbf{e}_1 \mathbf{e}_1^H) \mathbf{e}_1 = P_1 \mathbf{e}_1 (\mathbf{e}_1^H \mathbf{e}_1) = M P_1 \mathbf{e}_1$$

it follows that the nonzero eigenvalue is equal to  $M P_1$ , and that  $\mathbf{e}_1$  is the corresponding eigenvector. In addition, since  $\mathbf{R}_s$  is Hermitian then the remaining eigenvectors,  $\mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_M$ , will be orthogonal to  $\mathbf{e}_1$ .<sup>15</sup>

$$\mathbf{e}_1^H \mathbf{v}_i = 0 \quad ; \quad i = 2, 3, \dots, M \quad (8.141)$$

Finally, note that if we let  $\lambda_i^s$  be the eigenvalues of  $\mathbf{R}_s$ , then

$$\mathbf{R}_x \mathbf{v}_i = (\mathbf{R}_s + \sigma_w^2 \mathbf{I}) \mathbf{v}_i = \lambda_i^s \mathbf{v}_i + \sigma_w^2 \mathbf{v}_i = (\lambda_i^s + \sigma_w^2) \mathbf{v}_i \quad (8.142)$$

Therefore, the eigenvectors of  $\mathbf{R}_x$  are the same as those of  $\mathbf{R}_s$ , and the eigenvalues of  $\mathbf{R}_x$  are

$$\lambda_i = \lambda_i^s + \sigma_w^2$$

As a result, the largest eigenvalue of  $\mathbf{R}_x$  is

$$\lambda_{\max} = M P_1 + \sigma_w^2$$

and the remaining  $M - 1$  eigenvalues are equal to  $\sigma_w^2$ . Thus, it is possible to extract all of the parameters of interest about  $x(n)$  from the eigenvalues and eigenvectors of  $\mathbf{R}_x$  as follows:

<sup>15</sup>Recall that for a Hermitian matrix the eigenvectors corresponding to distinct eigenvalues are orthogonal. See Property 4 on p. 43 of Chapter 2.

1. Perform an eigendecomposition of the autocorrelation matrix,  $\mathbf{R}_x$ . The largest eigenvalue will be equal to  $M P_1 + \sigma_w^2$  and the remaining eigenvalues will be equal to  $\sigma_w^2$ .
2. Use the eigenvalues of  $\mathbf{R}_x$  to solve for the power  $P_1$  and the noise variance as follows:

$$\begin{aligned}\sigma_w^2 &= \lambda_{\min} \\ P_1 &= \frac{1}{M} (\lambda_{\max} - \lambda_{\min})\end{aligned}$$

3. Determine the frequency  $\omega_1$  from the eigenvector  $\mathbf{v}_{\max}$  that is associated with the largest eigenvalue using, for example, the second coefficient of  $\mathbf{v}_{\max}$ ,<sup>16</sup>

$$\omega_1 = \arg \left\{ v_{\max}(1) \right\}$$

The following example illustrates the procedure.

---

### **Example 8.6.1 Eigendecomposition of a Complex Exponential in Noise**

Let  $x(n)$  be a first-order harmonic process consisting of a single complex exponential in white noise,

$$x(n) = A_1 e^{jn\omega_1} + w(n)$$

with a  $2 \times 2$  autocorrelation matrix given by

$$\mathbf{R}_x = \begin{bmatrix} 3 & 2(1-j) \\ 2(1+j) & 3 \end{bmatrix}$$

The eigenvalues of  $\mathbf{R}_x$  are

$$\lambda_{1,2} = 3 \pm |2(1+j)| = 3 \pm 2\sqrt{2}$$

and the eigenvectors are

$$\mathbf{v}_{1,2} = \begin{bmatrix} 1 \\ \pm \frac{\sqrt{2}}{2}(1+j) \end{bmatrix}$$

Therefore, the variance of the white noise is

$$\sigma_w^2 = \lambda_{\min} = 3 - 2\sqrt{2}$$

and the power in the complex exponential is

$$P_1 = \frac{1}{M} (\lambda_{\max} - \lambda_{\min}) = 2\sqrt{2}$$

Finally, the frequency of the complex exponential is

$$\omega_1 = \arg \left\{ \frac{\sqrt{2}}{2}(1+j) \right\} = \pi/4$$

It should be pointed out that, for a first-order harmonic process, finding the parameters of the process is actually a bit easier than this. For example, note that

$$r_x(1) = 2(1+j) = 2\sqrt{2}e^{j\pi/4} = P_1 e^{j\omega_1}$$

<sup>16</sup>The components of the eigenvectors are numbered from  $v_i(0)$  to  $v_i(M-1)$ .

Therefore, we see immediately that  $P_1 = 2\sqrt{2}$  and  $\omega_1 = \pi/4$ . Furthermore, once  $P_1$  is known, then the variance of the white noise may be determined as follows:

$$\sigma_w^2 = r_x(0) - P_1 = 3 - 2\sqrt{2}$$

In practice, the approach described above is of limited value in frequency estimation since it requires that the autocorrelation matrix be known exactly. Although an estimated autocorrelation matrix could be used in place of  $\mathbf{R}_x$ , if this is done then the largest eigenvalue will only be approximately equal to  $P_1 + \sigma_w^2$  and the corresponding eigenvector will only be an approximation to  $\mathbf{e}_1$ . Since the eigenvalues and eigenvectors may be quite sensitive to small errors in  $r_x(k)$ , instead of estimating the frequency of the complex exponential from a single eigenvector, we may consider using a weighted average as follows. Let  $\mathbf{v}_i$  be a *noise eigenvector* of  $\mathbf{R}_x$ , i.e., one that has an eigenvalue of  $\sigma_w^2$ , and let  $v_i(k)$  be the  $k$ th component of  $\mathbf{v}_i$ . If we compute the discrete-time Fourier transform of the coefficients in  $\mathbf{v}_i$ ,

$$V_i(e^{j\omega}) = \sum_{k=0}^{M-1} v_i(k)e^{-jk\omega} = \mathbf{e}^H \mathbf{v}_i \quad (8.143)$$

then the orthogonality condition given in Eq. (8.141) implies that  $V_i(e^{j\omega})$  will be equal to zero at  $\omega = \omega_1$ , the frequency of the complex exponential. Therefore, if we form the *frequency estimation function*

$$\hat{P}_i(e^{j\omega}) = \frac{1}{\left| \sum_{k=0}^{M-1} v_i(k)e^{-jk\omega} \right|^2} = \frac{1}{|\mathbf{e}^H \mathbf{v}_i|^2} \quad (8.144)$$

then  $\hat{P}_i(e^{j\omega})$  will be large (in theory, infinite) at  $\omega = \omega_1$ . Thus, the location of the peak of this frequency estimation function may be used to estimate the frequency of the complex exponential. However, since Eq. (8.144) uses only a single eigenvector and, therefore, may be sensitive to errors in the estimation of  $\mathbf{R}_x$ , we may consider using a weighted average of all of the noise eigenvectors as follows:

$$\hat{P}(e^{j\omega}) = \frac{1}{\sum_{i=2}^M \alpha_i |\mathbf{e}^H \mathbf{v}_i|^2} \quad (8.145)$$

where  $\alpha_i$  are some appropriately chosen constants.

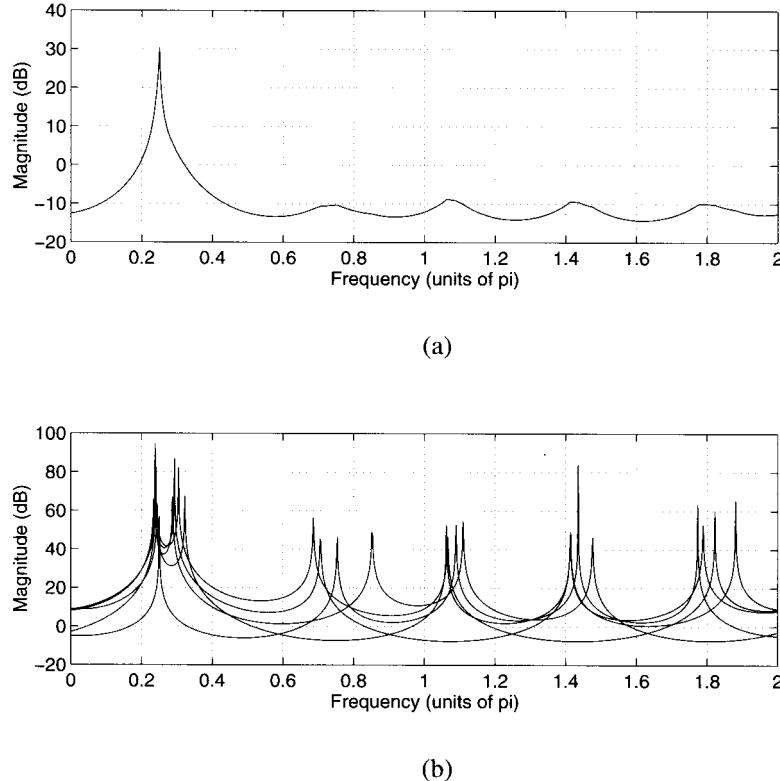
---

### Example 8.6.2 Eigendecomposition of a Complex Exponential in Noise (cont.)

Let  $x(n)$  be a WSS process consisting of a single complex exponential in unit variance white noise,

$$x(n) = 4 e^{j(n\pi/4+\phi)} + w(n)$$

where  $\phi$  is a uniformly distributed random variable. Using  $N = 64$  values of  $x(n)$ , a  $6 \times 6$  autocorrelation matrix was estimated and an eigendecomposition performed. Shown in Fig. 8.31a is a plot of the frequency estimation function defined in Eq. (8.145) with  $\alpha_i = 1$ . The peak of  $\hat{P}(e^{j\omega})$  occurs at frequency  $\omega = 0.2539\pi$  and the minimum eigenvalue is



**Figure 8.31** Frequency estimation functions of a single complex exponential in white noise. (a) The frequency estimation function that uses all of the noise eigenvectors with a weighting  $\alpha_i = 1$ . (b) An overlay plot of the frequency estimation functions  $V_i(e^{j\omega}) = 1/|\mathbf{e}^H \mathbf{v}_i|^2$  that are derived from each noise eigenvector.

$\lambda_{\min} = 1.08$ , which is close to the variance of the white noise. By contrast, shown in Fig. 8.31b is an overlay plot of the frequency estimation functions  $V_i(e^{j\omega})$  for each of the noise eigenvectors as defined in Eq. (8.143). As we see from this figure, although each plot has a peak that is close to  $\omega = 0.25\pi$ , with a single plot it is difficult to distinguish the correct peak from a spurious one.

Let us now consider what happens in the case of two complex exponentials in white noise,

$$x(n) = A_1 e^{jn\omega_1} + A_2 e^{jn\omega_2} + w(n)$$

where  $A_i = |A_i| e^{j\phi_i}$  for  $i = 1, 2$  are the amplitudes of the complex exponentials, and  $\omega_1$  and  $\omega_2$  are the frequencies with  $\omega_1 \neq \omega_2$ . If the variance of  $w(n)$  is  $\sigma_w^2$  then the autocorrelation of  $x(n)$  is

$$r_x(k) = P_1 e^{jk\omega_1} + P_2 e^{jk\omega_2} + \sigma_w^2 \delta(k)$$

where  $P_1 = |A_1|^2$  and  $P_2 = |A_2|^2$ . Thus, the autocorrelation matrix may again be written

as a sum

$$\mathbf{R}_x = P_1 \mathbf{e}_1 \mathbf{e}_1^H + P_2 \mathbf{e}_2 \mathbf{e}_2^H + \sigma_w^2 \mathbf{I}$$

where

$$\mathbf{R}_s = P_1 \mathbf{e}_1 \mathbf{e}_1^H + P_2 \mathbf{e}_2 \mathbf{e}_2^H$$

is a rank two matrix representing the component of  $\mathbf{R}_x$  that is due to the signal, and

$$\mathbf{R}_n = \sigma_w^2 \mathbf{I}$$

is a diagonal matrix that is due to the noise. A more concise way to express this decomposition is to write  $\mathbf{R}_x$  as follows:

$$\mathbf{R}_x = \mathbf{E} \mathbf{P} \mathbf{E}^H + \sigma_w^2 \mathbf{I} \quad (8.146)$$

where

$$\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2]$$

is an  $M \times 2$  matrix containing the two signal vectors  $\mathbf{e}_1$  and  $\mathbf{e}_2$  and  $\mathbf{P} = \text{diag}\{P_1, P_2\}$  is a diagonal matrix containing the signal powers.

In addition to decomposing  $\mathbf{R}_x$  into a sum of two autocorrelation matrices as in Eq. (8.146), we may also perform an eigendecomposition of  $\mathbf{R}_x$  as follows. Let  $\mathbf{v}_i$  and  $\lambda_i$  be the eigenvectors and eigenvalues of  $\mathbf{R}_x$ , respectively, with the eigenvalues arranged in decreasing order,

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$$

Since  $\mathbf{R}_x = \mathbf{R}_s + \sigma_w^2 \mathbf{I}$  then

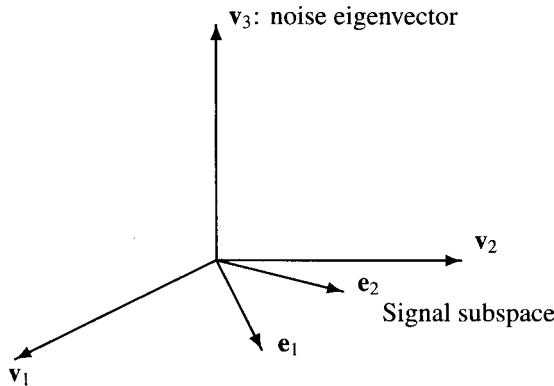
$$\lambda_i = \lambda_i^s + \sigma_w^2$$

where  $\lambda_i^s$  are the eigenvalues of  $\mathbf{R}_s$ . Since the rank of  $\mathbf{R}_s$  is equal to two, then  $\mathbf{R}_s$  has only two nonzero eigenvalues, and both of these are greater than zero ( $\mathbf{R}_s$  is nonnegative definite). Therefore, the first two eigenvalues of  $\mathbf{R}_x$  are greater than  $\sigma_w^2$  and the remaining eigenvalues are equal to  $\sigma_w^2$ . Thus, the eigenvalues and eigenvectors of  $\mathbf{R}_x$  may be divided into two groups. The first group, consisting of the two eigenvectors that have eigenvalues greater than  $\sigma_w^2$ , are referred to as *signal eigenvectors* and span a two-dimensional subspace called the *signal subspace*. The second group, consisting of those eigenvectors that have eigenvalues equal to  $\sigma_w^2$ , are referred to as the *noise eigenvectors* and span an  $(M - 2)$ -dimensional subspace called the *noise subspace*.<sup>17</sup> Since  $\mathbf{R}_x$  is Hermitian, the eigenvectors  $\mathbf{v}_i$  form an orthonormal set (see the discussion following Property 4 on p. 43). Therefore, the signal and noise subspaces are orthogonal. That is to say, for any vector  $\mathbf{u}$  in the signal subspace and for any vector  $\mathbf{v}$  in the noise subspace,  $\mathbf{u}^H \mathbf{v} = 0$ . The geometry of these subspaces is illustrated in Fig. 8.32.

Unlike the case for a single complex exponential, with a sum of two complex exponentials in noise, the signal eigenvectors will generally not be equal to  $\mathbf{e}_1$  and  $\mathbf{e}_2$ .<sup>18</sup> Nevertheless,  $\mathbf{e}_1$  and  $\mathbf{e}_2$  will lie in the signal subspace that is spanned by the signal eigenvectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , and since the signal subspace is orthogonal to the noise subspace, then  $\mathbf{e}_1$  and  $\mathbf{e}_2$  will be

<sup>17</sup>Note that this term is a bit misleading since the noise has components in both the noise and signal subspaces.

<sup>18</sup>This may be seen easily by noting that the signal vectors will not, in general, be orthogonal whereas the eigenvectors will always be orthogonal.



**Figure 8.32** Geometrical interpretation of the orthogonality of the signal and noise subspaces. The one-dimensional noise subspace is spanned by the noise eigenvector  $\mathbf{v}_3$  and the signal subspace which contains the signal vectors  $\mathbf{e}_1$  and  $\mathbf{e}_2$  is spanned by the signal eigenvectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$ .

orthogonal to the noise eigenvectors  $\mathbf{v}_i$ , i.e.,

$$\begin{aligned}\mathbf{e}_1^H \mathbf{v}_i &= 0 \quad ; \quad i = 3, 4, \dots, M \\ \mathbf{e}_2^H \mathbf{v}_i &= 0 \quad ; \quad i = 3, 4, \dots, M\end{aligned}\quad (8.147)$$

Therefore, as in the case of one complex exponential, the complex exponential frequencies,  $\omega_1$  and  $\omega_2$ , may be estimated using a frequency estimation function of the form

$$\hat{P}(e^{j\omega}) = \frac{1}{\sum_{i=3}^M \alpha_i |\mathbf{e}^H \mathbf{v}_i|^2}.$$

Let us now consider the general case of a wide-sense stationary process consisting of  $p$  distinct complex exponentials in white noise. The  $M \times M$  autocorrelation sequence is

$$r_x(k) = \sum_{i=1}^p P_i e^{jk\omega_i} + \sigma_w^2 \delta(k)$$

where  $P_i = |A_i|^2$  is the power in the  $i$ th component. Therefore, the autocorrelation matrix may be written as

$$\mathbf{R}_x = \mathbf{R}_s + \mathbf{R}_n = \sum_{i=1}^p P_i \mathbf{e}_i \mathbf{e}_i^H + \sigma_w^2 \mathbf{I} \quad (8.148)$$

where

$$\mathbf{e}_i = [1, e^{j\omega_i}, e^{j2\omega_i}, \dots, e^{j(M-1)\omega_i}]^T \quad ; \quad i = 1, 2, \dots, p$$

is a set of  $p$  linearly independent vectors. As was the case for two complex exponentials, Eq. (8.148) may be written concisely as follows:

$$\mathbf{R}_x = \mathbf{E} \mathbf{P} \mathbf{E}^H + \sigma_w^2 \mathbf{I}$$

(8.149)

where  $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_p]$  is an  $M \times p$  matrix containing the  $p$  signal vectors,  $\mathbf{e}_i$ , and  $\mathbf{P} = \text{diag}\{P_1, \dots, P_p\}$  is a diagonal matrix of signal powers. Since the eigenvalues of  $\mathbf{R}_x$  are  $\lambda_i = \lambda_i^s + \sigma_w^2$  where  $\lambda_i^s$  are the eigenvalues of  $\mathbf{R}_s$ , and since  $\mathbf{R}_s$  is a matrix of rank  $p$ , then the first  $p$  eigenvalues of  $\mathbf{R}_x$  will be greater than  $\sigma_w^2$  and the last  $M - p$  eigenvalues will be equal to  $\sigma_w^2$ . Therefore, the eigenvalues and eigenvectors of  $\mathbf{R}_x$  may again be divided into two groups: the signal eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_p$  that have eigenvalues greater than  $\sigma_w^2$ , and the noise eigenvectors  $\mathbf{v}_{p+1}, \dots, \mathbf{v}_M$  that have eigenvalues equal to  $\sigma_w^2$ . Assuming that the eigenvectors have been normalized to have unit norm, we may use the spectral theorem (p. 44) to decompose  $\mathbf{R}_x$  as follows:

$$\mathbf{R}_x = \sum_{i=1}^p (\lambda_i^s + \sigma_w^2) \mathbf{v}_i \mathbf{v}_i^H + \sum_{i=p+1}^M \sigma_w^2 \mathbf{v}_i \mathbf{v}_i^H$$

In matrix notation, this decomposition may be written as

$$\boxed{\mathbf{R}_x = \mathbf{V}_s \mathbf{V}_s^H + \mathbf{V}_n \mathbf{V}_n^H} \quad (8.150)$$

where

$$\mathbf{V}_s = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p]$$

is the  $M \times p$  matrix of signal eigenvectors and

$$\mathbf{V}_n = [\mathbf{v}_{p+1}, \mathbf{v}_{p+2}, \dots, \mathbf{v}_M]$$

is the  $M \times (M - p)$  matrix of noise eigenvectors, and where  $_s$  and  $_n$  are diagonal matrices that contain the eigenvalues  $\lambda_i = \lambda_i^s + \sigma_w^2$  and  $\lambda_i = \sigma_w^2$ , respectively. Later, we will be interested in projecting a vector onto either the signal subspace or the noise subspace. The projection matrices  $\mathbf{P}_s$  and  $\mathbf{P}_n$  that will perform these projections onto the signal and noise subspaces, respectively, are (see p. 34)

$$\boxed{\mathbf{P}_s = \mathbf{V}_s \mathbf{V}_s^H \quad ; \quad \mathbf{P}_n = \mathbf{V}_n \mathbf{V}_n^H} \quad (8.151)$$

As was the case for one and two complex exponentials in white noise, the orthogonality of the signal and noise subspaces may be used to estimate the frequencies of the complex exponentials. Specifically, since each signal vector  $\mathbf{e}_1, \dots, \mathbf{e}_p$  lies in the signal subspace, this orthogonality implies that  $\mathbf{e}_i$  will be orthogonal to each of the noise eigenvectors,

$$\mathbf{e}_i^H \mathbf{v}_k = 0 \quad ; \quad \begin{aligned} i &= 1, 2, \dots, p \\ k &= p + 1, p + 2, \dots, M \end{aligned}$$

Therefore, the frequencies may be estimated using a frequency estimation function such as

$$\hat{P}(e^{j\omega}) = \frac{1}{\sum_{i=p+1}^M \alpha_i |\mathbf{e}_i^H \mathbf{v}_i|^2} \quad (8.152)$$

In the following sections, we will develop several different types of frequency estimation algorithms that are based on Eq. (8.152). We begin with the Pisarenko harmonic decomposition, which uses a frequency estimator of this form with  $M = p + 1$  and  $\alpha_M = 1$ .

### 8.6.2 Pisarenko Harmonic Decomposition

In 1973, V. Pisarenko considered the problem of estimating the frequencies of a sum of complex exponentials in white noise [41]. Based on a theorem of Carathéodory, he demonstrated that the frequencies could be derived from the eigenvector corresponding to the minimum eigenvalue of the autocorrelation matrix. Although the resulting technique, referred to as the *Pisarenko harmonic decomposition*, is somewhat limited in its usefulness due to its sensitivity to noise, this decomposition is of theoretical interest, has led to important insights into the frequency estimation problem, and has provided the stimulus for the development of other eigenvalue decomposition methods that are more robust.

In the Pisarenko harmonic decomposition, it is assumed that  $x(n)$  is a sum of  $p$  complex exponentials in white noise, and that the number of complex exponentials,  $p$ , is known. It is also assumed that  $p + 1$  values of the autocorrelation sequence are either known or have been estimated. With a  $(p + 1) \times (p + 1)$  autocorrelation matrix, the dimension of the noise subspace is equal to one, and it is spanned by the eigenvector corresponding to the minimum eigenvalue,  $\lambda_{\min} = \sigma_w^2$ . Denoting this noise eigenvector by  $\mathbf{v}_{\min}$ , it follows that  $\mathbf{v}_{\min}$  will be orthogonal to each of the signal vectors,  $\mathbf{e}_i$ ,

$$\mathbf{e}_i^H \mathbf{v}_{\min} = \sum_{k=0}^p v_{\min}(k) e^{-jk\omega_i} = 0 \quad ; \quad i = 1, 2, \dots, p \quad (8.153)$$

Therefore,

$$V_{\min}(e^{j\omega}) = \sum_{k=0}^p v_{\min}(k) e^{-jk\omega}$$

is equal to zero at each of the complex exponential frequencies  $\omega_i$  for  $i = 1, 2, \dots, p$ . Consequently, the  $z$ -transform of the noise eigenvector, referred to as an *eigenfilter*, has  $p$  zeros on the unit circle,

$$V_{\min}(z) = \sum_{k=0}^p v_{\min}(k) z^{-k} = \prod_{k=1}^p (1 - e^{j\omega_k} z^{-1}) \quad (8.154)$$

and the frequencies of the complex exponentials may be extracted from the roots of the eigenfilter. As an alternative to rooting  $V_{\min}(z)$ , we may also form the *frequency estimation function*

$$\hat{P}_{PHD}(e^{j\omega}) = \frac{1}{|\mathbf{e}^H \mathbf{v}_{\min}|^2} \quad (8.155)$$

which is a special case of Eq. (8.152) with  $M = p + 1$  and  $\alpha_{p+1} = 1$ . Since  $\hat{P}_{PHD}(e^{j\omega})$  will be large (in theory, infinite) at the frequencies of the complex exponentials, the locations of the peaks in  $\hat{P}_{PHD}(e^{j\omega})$  may be used as frequency estimates. Although written in the form of a power spectrum,  $\hat{P}_{PHD}(e^{j\omega})$  is called a *pseudospectrum* (sometimes referred to as an *eigenspectrum*) since it does not contain any information about the power in the complex exponentials, nor does it contain a component due to the noise.

Once the frequencies of the complex exponentials have been determined, the powers  $P_i$  may be found from the eigenvalues of  $\mathbf{R}_x$  as follows. Let us assume that the signal subspace eigenvectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ , have been normalized so that  $\mathbf{v}_i^H \mathbf{v}_i = 1$ . With

$$\mathbf{R}_x \mathbf{v}_i = \lambda_i \mathbf{v}_i \quad ; \quad i = 1, 2, \dots, p \quad (8.156)$$

if both sides of Eq. (8.156) are multiplied on the left by  $\mathbf{v}_i^H$ , then

$$\mathbf{v}_i^H \mathbf{R}_x \mathbf{v}_i = \lambda_i \mathbf{v}_i^H \mathbf{v}_i = \lambda_i \quad ; \quad i = 1, 2, \dots, p \quad (8.157)$$

Substituting the expression for  $\mathbf{R}_x$  given in Eq. (8.148) into Eq. (8.157) we have

$$\mathbf{v}_i^H \mathbf{R}_x \mathbf{v}_i = \mathbf{v}_i^H \left\{ \sum_{k=1}^p P_k \mathbf{e}_k \mathbf{e}_k^H + \sigma_w^2 \mathbf{I} \right\} \mathbf{v}_i = \lambda_i$$

which may be simplified to

$$\boxed{\sum_{k=1}^p P_k |\mathbf{e}_k^H \mathbf{v}_i|^2 = \lambda_i - \sigma_w^2 \quad ; \quad i = 1, 2, \dots, p} \quad (8.158)$$

Note that the terms  $|\mathbf{e}_k^H \mathbf{v}_i|^2$  in the sum correspond to the squared magnitude of the DTFT of the signal subspace eigenvector  $\mathbf{v}_i$  evaluated at frequency  $\omega_k$ ,

$$|\mathbf{e}_k^H \mathbf{v}_i|^2 = |V_i(e^{j\omega_k})|^2$$

where

$$V_i(e^{j\omega}) = \sum_{l=0}^p v_i(l) e^{-jl\omega}$$

Therefore, Eq. (8.158) may also be written as

$$\sum_{k=1}^p P_k |V_i(e^{j\omega_k})|^2 = \lambda_i - \sigma_w^2 \quad ; \quad i = 1, 2, \dots, p \quad (8.159)$$

Equation (8.159) is a set of  $p$  linear equations in the  $p$  unknowns,  $P_k$ ,

$$\begin{bmatrix} |V_1(e^{j\omega_1})|^2 & |V_1(e^{j\omega_2})|^2 & \cdots & |V_1(e^{j\omega_p})|^2 \\ |V_2(e^{j\omega_1})|^2 & |V_2(e^{j\omega_2})|^2 & \cdots & |V_2(e^{j\omega_p})|^2 \\ \vdots & \vdots & & \vdots \\ |V_p(e^{j\omega_1})|^2 & |V_p(e^{j\omega_2})|^2 & \cdots & |V_p(e^{j\omega_p})|^2 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_p \end{bmatrix} = \begin{bmatrix} \lambda_1 - \sigma_w^2 \\ \lambda_2 - \sigma_w^2 \\ \vdots \\ \lambda_p - \sigma_w^2 \end{bmatrix} \quad (8.160)$$

which may be solved for the powers  $P_k$ . Thus, as summarized in Table 8.9, the Pisarenko

**Table 8.9 Pisarenko's Method for Frequency Estimation**

**Step 1:** Given that a process consists of  $p$  complex exponentials in white noise, find the minimum eigenvalue  $\lambda_{\min}$  and the corresponding eigenvector  $\mathbf{v}_{\min}$  of the  $(p+1) \times (p+1)$  autocorrelation matrix  $\mathbf{R}_x$ .

**Step 2:** Set the white noise power equal to the minimum eigenvalue,  $\lambda_{\min} = \sigma_w^2$ , and set the frequencies equal to the angles of the roots of the eigenfilter

$$V_{\min}(z) = \sum_{k=0}^p v_{\min}(k) z^{-k}$$

or the location of the peaks in the frequency estimation function

$$\hat{P}_{PHD}(e^{j\omega}) = \frac{1}{|\mathbf{e}^H \mathbf{v}_{\min}|^2}$$

**Step 3:** Compute the powers of the complex exponentials by solving the linear equations (8.160).

***The Pisarenko Harmonic Decomposition***

```

function [vmin,sigma] = phd(x,p)
%
x = x(:);
R = covar(x,p+1);
[v,d]=eig(R);
sigma=min(diag(d));
index=find(diag(d)==sigma);
vmin = v(:,index);
end;

```

**Figure 8.33** A MATLAB program to estimate the frequencies of  $p$  complex exponentials in white noise using Pisarenko's method.

harmonic decomposition estimates the complex exponential frequencies either from the roots of the eigenfilter  $V_{\min}(z)$  or from the locations of the peaks in the frequency estimation function  $\hat{P}_{PHD}(e^{j\omega})$ , and then solves the linear equations (8.160) for the powers of the complex exponentials. A MATLAB program for the Pisarenko decomposition is given in Fig 8.33.

***Example 8.6.3 Pisarenko's Method for Two Complex Exponentials in Noise***

Suppose that the first three autocorrelations of a random process consisting of two complex exponentials in white noise are

$$\begin{aligned} r_x(0) &= 6 \\ r_x(1) &= 1.92705 + j4.58522 \\ r_x(2) &= -3.42705 + j3.49541 \end{aligned}$$

Let us use Pisarenko's method to find the frequencies and powers of the complex exponentials. Since  $p = 2$ , we must perform an eigendecomposition of the  $3 \times 3$  autocorrelation matrix

$$\mathbf{R}_x = \begin{bmatrix} 6 & 1.92705 - j4.58522 & -3.42705 - j3.49541 \\ 1.92705 + j4.58522 & 6 & 1.92705 - j4.58522 \\ -3.42705 + j3.49541 & 1.92705 + j4.58522 & 6 \end{bmatrix}$$

The eigenvectors are

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = \begin{bmatrix} 0.5763 - 0.0000j & -0.2740 + 0.6518j & -0.2785 - 0.3006j \\ 0.2244 + 0.5342j & 0.0001 + 0.0100j & -0.3209 + 0.7492j \\ -0.4034 + 0.4116j & 0.2830 + 0.6480j & 0.4097 - 0.0058j \end{bmatrix}$$

and the eigenvalues are

$$\lambda_1 = 15.8951 ; \quad \lambda_2 = 1.1049 ; \quad \lambda_3 = 1.0000$$

Therefore, the minimum eigenvalue is  $\lambda_{\min} = 1$  and the corresponding eigenvector is

$$\mathbf{v}_{\min} = \begin{bmatrix} -0.2785 - 0.3006j \\ -0.3209 + 0.7492j \\ 0.4097 - 0.0058j \end{bmatrix}$$

Since the roots of the eigenfilter are

$$z_1 = 0.5 + j0.8660 = e^{j\pi/3} ; \quad z_2 = 0.3090 + j0.9511 = e^{j2\pi/5}$$

then the frequencies of the complex exponentials are

$$\omega_1 = \pi/3 ; \quad \omega_2 = 2\pi/5$$

To find the powers in the complex exponentials, we must compute the squared magnitude of the DTFT of the signal eigenvectors,  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , at the complex exponential frequencies  $\omega_1$  and  $\omega_2$ . With

$$|V_1(e^{j\omega_1})|^2 = 2.9685 ; \quad |V_1(e^{j\omega_2})|^2 = 2.9861$$

and

$$|V_2(e^{j\omega_1})|^2 = 0.0315 ; \quad |V_2(e^{j\omega_2})|^2 = 0.0139$$

then Eq. (8.160) becomes

$$\begin{bmatrix} 2.9685 & 2.9861 \\ 0.0315 & 0.0139 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 - \sigma_w^2 \\ \lambda_2 - \sigma_w^2 \end{bmatrix}$$

where  $\sigma_w^2 = \lambda_{\min} = 1$ . Solving for  $P_1$  and  $P_2$  we find

$$P_1 = 2 ; \quad P_2 = 3$$

In the previous example, we used Pisarenko's method to estimate the frequencies of two complex exponentials in white noise. In the following example we consider the problem of estimating the frequency of a single sinusoid in white noise. Although a sinusoid is a sum of two complex exponentials, the two frequencies are constrained to be negatives of each other,  $\omega_2 = -\omega_1$ . This constraint results in an autocorrelation sequence that is real-valued, which forces the eigenvectors to be real and, therefore, constrains the roots of the eigenfilter to occur in complex conjugate pairs.

#### **Example 8.6.4 Pisarenko's Method for One Sinusoid**

Let  $x(n)$  be a random phase sinusoid in white noise

$$x(n) = A \sin(n\omega_0 + \phi) + w(n)$$

with

$$r_x(0) = 2.2 ; \quad r_x(1) = 1.3 ; \quad r_x(2) = 0.8$$

The eigenvalues of the  $3 \times 3$  autocorrelation matrix  $\mathbf{R}_x = \text{toep}\{2.2, 1.3, 0.8\}$  are

$$\lambda_1 = 4.4815 ; \quad \lambda_2 = 1.4 ; \quad \lambda_3 = 0.7185$$

Therefore, the white noise power is

$$\sigma_w^2 = \lambda_{\min} = 0.7185$$

The eigenvectors of  $\mathbf{R}_x$  are

$$\mathbf{v}_1 = \begin{bmatrix} 0.5506 \\ 0.6275 \\ 0.5506 \end{bmatrix} ; \quad \mathbf{v}_2 = \begin{bmatrix} -0.7071 \\ 0 \\ 0.7071 \end{bmatrix} ; \quad \mathbf{v}_3 = \begin{bmatrix} 0.4437 \\ -0.7787 \\ 0.4437 \end{bmatrix}$$

Note the symmetry of the eigenvectors (See Problem 2.9). To estimate the frequency of the sinusoid we find the roots of the eigenfilter  $V_{\min}(z)$ , which is the  $z$ -transform of  $\mathbf{v}_3$ ,

$$V_{\min}(z) = 0.4437(1 - 1.755z^{-1} + z^{-2})$$

Rooting the eigenfilter we find that  $V_{\min}(z)$  has roots at  $z = e^{\pm j\omega_0}$  where

$$2 \cos \omega_0 = 1.755$$

or

$$\omega_0 = 0.159\pi$$

Finally, the power in the sinusoid may be estimated using Eq. (8.160). However, the computation may be simplified by taking into account the fact that  $x(n)$  contains a single sinusoid. Specifically, since the autocorrelation sequence for a single sinusoid in white noise is (see Section 3.6.4)

$$r_x(k) = \frac{1}{2} A^2 \cos(k\omega_0) + \sigma_w^2 \delta(k)$$

then, for  $k = 0$ ,

$$r_x(0) = \frac{1}{2} A^2 + \sigma_w^2$$

With  $r_x(0) = 2.2$  and  $\sigma_w^2 = 0.7185$  it follows that

$$A^2 = 2.963$$

In spite of the mathematical elegance of the Pisarenko harmonic decomposition, it is not commonly used in practice. One of the reasons for this is that it requires that the number of complex exponentials be known. In the ideal case of exact autocorrelations, the number of complex exponentials may be determined by overestimating the number of exponentials and examining the multiplicity of the minimum eigenvalue. However, with estimated autocorrelations, the multiplicity rule will not work since the minimum eigenvalue will rarely have a multiplicity greater than one. Another limitation with the Pisarenko decomposition is that it assumes that the additive noise is white. In practice, this will generally not be the case and the frequency estimates will be biased. Although Pisarenko's method may be modified to account for nonwhite noise, it is necessary that the power spectrum of the additive noise be known [46].

From a computational point of view, the Pisarenko harmonic decomposition requires finding the minimum eigenvalue and eigenvector of the signal autocorrelation matrix. For high-order problems, this may be computationally time consuming. It is possible, however, to find the minimum eigenvalue and eigenvector efficiently using an iterative algorithm based on the Levinson-Durbin recursion [13,15,19].

### 8.6.3 MUSIC

In 1979, an improvement to the Pisarenko harmonic decomposition known as the MULTiple SIgnal Classification method (MUSIC) was presented by Schmidt [48]. Like Pisarenko's method, the MUSIC algorithm is a frequency estimation technique. To see how the MUSIC algorithm works, assume that  $x(n)$  is a random process consisting of  $p$  complex exponentials

in white noise with a variance of  $\sigma_w^2$ , and let  $\mathbf{R}_x$  be the  $M \times M$  autocorrelation matrix of  $x(n)$  with  $M > p + 1$ .<sup>19</sup> If the eigenvalues of  $\mathbf{R}_x$  are arranged in decreasing order,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$ , and if  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M$  are the corresponding eigenvectors, then we may divide these eigenvectors into two groups: the  $p$  signal eigenvectors corresponding to the  $p$  largest eigenvalues, and the  $M - p$  noise eigenvectors that, ideally, have eigenvalues equal to  $\sigma_w^2$ . However, with inexact autocorrelations the smallest  $M - p$  eigenvalues will only be approximately equal to  $\sigma_w^2$ . Although we could consider estimating the white noise variance by averaging the  $M - p$  smallest eigenvalues

$$\hat{\sigma}_w^2 = \frac{1}{M - p} \sum_{k=p+1}^M \lambda_k \quad (8.161)$$

estimating the frequencies of the complex exponentials is a bit more difficult. Since the eigenvectors of  $\mathbf{R}_x$  are of length  $M$ , each of the noise subspace eigenfilters

$$V_i(z) = \sum_{k=0}^{M-1} v_i(k) z^{-k} \quad ; \quad i = p + 1, \dots, M$$

will have  $M - 1$  roots (zeros). Ideally,  $p$  of these roots will lie on the unit circle at the frequencies of the complex exponentials, and the eigenspectrum

$$|V_i(e^{j\omega})|^2 = \frac{1}{\left| \sum_{k=0}^{M-1} v_i(k) e^{-jk\omega} \right|^2}$$

associated with the noise eigenvector  $\mathbf{v}_i$  will exhibit sharp peaks at the frequencies of the complex exponentials. However, the remaining  $(M - p - 1)$  zeros may lie anywhere and, in fact, some may lie close to the unit circle, giving rise to spurious peaks in the eigenspectrum. Furthermore, with inexact autocorrelations, the zeros of  $V_i(z)$  that are on the unit circle may not remain on the unit circle. Therefore, when only one noise eigenvector is used to estimate the complex exponential frequencies, there may be some ambiguity in distinguishing the desired peaks from the spurious ones. As we saw in Example 8.6.2, for one complex exponential in white noise, the spurious peaks that are introduced from each of the noise subspace eigenfilters tend to occur at different frequencies. Therefore, in the MUSIC algorithm, the effects of these spurious peaks are reduced by averaging, using the frequency estimation function

$$\hat{P}_{MU}(e^{j\omega}) = \frac{1}{\sum_{i=p+1}^M |\mathbf{e}^H \mathbf{v}_i|^2} \quad (8.162)$$

The frequencies of the complex exponentials are then taken as the locations of the  $p$  largest peaks in  $\hat{P}_{MU}(e^{j\omega})$ . Once the frequencies have been determined the power of each complex exponential may be found using Eq. (8.160).

Instead of searching for the peaks of  $\hat{P}_{MU}(e^{j\omega})$ , an alternative is to use a method called root MUSIC, which involves rooting a polynomial. Since the  $z$ -transform equivalent of

<sup>19</sup>Recall that in the Pisarenko decomposition,  $M = p + 1$ . Thus, here we are using additional autocorrelations. As we will see, if  $M = p + 1$ , then the MUSIC algorithm is equivalent to Pisarenko's method.

Eq. (8.162) is

$$\hat{P}_{MU}(z) = \frac{1}{\sum_{i=p+1}^M V_i(z)V_i^*(1/z^*)}$$

then the frequency estimates may be taken to be the angles of the  $p$  roots of the polynomial

$$D(z) = \sum_{i=p+1}^M V_i(z)V_i^*(1/z^*) \quad (8.163)$$

that are closest to the unit circle. A MATLAB program for the MUSIC algorithm is given in Fig 8.34.

#### 8.6.4 Other Eigenvector Methods

In addition to the Pisarenko and Music algorithms, a number of other eigenvector methods have been proposed for estimating the frequencies of complex exponentials in noise. One of these, the EigenVector (EV) method [20], is closely related to the MUSIC algorithm. Specifically, the EV method estimates the exponential frequencies from the peaks of the eigenspectrum

$$\hat{P}_{EV}(e^{j\omega}) = \frac{1}{\sum_{i=p+1}^M \frac{1}{\lambda_i} |\mathbf{e}^H \mathbf{v}_i|^2} \quad (8.164)$$

where  $\lambda_i$  is the eigenvalue associated with the eigenvector  $\mathbf{v}_i$ . Note that if  $w(n)$  is white noise and if the autocorrelation sequence  $r_x(k)$  is known exactly for  $k = 0, 1, \dots, M - 1$ ,

#### The MUSIC Algorithm

```
function Px = music(x,p,M)
%
x = x(:);
if M<p+1 | length(x)<M, error('Size of R is inappropriate'), end
R = covar(x,M);
[v,d]=eig(R);
[y,i]=sort(diag(d));
Px=0;
for j=1:M-p
    Px=Px+abs(fft(v(:,i(j)),1024));
end;
Px=-20*log10(Px);
end;
```

**Figure 8.34** A MATLAB program for estimating the frequencies of  $p$  complex exponentials in white noise using the MUSIC algorithm.

then the eigenvalues in Eq. (8.164) are equal to the white noise variance,  $\lambda_i = \sigma_v^2$ , and the EV eigenspectrum will be the same as the MUSIC pseudospectrum to within a constant. However, with estimated autocorrelations the eigenvector method differs from the MUSIC algorithm and appears to produce fewer spurious peaks [20,34]. A MATLAB program for finding the eigenspectrum using the eigenvector method is given in Fig 8.35.

Another eigendecomposition-based method of interest is the *minimum norm algorithm* [29]. Instead of forming an eigenspectrum that uses all of the noise eigenvectors as in the MUSIC and eigenvector algorithms, the minimum norm algorithm uses a single vector  $\mathbf{a}$  that is constrained to lie in the noise subspace, and the complex exponential frequencies are estimated from the peaks of the frequency estimation function

$$\hat{P}_{MN}(e^{j\omega}) = \frac{1}{|\mathbf{e}^H \mathbf{a}|^2} \quad (8.165)$$

With  $\mathbf{a}$  constrained to lie in the noise subspace, if the autocorrelation sequence is known exactly, then  $|\mathbf{e}^H \mathbf{a}|^2$  will have nulls at the frequencies of each complex exponential. Therefore, the  $z$ -transform of the coefficients in  $\mathbf{a}$  may be factored as follows:

$$A(z) = \sum_{k=0}^{M-1} a(k)z^{-k} = \prod_{k=1}^p (1 - e^{j\omega_k} z^{-1}) \prod_{k=p+1}^{M-1} (1 - z_k z^{-1})$$

where  $z_k$  for  $k = p + 1, \dots, M - 1$  are the spurious roots that do not, in general, lie on the unit circle. The problem then is to determine which vector in the noise subspace minimizes the effects of the spurious zeros on the peaks of  $\hat{P}_{MN}(e^{j\omega})$ . The approach that is used in the minimum norm algorithm is to find the vector  $\mathbf{a}$  that satisfies the following three constraints:

1. The vector  $\mathbf{a}$  lies in the noise subspace.
2. The vector  $\mathbf{a}$  has minimum norm.
3. The first element of  $\mathbf{a}$  is unity.

The first constraint ensures that  $p$  roots of  $A(z)$  lie on the unit circle. The second constraint ensures that the spurious roots of  $A(z)$  lie *inside* the unit circle, i.e.,  $|z_k| < 1$ . Finally, the third constraint ensures that the minimum norm solution is not the zero vector.

#### *The Eigenvector Method*

```
function Px = ev(x,p,M)
%
x = x(:);
if M<p+1, error('Specified size of R is too small'), end
R=covar(x,M);
[v,d]=eig(R);
[y,i]=sort(diag(d));
Px=0;
for j=1:M-p
    Px=Px+abs(fft(v(:,i(j)),1024)).^2/abs(y(j));
end;
Px=-10*log10(Px);
end;
```

**Figure 8.35** A MATLAB program for estimating the frequencies of  $p$  complex exponentials in white noise using the eigenvector method.

To solve this constrained minimization problem, we begin by noting that the constraint that  $\mathbf{a}$  lies in the noise subspace may be written as

$$\mathbf{a} = \mathbf{P}_n \mathbf{v} \quad (8.166)$$

where  $\mathbf{P}_n = \mathbf{V}_n \mathbf{V}_n^H$  is the projection matrix that projects an arbitrary vector  $\mathbf{v}$  onto the noise subspace (see Eq. (8.151) and p. 34). The third constraint may be expressed as follows:

$$\mathbf{a}^H \mathbf{u}_1 = 1 \quad (8.167)$$

where  $\mathbf{u}_1 = [1, 0, \dots, 0]^T$ . This constraint may be combined with the constraint given in Eq. (8.166) as follows,

$$\mathbf{v}^H (\mathbf{P}_n^H \mathbf{u}_1) = 1 \quad (8.168)$$

Using Eq. (8.166) the norm of  $\mathbf{a}$  may be written as

$$\|\mathbf{a}\|^2 = \|\mathbf{P}_n \mathbf{v}\|^2 = \mathbf{v}^H (\mathbf{P}_n^H \mathbf{P}_n) \mathbf{v}$$

Since  $\mathbf{P}_n$  is a projection matrix then it is Hermitian,  $\mathbf{P}_n = \mathbf{P}_n^H$ , and idempotent,  $\mathbf{P}_n^2 = \mathbf{P}_n$ . Therefore,

$$\|\mathbf{a}\|^2 = \mathbf{v}^H \mathbf{P}_n \mathbf{v}$$

and it follows that minimizing the norm of  $\mathbf{a}$  is equivalent to finding the vector  $\mathbf{v}$  that minimizes the quadratic form  $\mathbf{v}^H \mathbf{P}_n \mathbf{v}$ . We may now reformulate the constrained minimization problem as follows:

$$\boxed{\min \mathbf{v}^H \mathbf{P}_n \mathbf{v} \quad \text{subject to} \quad \mathbf{v}^H (\mathbf{P}_n^H \mathbf{u}_1) = 1} \quad (8.169)$$

Once the solution to Eq. (8.169) has been found, the minimum norm solution is formed by projecting  $\mathbf{v}$  onto the noise subspace using Eq. (8.166).

In Section 2.3.10 we considered the problem of solving constrained minimization problems of the form given in Eq. (8.169). What we found was that the solution is

$$\mathbf{v} = \lambda \mathbf{P}_n^{-1} (\mathbf{P}_n^H \mathbf{u}_1) = \lambda \mathbf{u}_1$$

where

$$\lambda = \frac{1}{\mathbf{u}_1^H \mathbf{P}_n \mathbf{u}_1}$$

Therefore, the minimum norm solution is

$$\boxed{\mathbf{a} = \mathbf{P}_n \mathbf{v} = \lambda \mathbf{P}_n \mathbf{u}_1 = \frac{\mathbf{P}_n \mathbf{u}_1}{\mathbf{u}_1^H \mathbf{P}_n \mathbf{u}_1}} \quad (8.170)$$

which is simply the projection of the unit vector onto the noise subspace, normalized so that the first coefficient is equal to one. In terms of the eigenvectors of the autocorrelation matrix, the minimum norm solution may be written as

$$\boxed{\mathbf{a} = \frac{(\mathbf{V}_n \mathbf{V}_n^H) \mathbf{u}_1}{\mathbf{u}_1^H (\mathbf{V}_n \mathbf{V}_n^H) \mathbf{u}_1}} \quad (8.171)$$

A MATLAB program for the minimum norm algorithm is given in Fig 8.36. The frequency estimation algorithms are summarized in Table 8.10.

***The Minimum Norm Algorithm***

```

function Px = min_norm(x,p,M)
%
x = x(:);
if M<p+1, error('Specified size of R is too small'), end
R=covar(x,M);
[v,d]=eig(R);
[y,i]=sort(diag(d));
for j=1:M-p
    V=[V,v(:,i(j))];
end;
a=V*V(1,:)';
Px=-20*log10(abs(fft(a,1024)));
end;

```

**Figure 8.36** A MATLAB program for estimating the frequencies of  $p$  complex exponentials in white noise using the minimum norm algorithm.

**Table 8.10 Noise Subspace Methods for Frequency Estimation**

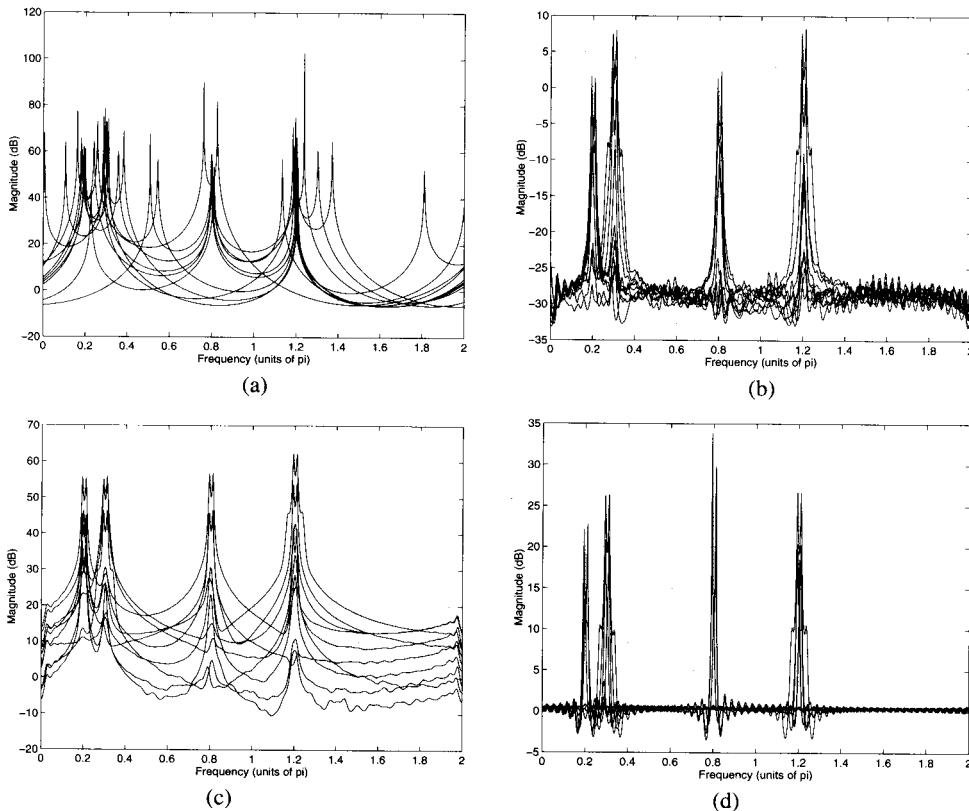
Pisarenko	$\hat{P}_{PHD}(e^{j\omega}) = \frac{1}{ \mathbf{e}^H \mathbf{v}_{\min} ^2}$
MUSIC	$\hat{P}_{MU}(e^{j\omega}) = \frac{1}{\sum_{i=p+1}^M  \mathbf{e}^H \mathbf{v}_i ^2}$
Eigenvector Method	$\hat{P}_{EV}(e^{j\omega}) = \frac{1}{\sum_{i=p+1}^M \frac{1}{\lambda_i}  \mathbf{e}^H \mathbf{v}_i ^2}$
Minimum Norm	$\hat{P}_{MN}(e^{j\omega}) = \frac{1}{ \mathbf{e}^H \mathbf{a} ^2} ; \quad \mathbf{a} = \lambda \mathbf{P}_n \mathbf{u}_1$

**Example 8.6.5 A Comparison of Frequency Estimation Methods**

Let  $x(n)$  be a process consisting of a sum of four complex exponentials in white noise

$$x(n) = \sum_{k=1}^4 A_k e^{j(n\omega_k + \phi_k)} + w(n)$$

where the amplitudes  $A_k$  are equal to one, the frequencies  $\omega_k$  are  $0.2\pi$ ,  $0.3\pi$ ,  $0.8\pi$ , and  $1.2\pi$ , the phases are uncorrelated random variables that are uniformly distributed over the interval  $[0, 2\pi]$ , and the variance of the white noise is  $\sigma_w^2 = 0.5$ . Using ten different realizations of  $x(n)$  with  $N = 64$  values, overlay plots of the frequency estimation functions using Pisarenko's method, the MUSIC algorithm, the eigenvector method, and the minimum norm algorithm are shown in Fig. 8.37. For Pisarenko's method, the frequency estimation



**Figure 8.37** The frequency estimation functions for a process consisting of four complex exponentials in white noise using (a) the Pisarenko harmonic decomposition, (b) the MUSIC algorithm, (c) the eigenvector method and (d) the minimum norm algorithm.

function was derived from the  $5 \times 5$  autocorrelation matrix that was estimated from the 64 values of  $x(n)$ . For the MUSIC, eigenvector, and minimum norm algorithms, the frequency estimation functions were formed from the  $64 \times 64$  autocorrelation matrix that was estimated from  $x(n)$ , i.e.,  $M = 64$ . Except for Pisarenko's method, the frequency estimation functions for this process produce accurate estimates of the exponential frequencies, with the most well-defined peaks being produced with the minimum norm algorithm. However, it is important to point out that not all of the frequency estimation functions shown in these overlay plots have four well-defined peaks. In some cases, for example, only two or three peaks are observed.

## 8.7 PRINCIPAL COMPONENTS SPECTRUM ESTIMATION

In the previous section, we saw how the orthogonality of the signal and noise subspaces could be used to estimate the frequencies of  $p$  complex exponentials in white noise. Since these methods only use vectors that lie in the noise subspace, they are often referred to as *noise subspace methods*. In this section, we consider another set of algorithms that use vectors that lie in the signal subspace. These methods are based on a principal components

analysis of the autocorrelation matrix and are referred to as *signal subspace methods*. The basic idea of these methods is as follows. Let  $\mathbf{R}_x$  be an  $M \times M$  autocorrelation matrix of a process that consists of  $p$  complex exponentials in white noise. With an eigendecomposition of  $\mathbf{R}_x$  we have

$$\mathbf{R}_x = \sum_{i=1}^M \lambda_i \mathbf{v}_i \mathbf{v}_i^H = \sum_{i=1}^p \lambda_i \mathbf{v}_i \mathbf{v}_i^H + \sum_{i=p+1}^M \lambda_i \mathbf{v}_i \mathbf{v}_i^H \quad (8.172)$$

where it is assumed that the eigenvalues have been arranged in decreasing order,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$ . Since the second term in Eq. (8.172) is due only to the noise, we may form a *reduced rank* approximation to the signal autocorrelation matrix,  $\mathbf{R}_s$ , by retaining only the principal eigenvectors of  $\mathbf{R}_x$ ,

$$\hat{\mathbf{R}}_s = \sum_{i=1}^p \lambda_i \mathbf{v}_i \mathbf{v}_i^H \quad (8.173)$$

This principal components approximation may then be used in the place of  $\mathbf{R}_x$  in a spectral estimator such as the minimum variance method or the maximum entropy method. The net effect of this approach is to *filter out* a portion of the noise, thereby enhancing the estimate of the spectral component due to the signal alone, i.e., the complex exponentials. Another way to view this approach is in terms of a constraint that is being placed on the autocorrelation matrix. Specifically, given that a process consists of  $p$  complex exponentials in noise, since the rank of the autocorrelation matrix due to the signal,  $\mathbf{R}_s$ , is  $p$ , then a principal components representation simply imposes this rank- $p$  constraint on  $\mathbf{R}_x$ . In the following subsections, we discuss how a principal components analysis of the autocorrelation matrix may be used in conjunction with the Blackman-Tukey method, the minimum variance method, and the maximum entropy method to form a principal components spectrum estimate.

### 8.7.1 Blackman-Tukey Frequency Estimation

The Blackman-Tukey estimate of the power spectrum is formed by taking the discrete-time Fourier transform of a windowed autocorrelation sequence

$$\hat{P}_{BT}(e^{j\omega}) = \sum_{k=-M}^M \hat{r}_x(k) w(k) e^{-jk\omega}$$

If  $w(k)$  is a Bartlett window, then the Blackman-Tukey estimate may be written in terms of the autocorrelation matrix  $\mathbf{R}_x$  as follows:

$$\hat{P}_{BT}(e^{j\omega}) = \frac{1}{M} \sum_{k=-M}^M (M - |k|) \hat{r}_x(k) e^{-jk\omega} = \frac{1}{M} \mathbf{e}^H \mathbf{R}_x \mathbf{e} \quad (8.174)$$

With an eigendecomposition of the autocorrelation matrix,  $\hat{P}_{BT}(e^{j\omega})$  may therefore be expressed as

$$\hat{P}_{BT}(e^{j\omega}) = \frac{1}{M} \sum_{i=1}^M \lambda_i |\mathbf{e}^H \mathbf{v}_i|^2$$

If  $x(n)$  is known to consist of  $p$  complex exponentials in white noise, and if the eigenvalues of  $\mathbf{R}_x$  are arranged in decreasing order,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$ , then a principal components

version of this spectrum estimate is

$$\hat{P}_{PC-BT}(e^{j\omega}) = \frac{1}{M} \mathbf{e}^H \hat{\mathbf{R}}_s \mathbf{e} = \frac{1}{M} \sum_{i=1}^p \lambda_i |\mathbf{e}^H \mathbf{v}_i|^2 \quad (8.175)$$

A MATLAB program for the principal components Blackman-Tukey spectrum estimate is given in Fig. 8.38.

### 8.7.2 Minimum Variance Frequency Estimation

Given the autocorrelation sequence  $r_x(k)$  of a process  $x(n)$  for lags  $|k| \leq M$ , the  $M$ th-order minimum variance spectrum estimate is

$$\hat{P}_{MV}(e^{j\omega}) = \frac{M}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}} \quad (8.176)$$

With an eigendecomposition of the autocorrelation matrix, the inverse of  $\mathbf{R}_x$  is

$$\mathbf{R}_x^{-1} = \sum_{i=1}^p \frac{1}{\lambda_i} \mathbf{v}_i \mathbf{v}_i^H + \sum_{i=p+1}^M \frac{1}{\lambda_i} \mathbf{v}_i \mathbf{v}_i^H \quad (8.177)$$

where  $p$  is the number of complex exponentials. Retaining only the first  $p$  principal components of  $\mathbf{R}_x^{-1}$  leads to the principal components minimum variance estimate

$$\hat{P}_{PC-MV}(e^{j\omega}) = \frac{M}{\sum_{i=1}^p \frac{1}{\lambda_i} |\mathbf{e}^H \mathbf{v}_i|^2} \quad (8.178)$$

It is interesting to compare Eq. (8.178) with the eigenvector method given in Eq. (8.164).

#### **Principal Components Frequency Estimation Using the Blackman-Tukey Method**

```

function Px = bt_pc(x,p,M)
%
x = x(:,1);
if M<p+1, error('Specified size of R is too small'), end
R=covar(x,M);
[v,d]=eig(R);
[y,i]=sort(diag(d));
Px=0;
for j=M-p+1:M
    Px=Px+abs(fft(v(:,i(j)),1024))*sqrt(real(y(j)));
end;
Px=20*log10(Px)-10*log10(M);
end;

```

**Figure 8.38** A MATLAB program for estimating the frequencies of  $p$  complex exponentials in white noise using a principal components analysis with the Blackman-Tukey method.

Whereas  $\hat{P}_{PC-MV}(e^{j\omega})$  is based on the first term in the decomposition given in Eq. (8.177), the EV method uses the second. Recall, however, that the EV algorithm produces a frequency estimation function rather than an estimate of the spectrum. Therefore, whereas  $\hat{P}_{EV}(e^{j\omega})$  simply produces peaks in the pseudospectrum at the complex exponential frequencies, the minimum variance method provides an estimate of the power spectrum.

### 8.7.3 Autoregressive Frequency Estimation

Autoregressive spectrum estimation using the autocorrelation, covariance, or modified covariance algorithms involves finding the solution to a set of linear equations of the form

$$\mathbf{R}_x \mathbf{a}_M = \epsilon_M \mathbf{u}_1 \quad (8.179)$$

where  $\mathbf{R}_x$  is an  $(M + 1) \times (M + 1)$  autocorrelation matrix. From the solution to these equations,

$$\mathbf{a}_M = \epsilon_M \mathbf{R}_x^{-1} \mathbf{u}_1$$

an estimate of the spectrum is formed as follows:

$$\hat{P}_{AR}(e^{j\omega}) = \frac{|b(0)|^2}{|\mathbf{e}^H \mathbf{a}_M|^2}$$

where  $b(0)$  is some appropriately chosen constant such as  $|b(0)|^2 = \epsilon_M$ . However, if  $x(n)$  is known to consist of  $p$  complex exponentials in noise, then we may form a principal components solution to Eq. (8.179) as follows:

$$\mathbf{a}_{pc} = \epsilon_M \left( \sum_{i=1}^p \frac{1}{\lambda_i} \mathbf{v}_i \mathbf{v}_i^H \right) \mathbf{u}_1$$

or,

$$\mathbf{a}_{pc} = \epsilon_M \sum_{i=1}^p \frac{v_i^*(0)}{\lambda_i} \mathbf{v}_i = \epsilon_M \sum_{i=1}^p \alpha_i \mathbf{v}_i$$

where  $v_i(0)$  is the first element of the normalized eigenvector  $\mathbf{v}_i$  and  $\alpha_i = v_i^*(0)/\lambda_i$ . Therefore, if we set  $|b(0)|^2 = \epsilon_M$ , then the principal components autoregressive spectrum estimate becomes

$$\hat{P}_{PC-AR}(e^{j\omega}) = \frac{1}{\left| \sum_{i=1}^p \alpha_i \mathbf{e}^H \mathbf{v}_i \right|^2}$$

Note that as the number of autocorrelations is increased, only  $p$  principal eigenvectors and eigenvalues are used in  $\hat{P}_{PC-AR}(e^{j\omega})$ . This allows for an increase in the model order without a corresponding increase in the spurious peaks that are due to the noise eigenvectors of the autocorrelation matrix.

We conclude this section with a summary of the signal subspace frequency estimation algorithms, which is given in Table 8.11. As we have seen, each technique is based on using the  $p$  principal eigenvectors of the autocorrelation matrix. The techniques differ, however, in the weighting function that is applied to the eigenfilters  $\mathbf{e}^H \mathbf{v}_k$  and in whether the linear combination of weighted eigenfilters is in the numerator or denominator of the spectrum estimate.

**Table 8.11 Signal Subspace Methods**

Blackman-Tukey	$\hat{P}_{PC-BT}(e^{j\omega}) = \frac{1}{M} \sum_{i=1}^p \lambda_i  \mathbf{e}^H \mathbf{v}_i ^2$
Minimum variance	$\hat{P}_{PC-MV}(e^{j\omega}) = \frac{M}{\sum_{i=1}^p \frac{1}{\lambda_i}  \mathbf{e}^H \mathbf{v}_i ^2}$
Autoregressive	$\hat{P}_{PC-AR}(e^{j\omega}) = \frac{1}{\left  \sum_{i=1}^p \alpha_i \mathbf{e}^H \mathbf{v}_i \right ^2}$

## 8.8 SUMMARY

In this chapter, we considered many different approaches for estimating the power spectrum of a wide-sense stationary process. We began with a discussion of the nonparametric methods that are based on computing the discrete-time Fourier transform of an estimate of the autocorrelation sequence. The first of these was the periodogram, which is easily evaluated from the DFT of the given values of the process. Unfortunately, however, the periodogram is not a consistent estimate of the power spectrum. Therefore, we looked at several modifications of the periodogram to improve its statistical properties. These included applying a window to the data, periodogram averaging, and periodogram smoothing. Although periodogram averaging and periodogram smoothing provide a consistent estimate of the power spectrum, they generally do not work well for short data records, and are limited in their ability to resolve closely spaced narrowband processes when the number of data samples is limited. An advantage of these methods, on the other hand, is that they do not make any assumptions or place any constraints on the process and, therefore, may be used on any type of process.

Next, we derived the minimum variance method, which may be viewed as a data-adaptive modification to the periodogram. The basic idea of this approach is to design a filter bank of bandpass filters, measure the power in the processes that are produced at the output of each filter, and estimate the power spectrum by dividing this power estimate by the bandwidth of the filter. Generally, the minimum variance spectrum estimate provides higher resolution than the periodogram and Blackman-Tukey methods.

The primary limitation with the nonparametric methods of spectrum estimation is that the estimate of the spectrum is based on a windowed autocorrelation sequence. In an attempt to overcome this limitation and improve the resolution, we then derived the maximum entropy method, which estimates the spectrum using a maximum entropy extrapolation of a given partial autocorrelation sequence. In other words, given  $r_x(k)$  for  $|k| \leq p$ , the values of  $r_x(k)$  for  $|k| > p$  are found that make the underlying process as white or as random as possible. What we discovered, however, is that this is equivalent to finding an all-pole model for the process that is consistent with the given autocorrelation sequence, and then computing the power spectrum from the all-pole model.

The next set of techniques that were discussed are the parametric methods of spectrum estimation. With a parametric approach, the first step is to select an appropriate model for the process. This selection may be based on a priori knowledge about how the process is generated or, perhaps, on experimental results indicating that a particular model "works well". Once a model has been selected, the next step is to estimate the model parameters from the given data. For example, if  $x(n)$  is assumed to be an autoregressive process, then the covariance method or some other algorithm may be used to estimate the all-pole parameters. The final step is to estimate the power spectrum by incorporating the estimated parameters into the parametric form for the spectrum. Although it is possible to significantly improve the performance of the spectrum estimate with a parametric approach, it is important that the model that is used be consistent with the process that is being analyzed. Otherwise inaccurate or misleading spectrum estimates may result.

The final set of techniques that were discussed are those that assume a harmonic model for the process, i.e., that  $x(n)$  is a sum of complex exponentials or sinusoids in white noise. For these processes, the goal is generally to estimate the frequencies of the complex exponentials and, possibly, to determine the powers. Two different approaches to the problem of frequency estimation were considered. The first involves defining a frequency estimation function that produces peaks at the frequencies of the complex exponentials. These frequency estimation functions are designed to take advantage of the fact that the signal and noise subspaces are orthogonal. The second set of approaches use a principal components analysis of the autocorrelation matrix. The resulting reduced-rank approximation to the autocorrelation matrix is then used in a spectrum estimation technique such as the minimum variance method or the maximum entropy method.

Although many different approaches to spectrum estimation have been presented in this chapter, the list is by no means complete. Many other methods have been proposed and, under certain conditions or set of assumptions, these approaches may be superior to the methods described here. For a more complete coverage of spectrum estimation, the reader may consult any one of a number of excellent books on the subject [10,16,25,28,34].

### References

---

1. H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst. Statist. Math.*, vol. 21, pp. 243–247, 1969.
2. H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, pp. 716–723, Dec. 1974.
3. N. O. Anderson, "Comments on the performance of the maximum entropy algorithm," *Proc. IEEE*, vol. 66, pp. 1581–1582, Nov. 1978.
4. R. Ash, *Information Theory*, Wiley Interscience, New York, 1965.
5. M. S. Bartlett, "Smoothing periodograms from time series with continuous spectra," *Nature* (London), vol. 161, pp. 686–687, May 1948.
6. R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*, Dover, New York, 1958.
7. J. P. Burg, "Maximum entropy spectral analysis," Ph.D. thesis, Stanford University, Stanford, CA., May 1975.
8. J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
9. W. Y. Chen and G. R. Stegen, "Experiments with maximum entropy power spectra of sinusoids," *J. Geophysics Rev.*, vol. 79, pp. 3019–3022, July 1974.
10. D. G. Childers, Ed., *Modern Spectrum Analysis*, IEEE Press, New York, 1978.

11. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., New York, 1991.
12. R. E. Davis and L. A. Regier, "Methods of estimating directional wave spectra from multi-element arrays," *Journal of Marine Research*, vol. 35, pp. 453–477, 1977.
13. C. Gueguen, Y. Grenier, and F. Giannella, "Factorial linear modeling, algorithms, and applications," *Proc. 1980 Int. Conf. Acoust., Speech, Signal Proc.*, pp. 618–621, 1980.
14. F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, pp. 51–83, Jan. 1978.
15. M. H. Hayes and M. A. Clements, "An efficient algorithm for computing Pisarenko's harmonic decomposition using Levinson's recursion," *IEEE Trans. on Acoustics, Speech, Sig. Proc.*, vol. ASSP-34, no. 3, pp. 485–491, June 1986.
16. S. Haykin, Ed., *Nonlinear Methods of Spectral Analysis*, Springer-Verlag, New York, 1979.
17. S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
18. S. Haykin and S. Kesler, "Prediction-error filtering and maximum-entropy spectral estimation," Chapter 2 in *Nonlinear Methods of Spectral Analysis*, S. Haykin, Ed., Springer-Verlag, New York, 1979.
19. Y. Hu and S. Y. Kung, "Highly concurrent Toeplitz eigen-system solver for high-resolution spectral estimation," *Proc. 1983 Int. Conf. Acoust., Speech, Signal Proc.*, pp. 1422–1425, 1983.
20. D. H. Johnson and S. R. DeGraaf, "Improving the resolution of bearing in passive sonar arrays by eigenvalue analysis," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-30, no. 4, pp. 638–647, Aug. 1982.
21. R. H. Jones, "Autoregression order selection," *Geophysics*, vol. 41, pp. 771–773, Aug. 1976.
22. R. H. Jones, "Identification and autoregressive spectrum estimation," *IEEE Trans. Autom. Control*, vol. AC-19, pp. 894–897, Dec. 1974.
23. R. L. Kashyap, "Inconsistency of the AIC rule for estimating the order of autoregressive models," *IEEE Trans. Autom. Control*, vol. AC-25, pp. 996–998, Oct. 1980.
24. M. Kaveh and S. P. Bruzzone, "Order determination for autoregressive spectral estimation," *Record of the 1979 RADC Spectral Estimation Workshop*, pp. 139–145, 1979.
25. S. Kay, *Modern Spectrum Estimation: Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
26. S. Kay, "Noise compensation for autoregressive spectral estimates," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-28, pp. 292–303, June 1980.
27. S. Kay and S. L. Marple Jr., "Sources and remedies for spectral line splitting in autoregressive spectrum analysis," *Proc. 1979 Int. Conf. on Acoust., Speech, Sig. Proc.*, pp. 151–154, 1979.
28. S. B. Kesler, Ed., *Modern Spectrum Analysis, II*, IEEE Press, New York, 1986.
29. R. Kumaresan and D. W. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Trans. on Aerospace and Elec. Syst.*, vol. AES-19, vol. 1, pp. 134–139, Jan. 1983.
30. R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, pp. 661–675, Aug. 1971.
31. M. A. Lagunas and A. Gasull-Llampalas, "An improved maximum likelihood method for power spectral density estimation," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-32, no. 1, pp. 170–172, Feb. 1984.
32. T. E. Landers and R. T. Lacoss, "Some geophysical applications of autoregressive spectral estimates," *IEEE Trans. Geosci. Electron.*, vol. GE-15, pp. 26–32, Jan. 1977.
33. H. L. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences: Vol. I, Random Variables and Stochastic Processes*, John Wiley & Sons, New York, 1979.
34. S. L. Marple Jr., *Digital Spectral Analysis with Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
35. S. L. Marple Jr., "A new autoregressive spectrum analysis algorithm," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-28, pp. 441–454, Aug. 1980.

36. T. L. Marzetta and S. W. Lang, "New interpretations for the MLM and DASE spectral estimators," *Proc. 1983 Int. Conf. Acoust., Speech, Sig. Proc.*, pp. 844–846, April 1983.
37. A. H. Nuttall, "Spectral analysis of a univariate process with bad data points via maximum entropy and linear predictive techniques," Tech. Rep. TR-5303, Naval Underwater Systems Center, New London, CT., March 1976.
38. A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
39. E. Parzen, "Some recent advances in time series modeling," *IEEE Trans. Autom. Control*, vol. AC-19, pp. 723–730, Dec. 1974.
40. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1984.
41. V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophysics J. Roy. Astron. Soc.*, vol. 33, pp. 347–366, 1973.
42. R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-34, no. 4, pp. 744–754, Aug. 1986.
43. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, NJ, 1978
44. I. S. Reed, "On a moment theorem for complex Gaussian processes," *Proc. IEEE Trans. Inf. Theory*, vol. IT-8, pp. 194–195, April 1962.
45. J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
46. R. A. Roberts and C. T. Mullis, *Digital Signal Processing*, Addison Wesley, Reading, MA. 1987.
47. E. A. Robinson, "A historical perspective of spectrum estimation," *Proc. IEEE*, vol. 70, pp. 885–907, Sept., 1982.
48. R. Schmidt, "Multiple emitter location and signal parameter estimation," *Proc. RADC Spectrum Estimation Workshop*, pp. 243–258, 1979.
49. A. Schuster, "On the investigation of hidden periodicities with application to a supposed twenty-six day period of meteorological phenomena," *Terr. Magn.*, vol. 3, no. 1, pp. 13–41, March 1898.
50. G. Schwartz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.
51. S. Shon and K. Mehrota, "Performance comparisons of autoregressive estimation methods," *Proc. 1984 Int. Conf. Acoust., Speech, Sig. Proc.*, pp. 14.3.1–14.3.4., 1984.
52. G. Strang, *Linear Algebra and its Applications*, Academic Press, New York, 1980.
53. D. N. Swigler, "A comparison between Burg's maximum entropy method and a non-recursive technique for the spectral analysis of deterministic signals," *J. Geophysics Res.*, vol. 84, pp. 679–685, Feb. 1979.
54. D. M. Thomas and M. H. Hayes, "A novel data-adaptive power spectrum estimation technique," *Proc. 1987 Int. Conf. on Acoustics, Speech, Sig. Proc.*, pp. 1589–1592, April 1987.
55. H. Tong, "Autoregressive model fitting with noisy data by Akaike's information criterion," *IEEE Trans. Inf. Theory*, vol. IT-21, pp. 476–480, July 1975.
56. T. J. Ulrych and M. Ooe, "Autoregressive and mixed autoregressive-moving average models," Chapter 3 in *Nonlinear Methods of Spectral Analysis*, S. Haykin, Ed., Springer-Verlag, New York, 1979.
57. T. J. Ulrych and R. W. Clayton, "Time series modeling and maximum entropy," *Phys. Earth Planet. Inter.*, vol. 12, pp. 188–200, Aug. 1976.
58. M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-33, no. 2, pp. 387–392, April 1985.
59. P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short modified periodograms," *IEEE Trans. Audio and Electroacoust.*, vol. AU-15, pp. 70–73, June 1967.

## 8.9 PROBLEMS

**8.1.** Given  $N = 10,000$  samples of a process  $x(n)$ , you are asked to compute the periodogram. However, with only a finite amount of memory resources, you are unable to compute a DFT any longer than 1024. Using these 10,000 samples, describe how you would be able to compute a periodogram that has a resolution of

$$\Delta\omega = 0.89 \frac{2\pi}{10000}$$

Hint: Consider how the decimation-in-time FFT algorithm works.

**8.2.** A continuous-time signal  $x_a(t)$  is bandlimited to 5 kHz, i.e.,  $x_a(t)$  has a spectrum  $X_a(f)$  that is zero for  $|f| > 5$  kHz. Only 10 seconds of the signal has been recorded and is available for processing. We would like to estimate the power spectrum of  $x_a(t)$  using the available data in a radix-2 FFT algorithm, and it is required that the estimate have a resolution of at least 10 Hz. Suppose that we use Bartlett's method of periodogram averaging.

- (a) If the data is sampled at the Nyquist rate, what is the minimum section length that you may use to get the desired resolution?
- (b) Using the minimum section length determined in part (a), with 10 seconds of data, how many sections are available for averaging?
- (c) How does your choice of the sampling rate affect the resolution and variance of your estimate? Are there any benefits to sampling above the Nyquist rate?

**8.3.** Bartlett's method is used to estimate the power spectrum of a process from a sequence of  $N = 2000$  samples.

- (a) What is the minimum length  $L$  that may be used for each sequence if we are to have a resolution of  $\Delta f = 0.005$ ?
- (b) Explain why it would not be advantageous to increase  $L$  beyond the value found in (a).
- (c) The *quality factor* of a spectrum estimate is defined to be the inverse of the variability,

$$Q = 1/\mathcal{V}$$

Using Bartlett's method, what is the minimum number of data samples,  $N$ , that are necessary to achieve a resolution of  $\Delta f = 0.005$ , and a quality factor that is five times larger than that of the periodogram?

**8.4.** A random process  $x(n)$  is generated by filtering unit variance white noise as shown in the figure below



where

$$A_1(z) = 1 + az^{-1} + 0.99z^{-2} \quad ; \quad A_2(z) = 1 - az^{-1} + 0.98z^{-2}$$

- (a) Prepare a carefully labeled sketch of the power spectrum of  $x(n)$  assuming that  $a$  is small, e.g.,  $0 < a < 0.1$ . Pay careful attention to the location and amplitude of the two spectral peaks and the value of  $P_x(e^{j\omega})$  at  $\omega = \pi/2$ .

- (b) If  $\alpha = 0.1$ , determine the section length  $L$  required to resolve the spectral peaks of  $P_x(e^{j\omega})$  using Bartlett's method. For this value of  $L$ , find an approximate value for the bias of the estimate at the peaks of the spectrum. How is the bias related to the area of the spectral peaks?
- (c) Consider the method of periodogram smoothing. How many lags of the autocorrelation must be used to obtain a resolution that is comparable to that of Bartlett's estimate considered in part (b)? How much data must be available if the variance of the estimate is to be comparable to that of a four-section Bartlett estimate?

**8.5.** Many commercial *Fourier analyzers* continuously update the estimate of the power spectrum of a process  $x(n)$  by exponential averaging of periodograms as follows,

$$\hat{P}_i(e^{j\omega}) = \alpha \hat{P}_{i-1}(e^{j\omega}) + \frac{1-\alpha}{N} \left| \sum_{n=0}^{N-1} x_i(n) e^{-jn\omega} \right|^2$$

where  $x_i(n) = x(n + Ni)$  is the  $i$ th sequence of  $N$  data values. This update equation is initialized with  $\hat{P}_{-1}(e^{j\omega}) = 0$ .

- (a) Qualitatively describe the philosophy behind this method, and discuss how the value for the weighting factor  $\alpha$  should be selected.
- (b) Assuming that successive periodograms are uncorrelated and that  $0 < \alpha < 1$ , find the mean and variance of  $\hat{P}_i(e^{j\omega})$  for a Gaussian random process.
- (c) Repeat the analysis in part (b) if the periodograms are replaced with modified periodograms.

**8.6.** The minimum variance method of spectrum estimation constrains the bandpass filter  $G_i(e^{j\omega})$  to have a gain of one at frequency  $\omega = \omega_i$ ,

$$G_i(e^{j\omega_i}) = 1$$

Another approach is to constrain the filter to have unit energy over a frequency band that is centered at  $\omega = \omega_i$  and has a bandwidth of  $\Delta$ ,

$$\frac{1}{\Delta} \int_{\omega_i - \Delta/2}^{\omega_i + \Delta/2} |G_i(e^{j\omega})|^2 d\omega = 1$$

With this constraint, the filter coefficients  $\mathbf{g}_i = [g_i(0), g_i(1), \dots, g_i(p)]^T$  that minimize the power in the filtered process,

$$E\{|y_i(n)|^2\} = \mathbf{g}_i^H \mathbf{R}_x \mathbf{g}_i$$

may be shown to be the solution to a generalized eigenvalue problem,

$$\mathbf{R}_x \mathbf{g}_i = \lambda(\omega_i, \Delta) \mathbf{T} \mathbf{g}_i$$

where  $\mathbf{T}$  is a matrix whose elements depend upon  $\omega_i$  and  $\Delta$ . The spectrum estimate, referred to as the DASE estimate, is

$$\hat{P}_{DASE}(e^{j\omega}) = \lambda_{\min}(\omega_i, \Delta)$$

where  $\lambda_{\min}(\omega_i, \Delta)$  is the minimum eigenvalue of the generalized eigenvalue problem.

- (a) Perform the minimization of  $E\{|y_i(n)|^2\}$  and determine the form of the matrix  $\mathbf{T}$ .
- (b) What happens to the matrix  $\mathbf{T}$  in the limit as  $\Delta \rightarrow 0$ ? What does the power spectrum estimate correspond to in this case?

- (c) Repeat part (b) for  $\Delta = 2\pi$ .  
(d) Find the DASE estimate for white noise.

**8.7.** Let  $x(n)$  be a random process consisting of a single complex exponential in white noise,

$$r_x(k) = P e^{j k \omega_0} + \sigma_w^2 \delta(k)$$

and let  $\mathbf{g}_i$  be the minimum variance bandpass filter

$$\mathbf{g}_i = \frac{\mathbf{R}_x^{-1} \mathbf{e}_i}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i}$$

that has a center frequency  $\omega_i$  with  $G(e^{j\omega_i}) = 1$ . Assuming that  $\omega_i \neq \omega_0$ , prove that  $G_i(z)$  has a zero that approaches  $z = e^{j\omega_0}$  as  $\sigma_w^2 \rightarrow 0$ .

**8.8.** A random process is known to consist of a single sinusoid in white noise,

$$x(n) = A \cos(n\omega_0 + \phi) + w(n)$$

Thus, the autocorrelation sequence for  $x(n)$  is

$$r_x(k) = \frac{1}{2} A^2 \cos(k\omega_0) + \sigma_w^2 \delta(k)$$

- (a) If  $\omega_0 = \pi/4$ ,  $A = \sqrt{2}$ , and  $\sigma_w^2 = 1$ , find the second-order MEM spectrum,  $\hat{P}_{mem}(e^{j\omega})$ .  
(b) Determine the location of the poles of  $\hat{P}_{mem}(z)$ .  
(c) Does the peak of  $\hat{P}_{mem}(e^{j\omega})$  provide an accurate estimate of  $\omega_0$ ? How does this estimate of  $\omega_0$  compare to that obtained using the Pisarenko Harmonic decomposition?

**8.9.** Suppose that we have determined the following values for the autocorrelation sequence of a real-valued random process  $x(n)$ :

$$r_x(0) = 1 \quad ; \quad r_x(1) = a \quad ; \quad r_x(2) = 0$$

- (a) Using the Blackman-Tukey method with a rectangular window, find and make a carefully labeled sketch of the estimated power spectrum,  $\hat{P}_{BT}(e^{j\omega})$ .  
(b) Repeat part (a) for a second-order MEM spectrum estimate,  $\hat{P}_{mem}(e^{j\omega})$ .  
(c) Repeat part (a) for a MV spectrum estimate,  $\hat{P}_{MV}(e^{j\omega})$ .  
(d) What can you say about the autocorrelation sequences that correspond to the spectrum estimates  $\hat{P}_{BT}(e^{j\omega})$ ,  $\hat{P}_{mem}(e^{j\omega})$ , and  $\hat{P}_{MV}(e^{j\omega})$  found in parts (a)-(c)?

**8.10.** Given that the sixth-order minimum variance spectrum estimate of a process  $x(n)$  is

$$\hat{P}_{MV}(e^{j\omega}) = \frac{1}{1 + a \cos 4\omega + 4a \cos 6\omega}$$

and the seventh-order estimate is

$$\hat{P}_{MV}(e^{j\omega}) = \frac{1}{1 - 2a \cos 2\omega - a \cos 7\omega}$$

find the seventh-order maximum entropy spectrum,  $\hat{P}_{mem}(e^{j\omega})$ .

**8.11.** The first-order ( $p = 1$ ) minimum variance spectrum estimate of a random process is

$$\hat{P}_{MV}(e^{j\omega}) = \frac{8}{3 - \cos \omega}$$

- (a) Find the autocorrelations,  $r_x(0)$  and  $r_x(1)$ , that produced this spectrum estimate.  
 (b) In general, given the  $p$ th-order minimum variance estimate  $\hat{P}_{MV}(e^{j\omega})$ , is it possible to recover the values of the autocorrelation sequence that produce this estimate?

**8.12.** The second-order maximum entropy spectrum of a process  $x(n)$  is

$$\hat{P}_{mem}(e^{j\omega}) = \frac{2}{|1 - 0.5e^{-j\omega} + 0.25e^{-2j\omega}|^2}$$

- (a) What is the first-order maximum entropy spectrum?  
 (b) Find the second-order minimum variance spectrum estimate.

**8.13.** From measurements of a process  $x(n)$ , we estimate the following values for the autocorrelation sequence:

$$r_x(k) = \alpha^{|k|} ; |k| \leq M$$

where  $|\alpha| < 1$ . Estimate the power spectrum using

- (a) The Blackman-Tukey method with a rectangular window.  
 (b) The minimum variance method.  
 (c) The maximum entropy method.

**8.14.** In Eq. (8.97), the entropy of a Gaussian random process was given as

$$H(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln P_x(e^{j\omega}) d\omega$$

In this problem, we derive another expression for the entropy. Let  $x(n)$  be a real-valued zero mean Gaussian random process, and let  $\mathbf{x} = [x(0), x(1), \dots, x(N-1)]^T$  be an  $N$ -dimensional Gaussian random vector that is formed from samples of this process. The probability density function for this random vector is

$$f_x(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} (\det \mathbf{R}_x)^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{x}^T \mathbf{R}_x^{-1} \mathbf{x}\right\}$$

where  $\mathbf{R}_x$  is the autocorrelation matrix.

- (a) The *average entropy* of a random vector  $\mathbf{x}$  is defined as

$$H_N(\mathbf{x}) = -\frac{1}{N} \int f_x(\mathbf{x}) \ln f_x(\mathbf{x}) d\mathbf{x}$$

Show that the average entropy of a zero mean Gaussian random vector is

$$H_N(\mathbf{x}) = \frac{1}{2} \ln(2\pi e) + \frac{1}{2N} \ln \det \mathbf{R}_x$$

- (b) Show that the average entropy of a Gaussian random vector may be written as

$$H_N(\mathbf{x}) = \frac{1}{2} \ln(2\pi e) + \frac{1}{N} \sum_{k=0}^{N-1} \ln \epsilon_k$$

where  $\epsilon_k$  is the prediction error sequence that is generated by the Levinson-Durbin recursion from the autocorrelation sequence  $r_x(k)$ .

- (c) The *entropy rate* of a process  $x(n)$  is the limit, as  $N \rightarrow \infty$ , of the average entropy [40],

$$\bar{H}(\mathbf{x}) = \lim_{N \rightarrow \infty} H_N(\mathbf{x})$$

Given a partial autocorrelation sequence,  $r_x(k)$ , for  $k = 0, 1, \dots, N - 1$ , find the spectrum  $P_x(e^{j\omega})$  that maximizes  $H(x)$  subject to the constraint that the spectrum is consistent with the given autocorrelations.

- (d) Let  $P_x(e^{j\omega})$  be the power spectrum of a wide-sense stationary process with an  $N \times N$  autocorrelation matrix  $\mathbf{R}_x$ , and let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N \geq 0$  be the eigenvalues. Szegö's theorem states that if  $g(\cdot)$  is a continuous real-valued function then

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N g(\lambda_k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g[P_x(e^{j\omega})] d\omega$$

Use Szegö's theorem to show that

$$H(x) = \frac{1}{2} \ln(2\pi e) + \bar{H}(x)$$

i.e., that the two entropy expressions are equal to within an additive constant.

- 8.15.** In this problem, we examine how the entropy of a process changes with the addition of a harmonic process. Let  $y(n)$  be a random process with a power spectrum

$$P_y(e^{j\omega}) = P_x(e^{j\omega}) + P_\epsilon(e^{j\omega})$$

where

$$P_\epsilon(e^{j\omega}) = \begin{cases} 1/\epsilon & ; \quad |\omega - \omega_0| < \epsilon \\ 0 & ; \quad \text{otherwise} \end{cases}$$

- (a) Find the entropy of  $y(n)$ .  
(b) What is the entropy of this process in the limit as  $\epsilon \rightarrow 0$ ?

- 8.16.** Given an autocorrelation sequence  $r_x(k)$  for  $k = 0, 1, \dots, p$ , the maximum entropy spectrum is

$$\hat{P}_{mem}(e^{j\omega}) = \frac{\epsilon_p}{\left| 1 + \sum_{k=1}^p a_p(k)e^{-jk\omega} \right|^2}$$

where the coefficients  $a_p(k)$  are the solution to the normal equations  $\mathbf{R}_x \mathbf{a}_p = \epsilon_p \mathbf{u}_1$ . If  $\Gamma_k$  are the reflection coefficients produced by the Levinson-Durbin recursion, show that the MEM spectrum may be upper and lower bounded in terms of  $\Gamma_k$  as follows,

$$r_x(0) \prod_{k=1}^p \frac{1 - |\Gamma_k|}{1 + |\Gamma_k|} \leq \hat{P}_{mem}(e^{j\omega}) \leq r_x(0) \prod_{k=1}^p \frac{1 + |\Gamma_k|}{1 - |\Gamma_k|}$$

Hint: Begin with the frequency domain version of the Levinson order-update equation and use the inequality,

$$||a| - |b|| \leq |a + b| \leq ||a| + |b||$$

- 8.17.** Let  $x(n)$  be a first-order Gaussian autoregressive process with power spectrum

$$P_x(z) = \frac{c}{(1 - az^{-1})(1 - az)}$$

where  $a$  and  $c$  are real numbers.

- (a) With the constraint that the total power in the process is equal to one, find the value or values of  $a$  and  $c$  that maximize the entropy of  $x(n)$ .  
(b) Repeat part (a) and find the value or values of  $a$  and  $c$  that minimize the entropy.

**8.18.** The estimated autocorrelation sequence of a random process  $x(n)$  for lags  $k = 0, 1, 2, 3, 4$  are

$$r_x(0) = 2 ; \quad r_x(1) = 1 ; \quad r_x(2) = 1 ; \quad r_x(3) = 0.5 ; \quad r_x(4) = 0$$

Estimate the power spectrum of  $x(n)$  for each of the following cases.

- (a)  $x(n)$  is an AR(2) process.
- (b)  $x(n)$  is an MA(2) process.
- (c)  $x(n)$  is an ARMA(1,1) process.
- (d)  $x(n)$  contains a single sinusoid in white noise.

**8.19.** The first three values of the autocorrelation sequence for a process  $x(n)$  are:

$$r_x(0) = 1 ; \quad r_x(1) = 0 ; \quad r_x(2) = -\alpha$$

where  $0 < \alpha < 1$ . The eigenvalues of the  $3 \times 3$  autocorrelation matrix that is formed from these autocorrelations are  $\lambda_1 = 1 + \alpha$ ,  $\lambda_2 = 1$ , and  $\lambda_3 = 1 - \alpha$ , and the corresponding eigenvectors are

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} ; \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} ; \quad \mathbf{v}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

- (a) Use the Blackman-Tukey method with a rectangular window to estimate the power spectrum of  $x(n)$ , and make a carefully labeled sketch of your estimate.
- (b) Suppose that  $x(n)$  is known to consist of two complex exponentials in white noise. Estimate the power spectrum of  $x(n)$  and make a carefully labeled sketch of your estimate.

**8.20.** Suppose that we would like to estimate the power spectrum of an AR(2) process

$$x(n) = a(1)x(n-1) + a(2)x(n-2) + w(n)$$

where  $w(n)$  is unit variance white noise. However, our measurements of  $x(n)$  are noisy, and what we observe is the process

$$y(n) = x(n) + v(n)$$

where the measurement noise,  $v(n)$ , is uncorrelated with  $x(n)$ . It is known that  $v(n)$  is a first-order moving average process,

$$v(n) = b(0)q(n) + b(1)q(n-1)$$

where  $q(n)$  is white noise. Based on measurements of  $v(n)$ , the power spectrum of  $v(n)$  is estimated to be

$$\hat{P}_v(e^{j\omega}) = 3 + 2 \cos \omega$$

From  $y(n)$  we estimate the following values of the autocorrelation sequence  $r_y(k)$ ,

$$\hat{r}_y(0) = 5 ; \quad \hat{r}_y(1) = 2 ; \quad \hat{r}_y(2) = 0 ; \quad \hat{r}_y(3) = -1 ; \quad \hat{r}_y(4) = 0.5$$

Using all of the given information, estimate the power spectrum of  $x(n)$  using the maximum entropy method.

**8.21.** Show that for  $N \gg 1$ , estimating the order of an autoregressive process by minimizing  $\text{FPE}(p)$  is equivalent to minimizing  $\text{AIC}(p)$ . Hint: Show that for large  $N$ ,

$$N \ln \text{FPE}(p) \approx \text{AIC}(p)$$

and use the fact that, if  $x$  is small, then  $\ln(1 + x) \approx x$ .

**8.22.** You are given the following values for the autocorrelation sequence of a wide-sense stationary process  $x(n)$ ,

$$r_x(0) = 2 \quad ; \quad r_x(1) = \sqrt{3}/2 \quad ; \quad r_x(2) = 0.5$$

The eigenvalues of the  $3 \times 3$  Toeplitz autocorrelation matrix are  $\lambda_1 = 3.5$ ,  $\lambda_2 = 1.5$ , and  $\lambda_3 = 1.0$  and the corresponding normalized eigenvectors are

$$\mathbf{v}_1 = \sqrt{2/5} \begin{bmatrix} \sqrt{3}/2 \\ 1 \\ \sqrt{3}/2 \end{bmatrix} \quad ; \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad ; \quad \mathbf{v}_3 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ -\sqrt{3} \\ 1 \end{bmatrix}$$

It is known that  $x(n)$  consists of a single sinusoid in white noise.

- (a) Estimate the frequency of the sinusoid using the Blackman-Tukey method of frequency estimation.
- (b) Use the MUSIC algorithm to estimate the frequency of the sinusoid.
- (c) Repeat part (b) using the minimum norm algorithm.

**8.23.** The Pisarenko harmonic decomposition provides a way to estimate the frequencies of a sum of complex exponentials in white noise. As described in Sect. 8.6.2, the powers of the complex exponentials may be found by solving the set of linear equations given in Eq. (8.160). Another method that may be used is based on the orthogonality of the trigonometric sine and cosine functions. This orthogonality condition implies that

$$\det \begin{bmatrix} \sin \omega_1 & \sin \omega_2 & \cdots & \sin \omega_p \\ \sin 2\omega_1 & \sin 2\omega_2 & \cdots & \sin 2\omega_p \\ \vdots & \vdots & & \vdots \\ \sin p\omega_1 & \sin p\omega_2 & \cdots & \sin p\omega_p \end{bmatrix} \neq 0$$

and

$$\det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \cos \omega_1 & \cos \omega_2 & \cdots & \cos \omega_p \\ \vdots & \vdots & & \vdots \\ \cos(p-1)\omega_1 & \cos(p-1)\omega_2 & \cdots & \cos(p-1)\omega_p \end{bmatrix} \neq 0$$

provided  $0 < \omega_i < \pi$  and  $\omega_i \neq \omega_j$ .

- (a) Given the autocorrelation sequence of a  $p$ th-order harmonic process,

$$r_x(k) = \sum_{i=1}^p P_i e^{j k \omega_i} + \sigma_w^2 \delta(k)$$

evaluate the imaginary part of  $r_x(k)$  and use the orthogonality of the sine functions to derive a set of linear equations that may be solved to find the signal powers  $P_i$ .

- (b) How would you modify this algorithm if some of the frequencies were equal to zero or  $\pi$ ?  
(c) How would you modify this approach for a sum of sinusoids in white noise?

**8.24.** The Pisarenko harmonic decomposition was derived for a process that consists of a sum of complex exponentials in white noise. In this problem we generalize the decomposition to nonwhite noise. To accomplish this, we begin with an alternate derivation of the Pisarenko decomposition for white noise. Let

$$x(n) = \sum_{k=1}^p A_k e^{jn\omega_k} + w(n)$$

where  $w(n)$  is noise that is uncorrelated with the complex exponentials.

- (a) If  $w(n)$  is white noise then, as we saw in Eq. (8.149), the autocorrelation matrix for  $x(n)$  may be written as

$$\mathbf{R}_x = \mathbf{E}\mathbf{P}\mathbf{E}^H + \sigma_w^2 \mathbf{I}$$

where  $\mathbf{E}$  is a matrix of complex exponentials and  $\mathbf{P}$  is a diagonal matrix of signal powers. If  $x(n)$  is filtered with a  $p$ th-order FIR filter  $\mathbf{a} = [a(0), a(1), \dots, a(p)]^T$ , then the power in the output process is

$$\xi = E\{|y(n)|^2\} = \mathbf{a}^H \mathbf{R}_x \mathbf{a}$$

If  $\mathbf{a}$  is constrained to have unit norm,  $\mathbf{a}^H \mathbf{a} = 1$ , show that the filter that minimizes  $\xi$  has  $p$  zeros on the unit circle at the frequencies  $\omega_k$  of the complex exponentials, and show that the minimum value of  $\xi$  is equal to  $\sigma_w^2$ .

- (b) Now assume that  $w(n)$  has an arbitrary power spectrum,  $P_w(e^{j\omega})$ . If the autocorrelation matrix of  $w(n)$  is  $\sigma_w^2 \mathbf{Q}$ , then the autocorrelation matrix for  $x(n)$  becomes

$$\mathbf{R}_x = \mathbf{E}\mathbf{P}\mathbf{E}^H + \sigma_w^2 \mathbf{Q}$$

Suppose that  $x(n)$  is filtered with a  $p$ th-order FIR filter  $\mathbf{a} = [a(0), a(1), \dots, a(p)]^T$  that is normalized so that

$$\mathbf{a}^H \mathbf{Q} \mathbf{a} = 1$$

Show that the filter that minimizes the power in the filtered process has  $p$  zeros on the unit circle at the frequencies  $\omega_k$  of the complex exponentials, and that the minimum value is equal to  $\sigma_w^2$ .

- (c) Show that minimizing  $\xi = \mathbf{a}^H \mathbf{R}_x \mathbf{a}$  subject to the constraint  $\mathbf{a}^H \mathbf{Q} \mathbf{a} = 1$  is equivalent to solving the generalized eigenvalue problem

$$\mathbf{R}_x \mathbf{a} = \lambda \mathbf{Q} \mathbf{a}$$

for the minimum eigenvalue and eigenvector. Thus, the frequencies of the complex exponentials correspond to the roots of the polynomial that is formed from the minimum eigenvector

$$V_{\min}(z) = \sum_{k=0}^p v_{\min}(k) z^{-k}$$

and  $\sigma_w^2$  corresponds to the minimum eigenvalue.

- (d) A random process consists of single sinusoid in nonwhite noise,

$$x(n) = A \sin(n\omega_0 + \phi) + w(n)$$

The first three values of the autocorrelation sequence for  $x(n)$  are

$$\mathbf{r}_x = [9.515, 7.758, 6.472]^T$$

It is known that the additive noise  $w(n)$  is a moving average process that is generated by filtering white noise  $v(n)$  as follows

$$w(n) = v(n) + 0.1v(n-1)$$

However, the variance of  $v(n)$  is unknown. Find the frequency  $\omega_0$  and the power,  $P = \frac{1}{2}A^2$ , of the sinusoid.

- 8.25.** A random process is known to consist of a single sinusoid in white noise

$$x(n) = A \sin(n\omega_0 + \phi) + w(n)$$

where the variance of  $w(n)$  is  $\sigma_w^2$ .

- (a) Suppose that the first three values of the autocorrelation sequence are estimated and found to be

$$r_x(0) = 1 ; r_x(1) = \beta ; r_x(2) = 0$$

Find and prepare a carefully labeled sketch of the spectrum estimate that is formed using the Blackman-Tukey method with a rectangular window.

- (b) With the autocorrelations given in part (a), use the Pisarenko harmonic decomposition to estimate the variance of the white noise,  $\sigma_w^2$ , the frequency of the sinusoid,  $\omega_0$ , and the sinusoid power,  $P = \frac{1}{2}A^2$ . How does your estimate of the white noise power and the sinusoid frequency depend upon  $\beta$ ? Does the sinusoid power depend upon  $\beta$ ?
- (c) Suppose that we compute the periodogram  $\hat{P}_{per}(e^{j\omega})$  using  $N$  samples of  $x(n)$ . Find and prepare a carefully labeled sketch of the expected value of this spectrum estimate. Is this estimate biased? Is it consistent?
- (d) Using the autocorrelations given in part (a), find the second-order MEM power spectrum.

- 8.26.** A random process may be classified in terms of the properties of the prediction error sequence  $\epsilon_k$  that is produced when fitting an all-pole model to the process. Listed below are five different classifications for the error sequence:

1.  $\epsilon_k = c > 0$  for all  $k \geq 0$ .
2.  $\epsilon_k = c > 0$  for all  $k \geq k_0$  for some  $k_0 > 0$ .
3.  $\epsilon_k \rightarrow c$  as  $k \rightarrow \infty$  where  $c > 0$ .
4.  $\epsilon_k \rightarrow 0$  as  $k \rightarrow \infty$ .
5.  $\epsilon_k = 0$  for all  $k \geq k_0$  for some  $k_0 > 0$ .

For each of these classifications, describe as completely as possible the characteristics that may be attributed to the process and its power spectrum.

**8.27.** In the MUSIC algorithm, finding the peaks of the frequency estimation function

$$\hat{P}_{MU}(e^{j\omega}) = \frac{1}{\sum_{i=p+1}^M |\mathbf{e}^H \mathbf{v}_i|^2}$$

is equivalent to finding the minima of the denominator. Show that finding the *minima* of the denominator is equivalent to finding the *maxima* of

$$\sum_{i=1}^p |\mathbf{e}^H \mathbf{v}_i|^2$$

Hint: Use the fact that

$$\mathbf{I} = \sum_{i=1}^M \mathbf{v}_i \mathbf{v}_i^H$$

**8.28.** The  $3 \times 3$  autocorrelation matrix of a harmonic process is

$$\mathbf{R}_x = \begin{bmatrix} 3 & -j & -1 \\ j & 3 & -j \\ -1 & j & 3 \end{bmatrix}$$

- (a) Using the Pisarenko harmonic decomposition, find the complex exponential frequencies and the variance of the white noise.
- (b) Repeat part (a) using the MUSIC algorithm, the eigenvector method, and the minimum norm method.

**8.29.** In this problem we prove that the spurious roots in the minimum norm method lie inside the unit circle. Let  $x(n)$  be a random process that is a sum of  $p$  complex exponentials in white noise, and let  $\mathbf{a}$  be an  $M$ -dimensional vector that lies in the noise subspace. The  $z$ -transform of  $\mathbf{a}$  may be factored as follows

$$A(z) = A_0(z)A_1(z)$$

where

$$A_0(z) = \prod_{k=1}^p (1 - e^{j\omega_k} z^{-1})$$

is a monic polynomial that has  $p$  roots on the unit circle at the frequencies of the complex exponentials in  $x(n)$ , and  $A_1(z)$  is a polynomial that contains the  $M - p - 1$  spurious roots.

- (a) Show that minimizing  $\|\mathbf{a}\|^2$  is equivalent to minimizing

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |A(e^{j\omega})|^2 d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} |A_0(e^{j\omega})|^2 |A_1(e^{j\omega})|^2 d\omega$$

where  $A_0(e^{j\omega})$  is fixed and  $A_1(e^{j\omega})$  is monic.

- (b) From the results of part (a), show that minimizing  $\|\mathbf{a}\|^2$  is thus equivalent to using the autocorrelation method to find the prediction error filter  $A_1(z)$  for the signal whose  $z$ -transform is  $A_0(z)$ .
- (c) From (c), argue that  $A_1(z)$  must therefore have all of its roots inside the unit circle.

**8.30.** In the minimum norm method, the spurious zeros in the polynomial  $A(z)$  are separated from those that lie on the unit circle by forcing the spurious roots to lie inside the unit circle. In some applications of eigenvector methods, such as system identification, some of the desired zeros may lie inside the unit circle. In this case, the desired roots cannot be distinguished from the spurious roots. The minimum norm method may be modified, however, to force the spurious zeros to lie *outside* the unit circle. This is done by constraining the *last* element of the vector  $\mathbf{a}$  to have a value of one, i.e.,

$$\mathbf{a}^H \mathbf{u}_M = 1$$

where  $\mathbf{u}_M = [0, 0, \dots, 0, 1]^T$  is a unit vector with the last element equal to one.

- (a) Derive the *modified minimum norm* algorithm that uses the constant that  $\mathbf{a}^H \mathbf{u}_M = 1$  instead of  $\mathbf{a}^H \mathbf{u}_1 = 1$  as in the minimum norm algorithm.
- (b) The  $3 \times 3$  autocorrelation matrix for a single complex exponential in white noise is

$$\mathbf{R}_x = \begin{bmatrix} 2 & 1-j & -j\sqrt{2} \\ 1+j & 2 & -j \\ j\sqrt{2} & j & 2 \end{bmatrix}$$

Find the frequency of the complex exponential and the locations of the spurious roots in the minimum norm frequency estimation function.

- (c) Repeat part (b) for the modified minimum norm method.



### Computer Exercises

**C8.1.** Consider the broadband AR(4) process  $x(n)$  that is produced by filtering unit variance white Gaussian noise with the filter

$$H(z) = \frac{1}{(1 - 0.5z^{-1} + 0.5z^{-2})(1 + 0.5z^{-2})}$$

- (a) Generate  $N = 256$  samples of this process and estimate the power spectrum using the autocorrelation method with  $p = 4$ . Make a plot of the estimate and compare it to the true power spectrum.
- (b) Repeat part (a) for 20 different realizations of the process  $x(n)$ . Generate an overlay plot of the 20 estimates and plot the ensemble average. Comment on the variance of the estimate and on how accurately the autocorrelation method is able to estimate the power spectrum.
- (c) Repeat part (b) using model orders of  $p = 6, 8$ , and  $12$ . Describe what happens when the model order becomes too large.
- (d) Repeat parts (b) and (c) for the covariance, modified covariance, and Burg algorithms. Which approach seems to be the best for a broadband autoregressive process?

**C8.2.** Repeat Problem C8.1 for the narrowband AR(4) process that is generated by filtering unit variance white noise with

$$H(z) = \frac{1}{(1 - 1.585z^{-1} + 0.96z^{-2})(1 - 1.152z^{-1} + 0.96z^{-2})}$$

**C8.3.** Consider the process

$$y(n) = x(n) + w(n)$$

where  $w(n)$  is white Gaussian noise with a variance  $\sigma_w^2$  and  $x(n)$  is an AR(2) process that is generated by filtering unit variance white noise with the filter

$$H(z) = \frac{1}{1 - 1.585z^{-1} + 0.96z^{-2}}$$

- (a) Plot the power spectrum of  $x(n)$  and  $y(n)$ .
- (b) For  $\sigma_w^2 = 0.5, 1, 2, 5$ , generate  $N = 100$  samples of the process  $y(n)$ , and estimate the power spectrum of  $x(n)$  from  $y(n)$  using the maximum entropy method with  $p = 2$ . What is the effect of the noise  $w(n)$  on the accuracy of the spectrum estimate?
- (c) Repeat part (b) using the maximum entropy method with  $p = 5$ . Describe your observations.
- (d) Since the autocorrelation sequence of  $y(n)$  is

$$r_y(k) = r_x(k) + \sigma_w^2 \delta(k)$$

investigate what happens to your estimate in (c) if the autocorrelation sequence is modified by subtracting  $\sigma_w^2$  from  $r_y(0)$  before the maximum entropy spectrum is computed. Does this improve the spectrum estimate?

**C8.4.** In this exercise, we look at what happens if an autoregressive spectrum estimation technique is used on a moving average process.

- (a) Let  $x(n)$  be the second-order moving average process that is formed by filtering unit variance white Gaussian noise  $w(n)$  as follows,

$$x(n) = w(n) - w(n - 2)$$

Examine the spectrum estimates that are produced using an autoregressive technique such as MEM or the covariance method, and discuss your findings. What happens as you let the model order become large?

- (b) Repeat part (a) for the MA(3) process that is formed by filtering unit variance white Gaussian noise with the filter

$$H(z) = (1 - 0.98z^{-1})(1 - 0.96z^{-2})$$

**C8.5.** Write a MATLAB program to estimate the order of an all-pole random process using the Akaike final prediction error, the minimum descriptor length, the Akaike information criterion, and Parzen's CAT. Evaluate the accuracy of the estimates that are produced with these methods if the modeling errors  $\mathcal{E}_p$  are derived using the autocorrelation method. Consider both broadband and narrowband AR processes. How much do the estimates change if  $\mathcal{E}_p$  is replaced with the Burg error,  $\mathcal{E}_p^B$ ? Discuss your findings.

**C8.6.** Consider the process  $x(n)$  consisting of two sinusoids in noise,

$$x(n) = 2 \cos(n\omega_1 + \phi_1) + 2 \cos(n\omega_2 + \phi_2) + w(n)$$

where  $\phi_1$  and  $\phi_2$  are uncorrelated random variables that are uniformly distributed over the interval  $[0, 2\pi]$ , and  $w(n)$  is a fourth-order moving average process that is generated by filtering unit variance white noise with a linear shift-invariant filter that has a system function

$$H(z) = (1 - z^{-1})(1 + z^{-1})^3$$

Let  $\omega_1 = \pi/2$  and  $\omega_2 = 1.1\pi/2$ .

- (a) Plot the power spectrum of  $x(n)$ .
- (b) Find the variance of the moving average process and compare the power in  $w(n)$  to the power in each of the sinusoids.
- (c) How many values of  $x(n)$  are necessary in order to resolve the frequencies of the two sinusoids using the periodogram?
- (d) Compare the expected value of the periodogram at the sinusoid frequencies  $\omega_1$  and  $\omega_2$  with the power spectrum of the noise  $P_w(e^{j\omega})$ . What record length  $N$  is required in order for the expected value of the periodogram at  $\omega_1$  and  $\omega_2$  to be twice the value of  $P_w(e^{j\omega})$ ?
- (e) Generate a sample realization of  $x(n)$  of length  $N = 256$  and plot the periodogram. Are the two sinusoidal frequencies clearly visible in the spectrum estimate? Repeat for 20 different realizations of the process and discuss your findings.
- (f) Estimate the expected value of the periodogram by averaging the periodograms that are formed from an ensemble of 50 realizations of  $x(n)$ . Plot the estimate and comment on its resolution and on how closely it estimates the moving average part of the spectrum.
- (g) Repeat parts (e) and (f) using Bartlett's method with  $K = 2, 4$ , and 8 sections.

**C8.7.** Let  $x(n)$  be the process defined in Problem C8.6 consisting of a fourth-order moving average process plus two sinusoids.

- (a) Generate a sample realization of  $x(n)$  of length  $N = 256$ , and estimate the spectrum using the maximum entropy method with  $p = 4, 8, 16$ , and 32. Repeat for 20 different realizations of the process, and generate an overlay plot of the estimates along with the ensemble average. Discuss your findings.
- (b) Repeat part (a) using the Burg algorithm and the modified covariance method.

**C8.8.** One of the methods we have seen for estimating the numerator and denominator coefficients of an ARMA process is iterative prefiltering. In this exercise we consider the use of iterative prefiltering for spectrum estimation.

- (a) Generate 100 samples of an ARMA(2,2) process  $x(n)$  by filtering unit variance white Gaussian noise with a filter that has a system function of the form

$$H(z) = \frac{1 + 1.5z^{-1} + z^{-2}}{1 + 0.9z^{-2}}$$

Using the method of iterative prefiltering, find the second-order ARMA model for  $x(n)$ , i.e.,  $p = q = 2$ . Make a plot of the power spectrum that is formed from this model, and compare it to the true power spectrum.

- (b) Repeat your experiment in part (a) for 20 different realizations of  $x(n)$ . How close are the estimated model coefficients to the true model, on the average? How accurate are the estimates of the power spectrum that are generated from these models? What happens if the process is over-modeled by setting  $p = q = 3$ ? What about  $p = q = 5$ ?

**C8.9.** Let  $P_x(e^{j\omega})$  be the power spectrum of a process that has an autocorrelation sequence  $r_x(k)$ , and let  $\lambda_k$  be the eigenvalues of the  $N \times N$  autocorrelation matrix  $\mathbf{R}_x$ .

- (a) Use Szegő's theorem (see Prob. 8.14) to estimate the eigenvalues of the  $128 \times 128$  autocorrelation matrix of the lowpass random process  $x(n)$  that has a power spectrum

$$P_x(e^{j\omega}) = \begin{cases} 1 & ; \quad |\omega| < \pi/2 \\ 0 & ; \quad \pi/2 \leq |\omega| \leq \pi \end{cases}$$

- (b) Generate the  $128 \times 128$  autocorrelation matrix for this process, find the eigenvalues, and plot them in increasing order. Given the autocorrelation sequence that is stored in the vector  $r$ , note that this plot may be generated with the single MATLAB command

```
plot(sort(eig(toeplitz(r)))).
```

Are your results consistent with the estimate found in part (a)?

- (c) Repeat parts (a) and (b) for the harmonic process

$$x(n) = A \cos(n\omega_0 + \phi) + w(n)$$

where  $w(n)$  is unit variance white noise.

**C8.10.** It has been said that the minimum variance spectrum estimate exhibits less variance than an AR spectrum estimate for long data records [30]. To investigate this claim, consider the autocorrelation sequence

$$r_x(k) = \cos(0.35\pi k) + \delta(k) + w(k) \quad ; \quad k = 0, 1, \dots, p$$

where  $w(k)$  is uniformly distributed white noise with zero mean and variance  $\sigma_w^2$ . This sequence represents a noisy estimate of the autocorrelation sequence of a random phase sinusoid of frequency  $\omega = 0.35\pi$  in unit variance white noise.

- (a) Compute the minimum variance spectrum estimate with  $p = 10$  and  $\sigma_w^2 = 0$ .
- (b) Generate an overlay plot of 25 minimum variance spectrum estimates with  $\sigma_w^2 = 0.1$  and compare these estimates to that generated in part (a).
- (c) Repeat parts (a) and (b) using the maximum entropy method and compare your results with the minimum variance estimates. Is the claim substantiated? Suppose the Burg algorithm is used instead of MEM. Are your results any different?

**C8.11.** In the minimum variance spectrum estimate method, the bandwidths of the bandpass filters are the same for each filter in the filter bank,  $\Delta\omega = 2\pi/p$ . However, since each filter  $\mathbf{g}_i$  has a bandwidth that depends, in general, on the center frequency of the filter, it may be possible to *improve* the power estimate by modifying the definition for the filter bandwidth. For example, consider the “equivalent filter bandwidth,” which is defined by

$$\Delta = \mathbf{g}_i^H \mathbf{g}_i$$

- (a) Show that the resulting *modified MV spectrum* is given by

$$\hat{P}_x(e^{j\omega}) = \frac{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}}{\mathbf{e}^H \mathbf{R}_x^{-2} \mathbf{e}}$$

- (b) What is the modified MV estimate of white Gaussian noise that has a variance of  $\sigma_v^2$ .
- (c) Modify the m-file `minvar.m` so that it finds the modified minimum variance spectrum estimate of a random process  $x(n)$ .
- (d) Let  $x(n)$  be a harmonic process consisting of a sum of two sinusoids in white noise. Let the sinusoid frequencies be  $\omega_1 = 0.2\pi$  and  $\omega_2 = 0.25\pi$ , and assume that the