2

Nonparametric Methods

2.1 INTRODUCTION

The nonparametric methods of spectral estimation rely entirely on the definitions (1.3.7) and (1.3.10) of PSD to provide spectral estimates. These methods constitute the "classical means" for PSD estimation [JENKINS AND WATTS 1968]. The present chapter reviews the main nonparametric methods, their properties, and the Fast Fourier Transform (FFT) algorithm for their implementation. A related discussion is to be found in Chapter 5, where the nonparametric approach to PSD estimation is given a filter-bank interpretation.

We first introduce two common spectral estimators, the *periodogram* and the *correlogram*, derived directly from (1.3.10) and (1.3.7), respectively. These methods are then shown to be equivalent under weak conditions. The periodogram and correlogram methods provide reasonably high resolution for sufficiently long data lengths, but are poor spectral estimators, because their variance is high and does not decrease with increasing data length. (In Chapter 5, we provide an interpretation of the periodogram and correlogram methods as a power estimate based on a *single* sample of a filtered version of the signal under study; it is thus not surprising that the periodogram or correlogram variance is large.)

The high variance of the periodogram and correlogram methods motivates the development of modified methods that have lower variance, at the cost of reduced resolution. Several modified methods have been introduced, and we present some of the most popular ones. We show them all to be, more or less, equivalent in their properties and performance for large data lengths.

2.2 PERIODOGRAM AND CORRELOGRAM METHODS

2.2.1 Periodogram

The periodogram method relies on the definition (1.3.10) of the PSD. Neglecting the expectation and the limit operation in (1.3.10), which cannot be performed when the only available information on the signal consists of the samples $\{y(t)\}_{t=1}^{N}$, we get

$$\hat{\phi}_p(\omega) = \frac{1}{N} \left| \sum_{t=1}^{N} y(t) e^{-i\omega t} \right|^2 \quad \text{(Periodogram)} \tag{2.2.1}$$

One of the first uses of the *periodogram spectral estimator*, (2.2.1), has been in determining possible "hidden periodicities" in time series, which may be seen as a motivation for the name of this method [SCHUSTER 1900].

2.2.2 Correlogram

The correlation-based definition (1.3.7) of the PSD leads to the *correlogram spectral estimator* [BLACKMAN AND TUKEY 1959]:

$$\hat{\phi}_c(\omega) = \sum_{k=-(N-1)}^{N-1} \hat{r}(k) e^{-i\omega k} \quad \text{(Correlogram)} \tag{2.2.2}$$

where $\hat{r}(k)$ denotes an estimate of the covariance lag $r(k)$, obtained from the available sample $\{y(1), \ldots, y(N)\}$. When no assumption is made on the signal under study, except for the stationarity assumption, there are two standard ways to obtain the sample covariances required in (2.2.2):

$$\hat{r}(k) = \frac{1}{N-k} \sum_{t=k+1}^{N} y(t) y^*(t-k), \qquad 0 \le k \le N-1 \tag{2.2.3}$$

and

$$\hat{r}(k) = \frac{1}{N} \sum_{t=k+1}^{N} y(t) y^*(t-k) \qquad 0 \le k \le N-1 \tag{2.2.4}$$

The sample covariances for negative lags are then constructed by using the property (1.3.3) of the covariance function:

$$\hat{r}(-k) = \hat{r}^*(k), \qquad k = 0, \ldots, N-1 \tag{2.2.5}$$

The estimator (2.2.3) is called the standard unbiased ACS estimate, and (2.2.4) is called the standard biased ACS estimate. The biased ACS estimate is most commonly used, for the following reasons:

- For most stationary signals, the covariance function decays rather rapidly, so that $r(k)$ is quite small for large lags k. Comparing the definitions (2.2.3) and (2.2.4), it can be seen that $\hat{r}(k)$ in (2.2.4) will be small for large k (provided N is reasonably large), whereas $\hat{r}(k)$ in (2.2.3) could take large and erratic values for large k, as it is obtained by averaging only a few products in such a case (in particular, only one product for $k = N - 1$!). This observation implies that (2.2.4) is likely to be a more accurate estimator of $r(k)$ than (2.2.3) for medium and large values of k (compared with N). For small values of k, the two estimators in (2.2.3) and (2.2.4) can be expected to behave in a similar manner.
- The sequence $\{\hat{r}(k), k = 0, \pm 1, \pm 2, \ldots\}$ obtained with (2.2.4) is guaranteed to be positive semidefinite (as it should, see equation (1.3.5) and the related discussion), but this is not the case for (2.2.3). This fact is especially important for PSD estimation, since a sample covariance sequence that is not positive definite, when inserted in (2.2.2), may lead to negative spectral estimates, and this is undesirable in most applications.

When the sample covariances (2.2.4) are inserted in (2.2.2), it can be shown that the spectral estimate so obtained is identical to (2.2.1). In other words, we have the following result:

$\hat{\phi}_c(\omega)$, when evaluated by using the standard biased ACS estimates, coincides with $\hat{\phi}_p(\omega)$.

(2.2.6)

A simple proof of (2.2.6) runs as follows: Consider the signal

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} y(k)e(t - k) \qquad (2.2.7)$$

where $\{y(k)\}$ are considered to be fixed (nonrandom) constants and $e(t)$ is a white noise of unit variance: $E\{e(t)e^*(s)\} = \delta_{t,s}$ (= 1 if $t = s$; and = 0 otherwise). Hence, $x(t)$ is the output of a filter with the following transfer function:

$$Y(\omega) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} y(k)e^{-i\omega k}$$

Since the PSD of the input to the filter is given by $\phi_e(\omega) = 1$, it follows from (1.4.5) that

$$\phi_x(\omega) = |Y(\omega)|^2 = \hat{\phi}_p(\omega) \qquad (2.2.8)$$

On the other hand, a straightforward calculation gives (for $k \geq 0$)

$$
\begin{aligned}
r_x(k) &= E\left\{x(t)x^*(t-k)\right\} \\
&= \frac{1}{N}\sum_{p=1}^{N}\sum_{s=1}^{N} y(p)y^*(s)E\left\{e(t-p)e^*(t-k-s)\right\} \\
&= \frac{1}{N}\sum_{p=1}^{N}\sum_{s=1}^{N} y(p)y^*(s)\delta_{p,k+s} = \frac{1}{N}\sum_{p=k+1}^{N} y(p)y^*(p-k) \\
&= \begin{cases} \hat{r}(k) \text{ given by (2.2.4),} & k = 0,\dots,N-1 \\ 0, & k \geq N \end{cases}
\end{aligned}
\tag{2.2.9}
$$

Inserting (2.2.9) in the definition (1.3.7) of PSD yields the following alternative expression for $\phi_x(\omega)$:

$$
\phi_x(\omega) = \sum_{k=-(N-1)}^{N-1} \hat{r}(k)e^{-i\omega k} = \hat{\phi}_c(\omega)
\tag{2.2.10}
$$

Comparing (2.2.8) and (2.2.10) concludes the proof of the claim (2.2.6).

The equivalence of the periodogram and correlogram spectral estimators can be used to derive their properties simultaneously. These two methods are shown in Section 2.4 to provide *poor estimates* of the PSD. There are two reasons for this, and both can be explained intuitively by using $\hat{\phi}_c(\omega)$:

- The estimation errors in $\hat{r}(k)$ are on the order of $1/\sqrt{N}$ for large N (see Exercise 2.4), at least for $|k|$ not too close to N. Because $\hat{\phi}_c(\omega) = \hat{\phi}_p(\omega)$ is a sum that involves $(2N-1)$ such covariance estimates, the difference between the true and estimated spectra will be a sum of "many small" errors. Hence, there is no guarantee that the total error will die out as N increases. The spectrum estimation error is even larger than what is suggested by the preceding discussion, because errors in $\{\hat{r}(k)\}$, for $|k|$ close to N, are typically of an order larger than $1/\sqrt{N}$. The consequence is that the variance of $\hat{\phi}_c(\omega)$ does not go to zero as N increases.
- In addition, if $r(k)$ converges slowly to zero, then the periodogram estimates will be biased. Indeed, for lags $|k| \simeq N$, $\hat{r}(k)$ will be a poor estimate of $r(k)$ since $\hat{r}(k)$ is the sum of only a few lag products that are divided by N (see equation (2.2.4)). Thus, $\hat{r}(k)$ will be much closer to zero than $r(k)$ is; in fact, $E\left\{\hat{r}(k)\right\} = [(N-|k|)/N]r(k)$, and the bias is significant for $|k| \simeq N$ if $r(k)$ is not close to zero in this region. If $r(k)$ decays rapidly to zero, the bias will be small and will not contribute significantly to the total error in $\hat{\phi}_c(\omega)$; however, the nonzero variance just discussed will still be present.

Both the bias and the variance of the periodogram are discussed more quantitatively in Section 2.4.

Another intuitive explanation for the poor statistical accuracy of the periodogram and correlogram methods is given in Chapter 5, where it is shown, roughly speaking, that these methods

can be viewed as procedures attempting to estimate the variance of a data sequence from a *single* sample.

In spite of their poor quality as spectral estimators, the periodogram and correlogram methods form the basis for the improved nonparametric spectral estimation methods, to be discussed later in this chapter. As such, computation of these two basic estimators is relevant to many other nonparametric estimators derived from them. The next section addresses this computational task.

2.3 PERIODOGRAM COMPUTATION VIA FFT

In practice, it is not possible to evaluate $\hat{\phi}_p(\omega)$ (or $\hat{\phi}_c(\omega)$) over a continuum of frequencies. Hence, the frequency variable must be sampled for the purpose of computing $\hat{\phi}_p(\omega)$. The following frequency sampling scheme is most commonly used:

$$\omega = \frac{2\pi}{N}k, \qquad k = 0, \ldots, N-1 \qquad (2.3.1)$$

Define

$$W = e^{-i\frac{2\pi}{N}} \qquad (2.3.2)$$

Then, evaluation of $\hat{\phi}_p(\omega)$ (or $\hat{\phi}_c(\omega)$) at the frequency samples in (2.3.1) basically reduces to the computation of the following Discrete Fourier Transform (DFT):

$$Y(k) = \sum_{t=1}^{N} y(t)W^{(t-1)k}, \qquad k = 0, \ldots, N-1 \qquad (2.3.3)$$

A direct evaluation of (2.3.3) would require about N^2 complex multiplications and additions, which might be a prohibitive burden for large values of N. Any procedure that computes (2.3.3) in less than N^2 flops (1 flop = 1 complex multiplication plus 1 complex addition) is called a Fast Fourier Transform (FFT) algorithm. In recent years, there has been significant interest in developing more and more computationally efficient FFT algorithms. In the following, we review one of the first FFT procedures—the so-called radix-2 FFT—which, while not being the most computationally efficient of all, is easy to program in a computer and yet quite computationally efficient [COOLEY AND TUKEY 1965; PROAKIS, RADER, LING, AND NIKIAS 1992].

2.3.1 Radix-2 FFT

Assume that N is a power of 2:

$$N = 2^m \qquad (2.3.4)$$

If this is not the case, then we can resort to *zero padding*, as described in the next subsection. By splitting the sum in (2.3.3) into two parts, we get

$$Y(k) = \sum_{t=1}^{N/2} y(t) W^{(t-1)k} + \sum_{t=N/2+1}^{N} y(t) W^{(t-1)k}$$

$$= \sum_{t=1}^{N/2} [y(t) + y(t+N/2) W^{\frac{Nk}{2}}] W^{(t-1)k} \qquad (2.3.5)$$

Next, note that

$$W^{\frac{Nk}{2}} = \begin{cases} 1, & \text{for even } k \\ -1, & \text{for odd } k \end{cases} \qquad (2.3.6)$$

Using this simple observation in (2.3.5), we obtain

For $k = 2p = 0, 2, \dots$

$$Y(2p) = \sum_{t=1}^{\bar{N}} [y(t) + y(t + \bar{N})] \bar{W}^{(t-1)p} \qquad (2.3.7)$$

For $k = 2p + 1 = 1, 3, \dots$

$$Y(2p+1) = \sum_{t=1}^{\bar{N}} \{[y(t) - y(t + \bar{N})] W^t\} \bar{W}^{(t-1)p} \qquad (2.3.8)$$

---

where $\bar{N} = N/2$ and $\bar{W} = W^2 = e^{-i2\pi/\bar{N}}$.

    These two equations are the core of the radix-2 FFT algorithm. Both of these equations represent DFTs for sequences of length equal to $\bar{N}$. Computation of the sequences transformed in (2.3.7) and (2.3.8) requires roughly $\bar{N}$ flops. Hence, the computation of an N-point transform has been reduced to the evaluation of two $N/2$-point transforms plus a sequence computation requiring about $N/2$ flops. This reduction process is continued until $\bar{N} = 1$ (which is made possible by requiring N to be a power of 2).

    In order to evaluate the number of flops required by a radix-2 FFT, let c_k denote the computational cost (expressed in flops) of a 2^k-point radix-2 FFT. According to the discussion in the previous paragraph, c_k satisfies the recursion

$$c_k = 2^k/2 + 2c_{k-1} = 2^{k-1} + 2c_{k-1} \qquad (2.3.9)$$

with initial condition $c_1 = 1$ (the number of flops required by a 1-point transform). By iterating (2.3.9), we obtain the solution

$$c_k = k2^{k-1} = \frac{1}{2}k2^k \tag{2.3.10}$$

from which it follows that $c_m = \frac{1}{2}m2^m = \frac{1}{2}N \log_2 N$; thus

> An N-point radix-2 FFT requires about $\frac{1}{2}N \log_2 N$ flops. (2.3.11)

As a comparison, the number of complex operations required to carry out an N-point *split-radix* FFT, which at present appears to be the most practical algorithm for general-purpose computers when N is a power of 2, is about $\frac{1}{3}N \log_2 N$. (See [PROAKIS, RADER, LING, AND NIKIAS 1992].)

2.3.2 Zero Padding

In some applications, N is not a power of 2 and so the previously described radix-2 FFT algorithm cannot be applied directly to the original data sequence. However, this is easily remedied; we may increase the length of the given sequence by means of zero padding $\{y(1), \dots, y(N), 0, 0, \dots\}$ until the length of the sequence so obtained is, say, L (which is generally chosen as a power of 2).

Zero padding is also useful when the frequency sampling (2.3.1) is considered to be too sparse to provide a good representation of the continuous-frequency estimated spectrum, for example $\hat{\phi}_p(\omega)$. Applying the FFT algorithm to the data sequence padded with zeroes gives

$$\hat{\phi}_p(\omega) \text{ at frequencies } \omega_k = \frac{2\pi k}{L}, \qquad 0 \le k \le L - 1$$

The finer sampling of $\hat{\phi}(\omega)$ can reveal details in the spectrum that were not visible without zero padding.

Since the *continuous-frequency* spectral estimate, $\hat{\phi}_p(\omega)$, is the same for both the original data sequence and the sequence padded with zeroes, zero padding cannot of course improve the spectral resolution of the periodogram methods. See [OPPENHEIM AND SCHAFER 1989; PORAT 1997] for further discussion.

In a zero-padded data sequence, the number of nonzero data points may be considerably smaller than the total number of samples—that is, $N \ll L$. In such a case, a significant time saving can be obtained by *pruning the FFT* algorithm by reducing or eliminating operations on zeroes. (See, for example, [MARKEL 1971].) FFT pruning, along with a decimation in time, can also be used to reduce the computation time when we want to evaluate the FFT only in a narrow region of the frequency domain. (See [MARKEL 1971].)

2.4 PROPERTIES OF THE PERIODOGRAM METHOD

The analysis of the statistical properties of $\hat{\phi}_p(\omega)$ (or $\hat{\phi}_c(\omega)$) is important, in that it shows the poor quality of the periodogram as an estimator of the PSD and, in addition, provides some insight into how we can modify the periodogram so as to obtain better spectral estimators. We split the analysis in two parts: bias analysis and variance analysis. (See also [PRIESTLEY 1981].)

The bias and variance of an estimator are two measures often used to characterize its performance. A primary motivation is that the total squared error of the estimate is the sum of the squared bias and the variance. To see this, let a denote any quantity to be estimated, and let $\hat{a}$ be an estimate of a. Then the mean squared error (MSE) of the estimate is

$$
\begin{aligned}
\text{MSE} \triangleq E\left\{|\hat{a} - a|^2\right\} &= E\left\{|\hat{a} - E\left\{\hat{a}\right\} + E\left\{\hat{a}\right\} - a|^2\right\} \\
&= E\left\{|\hat{a} - E\left\{\hat{a}\right\}|^2\right\} + |E\left\{\hat{a}\right\} - a|^2 \\
&\quad + 2\,\text{Re}\left[\left(E\left\{\hat{a} - E\left\{\hat{a}\right\}\right\}\right)^* \left(E\left\{\hat{a}\right\} - a\right)\right] \\
&= \text{var}\{\hat{a}\} + |\text{bias}\{\hat{a}\}|^2
\end{aligned}
\tag{2.4.1}
$$

By considering the bias and variance components of the MSE separately, we gain some additional insight into the source of error and ways to reduce the error.

2.4.1 Bias Analysis of the Periodogram

By using the result (2.2.6), we obtain

$$
E\left\{\hat{\phi}_p(\omega)\right\} = E\left\{\hat{\phi}_c(\omega)\right\} = \sum_{k=-(N-1)}^{N-1} E\left\{\hat{r}(k)\right\} e^{-i\omega k}
\tag{2.4.2}
$$

For $\hat{r}(k)$ defined in (2.2.4),

$$
E\left\{\hat{r}(k)\right\} = \left(1 - \frac{k}{N}\right) r(k), \qquad k \geq 0
\tag{2.4.3}
$$

and

$$
E\left\{\hat{r}(-k)\right\} = E\left\{\hat{r}^*(k)\right\} = \left(1 - \frac{k}{N}\right) r(-k), \qquad -k \leq 0
\tag{2.4.4}
$$

Hence,

$$
E\left\{\hat{\phi}_p(\omega)\right\} = \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N}\right) r(k) e^{-i\omega k}
\tag{2.4.5}
$$

Define

$$w_B(k) = \begin{cases} 1 - \dfrac{|k|}{N}, & k = 0, \pm 1, \ldots, \pm(N-1) \\ 0, & \text{otherwise} \end{cases} \tag{2.4.6}$$

The preceding sequence is called the *triangular window* or the *Bartlett window*. By using $w_B(k)$, we can write (2.4.5) as a DTFT:

$$E\left\{\hat{\phi}_p(\omega)\right\} = \sum_{k=-\infty}^{\infty} [w_B(k)r(k)]e^{-i\omega k} \tag{2.4.7}$$

The DTFT of the product of two sequences is equal to the convolution of their respective DTFTs. Hence, (2.4.7) leads to

$$\boxed{E\left\{\hat{\phi}_p(\omega)\right\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(\psi)W_B(\omega - \psi)d\psi} \tag{2.4.8}$$

where $W_B(\omega)$ is the DTFT of the triangular window. For completeness, we include a direct proof of (2.4.8). Inserting (1.3.8) in (2.4.7), we get

$$E\left\{\hat{\phi}_p(\omega)\right\} = \sum_{k=-\infty}^{\infty} w_B(k) \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(\psi)e^{i\psi k} d\psi\right] e^{-i\omega k} \tag{2.4.9}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(\psi) \left[\sum_{k=-\infty}^{\infty} w_B(k)e^{-ik(\omega-\psi)}\right] d\psi \tag{2.4.10}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(\psi)W_B(\omega - \psi)d\psi \tag{2.4.11}$$

which is (2.4.8).

We can find an explicit expression for $W_B(\omega)$ as follows:

$$W_B(\omega) = \sum_{k=-(N-1)}^{N-1} \frac{N - |k|}{N}e^{-i\omega k} \tag{2.4.12}$$

$$= \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} e^{-i\omega(t-s)} = \frac{1}{N} \left|\sum_{t=1}^{N} e^{i\omega t}\right|^2 \tag{2.4.13}$$

$$= \frac{1}{N} \left|\frac{e^{i\omega N} - 1}{e^{i\omega} - 1}\right|^2 = \frac{1}{N} \left|\frac{e^{i\omega N/2} - e^{-i\omega N/2}}{e^{i\omega/2} - e^{-i\omega/2}}\right|^2 \tag{2.4.14}$$
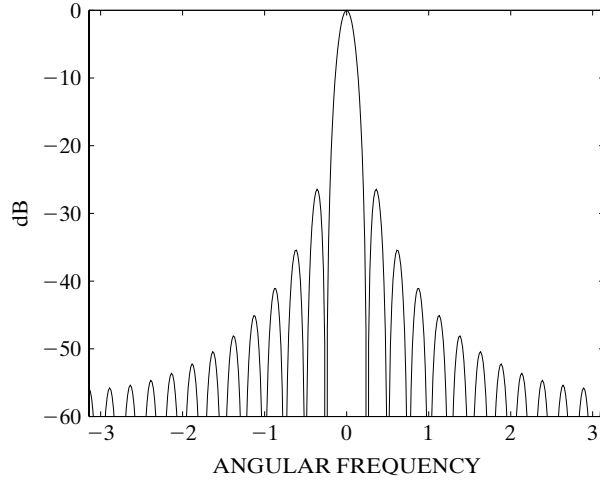
Figure 2.1 $W_B(\omega)/W_B(0)$, for $N = 25$.

or, in final form,

$$W_B(\omega) = \frac{1}{N} \left[\frac{\sin(\omega N/2)}{\sin(\omega/2)} \right]^2 \tag{2.4.15}$$

$W_B(\omega)$ is sometimes referred to as the *Fejer kernel*. As an illustration, $W_B(\omega)$ is displayed as a function of ω, in Figure 2.1.

The convolution formula (2.4.8) is the key equation for understanding the behavior of the mean estimated spectrum $E\{\hat{\phi}_p(\omega)\}$. In order to facilitate the interpretation of this equation, the reader may think of it as representing a dynamical system with "input" $\phi(\omega)$, "weighting function" $W_B(\omega)$, and "output" $E\{\hat{\phi}_p(\omega)\}$. Note that a similar equation would be obtained if the covariance estimator (2.2.3) were used in $\hat{\phi}_c(\omega)$, in lieu of (2.2.4). In that case, $E\{\hat{r}(k)\} = r(k)$, so the corresponding $W(\omega)$ function that would appear in (2.4.8) is the DTFT of the *rectangular window*

$$w_R(k) = \begin{cases} 1, & k = 0, \pm 1, \ldots, \pm(N-1) \\ 0, & \text{otherwise} \end{cases} \tag{2.4.16}$$

A straightforward calculation gives

$$W_R(\omega) = \sum_{k=-(N-1)}^{(N-1)} e^{-i\omega k} = 2\,\mathrm{Re}\left[\frac{e^{iN\omega} - 1}{e^{i\omega} - 1} \right] - 1$$

$$= \frac{2\cos\left[\frac{(N-1)\omega}{2} \right] \sin\left[\frac{N\omega}{2} \right]}{\sin\left[\frac{\omega}{2} \right]} - 1 = \frac{\sin\left[\left(N - \frac{1}{2}\right)\omega \right]}{\sin\left[\frac{\omega}{2} \right]} \tag{2.4.17}$$
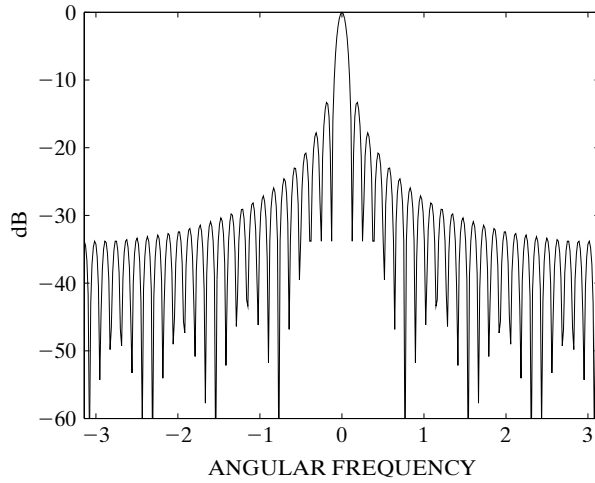
Figure 2.2 $W_R(\omega)/W_R(0)$, for $N = 25$.

which is displayed in Figure 2.2 (for $N = 25$; to facilitate comparison with $W_B(\omega)$). $W_R(\omega)$ is sometimes called the *Dirichlet kernel*.

As can be seen, there are no "essential" differences between $W_R(\omega)$ and $W_B(\omega)$. For conciseness, in the following we focus on the use of the triangular window.

Since we would like $E\{\hat{\phi}_p(\omega)\}$ to be as close to $\phi(\omega)$ as possible, it follows from (2.4.8) that $W_B(\omega)$ should be a close approximation to a Dirac impulse. The *half-power (3 dB) width* of the *main lobe* of $W_B(\omega)$ can be shown to be approximately $2\pi/N$ radians (see Exercise 2.15), so, in frequency units (with $f = \omega/2\pi$),

$$\boxed{\text{main lobe width in frequency } f \simeq 1/N} \qquad (2.4.18)$$

(Also, see the calculation of the time-bandwidth product for windows in the next section, which supports (2.4.18).) It follows from (2.4.18) that $W_B(\omega)$ is a poor approximation of a Dirac impulse for small values of N. In addition, unlike the Dirac delta function, $W_B(\omega)$ has a large number of *sidelobes*. It follows that the bias of the periodogram spectral estimate can basically be divided into two components. These two components correspond respectively to the nonzero main lobe width and the nonzero sidelobe height of the window function $W_B(\omega)$, as we explain next.

The principal effect of the *main lobe* of $W_B(\omega)$ is to smear or smooth the estimated spectrum. Assume, for instance, that $\phi(\omega)$ has two peaks separated in frequency f by less than $1/N$. Then these two peaks appear as a single broader peak in $E\{\hat{\phi}_p(\omega)\}$, because (per (2.4.8)), the "response" of the "system" corresponding to $W_B(\omega)$ to the first peak does not die out before the "response" to the second peak starts. This kind of effect of the main lobe on the estimated spectrum is called *smearing*. Smearing prevents the periodogram-based methods from resolving details in the studied spectrum that are separated in frequency by less than $1/N$ cycles per sampling interval. For this reason, $1/N$ is called the *spectral resolution limit* of the periodogram method.

Remark: The previous comments on resolution give us the occasion to stress that, in spite of the fact that we have seen the PSD as a function of the angular frequency (ω), we generally refer to *the resolution in frequency* (f) in units of cycles per sampling interval. Of course, the "resolution in angular frequency" is determined from the "resolution in frequency" by the simple relation $\omega = 2\pi f$. ∎

The principal effect of the *sidelobes* on the estimated spectrum consists of transferring power from the frequency bands that concentrate most of the power in the signal to bands that contain less or no power. This effect is called *leakage*. For instance, a dominant peak in $\phi(\omega)$ could, through convolution with the sidelobes of $W_B(\omega)$, lead to an estimated spectrum that contains power in frequency bands where $\phi(\omega)$ is zero. Note that the smearing effect associated with the main lobe can also be interpreted as a form of leakage from a local peak of $\phi(\omega)$ to neighboring frequency bands.

It follows from the previous discussion that smearing and leakage are particularly critical for spectra with large amplitude ranges, such as peaky spectra. For smooth spectra, these effects are less important. In particular, we see from (2.4.7) that, for *white noise* (which has a maximally smooth spectrum), the periodogram is an *unbiased* spectral estimator: $E\{\hat{\phi}_p(\omega)\} = \phi(\omega)$. (See also Exercise 2.9.)

The bias of the periodogram estimator, even though it might be severe for spectra with large dynamic ranges when the sample length is small, does not constitute the main limitation of this spectral estimator. In fact, if the bias were the only problem, then increasing N (assuming this is possible) would cause the bias in $\hat{\phi}_p(\omega)$ to be eliminated. In order to see this, note from (2.4.5) that

$$\lim_{N \to \infty} E\left\{\hat{\phi}_p(\omega)\right\} = \phi(\omega)$$

Hence, the periodogram is an *asymptotically unbiased spectral estimator*. The main problem of the periodogram method lies in its large variance, as is explained next.

2.4.2 Variance Analysis of the Periodogram

The finite-sample variance of $\hat{\phi}_p(\omega)$ can be easily established only in some specific cases, such as in the case of Gaussian white noise. The *asymptotic* variance of $\hat{\phi}_p(\omega)$, however, can be derived for more general signals. In the following, we present an *asymptotic (for $N \gg 1$) analysis* of the variance of $\hat{\phi}_p(\omega)$, since it turns out to be sufficient for showing the poor statistical accuracy of the periodogram. (For a finite-sample analysis, see Exercise 2.13.)

Some preliminary discussion is required. A sequence $\{e(t)\}$ is called *complex (or circular) white noise* if it satisfies

$$
\boxed{
\begin{aligned}
E\left\{e(t)e^*(s)\right\} &= \sigma^2 \delta_{t,s} \\
E\left\{e(t)e(s)\right\} &= 0, \qquad \text{for all } t \text{ and } s
\end{aligned}
}
\tag{2.4.19}
$$

Note that $\sigma^2 = E\left\{|e(t)|^2\right\}$ is the variance (or power) of $e(t)$. Equation (2.4.19) can be rewritten as

$$
\begin{cases}
E\left\{\operatorname{Re}[e(t)]\operatorname{Re}[e(s)]\right\} = \frac{\sigma^2}{2}\delta_{t,s} \\[2mm]
E\left\{\operatorname{Im}[e(t)]\operatorname{Im}[e(s)]\right\} = \frac{\sigma^2}{2}\delta_{t,s} \\[2mm]
E\left\{\operatorname{Re}[e(t)]\operatorname{Im}[e(s)]\right\} = 0
\end{cases}
\tag{2.4.20}
$$

Hence, the real and imaginary parts of a complex/circular white noise are real-valued white-noise sequences, of identical power equal to $\sigma^2/2$ and uncorrelated with one another. See Appendix B for more details on circular random sequences.

In what follows, we shall also make use of the symbol $\mathcal{O}(1/N^\alpha)$, for some $\alpha > 0$, to denote a random variable that is such that the square root of its second-order moment goes to zero at least as fast as $1/N^\alpha$, as N tends to infinity.

First, we establish the asymptotic variance/covariance of $\hat{\phi}_p(\omega)$ in the case of *Gaussian complex/circular white noise*. The following result holds:

$$
\lim_{N\to\infty} E\left\{[\hat{\phi}_p(\omega_1) - \phi(\omega_1)][\hat{\phi}_p(\omega_2) - \phi(\omega_2)]\right\} =
\begin{cases}
\phi^2(\omega_1), & \omega_1 = \omega_2 \\
0, & \omega_1 \neq \omega_2
\end{cases}
\tag{2.4.21}
$$

Note that, for white noise, $\phi(\omega) = \sigma^2$ (for all ω). Since $\lim_{N\to\infty} E\{\hat{\phi}_p(\omega)\} = \phi(\omega)$ (*cf.* the analysis in the previous subsection), in order to prove (2.4.21) it suffices to show that

$$
\lim_{N\to\infty} E\left\{\hat{\phi}_p(\omega_1)\hat{\phi}_p(\omega_2)\right\} = \phi(\omega_1)\phi(\omega_2) + \phi^2(\omega_1)\delta_{\omega_1,\omega_2}
\tag{2.4.22}
$$

From (2.2.1), we obtain

$$
E\left\{\hat{\phi}_p(\omega_1)\hat{\phi}_p(\omega_2)\right\} = \frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}\sum_{p=1}^{N}\sum_{m=1}^{N} E\left\{e(t)e^*(s)e(p)e^*(m)\right\}
$$
$$
\cdot e^{-i\omega_1(t-s)}e^{-i\omega_2(p-m)}
\tag{2.4.23}
$$

For general random processes, the evaluation of the expectation in (2.4.23) is relatively complicated. However, the following general result for Gaussian random variables can be used: If a, b, c, and d are jointly Gaussian (complex or real) random variables, then

$$
E\{abcd\} = E\{ab\}E\{cd\} + E\{ac\}E\{bd\} + E\{ad\}E\{bc\}
$$
$$
- 2E\{a\}E\{b\}E\{c\}E\{d\}
\tag{2.4.24}
$$

For a proof of (2.4.24), see, for example, [JANSSEN AND STOICA 1988] and references therein. Thus, if the white noise $e(t)$ is Gaussian as assumed, the fourth-order moment in (2.4.23) is found to be

$$
\begin{aligned}
E\left\{e(t)e^*(s)e(p)e^*(m)\right\} &= \left[E\left\{e(t)e^*(s)\right\}\right]\left[E\left\{e(p)e^*(m)\right\}\right] \\
&\quad + \left[E\left\{e(t)e(p)\right\}\right]\left[E\left\{e(s)e(m)\right\}\right]^* \\
&\quad + \left[E\left\{e(t)e^*(m)\right\}\right]\left[E\left\{e^*(s)e(p)\right\}\right] \\
&= \sigma^4(\delta_{t,s}\delta_{p,m} + \delta_{t,m}\delta_{s,p})
\end{aligned}
\tag{2.4.25}
$$

Inserting (2.4.25) in (2.4.23) gives

$$
\begin{aligned}
E\left\{\hat{\phi}_p(\omega_1)\hat{\phi}_p(\omega_2)\right\} &= \sigma^4 + \frac{\sigma^4}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}e^{-i(\omega_1-\omega_2)(t-s)} \\
&= \sigma^4 + \frac{\sigma^4}{N^2}\left|\sum_{t=1}^{N}e^{i(\omega_1-\omega_2)t}\right|^2 \\
&= \sigma^4 + \frac{\sigma^4}{N^2}\left\{\frac{\sin[(\omega_1-\omega_2)N/2]}{\sin[(\omega_1-\omega_2)/2]}\right\}^2
\end{aligned}
\tag{2.4.26}
$$

The limit of the second term in (2.4.26) is σ^4 when $\omega_1 = \omega_2$ and zero otherwise, and (2.4.22) follows at once.

Remark: Note that, in the previous case, it was indeed possible to derive the *finite-sample* variance of $\hat{\phi}_p(\omega)$. For colored noise, the preceding derivation becomes more difficult, and a different approach (presented shortly) is needed. See Exercise 2.13 for yet another approach that applies to general Gaussian signals. ■

Next, we consider the case of a much more general signal obtained by linear filtering of the Gaussian white-noise sequence $\{e(t)\}$ examined earlier. This signal is given by

$$
y(t) = \sum_{k=1}^{\infty}h_k e(t-k)
\tag{2.4.27}
$$

and its PSD is given by

$$
\phi_y(\omega) = |H(\omega)|^2\phi_e(\omega)
\tag{2.4.28}
$$

(*cf.* (1.4.9)). Here, $H(\omega) = \sum_{k=1}^{\infty}h_k e^{-i\omega k}$. The intermediate result that follows, concerned with signals of the foregoing type, appears to have an independent interest. (We will omit the index "p" of $\hat{\phi}_p(\omega)$ in order to simplify the notation.)

For $N \gg 1$,

$$\hat{\phi}_y(\omega) = |H(\omega)|^2 \hat{\phi}_e(\omega) + \mathcal{O}(1/\sqrt{N}) \tag{2.4.29}$$

Hence, the periodograms approximately satisfy an equation of the form of (2.4.28) that is satisfied by the true PSDs.

In order to prove (2.4.29), first observe that

$$
\frac{1}{\sqrt{N}} \sum_{t=1}^{N} y(t) e^{-i\omega t} = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \sum_{k=1}^{\infty} h_k e(t-k) e^{-i\omega(t-k)} e^{-i\omega k}
$$

$$
= \frac{1}{\sqrt{N}} \sum_{k=1}^{\infty} h_k e^{-i\omega k} \sum_{p=1-k}^{N-k} e(p) e^{-i\omega p}
$$

$$
= \frac{1}{\sqrt{N}} \sum_{k=1}^{\infty} h_k e^{-i\omega k}
$$

$$
\cdot \left[\sum_{p=1}^{N} e(p) e^{-i\omega p} + \sum_{p=1-k}^{0} e(p) e^{-i\omega p} - \sum_{p=N-k+1}^{N} e(p) e^{-i\omega p} \right]
$$

$$
\triangleq H(\omega) \left[\frac{1}{\sqrt{N}} \sum_{p=1}^{N} e(p) e^{-i\omega p} \right] + \rho(\omega) \tag{2.4.30}
$$

where

$$
\rho(\omega) = \frac{1}{\sqrt{N}} \sum_{k=1}^{\infty} h_k e^{-i\omega k} \left[\sum_{p=1-k}^{0} e(p) e^{-i\omega p} - \sum_{p=N-k+1}^{N} e(p) e^{-i\omega p} \right]
$$

$$
\triangleq \frac{1}{\sqrt{N}} \sum_{k=1}^{\infty} h_k e^{-i\omega k} \varepsilon_k(\omega) \tag{2.4.31}
$$

Next, note that

$$
E\{\varepsilon_k(\omega)\} = 0,
$$

$$
E\{\varepsilon_k(\omega)\varepsilon_j(\omega)\} = 0 \text{ for all } k \text{ and } j, \text{ and}
$$

$$
E\{\varepsilon_k(\omega)\varepsilon_j^*(\omega)\} = 2\sigma^2 \min(k,j)
$$

which imply

$$
E\{\rho(\omega)\} = 0, \qquad E\{\rho^2(\omega)\} = 0
$$

and

$$E\left\{|\rho(\omega)|^2\right\} = \frac{1}{N}\left|\sum_{k=1}^{\infty}\sum_{j=1}^{\infty} h_k e^{-i\omega k} h_j^* e^{i\omega j} E\left\{\varepsilon_k(\omega)\varepsilon_j^*(\omega)\right\}\right|$$

$$= \frac{2\sigma^2}{N}\left|\sum_{k=1}^{\infty} h_k e^{-i\omega k}\left\{\sum_{j=1}^{k} h_j^* e^{i\omega j} j + \sum_{j=k+1}^{\infty} h_j^* e^{i\omega j} k\right\}\right|$$

$$\leq \frac{2\sigma^2}{N}\sum_{k=1}^{\infty}|h_k|\left\{\sum_{j=1}^{\infty}|h_j|j + \sum_{j=1}^{\infty}|h_j|k\right\}$$

$$= \frac{4\sigma^2}{N}\left(\sum_{k=1}^{\infty}|h_k|\right)\left(\sum_{j=1}^{\infty}|h_j|j\right)$$

If $\sum_{k=1}^{\infty} k|h_k|$ is finite (which is true, for example, if $\{h_k\}$ is exponentially stable; see [SÖDERSTRÖM AND STOICA 1989]), we have

$$E\left\{|\rho(\omega)|^2\right\} \leq \frac{\text{constant}}{N} \tag{2.4.32}$$

Now, from (2.4.30), we obtain

$$\hat{\phi}_y(\omega) = |H(\omega)|^2\hat{\phi}_e(\omega) + \gamma(\omega) \tag{2.4.33}$$

where

$$\gamma(\omega) = H^*(\omega)E^*(\omega)\rho(\omega) + H(\omega)E(\omega)\rho^*(\omega) + \rho(\omega)\rho^*(\omega)$$

and where

$$E(\omega) = \frac{1}{\sqrt{N}}\sum_{t=1}^{N} e(t)e^{-i\omega t}$$

Both $E(\omega)$ and $\rho(\omega)$ are linear combinations of Gaussian random variables, so they are also Gaussian distributed. This means that the fourth-order moment formula (2.4.24) can be used to obtain the second-order moment of $\gamma(\omega)$. By doing so, and also by using (2.4.32) and the fact that, for example, (see (2.4.32)),

$$\left|E\left\{\rho(\omega)E^*(\omega)\right\}\right| \leq \left[E\left\{|\rho(\omega)|^2\right\}\right]^{1/2}\left[E\left\{|E(\omega)|^2\right\}\right]^{1/2}$$

$$= \frac{\text{constant}}{\sqrt{N}} \cdot \left[E\left\{|\hat{\phi}_e(\omega)|^2\right\}\right]^{1/2} = \frac{\text{constant}}{\sqrt{N}}$$

we can verify that $\gamma(\omega) = \mathcal{O}(1/\sqrt{N})$, and hence the proof of (2.4.29) is concluded.

The main result of this section is derived by combining (2.4.21) and (2.4.29):

> The asymptotic variance/covariance result (2.4.21) is also valid for a general linear signal as defined in (2.4.27).

(2.4.34)

Remark: In the introduction to Chapter 1, we mentioned that the analysis of a complex-valued signal is not always more general than the analysis of the corresponding real-valued signal; we supported this claim by the example of a complex sine wave. Here, we have another instance where the claim is valid. Much as in the complex sinusoidal signal case, the complex (or circular) white noise does not specialize, in a direct manner, to real-valued white noise. Indeed, if we would let $e(t)$ in (2.4.19) be real valued, then the two equations in (2.4.19) would conflict with each other (for $t = s$). The *real-valued white noise random process* is a stationary signal that satisfies

$$E\{e(t)e(s)\} = \sigma^2 \delta_{t,s}$$

(2.4.35)

If we try to carry out the proof of (2.4.21) under (2.4.35), then we find that the proof has to be modified. This was expected: both $\phi(\omega)$ and $\hat{\phi}_p(\omega)$ are even functions in the real-valued case; hence, (2.4.21) should be modified to include the case of both $\omega_1 = \omega_2$ and $\omega_1 = -\omega_2$. ∎

It follows from (2.4.34) that, for a fairly general class of signals, the periodogram values are asymptotically (for $N \gg 1$) uncorrelated random variables whose means and standard deviations are both equal to the corresponding true PSD values. Hence, the periodogram is an *inconsistent spectral estimator*, which continues to fluctuate around the true PSD, with a nonzero variance, even if the length of the processed sample increases without bound. Furthermore, the fact that the periodogram values $\hat{\phi}_p(\omega)$ are uncorrelated (for large N values) makes the periodogram exhibit an *erratic behavior* (similar to that of a white-noise realization). These facts constitute the main limitations of the periodogram approach to PSD estimation. In the next sections, we present several modified periodogram-based methods that attempt to cure the aforementioned difficulties of the basic periodogram approach. As we shall see, the "improved methods" decrease the variance of the estimated spectrum at the expense of increasing its bias (and, hence, decreasing the average resolution).

2.5 THE BLACKMAN–TUKEY METHOD

In this section, we develop the Blackman–Tukey method [BLACKMAN AND TUKEY 1959] and compare it to the periodogram. In later sections, we consider several other refined periodogram-based methods that, like the Blackman–Tukey (BT) method, seek to reduce the statistical variability of the estimated spectrum; we will compare these methods with one another and with the Blackman–Tukey method.

2.5.1 The Blackman–Tukey Spectral Estimate

As we have seen, the main problem with the periodogram is the high statistical variability of this spectral estimator, even for very large sample lengths. The poor statistical quality of the periodogram PSD estimator has been intuitively explained as arising from both the poor accuracy of $\hat{r}(k)$ in $\hat{\phi}_c(\omega)$ for extreme lags ($|k| \simeq N$) and the large number of (even if small) covariance estimation errors that are cumulatively summed up in $\hat{\phi}_c(\omega)$. Both these effects may be reduced by truncating the sum in the definition formula of $\hat{\phi}_c(\omega)$, (2.2.2). Following this idea leads to the Blackman–Tukey estimator, which is given by

$$\hat{\phi}_{BT}(\omega) = \sum_{k=-(M-1)}^{M-1} w(k)\hat{r}(k)e^{-i\omega k} \tag{2.5.1}$$

where $\{w(k)\}$ is an even function (i.e., $w(-k) = w(k)$), $w(0) = 1$, $w(k) = 0$ for $|k| \geq M$, and $w(k)$ decays smoothly to zero with k, and where $M < N$. Since $w(k)$ in (2.5.1) weights the lags of the sample covariance sequence, it is called a *lag window*.

If $w(k)$ in (2.5.1) is selected as the rectangular lag window, (i.e., $w(k) = 1$), then we simply obtain a truncated version of $\hat{\phi}_c(\omega)$. However, we may choose $w(k)$ in many other ways, and this flexibility may be employed to improve the accuracy of the Blackman–Tukey spectral estimator or to emphasize some of its characteristics that are of particular interest in a given application. In the next subsections, we address the principal issues that concern the problem of window selection. However, before doing so, we rewrite (2.5.1) in an alternative form that will be used in several places in the discussion that follows.

Let $W(\omega)$ denote the DTFT of $w(k)$,

$$W(\omega) = \sum_{k=-\infty}^{\infty} w(k)e^{-i\omega k} = \sum_{k=-(M-1)}^{M-1} w(k)e^{-i\omega k} \tag{2.5.2}$$

Making use of the DTFT property that led to (2.4.8), we can then write

$$\hat{\phi}_{BT}(\omega) = \sum_{k=-\infty}^{\infty} w(k)\hat{r}(k)e^{-i\omega k}$$

$$= \text{DTFT of the product of the sequences}$$

$$\{\ldots, 0, 0, w(-(M-1)), \ldots, w(M-1), 0, 0, \ldots\} \text{ and}$$

$$\{\ldots, 0, 0, \hat{r}(-(N-1)), \ldots, \hat{r}(N-1), 0, 0, \ldots\}$$

$$= \{\text{DTFT}(\hat{r}(k))\} * \{\text{DTFT}(w(k))\}$$

As DTFT$\{\ldots, 0, 0, \hat{r}(-(N-1)), \ldots, \hat{r}(N-1), 0, 0, \ldots\} = \hat{\phi}_p(\omega)$, we obtain

$$\hat{\phi}_{BT}(\omega) = \hat{\phi}_p(\omega) * W(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\phi}_p(\psi) W(\omega - \psi) d\psi \qquad (2.5.3)$$

This equation is analogous to (2.4.8) and can be interpreted in the same way. Since, for most windows in common use, $W(\omega)$ has a dominant, relatively narrow peak at $\omega = 0$, it follows from (2.5.3) that

> The Blackman–Tukey spectral estimator (2.5.1) corresponds to a "locally" weighted average of the periodogram. $\qquad (2.5.4)$

Since the function $W(\omega)$ in (2.5.3) acts as a window (or weighting) in the frequency domain, it is sometimes called a *spectral window*. As we shall see, several refined periodogram-based spectral estimators discussed in what follows can be given an interpretation similar to that afforded by (2.5.3).

The form (2.5.3) in which the Blackman–Tukey spectral estimator has been cast is quite appealing from an intuitive standpoint. The main problem with the periodogram lies in its large variations about the true PSD. The weighted average in (2.5.3), in the neighborhood of the current frequency point ω, should smooth the periodogram and hence eliminate its large fluctuations.

On the other hand, this smoothing by the spectral window $W(\omega)$ will also have the undesirable effect of reducing the resolution. We may expect that the smaller the M, the larger the reduction in variance and the lower the resolution. These qualitative arguments may be made exact by a statistical analysis of $\hat{\phi}_{BT}(\omega)$, similar to that in the previous section. In fact, it is clear from (2.5.3) that the mean and variance of $\hat{\phi}_{BT}(\omega)$ can be derived from those of $\hat{\phi}_p(\omega)$. Roughly speaking, the results that can be established by the analysis of $\hat{\phi}_{BT}(\omega)$, based on (2.5.3), show that the resolution of this spectral estimator is on the order of $1/M$, whereas its variance is on the order of M/N. The compromise between resolution and variance, which should be considered when choosing the window's length, is clearly seen from the preceding considerations. We will look at the resolution-variance tradeoff in more detail in what follows. The next discussion addresses some of the main issues that concern window design.

2.5.2 Nonnegativeness of the Blackman–Tukey Spectral Estimate

Since $\phi(\omega) \geq 0$, it is natural to also require that $\hat{\phi}_{BT}(\omega) \geq 0$. The lag window can be selected to achieve this desirable property of the estimated spectrum. The following result holds true:

> If the lag window $\{w(k)\}$ is positive semidefinite (i.e., $W(\omega) \geq 0$), then the windowed covariance sequence $\{w(k)\hat{r}(k)\}$ (with $\hat{r}(k)$ given by (2.2.4)) is positive semidefinite, too; this result implies that $\hat{\phi}_{BT}(\omega) \geq 0$ for all ω. $\qquad (2.5.5)$

In order to prove the previous result, first note that $\hat{\phi}_{BT}(\omega) \geq 0$ if and only if the sequence $\{\ldots, 0, 0, w(-(M-1))\hat{r}(-(M-1)), \ldots, w(M-1)\hat{r}(M-1), 0, 0, \ldots\}$ is positive semidefinite or, equivalently, the following Toeplitz matrix is positive semidefinite for all dimensions:

$$
\begin{bmatrix}
w(0)\hat{r}(0) & \ldots & w(M-1)\hat{r}(M-1) & & 0 \\
\vdots & & & \ddots & \\
w(-M+1)\hat{r}(-M+1) & & \ddots & & w(M-1)\hat{r}(M-1) \\
& \ddots & & & \vdots \\
0 & & w(-M+1)\hat{r}(-M+1) & \ldots & w(0)\hat{r}(0)
\end{bmatrix} =
$$

$$
\begin{bmatrix}
w(0) & \ldots & w(M-1) & & 0 \\
\vdots & & & \ddots & \\
w(-M+1) & & \ddots & & w(M-1) \\
& \ddots & & & \vdots \\
0 & & w(-M+1) & \ldots & w(0)
\end{bmatrix} \odot
\begin{bmatrix}
\hat{r}(0) & \ldots & \hat{r}(M-1) & & 0 \\
\vdots & & & \ddots & \\
\hat{r}(-M+1) & & \ddots & & \hat{r}(M-1) \\
& \ddots & & & \vdots \\
0 & & \hat{r}(-M+1) & \ldots & \hat{r}(0)
\end{bmatrix}
$$

The symbol $\odot$ denotes the Hadamard matrix product (i.e., element-wise multiplication). By a result in matrix theory, the Hadamard product of two positive semidefinite matrices is also a positive semidefinite matrix. (See Result R19 in Appendix A.) Thus, the proof of (2.5.5) is concluded.

Another, perhaps simpler, proof of (2.5.5) makes use of (2.5.3) in the following way: Since the sequence $\{w(k)\}$ is real and symmetric about the point $k = 0$, its DTFT $W(\omega)$ is an even, real-valued function. Furthermore, if $\{w(k)\}$ is a positive semidefinite sequence, then $W(\omega) \geq 0$ for all ω values. (See Exercise 1.8.) By (2.5.3), $W(\omega) \geq 0$ immediately implies $\hat{\phi}_{BT}(\omega) \geq 0$, as $\hat{\phi}_p(\omega) \geq 0$ by definition.

On one hand, it should be noted that some lag windows, such as the rectangular window, do not satisfy the assumption made in (2.5.5); hence, their use could lead to estimated spectra that take negative values. The Bartlett window, on the other hand, is positive semidefinite (as can be seen from (2.4.15)).

2.6 WINDOW DESIGN CONSIDERATIONS

The properties of the Blackman–Tukey estimator (and of other refined periodogram methods discussed in the next section) are related directly to the choice of the lag window. In this section, we discuss several relevant properties of windows that are useful in selecting or designing a window to use in a refined spectral estimation procedure.

2.6.1 Time-Bandwidth Product and Resolution-Variance Tradeoffs in Window Design

Most windows are such that they take only nonnegative values in both time and frequency domains (or, if they also take negative values, these are much smaller than the positive values of

the window). In addition, they peak at the origin in both domains. For this type of window, it is possible to define an *equivalent time width*, N_e, and an *equivalent bandwidth*, β_e, as follows:

$$N_e = \frac{\sum_{k=-(M-1)}^{M-1} w(k)}{w(0)} \tag{2.6.1}$$

and

$$\beta_e = \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega)d\omega}{W(0)} \tag{2.6.2}$$

From the definitions of direct and inverse DTFTs, we obtain

$$W(0) = \sum_{k=-\infty}^{\infty} w(k) = \sum_{k=-(M-1)}^{M-1} w(k) \tag{2.6.3}$$

and

$$w(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega)d\omega \tag{2.6.4}$$

Using (2.6.3) and (2.6.4) in (2.6.1) and (2.6.2) gives the following result:

> The (equivalent) time-bandwidth product equals unity:
> $$N_e \beta_e = 1$$
$$\tag{2.6.5}$$

As already indicated, the preceding result applies to window-like signals. Some extended results of the time-bandwidth product type, which apply to more general classes of signals, are presented in Complement 2.8.5.

It is clearly seen from (2.6.5) that a window cannot be both time limited and bandlimited. The more slowly the window decays to zero in one domain, the more concentrated it is in the other domain. The simple result (2.6.5) has several other interesting consequences, as explained next.

The equivalent temporal extent (or aperture), N_e, of $w(k)$ is essentially determined by the window's length. For example, for a rectangular window, we have $N_e \simeq 2M$, whereas for a triangular window, $N_e \simeq M$. This observation, together with (2.6.5), implies that the equivalent bandwidth β_e is basically determined by the window's length. More precisely, $\beta_e = \mathcal{O}(1/M)$. This fact lends support to a claim made previously that, for a window that concentrates most of its energy in its main lobe, the width of that lobe should be on the order of $1/M$. Since the main lobe's width sets a limit on the spectral resolution achievable (as explained in Section 2.4), the preceding observation shows that the spectral resolution limit of a windowed method should be on the order of $1/M$. On the other hand, as explained in the previous section, the statistical variance of such a method is essentially proportional to M/N. Hence, we reach the following conclusion:

> The choice of window's length should be based on a tradeoff between spectral resolution and statistical variance.
$$\tag{2.6.6}$$

As a rule of thumb, we should choose $M \leq N/10$ in order to reduce the standard deviation of the estimated spectrum by at least a factor of three as compared with the periodogram.

Once M is determined, we cannot decrease simultaneously the energy in the main lobe (to reduce smearing) and the energy in the sidelobes (to reduce leakage). This follows, for example, from (2.6.4), which shows that the area of $W(\omega)$ is fixed once $w(0)$ is fixed (such as $w(0) = 1$). In other words, if we want to decrease the main lobe's width then we should accept an increase in the sidelobe energy and vice versa. In summary,

> The selection of window's shape should be based on a tradeoff between smearing and leakage effects.   (2.6.7)

The tradeoff just mentioned is usually dictated by the specific application at hand. A number of windows have been developed to address this tradeoff. In some sense, each of these windows can be seen as a design at a specific point in the resolution–leakage tradeoff curve. We consider several such windows in the next subsection.

2.6.2 Some Common Lag Windows

In this section, we list some of the most common lag windows and outline their relevant properties. Our purpose is not to provide a detailed derivation or an exhaustive listing of such windows, but rather to provide a quick reference of common windows. More detailed information on these and other windows can be found in [HARRIS 1978; KAY 1988; MARPLE 1987; OPPENHEIM AND SCHAFER 1989; PRIESTLEY 1981; PORAT 1997], where many of the closed-form windows have been compiled. Table 2.1 lists some common windows along with some useful properties.

In addition to the fixed-window designs in Table 2.1, there are windows that contain a design parameter that may be varied to trade between resolution and sidelobe leakage. Two such common designs are the Chebyshev window and the Kaiser window. The Chebyshev window has the property that the peak level of the sidelobe "ripples" is constant. Thus, unlike in most other windows, the sidelobe level does not decrease as ω increases. The Kaiser window is defined by

$$w(k) = \frac{I_0\left(\gamma\sqrt{1 - [k/(M-1)]^2}\right)}{I_0(\gamma)}, \quad -(M-1) \leq k \leq M-1 \qquad (2.6.8)$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind. The parameter γ trades the main lobe width for the sidelobe leakage level; $\gamma = 0$ corresponds to a rectangular window, and $\gamma > 0$ results in lower sidelobe leakage at the expense of a broader main lobe. The approximate value of γ needed to achieve a peak sidelobe level of B dB below the peak value is

$$\gamma \simeq \begin{cases} 0, & B < 21 \\ 0.584(B-21)^{0.4} + 0.0789(B-21), & 21 \leq B \leq 50 \\ 0.11(B-8.7), & B > 50 \end{cases}$$

Table 2.1   Some Common Windows and their Properties

These windows satisfy $w(k) \equiv 0$ for $|k| \geq M$, and $w(k) = w(-k)$; the defining equations are valid for $0 \leq k \leq (M-1)$.

Window Name	Defining Equation	Approx. Main Lobe Width (radians)	Sidelobe Level (dB)
Rectangular	$w(k) = 1$	$2\pi/M$	-13
Bartlett	$w(k) = \frac{M-k}{M}$	$4\pi/M$	-25
Hanning	$w(k) = 0.5 + 0.5\cos\left(\frac{\pi k}{M}\right)$	$4\pi/M$	-31
Hamming	$w(k) = 0.54 + 0.46\cos\left(\frac{\pi k}{M-1}\right)$	$4\pi/M$	-41
Blackman	$w(k) = 0.42 + 0.5\cos\left(\frac{\pi k}{M-1}\right)$ $+ 0.08\cos\left(\frac{\pi k}{M-1}\right)$	$6\pi/M$	-57

The Kaiser window is an approximation of the optimal window described in the next subsection. It is often chosen over the fixed-window designs because it has a lower sidelobe level when γ is selected to give the same main lobe width as the corresponding fixed window (or narrower main lobe width for a given sidelobe level). The optimal window of the next subsection improves on the Kaiser design slightly.

Figure 2.3 shows plots of several windows with $M = 26$. The Kaiser window is shown for $\gamma = 1$ and $\gamma = 4$, and the Chebyshev window is designed to have a -40 dB sidelobe level. Figure 2.4 shows the corresponding normalized window transfer functions $W(\omega)$. Note the constant sidelobe ripple level of the Chebyshev design.

We remark that, except for the Bartlett window, none of the windows we have introduced (including the Chebyshev and Kaiser windows) has nonnegative Fourier transform. On the other hand, it is straightforward to produce such a nonnegative definite window by convolving the window with itself. Recall that the Bartlett window is the convolution of a rectangular window with itself. We will make use of the convolution of windows with themselves in the next two subsections, both for window design and for relating temporal windows to covariance lag windows.

2.6.3 Window Design Example

Assume a situation where the observed signal consists of a weak desired signal and a strong interference and where both the desired signal and the interference are narrowband signals that are well separated in frequency. However, there is no *a priori* quantitative information available on the frequency separation between the desired signal and the interference. It is required to design a lag window for use in a Blackman–Tukey spectral estimation method, with the purpose of detecting and locating in frequency the useful signal.

The main problem in the application just outlined lies in the fact that the (strong) interference might completely mask the (weak) desired signal, through leakage. In order to get rid of this
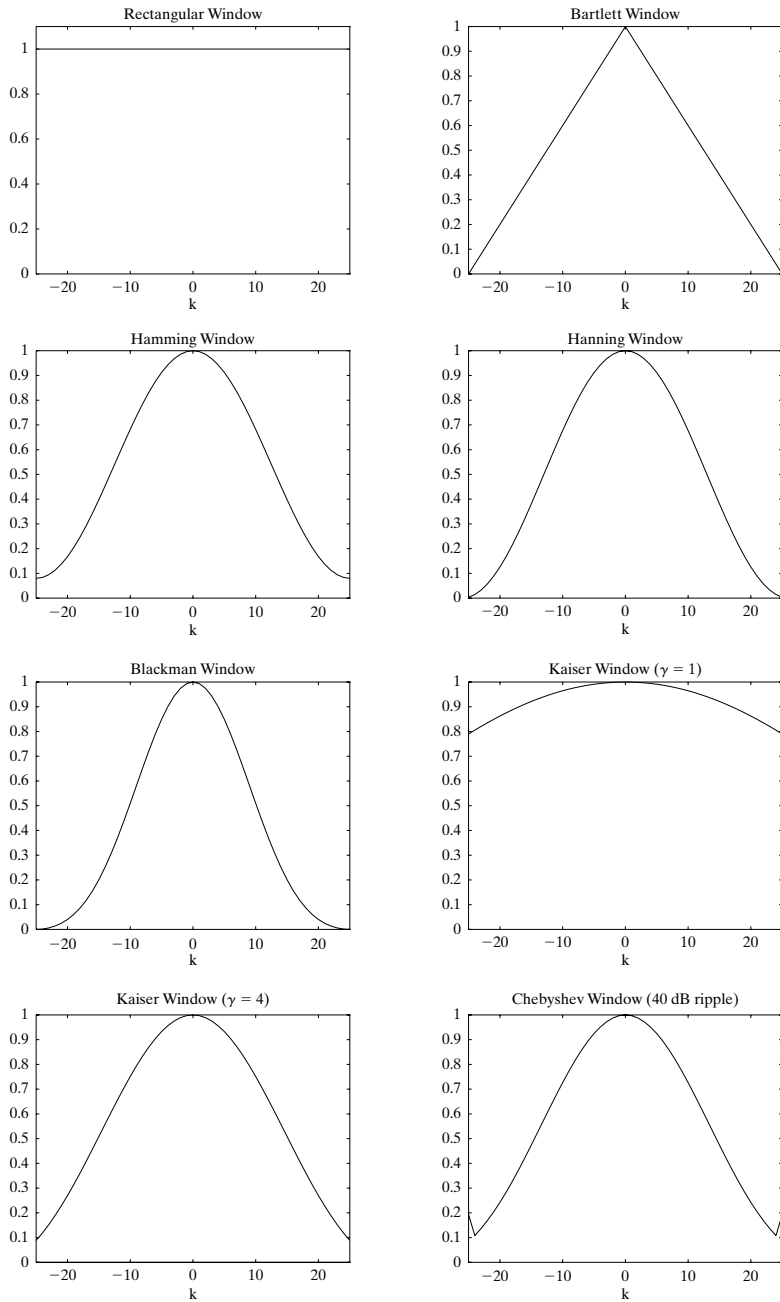
Figure 2.3 Some common window functions (shown for $M = 26$). The Kaiser window uses $\gamma = 1$ and $\gamma = 4$, and the Chebyshev window is designed for a -40 dB sidelobe level.
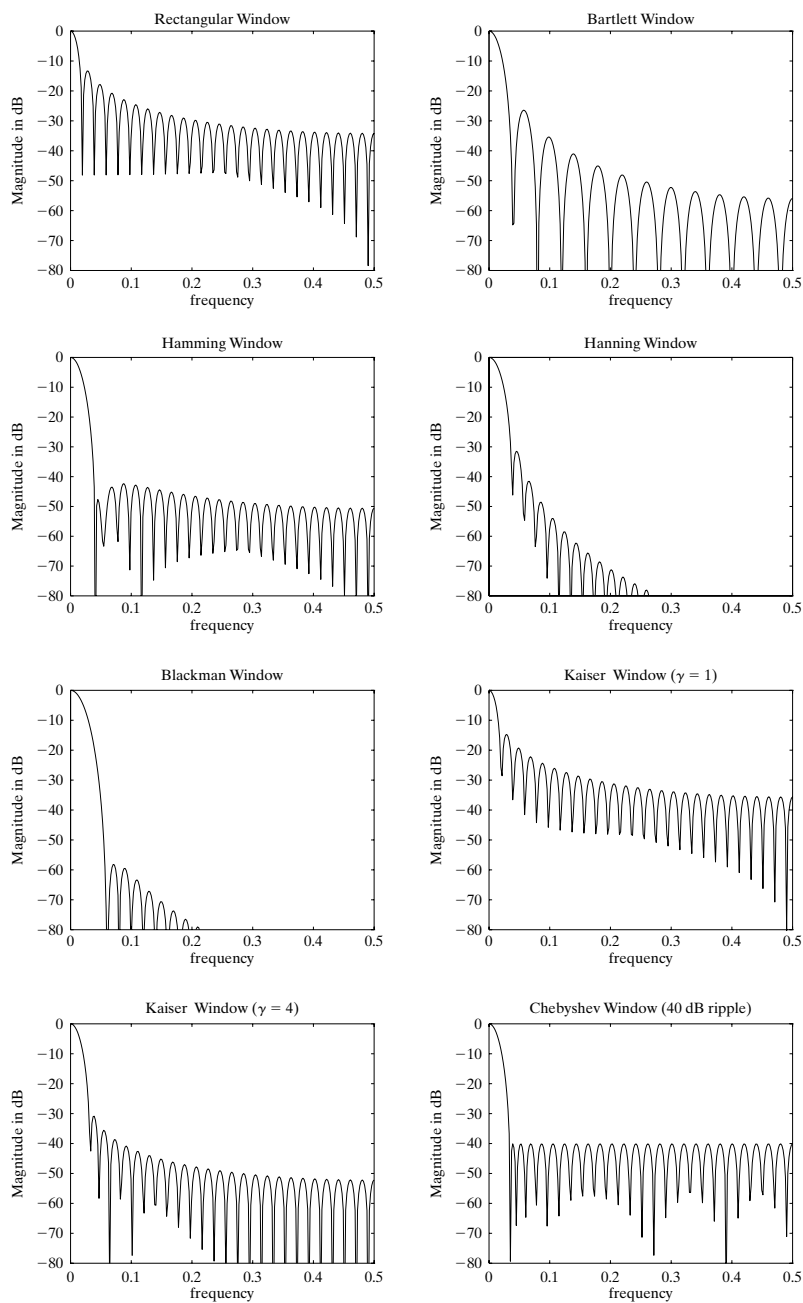
Figure 2.4   The DTFTs of the window functions in Figure 2.3.

problem, the window design should compromise smearing for leakage. Note that the smearing effect is not of main concern in this application, because the useful signal and the interference are well separated in frequency. Hence, smearing will not affect our ability to *detect* the desired signal although it will limit, to some degree, our ability to accurately *locate in frequency* the signal in question.

We consider a window sequence whose DTFT $W(\omega)$ is constructed as the squared magnitude of the DTFT of another sequence $\{v(k)\}$; in this way, we guarantee that the constructed window is positive semidefinite. Mathematically, the previous design problem can be formulated as follows: Consider a sequence $\{v(0), \ldots, v(M-1)\}$, and let

$$V(\omega) = \sum_{k=0}^{M-1} v(k)e^{-i\omega k} \tag{2.6.9}$$

The DTFT $V(\omega)$ can be rewritten in the more compact form

$$V(\omega) = v^* a(\omega) \tag{2.6.10}$$

where

$$v = [v(0) \ \ldots \ v(M-1)]^* \tag{2.6.11}$$

and

$$a(\omega) = [1 \ e^{-i\omega} \ldots \ e^{-i(M-1)\omega}]^T \tag{2.6.12}$$

Define the spectral window as

$$W(\omega) = |V(\omega)|^2 \tag{2.6.13}$$

The corresponding lag window can be obtained from (2.6.13) as follows:

$$\sum_{k=-(M-1)}^{M-1} w(k)e^{-i\omega k} = \sum_{n=0}^{M-1} \sum_{p=0}^{M-1} v(n)v^*(p)e^{-i\omega(n-p)}$$

$$= \sum_{n=0}^{M-1} \sum_{k=n}^{n-(M-1)} v(n)v^*(n-k)e^{-i\omega k}$$

$$= \sum_{k=-(M-1)}^{M-1} \left[\sum_{n=0}^{M-1} v(n)v^*(n-k) \right] e^{-i\omega k} \tag{2.6.14}$$

This gives

$$w(k) = \sum_{n=0}^{M-1} v(n)v^*(n-k) \tag{2.6.15}$$

The last equality in (2.6.14), and hence the equality (2.6.15), are valid under the convention that $v(k) = 0$ for $k < 0$ and $k \geq M$.

As already mentioned, this method of constructing $\{w(k)\}$ from the convolution of the sequence $\{v(k)\}$ with itself has the advantage that the lag window so obtained is always positive semidefinite or, equivalently, the corresponding spectral window satisfies $W(\omega) \geq 0$ (as is easily seen from (2.6.13)). Besides this, the design of $\{w(k)\}$ can be reduced to the selection of $\{v(k)\}$, which may be more conveniently done, as explained next.

In the present application, the design objective is to reduce the leakage incurred by $\{w(k)\}$ as much as possible. This objective can be formulated as the problem of minimizing the relative energy in the sidelobes of $W(\omega)$ or, equivalently, as the problem of maximizing the relative energy in the main lobe of $W(\omega)$:

$$\max_{v} \left\{ \frac{\int_{-\beta\pi}^{\beta\pi} W(\omega)d\omega}{\int_{-\pi}^{\pi} W(\omega)d\omega} \right\} \tag{2.6.16}$$

Here, β is a design parameter that quantifies how much smearing (or resolution) we can trade off for leakage reduction. The larger the β is, the more leakage free is the optimal window derived from (2.6.16), but also the more diminished is the spectral resolution associated with that window.

By writing the criterion in (2.6.16) in the form

$$\frac{\frac{1}{2\pi} \int_{-\beta\pi}^{\beta\pi} |V(\omega)|^2 d\omega}{\frac{1}{2\pi} \int_{-\pi}^{\pi} |V(\omega)|^2 d\omega} = \frac{v^* \left[\frac{1}{2\pi} \int_{-\beta\pi}^{\beta\pi} a(\omega)a^*(\omega)d\omega \right] v}{v^* v} \tag{2.6.17}$$

(*cf.* (2.6.10) and Parseval's theorem, (1.2.6)), the optimization problem (2.6.16) becomes

$$\max_{v} \frac{v^* \Gamma v}{v^* v} \tag{2.6.18}$$

where

$$\Gamma = \frac{1}{2\pi} \int_{-\beta\pi}^{\beta\pi} a(\omega)a^*(\omega)d\omega \triangleq [\gamma_{m-n}] \tag{2.6.19}$$

and where

$$\gamma_{m-n} = \frac{1}{2\pi} \int_{-\beta\pi}^{\beta\pi} e^{-i(m-n)\omega} d\omega = \frac{\sin[(m-n)\beta\pi]}{(m-n)\pi} \tag{2.6.20}$$

(note that $\gamma_0 = \beta$). By using the function

$$\text{sinc}(x) \triangleq \frac{\sin x}{x}, \qquad (\text{sinc}(0) = 1) \tag{2.6.21}$$

we can write (2.6.20) as

$$\gamma_{m-n} = \beta \mathrm{sinc}[(m-n)\beta\pi] \qquad (2.6.22)$$

The solution to the problem (2.6.18) is well known: the maximizing v is given by the dominant eigenvector of Γ, associated with the maximum eigenvalue of this matrix. (See Result R13 in Appendix A.) To summarize,

> The optimal lag window that minimizes the relative energy in the sidelobe interval $[-\pi, -\beta\pi] \cup [\beta\pi, \pi]$ is given by (2.6.15), where v is the dominant eigenvector of the matrix Γ defined in (2.6.19) and (2.6.22).   (2.6.23)

Regarding the choice of the design parameter β, it is clear that β should be larger than $1/M$ in order to allow for a significant reduction of leakage. Otherwise, by selecting, for example, $\beta \simeq 1/M$, we weight the resolution issue too much in the design problem, with unfavorable consequences for leakage reduction.

Finally, we remark that a problem quite similar to the previous problem, although derived from different considerations, will be encountered in Chapter 5. (See also [MULLIS AND SCHARF 1991].)

2.6.4 Temporal Windows and Lag Windows

As we have previously seen, the unwindowed periodogram coincides with the unwindowed correlogram. The Blackman–Tukey estimator is a windowed correlogram obtained by using a lag window. Similarly, we can define a windowed periodogram

$$\hat{\phi}_W(\omega) = \frac{1}{N} \left| \sum_{t=1}^{N} v(t)y(t)e^{-i\omega t} \right|^2 \qquad (2.6.24)$$

where the weighting sequence $\{v(t)\}$ may be called a *temporal window*. A temporal window is sometimes called a *taper*. Welch [WELCH 1967] was one of the first researchers who considered windowed periodogram spectral estimators (see Section 2.7.2 for a description of Welch's method); hence, the subscript "W" is attached to $\hat{\phi}(\omega)$ in (2.6.24). However, while the reason for windowing the correlogram is clearly motivated, the reason for windowing the periodogram is less obvious. In order to motivate (2.6.24), at least partially, write it as

$$\hat{\phi}_W(\omega) = \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} v(t)v^*(s)y(t)y^*(s)e^{-i\omega(t-s)} \qquad (2.6.25)$$

Next, take expectation of both sides of (2.6.25), to obtain

$$E\left\{ \hat{\phi}_W(\omega) \right\} = \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} v(t)v^*(s)r(t-s)e^{-i\omega(t-s)} \qquad (2.6.26)$$

Inserting

$$r(t - s) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(\omega) e^{i\omega(t-s)} d\omega \qquad (2.6.27)$$

in (2.6.26) gives

$$E\left\{\hat{\phi}_W(\omega)\right\} = \frac{1}{N2\pi} \int_{-\pi}^{\pi} \phi(\psi) \left[\sum_{t=1}^{N} \sum_{s=1}^{N} v(t) v^*(s) e^{-i(\omega-\psi)(t-s)} \right] d\psi$$

$$= \frac{1}{N2\pi} \int_{-\pi}^{\pi} \phi(\psi) \left| \sum_{t=1}^{N} v(t) e^{-i(\omega-\psi)t} \right|^2 d\psi \qquad (2.6.28)$$

Define

$$W(\omega) = \frac{1}{N} \left| \sum_{t=1}^{N} v(t) e^{-i\omega t} \right|^2 \qquad (2.6.29)$$

By using this notation, we can write (2.6.28) as

$$E\left\{\hat{\phi}_W(\omega)\right\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(\psi) W(\omega - \psi) d\psi \qquad (2.6.30)$$

As the equation (2.6.29) is similar to (2.6.13), the sequence whose DTFT is equal to $W(\omega)$ immediately follows from (2.6.15):

$$w(k) = \frac{1}{N} \sum_{n=1}^{N} v(n) v^*(n - k) \qquad (2.6.31)$$

Next, by comparing (2.6.30) and (2.5.3), we get the following result:

> The windowed periodogram and the windowed correlogram have the same *average* behavior, provided that the temporal and lag windows are related as in (2.6.31).   (2.6.32)

Hence, $E\{\hat{\phi}_W(\omega)\} = E\{\hat{\phi}_{BT}(\omega)\}$, provided that the temporal and lag windows are matched to one another. However, a similarly simple relationship between $\hat{\phi}_W(\omega)$ and $\hat{\phi}_{BT}(\omega)$ does not seem to exist. This makes it somewhat difficult to motivate the windowed periodogram as defined in (2.6.24). The Welch periodogram, though, does not weight all data samples as in (2.6.24) and is a useful spectral estimator (as is shown in the next section).

2.7 OTHER REFINED PERIODOGRAM METHODS

In Section 2.5, we introduced the Blackman–Tukey estimator as an alternative to the periodogram. In this section, we present three other modified periodograms: the Bartlett, Welch, and Daniell methods. Like the Blackman–Tukey method, they seek to reduce the variance of the periodogram by smoothing or averaging the periodogram estimates in some way. We will relate these methods to one another and to the Blackman–Tukey method.

2.7.1 Bartlett Method

The basic idea of the Bartlett method [BARTLETT 1948; BARTLETT 1950] is simple: to reduce the large fluctuations of the periodogram, split up the available sample of N observations into $L = N/M$ subsamples of M observations each, and then average the periodograms obtained from the subsamples for each value of ω. Mathematically, the Bartlett method can be described as follows: Let

$$y_j(t) = y((j-1)M + t), \qquad \begin{aligned} t &= 1, \ldots, M \\ j &= 1, \ldots, L \end{aligned} \qquad (2.7.1)$$

denote the observations of the jth subsample, and let

$$\hat{\phi}_j(\omega) = \frac{1}{M} \left| \sum_{t=1}^{M} y_j(t) e^{-i\omega t} \right|^2 \qquad (2.7.2)$$

denote the corresponding periodogram. The Bartlett spectral estimate is then given by

$$\hat{\phi}_B(\omega) = \frac{1}{L} \sum_{j=1}^{L} \hat{\phi}_j(\omega) \qquad (2.7.3)$$

Since the Bartlett method operates on data segments of length M, the resolution afforded should be on the order of $1/M$. Hence, the spectral resolution of the Bartlett method is reduced by a factor L, compared with the resolution of the original periodogram method. In return for this reduction in resolution, we can expect the Bartlett method to have a reduced variance. We show below that the Bartlett method reduces the variance of the periodogram by the same factor L. The compromise between resolution and variance when selecting M (or L) is thus evident.

An interesting way to look at the Bartlett method and its properties is by relating it to the Blackman–Tukey method. As we know, $\hat{\phi}_j(\omega)$ of (2.7.2) can be rewritten as

$$\hat{\phi}_j(\omega) = \sum_{k=-(M-1)}^{M-1} \hat{r}_j(k) e^{-i\omega k} \qquad (2.7.4)$$

where $\{\hat{r}_j(k)\}$ is the sample covariance sequence corresponding to the jth subsample. Inserting (2.7.4) in (2.7.3) gives

$$\hat{\phi}_B(\omega) = \sum_{k=-(M-1)}^{M-1} \left[\frac{1}{L} \sum_{j=1}^{L} \hat{r}_j(k) \right] e^{-i\omega k} = \sum_{k=-(M-1)}^{M-1} \hat{r}_B(k) e^{-i\omega k} \qquad (2.7.5)$$

We see that $\hat{\phi}_B(\omega)$ is similar in form to the Blackman–Tukey estimator that uses a rectangular window. Here, $\hat{r}_B(k)$ is an estimate of the ACS $r(k)$. However, $\hat{r}_B(k)$ in (2.7.5) does not make efficient use of the available lag products $y(t)y^*(t-k)$, especially for $|k|$ near $M-1$. (See Exercise 2.14.) In fact, for $k = M-1$, only about $1/M$th of the available lag products are used to form the ACS estimate in (2.7.5). We expect that the variance of $\hat{r}_B(k)$ will be higher than that of the corresponding $\hat{r}(k)$ lags used in the Blackman–Tukey estimate and, similarly, that the variance of $\hat{\phi}_B(\omega)$ will be higher than that of $\hat{\phi}_{BT}(\omega)$. In addition, the Bartlett method uses a fixed rectangular lag window and thus provides less flexibility in resolution–leakage tradeoff than does the Blackman–Tukey method. For these reasons, we conclude the following:

> The Bartlett estimate, as defined in (2.7.1)–(2.7.3), is similar in form to, but typically has a slightly higher variance than, the Blackman–Tukey estimate with a rectangular lag window of length M. $\qquad$ (2.7.6)

The reduction in resolution and the decrease of variance (both by a factor $L = N/M$) for the Bartlett estimate, as compared to the basic periodogram method, follow from (2.7.6) and from the properties of the Blackman–Tukey spectral estimator given previously.

The main lobe of the rectangular window is narrower than that associated with most other lag windows—this follows from the observation that the rectangular window clearly has the largest equivalent time width and from the fact that the time-bandwidth product is constant; see (2.6.5). Thus, it follows from (2.7.6) that, in the class of Blackman–Tukey estimates, the Bartlett estimator can be expected to have the least smearing (and hence the best resolution), but the most significant leakage.

2.7.2 Welch Method

The Welch method [WELCH 1967] is obtained by refining the Bartlett method in two respects. First, the data segments in the Welch method are allowed to overlap. Second, each data segment is windowed prior to computation of the periodogram. To describe the Welch method in a mathematical form, let

$$y_j(t) = y((j-1)K + t), \qquad \begin{array}{l} t = 1, \dots, M \\ j = 1, \dots, S \end{array} \qquad (2.7.7)$$

denote the jth data segment. In (2.7.7), $(j-1)K$ is the starting point for the jth sequence of observations. If $K = M$, then the sequences do not overlap (but are contiguous), and we get the sample splitting used by the Bartlett method (which leads to $S = L = N/M$ data subsamples).

However, the value recommended for K in the Welch method is $K = M/2$, in which case $S \simeq 2M/N$ data segments (with 50% overlap between successive segments) are obtained.

The windowed periodogram corresponding to $y_j(t)$ is computed as

$$\hat{\phi}_j(\omega) = \frac{1}{MP}\left|\sum_{t=1}^{M} v(t)y_j(t)e^{-i\omega t}\right|^2 \tag{2.7.8}$$

where P denotes the "power" of the temporal window $\{v(t)\}$:

$$P = \frac{1}{M}\sum_{t=1}^{M}|v(t)|^2 \tag{2.7.9}$$

The Welch estimate is found by averaging the windowed periodograms in (2.7.8):

$$\hat{\phi}_W(\omega) = \frac{1}{S}\sum_{j=1}^{S}\hat{\phi}_j(\omega) \tag{2.7.10}$$

The reasons for these modifications to the Bartlett method, which led to the Welch method, are simple to explain. By allowing overlap between the data segments and hence getting more periodograms to be averaged in (2.7.10), we hope to decrease the variance of the estimated PSD. By introducing the window in the periodogram computation, we hope to get more control over the bias/resolution properties of the estimated PSD (see Section 2.6.4). Additionally, the temporal window may be used to give less weight to the data samples at the ends of each subsample, hence, making the consecutive subsample sequences less correlated to one another, even though they are overlapping. The principal effect of this "decorrelation" should be a more effective reduction of variance via the averaging in (2.7.10).

The analysis that led to the results (2.6.30)–(2.6.32) can be modified to show that the use of windowed periodograms in the Welch method, as contrasted to the unwindowed periodograms in the Bartlett method, indeed offers more flexibility in controlling the bias properties of the estimated spectrum. The variance of the Welch spectral estimator is more difficult to analyze (except in some special cases). However, there is empirical evidence that the Welch method can offer lower variance than the Bartlett method, but the difference in the variances corresponding to the two methods is not dramatic.

We can relate the Welch estimator to the Blackman–Tukey spectral estimator by a straightforward calculation, as we did for the Bartlett method. By inserting (2.7.8) in (2.7.10), we obtain

$$\hat{\phi}_W(\omega) = \frac{1}{S}\sum_{j=1}^{S}\frac{1}{MP}\sum_{t=1}^{M}\sum_{k=1}^{M}v(t)v^*(k)y_j(t)y_j^*(k)e^{-i\omega(t-k)}$$

$$= \frac{1}{MP}\sum_{t=1}^{M}\sum_{k=1}^{M}v(t)v^*(k)\underbrace{\left[\frac{1}{S}\sum_{j=1}^{S}y_j(t)y_j^*(k)\right]}_{\overset{\Delta}{=}\tilde{r}(t,k)}e^{-i\omega(t-k)} \tag{2.7.11}$$

For large values of N and for $K \leq M/2$, S turns out to be sufficiently large for $\tilde{r}(t, k)$ to be a reasonable estimate of the covariance $r(t - k)$. We assume that $\tilde{r}(t, k)$ depends only on the difference $(t - k)$, at least approximately:

$$\tilde{r}(t, k) \simeq \tilde{r}(t - k) \tag{2.7.12}$$

Using (2.7.12) in (2.7.11) gives

$$\hat{\phi}_W(\omega) \simeq \frac{1}{MP} \sum_{t=1}^{M} \sum_{k=1}^{M} v(t)v^*(k)\tilde{r}(t - k)e^{-i\omega(t-k)}$$

$$= \frac{1}{MP} \sum_{t=1}^{M} \sum_{\tau=t-1}^{t-M} v(t)v^*(t - \tau)\tilde{r}(\tau)e^{-i\omega\tau}$$

$$= \sum_{\tau=-(M-1)}^{M-1} \left[\frac{1}{MP} \sum_{t=1}^{M} v(t)v^*(t - \tau) \right] \tilde{r}(\tau)e^{-i\omega\tau} \tag{2.7.13}$$

By introducing

$$w(\tau) = \frac{1}{MP} \sum_{t=1}^{M} v(t)v^*(t - \tau) \tag{2.7.14}$$

(under the convention that $v(k) = 0$ for $k < 1$ and $k > M$), we can write (2.7.13) as

$$\hat{\phi}_W(\omega) \simeq \sum_{\tau=-(M-1)}^{M-1} w(\tau)\tilde{r}(\tau)e^{-i\omega\tau} \tag{2.7.15}$$

which is to be compared with the form of the Blackman–Tukey estimator. To summarize, the Welch estimator has been shown to approximate a Blackman–Tukey-*type* estimator for the estimated covariance sequence (2.7.12) (which may be expected to have finite-sample properties different from those of $\hat{r}(k)$).

The Welch estimator can be efficiently computed via the FFT and is one of the most frequently used PSD estimation methods. Its previous interpretation is pleasing, even though approximate, because the Blackman–Tukey form of spectral estimator is theoretically the most favored one. This interpretation also shows that we may think of replacing the usual covariance estimates $\{\hat{r}(k)\}$ in the Blackman–Tukey estimator by other sample covariances, with the purpose of either reducing the computational burden or improving the statistical accuracy.

2.7.3 Daniell Method

As shown in (2.4.21), the periodogram values $\hat{\phi}(\omega_k)$ corresponding to different frequency values ω_k are (asymptotically) uncorrelated random variables. One may then think of reducing the large variance of the basic periodogram estimator by averaging the periodogram over small intervals

centered on the current frequency ω. This is the idea behind the Daniell method [DANIELL 1946]. The practical form of the Daniell estimate, which can be implemented by means of the FFT, is

$$\hat{\phi}_D(\omega_k) = \frac{1}{2J+1} \sum_{j=k-J}^{k+J} \hat{\phi}_p(\omega_j) \tag{2.7.16}$$

where

$$\omega_k = \frac{2\pi}{\tilde{N}} k, \qquad k = 0, \ldots, \tilde{N} - 1 \tag{2.7.17}$$

and where $\tilde{N}$ is (much) larger than N to ensure a fine sampling of $\hat{\phi}_p(\omega)$. The periodogram samples needed in (2.7.16) can be obtained, for example, by using a radix-2 FFT algorithm applied to the zero-padded data sequence, as described in Section 2.3. The parameter J in the Daniell method should be chosen sufficiently small to guarantee that $\phi(\omega)$ is nearly constant on the interval(s):

$$\left[\omega - \frac{2\pi}{\tilde{N}} J, \ \omega + \frac{2\pi}{\tilde{N}} J \right] \tag{2.7.18}$$

$\tilde{N}$ can, in principle, be chosen as large as we want, so we can choose J to be fairly large without violating the requirement that $\phi(\omega)$ be nearly constant over the interval in (2.7.18). For the sake of illustration, let us assume that we keep the ratio $J/\tilde{N}$ constant, but increase both J and $\tilde{N}$ significantly. As $J/\tilde{N}$ is constant, the resolution/bias properties of the Daniell estimator should be basically unaffected. On the other hand, the fact that the number of periodogram values averaged in (2.7.16) increases with increased J might suggest that the variance decreases. However, we know that this should not be possible, because the variance can be decreased only at the expense of increasing the bias (and vice versa). Indeed, in the case under discussion, the periodogram values averaged in (2.7.16) become more and more correlated as $\tilde{N}$ increases; hence, the variance of $\hat{\phi}_D(\omega)$ does not necessarily decrease with J if $\tilde{N}$ is larger than N. (See, for example, Exercise 2.13.) We will return to the bias and variance properties of the Daniell method a bit later.

By introducing $\beta = 2J/\tilde{N}$, one can write (2.7.18) in a form that is more convenient for the discussion that follows, namely

$$[\omega - \pi\beta, \omega + \pi\beta] \tag{2.7.19}$$

Equation (2.7.16) is a discrete approximation of the continuous version of the Daniell estimator, which is given by

$$\hat{\phi}_D(\omega) = \frac{1}{2\pi\beta} \int_{\omega-\beta\pi}^{\omega+\beta\pi} \hat{\phi}_p(\psi) d\psi \tag{2.7.20}$$

The larger the $\tilde{N}$, the smaller the difference between the approximation (2.7.16) and the continuous version (2.7.20) of the Daniell spectral estimator.

It is intuitively clear from (2.7.20) that, as β increases, the resolution of the Daniell estimator decreases (so the bias increases) and the variance decreases. In fact, if we introduce

$$M = 1/\beta \tag{2.7.21}$$

(in an approximate sense, as $1/\beta$ is not necessarily an integer) then we may expect that the resolution and the variance of the Daniell estimator are both decreased by a factor M, compared with the basic periodogram method. In order to support this claim, we relate the Daniell estimator to the Blackman–Tukey estimation technique. By comparing (2.7.20) and (2.5.3), we obtain the following result:

> The Daniell estimator is a particular case of the Blackman–Tukey class of spectral estimators, one corresponding to a rectangular *spectral* window:
> $$W(\omega) = \begin{cases} 1/\beta, & \omega \in [-\beta\pi, \beta\pi] \\ 0, & \text{otherwise} \end{cases}$$
(2.7.22)

This observation, along with the time-bandwidth product result and the properties of the Blackman–Tukey spectral estimator, lends support to the claim previously made for the Daniell estimator. Note that the Daniell estimate of PSD is a nonnegative function by its very definition, (2.7.20); such is not necessarily the case for several members of the Blackman–Tukey class of PSD estimators.

The lag window corresponding to the $W(\omega)$ in (2.7.22) is readily evaluated as follows:

$$w(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega)e^{i\omega k}d\omega = \frac{1}{2\pi\beta} \int_{-\pi\beta}^{\pi\beta} e^{i\omega k}d\omega$$

$$= \frac{\sin(k\pi\beta)}{k\pi\beta} = \text{sinc}(k\pi\beta) \tag{2.7.23}$$

Note that $w(k)$ does not vanish as k increases; this effect leads to a subtle (but not essential) difference between the lag-windowed forms of the Daniell and Blackman–Tukey estimators. Since the inverse DTFT of $\hat{\phi}_p(\omega)$ is given by the sequence $\{\ldots, 0, 0, \hat{r}(-(N-1)), \ldots, \hat{r}(N-1), 0, 0, \ldots\}$, it follows immediately from (2.7.20) that $\hat{\phi}_D(\omega)$ can also be written as

$$\hat{\phi}_D(\omega) = \sum_{k=-(N-1)}^{N-1} w(k)\hat{r}(k)e^{-i\omega k} \tag{2.7.24}$$

It is seen from (2.7.24) that, like the Blackman–Tukey estimator, $\hat{\phi}_D(\omega)$ is a windowed version of the correlogram but, unlike the Blackman–Tukey estimator, the sum in (2.7.24) is not truncated to a value $M < N$. Hence, contrary to what might have been expected intuitively, the parameter

M defined in (2.7.21) cannot be exactly interpreted as a "truncation point" for the lag-windowed version of $\hat{\phi}_D(\omega)$. However, the equivalent bandwidth of $W(\omega)$ is clearly equal to β:

$$\beta_e = \beta$$

Thus, it follows that the equivalent time width of $w(k)$ is

$$N_e = 1/\beta_e = M$$

which shows that M plays essentially the same role here as the "truncation point" in the Blackman–Tukey estimator (and, indeed, it can be verified that $w(k)$ in (2.7.23) takes small values for $|k| > M$).

In closing this section and this chapter, we point out that the periodogram-based methods for spectrum estimation are all variations on the same theme. These methods attempt to reduce the variance of the basic periodogram estimator, at the expense of some reduction in resolution, by various means: averaging periodograms derived from data subsamples (Bartlett and Welch methods); averaging periodogram values locally around the frequency of interest (Daniell method); and smoothing the periodogram (Blackman–Tukey method). The unifying theme of these methods is seen in that they are essentially special forms of the Blackman–Tukey approach. In Chapter 5, we will push the unifying theme one step further, by showing that the periodogram-based methods can also be obtained as special cases of the filter-bank approach to spectrum estimation described there. (See also [Mullis and Scharf 1991].)

Finally, it is interesting to note that, although the modifications of the periodogram described in this chapter are indeed required when estimating a continuous PSD, the *unmodified periodogram* can be shown to be a satisfactory estimator (actually, the best one in large samples) for discrete (or line) spectra corresponding to sinusoidal signals. This is shown in Chapter 4.

2.8 COMPLEMENTS

2.8.1 Sample Covariance Computation via FFT

Computation of the sample covariances is a ubiquitous problem in spectral estimation and in signal processing applications. In this complement, we make use of the DTFT-like formula (2.2.2), relating the periodogram and the sample covariance sequence, to devise an FFT-based algorithm for computation of the $\{\hat{r}(k)\}_{k=0}^{N-1}$. We also compare the computational requirements of such an algorithm with those corresponding to the evaluation of $\{\hat{r}(k)\}$ via the temporal averaging formula (2.2.4), and we show that the former could be computationally more efficient than the latter if N is larger than a certain value.

From (2.2.2) and (2.2.6), we have (omitting the subscript p of $\hat{\phi}_p(\omega)$ for notational simplicity)

$$\hat{\phi}(\omega) = \sum_{k=-N+1}^{N-1} \hat{r}(k)e^{-i\omega k} = \sum_{p=1}^{2N-1} \hat{r}(p-N)e^{-i\omega(p-N)}$$

or, equivalently,

$$e^{-i\omega(N-1)}\,\hat{\phi}(\omega) = \sum_{p=1}^{2N-1} \rho(p)e^{-i\omega(p-1)} \tag{2.8.1}$$

where $\rho(p) \triangleq \hat{r}(p-N)$. Equation (2.8.1) has the standard form of a DFT (see (2.3.3)). It is evident from (2.8.1) that, in order to determine the sample covariance sequence, we need at least $(2N-1)$ values of the periodogram. This is expected; the sequence $\{\hat{r}(k)\}_{k=0}^{N-1}$ contains $(2N-1)$ real-valued unknowns, for the computation of which at least $(2N-1)$ periodogram values should be necessary (because $\hat{\phi}(\omega)$ is real valued).

Let

$$\omega_k = \frac{2\pi}{2N-1}\,(k-1), \qquad k = 1, \dots, 2N-1$$

Also, let the sequence $\{y(t)\}_{t=1}^{2N-1}$ be obtained by padding the raw data sequence with $(N-1)$ zeroes. Compute

$$Y_k = \sum_{t=1}^{2N-1} y(t)e^{-i\omega_k(t-1)} \qquad (k = 1, 2, \dots, 2N-1) \tag{2.8.2}$$

by means of a $(2N-1)$-point FFT algorithm. Next, evaluate

$$\tilde{\phi}_k = e^{-i\omega_k(N-1)}\,|Y_k|^2/N \qquad (k = 1, \dots, 2N-1) \tag{2.8.3}$$

Finally, calculate the sample covariances via the "inversion" of (2.8.1):

$$\rho(p) = \sum_{k=1}^{2N-1} \tilde{\phi}_k e^{i\omega_k(p-1)}/(2N-1)$$

$$= \sum_{k=1}^{2N-1} \tilde{\phi}_k e^{i\omega_p(p-1)}/(2N-1) \tag{2.8.4}$$

The previous computation may once again be done by using a $(2N-1)$-point FFT algorithm. The bulk of the procedure just outlined consists of the FFT-based computation of (2.8.2) and (2.8.4). That computation requires about $2N\log_2(2N)$ flops (assuming that the radix-2 FFT algorithm is used; the required number of operations is larger than the one previously given whenever N is not a power of two). The direct evaluation of the sample covariance sequence via (2.2.4) requires

$$N + (N-1) + \cdots + 1 \simeq N^2/2 \qquad \text{flops}$$

Hence, the FFT-based computation would be more efficient whenever

$$N > 4 \log_2(2N)$$

This inequality is satisfied for $N \geq 32$. (Actually, N needs to be greater than 32, because we neglected the operations needed to implement equation (2.8.3).)

The previous discussion assumes that N is a power of two. If this is not the case, then the relative computational efficiency of the two procedures might be different. Note, also, that there are several other issues that could affect this comparison. For instance, if only the lags $\{\hat{r}(k)\}_{k=0}^{M-1}$ (with $M \ll N$) are required, then the number of computations required by (2.2.4) is drastically reduced. On the other hand, the FFT-based procedure can also be implemented in a more efficient way in such a case, and so it would remain more efficient computationally than a direct calculation when, for example, $N \geq 100$ [OPPENHEIM AND SCHAFER 1989]. We conclude that the various implementation details could change the value of N beyond which the FFT-based procedure is more efficient than the direct approach and hence might influence the decision as to which of the two procedures should be used in a given application.

2.8.2 FFT-Based Computation of Windowed Blackman–Tukey Periodograms

The windowed Blackman–Tukey periodogram (2.5.1), unlike its unwindowed version, is not amenable to a direct computation via a single FFT. In this complement, we show that three FFTs are sufficient to evaluate (2.5.1): two FFTs for the computation of the sample covariance sequence entering the equation (2.5.1) (as described in Complement 2.8.1), and one FFT for the evaluation of (2.5.1). We also show that the computational formula for $\{\hat{r}(k)\}$ derived in Complement 2.8.1 can be used to obtain an FFT-based algorithm for evaluation of (2.5.1) directly in terms of $\hat{\phi}_p(\omega)$. We relate the latter way of computing (2.5.1) to the evaluation of $\hat{\phi}_{BT}(\omega)$ from the integral equation (2.5.3). Finally, we compare the two ways just outlined for evaluating the windowed Blackman–Tukey periodogram.

The windowed Blackman–Tukey periodogram can be written as

$$
\begin{aligned}
\hat{\phi}_{BT}(\omega) &= \sum_{k=-(N-1)}^{N-1} w(k)\hat{r}(k)e^{-i\omega k} \\
&= \sum_{k=0}^{N-1} w(k)\hat{r}(k)e^{-i\omega k} + \sum_{k=0}^{N-1} w(k)\hat{r}^*(k)e^{i\omega k} - w(0)\hat{r}(0) \\
&= 2\,\mathrm{Re}\left\{\sum_{k=0}^{N-1} w(k)\hat{r}(k)e^{-i\omega k}\right\} - w(0)\hat{r}(0)
\end{aligned}
\tag{2.8.5}
$$

where we made use of the facts that the window sequence is even and $\hat{r}(-k) = \hat{r}^*(k)$. It is now evident that an N-point FFT can be used to evaluate $\hat{\phi}_{BT}(\omega)$ at $\omega = 2\pi k/N$ $(k = 0, \ldots, N-1)$. This requires about $\frac{1}{2}N \log_2(N)$ flops that should be added to the $2N \log_2(2N)$ flops required

to compute $\{\hat{r}(k)\}$ (as in Complement 2.8.1), giving a total of about $N[\frac{1}{2}\log_2(N) + 2\log_2(2N)]$ flops for this way of evaluating $\hat{\phi}_{BT}(\omega)$.

Next, we make use of the expression (2.8.4) for $\{\hat{r}(k)\}$ that is derived in Complement 2.8.1. We have

$$\hat{r}(p - N) = \frac{1}{2N - 1} \sum_{k=1}^{2N-1} \hat{\phi}(\bar{\omega}_k)e^{i\bar{\omega}_k(p-N)} \qquad (p = 1, \ldots, 2N - 1) \qquad (2.8.6)$$

where $\bar{\omega}_k = 2\pi(k - 1)/(2N - 1)$ for $(k = 1, \ldots, 2N - 1)$ and where $\hat{\phi}(\omega)$ is the unwindowed periodogram. Inserting (2.8.6) into (2.5.1), we obtain

$$\hat{\phi}_{BT}(\omega) = \frac{1}{2N - 1} \sum_{s=-(N-1)}^{N-1} w(s)e^{-i\omega s} \sum_{k=1}^{2N-1} \hat{\phi}(\bar{\omega}_k)e^{i\bar{\omega}_k s}$$

$$= \frac{1}{2N - 1} \sum_{k=1}^{2N-1} \hat{\phi}(\bar{\omega}_k) \left[\sum_{s=-(N-1)}^{N-1} w(s)e^{-i(\omega-\bar{\omega}_k)s} \right] \qquad (2.8.7)$$

which gives

$$\hat{\phi}_{BT}(\omega) = \frac{1}{2N - 1} \sum_{k=1}^{2N-1} \hat{\phi}(\bar{\omega}_k)\, W(\omega - \bar{\omega}_k) \qquad (2.8.8)$$

where $W(\omega)$ is the spectral window.

It might be thought that the last step in the preceding derivation requires that $\{w(k)\}$ be a "truncated-type" window (i.e., $w(k) = 0$ for $|k| \geq N$). However, no such requirement on $\{w(k)\}$ is needed: By inserting the usual expression for $\hat{\phi}(\omega)$ into (2.8.6), we obtain

$$\hat{r}(p - N) = \frac{1}{2N - 1} \sum_{k=1}^{2N-1} \left[\sum_{s=-(N-1)}^{N-1} \hat{r}(s)e^{-i\bar{\omega}_k s} \right] e^{i\bar{\omega}_k(p-N)}$$

$$= \frac{1}{2N - 1} \sum_{s=-(N-1)}^{N-1} \hat{r}(s) \left[\sum_{k=1}^{2N-1} e^{i\bar{\omega}_k(p-N-s)} \right]$$

$$\triangleq \frac{1}{2N - 1} \sum_{s=-(N-1)}^{N-1} \hat{r}(s)\Delta(s, p)$$

where

$$\Delta(s,p) = \sum_{k=1}^{2N-1} e^{i\bar{\omega}_{p-N-s}k} = e^{i\bar{\omega}_{p-N-s}} \frac{e^{i(2N-1)\bar{\omega}_{p-N-s}} - 1}{e^{i\bar{\omega}_{p-N-s}} - 1}$$

Now, $(2N-1)\bar{\omega}_{p-N-s} = 2\pi(p-N-s)$, so it follows that

$$\Delta(s,p) = (2N-1)\delta_{p-N,s}$$

from which we immediately get

$$\frac{1}{2N-1} \sum_{s=-(N-1)}^{N-1} \hat{r}(s)\Delta(s,p) = \begin{cases} \hat{r}(p-N) & p = 1,\ldots,2N-1 \\ 0, & \text{otherwise} \end{cases} \tag{2.8.9}$$

First, this calculation provides a cross-checking of the derivation of equation (2.8.6) in Complement 2.8.1. Second, the result (2.8.9) implies that the values of $\hat{r}(p-N)$ calculated with the formula (2.8.6) are equal to zero for $p < 1$ or $p > 2N - 1$. It follows that the limits for the summation over s in (2.8.7) can be extended to $\pm\infty$, which shows that (2.8.8) is valid for an arbitrary window.

In the general case, there seems to be no way of evaluating (2.8.8) by means of an FFT algorithm. Hence, it appears that, for a general window, it is more efficient to base the computation of $\hat{\phi}_{BT}(\omega)$ on (2.8.5) rather than on (2.8.8). For certain windows, however, (2.8.8) could be more efficient than (2.8.5) computationally. For instance, in the case of the Daniell method, which corresponds to a rectangular spectral window, (2.8.8) takes a very convenient computational form and should be preferred to (2.8.5). It should be noted that (2.8.8) can be viewed as an *exact formula* for evaluation of the integral in equation (2.5.3). In particular, (2.8.8) provides an *exact* implementation formula for the Daniell periodogram (2.7.20) (whereas (2.7.16) is only an approximation of the integral (2.7.20) that is valid for sufficiently large values of N).

2.8.3 Data- and Frequency-Dependent Temporal Windows: The Apodization Approach

All windows discussed so far are both data- and frequency-independent; in other words, the window used is the same at any frequency of the spectrum and for any data sequence. Apparently, this is a rather serious restriction. A consequence of this restriction is that, for such nonadaptive windows (i.e., windows that do not adapt to the data under analysis) any attempt to reduce the leakage effect (by keeping the sidelobes low) inherently leads to a reduction of the resolution (due to the widening of the main lobe), and vice versa; see Section 2.6.1.

In this complement, we show how to design a *data- and frequency-dependent temporal window* that has the following desirable properties:

- It mitigates the leakage problem of the periodogram without compromising its resolution; and
- It does so with only a very marginal increase in the computational burden.

Our presentation is based on the *apodization approach* of [STANKWITZ, DALLAIRE, AND FIENUP 1994], even though in some places we will deviate from it to some extent. Apodization is a term borrowed from optics where it has been used to mean a reduction of the sidelobes induced by diffraction.

We begin our presentation with a derivation of the temporally windowed periodogram, (2.6.24), in a least-squares (LS) framework. Consider the weighted LS fitting problem

$$\min_a \sum_{t=1}^{N} \rho(t) \left| y(t) - ae^{i\omega t} \right|^2 \tag{2.8.10}$$

where ω is given and so are the weights $\rho(t) \geq 0$. It can readily be verified that the minimizer of (2.8.10) is given by

$$\hat{a} = \frac{\sum_{t=1}^{N} \rho(t)y(t)e^{-i\omega t}}{\sum_{t=1}^{N} \rho(t)} \tag{2.8.11}$$

If we let

$$v(t) = \frac{\rho(t)}{\sum_{t=1}^{N} \rho(t)} \tag{2.8.12}$$

then we can rewrite (2.8.11) as a windowed DFT:

$$\hat{a} = \sum_{t=1}^{N} v(t)y(t)e^{-i\omega t} \tag{2.8.13}$$

The squared magnitude of (2.8.13) appears in the windowed periodogram formula (2.6.24)—of course, not accidentally, as $|\hat{a}|^2$ should indicate the power in $y(t)$ at frequency ω (*cf.* (2.8.10)).

The usefulness of the LS-based derivation of (2.6.24) lies in the fact that it reveals two constraints that must be satisfied by a temporal window, namely,

$$v(t) \geq 0 \tag{2.8.14}$$

which follows from $\rho(t) \geq 0$, and

$$\sum_{t=1}^{N} v(t) = 1 \tag{2.8.15}$$

which follows from (2.8.12). The constraint (2.8.15) can also be obtained by inspection of (2.6.24); indeed, if $y(t)$ had a component with frequency ω, then that component would pass undistorted (or unbiased) through the DFT in (2.6.24) if and only if (2.8.15) holds. For this reason, (2.8.15) is

sometimes called the unbiasedness condition. On the other hand, the constraint (2.8.14) appears to be more difficult to obtain directly from (2.6.24).

Next, we turn our attention to window design, which is the problem of main interest here. To emphasize the dependence of the temporally windowed periodogram in (2.6.24) on $\{v(t)\}$, we use the notation $\hat{\phi}_v(\omega)$:

$$\hat{\phi}_v(\omega) = N \left| \sum_{t=1}^{N} v(t) y(t) e^{-i\omega t} \right|^2 \tag{2.8.16}$$

Note that, in (2.8.16), the squared modulus is multiplied by N, whereas, in (2.6.24), it is divided by N; this difference is due to the fact that the window $\{v(t)\}$ in this complement is constrained to satisfy (2.8.15), whereas in Section 2.6 it is implicitly assumed to satisfy $\sum_{t=1}^{N} v(t) = N$.

In the apodization approach, the window is selected such that

$$\boxed{\hat{\phi}_v(\omega) = \text{minimum}} \tag{2.8.17}$$

for *each* ω and for the given data sequence. Evidently, the apodization window will in general be both frequency and data dependent. Sometimes, such a window is said to be *frequency and data adaptive*. Let $\mathcal{C}$ denote the class of windows over which we perform the minimization in (2.8.17). Each window in $\mathcal{C}$ must satisfy the constraints (2.8.14) and (2.8.15). Usually, $\mathcal{C}$ is generated by an archetype window that depends on a number of unknown or free parameters, most commonly in a linear manner. It is important to observe that *we should not use more than two free parameters* to describe the windows $v(t) \in \mathcal{C}$. Indeed, one parameter is needed to satisfy the constraint (2.8.15) and the remaining one(s) to minimize the function in (2.8.17) under the inequality constraint (2.8.14); if, in the minimization operation, $\hat{\phi}_v(\omega)$ depends quadratically on more than one parameter, then in general the minimum value will be zero, $\hat{\phi}_v(\omega) = 0$ for all ω, which is not acceptable. We postpone a more detailed discussion on the parameterization of $\mathcal{C}$ until we have presented a motivation for the apodization design criterion in (2.8.17).

To understand *intuitively* why (2.8.17) makes sense, consider an example in which the data consists of two noise-free sinusoids. In this example, we use a rectangular window $\{v_1(t)\}$ and a Kaiser window $\{v_2(t)\}$. The use of these windows leads to the windowed periodograms in Figure 2.5. As is apparent from this figure, $v_1(t)$ is a "high-resolution" window that trades off leakage for resolution, whereas $v_2(t)$ compromises resolution (the two sinusoids are not resolved in the corresponding periodogram) for less leakage. By using the apodization principle in (2.8.17) to choose between $\hat{\phi}_{v_1}(\omega)$ and $\hat{\phi}_{v_2}(\omega)$, at each frequency ω, we obtain the spectral estimate shown in Figure 2.5, which inherits the high resolution of $\hat{\phi}_{v_1}(\omega)$ and the low leakage of $\hat{\phi}_{v_2}(\omega)$.

A *more formal* motivation of the apodization approach can be obtained as follows. Let

$$h_t = v(t) e^{-i\omega t}$$

In terms of $\{h_t\}$ the equality constraint (2.8.15) becomes

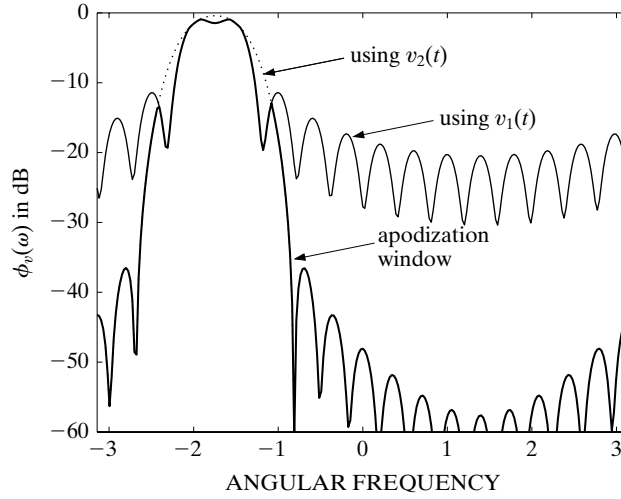$$\sum_{t=1}^{N} h_t e^{i\omega t} = 1 \tag{2.8.18}$$

Figure 2.5 An apodization window design example using a rectangular window ($v_1(t)$) and a Kaiser window ($v_2(t)$). Shown are the periodograms corresponding to $v_1(t)$ and $v_2(t)$, and to the apodization window $v(t)$ selected using (2.8.17), for a data sequence of length 16 consisting of two noise-free sinusoids.

and hence the apodization design problem is to minimize

$$\left| \sum_{t=1}^{N} h_t y(t) \right|^2 \tag{2.8.19}$$

subject to (2.8.18), (2.8.14), and any other conditions resulting from the parameterization used for $\{v(t)\}$ (and therefore for $\{h_t\}$). We can interpret $\{h_t\}$ as an FIR filter of length N; consequently, (2.8.19) is the "power" of the filter output, and (2.8.18) is the (complex) gain of the filter at frequency ω. Therefore, by making use of $\{h_t\}$, we can describe the apodization principle in words as follows: Find the (parameterized) FIR filter $\{h_t\}$ that passes without distortion the sinusoid with frequency ω (see (2.8.18)) and minimizes the output power (see (2.8.19)) and thus attenuates any other frequency components in the data as much as possible. The (normalized) power at the output of the filter is taken as an estimate of the power in the data at frequency ω. This interpretation clearly can serve as a motivation of the apodization approach, and it sheds more light on the apodization principle. In effect, minimizing (2.8.19) subject to (2.8.18) (along with the other constraints on $\{h_t\}$ resulting from the parameterization used for $\{v(t)\}$) is a special case of a sound approach to spectral analysis that will be described in Section 5.4.1 (a fact noted, apparently for the first time, in [LEE AND MUNSON JR. 1995]).

As already stated previously, an important aspect that remains to be discussed is *the parameterization* of $\{v(t)\}$. For the apodization principle to make sense, the class $\mathcal{C}$ of windows must be chosen carefully. In particular, as just explained, we should not use more than two parameters to describe $\{v(t)\}$ (to prevent the meaningless "spectral estimate" $\hat{\phi}_v(\omega) \equiv 0$). The choice of the

class $\mathcal{C}$ is also important from a *computational standpoint*. Indeed, the task of solving (2.8.17), for each ω and then computing the corresponding $\hat{\phi}_v(\omega)$ could be computationally demanding unless $\mathcal{C}$ is chosen carefully.

In what follows, we will consider the class of temporal windows used in [STANKWITZ, DALLAIRE, AND FIENUP 1994]:

$$v(t) = \frac{1}{N}\left[\alpha - \beta \cos\left(\frac{2\pi}{N}t\right)\right], \quad t = 1, \ldots, N \tag{2.8.20}$$

It can readily be checked that (2.8.20) satisfies the constraints (2.8.14) and (2.8.15) if and only if

$$\alpha = 1 \text{ and } |\beta| \leq 1 \tag{2.8.21}$$

In addition, we require that

$$\beta \geq 0 \tag{2.8.22}$$

to ensure that the peak of $v(t)$ occurs in the middle of the interval $[1, N]$; this condition guarantees that the window in (2.8.20) (with $\beta > 0$) has lower sidelobes than the rectangular window corresponding to $\beta = 0$. (The window (2.8.20) with $\beta < 0$ generally has higher sidelobes than the rectangular window; hence, $\beta < 0$ cannot be a solution to the apodization design problem.)

Remark: The *temporal* window (2.8.20) is of the same type as the *lag* Hanning and Hamming windows in Table 2.1. For the latter windows, the interval of interest is $[-N, N]$; hence, for the peak of these windows to occur in the middle of the interval of interest, we need $\beta \leq 0$ (*cf.* Table 2.1). This observation explains the difference between (2.8.20) and the lag windows in Table 2.1. ∎

Combining (2.8.20), (2.8.21), and (2.8.22) leads to the following (constrained) parameterization of the temporal windows:

$$\begin{aligned} v(t) &= \frac{1}{N}\left[1 - \beta \cos\left(\frac{2\pi}{N}t\right)\right] \\ &= \frac{1}{N}\left[1 - \frac{\beta}{2}\left(e^{i\frac{2\pi}{N}t} + e^{-i\frac{2\pi}{N}t}\right)\right], \quad \beta \in [0, 1] \end{aligned} \tag{2.8.23}$$

Assume, for simplicity, that N is a power of two (for the general case, we refer to [DEGRAAF 1994]) and that a RADIX-2 FFT algorithm is used to compute

$$Y(k) = \sum_{t=1}^{N} y(t)e^{-i\frac{2\pi k}{N}t}, \quad k = 1, \ldots, N \tag{2.8.24}$$

as discussed in Section 2.3. Then the windowed periodogram corresponding to (2.8.23) can conveniently be computed, as follows:

$$\hat{\phi}_v(k) = \frac{1}{N} \left| Y(k) - \frac{\beta}{2} \left[Y(k-1) + Y(k+1) \right] \right|^2, \quad k = 2, \ldots, N-1 \qquad (2.8.25)$$

Furthermore, in (2.8.25), β is the solution to the following apodization design problem:

$$\min_{\beta \in [0,1]} \left| Y(k) - \frac{\beta}{2} \left[Y(k-1) + Y(k+1) \right] \right|^2 \qquad (2.8.26)$$

The unconstrained minimizer of the preceding function is given by

$$\beta_0 = \text{Re} \left[\frac{2Y(k)}{Y(k-1) + Y(k+1)} \right] \qquad (2.8.27)$$

Because the function in (2.8.26) is quadratic in β, it follows that the constrained minimizer of (2.8.26) is given by

$$\beta = \begin{cases} 0, & \text{if } \beta_0 < 0 \\ \beta_0, & \text{if } 0 \le \beta_0 \le 1 \\ 1, & \text{if } \beta_0 > 1 \end{cases} \qquad (2.8.28)$$

Remark: It is interesting to note, from (2.8.28), that a change of the value of α in the window expression (2.8.20) will affect the apodization (optimal) window in a more complicated way than just a simple scaling. Indeed, if we change the value of α, for instance to $\alpha = 0.75$, then the interval for β becomes $\beta \in [0, 0.75]$ and this modification will affect the apodization window *nonlinearly* via (2.8.28). ∎

    The apodization-based windowed periodogram is obtained simply, by using the β given by (2.8.28) *in* (2.8.25). Hence, despite the fact that the apodization window is both frequency- and data-dependent (via β in (2.8.27), (2.8.28)), the implementation of the corresponding spectral estimate is only marginally more demanding computationally than is the implementation of an unwindowed periodogram. Compared with the latter, however, the apodization-based windowed periodogram has a considerably reduced leakage problem and essentially the same resolution. (See [STANKWITZ, DALLAIRE, AND FIENUP 1994; DEGRAAF 1994] for numerical examples illustrating this fact.)

2.8.4 Estimation of Cross-Spectra and Coherency Spectra

As can be seen from Complement 1.6.1, the estimation of the cross-spectrum $\phi_{yu}(\omega)$ of two stationary signals, $y(t)$ and $u(t)$, is a useful operation in the study of possible linear (dynamic)

relations between $y(t)$ and $u(t)$. Let $z(t)$ denote the bivariate signal

$$z(t) = [y(t) \ u(t)]^T$$

and let

$$\hat{\phi}(\omega) = \frac{1}{N} \, Z(\omega) Z^*(\omega) \tag{2.8.29}$$

denote the unwindowed periodogram estimate of the spectral density matrix of $z(t)$. In equation (2.8.29),

$$Z(\omega) = \sum_{t=1}^{N} z(t) e^{-i\omega t}$$

is the DTFT of $\{z(t)\}_{t=1}^{N}$. Partition $\hat{\phi}(\omega)$ as

$$\hat{\phi}(\omega) = \begin{bmatrix} \hat{\phi}_{yy}(\omega) & \hat{\phi}_{yu}(\omega) \\ \hat{\phi}_{yu}^*(\omega) & \hat{\phi}_{uu}(\omega) \end{bmatrix} \tag{2.8.30}$$

As indicated by the notation previously used, estimates of $\phi_{yy}(\omega)$, of $\phi_{uu}(\omega)$, and of the cross-spectrum $\phi_{yu}(\omega)$ may be obtained from the corresponding elements of $\hat{\phi}(\omega)$.

We first show that the estimate of the coherence spectrum obtained from (2.8.30) is always such that

$$\left| \hat{C}_{yu}(\omega) \right| = 1 \qquad \text{for all } \omega \tag{2.8.31}$$

and, hence, it is useless. To see this, note that the rank of the 2×2 matrix in (2.8.30) is equal to one (see Result R22 in Appendix A), so we must have

$$\hat{\phi}_{uu}(\omega) \hat{\phi}_{yy}(\omega) = \left| \hat{\phi}_{yu}(\omega) \right|^2$$

which readily leads to the conclusion that the coherency spectrum estimate obtained from the elements of $\hat{\phi}(\omega)$ is bound to satisfy (2.8.31) and hence is meaningless. This result is yet another indication that the unwindowed periodogram is a poor estimate of the PSD.

Consider next a windowed Blackman–Tukey periodogram estimate of the cross-spectrum:

$$\hat{\phi}_{yu}(\omega) = \sum_{k=-M}^{M} w(k) \hat{r}_{yu}(k) e^{-i\omega k} \tag{2.8.32}$$

where $w(k)$ is the lag window, and $\hat{r}_{yu}(k)$ is some usual estimate of $r_{yu}(k)$. Unlike $r_{yy}(k)$ or $r_{uu}(k)$, $r_{yu}(k)$ does not necessarily peak at $k = 0$; moreover, it is in general not an even function.

The choice of the lag window for estimating cross-spectra may hence be governed by different rules from those commonly used in the autospectrum estimation.

   The main task of a lag window is to retain the "essential part" of the covariance sequence in the defining equation for the spectral density. In this way, the bias is kept small and the variance is also reduced as the noisy tails of the sample covariance sequence are weighted out. For simplicity of discussion, assume that most of the area under the plot of $\hat{r}_{yu}(k)$ is concentrated about $k = k_0$, with $|k_0| \ll N$. As $\hat{r}_{yu}(k)$ is a reasonably accurate estimate of $r_{yu}(k)$, provided $|k| \ll N$, we can assume that $\{\hat{r}_{yu}(k)\}$ and $\{r_{yu}(k)\}$ have similar shapes. In such a case, one can redefine (2.8.32) as

$$\hat{\phi}_{yu}(\omega) = \sum_{k=-M}^{M} w(k - k_0)\hat{r}_{yu}(k)e^{-i\omega k}$$

where the lag window $w(s)$ is of the type recommended for autospectrum estimation. The choice of an appropriate value for k_0 in the preceding cross-spectral estimator is essential—if k_0 is selected poorly, the following situations can occur:

- If M is chosen small (to reduce the variance), the bias could be significant, because "essential" lags of the cross-covariance sequence could be left out.
- If M is chosen large (to reduce the bias), the variance could be significantly inflated when poorly estimated high-order "nonessential" lags are included in the spectral estimation formula.

   Finally, let us look at the cross-spectrum estimators derived from (2.8.30) and (2.8.32), respectively, with a view to establishing a relation between them. Partition $Z(\omega)$ as

$$Z(\omega) = \begin{bmatrix} Y(\omega) \\ U(\omega) \end{bmatrix}$$

and observe that

$$\frac{1}{2\pi N} \int_{-\pi}^{\pi} Y(\omega)U^*(\omega)e^{i\omega k} \, d\omega$$

$$= \frac{1}{2\pi N} \int_{-\pi}^{\pi} \sum_{t=1}^{N} \sum_{s=1}^{N} y(t)u^*(s)e^{-i\omega(t-s)} \, e^{i\omega k} \, d\omega$$

$$= \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} y(t)u^*(s)\delta_{k,t-s}$$

$$= \frac{1}{N} \sum_{t\in[1,N]\cap[1+k,N+k]} y(t)u^*(t-k) \triangleq \hat{r}_{yu}(k) \qquad (2.8.33)$$

where $\hat{r}_{yu}(k)$ can be rewritten in the following more familiar form:

$$
\hat{r}_{yu}(k) = \begin{cases} \dfrac{1}{N} \displaystyle\sum_{t=k+1}^{N} y(t)u^*(t-k), & k = 0, 1, 2, \ldots \\[2em] \dfrac{1}{N} \displaystyle\sum_{t=1}^{N+k} y(t)u^*(t-k), & k = 0, -1, -2, \ldots \end{cases}
$$

Let

$$
\hat{\phi}_{yu}^p(\omega) = \frac{1}{N} \, Y(\omega)U^*(\omega)
$$

denote the unwindowed cross-spectral periodogram-like estimator, given by the off-diagonal element of $\hat{\phi}(\omega)$ in (2.8.30). With this notation, (2.8.33) can be written more compactly as

$$
\hat{r}_{yu}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\phi}_{yu}^p(\mu)e^{i\mu k} \, d\mu
$$

By using the previous equation in (2.8.32), we obtain

$$
\hat{\phi}_{yu}(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\phi}_{yu}^p(\mu) \sum_{k=-M}^{M} w(k)e^{-i(\omega-\mu)k} \, d\mu
$$

$$
= \frac{1}{2\pi} \int_{-\pi}^{\pi} W(\omega-\mu)\hat{\phi}_{yu}^p(\mu) \, d\mu \tag{2.8.34}
$$

where $W(\omega) = \sum_{k=-\infty}^{\infty} w(k)e^{-i\omega k}$ is the spectral window. The previous equation should be compared with the similar equation, (2.5.3), that holds in the case of autospectra.

For implementation purposes, one can use the discrete approximation of (2.8.34) given that

$$
\boxed{\hat{\phi}_{yu}(\omega) = \frac{1}{N} \sum_{k=-N}^{N} W(\omega-\omega_k)\hat{\phi}_{yu}^p(\omega_k)}
$$

where $\omega_k = \frac{2\pi}{N}k$ are the Fourier frequencies. The periodogram (cross-spectral) estimate that appears in the above equation can be computed efficiently by means of an FFT algorithm.

2.8.5 More Time-Bandwidth Product Results

The time (or duration)-bandwidth product result (2.6.5) relies on the assumptions that both $w(t)$ and $W(\omega)$ have a dominant peak at the origin, that they both are real valued, and that they take

on nonnegative values only. While most window-like signals (nearly) satisfy these assumptions, many other signals do not satisfy them. In this complement, we obtain time-bandwidth product results that apply to a much broader class of signals.

We begin by showing how the result (2.6.5) can be extended to a more general class of signals. Let $x(t)$ denote a general discrete-time sequence, and let $X(\omega)$ denote its DTFT. Both $x(t)$ and $X(\omega)$ are allowed to take negative or complex values, and neither is required to peak at the origin. Let t_0 and ω_0 denote the maximum points of $|x(t)|$ and $|X(\omega)|$, respectively. The time width (or duration) and bandwidth definitions in (2.6.1) and (2.6.2) are modified as follows:

$$\bar{N}_e = \frac{\sum_{t=-\infty}^{\infty} |x(t)|}{|x(t_0)|}$$

and

$$\bar{\beta}_e = \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\omega)| \, d\omega}{|X(\omega_0)|}$$

Because $x(t)$ and $X(\omega)$ form a Fourier transform pair, we obtain

$$|X(\omega_0)| = \left| \sum_{t=-\infty}^{\infty} x(t) e^{-i\omega_0 t} \right| \leq \sum_{t=-\infty}^{\infty} |x(t)|$$

and

$$|x(t_0)| = \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{i\omega t_0} \, d\omega \right| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\omega)| \, d\omega$$

which implies that

$$\boxed{\bar{N}_e \bar{\beta}_e \geq 1} \qquad (2.8.35)$$

This result, similar to (2.6.5), can be used to conclude the following:

$$\boxed{\text{A sequence } \{x(t)\} \text{ cannot be narrow in both time and frequency.}} \qquad (2.8.36)$$

More precisely, if $x(t)$ is narrow in one domain, it must be wide in the other domain. However, the inequality result (2.8.35), unlike (2.6.5), does not necessarily imply that $\bar{\beta}_e$ decreases whenever $\bar{N}_e$ increases (or vice versa). Furthermore, the result (2.8.35)—again unlike (2.6.5)—does not exclude the possibility that the signal is broad in both domains. In fact, in the general class of signals to which (2.8.35) applies, there are signals that are broad both in the time and in the frequency domain. (For such signals $\bar{N}_e \bar{\beta}_e \gg 1$; see, for example, [PAPOULIS 1977].) Evidently, the significant consequence of (2.8.35) is (2.8.36), which is precisely what makes the duration-bandwidth result an important one.

The duration-bandwidth product type of result (such as (2.6.5) or (2.8.35), and later (2.8.40)) has sometimes been referred to by using the generic name *uncertainty principle*, in an attempt

to relate it to the Heisenberg Uncertainty Principle in quantum mechanics. (Briefly stated, the Heisenberg Uncertainty Principle asserts that the position and velocity of a particle cannot simultaneously be specified to arbitrary precision.) To support the relationship, one can argue as follows: Suppose that we are given a sequence with (equivalent) duration equal to N_e and that we are asked to use a linear filtering device to determine the sequence's spectral content in a certain narrow band. Because the filter impulse response cannot be longer than N_e (in fact, it should be (much) shorter!), it follows from the time-bandwidth product result that the filter's bandwidth can be on the order of $1/N_e$, *but not smaller*. Hence, the sequence's spectral content in fine bands on an order smaller than $1/N_e$ cannot be ascertained exactly and therefore is "uncertain." This is, in effect, the type of limitation that applies to the nonparametric spectral methods discussed in this chapter. However, this way of arguing is related to a *specific approach to spectral estimation and not to a fundamental limitation associated with the signal itself.* (As we will see in later chapters of this text, there are parametric methods of spectral analysis that can provide the "high resolution" necessary to determine the spectral content in bands that are on an order less than $1/N_e$.)

Next, we present another, slightly more general form of time-bandwidth product result. The definitions of duration and bandwidth used to obtain (2.8.35) make full sense whenever $|x(t)|$ and $|X(\omega)|$ are single pulse-like waveforms, though these definitions might give reasonable results in many other instances as well. There are several other possible definitions of the broadness of a waveform in either the time or the frequency domain. The definition used next and the corresponding time-bandwidth product result appear to be among the most general.

Let

$$\tilde{x}(t) = \frac{x(t)}{\sqrt{\sum_{t=-\infty}^{\infty} |x(t)|^2}} \tag{2.8.37}$$

and

$$\tilde{X}(\omega) = \frac{X(\omega)}{\sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\omega)|^2 d\omega}} \tag{2.8.38}$$

By Parseval's theorem (see (1.2.6)), the denominators in (2.8.37) and (2.8.38) are equal to each other. Therefore, $\tilde{X}(\omega)$ is the DTFT of $\tilde{x}(t)$, as is already indicated by notation. Observe that

$$\sum_{t=-\infty}^{\infty} |\tilde{x}(t)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\tilde{X}(\omega)|^2 d\omega = 1$$

Hence, both $\{|\tilde{x}(t)|^2\}$ and $\{|\tilde{X}(\omega)|^2/2\pi\}$ can be interpreted as probability density functions in the sense that they are nonnegative and that they sum or integrate to one. The means and variances associated with these two "probability" densities are given by the following equations:

Time Domain:

$$\mu = \sum_{t=-\infty}^{\infty} t|\tilde{x}(t)|^2$$

$$\sigma^2 = \sum_{t=-\infty}^{\infty} (t-\mu)^2|\tilde{x}(t)|^2$$

Frequency Domain:

$$v = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \omega |\tilde{X}(\omega)|^2 d\omega$$

$$\rho^2 = \frac{1}{(2\pi)^3} \int_{-\pi}^{\pi} (\omega - 2\pi v)^2 |\tilde{X}(\omega)|^2 d\omega$$

The values of the "standard deviations" σ and ρ show whether the normalized functions $\{|\tilde{x}(t)|\}$ and $\{|\tilde{X}(\omega)|\}$, respectively, are narrow or broad. Hence, *we can use σ and ρ as definitions for the duration and bandwidth, respectively,* of the original functions $\{x(t)\}$ and $\{X(\omega)\}$.

In what follows, we assume that

$$\mu = 0, \qquad v = 0 \tag{2.8.39}$$

For continuous-time signals, the zero-mean assumptions can always be made to hold by translating the origin appropriately on the time and frequency axes; see, for example, [COHEN 1995]. However, doing the same in the case of the discrete-time sequences considered here does not appear to be possible. Indeed, μ might not be integer valued, and the support of $X(\omega)$ is finite and, hence, affected by translation. Consequently, in the present case, the zero-mean assumption introduces some restriction; nevertheless, we impose it to simplify the analysis.

According to this discussion and assumption (2.8.39), we define the (equivalent) time width and bandwidth of $x(t)$ as follows:

$$\tilde{N}_e = \left[\sum_{t=-\infty}^{\infty} t^2 |\tilde{x}(t)|^2 \right]^{1/2}$$

$$\tilde{\beta}_e = \frac{1}{2\pi} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \omega^2 |\tilde{X}(\omega)|^2 d\omega \right]^{1/2}$$

In the remainder of this complement, we prove the following time–bandwidth-product result:

$$\boxed{\tilde{N}_e \tilde{\beta}_e \geq \frac{1}{4\pi}} \tag{2.8.40}$$

which holds true under (2.8.39) and the weak additional assumption that

$$|\tilde{X}(\pi)| = 0 \tag{2.8.41}$$

To prove (2.8.40), first we note that

$$\tilde{X}'(\omega) \triangleq \frac{d\tilde{X}(\omega)}{d\omega} = -i \sum_{t=-\infty}^{\infty} t\tilde{x}(t) e^{-i\omega t}$$

Hence, $i\tilde{X}'(\omega)$ is the DTFT of $\{t\tilde{x}(t)\}$, which implies (by Parseval's theorem) that

$$\sum_{t=-\infty}^{\infty} t^2|\tilde{x}(t)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}|\tilde{X}'(\omega)|^2 d\omega \tag{2.8.42}$$

Consequently, by the Cauchy–Schwartz inequality for functions (see Result R23 in Appendix A),

$$\begin{aligned}
\tilde{N}_e\tilde{\beta}_e &= \left[\frac{1}{2\pi}\int_{-\pi}^{\pi}|\tilde{X}'(\omega)|^2 d\omega\right]^{1/2}\left[\frac{1}{(2\pi)^3}\int_{-\pi}^{\pi}\omega^2|\tilde{X}(\omega)|^2 d\omega\right]^{1/2} \\
&\geq \frac{1}{(2\pi)^2}\left|\int_{-\pi}^{\pi}\omega\tilde{X}^*(\omega)\tilde{X}'(\omega)d\omega\right| \\
&= \frac{1}{2(2\pi)^2}\left\{\left|\int_{-\pi}^{\pi}\omega\tilde{X}^*(\omega)\tilde{X}'(\omega)d\omega\right|\right. \\
&\qquad\qquad \left.+ \left|\int_{-\pi}^{\pi}\omega\tilde{X}(\omega)\tilde{X}^{*\prime}(\omega)d\omega\right|\right\}
\end{aligned} \tag{2.8.43}$$

(The first equality above follows from (2.8.42) and the last one from a simple calculation.) Hence,

$$\begin{aligned}
\tilde{N}_e\tilde{\beta}_e &\geq \frac{1}{2(2\pi)^2}\left|\int_{-\pi}^{\pi}\omega\left[\tilde{X}^*(\omega)\tilde{X}'(\omega) + \tilde{X}(\omega)\tilde{X}^{*\prime}(\omega)\right]d\omega\right| \\
&= \frac{1}{2(2\pi)^2}\left|\int_{-\pi}^{\pi}\omega\left[|\tilde{X}(\omega)|^2\right]' d\omega\right|
\end{aligned}$$

which, after integration by parts and use of (2.8.41), yields

$$\tilde{N}_e\tilde{\beta}_e \geq \frac{1}{2(2\pi)^2}\left|\omega|\tilde{X}(\omega)|^2\Big|_{-\pi}^{\pi} - \int_{-\pi}^{\pi}|\tilde{X}(\omega)|^2 d\omega\right| = \frac{1}{2(2\pi)}$$

and the proof is concluded.

Remark: There is an alternative way to complete the proof above, starting from the inequality in (2.8.43). In fact, as we will see, this alternative proof yields a tighter inequality than (2.8.40). Let $\varphi(\omega)$ denote the phase of $\tilde{X}(\omega)$:

$$\tilde{X}(\omega) = |\tilde{X}(\omega)|e^{i\varphi(\omega)}$$

Then,

$$\begin{aligned}
\omega\tilde{X}^*(\omega)\tilde{X}'(\omega) &= \omega|\tilde{X}(\omega)|\left[|\tilde{X}(\omega)|\right]' + i\omega\varphi'(\omega)|\tilde{X}(\omega)|^2 \\
&= \frac{1}{2}\left[\omega|\tilde{X}(\omega)|^2\right]' - \frac{1}{2}|\tilde{X}(\omega)|^2 + i\omega\varphi'(\omega)|\tilde{X}(\omega)|^2
\end{aligned} \tag{2.8.44}$$

Inserting (2.8.44) into (2.8.43) yields

$$\tilde{N}_e \tilde{\beta}_e \geq \frac{1}{(2\pi)^2} \left| \frac{\omega}{2} |\tilde{X}(\omega)|^2 \Big|_{-\pi}^{\pi} - \pi + i2\pi\gamma \right| \tag{2.8.45}$$

where

$$\gamma = \frac{1}{2\pi} \int_{-\pi}^{\pi} \omega \varphi'(\omega) |\tilde{X}(\omega)|^2 d\omega$$

can be interpreted as the "covariance" of ω and $\varphi'(\omega)$ under the "probability density function" given by $|\tilde{X}(\omega)|^2/(2\pi)$. From (2.8.45), we at once obtain

$$\tilde{N}_e \tilde{\beta}_e \geq \frac{1}{4\pi} \sqrt{1 + 4\gamma^2} \tag{2.8.46}$$

which is a slightly stronger result than (2.8.40).   ∎

The results (2.8.40) and (2.8.46) are similar to (2.8.35); hence, the type of comment previously made about (2.8.35) applies to (2.8.40) and (2.8.46) as well.

For a time-bandwidth product result more general than (2.8.46), see [DOROSLOVACKI 1998]; the papers [CALVEZ AND VILBÉ 1992] and [ISHII AND FURUKAWA 1986] contain results similar to the one presented in this complement.

2.9 EXERCISES

Exercise 2.1: Covariance Estimation for Signals with Unknown Means

The sample-covariance estimators (2.2.3) and (2.2.4) are based on the assumption that the signal mean is equal to zero. A simple calculation shows that, under the zero-mean assumption,

$$E\{\tilde{r}(k)\} = r(k) \tag{2.9.1}$$

and

$$E\{\hat{r}(k)\} = \frac{N - |k|}{N}\, r(k) \tag{2.9.2}$$

where $\{\tilde{r}(k)\}$ denotes the sample covariance estimate in (2.2.3). Equations (2.9.1) and (2.9.2) show that $\tilde{r}(k)$ is an unbiased estimate of $r(k)$, whereas $\hat{r}(k)$ is a biased one. (Note, however, that the bias in $\hat{r}(k)$ is small for $N \gg |k|$.) For this reason, $\{\tilde{r}(k)\}$ and $\{\hat{r}(k)\}$ are often called, respectively, the unbiased and the biased sample covariances.

Whenever the signal mean is unknown, a most natural modification of the covariance estimators (2.2.3) and (2.2.4) is as follows:

$$\tilde{r}(k) = \frac{1}{N - k} \sum_{t=k+1}^{N} [y(t) - \bar{y}] [y(t - k) - \bar{y}]^* \tag{2.9.3}$$

and

$$\hat{r}(k) = \frac{1}{N} \sum_{t=k+1}^{N} [y(t) - \bar{y}]\, [y(t-k) - \bar{y}]^* \tag{2.9.4}$$

where $\bar{y}$ is the sample mean

$$\bar{y} = \frac{1}{N} \sum_{t=1}^{N} y(t) \tag{2.9.5}$$

Show that, in the unknown-mean case, the usual names, "unbiased" and "biased" sample covariance, associated with (2.9.3) and (2.9.4), respectively, might no longer be appropriate. Indeed, in such a case *both estimators could be biased*; furthermore, $\hat{r}(k)$ *could be less biased than* $\tilde{r}(k)$. To simplify the calculations, assume that $y(t)$ is white noise.

Exercise 2.2: Covariance Estimation for Signals with Unknown Means (cont'd)

Show that the sample covariance sequence $\{\hat{r}(k)\}$ in equation (2.9.4) of Exercise 2.1 satisfies the following equality:

$$\sum_{k=-(N-1)}^{N-1} \hat{r}(k) = 0 \tag{2.9.6}$$

This equality might seem somewhat surprising. (Why should the $\{\hat{r}(k)\}$ satisfy such a constraint, which the true covariances do not necessarily satisfy? Note, for instance, that the latter covariance sequence could well comprise only positive elements.) However, the equality in (2.9.6) has a natural explanation when viewed in the context of periodogram-based spectral estimation. Derive and explain formula (2.9.6) in the aforementioned context.

Exercise 2.3: Unbiased ACS Estimates Can Lead to Negative Spectral Estimates

We stated in Section 2.2.2 that, if unbiased ACS estimates, given by equation (2.2.3), are used in the correlogram spectral estimate (2.2.2), then negative spectral estimates could result. Find an example data sequence $\{y(t)\}_{t=1}^{N}$ that gives such a negative spectral estimate.

Exercise 2.4: Variance of Estimated ACS

Let $\{y(t)\}_{t=1}^{N}$ be real Gaussian (for simplicity), with zero mean, ACS equal to $\{r(k)\}$, and ACS estimate (either biased or unbiased) equal to $\{\hat{r}(k)\}$ (given by equation (2.2.3) or (2.2.4); we treat both cases simultaneously). Assume, without loss of generality, that $k \geq 0$.

(a) Make use of equation (2.4.24) to show that

$$\text{var}\{\hat{r}(k)\} = \alpha^2(k) \sum_{m=-(N-k-1)}^{N-k-1} (N - k - |m|)\left[r^2(m) + r(m+k)r(m-k)\right]$$

where

$$\alpha(k) = \begin{cases} \dfrac{1}{N-k} & \text{for unbiased ACS estimates} \\[2mm] \dfrac{1}{N} & \text{for biased ACS estimates} \end{cases}$$

Hence, for large N, the standard deviation of the ACS estimate is $\mathcal{O}(1/\sqrt{N})$ under weak conditions on the true ACS $\{r(k)\}$.

(b) For the special case that $y(t)$ is white Gaussian noise, show that $\text{cov}\{\hat{r}(k), \hat{r}(l)\} = 0$ for $k \neq l$, and find a simple expression for $\text{var}\{\hat{r}(k)\}$.

Exercise 2.5: Another Proof of the Equality $\hat{\phi}_p(\omega) = \hat{\phi}_c(\omega)$

The proof of the result (2.2.6) in the text introduces an auxiliary random sequence and treats the original data sequence as deterministic (nonrandom). That proof relies on several results previously derived. A more direct proof of (2.2.6) can be found using only (2.2.1), (2.2.2), and (2.2.4). Find such a proof.

Exercise 2.6: A Compact Expression for the Sample ACS

Show that the expressions for the sample ACS given in the text (equations (2.2.3) or (2.2.4) for $k \geq 0$ and (2.2.5) for $k < 0$) can be rewritten by using a single formula as follows:

$$\hat{r}(k) = \rho \sum_{p=1}^{N} \sum_{s=1}^{N} y(p)y^*(s)\delta_{s,p-k}, \quad k = 0, \pm 1, \ldots, \pm(N-1) \tag{2.9.7}$$

where $\rho = \frac{1}{N}$ for (2.2.4) and $\rho = \frac{1}{N-|k|}$ for (2.2.3).

Exercise 2.7: Yet Another Proof of the Equality $\hat{\phi}_p(\omega) = \hat{\phi}_c(\omega)$

Use the compact expression for the sample ACS derived in Exercise 2.6 to obtain a very simple proof of (2.2.6).

Exercise 2.8: Linear Transformation Interpretation of the DFT

Let F be the $N \times N$ matrix whose (k, t)th element is given by W^{kt}, where W is as defined in (2.3.2). Then the DFT, (2.3.3), can be written as a linear transformation of the data vector $y \triangleq [y(1) \ldots y(N)]^T$,

$$Y \triangleq [Y(0) \ldots Y(N-1)]^T = Fy \tag{2.9.8}$$

Show that F is an orthogonal matrix that satisfies

$$\frac{1}{N} FF^* = I \tag{2.9.9}$$

and, as a result, that the *inverse transform* is

$$y = \frac{1}{N} \, F^* Y \tag{2.9.10}$$

Deduce from these results that the DFT is nothing but a representation of the data vector y via an orthogonal basis in $\mathbf{C}^n$ (the basis vectors are the columns of F^*). Also, deduce that, if the sequence $\{y(t)\}$ is periodic with a period equal to N, then the Fourier coefficient vector, Y, determines the whole sequence $\{y(t)\}_{t=1,2,\ldots}$; and that, in effect, the inverse transform (2.9.10) can be extended to include all samples $y(1), \ldots, y(N), \; y(N+1), \; y(N+2), \ldots$

Exercise 2.9: For White Noise, the Periodogram Is an Unbiased PSD Estimator
Let $y(t)$ be a zero-mean white noise with variance σ^2 and let

$$Y(\omega_k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} y(t) e^{-i\omega_k t} \; ; \quad \omega_k = \frac{2\pi}{N} k \qquad (k = 0, \ldots, N-1)$$

denote its (normalized) DFT evaluated at the Fourier frequencies.

  (a) Derive the covariances

$$E\left\{Y(\omega_k)Y^*(\omega_r)\right\}, \qquad k, r = 0, \ldots, N-1$$

  (b) Use the result of the previous calculation to conclude that the periodogram $\hat{\phi}(\omega_k) = |Y(\omega_k)|^2$ is an unbiased estimator of the PSD of $y(t)$.
  (c) Explain whether the unbiasedness property holds for $\omega \neq \omega_k$ as well. Present an intuitive explanation for your finding.

Exercise 2.10: Shrinking the Periodogram
First, we introduce a simple general result on mean squared error (MSE) reduction by shrinking. Let $\hat{x}$ be some estimate of a true (and unknown) parameter x. Assume that $\hat{x}$ is unbiased (i.e., $E\{\hat{x}\} = x$), and let $\sigma_{\hat{x}}^2$ denote the MSE of $\hat{x}$:

$$\sigma_{\hat{x}}^2 = E\left\{(\hat{x} - x)^2\right\}$$

(Since $\hat{x}$ is unbiased, $\sigma_{\hat{x}}^2$ also equals the variance of $\hat{x}$.) For a fixed (nonrandom) ρ, let

$$\tilde{x} = \rho\hat{x}$$

be another estimate of x. The "shrinkage coefficient" ρ can be chosen so as to make the MSE of $\tilde{x}$ (much) smaller than $\sigma_{\hat{x}}^2$. (Note that $\tilde{x}$, for $\rho \neq 1$, is a biased estimate of x; hence $\tilde{x}$ trades off bias for variance.) More precisely, show that the MSE of $\tilde{x}$, $\sigma_{\tilde{x}}^2$, achieves its minimum value (with respect to ρ),

$$\sigma_{\tilde{x}_o}^2 = \rho_o \, \sigma_{\hat{x}}^2$$

for

$$\rho_o = \frac{x^2}{x^2 + \sigma_{\hat{x}}^2}$$

Next, consider the application of the previous result to the periodogram. As we explained in the chapter, the periodogram-based spectral estimate is asymptotically unbiased and has an asymptotic MSE equal to the squared PSD value:

$$E\left\{\hat{\phi}_p(\omega)\right\} \to \phi(\omega), \qquad E\left\{\left(\hat{\phi}_p(\omega) - \phi(\omega)\right)^2\right\} \to \phi^2(\omega) \qquad \text{as } N \to \infty$$

Show that the "optimally shrunk" periodogram estimate is

$$\tilde{\phi}(\omega) = \hat{\phi}_p(\omega)/2$$

and that the MSE of $\tilde{\phi}(\omega)$ is $1/2$ the MSE of $\hat{\phi}_p(\omega)$.

Finally, comment on the general applicability of this extremely simple tool for MSE reduction.

Exercise 2.11: Asymptotic Maximum Likelihood Estimation of $\phi(\omega)$ from $\hat{\phi}_p(\omega)$
It follows from the calculations in Section 2.4 that, asymptotically in N, $\hat{\phi}_p(\omega)$ has mean $\phi(\omega)$ and variance $\phi^2(\omega)$. In this exercise, we assume that $\hat{\phi}_p(\omega)$ is (asymptotically) Gaussian distributed (which is *not* necessarily the case; however, the spectral estimator derived here under the Gaussian assumption may also be used when this assumption does not hold). Hence, the asymptotic probability density function of $\hat{\phi}_p(\omega)$ is (we omit both the index p and the dependence on ω to simplify the notation):

$$p_\phi(\hat{\phi}) = \frac{1}{\sqrt{2\pi\phi^2}} \exp\left[-\frac{(\hat{\phi} - \phi)^2}{2\phi^2}\right]$$

Show that the maximum likelihood estimate (MLE) of ϕ based on $\hat{\phi}$, which by definition is equal to the maximizer of $p_\phi(\hat{\phi})$ (see Appendices B and C for a short introduction of maximum likelihood estimation), is given by

$$\tilde{\phi} = \frac{\sqrt{5} - 1}{2} \, \hat{\phi}$$

Compare $\tilde{\phi}$ with the "optimally shrunk" estimate of ϕ derived in Exercise 2.10.

Exercise 2.12: Plotting the Spectral Estimates in dB
It has been shown in this chapter that the spectral estimate $\hat{\phi}(\omega)$, obtained via an improved periodogram method, is asymptotically unbiased with a variance of the form $\mu^2\phi^2(\omega)$, where μ is a constant that can be made (much) smaller than 1 by choosing the window appropriately. This fact implies that the *confidence interval* $\hat{\phi}(\omega) \pm \mu\phi(\omega)$, constructed around the estimated PSD, should include the true (and unknown) PSD with a large probability. Now, obtaining a confidence

interval as just shown has a twofold drawback: First, $\phi(\omega)$ is unknown; secondly, the interval could have significantly different widths for different frequency values.

Show that plotting $\hat{\phi}(\omega)$ in decibels eliminates the previous drawbacks. More precisely, show that, when $\hat{\phi}(\omega)$ is expressed in dB, its asymptotic variance is $c^2\mu^2$ (with $c = 10\log_{10} e$) and hence, that the confidence interval for a log-scale plot has the same width (independent of $\phi(\omega)$) for all ω.

Exercise 2.13: Finite-Sample Variance/Covariance Analysis of the Periodogram

This exercise has two aims. *First*, it shows that, in the Gaussian case, the variance/covariance analysis of the periodogram can be done in an extremely simple manner (even *without* the assumption that the data comes from a linear process, as in (2.4.26)). *Secondly*, the exercise asks for a finite-sample analysis which, for some purposes, might be more useful than the asymptotic analysis presented in the text. Indeed, the asymptotic-analysis result (2.4.21) can be misleading if not interpreted with care. For instance, (2.4.21) says that, asymptotically (for $N \to \infty$), $\hat{\phi}(\omega_1)$ and $\hat{\phi}(\omega_2)$ are uncorrelated with one another, no matter how close ω_1 and ω_2 are. This cannot be true in finite samples, and so the following question naturally arises: For a given N, how close can ω_1 be to ω_2 such that $\hat{\phi}(\omega_1)$ and $\hat{\phi}(\omega_2)$ are (nearly) uncorrelated with each other? The finite-sample analysis of this exercise can provide an answer to such questions, whereas the asymptotic analysis cannot.

Let

$$a(\omega) = [e^{i\omega} \ldots e^{iN\omega}]^T$$

$$y = [y(1) \ldots y(N)]^T$$

Then the periodogram, (2.2.1), can be written as follows (omitting the subindex p of $\hat{\phi}_p(\omega)$ in this exercise):

$$\hat{\phi}(\omega) = |a^*(\omega)y|^2/N \tag{2.9.11}$$

Assume that $\{y(t)\}$ is a zero-mean, stationary, circular Gaussian process. The "circular Gaussianity" assumption (see, for example, Appendix B) allows us to write the fourth-order moments of $\{y(t)\}$ (see equation (2.4.24)) as

$$E\left\{y(t)y^*(s)y(u)y^*(v)\right\} = E\left\{y(t)y^*(s)\right\} E\left\{y(u)y^*(v)\right\}$$
$$+ E\left\{y(t)y^*(v)\right\} E\left\{y(u)y^*(s)\right\} \tag{2.9.12}$$

Make use of (2.9.11) and (2.9.12) to show that

$$\text{cov}\{\hat{\phi}(\mu),\ \hat{\phi}(\nu)\} \triangleq E\left\{\left[\hat{\phi}(\mu) - E\{\hat{\phi}(\mu)\}\right]\left[\hat{\phi}(\nu) - E\{\hat{\phi}(\nu)\}\right]\right\}$$
$$= |a^*(\mu)Ra(\nu)|^2/N^2 \tag{2.9.13}$$

where $R = E\{yy^*\}$. Deduce from (2.9.13) that

$$\text{var}\{\hat{\phi}(\mu)\} = |a^*(\mu)Ra(\mu)|^2/N^2 \tag{2.9.14}$$

Use (2.9.14) to readily rederive the variance part of the asymptotic result (2.4.21). Next, use (2.9.14) to show that *the covariance between $\hat{\phi}(\mu)$ and $\hat{\phi}(\nu)$ is not significant if*

$$|\mu - \nu| > 4\pi/N$$

and also that it can be significant otherwise. **Hint:** To show this inequality, make use of the Carathéodory parameterization of a covariance matrix in Section 4.9.2.

Exercise 2.14: Data-Weighted ACS Estimate Interpretation of Bartlett and Welch Methods
Consider the Bartlett estimator, and assume $LM = N$.

(a) Show that the Bartlett spectral estimate can be written as

$$\hat{\phi}_B(\omega) = \sum_{k=-(M-1)}^{M-1} \tilde{r}(k)e^{-i\omega k}$$

where

$$\tilde{r}(k) = \sum_{t=k+1}^{N} \alpha(k,t)y(t)y^*(t-k), \qquad 0 \le k < M$$

for some $\alpha(k,t)$ to be derived. Note that this is nearly in the form of the Blackman–Tukey spectral estimator, with the exception that the "standard" biased ACS estimate that is used in the Blackman–Tukey estimator is replaced by the "generalized" ACS estimate $\tilde{r}(k)$.

(b) Make use of the derived expression for $\alpha(k,t)$ to conclude that the Bartlett estimator is inferior to the Blackman–Tukey estimator (especially for small N) because it fails to use all available lag products in forming ACS estimates.

(c) Find $\alpha(k,t)$ for the Welch method. What overlap values (K in equation (2.7.7)) give lag product usage similar to that in the Blackman–Tukey method?

Exercise 2.15: Approximate Formula for Bandwidth Calculation
Let $W(\omega)$ denote a general spectral window that has a peak at $\omega = 0$ and is symmetric about that point. In addition, assume that the peak of $W(\omega)$ is narrow (as it usually should be). Under these assumptions, make use of a Taylor series expansion to show that an approximate formula for calculating *the bandwidth B of the peak of $W(\omega)$* is the following:

$$\boxed{B \simeq 2\sqrt{|W(0)/W''(0)|}} \tag{2.9.15}$$

The spectral-peak bandwidth B is defined mathematically as follows: Let ω_1 and ω_2 denote the "half-power points," defined through

$$W(\omega_1) = W(\omega_2) = W(0)/2, \qquad \omega_1 < \omega_2$$

(hence, the ratio $10 \log_{10}\left(W(0)/W(\omega_j)\right) \simeq 3$ dB for $j = 1, 2$; we use $10 \log_{10}$ rather than $20 \log_{10}$ because the spectral window is applied to a power quantity, $\phi(\omega)$). Then, since $W(\omega)$ is symmetric, $\omega_2 = -\omega_1$, and

$$B \triangleq \omega_2 - \omega_1 = 2\omega_2$$

As an application of (2.9.15), show that

$$B \simeq 0.78 \cdot 2\pi/N \quad \text{(in radians per sampling interval)}$$

or, equivalently, that

$$B \simeq 0.78/N \quad \text{(in cycles per sampling interval)}$$

for the Bartlett window (2.4.15).

   Note that this formula remains approximate even as $N \to \infty$. Even though the half-power bandwidth of the window gets smaller as N increases (so that one would expect the Taylor series expansion to be more accurate), the curvature of the window at $\omega = 0$ increases without bound as N increases. For the Bartlett window, verify that $B \simeq 0.9 \cdot 2\pi/N$ for N large, which differs from the prediction in this exercise by about 16%.

Exercise 2.16: A Further Look at the Time-Bandwidth Product
We saw in Section 2.6.1 that the product between the equivalent time and frequency widths of a regular window equals unity. Use the formula (2.9.15) derived in Exercise 2.15 to show that the spectral-peak bandwidth B of a window $w(k)$ that is nonzero only for $|k| < N$ satisfies

$$\boxed{B \cdot N \geq 1/\pi \quad \text{(in cycles per sampling interval)}} \qquad (2.9.16)$$

This once again illustrates the "time-bandwidth product" type of result. Note that (2.9.16) involves the *effective* window time length and spectral *peak width*, as opposed to (2.6.5), which is concerned with *equivalent* time and frequency widths.

Exercise 2.17: Bias Considerations in Blackman–Tukey Window Design
The discussion in this chapter treated the bias of a spectral estimator and its resolution as two interrelated properties. This exercise illustrates further the strong relationship between bias and resolution.

   Consider $\hat{\phi}_{BT}(\omega)$ as in (2.5.1), and, without loss of generality, assume that $E\{\hat{r}(k)\} = r(k)$. (Generality is not lost because, if $E\{\hat{r}(k)\} = \alpha(k)r(k)$, then replacing $w(k)$ by $\alpha(k)w(k)$ and $\hat{r}(k)$

by $\hat{r}(k)/\alpha(k)$ results in an equivalent estimator with unbiased ACS estimates.) Find the weights $\{w(k)\}_{k=-M+1}^{M-1}$ that minimize the squared bias, as given by the error measure

$$\epsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\phi(\omega) - E\{\hat{\phi}_{BT}(\omega)\} \right]^2 d\omega \qquad (2.9.17)$$

In particular, show that the weight function that minimizes ϵ is the rectangular window. Recall that the rectangular window also has the narrowest main lobe and, hence, the best resolution.

Exercise 2.18: A Property of the Bartlett Window

Let the window length, M, be given. Then, in the general case, the rectangular window can be expected to yield the windowed spectral estimate with the most favorable bias properties, because the sample covariance lags $\{\hat{r}(k)\}_{k=-(M-1)}^{M-1}$, appearing in (2.5.1), are left unchanged by this window (as in Exercise 2.17). The rectangular window, however, has the drawback that it is not positive definite and hence could produce negative spectral estimates. The Bartlett window, on the other hand, is positive definite and therefore yields a spectral estimate that is positive for all frequencies. Show that the latter window is the positive definite window that is closest to the rectangular one, in the sense of minimizing the following criterion:

$$\min_{\{w(k)\}} \sum_{k=0}^{M-1} |1 - w(k)| \quad \text{subject to:} \qquad (2.9.18)$$

$$1) \; w(k) \equiv 0 \text{ for } |k| \geq M$$
$$2) \; \{w(k)\}_{k=-\infty}^{\infty} \text{ is a positive definite sequence}$$
$$3) \; w(0) = 1$$

Conclude that the Bartlett window is the positive definite window that distorts the sample covariances $\{\hat{r}(k)\}_{k=-(M-1)}^{M-1}$ *least* in the windowed spectral estimate formula. **Hint:** Any positive definite real window $\{w(k)\}_{k=-(M-1)}^{M-1}$ can be written as

$$w(k) = \sum_{i=0}^{M-1} b_i \, b_{i+k} \qquad (b_i = 0 \; \text{for} \; i \geq M) \qquad (2.9.19)$$

for some real-valued parameters $\{b_i\}_{i=0}^{M-1}$. Make use of the preceding parameterization of the set of positive definite windows to transform (2.9.18) into an optimization problem without constraints.

COMPUTER EXERCISES

Tools for Periodogram Spectral Estimation:

The text website www.prenhall.com/stoica contains the following MATLAB functions for use in computing periodogram-based spectral estimates. In each case, y is the input data vector, L

controls the frequency-sample spacing of the output, and the output vector is $\texttt{phi}= \phi(\omega_k)$, where $\omega_k = \frac{2\pi k}{L}$. MATLAB functions that generate the Correlogram, Blackman–Tukey, Windowed Periodogram, Bartlett, Welch, and Daniell spectral estimates are as follows:

- $\texttt{phi = correlogramse(y,L)}$
  Implements the correlogram spectral estimate in equation (2.2.2).
- $\texttt{phi = btse(y,w,L)}$
  Implements the Blackman–Tukey spectral estimate in equation (2.5.1); $\texttt{w}$ is the vector $[w(0), \ldots, w(M-1)]^T$.
- $\texttt{phi = periodogramse(y,v,L)}$
  Implements the windowed periodogram spectral estimate in equation (2.6.24); $\texttt{v}$ is a vector of window function elements $[v(1), \ldots, v(N)]^T$ and should be the same size as $\texttt{y}$. If $\texttt{v}$ is a vector of ones, this function implements the unwindowed periodogram spectral estimate in equation (2.2.1).
- $\texttt{phi = bartlettse(y,M,L)}$
  Implements the Bartlett spectral estimate in equations (2.7.2) and (2.7.3); $\texttt{M}$ is the size of each subsequence, as in equation (2.7.2).
- $\texttt{phi = welchse(y,v,K,L)}$
  Implements the Welch spectral estimate in equation (2.7.8); $\texttt{M}$ is the size of each subsequence, $\texttt{v}$ is the window function $[v(1), \ldots, v(M)]^T$ applied to each subsequence, and $\texttt{K}$ is the overlap parameter, as in equation (2.7.7).
- $\texttt{phi = daniellse(y,J,Ntilde)}$
  Implements the Daniell spectral estimate in equation (2.7.16); $\texttt{J}$ and $\texttt{Ntilde}$ correspond to J and $\tilde{N}$ there.

Exercise C2.19: Zero-Padding Effects on Periodogram Estimators

In this exercise, we study the effect that zero padding has on the periodogram.

Consider the sequence

$$y(t) = 10\sin(0.2 \cdot 2\pi t + \phi_1) + 5\sin((0.2 + 1/N)2\pi t + \phi_2) + e(t), \qquad (2.9.20)$$

where $t = 0, \ldots, N-1$, and $e(t)$ is white Gaussian noise with variance 1. Let $N = 64$ and $\phi_1 = \phi_2 = 0$.

From the results in Chapter 4, we find the spectrum of $y(t)$ to be

$$\phi(\omega) = 50\pi \left[\delta(\omega - 0.2 \cdot 2\pi) + \delta(\omega + 0.2 \cdot 2\pi)\right]$$

$$+12.5\pi \left[\delta(\omega - (0.2 + 1/N) \cdot 2\pi) + \delta(\omega + (0.2 + 1/N) \cdot 2\pi)\right] + 1$$

Plot the periodogram for the sequence $\{y(t)\}$ and for the sequence $\{y(t)\}$ zero padded with N, $3N$, $5N$, and $7N$ zeroes.

Explain the difference between the five periodograms. Why does the first periodogram not give a good description of the spectral content of the signal? Note that zero padding does not change the resolution of the estimator.

Exercise C2.20: Resolution and Leakage Properties of the Periodogram

We have seen from Section 2.4 that the expected value of the periodogram is the convolution of the true spectrum $\phi_y(\omega)$ with the Fourier transform of a Bartlett window, denoted $W_B(\omega)$. (See equation (2.4.15).) The shape and size of the $W_B(\omega)$ function determines the amount of *smearing* and *leakage* in the periodogram. Similarly, in Section 2.5, we introduced a windowed periodogram in (2.6.24) whose expected value is equal to the expected value of a corresponding Blackman–Tukey estimate with weights $w(k)$ given by (2.6.31). Window functions different from the rectangular window could be used in the periodogram estimate, giving rise to correspondingly different windows in the correlogram estimate. The choice of window affects the resolution and leakage properties of the periodogram (correlogram) spectral estimate.

Resolution Properties: The amount of smearing of the spectral estimate is determined by the width of the main lobe, and the amount of leakage is determined by the energy in the sidelobes. The amount of smearing is what limits the resolving power of the periodogram; it is studied empirically next.

   We first study the resolution properties by considering a sequence made up of two sinusoids in noise, where the two sinusoidal frequencies are "close". Consider

$$y(t) = a_1 \sin(f_0 \cdot 2\pi t + \phi_1) + a_2 \sin((f_0 + \alpha/N)2\pi t + \phi_2) + e(t), \qquad (2.9.21)$$

where $e(t)$ is real-valued Gaussian white noise having mean zero and variance σ^2. We choose $f_0 = 0.2$ and $N = 256$, but the results are nearly independent of f_0 and N.

(a) Find empirically the 3 dB width of the main lobe of $W_B(\omega)$ as a function of N, and verify equation (2.4.18). Also, compute the peak sidelobe height (in dB) as a function of N. Note that the sidelobe level of a window function generally is independent of N. Verify this by examining plots of the magnitude of $W_B(\omega)$ for several values of N; try both linear and dB scales in your plots.

(b) Set $\sigma^2 = 0$ (to eliminate the statistical variation in the periodogram, so that the bias properties can be isolated and studied). Set $a_1 = a_2 = 1$ and $\phi_1 = \phi_2 = 0$. Plot the (zero-padded) periodogram of $y(t)$ for various α and determine the resolution threshold (i.e., the minimum value of α for which the two frequency components can be resolved). How does this value of α compare with the predicted resolution in Section 2.4?

(c) Repeat part (b) for a Hamming-windowed correlogram estimate.

(d) For reasonably high signal-to-noise ratio (SNR) values and reasonably close signal amplitudes, the resolution thresholds in parts (b) and (c) are not very sensitive to variations in the signal amplitudes and frequency f_0. However, these thresholds are sensitive to the phases ϕ_1 and ϕ_2, especially if α is smaller than 1. Try two pairs (ϕ_1, ϕ_2) such that the two sinusoids are in phase and out of phase, respectively, at the center of the observation interval, and compare the resolution thresholds. Also, try different values of a_1, a_2, and σ^2 to verify that their values have relatively little effect on the resolution threshold.

Spectral Leakage: In this part, we analyze the effects of leakage on the periodogram estimate. Leakage properties can be seen clearly when one is trying to estimate two sinusoidal terms that are well separated, but have greatly differing amplitudes.

(a) Generate the sinusoidal sequence for $\alpha = 4$, $\sigma^2 = 0$, and $\phi_1 = \phi_2 = 0$. Set $a_1 = 1$, and vary a_2. (Choose $a_2 = 1$, 0.1, 0.01, and 0.001, for example.) Compute the periodogram (using a rectangular data window), and comment on the ability to identify the second sinusoidal term from the spectral estimate.

(b) Repeat part (a) for $\alpha = 12$. Does the amplitude threshold for identifiability of the second sinusoidal term change?

(c) Explain your results in parts (a) and (b) by looking at the amplitude of the Bartlett window's Fourier transform at frequencies corresponding to α/N for $\alpha = 4$ and $\alpha = 12$.

(d) The Bartlett window has the property (as do many other windows) that the leakage level depends on the distance between spectral components in the data, as seen in parts (a) and (b). For many practical applications, it could be known what dynamic range the sinusoidal components in the data may have, and it is thus desirable to use a data window with a constant sidelobe level that can be chosen by the user. The Chebyshev window is a good choice for these applications, because the user can select the (constant) sidelobe level in the window design. (See the MATLAB command `chebwin`.)

Assume we know that the maximum dynamic range of sinusoidal components is 60 dB. Design a Chebyshev window $v(t)$ and corresponding Blackman–Tukey window $w(k)$, using (2.6.31), so that the two sinusoidal components of the data can be resolved for this dynamic range by using (i) the Blackman–Tukey spectral estimator with window $w(k)$, and (ii) the windowed periodogram method with window $v(t)$. Plot the Fourier transform of the window and determine the spectral resolution of the window.

Test your window design by computing the Blackman–Tukey and windowed periodogram estimates for two sinusoids whose amplitudes differ by 50 dB in dynamic range and whose frequency separation is the minimum value you predicted. Compare the resolution results with your predictions. Explain why the smaller-amplitude sinusoid can be detected by using one of the methods but not the other.

Exercise C2.21: Bias and Variance Properties of the Periodogram Spectral Estimate

In this exercise, we verify the theoretical predictions about bias and variance properties of the periodogram spectral estimate. We use autoregressive moving average (ARMA) signals (see Chapter 3) as test signals.

Bias Properties—Resolution and Leakage: We consider a random process

$$y(t) = H(z)e(t)$$

generated by filtering white noise, where $e(t)$ is zero-mean Gaussian white noise with variance $\sigma^2 = 1$ and the filter $H(z)$ is given by

$$H(z) = \sum_{k=1}^{2} A_k \left[\frac{1 - z_k z^{-1}}{1 - p_k z^{-1}} + \frac{1 - z_k^* z^{-1}}{1 - p_k^* z^{-1}} \right] \tag{2.9.22}$$

with

$$\begin{aligned} p_1 &= 0.99 e^{i2\pi 0.3} & p_2 &= 0.99 e^{i2\pi(0.3+\alpha)} \\ z_1 &= 0.95 e^{i2\pi 0.3} & z_2 &= 0.95 e^{i2\pi(0.3+\alpha)} \end{aligned} \tag{2.9.23}$$

We first let $A_1 = A_2 = 1$ and $\alpha = 0.05$.

(a) Plot the true spectrum $\phi(\omega)$. Using a sufficiently fine grid for ω so that approximation errors are small, plot the ACS, using an inverse FFT of $\phi(\omega)$.

(b) For $N = 64$, plot the Fourier transform of the Bartlett window, and also plot the expected value of the periodogram estimate $\hat{\phi}_p(\omega)$, as given by equation (2.4.8). We see that, for this example and data length, the main lobe width of the Bartlett window is wider than the distance between the spectral peaks in $\phi(\omega)$. Discuss how this relatively wide main lobe width affects the resolution properties of the estimator.

(c) Generate 50 realizations of $y(t)$, each of length $N = 64$ data points. You can generate the data by passing white noise through the filter $H(z)$ (see the MATLAB commands `dimpulse` and `filter`); be sure to discard a sufficient number of initial filter output points to effectively remove the transient part of the filter output. Compute the periodogram spectral estimates for each data sequence; plot 10 spectral estimates overlaid on a single plot. Also, plot the average of the 50 spectral estimates. Compare the average with the predicted expected value found in part (b).

(d) The resolution of the spectral peaks in $\phi(\omega)$ will depend on their separation relative to the width of the Bartlett-window main lobe. Generate realizations of $y(t)$ for $N = 256$, and find the minimum value of α such that the spectral peaks can be resolved in the averaged spectral estimate. Compare your results with the predicted formula (2.4.18) for spectral resolution.

(e) Leakage from the Bartlett window will affect the ability to identify peaks of different amplitudes. To illustrate this, generate realizations of $y(t)$ for $N = 64$, for both $\alpha = 4/N$ and $\alpha = 12/N$. For each value of α, set $A_1 = 1$, and vary A_2 to find the minimum amplitude for which the lower-amplitude peak can reliably be identified from the averaged spectral estimate. Compare this value with the Bartlett window sidelobe level for $\omega = 2\pi\alpha$ and for the two values of α. Does the window sidelobe level accurately reflect the amplitude separation required to identify the second peak?

Variance Properties: In this part, we will verify that the variance of the periodogram is almost independent of the data length and will compare the empirical variance with theoretical predictions. For this part, we consider a broadband signal $y(t)$ for which the Bartlett window smearing and leakage effects are small.

Consider the broadband ARMA process

$$y(t) = \frac{B_1(z)}{A_1(z)} e(t)$$

with

$$A_1(z) = 1 - 1.3817z^{-1} + 1.5632z^{-2} - 0.8843z^{-3} + 0.4096z^{-4}$$

$$B_1(z) = 1 + 0.3544z^{-1} + 0.3508z^{-2} + 0.1736z^{-3} + 0.2401z^{-4}$$

(a) Plot the true spectrum $\phi(\omega)$.

(b) Generate 50 Monte Carlo data realizations, using different noise sequences, and compute the corresponding 50 periodogram spectral estimates. Plot the sample mean, the sample mean plus one sample standard deviation, and sample mean minus one sample standard deviation spectral estimate curves. Do this for $N = 64$, 256, and 1024. Note that the variance does not decrease with N.

(c) Compare the sample variance with the variance predicted in equation (2.4.21). It may help to plot stdev$\{\hat{\phi}(\omega)\}/\phi(\omega)$ and determine to what degree this curve is approximately constant. Discuss your results.

Exercise C2.22:   Refined Methods: Variance–Resolution Tradeoff

In this exercise, we apply the Blackman–Tukey and Welch estimators to both a narrowband and a broadband random process. We consider the same processes in Chapters 3 and 5, to facilitate comparison with the spectral estimation methods developed in those chapters.

Broadband ARMA Process: Generate realizations of the broadband autoregressive moving-average (ARMA) process

$$y(t) = \frac{B_1(z)}{A_1(z)} \, e(t)$$

with

$$A_1(z) = 1 - 1.3817z^{-1} + 1.5632z^{-2} - 0.8843z^{-3} + 0.4096z^{-4}$$
$$B_1(z) = 1 + 0.3544z^{-1} + 0.3508z^{-2} + 0.1736z^{-3} + 0.2401z^{-4}$$

Choose the number of samples as $N = 256$.

(a) Generate 50 Monte Carlo data realizations, using different noise sequences, and compute the corresponding 50 spectral estimates by using the following methods:
  - The Blackman–Tukey spectral estimate using the Bartlett window $w_B(t)$. Try both $M = N/4$ and $M = N/16$.
  - The Welch spectral estimate using the rectangular window $w_R(t)$, and using both $M = N/4$ and $M = N/16$ and overlap parameter $K = M/2$.
    Plot the sample mean, the sample mean plus one sample standard deviation, and sample mean minus one sample standard deviation spectral estimate curves. Compare with the periodogram results from Exercise C2.21 and with each other.

(b) Judging from the plots you have obtained, how has the variance decreased in the refined estimates? How does this variance decrease as compared with the theoretical expectations?

(c) As discussed in the text, the value of M should be chosen to compromise between low "smearing" and low variance. For the Blackman–Tukey estimate, experiment with different values of M and different window functions to find a "best design" (in your judgment), and plot the corresponding spectral estimates.

Narrowband ARMA Process: Generate realizations of the narrowband ARMA process

$$y(t) = \frac{B_2(z)}{A_2(z)} \, e(t)$$

with

$$A_2(z) = 1 - 1.6408z^{-1} + 2.2044z^{-2} - 1.4808z^{-3} + 0.8145z^{-4}$$

$$B_2(z) = 1 + 1.5857z^{-1} + 0.9604z^{-2}$$

and $N = 256$.

Repeat the experiments and comparisons in the broadband example for the narrowband process.

Exercise C2.23: Periodogram-Based Estimators Applied to Measured Data
Consider the data sets in the files sunspotdata.mat and lynxdata.mat. These files can be obtained from the text website www.prenhall.com/stoica. Apply periodogram-based estimation techniques (possibly after some preprocessing as in part (b)) to estimate the spectral content of these data. Try to answer the following questions:

(a) Are there sinusoidal components (or periodic structure) in the data? If so, how many components, and at what frequencies?
(b) Nonlinear transformations and linear or polynomial trend removal are often applied before spectral analysis of a time series. For the lynx data, compare your spectral analysis results from the original data with those from the data transformed first by taking the logarithm of each sample and then by subtracting the sample mean of this logarithmic data. Does the logarithmic transformation make the data more sinusoidal in nature?