# 7

# Multitaper Spectral Estimation

## 7.0 Introduction

In Chapter 6 we introduced the important concept of tapering a time series as a way of obtaining a spectral estimator with acceptable bias properties. While tapering does reduce bias due to leakage, there is a price to pay in that the sample size is effectively reduced. When we also smooth across frequencies, this reduction translates into a loss of information in the form of an increase in variance (recall the $C_h$ factor in Equation (248b) and Table 248). This inflated variance is acceptable in some practical applications, but in other cases it is not. The loss of information inherent in tapering can often be avoided either by prewhitening (see Sections 6.5 and 9.10) or by using Welch's overlapped segment averaging (WOSA – see Section 6.17).

In this chapter we discuss another approach to recovering information lost due to tapering. This approach was introduced in a seminal paper by Thomson (1982) and involves the use of multiple orthogonal tapers. As we shall see, multitaper spectral estimation has a number of interesting points in its favor:

[1] In contrast to either prewhitening which typically requires the careful design of a prewhitening filter or the conventional use of WOSA (i.e., a Hanning data taper with 50% overlap of blocks) which can still suffer from leakage for spectra with very high dynamic ranges, the multitaper scheme can be used in a fairly 'automatic' fashion. Hence it is useful in situations where thousands – or millions – of individual time series must be processed so that the pure volume of data precludes a careful analysis of individual series (this occurs routinely in exploration geophysics).

[2] The bias of multitaper spectral estimators can be broken down into two quantifiable components: the local bias (due to frequency components within a user-selectable passband $[-W, W]$) and the broad-band bias (due to components outside this passband).

[3] There is a natural nonsubjective way to define the *resolution* of multitaper estimators (it is just the bandwidth $2W$ of the passband) – this is a rather sticky concept to define for spectral estimators in general (Kay and Demeure, 1984).

[4] The tradeoff between the bias and variance for multitaper spectral estimators is also relatively easy to quantify.

[5] Although either prewhitening or WOSA can usually compensate for bias just as well as multitapering can, Thomson (1990a, p. 614) argues that there are examples of processes with highly structured sdf's for which prewhitening is inferior to multitapering; on the other hand, Bronez (1992) shows, using bounds for bias, variance and resolution, that, if a WOSA spectral estimator and a multitaper estimator are fixed so that two of the three bounds are identical, the remaining bound favors the multitaper estimator.

[6] The multitaper approach can be regarded as a way of producing direct spectral estimates with more than just 2 degrees of freedom (typically 4 to 10). Figure 257 shows that such modest increases are enough to noticeably shrink the width of a 95% confidence interval for an sdf at a fixed frequency.

[7] It is possible to obtain an 'internal' estimate of the variance of a multitaper spectral estimator by means of a technique called 'jackknifing' (Thomson and Chave, 1991).

[8] As we shall discuss in Section 10.11, there is an appealing statistical test based upon the multitaper spectral estimator for the presence of sinusoidal (line) components in a time series. The multitaper technique thus offers a unified approach to estimating both mixed spectra and sdf's (processes with mixed spectra were defined in Section 4.4 and are discussed in detail in Chapter 10).

[9] There are natural extensions to the multitaper approach for both spectral estimation of time series with missing or irregularly sampled observations and the problem of estimating multidimensional spectra (Bronez, 1985, 1986, 1988; Liu and Van Veen, 1992).

The remainder of this chapter is organized as follows. In Section 7.1 we motivate the multitaper approach by presenting a computational recipe for it and then illustrating how it works on the AR(4) time series in the lower plot of Figure 45. We investigate in Section 7.2 the statistical properties of the multitaper spectral estimator for the simple case of white noise. We formally justify the multitaper approach in Section 7.3 by showing how it arises in the context of quadratic spectral estimators. We present a second argument for the technique in Section 7.4 based

upon projections and the concept of regularization. We conclude with an example in Section 7.5.

In addition to the papers by Thomson and Bronez we have already cited, the reader should consult Park *et al.* (1987b), Walden (1990b) and Mullis and Scharf (1991) for introductory discussions on multitapering. This estimation technique has been used on a number of interesting time series, including ones concerning terrestrial free oscillations (Park *et al.*, 1987a, and Lindberg and Park, 1987), the relationship between atmospheric carbon dioxide and global temperature (Kuo *et al.* 1990), oxygen isotope ratios from deep-sea cores (Thomson, 1990b) and the rotation of the earth (King, 1990). The relationship between structured covariance matrices and multitapering is discussed in Van Veen and Scharf (1990).

## 7.1 A Motivating Example

Here we introduce the multitaper spectral estimator $\hat{S}^{(mt)}(\cdot)$ by sketching the rationale for it and by discussing in detail the steps that must be taken to compute it. As an example, we compute $\hat{S}^{(mt)}(\cdot)$ for the AR(4) time series shown in the bottom of Figure 45. A more detailed discussion of the rationale for multitaper estimators is given in Sections 7.2 to 7.4.

Suppose we have a time series that is a realization of a portion $X_1, X_2, \ldots, X_N$ of the stationary process $\{X_t\}$ with zero mean, variance $\sigma^2$ and sdf $S(\cdot)$. We assume that the sampling interval between observations is $\Delta t$ so that the Nyquist frequency is $f_{(N)} \equiv 1/(2\,\Delta t)$. For the sample size $N$, the *fundamental Fourier frequency* is defined to be $1/(N\,\Delta t)$ (note that this is just the smallest nonzero Fourier frequency).

As its name implies, the multitaper spectral estimator utilizes several different data tapers. In its simplest formulation, this estimator is the average of $K$ direct spectral estimators and hence takes the form

$$\hat{S}^{(mt)}(f) \equiv \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_k^{(mt)}(f) \text{ for } \hat{S}_k^{(mt)}(f) \equiv \Delta t \left| \sum_{t=1}^{N} h_{t,k} X_t e^{-i2\pi f t\,\Delta t} \right|^2 , \tag{333}$$

where $\{h_{t,k}\}$ is the data taper for the $k$th direct spectral estimator $\hat{S}_k^{(mt)}(\cdot)$ (as usual, we assume the normalization $\sum_{t=1}^{N} h_{t,k}^2 = 1$; see the discussion about Equation (208a)). In Thomson (1982) the estimator $\hat{S}_k^{(mt)}(\cdot)$ is called the $k$th *eigenspectrum*, but note that it is just a special case of the familiar direct spectral estimator $\hat{S}^{(d)}(\cdot)$ of Equation (206c). For each data taper, we can define an associated spectral window

$$\mathcal{H}_k(f) \equiv \Delta t \left| \sum_{t=1}^{N} h_{t,k} e^{-i2\pi f t\,\Delta t} \right|^2 .$$

From Equation (207a) the first moment properties of $\hat{S}_k^{(mt)}(\cdot)$ are given by

$$E\{\hat{S}_k^{(mt)}(f)\} = \int_{-f_{(N)}}^{f_{(N)}} \mathcal{H}_k(f - f')S(f')\,df'.$$

It follows readily that

$$E\{\hat{S}^{(mt)}(f)\} = \int_{-f_{(N)}}^{f_{(N)}} \overline{\mathcal{H}}(f - f')S(f')\,df' \text{ with } \overline{\mathcal{H}}(f) \equiv \frac{1}{K}\sum_{k=0}^{K-1} \mathcal{H}_k(f). \tag{334a}$$

The function $\overline{\mathcal{H}}(\cdot)$ is the spectral window for the estimator $\hat{S}^{(mt)}(\cdot)$. Because the sidelobe level of $\mathcal{H}_k(\cdot)$ dictates whether or not $\hat{S}_k^{(mt)}(\cdot)$ is approximately free of leakage, the above equation tells us that the preponderance of the $K$ spectral windows must provide good protection against leakage if $\overline{\mathcal{H}}(\cdot)$ is to produce a leakage-free $\hat{S}^{(mt)}(\cdot)$. On the other hand, the reason for averaging the $K$ different eigenspectra is to produce an estimator of $S(f)$ with smaller variance than that of any individual $\hat{S}_k^{(mt)}(f)$. For example, if the $\hat{S}_k^{(mt)}(f)$ terms are pairwise uncorrelated with a common variance, then the variance of $\hat{S}^{(mt)}(f)$ would be smaller than that of $\hat{S}_k^{(mt)}(f)$ by a factor of $1/K$.

We thus need a set of $K$ data tapers such that each one provides good protection against leakage and such that the resulting individual eigenspectra are as nearly uncorrelated as possible. As we shall see, if the sdf $S(\cdot)$ satisfies certain conditions, approximate uncorrelatedness follows if the data tapers are orthogonal in the sense that

$$\sum_{t=1}^{N} h_{t,j}h_{t,k} = 0 \text{ for all } j \neq k.$$

A set of $K$ orthogonal data tapers with good leakage properties is given by portions of the *discrete prolate spheroidal sequences* (dpss) with parameter $W$ and orders $k = 0$ to $K - 1$ if $K$ is chosen to be less than the Shannon number $2NW\,\Delta t$ – recall that $2W$ defines the bandwidth for the concentration problems posed in Section 3.9, for which the zeroth-order dpss is the optimum solution (the reader should note that, in that section, $W$ is expressed in standardized units so that $0 < W < 1/2$, whereas here $W$ has physically meaningful units so that now $0 < W < 1/(2\,\Delta t) = f_{(N)}$). Hence we now define $\{h_{t,k}\}$ to be

$$h_{t,k} \equiv \begin{cases} v_{t-1,k}(N, W), & t = 1, \ldots, N; \\ 0, & \text{otherwise}, \end{cases} \tag{334b}$$

where $\{v_{t,k}(N, W)\}$ is the notation we used in Section 3.9 for the $k$th order dpss. We refer to $\{h_{t,k}\}$ as the *kth order dpss data taper*.

We can now outline a recipe for computing $\hat{S}^{(mt)}(\cdot)$.

[1] We first select the *resolution bandwidth* $2W$. Typically $W$ is taken to be a small multiple $j > 1$ of the fundamental frequency so that $W = j/(N\,\Delta t)$ for, say, $j = 2$, 3 or 4. We note, however, that noninteger multiples are sometimes used (for example, Thomson, 1982, p. 1086, uses 3.5) and that larger values than 4 are sometimes of interest. The value of $W$ is usually expressed indirectly via the duration $\times$ half bandwidth product $NW = j/\Delta t$ (or just $NW = j$ with $\Delta t$ standardized to unity). In the example to follow, we picked $NW = 4/\Delta t$, but a reasonable choice of $W$ must take into account the following tradeoff. If we make $W$ large, the number of tapers with good leakage properties increases – hence we can make $K$ large and decrease the variance of $\hat{S}^{(mt)}(f)$. On the other hand, we shall see that the resolution of $\hat{S}^{(mt)}(\cdot)$ decreases as $W$ increases – if we make $W$ too large, we can inadvertently smear out fine features in the sdf.

[2] With $W$ so specified, we compute the eigenspectra $\hat{S}_k^{(mt)}(\cdot)$ for $k = 0$ to $K - 1$ with $K < 2NW\,\Delta t$. Unless we are dealing with a process that is very close to white noise, these will be all of the eigenspectra with potentially good first moment properties. Additionally, for interest, we compute here the eigenspectrum for $k = K$. For this step we obviously must first compute the dpss data tapers, a subject that is discussed in detail in Chapter 8.

[3] Finally we average together $K$ of these eigenspectra, where $K$ is determined by examining the individual eigenspectra for evidence of leakage.

We note here that there are more sophisticated ways of forming a multitaper spectral estimator than simply averaging the eigenspectra. These involve various weighted averages of the eigenspectra – with the weights in general being both frequency and data dependent – and can be implemented in a fairly automatic fashion. These are introduced briefly in Section 7.4 and discussed in detail in Thomson (1982).

As an example, let us construct multitaper spectral estimates for the AR(4) time series shown in the bottom of Figure 45 (see Section 6.18 for a conventional nonparametric spectral analysis of this series). Here $N = 1024$ and $\Delta t = 1$, and we have set $W$ by letting $NW = 4$; i.e., $W = 4/N = 1/256$. Since $2NW = 8$, we need to compute the eigenspectra of orders $k = 0$ to 6 because these are the ones with potentially good bias properties. Additionally, we compute the eigenspectrum for $k = 7$. The left-hand plots of Figures 336 and 338 show the dpss data tapers $\{h_{t,k}\}$ of orders 0 to 7; the corresponding right-hand plots show the product $\{h_{t,k}X_t\}$ of these tapers with the AR(4) time series. Note carefully what happens to the values at the beginning and end of this time series as $k$ varies from 0 to 7: for $k = 0$, these extremes are severely

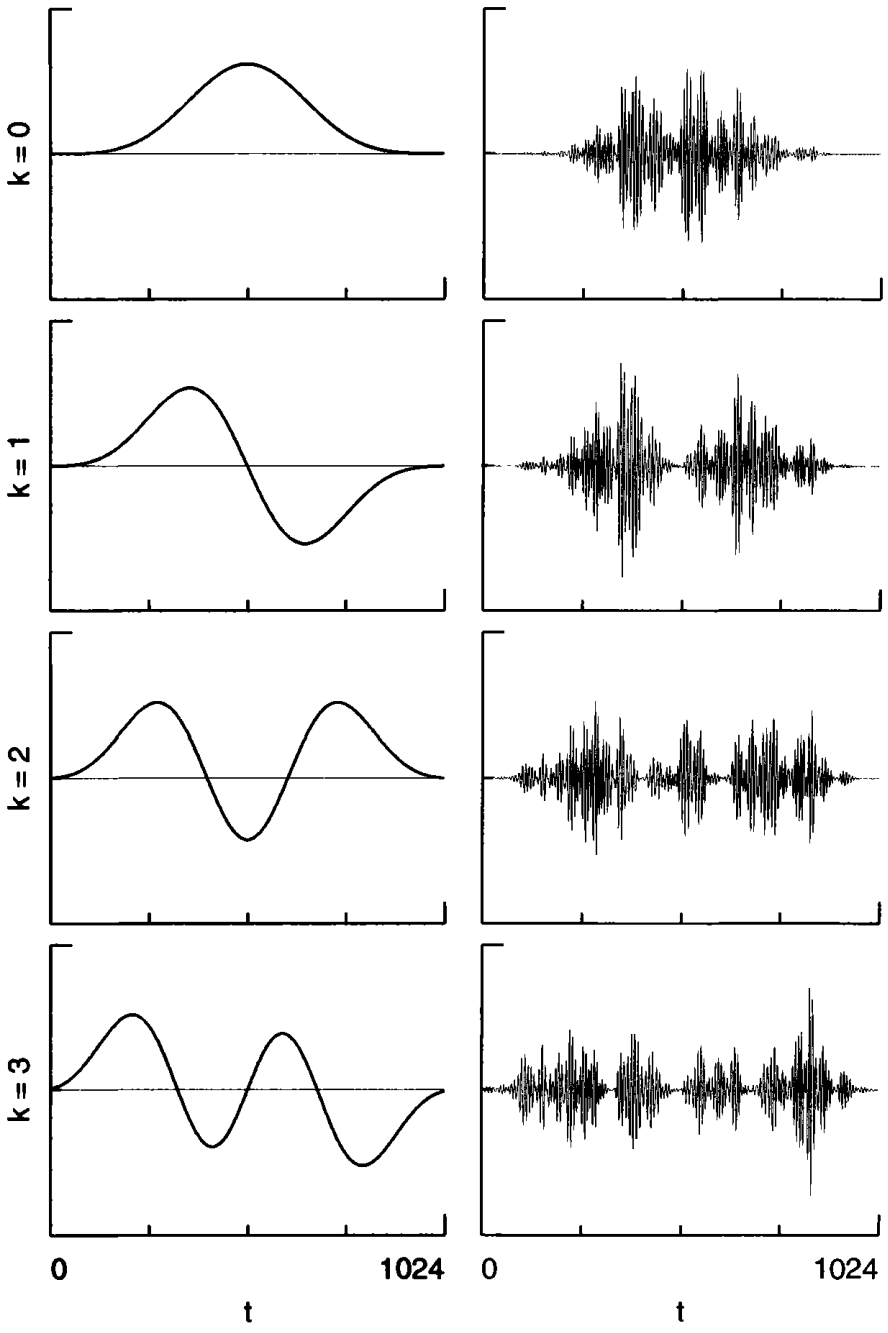


**Figure 336.** Discrete prolate spheroidal sequence data tapers $\{h_{t,k}\}$ (left-hand plots) and product $\{h_{t,k}X_t\}$ of these tapers and a time series $\{X_t\}$ (right-hand plots), part 1. Here $NW = 4$ with $N = 1024$. The orders $k$ for the dpss data tapers are 0, 1, 2 and 3 (top to bottom rows). The time series is the realization of the AR(4) process shown in the bottom of Figure 45.

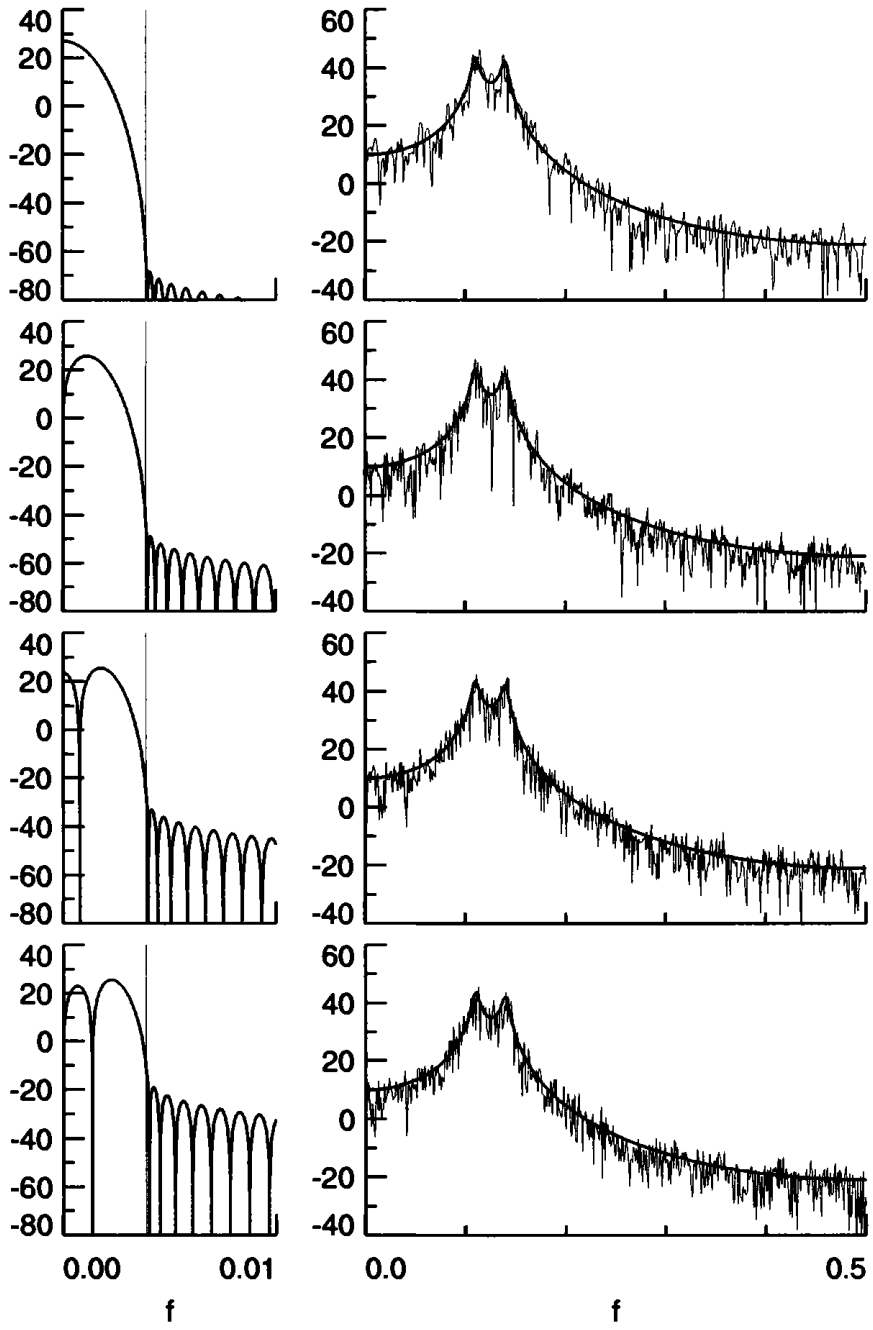


**Figure 337.** Spectral windows $\mathcal{H}_k(\cdot)$ corresponding to $\{h_{t,k}\}$ (left-hand plots) and eigenspectra $\hat{S}_k^{(mt)}(\cdot)$ for AR(4) time series (right-hand plots, thin curves), part 1. The orders $k$ are the same as in Figure 336. The true sdf for the AR(4) process is the thick curve in each right-hand plot. The vertical scale for all plots is in decibels.

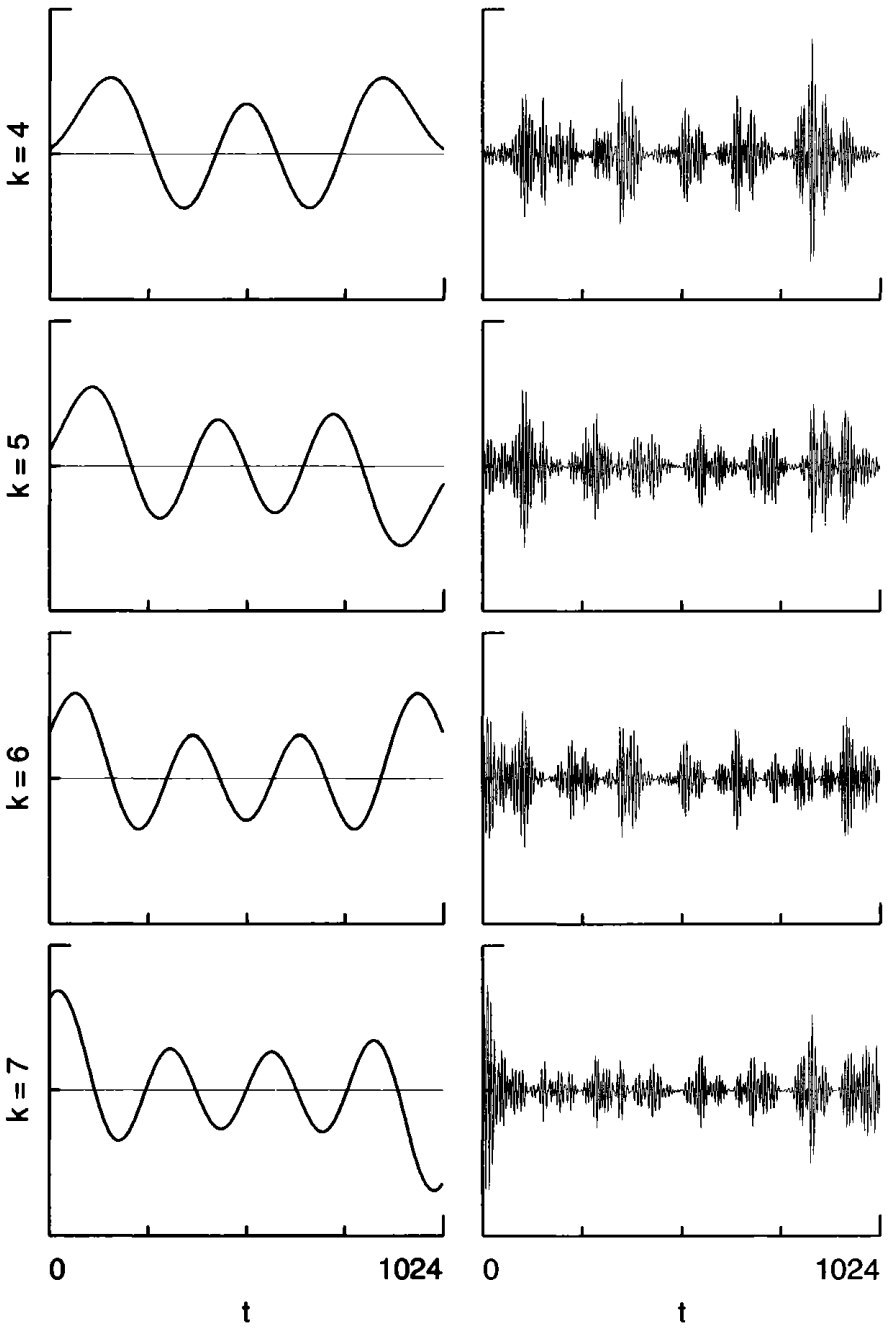


**Figure 338.** Discrete prolate spheroidal sequence data tapers and product of these tapers and a time series, part 2 (cf. Figure 336). Here the orders $k$ for the tapers are 4, 5, 6 and 7 (top to bottom rows). Note that, for $k = 0$ in the top row of Figure 336, the extremes of the time series are severely attenuated, but that, as $k$ increases, portions of the extremes are accentuated.

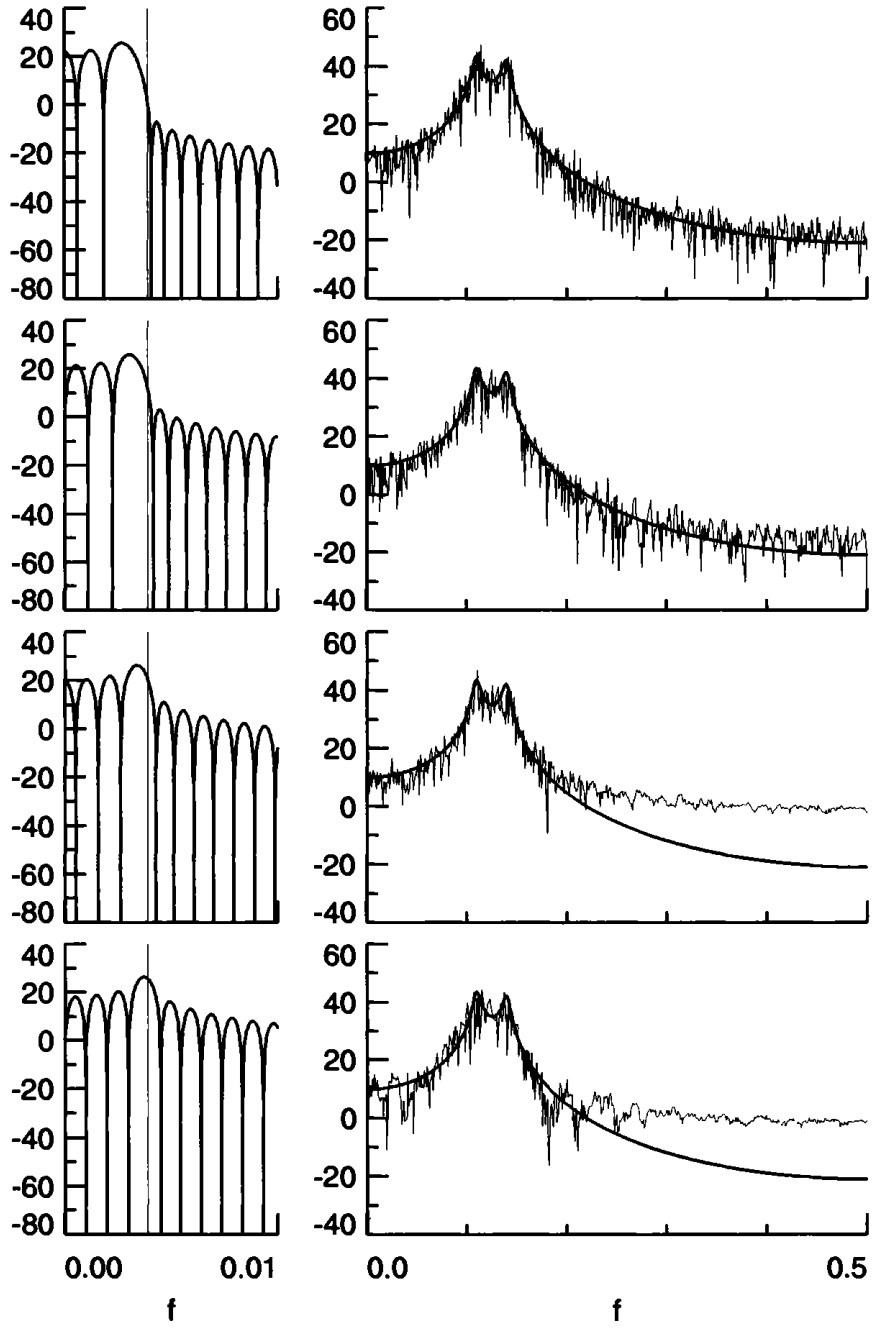


**Figure 339.** Spectral windows $\mathcal{H}_k(\cdot)$ corresponding to $\{h_{t,k}\}$ (left-hand plots) and eigenspectra $\hat{S}_k^{(mt)}(\cdot)$ for AR(4) time series (right-hand plots, thin curves), part 2. The orders $k$ are the same as in Figure 338. Note the marked increase in bias in the eigenspectra as $k$ increases.

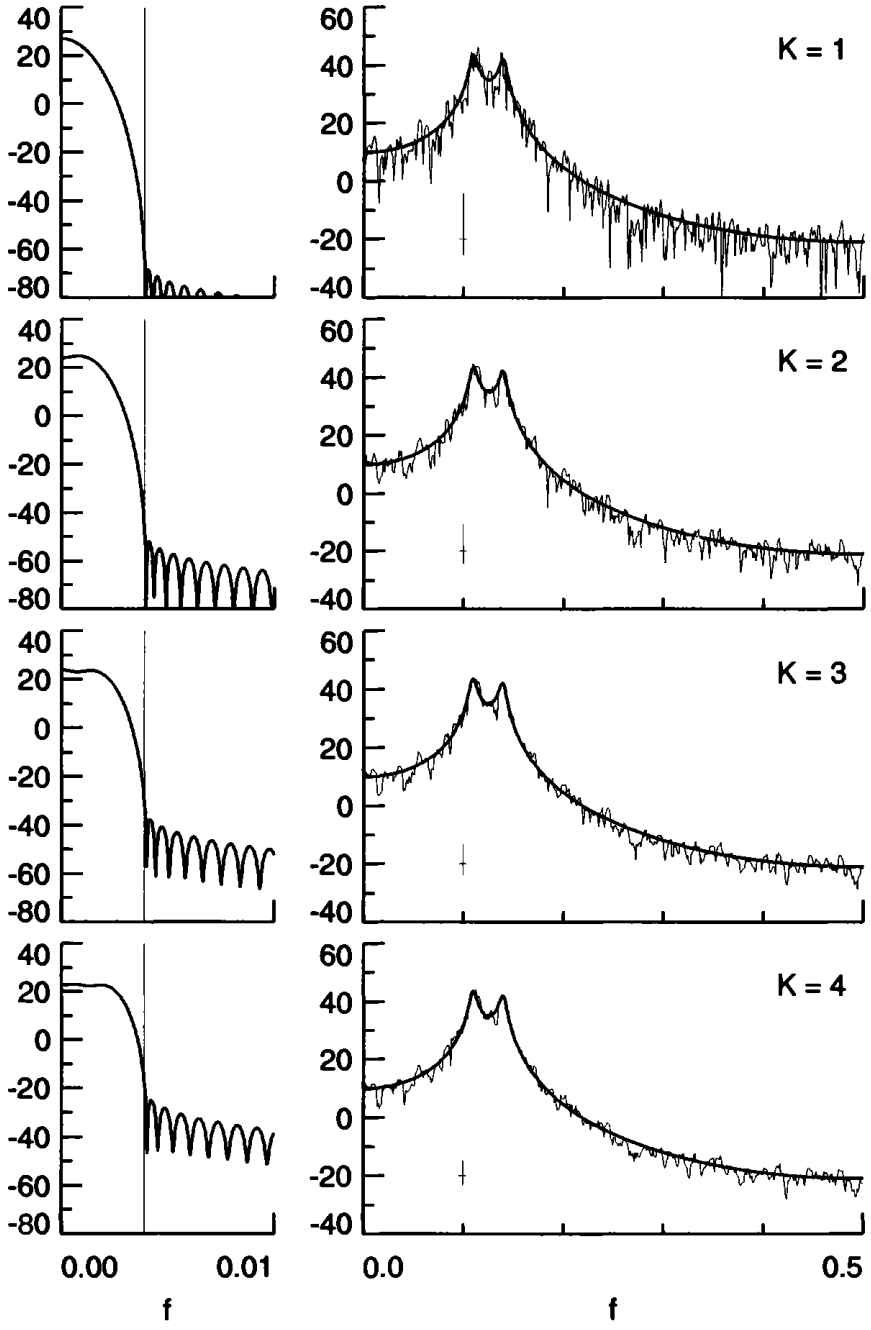


**Figure 340.** Multitaper spectral estimates $\hat{S}^{(mt)}(\cdot)$ formed by averaging $K$ eigenspectra (right-hand plots) and corresponding spectral windows $\overline{\mathcal{H}}(\cdot)$ (left-hand plots) for $K = 1$, 2, 3 and 4. The thin vertical lines on the left-hand plots indicate the frequency $W = 4/N = 1/256$. The vertical scale for all plots is in decibels.

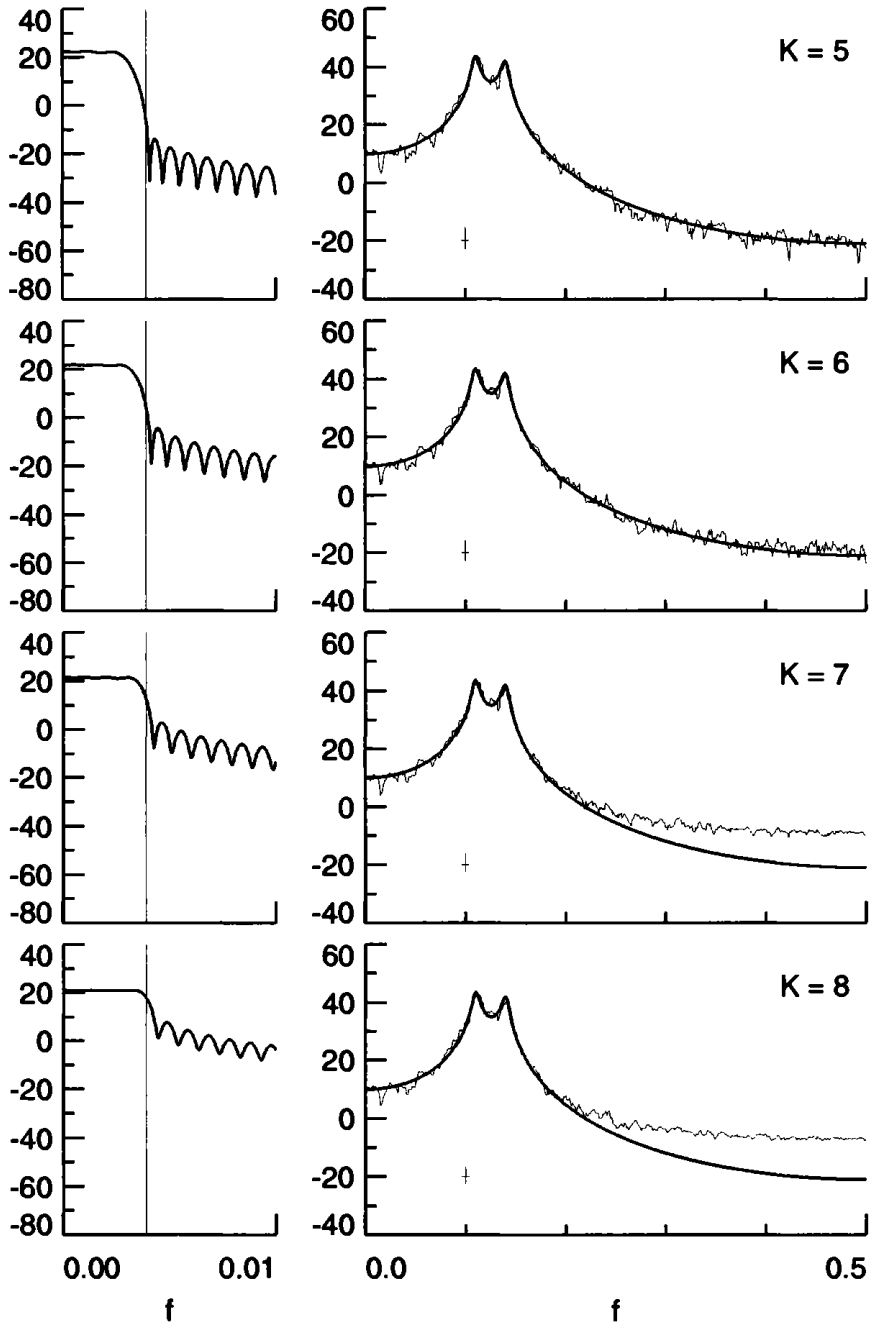


**Figure 341.** Multitaper spectral estimates $\hat{S}^{(mt)}(\cdot)$ formed by averaging $K$ eigenspectra (right-hand plots) and corresponding spectral windows $\overline{\mathcal{H}}(\cdot)$ (left-hand plots) for $K = 5$, 6, 7 and 8. Note the increasing discrepancy at high frequencies between $\hat{S}^{(mt)}(\cdot)$ and the true sdf as $K$ gets large.

attenuated, but, as $k$ increases, the attenuation becomes less and less until the extremes are actually *accentuated* for $k = 7$. One interpretation of the multitapering scheme is that the higher order tapers pick up 'information' that is lost by just using the zeroth-order dpss taper alone (this point is discussed in more detail in item [1] of the Comments and Extensions below).

The left-hand plots of Figures 337 and 339 show the low frequency portion of the spectral windows $\mathcal{H}_k(\cdot)$ for $k = 0$ to 7. The thin vertical line in each plot indicates the location of the frequency $W = 1/256 \approx 0.004$. Note that, as $k$ increases, the level of the sidelobes of $\mathcal{H}_k(\cdot)$ also increases until at $k = 7$ the main sidelobe level is just barely below the lowest lobe in $[-W, W]$. In fact, the spectral windows for dpss data tapers of higher order than 7 resemble the squared modulus of a band-pass filter with a center frequency outside the interval $[-W, W]$ – see item [2] of the Comments and Extensions below. Use of these tapers would yield direct spectral estimates $\hat{S}_k^{(mt)}(f)$ whose expectation is controlled by values of the true sdf *outside* $[f - W, f + W]$!

The jagged curves in the right-hand plots of Figures 337 and 339 show the eigenspectra $\hat{S}_k^{(mt)}(\cdot)$. The thick smooth curve in each of these plots shows the true sdf for the AR(4) process. Note that, whereas the eigenspectra of orders $k = 0$ to 3 (Figure 337) show no indication of leakage, those of order 4 and higher (Figure 339) show increasing evidence of leakage at high frequencies. For example, the $k = 4$ estimate (top right-hand plot of Figure 339) is about 5 dB above the true sdf for frequencies in the range 0.3 to 0.5, whereas the $k = 7$ estimate is elevated there by about 20 dB (had we not known the true sdf, we could have discovered this leakage by comparison with the eigenspectra of orders 0 to 4). The lack of homogeneity of variance across frequencies in the $k = 6$ and 7 eigenspectra is another indication of leakage in the high frequencies (see the discussion in item [4] of the Comments and Extensions to Section 6.4).

Let us now consider the multitaper spectral estimates $\hat{S}^{(mt)}(\cdot)$ and corresponding spectral windows $\overline{\mathcal{H}}(\cdot)$ for different choices of $K$, the number of eigenspectra to be averaged together. Figures 340 and 341 show these two functions for $K = 1$ up to 8 (the case $K = 1$ corresponds to using just a single data taper, so the two plots in the first row of Figure 340 are identical to the ones in the first row of Figure 337). The thick curves in the left-hand plots are the low frequency portions of the spectral windows, while – as before – the thin vertical lines mark the frequency $W = 1/256$. The jagged curves in the right-hand plots are the corresponding multitaper spectral estimates, and the thick smooth curve in each of these plots is the true sdf. Note that, as $K$ increases, the level of the sidelobes in $\overline{\mathcal{H}}(\cdot)$ increases (as we shall see in Sections 7.3 and 7.4, this is related to an increase in what we refer to there as broad-band

bias); note also that $\overline{\mathcal{H}}(\cdot)$ becomes noticeably closer to being constant over frequencies less than $W$ in magnitude (this is related to a decrease in the – yet to be defined – local bias).

Let us now study how well the $\hat{S}^{(mt)}(\cdot)$ in the right-hand plots do in terms of bias and variance. For $K = 1$ up to 5, we see that there is little evidence of any bias in $\hat{S}^{(mt)}(\cdot)$ at any frequency; for $K = 6$ there is evidence of a small bias of about 3 dB for frequencies between 0.4 and 0.5; and, for $K = 7$ and 8, there is significant bias in the high frequencies. On the other hand, as $K$ increases, we see that the variability in $\hat{S}^{(mt)}(\cdot)$ steadily decreases. We can quantify this variability in the following way. Since each eigenspectrum is simply a direct spectral estimate, we know from Equation (223b) that, if we exclude the zero and Nyquist frequencies, $\hat{S}_k^{(mt)}(f)$ is approximately distributed as a rescaled chi-square rv with 2 degrees of freedom, with the scaling factor being $S(f)/2$. If the different eigenspectra are approximately independent (as the theory in the sections following suggests they should be), then the multitaper spectral estimator $\hat{S}^{(mt)}(f)$ should approximately follow a chi-square distribution with $2K$ degrees of freedom and a scaling factor of $S(f)/(2K)$. Just as we did using the lag window spectral estimates in Section 6.10, we can use this approximation to compute a, say, 95% confidence interval for $10 \log_{10}(S(f))$ based upon $10 \log_{10}(\hat{S}^{(mt)}(f))$. The form of the confidence interval is the same as that shown in Equation (258) with $\hat{S}^{(mt)}(f)$ substituted for $\hat{S}^{(lw)}(f)$, $\nu$ set equal to $2K$, and $\lambda$ set to $10 \log_{10}(e) = 4.343$. The vertical height of the crisscross in the lower left-hand portion of the plots with $\hat{S}^{(mt)}(\cdot)$ shows the width of such a 95% confidence interval. Note that, as $K$ increases, this height decreases and that, except in the high frequency regions of $\hat{S}^{(mt)}(\cdot)$ for $K = 7$ and 8 where bias is dominant, the amount of variability in $\hat{S}^{(mt)}(\cdot)$ is roughly consistent with these heights. The horizontal width of each crisscross is just $2W$, a natural measure of the bandwidth of $\hat{S}^{(mt)}(\cdot)$ (this measure is more appropriate for large $K$).

To conclude, it appears that, for this particular example, the multitaper spectral estimator with $K = 5$ data tapers gives us the best compromise between good bias and variance properties. Since the variations in the true sdf are all on a larger scale than our selected bandwidth $2W$ (compare the width of the crisscrosses with the true sdf's in Figure 340 or 341), we could obtain a $\hat{S}^{(mt)}(\cdot)$ with similar bias properties and reduced variance by increasing $W$ by a factor of, say, 2 so that $NW = 8$. Finally, we note that this particular time series could be handled just as well using either prewhitening (all stationary AR processes can be prewhitened *perfectly*, a nicety that never happens with real data!) or a WOSA spectral estimator (see the bottom plot of Figure 316; this estimate is approximately bias free, has 7.6 degrees of freedom, and hence is roughly comparable to the $K = 4$ multitaper estimate in the lower

right-hand plot of Figure 340).

### Comments and Extensions to Section 7.1

[1] We present here support for the statement that multitapering is a scheme for recovering information usually lost when but a single data taper is used (Walden, 1990b). In Section 3.9 we found that the vector

$$\begin{aligned}\mathbf{v}_k(N,W) &\equiv [v_{0,k}(N,W), v_{1,k}(N,W), \ldots, v_{N-1,k}(N,W)]^T \\ &= [h_{1,k}, h_{2,k}, \ldots, h_{N,k}]^T\end{aligned}$$

is the $k$th order eigenvector for the $N \times N$ real symmetric matrix $A$ with $(t',t)$th element given by $\sin[2\pi W(t'-t)]/[\pi(t'-t)]$. Now the matrix $V \equiv [\mathbf{v}_0(N,W), \mathbf{v}_1(N,W), \ldots, \mathbf{v}_{N-1}(N,W)]$ whose columns are these eigenvectors is an orthogonal matrix. Since each eigenvector is assumed to be normalized to have a unit sum of squares, we have $V^TV = I$, where $I$ is the $N \times N$ identity matrix. Because the transpose of $V$ is its inverse, it follows that $VV^T = I$ also. This yields the interesting result that

$$\sum_{k=0}^{N-1} h_{t,k}^2 = 1 \text{ for } t = 1, \ldots, N;$$

i.e., the sum over all orders $k$ of the squares of the $t$th element of each taper is unity. The energy from all $N$ tapered series $\{h_{t,k}X_t\}$ is thus

$$\sum_{k=0}^{N-1}\sum_{t=1}^{N}(h_{t,k}X_t)^2 = \sum_{t=1}^{N} X_t^2 \sum_{k=0}^{N-1} h_{t,k}^2 = \sum_{t=1}^{N} X_t^2;$$

i.e., the total energy using all $N$ possible tapers equals the energy in the time series. Figure 345 shows the build-up of $\sum_{k=0}^{K-1} h_{t,k}^2$ for $t = 1, \ldots, N$ as $K$ increases from 1 to 8 (as before, $N = 1024$, and $W$ is set such that $NW = 4$). Loosely speaking, the plot for a particular $K$ shows the relative influence of $X_t$ – as $t$ varies – on the multitaper spectral estimator $\hat{S}^{(mt)}(\cdot)$ formed using $K$ eigenspectra. The $K = 1$ estimator (i.e., a single dpss taper) is not influenced much by values of $\{X_t\}$ near $t = 1$ or 1024. As $K$ increases to 7, however, the amount of data in the extremes that has little influence decreases steadily, supporting the claim that multitapering is picking up information from the extremes of a time series that is lost with use of a single data taper. Increasing $K$ further to 8 shows a breakdown in which certain values near the extremes start to have *greater* influence than the central portion of the time series (see the case $K = 9$ in the bottom plot of Figure 346).

[2] The previous figures in this chapter refer to the dpss data tapers $\{h_{t,k}\}$ of orders $k = 0$ to 7 for $N = 1024$ and $NW = 4$. For this particular choice of $W$, these are all the tapers whose associated spectral

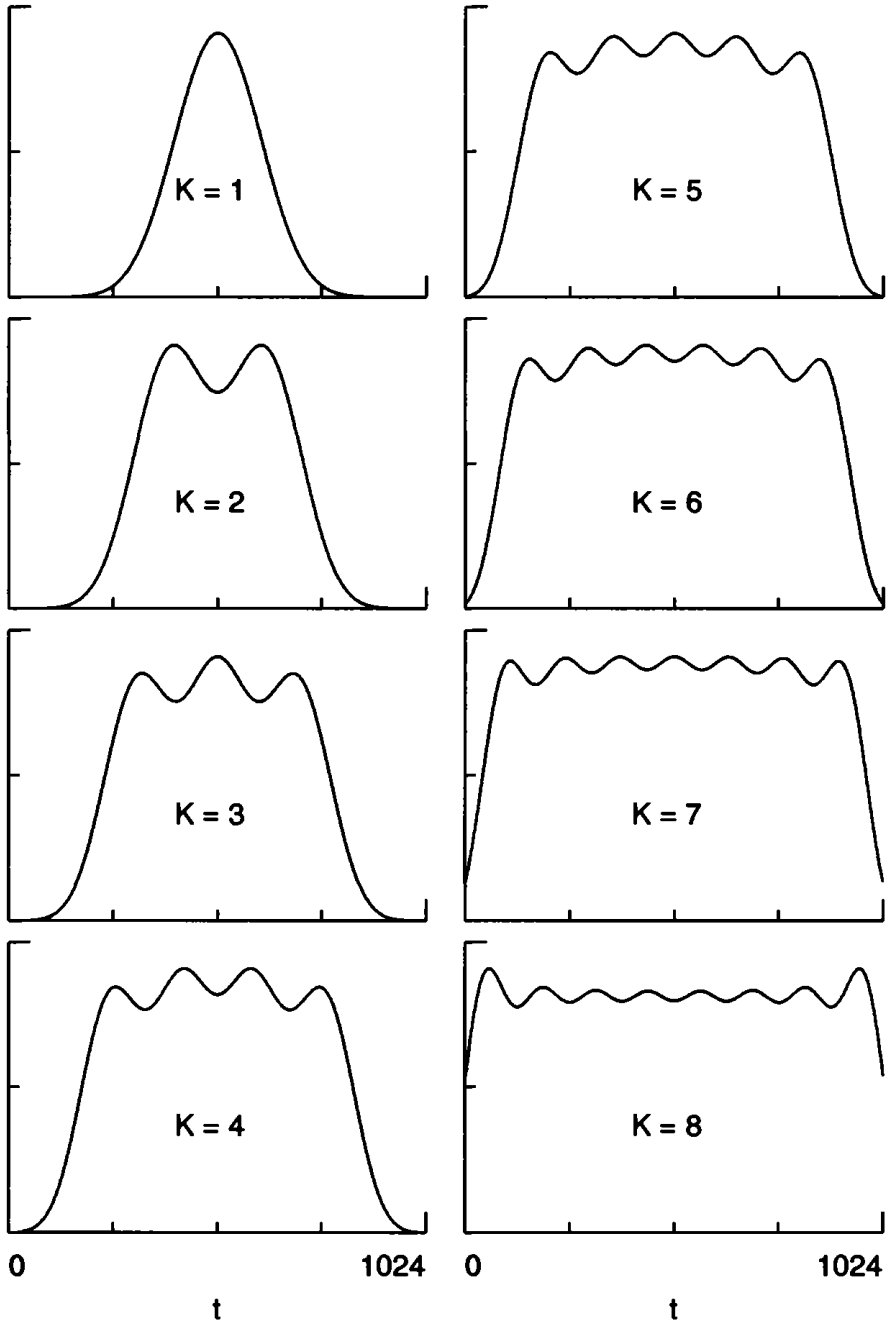


**Figure 345.** Decomposition of dpss taper energy across $t$ as $K$ – the number of tapers used in the multitapering scheme – increases (the curve on each plot is normalized such that its peak value occurs at the same relative height).

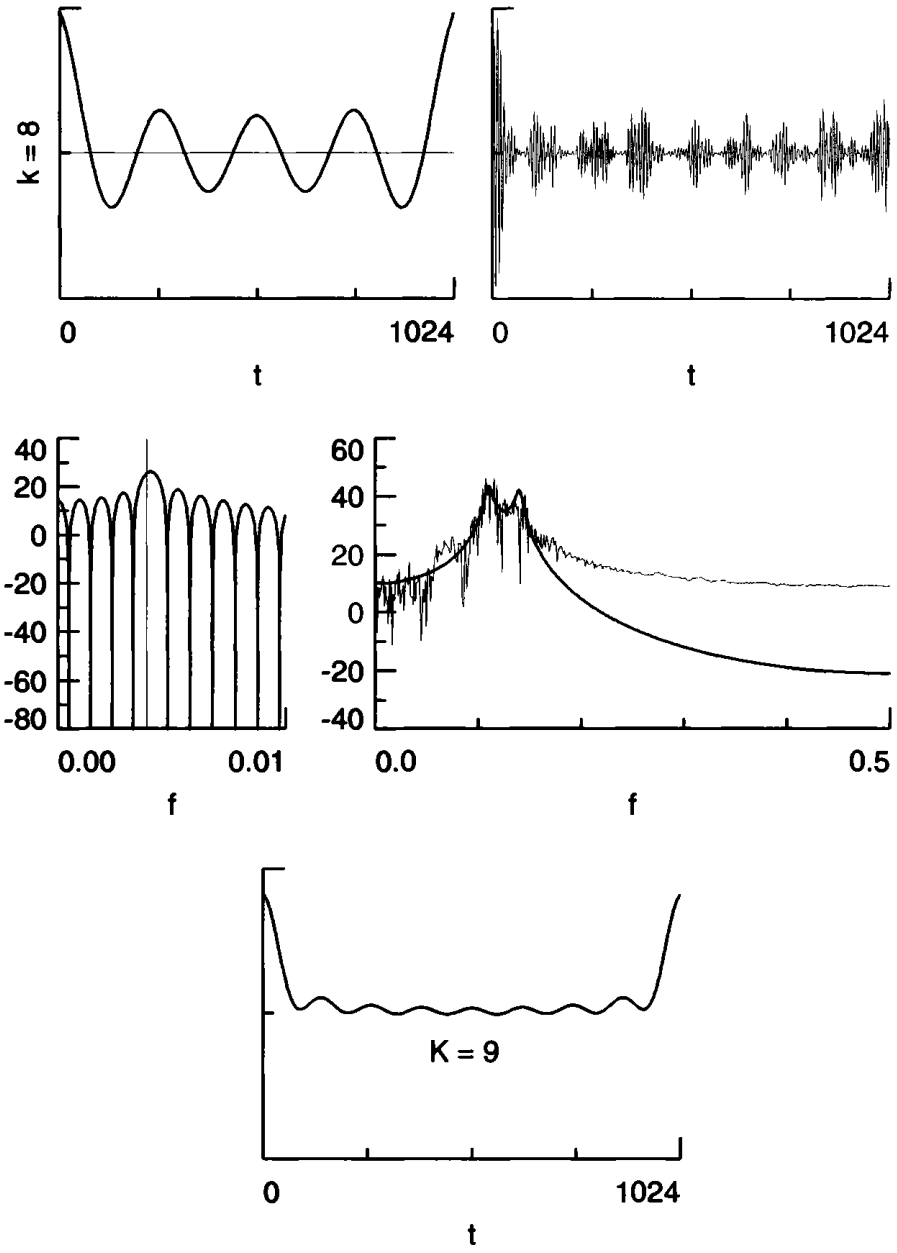


**Figure 346.** Plots for $k = 8$ dpss data taper (see text for details).

windows $\mathcal{H}_k(\cdot)$ have the majority of their energy concentrated in the frequency interval $[-W, W]$; i.e., we have

$$\frac{\int_{-W}^{W} \mathcal{H}_k(f)\, df}{\int_{-f_{(N)}}^{f_{(N)}} \mathcal{H}_k(f)\, df} = \lambda_k(N, W) > 1/2 \text{ for } k = 0, \ldots, 7,$$

where $\lambda_k(N,W)$ is the $k$th order eigenvalue defined in Section 3.9 (the normalization $\sum_{t=1}^{N} h_{t,k}^2 = 1$ implies that $\int_{-f_{(N)}}^{f_{(N)}} \mathcal{H}_k(f)\,df = 1$, so the denominator in the ratio above is just unity). The eigenvalue $\lambda_k(N,W)$ is thus a useful measure of concentration of $\mathcal{H}_k(\cdot)$ over the frequency range $[-W,W]$. For our particular example, we have the following values:

$$\begin{aligned}
\lambda_0(N,W) &= 0.999\,999\,999\,7 & \lambda_4(N,W) &= 0.999\,4\\
\lambda_1(N,W) &= 0.999\,999\,97 & \lambda_5(N,W) &= 0.993\\
\lambda_2(N,W) &= 0.999\,998\,8 & \lambda_6(N,W) &= 0.94\\
\lambda_3(N,W) &= 0.999\,97 & \lambda_7(N,W) &= 0.70.
\end{aligned}$$

For the sake of comparison, Figure 346 shows corresponding plots for the eighth-order dpss data taper, for which $\lambda_8(N,W) = 0.30$ (for $k > 8$, the concentration measures $\lambda_k(N,W)$ are all close to zero – see Figure 110). The two plots in the top row show the taper itself and the product of the taper and the AR(4) series. Note, in particular, the accentuation of the extremes of the series. The left-hand plot in the middle row shows that the peak value of the spectral window $\mathcal{H}_8(\cdot)$ occurs just *outside* the frequency interval $[-W,W]$; the right-hand plot shows the poor leakage properties of $\hat{S}_8^{(mt)}(\cdot)$. Finally, as noted previously, the plot on the last row shows that the distribution of energy over $t$ is skewed to accentuate the extremes of $\{X_t\}$ when $K = 9$ tapers are used to form $\hat{S}^{(mt)}(\cdot)$.

## 7.2 Multitapering of Gaussian White Noise

The statistical properties of the multitaper estimator $\hat{S}^{(mt)}(\cdot)$ are quite easy to derive in the case of Gaussian white noise and serve to illustrate the decrease in variability that multitapering affords. Accordingly, for this section only, let us assume that the time series we are dealing with is a realization of a portion $X_1, X_2, \ldots, X_N$ of a Gaussian white noise process with zero mean and unknown variance $s_0$. The sdf is thus just $S(f) = s_0\,\Delta t$. Standard statistical theory suggests that the best (in a number of different senses) estimator of $s_0$ for this model is just the sample variance

$$\frac{1}{N}\sum_{t=1}^{N} X_t^2 = \hat{s}_0^{(p)}.$$

It follows that the best estimator for the sdf is just $\hat{s}_0^{(p)}\,\Delta t$. Equation (278b) tells us that

$$\text{var}\,\{\hat{s}_0^{(p)}\,\Delta t\} = \frac{2s_0^2\,(\Delta t)^2}{N}. \qquad (347)$$

As we discussed in Section 6.3, the periodogram $\hat{S}^{(p)}(f)$ is an unbiased estimator of $S(f)$ in the case of white noise, and hence a data taper is certainly not needed to control bias. Moreover, because $S(\cdot)$ is flat, we can smooth $\hat{S}^{(p)}(\cdot)$ with a rectangular smoothing window of width $2f_{(N)}$ and height $1/(2f_{(N)}) = \Delta t$ to obtain

$$\int_{-f_{(N)}}^{f_{(N)}} \hat{S}^{(p)}(f)\,\Delta t\,df = \hat{s}_0^{(p)}\,\Delta t$$

(this follows from Exercise [6.6a]). We can thus easily recover the best spectral estimator $\hat{s}_0^{(p)}\,\Delta t$ by averaging $\hat{S}^{(p)}(\cdot)$ over $[-f_{(N)}, f_{(N)}]$. On the other hand, suppose that we use any nonrectangular taper $\{\tilde{h}_{t,0}\}$ – with the usual normalization $\sum_{t=1}^{N} \tilde{h}_{t,0}^2 = 1$ – to produce the direct spectral estimator $\hat{S}^{(d)}(\cdot)$. If we smooth this estimator, we obtain

$$\int_{-f_{(N)}}^{f_{(N)}} \hat{S}^{(d)}(f)\,\Delta t\,df = \sum_{t=1}^{N} \tilde{h}_{t,0}^2 X_t^2\,\Delta t = \hat{s}_0^{(d)}\,\Delta t.$$

Now

$$\text{var}\,\{\hat{s}_0^{(d)}\,\Delta t\} = (\Delta t)^2 \sum_{t=1}^{N} \text{var}\,\{\tilde{h}_{t,0}^2 X_t^2\} = 2s_0^2\,(\Delta t)^2 \sum_{t=1}^{N} \tilde{h}_{t,0}^4$$

since we have $\text{var}\,\{X_t^2\} = 2s_0^2$ from the Isserlis theorem (Equation (40)). The Cauchy inequality, namely,

$$\left|\sum_{t=1}^{N} a_t b_t\right|^2 \le \sum_{t=1}^{N} |a_t|^2 \sum_{t=1}^{N} |b_t|^2,$$

with $a_t = \tilde{h}_{t,0}^2$ and $b_t = 1$, tells us that $\sum_{t=1}^{N} \tilde{h}_{t,0}^4 \ge 1/N$ with equality holding if and only if $\tilde{h}_{t,0} = \pm 1/\sqrt{N}$ (i.e., $\{\tilde{h}_{t,0}\}$ is – to within an innocuous change of sign – the rectangular data taper). We can conclude that $\text{var}\,\{\hat{s}_0^{(p)}\,\Delta t\} < \text{var}\,\{\hat{s}_0^{(d)}\,\Delta t\}$ for any nonrectangular data taper. Use of a single data taper on white noise yields a smoothed spectral estimator with increased variance and thus amounts to throwing away a certain portion of the data (Sloane, 1969).

We now show that we can compensate for this loss through multitapering. Accordingly, let $\{\tilde{h}_{t,0}\}$, $\{\tilde{h}_{t,1}\}$, ..., $\{\tilde{h}_{t,N-1}\}$ be *any* set of orthonormal data tapers, each of length $N$ (note that these need not be dpss tapers). If we let $\tilde{V}$ be the $N \times N$ matrix whose columns are formed from these tapers, i.e.,

$$\tilde{V} \equiv \begin{bmatrix} \tilde{h}_{1,0} & \tilde{h}_{1,1} & \cdots & \tilde{h}_{1,N-1} \\ \tilde{h}_{2,0} & \tilde{h}_{2,1} & \cdots & \tilde{h}_{2,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{h}_{N,0} & \tilde{h}_{N,1} & \cdots & \tilde{h}_{N,N-1} \end{bmatrix},$$

orthonormality says that $\tilde{V}^T\tilde{V} = I$, where $I$ is the $N \times N$ identity matrix. Since the transpose of $\tilde{V}$ is thus its inverse, we also have $\tilde{V}\tilde{V}^T = I$; i.e.,

$$\sum_{k=0}^{N-1} \tilde{h}_{t,k}\tilde{h}_{u,k} = \begin{cases} 1, & \text{if } t = u; \\ 0, & \text{otherwise} \end{cases} \tag{349a}$$

(see item [1] of the Comments and Extensions to the previous section). Let

$$\tilde{S}_k^{(mt)}(f) \equiv \Delta t \left| \sum_{t=1}^{N} \tilde{h}_{t,k} X_t e^{-i2\pi f t\,\Delta t} \right|^2 \tag{349b}$$

be the $k$th direct spectral estimator, and consider the multitaper estimator formed by averaging all $N$ of the $\tilde{S}_k^{(mt)}(f)$:

$$\tilde{S}^{(mt)}(f) \equiv \frac{1}{N} \sum_{k=0}^{N-1} \tilde{S}_k^{(mt)}(f).$$

We can write

$$\begin{aligned} \tilde{S}^{(mt)}(f) &= \frac{\Delta t}{N} \sum_{k=0}^{N-1} \left( \sum_{t=1}^{N} \tilde{h}_{t,k} X_t e^{-i2\pi f t\,\Delta t} \right) \left( \sum_{u=1}^{N} \tilde{h}_{u,k} X_u e^{i2\pi f u\,\Delta t} \right) \\ &= \frac{\Delta t}{N} \sum_{t=1}^{N} \sum_{u=1}^{N} X_t X_u \left( \sum_{k=0}^{N-1} \tilde{h}_{t,k}\tilde{h}_{u,k} \right) e^{-i2\pi f (t-u)\,\Delta t}. \end{aligned}$$

From Equation (349a), the term in the parentheses is unity if $t = u$ and is zero otherwise, so we obtain

$$\tilde{S}^{(mt)}(f) = \frac{\Delta t}{N} \sum_{t=1}^{N} X_t^2 = \hat{s}_0^{(p)}\,\Delta t. \tag{349c}$$

A multitaper spectral estimator using $N$ orthogonal data tapers is thus identical to the best spectral estimator for a white noise process, so multitapering effectively restores the information normally lost when but a single taper is used.

We note that the above argument is valid for *any* set of orthogonal data tapers. We can get away with such a cavalier choice only in the very special case of white noise because then any data taper yields an unbiased direct spectral estimator. For nonwhite processes, this is certainly not the case – we must then insist upon a set of orthogonal tapers, at least some of which provide decent leakage protection. With

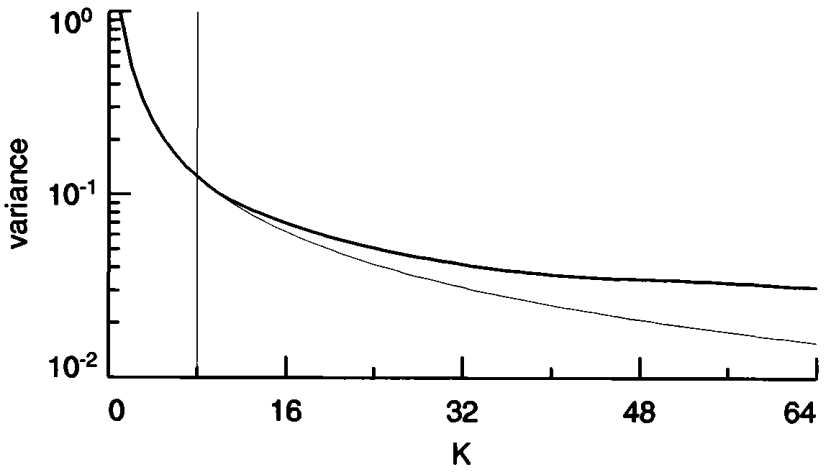


**Figure 350.** Decrease in $\mathrm{var}\{\hat{S}^{(mt)}(f)\}$ for $f = 1/4$ as the number $K$ of eigenspectra used to form the multitaper spectral estimator increases from 1 up to the sample size $N = 64$ (thick curve). Here a Gaussian white noise process is assumed with unit variance and a sampling time $\Delta t = 1$. The data tapers are dpss tapers with $NW = 4$. The thin curve indicates the decrease in variance to be expected if the eigenspectra were all uncorrelated. The thin vertical line marks the Shannon number $2NW = 8$.

this restriction, we are led (via the concentration measure) to the dpss data tapers as a natural set of choice.

Finally, let us study the rate at which the variance of the multitaper estimator decreases as the number of eigenspectra $K$ that are averaged together increases. To do so, we now assume that we use the dpss data tapers $\{h_{t,k}\}$ to form the eigenspectra. The variance of $\hat{S}^{(mt)}(f)$ can be expressed as follows (see Exercise [2.8]):

$$\begin{aligned}\mathrm{var}\{\hat{S}^{(mt)}(f)\} &= \mathrm{var}\left\{\frac{1}{K}\sum_{k=0}^{K-1}\hat{S}_k^{(mt)}(f)\right\} \\ &= \frac{1}{K^2}\sum_{j=0}^{K-1}\sum_{k=0}^{K-1}\mathrm{cov}\{\hat{S}_j^{(mt)}(f), \hat{S}_k^{(mt)}(f)\}.\end{aligned}$$

Exercise [7.1b] indicates how to compute the covariance terms in the above expression for a Gaussian white noise process. The exercise also shows that $\mathrm{var}\{\hat{S}^{(mt)}(f)\}$ depends on $f$. As a specific example, let us set $f = 1/4$. The thick curve in Figure 350 shows $\mathrm{var}\{\hat{S}^{(mt)}(1/4)\}$ versus the number of eigenspectra $K$ for $N = 64$, $NW = 4$, $\Delta t = 1$ and $s_0 = 1$. Note that, under these conditions, Equation (223c) says that $\mathrm{var}\{\hat{S}^{(mt)}(1/4)\} \approx 1$ for $K = 1$ because $S(1/4) = s_0 \Delta t = 1$, while Equations (349c) and (347) tell us that $\mathrm{var}\{\hat{S}^{(mt)}(1/4)\} = 2/64 \approx$

0.03 for $K = N = 64$. The extremes of the thick curve agree with these values. The thin curve shows $\text{var}\{\hat{S}^{(mt)}(f)\}$ versus $K$ under the incorrect assumption that $\text{cov}\{\hat{S}_j^{(mt)}(f), \hat{S}_k^{(mt)}(f)\}$ is zero for all $j \neq k$. The thin vertical line marks the Shannon number $2NW = 8$. We see that, for $K \leq 8$, the variance decreases at a rate consistent with $\text{cov}\{\hat{S}_j^{(mt)}(f), \hat{S}_k^{(mt)}(f)\} = 0$ to a good approximation for all $j \neq k$ such that $0 \leq j, k < 8$. As $K$ increases beyond the Shannon number, the rate of decrease in variance becomes slower than that implied by uncorrelatedness of the $\hat{S}_k^{(mt)}(f)$ terms. For nonwhite processes, this result suggests that the modest decrease in the variance of $\hat{S}^{(mt)}(f)$ afforded by including high-order eigenspectra will be readily offset by the poor bias properties of these eigenspectra.

## 7.3 Quadratic Spectral Estimators and Multitapering

We now present two complementary approaches to justifying the multitaper spectral estimator. The first is due to Bronez (1985, 1988) and is based upon a study of the first and second moments of quadratic spectral estimators (defined below); the second is closely related to the integral equation approach in Thomson's 1982 paper and is discussed in the next section.

Suppose that $X_1, \ldots, X_N$ is a segment of length $N$ of a real-valued stationary process with zero mean and sdf $S(\cdot)$. For a fixed frequency $f$ such that $0 \leq f \leq f_{(N)}$, let us define the complex-valued process

$$Z_t \equiv X_t e^{-i2\pi f t \Delta t}.$$

The results of Exercise [5.7a] tell us that $\{Z_t\}$ is a stationary process with zero mean and sdf given by

$$S_Z(f') = S(f + f') \text{ for } |f'| \leq f_{(N)}$$

(recall that $S(\cdot)$ is a periodic function with period $2f_{(N)}$). In particular, $S_Z(0) = S(f)$, so we can obtain an estimate of $S(f)$ via an estimate of $S_Z(\cdot)$ at zero frequency.

Let $\mathbf{Z}$ be an $N$-dimensional column vector whose $t$th element is $Z_t$, and let $\mathbf{Z}^H$ be its Hermitian transpose; i.e., $\mathbf{Z}^H \equiv (Z_1^*, \ldots, Z_N^*)$, where the asterisk denotes complex conjugation (for what follows, the reader should note that, if $A$ is a matrix with real-valued elements, its Hermitian transpose $A^H$ is equal to its usual transpose $A^T$). If $X_t$ has units of, say, meters and if the sampling interval $\Delta t$ is, say, 1 second, then the sdf $S(\cdot)$ has units of (meters)$^2$ per Hertz or, equivalently, seconds $\times$ (meters)$^2$/cycle. Since $S(\cdot)$ is real-valued, an estimator for $S(f)$ with the correct dimensionality is

$$\hat{S}^{(q)}(f) \equiv \hat{S}_Z^{(q)}(0) \equiv \Delta t \sum_{s=1}^{N} \sum_{t=1}^{N} Z_s^* Q_{s,t} Z_t = \Delta t \, \mathbf{Z}^H Q \mathbf{Z}, \qquad (351)$$

where $Q_{s,t}$ is the $(s,t)$th element of the $N \times N$ matrix $Q$ of weights. For simplicity, we assume that $Q_{s,t}$ is real-valued; with this proviso, there is no loss of generality in assuming that $Q$ is a symmetric matrix (since $\mathbf{Z}^H Q \mathbf{Z} = \mathbf{Z}^H \left(Q + Q^H\right) \mathbf{Z}/2$ with $Q + Q^H$ being necessarily symmetric). We assume that $Q_{s,t}$ does not depend on $\{Z_t\}$. An estimator of form (351) is known in the literature as a *quadratic estimator*. Note that, if the matrix $Q$ is positive semidefinite, we have the desirable property $\hat{S}^{(q)}(f) \geq 0$.

We have already met an important example of a quadratic estimator in Chapter 6, namely, the lag window spectral estimator

$$\hat{S}^{(lw)}(f) \equiv \Delta t \sum_{\tau=-(N-1)}^{N-1} w_{\tau,m} \hat{s}_\tau^{(d)} e^{-i2\pi f \tau \, \Delta t}, \tag{352a}$$

where $\{w_{\tau,m}\}$ is a lag window;

$$\hat{s}_\tau^{(d)} \equiv \begin{cases} \sum_{t=1}^{N-|\tau|} h_t X_t h_{t+|\tau|} X_{t+|\tau|}, & \tau = 0, \pm 1, \ldots, \pm(N-1); \\ 0, & \text{otherwise;} \end{cases}$$

and $\{h_t\}$ is a data taper. Since Equation (352a) can be rewritten as

$$\begin{aligned} \hat{S}^{(lw)}(f) &= \Delta t \sum_{s=1}^{N} \sum_{t=1}^{N} h_s X_s h_t X_t w_{t-s,m} e^{-i2\pi f (t-s) \, \Delta t} \\ &= \Delta t \sum_{s=1}^{N} \sum_{t=1}^{N} Z_s^* h_s w_{t-s,m} h_t Z_t, \end{aligned}$$

it follows that $\hat{S}^{(lw)}(f)$ can be expressed as a quadratic estimator with $Q_{s,t} = h_s w_{t-s,m} h_t$. Since a lag window spectral estimator can sometimes be negative (consider the reshaped lag window estimate of Figure 270(c)), it follows that the matrix $Q$ for such an estimator need not be positive semidefinite. By letting $w_{\tau,m} = 1$ for all $\tau$, a lag window spectral estimator reduces to a direct spectral estimator

$$\hat{S}^{(d)}(f) \equiv \Delta t \left| \sum_{t=1}^{N} h_t X_t e^{-i2\pi f t \, \Delta t} \right|^2, \tag{352b}$$

for which $Q_{s,t} = h_s h_t$. Since a direct spectral estimator can never be negative, the matrix $Q$ for such an estimator must be positive semidefinite. The WOSA spectral estimator $\hat{S}^{(\text{WOSA})}(\cdot)$ of Equation (291b) is another example of a quadratic estimator whose $Q$ matrix is necessarily positive semidefinite.

We want to find a quadratic estimator $\hat{S}^{(q)}(f)$ for $S(f)$ with good bias and variance properties. To ensure the condition $\hat{S}(f) \geq 0$, we insist

from now on that $Q$ in Equation (351) be positive semidefinite. If $Q$ were positive definite, its rank would be $N$; since we are only assuming it to be positive semidefinite, its rank – call it $K$ – can be taken to satisfy $1 \le K \le N$ (the case $K = 0$ is uninteresting). Based upon some standard results in linear algebra (see Exercise [7.2]), we can decompose $Q$ as

$$Q = AA^T = AA^H,$$

where $A$ is a real-valued $N \times K$ matrix such that $A^T A$ is a $K \times K$ diagonal matrix; i.e., if we let the $N$-dimensional column vector $\mathbf{a}_k$ represent the $k$th column of $A$, we have $\mathbf{a}_j^T \mathbf{a}_k = 0$ for $j \neq k$ so that the columns of $A$ are mutually orthogonal. We can now write

$$\begin{aligned}
\hat{S}^{(q)}(f) &= \Delta t \, \mathbf{Z}^H A A^T \mathbf{Z} \qquad (353a)\\
&= \Delta t \, \mathbf{Z}^H \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_K \end{bmatrix} \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_K^T \end{bmatrix} \mathbf{Z} \\
&= \Delta t \begin{bmatrix} \mathbf{Z}^H \mathbf{a}_1 & \mathbf{Z}^H \mathbf{a}_2 & \dots & \mathbf{Z}^H \mathbf{a}_K \end{bmatrix} \begin{bmatrix} \mathbf{a}_1^T \mathbf{Z} \\ \mathbf{a}_2^T \mathbf{Z} \\ \vdots \\ \mathbf{a}_K^T \mathbf{Z} \end{bmatrix} \\
&= \Delta t \sum_{k=1}^{K} \mathbf{Z}^H \mathbf{a}_k \mathbf{a}_k^T \mathbf{Z} = \Delta t \sum_{k=1}^{K} (\mathbf{a}_k^T \mathbf{Z})^* \mathbf{a}_k^T \mathbf{Z} = \Delta t \sum_{k=1}^{K} \left| \mathbf{a}_k^T \mathbf{Z} \right|^2 \\
&= \Delta t \sum_{k=1}^{K} \left| \sum_{t=1}^{N} a_{t,k} Z_t \right|^2 = \frac{\Delta t}{K} \sum_{k=0}^{K-1} \left| \sum_{t=1}^{N} \tilde{h}_{t,k} X_t e^{-i2\pi f t \, \Delta t} \right|^2 ,
\end{aligned}$$

where $a_{t,k}$ is the $t$th element of $\mathbf{a}_k$, and we define $\tilde{h}_{t,k} \equiv a_{t,k+1}\sqrt{K}$. Note that the above can be written as

$$\hat{S}^{(q)}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_k^{(q)}(f) \text{ with } \hat{S}_k^{(q)}(f) \equiv \Delta t \left| \sum_{t=1}^{N} \tilde{h}_{t,k} X_t e^{-i2\pi f t \, \Delta t} \right|^2 . \qquad (353b)$$

We have thus the important result that all quadratic estimators with a real-valued, symmetric, positive semidefinite matrix $Q$ of weights can be written as an average of $K$ direct spectral estimators $\hat{S}_k^{(q)}(f)$ ($K$ being the rank of $Q$), the $k$th one of which uses $\{\tilde{h}_{t,k}\}$ as its data taper; moreover, because the $\mathbf{a}_k$ vectors are mutually orthogonal, it follows that the data tapers $\{\tilde{h}_{t,j}\}$ and $\{\tilde{h}_{t,k}\}$ are also orthogonal for $j \neq k$.

We need expressions for the expected value of $\hat{S}^{(q)}(f)$. It follows

from Equations (353b) and (207a) that

$$E\{\hat{S}^{(q)}(f)\} = \frac{1}{K}\sum_{k=0}^{K-1} E\{\hat{S}_k^{(q)}(f)\} = \frac{1}{K}\sum_{k=0}^{K-1}\int_{-f_{(N)}}^{f_{(N)}} \widetilde{\mathcal{H}}_k(f-f')S(f')\,df'$$
$$= \int_{-f_{(N)}}^{f_{(N)}} \widetilde{\mathcal{H}}(f-f')S(f')\,df', \quad (354a)$$

where

$$\widetilde{\mathcal{H}}_k(f) \equiv \Delta t \left|\sum_{t=1}^{N} \tilde{h}_{t,k} e^{-i2\pi f t\,\Delta t}\right|^2 \text{ and } \widetilde{\mathcal{H}}(f) \equiv \frac{1}{K}\sum_{k=0}^{K-1} \widetilde{\mathcal{H}}_k(f). \quad (354b)$$

The function $\widetilde{\mathcal{H}}(\cdot)$ is the spectral window of the quadratic spectral estimator $\hat{S}^{(q)}(\cdot)$, while $\widetilde{\mathcal{H}}_k(\cdot)$ is the spectral window of the $k$th direct spectral estimator.

It is also convenient to have a 'covariance domain' expression for $E\{\hat{S}^{(q)}(f)\}$. If we let $\{s_\tau\}$ represent the acvs for $\{X_t\}$, it follows from Exercise [5.7a] that the acvs $\{s_{\tau,Z}\}$ for $\{Z_t\}$ is given by $s_{\tau,Z} = s_\tau e^{-i2\pi f\tau\,\Delta t}$. If we let $\Sigma_Z$ be the $N \times N$ covariance matrix for $Z_1$, $\ldots$, $Z_N$ (i.e., its $(j,k)$th element is given by $s_{j-k,Z}$), it follows from Exercise [7.3] that

$$E\{\hat{S}^{(q)}(f)\} = \Delta t\ \mathrm{tr}\{Q\Sigma_Z\} = \Delta t\ \mathrm{tr}\{AA^T\Sigma_Z\} = \Delta t\ \mathrm{tr}\{A^T\Sigma_Z A\}, \quad (354c)$$

where $\mathrm{tr}\{\cdot\}$ is the matrix trace operator. If we equate the above to Equation (354a), we obtain

$$\int_{-f_{(N)}}^{f_{(N)}} \widetilde{\mathcal{H}}(f-f')S(f')\,df' = \Delta t\ \mathrm{tr}\{A^T\Sigma_Z A\}. \quad (354d)$$

We can make substantial progress in obtaining a quadratic estimator with quantifiably good first moment properties for a wide class of sdf's if we adopt a slightly different definition of what we mean by 'good.' If we were to use the same approach as in Chapter 6, we would take this to mean $E\{\hat{S}^{(q)}(f)\} \approx S(f)$. Here we adopt a criterion that incorporates the important notion of *resolution*: for a selected $W > 0$, we want

$$E\{\hat{S}^{(q)}(f)\} \approx \frac{1}{2W}\int_{f-W}^{f+W} S(f')\,df' \equiv \overline{S}(f), \quad (354e)$$

where $2W > 0$ defines a *resolution bandwidth*. Note that $\overline{S}(f)$ is the average value of $S(\cdot)$ over the interval from $f - W$ to $f + W$ (see Equation (86)).

The requirement that (354e) holds rather than just $E\{\hat{S}^{(q)}(f)\} \approx S(f)$ can be justified from two points of view. First, we noted in Section 5.6 that one interpretation of $S(f)$ is that $2S(f)\,df$ is approximately the variance of the process created by passing $\{X_t\}$ through a sufficiently narrow band-pass filter. If we were to actually use such a filter with bandwidth $2W$ and center frequency $f$, the spectral estimate we would obtain via the sample variance of the output from the filter would be more closely related to $\overline{S}(f)$ than to $S(f)$. The requirement $E\{\hat{S}^{(q)}(f)\} \approx \overline{S}(f)$ thus incorporates the limitations imposed by the filter bandwidth. Second, because we must base our estimate of the function $S(\cdot)$ on the finite sample $X_1, \ldots, X_N$, we face an inherently ill-posed problem in the sense that, with a finite number of observations, we cannot hope to determine $S(\cdot)$ over an infinite number of frequencies unless $S(\cdot)$ is sufficiently smooth. Since $\overline{S}(\cdot)$ is a smoothed version of $S(\cdot)$, it should be easier to find an approximately unbiased estimator for $\overline{S}(f)$ than for $S(f)$; moreover, if $S(\cdot)$ is itself sufficiently smooth, then $\overline{S}(\cdot)$ and $S(\cdot)$ should be approximately equal for $W$ small enough. In effect, we are redefining the function to be estimated from one that need not be smooth, namely, $S(\cdot)$, to one that is guaranteed to be smooth to a certain degree, namely, $\overline{S}(\cdot)$.

With these comments in mind, let us see what constraints we need to impose on the matrix $A$ in Equation (353a) so that $\hat{S}^{(q)}(f)$ is an approximately unbiased estimator of $\overline{S}(f)$, i.e., that $E\{\hat{S}^{(q)}(f)\} = \overline{S}(f)$ to some level of approximation. Can we insist upon having

$$E\{\hat{S}^{(q)}(f)\} = \overline{S}(f) = \frac{1}{2W}\int_{f-W}^{f+W} S(f')\,df'$$

for all possible $\overline{S}(\cdot)$? From Equation (354a) we would then need

$$\widetilde{\mathcal{H}}(f') = \begin{cases} 1/(2W), & |f'| \le W; \\ 0, & \text{otherwise.} \end{cases} \tag{355}$$

Since $\widetilde{\mathcal{H}}(\cdot)$ is the average of the $\widetilde{\mathcal{H}}_k(\cdot)$ terms, each one of which is proportional to the modulus squared of the Fourier transform of a finite sequence of values $\{\tilde{h}_{t,k}\}$, we cannot attain the simple form stated in (355) (see the discussion at the beginning of Section 5.8). We must therefore be satisfied with unbiasedness to some level of approximation. The approach that we take is, first, to insist that $\hat{S}^{(q)}(f)$ be an unbiased estimator of $\overline{S}(f)$ for white noise processes and, second, to develop bounds for two components of the bias for colored noise.

If $\{X_t\}$ is a white noise process with variance $\sigma^2$, then $S(f) = \sigma^2\,\Delta t$ for all $f$, so

$$\overline{S}(f) = \frac{1}{2W}\int_{f-W}^{f+W} \sigma^2\,\Delta t\,df' = \sigma^2\,\Delta t.$$

If $\hat{S}^{(q)}(f)$ is to be an unbiased estimator of $\overline{S}(f)$ for white noise, Equation (354a) tells us that

$$E\{\hat{S}^{(q)}(f)\} = \sigma^2 \Delta t \int_{-f_{(N)}}^{f_{(N)}} \widetilde{\mathcal{H}}(f - f')\, df' = \sigma^2 \Delta t,$$

i.e., that

$$\int_{-f_{(N)}}^{f_{(N)}} \widetilde{\mathcal{H}}(f - f')\, df' = \int_{-f_{(N)}}^{f_{(N)}} \widetilde{\mathcal{H}}(f')\, df = 1$$

(see Equation (208c)). Since $\Sigma_Z$ for a white noise process is a diagonal matrix, all of whose elements are equal to $\sigma^2$, it follows from Equation (354d) that the above requirement is equivalent to

$$\mathrm{tr}\,\{A^T A\} = 1.$$

For a colored noise process let us define the bias, $b$, in the estimator $\hat{S}^{(q)}(f)$ as

$$\begin{aligned} b\{\hat{S}^{(q)}(f)\} &\equiv E\{\hat{S}^{(q)}(f)\} - \overline{S}(f) \\ &= \int_{-f_{(N)}}^{f_{(N)}} \widetilde{\mathcal{H}}(f - f')S(f')\, df' - \frac{1}{2W}\int_{f-W}^{f+W} S(f')\, df'. \end{aligned}$$

It is impossible to give any sort of reasonable bound for $b\{\hat{S}^{(q)}(f)\}$ since the second integral above does not depend upon the matrix $A$. It is useful, however, to split the bias into two components, one due to frequencies from $f - W$ to $f + W$ – the *local bias* – and the other, to frequencies outside this interval – the *broad-band bias*. Accordingly, we write

$$b\{\hat{S}^{(q)}(f)\} = b^{(l)}\{\hat{S}^{(q)}(f)\} + b^{(b)}\{\hat{S}^{(q)}(f)\},$$

where

$$\begin{aligned} b^{(l)}\{\hat{S}^{(q)}(f)\} &\equiv \int_{f-W}^{f+W} \widetilde{\mathcal{H}}(f - f')S(f')\, df' - \frac{1}{2W}\int_{f-W}^{f+W} S(f')\, df' \\ &= \int_{f-W}^{f+W} \left(\widetilde{\mathcal{H}}(f - f') - \frac{1}{2W}\right) S(f')\, df' \end{aligned}$$

and

$$b^{(b)}\{\hat{S}^{(q)}(f)\} \equiv \int_{-f_{(N)}}^{f-W} \widetilde{\mathcal{H}}(f - f')S(f')\, df' + \int_{f+W}^{f_{(N)}} \widetilde{\mathcal{H}}(f - f')S(f')\, df'.$$

To obtain useful upper bounds on the local and broad-band bias, let us assume that $S(\cdot)$ is bounded by $S_{\max}$, i.e., that $S(f) \leq S_{\max} < \infty$

for all $f$ in the interval $[-f_{(N)}, f_{(N)}]$. For the local bias we then have

$$\left|b^{(l)}\{\hat{S}^{(q)}(f)\}\right| \le \int_{f-W}^{f+W} \left|\widetilde{\mathcal{H}}(f-f') - \frac{1}{2W}\right| S(f')\,df'$$
$$\le S_{\max} \int_{-W}^{W} \left|\widetilde{\mathcal{H}}(-f'') - \frac{1}{2W}\right| df'',$$

where $f'' \equiv f' - f$. We thus can take the quantity

$$\int_{-W}^{W} \left|\widetilde{\mathcal{H}}(f'') - \frac{1}{2W}\right| df'' \tag{357}$$

as a useful measure of the magnitude of the local bias in $\hat{S}^{(q)}(f)$. Note that we can control the local bias by approximating the unattainable ideal $\widetilde{\mathcal{H}}(\cdot)$ of Equation (355) as closely as possible over the resolution band.

Let us now obtain a bound for the broad-band bias (usually a more important source of concern than the local bias). With $f'' \equiv f' - f$ as before, we have

$$b^{(b)}\{\hat{S}^{(q)}(f)\} \le S_{\max}\left(\int_{-f_{(N)}}^{f-W} \widetilde{\mathcal{H}}(f-f')\,df' + \int_{f+W}^{f_{(N)}} \widetilde{\mathcal{H}}(f-f')\,df'\right)$$
$$= S_{\max}\left(\int_{-f_{(N)}}^{f_{(N)}} \widetilde{\mathcal{H}}(f-f')\,df' - \int_{f-W}^{f+W} \widetilde{\mathcal{H}}(f-f')\,df'\right)$$
$$= S_{\max}\left(\int_{-f_{(N)}}^{f_{(N)}} \widetilde{\mathcal{H}}(f-f')\,df' - \int_{-W}^{W} \widetilde{\mathcal{H}}(-f'')\,df''\right).$$

By considering Equation (354d) for a white noise process with unit variance, we can rewrite the first integral above as

$$\int_{-f_{(N)}}^{f_{(N)}} \widetilde{\mathcal{H}}(f-f')\,df' = \mathrm{tr}\{A^T A\}$$

(this follows because $\Sigma_Z$ then becomes the identity matrix). The second integral can be rewritten using (354d) again, but this time with an sdf $S^{(bl)}(\cdot)$ for band-limited white noise; i.e.,

$$S^{(bl)}(f) = \begin{cases} \Delta t, & |f| \le W; \\ 0, & W < |f| \le f_{(N)}. \end{cases}$$

The corresponding acvs $s_\tau^{(bl)}$ at lag $\tau$ is given by Equation (134a):

$$s_\tau^{(bl)} = \int_{-f_{(N)}}^{f_{(N)}} e^{i2\pi f'\tau\,\Delta t} S^{(bl)}(f')\,df'$$
$$= \int_{-W}^{W} e^{i2\pi f'\tau\,\Delta t}\Delta t\,df' = \frac{\sin(2\pi W\tau\,\Delta t)}{\pi\tau},$$

where, as usual, this ratio is defined to be $2W\,\Delta t$ when $\tau = 0$. If we let $\Sigma^{(bl)}$ be the $N \times N$ matrix whose $(j,k)$th element is $s^{(bl)}_{j-k}$, we obtain from Equation (354d) with $f = 0$

$$\int_{-W}^{W} \widetilde{\mathcal{H}}(-f'')\,df'' = \mathrm{tr}\,\{A^T \Sigma^{(bl)} A\}.$$

We thus have that

$$b^{(b)}\{\hat{S}^{(q)}(f)\} \le S_{\max}\left(\mathrm{tr}\,\{A^T A\} - \mathrm{tr}\,\{A^T \Sigma^{(bl)} A\}\right),$$

so we can take the quantity

$$\mathrm{tr}\,\{A^T A\} - \mathrm{tr}\,\{A^T \Sigma^{(bl)} A\} \tag{358}$$

to be a useful measure of the broad-band bias in $\hat{S}^{(q)}(f)$.

Suppose for the moment that our only criterion for choosing the matrix $A$ is that our broad-band bias measure be made as small as possible, subject to the constraint that the resulting estimator $\hat{S}^{(q)}(f)$ be unbiased for white noise processes. This means that we want to

$$\text{maximize } \mathrm{tr}\,\{A^T \Sigma^{(bl)} A\} \text{ subject to } \mathrm{tr}\,\{A^T A\} = 1.$$

Exercise [7.4] indicates that, if $\Sigma^{(bl)}$ were positive definite with a distinct largest eigenvalue, then a solution to this maximization problem would be to set $A$ equal to a normalized eigenvector associated with the largest eigenvalue of the matrix $\Sigma^{(bl)}$. In fact, $\Sigma^{(bl)}$ satisfies the stated condition: because its $(j,k)$th element is just $\sin[2\pi W(j-k)\,\Delta t]/[\pi(j-k)]$, its eigenvalues and eigenvectors are the solutions to the problem posed previously by Equation (105b) in Section 3.9 (note that $W$ on the left-hand side of (105b) must be replaced by $W\,\Delta t$; however, we use the same notation for the eigenvalues $\lambda_k(N,W)$ as before). From that discussion, we know that all of the eigenvalues are positive (and hence $\Sigma^{(bl)}$ is positive definite) and that there is only one eigenvector corresponding to the largest eigenvalue $\lambda_0(N,W)$. Moreover, the elements of the normalized eigenvector corresponding to $\lambda_0(N,W)$ can be taken to be a finite subsequence of the zeroth-order dpss, namely, $v_{0,0}(N,W)$, $v_{1,0}(N,W)$, $\ldots$, $v_{N-1,0}(N,W)$. Because of the definition for the zeroth-order dpss data taper $\{h_{t,0}\}$ in Equation (334b), the solution to the maximization problem is thus to set $A$ to the $N \times 1$ vector whose elements are $h_{1,0}$, $h_{2,0}$, $\ldots$, $h_{N,0}$, i.e., the nonzero portion of $\{h_{t,0}\}$. If we denote this vector as $\mathbf{h}_0$, the constraint $\mathrm{tr}\,\{\mathbf{h}_0^T \mathbf{h}_0\} = 1$ implies that

$$\sum_{t=1}^{N} h_{t,0}^2 = 1,$$

consistent with the usual normalization for a data taper (see Equation (208a), and recall that the trace of a scalar is just the scalar itself). Note also that, because $\mathbf{h}_0$ is an eigenvector of $\Sigma^{(bl)}$ corresponding to the eigenvalue $\lambda_0(N, W)$, we have

$$\Sigma^{(bl)}\mathbf{h}_0 = \lambda_0(N, W)\mathbf{h}_0.$$

Using this fact, we see that the minimum value of the broad-band bias measure in (358) is simply

$$\begin{aligned}\mathrm{tr}\,\{\mathbf{h}_0^T\mathbf{h}_0\} - \mathrm{tr}\,\{\mathbf{h}_0^T\Sigma^{(bl)}\mathbf{h}_0\} &= 1 - \mathrm{tr}\,\{\mathbf{h}_0^T(\lambda_0(N, W)\mathbf{h}_0)\} \\ &= 1 - \mathrm{tr}\,\{\lambda_0(N, W)\mathbf{h}_0^T\mathbf{h}_0\} \\ &= 1 - \lambda_0(N, W).\end{aligned}$$

In summary, with the constraint $\mathrm{tr}\,\{A^TA\} = 1$, the broad-band bias measure (358) is minimized by setting $A$ equal to the $N$-dimensional vector $\mathbf{h}_0$ whose elements are the nonzero portion of a zeroth-order dpss data taper. The resulting quadratic estimator $\hat{S}^{(q)}(\cdot)$ has a weight matrix $Q = \mathbf{h}_0\mathbf{h}_0^T$ of rank $K = 1$, and the estimator amounts to a direct spectral estimator of the form of Equation (352b). Note that this method of specifying a quadratic estimator obviously does *not* take into consideration either the variance of the estimator or its local bias, both of which we now examine in more detail.

Suppose now that we wish to obtain a quadratic estimator (with a positive semidefinite weight matrix $Q = AA^T$) that has good variance properties, subject to the mild condition that the estimator be unbiased in the case of white noise, i.e., that $\mathrm{tr}\,\{A^TA\} = 1$. From Equation (353b) we know that a rank $K$ quadratic estimator $\hat{S}^{(q)}(f)$ can be written as the average of $K$ direct spectral estimators $\hat{S}_k^{(q)}(f)$ employing a set of $K$ orthonormal data tapers. In Section 7.2 we found that, for Gaussian white noise, the variance of such an estimator is minimized by using *any* set of $K = N$ orthonormal data tapers. This result is in direct conflict with the recommendation we came up with for minimizing our broad-band bias measure, namely, that we use just a single zeroth-order dpss data taper.

An obvious compromise estimator that attempts to balance this conflict between variance and broad-band bias is to use as many members $K$ of the set of orthonormal dpss data tapers as possible. With this choice, our quadratic estimator $\hat{S}^{(q)}(f)$ becomes the multitaper estimator $\hat{S}^{(mt)}(f)$ defined by Equations (333) and (334b). Let us now discuss how this compromise affects the broad-band bias measure, the variance and the local bias measure for the resulting quadratic estimator.

[1] If we use $K$ orthonormal dpss data tapers, the broad-band bias

measure of Equation (358) becomes

$$1 - \frac{1}{K} \sum_{k=0}^{K-1} \lambda_k(N, W) \tag{360}$$

(this is Exercise [7.5]). Because the $\lambda_k(N, W)$ terms are all close to unity as long as we set $K$ to an integer less than the Shannon number $2NW\,\Delta t$, the broad-band bias measure must be close to zero if we impose the restriction $K < 2NW\,\Delta t$.

[2] For Gaussian white noise, we saw in Figure 350 that, for $K \leq 2NW\,\Delta t$, the variance of the average $\hat{S}^{(mt)}(f)$ of $K$ direct spectral estimators $\hat{S}_k^{(mt)}(f)$ is consistent with the approximation that all $K$ of the $\hat{S}_k^{(mt)}(f)$ rv's are pairwise uncorrelated. Since $\hat{S}_k^{(mt)}(f)$ is approximately distributed as $S(f)\chi_2^2/2$ for $k = 0, \ldots, K-1$ (as long as $f$ is not too close to 0 or $f_{(N)}$), this figure suggests that, with $K$ so chosen, the estimator $\hat{S}^{(mt)}(f)$ should be approximately distributed as $S(f)\chi_{2K}^2/2K$. The variance of $\hat{S}^{(mt)}(f)$ is hence approximately $S^2(f)/K$. For colored Gaussian processes, this approximation is still good as long as $S(f)$ does not vary too rapidly over the interval $[f-W, f+W]$ (see the next section or Thomson, 1982, Section IV).

[3] The local bias measure of Equation (357) is small if the spectral window for $\hat{S}^{(mt)}(\cdot)$ is as close to $1/(2W)$ as possible over the resolution band. This window is denoted as $\overline{\mathcal{H}}(\cdot)$ in Equation (334a) and is shown in the left-hand plots of Figures 340 and 341 for $K = 1$ to 8, $N = 1024$ and $W = 4/N$. Since $10 \log_{10}(1/(2W)) = 21$ dB here, these plots indicate that, as $K$ increases, the spectral window $\overline{\mathcal{H}}(\cdot)$ becomes closer to $1/(2W)$ over the resolution band and hence that the local bias decreases as $K$ increases.

In summary, we have shown that all quadratic estimators with a symmetric real-valued positive semidefinite weight matrix $Q$ of rank $K$ can be written as an average of $K$ direct spectral estimators with orthogonal data tapers. The requirement that the quadratic estimator be unbiased for white noise is satisfied if the tapers are in fact orthonormal. We then redefine (or 'regularize') our spectral estimation problem so that the quantity to be estimated is $S(f)$ averaged over an interval of selectable width $2W$ (i.e., $\overline{S}(f)$ of Equation (354e)) rather than just $S(f)$ itself. This redefinition allows us to profitably split the bias of a quadratic estimator into two parts, denoted as the local bias and the broad-band bias. Minimization of a measure of broad-band bias dictates that we set $K = 1$, with the single data taper being a zeroth-order dpss data taper; on the other hand, the variance of a quadratic estimator is minimized in the special case of Gaussian white noise by choosing $K$ to be $N$, but any set of $N$ orthonormal data tapers yields the same minimum value.

The obvious compromise for colored noise is thus to use as many of the low-order members of the family of dpss data tapers as possible, yielding a multitaper spectral estimator $\hat{S}^{(mt)}(f)$. The broad-band bias measure suggests restricting $K$ to be less than the Shannon number $2NW\,\Delta t$. With $K$ thus restricted, the estimator $\hat{S}^{(mt)}(f)$ follows approximately a rescaled $\chi^2_{2K}$ distribution (provided that $f$ is not too close to either 0 or $f_{(N)}$ and that $S(f)$ does not vary too rapidly over the interval $[f - W, f + W]$). The variance of $\hat{S}^{(mt)}(f)$ is inversely proportional to $K$ – increasing $K$ thus decreases the variance. A bound on the magnitude of the local bias also decreases as $K$ increases, so an increase in $K$ above unity improves both the variance and local bias of $\hat{S}^{(mt)}(f)$ at the cost of increasing its broad-band bias. (The reader should consult Bronez (1985, 1988) for an extension of the development in this section to the case of continuous parameter complex-valued stationary processes sampled at irregular time intervals.)

## 7.4 Regularization and Multitapering

In this section we present a rationale for the multitaper spectral estimator that closely follows Thomson's original 1982 approach. It makes extensive use of the spectral representation theorem and presents the spectral estimation problem as a search for a calculable approximation to a desirable quantity that unfortunately cannot be calculated from observable data alone. We begin with a review of some key concepts.

A way of representing any stationary process $\{X_t\}$ with zero mean is given by the spectral representation theorem introduced in Chapter 4:

$$X_t = \int_{-1/2}^{1/2} e^{i2\pi ft}\, dZ(f)$$

(this is Equation (130b); for convenience, we assume that $\Delta t = 1$ in this section, yielding a Nyquist frequency of $f_{(N)} = 1/2$). The increments of the orthogonal process $\{Z(f)\}$ define the sdf of $\{X_t\}$ as

$$S(f)\, df = E\{|dZ(f)|^2\}. \tag{361a}$$

Thomson (1982) suggests – purely for convenience – changing the definition of $dZ(\cdot)$ by a phase factor, which of course does not affect (361a):

$$X_t = \int_{-1/2}^{1/2} e^{i2\pi f[t-(N+1)/2]}\, dZ(f). \tag{361b}$$

Given $X_1, \ldots, X_N$, we want to relate the Fourier transform of these $N$ values to the spectral representation for $\{X_t\}$. For convenience, we shall work with the following phase-shifted Fourier transform:

$$\Upsilon(f) \equiv \sum_{t=1}^{N} e^{-i2\pi f[t-(N+1)/2]} X_t. \tag{361c}$$

Thus, from Equations (361b) and (361c), we have

$$\Upsilon(f) = \int_{-1/2}^{1/2} \left( \sum_{t=1}^{N} e^{-i2\pi(f-f')[t-(N+1)/2]} \right) dZ(f'). \tag{362a}$$

The summation in the parentheses is proportional to Dirichlet's kernel (see Exercise [1.3c]). Hence we arrive at a Fredholm integral equation of the first kind:

$$\Upsilon(f) = \int_{-1/2}^{1/2} \frac{\sin[N\pi(f-f')]}{\sin[\pi(f-f')]} \, dZ(f') \tag{362b}$$

(Thomson, 1982). Since $dZ(\cdot)$ is by no stretch of the imagination a smooth function, it is not possible to solve for $dZ(\cdot)$ using standard inverse theory. We adopt an approach that emphasizes regularization of the spectral estimation problem and uses the same building blocks as Thomson. This leads us to consider projecting the Fourier transform $\Upsilon(\cdot)$ onto an appropriate set of basis functions, to which we now turn.

We introduced the discrete prolate spheroidal wave functions (dpswf's) in Section 3.9, where we denoted them as $U_k(\cdot; N, W)$, but here we simplify the notation to just $U_k(\cdot)$. As indicated in Exercise [3.12],

$$U_k(f) = (-1)^k \epsilon_k \sum_{t=0}^{N-1} v_{t,k}(N,W) e^{-i2\pi f[t-(N-1)/2]},$$

where

$$\epsilon_k \equiv \begin{cases} 1, & \text{if } k \text{ is even;} \\ \sqrt{-1}, & \text{if } k \text{ is odd;} \end{cases}$$

and $\{v_{t,k}(N,W)\}$ is the $k$th order dpss. If we define the $k$th order dpss data taper $\{h_{t,k}\}$ as in Equation (334b), we have

$$U_k(f) = (-1)^k \epsilon_k \sum_{t=1}^{N} h_{t,k} e^{-i2\pi f[t-(N+1)/2]}, \quad k = 0, \ldots, N-1. \tag{362c}$$

If we recall that the $U_k(\cdot)$ functions are orthogonal over $[-W, W]$, i.e.,

$$\int_{-W}^{W} U_j(f) U_k(f) \, df = \begin{cases} \lambda_k, & j = k; \\ 0, & \text{otherwise,} \end{cases} \tag{362d}$$

then the rescaled functions $U_k(\cdot)/\sqrt{\lambda_k}$ are orthonormal:

$$\int_{-W}^{W} \frac{U_j(f)}{\sqrt{\lambda_j}} \frac{U_k(f)}{\sqrt{\lambda_k}} \, df = \begin{cases} 1, & j = k; \\ 0, & \text{otherwise} \end{cases}$$

(here $\lambda_k$ is shorthand notation for $\lambda_k(N,W)$ defined in Section 3.9). Slepian (1978, Section 4.1) shows that the finite-dimensional space of functions of the form of Equation (361c) is spanned over the interval $[-W,W]$ by the rescaled dpswf's $U_k(\cdot)/\sqrt{\lambda_k}$. Hence we can write $\Upsilon(\cdot)$ as

$$\Upsilon(f) = \sum_{k=0}^{N-1} \Upsilon_k \frac{U_k(f)}{\sqrt{\lambda_k}} \text{ for } \Upsilon_k \equiv \int_{-W}^{W} \Upsilon(f) \frac{U_k(f)}{\sqrt{\lambda_k}}\, df$$

(see the discussion in item [2] of the Comments and Extensions to Section 5.1). If, for a given frequency $f'$, we now define a new function $\Upsilon_{f'}(\cdot)$ via $\Upsilon_{f'}(f) \equiv \Upsilon(f+f')$, we can express it as

$$\Upsilon_{f'}(f) = \sum_{k=0}^{N-1} \Upsilon_k(f') \frac{U_k(f)}{\sqrt{\lambda_k}} \text{ for } \Upsilon_k(f') \equiv \int_{-W}^{W} \Upsilon_{f'}(f) \frac{U_k(f)}{\sqrt{\lambda_k}}\, df;$$

i.e., we have

$$\Upsilon(f+f') = \sum_{k=0}^{N-1} \Upsilon_k(f') \frac{U_k(f)}{\sqrt{\lambda_k}} \text{ for } \Upsilon_k(f') = \int_{-W}^{W} \Upsilon(f+f') \frac{U_k(f)}{\sqrt{\lambda_k}}\, df. \tag{363a}$$

If we substitute the expression for $\Upsilon(\cdot)$ given in Equation (361c) into the above (after switching the roles of the variables $f$ and $f'$), we obtain

$$\begin{aligned}\Upsilon_k(f) &= \int_{-W}^{W} \left( \sum_{t=1}^{N} e^{-i2\pi(f'+f)[t-(N+1)/2]} X_t \right) \frac{U_k(f')}{\sqrt{\lambda_k}}\, df' \\ &= \sum_{t=1}^{N} e^{-i2\pi f[t-(N+1)/2]} X_t \int_{-W}^{W} e^{-i2\pi f'[t-(N+1)/2]} \frac{U_k(f')}{\sqrt{\lambda_k}}\, df'.\end{aligned}$$

From Exercise [3.12] we have the following relationship between the $k$th order dpss $\{v_{t,k}(N,W)\}$ and the $k$th order dpswf $U_k(\cdot)$:

$$v_{t-1,k}(N,W) = \frac{(-1)^k}{\epsilon_k \sqrt{\lambda_k}} \int_{-W}^{W} e^{i2\pi f'[t-(N+1)/2]} \frac{U_k(f')}{\sqrt{\lambda_k}}\, df'.$$

Because both $v_{t-1,k}(N,W)$ and $U_k(\cdot)$ are real-valued and because $\epsilon_k^* = 1/\epsilon_k$, we also have

$$v_{t-1,k}(N,W) = \frac{(-1)^k \epsilon_k}{\sqrt{\lambda_k}} \int_{-W}^{W} e^{-i2\pi f'[t-(N+1)/2]} \frac{U_k(f')}{\sqrt{\lambda_k}}\, df'. \tag{363b}$$

This yields

$$\begin{aligned}\Upsilon_k(f) &= \frac{\sqrt{\lambda_k}}{(-1)^k \epsilon_k} \sum_{t=1}^{N} e^{-i2\pi f[t-(N+1)/2]} v_{t-1,k}(N,W) X_t \\ &= \epsilon_k \sqrt{\lambda_k} \sum_{t=1}^{N} h_{t,k} X_t e^{-i2\pi f[t-(N+1)/2]}.\end{aligned} \tag{363c}$$

(here we use the fact that $1/((-1)^k \epsilon_k) = \epsilon_k$). Thus the projection $\Upsilon_k(\cdot)$ of the phase-shifted Fourier transform $\Upsilon(\cdot)$ onto the $k$th basis function $U_k(\cdot)/\sqrt{\lambda_k}$ is just the (phase-shifted and rescaled) Fourier transform of $X_1, \ldots, X_N$ multiplied by the $k$th order dpss data taper.

It is shown in Exercise [7.6] that another representation for $\Upsilon_k(f)$ is

$$\Upsilon_k(f) = \sqrt{\lambda_k} \int_{-1/2}^{1/2} U_k(f')\, dZ(f + f'). \tag{364a}$$

In Equation (363a) the phase-shifted Fourier transform $\Upsilon_k(\cdot)$ is 'seen' through the dpswf with smoothing carried out only over the interval $[-W, W]$, while, in the alternative representation of Equation (364a), $dZ(\cdot)$ is seen through the dpswf with smoothing carried out over the whole Nyquist interval $[-1/2, 1/2]$.

For $\Delta t = 1$ (as we assume in this section), we have

$$\hat{S}_k^{(mt)}(f) \equiv \left| \sum_{t=1}^{N} h_{t,k} X_t e^{-i2\pi ft} \right|^2$$

from Equation (333). The right-hand side of the above can be rewritten as

$$\left| \epsilon_k \sum_{t=1}^{N} h_{t,k} X_t e^{-i2\pi f[t-(N+1)/2]} \right|^2 .$$

It now follows from Equation (363c) that

$$\hat{S}_k^{(mt)}(f) = \left| \Upsilon_k(f)/\sqrt{\lambda_k} \right|^2 . \tag{364b}$$

As we noted in Section 7.1, the estimator $\hat{S}_k^{(mt)}(\cdot)$ is called the *kth eigenspectrum*. The terminology 'eigenspectrum' is motivated by the fact that the taper $\{h_{t,k}\}$ is an eigenvector for Equation (105b). The taper $\{h_{t,k}\}$ has the usual normalization $\sum_{t=1}^{N} h_{t,k}^2 = 1$ (this follows from Parseval's theorem of Equation (89c) because the functions $U_k(\cdot)$ are orthonormal over the interval $[-1/2, 1/2]$). When $\{X_t\}$ is a white noise process, we noted in Section 7.2 that $\hat{S}_k^{(mt)}(f)$ is an unbiased estimator of $S(f)$; i.e., $E\{\hat{S}_k^{(mt)}(f)\} = S(f)$ for all $f$ and $k$.

We are now in a position to consider a method of 'regularizing' the spectral estimation problem. Our approach is first to introduce a set of desirable (but unobservable) projections and then to outline how these can best be approximated by the $\Upsilon_k(\cdot)$ terms – recall that these were formed in Equation (363a) by projecting the rescaled Fourier transform $\Upsilon(\cdot)$ onto the $k$th basis function $U_k(\cdot)/\sqrt{\lambda_k}$. In Equation (364a) the projection $\Upsilon_k(\cdot)$ was then expressed in terms of $dZ(\cdot)$, which determines the

sdf via Equation (361a). Now consider the equivalent *direct* projection for $dZ(\cdot)$ onto the same $k$th basis function:

$$Z_k(f) \equiv \int_{-W}^{W} \frac{U_k(f')}{\sqrt{\lambda_k}}\, dZ(f + f'). \tag{365a}$$

Comparison of Equations (364a) and (365a) shows that – apart from different uses of the scaling factor $\sqrt{\lambda_k}$ – the difference between $\Upsilon_k(\cdot)$ and $Z_k(\cdot)$ lies in the integration limits. Of course, there is a large practical difference – $\Upsilon_k(\cdot)$ is calculable from $X_1, \ldots, X_N$ via Equation (363c), while $Z_k(\cdot)$ is not! Also $dZ(\cdot)$ cannot be written as a finite linear combination of the dpswf's: unlike $\Upsilon(\cdot)$, $dZ(\cdot)$ does not fall in an $N$-dimensional space due to its highly discontinuous nature. Nevertheless, we are still at liberty to project $dZ(\cdot)$ onto each of the $N$ dpswf's to obtain $Z_k(\cdot)$ in Equation (365a).

Note that Equation (365a) can be rewritten in terms of a convolution:

$$Z_k(f) = (-1)^k \int_{f-W}^{f+W} \frac{U_k(f - f')}{\sqrt{\lambda_k}}\, dZ(f'),$$

so that clearly $Z_k(\cdot)$ represents smoothing $dZ(\cdot)$ by a rescaled $k$th order dpswf over an interval of width $2W$. For a chosen $W$, $Z_k(\cdot)$ represents the best – in the sense that the dpswf's maximize the concentration measure (see Section 3.9) – glimpse at $dZ$ in each orthogonal direction. It is straightforward to show that

$$E\{|Z_k(f)|^2\} = \int_{f-W}^{f+W} \left[\frac{U_k(f - f')}{\sqrt{\lambda_k}}\right]^2 S(f')\, df'. \tag{365b}$$

For a general spectral shape this expected value can be interpreted as a smoothed version of the spectrum, with the smoothing by the scaled dpswf (squared) over the interval of width $2W$. When the true sdf is that of a white noise process, we see, using Equation (362d), that

$$E\{|Z_k(f)|^2\} = S(f);$$

i.e., $|Z_k(f)|^2$ is an unbiased (but uncalculable) estimator of $S(f)$ for white noise.

Thomson's 1982 approach puts emphasis on finding a calculable expression close to $Z_k(f)$. Why is $|Z_k(f)|^2$ so appealing when its expected value is a smoothed version of the sdf? The answer illuminates a key difficulty with spectral estimation. Estimation of $S(\cdot)$ from a time series $X_1, \ldots, X_N$ is an ill-posed problem because, given a finite number $N$ of observations, we cannot uniquely determine $S(\cdot)$ over the infinite number of points for which it is defined. As we indicated in the previous section, the problem can be 'regularized' by instead calculating an

average of the function over a small interval, i.e., the integral convolution of the unknown quantity with a good smoother. The smoother in Equation (365b) is good because it smooths only over the main lobe of width $2W$ and thus avoids sidelobe leakage.

Let us now establish a connection between the projections $Z_k(\cdot)$ and $\Upsilon_k(\cdot)$. As has already been pointed out, $Z_k(\cdot)$ cannot be computed from the data whereas $\Upsilon_k(\cdot)$ can. Following the approach of Thomson (1982), we introduce the weight $b_k(f)$ and look at the difference

$$Z_k(f) - b_k(f)\Upsilon_k(f).$$

What weight should we use to make the right-hand side – incorporating the calculable but defective quantity $\Upsilon_k(f)$ – most like the uncalculable but desirable (leakage-free) $Z_k(f)$? From Equations (365a) and (364a) we can write this difference in the following way:

$$\begin{aligned}
&Z_k(f) - b_k(f)\Upsilon_k(f) \\
&\quad = \int_{-W}^{W} \frac{U_k(f')}{\sqrt{\lambda_k}}\, dZ(f+f') - b_k(f)\sqrt{\lambda_k} \int_{-1/2}^{1/2} U_k(f')\, dZ(f+f') \\
&\quad = \left(\frac{1}{\sqrt{\lambda_k}} - b_k(f)\sqrt{\lambda_k}\right) \int_{-W}^{W} U_k(f')\, dZ(f+f') \\
&\qquad - b_k(f)\sqrt{\lambda_k} \int_{f \notin [-W,W]} U_k(f')\, dZ(f+f'),
\end{aligned}$$

where the second integral is over all frequencies in the disjoint intervals $[-1/2, -W]$ and $[W, 1/2]$. Since both these integrals are with respect to $dZ(f)$, they are uncorrelated over these disjoint domains of integration (because for $f \neq f'$ we know $E\{dZ^*(f)\, dZ(f')\} = 0$). Thus the mean square error between $Z_k(f)$ and $b_k(f)\Upsilon_k(f)$, namely,

$$\mathrm{mse}_k(f) \equiv E\{|Z_k(f) - b_k(f)\Upsilon_k(f)|^2\},$$

is given by

$$\begin{aligned}
&\left(\frac{1}{\sqrt{\lambda_k}} - b_k(f)\sqrt{\lambda_k}\right)^2 E\left\{\left|\int_{-W}^{W} U_k(f')\, dZ(f+f')\right|^2\right\} \\
&\qquad + b_k^2(f)\lambda_k E\left\{\left|\int_{f \notin [-W,W]} U_k(f')\, dZ(f+f')\right|^2\right\}.
\end{aligned} \tag{366}$$

Let us look at the first expectation above. We have

$$E\left\{\left|\int_{-W}^{W} U_k(f')\,dZ(f+f')\right|^2\right\}$$
$$= \int_{-W}^{W}\int_{-W}^{W} U_k(f')U_k(f'')E\{dZ^*(f+f')\,dZ(f+f'')\}$$
$$= \int_{-W}^{W} U_k^2(f')S(f+f')\,df' \approx S(f)\int_{-W}^{W} U_k^2(f')\,df' = \lambda_k S(f),$$

provided the sdf $S(\cdot)$ is slowly varying in $[f-W, f+W]$ (the approximation becomes an equality for a white noise sdf).

The second expectation is the expected value of the part of the smoothed $dZ(\cdot)$ outside the primary smoothing band $[-W, W]$. As in the previous section, this can be described as the broad-band bias of the $k$th eigenspectrum. This bias depends on details of the sdf outside $[f-W, f+W]$, but, as a useful approximation, let us consider its average value – in the sense of Equation (86) – over the frequency interval $[-1/2, 1/2]$ of unit length:

$$\int_{-1/2}^{1/2} E\left\{\left|\int_{f\notin[-W,W]} U_k(f')\,dZ(f+f')\right|^2\right\} df$$
$$= \int_{f\notin[-W,W]} U_k^2(f')\int_{-1/2}^{1/2} S(f+f')\,df\,df'$$
$$= \sigma^2\int_{f\notin[-W,W]} U_k^2(f')\,df',$$

where we use the fundamental properties that $S(\cdot)$ is periodic with period 1 and that $\int_{-1/2}^{1/2} S(f)\,df = \sigma^2 = \mathrm{var}\{X_t\}$. Because the $U_k(\cdot)$ functions are orthonormal over $[-1/2, 1/2]$ and orthogonal over $[-W, W]$, we have

$$\int_{f\notin[-W,W]} U_k^2(f')df' = \int_{-1/2}^{1/2} U_k^2(f')\,df' - \int_{-W}^{W} U_k^2(f')\,df' = 1-\lambda_k.$$

Hence the average value of the broad-band bias for general spectra is given by $(1-\lambda_k)\sigma^2$; for white noise the second expectation in Equation (366) need not be approximated by averaging over $[-1/2, 1/2]$, but can be evaluated directly, giving $(1-\lambda_k)\sigma^2$ again, so that this expression is exact for the broad-band bias for white noise. For general spectra a useful approximation to the mean square error is thus

$$\mathrm{mse}_k(f) \approx \left(\frac{1}{\sqrt{\lambda_k}} - b_k(f)\sqrt{\lambda_k}\right)^2 \lambda_k S(f) + b_k^2(f)\lambda_k(1-\lambda_k)\sigma^2, \quad (367)$$

but note that this approximation is an equality for a white noise sdf. To find the value of $b_k(f)$ that minimizes the approximate mean square error, we differentiate with respect to $b_k(f)$ and set the result to zero to obtain

$$b_k(f) = \frac{S(f)}{\lambda_k S(f) + (1 - \lambda_k)\sigma^2}. \tag{368a}$$

With this optimum value for $b_k(f)$, we can substitute it into the approximation to the mean square error to obtain

$$\mathrm{mse}_k(f) \approx \frac{S(f)(1 - \lambda_k)\sigma^2}{\lambda_k S(f) + (1 - \lambda_k)\sigma^2}. \tag{368b}$$

Let us now consider the special case in which $\{X_t\}$ is white noise; i.e., $S(f) = \sigma^2$. From Equation (368a) we see that $b_k(f)$ is identically unity for all frequencies. Hence, for white noise,

$$\Upsilon_k(f) \text{ estimates } Z_k(f)$$

in a minimum mean square error sense. Since the approximation in Equation (367) is an equality in the case of white noise, we have

$$\mathrm{mse}_k(f) = \left(\frac{1}{\sqrt{\lambda_k}} - \sqrt{\lambda_k}\right)^2 \lambda_k\sigma^2 + \lambda_k(1 - \lambda_k)\sigma^2 = (1 - \lambda_k)\sigma^2;$$

i.e., the mean square error is identical to the broad-band bias. If $k$ is less than $2NW - 1$ and hence $\lambda_k \approx 1$, then the mean square error is negligible.

What is the best way to combine the individual eigenspectra in the special case of white noise? From the expression for $\hat{S}_k^{(mt)}(\cdot)$ in Equation (363c) we have

$$\lambda_k \hat{S}_k^{(mt)}(f) = |\Upsilon_k(f)|^2.$$

Since $\Upsilon_k(f)$ estimates $Z_k(f)$, it follows that

$$\frac{1}{K}\sum_{k=0}^{K-1} \lambda_k \hat{S}_k^{(mt)}(f) \text{ estimates } \frac{1}{K}\sum_{k=0}^{K-1} |Z_k(f)|^2.$$

What is $E\{\hat{S}_k^{(mt)}(f)\}$? By definition, this can be obtained by rescaling $E\{|\Upsilon_k(f)|^2\}$. From Equation (364a) it follows that

$$\begin{aligned} E\{|\Upsilon_k(f)|^2\} &= E\{\Upsilon_k^*(f)\Upsilon_k(f)\} \\ &= \lambda_k \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} U_k(f')U_k(f'')E\{dZ^*(f + f')\, dZ(f + f'')\} \\ &= \lambda_k \int_{-1/2}^{1/2} U_k^2(f')S(f + f')\, df' \\ &= \lambda_k \int_{-1/2}^{1/2} U_k^2(f - f')S(f')\, df'. \end{aligned} \tag{368c}$$

For a white noise process, because of the orthonormality of the $U_k(\cdot)$ functions over $[-1/2, 1/2]$, it follows that

$$E\{\hat{S}_k^{(mt)}(f)\} = S(f).$$

Hence a natural spectrum estimator when the sdf is white – based on the first few eigenspectra (i.e., those with least sidelobe leakage) – is given by

$$\bar{S}^{(mt)}(f) \equiv \frac{\sum_{k=0}^{K-1} \lambda_k \hat{S}_k^{(mt)}(f)}{\sum_{k=0}^{K-1} \lambda_k}. \tag{369a}$$

The denominator makes the expression unbiased. Note also that $K \approx \sum_{k=0}^{K-1} \lambda_k$ provided $K$ is chosen to be less than or equal to $2NW - 1$ so that each $\lambda_k$ for $k = 0, \ldots, K-1$ is close to unity.

The estimator in Equation (369a) is considered here to be the initial or basic spectrum estimator resulting from the theory. It is intuitively attractive since, as the order $k$ of the dpss tapers increases, the corresponding eigenvalues will decrease, and the eigenspectra will become more contaminated with leakage; i.e., $\text{mse}_k(f) = (1-\lambda_k)\sigma^2$ will become larger. The eigenvalue weights in Equation (369a) will help to lessen the contribution of the higher leakage eigenspectra. In practice this effect will be negligible provided $K$ is chosen no larger than $2NW - 1$.

The estimator in Equation (369a) can be refined to take into account a colored sdf, as follows. If we combine Equations (364b), (368c) and (369a), we obtain

$$E\{\bar{S}^{(mt)}(f)\} = \int_{-1/2}^{1/2} \left( \frac{1}{\sum_{k=0}^{K-1} \lambda_k} \sum_{k=0}^{K-1} \lambda_k U_k^2(f - f') \right) S(f')\, df'. \tag{369b}$$

Note that this expectation integrates over the full band $[-1/2, 1/2]$ and not just $[-W, W]$. If $S(\cdot)$ has a large dynamic range, notable spectral leakage could occur in this smoothing expression due to sidelobes of the dpswf outside $[-W, W]$ for larger values of $k$, e.g., those values of $k$ approaching $2NW - 1$. Hence Equation (369a) might not be a satisfactory estimator for nonwhite sdf's. By way of contrast, the integration range for the expectation of $|Z_k(f)|^2$ in Equation (365b) is only $[-W, W]$, so that such sidelobe problems are avoided for $|Z_k(f)|^2$. Now

$$b_k^2(f)|\Upsilon_k(f)|^2 \text{ estimates } |Z_k(f)|^2,$$

where $b_k(f)$ will take the general form of Equation (368a), involving the true sdf $S(f)$. Hence

$$\frac{1}{K} \sum_{k=0}^{K-1} b_k^2(f) \lambda_k \hat{S}_k^{(mt)}(f) \text{ estimates } \frac{1}{K} \sum_{k=0}^{K-1} |Z_k(f)|^2$$

for a nonwhite sdf. By noting that $K \approx \sum_{k=0}^{K-1} b_k^2(f)\lambda_k$ provided $K$ does not exceed $2NW - 1$, we arrive at the *adaptive multitaper spectral estimator*

$$\hat{S}^{(amt)}(f) \equiv \frac{\sum_{k=0}^{K-1} b_k^2(f)\lambda_k \hat{S}_k^{(mt)}(f)}{\sum_{k=0}^{K-1} b_k^2(f)\lambda_k}. \tag{370a}$$

This is, of course, of the same form as Equation (369a) when the weights are set to unity. It cannot however be claimed that Equation (370a) is exactly unbiased since $E\{\hat{S}_k^{(mt)}(f)\}$ is not identically $S(f)$ for a nonwhite sdf (see Equation (368c)), but it will generally be close provided $S(f)$ does not vary rapidly over the interval $[f - W, f + W]$. The arguments leading to Equation (370a) justify the same result in Thomson (1982, Equation (5.3)), but the definition of the weights $b_k(f)$ used here is more appealing since these weights are unity for a white noise sdf.

Now the weight formula of Equation (368a) involves the true unknown spectrum and variance. Spectral estimation via Equation (370a) must thus be carried out in an iterative fashion such as the following. To proceed, we assume $NW \geq 2$. We start with a spectral estimate of the form of Equation (369a) with $K$ set equal to 1 or 2. This initial estimate involves only the one or two tapers with lowest sidelobe leakage and hence will preserve rapid spectral decays. This spectral estimate is then substituted – along with the estimated variance – into Equation (368a) to obtain the weights for orders $k = 0, \ldots, K - 1$ with $K$ typically $2NW - 1$. These weights are then substituted into Equation (370a) to obtain the new spectral estimate with $K$ again $2NW - 1$. The spectral estimate given by Equation (370a) is next substituted back into Equation (368a) to get new weights and so forth. Usually two executions of Equation (370a) are sufficient. Thomson (1982) describes a method for estimating the broad-band bias (rather than its average over $[-1/2, 1/2]$) within the iterative stage, but this additional complication often leads to very marginal changes and suffers from additional estimation problems.

An 'equivalent degrees of freedom' argument says that $\hat{S}^{(amt)}(f)$ is approximately equal in distribution to the rv $S(f)\chi_\nu^2/\nu$, where the chi-square rv has degrees of freedom $\nu$ given by

$$\nu = \frac{2\left(\sum_{k=0}^{K-1} b_k^2(f)\lambda_k\right)^2}{\sum_{k=0}^{K-1} b_k^4(f)\lambda_k^2} \tag{370b}$$

for $f \neq 0$ or $f_{(N)}$ (this is Exercise [7.8]).

### Comments and Extensions to Section 7.4

[1] The estimators of Equations (369a) and (370a) differ significantly from each other in two important properties. First, the weights in the spectral estimator of Equation (369a) do not depend on frequency – as a result, Parseval's theorem is satisfied in expected value (this is also true for the simple average of Equation (333)). To see this, use Equation (369b) to write

$$\int_{-1/2}^{1/2} E\{\bar{S}^{(mt)}(f)\}\, df = \sum_{k=0}^{K-1} \lambda_k \int_{-1/2}^{1/2} \frac{1}{\sum_{k=0}^{K-1} \lambda_k} \left( \int_{-1/2}^{1/2} U_k^2(f - f')\, df \right) S(f')\, df'.$$

The integral in the parentheses is unity, so we obtain

$$\int_{-1/2}^{1/2} E\{\bar{S}^{(mt)}(f)\}\, df = \int_{-1/2}^{1/2} S(f')\, df' = \sigma^2,$$

as required. In contrast, Exercise [7.10] shows that Parseval's theorem is in general *not* satisfied exactly in expected value for the adaptive multitaper spectral estimator of Equation (370a).

Second, the simply weighted estimator of Equation (369a) (and also Equation (333)) has approximately constant variance across all the Fourier frequencies (excluding 0 and Nyquist), whereas the adaptive weighting scheme of Equation (370a) can give rise to appreciable variations of the variance throughout the spectrum, making interpretation of the plot of the spectral estimate more difficult (see the example in the next section).

## 7.5 Multitaper Spectral Analysis of Ocean Wave Data

As an example of the multitaper approach to spectral analysis, let us reconsider the ocean wave data of Section 6.18 (this time series is plotted in Figure 295, and various nonparametric spectral estimates for the series are shown in Figures 296 and 301). As discussed previously, we are mainly interested in the rate at which the sdf decreases over the range 0.2 to 1.0 Hz, but the sdf at higher frequencies is also of some marginal interest. Because there are unlikely to be sharp spectral features in these frequency ranges (an assumption that is verified by our analysis in Section 6.18), we can set the resolution bandwidth $2W$ to be fairly large. We thus let $W = 4/(N\,\Delta t)$ initially, yielding $2W = 0.03125$ Hz (recall that $N = 1024$ and $\Delta t = 1/4$ second). Since $2NW\,\Delta t - 1 = 7$, we compute the eigenspectra $\hat{S}_k^{(mt)}(\cdot)$ for orders $k = 0, 1, \ldots, 6$. A comparison of the high-order eigenspectra with the order $k = 0$ and $k = 1$

eigenspectra indicates that $\hat{S}_6^{(mt)}(\cdot)$ has unacceptable bias in the high frequency range (i.e., 1 Hz $\leq f \leq$ 2 Hz) but that all the eigenspectra of lower order than 6 are acceptable. We thus form the simple multitaper estimate $\hat{S}^{(mt)}(\cdot)$ of Equation (333) by averaging $K = 6$ eigenspectra together. This estimate is plotted in Figure 373(a), along with a crisscross in the lower left-hand corner depicting the bandwidth $2W$ and the length (in decibels) of a 95% confidence interval for $10 \log_{10}(S(f))$ based on $10 \log_{10}(\hat{S}^{(mt)}(f))$ with $2K = 12$ degrees of freedom.

If we now let $W = 6/(N \Delta t)$ and hence increase the resolution bandwidth to $2W = 0.046875$ Hz, we find that there are $K = 10$ eigenspectra with acceptable bias properties. When these are averaged together, we obtain the multitaper estimate $\hat{S}^{(mt)}(\cdot)$ shown by the thick curve in Figure 373(b). This estimate has $2K = 20$ degrees of freedom. A comparison between this estimate and the one for $W = 4/(N \Delta t)$ in the top plot shows that the $W = 6/(N \Delta t)$ estimate is smoother in appearance, as we would expect. It is also of interest to compare this estimate with the $m = 150$ Parzen lag window estimate $\hat{S}^{(lw)}(\cdot)$ shown by the thick curve in the top plot of Figure 301. This lag window estimate has a bandwidth of 0.0494 Hz, which is nearly the same as the bandwidth for the $W = 6/(N \Delta t)$ multitaper spectral estimate ($2W = 0.046875$ Hz). We have replotted $\hat{S}^{(lw)}(\cdot)$ as a thin curve in Figure 373(b), but it is just barely visible because it is in such good agreement with $\hat{S}^{(mt)}(\cdot)$ for $W = 6/(N \Delta t)$. Thus, for this time series, lag window and multitaper spectral estimates are quite comparable.

Next, we examine how well the adaptive multispectral estimator $\hat{S}^{(amt)}(\cdot)$ of Equation (370a) performs here. With $W = 4/(N \Delta t)$ as in Figure 373(a), we obtain the adaptive estimate shown as the thick curve in Figure 373(c). A comparison of this estimate with $\hat{S}^{(mt)}(\cdot)$ of Figure 373(a) indicates very good agreement for frequencies between 0 and 1 Hz. For higher frequencies, the estimate $\hat{S}^{(amt)}(\cdot)$ shows more structure than $\hat{S}^{(mt)}(\cdot)$. What accounts for this difference? The estimator $\hat{S}^{(amt)}(\cdot)$ is based upon weighted averages of the eigenspectra $\hat{S}_0^{(mt)}(\cdot)$, $\ldots$, $\hat{S}_6^{(mt)}(\cdot)$. As indicated by Equation (370b), the degrees of freedom $\nu$ for $\hat{S}^{(amt)}(\cdot)$ are frequency-dependent – we have plotted $\nu$ as a function of frequency in Figure 373(d). Note that $\nu$ drops down to about 6 for $f$ between 1 and 2 Hz, which is half the degrees of freedom associated with $\hat{S}^{(mt)}(\cdot)$ in the top plot. At least part of the additional structure in $\hat{S}^{(amt)}(\cdot)$ can be attributed to this decrease in $\nu$. To explore this point further, the upper and lower thin curves in Figure 373(c) show upper and lower 95% confidence limits for each $10 \log_{10}(S(f))$ based upon $10 \log_{10}(\hat{S}^{(amt)}(f))$ (the thick curve). As $\nu$ decreases in the high frequency region, these confidence intervals increase in width, indicating an increase in variability in $\hat{S}^{(amt)}(\cdot)$. The apparent additional spectral structure is in accordance with this increased variability.

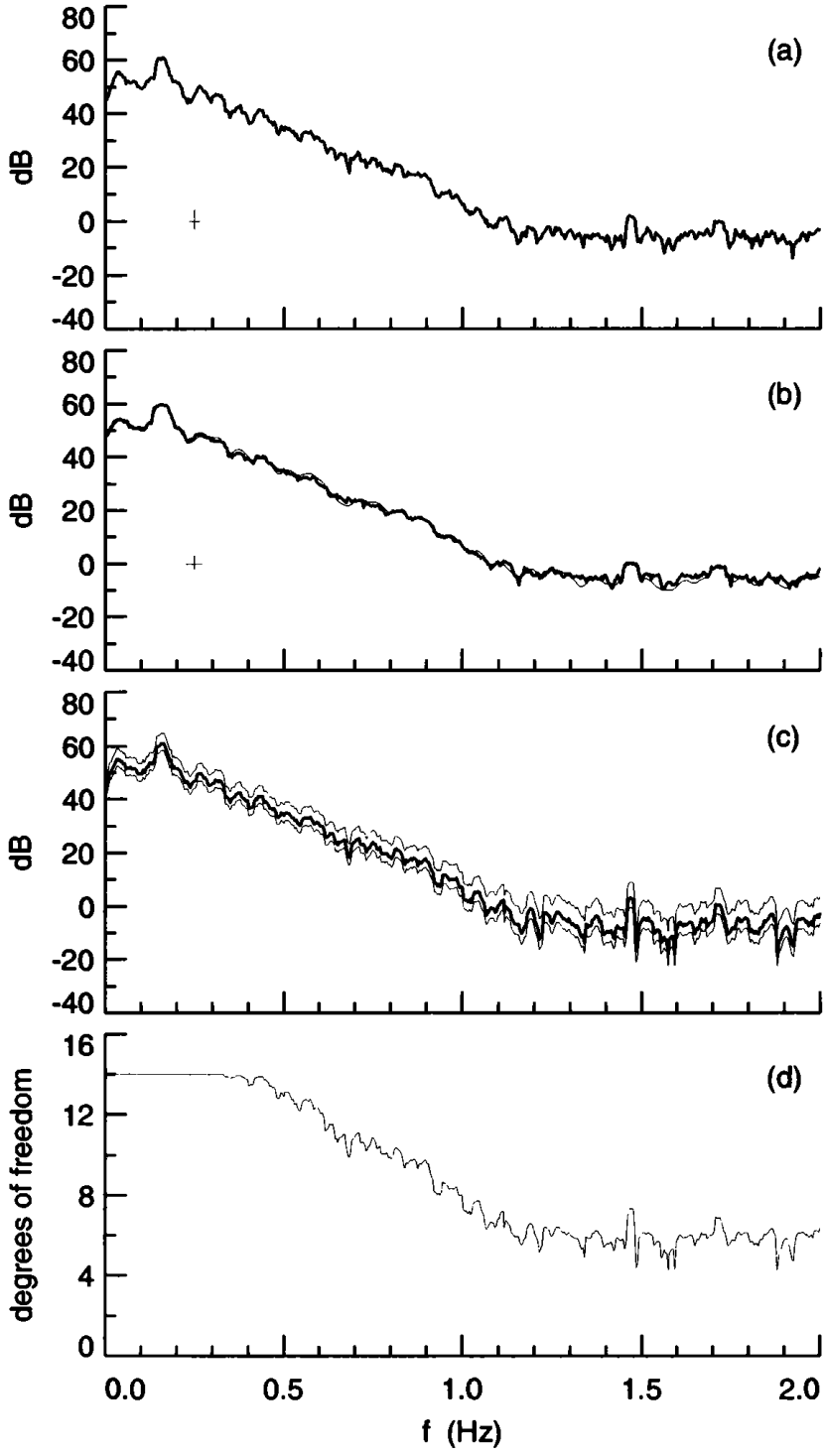


**Figure 373.** Multitaper spectral estimates for ocean wave data.

Finally, we note two additional differences between the simple multitaper estimator $\hat{S}^{(mt)}(\cdot)$ and the adaptive estimator $\hat{S}^{(amt)}(\cdot)$. Once the resolution bandwidth $2W$ is set, the estimate $\hat{S}^{(amt)}(\cdot)$ follows automatically; i.e., we do not have to carefully examine the individual eigenspectra as we must to select a suitable $K$ for $\hat{S}^{(mt)}(\cdot)$. This illustrates the point that the adaptive multitaper estimator can be 'turned loose' on sdf's with high dynamic ranges in an automatic fashion. Second, note that $\hat{S}^{(amt)}(f)$ has two more degrees of freedom than $\hat{S}^{(mt)}(f)$ for $f$ between 0 and 0.3 Hz ($\nu = 14$ as opposed to $\nu = 12$). This is because the adaptive estimator makes use of $\hat{S}_6^{(mt)}(\cdot)$, which we rejected in forming $\hat{S}^{(mt)}(\cdot)$ only because of its poor bias properties at *high* frequencies. The inclusion of $\hat{S}_6^{(mt)}(\cdot)$ for low frequencies makes sense, a point clearly in favor of $\hat{S}^{(amt)}(\cdot)$ (there is, however, very little *actual* difference between the two spectral estimates at low frequencies).

## 7.6 Summary of Multitaper Spectral Estimation

We assume that $X_1, X_2, \ldots, X_N$ is a sample of length $N$ from a zero mean real-valued stationary process $\{X_t\}$ with unknown sdf $S(\cdot)$ defined over the interval $[-f_{(N)}, f_{(N)}]$, where $f_{(N)} \equiv 1/(2\,\Delta t)$ is the Nyquist frequency and $\Delta t$ is the sampling interval between observations. (If $\{X_t\}$ has an unknown mean, we need to replace $X_t$ with $X_t' \equiv X_t - \bar{X}$ in all computational formulae, where $\bar{X} \equiv \sum_{t=1}^{N} X_t/N$ is the sample mean.)

[1] *Simple multitaper spectral estimator* $\hat{S}^{(mt)}(\cdot)$
This estimator is defined in Equation (333) as the average of $K$ eigenspectra $\hat{S}_k^{(mt)}(\cdot)$, $k = 0, \ldots, K-1$, the $k$th of which is a direct spectral estimator employing a dpss data taper $\{h_{t,k}\}$ with parameter $W$. A discussion on how to set $W$ and pick $K$ is given in Section 7.1. The estimator $\hat{S}^{(mt)}(f)$ is approximately equal in distribution to $S(f)\chi_{2K}^2/2K$.

[2] *Adaptive multitaper spectral estimator* $\hat{S}^{(amt)}(\cdot)$
This estimator uses the same eigenspectra as $\hat{S}^{(mt)}(\cdot)$, but it now adaptively weights the $\hat{S}_k^{(mt)}(\cdot)$ terms. The weight for the $k$th eigenspectrum is proportional to $b_k^2(f)\lambda_k$, where $\lambda_k$ is the eigenvalue corresponding to the eigenvector with elements $\{h_{t,k}\}$, while $b_k(f)$ is given by

$$b_k(f) = \frac{S(f)}{\lambda_k S(f) + (1-\lambda_k)\sigma^2\,\Delta t}$$

(this is Equation (368a) for the case where $\Delta t$ is not necessarily unity). The $b_k(f)$ term depends on the unknown sdf $S(f)$, but it can be estimated using the iterative scheme outlined at the end of Section 7.4. The estimator $\hat{S}^{(mt)}(f)$ is approximately equal in distribution to $S(f)\chi_\nu^2/\nu$, where $\nu$ is given by Equation (370b).

## 7.7 Exercises

[7.1] Suppose that $X_1, \ldots, X_N$ is a sample of a zero mean white noise process $\{X_t\}$ with variance $s_0$ and sampling time $\Delta t$. Let $\{\tilde{h}_{t,j}\}$ and $\{\tilde{h}_{t,k}\}$ be any two orthonormal data tapers; i.e.,

$$\sum_{t=1}^{N} \tilde{h}_{t,j}\tilde{h}_{t,k} = 0 \text{ and } \sum_{t=1}^{N} \tilde{h}_{t,j}^2 = \sum_{t=1}^{N} \tilde{h}_{t,k}^2 = 1.$$

For $l = j$ or $k$, define

$$\tilde{J}_l(f) \equiv (\Delta t)^{1/2} \sum_{t=1}^{N} \tilde{h}_{t,l} X_t e^{-i2\pi f t\,\Delta t} \text{ so that } \left|\tilde{J}_l(f)\right|^2 = \tilde{S}_l^{(mt)}(f)$$

(see Equation (349b)).

a) Show that $\operatorname{cov}\{\tilde{J}_j(f), \tilde{J}_k(f)\} = 0$.

b) Under the additional assumption that $\{X_t\}$ is a Gaussian process, use Equation (330d) to show that, for $l = j$ or $k$,

$$\operatorname{var}\{\tilde{S}_l^{(mt)}(f)\} = s_0^2(\Delta t)^2 \left(1 + \left|\sum_{t=1}^{N} \tilde{h}_{t,l}^2 e^{-i4\pi f t\,\Delta t}\right|^2\right)$$

and that

$$\operatorname{cov}\{\tilde{S}_j^{(mt)}(f), \tilde{S}_k^{(mt)}(f)\} = s_0^2(\Delta t)^2 \left|\sum_{t=1}^{N} \tilde{h}_{t,j}\tilde{h}_{t,k} e^{-i4\pi f t\,\Delta t}\right|^2.$$

What do the above equations reduce to if $f = 0$, $f_{(N)}/2$ or $f_{(N)}$?

[7.2] Let $Q$ be an $N \times N$ symmetric matrix of real-valued numbers. The following standard results can be found in any number of books on matrix theory (see, for example, Graybill, 1983):

a) There exists an $N \times N$ orthogonal matrix $P$ such that $P^T Q P = D_N$, where $D_N$ is an $N \times N$ diagonal matrix with diagonal elements $d_1, \ldots, d_N$. Each $d_j$ is an eigenvalue of $Q$ and is necessarily real-valued. We can assume that they are ordered such that $d_1 \geq d_2 \geq \cdots \geq d_{N-1} \geq d_N$.

b) Because $P$ is an orthogonal matrix, we have $P^T P = P P^T = I$, where $I$ is the $N \times N$ identity matrix. This implies that we can write $Q = P D_N P^T$.

c) If $Q$ is positive semidefinite with rank $K$, then the eigenvalues of $Q$ are all nonnegative; the number of positive eigenvalues is equal to $K$ so that $d_1 \geq d_2 \geq \cdots \geq d_K > 0$ and $d_j = 0$

for $K+1 \leq j \leq N$; and we can write $Q = P\sqrt{D_N}\sqrt{D_N}P^T$, where $\sqrt{D_N}$ refers to the diagonal matrix whose elements are the square roots of the elements of $D_N$; i.e., $\sqrt{D_N}\sqrt{D_N} = D_N$.

Using these facts, prove that, if $Q$ is positive semidefinite with rank $1 \leq K \leq N$, we can write $Q = AA^T$, where $A$ is an $N \times K$ matrix such that $A^TA = D_K$, with $D_K$ being a $K \times K$ diagonal matrix whose diagonal elements are the nonzero eigenvalues of $Q$, namely, $d_1, \ldots, d_K$.

[7.3] Show that the first part of Equation (354c) holds, namely,

$$E\{\hat{S}^{(q)}(f)\} = \Delta t \ \mathrm{tr}\{Q\Sigma_Z\}.$$

[7.4] Suppose that $\Sigma$ is an $N \times N$ symmetric positive definite matrix with eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{N-1}$ (ordered from largest to smallest). Show that, if the largest eigenvalue $\lambda_0$ is distinct (i.e., $\lambda_0 > \lambda_1 \geq \cdots \geq \lambda_{N-1}$) and if $A$ is any $N \times K$ matrix such that $\mathrm{tr}\{A^TA\} = 1$, then $\mathrm{tr}\{A^T\Sigma A\}$ is maximized when $A$ is a normalized eigenvector associated with the eigenvalue $\lambda_0$. Hints:

a) the eigenvalues of $\Sigma$ must be positive; and
b) if $\mathbf{V}_0, \mathbf{V}_1, \ldots, \mathbf{V}_{N-1}$ are an orthonormal set of eigenvectors corresponding to the eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{N-1}$, then each column of $A$ can be expressed as a unique linear combination of the $\mathbf{V}_k$ terms.

[7.5] If $A$ is an $N \times K$ matrix whose $k$th column is the rescaled dpss data taper $\{h_{t,k}/\sqrt{K}\}$ of order $(k-1)$, verify that the broad-band bias measure of Equation (358) can be rewritten as

$$\mathrm{tr}\{A^TA\} - \mathrm{tr}\{A^T\Sigma^{(bl)}A\} = 1 - \frac{1}{K}\sum_{k=0}^{K-1}\lambda_k(N,W)$$

(this result was stated in Equation (360)).

[7.6] Show that Equation (364a) holds. Hints: start with the expression for $\Upsilon_k(f')$ in Equation (363a), use Equation (362a), interchange the order of the integrals and the summation such that the innermost integral equals the integral on the right-hand side of Equation (363b), and then use the complex conjugate of Equation (362c) (recall that the dpswf $U_k(\cdot)$ is a real-valued function). Finally, use the fact that both $dZ(\cdot)$ and $U_k(\cdot)$ are periodic functions with a period of unity.

[7.7] If we make use of the fact that the $U_k(\cdot)$ functions are orthonormal over $[-1/2, 1/2]$, we can replace the definition of the projections $\Upsilon_k(f')$ in Equation (363a) with

$$\tilde{\Upsilon}_k(f') \equiv \int_{-1/2}^{1/2} \Upsilon(f+f')U_k(f)\,df.$$

a) Show that $\tilde{\Upsilon}_k(f') = \Upsilon_k(f)/\sqrt{\lambda_k}$.
b) Show that the $\tilde{\Upsilon}_k(f')$ projections lead to the same multitaper spectral estimators we obtained using $\Upsilon_k(f')$, i.e., Equation (369a) for white noise and Equation (370a) for colored noise.

[7.8] Under the assumption that the $\hat{S}_k^{(mt)}(f)$ terms in Equation (370a) are uncorrelated and have the usually assumed distribution for direct spectral estimators (see Equation (223b)), use the 'equivalent degrees of freedom' argument of Section 6.10 to show that the adaptive multitaper estimator $\hat{S}^{(amt)}(f)$ is approximately distributed as $S(f)\chi_\nu^2/\nu$, where $\nu$ is given by Equation (370b).

[7.9] Suppose that the eigenspectra $\hat{S}_k^{(mt)}(f)$ used in forming the multitaper estimator $\hat{S}^{(mt)}(f)$ of Equation (333) are approximately unbiased and uncorrelated (this is a reasonable assertion with restrictions on $f$, the number of tapers and an assumption that the true sdf does not vary too rapidly). Describe a way of estimating the variability in the multitaper estimator that does not make use of the argument that $\hat{S}^{(mt)}(f)$ is proportional to a chi-square random variable with $2K$ degrees of freedom (Thomson and Chave, 1991).

[7.10] Show that, for the adaptive multitaper spectral estimator of Equation (370a), we have

$$\int_{-1/2}^{1/2} E\{\hat{S}^{(amt)}(f)\}\, df$$
$$= \int_{-1/2}^{1/2} S(f') \left( \sum_{k=0}^{K-1} \lambda_k \int_{-1/2}^{1/2} \frac{b_k^2(f)}{\sum_{k=0}^{K-1} \lambda_k b_k^2(f)} U_k^2(f - f')\, df \right) df'.$$

Since the term in the parentheses above is not unity in general, it follows that, in contrast to both $\hat{S}^{(mt)}(\cdot)$ and $\bar{S}^{(mt)}(\cdot)$, Parseval's theorem is in general *not* satisfied exactly in expected value for $\hat{S}^{(amt)}(\cdot)$.