Optimal Bandwidth for Multitaper Spectrum Estimation

Charlotte L. Haley, *Member, IEEE,* and Mihai Anitescu, *Member, IEEE*

***Abstract*—A systematic method for bandwidth parameter selection is desired for Thomson multitaper spectrum estimation. We give a method for determining the optimal bandwidth based on a mean squared error (MSE) criterion. When the true spectrum has a second-order Taylor series expansion, one can express quadratic local bias as a function of the curvature of the spectrum, which can be estimated by using a simple spline approximation. This is combined with a variance estimate, obtained by jackknifing over individual spectrum estimates, to produce an estimated MSE for the log spectrum estimate for each choice of time-bandwidth product. The bandwidth that minimizes the estimated MSE then gives the desired spectrum estimate. Additionally, the bandwidth obtained using our method is also optimal for cepstrum estimates. We give an example of a damped oscillatory (Lorentzian) process in which the approximate optimal bandwidth can be written as a function of the damping parameter. The true optimal bandwidth agrees well with that given by minimizing estimated the MSE in these examples.**

***Index Terms*—Spectrum Analysis, Multitaper, Bandwidth, Cepstrum**

I. Introduction

NONPARAMETRIC spectrum estimation is an indispensable tool in exploratory data analysis of univariate time series. No unbiased spectral estimator exists for a discrete, finite time series, because of uncertainty principles, however, *bandwidth* is a quantity through which the statistical properties of a nonparametric spectrum estimator can be understood. That is, bandwidth controls the natural tradeoff between bias and variance of the estimator. We are concerned with bandwidth optimality in nonparametric spectrum estimation, specifically for multitaper estimates [1, 2] that require a user-defined bandwidth.

The multitaper method offers considerable performance advantages over nonparametric single-taper [3, Fig. 2] and parametric estimates [4, Figs. 3, 6, and 9] [5]. Stoica and Sundin [6] showed that multitaper estimators are maximum likelihood when the spectrum is the sum of step functions with "treads" no smaller than the chosen bandwidth. Bronez [7] showed that when two of bias, variance, and bandwidth are fixed, the multitaper method produces dramatically smaller values for the third quantity than does the Welch method [8], which generalizes the periodogram by averaging spectrum estimates computed on overlapped segments of data. Prieto *et al.* [9] gave a procedure for eliminating a local quadratic bias term in the multitaper estimate; see also Riedel and Siderenko [10]. No systematic method has been described for bandwidth choice. Most computer implementations use a default time-bandwidth product fixed at 4, regardless of the context in which the spectrum estimation is being applied.

The main contribution of this paper, beginning in §VI, is a new methodology to estimate optimal bandwidth by minimizing the estimated mean squared error (MSE) of the log spectrum as a function of time-bandwidth parameter. The MSE is obtained from the data by separately estimating bias, §IV, and variance, §V. Using a second-order Taylor expansion for the spectrum as in [9, 1], bias is described by the second derivative of the spectrum. In §IV-B, the second derivative is estimated by using splines. Variance estimation is done in §V by using the jackknife technique described in [11]. In §VII an illustrative example is given where the optimal bandwidth is known; in this case, minimization of estimated MSE produces bandwidth close to the optimal value. This method greatly reduces the subjectivity in bandwidth choice and is easily automated for simple implementation. We note that optimal bandwidth for the log-spectrum is also optimal for the multitaper cepstrum [12].

II. Multitaper Spectrum Estimation

Let $x(t)$, for $t = 0, 1, \ldots, N-1$, be a zero-mean stationary time series with unit sampling and spectral density $S(f)$ [13]. Consider the simple *multitaper* estimate of the spectrum, $\bar{S}(f)$, computed by

$$\widehat{S^{(k)}}(f) = \left| \sum_{t=0}^{N-1} v_t^{(k)} x_t e^{-i2\pi f t} \right|^2 ; \quad \bar{S}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \widehat{S^{(k)}}(f), \tag{1}$$

which is an average of K approximately independent estimates of the spectrum. The averaging of independent estimators decreases the variance of the overall estimate. The *discrete prolate spheroidal sequences*, or Slepian sequences $v_t^{(k)}(N, W)$, where W is the bandwidth [14], solve the eigenvalue problem

$$\lambda_k(N, W) v_s^{(k)}(N, W) = \sum_{t=0}^{N-1} \frac{\sin 2\pi W(t-s)}{\pi(t-s)} v_t^{(k)}(N, W)$$

for $s = 0, \ldots, N-1$. For ease of notation, $v_t^{(k)}(N, W)$ and $\lambda_k(N, W)$ will be written $v_t^{(k)}$ and λ_k, respectively. The eigenvalues (concentrations) in decreasing order $1 > \lambda_0 >$

Copyright (c) 2017 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

C. L. Haley and M. Anitescu are with the Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439, USA; e-mail: haley@anl.gov.

M. Anitescu is also in the Department of Statistics at the University of Chicago.

Manuscript received April 2, 2017; revised May 26, 2017.

$\lambda_1 > \ldots > \lambda_{N-1} > 0$ have only $K \approx 2NW$ large values close to one, after which they drop sharply to zero. The Slepians form an orthonormal set in time; and the *Slepian functions*,

$$U^{(k)}(f) = \sum_{t=0}^{N-1} v_t^{(k)} e^{-i2\pi ft}, \tag{2}$$

are orthonormal on $(-1/2, 1/2)$ and orthogonal on $(-W, W)$. This orthogonality property ensures the approximate independence of the quantities $\hat{S}^{(k)}(f)$ in (1), thus controlling the variance of $\bar{S}(f)$. The approximate bandlimitedness of the tapers controls the bias.

III. Bandwidth Choice

As mentioned, the bandwidth controls the bias-variance tradeoff in spectrum estimation, so bandwidth choice can be based on either the desired resolution or desired statistical properties. In the former case, if one wishes to resolve two peaks spaced approximately $2W_0$ Hz apart, one should ensure that W is less than W_0. In the latter case, larger bandwidth can generally be used with larger sample size (to favor variance over bias). But in this case many authors prefer to control variance using a multitaper-Welch method [8, 3].

Most often, bandwidth choice is often carried out by simply using $NW = 4$. The reasons for this choice may be that (i) Thomson uses this choice in the original paper [1] for a short series ($N = 100$), (ii) larger NW results in additional computation time, (iii) canned multitaper estimators in standard software packages generally use $NW = 4$ by default, and (iv) since small bandwidth favors bias over variance, we may incorrectly assume there is little harm in using a variable estimate if it is not badly biased.

Unfortunately, severe consequences can result from choosing too large or too small a bandwidth (see supplement §XI). Most importantly, no systematic method currently exists for bandwidth selection. We propose such a method in this paper.

IV. Bias of Multitaper Estimates

Multitaper estimates are asymptotically unbiased and are distributed according to a scaled χ^2_{2K} distribution. The bias of a multitaper spectrum estimator can be written as

$$\mathbf{E}\{\bar{S}(f)\} - S(f) = B_B(f) + B_W(f), \tag{3}$$

where $B_B(f)$ denotes *broadband* expectation, or leakage, and $B_W(f)$ denotes the *local* bias, namely,

$$B_B(f) = \frac{1}{K} \int_{\mathcal{D}} \sum_{k=0}^{K-1} |U^{(k)}(f-\nu)|^2 S(\nu) d\nu; \tag{4}$$

$$B_W(f) = \frac{1}{K} \int_{-W}^{W} \sum_{k=0}^{K-1} |U^{(k)}(f-\nu)|^2 S(\nu) d\nu - S(f). \tag{5}$$

where the set $\mathcal{D} = [-1/2, 1/2) \setminus (-W, W)$.

Broadband expectation is bounded above,

$$B_B(f) \quad \leq \quad \frac{1}{K} \sum_{k=0}^{K-1} (1-\lambda_k)\sigma^2, \tag{6}$$

by using the Cauchy inequality [1, §IV, Eq. (4.7)], where σ^2 is the variance of the process $x(t)$. The quantity $1-\lambda_k$ is very small for $k = 0, \ldots, K-1$, (see [1, Table II]) so the overall leakage is small compared with the local bias.

Replacing the sum in (5), called the *spectral window*, by the simple rectangular approximation

$$\frac{1}{K} \sum_{k=0}^{K-1} |U^{(k)}(f)|^2 \lessapprox \begin{cases} \frac{1}{2W} & f \in (-W, W) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

and assuming the spectrum is twice continuously differentiable, a second-order Taylor series approximation for the spectrum, valid where $\nu \in (f-W, f+W)$, can be employed

$$S(\nu) \approx S(f) + (\nu - f)S'(f) + \frac{(\nu-f)^2}{2!} S''(f). \tag{8}$$

The dominant local bias expression (5) can then be approximated as follows, [9, 10]

$$B_W(f) \approx \frac{1}{2W} \int_{f-W}^{f+W} S(\nu) d\nu - S(f) \tag{9}$$

$$\approx \frac{W^2}{6} \cdot S''(f), \tag{10}$$

where all odd-order derivative terms have disappeared.

A. Bias of the log spectrum

The mean squared error of the log spectrum can be related to the error for the cepstrum [15], defined as the inverse Fourier transform of the log spectrum. Proceeding in the spirit of [12, Eqns. (15) & (18)], we see that

$$\Lambda(f; N, W) = \mathbf{E}\{\big(\log \bar{S}(f) - \log S(f)\big)^2\} \tag{11}$$

$$\approx \left[\log \frac{\mathbf{E}\{\bar{S}(f)\}}{S(f)}\right]^2 + \mathbf{Var}\{\log \bar{S}(f)\} \tag{12}$$

$$\approx \left[\frac{\mathbf{E}\{\bar{S}(f)\} - S(f)}{\mathbf{E}\{\bar{S}(f)\}}\right]^2 + \frac{\mathbf{Var}\{\bar{S}(f)\}}{\mathbf{E}\{\bar{S}(f)\}^2}, \tag{13}$$

where the approximation (13) is appropriate for any consistent estimator of $S(f)$, since it ensures that $\mathbf{E}\{\bar{S}(f)\} \approx S(f)$. That is, the mean squared error of the log-spectrum estimator can be written in terms of the mean squared error of the spectrum estimator and its expectation.

B. Estimation of multitaper local bias of the log spectrum

When the spectrum is twice continuously differentiable, one estimates local bias by estimating the second derivative of the spectrum, as in (10). Local second-derivative estimates of the spectrum have been computed by using Chebyshev polynomials as basis functions as in [9]; however, here we apply an alternative "global" method using splines.

First one selects a set of M frequencies of interest, say, $\{f_m\}_{m=0}^{M}$, with $f_m = mW/P$ for some $P > 0$ and $M = \lfloor P/W \rfloor$, then computes an estimate of the spectrum at each of the nodes $\bar{S}(f_m), m = -M/2, \ldots, M/2-1$ and fits a spline, defined piecewise over $[-1/2, 1/2)$. Choosing $P = 1/2$ results in $2W$ frequency separation between the samples and ensures approximate independence of the $\bar{S}(f_m)$ values. The spline tuning parameter, $\approx \sqrt{K}$, is selected as the

reciprocal of the standard deviation of the estimator [16] and determines the number of knots. One differentiates the spline twice to obtain an estimate of the second derivative $\tilde{S}''(f)$ of the spectrum [16]. Since the estimate of the spectrum is asymptotically unbiased and consistent, the limiting behavior of the spline estimate of the second derivative is convergence to its true value.

The result is the following expression for the estimated squared bias of the log spectrum (first term of (13)),

$$\left[\frac{\frac{W^2}{6}\tilde{S}''(f_m)}{\bar{S}(f_m)}\right]^2, \tag{14}$$

evaluated at the spline frequencies.

V. Variance of Multitaper Spectrum Estimators

The variance expression for the multitaper is quartic in $x(t)$ and can be found in [1, §VII] for the simple case where the data $x(t)$ are Gaussian. The multitaper approach gives consistent estimates of the spectrum provided that the number of tapers, K, increases with the number of samples, N. The variance can be simply estimated as follows.

Jackknife Variance Estimation for Multitaper Spectra

The spectra, $\widehat{S}^{(k)}(f)$, produced as intermediates of the multitaper computation, are approximately independent and hence can be used to produce an estimate of variance. The *delete one jackknife* method [17, 18, 11], gives an estimate having less bias than one computed directly as a sample variance of the individual eigenspectra [19] and is statistically efficient for computation on a small number of independent samples. Here we jackknife the logarithms of the spectra, in lieu of the spectrum itself, in order to stabilize the variance of the resulting estimate [20, §5.3]. The details of this computation are reviewed in the supplement. Denote by $\hat{\sigma}_J^2(f)$ the estimate of the variance of the log spectrum at frequency f. The result roughly follows a $t-$distribution, so one can use the variance estimate to construct $100(1-\alpha_0)\%$ confidence intervals for the log spectrum as in [11, Eq. (2.49)]. Note that since the multitaper method produces approximately independent estimates, $\widehat{S}^{(k)}(f)$, it is the only method for spectrum estimation that is accompanied by confidence intervals [11].

VI. Optimal Bandwidth Estimate

When the true log spectrum is unknown, the MSE is estimated by squaring the spline estimate of the bias in (14) and adding the jackknife estimate of the variance $\hat{\sigma}_J^2(f)$ of the log spectrum:

$$\hat{\Lambda}(f_m; N, W) = \frac{W^4}{36}\left[\frac{\tilde{S}''(f_m)}{\bar{S}(f_m)}\right]^2 + \hat{\sigma}_J^2(f_m). \tag{15}$$

Then, by summing over all frequencies to get an overall estimate of the error, the optimal time-bandwidth product can be written

$$\arg\min_{NW} \sum_{m=0}^{M-1} \hat{\Lambda}(f_m; N, W), \tag{16}$$

where N is assumed fixed. By computing (16) for various choices of NW, usually restricting oneself to half-integer values, one can select the bandwidth that produces the estimate with the least estimated MSE.

Summary of bandwidth selection

In summary, for each NW, one computes an integrated MSE estimate as follows:

1) Compute the multitaper estimate of the spectrum (1).
2) Assuming a second-order Taylor's series is appropriate, estimate the local bias as described in §IV-B.
   a) Fit a spline to $\bar{S}(f)$ on a grid of frequencies $\{f_m\}$ spaced $2W$ apart.
   b) Evaluate the second derivative of the spline estimate at these frequencies, denoted $\bar{S}''(f_m)$.
   c) Estimate $\widehat{B_W}(f)$ as in (10) as the sum $\frac{W^2}{6}\bar{S}''(f_m)$.
   d) Estimate the squared bias of the log spectrum using (14).
3) Compute the jackknife estimate of the variance of the log spectrum, $\hat{\sigma}_J^2(f)$, from §V.
4) Substitute the quantities from 2d) and 3) into (16), and sum over the set of frequencies f_m.

Then select NW such that the quantity from 4) is minimized.

VII. Lorentzian Optimal Bandwidth

As an example, we introduce the *Lorentzian* spectrum which models damped oscillatory random processes,

$$L(f) = \frac{A\alpha^2}{\alpha^2 + (f - f_0)^2}, \tag{17}$$

where f_0 is the frequency (in Hz) of the undamped oscillator, the *half-width at half-maximum*, α, of the oscillator represents the rate of exponential decay, or damping; and A is the gain. In the simulations that follow, parameters $A = 100$, $\alpha = 0.01$, and $f_0 = 0.2$ have been used.

The relationship between time-bandwidth product and MSE is shown in Fig. 1 by simulating the Lorentzian process with sample size $N = 1,000$ twenty-five times. If one computes as the integrated squared distance between the spectrum estimate (1) and true spectrum (17), one gets the solid curve in Fig. 1. Using our method to estimate (16) one obtains the dashed curve, the result (marked with "x") is the mean of the bandwidth obtained over the 25 simulated processes. Splines were estimated using a grid of frequencies placed $W/2$ Hz apart, allowing slight correlation between samples, while ensuring dense enough sampling to capture overall behavior. Boxplots show the robust summary statistics for the optimal NW values obtained over the 25 simulations. Note that the minima of these two mean curves are close: the minimum of the estimated MSE occurs at $NW = 16$ on average, while the minimum MSE obtained by simulation occurs near $NW = 14.5$.

The optimal bandwidth choice for the Lorentzian spectrum depends only on sample size and α and is derived in the supplement. Note that this is for the case where the logarithm of the spectrum is not taken. In summary, the bandwidth that

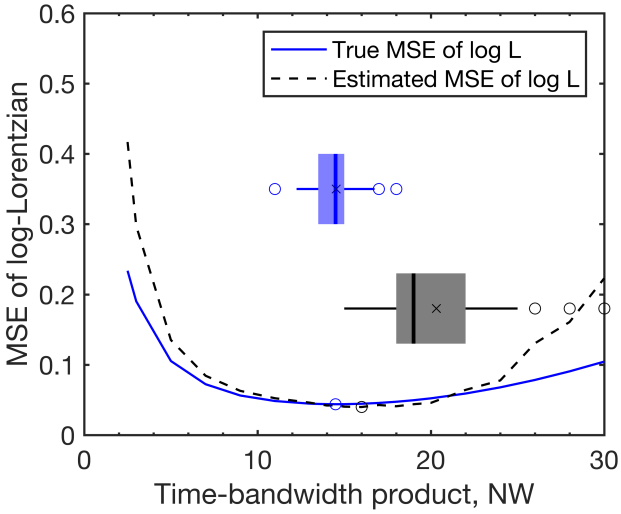

Fig. 1. MSE estimates for the log-Lorentzian (16) vs NW for $N = 1,000$ averaged across 25 realizations and all frequencies. Two quantities plotted are the integrated MSE of Figthe log-Lorentzian, obtained by simulating the process 25 times, and computing the squared distance between multitaper spectrum and true value (blue, solid), and its estimated value (black, dashed line). Boxplots show values of the time-bandwidth corresponding to the smallest MSE obtained by simulation (upper boxplot, corresponds to solid blue curve) using the true spectrum, and by estimating the MSE using our method (lower boxplot, corresponds to dashed curve). Open blue circles on the curves represent minima of the means of the MSE curves (of 25 samples). While mean values are in near-agreement (marked with "x") at $NW = 16$, estimated and $NW = 14.5$, simulated, the estimation procedure tends to give minima of the MSE at a slightly larger time-bandwidth product.

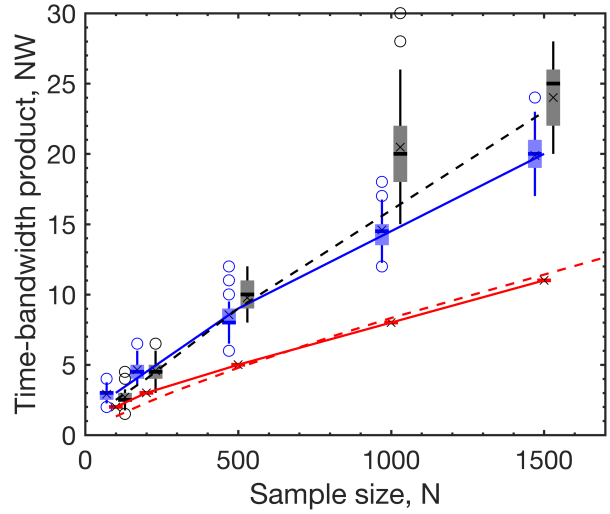

Fig. 2. Optimal time-bandwidth product for the Lorentzian (red curves) and log-Lorentzian (blue curves) as a function of sample size. Here, the same parameters are used as in Fig. 1, and mean values over 25 simulations of the process for the optimal time-bandwidth are shown. Optimal time-bandwidth computed by using our method agrees within half of NW for all N shown. When the logarithm is not taken, an analytic expression for the optimal bandwidth is derived in the supplement, and is shown as the dashed red curve, scaled to approximately match the solid red line.

minimizes the integrated MSE for the Lorentzian spectrum is asymptotically

$$W_{opt} = c_0 \left(\frac{\alpha^4}{N} \right)^{1/5}, \tag{18}$$

where c_0 is some positive constant.

Figure 2 shows the optimal NW values for the Lorentzian and log-Lorentzian as N increases. Red curves represent the optimal NW based on minimizing the MSE computed on the spectrum itself and not its logarithm (18) (not obtained experimentally, but compare with (18)), and blue curves represent optimal NW values obtained by minimizing the MSE of the log spectrum. Solid lines represent the difference between the spectrum estimate and its true value, as in Fig. 1. If the MSE of the log spectrum is minimized, one arrives at a larger optimal time-bandwidth product than if one were to minimize the MSE of the spectrum itself.

We conclude that, for this example, the method we described for computing the MSE from the data yields approximately the same curve (on average) that one can obtain by integrating the squared difference between the log spectrum estimate and its true value, however as Figs 1 and 2 show, one can easily obtain large values for time-bandwidth parameter, since large NW can produce spuriously small values for estimated bias.

VIII. Discussion and Conclusions

We have described a technique for estimating the mean squared error of a multitaper spectrum estimate in terms of its bandwidth. This procedure greatly reduces the subjectivity in the choice of bandwidth for the multitaper method. As shown in the supplement, inappropriate choice of bandwidth has severe consequences; and in none of these examples is a time-bandwidth product of 4 optimal. Fortunately, our method can be implemented automatically so the user will not have to manually choose the bandwidth. We caution that this method is appropriate on regions in frequency where the true spectrum is sufficiently smooth. When conditions are violated, often from discontinuities resulting from line components, one can identify these using a pilot spectrum estimate and F-test and replace them in a "reshaped" version of the spectrum [1, §XIII B]. By virtue of being squared magnitudes of linear combinations of complex exponentials, we remind that all spectrum estimates are infinitely continuously differentiable. As such, no spectrum estimator will be able to faithfully reproduce cusps or large jump discontinuities except in the limit of large sample sizes. As is seen in the supplement, the multitaper method performs well as compared to other nonparametric estimators when the true spectrum has large range, and in this respect is ideally suited to estimating the size of large discontinuities.

Acknowledgments

We thank C. J. Geoga, D. L. Riegert, M. Sandsten, and D. J. Thomson for their suggestions and careful reading.

M. A. acknowledges partial NSF funding through awards FP061151-01-PR and CNS-1545046.

This material was based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under Award DE-SC0008603 and Contract DE-AC02-06CH11347.

References

[1] D. J. Thomson. "Spectrum estimation and harmonic analysis". *Proceedings of the IEEE* 70 (1982), pp. 1055–1096.

[2] J. Park, C. R. Lindberg, and F. L. Vernon. "Multitaper spectral analysis of high-frequency seismograms". *J Geophys. Res.* 92 (1987), pp. 12, 675–12, 684.

[3] D. J. Thomson and C. L. Haley. "Spacing and shape of random peaks in non–parametric spectrum estimates". *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 470.2167 (2014), p. 20140101.

[4] J. M. Lees and J. Park. "Multiple-taper spectral analysis: A stand-alone C-subroutine". *Computers & Geosciences* 21.2 (1995), pp. 199–236.

[5] D. B. Percival and A. T. Walden. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge: Cambridge Univ. Press, 1993.

[6] P. Stoica and T. Sundin. "On nonparametric spectral estimation". *Circuits Systems Signal Process.* 18 (1999), pp. 169–181.

[7] T. P. Bronez. "On the performance advantage of multitaper spectral analysis". *IEEE Trans. on Signal Processing* 40 (1992), pp. 2941–2946.

[8] P. D. Welch. "The use of the fast Fourier transform for estimation of spectra: A method based on time averaging over short, modified periodograms". *IEEE Trans. on Audio and Acoustics* 15 (1967), pp. 70–74.

[9] G. A. Prieto, R. L. Parker, D. J. Thomson, F. L. Vernon, and R. L. Graham. "Reducing the bias of multitaper spectrum estimates". *Geophys. J. Inter.* 171 (2007), pp. 1269–1281.

[10] K. S. Riedel and A. Sidorenko. "Minimum bias multiple taper spectral estimation". *IEEE Trans. on Signal Processing* 43 (1995), pp. 188–195.

[11] D. J. Thomson and A. D. Chave. "Jackknifed error estimates for spectra, coherences, and transfer functions". In: *Advances in Spectrum Analysis and Array Processing*. Ed. by Simon Haykin. Vol. 1. Upper Saddle River, NJ: Prentice-Hall, 1991. Chap. 2, pp. 58–113.

[12] Maria Hansson-Sandsten. "Mean square error optimal weighting for multitaper cepstrum estimation". *EURASIP Journal on Advances in Signal Processing* 2013.1 (2013), p. 158.

[13] H. Cramér and M.R. Leadbetter. *Stationary and Related Stochastic Processes; Sample Function Properties and Their Applications*. New York: Wiley, 1967.

[14] D. Slepian. "Prolate spheroidal wave functions, Fourier analysis, and uncertainty V: The discrete case". *Bell System Tech. J.* 57 (1978), pp. 1371–1429.

[15] Y. Pawitan and F. O'Sullivan. "Nonparametric spectral density estimation using penalized Whittle likelihood". *Journal of the American Statistical Association* 89.426 (1994), pp. 600–610.

[16] P. Craven and G. Wahba. "Smoothing noisy data with spline functions". *Numerische Mathematik* 31.4 (1978), pp. 377–403.

[17] B. Efron and C. Stein. "The jackknife estimate of variance". *Ann. Statist.* 9 (1981), pp. 586–596.

[18] D. J. Thomson. "Quadratic-inverse spectrum estimates: applications to paleoclimatology". *Phil. Trans. Royal Soc. London A* 332 (1990), pp. 539–597.

[19] N. Cressie. "Transformations and the jackknife". *J. Royal Statist. Soc. B* 43 (1981), pp. 177–182.

[20] R. G. Miller. "The jackknife: A review". *Biometrika* 61 (1974), pp. 1–15.

Optimal Bandwidth for Multitaper Spectrum Estimation - Supplementary Material

Charlotte L. Haley, *Member, IEEE,* and Mihai Anitescu, *Member, IEEE*

This supplement serves to (i) motivate the use of multitaper spectrum analysis over single-taper estimates, (ii) summarize the jackknife method for multitaper variance estimation, described in the literature and referenced in the body of the text, (iii) show the consequences of poor multitaper bandwidth choice on the final estimate, and (iv) provide a derivation for the optimization of MSE in our motivating example.

IX. Note on asymptotic unbiasedness of the periodogram

Schuster's 1898 periodogram estimator for the spectrum of a stationary process $x(t)$,

$$P(f) = \left| \sum_{n=0}^{N-1} x(t) e^{-i2\pi f t} \right|^2, \tag{19}$$

is an unbiased estimator of the spectrum as $N \to \infty$ at the Fourier frequencies. This statement has justified the use of the periodogram (or smoothed periodogram) for scientific inference when large samples are available, but we caution that this statement is grossly misleading because convergence can be too slow for practical purposes.

In general, processes that have large spectral range will have poor periodogram and direct estimates since high sidelobes of the periodogram spectral window (Fejér kernel) cannot resolve such a large range. Consider the example in [3, Fig. 1(a)], in which $N = 1,000$ samples of barometric pressure data are used to compute a periodogram estimate that differs from the multitaper with $NW = 8$ by a factor of at least 1,000 across most of the frequency range, and by 10^7 in higher frequencies. Since the multitaper estimator has leakage bounded above (see (6)) by 8.3×10^{-17}, such a discrepancy shows that the sample size $N = 1,000$ is not large enough to use the periodogram for scientific inference. Furthermore, the estimate would not benefit from smoothing. If a multitaper estimate had not been computed on these data, it would not be known that these data have such a large spectral range. In fact, any effort to fit a parametric model to these data based on a periodogram [21, 15] or a Bartlett covariance would result in more high-frequency power in the model than actually exists.

Copyright (c) 2017 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

C. L. Haley and M. Anitescu are with the Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439, USA; e-mail: haley@anl.gov.

M. Anitescu is also in the Department of Statistics at the University of Chicago.

Manuscript received April 2, 2017; revised May 26, 2017.

Although asymptotic unbiasedness of periodograms and single-tapered estimates is important, this fact is of little use when doing analyses on finite samples, since no sample size is large enough to sufficiently reduce the periodogram bias. More examples of bias in periodogram and direct estimates can be found in [5, p. 227].

X. Jackknifing Multitaper Spectrum Estimates

In this section, we give an overview of jackknifing multitaper spectrum estimates in order to obtain variance estimates and confidence intervals. The spectra, $\widehat{S}^{(k)}(f)$, produced as intermediates of the multitaper computation, are approximately independent, by virtue of having been computed with orthogonal tapers, and hence can be used to produce an estimate of variance. The *delete one jackknife* method [17, 18, 11, 22], described here gives an estimate having less bias than one computed directly as a sample variance of the individual spectra [19] and is statistically efficient for computation on a small number of independent samples.

Beginning with the spectra, $\hat{S}^{(k)}(f)$ from (1) for $k = 0, \ldots, K-1$, one computes *delete one* estimates of the log spectrum

$$\ln \widehat{S}_{\setminus j}(f) = \ln \left[\frac{1}{K-1} \sum_{\substack{k=0 \\ k \neq j}}^{K-1} \hat{S}^{(k)}(f) \right] \tag{20}$$

$$= \ln \left[\frac{1}{K-1} (K \bar{S}(f) - \hat{S}^{(j)}(f)) \right]. \tag{21}$$

The average of the delete one estimates is denoted

$$\ln \widehat{S}_{\setminus \bullet}(f) = \frac{1}{K} \sum_{j=0}^{K-1} \ln \widehat{S}_{\setminus j}(f). \tag{22}$$

The jackknife variance estimate is computed as

$$\hat{\sigma}_J^2(f) = \frac{(K-1)^2}{K(K-\frac{1}{2})} \sum_{j=0}^{K-1} [\ln \widehat{S}_{\setminus j}(f) - \ln \widehat{S}_{\setminus \bullet}(f)]^2. \tag{23}$$

The expected value of this last is

$$\mathbf{E}\{\hat{\sigma}_J^2(f)\} = \frac{(K-1)^3}{K(K-\frac{1}{2})} \left[\frac{2}{(K-2)^2} + \frac{1}{2} \left[\psi'\left(\frac{K-1}{2}\right) - \psi'\left(\frac{K-2}{2}\right) \right] \right]. \tag{24}$$

Where ψ' is the trigamma function,

$$\psi'(t) = \frac{d^2}{dt^2} \ln \Gamma(t); \quad \Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx. \tag{25}$$

In general, jackknife variance agrees well with its expected value when the data are stationary. One can then construct confidence intervals for the log spectrum according to

$$\bar{S}(f)e^{t_{K-1}(\alpha_0/2)\hat{\sigma}_J} < S(f) \leq \bar{S}(f)e^{t_{K-1}(1-\alpha_0/2)\hat{\sigma}_J}. \quad (26)$$

In the Lorentzian example, the variances follow (24) almost exactly. Examples of jackknife standard deviation are illustrated in the next section. We note that the multitaper method is the only nonparametric method for which one can compute confidence intervals for the spectrum estimate.

XI. Motivation for the Optimization of Multitaper Bandwidth

As an example, we use the Lorentzian spectrum (17) with parameters $A = 100$, $\alpha = 0.01$, $f_0 = 0.2$, and $N = 100$ introduced in the main paper. Simulating this process and estimating the spectrum by using the multitaper method with various choices of bandwidth parameter, we obtain the spectrum estimates shown in Figs. 3 and 4, where the true spectrum is shown in dashed red. Dash-dotted vertical lines show one bandwidth away from the peak and have the same color as the associated spectral estimate. The overall impression is that the shape of the peak reflects the shape of the multitaper spectral window (7), which is square with width $2W$.

In Figs. 3 and 4, two spectrum estimates with different bandwidth choices are shown together with shaded 5% and 95% confidence intervals (26).

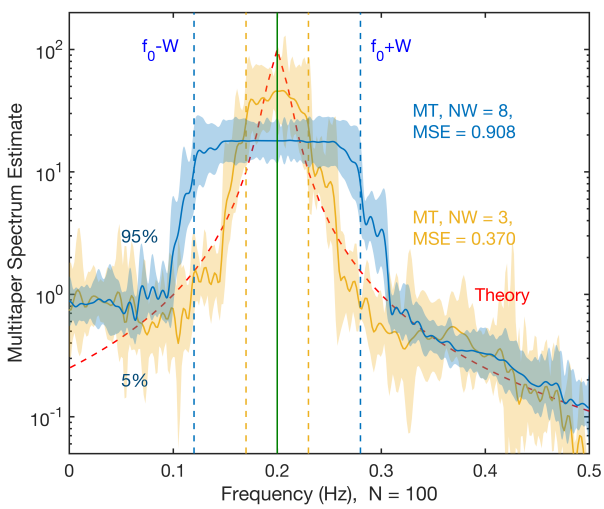

Fig. 3. Two multitaper spectrum estimates computed from simulated Lorentzian data using $NW = 3$ in yellow (lighter shading) and $NW = 8$ in blue (darker shading) along with their jackknife 90% confidence intervals. These are single realizations and have not been averaged. Dot-dashed vertical lines show the bandwidth. Expected jackknife variances for these log spectra are 0.1807 and 0.0645, and MSE are $0.370, 0.908$ for $NW = 3$ and for $NW = 8$, respectively.

Figure 3 shows that choosing too large an NW results in a spectrum with large bias. Comparing the time-bandwidth $NW = 8$, blue line with $NW = 3$, yellow line, the latter much better represents the peak; but it is accompanied by larger jackknife variance (0.1807 vs 0.0645, respectively, given in log-scale units). Here the superior choice is $NW = 3$, since the empirical MSE is reduced by about 40% (0.370 vs 0.908, respectively) on a log scale. Note that with a series of only $N = 100$ data points, no spectrum estimate looks acceptably close to its true value; but from this example it is clear that one dearly pays the price in bias if one chooses too large a bandwidth. The periodogram of this series (not shown) has a range three orders of magnitude larger.

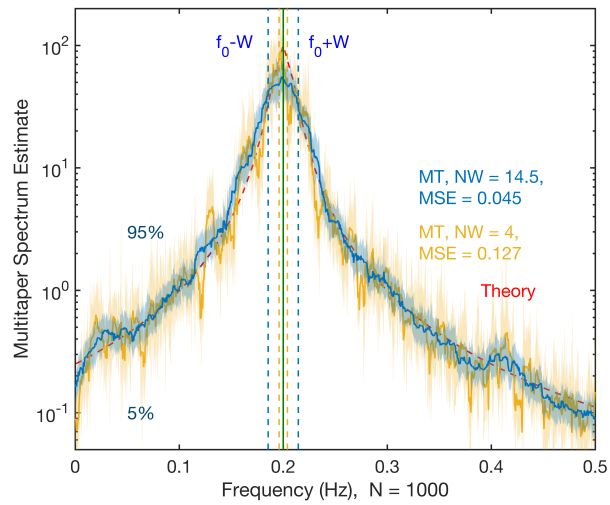

Fig. 4. Two spectrum estimates using $NW = 4$ in yellow (lighter shading) and $NW = 14.5$ in blue (darker shading) along with their jackknife 90% confidence intervals and mean squared errors. Here the sample size has been increased to $N = 1,000$ in comparison with the spectra shown in Fig. 3, while all other parameters remain the same. The expected jackknife variances of these log spectra are 0.1330, and 0.0351, and the mean squared error between the true log spectrum and the two log estimates, integrated over all frequencies, is 0.127 for $NW = 4$ and 0.045 for $NW = 14.5$.

An overly variable estimate results when one chooses a bandwidth that is too small. Using the same Lorentzian parameters as before but increasing the sample size to $N = 1,000$, we show the comparison of time-bandwidth $NW = 4$ with $NW = 14.5$ in Fig. 4. The choice of smaller bandwidth results in larger confidence intervals (expected jackknife variance of 0.1330 and 0.0351 log spectrum units, respectively) wanders farther about the true spectrum than the other and has almost triple the mean squared error (0.127 vs 0.045, respectively). The optimal time-bandwidth product, as determined by our method is $NW = 14.5$. The choice of $NW = 4$, the "default" time-bandwidth product, performs poorly on this example.

In summary, adjustment of the bandwidth parameter allows us to exchange bias for variance, with small bandwidth resulting in small bias but large variance. As these examples show, unwise bandwidth parameter selection can result in a poor spectrum estimate.

XII. Lorentzian Optimal Bandwidth

In this section, we derive an asymptotic result for optimal bandwidth based on the MSE of the spectrum in terms of the derivatives of the spectrum, and we determine this optimal bandwidth for the toy example when the spectrum is known to be Lorentzian. Note that this result is valid when the MSE

of the spectrum and not its logarithm is used to find optimal bandwidth, the relationship between which is illustrated in Fig. 2.

A. Theoretical Variance of the Multitaper

One can express in numerous ways the variance of the simple multitaper estimate (1) in full generality; but when the spectrum is C^∞, see [23, Eqs. (4.20, 4.21)], we can write

$$\begin{aligned}\mathbf{Var}\{\bar{S}(f)\} &= \frac{1}{K^2}\sum_{j,k=0}^{K-1}\mathbf{Cov}\{\widehat{S}^{(k)}(f),\hat{S}^{(j)}(f)\}\\ &= \frac{1}{K^2}\sum_{p,q=0}^{\infty}D_{p,q}S^{(p)}(f)S^{(q)}(f),\end{aligned} \tag{27}$$

where $S^{(p)}(f)$ is the pth derivative of the true spectrum. The coefficients $D_{p,q}$ are

$$D_{p,q} = \int_{-W}^{W}\int_{-W}^{W}\frac{\nu^p}{p!}\frac{f^q}{q!}\left|\sum_k \lambda_k U^{(k)}(\nu)U^{*(k)}(f)\right|^2 d\nu df. \tag{28}$$

From (28), we see that $D_{0,0}$ is K, $D_{p,q} = D_{q,p}$, and $D_{1,0} = 0$ from orthogonality of the Slepian functions and symmetry $U_k(f) = U_k^*(-f)$. Approximating $D_{1,1}$, we obtain [23, Eq. (4.19)]

$$D_{1,1} \approx \frac{4}{15}\cdot\frac{\sqrt{6}}{\pi}\cdot 2NW^3. \tag{29}$$

This results in the following first-order Taylor formulation for multitaper variance,

$$\mathbf{Var}\{\bar{S}(f)\} \approx \frac{1}{2NW}S(f)^2 + \frac{2}{15}\frac{\sqrt{6}}{\pi}\frac{W}{N}S'(f)^2, \tag{30}$$

where $K = 2NW$ is used.

B. Theoretical Expressions for MSE

We can express the MSE of the spectrum in terms of the derivatives of the true spectrum by combining expressions (3) and (30), ignoring broadband bias,

$$\begin{aligned}\mathbf{E}\{(\bar{S}(f)-S(f))^2\} \approx \frac{W^4(S''(f))^2}{36} + \frac{S^2(f)}{2NW}\\ +\frac{2}{15}\frac{\sqrt{6}}{\pi}\frac{W}{N}\left(S'(f)\right)^2,\end{aligned} \tag{31}$$

where the first term is the squared bias, the last two terms are variance (27), and we have used the first-order Taylor series approximation.

Explicitly, formula (31) gives the MSE at frequency f for a multitaper estimate $\bar{S}(f)$ with bandwidth W estimated from a finite sample of length N, when the true spectrum of the process is $S(f)$. To obtain a single index that measures the error, we simply integrate (31) over $f \in [-1/2, 1/2)$. In the next section we obtain an analytic expression for the integrated MSE for the Lorentzian spectrum. This expression depends on the half-width at half-maximum, α; the number of samples, N; and the bandwidth of the estimator, W. We optimize the integrated MSE over the bandwidth.

C. Bandwidth Optimality

To find the optimal bandwidth for the Lorentzian example, we begin by deriving an asymptotic expression for the integrated MSE. Equation (31) contains squared derivatives of the spectrum. Thus, referring to (17) and substituting $f_0 = 0, A = 1$ for simplicity (since the optimal bandwidth will depend only on α and N) one gets

$$\frac{d}{df}L(f) = \frac{-2\alpha^2 f}{(\alpha^2+f^2)^2} \tag{32}$$

$$\frac{d^2}{df^2}L(f) = \frac{2\alpha^2\left(3f^2-\alpha^2\right)}{(\alpha^2+f^2)^3}. \tag{33}$$

The quantity of interest is (31). Integrating the squared quantities over the entire frequency range and retaining only the leading term of a Taylor series in α centered at 0, one gets

$$\int_{-\frac{1}{2}}^{\frac{1}{2}}(L(f))^2 df \asymp \Gamma_0\alpha \tag{34}$$

$$\int_{-\frac{1}{2}}^{\frac{1}{2}}(L'(f))^2 df \asymp \Gamma_1\frac{1}{\alpha} \tag{35}$$

$$\int_{-\frac{1}{2}}^{\frac{1}{2}}(L''(f))^2 df \asymp \Gamma_2\frac{1}{\alpha^3} \tag{36}$$

where Γ_0, Γ_1, and Γ_2 are positive constants. The integrated MSE then becomes asymptotically

$$\frac{\pi W^4}{48\,\alpha^3} + \frac{\pi\alpha}{4\,NW} + \frac{\sqrt{6}W}{30\,N\alpha}. \tag{37}$$

To find the bandwidth that minimizes integrated MSE, one sets to zero the derivative of the above expression with respect to W, and solves to obtain

$$W_{opt} = c_0\left(\frac{\alpha^4}{N}\right)^{1/5}, \tag{38}$$

where c_0 is some positive constant. Back-substituting into the MSE expression (37) results in an integrated squared error on the order of

$$c_1\left(\frac{1}{N^6\alpha}\right)^{1/5}, \tag{39}$$

where c_1 is some positive constant. As long as $N > \alpha^{-1/6}$ the error is under control, which implies a bandwidth choice of $W < \alpha^{5/6}$, asymptotically.

References

[21] P. Whittle. "The analysis of multiple stationary time series". *Journal of the Royal Statistical Society. Series B (Methodological)* (1953), pp. 125–139.

[22] D. J. Thomson. "An overview of multiple-window and quadratic-inverse spectrum estimation methods". *In Proceedings of ICASSP* VI (1994). Invited plenary lecture, pp. 185–94.

[23] D. J. Thomson. "Multitaper analysis of nonstationary and nonlinear time series data". In: *Nonlinear and Nonstationary Signal Processing*. Ed. by W. Fitzgerald, R. Smith, A. Walden, and P. Young. Cambridge: Cambridge Univ. Press, 2001, pp. 317–394.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan. http://energy.gov/downloads/doe-public-access-plan.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, under contract number DE-AC02-06CH11357.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan. http://energy.gov/downloads/doe-public-access-plan.