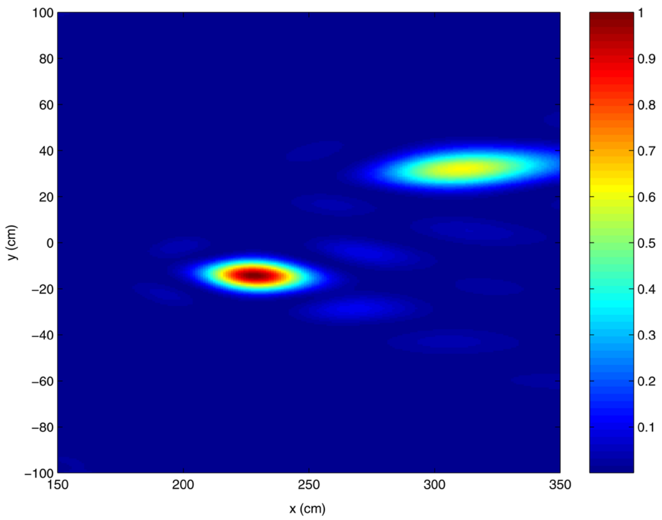

Fig. 3. MLE of scattering potential based image using imaging function (16) and the measurement at 5 GHz.

by (16). To make the two images comparable, we normalized the functions of the two images to $[0, 1]$. It can be seen that the scatterer located at (202.7,32.3) appears in Fig. 2 to be much darker compared with the scatterer at $(224.7, -14.0)$, whereas it is much brighter in Fig. 3. This difference verifies the near–far problem of the basic time-reversal and confirms that the MLE of scattering potential imaging is more balanced due to the proper scaling. In addition, many spurious local peaks can be observed in both images, which are the grating lobes since both the transmit and receive arrays in the experiment have antenna spacing much larger than half of the wavelength. We proposed a wideband imaging approach to exploit frequency diversity and resolved this spatial ambiguity under the sparse array setup. Interested readers are referred to [1].

VI. Conclusion

We demonstrated that basic time-reversal imaging is related to an MLE of the scattering potential under the assumption of a simplified single-scatterer physical model. We showed that the two imaging functions differ by a scaling factor, which is function of the imaging position. The basic time-reversal imaging exhibits the near–far problem, producing a weaker image for the area further away from the imaging array, whereas the MLE-based imaging is balanced due to the proper scaling.

Acknowledgment

The authors would like to thank J. M. F. Moura, D. Stancil, and J. Zhu for providing them with the experimental data in Section V.

References

[1] G. Shi and A. Nehorai, "Maximum likelihood estimation of point scatterers for computational time-reversal imaging," *Commun. Inf. Syst.*, vol. 5, no. 2, pp. 227–256, 2005.

[2] M. Fink, "Time-reversed acoustics," *Sci. Amer.*, vol. 281, pp. 91–97, Nov. 1999.

[3] M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, M. Tanter, J. L. Thomas, and F. Wu, "Time-reversed acoustics," *Rep. Progr. Phys.*, vol. 63, pp. 1933–1995, 2000.

[4] M. Fink and C. Prada, "Acoustic time-reversal mirrors, topical review," *Inverse Problems*, vol. 17, pp. R1–R38, 2001.

[5] L. Borcea, G. Papanicolaou, and C. Tsogka, "Theory and applications of time reversal and interferometric imaging," *Inverse Problems*, vol. 19, pp. 5139–5164, 2003.

[6] L. Borcea, C. Tsogka, G. Papanicolaou, and J. Berryman, "Imaging and time reversal in random media," *Inverse Problems*, vol. 18, pp. 1247–1279, 2002.

[7] J. H. Taylor, *Scattering Theory*. New York: Wiley, 1972.

[8] R. G. Newton, *Scattering Theory of Waves and Particles*. New York: McGraw-Hill, 1966.

[9] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. cas-52, no. 9, pp. 772–781, Sep. 1978.

[10] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, 6th ed. Burlington, MA: Elsevier Academic, 2005.

[11] A. J. Devaney, Super-Resolution Processing of Multi-Static Data Using Time Reversal and MUSIC, 2000 [Online]. Available: http://www.ece.neu.edu/faculty/devaney

[12] C. Prada and J. Thomas, "Experimental subwavelength localization of scatterers by decomposition of the time reversal operator interpreted as a covariance matrix," *J. Acoust. Soc. Amer.*, vol. 114, pp. 235–243, 2003.

[13] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramér–Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-37, pp. 720–741, May 1989.

[14] A. J. Devaney, "Basic time reversal imaging of CMU data set," 2004, DARPA progress rep.

Spectral Density Estimation Using Sharpened Periodograms

Fang Yao and Thomas C. M. Lee, *Senior Member, IEEE*

***Abstract*—This correspondence introduces the use of sharpened periodograms for spectral density estimation. It is shown analytically that spectrum estimates obtained from smoothing the sharpened periodograms enjoy higher order bias reduction when compared with the ordinary smoothed periodogram estimates. The promising numerical performances of using sharpened periodograms for spectral density estimation are illustrated via numerical experiments.**

***Index Terms*—Bias reduction, data sharpening, periodogram smoothing, sharpened periodograms, spectral density estimation, unbiased risk estimation.**

I. Introduction

In this correspondence, we study the problem of nonparametric spectral density estimation. In particular we propose using the so-called *data sharpening* technique [2], [3], [6] to help reduce the estimation bias. Data sharpening can be seen as a data preprocessing step that aims to achieve the following. It produces preprocessed data in such a way that when these preprocessed data are fed into certain standard and relatively simple estimation methods, the results are improved relative to the case when the original raw data were used. In other words, data sharpening can be applied to boost the performances of simple estimation methods while at the same time preserve the simplicity of such methods.

Manuscript received May 10, 2006; revised December 20, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Peter Handel.

F. Yao is with the Department of Statistics, University of Toronto, Toronto, ON M5S 3G3, Canada (e-mail: fyao@utstat.toronto.edu).

T. C. M. Lee is with the Department of Statistics, the Chinese University of Hong Kong, Shatin, N.T., Hong Kong and Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877 USA (e-mail: tlee@sta.cuhk.edu.hk).

Digital Object Identifier 10.1109/TSP.2007.896297

Data sharpening procedures have been constructed and studied in the contexts of probability density estimation [2], [6] and nonparametric regression [3]. Here, we extend and improve this technique in the following three directions. First, we apply data sharpening to a different estimation problem, namely, spectral density estimation, by proposing a sharpening procedure for preprocessing the periodogram of a stationary series. Second, unlike those "fixed" sharpening procedures previously studied in [2], [3], and [6], we introduce a tuning parameter in this new periodogram sharpening procedure that allows the data to be "sharpened" to different degrees. "Sharpened periodograms" produced from this procedure can be smoothed, say, by kernel methods to obtain nonparametric estimates of the spectrum of the series. Lastly, based on the idea of unbiased risk estimation, we develop an automatic method for simultaneously choosing the amounts of smoothing and sharpening. To the best of our knowledge, no automatic method has been proposed in the literature for selecting the amount of smoothing for any type of sharpened data. Furthermore, under some mild regularity conditions, we show that the estimate obtained from smoothing the sharpened periodogram enjoys a higher order bias reduction relative to the estimate obtained from smoothing the raw periodogram.

The rest of this correspondence is organized as follows. Background material is reviewed in Section II. Section III defines sharpened periodograms and illustrates its uses for spectral density estimation. Section IV discusses implementation issues. Theoretical and numerical results of our work are presented in Sections V and VI, respectively. Concluding remarks are offered in Section VII. Lastly proofs and technical details are deferred to the Appendix.

II. Background

Suppose $x_0, \dots, x_{2n-1}$ is a finite-sized realization of a real-valued, zero-mean stationary process $\{x_t\}$ with unknown spectral density f. Given the observations x_t's, our aim is to estimate f nonparametrically. This typically starts by computing the periodogram, defined as

$$I(\omega) = \frac{1}{2\pi \times 2n} \left| \sum_{t=0}^{2n-1} x_t \exp(-i\omega t) \right|^2, \quad i = \sqrt{-1}, \quad \omega \in [0, 2\pi).$$

To simplify notation, write $\omega_j = \pi j/n$, $f_j = f(\omega_j)$, and $I_j = I(\omega_j)$. As the spectral density f is symmetric about $\omega = \pi$, in this correspondence we shall focus on f_j for $j = 0, \dots, n-1$. Also, since f is periodic with period 2π, we have $f_{-j} = f_j$ and $I_{-j} = I_j$ for $j = 1, \dots, n-1$.

For the rest of this correspondence, we adopt the following model for I_j:

$$I_j = f_j \epsilon_j, \quad j = 0, \dots, n-1 \tag{1}$$

where the ϵ_j's are independent standard exponential random variables. We note that this model is only an approximation and does not reflect the exact relationship between f_j and I_j (e.g., see [1, Ch. 10]). However, due to its simplicity and accuracy, this model has been used by many previous authors (e.g., see [8], [9], [11], and [14]).

Under model (1), we have $E(I_j) = f_j$ and $\text{var}(I_j) = f_j^2$. Due to its unacceptably large variance, I_j is seldom used as an estimate of f_j. In order to reduce the variance and obtain a consistent estimate for f, one could compute the kernel estimator $\hat{f}_j$ for f_j by smoothing the periodogram

$$\hat{f}_j = \sum_{m=-n}^{2n-1} K_h(\omega_m - \omega_j) I_m \Bigg/ \sum_{l=-n}^{2n-1} K_h(\omega_l - \omega_j), \quad j = 0, \dots, n-1. \tag{2}$$

In the above, $K_h(\cdot) = (1/h)K(\cdot/h)$, where the $K(\cdot)$ is a univariate kernel function which is often taken as a symmetric density and the bandwidth h is a nonnegative smoothing parameter that controls the amount of smoothing. Automatic methods for selecting h can be found, for example, in [8] and [10]. Notice that $\hat{f}_j$ is a function of h, but for simplicity this dependence is suppressed in its notation. In many other kernel smoothing problems, the limits of the two summations in (2) are 0 and $n-1$. However, since in the present setting boundary effects can be handled by periodic smoothing, the limits are changed from 0 and $(n-1)$ to $(-n)$ and $(2n-1)$, respectively. Observe that the estimator $\hat{f}_j$ can also be interpreted as a weighted average of the I_j's, as it can be expressed as

$$\hat{f}_j = \sum_{m=-n}^{2n-1} W_{m-j} I_m \text{ with } W_{m-j} = \frac{K_h(\omega_m - \omega_j)}{\sum_{l=-n}^{2n-1} K_h(\omega_l - \omega_j)}. \tag{3}$$

Also observe that the weights W_m's sum to unity. Since $\hat{f}_j$ can be expressed a linear combination of I_j, this can also be seen as a form of the classical Blackman–Tukey estimator.

III. Sharpened Periodograms

It is straightforward to see that, due to the effects of local averaging, $\hat{f}_j$ tends to overestimate (or underestimate) f_j whenever f_j is near a local minimum (or maximum). In order to reduce such biases in the estimation of f, we propose smoothing the *sharpened periodogram*, to be defined next, instead of smoothing the periodogram I_j.

For a given $0 \le \alpha \le 1$, we define the corresponding sharpened periodogram as

$$\tilde{I}_{j,\alpha} = (1+\alpha)I_j - \alpha \hat{f}_j = (1+\alpha)I_j - \alpha \sum_{m=-n}^{2n-1} W_{m-j} I_m, \quad j = 0, \dots, n-1. \tag{4}$$

We call α the *sharpening parameter*. Notice that when $\alpha = 0$ the sharpened periodogram $\tilde{I}_{j,0}$ becomes the original periodogram I_j. We propose estimating f_j by smoothing the sharpened periodogram and denote the resulting estimator as $\tilde{f}_{j,\alpha}$, as follows:

$$\begin{aligned}\tilde{f}_{j,\alpha} &= \sum_{m=-n}^{2n-1} W_{m-j} \tilde{I}_{m,\alpha} \\ &= \sum_{m=-n}^{2n-1} W_{m-j} \left\{ (1+\alpha)I_m - \alpha \sum_{l=-n}^{2n-1} W_{l-m} I_l \right\}, \quad j = 0, \dots, n-1.\end{aligned} \tag{5}$$

Under some mild regularity conditions, we have shown in Theorem 1 that, when comparing to the ordinary kernel estimator $\hat{f}_j$ as in (2) or (3), this sharpened estimator $\tilde{f}_{j,\alpha}$ is able to reduce the bias to a higher order while, at the same time, it only inflates the variance by a constant factor.

In practice, the calculation of $\tilde{f}_{j,\alpha}$ requires the selection of two parameters: i) the bandwidth h (which determines the weights W_m's) and ii) the sharpening parameter α. The selection of these two parameters is critical as the quality of $\tilde{f}_{j,\alpha}$ is highly dependent on it. One sensible way for choosing these parameters is to choose the pair that aims to minimize the following L_2 risk $R(h, \alpha)$:

$$R(h,\alpha) = \frac{1}{n} E\left\{ \sum_{j=0}^{n-1} (f_j - \tilde{f}_{j,\alpha})^2 \right\} = \frac{1}{n} E\left(\|\boldsymbol{f} - \tilde{\boldsymbol{f}}_\alpha\|^2 \right) \tag{6}$$

where $\boldsymbol{f} = (f_0, \dots, f_{n-1})^T$, $\tilde{\boldsymbol{f}}_\alpha = (\tilde{f}_{0,\alpha}, \dots, \tilde{f}_{n-1,\alpha})^T$, and $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$ is the Euclidean norm associated with the corresponding inner product. Of course this unknown risk function $R(h, \alpha)$ cannot be directly minimized. One common strategy to overcome this problem is first to construct an unbiased estimator for $R(h, \alpha)$ and then choose

(h, α) as the pair that minimizes the resulting estimator. We have developed such an estimator, as follows:

$$\widehat{R}(h,\alpha) = \frac{\mathrm{RSS}(h,\alpha)}{n} + \frac{1}{n}\sum_{j=0}^{n-1}\left[2\left\{(1+\alpha)W_0 - \alpha\sum_{m=-n}^{2n-1}W_{m-j}^2\right\} - 1\right]\frac{I_j^2}{2} \quad (7)$$

where $\mathrm{RSS}(h,\alpha) = \sum_{j=0}^{n}(I_j - \tilde{f}_{j,\alpha})^2$ is the residual sum of squares. The first term can be treated as a measure for the bias of $\tilde{f}_{j,\alpha}$, while the second term is for the variance. It is shown in Theorem 2 (see Section V) that $\widehat{R}(h,\alpha)$ is an unbiased estimator for $R(h,\alpha)$. We propose choosing (h,α) as the joint minimizer of $\widehat{R}(h,\alpha)$.

The above idea of data sharpening has been applied to the context of nonparametric function estimation by [3]. However, in [3], the authors only considered the case when $\alpha = 1$ and did not provide any automatic method for choosing h. Thus, in addition to extending the data sharpening technique to periodogram smoothing, we have also advanced the sharpening technique by i) allowing the extent of sharpening to be varied via the introduction of the sharpening parameter α in the sharpening formula (4) and ii) making the whole estimation procedure completely automatic through the development of a practical method for choosing (h,α).

IV. Implementation Issues

In summary, the sharpened estimator $\tilde{f}_{j,\alpha}$ can be computed by the following steps.

1) Preselect a two-dimensional grid $\Theta = [h_{\min},\ldots,h_{\max}] \times [\alpha_{\min},\ldots,\alpha_{\max}]$ as the search space for the minimization of $\widehat{R}(h,\alpha)$.
2) Calculate the periodogram $I_0,\ldots,I_{n-1}$.
3) For each $(h,\alpha) \in \Theta$, calculate $\tilde{f}_{j,\alpha}$ using (5) and then $\widehat{R}(h,\alpha)$ using (7).
4) Denote the pair of (h,α) that minimizes $\widehat{R}(h,\alpha)$ over Θ as (h_0,α_0). The final estimator is obtained by (5) with $h = h_0$ and $\alpha = \alpha_0$.

Lastly, we compare the computational requirements of the unsharpened estimator $\hat{f}_j$ (2) and the sharpened estimator $\tilde{f}_{j,\alpha}$ (5). First, we note that the length of time, say η, required for computing $\hat{f}_j$ for a given h is roughly the same as the time for computing $\tilde{f}_{j,\alpha}$ for a given pair of (h,α). Denote, respectively, that numbers of elements in $[h_{\min},\ldots,h_{\max}]$ and $[\alpha_{\min},\ldots,\alpha_{\max}]$ as N_h and N_α. Typically a "best" h for calculating $\hat{f}_j$ is defined as the optimizer of some criterion. If this criterion is to be optimized over $[h_{\min},\ldots,h_{\max}]$ by a grid search, then the computation time for calculating $\hat{f}_j$ with an automatic selected h is approximately $N_h\eta$. Similarly, the computation time for calculating $\tilde{f}_{j,\alpha}$ with an automatic selected pair of (h,α) is approximately $N_h N_\alpha \eta$. Therefore, the computational time for calculating the sharpened estimator $\tilde{f}_{j,\alpha}$ is roughly N_α times the computational time for the unsharpened estimator $\hat{f}_j$. In practice, we use $\alpha = [0.0, 0.1, \ldots, 1.0]$, i.e., $N_\alpha = 11$.

V. Theoretical Properties

This section summarizes our theoretical findings. Proofs are deferred to the Appendix.

We first derive and compare the asymptotic biases and variances of the unsharpened estimator $\hat{f}_j$ defined in (3) and the sharpened estimator $\tilde{f}_{j,\alpha}$ defined in (5). We show that it is possible to reduce the bias of $\tilde{f}_{j,\alpha}$ to a higher order when comparing to the bias of $\hat{f}_j$, while the variance of $\tilde{f}_{j,\alpha}$ is only inflated by a multiplicative constant.

We begin with the asymptotic bias and variance of $\hat{f}_j$. We require the following assumptions for the spectral density $f(\cdot)$, the kernel function $K(\cdot)$, and the sequence of bandwidths $h = h(n)$, as follows:

A1) the second derivative of $f(\cdot)$, $f^{(2)}(\cdot)$, is bounded on $[0, 2\pi)$;
A2) K is a compactly supported, symmetric probability density with $\sigma_K^2 = \int u^2 K(u)du < \infty$, $\|K\|^2 = \int K^2(u)du < \infty$, and the first derivative $K^{(1)}(\cdot)$ is bounded on its support;
A3) $h \to 0, nh \to \infty$ and $\liminf nh^2 > 0$, as $n \to \infty$.

Lemma 1: Under the assumptions A1)–A3)

$$\mathrm{Bias}(\hat{f}_j) = \frac{1}{2}\sigma_K^2 f_j^{(2)} h^2 + o(h^2)$$
$$\mathrm{Var}(\hat{f}_j) = \frac{\|K\|^2 f_j^2}{nh} + o\left(\frac{1}{nh}\right) \quad (8)$$

where $\hat{f}_j$ is defined in (2).

Next, we consider the sharpened estimator $\tilde{f}_{j,\alpha}$. We first replace assumptions A1) and A2) with the following:

A1†) the fourth derivative of $f(\cdot)$, $f^{(4)}(\cdot)$, is bounded on $[0, 2\pi)$;
A2†) in addition to the conditions in A2), $\mu_4(K) = \int u^4 K(u)du < \infty$.

Theorem 1: Under the assumptions A1†), A2†) and A3)

$$\mathrm{Bias}(\tilde{f}_{j,\alpha}) = \frac{1}{2}(1-\alpha)\sigma_K^2 f_j^{(2)} h^2 + \frac{1}{4}\left\{\frac{1}{6}(1-\alpha)\mu_4(K) - \alpha\sigma_K^4\right\} f_j^{(4)} h^4 + o(h^4) \quad (9)$$

$$\mathrm{Var}(\tilde{f}_{j,\alpha}) = (1+\alpha)^2 \frac{\|K\|^2 f_j^2}{nh} + o\left(\frac{1}{nh}\right) \quad (10)$$

where $\tilde{f}_{j,\alpha}$ is defined in (5).

From (9), one can see that, for the sole purpose of bias reduction, the best choice for α is $\alpha = 1$. Such a choice of α gives $\mathrm{Bias}(\tilde{f}_{j,1}) = -\sigma_K^4 f_j^{(4)} h^4/4 + o(h^4)$; that is, $\mathrm{Bias}(\tilde{f}_{j,\alpha})$ is reduced to $O(h^4)$. However, instead of fixing $\alpha = 1$, in practice we advocate allowing the data to select α through the minimization of (7). It is because (9) is only an asymptotic expression and also the variance term (10) increases as α increases.

Finally, we have established the following property of our risk estimator $\hat{R}(h,\alpha)$.

Theorem 2: Under the model (1), the risk estimator $\hat{R}(h,\alpha)$ defined in (7) is an unbiased estimator of the risk function $R(h,\alpha)$.

VI. Numerical Experiments

This section reports results of numerical experiments that were conducted for evaluating the finite sample performance of the sharpened estimator $\tilde{f}_{j,\alpha}$.

Altogether three testing spectra and four different sample sizes were used. The three testing spectra are the mobile radio communication example of [9], the broadband MA(3) and the narrowband ARMA(4,4) examples of [13]. These three testing spectra are displayed in Fig. 1, and their complete specifications can be found in [9], [13]. The four sample sizes were $n = 128, 256, 512$ and 1024. For all experiments, the following kernel function was used: $K(x) = (3/4)(1-x^2), x \in [0,1]$. It is the optimal kernel of order (0, 2) derived in [4]. It is optimal in the sense that, under certain regularity conditions, it minimizes the L_2 distance between the true and the estimated f.

For each combination of testing spectrum and sample size (in total there are 12 such combinations), 200 independent series were simulated, and the corresponding periodograms were computed. Then, from each periodogram, the sharpened estimator $\tilde{f}_{j,\alpha}$ (5) was computed, where the pair of free parameters (h,α) was chosen as the joint minimizer of the risk estimator $\hat{R}(h,\alpha)$ (7). For comparison purposes, the unsharpened estimator $\hat{f}_j$ defined in (3) was also computed twice, with the following two bandwidths. The first one was chosen using the

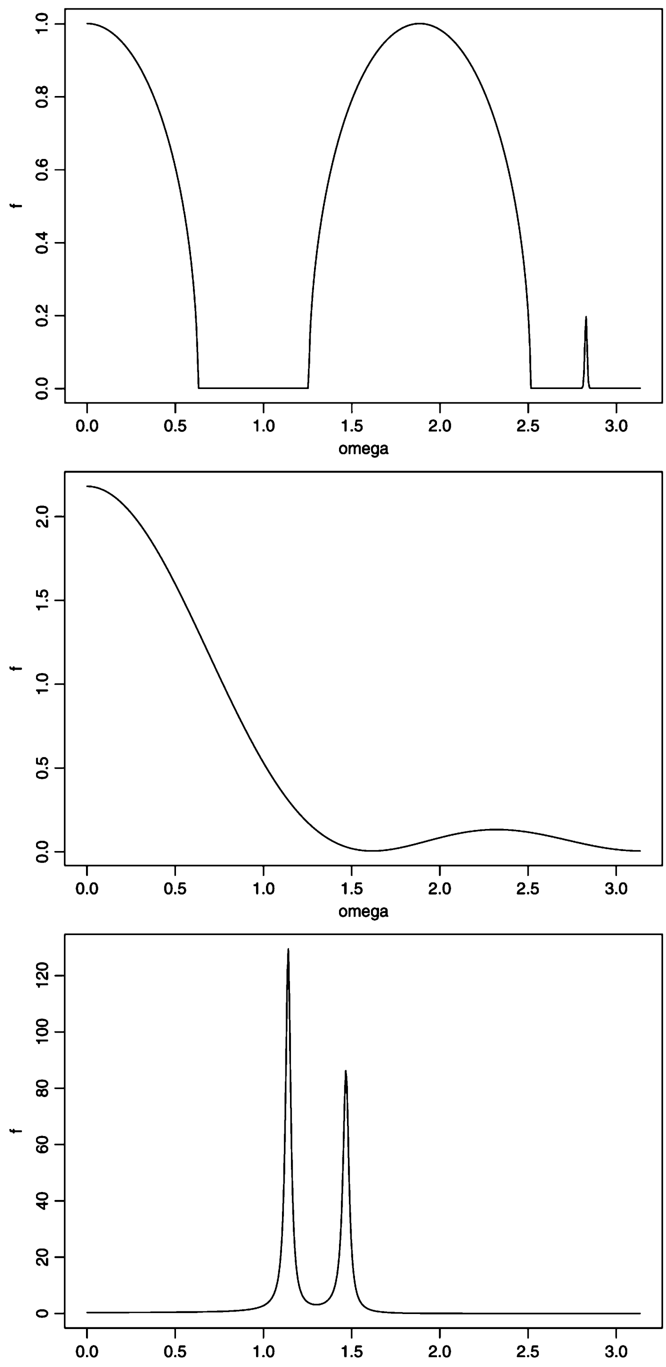

Fig. 1. Three testing spectra used in the numerical experiments. From top to bottom: The mobile radio communication example of [9], the broadband MA(3) example of [13], and the narrowband ARMA(4,4) example of [13].

Kullback–Leibler (KL) distance-based method of [7] while the second was the generalized cross-validation (GCV) method of [10]. Lastly, we also applied the recent cepstrum thresholding technique of [5] (see also [12]) to obtain a fourth estimate of f.

For each estimated spectrum, we calculated the corresponding mean-squared error (MSE), i.e., $(1/n)\sum(f_j - \tilde{f}_{j,\alpha})^2$, for the sharpened estimate and similarly for the other three unsharpened estimates. The averaged MSEs together with their standard errors for each of the 12 combinations of experimental setups are tabulated in Table I. To summarize the relative performances of the above four estimators, we ranked them in the following manner. First, paired t-tests were applied to test if the difference between the averaged MSE values of any two estimators is significant or not. The significance level used was 1.25%. If the averaged MSE value of a method is significantly less than the remaining three, it will be assigned a rank 1. If the averaged MSE value of a method is significantly larger than one but less than two methods, it will be assigned a rank 2, and similarly for ranks 3 and 4. Methods having nonsignificantly different averaged MSE values will share the same averaged rank. The resulting rankings are also tabulated in Table I.

The following empirical conclusions can be drawn. First, as the sample size increases, the performances of all estimators improve. Second, as the overall averaged pairwise t-test rankings for the proposed sharpened estimator, the KL estimator of [7], the GCV estimator of [10], and the cepstrum thresholding estimator of [5] are 1.08, 2.5, 2.42, and 4.0, respectively, there is some evidence suggesting that the proposed sharpened estimator is preferred. Third, for the mobile radio communication spectrum, the performances of the proposed estimator is much better than the rest, especially for large n. It is most likely due to the sharp spike feature of the spectrum, which, as a result of smoothing, can potentially cause large bias and the sharpening has successfully reduced it. Lastly, we want to comment on the poor performance of the cepstrum thresholding approach. As with most cepstrum approaches, this estimator aims to obtain a good estimate for the log of the spectrum; hence, it has a tendency to oversmooth sharp features in the spectrum.

VII. Concluding Remarks

In this correspondence, we have developed a new method for spectral density estimation via the smoothing of sharpened periodograms. We have shown theoretically that the smoothing of sharpened periodograms can reduce the bias to a higher order while at the same time only inflate the variance by a constant multiple. Using the idea of unbiased risk estimation, we have also constructed a method for choosing the two free tuning parameters involved in the estimation procedure, namely the bandwidth that determines the amount of smoothing and the sharpening parameter that controls the extent for sharpening. Numerical results suggest that sharpened estimates are superior to their unsharpened counterparts. One possible extension of the current work is to adopt other smoothing methods, such as wavelet shrinkage, to smooth the sharpened periodograms. This would require the development of new methods for bandwidth and sharpening parameter selection.

Appendix
Proofs

Proof of Lemma 1: We first show that A1) and A2) imply that, when $n \to \infty$

$$\sum_{m=-n}^{2n-1} W_{m-j}^2 = \frac{\|K\|^2}{nh} + o\left(\frac{1}{nh}\right)$$
$$\sum_{m=-n}^{2n-1} W_{m-j}(\omega_m - \omega_j)^2 = \sigma_K^2 h^2 + o(h^2). \tag{11}$$

Let $s_l(\omega_j;h) = (\pi/n)\sum_{m=-n}^{2n-1}(\omega_m - \omega_j)^l K_h(\omega_m - \omega_h)$, where $K_h(\cdot) = (1/h)K(\cdot/h)$. Since $K^{(1)}$ is bounded on its compact support, say $[-M, M]$, for large n, $s_l(\omega_j;h)$ can be well approximated by

$$\begin{aligned} s_l(\omega_j;h) &= \int_{-\pi}^{2\pi}(y-\omega_j)^l K_h(y-\omega_j)dy + o(1/n) \\ &= h^l \int_{(-\pi-\omega_j)/h}^{(2\pi-\omega_j)/h} u^l K(u)du + o(1/n) \\ &= h^l \int_{-M}^{M} u^l K(u)du + o(1/n) = \mu_l(K)h^l + o(1/n) \end{aligned}$$

where $\mu_l(K) = \int u^l K(u)du$. Analogously, let $s_l^*(\omega_j;h) = (\pi/n)\sum_{m=-n}^{2n-1}(\omega_m - \omega_j)^l K_h^2(\omega_m - \omega_h)$, one has $s_l^*(\omega_j;h) =$

TABLE I

AVERAGE MSEs AND PAIRWISE t-TEST RANKINGS (IN ITALICS) OBTAINED FROM THE NUMERICAL EXPERIMENTS. NUMBERS IN PARENTHESES ARE STANDARD ERRORS, MULTIPLIED BY 10^4, OF THE MSEs. THE FOUR ESTIMATORS WERE SHARP—THE PROPOSED SHARPENED ESTIMATOR, KL—THE UNSHARPENED ESTIMATOR WITH KL CHOICE OF BANDWIDTH [7], GCV—THE UNSHARPENED ESTIMATOR WITH GCV CHOICE OF BANDWIDTH [10], AND CEPS—THE CEPSTRUM THRESHOLDING APPROACH OF [5]. NOTICE THAT FOR BROADBAND MA(3) WITH $n = 128$, ALTHOUGH THE **OVERALL** AVERAGED MSE OF SHARP IS LARGER THAN THOSE OF KL AND GCV, THE **PAIRWISE** t-TEST HAS ASSIGNED IT A RANK OF 1

	mobile radio communication				broadband MA(3)				narrowband ARMA(4,4)			
n	Sharp	KL	GCV	Ceps	Sharp	KL	GCV	Ceps	Sharp	KL	GCV	Ceps
128	0.0376	0.0451	0.0806	0.365	0.0863	0.0638	0.0741	0.229	100	130	108	166
	(1.27)	(1.07)	(1.74)	(14.2)	(7.97)	(2.76)	(3.13)	(7.42)	(4170)	(1070)	(1900)	(703)
	1	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*1*	*3*	*2*	*4*
256	0.0228	0.0410	0.0622	0.217	0.0454	0.0459	0.0437	0.167	78.9	95.5	86.0	166
	(0.830)	(0.799)	(1.14)	(6.14)	(3.70)	(1.55)	(1.53)	(4.59)	(4070)	(1080)	(1460)	(818)
	1	*2*	*3*	*4*	*1.5*	*3*	*1.5*	*4*	*1.5*	*3*	*1.5*	*4*
512	0.0132	0.0384	0.0479	0.162	0.0209	0.0291	0.0261	0.0869	50.1	56.7	59.0	143
	(0.373)	(0.524)	(0.684)	(3.60)	(1.42)	(1.04)	(0.961)	(2.54)	(2020)	(719)	(755)	(920)
	1	*2*	*3*	*4*	*1*	*3*	*2*	*4*	*1*	*2*	*3*	*4*
1024	0.00798	0.0355	0.0344	0.105	0.0116	0.0171	0.0149	0.0460	29.8	36.2	40.6	115
	(0.185)	(0.332)	(0.381)	(1.87)	(0.734)	(0.484)	(0.422)	(1.33)	(1130)	(547)	(554)	(783)
	1	*3*	*2*	*4*	*1*	*3*	*2*	*4*	*1*	*2*	*3*	*4*

$\mu_l(K^2)h^{l-1} + o(1/n)$. Note from A3) $\liminf nh^2 > 0$ which allows $o(1/n)$ to be replaced by $o(h^2)$, $\sum_{m=-n}^{2n-1} W_{m-j}^2 = s_0^*(\omega_j;h)/\{ns_0(\omega_j;h)\}$ and $\sum_{m=-n}^{2n-1} W_{m-j}(\omega_m - \omega_j)^2 = s_2(\omega_j;h)/s_0(\omega_j;h)$; hence, (11) is proved.

We now derive (8). A direct application of the Taylor expansion gives

$$f_m = f_j + (\omega_m - \omega_j)f_j^{(1)} + \frac{1}{2}(\omega_m - \omega_j)^2 f_j^{(2)} + o(n^{-2}).$$

For convenience, we shall write "$\sum_{m=-n}^{2n-1}$" as "$\sum_m$" in the sequel unless defined otherwise. Thus

$$\begin{aligned}
\mathrm{Bias}(\hat{f}_j) &= \sum_m W_{m-j}\{E(f_m\epsilon_m) - f_j\} \\
&= \sum_m W_{m-j}(f_m - f_j) \\
&= \sum_m W_{m-j}\left\{(\omega_m - \omega_j)f_j^{(1)}\right. \\
&\qquad \left. + \frac{1}{2}(\omega_m - \omega_j)^2 f_j^{(2)} + o\left(\frac{1}{n^2}\right)\right\} \\
&= \frac{1}{2}\sigma_K^2 f_j^{(2)} h^2 + o(h^2) \\
\mathrm{Var}(\hat{f}_j) &= \mathrm{Var}\left(\sum_m W_{m-j} I_m\right) = \sum_m W_{m-j}^2 f_m^2 \\
&= \sum_m W_{m-j}^2\left\{f_j + (\omega_m - \omega_j)f_j^{(1)} + o\left(\frac{1}{n}\right)\right\}^2 \\
&= \sum_m W_{m-j}^2 f_j^2 + o\left(\frac{1}{nh}\right) = \frac{\|K\|^2 f_j^2}{nh} + o\left(\frac{1}{nh}\right).
\end{aligned}$$

Proof of Theorem 1: Under assumptions A1†) and A2†), by similar derivation of Lemma 1, and by applying the Taylor expansion up to the term h^4, it is straightforward to show that

$$\hat{f}_m = f_m + \frac{1}{2}\sigma_K^2 f_m^{(2)} h^2 + \frac{1}{24}\mu_4(K) f_m^{(4)} h^4 + R_{n,m}$$

where $\mu_4(K) = \int u^4 K(u)du$, and $R_{n,m}$ is the remaining term satisfying $E(R_{n,m}) = o(h^4)$ and $\mathrm{Var}(R_{n,m}) = O\{1/(nh)\}$. This implies that the sharpened periodograms can be expressed by

$$\begin{aligned}
\tilde{I}_{m,\alpha} = \{(1+\alpha)\epsilon_m - \alpha\}f_m &- \frac{1}{2}\alpha\sigma_K^2 f_m^{(2)} h^2 \\
&- \frac{1}{24}\alpha\mu_4(K) f_m^{(4)} h^4 - \alpha R_{n,m}. \qquad (12)
\end{aligned}$$

One also has $f_m^{(2)} = f_j^{(2)} + (\omega_m - \omega_j)f_j^{(3)} + (\omega_m - \omega_j)^2 f_j^{(4)}/2 + o(n^{-2})$, and $f_m^{(4)} - f_j^{(4)} = o(1)$. Therefore

$$\begin{aligned}
&\mathrm{Bias}(\tilde{f}_{j,\alpha}) \\
&= \sum_m W_{m-j}\left[E\{(1+\alpha)\epsilon_m - \alpha\}f_m - \frac{1}{2}\alpha\sigma_K^2 f_m^{(2)} h^2\right. \\
&\qquad \left. - \frac{1}{24}\alpha\mu_4(K) f_m^{(4)} h^4 - \alpha E(R_{n,m}) - f_j\right] \\
&= \sum_m W_{m-j}\left\{f_m - f_j - \frac{1}{2}\alpha\sigma_K^2 f_m^{(2)} h^2\right. \\
&\qquad \left. - \frac{1}{24}\alpha\mu_4(K) f_m^{(4)} h^4 + o(h^4)\right\} \\
&= \sum_m W_{m-j}\left[\left\{\frac{1}{2}(\omega_m - \omega_j)^2 f_j^{(2)} + \frac{1}{24}(\omega_m - \omega_j)^4 f_j^{(4)}\right\}\right. \\
&\qquad - \frac{1}{2}\alpha\sigma_K^2 h^2\left\{f_j^{(2)} + \frac{1}{2}(\omega_m - \omega_j)^2 f_j^{(4)} + o(n^{-2})\right\} \\
&\qquad \left. - \frac{1}{24}\alpha\mu_4(K) f_j^{(4)} h^4 + o(h^4)\right] \\
&= \frac{1}{2}(1-\alpha)\sigma_K^2 f_j^{(2)} h^2 \\
&\quad + \frac{1}{4}\left\{\frac{1}{6}(1-\alpha)\mu_4(K) - \alpha\sigma_K^4\right\} f_j^{(4)} h^4 + o(h^4).
\end{aligned}$$

The asymptotic variance of $\tilde{f}_{j,\alpha}$ is given by

$$\begin{aligned}
&\mathrm{Var}(\hat{f}_j) \\
&= \mathrm{Var}\left[\sum_m W_{m-j}\left[\{(1+\alpha)\epsilon_m - \alpha\}f_m - \frac{1}{2}\alpha\sigma_K^2 f_m^{(2)} h^2\right.\right. \\
&\qquad \left.\left. - \frac{1}{24}\alpha\mu_4(K) f_m^{(4)} h^4 - \alpha R_{n,m}\right]\right] \\
&= \sum_m W_{m-j}^2 (1+\alpha)^2 f_m^2 + o\left(\frac{1}{nh}\right) \\
&= \sum_m W_{m-j}^2 (1+\alpha)^2\left\{f_j + (\omega_m - \omega_j)f_j^{(1)}\right\}^2 + o\left(\frac{1}{nh}\right) \\
&= (1+\alpha)^2\frac{\|K\|^2 f_j^2}{nh} + o\left(\frac{1}{nh}\right).
\end{aligned}$$

Proof of Theorem 2: First, we stress that the unbiasness stated in this theorem only holds under model (1). The proof begins by noting that $E\{\mathrm{RSS}(h,\alpha)\} = E\{\sum_j (I_j - \tilde{f}_{j,\alpha})^2\} = \sum_j E(I_j^2 - 2I_j\tilde{f}_{j,\alpha} + \tilde{f}_{j,\alpha}^2)$. Since the ϵ_j's are independent standard exponentials, then $E(I_j) = f_j, E(I_j^2) = E(f_j^2\epsilon_j^2) = 2f_j^2$. From (3) and (5), one has $\tilde{I}_{j,\alpha} = (1+\alpha)I_j - \alpha\sum_m W_{m-j}I_m$ and

$$\tilde{f}_{j,\alpha} = (1+\alpha)\sum_{m=-n}^{2n-1} W_{m-j}I_m - \alpha\sum_{m,k=-n}^{2n-1} W_{m-j}W_{k-m}I_k.$$

Then

$$\begin{aligned}
E(I_j\tilde{f}_{j,\alpha}) &= E\left\{(1+\alpha)f_j\epsilon_j\sum_m W_{m-j}f_m\epsilon_m\right\} \\
&\quad - \alpha E(f_j\epsilon_j\sum_{m,k} W_{m-j}W_{k-m}f_k\epsilon_k) \\
&= (1+\alpha)\left(f_j\sum_{m\neq j} W_{m-j}f_m + 2W_0f_j^2\right) \\
&\quad - \alpha\left(2\sum_m W_{m-j}^2 f_j^2 + f_j\sum_{k\neq j} W_{m-j}W_{k-m}f_k\right) \\
&= \left\{(1+\alpha)W_0 - \alpha\sum_m W_{m-j}^2\right\}f_j^2 + f_j \\
&\quad \times\left\{(1+\alpha)\sum_m W_{m-j}f_m - \alpha\sum_{m,k} W_{m-j}W_{k-m}f_k\right\} \\
&= \left\{(1+\alpha)W_0 - \alpha\sum_m W_{m-j}^2\right\}f_j^2 + f_jE(\tilde{f}_{j,\alpha}).
\end{aligned}$$

Therefore

$$\begin{aligned}
E\{(I_j - \tilde{f}_{j,\alpha})^2\} &= 2f_j^2 - 2\left[\left\{(1+\alpha)W_0 - \alpha\sum_m W_{m-j}^2\right\}f_j^2\right. \\
&\qquad \left. + f_jE(\tilde{f}_{j,\alpha})\right] + E(\tilde{f}_{j,\alpha}^2) \\
&= E\{(f_j - \tilde{f}_{j,\alpha})^2\} \\
&\quad - \left[2\left\{(1+\alpha)W_0 - \alpha\sum_m W_{m-j}^2\right\} - 1\right]f_j^2,
\end{aligned}$$

and

$$\begin{aligned}
E\{\mathrm{RSS}(h,\alpha)\} &= nR(h,\alpha) - \sum_{j=0}^{n-1}\left[2\left\{(1+\alpha)W_0 - \alpha\sum_m W_{m-j}^2\right\} - 1\right]f_j^2.
\end{aligned}$$

Thus, $\hat{R}(h,\alpha)$ defined in (7) is an unbiased estimator of $R(h,\alpha)$.

Acknowledgment

The authors dedicate this correspondence to P. Brockwell for his retirement. Comments from the referees are gratefully acknowledged.

References

[1] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag, 1991.

[2] E. Choi and P. Hall, "Data sharpening as a prelude to density estimation," *Biometrika*, vol. 86, pp. 941–947, 1999.

[3] E. Choi, P. Hall, and V. Rousson, "Data sharpening methods for bias reduction in nonparametric regression," *Ann. Stat.*, vol. 28, pp. 1339–1355, 2000.

[4] T. Gasser, H.-G. Műller, and V. Mammitzsch, "Kernels for nonparametric curve estimation," *J. Roy. Stat. Soc. Series B*, vol. 47, pp. 238–252, 1985.

[5] E. Gudmundson, N. Sandgren, and P. Stoica, "Automatic smoothing of periodograms," in *Proc. 31st Int. Conf. Acoustics, Speech, Signal Processing*, 2006, pp. 504–507.

[6] P. Hall and M. C. Minnotte, "High order data sharpening for density estimation," *J. Roy. Stat. Soc. Series B*, vol. 64, pp. 141–157, 2002.

[7] J. Hannig and T. C. M. Lee, "Kernel smoothing of periodograms under Kullback–Leibler discrepancy," *Signal Process.*, vol. 84, pp. 1255–1266, 2004.

[8] T. C. M. Lee, "A simple span selector for periodogram smoothing," *Biometrika*, vol. 84, pp. 965–969, 1997.

[9] P. Moulin, "Wavelet thresholding techniques for power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 3126–3136, Nov. 1994.

[10] H. C. Ombao, J. A. Raz, R. L. Strawderman, and R. von Sachs, "A simple generalised crossvalidation method of span selection for periodogram smoothing," *Biometrika*, vol. 88, pp. 1186–1192, 2001.

[11] K. S. Riedel and A. Sidorenko, "Adaptive smoothing of the log-spectrum with multiple tapering," *IEEE Trans. Signal Process.*, vol. 44, no. 7, pp. 1794–1800, Jul. 1996.

[12] P. Stoica and N. Sandgren, "Smoothed nonparametric spectral estimation via cepstrum thresholding," *IEEE Signal Process. Mag.*, 2006, to be published.

[13] P. Stoica and T. Sundin, "Optimally smoothed periodogram," *Signal Process.*, vol. 78, pp. 253–264, 1999.

[14] G. Wahba, "Automatic smoothing of the log periodogram," *J. Amer. Stat. Assoc.*, vol. 75, pp. 122–132, 1980.