Multiple Window Time-Frequency and Time-Scale Analysis

Metin Bayram and Richard G. Baraniuk *

Department of Electrical and Computer Engineering
Rice University
Houston, TX 77251–1892, USA
Email: mebay@rice.edu, richb@rice.edu
URL: http://www-dsp.rice.edu

ABSTRACT

We propose an extension of Thomson's multiple window spectrum estimation for stationary random processes to the time-varying spectrum estimation of non-stationary random processes. Unlike previous extensions of Thomson's method, in this paper we identify and utilize optimally concentrated window and wavelet functions for the time-frequency and time-scale planes respectively. Moreover, we develop a statistical test for detecting and extracting chirping line components.

Keywords: multiple window method, spectrum estimation, non-stationary processes

1 INTRODUCTION

The theory of power spectrum estimation of stationary signals is well understood and widely applied.[1] However, these methods are inappropriate for the non-stationary signals that occur in important applications such as radar, sonar, acoustics, biology, and geophysics. These applications demand time-frequency representations that indicate how the spectral content of the process changes over time. Most research in time-frequency analysis has focused on deterministic signals. Only recently has attention turned to non-stationary random processes.[2–6]

Unlike the power spectrum for stationary random processes, there is no unique definition for the time-varying spectrum of a non-stationary random process. Because it satisfies a number of desirable properties, we choose the *Wigner-Ville spectrum* (WVS)[5] as our definition of the time-varying spectrum in this paper. The WVS is defined as the Fourier transform of the non-stationary auto-correlation function

$$r(t,\tau) = E[x^*(t-\tau/2)x(t+\tau/2)], \tag{1}$$

and can also be written as the expected value of the Wigner distribution[7] $W_x(t,f)$ of one realization of the process $x(t)$

$$\mathbf{W}_x(t,f) = \int_{-\infty}^{\infty} r(t,\tau)\, e^{-j2\pi f\tau}\, d\tau = E\{W_x(t,f)\}. \tag{2}$$

The problem of time-varying spectrum estimation can be stated as the estimation of $\mathbf{W}_x(t,f)$ given only one realization of the non-stationary process $x(t)$. Figure 1 shows a test signal composed of a chirp with sinusoidal instantaneous frequency in an additive bandpass Gaussian noise of linearly rising center frequency, and its ideal time-varying spectrum together with four different estimates of the WVS of the test signal.

A number of different WVS estimates have been proposed. The simplest is the empirical Wigner distribution $W_x(t,f)$ (WD) itself. However, while it is unbiased, it has very large (infinite in theory) variance and cross-

* This work was supported by the National Science Foundation, grant no. MIP–9457438, the Office of Naval Research, grant no. N00014–95–1–0849, and the Texas Advanced Technology Program, grant no. TX–ATP 003604–002.

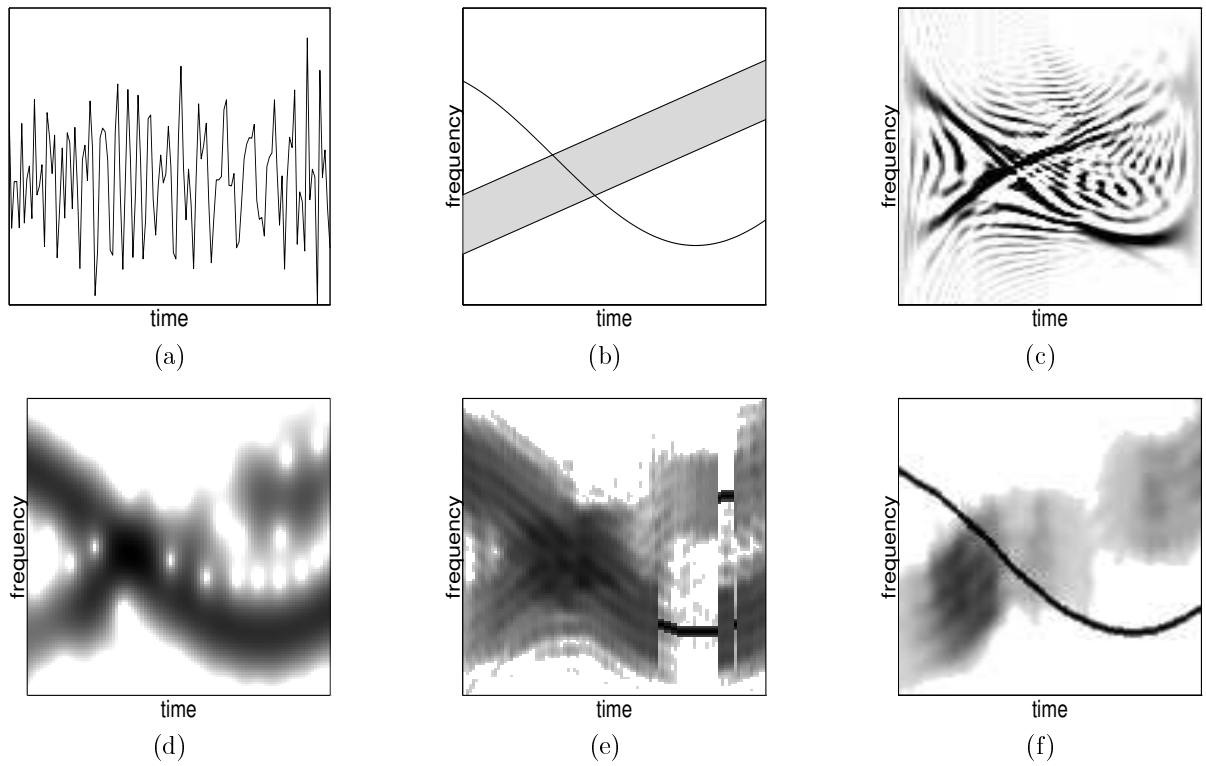


Figure 1: *(a) Test signal composed of a chirp with sinusoidal instantaneous frequency in an additive bandpass Gaussian noise of linearly rising center frequency. (b) Ideal representation. (c) Empirical Wigner distribution. (d) Spectrogram using a Gaussian window. (e) Sliding window Thomson's method as in.*[2–4] *(f) Multiple window (MW) method.*

components which are artifacts produced due to its bilinear nature. The large variance and the cross components of the empirical WD make it very difficult to interpret the WVS estimate, as can be seen in Figure 1(c).

To overcome the large variance problem of the WD, two-dimensional convolution of the Wigner distribution with a signal-independent smoothing kernel can be performed.[5] Sayeed and Jones[6] have developed a method for optimal kernel design for WVS estimation when the statistics of the process are known. They minimize the mean-squared error between the true WVS and the estimate, and also address the problem of local WVS estimation by allowing the smoothing kernel to be signal dependent. Whether the kernels are signal independent or signal dependent, the large amount of smoothing required to obtain a low variance WVS estimate can damage the resolution of line components in the data. Line components are deterministic chirping signals of the form $e^{j2\pi\gamma(t)}$, whose ideal time-frequency representations have the form $\delta(f-\gamma'(t))$. Figure 1(d) shows the effect of smoothing on the line components using the spectrogram with a Gaussian window.

Realizing that random and deterministic spectral components must be dealt with separately, Thomson introduced a powerful multiple window (MW) spectrum estimator for stationary signals in[8] to obtain a low variance spectrum without degrading the resolution of line components. The method uses a statistical significance test to detect and extract all sinusoids (the only stationary deterministic line components) from the data, computes a MW spectrum estimate of the sinusoid-free data with optimal windows, and reshapes the spectrum to account for the excised sinusoids. Because of its excellent performance, several groups have applied this technique, ad hoc, to non-stationary signals in a piecewise fashion.[2–4,9] There are two potential problems in doing this: 1) The windows used by Thomson are not optimal in a time-frequency setting, 2) The chirping rate of the line components must be very small so that they can be approximated as piece-wise sinusoids. Figure 1(e) shows the sliding window Thomson's method applied to the test signal. It can be seen that the method comes short of detecting the line component.

In this paper we refine the previous extensions of Thomson's method into an improved time-varying MW spectrum

estimate for non-stationary signals by identifying the optimal windows to use, and by developing a statistical test to detect line components of the form $e^{j2\pi\gamma(t)}$. Our method preserves the resolution of line components, has low variance, and offers fine control over the bias-variance trade-off. Figure 1(f) shows the method applied to the test signal.

This paper is organized as follows. In Section 2 we give a brief review of Thomson's MW method for stationary signals, and explain the essence of his significance test for sinusoids. Section 3 discusses MW time-frequency analysis, and identifies the optimal windows to use in the MW method. Section 3.4 extends the significance test to include all line components of the form $e^{j2\pi\gamma(t)}$. In Section 4, we extend the ideas in Section 3 to the time-scale plane, again identifying the optimal windows. Finally, Section 5 demonstrates the performance of our method. We begin with a review of Thomson's MW method for stationary signals.

2 THOMSON'S MULTIPLE WINDOW METHOD

The classical spectrum estimator for stationary signals, the periodogram, is defined as simply the squared magnitude of the Fourier transform of the data. While the periodogram suffers from a large variance, this variance can be reduced by cutting the data into blocks, computing a periodogram of each block, and then averaging the periodograms. However, this procedure also smoothes and biases the resulting spectrum estimate.

Inspired by the notion of averaging but displeased with the resulting bias, Thomson suggested computing several periodogram estimates of the *entire signal* using a set of different windows and then averaging the resulting periodograms to construct a spectrum estimate.[8] For a low variance, low bias estimate, he demanded that the windows be orthogonal (to minimize variance), and optimally concentrated in frequency (to minimize bias). The optimal windows satisfying these requirements for signals of finite extent are *the prolate spheroidal functions.*[10] In addition to multiple windows, Thomson also introduced into his estimate a separate procedure for deterministic sinusoidal components as mentioned above.

2.1 Summary of Thomson's Method

Thomson's MW method can be summarized in three steps:[8]

1. Detect and extract all significant sinusoids (stationary deterministic line components) in the data $x(t)$ using a statistical significance test (see Section 2.2) to obtain the part $y(t)$ of the data having a continuous spectrum (Thomson sets up the spectrum estimation problem in discrete-time.[8])
$$y(t) = x(t) - \{sinusoids\}. \tag{3}$$
2. Average K "orthogonal" periodogram estimates of $y(t)$ using prolate spheroidal data windows $\{v_k(t)\}$[10]
$$P_T(f) = \frac{1}{K}\sum_{k=0}^{K-1}\left|\int y(t)\, v_k(t)\, e^{-j2\pi f t}\, dt\right|^2. \tag{4}$$
These orthogonal functions are the eigenfunctions of a localization operator that band limits and then time limits functions. As windows, they are perfectly suited to stationary spectrum estimation, because they are simultaneously compactly supported in time and optimally concentrated in frequency. This concentration property results in a low bias estimate of the spectrum. In the above estimate the first K prolate windows are used for which the eigenvalues are very close to one. This guarantees a minimized introduced bias due to the averaging.[8]
3. Reshape the spectrum $P_T(f)$ to account for the sinusoids excised in Step 1.

2.2 Thomson's F-test For Sinusoids

Before we can extract the significant sinusoids from the data $x(t)$ as in (3), we must detect their presence and estimate their parameters.

We assume the signal model

$$x(t) = y(t) + \sum_i \mu(f_i)\, e^{j2\pi f_i t} \tag{5}$$

with $y(t)$ zero mean and Gaussian. Assuming that a sinusoid is present at frequency f, we estimate its complex amplitude and subtract it from the data to obtain an estimate of the "background" continuous spectrum around f. Comparing this power in the background spectrum with the power in the assumed sinusoid results in an F variance-ratio test $F(f)$.[8] If $F(f)$ exceeds a significance threshold, we say that a sinusoid exists at frequency f. When $F(f)$ exceeds the threshold when no sinusoid is present at frequency f, we say a *spurious peak* occurs.

Averaging orthogonal periodogram estimates reduces the variance of the MW power spectrum estimate by approximately K times compared to the variance of a single periodogram (in which $K = 1$). Furthermore, concentrated windows and sinusoid extraction keep resolution very high. These properties make Thomson's MW method the tool of choice for estimating the power spectrum of stationary random processes.

3 MULTIPLE WINDOW TIME-FREQUENCY ANALYSIS

The excellent performance of Thomson's MW method has led several groups to apply the method to time-varying spectrum estimation by simply sliding the estimate (4) along the signal and computing a MW spectrogram estimate about each time point.[2–4,9] While reasonably effective on certain classes of piecewise stationary signals, this approach suffers from two primary drawbacks. First, prolate spheroidal window functions have no inherent optimality properties in the joint time-frequency domain. Second, Thomson's F-test sinusoid extraction procedure fails on chirping line components of rapidly changing instantaneous frequency as we saw in Figure 1(e). In this and next sections, we will extend Thomson's MW method to the time-frequency and time-scale planes by identifying sets of optimal windows/wavelets, and by developing a linear-chirp extraction algorithm that better matches non-stationary line components.

3.1 Hermite Windows

The foundation of the stationary MW method rests on the fact that the prolate spheroidal functions are optimal windows for estimating the spectrum of a time-limited signal. This optimality does not carry over into time-frequency, however, since the prolate spheroidal functions treat the time-frequency plane as two separate spaces rather than as one geometric whole.[11–13]

For time-frequency signal analysis, it is natural to average over multiple orthogonal windows that are optimally concentrated in an appropriate time-frequency domain. To date, optimal orthogonal functions of this kind have been found only for a few very special domains. For instance, the Hermite functions are optimally concentrated in the circular time-frequency region of Figure 2(a)

$$\left\{(t,f): \quad t^2 + f^2 \le R^2\right\} \tag{6}$$

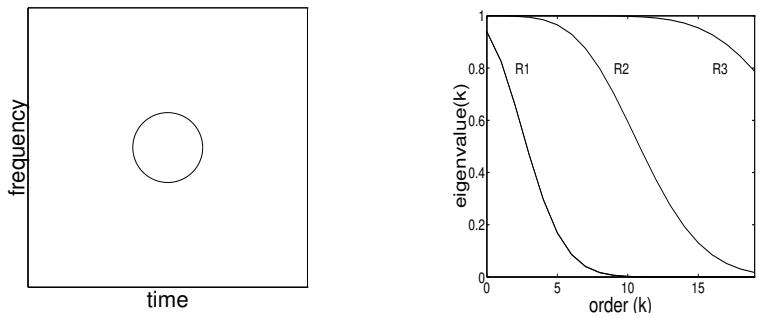
with R a constant, and thus treat all time-varying spectral features in the same fashion.[11–13] The Hermite functions, which are the eigenfunctions of the Fourier transform, are also the eigenfunctions of a localization operator over region (6).[11] This operator localizes functions in joint time-frequency as opposed to the localization operator for the prolate functions that localizes functions in time and frequency separately. The k-th order Hermite function is defined as

$$h_k(x) = \pi^{-1/4}(2^k k!)^{-1/2}\left(x - \frac{d}{dx}\right)^k e^{-x^2/2}, \qquad k = 0, 1, 2, \cdots. \tag{7}$$

The eigenvalues of the localization operator over the region (6) are given by

$$\lambda_k(R) = 1 - e^{\frac{-R^2}{2}} \sum_{i=0}^{k} \frac{1}{i!}\, 2^{-i}\, R^{2i}. \tag{8}$$

In Figure 2(b) we show the behavior of these eigenfunctions. Since these are the eigenfunctions of a localization operator, the closer the k-th eigenvalue is to one the better the concentration of the k-th order Hermite function. Hence, for a given R, there are only a few Hermite functions with good concentration in region (6).



(a) (b)

Figure 2: *(a) The circular concentration region (6) for the Hermite functions. (b) Eigenvalues of the localization operator over region (6)*[11] *for different values of R ($R_1 < R_2 < R_3$).*

Figure 3 shows the first three Hermite functions in time, their Fourier transforms, and their Wigner distributions. Comparing this figure with Figure 2(a), we see how the shape of the Wigner distribution of each Hermite function matches the shape of the concentration region. Hermite functions concentrated in elliptical regions are easily obtained by compressing or dilating the above functions $h_k(t)$.

3.2 Multiple Window WVS Estimate

Thomson's MW spectrum average (4) estimates the energy content of the signal at frequency f by projecting onto the windowed sinusoids $v_k(t)\, e^{j2\pi f t}$. By analogy, we estimate the energy content of a non-stationary signal at time t and frequency f by projecting onto the sliding windowed sinusoids $h_k(\tau - t)\, e^{j2\pi f \tau}$. The result can be written as the average of K Hermite-windowed spectrograms of the data

$$\widehat{\mathbf{W}}_x(t,f) = \frac{1}{K} \sum_{k=0}^{K-1} \left| \int y(\tau)\, h_k(\tau - t)\, e^{-j2\pi f \tau} d\tau \right|^2 . \tag{9}$$

The value of K is such that for a chosen R in (6) the first K eigenvalues in (8) are very close to one. The smaller the value of R, the smaller the bias of the estimate, and the smaller the value of K. Therefore, we see the bias-variance trade-off reflecting in the choice of K and R. The WVS estimator has low variance thanks to the averaging but also minimized bias due to the optimal concentration of the Hermite windows. The bias-variance trade-off can easily be controlled and optimized by changing the number of windows K.

3.3 Cohen's Class Interpretation

The MW WVS estimate (9) belongs to Cohen's class of time-frequency distributions.[7] We recall this class here as distributions that can be written as

$$C_x(t,f) = W_x(t,f) \;**\; \phi(t,f) \tag{10}$$

with $\phi(t,f)$ a kernel function. The kernel generating the spectrogram is precisely the Wigner distribution of the window function. Furthermore, the Wigner distribution of the k-th order Hermite function is the k-th order Laguerre function[14]

$$W_{h_k}(t,f) = L_k(t^2+f^2) = e^{-\frac{\pi}{2}\left(t^2+f^2\right)} \sum_{m=0}^{k} \frac{k!}{(k-m)!\, m!} \frac{\left(-\left(\pi\left(t^2+f^2\right)\right)\right)^m}{m!}. \tag{11}$$

Therefore, we have a closed form expression for the kernel ϕ corresponding to the MW WVS estimate (9) as a weighted sum of K Laguerre functions; in this interpretation, the MW WVS estimate reads

$$\widehat{\mathbf{W}}_x(t,f) = W_x(t,f) \;**\; \frac{1}{K} \sum_{k=0}^{K-1} L_k(t^2+f^2). \tag{12}$$

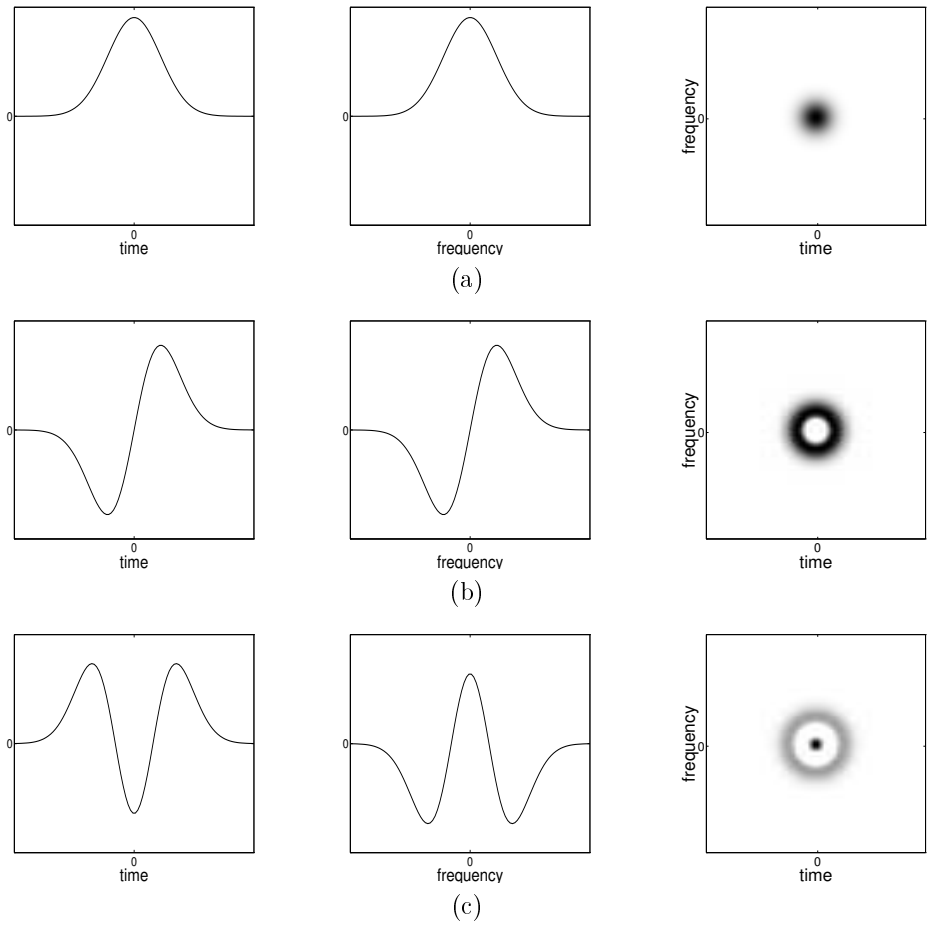


Figure 3: *From left to right the first three Hermite functions in the time domain, the frequency domain, and the time-frequency plane (WD): (a) 0-th order, (b) 1-st order (since the Fourier transform of the 1-st order Hermite function is purely imaginary, we plot the imaginary part), (c) 2-nd order.*

3.4 Extraction Of Line Components

As in Thomson's method for stationary signals, the averaging inherent in (9) will degrade the resolution of line components. Following Thomson's programme, we will first detect and extract all line components in the data before performing (9), and then reshape the estimate. We assume the signal model

$$x(t) = y(t) + \sum_i \mu_i(t)\, e^{j2\pi\gamma_i(t)} \tag{13}$$

with $y(t)$ zero mean and Gaussian.

A straightforward application of Thomson's sinusoid extraction algorithm to $x(t)$ as in[3,4] relies on an assumption that the chirp functions $e^{j2\pi\gamma_i(t)}$ can be closely approximated locally as sinusoids. Unfortunately, this is not the case for most chirping components; in these cases, the approach fails. In order to detect and extract highly non-stationary chirps, we have developed a statistical significance test for linear chirps of the form $e^{j2\pi(ft+ct^2)}$.[15] Linear chirps can closely approximate all but the most rapidly changing chirp functions.

The test for linear chirp components flows as in Section 2.2, except that the F variance-ratio test must be performed at each time t, frequency f, and chirp rate c. This results in a three dimensional F-test statistic $F(t, f, c)$.[15]

Due to the repeated application of the test, the number of spurious peaks in F increases far beyond that seen

in Thomson's method for stationary signals. These peaks must be suppressed in order to create a readable time-frequency image.

To suppress spurious peaks that peek above the significance threshold, we employ the following nonlinear cleaning algorithm:[15]

1. Threshold the data volume $F(t, f, c)$ and slice it along the chirp-rate dimension c.
2. For each fixed c_i, apply a nonlinear filter to $F(t, f, c_i)$ to remove peaks that have not coalesced into a region larger than the Heisenberg uncertainty principle mandates. (Intuition: spurious peaks are isolated in $F(t, f, c_i)$; true peaks lie along curves in $F(t, f, c_i)$.)
3. Combine the results from each c_i to obtain the final test statistic.

Although the linear chirp detection and extraction algorithm is computationally expensive, it is readily parallelizable.

4 MULTIPLE WINDOW TIME-SCALE ANALYSIS

For random processes containing high frequency components of short duration and low frequency components of long duration, time-frequency techniques are not appropriate. These types of processes are better matched by time-scale representations from the affine class.[16] The smoothing kernels in the affine class change with frequency to accommodate high frequency components of short duration and low frequency components of long duration. The regions of smoothing at different parts in the time-frequency plane for Cohen's class and the affine class are shown in Figure 4.

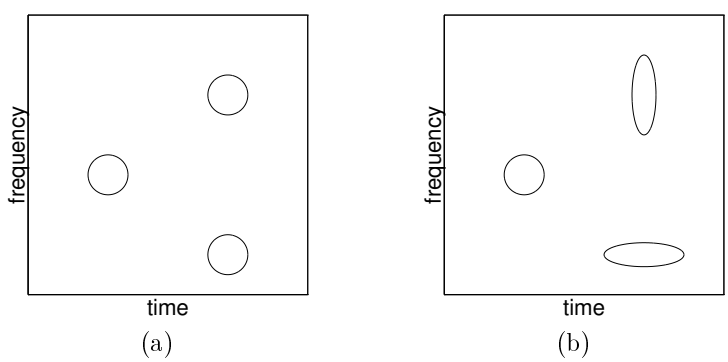


Figure 4: *Smoothing regions in the time-frequency plane in (a) Cohen's class, (b) Affine class.*

4.1 Morse Wavelets

The Morse wavelets[17–19] form a set of orthogonal functions that play a rôle in time-scale analogous to that of the Hermite windows in time-frequency. The k-th order Morse wavelet (In fact, Morse defined a special case of these wavelets for $\gamma = 1$,[18] but we still call the general class that is due to Daubechies and Paul[19] the Morse wavelets.) is defined in the frequency domain as

$$\Psi_k(f) = f^{\beta/2}\, e^{-f^\gamma/2}\, \frac{d^\beta}{df^\beta}\left[e^{f^\gamma}\, \frac{d^{\beta+k}}{df^{\beta+k}}\left(f^{\beta+k}\, e^{-f^\gamma}\right)\right], \tag{14}$$

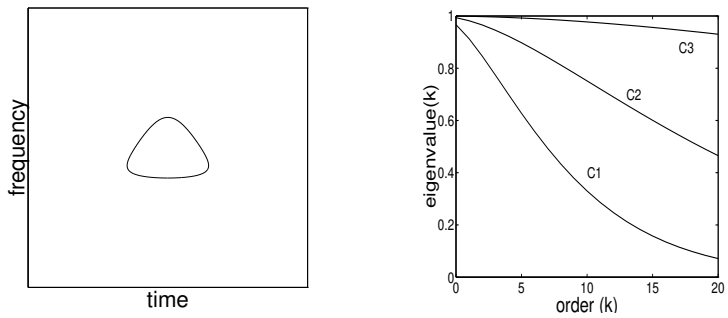
with $k = 0, 1, 2, \cdots$, $\beta > 0$ the degree of flatness at $f = 0$, and $\gamma > 0$. The Morse wavelets are the eigenfunctions of a localization operator over a tear-drop shaped region whose exact formula for any β and γ can be found in.[19] For the special case $\beta = \gamma = 1$, the Morse functions are mutually orthogonal and maximally concentrated in the time-frequency region which can be written as[17,19]

$$\left\{(t, f): \quad t^2 + \frac{9}{4f^2} + 1 \le \frac{3C}{|f|}\right\} \tag{15}$$

with C a constant. We show this region in Figure 5(a). For β and γ equal one, the eigenvalues of the bandpass localization operator corresponding to the Morse wavelets are given by

$$\lambda_k(C) = (k+1)\left(\frac{C-1}{C+1}\right)^{k+1}\left(\frac{2}{C+1}\,\frac{1}{k+1}\right). \tag{16}$$

There is no closed form expression for the eigenvalues for any other choices of β and γ. We show the behavior of the eigenvalues in Figure 5(b). We can see that for a given C, only a certain number of the eigenvalues are close to one, therefore only the first few Morse wavelets corresponding to those eigenvalues have excellent concentration. Figure 6 shows the first three Morse wavelets in time, their Fourier transforms, and their Wigner distributions.



(a) (b)

Figure 5: *(a) The tear-drop shaped concentration region (15) for the Morse wavelets for $\beta = \gamma = 1$. (b) Eigenvalues of the localization operator for different values of C ($C_1 < C_2 < C_3$) for $\beta = \gamma = 1$.*

As we stated earlier, for $\beta = \gamma = 1$ the Morse wavelets are the most concentrated functions on the time-frequency region (15). A comparison of the concentration properties of the Morse wavelets and the Hermite functions is relevant to see just how well-concentrated the Morse wavelets are in region (15).

The area of the concentration region (6) for the Hermite functions is given by

$$A_H = \pi R^2, \qquad R > 0. \tag{17}$$

The area of the concentration region (15) for the Morse wavelets on the other hand is given by

$$A_M = 3\pi(C-2), \qquad C > 0. \tag{18}$$

In Figure 7 we plot the eigenvalues of the Hermite functions and the Morse wavelets for three different values of the area of the concentration regions. The Hermite functions have more of their eigenvalues close to one compared to the Morse wavelets for a given area of concentration region. Therefore, the concentration properties of the Hermite functions on region (6) are much better than the concentration properties of the Morse wavelets on region (15). This means that the introduced bias due to the averaging using K Morse wavelets is larger than that due to the averaging using K Hermite functions.

4.2 Multiple Window WVS Estimate

We form a time-scale MW WVS estimate of the data $x(t)$ as the weighted average of the squares of K wavelet transforms (The squared magnitude of the continuous wavelet transform is called a *scalogram.*) using the Morse wavelets

$$\widehat{\mathbf{W}}_x(t,a) \;=\; \frac{1}{K}\sum_{k=0}^{K-1}\left| a^{-1/2}\int x(\tau)\,\psi_k\left(\frac{\tau-t}{a}\right)d\tau\right|^2 \tag{19}$$

where ψ_k is the k-th order Morse wavelet expressed in the time domain. The value of K determines how large the concentration region is. Therefore, we again see the bias-variance trade-off. For small variance a large K is needed, whereas for small bias a small C hence a small K is needed. The resulting estimate belongs to the affine class

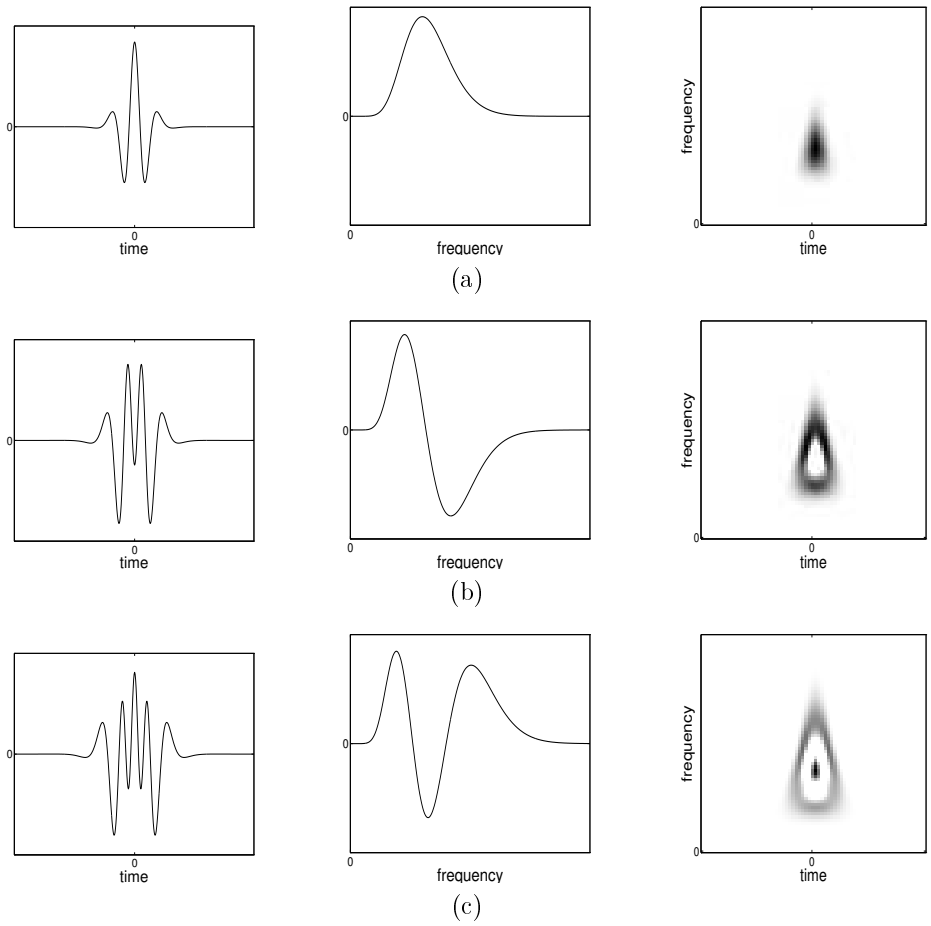


Figure 6: *From left to right the first three Morse wavelets in the time domain, the frequency domain (only the positive frequencies), and the time-frequency plane (WD): (a) 0-th order, (b) 1-st order, (c) 2-nd order.*

of time-scale covariant distributions.[16] Its kernel is the weighted sum of the Wigner distributions of the K Morse wavelets.

A similar line detection algorithm can be performed for the time-scale MW method. Due to the implementation of the scalogram, the line detection and extraction algorithm is much more expensive than that for the time-frequency case. Nonetheless, it can be done.

Lilly and Park have also considered multi-wavelet spectrum estimation using different wavelets.[20]

5 EXAMPLES

For our first example we refer to Figure 1. In this figure we illustrate the performance of the MW WVS estimate using a test signal composed of a chirp with sinusoidal instantaneous frequency in an additive bandpass Gaussian noise of linearly rising center frequency. The signal in time domain and its ideal representation in time-frequency is shown in 1(a) and 1(b). It is not possible to identify the components of the test signal from the empirical Wigner distribution due to its large variance. The spectrogram smoothes the Wigner distribution. Unfortunately the amount of smoothing needed to reduce the variance smears the line components excessively. A sliding version of Thomson's method as proposed in[2–4] does not perform well for this non-stationary data, since a local sine approximation to the chirping line component is inadequate. The time-frequency MW estimate of Figure 1(f) on the other hand has both

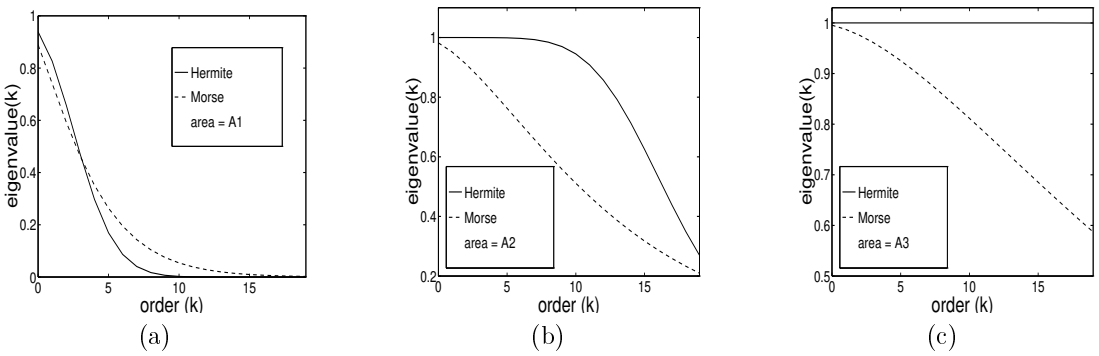


Figure 7: *The eigenvalues of the Hermite functions and the Morse wavelets for three different areas of the concentration regions ($A_1 < A_2 < A_3$).*

high resolution and low variance simultaneously. The computed variance of the MW WVS estimate is approximately $\frac{1}{4}$ that of the spectrogram, which agrees with the fact that four windows were used in the computation of the MW WVS estimate.

In Figure 8, we demonstrate the ability of the linear chirp detection algorithm to detect four hyperbolic chirps simultaneously. The data is a digitized 2.5 microsecond echo-location pulse emitted by the Large Brown Bat, Eptesicus Fuscus. There are 400 samples and the sampling period is 7 microseconds. Comparing the multiple window method with the other two plots, we see that the detection algorithm is able to pull out even the weakest high frequency line component successfully. The method even reveals the aliasing of the hyperbolic chirp due to under-sampling.

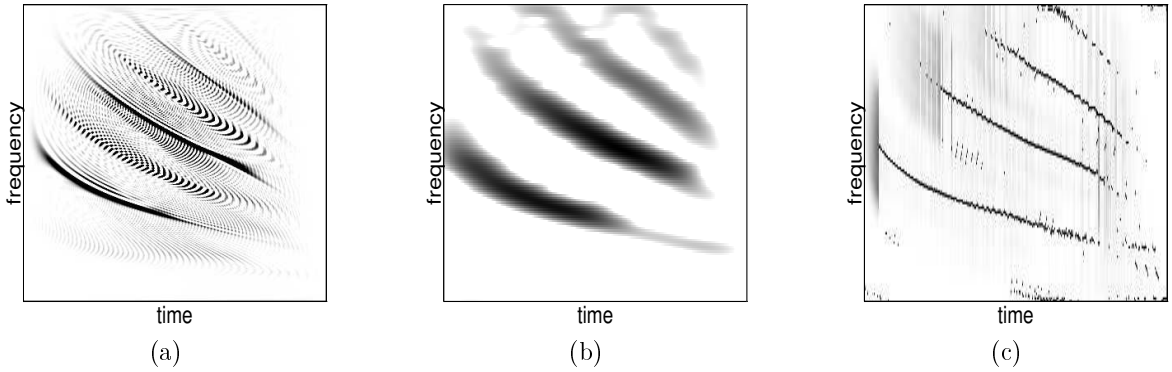


Figure 8: *Three spectrum estimates of the echo-location pulse emitted by the Large Brown Bat, Eptesicus Fuscus: (a) Empirical Wigner distribution. (b) Spectrogram. (c) Multiple window method.*

In Figure 9, we illustrate the performance of the time-scale MW method using a test signal of 256 points composed of two singularities in additive white Gaussian noise $n(t)$

$$x(t) = |t-64|^{-0.1} + |t-180|^{-0.1} + n(t). \tag{20}$$

The scalograms in Figure 9(c) and Figure 9(d) are incapable of capturing the structure of the signal shown in Figure 9(b). The MW method of Figure 9(e) on the other hand captures the structure of the signal in the presence of noise.

6 CONCLUSIONS

In this paper, we have motivated and developed multiple-window time-frequency and time-scale analysis for time-varying signals by extending Thomson's work[8] on multiple-window spectrum estimation for stationary signals. Our

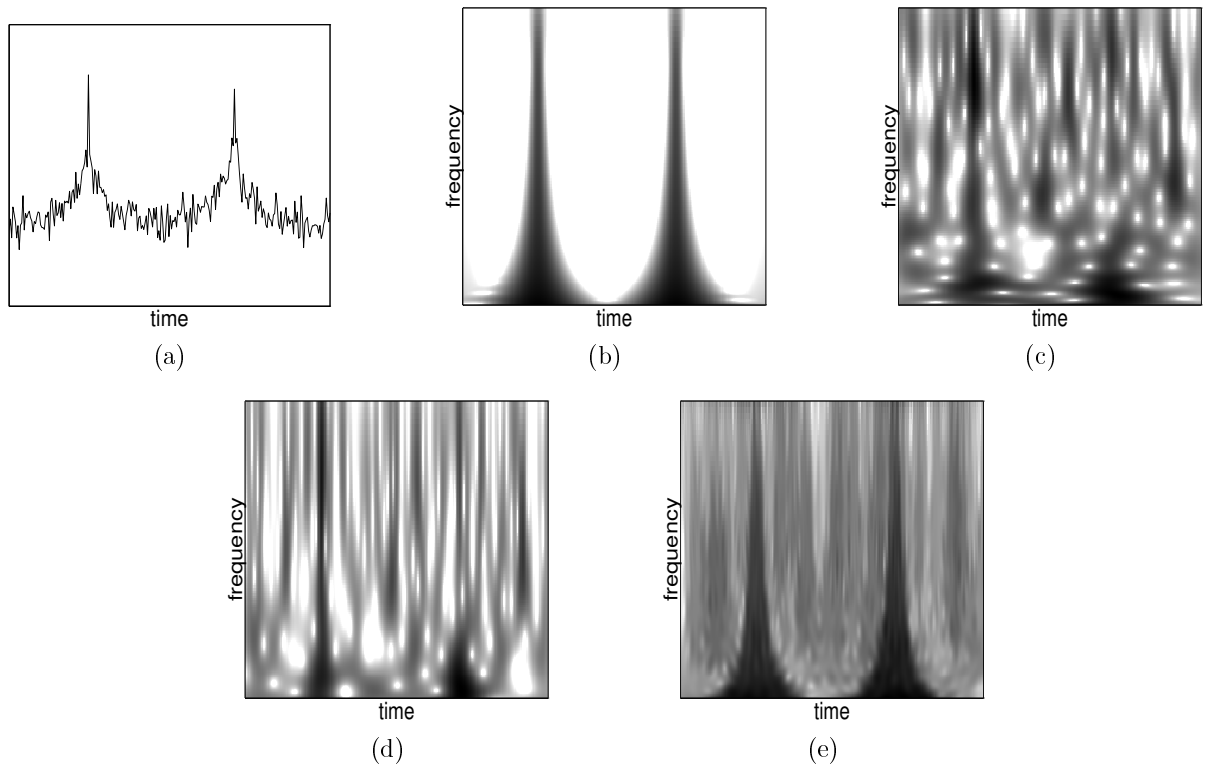


Figure 9: *(a) Test signal composed of two singularities in additive white Gaussian noise. (b) Scalogram of noise-free signal with a Morlet wavelet. (c) Scalogram of signal with a Morlet wavelet. (d) Scalogram of signal with the zeroth order Morse wavelet. (e) Multiple window (MW) method.*

contribution differs from the previous work done in multiple-window method spectrum estimation of time-varying signals in two major ways:

1. We have identified the optimal windows to use for time-varying spectrum estimation. They are the Hermite functions[11–13] for time-frequency analysis, and the Morse wavelets[17–19] for time-scale analysis. These windows are optimal in the sense that they are the most concentrated set of orthogonal functions in the time-frequency and time-scale planes resulting in low bias spectral estimates.

2. We have developed an algorithm to detect and extract non-stationary line components from the data by approximating them as piece-wise linear chirps. We then form the MW WVS estimate of the chirp-free data and reshape the spectrum to account for the excised line components. This preserves the resolution of the line components.

7 ACKNOWLEDGMENTS

We thank Patrick Flandrin for encouraging us to publish these results, and David Thomson, Paulo Gonçalvès, and Robert D. Nowak for stimulating discussions.

8 REFERENCES

[1] S. M. Kay, *Modern Spectral Estimation.* Prentice Hall, 1988.

[2] G. Frazer and B. Boashash, "Multiple window spectrogram and time-frequency distributions," in *Proc. IEEE Int. Conf.*

Acoust., Speech, Signal Processing — ICASSP '94, vol. IV, pp. 293–296, 1994.

[3] K. A. Farry, *Issues in Myoelectric Teleoperation of Complex Artificial Hands.* Ph.D. dissertation, Dep. Elec. Comput. Eng., Rice University, 1994.

[4] K. A. Farry, R. G. Baraniuk, and I. D. Walker, "Nonparametric, low bias, and low variance time-frequency analysis of myoelectric signals," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), Montreal, Canada*, Sept. 1995.

[5] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary random processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 1461–1470, Dec. 1985.

[6] A. M. Sayeed and D. L. Jones, "Optimal kernels for nonstationary spectral estimation," *IEEE Trans. Signal Processing*, vol. 43, pp. 478–491, Feb. 1995.

[7] L. Cohen, *Time-Frequency Analysis.* Englewood Cliffs, NJ: Prentice-Hall, 1995.

[8] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, pp. 1055–1096, Sept. 1982.

[9] D. J. Thomson. Personal Communication.

[10] D. Slepian and H. O. Pollack, "Prolate spheroidal wave functions, Fourier analysis and uncertainty," *Bell Syst. Tech. J.*, vol. 40, pp. 43–64, Jan. 1961.

[11] I. Daubechies, "Time-frequency localization operators: A geometric phase space approach," *IEEE Trans. Inform. Theory*, vol. 34, pp. 605–612, July 1988.

[12] P. Flandrin, "Maximum signal energy concentration in a time-frequency domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '88*, pp. 2176–2179, 1988.

[13] T. W. Parks and R. G. Shenoy, "Time-frequency concentrated basis functions," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '90*, vol. 5, pp. 2459–2462, 1990.

[14] G. B. Folland, *Harmonic Analysis in Phase Space.* Princeton, NJ: Princeton University Press, 1989.

[15] M. Bayram, "Multiple window time-frequency analysis," Master's thesis, Rice University, May 1996.

[16] O. Rioul and P. Flandrin, "Time-scale energy distributions: A general class extending wavelet transforms," *IEEE Trans. Signal Processing*, vol. 40, pp. 1746–1757, July 1992.

[17] I. Daubechies, *Ten Lectures on Wavelets.* New York: SIAM, 1992.

[18] P. Morse, "Diatomic molecules according to the wave mechanics II. Vibrational levels," *Physical Review*, vol. 34, pp. 57–64, July 1929.

[19] I. Daubechies and T. Paul, "Time-frequency localization operators — a geometric phase space approach: II. The use of dilations," *Inverse Problems*, no. 4, pp. 661–680, 1988.

[20] J. M. Lilly and J. Park, "Multiwavelet spectral and polarization analyses of seismic records," *Geophys. J. Int.*, no. 122, pp. 1001–1021, 1995.