# Multiple Window Time-Varying Spectrum Estimation

*Metin Bayram and Richard Baraniuk*

## 1 Summary

We overview a new non-parametric method for estimating the time-varying spectrum of a non-stationary random process. Our method extends Thomson's powerful multiple window spectrum estimation scheme to the time-frequency and time-scale planes. Unlike previous extensions of Thomson's method, we identify and utilize optimally concentrated Hermite window and Morse wavelet functions and develop a statistical test for extracting chirping line components. Examples on synthetic and real-world data illustrate the superior performance of the technique.

## 2 Introduction

Many methods exist for estimating the power spectra of stationary signals [1]. However, these methods are insufficient 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 power spectrum changes over time. To date, research in time-frequency analysis has mainly focused on deterministic signals. Only more recently has attention turned to non-stationary random processes [2–13].

Unlike the power spectrum for stationary random processes, there is no unique definition for the time-varying spectrum of a nonstationary random process $\mathbf{x}$. Perhaps the best compromise is the *Wigner-Ville spectrum* (WVS) $\mathsf{W}_{\mathbf{x}}$ [9]. Given the instantaneous auto-correlation function

$$r_{\mathbf{x}}(t,\tau) := \mathsf{E}[\mathbf{x}^*(t-\tau/2)\,\mathbf{x}(t+\tau/2)], \tag{2.1}$$

the WVS is defined as its Fourier transform

$$\mathsf{W}_{\mathbf{x}}(t,f) := \int r_{\mathbf{x}}(t,\tau)\,e^{-j2\pi f\tau}\,d\tau. \tag{2.2}$$

Alternatively, the WVS can be defined as the expected value of the empirical *Wigner distributions* (WDs) $\mathbf{W}_{\mathbf{x}}$ [14, 15] of the realizations of the process

$$\mathsf{W}_{\mathbf{x}}(t,f) = \mathsf{E}[\mathbf{W}_{\mathbf{x}}(t,f)] = \mathsf{E}\left[\int \mathbf{x}^*(t-\tau/2)\,\mathbf{x}(t+\tau/2)e^{-j2\pi f\tau}\,d\tau\right]. \tag{2.3}$$

In this framework, the problem of time-varying spectrum estimation can be stated as estimating the WVS $\mathsf{W}_{\mathbf{x}}$ given only one realization of the nonstationary process $\mathbf{x}$.

A number of different WVS estimates have been proposed. The simplest is the empirical WD $\mathbf{W}_{\mathbf{x}}$ itself. However, while it is unbiased, it has very large (infinite in theory) variance and cross-components — artifacts produced due to its quadratic nature [9]. Figure 1 plots a test signal composed of a deterministic FM chirp with sinusoidal instantaneous frequency [14, 15] submerged in a realization of an additive bandpass Gaussian process of linearly rising center frequency, its ideal time-varying spectrum, and four different WVS estimates. The large variance and the cross-components cloud the WD's interpretation, as can be seen in Figure 1(c).

To reduce the variance of the WVS estimate at the expense of some bias, the WD can be smoothed through convolution with a 2-D kernel function [9]. This bias-variance tradeoff is well illustrated by the *spectrogram* estimator of Figure 1(d). (The spectrogram smoothing kernel is the Wigner distribution of the analysis window [14, 15].)

Unfortunately, the large amount of smoothing required to obtain a low variance WVS estimate damages the resolution of deterministic chirping signals of the form $e^{j2\pi\gamma(t)}$, whose ideal time-frequency representations have the form $\delta(f-\gamma'(t))$. In this paper, we focus on estimating the WVS of mixed stochastic/deterministic signals of the form

$$\mathbf{x}(t) = \mathbf{y}(t) + \sum_i \mu_i(t)\, e^{j2\pi\gamma_i(t)}, \tag{2.4}$$

with $\mathbf{y}$ a zero-mean, nonstationary, Gaussian random process and $\mu_i(t)\, e^{j2\pi\gamma_i(t)}$ a deterministic "chirp" signal with instantaneous amplitude $\mu_i(t)$ and instantaneous phase $\gamma_i(t)$. For inspiration, we turn to the seminal stationary spectrum estimation work of Thomson [16].

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 [16] to obtain a low variance spectrum without degrading the resolution of line components. His method uses a statistical significance test to detect and extract stationary deterministic line components (sinusoids) from the data, computes a MW spectrum estimate of the sinusoid-free data using a set of 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 nonstationary signals in a sliding-window fashion [2–4, 17]. There are two potential problems to such an approach: 1) the windows used by Thomson are not optimal in a joint time-frequency setting and 2) the chirping rates of the line components must be very low (for them to be well approximated as piecewise sinusoidal). Figure 1(e) shows the sliding-window

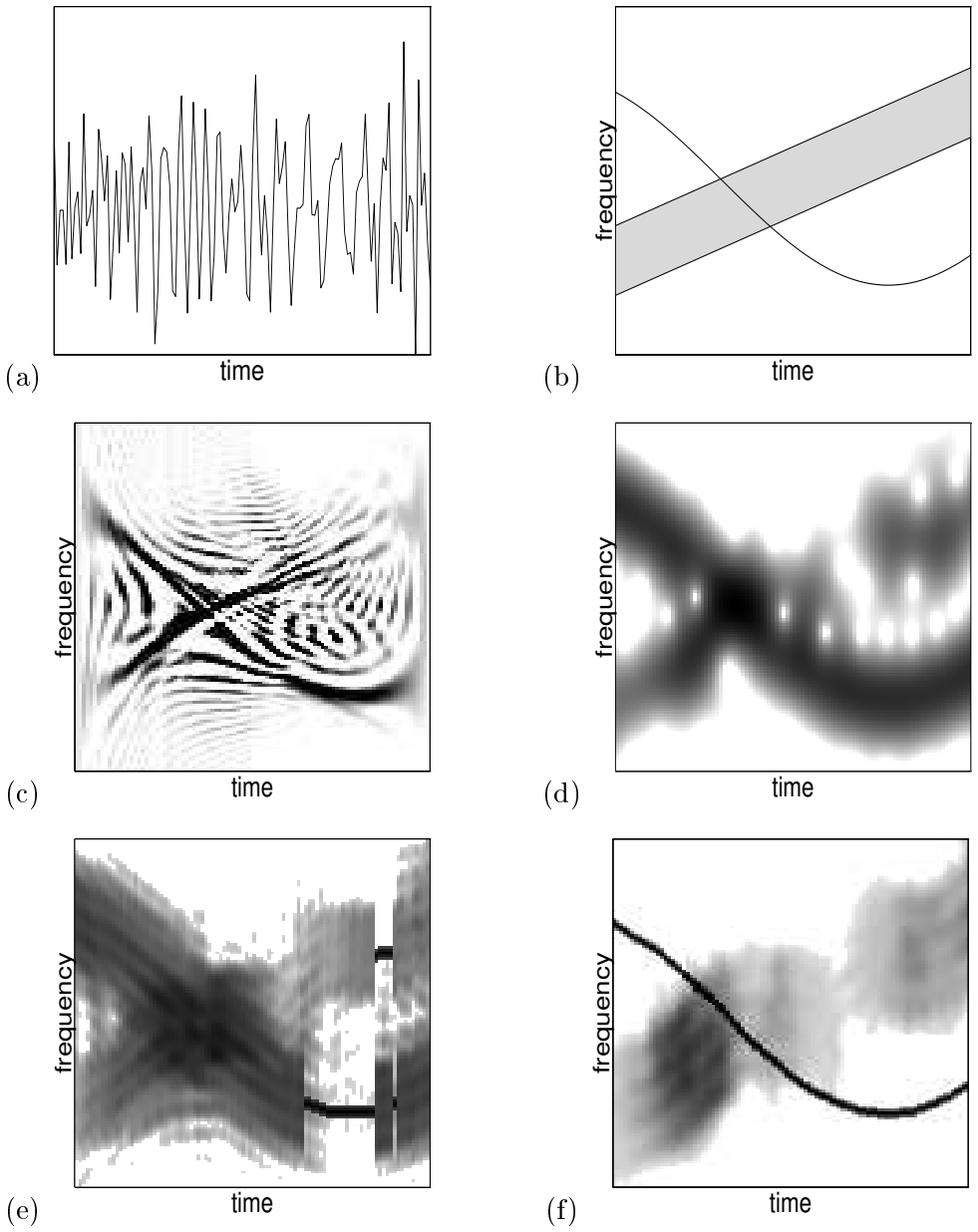



Figure 1: *(a) Test signal* $\mathbf{x}$ *composed of a chirp with sinusoidal instantaneous frequency plus an additive bandpass Gaussian noise of linearly rising center frequency. (b) Ideal representation. (c) Empirical Wigner distribution* $\mathbf{W_x}$*. (d) Spectrogram using a Gaussian window. (e)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

Thomson's method applied to the test signal. We see clearly that the method comes short of detecting and extracting the nonstationary line component (except where its chirp rate becomes small).

In this chapter we overview a refinement of the previous extensions of Thomson's method into an improved, unified time-varying MW spectrum estimate for nonstationary signals of the form (2.4) [5–8]. We will identify the optimal windows and develop a statistical test to detect and extract the time-varying line components. 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 chapter is organized as follows. In Section 3 we give a brief review of Thomson's MW method for stationary signals and explain the essence of his statistical test for sinusoids. Section 4 discusses MW time-frequency analysis and identifies the optimal (Hermite) windows to use in the MW method. Section 5 extends the significance test to include rapidly varying line components of the form $e^{j2\pi\gamma(t)}$. Section 6 extends the ideas in Section 4 to the time-scale plane, again identifying the optimal (Morse) windows. In Section 7 we demonstrate the performance of the estimators and, in Section 8, we discuss a number of related techniques and recent extensions of our approach. Section 9 concludes with a discussion.

## 3 Thomson's Multiple Window Method

Here we overview the key elements of Thomson's method for stationary power spectrum estimation, reformulated in continuous time. The stationary version of the signal model (2.4) reads

$$\mathbf{x}(t) = \mathbf{y}(t) + \sum_i \mu(f_i)\, e^{j2\pi f_i t + \rho_i} \tag{3.1}$$

with $\mathbf{y}$ a zero-mean, stationary, Gaussian random process having a continuous power spectrum and $\mu(f_i)\, e^{j2\pi f_i t + \rho_i}$ a deterministic sinusoid having a line power spectrum.

The classical spectrum estimator for stationary signals, the *periodogram*, is defined as simply the squared magnitude of the Fourier transform of a windowed version of the data

$$\widehat{\mathsf{P}}_{\mathbf{x}}(f) := \left| \int \mathbf{x}(t)\, w(t)\, e^{-j2\pi f t}\, dt \right|^2 \tag{3.2}$$

with $w(t)$ the window function. While the periodogram suffers from large variance, this variance can be reduced by cutting the data into blocks, computing a periodogram of each block, and then averaging the periodograms [18]. However, this procedure also smears and biases the resulting spectrum estimate. The bias/variance tradeoff is clear: reducing the variance necessitates averaging over a larger number of shorter blocks, which increases the bias.

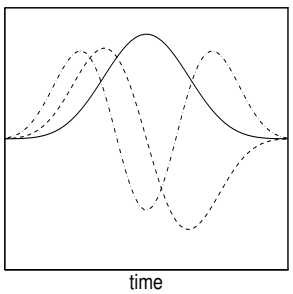



Figure 2: *The first three prolate spheroidal wave functions in the time domain.*

### 3.1 Summary of Thomson's Method

Inspired by the notion of averaging but displeased with the resulting bias, Thomson suggested computing several periodograms of the *entire signal* using a set of different windows and then averaging the resulting periodograms [16]. 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 wave functions* [16, 19] (see Figure 2). 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. In addition to averaging over multiple windows, Thomson also introduced a separate pre-estimate for deterministic sinusoidal components.

Thomson's MW method consists of three main steps [16]:

1. Detect and extract all significant sinusoids (stationary deterministic line components) in the data $\mathbf{x}$ using a statistical significance test (see Section 3.2) to obtain an estimate $\widehat{\mathbf{y}}$ of the part having a continuous spectrum

$$\widehat{\mathbf{y}} = \mathbf{x} - \{\mathit{sinusoids}\}. \tag{3.3}$$

2. Average $K$ "orthogonal" periodogram estimates of $\widehat{\mathbf{y}}$ using prolate spheroidal data windows $\{v_k\}$ [16, 19][1]

$$\widehat{\mathsf{P}}_T(f) := \frac{1}{K} \sum_{k=0}^{K-1} \left| \int \widehat{\mathbf{y}}(t)\, v_k(t)\, e^{-j2\pi f t}\, dt \right|^2. \tag{3.4}$$

   The concentration of the prolate windows in frequency results in a low bias estimate of the spectrum.

[1] Thomson actually weights the periodogram computed with window $v_k$ with the reciprocal of the corresponding prolate spheroidal eigenvalue $\lambda_k$ [16]. However, since these are typically very close to 1, we will neglect them until Section 8.

3. Reshape the spectrum estimate $\widehat{\mathsf{P}}_T$ to account for the sinusoids excised in Step 1.

### 3.2 Thomson's $F$-test For Sinusoids

Before we can extract the significant sinusoids from the data $\mathbf{x}$ as in (3.3), we must detect their presence and estimate their parameters. Assume the signal model (3.1) and define the $k$-th *eigenspectrum* $\chi_k$ as the Fourier transform of the windowed original data

$$\chi_k(f) := \int \mathbf{x}(t)\, v_k(t)\, e^{-j2\pi f t}\, dt. \tag{3.5}$$

The expected value of $\chi_k$ at frequency $f_i$ is given by

$$\mathsf{E}[\chi_k(f_i)] = \mu(f_i)\, V_k(0), \tag{3.6}$$

with $V_k$ the Fourier transform of $v_k$. Using a simple linear regression, the complex amplitude $\mu(f_i)$ of each possible sinusoid can thus be estimated as

$$\widehat{\mu}(f_i) = \frac{\sum_{k=0}^{K-1} V_k(0)\, \chi_k(f_i)}{\sum_{k=0}^{K-1} V_k^2(0)}. \tag{3.7}$$

The eigenspectra yield a simple statistical test for whether sinusoids are actually present in the data. Assuming that a sinusoid is present at frequency $f_i$ with complex amplitude $\widehat{\mu}(f_i)$, we subtract it from the data to obtain an estimate of the "background" continuous spectrum around $f_i$. Comparing the power in the background spectrum with the power in the assumed sinusoid yields an $F$ variance-ratio test with 2 and $2K-2$ degrees of freedom for the significance of the estimated line component [16]. Defining

$$F(f_i) := \frac{(K-1)\ |\widehat{\mu}(f_i)|^2 \sum_{k=0}^{K-1} V_k(0)^2}{\sum_{k=0}^{K-1} |\chi_k(f_i) - \widehat{\mu}(f_i)\, V_k(0)|^2}, \tag{3.8}$$

if $F(f_i)$ exceeds a significance threshold then we say that a sinusoid exists at frequency $f_i$.

The probability of missing a sinusoid increases with the threshold. On the other hand, the false alarm probability increases with decreasing threshold. False alarms give rise to *spurious peaks* in the spectrum estimate. For very closely spaced sinusoids, the above $F$ test fails. For such situations Thomson

suggests a double $F$ test that searches for the existence of two sinusoids at a time instead of one [16].

*Summary:* 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 $K = 1$) [16]. Furthermore, the concentrated prolate windows and sinusoid extraction ensure high resolution. These properties make Thomson's MW method the tool of choice for estimating the power spectra of stationary random processes.

# 4 Multiple Window Time-Frequency Analysis

The excellent performance of Thomson's MW method for stationary signals has led several groups to apply the method to time-varying spectrum estimation by simply sliding the estimator along the signal [2–4,17]. While reasonably effective for 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 Section, we will begin a more thorough extension of Thomson's MW method to the time-frequency plane by first identifying an optimal set of windows [5–8].

## 4.1 Hermite Windows

One of Thomson's key insights is to smooth the spectrum estimate using orthogonal windows that are concentrated in the smallest possible region in the domain of interest (the frequency domain in his case). The foundation of the stationary MW method rests on the fact that the prolates are the most frequency-concentrated of all sets of orthogonal, time-limited windows. For time-frequency signal analysis, it is clear that we should average over multiple orthogonal windows that are optimally concentrated in an appropriate time-frequency domain. The optimality of the prolate functions is not as natural in time-frequency, since these functions treat the time-frequency plane as two separate spaces rather than as one geometric whole [20–23].

More natural for time-frequency are the *Hermite functions*, defined by

$$h_k(t) := \pi^{-1/4}(2^k k!)^{-1/2} \left(t - \frac{d}{dt}\right)^k e^{-t^2/2}, \qquad k = 0, 1, 2, \ldots \tag{4.1}$$

The zeroth-order Hermite function is the Gaussian. The Hermite functions are optimally concentrated in the circular time-frequency region

$$\left\{(t, f): \quad t^2 + f^2 \leq R^2\right\} \tag{4.2}$$

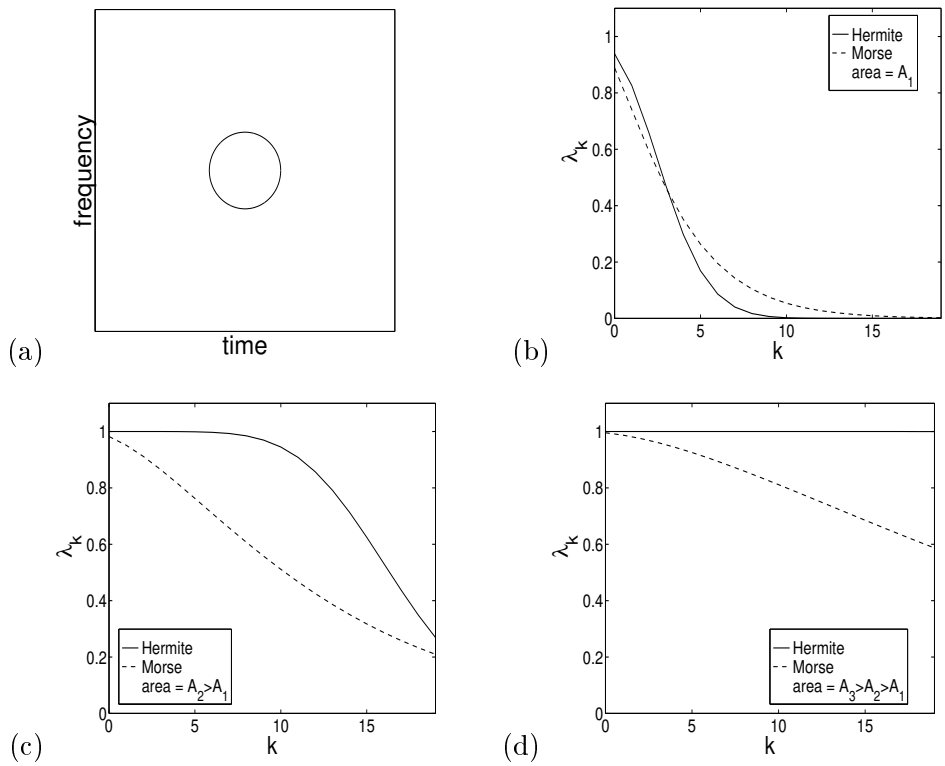


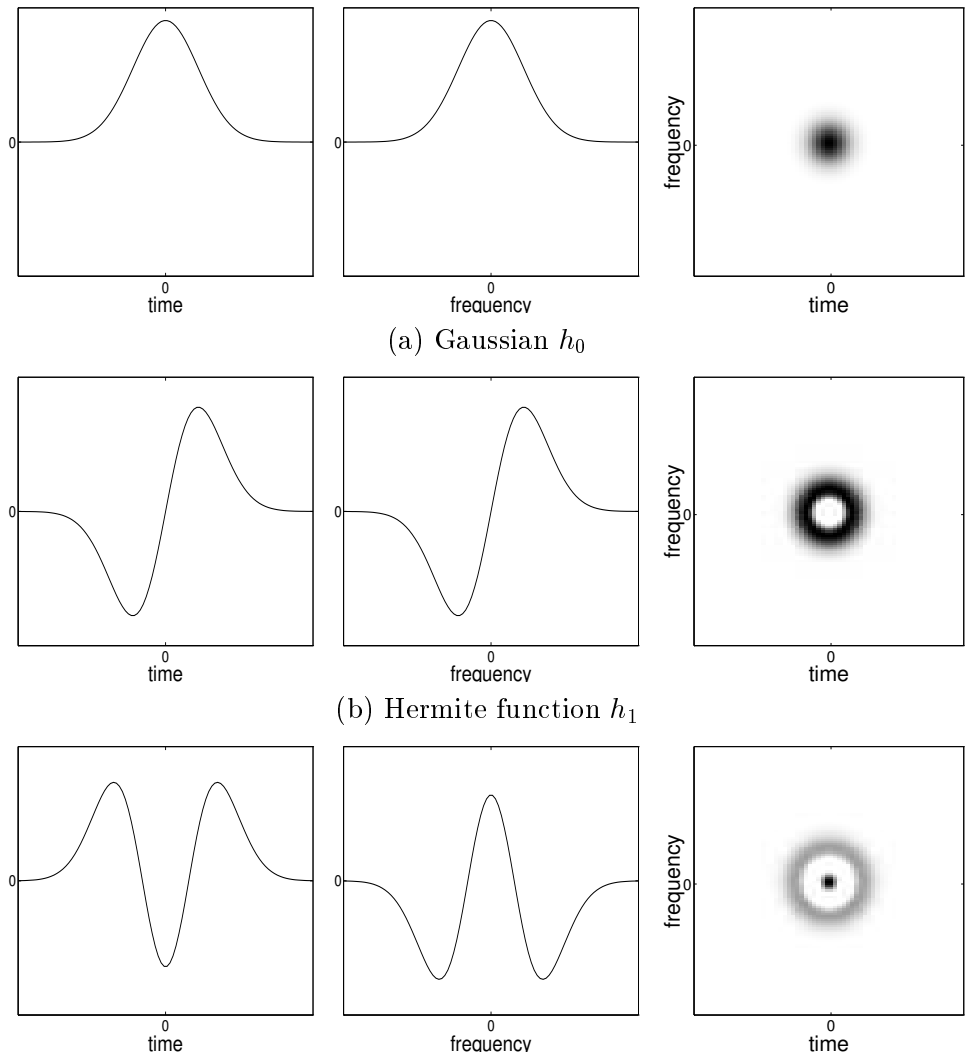

Figure 3: *(a) Circular concentration region (4.2) for the Hermite functions. (b)–(d) Solid: Eigenvalues of the Hermite functions over this circular region for three different areas $A_1 < A_2 < A_3$. Dashed: Eigenvalues of the Morse wavelets (to come in Section 6.1) over the tear-drop region (6.2) shown in Figure 9 for the same three areas. Clearly more Hermite functions than Morse wavelets are concentrated in a region of area $A$.*

of area $A = \pi R^2$ (see Figure 3(a)) and thus treat all time-varying spectral features in the same fashion [20–23]. Hermite functions optimally concentrated in elliptical time-frequency regions are easily obtained by compressing or dilating the $h_k$.

Figure 4 plots the first three Hermite functions, their Fourier transforms, and their Wigner distributions. Note the circular symmetry of the Wigner distributions that matches the circular shape of the concentration region.

The Hermite functions are eigenfunctions of the Fourier transform and also of a time-frequency localization operator over the circular time-frequency region (4.2) [20]. The eigenvalues in this latter case are a function of the area $A = \pi R^2$ of the region [20]

$$\lambda_k(R) := 1 - e^{\frac{-R^2}{2}} \sum_{i=0}^{k} \frac{1}{i!} 2^{-i} R^{2i}. \tag{4.3}$$




(a) Gaussian $h_0$

(b) Hermite function $h_1$

(c) Hermite function $h_2$

Figure 4: *Hermite functions in the (left) time domain, (middle) frequency domain, and (right) time-frequency domain via Wigner distribution. (a) $h_0$ (Gaussian), (b) $h_1$ (since the Fourier transform of $h_1$ is purely imaginary, we plot the imaginary part), (c) $h_2$.*

In Figure 3(b)–(d) we plot the behavior of these eigenvalues with $k$ and area $A$. The closer $\lambda_k$ is to one, the better concentrated $h_k$ is in the circular region (4.2). Thus, as the area increases, more Hermite functions are concentrated within the circular region.

### 4.2 Multiple Window WVS Estimate

Under the stationary signal model (3.1), Thomson's MW spectrum average (3.4) estimates the energy content of the stationary signal $\mathbf{y}$ at frequency $f$ by projecting onto the prolate-windowed sinusoids $v_k(t)\, e^{j2\pi f t}$. By analogy, under the nonstationary signal model (2.4), we estimate the energy content of the non-stationary signal $\mathbf{y}$ at time $t$ and frequency $f$ by projecting onto the sliding Hermite-windowed sinusoids $h_k(\tau - t)\, e^{j2\pi f \tau}$. The estimate can be written as the average of $K$ Hermite-windowed *eigenspectrograms* of the data

$$\widehat{\mathsf{W}}_{\mathbf{y}}(t,f) = \frac{1}{K} \sum_{k=0}^{K-1} \left| \int \widehat{\mathbf{y}}(\tau)\, h_k(\tau - t)\, e^{-j2\pi f \tau}\, d\tau \right|^2 . \tag{4.4}$$

We choose $K$ such that for a given radius $R$ in (4.2) the first $K$ eigenvalues in (4.3) are very close to one. The bias/variance tradeoff is clear: smaller $R$ mean smaller $K$ and thus lower bias in the estimate at the expense of less averaging and hence higher variance.

### 4.3 Cohen's Class Interpretation

The MW WVS estimate (4.4) belongs to Cohen's class of time-frequency distributions. All distributions $\mathbf{C}$ in Cohen's class can be written as [14, 15]

$$\mathbf{C}_{\mathbf{y}}(t,f) = \mathbf{W}_{\mathbf{y}}(t,f) \;**\; \phi(t,f) \tag{4.5}$$

with $\phi$ a kernel function and $**$ 2-D convolution. The kernel generating the spectrogram is precisely the Wigner distribution of the window function.

The Wigner distribution of the $k$-th order Hermite function is the $k$-th order Laguerre function [24–26]

$$\mathbf{W}_{h_k}(t,f) = L_k(t^2+f^2) := e^{-\frac{\pi}{2}(t^2+f^2)} \sum_{m=0}^{k} \frac{k!}{(k-m)!\, m!} \frac{\left[-\pi\,(t^2+f^2)\right]^m}{m!}. \tag{4.6}$$

Therefore, we have a closed form expression for the kernel $\phi$ corresponding to the MW WVS estimate (4.4) as a weighted sum of $K$ Laguerre functions. In this interpretation, the MW WVS estimate (4.4) reads

$$\widehat{\mathsf{W}}_{\mathbf{y}}(t,f) = \mathbf{W}_{\widehat{\mathbf{y}}}(t,f) \;**\; \frac{1}{K} \sum_{k=0}^{K-1} L_k(t^2+f^2). \tag{4.7}$$

In Figure 5, we plot the kernel for a MW spectrum estimate using four Hermite windows.

The fact that thus far the MW WVS estimate is just a distribution from Cohen's class seems to imply that we could and should just smooth the empirical Wigner distribution with a "top hat" function (which is one inside a

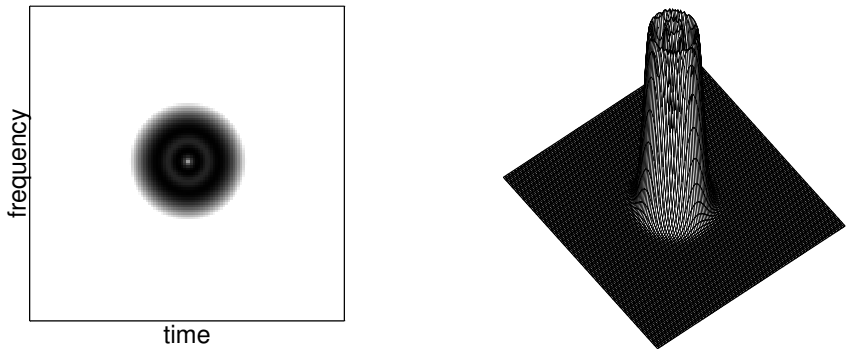



Figure 5: *Cohen's class kernel function $\phi$ corresponding to the MW WVS estimate using the first four Hermite windows (sum of the first four Laguerre functions). Note how closely it approximates a "top hat."*

circle and zero outside) rather than go through the rigmarole of (4.4). As we will see in the next section, however, the eigenspectrograms play a key rôle in detecting/extracting the chirping line components from the signal (recall the model (2.4)).

Note that, unlike most Cohen's class time-frequency distributions [14, 15], the MW WVS estimate is manifestly positive for all signals. And the connection with positive distributions does not stop here. Computation of the jackknife estimate of the variance of the MW WVS estimate [27] leads naturally to the concept of combining eigenspectrograms using a geometric rather than arithmetic mean. This is closely related to Loughlin, Pitton, and Hannaford's generation of positive time-frequency distributions using products of spectrograms [28].

## 5 Extracting Line Components

As in Thomson's method for stationary signals, the averaging inherent in (4.4) will degrade the resolution of chirping line components. Following Thomson's programme, we will first detect and extract all line components in the data before performing (4.4) and then reshape the estimate accordingly.

A straightforward application of Thomson's sinusoid extraction algorithm to a signal from the model (2.4) 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 rapidly chirping components, as we saw in Figure 1(e). In order to detect and extract highly non-stationary chirps, we now present a simple statistical significance test for *linear chirps* of the form $e^{j2\pi(f_0 t + ct^2)}$ [7]. Linear chirps can closely approximate locally all but the most rapidly changing chirp functions. Our approach can thus be interpreted as a first extension of Thomson's technique to time-frequency line components.

### 5.1 Algorithm to Detect and Extract Chirp Components

We make two basic assumptions about the chirp components. We assume 1) that not more than one chirp is present within the elliptical analysis region of the Hermite windows $h_k$ and 2) that the highest chirp rate $c$ attained by any chirp $e^{j2\pi(f_0 t + ct^2)}$ is $\frac{1}{4T}$, with $T$ the effective time support of the Hermite windows. Within the windows' support, we will approximate the line components as piecewise linear chirps

$$\sum_i \mu_i(t)\, e^{j2\pi\gamma_i(t)} \approx \sum_i \mu_i(t)\, e^{j2\pi(f_i(t)t + c_i(t)t^2)} \tag{5.1}$$

with time-varying offset frequency $f_i(t)$ and chirp rate $c_i(t)$.

The chirp detection and extraction algorithm runs as follows:

1. Project the data $\mathbf{x}$ onto linear chirps of the form $e^{j2\pi(ft+ct^2)}$ for a fine grid of offset frequencies $f$ and chirp rates $c$. This is equivalent to Thomson's $F$-test (3.8) applied at each frequency and each chirp rate. A two-dimensional test statistic $F(f, c)$ results.

2. Repeat the above steps at each time point to obtain the 3-D test statistic $F(t, f, c)$.

If the chirp rates of the line components in (2.4) are too high, we must use shorter Hermite windows to deal with (5.1). However, short windows will not detect line components with low chirp rates, because there may not be enough oscillations within the windows for the $F$ test to be reliable. Therefore, for signals containing line components of both high and low chirp rates, it may be necessary to run the above algorithm for different sized windows and combine the results into one test statistic.

### 5.2 Suppressing Spurious Peaks

Due to the repeated application of the test (3.8), the number of spurious peaks in $F$ increases far beyond that seen with stationary signals. (For stationary signals, Thomson applies the test at only one chirp rate $c = 0$, whereas we apply it for each time and chirp rate.) Roughly, if the $F$ test is performed at $M$ chirp rates for each frequency $f$, then $M$ times more spurious peaks will appear compared to when the test is performed at only one chirp rate ($c = 0$). These peaks must be suppressed to create a readable time-frequency image.

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

1. Slice $F(t, f, c)$ along the chirp-rate dimension at several $\{c_j\}$.

2. For each $c_j$, apply a nonlinear order-statistic filter to $F(t, f, c_j)$ to remove peaks that have not coalesced into a region larger than the Heisenberg

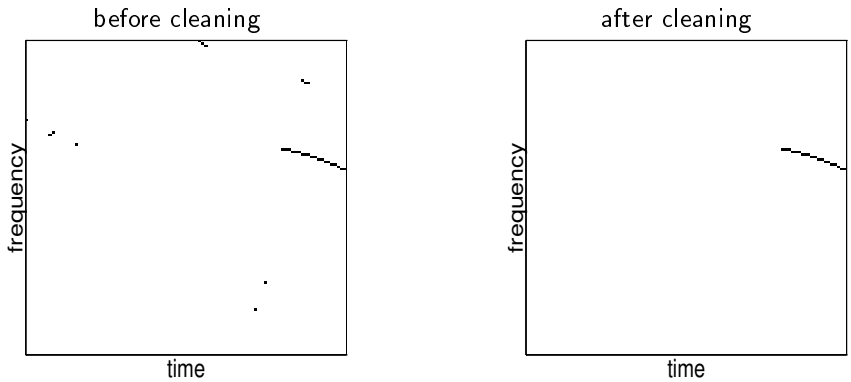



Figure 6: *Test statistic $F(t, f, c_0)$ for a fixed $c_0$ before and after nonlinear cleaning.*

uncertainty principle mandates. (Intuition: spurious peaks are isolated in $F(t, f, c_j)$, while true peaks lie along curves in $F(t, f, c_j)$.) The nonlinear filter essentially counts the number of peaks that lie on a line within a region and compares the count to a threshold.

3. Combine the results from each $c_j$ to obtain the cleaned $F$ test statistic.

While the linear chirp detection/cleaning/extraction algorithm is computationally expensive, it is readily parallelizable.

In Figure 6, we demonstrate the performance of this algorithm on the signal $\mathbf{x}(t) = e^{j2\pi(\frac{a}{b}\sin(bt)+f_0 t)} + n(t)$, with $a$, $b$, and $f_0$ constants and $n(t)$ a stationary Gaussian white noise. The signal-to-noise ratio was set to 0.4 dB. The figure plots $F(t, f, c_0)$ for a fixed chirp rate $c_0$ that corresponds to spurious peaks for early $t$ and intersects the true line component for late $t$. The nonlinear cleaning algorithm exploits the fact that the peaks corresponding to true line components form curves, whereas the spurious peaks lie isolated.

### 5.3 Impulses and Closely Spaced Chirps

In our signal model (2.4) and the above line detection/extraction algorithm, we have not explicitly addressed impulses, which are vertical time-frequency line components of the form $\delta(t-t_0)$. Since the Fourier transform of an impulse is a complex sinusoid, we can detect and extract impulses from the data by applying the above algorithm in the Fourier domain with $c = 0$.

Multiple, crossing chirps will confuse any algorithm that seeks to extract them one at a time. The above approach can be extended à la Thomson to seek and destroy more than one chirp at each time-frequency point. However, as in the stationary case, the detection power of such a test will suffer.

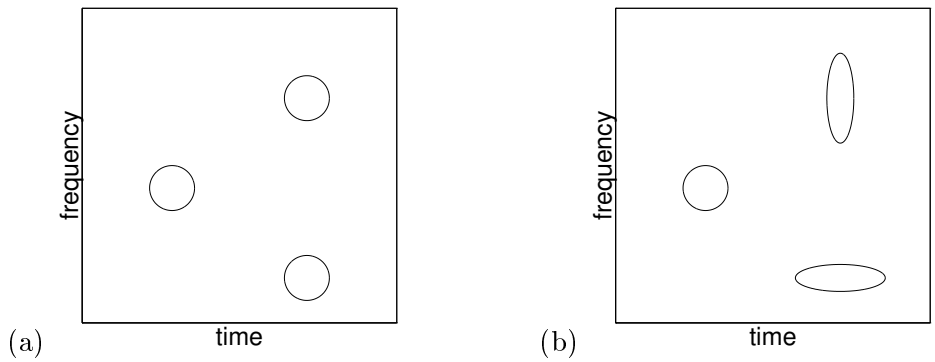



Figure 7: *(a) Cohen's class kernels smooth by translating in time-frequency. (b) Affine class kernels smooth by scaling and translating in time-frequency.*

# 6 Multiple Window Time-Scale Analysis

For random processes containing scaling phenomena (high frequency components of short duration and low frequency components of long duration), standard time-frequency techniques are not appropriate. These types of processes are better matched by the *time-scale representations* from the *affine class* [15, 29]. The smoothing kernels in the affine class change with frequency to accommodate component scaling. The smoothing regions in different parts in the time-frequency plane for Cohen's class and the affine class are shown in Figure 7. To reach higher frequencies, Cohen's class kernels translate, whereas affine class kernels scale.

To estimate the time-varying frequency spectrum of scaling processes, we will modify our Thomson-inspired estimation procedure to incorporate the wavelet transform (replaces the spectrogram) and the Morse wavelets (replace the Hermite windows) that are optimally concentrated in tear-drop shaped regions in time-frequency matched to the scaling behavior in Figure 7(b) [5–8].

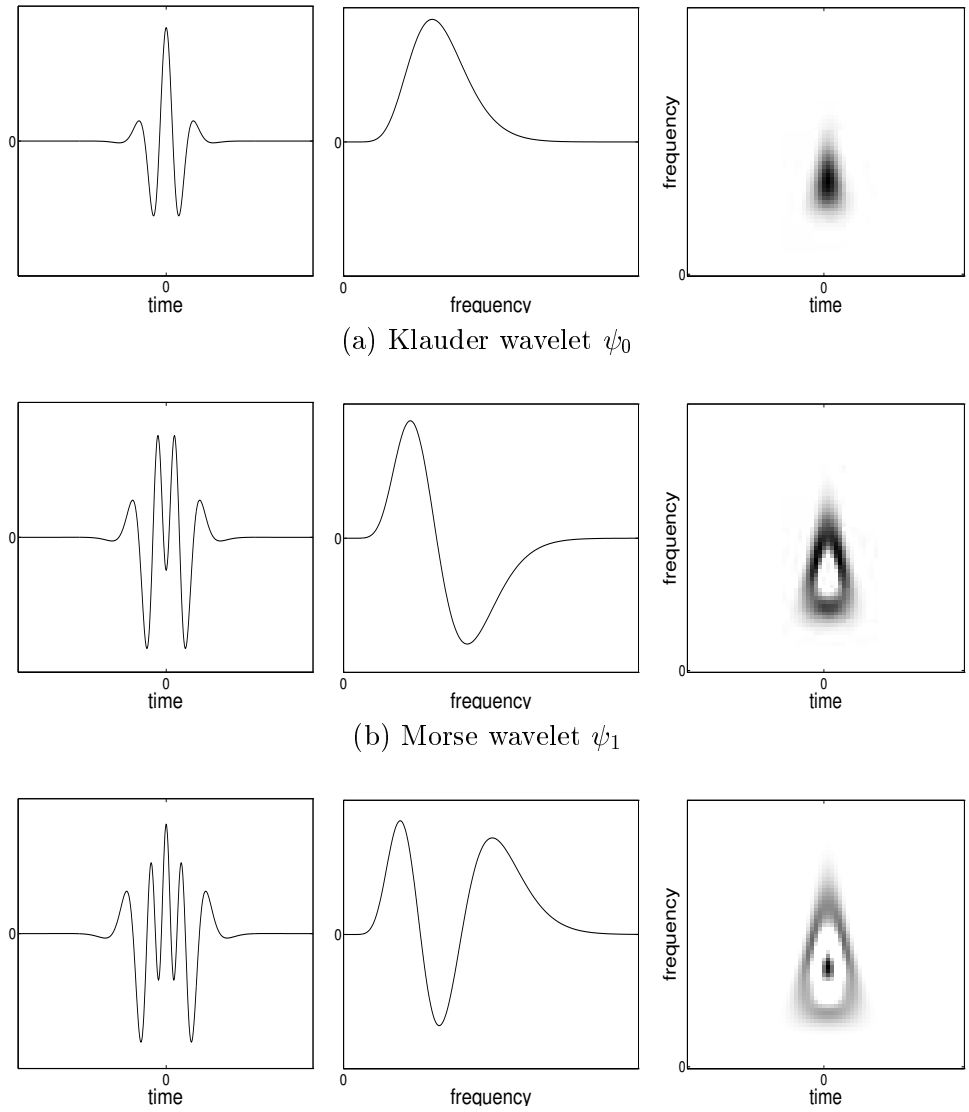
## 6.1 Morse Wavelets

The *Morse wavelets* [30–33] play a rôle in time-scale analogous to that of the Hermite windows in time-frequency. The $k$-th order Morse wavelet[2] $\psi_k(t)$ is defined in the frequency domain as

$$\Psi_k(f) := f^{\beta/2}\, e^{-f^\gamma/2}\, \frac{d^\beta}{df^\beta}\left[e^{f^\gamma}\, \frac{d^{\beta+k}}{df^{\beta+k}}\left(f^{\beta+k}\, e^{-f^\gamma}\right)\right], \qquad k = 0, 1, 2, \ldots \tag{6.1}$$

with $\beta > 0$ the degree of flatness at $f = 0$ and $\gamma > 0$. The zero-th order Morse wavelet is commonly known as the Klauder wavelet [34], although it goes by

[2]While Morse defined only a special case of these wavelets for $\gamma = 1$ [33], we will refer to the entire class derived in [31, 32] as the Morse wavelets.




(a) Klauder wavelet $\psi_0$

(b) Morse wavelet $\psi_1$

(c) Morse wavelet $\psi_2$

Figure 8: *Morse wavelets in the (left) time domain, (middle) frequency domain, and (right) time-frequency domain via Wigner distribution. (a) $\psi_0$ (Klauder wavelet), (b) $\psi_1$ (since the Fourier transform of $\psi_1$ is purely imaginary, we plot the imaginary part), (c) $\psi_2$.*

other names as well [35, p. 25], [36, 37]. Figure 8 shows the first three Morse wavelets in time, their Fourier transforms, and their Wigner distributions.

The Morse wavelets are the eigenfunctions of a localization operator over a tear-drop shaped region whose exact formula for any $\beta$ and $\gamma$ can be found in [32]. For the special case $\beta = \gamma = 1$, the Morse functions are mutually

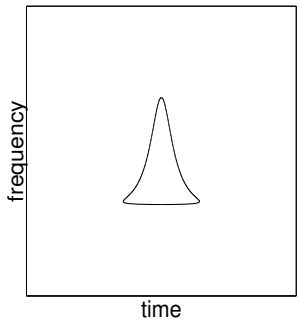



Figure 9: *The tear-drop shaped concentration region (6.2) for the Morse wavelets for* $\beta = \gamma = 1$.

orthogonal and maximally concentrated in the time-frequency region [30, 32]

$$\left\{(t, f): \quad t^2 + \frac{9}{4f^2} + 1 \leq \frac{3C}{|f|}\right\} \tag{6.2}$$

of area $A = 3\pi(C - 2)$ [32]. Figure 9 depicts this region. Just as a circular disk contains all points equidistant from the center point in the Euclidean distance, this region contains all points equidistant from the center point in the (scale-invariant) *Lobachevsky distance* [31].

For $\beta = \gamma = 1$, the eigenvalues of the bandpass localization operator corresponding to the Morse wavelets are given by

$$\lambda_k(C) := \frac{2(k+1)}{(C+1)(k+1)} \left(\frac{C-1}{C+1}\right)^{k+1}. \tag{6.3}$$

(No closed form expression exists for the eigenvalues for any other choices of $\beta$ and $\gamma$.) As in the Hermite case, these eigenvalues indicate the degree of concentration of the corresponding Morse wavelet in the tear-drop region (6.2).

A simple comparison of the concentration properties of the Morse wavelets and the Hermite functions is easily made. In Figure 3(b)–(d), we plot the eigenvalues of the Morse wavelets and Hermite functions for three different areas of their concentration regions. Clearly, for a given area, the Hermite functions have more eigenvalues close to one compared to the Morse wavelets. Therefore, the concentration properties of the Hermite functions on the circular region (4.2) are much better than the concentration properties of the Morse wavelets on the tear-drop region (6.2). Roughly speaking, this means that the bias introduced by averaging over $K$ Morse wavelets should be larger than that due to averaging over $K$ Hermite functions.

### 6.2 Multiple Window WVS Estimate

We form our time-scale MW WVS estimate as the weighted average of the squares of $K$ *eigenscalograms* (squares of wavelet transforms) using Morse wavelets

$$\widehat{\mathsf{W}}_{\mathbf{y}}(t,f) \;=\; \frac{1}{K}\sum_{k=0}^{K-1}\left|\left(\frac{f}{f_0}\right)^{1/2}\int \widehat{\mathbf{y}}(\tau)\,\psi_k\left(\frac{f}{f_0}(\tau-t)\right)d\tau\right|^2. \tag{6.4}$$

Here $f_0$ is a reference frequency (the peak frequency of $\Psi_0$, for example). Again we see a clear bias/variance tradeoff: larger $K$ implies more averaging (smaller variance) but a larger concentration region (larger bias).

A chirp line detection algorithm can be performed similarly to the time-frequency case.

### 6.3 Affine Class Interpretation

The time-scale MW WVS estimate (6.4) belongs the *affine class* of time-scale covariant distributions [15, 29]. Each distribution $\Omega$ in this class can be interpreted as an affine-smoothed version of the Wigner distribution

$$\Omega_{\mathbf{y}}(t,f) = \iint \mathbf{W}_{\mathbf{y}}(\tau,\nu)\,\Pi\left(f(\tau-t),\frac{\nu}{f}\right)d\tau\,d\nu \tag{6.5}$$

with kernel $\Pi$ centered at time zero and frequency $f_0 = 1$. The affine class can also be defined in terms of the unitary Bertrand distribution [38, 39]. Interestingly, the Klauder wavelet $\psi_0$ has a positive Bertrand distribution, just as the Gaussian $h_0$ has a positive Wigner distribution [40].

As in the time-frequency, Cohen's class case, the MW time-scale spectrum estimator can be interpreted as a member of the affine class with a kernel that is a weighted sum of Wigner (or Bertrand) distributions of Morse wavelets (see Figure 10). Unfortunately, closed form formulas for the Wigner and Bertrand distributions of any of the Morse wavelets have not been found. Nevertheless, like the time-frequency MW WVS estimate, the time-scale estimate is manifestly positive.

## 7 Examples

For a first example, refer to Figure 1, where we illustrate the performance of the time-frequency 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 time-domain signal and its ideal representation in time-frequency are shown in Figure 1(a) and (b). It is not possible to identify the components of the test signal from the empirical

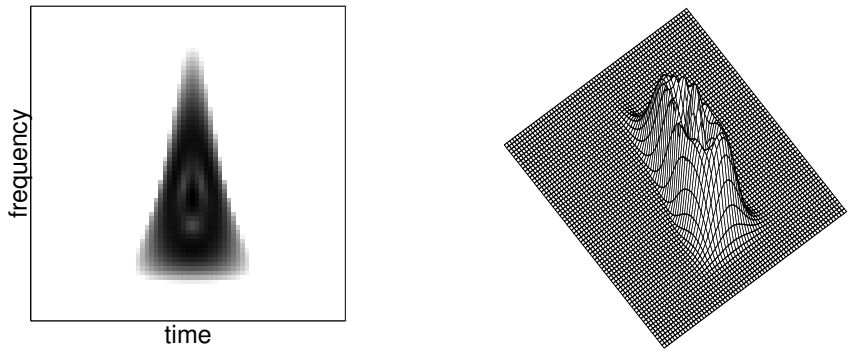



Figure 10: *Affine class kernel function corresponding to the time-scale MW WVS estimate using four Morse wavelets (sum of their Wigner distributions).*

Wigner distribution due to its high variance. The spectrogram smooths the Wigner distribution to reduce the variance, but smears the line component 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. In contrast, the MW estimate of Figure 1(f) has both high resolution and low variance. The empirical variance of the MW WVS estimate is approximately $\frac{1}{4}$ that of the spectrogram, which agrees with the fact that four windows were employed in its computation.

In Figure 11, we demonstrate the ability of the linear chirp detection/extraction algorithm to detect four hyperbolic chirps simultaneously. The data is a digitized 2.5 msec echo-location pulse emitted by the Large Brown Bat, *Eptesicus Fuscus*. There are 400 samples, and the frequency range spanned is approximately $[0, 70]$ kHz. Comparing the time-frequency MW method against the Wigner distribution and spectrogram, we see that the detection algorithm successfully pulls out even the weakest high frequency line component. The method even reveals aliasing in this component due to under-sampling in data acquisition (note the "wraparound" in frequency). Inspection of the early part of each chirp in the MW estimate reveals a threshold effect before which the line extraction algorithm locks on to each component.

In Figure 12, we illustrate the performance of the time-scale MW method using a 256-point test signal containing two Hölder singularities [41, 42] in additive white Gaussian noise $n(t)$

$$\mathbf{x}(t) = |t - 64|^{-0.1} + |t - 180|^{-0.1} + n(t). \tag{7.1}$$

Unlike the scalogram in Figure 12(c), the MW estimate of Figure 12(d) clearly captures the cone-like time-frequency structure of the singularities even in the presence of significant noise.

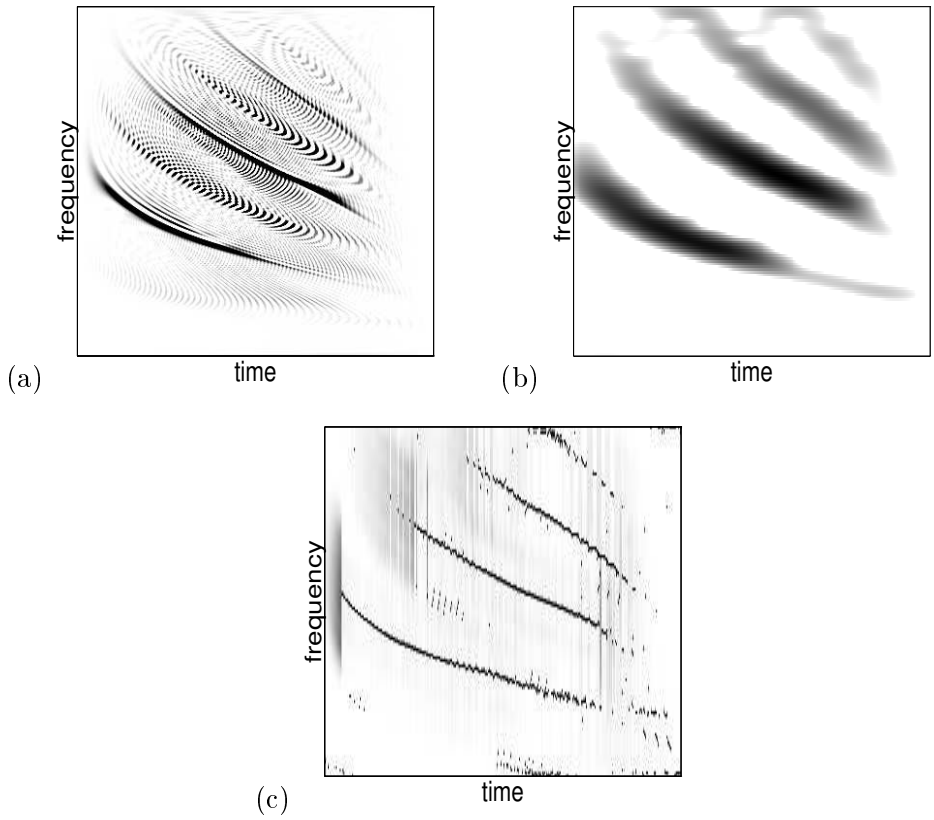



Figure 11: *Three WVS estimates of the echo-location pulse emitted by the Large Brown Bat: (a) empirical Wigner distribution, (b) spectrogram with Gaussian window, and (c) time-frequency MW estimate.*

## 8 Related Work

Our primary contributions to MW time-varying spectral analysis have been identifying the Hermite windows for the time-frequency estimate, introducing the time-scale estimate and its Morse wavelets, and extending the line component $F$ test to the linear chirps [5–8]. Since our original papers were published, a number of interesting extensions and improvements have been made to the technique. We will review these here, as well as point to a large body of related work.

*Different sets of orthogonal window functions:* Xu, Haykin, and Racine compare and contrast prolate and Hermite windows in [43]. Pitton has developed new sets of concentrated time-frequency windows functions that balance the advantages of prolate and Hermite windows [44–46]. Lilly and Park have also considered multi-wavelet time-scale spectrum estimation using a specially designed set of wavelets [47].

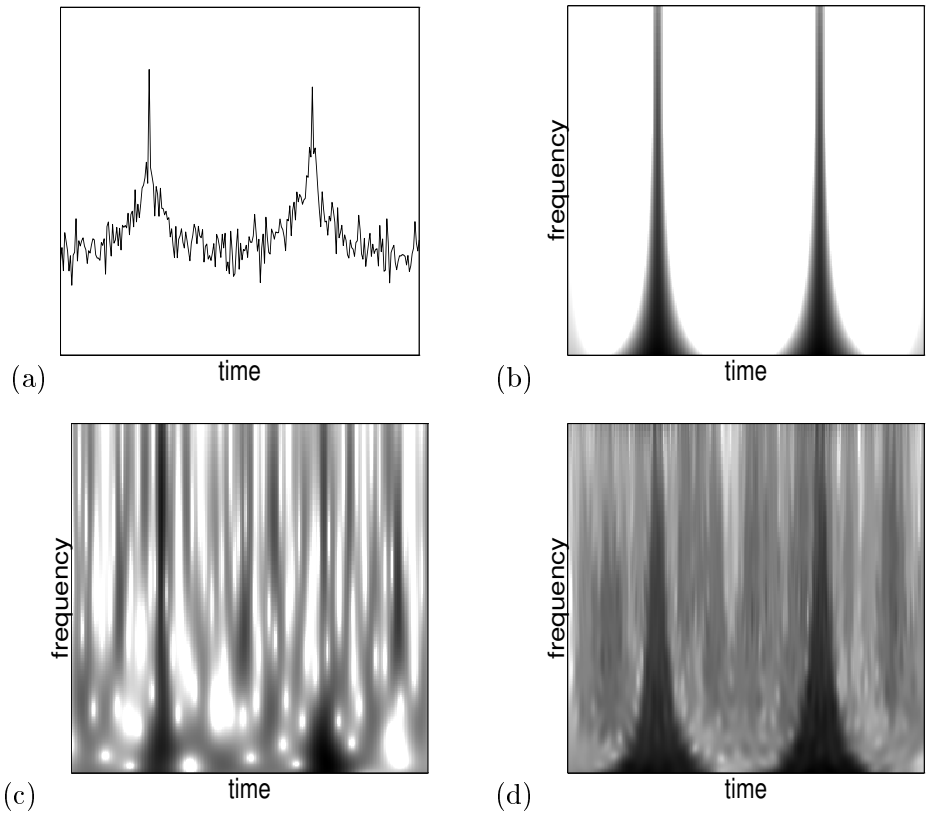



Figure 12: *(a) Test signal composed of two singularities in additive white Gaussian noise. (b) Scalogram of noise-free signal with Klauder wavelet. (c) Scalogram of noisy signal with Klauder wavelet. (d) MW time-scale estimate of noisy signal.*

*Adaptive weighting algorithms:* We have oversimplified our explanation of Thomson's method. Thomson does not weight each eigenspectrum by $K^{-1}$ as we have; rather, he adaptively changes the weights to optimize the bias/variance tradeoff of the estimator. The improvement can be dramatic [16]. In the time-frequency setting, Çakrak and Loughlin adaptively weight Hermite eigenspectra using a least squares procedure [48, 49], while Pitton exploits knowledge of the windows' leakage characteristics [45]. Other stationary adaptation schemes such as that of Hansson [50] and Walden et al [51] could also prove useful in the time-frequency setting.

*Extended chirp extraction algorithms:* Many alternatives exist to our simple linear chirp extraction algorithm. Çakrak and Loughlin employ a multiple window estimate to estimate the instantaneous frequencies of polynomial-phase line components [52]. Pitton has furthermore extended the $F$ test in [46]. Further afield, we could extract chirps using the polynomial phase transform [53], the reassignment method [54–57], the ridge and snakes method [58], or the squeezing method [59].

*Alternative frameworks:* Multiple window estimates succeed when the time-frequency spectrum can be approximated as "locally stationary" within time-frequency regions larger than the concentration region of the orthogonal windows. As such, they can be viewed as special cases of the more general estimation frameworks of Sayeed and Jones [10] and Kozek et al [11, 12, 60, 61]. Rather than assuming a parametric Cohen's class kernel that is the sum of several Wigner distributions of Hermite functions (recall (4.5)–(4.7)), Sayeed and Jones [10] design an optimal kernel $\phi$ that minimizes the mean-square error between the true WVS and the estimate. Other estimation procedures for locally stationary time-frequency spectra that fit within these general frameworks include those of Mallat et al [13] and von Sachs et al [62, 63].

## 9 Conclusions

In this chapter, we have overviewed two multiple-window time-frequency and time-scale spectrum estimators that extend Thomson's seminal work [16] on multiple-window spectrum estimation. The hallmarks of our approach are:

1. Averaging over sets of orthogonal, optimally concentrated windows, the Hermite functions for time-frequency analysis and the Morse wavelets for time-scale analysis. A low bias/low variance estimate results.
2. Detecting and extracting non-stationary line components by approximating them as piece-wise linear chirps. This pre-processing preserves the resolution of the line components.

As we saw in Section 8, much progress has been made recently in extending this framework. But many interesting open issues remain in the theory, implementation, and application of these time-varying spectrum estimators.

## Acknowledgements

Thanks to Patrick Flandrin, Paulo Gonçalvès, Akbar Sayeed, and David Thomson for many stimulating discussions and to Al Feng for the bat data and permission to use it in this chapter. Kudos to Douglas Jones for discussions involving the application of the JAM in this context. This work was supported by the Isaac Newton Institute of Cambridge University (Rosenbaum Fellowship), NSF grants MIP–9457438 and CCR–9973188, DARPA/AFOSR grant F49620–97–1–0513, ONR grants N00014–99–1–0813 and N00014–99–1–0813, the State of Texas Advanced Technology Program, and the Texas Instruments Leadership University Program.

Contact information: c/o RB, Department of Electrical and Computer Engineering, Rice University, 6100 Main Street, Houston, TX 77005, USA. Email: {mebay,richb}@ece.rice.edu, Internet: www.dsp.rice.edu

## References

[1] S. M. Kay, *Modern Spectral Estimation: Theory and Application.* Englewood Cliffs, NJ: 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, I. D. Walker, and R. G. Baraniuk, "Myoelectric teleoperation of a complex robotic hand," *IEEE Trans. Robotics, Automation*, vol. 12, pp. 775–788, Aug. 1996.

[5] M. Bayram and R. G. Baraniuk, "Multiple window time-frequency analysis," in *Proc. IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, (Paris, France), pp. 511–514, June 1996.

[6] M. Bayram and R. G. Baraniuk, "Multiple window time-varying spectrum estimation," in *Conf. Info. Sci. and Sys. (CISS)*, vol. 30, Mar. 1996.

[7] M. Bayram and R. G. Baraniuk, "Multiple window time-frequency and time-scale analysis," in *Proc. SPIE Int. Soc. Opt. Eng.*, 1996.

[8] M. Bayram, "Multiple window time-frequency analysis," Master's thesis, Dep. Elec. Comput. Eng., Rice University, May 1996.

[9] 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.

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

[11] W. Kozek, "Optimally Karhunen-Loève-like STFT expansion of nonstatonary processes," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '93*, vol. IV, (Minneapolis), pp. 428–431, 1993.

[12] W. Kozek and K. Riedel, "Quadratic time-varying spectral estimation for underspread processes," in *Proc. IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, pp. 460–463, 1994.

[13] S. Mallat, G. C. Papanicolaou, and Z. Zhang, "Adaptive covariance estimation of locally stationary processes," *Ann. Statist.*, vol. 26, pp. 1–47, 1998.

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

[15] P. Flandrin, *Time-Frequency and Time-Scale Analysis.* Academic Press, 1999.

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

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

[18] S. M. Kay and G. F. Boudreaux-Bartels, "On the optimality of the Wigner distribution for detection," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '85*, pp. 1017–1020, 1985.

[19] 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.

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

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

[22] 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.

[23] J. Ramanathan and P. Topiwala, *in Wavelets and Their Applications*, pp. 313–325. Kluwer Academic Pulishers, 1994.

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

[25] H. J. Groenewold, "On the principles of elementary quantum mechanics," *Physica*, vol. 21, pp. 405–460, 1946.

[26] N. G. de Bruijn, "Uncertainty principles in Fourier analysis," in *Inequalities*, Academic Press, 1967.

[27] D. J. Thomson, "Jackknifing multiple-window spectra," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '94*, vol. 1, pp. 73–76, 1994.

[28] P. Loughlin, J. Pitton, and B. Hannaford, "Approximating time-frequency density functions via optimal combinations of spectrograms," *IEEE Signal Processing Letters*, vol. 1, pp. 199–202, Dec. 1994.

[29] 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.

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

[31] I. Daubechies, J. R. Klauder, and T. Paul, "Wiener measures for path integrals with affine kinematic variables," *J. Math. Phys.*, vol. 28, pp. 85–102, Jan. 1987.

[32] 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.

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

[34] J. R. Klauder, "Path integrals for affine variables," in *Functional Integration Theory and Applications* (J. P. Antoine and E. Tirapagui, eds.), Plenum, New York, 1980.

[35] Y. Meyer, *Wavelets: Algorithms and Applications.* SIAM, 1993.

[36] H. O. Rasmussen, "The wavelet Gibbs phenomenon," in *Wavelets, Fractals, and Fourier Transforms* (M. Farge, J. C. R. Hunt, and J. C. Vassilicos, eds.), pp. 123–142, Oxford, UK: Clarendon Press, 1993.

[37] M. Holschneider, "On the wavelet transformation of fractal objects," *J. Stat. Phys.*, vol. 50, no. 5/6, pp. 963–993, 1988.

[38] J. Bertrand and P. Bertrand, "A class of affine Wigner functions with extended covariance properties," *J. Math. Phys.*, vol. 33, pp. 2515–2527, July 1992.

[39] R. G. Baraniuk, "Beyond time-frequency analysis: Energy densities in one and many dimensions," *IEEE Trans. Signal Processing*, vol. 46, pp. 2305–2314, Sept. 1998.

[40] P. Flandrin, "On separability, positivity and minimum uncertainty in time-frequency energy distributions," *J. Math. Phys.*, 1998.

[41] P. Gonçalvès and P. Flandrin, "Sur la localisation et la géométrie des distributions affines," in *14ème Colloque GRETSI*, (Juan Les Pins, France), pp. 355–367, 1993.

[42] P. Gonçalvès and R. G. Baraniuk, "Pseudo affine Wigner distributions: Definition and kernel formulation," *IEEE Trans. Signal Processing*, vol. 46, pp. 1505–1516, June 1998.

[43] Y. Xu, S. Haykin, and R. J. Racine, "Multiple window time-frequency analysis of EEG using Slepian sequences and Hermite functions," Nov. 1997. Preprint.

[44] J. W. Pitton, "Nonstationary spectrum estimation and time-frequency concentration," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '98*, 1998.

[45] J. W. Pitton, "Time-frequency spectrum estimation: An adaptive multitaper method," in *Proc. IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, (Pittsburgh), pp. 665–668, 1998.

[46] J. W. Pitton, "Adaptive multitaper time-frequency spectrum estimation," in *Proc. SPIE Int. Soc. Opt. Eng.*, (Denver), 1999.

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

[48] F. Çakrak and P. Loughlin, "Multiple window nonlinear time-frequency analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '98*, pp. 2409–2412, 1998.

[49] F. Çakrak and P. Loughlin, "Multiple window time-varying spectral analysis," *IEEE Trans. Signal Processing*, 1999. Submitted.

[50] M. Hansson, "Optimized weighted average of peak matched multiple window spectrum estimates," *IEEE Trans. Signal Processing*, vol. 47, pp. 1141–1146, Apr. 1999.

[51] A. T. Walden, D. B. Percival, and E. J. McCoy, "Spectrum estimation by wavelet thresholding of multitaper estimators," *IEEE Trans. Signal Processing*, vol. 46, pp. 3153–3165, Dec. 1998.

[52] F. Çakrak and P. Loughlin, "Instantaneous frequency estimation of polynomial phase signals," in *Proc. IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, pp. 549–552, 1998.

[53] S. Peleg and B. Friedlander, "The discrete polynomial-phase transform," *IEEE Trans. Signal Processing*, vol. 43, pp. 1901–1914, Aug. 1995.

[54] K. Kodera, C. D. Villedary, and R. Gendrin, "A new method for the numerical analysis of non-stationary signals," *Phys. Earth and Plan. Int.*, vol. 12, pp. 142–150, 1976.

[55] K. Kodera, R. Gendrin, and C. D. Villedary, "Analysis of time-varying signals with small BT values," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, no. 1, pp. 64–76, 1978.

[56] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. Signal Processing*, vol. 43, pp. 1068–1089, May 1995.

[57] F. Auger, E. Chassande-Mottin, I. Daubechies, and P. Flandrin, "Differential reassignment," *IEEE Signal Processing Letters*, vol. 4, pp. 293–294, Oct. 1997.

[58] R. A. Carmona, W. L. Hwang, and B. Torrèsani, "Characterization of signals by the ridges of their wavelet transforms." Mar. 1995.

[59] I. Daubechies and S. Maes, "A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models," in *Wavelets in Medicine and Biology* (A. Aldroubi and M. Unser, eds.), pp. 527–546, CRC Press, 1996.

[60] W. Kozek, F. Hlawatsch, H. Kirchauer, and U. Trautwein, "Correlative time-frequency analysis and classification of nonstationary random processes," in *Proc. IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, pp. 417–420, 1994.

[61] G. Matz, F. Hlawatsch, and W. Kozek, "Generalized evolutionary spectrum analysis and the Weyl spectrum of nonstationary random processes," *IEEE Trans. Signal Processing*, vol. 45, pp. 1520–1533, June 1997.

[62] R. von Sachs and K. Schneider, "Wavelet smoothing of evolutionary spectra by nonlinear thresholding," *Appl. Comp. Harm. Anal.*, vol. 3, pp. 268–282, 1996.

[63] M. H. Neumann and R. von Sachs, "Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra," *Ann. Statist.*, vol. 25, pp. 38–76, 1997.