Signal Processing for Multimedia
J.S. Byrnes (Ed.)
IOS Press, 1999

Time-Frequency and Time-Scale Analysis

Bruno Torrésani

Centre de Physique Théorique
CNRS-Luminy, case 907
13288 Marseille Cedex 09
France

Laboratoire d'Analyse
Topologie et Probabilités
CMI
39 rue F. Joliot-Curie
13453 Marseille Cedex 13
France

ABSTRACT. Current multimedia technologies call for efficient ways of representing signals. We review several efficient methods for signal representation, emphasizing potential applications in signal compression and denoising. We pay special attention to the representations which are adapted to "non-stationary" features of signals, in particular the classes of bilinear representations, and their approximations using time-frequency atoms (mainly wavelet transforms and Gabor transforms).

1. Introduction

In various instances in signal processing, an important part of the processing is achieved by an *efficient representation* of the considered signal. This is the case, for example, in signal compression, where coding and bit allocation often come after a transform which expresses the signal in an adapted basis, with respect to which a large number of coefficients may be discarded. This is true for signal denoising as well, for an efficient representation "concentrates" the useful signal within a few significant coefficients, while noise remains distributed over all coefficients. Therefore, an efficient representation, followed by simple operations such as thresholding, generally yield good denoising algorithms.

The goal of this contribution is to describe a number of simple efficient representations that are generated by using "time-frequency" decompositions, and to show how these may be used for practical purpose. Special attention is paid to problems of spectral estimation, for non stationary time series. We shall mainly limit our discussion to simple decompositions such as wavelet or Gabor decompositions, in order to emphasize the difficulties of such approaches, but we will also devote a few words to more sophisticated tools.

2. "Non-Stationary Tools"

Let us start by describing a few tools which we will use in the following. The most usual tool is the Fourier transform. It is well known that Fourier analysis is well adapted to "stationary situations", i.e. signals which possess some translation invariance properties (we use the following convention for Fourier transformation: $\hat{f}(\omega) = \int f(t)e^{-i\omega t}dt$.) When translation invariance assumptions are relaxed the Fourier transform is no longer a well-adapted tool, and alternatives are needed. Among them, time-frequency and time-scale methods have become quite popular in recent years, as they provide simple approximations of optimal, Karhunen-Loève-type decompositions.

Briefly, the Karhunen-Loève (KL for short) decomposition is obtained by diagonalizing the covariance of a second order random time series. Let $\{X_t, t \in \mathbb{R}\}$ be a second order zero mean time series, and let $\mathcal{C}$ be its covariance operator, defined by its matrix elements $\langle \mathcal{C}f, g\rangle = \mathbb{E}\left\{\overline{\langle X, f\rangle}\langle X, g\rangle\right\}$. $\mathcal{C}$ is non-negative definite (and self-adjoint.) Assume for the sake of simplicity that $\mathcal{C}$ has discrete spectrum, and denote by $\{\varphi_k, \lambda_k\}$ the eigenfunctions and eigenvalues of $\mathcal{C}$. This yields an expansion of the time series as a (random) linear combination of the functions φ_k:

$$X_t = \sum_k \sqrt{\lambda_k} w_k \varphi_k(t) \tag{1}$$

Such an expansion is "doubly orthogonal" in the sense that

$$\langle \varphi_k, \varphi_\ell \rangle = \delta_{k\ell} \tag{2}$$

$$\mathbb{E}\{w_k \overline{w_\ell}\} = \delta_{k\ell} \tag{3}$$

However, KL-type decompositions are sometimes of poor practical interest. Indeed, diagonalizing the covariance becomes in practice a matrix diagonalization problem, which becomes cumbersome as the matrix size increases. In addition, prior to diagonalization, the covariance matrix has to be estimated, generally from one or (in rare cases) a few realizations of the time series. All together, performing a KL decomposition may become a difficult practical problem, and it makes sense to seek alternative methods, at least in some specific situations where some *a priori* information about the time series is available. One particular case is that of time series which are "not far from stationary", i.e., to which one may want to associate a sort of time dependent spectral representation. Studying such time series leads to the notion of time-frequency representations. Another example is provided by time series which present some sort of scale invariance. This is the realm of time-scale analysis. In what follows, we briefly describe these topics.

2.1. *Bilinear representations*

Signals are often modeled either as deterministic signals, or more generally as (second order) random time series. In what follows we will focus on the random situations (unless otherwise specified), the deterministic case being easily obtained. The simplest model to consider is that of (weakly) stationary time series. However, in many situations signals can hardly be considered stationary, and it is necessary to turn to alternative tools which generalize the usual ones in non stationary situations. Several such tools have been developed in the literature, the most commonly used being probably the KL-based methods. However, there are many situations in which the signals to be analyzed possess some characteristics which may be better understood in terms of joint time-frequency representations. The prototypes of such time-frequency representations are the so-called *ambiguity function* and *Wigner function* (or *Wigner-Ville function*). The ambiguity function was introduced by Woodward in a radar context. The ambiguity function is essentially obtained by taking scalar products of a function with a time-frequency shifted copy of itself. More precisely:

DEFINITION 2.1. 1. *Let $f \in L^2(\mathbb{R})$. Its ambiguity function is defined by*

$$\mathcal{A}_f(\tau, \xi) = \int f(t+\tau/2)\overline{f}(t-\tau/2)e^{-i\xi t}dt\ . \tag{4}$$

2. *Let* $\{X_t, t \in \mathbb{R}\}$ *be a second order time series. Then its ambiguity function is defined by*

$$\mathcal{A}_X(\tau, \xi) = \mathbb{E}\left\{ \int X_{t+\tau/2}\overline{X}_{t-\tau/2} e^{i\xi t} dt \right\} . \tag{5}$$

The ambiguity function was originally introduced in a deterministic context. The deterministic ambiguity function may be seen as a scalar product of f with a translated and modulated copy of f (up to a trivial factor). It is easily seen that if $f \in L^2(\mathbb{R})$, then $\mathcal{A}_f$ is a bounded function (with $||\mathcal{A}_f||_\infty \leq ||f||^2$). In addition, a direct calculation shows that if $f \in L^2(\mathbb{R})$, then $\mathcal{A}_f \in L^2(\mathbb{R}^2)$, and that $||\mathcal{A}_f||_2^2 = 2\pi ||f||^4$. More generally, it may be shown that $\mathcal{A}_f \in L^p(\mathbb{R}^2)$ for all $p \in [1, \infty]$ (bounds for the corresponding $L^p(\mathbb{R}^2)$ norms have been derived by E. Lieb).

The non-deterministic version may be given a similar interpretation. Its properties depend on the properties of the covariance operator $\mathcal{C}$ of the process, defined by its matrix elements: for all $f, g \in \mathcal{D}(\mathbb{R})$,

$$\langle \mathcal{C}f, g \rangle = \mathbb{E}\left\{ \langle X, g \rangle \overline{\langle X, f \rangle} \right\} . \tag{6}$$

For example, if $\mathcal{C}$ extends to a Hilbert-Schmidt operator, which we denote by $\mathcal{C} \in \mathcal{L}^2$, then $\mathcal{A} \in L^2(\mathbb{R}^2)$.

REMARK 2.2. *Ambiguity functions, or* cross-ambiguity functions *of the form*

$$\mathcal{A}_{f,g}(\tau, \xi) = \int f(t + \tau/2)\overline{g}(t - \tau/2) e^{-i\xi t} dt \tag{7}$$

are widely used in the context of radar detection. There, f *is an incident waveform and* g *is the observation, supposed to be an attenuated time-frequency shifted copy of* g*, of the form* $g(t) = Ae^{i\omega_0(t-t_0)} f(t - t_0) +$ *noise. Here,* A, ω_0 *and* t_0 *are constants of practical interest, to be estimated. The maxima of the cross ambiguity function provide estimates for these constants.*

As we shall see, ambiguity functions only provide estimates for the spreading of the analyzed object in the joint time-frequency plane, but not on its localization in that space. Such an analysis is done by the *Wigner-Ville* function, defined by

DEFINITION 2.3. 1. *Let* $f \in L^2(\mathbb{R})$*. Its Wigner-Ville function is defined by*

$$\mathcal{E}_f(b, \omega) = \int f(b + \tau/2)\overline{f}(b - \tau/2) e^{-i\omega\tau} d\tau \tag{8}$$

2. *Let* $\{X_t, t \in \mathbb{R}\}$ *be a second order time series. Then its Wigner-Ville function is defined by*

$$\mathcal{E}_X(b, \omega) = \mathbb{E}\left\{ \int X_{b+\tau/2}\overline{X}_{b-\tau/2} e^{-i\omega\tau} d\tau \right\} \tag{9}$$

REMARK 2.4. *It is readily seen that the Wigner-Ville function and the ambiguity function are related via a symplectic Fourier transform*

$$\mathcal{A}(\tau, \xi) = \frac{1}{2\pi} \int\int \mathcal{E}(b, \omega) e^{-i(\xi b - \omega\tau)} db\, d\omega , \tag{10}$$

$$\mathcal{E}(b, \omega) = \frac{1}{2\pi} \int\int \mathcal{A}(\tau, \xi) e^{i(\xi b - \omega\tau)} d\tau\, d\xi . \tag{11}$$

The same holds true in the deterministic context. Therefore, the Wigner function is square-integrable as soon as the ambiguity function is.

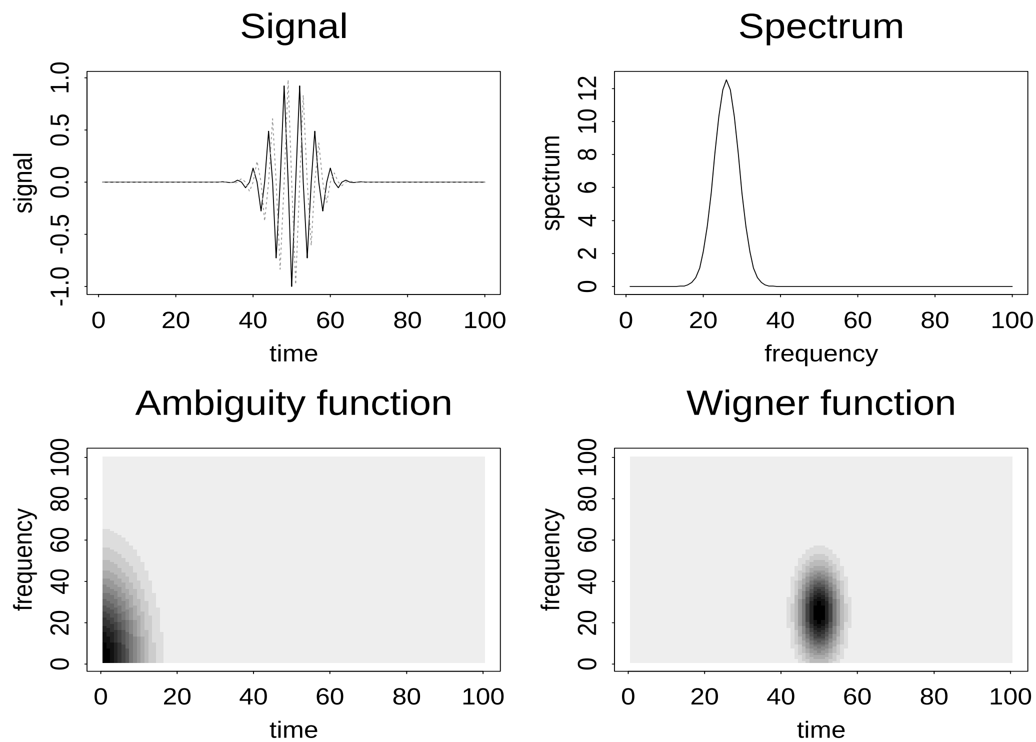

FIGURE 1. Example of a "Time-Frequency Atom" (left top), together with its Fourier transform (right top); its Ambiguity function (left bottom) and its Wigner function (right bottom) are displayed as gray levels images.

It is important to notice the major difference between the ambiguity function and the Wigner function (even though their expressions are quite close). As we have seen, the ambiguity function is a scalar product between two time and frequency shifted copies of the signal: if $f \in L^2(\mathbb{R})$:

$$\mathcal{A}_f(\tau, \xi) = \langle T_{-\tau/2} E_{-\xi/2} f, T_{\tau/2} E_{\xi/2} f \rangle$$

where T and E are translation and modulation operators respectively, defined by $T_b f(t) = f(t-b)$, and $E_\omega f(t) = e^{i\omega t} f(t)$. For $f \in L^2(\mathbb{R})$, one has $|\mathcal{A}_f(\tau, \xi)| \leq \mathcal{A}_f(0,0) = ||f||^2$, and the decay of $\mathcal{A}$ gives indications about the localization properties of the analysed object (process or function) in the time-frequency plane.

On the other hand, the Wigner-Ville function has a more complex structure, i.e.,

$$\mathcal{E}_f(b, \omega) = \frac{1}{2} \langle \Pi T_{-b} E_{-\omega} f, T_{-b} E_{-\omega} f \rangle \ ,$$

where Π is the parity, defined by $\Pi f(t) = f(-t)$, and actually provides estimates for time-frequency localization of signals. An example stressing the difference is given in Figure 1, for the particular case where $f(t)$ is a modulated Gaussian function. As expected, the ambiguity function is localized near the origin in the time-frequency domain, while the Wigner function is concentrated near a specific point in the time-frequency plane, yielding estimates for the time and frequency content of the analyzed function.

Of interest too is the *cross Wigner-Ville function*, defined for all $f, g \in L^2(\mathbb{R})$ by

$$\mathcal{E}_{f,g}(b, \omega) = \int f(b + \tau/2) \overline{g}(b - \tau/2) e^{-i\omega\tau} d\tau \ . \tag{12}$$

2.2. *Properties*

The Wigner function possesses a large number of important properties. We list here a number of simple ones, refering to [9] for a detailed account.

1. *Marginals:* The first two properties we mention deal with the behavior of marginals of the Wigner function. Namely, the Wigner function integrated with respect to the time variable or the frequency variable reproduces the power spectrum and the (square modulus of the) signal. More precisely, we have the following:

   Let $f \in L^2(\mathbb{R})$. Then $\int \mathcal{E}_f(b,\omega)d\omega = 2\pi|f(b)|^2$, and $\int \mathcal{E}_f(b,\omega)db = |\hat{f}(\omega)|^2$.
2. *Orthogonality relations:* Let $f, f', g, g' \in L^2(\mathbb{R})$. Then we know that $\mathcal{E}_{f,g}, \mathcal{E}_{f',g'} \in L^2(\mathbb{R}^2)$, and a simple calculation shows that

   $$\langle \mathcal{E}_{f,g}, \mathcal{E}_{f',g'} \rangle = 2\pi \langle f, f' \rangle \overline{\langle g, g' \rangle}$$

   Such relations are known as orthogonality relations, or as *Moyal's formula*.
3. *Time-frequency localization:* The second set of properties deals with localization. It is well known that the Fourier transform is "optimal" in the case of sine waves, in the sense that the Fourier transform of a sine wave is a delta distribution, which is "optimally localized". Since the Wigner function plays the role of a generalized spectrum, it makes sense to search for signals with "perfect" localization in the time-frequency plane. In the case of Wigner's function, such signals are provided by the class of (generalized) "linear chirps". A correct treatment of such cases (which involves Wigner functions defined as distributions) is out of the scope of the present discussion, and we limit ourselves to a formal discussion. Suppose that $f(t)$ is defined as $f(t) = \exp(i\omega_0 t + \alpha t^2/2))$, for some parameters ω_0, α. Then $\mathcal{E}_f(b,\omega) = \delta(\omega - (\omega_0 + \alpha t))$, i.e., has "perfect localization" on a straight line in the time-frequency plane. Such signals may be viewed as time-frequency rotated copies of sine waves, and include as limiting cases Dirac deltas, which are optimally localized too. Unfortunately, such a property does not generalize to frequency modulations different from linear ones (see Remark 2.5 below.)
4. *Bilinearity:* Inherent to the bilinear nature of Wigner's function is the existence of "cross terms". More precisely, let $f \in L^2(\mathbb{R})$ be of the form $f(t) = f_1(t) + f_2(t)$, with $f_1, f_2 \in L^2(\mathbb{R})$. Then

   $$\mathcal{E}_f(b,\omega) = \mathcal{E}_{f_1}(b,\omega) + \mathcal{E}_{f_2}(b,\omega) + 2\Re\mathcal{E}_{f_1,f_2}(b,\omega)\ ,$$

   where the cross Wigner-Ville function has been defined in (12). The presence of such interference terms (sometimes called "ghost terms") is generally considered a serious difficulty when it comes to interpreting a Wigner representation. One classical method amounts to getting rid of ghosts by appropriate smoothings of the representation (see below). However, smoothing modifies the localization properties of the representation. In a few specific cases, it is possible to analyze and understand completely the geometric properties of ghost terms. But this is limited to very specific situations.

REMARK 2.5. *As we have seen, the Wigner-Ville representation is "optimal" for linear chirps, in terms of time-frequency localization. It is worth mentioning that other classes of bilinear time-frequency representations have been proposed, which are optimal for some specific frequency modulations. More generally, bilinear time-frequency representations may be designed which enforce specific properties (optimal energy localization for given frequency modulations, positivity, unitarity,...) We refer to [6,9] for a detailed account of recent contributions in that area.*

2.3. *Estimation*

The practical problem is often that of estimating the spectral characteristics of a function or a process from one or a few realizations. The simplest estimators are the sample estimators: for example, given N independent realizations $X^{(1)}, \ldots X^{(N)}$ of the time series, consider (throughout this paper, we use the notation "$\tilde{x}$" to denote an estimator for the quantity "x", reserving the notation “$\hat{x}$” –more standard in the statistics literature–to denote Fourier transform)

$$\tilde{\mathcal{E}}_X(b,\omega) = \frac{1}{N}\sum_{n=1}^{N}\int X_{b+\tau/2}\overline{X}_{b-\tau/2}e^{-i\omega\tau}d\tau \ . \tag{13}$$

In addition, real data are most often discrete and of finite length, so that the integral defining $\tilde{\mathcal{E}}_X(b,\omega)$ in (13) has to be replaced with a finite sum. The limits of the estimator as the sample length and the sampling frequency increase are an important issue. For the sake of simplicity, we shall not address those technical issues here. We just notice that such sample estimators generally turn out to have a large variance and poor smoothness. Therefore, one generally turns to smoothed versions (see the discussions in [9] for example). We shall see below that the use of wavelet or Gabor transforms provides examples of such smoothings. A more general class of smoothings of the Wigner-Ville function has been introduced by L. Cohen, and is known as the *Cohen's class*. See [6, 9] for a detailed account.

3. Approximating Bilinear Representations

Let us now address a slightly different point of view, and discuss somewhat simpler objects, namely the so-called *linear time-frequency representations*. As we shall see, such representations may be seen as alternatives to the bilinear representations we just described, but also as approximations. The simplest examples of such linear transforms are the continuous wavelet and Gabor transforms, which we describe now. However, several variants have been proposed, which we shall briefly discuss later.

3.1. *Windowed Fourier Transform and Wavelet transform*

We describe here the simplest two examples of time-frequency linear decompositions. We first focus on the case of continuous transforms, and postpone the description of the discretization problem to a subsequent section. We first describe the deterministic situation.

The simplest localized version of Fourier analysis is provided by the windowed Fourier transform, where the main idea is to localize the signal first by multiplying it by a smooth and localized window, and then perform a Fourier transform. More precisely, the construction goes as follows. Start from a function $g \in L^2(\mathbb{R})$ such that $||g|| \neq 0$, and associate with it the following family of *Gaborlets*

$$g_{(b,\omega)}(t) = e^{i\omega(t-b)}g(t-b) \ . \tag{14}$$

DEFINITION 3.1. *Let $g \in L^2(\mathbb{R})$ be a window. The continuous Gabor transform of a finite-energy signal $f \in L^2(\mathbb{R})$ is defined by the integral transform*

$$G_f(b,\omega) = \langle f, g_{(b,\omega)}\rangle = \int f(t)\,\overline{g(t-b)}e^{-i\omega(t-b)}\,dt \ . \tag{15}$$

Gaborlets yield decomposition formulas for functions in $L^2(\mathbb{R})$, as follows.

THEOREM 3.2. *Let $g \in L^2(\mathbb{R})$ be a non trivial window (i.e., $||g|| \neq 0$). Then every $f \in L^2(\mathbb{R})$ admits the decomposition*

$$f(t) = \frac{1}{2\pi||g||^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_f(b,\omega)g_{(b,\omega)}(t)dbd\omega \ , \tag{16}$$

where equality holds in the weak $L^2(\mathbb{R})$ sense.

In other words, the mapping

$$L^2(\mathbb{R}) \ni f \hookrightarrow \frac{1}{||g||\sqrt{2\pi}}G_f \in L^2(\mathbb{R}^2)$$

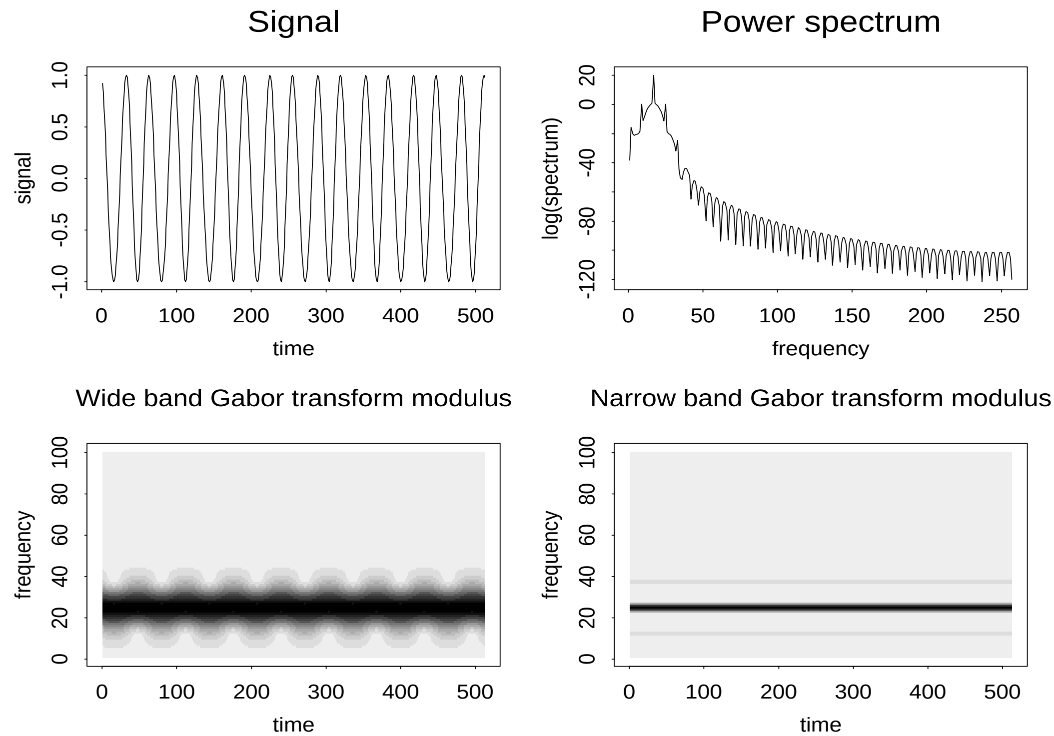

FIGURE 2. Example of a frequency modulated signal, with periodic frequency modulation (left top plot). The power spectrum (right top plot) exhibits a main frequency and a few harmonics and subharmonics. A Gabor transform with a wide band window (left bottom) exhibits the frequency modulation, while a Gabor transform with a narrow band window (right bottom) reproduces the harmonic structure of the signal. The window was a Gaussian function, with two different scales.

is an isometry between $L^2(\mathbb{R})$ and $L^2(\mathbb{R}^2)$.

The Gabor transform of a signal gives indications of its "time-frequency content". Unlike the Wigner transform, it does not have sharp localization properties for specific frequency modulations (this is due to the fact that the Gabor transform is closely related to a smoothing of the Wigner transform). Nevertheless, it may be used to study frequency modulations. For example, consider a function of the form $f(t) = A(t)e^{i\phi(t)}$, and assume that $A \in C^1(\mathbb{R})$, $\phi \in C^2(\mathbb{R})$, and that both $A(t)$ and $\phi'(t)$ are slowly varying. Then, it follows directly from Taylor's formula that $G_f(b,\omega) = A(b)e^{i\phi(b)}\overline{\hat{g}(\phi'(b) - \omega)} + R(b,\omega)$, where $|R(b,\omega)| = O(|A'|, |\phi''|)$. Therefore, if $g(t)$ is a smooth function whose Fourier transform is peaked at the origin of frequencies, and assuming that $R(b,\omega)$ is small enough to be neglected in a first order approximation, $|G_f(b,\omega)|$ is peaked near a curve (the so-called *ridge*) of equation $\omega = \phi'(b)$, which reproduces the frequency modulation of the signal.

An example of such time-frequency localization is given in Figure 2, for the case of a periodically frequency modulated signal. This illustrates the two main features of the Gabor transform. The left bottom image is a gray level representation of the modulus of the Gabor transform, in the case where the window $g(t)$ (here a Gaussian window) is "local enough"; such windows allow us to "see" the changes in the frequencies of the signal, thereby giving a meaning to the notion of "local frequency". To obtain such local quantities, we have to give up frequency resolution, i.e., the localization near the ridge is not as sharp as one would naively expect. This is especially clear on the right bottom image of Figure 2, where a "narrow band window" (again a Gaussian function) has been used. In that case, the window is not enough local, and cannot carefully analyze the frequency changes. However, it is extremely precise in the frequency domain, and reproduces the harmonics and subharmonics which appear in the Fourier spectrum with great precision.

An alternative to the Gabor transform was proposed more recently by Grossmann and Morlet [11]. The main idea was to improve the time resolution of the Gabor transform, by changing the rule for generating the "basis functions". This may be done by replacing the modulation operation used to generate Gaborlets by a scaling operation. Let $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ be a fixed function (in fact, it is sufficient to assume $\psi \in L^1(\mathbb{R})$, but for convenience also assume that $\psi \in L^2(\mathbb{R})$. This extra assumption ensures the boundedness of the wavelet transform). From now on it will be called the *analyzing wavelet*. It is also sometimes called the *mother wavelet* of the analysis. The corresponding family of wavelets is the family $\{\psi_{(b,a)};\ b \in \mathbb{R}, a \in \mathbb{R}_+^*\}$ of shifted and scaled copies of ψ defined as follows: If $b \in \mathbb{R}$ and $a \in \mathbb{R}_+^*$ we set:

$$\psi_{(b,a)}(t) = \frac{1}{a}\psi\left(\frac{t-b}{a}\right), \quad t \in \mathbb{R}\,. \tag{17}$$

The wavelet $\psi_{(b,a)}$ can be viewed as a copy of the original wavelet ψ rescaled by a and centered around the "time" b. Given an analyzing wavelet ψ, the associated continuous wavelet transform is defined as follows

DEFINITION 3.3. *Let $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ be an analyzing wavelet. The continuous wavelet transform (CWT for short) of a finite-energy signal $f(t)$ is defined by the integral:*

$$T_f(b,a) = \langle f, \psi_{(b,a)}\rangle = \frac{1}{a}\int f(t)\overline{\psi}\left(\frac{t-b}{a}\right)dt\,. \tag{18}$$

Like Gaborlets, wavelets may form complete sets of functions in $L^2(\mathbb{R})$, and we have in particular

THEOREM 3.4. *Let $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, be such that the number c_ψ defined by:*

$$c_\psi = \int_0^\infty |\hat{\psi}(a\xi)|^2\frac{da}{a} \tag{19}$$

is finite, nonzero and independent of $\xi \in \mathbb{R}$. Then every $f \in L^2(\mathbb{R})$ admits the decomposition

$$f(t) = \frac{1}{c_\psi}\int_{-\infty}^\infty\int_0^\infty T_f(b,a)\psi_{(b,a)}(t)\frac{da}{a}db\,, \tag{20}$$

where the convergence holds in the strong $L^2(\mathbb{R})$ sense.

In particular, we also have "energy conservation": if $f \in L^2(\mathbb{R})$, then $T_f \in L^2(\mathbb{R} \times \mathbb{R}_+^*, db\frac{da}{a})$, and $||T_f||^2 = c_\psi||f||^2$. Notice that in (19) the constant c_ψ can only depend on the sign of $\xi \in \mathbb{R}$. Therefore, independence wrt ξ is a simple symmetry assumption. The fact that $0 < c_\psi < \infty$ implies that $\hat{\psi}(0) = 0$, so that the wavelet $\psi(t)$ has to oscillate enough to be of zero mean.

Wavelet analysis may be used as a time-frequency analysis method, though in a slightly different way. To see that, let us consider again the same example as before: $f(t) = A(t)e^{i\phi(t)}$, with $A \in C^1(\mathbb{R})$, $\phi \in C^2(\mathbb{R})$, and both $A(t)$ and $\phi'(t)$ are slowly varying. With the same arguments as before, we obtain $T_f(b,a) = A(b)e^{i\phi(b)}\overline{\hat{\psi}(a\phi'(b))} + R(b,a)$, where $|R(b,\omega)|$ is again controlled by the speed of variation of A and ϕ'. Assuming that $\psi(t)$ is a smooth function whose Fourier transform is peaked at a particular frequency ω_0 (by definition of a wavelet, $\omega_0 \neq 0$), and assuming that $R(b,a)$ is small enough to be neglected, $|T_f(b,a)|$ is peaked near a curve (the *ridge* of the wavelet transform) of equation $a = \omega_0/\phi'(b)$, which again reproduces the frequency modulation of the signal, in a slightly different way (multiplicative instead of additive).

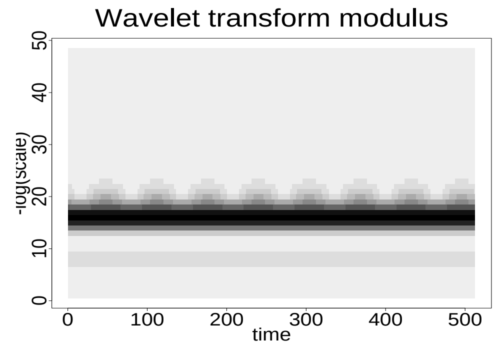

FIGURE 3. Continuous wavelet transform of the frequency modulated signal of Figure 2.

There is a major difference between wavelet systems and Gaborlet systems: Gaborlets are functions of *constant size*, and *variable shape*, while wavelets have *constant shape*, and *variable size*. An illustration of this fact may be seen in Figure 3, where we display the wavelet transform modulus of the frequency modulated signal analyzed in Figure 2. The wavelets have sharp frequency localization at low frequencies, and sharp time localization at high frequencies. Therefore, wavelets analyze this particular signal as follows: at low frequencies the time resolution is not good enough to capture the frequency changes, but the frequency resolution is extremely good, and the analysis exhibits clearly the fundamental frequency of the signal. At higher frequencies the wavelets have smaller support, and exhibit time dependent frequencies.

REMARK 3.5. *Figures 2 and 3 exhibit striking differences between wavelets and Gaborlets, and open the problem of selecting the "best" representation for a given signal. We shall address this problem briefly in a subsequent section.*

REMARK 3.6. *Time-frequency analysis is certainly not the main application of wavelet analysis. Wavelets are particularly well adapted to all problems which present some scale invariance properties. A good illustration is provided by the characterization of singularities and its applications to multifractal analysis (see e.g. [3] for a review). More practical applications include the characterization of long range correlations in* $1/f$*-type processes. We shall see a few examples below.*

3.2. *Weighted spectra of second order random time series, and sample estimates*

Wavelet transforms and Gabor transforms may yield alternatives to the Wigner function, i.e. bilinear time-frequency (or time-scale) representations. We shall use the generic term *weighted spectra* for alternatives obtained in such ways. Let us consider a second order random time series $\{X_t, t \in \mathbb{R}\}$. Its continuous Gabor transform is defined analogously to (15), as the *time-frequency series*

$$G_X(b,\omega) = \langle X, g_{(b,\omega)} \rangle \ . \tag{21}$$

Similarly, one introduces the continuous wavelet transform of $\{X_t, t \in \mathbb{R}\}$ by

$$T_X(b,a) = \langle X, \psi_{(b,a)} \rangle \ . \tag{22}$$

It may be proved that this defines a second order stochastic processes. This motivates the following definition:

DEFINITION 3.7. *Let* $\{X_t, t \in \mathbb{R}\}$ *be a second order time series.*

1. *Let* $g \in L^2(\mathbb{R})$, *normalized so that* $||g|| = 1$. *The* Gabor spectrum *of the time series* $\{X_t, t \in \mathbb{R}\}$ *is defined by*

$$\mathcal{E}_X^{(G)}(b,\omega) = \mathbb{E}\left\{ |G_X(b,\omega)|^2 \right\} \ . \tag{23}$$

2. *Let* $\psi \in L^1(\mathbb{R})$ *be a wavelet, normalized so that* $||\psi|| = 1$. *The* wavelet spectrum *of the process is defined by*

$$\mathcal{E}_X^{(W)}(b,a) = \mathbb{E}\left\{|T_X(b,a)|^2\right\} . \tag{24}$$

Notice that in the particular case of (weakly) stationary time series, these quantities do not depend on b anymore (like the Wigner spectrum).

REMARK 3.8. *For a given window* $g \in L^2(\mathbb{R})$, $||g|| \neq 0$, *the Gaborlets* $g_{(b,\omega)}(t)$ *form a complete set in* $L^2(\mathbb{R})$. *Therefore, the covariance* $\mathcal{C}$ *of a second order time series is completely characterized by the matrix elements* $\langle \mathcal{C} g_{(b,\omega)}, g_{(b',\omega')}\rangle$. *However, we have by definition*

$$\mathcal{E}_X^{(G)}(b,\omega) = \langle \mathcal{C} g_{(b,\omega)}, g_{(b,\omega)}\rangle \ .$$

Therefore, one cannot expect to characterize $\mathcal{C}$ *by its Gabor spectrum* unless the "matrix" $\langle \mathcal{C} g_{(b,\omega)}, g_{(b',\omega')}\rangle$ is sharply localized near its diagonal. *The time series for which such a property holds true may be termed "locally stationary", and Gabor analysis may be used to study them. Such time series have been considered by several authors in the literature. See e.g. [14,16,17].*

A similar remark holds when Gaborlets are replaced with wavelets. The class of time series which are well characterized by the wavelet spectrum is different however, and basically corresponds to time series whose covariance has a simple behavior under rescalings and translations. We shall see some examples below.

The simplest estimators for weighted spectra are again the sample estimators. Let us consider, for example, the wavelet case. Let $X^{(1)}, \dots X^{(N)}$ be N independent realizations of the time series $\{X_t, t \in \mathbb{R}\}$, let $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ be a wavelet such that $||\psi|| = 1$, and set

$$\tilde{\mathcal{E}}^{(W)}(b,a) = \frac{1}{N}\sum_{n=1}^{N} |T_{X^{(n)}}(b,a)|^2 \ . \tag{25}$$

Similarly, if $g \in L^2(\mathbb{R})$ is such that $||g|| = 1$, we set

$$\tilde{\mathcal{E}}^{(G)}(b,\omega) = \frac{1}{N}\sum_{n=1}^{N} |G_{X^{(n)}}(b,\omega)|^2 \tag{26}$$

We shall see below that such estimators may be used as estimators for the Wigner-Ville spectrum, or even of the usual power spectrum in the stationary case (in which case they are close to classical Welsh-Bartlett estimators). This generally leads to biased, but smoother, estimates.

3.3. *Weighted spectra as approximations*

We have seen in the previous subsection how wavelet or Gabor transforms may be used to introduce local versions of power spectra. The following result is well known (see for example [9]) and follows from a simple calculation

PROPOSITION 3.9. *Let* $\{X_t, t \in \mathbb{R}\}$ *be a second order time series, and let* $\mathcal{E}_X$ *and* $\mathcal{E}_X^{(G)}$ *denote respectively its Wigner spectrum and its Gabor spectrum. Then we have the following*

$$\mathcal{E}^{(G)}(b,\omega) = \frac{1}{2\pi}\int \mathcal{E}_X(b',\omega')\overline{\mathcal{E}_g}(b'-b,\omega'-\omega)\, db' d\omega' \tag{27}$$

where $\mathcal{E}_g(b,\omega)$ *is the Wigner function of the window* $g(t)$. *A similar result holds true in the case of the wavelet spectrum:*

$$\mathcal{E}^{(W)}(b,a) = \frac{1}{2\pi}\frac{1}{a}\int \mathcal{E}_X(b+a\tau,\omega/a)\overline{\mathcal{E}_\psi}(\tau,\omega)\, d\tau\, d\omega \ . \tag{28}$$

This shows that the two weighted spectra we have described above may be seen as approximations, or smoothings, of the Wigner-Ville spectrum. The corresponding sample estimators may therefore be expected to be smoother than the sample estimator given in (13).

The case of weakly stationary second order random time series is an interesting particular case. Let $\{X_t, t \in \mathbb{R}\}$ be such a stationary time series. Then one easily verifies that

• The random time series $\{G_x(b,\omega), b \in \mathbb{R}\}$ and $\{T_X(b,a), b \in \mathbb{R}\}$ are second order, weakly stationary, random time series.

• The two weighted spectra $\mathcal{E}^{(W)}(b,a)$ and $\mathcal{E}^{(G)}(b,\omega)$ are, respectively, functions of a and ω only.

This suggests to modify the sample estimators by smoothing with respect to the b variable to improve their smoothness. Cases using the simplest smoothing, for example

$$\tilde{\mathcal{E}}^{(G)}(\omega) = \frac{1}{B}\int_{b_0-B/2}^{b_0+B/2} \tilde{\mathcal{E}}^{(G)}(b,\omega)db \ , \tag{29}$$

$$\tilde{\mathcal{E}}^{(W)}(a) = \frac{1}{B}\int_{b_0-B/2}^{b_0+B/2} \tilde{\mathcal{E}}^{(W)}(b,a)db \ , \tag{30}$$

where B and b_0 are fixed parameters, have been studied in [4]. In particular, one easily shows that

$$\mathbb{E}\left\{\tilde{\mathcal{E}}^{(G)}(\omega)\right\} = \mathbb{E}\left\{\tilde{\mathcal{E}}^{(G)}(b,\omega)\right\} = \frac{1}{2\pi}\int \mathcal{E}(\omega')|\hat{g}(\omega-\omega')|^2\, d\omega' \ , \tag{31}$$

$$\mathbb{E}\left\{\tilde{\mathcal{E}}^{(W)}(a)\right\} = \mathbb{E}\left\{\tilde{\mathcal{E}}^{(W)}(b,a)\right\} = \frac{1}{2\pi}\int \mathcal{E}(\omega)|\hat{\psi}(a\omega)|^2\, d\omega \tag{32}$$

This shows that the two estimators $\tilde{\mathcal{E}}^{(G)}(\omega)$ and $\tilde{\mathcal{E}}^{(W)}(a)$ yield smoothed versions of the power spectrum $\mathcal{E}(\omega)$, the smoothing being a standard convolution in the Gabor case and a multiplicative convolution in the wavelet case.

REMARK 3.10. *Notice that the estimator $\tilde{\mathcal{E}}^{(G)}(\omega)$ is very much in the spirit of the so-called* Welsh-Bartlett estimator, *a standard tool for spectral estimation. The Welsh-Bartlett estimator is obtained by computing local (tapered) periodograms of the signal, and then taking the average of these local spectra. This is basically what the averaged Gabor spectral estimator $\tilde{\mathcal{E}}^{(G)}(\omega)$ does. The averaged wavelet estimator $\tilde{\mathcal{E}}^{(W)}(a)$ (which may be seen as a spectral estimator by considering the scale as an inverse frequency variable) does a similar job, the difference being that the window size changes proportionally to the inverse of the frequency (the higher the frequency, the smaller the window). Examples of spectral estimation are presented in Section 4 below.*

3.4. *Adaptive Decompositions*

We have seen that different types of weighted spectra are adapted to different types of time series, i.e., different types of covariances. In practice, the covariance is not known, and has to be estimated from one or a few realizations of the time series. In such a context the choice of the method, or say the choice of the window $g(t)$ in the Gabor case, is not innocuous. Several approaches have been proposed to solve such problems (see, for example, [14, 16]). In most cases the objective is to find the decomposition which make the corresponding weighted spectrum "as diagonal as possible". Such a requirement may be realized in different ways.

Let us describe here the solution proposed in [14], in a slightly more general context than in [14]. Let us assume that we are given a family of functions (for example wavelets, or Gaborlets, or more general functions), denoted by $\mathcal{L} = \{\psi_\lambda, \lambda \in \Lambda\}$, where Λ is some measure space, normalized so that $||\psi_\lambda|| = 1$ for all $\lambda \in \Lambda$. Assume further that there exists an associated reproducing formula: there exists a measure $d\mu$ on Λ such that $\forall f \in L^2(\mathbb{R})$,

$$f = \int_\Lambda \langle f, \psi_\lambda \rangle \psi_\lambda d\mu(\lambda) \tag{33}$$

weakly in $L^2(\mathbb{R})$. Continuous Gabor and wavelet transforms provide examples of such reproducing formulas. Other examples have been studied, in particular in [12, 18].

Let $\mathcal{C}$ be the covariance of a second order time series $\{X_t, t \in \mathbb{R}\}$, and consider the generalized weighted spectrum

$$\mathcal{E}^{(\mathcal{L})}(\lambda) = \mathbb{E}\left\{ |\langle X, \psi_\lambda \rangle|^2 \right\} = \langle \mathcal{C}\psi_\lambda, \psi_\lambda \rangle \ . \tag{34}$$

Let us write

$$\mathcal{C}\psi_\lambda = \mathcal{E}^{(\mathcal{L})}(\lambda)\psi_\lambda + r_\lambda$$

where r_λ is some remainder such that $r_\lambda \perp \psi_\lambda$. In order to "almost diagonalize" the covariance $\mathcal{C}$, a possible approach amounts to making $||r_\lambda||$ as small as possible for all λ. The solution proposed in [14] amounts to searching for the optimal decomposition $\mathcal{L}$ by solving

$$\mathcal{L}_{opt} = \arg \min_{\mathcal{L}} \int ||r_\lambda||^2 \, d\mu(\lambda). \tag{35}$$

This program may be justified in the following situation:

PROPOSITION 3.11. *Assume that $\mathcal{C}$ is Hilbert-Schmidt (i.e. $||\mathcal{C}||_{HS}^2 = Tr(\mathcal{C}^*\mathcal{C}) < \infty$). Then $\int ||r_\lambda||^2 \, d\mu(\lambda) = ||\mathcal{C}||_{HS}^2 - ||\mathcal{E}^{(\mathcal{L})}||_{L^2(\Lambda)}^2$, and problem (35) is equivalent to*

$$\mathcal{L}_{opt} = \arg \max_{\mathcal{L}} ||\mathcal{E}^{(\mathcal{L})}||_{L^2(\Lambda)}^2 \tag{36}$$

The proposition follows from a simple calculation. The fact that $\mathcal{E}^{(\mathcal{L})} \in L^2(\Lambda)$ is also verified directly from the reproducing formula (33).

This program has been carried out by W. Kozek in [14] in the particular case of Gaborlets. The families $\mathcal{L}$ are families of Gaborlets $\mathcal{L} = \mathcal{L}_g = \{g_{(b,\omega)}, b, \omega \in \mathbb{R}\}$ with different window functions $g(t)$, $||g|| = 1$. In that particular case it may be shown that the "optimal window" is that one which maximizes the scalar products of the square moduli of the ambiguity functions of the time series $|\mathcal{A}_X|^2$ and of the window $|\mathcal{A}_g|^2$, i.e. the problem (35) becomes

$$g_{opt} = \arg \max_{g \in L^2(\mathbb{R}), ||g||=1} ||\mathcal{E}^{(G)}||_{L^2(\mathbb{R}^2)}^2 = \arg \max_{g \in L^2(\mathbb{R}), ||g||=1} \langle |\mathcal{A}_X|^2, |\mathcal{A}_g|^2 \rangle \tag{37}$$

This provides a simple interpretation to this problem, in the light of the example of Figure 2: the optimization searches a window whose spreading in the time-frequency domain matches best that of the time series.

More generally, there is no reason for a time series to be well described by constant size Gaborlets, even if the time series may be considered locally stationary. This implies that most of the time it is not sufficient to limit oneself to Gaborlets, and the decomposition has to be looked for in larger families. Examples have been studied e.g. in [12, 18]. Another approach may be found in [16], in a different context which we briefly mention now.

3.5. *Remark: Discretization, and Adaptive Spectral Decomposition*

So far we have limited our analysis to the case of time series defined for continuous time, and considered only continuously labeled decompositions. In practice, one clearly needs to develop discrete analogs of these techniques.

Discretizations of the continuous time-frequency decompositions have been discussed in several places (see, for example, [4, 5, 7, 13, 22]). The first main result (existence of wavelet and Gabor frames, see, for example, [7]) is that the continuous formulae in (16) and (20) may be discretized without introducing important distortions, as long as the discretization step is small enough. By doing so, one generally keeps some redundancy in the decompositions.

However, it is desirable in some contexts to go further and get rid of redundancy as much as possible, and in particular use (orthonormal) bases when possible. This is the case in particular for all applications related to signal compression. As we have seen above, the optimal representation, in terms of reduction of variance, is the Karhunen-Loeve decomposition (see equation (1)). However, we have stressed already at the beginning of Section 2 that such decompositions may sometimes be of poor interest in practice. A possible alternative, in the spirit of our previous discussion, amounts to looking for the “optimal basis decomposition” within a *library of bases*, generated in a simple and systematic way. This best basis paradigm, proposed by Coifman and Wickerhauser first and developed systematically since then (see in particular [22]), has been applied recently by Mallat, Papanicolaou and Zhang in [16] for constructing approximate spectral decompositions for locally stationary processes. The construction makes use of the *local trigonometric bases*, which may be understood as (generalized) bases of Gaborlets. We refer to [16] for the details, and to [8] for an alternative approach.

4. Examples

We have seen in the previous section how time-frequency or time-scale transforms may be used to build spectral estimators, adapted to stationary or non-stationary situations. We now illustrate the ideas with a couple of examples of spectral estimation based upon wavelet and Gabor transforms. Our purpose is *not* to provide a systematic comparison of estimation techniques in a variety of situations. We just present two situations for which the methods we just explained are well adapted.

It is readily seen from equation (31) that if $\mathcal{E}(\omega) = Ce^{-\lambda\omega}$ for some constants λ, C, and if $g(t)$ is chosen in such a way that $K = \frac{1}{2\pi}\int e^{-\lambda\omega}|\hat{g}(\omega)|^2 d\omega < \infty$, then $\mathbb{E}\left\{\tilde{\mathcal{E}}^{(G)}(\omega)\right\} = K\mathcal{E}(\omega)$. Of course, such a choice for $\mathcal{E}(\omega)$ is not suitable for the spectral density of a stationary time series. However, if $\mathcal{E}(\omega) \sim e^{-\lambda\omega}$ within a given frequency domain, we may expect $\tilde{\mathcal{E}}^{(G)}(\omega)$ to provide an unbiased estimation of $\mathcal{E}(\omega)$ within this frequency domain.

Figure 4 is an illustration of this case. We consider an example of discrete (weakly stationary) time series, whose spectral density is a decaying exponential function of the frequency:

$$\mathcal{E}(\omega) = \exp\{-\lambda|\omega|\} ,$$

for $-\pi \leq \omega \leq \pi$. In such a case one expects $\tilde{\mathcal{E}}^{(G)}$ to provide a good estimation, except perhaps in the neighborhood of the origin of frequencies.

In the example, λ has been set to $\lambda = 2$. A standard periodogram-based estimation (shown in the right top plot of the figure) followed by a linear regression yields an estimate $\tilde{\lambda} = 2.0059$. A part of the Gabor spectrum $\tilde{\mathcal{E}}^{(G)}(b,\omega)$ is represented in the left bottom figure, and the corresponding b-averaged version $\tilde{\mathcal{E}}^{(G)}(\omega)$ is displayed in the right bottom plot. The estimate for λ obtained in this case was $\tilde{\lambda} = 2.003$. Several simulations in the same “experimental” situations have shown that the Gabor estimate has a variance about twice smaller than the estimate based on the periodogram. This is not surprising, for we have already remarked on the similarity of this approach with the Welsh-Bartlett estimator, which was introduced for that purpose.

Our second example illustrates the behavior of wavelet-based spectral estimation. As we have remarked already, the wavelet spectrum is more adapted to time series whose covariance possesses some scale invariance properties. This is the case for fractional Brownian motion,

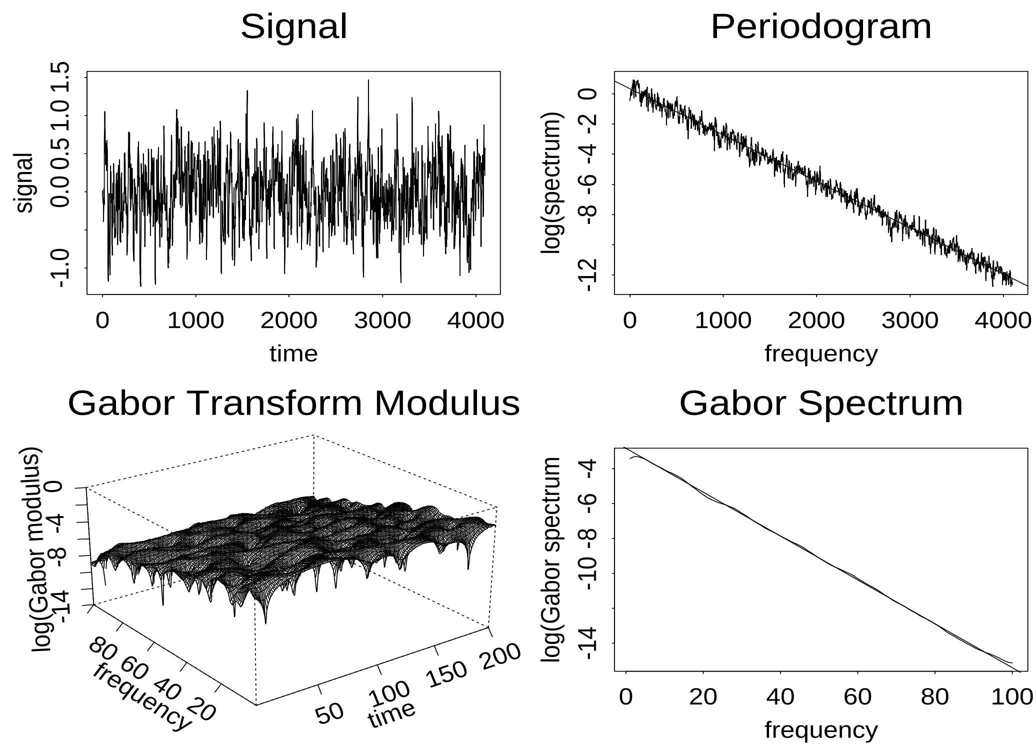

FIGURE 4. Example of spectral estimation using the Gabor spectrum. The signal (left top plot) is a Gaussian process, whose spectral density is an exponential function of the frequency. The logarithm of the periodogram estimate of the spectral density is shown in the right top plot, superimposed with a regression line. The (estimated) Gabor spectrum (on a small part of the signal) is displayed as a surface plot in the left bottom figure, and its average with respect to the time variable is presented in the right bottom plot, together with a corresponding regression line.

which we now consider. Fractional Brownian motion (fBm) of Hurst exponent h is a Gaussian process $\{X_t, t \in \mathbb{R}\}$ with zero mean and covariance

$$\mathbb{E}\{X_t X_s\} = \frac{\sigma^2}{2}\left\{|t|^{2h} + |s|^{2h} - |t-s|^{2h}\right\} .$$

This is an interesting example of a non stationary process. It has received great attention lately because it exhibits long range correlations, a phenomenon which has been observed in various contexts. However, it is known that because of these long range correlations, the estimation of the Hurst exponent h (and the variance parameter σ) is a difficult task, for sample estimators turn out to have a large variance.

Remarkably enough, the fixed-scale wavelet transform $\{T_X(b,a), b \in \mathbb{R}\}$ of such a time series is a weakly stationary time series, so that the discussion of the previous section may be extended to the present situation. We show how the wavelet spectrum we described in the previous section may be used for estimating h from a single realization.

We illustrate it with one realization, with Hurst exponent $h = 0.2$. Figure 5 shows the realization of the time series (left top plot), together with a periodogram-based estimation. As can be seen from the right top plot, the lack of smoothness of the periodogram makes it difficult to estimate the Hurst exponent. This is to be compared with the right bottom plot, representing the (logarithm of the) wavelet spectrum $\tilde{\mathcal{E}}^{(W)}(a)$. For such processes, it may be shown (see, for example, [4]) that

$$\mathbb{E}\left\{\tilde{\mathcal{E}}^{(W)}(a)\right\} \sim a^{2h} ,$$

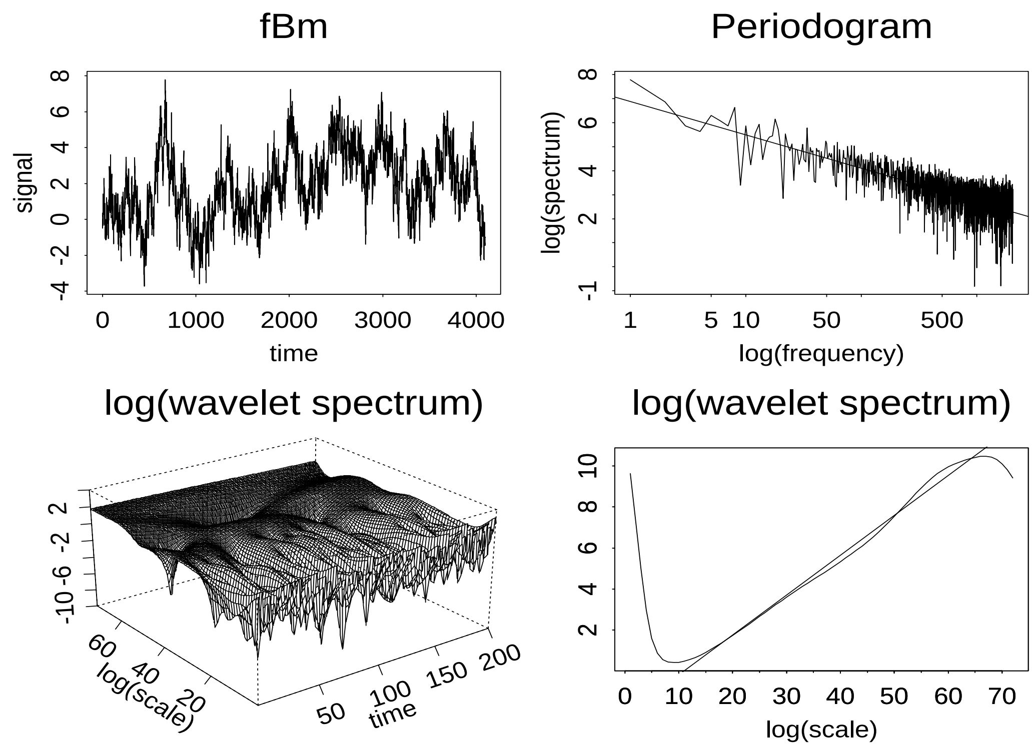

FIGURE 5. Example of spectral estimation using the wavelet spectrum. The signal (left top plot) is a realization of a fractional Brownian motion, with Hurst exponent $h = .2$. The logarithm of the periodogram estimate of the spectral density is shown in the right top plot, superimposed with a regression line. The (estimated) wavelet spectrum (on a small part of the signal) is displayed as a surface plot in the left bottom figure, and its average with respect to the time variable is presented in the right bottom plot, together with a corresponding regression line.

as soon as $\psi(t)$ has been chosen in such a way that $\int |\omega|^{2h-1}|\hat{\psi}(\omega)|^2\, d\omega < \infty$, so that a linear regression on a *log-log* plot of the wavelet spectrum directly provides an estimate for the exponent. In our case the estimate was $\tilde{h} = 0.196$ (notice that, as in the case of the periodogram, all the scales could not be used, the smallest scales being corrupted by additional noise).

Such methods have been carefully analyzed (see e.g. [1, 4]). In particular, P. Abry has shown that estimators of the type we are considering are unbiased and of minimal variance.

5. Conclusions

We have described a series of methods designed to provide simple representations of deterministic and random signals, emphasizing the so-called *time-frequency methods*. Besides more sophisticated tools such as Wigner's functions and its generalizations, we have shown that simple decompositions such as Gabor or wavelet transforms yield efficient algorithms, in particular for spectral estimation. These two particular methods are well adapted to specific situations, and extensions to more general contexts require using more general (adaptive) decomposition methods.

The illustrations of this papers have been generated using the `Swave` package developed by R. Carmona, W.L. Hwang and the author (see [4]). `Swave` (based on the `Splus` environment), is available by anonymous `ftp` at the site:

`http://soil.princeton.edu/ rcarmona`

and documented in [4].

References

[1] P. Abry, P. Gonçalves and P. Flandrin (1994): Wavelets, spectrum analysis and $1/f$ processes. *in Wavelets and Statistics*, A. Antoniadis & G. Oppenheim Eds, *Lecture Notes in Statistics, 103*, 15-29.

[2] A. Antoniadis & G. Oppenheim Eds (1994): Proceedings of the Conference *Wavelets and Statistics*, Villard de Lans, France, *Lecture Notes in Statistics. 103*, Springer Verlag, New York, N.Y.

[3] A. Arneodo, F. Argoul, E. Bacry, J. Elezgaray and J.F. Muzy (1995): *Ondelettes, Multifractales et Turbulence*, Diderot éditeur des Arts et Sciences, Paris.

[4] R. Carmona, W.L. Hwang and B. Torrésani (1998): *Practical Time-Frequency Analysis: Wavelet and Gabor transforms, with an implementation in S*, Academic Press, to appear.

[5] C.K. Chui (1992): *An Introduction to Wavelets*. Wavelet Analysis and Its Applications, 1, Academic Press, San Diego.

[6] L. Cohen (1995): *Time–frequency Analysis*. Prentice Hall, Englewood Cliffs (NJ).

[7] I. Daubechies (1992): *Ten Lectures on Wavelets*. CBMS-NFS Regional Series in Applied Mathematics 61.

[8] D. Donoho, S. Mallat and R. Von Sachs (1998): Estimating Covariances of Locally Stationary Processes: Rates of Convergence of Best Basis Methods. Technical report, Dept of Statistics, Stanford University.

[9] P. Flandrin (1993): Temps-Fréquence. *Traité des Nouvelles Technologies, série Traitement du Signal, Hermes*.

[10] D. Gabor (1946): Theory of communication. *J. Inst. Elec. Eng.* 903, 429.

[11] A. Grossmann, J. Morlet (1984): Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape. *SIAM J. of Math. An.* 15, p.723.

[12] J.A. Hogan and J.D. Lakey (1995): Extensions of the Heisenberg Group by Dilations and Frames. *Appl. and Comp. Harm. Anal* 2, pp. 174-199.

[13] M. Holschneider (1995): *Wavelets; An Analysis Tool*. Oxford Mathematical Monographs, Clarendon Press, Oxford.

[14] W. Kozek (1996): *Spectral Estimation in Non-Stationary Environments*. PhD Thesis, Vienna.

[15] S. Mallat (1998): *A Wavelet Tour of Signal Processing*. Academic Press.

[16] S. Mallat, G. Papanicolaou and Z. Zhang (1995): Adaptive Covariance Estimation of Locally Stationary Processes, Preprint (available by anonymous ftp at cs.nyu.edu/pub/wave), to appear in *Advances in Statistics*.

[17] M.B. Priestley (1965): Evolutionary Spectra and Non-Stationary Processes, *J. Roy. Stat. Soc.* B27, 204-237. Design relations for Non-Stationary Processes, *J. Roy. Stat. Soc.* B28, 228-240.

[18] B. Torrésani (1992): Time-Frequency Distributions: Wavelet Packets and Optimal Decompositions. *Annales de l'Institut Henri Poincaré 56#2*, 215–234.

[19] B. Torrésani (1996): *Analyse Continue par Ondelettes*. Inter Editions/Editions du CNRS (in French). English translation to appear at SIAM (translated and revised by R.D. Ryan).

[20] M. Vetterli and J. Kovacevic (1996): *Wavelets and SubBand Coding*, Prentice Hall.

[21] J. Ville (1948): Théorie et Applications, de la Notion de Signal Analytique. *Cables et Transmissions 2*, pp. 61-74. Translated into English by I. Selin, RAND Corp. Report T-92, Santa Monica, CA (August 1958).

[22] M.V. Wickerhauser (1994): *Adapted Wavelet Analysis, from Theory to Software*. A.K. Peters Publ.

[23] E.P. Wigner (1932): On the Quantum Corrections for the Thermodynamic Equilibrium. *Phys. Rev. 40*, 749-759