Multiridge Detection and Time–Frequency Reconstruction

René A. Carmona, *Member, IEEE*, Wen L. Hwang, and Bruno Torrésani

***Abstract*— The ridges of the wavelet transform, the Gabor transform, or any time–frequency representation of a signal contain crucial information on the characteristics of the signal. Indeed, they mark the regions of the time–frequency plane where the signal concentrates most of its energy. We introduce a new algorithm to detect and identify these ridges. The procedure is based on an original form of Markov chain Monte Carlo algorithm especially adapted to the present situation. We show that this detection algorithm is especially useful for noisy signals with multiridge transforms. It is a common practice among practitioners to reconstruct a signal from the skeleton of a transform of the signal (i.e., the restriction of the transform to the ridges.) After reviewing several known procedures, we introduce a new reconstruction algorithm, and we illustrate its efficiency on speech signals and its robustness and stability on chirps perturbed by synthetic noises at different SNR's.**

***Index Terms*—Continuous wavelet transform, redundancy, signal detection, signal reconstruction, stochastic relaxation methods, time–frequency analysis.**

I. Introduction

A WIDE CLASS of signals may be conveniently described in terms of time-dependent amplitude and frequency (see, for example, [9] or [2]) or sums of such amplitude and frequency-modulated components (like in speech analysis and synthesis applications [12], [13], [17].) However, the main problem is still the numerical estimation of these time-dependent characteristics. Time–frequency representations [9] offer a convenient setup, and the problem of the estimation of the local amplitude and frequency is well understood in the noise-free case with only one component. However, the situation becomes more problematic in the presence of noise and/or of several components.

We proposed in [5] a new constrained optimization approach for processing one-component signals in very noisy situations. At high signal-to-noise ratios (SNR's), one-component signals can be analyzed by means of their instantaneous amplitude and frequency, for which the estimation theory is well developed [2]. The nonparametric approach used in [5] was based on the solution of variational problems. This strategic choice was motivated by the desire to handle low SNR's. We also note that bilinear representations such as the (generalized) Wigner distributions [3], [9] can be extremely precise in the one-component case but may completely fail in the multicomponent situation because of interference terms. We describe here a new approach capable of handling multiple component signals. This new approach is based on a new Markov chain Monte Carlo (MCMC) approach, which uses the energetic distribution provided by a time–frequency representation of the signal. Given the fact that the energy of the signal concentrates around curves in the time–frequency plane, which we shall call "ridges," the Markov chain is constructed in such a way that the random walkers (hereafter called "crazy climbers") are attracted by these 1-D structures. The analysis is complemented by an "synthesis step" devoted to the reconstruction of the "part(s)" of the signal that produced the ridge(s).

Throughout the paper, the discussion is restricted to the cases of the Gabor and wavelet transforms. Notice that since our detection algorithm is only a special postprocessing of a time–frequency transform, it can be used with other time–frequency energetic representations, for example, the family of Wigner distributions discussed in [1]. On the contrary, the reconstruction algorithm is specific to the representations. We develop it for wavelet and Gabor transforms, but the modifications required to extend it to other linear time–frequency representations are straightforward.

The thrust of the present paper is twofold. First, we give a new ridge detection algorithm that can efficiently detect multiple ridges in the modulus of a transform, and second, we propose a signal reconstruction procedure from the knowledge of the skeleton of the transform on arbitrary points of its ridges. Our detection procedure is based on an original Markov chain Monte Carlo algorithm. It is designed in such a way that weighted occupation densities draw the ridges on the time–frequency plane. Most importantly, its robustness to noise is remarkable. The reconstruction procedure is restricted to linear transforms. It is based on the classical idea of spline smoothing as presented in [20]. We already used it in the case of the wavelet transform (without much explanation) in the companion correspondence [5]. We present it in full detail here. As for the ridge detection, it performs very well in noisy situations. Both components of our work (ridge detection and reconstruction) are illustrated on two specific data sets. The first one is the superposition of a real life sonar bat signal with an artificially generated chirp. It is analyzed with the wavelet transform. The second one is the speech recording considered in [14]. We analyze it using the Gabor transform.

Manuscript received August 15, 1996; revised May 17, 1998. This work was supported in part by the Office of Naval Research under Grant N00014-96-1-0367. The associate editor coordinating the review of this paper and approving it for publication was Prof. Moeness Amin.

R. A. Carmona is with the Statistics and Operations Research Program and the Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544 USA.

W. L. Hwang is with the Institute of Information Science, Academia Sinica, Nanking, Taipei, Taiwan, R.O.C.

B. Torrésani is with CPT, CNRS-Luminy, Marseille Cedex, France.

Publisher Item Identifier S 1053-587X(99)00731-X.

We close this introduction with a short summary of the contents of the paper. After a short section devoted to notation and discretization issues, our detection algorithm is presented in a general setting in Section III. We also try to emphasize the similarities and the differences with known procedures such as minimization by simulated annealing or the more recent "reassignment" procedures advocated in [1] and [14]. Section IV contains a detailed discussion of a first numerical example: a discussion of the specifics of noisy signals and a Monte Carlo analysis of the robustness of the ridge detection algorithm. Section V gives our reconstruction procedure of the original signal from the estimates of the transform on the ridges. Some details of the penalization procedure are postponed to the Appendix. Because we choose to illustrate the efficiency of this reconstruction on speech signals, a short Section VI is devoted to the specifics of the sinusoidal model for speech.

The algorithms presented in this paper have been implemented in the S-plus environment. The data files and the S-code needed to produce all the numerical results and figures given as examples in this paper are available on the Internet [6].

II. Notation for the Continuous Gabor and Wavelet Transforms

We set the stage by introducing the time–frequency representations that we use to illustrate the crazy climber algorithm. Even though the latter may be used as post processing of any time–frequency representation, we shall restrict the present discussion to the cases of the continuous wavelet and Gabor transforms for the sake of simplicity. Indeed, the behavior of these transforms is easy to understand for amplitude- and frequency-modulated signals.

We work with the complex Hilbert space $L^2(\mathbb{R})$ of square-integrable functions. Our convention for the Fourier transform is $\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-i\xi x}\,dx$, and consequently, the Plancherel formula reads $\|\hat{f}\|^2 = 2\pi\|f\|^2$.

A. Continuous Time–Frequency Transforms

Let $\psi(x)$ be a fixed integrable function such that $0 < c_\psi = \int_0^\infty |\hat{\psi}(\xi)|^2\ (d\xi/\xi) < \infty$. Such a function ψ is called an analyzing wavelet. We use the notation

$$\psi_{(b,a)}(x) = \frac{1}{a}\,\psi\left(\frac{x-b}{a}\right) \tag{1}$$

for the wavelet with scale a and location b. The wavelet transform of a signal $f \in L^2(\mathbb{R})$ with respect to ψ is defined by

$$T_f(b,a) = \langle f, \psi_{(b,a)}\rangle = \frac{1}{a}\int_{-\infty}^{\infty} f(x)\overline{\psi}\left(\frac{x-b}{a}\right)dx. \tag{2}$$

Throughout this paper, we restrict ourselves to complex-valued wavelet belonging to the Hardy space $H^2(\mathbb{R})$, i.e., such that $\hat{\psi}(\xi) = 0\ \forall \xi \le 0$. In addition, we find convenient to introduce the auxiliary variable $\varphi = \log(a)$. With such notation, it follows directly from Taylor's formula that if we consider signals of the form

$$f(x) = \sum_{k=1}^{N} A_k(x)\,\cos(\phi_k(x)) \tag{3}$$

where the amplitudes $A_k(x)$ are continuously differentiable and the phases $\phi_k(x)$ are twice continuously differentiable, then their wavelet transform can be written in the form

$$T_f(b,a) = \frac{1}{2}\sum_{k=1}^{N} A_k(x)e^{i\phi_k(b)}\overline{\hat{\psi}(a\phi_k'(b))} + r(b,a) \tag{4}$$

with $r(b,a) \sim O(|A_k'|, |\phi_k''|)$. Therefore, if $\hat{\psi}(\xi)$ is localized near a certain value $\xi = \omega_0$ in the frequency domain, the wavelet transform square modulus $M(b,a) = |T_f(b,a)|^2$ is localized near the N curves with equations $a = a_k(b) = \omega_0/\phi_k'(b)$. These curves are called the *ridges of the transform*. Our goal is to present efficient algorithms to estimate ridges.

Next, we describe the case of the Gabor transform (also called STFT.) Although Gabor's original representation was discrete, we use a continuous version which we still call the Gabor transform. The Gabor transform of $f \in L^2(\mathbb{R})$ is defined as

$$G_f(b,\omega) = \int_{-\infty}^{\infty} f(x)g(x-b)e^{-i\omega(x-b)}\,dx \tag{5}$$

where $g(x)$ is a window function with a good time–frequency localization. We shall use the notation

$$g_{(b,\omega)}(x) = g(x-b)e^{i\omega(x-b)} \tag{6}$$

for the time–frequency atoms used in the Gabor transform. The same argument as before shows that the continuous Gabor transform of signals of the type (3) may be written as

$$G_f(b,\omega) = \frac{1}{2}\sum_{k=1}^{N} A_k(x)e^{i\phi_k(b)}\overline{\hat{g}(\phi_k'(b)-\omega)} + r(b,\omega). \tag{7}$$

Again, the remainder term $r(b,\omega)$ depends on the derivatives of the amplitudes and the local frequencies. Assuming, for simplicity (this is the case for the Gaussian windows as well as for the Hamming windows), that the Fourier transform of the window has fast decay away from the origin of frequencies, we end up again with a Gabor transform square modulus $M(b,\omega) = |G_f(b,\omega)|^2$ exhibiting a certain number of ridges.

B. Discretizations

In practice, we have only access to discrete data, and the wavelet and Gabor transforms have to be discretized. The discretization of the Gabor transform is traditionally a regular sampling of the continuous formulas. In practice, we sample the Gabor transform at the same rate as the signal.

Sampling the wavelet transform is somewhat more problematic. Indeed, the classical discretization scheme for wavelet transforms involves the so-called *dyadic grid* of the form $(kb_0a_0^j, a_0^j), j, k \in \mathbb{Z}$. Such a choice is not convenient for our purpose because it is *not* shift-invariant. The discrete transform of a shifted copy of the signal is not the shifted copy of the original signal's transform. To overcome such a drawback, it is convenient to introduce more redundancy

into the transform and work with discretization grids of the form $(kb_0, a_0^j), j, k \in \mathbb{Z}$, or in terms of the auxiliary variable φ, we will use a grid of the form $(kb_0, j\varphi_0), j, k \in \mathbb{Z}$, with $\varphi_0 = \log(a_0)$, i.e., a regular grid again.

III. Drawing the Ridges of a Surface

As we have seen, characterization of the signal's instantaneous frequency by the wavelet or the Gabor transform can be achieved by estimating numerically the ridges as the sets of local maxima of the transforms square moduli, which are denoted hereafter by $M(b, \varphi)$, where φ is either the log of the scale variable (wavelet case) or the frequency variable (Gabor case). The problem we are addressing in this section is to determine the ridges of the surface $M(b, \varphi)$. When the surface is the energy landscape over the time–frequency plane given by a specific transform, a second challenging problem is, for each ridge, to estimate the corresponding component of the signal and reconstruct it. We shall address this second problem in Section V.

According to [10], ridges may be defined in a very general setting as curves on a surface $z = f(x, y)$. They are completely characterized by their projection on the (x, y) plane. We also call these curves ridges by a convenient abuse of language. We will be interested in the special class of ridges we describe below.

We start with a subset D of the upper (time–frequency) half plane. D will be bounded in the applications, but we can think of D as the whole upper halfplane for the purpose of the present discussion. We shall use the notation (b, φ) for the points of the domain D. We consider a non-negative function $M(b, \varphi)$ defined on D. We define the ridge set R as the set of local maxima in φ of the functions $\varphi \hookrightarrow M(b, \varphi)$ when the variable b is held fixed. We assume that the surface $M(b, \varphi)$ is smooth enough so that the ridge set is the finite union of the graphs of smooth functions slowly varying on their respective domains. In other words, we assume that

$$R = \cup_{\ell=1}^{L} R_\ell \tag{8}$$

where each R_ℓ is the graph of a smooth function $[b_{\ell,\min}, b_{\ell,\max}] \ni b \hookrightarrow \varphi_\ell(b)$ defined on a subset of the domain of the variable b. In the practical applications we have in mind, the ridge functions $\varphi_\ell(b)$ are slowly varying. Notice that we do not make any assumption on the lengths of the individual ridges R_ℓ or even the fact that they could cross.[1]

A. The Crazy Climber Detection Algorithm

The main idea of the "crazy climber" algorithm is as follows. A large number of particles (the climbers) are initially randomly seeded on the domain D at step 0. Then, each climber evolves according to a Markov chain on D with a transition mechanism depending on the local values of the $M(b, \varphi)$ function. This chain is designed to relax to a steady distribution that is essentially concentrated on the ridges. The projection of the motion on the b axis is the standard symmetric random walk, say, with elastic reflection at the boundary points of the interval, and consequently, the projection of the steady distribution on the b-axis is the uniform distribution. Vertically, the climbers are encouraged to "climb on the hills" to reach the ridges by a Hastings–Metropolis penalization and a temperature schedule similar to the simulated annealing algorithm. However, contrary to simulated annealing, the crazy climber algorithm looks for all the local maxima when the time variable is fixed instead of searching only for the global maxima. In fact, a natural implementation of the simulated annealing would lead to the simulation of one single sample path of a Markov chain in the very complex space of all the ridge candidates (see, for example, [5] for an implementation in this spirit), whereas the crazy climber detection algorithm requires the simulation of many sample paths of a Markov chain in the time–frequency plane (or, more precisely, its discretized version), which is a state space of a much lower complexity indeed.

Obviously, the implementation involves a discretized version of the time–frequency plane. We assume that the time interval over which the signal is analyzed is discretized into a finite set $\{b_0, b_1, \cdots, b_{B-1}\}$ with B elements. We also assume that the values of the frequency variable φ are discretized into a finite set $\{\varphi_0, \varphi_1, \cdots, \varphi_{K-1}\}$. We thus reduce the analysis of the modulus of the transform to the analysis of a finite $B \times K$ matrix with nonnegative entries.

1) Crazy Climbers: At time $t = 0$, we initialize the positions $X_\alpha(0)$ of N climbers on the grid $\Gamma = \{0, \cdots, B-1\} \times \{0, \cdots, K-1\}$. The climbers are labeled by the parameter $\alpha = 1, \cdots, N$. The initial positions are chosen independently of each other uniformly over the grid Γ. The climbers evolve independently of each other according to the same law. If a climber is at the point (j, k) at time t, i.e., if $X_\alpha(t) = (j, k)$, then its position at time $t + 1$, say, $X_\alpha(t + 1) = (j', k')$, is determined according to the following law: $j' = j - 1$ with probability 1/2 and $j' = j + 1$ with probability 1/2 (we do not discuss the particular cases $j = 0$ and $j = B - 1$ involving boundary conditions so that we not to confuse the issue). Then, when the climber has decided to move to the left (when $j' = j - 1$) or to the right (when $j' = j + 1$) in the horizontal direction, a possible vertical move is considered. As for the horizontal component, the climber tries to move up, i.e., $k' = k+1$, or down, i.e., $k' = k-1$, with equal probabilities. Again, we ignore the boundary conditions for the sake of simplicity. Unlike in the case of the horizontal direction, the move does not always take place. The transition from (j', k) to (j', k') takes place if the value of the function increases, i.e., if its so-called Delta $\Delta M = M(j', k') - M(j', k)$ is nonnegative. On the other hand, the move does not necessarily take place if the function decreases, i.e., if $\Delta M < 0$. Indeed, in this case, the transition is made, i.e., $X_\alpha(t + 1) = (j', k')$ with probability $\exp[\Delta M/T(t)]$, and the climber does not move vertically, i.e., $X_\alpha(t + 1) = (j', k)$ with probability $1 - \exp[\Delta M/T(t)]$.

[1] In fact, in the cases under consideration, namely, wavelet and Gabor transforms, ridges never cross; even if we generate a signal whose analytical expression is the sum of two frequency-modulated components with intersecting frequency-modulation curves, the ridges will not reproduce exactly the frequency modulations near their expected intersection. We have no simple explanation for this "experimental" fact. Notice that such a remark is specific to wavelet or Gabor transforms. Indeed, ridges of Wigner–Ville representations do cross.

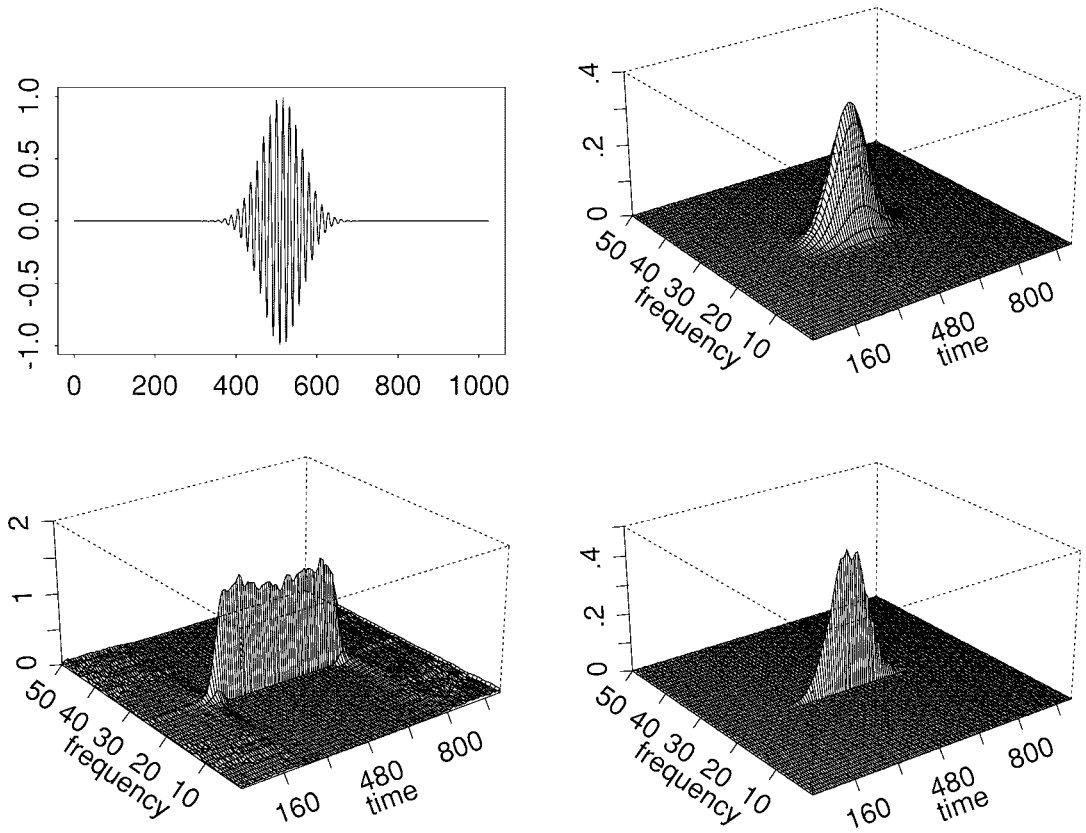


Fig. 1. Occupation measures for a simple windowed sine wave. Top left: The signal. Top right: Its Gabor transform modulus. Bottom left: Unweighted occupation measure. Bottom right: Weighted occupation measure.

At each time t, we consider two occupation measures. The first one is defined by

$$\mu_t^{(0)} = \frac{1}{N} \sum_{\alpha=1}^{N} \delta_{X_\alpha(t)}.$$

It is obtained by putting a mass $1/N$ at the location of each of the climbers on the grid. In other words, $\mu_t^{(0)}(A)$ is the proportion of climbers in A at time t. The second one is the "weighted" occupation measure μ_t obtained by putting a mass equal to the value of the function M at the current location of the climber

$$\mu_t = \sum_{\alpha=1}^{N} M(X_\alpha(t))\, \delta_{X_\alpha(t)}.$$

We finally consider the corresponding "integrated occupation measures," which are defined by ergodic averages as

$$\mu_I^0 = \frac{1}{T} \sum_{t=1}^{T} \mu_t^0 \quad \text{and} \quad \mu_I = \frac{1}{T} \sum_{t=1}^{T} \mu_t. \tag{9}$$

The occupation measure μ_I^0 is only given here for the sake of completeness. Indeed, its main shortcoming is the fact that it assigns nonzero mass to regions without ridges if the lengths of the ridges are smaller than the length of the window. This is due to the very nature of the unrestricted horizontal motions of the climbers. Because the modulus of the *denoised* versions of the functions M, which we use in the applications, are essentially zero away from the ridges, the occupation measure μ_I gives much better results when it comes to detecting ridges.

Further Remark: The climbers evolve independently of each other without interaction and with the same distribution so that the computer code generating the motions of the climbers is the same for all the climbers. This indicates that the algorithm can naturally be parallelized on a SIMDIM machine (such as, for example, the massively parallel computers MASPAR I and II.) We will not report of such an implementations here.

2) Simple Examples of Occupation Measures: To illustrate the crazy climber algorithm, we first present, in Fig. 1, a simple example, namely, that of the Gabor transform of a sine wave multiplied by a Gaussian envelope (top left). The square modulus of the Gabor transform with a Gaussian window is displayed at the top right of the figure, and the two integrated occupation measures are displayed in the bottom of the figure. We clearly see in the figure the different meanings of the two measures; in particular, the weighted occupation measure appears as a shrunk copy of the Gabor square modulus. In addition, the time dependence is essentially the same as that of the square modulus. This suggests that thresholding the weighted measure is likely to yield good results for ridge detection. We will not elaborate on that here.

3) Chaining: The output of the algorithm described above is a measure on the domain D. We identify it with its density, which is a function on D. The next step of the algorithm is

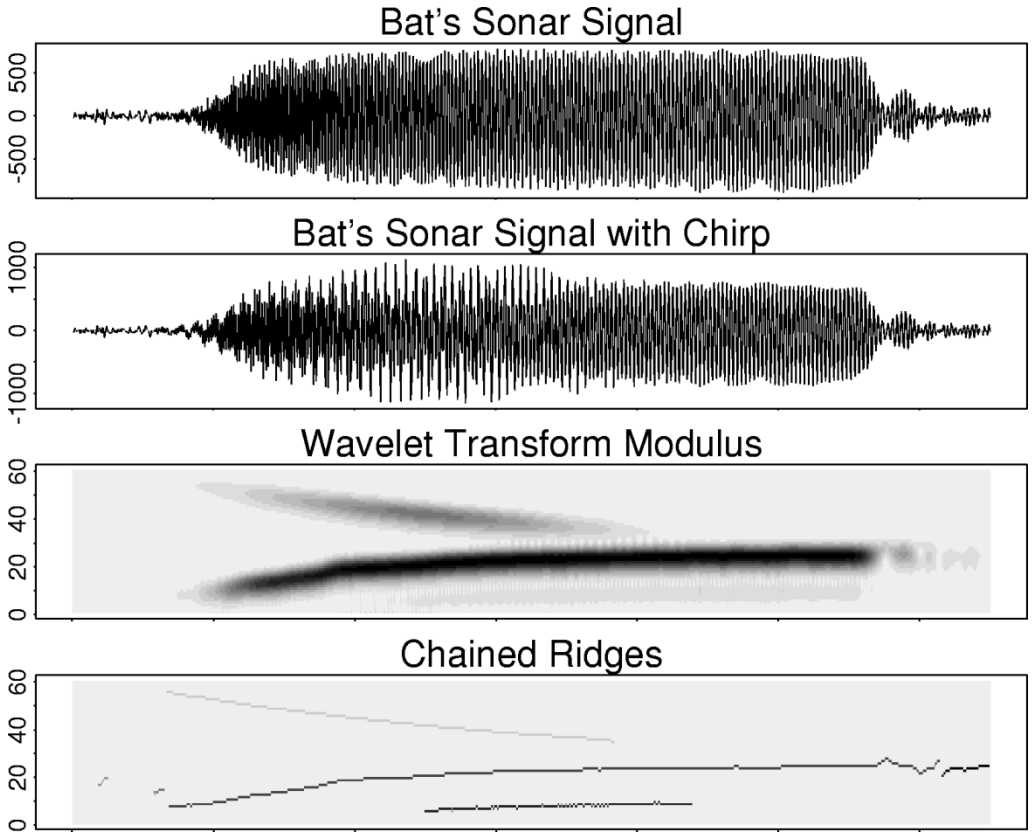


Fig. 2. Bat sonar signal with an additional chirp.

to associate with it the various ridges R_ℓ. This is done via a chaining procedure that replaces the occupation density by one-dimensional (1-D) curves. This procedure is based on the two steps:

1) Thresholding of the density function obtained as the output of the crazy climbers algorithm;
2) "propagation" in the b direction: Given a point (b, φ) belonging to a given ridge R_ℓ, look for the "best neighbor" among $(b+\epsilon_b, \varphi)$ and $(b+\epsilon_b, \varphi \pm \epsilon_\varphi)$ (here, ϵ_b and ϵ_φ are parameters fixed in advance), and then, iterate the process until only values below the threshold can be reached.

The result is a series of ridges, which are graphs of curves $\varphi = \varphi_\ell(b), b = b_0^\ell \cdots b_{k_\ell}^\ell$.

IV. Numerical Results and Examples

The crazy climber algorithm was tested on several signals containing multiple ridges. We restrict the present discussion to two examples: one treated with the wavelet transform and the other with the Gabor transform.

A. First Numerical Results

The first signal (which is displayed at the top of Fig. 2) is the sum of a (real) sonar signal emitted by a bat (this particular signal, which is displayed at the very top of Fig. 2, together with noisy versions, was intensively studied in [5]) and a "linear chirp," i.e., a function of the form $A(x)\cos(\phi(x))$ with $A(x)$ a Gaussian function and $\phi(x)$ a quadratic phase. We used the Morlet wavelet $\psi(x) = e^{-x^2/2}e^{i\omega_0 x}$. Even though it is not strictly admissible, $\hat{\psi}(0)$ is small enough for the wavelet to be considered to be "numerically admissible." The continuous wavelet transform was computed for 60 different values of the scale, more precisely, for scales of the form $2 \times 2^{n/15}, n = 0, \cdots 59$. The time variable was sampled at the same rate as the original signal. Fig. 2 shows the modulus of the wavelet transform of the signal, together with the three different ridges found by the crazy climber method: the main ridge of the bat signal, the first harmonic component, and the chirp signal. The horizontal axis is the time axis, and the vertical axis corresponds to the logarithm of the scale, i.e., the variable φ alluded to above. The modulus is represented with gray levels proportional to the values of the modulus of the wavelet transform. The different ridges are displayed with different gray levels.

The second signal is a speech signal, namely 250 ms of the word /one/ sampled at 8 kHz. The signal is displayed at the very top of Fig. 5, and the squared modulus of its Gabor transform is the third item of the figure. We used a Gaussian window $g(x) = g_s(x) = (1/s\sqrt{2\pi})e^{-x^2/2s^2}$, where s is a scale parameter, but other choices, such as the Hamming windows (which are very popular in speech processing), would be as convenient. The scale parameter s was set so that the window has a size approximately equal to 16 ms (following the lines of [12]). The Gabor transform was computed over the range 0–4000 Hz with 100 different values for the regularly sampled frequency; see the discussion above. The horizontal axis is the time axis, and the vertical axis is the frequency axis

(the convention for the square modulus and ridge displays are the same as before). The crazy climbers algorithm (500 climbers, 10 000 time steps for each) found 18 different ridges, which are displayed at the bottom of Fig. 5.

We shall come back to these examples when discussing the reconstruction from ridges, which are displayed on the same figures.

B. The Case of Noisy Signals

In many applications, we can assume that we have observations $f(x)$ of an unknown signal $f_0(x)$ in the presence of an additive noise $\epsilon(x)$ with mean zero. In other words, we work with the model $f(x) = f_0(x) + \epsilon(x)$, where the noise is given by a mean zero stationary process with (unknown) auto-covariance function $\mathbb{E}\{\epsilon(x)\epsilon(y)\} = \gamma(x-y)$. The case $\gamma(x-y) = \sigma^2\delta(x-y)$ corresponds to an additive white noise with variance σ^2. In some situations, "*a priori*" knowledge of the noise is available. For instance, it may happen that the power spectrum of the noise is known, or a piece of the signal is known to contain only noise, which gives us the chance to learn about the statistics of this noise. Then the detection algorithm may be improved by "renormalizing" the time–frequency representation, i.e., subtracting what is supposed to be the "typical" contribution of the noise. This contribution could be chosen to be the expectation $\mathbb{E}\{M_\epsilon(b,\varphi)\}$. Moreover, if an *a priori* model for the noise is available, such a quantity may be estimated by Monte Carlo simulations or sometimes by a direct computation.

Example: Assume that $\{\epsilon(x)\}$ is a second-order mean zero stationary process with power spectrum of the form $p(\xi) \sim \sigma^2\xi^\alpha$ and that we are using the continuous wavelet transform. In this case, $\mathbb{E}\{M_\epsilon(b,a)\} = \mathbb{E}\{|T_\epsilon(b,a)|^2\} \sim K_\alpha\sigma^2 a^{-\alpha-1}$ provided the analyzing wavelet $\psi(x)$ is such that $K_\alpha = \int u^\alpha|\hat{\psi}(u)|^2\, du < \infty$. In most practical applications, we only have one realization of the noise component, and it is impossible to compute directly this expectation, but a simple ergodic argument justifies the use of the estimate

$$V(\varphi) = \frac{1}{B}\int_0^B M(b,\varphi)\, db. \tag{10}$$

Then, in the penalty term used to define the φ motion of the climbers, the squared modulus of the time–frequency transform may be replaced by

$$\tilde{M}(b,\varphi) = M(b,\varphi) - V(\varphi). \tag{11}$$

Remark: In situations where a reasonable estimate of the power spectrum $p(\xi)$ is available, a *prewhitened transform* may be preferred to the usual transform. For example, in the case of the continuous wavelet transform, the following transform was used in [11]:

$$W_f(b,a) = \langle\Gamma^{-1/2}f, \psi_{(b,a)}\rangle = \frac{1}{2\pi}\int \hat{f}(\xi)\overline{\widehat{\psi_{(b,a)}}}(\xi)\,\frac{d\xi}{\sqrt{p(\xi)}} \tag{12}$$

where Γ denotes the covariance operator of the noise.

The modifications (11) and (12) avoid "trapping" the ridge in regions dominated by the noise. Notice that they can always be used, as long as the noise may be modeled as a weakly stationary process.

As an illustration, we display in Fig. 4 the ridges of the same signal as before, embedded into a Gaussian white noise with SNR $= 1$ dB. We can see that the main ridge is quite well reconstructed and that the ridge of the chirp is also recovered, although only its most energetic part has been detected. The first harmonic component of the bat signal has not been detected. (The corresponding wavelet transform square modulus was too low compared with the typical size of the noise, which, here, is of the form K/a for some constant K.)

C. Robustness

In the analysis of noisy signals, one of the main difficulties is the dependence of many algorithms on the realization of the noise perturbation. It happens too often that the parameters and the results of the analysis are too sensitive to the noise, limiting the realm of applicability. The crazy climber algorithm (together with the associated chaining) seems to be very stable when facing many noise realizations. In order to illustrate its robustness, we ran the following experiment with another speech signal borrowed from Maes [14]. We added a Gaussian white noise with variance σ^2 to the signal /How are you?/ of [14], we computed the Gabor transform of the noisy signal so obtained, and we ran the crazy climber algorithm to detect the ridges in the modulus of the transform. We then measured the performance of the detection algorithm by computing the following measure of discrepancy between the set R_0 of the ridges detected in the original Gabor transform and the set R_1 of ridges found in the modulus of the transform of the noisy signal

$$D = \frac{\#(R_1\backslash(R_1\cap R_2)) + \#(R_2\backslash(R_2\cap R_1))}{\#R_1 + \#R_2}. \tag{13}$$

This distance D is always between 0 and 1, it is equal to 0 when the two sets R_0 and R_1 are identical, and it is equal to 1 when the two set have no point in common. We calibrated the values of the variance σ^2 of the white noise so that the corresponding SNR would be equal to 20, 10, 5, 0, and -5 dB, respectively. For each value of the SNR from this list, we generated 32 samples of the white noise. For each of them, we computed the Gabor transform, determined the ridge set R_1 with the crazy climber algorithm, and computed the corresponding value of the distance D, as shown in Table I. These numerical results show that only 3% of the ridge points are misdetected at SNR $= 20$ dB and 25% at SNR $= 0$ dB. Moreover the 95% confidence intervals are relatively short. The situation deteriorates at low SNR's because the definition of D penalizes heavily all the mismatches, even if a ridge point is detected but misplaced by a very small amount.

V. Reconstructions From the Ridges

We now address the problem of the reconstruction of a signal from the knowledge of a transform on its ridge(s). This problem was already discussed in [8] as well as many articles on speech processing (see, e.g., [12]). We describe here an approach based on a penalization procedure, which was alluded

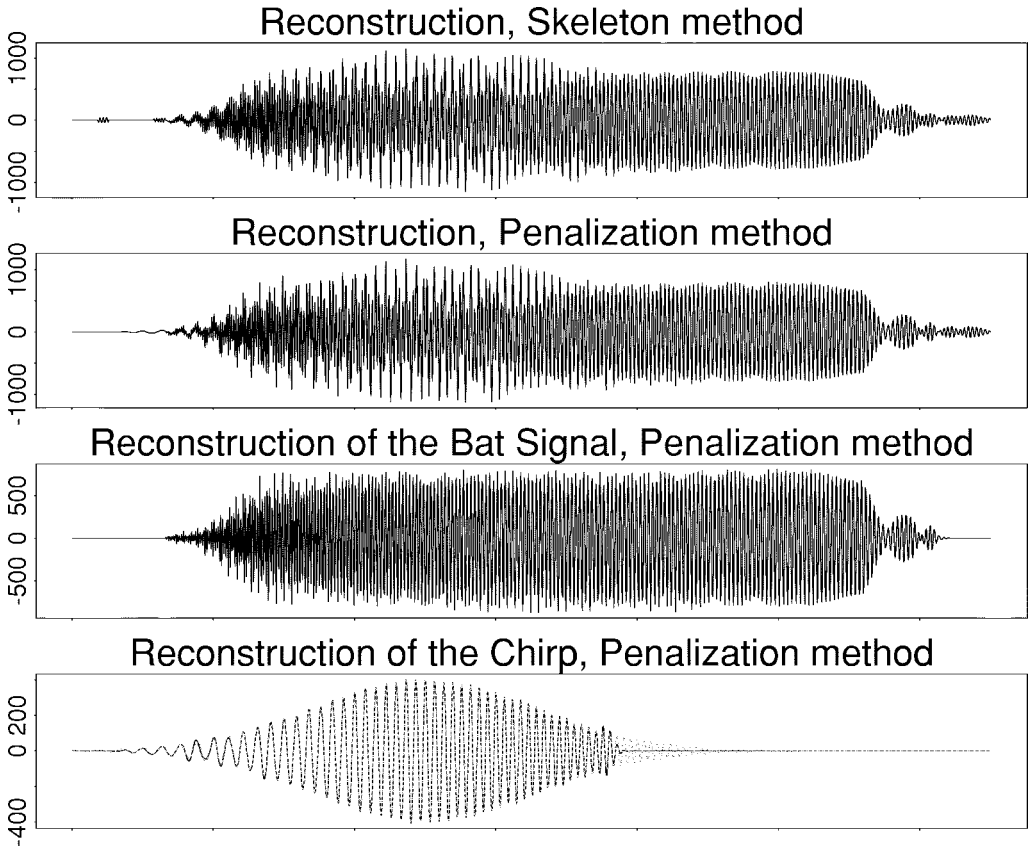


Fig. 3. Reconstruction from the ridges. Last plot: Full curve, reconstructed chirp. Dashed curve: The original chirp.

TABLE I

dB	average D	95% confidence interval
20	0.0313	[0.0229 0.0423]
10	0.0632	[0.0437 0.0841]
5	0.1752	[0.1427 0.2203]
0	0.2670	[0.2047 0.3315]

to in [5]. Even though the ridge-detection part of the algorithm was independent of the time–frequency representation, this is not the case for the reconstruction. In particular, our approach is not adapted to bilinear time–frequency representations such as those given by the Wigner–Ville transform. We shall restrict ourselves to the cases of the wavelet and the Gabor transforms. Nevertheless, our approach extends to linear representations such as those obtained from matching pursuit as discussed in [15] or those developed in [19].

A. General Discussion

In order to put our reconstruction algorithm in perspective, we first review the procedures currently used (a large number of illustrations may be found in [6]). The methods outlined in Section V-A3 will be developed with more detail in Section V-B.

1) The Transform Skeleton: The first reconstruction is the simplest one (once the ridges have been estimated). It consists of restricting the transform (whether we are working with the wavelet transform of the Gabor transform) to the ridges. It is motivated by the approximate formulas (4) and (7). More precisely, using the notation used throughout the paper, this reconstruction is given in the Gabor case by

$$\hat{f}(x) = 2\Re \sum_{\ell=1}^{L} G_f(x, \omega_\ell(x)) \tag{14}$$

where the summation in the right-hand side is restricted to the ℓ's for which $\omega_\ell(x)$ makes sense (in particular, $\hat{f}(x) = 0$ when there is no ridge at "time" x) and by a similar formula in the wavelet case. The restriction of a transform to a ridge is sometimes called the "skeleton" of the transform [8].

This is a very simple scheme, and as Fig. 3 shows, the results of this naive reconstruction can be extremely good. Its main shortcoming is that it requires the knowledge of the transform at *all* the points of the ridges. This limitation makes it impossible to subsample the ridge (for compression purposes, for example.)

2) Parametric Reconstruction: For the sake of completeness, we quote a nonlinear reconstruction scheme that has been successfully used in speech processing in the framework of the so-called sinusoidal model. See [12] for a review. The main observation is that although the wavelet and the Gabor transforms restricted to ridges are often very oscillatory (and then uneasy to compress), the corresponding amplitudes $A_\ell(b)$ and frequencies $\nu_\ell(b) = (1/2\pi)\phi'_\ell(b)$ are sometimes slowly varying. This is true, for instance, for the Gabor transform of speech signals when the window is broadband.

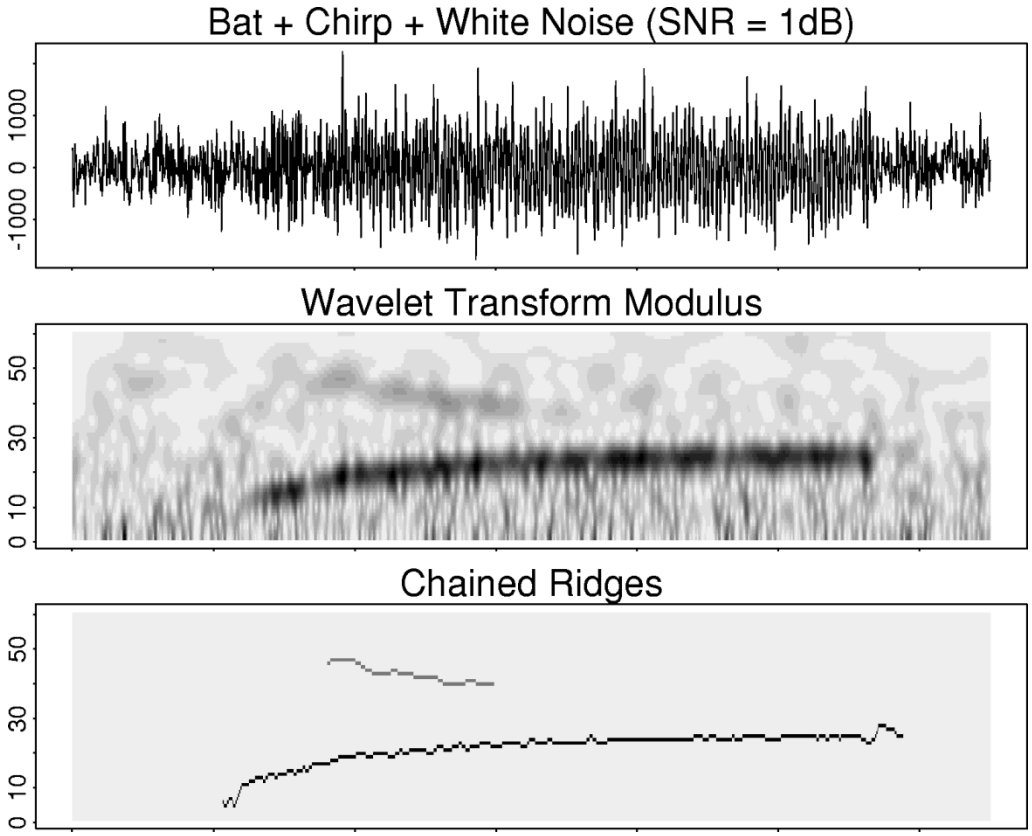


Fig. 4. Bat's sonar signal and chirp embedded into a Gaussian white noise (SNR = 1 dB).

Whenever amplitudes and frequencies are slowly varying, they may be very well approximated by a class of functions characterized by a small number of numerical parameters. The reconstructed signal is then obtained via parametric estimation of the amplitude and the frequency and then by integrating the estimate of the frequency to recover the phase

$$\phi_\ell(b) = \phi_\ell(b_0) + \frac{1}{2\pi} \int_{b_0}^{b} \nu_\ell(x)\, dx \tag{15}$$

and computing $A_\ell(b)\ \cos(\phi_\ell(b))$. This may be done for all the ridges of a given signal. The reconstructed signal is the sum of all the components of the form $A_\ell(b)\ \cos(\phi_\ell(b))$.

3) Ridge Penalization: To reduce the amount of data necessary for the reconstruction, it is tempting to start by selecting a few sample points from each ridge and to use such points and the corresponding values of the representation to reconstruct a signal. The reconstructed signal takes the form of a superposition of elementary waveforms located at the sample points of the ridge. Such waveforms are, in general, mere perturbations of wavelets or Gabor functions.

As we shall see, such solutions may be derived from general principles, namely, by minimizing a suitably chosen quadratic functional and using the values of the time–frequency representation at the sample points of the ridge as linear constraints.

B. The Penalization Approach

We now focus on the penalization approach, and we treat the Gabor case and the wavelet case in the same setting. We present a reconstruction algorithm that produces, from the mere knowledge of a time–frequency transform at sample points of the ridges, a very good approximation of the original signal (this reconstruction procedure was implemented and tested in [5] in the case of the wavelet transform of a signal in the "one-ridge case.")

We assume that the ridges can be parameterized by continuous functions $[b_{\ell,\min}, b_{\ell,\max}] \ni b \hookrightarrow \varphi_\ell(b) \in (0,\infty)$, where $\ell \in \{1,\cdots,L\}$ is the ridge label. These ridges are usually constructed as smooth functions resulting from fitting procedures (spline smoothing is an example we are using in practical applications) from the sample points obtained from ridge estimation algorithms such as the crazy climbers algorithm presented in this paper or the snake annealing described in [5].

1) Statement of the Problem: We consider a (linear) time–frequency transform of an unknown signal of finite energy $f_0(x)$, which we denote generically by $\mathcal{T}_f(b,\varphi) = \langle f_0, e_{(b,\varphi)}\rangle$, where the time–frequency atoms $e_{(b,\varphi)}$ are either wavelets or Gabor functions. We assume that the values of $\mathcal{T}_f(b,\varphi)$ are known at sample points $(b_{\ell,j},\varphi_{\ell,j})$ with $j = 1,\cdots n_\ell, \ell = 1,\cdots L$, which are regarded as representative of the ridges associated with the (unknown) signal $f_0(x)$. We use the notation $z_{\ell,j}$ for the value of the transform of f_0 at the point $(b_{\ell,j},\varphi_{\ell,j})$ and $e_{\ell,j}(x)$ for the corresponding function $e_{(b_{\ell,j},\varphi_{\ell,j})}(x)$. The set of sample points $(b_{\ell,j},\varphi_{\ell,j})$, together with the values $z_{\ell,j}$, constitute what we call the *skeleton of the transform* of the signal to be reconstructed. Let us denote by $L_{\ell,j}$ the linear form defined by $L_{\ell,j} f = \langle f, e_{\ell,j}\rangle$. As we already explained, we

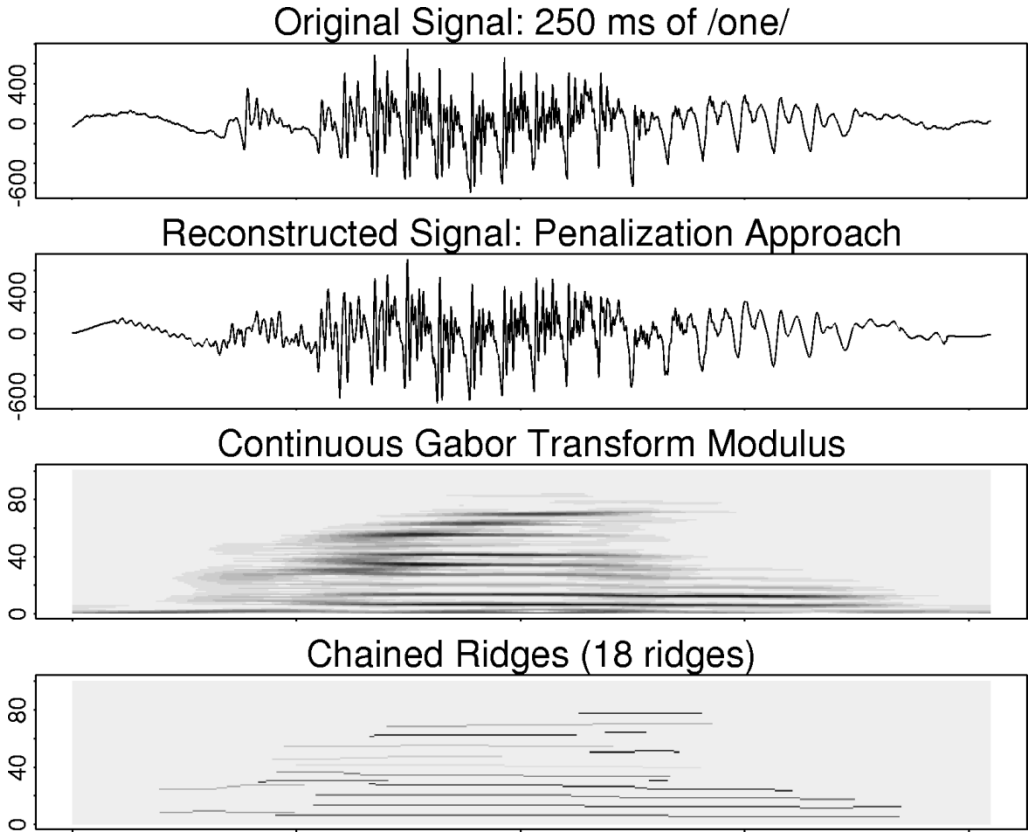


Fig. 5. Two hundred fifty milliseconds of the speech signal /one/ (sampling frequency: 8 kHz).

use smooth functions $b \hookrightarrow \varphi_\ell(b)$, which fit the sample points, and we use the graphs of these functions as our best guesses for the ridges of the modulus of the transform of f_0. The reconstruction problem is to find a signal $f(x)$ of finite energy whose transform $\mathcal{T}_f(b,\varphi)$ satisfies

$$L_{\ell,j} f = z_{\ell,j}, \qquad \ell = 1, \cdots L, \quad j = 1, \cdots, n_\ell \tag{16}$$

and has the union R of the graphs of the functions $\varphi_\ell(b)$ as a set of ridges. We recall that this means that for each b, the points $(b, \varphi_\ell(b))$ of the time–frequency plane are the local maxima of the function $\varphi \hookrightarrow |\mathcal{T}_f(b,\varphi)|^2$.

2) The Penalty Functions: We start with the case of the wavelet transform. Let us assume that we are given a series of L ridges of the form $b \hookrightarrow a_\ell(b), \ell = 1, \cdots L$. As explained in the Introduction, the reconstructed signal is obtained as the function $\tilde{f}(x)$ minimizing a quadratic functional $F[f] = \langle f, Qf\rangle$ with the constraints $T_{\tilde{f}}(b_{\ell,j}, a_{\ell,j}) = z_{\ell,j}$. In the case of the wavelet transform, we use Q given by the integral operator defined by the kernel

$$\mathcal{W}(x,y) = \delta(x-y) + \epsilon \mathcal{W}_2(x,y).$$

The aim of the term $\mathcal{W}_2(x,y)$ is to enforce smoothness of $|T_{\tilde{f}}|$ on the ridges, i.e., of the functions $b \to |T_{\tilde{f}}(b, a_\ell(b))|$. It is given by

$$\begin{aligned}\mathcal{W}_2(x,y) = \sum_\ell \int \frac{db}{a_\ell(b)^4} \Bigg(&\overline{\psi}\left(\frac{x-b}{a_\ell(b)}\right)\psi\left(\frac{y-b}{a_\ell(b)}\right) \\ &\cdot [a'_\ell(b)^2 - \omega_0^2] + \overline{\psi}'\left(\frac{x-b}{a_\ell(b)}\right)\psi'\left(\frac{y-b}{a_\ell(b)}\right) \\ &\cdot \left[\frac{(x-b)(y-b)}{a_\ell(b)^2} + 1 + \frac{x-2b+y}{a_\ell(b)}\right] \\ &+ \overline{\psi}\left(\frac{x-b}{a_\ell(b)}\right)\psi'\left(\frac{y-b}{a_\ell(b)}\right) a'_\ell(b)\left[1 + \frac{y-b}{a_\ell(b)}\right] \\ &+ \overline{\psi}'\left(\frac{x-b}{a_\ell(b)}\right)\psi\left(\frac{y-b}{a_\ell(b)}\right) a'_\ell(b)\left[1 + \frac{x-b}{a_\ell(b)}\right]\Bigg).\end{aligned} \tag{17}$$

In the case of the Gabor transform, the quadratic form Q is given by the integral kernel

$$\mathcal{G}(x,y) = \delta(x-y) + \epsilon \mathcal{G}_2(x,y) \tag{18}$$

where $\mathcal{G}_2(x,y)$ enforces smoothness of $|G_{\tilde{f}}|$ on the ridges, i.e., of the functions $b \to |G_f(b, \omega_\ell(b))|$. It is given by

$$\begin{aligned}\mathcal{G}_2(x,y) &= \sum_\ell \Bigg(\int \Bigg(g'(x-b)g'(y-b) + g(x-b)g(y-b) \\ &\quad \cdot \left[(x-b)(y-b)\omega'_\ell(b)^2 \right. \\ &\quad \left. - (x+y-2b)\omega_\ell(b)\omega''_\ell(b)\right] \Bigg) \\ &\quad \cdot \cos(\omega_\ell(b)(x-y))\, db \\ &\quad + \int \Bigg(g'(x-b)g(y-b)\left[\omega_\ell(b) - (y-b)\omega'_\ell(b)\right] \\ &\quad - g(x-b)g'(y-b)\left[\omega_\ell(b) - (x-b)\omega'_\ell(b)\right] \Bigg) \\ &\quad \cdot \sin(\omega_\ell(b)(x-y)) \Bigg)\, db.\end{aligned} \tag{19}$$

For the sake of simplicity, we have assumed that the window $g(x)$ is real valued.

C. Solution of the Optimization Problem

The reconstruction problem has been reformulated as a problem of minimization of a sesquilinear functional with linear constraints. This is a classical problem. We outline the solution for the sake of completeness. It may be conveniently reformulated as a minimization problem in the real domain rather than the complex domain by noticing that since we restrict ourselves to real-valued signals, both kernels are sesquilinear ($\mathcal{G}(x,y)$ is even real valued) and may be be replaced by their real part (which we shall still denote by the same letter). For each $\ell = 1, \cdots L$, the n_ℓ complex constraints (15) may be replaced by the $2n_\ell$ real constraints as follows. Set

$$\begin{cases} \rho_{\ell,j}(x) = \Re e_{\ell,j}(x), & j = 1, \cdots, n_\ell \\ \rho_{\ell,j}(x) = \Im e_{\ell,j}(x), & j = n+1, \cdots, 2n_\ell \end{cases} \quad (20)$$

$r_j = \Re z_j$, and $r_{n+j} = \Im z_j$. (Here, $\Re$ and $\Im$ stand for the real and imaginary parts, respectively.) Then, the new constraints read

$$\mathcal{R}_j(f) = \langle f, \rho_{\ell,j} \rangle = r_j, \qquad j = 1, \cdots, 2n. \quad (21)$$

Consequently, there exist real numbers $\lambda_{\ell,j}, j = 1, \cdots 2n_\ell, \ell = 1, \cdots L$ (the Lagrange multipliers of the problem) such that the solution $\tilde{f}(x)$ of the optimization problem is given by

$$\tilde{f}(x) = \sum_{\ell=1}^{L} \sum_{j=1}^{2n_\ell} \lambda_{\ell,j}\, \tilde{\rho}_{\ell,j}(x) \quad (22)$$

where the functions $\tilde{\rho}_{\ell,j}(x)$ are defined by $\tilde{\rho}_{\ell,j} = Q^{-1}\rho_{\ell,j}$. The Lagrange multipliers are determined by requiring that the constraints (21) be satisfied. In other words, we must demand that the wavelet transform of the function $\tilde{f}$ given in (22) be equal to the z_j's at the sample points (b_j, φ_j) of the time-scale plane. This gives a system of linear equations from which the Lagrange multipliers λ_j's can be computed. More precisely, if we denote by Λ the column vector consisting of the Lagrange multipliers, and by R the vector of values $r_{\ell,j}$, we obtain $\Lambda = \mathcal{M}^{-1}R$ and $\mathcal{M}_{\ell,j;\ell',j'} = \langle Q^{-1}\rho_{\ell,j}, \rho_{\ell',j'} \rangle$.

Remark: If we denote by $N = \Sigma_\ell\, n_\ell$ the total number of ridge samples, we have to deal with $N \times N$ matrices, i.e., very large matrices. However, such matrices are, in general, sparse because of the localization properties of the functions. In particular, if the different ridges of the transform are "well separated," i.e., if they are located in different regions of the time–frequency domain, the matrix $\mathcal{M}$ may be approximated by a block diagonal matrix. In other words, the contributions of all ridges may be reconstructed independently. This reduces the complexity from N^3 to Ln^3, where n is an average number of sample points per ridge. The Lagrange multipliers are then obtained as follows. If we denote by Λ_ℓ the column vector consisting of the Lagrange multipliers for fixed ℓ and by R_ℓ the vector of values $r_{\ell,j}$ for fixed ℓ, we obtain the expression for the Lagrange multipliers $\lambda_{\ell,j}$ with fixed $\ell, \Lambda_\ell = \mathcal{M}_\ell^{-1}R_\ell$ with a new $n_\ell \times n_\ell$ matrix $\mathcal{M}_\ell$, whose elements are given by $(\mathcal{M}_\ell)_{j,j'} = \langle Q^{-1}\rho_{\ell,j}, \rho_{\ell,j'} \rangle$.

D. The Reconstruction Algorithm

The results of the discussion of this section may be summarized in an algorithmic walk through our solution to the reconstruction problem.

- Determine a finite set $\{R_\ell\}_{\ell=1,\cdots,L}$ of ridges and, on each of them, a set of sample points $(b_1, \varphi_{\ell,1}) = \varphi_\ell(b_1), \cdots, (b_n, \varphi_{\ell,n(\ell)} = \varphi_\ell(b_n(\ell))$ on the ridge.
- Construct smooth estimates $b \hookrightarrow \varphi_\ell(b)$ of the ridges from the sample points.
- Compute the matrix, i.e., $\mathcal{W}(x,y)$ or $\mathcal{G}(x,y)$, of the smoothness penalty along the ridge estimate.
- Compute the reconstruction time–frequency atoms $\tilde{e}_{\ell,j} = Q^{-1}e_{(b_{\ell,j},\varphi_{\ell,j})}$ localized (in the time–frequency plane) at the ridge sample points.
- Compute the coefficients $\lambda_{\ell,j}$.

The solution $\tilde{f}$ of the reconstruction problem is then given by (22). Numerical examples are discussed below.

E. The Case of Noisy Signals: Smoothing Spline-Type Reconstruction

The synthesis presented above requires the knowledge of the exact values of the transform at a finite sample of points of the time-scale or time–frequency plane. We now consider the possibility of an additive (possibly colored) noise in the observations of the input signal and the possibility of noise in the computation of the transform of the signal. As before, our approach is motivated by the smoothing splines technique, as presented in [20]. The generalization presented in this paper was alluded to as a possible extension to the reconstruction algorithm derived and used in [4] and [5]. The motivation of [4] was to simplify the algorithm given in [16] to reconstruct a signal from the extrema of its dyadic wavelet transform. The motivation of [5] was to generalize this approach to the case of the continuous wavelet transform, where the role of the extrema of the dyadic wavelet transform was played by the ridges of the continuous wavelet transform.

The reconstruction we present now is based on a variational approach involving a penalty on the smoothness of the transform along the estimated ridges. However, contrary to [5], the observations of the transform along the ridges are used to define a second penalty component. This form of the variational problem allows for a delicate balance between the fit of the transform of the solution to the observations and the smoothness of the modulus of the transform along the ridges. As before, we use the notation $\mathcal{T}_f(b,\varphi)$ for a transform that could be the wavelet transform $T_f(b,a)$ as well as for the Gabor transform $G_f(b,\omega)$.

We assume that we are dealing with the signal+noise model introduced earlier in Section IV-B; after computation of the transform of the observations, estimation of the ridges of the transform, and sampling these estimates, we end up with a discrete set $\{(b_{\ell,j}, \varphi_{\ell,j}), \ell = 1, \cdots L, j = 1, \cdots n_\ell\}$, in the transform plane and observations $\mathcal{T}_f(b_{\ell,j}, \varphi_{\ell,j})$ of the transform of the unknown signal at these points. We assume that the observations follow the usual linear model

$$z_{\ell,j} = \mathcal{T}_f(b_{\ell,j}, \varphi_{\ell,j}) + \epsilon'_{\ell,j}$$

where the computational noise terms $\epsilon'_{\ell,j}$ are assumed to be identically distributed and uncorrelated between themselves and with the observation noise terms $\epsilon(x)$. Hence, the final model is of the form

$$z_{\ell,j} = L_{\ell,j} f_0 + \epsilon_{\ell,j}, \qquad \ell = 1, \cdots L, \quad j = 1, \cdots, n_\ell \quad (23)$$

where $L_{\ell,j}$ is the linear form representing the value of the transform at the point $(b_{\ell,j}, \varphi_{\ell,j})$ and where

$$\epsilon_{\ell,j} = \mathcal{T}_\epsilon(b_{\ell,j}, \varphi_{\ell,j}) + \epsilon'_{\ell,j}.$$

The assumption that the two sources of noise are uncorrelated implies that the covariance matrix Σ of the $\epsilon_{\ell,j}$ is the sum of the covariance of the $\mathcal{T}_\epsilon(b_{\ell,j}, \varphi_{\ell,j})$ and the covariance of the $\epsilon'_{\ell,j}$. Because the latter is of the form $\sigma'^2 I$, we have $\Sigma = \sigma'^2 I + \Sigma^{(1)}$, where the entries of the matrix $\Sigma^{(1)}$ are given by

$$\Sigma^{(1)}_{\ell,j;\ell',j'k} = \int e_{\ell,j}(x)\, \Gamma(x-y)\, \overline{e}_{\ell',j'}(x)\, dx\, dy.$$

The reconstruction algorithm is formulated as the solution of the minimization problem

$$\min_f \frac{1}{n} \|\Sigma^{-1/2}(Z - \mathcal{T}_f(\cdot,\cdot)\|^2 + \lambda \langle Qf, f\rangle \qquad (24)$$

where Z denotes the vector of observations $z_{\ell,j}$, $\mathcal{T}_f(\cdot,\cdot)$ denotes the vector of values of the transform of the candidate function f at the points $(b_{\ell,j}, \varphi_{\ell,j})$, and the constant $\lambda > 0$ is introduced to balance the effects of the two components of the penalty. In [20], Theorem 1.3.1 implies that the solution is given by

$$\hat{f}_\lambda(x) = \sum_{j=1}^{n} \lambda_{\ell,j}\, \tilde{e}_{\ell,j}(x) = \sum_{j=1}^{n} \lambda_{\ell,j}\, Q^{-1} e_{\ell,j}(x) \qquad (25)$$

where the coefficients $\lambda_{\ell,j}$ are the entries of the matrix $\Lambda = (n\lambda I + \tilde{\Sigma})^{-1}\Sigma^{-1/2}Z$, and the matrix $\tilde{\Sigma}$ is defined by its entries $\tilde{\Sigma}_{j,k} = \langle \tilde{\psi}_j, \psi_k \rangle$.

Remarks:

- Notice that we did not use the full generality of the smoothing spline problem as defined in [20]. Indeed, we could have chosen a quadratic penalty of the form $\|Q^{1/2}P_1 f\|^2$, where P_1 is the projection onto the orthogonal complement of a subspace of finite dimension. In this generality, it is possible to avoid penalizing special subspaces of functions (for example, the space of polynomial functions of degree smaller than a fixed number, ...). Since the form of the solution is much more involved and since we did not find an application justifying this level of generality, we decided to use the smoothing spline approach in our simpler context. Consult [20] to see the specific feature of the smoothing splines technique.
- The approach presented here was alluded to as a possible extension to the reconstruction algorithm derived and used in [4] and [5]. The latter corresponds to the case where the knowledge of the wavelet transform of the unknown signal is assumed to be perfect. In other words, it corresponds to the case where both Γ and σ'^2 are assumed to be zero. It is easy to see that under these extra assumptions, the reconstruction procedure given by the above minimization problem reduces to the minimization of the quadratic form $\langle f, Qf\rangle$ under the constraints (15). This is the problem that was solved in [4] and [5]. It appears as a particular case of the more general procedure presented here. The advantages of the latter were explained in the introduction. We will not reproduce this discussion here.
- Notice that the reconstructed signal appears as a linear function of the observations. Nevertheless, our whole analysis is nonlinear because of the ridge estimation and the sampling of the latter.

F. An Example

Let us return to the wavelet analysis of the bat signal with the additional chirp. We used $L = 3$ ridges, say, R_1, R_2, and R_3, and we chose on each ridge estimate a number of samples proportional to the length of the ridge and inversely proportional to the corresponding scale according to the sampling theory of wavelet transforms; see [7].

We used the value $\epsilon = 0.5$ to reconstruct the signal. The result of the reconstruction is given in the second part of Fig. 3. The last two plots of Fig. 3 give the reconstructions of the two components: the bat signal, reconstructed from two ridges (to be compared with the top plot in Fig. 2) and the chirp (the original chirp and the reconstructed one are displayed on the same plot: the bottom of Fig. 3). As we can see, the agreement is very good (except at the end of the chirp, where the ridge was a bit smaller than the true signal. In addition, we stress that the number of coefficients needed to characterize such a signal, i.e., twice the number of complex constraints, was approximately one fifth of the number of samples. Although compression was not our goal, the method seems to have a definite potential.

VI. Ridges and the Sinusoidal Model for Speech Signal

A popular representation of speech signals is to view the signal as the output of a slowly time-varying filter excited by a glottal waveform. The filter models the resonant characteristics of the vocal tract. We will not go into the details of speech modeling here (see [12] and [14] for a detailed presentation), but we notice that the resulting model for speech signal is of the form given in (3). Hence, it is natural to use a time–frequency representation in order to separate the components of the signal. Since those components are close to harmonic, the Gabor transform is better suited than the wavelet transform for the description of these signals. Indeed, since the wavelet processing may be viewed as a filter bank of constant *relative* frequency, it is not able to separate the components of the high frequencies (nevertheless, see [14] for a method to separate the first low-frequency components).

Hence, we use the continuous Gabor transform with a Gaussian window of length approximately equal to 160 ms. Our approach is, at least in spirit, similar to that of [12]. However, there is a major difference; since the detection algorithm

described in Section III returns ridges, i.e., 1-D structures, the chaining method required by the McAulay–Quatieri approach is not needed here.

We illustrate this discussion on the example of the /one/ signal displayed at the top of Fig. 5. Our results were obtained using approximately 200 ridge samples, i.e., 400 real constraints, whereas the signal's length is 2048. As can be seen on the top two plots of the figure, the reconstructed signal is very close to the true one. Of course, such a comparison is not significant from the speech processing point of view. However, we stress the fact that the main features of the signal are preserved (in particular, the pitch.) Since the L^2-norm (or any other norm for that matter) of the difference between the original and the reconstructed signal cannot be used as a measure of quality as far as speech is concerned, listening to the two sounds remains one of the best ways to evaluate the performance of the synthesis algorithm. We did so in the example discussed above: We could not hear any significant difference. The sounds turn out to be almost indistinguishable.

VII. Conclusions

We presented a new technique to detect ridges on a surface. This algorithm is based on the stochastic relaxation of a particle system of a new type. Our detection technique performs extremely well, especially at very low SNR's. It can be used to detect ridges in all the energetic distribution representations of a signal, and it is especially useful for multicomponent signals. Moreover, we used a Monte Carlo analysis to show that it is extremely robust. We also presented a reconstruction procedure from the knowledge of a linear transform (such as the wavelet or the Gabor transform) on the ridges. In the case of the Gabor transform, we showed that it was performing very well on speech signals, even in the presence of significant noise disturbances.

The most important extension to the results presented in this paper would be a real-time implementation. It is relatively easy to find approximations of the reconstruction procedure, which would be amenable to on-line implementations. It seems more difficult to modify the ridge detection algorithm to accommodate frequent updates.

Appendix
Derivation of the Reconstruction Kernels

We give a derivation of (16) and (17) and their counterparts (18) and (19) in the case of the Gabor transform. In both cases, the first term aims to enforce the localization of the transform near the ridges. This is achieved by minimizing $F_1[f] = ||f||^2$ with the constraints (15). This term alone would yield a solution of the form

$$f(x) = \sum_{\ell,j} \lambda_{\ell,j} \psi_{\ell,j}(x)$$

where the time–frequency atoms $e_{\ell,j}(x) = e_{(b_{\ell,j}, a_{\ell,j})}(x)$ are either the wavelets or the Gabor functions, and where the coefficients $\lambda_{\ell,j}$ are obtained from the $z_{\ell,j}$ by multiplication with the inverse of the matrix $[\langle e_{\ell,j}, e_{\ell',j'} \rangle]_{\ell,j,\ell',j'}$. Numerical tests show that such a solution gives accurate results if the sampling of the ridge is fine enough. Otherwise, an extra term has to be introduced in order to enforce the smoothness of $|T_f|^2$ on the ridges. An adequate candidate for such a term could be given by the H^1-norm of the restriction of the modulus of the transform to the ridges. In the case of the wavelet transform, such a term reads $\Sigma_\ell \int |(d/db)|T_f(b, a_\ell(b))||^2 \, db$, but unfortunately, this does not define a sesquilinear form. However, if we set $\Omega(b,a) = \arg T_f(b,a)$, it follows from the analysis of [8] that near the ridges number ℓ, we have $(d/db)\Omega(b,a) \approx (\omega_0/a_\ell(b))$. This suggests the approximation

$$\frac{d}{db} T_f(b, a_\ell(b)) \approx \frac{d}{db} |T_f(b, a_\ell(b)| e^{i\Omega(b, a_\ell(b)} + i \frac{\omega_0}{a_\ell(b)} |T_f(b, a_\ell(b))| e^{i\Omega(b, a_\ell(b))} \tag{26}$$

and, hence, the use the sesquilinear functional

$$F[f] = \langle f, Qf \rangle = ||f||^2 + \epsilon \sum_\ell \int \left(\left| \frac{d}{db} T_f(b, a_\ell(b)) \right|^2 - \frac{\omega_0^2}{a_\ell(b)^2} |T_f(b, a_\ell(b))|^2 \right) db. \tag{27}$$

An explicit computation shows that the corresponding kernel is as in (16) and (17).

The derivation in the Gabor case goes along the same lines. Instead of minimizing

$$||f||^2 + \epsilon \sum_\ell \int \left| \frac{d}{db} |G_f(b, \omega_\ell(b)| \right|^2 db$$

we minimize the approximate objective function

$$F[f] = \langle f, Qf \rangle = ||f||^2 + \epsilon \sum_\ell \int \left(\left| \frac{d}{db} G_f(b, \omega_\ell(b)) \right|^2 - \omega_\ell(b)^2 |G_f(b, \omega_\ell(b))|^2 \right) db \tag{28}$$

which is sesquilinear. The complete derivation is given in [6].

Acknowledgment

This work was done during visits of the second and third authors to the Department of Mathematics, University of California at Irvine and the C.E.O.R. Department, Princeton University. The hospitality of both institutions is gratefully acknowledged.

References

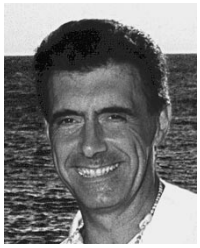
[1] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," Tech. Rep. 93.05, Lab. d'automatique, Ecole Centrale de Nantes, Nantes, France, 1993.

[2] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal," *Proc. IEEE*, part I, vol. 80, pp. 520–538; part II, vol. 80, pp. 540–568, 1992.

[3] B. Boashash and B. Ristic, "Polynomial Wigner-Ville distributions and time-varying higher order spectra," in *IEEE-SP Int. Symp. Time-Scale Time-Freq. Anal.*, Victoria, B.C., Canada, 1992, pp. 31–34.

[4] R. Carmona, "Spline smoothing & extrema representation: Variations on a reconstruction algorithm of Mallat and Zhong," in *Wavelets and*

Statistics, A. Antoniadis and G. Oppenheim, Eds., *Lecture Notes in Statistics*, 1992.
[5] R. Carmona, W. L. Hwang, and B. Torrésani, "Characterization of signals by the ridges of their wavelet transforms," *IEEE Trans. Signal Processing*, vol. 45, pp. 2586–2590, 1997.
[6] ———, "Practical time/frequency analysis: Wavelet and gabor transforms with an implementation in S. Academic Press (in print). Swave: a Public Domain Package of S Functions for Time-Frequency Analysis. Available at http://soil.princeton.edu/ rcarmona/Publications, or anonymous ftp from chelsea.princeton.edu/pub/outgoing/Swave, 1997.
[7] I. Daubechies, *Ten Lectures on Wavelets.* Philadelphia, PA: SIAM, 1992.
[8] N. Delprat *et al.*, "Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies," *IEEE Trans. Inform. Theory*, vol. 38, pp. 644–664, 1992.
[9] P. Flandrin, "Temps-fréquence," in *Traité des Nouvelles Technologies, série Traitement du Signal.* Paris, France: Hermès, 1993.
[10] P. Hall, W. Qian, and D. M. Titterington, "Ridge finding from noisy data," *J. Comput. Graph. Statist.*, vol. 1, pp. 197–211, 1992.
[11] J. M. Innocent and B. Torrésani, "Wavelet transforms and binary coalescences detection," *Appl. Comput. Harmon. Anal.*, vol. 4, pp. 113–116, 1996.
[12] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 744–754, 1986.
[13] ———, "Low rate speech coding based on the sinusoidal model," in *Advances in Speech Signal Processing*, S. Furui and M. Mohan, eds. Tokyo, Japan: Sondui, 1992.
[14] S. Maes, Ph.D. dissertation, Rutgers Univ., New Brunswick, NJ, 1994.
[15] S. Mallat and S. Zhang, "Matching pursuit with time-frequency dictionaries," preprint, 1993.
[16] ———, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 710–732, 1992.
[17] T. F. Quatieri and R. J. McAulay, "Phase coherence in speech reconstruction for enhancement and coding applications," in *Proc. IEEE Int. Conf. Audio, Speech Signal Process.*, Glasgow, U.K., 1989, pp. 207–209.
[18] Ph. Tchamitchian and B. Torrésani, "Ridge and skeleton extraction from the wavelet transform," in *Wavelets and Applications*, M. B. Ruskai *et al.* eds. Boston, MA: Jones & Bartlett, 1992, pp. 123–151.
[19] B. Torrésani, "Time-frequency distributions: Wavelet packets and optimal decompositions," *Ann. l'Institut Henri Poincaré*, vol. 56, no. 2, pp. 215–234, 1992.
[20] G. Wahba, "Spline models for observational data," in *Proc. SIAM CBMS-NSF Reg. Conf. Ser. Applied Math.*, 1988, no. 59.



René A. Carmona (M'92) is a Professor with the Statistics and Operations Research Program, Civil Engineering and Operations Research Department, Princeton University, Princeton, NJ. His interests include image analysis, the statistical time–frequency analysis of signals, and the theory and the numerics of stochatic processes.

Dr. Carmona is a Fellow of the Institute of Mathematical Statistics and a member of AMS, IMS, and SIAM.

Wen L. Hwang was born in Kaoshiung, Taiwan, R.O.C. He received the Ph.D. degree in computer science from New York University, New York, in 1993.

Since January 1995, he has been as Assistant Research Fellow with the Institute of Information Science of the Academia Sinica, Nanking, Taipei, Taiwan. His research interests include signal/image processing, computer vision, and pattern recognition.

Bruno Torrésani, photograph and biography not available at the time of publication.