SIAM J. SCI. COMPUT.
Vol. 31, No. 5, pp. 3922–3945

CHIRPLET APPROXIMATION OF BAND-LIMITED, REAL SIGNALS MADE EASY*

J. M. GREENBERG† AND LAURENT GOSSE‡

Abstract. In this paper we present algorithms for approximating real band-limited signals by multiple Gaussian chirps. These algorithms do not rely on matching pursuit ideas. They are hierarchical, and, at each stage, the number of terms in a given approximation depends only on the number of positive-valued maxima and negative-valued minima of a signed amplitude function characterizing part of the signal. Like the algorithms used in [J. M. Greenberg, Z. Wang, and J. Li, *IEEE Trans. Signal Process.*, 55 (2007), pp. 734–741] and unlike previous methods, our chirplet approximations require neither a complete dictionary of chirps nor complicated multidimensional searches to obtain suitable choices of chirp parameters.

Key words. band-limited signals, chirplet decomposition, Paley–Wiener class

AMS subject classifications. 41A30, 94A12, 94A20, 65T99

DOI. 10.1137/080734674

1. Introduction.

A real-valued signal $f : \mathbb{R} \to \mathbb{R}$ is said to be band-limited if its Fourier transform, usually denoted by $\hat{f}$, has compact support. This constitutes a broad class of signals sometimes referred to as the *Paley–Wiener space* being of paramount importance in applications as most human and natural phenomena should exclude infinite frequencies [31]. The classical Paley–Wiener theorem states that any band-limited signal with finite energy (i.e., such that $\int_{\mathbb{R}} |f(t)|^2 \, dt$ is finite) is indeed the restriction of an entire function of exponential type defined in the whole complex plane [30]. Consequently, it would suffice to know a small part of such a signal in order to be able to extend it arbitrarily to any open set of the complex plane. However, this is unrealistic at present (even if some progress has been made recently; see, for instance, [32] and references therein), and, in practice, one has still to focus on their analysis.

As the Fourier transform is supposed to be of compact support, it may seem a good idea to express numerically band-limited signals relying on optimized algorithms like the well-known *fast Fourier transform* (FFT). In particular, Shannon's sampling theorem ensures that all the signal's information can be recovered from the knowledge of a countable collection of samples. However, we shall explain in section 2.2 that this idea does not lead to economical (so-called *sparse*) representations (consult especially [17]). An alternative can come from the decomposition into *chirps* which read like $A(t)\exp(j\phi(t))$; the term *chirplets* has been coined by Mann and Haykin [24] (see [6] for a precise definition) in an attempt to derive an object more sophisticated than wavelets. Briefly, if Fourier transform gives a full resolution of frequencies but a null time-resolution, wavelets offer a good compromise according to the uncertainty principle but still decompose signals onto horizontal rectangles in the time-frequency (TF) plane [11]. As chirps are endowed with a time-varying phase function $t \mapsto \phi(t)$,

*Received by the editors September 7, 2008; accepted for publication (in revised form) July 9, 2009; published electronically October 22, 2009.
http://www.siam.org/journals/sisc/31-5/73467.html
†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213 (greenber@andrew.cmu.edu).
‡IAC–CNR "Mauro Picone" (sezione di Bari), Via Amendola 122/D, 70126 Bari Italy (l.gosse@ba.iac.cnr.it).

decomposing a signal into a superposition of chirps means that time-frequency *curves* are now involved. The most usual chirps are the Gaussian/quadratic ones, for which both $\ln(A)$ and ϕ are polynomials of degree 2: we shall work in this framework hereafter.

Clearly, since chirps are more complex objects, they need more parameters to be defined correctly. Indeed, one can roughly count that six parameters are needed for the definition of both functions, plus two other parameters dealing with the time and the mean frequency around which the chirp is localized. All in all, we found at the end of section 2.3 that a signal being written as the sum of $p+1$ chirps requires at most $6p+3$ parameters to evaluate. In many cases, this constitutes a great improvement with respect to classical Fourier techniques.

The main drawback in the setup of chirp decomposition techniques is the weight of the computational effort required: according to the literature, most of the algorithms rely on matching pursuit techniques [23] where one first considers a rather complete dictionary of chirps in order to find iteratively the ones matching the signal as best as possible (see [3, 9, 13, 19, 20, 21, 26] and the recent paper [5] motivated by gravitational waves detection). However, in a recent paper, Greenberg and coauthors [12] proposed a completely different approach whereby one gets rid of the dictionary and constructs the chirp approximation by means of a simple and easy-to-implement procedure. Loosely speaking, given a complex signal in polar form $\omega \mapsto a(\omega)\exp(j\varphi(\omega))$, it amounts to seeking local maxima of a (we call them, for instance, ω_n) and approximating both $\ln(a)$ and φ by polynomials of degree 2 admitting an extremum around ω_n. The procedure is first applied to ω_1, the first maximum point of a, in order to derive two polynomials A_1 and ϕ_1, and then it applies to the remaining signal $a(\omega)\exp(j\varphi(\omega)) - A_1(\omega)\exp(j\phi_1(\omega))$ until residues become low enough. In the present paper, we propose a more elaborate algorithm than the one in [12]; it will be presented in detail in section 3. First, section 3.1 will deal with a pointwise selection procedure to compute chirps' parameters. Then, in section 3.2, a mean squares L^2 procedure will be proposed. Sparsity issues for chirplets decompositions are discussed in section 3.3. Both approaches will be tested on an academic example in section 4 which presents a numerical validation of these algorithms. We insist on the fact that the approach is numerically efficient and computationally cheap. Finally, section 5 will be devoted to more "real-life" experiments—trying to decompose a signal slightly corrupted by white noise and seeking a chirp inside a stock market index.

2. Preliminaries.

2.1. Specifics of real, band-limited signals. Our interest lies in an efficient, approximate representation of real-valued, band-limited signals $t \mapsto f(t)$. If $f(\cdot)$ is such a time-dependent signal, it may be written in a very general way as

$$\forall t \in \mathbb{R}, \qquad f(t) = \frac{1}{\pi}\int_{-\Omega}^{\Omega} e^{jwt}\mathcal{H}(\omega)d\omega, \tag{2.1}$$

where $j = \sqrt{-1}$ is the unit of the imaginary axis, ω stands from now on for the Fourier dual variable, and the function $\mathcal{H}$ rewrites as

$$\forall \omega \in (-\Omega, \Omega), \qquad \mathcal{H}(\omega) = H_e(\omega) - jH_o(\omega).$$

The corresponding even and odd components of $\mathcal{H}$, denoted by $H_e(\cdot)$ and $H_o(\cdot)$, are smooth real-valued functions satisfying

$$\forall \omega \in (-\Omega, \Omega), \qquad H_e(-\omega) = H_e(\omega) \ \text{ and } \ H_o(-\omega) = -H_o(\omega). \tag{2.2}$$

By definition of band-limited, both H_e and H_o vanish identically outside of the interval $(-\Omega, \Omega)$; moreover, they are assumed to satisfy

$$\lim_{\epsilon \to 0^+} H_e(\Omega - \epsilon) = \lim_{\epsilon \to 0^+} H_o(\Omega - \epsilon) = 0. \tag{2.3}$$

The symmetries of $H_e(\cdot)$ and $H_o(\cdot)$ imply that

$$\forall \omega \in (-\Omega, \Omega), \qquad \overline{\mathcal{H}}(\omega) = \mathcal{H}(-\omega), \tag{2.4}$$

which is a well-known property of the Fourier transform for real signals (the overbar stands hereafter for complex conjugate). The function $\mathcal{H}(\cdot)$ can be recovered from the signal via the classical Fourier transform:

$$\mathcal{H}(\omega) = \int_{-\infty}^{\infty} e^{-j\omega t} f(t) dt. \tag{2.5}$$

This last identity further implies that for all ω,

$$H_e(\omega) = 2\int_0^{\infty} \cos(\omega t) f_e(t) dt, \qquad H_o(\omega) = 2\int_0^{\infty} \sin(\omega t) f_o(t) dt \tag{2.6}$$

hold for f_e and f_o, the even/odd parts of the signal f defined for all $t \in \mathbb{R}$ as follows:

$$f_e(t) \stackrel{\text{def}}{=} \frac{1}{2}\Big(f(t) + f(-t)\Big) = f_e(-t) \text{ and } f_o(t) \stackrel{\text{def}}{=} \frac{1}{2}\Big(f(t) - f(-t)\Big) = -f_o(-t).$$

We regard (2.5)–(2.6) as a sanity check on how well we are doing in synthesizing $f(\cdot)$ from $\mathcal{H}(\cdot)$; that is, if we employ some algorithm to compute (2.1) approximately, we then use that computed $f(\cdot)$ to approximately evaluate (2.5) and see how well the result agrees with the original function $\mathcal{H}(\cdot)$.

2.2. Standard representations of band-limited signals. Standard representations of band-limited signals may be obtained by discretizing the integral defined in (2.1). If we introduce a discrete (and finite) set of frequencies ω_k

$$\omega_k = \frac{k\Omega}{N}, \quad k \in \{-N, -N+1, \ldots, N\},$$

and exploit (2.3), we find that the trapezoidal rule, applied to (2.1), yields the approximate band-limited function, $f_N(\cdot)$, whose value reads

$$\forall t \in \mathbb{R}, \qquad f_N(t) = \frac{\Omega}{N\pi} \sum_{k=-N+1}^{N-1} e^{j\frac{k\Omega t}{N}} \mathcal{H}\left(\frac{k\Omega}{N}\right).$$

The function is real-valued and satisfies $|f(t) - f_N(t)| = 0(1/N^2)$ provided the functions $H_e(\cdot)$ and $H_o(\cdot)$ are C^2 on $[-\Omega, \Omega]$. More interestingly, $f_N(\cdot)$ is periodic with period $T_N = \frac{2N\pi}{\Omega}$ and is completely determined by its values at times $t_n = \frac{n\pi}{\Omega}$, $-N \leq n \leq N-1$ (this is the classical Shannon sampling theorem). To see this we note that

$$f_N\left(\frac{n\pi}{\Omega}\right) \stackrel{\text{def}}{=} \frac{\Omega}{N\pi} \sum_{k=-N+1}^{N-1} e^{j\frac{kn\pi}{N}} \mathcal{H}\left(\frac{k\Omega}{N}\right), \tag{2.7}$$

which yields

$$\sum_{n=-N}^{N-1} e^{-j\frac{pn\pi}{N}} f_N\left(\frac{n\pi}{\Omega}\right) = \frac{\Omega}{N\pi} \sum_{k=-N+1}^{N} \mathcal{H}\left(\frac{k\Omega}{N}\right) \left(\sum_{n=-N}^{N-1} e^{j\frac{(k-p)n\pi}{N}} \right). \tag{2.8}$$

Now, one observes the following property of the exponentials:

$$\sum_{n=-N}^{N-1} e^{j\frac{(k-p)n\pi}{N}} = \begin{cases} 0 & \text{if } k \neq p, \\ 2N & \text{if } k = p. \end{cases} \tag{2.9}$$

The identities (2.7)–(2.9) imply that for $-N+1 \leq p \leq N-1$

$$\mathcal{H}\left(\frac{p\Omega}{N}\right) = \frac{\pi}{2\Omega} \sum_{n=-N}^{N-1} e^{-j\frac{pn\pi}{N}} f_N\left(\frac{n\pi}{\Omega}\right), \tag{2.10}$$

while for $p = \pm N$ they yield

$$0 = \sum_{n=-N}^{N-1} e^{-jn\pi} f_N\left(\frac{n\pi}{\Omega}\right) = \sum_{n=-N}^{N-1} e^{jn\pi} f_N\left(\frac{n\pi}{\Omega}\right). \tag{2.11}$$

This identity (2.11) shows that if we extend (2.10) to the indices $p = \pm N$, then the extension is consistent with the constraint (2.3). This last set of identities gives an alternative means of computing $\mathcal{H}(\cdot)$ at the lattice points $\frac{p\Omega}{N}$, $-N \leq p \leq N$. For completeness, we record relevant identities for the coefficients $(H_e(\frac{p\Omega}{N}),\ H_o(\frac{p\Omega}{N}))$, $0 \leq p \leq N-1$. They are

$$H_e\left(\frac{p\Omega}{N}\right) = \frac{\pi}{2\Omega} f_N(0) + \frac{\pi}{\Omega} \sum_{n=1}^{N} \cos\left(\frac{pn\pi}{N}\right) f_{N,e}\left(\frac{n\pi}{\Omega}\right), \quad 0 \leq p \leq N-1, \tag{2.12}$$

$$H_o\left(\frac{p\Omega}{N}\right) = \frac{\pi}{\Omega} \sum_{n=1}^{N} \sin\left(\frac{pn\pi}{N}\right) f_{N,o}\left(\frac{n\pi}{\Omega}\right), \quad 1 \leq p \leq N-1, \tag{2.13}$$

where

$$f_{N,e}\left(\frac{n\pi}{\Omega}\right) \stackrel{\text{def}}{=} \frac{1}{2}\left(f_N\left(\frac{n\pi}{\Omega}\right) + f_N\left(\frac{-n\pi}{\Omega}\right)\right) = f_{N,e}\left(\frac{-n\pi}{\Omega}\right), \quad 1 \leq n \leq N,$$

and

$$f_{N,o}\left(\frac{n\pi}{\Omega}\right) \stackrel{\text{def}}{=} \frac{1}{2}\left(f_N\left(\frac{n\pi}{\Omega}\right) - f_N\left(\frac{-n\pi}{\Omega}\right)\right) = -f_{N,o}\left(\frac{-n\pi}{\Omega}\right), \quad 1 \leq n \leq N.$$

The $\frac{2N\pi}{\Omega}$ periodicity of $f_N(\cdot)$ guarantees that $f_{N,e}\left(\frac{N\pi}{\Omega}\right) = f_N\left(\frac{-N\pi}{\Omega}\right)$ and that $f_{N,o}\left(\frac{N\pi}{\Omega}\right) = 0$. Moreover, if we extend (2.12) to $p = N$, then (2.11) guarantees that the extension is consistent with (2.2). Similarly, if we extend (2.13) to $p = N$, we find the extension is also consistent with (2.3). The standard approach outlined above will, in the limit as $N \to \infty$, yield the desired signal $f(\cdot)$ but is computationally intensive and requires a significant amount of data. Our goal here is a more economical approximate representation of $f(\cdot)$, which, in many circumstances, requires substantially less data.

2.3. Chirplet representation of band-limited signals. We first note that if H_e and H_o are defined as in section 2.1 and satisfy (2.3), then $\mathcal{H}(\cdot)$ can be written in polar form (at least at points where it does not vanish):

$$\forall \omega \in (-\Omega, \Omega), \qquad \mathcal{H}(\omega) = A(\omega)e^{-j\psi(\omega)},$$

where the "amplitude" is real and nonnegative and satisfies

$$A(\omega) \stackrel{\text{def}}{=} \left(H_e^2(\omega) + H_o^2(\omega)\right)^{1/2} = A(-\omega).$$

Furthermore, $A(\omega)$ is identically zero on $|\omega| > \Omega$ and satisfies $\lim_{\epsilon \to 0^+} A(\Omega - \epsilon) = 0$. The "phase" $\omega \mapsto \psi(\omega)$ is a smooth, odd function satisfying

$$\cos\psi(\omega) = \frac{H_e(\omega)}{A(\omega)} \quad \text{and} \quad \sin\psi(\omega) = \frac{H_o(\omega)}{A(\omega)}.$$

As our *canonical model* for $A(\cdot)$ we assume that it has exactly p local maxima at the distinct points

$$0 < \Omega_1 < \Omega_2 < \cdots < \Omega_p < \Omega \tag{2.14}$$

and that each of these maxima is nondegenerate; i.e., $A^{(2)}(\Omega_k) < 0,\ 1 \le k \le p$. The fact that $A(-\omega) = A(\omega)$ guarantees that $A(\cdot)$ also has nondegenerate local maxima located at the symmetric set of points

$$-\Omega < -\Omega_p < \cdots < -\Omega_2 < -\Omega_1 < 0. \tag{2.15}$$

The origin, $\omega = 0$, will be either a local maximum or a local minimum of $A(\cdot)$. In the former case we shall assume that $A^{(2)}(0) < 0$. Given the structural properties of the amplitude $A(\cdot)$, we attempt to approximate it by means of functions $A_p(\cdot)$ of the following Gaussian form:

$$A_p(\omega) = \alpha_0 e^{-\frac{\omega^2}{2\sigma_0}} + \sum_{k=1}^{p} \alpha_k \left(e^{-\frac{(\omega-\omega_k)^2}{2\sigma_k}} + e^{-\frac{(\omega+\omega_k)^2}{2\sigma_k}} \right), \tag{2.16}$$

where $\alpha_k > 0,\ \sigma_k > 0$, and

$$0 < \omega_1 < \omega_2 < \cdots < \omega_p.$$

When $\omega = 0$ corresponds to a minimum of $A(\cdot)$, we choose $\alpha_0 = 0$. Approximate $A(\cdot)$'s of the form (2.16) are not band-limited, but they have tails which decay rapidly as $|\omega| \to \infty$, and this is adequate for our purposes.

In section 3 we present two algorithms for choosing the parameters $\alpha_k > 0$ and $\sigma_k > 0$ and the numbers ω_k's and show how these selection procedures perform on various examples. For the time being, we assume the relevant parameters are known and instead of working with $\mathcal{H}(\omega) = A(\omega)e^{-j\psi(\omega)}$ we replace it with

$$\mathcal{H}_{p,1}(\omega) \stackrel{\text{def}}{=} A_p(\omega)e^{-j\psi(\omega)}. \tag{2.17}$$

Using this approximation to $\mathcal{H}(\cdot)$, we arrive at the following approximation for the signal defined by (2.1):

$$
\begin{aligned}
f_{p,1}(t) &= \alpha_0 \int_{-\infty}^{\infty} e^{-\frac{\omega^2}{2\sigma_0}+j(\omega t-\psi(\omega))}d\omega \\
&\quad + \sum_{k=1}^{p}\alpha_k \left(e^{j\omega_k t}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2\sigma_k}+j(ut-\psi(\omega_k+u))}du \right. \\
&\quad \left. + e^{-j\omega_k t}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2\sigma_k}+j(ut-\psi(-w_k+u))}du \right).
\end{aligned} \tag{2.18}
$$

To carry this process further, we replace the terms $\psi(\pm w_k + u)$ in (2.18) by their local quadratic Taylor approximations around $\pm w_k$. The result is

$$
\psi(\omega_k + u) = \gamma_k + t_k u + \frac{\kappa_k u^2}{2}, \qquad \psi(-\omega_k + u) = -\gamma_k + t_k u - \frac{\kappa_k u^2}{2};
$$

hence

$$
\begin{cases}
\gamma_k = \psi(\omega_k) = -\psi(-\omega_k), \\
t_k = \psi'(\omega_k) = \psi'(-\omega_k), \\
\kappa_k = \psi''(\omega_k) = -\psi''(-\omega_k).
\end{cases} \tag{2.19}
$$

We note that this last step is equivalent to replacing $\mathcal{H}_{p,1}(\omega)$ with

$$
\begin{aligned}
\mathcal{H}_{p,2}(\omega) &= \alpha_0 e^{-\frac{\omega^2}{2\sigma_0}-jt_0\omega} \\
&\quad + \sum_{k=1}^{p}\alpha_k e^{-\frac{(\omega-\omega_k)^2}{2\sigma_k}-j\left(\gamma_k+t_k(\omega-\omega_k)+\frac{\kappa_k}{2}(\omega-\omega_k)^2\right)} \\
&\quad + \sum_{k=1}^{p}\alpha_k e^{-\frac{(\omega+\omega_k)^2}{2\sigma_k}+j\left(\gamma_k-t_k(\omega+\omega_k)+\frac{\kappa_k}{2}(\omega+\omega_k)^2\right)}
\end{aligned} \tag{2.20}
$$

and computing (2.1) with $\mathcal{H}(\cdot)$ replaced by $\mathcal{H}_{p,2}(\cdot)$. The function $\mathcal{H}_{p,2}(\cdot)$ satisfies $\overline{\mathcal{H}}_{p,2}(\omega) = \mathcal{H}_{p,2}(-\omega)$, as required by (2.4) of $\mathcal{H}(\cdot)$. The reader will note that $\mathcal{H}_{p,2}(\cdot)$ is merely the sum of $2p+1$ complex *Gaussian (or quadratic) chirps*. Exploiting the quadratic approximations defined in (2.19) when evaluating (2.18) yields, after a tedious calculation, the approximation $f_{p,2}(\cdot)$ to $f(\cdot)$:

$$
\begin{aligned}
f_{p,2}(t) &= (2\pi\sigma_0)^{1/2}\alpha_0 e^{-\frac{\sigma_0(t-t_0)^2}{2}} \\
&\quad + 2^{3/2}\pi^{1/2}\sum_{k=1}^{p}\alpha_k \frac{\sigma_k^{1/2} e^{\frac{-\sigma_k(t-t_k)^2}{2(1+\kappa_k^2\sigma_k^2)}}}{(1+\sigma_k^2\kappa_k^2)^{1/4}} \cos\left(\frac{\kappa_k\sigma_k^2(t-t_k)^2}{2(1+\kappa_k^2\sigma_k^2)} + \omega_k t - \gamma_k - \phi_k\right),
\end{aligned} \tag{2.21}
$$

where

$$
(1 + j\kappa_k\sigma_k) = (1+\kappa_k^2\sigma_k^2)^{1/2} e^{2j\phi_k}
$$

or equivalently

$$\cos 2\phi_k = \frac{1}{(1+\kappa_k^2\sigma_k^2)^{1/2}} \quad \text{and} \quad \sin 2\phi_k = \frac{\kappa_k\sigma_k}{(1+\kappa_k^2\sigma_k^2)^{1/2}}. \tag{2.22}$$

The aforementioned approximation $f_{p,2}(\cdot)$ is simply the sum of $p+1$ real Gaussian chirps which are characterized by $6p+3$ parameters

- $(\alpha_0, \sigma_0, t_0)$,
- $(\alpha_k, \omega_k, \sigma_k, \gamma_k, t_k, \kappa_k)$ for $k = \{1, \dots, p\}$.

Finally, ϕ_k is computed using (2.22).

2.4. The objective of the paper. Approximations of band-limited signals by real-valued Gaussian chirps of the type described in (2.21) offer efficient representation of radar, seismic, and some biological signals. A sampling of the literature on this subject may be found in [3, 5, 13, 17, 19, 20, 21, 26, 28] and the references contained therein. The central problem in establishing such a formula is an efficient and natural method for obtaining the parameters characterizing the chirps. The "matching-pursuit" and "ridge-pursuit" algorithms have been used by a number of authors; for details see, e.g., [3, 9, 13, 23, 33]. These typically require complicated multidimensional searches to obtain the chirp parameters, and they also demand a complete a priori dictionary of chirps. The procedures we advance in section 3 to capture the parameters $(\alpha_k, \sigma_k, \omega_k)_{k=1}^p$ require no such complete dictionary and rely only on information about the signal amplitude $A(\cdot)$. The remaining chirp parameters $(\gamma_k, t_k, \kappa_k)_{k=1}^p$ are obtained from local information about the phase, $\psi(\cdot)$, near the points ω_k. The algorithms are also hierarchical and give our approximations a multiresolution character. Less refined versions of these procedures were advanced in [12].

3. Parameter selection for $A_p(\cdot)$. In this section we present two different algorithms for the selection of the parameters defining $A_p(\cdot)$; recall (2.16). For definiteness, we assume $\omega = 0$ is a local minimum of $A(\cdot)$ and thus choose $\alpha_0 = 0$.

3.1. Pointwise selection procedure. We attempt to choose the parameters so that at the positive local maxima, $\Omega_1 < \Omega_2 < \cdots < \Omega_p$, one gets

$$(3.1)_j \quad A_p(\Omega_j) = A(\Omega_j),\ A_p^{(1)}(\Omega_j) = A^{(1)}(\Omega_j) = 0,\ \text{and}\ A_p^{(2)}(\Omega_j) = A^{(2)}(\Omega_j) < 0$$

for $1 \le j \le p$. Hereafter, we denote by g the Gaussian function centered in $\pm\omega_k$:

$$g(\omega; \omega_k, \sigma_k) = \left(e^{-\frac{(\omega-\omega_k)^2}{2\sigma_k}} + e^{-\frac{(\omega+\omega_k)^2}{2\sigma_k}} \right). \tag{3.1}$$

Then, its first derivative reads

$$g^{(1)}(\omega, \omega_k, \sigma_k) = -\left(\frac{(\omega-\omega_k)}{\sigma_k} e^{-\frac{(\omega-\omega_k)^2}{2\sigma_k}} + \frac{(\omega+\omega_k)}{\sigma_k} e^{-\frac{(\omega-\omega_k)^2}{2\sigma_k}} \right),$$

and its second derivative is

$$g^{(2)}(\omega; \omega_k, \sigma_k) = \left(\left(-\frac{1}{\sigma_k} + \frac{(\omega-\omega_k)^2}{\sigma_k^2} \right) e^{-\frac{(\omega-\omega_k)^2}{2\sigma_k}} + \left(-\frac{1}{\sigma_k} + \frac{(\omega+\omega_k)^2}{\sigma_k^2} \right) e^{-\frac{(\omega+\omega_k)^2}{2\sigma_k}} \right).$$

Hence, for $1 \le j \le p$, the set of equations $(3.1)_j$ rewrites as

$$\alpha_j g(\omega; \omega_j, \sigma_j) = A(\Omega_j) - \sum_{\substack{k=1 \\ k \ne j}}^{p} \alpha_k g(\Omega_j; \omega_k, \sigma_k), \tag{3.2}$$

which boils down to

$$\alpha_j e^{-\frac{(\Omega_j-\omega_j)^2}{2\sigma_j}} = A(\Omega_j) - \sum_{\substack{k=1\\k\neq j}}^{p} \alpha_k g(\Omega_j;\omega_k,\sigma_k) - \alpha_j e^{-\frac{(\Omega_j+\omega_j)^2}{2\sigma_j}}.$$

We differentiate this equality once,

$$\alpha_j \frac{(\Omega_j-\omega_j)}{\sigma_j} e^{-\frac{(\Omega_j-\omega_j)^2}{2\sigma_j}} = \sum_{\substack{k=1\\k\neq j}}^{p} \alpha_k g^{(1)}(\Omega_j;\omega_k,\sigma_k) - \alpha_j \frac{(\Omega_j+\omega_j)}{\sigma_j} e^{-\frac{(\Omega_j+\omega_j)^2}{2\sigma_j}},$$

and twice,

$$\begin{aligned}\alpha_j \left(-\tfrac{1}{\sigma_j} + \tfrac{(\Omega_j-\omega_j)^2}{\sigma_j^2}\right) e^{-\frac{(\Omega_j-\omega_j)^2}{2\sigma_j}} &= A^{(2)}(\Omega_j) - \sum_{\substack{k=1\\k\neq j}}^{p} \alpha_k g^{(2)}(\Omega_j;\omega_k,\sigma_k)\\ &\quad + \alpha_j \left(\tfrac{1}{\sigma_j} - \tfrac{(\Omega_j+\omega_j)^2}{\sigma_j^2}\right) e^{-\frac{(\Omega_j+\omega_j)^2}{2\sigma_j}}.\end{aligned}$$

We have in mind to propose an iterative algorithm to compute a solution to the above system of three equations in order to get the value of the three parameters; the index n will refer to the successive approximations built out of this iterative process.

- For the first index $n=0$, we let

$$\alpha_j^0 = A(\Omega_j),\ \omega_j^0 = \Omega_j,\ \text{and}\ \sigma_j^0 = -\frac{A(\Omega_j)}{A^{(2)}(\Omega_j)}\ \text{for}\ 1\le j\le p. \tag{3.3}$$

- We now suppose the coefficients $\left(\alpha_j^n,\omega_j^n,\sigma_j^n\right)_{j=1}^p$ are known; we want to compute these parameters at step $n+1$. For the first index $j=1$, we let

$$\begin{cases} \mathcal{A}_1^{n+1} = A(\Omega_1) - \sum_{k=2}^{p} \alpha_k^n g\left(\Omega_1;\omega_k^n,\sigma_k^n\right) - \alpha_1^n e^{-\frac{(\Omega_1+\omega_1^n)^2}{2\sigma_1^n}},\\ \mathcal{B}_1^{n+1} = \sum_{k=2}^{p} \alpha_k^n g^{(1)}\left(\Omega_1;\omega_k^n,\sigma_k^n\right) - \frac{\alpha_1^n(\Omega_1+\omega_1^n)}{\sigma_1^n} e^{-\frac{(\Omega_1+\omega_1^n)^2}{2\sigma_1^n}},\\ \mathcal{C}_1^{n+1} = A^{(2)}(\Omega_1) - \sum_{k=2}^{p} \alpha_k^n g^{(2)}(\Omega_1;\omega_k^n,\sigma_k^n)\\ \qquad\quad + \alpha_1^n \left(\frac{1}{\sigma_1^n} - \frac{(\Omega_1+\omega_1^n)^2}{(\sigma_1^n)^2}\right) e^{-\frac{(\Omega_1+\omega_1^n)^2}{2\sigma_1^n}}. \end{cases}$$

If $\mathcal{A}_1^{n+1}>0$ and $\mathcal{C}_1^{n+1}<0$, then we can compute an auxiliary quantity,

$$\mathcal{D}_1^{n+1} = \mathcal{B}_1^{n+1}/\mathcal{A}_1^{n+1},$$

and $\sigma_1^{n+1},\omega_1^{n+1}$, and α_1^{n+1} are deduced as follows:

$$\begin{cases} \sigma_1^{n+1} = \dfrac{\mathcal{A}_1^{n+1}}{\mathcal{A}_1^{n+1}(\mathcal{D}_1^{n+1})^2 - \mathcal{C}_1^{n+1}} > 0,\\ \omega_1^{n+1} = \Omega_1 - \sigma_1^{n+1}\mathcal{D}_1^{n+1},\\ \alpha_1^{n+1} = \mathcal{A}_1^{n+1} e^{\frac{\sigma_1^{n+1}(\mathcal{D}_1^{n+1})^2}{2}} > 0. \end{cases} \tag{3.4}$$

- We now assume that we have determined $\left(\alpha_j^{n+1},\ \omega_j^{n+1}, \sigma_j^{n+1}\right)$ for indices $1 \leq j \leq j_0 \leq p-1$. We then fix the index j_o and compute[1] at step j_o+1

$$\begin{aligned}\mathcal{A}_{j_0+1}^{n+1} &= A\left(\Omega_{j_0+1}\right) - \sum_{k=1}^{j_0} \alpha_k^{n+1} g\left(\Omega_{j_0+1}; \omega_k^{n+1}, \sigma_k^{n+1}\right) \\ &\quad - \alpha_{j_0+1}^n e^{-\frac{(\Omega_{j_0+1}+\omega_{j_0+1}^n)^2}{2\sigma_{j_0+1}^n}} - \sum_{k=j_0+2}^{p} \alpha_k^n g\left(\Omega_{j_0+1}; \omega_k^n, \sigma_k^n\right).\end{aligned} \tag{3.5}$$

Then,

$$\begin{aligned}\mathcal{B}_{j_o+1}^{n+1} &= \sum_{k=1}^{j_o} \alpha_k^{n+1} g^{(1)}\left(\Omega_{j_o+1}; \omega_k^{n+1}, \sigma_k^{n+1}\right) \\ &\quad - \alpha_{j_0+1}^n \frac{\left(\Omega_{j_o+1}+\omega_{j_o+1}^n\right)}{\sigma_{j_o+1}^n} e^{-\frac{(\Omega_{j_o+1}+\omega_{j_0+1})^2}{2\sigma_{j_o+1}}} \\ &\quad + \sum_{k=j_0+2}^{p} \alpha_k^n g^{(1)}\left(\Omega_{j_o+1}; \omega_k^n, \sigma_k^n\right),\end{aligned} \tag{3.6}$$

and

$$\begin{aligned}\mathcal{C}_{j_o+1}^{n+1} &= A^{(2)}(\Omega_{j_o+1}) - \sum_{k=1}^{j_0} \alpha_k^{n+1} g^{(2)}(\Omega_{j_o+1}; \omega_k^{n+1}, \sigma_k^{n+1}) \\ &\quad + \alpha_{j_0+1}^n \left(\frac{1}{\sigma_{j_o+1}^{n+1}} - \frac{(\Omega_{j_o+1}+\omega_{j_0+1}^n)^2}{(\sigma_{j_0+1}^n)^2}\right) e^{-\frac{(\Omega_{j_o+1}+\omega_{j_0+1}^n)^2}{2\sigma_{j_o+1}^n}} \\ &\quad - \sum_{k=j_o+2}^{p} \alpha_k^n g^{(2)}\left(\Omega_{j_o+1}; \omega_k^n, \sigma_2^n\right).\end{aligned} \tag{3.7}$$

If $\mathcal{A}_{j_o+1}^{n+1} > 0$ and $\mathcal{C}_{j_o+1}^{n+1} < 0$, then the auxiliary quantity reads

$$\mathcal{D}_{j_o+1}^{n+1} = \mathcal{B}_{j_o+1}^{n+1} / \mathcal{A}_{j_o+1}^{n+1},$$

and we compute $\sigma_{j_o+1}^{n+1},\ \omega_{j_o+1}^{n+1}$, and $\alpha_{j_o+1}^{n+1}$ as follows:

$$\begin{cases} \sigma_{j_o+1}^{n+1} = \dfrac{\mathcal{A}_{j_o+1}^{n+1}}{\mathcal{A}_{j_o+1}^{n+1}(\mathcal{D}_{j_o+1}^{n+1})^2 - \mathcal{C}_{j_o+1}^{n+1}} > 0, \\ \omega_{j_o+1}^{n+1} = \Omega_{j_o+1} - \sigma_{j_o+1}^{n+1} \mathcal{D}_{j_o+1}^{n+1}, \\ \alpha_{j_o+1}^{n+1} = \mathcal{A}_{j_o+1}^{n+1} e^{\frac{\sigma_{j_o+1}^{n+1} (\mathcal{D}_{j_o+1}^{n+1})^2}{2}} > 0. \end{cases} \tag{3.8}$$

We thus iterate until satisfactory convergence is obtained.

The aforementioned algorithm realizes a compromise between simplicity and efficiency because it stems on inverting only half of the nonlinearities appearing in (3.2) and its derivatives in order to keep computations easy as it suffices to define the auxiliary quantity $\mathcal{D}$ to derive (3.4) and (3.8). The nonlinear problem is thus solved by a

[1] If $j_o = p-1$, the last sums in (3.5)–(3.7) are not present.

fixed-point algorithm in $\mathbb{R}^{3p}$, whose convergence is ensured as soon as the underlying mapping has a Lipschitz constant $K < 1$. Hence it realizes a contraction, and thus the approximation error decays by a factor K at each iteration. Such a convergence rate does not hold in the case in which the problem becomes degenerate, that is, if the second derivative of the amplitude A vanishes at a local maximum point (for instance, there is an issue with (3.3)). This way of proceeding is related to standard iterative methods in numerical linear algebra, such as, for instance, Jacobi or Gauss–Seidel for which convergence results are well known [8].

Remark 1. We note that with this method of parameter selection, standard Taylor expansions imply that the following estimate holds: $|A(\omega) - A_p(\omega)| = O(|\omega - \Omega_j|^3)$ in a neighborhood of each of the local maxima, Ω_j, of $A(\cdot)$. This pointwise selection procedure generalizes the procedure used in [12] and captures the behavior of $A(\cdot)$ near all maxima simultaneously.

3.2. L^2 parameter selection. Again, we seek to approximate the amplitude $A(\cdot)$ by a function of the form $A_p(\cdot)$ (recall (2.16)) and choose the parameters $(\alpha_k, \omega_k, \sigma_k)$ so as to minimize the mean-squares error:

$$\begin{aligned} ||A(\cdot) - A_p(\cdot)||^2 &\stackrel{\text{def}}{=} \int_{-\infty}^{\infty} |A(\omega) - A_p(\omega)|^2 \, d\omega \\ &= \int_{-\Omega}^{\Omega} A^2(\omega) d\omega - 2\int_{-\Omega}^{\Omega} A(\omega) A_p(\omega) d\omega + \int_{-\infty}^{\infty} A_p^2(\omega) d\omega. \end{aligned}$$

For definiteness, we again assume that $\omega = 0$ is a local minimum of $A(\cdot)$ and choose $\alpha_0 = 0$. We also adopt the short-hand notation

$$G_i(\omega) = \left(e^{-\frac{(\omega-\omega_i)^2}{2\sigma_i}} + e^{-\frac{(\omega+\omega_i)^2}{2\sigma_i}} \right) = g(\omega; \omega_i, \sigma_i), \qquad 1 \le i \le p, \tag{3.9}$$

and note, for future reference, that for all $1 \le i$ and $j \le p$,

$$\begin{aligned} \langle G_i, G_j \rangle = \langle G_j, G_i \rangle &\stackrel{\text{def}}{=} \int_{-\infty}^{\infty} G_i(\omega) G_j(\omega) d\omega \\ &= 2 \left(\frac{2\sigma_i \sigma_j \pi}{\sigma_i + \sigma_j} \right)^{1/2} \left(e^{-\frac{(\omega_i - \omega_j)^2}{2(\sigma_i + \sigma_j)}} + e^{-\frac{(\omega_i + \omega_j)^2}{2(\sigma_i + \sigma_j)}} \right). \end{aligned} \tag{3.10}$$

A quick calculation shows that $||A(\cdot) - A_p(\cdot)||^2$ is given by

$$||A(\cdot) - A_p(\cdot)||^2 = \int_{-\Omega}^{\Omega} A^2(\omega) d\omega - 2\sum_{k=1}^{p} f_k \alpha_k + \sum_{i,j=1}^{p} \alpha_i \langle G_i, G_j \rangle \alpha_j \ge 0, \tag{3.11}$$

where the value of $\langle G_i, G_j \rangle$ has been given above and

$$f_k \stackrel{\text{def}}{=} \langle A, G_k \rangle = \int_{-\Omega}^{\Omega} A(\omega) G_k(\omega) d\omega, \qquad 1 \le k \le p. \tag{3.12}$$

For a given choice of parameters $(\omega_i, \sigma_i)_{i=1}^p$, the $\boldsymbol{\alpha}$'s which minimize (3.11) satisfy

$$\mathcal{G}\boldsymbol{\alpha} = \boldsymbol{f}, \tag{3.13}$$

where $\mathcal{G}$ is the symmetric, positive-definite, $p \times p$ matrix whose (i,j)th entry reads

$$\mathcal{G}_{i,j} = \langle G_i, G_j \rangle = \mathcal{G}_{j,i}, \tag{3.14}$$

$\boldsymbol{\alpha}$ is the $p \times 1$ column vector whose ith entry is α_i, and $\boldsymbol{f}$ is the $p \times 1$ column vector whose ith entry is f_i. The solution to (3.13) is given by $\boldsymbol{\alpha} = \mathcal{G}^{-1}\boldsymbol{f}$, and this implies

$$||A(\cdot) - A_p(\cdot)||^2 = \int_{-\Omega}^{\Omega} A^2(\omega)d\omega - \boldsymbol{f}^T\mathcal{G}^{-1}\boldsymbol{f} \geq 0.$$

Thus, the task before us is to choose $(\omega_i, \sigma_i)_{i=1}^p$ to maximize $\boldsymbol{f}^T\mathcal{G}^{-1}\boldsymbol{f}$, where again $\mathcal{G}^{-1}$ is the symmetric, positive-definite inverse of $\mathcal{G}$ defined in (3.14) above. At first blush, the problem of maximizing $Q = \boldsymbol{f}^T\mathcal{G}^{-1}\boldsymbol{f}$ looks rather daunting, but, in fact, this problem has a rather simple structure.

- For any integer k between 1 and p, we let β_k be one of the two parameters ω_k or σ_k, and we observe that (the notation $.,_{\beta_k}$ stands for partial derivative with respect to β_k)

$$\frac{\partial Q}{\partial \beta_k} = \boldsymbol{f}^T\mathcal{G}^{-1},_{\beta_k}\boldsymbol{f} + 2\boldsymbol{f}^T\mathcal{G}^{-1}\boldsymbol{f},_{\beta_k}. \tag{3.15}$$

  Noting that $\boldsymbol{f}^T\mathcal{G}^{-1} = \boldsymbol{\alpha}^T$ and that $\boldsymbol{f},_{\beta_k}$ is the $p \times 1$ column vector whose kth component is $f_{k,\beta_k} = \langle A, G_{k,\beta_k} \rangle$ and other components zero, we see that the second term on the right-hand side of (3.15) is $2\alpha_k f_{k,\beta_k}$. Moreover, if we exploit the identity

$$\mathcal{G}^{-1},_{\beta_k} = -\mathcal{G}^{-1}\mathcal{G},_{\beta_k}\mathcal{G}^{-1}, \tag{3.16}$$

  we find that the first term on the right-hand side of (3.15) is

$$-2\alpha_k \sum_{\substack{j=1 \\ j\neq k}}^{p} \langle G_k, G_j \rangle_{,\beta_k} \alpha_j - \alpha_k^2 \langle G_k, G_k \rangle_{,\beta_k},$$

  and again the $\boldsymbol{\alpha}$'s are the solution of (3.13). These observations imply that

$$\frac{\partial Q}{\partial \beta_k} = 2\alpha_k \left(\frac{\partial f_k}{\partial \beta_k} - \sum_{\substack{j=1 \\ j\neq k}}^{p} \frac{\partial \langle G_k, G_j \rangle}{\partial \beta_k} \alpha_j \right) - \alpha_k^2 \langle G_k, G_k \rangle_{,\beta_k},$$

  where again $\mathcal{G}\boldsymbol{\alpha} = \boldsymbol{f}$. Given the particularly simple structure of $\frac{\partial Q}{\partial \beta_i}$, we solve the problem of maximizing Q by a "steepest ascent" type algorithm, as we explain now.
- So we assume now that $(\omega_i^n, \sigma_i^n)_{i=1}^p$ are known. With these data, and the explicit formula (3.9) for $G_i(\omega)$, we explicitly compute the $\langle G_i, G_j \rangle$'s and $\frac{\partial}{\partial \beta_i}\langle G_i, G_j \rangle$'s at the nth data set. Once again, $\beta_i = \omega_i$ or σ_i. We denote the results by $\langle G_i, G_j \rangle^n$ and $\frac{\partial}{\partial \beta_i}\langle G_i, G_j \rangle^n$, respectively. We sample the amplitude $A(\cdot)$ at points

$$\omega_p = \frac{p\Omega}{N}, \quad -N+1 \leq p \leq N-1,$$

and do the same for the functions $G_i(\cdot)$ and $\frac{\partial G_i}{\partial \beta_i}(\cdot)$. These are of course also evaluated with the nth data set, and the results are superscripted with the index n. We then derive approximate values for f_i and $\frac{\partial f_i}{\partial \beta_i}$ using the discrete inner products defined below to replace the integrals (see (3.12))

$$f_i^n \approx \frac{\Omega}{N} \sum_{p=-N+1}^{N+1} G_i^n\left(\frac{p\Omega}{N}\right) A\left(\frac{p\Omega}{N}\right)$$

and

$$\frac{\partial f_i^n}{\partial \beta_i} \approx \frac{\Omega}{N} \sum_{p=-N+1}^{N-1} \frac{\partial G_i^n}{\partial \beta_i}\left(\frac{p\Omega}{N}\right) A\left(\frac{p\Omega}{N}\right).$$

Next we obtain $\boldsymbol{\alpha}^n$ by solving $\mathcal{G}^n \boldsymbol{\alpha}^n = \boldsymbol{f}^n$ in order to compute

$$\frac{\partial Q^n}{\partial \beta_i} = 2\alpha_i^n \left(\frac{\partial f_i^n}{\partial \beta_i} - \sum_{\substack{j=1 \\ j\neq 1}}^{p} \frac{\partial \langle G_i, G_j\rangle^n}{\partial \beta_i} \alpha_j^n \right) - (\alpha_i^n)^2 \langle G_i, G_i\rangle^n_{,\beta_i}, \tag{3.17}$$

and we use (3.17) to update the parameters as follows:

$$\omega_i^{n+1} = \omega_i^n + \Delta \frac{\partial Q^n}{\partial \omega_i}, \qquad \sigma_i^{n+1} = \sigma_i^n + \Delta \frac{\partial Q^n}{\partial \sigma_i}.$$

Unless $\frac{\partial Q^n}{\partial \omega_i} = \frac{\partial Q^n}{\partial \sigma_i} = 0,\ 1 \le i \le p,\ Q$ will increase for Δ small enough. We iterate the aforementioned process until convergence.

Remark 2 (convergence rate of steepest descent). Following, e.g., [18], we can state that the standard steepest descent algorithm for the minimization of a quadratic function $\frac{1}{2}x^T Mx + q^T x$, where M is a symmetric positive definite matrix, satisfies the error estimate

$$\|\varepsilon^n\| \le \sqrt{\delta}\left(\frac{\delta - 1}{\delta + 1}\right)^n \|\varepsilon^0\|,$$

where δ is the condition number (the ratio of the biggest eigenvalue of A over its lowest) and $\varepsilon^n = x^n - x^*$, with $x^* = -M^{-1}q$ the optimal point. Here, the situation is more intricate because we must minimize the function (3.11), which boils down to finding $(\omega_i, \sigma_i)_{i=1}^p$ which maximize $\boldsymbol{f}^T \mathcal{G}^{-1} \boldsymbol{f}$, under the constraints (3.12)–(3.14).

3.3. About sparsity of chirplet decomposition. A signal's representation is said to be *sparse in a given base* (or, even more generally, in an overcomplete dictionary) when it happens that most of its information can be represented by only a small number of coefficients (typically, the number of meaningful coefficients has to be much smaller than the length of the signal sample). A canonical example is the one of a sine wave in the Fourier base. This notion of sparsity is somehow vague as, intuitively, it refers to the cardinality of the volume of data; however, in realistic situations where samples are generally corrupted by noise, more elaborate *sparsity measures* come into play: for a signal f with coefficients f_j, $j = 1, 2, \ldots$, there are

- ℓ^p-norms for $p < 2$: $\|f\|_p = (\sum_j |f_j|^p)^{\frac{1}{p}}$,
- Shannon entropy: $S(f) = -\sum_j |f_j|^2 \log_2(|f_j|^2)$, and
- Renyi entropy for $\alpha > 1$: $R_\alpha(f) = \frac{1}{1-\alpha} \log_2(\|f\|_\alpha^\alpha)$.

All these quantities decay as much as the signal's representation involves fewer nonzero coefficients. The ℓ^1 norm is considered in the Basis Pursuit (BP) algorithm; see [7]. A result in [14] states that if the signal admits a "very sparse" representation under a given ℓ^p norm, then it will also be the sparsest with respect to all sparsity measures in ℓ^p, $0 \leq p \leq 1$. However, each criterion leads to a particular optimal representation, and these representations in general differ from one another [15].

Chirplets are believed to yield sparser decompositions for certain signals[2] because they are represented through pieces of parabolic curves in the time-frequency (TF) plane (see Figures 2 and 3 in [24]). In comparison, discrete wavelets decompose signals through rectangles in the TF plane, whereas Fourier exploits horizontal lines (see, for instance, Figure 2 in [11]). This has advantages in certain applications: wolf howls and whale whistles seem to admit a sparse representation when expressed by means of chirplets (see [9, 27]). Concerning the detection of relativistic gravitational waves, several algorithms have been developed as chirplets seem to match the adequate form; see, e.g., [5, 6].

In [27] (see also [2]), the authors raise the problem of building an efficient algorithm generating sparse representations with chirplets; they especially compare their results with the results obtained with matching pursuit. Another important step in this direction is the development of chirp-sensing codes [1] which follow a deterministic compressed sensing methodology in order to acquire a sparse signal with very few measurements (see also [10, 22]). One application in the field of image processing is proposed in [25], where chirps are used instead of random Gaussian atoms.

At the top of Figure 5.5, we present the continuous Morlet wavelet transform of the fluctuation of the HSCEI stock index (cf. bottom of Figure 5.4) together with the one of its one-chirp approximations obtained with the algorithm of section 3.1.

4. A first "academic" numerical example.

4.1. First test. To see how these selection processes perform, we consider the following example. We let

$$A(\omega) = \begin{cases} 0, & -\infty < \omega \leq -2, \\ (4-\omega^2)^2 \left(\frac{1}{2} + \omega^2\right), & -2 \leq \omega \leq 2, \\ 0, & 2 \leq \omega < \infty. \end{cases} \tag{4.1}$$

The origin $\omega = 0$ is a local minimum of A, and the points $\omega = \pm 1$ are the maxima of $A(\cdot)$ with $A(\pm 1) = 13.5$. The second derivative at these points is $A^{(2)}(\pm 1) = -36$. Our approximations are of the form

$$A_1(\omega) = \alpha_1 \left(e^{-\frac{(\omega-\omega_1)^2}{2\sigma_1}} + e^{-\frac{(\omega+\omega_1)^2}{2\sigma_1}} \right), \qquad \omega \in \mathbb{R}. \tag{4.2}$$

The pointwise selection procedure yielded the following results: $\alpha_1 = 13.4515$, $\omega_1 = 1.0074$, and $\sigma_1 = 0.3595$, which in turn yield $\max_\omega A_1(\omega) = A_1(\pm 1) = 13.5$. The L^2 procedure furnished $\alpha_1 = 13.8189$, $\omega_1 = .8974$, and $\sigma_1 = 0.2942$; in this case $\max_\omega A_1(\omega) = A_1(\pm .9) = 13.8782$. The maximal value of Q is 390.9413, and the L^2 norm of $A(\cdot)$ is 395.0055. Figure 4.1 shows a graph of $A(\cdot)$ along with graphs of $A_1(\cdot)$ for each selection procedure. The results for this simple example are not atypical of the general case; namely, one application of either selection process typically yields

[2] In particular, this is known to be the case for multicomponent AM-FM signals; see [2].

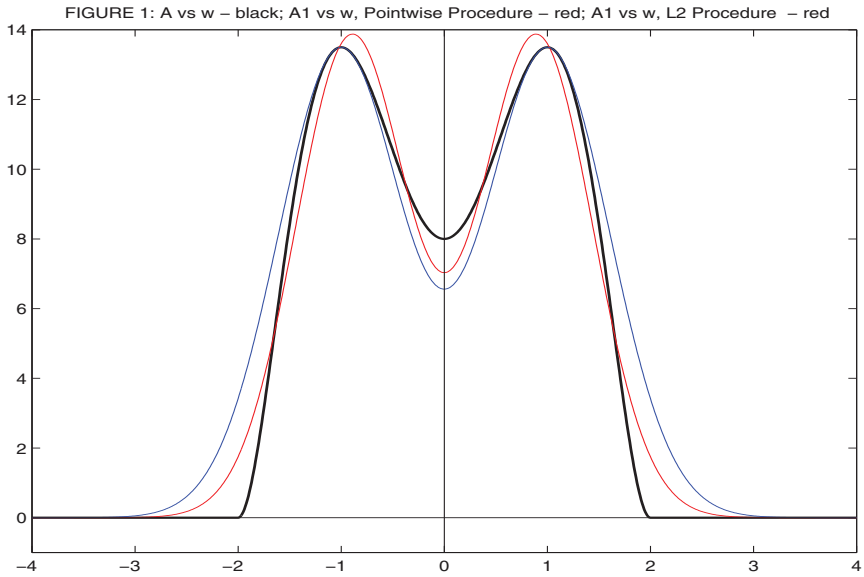

FIG. 4.1. *Comparison between pointwise and L^2 selection procedures with example* (4.1)–(4.2)*: Original signal A is in black, pointwise algorithm is in blue, and L^2 is in red.*

an approximation, $A_p(\cdot)$, which is qualitatively similar to the given amplitude, $A(\cdot)$, but may differ quantitatively from $A(\cdot)$. This defect may be overcome by repeated applications of the procedure.

4.2. Hierarchical refinements of selection procedures. We assume that we have already applied either the pointwise or L^2 selection procedures $n-1$ times and constructed the functions $A_{N_k,k}(\cdot)$, $1 \le k \le n-1$, each of which are sums of N_k Gaussians. We let

$$\forall \omega \in (-\Omega, \Omega), \qquad A_n(\omega) = A_0(\omega) - \sum_{k=1}^{n-1} A_{p_k,k}(\omega),$$

where $\omega \mapsto A_o(\omega)$ is the original amplitude, previously denoted by $A(\cdot)$. As defined $A_n(\cdot)$ is even and rapidly decreasing as $|\omega| \to \infty$. The principal qualitative difference between $A_n(\cdot)$ and $A_o(\cdot)$ is that $A_n(\cdot)$ takes on both positive and negative values. For definiteness, we assume now that $\omega = 0$ is a maximum of $A_n(\cdot)$ and that the typical structure of $A_n(\cdot)$ is as follows:

- there are exactly q_n negative-valued, local minima of $A_n(\cdot)$ at points

$$0 < \underline{\Omega}_1^n < \underline{\Omega}_2^n < \cdots < \underline{\Omega}_{q_n}^n,$$

  and each minima is nondegenerate;
- there are exactly p_n positive-valued, local maxima of $A_n(\cdot)$ at points

$$0 < \overline{\Omega}_1^n < \overline{\Omega}_2^n < \cdots < \overline{\Omega}_{p_n}^n,$$

  and each maxima is nondegenerate;
- the minima and maxima are not necessarily interlaced since $A_n(\cdot)$ may have positive-valued local minima and negative-valued local maxima.

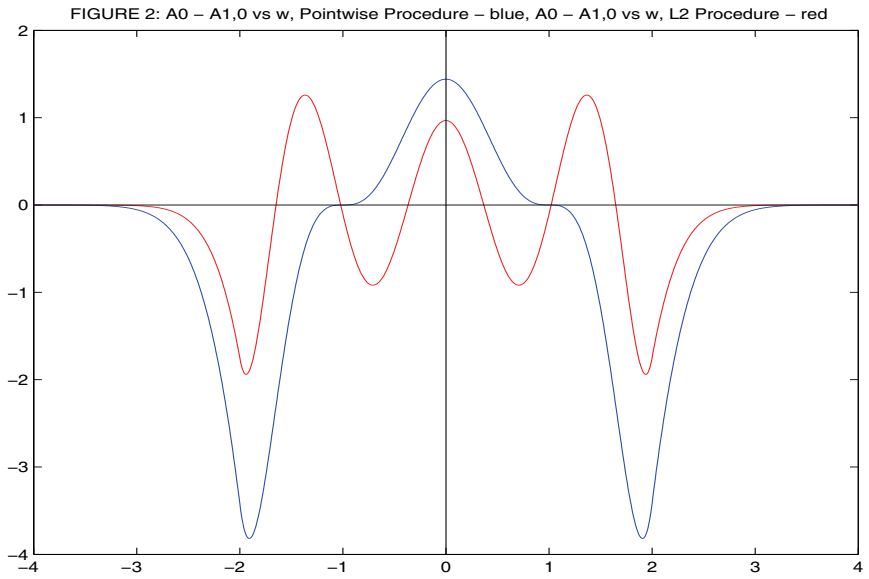

FIG. 4.2. *Difference $A_0(\omega) - A_{1,0}(\omega)$: pointwise selection is in blue, and L^2 selection is in red.*

Graphs of $A_1(\cdot) = A_0(\cdot) - A_{1,0}(\cdot)$ for the two selection procedures used in our previous example are shown in Figure 4.2. Our induction step is to replace $A_n(\cdot)$ by a sum of $N_n = q_n + p_n + 1$ even Gaussians:

$$
\begin{aligned}
A_{N_n,n}(\omega) = \alpha_0^n e^{-\frac{\omega^2}{2\overline{\sigma}_0^n}} + \sum_{k=1}^{p_n} \alpha_k^n \left(e^{-\frac{(\omega-\overline{\omega}_k^n)^2}{2\overline{\sigma}_k^n}} + e^{-\frac{(\omega+\overline{\omega}_k^n)^2}{2\overline{\sigma}_k^n}} \right) \\
- \sum_{k=1}^{q_n} \beta_k^n \left(e^{-\frac{(\omega-\underline{\omega}_k^n)^2}{2\underline{\sigma}_k^n}} + e^{-\frac{(\omega+\underline{\omega}_k^n)^2}{2\underline{\sigma}_k^n}} \right),
\end{aligned} \tag{4.3}
$$

where the α's, β's, $\underline{\sigma}$'s, and $\overline{\sigma}$'s are positive and the $\underline{\omega}$'s and $\overline{\omega}$'s satisfy

$$
0 \le \underline{\omega}_1^n < \underline{\omega}_2^n < \cdots < \underline{\omega}_{q_n}^n, \qquad 0 < \overline{\omega}_1^n < \overline{\omega}_2^n < \cdots < \overline{\omega}_{p_n}^n.
$$

The pointwise selection procedure generates the unknown parameters by insisting that $A_{N_n,n}(\cdot)$, $A^{(1)}_{N_n,n}(\cdot)$, and $A^{(2)}_{N_n,n}(\cdot)$ match $A_n(\cdot)$, $A_n^{(1)}(\cdot)$, and $A_n^{(2)}(\cdot)$ at the local maxima $\{0, \{\overline{\Omega}_k^n\}_{k=1}^{p_n}\}$ and local minima $\{\underline{\Omega}_k^n\}_{k=1}^{q_n}$. The L^2 selection procedure chooses the coefficients to minimize $||A_n(\cdot) - A_{N_n,n}(\cdot)||^2$. This problem has the same structure as the optimization problem presented in detail earlier and may be solved by the "steepest-ascent" algorithm. The optimal solution satisfies

$$
0 \le ||A_{n+1}(\cdot)||^2 = ||A_n(\cdot) - A_{N_n,n}(\cdot)||^2 = ||A_n(\cdot)||^2 - Q_{n,\max}, \tag{4.4}
$$

where $(\underline{\omega}^n, \underline{\sigma}^n; \overline{\omega}^n, \overline{\sigma}^n) \to Q_n = ||A_{N_n,n}(\cdot)||^2$ when the $\boldsymbol{\alpha}$'s and $\boldsymbol{\beta}$'s are the least squares parameters corresponding to the given choice $\underline{\omega}^n, \underline{\sigma}^n, \overline{\omega}^n$, $\overline{\sigma}^n$, and $Q_{n,\max} = \max_{(\underline{\omega}^n, \underline{\sigma}^n, \overline{\omega}^n, \overline{\sigma}^n)} Q_n$. The equality (4.4) implies that

$$
0 \le \|A_{n+1}(\cdot)\|^2 = \left\| A_0(\cdot) - \sum_{j=0}^{n} A_{N_j,j}(\cdot) \right\|^2 = \|A_0(\cdot)\|^2 - \sum_{j=0}^{n} Q_{j,\max}
$$

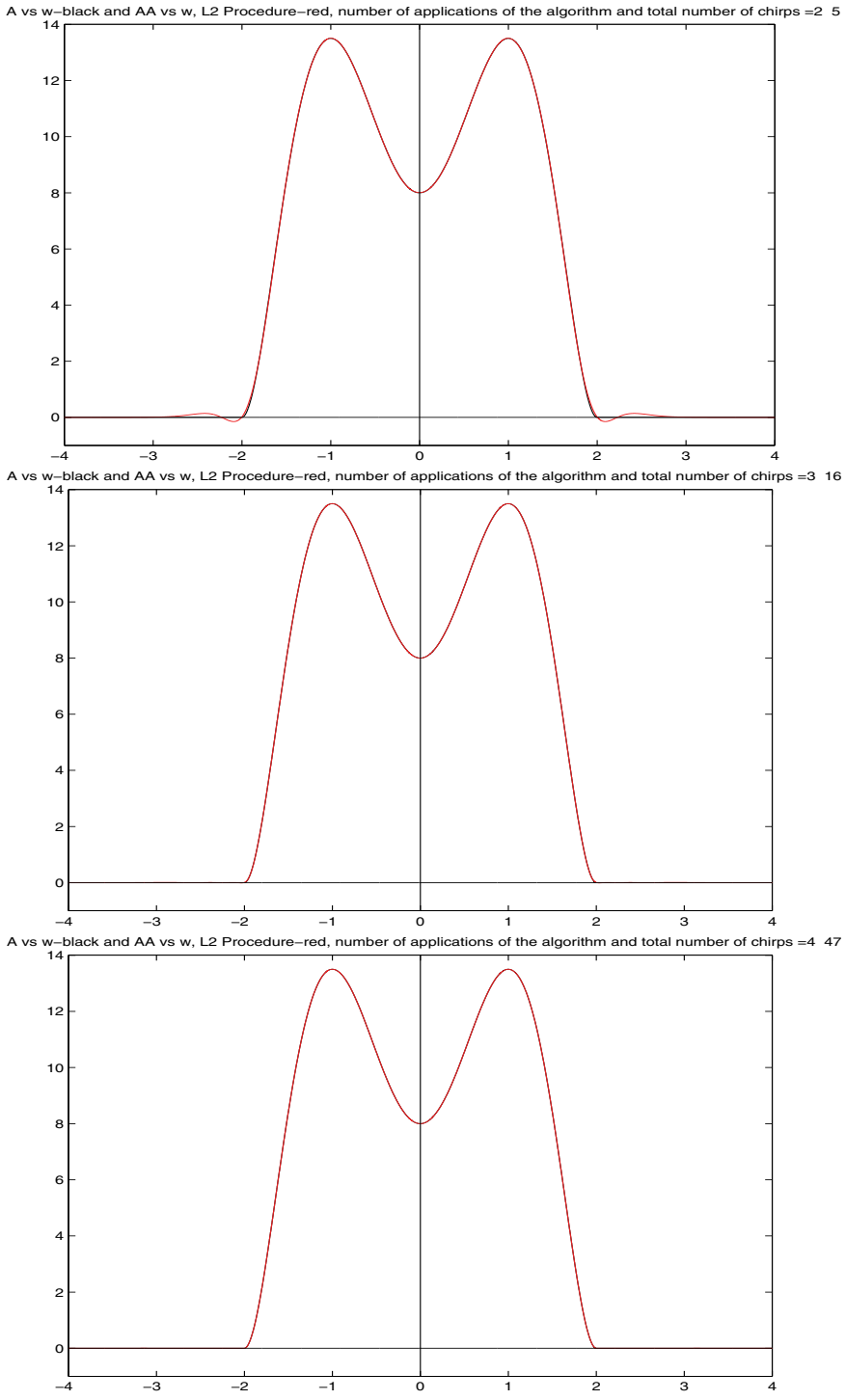

FIG. 4.3. *L^2 hierarchical refinement procedure (top to bottom): Original signal A is in black, and L^2 is in red.*

and provides us with a stopping criterion for the number of applications of the L^2 selection procedure. Specifically, we stop when $||A_{n+1}(\cdot)||^2 \leq \epsilon_{\text{stop}}$, a preassigned tolerance. The stopping criteria for the pointwise selection procedure is equally easy; we stop when

$$\max_{\omega}(A_{n+1}(\omega)) \leq \epsilon_{\text{stop}} \quad \text{and} \quad \min_{\omega}(A_{n+1}(\omega)) \geq -\epsilon_{\text{stop}}. \tag{4.5}$$

Figure 4.3 shows the results of applying the hierarchical L_2 procedure two, three, and four times, respectively. The "black" curves on each figure are the original amplitude, $A(\cdot)$, and the "red"curves are the composite hierarchical Gaussian approximations. In applying the L^2 procedure twice we went from 1 to 5 even Gaussians; in the third application we went from 5 to 16 even Gaussians, and in the fourth application from 16 to 47 even Gaussians. After four applications of the algorithm, the curves become indistinguishable.

5. Two "real-life" applications.

5.1. Chirp decomposition of a noisy signal. A band-limited signal slightly corrupted by white noise is not band limited anymore; however, if we set up the preceding algorithms with a value of Ω being large enough, it may be possible to recover a correct approximation. First, we set up the following data: consider the even amplitude function displaying only two bumps,

$$A(\omega) = \frac{\exp(-a|\omega|^3) - \exp(-b|\omega|^3)}{b-a}, \qquad a = 0.8,\ b = 1.3. \tag{5.1}$$

We couple it with different phase functions of increasing complexity: first, we choose a cubic one, $\phi(\omega) = \frac{\omega^3}{50}$, and then $\phi(\omega) = \pi(1-\exp(-\omega^2))\sin(2\omega)$. The discretization grid is 512 points from $t = -5.1$ to $t = 5.12$. We do not insist on the fact that in case the original phase function ϕ is a polynomial of degree 2, its recovery is exact.

The specificity of these runs is that now, we test the algorithm "blind," which means that we just furnish the collection of sample data for both amplitude and phase, and we look after the maximum numerically. Similarly, the derivatives involved in the parameters' calculation are computed with finite differences. Corresponding results are displayed in the Figures 5.1–5.3. In the left column, one sees the original amplitudes and phases (in black) and their approximations (in blue); in the right one, there are the signal as a function of t (in black), its chirplet approximation (in blue), and finally the absolute error in log-scale. The recovery of the amplitude (5.1) by means of two Gaussians looks very satisfying and errors are noticeable only in the last example, where the original signal is corrupted by *white noise*; the cubic phase (Figure 5.1) is approximated in a correct way by second degree polynomials around the maxima of A which are well identified. Of course, in case A would display several local extrema, a more involved process would have to be set up in order to localize them properly inside the vector of samples. Concerning the sinusoidal phase model (Figure 5.2), its recovery by means of polynomials of degree two is of course rather poor; however, what really matters is that it is correct in the vicinity of maximum points of A, and this is just what happens (see [29]). The behavior of these approximations in the t variable is good, and the absolute error remains reasonable.

Figure 5.3 deals with a more difficult test-case; namely, we set up the amplitude (5.1) with the sinusoidal phase, we perform the inverse Fourier transform to get it as a function of t, and we corrupt the resulting signal with white noise (which makes it not band-limited anymore). This is the data we furnish to the algorithm of

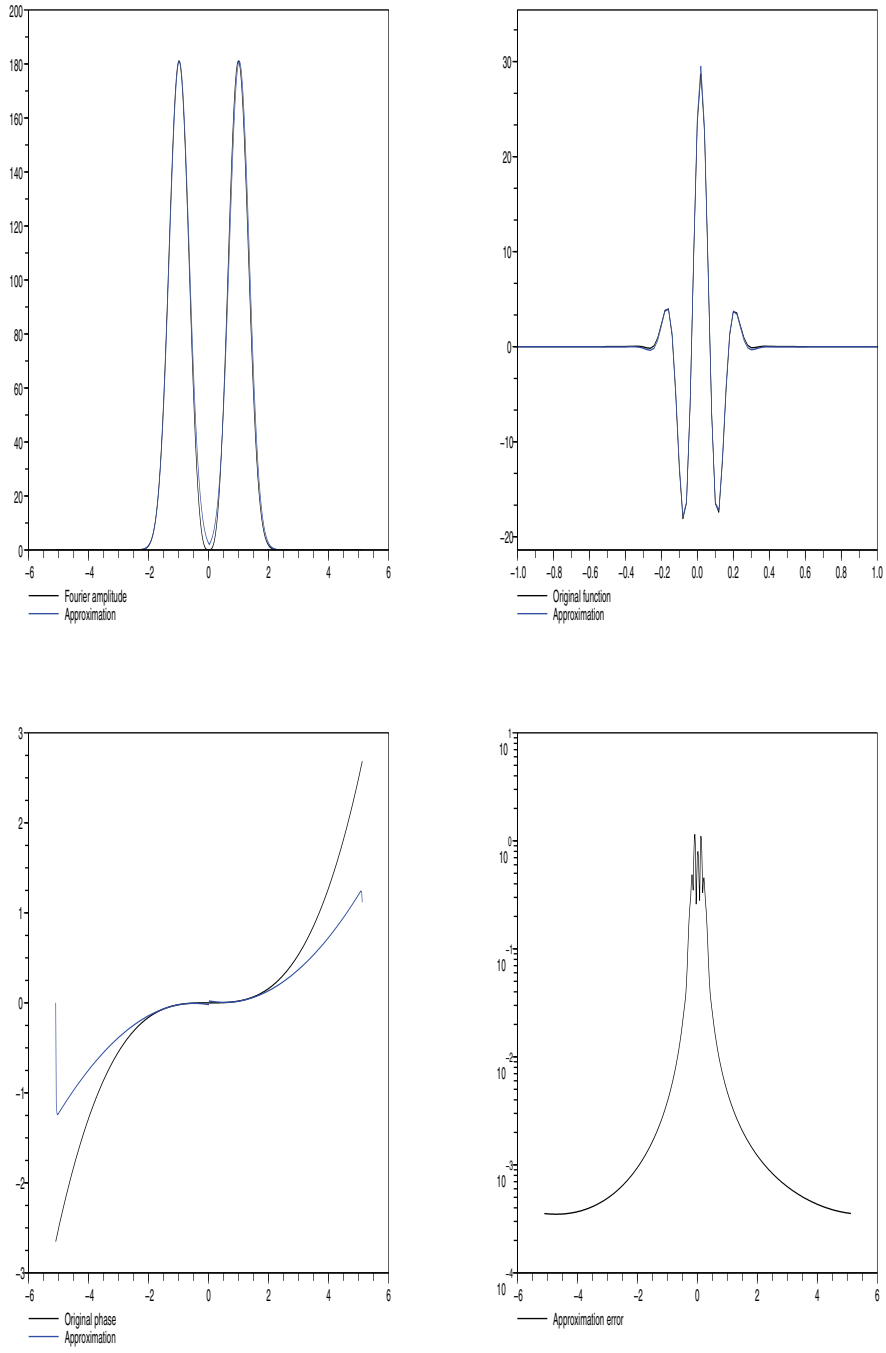

FIG. 5.1. *Pointwise procedure selection on example* (5.1) for cubic phase.

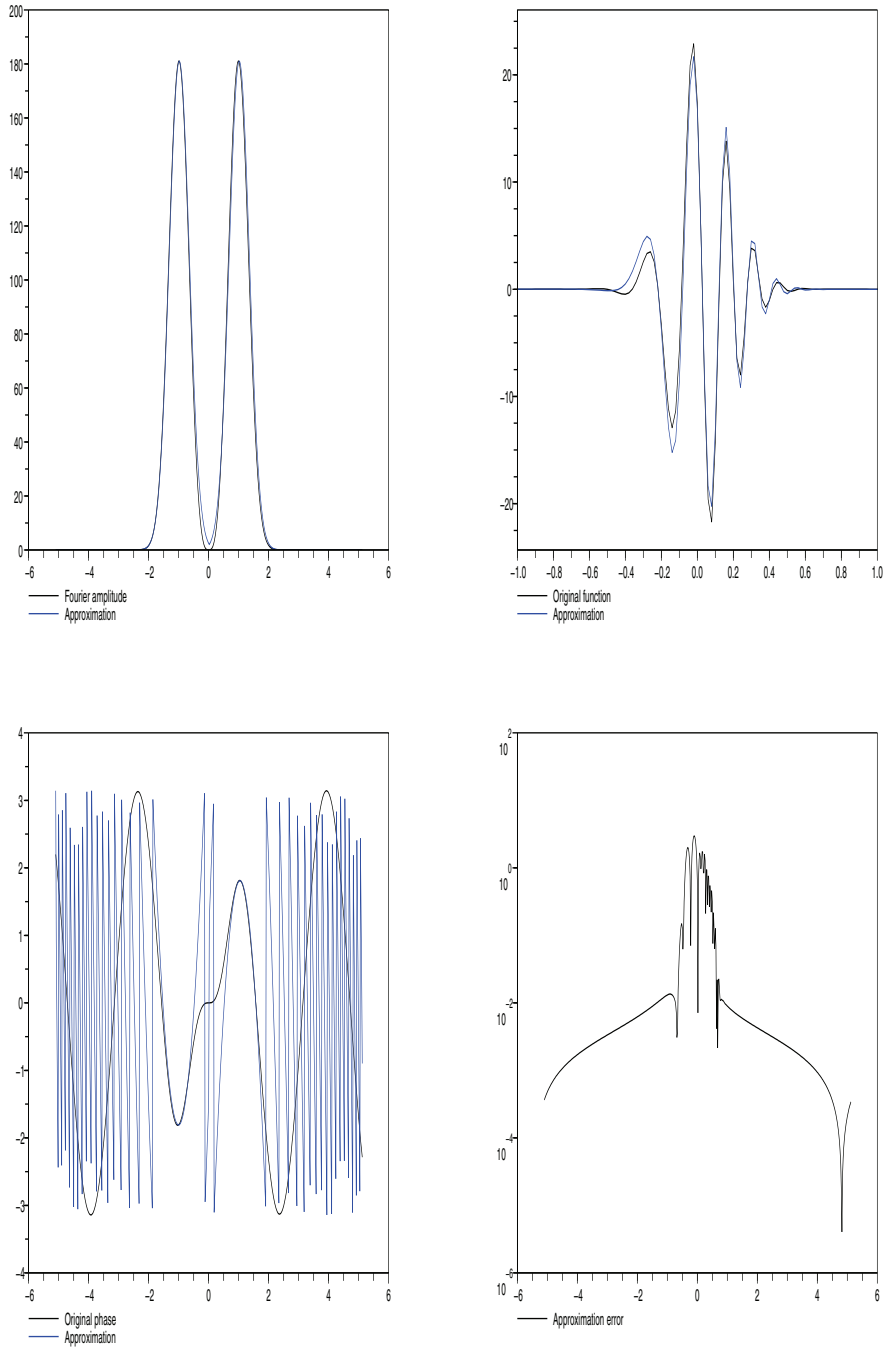

FIG. 5.2. *Pointwise procedure selection on example* (5.1) *for sinusoidal phase.*

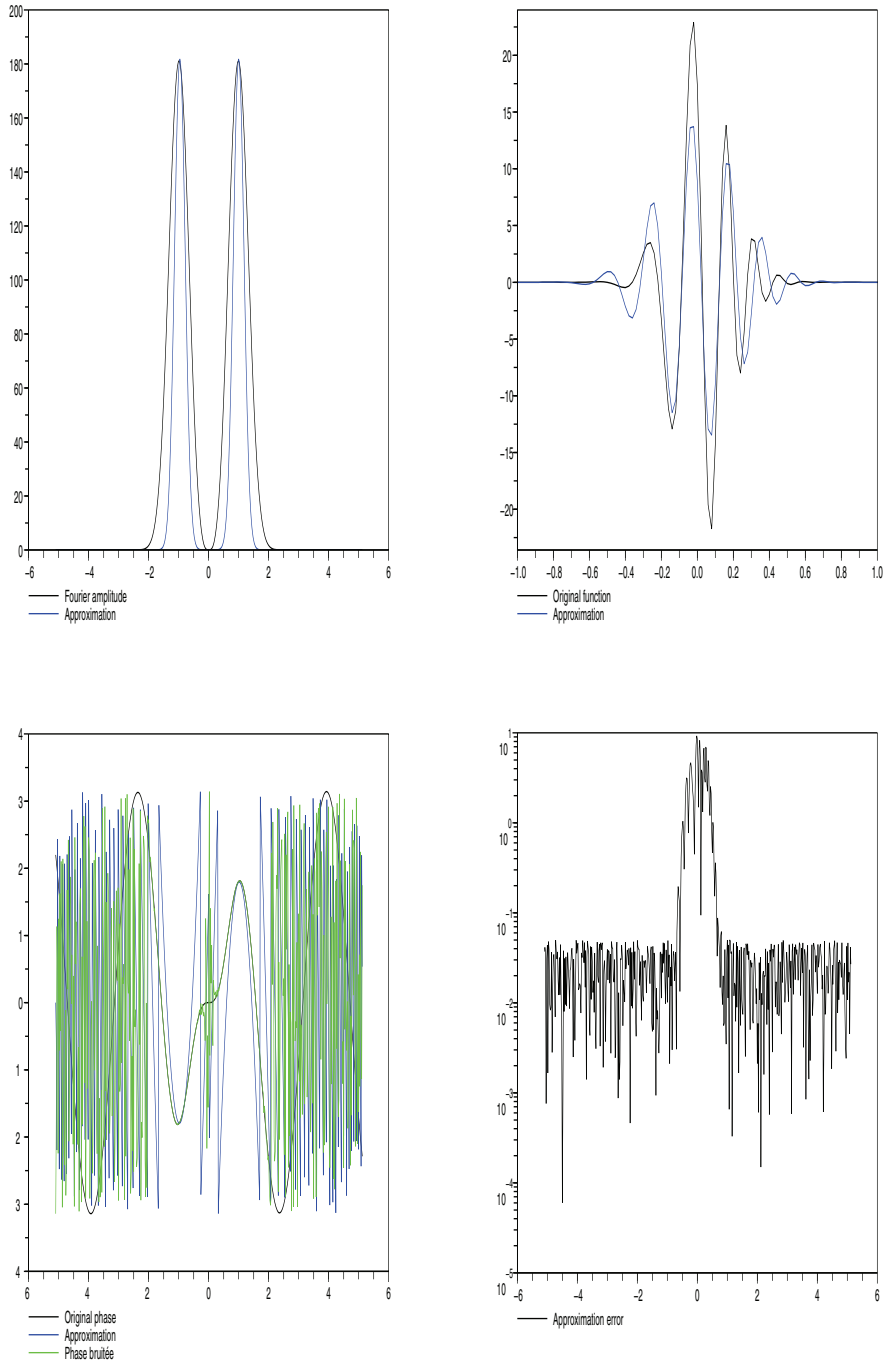

FIG. 5.3. *Pointwise procedure selection on example* (5.1) *slightly corrupted by white noise with sinusoidal phase.*

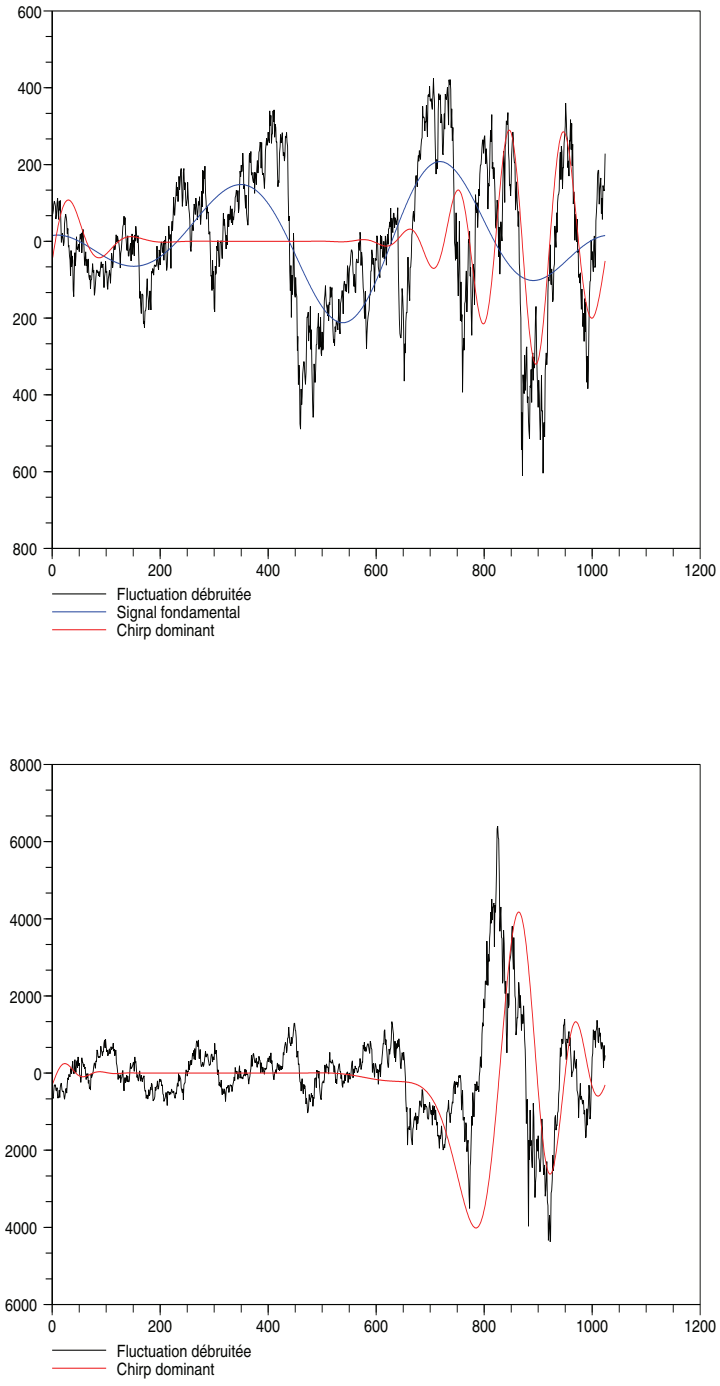

FIG. 5.4. *Chirp patterns in the daily price fluctuations of CAC* 40 *(top) and HSCEI (bottom).*

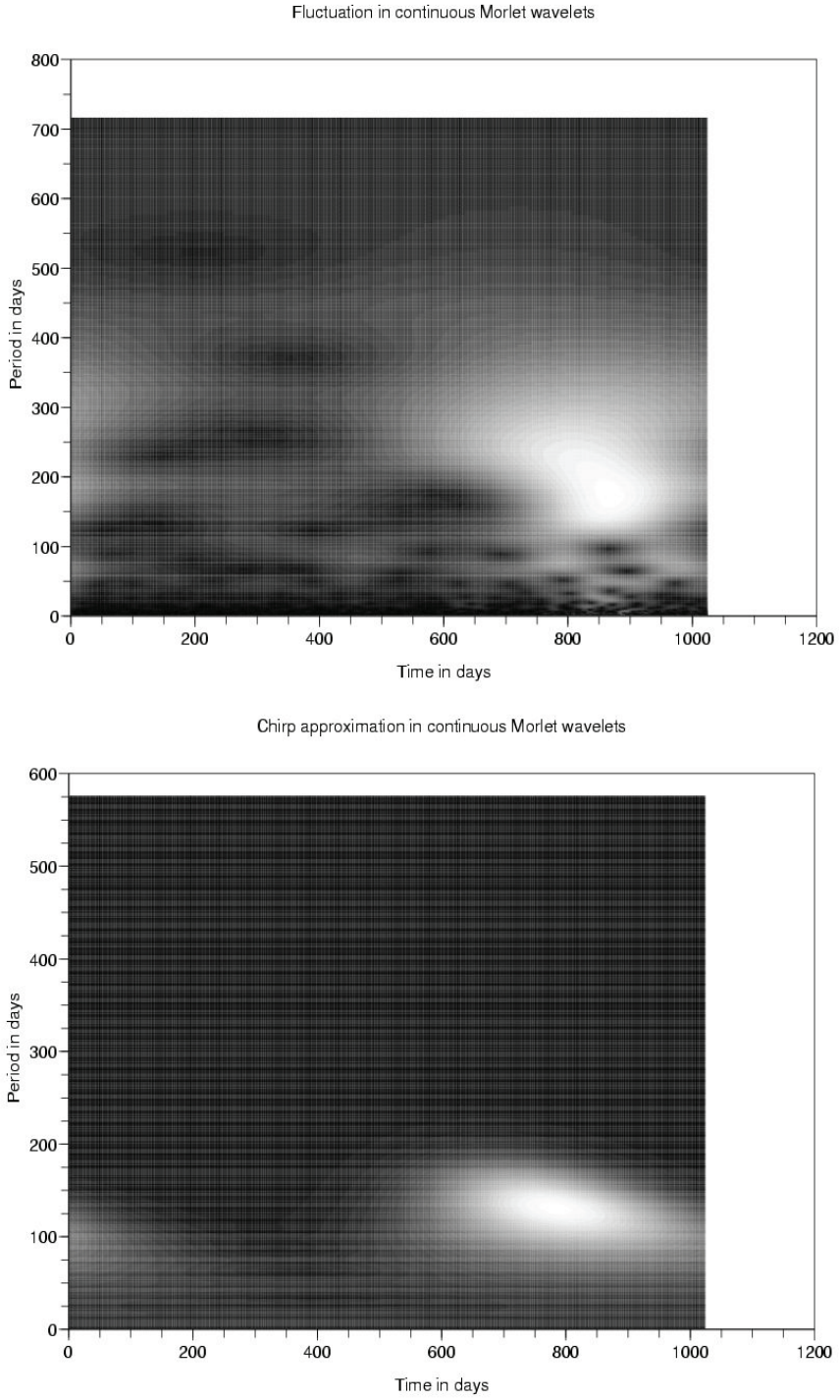

FIG. 5.5. *Continuous Morlet wavelet transform for HSCEI fluctuation (top) and its one chirp approximation (bottom). The abscissa is time, and the ordinate is an instantaneous period; both are expressed in days.*

section 3.1. Obviously, the white noise has to remain small in order not to perturb the maxima of A. What we found is that even this requirement is not enough as the phase function should not be corrupted too much in order to let the algorithm approximate it locally by a parabola correctly. In this case, the absolute recovery error is bigger compared to the noise-free case (see again Figure 5.2).

5.2. Finding chirp patterns in stock market indices. The Hang–Seng Composite Enterprise Index is one of the main indices for the Chinese stock market, and the CAC 40 is the leading index for the Parisian place. We used the corresponding daily price fluctuations on an interval of 1024 days (ending on August 29th, 2008) in order to check in a real-life case whether or not the level of tolerance to noise was acceptable. Clearly, in order to avoid as much as possible spurious local extrema in the Fourier spectrum of the data, we *detrended* the prices by a global least-squares interpolation[3]; polynomials of degree 5 and 6 have been used. Moreover, for the CAC 40 only, the detrended fluctuation was displaying quite a sharp peak corresponding to a periodic cycle. We also used a BP algorithm to remove as much as possible the white noise components of these fluctuations; see [7]. In Figure 5.4, we display the detrended data (in black), the long-lasting periodic cycle of the CAC 40 (in blue), and the chirps we got out of the algorithm of section 3.1. The case of the Chinese index seem to be quite interesting as the recovery seems to be quite sharp. Results on both CAC 40 and HSCEI suggest that a qualitative change in the behavior of blue chip quotations occurred after the ignition of the so-called *subprime crisis/credit crunch* (corresponding to abscissa $t \simeq 700$). Lately, we have become aware of a similar approach which has been published in [4]: the authors seek particular chirps (inspired by earthquake models [16] inside Brazilian iBOVESPA index after convenience detrending and cyclic terms removal. Despite a slightly different methodology ([4] restricts itself to linear trends only, i.e., polynomials of degree 1 in time), they reach the same conclusion: chirps appear before strong downward corrections.[4]

REFERENCES

[1] L. Applebaum, S.D. Howard, S. Searle, and R. Calderbank, *Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery*, Appl. Comput. Harmon. Anal., 26 (2009), pp. 283–290.

[2] P. Borgnat and P. Flandrin, *Time-frequency localization from sparsity constraints*, in Proceedings of the IEEE International Conference on Acoustic Speech and Signal Processing, Las Vegas, NV, 2008, pp. 3785–3788.

[3] A. Bultan, *A four-parameter atomic decomposition of chirplets*, IEEE Trans. Signal Process., 47 (1999), pp. 731–745.

[4] M.A.L. Caetano and T. Yoneyama, *Characterizing abrupt changes in the stock prices using a wavelet decomposition method*, Phys. A, 383 (2007), pp. 519–526.

[5] E. Candès, P.R. Charlton, and H. Helgason, *Detecting highly oscillatory signals by chirplet path pursuit*, Appl. Comput. Harmon. Anal., 24 (2008), pp. 14–40.

[6] E. Chassande-Mottin and P. Flandrin, *On the time-frequency detection of chirps*, Appl. Comput. Harmon. Anal., 24 (2008), pp. 14–40.

[7] S. Shaobing Chen, D.L. Donoho, and M.A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20 (1998), pp. 33–61.

[8] P.G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*, Masson, Paris, 1982.

[3] This construction ensures that the fluctuation has some number of vanishing moments.

[4] We insist on the fact that this paper was submitted for publication at the beginning of September 2008—hence before Lehman Brothers' collapse.

[9] B. Dugnol, G. Fernandez, G. Galiano, and J. Velasco, *On a chirplet transform-based method applied to separating and counting wolf howls*, Signal Processing, 88 (2008), pp. 1817–1826.
[10] P. Flandrin and P. Borgnat, *Sparse time-frequency distribution of chirps from a compressed sensing perspective*, in Proceeding of the 8th IMA International Conference on Mathematics in Signal Processing, Cirencester, UK, 2008, pp. 16–19.
[11] G.B. Folland and A. Sitaram, *The uncertainty principle: A mathematical survey*, J. Fourier Anal. Appl., 3 (1997), pp. 207–238.
[12] J.M. Greenberg, Z. Wang, and J. Li, *New approaches for chirplet approximations*, IEEE Trans. Signal Process., 55 (2007), pp. 734–741.
[13] R. Gribonval, *Fast ridge pursuit with multiscale dictionary of Gaussian chirps*, IEEE Trans. Signal Process., 49 (2001), pp. 994–1001.
[14] X. Huo and D. Donoho, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inform. Theory, 47 (2001), pp. 2845–2862.
[15] F. Jaillet and B. Torrésani, *Adaptive time-frequency representation for sound analysis and processing*, in Proceeding of the Joint Symposium SFA/DAGA (Strasbourg, 2004), Vol. 2, 2004, pp. 711–712.
[16] A. Johansen, *Characterization of large price variations in financial markets*, Phys. A, 324 (2003), pp. 157–166.
[17] C.J. Kicey and C.J. Lennard, *Unique reconstruction of band-limited signals by a Mallat–Zhong wavelet transform algorithm*, J. Fourier Anal. Appl., 3 (1997) pp. 63–82.
[18] A.V. Knyzaev and A.L. Skorokhodov, *The rate of convergence of steepest descent in a Euclidian norm*, USSR Comp. Maths. Math. Phys., 28 (1988), pp. 195–196.
[19] H. Kwok and D. Jones, *Improved FM demodulation in a fading environment*, in Proceedings of the IEEE Conference on Time-Frequency and Time-Scale Analysis, Paris, France, 1996, pp. 9–12.
[20] J. Li and P. Stoica, *Efficient mixed-spectrum estimation with applications to target feature extraction*, IEEE Trans. Signal Process., 44 (1996), pp. 281–295.
[21] J. Li, D. Zheng, and P. Stoica, *Angle and waveform estimation via RELAX*, IEEE Trans. Aerospace and Electronic Systems, 33 (1977), pp. 1077–1087.
[22] G. Lopez-Risueno, J. Grajal, and O.A. Yeste-Ojeda, *Signal detection and estimation using atomic decomposition and information theoretic criteria*, in Proceedings of the IEEE International Conference on Acoustic Speech and Signal Processing, 2004; available online from http://cat.inist.fr/?aModele=afficheN&cpsidt=17610286.
[23] S. Mallat and Z. Zhong, *Matching pursuit with time-frequency dictionaries*, IEEE Trans. Signal Process., 41 (1993), pp. 3397–3415.
[24] S. Mann and S. Haykin, *The chirplet transform: Physical considerations*, IEEE Trans. Signal Process., 43 (1995), pp. 2745–2761.
[25] K. Ni, S. Datta, S. Roudenko, and D. Cochran, *Image Reconstruction by Deterministic Compressive Sensing*, preprint; available online from http://math.asu.edu/~svetlana/Papers.html.
[26] J.C. O'Neill and P. Flandrin, *Chirp hunting*, in Proceedings of the IEEE International Symposium on Time-Frequency and Time-Scale Analysis, 1998, pp. 425–428.
[27] J.C. O'Neill, P. Flandrin, and W.C. Karl, *Sparse Representations with Chirplets via Maximum Likelihood Estimation*, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.33.2204.
[28] A. Olevskii and A. Ulanovskii, *Universal sampling and interpolation of band-limited signals*, Geom. Funct. Anal., 18 (2008), pp. 1029–1052.
[29] A.V. Openheim and J. S. Lim, *The importance of phase in signals*, Proc. IEEE, 69 (1981), pp. 529–541.
[30] W. Rudin, *Analyse réelle et complexe*, Masson, Paris, 1975, Chapter 19.
[31] D. Slepian, *On bandwidth*, Proc. IEEE, 64 (1976), pp. 292–300.
[32] Th. Strohmer, *On discrete band-limited signal extrapolation*, Contemp. Math., 190 (1995), pp. 323–337.
[33] Q. Yin, H. Qian, and A. Feng, *A fast refinement for adaptive Gaussian chirplet decomposition*, IEEE Trans. Signal Process., 50 (2002), pp. 1298–1306.