Time-Frequency Energy Distributions Meet Compressed Sensing

Patrick Flandrin, Pierre Borgnat

Time-Frequency Energy Distributions Meet Compressed Sensing

Patrick Flandrin, *Fellow, IEEE* and Pierre Borgnat, *Member, IEEE*

***Abstract*—In the case of multicomponent signals with amplitude and frequency modulations, the idealized representation which consists of weighted trajectories on the time-frequency (TF) plane, is intrinsically sparse. Recent advances in optimal recovery from sparsity constraints thus suggest to revisit the issue of TF localization by exploiting sparsity, as adapted to the specific context of (quadratic) TF distributions. Based on classical results in TF analysis, it is argued that the relevant information is mostly concentrated in a restricted subset of Fourier coefficients of the Wigner-Ville distribution neighbouring the origin of the ambiguity plane. Using this incomplete information as the primary constraint, the desired distribution follows as the minimum ℓ_1-norm solution in the transformed TF domain. Possibilities and limitations of the approach are demonstrated via controlled numerical experiments, its performance is assessed in various configurations and the results are compared with standard techniques. It is shown that improved representations can be obtained, though at a computational cost which is significantly increased.**

***Index Terms*—time-frequency, localization, sparsity.**

EDICS Category: SSP-NSSP

I. Introduction

CHIRPS (i.e., transient amplitude and frequency modulated (AM-FM) signals) are ubiquitous in nature and man-made systems. They are commonly encountered in a variety of fields ranging from audio (speech, music) to wave physics (whistling atmospherics, gravitational waves), from engineering (radar, sonar) to biology and medicine (EEG epilectic seizures, EMG uterine contractions), or even in mathematics (Weierstrass and Riemann functions), see, e.g., [3] and references therein. From the point of view of exploratory data analysis, a fine and non parametric characterization of chirps is often needed, e.g., as a hint for some further modeling. Since chirps are waveforms whose structure explicitly involves time-varying properties with respect to both amplitude and spectral content, they naturally call for time-frequency (TF) analysis tools. Because of the Fourier uncertainty between time and frequency, any TF approach is however faced with difficulties which become all the more challenging as signals are of a shorter duration. Albeit standard in its motivation, this problem has not yet received its ultimate solution, and the purpose of this paper is to revisit it in a new way that combines basics of "classical" TF analysis with recent advances from Compressed Sensing (CS) (for a general overview of CS, the interested reader is referred to [4], [5], as well as to the resources webpage [6]).

P. Flandrin and P. Borgnat are with Université de Lyon, Physics Department (UMR 5672 CNRS) of École Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07 France. Tel.: +33 (0)4 72 72 81 60; Fax: +33 (0)4 72 72 80 80; E-mail: {flandrin,pborgnat}@ens-lyon.fr. Part of this paper was first presented at the IEEE Int. Conf. on Acoust. Speech and Signal Proc. ICASSP-08 (Las Vegas, NV) [1] and at the 8th IMA Int. Conf. on Math. in Signal Proc. (Cirencester, UK) [2].

It is worth stressing the fact that the approach reported here differs in several respects from related attempts. First, as compared to Gabor-based approaches (see, e.g., [7]) which assume sparsity in *linear* representations, we make here the choice of working with *quadratic* energy distributions as TF representations, the rationale being to favor structured components in the form of localized chirps along TF trajectories rather than independent atoms. Second, whereas most "classical" CS-based methods involve random measurements, a deterministic approach will be followed here for the selection of a small subset of observed values used as constraints. For sake of incoherence between the targeted representation and its measurements, such a selection will be done in the Fourier domain in the spirit of the emblematic examples of phantom recontruction from Fourier slices discussed in [9] or of "Compressed Sensing MRI" [11], [12], thus contrasting with other possibilities of deterministic sensing [13]. A major difference exists however with such Fourier-based CS approaches since, as it will be made explicit in Sect. IV-B2, the addressed problem will be shown to differ from a classical *reconstruction* by moving to the *construction* of a TF distribution which would not be attainable otherwise, regardless of the number of measurements in the Fourier domain. Finally, whereas specific improvements to Fourier-like TF methods could be obtained by means of model-based approaches, we will adopt here a model-free perspective, the only assumption being that the representation which of interest for is intrinsically sparse in the TF domain since the energy of a chirp is expected to be essentially distributed along a 1D curve of the 2D plane.

More precisely, the paper is organized as follows. In Section II, the issue of chirp localization is addressed and classical TF solutions are briefly recalled. Considering that localized TF distributions of chirps are sparse in the TF plane, Section III gives the rationale for revisiting the problem from a CS perspective. This is detailed further in Section IV where the effectiveness and performance of the approach are discussed, in particular with respect to the selection of Fourier samples for which a rule-of-thumb criterion is proposed. Whereas this paper is not focused on the algorithmic part of the method, some computational issues are addressed, and limitations as well as possible extensions are finally discussed in Section V.

II. Time-frequency localization

A. AM-FM signals as time-frequency trajectories

If we consider a signal made of the superimposition of a finite number of AM-FM components:

$$x(t) = \sum_{k=1}^{K} a_k(t)\, e^{i\varphi_k(t)},$$

it is natural to attach to it an *idealized* TF distribution (TFD) $\rho_x(t,f)$ which essentially distributes the total energy along TF trajectories according to:

$$\rho_x(t,f) = \sum_{k=1}^{K} a_k^2(t)\, \delta\left(f - \dot{\varphi}_k(t)/2\pi\right). \qquad (1)$$

In such a picture, each component is characterized at each time instant by essentially one instantaneous frequency (which, in a first approximation, can be identified to the phase derivative), weighted by the corresponding instantaneous power.

Except for very special cases, there is no general methodology to automatically get a distribution as in (1). In the case of a single component ($K = 1$), it is well-known [14] that a perfect localization can be attained for pure FM signals with a linear modulation ($a_1(t) = 1$ and $\dot{\varphi}_1(t) = 2\pi(f_0 + \alpha\, t)$) by using the Wigner-Ville Distribution (WVD):

$$W_x(t,f) = \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right)\, x^*\left(t - \frac{\tau}{2}\right)\, e^{-i2\pi f\tau}\, d\tau. \qquad (2)$$

Although this property can be extended to some forms of nonlinear FMs (e.g., Bertrands' distributions for power-laws [15]), it is generally at the expense of a substantially increased complexity in the definition (and the computation) of the distributions, with furthermore the limitation of being adapted to some specific type of FM only and to not extend to multicomponent situations. For this last point, the well-known drawback of energy distributions is to obey a quadratic superposition principle which creates cross-terms in between any two components of a signal, and thus significantly reduces the readability of Wigner-type distributions [16].

B. Classical techniques of TF localization

The aforementioned difficulties have led to many developments during the last 20 years but, unfortunately, since both localization and creation of cross-terms result from the very same mechanism [17], it turns out that trying to impose simultaneously localization and cross-terms reduction is faced with a trade-off that can be viewed as a form of time-frequency uncertainty principle. The simplest way to understand where this trade-off comes from (and how to manage it) is to interpret the WVD in its 2D Fourier transform domain.

By definition, the WVD admits a 2D Fourier transform which is referred to as the *ambiguity function* (AF) and reads

$$A_x(\xi,\tau) = \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right)\, x^*\left(t - \frac{\tau}{2}\right)\, e^{i2\pi\xi t}\, dt.$$

If we introduce the TF shift operator $\mathbf{T}_{\xi,\tau}$ which acts on signals $x(t) \in L^2(\mathbb{R})$ as

$$\left(\mathbf{T}_{\xi,\tau} x\right)(t) := x(t-\tau)\, e^{-i2\pi\xi(t-\tau/2)},$$

we readily get that $A_x(\xi,\tau) = \langle x, \mathbf{T}_{\xi,\tau} x\rangle$, i.e., that the AF is nothing but the inner product between the analyzed signal and its TF shifted versions. By construction, the AF can thus be viewed as a TF correlation function and, as such, it exhibits most properties of a correlation function, including hermitian symmetry and the fact that its modulus is maximum at the origin. Moreover, in the case of multicomponent signals, the total AF consists of both *auto-components* neighbouring the origin of the plane and *cross-components* mostly located at a TF distance from the origin which directly depends on the TF separation between the individual components and that are the Fourier images of the undesired cross-terms in the TF plane. This observation early prompted [18] to propose improvements upon the WVD by weighting the AF around the origin of the plane prior to applying an inverse 2D Fourier transform: the more restricted the weighting domain, the more effective the cross-terms suppression but, at the same time, the more decreased the TF localization. This procedure (which defines Cohen's class [14] on geometrical grounds and can be improved by adapting the weighting domain to the signal structure [19]) proved reasonably effective, but it basically faces a trade-off between cross-terms reduction and localization. This motivated the search for improved techniques and, amongst the other approaches that have been proposed, the one that is referred to as *reassignment* plays a prominent role. The rationale behind reassignment (see [20] for a comprehensive presentation of reassignment techniques) is that it basically exploits the phase information that is usually discarded in simple quadratic distributions such as spectrograms (or scalograms). In this respect, *reassigned spectrograms* (RSP) proved in particular extremely efficient to approach (1) and will thus be used in Section IV for the sake of comparison.

III. Localization from sparsity constraints

If the analyzed signal is given in discrete-time and supposed to be of dimension N in time, its TFD is of dimension N^2 when computed over N frequency bins. However, assuming that $K \ll N$, i.e., that the number of components is much smaller than the dimension of the signal, the targeted TFD which is supposed to satisfy (1) is distributed over the plane in a very sparse way, with only K 1D trajectories where at most KN values are expected to be non-zero. Imposing such a sparsity is therefore a new way of approaching the problem.

A. Principle

Based on the arguments recalled in Section II-B, the principle of the proposed sparsity-based approach to TF localization is thus very simple: it simply consists of selecting a suitable collection of AF samples neighbouring the origin of the plane in a given domain $\Omega(\xi,\tau)$ and searching for the sparsest TFD ρ such that its 2D Fourier transform $\mathcal{F}\{\rho\}$ coincides with the original AF over Ω. As has been mentioned previously, masking AF around the origin has the desired effect of reducing cross-terms, but at the expense of degrading localization if a crude Fourier inversion is performed without any constraint. Selecting a few samples only of the AF around

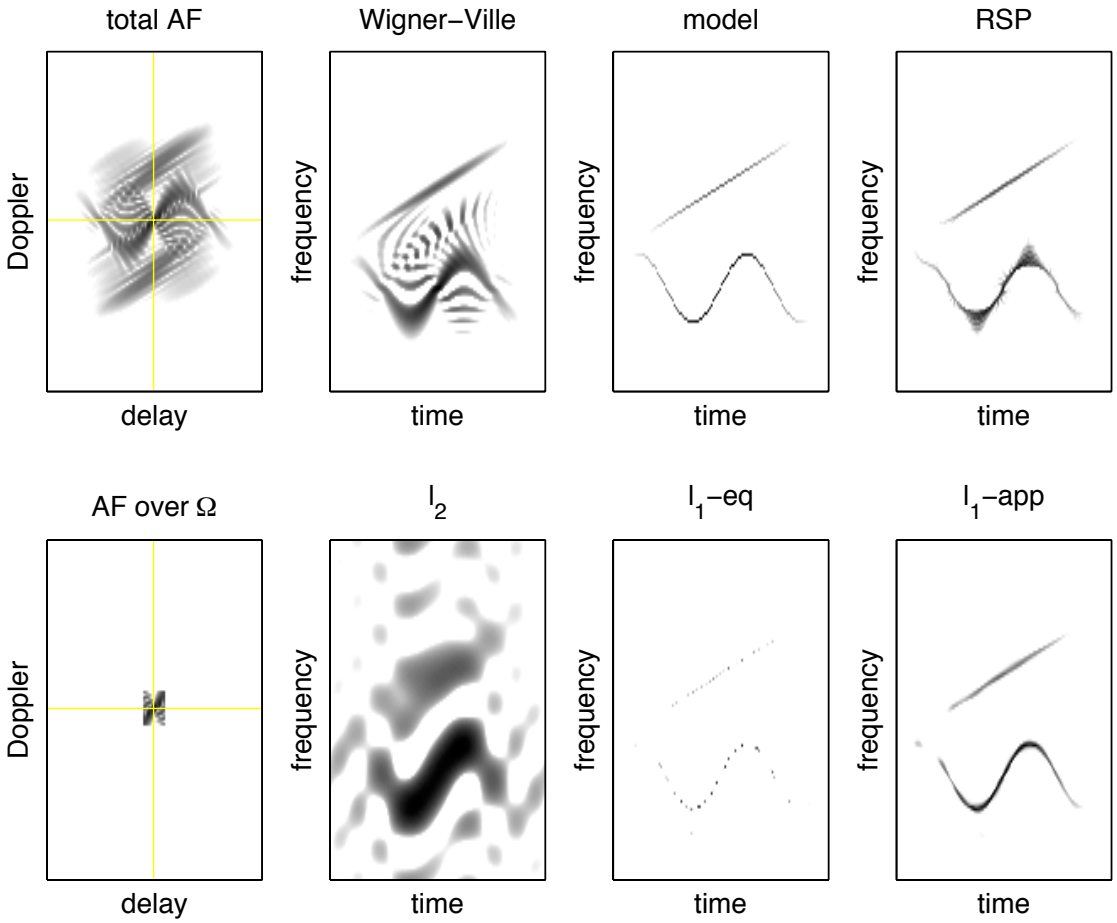

Fig. 1. *Example with a 128 points 2-component signal* — Different TFDs are displayed in the case of a 128 points signal whose TF model (1) is given in the top row, in between the Wigner-Ville Distribution (WVD) and the reassigned spectrogram (RSP) that has been "optimized" as explained in Sect. IV-A. The corresponding total ambiguity function (AF) of size 128×128 (defined as the 2D Fourier transform of the WVD) is given on the left of the top row, with its restriction to a domain Ω consisting of 13×13 samples neighbouring the origin of the plane on the left of the bottom row. The rest of the bottom row presents the results obtained from optimizations based on the only knowledge of this restriction of the AF, with the exact minimum ℓ_1-norm solution according to (3) in between the minimum ℓ_2-norm solution and the approximate minimum ℓ_1-norm solution according to (4) with $\epsilon = 0.05\,\|x\|_2$. For all diagrams, amplitudes are grey coded logarithmically, with a dynamic range of 18 dB.

the origin is therefore the first ingredient for ensuring cross-terms reduction, and it is the further assumption of few chirp components that is instrumental for guaranteeing TF localization. In this respect, the recourse to CS algorithms can be viewed as a form of constrained AF extrapolation.

B. Constraints

Looking for a perfectly localized solution such as (1) would require minimizing the total number of non-zero coefficients, i.e., the ℓ_0-"norm" of the TFD. While this turns out not to be practicable from a computational viewpoint, a series of recent works (see, e.g., [8], [9], [10], and [6] for a comprehensive list of references) have shown that a near-optimal solution can be attained at a sensibly more affordable cost by minimizing the ℓ_1-norm, reducing the problem to the solving of a linear program. It is this technique which is proposed to be followed here, the desired localized TFD $\rho_x(t,f)$ being therefore the solution of the constrained minimization problem:

$$\rho_x = \arg\min_{\rho} \|\rho\|_1 \,;\, \mathcal{F}\{\rho\} - A_x = 0|_{(\xi,\tau)\in\Omega}\,. \qquad (3)$$

The primary constraint which is given by (3) imposes a strict equality over Ω in the AF domain. This however can be relaxed [10] according to

$$\rho_x = \arg\min_{\rho} \|\rho\|_1 \,;\, \|\mathcal{F}\{\rho\} - A_x\|_2 \le \epsilon|_{(\xi,\tau)\in\Omega}\,, \qquad (4)$$

where ϵ is a user-specified bound. Both possibilities (3) and (4) will be considered in the following, with the corresponding ℓ_1 solutions referred to as, respectively, ℓ_1-eq (for ℓ_1 with equality) and ℓ_1-app (for ℓ_1 with approximation).

One can remark that, the present approach being based on optimization, it is possible to easily impose further constraints besides (3) or (4). This has not been followed in this work,

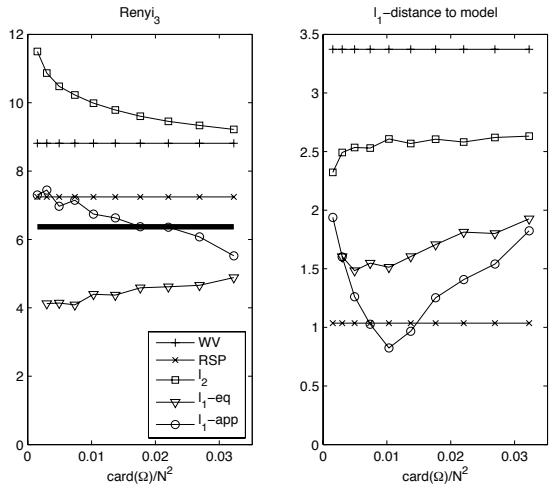

Fig. 2. *Rényi entropy and ℓ_1-distance* — Using the same signal as in Figure 1, the localization properties and proximity from the model of the different TFDs are quantified in terms of the Rényi entropy of order 3 (left, with the model entropy in thick black line) and ℓ_1-distance (right), as a function of the relative number of AF samples $\mathrm{card}(\Omega)/N^2$ used in the optimizations.

but it should certainly be interesting to explore further such a possibility (with respect, e.g., to regularity, marginals, ...).

IV. CS-BASED TF ANALYSIS IN ACTION

The feasability of the method has been tested on simple, yet informative examples. Since the purpose of this paper is not to contribute in a new way to the solving of optimization problems such as those stated in (3) and (4), the reported examples are based on a straightforward application of existing solutions. For the sake of completeness, basics of the corresponding methods are however briefly recalled in the Appendix. All the computations presented in this section have been made in MATLAB, with the TIME-FREQUENCY TOOLBOX [21] for the TF computations and the ℓ_1-MAGIC TOOLBOX [22] for the optimization, but alternative algorithmic procedures have been considered as well, with similar results: this will be detailed further in Section IV-D.

A. A first comparison

Figure 1 illustrates the principle of the proposed approach and compares different TFDs in the case of a 128 points waveform made of the superimposition of a linear and of a sinusoidal FM, both modulated in amplitude with a Gaussian. In this example, the different optimizations have been based on the only knowledge of the 13×13 Fourier samples of the WVD neighbouring the origin of the AF plane, i.e., on a subset of about 1% only of the total number of AF coefficients. With so few AF coefficients, it is clear that a crude Fourier inversion (which corresponds to the ℓ_2-norm solution of the optimization problem) ends up with a very poor TFD. From a qualitative point of view, it then turns out from this figure that the minimum ℓ_1-norm solutions (either exact, according to (3) or approximate, according to (4)) are much more effective in terms of localization, even as compared to the reassigned spectogram which is known to usually behave best for this kind of signal. It has to be noted that, since a reassigned spectrogram depends on the choice of a window length, experiments have been conducted with various lengths and the reported quantitative results correspond to the best performance that has been obtained. In practice, such an optimization cannot be achieved, but making use of this "optimum" reassigned spectrogram has been chosen so as to get a form of bounds on the achievable performance (such bounds are likely not to be attained in real situations). Comparing the exact and approximate ℓ_1 solutions, it also appears that the former is in some sense "too sparse" as compared to the latter, with energy contributions well localized on the TF trajectories, but in a discontinuous, spiky fashion.

B. Selection of Fourier samples

The result reported in Figure 1 can be thought of as very peculiar in the sense that it heavily relies on the specific choice of a domain, regarding area (why 13×13 Fourier samples?) as well as shape (why a fixed square domain?). All the appreciations commented above can then be quantified further in terms of both the achieved performance with respect to the actual model, and the influence of the AF domain from which the optimization is conducted.

1) Heisenberg cardinality: Concerning the cardinality of the domain (first assumed to be square), the result is reported in Figure 2 which displays a localization measure (the Rényi entropy of order 3, which has proved to be a good measure of localization for energy TFDs [23]) together with the ℓ_1-distance to the model, both as functions of the relative sparsity measure $\mathrm{card}(\Omega)/N^2$. What is revealed by this figure is that both minimum ℓ_1-norm solutions are generally better localized (i.e., have a smaller Rényi entropy) than the other considered TFDs, with even an entropy that might be smaller than the model one whose value is in this case 6.37. In particular, the exact ℓ_1 solution has always the minimum entropy but, as evidenced by Figure 1, this is due to an oversparse, discontinuous structure which results in a larger distance to the actual model. The evolution of this distance shows that the best behavior is obtained with the approximate minimum ℓ_1-norm solution, the minimum being obtained when using only about 1% of the N^2 AF samples. The small value of this sparsity measure is indeed a consequence of the fact that, as it will be justified further in Sect. IV-B2, the required number of measurements in the Fourier domain scales as $\mathrm{card}(\Omega) \approx N$. Additionnally, it is worth noticing that, while the WVD we started with is known to attain negative values and whereas no positivity constraint has been imposed, the minimum ℓ_1-norm solutions happen to be almost positive.

A qualitative interpretation of why a cardinality of the order of the signal length leads to the best result can come from an uncertainty argument. Indeed, AFs (as TFDs) are known to obey uncertainty principles which prevent them to be pointwise localized in their respective planes [24]. Auto-terms of a multicomponent signal are thus expected to exist over a domain (neighbouring the origin of the plane, see Section II-B) whose area is at least of the order of a "Heisenberg cell",

i.e., roughly N if the AF of a N points signal is computed over a $N \times N$ grid. Most of the information in such a cell being coded by the modulus and the phase of the AF, there is no point in enlarging further the domain, since the main effect of such an enlargement would be to possibly include undesired cross-terms influences in the constraints.

2) Oracle: Some further appreciation can be gained from a comparison with an "oracle", based on the assumed knowledge of the ideal distribution. In this case, the linear program (4) is run exactly as in Figure 1, while replacing *mutatis mutandis* the WVD by the TF model (1). This is illustrated in Figure 3, where an optimum cardinality (namely $|\Omega|_* \approx N^2/100 \approx N$) is shown to exist, at least qualitatively, for trading off auto-terms localization and cross-terms reduction.

The compared evolutions of the ℓ_1 solutions derived from the WVD and the model are clearly similar in a first regime, whereas they diverge after the domain cardinality has attained some critical value, the higher localization expected for auto-terms being hampered by the off-spring of cross-terms that appear as spikes in the representation due to their oscillatory nature. If no such spiky terms would be present, a useful criterion for choosing the cardinality of Ω could be the entropy: the smaller the entropy, the more localized the TFD. In order to prevent possible values of entropy due the spiky structure stemming from undesired additional cross-terms, the sole entropy has to be penalized in some way. This has been investigated in [2] (but not reported here) by using total variation as such a penalty function, ending up with the identification of an optimum cardinality $|\Omega|_*$ that is consistent with the results displayed in Figures 2 and 3.

At this point, it might be worth stressing the fact that the approach described here differs from a *reconstruction* problem (as those considered, e.g., in [9], [11], [12]) in the sense that there would be no point in exactly recovering the whole WVD from a limited set of measurements (in fact, such a perfect reconstruction is of course obtained when the entire ambiguity plane is chosen as the measurement domain, see the far right solutions of Figure 3). The situation is much more that of a *construction* problem in which is created some idealized object which does not exist *per se* prior optimization. From the oracle perspective, a suitable selection of AF samples should thus explicitly correspond to a subset of the Fourier transform of the ideal localized TFD, and no more of the WVD. Based on fundamental CS results for partial Fourier coefficients (see [9] and the review in [5]), the cardinality of the domain Ω should be therefore $\mathrm{card}(\Omega) = \mathcal{O}\left(KN\log(N^2)\right)$ for the recovery of K AM-FM trajectories of N points each in a TF domain of size N^2. In practice, the typical value obtained from this formula seems to be higher that the previous findings reported in Figure 2 and those based on the oracle. This is first related to the *construction* nature of our problem, when the CS results are given for a *reconstruction* problem. Second, a lower limit for a ℓ_1 minimization problem to find the sparsest solution is given by a phase transition behaviour, as discussed in [5]. In that case, the cardinality of the domain Ω could be as low as $\mathrm{card}(\Omega) = \mathcal{O}\left(KN\log(N/K)\right)$ in this problem. With the numerical values used in Figure 1, this leads to a magnitude of order $10KN$ (up to a constant of $\mathcal{O}(1)$), that is of the same order of magnitude as the Heisenberg cardinality discussed before.

3) Domain geometry: Up to now, emphasis has been put on the area of a domain Ω assumed to be square. This is of course an unnecessary restriction that has to be questioned. To this end, Figure 4 displays ℓ_1 distances to the model, as in Figure 2, with a domain Ω whose area is kept fixed and equal to its "optimum" value 169, but with a rectangular or elliptic shape with a varying aspect ratio. The latter being computed as "$r =$ Doppler semi-axis/delay semi-axis", it turns out that the best performance is obtained when the selected ambiguity domain is slightly elongated along the delay axis, with almost uniformly improved results within the range $1/3 \lesssim r \lesssim 1$ (i.e., equivalently, $-0.4 \lesssim \log_{10}(r) \lesssim 0$). While such numerical values are likely not to be universal and might depend on the analyzed signal, the interpretation is believed to be general in the sense that aspect ratios $r < 1$ tend to favor TF coherence in the time direction, a feature that is characteristic of most chirps which are naturally considered as "frequencies that vary with time". Would the signal TF structure be organized differently as, e.g., in dispersive systems where frequency-dependent epochs are the most relevant signatures, aspect ratios greater than one would be better. When some anisotropy is expected in a TFD, the recipe is that this should be taken into account as a prior when choosing Ω.

A further remark can be done in the same direction, which points to a basic difference between the approaches based on linear decompositions and the quadratic one proposed here. If sparsity was assumed to exist with respect to some dictionary of TF atoms, their identification (via matching or basis pursuit) would end up with a collection of elementary components (typically, Gabor logons) that could be located along the TF trajectories of interest but as independent components rather than as a globally coherent structure. From a quadratic TF perspective [17], and in the spirit of the "cross-terms deleted representations" [25], this corresponds to a situation where the *inner* interference terms that guarantee the coherently organized TF structure of chirps are just ignored. In this respect, a flexible improvement can come from the taking into account of such terms that, by nature, have short (TF) distance interactions [26]. This is also what is done in the present approach since, by nature, inner interference terms mix up with auto-terms in the vicinity of the origin of the ambiguity plane, without any clear-cut separation. They are thus automatically taken into account when restricting Ω to an Heisenberg cell, with a representation all the more satisfactory in terms of chirp continuity as the geometry of the domain is adapted to the TF correlation structure of the analyzed signal.

This last remark prompts to look for even more adaptivity in the domain geometry, beyond convex domains whose main axes would necessarily coincide with those of the ambiguity plane. At first sight, it might be expected that the use of adapted kernels (as proposed, e.g., in [19]) would prove useful. This however seems not to be the case, which does not necessarily come as a surprise. Indeed, besides an assumed *sparsity* of the solution, the other ingredient for a successful CS-based approach is that of its *incoherence* with the measurements on which it is based. Operating in the Fourier domain clearly

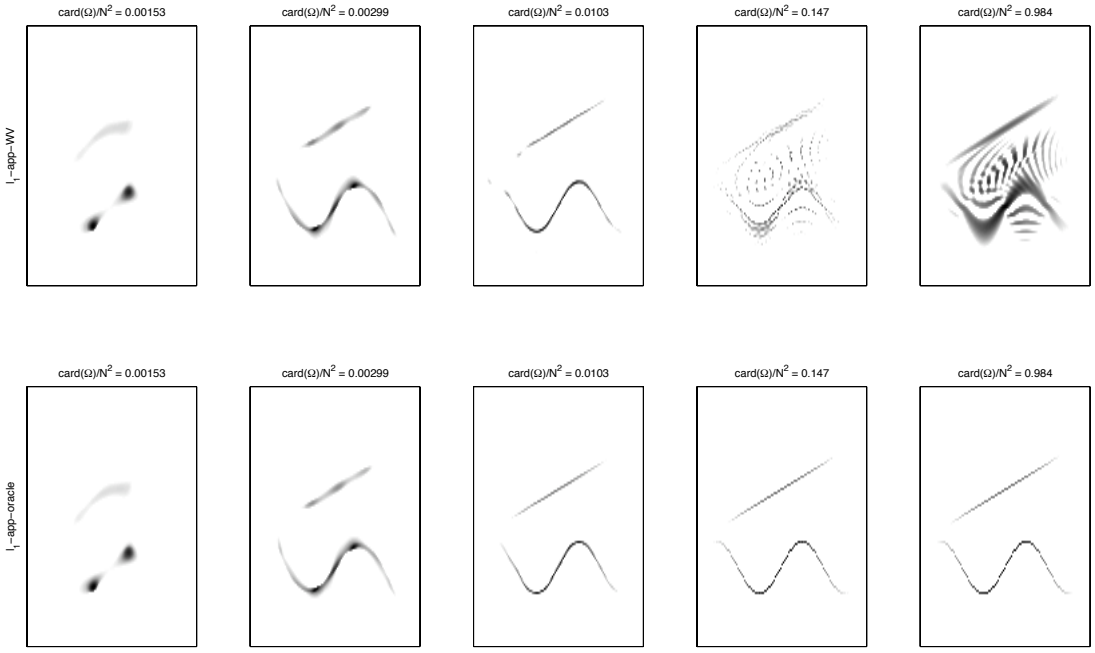

Fig. 3. *Qualitative selection of the optimal cardinality of* Ω — Top row: TF solutions obtained from the WVD as in Figure 1, for increasing values of the relative cardinality $\mathrm{card}(\Omega)/N^2$. The evolution shows that too small a cardinality does not permit to gain enough information about auto-terms whereas a too large one leads to take into account more and more undesired cross-terms, converging eventually to the actual WVD when the support of Ω identifies to the whole AF plane. Bottom row: companion "oracle" solutions obtained the same way, but based on the ideal distribution in place of the WVD. Comparison of both evolutions evidences that an optimum cardinality $|\Omega|_*$ exists for trading off auto-terms localization and cross-terms reduction (about 1% of the AF support in the present case). For all diagrams, amplitudes are grey coded logarithmically, with a dynamic range of 18 dB.

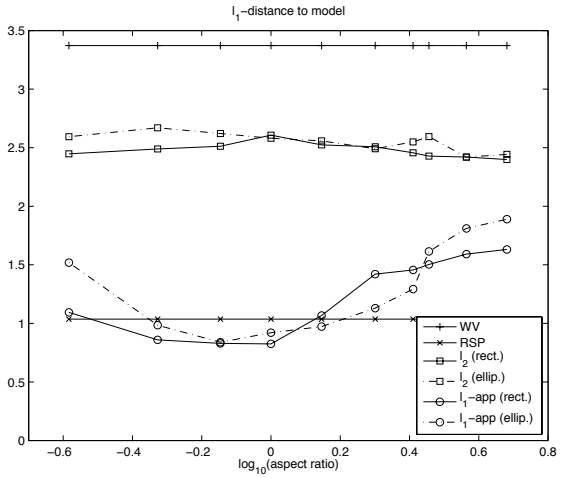

Fig. 4. *Influence of the anisotropy of* Ω — Given a domain Ω (rectangular or elliptic) whose area is kept fixed (and equal to 169), the ℓ_1 distances of the resulting TFDs to the actual model are plotted as a function of the aspect ratio "Doppler semi-axis/delay semi-axis" of Ω in the ambiguity plane.

goes this way, but *adapting* the selected subset to the signal structure goes the opposite.

C. More examples

Another example is shown in Figure 5 about the application of the method to a synthetic signal with crossing chirps, a case that is known to be difficult to recover from quadratic TF distributions (due to the interferences, [16]). The optimum magnitude for the Heisenberg cardinality is, for this example with two linear chirps with Gaussian modulated amplitudes, $\mathrm{card}(\Omega) \simeq 0.005\,N^2$, in agreement with the previous discussions. The distribution resulting of ℓ_1-app is satisfactorily close to the model one, see the quantitative results given in the following Table:

	model	RSP	ℓ_1-app
ℓ_1-distance to model	0	1.016	1.164
Rényi entropy	7.435	7.769	7.853

Simulation examples of this kind could of course be multiplied, and we will give only one more (already considered as a test-signal for reassignment methods in [20]). It consists of a signal that is similar to the one used in the first part of this paper, while having a structure that is more complex (and difficult to deal with) in three respects:

1) it has three components, with one of them embedded in the cross-terms resulting from the interference between the two other ones;

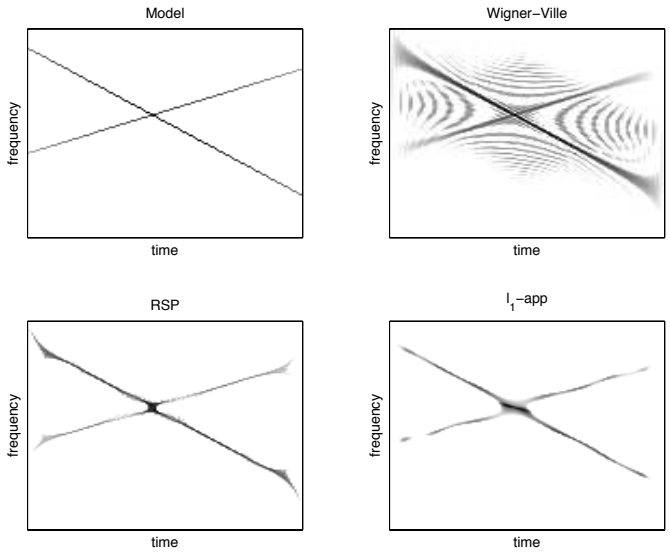

Fig. 5. *Crossing chirps example* — This Figure compares, for a signal composed of the 2 linear crossing chirps represented in the model (top left), its Wigner-Ville Distribution (top right), its reassigned spectrogram (bottom left) and the approximate minimum ℓ_1-norm solution (4) with $\epsilon = 0.05\,\|x\|_2$ (bottom right). The number of data points is $N = 128$ and the optimization is based on the knowledge of the 9×9 Fourier samples of the WVD neighbouring the origin of the AF plane, i.e, $\mathrm{card}(\Omega) \simeq 0.005\,N^2$. For all diagrams, amplitudes are grey coded logarithmically, with a dynamic range of 18 dB.

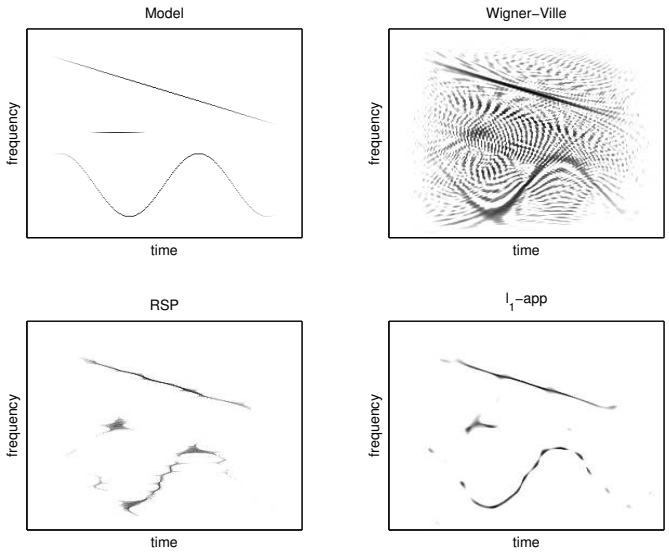

Fig. 6. *Noisy 3-components example* — This Figure compares, for a signal composed of 2 chirps and 1 Gabor logon represented in the model (top left), its Wigner-Ville Distribution (top right), its reassigned spectrogram (bottom left) and the approximate minimum ℓ_1-norm solution (4) with $\epsilon = 0.05\,\|x\|_2$ (bottom right), when the observation is the signal embedded in white Gaussian noise with SNR $= 10$ dB. The number of data points is $N = 256$ and the optimization is based on the knowledge of the 14×14 Fourier samples of the WVD neighbouring the origin of the AF plane, i.e, $\mathrm{card}(\Omega) \simeq 0.0035\,N^2$. For all diagrams, amplitudes are grey coded logarithmically, with a dynamic range of 18 dB.

2) the newly introduced component is a Gabor logon, which *a priori* does not enter directly the AM-FM model considered so far (no TF "trajectory" in the plane);
3) some noise is added (with a SNR fixed to 10 dB).

As for the previous example, results are plotted in Figure 6, with a performance summarized in the following Table:

	model	RSP	ℓ_1-app
ℓ_1-distance to model	0	1.599	1.554
Rényi entropy	8.065	9.298	9.406

As a complement to the synthetic examples of Figures 1, 5 and 6, Figure 7 is concerned with some real data, namely the classical benchmark signal of a bat echolocation call[1] of effective length 400, zero-padded to $N = 512$ data points. In this case too, the approximate minimum ℓ_1-norm solution (4) with $\mathrm{card}(\Omega) = 23 \times 23 \approx N$ compares very favorably with a reassigned spectrogram in terms of localization, with even some smoother regularity along TF trajectories.

D. Computational cost

Given the above reported findings, the new discussed approach is no doubt attractive in terms of its ability to give sharply localized TFDs in the case of transient, short duration AM-FM multicomponent signals. There is however a price to pay for this performance, which is a quite heavy computational cost. To illustrate this point, one can report the average computation times associated to the simulations of Figure 2, all computations having been performed with MATLAB R2008a in similar conditions (Intel Core 2 Duo 3.06 GHz on Mac OS 10.5). The result is that, whereas such an average time is 0.007 s for a WVD and 0.23 s for a reassigned spectrogram, it goes up to 50 s for the exact ℓ_1 solution and even to 152 s for the approximate one, using the implementation of the ℓ_1-MAGIC TOOLBOX [9]. It thus turns out that computing a CS-based TFD may differ from a classical approach by several orders of magnitude.

There is an increasing literature about alternative algorithms that could be used to solve programs having the form of (4) and this could lead to improvement in that respect. For instance, using the GPSR algorithm proposed in [27] to solve the quadratic program associated to (4), the computational cost goes down to 43 s (using the GPSR-BB with continuation) and, for a suitable choice of the regularization parameter (see Appendix), the results are identical in terms of ℓ_1 distance to the model and Rényi entropy of the resulting sparse TF distribution. Other approaches (see a comprehensive list on the CS resource page [6]) based on greedy strategies (for instance refinements on Matching Pursuit, e.g., [28], [29], [30]), or on different strategies for convex relaxation such as Iterative Thresholding (e.g., [31], [32], [33] and contributions improving those), could lead to more efficient recovery procedures in the current context, though we do not pursue here this question further on. Our discussion about the obtained non-negativity of the minimum ℓ_1-norm solutions can be complemented by

[1]The authors wish to thank Curtis Condon, Ken White, and Al Feng of the Beckman Institute of the University of Illinois for the bat data and for permission to use it in this paper.

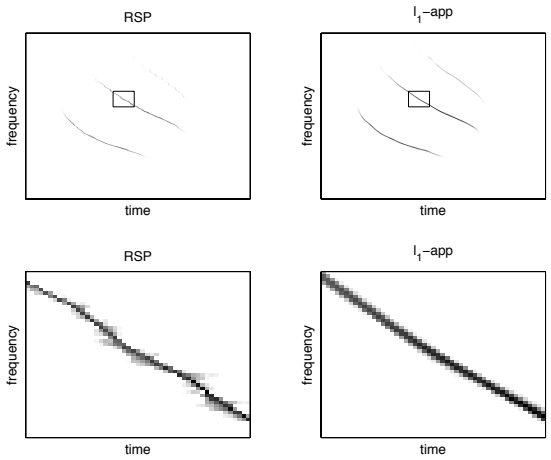

Fig. 7. *Real data example* — This Figure compares a reassigned spectrogram (left column) and the approximate minimum ℓ_1-norm solution (4) with $\epsilon = 0.05\,\|x\|_2$ (right column) in the case of a bat echolocation call of effective length $N = 400$, the optimization being based on the knowledge of the 23×23 Fourier samples of the WVD neighbouring the origin of the AF plane. The bottom row displays enlarged versions of the distributions within the boxes in the top row. For all diagrams, amplitudes are grey coded logarithmically, with a dynamic range of 18 dB.

noting that using some algorithms would make it possible to impose a condition on non-negativity of the solution, as proposed, for instance, in the GPSR method [27]. The issue of the choice of the better algorithm for the study proposed here is only sketched and will be part of further work.

V. Conclusion and perspectives

A new approach has been proposed for getting sharply localized TFDs in the case of multicomponent AM-FM signals by making use of the assumed sparsity of their energy distribution in the TF plane. Once the principle of the method has been given, there is clearly plenty of room for more thorough investigations and further developments.

As has been briefly mentioned in Section III-B, one particular interest of the present approach which is based on optimization, is that further constraints can be imposed besides (3) or (4). One can think, e.g., of the marginalization properties attached to unit cross-sections in the AF plane [14]. This however is not necessarily relevant in the context of a sharp TF localization since, in the case of multicomponent signals, highly oscillatory behaviours along TF trajectories will be favored. A more interesting variation is to favor regular time evolutions, what can be achieved explicitly by imposing specific AF cross-sections or, implicitly, by using as a starting point a time smoothed WVD.

Such extensions would also suggest to conduct further experiments in comparison with distributions within Cohen's class (with fixed or adapted kernels): a comprehensive treatment of this point is clearly beyond the scope of this paper, but this is a perspective that is certainly worth investigating in future developments. In a different direction, one can mention that CS-based approaches have been very recently proposed for the purpose of nonstationary spectral estimation in the case of stochastic processes [34], [35], in place of the time-frequency localization problem that has been considered here for deterministic chirp signals. While their objectives are somehow different, it turns out that the CS-based rationale of the two approaches is quite similar, and it would be interesting to merge them in a common perspective.

Whereas the heavy computational load can be considered as a severe drawback, it has been mentioned that improving upon this limitation is a current topic of very active research, with advances in this direction expected to be most benefitial for the present method. Moreover, it is worth stressing that the proposed optimization-based approach is not primarily intended to long duration signals. It is rather relevant for transient, highly nonstationary signals and, as such, it is expected to provide some nice addition to the existing toolbox of non-parametric TF methods aimed at exploratory data analysis.

Appendix

There has been a considerable interest in the solving of optimization problems such as the ones in eqs. (3) and (4) in the past years, thanks to the emergence of the "Compressed Sensing" (also called "Compressive Sampling", and referred to as CS) [4], [8], [9]. The purpose of this Appendix is to provide the reader with an overview of this domain and of some methods used to solve these optimization problems, and to refer to the articles proposing them.

A. Review on Compressed Sensing and optimization problems with ℓ_1 norm

The minimization problem ℓ_1-eq of eq. (3) is known as "Basis Pursuit" since it corresponds to [37]:

$$\min_{\rho} \|\rho\|_1 \,;\ \text{subject to}\ \ \mathcal{O}\{\rho\} = b, \tag{5}$$

where $\rho \in \mathbb{R}^n$, $b \in \mathbb{R}^k$ and $\mathcal{O}$ is an operator associated to a $k \times n$ matrix representation. The ℓ_1-app problem of eq. (4) is a relaxed version of ℓ_1-eq, where the equality constraint is replaced by a quadratic constraint:

$$\min_{\rho} \|\rho\|_1 \,;\ \text{subject to}\ \ \|\mathcal{O}\{\rho\} - b\|_2 \leq \epsilon. \tag{6}$$

Both problems are related in the sense that eq. (5) has the form (6) with $\epsilon = 0$.

Compressed Sensing (CS) deals with these problems when the number k of linear measurements available is much smaller than the number n of unknowns [4], and many methods have been proposed to solve both of them. Note that, in our context, $\mathcal{O}$ being a Fourier transform on the limited set Ω in the ambiguity domain, the number of measurements $k = \mathrm{card}\,(\Omega)$ is indeed much smaller than $n = N^2$, the number of points of the TFD; the situation is thus equivalent to a CS one.

B. A short lecture guide for optimization with ℓ_1-norm

A review of all the algorithmic procedures that have been proposed so far is outside the scope of this Appendix, see for instance discussions in [27], [28], [22] or the PhD thesis [36] for a comprehensive discussion on modern algorithms. Here, we will discuss how the two optimization problems stated

above can be solved for our work, using classical techniques from the CS field.

When b is real, the problem ℓ_1-eq can be recast and solved as a Linear Program, as shown in the seminal work on Basis Pursuit [37], and solved using interior point methods. In our context, the operator $\mathcal{O}$ is the Fourier transform and the constraint b corresponds to a restriction to some observations in the Fourier domain. This situation was considered in [9], proving that sparse vectors can then be recovered exactly through ℓ_1 minimization (see also [11], [12] for use of CS principles in the Fourier domain). When separating on different coordinates in b the mean and the real and imaginary parts of the Fourier transform, b turns out to be real and the primal-dual interior point method proposed in `l1eq_pd.m` of [22] can be used to solve (5), once recast as a constrained linear program.

As argued in the article, the solutions of (5) being often too sparse to be considered as a good TFD, solving the problem (6) is more relevant. It is a convex optimization problem, an instance of a second order cone program as considered in [38], [39]. Many approaches are now standard to solve it, including the log-barrier algorithm implemented in [22], interior-point algorithms as proposed in [37] or Iteratively-Reweighted Least Squares (see Appendix B of [38]). Also, using results from convex optimization, it is shown that the problem ℓ_1-app is related to the LASSO problem (minimization of the error with a ℓ_1-norm constraint) [40] and to the Basis Pursuit Denoising (BPDN) [37]:

$$\min_{\rho} \quad \lambda\|\rho\|_1 + \frac{1}{2}\,\|\mathcal{O}\{\rho\} - b\|_2^2. \tag{7}$$

Namely, solutions of (6) are either 0 or a minimizer of (7) for some $\lambda \geq 0$ [27]. As a consequence, any method that can find a solution to the BPDN problem of eq. (7) can be adapted for solving our original ℓ_1-app problem. For instance, homotopy methods originally proposed for the LASSO problem were considered in [28], [36]. Various modern instances of (Orthogonal) Matching Pursuit were proposed also to solve BPDN, e.g. [28], [29], [30], as well as iterative Shrinkage and Thresholding methods, e.g. [31], [32], [33].

Given the tremendous amount of works done on the algorithmic part of BP and BPDN problems, it was not possible to compare the behaviours of all the algorithms in our context. We limited ourselves to two approaches for the solving of ℓ_1-app with Fourier domain measurements:

- The `l1qc_logbarrier.m` software of the ℓ_1-MAGIC toolbox [22] that looks for a solution of ℓ_1-app as a second order cone program, solved with a log-barrier algorithm.
- The `GPSR_BB.m` software of the GPSR toolbox [27] as an instance of an efficient gradient projection algorithm that solves BPDN (expressed as a bound-constrained quadratic program) for a sequence of values of λ. This allows a reduction of the computational time to obtain solutions of ℓ_1-app, at the expense of not knowing a priori how to select λ.

When properly selecting λ, the results obtained for the TFD built from solving ℓ_1-app, as formulated in eq. (4), were similar. Such a proper choice can be made by following the guidelines reported in [27], i.e.,

$$\lambda = \alpha\|\mathcal{F}^{-1}\{A_x\}\|_\infty,$$

with $\alpha \approx 0.1$. In practice, it happened that a smaller value $\alpha \approx 0.02$ gave the best results for the considered examples. The question of finding in an automatic, data-driven way, the best regularization parameter and, more generally, the best algorithm for the specific context of TF analysis (especially in terms of trade-off between computational cost and localization), is left open to future works.

REFERENCES

[1] P. Borgnat and P. Flandrin, "Time-frequency localization from sparsity constraints," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-08*, pp. 3785–3788, Las Vegas (NV), 2008.

[2] P. Flandrin and P. Borgnat, "Sparse time-frequency distributions of chirps from a compressed sensing perspective," in *Proc. 8th IMA Int. Conf. on Math. in Signal Proc.*, pp. 84–87, Cirencester (UK), 2008.

[3] P. Flandrin, "Time-frequency and chirps," in *Proc. SPIE Meeting Wavelet Applications VIII*, Conf. No. 8, Vol. 4391, pp. 161–175, Orlando (FL), 2001.

[4] R.G. Baraniuk, E. Candès, R. Nowak and M. Vetterli (eds.), "Special Section on Compressive Sampling," *IEEE Sig. Proc. Mag.*, Vol. 25, No. 2, pp. 12–101, 2008.

[5] A.M. Bruckstein, D.L. Donoho and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, Vol. 51, No. 1., pp. 34–81, 2009.

[6] `http://www.dsp.ece.rice.edu/cs/`

[7] G.E. Pfander and H. Rauhut, "Sparsity in time-frequency representations," arXiv:0711.2503, 2007. To appear in *J. Fourier Anal. Appl.*, 2009.

[8] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Info. Theory*, Vol. 52, No. 4, pp. 1289–1306, 2006.

[9] E. Candès, J. Romberg and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Info. Theory*, Vol. 52, No. 2, pp. 489–509, 2006.

[10] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. on Info. Theory*, Vol. 52, No. 12, pp. 5406–5425, 2006.

[11] M. Lustig, D. Donoho and J.M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, Vol. 58, No. 6, pp. 1182–1195, 2007.

[12] M. Lustig, D. Donoho, J.M. Santos and J.M. Pauly, "Compressed Sensing MRI," *IEEE Signal Proc. Mag.*, Vol. 25, No. 2, pp. 72–82, 2008.

[13] L. Applebaum, S.D. Howard, S. Searle and R. Calderbank , "Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery," *Appl. Comp. Harmonic Anal.*, Vol. 26, No. 2, pp. 283–290, 2009.

[14] P. Flandrin, *Time-Frequency/Time-Scale Analysis*. San Diego: Academic Press, 1999.

[15] P. Flandrin, P. Gonçalvès, "Geometry of affine time-frequency distributions," *Appl. Comp. Harmonic Anal.*, Vol. 3, pp. 10–39, 1996.

[16] F. Hlawatsch, P. Flandrin, "The interference structure of Wigner distribution and related time-frequency representations ," in *The Wigner Distribution — Theory and Applications in Signal Processing* (W. Mecklenbräuker and F. Hlawatsch, eds.), pp. 59-133, Elsevier, Amsterdam, 1997.

[17] P. Flandrin, "Cross-terms and localization in time-frequency energy distributions," in: *Time-Frequency Signal Analysis and Processing* (B. Boashash, *ed.*), Article 4.2, pp. 94–101, Elsevier, 2003.

[18] P. Flandrin, "Some features of time-frequency representations of multicomponent signals," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-84*, San Diego (CA), pp. 41B.4.1–41B.4.4, 1984.

[19] R.G. Baraniuk and D.L. Jones, "Signal-dependent time-frequency analysis using a radially Gaussian kernel," *Signal Proc.*, Vol. 32, No. 3, pp. 263–284, 1993.

[20] P. Flandrin, F. Auger and E. Chassande-Mottin, "Time-Frequency Reassignment — From Principles to Algorithms," in *Applications in Time-Frequency Signal Processing* (A. Papandreou-Suppappola, *ed.*), Chap. 5, pp. 179–203, CRC Press, 2003.

[21] `http://tftb.nongnu.org`

[22] E. Candès and J. Romberg, "ℓ_1-MAGIC: Recovery of Sparse Signals via Convex Programming", User's Guide of the ℓ_1-MAGIC toolbox for MATLAB, `http://www.acm.caltech.edu/l1magic/`, 2005.
[23] R.G. Baraniuk, P. Flandrin, A.J.E.M. Janssen and O. Michel, "Measuring time-frequency information content using the Rényi entropies," *IEEE Trans. on Info. Theory*, Vol. 47, No. 4, pp. 1391–1409, 2001.
[24] R. Price and E.M. Hofstetter, "Bounds on the volume and height distribution of the ambiguity function," *IEEE Trans. on Info. Theory*, Vol. 11, pp. 207–214, 1965.
[25] S. Qian and J.M. Morris, "Wigner distribution decomposition and cross-terms deleted representation," *Sig. Proc.*, Vol. 27, pp. 125–144, 1992.
[26] S. Qian and D. Chen, "Decomposition of the Wigner distribution and time-frequency distribution series," *IEEE Trans. on Sig. Proc.*, Vol. 42, pp. 2836–2842, 1994.
[27] M. Figueiredo, R. Nowak and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. of Select. Topics in Signal Proc.: Special Issue on Convex Optimization Methods for Signal Processing*, Vol. 1, No 4, pp. 586–598, 2007.
[28] D. Donoho, Y. Tsaig, I. Drori and J.-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," Report preprint available from `http://www-stat.stanford.edu/~donoho/reports.html`, 2006.
[29] J. A. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Info. Theory*, Vol. 53, No. 12, pp. 4655–4666, 2007.
[30] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harm. Anal.*, Vol. 26, No. 3, pp. 301–321, 2009.
[31] I. Daubechies, M. Defrise and C. DeMol, "An iterative thresholding algorithm for linear inverse problems," *Comm. Pure Appl. Math.*, Vol. 57, No. 11, pp. 1413–1457, 2004.
[32] I. Daubechies, M. Fornasier and I. Loris, "Accelerated Projected Gradient Method for Linear Inverse Problems with Sparsity Constraints," *J. Fourier Anal. Appl.*, Vol. 14, No. 5–6, pp. 764-792, 2008.
[33] T. Blumensath and M. E. Davies, "Iterative Thresholding for Sparse Approximations," *J. Fourier Anal. Appl.*, Vol.14, No 5, pp. 629–654, 2008.
[34] A. Jung, G. Taubck, and F. Hlawatsch, "Compressive spectral estimation for nonstationary random processes," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-09*, pp. 3029–3032, Taipei (Taiwan, R.O.C.), 2009.
[35] A. Jung, G. Taubck, and F. Hlawatsch, "Compressive nonstationary spectral estimation using parsimonious random sampling of the ambiguity function," in *Proc. IEEE Statistical Signal Processing Workshop SSP-09*, pp. 642–645, Cardiff (Wales UK), 2009.
[36] Y. Tsaig, "Sparse solution of underdetermined linear systems: Algorithms and applications," PhD Thesis, Stanford University, 2007.
[37] S. S. Chen, D. L. Donoho and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, Vol. 20, pp. 33–61, 2001.
[38] D. L. Donoho, M. Elad and V. N. Temlyakov, "Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise," *IEEE Trans. on Info. Theory*, Vol. 52, No. 1, pp. 6–18, 2006.
[39] E. Candès, J. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements" *Comm. Pure Appl. Math.*, Vol. 59, No. 8, pp. 1207–1223, 2006.
[40] R. Tibshirani, "Regression shrinkage and selection via the lasso", *J. Royal Statist. Soc. B.*, Vol. 58, No. 1, pp. 267–288, 1996.

Pierre Borgnat was born in Poissy, France, in 1974. He made his studies at the École Normale Supérieure de Lyon, France, receiving the Professeur-Agrégé de Sciences Physiques degree in 97, a Ms. Sc. in Physics in 99 and defended a Ph.D. degree in Physics and Signal Processing in 2002. In 2003-2004, he spent one year in the Signal and Image Processing group of the IRS, IST (Lisbon, Portugal). Since October 2004, he has been a full-time CNRS researcher with the Laboratoire de Physique, ÉNS Lyon. His research interests are in statistical signal processing of non-stationary processes (time-frequency representations, time deformations, stationarity tests) and scaling phenomena (time-scale, wavelets) for complex systems (turbulence, networks,...). He is also working on Internet traffic measurements and modeling, and in analysis and modeling of dynamical complex networks.

Patrick Flandrin (M'85–SM'01–F'02) received the engineer degree from ICPI Lyon, France, in 1978, and the Doct.-Ing. and "Docteur d'État" degrees from INP Grenoble, France, in 1982 and 1987, respectively. He joined CNRS in 1982, where he is currently Research Director. Since 1991, he has been with the "Signals, Systems and Physics" Group, within the Physics Department at École Normale Supérieure de Lyon, France. In 1998, he spent one semester in Cambridge, UK, as an invited long-term resident of the Isaac Newton Institute for Mathematical Sciences and, from 2002 to 2005, he has been Director of the CNRS national cooperative structure "GdR ISIS." His research interests include mainly nonstationary signal processing (with emphasis on time-frequency and time-scale methods) and the study of self-similar stochastic processes. He published many research papers in those areas and he is the author of the book *Temps-Fréquence* (Paris: Hermès, 1993 and 1998), translated into English as *Time-Frequency/Time-Scale Analysis* (San Diego: Academic Press, 1999). He has been a guest co-editor of the Special Issue "Wavelets and Signal Processing" of the IEEE TRANSACTIONS ON SIGNAL PROCESSING in 1993, the Technical Program Chairman of the 1994 IEEE-SP Int. Symp. on Time-Frequency and Time-Scale Analysis and, since 2001, he is the Program Chairman of the French GRETSI Symposium on Signal and Image Processing. He is currently an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and he has been a member of the "Signal Processing Theory and Methods" Technical Committee of the IEEE Signal Processing Society from 1993 to 2004.

Dr. Flandrin was awarded the Philip Morris Scientific Prize in Mathematics in 1991, the SPIE Wavelet Pioneer Award in 2001 and the Prix Michel Monpetit from the French Academy of Sciences in 2001. He is Fellow of IEEE (2002) and of EURASIP (2009).