On Wigner-based sparse time-frequency distributions

Patrick Flandrin, Nelly Pustelnik, and Pierre Borgnat
CNRS & ENS de Lyon
46 allée d'Italie, 69364 Lyon Cedex 07, France
{firstname.lastname}@ens-lyon.fr

***Abstract*—Signals made of the superimposition of a reduced number of AM-FM components can be characterized by a time-frequency signature which consists of weighted trajectories in the plane, thus ending up with an ideal representation of their energy distribution that is intrinsically sparse. Elaborating on first studies that pioneered a compressed sensing solution to the question of approaching such an ideally localized distribution by selecting samples in the ambiguity domain and imposing sparsity in the time-frequency domain, the present paper discusses new advances aimed at achieving better performance in the construction of "cross-term-free" Wigner-type distributions. Improved optimization schemes are first proposed, that both speed up the computation and prove more versatile to accommodate for side constraints such as positivity. A special attention is then paid to the choice of the necessary measurements in the ambiguity plane (in fixed or adapted geometries), emphasizing the key role played by the Heisenberg minimum area, regardless of the signal complexity.**

I. Introduction

Sparsity-based signal processing has received a considerable attention over the last 10 years. Observing that many data are sparse when transformed in an appropriate representation space, *compressed sensing* techniques [1], [2] have been developed on the basis that sparse data can be reconstructed almost perfectly with a remarkably small number of measurements made in a dual domain. Wavelets are emblematic of such a sparsifying transformation, but one can also think of the Fourier transformation when applied to a limited number of tones and, by extension, to time-frequency (TF) representations in the case of multicomponent AM-FM signals. This observation prompted to reconsider the question of getting a sharply localized TF distribution from the fresh perspective of imposing sparsity in the TF domain while using only few measurements in some Fourier transform domain. The first work in this direction [3] has been followed by a few studies in both deterministic [4], [5] and stochastic [6] settings. The purpose of this paper is to go further along such lines by (*i*) improving the actual implementation of the optimization-based technique and (*ii*) better understanding the performance that can be expected from this alternative approach.

II. Rationale

A. Time-frequency sparsity

Many natural and man-made signals can be modeled as a superimposition of AM-FM components of the form

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

that generalizes locally (in time) a Fourier decomposition by attaching "instantaneous amplitudes" $a_k(t)$ to "instantaneous frequencies" $\dot\varphi_k(t)/2\pi$. From an energetic point of view, this amounts to consider as an ideal TF representation the quantity

$$\rho_x(t,f) = \sum_{k=1}^{K} a_k^2(t)\delta(f-\dot\varphi_k(t)/2\pi) \tag{2}$$

that consists of a finite collection of weighted trajectories in the TF plane [7], [8], [9]. In the common situation where the analyzed signal is observed as a sequence of N samples in time and is analyzed over N frequency bins, the resulting TF matrix is of size $N \times N$ whereas (2) suggests an occupancy with at most $N \times K$ non-zero entries. Therefore, provided that $K \ll N$, the TF matrix is expected to be intrinsically sparse and, in turn, exploiting explicitly such a sparsity assumption is a way to approach the ideal TF representation given in (2).

B. Localization vs. interference

It is well-known that quadratic TF energy distributions may achieve a perfect localization as in (2) for some specific phase functions $\varphi_k(t)$, but only in the monocomponent case ($K = 1$) [8]. As soon as $K > 1$, the localization of individual components (e.g., linear chirps with a quadratic phase in the Wigner case) is hampered by the existence of cross-terms [10]. A convenient way of handling this problem is to switch from the TF plane to its Fourier transformed domain. Indeed, the 2D Fourier transform of the Wigner distribution identifies to the so-called ambiguity function (AF), defined as $A_x(\xi,\tau) = \langle x, \mathbf{T}_{\xi,\tau} x\rangle$, with $\mathbf{T}_{\xi,\tau}$ a TF shift operator. The AF appears therefore as a form of TF correlation function and, as such, it contains interaction terms (between distinct components) that lie away from the origin of the AF plane, at a distance that roughly corresponds to the TF distance between those components. This suggested long ago [11] to reduce the influence of such cross-terms by windowing $A_x(\xi,\tau)$ with some appropriate function $w(\xi,\tau)$ prior Fourier inversion:

$$\hat\rho_x(t,f) = \underset{\xi\to t,\tau\to f}{\mathcal{F}^{-1}} \left\{ w(\xi,\tau) A_x(\xi,\tau) \right\}. \tag{3}$$

While such a procedure proved successful in many ways, it is faced with a severe limitation: an effective reduction of cross-terms requires the AF window to be of limited support, which in turn degrades the localization of the surviving contributions when applying the inverse Fourier transformation. An illustration of this behavior is reported in Figure 1 (second row).

Different improvements have been proposed in order to overcome this limitation, either by designing a data-driven weighting function $w(\xi,\tau)$ (this will be discussed further in Sect. III-B), or by finely exploiting phase informations so as to sharpen a classical TF distribution. The latter approach

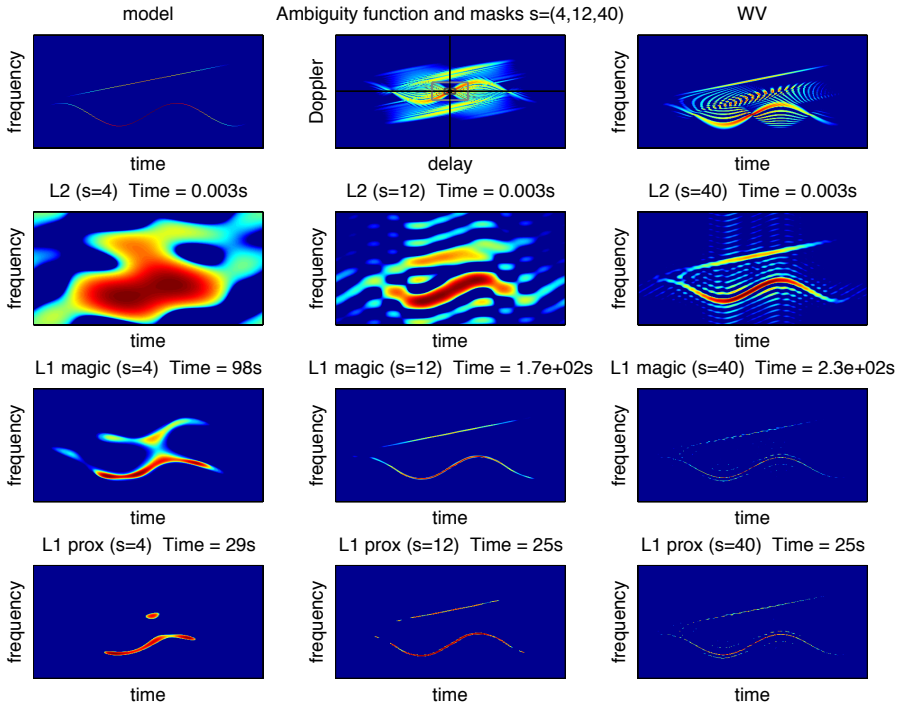

Fig. 1. Principle of sparse reconstruction — On the first line, we present the model (left), the ambiguity function (middle) with 3 differents masks Ω of size $(s+1)^2$, with $s = 4$ (black), $s = 12$ (dark gray), $s = 40$ (light gray), and the Wigner distribution (right). On the second line, we present the L2 inversion for the 3 different sizes of the mask, i.e., the classical solution given by (3). On the third (resp., fourth) line, the optimization-based results associated with the L1-magic (reps., L1-prox with $\lambda = 0.01$ and $5,000$ iterations) procedure are illustrated for the different masks. Computation times are given in each case, for a sake of comparison.

encompasses the so-called *reassignment technique* [12] that is known to achieve an almost perfect localization in many realistic situations, and that will be used as a benchmark in the sequel. Whereas reassigned spectrograms can be considered as state-of-the-art distributions in terms of localization, it must however be noted that their optimal performance depends on the choice of a well-chosen short-time window, which might prove difficult in blind situations.

C. An optimization framework

The observation above about the "localization vs. cross-terms" trade-off, together with the idea that the target distribution (2) should be intrinsically sparse, was the key for proposing in [3] an alternative approach to (3). More precisely, $\hat{\rho}_x(t,f)$ was proposed to be defined as the solution of an optimization problem in which sparsity is imposed by minimizing the ℓ_1-norm of the distribution, while the values of its 2D Fourier transform are constrained to equal those of the actual AF in a domain Ω of reduced support, that is neighboring the origin of the plane so as to reject most of cross-terms. This initial formulation, with the exact constraint $\mathcal{F}\{\rho\} = A_x$ over the set Ω, has been shown [3] to be too strict, leading to discontinuous, over-spiky solutions. This justified the use of a relaxed constraint that will be used in the present work. More specifically, considering from now on a discrete-time setting in which TF distributions as well as AFs are defined as matrices, the approach we will adopt can be formulated as

$$\hat{\rho}_\epsilon \in \underset{\rho\in\mathbb{R}^{N\times N}}{\operatorname{Argmin}} \|\rho\|_1 \quad \text{subj. to} \quad \|M_\Omega \circ (\mathcal{F}\{\rho\} - A_x)\|_{\mathrm{F}}^2 \leq \epsilon \tag{4}$$

where $\circ$ denotes the pointwise matrix product and $\|\cdot\|_{\mathrm{F}}$ the Frobenius norm, $\|\cdot\|_1$ modeling the TF ℓ_1-norm defined as

$$\|\rho\|_1 = \sum_{n=1}^{N}\sum_{m=1}^{N} |\rho_{nm}|,$$

and M_Ω the AF mask such that $M_\Omega(\xi,\tau) = 1$ if $(\xi,\tau) \in \Omega$ and 0 otherwise. An equivalent formulation of (4) reads

$$\hat{\rho}_\lambda \in \underset{\rho}{\operatorname{Argmin}} \|\rho\|_1 + \lambda \|M_\Omega \circ (\mathcal{F}\{\rho\} - A_x)\|_{\mathrm{F}}^2, \tag{5}$$

where λ, as in any inverse problem, is a regularization parameter that balances the constraints regarding sparsity in the TF domain and fidelity in the AF domain. In order to impose some additional constraint (such as, e.g., positivity), we can further modify (5) as

$$\hat{\rho}_{\lambda,C} \in \underset{\rho}{\operatorname{Argmin}} \|\rho\|_1 + \iota_C(\rho) + \lambda \|M_\Omega \circ (\mathcal{F}\{\rho\} - A_x)\|_{\mathrm{F}}^2, \tag{6}$$

with C a non-empty closed convex subset of $\mathbb{R}^{N\times N}$ and $\iota_C(\rho)$ the indicator function that is equal to 0 if $\rho \in C$ and $+\infty$ otherwise.

The algorithmic solution[1] proposed in [3] was based on the L1-magic toolbox[2] whose algorithmic core consists in reformulating the sparse minimization problem as a second-order cone program and uses a log barrier algorithm [1]. While the achieved results were convincing, the computational cost proved to be high. Numerous algorithms, which belong to the

[1] perso.ens-lyon.fr/patrick.flandrin/sparseTFR.html
[2] users.ece.gatech.edu/~justin/l1magic/

class of proximal methods [13], [14], [15], have been recently developed to deal more efficiently with large scale problems and non-smooth criteria. In order to obtain a faster algorithm and to easily deal with additional constraints as in (6), we can use the forward-backward algorithm [16]. This algorithm is well suited for a minimization problem involving a smooth data term having a Lipschitz gradient (the most right term of (6)) and a non-smooth penalization (the sum of the ℓ_1-norm and of the indicator function) [17].

In Figure 1, we contrast the proposed L1-prox[3] solution of (5) with that obtained with L1-magic for the 256-points signal used as a test example in [3]. The results are very similar for several choices of Ω (here chosen as a square domain of size $(s+1)^2$, with $s = \{4, 12, 40\}$), with the notable difference that the computational time is significantly lower and that the convergence to a sparse solution is better when s is small.

III. Constraints geometry and performance

A. Measurements vs. uncertainty

Reasoning along a "compressed sensing" line, the sparsity assumption attached to the model (2) would suggest that the number of measurements, i.e., the size of the domain Ω in the AF plane, be dependent on the number K of components. However, it turns out that the TF situation is different from a classical compressed sensing problem in which the issue would be to *reconstruct* some sparse function from a reduced number of measurements. In the present case where measurements are done in the AF domain, there would be no point in addressing such a *reconstruction* problem, since this would lead to the Wigner distribution as the solution. What is aimed at is indeed the *construction* of a new object that best approaches a "cross-term-free" distribution which can be thought of as the skeleton of the Wigner distribution. By Fourier duality, the 2D transforms of the components of this TF skeleton all collapse around the origin of the AF plane, with a phase structure that encodes their respective locations while the magnitude of their combination extends over a domain whose area is of the order of Heisenberg lower bound. It can be therefore expected that, in the present context, it is this Heisenberg limit that will play a key role. Such an interpretation, that is supported by the behavior reported in the example of Figure 1, will now be precised further.

B. On adaptive kernels

The selection of a relevant domain Ω in the AF plane involves two degrees of freedom, namely shape and area. Without any specific prior information about the signal structure, simple shapes such as squares or disks can be adopted, while one might think of possible improvements based on an adaptive choice of the domain that would be data-driven. A classical solution in this direction [18] is provided by choosing for $w(\xi,\tau)$ in (3) a weighting function such that

$$w_*(\xi,\tau) = \arg\max_{w} \iint_{-\infty}^{+\infty} |w(\xi,\tau) A_x(\xi,\tau)|^2 \, d\xi \, d\tau, \quad (7)$$

under the constraint $\|w\|_2 = 1$. The rationale of this technique is to favor the selection of AF contributions nearby the origin of the plane, for a constrained total volume that makes of the resulting distribution an adapted spectrogram. An efficient solution has been proposed in [18] in the case where $w(\xi,\tau)$ is assumed to be radially Gaussian.

[3] perso.ens-lyon.fr/patrick.flandrin/proxTFR.html

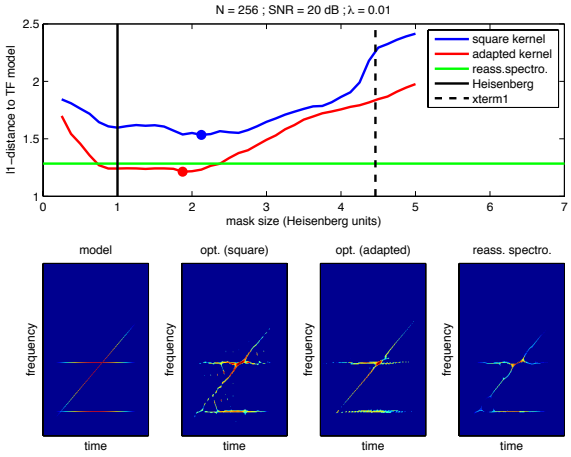

Fig. 2. Fixed vs. adaptive kernels (1) — In the top diagram, the ℓ_1-distance to the model is plotted for fixed (square) and adapted kernels of the same size (in Heisenberg units), together with the "best" result obtained with a reassigned spectrogram. The subplots of the bottom row display the TF model and the corresponding optimal solutions.

Figure 2 compares the results obtained with either square or adapted masks of the same size. In both cases, the ℓ_1-distance to the model is plotted as a function of the typical length attached to the mask, measured in Heisenberg units[4]. For a sake of comparison, the best performance of a reassigned spectrogram is displayed too. What can be learnt from this example is at least threefold. First, the performance curves attain a minimum, in both cases of fixed and adapted masks, for a typical size which is of the order of Heisenberg's. Second, the value of this minimum evidences that combining the optimization-based approach with an adapted kernel is optimal as compared to reassigned spectrograms (let us recall that the best reassigned spectrogram requires the use of an optimal window that is unknown in practice). Third, one can observe for the fixed kernel a jump for a size such that the cross-term between the 2 tones enters the mask, thus starting reconstructing unwanted oscillations in the TF solution. Finally, one can remark that the model chosen for this example is particularly favorable to the radially Gaussian adaptation of the kernel, while supporting the versatility of a fixed square kernel whose effectiveness is almost comparable.

A companion example is given in Figure 3, where a 3-tone signal is considered. The 3 chosen frequencies being not harmonically related, the AF structure is made of 3 distinct cross-terms ("xterm1", "xterm2" and "xterm3") whose locations are identified on the diagram. When increasing enough the number of iterations in the algorithm (from 100 for thin lines to $5,000$ for thick ones), we observe again the distinctive features of transitions in the error that are coupled with the locations of cross-terms. Fixed kernels happen to attain the optimal localization (i.e., that of the best reassigned spectrogram). The

[4] In this system of units, lengths are normalized so that a square of length 1 has an area which equals the minimum TF occupancy authorized by Heisenbergs's uncertainty principle.

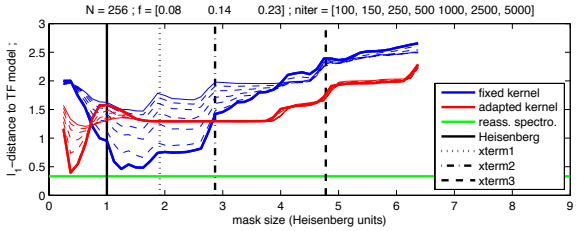

Fig. 3. Fixed vs. adaptive kernels (2) — In the case of a 3-tone signal, the ℓ_1-distance to the model is plotted for square and adapted kernels of the same size (in Heisenberg units), together with the best result achievable with a reassigned spectrogram. The different curves correspond to different numbers of iterations in the algorithm (from 100 for thin lines to $5,000$ for thick ones), the regularization parameter being fixed to $\lambda = 0.01$.

same behavior is reported for the adapted kernel with a much smaller mask, but this is quite specific of the chosen model.

C. On positivity

Given a mask in the AF domain, neither a Fourier inversion nor a simple ℓ_1 minimization guarantees the positivity of the solution, a property that is satisfied by (reassigned) spectrograms and that might be important for a sake of interpretation. As suggested above, such a property can however be imposed by resorting to (6) in place of (5) in the optimization process. As illustrated in Figure 4 (same example as in Figure 1), investigating this issue leads to two main conclusions. On the one hand, unconstrained solutions end up for free with positive solutions as long as the mask is of the order of Heisenberg. On the other hand, solutions constrained to be positive (with $C = [0, +\infty)^{\mathbb{N}\times\mathbb{N}}$ in (6)) behave the same, but with an error that stabilizes to a plateau for larger values of the mask size.

IV. Conclusion

An improved version of the sparsity-based approach to TF energy distributions has been proposed, and elements supporting its versatility have been provided. The main finding is that, for a tuning of the free parameters that is almost signal-independent (namely, a mask size of the order of Heisenberg's uncertainty, a regularization parameter $\lambda \approx 0.01$, and about 5,000 iterations), the method is able to attain state-of-the-art performance of signal-dependent techniques. As compared to the original work reported in [3], resorting to proximal methods is more versatile, it easily allows for accommodating additional constants such as positivity, and it also guarantees a moderate computational time, making of the proposed approach an effective addition to the toolkit of TF practitioners.

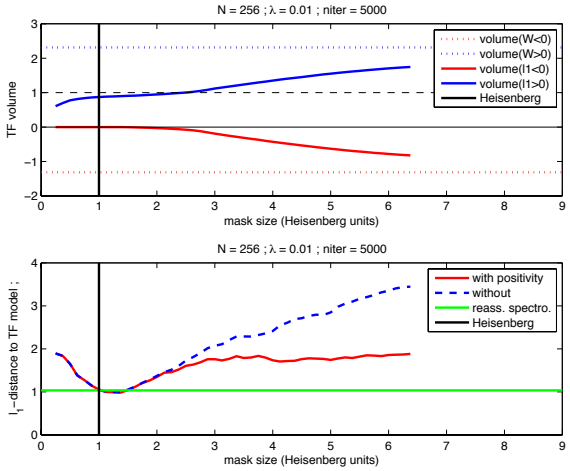

Fig. 4. Positivity — Top diagram: non-positivity characterization in terms of TF volumes for the normalized unconstrained L1-prox solution and the Wigner distribution (positivity corresponds to a zero volume for negative values, while positive and negative contributions are always expected to contribute to volumes that add up to unity). Bottom diagram: comparison of the unconstrained and positivity constrained L1-prox solutions. In both diagrams, the regularization parameter ($\lambda = 0.01$) is optimal, the number of iterations being fixed to $5,000$.

References

[1] E.J. 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.

[2] E.J. Candès and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Proc. Mag.*, Vol. 25, No. 2, pp. 21–30, 2008.

[3] P. Flandrin and P. Borgnat, "Time-frequency energy distributions meet compressed sensing," *IEEE Trans. on Signal Proc.*, Vol. 58, No. 6, pp. 2974–2982, 2010.

[4] L. Stankovic, I. Orovic, S. Stankovic, and M.G. Amin, "Compressive sensing based separation of non-stationary and stationary signals overlapping in time-frequency," *IEEE Trans. on Signal Proc.*, Vol. 61, No. 18, pp. 4562–4572, 2013.

[5] Y.D. Zhang, M.G. Amin, and B. Himed, "Reduced interference time-frequency representations and sparse reconstruction of under sampled data," in Proc. of the 21st European Signal Proc. Conf. EUSIPCO-13, Marrakech (MO), 2013.

[6] A. Jung, G. Tauböck, and F. Hlawatsch, "Compressive spectral estimation for nonstationary random processes," *IEEE Trans. on Info. Theory*, Vol. 59, No. 5, pp. 3117–3138, 2013.

[7] B. Boashash (ed.), *Time-Frequency Signal Analysis and Processing. A Comprehensive Reference*, Elsevier, Oxford (UK), 2003.

[8] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, Academic Press, San Diego (CA), 1999.

[9] P. Flandrin, M.G. Amin, S. McLaughlin, and B. Torrésani (eds.), Special Issue on "Time-frequency analysis and applications," *IEEE Signal Proc. Mag.*, Vol. 30, No. 6, pp. 19–150, 2013.

[10] P. Flandrin, "Cross-terms and localization in time-frequency energy distributions," in :[7], Article 4.2, pp. 94–101, 2003.

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

[12] P. Flandrin, F. Auger, and E. Chassande-Mottin, "Time-frequency reassignment—From principles to algorithms," in *Applications in Time-Frequency Signal Processing*, A. Papandreaou-Suppappola, Ed., Chap. 5, pp. 179–203, Boca Raton (FL), 2003.

[13] P.L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H.H. Bauschke *et al.* (eds.), pp. 185–212, 2010.

[14] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011.

[15] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, Vol. 1, No. 3, pp.123–231, 2014.

[16] P.L. Combettes and V.R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling and Simulation*, Vol. 4, No. 4, pp.1168–1200, 2005.

[17] C. Chaux, J.-C. Pesquet, and N. Pustelnik, "Nested iterative algorithms for convex constrained image recovery problems," *SIAM Journal on Imaging Sciences*, Vol. 2, No. 2, pp. 730-762, 2009.

[18] D.L. Jones and R.G. Baraniuk, "An adaptive optimal-kernel time-frequency representation, " *IEEE Trans. on Signal Proc.*, Vol. 43, No. 10, pp. 2361–2371, 1995.