TIME-FREQUENCY RIDGE ANALYSIS BASED ON THE REASSIGNMENT VECTOR

S. Meignen [*], *T. Gardner* [†], *T. Oberlin* [‡]

[*] Laboratoire Jean Kuntzmann, Grenoble, France
[†] Boston University, Boston, MA, USA
[‡] INP–ENSEEIHT and IRIT, University of Toulouse, Toulouse, France

ABSTRACT

This paper considers the problem of detecting and estimating AM/FM components in the time-frequency plane. It introduces a new algorithm to estimate the ridges corresponding to the instantaneous frequencies of the components, and to segment the time-frequency plane into different "basins of attraction", each basin corresponding to one mode. The technique is based on the structure of the reassignment vector, which is commonly used for sharpening time-frequency representations. Compared with previous approaches, this new method does not need extra parameters, exhibits less sensitivity to the choice of the window and shows better reconstruction performance. Its effectiveness is demonstrated on simulated and real datasets.

Index Terms— multicomponent signals; short-time Fourier transform; reassignment; time-frequency; AM/FM; ridges

1. INTRODUCTION

Many signals from the physical world can be modelled as a sum of amplitude- and frequency-modulated (AM/FM) components, and are therefore designated as multicomponent signals (MCS). The short-time Fourier transforms (STFT) of MCS exhibits particular time-frequency (TF) structures which are of great importance for analysing the constituent modes of the signal. Indeed, the components are associated with local maxima of the TF representation that are known to make up curves, called *ridges* in the TF plane. Many techniques for modes analysis from their TF representation are based on ridge analysis. Among these techniques, those based on reassignment have been extensively used [1, 2, 3] to enhance the TF representation. The basis of the reassignment technique is to move the points of a given TF representation onto a ridge following the direction given by the so-called *reassignment vector* (RV). One limitation of reassignment technique is that it does not allow for mode reconstruction. Recently, a technique, called *Synchrosqueezing transform*,

has been introduced based on reassignment along the frequency axis that enhances TF representation while enables mode reconstruction [4].

Every previously described technique is nevertheless tied to a robust ridge estimator. In this paper, we first show the central role played by RV to determine the ridges which will appear as a subclass of objects called *contours* and then to subsequently compute the so-called *basins of attraction* associated with these ridges. The novelty of our approach is that both ridges detection and basins of attraction determination are fully adaptive. Indeed, to compute such entities some results were already obtained in [5] considering the inner product of RV with vectors having a predefinite orientation. After having explained the specificity of ridges in the contours' class, we focus on how to compute the basin of attraction associated with a given ridge using RV and the knowledge on the contour location computed just before. Finally, we introduce a new mode reconstruction technique based on the basin of attraction which we numerically test on noisy simulated signals and real ones.

2. DEFINITIONS

Given a signal $f \in L^1(\mathbb{R})$, the space of real integrable functions, we define its Fourier transform by:

$$\hat{f}(\eta) = \int_{\mathbb{R}} f(t)\, e^{-i\eta t}\, dt. \tag{1}$$

The *STFT* of a signal is then defined by

$$V_f^g(\eta, t) = \int_{\mathbb{R}} f(u)g(u-t)e^{-i\eta u}\, du, \tag{2}$$

where g is assumed to be a real-valued window with L^1 norm equal to 1. The spectrogram is then defined as $|V_f(\eta, t)|^2$. Finally, multicomponent signals f to be considered in the sequel are defined by:

$$f(t) = \sum_{k=1}^{K} f_k(t), \text{ with } f_k(t) = a_k(t)e^{i\phi_k(t)}, \tag{3}$$

for some finite K, where $a_k(t) > 0$ is a continuously differentiable function, ϕ_k is a two times continuously differentiable

function satisfying $\phi'_k(t) > 0$ and $\phi'_{k+1}(t) > \phi'_k(t)$ for all t. Note that this last condition enforces the frequency *separation* of the modes, which is needed to ensure the uniqueness of the decomposition. In the following, f_k will be referred to as an AM-FM component.

3. DEFINING RIDGES AND CONTOURS

3.1. Reassignment of the spectrogram

The principle of the *reassignment method* (RM) is to compensate for the TF shifts induced by the 2D smoothing defining the spectrogram. To do so, a meaningful TF location to which to assign the local energy given by the spectrogram is first determined. It corresponds to the *centroid* of the distribution, whose coordinates are defined by

$$\hat{\tau}_f(\eta, t) := -\partial_\eta \arg V_f^g(\eta, t)$$
$$\hat{\omega}_f(\eta, t) := \eta + \partial_t \arg V_f^g(\eta, t).$$

Both quantities, which *locally* define an instantaneous frequency and a group delay, enable perfect localization of linear chirps [3]. An efficient procedure computes them according to:

$$\hat{\tau}_f(\eta, t) = t + \Re\left\{\frac{V_f^{tg}(\eta, t)}{V_f^g(\eta, t)}\right\}, \qquad (4)$$

$$\hat{\omega}_f(\eta, t) = \eta - \Im\left\{\frac{V_f^{g'}(\eta, t)}{V_f^g(\eta, t)}\right\} \qquad (5)$$

where tg stands for the function $tg(t)$ and $\Re\{Z\}$ (resp. $\Im\{Z\}$) is the real (resp. imaginary) part of the complex number Z.

3.2. Definition of Contour Points

Detecting ridge points and linking them to build smooth contours is a challenging problem that has already been considered in the STFT and wavelet settings [6]. Let us here present the technique introduced in [5] to define ridges in the TF plane. It first considers the reassignment vector (RV) defined by:

$$RV(\eta, t) = \begin{pmatrix} \hat{\tau}_f(\eta, t) - t \\ \hat{\omega}_f(\eta, t) - \eta \end{pmatrix}. \qquad (6)$$

Using that formalism, ridge points correspond to locations where RV changes sign rapidly, which happens when one crosses a ridge, corresponding to the instantaneous frequency of a mode. However, to find out those crossings in a discrete setting is rather unstable, therefore one prefers to project RV onto a specific direction and then determines the location of the change of sign of the projection. These points are called *contour points* (CPs), and were first introduced in [5]. In practice, they are defined as the zeros of the inner product of RV

with a unitary vector of orientation θ as explained hereafter. Indeed, assuming that the window g is Gaussian with unit variance, then one has $V_f^{tg}(\eta, t) = -V_f^{g'}(\eta, t)$, and RV reads

$$\left(\Re\left\{\frac{V_f^{tg}(\eta, t)}{V_f^g(\eta, t)}\right\}, \Im\left\{\frac{V_f^{tg}(\eta, t)}{V_f^g(\eta, t)}\right\}\right). \qquad (7)$$

CPs associated with angle θ are then said to satisfy:

$$\Im\left\{\frac{V_f^{tg}(\eta, t)}{V_f^g(\eta, t)}e^{-i\theta}\right\} = 0. \qquad (8)$$

One can then remark that expression (8) corresponds to the zero of an inner product since:

$$\Im\left\{\frac{V_f^{tg}(\eta, t)}{V_f^g(\eta, t)}e^{-i\theta}\right\} = \langle RV(\eta, t), v_{\frac{\pi}{2}+\theta}\rangle, \qquad (9)$$

where v_λ is the unit vector in the direction λ. CPs thus correspond to the points where the inner product of RV with the unit vector in the direction $\frac{\pi}{2} + \theta$ changes signs.

They can also be related to the differential of $\log|V_f^g(\eta, t)|$ in the direction $v_{\frac{\pi}{2}+\theta}$. Indeed, when STFT using a Gaussian window with unit variance is computed on a linear chirp, we get [3]:

$$RV(\eta, t) = \nabla \log|V_f^g(\eta, t)| \qquad (10)$$

where ∇ is the operator $(\partial_t, \partial_\eta)$, so that (9) rewrites:

$$\langle RV(\eta, t), v_{\frac{\pi}{2}+\theta}\rangle = \nabla \log|V_f^g(\eta, t)|^T v_{\frac{\pi}{2}+\theta}.$$

CPs are thus associated with some predefinite angle θ. It has been noted [5] that the best results are obtained when θ corresponds to the direction of the ridge, meaning $\theta + \frac{\pi}{2}$ roughly corresponds to the orientation of RV.

Once CPs are determined, authors in [5] chain them by considering the contours of level zero of the function defined in (9) (contourc MATLAB function). Then, the contours are segmented depending on the zeros of the spectrogram, and finally ordered by decreasing energy (corresponding to the sum of the squared spectrogram coefficients along the chain). One ends up with a series of ridges corresponding to modes with deacreasing energy, that can finally be reconstructed in the temporal domain.

3.3. Case of a non-unitary window

Since RV points towards a ridge (see (10)), (9) keeps a fixed but different sign on each side of a ridge provided $v_{\frac{\pi}{2}+\theta}$ is not in the direction of the ridge. Then, if one uses a Gaussian window with standard deviation σ, one shall consider the following renormalisation of STFT:

$$\tilde{V}_f^g(\eta, t) = V_f^g(\frac{\eta}{\sigma}, \sigma t), \qquad (11)$$

and then the following generalisation of vector (7)

$$R_\sigma(\eta,t) = \left(\Re\left\{ \frac{\tilde{V}_f^{tg}(\eta,t)}{\tilde{V}_f^{g}(\eta,t)} \right\}, \Im\left\{ \frac{\tilde{V}_f^{tg}(\eta,t)}{\tilde{V}_f^{g}(\eta,t)} \right\} \right) \quad (12)$$

which is no longer equal to RV. However, it shares the nice property with RV computed using a Gaussian window of unit variance to correspond to the gradient of the logarithm of the modulus of the renormalised STFT. Indeed, remarking that $V_f^{tg}(\eta,t) = -\sigma^2 V_f^{g'}(\eta,t)$, one has:

$$\left(\frac{\partial_t|\tilde{V}_f^{g}(\eta,t)|}{|\tilde{V}_f^{g}(\eta,t)|}, \frac{\partial_\eta|\tilde{V}_f^{g}(\eta,t)|}{|\tilde{V}_f^{g}(\eta,t)|} \right)^T = \frac{1}{\sigma} \frac{\tilde{V}_f^{tg}(\eta,t)}{\tilde{V}_f^{g}(\eta,t)}.$$

From this, we deduce that:

$$R_\sigma(\eta,t) = \sigma\nabla\log|\tilde{V}_f^{g}(\eta,t)|. \quad (13)$$

To interpret this result, one can finally note that denoting g_1 the Gaussian window with unit variance:

$$\tilde{V}_f^{g}(\eta,\tau) = \int_{\mathbb{R}} f(t)g(t - \sigma\tau)e^{-i\frac{\eta}{\sigma}t}dt$$
$$= \int_{\mathbb{R}} f(\sigma t)g_1(t - \tau)e^{-i\eta t}dt.$$

So, when f is a linear chirp, $R_\sigma(\eta,t)$ is orthogonal to the ridge associated with the chirp $f(\sigma t)$, then one needs to renormalise $R_\sigma(\eta,t)$ as follows to get a vector orthogonal to $f(t)$:

$$\tilde{R}_\sigma(\eta,t) = \left(\frac{1}{\sigma}\Re\left\{ \frac{\tilde{V}_f^{tg}(\eta,t)}{\tilde{V}_f^{g}(\eta,t)} \right\}, \sigma\Im\left\{ \frac{\tilde{V}_f^{tg}(\eta,t)}{\tilde{V}_f^{g}(\eta,t)} \right\} \right) \quad (14)$$

so that the contours associated with angle θ can be defined by:

$$\langle \tilde{R}_\sigma(\eta,t), v_{\frac{\pi}{2}+\theta} \rangle = 0. \quad (15)$$

Note that due to the renormalisation factor, the formula giving the contours can no longer be defined using the imaginary part of some complex number.

4. ADAPTIVE COMPUTATION OF CONTOURS AND BASINS OF ATTRACTION

4.1. Adaptive contour estimation

The previous approach requires the direction θ to be given a priori and is therefore not well suited to determine contours with varying orientations. As already noticed, vector $\tilde{R}_\sigma(\eta,t) = r_\sigma(\eta,t)e^{i\theta_\sigma(\eta,t)}$ is oriented perpendicularly to the ridge of f (if f is a linear chirp) in its vicinity and changes orientation when one moves away from it due to the presence of noise. Instead of imposing an orientation θ to compute the contours as in (15), we define these as the points where:

$$\alpha(\eta,t) := \langle \tilde{R}_\sigma(\eta,t), v_{\theta_\sigma(\eta,t) \mod \pi} \rangle = 0 \quad (16)$$

with $(\theta_\sigma(\eta,t) \mod \pi) \in [0,\pi[$.

The rationale behind this formula is that $\alpha(\eta,t)$ corresponds to the signed magnitude of the (renormalised) RV: its magnitude equals $r_\sigma(\eta,t)$ and it has a different sign on each side of the ridge. More precisely, one can check that if the slope of the ridge is positive (resp. negative) $\alpha(\eta,t)$ is negative (resp. positive) above the ridge and positive (resp. negative) under. This way, one defines a new type of CPs that no longer depend on a fixed angle θ and that are chained together to make up ridges the same way as with a fixed θ.

4.2. Determination of basins of attraction using RV

We consider here an alternative way of defining the basin of attraction of a ridge, i.e. the set of coefficients associated to a given contour. This is a necessary step for a full segmentation of the time-frequency plane, and the reconstruction of all the modes making up the signal.

Since RV points towards a ridge in its vicinity, it is natural to determine the basin of attraction of a given ridge as the set of points such that RV points towards that ridge. Note that here we no longer need to consider the renormalised version of $RV(\eta,t)$. Because the localization property of RV is only valid for linear chirps, and also because of the presence of noise, the RV does not point exactly to a ridge. We thus propose to associate to a given coefficient (η,t) the closest ridge to where its RV points, i.e. the closest ridge of point $(\hat{\tau}_f(\eta,t), \hat{\omega}_f(\eta,t))$. An illustration of this procedure is given on Figure 1 (bottom), on a sinusoidal phase signal.

4.3. New mode reconstruction procedure based on basins of attraction

We here investigate two different reconstruction techniques, based on two synthesis formula for STFT and on the knowledge on the basins of attraction. Let $\mathcal{B}_i \subset \mathbb{R}^2$ be the basin of attraction associated with ridge i, then a local reconstruction technique of mode f_i corresponding to ridge i can be achieved by:

$$f_i(t) = \frac{1}{g(0)} \int_{(t,\eta)\in\mathcal{B}_i} V_f^{g}(\eta,t)d\eta. \quad (17)$$

This reconstruction technique will be denoted by R_1 in the following. A second integration technique consists in integrating the information on the whole basin of attraction by considering the more standard formula:

$$f_i(\tau) = \int\int_{\mathcal{B}_i} V_f^{g}(\eta,t)g(\tau - t)e^{2i\pi\eta(\tau-t)}dtd\eta. \quad (18)$$

This reconstruction technique will be denoted by R_2 in the following. The main difference between the two synthesis formulae is that equation (17) is pointwise and does not guarantee a smooth reconstruction (see [7] for a more detailed explanation).

5. NUMERICAL RESULTS

5.1. Basin of attraction determination and mode reconstruction for synthetic data

We first consider a toy example consisting of the AM/FM signal represented in Figure 1 (top), with an additive white Gaussian noise (input SNR = 17 dB). Figure 1 (bottom) shows the contours and basins associated with this example, with $\sigma = 0.04$. We then perform the reconstruction by selecting the coefficients associated with the first basin (with maximal energy), and applying synthesis formula (17) or (18). The SNR after reconstruction (output SNR) is 28 and 29 dB respectively.
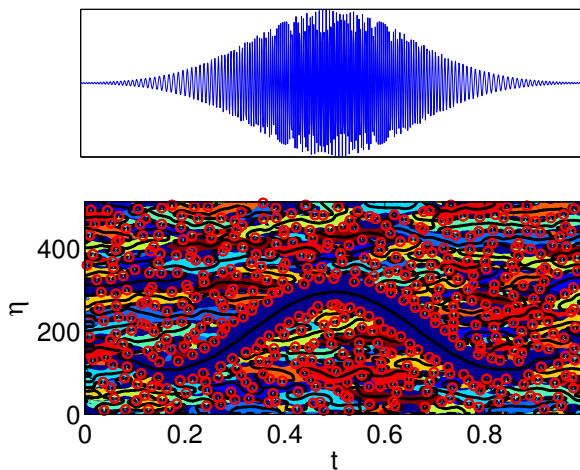

Fig. 1. top: the noise-free toy signal. bottom: the contours (black lines), basins of attraction (colored areas) and the zeros of the spectrogram (red dots).

Since the choice of the window's size is a common problem in TF analysis, we now investigate the sensitivity of the reconstruction procedure to that parameter. To this end, we carry out the same test with different values for σ : 0.02 and 0.06. The basins of attraction and ridges are depicted on Figure 2, showing that the size of the window alters the structure of the basins. However, we observe that the value of σ has almost no influence on the reconstruction results. On the one hand, it is expected that the fact the ridge is not detected close to time $t = 0s$ and $t = 1s$ when $\sigma = 0.06$ does not affect the mode reconstruction since the signal has very low amplitude at these times. On the other hand, we note that the width of the basin of attraction is considerably reduced when σ is large and when the modulation is low but this has very little incidence on the quality of the mode reconstruction.

Now, we aim to quantitatively evaluate the performance of methods R_1 and R_2 in a noisy context. In this regard, Figure 3 shows the output SNR as a function of input SNR, for the same synthetic signal and for $\sigma = 0.04$. On that figure, we also compare the two reconstruction methods with a TF-
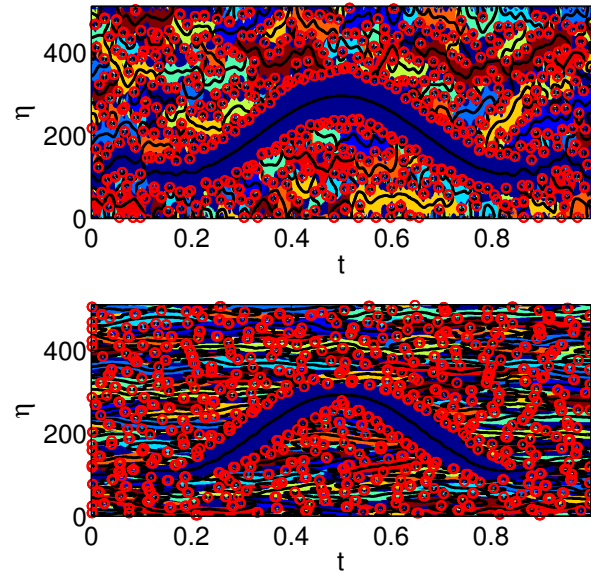

Fig. 2. top: basin of attraction computed with $\sigma = 0.02$, the ouput SNR R_1 and R_2 methods equals 26 and 28 dB, respectively. bottom: idem but with $\sigma = 0.06$, the output SNR is 27 and 29 dB respectively.

based denoising method named block-thresholding [8]. These results show good denoising performance of both proposed reconstruction techniques, R_2 behaving slightly better. The contour-based approach outperforms the block-thresholding of a few decibels, on a wide range of input SNR. Note that for denoising purpose the reconstruction from the basins of attractions may not be optimal, and can be improved if we select a strip of coefficients around the ridge with a width adapted to the noise level, such as in [9].
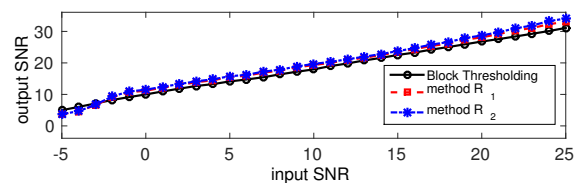

Fig. 3. Analysis of denoising performance of the proposed approach, and comparison with the block-thresholding.

5.2. Application to real data

Finally, we test our method on a real bat echolocation signal, made of 400 samples recorded at 143 Hz, to which we add a white Gaussian noise with input SNR of 8.0 dB. Figure 4 (top) shows the spectrogram of the signal, then the contours and basins of attraction are depicted underneath. The three contour chains corresponding to the largest energies are then

used to reconstruct the modes. The three main components of the echolocation signal seem to be well estimated. Figure 4 also shows the reconstruction of the three modes using method R_2, and compares the final reconstruction obtained by summing those three modes, with the original noise-free signal. The output SNR of the final reconstruction being 11.5 dB, it means that while estimating the modes the algorithm performs some kind of denoising.
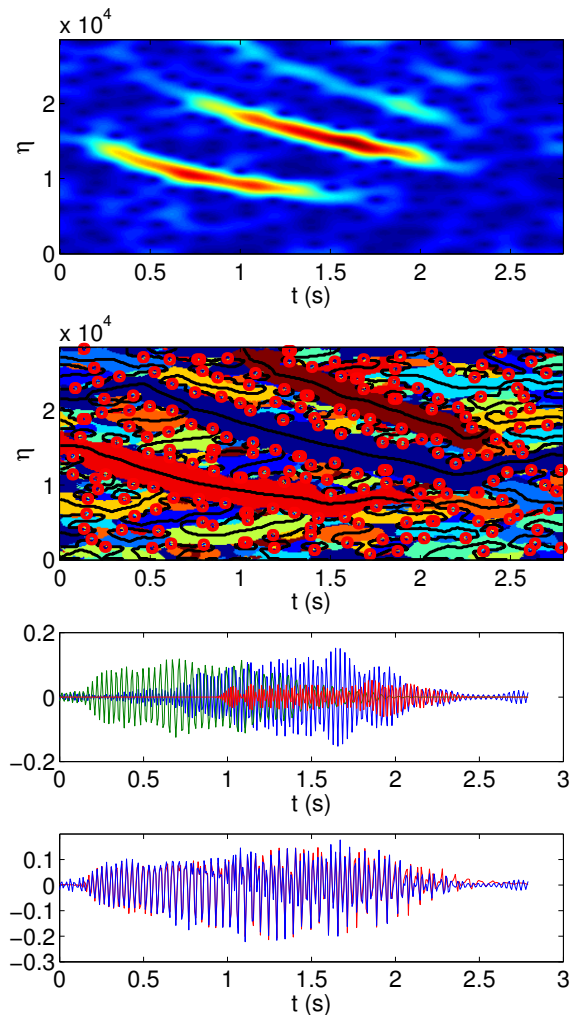

Fig. 4. From top to bottom: the spectrogram of the noisy bat signal, ridges and basins of attraction, three reconstructed modes, original noise-free signal (red) and the reconstruction (blue).

6. CONCLUSION

In this paper, we have presented a new adaptive algorithm to estimate some structure called contours associated with TF representation of multicomponent signals. This algorithm is based on an extensive use of the reassignment vector and is fully adaptive compared with previously proposed approach proposed in [5]. Having determined the basins of attraction associated with some kind of contours called ridges, we have proposed a new technique for mode reconstruction. An interesting aspect of the proposed ridge detector is that it does not require any assumption on the co-existence of a certain number of modes at a given time t. This will be further exploited in a future work.

REFERENCES

[1] K. Kodera, C. De Villedary, and R. Gendrin, "A new method for the numerical analysis of non-stationary signals," *Phys. Earth Plan. Inter.*, vol. 12, no. 2-3, pp. 142–150, 1976.

[2] K. Kodera, R. Gendrin, and C. Villedary, "Analysis of time-varying signals with small BT values," *IEEE Trans. Acoust., Speech and Sig. Proc.*, vol. 26, no. 1, pp. 64–76, 1978.

[3] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *Signal Processing, IEEE Transactions on*, vol. 43, no. 5, pp. 1068–1089, 1995.

[4] I. Daubechies, J. Lu, and H.T. Wu, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Applied and Computational Harmonic Analysis*, 2010.

[5] Y. Lim, B.G. Shinn-Cunningham, and T.J. Gardner, "Sparse contour representations of sound," *IEEE Sig. Proc. Let.*, vol. 19, no. 10, pp. 684–687, 2012.

[6] N. Delprat, B. Escudie, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torresani, "Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies," *Information Theory, IEEE Transactions on*, vol. 38, no. 2, pp. 644–664, 1992.

[7] T. Oberlin, S. Meignen, and V. Perrier, "On the mode synthesis in the synchrosqueezing method," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. IEEE, 2012, pp. 1865–1869.

[8] G. Yu, S. Mallat, and E. Bacry, "Audio denoising by time-frequency block thresholding," *Signal Processing, IEEE Transactions on*, vol. 56, no. 5, pp. 1830–1839, 2008.

[9] T Oberlin, S Meignen, and S McLaughlin, "A novel time-frequency technique for multicomponent signal denoising," in *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*. IEEE, 2013, pp. 1–5.