Unsupervised classification of the spectrogram zeros with an application to signal detection and denoising

Juan Miramont, François Auger, Marcelo Colominas, Nils Laurent, Sylvain Meignen

Unsupervised Classification of the Spectrogram Zeros with an Application to Signal Detection and Denoising

Juan M. Miramont, François Auger, Marcelo A. Colominas, Nils Laurent, and Sylvain Meignen

Abstract

Spectrogram zeros, originated by the destructive interference between the components of a signal in the time-frequency plane, have proven to be a relevant feature to describe the time-varying frequency structure of a signal. In this work, we first introduce a classification of the spectrogram zeros in three classes that depend on the nature of the components that interfere to produce them. Then, we describe an algorithm to classify these points in an unsupervised way, based on the analysis of the stability of their location with respect to additive noise. Finally, potential uses of the classification of zeros of the spectrogram for signal detection and denoising are investigated, and compared with other methods on both synthetic and real-world signals.

Keywords: Zeros of the spectrogram, time-frequency analysis, non-stationary signals, noise-assisted methods.

1 Introduction

Time-frequency (TF) representations, such as the well-known *spectrogram* [18], are powerful tools to reveal the time-varying frequency structure of a signal. A common interpretation of the spectrogram is that of a TF energy *distribution*, a conception that naturally leads to consider that more information about the signal is located where the energy is highly concentrated in the TF plane [1, 9, 10, 23, 36].

Nevertheless, a paradigm-shifting idea was proposed in [19], where spectrogram zeros (SZs) were shown to play a central role to characterize the TF structures associated with different types of signals [19, 20]. In addition, Bardenet et al. [4, 5] provided a thorough mathematical description of the distribution of the SZs for complex white Gaussian noise, establishing promising links between TF analysis and point processes, further motivating the study of the SZs [22] and even the zeros of other signal representations [27, 33].

In this paper, we shall focus on the SZs and their applications. In particular, the SZs have been the basis of some signal denoising and signal detection approaches. For instance, strategies aiming to disentangle signal and noise, using the SZs in order to approximate the signal domain, were introduced in [19] and [4]. However, as we will show later in the paper, these SZs based approaches face some limitations when the signal has several components close to each other in the TF plane, mainly due to the SZs that appear between the components.

In addition to signal denoising, authors in [4] also deal with signal detection, introducing Monte Carlo based tests to determine whether a signal is present in a given noisy time series. These tests, however, suffer from a low true positive rate when the signal-to-noise ratio (SNR) is also low.

In this paper, we make three main contributions. First, we propose a novel approach to automatically classify the SZs, after having explained that they can be thought as belonging to three different classes, namely SZs created either by the interference between signal components, by the interaction between signal and noise, or by interference of pure noise components. Second, based on this classification, we derive a method for signal detection. Finally, we develop a novel strategy for signal denoising which exploits the proposed classification scheme.

In order to classify SZs in the three proposed categories, we follow an unsupervised strategy inspired by noise-assisted techniques. In a nutshell, we add noise with a specific variance to a signal, and explore how this alters the position of the SZs of the resulting mixture of signal and noise, to conclude that their location varies quite

---

Juan M. Miramont and François Auger are with Nantes Université, Institut de Recherche en Énergie Électrique de Nantes Atlantique (IREENA, UR 4642), F-44600 Saint-Nazaire, France (email: juan.miramont@univ-nantes.fr and francois.auger@univ-nantes.fr).

Marcelo A. Colominas is with the Institute of Research and Development in Bioengineering and Bioinformatics (IBB, UNER - CONICET), and with the Faculty of Engineering (UNER), Oro Verde, Entre Rios, Argentina (email: macolominas@conicet.gov.ar).

Nils Laurent and Sylvain Meignen are with Jean Kuntzmann Laboratory, University of Grenoble-Alpes and CNRS UMR 5224, F-38401 Grenoble, France (email: nils.laurent1@univ-grenoble-alpes.fr and sylvain.meignen@univ-grenoble-alpes.fr).

differently depending on the nature of the SZs of the original noisy signal. This enables us to find relevant features to classify SZs into the three different categories by means of a clustering algorithm.

The classified SZs are then the basic ingredients to, on the one hand, a signal detection algorithm and, on the other hand, a signal denoising strategy that uses the Delaunay triangulation of SZs, but in a different way as in [19]. To highlight the relevance of our contributions, we then compare the performance of the proposed techniques with that of other SZs based methods using synthetic and real-world signals.

The rest of the paper is organized as follows. In Secs. 2 and 3 we give some relevant definitions and outline the state-of-the-art. In Sec. 4 we describe the three proposed types of SZs, and explain that we aim to classify them using their stability under additive noise. In Secs. 5 and 6, we show how to classify the SZs using an unsupervised approach based on a stability analysis. In Sec. 7, we introduce the novel approaches for signal detection and denoising based on SZs classification, while, in Sec. 8, we illustrate the performance of these techniques on both simulated and real signals. Finally, we comment on the obtained results in Sec. 9 and draw some conclusions in Sec. 10.

2 Context

2.1 Short-Time Fourier Transform

The short-time Fourier transform (STFT) is a fundamental tool for TF signal analysis. Given a real signal $x \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, its STFT is defined by

$$V_x^g(t, f) := \int_{-\infty}^{+\infty} x(u)\, g(u - t)\, e^{-i2\pi f u} du, \tag{1}$$

where $t, f \in \mathbb{R}$ are, respectively, the time and frequency variables, and $g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is the analysis window. In the context of this paper, we restrict ourselves to Gaussian windows defined by:

$$g(t) = \frac{2^{1/4}}{\sqrt{\sigma}}\, e^{-\frac{\pi t^2}{\sigma^2}}, \tag{2}$$

with $\sigma > 0$, a window width parameter. The spectrogram of x is then defined as $S_x^g(t, f) := |V_x^g(t, f)|^2$. Henceforth, we will use $\sigma = 1$, corresponding to the *circular* Gaussian window, i.e. with equal supports in time and frequency [21].

Following closely the work done in [4, 17, 19], a SZ corresponds to a local minimum on a 3×3 grid centered at each point of the (discrete) spectrogram $S[n, m] \approx S(nT_s, \frac{m}{MT_s})$, where $n = 0, ..., N - 1$ (resp. $m = 0, ..., M - 1$) is the time (resp. frequency) discrete variable, and T_s is the sampling period.

2.2 Signal and Noise Mixtures

In the following, we shall consider signal and noise mixtures, in which the noise $\xi(t)$ is a zero-mean white Gaussian noise (WGN), namely distributed as $\mathcal{N}(0, \gamma_0^2)$ for all values of t, satisfying:

$$\mathbb{E}\{\xi(t)\xi(t - \tau)\} = \gamma_0^2 \delta(\tau), \tag{3}$$

where γ_0^2 is the noise variance and $\delta(t)$ is the Dirac distribution. We shall express the SNR between a deterministic signal x and $\xi(t)$ as:

$$\mathrm{SNR}(x, \xi) = 10 \log_{10}\left(\frac{P_x}{\gamma_0^2}\right) \text{ (dB)}, \tag{4}$$

where P_x is the power of the signal.

3 Signal Denoising Based on Spectrogram Zeros

The SZs of complex white Gaussian noise (CWGN) are homogeneously distributed in the TF plane [4,5] (see [4, Sec. 3.1] for a rigorous definition of CWGN). In the case of WGN, the distribution of the SZs in the TF plane is equivalent to that of a CWGN for sufficiently large frequencies [4]. As an example, Fig. 1 (a) shows the distribution of SZs for a WGN realization. These SZs are, by far, the most studied ones. In particular, the lattice-like structure they

generate in the TF plane is the basis to some recently developed denoising methods [4, 19, 22]. When adding a signal to a noise realization, like the sinusoidal chirp shown in Fig. 1 (b), one remarks that its presence only affects SZs *locally* [4, 19]. This is shown in Fig. 1 (c), where the energy corresponding to the signal appears as a sinusoidal trajectory in the TF plane, and the SZs outside the signal domain are not modified. This can be seen by comparing, for example, the position of SZs in Figs. 1 (a) and (c).
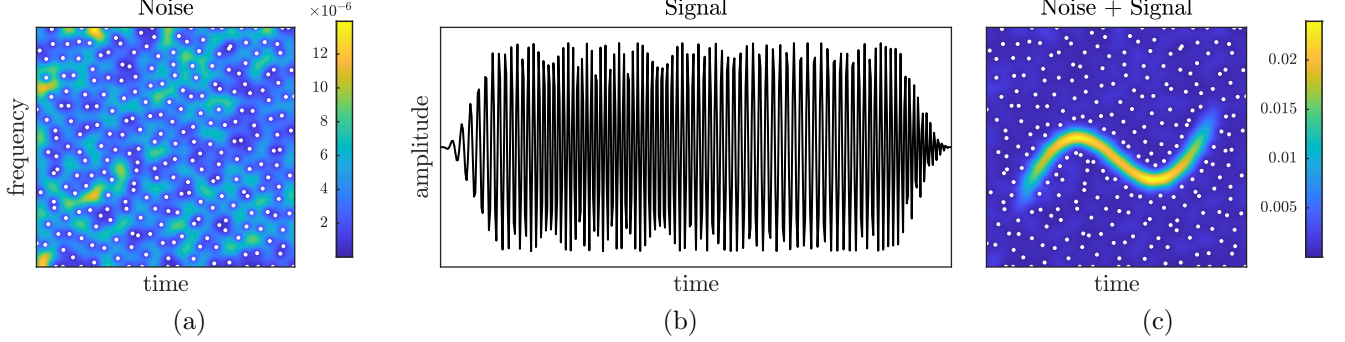

Figure 1: (a) Spectrogram of a white Gaussian noise. The white dots mark the positions of the spectrogram zeros. The yellow (resp. blue) areas represent the regions with high (resp. low) energy. (b) A sinusoidal chirp. (c) Spectrogram of the mixture of the same noise realization used in (a) and the sinusoidal chirp shown in (b) with SNR = 15 dB.

Based on this, a signal denoising strategy was introduced in [19, 20], and summarized in the following steps:

1. Compute the spectrogram of x and detect SZs.

2. Compute the Delaunay triangulation based on SZs.

3. Find triangles with an edge longer than a threshold $\ell_{\max}$.

4. Gather these triangles together to obtain an estimate of the signal domain in the TF plane, denoted by $\mathcal{D}_s$.

5. Compute an estimation of the noiseless signal using

$$\tilde{x}(t) = \frac{1}{g(0)} \int_{-\infty}^{+\infty} V_x^g(t, f) \mathbb{1}_{\mathcal{D}_s}(t, f) e^{i2\pi f t} df \tag{5}$$

where $\mathbb{1}_{\mathcal{D}_s}(t, f)$ is the indicator function of $\mathcal{D}_s$.

From now on, we refer to this approach as the *DT method*. The rationale behind the DT method is that the presence of a signal creates larger regions without SZs than in the noise-only case. Therefore, the Delaunay triangles corresponding to the signal are those with at least one edge longer than some threshold $\ell_{\max}$, the remaining triangles being then considered as related to noise. The value of $\ell_{\max}$ is determined by studying the Delaunay triangulation for noise only, so that a triangle with an edge longer than $\ell_{\max}$ is not generated by noise with a high probability. For instance, considering the circular Gaussian window, $\ell_{\max} = 2.0$ [19] is a satisfactory threshold. With such a choice for $\ell_{\max}$, the procedure described in points 4) and 5) works well to reconstruct monocomponent signals in the presence of some mild noise, but some limitations appear when dealing with multicomponent signals. Indeed, consider that x is made of three pure tones, namely

$$x(t) = s_1(t) + s_2(t) + s_3(t), \tag{6}$$

with $s_i(t) = \cos(2\pi f_i t)$, and $f_1 < f_2 < f_3$. As the three components interfere, SZs are created between the modes, which are located at specific time instants which depend only on the difference between the frequency of the components [24, 25]. Let us define the separation between adjacent tones as $\Delta f_{1,2} = f_2 - f_1$ and $\Delta f_{2,3} = f_3 - f_2$, and assume that

$$\Delta f_{1,2} = \Delta f_{2,3} = \frac{\theta}{\sqrt{2\pi}}, \tag{7}$$

where $\frac{1}{\sqrt{2\pi}}$ is the standard deviation of the Fourier transform of the circular Gaussian window g, and $\theta \in \mathbb{R}$ defines the frequency separation.

3

In this specific case, the SZs created by the interference between the first and second modes have the same time coordinates as the ones generated by the second and third modes, thus producing orthogonal Delaunay triangles associated with the second mode in the TF plane (see Fig. 2 (b)). Let us now take a closer look at these triangles. Neglecting the interaction between the STFTs of non-adjacent tones, successive SZs corresponding to interference between two modes are separated by a time-lapse of $\frac{\sqrt{2\pi}}{\theta}$ [31], while, as the studied modes have the same amplitude, the SZs created by the interference between modes are located at frequency $\frac{f_1+f_2}{2}$ and $\frac{f_2+f_3}{2}$. Thus, a triangle overlapping the second mode has a vertical edge of length $\frac{\theta}{\sqrt{2\pi}}$ and an horizontal edge of length $\frac{\sqrt{2\pi}}{\theta}$, the length of the hypotenuse being $\sqrt{\frac{\theta^2}{2\pi} + \frac{2\pi}{\theta^2}}$. Taking $\theta = 3$, one obtains a length $\ell \approx 1.46$ for the hypotenuse. Thus, a threshold $\ell_{\max} < 1.46$ must be used in order to select the triangles associated with the second mode as part of the signal domain. An illustration is proposed in Fig. 2, where the triangles corresponding to $\mathcal{D}_s$ are selected using two different thresholds, one below 1.46 and one above. As expected using $\ell_{\max} > 1.46$ one misses the middle tone (Fig. 2 (a)), whereas with $\ell_{\max} = 1.30$ the entire signal is recovered, but more triangles in the noise region are considered as part of $\mathcal{D}_s$ (Fig. 2 (b)). Consequently, no value for $\ell_{\max}$ can properly separate signal and noise domains in that case.
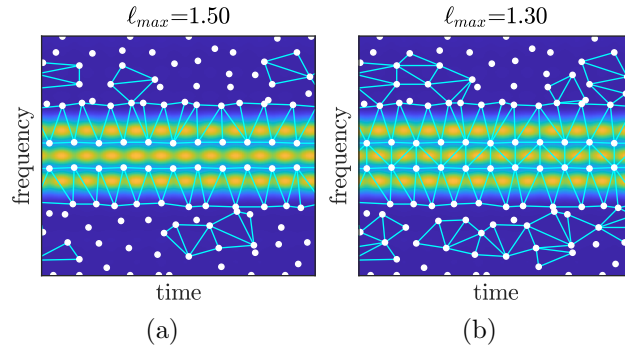

Figure 2: The two panels show the spectrogram of the signal described in Eq. (6), with the SZs and the selected triangles superimposed. (a) Triangles selected by applying a threshold of $\ell_{\max} = 1.50$. (b) Triangles selected by applying a threshold of $\ell_{\max} = 1.30$.

This very simple example illustrates how interference between close components can challenge the DT method. Note that the same limitations would arise with the method introduced in [4], which considers disks equivalent to the circumcircles of the Delaunay triangles rather than the triangles themselves. One of the motivations of the paper will thus be to introduce a different approach to signal denoising, still using the Delaunay triangulation but based on a careful analysis of the nature of SZs.

4   Three Kinds of Spectrogram Zeros

In this section, we first explain that SZs are of three different kinds, and, then, introduce a principle to determine these three categories, based on the stability of SZs with respect to additional noise.

4.1   Characterizing SZs

To start with, we notice that SZs can result from three kinds of interference:

1. Signal-Signal Zeros (SS): these zeros are generated by the interference between signal components.

2. Signal-Noise Zeros (SN) : these are produced by the interference between signal components and noise, and surround the signal domain.

3. Noise-Noise Zeros (NN): these zeros are generated by the noise only and can be viewed, as explained later, as the result of interference between randomly located Gaussian *logons*.

To better describe the origin of SZs, we resort to the *destructive interference* between components in the TF plane [19, 21]. Indeed, the conditions for SZs to appear are given by the following:

Proposition 4.1: *Let us consider the spectrogram of the sum of x_1 and x_2 [24, 25]:*

$$S^g_{x_1+x_2}(t,f) = S^g_{x_1}(t,f) + S^g_{x_2}(t,f) + 2|V^g_{x_1}(t,f)V^g_{x_2}(t,f)|\cos\left(\Phi^g_{x_2}(t,f) - \Phi^g_{x_1}(t,f)\right), \quad (8)$$

where $\Phi^g_x(t,f)$ is the phase of $V^g_x(t,f)$. Then, $S^g_{x_1+x_2}(t,f) = 0$ if and only if:

1. *$\Phi_{x_1}(t,f)$ and $\Phi_{x_2}(t,f)$ differ by an odd factor of π.*

2. *The modulus of $V^g_{x_1}(t,f)$ and $V^g_{x_2}(t,f)$ are equal.*

The proof being straightforward, it is left to the reader. In the case of SS zeros, x_1 and x_2 in Prop. 4.1 are two signal components. These zeros are therefore *deterministic*, being only dependent on the signal components in the TF plane. Also, for multicomponent signals, such SZs are always located between the components.

Being partly dependent on the noise, the location of SZs of type SN cannot be precisely given. However, considering that x_1 is a signal and x_2 is a WGN noise realization, one can approximate the location of these SZs by taking the expected value of the spectrogram of noise (constant and equal to γ_0^2 in the case of a WGN [18, 21]), and replacing it in condition 2 of Prop. 4.1:

$$S^g_{x_1}(t,f) = \gamma_0^2. \quad (9)$$

Equation (9) then defines a level curve

$$\Gamma = \left\{(t,f): \ S^g_{x_1}(t,f) = \gamma_0^2\right\}. \quad (10)$$

SZs of SN type are very close to that curve, as illustrated in Fig. 3, where the dashed line corresponds to Γ. The location of SN and SS zeros are thus constrained differently.
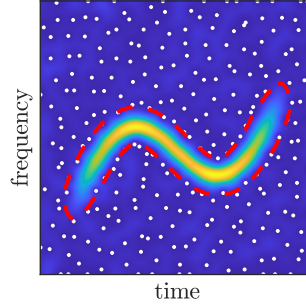

Figure 3: The same spectrogram as the one shown in Fig. 1(b) with the superposition of the curve Γ defined in Eq. (9), in dashed red line.

SZs of NN kind can also be related to Prop. 4.1 by recalling that a CWGN can be modeled as the sum of multiple, randomly-located Gaussian atoms $h_k(t) = a_k \, g(t - t_k) \, e^{i(2\pi f_k + \varphi_k)}$, also called *logons*, with random amplitude a_k and position (t_k, f_k), and phase φ_k [21]. The spectrogram of a CWGN can then be expressed as [21, Chapter 13]:

$$S^g_\xi(t,f) = \left|\sum_k V^g_{h_k}(t,f)e^{i\varphi_k}\right|^2. \quad (11)$$

Note that, in this model, the logons are randomly located, but not arbitrarily so, because of the restrictions imposed by the reproducing kernel of the STFT, which gives rise to the *honeycomb* structure of spectrogram maxima and zeros [21]. SZs of NN kind can then be explained by Prop. 4.1 considering x_1 and x_2 as a pair of logons from the mixture given in Eq. (11).

Remark. Following Prop. 4.1, SS zeros must appear between components for any multicomponent signal. In practice, however, two components separated in frequency by a distance longer than the essential support of the Fourier transform of g can be considered independent [18, 21]. In such a case, the third term in the right hand of Eq. (8) is usually negligible when compared with the spectrogram of noise, hence the zeros between such two signal components are no longer of SS kind.

Note that SZs of SS type were not considered in previous signal denoising approaches such as the DT method, and constitute one of the main contributions of this article. The usefulness of SZs classification to improve the DT method will be exemplified in detail later.

4.2 Stability to Noise Addition

In the following, we hypothesize that the three kinds of SZs can be discriminated by analyzing their stability to noise addition. In this regard, we expect SZs of SS type to be the most stable, since they depend only on the location of the signal components, and higher energy components are less likely to be affected by noise. Following this logic, the least stable ones should be SZs of NN type, assuming that the noise has, on average, a lower amplitude than the signal. Finally, because they partially depend on the signal, the stability of the position of SZs of SN type should be between that of the SS and NN kinds.

Though one needs a precise criterion to measure this stability, to motivate our analysis, let us first consider the simple example of a signal composed of two parallel linear chirps, and contaminated by a WGN, the spectrogram of which is shown in Fig. 4 (a), as well as the expected classification for the different types of zeros. When a new realization of the noise is added to the (already noisy) signal, the SZs appear to be *displaced* from their original positions. This effect can be seen in Fig. 4 (b), where the circles and dots indicate the positions of the SZs after adding two different noise realizations to the original noisy signal. The SZs between the chirps, generated by the interference between the two components and hence of SS type, do not move that much under noise addition. This is not true for the other SZs of the figure: the new SZs created near the border of the signal domain, however, seem to stay close, which is not the case of SZs located in the noise-dominated regions of the TF plane. These observations make us think that, to discriminate the different types of SZs, it should be of interest to study more in detail the location of the new SZs around the original ones after noise addition.
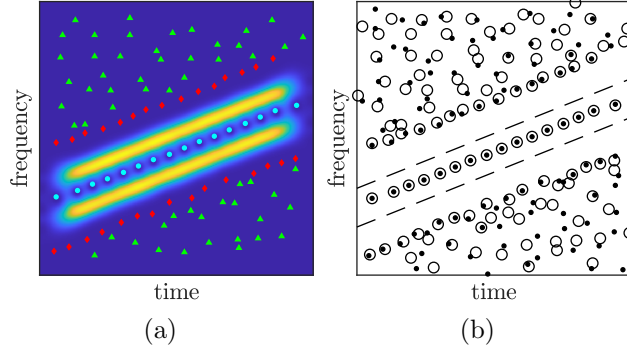

Figure 4: (a) Spectrogram of a signal with two parallel linear chirps contaminated by real WGN. Cyan dots correspond SZs of the SS kind, green triangles to those of the NN kind, while red diamonds represent SZs of the SN kind. (b) Position of SZs: Circles and dots indicate the position of the SZs corresponding to two different noise realizations. Dashed lines denote the instantaneous frequency associated with each chirp.

5 Assessing the Stability of SZs

Drawing inspiration from noise-assisted methods [11, 12, 39, 41], we propose to *add* several new noise realizations to a noisy signal to modify the position of SZs. More precisely, let us call $y(t) = x(t) + \xi(t)$ the *original* signal mixture, where $\xi(t)$ is the noise already present in the signal, and consider J *new* mixtures $y_j(t) = x(t) + \xi(t) + \eta_j(t)$, $j = 1, \ldots, J$, where $\eta_j(t)$ is a WGN realization with zero mean and variance γ_j^2. Moreover, denote by $\mathcal{Z}_0$ the set of *original* SZs of $y(t)$, as opposed to the sets $\mathcal{Z}_j$ of *new* SZs of each $y_j(t)$, $j = 1, 2, ...J$.

By computing $\mathcal{Z}_j$, $j = 1, \ldots, J$, we can then build a 2D histogram of the positions of the new SZs, by counting the number of new SZs that have fallen in a given TF point. The binning of the aforementioned 2D histogram, denoted here by $G[n, m]$, corresponds to that of the TF plane discretization. Examples of $G[n, m]$ with different values of γ_j are shown in Fig. 5, using the same signal as in Fig. 4 (a). The brighter the histogram, the higher the number of zeros that are located at a particular TF point. Fig. 5 (a) (resp. Fig. 5 (c)) shows a 2D histogram of SZs where the added noise realizations has a lower (resp. higher) variance. For comparison, Fig. 5 (b) shows the histogram corresponding to $\gamma_j^2 = \gamma_0^2$. From these examples, one can see that the value of γ_j, which we will study in detail later, directly influences the distribution of the new SZs around the original ones.
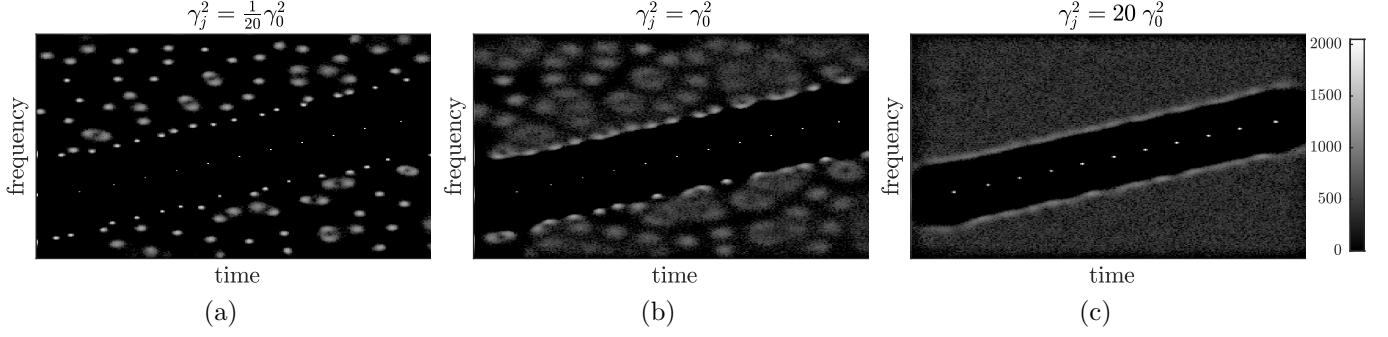
Figure 5: 2D histograms of new SZs after noise perturbation, corresponding to the signal shown in Fig. 4(a). (a): Noise realizations with a lower variance, $\gamma_j^2 = \frac{1}{20} \gamma_0^2$. (b) Added noise realizations with the same variance as the original one, $\gamma_j^2 = \gamma_0^2$. (c) Noise realizations with a higher variance $\gamma_j^2 = 20\, \gamma_0^2$.

5.1 Local Feature Extraction from Voronoï Cells

A noteworthy aspect of the histograms shown in Fig. 5 is that the distribution of new SZs in a neighborhood of an original SZ appears to vary according to what we hypothesized in Section 4.2, and hence local features accounting for these variations should be useful to discriminate between the proposed categories of SZs. To compute them, we first need to define a neighborhood of an original SZ. With this in mind, and based on the principle we applied in [28], we consider the Voronoï tessellation associated with $\mathcal{Z}_0$, which provides a partition of the TF plane that not only naturally defines a region associated with each original SZ, i.e. the Voronoï Cell (VC), but also adapts to the changes in the analysis window used to compute the STFT. In this regard, we define

$$\mathcal{V}(\mathbf{z}) = \{(n, m) : \ (n, m) \text{ is inside the VC of } \mathbf{z} \in \mathcal{Z}_0\}, \tag{12}$$

where $\mathbf{z}$ represents the discrete time and frequency coordinates of an original zero. Still considering the signal of Fig. 4 (a), Fig. 6 (a) shows the Voronoï tessellation with white lines superimposed to the 2D histogram, computed with $J = 2048$ and in which we used $\gamma_j^2 = \gamma_0^2$ (we shall discuss later how to properly set this parameter).
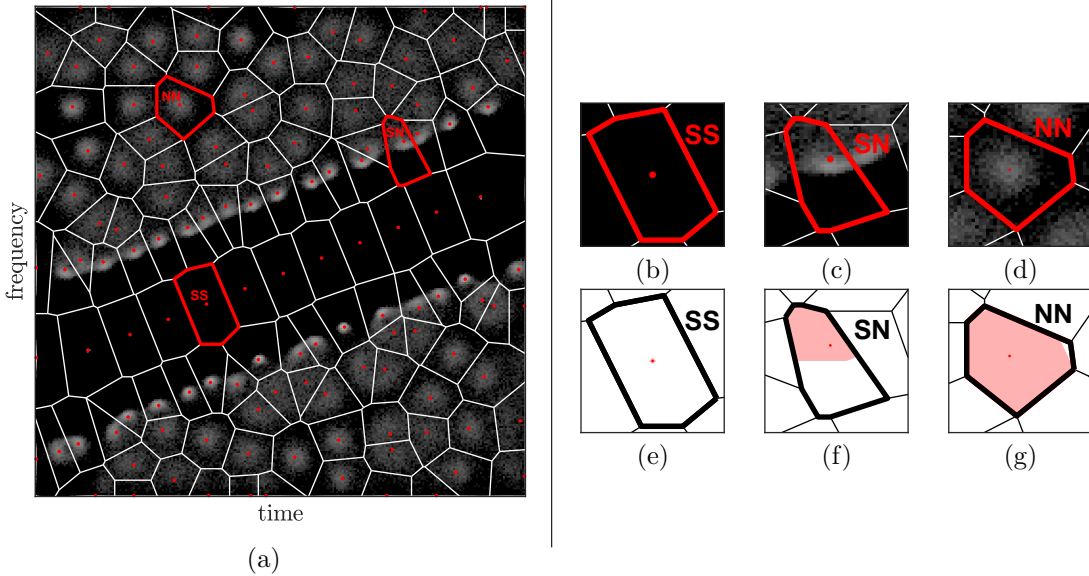

Figure 6: Panel 1: (a) 2D histogram of new SZs after noise perturbation corresponding to the signal shown in Fig. 4 (a), with $J = 2048$ and $\gamma_j^2 = \gamma_0^2$. The Voronoï tessellation is superimposed (white line segments). The red dots indicate the position of the original SZs. Panel 2, first row : (b): Detail of a Voronoï cell of a SZ of SS kind; (c): Same as (b), but for a SZ of SN kind; (d): Same as (b), but for a SZ of NN kind. Panel 2, second row: (e) The pink-colored area indicates the convex hull of the distribution of new SZs in the Voronoï cell of a SZ of SS kind. (f) Same as (e), but for a SZ of SN kind. (g): Same as (e), but for a SZ of NN kind.

Figs. 6 (b) to (d) display VCs associated with each kind of SZs. From these figures, one notices that the distribution of the new SZs after noise addition varies greatly from one type of zeros to another: it can be reduced

to a unique position (see Fig. 6 (b), SZs of SS type), cover almost all the VC (see Fig. 6 (d), SZs of NN type), or have an in-between behavior as embodied by Fig. 6 (c). Therefore, in order to estimate the support of the distribution of new SZs, we first compute the convex hull of the distribution of the new SZs within a given VC, i.e. the set of points given by $\{(n, m) : (n, m) \in \mathcal{V}(\mathbf{z}) \wedge G[n, m] > 0\}$.

An illustration of these convex hulls is given in Figs. 6 (e) to (g) for the three different types of SZs. We then derive a simple feature, by computing the ratio between the area of this convex hull, denoted by $A_{\mathrm{CH}}(\mathbf{z})$, and that of the corresponding VC, denoted by $A_{\mathrm{VC}}(\mathbf{z})$:

$$\mathrm{AR}(\mathbf{z}) = \frac{A_{\mathrm{CH}}(\mathbf{z})}{A_{\mathrm{VC}}(\mathbf{z})}. \tag{13}$$

As seen before, this area ratio varies with the kind of SZ. SZs of SS kind should have a small ratio, whereas SZs of NN kind, a ratio close to 1, since the distribution of new SZs spreads more within the VC. Finally, SZs of SN kind should be associated with a ratio between the one of SS and NN kinds. The area ratio has the advantage of being simple, and its estimation does not require many noise realizations. However, it does not provide with information about the distribution of SZs inside the convex hulls.

To account for this, we consider, as an extra feature, the normalized maximum value of the histogram inside the VC associated with a given $\mathbf{z}$:

$$\mathrm{Max}(\mathbf{z}) = \max_{(n,m) \in \mathcal{V}(\mathbf{z})} \frac{G[n, m]}{J}. \tag{14}$$

Because new SZs are more concentrated near the original SZs of the SS kind than in the other two categories, it is more likely that new zeros fall in the same TF position in such case. Therefore, the value of $\mathrm{Max}(\mathbf{z})$ should be higher for the SS kind than for the other two categories, and then SZs of SN kind should lead to higher normalized maxima than those of NN kind, for which the distribution of new SZs is nearly uniform throughout the VC (see, for example, Figs. 6 (b) to (d)).

Since the number of TF points inside the VCs is proportional to the length of the signal [5], the larger the analyzed time series, the larger the value of J needed to obtain a representative histogram. Because of this, we choose to consider the normalized maximum of the histogram as a feature, since it converges to a good estimation using a number of noise realizations much lower than the signal length. This aspect will be studied in detail in the next section.

An example of the feature space of $\mathrm{AR}(\mathbf{z})$ and $\mathrm{Max}(\mathbf{z})$ can be seen in Fig. 7 (a), where each point represents a SZ of the signal used in Fig. 4 (a). It can be seen that the SZs form three groups that correspond to the three categories of SZ previously discussed, which is illustrated in Fig. 4 (a), each group having different colors and markers. Our concern, in the next section, is to explain more in detail how the classification into the three different kinds of SZs is carried out.
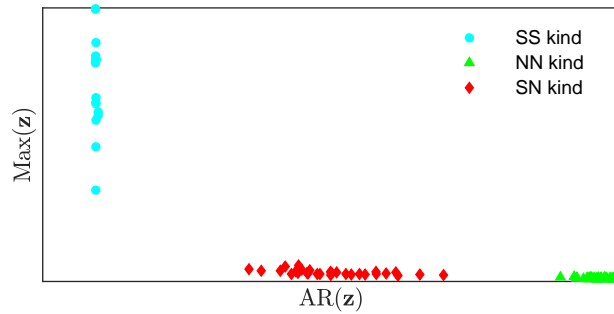
Figure 7: (a): Feature space of $\mathrm{AR}(\mathbf{z})$ and $\mathrm{Max}(\mathbf{z})$. Each point represents one SZ of the spectrogram of the parallel chirps signal displayed in Fig. 4 (a). The markers and colors of the SZs correspond to SZ shown in Fig. 4 (a).

6 Unsupervised Classification of the SZs

We now describe how to automatically classify SZs in the three categories described in Sec. 4, using the local features we just introduced.

6.1 SZs Clustering and Labeling

Based on the previous observations, one can expect SZs of the same kind to aggregate and form clusters in the two-dimensional space spanned by $\mathrm{Max}(\mathbf{z})$ and $\mathrm{AR}(\mathbf{z})$. Because one cannot guarantee that there will always be SZs of the three kinds, it is also necessary to decide whether $K = 1$ (the data is noise only, and the SZs are all of the NN kind), $K = 2$ (there is a mixture of noise and signal with no strong interference between signal components) or $K = 3$ (noise and signal with strong interference between signal components).

For clustering, we consider Gaussian mixture models (GMM) [6] and the known K-means algorithm, assuming a prescribed number of groups $K \in \mathbb{N}$. As for the automatic estimation of the number of clusters K, there exist several algorithms described in the literature able to do this (see [8, 13, 38]). In this section, we pick the GAP criterion [38] in order to illustrate one of such methods. This approach is based on finding the optimal number of clusters K as

$$K = \arg\max_{k \in \{1,2,3\}} \mathbb{E}\{\log(W_k^\star)\} - \log(W_k), \tag{15}$$

where $W_k = \sum_{r=1}^{k} \frac{1}{N_r} D_r$, in which D_r is the sum of all the pairwise distances for the points in the r-th cluster, and N_r the number of elements in the cluster. Meanwhile, $W_k^\star$ is computed from a surrogate set of points drawn from a *reference distribution*, considered here as a uniform distribution over a box aligned with the principal components of the data [38], and $\mathbb{E}\{\log(W_k^\star)\}$ is estimated by averaging over a number of Monte Carlo samples. An advantage of the GAP criterion is that it can decide whether there is only one cluster or more, which will be useful later in Sec. 7.1, where we outline a signal detection strategy. In contrast, other approaches, like the Calinski-Harabasz (CaHa) criterion [8], are designed to decide if the number of clusters is two or more, but cannot decide if there is only one group. As a disadvantage, since the GAP criterion is based on Monte Carlo simulations, its computation time is larger than other approaches. We will discuss the impact of this drawback on the overall execution time of a denoising strategy in Sec. 8.2, where we will compare it with the CaHa criterion as well.

Having estimated the number of clusters and their composition, we use the value of $\mathrm{AR}(\mathbf{z})$ to assign a label to the clusters, as discussed in Sec. 5.1. If $K = 1$, then all the zeros are assigned to the NN class. When $K = 2$ or $K = 3$, we label the clusters by sorting their $\mathrm{AR}(\mathbf{z})$ coordinate in increasing order. Then, if $K = 2$, we assign the label SN to the cluster with the lowest value of $\mathrm{AR}(\mathbf{z})$ and NN to the one with the highest value $\mathrm{AR}(\mathbf{z})$. If $K = 3$, then the cluster with the lowest value of $\mathrm{AR}(\mathbf{z})$ corresponds to the SS kind, and the remaining two groups are labeled SN and NN following the increasing value of $\mathrm{AR}(\mathbf{z})$.

6.2 Parameters Setting

The just described procedure to automatically cluster and classify the SZs depends on two parameters: 1) the variance γ_j^2 of the added noise, and 2) the number J of noise realizations. We consider the clustering method and the algorithm used to estimate the correct number of clusters as hyper-parameters. In this section, we will use GMM for the first, and GAP for the latter. Results for K-means do not significantly differ from the ones shown here, and can be consulted as supplementary material in the public repository associated with the code used in this paper [1].

To find out what would be a good choice for the remaining parameters, we investigate their respective impact on the classification results. Therefore, we compute the accuracy of the classification, i.e. the rate of correctly classified SZs, for different values of γ_j and J in order to select a combination of parameters that yields optimal results. To do so, we first define γ_j as:

$$\gamma_j = \beta \, \gamma_0, \tag{16}$$

where γ_0 is known for a synthetic time series, and $\beta \in \{0.70, 0.85, 1.00, 1.15, 1.30\}$. Then, we compute the accuracy of the classification of SZs for the signal defined in Eq. (6) with a length of $N = 512$, for the just defined range of values for β, and when J also varies. A *ground truth* set of correctly classified SZs according to the types described in Sec. 4 is manually generated, making use of the known frequencies of the three tones of the signal. Then, optimal values of β and J are empirically found by solving

$$\arg\max_{\beta, J} \mathrm{Acc}(\beta, J), \tag{17}$$

where $\mathrm{Acc}(\beta, J)$ is the computed accuracy. Fig. 8 shows the accuracy of the classification of SZs for different values of β, J and SNRs. It can be seen that the proposed approach is robust to the choice of β and J, the best results

---

[1]https://github.com/jmiramont/spectrogram-zeros-classification

obtained for $\beta = 1.00$ and $\beta = 1.15$, and $J = N = 512$, though with a smaller J the results are almost the same, which confirms that meaningful features are obtained with a lot fewer noise realizations than the signal length.
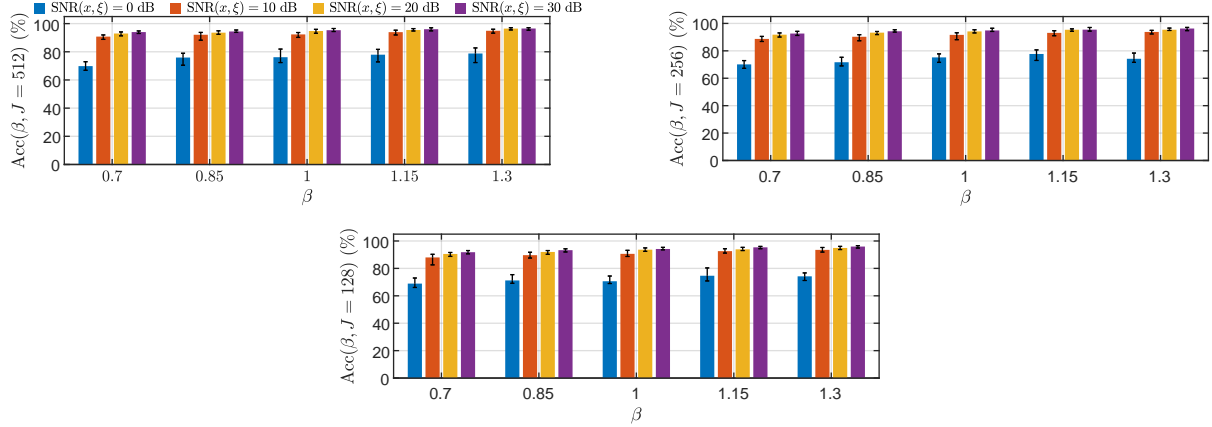

Figure 8: Accuracy of the classification of SZs (in %) of the signal shown in Fig. 2, and described by Eq. (6), with a length of $N = 512$, for several SNRs and values of β and J. The bars indicate the median over 50 realizations. The errorbars indicate the 25th and 75th percentiles.

In practice, γ_0 can be estimated by means of a robust median absolute deviation estimator [16, 30, 35] given by

$$\hat{\gamma}_0 = \frac{\sqrt{2}}{0.6745} \text{ median} \left(\left| \Re \left\{ V_y^g(t,f) \right\} \right| \right).$$ (18)

Then, γ_j is fixed to $\gamma_j = \beta \, \hat{\gamma}_0$, with $\beta = 1.00$ based on the results of Fig. 8. Note that this value for β can be viewed as the default parameter value, and can be adapted depending on the application and to compensate for the estimation of $\hat{\gamma}_0$, if needed. The steps to automatically group and label the SZs in the proposed kinds are summarized in Algorithm 1.

Algorithm 1 Classification of SZs.

Input: A noisy signal y, J, β.
1: Compute the short-time Fourier transform $V_y^g(t,f)$.
2: Estimate the noise variance $\hat{\gamma}_0$ as in Eq. (18) using $V_y^g(t,f)$.
3: From $V_y^g(t,f)$, compute spectrogram of y and the set $\mathcal{Z}_0$ of original SZs.
4: Compute the Voronoï tessellation of $\mathcal{Z}_0$.
5: Compute $G[n,m]$ using J independent WGN realizations with $\gamma_j = \beta \, \hat{\gamma}_0$.
6: **for** each $\mathbf{z} \in \mathcal{Z}_0$ **do**
7:     Compute $\text{AR}(\mathbf{z})$ (Eq. (13)) and $\text{Max}(\mathbf{z})$ (Eq. (14)).
8: **end for**
9: Determine the best GMM in the feature space spanned by $\text{AR}(\mathbf{z})$ and $\text{Max}(\mathbf{z})$, for $K = 2$ and $K = 3$.
10: Decide whether $K = 1, 2$ or 3 (for example, using the GAP criterion [38]) .
11: Compute centroids C_i, $i = 1, \ldots, K$ of detected clusters.
12: Sort clusters in ascending order according to $\text{AR}(\mathbf{z})$, to obtain $C_{(i)}$, $i = 1, \ldots, K$.
13: **if** K=1 **then**
14:     Label the only cluster as: $C_{(1)} \to \text{NN}$.
15: **end if**
16: **if** K=2 **then**
17:     Label clusters as: $(C_{(1)}, C_{(2)}) \to (\text{SN}, \text{NN})$.
18: **end if**
19: **if** K=3 **then**
20:     Label clusters as: $(C_{(1)}, C_{(2)}, C_{(3)}) \to (\text{SS}, \text{SN}, \text{NN})$.
21: **end if**
22: **return** K and the labels of each SZ in $\mathcal{Z}_0$.

Let us now give some simple illustrations of the behavior of Algorithm 1 for various signals. Fig. 9 (a) displays the classification results for a monocomponent signal, the spectrogram of which has only SZs of the SN and NN

kinds. In such case, as expected, only two clusters are well identified by Algorithm 1. Figs. 9 (b) and 9 (c) show classification results for more complicated signals. In Fig. 9 (b), the algorithm is able to identify the zeros of the SS kind created when a component interferes with *itself* due to rapid changes in its instantaneous frequency, as it happens with the sinusoidal chirp in the signal. Fig. 9 (c) shows the results for a signal made of three components with frequency and amplitude modulations. In all these examples, the SZs of the border of the signal domain are well classified as zeros of the SN kind. In the following section, we describe some potential applications of the proposed classification of SZs.
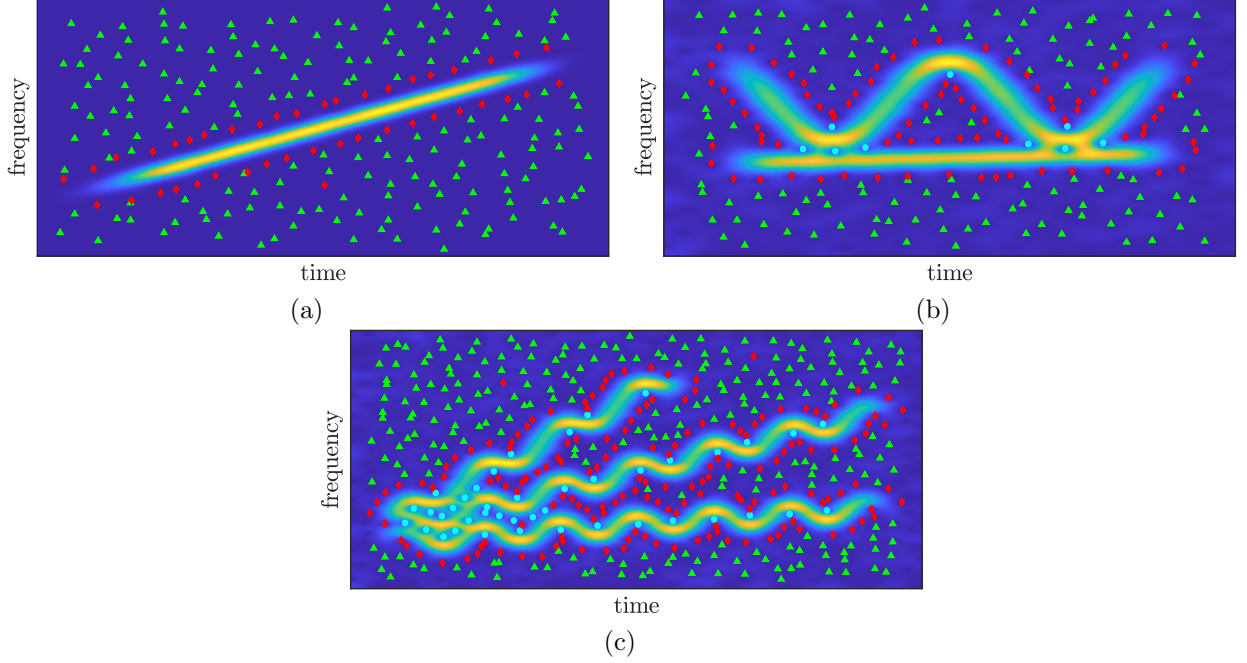

(a)

(b)

(c)

Figure 9: Panel (a) classification results displayed on the spectrogram of a monocomponent signal for which only two types of zeros are present. Panels (b) and (c) same as (a) but for two different multicomponent signals with length $N = 1024$. All signals have $SNR(x, \xi) = 20$ dB. $J = 256$ was used for SZs classification with Algorithm 1. Cyan dots correspond SZs of the SS kind, green triangles to those of the NN kind, while red diamonds represent SZs of the SN kind.

7   Signal Detection and Denoising Based on Classified SZs

The classification of SZs in three types can be useful to estimate $\mathcal{D}_s$ by combining it with the Delaunay triangulation, which naturally leads to a novel signal denoising technique. Additionally, by using the GAP criterion [38] to decide whether the number of clusters in the feature space is greater than one, a signal detection procedure can be obtained.

7.1   Signal Detection

Signal detection can be stated as a binary hypothesis test [4, 18]:

$$\begin{cases} H_0 : y(t) = \xi(t) \\ H_1 : y(t) = x(t) + \xi(t) \end{cases}, \tag{19}$$

where H_0 simply states that a given signal is only noise, whereas the alternative hypothesis states that the data is actually a mixture of signal and noise. A test aiming to detect a signal can then be characterized by:

1. Specificity: the probability of correctly keeping H_0 or $1 - \alpha$, where α is the so-called *significance* of the test, i.e. the probability of wrongfully rejecting H_0.

2. Sensitivity: the probability of correctly rejecting H_0, also known as the *statistical power of the test*.

Considering the classification of SZs described in the previous section, a detection procedure can be set as follows. Assuming some noise is always present in the signal, one can apply Algorithm 1 to a signal $y(t)$ in order to

obtain K, i.e. the number of clusters formed by the zeros of its spectrogram. If there is only one cluster, i.e. $K = 1$ in Eq. (15), then $y(t) = \xi(t)$. Correspondingly, all the SZs are of the NN kind. Otherwise, if $K > 1$ and there are zeros of the SN (and potentially SS) kind, then $y(t) = x(t) + \xi(t)$, and H_0 is rejected.

The results of this simple procedure are provided in Sec. 8.1, where we estimate the specificity and the sensitivity of the detection algorithm, and compare them with the tests described in [4].

7.2   Signal Denoising

We now describe a method that combines the information given by the classification of the SZs with the Delaunay triangulation used in [19] in order to estimate the signal domain $\mathcal{D}_s$ and disentangle signal and noise.

After having performed the SZs classification using Algorithm 1, one estimates $\mathcal{D}_s$ by selecting the Delaunay triangles that satisfy any of the following two criteria:

1. At least one of their vertices is a SZ of SS kind.

2. All their vertices are SZs of SN kind.

The first criterion aims to always select the Delaunay triangles associated with SZs of the SS kind. This will prevent the algorithm from rejecting triangles that correspond to the signal domain but would otherwise be neglected by the DT method [19] for not being triangles with a large enough edge, as illustrated previously in Sec. 3. Meanwhile, the second criterion aims to keep the triangles associated with an isolated component in the signal domain.

Once the estimation of $\mathcal{D}_s$ is found, the denoised signal is obtained by inverting the STFT as in the DT method, using Eq. (5). Given that these criteria are not based on the length of the edges of the triangles, they should allow to more efficiently identify the triangles corresponding to $\mathcal{D}_s$ when there exists a strong interference between the components of the signal.

Denoising methods can be applied after the detection of a signal, therefore the approach just described will consider that a signal is actually present in the data (as it is customary when studying the performance of denoising/reconstruction approaches [19, 32]). We implement this assumption by making a small modification in Algorithm 1. In the context of denoising, we will not consider the case $K = 1$ in step 10 of Algorithm 1, since we would expect $K = 2$ or $K = 3$ if a signal is present and contaminated with noise. Hence, the algorithm will only consider $K \in \{2, 3\}$ in such case. This, in turn, allows us to use the CaHa criterion instead of GAP to select the optimal number of clusters, reducing the execution time of the denoising procedure described above. Results supporting this approach for signal denoising are reported in the following section.

8   Numerical Experiments

In this section, we study the performance of the signal detection method proposed in the previous section. Then, we report the results for the signal denoising approach based on SZs classification, using both synthetic and real-world signals.

8.1   Signal Detection

We estimate the specificity of the proposed signal detection approach described in Sec. 7.1 using 200 realizations of WGN, and computing the proportion of times $K = 1$ is found when applying Algorithm 1. Similarly, the sensitivity is estimated as the proportion of times Algorithm 1 outputs $K > 1$ for 200 signals made of the sum of a WGN realization and a linear chirp (similar to the one shown in Fig. 9 (a), but using $N = 256$). To complete the performance evaluation, we assess the sensitivity for a range of SNR (between -5 and 10 dB). These results are then compared with the Monte Carlo test proposed by Bardenet et al. [2, 4]. Such a test, and the proposed approach for detection, are based on the principle of finding significant differences between the data and the noise-only case in order to infer the presence of signal. Hence, these detection methods do not make any assumption on the signal, but instead consider a particular type of noise (for example, WGN). In this sense, they are different from other types of detection tests described in the literature, where a signal model is often used [26, 40].

A specificity of 0.88 was estimated for the proposed approach. In order to make a fair comparison, we fixed the significance of the detection test proposed in [4] to 0.12 so that both methods have a similar specificity, and proceeded to estimate the sensitivity. Fig. 10 shows that the proposed approach has a higher sensitivity for SNR ≤ 0 dB (e.g. 0.81 vs. 0.61 for SNR $= 0$ dB) and similar results for SNR > 0 dB. These findings indicate that

the proposed procedure is more effective than the competing method in the more difficult case of low SNR, while keeping a very high specificity.
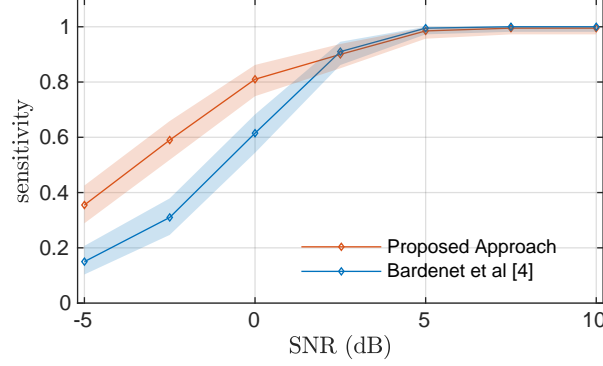

Figure 10: Comparison of the detection performance of Algorithm 1 (detecting if $K = 1$ or if $K > 1$, with $\beta = 1.0$ and $J = 256$) and the detection tests described in [2, 4] Sensitivity was estimated using 200 mixtures of WGN and a linear chirp of $N = 256$ samples, for several SNRs. The shaded regions around the curves show the Clopper-Pearson 95% confidence intervals.

8.2   Signal Denoising

We report here a characterization of the performance of the proposed method for signal denoising. We first describe results using synthetic signals. Then we discuss the computational cost of the method and its variations. We end the section by showing examples of applications on real-world signals.

8.2.1   Examples with Synthetic Signals

We now go back to the example shown in Fig. 2, in order to illustrate how the denoising procedure introduced in Sec. 7.2 leads to a better selection of the Delaunay triangles that form $\mathcal{D}_s$ and thus better denoising performance than the DT method. Fig. 11 shows the triangles selected by considering the approach detailed in Sec. 7.2, which are much more relevant than those obtained with the DT method displayed in Figs. 2 (a) and 2 (b). In such a simple case, the method we propose enables the full disentanglement between signal and noise.
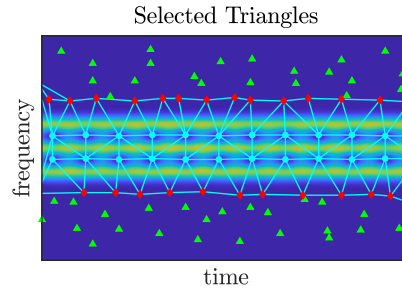

Figure 11: Delaunay triangles selected by applying the technique described in Sec. 7.2, based on the classification of spectrogram zeros. Blue dots denote SZs of SS kind, green triangles those of NN kind, and red diamonds those of SN kind.

We now describe the results for a more challenging signal. We compared the performance of the denoising strategy proposed in Sec. 7.2 with the DT method [19], an approach [7, 37] based on synchrosqueezing and ridge detection (SST+RD), a simple hard thresholding (HT) [30] of the STFT based on the spectrogram, and a contours-based reconstruction [29, 34] (denoted "Contours").

As performance metrics, we used the so-called Quality Reconstruction Factor (QRF)

$$\text{QRF} := 10 \log_{10} \left(\frac{\|x\|_2^2}{\|x - \tilde{x}\|_2^2} \right) \ (\text{dB}), \tag{20}$$

and a correlation coefficient (CC):

$$\text{CC} := \frac{\langle x, \tilde{x} \rangle}{\|x\|_2 \|\tilde{x}\|_2}, \tag{21}$$

13

where x is the noiseless signal, $\tilde{x}$ is the signal obtained after applying a denoising method, and $\langle x, \tilde{x} \rangle$ is the inner product between x and $\tilde{x}$.

The QRF can be interpreted as the SNR of $\tilde{x}$, whereas the value of CC reflects the similarity of the wave-shape between x and $\tilde{x}$. Note that we are not interested here in discriminating the individual components, but in correctly determining $\mathcal{D}_s$ to effectively separate the whole signal from noise.

First, we study the performance of the proposed approach, denoted by "SZC" (SZs Classification), for different combinations of the hyper-parameters (GMM or KMEANS for clustering, GAP or CaHa criterion for determining the correct number of clusters). The results are shown in Fig. 12 (a), where the signal from Fig. 9 (c) is used in the simulations. In particular, the combinations using GMM resulted in a higher QRF, the difference between both approaches being more important for lower SNRs. The combination SZC-KMEANS-CaHa is the one with the worst performance for low SNR. Combining CaHa with GMM, instead of GAP, does not lead to a significant difference in terms of QRF (see SZC-GMM-GAP and SZC-GMM-CaHa in Fig. 12 (a)) although, as we will show in the next section, it can reduce the computational burden of the technique. For the sake of clarity, we shall use only SZC-GMM-GAP as a representative of our method when comparing with other approaches, keeping in mind that, excepting for SZC-KMEANS-CaHa, the performance in terms of QRF is equivalent to other choices of hyperparameters.

The results of the comparison with other denoising strategies can be seen in Figs. 12 (b) and (c), where QRF and CC are respectively considered. For high SNRs (SNR > 0 dB), our approach yields higher QRF values than the other methods considered. The performance in terms of CC is mostly similar for all methods. For the more challenging case of low SNR, the thresholding method HT provides with the best results in terms of the QRF. However this is mainly because, for such low SNRs, most of the signal is also filtered out. This is reflected in the poor performance of the same method when CC is considered instead. For that performance metric, DT-$\ell_{max} = 1.2$ is much better (see "DT-1.2" in Fig. 12 (b)), although this method fails from the point of view of the QRF. In contrast, still in the low SNR case, our method provides with a better signal reconstruction than most of the other studied approaches, with similar results to Contours and losing only to HT when QRF is considered. From the point of view of CC, however, the results are comparable to DT-$\ell_{max} = 1.2$. All in all, the proposed approach have a more balanced performance than other methods when both metrics, i.e. QRF and CC, are taken into consideration.
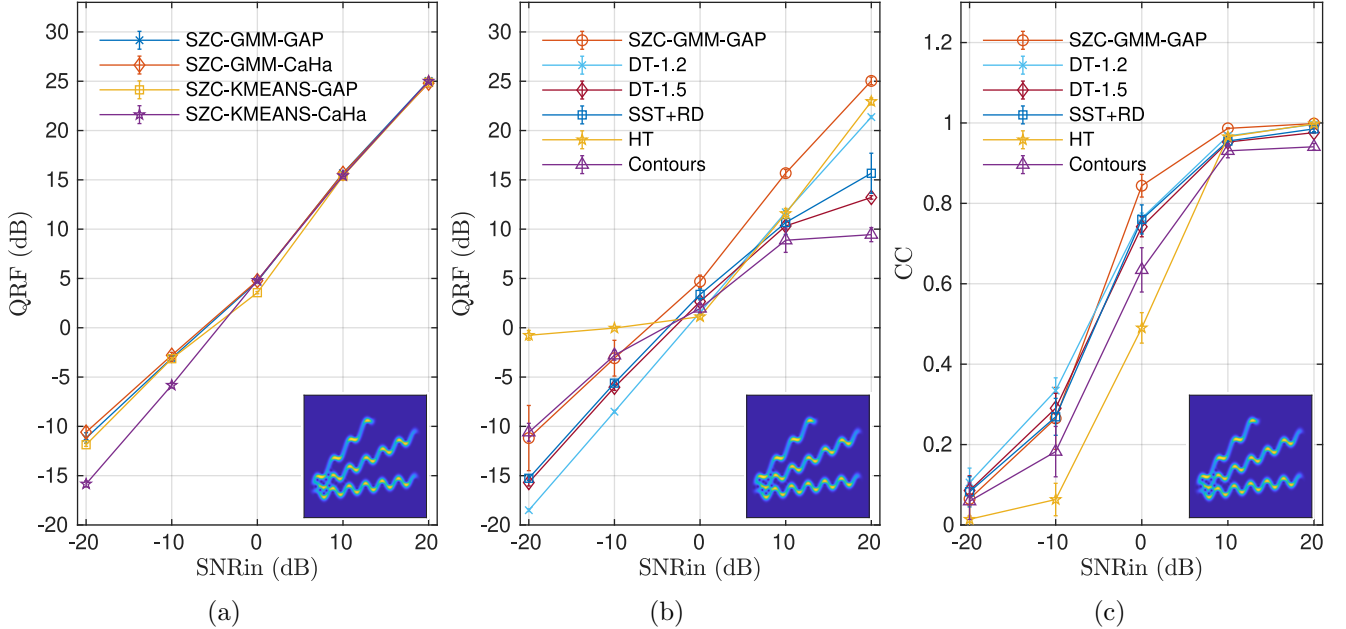

Figure 12: Signal reconstruction performance of different denoising approaches for the three-component signal corresponding to the spectrogram of Fig. 9 (c) with length $N = 1024$. The markers and error bars indicate, respectively, the mean and the standard deviation over 150 repetitions of the experiment. (a) Quality Reconstruction Factor (QRF) of the proposed approach for four combinations of hyper-parameters. "SZC" (SZs classification) is the proposed strategy based on Algorithm 1 with $\beta = 1.0$ and $J = 256$, and using either GMM or KMEANS for clustering, and either GAP or CaHa criterion for the selection of the optimal number of clusters. (b) Comparison of the QRF obtained for different denoising techniques and the proposed approach using GMM and GAP (SZS-GMM-GAP). (c) Same as (b) but considering the correlation coefficient (CC) as performance metric.

Figs. 13 (a) to (d) give us a hint into why the proposed approach produces better results. Fig. 13 (a) shows

the spectrogram of the recovered signal $\tilde{x}$ obtained after applying the denoising strategy based on SZs classification (SZC-GMM-GAP). Figs. 13 (b) to (d) show, respectively, the spectrogram of $\tilde{x}$ obtained with SST+RD, DT method with $\ell_{max} = 1.2$ and with $\ell_{max} = 1.5$, for SNR = 15 dB. Indeed, by comparing these figures, one can see that the main source of errors of other approaches, for example SST+RD and DT with $\ell_{max} = 1.5$, is the strong interference between the components in the leftmost part of the signal, where the modes are more closely packed (indicated by a red rectangle in Fig. 13 (d)). Fig. 13 (e) and (f) display the signals in the time domain, showing the original signal x, the recovered version $\tilde{x}$ using SZC-GMM-GAP or DT-$\ell_{max} = 1.5$ respectively, and the error signal $x - \tilde{x}$. The red rectangle in Fig. 13 (f) corresponds to the one shown in Fig. 13 (d), highlighting the increased error committed by using DT-$\ell_{max} = 1.5$. As previously discussed in Sec. 3, reducing the value of ℓ_{max} to 1.2 improves the estimation of $\mathcal{D}_s$, but at the cost of introducing more noise in the recovered signal, which is why the proposed approach is more efficient than the DT method even in this case.
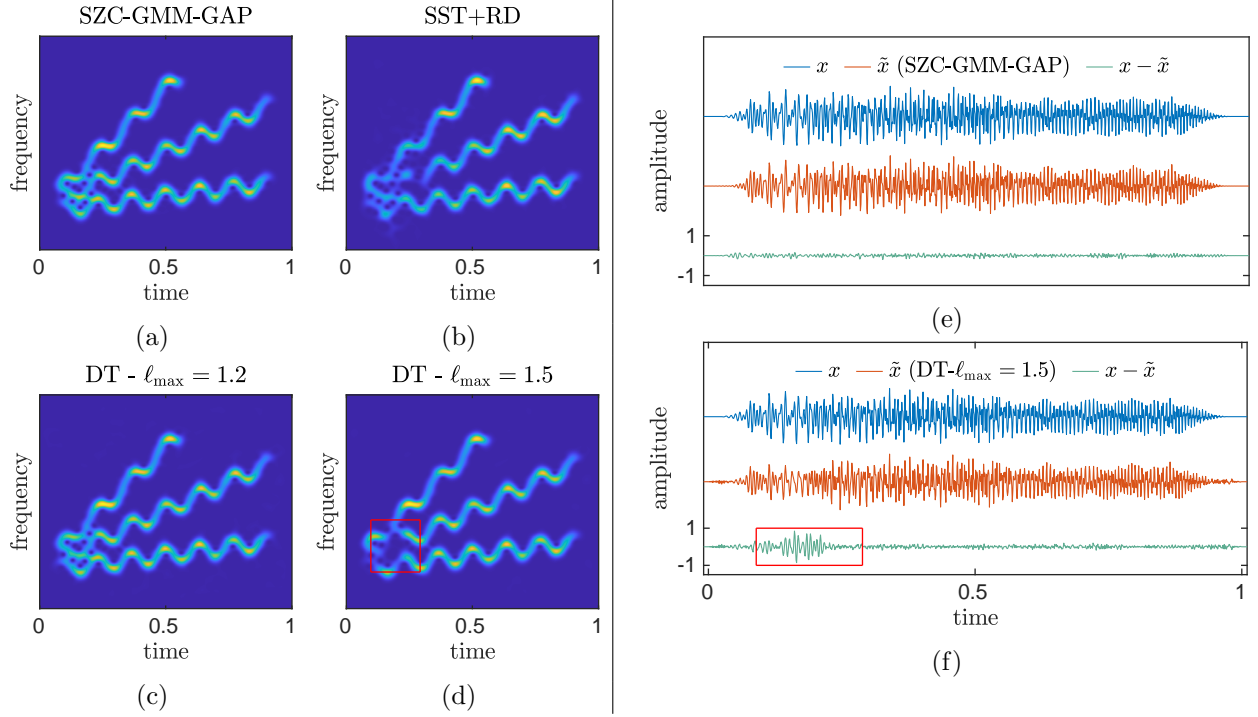

Figure 13: (a) - (d): Spectrograms of the recovered signal obtained with different methods, with SNR = 15 dB. DT method [19] was applied using $\ell_{max} = 1.2$ and $\ell_{max} = 1.5$, as indicated in the title of the sub-figures. (e) Time domain representation of the original signal x, the recovered signal $\tilde{x}$, and the error signal $x - \tilde{x}$, after applying the proposed approach (SZC-GMM-GAP). (f) is the similar to (e) but for the DT method with $\ell_{max} = 1.5$. The red rectangle corresponds to the time interval of the red rectangle shown in (d). Amplitude is given in arbitrary units.

8.2.2 Computational Cost

We now study the computational cost of the proposed algorithm by contrasting its execution time. Because it is a noise-assisted approach needing the computation of the STFT for several noise realizations, one can expect a larger execution time than with other methods, like the closely related DT approach. Table 1 compares the average execution time, in seconds, of the proposed method using GAP and CaHa for the determination of the number of clusters. All the simulations were done in MATLAB 2022b running on a notebook PC, Intel i7 12th Gen., 16 GB RAM, using MATLAB's parallel toolbox to accelerate the computation of the 2D histograms of SZs. In Table 1, one can see that using CaHa reduces the computational burden when compared with the GAP criterion. This is mainly because the latter needs more costly Monte Carlo simulations, the combination of which with GMM greatly increases the execution time. We stress that we are able to select either GAP or CaHa here because, for the denoising application, a signal is supposed to be present in the data and, therefore, there is no need to consider the case where $K = 1$, i.e. when the data is noise only.

When compared with the DT method (last row in Table 1), the execution time of the proposed approach is at least one order of magnitude greater. This illustrates a limitation of the introduced denoising method, although it is

15

out of the scope of the technique to be used in *online* processing, for example, where fast algorithms are necessary. Nevertheless, this is an aspect of the proposed method that calls for further improvement, and could benefit from the study of the underlying distribution of the 2D histograms in order to reduce the number of noise realizations used to compute them.

Table 1: Average execution time, in seconds, over 50 simulations of the denoising strategy described in Sec. 7.2 and DT method [19]. The signal corresponding to Fig. 9(c) was used, for $N \in \{512, 1024\}$. SZC stands for the denoising approach using classified SZs, using either GMM or KMEANS for clustering, and either GAP criterion or Calinski-Harabasz (CaHa) for determining the number of clusters. DT stands for the Delaunay Triangulation approach [19].

Method	Execution Time	
	$N = 512$	$N = 1024$
SZC-GMM-GAP-$J = 256$	6.69	18.12
SZC-GMM-CaHa-$J = 256$	1.94	9.87
SZC-KMEANS-GAP-$J = 256$	2.54	10.76
SZC-KMEANS-CaHa-$J = 256$	1.92	9.82
DT - $\ell_{\max} = 1.2$	0.16	1.13

8.2.3   Examples with Real-World Signals

To end this section, we exemplify the potential of the denoising approach introduced in Sec. 7.2 on audio signals. Since we cannot compute the QRF for real-world signals, we limit ourselves to a more qualitative assessment of the extraction masks obtained with the proposed approach and DT method. Fig. 14 (a) shows the spectrogram of a cello recording, while Fig. 14 (b) displays the extraction mask obtained after SZs classification, where GMM was used for clustering and GAP criterion for determining the optimal number of clusters (SZC-GMM-GAP). Two extraction masks for DT method using $\ell_{\max} = 1.75$ and $\ell_{\max} = 1.85$ are also shown for comparison in Figs. 14 (c) and 14 (d), respectively. Both approaches seem comparable, although the mask corresponding to the proposed strategy based on the classification of SZs has less noise-related components in the uppermost and lowermost parts.

Similarly, Fig. 15 (a) shows the spectrogram of an excerpt of the aria "Alto Giove" by N. A. Porpora. Comparing Fig. 15 (b) with Figs. 15 (c) and Fig. 15 (d), it is clear that the proposed approach produces a mask where the components are complete and which contains fewer spurious isolated domains related to noise.
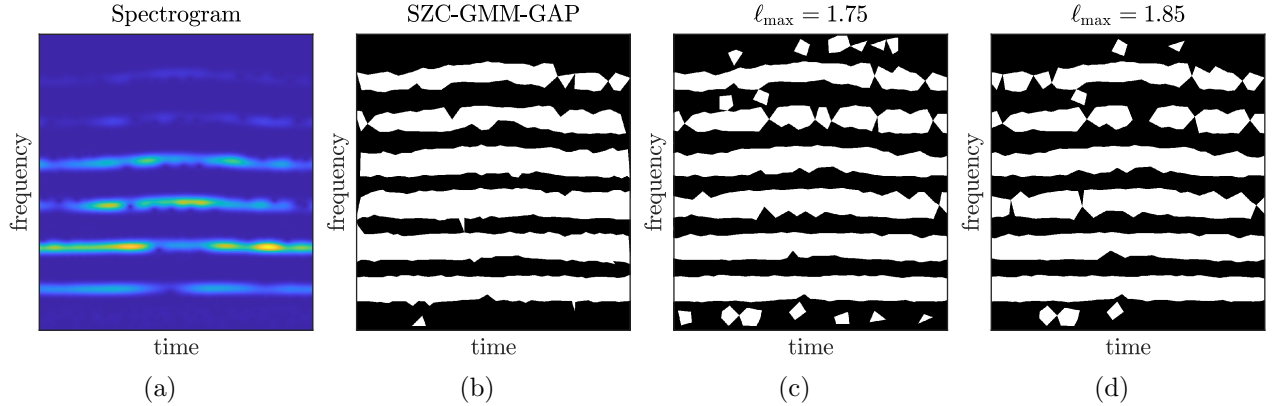

Figure 14: (a) Spectrogram of a cello recording, with length $N = 2048$. (b) Extraction mask for the proposed approach based on the classification of SZs (SZC-GMM-GAP), using $\beta = 1.5$ and $J = 512$. (c) Extraction mask for the DT method with $\ell_{\max} = 1.75$. (d) Idem with $\ell_{\max} = 1.85$.
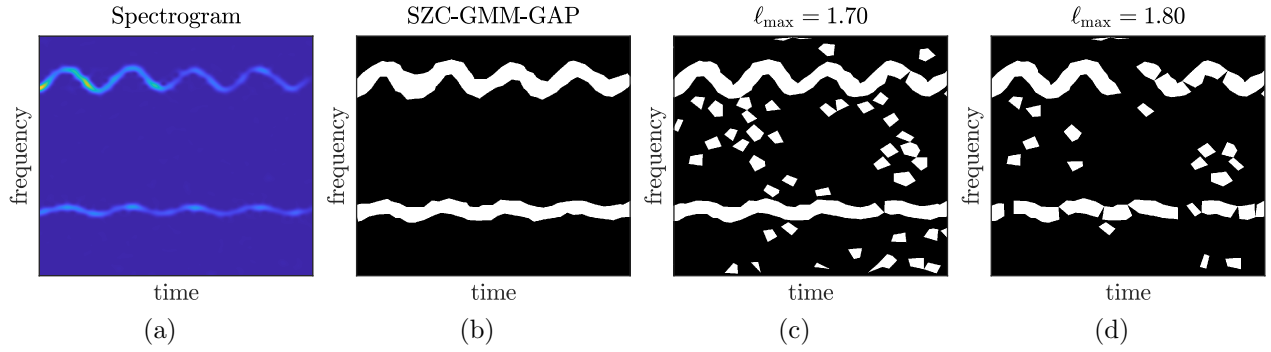
Figure 15: (a) Spectrogram of an excerpt of the aria "Alto Giove", with length $N = 2048$. (b) Extraction mask for the proposed approach based on the classification of SZs (SZC-GMM-GAP), using $\beta = 1.9$ and $J = 512$. (c) Extraction mask for the DT method with $\ell_{\max} = 1.70$. (d) Idem with $\ell_{\max} = 1.8$.

9 Discussion

We introduced an unsupervised algorithm to classify the SZs in three classes by labeling clusters of SZs in the two-dimensional space spanned by $\mathrm{AR}(\mathbf{z})$ and $\mathrm{Max}(\mathbf{z})$. On top of this algorithm, we proposed novel approaches for 1) signal detection and 2) signal denoising, that outperform previous approaches in the more challenging scenarios of low SNR and the presence of strong interfering signal components.

As mentioned before, to decide whether the number of clusters is one or more is akin to a signal detection test. On the other hand, to detect whether the number of clusters is two or three amounts to identify the presence of strong interference between the components of the signal for a given analysis window. When two components are very close to each other, interference can produce the so-called *time-frequency bubbles* [14, 15], which have been shown to be a hindrance in ridge detection and instantaneous frequency estimation. Future work will continue to explore other uses of SZs classification, like the detection of time-frequency bubbles and the regularization of spectrogram concentration measures, such as the well-known Rényi entropy [3], for proper selection of the analysis window. Additionally, further work will also explore the classification of zeros corresponding to other signal representations, such as the continuous wavelet transform, aiming to generalize the ideas introduced in this article.

10 Conclusion

Focusing on the destructive interference between components that produces spectrogram zeros, a classification scheme of these points in three categories was introduced, along with a noise-assisted algorithm to automatically classify them in the proposed kinds. Applications to signal denoising and detection based on the proposed classification were then introduced. Numerical experiments showed that these approaches produce better results than state-of-the-art techniques using SZs, especially in, but not limited to, the cases when the components of the signal are very close to each other and for low SNR.

Conflicts of Interest

The authors have no conflicts of interest to declare.

References

[1] François Auger and Patrick Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Transactions on Signal Processing*, 43(5):1068–1089, 1995.

[2] Adrian Baddeley, Peter J Diggle, Andrew Hardegen, Thomas Lawrence, Robin K Milne, and Gopalan Nair. On tests of spatial pattern based on simulation envelopes. *Ecological Monographs*, 84(3):477–489, 2014.

[3] Richard G Baraniuk, Patrick Flandrin, Augustus JEM Janssen, and Olivier JJ Michel. Measuring time-frequency information content using the Rényi entropies. *IEEE Transactions on Information Theory*, 47(4):1391–1409, 2001.

[4] Rémi Bardenet, Julien Flamant, and Pierre Chainais. On the zeros of the spectrogram of white noise. *Applied and Computational Harmonic Analysis*, 48(2):682–705, 2020.

[5] Rémi Bardenet and Adrien Hardy. Time-frequency transforms of white noises and Gaussian analytic functions. *Applied and computational harmonic analysis*, 50:73–104, 2021.

[6] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[7] Eugene Brevdo, Neven S Fuckar, Gaurav Thakur, and Hau-Tieng Wu. The synchrosqueezing algorithm: a robust analysis tool for signals with time-varying spectrum. *arXiv preprint arXiv:1105.0010*, 2011.

[8] Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

[9] E Chassande-Mottin, Ingrid Daubechies, Francois Auger, and Patrick Flandrin. Differential reassignment. *IEEE Signal Processing Letters*, 4(10):293–294, 1997.

[10] Leon Cohen. Generalized phase-space distribution functions. *Journal of Mathematical Physics*, 7(5):781–786, 1966.

[11] Marcelo A Colominas, Gastón Schlotthauer, and María E Torres. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomedical Signal Processing and Control*, 14:19–29, 2014.

[12] Marcelo A Colominas, Gastón Schlotthauer, María E Torres, and Patrick Flandrin. Noise-assisted EMD methods in action. *Advances in Adaptive Data Analysis*, 4(04):1250025, 2012.

[13] David L. Davies and Donald W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227, 1979.

[14] Nathalie Delprat. Global frequency modulation laws extraction from the gabor transform of a signal: A first study of the interacting components case. *IEEE Transactions on Speech and Audio Processing*, 5(1):64–71, 1997.

[15] Nathalie Delprat, Bernard Escudié, Philippe Guillemain, Richard Kronland-Martinet, Philippe Tchamitchian, and Bruno Torresani. Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies. *IEEE Transactions on Information Theory*, 38(2):644–664, 1992.

[16] David L Donoho and Jain M Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[17] Luis Alberto Escudero, Naomi Feldheim, Günther Koliander, and José Luis Romero. Efficient computation of the zeros of the Bargmann transform under additive white noise. *arXiv preprint arXiv:2108.12921*, 2021.

[18] Patrick Flandrin. *Time-frequency/time-scale analysis*. Academic press, 1998.

[19] Patrick Flandrin. Time–frequency filtering based on spectrogram zeros. *IEEE Signal Processing Letters*, 22(11):2137–2141, 2015.

[20] Patrick Flandrin. The sound of silence: Recovering signals from time-frequency zeros. In *50th Asilomar Conference on Signals, Systems and Computers*, pages 544–548. IEEE, 2016.

[21] Patrick Flandrin. *Explorations in time-frequency analysis*. Cambridge University Press, 2018.

[22] Subhroshekhar Ghosh, Meixia Lin, and Dongfang Sun. Signal analysis via the stochastic geometry of spectrogram level sets. *IEEE Transactions on Signal Processing*, 70:1104–1117, 2022.

[23] Philippe Guillemain and Richard Kronland-Martinet. Characterization of acoustic signals through continuous linear time-frequency representations. *Proceedings of the IEEE*, 84(4):561–585, 1996.

[24] Jechang Jeong and William J Williams. Mechanism of the cross-terms in spectrograms. *IEEE Transactions on Signal Processing*, 40(10):2608–2613, 1992.

[25] Shubha Kadambe and G Faye Boudreaux-Bartels. A comparison of the existence of 'cross terms' in the Wigner distribution and the squared magnitude of the wavelet transform and the short-time Fourier transform. *IEEE Transactions on Signal Processing*, 40(10):2498–2517, 1992.

[26] SM Kay. Fundamentals of statistical signal processing volume i, and ii: Estimation theory and detection theory. *Beijin: Publishing House of Electronics Industry*, 41, 2006.

[27] Günther Koliander, Luis Daniel Abreu, Antti Haimi, and José Luis Romero. Filtering the continuous wavelet transform using hyperbolic triangulations. In *2019 13th International conference on Sampling Theory and Applications (SampTA)*, pages 1–4. IEEE, 2019.

[28] N Laurent, S Meignen, MA Colominas, JM Miramont, and F Auger. A novel approach based on voronoï cells to classify spectrogram zeros of multicomponent signals. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[29] Yoonseob Lim, Barbara Shinn-Cunningham, and Timothy J Gardner. Sparse contour representations of sound. *IEEE Signal Processing Letters*, 19(10):684–687, 2012.

[30] Stephane Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 2008.

[31] Sylvain Meignen and Thomas Oberlin. On the study of interference and their localization in the time-frequency plane. *HAL (preprint), https://hal.science/hal-03327526*, 2021.

[32] Sylvain Meignen, Thomas Oberlin, Philippe Depalle, Patrick Flandrin, and Stephen McLaughlin. Adaptive multimode signal reconstruction from time–frequency representations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150205, 2016.

[33] Barbara Pascal and Rémi Bardenet. A covariant, discrete time-frequency representation tailored for zero-based signal detection. *IEEE Transactions on Signal Processing*, 70:2950–2961, 2022.

[34] Duong-Hung Pham and Sylvain Meignen. An adaptive computation of contour representations for mode decomposition. *IEEE Signal Processing Letters*, 24(11):1596–1600, 2017.

[35] Duong-Hung Pham and Sylvain Meignen. A novel thresholding technique for the denoising of multicomponent signals. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4004–4008. IEEE, 2018.

[36] Ervin Sejdić, Igor Djurović, and Jin Jiang. Time–frequency feature representation using energy concentration: An overview of recent advances. *Digital Signal Processing*, 19(1):153–183, 2009.

[37] Gaurav Thakur, Eugene Brevdo, Neven S Fučkar, and Hau-Tieng Wu. The synchrosqueezing algorithm for time-varying spectral analysis: Robustness properties and new paleoclimate applications. *Signal Processing*, 93(5):1079–1094, 2013.

[38] Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001.

[39] María E Torres, Marcelo A Colominas, Gaston Schlotthauer, and Patrick Flandrin. A complete ensemble empirical mode decomposition with adaptive noise. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4144–4147. IEEE, 2011.

[40] Harry L Van Trees. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.

[41] Zhaohua Wu and Norden E Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(01):1–41, 2009.