Improving methods based on spectrogram zeros using unsupervised classification

Juan M. Miramont
jmiramontt@univ-lille.fr

November 9, 2023

Agenda:

Agenda:

Introduction

Short-time Fourier transform (STFT)

Let $x(t)$ be a real -or complex- signal and its STFT be defined as:

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

with $g(t) = 2^{1/4} e^{-\pi t^2}$. Then, its **spectrogram** is defined as:

$$S_x^g(t,f) := |V_x^g(t,f)|^2.$$

Introduction

Short-time Fourier transform (STFT)

Let $x(t)$ be a real -or complex- signal and its STFT be defined as:

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

with $g(t) = 2^{1/4} e^{-\pi t^2}$. Then, its **spectrogram** is defined as:

$$S_x^g(t,f) := |V_x^g(t,f)|^2.$$

- $S_x^g(t,f)$ can be *interpreted* as a time-frequency distribution:

$$\iint_{\mathbb{R}^2} S_x^g(t,f)dt\, df = E_x$$

- *Largest values* of $S_x^g(t,f)$ contain more information about $x(t)$.

Short-time Fourier transform (STFT)

Let $x(t)$ be a real -or complex- signal and its STFT be defined as:

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

with $g(t) = 2^{1/4}e^{-\pi t^2}$. Then, its **spectrogram** is defined as:

$$S_x^g(t,f) := |V_x^g(t,f)|^2.$$

- $S_x^g(t,f)$ can be *interpreted* as a time-frequency distribution:

$$\iint_{\mathbb{R}^2} S_x^g(t,f)dt\,df = E_x$$

- *Largest values* of $S_x^g(t,f)$ contain more information about $x(t)$.
  $\rightarrow$ Classical TF paradigm.

Introduction

In the following, we will consider signal and noise mixtures, in which:

- $\xi(t)$ is a zero-mean white Gaussian noise (WGN) satisfying:

$$\mathbb{E}\left\{\xi(t)\overline{\xi(t-\tau)}\right\} = \gamma_0^2 \delta(\tau),$$

  where γ_0^2 is the noise variance.

- The Signal-to-Noise Ratio (SNR) between a signal x and ξ as:

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

  where P_x is the power of the signal.

The zeros of the spectrogram

Considering $z = t + if$, then the STFT can be written as[1]

$$V_x^g(t,f) = \mathcal{F}_x(z)\exp\left(-|z|^2\right)\exp(-i\pi tf),$$

where $\mathcal{F}_x(z)$ is the *Bargmann transform*.

[1] Karlheinz Gröchenig. *Foundations of time-frequency analysis*. Springer Science & Business Media, 2001.

[2] Patrick Flandrin. "Time–frequency filtering based on spectrogram zeros". In: *IEEE Signal Processing Letters* 22.11 (2015), pp. 2137–2141.

The zeros of the spectrogram

Considering $z = t + if$, then the STFT can be written as[1]

$$V_x^g(t, f) = \mathcal{F}_x(z) \exp\left(-|z|^2\right) \exp(-i\pi t f),$$

where $\mathcal{F}_x(z)$ is the *Bargmann transform*.
$\mathcal{F}_x(z)$ admits a Haddamard-Weierstrass factorization[2]:

$$\mathcal{F}_x(z) = z^m e^{Q(z)} \prod_n \left(1 - \frac{z}{z_n}\right) \exp\left(\frac{z}{z_n} + \frac{z^2}{2z_n^2}\right),$$

where z_n are the zeros of $\mathcal{F}_x(z)$, m is the order of a (possible) zero at the origin, and $Q(z)$ is a quadratic polynomial.

[1] Gröchenig, *Foundations of time-frequency analysis*.
[2] Flandrin, "Time–frequency filtering based on spectrogram zeros".

The zeros of the spectrogram

Considering $z = t + if$, then the STFT can be written as[1]

$$V_x^g(t,f) = \mathcal{F}_x(z)\exp\left(-|z|^2\right)\exp(-i\pi tf),$$

where $\mathcal{F}_x(z)$ is the *Bargmann transform*.
$\mathcal{F}_x(z)$ admits a Haddamard-Weierstrass factorization[2]:

$$\mathcal{F}_x(z) = z^m e^{Q(z)} \prod_n \left(1 - \frac{z}{z_n}\right)\exp\left(\frac{z}{z_n} + \frac{z^2}{2z_n^2}\right),$$

where z_n are the zeros of $\mathcal{F}_x(z)$, m is the order of a (possible) zero at the origin, and $Q(z)$ is a quadratic polynomial.
$\rightarrow$ $S_x^g(t,f)$ is characterized by the distribution of its zeros.

[1] Gröchenig, *Foundations of time-frequency analysis*.
[2] Flandrin, "Time–frequency filtering based on spectrogram zeros".

The zeros of the spectrogram of white noise

- The zeros of the spectrogram of complex white Gaussian noise (CWGN) are homogeneously distributed.
- Its distribution corresponds to that of the roots of a *planar Gaussian Analytic Function* (planar GAF)[3].
- The expected number of zeros in the TF plane can be rigorously deduced from the properties of the planar GAF.
- For a discrete signal with N time samples, the expected number of zeros of the spectrogram is N.

[3] Rémi Bardenet, Julien Flamant, and Pierre Chainais. "On the zeros of the spectrogram of white noise". In: *Applied and Computational Harmonic Analysis* 48.2 (2020), pp. 682–705.

The zeros of the spectrogram of white noise

- The zeros of the spectrogram of complex white Gaussian noise (CWGN) are homogeneously distributed.
- Its distribution corresponds to that of the roots of a *planar Gaussian Analytic Function* (planar GAF)[3].
- The expected number of zeros in the TF plane can be rigorously deduced from the properties of the planar GAF.
- For a discrete signal with N time samples, the expected number of zeros of the spectrogram is N.
- When a signal is present, the zeros *surround* the signal domain.

[3] Bardenet, Flamant, and Chainais, "On the zeros of the spectrogram of white noise".

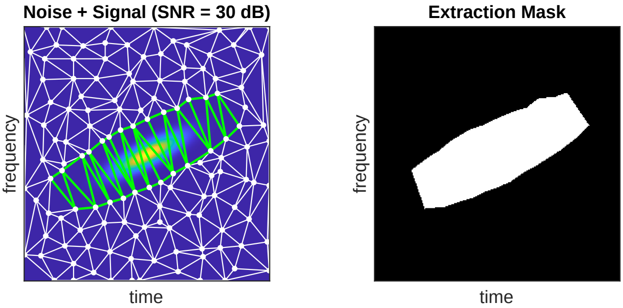
The zeros of the spectrogram of white noise

- The zeros of the spectrogram of complex white Gaussian noise (CWGN) are homogeneously distributed.
- Its distribution corresponds to that of the roots of a *planar Gaussian Analytic Function* (planar GAF)[3].
- The expected number of zeros in the TF plane can be rigorously deduced from the properties of the planar GAF.
- For a discrete signal with N time samples, the expected number of zeros of the spectrogram is N.
- When a signal is present, the zeros *surround* the signal domain.
  $\rightarrow$ Larger-than-expected regions without zeros are created.

[3] Bardenet, Flamant, and Chainais, "On the zeros of the spectrogram of white noise".

Signal estimation based on spectrogram zeros

Using Delaunay Triangulation (Flandrin, 2015)



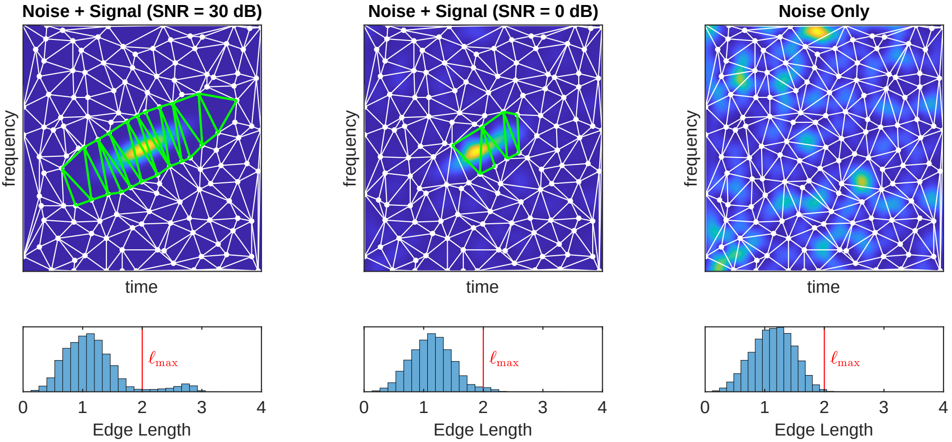
1. Compute the Delaunay triangulation on the SZ.
2. Find triangles with at least one edge length larger than ℓ_{max}.
3. Approximate the signal's TF domain $\mathcal{D}_x$.
4. Estimate $x(t)$ using:

$$\tilde{x}(t) = \frac{1}{g(0)} \int_{-\infty}^{+\infty} V_y^g(t,f) \mathbb{1}_{\mathcal{D}_x}(t,f) e^{i2\pi ft} df.$$

Signal estimation based on spectrogram zeros

Limitations

- Low SNR makes more difficult to identify the triangles.
- ℓ_{max} depends on the Signal-to-Noise Ratio (SNR).



Signal estimation based on spectrogram zeros

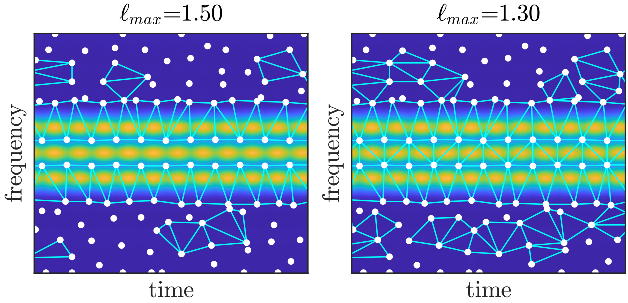
A pathological case

- Consider a signal given by:

$$x(t) = s_{f_1}(t) + s_{f_2}(t) + s_{f_3}(t),$$

where $s_{f_i}(t) = \cos(2\pi f_i t)$, and $f_1 < f_2 < f_3$.

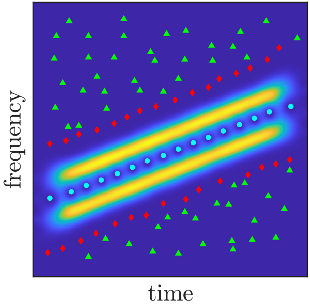
- If $\Delta f_{1,2} = \Delta f_{2,3} = \frac{3}{\sqrt{2\pi}}$, the longest edge of the Delaunay triangles covering the middle tone is $\ell \approx 1.46$.



Agenda:

Three Kinds of SZ

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 the result of interference between randomly located Gaussian *logons*.



Three Kinds of SZ

Destructive Interference

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

where $\Phi^g_x(t,f)$ is the phase of the complex valued $V^g_x(t,f)$.
Then, $|V^g_{x_1+x_2}(t,f)|^2 = 0$ 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.

Signal-Signal Spectrogram Zeros

Let x_1 and x_2 be deterministic signals. SS zeros appear where x_1 and x_2 fulfill the conditions of destructive interference.

Three Kinds of SZ

Noise-Noise Spectrogram Zeros

The spectrogram of (complex) white Gaussian noise can be expressed as[a]:

$$S_{\xi}^{g}(t,f) \approx \left| \sum_{k} V_{h_k}^{g}(t,f) e^{i\varphi_k} \right|^2 .$$

where $h_k(t) = a_k \, g(t - t_k) \, e^{i(2\pi f_k + \varphi_k)}$, also called *logons*, with random amplitude a_k, position (t_k, f_k), and phase φ_k.

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

Three Kinds of SZ

Signal-Noise Spectrogram Zeros

Let x_1 be a deterministic signal, and x_2 be a realization of noise.
If x_2 is real white Gaussian noise:

$$\mathbb{E}\left\{|V_{x_2}^{g}(t,f)|^2\right\} = \gamma_0^2. \tag{1}$$

We can then define a level curve

$$\Gamma = \left\{(t,f):\ S_{x_1}^{g}(t,f) = \gamma_0^2\right\}. \tag{2}$$

Three Kinds of SZ

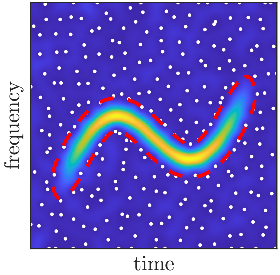
Signal-Noise Spectrogram Zeros

Let x_1 be a deterministic signal, and x_2 be a realization of noise.
If x_2 is real white Gaussian noise:

$$\mathbb{E}\left\{|V_{x_2}^{g}(t,f)|^2\right\} = \gamma_0^2. \tag{1}$$

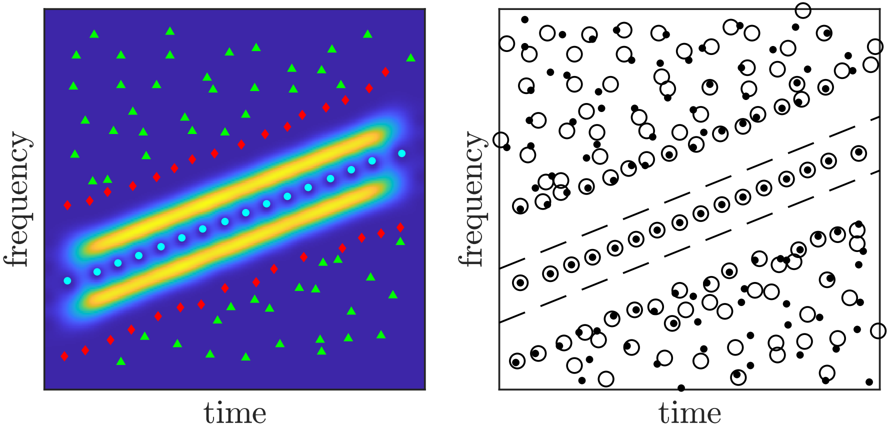
We can then define a level curve

$$\Gamma = \left\{(t,f):\ S_{x_1}^{g}(t,f) = \gamma_0^2\right\}. \tag{2}$$



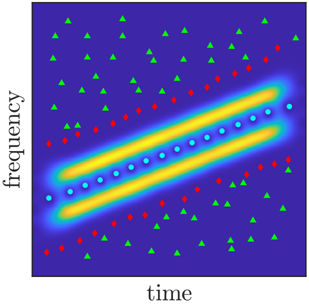
Agenda:

Stability to Noise Addition



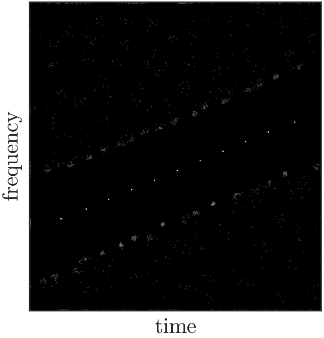
2D Histograms of the position of zeros

1. Get a new mixture $y_j(t) = x(t) + \xi(t) + \eta_j(t)$, where $\eta_j(t)$ is white Gaussian noise with variance $\gamma_j^2 = \beta\hat{\gamma}_0^2$, and:
   $\hat{\gamma}_0 = \frac{\sqrt{2}}{0.6745}$ median $\left(\left|\Re\left\{V_y^g(t,f)\right\}\right|\right)$.
2. Repeat this procedure for $j = 1, ..., J$ independent noise realizations.



2D Histograms of the position of zeros

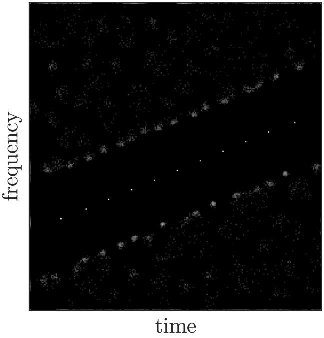
1. Get a new mixture
$y_j(t) = x(t) + \xi(t) + \eta_j(t)$,
where $\eta_j(t)$ is white Gaussian noise with variance $\gamma_j^2 = \beta\hat{\gamma}_0^2$, and:
$\hat{\gamma}_0 = \frac{\sqrt{2}}{0.6745}$ median $\left(\left|\Re\left\{V_y^g(t,f)\right\}\right|\right)$.
2. Repeat this procedure for $j = 1, ..., J$ independent noise realizations.
3. A 2D histogram $G[n, m]$ can then be obtained by counting the number of zeros that have fallen in each TF position (n, m).



$(J = 64)$

2D Histograms of the position of zeros

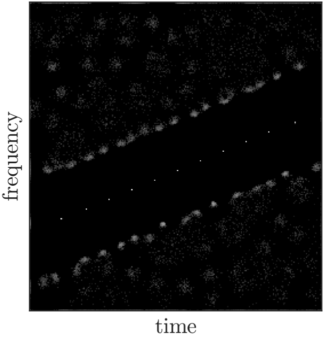
1. Get a new mixture
   $y_j(t) = x(t) + \xi(t) + \eta_j(t)$,
   where $\eta_j(t)$ is white Gaussian noise with variance $\gamma_j^2 = \beta \hat{\gamma}_0^2$, and:
   $\hat{\gamma}_0 = \frac{\sqrt{2}}{0.6745}$ median $\left(\left| \Re \left\{ V_y^g(t,f) \right\} \right| \right)$.
2. Repeat this procedure for $j = 1, ..., J$ independent noise realizations.
3. A 2D histogram $G[n,m]$ can then be obtained by counting the number of zeros that have fallen in each TF position (n,m).



$(J = 128)$

2D Histograms of the position of zeros

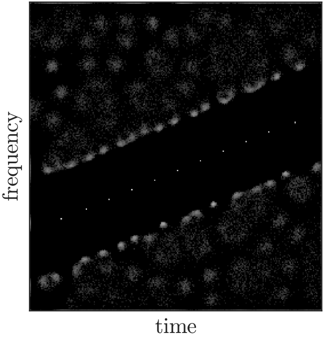
1. Get a new mixture
   $y_j(t) = x(t) + \xi(t) + \eta_j(t)$,
   where $\eta_j(t)$ is white Gaussian noise with variance $\gamma_j^2 = \beta\hat{\gamma}_0^2$, and:
   $\hat{\gamma}_0 = \frac{\sqrt{2}}{0.6745}$ median $\left(\left|\Re\left\{V_y^g(t,f)\right\}\right|\right)$.
2. Repeat this procedure for $j = 1, ..., J$ independent noise realizations.
3. A 2D histogram $G[n,m]$ can then be obtained by counting the number of zeros that have fallen in each TF position (n,m).



$(J = 256)$

2D Histograms of the position of zeros

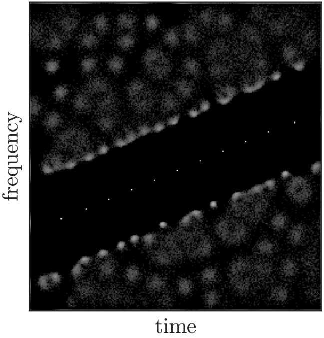
1. Get a new mixture
   $y_j(t) = x(t) + \xi(t) + \eta_j(t)$,
   where $\eta_j(t)$ is white Gaussian noise with variance $\gamma_j^2 = \beta\hat{\gamma}_0^2$, and:
   $\hat{\gamma}_0 = \frac{\sqrt{2}}{0.6745}$ median $\left(\left|\Re\left\{V_y^g(t,f)\right\}\right|\right)$.
2. Repeat this procedure for $j = 1, ..., J$ independent noise realizations.
3. A 2D histogram $G[n, m]$ can then be obtained by counting the number of zeros that have fallen in each TF position (n, m).



$(J = 512)$

2D Histograms of the position of zeros

1. Get a new mixture
   $y_j(t) = x(t) + \xi(t) + \eta_j(t)$,
   where $\eta_j(t)$ is white Gaussian noise with variance $\gamma_j^2 = \beta\hat{\gamma}_0^2$, and:
   $\hat{\gamma}_0 = \frac{\sqrt{2}}{0.6745}\ \text{median}\left(\left|\Re\left\{V_y^g(t,f)\right\}\right|\right)$.
2. Repeat this procedure for $j = 1, ..., J$ independent noise realizations.
3. A 2D histogram $G[n,m]$ can then be obtained by counting the number of zeros that have fallen in each TF position (n,m).

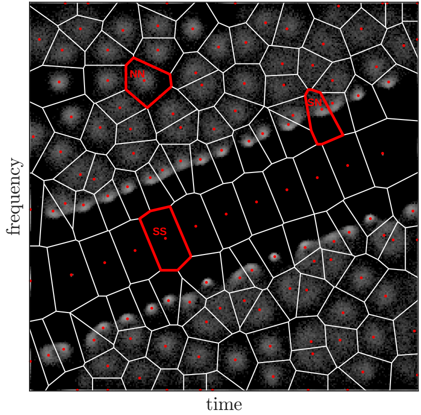


$(J = 1024)$

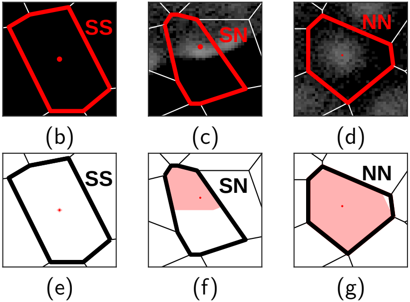
Agenda:

Local features for SZ classification

- Extracting features describing the local distribution of $G[n, m]$.



(a) Voronoi tessellation superimposed on the 2D histogram of spectrogram zeros.



Local features for SZ classification

Convex Hull & Voronoi Cell Area Ratio

For each $\mathbf{z} = (n_z, m_z)$ compute the area ratio:

$$\mathrm{AR}(\mathbf{z}) = \frac{A_{\mathrm{CH}}(\mathbf{z})}{A_{\mathcal{V}(\mathbf{z})}(\mathbf{z})},$$

where

1. $\mathcal{V}(\mathbf{z})$ is the Voronoi cell corresponding to $\mathbf{z}$.
2. CH is the convex hull of $\{(n, m) : (n, m) \in \mathcal{V}(\mathbf{z}) \wedge G[n, m] > 0\}$.

Maximum in Voronoi Cell

For each $\mathbf{z}$, compute the maximum of the histogram in $\mathcal{V}(\mathbf{z})$:

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

Unsupervised Classification

- Based on first *clustering*, and then *labeling*.
- We use Gaussian Mixture Models (GMM) for clustering in the feature space spanned by AR($\mathbf{z}$) and Max($\mathbf{z}$).
- The number K of clusters to search must be determined. For example, using the GAP criterion[4]:

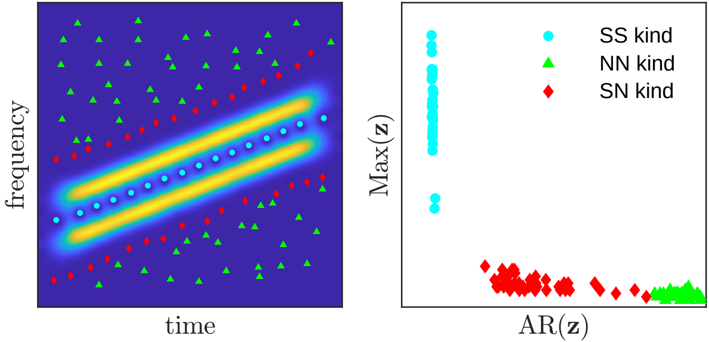
$$K = \underset{k \in \{1,2,3\}}{\arg\max} \ \mathbb{E}\{\log(W_k^{\star})\} - \log(W_k), \tag{3}$$

  - $W_k = \sum_{r=1}^{k} \frac{1}{N_r} D_r$
  - D_r is the sum of all the pairwise distances for the points in the r-th cluster.
  - N_r the number of elements in the cluster.
  - W_k^* is computed using Monte-Carlo simulations.

[4] Robert Tibshirani, Guenther Walther, and Trevor Hastie. "Estimating the number of clusters in a data set via the gap statistic". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.2 (2001), pp. 411–423.

Unsupervised Classification

- The clusters are then labeled according to the value of AR($\mathbf{z}$).



Algorithm Summary

Require: A noisy signal y, J, β.

Compute $V_y^g(t,f)$, the set $\mathcal{Z}_0$ of original SZs and Voronoi tessellation.

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

Compute the 2D histograms using J realizations and $\gamma_j^2 = \beta\gamma_0^2$.

Extract AR($\mathbf{z}$) and Max($\mathbf{z}$).

Apply GMM clustering, for $K \in \{1,2,3\}$.

Decide whether $K = 1$, 2 or 3.

Find the centroids C_i, $i = 1, \ldots, K$ of the detected clusters.

Sort clusters $C_{(i)}$ in ascending order of AR($\mathbf{z}$).

If $K = 1$, $C_{(1)} \to$ NN.

If $K = 2$, $(C_{(1)}, C_{(2)}) \to (\mathrm{SN}, \mathrm{NN})$.

If $K = 3$, $(C_{(1)}, C_{(2)}, C_{(3)}) \to (\mathrm{SS}, \mathrm{SN}, \mathrm{NN})$.

return K and the labels of each SZ in $\mathcal{Z}_0$

Signal domain estimation based on classified zeros

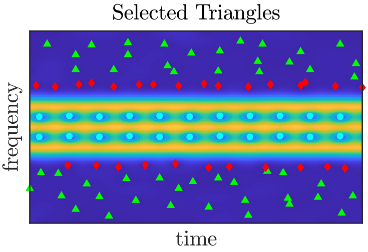
Different criteria to select Delaunay triangles

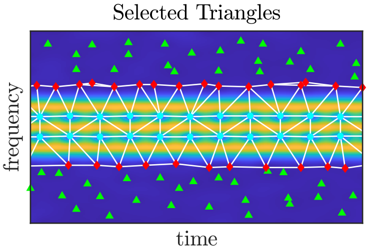
- Based on the classification of the spectrogram zeros, one can select Delaunay triangles that satisfy any of the following criteria:
  1. All its vertices are zeros of the SN kind (i.e. zeros located at the border of the signal domain).
  2. At least one of its vertices is a zero of the SS kind.

Signal domain estimation based on classified zeros

Different criteria to select Delaunay triangles

- Based on the classification of the spectrogram zeros, one can select Delaunay triangles that satisfy any of the following criteria:
  1. All its vertices are zeros of the SN kind (i.e. zeros located at the border of the signal domain).
  2. At least one of its vertices is a zero of the SS kind.

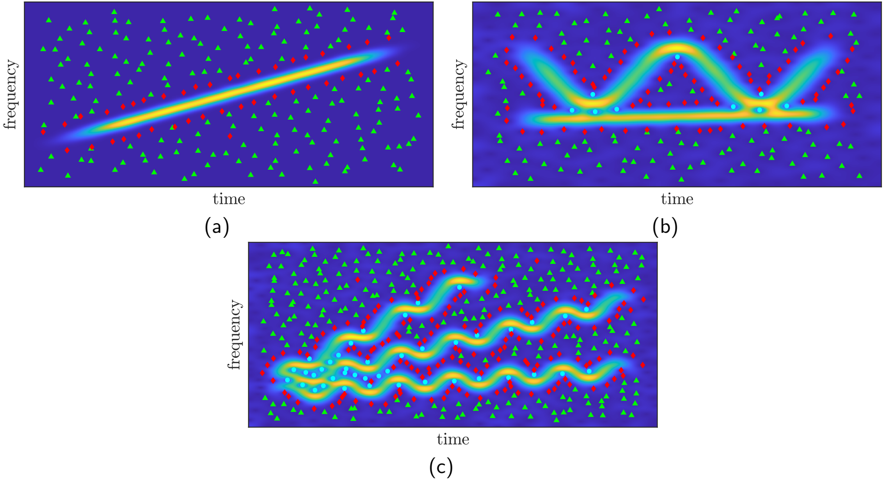




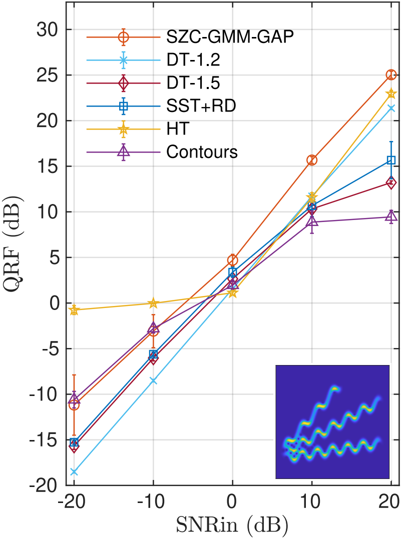
Agenda:

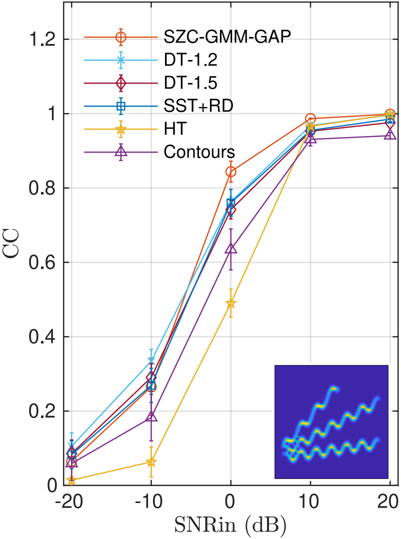
Results - Spectrogram Zeros

- Spectrogram zeros classification (using $\beta = 1.0$ and $J = N/4$):



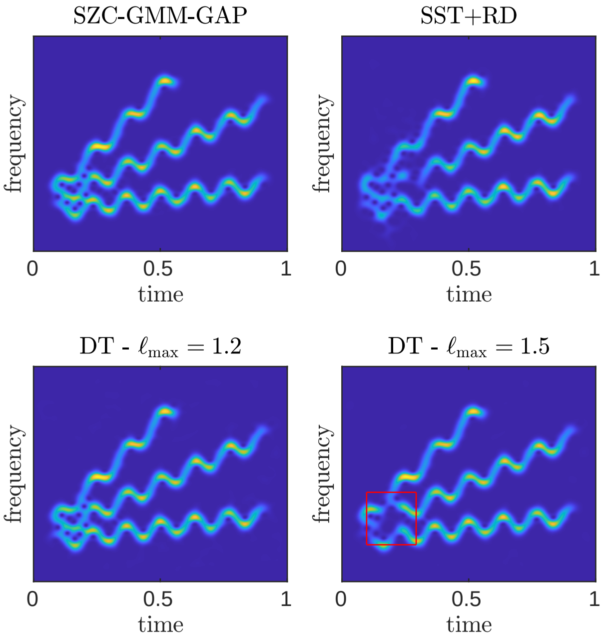
Results - Signal estimation





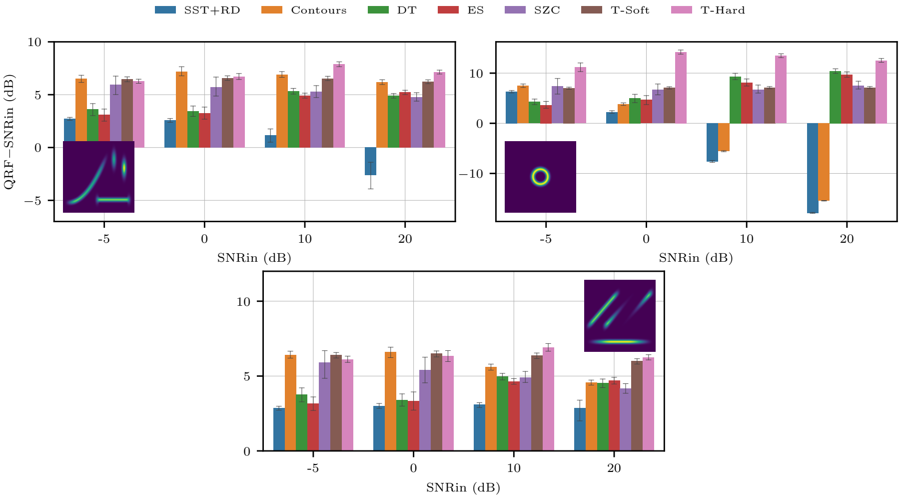
$$\mathrm{QRF} := 10\log_{10}\left(\frac{\|x\|_2^2}{\|x-\tilde{x}\|_2^2}\right)\ (\mathrm{dB}) \qquad \mathrm{CC} := \frac{\langle x,\tilde{x}\rangle}{\|x\|_2\|\tilde{x}\|_2}$$

Results - Signal estimation

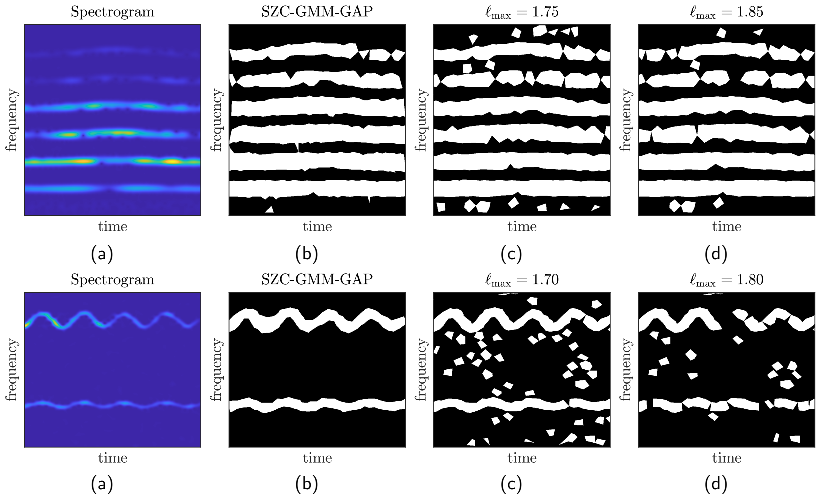


Results - Spectrogram Zeros

- Spectrogram zeros classification (using $\beta = 1.0$ and $J = N/4$):



Results - Audio signals



Results

- Average execution time (in seconds):

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

† CaHa stands for Calinsky-Harabasz criterion.

Agenda:

Conclusions

- We proposed a study of the spectrogram zeros from the perspective of interference in the TF.

Conclusions

- We proposed a study of the spectrogram zeros from the perspective of interference in the TF.
- The classification of the spectrogram zeros provides different criteria to estimate the signal domain using the Delaunay triangulation.

Conclusions

- We proposed a study of the spectrogram zeros from the perspective of interference in the TF.
- The classification of the spectrogram zeros provides different criteria to estimate the signal domain using the Delaunay triangulation.
- The approach works better in low SNR scenarios and in the presence of interference.

Conclusions

- We proposed a study of the spectrogram zeros from the perspective of interference in the TF.
- The classification of the spectrogram zeros provides different criteria to estimate the signal domain using the Delaunay triangulation.
- The approach works better in low SNR scenarios and in the presence of interference.
- Improvement comes with the cost of an increased computation time, mainly due to the 2D histograms.
- The noise is assumed to be broadband.

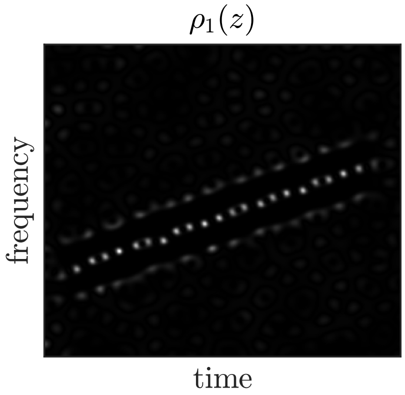
Future work

- Zeros of the SS kind implies strong interference between components.
  - Detecting *time-frequency bubbles*.
  - Improving window selection to reduce interference between signal components.

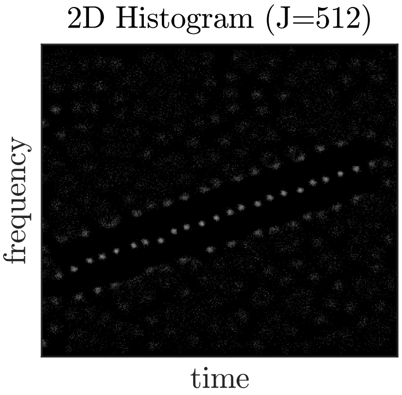
Future work

- Zeros of the SS kind implies strong interference between components.
  - Detecting *time-frequency bubbles*.
  - Improving window selection to reduce interference between signal components.
- The introduced 2D histograms are interesting objects by themselves.
  - Describing/ modeling their underlying distribution.
  - Reducing the number of noise realizations needed for computation.

Future work



($\approx$ 0.01 s)



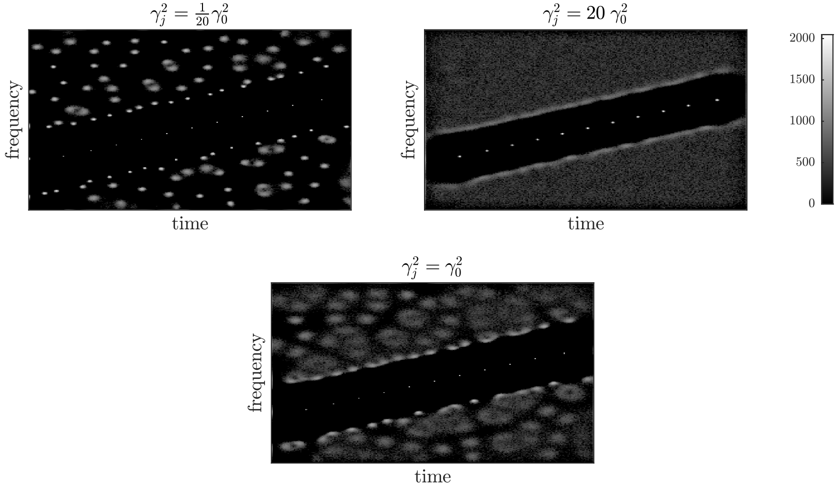
($\approx$ 2 s)

Thank you for your attention!

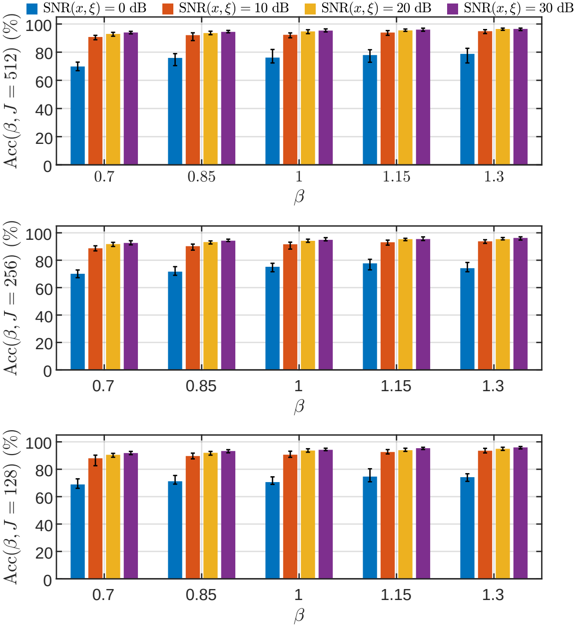
`https://github.com/jmiramont/spectrogram-zeros-classification`

Juan M. Miramont, François Auger, Marcelo A. Colominas, Nils Laurent, and Sylvain Meignen. *"Unsupervised classification of the spectrogram zeros with an application to signal detection and denoising*", Signal Processing (2023).

2D Histograms of the position of zeros



Setting parameters



Variations of the proposed approach