A New Ridge Detector Localizing Strong Interference in Multicomponent Signals in the Time-Frequency Plane

Sylvain Meignen, Marcelo Colominas

A New Ridge Detector Localizing Strong Interference in Multicomponent Signals in the Time-Frequency Plane

Sylvain Meignen and Marcelo A. Colominas

***Abstract*—In this paper, we define a new ridge detector that enables to localize strong interference in multicomponent signals in the time-frequency (TF) plane. Each mode of a multicomponent signal can usually be associated with a ridge in the TF plane, but this is no longer the case when strong interferences occur in the signal. The new ridge detector we propose is thus designed to determine when such situations happen in the TF plane. We show that this knowledge helps to determine an appropriate window length in the definition of the spectrogram, as well as the nature of the strong interference detected. An application of the proposed approach to voice signals concludes the paper.**

***Index Terms*—Time-frequency analysis, short-time Fourier transform, spectrogram ridges.**

I. Introduction

The analysis of *multicomponent signals* (MCSs) using *time-frequency representations* (TFRs) has been the subject of intense research in the last few decades, in particular because the modes making up an MCS are associated with specific time-frequency (TF) regions located around curves, called *ridges* [1], [2]. These are estimates of the *instantaneous frequencies* (IFs) of the modes [3], [4], the quality of which is tightly related to how well the modes are separated in the TF plane. To detect such ridges, there exist a whole set of techniques based on various assumptions on the modes to be extracted, and on the type of TFR used to represent the signal. For instance, an algorithm was proposed in [5], using the so-called B-distribution, and based on the fact that, at each time instant, a mode corresponds to a local maximum of that TFR along the frequency axis, and a ridge to a chain linking all these local maxima. However, in that paper, such a chain is obtained in an ad-hoc manner. Furthermore, the question of crossing modes was not addressed. In [6], a similar technique was proposed that applied precisely to that case, but assuming the modes last for the whole time-span. At the same time, approaches based on blind source separation were also introduced [7]. All these approaches, though interesting, rely on ad-hoc parameters to link the ridge points. To circumvent this limitation, more adaptive techniques were proposed, in which the ridge points are linked using a signal-based modulation operator [8], [9]. In these techniques, the question of crossing modes was however not considered and, for that purpose, a three-dimensional time-frequency chirp-rate decomposition was introduced in [10]. Indeed, two crossing modes in the TF plane have different chirp-rates, and, in that context, the ridges associated with the modes are curves that no longer cross each other in the three-dimensional space associated with the decomposition. In spite this approach is definitely of interest, to detect ridges that way is very complicated in noisy situations, and also time-consuming. Another important issue faced by ridge detection is when the modes are close in the TF plane. Indeed, TFRs commonly depend on an analysis window, the choice of which is essential to ridge detection. As to consider a single analysis window may be troublesome, approaches adapting the window length to time and components were proposed [11], [12]. The separation of close modes was also recently investigated using fractional wavelet transform [13] or short-time fractional Fourier transform [14], instead of the classical wavelet or short-time Fourier transforms.

To introduce the topic of this paper, we first remark that mode crossings or close modes situations only occur at very specific location of the TF plane, and, to assume in the ridge detection model that these may happen everywhere in the TF plane is definitely not relevant. Our concern in this paper is thus to characterize precisely where such interferences take place, using the spectrogram as TFR. Note that the ideas developed in this paper could certainly be extended to other types of quadratic TFR, but the theoretical results derived here would certainly be harder to obtain.

When two modes are close in the TF plane, it is sometimes impossible to associate a ridge with each of them. A ridge is often defined as a chain of *local maxima of the spectrogram along the frequency axis* (LMFs) [9], and two close modes may result in a single LMF at some time instant. Similarly, two crossing modes will result in a single LMF at some time instant. In the present paper, one of our goals is to find where such situations occur. To do so, we first build an adaptive *ridge detector* (RD) that can follow a chain of LMFs even when it merges with another chain. We found our new RD on previous ones that build each ridge point by point, following a direction in the TF plane given by a modulation operator based on a local linear chirp approximation for the modes [8], [9]. However, in those approaches, the detected ridges cannot merge, and a separability hypothesis has to be enforced on the modes, which prevents the analysis of very close or crossing modes in the TF plane. By slightly modifying the RD

S. Meignen is with the Jean Kuntzmann Laboratory, University Grenoble Alpes and CNRS 5225, Grenoble 38401, France (email: sylvain.meignen@univ-grenoble-alpes.fr), while M. A. Colominas is with the Institute for Research and Development in Bioengineering and Bioinformatics (IBB), CONICET, Ruta Prov. 11 Km. 10, Oro Verde, Entre Ríos, Argentina (e-mail: macolominas@conicet.gov.ar). This work was supported in part by the ANR ASCETE project with grant number ANR-19-CE48-0001-01, and by the STIC AmSud ASPMLM-Voice grant.

proposed in [9], we allow the detected ridges to merge locally, the merging points being called singular points. A particular focus is then put on the ability of our new RD to detect a special type of interference, called *time-frequency bubbles* (TFB), first mentioned in [15], that arise when two modes are crossing, or are very close in the TF plane. Inspired by a recent work on the existence of TFBs in simple situations [16], we analyze the singularities of the modulation operator when TFBs are present. Then, we propose different applications of the detection of TFBs to the localization of the crossing of components and of other types of strong interference. We finally explain how the proposed TFB detector can be used to determine an appropriate analysis window length in the definition of the spectrogram.

The paper is organized as follows. In the following section, we introduce the notation used throughout the paper, and then briefly describe adaptive RDs which are the basis to our new RD [8], [9], defined in Sec. III. After having introduced the concept of TFBs in Sec. IV, we explain how the new RD we propose can be used to detect them, in Sec. V-A, but also singular points, in Sec. V-B. Then, we study the singularities of the modulation operator used in the definition of our new RD, in a simple case, when TFBs are present, in Sec. VI.

In the Results section, we illustrate the procedure for the detection of TFBs in the presence of mode crossings or other types of interferences, and also in what way to detect these interferences enables us to find an appropriate window length to represent the spectrogram of the signal. Finally, we illustrate the interest of the detection of singular points on a voice signal to determine an appropriate window length in the spectrogram, and then to help improve the accuracy of its reassigned version.

II. Notation, Ridge Definition, and Adaptive Ridge Detection Basics

A. STFT and Model Signal

Considering a signal $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and a real window $h \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, the STFT of f is defined as:

$$\begin{aligned} V_f^h(t,\eta) = & \int_{\mathbb{R}} f(\tau) h(\tau - t) e^{-2i\pi(\tau - t)\eta} d\tau \\ = & \quad |V_f^h(t,\eta)| e^{2i\pi\Psi(t,\eta)}, \end{aligned} \tag{1}$$

its spectrogram being the square modulus of V_f^h. The MCSs we study in this paper are defined as the superimposition of P AM-FM components, namely

$$f(t) = \sum_{p=1}^{P} f_p(t), \text{ with } f_p(t) = A_p(t) e^{2i\pi\phi_p(t)}, \tag{2}$$

in which the *instantaneous amplitudes* (IAs) $A_p(t)$ are positive, as well as the *instantaneous frequencies* (IFs) $\phi_p'(t)$.

B. Ridge Definition

There exist several ways to characterize the ridges associated with an MCS in the TF plane. A first characterization was proposed in [15], stating that a ridge point should be a solution to

$$\partial_t \Psi(t,\eta) = \eta. \tag{3}$$

To make the connection between that definition and the more classical one that views ridge points as LMFs [4], [8], [9], namely TF points such that $\partial_\eta |V_f^h(t,\eta)|^2 = 0$, one remarks that when $h(t) = e^{-\pi \frac{t^2}{\sigma^2}}$, then $h'(t) = -\frac{2\pi}{\sigma^2} t h(t)$, and thus:

$$\begin{aligned} \partial_t \Psi(t,\eta) &= \frac{1}{2\pi} \Im \left\{ \frac{\partial_t V_f^h(t,\eta)}{V_f^h(t,\eta)} \right\} \\ &= \eta - \frac{1}{2\pi} \Im \left\{ \frac{V_f^{h'}(t,\eta)}{V_f^h(t,\eta)} \right\} = \eta + \Im \left\{ \frac{1}{\sigma^2} \frac{V_f^{th}}{V_f^h} \right\} \\ &= \eta - \Im \left\{ \frac{\partial_\eta V_f^h}{2i\pi\sigma^2 V_f^h} \right\} = \eta + \frac{1}{4\pi\sigma^2} \frac{\partial_\eta |V_f^h|^2}{|V_f^h|^2}, \end{aligned} \tag{4}$$

in which $\Im\{X\}$ denotes the imaginary part of the complex number X. This means that the solutions to Eq. (3) actually correspond to LMFs. With this in mind, it is natural to define a ridge as a chain of LMFs in the TF plane. In the next section, we recall some recent approaches that were used to detect such ridges, on which we base our new approach.

C. Adaptive Ridge Detector Based on Modulation Operator

The new RD we propose, whose description is detailed in Sec. III, is inspired by some adaptive RDs respectively developed in [8] and [9], which we briefly describe hereafter.

The RD called *modulation based ridge detector* (MB-RD) introduced in [8], uses the following complex modulation operator based on a linear chirp approximation for the modes [17]

$$\tilde{q}_f(t,\eta) = \frac{1}{2i\pi} \frac{V_f^{h''}(t,\eta) V_f^h(t,\eta) - (V_f^{h'}(t,\eta))^2}{V_f^{th}(t,\eta) V_f^{h'}(t,\eta) - V_f^{th'}(t,\eta) V_f^h(t,\eta)}, \tag{5}$$

in which $V_f^{h'}, V_f^{th}, V_f^{h''}, V_f^{th'}$ are respectively the STFTs of f computed with windows $h'(t), th(t), h''(t)$ and $th'(t)$. It is shown in [17] that $\widehat{q_f}(t,\eta) = \Re\{\tilde{q}_f(t,\eta)\}$, where $\Re\{X\}$ is the real part of complex number X, consists of an estimate of the frequency modulation of the closest mode to the point (t,η) in the TF plane. Note that when $f(t) = A(t) e^{2i\pi\phi(t)}$ is a Gaussian modulated linear chirp, then $\widehat{q_f}(t,\eta) = \phi''(t)$. Now, if one considers the model signal described Eq. (2), one has $\widehat{q_f}(t,\eta) \approx \phi_p''(t)$ if the modes are separated at time t, if (t,η) is in the TF domain associated with f_p, and if the latter is well approximated by a linear chirp at time t.

The idea of MB-RD is then to extract the ridges one after the other using the modulation operator $\widehat{q_f}$. In practice, the latter is discretized on a TF grid, on which n is the time index, ranging from 0 to $N-1$ and k is the frequency index, ranging from 0 to $M-1$. In that context, $\widehat{q_f}$ is associated with an $N \times M$ matrix $\widehat{\mathbf{q}_f}$. More precisely, if the signal f lasts for T seconds then the time index n corresponds to the time $\frac{n}{N}T$, and then, the frequency index k to the frequency $\frac{k}{M}\frac{N}{T}$. With this in mind, to build the first ridge with MB-RD,

denoted here by the vector $\boldsymbol{\varphi}$ of frequency indices of length N, one picks a time index n, and computes

$$\boldsymbol{\varphi}[n] = \underset{0\leq k\leq M-1}{\operatorname{argmax}} |\mathbf{V}_f^h[n,k]|, \tag{6}$$

where $\mathbf{V}_f^h$ is the matrix corresponding to the discretization of V_f^h on the TF grid. Then, the next ridge point is computed through:

$$\boldsymbol{\varphi}[n+1] := \underset{k,\ |k-\boldsymbol{\varphi}[n]-\frac{MT^2}{N^2}\widehat{\mathbf{q}_f}[n,\boldsymbol{\varphi}[n]]|\leq C}{\operatorname{argmax}} |\mathbf{V}_f^h[n+1,k]|. \tag{7}$$

At this stage, it is important to understand the meaning of the set for the index k in this expression. Indeed, if one performs a first order Taylor expansion of ϕ_p', assuming f_p is a linear chirp, one obtains:

$$\phi_p'\left(\frac{n+1}{N}T\right) = \phi_p'\left(\frac{n}{N}T\right) + \frac{T}{N}\phi_p''\left(\frac{n}{N}T\right).$$

Now, if one assumes $\phi_p'(\frac{n}{N}T) \approx \frac{\boldsymbol{\varphi}[n]}{M}\frac{N}{T}$ for some p (meaning the ridge point detected is associated with mode p), one has:

$$\begin{aligned}\frac{\boldsymbol{\varphi}[n+1]}{M}\frac{N}{T} &\approx \frac{\boldsymbol{\varphi}[n]}{M}\frac{N}{T} + \frac{T}{N}\phi_p''\left(\frac{n}{N}T\right)\\ &\approx \frac{\boldsymbol{\varphi}[n]}{M}\frac{N}{T} + \frac{T}{N}\widehat{\mathbf{q}_f}[n,\boldsymbol{\varphi}[n]]\\ \Leftrightarrow \boldsymbol{\varphi}[n+1] &\approx \boldsymbol{\varphi}[n] + \frac{MT^2}{N^2}\widehat{\mathbf{q}_f}[n,\boldsymbol{\varphi}[n]],\end{aligned} \tag{8}$$

which justifies the range for k in Eq. (7). The parameter C is then used to cope with potential errors in the approximation given by Eq. (8). Then, ridge detection continues from time index $n+1$ until time index $N-1$, and then also backward from time index $n-1$ to time index 0.

Unfortunately, the ridge obtained that way is not necessarily made of discrete LMFs, i.e. local maxima of along the frequency axis of the discretized spectrogram. In the sequel, and for the sake of simplicity, we also denote by LMF a discrete LMF. In addition, $\widehat{\mathbf{q}_f}$ is not necessarily accurate enough because, strictly speaking, it indicates a good direction to pursue ridge detection only if the associated mode is a linear chirp. Furthermore, ridge detection is continued even if the detected points are irrelevant, and this often happens in noisy situations. Finally, the detected ridge depends on the initialization time index n [8].

To deal with all these issues, the concept of *relevant ridge portions* (RRPs) was introduced in [9]. In a nutshell, an RRP is a set of frequency indices associated with a set of successive time indices, meaning an RRP does not necessarily last for the whole time-span. Thus, to define a generic RRP, still denoted by $\boldsymbol{\varphi}$, one first selects a time index n and a frequency index k, such that $[n,k]$ is an LMF and sets $\boldsymbol{\varphi}[n] := k$. Then, following (8), one defines:

$$F[k] := k + \frac{MT^2}{N^2}\widehat{\mathbf{q}_f}[n,k], \tag{9}$$

meaning that $\boldsymbol{\varphi}[n+1] \approx F[\boldsymbol{\varphi}[n]]$, when the modulation operator accurately estimates the chirp-rate on the ridge (the letter F being chosen to mean "forward"). Now, assuming $\boldsymbol{\varphi}[n+1]$ is known, and considering as previously a first order Taylor expansion of the frequency, one also have the relation $\boldsymbol{\varphi}[n] \approx B[\boldsymbol{\varphi}[n+1]]$, with

$$B[k] := k - \frac{MT^2}{N^2}\widehat{\mathbf{q}_f}[n+1,k], \tag{10}$$

where the letter B is used to mean "backward". Using these notations, $\boldsymbol{\varphi}[n+1]$ is then defined as satisfying $[n,\boldsymbol{\varphi}[n]] \sim [n+1,\boldsymbol{\varphi}[n+1]]$, where the relation $\sim$ corresponds to the following definition [9]:

Definition II.1. *Let $\mathbf{m}$ be a vector with values in $[\![0,M-1]\!]$, and $[n,\mathbf{m}[n]]$ and $[n+1,\mathbf{m}[n+1]]$ two LMFs, then:*

$$\begin{aligned}&[n,\mathbf{m}[n]] \sim [n+1,\mathbf{m}[n+1]]\\ \Leftrightarrow &\begin{cases}\mathbf{m}[n+1] &:= \underset{k,\ [n+1,k]\ LMF}{\operatorname{argmin}} |k - F[\mathbf{m}[n]]|\\ \mathbf{m}[n] &:= \underset{k,\ [n,k]\ LMF}{\operatorname{argmin}} |B[\mathbf{m}[n+1]] - k|,\end{cases}\end{aligned}$$

with the functions F and B defined in Eq. (9) *and* (10)*, respectively.*

This tells us that $[n+1,\boldsymbol{\varphi}[n+1]]$ (resp. $[n,\boldsymbol{\varphi}[n]]$) is the closest LMF to $[n,\boldsymbol{\varphi}[n]]$ (resp. $[n+1,\boldsymbol{\varphi}[n+1]]$) at time index $n+1$ (resp. n) in the direction given by $\widehat{\mathbf{q}_f}[n,\boldsymbol{\varphi}[n]]$ (resp. $-\widehat{\mathbf{q}_f}[n+1,\boldsymbol{\varphi}[n+1]]$). So, $[n,\boldsymbol{\varphi}[n]] \sim [n+1,\boldsymbol{\varphi}[n+1]]$ also means that $\widehat{\mathbf{q}_f}$ computed at these LMFs correspond to a stable orientation. The relation $\sim$ is then used to define $\boldsymbol{\varphi}$ iterating the procedure forward and backward, from time index n. Note that, when the relation $\sim$ cannot be satisfied at a time index, the detection procedure stops, which is why one uses the term "ridge portion", hence the notation RRP. In this approach, as the ridge points are linked by means of relation $\sim$, the stability of the modulation operator along a ridge is ensured.

III. Definition of the New Ridge Detector

Because RRPs are defined by means of the relation $\sim$, a ridge point cannot belong to several RRPs, and we propose, in the following section, to slightly change the definition of ridge portions to allow for their merging.

A. Definition of the New Ridge Detector: Noiseless Case

The ridge portions we intend to detect are denoted by ERRPs (for *extended relevant ridge portions*) and are slightly different from RRPs introduced in the previous section, in the following way. To build the first ERRP, which we denote by $\boldsymbol{\varphi}_1$, we consider the set

$$\mathcal{D}_1 = \{[n,k]\ LMF,\ n\in[\![0,N-1]\!],\ k\in[\![0,M-1]\!]\}, \tag{11}$$

and then an initial point:

$$[n_1,\boldsymbol{\varphi}_1[n_1]] = \underset{[n,k]\in\mathcal{D}_1}{\operatorname{argmax}} |\boldsymbol{V}_f^h[n,k]|. \tag{12}$$

Starting with $[n_1,\boldsymbol{\varphi}_1[n_1]]$, the associated ERRP corresponds to the RRP detected using the procedure described in the previous section, plus the two ending points at which the relation $\sim$ is no longer satisfied. Such points can belong to several chains of LMFs, contrary to the other points on the ERRP. Having defined $\boldsymbol{\varphi}_1$, its associated TF domain reads:

$$\mathcal{E}(\boldsymbol{\varphi}_1) := \{[n,\boldsymbol{\varphi}_1[n]],\ \boldsymbol{\varphi}_1[n] \text{ defined}\}, \tag{13}$$

and then, we introduce

$$\mathcal{D}_2 := \mathcal{D}_1 \setminus \mathcal{E}(\boldsymbol{\varphi}_1). \tag{14}$$

To detect the next ERRP, we first consider

$$[n_2, \boldsymbol{\varphi}_2[n_2]] := \underset{[n,k]\in\mathcal{D}_2}{\operatorname{argmax}} |\boldsymbol{V}_f^h[n,k]|, \tag{15}$$

and $\boldsymbol{\varphi}_2$ is detected following the same procedure as for $\boldsymbol{\varphi}_1$, with its associated TF domain being

$$\mathcal{E}(\boldsymbol{\varphi}_2) := \{[n, \boldsymbol{\varphi}_2[n]],\ \boldsymbol{\varphi}_2[n] \text{ defined}\}. \tag{16}$$

To compute the following ERRPs (for $p \geq 3$), we proceed iteratively by defining:

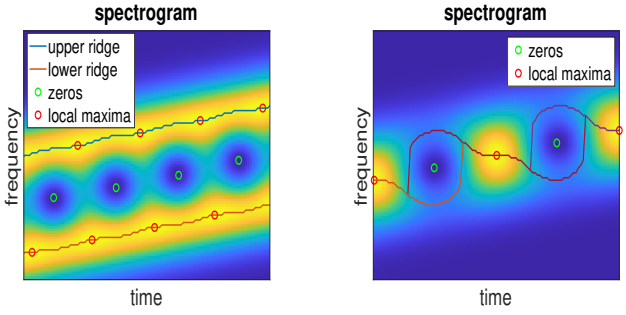
$$\mathcal{D}_p := \mathcal{D}_{p-1} \setminus \mathcal{E}(\boldsymbol{\varphi}_{p-1}), \tag{17}$$

and then, starting with

$$[n_p, \boldsymbol{\varphi}_p[n_p]] := \underset{[n,k]\in\mathcal{D}_p}{\operatorname{argmax}} |\boldsymbol{V}_f^h[n,k]|,$$

the p^{th} ERRP is detected in the same way as for $\boldsymbol{\varphi}_1$, and $\mathcal{E}(\boldsymbol{\varphi}_p)$ correspond to its TF domain. Such a procedure is carried out until the detected ERRP have a length below some predefined threshold.

An illustration of the behavior of the detector of ERRPs is given in Fig. 1 (a) and (b) on a signal made of the sum of two parallel linear chirps, respectively when the two modes are sufficiently far apart to enable separate ridge detection and when the latter are so close that the ridges associated with the modes merge for some time indices (in that case, the minimal length that stops the detection procedure is set to 10). From now on, we denote by ERRP-RD (for ERRP ridge detector) the just introduced RD. It is important to recall here that local minima of the spectrogram are necessarily zeros, since the spectrogram can be viewed as an analytic function [18]. The zeros in Fig. 1. (a) and (b) are computed that way.



(a) (b)

Fig. 1: (a): Spectrogram of two parallel linear chirps with the same amplitude, when each of them can be associated with a chain of LMFs (upper and lower ridges); (b): Spectrogram of two parallel linear chirps when TFBs are present. We plot the ERRPs computed with ERRP-RD detailed in Sec. III, as well as the zeros and the local maxima of the spectrogram.

B. Definition of the New Ridge Detector: Noisy Case

In the case of noisy signals, we consider $\tilde{f} = f + \varepsilon$, with ε a complex Gaussian white noise. To adapt ERRP-RD to that context, we propose to initialize the detection of ERRPs using only LMFs that are associated with the signal part of the spectrogram with a high probability. To do this, we first recall that, as the added complex noise ε is Gaussian white with variance σ_ε^2, $\boldsymbol{V}_\varepsilon^h[n,k]$ is also Gaussian with zero mean and satisfies [19]:

$$\text{Var}\left(\Re\{\boldsymbol{V}_\varepsilon^h[n,k]\}\right) = \text{Var}\left(\Im\{\boldsymbol{V}_\varepsilon^h[n,k]\}\right) = \sigma_\varepsilon^2 \|h\|_2^2,$$

with $\|h\|_2$ the l^2-norm of h. Then, remarking that $\frac{|\boldsymbol{V}_\varepsilon^h|^2}{\sigma_\varepsilon^2\|h\|_2^2}$ is χ_2 distributed with two degrees of freedom, and assuming the variance of the noise σ_ε^2 is known, the probability that $|\boldsymbol{V}_\varepsilon^h[n,k]| \geq \beta\sigma_\varepsilon\|h\|_2$ is lesser than 10% and 1% if $\beta = 2$ and 3, respectively. To estimate $\gamma = \sigma_\varepsilon\|h\|_2$, we use the robust estimator proposed in [20]:

$$\hat{\gamma} := \frac{\text{median}\left|\Re\left\{\boldsymbol{V}_{\tilde{f}}^h[n,k]\right\}_{n,k}\right|}{0.6745},$$

where median represents the median of the coefficients. Based on this analysis, we define:

$$\mathcal{S}(\beta) := \left\{[n,k], |\boldsymbol{V}_{\tilde{f}}^h[n,k]| \geq \beta\hat{\gamma}\right\}, \tag{18}$$

and, referring to [9], we consider that the LMFs involved in the detection of ERRPs belong to $\mathcal{S}(3)$, which guarantees that the detected ERRPs do not propagate inside the noisy part of the spectrogram. Thus, to detect the first ERRP, we use the same algorithm as previously, but starting this time with

$$[n_1, \boldsymbol{\varphi}_1[n_1]] := \underset{[n,k]\in\tilde{\mathcal{D}}_1}{\operatorname{argmax}} |\boldsymbol{V}_{\tilde{f}}^h[n,k]|, \tag{19}$$

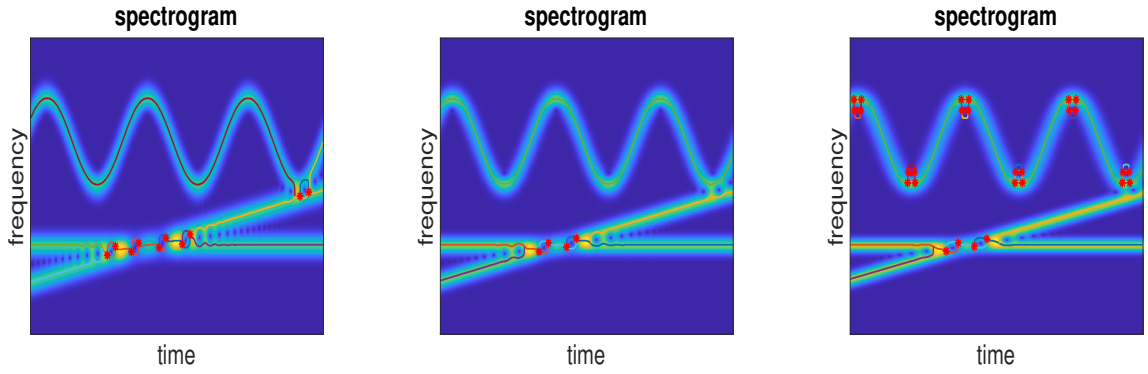
with $\tilde{\mathcal{D}}_1 = \mathcal{D}_1 \cap \mathcal{S}(3)$, and then performing ERRP detection as previously with the additional constraint that the points on the ERRP belong to $\tilde{\mathcal{D}}_1$. The set $\mathcal{E}(\boldsymbol{\varphi}_1)$ being defined as in the noiseless case, we put:

$$\tilde{\mathcal{D}}_2 = \tilde{\mathcal{D}}_1 \setminus \mathcal{E}(\boldsymbol{\varphi}_1). \tag{20}$$

Then the detection of $\boldsymbol{\varphi}_2$ follows the same procedure as that in the noiseless case, replacing $\mathcal{D}_2$ by $\tilde{\mathcal{D}}_2$ to find the initial points, and bearing in mind that the LMFs on the detected ERRP have to be all in $\tilde{\mathcal{D}}_1$. Finally, the detection of the next ERRPs ($p \geq 3$) involves points in $\tilde{\mathcal{D}}_1$ and is based on the same framework as in the noiseless case, replacing $\mathcal{D}_p$ by $\tilde{\mathcal{D}}_p := \tilde{\mathcal{D}}_{p-1} \setminus \mathcal{E}(\boldsymbol{\varphi}_{p-1})$. From now on, we denote this RD by ERRP-RD$_{noise}$.

IV. Time-Frequency Bubbles Definition

A particular type of strong interference in the TF plane corresponds to the notion of *Time-Frequency Bubbles* (TFBs), first mentioned in [15], but without a clear definition. They occur when the signal is locally associated with a "circular" set of LMFs in the TF plane. Note that such structures can also be present in noise, as a result of the interaction between two *logons* [21]. To clarify this notion of TFBs in our discrete TF setting, we propose the following definition using ERRPs:



(a) (b) (c)

Fig. 2: (a): spectrogram of a three mode signal along with the ridges superimposed, in the definition of the Gaussian window $\sigma = 0.03$, the stars correspond to TFB points (ERRP-RD is used as RD); (b): same as (a) but with $\sigma = 0.04$; (c): same as (a) but with $\sigma = 0.05$.

Definition IV.1. *Two ERRPs create a TFB when they have two points in common, and when, in the region delimited by these ERRPs between these two points, there is a single zero of the spectrogram.*

As the zeros correspond to local minima of the spectrogram, in Definition IV.1, we could have alternatively said that, in the region delimited by the two RRPs between the two merging points, there is a unique local minimum. An illustration of such TFBs is given in Fig. 1 (b), in the case of two close parallel linear chirps (ERRPs are detected using ERRP-RD).

When the signal is the sum of two pure tones, the existence of TFBs on the spectrogram is equivalent to the two modes being associated with a single LMF for some time instant. In a continuous TF setting, and for the Gabor transform, the condition for such a situation to occur was studied in detail in [16], and lead to:

Proposition IV.1. *Assume* $f(t) = f_1(t) + f_2(t)$ *with* $f_1(t) = Ae^{i2\pi\xi_1 t}$ *and* $f_2(t) = e^{i2\pi\xi_2 t}$*, where* $\xi_1 < \xi_2$*, and* $h(t) = e^{-\pi\frac{t^2}{\sigma^2}}$*. Then* $|V_f^h(t,\cdot)|^2$ *has two LMFs and a local minimum if and only if* $\alpha := \sqrt{\frac{\pi}{2}}\sigma(\xi_2 - \xi_1) > 1$ *and*

$$|\log(A)| < -2\,\mathrm{arcosh}(\alpha) + 2\alpha\sqrt{\alpha^2 - 1}.$$

In any other cases, there exists some time t *where* $|V_f^h(t,\cdot)|^2$ *has a unique extremum which is an LMF.*

In Proposition IV.1, $|V_f^h(t,\cdot)|^2$ means one considers the spectrogram with respect to the frequency variable at time t. The proof of this proposition is available in [16]. Though this result is derived for a specific Gaussian window h, it can be generalized to any Gaussian window using an appropriate renormalization. To better understand the meaning of Proposition IV.1, we may say that as soon as $\alpha \leq 1$ or $|\log(A)| < -2\,\mathrm{arcosh}(\alpha) + 2\alpha\sqrt{\alpha^2 - 1}$, TFBs appear. In particular, whatever the amplitude A, if the frequency difference between the two modes is smaller than $\frac{1}{\sigma}\sqrt{\frac{2}{\pi}}$, then TFBs are present, and, if this condition is not fulfilled, TFBs also appear provided A is large enough. Note that it was also shown in [16] that such a result can easily be extended to parallel linear chirps.

Nevertheless, the conditions put forward in Proposition IV.1 are of limited practical interest, since they are restricted to the case of pure tones (or parallel linear chirps), of which the parameters are known. On the contrary, Definition IV.1, is interesting from a practical point of view, since it is only based on the detected ERRPs and does not need the knowledge of the parameters of the modes. To prove the relevance of Definition IV.1, we will check later that it is in accordance with Proposition IV.1, in the case of the sum of two pure tones with known parameters.

V. Detecting Strong Interference with ERRP-RD

Interference between the modes of an MCS are ubiquitous in the TF plane. Here, we investigate the detection of strong interference, corresponding to the case where some ERRPs associated with the signal merge.

A. TFBs Detection Using ERRP-RD

In this section, we explain how to use ERRP-RD to detect TFBs. To find out the pairs of TF points associated with a TFB following Definition IV.1, we apply the following procedure:

- Find the ERRPs that have two points in common.
- Compute the number of zeros of the spectrogram inside the TF domain delimited by these two points and the associated ERRPs.
- If this number equals one, this pair of points is associated with a TFB.

We coin such a pair of points *TFB points*, as they localize a TFB in the TF plane. An illustration of the detection of TFBs using such an approach is given in Fig. 2, in which we see that, depending on σ, the number of TFB points vary. A small σ, i.e. a small window length, leads to more TFB points in the vicinity of crossing or close modes than a large σ. But, as σ increases, one also notices that new TFB points appear, associated with the mode with an oscillating phase. Indeed, in that latter case, interferences arise where the IF of that

mode passes through an extremum. In the Results section, we are going to investigate more in detail the detection of TFB points in relation with the choice for σ.

B. Singular Points Detection in Spectrograms of Polyharmonic Signals

Another application of ERRP-RD is to localize *singular points* in polyharmonic signals, such as voice signals, that a priori do not contain any mode crossings. By singular points, we recall that we mean TF locations where two ERRPs merge (without being necessarily a TFB point). ERRP-RD enables to localize such points, the number of which varies with respect to σ, and we will see that to study this variation is of great practical interest.

VI. On the Relation Between Singularities of the Modulation Operator and TFBs

In this section, we propose to investigate the singularities of the modulation operator $\widehat{q_f}$, when TFBs are present, and in the simple situation of two interfering pure tones. Considering that the analysis window is the same as in the previous section (though a similar result could be obtained with any Gaussian window), one has $h'(t) = -\frac{2\pi}{\sigma^2} t h(t)$, and then the following

Proposition VI.1. *Let $f(t) = f_1(t) + f_2(t)$ with $f_1(t) = Ae^{i2\pi\xi_1 t}$ and $f_2(t) = e^{i2\pi\xi_2 t}$, where $\xi_1 < \xi_2$, the singularities of $\widehat{q_f}$ are the zeros of the denominator in Eq.* (5)*, namely the TF points (t, η) satisfying:*

$$V_f^h(t,\eta) V_f^{t^2 h}(t,\eta) - (V_f^{th}(t,\eta))^2 = 0. \tag{21}$$

Now, if we further assume that $\alpha := \sqrt{\frac{\pi}{2}}\sigma(\xi_2 - \xi_1) \leq 1$, the singularities of $\widehat{q_f}$ are located at $(t_{k,1}, \eta^*)$ or $(t_{k,2}, \eta^*)$, with

$$\eta^* = \frac{\xi_1 + \xi_2}{2} + \frac{\log(A)}{2\pi\sigma^2(\xi_2 - \xi_1)}$$

$$t_{k,1} = \frac{k - \frac{\arccos(-1+2\alpha^2)}{2\pi}}{\xi_2 - \xi_1} \text{ and } t_{k,2} = \frac{k + \frac{\arccos(-1+2\alpha^2)}{2\pi}}{\xi_2 - \xi_1}$$

with $k \in \mathbb{Z}$.

The proof is available in Appendix A. Then, it is interesting to analyze where these singularities are located with respect to the zeros and the local maxima of the spectrogram. As, for the signal f of Proposition VI.1, the spectrogram reads:

$$\begin{aligned} |V_f^h(t,\eta)|^2 =& A^2\hat{h}^2(\eta - \xi_1) + \hat{h}^2(\eta - \xi_2) \\ &+ 2A\hat{h}(\eta - \xi_1)\hat{h}(\eta - \xi_2)\cos(2\pi(\xi_2 - \xi_1)t), \end{aligned} \tag{22}$$

we can derive the following:

Proposition VI.2. *Considering the signal f of Proposition VI.1, the zeros of its spectrogram are located at $(\tilde{t}_k, \eta^*)$, with $\tilde{t}_k = \frac{k+1/2}{\xi_2 - \xi_1}$, $k \in \mathbb{Z}$, meaning they are aligned with the singularities of $\widehat{q_f}$.*

The proof is available in Appendix B. Now, investigating the position of the singularities of $\widehat{q_f}$ with respect to that of local maxima, we show the following:

Proposition VI.3. *Considering the signal f of Proposition VI.1, the zeros and the local maxima of its spectrogram are aligned only if $A = 1$.*

The proof is given in Appendix C. This means that the singularities of $\widehat{q_f}$ are never aligned with local maxima except if $A = 1$.

Now, investigating further the singularities of $\widehat{q_f}$ we show that they correspond to LMFs only when $A = 1$. For that purpose, we first show that at times $t_{k,i}$ the spectrogram has a unique maximum along the frequency axis, since we have the following

Proposition VI.4. *Assume α is defined as in Proposition VI.1 and define $\gamma = \cos(2\pi(\xi_2 - \xi_1)t)$, then $|V_f^h(t,\cdot)|^2$ has three local extrema if and only if $\alpha > \sqrt{\frac{1+\gamma}{2}}$ and*

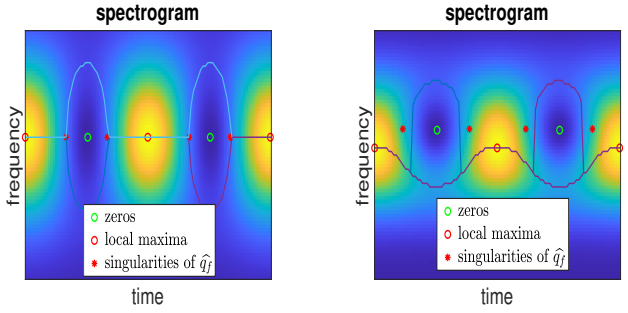
$$|\log(A)| < -\operatorname{arcosh}(X_2) + 2\alpha^2 \frac{\sqrt{X_2^2 - 1}}{X_2 + \gamma},$$

with $X_2 = \gamma(\alpha^2 - 1) + \alpha\sqrt{\gamma^2(\alpha^2 - 2) + 2}$. In any other case, $|V_f^h(t,\cdot)|^2$ has a unique extremum (which is a maximum).

The proof is available in Appendix D. In the case of Proposition VI.1, at $t = t_{k,i}$, $i = 1$ or 2, we have that $\gamma := \cos(2\pi(\xi_2 - \xi_1)t_{k,i}) = -1 + 2\alpha^2$ or equivalently $\alpha = \sqrt{\frac{1+\gamma}{2}}$, which means that, from Proposition VI.4, $|V_f^h(t_{k,i},\cdot)|^2$ has a single maximum whatever the value of A, which corresponds to a LMF. Then, we show the following:

Proposition VI.5. *Considering the signal of Proposition VI.1, then $(t_{k,i}, \eta^*)$ is not an LMF except if $A = 1$.*

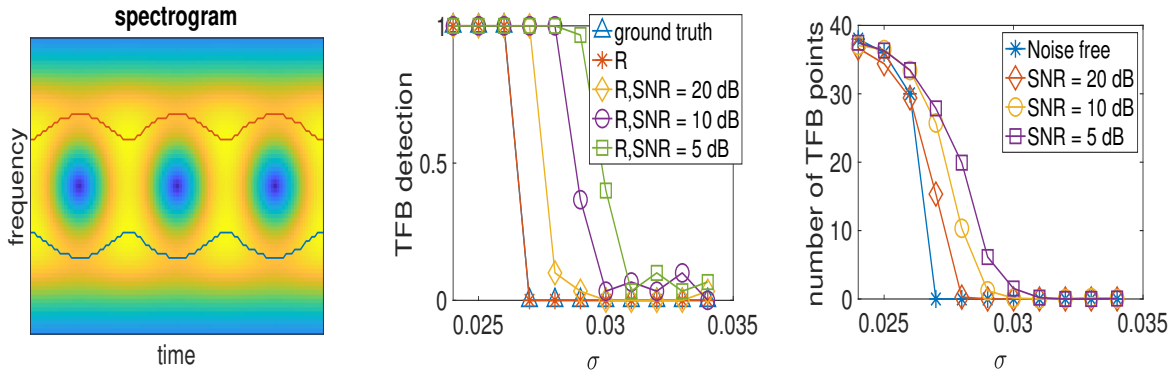
The proof is available in Appendix E.



(a) (b)

Fig. 3: (a): Spectrogram of two close pure tones where TFBs are present and when $A = 1$. The locations of the singularities of $\widehat{q_f}$, of the zeros and of the local maxima of the spectrogram are also displayed, as well as the ERRPs computed with the algorithm detailed in Sec. III; (b): Same as in (a), except that the amplitude of the high-frequency mode is larger than that of the low-frequency mode.

This proposition means, that when $A = 1$, $(t_{k,i}, \eta^*)$ is the unique LMF at time $t_{k,i}$, and when $A \neq 1$ the unique LMF at time $t_{k,i}$ is not a singularity of $\widehat{q_f}$. To illustrate this, we display, in Fig. 3 (a), the spectrogram of two pure tones with the same amplitude in a case such that $\sqrt{\frac{\pi}{2}}\sigma(\xi_2 - \xi_1) \leq 1$, meaning



(a) (b) (c)

Fig. 4: (a): spectrogram of the signal defined in Eq. (23), for $t \in [0.4, 0.5]$, the frequency range for the plot is $[210, 280]$, and $\sigma = 0.03$. The ridges associated with the two modes are also superimposed; (b): TFB detection for the signal defined in Eq. 23, using either R or the ground truth. The results are averaged over 30 noise realizations; (c): number of TFB points detected as a function of σ for different SNRs.

TFBs are present (ERRPs are computed using ERRP-RD introduced in the previous section). In that case, the singularities of the modulation operator are LMFs and correspond to the TF points where two ERRPs merge. On that figure, we also display the local maxima and zeros of the spectrogram. These are all aligned with the singularities of $\widehat{q_f}$. In Fig. 3 (b), we display the spectrogram of two pure tones exhibiting TFBs, and in which case the amplitude of f_2 is larger than that of f_1. We notice that, as expected, the singularities of $\widehat{q_f}$ are not LMFs, they are aligned with the zeros of the spectrogram, but not with its local maxima (ERRPs are computed using ERRP-RD introduced in the previous section). From this study, one can also estimate the time width of TFBs as the distance between two successive singularities surrounding a zero, as well as the distance separating the TFBs. Indeed, we have the following:

Proposition VI.6. *Considering the signal f of Proposition VI.1, assuming $\alpha \leq 1$, then the time interval between the end and the beginning of two successive TFBs has length $\frac{\arccos(-1+2\alpha^2)}{\pi(\xi_2-\xi_1)}$, and each TFB is almost centered around a zero with respect to time and its time width is $\frac{1-\frac{\arccos(-1+2\alpha^2)}{\pi}}{\xi_2-\xi_1}$.*

The proof is available in Appendix F. Analyzing the result of the above proposition, it transpires that when σ tends to 0, then α tends to 0 and thus the time width of the TFBs goes to 0, while when α tends to 1, the distance between the end and the beginning of two successive TFBs tends to 0 and its width to $\frac{1}{\xi_2-\xi_1}$.

VII. Results

To start with we shall remark that a repository enabling the reproduction of all the figures of the paper is available at https://github.com/meignen/interference_detection.

A. Validation of TFB Detector

Before we use the TFB detector based on TFB points proposed in Sec. V-A on complex examples, it is wise to check its accuracy on a simple example. For that purpose, consider the two-tone signal

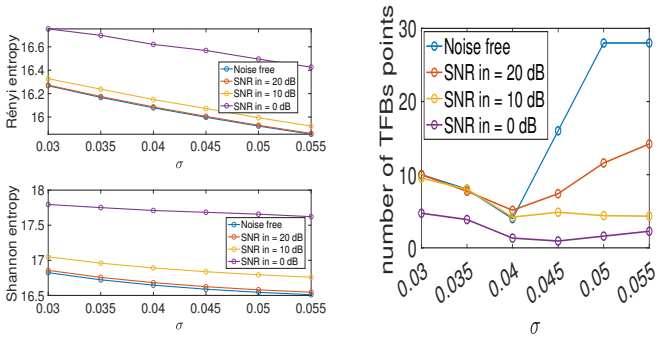
$$f(t) = e^{2i\pi 240t} + e^{2i\pi 260t}, \tag{23}$$

for $t \in [0, 1]$ whose spectrogram is displayed in Fig. 4 (a), for $t \in [0.4, 0.5]$. In such a case, we know from Proposition IV.1 that TFBs are present when $\alpha := \sqrt{\frac{\pi}{2}}\sigma(\xi_2 - \xi_1) \leq 1$ and not otherwise. As ξ_1 and ξ_2 are fixed parameters, this defines a function of σ equal to 1 when a TFB is present and to 0 otherwise. This function is denoted by *ground truth* (GT) hereafter. Then, we define a function R such that, $R(\sigma)$ equals 1 if at least one TFB is detected by means of the algorithm described in Sec. V-A, and zero otherwise. So R should be close to GT if it is relevant to detect the presence of TFBs. To check this, we display, in Fig. 4 (b), R along with GT, in noiseless and noisy cases (ERRP-RD and ERRP-RD$_{noise}$ being used as RD respectively). We notice that R equals GT in the noiseless case, which is precisely what we expected. We also remark that the spectrograms of noisy interfering modes may contain TFB points when there are none in the noiseless case. To investigate the stability of the detected TFBs, we also compute in Fig. 4 (c) the number of TFB points as a function of σ for different noise levels. We notice that most TFBs detected in the noiseless case are still detected when noise is added (behavior for $\sigma \leq 0.026$), while for σ larger than 0.03, no TFBs are detected, and this remains the case when noise is added. This study also highlights the fact that in the range between these two values of σ, the noise can create TFBs that are not present in the noiseless case.

B. Detecting Mode Crossings with TFB points

In this section, we investigate in what way the knowledge of TFB points can help us determine whether there are some mode crossings in the TF plane, and find an appropriate window length parameter σ. To do this, let us consider the signal of Fig. 2, for which we determine the number of TFB points when both σ and the input SNR varies.

The results depicted in Fig. 5 (b) tell us that, in the noiseless case, the number of TFB points attains a minimum when $\sigma = 0.04$, and equals 4 (see Fig. 2 (b)). We remark that these TFB points correspond to the location of the crossing of the two modes (see Fig. 2 (b)) and that they are stable as the noise level increases, provided the latter is not too strong, that is the SNR is larger than 0 dB. When σ is smaller than this value, the crossing of the two modes as well as the interference between the oscillatory modes and the linear chirp generate more TFB points (see Fig. 2 (a)), which are stable under the addition of a reasonable amount of noise (SNR larger than 0 dB). For these values of σ and when the SNR equals 0 dB, we notice that the domain $\mathcal{S}(3)$ does not contain, for some noise realizations, the ridge portions involved in the definition of the TFBs at a lower noise level, therefore fewer TFB points are detected in that case. On the contrary, when σ is larger than 0.04, many TFB points arise due to interference in the oscillatory mode (see Fig. 2 (c)), but these mostly disappear when some noise is added. We shall also remark that as the number of TFB points is not null for all the tested σ, and, by considering the regions where TFB points are present regardless of σ, one can localize mode crossings. Furthermore, we get from this study that to choose the value of σ that minimizes the number of detected TFB points leads to an appropriate window length, because it minimizes the number of locations where mode mixing arise, and thus corresponds to the value that best separates the modes.



(a) (b)

Fig. 5: (a): Average Rényi and Shannon entropies computed on the spectrogram corresponding to the signal of Fig. 2 for different values of σ and input SNRs; (b): Number of TFB points with respect to σ and noise level still for the signal of Fig. 2. The results are averaged over 30 noise realizations.

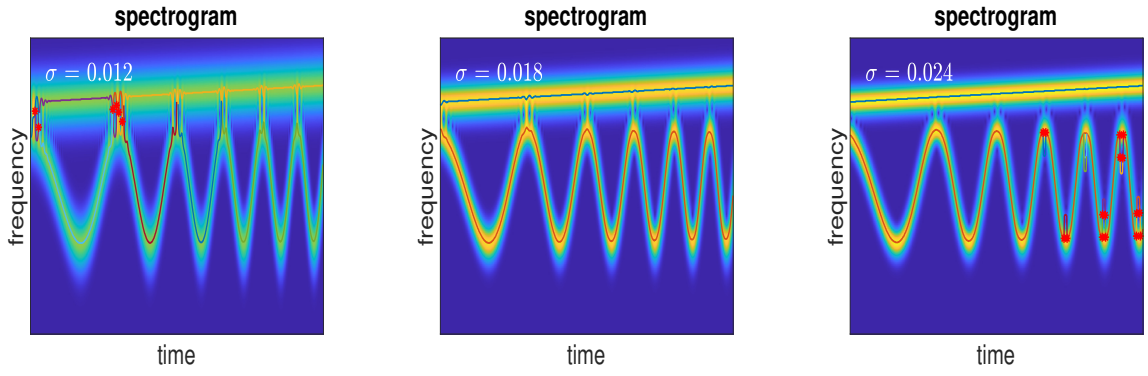
In most applications, the window length parameter σ is however found by considering the value minimizing the Rnyi entropy over a certain range of the length parameter [22]. The main motivation for such an approach is that the Rnyi entropy is an overall measure of the dispersion of the information in the TF plane [23]. Also, when considering a linear chirp, the window length associated with the minimal Rnyi entropy automatically adapts to the chirp rate of the signal [24]. However, when the signal is made of modes with different frequency modulations, what matters most is not the dispersion but the separability. Furthermore, it is not guaranteed that the Rnyi entropy actually passes through a minimum in the range of interest for σ. In order to check the difference between separability and dispersion, we also compute the Rnyi entropy, for the signal depicted in Fig. 2, under the same conditions as for TFB points. The results depicted in Fig. 5 (a) reveals that the parameter value leading to the minimal Rnyi entropy over the studied range for σ is irrelevant in that case. Note finally that to consider other types of entropies, like the Shannon entropy, leads to the same conclusion.

C. Finding the Best Window Length Based on TFBs Detection in the Case of Interfering Modes

In this section, our goal is first to show that to choose the value for σ that minimizes the number of TFB points is a good strategy for mode separation also in situations where there are no mode crossings.

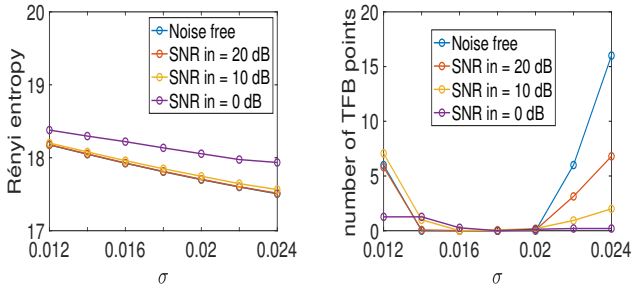
To illustrate the impact of the value of σ on a signal containing close but not intersecting modes, we display in Fig. 6, the spectrogram, for different σ, of the sum of a sinusoidal mode and a linear chirp. In each case, we also plot the ERRPs computed with ERRP-RD along with the TFB points. We notice that a small value of σ results in TFB points associated with strong interferences between the two modes, while for intermediate values of σ, no TFB points are detected, and larger values of σ generate TFB points associated with self-interferences in the oscillatory mode. For such a signal, $\sigma = 0.018$ seem to ensure the separation of the two modes while avoiding interference in the oscillatory mode. To investigate this more in depth, we study as previously the number of TFB points with varying σ and input SNR, the results being depicted in Fig. 7 (b). We notice that, when σ is small, there are many TFB points related to inferences between the two modes and that these are stable under the addition of a reasonable amount of noise. For intermediate values of σ the number of TFB points is null and this remains true for all the tested noise levels, while for large values of σ, interference in the oscillatory mode creates many TFB points, which are unstable through noise addition. From this study, we conclude that the values of σ that correspond to the absence of TFB points are good choices to well separate the modes and avoid interference in the oscillatory mode. The fact that the minimal number of TFB points is zero also means there are no ridge crossings. Though it is not the purpose of the present paper, it is worth also noting here that such values of σ lead to two separated ridges associated with the two modes making up the signal, which is a crucial point if one wants to perform mode reconstruction from ridge detection.

To compare with, we also plot the Rnyi entropies of the spectrogram corresponding to the same noise level and values of σ, and we notice that it keeps decreasing when σ increases. From this, one can draw the conclusion that to use the minimal Rnyi entropy in the relevant range for σ is not a good strategy. To consider other entropies, like the Shannon entropy, would lead to the same results, which are therefore not displayed here.



(a) (b) (c)

Fig. 6: Spectrogram of a sum of a sinusoidal chirp and a linear chirp for different values of σ, along with the detected ERRPs (ERRP-RD is used as RD), the red stars corresponding to TFB points.



(a) (b)

Fig. 7: (a): Average Rényi entropy computed on the spectrogram corresponding to the signal of Fig. 6; (b): Number of TFB points. The results are averaged over 30 noise realizations.

D. Singular Points Detection: Application to Voice Signals

In this section, our goal is to give an illustration on a voice signal of the potential interest of the procedure we propose in Sec. V-B to detect singular points to determine an appropriate window length to analyze such a signal. Indeed, for that kind of signals, it is of interest to have access to its different harmonics, which can be done by extracting the ERRPs. A natural question one then asks is to what extent the extracted ERRPs are related to the harmonics. We are going to show that, by finding the value of σ that minimizes the number of detected singular points, we obtain a satisfactory answer to that question.

Let us consider the voice signal of a 22-year-old female taken from the recording 62 of the Saarbruken Voice Database [25]. This signal contains 126902 samples for a length of 2.538 seconds and for efficiency purpose, we downsample it by a factor of 10. It contains a sustained /a/ vowel of the type "low-high-low" meaning there is a change on the pitch of the signal. The spectrograms of such a signal for different values of σ are depicted on the first column of Fig. 8, along with the detected ERRPs, the singular points corresponding to red stars. We notice that when $\sigma = 0.007$, ERRP-RD detects 6 uninterrupted and meaningful ridges, which is not the case for the other displayed cases, in which the harmonics interfere, creating singular points. So, we check that this value of σ actually corresponds to the detection of a minimal number of singular points even in noisy situations. To do so, we compute the number of singular points as a function of σ in Fig. 9 (b), and for different input SNRs. We notice that when this number equals 0 in this absence of noise, it corresponds to cases where the harmonics are well separated, and that, in such cases, the number of singular points remains small in noisy situations. Finally, it is important to remark that the sharp increase in the number of singular points when σ is small is due to interference appearing between the modes where these are not modulated.

To compare with, the values of the Rényi entropy as a function of σ are shown in Fig. 9 (a), and, as expected, the value of σ corresponding to the minimum of the Rényi entropy over the tested range is not associated with the best result in terms of mode separability. We are now interested in reassigning the STFT of this signal using *second order synchrosqueezing transform* (FSST2) [17], [26] (see in that paper for the details of the technique), to obtain a sharper TFR. The key point for FSST2 to be successful is that the STFT of the signal can be locally accurately approximated by that of a linear chirp. However, when the harmonics are interfering due to a wrong window choice, some interferences appear in the spectrogram, causing the failure of FSST2. We investigate visually the effect of a wrong choice for σ on the spectrogram and FSST2 in an area of interference. Zoomed in versions of the spectrogram and FSST2 around the interference area are shown in the second and third columns of Fig. 8. In that case, we see that, when no singular points are detected, as in the case $\sigma = 0.007$, FSST2 leads to a much sharper representation in the area of interference. On the contrary, when $\sigma = 0.003$ (first row of Fig. 8), interference along the frequency axis creates undesirable oscillations on the synchrosqueezed representation, while for $\sigma = 0.013$ (third row of Fig. 8), interference along the time axis are present, creating spurious horizontal structures that make ridge detection difficult, and therefore mode separation. This simple study confirms that what monitors the quality of FSST2 locally is whether singular points are present in the TF area of interest.

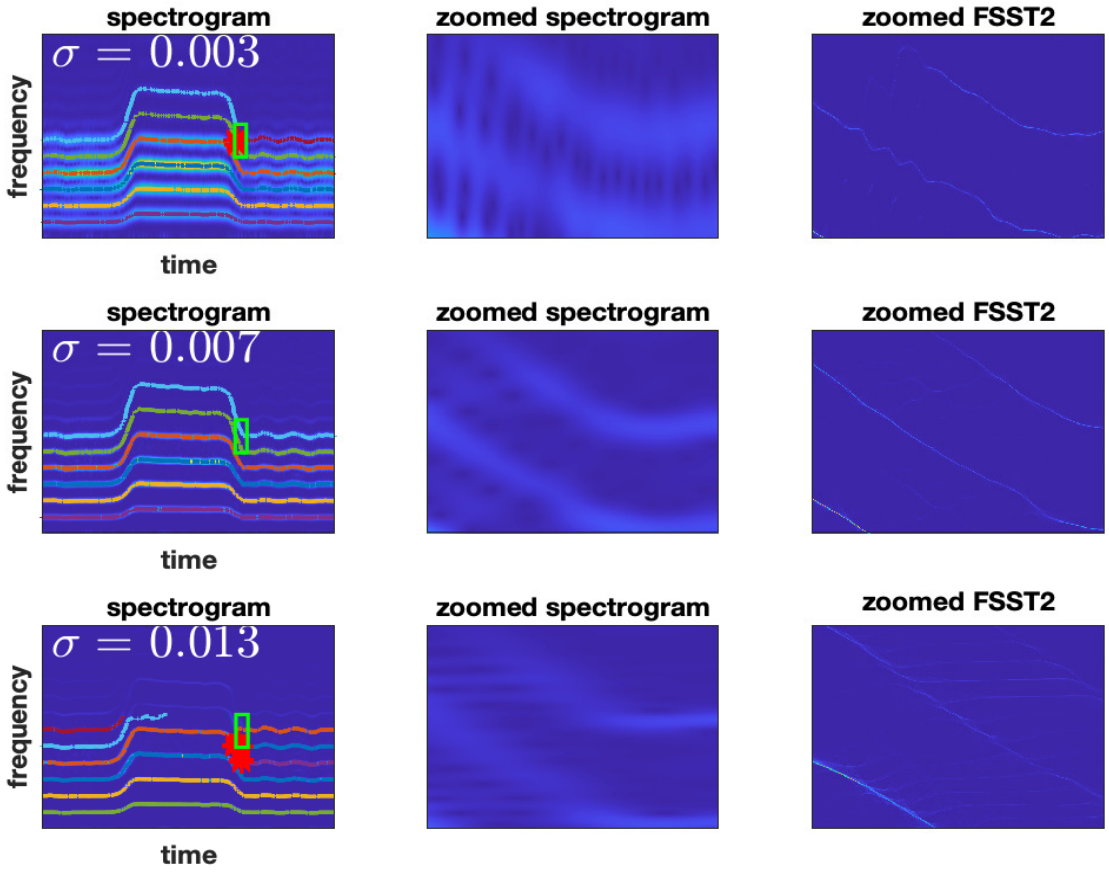
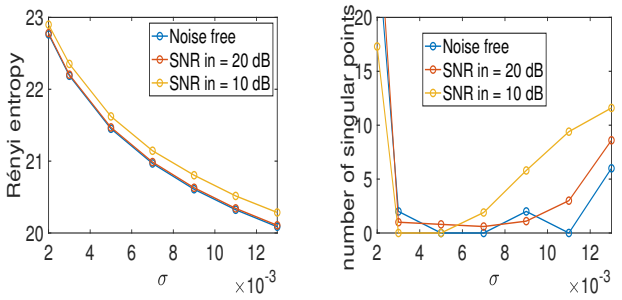


Fig. 8: Left column: spectrogram of a voice signal (recording 62 of the Saarbruken Voice Database [25]) along with the ERRPs associated with different values of σ; Middle and right columns : zoomed in versions of the spectrogram and FSST2 corresponding to the green rectangular regions of the spectrogram on the left column.



(a) (b)

Fig. 9: (a): Rnyi entropies corresponding to the signal of Fig. 8, when σ varies; (b): Number of singular points detected using ERRP-RD or ERRP-RD$_{noise}$ depending on whether some noise is present. The results are averaged over 10 noise realizations.

VIII. Conclusion

In this paper, our goal was to introduce a new ridge detector to enable the detection of strong interference in the time-frequency plane. We found our analysis on some existing adaptive ridge detectors that used a modulation operator to construct the ridges. We explained how the intersection of the different ridges could help us localize some specific structures called time-frequency bubbles, which proved to be useful to detect crossing modes and where two modes are strongly interfering in the TF plane. Application to voice signals on how to use singular points to find an appropriate window length in the definition of the spectrogram was also presented. In a near future, such a window determination technique will be applied to find locally the best window length and improve the behavior of synchrosqueezing transforms.

Appendix

A. Proof of Proposition VI.1

Let us write:

$$\widehat{q_f}(t,\eta) = -\frac{1}{2\pi}\Im\left\{\frac{(V_f^h(t,\eta))^2}{V_f^{t^2h}(t,\eta)V_f^h(t,\eta) - (V_f^{th}(t,\eta))^2}\right\} = -\frac{1}{2\pi}\Im\left\{\frac{N(t,\eta)}{D(t,\eta)}\right\}.$$

For the signal f of the proposition, one has:

$$\begin{aligned}D(t,\eta) = \frac{\sigma^2}{2\pi}\Big(&f_1^2(t)\hat{h}^2(\eta-\xi_1) + f_2^2(t)\hat{h}^2(\eta-\xi_2)\\ &+2f_1(t)f_2(t)\hat{h}(\eta-\xi_1)\hat{h}(\eta-\xi_2)\Big)\\ &-\sigma^4(\xi_2-\xi_1)^2 f_1(t)f_2(t)\hat{h}(\eta-\xi_1)\hat{h}(\eta-\xi_2)\\ = \frac{\sigma^2}{2\pi}N(t,\eta) &- \sigma^4(\xi_2-\xi_1)^2 f_1(t)f_2(t)\hat{h}(\eta-\xi_1)\hat{h}(\eta-\xi_2)).\end{aligned}$$

Then, $N(t,\eta)$ and $D(t,\eta)$ have no common zeros, since $N(t,\eta) = 0$ implies that $D(t,\eta) \neq 0$, since $\xi_1 \neq \xi_2$. Therefore, the singularity of $\widehat{q_f}$, located at (t^*,η^*) corresponds to $D(t^*,\eta^*) = 0$.

Then, we may write

$$\begin{aligned}&D(t,\eta) = 0\\ \Leftrightarrow\ &A^2 e^{i4\pi\xi_1 t}\hat{h}^2(\eta-\xi_1) + e^{i4\pi\xi_2 t}\hat{h}^2(\eta-\xi_2)\\ &+2(1-\pi\sigma^2(\xi_2-\xi_1)^2)Ae^{i2\pi(\xi_1+\xi_2)t}\hat{h}(\eta-\xi_1)\hat{h}(\eta-\xi_2) = 0\\ \Leftrightarrow\ &A^2\hat{h}^2(\eta-\xi_1)e^{i2\pi(\xi_1-\xi_2)t} + \hat{h}^2(\eta-\xi_2)e^{i2\pi(\xi_2-\xi_1)t}\\ &+2(1-\pi\sigma^2(\xi_2-\xi_1)^2)A\hat{h}(\eta-\xi_1)\hat{h}(\eta-\xi_2) = 0\end{aligned}$$

A necessary condition for the equality to be true is that $A^2\hat{h}^2(\eta-\xi_1)e^{i2\pi(\xi_1-\xi_2)t} + \hat{h}^2(\eta-\xi_2)e^{i2\pi(\xi_2-\xi_1)t}$ belongs to $\mathbb{R}$, which is true if and only if $A\hat{h}(\eta-\xi_1) = \hat{h}(\eta-\xi_2)$. Due to the choice for h, it is easy to check that this equation in η has a unique solution corresponding to:

$$\eta^* = \frac{\xi_1+\xi_2}{2} + \frac{\log(A)}{2\pi\sigma^2(\xi_2-\xi_1)}.$$

For $\eta = \eta^*$, the zeros of $D(\eta^*,t)$ would be such that:

$$2e^{i2\pi(\xi_1+\xi_2)t}(\cos(2\pi(\xi_2-\xi_1)t) + 1 - \pi\sigma^2(\xi_2-\xi_1)^2) = 0$$

meaning $\cos(2\pi(\xi_2-\xi_1)t) = -1+\pi\sigma^2(\xi_2-\xi_1)^2$. Since we further assume that $0 < \sqrt{\frac{\pi}{2}}\sigma(\xi_2-\xi_1) \leq 1$, the solutions to this equation correspond to:

$$\begin{aligned}t_{k,1} &= \frac{k - \frac{\arccos(-1+\pi\sigma^2(\xi_2-\xi_1)^2)}{2\pi}}{\xi_2-\xi_1}\\ t_{k,2} &= \frac{k + \frac{\arccos(-1+\pi\sigma^2(\xi_2-\xi_1)^2)}{2\pi}}{\xi_2-\xi_1}\end{aligned}$$

for $k \in \mathbb{Z}$.

B. Proof of Proposition VI.2

The function $|V_f^h(\cdot,\eta)|^2$, where '$\cdot$' means we consider this variable, attains its minimum at $\tilde{t}_k = \frac{k+1/2}{\xi_2-\xi_1}$, at which:

$$|V_f^h(\tilde{t}_k,\eta)|^2 = (A\hat{h}(\eta-\xi_1) - \hat{h}(\eta-\xi_2))^2.$$

Then, the zeros of the spectrogram correspond to frequencies η such that $A\hat{h}(\eta-\xi_1) = \hat{h}(\eta-\xi_2)$, meaning $\eta = \eta^*$.

C. Proof of Proposition VI.3

$|V_f^h(\cdot,\eta)|^2$ attains its maximum at $t_k = \frac{k}{\xi_2-\xi_1}, k \in \mathbb{Z}$ at which:

$$|V_f^h(t_k,\eta)|^2 = \sigma^2(Ae^{-\pi\sigma^2(\eta-\xi_1)^2} + e^{-\pi\sigma^2(\eta-\xi_2)^2})^2.$$

Now, local maxima of the spectrogram correspond to frequencies η such that $|V_f^h(t_k,\eta)|^2$ passes through a local maximum. Computing the derivative of $|V_f^h(t_k,\eta)|^2$, one obtains:

$$\begin{aligned}-2\pi(&A(\eta-\xi_1)\hat{h}(\eta-\xi_1) + (\eta-\xi_2)\hat{h}(\eta-\xi_2))\\ &(A\hat{h}(\eta-\xi_1) + \hat{h}(\eta-\xi_2)).\end{aligned}$$

The second term never vanishes and, for the first term, at $\eta = \eta^*$, one has:

$$\begin{aligned}&A(\eta^*-\xi_1)\hat{h}(\eta^*-\xi_1) + (\eta^*-\xi_2)\hat{h}(\eta^*-\xi_2)\\ &\quad= [(\eta^*-\xi_1) + (\eta^*-\xi_2)]\,\hat{h}(\eta^*-\xi_2),\end{aligned}$$

which is null only if $A = 1$, in which case $\eta^* = \frac{\xi_1+\xi_2}{2}$.

D. Proof of Proposition VI.4

Let us consider the function:

$$\begin{aligned}l(\eta) = &A^2 e^{-2\pi\sigma^2(\eta-\xi_1)^2} + e^{-2\pi\sigma^2(\eta-\xi_2)^2}\\ &+2Ae^{-\pi\sigma^2[(\eta-\xi_1)^2+(\eta-\xi_2)^2]}\cos(2\pi(\xi_2-\xi_1)t),\end{aligned}$$

and then put $\gamma = \cos(2\pi(\xi_2-\xi_1)t)$. Making the change of variables $\eta = \nu + \frac{\xi_1+\xi_2}{2}$ and putting $\xi = \frac{\xi_2-\xi_1}{2}$, one may define:

$$\begin{aligned}f_1(\nu) =&l(\nu + \frac{\xi_1+\xi_2}{2})\\ =&e^{-2\pi\sigma^2(\nu^2+\xi^2)}(A^2e^{-4\pi\sigma^2\xi\nu} + e^{4\pi\sigma^2\xi\nu} + 2A\gamma).\end{aligned}$$

Putting $\nu = \frac{\log(A)+\mu}{4\pi\sigma^2\xi}$, enables us to define:

$$\begin{aligned}f_1(\nu) =&f_2(\mu)\\ =&e^{-2\pi\sigma^2\left[\left(\frac{\log(A)+\mu}{4\pi\sigma^2\xi}\right)^2+\xi^2\right]}2A(\cosh(\mu)+\gamma).\end{aligned}$$

The derivative of f_2 reads:

$$\begin{aligned}f_2'(\mu) =&e^{-2\pi\sigma^2\left[\left(\frac{\log(A)+\mu}{4\pi\sigma^2\xi}\right)^2+\xi^2\right]}\\ &2A\left[-\frac{\log(A)+\mu}{4\pi\sigma^2\xi^2}(\cosh(\mu)+\gamma) + \sinh(\mu)\right]\end{aligned}$$

which has the sign of (assuming $\gamma \neq -1$):

$$g(\mu) \quad = \quad -\frac{\log(A)+\mu}{2\alpha^2} + \frac{\sinh(\mu)}{\cosh(\mu)+\gamma}.$$

Differentiating g we get

$$g'(\mu) = \quad -\frac{1}{2\alpha^2} + \frac{1+\gamma\cosh(\mu)}{(\cosh(\mu)+\gamma)^2}$$

which has the same sign as

$$\begin{aligned}h(\cosh(\mu)) &= -\frac{(\cosh(\mu)+\gamma)^2}{2\alpha^2} + 1 + \gamma\cosh(\mu)\\ = \frac{1}{\alpha^2}&\left(-\frac{\cosh(\mu)^2}{2} + \gamma(\alpha^2-1)\cosh(\mu) + \alpha^2 - \frac{\gamma^2}{2}\right).\end{aligned}$$

The factor inside the parentheses is a second order polynomial in $\cosh(\mu)$ whose discriminant reads: $\Delta = \alpha^2(\gamma^2(\alpha^2-2)+2) > 0$. The roots of this polynomial are denoted by X_1 and X_2 with $X_1 < X_2$, and $h(X) < 0$ if $X \in]-\infty, X_1[\bigcup]X_2, +\infty[$ and $h(X) > 0$ if $X \in]X_1, X_2[$. To know the location of $\cosh(\mu)$ with respect to X_1 and X_2, one

first compute $h(1) = (\gamma+1)(1-\frac{\gamma+1}{2\alpha^2})$ which has the same sign as $\alpha^2 - \frac{\gamma+1}{2}$. Assuming $\alpha \leq \sqrt{\frac{\gamma+1}{2}}$, $h(1) \leq 0$ and 1 belongs to $]-\infty, X_1]$ or $[X_2, \infty[$. As $\frac{X_1+X_2}{2} = \gamma(\alpha^2-1) < 1$, 1 belongs to $]X_2, +\infty[$. Finally, as $\cosh(\mu) \geq 1$, $h(\cosh(\mu))$ and thus $g'(\mu)$ are negative. From this study one deduces the following table of variations:

μ	$-\infty$		μ_1		$+\infty$
$g(\mu)$	$+\infty$	$\searrow$		$\searrow$	$-\infty$
$f_2'(\mu)$		$+$	0	$-$	
$f_2(\mu)$		$\nearrow$		$\searrow$	

and $l(\eta)$ has a unique extremum which is a maximum when $\alpha \leq \sqrt{\frac{\gamma+1}{2}}$.

If one now assumes that $\alpha > \sqrt{\frac{\gamma+1}{2}}$, then $h(1) > 0$ and thus 1 belongs to $]X_1, X_2[$. As $\cosh(\mu) \in]X_1, X_2[$ is equivalent to $|\mu| < \operatorname{arcosh}(X_2)$, $h(\cosh(\mu)) < 0$ if

$$\mu \in]-\infty, -\operatorname{arcosh}(X_2)[\bigcup] \operatorname{arcosh}(X_2), +\infty[,$$

and $h(\cosh(\mu))$ is positive otherwise, leading to the following table of variations:

μ	$-\infty$		$-\operatorname{arcosh}(X_2)$		$\operatorname{arcosh}(X_2)$		$+\infty$
$g'(\mu)$		$-$	0	$+$	0	$-$	
$g(\mu)$	$+\infty$	$\searrow$		$\nearrow$		$\searrow$	$-\infty$

$g(\mu)$ vanishes with a change of sign only once at some $\mu = \mu_1$ if and only if $g(-\operatorname{arcosh}(X_2)) \geq 0$ or $g(\operatorname{arcosh}(X_2)) \leq 0$. In this case, we deduce that

μ	$-\infty$		μ_1		$+\infty$
$f_2'(\mu)$		$+$	0	$-$	
$f_2(\mu)$		$\nearrow$		$\searrow$	

meaning $l(\eta)$ has a unique extremum (which is a maximum).

If $g(-\operatorname{arcosh}(X_2)) < 0$ and $g(\operatorname{arcosh}(X_2)) > 0$, $g(\mu)$ vanishes and changes signs three times at some $\mu = \mu_1$, μ_2 and μ_3 leading to the following table of variations for f_2:

μ	$-\infty$		μ_1		μ_2		μ_3		$+\infty$
$g(\mu)$		$+$	0	$-$	0	$+$	0	$-$	
$f_2(\mu)$		$\nearrow$		$\searrow$		$\nearrow$		$\searrow$	

In such a case, $l(\eta)$ has three extrema: 2 maxima and a minimum.

Finally, to specify in which situations l has three extrema, we remark that

$$\begin{aligned} & g(\operatorname{arcosh}(X_2)) > 0 \\ \Leftrightarrow & \log(A) < -\operatorname{arcosh}(X_2) + 2\alpha^2 \frac{\sqrt{X_2^2-1}}{X_2+\gamma} \\ & g(-\operatorname{arcosh}(X_2) < 0 \\ \Leftrightarrow & -\log(A) < -\operatorname{arcosh}(X_2) + 2\alpha^2 \frac{\sqrt{X_2^2-1}}{X_2+\gamma}. \end{aligned}$$

So $l(\eta)$ has three extrema if and only if $\alpha > \sqrt{\frac{1+\gamma}{2}}$ and then if $|\log(A)| < -\operatorname{arcosh}(X_2)) + 2\alpha^2 \frac{\sqrt{X_2^2-1}}{X_2+\alpha}$ with $X_2 = \gamma(\alpha^2-1) + \alpha\sqrt{\gamma^2(\alpha^2-2)+2}$.

E. Proof of Proposition VI.5

Let us consider $\gamma = \cos(2\pi(\xi_2-\xi_1)t_{k,i}) = -1 + 2\alpha^2$. We shall then remark that

$$\begin{aligned} & \partial_\eta |V_f^h(t_{k,i}, \eta)|^2 = \\ & -4\pi\sigma^2 \Big[(\eta-\xi_1) A^2 \hat{h}^2(\eta-\xi_1) + (\eta-\xi_2)\hat{h}^2(\eta-\xi_2) \\ & \qquad + \gamma(2\eta-\xi_1-\xi_2) A \hat{h}(\eta-\xi_1)\hat{h}(\eta-\xi_2) \Big] \end{aligned}$$

Evaluating this quantity at $\eta = \eta^*$, and since $A\hat{h}(\eta^*-\xi_1) = \hat{h}(\eta^*-\xi_2)$, we get:

$$\begin{aligned} & \partial_\eta |V_f^h(t_{k,i}, \eta^*)|^2 \\ & = -4\pi\sigma^2 (2\eta^* - \xi_1 - \xi_2)(1+\gamma) A^2 \hat{h}^2(\eta^*-\xi_1) \\ & = -\frac{8\log(A)}{\xi_2-\xi_1} \alpha^2 A^2 \hat{h}^2(\eta^*-\xi_1) \end{aligned}$$

which is null if $A = 1$ and non-zero otherwise.

F. Proof of Proposition VI.6

Let us recall that the zeros are located at time $\tilde{t}_k = \frac{k+1/2}{\xi_2-\xi_1}$, and that the local maxima are located at $t_k = \frac{k}{\xi_2-\xi_1}$, for $k \in \mathbb{Z}$. Since $0 < \sqrt{\frac{\pi}{2}}\sigma(\xi_2-\xi_1) \leq 1$, we have $\pi > \arccos(K) \geq 0$, then we have $t_{k,1} \leq t_k \leq t_{k,2}$, which means that the distance between the end and the beginning of two successive TFBs is

$$t_{k,2} - t_{k,1} = \frac{\arccos(-1+2\alpha^2)}{\pi(\xi_2-\xi_1)}.$$

Thus, the width is $t_{k+1,1} - t_{k,2} = \frac{1-\frac{\arccos(-1+2\alpha^2)}{\pi}}{\xi_2-\xi_1}$.

REFERENCES

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

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

[3] L. Stankovic, M. Dakovic, and V. Ivanovic, "Performance of spectrogram as IF estimator," *Electronics Letters*, vol. 37, no. 12, pp. 797–799, 2001.

[4] R. Carmona, W. Hwang, and B. Torresani, "Multiridge detection and time-frequency reconstruction," *IEEE Transactions on Signal Processing*, vol. 47, no. 2, pp. 480–492, Feb 1999.

[5] L. Rankine, M. Mesbah, and B. Boashash, "If estimation for multicomponent signals using image processing techniques in the time–frequency domain," *Signal Processing*, vol. 87, no. 6, pp. 1234–1250, 2007.

[6] B. Barkat and K. Abed-Meraim, "Algorithms for blind components separation and extraction from the time-frequency distribution of their mixture," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, no. 13, pp. 1–9, 2004.

[7] K. Abed-Meraim, N. Linh-Trung, V. Sucic, F. Tupin, and B. Boashash, "An image processing approach for underdetermined blind separation of nonstationary sources," in *3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the*, vol. 1. IEEE, 2003, pp. 347–352.

[8] M. A. Colominas, S. Meignen, and D.-H. Pham, "Fully adaptive ridge detection based on stft phase information," *IEEE Signal Processing Letters*, vol. 27, pp. 620–624, 2020.

[9] N. Laurent and S. Meignen, "A novel ridge detector for nonstationary multicomponent signals: development and application to robust mode retrieval," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3325–3336, 2021.

[10] X. Zhu, H. Yang, Z. Zhang, J. Gao, and N. Liu, "Frequency-chirprate reassignment," *Digital Signal Processing*, vol. 104, p. 102783, 2020.

[11] L. Li, H. Cai, H. Han, Q. Jiang, and H. Ji, "Adaptive short-time Fourier transform and synchrosqueezing transform for non-stationary signal separation," *Signal Processing*, vol. 166, p. 107231, 2020.

[12] L. Li, C. K. Chui, and Q. Jiang, "Direct signal separation via extraction of local frequencies with adaptive time-varying parameters," *IEEE Transactions on Signal Processing*, vol. 70, pp. 2321–2333, 2022.

[13] J. Shi, G. Chen, Y. Zhao, and R. Tao, "Synchrosqueezed fractional wavelet transform: A new high-resolution time-frequency representation," *IEEE Transactions on Signal Processing*, vol. 71, pp. 264–278, 2023.

[14] Z. Zhao and G. Li, "Synchrosqueezing-based short-time fractional fourier transform," *IEEE Transactions on Signal Processing*, vol. 71, pp. 279–294, 2023.

[15] N. 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*, vol. 5, no. 1, pp. 64–71, 1997.

[16] S. Meignen, N. Laurent, and T. Oberlin, "One or two ridges? An exact mode separation condition for the Gabor transform," *IEEE Signal Processing Letters*, vol. 29, pp. 2507–2511, 2022.

[17] R. Behera, S. Meignen, and T. Oberlin, "Theoretical analysis of the second-order synchrosqueezing transform," *Applied and Computational Harmonic Analysis*, vol. 45, no. 2, pp. 379–404, 2018.

[18] P. Flandrin, "Time–frequency filtering based on spectrogram zeros," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2137–2141, 2015.

[19] D.-H. Pham and S. Meignen, "A novel thresholding technique for the denoising of multicomponent signals," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 4004–4008.

[20] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[21] R. Bardenet, J. Flamant, and P. Chainais, "On the zeros of the spectrogram of white noise," *Applied and Computational Harmonic Analysis*, vol. 48, no. 2, pp. 682–705, 2020.

[22] S. Meignen and D.-H. Pham, "Retrieval of the modes of multicomponent signals from downsampled short-time Fourier transform," *IEEE Transactions on Signal Processing*, vol. 66, no. 23, pp. 6204–6215, 2018.

[23] R. G. Baraniuk, P. Flandrin, A. J. Janssen, and O. J. Michel, "Measuring time-frequency information content using the Rényi entropies," *IEEE Transactions on Information theory*, vol. 47, no. 4, pp. 1391–1409, 2001.

[24] S. Meignen, M. Colominas, and D.-H. Pham, "On the use of Rényi entropy for optimal window size computation in the short-time Fourier transform," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5830–5834.

[25] B. Woldert-Jokisz. (2007) Saarbruecken voice database. [Online]. Available: http://stimmdb.coli.uni-saarland.de

[26] T. Oberlin, S. Meignen, and V. Perrier, "Second-order synchrosqueezing transform or invertible reassignment? Towards ideal time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1335–1344, March 2015.

Dr. Sylvain Meignen received his Ph.D. degree in applied mathematics in 2001, and its "Habilitation à Diriger Des Recherches" in 2011, both from the University of Grenoble, France. Since 2002, he has been an Associate Professor with Grenoble INP. His research interests include nonlinear multiscale image and signal processing, time-frequency analysis (empirical mode decomposition, synchrosqueezing) and approximation theory. In 2010, he was a visitor at the GIPSA-Lab Grenoble, at the IDCOM of the university of Edinburgh, U.K., in 2011 and at the university of Pisa, Italy, in 2018.

Marcelo A. Colominas received the Bioengineering degree from the Universidad Nacional de Entre Ríos (UNER, Argentina) in 2011, and the Ph.D. degree in Engineering from the Universidad Nacional del Litoral (UNL, Argentina) in 2016. He was a postdoctoral fellow with the Université d'Angers (France) in 2017 and 2018. He joined the Department of Mathematics, Faculty of Engineering UNER, in 2008, became a Teaching Assistant in 2011, and an Adjunct Professor in 2019. Since 2019, he is an Assistant Researcher with the Consejo Nacional de Investigaciones Cientificas y Técnicas (CONICET, Argentina). His research interests are time-frequency/time-scale signal analysis, synchrosqueezing, data-driven methods and biomedical signal processing.