Novel Instantaneous Frequency Estimator in Multicomponent Signals with Interfering or Crossing Modes

S. Meignen, and K. Polisano

***Abstract*—In this paper, we introduce a novel estimator for the instantaneous frequencies of the modes making up multicomponent signals, particularly when they are close or crossing in the time-frequency plane. Our approach is based on locally optimizing the spectrogram window in the time-frequency domain to enhance signal representation, depending on whether the modes are isolated or interfering. To achieve this, we employ an adaptive criterion: for isolated modes, the window is selected to minimize time dependency, while for close or interfering modes, it is chosen to reduce interference. By integrating this adaptive time-frequency representation with ridge detection, and then by using spline fitting, we obtain highly robust instantaneous frequency estimates. Our proposed method demonstrates strong performance in handling MCSs with close or crossing modes, and is shown to outperform state-of-the-art instantaneous frequency estimation techniques, including methods based on synchrosqueezing and the chirplet transform.**

***Index Terms*—Time-frequency, multicomponent signal, interference, spline approximation.**

I. Introduction

NON-stationary signals such as audio (music, speech, bird songs) [1], electrocardiogram [2] or thoracic and abdominal movement signals [3] can be modelled as a superimposition of amplitude and frequency-modulated (AM/FM) modes, called *multicomponent signal* (MCS), and defined as

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

where the *instantaneous amplitudes* (IAs) $A_p(t)$, and the *instantaneous frequencies* (IFs) $\phi_p'(t)$ are set to be positive. Associated with this representation is the ideal TFR defined by:

$$IT_f(t,\eta) = \sum_{p=1}^{P} A_p(t)\delta(\eta - \phi_p'(t)). \qquad (2)$$

To estimate IT_f, the *short-time Fourier transform* (STFT)

$$V_f^h(t,\eta) = \int_{\mathbb{R}} f(x)h(x-t)\mathrm{e}^{-i2\pi\eta(x-t)}\mathrm{d}x, \qquad (3)$$

with h assumed to be a real window, is commonly used along with its squared modulus $S_f^h := |V_f^h|^2$, the *spectrogram*. The ideal representation associated with the spectrogram representation is thus the same as (2) except that the amplitude is replaced by the squared amplitude.

Traditionally, IFs are estimated from the spectrogram by considering *ridge detection*, which consists in detecting and connecting, along the time axis, local maxima along the frequency axis of the spectrogram [4]. Such a technique has three main drawbacks: first, it is constrained by the frequency resolution of the spectrogram, which limits its accuracy; second, it becomes irrelevant in the presence of strong interference between modes, especially when they are crossing; and finally, it is sensitive to the presence of noise.

Interference between modes, letting apart mode crossing, are strongly related to the window choice [5], and to automatically determine an appropriate window length to separate the modes is challenging. Unfortunately, such a window length may not be relevant to represent strongly frequency modulated modes. Indeed, the spectrogram reflects the frequency modulation of a signal in such a way that, on a ridge associated with a mode, its magnitude is smaller where a mode exhibits strong frequency modulation. As a consequence, ridges associated with a mode are less robust to noise where the frequency modulation of that mode is strong. To find a criterion to automatically determine an appropriate window length depending on whether the modes are frequency modulated or interfering is one objective of the present paper.

For that purpose, we propose an algorithm to locally determine, in the *time-frequency* (TF) plane, an appropriate window length associated with the spectrogram representation, and then define the corresponding *time-frequency representation* (TFR), which we call spectrogram approximation. We then explain how to adapt *ridge detection* (RD) based on the notion of *relevant ridge portions* (RRPs), introduced in [6], to that new TFR. Spline fitting is finally used to obtain the IF estimations of the modes making up the MCS, and this final step is specifically designed to handle potential mode crossings.

The paper is organized as follows. In Section II, we detail the spectrogram model, and emphasize how frequency modulation and interference affect that representation. This gives us a direction to follow to automatically determine the window length in the TF plane, which is done in Section III, in which we also describe the novel ridge detection and spline fitting used to perform the IF estimation of the modes. In Section IV, we recall different state-of-the-art techniques for IF estimation which we compare with in the final Section V.

The authors are with the Jean Kuntzmann Laboratory, University Grenoble Alpes and CNRS 5225, Grenoble 38401, France (emails: sylvain.meignen@univ-grenoble-alpes.fr, kevin.polisano@univ-grenoble-alpes.fr).

II. Model and Discussion

A. Spectrogram of Separated Modes

In the sequel, to allow for an analytic expression of the spectrogram, the Gaussian window $h(t) = h_\sigma(t) = \frac{1}{\sigma}\mathrm{e}^{-\pi\frac{t^2}{\sigma^2}}$ is used to compute the STFT and the spectrogram. The Fourier transform of that window is $\hat{h}_\sigma(\eta) = \mathrm{e}^{-2\pi\sigma^2\eta^2}$, denoted by g_σ. When the modes making up an MCS are well-separated pure tones $f_p(t) = A_p(t)\mathrm{e}^{i2\pi\omega_p t}$, the spectrogram can be approximated by [7] (provided the amplitudes of the modes vary slowly in time):

$$\begin{aligned} S_f^{h_\sigma}(t,\eta) \approx & \sum_{p=1}^{P} A_p(t)^2 g_\sigma(\eta-\omega_p) \\ = & \sum_{p=1}^{P} a_p(t) g_\sigma(\eta-\omega_p), \end{aligned} \tag{4}$$

which corresponds to the convolution along the frequency axis of IT_f with g_σ. To estimate $(\omega_p)_{p=1,\cdots,P}$, the *Fourier-based synchrosqueezing transform* (FSST) combines with ridge detection was proposed in [7]. As will be recalled later, it basically consists in reassigning the STFT along the frequency axis, and then computing the IF estimations by considering the value of the reassignment operator on the ridge of FSST. Alternatively, techniques based on *finite rate of innovation* (FRI) were developed in which the convolution kernel can either be a cardinal sine, a Gaussian function [8], or a more general compactly supported functions such as those satisfying the Strang–Fix conditions [9], [10].

A first important limitation of such approaches is that the spectrogram model assumes the modes are pure tones. When this is not the case, the spectrogram at time t is no longer the convolution along the frequency axis of IT_f with a fixed kernel. Therefore, FRI-based models for IF estimation can be highly inaccurate when the modes contain strong frequency modulation, though used with success in low frequency modulation situations [11]. To cope with that modulation issue, the *second order synchrosqueezing transform* (FSST2), considering a local second order polynomial approximation for the phase of the modes, was proposed to replace FSST in the IF estimation process [12], [13]. Such an approach was then extended to higher-order local polynomial phase approximation in [14], unfortunately increasing the computational cost of the transform and its sensitivity to noise. Indeed, during the reassignment process the noise contained in STFT coefficients is also reassigned, but it is very challenging to figure out how the noise is reassigned [15]. To cope for the noise issue, new techniques have been recently developed to locally adapt the local polynomial phase order depending on specific criteria, either by finding were a mode can be locally approximated by a linear chirp [16] or by locally changing the order of polynomial approximation [17].

B. Spectrogram of MCS Made of Modulated Modes

In this paper, we propose an alternative to reassignment operators to obtain IF estimations in the case of modulated modes. Our primary goal is to better understand the importance of modulation in the spectrogram representation. For that purpose, let us assume that the p^{th} mode deviates from pure harmonicity, and can be locally approximated around time t, by a linear chirp, namely $f_p(\tau) \approx A_p(t)\mathrm{e}^{i2\pi\left(\phi_p'(t)(\tau-t)+\frac{\phi_p''(t)}{2}(\tau-t)^2\right)}$. Then, the spectrogram of that mode satisfies [13]

$$S_{f_p}^{h_\sigma}(t,\eta) \approx A_p(t)^2 \frac{\mathrm{e}^{-2\pi\frac{\sigma^2(\eta-\phi_p'(t))^2}{1+\sigma^4\phi_p''(t)^2}}}{\sqrt{1+\sigma^4\phi_p''(t)^2}}, \tag{5}$$

where this approximation remains valid as long as the amplitude A_p varies slowly in the vicinity of t. Under this assumption, the spectrogram can be modeled by considering separated modes as follows:

$$S_f^{h_\sigma}(t,\eta) \approx \sum_{p=1}^{P} A_p(t)^2 g_{\sigma,\phi_p''(t)}(\eta-\phi_p'(t)), \tag{6}$$

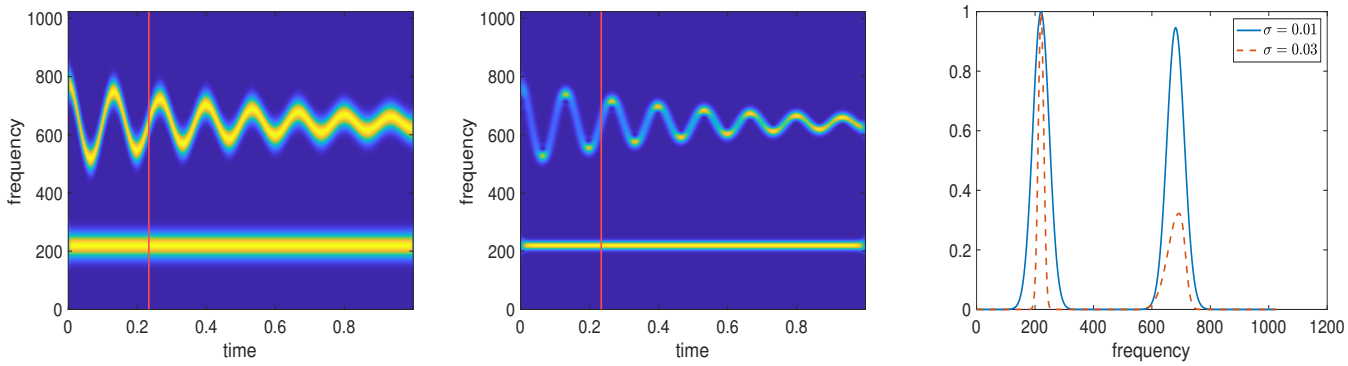
with

$$g_{\sigma,\phi_p''(t)}(\eta) = \frac{\mathrm{e}^{-2\pi\frac{\sigma^2\eta^2}{1+\sigma^4\phi_p''(t)^2}}}{\sqrt{1+\sigma^4\phi_p''(t)^2}}, \tag{7}$$

meaning $g_\sigma = g_{\sigma,0}$. So, considering $g_{\sigma,\phi_p''(t)}$ instead of g_σ, this formulation is similar to that involving pure tones, except that the convolution kernel now depends on the frequency modulation (or *chirp rate* (CR)) of each mode.

Frequency modulation not only affects the width of the Gaussian kernel but also its magnitude, which has a great impact on the quality of representation in noisy situations. To illustrate the error one makes by not taking into account the frequency modulation in the spectrogram model, we consider the two mode signal, both with amplitude 1, of Fig. 1, in which the left subfigure corresponds to the spectrogram obtained with a large σ, the middle one to a small σ, and the graph to the right represents a slice of the spectrogram associated with the time index corresponding to the vertical red line. This exemplifies that when σ is small, the spectrogram of a modulated mode is similar to that of a pure harmonic, which can be interpreted by remarking that $g_{\sigma,\phi_p''(t)}$ is closer to g_σ as σ goes to 0.

To obtain the same response in the spectrogram regardless of the modulation, it is clear that one had rather choose a small σ, in which case the spectrogram of the MCS resembles that of separated pure tones described in (4). Note that to choose such a small value for σ is somehow in contradiction with a choice commonly made in the literature which consists of minimizing the *Rényi entropy* associated with the spectrogram to find an appropriate σ to compute STFT [18]. Such an entropy has been extensively studied and has shown to be an interesting measure [19], [20], with very useful properties such as component counting [20]. As remarked in [21], for a linear chirp model [21] and when h_σ is used to compute the STFT, Rényi entropy is minimized for $\sigma = \frac{1}{\sqrt{\phi''(t)}}$. Such a value for σ minimizes the spreading of the kernel $g_{\sigma,\phi''(t)}$, which, in this case, is equal to $g_{\frac{\sigma}{\sqrt{2}}}$. However, it is not optimal for mitigating the impact of frequency modulation on the spectrogram.



(a) (b) (c)

Fig. 1: (a): spectrogram of a two mode signal (each mode has amplitude 1), $\sigma = 0.01$, (b): same as (a) but for $\sigma = 0.03$, (c): plot of the spectrograms restricted to the red line drawn on each of the first two figures

C. Spectrogram of MCS Made of Close Modes

In this section, we take a closer look at the case of two close modes, starting with the case of two pure tones.

1) The case of pure tones: Consider a signal f composed of only two pure tones $f(t) = A(t)\mathrm{e}^{i2\pi\omega_1 t} + \mathrm{e}^{i2\pi\omega_2 t}$, with $A(t)$ varying slowly. Computing its STFT gives the following approximation for the spectrogram:

$$S_f^{h_\sigma}(t,\eta) \approx \overbrace{A(t)^2\mathrm{e}^{-2\pi\sigma^2(\eta-\omega_1)^2} + \mathrm{e}^{-2\pi\sigma^2(\eta-\omega_2)^2}}^{\text{Modes part}} + \underbrace{2A(t)\mathrm{e}^{-\pi\sigma^2\frac{(\omega_2-\omega_1)^2}{2}}\mathrm{e}^{-2\pi\sigma^2\left(\eta-\frac{\omega_1+\omega_2}{2}\right)^2}\cos(2\pi(\omega_2-\omega_1)t)}_{\text{Interference part}}$$
$$= \sum_{p=1}^{3} a_p(t) g_\sigma(\eta-\omega_p), \tag{8}$$

with $\omega_3 = \frac{\omega_1+\omega_2}{2}$. An important remark is that $a_3(t)$ varies with time and can be negative. For the other modes, a_p approximates the squared amplitude of the modes. This result generalizes to a signal made of P harmonics $f(t) = \sum_{p=1}^{P} A_p(t)e^{i2\pi\omega_p t}$ where the spectrogram at time t can be approximated by (provided each amplitude A_p varies slowly)

$$S_f^{h_\sigma}(t,\eta) \approx \sum_{p=1}^{Q} a_p(t) g_\sigma(\eta-\phi_p'(t)) \text{ with } Q = \frac{P(P+1)}{2}. \tag{9}$$

Due to the quadratic nature of the spectrogram, each pair of modes in the signal will thus produce an additional term (hence the Q in Eq. (9)), responsible for some interference patterns observed when those modes get too close in frequency. It is thus clear that interference plays an important role in the spectrogram representation, but to take into account all the cross-terms in the spectrogram model is impractical, since the number Q of modes increases rapidly with P, and most of them are associated with very small amplitude. Nevertheless, we will see later in which circumstances we should bear in mind this interference term to improve IF estimation.

In the studied context, it is clear that interference terms are dumped by choosing a large σ as opposed to the choice of a small σ when the modes are separated to minimize the impact of frequency modulation on the spectrogram representation. This pleads in favor of adapting the window choice to the TF content of the signal. However, the situation becomes more complex when considering close but non-harmonic modes, as we shall now see.

2) The case of two parallel linear chirps: Let us consider the sum of two linear parallel linear chirps $A(t)\mathrm{e}^{i2\pi\phi_1(t)} + \mathrm{e}^{i2\pi\phi_2(t)}$ with $A(t)$ varying slowly. Then, one has [5]:

$$S_f^{h_\sigma}(t,\eta) \approx A(t)^2 g_{\sigma,c}(\eta-\phi_1'(t)) + g_{\sigma,c}(\eta-\phi_2'(t)) + I(t,\eta), \tag{10}$$

with $c = \phi_1''(t) = \phi_2''(t)$, and where the interference term takes the following form:

$$I(t,\eta) := 2A(t)\mathrm{e}^{-\pi\sigma_c^2\frac{(\phi_2'(t)-\phi_1'(t))^2}{2}}\cos(2\pi\phi(t,\eta)) \; g_{\sigma,c}\left(\eta - \frac{\phi_1'(t)+\phi_2'(t)}{2}\right), \tag{11}$$

with $\sigma_c = \frac{\sigma}{\sqrt{1+\sigma^4c^2}}$, and

$$\phi(t,\eta) := \left(\frac{\sigma_c}{\sigma}\right)^2(\phi_2(t)-\phi_1(t)) - c\sigma^2\sigma_c^2(\phi_2'(t)-\phi_1'(t))\left(\eta - \frac{\phi_1'(t)+\phi_2'(t)}{2}\right).$$

In this case, the non-oscillatory factor of the interference reads

$$a(t,\eta) = 2A\mathrm{e}^{-\pi\sigma_c^2\frac{(\phi_2'(t)-\phi_1'(t))^2}{2}}\cos(2\pi\phi(t,\eta)). \tag{12}$$

So, whatever the choice for σ, the interference term cannot be made negligible (except at TF points (t,η) where the cosine vanishes). More precisely, minimizing the absolute value of $a(t,\eta)$ requires maximizing σ_c, which corresponds to $\sigma = \frac{1}{\sqrt{c}}$, the same value obtained by minimizing the Rényi entropy. With this choice of σ, one has $g_{\sigma,c} = g_{\frac{\sigma}{\sqrt{2}}}$. Thus modulation alters the spectrogram representation, and interference is still present.

Based on this simple analysis, we conclude that it is not possible to get rid of interference simply by seeking an optimal parameter σ locally in the TF plane. The best we can do is to choose a σ that minimizes the importance of interference in the spectrogram and then improve IF estimation by taking

into account the remaining interference, as we shall see in the next section.

III. Algorithm for IF estimation

A. Spectrogram Approximation

To begin, we first discretize STFT in time and frequency, to obtain the discrete spectrogram $s_f^\sigma[n,k] \approx S_f^{h_\sigma}(\frac{n}{F_s}, \frac{k}{K}F_s)$, where $n \in [\![0, N-1]\!]$, $k \in [\![0, K-1]\!]$ and F_s is the sampling frequency.

In what follows, we consider that the discrete spectrogram is approximated by:

$$s_f^\sigma[n,k] \approx \sum_{p=1}^{R_n^\sigma} a_p[n] g_\sigma\left(\frac{k}{K}F_s - \eta_{p,n}\right), \tag{13}$$

with R_n^σ varying with n and σ, and being defined later.

To compute this approximation, let us consider the case of a noisy signal $\tilde{f} = f + \varepsilon$, with ε a complex Gaussian white noise with variance σ_ε^2. v_ε^σ, the discrete STFT of the noise, is also Gaussian with zero mean and satisfies [22]:

$$\mathrm{Var}\left(\Re\{v_\varepsilon^\sigma[n,k]\}\right) = \mathrm{Var}\left(\Im\{v_\varepsilon^\sigma[n,k]\}\right) = \sigma_\varepsilon^2 \|h_\sigma\|_2^2,$$

where $\|h_\sigma\|_2$ denotes the l^2-norm of h_σ. The ratio $\frac{s_\varepsilon^\sigma}{\sigma_\varepsilon^2\|h_\sigma\|_2^2}$ follows a χ_2 distribution with two degrees of freedom. Assuming the noise variance σ_ε^2 is known, the probability that $s_\varepsilon^\sigma[n,k] \geq 9\sigma_\varepsilon^2\|h_\sigma\|_2^2$ is less than 1%. To estimate $\gamma = \sigma_\varepsilon\|h_\sigma\|_2$, we use the robust estimator proposed in [23]:

$$\hat{\gamma} := \frac{\text{median}\left|\Re\left\{v_{\tilde{f}}^\sigma[n,k]\right\}_{n,k}\right|}{0.6745}. \tag{14}$$

For each time index n, we consider the local maxima along the frequency index of $s_{\tilde{f}}^\sigma[n,.]$ that are above the noise threshold $9\hat{\gamma}^2$, the location of which corresponds to the set of indices $\mathcal{L}_n^\sigma$. The coarse approximation of $s_f^\sigma[n,.]$ defined in (13) then corresponds to:

$$\tilde{s}_f^\sigma[n,k] = \sum_{q \in \mathcal{L}_n^\sigma} s_{\tilde{f}}^\sigma[n,q] g_\sigma(k-q). \tag{15}$$

In this model, modes associated with interference terms that do not correspond to extrema of the spectrogram are not considered. To take them into account when necessary, we proceed as follows. First, we define the essential support $[\![-\Omega^\sigma, \Omega^\sigma]\!]$ of the Gaussian g_σ as the interval where $|g_\sigma| > 10^{-3}$ restricted on the grid, hence

$$\Omega^\sigma = \left\lfloor \frac{K}{F_s}\sqrt{\frac{3\log(10)}{2\pi\sigma^2}} \right\rfloor, \tag{16}$$

where $\lfloor . \rfloor$ is the entire part. Next, we define the following sets of intervals

$$\left\{ I_q^\sigma = [\![q - \Omega^\sigma, q + \Omega^\sigma]\!],\ q \in \mathcal{L}_n^\sigma \right\}. \tag{17}$$

We then distinguish three cases which we detail hereafter:

- $I_q^\sigma \bigcap I_{q'}^\sigma = \emptyset, \forall q' \neq q$, the extremum located at q is isolated, we denote this set of locations as $\mathcal{S}_{1,n}^\sigma$.
- (q,q'), $I_q^\sigma \bigcap (\bigcup_{l \neq q} I_l^\sigma) = I_{q'}^\sigma \bigcap (\bigcup_{l \neq q'} I_l^\sigma) \neq \emptyset$, meaning I_q^σ and $I_{q'}^\sigma$ are intersecting but with no other intervals in the set defined by (17). We denote this set of pairs of locations (q,q') as $\mathcal{S}_{2,n}^\sigma$.
- I_q^σ intersects $I_{q'}^\sigma$ for several q' in $\mathcal{L}_n^\sigma$.

When q corresponds to the third situation we remove it from the spectrogram approximation, because it corresponds to a complex interference pattern that we shall not consider. When q is in $\mathcal{S}_{1,n}^\sigma$, we keep this extremum in the spectrogram approximation, and when $(q,q') \in \mathcal{S}_{2,n}^\sigma$, we take into account the interference term using the following strategy. As shown earlier, when two harmonic modes or two parallel linear chirps interfere, the mode created by the interference is associated with a Gaussian function centered at the average of the centers of the Gaussian creating the interference (see Eq. (11)). Consequently, we propose to exploit this property when identifying the position and amplitude of the Gaussian functions involved in (10). Specifically, assuming the modes interfering correspond to frequency indices q and q' assuming without loss of generality $q < q'$, we perform a refitting of the spectrogram approximation in the union of interval $I_q^\sigma \bigcup I_{q'}^\sigma$. This refitting procedure accounts for the interference term (incorporating the position constraint for this term) as well as the modulation (allowing flexibility in the standard deviation of the Gaussian function).

More precisely, we seek for the optimal vector of parameters $\boldsymbol{p}_0 = (a_1, a_2, a_3, \eta_1, \eta_2, \sigma)$ associated with the following Gaussian mixture model:

$$s_{\boldsymbol{p}}(x) = a_1 g_\sigma(x-\eta_1) + a_2 g_\sigma(x-\eta_2) + a_3 g_\sigma\left(x - \frac{\eta_1+\eta_2}{2}\right), \tag{18}$$

by minimizing the least-square error with the noisy spectrogram on $I_q^\sigma \bigcup I_{q'}^\sigma$, as summarized in Algorithm 1.

Algorithm 1: Refitting

Input: $s_{\tilde{f}}^\sigma$, F_s, K, q, q', σ

1: **Initialization**: $\boldsymbol{p}_0 = (1,1,0,q,q',\sigma)$

2: $\boldsymbol{p}^\star \leftarrow \underset{\boldsymbol{p}\in\mathbb{R}^6}{\text{argmin}} \|s_{\tilde{f}}^\sigma[n,.] - s_{\boldsymbol{p}}(\frac{F_s}{K}.)\|_{2, I_q^\sigma \bigcup I_{q'}^\sigma}^2$

3: $\tilde{a}_1, \tilde{a}_2, \tilde{\eta}_1, \tilde{\eta}_2 \leftarrow p_1^\star, p_2^\star, p_4^\star, p_5^\star$

Output: Positions and amplitudes $(\tilde{\eta}_p, \tilde{a}_p)_{p=1}^2$ of the refitted Gaussian functions.

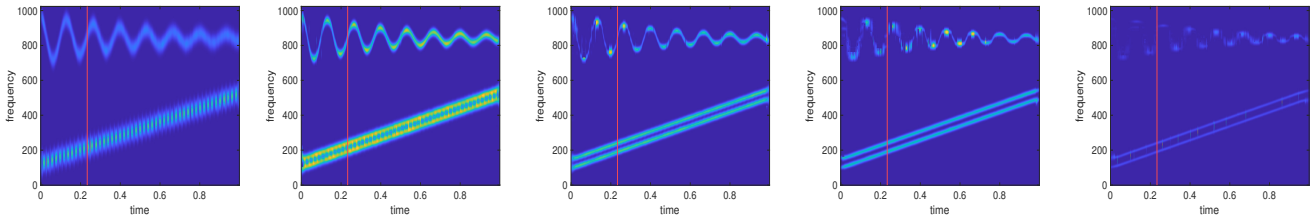
We then assign to the interval I_q^σ for $q \in \mathcal{S}_{1,n}^\sigma$ the following error, namely:

$$E^\sigma[n,k] = \sqrt{\frac{\sum_{r \in I_q^\sigma} (\tilde{s}_f^\sigma[n,r] - s_{\tilde{f}}^\sigma[n,r])^2}{\sum_{r \in I_q^\sigma} s_{\tilde{f}}^\sigma[n,r]^2}}, \forall k \in I_q^\sigma. \tag{19}$$

For $(q,q') \in \mathcal{S}_{2,n}^\sigma$, one defines $s_{\tilde{f},(q,q')}^\sigma(x) = \tilde{a}_1 g_\sigma(x-\tilde{\eta}_1) + \tilde{a}_2 g_\sigma(x-\tilde{\eta}_2)$, and then:

$$E^\sigma[n,k] = \sqrt{\frac{\sum_{r \in I_q^\sigma \bigcup I_{q'}^\sigma} (s_{\tilde{f},(q,q')}^\sigma(\frac{F_s}{K}r) - s_{\tilde{f}}^\sigma[n,r])^2}{\sum_{r \in I_q^\sigma \bigcup I_{q'}^\sigma} s_{\tilde{f}}^\sigma[n,r]^2}}, \forall k \in I_q^\sigma \bigcup I_{q'}^\sigma. \tag{20}$$

$s_{\tilde{f},(q,q')}^\sigma$ corresponds to the approximation without taking the interference term into account (though interference is



(a) (b) (c) (d) (e)

Fig. 2: (a): spectrogram approximation defined in Eq. (21) of a three mode signal (input SNR 20 dB), $\sigma = 0.01$, (b): same as (a) but for $\sigma = 0.02$, (c): same as (a) but for $\sigma = 0.03$;(d): same as (a) but for $\sigma = 0.04$; same as (e) but for $\sigma = 0.05$

used in the refitting process). The rationale to use such an approximation is that, while interference is used to improve IF estimation, and the error shall then measure the relative importance of the interference in the spectrogram. Finally, for any frequency index k other than those involved in Eq. (19) and Eq. (20), $E^\sigma[n,k]$ is set to 1. For that parameter value σ, the final approximation of the spectrogram reads:

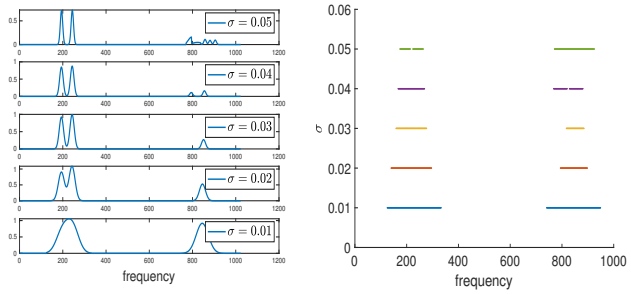
$$s^\sigma_{app}[n,.] = \sum_{q \in \mathcal{S}^\sigma_{1,n}} \tilde{s}^\sigma_f[n,.]\mathbb{1}_{I^\sigma_q} + \sum_{(q,q') \in \mathcal{S}^\sigma_{2,n}} s^\sigma_{\tilde{f},(q,q')}[n,.]\mathbb{1}_{I^\sigma_q \bigcup I^\sigma_{q'}}, \tag{21}$$

and R^σ_n corresponds to the cardinal of the extrema involved in that sum.

B. Optimizing the Window Size

A natural advantage of the just-defined approximation of the spectrogram is that the noise is naturally eliminated in (21), because only the extrema above the noise threshold are considered in the approximation. However, to illustrate the limitation of considering an approximation based on a single window, we plot s^σ_{app} in Fig. 2 for different values of σ, for a signal containing both interfering modes and strongly frequency-modulated modes. As expected, with a small value for σ, the strongly modulated mode is well represented in the spectrogram, but the two parallel linear chirp are entangled. When σ increases, the readability of the strongly modulated mode decreases while the linear chirps are better separated. So, considering an approximation (21) with a single window is not relevant.

Our strategy here is to compute the spectrogram approximation for different value of σ, and then find out what is locally the best approximation based on some criterion defined hereafter. For that purpose, we depict in Fig. 3 (a), the spectrogram approximation associated with the red lines drawn on the subfigures of Fig. 2. As far as the two linear chirps are concerned, we see that for a small σ, the two modes result in a single extremum along the frequency axis at each time instant. Thus, we expect the error E^σ to be huge at the corresponding TF locations because the interference is not taken into account in the approximation. Then, when σ increases, two close extrema associated with the two linear chirps appear, the approximation takes into account the interference, and the value of E^σ at the corresponding locations is expected to tell us which σ corresponds to the minimal contribution of the interference in the spectrogram. Switching to the strongly modulated mode,

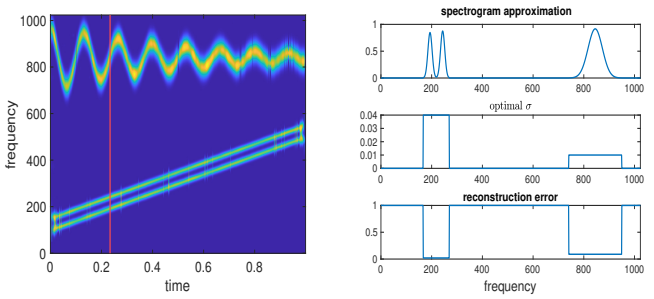


(a) (b)

Fig. 3: (a): spectrogram approximation corresponding to red lines on subfigures of Fig. 2; (b): corresponding non zero coefficients in s^σ_{app} for the studied values of σ

a small value for σ should lead to a small error E^σ at the corresponding locations because, on the one hand, it minimizes the error between the spectrogram model,i.e. without taking frequency modulation into account, and the true spectrogram in terms of amplitude, and on the other hand, for a large σ intra-mode time interference may occur, and the spectrogram departs from the spectrogram model compared with in the error estimation. Another important aspect we use for finding the optimal σ, is that as, when $\sigma_1 < \sigma_2$, $\Omega^{\sigma_1} > \Omega^{\sigma_2}$, the domain associated with an isolated mode shrinks when σ increases. This is also true in the case of two interfering modes, namely if $(q,q') \in \mathcal{S}^\sigma_{2,n}$, then the length of $I^\sigma_q \bigcup I^\sigma_{q'}$ decreases as σ increases. This is illustrated in Fig. 3 (b), in which we plot the support of s^σ_{app} for different σ. We also remark that when σ varies, the intervals corresponding to the spectrogram approximation of a mode intersect, so that one can easily make the connection between different approximation obtained with close σ.

Based on these remarks, our strategy is to define an approximation of the spectrogram based on an adapted window in the TF plane. This is done by considering an initial approximation of the spectrogram obtained with a small value of σ, and then by modifying it by considering spectrogram approximations obtained with larger values of σ. To clarify this, let us assume that $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_R)$ is a collection of window lengths ranked in increasing order. One considers as initial approximation $s^{\sigma_1}_{app}[n,.]$, associated with the error E^{σ_1}, and



(a) (b)

Fig. 4: (a): spectrogram approximation s_{app} of the signal of Fig. 2, obtained with Algorithm 2; (b): from top to bottom : coefficients corresponding to the red line in (a), corresponding σ values and approximation error

supported at time index n on the set

$$\mathcal{M}_n^{\sigma_1} = \left\{ I_q^{\sigma_1}, s.t.\ q \in \mathcal{S}_{1,n}^{\sigma_1}, I_q^{\sigma_1} \bigcup I_{q'}^{\sigma_1}, s.t. (q,q') \in \mathcal{S}_{2,n}^{\sigma_1} \right\} \tag{22}$$

For any other σ_i, $\mathcal{M}_n^{\sigma_i}$ is defined the same way. The idea we use to select locally the best window length is to investigate whether using a larger σ_i locally decreases the approximation error. If this is actually the case, the approximation spectrogram is changed accordingly, following the framework detailed in Algorithm 2.

Algorithm 2: WINDOW ADAPTED SPECTROGRAM APPROXIMATION

Input: $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_R)$, $(s^{\sigma_i}, E^{\sigma_i}, \mathcal{M}_n^{\sigma_i})_{i=1,\ldots,R}$

Initialization: $s_{app} = s^{\sigma_1}$, $E_{app} = E^{\sigma_1}$,
for $i = 1, \ldots, R-1$ **do**
  for $n \in [\![0, N-1]\!]$ **do**
    $C = \mathcal{M}_n^{\sigma_i}$
    for $X \in \mathcal{M}_n^{\sigma_{i+1}}$ **do**
      for $Y \in \mathcal{M}_n^{\sigma_i}$ **do**
        if $I = X \bigcap Y \neq \emptyset$ **then**
          if $\min_{k \in I} E^{\sigma_{i+1}}[n,k] < \min_{k \in I} E_{app}[n,k]$ **then**
            $E_{app}[n,k] = E^{\sigma_{i+1}}[n,k]$, **for** $k \in X \bigcup Y$
            $s_{app}[n,k] = s^{\sigma_{i+1}}[n,k]$, **for** $k \in X \bigcup Y$
            $C = \{C \setminus Y, X\}$
          end if
        end if
      end for
      if $X \bigcap (\bigcup_{Y \in \mathcal{M}_n^{\sigma_i}} Y) = \emptyset$ **then**
        $E_{app}[n,k] = E^{\sigma_{i+1}}[n,k]$, **for** $k \in X$
        $s_{app}[n,k] = s^{\sigma_{i+1}}[n,k]$, **for** $k \in X$
        $C = \{C, X\}$
      end if
    end for
    $\mathcal{M}_n^{\sigma_{i+1}} \leftarrow C$
  end for
end for

Output: $s_{app}, E_{app}, (\mathcal{M}_n^{\sigma_R})_{n=0,\ldots,N-1}$.

In Fig. 4 (a), we display s_{app} obtained as the output of Algorithm 2 applied to the three mode signal of Fig. 2, and, in Fig. 4 (b), the coefficients corresponding to the red lines in Fig. 4 (a), the corresponding optimal σ value and the approximation error. As expected, the algorithm selects a small value of σ for the isolated mode and finds an appropriate value of that parameter to well separate the two lower frequency parallel linear chirps.

C. Ridge Portions Detection

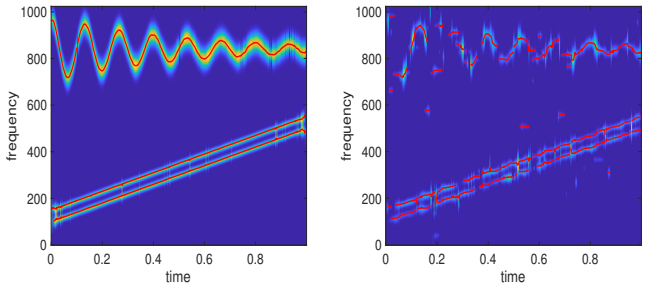
Once the approximation of the spectrogram s_{app} is computed, the next step towards IF estimation is ridge detection (RD), for which several techniques have been developed using the idea that the ridges correspond to *local maxima along the frequency axis* (LMFs) of the spectrogram. Indeed, as shown in [24], [25], the locations of the LMFs are estimates of the IFs of the modes, the quality of estimation depending on the noise level and on the length of the analysis window. A very important aspect is that RD should perform well in noisy situations, and, inspired by the work of [6], we remark that the ridge associated with one mode may be interrupted due to noise, mode interference or mode crossing. As a result, it is more relevant to use the concept of *relevant ridge portions* (RRPs) associated with a mode, rather than a single ridge.

In [6], RRPs consisted in LMFs above the noise level that were connected by means of the modulation operator used in the reassignment of the STFT. In our context, the spectrogram approximation at each time n is the sum of Gaussian functions associated with different values of σ, and whose maxima are above the noise threshold associated with the corresponding σ. So, every LMF in the spectrogram approximation is meaningful, and RD can be directly performed on that set of points.

Our approach to build the first RRP in the TF plane, consists in first picking a time index n_0, and then finding $c_1[n_0] = \text{argmax}_k\, s_{app}[n_0, k]$. The first candidate for the ridge portion is then defined as:

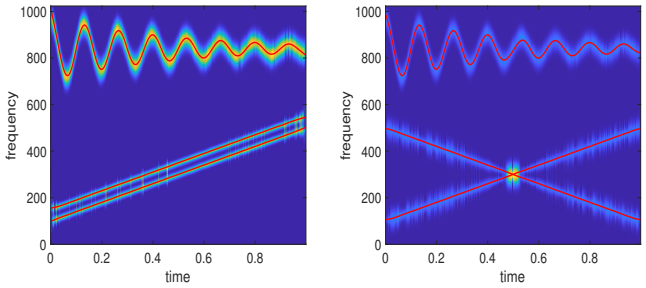
$$\max_{c_1} \sum_{n \in [n_0^-, n_0^+]} s_{app}[n, c_1[n]], \text{ s.t.} \begin{cases} |c_1[n+1] - c_1[n]| \leq B_f, \\ s_{app}[n, c_1[n]] \text{ is a LMF} \end{cases} \tag{23}$$

where B_f is a *jump parameter*, fixed a priori, that corresponds to the largest frequency modulation assumed on the modes, and the interval $[n_0^-, n_0^+]$ corresponds to the lower and upper bound where the conditions in (23) are no longer satisfied, and starting from n_0. Since the choice of n_0 is arbitrary, the first ridge portion is defined by considering not only the most energetic ridge portion as described above, but the most energetic one among all possible ridge portions starting from M_0 random initializations. Once this ridge portion is built, the set of corresponding LMFs are removed from the set of LMFs that can be used to build the next ridge portions. The process is then iterated until the length of the ridge portions is below a minimal length $l_{\min}$. An illustration of the procedure is given Fig. 5 (a) and (b) for an input signal-to-noise ratios (SNR) of 20 and 0 dB, respectively. It can be observed that for an SNR of 20 dB, the algorithm extracts a ridge associated with each mode, while for an SNR of 0 dB, a mode is associated with a



(a) (b)

Fig. 5: (a): spectrogram approximation of a three mode signal (input SNR 20 dB) along with the associated ridge portions ($l_{\min} = 10$); (b): same as (a) but for an input SNR of 0 dB.



(a) (b)

Fig. 6: (a): SA-IF estimation of a three mode signal (input SNR 20 dB); (b): same as (a) but for a three mode signal with cross-over frequencies.

series of ridge portions, most of which remain relevant even at that high noise level.

D. Spline Fitting

It is important to note that up to this point, neither in the definition of the spectrogram approximation nor in the construction of the ridge portions, we have made any assumptions about the number of modes present in the signal. However, to build IF estimations based on the detected ridge portions, we now assume that we want to compute P IF estimations, each of which is associated with a mode.

An important aspect is that our IF estimator should be able to adapt to the case of crossing modes. It is well-known that IF estimation based on RD in the TF plane analysis maybe not relevant in the case of mode crossing. For that purpose, *chirplet transform* (CT) [26] offers the possibility to separate modes with cross-over frequency by modifying the STFT using the *chirp rate* (CR) as an extra parameter. To improve the localization of modes making up a MCS in the time-frequency-CR domain, the reassignment of the modulus of CT was proposed in [27] and then further detailed in [28]. However, such approaches are time consuming and sensitive to noise, as we will demonstrate in the numerical section.

Our goal is to propose an IF estimator that outperforms CT-based approaches, recalled in the next section, when the signal contains modes with cross-over frequencies. We first gather the ridge portions together, starting with the set $\mathcal{I}_0$ of time indices corresponding to the longest interval where P ridge portions coexist. Denoting by $(n, c_p[n])_{n\in\mathcal{I}_0}$ the p^{th} ridge portion, we construct an initial spline approximation of the the P modes on $\mathcal{I}_0$ by minimizing:

$$\boldsymbol{\psi}_p^0 := \underset{\varphi}{\operatorname{argmin}} \sum_{n\in\mathcal{I}_0} (1-r)\left(c_p[n] - \varphi(\frac{n}{N})\right)^2 + r|\varphi^{(2)}(\frac{n}{N})|^2, \tag{24}$$

with $r \in [0,1]$. Then, let $\mathcal{I}_1$ be the set of time indices corresponding to the second longest time interval where P ridge portions coexist. To take into account potential mode crossings in the gathering of ridge portions, we consider any permutation m of $[\![1, P]\!]$ to compute:

$$\psi_{m,p}^1 := \underset{\varphi}{\operatorname{argmin}}(1-r)\left[\sum_{n\in\mathcal{I}_0}\left(c_p[n] - \varphi(\frac{n}{N})\right)^2 + \sum_{n\in\mathcal{I}_1}\left(c_{m[p]}[n] - \varphi(\frac{n}{N})\right)^2\right] + r\sum_{n\in\mathcal{I}_0\bigcup\mathcal{I}_1}|\varphi^{(2)}(\frac{n}{N})|^2. \tag{25}$$

and then the spline approximation we consider on $\mathcal{I}_0\bigcup\mathcal{I}_1$, is defined by:

$$\psi_p^1 := \underset{m}{\operatorname{argmin}} \sum_{n\in\mathcal{I}_0\bigcup\mathcal{I}_1} |(\psi_{m,p}^1)^{(2)}(\frac{n}{N})|^2. \tag{26}$$

Each of the P modes is thus associated with a set of two ridge portions and the process is iterated considering smaller interval length on which P modes coexist. At the end of this process, we obtain a set of P IF estimates which are denoted by $(\psi_p)_{p=1,\dots,P}$ hereafter. In the remainder of this paper, we refer to this technique as *SA-IF* estimation (for "spectrogram approximation instantaneous frequency" estimation). An illustration of the proposed SA-IF estimation is shown in Fig. 6 (a) for the case of two close modes plus a strongly frequency modulated one, and for an input SNR of 20 dB. Fig. 6 (b) then illustrates that SA-IF estimation technique perfectly handles modes with crossing over frequencies. A more systematic analysis of the behavior of this new IF estimation technique is carried out in Sec. V, along with a comparison with state-of-the-art methods.

IV. Other State-of-the-Art Methods for IF Estimation

In this section, we introduce different recent IF estimation techniques, based either on synchrosqueezing or chirplet transforms.

A. Synchrosqueezing Transform-Based Methods

Synchrosqueezing transform was originally proposed within the wavelet context [29], [30] and adapted to the STFT setting [31], [7]. In that latter context, it aims at "sharpening" the

STFT by vertically reassigning the coefficients above the noise threshold (defined in see Sec. III-A) through:

$$T_{\tilde{f}}^{h_\sigma,N}(t,\omega) := \frac{1}{h_\sigma(0)}\int_{\{\eta,|V_{\tilde{f}}^{h_\sigma}(t,\eta)|\geq 3\hat{\gamma}\}} V_{\tilde{f}}^{h_\sigma}(t,\eta)\delta(\omega-\omega_{\tilde{f}}^{[N]}(t,\eta))d\eta, \quad (27)$$

so that the reassigned transform is closer to the ideal TF signature. In that context, $\omega_{\tilde{f}}^{[N]}(t,\eta)$ is called the N-th order local IF estimation at time t and frequency η, and is built assuming the phases of the modes making up the signal locally behave as a polynomial of order N [14]. $\omega_{\tilde{f}}^{[N]}(t,\eta)$ has been proven to well approximate the frequency of the p^{th} mode in the TF plane when (t,η) is close to $(t,\phi_p'(t))$ and when that mode is well separated from the other modes [14]. In what follows, FSSTN denotes the synchrosqueezing transform of order N. The case of $N=1$ was addressed in [31], [7], $N=2$ in [12], and $N\geq 3$ was discussed in [14], giving birth to the so-called *high-order* synchrosqueezing transform. The computation of $\omega_{\tilde{f}}^{[N]}(t,\eta)$ can be carried out using the following two matrices of STFTs [32] (we omit (t,η) in the definition of the matrix for simplicity):

$$D^{[N]}(t,\eta) = \begin{bmatrix} V_{\tilde{f}}^{h_\sigma} & V_{\tilde{f}}^{th_\sigma} & \cdots & V_{\tilde{f}}^{t^{N-1}h_\sigma} \\ V_{\tilde{f}}^{th_\sigma} & V_{\tilde{f}}^{t^2h_\sigma} & \cdots & V_{\tilde{f}}^{t^Nh_\sigma} \\ \vdots & \vdots & \ddots & \vdots \\ V_{\tilde{f}}^{t^{N-1}h_\sigma} & V_{\tilde{f}}^{t^Nh_\sigma} & \cdots & V_{\tilde{f}}^{t^{2(N-1)}h_\sigma} \end{bmatrix},$$

and

$$U^{[N]}(t,\eta) = \begin{bmatrix} 0 & V_{\tilde{f}}^{th_\sigma} & \cdots & V_{\tilde{f}}^{t^{N-1}h_\sigma} \\ V_{\tilde{f}}^{h_\sigma} & V_{\tilde{f}}^{t^2h_\sigma} & \cdots & V_{\tilde{f}}^{t^Nh_\sigma} \\ \vdots & \vdots & \ddots & \vdots \\ (N-1)V_{\tilde{f}}^{t^{N-2}h_\sigma} & V_{\tilde{f}}^{t^Nh_\sigma} & \cdots & V_{\tilde{f}}^{t^{2(N-1)}\tilde{f}} \end{bmatrix},$$

where $V_{\tilde{f}}^{t^nh_\sigma}$ stands for the STFT with window $t^nh_\sigma(\cdot)$. With these notations one has [32]:

$$\tilde{\omega}_{\tilde{f}}^{[N]}(t,\eta) = \eta - \frac{1}{2\pi}\Im\left\{\frac{det(U^{[N]}(t,\eta))}{det(D^{[N]}(t,\eta))}\right\}. \quad (28)$$

To compute the IF estimates in practice, one first performs RD on FSSTN to obtain a first set of crude IF estimates $(t,\omega_p^{r,N}(t))_{p=1,\dots,P}$, on the grid (the superscript r standing for ridges, and N denoting the order of the synchrosqueezing transform), the final off-grid IF estimates being obtained as:

$$(\tilde{\omega}_{\tilde{f}}^{[N]}(t,\omega_p^{r,N}(t)))_{p=1,\dots,P}. \quad (29)$$

The quality of the IF estimation is thus tightly related to the quality of the ridge detection, and is not adapted to the case of cross-over frequencies. This will be further discussed in Section V. Finally note that FSSTs can either be computed using a single window, often associated with the minimal Rényi entropy [14], or using different time dependent windows [33], but cannot cope with a window whose length varying both in time in frequency, contrary to the novel IF estimation technique we propose.

B. Methods Based on Chirplet Transform and Ressignment

The *chirplet transform* (CT) is a generalization of the STFT and is defined for a signal f in $L^1(\mathbb{R})\bigcap L^2(\mathbb{R})$, the window h_σ, the chirp rate β in $\mathbb{R}$, and η in $\mathbb{R}_+$, by:

$$C_f^{h_\sigma}(t,\eta,\beta) := \int_{\mathbb{R}} f(\tau)h(\tau-t)e^{-i2\pi(\eta(\tau-t)+\beta\frac{(\tau-t)^2}{2})}d\tau. \quad (30)$$

The extra *chirp rate* (CR) parameter β enables to separate modes with cross-over frequencies, provided the noise level is relatively low. Such a separation, however, requires the definition of a ridge detector in the three-dimensional space indexed by time, frequency and CR. Going further, the reassignment of the modulus of CT was proposed in [27], a mathematical analysis being available in [28]. For the sake of consistency, we just recall how reassignment is performed in that context. Assume f is a local linear chirp namely, for τ in the vicinity of t, $f(\tau) = A(t)e^{i2\pi(\phi(t)+(\tau-t)\phi'(t)+\frac{(\tau-t)^2}{2}\phi''(t))}$. One can easily show that:

$$C_f^{h_\sigma}(t,\eta,\beta) = \frac{f(t)}{\sqrt{1+i\sigma^2(\beta-\phi''(t))}}e^{-\pi\frac{\sigma^2(\eta-\phi'(t))^2}{1+i\sigma^2(\beta-\phi''(t))}}. \quad (31)$$

Differentiating (31) with respect to η, and remarking that $\partial_\eta C_f^{h_\sigma} = -2i\pi C_f^{th_\sigma}$, a simple calculation leads to (we omit the variable (t,η,β) for the sake of simplicity):

$$\begin{aligned} \Re\left\{\frac{\partial_\eta C_f^{h_\sigma}}{C_f^{h_\sigma}}\right\} &= -\frac{2\pi\sigma^2(\eta-\phi'(t))}{1+\sigma^4(\beta-\phi''(t))^2} \\ \Leftrightarrow \phi'(t) &= \eta + \frac{1+\sigma^4(\beta-\phi''(t))^2}{\sigma^2}\Im\left\{\frac{C_f^{th_\sigma}}{C_f^{h_\sigma}}\right\}. \end{aligned} \quad (32)$$

Differentiating twice (31) with respect to η, one easily obtains:

$$\begin{aligned} \partial_\eta^2 C_f^{h_\sigma} &= -\frac{2\pi\sigma^2 C_f^{h_\sigma}}{1+i\sigma^2(\beta-\phi''(t))} + \frac{\left(\partial_\eta C_f^{h_\sigma}\right)^2}{C_f^{h_\sigma}} \\ \Leftrightarrow \phi''(t) &= \beta + \frac{1}{2\pi}\Im\left\{\frac{(C_f^{h_\sigma})^2}{\left(C_f^{th_\sigma}\right)^2 - C_f^{t^2h_\sigma}C_f^{h_\sigma}}\right\}. \end{aligned}$$

The reassigned CR, which equals $\phi''(t)$ in that case is thus

$$\tilde{\beta}^{CT}(t,\eta,\beta) = \beta + \frac{1}{2\pi}\Im\left\{\frac{(C_f^{h_\sigma})^2}{\left(C_f^{th_\sigma}\right)^2 - C_f^{t^2h_\sigma}C_f^{h_\sigma}}\right\}, \quad (33)$$

Then, replacing β by this expression in (32), one obtains

$$\phi'(t) = \eta + \frac{1}{\sigma^2}\Im\left\{\frac{C_f^{th_\sigma}(t,\eta,\tilde{\beta})}{C_f^{h_\sigma}(t,\eta,\tilde{\beta})}\right\},$$

and the frequency reassignment operator reads, in that case:

$$\tilde{\omega}^{CT}(t,\eta,\beta) = \eta + \Im\left\{\frac{C_f^{th_\sigma}(t,\eta,\tilde{\beta})}{C_f^{h_\sigma}(t,\eta,\tilde{\beta})}\right\}, \quad (34)$$

Finally, the modulus of CT is reassigned through:

$$\begin{aligned} D_f^{h_\sigma}(t,\eta,\beta) := \frac{1}{h_\sigma(0)}\int_{\mathbb{R}^+\times\mathbb{R}} |C_f^{h_\sigma}(t,v,\lambda)|^2 \\ \delta(\eta-\tilde{\omega}^{CT}(t,v,\lambda),\beta-\tilde{\beta}^{CT}(t,v,\lambda))dvd\lambda, \end{aligned} \quad (35)$$

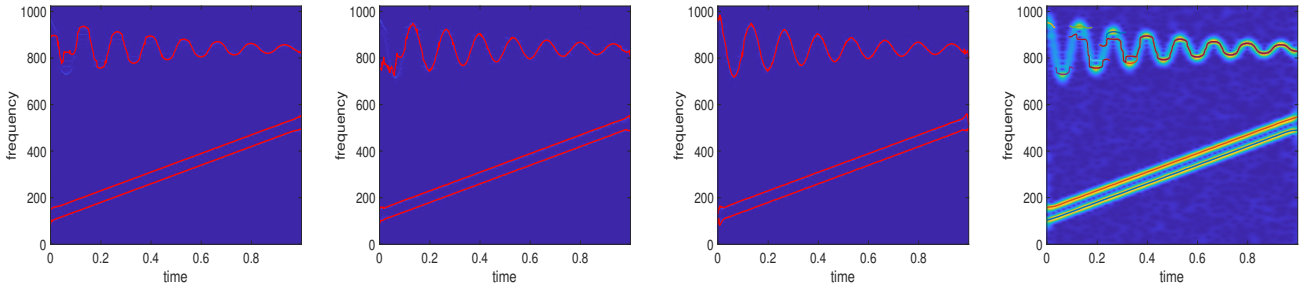


Fig. 7: (a): FSST2 based on STFT computed with optimal window length given by Rényi entropy with the first 3 most energetic ridges superimposed; (b): same as (a) but for FSST3; (c): same as (a) but for FSST4; (d): ridge portions used in RRP-RD ridge detector

where δ is the bidimensional Dirac distribution.

To obtain IF estimates, the approach used in [27] consists of computing CT ridges, corresponding to the P curves $(t, \omega_p^{r,CT}(t), \beta_p^{r,CT}(t))_{p=1,\dots,P}$, which are defined on the grid of CT (the superscripts r, CT standing for ridge, and chirplet transform), while the final IF estimations off-grid are obtained as:

$$\begin{aligned}&(\tilde{\beta}^{CT}(t, \omega_p^{r,CT}(t), \beta_p^{r,CT}(t)),\\ &\tilde{\omega}^{CT}(t, \omega_p^{r,CT}(t), \beta_p^{r,CT}(t)))_{p=1,\dots,P}.\end{aligned} \tag{36}$$

V. Results

In this section, we analyze the quality of the technique for IF estimation, called SA-IF, we propose by comparing it with state-of-the-art techniques, in the case of close mode, mode with strong frequency modulation or crossing modes.

A. IF Estimation in the Presence of Interference and Strong Frequency Modulation

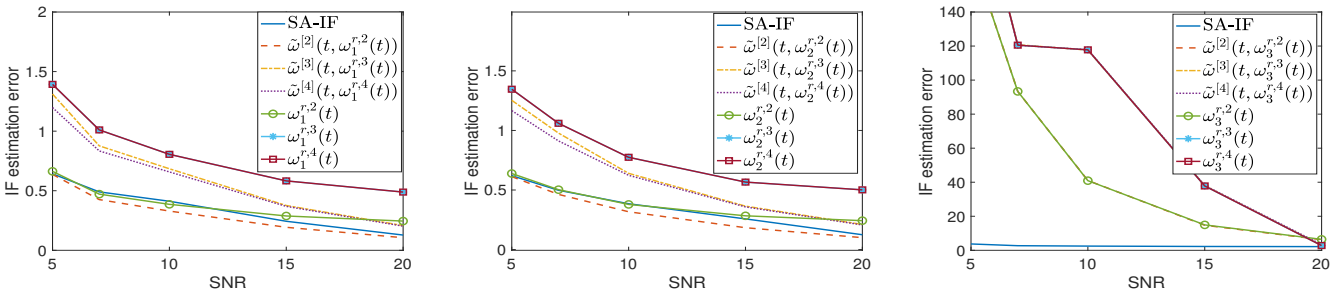
In this section, our goal is first to compare the relevance of the ridge portions computed on the spectrogram approximation compared with classical RD based on the synchrosqueezing transforms or ridge portions used in RRP-RD technique [6]. We believe that this will help to understand why SA-IF is ultimately better.

To illustrate the limitations of techniques based on synchrosqueezing, we again consider the signal first introduced in Fig. 2, with input SNR 20 dB, and perform different synchrosqueezing transforms on the STFT computed with the window length associated with the minimization of the Rényi entropy on the spectrogram with $\alpha = 3$ [14]. For each of the computed transforms, we extract the P most energetic ridges from these transforms, and superimpose them on Fig. 7 (a) to (c) to the synchrosqueezing transforms. As expected, it is necessary to increase the order of the synchrosqueezing transforms to capture the oscillations in the phase of the upper mode, and using estimate (29) with $N = 2$ or 3 will inevitably lead to inaccurate IF estimate for that mode. Comparing with Fig. 5 (a), only RD detection based on FFST4 seems to be able to compete with the approach we propose, but we will see hereafter that the former technique is very sensitive to noise. Finally, in Fig. 7 (d), we also plot the RRPs used in RRD-RD technique [6], computed on the same STFT as the one used in synchrosqueezing transforms. In that figure, we observe some time interference in the upper mode, and this results in multiple RRPs associated with that mode. This causes the failure of RRP-RD since this configuration is interpreted as the crossing of several modes, which is not handled by the algorithm for IF estimation proposed in [6],

To evaluate more precisely the benefits of SA-IF technique, we compare it with the different techniques based on synchrosqueezing transforms for the signal of Fig. 2 and for different SNRs. As a measure of the quality of IF estimation, we consider for each p the following error:

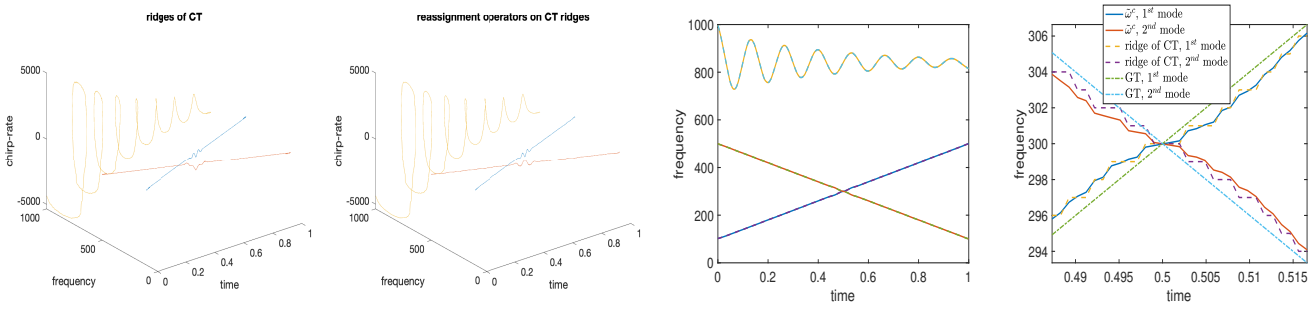
$$Er(p) = \frac{1}{N}\sum_{n=0}^{N-1} |\phi_p'(\frac{n}{N}) - \psi_p[n]|, \tag{37}$$

where $\psi_p[n]$ an IF estimate at time index n. The results are depicted in Fig. 8, for the three mode signal of Fig. 2 and, for each input SNR, we consider 10 realizations of the noise. Looking at the results of Fig. 8 (a) and (b), for the two linear chirps performance of SA-IF are similar to the IF estimation based on FSST2, which is known to be optimal for that type of modes. We shall also remark that, as expected, considering the reassignment operator on the ridges of FSST2 improves the results for that type of mode. Finally, it is important to note that the window length minimizing the Rényi entropy well separates the two linear chirps which explains why IF based on FFST2 works well in that situation, contrary to those based on FFST3 or FFST4, which are much more sensitive to noise. Regarding the oscillatory mode, the choice made on the window length creates time interference in some part of the signal, and the local linear chirp approximation used in FFST2 is no longer valid. Therefore, for such a mode and at very low noise level (around SNR equal to 20 dB), IF based on FFST3 or FFST4 behave better than FSST2. But, as the noise level increases, taking into account high order phase oscillations is no longer relevant. In any case, IF estimations based on FSSTs are not efficient for that type of mode compared with SA-IF, which is associated with a remarkably small estimation error regardless of the noise level.



(a) (b) (c)

Fig. 8: (a): IF estimation errors with respect to input SNR for the lowest frequency mode of Fig. 2;(b): same as (a) but for the other linear chirp; (c): same as (a) but for the mode with oscillatory phase.The results are averaged over 10 noise realizations



(a) (b) (c) (d)

Fig. 9: (a): ridges of CT for the signal of Fig. 6 (b); (b): reassigned operators on the ridges of CT; (c): IF estimations $\tilde{\omega}^{CT}$ on the ridges of CT and $\omega_p^{r,CT}$ for $p = 1, \ldots, 3$; (d): zoom in around the crossing of the two linear chirps

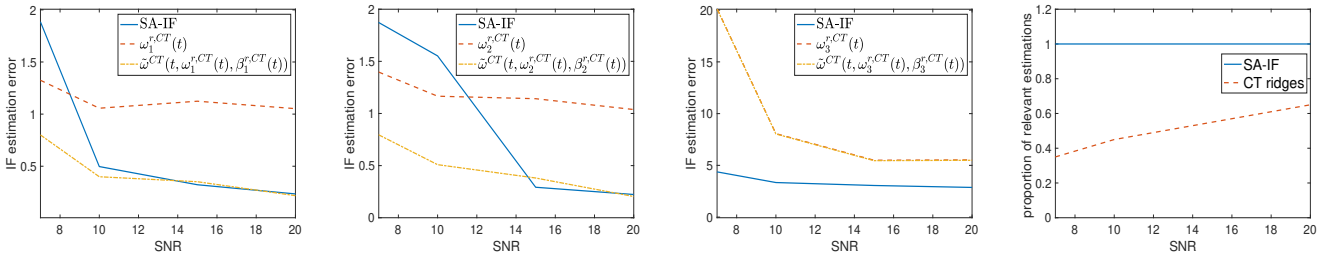
B. IF Estimation in the Presence of Crossing Modes and Strong Frequency Modulation

In this section we investigate the performance of SA-IF in the presence of crossing modes by comparing it with the techniques based on CT. To this end, we consider the signal of Fig. 6 (b) for which we already showed the nice behavior of SA-IF for a particular noise level.

To start with, we recall why mode separation can be performed using CT on that simple example, highlighting the limitation of such a technique. A first constraint is that the magnitude of CR can be very high for some modes: in the studied example, the magnitude of ϕ'' attains a maximum value of 5000 for the mode with oscillatory phase. Thus, the range for β must include the interval $[-5000, 5000]$. To keep a reasonable tensor size for the CT representation, the discretization along the CR axis has to be chosen appropriately. Then, when CT is used for IF estimation based on ridge detection, one needs to properly define a three dimensional ridge detector. For that purpose, once a ridge is extracted using the classical ridge detector based on peeling algorithm [34], one removes a neighborhood of this ridge in the time-frequency-CR three-dimensional space before extracting the following ridges. However, the size of this neighborhood, a frequency-CR rectangle at each time index n, is challenging to define, and depends on how close the CR of the different modes are. However, such an information is not available a priori. Furthermore, similarly to what happens in the TF plane where close modes result in oscillatory ridges, two modes with close CR lead to oscillatory ridges in the time-frequency-CR space. Such oscillations can end up in mode mixing, and all the more so that noise is present in the signal.

To illustrate this, in Fig. 9 (a), in the time-frequency-CR space, we plot the three-ridges associated with the studied signal (namely $(t, \omega_p^{r,CT}, \beta_p^{r,CT}(t))$ introduced before Eq. (36)), with input SNR 20 dB the window h_σ being still such that σ minimizes the Rényi entropy of the spectrogram. Also, we consider as IF estimates the values on these ridges of the reassignment operators (see Eq. (36)), in Fig. 9 (b). One clearly see that, oscillations are present in the detected ridges in the vicinity of the crossing. In Fig. 9 (c), we plot the different TF curves $(t, \omega_p^{r,CT})_{p=1,\ldots,P}$ and $(t, \tilde{\omega}^{CT}(t, \omega_p^{r,CT}(t), \beta_p^{r,CT}(t)))_{p=1,\ldots,P}$, exhibiting apparently no significant difference, but zooming in the crossing location in Fig. 8 (d), one observes a staircase effect when the IF estimation is performed using the ridges of the CT, as a result of frequency discretization, and that to consider the reassigned operator on these ridges does not lead to an improved IF estimation. It is worth finally noting that, while RD and subsequent IF estimation succeed in this particular example, as shown later, this technique lacks robustness to noise.

To further assess the quality of IF estimation under varying noise levels for the signal in Fig. 6 (b), we compare SA-IF with methods based on CT ridges and reassignment operators computed on these ridges. To make this comparison possible, we only compute the IF estimations when the RD they are based on is successful. To do so, we consider that, for a given IF estimation technique, RD has failed when the error



(a) (b) (c) (d)

Fig. 10: (a): IF estimation results with the different techniques for the first linear chirp; (b): same as (a) but for the second linear chirp; (c): same as (a) but for the upper mode; (d): proportion relevant IF estimations. In all these figures, the results are averaged over 20 noise realizations.

defined in (37) is larger than 40 for one of the mode at least. This means that the IF estimation error for given time n is larger than 40 Hz, which is huge compared to the error when the RD actually works fine. The IF estimation results are displayed in Fig. 10 (a) to (c). For the two crossing linear chirps, the three tested techniques exhibit similar performance (see Fig. 10 (a) and (b)), since a difference of 1 Hz in the estimation is negligible given the 1 Hz frequency sampling step. However, this similar behavior occurs only when RD has been successful, which is the case in only 40 % of the tested case at 10 dB for the techniques based on CT ridges. On the contrary, the ridges based on spectrogram approximation is always successful for the studied noise levels. Finally, when one considers the upper mode we see that techniques based on CT ridges are less accurate than SA-IF when ridge detection is successful.

VI. Conclusion

In this paper, we proposed a novel technique for instantaneous frequency estimation. Our approach first locally determines in the time-frequency plane an optimal window length for spectrogram approximation by minimizing interference in the case of close modes, and limiting frequency modulation effects in the representation of isolated modes. We then adapted an existing ridge detector to this optimized time-frequency representation and introduced a new IF estimation method based on ridge detector outputs and spline fitting. The effectiveness of our proposed estimator was evaluated on signals with close, crossing, and highly modulated frequency modes and was compared against state-of-the-art techniques. The results clearly demonstrate the advantages of our approach. For future work, the computational efficiency of our method could be significantly improved by parallelizing the spectrogram approximation, as each time index is independent of its neighbors. Once this optimization is achieved, we plan to explore its application to voice signal processing.

References

[1] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 101–111, 2003.

[2] C. L. Herry, M. Frasch, A. J. Seely, and H.-T. Wu, "Heart beat classification from single-lead ECG using the synchrosqueezing transform," *Physiological Measurement*, vol. 38, no. 2, pp. 171–187, 2017.

[3] Y.-Y. Lin, H.-T. Wu, C.-A. Hsu, P.-C. Huang, Y.-H. Huang, and Y.-L. Lo, "Sleep apnea detection based on thoracic and abdominal movement signals of wearable piezoelectric bands," *IEEE journal of biomedical and health informatics*, vol. 21, no. 6, pp. 1533–1545, 2017.

[4] 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.

[5] 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.

[6] 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.

[7] T. Oberlin, S. Meignen, and V. Perrier, "The fourier-based synchrosqueezing transform," in *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2014, pp. 315–319.

[8] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[9] P. L. Dragotti, M. Vetterli, and T. Blu, "Exact sampling results for signals with finite rate of innovation using strang-fix conditions and local kernels," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 4. IEEE, 2005, pp. iv–233.

[10] ——, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang–fix," *IEEE Transactions on signal processing*, vol. 55, no. 5, pp. 1741–1757, 2007.

[11] Q. Legros, D. Fourer, S. Meignen, and M. A. Colominas, "Instantaneous frequency and amplitude estimation in multi-component signals using an em-based algorithm," *IEEE Transactions on Signal Processing*, 2024.

[12] 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.

[13] 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.

[14] D. H. Pham and S. Meignen, "High-order synchrosqueezing transform for multicomponent signals analysis-with an application to gravitational-wave signal." *IEEE Trans. Signal Processing*, vol. 65, no. 12, pp. 3168–3178, 2017.

[15] M. Sourisseau, H.-T. Wu, and Z. Zhou, "Asymptotic analysis of synchrosqueezing transform—toward statistical inference with nonlinear-type time-frequency analysis," *The Annals of Statistics*, vol. 50, no. 5, pp. 2694–2712, 2022.

[16] M. A. Colominas and S. Meignen, "Instantaneous frequency estimation based on reassignment operators and linear chirp points detection," *IEEE Signal Processing Letters*, 2024.

[17] ——, "Adaptive order synchrosqueezing transform," *Signal Processing*, p. 109881, 2025.

[18] S. Meignen, D.-H. Pham, and M. Colominas, "On the use of short-time Fourier transform and synchrosqueezing-based demodulation for the retrieval of the modes of multicomponent signals," *Signal Processing*, vol. 178, p. 107760.

[19] W. J. Williams, M. L. Brown, and A. O. Hero III, "Uncertainty, information, and time-frequency distributions," in *Advanced signal processing algorithms, architectures, and implementations ii*, vol. 1566. SPIE, 1991, pp. 144–156.

[20] 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.

[21] 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.

[22] 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.

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

[24] L. Stanković, "A measure of some time–frequency distributions concentration," *Signal Processing*, vol. 81, no. 3, pp. 621–631, 2001.

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

[26] S. Mann and S. Haykin, "The chirplet transform: Physical considerations," *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2745–2761, 1995.

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

[28] Z. Chen and H.-T. Wu, "Disentangling modes with crossover instantaneous frequencies by synchrosqueezed chirplet transforms, from theory to application," *Applied and Computational Harmonic Analysis*, vol. 62, pp. 84–122, 2023.

[29] I. Daubechies and S. Maes, "A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models," *Wavelets in medicine and biology*, pp. 527–546, 1996.

[30] I. Daubechies, J. Lu, and H.-T. Wu, "Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 243–261, 2011.

[31] H.-T. Wu, "Adaptive analysis of complex data sets," Ph.D. dissertation, Princeton, 2011.

[32] S. Meignen and N. Singh, "Analysis of reassignment operators used in synchrosqueezing transforms: With an application to instantaneous frequency estimation," *IEEE Transactions on Signal Processing*, vol. 70, pp. 216–227, 2021.

[33] 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.

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