



HAL
open science

Instantaneous Frequency and Amplitude Estimation in Multi-Component Signals Using an EM-based Algorithm

Quentin Legros, Dominique Fourer, Sylvain Meignen, Marcelo Colominas

► To cite this version:

Quentin Legros, Dominique Fourer, Sylvain Meignen, Marcelo Colominas. Instantaneous Frequency and Amplitude Estimation in Multi-Component Signals Using an EM-based Algorithm. *IEEE Transactions on Signal Processing*, 2024, 72, pp.1130–1140. 10.1109/TSP.2024.3361713 . hal-04436653

HAL Id: hal-04436653

<https://hal.science/hal-04436653>

Submitted on 3 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Instantaneous Frequency and Amplitude Estimation in Multi-Component Signals Using an EM-based Algorithm

Quentin Legros, Dominique Fourer, Sylvain Meignen and Marcelo A. Colominas

Abstract—This paper addresses the problem of estimating the instantaneous frequency (IF) and amplitude of the modes composing a non-stationary multi-component signal in the presence of noise. A novel observation model for the signal spectrogram is developed within a Bayesian framework to handle intricate configurations involving noise or overlapping components. The model parameters are estimated using a stochastic variant of the Expectation-Maximization algorithm, bypassing the computationally expensive joint parameter estimation from the posterior distribution. We then design an algorithm for instantaneous amplitude and frequency estimation that accounts for overlap and amplitude variations of the components. To assess the performance of the proposed method, we conduct experiments on both real-world and simulated signals, involving separated or crossing modes. The benefits of our method in terms of efficiency compared with several state-of-the-art techniques appear to be significant in that latter case, but also when the amplitude of the components are varying across time.

Index Terms—ridge extraction, time-frequency, Bayesian estimation, Expectation-Maximization algorithm, Monte Carlo sampling.

I. INTRODUCTION

DISENTANGLING natural or synthesized signals in a reliable and fast way is often necessary for a wide range of applications, including biomedicine [1], audio [2], seismic [3] or radar [4]. For that purpose, a common model for representing such signals, referred to as non-stationary Multi-component Signal (MCS), is a superimposition of Amplitude- and Frequency-Modulated (AM-FM) components (or modes). For the last decade, numerous methods have been developed to extract the modes of a MCS, which is of paramount importance to understand the signal content.

One can mention the Empirical Mode Decomposition (EMD) [5], [6] and the Singular Spectrum Analysis (SSA) [7] which belong to data-driven approaches designed to estimate the modes directly from the observed waveform. Other techniques consist in projecting the signal onto the Time-Frequency (TF) plane using a linear transform such as the

Short-Time Fourier Transform (STFT) or the Continuous Wavelet Transform (CWT) [8], revealing TF curves (also called ridges) associated with modes, which correspond to estimates of the components Instantaneous Frequency (IF). These techniques combine a Time-Frequency Representation (TFR) with a ridge detection method, based for instance on Bayesian inference [9], to disentangle the components. Hence, the data extracted from the ridges (i.e. TF coordinates, STFT magnitude and phase, chirp rate, etc.) provide with valuable inputs for a variety of applications, such as source separation [10], signal disentangling [11] or to enhance the readability of the signal in the presence of noise [9], [12].

A wide range of methods are designed to deal explicitly with the presence of spurious noise [13], [14]. These methods highlight the importance of regularization in ridge detection in that context [13], [15]. In [16] for instance, a Total Variation (TV) regularization is applied to the signal synchrosqueezing transform [17] to estimate the modes IF. A Pseudo-Bayesian (PB) framework is introduced in [9], where the IF of each mode is sequentially estimated at each time instant using alternative divergences and a regularization term based on a simple prior model.

Nonetheless, these methods are not designed to perform Instantaneous Amplitude (IA) estimation, and their performance is limited when the modes overlap in the TF plane. While existing methods tackle this problem, as in [18] where the signal is projected onto a three-dimensional time-frequency-chirp-rate space to disentangle its modes, most of them are computationally expensive.

The present work aims to bridge this gap by introducing a novel observation model in a Bayesian framework, based on the analysis of the MCS spectrogram. Our mixture model introduced in Section II, is parameterized by the IA and IF of the modes, and accounts for the possible presence of noise. The prior distributions assigned to each model parameter are presented in Section III, before formulating the estimation strategy in Section IV, where the model parameters are estimated using an Expectation-Maximization (EM)-based algorithm [19], [20] to avoid their computationally heavy estimation from the joint posterior. Our algorithm is designed for the use of Markovian prior models at a low computational cost. A sequential Maximum a Posteriori (MAP) algorithm is formulated to perform the estimation of overlapping modes at each iteration of the EM algorithm, significantly reducing the computational complexity of the method. Finally, the performance of the proposed algorithm is evaluated in Section V,

Q. Legros (quentin.legros@univ-orleans.fr) and S. Meignen (sylvain.meignen@univ-grenoble-alpes.fr) are with the Jean Kuntzmann Laboratory, University Grenoble Alpes, Grenoble 38401, France. D. Fourer (dominique.fourer@univ-evry.fr) is with the IBISC Laboratory, University of Evry Paris-Saclay, Évry-Courcouronnes 91020, France. Marcelo A. Colominas (macolominas@conicet.gov.ar) is with the Institute for Research and Development in Bioengineering and Bioinformatics (IBB), CONICET, and with Facultad de Ingeniería, Universidad Nacional de Entre Ríos (UNER), Oro Verde, Argentina. This research was supported by the French ANR ASCETE project (ANR-19-CE48-0001). The authors thank Curtis Condon, Ken White, and Ai Feng of the Beckman Institute of the University of Illinois for the bat data and for permission to use it in this paper.

through a comparison with several state-of-the-art methods applied to both synthetic and real-world signals. The main contributions of the paper can be summarized as follows:

- A novel observation model to infer both the IA and IF of the modes of a MCS from its spectrogram using an EM-based algorithm.
- A stochastic approach designed to keep a reasonable computational cost for the method.
- An estimation strategy adapted to MCS with overlapping modes.

II. OBSERVATION MODEL

A. Time-Frequency Analysis

Let x be a signal made of K superimposed AM-FM components expressed, $\forall n \in \mathbb{Z}$, as:

$$x(n) = \sum_{k=1}^K x_k(n), \quad \text{with } x_k(n) = \alpha_k(n) e^{j2\pi\phi_k(n)}, \quad (1)$$

where $\alpha_k(n)$ and $\phi_k(n)$ are, respectively, the time-varying amplitude and phase of the k -th component at time n . The STFT of signal x , using a real-valued analysis window θ , is defined as:

$$F_x^\theta(n, m) := \sum_{l \in \mathbb{Z}} x(l) \theta(l - n) e^{-j \frac{2\pi m(l-n)}{M}}, \quad (2)$$

at each time instant $n \in \llbracket 0, N-1 \rrbracket$ and each frequency bin $m \in \llbracket 0, M-1 \rrbracket$. The spectrogram is then defined as a matrix $\mathbf{S} = \{|F_x^\theta(n, m)|^2\}_{n,m}$ where each column at time index n is denoted by $\mathbf{s}_n = [s_{n,0}, \dots, s_{n,M-1}]^\top$. Now, we assume that x is a sum of pure harmonics with constant IF and IA, namely:

$$x(n) = \sum_{k=1}^K A_k e^{2j\pi\eta_k n}, \quad (3)$$

with IF $\frac{d\phi_k}{dn}(n) = \eta_k \in [0, 1]$, and IA $\alpha_k(n) = A_k$. It follows from Eq. (2) that:

$$\begin{aligned} F_x^\theta(n, m) &= \sum_{k=1}^K A_k e^{2j\pi\eta_k n} \sum_{l \in \mathbb{Z}} \theta(l) e^{-2j\pi(\frac{m}{M} - \eta_k)l} \\ &\approx \sum_{k=1}^K A_k e^{2j\pi\eta_k n} F_\theta\left(\frac{m}{M} - \eta_k\right), \end{aligned} \quad (4)$$

where F_θ is the Fourier transform of θ in $L^1(\mathbb{R})$. Assuming the modes to be sufficiently far apart with respect to the essential support of F_θ , it follows:

$$|F_x^\theta(n, m)|^2 \approx \sum_{k=1}^K A_k^2 \left(F_\theta\left(\frac{m}{M} - \eta_k\right) \right)^2, \quad (5)$$

which does not depend on n . Further, assuming that both IF and IA vary over time, Eq. (5) can be generalized into:

$$|F_x^\theta(n, m)|^2 \approx \sum_{k=1}^K \alpha_k^2(n) \left(F_\theta\left(\frac{m}{M} - \frac{d\phi_k}{dn}(n)\right) \right)^2, \quad (6)$$

on which our observation model is based.

B. Likelihood

We consider the following observation model built from Eq. (6):

$$p(s_{n,m} | \mathbf{w}_n, \mathbf{m}_n) = \sum_{k=1}^K w_{n,k} g(m - m_{n,k}) + \frac{1 - \sum_{k=1}^K w_{n,k}}{M} \quad (7)$$

where $g(m) = \left(\frac{F_\theta(\frac{m}{M})}{\|F_\theta\|_2} \right)^2$, with $\|\cdot\|_2$ the L^2 -norm, and $\frac{m_{n,k}}{M}$ an approximation of $\frac{d\phi_k}{dn}(n)$, such that $m_{n,k} \in \llbracket 0, M-1 \rrbracket$ is the k -th ridge position at time n . Considering a Gaussian analysis window $\theta(n) = \frac{1}{\sqrt{2\pi L}} e^{-\frac{n^2}{2L^2}}$, with L the time-spread parameter [11], we obtain $g(m) = \frac{2\sqrt{\pi L}}{M} e^{-(\frac{2\pi m L}{M})^2}$ that is used in the remainder of this work.

Moreover, the weights $w_{n,k}$ are constrained to belong to $[0, 1]$ such that $\sum_k w_{n,k} \leq 1$. In what follows, we set $\mathbf{w}_n = [w_{n,1}, \dots, w_{n,K}]^\top$, $\mathbf{m}_n = [m_{n,1}, \dots, m_{n,K}]^\top$, and the vector $[m_{0,k}, \dots, m_{N-1,k}]$ is called the k -th ridge in the sequel.

The observation model in Eq. (7) describes the probability distribution of each single element $s_{n,m}$ in column \mathbf{s}_n , for any time n . As can be seen from Eq. (7), $\sum_k w_{n,k} = 1$ in the absence of noise, reducing the model to a mixture of K Gaussian functions. Conversely, Eq. (7) reduces to a uniform distribution in pure noise situations since $\sum_k w_{n,k} = 0$. Despite the estimation of the number of components K is out of the scope of this paper, we assume it to be known in the remainder of this paper and refer to [21], [22] for further analysis.

The model in Eq. (7) is tightly related to how well the spectrogram can be approximated by Eq. (5). Such an approximation is proved to be valid only when both the frequency modulations of the components and the IA variations are small [23]. This can be circumvented by enforcing smoothness of the IF estimates, through the use of prior models as discussed in the next section.

For further developments, we set $\mathbf{W} = \{\mathbf{w}_n\}_{n=0}^{N-1}$ and $\mathbf{M} = \{\mathbf{m}_n\}_{n=0}^{N-1}$, which are the model parameters we seek to estimate. Assuming independence between each TF point conditioned on the value of $(\mathbf{w}_n, \mathbf{m}_n)$, we express the joint likelihood function as:

$$\begin{aligned} p(\mathbf{S} | \mathbf{W}, \mathbf{M}) &= \prod_n \prod_m p(s_{n,m} | \mathbf{w}_n, \mathbf{m}_n). \\ &= \prod_n p(\mathbf{s}_n | \mathbf{w}_n, \mathbf{m}_n). \end{aligned} \quad (8)$$

III. PRIOR MODELS

In order to complete the Bayesian model, prior distributions are assigned to the model parameters to account for the available a priori information [24]. A weak uniform prior model is assigned to the mixture weights \mathbf{W} to deal with the lack of prior information. In the next section, we introduce the two prior models we consider for \mathbf{M} .

A. Total Variation Prior

In the presence of a strong level of noise, constructing smooth ridges can be accomplished by enforcing spatial

smoothness between successive IF estimates, for instance by constraining the derivative of the estimates [13]. In our Bayesian framework, a way to constrain such derivatives is to consider the following TV Markov Random Field (MRF) prior model for \mathbf{M} which is known to preserve sharp transitions [25], [26]

$$p(\mathbf{M}|\epsilon) \propto \exp \left[-\epsilon \sum_{k=1}^K \|\Delta^1 \mathbf{m}_{\cdot,k}\|_1 \right], \quad (9)$$

with Δ^1 denoting the first order finite difference, $\mathbf{m}_{\cdot,k}$ the k -th row of \mathbf{M} , i.e. the k -th ridge position, $\|\cdot\|_1$ the L^1 -norm, and the prior weight ϵ a user-defined hyperparameter. Also, using a TV prior enables the estimates to make sharp transitions when the ridges are split, as can be observed in Fig. 1 where we apply our IF estimation method described in the following section.

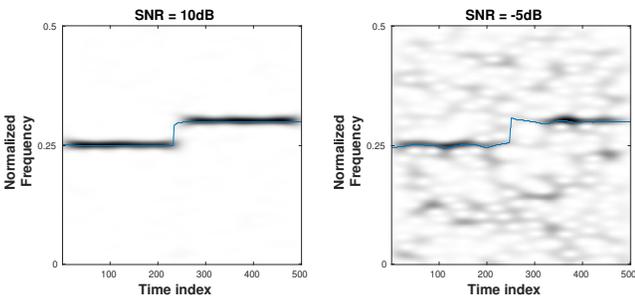


Fig. 1. Spectrograms of a signal made of two portions of noisy sinusoidal waves, resulting in a SNR of 10 dB (left) and -5 dB (right). The blue curve corresponds to the IF estimation performed using the method described in Section IV using the TV prior model with $\epsilon = 10^{-1}$.

B. Laplacian Prior

Another possible choice for the prior of \mathbf{M} is to constrain the mean curvature of the estimated ridge using a MRF Laplacian prior model [27], [28] by setting an L^2 -norm penalization, namely to define:

$$p(\mathbf{M}|\lambda) \propto \exp \left[-\frac{\lambda}{2} \sum_{k=1}^K \|\Delta^2 \mathbf{m}_{\cdot,k}\|_2^2 \right], \quad (10)$$

which is a log-concave and differentiable operator controlling the smoothness of the estimation, in which Δ^2 denotes the second order finite difference and λ a user-defined hyperparameter.

IV. ESTIMATION STRATEGY

TABLE I
SUMMARY OF THE PARAMETERS.

Notations	
\mathbf{S}	Signal spectrogram
\mathbf{W}	Observed model parameter (mixture weights)
\mathbf{M}	Latent model parameter (IF)
ϵ, λ	Prior hyperparameters

For the sake of clarity, we omit in the sequel the hyperparameters ϵ and λ related to the prior models. Using the

Bayes rule, the joint posterior distribution of (\mathbf{W}, \mathbf{M}) can be expressed as:

$$p(\mathbf{W}, \mathbf{M}|\mathbf{S}) \propto p(\mathbf{S}|\mathbf{W}, \mathbf{M})p(\mathbf{M})p(\mathbf{W}). \quad (11)$$

Thus, we present in Table. I a reminder of the Bayesian model parameters.

A. Expectation-Maximization Based Algorithm

Jointly estimating (\mathbf{W}, \mathbf{M}) is a challenging task due to the latent parameter \mathbf{M} making the function in Eq. (8) multimodal when $K > 1$. EM-based algorithms [19], [29] are particularly adapted to address this problem, and the shape of the model in Eq. (7) is well suited for such methods. Therefore, we propose to marginalize over the latent parameter \mathbf{M} to estimate \mathbf{W} as:

$$\hat{\mathbf{W}}_{\text{MMAP}} = \underset{\mathbf{W}}{\operatorname{argmax}} \sum_{\mathbf{M}} p(\mathbf{W}, \mathbf{M}|\mathbf{S}) = \underset{\mathbf{W}}{\operatorname{argmax}} p(\mathbf{W}|\mathbf{S}). \quad (12)$$

The EM algorithm is a two-steps iterative algorithm that computes a MAP estimate. More precisely, given $\mathbf{W}^{(i)}$ the current estimate of \mathbf{W} at iteration i , the E-step consists of computing:

$$\begin{aligned} Q(\mathbf{W}|\mathbf{W}^{(i)}) &= E_{\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S}} [\log(p(\mathbf{W}, \mathbf{M}|\mathbf{S}))] \\ &= E_{\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S}} [\log(\prod_n p(\mathbf{w}_n, \mathbf{m}_n|\mathbf{s}_n))] \\ &= \sum_n E_{\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S}} [\log(p(\mathbf{w}_n, \mathbf{m}_n|\mathbf{s}_n))], \end{aligned} \quad (13)$$

followed by the M-step which finds the parameters maximizing $Q(\mathbf{W}|\mathbf{W}^{(i)})$:

$$\mathbf{W}^{(i+1)} = \underset{\mathbf{W}}{\operatorname{argmax}} Q(\mathbf{W}|\mathbf{W}^{(i)}). \quad (14)$$

Nonetheless, the expectation in Eq. (13) is intractable due to the Markovian nature of the prior model. To deal with this issue, a classical approach consists in resorting to Stochastic Expectation-Maximization (SEM) algorithms [19], [20]. Nevertheless, the multimodality of the conditional distribution $p(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$, as well as the overlapping of components would imply long simulation steps to get an accurate approximation of $p(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$. Therefore we adopt instead a strategy similar to [30] as detailed in the next section.

B. Stochastic Approach to Approximate $p(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$

In order to replace $p(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$ used in Eq. (13) by an approximating distribution such that the expectation becomes tractable, we compute an approximation $\tilde{p}(\mathbf{M}|\bar{\mathbf{M}})$ of $p(\mathbf{M})$, by generating an auxiliary random sample $\bar{\mathbf{M}}$ through Markov chain Monte Carlo (MCMC) simulations such that

$$\tilde{p}(\mathbf{M}|\bar{\mathbf{M}}) = \prod_{n=0}^{N-1} p(\mathbf{m}_n|\bar{\mathbf{m}}_{\setminus n}), \quad (15)$$

where $\bar{\mathbf{m}}_{\setminus n}$ is the matrix $\bar{\mathbf{M}}$ whose n -th row has been removed. The approximating prior can then be used to compute an approximation of the conditional distribution $p(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$ using Bayes rule as:

$$\tilde{p}(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S}) = \frac{p(\mathbf{S}|\mathbf{W}^{(i)}, \mathbf{M})\tilde{p}(\mathbf{M}|\bar{\mathbf{M}})}{p(\mathbf{S})}. \quad (16)$$

The approximation of $p(\mathbf{M})$ is performed using a 2-step Gibbs sampler. At each iteration, we hot-start the Gibbs sampler using the current estimate of \mathbf{M} (discussed in Section IV-D) to avoid the required long burn-in period due to random initialization of the Gibbs sampler. Since we are sampling from the prior, each iteration of the Gibbs sampling increases the importance of the prior regularization. Hence, a single iteration of the Gibbs sampling is performed at each iteration of the algorithm (discussed in Section V-E). If not stated otherwise, we set the prior weight of our method to $\epsilon = \lambda = 10^{-2}$.

While this approach is similar to the classical SEM [19], [20], here we sample from $p(\mathbf{M})$ instead of $p(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$ to compute an approximating distribution $\tilde{p}(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$. This choice is motivated by the necessity to reduce the computational cost of the M-step (discussed in the next section). To do so, we propose to estimate \mathbf{M} from $p(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$ at each iteration of the algorithm. To ensure a reasonable computational complexity of the method, a designed method (discussed in the next section) is preferred to MCMC simulations.

C. EM-based Algorithm

Using the approximation of $p(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$ at step i , we compute the MAP estimate of \mathbf{M} , denoted by $\mathbf{M}^{(i+1)}$. The whole algorithm can then be significantly fastened using this current estimate, since the M-step reduces to:

$$\begin{aligned} \mathbf{W}^{(i+1)} &= \underset{\mathbf{W}}{\operatorname{argmax}} Q(\mathbf{W}|\mathbf{W}^{(i)}) \\ &\approx \underset{\mathbf{W}}{\operatorname{argmax}} \log \left(p(\mathbf{W}, \mathbf{M}^{(i+1)}|\mathbf{S}) \right) \\ &\approx \underset{\mathbf{W}}{\operatorname{argmax}} \log \left(p(\mathbf{S}|\mathbf{W}, \mathbf{M}^{(i+1)}) \right) + C, \end{aligned} \quad (17)$$

where $C = \log \left(\frac{p(\mathbf{M})p(\mathbf{W})}{p(\mathbf{S})} \right)$ is a constant since a uniform prior model is assigned to \mathbf{W} . From Eq. (7), the likelihood is log-concave with respect to \mathbf{W} which ensures that $Q(\mathbf{W}|\mathbf{W}^{(i)})$ is also concave, allowing for the use of time-wise convex optimization approaches to solve the M-step in Eq. (14), and thus to perform an estimation of \mathbf{W} .

The M-step is solved using the Newton-Raphson algorithm [19], whose computational complexity only depends on the number K of components. While the convergence of the algorithm is fast (less than 5-10 iterations), the estimates oscillate around the Marginal Maximum a Posteriori (MMAP) estimator of \mathbf{W} instead of converging to it, due to the stochastic nature of the method. The oscillations being small in practice, we empirically stop iterating after $N_{\text{iter}} = 20$ iterations.

A summary of the proposed EM algorithm is available in Algo. 1 which uses Algo. 2 (discussed in Section IV-D) for estimating \mathbf{M} at each iteration.

ALGORITHM 1

EM-based estimation of \mathbf{M} and \mathbf{W}

- 1: **Input:** \mathbf{S} , ϵ or λ , number of EM iterations N_{iter}
 - 2: **Initialization:** Set $\mathbf{W}^{(0)}$
 - 3: **for** $i = 0, \dots, N_{\text{iter}} - 1$ **do**
 - 4: Compute $p(\mathbf{S}|\mathbf{W}, \mathbf{M})$, using Eq. (7)
 - 5: Sample $\bar{\mathbf{M}} \sim p(\mathbf{M})$
 - 6: Compute $\tilde{p}(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$, using Eq. (16)
 - 7: Estimate $\mathbf{M}^{(i+1)}$, using Algo. 2
 - 8: Compute $\mathbf{W}^{(i+1)}$, using Eq. (17).
 - 9: **end for**
 - 10: **Output:** $\hat{\mathbf{M}} = \mathbf{M}^{(N_{\text{iter}})}$ and $\hat{\mathbf{W}} = \mathbf{W}^{(N_{\text{iter}})}$
-
-

ALGORITHM 2

Estimation of $\mathbf{M}^{(i+1)}$

- 1: **Input:** K , $\tilde{p}(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$
 - 2: **Initialization:** Set $v_r \leftarrow 3 \left[\sqrt{M/(\pi L)} \right] + 1$, and $v \leftarrow 4v_r$
 - 3: **for** $k = 1, \dots, K$ **do**
 - 4: Compute $\mathbf{m}_{n,k}^{(i+1)}$ the MAP estimate from $\tilde{p}(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$
 - 5: **for** $l = n - 1, \dots, 0$ **do**
 - 6: Compute $\mathbf{m}_{l,k}^{(i+1)}$ the MAP estimate at time l in frequency range $[\mathbf{m}_{l+1,k} - v, \mathbf{m}_{l+1,k} + v]$ from $\tilde{p}(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$
 - 7: **end for**
 - 8: **for** $l = n + 1, \dots, N - 1$ **do**
 - 9: Compute $\mathbf{m}_{l,k}^{(i+1)}$ the MAP estimate at time l in frequency range $[\mathbf{m}_{l-1,k} - v, \mathbf{m}_{l-1,k} + v]$ from $\tilde{p}(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$
 - 10: **end for**
 - 11: Update $\tilde{p}(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$ by setting to zero-value a neighborhood of size $[-v_r, v_r]$ around the k -th estimated ridge.
 - 12: **end for**
 - 13: **Output:** $\mathbf{M}^{(i+1)}$
-
-

D. IF Estimation

As previously discussed, the multimodal nature of $\tilde{p}(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$ challenges the parameters estimation. Thus, we estimate the ridges associated with the different modes one at a time, as explained in Algo. 2. In that algorithm, a local IF point estimate is computed through MAP estimation. Sequential MAP estimations are then obtained by moving backward in time. In order to constrain the estimates to correspond to the same ridge, the current MAP estimate is restricted to belong to an interval of radius v centered around the previous estimate. The same procedure is then iterated forward in time using the same starting point.

In the presence of multiple components, the approximate distribution $\tilde{p}(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$ needs to be updated after the estimation of each ridge (see line 11 in Algo. 2) to avoid estimating it several times. Similarly to [16], $\tilde{p}(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$ is set to zero in a frequency neighborhood of radius v_r around each estimated ridge. The procedure is then repeated until K ridges are estimated.

While both v and v_r could be defined as arbitrary user-defined hyperparameters, some assumptions have to be made to automatize ridge estimation. Setting v_r to three standard deviations (three-sigma rule of thumb) of the data distribution

$\sigma_d = \sqrt{M/(\pi L)}$ leads to proper discard when dealing with purely harmonic modes, therefore we set $v_r = \lceil 3\sigma_d \rceil$. Setting the value of v is a challenging task, as it requires finding a balance between two conflicting aspects. First, successive IF estimates have to be close to each other, suggesting the use of a small value for v to prevent sudden changes in the estimated ridge. Nonetheless, v needs to be greater than v_r since overlapping modes result in the creation of gaps in the region of overlap once one of the associated ridges has been estimated. In order to handle such situations, we set $v = 4v_r$, which enables the algorithm to jump over a removed region of overlap.

E. Amplitude Estimation

In this section, we propose an amplitude estimation process, performed after the computation of \hat{M} and \hat{W} in Algo. 1. Indeed, the relation between $w_{n,k}$, $\alpha_k(n)$ and the analysis window θ can be expressed as:

$$w_{n,k} = \frac{\alpha_k^2(n) \|F_\theta\|_2^2}{\sum_{k=1}^K \alpha_k^2(n) \|F_\theta\|_2^2 + Mb_n}, \quad (18)$$

with $\|F_\theta\|_2^2 = \frac{M}{2L\sqrt{\pi}}$ (for θ Gaussian) and b_n the average noise amplitude at time n . Nonetheless, the amplitude cannot be directly estimated from Eq. (18) due to the shape of the denominator. Setting $\bar{s}_n = \sum_{m=0}^{M-1} s_{n,m}$, it can be seen from Eq. (7) that:

$$E_{s_n | a_n, m_n, b_n} [\bar{s}_n] = \sum_{k=1}^K \alpha_k^2(n) \|F_\theta\|_2^2 + Mb_n, \quad (19)$$

so the denominator in Eq. (18) can thus be approximated by the expected values of \bar{s}_n . From Eq. (18) and (19), it directly follows that $\alpha_k^2(n) = \frac{w_{n,k} E[\bar{s}_n]}{\|F_\theta\|_2^2}$. This statement allows to perform amplitude estimation as $\alpha_k^2(n) = \frac{w_{n,k} \bar{s}_n}{\|F_\theta\|_2^2}$, by assuming \bar{s}_n to be a good approximation of $E[\bar{s}_n]$. However, the performance of the resulting estimator depends both on the quality of the approximation \bar{s}_n of $E[\bar{s}_n]$, and on that of \hat{W} . The final amplitude estimate thus reads:

$$\hat{\alpha}_k(n) = \sqrt{\frac{\hat{w}_{n,k} \bar{s}_n}{\|F_\theta\|_2^2}}. \quad (20)$$

F. Post-Processing

While the iterative procedure outlined in Section IV-D enables sequential IF estimation, the discarding step (see line 11 in Algo. 2) can be troublesome when components overlap in the TF plane. Indeed, by setting to zero the posterior values already assigned to a ridge, it becomes difficult to properly perform estimation of other ridges in the regions of overlap. In such a case, we would expect the estimation process to link distant local maxima in $\tilde{p}(M | \mathbf{W}^{(i)}, \mathcal{S})$ belonging to the same ridge, even if it is split during the discard step. Such an objective is challenging and would require complex estimation strategy, involving a significant increase in the computational cost. For this reason, we propose an additional step to get rid of this limitation. Once \hat{M} is computed, we detect the

regions of overlap. This is done by looking at the absolute difference between the positions of the ridges, assuming that the components overlap below a given threshold which is defined as three standard deviations of g (three-sigma rule of thumb). We then interpolate the remaining pieces (those that are far enough apart) using polynomial interpolation. While other types of interpolation such as spline interpolation could be used here, a thorough study of the different interpolation strategies is out of the scope of this paper. The advantage of this interpolation step can be observed in Fig. 2 where we compare the estimation performed with (left) and without (right) this step on a MCS made of two overlapping components. Note that both estimations have been achieved using EM method with a Laplacian prior ($\lambda = 10^{-2}$).

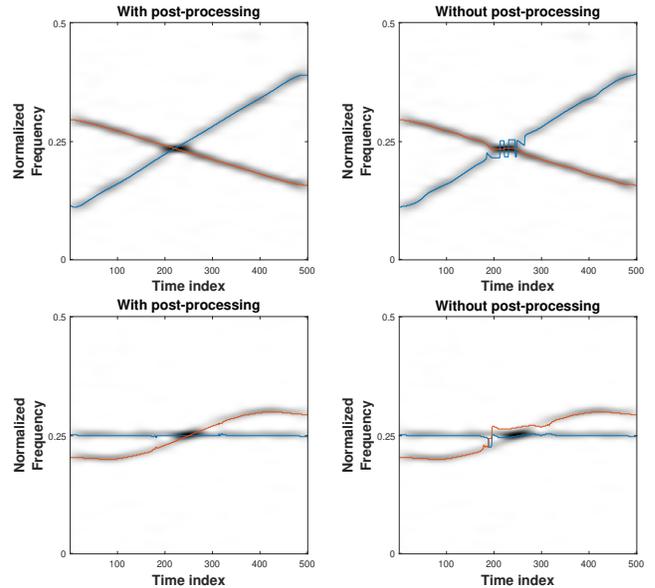


Fig. 2. Top: spectrograms of a MCS made of two overlapping linear chirps; bottom: spectrogram of a pure tone overlapping a FM chirp. Estimations performed using EM method (straight lines) with a Laplacian prior with a SNR = 10 dB (additive white Gaussian noise).

From Fig. 2 (right column), it can be observed that the second estimated ridge significantly oscillates around the crossing point, since the posterior probability does not provide any information in this region. On the left column of Fig. 2, the two IFs are well estimated.

V. RESULTS

In this section, we first assess the IF estimation performance of EM approach on MCSs either made of separated or crossing modes, in the presence of an additive white Gaussian noise¹. Then, we study in detail the estimation of the amplitude. A section is dedicated to the computational complexity of the proposed EM approach, while its performance is evaluated on two real-world signals, a bat echolocation call and a speech signal.

¹Matlab codes are freely available for the sake of reproducible research at: <https://github.com/QuentinLEGROS/TSP2023>.

A. Synthetic Data - Without Component Overlap

Here, we consider a MCS, made of a linear chirp and a FM component, whose spectrogram is depicted in Fig. 3. We now

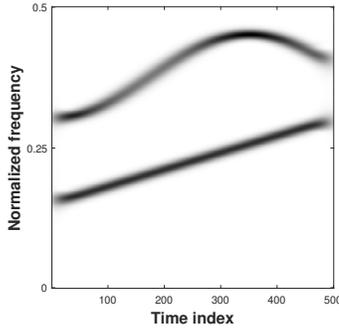


Fig. 3. Spectrogram of the analyzed multicomponent signal.

evaluate EM approach in terms of IF estimation performance by means of M and compare it with the following approaches.

1) *Pseudo-Bayesian (PB) method* [9]: The PB method sequentially estimates the position of the ridges at a time instant by computing a pseudo-posterior distribution based on a parametric divergence, and uses these positions to perform ridge detection at the following time instant. This method depends on two hyperparameters: the variance of the Gaussian random walk used to bound the IF derivative of the estimated modes, and the divergence parameter. The latter is denoted by β (resp. α), with values in range $[0, 1]$, if the β - (resp. α -) divergence is used to perform the estimation. To set the parameter values close to 0 improves the robustness of the estimation to the detriment of its accuracy at low noise levels. In all the experiments, we empirically set the variance of the Gaussian random walk to 2, which was shown to provide the best results.

2) *Ridge Detector (RD) algorithm* [13]: The RD algorithm aims at estimating the position of the ridges in the TF plane in the presence of strong noise. The method first estimates the ridge portions that are informative enough, before gathering them when they correspond to the same mode. The main parameters of the method are the frequency clearing window avoiding sequential estimations of the same ridge, and λ_s (resp. β_s) which constrains the first (resp. second) IF derivative. Moreover, we set $\lambda_s = 0.2$ and $\beta_s = 0.4$ as it provides with the best performance on the synthetic signals with low varying frequency used in our evaluations.

3) *Carmona method* [16], [31], [32]: This technique estimates the ridges position associated with the modes of a MCS in the TF plane by solving an optimization problem. More precisely, a cost function, involving a data term and a least-square constraint on the IF derivative with respect to time, is minimized to perform IF estimation. The method depends on three parameters: λ_b which controls the smoothness of the estimates (set to $\lambda_b = 10^{-2}$ as suggested by the authors and which empirically provides the best results), the clearing window size parameter, again corresponding to the parameter v_r introduced in Section IV-D, and a so-called jump parameter controlling the maximum interval between two successive points on a ridge. Unless stated otherwise, the experiments

presented in this section are performed using the same settings as in [16].

The estimation performance is measured using the relative mean squared error $RMSE = \frac{1}{NM^2} \sum_{k=1}^K \sum_{n=0}^{N-1} (\bar{m}_{n,k} - \hat{m}_{n,k})^2$, where $\bar{m}_{n,k}$ (resp. $\hat{m}_{n,k}$) is the actual (resp. estimated) normalized IF of the k -th component at the n -th time instant, projected onto the discrete frequency grid. As the EM method produces estimations on the spectrogram grid, it cannot be directly compared with methods like RD [13] that yield real-valued estimates. In the remainder of this paper, we thus project the estimates obtained with all the tested methods onto the closest TF bin along the frequency axis. It is worth noting that this does not penalize methods that provide real-valued estimates, as the ground truth IF is also quantized. In all the experiments performed on synthetic data, the term noise refers to complex additive white Gaussian noise.

The RMSEs obtained with the different tested methods are displayed in Fig. 4 for a varying SNR. It can be observed that

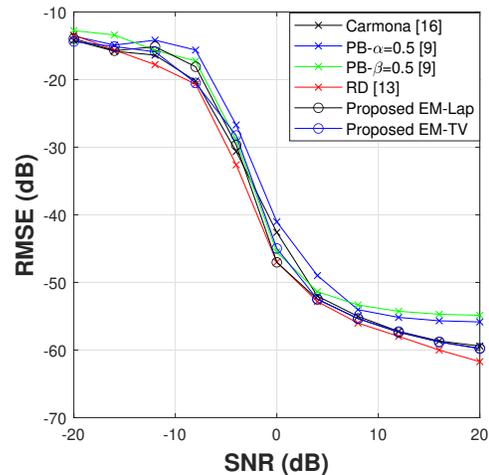


Fig. 4. RMSE of the normalized IF (averaged over 100 realizations of the noise) estimated from the MCS displayed in Fig. 3 obtained with the different competing methods for a varying SNR.

all the approaches behave similarly in terms of RMSE, except the PB methods that are outperformed at high SNR. Such a worse behavior is due to the hyperparameters choice, which improves the robustness of the PB method to the detriment of its efficiency at low noise levels [9], [33]. While RD performs slightly better than the other methods, EM using either prior models exhibits very similar performance. The next section is dedicated to the study of crossing modes scenarios.

B. Overlapping Components

Here, we evaluate the performance of the EM approach in estimating a MCS with overlapping modes in the TF plane. For comparison, we exclude the RD algorithm proposed in [13] since it explicitly assumes that the modes are separated in the TF plane.

A more relevant technique to study overlapping components is the so-called time-frequency-chirp-rate method [18], denoted by 3DRD in the sequel. This method first computes a three dimensional time-frequency and chirp-rate transform

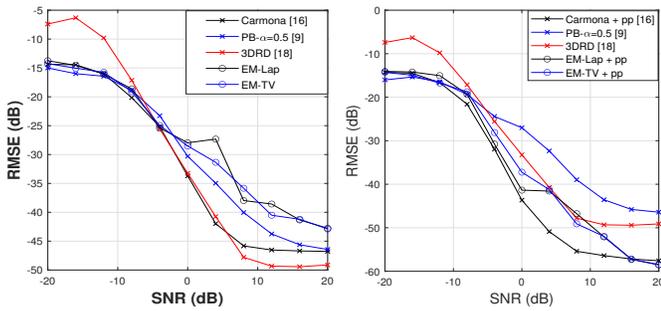


Fig. 5. Left: RMSEs of the normalized IF (averaged over 50 realizations of the noise) estimated from the MCS displayed in Fig. 2 (top) obtained with the different competing methods for a varying SNR; right: same as the left sub-figure but with the post-processing treatment.

of the observed signal, in which the axes of the transform respectively correspond to time, frequency and chirp-rate. In a second time, a tracking algorithm extracts the 3D ridges associated with each of the signal components. This approach can deal with components overlapping in the TF plane, since these components are perfectly separated in the time-frequency chirp-rate 3D space. This method relies on two parameters tuning the level of frequency and chirp-rate regularization, and on two others constraining the spatial variations of the IF estimates.

We also consider the Carmona method [31] (whose code made by Brevdo [16] can be found at <https://ebrevdo.github.io/>) setting its respectively so-called jump parameters and the clearing window size to v and v_r , for the sake of a fair comparison with the EM method (see Section. IV-D for the definition of v and v_r). In that method, we significantly decrease the value of λ_b compared with the previous example, to allow for higher variations of the ridges in the region of overlap (we set $\lambda_b = 10^{-5}$).

In the initial set of experiments, we assess the IF estimation performance of the different methods without resorting to the post-processing step. For that purpose, we consider the MCS (displayed in Fig. 2 (top)), whose components overlap in the TF plane. The obtained results are presented in Fig. 5 (left) for varying SNR. From this figure, we observe that all the methods perform similarly at low SNR, except for 3DRD that is outperformed by all the other tested methods. The similar behavior of all the tested methods at low SNR can be explained by the presence of strong noise around the overlapping region, which challenges the IF estimation since the signal information disappears as the noise level increases (around $\text{SNR} \approx -5$ dB). At high SNR, 3DRD performs the best since it provides smoother estimates than both EM and Carmona methods which exhibit oscillations in the region of overlap. The higher performance of Carmona method compared to EM is explained by the role of λ_b , which removes almost all the oscillations in the region of interest. Finally, even PB approach performs better than the proposed EM, since the Gaussian random walk prior enforces the smoothness of the estimates in the region of overlap.

An in-depth analysis of the suboptimal behavior obtained with the proposed EM technique suggests that the error

comes from strong oscillations in the region of overlap, which motivates the use of the post-processing step. Therefore, we conduct a second set of experiments in which we evaluate the effect of the post-processing step discussed in Section IV-F on the quality of IF estimation with the different techniques (denoted as ‘method + pp’). The results, displayed in Fig. 5 (right) for varying SNR, show that smoothing the estimates significantly improves the performance of both the EM method and Carmona approach, which now provide with the best performance for all SNRs. However, the Carmona method outperforms the proposed EM within the SNR range of $[-5, 10]$ dB.

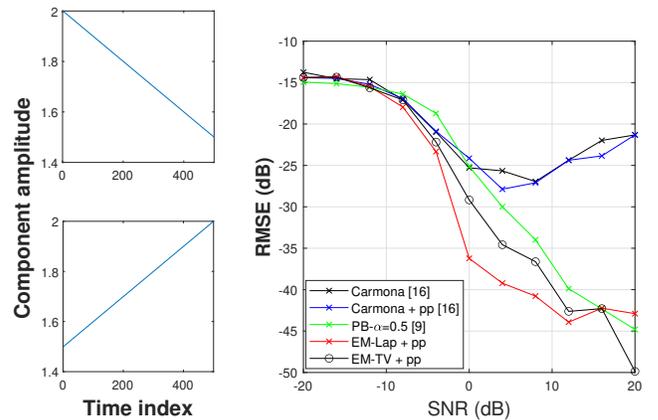


Fig. 6. Left: amplitude profiles assigned to the linear chirps of Fig. 2; right: RMSE of the normalized IF (averaged over 50 realizations of the noise) estimated from the MCS displayed in Fig. 2(top) with amplitude profiles displayed in the left sub-figures.

We aim to show that the effectiveness of the Carmona method decreases as the amplitudes of the modes fluctuate. This technique indeed focuses on maximizing the energy concentration on the ridge, rather than ensuring the smoothness of the components amplitude. In contrast, the observation model used in the EM technique guarantees the continuity of the estimated amplitude. To illustrate this, we assess the performance of various methods on amplitude modulated components, by replicating the experiment of Fig. 5 where an amplitude function is assigned to each component. More precisely, the amplitude function of Fig. 6 (top left) (resp. (bottom left)) is assigned to the red decreasing (resp. blue increasing) linear chirp in Fig. 2 (top). The results shown in Fig. 6 (right) highlight the limitations of Carmona method in handling overlapping components when their amplitudes change over time, unlike the EM technique which still performs well under such conditions. Except in the range $[-20, -7]$ dB, where all the methods perform similarly, the EM technique achieves the best performance at all SNRs. Although the PB method appears to be robust to amplitude modulation, it performs worse than the EM technique. In order to prevent misinterpretation of the results obtained with Carmona method, we present the results with and without post-processing in Fig. 6 (right). Regardless of whether a post-processing step is applied, the amplitude modulation significantly alters the IF estimation with that method, which hinders an efficient tracking of the ridges.

C. Amplitudes

We now assess the performance of the amplitude estimation method discussed in Section IV-E for the proposed EM approach. To this end, we introduce the relative mean absolute error $\text{RMAE} = \frac{1}{NK} \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} |\bar{\alpha}_k(n) - \hat{\alpha}_k(n)|$, where $\bar{\alpha}_k(n)$ (resp. $\hat{\alpha}_k(n)$) is the actual (resp. estimated) amplitude of the k -th component at time instant n . In our comparison, we use two other amplitude estimation methods. PB method proposed in [34] (denoted by PB), and the deterministic approach proposed in [12] (denoted by Local) where an amplitude estimator is derived assuming a local-log Gaussian approximation for the amplitude of the modes. The hyperparameters of the PB method are still set to $\alpha = 0.2$ and $\beta = 0.4$ as they provide the best trade-off between efficiency and robustness to outliers. As far as the EM method is concerned, since similar results are obtained using both prior models, we only display those obtained using the Laplacian prior model. Additionally, for each of the tested techniques, we associate an Oracle estimator, which involves estimating the amplitude with knowledge of the ground truth IF. These Oracles enable to assess the ability of the tested methods to estimate the amplitude without the bias potentially induced by an incorrect IF estimation. The Oracle methods are respectively named Oracle PB, Oracle Local and Oracle EM.

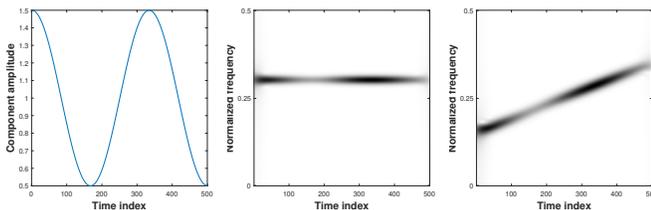


Fig. 7. Left: amplitude profile; middle: (resp. right:) spectrogram of a pure tone (linear chirp) with the amplitude profile given by the left subfigure

Our first goal is to investigate the influence of frequency modulation on amplitude estimation. To this end, we compute the RMAEs obtained with the tested methods, first for the pure tone signal whose spectrogram is displayed in Fig. 7 (middle), and also for the linear chirp displayed in Fig. 7 (right), the amplitude profile being displayed on the left-hand side of that figure.

In Fig. 8 (left), we observe a similar behavior for the different methods for a SNR higher than 0 dB, though the Local method performs slightly better at SNR above 15 dB. Conversely, this method is outperformed by the other approaches at SNR below 0 dB, as the constraints imposed by the signal model become too stringent. Furthermore, the slight differences between PB, EM and their Oracle versions illustrate the small effect incorrect IF estimation has on the amplitude estimation.

Though PB and EM methods appear to behave similarly in that case, the limitations of the former clearly arise when a frequency modulation is introduced in the mode, as illustrated in Fig. 8 (right). Indeed, when dealing with a linear chirp, the results obtained with EM and Local methods are similar to those observed in Fig. 8 (left), PB, as well as its Oracle

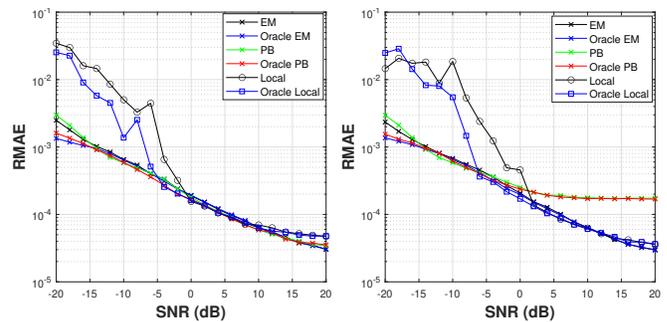


Fig. 8. RMAE of estimated amplitudes with the different techniques, averaged over 100 realizations of the noise, for the signal whose spectrogram is displayed in Fig. 7 (middle) (resp. Fig. 7 (right)) in the left-hand side (resp. right-hand side).

version, are less efficient for SNR above 0 dB. Despite frequency modulation is not considered in the models utilized in PB or EM, the superior performance of the EM method can be attributed to the method used for amplitude estimation. In the PB technique, the amplitude is estimated by computing the energy on the ridge and adjusting it with an estimation of the average noise level. However, when dealing with a frequency modulated mode, the energy on the ridge underestimates the amplitude of the mode.

Next, we assess the ability of EM method to estimate the amplitudes for the MCS associated with Fig. 3. In this MCS, the linear chirp component exhibits a linearly decreasing amplitude from 1 to 0.5 over the entire time range, and the mode with a sinusoidal frequency is assigned the sinusoidal amplitude depicted in Fig. 7 (left). The results obtained in that two-component case, displayed in Fig. 9 (left), are similar to those corresponding to mono-component signal except for the Local method which appears to be more sensitive to noise. Lastly, we examine the performance of the proposed method in a specific scenario involving two overlapping components depicted in Fig. 2 (top). Once again, each component is assigned an amplitude function such that the linear chirp with increasing (or decreasing) frequency has a linearly decreasing amplitude from 1 to 0.5 (or the amplitude profile of Fig. 7 (left)). The results in Fig. 9 (right) show that the presence of overlapping components does not reduce the accuracy of the amplitude estimation with the EM method, as the results are similar to those obtained in Fig. 9 (left).

D. Computational Time

Since the proposed algorithm depends on the number of components K , the number of iterations N_{iter} and the dimension of the observed TFR $M \times N$, we estimate the runtime complexity expressed in units of time using the “big O” notation as $\mathcal{O}(KN_{\text{iter}}NM^2)$.

Moreover, the computational times of the tested methods, when applied to the two-component signal shown in Fig. 3 with $N = 500$ is reported in Table II, for different frequency resolutions (indexed by M). All the experiments have been conducted using Matlab R2021b with an Intel(R) Xeon(R) W-2123 CPU @ 3.60GHz. The time needed to perform spectrograms computation is not included in the following

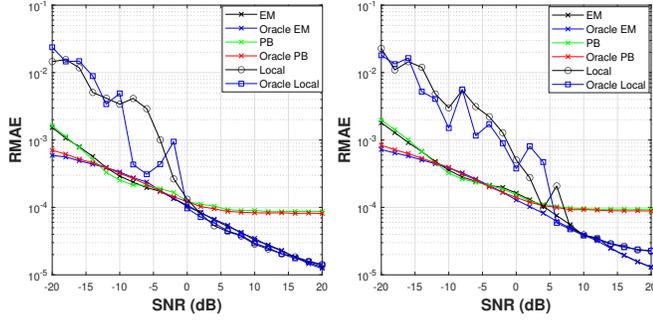


Fig. 9. RMAE of estimated amplitudes, averaged over 100 realizations of the noise, estimated from the MCS displayed in Fig. 3 (resp. Fig. 2 (top)) obtained with the different methods for a varying SNR in the left-hand side (resp. right-hand side).

comparison, which only gathers the execution time of the main algorithms. Despite the computational time required by the EM method is considerably longer than that of Carmona, RD, and PB, it still remains significantly shorter than that of 3DRD. Furthermore, it should be noted that this additional computational time allows for much more accurate IF and IA estimations, especially in the case of crossing modes, compared with the other tested methods.

TABLE II
COMPUTATIONAL TIME (IN SECONDS) OF THE COMPETING APPROACHES FOR SYNTHETIC DATA ANALYSIS, AVERAGED OVER 50 REALIZATIONS, FOR $N = 500$.

M	500	1000	2000
Carmona [16]	0.03	0.04	0.05
PB [9]	0.07	0.15	0.51
RD [13]	0.41	0.63	1.25
Proposed EM-Lap	4.89	18.73	40.30
Proposed EM-TV	4.90	16.74	42.46
TF chirp-rate [18]	30.17	67.61	5220

E. Simulation efficiency

We now evaluate the behavior of the method according to the number of Gibbs sampler iterations denoted N_s . We consider the MCS (displayed in Fig. 3), without amplitude modulation (amplitudes set to 1), and perform estimation of the model parameters using different values for N_s .

The obtained results using the TV prior with $\epsilon = 10^{-2}$ and $\epsilon = 1$ are displayed in Fig. 10 for varying SNR. From Fig. 10 (left), we observe that N_s has a low impact on the final estimation, while in Fig. 10 (right) the performance of the method depends on the choice of N_s at high SNR. The latter result is explained by the over-weighted prior model, giving too much emphasis to the prior and corrupting the estimation. These results highlight the difference between our approach and the classical SEM algorithm, as we are sampling from $p(\mathbf{M})$ and not from the conditional distribution $p(\mathbf{M}|\mathbf{W}^{(i)}, \mathbf{S})$. Numerous update of the prior thus leads a stronger regularization.

Both for estimation performance and computational cost, a small number of iterations is required to perform a satisfying estimation of the model parameters if no prior knowledge is available on the IF of the modes.

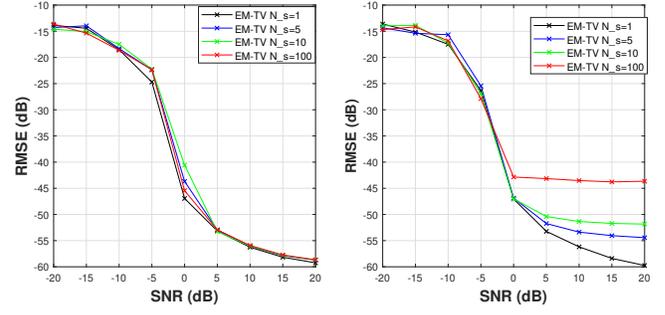


Fig. 10. RMSE of the normalized IF (averaged over 50 realizations of the noise) estimated from the MCS displayed in Fig. 3 with $\epsilon = 10^{-2}$ (resp. $\epsilon = 1$) on the left-hand side (resp. right-hand side) obtained with varying N_s and SNR.

F. Real-World Data

Here, we evaluate the performance of the proposed method on real-world signals. Since the signals used in this section do not contain overlapping modes, we are considering RD [13] for comparison with the EM method. Since the ground truth IF is not available, we compare the methods by assessing the results through informal observation. As in the previous experiments, we set $\lambda = 10^{-1}$ when using EM method, in order to enforce the regularization and limit the impact of outliers. First, we consider a speech signal of Japanese native speakers [35] whose spectrogram (STFT computed using $L = 40$) is displayed in Fig. 11. We look only for two components to enhance the readability of the results and to assess the behavior of our method in a scenario involving an underestimated number of components K . In Fig. 11, the IF estimates obtained using respectively the proposed EM-Lap and the estimation performed with RD method [13] are superimposed to the signal spectrogram.

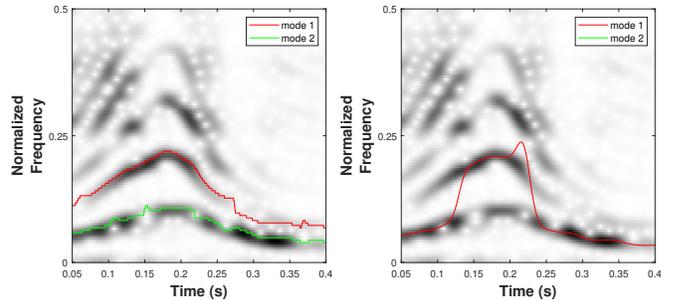


Fig. 11. Estimation of the first $K = 2$ signal components of the speech signal using the proposed EM-Lap method (left-hand side) and using the RD proposed in [13] (right-hand side).

One of the strengths of the EM method is its ability to handle the presence of gaps in the ridges, which can be observed, for example, on the lowest frequency component associated with the estimated mode 2 in Fig. 11. These gaps correspond to regions where the modes cannot be identified as a continuous chain of local maxima along the frequency axis, and are often caused by interferences between the signal and noise. On the contrary, RD method is not able to track the ridges when the chains of local maxima are broken. Moreover, RD method completely misses the second mode.

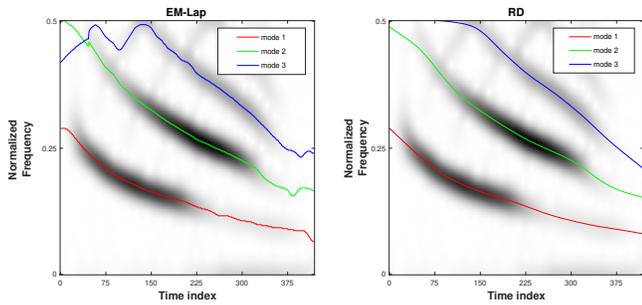


Fig. 12. Estimation of $K = 3$ signal components of the bat signal using the proposed EM-Lap method (left-hand side) and using the RD proposed in [13] (right-hand side).

We then conduct a test on an echolocation pulse signal emitted by an *Eptesicus Fuscus* bat [36]. We compare the results of the EM approach with those of RD and display both the spectrogram and estimations in Fig. 12. Our observations reveal that although both methods detect ridges and provide satisfactory IF estimations, the results are different for the two methods. Specifically, the estimates obtained using the EM approach (left-hand side) exhibit oscillations around the ridge positions, whereas those obtained using RD (right-hand side) provide smoother estimates that are centered around the expected IF of the modes. It is important to note that the RD approach does not accurately estimate mode 3, particularly around time index 100, where the associated ridge presents a strong frequency variation due to aliasing. In contrast, the EM approach captures this sharp transition.

VI. CONCLUSION

In conclusion, this study introduced a novel observation model for estimating the instantaneous frequency and amplitude of the modes of a multicomponent signal in the presence of noise. The proposed approach outperformed state-of-the-art methods in terms of both instantaneous frequency and amplitude estimation in the case of crossing modes and amplitude modulation. By leveraging EM algorithms, our method achieves a reduction in problem complexity and reasonable computational time for estimating mixture weights. Future work includes extending the approach to estimate chirp rate and generalizing the method to account for hyperparameter estimation.

REFERENCES

- [1] M. Liebling, T.-F. Bernhard, A. H. Bachmann, L. Froehly, T. Lasser, and M. Unser, "Continuous wavelet transform ridge extraction for spectral interferometry imaging," in *Coherence Domain Optical Methods and Optical Coherence Tomography in Biomedicine IX*, vol. 5690, pp. 397–402, International Society for Optics and Photonics, 2005.
- [2] S. Jo and C. D. Yoo, "Melody extraction from polyphonic audio based on particle filter," in *Proc. International Society for Music Information Retrieval (ISMIR'10)*, pp. 357–362, 2010.
- [3] R. H. Herrera, J. Han, and M. van der Baan, "Applications of the synchrosqueezing transform in seismic time-frequency analysis," *Geophysics*, vol. 79, no. 3, pp. V55–V64, 2014.
- [4] K. Abratkiewicz, P. Samczyński, and D. Fourer, "A comparison of the recursive and FFT-based reassignment methods in micro-doppler analysis," in *Proc. IEEE Radar Conference*, pp. 1–6, 2020.

- [5] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Lett.*, vol. 11, no. 2, pp. 112–114, 2004.
- [6] G. Rilling and P. Flandrin, "One or two frequencies? The empirical mode decomposition answers," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 85–95, 2007.
- [7] H. Hassani, "Singular spectrum analysis: methodology and comparison," *Journal of Data Science*, vol. 5, no. 2, pp. 239–257, 2007.
- [8] P. Flandrin, *Time-Frequency/Time-Scale Analysis. Wavelet analysis and its applications*, vol. 10. Academic Press, 1998.
- [9] Q. Legros and D. Fourer, "A novel pseudo-Bayesian approach for robust multi-ridge detection and mode retrieval," in *Proc. European Signal Processing Conference (EUSIPCO'21)*, Aug. 2021.
- [10] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.
- [11] D. Fourer, J. Harmouche, J. Schmitt, T. Oberlin, S. Meignen, F. Auger, and P. Flandrin, "The ASTRES toolbox for mode extraction of non-stationary multicomponent signals," in *Proc. European Signal Processing Conference (EUSIPCO'17)*, pp. 1130–1134, Aug. 2017.
- [12] D. Fourer, F. Auger, and G. Peeters, "Local AM/FM parameters estimation: application to sinusoidal modeling and blind audio source separation," *IEEE Signal Processing Lett.*, vol. 25, Oct. 2018.
- [13] N. Laurent and S. Meignen, "A novel ridge detector for nonstationary multicomponent signals: Development and application to robust mode retrieval," *IEEE Trans. Signal Process.*, vol. 69, pp. 3325–3336, 2021.
- [14] M. Lagrange, S. Marchand, and J.-B. Rault, "Enhancing the tracking of partials for the sinusoidal modeling of polyphonic sounds," *IEEE Trans. Audio, Speech, Language Processing*, vol. 15, no. 5, pp. 1625–1634, 2007.
- [15] K. Polissano, L. Condat, M. Clausel, and V. Perrier, "A convex approach to superresolution and regularization of lines in images," *SIAM J. Imaging Sci.*, vol. 12, no. 1, pp. 211–258, 2019.
- [16] E. Brevdo, N. S. Fuckar, G. Thakur, and H.-T. Wu, "The synchrosqueezing algorithm: a robust analysis tool for signals with time-varying spectrum," *Computing Research Repository - CORR*, 01 2011.
- [17] I. Daubechies and S. Maes, "A nonlinear squeezing of the continuous wavelet transform," *Wavelets in Medicine and Bio.*, pp. 527–546, 1996.
- [18] X. Zhu, H. Yang, Z. Zhang, J. Gao, and N. Liu, "Frequency-chirp rate reassignment," *Digital Signal Processing*, vol. 104, p. 102783, 2020.
- [19] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*, vol. 382. John Wiley & Sons, 2007.
- [20] G. Celeux, D. Chauveau, and J. Diebolt, "Stochastic versions of the EM algorithm: an experimental study in the mixture case," *Journal of Stat. Comp. and Simulation*, vol. 55, no. 4, pp. 287–314, 1996.
- [21] V. Susic, N. Saulig, and B. Boashash, "Estimating the number of components of a multicomponent nonstationary signal using the short-term time-frequency rényi entropy," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, pp. 1–11, 2011.
- [22] N. Saulig, N. Pustelnik, P. Borgnat, P. Flandrin, and V. Susic, "Instantaneous counting of components in nonstationary signals," in *Proc. European Signal Processing Conference (EUSIPCO'13)*, pp. 1–5, 2013.
- [23] 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.
- [24] D. Iatsenko, P. McClintock, and A. Stefanovska, "Extraction of instantaneous frequencies from ridges in time-frequency representations of signals," *Signal Processing*, vol. 125, pp. 290–303, 2016.
- [25] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical imaging and vision*, vol. 20, no. 1, pp. 89–97, 2004.
- [26] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [27] X. Wang, "Laplacian operator-based edge detectors," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 29, pp. 886–90, June 2007.
- [28] M. Meyer, M. Desbrun, P. Schroder, and A. H. Barr, "Discrete differential-geometry operators for triangulated 2-manifolds," in *Visualization and Math. III*, pp. 35–57, Springer Berlin Heidelberg, 2003.
- [29] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
- [30] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognition*, vol. 36, no. 1, pp. 131 – 144, 2003.

- [31] R. A. Carmona, W. L. Hwang, and B. Torrèsani, "Characterization of signals by the ridges of their wavelet transforms," *IEEE Trans. Signal Process.*, vol. 45, no. 10, pp. 2586–2590, 1997.
- [32] R. A. Carmona, W. L. Hwang, and B. Torrèsani, "Multiridge detection and time-frequency reconstruction," *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 480–492, 1999.
- [33] A. Cichocki, S. Cruces, and S.-I. Amari, "Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization," *Entropy*, vol. 13, no. 1, pp. 134–170, 2011.
- [34] Q. Legros and D. Fourer, "Pseudo-Bayesian approach for robust mode detection and extraction based on the STFT," *Sensors*, vol. 23, no. 1, p. 85, 2022.
- [35] D. Fourer, T. Shochi, J.-L. Rouas, and A. Riiliard, "Perception and manipulation of Japanese attitudes," in *Proc. Speech Prosody (SP'16)*, June 2016.
- [36] R. Baraniuk, "Beckman bat pulse." <https://www.ece.rice.edu/dsp/software/bat.shtml>, 2017.



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. In 2019, he joined the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina) as an Assistant Researcher, becoming an Adjunct Researcher in 2022. In 2021, he was a Fulbright Visiting Scholar at Duke University, USA; and he visited the Université Grenoble-Alpes. His research interests are time-frequency/time-scale signal analysis, synchrosqueezing, data-driven methods and biomedical signal processing.



Quentin Legros was born in Orange, France, in 1992. He received a Licence degree in mathematics in 2015 from Avignon University and a Master's degree in applied mathematics in 2017 from University of Grenoble Alpes. He obtained his Ph.D. degree in electrical engineering in 2021 from Heriot-Watt University for a study on sequential approaches for adaptive image acquisition and processing. He is currently with the signal team within the PRISME laboratory at Polytech Orléans (France). His interests include inverse problems, time-frequency analysis,

and Bayesian inference.



Dominique Fourer (M'11) received the M.Sc degree in Computer Sciences in 2008 and the M.Sc degree in Applied Mathematics in 2009, both from University of Bordeaux 1, France. He defended a Ph.D thesis in the field of audio signal processing at LABRI Laboratory of Bordeaux in 2013. Following his Ph.D, he served as a research fellow in the sound analysis-synthesis team at the Institute of Research and Coordination Acoustic-Music (IRCAM) in Paris until 2018. Currently, he holds a tenured position as associate professor at the IBISC laboratory in

the University of Evry Paris-Saclay. His research focuses on time-frequency analysis theory and applications in digital speech, music and audio signal processing. In addition to his academic pursuits, he is an accomplished jazz musician (pianist) actively participating in various music bands and leading his own trio in the Paris area.



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 at Grenoble institute of technology. 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 Laboratory of the university of

Edinburgh, U.K., in 2011, and at the university of Pisa, Italy, in 2018.