A greedy algorithm for the estimation of instantaneous frequencies in the time-frequency plane

Sajjad KHODAVERDI Jean-Baptiste COURBOT Ali MOUKADEM
IRIMAS UR 7499, 61 rue Albert Camus, Mulhouse, France

Résumé – Dans cette contribution, nous proposons d'étudier l'estimation de modes dans le plan temps-fréquence en adoptant un modèle paramétrique. Le problème inverse à résoudre est celui de l'estimation du nombre, et des paramètres, des composantes. Pour le résoudre, nous proposons un algorithme glouton qui ajoute de manière itérative des composantes, tout en effectuant une optimisation *coarse-to-fine* à chaque étape. Les résultats numériques permettent d'illustrer les performances de la méthode proposée, pour la restitution des composantes comme des fréquences instantanées.

Abstract – In this paper, we propose to study mode recovery in the time-frequency plane from a parametric perspective. Formulating the inverse problem to solve, we seek to retrieve the number of components as well as their parameter. To that end, we propose a greedy algorithm that iteratively add new components to the solution, while providing at each step a coarse-to-fine optimization to avoid local minima. Numerical results allow assessing the performance of our method, with respect to component and instantaneous frequency retrieval.

1 Introduction

1.1 Context and problems

Most of the signals we encounter in nature can be modeled by a multi-component harmonic model, which consists of superimposing signals modulated in amplitude and frequency [9]. This is for instance the case of audio sounds (music, speech), physiological signals (phonocardiogram, electrocardiogram), gravitational waves, just to name a few. Nonstationary, in the sense of frequency variability as a function of time, is therefore an inherent property of these signals. Estimating and tracking these components as a function of time has been an issue in signal processing for several decades. Estimating these components will enable us to better understand the signals we measure, and also to obtain compact (sparse) time-frequency representations that are as faithful as possible to the physical phenomena that generate them.

The standard method for estimating signal components is the ridge estimation from the time–frequency representation [6]. The ridge method is based on the detection of local maxima at each time t of time–frequency representation, which enables the tracking of signal components over time. Several variants of the ridge methods attempt to optimize and refine detection by designing a robust peak detector [12]. However, these approaches reach their limits when the components are very close, which will create interference in the time-frequency plane, and when the components cross, and also in the presence of high noise levels [13]. Hence, the interest in exploring other approaches to estimating the components of non-stationary signals, since it remains an open question, which is the main aim of this paper. In this work, we propose to approach the component retrieval by posing an optimization problem that estimate the optimal parameters of the components that best fit a given observation, *i.e.* the time-frequency representation.

1.2 Prior works

Recently, a new approach was introduced based on sparse inverse problem to estimate linear chirps in time–frequency plane [14]. Authors propose a gridless approach by using the Sliding Frank-Wolfe [7] algorithm to solve the problem. Compared with conventional methods based on grid-based ridge estimation, this approach enabled robust estimation of linear chirps even in the presence of very high noise levels. However, there remains the question of more complex signals, which can contains one or more components that are not linearly modulated in time and frequency.

Taking a step back, the problem can be related to curve estimation in image, which has been addressed in a variety of contexts (*e.g.*, remote sensing [3] or vascular imaging [11]). However, the formulation in the TF plane, while sharing some aspects with these problems (one can, *e.g.*, write an inverse problem that describe the optimal parameter set), there are some peculiarities of the TF variant.

In the same vein as [14], in this work, we are interested in the detection of ridges in the time-frequency plane using a parametric approach that consists in finding the signal components and their parameters by solving an inverse problem.

2 Methods

2.1 Inverse problem and criterion

At first, we define the assumptions made about the signals of interest, as well as the criterion we seek to minimize. We assume the observed signal $x(t)$ is of the form:

$$x(t) = \sum_{n=1}^{N} c(t, \boldsymbol{\theta}_n) + \epsilon, \tag{1}$$

with ϵ an i.i.d. Gaussian noise, and $c(t, \boldsymbol{\theta})$ the chirp determined by its parameters $\boldsymbol{\theta}$. For the sake of prototyping

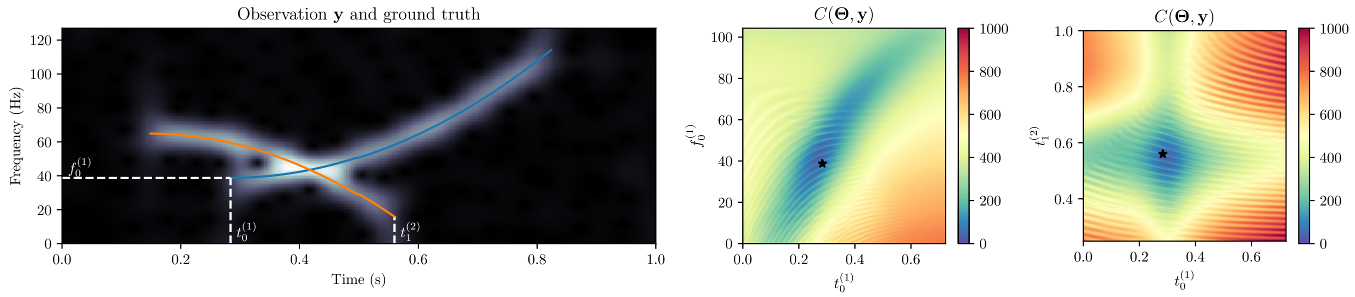

Figure 1 – Toy case depiction of the problem considered in this paper. Left panel: representation of $\mathbf{y} = |V_g x|$ in the case of $N = 2$ components, whose instantaneous frequencies are in color (blue for component C1, orange for component C2). Middle panel: depiction of the values taken by $C(\mathbf{\Theta}, \mathbf{y})$ when moving the left extremity of C1, $(t_0^{(1)}, f_0^{(1)})$. Right panel: values of $C(\mathbf{\Theta}, \mathbf{y})$ when varying the starting time of C1, noted $t_0^{(1)}$, together with the ending time of C2, noted $t_1^{(2)}$, (right). In both middle and right panels, all other values of $\mathbf{\Theta}$ fixed remain fixed, and the star depicts the true parameters location.

the optimization method, we focus on quadratic chirps, so $\boldsymbol{\theta} \overset{\text{def.}}{=} \{a, t_0, f_0, t_1, f_1\}$ such that a is its intensity and its extremities are in time are (t_0, t_1) and in frequency (f_0, f_1).

$$c(\boldsymbol{\theta}, t) = a \cos\left(2\pi t \left(f_0 + \frac{f_1 - f_0}{(t_1 - t_0)^2} t^2\right)\right) \mathbb{1}_{\{t_0 < t < t_1\}}. \quad (2)$$

Then, the problem we handle is the retrieval of the number N and parameters $\boldsymbol{\theta}$ of the components within x.

We study the casting and inversion of (1) through time-frequency representations, focusing here on the Short Time Fourier Transform (STFT), defined as

$$V_g x(\tau, f) = \langle x, \mathbf{M}_f \mathbf{T}_\tau g \rangle = \int_{\mathbb{R}} x(t) \overline{g(t - \tau)} e^{-2\pi i t f} \mathrm{d}t, \quad (3)$$

i.e., the STFT locates the signal in the plane at time τ and frequency f following a modulation by f and a translation by τ of the signal seen by a Gaussian windows g. The STFT being additive, one can recast (1) as:

$$V_g x = \sum_{n=1}^{N} V_g c(\boldsymbol{\theta}) + V_g \epsilon. \quad (4)$$

We denote $\mathbf{y} = |V_g x|$ the modulus of the STFT of x after sampling (i.e., in the discretized TF plane), and the set of parameters piloting x as $\mathbf{\Theta} = \{\boldsymbol{\theta}_n\}_{n=1}^{N}$. Then, the forward operator is $\Phi(\mathbf{\Theta}) = \left|\sum_{n=1}^{N} V_g c(\boldsymbol{\theta})\right|$, *i.e.*, $\Phi(\mathbf{\Theta})$ is the modulus of the STFT of the signal span by $\mathbf{\Theta} = \{\boldsymbol{\theta}_n\}_{n=1}^{N}$. We can then formulate the retrieval of the optimal parameter set as the minimization of:

$$C(\mathbf{\Theta}, \mathbf{y}) = \|\Phi(\mathbf{\Theta}) - \mathbf{y}\|_2^2. \quad (5)$$

Remark. This criterion is similar in formulation to other ℓ_2-minimization problems, but the following points need to be acknowledged:

- (5) assumes implicitly, through ℓ_2 minimization, an additive i.i.d. Gaussian noise *in the TF plane*. This is a convenient assumption that remains inexact, see the numerous studies on the nature of the STFT of white noise [8, 1].
- Φ is not a linear operator, as the modulus operates outside the sum ; in other words $\Phi(\mathbf{\Theta}_1) + \Phi(\mathbf{\Theta}_2) \neq \Phi(\mathbf{\Theta}_1 \cup \mathbf{\Theta}_2)$. This is in fact due to interferences between components of the signal.

2.2 Estimation

Minimizing (5) is not a trivial problem: the number of component N is unknown, and even for a known N, C possesses several local and global minima. This is depicted in Fig. 1, which shows that even for a toy example, the criterion exhibits the effects of interferences. This tormented landscape makes optimization a difficult task, which is why we adopt the two following ideas:

1. Greediness, in order to add components one by one to the solution. Each addition delimits a new iteration of the algorithm, until no new component can be found.
2. Coarse-to-fine optimization: in order to avoid local minima, we seek solutions in an extended neighborhood at first, then locally and more finely. Besides, we do not assume the estimations performed earlier are definitive, so they are brought back into optimization at each step.

So, the algorithm we propose is based on the repetition of three main steps.

1) *Addition of new component.* From the residual $\mathbf{r}(\mathbf{y}, \hat{\mathbf{\Theta}}) \overset{\text{def.}}{=} \mathbf{y} - \Phi(\hat{\mathbf{\Theta}})$, we select the largest connected components in the set $\left\{\tau, f : \mathbf{r}\big(\mathbf{y}(\tau, f), \hat{\mathbf{\Theta}}\big) > \mathrm{mean}(\mathbf{y}) + \mathrm{std}(\mathbf{y})\right\}$. The boundaries in time and frequency of the component provides an initial estimate of the new $\boldsymbol{\theta}$ added to the solution. Note that the sought-after solution is bounded such that $|t_1 - t_0| > \Delta_t$, *i.e.*, there is a lower bound on the duration of the component. If no component long enough is found, the algorithm is stopped.

2) *Coarse optimization.* We need to first explore a large panel of solution, possibly belonging to different basins of attraction (see Fig. 1). To do so, we perform random walks based on the Metropolis-Hastings (MH) algorithm [4], which allows a cheap and large domain exploration. We choose to run Q parallel MH chain and to retain the one providing the lower value of C. As we do not deal with convergence, a fixed number of MH step and variance of the perturbation (1/10 of the size of the parameter space) are used.

3) *Fine optimization.* From the best MH result, we finally run the L-BFGS-B algorithm [2], which allows managing efficiently both boundaries in the parameter space and computational resources. At this step, the gradient is numerically approximated.

The algorithm stops either when it is unable to find new

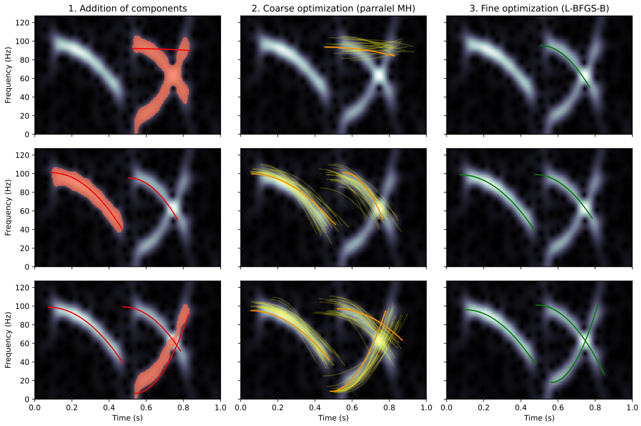

Figure 2 – Step-by-step depiction of the proposed algorithm. Each line corresponds to a new iteration, while the columns depicts the steps described in Alg. 1. In the left panels, the light red region is the largest connected component found in the residual, and in the middle panels the transparent curves are other outcomes of MH that were not selected at step 2. The algorithm stops as it cannot found long enough connected components after 3 iterations (not shown).

components, or when C does not diminish between iterations. The overall procedure is summarized in Alg. 1, and Fig. 2 depicts the stepwise course of the algorithm.

Algorithm 1: Proposed algorithm

```
Require: $\mathbf{y} = |V_g x|$, minimum time duration $\Delta_t$,
    number of parallel MH chain $Q$.
Ensure: $\hat{\boldsymbol{\Theta}}$
    do (iteration $n$):
        1. Add a new components from $\mathbf{r}(\mathbf{y}, \hat{\boldsymbol{\Theta}}_{n-1})$
        If no component last longer than $\Delta_t$: stop
        2. Coarse optimisation ($Q$ parallel MH chains).
        3. Fine optimization (L-BFGS-B), yielding $\hat{\boldsymbol{\Theta}}_n$.
    While $C(\hat{\boldsymbol{\Theta}}_n, \mathbf{y}) < C(\hat{\boldsymbol{\Theta}}_{n-1}, \mathbf{y})$
```

Remark. The proposed algorithm shares some aspects with existing methods in the literature, namely:

- Matching pursuit and extensions for component unmixing. We however do not rely on a dictionary of $\boldsymbol{\theta}$ to perform inference, and that would be suboptimal in our context, as there is no reason for the $\boldsymbol{\theta}$ to live on a predefined grid.
- Sliding Frank-Wolfe [7] regarding the iterative and continuous aspect of our optimization. However, we do not enforce sparsity: it would require the setting of some regularization parameter, whose choice can itself be the object of a homotopy method, which in turns need to be problem-tailored, as in [5].

3 Numerical results

We now propose an evaluation study for Alg. 1. Sampling random $\boldsymbol{\theta}$ parameters and ensuring the resulting quadratic chirps cross two by two, we vary the number of components N and the signal-to-noise ratio (SNR), defined as $20\log(\overline{x}/\overline{\epsilon})$. We are then interested in evaluating our ability to:

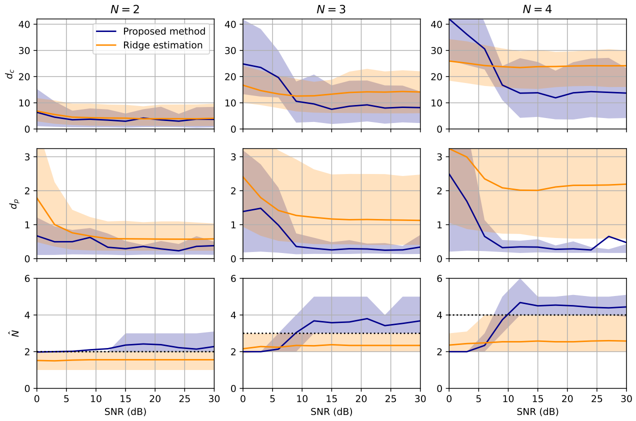

Figure 3 – Summary of experiments. From top to bottom are represented the distance between estimation and ground truth according to d_c (7), to d_p (6), and the estimated number of components $\hat{N}$. Experiments are repeated over 50 random sampling of chirps (see Fig. 4) and the solid lines represent the average, while the area depicts the boundary between the first and 9th decile.

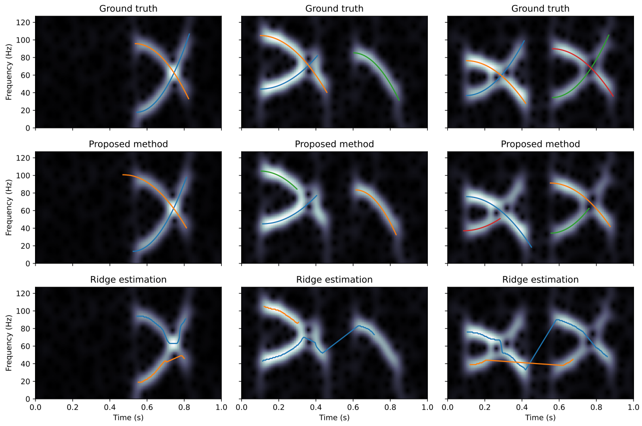

Figure 4 – Example of results obtained at SNR = 15 dB. The background corresponds to $\mathbf{y} = |Vx|$, and the results corresponds from left to right to $N = 2$, $N = 3$ and $N = 4$.

- accurately find local maxima at any given time, *i.e.*, estimate instantaneous frequencies correctly,
- recover N accurately,
- attribute the local maxima to the correct component.

We evaluate the quality of estimation through their curve representation in the TF plane, so we propose to use specific distances to that end. This representation is assumed to be discretized for all time step at which it exists.

At first, we define a *pointwise* distance between a reference curve $\mathfrak{r}$ and an estimated curve $\boldsymbol{e}$ as:

$$d_p(\mathfrak{r}, \boldsymbol{e}) = \sum_t \min_{t'} \|\mathfrak{r}[t] - \boldsymbol{e}[t']\| \,, \tag{6}$$

with t and t' discrete time indexes. Then, we can define a distance between set of curves. Let us assume we have a set of reference curves $\mathcal{R} = \{\mathfrak{r}_k\}_{k=1}^K$ and a set of estimations $\mathcal{E} = \{e_l\}_{l=1}^L$, then the component-wise distance between $\mathcal{R}$ and $\mathcal{E}$ is:

$$d_c(\mathcal{R}, \mathcal{E}) = \sum_{k=1}^{K} \min_{1\leq l\leq L} d_p(\mathfrak{r}_k, e_l). \tag{7}$$

In other words, from a curve in the reference set, one first find its match in the estimation in terms of pointwise distance, and this contributes to the distance between sets.

Alternatively, we consider a pointwise distance between sets that is indifferent to curve attribution, *i.e.* computing d_p (6) assuming all points in $\mathscr{R}$ and $\mathscr{E}$ belong to the same component. These distances will help us assert the quality of frequency retrieval, regardless of component attribution.

We compare our method with the *ridge tracking* (RT) proposed in [10] as a baseline. This method requires a prior knowledge of the number of component to seek, and provide a frequency estimate per component at *all* time step. To properly delimit components, we threshold them above the spectrogram mean, thus yielding possibly fewer components than expected.

Numerical results are summarized in Fig. 3 and some examples are depicted in Fig. 4. First observations yield an expected lower quality results with lower SNR, or with more components. More generally, we make the following observations:

- the RT method identifies quite well instantaneous frequencies, at the cost of label switching: sets of instantaneous frequencies are sometimes gathered together despite being separate in $|V_g x|$.
- the method we propose is by comparison more able to follow components, but remains affected by negative interferences (chirps are "too short" when crossing).
- the RT method most often underestimates the number of component, while our method generally overestimate it: in some cases, two chirp crossing can be seen as two, three or four chirps.
- overall, our proposed method yields less variability with respect to the proposed metrics, and seems more robust to both noise and the increasing of N.

4 Discussion

In this paper, we modeled the instantaneous frequency estimation problem as a parametric estimation problem, and proposed a greedy approach to solve it. While this work is still preliminary (*e.g.*, the impact of the algorithm parameter has to be thoroughly assessed), it has shed light on several aspects of the problem. We have shown that even in a simple case, component recovery can be surprisingly difficult, due to non-linearities induced by interferences. In that regard, minimizing a ℓ_2 norm, while being straightforward, is perhaps not the best choice. Considering the relation between the STFT and random fields (see, *e.g.*, [8] describing the covariance structure of the STFT), a specific norm accounting, *e.g.*, for the covariance structure of the STFT might be promising. We also aim at comparing, and using, more recent ridge tracking method either for comparison or as part of the optimization process.

Notably, the parametric model and estimation allows providing, for any component, an off-the grid estimation of the instantaneous frequency. In other words, we do not depend on the grid used in the TF plane, as long as the transform and its parameter allows for a sufficient resolution. This aspect of the problem is in link with the main step forward regarding the formulation of the problem, that can be recast as an instance of a sparse off-the-grid inverse problem, benefiting from relevant algorithms [7] and recovery guarantees.

References

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

[2] Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208, 1995.

[3] Ziyi Chen, Liai Deng, Yuhua Luo, Dilong Li, José Marcato Junior, Wesley Nunes Gonçalves, Abdul Awal Md Nurunnabi, Jonathan Li, Cheng Wang, and Deren Li. Road extraction in remote sensing data: A survey. *International journal of applied earth observation and geoinformation*, 112:102833, 2022.

[4] Siddhartha Chib and Edward Greenberg. Understanding the Metropolis-Hastings algorithm. *The american statistician*, 49(4):327–335, 1995.

[5] Jean-Baptiste Courbot, Ali Moukadem, Bruno Colicchio, and Alain Dieterlen. Sparse off-the-grid computation of the zeros of STFT. *IEEE Signal Processing Letters*, 30:788–792, 2023.

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

[7] Quentin Denoyelle, Vincent Duval, Gabriel Peyré, and Emmanuel Soubies. The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy. *Inverse Problems*, 36(1):014001, 2019.

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

[9] Patrick Flandrin. *Explorations in Time-Frequency Analysis*. Cambridge University Press, 2018.

[10] Dmytro Iatsenko, Peter VE McClintock, and Aneta Stefanovska. Extraction of instantaneous frequencies from ridges in time–frequency representations of signals. *Signal Processing*, 125:290–303, 2016.

[11] Chetan L Srinidhi, P Aparna, and Jeny Rajan. Recent advancements in retinal vessel segmentation. *Journal of medical systems*, 41:1–22, 2017.

[12] Nils Laurent and Sylvain Meignen. A novel ridge detector for nonstationary multicomponent signals: Development and application to robust mode retrieval. *IEEE Transactions on Signal Processing*, 69:3325–3336, 2021.

[13] Gi-Ren Liu, Yuan-Chung Sheu, and Hau-Tieng Wu. Analyzing scalogram ridges in the presence of noise. *arXiv preprint arXiv:2501.00270*, 2024.

[14] Kévin Polisano, Basile Dubois-Bonnaire, and Sylvain Meignen. Gridless 2D recovery of lines using the Sliding Frank-Wolfe algorithm. In *2024 32nd European Signal Processing Conference (EUSIPCO)*, pages 2697–2701. IEEE, 2024.