Convex Optimization approach to signals with fast varying instantaneous frequency

Matthieu Kowalski, Adrien Meynard, Hau-Tieng Wu

Convex Optimization approach to signals with fast varying instantaneous frequency

Matthieu Kowalski[a,b], Adrien Meynard[a], Hau-tieng Wu[c]

[a] *Laboratoire des Signaux et Systèmes – Univ Paris-Sud – CNRS – CentraleSupelec*
[b] *Parietal project-team, INRIA, Neurospin, CEA-Saclay, France*
[c] *Department of Mathematics, University of Toronto, Toronto, Ontario, Canada*

Abstract

Motivated by the oscillatory signals with components with fast-varying instantaneous frequency, we approach the time-frequency analysis problem by optimization. Based on the proposed adaptive harmonic model, the time-frequency representation of a signal is obtained by directly minimizing a functional, which involves few properties an ideal time-frequency representation should satisfy, for example, the signal reconstruction and concentrative time frequency representation. FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) is applied to achieve an efficient numerical approximation of the functional. We coin the algorithm as *Time-frequency bY COnvex OptimizatioN* (Tycoon). The numerical results confirm the potential of the Tycoon algorithm.

Keywords: Time-frequency analysis, Convex optimization, FISTA, Instantaneous frequency, Chirp factor

1. Introduction

Extracting proper features from the collected dataset is the first step toward data analysis. Take an oscillatory signal as an example. We might ask how many oscillatory components inside the signal, how fast each component oscillates, how strong each component is, etc. Traditionally, Fourier transform is commonly applied to answer this question. However, it has been well known for a long time that when the signal is not composed of harmonic functions, then Fourier transform might not perform correctly. Specifically, when the signal satisfies $f(t) = \sum_{k=1}^{K} A_k(t)\cos(2\pi\phi_k(t))$, where $K \in \mathbb{N}$, $A_k(t) > 0$ and $\phi_k'(t) > 0$ but $A_k(t)$ and $\phi_k'(t)$ are not constants, the momentary behavior of the oscillation can not be captured by the Fourier transform. A lot of efforts have been made in the past few decades to handle this problem. Time frequency analysis based on different principals [15] has attracted a lot of attention in the field and many variations are available. Well known examples include short time Fourier transform (STFT), continuous wavelet transform (CWT), Wigner-Ville distribution (WVD), chirplet transform [29], S-transform [33], etc. Empirical mode decomposition [20] and ensemble empirical

Email addresses: matthieu.kowalski@lss.supelec.fr (Matthieu Kowalski), hauwu@math.toronto.edu (Hau-tieng Wu)

mode decomposition [44] were proposed as alternative solutions aiming to alleviate the shortage of these analyses; however, its mathematical foundation is still lacking at this moment and several numerical limitations cannot be ignored. The sparsity approach [34], approximation approach [9] and time-varying autoregression and moving average approach [13] are also proposed in the literature, to name but a few.

Among these approaches, the reassignment technique [23, 2, 6] has attracted more and more attention in the past few years. The main motivation of the reassignment technique is to improve the resolution issue introduced by the Heisenberg principal. Precisely, the STFT coefficients are reallocated in *both* frequency axis and time axis according to their local phase information, which leads to the reassignment technique. The same reassignment idea can be applied to a very general settings like Cohen's class, affine class, etc [16]. It has been shown that the reassignment technique shares the same flavor as the non-local mean technique commonly applied in image processing [17]. Among several possible reassignment techniques [2, 6, 1], synchrosqueezing transform (SST) is a special reassignment technique. To be more precise, in SST, the STFT or CWT coefficients are reassigned *only* on the frequency axis [12, 11, 40, 30, 35, 41, 7] so that the causality is preserved and hence a real time algorithm is possible [8]. As a special reassignment technique, the same trick could be applied to different time-frequency representation. For example, the SST based on wave packet transform is recently considered in [45]. After that, SST is applied to different fields, ranging from medicine [21, 28, 43, 42, 3, 27], mechanics [24, 14], finance [18, 38], high energy physics [25, 32] to geographics [19, 39].

As useful as the reassignment and the SST approach, they are limited to the analysis of signals of "slowly varying instantaneous frequency". Precisely, the conditions $|A_k'(t)| \leq \epsilon\phi_k'(t)$ and $|\phi_k''(t)| \leq \epsilon\phi_k'(t)$ are essential if we want to study the model $f(t) = \sum_{k=1}^{K} A_k(t)\cos(2\pi\phi_k(t))$ by the current proposed reassignment and SST techniques. However, the oscillatory signals with fast varying instantaneous frequency is commonly encountered in practice, for example, the chirp signal generated by bird's song, bat's vocalization and wolf's howl, the uterine electromyogram signal, the heart rate time series of a subject with atrial fibrillation, to name but a few. Thus, finding a way to study this kind of signal is fundamentally important in data analysis. In this paper, based on previous works and the goal of having a time-frequency representation, we would consider an optimization approach to study the signals with fast varying instantaneous frequency. In brief, based on the relationship among the oscillatory components, the reconstruction property and the sparsity requirement on the time-frequency representation, we suggest to evaluate the optimal time-frequency representation F by optimizing the following functional

$$\begin{aligned}\mathcal{H}(F,G) := &\int \left|\Re\int F(t,\omega)\mathrm{d}\omega - f(t)\right|^2 \mathrm{d}t \\ &+ \mu\iint |\partial_t F(t,\omega) - i2\pi\omega F(t,\omega) + G(t,\omega)\partial_\omega F(t,\omega)|^2 \mathrm{d}t\mathrm{d}\omega \qquad (1)\\ &+ \lambda\|F\|_{L^1} + \gamma\|G\|_{L^2},\end{aligned}$$

where G is an auxiliary function which quantifies the potentially fast varying instantaneous frequency. It is clear that although $\mathcal{H}$ is not strictly convex, it is convex, so finding the minimizer is guaranteed. To solve this optimization problem, we propose to apply the widely applied and well studied algorithm *Fast Iterative Shrinkage-Thresholding Al-*

gorithm (FISTA). We coin the algorithm as *Time-frequency bY COnvex OptimizatioN* (Tycoon).

The paper is organized in the following way. In Section 2, we discuss the adaptive harmonic model to model the signals with a fast varying instantaneous frequency and its identifiability problem; in Section 3, the motivation of the optimization approach based on the functional (1) is provided; in Section 4, we discuss the numerical details of Tycoon. In particular, how to apply the FISTA algorithm to solve the optimization problem; in Section 5, numerical results of Tycoon are provided.

2. Adaptive Harmonic Model

We start from introducing the model which we use to capture the signal with "fast varying instantaneous frequency". First introduce the following model, which generalizes the $\mathcal{A}_{\epsilon,d}^{c_1,c_2}$ class considered in [11, 7]:

Definition 2.1 (Generalized intrinsic mode type function (gIMT))**.** *Fix constants* $0 \leq \epsilon \ll 1$, $0 < d < 1$ *and* $c_2 > c_1 > 0$. *Consider the functional set* $\mathcal{Q}_{\epsilon}^{c_1,c_2}$, *which consists of functions in* $C^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ *with the following format:*

$$g(t) = A(t)\cos(2\pi\phi(t)),$$

where the following conditions hold:

$$\begin{cases} A \in C^1(\mathbb{R}) \cap L^\infty(\mathbb{R}), \quad \phi \in C^3(\mathbb{R}), \\ \inf_{t\in\mathbb{R}} A(t) > c_1, \quad \inf_{t\in\mathbb{R}} \phi'(t) > c_1, \\ \sup_{t\in\mathbb{R}} A(t) \leq c_2, \quad \sup_{t\in\mathbb{R}} \phi'(t) \leq c_2, \quad \sup_{t\in\mathbb{R}} |\phi''(t)| \leq c_2, \\ |A'(t)| \leq \epsilon\phi'(t), \quad |\phi'''(t)| \leq \epsilon\phi'(t) \quad \text{for all } t \in \mathbb{R}, \end{cases}$$

Definition 2.2 (Adaptive harmonic model)**.** *Fix constants* $0 \leq \epsilon \ll 1$, $d > 0$ *and* $c_2 > c_1 > 0$. *Consider the functional set* $\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$, *which consists of functions in* $C^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ *with the following format:*

$$g(t) = \sum_{\ell=1}^{K} g_k(t),$$

where K *is finite and* $g_k(t) = A_\ell(t)\cos(2\pi\phi_\ell(t)) \in \mathcal{Q}_{\epsilon}^{c_1,c_2}$; *when* $K > 1$, *the following condition is satisfied:*

$$\phi'_{\ell+1}(t) - \phi'_\ell(t) > d \tag{2}$$

for all $\ell = 1, \ldots, K-1$.

We call ϵ, d, c_1 and c_2 *model parameters* of the $\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$ model. Clearly, $\mathcal{Q}_{\epsilon}^{c_1,c_2} \subset \mathcal{Q}_{\epsilon,d}^{c_1,c_2}$ and both $\mathcal{Q}_{\epsilon}^{c_1,c_2}$ and $\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$ are not vector spaces. Note that in the $\mathcal{A}_{\epsilon,d}^{c_1,c_2}$ model, the condition "$\phi_\ell \in C^3(\mathbb{R})$, $\sup_{t\in\mathbb{R}} |\phi''_\ell(t)| \leq c_2$ and $|\phi'''_\ell(t)| \ll \epsilon\phi'_\ell(t)$ for all $t \in \mathbb{R}$" is replaced by "$\phi_\ell \in C^2(\mathbb{R})$ and $|\phi''_\ell(t)| \ll \epsilon\phi'_\ell(t)$ for all $t \in \mathbb{R}$". Thus, we say that the signals in $\mathcal{A}_{\epsilon,d}^{c_1,c_2}$ are oscillatory with slowly varying instantaneous frequency. Also note that $\mathcal{A}_{\epsilon,d}^{c_1,c_2}$ is not a subset of $\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$. Indeed, for $A_\ell(t)\cos(2\pi\phi_\ell(t)) \in \mathcal{A}_{\epsilon,d}^{c_1,c_2}$, even if $\phi_\ell \in C^3(\mathbb{R})$, the third order derivative of ϕ_ℓ is not controlled.

Before proceeding to say what it means by "instantaneous frequency" or "amplitude modulation", we immediately encounter a problem which is understood as the *identifiability problem*. Indeed, we might have infinitely many different ways to represent a cosine function $g_0(t) = \cos(2\pi t)$ in the format $a(t)\cos(2\pi\phi(t))$ so that $a > 0$ and $\phi' > 0$, even though it is well known that $g_0(t)$ is a harmonic function with amplitude 1 and frequency 1. Precisely, there exist infinitely many smooth functions α and β so that $g_0(t) = \cos(t) = (1 + \alpha(t))\cos(2\pi(t + \beta(t)))$, and in general there is no reason to favor $\alpha(t) = \beta(t) = 0$. Before resolving this issue, we could not take amplitude 1 and frequency 1 as reliable features to quantify the signal g_0 when we view it as a component in $\mathcal{Q}_\epsilon^{c_1,c_2}$. In [7], it is shown that if $g(t) = A(t)\cos(\phi(t)) = [A(t) + \alpha(t)]\cos(2\pi[\phi(t) + \beta(t)])$ are both in $\mathcal{A}_{\epsilon,d}^{c_1,c_2}$, then $|\alpha(t)| \le C\epsilon$ and $|\beta'(t)| \le C\epsilon$, where C is a constant depending only on the model parameters c_1, c_2, d. Therefore, A_ℓ and ϕ'_ℓ are unique locally up to an error of order ϵ, and hence we could view them as features of an oscillatory signal in $\mathcal{A}_{\epsilon,d}^{c_1,c_2}$. Here, we show a parallel theorem describing the identifiability property for the functions in the $\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$ model.

Theorem 2.1 (Identifiability of $\mathcal{Q}_\epsilon^{c_1,c_2}$)**.** *Suppose a gIMT $a(t)\cos\phi(t) \in \mathcal{Q}_\epsilon^{c_1,c_2}$ can be represented in a different form which is also a gIMT in $\mathcal{Q}_\epsilon^{c_1,c_2}$; that is, $a(t)\cos\phi(t) = A(t)\cos\varphi(t) \in \mathcal{Q}_\epsilon^{c_1,c_2}$. Define $t_m := \phi^{-1}((m + 1/2)\pi)$ and $s_m := \phi^{-1}(m\pi)$, $m \in \mathbb{Z}$, $\alpha(t) := A(t) - a(t)$, and $\beta(t) := \varphi(t) - \phi(t)$. Then we have the following controls of α and β at t_m and s_m*

1. *Up to a global factor $2l\pi$, $l \in \mathbb{Z}$, $\beta(t_n) = 0$ for all $n \in \mathbb{Z}$;*
2. $\frac{a(t_n)}{a(t_n)+\alpha(t_n)} = \frac{\phi'(t_n)+\beta'(t_n)}{\phi'(t_n)}$ *for all $n \in \mathbb{Z}$. In particular, $\alpha(t_n) = 0$ if and only if $\beta'(t_n) = 0$ for all $n \in \mathbb{Z}$;*
3. *$\alpha(s_m) \ge 0$ for all $m \in \mathbb{Z}$. If $\alpha(s_m) = 0$, then $\beta(s_m) = 0$; if $\alpha(s_m) > 0$, then $|\beta(s_m)| < \pi/2$.*

Furthermore, the size of α and β are bounded by

1. *$|\alpha(t)| < 2\pi\epsilon$ for all $t \in \mathbb{R}$;*
2. *$|\beta''(t)| \le 2\pi\epsilon$, $|\beta'(t)| \le \frac{2\pi\epsilon}{c_1}$ and $|\beta(t)| \le \frac{2\pi\epsilon}{c_1^2}$ for all $t \in \mathbb{R}$.*

Theorem 2.2 (Identifiability of $\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$)**.** *Suppose $f(t) \in \mathcal{Q}_{\epsilon,d}^{c_1,c_2}$ can be represented in a different form which is also in $\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$; that is,*

$$f(t) = \sum_{l=1}^{N} a_l(t)\cos\phi_l(t) = \sum_{l=1}^{M} A_l(t)\cos\varphi_l(t) \in \mathcal{Q}_{\epsilon,d}^{c_1,c_2}.$$

Then, when d is big enough as is described in (B.2), $M = N$ and for all $t \in \mathbb{R}$ and for all $l = 1, \dots, N$, the following holds:

1. *$|\phi_l(t) - \varphi_l(t)| = O(\sqrt{\epsilon})$ up to a global factor $2n\pi$, $n \in \mathbb{Z}$;*
2. *$|\phi'_l(t) - \varphi'_l(t)| = O(\sqrt{\epsilon})$;*
3. *$|\phi''_l(t) - \varphi''_l(t)| = O(\sqrt{\epsilon})$;*
4. *$|a_l(t) - A_l(t)| = O(\sqrt{\epsilon})$,*

where the constants on the right hand side are universal constants depending on the model parameters of $\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$.

Note that in this theorem, we do not concern ourselves with the optimal constants or the optimal d bound. As a result, we have the following definitions, which generalize the notion of amplitude and frequency.

Definition 2.3. *[Phase function, instantaneous frequency, chirp factor and amplitude modulation] Take a function $f(t) = \sum_{\ell=1}^{N} a_\ell(t)\cos\phi_\ell(t) \in \mathcal{Q}_{\epsilon,d}^{c_1,c_2}$. For each $\ell = 1,\dots,N$, the monotonically increasing function $\phi_\ell(t)$ is called the phase function of the ℓ-th gIMT; the first derivative of the phase function, $\phi'_\ell(t)$, is called the instantaneous frequency (IF) of the ℓ-th gIMT; the second derivative of the phase function, $\phi''_\ell(t)$, is called the chirp factor (CF) of the ℓ-th gIMT; the positive function $A_\ell(t)$ is called the amplitude modulation (AM) of the ℓ-th gIMT.*

Note that the IF and AM are always positive, but usually not constant. On the other hand, the CF might be negative and non-constant. Clearly, when ϕ_ℓ are all linear functions with positive slopes and A_ℓ are all positive constants, then the model is reduced to the harmonic model and the IF is equivalent to the notion frequency in the ordinary Fourier transform sense. The conditions $|A'_\ell(t)| \le \epsilon\phi'_\ell(t)$ and $|\phi'''_\ell(t)| \le \epsilon\phi'_\ell(t)$ force the signal to locally behave like a harmonic function or a chirp function, and hence the nominations. By Theorem 2.1 and Theorem 2.2, we know that the definition of these quantities are unique up to an error of order ϵ.

We could also model the commonly encountered ingredient in signal processing – the shape function, trend and noise as those considered in [41, 7]. However, to concentrate the discussion on the optimization approach to the problem, in this paper we focus only on the $\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$ functional class.

3. Optimization Approach

In general, given a function $f(t) = \sum_{k=1}^{K} A_k(t)\cos(2\pi\phi_k(t))$ so that $A_k(t) > 0$ and $\phi'_k(t) > 0$ for $t \in \mathbb{R}$, we would expect to have the *ideal time-frequency representation* (iTFR), denoted as $R_f(t,\omega)$, satisfying

$$R_f(t,\omega) = \sum_{k=1}^{K} A_k(t) e^{i2\pi\phi_k(t)} \delta_{\phi'_k(t)}(\omega),$$

where $\delta_{\phi'_k(t)}$ is the Dirac measure supported at $\phi'_k(t)$, so that we could well extract the features $A_k(t)$ and $\phi'_k(t)$ describing the oscillatory signal from R_f. Note that the iTFR is a distribution. In addition, the reconstruction and visualization of each component are possible. Indeed, we can reconstruct the k-th component by integrating along the frequency axis on the period near $\phi'_k(t)$. Indeed,

$$A_k(t)\cos(2\pi\phi_k(t)) = \Re \int_{\mathbb{R}} R_f(t,\omega) h\left(\frac{\omega - \phi'_k(t)}{\theta}\right) \mathrm{d}\omega,$$

where $\Re$ means taking the real part, $\theta \ll 1$, h is a Schwartz function supported on $[-\Delta,\Delta]$, $\Delta > 0$, so that $h(0) = 1$. Further, the visualization is realized via displaying the "time-varying power spectrum" of f, which is defined as

$$S_f(t,\omega) := \sum_{k=1}^{K} A_k^2(t) \delta_{\phi'_k(t)}(\omega),$$

and we call it the *ideal time-varying power spectrum* (itvPS) of f, which is again a distribution.

To evaluate the iTFR for a function $f = \sum_{k=1}^{K} A_k(t)\cos(2\pi\phi_k(t))$ restricted on a compact and connected interval $I \subset \mathbb{R}$ numerically, we fix $0 < \theta \ll 1$ and consider the following *approximative iTFR with resolution* θ

$$\tilde{R}_f(t,\omega) = \sum_{k=1}^{K} A_k(t) e^{i2\pi\phi_k(t)} \frac{1}{\theta} h\left(\frac{\omega - \phi_k'(t)}{\theta}\right),$$

where $t \in I$, $\omega \in \mathbb{R}$ and h is a Schwartz function supported on $[-\Delta, \Delta]$, $\Delta > 0$, so that $\int h = 1$ and $\frac{1}{\epsilon} h\left(\frac{\cdot}{\epsilon}\right)$ converges to Dirac measure δ supported at 0 weakly as $\epsilon \to 0$ and $\int h(x)\mathrm{d}x = 1$. Clearly, we know that $\tilde{R}_f$ is essentially supported around $(t, \phi_k'(t))$ for $k = 1, \ldots, K$ and as $\epsilon \to 0$, $\tilde{R}_f$ converges to the iTFR in the weak sense. Also, we have for all $t \in I$ and $k = 1, \ldots, K$, when θ is small enough so that $\Delta\theta > d$, we have

$$\Re \int_{\phi_k'(t)-\Delta\theta}^{\phi_k'(t)+\Delta\theta} \tilde{R}_f(t,\omega)\mathrm{d}\omega = A_k(t)\cos(2\pi\phi_k(t)).$$

Thus, the reconstruction property of iTFR is satisfied. In addition, the visualization property of itvPS can be achieved by taking

$$\tilde{S}_f(t,\omega) = \left|\tilde{R}_f(t,\omega)\right|^2 = \sum_{k=1}^{K} |A_k(t)|^2 \frac{1}{\theta^2} \left| h\left(\frac{\omega - \phi_k'(t)}{\theta}\right)\right|^2,$$

where the equality holds due to the facts that ϕ_k' are separated and $\theta \ll 1$. Next we need to find other conditions about $\tilde{R}_f$. A natural one is observing its differentiation. By a direct calculation, we know $\frac{1}{\theta^2} h'\left(\frac{\omega-\phi_k'(t)}{\theta}\right) = \partial_\omega \frac{1}{\theta} h\left(\frac{\omega-\phi_k'(t)}{\theta}\right)$, and hence we have

$$\begin{aligned}
\partial_t \tilde{R}_f(t,\omega) =& \sum_{k=1}^{K} A_k'(t) e^{i2\pi\phi(t)} \frac{1}{\theta} h\left(\frac{\omega - \phi_k'(t)}{\theta}\right) \\
&+ i2\pi \sum_{k=1}^{K} A_k(t)\phi_k'(t) e^{i2\pi\phi_k(t)} \frac{1}{\theta} h\left(\frac{\omega - \phi_k'(t)}{\theta}\right) \\
&- \sum_{k=1}^{K} A_k(t) e^{i2\pi\phi(t)} \phi_k''(t) \frac{1}{\theta^2} h'\left(\frac{\omega - \phi_k'(t)}{\theta}\right) \\
=& \sum_{k=1}^{K} A_k'(t) e^{i2\pi\phi(t)} \frac{1}{\theta} h\left(\frac{\omega - \phi_k'(t)}{\theta}\right) \\
&+ i2\pi \sum_{k=1}^{K} A_k(t)\phi_k'(t) e^{i2\pi\phi_k(t)} \frac{1}{\theta} h\left(\frac{\omega - \phi_k'(t)}{\theta}\right) \\
&+ \partial_\omega \sum_{k=1}^{K} A_k(t) e^{i2\pi\phi(t)} \phi_k''(t) \frac{1}{\theta} h\left(\frac{\omega - \phi_k'(t)}{\theta}\right).
\end{aligned}$$

By the fact that $\omega\tilde{R}_f(t,\omega) = \sum_{k=1}^K A_k(t)\omega e^{i2\pi\phi_k(t)}\frac{1}{\theta}h\left(\frac{\omega-\phi_k'(t)}{\theta}\right)$, we have

$$\begin{aligned}
&\partial_t\tilde{R}_f(t,\omega) - i2\pi\omega\tilde{R}_f(t,\omega) \qquad (3)\\
=&\sum_{k=1}^K A_k'(t)e^{i2\pi\phi(t)}\frac{1}{\theta}h\left(\frac{\omega-\phi_k'(t)}{\theta}\right)\\
&- i2\pi\sum_{k=1}^K A_k(t)(\omega-\phi_k'(t))e^{i2\pi\phi_k(t)}\frac{1}{\theta}h\left(\frac{\omega-\phi_k'(t)}{\theta}\right)\\
&+ \partial_\omega\sum_{k=1}^K A_k(t)e^{i2\pi\phi_k(t)}\phi_k''(t)\frac{1}{\theta}h\left(\frac{\omega-\phi_k'(t)}{\theta}\right).
\end{aligned}$$

We first discuss the case when $f \in \mathcal{A}_{\epsilon,d}^{c_1,c_2}$; that is, $|\phi_k''(t)| \le \epsilon|\phi_k'(t)|$ for all $t \in I$. Note that by the assumption of frequency separation (2) and the fact that $\theta \ll 1$, $[\phi_l'(t)-\theta\Delta, \phi_l'(t)+\theta\Delta] \cap [\phi_k'(t)-\theta\Delta, \phi_k'(t)+\theta\Delta] = \emptyset$ when $l \neq k$. Thus we have

$$\left|\sum_{k=1}^K A_k'(t)e^{i2\pi\phi_k(t)}\frac{1}{\theta}h\left(\frac{\omega-\phi_k'(t)}{\theta}\right)\right|^2 = \sum_{k=1}^K |A_k'(t)|^2\frac{1}{\theta^2}h^2\left(\frac{\omega-\phi_k'(t)}{\theta}\right).$$

Indeed, when $\omega \in [\phi_l'(t)-\theta\Delta, \phi_l'(t)+\theta\Delta]$, we have

$$\left|\sum_{k=1}^K A_k'(t)e^{i2\pi\phi(t)}\frac{1}{\theta}h\left(\frac{\omega-\phi_k'(t)}{\theta}\right)\right|^2 = |A_l'(t)|^2\frac{1}{\theta^2}h^2\left(\frac{\omega-\phi_l'(t)}{\theta}\right).$$

The same argument holds for the other terms on the right hand side of (3). As a result, by a direct calculation, we have

$$\begin{aligned}
&\left\|\sqrt{\theta}\left(\partial_t\tilde{R}_f(t,\omega) - i2\pi\omega\tilde{R}_f(t,\omega)\right)\right\|_{L^2}^2 \qquad (4)\\
\le&\left(\epsilon^2 J_{0,0,2} + 2\pi\theta\epsilon J_{1,0,2} + 4\pi^2\theta^2 J_{2,0,2} + \frac{\epsilon^2 c_2^2}{\theta^2}J_{0,1,2}\right)c_2^2 I,
\end{aligned}$$

where $J_{n,m,l} := \int \eta^n[\partial_\eta^m h(\eta)]^l d\eta$, where $n, m, l = 0, 1, \ldots$. Thus, when ϵ is small enough, $\left\|\sqrt{\theta}\left(\partial_t\tilde{R}_f(t,\omega) - i2\pi\omega\tilde{R}_f(t,\omega)\right)\right\|_{L^2}^2$ is small. This observation leads to a variational approach discussed in [11]. Precisely, the authors in [11] considered to minimize the following functional

$$\begin{aligned}
\mathcal{H}_0(F) := &\int\left|\Re\int F(t,\omega)d\omega - f(t)\right|^2 dt\\
&+ \mu\iint|\partial_t F(t,\omega) - i2\pi\omega F(t,\omega)|^2\, dtd\omega.
\end{aligned}$$

The optimal F would be expected to approximate the iTFR of $f \in \mathcal{A}_{\epsilon,d}^{c_1,c_2}$ well. However, that optimization was not numerically carried out in [11].

Now we come back to the case we have interest; that is, $f \in \mathcal{Q}_{\epsilon,d}^{c_1,c_2}$ restricted on I. Since the condition on the CF terms, that is, $|\phi_k''(t)| \leq \epsilon|\phi_k'(t)|$, no longer holds, the above bound (4) does not hold and minimizing the functional $\mathcal{H}_0$ might not lead to the right solution. In this case, however, we still have the following bound by the same argument as that of (4):

$$\begin{aligned}
&\left\| \sqrt{\theta}\left(\partial_t \tilde{R}_f(t,\omega) - i2\pi\omega\tilde{R}_f(t,\omega) - \partial_\omega \sum_{k=1}^{K} A_k(t)e^{i2\pi\phi_k(t)}\phi_k''(t)\frac{1}{\theta}h\left(\frac{\omega-\phi_k'(t)}{\theta}\right)\right)\right\|_{L^2(I)}^2 \\
&\leq \left\| \sqrt{\theta}\sum_{k=1}^{K}\left(A_k'(t) - i2\pi A_k(t)(\omega-\phi_k'(t))\right)e^{i2\pi\phi(t)}\frac{1}{\theta}h\left(\frac{\omega-\phi_k'(t)}{\theta}\right)\right\|_{L^2(I)}^2 \\
&\leq \left(\epsilon^2 J_{0,0,2} + 2\pi\epsilon\theta J_{1,0,2} + 4\pi^2\theta^2 J_{2,0,2}\right)c_2^2 I,
\end{aligned}$$

Thus, once we find a way to express the extra term $\partial_\omega \sum_{k=1}^{K} A_k(t)e^{i2\pi\phi_k(t)}\phi_k''(t)\frac{1}{\theta}h\left(\frac{\omega-\phi_k'(t)}{\theta}\right)$ in a convenient formula, we could introduce another conditions on F.

In the special case when $K = 1$; that is, $f = A(t)\cos(2\pi\phi(t))$, we know that

$$\partial_\omega\left[A(t)e^{i2\pi\phi(t)}\phi''(t)\frac{1}{\theta}h\left(\frac{\omega-\phi'(t)}{\theta}\right)\right] = \phi''(t)\partial_\omega\tilde{R}_f(t,\omega). \tag{5}$$

Thus, we have

$$\theta\int\int_I |\partial_t\tilde{R}_f(t,\omega) - i2\pi\omega\tilde{R}_f(t,\omega) + \phi''(t)\partial_\omega\tilde{R}_f(t,\omega)|^2 \mathrm{d}t\mathrm{d}\omega = O(\theta^2, \theta\epsilon, \epsilon^2).$$

Thus, we could consider the following functional

$$\theta\iint |\partial_t F(t,\omega) - i2\pi\omega F(t,\omega) + \alpha(t)\partial_\omega F(t,\omega)|^2 \mathrm{d}t\mathrm{d}\omega, \tag{6}$$

where $\alpha(t) \in \mathbb{R}$ is used to capture the CF term associated with the "fast varying instantaneous frequency". Thus, when $K = 1$, we can capture more general oscillatory signals by considering the following functional

$$\begin{aligned}
\mathcal{H}(F,\alpha) := &\int_I \left|\Re\int F(t,\omega)\mathrm{d}\omega - f(t)\right|^2 \mathrm{d}t \\
&+ \mu\theta\iint_I |\partial_t F(t,\omega) - i2\pi\omega F(t,\omega) + \alpha(t)\partial_\omega F(t,\omega)|^2 \mathrm{d}t\mathrm{d}\omega \\
&+ \lambda\|F\|_{L^1(I)} + \gamma\|\alpha\|_{L^2(I)},
\end{aligned} \tag{7}$$

where $F \in L^2(I\times\mathbb{R})$ is the function defined on the time-frequency plane restricted on $I\times\mathbb{R}$. The L^1 norm is added since we would expect to introduce a sparse time-frequency representation when the signal is composed of several intrinsic mode functions.

In general when $K > 1$, we cannot link $\partial_\omega \sum_{k=1}^{K} A_k(t)e^{i2\pi\phi_k(t)}\phi_k''(t)\frac{1}{\theta}h\left(\frac{\omega-\phi_k'(t)}{\theta}\right)$ to $\partial_\omega\tilde{R}_f(t,\omega)$ by any function on t like that in (5). In this case, we could expect to find another function $G \in L^2(I\times\mathbb{R})$ so that

$$G(t,\omega) = \begin{cases} \phi_k''(t) & \text{when } \omega \in [\phi_k'(t) - \theta\Delta, \phi_k'(t) + \theta\Delta] \\ 0 & \text{o.w..} \end{cases} \tag{8}$$

and hence $G(t,\omega)\partial_\omega \tilde{R}_f(t,\omega) = \partial_\omega \sum_{k=1}^K A_k(t)e^{i2\pi\phi_k(t)}\phi_k''(t)\frac{1}{\theta}h\left(\frac{\omega-\phi_k'(t)}{\theta}\right)$. Thus, we could consider minimizing the following functional

$$\begin{aligned}\mathcal{H}(F,G) := &\int \left|\Re\int F(t,\omega)\mathrm{d}\omega - f(t)\right|^2 \mathrm{d}t \\ &+ \mu\theta\iint_I |\partial_t F(t,\omega) - i2\pi\omega F(t,\omega) + G(t,\omega)\partial_\omega F(t,\omega)|^2 \mathrm{d}t\mathrm{d}\omega \\ &+ \lambda\|F\|_{L^1(I)} + \frac{\gamma}{\sqrt{\theta}}\|G\|_{L^2(I)}.\end{aligned} \tag{9}$$

Here, the L^2 penalty term $\|G\|_{L^2}$ has $1/\sqrt{\theta}$ in front of it since

$$\|G\|_{L^2(I)} = \sqrt{2\theta\Delta}\sum_{k=1}^K \|\phi_k''\|_{L^2(I)}. \tag{10}$$

Thus, the L^2 penalty term does not depend on θ. It is also clear that the L^1 penalty term in the above functional does not depend on θ as we have

$$\int\int_I \left|\sum_{k=1}^K A_k(t)e^{i2\pi\phi(t)}\frac{1}{\theta}h\left(\frac{\omega-\phi_k'(t)}{\theta}\right)\right|\mathrm{d}t\mathrm{d}\omega = \sum_{k=1}^K \|A_k(t)\|_{L^1(I)}. \tag{11}$$

Next theorem shows that we can always find a minimizer of the functional $\mathcal{H}$ which has an Hermitian property.

Theorem 3.1. *Let f be a real signal in $\mathcal{H}$. Then, for a fixed G such that for all t,ω, $G(t,\omega) = -\overline{G(t,-\omega)}$, there exists a minimizer F of $\mathcal{H}(\cdot,G)$, which has a Hermitian symmetry property with respect to the frequencies ω, i.e. for all t,ω, $\overline{F(t,\omega)} = F(t,-\omega)$.*

Proof. Let F be a minimizer of $\mathcal{H}(\cdot,G)$. We claim that $\check{F} : L^2(I\times\mathbb{R}) \to \mathbb{C}, (t,\omega)\mapsto \overline{F(t,-\omega)}$ is also a minimizer of $\mathcal{H}(\cdot,G)$. Indeed:

$$\begin{aligned}\mathcal{H}(\check{F},G) = &\int \left|\Re\int \overline{F(t,-\omega)}\mathrm{d}\omega - f(t)\right|^2 \mathrm{d}t \\ &+ \mu\theta\iint \left|\partial_t\overline{F(t,-\omega)} - i2\pi\omega\overline{F(t,-\omega)} + G(t,\omega)\partial_\omega\left(\overline{F(t,-\omega)}\right)\right|^2 \mathrm{d}t\mathrm{d}\omega \\ &+ \lambda\iint \left|\overline{F(t,-\omega)}\right|\mathrm{d}t\mathrm{d}\omega + \frac{\gamma}{\sqrt{\theta}}\|G\|_{L^2}.\end{aligned}$$

Then, using a change of variable $\eta = -\omega$ and the assumption $G(t,\omega) = -\overline{G(t,-\omega)}$:

$$\begin{aligned}\mathcal{H}(\check{F},G) = &\int \left|\Re\int \overline{F(t,\eta)}\mathrm{d}\eta - f(t)\right|^2 \mathrm{d}t \\ &+ \mu\sqrt{\theta}\int \left|\left(\partial_t + i2\pi\eta + \overline{G(t,\eta)}\partial_\eta\right)\overline{F(t,\eta)}\right|^2 \mathrm{d}t\mathrm{d}\eta \\ &+ \lambda\iint \left|\overline{F(t,\eta)}\right|\mathrm{d}t\mathrm{d}\eta + \frac{\gamma}{\sqrt{\theta}}\|G\|_{L^2(I)}.\end{aligned}$$

Here, note that $\partial_\omega \left(\overline{F(t,-\omega)}\right) = -\partial_\eta \left(\overline{F(t,\eta)}\right)$. We obviously have

$$\Re \int \overline{F(t,\eta)} \mathrm{d}\eta = \Re \int F(t,\eta) \mathrm{d}\eta$$

and

$$\iint \left|\overline{F(t,\eta)}\right| \mathrm{d}t\mathrm{d}\eta = \iint |F(t,\eta)| \mathrm{d}t\mathrm{d}\eta.$$

Moreover, one has

$$\begin{aligned}
&\Re\left[\left(\partial_t + i2\pi\omega + \overline{G(t,\omega)}\partial_\omega\right)\overline{F(t,\omega)}\right] \\
&= \Re(\partial_t \overline{F(t,\omega)}) - 2\pi\omega\Im(\overline{F(t,\omega)}) + \Re(\overline{G(t,\omega)})\Re(\partial_\omega \overline{F(t,\omega)}) - \Im(\overline{G(t,\omega)})\Im(\partial_\omega \overline{F(t,\omega)}) \\
&= \Re(\partial_t F(t,\omega)) + 2\pi\omega\Im(F(t,\omega)) + \Re(G(t,\omega))\Re(\partial_\omega F(t,\omega)) - \Im(G(t,\omega))\Im(\partial_\omega F(t,\omega)) \\
&= \Re\left[\left(\partial_t - i2\pi\omega + G(t,\omega)\partial_\omega\right) F(t,\omega)\right]
\end{aligned}$$

and

$$\begin{aligned}
&\Im\left[\left(\partial_t + i2\pi\omega + \overline{G(t,\omega)}\partial_\omega\right)\overline{F(t,\omega)}\right] \\
&= \Im(\partial_t \overline{F(t,\omega)}) + 2\pi\omega\Re(\overline{F(t,\omega)}) + \Im(\overline{G(t,\omega)})\Re(\overline{F(t,\omega)}) + \Re(\overline{G(t,\omega)})\Im(\overline{F(t,\omega)}) \\
&= -\Im(\partial_t F(t,\omega)) + 2\pi\omega\Re(F(t,\omega)) - \Im(G(t,\omega))\Im(F(t,\omega)) - \Re(G(t,\omega))\Im(F(t,\omega)) \\
&= -\Im\left[\left(\partial_t - i2\pi\omega + \overline{G(t,\omega)}\right) F(t,\omega)\right].
\end{aligned}$$

Thus, one has

$$\begin{aligned}
&\iint \left|\partial_t F(t,\omega) - i2\pi\omega F(t,\omega) + G(t,\omega)\partial_\omega\left(F(t,\omega)\right)\right|^2 \mathrm{d}t\mathrm{d}\omega \\
&= \iint \left|\partial_t \overline{F(t,-\omega)} - i2\pi\omega\overline{F(t,-\omega)} + G(t,\omega)\partial_\omega\left(\overline{F(t,-\omega)}\right)\right|^2 \mathrm{d}t\mathrm{d}\omega,
\end{aligned}$$

and hence $\check{F}$ is also a minimizer of $\mathcal{H}(\cdot, G)$. Consequently, $\mathcal{H}(\cdot, G)$ being convex, $cF + (1-c)\check{F}$, where $0 \le c \le 1$, is also a minimizer of $\mathcal{H}(\cdot, G)$, hence the conclusion. □

Remark. There exist a *scaling ambiguity* between G and $\partial_\omega F$, as we can alway rewrite $G(t,\omega)\partial_\omega F(t,\omega)$ as $G(t,\omega)cc^{-1}\partial_\omega F(t,\omega)$ with $c \neq 0$. This scaling ambiguity is well known in the blind source separation problem [31], and the L^2 regularization on G allows one to deal with it during the optimization process. However, we could not obtain G without this scaling ambiguity and its correct scaling factor will be estimated separately.

4. Numerical Algorithm

We consider the following functionals associated with (7):

$$\begin{aligned}
\mathcal{G}(F,\alpha) := &\int_{\mathbb{R}} \left|\Re \int_{\mathbb{R}} F(t,\omega) d\omega - f(t)\right|^2 \mathrm{d}t \\
&+ \mu \iint_{\mathbb{R}} \left|\partial_t F(t,\omega) - i2\pi\omega F(t,\omega) + \alpha(t)\partial_\omega F(t,\omega)\right|^2 \mathrm{d}t\mathrm{d}\omega,
\end{aligned}$$

and

$$\Psi(F,\alpha) := \lambda\|F\|_{L^1} + \gamma\|\alpha\|_{L^2}^2,$$

where t is the time and ω is the frequency. The numerical implementation of (9) follows the same lines while we have to discretize a two dimensional function G.

4.1. Numerical discretization

Numerically, we consider the following discretization of F by taking $\Delta_t > 0$ and $\Delta_\omega > 0$ as the sampling periods in the time axis and frequency axis. We also restrict F to time $[0, M\Delta_t]$ and, because of Theorem 3.1, to the positive frequencies $[0, N\Delta_\omega]$, where $N, M \in \mathbb{N}$. Then, we discretize F as $\boldsymbol{F} \in \mathbb{C}^{(N+1)\times(M+1)}$ and α as $\boldsymbol{\alpha} \in \mathbb{R}^{M+1}$, where

$$\boldsymbol{F}_{n,m} = F(t_m, \omega_n), \quad \boldsymbol{\alpha}_m = \alpha(t_m),$$

$t_m := m\Delta_t$, $\omega_n := n\Delta_\omega$, $n = 0, 1, \ldots, N$ and $m = 0, 1, \ldots, M$. The observed signal $f(t)$ is discretized as a $(M+1)$-dim vector $\boldsymbol{f}$, where

$$\boldsymbol{f}_l = f(t_l).$$

Note that the sampling period of the signal Δ_t and M most of time are determined by the data collection procedure. We could set $\Delta_\omega = \frac{1}{M\Delta_t}$ and $N = \lceil M/2 \rceil$ suggested by the Nyquist rate in the sampling theory.

Next, using the rectangle method, we could discretize $\mathcal{G}(F,\alpha)$ directly by

$$\begin{aligned}\mathcal{G}(\boldsymbol{F},\boldsymbol{\alpha}) := &\sum_{m=0}^{M}\left|\sum_{n=0}^{N} 2\Re\left(F(t_m,\omega_n)\right)\Delta_\omega - f(t_m)\right|^2 \Delta_t \\ &+ \mu\sum_{m=0}^{M}\sum_{n=0}^{N}\left|\partial_t F(t_m,\omega_n) - i2\pi\omega_n F(t_m,\omega_n) + \alpha(t_m)\partial_\omega F(t_m,\omega_n)\right|^2 \Delta_t\Delta_\omega.\end{aligned}$$

The partial derivative $\partial_t F$ can be implemented by the straight finite difference; that is, take a $(M+1)\times(M+1)$ finite difference matrix $\boldsymbol{D}_{M+1}$ so that $\boldsymbol{F}\boldsymbol{D}_{M+1}$ approximates the discretization of $\partial_t F$. However, this choice may lead to numerical instability. Instead, one can implement the partial derivative in the Fourier domain, using that $\partial_t F(t_m,\omega_n) = \mathcal{F}^{-1}\left(i2\pi\xi_k\hat{F}(\xi_k,\omega_n)\right)[m]$, where $\hat{F} = \mathcal{F}(F)$ and $\mathcal{F}$ denotes the finite Fourier transform. For the sake of simplicity, we still denote by ∂_t or ∂_ω the discretization operator in the discret domain, whatever the chosen method (finite difference or in the Fourier domain). Also denote $\mathbf{1} = (1,\ldots,1)^T \in \mathbb{R}^{N+1}$. In the matrix form, the functional $\mathcal{G}(F,\alpha)$ is thus discretized as

$$\mathcal{G}(\boldsymbol{F},\boldsymbol{\alpha}) = \Delta_t\left\|\mathcal{A}\boldsymbol{F} - \boldsymbol{F}\right\|^2 + \Delta_t\Delta_\omega\mu\left\|\mathcal{B}(\boldsymbol{F},\boldsymbol{\alpha})\right\|^2,$$

where

$$\begin{aligned}\mathcal{A}: \quad \mathbb{C}^{(N+1)\times(M+1)} &\to \mathbb{R}^{M+1} \\ \boldsymbol{F} &\mapsto 2\Re\left(\mathbf{1}^T\boldsymbol{F}\right)\Delta_\omega,\end{aligned}$$

$$\begin{aligned}\mathcal{B}: \quad \mathbb{C}^{(N+1)\times(M+1)}\times\mathbb{C}^{M+1} &\to \mathbb{C}^{(N+1)\times(M+1)} \\ (\boldsymbol{F},\boldsymbol{\alpha}) &\mapsto \partial_t\boldsymbol{F} - i2\pi\boldsymbol{\omega}\boldsymbol{F} + \partial_\omega\boldsymbol{F}\mathrm{diag}(\boldsymbol{\alpha}),\end{aligned}$$

and $\boldsymbol{\omega} = \mathrm{diag}(0, \Delta_\omega, 2\Delta_\omega, \ldots, N\Delta_\omega) \in \mathbb{R}^{(N+1)\times(N+1)}$.

4.2. Expression of the gradient operator

Denote $\mathcal{G}_{\boldsymbol{\alpha}}(\boldsymbol{F}) := \mathcal{G}(\boldsymbol{F}, \boldsymbol{\alpha})$ and $\mathcal{B}_{\boldsymbol{\alpha}}(\boldsymbol{F}) := \mathcal{B}(\boldsymbol{F}, \boldsymbol{\alpha})$; that is, $\boldsymbol{\alpha}$ is fixed. Similarly, define $\mathcal{G}_{\boldsymbol{F}}(\boldsymbol{\alpha}) := \mathcal{G}(\boldsymbol{F}, \boldsymbol{\alpha})$ and $\mathcal{B}_{\boldsymbol{F}}(\boldsymbol{\alpha}) := \mathcal{B}(\boldsymbol{F}, \boldsymbol{\alpha})$; that is, $\boldsymbol{F}$ is fixed. We will evaluate the gradient of $\mathcal{G}_{\boldsymbol{\alpha}}$ and $\mathcal{G}_{\boldsymbol{F}}$ after discretization for the gradient decent algorithm. Take $\boldsymbol{G} \in \mathbb{C}^{(M+1)\times(N+1)}$. The gradient of $\mathcal{G}_{\boldsymbol{\alpha}}$ after discretization is evaluated by

$$\begin{aligned}
\nabla \mathcal{G}_{\boldsymbol{\alpha}}|_{\boldsymbol{F}} \boldsymbol{G} &= \lim_{h\to 0} \frac{\mathcal{G}_{\boldsymbol{\alpha}}(\boldsymbol{F} + h\boldsymbol{G}) - \mathcal{G}_{\boldsymbol{\alpha}}(\boldsymbol{F})}{h} \\
&= 2\Delta_t (\mathcal{A}\boldsymbol{F} - \boldsymbol{f})^T \mathcal{A}\boldsymbol{G} + 2\Delta_t \Delta_\omega \mu \langle \mathcal{B}_{\boldsymbol{\alpha}} \boldsymbol{F}, \mathcal{B}_{\boldsymbol{\alpha}} \boldsymbol{G} \rangle \\
&= \langle 2\Delta_t \mathcal{A}^*(\mathcal{A}\boldsymbol{F} - \boldsymbol{f}) + 2\Delta_t \Delta_\omega \mu \mathcal{B}_{\boldsymbol{\alpha}}^* \mathcal{B}_{\boldsymbol{\alpha}} \boldsymbol{F}, \boldsymbol{G} \rangle.
\end{aligned}$$

As a result, we have

$$\nabla \mathcal{G}_{\boldsymbol{\alpha}}|_{\boldsymbol{F}} = 2\Delta_t \mathcal{A}^*(\mathcal{A}\boldsymbol{F} - \boldsymbol{f}) + 2\Delta_t \Delta_\omega \mu \mathcal{B}_{\boldsymbol{\alpha}}^* \mathcal{B}_{\boldsymbol{\alpha}} \boldsymbol{F}.$$

where $\mathcal{A}^*$ and $\mathcal{B}_{\boldsymbol{\alpha}}^*$ are adjoint operators of $\mathcal{A}$ and $\mathcal{B}_{\boldsymbol{\alpha}}$ respectively. Now we expand $\mathcal{A}^*$ and $\mathcal{B}_{\boldsymbol{\alpha}}^*$. Take $\boldsymbol{g} \in \mathbb{R}^{M+1}$. We have

$$\begin{aligned}
\langle \mathcal{A}\boldsymbol{F}, \boldsymbol{g} \rangle &= \sum_{m=0}^{M} \left(\sum_{n=0}^{N} 2\Re \boldsymbol{F}_{n,m} \Delta_\omega \right) \boldsymbol{g}_m \\
&= \sum_{m=0}^{M} \sum_{n=0}^{N} 2\Re \boldsymbol{F}_{n,m} \Re(\Delta_\omega \boldsymbol{g}_m),
\end{aligned}$$

and

$$\begin{aligned}
\langle \boldsymbol{F}, \mathcal{A}^* \boldsymbol{g} \rangle &= \sum_{m=0}^{M} \sum_{n=0}^{N} \boldsymbol{F}_{n,m} \overline{(\mathcal{A}^* \boldsymbol{g})}_{n,m} \\
&= \sum_{m=0}^{M} \sum_{n=0}^{N} 2\Re \boldsymbol{F}_{n,m} \Re(\mathcal{A}^* \boldsymbol{g})_{n,m} + \sum_{m=0}^{M} \sum_{n=0}^{N} \Im \boldsymbol{F}_{n,m} \Im(\mathcal{A}^* \boldsymbol{g})_{n,m} \\
&\quad + i \sum_{m=0}^{M} \sum_{n=0}^{N} \Im \boldsymbol{F}_{n,m} 2\Re(\mathcal{A}^* \boldsymbol{g})_{n,m} - i \sum_{m=0}^{M} \sum_{n=0}^{N} \Re \boldsymbol{F}_{n,m} \Im(\mathcal{A}^* \boldsymbol{g})_{n,m}.
\end{aligned}$$

Since $\langle \mathcal{A}\boldsymbol{F}, \boldsymbol{g} \rangle = \langle \boldsymbol{F}, \mathcal{A}^* \boldsymbol{g} \rangle$ for all $\boldsymbol{F}$ and $\boldsymbol{g}$, we conclude that

$$\begin{array}{rccc}
\mathcal{A}^* : & \mathbb{R}^{M+1} & \to & \mathbb{C}^{(N+1)\times(M+1)} \\
& \boldsymbol{g} & \mapsto & 2\Delta_\omega \begin{pmatrix} \boldsymbol{g}_1 & \cdots & \boldsymbol{g}_{M+1} \\ \vdots & & \vdots \\ \boldsymbol{g}_1 & \cdots & \boldsymbol{g}_{M+1} \end{pmatrix}.
\end{array} \tag{12}$$

To calculate $\mathcal{B}_{\boldsymbol{\alpha}}^*$, by a direct calculation we have

$$\begin{aligned}
\langle \mathcal{B}_{\boldsymbol{\alpha}} \boldsymbol{F}, \boldsymbol{G} \rangle &= \langle \partial_t \boldsymbol{F} - i2\pi \boldsymbol{\omega} \boldsymbol{F} + \partial_\omega \boldsymbol{F} \mathrm{diag}(\boldsymbol{\alpha}), \boldsymbol{G} \rangle \\
&= \langle \boldsymbol{F}, -\partial_t \boldsymbol{G} + i2\pi \boldsymbol{\omega} \boldsymbol{G} - \partial_\omega \boldsymbol{G} \mathrm{diag}(\boldsymbol{\alpha}) \rangle \\
&= \langle \boldsymbol{F}, \mathcal{B}_{\boldsymbol{\alpha}}^* \boldsymbol{G} \rangle,
\end{aligned}$$

where $\boldsymbol{G} \in \mathbb{C}^{(N+1)\times(M+1)}$. Thus, we conclude that

$$\begin{array}{rccc} \mathcal{B}^*_{\boldsymbol{\alpha}}: & \mathbb{C}^{(M+1)\times(N+1)} & \to & \mathbb{C}^{(M+1)\times(N+1)} \\ & \boldsymbol{G} & \mapsto & -\partial_t \boldsymbol{G} + i2\pi\boldsymbol{\omega}\boldsymbol{G} - \partial_\omega \boldsymbol{G}\mathrm{diag}(\boldsymbol{\alpha}). \end{array} \tag{13}$$

As a result, the first part of $\nabla\mathcal{G}_{\boldsymbol{\alpha}}|_{\boldsymbol{F}}$, $2\Delta_t\mathcal{A}^*(\mathcal{A}\boldsymbol{F} - \boldsymbol{f})$, can be numerically expressed as

$$4\Delta_t\Delta_\omega \begin{pmatrix} \Delta_\omega \Re \sum_{n=0}^{N} \boldsymbol{F}_{n,1} - \boldsymbol{f}_1 & \dots & \Delta_\omega \Re \sum_{n=0}^{N} \boldsymbol{F}_{n,M+1} - \boldsymbol{f}_{M+1} \\ \vdots & & \vdots \\ \Delta_\omega \Re \sum_{n=0}^{N} \boldsymbol{F}_{n,1} - \boldsymbol{f}_1 & \dots & \Delta_\omega \Re \sum_{n=0}^{N} \boldsymbol{F}_{n,M+1} - \boldsymbol{f}_{M+1} \end{pmatrix} \in \mathbb{R}^{(N+1)\times(M+1)}. \tag{14}$$

and the second term

$$\begin{aligned} 2\Delta_t\Delta_\omega\mu\mathcal{B}^*\mathcal{B}\boldsymbol{F} = 2\Delta_t\Delta_\omega\mu \big(& -\partial_t\partial_t\boldsymbol{F} + i4\pi\boldsymbol{\omega}\partial_t\boldsymbol{F} - \partial_t\partial_\omega\boldsymbol{F}\mathrm{diag}(\boldsymbol{\alpha}) + i4\pi^2\boldsymbol{\omega}^2\boldsymbol{F} \\ & + i2\pi\boldsymbol{\omega}\partial_\omega\boldsymbol{F}\mathrm{diag}(\boldsymbol{\alpha}) - \partial_\omega\partial_t\boldsymbol{F}\mathrm{diag}(\boldsymbol{\alpha}) + i2\pi\partial_\omega\boldsymbol{\omega}\boldsymbol{F}\mathrm{diag}(\boldsymbol{\alpha}) - \partial_\omega\partial_\omega\boldsymbol{F}\mathrm{diag}(\boldsymbol{\alpha})\big) \end{aligned} \tag{15}$$

Similarly, by taking $\boldsymbol{\beta} \in \mathbb{C}^{M+1}$, the gradient of $\mathcal{G}_{\boldsymbol{F}}$ at $\boldsymbol{\alpha}$ after discretization is evaluated by

$$\begin{aligned} \nabla\mathcal{G}_{\boldsymbol{F}}|_{\boldsymbol{\alpha}}\boldsymbol{\beta} &= \lim_{h\to 0} \frac{\mathcal{G}_{\boldsymbol{F}}(\boldsymbol{\alpha} + h\boldsymbol{\beta}) - \mathcal{G}_{\boldsymbol{F}}(\boldsymbol{\alpha})}{h} \\ &= \mathrm{tr}((\partial_\omega\boldsymbol{F}\mathrm{diag}(\boldsymbol{\beta}))^*(\partial_\omega\boldsymbol{F}\mathrm{diag}(\boldsymbol{\alpha})) \\ &= -\boldsymbol{\beta}^*\mathrm{tr}(\boldsymbol{F}^*\partial_\omega\partial_\omega\boldsymbol{F}\mathrm{diag}(\boldsymbol{\alpha})). \end{aligned}$$

Thus, we have

$$\nabla\mathcal{G}_{\boldsymbol{F}}|_{\boldsymbol{\alpha}} = -\mathrm{tr}(\boldsymbol{F}^*\partial_\omega\partial_\omega\boldsymbol{F}\mathrm{diag}(\boldsymbol{\alpha})) \in \mathbb{C}^{M+1},$$

where

$$(\nabla\mathcal{G}_{\boldsymbol{F}}|_{\boldsymbol{\alpha}})_m = \alpha(t_m)\sum_{n=1}^{N+1}[\partial_\omega F(t_m,\omega_n)]^2.$$

4.3. Minimize the functional $\mathcal{H}(F,\alpha)$

We now have all the results needed to propose an optimization algorithm to minimize the functional $\mathcal{H}(F,\alpha)$. The minimization of $\mathcal{H}(F,G)$ in (9) is the same so we skip it. The functional we would like to minimize depends on two terms, F and α. Since the functional spaces F and α live are convex, we will therefore minimize the functional alternately by optimizing one of these two terms when the other one is fixed; that is,

$$\begin{cases} F_{k+1} = \arg\min_F \mathcal{H}(F,\alpha_k) \\ \alpha_{k+1} = \arg\min_\alpha \mathcal{H}(F_{k+1},\alpha). \end{cases} \tag{16}$$

with $\alpha_0 = 0$ and $F_0 = 0$ are used to initialize the algorithm. Convergence results of this classical Gauss-Seidel method can be found for example in [46, 36].

4.4. Minimization of $\mathcal{H}_\alpha := \mathcal{H}(\cdot, \alpha)$

When α is fixed, $\mathcal{H}_\alpha$ is a convex non smooth functional, involving a convex Lipschitz differentiable term (the function $\mathcal{G}_\alpha$), and a convex non smooth term (the Ψ_α regularizer). Popular proximal algorithms such as forward-bacward [10] or the Fast Iterative Shrinkage/Thresholding Algorithm (FISTA) [4] can then be employed. FISTA has the great advantage to reach the optimal rate of convergence; that is, if $\check{\boldsymbol{F}}$ is the convergence point, $\mathcal{H}_\alpha(\boldsymbol{F}_k) - \mathcal{H}_\alpha(\check{\boldsymbol{F}}) = \mathcal{O}\left(\frac{1}{k^2}\right)$, while the forward-backward procedure converge in $\mathcal{O}\left(\frac{1}{k}\right)$ (see [37] for a great review of proximal methods and their acceleration). Contrary to the forward-backward, one limitation of FISTA is that the convergence is proven only on the sequence $(\mathcal{H}_\alpha(\boldsymbol{F}_k))_k$ rather than on the iterates $(\boldsymbol{F}_k)_k$. However, the latest study [5] gives a version of FISTA which fulfill this gap while maintaining the same convergence rate. As far as we know, it is the only algorithm with these two properties, and then will be use in the following.

In short, FISTA relies on three steps

1. A gradient descent step on the smooth term $\mathcal{G}_{\boldsymbol{\alpha}}$;
2. A soft-shrinkage operation, known as the proximal step;
3. A relaxation step.

The algorithm is summarized in Algorithm 1. In practice, the Lipschitz constant can be evaluated using a classical power iteration procedure, or using a backtracking step inside the algorithm (see [4] for details). $\nabla\mathcal{G}_\alpha$ is given by Eq. (14) and Eq. (15).

Algorithm 1 FISTA algorithm for $\mathcal{H}_\alpha$.

Choose the number of iterations n_{it}.
The initial values are $\boldsymbol{F}_0 \in \mathbb{C}^{(N+1)\times(M+1)}$, $z_0 = \boldsymbol{F}_0$ and the Lipschitz constant $L = \|\nabla\mathcal{G}_\alpha\|^2$.
for $k = 0$ **to** $n_{\text{it}} - 1$ **do**
  Gradient step: $\boldsymbol{F}_{k+1/2} \leftarrow z_k - \frac{1}{L}\nabla\mathcal{G}_\alpha|_{z_k}$;
  Proximal step: $\boldsymbol{F}_{k+1} \leftarrow \boldsymbol{F}_{k+1/2}\left(1 - \frac{\lambda/L}{|F_{k+1/2}|}\right)^+$;
  Relaxation step: $z_{k+1} \leftarrow \boldsymbol{F}_{k+1} + \frac{k+5}{k+6}(\boldsymbol{F}_{k+1} - \boldsymbol{F}_k)$;
end for
Output $\boldsymbol{F}_{n_{\text{it}}}$.

4.5. Minimization of $\mathcal{H}_F := \mathcal{H}(F, \cdot)$

Once $\boldsymbol{F}_k$ is estimated, the minimization of $\mathcal{H}_{\boldsymbol{F}_k}$ reduces to a simple quadratic minimization:

$$\boldsymbol{\alpha}_{k+1} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left\{ \mu \sum_{m=0}^{M}\sum_{n=0}^{N} |\partial_t F(t_m,\omega_n) - i2\pi\omega_n F(t_m,\omega_n) + \alpha(t_m)\partial_\omega F(t_m,\omega_n)|^2 + \gamma \sum_{m=0}^{M} |\alpha(t_m)|^2 \right\}.$$

Thus, α can be estimated in a closed form as, for all $m = 0, \ldots, M$,

$$\alpha_{k+1}(t_m) = \frac{2\sum_{n=0}^{N} \Re\left(\partial_\omega \overline{F(t_m,\omega_n)}\left[\partial_t F(t_m,\omega_n) - i2\pi\omega_n F(t_m,\omega_n)\right]\right)}{\sum_{n=0}^{N} |\partial_\omega F(t_m,\omega_n)|^2 + \gamma/\mu}.$$

5. Numerical Results

In this section we show numerical simulation results of the proposed algorithm. The code and simulated data are available via request. In this section, we take W to be the standard Brownian motion defined on $[0,\infty)$ and define a *smoothed Brownian motion with bandwidth* $\sigma > 0$ as

$$\Phi_\sigma := W \star K_\sigma,$$

where K_σ is the Gaussian function with the standard deviation $\sigma > 0$ and $\star$ denotes the convolution operator.

The first example is a semi-real example which is inspired from a medical challenge. Atrial fibrillation (Af) is a pathological condition associated with high mortality and morbidity [22]. It is well known that the subject with Af would have irregularly irregular heart beats. In the language under our framework, the instantaneous frequency of the electrocardiogram signal recorded from an Af patient varies fast. To study this kind of signal with fast varying instantaneous frequency, we pick a patient with Af and determine its instantaneous heart rate by evaluating its R peak to R peak intervals. Precisely, if the R peaks are located on t_i, we generate a non-uniform sampling of the instantaneous heart rate and denote it as $(t_i, 1/(t_{i+1} - t_i))$. Then the instantaneous heart rate, denoted as $\phi_1'(t)$, is approximated by the cubic spline interpolation. Next, define another a random process A_1 on $[0, L]$ by

$$A_1(t) = 1 + \frac{\Phi_{\sigma_1}(t) + \|\Phi_{\sigma_1}\|_{L^\infty[0,L]}}{2\|\Phi_{\sigma_1}\|_{L^\infty[0,L]}}, \tag{17}$$

where $t \in [0, L]$ and $\sigma_1 > 0$. Note that A_1 is a positive random process and in general there is no close form expression of $A_1(t)$ and $\phi_1(t)$. The dynamic of both components can be visually seen from the signal.

We then generate an oscillatory signal with fast varying instantaneous frequency

$$f_1(t) = A_1(t)\cos(2\pi\phi_1(t)),$$

where $A_1(t)$ is a realization of the random process defined in (17). We take $L = 80$, sample f_1 with the sampling rate $\Delta t = 1/10$, $\sigma_1 = 100$, $\sigma_2 = 200$. In Tycoon, we take $\mu = 0.003$, $\nu = 0.007$ and $\gamma = 1$. To compare the result with other methods, in addition to showing the result of the proposed algorithm, we also show the analysis results of short time Fourier transform (STFT) and synchrosqueezed STFT. In the short time Fourier transform (STFT) and synchrosqueezed STFT, we take the window function g as a Gaussian function with the standard deviation $\sigma = 1$. Please see Figure 1 for an example. In this example, we see that the proposed algorithm Tycoon could extract this

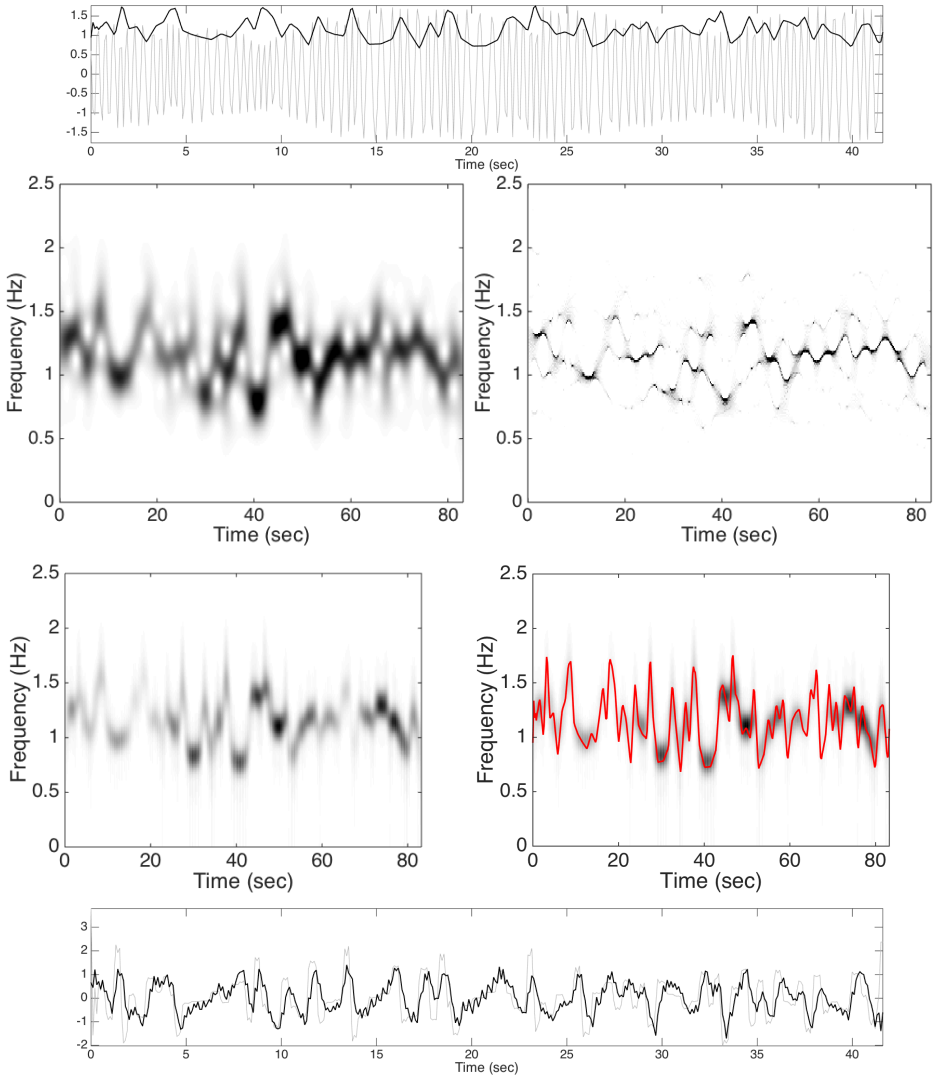

Figure 1: Top: the signal f_1 is shown as the gray curve with the instantaneous frequency superimposed as the black curve. It is clear that the instantaneous frequency varies fast. In the second row, the short time Fourier transform with the Gaussian window with the standard deviation 1 is shown on the left and the synchrosqueezed short time Fourier transform is shown on the right. In the third row, the Tycoon result is shown on the left and our result with the instantaneous frequency superimposed as a red curve is shown on the right. In the bottom, the chirp factor, $\phi_2''(t)$, is shown as the gray curve and the estimated $\phi_2''(t)$; that is, the $\alpha(t)$, is properly normalized to remove the scaling ambiguity and superimposed as the black curve.

kind of fast varying instantaneous frequency, while STFT and synchrosqueezed STFT fail. In addition, the chirp factor can be approximated up to some extent.

In the second example, we consider an oscillatory signal with two gIMTs. Define

random processes $A_2(t)$ and $\phi_2(t)$ on $[0, L]$ by

$$A_2(t) = 1 + \frac{\Phi_{\sigma_1}(t) + 2\|\Phi_{\sigma_1}\|_{L^\infty[0,L]}}{3\|\Phi_{\sigma_1}\|_{L^\infty[0,L]}},$$
$$\phi_2(t) = \pi t + \int_0^t \left[\frac{\Phi_{\sigma_2}(s) + 0.5\|\Phi_{\sigma_2}\|_{L^\infty[0,L]}}{1.5\|\Phi_{\sigma_2}\|_{L^\infty[0,L]}} - \sin(s)\right] \mathrm{d}s,$$

where $t \in [0, L]$ and $\sigma_2 > 0$. Note that by definition ϕ_2 are both monotonically increasing random processes. The signal is constructed as

$$f = f_1 + f_2, \tag{18}$$

where $f_2 = A_2 \cos(2\pi\phi_2(t))\chi_{[20,80]}(t)$ and χ is the indicator function. Again, we take $\sigma_1 = 100$, $\sigma_2 = 200$, $L = 80$ and sample f with the sampling rate $\Delta t = 1/10$. In Tycoon, we take $\mu = 0.003$, $\nu = 0.007$ and $\gamma = 1$. In STFT and synchrosqueezed STFT, the window function is the same as that in the first example – the Gaussian window with the standard deviation $\sigma = 1$. The results of STFT, synchrosqueezed STFT and Tycoon are shown in Figure 2. It is clear that the proposed convex optimization approach provides the dynamical information hidden inside the signal f.

Lastly, we add noise to the signal f and see how the proposed algorithm performs. To model the noise, we define the signal to noise ratio (SNR) as

$$\mathrm{SNR} := 20 \log_{10} \frac{\mathrm{std}(f)}{\mathrm{std}(\Phi)},$$

where f is the clean signal, Φ is the added noise and std means the standard deviation. In this simulation, we add the Gaussian white noise with SNR 7.25 to the clean signal f, and obtain a noisy signal Y. The result is shown in Figure 3. Clearly, we see that even when noise exists, the algorithm provides a reasonable result.

6. Discussion and future work

In this paper we propose a generalized intrinsic mode functions and adaptive harmonic model to model oscillatory functions with fast varying instantaneous frequency. A convex optimization approach to find the time-frequency representation, referred to as Tycoon algorithm, is proposed. While the numerical results are encouraging, there are several things we should discuss.

1. While with the help of FISTA the optimization process can be carried out, it is still not numerically efficient enough for practical usage. For example, it takes about 3 minutes to finish analyzing a time series with 512 points in the laptop, but in many problems the data length is of order 10^5 or longer. Finding a more efficient strategy to carry out the optimization is an important future work.
2. While the Tycoon algorithm is not very much sensitive to the choice of parameters μ, λ and γ, how to choose an optimal set of parameters is left unanswered in the current paper.

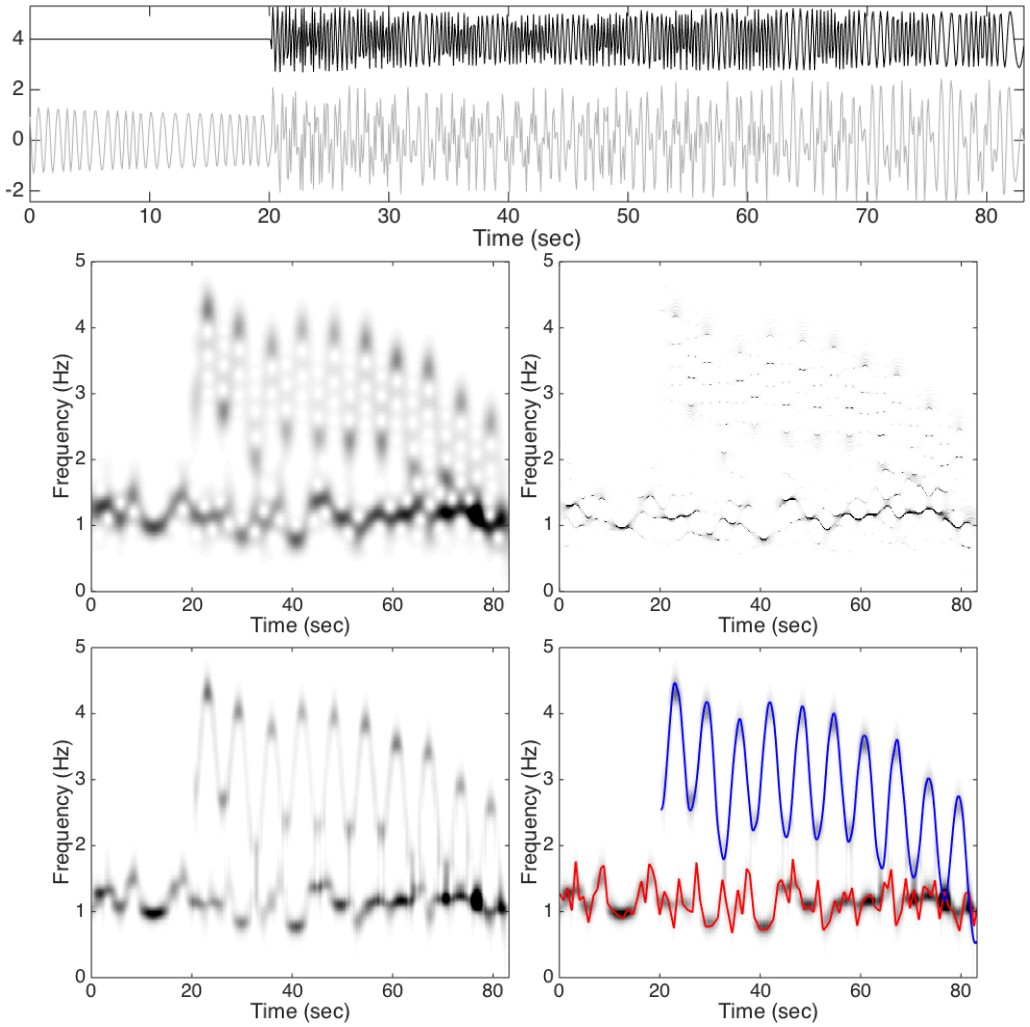

Figure 2: Top: the signal f is shown as the gray curve with f_2 superimposed as the black curve which is shifted up by 4 to increase the visualization. It is clear that the instantaneous frequency varies fast in the second component. In the second row, the short time Fourier transform (STFT) with a Gaussian window with the standard deviation $\sigma = 1$ is shown on the left and the synchrosqueezed STFT is shown on the right. In the third row, the intensity of the time frequency representation, $|\tilde{R}_f|^2$, determined by the proposed Tycoon algorithm is shown on the left; on the right hand side, the instantaneous frequencies associated with the two components are superimposed on $|\tilde{R}_f|^2$ as a red curve and a blue curve.

3. The noise behavior and influence on the Tycoon algorithm is not clear at this moment, although we could see that it is robust to the existence of noise in the numerical section. Theoretically studying the noise influence on the algorithm is important for us to better understand what we see in practice.

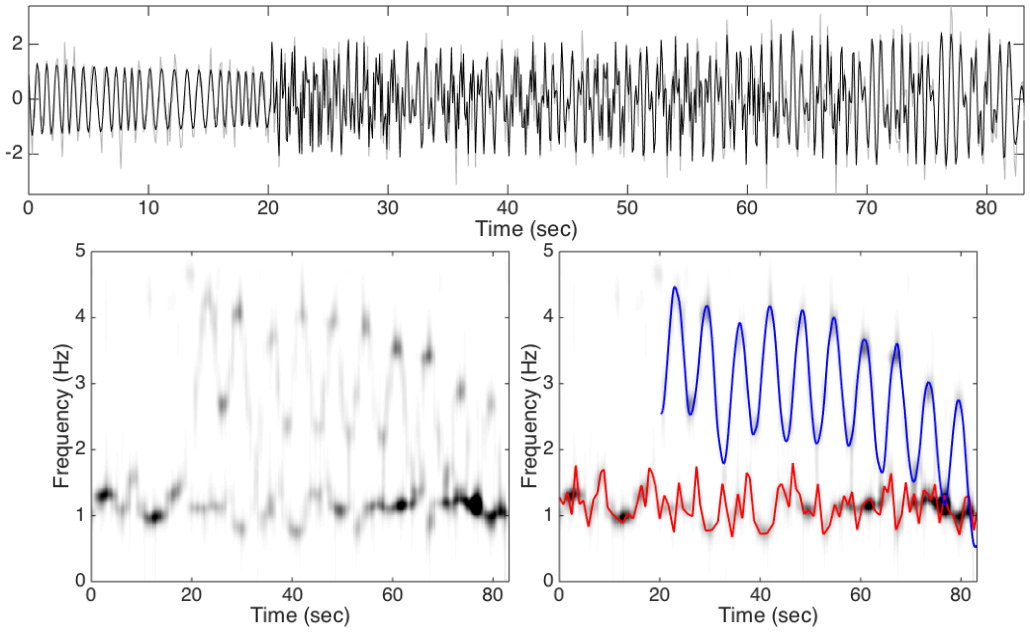

Figure 3: Top: the noisy signal Y is shown as the gray curve with the clean signal f superimposed as the black curve.In the second row, the intensity of the time frequency representation, $|\tilde{R}_Y|^2$, determined by our proposed Tycoon algorithm is shown on the left; on the right hand side, the instantaneous frequencies associated with the two components are superimposed on $|\tilde{R}_Y|^2$ as a red curve and a blue curve.

Before closing the paper, we would like to indicate an interesting finding about SST which is related to our current study. When an oscillatory signal is composed of intrinsic mode type function with slowly varying IF, it has been studied that the time-frequency representation of a function depends "weakly" on a chosen window, when the window has a small support in the Fourier domain [11, 7]. Precisely, the result depends only on the first three absolute moments of the chosen window and its derivative, but not depends on the profile of the window itself. However, the situation is different when we consider an oscillatory signal composed of gIMT function with fast varying IF. As we have shown in Figure 2, when the window is chosen to have a small support in the Fourier domain, the STFT and synchrosqueezed STFT results are not ideal. Nevertheless, nothing prevents us from trying a window with a small support in the time domain; that is, a wide support in the Fourier domain. As is shown in Figure 4, by taking the window to be a Gaussian function with the standard deviation 0.4, STFT and synchrosqueezed STFT provide reasonable results for the signal f considered in (18). Note that while we could start to see the dynamics in both STFT and synchrosqueezed STFT, the overall performance is not as good as that provided by Tycoon. Since it is not the focus of the current paper, we just indicate the possibility of achieving a better time-frequency representation by choosing a suitable window in SST, but not make effort to determine the optimal window. This kind of approach has been applied to the high energy physics field [26], where the window is manually but carefully chosen to extract the physically meaningful dynamics.

A theoretical study regarding this topic will be reported in the near future.

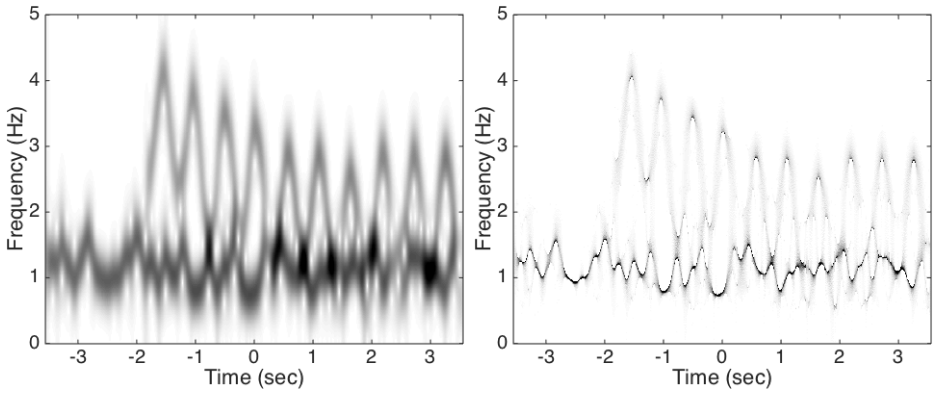

Figure 4: Left: the intensity of the short time Fourier transform (STFT) with a Gaussian window with the standard deviation $\sigma = 0.4$ is shown on the left and the intensity of the synchrosqueezed STFT is shown on the right.

7. Acknowledgement

Hau-tieng Wu would like to thank Professor Ingrid Daubechies and Professor Andrey Feuerverger for their valuable discussion.

Matthieu Kowalski benefited from the support of the "FMJH Program Gaspard Monge in optimization and operation research", and from the support to this program from EDF.

[1] F. Auger, E. Chassande-Mottin, and P. Flandrin. Making reassignment adjustable: The levenberg-marquardt approach. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 3889–3892, March 2012.

[2] F. Auger and P. Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Trans. Signal Process.*, 43(5):1068 –1089, may 1995.

[3] F. Baudin, H.-T. Wu, A. Bordessoule, J. Beck, P. Jouvet, M. Frasch, and G. Emeriaud. Impact of ventilatory modes on the breathing variability in mechanically ventilated infants'. *Frontiers in Pediatrics, section Neonatology*, page accepted for publication, 2014.

[4] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009.

[5] A. Chambolle and C. Dossal. On the convergence of the iterates of FISTA. *Preprint hal-01060130, September*, 2014.

[6] E. Chassande-Mottin, F. Auger, and P. Flandrin. Time-frequency/time-scale reassignment. In *Wavelets and signal processing*, Appl. Numer. Harmon. Anal., pages 233–267. Birkhäuser Boston, Boston, MA, 2003.

[7] Y.-C. Chen, M.-Y. Cheng, and H.-T. Wu. Nonparametric and adaptive modeling of dynamic seasonality and trend with heteroscedastic and dependent errors. *J. Roy. Stat. Soc. B*, 76:651–682, 2014.

[8] C. K. Chui, Y.-T. Lin, and H.-T. Wu. Real-time dynamics acquisition from irregular samples – with application to anesthesia evaluation. *Analysis and Applications, accepted for publication*, 2015.

[9] C. K. Chui and H.N. Mhaskar. Signal decomposition and analysis via extraction of frequencies. *Appl. Comput. Harmon. Anal.*, (0):–, 2015.

[10] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.

[11] I. Daubechies, J. Lu, and H.-T. Wu. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Appl. Comput. Harmon. Anal.*, 30:243–261, 2011.
[12] I. Daubechies and S. Maes. A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models. *Wavelets in Medicine and Biology*, pages 527–546, 1996.
[13] A. M. De Livera, R. J. Hyndman, and R. D. Snyder. Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *J. Am. Stat. Assoc.*, 106(496):1513–1527, 2011.
[14] Z. Feng, X. Chen, and M. Liang. Iterative generalized synchrosqueezing transform for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions. *Mechanical Systems and Signal Processing*, 52-53(0):360 – 375, 2015.
[15] P. Flandrin. *Time-frequency/time-scale analysis*, volume 10 of *Wavelet Analysis and its Applications*. Academic Press Inc., 1999.
[16] P. Flandrin. Time frequency and chirps. In *Proc. SPIE*, volume 4391, pages 161–175, 2001.
[17] G. Galiano and J. Velasco. On a non-local spectrogram for denoising one-dimensional signals. *Applied Mathematics and Computation*, 244:1–13, 2014.
[18] S. Guharay, G. Thakur, F. Goodman, S. Rosen, and D. Houser. Analysis of non-stationary dynamics in the financial system. *Economics Letters*, 121:454–457, 2013.
[19] R. H. Herrera, J. Han, and M. van der Baan. Applications of the synchrosqueezing transform in seismic time-frequency analysis. *Geophysics*, 79(3):V55–V64, 2014.
[20] N. E. Huang, Z. Shen, S. R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A*, 454(1971):903–995, 1998.
[21] D. Iatsenko, A. Bernjak, T. Stankovski, Y. Shiogai, P.J. Owen-Lynch, P. B. M. Clarkson, P. V. E. McClintock, and A. Stefanovska. Evolution of cardiorespiratory interactions with age Evolution of cardiorespiratory interactions with age. *Phil. Trans. R. Soc. A*, 371(20110622):1–18, 2013.
[22] A Jahangir, Lee V, PA Friedman, JM Trusty, DO Hodge, and et al. Long-term progression and outcomes with aging in patients with lone atrial fibrillation: a 30-year follow-up study. *Circulation*, 115:3050–3056, 2007.
[23] K. Kodera, R. Gendrin, and C. Villedary. Analysis of time-varying signals with small bt values. *IEEE Trans. Acoust., Speech, Signal Processing*, 26(1):64 – 76, feb 1978.
[24] C. Li and M. Liang. A generalized synchrosqueezing transform for enhancing signal time-frequency representation. *Signal Processing*, 92(9):2264 – 2274, 2012.
[25] P.-C. Li, Y.-L. Sheu, C. Laughlin, , and S.-I Chu. Role of laser-driven electron-multirescattering in resonance-enhanced below-threshold harmonic generation in he atoms. *Phy. Rev. A.*, 90:041401(R), 2013.
[26] P.-C. Li, Y.-L. Sheu, C. Laughlin, and S.-I Chu. Dynamical origin of near- and below-threshold harmonic generation of cs in an intense mid-infrared laser field. 2014.
[27] Y.-T. Lin. *The Modeling and Quantification of Rhythmic to Non-rhythmic Phenomenon in Electrocardiography during Anesthesia*. PhD thesis, National Taiwan University, 2015. ArXiv 1502.02764.
[28] Y.-T. Lin, H.-T. Wu, J. Tsao, H.-W. Yien, and S.-S. Hseu. Time-varying spectral analysis revealing differential effects of sevoflurane anaesthesia: non-rhythmic-to-rhythmic ratio. *Acta Anaesthesiologica Scandinavica*, 58:157–167, 2014.
[29] S. Mann and S. Haykin. The chirplet transform: physical considerations. *Signal Process. IEEE Trans.*, 43(11):2745–2761, 1995.
[30] S. Meignen, T. Oberlin, and S. McLaughlin. A new algorithm for multicomponent signals analysis based on synchrosqueezing: With an application to signal sampling and denoising. *IEEE Trans. Signal Process.*, 60(12):5787–5798, 2012.
[31] N. Murata, S. Ikeda, and A. Ziehe. An approach to blind source separation based on temporal structure of speech signals. *Neurocomputing*, 41(1):1–24, 2001.
[32] Y.-L. Sheu, L. Y. Hsu, H. T. Wu, P.-Ch. Li, and S.-I Chu. A new time-frequency method to reveal quantum dynamics of atomic hydrogen in intense laser pulses: Synchrosqueezing transform. *AIP Advances*, 4:117138, 2014.
[33] R.G. Stockwell, L. Mansinha, and R.P. Lowe. Localization of the complex spectrum: the S transform. *Signal Process. IEEE Trans.*, 44(4):998–1001, 1996.
[34] P. Tavallali, T. Hou, and Z. Shi. Extraction of intrawave signals using the sparse time-frequency representation method. *Multiscale Modeling & Simulation*, 12(4):1458–1493, 2014.
[35] G. Thakur. The synchrosqueezing transform for instantaneous spectral analysis. In *Excursions in Harmonic Analysis vol. 3*. Springer, 2014.
[36] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of optimization theory and applications*, 109(3):475–494, 2001.

[37] P. Tseng. Approximation accuracy, gradient methods, and error bound for structured convex optimization. *Mathematical Programming*, 125(2):263–295, 2010.
[38] T. Vatter, H.-T. Wu, V. Chavez-Demoulin, and B. Yu. Non-parametric estimation of intraday spot volatility: disentangling instantaneous trend and seasonality. *SSRN e-prints*, 2013. 2330159.
[39] P. Wang, Gao. J., and Z. Wang. Time-frequency analysis of seismic data using synchrosqueezing transform. *Geoscience and Remote Sensing Letters, IEEE*, 11(12):2042–2044, 2014.
[40] H.-T. Wu. *Adaptive Analysis of Complex Data Sets*. PhD thesis, Princeton University, 2011.
[41] H.-T. Wu. Instantaneous frequency and wave shape functions (I). *Appl. Comput. Harmon. Anal.*, 35:181–199, 2013.
[42] H.-T. Wu, Y.-H. Chan, Y.-T. Lin, and Y.-H. Yeh. Using synchrosqueezing transform to discover breathing dynamics from ecg signals. *Appl. Comput. Harmon. Anal.*, 36:354–359, 2014.
[43] H.-T. Wu, S.-S. Hseu, M.-Y. Bien, Y. R. Kou, and I. Daubechies. Evaluating physiological dynamics via synchrosqueezing: Prediction of ventilator weaning. *IEEE Trans. Biomed. Eng.*, 61:736–744, 2013.
[44] Z. Wu and N. E. Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Adapt. Data Anal.*, 1:1 – 41, 2009.
[45] H. Yang. Synchrosqueezed Wave Packet Transforms and Diffeomorphism Based Spectral Analysis for 1D General Mode Decompositions. *Appl. Comput. Harmon. Anal.*, 2014.
[46] W. I. Zangwill. *Nonlinear programming: a unified approach*, volume 196. Prentice-Hall Englewood Cliffs, NJ, 1969.

Appendix A. Proof of Theorem 2.1

Suppose

$$g(t) = a(t)\cos\phi(t) = (a(t)+\alpha(t))\cos(\phi(t)+\beta(t)) \in \mathcal{Q}_\epsilon^{c_1,c_2}. \tag{A.1}$$

Clearly we know $\alpha \in C^1(\mathbb{R})$, $\beta \in C^3(\mathbb{R})$. By the definition of $\mathcal{Q}_\epsilon^{c_1,c_2}$, we have

$$\inf_{t\in\mathbb{R}} a(t) > c_1, \ \sup_{t\in\mathbb{R}} a(t) < c_2, \tag{A.2}$$

$$\inf_{t\in\mathbb{R}} \phi'(t) > c_1, \ \sup_{t\in\mathbb{R}} \phi'(t) < c_2, \ |\phi''(t)| \le c_2 \tag{A.3}$$

$$|a'(t)| \le \epsilon\phi'(t), \ |\phi'''(t)| \le \epsilon\phi'(t) \tag{A.4}$$

and

$$\inf_{t\in\mathbb{R}} [a(t)+\alpha(t)] > c_1, \ \sup_{t\in\mathbb{R}} [a(t)+\alpha(t)] < c_2, \tag{A.5}$$

$$\inf_{t\in\mathbb{R}} [\phi'(t)+\beta'(t)] > c_1, \ \sup_{t\in\mathbb{R}} [\phi'(t)+\beta'(t)] < c_2, \ |\phi''(t)+\beta''(t)| \le c_2 \tag{A.6}$$

$$|a'(t)+\alpha'(t)| \le \epsilon(\phi'(t)+\beta'(t)), \ |\phi'''(t)+\beta'''(t)| \le \epsilon(\phi'(t)+\beta'(t)). \tag{A.7}$$

The proof is divided into two parts. The first part is determining the restrictions on the possible β and α based on the positivity condition of $\phi'(t)$ and $a(t)$, which is independent of the conditions (A.4) and (A.7). The second part is to control the amplitude of β and α, which depends on the conditions (A.4) and (A.7).

First, based on the conditions (A.2), (A.3), (A.5) and (A.6), we show how β and α are restricted. By the monotonicity of $\phi(t)$ based on the condition (A.3), define $t_m \in \mathbb{R}$, $m \in \mathbb{Z}$, so that $\phi(t_m) = (m+1/2)\pi$ and $s_m \in \mathbb{R}$, $m \in \mathbb{Z}$, so that $\phi(s_m) = m\pi$. In other words, we have

$$g(t_m) = 0 \quad \text{and} \quad g(s_m) = (-1)^m a(s_m).$$

Thus, for any $n \in \mathbb{Z}$, when $t = t_n$, we have

$$\begin{aligned} &(a(t_n) + \alpha(t_n)) \cos(\phi(t_n) + \beta(t_n)) \\ &= (a(t_n) + \alpha(t_n)) \cos[n\pi + \pi/2 + \beta(t_n)] \\ &= a(t_n) \cos(n\pi + \pi/2) = 0, \end{aligned} \tag{A.8}$$

where the second equality comes from (A.1). This leads to $\beta(t_n) = k_n\pi$, $k_n \in \mathbb{Z}$, since $a(t_n) + \alpha(t_n) > 0$ by (A.6).

Claim Appendix A.1. *k_n are the same for all $n \in \mathbb{Z}$ and k_n are even. As changing the phase function globally by $2l\pi$ will not change the value of $g(t_l)$, where $l \in \mathbb{Z}$, we could assume that $\beta(t_l) = 0$ for all $l \in \mathbb{Z}$.*

Proof. Suppose there exists t_n so that $\beta(t_n) = k\pi$ and $\beta(t_{n+1}) = (k+l)\pi$, where $k, l \in \mathbb{Z}$ and $l > 0$. In other words, we have $\phi(t_{n+1}) = \phi(t_n) + (l+1)\pi$. By the smoothness of β, we know there exists at least one $t' \in (t_n, t_{n_1})$ so that $\phi(t') + \beta(t') = (n+3/2)\pi$, but this is absurd since it means that $(a(t) + \alpha(t)) \cos(\phi(t) + \beta(t))$ will change sign in (t_n, t_{n+1}) while $a(t)\cos(\phi(t))$ will not.

Suppose k_n is a fixed odd integer k, then since $\beta \in C^3(\mathbb{R})$ and $\beta(t_n) = \beta(t_{n+1}) = k\pi$, there exists $t' \in (t_n, t_{n+1})$ so that $\beta(t') = k\pi$ and hence

$$a(t')\cos(\phi(t')) \ = (a(t') + \alpha(t'))\cos(\phi(t') + \beta(t')) = -(a(t') + \alpha(t'))\cos(\phi(t')),$$

which is again absurd since $\cos(\phi(t')) \neq 0$ and the amplitudes are positive by (A.2) and (A.5). We thus obtain the second claim. □

Claim Appendix A.2. *$\beta'(t)$ is constant or changes sign inside $[t_n, t_{n+1}]$ for all $n \in \mathbb{Z}$. Furthermore, $|\beta(t') - \beta(t'')| < \pi$ for any $t', t'' \in [t_m, t_{m+1}]$ for all $m \in \mathbb{Z}$.*

Proof. By the fundamental theorem of calculus and the fact that $\beta(t_n) = \beta(t_{n+1}) = 0$, we know that

$$0 = \beta(t_{n+1}) - \beta(t_n) = \int_{t_n}^{t_{n+1}} \beta'(u)\mathrm{d}u.$$

which implies the first argument. Also, due to the monotonicity of $\phi + \beta$ (A.6), that is, $(n+1/2)\pi = \phi(t_n) + \beta(t_n) < \phi(t') + \beta(t') < \phi(t_{n+1}) + \beta(t_{n+1}) = (n+3/2)\pi$ for all t, we have the second claim

$$|\beta(t') - \beta(t'')| < \pi.$$

Indeed, if $|\beta(t') - \beta(t'')| \geq \pi$, for some $t' < t'' \in [t_n, t_{n+1}]$, we get an contradiction since $\phi(t'') + \beta(t'') \notin [(n+1/2)\pi, (n+3/2)\pi]$ while $\phi(t') + \beta(t') \in [(n+1/2)\pi, (n+3/2)\pi]$. □

Claim Appendix A.3. *$\alpha(s_m) \geq 0$ for all $m \in \mathbb{Z}$. If $\alpha(s_m) = 0$, then $\beta(s_m) = 0$; if $\alpha(s_m) > 0$, then $|\beta(s_m)| < \pi/2$.*

Proof. When $t = s_m$, we have

$$\begin{aligned} (-1)^m a(s_m) &= a(s_m)\cos(m\pi) \\ &= (a(s_m) + \alpha(s_m))\cos[m\pi + \beta(s_m)] \\ &= (-1)^m (a(s_m) + \alpha(s_m))\cos(\beta(s_m)), \end{aligned} \tag{A.9}$$

where the second equality comes from (A.1), which leads to $\alpha(s_m) \geq 0$ since $|\cos(\beta(s_m))| \leq 1$.

Notice that (A.9) implies that $\beta(s_m) = 2k_m\pi$, where $k_m \in \mathbb{Z}$, if and only if $\alpha(s_m) = 0$. Without loss of generality, assume $k_m > 0$. Since $\beta \in C^3(\mathbb{R})$, there exists $t' \in (t_{m-1}, s_m)$ so that $\beta(t') = \pi$ and hence

$$a(t')\cos(\phi(t')) = (a(t') + \alpha(t'))\cos(\phi(t') + \beta(t')) = -(a(t') + \alpha(t'))\cos(\phi(t')),$$

which is absurd since $\cos(\phi(t')) \neq 0$ and the positive amplitudes by (A.2) and (A.5). Thus we conclude that $\beta(s_m) = 0$.

To show the last part, note that when $\alpha(s_m) > 0$, $0 < \cos(\beta(s_m)) = \frac{a(s_m)}{a(s_m)+\alpha(s_m)} < 1$ by (A.9). Thus, we know $\beta(s_m) \in (-\pi/2, \pi/2) + 2n_m\pi$, where $n_m \in \mathbb{Z}$. By the same argument as in the above, if $n_m > 0$, there exists $t' \in (t_{m-1}, s_m)$ so that $\beta(t') = \pi$ and hence

$$a(t')\cos(\phi(t')) = (a(t') + \alpha(t'))\cos(\phi(t') + \beta(t')) = -(a(t') + \alpha(t'))\cos(\phi(t')),$$

which is absurd since $\cos(\phi(t')) \neq 0$ and the positive amplitudes by (A.2) and (A.5). □

Claim Appendix A.4. $\frac{a(t_n)}{a(t_n)+\alpha(t_n)} = \frac{\phi'(t_n)+\beta'(t_n)}{\phi'(t_n)}$ *for all* $n \in \mathbb{Z}$. *In particular,* $\alpha(t_n) = 0$ *if and only if* $\beta'(t_n) = 0$ *for all* $n \in \mathbb{Z}$.

Proof. For $0 < x \ll 1$, we have

$$(a(t_n + x) + \alpha(t_n + x))\cos(\phi(t_n + x) + \beta(t_n + x)) = a(t_n + x)\cos(\phi(t_n + x)),$$

which means that

$$\frac{a(t_n + x)}{a(t_n + x) + \alpha(t_n + x)} = \frac{\cos(\phi(t_n + x) + \beta(t_n + x))}{\cos(\phi(t_n + x))}.$$

By the smoothness of ϕ and β, as $x \to 0$, the right hand side becomes

$$\begin{aligned}
&\lim_{x\to 0} \frac{\cos(\phi(t_n + x) + \beta(t_n + x))}{\cos(\phi(t_n + x))} \\
&= \lim_{x\to 0} \frac{(\phi'(t_n + x) + \beta'(t_n + x)\sin(\phi(t_n + x) + \beta(t_n + x))}{\phi'(t_n + x)\sin(\phi(t_n + x))} \\
&= \frac{\phi'(t_n) + \beta'(t_n)}{\phi'(t_n)}.
\end{aligned}$$

Thus, since $a(t_n + x) + \alpha(t_n + x) > 0$ and $a(t_n + x) > 0$ for all x, we have

$$\frac{a(t_n)}{a(t_n) + \alpha(t_n)} = \frac{\phi'(t_n) + \beta'(t_n)}{\phi'(t_n)}.$$

□

Claim Appendix A.5. $\beta''(t)$ *is* 0 *or changes sign inside* $[t_n, t_{n+1}]$ *for all* $n \in \mathbb{Z}$.

Proof. This is clear since $\beta'(t)$ is constant or changes sign inside $[t_n, t_{n+1}]$ for all $n \in \mathbb{Z}$. □

To finish the second part of the proof, we have to consider the conditions (A.4) and (A.7).

Claim Appendix A.6. $|\alpha(t)| \leq 2\pi\epsilon$ *for all* $t \in \mathbb{R}$.

Proof. Suppose there exists t' so that $\alpha(t') > 2\pi\epsilon$. The case $\alpha(t') < -2\pi\epsilon$ can be proved in the same way. Take $m \in \mathbb{Z}$ so that $t' \in (s_m, s_{m+1})$. Without loss of generality, we assume $t' < t_m$. From (A.4) and (A.7) we have

$$|\alpha'(t)| \leq \epsilon(2\phi'(t) + \beta'(t)).$$

Thus, if we take $t \in (s_m, t')$, we have by the fundamental theorem of calculus

$$\begin{aligned} |\alpha(t') - \alpha(t)| \ &\leq \int_t^{t'} |\alpha'(u)|\mathrm{d}u \leq \epsilon[2\phi(t') - 2\phi(t) + \beta(t') - \beta(t)] \\ &\leq \epsilon[(\phi(s_{m+1}) + \beta(s_{m+1}) - \phi(s_m) - \beta(s_m)) + (\phi(s_{m+1}) - \phi(s_m))] \leq 2\pi\epsilon, \end{aligned}$$

where the last inequality holds due to the fact that $\phi + \beta$ and ϕ are both monotonic and Claim Appendix A.2. This fact leads to $\alpha(t_m) > 0$. For each $m \in \mathbb{Z}$, since $|\beta(s_m)| < \pi/2$ by Claim Appendix A.3, we have

$$1 - \frac{1}{2}\beta(s_m)^2 < \frac{1}{1 + \frac{\alpha(s_m)}{a(s_m)}} < 1 - \frac{1}{2}\beta(s_m)^2 + \frac{1}{24}\beta(s_m)^4,$$

which comes from Taylor's expansion of the cosine function around $\phi(s_m)$. Thus, we know

$$|\beta(t_m)| > \sqrt{\frac{2\alpha(t_m)}{a(t_m) + \alpha(t_m)}} > 0,$$

which contradicts to the fact that $\beta(t_m) = 0$ shown in Claim Appendix A.1. □

Thus we obtain the control of the amplitude. Note that the proof does not depend on the condition about β''.

Claim Appendix A.7. $|\beta''(t)| \leq 2\pi\epsilon$, $|\beta'(t)| \leq \frac{2\pi\epsilon}{c_1}$ *and* $|\beta(t)| \leq \frac{2\pi\epsilon}{c_1^2}$ *for all* $t \in \mathbb{R}$.

Proof. Suppose there existed $t' \in (t_m, t_{m+1})$ for some $m \in \mathbb{Z}$ so that $|\beta''(t')| > 3\pi\epsilon$. Without loss of generality, we assume $\beta''(t') > 0$. From (A.4) and (A.7) we have

$$|\beta'''(t)| \leq \epsilon(2\phi'(t) + \beta'(t)).$$

Thus, by the fundamental theorem of calculus, for any $t \in (t_m, t')$, we know

$$|\beta''(t') - \beta''(t)| \leq \int_t^{t'} |\beta'''(u)|\mathrm{d}u \leq \epsilon \int_t^{t'} (2\phi'(t) + \beta'(t))\mathrm{d}u \leq 2\pi\epsilon,$$

where the last inequality holds due to Claim Appendix A.2 and the fact that $\phi(t') - \phi(t) \leq \phi(t_{m+1}) - \phi(t_m) = \pi$ and $|\beta(t') - \beta(t)| < \pi$ from Claim Appendix A.2. Similarly, we have that for all $t \in (t', t_{m+1})$, $|\beta''(t') - \beta''(t)| \leq 2\pi\epsilon$. Thus, $\beta''(t) > 0$ for all

$t \in [t_m, t_{m+1}]$, which contradicts the fact that $\beta''(t)$ must change sign inside $[t_m, t_{m+1}]$ by Claim Appendix A.5.

With the upper bound of $|\beta''|$, we immediately have for all $t \in [t_m, t_{m+1}]$ that

$$|\beta'(t)| = |\beta'(t) - \beta'(t_m)| \leq \int_{t_m}^{t} |\beta''(u)| \mathrm{d}u \leq 2\pi(t - t_m)\epsilon \leq \frac{2\pi\epsilon}{c_1},$$

where the last inequality holds by $t - t_m \leq t_{m+1} - t_m \leq 1/c_1$. Similarly, we have the bound for β. □

Appendix B. Proof of Theorem 2.2

When there are more than one gIMT in a given oscillatory signal $f \in \mathcal{Q}_{\epsilon,d}^{c_1,c_2}$, we loss the control of the hinging points for each gIMT like those, t_m and s_m, in Theorem 2.1. So the proof will be more qualitative. Suppose $f = \tilde{f} \in \mathcal{Q}_{\epsilon,d}^{c_1,c_2}$, where

$$f(t) = \sum_{l=1}^{N} a_l(t) \cos[2\pi\phi_l(t)], \quad \tilde{f}(t) = \sum_{l=1}^{M} A_l(t) \cos[2\pi\varphi_l(t)].$$

Fix $t_0 \in \mathbb{R}$. Denote $f_{t_0} := \sum_{l=1}^{N} f_{t_0,l}$, $\tilde{f}_{t_0} := \sum_{l=1}^{M} \tilde{f}_{t_0,l}$,

$$f_{t_0,l}(t) := a_l(t_0) \cos\left[2\pi\left(\phi_l(t_0) + \phi_l'(t_0)(t - t_0) + \phi_l''(t_0)\frac{(t - t_0)^2}{2}\right)\right]$$

and

$$\tilde{f}_{t_0,l}(t) := A_l(t_0) \cos\left[2\pi\left(\varphi_l(t_0) + \varphi_l'(t_0)(t - t_0) + \varphi_l''(t_0)\frac{(t - t_0)^2}{2}\right)\right].$$

Note that $f_{t_0,l}$ is an approximation of $a_l(t)\cos[2\pi\phi_l(t)]$ near t_0 based on the assumption of $\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$, where we approximate the amplitude $a_l(t)$ by the zero-th order Taylor expansion and the phase function $\phi_l(t)$ by the second order Taylor expansion. To simplify the proof, we focus on the case that $|\phi_l''(t_0)| > \epsilon|\phi_l'(t_0)|$ and $|\varphi_l''(t_0)| > \epsilon|\varphi_l'(t_0)|$ for all l. For the case when there is one or more l so that $|\phi_l''(t_0)| \leq \epsilon|\phi_l'(t_0)|$, the proof follows the same line while we approximate the phases of these oscillatory components by the first order Taylor expansion.

Recall that the short time Fourier transform (STFT) of a given tempered distribution $\mathsf{f} \in \mathcal{S}'$ associated with a Schwartz function $g \in \mathcal{S}$ as the window function is defined as

$$V_{\mathsf{f}}^{(\mathsf{g})}(t, \eta) := \int_{\mathbb{R}} \mathsf{f}(x) g(x - t) e^{-i2\pi\eta x} \mathrm{d}x.$$

Note that by definition $f, f_{t_0}, \tilde{f}_{t_0} \in \mathcal{S}'$. To prove the theorem, we need the following claim about the STFT.

Claim Appendix B.1. *For a fixed $t_0 \in \mathbb{R}$, we have*

$$\left|V_f^{(g)}(\tau, \eta) - V_{f_{t_0}}^{(g)}(\tau, \eta)\right| = O(\epsilon).$$

where C is a universal constant depending on c_1, c_2 and d.

Proof. Fix a time $t_0 \in \mathbb{R}$. By the same argument as that in [11, 7] and the conditions of $\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$, we immediately have

$$\begin{aligned}
&|V_f^{(g)}(\tau,\eta) - V_{f_{t_0}}^{(g)}(\tau,\eta)| \\
=& \left| \int_{\mathbb{R}} (f(t) - f_{t_0}(t)) g(t-\tau) e^{-i2\pi\eta t} \mathrm{d}t \right| \\
\leq& \sum_{l=1}^{N} \left| \int_{\mathbb{R}} (a_l(t) - a_l(t_0)) \cos[2\pi\phi_l(t)] g(t-t_0) e^{-i2\pi\eta t} \mathrm{d}t \right| \\
&+ \sum_{l=1}^{N} \left| \int_{\mathbb{R}} a_l(t_0) \big(\cos[2\pi\phi_l(t)] - \cos[2\pi(\phi(t_0) + \phi'(t_0)(t-t_0) \right. \\
&\qquad \left. + \frac{1}{2}\phi''(t_0)(t-t_0)^2)] \big) g(t-t_0) e^{-i2\pi\eta t} \mathrm{d}t \right| \\
=& O(\epsilon),
\end{aligned}$$

where the last term depends only on the first few absolute moments of g and g', d, c_1 and c_2. □

With this claim, we know in particular that $V_f^{(g)}(t_0,\eta) = V_{f_{t_0}}^{(g)}(t_0,\eta) + O(\epsilon)$. As a result, the spectrogram of f and f_{t_0} are related by

$$|V_f^{(g)}(t_0,\eta)|^2 = |V_{f_{t_0}}^{(g)}(t_0,\eta)|^2 + O(\epsilon).$$

Indeed, we have

$$|V_{f_{t_0}}^{(g)}(t_0,\eta)| \leq \sum_{l=1}^{N} a_l(t_0) I_{1,0}.$$

Next, recall that the spectrogram of a signal is intimately related to the Wigner-Ville distribution in the following way

$$|V_{f_{t_0}}^{(g)}(\tau,\eta)|^2 = \int\int WV_{f_{t_0}}(x,\xi) WV_g(x-\tau,\xi-\eta) \mathrm{d}x \mathrm{d}\xi,$$

where the Wigner-Ville distribution of a function h in the suitable space is defined as

$$WV_h(x,\xi) := \int h(x+\tau/2) h^*(x-\tau/2) e^{-i2\pi\tau\xi} \mathrm{d}\tau.$$

Claim Appendix B.2. *Take $g(t) = (2\sigma)^{1/4} \exp\{-\pi\sigma t^2\}$, where $\sigma > 0$. When d is large enough described in (B.2) with the properly chosen $\sigma_0 > 0$, where σ_0 is the minimizer of (B.2), we have*

$$|V_f^{(g)}(t_0,\eta)|^2 = L(t_0,\eta) + \epsilon \quad \textit{and} \quad |V_{\tilde{f}}^{(g)}(t_0,\eta)|^2 = \widetilde{L}(t_0,\eta) + \epsilon,$$

where

$$L(t_0,\eta) := \sum_{l=1}^{N} a_l^2(t_0) \sqrt{\frac{\sigma}{2(\sigma^2 + \phi_l''(t_0)^2)}} \exp\left\{ -\frac{2\pi\sigma(\phi_l'(t_0) - \eta)^2}{\sigma^2 + \phi_l''(t_0)^2} \right\}$$

and

$$\widetilde{L}(t_0,\eta) := \sum_{l=1}^{M} A_l^2(t_0)\sqrt{\frac{\sigma}{2(\sigma^2+\varphi_l''(t_0)^2)}}\exp\left\{-\frac{2\pi\sigma(\varphi_l'(t_0)-\eta)^2}{\sigma^2+\varphi_l''(t_0)^2}\right\}.$$

Proof. By a direct calculation, the Wigner-Ville distribution of the Gaussian function $g(t) = (2\sigma)^{1/4}\exp\left\{-\pi\sigma t^2\right\}$ with the unit energy, where $\sigma > 0$, is

$$WV_g(x,\xi) = 2\exp\left\{-2\pi\left(\sigma x^2 + \frac{\xi^2}{\sigma}\right)\right\};$$

similarly, the Wigner-Ville distribution of $f_{t_0,l}$ is

$$WV_{f_{t_0,l}}(x,\xi) = a_l^2(t_0)\delta_{\phi_l'(t_0)+\phi_l''(t_0)(x-t_0)}(\xi).$$

Thus, we know

$$\begin{aligned}
&\left|V_{f_{t_0,l}}^{(g)}(t_0,\eta)\right|^2 \qquad\qquad (B.1)\\
=&\int\int WV_{f_{t_0,l}}(x,\xi)WV_g(x-t_0,\xi-\eta)\mathrm{d}x\mathrm{d}\xi\\
=&\int\int\left(a_l^2(t_0)\delta_{\phi_l'(t_0)+\phi_l''(t_0)(x-t_0)}(\xi)\right)2\exp\left\{-2\pi\left(\sigma(x-t_0)^2+\frac{(\xi-\eta)^2}{\sigma}\right)\right\}\mathrm{d}\xi\mathrm{d}x\\
=&2a_l^2(t_0)\int\exp\left\{-2\pi\left(\sigma(x-t_0)^2+\frac{(\phi_l'(t_0)+\phi_l''(t_0)(x-t_0)-\eta)^2}{\sigma}\right)\right\}\mathrm{d}x\\
=&a_l^2(t_0)\sqrt{\frac{\sigma}{2(\sigma^2+\phi_l''(t_0)^2)}}\exp\left\{-\frac{2\pi\sigma}{\sigma^2+\phi_l''(t_0)^2}(\phi_l'(t_0)-\eta)^2\right\}.
\end{aligned}$$

Thus, we have the expansion of $\sum_{l=1}^{N}|V_{f_{t_0,l}}^{(g)}(t_0,\eta)|^2$, which is $L(t_0,\eta)$. Next, we clearly have

$$\begin{aligned}
&\left|V_{f_{t_0}}^{(g)}(\tau,\eta)|^2 - \sum_{l=1}^{N}|V_{f_{t_0,l}}^{(g)}(\tau,\eta)|^2\right| = \left|\Re\sum_{k\neq l}V_{f_{t_0,l}}^{(g)}(\tau,\eta)\overline{V_{f_{t_0,k}}^{(g)}(\tau,\eta)}\right|\\
\leq&\sum_{k\neq l}\left|V_{f_{t_0,l}}^{(g)}(\tau,\eta)\right|\left|V_{f_{t_0,k}}^{(g)}(\tau,\eta)\right|.
\end{aligned}$$

To bound the right hand side, note that (B.1) implies

$$\left|V_{f_{t_0,l}}^{(g)}(t_0,\eta)\right| = a_l(t_0)\left(\frac{\sigma}{2(\sigma^2+\phi_l''(t_0)^2)}\right)^{1/4}\exp\left\{-\frac{\pi\sigma(\phi_l'(t_0)-\eta)^2}{\sigma^2+\phi_l''(t_0)^2}\right\}.$$

As a result, $\sum_{k\neq l}\left|V_{f_{t_0,l}}^{(g)}(\tau,\eta)\right|\left|V_{f_{t_0,k}}^{(g)}(\tau,\eta)\right|$ becomes

$$\sum_{k\neq l}\frac{a_k(t_0)a_l(t_0)\sigma^{1/2}}{\left(4(\sigma^2+\phi_k''(t_0)^2)(\sigma^2+\phi_l''(t_0)^2)\right)^{1/4}}\exp\left\{-\pi\sigma\left(\frac{(\phi_k'(t_0)-\eta)^2}{\sigma^2+\phi_k''(t_0)^2}+\frac{(\phi_l'(t_0)-\eta)^2}{\sigma^2+\phi_l''(t_0)^2}\right)\right\},$$

which is a smooth and bounded function of η. As each component is product of two Gaussian functions, the maximum of each component should be achieved around ϕ_l' for some l. Thus, we have the following bound

$$\begin{aligned}&\sum_{k\neq l}^{N}\left|V^{(g)}_{f_{t_0,l}}(\tau,\eta)\right|\left|V^{(g)}_{f_{t_0,k}}(\tau,\eta)\right|\\ \leq&\frac{N^2}{2}\frac{a_k(t_0)a_l(t_0)\sigma^{1/2}}{(4(\sigma^2+\phi_k''(t_0)^2)(\sigma^2+\phi_l''(t_0)^2))^{1/4}}\exp\left\{-\pi\sigma\left(\frac{(\phi_k'(t_0)-\phi_l'(t_0))^2}{\sigma^2+\phi_k''(t_0)^2}\right)\right\}\end{aligned}$$

for some $k,l=1,\ldots,N$, which is by ϵ. Thus, by a direct calculation, we have the following bound

$$\frac{(c_2-c_1)^2}{2d^2}c_2^2\frac{1}{\sqrt{2\sigma}}\exp\left\{\frac{-\pi\sigma d^2}{\sigma^2+c_2^2}\right\}\leq\epsilon,$$

which leads to the following rough bound of d:

$$d^2\geq-\frac{\sigma^2+c_2^2}{\pi\sigma}\ln\frac{2\sqrt{2\sigma}\epsilon}{(c_2-c_1)^2c_2^2}. \tag{B.2}$$

Thus, by minimizing the right hand side by taking a suitable $\sigma>0$, we have shown that when d is large enough, the interference term is of order ϵ. We have finished the proof. □

Since t_0 is arbitrary in the above argument and the spectrogram of a function is unique, we have $|V_f^{(g)}(t_0,\eta)|^2=|V_{\tilde{f}}^{(g)}(t_0,\eta)|^2$ and hence

$$|L(t_0,\eta)-\widetilde{L}(t_0,\eta)|=O(\epsilon). \tag{B.3}$$

With the above claim, we now show $M=N$.

Claim Appendix B.3. *$M=N$.*

Proof. With σ_0, by Claim Appendix B.2, for each $l=1,\ldots,N$, there exists a subinterval $I_l(t_0)$ around $\phi_l'(t_0)$ so that on $I_l(t_0)$, $L(t_0,\eta)>\frac{a_l^2(t_0)}{2\sqrt{2(1+\phi_l''(t_0)^2)}}>\frac{c_1^2}{2\sqrt{2+2c_2^2}}$. Similarly, for each $l=1,\ldots,M$, there exists a subinterval $J_l(t_0)$ around $\varphi_l'(t_0)$ so that on $J_l(t_0)$, $\widetilde{L}(t_0,\eta)>\frac{A_l^2(t_0)}{2\sqrt{2(1+\varphi_l''(t_0)^2)}}>\frac{c_1^2}{2\sqrt{2+2c_2^2}}$. Thus, when ϵ is small enough, in particular, $\epsilon\ll\frac{c_1^2}{2\sqrt{2+2c_2^2}}$, the equality in (B.3) cannot hold if $M\neq N$. □

With this claim, we obtain the first part of the proof, and hence the equality

$$f(t)=\sum_{l=1}^{N}a_l(t)\cos[2\pi\phi_l(t)]=\sum_{l=1}^{N}A_l(t)\cos[2\pi\varphi_l(t)]\in\mathcal{Q}_{\epsilon,d}^{c_1,c_2}. \tag{B.4}$$

Now we proceed to finish the proof. Note that it is also clear that the sets $I_l(t_0)$ and $J_l(t_0)$ defined in the proof of Claim Appendix B.3 satisfy that $I_l(t_0)\cap I_k(t_0)=\emptyset$ for all $l\neq k$. Also, $I_l(t_0)\cap J_l(t_0)\neq\emptyset$ and $I_l(t_0)\cap J_k(t_0)=\emptyset$ for all $l\neq k$. Indeed, if $k=l+1$

and we have $I_l(t_0) \cap J_{l+1}(t_0) \neq \emptyset$, then $L(t_0, \eta) > \frac{c_1^2}{2\sqrt{2+2c_2^2}}$ on $J_{l+1}(t_0) \backslash I_l(t_0)$, which leads to the contradiction. By the ordering of $\phi_l'(t_0)$ and hence the ordering of $I_l(t_0)$, we have the result.

Take $\ell = 1$, $\sigma = 1$ and $\eta = \phi_1'(t_0)$. By Claim Appendix B.2, when d is large enough, on $I_1(t_0)$ we have

$$\frac{a_1^2(t_0)}{\sqrt{2(1+\phi_1''(t_0)^2)}} = \frac{A_1^2(t_0)}{\sqrt{2(1+\varphi_1''(t_0)^2)}} \exp\left\{-\frac{2\pi(\varphi_1'(t_0)-\phi_1'(t_0))^2}{1+\varphi_1''(t_0)^2}\right\} + O(\epsilon), \tag{B.5}$$

which leads to the fact that

$$\left|\frac{a_1^2(t_0)}{\sqrt{2(1+\phi_1''(t_0)^2)}} - \frac{A_1^2(t_0)}{\sqrt{2(1+\varphi_1''(t_0)^2)}}\right| = O(\epsilon). \tag{B.6}$$

Indeed, without loss of generality, assume $\frac{a_1^2(t_0)}{\sqrt{2(1+\phi_1''(t_0)^2)}} \geq \frac{A_1^2(t_0)}{\sqrt{2(1+\varphi_1''(t_0)^2)}}$ and we have

$$\begin{aligned}
&\frac{a_1^2(t_0)}{\sqrt{2(1+\phi_1''(t_0)^2)}} - \frac{A_1^2(t_0)}{\sqrt{2(1+\varphi_1''(t_0)^2)}} \\
\leq &\frac{a_1^2(t_0)}{\sqrt{2(1+\phi_1''(t_0)^2)}} - \frac{A_1^2(t_0)}{\sqrt{2(1+\varphi_1''(t_0)^2)}} \exp\left\{-\frac{2\pi(\varphi_1'(t_0)-\phi_1'(t_0))^2}{1+\varphi_1''(t_0)^2}\right\} = O(\epsilon)
\end{aligned}$$

by (B.5) since 0 is the unique maximal point of the chosen Gaussian function.

Claim Appendix B.4. *$|\phi_\ell'(t) - \varphi_\ell'(t)| = O(\sqrt{\epsilon})$ for all time $t \in \mathbb{R}$ and $\ell = 1, \ldots, N$.*

Proof. Fix $t_0 \in \mathbb{R}$ and $\ell = 1$. By (B.5), (B.6) and the conditions of $\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$, on $I_1(t_0)$ we have

$$\frac{A_1^2(t_0)}{\sqrt{2(1+\varphi_1''(t_0)^2)}} \left|1 - \exp\left\{-\frac{2\pi(\varphi_1'(t_0)-\phi_1'(t_0))^2}{1+\varphi_1''(t_0)^2}\right\}\right| = O(\epsilon).$$

Due to the fact that the Gaussian function monotonically decreases as $\frac{2\pi(\varphi_1'(t_0)-\phi_1'(t_0))^2}{1+\varphi_1''(t_0)^2} > 0$, we have

$$\frac{(\varphi_1'(t_0)-\phi_1'(t_0))^2}{1+\varphi_1''(t_0)^2} = O(\epsilon).$$

Since φ_1'' is uniformly bounded by c_2, we know

$$|\varphi_1'(t_0) - \phi_1'(t_0)| = O(\sqrt{\epsilon}).$$

By the same argument, we know that $|\varphi_l'(t) - \phi_l'(t)| = O(\sqrt{\epsilon})$ for all $l = 1, \ldots, N$ and $t \in \mathbb{R}$. □

Claim Appendix B.5. *$|\phi_\ell''(t) - \varphi_\ell''(t)| = O(\sqrt{\epsilon})$ for all time $t \in \mathbb{R}$ and $\ell = 1, \ldots, N$.*

Proof. Fix $t_0 \in \mathbb{R}$ and $\ell = 1$. By the assumption that $\phi_1'''(t_0) = O(\epsilon)$ and $\varphi_1'''(t_0) = O(\epsilon)$, we claim that $|\phi_1''(t_0) - \varphi_1''(t_0)| = O(\sqrt{\epsilon})$ holds. Indeed, we have

$$\phi_1'(t_0+1) = \phi_1'(t_0) + \int_{t_0}^{t_0+1} \phi_1''(s)\mathrm{d}s \quad \text{and} \quad \varphi_1'(t_0+1) = \varphi_1'(t_0) + \int_{t_0}^{t_0+1} \varphi_1''(s)\mathrm{d}s,$$

which leads to the relationship

$$\phi_1'(t_0+1)-\varphi_1'(t_0+1)=\phi_1'(t_0)-\varphi_1'(t_0)+\int_{t_0}^{t_0+1}(\phi_1''(s)-\varphi_1''(s))\mathrm{d}s.$$

Therefore, by the assumption that $\phi_1'''(t_0)=O(\epsilon)$ and $\varphi_1'''(t_0)=O(\epsilon)$, we have

$$\begin{aligned}&\int_{t_0}^{t_0+1}(\phi_1''(s)-\varphi_1''(s))\mathrm{d}s\\ =&\int_{t_0}^{t_0+1}\left(\phi_1''(t_0)-\varphi_1''(t_0)+\int_{t_0}^{s}(\phi_1'''(x)-\varphi_1'''(x))\,\mathrm{d}x\right)\mathrm{d}s\\ =&\phi_1''(t_0)-\varphi_1''(t_0)+O(\epsilon),\end{aligned}$$

which means that $|\phi_1''(t_0)-\varphi_1''(t_0)|=O(\sqrt{\epsilon})$ since $|\phi_1'(t_0+1)-\varphi_1'(t_0+1)|=O(\sqrt{\epsilon})$ and $|\phi_1'(t_0)-\varphi_1'(t_0)|=O(\sqrt{\epsilon})$. By the same argument, we know that $|\varphi_l''(t)-\phi_l''(t)|=O(\sqrt{\epsilon})$ for all $l=1,\dots,N$ and $t\in\mathbb{R}$. □

Claim Appendix B.6. *$|a_\ell(t)-A_\ell(t)|=O(\sqrt{\epsilon})$ for all time $t\in\mathbb{R}$ and $\ell=1,\dots,N$.*

Proof. Fix $t_0\in\mathbb{R}$ and $\ell=1$. From (B.6), it is clear that $|a_1(t_0)-A_1(t_0)|=O(\sqrt{\epsilon})$ if and only if $|\phi_1''(t_0)-\varphi_1''(t_0)|=O(\sqrt{\epsilon})$, so we obtain the claim by Claim Appendix B.5. Similar argument holds for all time $t\in\mathbb{R}$ and $\ell=2,\dots,N$. □

Lastly, we show the difference of the phase functions.

Claim Appendix B.7. *$|\phi_\ell(t)-\varphi_\ell(t)|=O(\sqrt{\epsilon})$ for all time $t\in\mathbb{R}$ and $\ell=1,\dots,N$.*

Proof. By (B.4) and the fact that $|a_l(t)-A_l(t)|=O(\sqrt{\epsilon})$, we have for all $t\in\mathbb{R}$,

$$\sum_{l=1}^{N}a_l(t)\cos[2\pi\phi_l(t)]=\sum_{l=1}^{N}a_l(t)\cos[2\pi(\phi_l(t)+\alpha_l(t))]+O(\sqrt{\epsilon}),$$

where $\alpha_l\in C^3(\mathbb{R})$. Note that $\sum_{l=1}^{N}a_l(t)\cos[2\pi(\phi_l(t)+\alpha_l(t))]\in\mathcal{Q}_{\epsilon,d}^{c_1,c_2}$. Fix $t_0\in\mathbb{R}$. Suppose there exists t_0 and the smallest number k so that $\alpha_k(t_0)=O(\sqrt{\epsilon})$ up to multiples of 2π does not hold. Then there exists at least one $\ell\neq k$ so that $\alpha_\ell(t_0)=O(\sqrt{\epsilon})$ does not hold. Suppose $L>k$ is the largest integer that $\alpha_L(t_0)=O(\sqrt{\epsilon})$ does not hold. In this case, there exists $t_1>t_0$ so that $\sum_{l=1}^{N}a_l(t_1)\cos[2\pi\phi_l(t_1)]=\sum_{l=1}^{N}a_l(t_1)\cos[2\pi(\phi_l(t_1)+\alpha_l(t_1))]+O(\sqrt{\epsilon})$ does not hold. Indeed, as $\phi_L'(t_0)$ is higher than $\phi_k'(t_0)$ by at least d, we could find $t_1=\phi_k^{-1}(\phi_k(t_0)+c)$, where $0<c<\pi$, so that $\cos[2\pi(\phi_L(t_1)+\alpha_L(t_1))]-\cos[2\pi(\phi_L(t_1))]=\cos[2\pi(\phi_L(t_0)+\alpha_L(t_0))]-\cos[2\pi(\phi_L(t_0))]+O(\sqrt{\epsilon})$ does not hold while $\sum_{l\neq L}^{N}a_l(t)\cos[2\pi\phi_l(t)]=\sum_{l\neq L}^{N}a_l(t)\cos[2\pi(\phi_l(t)+\alpha_l(t))]+O(\sqrt{\epsilon})$ holds. We thus get a contradiction and hence the proof. □