Appl. Comput. Harmon. Anal. 37 (2014) 235–270

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Convergence of a data-driven time–frequency analysis method

Thomas Y. Hou [a], Zuoqiang Shi [b,*], Peyman Tavallali [a]

[a] Applied and Comput. Math., Caltech, Pasadena, CA 91125, United States
[b] Mathematical Sciences Center, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Article history:
Received 28 March 2013
Received in revised form 1 August 2013
Accepted 29 December 2013
Available online 2 January 2014
Communicated by Stephane G. Mallat

Keywords:
Sparse representation
Data-driven
Time–frequency analysis
Matching pursuit

ABSTRACT

In a recent paper [11], Hou and Shi introduced a new adaptive data analysis method to analyze nonlinear and non-stationary data. The main idea is to look for the sparsest representation of multiscale data within the largest possible dictionary consisting of intrinsic mode functions of the form $\{a(t)\cos(\theta(t))\}$, where $a \in V(\theta)$, $V(\theta)$ consists of the functions that are less oscillatory than $\cos(\theta(t))$ and $\theta' \geqslant 0$. This problem was formulated as a nonlinear L^0 optimization problem and an iterative nonlinear matching pursuit method was proposed to solve this nonlinear optimization problem. In this paper, we prove the convergence of this nonlinear matching pursuit method under some scale separation assumptions on the signal. We consider both well-resolved and poorly sampled signals, as well as signals with noise. In the case without noise, we prove that our method gives exact recovery of the original signal.

1. Introduction

Developing a truly adaptive data analysis method is important for our understanding of many natural phenomena. Although a number of effective data analysis methods such as the Fourier transform or windowed Fourier transform have been developed, these methods use pre-determined bases and are mostly used to process linear and stationary data. Applications of these methods to nonlinear and non-stationary data tend to give many unphysical harmonic modes. To overcome these limitations of the traditional techniques, time–frequency analysis has been developed by representing a signal with a joint function of both time and frequency [9]. The recent advances of wavelet analysis have led to the development of several powerful wavelet-based time–frequency analysis techniques [13,7,19,17]. But they still cannot remove the artificial harmonics completely and do not give satisfactory results for nonlinear signals.

Another important approach in the time–frequency analysis is to study instantaneous frequency of a signal. Some of the pioneering work in this area was due to Van der Pol [25] and Gabor [10], who introduced the so-called Analytic Signal (AS) method that uses the Hilbert transform to determine instantaneous frequency of a signal. However, this method works mostly for monocomponent signals in which the number of zero-crossings is equal to the number of local extrema [1]. There were other attempts to define instantaneous

* Corresponding author.
E-mail addresses: hou@cms.caltech.edu (T.Y. Hou), zqshi@math.tsinghua.edu.cn (Z. Shi), ptavalla@caltech.edu (P. Tavallali).

1063-5203/$ – see front matter
http://dx.doi.org/10.1016/j.acha.2013.12.004

frequency such as the zero-crossing method [22,23,18] and the Wigner–Ville distribution method [1,15,21,9,14,20]. Most of these methods suffer from various limitations. The main limitation is that they all assume that there is a single instantaneous frequency which implies that the performance of these method is not good for multi-component signals.

More substantial progress has been made recently with the introduction of the Empirical Mode Decomposition (EMD) method [12]. Through a sifting process, the EMD method decomposes a signal into a collection of oscillating functions with possibly modulated amplitudes and frequencies, which are called intrinsic mode functions (IMFs) in the literatures of EMD [12]. On the other hand, since the EMD method relies on the information of local extrema of a signal, it is unstable to noise perturbation. Recently, an ensemble EMD method (EEMD) was proposed to make it more stable to noise perturbation [26]. But some fundamental issues remain unresolved.

1.1. A brief review of the data-driven time–frequency analysis method

Inspired by EMD/EEMD and the recently developed compressive sensing theory [5,4,8,2], Hou and Shi proposed a data-driven time–frequency analysis method in a recent paper [11]. The main idea of this method is to look for the sparsest decomposition of a signal over the largest possible dictionary. The dictionary is chosen to be:

$$\mathcal{D} = \left\{a\cos\theta\colon\ a \in V(\theta),\ \theta' \in V(\theta),\ \text{and } \theta'(t) > 0,\ \forall t \in \mathbb{R}\right\}, \tag{1}$$

where $V(\theta)$ is a collection of all the functions that are less oscillatory than $\cos\theta(t)$. By saying that f is less oscillatory than g, we mean that either f contains fewer high frequency Fourier modes than g or the high frequency Fourier coefficients of f decay faster than those of g. In many cases, this would imply that the H^1-norm of f is smaller that of g. In general, it is most effective to construct $V(\theta)$ using overcomplete Fourier modes in the θ-space, which is the function space with θ as a coordinate.

In this paper, we only consider the periodic data. We can use standard Fourier modes in the θ-space to construct the $V(\theta)$ space. More precisely, we will define $V(\theta)$ as follows:

$$V(\theta) = \operatorname{span}\left\{1, \left(\cos\left(\frac{k\theta}{L_\theta}\right)\right)_{1\leqslant k\leqslant \lambda L_\theta}, \left(\sin\left(\frac{k\theta}{L_\theta}\right)\right)_{1\leqslant k\leqslant \lambda L_\theta}\right\}, \tag{2}$$

where $L_\theta = (\theta(T) - \theta(0))/2\pi$ is a positive integer and $\lambda \leqslant 1/2$ is a control parameter, which enforces that the functions in $V(\theta)$ are less oscillatory that $\cos\theta(t)$. In the analysis and computations of this paper, we set $\lambda = 1/2$.

We then formulate the problem as a nonlinear version of the L^0 minimization problem.

$$\begin{array}{lcl} P: & \underset{(a_k)_{1\leqslant k\leqslant M},(\theta_k)_{1\leqslant k\leqslant M}}{\text{Minimize}} & M \\ & \text{Subject to} & \begin{cases} f = \sum_{k=1}^M a_k\cos\theta_k, \\ a_k\cos\theta_k \in \mathcal{D}, \quad k = 1,\ldots,M. \end{cases} \end{array} \tag{3}$$

The constraint $f = \sum_{k=1}^M a_k\cos\theta_k$ can be replaced by an inequality when the signal is polluted by noise. This kind of optimization problem is known to be very challenging to solve since both a_k and θ_k are unknown. Inspired by matching pursuit [16,24], Hou and Shi [11] proposed a nonlinear matching pursuit method to solve this nonlinear optimization problem. The basic idea is to decompose the signal sequentially into two parts by solving a nonlinear least square problem:

$$\min_{a,\theta} \|f - a\cos\theta\|_{l^2}^2, \qquad \text{Subject to} \quad a\cos\theta \in \mathcal{D}.$$

This nonlinear least square problem is solved by a Gauss–Newton type iteration. In each step of this algorithm, we need to solve the following least square problem:

$$\min_{a,b} \quad \left\|f - a\cos\theta^n - b\sin\theta^n\right\|_{l^2}^2,$$
$$\text{Subject to} \quad a, b \in V\left(\theta^n\right). \tag{4}$$

When the signal has sufficient samples, this nonlinear least square problem can be solved approximately by first interpolating f to a uniform mesh in the θ^n-space and then applying FFT. This gives rise to a very efficient algorithm with complexity of order $O(N\log(N))$, where N is the number of sample points of the signal, see Section 2 for more details.

If the signal is poorly sampled, then we cannot apply FFT. In this case, we need to solve an l^1 minimization problem to obtain the Fourier coefficients of f in the θ^n-space, where θ^n is a given approximate phase function.

$$\min_x \|x\|_1, \quad \text{subject to} \quad \Phi_{\theta^n} x = f, \tag{5}$$

where each column of matrix Φ_{θ^n} is a Fourier mode in the θ^n-space, i.e. each column of matrix Φ_{θ^n} is of the type $e^{i2k\pi\bar{\theta}^n}$, where $k \in \mathbb{Z}$ and $\bar{\theta}^n = \frac{\theta^n - \theta^n(0)}{\theta^n(T) - \theta^n(0)}$. We then use this coefficient x to update θ^n, and repeat this process until it converges. We refer to Section 3 for more details of this algorithm.

The objective of this paper is to analyze the convergence and stability of the algorithms in two cases: periodic signals with well-resolved samples and periodic signals with poor samples.

1.2. Main results

Our first result is for well-resolved periodic signals of the form $f(t) = f_0(t) + f_1(t)\cos\theta(t)$. By a well-resolved signal, we mean that the signal is measured over a set of grid points that are fine enough such that we can interpolate the signal to any other grid points with very little loss of accuracy.

We ignore the interpolation error and assume that $f(t)$ is given for all $t \in [0, T]$. We further assume that the non-zero Fourier coefficients of θ' in the physical space are confined in the first M_0 modes, i.e.

$$\theta'(t) \in \text{span}\left\{e^{i2k\pi t/T}, \ |k| \leqslant M_0\right\},$$

and f_0, f_1 have M_1 low frequency modes in the $\bar{\theta}$-space, i.e.

$$f_0, f_1 \in \text{span}\left\{e^{i2k\pi\bar{\theta}}, \ |k| \leqslant M_1\right\},$$

where $\bar{\theta} = \frac{\theta(t) - \theta(0)}{\theta(T) - \theta(0)}$ is the normalized phase function. Later on, we refer to this property for f_0, f_1 and θ' as the "low frequency confinement property".

For this type of signals, we can prove that the iterative algorithm converges to the exact solution under some scale separation assumption on the signal. More precisely, if θ^0, the initial guess of θ, satisfies

$$\left\|\mathcal{F}\left(\left(\theta^0 - \theta\right)'\right)\right\|_1 \leqslant \pi M_0/2, \tag{6}$$

where $\mathcal{F}$ is the Fourier transform in the physical space, then there exists $\eta_0 > 0$ such that

$$\left\|\mathcal{F}\left(\left(\theta^{m+1} - \theta\right)'\right)\right\|_1 \leqslant \frac{1}{2}\left\|\mathcal{F}\left(\left(\theta^m - \theta\right)'\right)\right\|_1, \quad \forall m = 0, 1, 2, \ldots \tag{7}$$

provided that $L \geqslant \eta_0$, where $L = \frac{\theta(T)-\theta(0)}{2\pi}$ and η_0 is a constant determined by M_0, M_1, $\min f_1$ and $\bar{\theta}'$. The precise statement of the theorem can be found in Theorem 2.1. We remark that $1/L$ can be used to measure the smallest scale of the signal. By smallest scale, we mean the length of the smallest interval over which the signal has $O(1)$ change. Similarly, we can use $1/M_1$ and $1/M_0$ to measure the smallest scale of f_0, f_1, and θ' respectively. The requirement $L \geqslant \eta_0$ is actually a mathematical formulation of the scale separation property. By scale separation, we mean that the mean f_0 and the amplitude f_1 are less oscillatory than $\cos\theta$.

The key idea of the proof is to estimate the decay rate of the coefficients over the Fourier basis in the θ^n-space, where θ^n is the approximate phase function in each step. We show that the Fourier coefficients of the signal in the θ^n-space have a very fast decay as long as that θ^n is a smooth function. Using this estimate, we can show that the error of the phase function in each step is a contraction and the iteration converges to the exact solution.

We have also proved a similar convergence result for signals that are polluted by noise, see Section 2.2. In many problems, f_0, f_1 and θ' may not be exactly low frequency confined. A more general setting is that the Fourier coefficients of f_0, f_1, and θ' decay according to some power law as the wave number increases. In this case, we can prove that our method will converge to an approximate solution with an error determined by the truncated error of f_0, f_1 and θ'. The detailed analysis will be presented in Section 2.3.

For signals with poor samples, we can also prove similar convergence results with an extra condition on the matrix Φ_{θ^n}. In this case, we need to use the l^1 minimization even with periodic signals. Suppose S is the largest number such that $\delta_{3S}(\Phi_{\theta^n}) + 3\delta_{4S}(\Phi_{\theta^n}) < 2$, and $\delta_S(A)$ is the S-restricted isometry constant of matrix A given in [3]. Under the same sparsity assumption on the instantaneous frequency, the mean and the amplitude as before, we can prove that there exist $\eta_L > 0$, $\eta_S > 0$, such that

$$\big\|\mathcal{F}\big(\big(\theta^{m+1}-\theta\big)'\big)\big\|_1 \leqslant \frac{1}{2}\big\|\mathcal{F}\big(\big(\theta^{m}-\theta\big)'\big)\big\|_1, \tag{8}$$

provided that $L \geqslant \eta_L$ and $S \geqslant \eta_S$.

Further, we show that if the sample points $\{t_j\}_{j=1}^{N_s}$ are selected at random from a set of uniformly distributed points $\{t_l\}_{j=1}^{N_f}$, the condition $\delta_{3S}(\Phi_{\theta^n}) + 3\delta_{4S}(\Phi_{\theta^n}) < 2$ holds with an overwhelming probability provided that $S \leqslant CN_s/(\max(\bar{\theta}')(\log N_b)^6)$ and $N_f \geqslant \max\{C\|\hat{\bar{\theta}}'\|_1 N_b, 2M_0\}$, where N_s is the number of the samples, N_b is the number of the basis. If $M_0 = 0$, which implies that $\bar{\theta}' = 1$, then the above result is reduced to the well-known theorem for the standard Fourier basis in [6].

The rest of the paper is organized as follows. In Section 2, we establish the convergence and stability of our method for well-resolved signals. In Section 3, we propose an algorithm for signals with poor samples and prove its convergence and stability. In Section 4, some numerical results are presented to demonstrate the performance of the algorithm and confirm the theoretical results. Some concluding remarks are made in Section 5.

2. Well-resolved periodic signal

In this section, we will analyze the convergence and stability of the algorithm proposed in [11] for signals which are well-resolved by the samples. In the analysis, we assume that the signal is periodic in the sample domain. Without loss of generality, we assume that the signal f is periodic over $[0,1]$.

As we mentioned in the introduction, our nonlinear matching pursuit method solves a least square problem (4) iteratively. Since we require that the phase function θ^n be monotonically increasing, we can use θ^n as a coordinate instead of the physical coordinate t. In this new coordinate, $\cos\theta^n$, $\sin\theta^n$ and the basis functions in $V(\theta^n)$ are simple Fourier modes. We can solve the least-square problem (4) easily by using the Fourier transform.

In the θ-space, the signal is sampled over a non-uniform grid. In order to employ the Fast Fourier Transform to accelerate the computation, we have to interpolate the signal to an uniform grid in the θ-space which will introduce some interpolation error. This is why we require that the signal is well-resolved to make sure that the interpolation error is very small. We can also utilize a non-uniform Fast Fourier Transform to avoid interpolation errors. In this paper, we do not consider this approach since the implementation of non-uniform FFT is more complicated. In our practical implementation, we use the interpolation-FFT approach to calculate the Fourier transform in the θ-space. However, in the analysis, we neglect the interpolation error since the signal is assumed to be well-resolved and the interpolation error is negligible.

In order to make this paper self-contained, we also state the algorithm here. Suppose the signal f is given over a uniform grid $t_j = j/N$ for $j = 0, \dots, N-1$.

Algorithm 1 *(Data-driven time–frequency analysis for periodic signal with well-resolved samples).*
Input: original signal f; initial guess of the phase functions θ^0.
Output: phase function θ, amplitude a_1, residual r.
Main iteration:

Initialization: $n = 0$ and $\theta^n = \theta^0$.

S1: Interpolate f from the grid in the time domain to a uniform mesh in the θ^n-coordinate to get f_{θ^n} and compute the Fourier transform $\hat{f}_{\theta^n}$:

$$f_{\theta^n,\,j} = \text{Interpolate}\ \big(\theta^n(t_i), f, \theta^n_j\big), \tag{9}$$

where θ^n_j, $j = 0, \dots, N-1$ are uniformly distributed in the θ^n-coordinate, i.e. $\theta^n_j = 2\pi L_{\theta^n} j/N$. We use the cubic spline to perform the interpolation.

Apply the Fourier transform to f_{θ^n} as follows:

$$\hat{f}_{\theta^n}(\omega) = \sum_{j=1}^{N} f_{\theta^n,j} e^{-i2\pi\omega\bar{\theta}^n_j}, \quad \omega = -N/2+1, \dots, N/2, \tag{10}$$

where $\bar{\theta}^n_j = \frac{\theta^n_j - \theta^n_0}{2\pi L_{\theta^n}}$.

S2: Apply a cutoff function to the Fourier transform of f_{θ^n} to compute a and b on the mesh in the θ^n_k-coordinate, denoted by a_{θ^n} and b_{θ^n}:

$$a_{\theta^n}(\omega) = \mathcal{F}^{-1}_{\theta^n}\big[\big(\hat{f}_{\theta^n}(\omega + L_{\theta^n_k}) + \hat{f}_{\theta^n}(\omega - L_{\theta^n_k})\big) \cdot \chi(\omega/L_{\theta^n_k})\big], \tag{11}$$

$$b_{\theta^n}(\omega) = -i \cdot \mathcal{F}^{-1}_{\theta^n}\big[\big(\hat{f}_{\theta^n}(\omega + L_{\theta^n_k}) - \hat{f}_{\theta^n}(\omega - L_{\theta^n_k})\big) \cdot \chi(\omega/L_{\theta^n_k})\big], \tag{12}$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform defined in the θ^n coordinate:

$$\mathcal{F}^{-1}_{\theta^n}(\hat{f}_{\theta^n}) = \frac{1}{N}\sum_{\omega=-N/2+1}^{N/2} \hat{f}_{\theta^n} e^{i2\pi\omega\bar{\theta}^n_j}, \quad j = 0, \dots, N-1, \tag{13}$$

and χ is the cutoff function, which is defined implicitly by the definition of $V(\theta)$ in (2),

$$\chi(\omega) = \begin{cases} 1, & -1/2 < \omega < 1/2, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

S3: Interpolate a_{θ^n} and b_{θ^n} back to the uniform mesh in the time domain by the cubic spline:

$$a^{n+1} = \text{Interpolate }\left(\theta_j^n, a_{\theta^n}, \theta^n(t_i)\right), \quad i = 0, \ldots, N-1, \tag{15}$$

$$b^{n+1} = \text{Interpolate }\left(\theta_j^n, b_{\theta^n}, \theta^n(t_i)\right), \quad i = 0, \ldots, N-1. \tag{16}$$

S4: Update θ^n in the t-coordinate:

$$\Delta\theta' = P_{V_{M_0}}\left(\frac{d}{dt}\left(\arctan\left(\frac{b_k^{n+1}}{a_k^{n+1}}\right)\right)\right), \qquad \Delta\theta(t_i) = \int_0^{t_i} \Delta\theta'(s)\,ds, \quad i = 0, \ldots, N-1$$

and

$$\theta^{n+1}(t_i) = \theta^n(t_i) + \beta\Delta\theta(t_i), \quad i = 0, \ldots, N-1,$$

where $\beta \in [0,1]$ is chosen to make sure that θ^{n+1} is monotonically increasing:

$$\beta = \max\left\{\alpha \in [0,1]\colon \frac{d}{dt}\left(\theta^n + \alpha\Delta\theta\right) \geqslant 0\right\}, \tag{17}$$

and $P_{V_{M_0}}$ is the projection operator to the space $V_{M_0} = \text{span}\{e^{i2k\pi t/T},\ k = -M_0, \ldots, 0, \ldots, M_0\}$ and M_0 is chosen *a priori*.

S5: If $\|\theta^{n+1} - \theta^n\|_2 < \epsilon_0$, stop. Set

$$\theta = \theta^{n+1}, \qquad a_1 = \sqrt{\left(a^{n+1}\right)^2 + \left(b^{n+1}\right)^2}, \qquad r = f - a^{n+1}\cos\theta^n - b^{n+1}\sin\theta^n. \tag{18}$$

Otherwise, set $n = n + 1$ and go to **S1**.

After the first component is obtained, treat the residual r as the input signal and apply the above algorithm to r with another initial guess of the phase function to get the second component. Repeat this process sequentially until the residual is small enough. This will decompose the original signal f to several components in the dictionary $\mathcal{D}$.

In the previous paper [11], we demonstrated that this algorithm works very effectively for periodic signals and is stable to noise perturbation. In this paper, we will analyze its convergence and stability. Our main results can be summarized as follows. For periodic signals that have the exact low-frequency sparsity structure, we can prove that the above algorithm will converge to the exact decomposition. For periodic signals that have an approximate low-frequency sparsity structure, the above algorithm will give an approximate result withe accuracy determined by the truncated error of the signal. The precise definition of low-frequency sparsity structure will be given in the convergence theorems. In the following three subsections, we will present these results separately.

2.1. Exact recovery

In this subsection, we consider a periodic signal $f(t)$ that has the following decomposition:

$$f(t) = f_0(t) + f_1(t)\cos\theta(t), \quad f_1(t) > 0,\ \theta'(t) > 0,\ t \in [0,1], \tag{19}$$

where f_0, f_1 and θ are the exact local mean, the amplitude and the phase function that we want to recover from the signal.

First, we introduce some notations. Let $L = \frac{\theta(1)-\theta(0)}{2\pi}$ be the number of periods of the signal which is a measurement of the scale of the signal. We denote $\bar{\theta} = \frac{\theta-\theta(0)}{2\pi L}$ as the normalized phase function and $\hat{f}_{0,\theta}(k), \hat{f}_{1,\theta}(k)$ as the Fourier coefficients of f_0, f_1 in the $\bar{\theta}$-coordinate, i.e.

$$\hat{f}_{0,\theta}(k) = \int_0^1 f_0\, e^{-i2\pi k\bar{\theta}}\, d\bar{\theta}, \qquad \hat{f}_{1,\theta}(k) = \int_0^1 f_1\, e^{-i2\pi k\bar{\theta}}\, d\bar{\theta}. \tag{20}$$

We also use the notation $\mathcal{F}_\theta(\cdot)$ to represent the Fourier transform in the θ-space and $\mathcal{F}(\cdot)$ to represent the Fourier transform in the original t-coordinate.

Now we can state the theorem as follows:

Theorem 2.1. *Assume that the instantaneous frequency θ' satisfies $\min\bar{\theta}' > M_0\pi$ and the non-zero Fourier coefficients of θ' in the physical space are confined in the first M_0 modes, i.e.*

$$\theta' \in V_{M_0} = span\bigl\{e^{i2k\pi t/T},\ k = -M_0, \ldots, 1, \ldots, M_0\bigr\}. \tag{21}$$

Further, we assume that the non-zero Fourier coefficients of f_0 and f_1 in the θ-space are confined in the first M_1 modes, i.e.

$$\hat{f}_{0,\theta}(k) = \hat{f}_{1,\theta}(k) = 0, \quad \forall|k| > M_1. \tag{22}$$

If the initial guess of the phase function, θ^0, satisfies

$$\bigl\|\mathcal{F}\bigl(\bigl(\theta^0 - \theta\bigr)'\bigr)\bigr\|_1 \leqslant \pi M_0/2, \tag{23}$$

then there exist $\eta_0 > 0$ such that

$$\bigl\|\mathcal{F}\bigl(\bigl(\theta^{m+1} - \theta\bigr)'\bigr)\bigr\|_1 \leqslant \frac{1}{2}\bigl\|\mathcal{F}\bigl(\bigl(\theta^m - \theta\bigr)'\bigr)\bigr\|_1, \tag{24}$$

provided that $L \geqslant \eta_0$.

We first introduce some notations for the convenience of the representation. Let θ^m be the approximate phase function in the mth step, and $\Delta\theta^m = \theta - \theta^m$ be the error of the phase function in the current step, $L^m = \frac{\theta^m(1)-\theta^m(0)}{2\pi}$ be the number of periods in mth step and $\Delta L^m = L - L^m$. Let $\tilde{a}^m$, $\tilde{b}^m$ be the approximate amplitude functions, which are obtained by using Step 3 of the algorithm. Further, we define $a^m = f_1 \cos\Delta\theta^m$, $b^m = f_1 \sin\Delta\theta^m$, and $\Delta a^m = a^m - \tilde{a}^m$, $\Delta b^m = b^m - \tilde{b}^m$. The quantities a^m and b^m can be considered as the "exact" amplitude functions at the mth iteration since $\Delta\theta^m = \arctan(\frac{b^m}{a^m})$. Thus, we would obtain the exact phase starting from θ^m in one iteration. In our analysis, we need to establish a relationship among Δa^m, Δb^m and $\Delta\theta^m$.

One key ingredient of the proof is to estimate the integral $\int_0^1 e^{i2\pi(\omega\bar{\theta}-k\bar{\theta}^m)}\, d\bar{\theta}^m$. Fortunately, for this type of integral, we have the following lemma.

Lemma 2.1. *Suppose $\phi'(t) > 0$, $t \in [0,1]$, $\phi(0) = 0$, $\phi(1) = 1$, and $\psi', \phi' \in V_{M_0} = span\{e^{i2k\pi t},\ k = -M_0, \ldots, 1, \ldots, M_0\}$. Then we have, for any $n \in \mathbb{N}$, there is a $(n-1)$th order polynomial $P(x,n)$, such that*

$$\left|\int_0^1 e^{i\psi} e^{-i2\pi\omega\phi}\, d\phi\right| \leqslant \frac{P\bigl(\frac{\|\widehat{\phi'}\|_1}{\min\phi'}, n\bigr) M_0^n}{|\omega|^n (\min\phi')^n} \sum_{j=1}^n (2\pi M_0)^{-j} \bigl\|\widehat{\psi'}\bigr\|_1^j, \tag{25}$$

provided that $e^{i\psi}e^{-i2\pi\omega\phi}$ is a periodic function. Here $P(x,n)$ is a $(n-1)$th order polynomial of x and the coefficients are non-negative and depend on n.

Remark 2.1. Lemma 2.1 is valid for any $n \in \mathbb{N}$. The integral that we would like to estimate in Lemma 2.1 is actually the Fourier transform of $e^{i\psi}$. Since ψ is a smooth function, we expect that the Fourier transform of $e^{i\psi}$ has a rapid decay for large $|\omega|$. In Lemma 2.1, we give a more delicate decay estimate of the Fourier transform of $e^{i\psi}$. Such estimate is required in our proof of Theorem 2.1.

Proof. Using integration by parts, we have

$$\left|\int_0^1 e^{i\psi}e^{-i2\pi\omega\phi}\,d\phi\right| = \frac{1}{|2\pi\omega|^n}\left|\int_0^1 \frac{d^n(e^{i\psi})}{d\phi^n}e^{-i2\pi\omega\phi}\,d\phi\right| \leqslant \frac{1}{|2\pi\omega|^n}\max_{t\in[0,1]}\left|\frac{d^n(e^{i\psi})}{d\phi^n}\right|.$$

Since $e^{i\psi}e^{-i2\pi\omega\phi}$ is periodic, there is no contribution from the boundary terms when performing integration by parts. Using the fact that $\psi', \phi' \in V_{M_0}$ for any $g \in V_{M_0}$ we have

$$\max_t\left|g^{(n)}(t)\right| \leqslant \sum_k\left|(2\pi k)^{n-1}\hat{g}'(k)\right| \leqslant (2\pi M_0)^{n-1}\sum_k\left|\hat{g}'(k)\right| = (2\pi M_0)^{n-1}\left\|\hat{g}'\right\|_1, \tag{26}$$

where $g^{(n)}(t)$ means the nth order derivative of g with respect to t.

Direct calculations give

$$\left|\frac{d^n(e^{i\psi})}{d\phi^n}\right| \leqslant \frac{P\left(\frac{\|\widehat{\phi'}\|_1}{\min\phi'}, n\right)}{(\min\phi')^n}\sum_{j=1}^n(2\pi M_0)^{n-j}\left\|\widehat{\psi'}\right\|_1^j. \tag{27}$$

Thus, we get

$$\left|\int_0^1 e^{i\psi}e^{-i2\pi\omega\phi}\,d\phi\right| \leqslant \frac{P\left(\frac{\|\widehat{\phi'}\|_1}{\min\phi'}, n\right)M_0^n}{|\omega|^n(\min\phi')^n}\sum_{j=1}^n(2\pi M_0)^{-j}\left\|\widehat{\psi'}\right\|_1^j. \tag{28}$$

This proves Lemma 2.1. □

Remark 2.2. Regarding the polynomial $P(x,n)$, we can get an explicit expression for small n. For example, when $n=2$, we have

$$\begin{aligned}
\left|\frac{d^2}{d\phi^2}e^{i\psi}\right| &= \left|i\left(\frac{\psi''}{\phi'^{\,2}} - \frac{\psi'\phi''}{\phi'^{\,3}} + i\frac{\psi'^{\,2}}{\phi'^{\,2}}\right)e^{i\psi}\right| \leqslant \left|\frac{\psi''}{\phi'^{\,2}}\right| + \left|\frac{\psi'\phi''}{\phi'^{\,3}}\right| + \left|\frac{\psi'^{\,2}}{\phi'^{\,2}}\right| \\
&\leqslant \frac{\max|\psi''|}{(\min\phi')^2} + \frac{\max|\psi'|\max|\phi''|}{(\min\phi')^3} + \frac{(\max|\psi'|)^2}{(\min\phi')^2} \\
&\leqslant \frac{1}{(\min\phi')^2}\left[\left(1 + \frac{\|\widehat{\phi'}\|_1}{\min\phi'}\right)2\pi M_0\left\|\widehat{\psi'}\right\|_1 + \left\|\widehat{\psi'}\right\|_1^2\right],
\end{aligned} \tag{29}$$

where we have used $\Delta\theta, \bar{\theta} \in V_{M_0}$ in deriving the last inequality. Then, we have $P(x,2) = x+1$. Similarly, we can also get $P(x,3) = 3x^2 + 4x + 3$.

Now we are ready to prove Theorem 2.1.

Proof of Theorem 2.1. First, we need to establish the relationship among $\Delta\theta^{m+1}$ and Δa^m, Δb^m. Recall that $\Delta\theta^m = \arctan(\frac{b^m}{a^m})$. Thus, we have $\widetilde{\Delta\theta} = \Delta\theta^m - \arctan(\frac{\tilde{b}^m}{\tilde{a}^m}) = \arctan(\frac{b^m}{a^m}) - \arctan(\frac{\tilde{b}^m}{\tilde{a}^m})$. Using the differential mean value theorem, we know that there exists $\xi \in [0,1]$ such that

$$\begin{aligned}
|\widetilde{\Delta\theta}| &= \left|\arctan\left(\frac{b^m}{a^m}\right) - \arctan\left(\frac{\tilde{b}^m}{\tilde{a}^m}\right)\right| = \left|\frac{(a^m+\xi\Delta a^m)\Delta b^m - (b^m+\xi\Delta b^m)\Delta a^m}{(a^m+\xi\Delta a^m)^2 + (b^m+\xi\Delta b^m)^2}\right| \\
&\leqslant \frac{(|a^m|+|\Delta a^m|)|\Delta b^m| + (|b^m|+|\Delta b^m|)|\Delta a^m|}{((a^m)^2+(b^m)^2)/2 - ((\Delta a^m)^2+(\Delta b^m)^2)} \\
&\leqslant D_1|\Delta a^m| + D_2|\Delta b^m|,
\end{aligned} \tag{30}$$

where

$$D_1 = \max_t\left\{\frac{f_1+|\Delta b^m|}{f_1^2/2 - ((\Delta a^m)^2+(\Delta b^m)^2)}\right\}, \qquad D_2 = \max_t\left\{\frac{f_1+|\Delta a^m|}{f_1^2/2 - ((\Delta a^m)^2+(\Delta b^m)^2)}\right\}, \tag{31}$$

and we have used the relations that $f_1^2 = (a^m)^2 + (b^m)^2$ and $|a^m|, |b^m| \leqslant f_1$.

In the algorithm, there is another smooth process when updating θ, which gives the following result for $\Delta\theta^{m+1}$,

$$\Delta\theta^{m+1} = 2\pi\Delta L^{m+1}t + \widetilde{\Delta\theta}_{p,M_0}, \tag{32}$$

where $\widetilde{\Delta\theta}_{p,M_0} = P_{V_{M_0}}(\widetilde{\Delta\theta}_p)$ is the projection of $\widetilde{\Delta\theta}_p$ over the space V_{M_0}, $\widetilde{\Delta\theta}_p$ and $2\pi\Delta L^{m+1}t$ are the periodic part and the linear part of $\widetilde{\Delta\theta}$ respectively:

$$\widetilde{\Delta\theta} = 2\pi\Delta L^{m+1}t + \widetilde{\Delta\theta}_p. \tag{33}$$

Using (32), we can estimate $(\Delta\theta^{m+1})'$ as follows,

$$\begin{aligned}
\left\|\mathcal{F}\left((\Delta\theta^{m+1})'\right)\right\|_1 &\leqslant 2\pi\Delta L^{m+1} + \left\|\widehat{\widetilde{\Delta\theta}'}_{p,M_0}\right\|_1 \leqslant 2\pi\Delta L + M_0\|\widehat{\widetilde{\Delta\theta}}_{p,M_0}\|_1 \\
&\leqslant 2\|\widetilde{\Delta\theta}\|_\infty + M_0^2\|\widetilde{\Delta\theta}_p\|_\infty \leqslant \left(3M_0^2+2\right)\|\widetilde{\Delta\theta}\|_\infty,
\end{aligned} \tag{34}$$

where we have used the fact that

$$2\pi\left|\Delta L^{m+1}\right| = \left|\widetilde{\Delta\theta}(1) - \widetilde{\Delta\theta}(0)\right| \leqslant 2\|\widetilde{\Delta\theta}\|_\infty, \tag{35}$$

$$\|\widetilde{\Delta\theta}_p\|_\infty = \|\widetilde{\Delta\theta}\|_\infty + 2\pi\Delta L \leqslant 3\|\widetilde{\Delta\theta}\|_\infty. \tag{36}$$

Combining (34) with (30), we get

$$\left\|\mathcal{F}\left((\Delta\theta^{m+1})'\right)\right\|_1 \leqslant \left(3M_0^2+2\right)\left(D_1\left\|\Delta a^m\right\|_\infty + D_2\left\|\Delta b^m\right\|_\infty\right). \tag{37}$$

Next, we will establish the relationship among Δa^m, Δb^m and $\Delta\theta^m$. This can be done by estimating the Fourier coefficients of $\bar{a}^m$, $\bar{b}^m$ in the θ^m-space.

In Appendix A, we will prove the following estimates of Δa^m and Δb^m (see (156), (157)),

$$\left|\Delta a^m\right| \leqslant 2\sum_{\frac{1}{2}L^m<k<\frac{3}{2}L^m}\left|\hat{f}_{0,\theta^m}(k)\right| + \sum_{\frac{3}{2}L^m<k<\frac{5}{2}L^m}\left(\left|\hat{a}_{\theta^m}(k)\right| + \left|\hat{b}_{\theta^m}(k)\right|\right) + \sum_{|k|>\frac{L^m}{2}}\left|\hat{a}_{\theta^m}(k)\right|, \tag{38}$$

$$\left|\Delta b^m\right| \leqslant 2\sum_{\frac{1}{2}L^m<k<\frac{3}{2}L^m}\left|\hat{f}_{0,\theta^m}(k)\right| + \sum_{\frac{3}{2}L^m<k<\frac{5}{2}L^m}\left(\left|\hat{a}_{\theta^m}(k)\right| + \left|\hat{b}_{\theta^m}(k)\right|\right) + \sum_{|k|>\frac{L^m}{2}}\left|\hat{b}_{\theta^m}(k)\right|, \tag{39}$$

where $\hat{f}_{0,\theta^m}$, $\hat{a}^m_{\theta^m}$ and $\hat{b}^m_{\theta^m}$ are the Fourier transform of f_0, a^m and b^m in the θ^m-space.

To obtain the desired estimates, we need to use Lemma 2.1 to estimate the Fourier coefficients of f_0, a^m, b^m in the θ^m-space. In an effort to make the proof concise and easy to follow, we defer the derivation of the estimates (40), (41) and (42) to Appendix C. The main results of Appendix C are summarized as follows. As long as $\gamma = \frac{\|\mathcal{F}[(\Delta\theta^m)']\|_1}{2\pi M_0} \leqslant 1/4$ and $L \geqslant 4M_1$, we have

$$\left|\hat{f}_{0,\theta^m}(\omega)\right| \leqslant C_0 Q\left(\frac{|\omega|}{2}\right)^{-n} M_0^n M_1 \gamma, \quad \forall|\omega| > L/2, \tag{40}$$

$$\left|\hat{a}^m_{\theta^m}(\omega)\right| \leqslant 4C_0 Q\left(\frac{|\omega|}{2}\right)^{-n} M_0^n (2M_1+1)\gamma, \quad \forall|\omega| \geqslant L/2, \tag{41}$$

$$\left|\hat{b}^m_{\theta^m}(\omega)\right| \leqslant 4C_0 Q\left(\frac{|\omega|}{2}\right)^{-n} M_0^n (2M_1+1)\gamma, \quad \forall|\omega| \geqslant L/2, \tag{42}$$

where $C_0 = \max_{|k|\leqslant M_1}(|\hat{f}_{0,\theta}(k)|, |\hat{f}_{1,\theta}(k)|)$ and

$$Q = \frac{P(z,n)}{(\min(\bar{\theta}^m)')^n}, \qquad z = \frac{\|\mathcal{F}[(\bar{\theta}^m)']\|_1}{\min(\bar{\theta}^m)'}, \qquad \gamma = \frac{\|\mathcal{F}[(\Delta\theta^m)']\|_1}{2\pi M_0}. \tag{43}$$

Using (38)–(42) and the fact that $\sum_{k=1}^{\infty} k^{-n}$ converges as long as $n \geqslant 2$, we conclude that

$$\left|\Delta a^m\right| \leqslant \Gamma_0 Q(\alpha L)^{-n+1}\gamma, \tag{44}$$

$$\left|\Delta b^m\right| \leqslant \Gamma_0 Q(\alpha L)^{-n+1}\gamma, \tag{45}$$

where $\alpha = L^m/L$ and Γ_0 is a constant that depends on M_0, M_1, n and C_0 (the magnitude of f_0 and f_1). It follows from (37), (43), (44) and (45) that

$$\left\|\mathcal{F}\left(\left(\Delta\theta^{m+1}\right)'\right)\right\|_1 \leqslant \Gamma_1(D_1+D_2)Q(\alpha L)^{-n+1}\left\|\mathcal{F}\left(\left(\Delta\theta^m\right)'\right)\right\|_1, \tag{46}$$

where Γ_1 is a constant that depends on M_0, M_1, n and C_0.

To complete the proof, we need to show that there exists a constant $\eta_0 > 0$ which does not change in the iterative process, such that $\tilde{\beta} = \Gamma_1(D_1+D_2)Q(\alpha L)^{-n+1} \leqslant 1/2$ provided that $L \geqslant \eta_0$. This seems to be trivial, simply choosing $\eta_0 = \frac{1}{\alpha}(2\Gamma_1(D_1+D_2)Q)^{1/(n-1)}$ would make $\tilde{\beta} \leqslant 1/2$ provided that $L \geqslant \eta_0$. The problem is that D_1, D_2, Q, α vary during the iteration. We need to show that they are uniformly bounded during the iteration.

It is relatively easy to show that α is bounded,

$$|1-\alpha| = \left|1 - \frac{\theta^m(1)-\theta^m(0)}{\theta(1)-\theta(0)}\right| = \left|\frac{\Delta\theta^m(1)-\Delta\theta^m(0)}{2\pi L}\right| \leqslant \frac{\|(\Delta\theta^m)'\|_\infty}{2\pi L} \leqslant \frac{\|\mathcal{F}[(\Delta\theta^m)']\|_1}{2\pi L} \leqslant \frac{M_0}{4L},$$

which implies that $7/8 \leqslant \alpha \leqslant 9/8$, provided that $L \geqslant 2M_0$ and $\gamma \leqslant 1/4$.

It is more involved to show that Q is bounded. We need to first estimate $|(\bar{\theta}^m)'|$ and $\|\mathcal{F}[(\bar{\theta}^m)']\|_1$,

$$\left|\left(\bar{\theta}^m\right)'\right| = \left|\bar{\theta}'/\alpha - \left(\Delta\theta^m\right)'/\left(2\pi L^m\right)\right| \geqslant \frac{1}{\alpha}\left(\bar{\theta}' - \left\|\mathcal{F}\left[\left(\Delta\theta^m\right)'\right]\right\|_1/(2\pi L)\right) \geqslant \frac{8}{9}\left(\bar{\theta}' - \frac{M_0}{4L}\right), \tag{47}$$

and

$$\begin{aligned}\left\|\mathcal{F}\left[\left(\bar{\theta}^m\right)'\right]\right\|_1 &= \frac{1}{\alpha}\left\|\widehat{\bar{\theta}'} - \mathcal{F}\left[\left(\Delta\theta^m\right)'\right]/(2\pi L)\right\|_1 \leqslant \frac{1}{\alpha}\left(\left\|\widehat{\bar{\theta}'}\right\|_1 + \left\|\mathcal{F}\left[\left(\Delta\theta^m\right)'\right]\right\|_1/(2\pi L)\right) \\ &\leqslant \frac{8}{7}\left(\left\|\widehat{\bar{\theta}'}\right\|_1 + M_0/(4L)\right),\end{aligned} \tag{48}$$

where we have used the assumption that $\gamma \leqslant \frac{1}{4}$. If L satisfies the following condition,

$$\frac{M_0}{L} \leqslant 2\min(\bar{\theta}'), \tag{49}$$

then we can get

$$\left|(\bar{\theta}^m)'\right| \geqslant \frac{4}{9}\bar{\theta}', \qquad \left\|\mathcal{F}\left[(\bar{\theta}^m)'\right]\right\|_1 \leqslant \frac{12}{7}\left\|\widehat{\bar{\theta}'}\right\|_1, \tag{50}$$

where we have used the fact that $\min(\bar{\theta}') \leqslant \max(\bar{\theta}') \leqslant \|\widehat{\bar{\theta}'}\|_1$. It follows from (50) that the term z defined in (43) is uniformly bounded,

$$z \leqslant z_0, \tag{51}$$

where z_0 is a constant depending on $\bar{\theta}'$.

Based on the above estimation of z, the term Q in (43) can be bounded by a constant,

$$Q = \frac{P(z,n)}{(\min(\bar{\theta}^m)')^n} \leqslant \left(\frac{9}{4}\right)^n \frac{P(z_0,n)}{(\min\bar{\theta}')^n} = Q_0, \tag{52}$$

where Q_0 is a constant that depends on $\bar{\theta}'$ and n. Here we have used the fact that $P(z,n)$ is a non-decreasing function of z, since it is a $(n-1)$th polynomial of z with non-negative coefficients.

We now proceed to bound D_1 and D_2. Note that if $|\Delta a^m|, |\Delta b^m| \leqslant \frac{\sqrt{2}}{4}\min f_1$, we can bound D_1 as follows:

$$\begin{aligned} D_1 &= \max\left\{\frac{|b^m| + |\Delta b^m|}{((a^m)^2 + (b^m)^2)/2 - ((\Delta a^m)^2 + (\Delta b^m)^2)}\right\} \\ &\leqslant \max \frac{|f_1| + |\Delta b^m|}{(f_1)^2/2 - ((\Delta a^m)^2 + (\Delta b^m)^2)} \\ &\leqslant \frac{4+\sqrt{2}}{\min f_1} = E_0. \end{aligned} \tag{53}$$

Similarly, we can show that $D_2 \leqslant E_0$.

It is not difficult to see that the condition $|\Delta a^m|, |\Delta b^m| \leqslant \frac{\sqrt{2}}{4}\min f_1$ is valid if L satisfies

$$\Gamma_0 Q_0 (7L/8)^{-n+1} \leqslant \sqrt{2}\min f_1, \tag{54}$$

since we have

$$|\Delta a| \leqslant \Gamma_0 Q(\alpha L)^{-n+1}\gamma \leqslant \frac{1}{4}\Gamma_0 Q_0 (7L/8)^{-n+1}, \tag{55}$$

$$|\Delta b| \leqslant \Gamma_0 Q(\alpha L)^{-n+1}\gamma \leqslant \frac{1}{4}\Gamma_0 Q_0 (7L/8)^{-n+1}, \tag{56}$$

where we have used $\alpha \geqslant 7/8$, $Q \leqslant Q_0$, the assumption $\gamma \leqslant \frac{1}{4}$ and the estimates (72), (73).

Finally, we have derived the following estimate for the error of the instantaneous frequency,

$$\left\|\mathcal{F}\left((\Delta\theta^{m+1})'\right)\right\|_1 \leqslant \beta\left\|\mathcal{F}\left((\Delta\theta^m)'\right)\right\|_1, \tag{57}$$

where $\beta = \Gamma_1 E_0 Q_0 (7L/8)^{-n+1}$, Γ_1 is a constant depends on M_0, M_1, n, E_0 depends on $\min f_1$, and Q_0 depends on $\bar{\theta}'$ and n.

Now, we would like to prove that if $\gamma = \frac{\|\mathcal{F}[(\Delta\theta^m)']\|_1}{2\pi M_0} \leqslant \frac{1}{4}$, then we have

$$\left\|\mathcal{F}\left(\left(\Delta\theta^{m+1}\right)'\right)\right\|_1 \leqslant \frac{1}{2}\left\|\mathcal{F}\left(\left(\Delta\theta^{m}\right)'\right)\right\|_1, \tag{58}$$

as long as L satisfies the following conditions

$$L \geqslant 4M_1, \qquad \frac{M_0}{L} \leqslant \min\left\{\frac{1}{2}, 2\min\left(\bar{\theta}'\right)\right\}, \tag{59}$$

$$\Gamma_0 Q_0 (7L/8)^{-n+1} \leqslant \sqrt{2}\min f_1, \tag{60}$$

$$\Gamma_1 E_0 Q_0 (7L/8)^{n-1} \leqslant \frac{1}{2}. \tag{61}$$

It is obvious that there exist $\eta_0 > 0$, such that conditions (59)–(61) are satisfied provided that $L \geqslant \eta_0$. Here η_0 is determined by $M_0, M_1, \bar{\theta}', \min f_1$ and n which does not change during the iteration process.

Using (57) and by induction, it is easy to show that if the initial condition satisfies

$$\frac{\|\mathcal{F}[(\theta^0 - \theta)']\|_1}{2\pi M_0} \leqslant \frac{1}{4},$$

then there exists $\eta_0 > 0$ which is determined by $M_0, M_1, \bar{\theta}', \min f_1$ and n, such that

$$\left\|\mathcal{F}\left(\left(\Delta\theta^{m+1}\right)'\right)\right\|_1 \leqslant \frac{1}{2}\left\|\mathcal{F}\left(\left(\Delta\theta^{m}\right)'\right)\right\|_1, \tag{62}$$

as long as $L \geqslant \eta_0$. This completes the proof of Theorem 2.1. □

Remark 2.3. The above proof is valid for any $n \geqslant 2$. Note that η_0 depends on n. Theoretically, there exists an optimal choice of n to make η_0 the smallest. By carefully tracking the constants in the proof, we can show that as n going to $+\infty$, η_0 tends to $\delta C(n)^{1/(n-1)} M_0$, where δ is a constant independent on n, and $C(n)$ is the maximum of the coefficients of polynomial $P(x, n)$ appears in Lemma 2.1. We conjecture that $C(n)^{1/(n-1)}$ is bounded for $n \geqslant 2$. If this is the case, then η_0 is proportional to M_0.

Remark 2.4. Classical time–frequency analysis methods, such as the windowed Fourier transform or wavelet transform, in general cannot extract the instantaneous frequency exactly for any signal due to the uncertainty principle. For a single linear chirp signal without amplitude modulation, the Wigner–Ville distribution can extract the exact instantaneous frequency, but it fails if the signal consists of several components due to the interference. Theorem 2.1 shows that our data-driven time–frequency analysis method has the capability to recover the exact instantaneous frequency for a much larger range of signals even if the signals consist of multi-components.

2.2. Recovery of signals polluted by noise

Now, we turn to consider the case when the signal is polluted by noise, which we model as follows:

$$f = f_0 + f_1 \cos\theta + s, \tag{63}$$

where s is a perturbation to the original signal.

Using techniques similar to those in Theorem 2.1, we can prove that our method is stable to small perturbation. More precisely, we have the following theorem

Theorem 2.2. *Under the same assumptions as Theorem 2.1, if the initial guess of the phase function, θ^0, satisfies*

$$\left\|\mathcal{F}\left(\left(\theta^0-\theta\right)'\right)\right\|_1 \leqslant \pi M_0/2, \tag{64}$$

then there exist $\eta_0>0$ and $\epsilon_0>0$, $\Gamma_s>0$ such that

$$\left\|\mathcal{F}\left(\left(\theta^{m+1}-\theta\right)'\right)\right\|_1 \leqslant \Gamma_s\|s\|_\infty+\frac{1}{2}\left\|\mathcal{F}\left(\left(\theta^m-\theta\right)'\right)\right\|_1, \tag{65}$$

provided that $L\geqslant\eta_0$ and $\|s\|_\infty\leqslant\epsilon_0$. Here η_0 is a constant determined by $M_0, M_1, f_1, \bar{\theta}'$, ϵ_0 is a constant depends on f_1 and Γ_s is an absolute constants.

To prove this theorem, we need the following technical lemma,

Lemma 2.2. *Suppose $s(t)$ is a periodic function over $[0,1]$,*

$$s_L(t)=\mathcal{F}^{-1}\left[\left(\chi(1+k/L)+\chi(1-k/L)\right)\cdot\mathcal{F}[s](k)\right],\quad L\in\mathbb{N}, \tag{66}$$

and χ is the cutoff function,

$$\chi(\omega)=\begin{cases}1, & -1/2<\omega<1/2,\\ 0, & \textit{otherwise.}\end{cases} \tag{67}$$

Then, there exists $\Gamma_s>0$ independent on L such that

$$\|s_L\|_\infty\leqslant\Gamma_s\|s\|_\infty. \tag{68}$$

The proof of this lemma is deferred to Appendix B.

Now we are ready to prove Theorem 2.2

Proof. Using the same estimate as that in Theorem 2.1, we can get

$$\left|\Delta a^m\right|\leqslant 2\left\|s_{L^m}^m\right\|_\infty+\Gamma_0 Q(\alpha L)^{-n+1}\gamma, \tag{69}$$

$$\left|\Delta b^m\right|\leqslant 2\left\|s_{L^m}^m\right\|_\infty+\Gamma_0 Q(\alpha L)^{-n+1}\gamma, \tag{70}$$

where $\alpha=L^m/L$ and Γ_0 is a constant that depends on M_0, M_1, n and C_0 (the magnitude of f_0 and f_1), Q is defined in (43).

$$s_{L^m}^m=\mathcal{F}_{\theta^m}^{-1}\left[\chi(k/L)\cdot\mathcal{F}_{\theta^m}[s](k)\right]. \tag{71}$$

Then using Lemma 2.2, we have

$$\left|\Delta a^m\right|\leqslant\Gamma_s\|s\|_\infty+\Gamma_0 Q(\alpha L)^{-n+1}\gamma, \tag{72}$$

$$\left|\Delta b^m\right|\leqslant\Gamma_s\|s\|_\infty+\Gamma_0 Q(\alpha L)^{-n+1}\gamma, \tag{73}$$

where $\Gamma_s>0$ is an absolute constant.

By following the same argument as that in Theorem 2.1, we have

$$\left\|\mathcal{F}\left(\left(\Delta\theta^{m+1}\right)'\right)\right\|_1\leqslant\Gamma_s\|s\|_\infty+\Gamma_1 E_0 Q_0(7L/8)^{-n+1}\left\|\mathcal{F}\left(\left(\Delta\theta^m\right)'\right)\right\|_1, \tag{74}$$

as long as $\gamma = \frac{\|\mathcal{F}[(\Delta\theta^m)']\|_1}{2\pi M_0} \leqslant 1/4$ and the following conditions are satisfied

$$L \geqslant 2M_0, \qquad \frac{M_0}{L} \leqslant \min\{1/2, 2\min(\bar{\theta}')\}, \tag{75}$$

$$\Gamma_s\|s\|_\infty + \frac{1}{4}\Gamma_0 Q_0 (7L/8)^{-n+1} \leqslant \frac{\sqrt{2}}{4}\min f_1, \tag{76}$$

$$\Gamma_s\|s\|_\infty + \Gamma_1 Q_0 E_0 (7L/8)^{-n+1} \leqslant \frac{\pi M_0}{2}, \tag{77}$$

$$\Gamma_1 Q_0 E_0 (7L/8)^{-n+1} \leqslant \frac{1}{2}. \tag{78}$$

It is obvious that there exist $\eta_0 > 0$ and $\epsilon_0 > 0$, such that conditions (75)–(78) are satisfied provided that $L \geqslant \eta_0$ and $\|s\|_\infty \leqslant \epsilon_0$. Here η_0 is determined by $M_0, M_1, \bar{\theta}', \min f_1$ and n which does not change during the iteration process and ϵ_0 is a constant depends on $\min f_1$.

By induction, it is easy to show that if initially

$$\frac{\|\mathcal{F}[(\theta^0 - \theta)']\|_1}{2\pi M_0} \leqslant \frac{1}{4},$$

then there exist $\eta_0 > 0$ which is determined by $M_0, M_1, \bar{\theta}', \min f_1$ and n and an absolute constant $\epsilon_0 > 0$, such that

$$\big\|\mathcal{F}\big((\Delta\theta^{m+1})'\big)\big\|_1 \leqslant \Gamma_s\|s\|_\infty + \frac{1}{2}\big\|\mathcal{F}\big((\Delta\theta^m)'\big)\big\|_1, \tag{79}$$

as long as $L \geqslant \eta_0$ and $\|s\|_\infty \leqslant \epsilon_0$. This completes the proof of the theorem. □

2.3. Approximate recovery

If the signal does not have an exact low-frequency confined structure in the θ-space as required in Theorem 2.1, our method cannot reproduce the exact decomposition. But the analysis in this subsection shows that we can still get an approximate result and the accuracy is determined by the truncated error of the signal. The main result is stated below.

Theorem 2.3. *Assume that the non-zero Fourier coefficients of θ' in the physical space are confined in the first M_0 modes, i.e.*

$$\theta'(t) \in V_{M_0} = span\{e^{i2k\pi t/T},\ k = -M_0, \ldots, 1, \ldots, M_0\}, \tag{80}$$

and the Fourier coefficients of f_0 and f_1 in the $\bar{\theta}$-space have a fast decay, i.e. there exists $C_0 > 0$, $p \geqslant 4$ such that

$$\big|\hat{f}_{0,\theta}(k)\big| \leqslant C_0|k|^{-p}, \qquad \big|\hat{f}_{1,\theta}(k)\big| \leqslant C_0|k|^{-p}. \tag{81}$$

Then, there exists $\eta_0 > 4$ such that if $L > \eta_0$ and the initial guess satisfies

$$\big\|\mathcal{F}\big((\theta^0 - \theta)'\big)\big\|_1 \leqslant \pi M_0/2, \tag{82}$$

then we have

$$\big\|\mathcal{F}\big((\theta^{m+1} - \theta)'\big)\big\|_1 \leqslant \Gamma_0(L/4)^{-p+2} + \frac{1}{2}\big\|\mathcal{F}\big((\theta^m - \theta)'\big)\big\|_1, \tag{83}$$

where $\Gamma_0 > 0$ is a constant determined by C_0, p, M_0 $\min f_1$ and $\bar{\theta}'$.

Remark 2.5. This theorem shows that our iterative method will converge to the exact solution up to the truncation error determined by the scale separation property.

Proof. The proof is very similar to that of Theorem 2.1. The only difference is that the estimates of $\hat{f}_{0,\theta^m}(k)$, $\hat{a}^m_{\theta^m}$ and $\hat{b}^m_{\theta^m}$ are more complicated since they are not exactly confined in low frequency modes in the $\bar{\theta}$-space. Here we only give the key estimates.

For $\hat{f}_{0,\theta^m}(\omega)$, $\omega \neq 0$, we have

$$\begin{aligned} |\hat{f}_{0,\theta^m}| &= \left| \int_0^1 f_0 e^{-i2\pi\omega\bar{\theta}^m}\, d\bar{\theta}^m \right| \\ &= \left| \int_0^1 \sum_{k\neq 0} \hat{f}_{0,\theta}(k) e^{i2\pi k\bar{\theta}} e^{-i2\pi\omega\bar{\theta}^m}\, d\bar{\theta}^m \right| \\ &= \left| \sum_{k\neq 0} \hat{f}_{0,\theta}(k) \int_0^1 e^{i2\pi(\alpha k-\omega)\bar{\theta}^m} e^{ik\Delta\theta^m/L}\, d\bar{\theta}^m \right| \end{aligned} \tag{84}$$

where $\alpha = L^m/L$ and $\hat{f}_{0,\theta}(k)$ are the Fourier coefficients of f_0 as a function of $\bar{\theta}$. Note that the integral is 0 when $k = 0$ and $\omega \neq 0$. Thus we exclude the case $k = 0$ in the above summation. In the derivation of the last equality, we have used the relationship that $\bar{\theta} = \theta/L = (\theta^m + \Delta\theta^m)/L = \theta^m/L + \Delta\theta^m/L = \alpha\bar{\theta}^m + \Delta\theta^m/L$.

As in the proof of Theorem 2.1, we also need to use Lemma 2.1. In the previous proof, we can choose n to be any positive integer that is greater than 2. In the current theorem, the Fourier coefficients $|\hat{f}_{0,\theta}|$ and $|\hat{f}_{1,\theta}|$ decay according to some power law. To obtain the desired estimates, we need to take $2 \leqslant n \leqslant p-2$. This is why we require $p \geqslant 4$.

Applying Lemma 2.1 to the last equality of (84), we have

$$\begin{aligned} \left|\hat{f}_{0,\theta^m}(\omega)\right| &\leqslant \sum_{k\neq 0} \left|\hat{f}_{0,\theta}(k)\right| \left| \int_0^1 e^{i2\pi(\alpha k-\omega)\bar{\theta}^m} e^{ik\Delta\theta^m/L}\, d\bar{\theta}^m \right| \\ &\leqslant \sum_{|k|>\frac{|\omega|}{2\alpha}} \left|\hat{f}_{0,\theta}(k)\right| + \sum_{0<|k|\leqslant\frac{|\omega|}{2\alpha}} \left|\hat{f}_{0,\theta}(k)\right| \left| \int_0^1 e^{i2\pi(\alpha k-\omega)\bar{\theta}^m} e^{ik\Delta\theta^m/L}\, d\bar{\theta}^m \right| \\ &\leqslant C_0 \sum_{|k|>\frac{|\omega|}{2\alpha}} |k|^{-p} + C_0 \sum_{0<|k|\leqslant\frac{|\omega|}{2\alpha}} \frac{QM_0^n|k|^{-p}}{|\omega-\alpha k|^n} \sum_{j=1}^n \left|\frac{k}{L}\right|^j \left(\frac{\|\mathcal{F}[(\Delta\theta^m)']\|_1}{2\pi M_0}\right)^j \\ &\leqslant C_0 \int_{|\omega|/(2\alpha)}^\infty x^{-p}\, dx + C_0 Q\left(\frac{|\omega|}{2}\right)^{-n} M_0^n \left(\sum_{0<|k|\leqslant\frac{|\omega|}{2\alpha}} |k|^{-p+n}\right)\left(\sum_{j=1}^n (\gamma/L)^j\right) \\ &\leqslant C_0\left(\frac{|\omega|}{2\alpha}\right)^{-p+1} + C_0 Q\left(\frac{|\omega|}{2}\right)^{-n} M_0^n \gamma/L, \end{aligned} \tag{85}$$

where we have used the assumption $n \leqslant p-2$, $\gamma \leqslant 1/4$, and the fact that $L \geqslant 1$ is the number of the periods within the time interval $[0,1]$. Here C_0 is a generic constant, Q, z and γ are defined below:

$$Q = \frac{P(z,n)}{(\min(\bar{\theta}^m)')^n}, \qquad z = \frac{\|\mathcal{F}[(\bar{\theta}^m)']\|_1}{\min(\bar{\theta}^m)'}, \qquad \gamma = \frac{\|\mathcal{F}[(\Delta\theta^m)']\|_1}{2\pi M_0}. \tag{86}$$

Using an argument similar to that as in the derivation of (85), we can get the desired estimates for $\hat{a}_{\theta m}^m$ and $\hat{b}_{\theta m}^m$ as follows:

$$\left|\hat{a}_{\theta^m}^m(\omega)\right| \leqslant C_0\left(\frac{|\omega|}{2\alpha}\right)^{-p+1} + Q\left|\hat{f}_{1,\theta}(0)\right||\omega|^{-n}M_0^n\gamma + C_0 Q\left(\frac{|\omega|}{2}\right)^{-n}M_0^n\gamma, \tag{87}$$

$$\left|\hat{b}_{\theta^m}^m(\omega)\right| \leqslant C_0\left(\frac{|\omega|}{2\alpha}\right)^{-p+1} + Q\left|\hat{f}_{1,\theta}(0)\right||\omega|^{-n}M_0^n\gamma + C_0 Q\left(\frac{|\omega|}{2}\right)^{-n}M_0^n\gamma. \tag{88}$$

The estimates (38) and (39) remain valid in this case. Thus we obtain upper bounds for Δa^m and Δb^m by substituting (87) and (88) into (38) and (39),

$$\left|\Delta a^m\right| \leqslant \Gamma_1 L^{-p+2} + \Gamma_2 Q(\alpha L)^{-n+1}\gamma, \tag{89}$$

$$\left|\Delta b^m\right| \leqslant \Gamma_1 L^{-p+2} + \Gamma_2 Q(\alpha L)^{-n+1}\gamma, \tag{90}$$

where Γ_1 is a constant depending on C_0, Γ_2 depends on p and $\max(C_0, |\hat{f}_{1,\theta}(0)|)$.

Moreover, by following the same argument as we did in the proof of Theorem 2.1, we can obtain an error estimate for the instantaneous frequency,

$$\left\|\mathcal{F}\left(\left(\Delta\theta^{m+1}\right)'\right)\right\|_1 \leqslant \Gamma_3 E_0(L/4)^{-p+2} + \Gamma_4 E_0 Q_0(7L/8)^{-n+1}\left\|\mathcal{F}\left(\left(\Delta\theta^m\right)'\right)\right\|_1, \tag{91}$$

as long as $\gamma \leqslant 1/4$ and the following conditions are satisfied

$$L \geqslant 2M_0, \qquad \frac{M_0}{L} \leqslant \min\left\{1/2, 2\min\left(\bar{\theta}'\right)\right\}, \tag{92}$$

$$\Gamma_1(L/4)^{-p+2} + \Gamma_2 Q_0(7L/8)^{-n+1} \leqslant \sqrt{2}\min f_1, \tag{93}$$

$$\Gamma_3 E_0(L/4)^{-p+2} + \Gamma_4 Q_0 E_0(7L/8)^{-n+1} \leqslant \frac{\pi M_0}{2}, \tag{94}$$

$$\Gamma_4 Q_0 E_0(7L/8)^{-n+1} \leqslant \frac{1}{2}, \tag{95}$$

where Γ_3, Γ_4 are constants that depend on $C_0, p, M_0, \min f_1$ and $\bar{\theta}'$. Using these four constraints, we can easily derive a constant η_0, such that all these conditions are satisfied provided that $L \geqslant \eta_0$. This proves

$$\left\|\mathcal{F}\left(\left(\Delta\theta^{m+1}\right)'\right)\right\|_1 \leqslant \Gamma_3 E_0(L/4)^{-p+2} + \frac{1}{2}\left\|\mathcal{F}\left(\left(\Delta\theta^m\right)'\right)\right\|_1. \tag{96}$$

This completes the proof of Theorem 2.3 by setting $\Gamma_0 = \Gamma_3 E_0$. □

Remark 2.6. The constraint $n \leqslant p-2$ in the above proof can be relaxed to $p \geqslant 3$ by using a more delicate calculation.

If we further consider a more general case: the instantaneous frequency is also approximately low frequency confined instead of exactly low frequency confined as we assume in Theorems 2.1 and 2.3. In this case, we can prove that the iterative algorithm also converges to an approximate result. However, we cannot apply Lemma 2.1 here and need the following lemma instead.

Lemma 2.3. *Suppose* $\phi'(t) > 0$, $t \in [0,1]$, $\phi(0) = 0$, $\phi(1) = 1$, *and*

$$\left|\widehat{\phi'}(k)\right|, \left|\widehat{\psi'}(k)\right| \leqslant C|k|^{-p}, \quad \forall |k| > M_0.$$

Then for $n \leqslant p-1$, we have

$$\left|\int_0^1 e^{i\psi} e^{-i2\pi\omega\phi}\, d\phi\right| \leqslant \frac{P\left(\frac{\|\widehat{\phi'}\|_{1,M_0}+CM_0^{-p+1}}{\min\phi'}, n\right)}{|\omega|^n(\min\phi')^n} M_0^n \sum_{j=1}^{n}(2\pi M_0)^{-j}\left(\|\widehat{\psi'}\|_{1,M_0} + CM_0^{-p+1}\right)^j,$$

provided that $e^{i\psi}e^{-i2\pi\omega\phi}$ is a periodic function. Here $\|\widehat{\psi'}\|_{1,M_0} = \sum_{|k|\leqslant M_0} |\widehat{\psi'}(k)|$ and $P(x,n)$ is the same $(n-1)$th order polynomial as in Lemma 2.1.

Proof. The proof is similar to the proof of Lemma 2.1. The only difference is that we need the following estimate instead of (26),

$$\begin{aligned}\max_t\left|\psi^{(n)}(t)\right| &\leqslant \sum_k \left|(2\pi k)^{n-1}\widehat{\psi'}(k)\right| \leqslant (2\pi M_0)^{n-1}\sum_{|k|\leqslant M_0}\left|\widehat{\psi'}(k)\right| + (2\pi)^{n-1}C\sum_{|k|>M_0}|k|^{-p+n-1} \\ &\leqslant (2\pi M_0)^{n-1}\left(\|\widehat{\psi'}\|_{1,M_0} + CM_0^{-p+1}\right). \qquad \square\end{aligned} \tag{97}$$

Using this lemma and following an argument similar to that as in the previous two theorems, we can prove the following theorem:

Theorem 2.4. *Assume that the Fourier coefficients of the instantaneous frequency θ', the local mean f_0 and the amplitude f_1 all have fast decay, i.e. there exists $C_0 > 0$, $p \geqslant 4$ such that*

$$\left|\mathcal{F}(\theta')(k)\right| \leqslant C_0|k|^{-p}, \qquad \left|\mathcal{F}_\theta(f_0)(k)\right| \leqslant C_0|k|^{-p}, \qquad \left|\mathcal{F}_\theta(f_1)(k)\right| \leqslant C_0|k|^{-p}. \tag{98}$$

If L is large enough and the initial guess satisfies

$$\left\|\mathcal{F}\left((\theta^0-\theta)'\right)\right\|_1 \leqslant \pi M_0/2, \tag{99}$$

then, we have

$$\left\|\mathcal{F}\left((\theta^{m+1}-\theta)'\right)\right\|_1 \leqslant \Gamma_0(L/4)^{-p+2} + \frac{1}{2}C_0 M_0^{-p+1} + \frac{1}{2}\left\|\mathcal{F}\left((\theta^m-\theta)'\right)\right\|_1, \tag{100}$$

where $\Gamma_0 > 0$ is a constant determined by C_0, M_0 and f_1.

Remark 2.7. In the analysis presented in this section, we have assumed that the Fourier transform in the θ^m-space, $\mathcal{F}_{\theta^m}(\cdot)$, is exact. In real computations, we need to first interpolate the signal from a uniform grid in the physical space to a uniform grid in the θ^m-space, then apply the Fast Fourier Transform. This interpolation process would introduce some error. However, the interpolation error should be very small since we assume that the signal is well-resolved by the sample points.

3. Periodic signal with poor samples

In this section, we will consider a more challenging case, if the signal is poorly sampled. More specifically, we consider the case that the sample points t_j, $j = 1, \dots, N$ are too few to resolve the signal. In this case, the algorithm presented in the last section does not apply directly. The reason is that the Fourier transform in the θ^m-space, $\mathcal{F}_{\theta^m}(\cdot)$, cannot be computed accurately by the interpolation-FFT method. One way to obtain the Fourier transform in the θ^m-space is to apply non-uniform FFT without interpolation. However, for the signals we consider in this section, the number of samples is very small, for example, 1.2 samples per period on average (see Examples 2 and 3, Section 4). For this kind of signals, neither FFT nor non-uniform FFT could give accurate Fourier transform.

Notice that for the signal we consider in Theorem 2.1, its Fourier spectral would consist of two parts: low frequency part corresponding to the mean f_0 and high frequency part corresponding to $f_1 \cos\theta$. Since we assume that the non-zero Fourier coefficients of f_0 and f_1 are confined in the first M_1 modes in the θ-space, the non-zero Fourier coefficients of the original signal f in the θ-space should be confined in the first $4M_1$ modes, which implies that the signal is sparse in the Fourier space of θ if M_1 is small. Thanks to the recent developments of compressive sensing, we know that if the Fourier coefficients are sparse, then l^1 minimization would give an approximate solution from very few sample points. Hence, we can use a l^1 minimization problem to generate the Fourier coefficients in the θ^m-space in each step. This observation leads to the following algorithm:

Algorithm 2 *(Data-driven time–frequency analysis for periodic signal with sparse samples).*
Input: original signal: f; initial guess of the phase functions: θ^0.
Output: phase function θ, amplitude a_1, residual r.
Main iteration:

Initialization: $m = 0$ and $\theta^m = \theta^0$.
S1: Solve the l_1 minimization problem to get the Fourier transform of the signal f in the θ^m-coordinate:

$$\hat{f}_{\theta^m} = \arg\min_{x\in\mathbb{R}^{N_b}} \|x\|_1, \quad \text{subject to} \quad A_{\theta^m}\cdot x = f \tag{101}$$

where $A_{\theta^m} \in \mathbb{R}^{N_s\times N_b}$, $N_s < N_b$, N_s is the number of samples and N_b is the number of Fourier modes. $A_{\theta^m}(j,k) = e^{i2\pi k\bar{\theta}^m(t_j)}$, $j = 1,\ldots,N_s$, $k = -N_b/2+1,\ldots,N_b/2$ and $\bar{\theta}^m = \frac{\theta^m-\theta^m(0)}{\theta^m(T)-\theta^m(0)}$.
S2: Apply a cutoff function to the Fourier transform of f_{θ^m} to compute a^{m+1} and b^{m+1}:

$$a^{m+1} = \mathcal{F}^{-1}_{\theta^m}\big[\big(\hat{f}_{\theta^m}(\omega + L_{\theta^m}) + \hat{f}_{\theta^m}(\omega - L_{\theta^m})\big)\cdot\chi(\omega/L_{\theta^m})\big], \tag{102}$$

$$b^{m+1} = -i\cdot\mathcal{F}^{-1}_{\theta^m}\big[\big(\hat{f}_{\theta^m}(\omega + L_{\theta^m}) - \hat{f}_{\theta^m}(\omega - L_{\theta^m})\big)\cdot\chi(\omega/L_{\theta^m})\big], \tag{103}$$

where $\mathcal{F}^{-1}_{\theta^m}$ is the inverse Fourier transform defined in the θ^m-coordinate:

$$\mathcal{F}^{-1}_{\theta^m}(\hat{f}_{\theta^m})(t_j) = \sum_{\omega=-N_b/2+1}^{N_b/2} \hat{f}_{\theta^m}(\omega)e^{i2\pi\omega\bar{\theta}^m(t_j)}, \quad j = 1,\ldots,N_s, \tag{104}$$

and χ is the cutoff function,

$$\chi(\omega) = \begin{cases} 1, & -1/2 < \omega < 1/2, \\ 0, & \text{otherwise.} \end{cases} \tag{105}$$

S3: Update θ^m in the t-coordinate:

$$\Delta\theta' = P_{V_{M_0}}\left(\frac{d}{dt}\left(\arctan\left(\frac{b^{m+1}}{a^{m+1}}\right)\right)\right), \qquad \Delta\theta(t_j) = \int_0^{t_j} \Delta\theta'(s)\,ds, \quad j = 1,\ldots,N_s,$$

and

$$\theta^{m+1}(t_j) = \theta^m(t_j) + \beta\Delta\theta(t_j), \quad j = 1,\ldots,N_s, \tag{106}$$

where $\beta \in [0,1]$ is chosen to make sure that θ^{m+1} is monotonically increasing:

$$\beta = \max\left\{\alpha \in [0,1]:\ \frac{d}{dt}\left(\theta^m + \alpha\Delta\theta\right) \geqslant 0\right\}, \tag{107}$$

and $P_{V_{M_0}}$ is the projection operator to the space $V_{M_0} = \text{span}\{e^{i2k\pi t/T},\ k = -M_0, \ldots, 0, \ldots, M_0\}$ and M_0 is chosen *a priori*.

S4: If $\|\theta^{m+1} - \theta^m\|_2 < \epsilon_0$, stop. Set

$$\theta = \theta^{m+1}, \qquad a_1 = \sqrt{\left(a^{m+1}\right)^2 + \left(b^{m+1}\right)^2}, \qquad r = f - a^{m+1}\cos\theta^m - b^{m+1}\sin\theta^m. \tag{108}$$

Otherwise, set $m = m+1$ and go to **S1**.

Suppose the sample points t_j, $j = 1, \ldots, N_s$ are selected at random from a set of uniform grid l/N_f, $l = 0, \ldots, N_f - 1$, then the optimization problem (101) in Step 1 can be rewritten in the following form:

$$\min \|x\|_1, \quad \text{subject to} \quad \Phi_{\theta^m} \cdot x = \tilde{f}, \tag{109}$$

where $\tilde{f} = \sqrt{\frac{(\bar{\theta}^m)'}{N_f}} f$ and Φ_{θ^m} is obtained by selecting N_s rows from an N_f by N_b matrix U_{θ^m} which is defined as $U_{\theta^m}(j,k) = \sqrt{\frac{(\bar{\theta}^m)'}{N_f}} \cdot e^{i2\pi k\bar{\theta}^m(t_j)}$, $j = 1, \ldots, N_f$, $k = -N_b/2 + 1, \ldots, N_b/2$. As we will show later, the columns of U_{θ^m} are approximately orthogonal to each other. This property will play an important role in our convergence and stability analysis.

We remark that our problem is more challenging than the compressive sensing problem in the sense that we need not only to find the sparsest representation but also a basis parametrized by a phase function θ over which the signal has the sparsest representation. To overcome this difficulty, we propose an iterative algorithm to solve this nonlinear optimization problem.

3.1. Exact recovery

Theorem 3.1. *Under the same assumption as in Theorem 2.1, there exist $\eta_0 > 0$, $\eta_1 > 0$, such that*

$$\left\|\mathcal{F}\left(\left(\theta^{m+1} - \theta\right)'\right)\right\|_1 \leqslant \frac{1}{2}\left\|\mathcal{F}\left(\left(\theta^m - \theta\right)'\right)\right\|_1, \tag{110}$$

provided that $L \geqslant \eta_0$ and $S \geqslant \eta_1$, where S be the largest number such that $\delta_{3S}(\Phi_{\theta^m}) + 3\delta_{4S}(\Phi_{\theta^m}) < 2$. Here $\delta_S(A)$ is the S-restricted isometry constant of matrix A given in [3], which is the smallest number such that

$$(1 - \delta_S)\|c\|_{l^2}^2 \leqslant \|A_T c\|_{l^2}^2 \leqslant (1 + \delta_S)\|c\|_{l^2}^2,$$

for all subsets T with $|T| \leqslant S$ and coefficients sequences $(c_j)_{j \in T}$.

To prove this theorem, we need to use the following theorem of Candes, Romberg, and Tao [6].

Theorem 3.2. *Let S be such that $\delta_{3S}(A) + 3\delta_{4S}(A) < 2$, where $A \in \mathbb{R}^{n\times m}$, $n < m$. Suppose that x_0 is an arbitrary vector in $\mathbb{R}^m$ and let $x_{0,S}$ be the truncated vector corresponding to the S largest values of x_0. Then the solution x^* to the l_1 minimization problem*

$$\min \|x\|_1, \quad \textit{subject to} \quad Ax = f \tag{111}$$

satisfies

$$\left\|x^* - x_0\right\|_1 \leqslant C_{2,S} \cdot \|x_0 - x_{0,S}\|_1. \tag{112}$$

Now we present the proof of Theorem 3.1.

Proof of Theorem 3.1. Using (158) and (159) in Appendix A, we have

$$\begin{aligned} \left|\Delta a^m\right| &\leqslant 2 \sum_{\frac{L^m}{2} < k < \frac{3}{2}L^m} \left|\hat{f}_{0,\theta^m}(k)\right| + \sum_{\frac{3}{2}L^m < k < \frac{5}{2}L^m} \left(\left|\hat{a}^m_{\theta^m}(k)\right| + \left|\hat{b}^m_{\theta^m}(k)\right|\right) \\ &\quad + \sum_{|k| > \frac{L^m}{2}} \left|\hat{a}^m_{\theta^m}(k)\right| + 2 \sum_{\frac{L^m}{2} < k < \frac{3}{2}L^m} \left|\hat{f}_{\theta^m}(k) - \hat{\hat{f}}_{\theta^m}(k)\right| \\ &\leqslant \Gamma_0 Q(\alpha L)^{-n+1}\gamma + C_{2,S} \cdot \|\hat{f}_{\theta^m} - \hat{f}_{\theta^m,S}\|_1, \end{aligned} \tag{113}$$

where Γ_0 is a constant depending on M_0, M_1, n and $\hat{f}_{\theta^m,S}$ is the truncated vector corresponding to the S largest values of $\hat{f}_{\theta^m}$.

Without loss of generality, we assume that $L^m > S/3$, and define $\hat{\hat{f}}_{\theta^m,S}$ to be

$$\hat{\hat{f}}_{\theta^m,S}(k) = \begin{cases} \hat{f}_{\theta^m}(k), & k \in [-L^m - S/6, -L^m + S/6] \cup [-S/6, S/6] \cup [L^m - S/6, L^m + S/6], \\ 0, & \text{otherwise}. \end{cases}$$

Then by the definition of $\hat{f}_{\theta^m,S}$ and $\hat{\hat{f}}_{\theta^m,S}$, we have

$$\begin{aligned} \|\hat{f}_{\theta^m} - \hat{f}_{\theta^m,S}\|_1 &\leqslant \|\hat{f}_{\theta^m} - \hat{\hat{f}}_{\theta^m,S}\|_1 \\ &= \sum_{S/6 < |k| < L^m - S/6} \left|\hat{f}_{\theta^m}(k)\right| + \sum_{|k| > L^m + S/6} \left|\hat{f}_{\theta^m}(k)\right| \\ &\leqslant \sum_{|k| > S/6} \left|\hat{f}_{0,\theta^m}(k)\right| + \sum_{|k| > S/6} \left|\hat{a}_{\theta^m}(k)\right| + \sum_{|k| > S/6} \left|\hat{b}_{\theta^m}(k)\right| \\ &\leqslant \Gamma_1 Q S^{-n+1}\gamma. \end{aligned} \tag{114}$$

Substituting (114) into (113), we get

$$\left|\Delta a^m\right| \leqslant \left(\Gamma_0(\alpha L)^{-n+1} + C_{2,S}\Gamma_1 S^{-n+1}\right) Q\gamma. \tag{115}$$

Similarly, we obtain

$$\left|\Delta b^m\right| \leqslant \left(\Gamma_0(\alpha L)^{-n+1} + C_{2,S}\Gamma_1 S^{-n+1}\right) Q\gamma. \tag{116}$$

Using these two key estimates and following the same argument as that in the proof of Theorem 2.1, we can complete the proof of Theorem 3.1. □

Remark 3.1. The above result on the exact recovery of signals with sparse samples can be generalized to the case that we consider in Theorem 2.3 by combining the argument of the above theorem with the idea presented in the proof of Theorem 2.3. In this case, we can recover the signal with an error which is determined by L, S and the decay rates of $\hat{f}_{0,\theta}$, $\hat{f}_{1,\theta}$ and $\hat{\theta}'$.

In Theorem 3.1, we assume that in each step, the condition $\delta_{3S}(\Phi_{\theta^m}) + 3\delta_{4S}(\Phi_{\theta^m}) < 2$ is satisfied. Using the definition of δ_S, it is easy to see that $\delta_{3S} \leqslant \delta_{4S}$. Thus, a sufficient condition to satisfy $\delta_{3S}(\Phi_{\theta^m}) + 3\delta_{4S}(\Phi_{\theta^m}) < 2$ is to require $\delta_{4S}(\Phi_{\theta^m}) < 1/2$.

In compressive sensing, there is a well-known result by Candes and Tao in [4]. This result states that if the matrix $\Phi \in \mathbb{R}^{M\times N}$ is obtained by selecting M rows at random from an $N \times N$ Fourier matrix U where $U_{j,k} = \frac{1}{\sqrt{N}} e^{i2\pi jk/N}$, $j, k = 1, \dots, N$, then the condition $\delta_S(\Phi) < 1/2$ is satisfied with an overwhelming probability provided that

$$S \leqslant C \frac{M}{(\log N)^6}, \tag{117}$$

where C is a constant.

In our formulation (see (109)), the matrix Φ_{θ^m} also consists of N_s rows of an N_f-by-N_b matrix U_{θ^m}. The main difference is that the matrix U_{θ^m} is not a standard Fourier matrix. Instead it is a Fourier matrix in the θ^m-space which makes it non-orthonormal. As a result, we cannot apply the result of Candes and Tao in [4] directly. Fortunately, we have the following result by slightly modifying the arguments used in [4] which can be applied to matrix U_{θ^m}.

Theorem 3.3. *If $\nu_0 = \max_{k,j} |(U_\theta^* U_\theta - I)_{k,j}| \leqslant \frac{1}{16N_b}$, where U_θ^* is the conjugate transpose of U_θ, the condition $\delta_S(\Phi_\theta) < 1/2$ holds with probability $1-\delta$ provided that*

$$N_s \geqslant C \cdot \max(\bar{\theta})' \big(S \log^2 N_b - \log \delta\big) \log^4 N_b, \tag{118}$$

where N_s is the number of the samples, N_b is the number of elements in the basis.

This theorem shows that if the columns of U_{θ^m} are approximately orthogonal to each other, it has a property similar to the standard Fourier matrix. Consequently, we need only to estimate the mutual coherence of the columns of the matrix U_{θ^m} for $\theta^m \in V_{M_0}$.

Lemma 3.1. *Let $\phi'(t) \in V_{M_0}$, $t \in [0,1]$ and $\phi(0) = 0$, $\phi(1) = 1$, $\phi' > 0$, $t_j = j/L$, $j = 0, \cdot, L-1$ is a uniform grid over $[0,1]$, then for any $n \in \mathbb{N}$, there exists $C(n) > 0$, such that*

$$\frac{1}{L} \sum_{j=0}^{L-1} \phi'(t_j) e^{i2\pi k\phi(t_j)} \leqslant C(n) \max\left\{ \left(\frac{k\|\widehat{\phi'}\|_1}{L} \right)^n, \left(\frac{2M_0}{L} \right)^n \right\}. \tag{119}$$

The proof of this lemma is deferred to Appendix D.

Using this lemma, we can show that the condition $\nu_0 = \max_{k,j} |(U_{\theta^m}^* U_{\theta^m} - I)_{k,j}| \leqslant \frac{1}{16N_b}$ is satisfied as long as $N_f \geqslant C\|\mathcal{F}((\bar{\theta}^m)')\|_1 N_b$ where C is a constant determined by N_b. This leads to the following theorem.

Theorem 3.4. *Suppose the sample points t_j, $j = 1, \dots, N_s$ are selected at random from a set of uniform grid l/N_f, $l = 0, \dots, N_f - 1$. If*

$$N_f \geqslant C \big\| \mathcal{F}\big((\bar{\theta}^m)' \big) \big\|_1 N_b$$

in $(m+1)$st step, we have $\delta_S(\Phi_{\theta^m}) < 1/2$ holds with probability $1-\delta$ provided that

$$N_s \geqslant C \cdot \max\big[(\bar{\theta}^m)' \big] \big(S \log^2 N_b - \log \delta\big) \log^4 N_b, \tag{120}$$

where N_s is the number of the samples, N_b is the number of elements in the basis.

The above result shows that if the sample points are selected at random, in each step, with probability $1-\delta$, we can get the right answer. This does not mean that the whole iteration converges to the right solution with an overwhelming probability. If the iteration is run up to the nth step, the probability that all these n steps are successful is $1-n\delta$. If n is large, the probability could be small even if δ is very small.

3.2. Uniform estimate of $\delta_S(\Phi_{\theta^m})$ during the iteration

In order to make sure that the iterative algorithm would converge with a high probability, we have to obtain an uniform estimate of $\delta_S(\Phi_{\theta^m})$ during the iteration. More precisely, we need to prove that with an overwhelming probability,

$$\sup_{\theta\in W_{M_0}} \delta_S(\Phi_\theta) \leqslant 1/2, \tag{121}$$

where $W_{M_0}=\{\phi\in C^\infty[0,1]\colon\ \phi(0)=0,\ \phi(1)=1,\ \phi'\in V_{M_0},\ \phi'(t)>0,\ \forall t\in[0,1]\}$.

The analysis below shows that this is true even if the number of sample points is in the same order as that required by Theorem 3.4. There are two key observations in this analysis. The first one is that the difference between $\delta_S(\Phi_{\bar\theta})$ and $\delta_S(\Phi_{\bar\phi})$ would be small if $\bar\theta,\bar\phi\in W_{M_0}$ and $\|\bar\theta-\bar\phi\|_\infty$ is small. Actually, we can make $|\delta_S(\Phi_{\bar\theta})-\delta_S(\Phi_{\bar\phi})|\leqslant\frac{1}{4}$ as long as $\|\bar\theta'-\bar\phi'\|_\infty\leqslant r=O(N_b^{-5/2}M_0^{-1})$. The second observation is that W_{M_0} is bounded and finite dimensional which implies that its closure is compact. Then for any $r>0$, there exists a finite subset $A_r\subset W_{M_0}$, such that for any $\bar\theta\in W_{M_0}$, there exists $\bar\phi_j\in A_r$, such that $\|\bar\theta'-\bar\phi_j'\|_\infty\leqslant r$.

Based on these two observations, we can show that

$$\sup_{\theta\in W_{M_0}} \delta_S(\Phi_\theta) \leqslant \sup_{\phi\in A_r} \delta_S(\Phi_\phi)+1/4. \tag{122}$$

Then by the union bound, we have

$$P\Big(\sup_{\theta\in W_{M_0}} \delta_S(\Phi_\theta)>1/2\Big) \leqslant P\Big(\sup_{\phi\in A_r} \delta_S(\Phi_\phi)>1/4\Big) \leqslant |A_r| \sup_{\phi\in A_r} P\big(\delta_S(\Phi_\phi)>1/4\big). \tag{123}$$

It is sufficient to prove that

$$P\big(\delta_S(\Phi_\phi)>1/4\big)\leqslant \delta/|A_r|, \quad \forall\phi\in A_r\subset W_{M_0}, \tag{124}$$

which is true as long as

$$N_s\geqslant C\cdot\max_{\theta\in A_r}\big\|\theta'\big\|_\infty\big(S\log^2 N_b+\log|A_r|-\log\delta\big)\log^4 N_b. \tag{125}$$

Now, we need only to choose a proper r and estimate the corresponding $|A_r|$.

Lemma 3.2. *Let $W=\{\phi\in C^\infty[0,1]\colon\ \phi(0)=0,\ \phi(1)=1,\ \phi'\in V_{M_0},\ \phi'(t)>0,\ \forall t\in[0,1]\}$. For any $r>0$, one can find a finite subset A_r of W with cardinality*

$$|A_r|\leqslant\left(\frac{16\pi M_0^2}{r}+1\right)^{2M_0}, \tag{126}$$

such that for all $\psi\in W$, there exists $\phi\in A_r$ such that $\|\psi'-\phi'\|_\infty\leqslant r$ and $\|\psi-\phi\|_\infty\leqslant r$.

Proof. Let $\overline{W} = \{\phi' \colon \phi \in W\}$. Then for all $\bar{\psi} \in \overline{W}$, we have the following Fourier representation

$$\bar{\psi}(t) = 1 + \sum_{j=1}^{M_0} \bigl(c_j \cos(2\pi jt) + d_j \sin(2\pi jt)\bigr) > 0, \quad \forall t \in [0,1]. \tag{127}$$

Since $\int_0^t \psi(s)\,ds \in W$ according to the definition of $\overline{W}$, then $\int_0^1 \psi(s)\,ds = 1$, so the constant in the above Fourier representation is 1.

By multiplying $1 + \cos(2\pi jt)$ to both sides of (127) and integrating over $[0,1]$ with respect to t, we get

$$1 + c_j/2 \geqslant 0,$$

which implies that $c_j \geqslant -2$, where we have used the fact that $1 + \cos(2\pi jt) \geqslant 0$.

On the other hand, multiplying $-1 + \cos(2\pi jt)$ to both sides of (127) and taking integral over $[0,1]$ with respect to t, we have $c_j \leqslant 2$. Combining these two results, we have

$$|c_j| \leqslant 2. \tag{128}$$

Similarly, by multiplying $\sin(2\pi jt) \pm 1$ to both sides of (127) and taking integral over $[0,1]$ with respect to t, we obtain

$$|d_j| \leqslant 2. \tag{129}$$

Now, we have proven that for any function in $\overline{W}$, its Fourier coefficients are bounded by 2. Let $h = r/(2M_0)$, $L_r = \lceil 4/h \rceil$, $Z_r = \{-2, -2+h, -2+2h, \ldots, -2+(L_r-1)h\}$. For any $\bar{\psi} \in \overline{W}$, we know that its Fourier coefficients $c_j, d_j \in [-2,2]$, $j = 1, \ldots, M_0$, then one can find $a_j, b_j \in Z_r$ correspondingly such that

$$|a_j - c_j| \leqslant h/2 = r/(4M_0), \quad j = 1, \ldots, M_0,$$
$$|b_j - d_j| \leqslant h/2 = r/(4M_0), \quad j = 1, \ldots, M_0,$$

which implies that there exists $y \in \overline{Y}_r$ such that

$$\|\psi - y\|_\infty \leqslant \sum_{j=1}^{M_0} \bigl(|a_j - c_j| + |b_j - d_j|\bigr) \leqslant 2\pi M_0^2 h = r/2, \tag{130}$$

where $\overline{Y}_r$ is defined as follows

$$\overline{Y}_r = \left\{ y = \sum_{j=1}^{M_0} \bigl(a_j \cos(2\pi jt) + b_j \sin(2\pi jt)\bigr) \colon\ a_j, b_j \in Z_r,\ B_{r/2}(y) \cap \overline{W} \neq \emptyset \right\},$$

and $B_{r/2}(y) = \{z \in V_{M_0} \colon\ \|z - y\|_\infty \leqslant r/2\}$.

By the definition of $\overline{Y}_r$, one can get

$$|\overline{Y}_r| \leqslant |Z_r|^{2M_0} = L_r^{2M_0} \leqslant \left(\frac{8M_0}{r} + 1\right)^{2M_0}. \tag{131}$$

Suppose $\overline{Y}_r = \{y_1, y_2, \ldots, y_{|\overline{Y}_r|}\}$, by the definition of $\overline{Y}_r$, for each y_j, there exists $\bar{\phi}_j \in \overline{W}$ such that $\bar{\phi}_j \in B_{r/2}(y_j)$. We can get a finite subset $\overline{A}_r$ of $\overline{W}$ by collecting all these $\bar{\phi}_j$ together and obviously $|\overline{A}_r| = |\overline{Y}_r|$.

Finally, let

$$A_r = \left\{ \int_0^t \bar{\phi}(s)\,ds\colon\ \bar{\phi} \in \bar{A}_r \right\}. \tag{132}$$

Then, for any $\psi \in W$, there exists $\phi_j \in A_r$ and $y_j \in \overline{Y}_r$, such that

$$\left\|\psi' - \phi_j'\right\|_\infty \leqslant \left\|\psi' - y_j\right\|_\infty + \left\|y_j - \phi_j'\right\|_\infty \leqslant r/2 + r/2 = r. \tag{133}$$

Moreover, we have

$$\|\psi - \phi_j\|_\infty \leqslant \int_0^1 \left|\psi'(s) - \phi_j'(s)\right| ds \leqslant r, \tag{134}$$

where we have used the fact that $\psi(0) = \phi_j(0) = 0$ to eliminate the integral constant. □

Remark 3.2. By multiplying $c_j \cos(2\pi jt) + d_j \sin(2\pi jt) \pm \sqrt{c_j^2 + d_j^2}$ to both sides of (127) and taking integral over $[0, 1]$ with respect to t, we have

$$c_j^2 + d_j^2 \leqslant 4, \quad j = 1, \ldots, M_0. \tag{135}$$

This implies a sharper estimate of $|A_r|$,

$$|A_r| \leqslant \left(\frac{8\pi M_0^2}{r^2}\right)^{M_0}. \tag{136}$$

Also, (135) gives us a bound for $\|\phi'\|_\infty$ in W_{M_0},

$$\sup_{\phi \in W_{M_0}} \left\|\phi'\right\|_\infty \leqslant 4M_0 + 1, \tag{137}$$

which will be used later.

It remains to choose a proper r. First, we show that the difference of δ_S between two matrices can be controlled by the difference of each element.

Proposition 3.1. *Let A, B are two M by N matrices, $M < N$ and the columns of A are normalized to be unit vectors in l^2 norm. Then, for any $S \in \mathbb{N}$, we have*

$$\left|\delta_S(A) - \delta_S(B)\right| \leqslant \left(2\epsilon\sqrt{M} + \epsilon^2 M\right)S, \tag{138}$$

where $\epsilon = \max_{i,j} |A_{ij} - B_{ij}|$.

Proof. By the definition of δ_S, we need only to prove that for all subsets T with $|T| \leqslant S$ and coefficients sequences $(c_j)_{j\in T}$,

$$\left|\|A_T c\|_2^2 - \|B_T c\|_2^2\right| \leqslant \left(2\epsilon\sqrt{M} + \epsilon^2 M\right)S\|c\|_2^2. \tag{139}$$

This can be verified by a direct calculation:

$$\begin{aligned}
\left|\|A_T c\|_2^2 - \|B_T c\|_2^2\right| &= \left|\sum_{i,j\in T} c_i c_j \left(A_i^T A_j - B_i^T B_j\right)\right| \\
&= \left|\sum_{i,j\in T} c_i c_j \left(D_i^T A_j + A_i^T D_j + D_i^T D_j\right)\right| \\
&\leqslant \max_{i,j\in T}\left|D_i^T A_j + A_i^T D_j + D_i^T D_j\right| \sum_{i,j\in T} |c_i c_j| \\
&\leqslant |T|\|c\|_2^2 \max_{i,j\in T}\left(\|D_i\|_2\|A_j\|_2 + \|A_i\|_2\|D_j\|_2 + \|D_i\|_2\|D_j\|_2\right) \\
&\leqslant \left(2\epsilon\sqrt{M}\max_{i\in\mathbb{Z}_N}\|A_i\|_2 + \epsilon^2 M\right) S\|c\|_2^2. \qquad (140)
\end{aligned}$$

In the above derivation, $D = B - A$, A_i, A_j are ith and jth columns of A. □

Using the above proposition, we obtain the following result:

Corollary 3.1. *Let $\bar{\theta}, \bar{\phi} \in W$, then*

$$\left|\delta_S(\Phi_{\bar{\theta}}) - \delta_S(\Phi_{\bar{\phi}})\right| \leqslant \frac{1}{8}, \qquad (141)$$

provided that $|\bar{\theta}' - \bar{\phi}'| \leqslant C N_b^{-2} M_0^{-1/2}$, where C is an absolute constant.

Proof. Assume that $|\bar{\theta}' - \bar{\phi}'| \leqslant \epsilon$. We need only to show that the difference between $\Phi_{\bar{\theta}}$ and $\Phi_{\bar{\phi}}$ can be controlled by ϵ. This is quite straightforward using the definition of $\Phi_{\bar{\theta}}$ and $\Phi_{\bar{\phi}}$:

$$\begin{aligned}
\left|\Phi_{\bar{\theta}}(j,k) - \Phi_{\bar{\phi}}(j,k)\right| &= \frac{1}{\sqrt{N_s}}\left|\sqrt{\bar{\theta}'(t_j)} e^{i2\pi k\bar{\theta}(t_j)} - \sqrt{\bar{\phi}'(t_j)} e^{i2\pi k\bar{\phi}(t_j)}\right| \\
&\leqslant \frac{\left|\sqrt{\bar{\theta}'(t_j)} - \sqrt{\bar{\phi}'(t_j)}\right|}{\sqrt{N_s}} + \frac{\sqrt{\bar{\theta}'(t_j)}}{\sqrt{N_s}}\left|e^{i2\pi k(\bar{\theta}(t_j)-\bar{\phi}(t_j))} - 1\right| \\
&\leqslant \frac{\left|\sqrt{\bar{\theta}'(t_j)} - \sqrt{\bar{\phi}'(t_j)}\right|}{\sqrt{N_s}} + \frac{\sqrt{\bar{\theta}'(t_j)}}{\sqrt{N_s}} 2\pi k\left|\bar{\theta}(t_j) - \bar{\phi}(t_j)\right| \\
&\leqslant \frac{\sqrt{\epsilon}}{\sqrt{N_s}} + \frac{2\pi N_b \epsilon\sqrt{4M_0+1}}{\sqrt{N_s}}, \qquad (142)
\end{aligned}$$

where we have used the estimate $\|\bar{\theta}'\|_\infty \leqslant 4M_0 + 1$ given in (137). Using Proposition 3.1 and the fact that $S \leqslant N_b$, we can complete the proof. □

Combining Lemma 3.2, Corollary 3.1 and (125), we have the following theorem,

Theorem 3.5. $\sup_{\theta\in W_{M_0}} \delta_S(\Phi_\theta) \leqslant 1/2$ *holds with probability $1-\delta$ provided that*

$$N_s \geqslant C \cdot (4M_0+1)\left(S\log^2 N_b + M_0 \log N_b - \log\delta\right)\log^4 N_b, \qquad (143)$$

where N_s is the number of the samples, N_b is the number of elements in the basis.

Remark 3.3. Comparing with the condition stated in Theorem 3.4, we require extra $M_0 \log^5 N_b$ samples in order to get the uniform estimate. But this number $M_0 \log^5 N_b$ can be absorbed by $S\log^6 N_b$, since S is larger than M_0. Thus the condition to get an uniform estimate is essentially the same as that in Theorem 3.4.

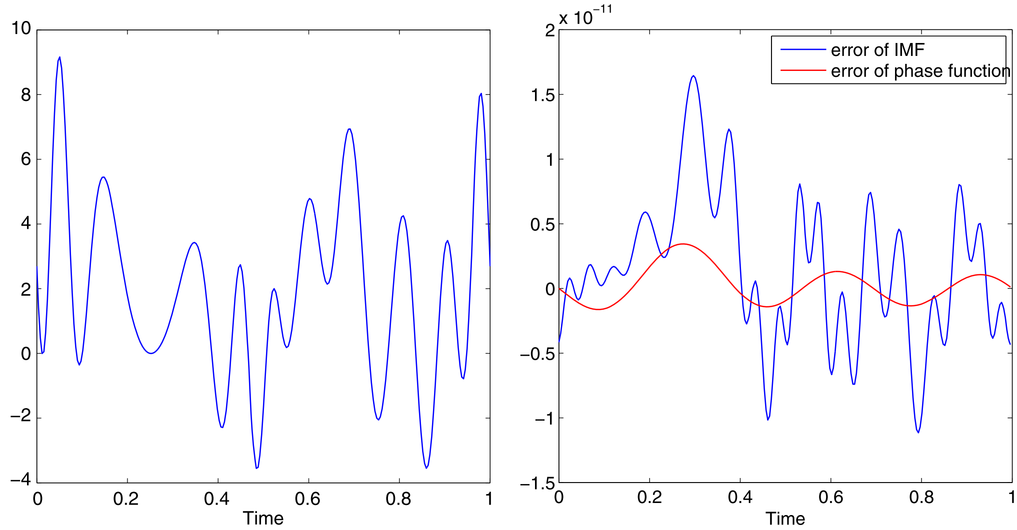

Fig. 1. Left: Original signal; Right: Error of the IMF and the phase function.

4. Numerical results

In this section, we will perform several numerical experiments to confirm our theoretical results presented in the previous section and to demonstrate the performance of the algorithm based on the weighted l^1 optimization.

Example 1 *(Exact recovery for a well-resolved signal).* The first example is a well-resolved periodic signal. The signal we use is generated by

$$f = (2 + \cos\bar{\theta} + 2\sin 2\bar{\theta} + \cos 3\bar{\theta}) + (3 + \cos\bar{\theta} + \sin 3\bar{\theta})\cos\theta, \tag{144}$$

where the phase function θ is given by

$$\theta = 20\pi t + 2\cos 2\pi t + 2\sin 4\pi t, \quad \bar{\theta} = \theta/10.$$

This signal is sampled over a uniform mesh of 256 points such that there are about 12 samples in each period of the signal on average to make sure that the signal is well-resolved by the samples.

In this example, the non-zero Fourier coefficients of mean a_0 and the amplitude a_1 in the θ-space are confined to the low frequency band. In this case, the corresponding $M_1 = 3$. The instantaneous frequency also consists of the low frequency Fourier modes, $M_0 = 2$ in this example. The parameter L, which is the number of periods, is equal to 10. In the computation, the initial guess of the phase function, θ^0, is chosen to be $20\pi t$.

From this example, we can see that the estimate in Theorem 2.1 is far from being sharp. It is easy to check that the initial condition does not satisfy the condition (64) in Theorem 2.1. Moreover, L is not as large as that required in (59)–(61) ($L < 4M_1$ and $\eta_0 \approx 30 > L$ when $n = 3$). But in the computation, Algorithm 1 is still capable to recover the exact result up to the interpolation error.

The numerical results are shown in Figs. 1 and 2. In Fig. 1, we can see that our algorithm indeed recovers the exact decomposition of this signal. This is also consistent with the theoretical result we obtained in Theorem 2.1. The result shown in Fig. 1 is obtained by applying the non-uniform Fourier transform directly by solving a linear system. As we proposed in our algorithm, for a well-resolved signal, it is more efficient to use a combination of interpolation and FFT. This procedure would introduce some interpolation error,

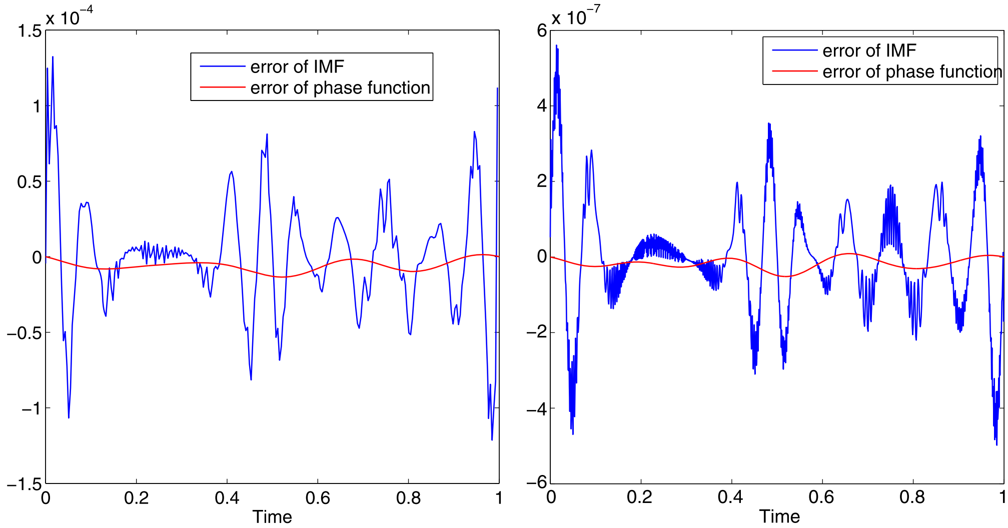

Fig. 2. Left: Error of the IMF and the phase function with 256 uniform samples; Right: Error of the IMF and the phase function with 1024 uniform samples.

however the computation is accelerated tremendously. As we see in Fig. 2, if we use the FFT-based algorithm, the error increase to the order of 10^{-4} instead of 10^{-11} in the previous result when we used the non-uniform Fourier transform. If we increase the number of sample points to 1024, the order of error decreases to 10^{-7}. This indicates that the main source of error comes from the interpolation error.

In our previous paper [11], we have shown many numerical results to demonstrate the stability of our algorithm. These numerical examples confirm the theoretical results presented in Theorems 2.3 and 2.4. We will not reproduce these numerical examples in this paper.

Example 2 *(Exact recovery for a signal with random samples).* The second example is designed to confirm the result of Theorem 3.1. This example shows that for a signal with a sparse structure, our algorithm is capable of producing the exact decomposition even if it is poorly sampled. The signal is given below:

$$f = \cos\bar{\theta} + (3 + \cos\bar{\theta} + \sin 2\bar{\theta})\cos\theta, \tag{145}$$

where the phase function θ is

$$\theta = 200\pi t - 10\cos 2\pi t - 2\sin 4\pi t, \quad \bar{\theta} = \theta/(100).$$

In this case, the corresponding parameters are $M_0 = 2$, $M_1 = 2$ and $L = 100$. The ratio between L and M_0, M_1 is much larger than that in the previous example. The initial guess is given by $\theta^0 = 200\pi t$.

The number of sample points is set to be 120. These sample points are selected at random over 4096 uniformly distributed points. On average, there are only 1.2 points in each period of the signal. We test 100 independent samples and our algorithm is able to recover the signal for 97 samples, which gives 97% success rate. Fig. 3 gives one of the successful samples.

The right panel of Fig. 3 shows that the order of error is 10^{-2} for IMF and 10^{-3} for the phase function. In the computation, the l^1 optimization problem is solved approximately in each step of the iteration. This is the reason that the error is much larger than the round-off error of the computer. If we increase the accuracy in solving the l^1 optimization problem, the algorithm would give a more accurate result. However the computational cost also increases as a consequence. We also reduce the number of sample points to 80 and carry out the same test for 100 times. In this case, the recovery rate was 46 out of 100.

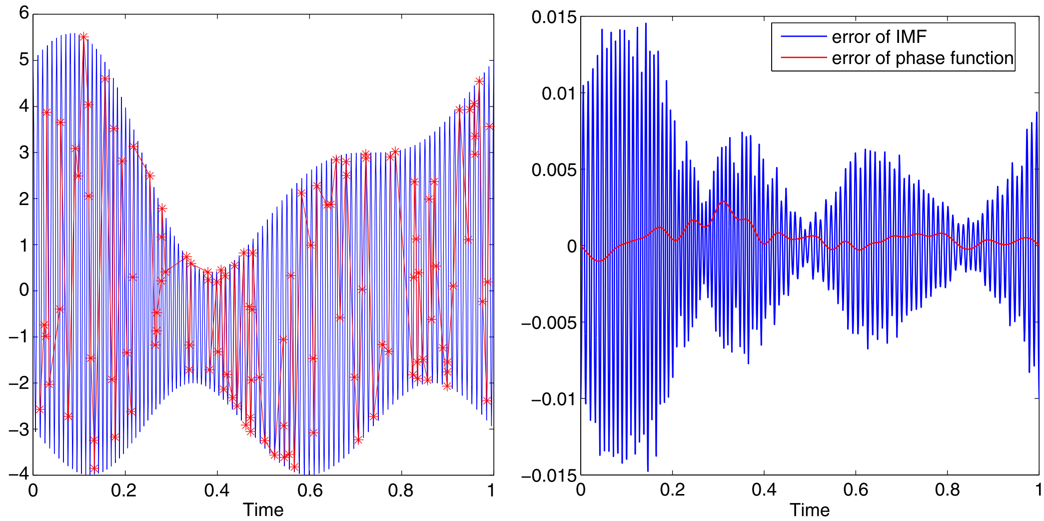

Fig. 3. Left: Original signal and the sample points; Right: Error of the IMF and phase function.

Example 3 *(Approximate recovery for a signal with random samples).* In this example, we will check the stability of our algorithm for a poorly sampled signal. The signal is generated by

$$f = \cos(2\pi t) + \big(3 + \cos(2\pi t) + \sin(4\pi t)\big)\cos\theta + 0.1X(t), \tag{146}$$

where the phase function θ is

$$\theta = \tilde{\theta} + 0.1\sin(120\pi t)$$

and $\tilde{\theta}$ is the phase function in Example 2, and $X(t)$ is the Gaussian noise with standard deviation $\sigma^2 = 1$.

Comparing with the signal in Example 2, we add one small high frequency component on the phase function, whose wave number is 60. In S3 of Algorithm 2, M_0 is set to be 20, which implies that the high frequency component of the phase function can not be captured in the computation. Moreover, the mean and amplitude are not exactly low frequency confined over the Fourier basis in the θ-space. This would also introduce some truncation error in the computation. We also add white noise to the original signal to make it even more challenging to decompose. The initial guess of the phase function is also $200\pi t$.

In this example, when the number of sample points is 120, our method can give 92 successful recoveries in 100 independent tests. Fig. 4 gives one of the successful recoveries obtained by our algorithm. Due to the truncation error and the noise, the error becomes much larger than that in the previous example. But all the errors are comparable with the magnitude of the truncation error and noise, which shows that our method has good stability even for signals with poor samples. When the number of samples is reduced to 80, the recovery rate drops to 40 out of 100.

5. Concluding remarks

In this paper, we analyze the convergence of the data-driven time–frequency analysis method proposed in [11]. First, we considered the case when the number of sample points is large enough. We proved that the algorithm we developed would converge to the exact decomposition if the signal satisfies some low frequency mode confinement condition in the coordinate determined by the phase function. Our convergence analysis has been extended to cover signals polluted by noise. We also proved the convergence of our method with an approximate decomposition when the signal does not have the exact low frequency confinement property but its spectral coefficients have a fast decay.

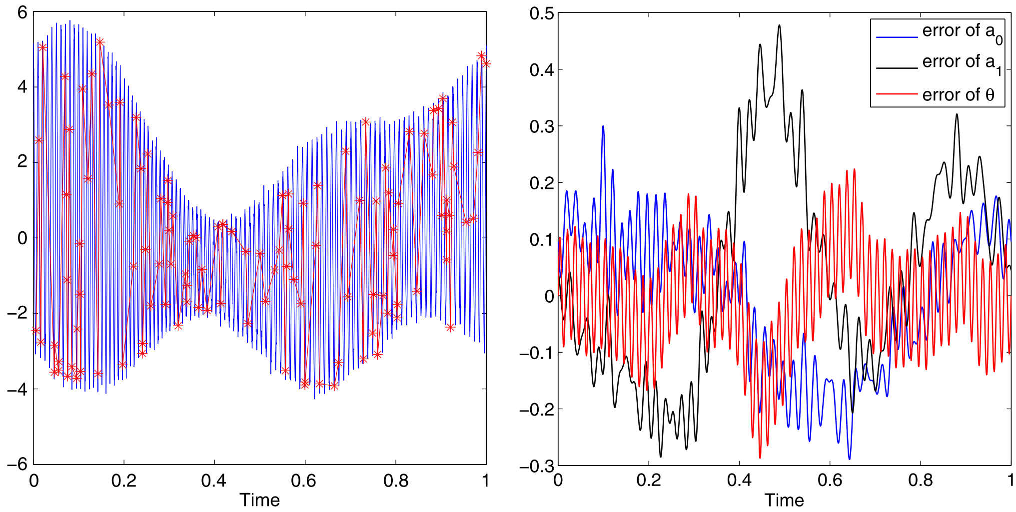

Fig. 4. Left: Original signal (blue) and the sample points (red) in Example 3; Right: Errors of a_0, a_1 and θ. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We further considered the more challenging case when only a few number of samples are given which do not resolve the original signal accurately. In this case, we need to solve a l^1 minimization problem which is computationally more expensive. We proved the stability and convergence of our method by using some results developed in compressive sensing. As in compressive sensing, the convergence and stability of our method assumes that certain S-restricted isometry condition is satisfied. We proved that for each fixed step in the iteration, this S-restricted isometry condition is satisfied with an overwhelming probability if the sample points are selected at random.

We presented numerical evidence to support our theoretical results. Our numerical results confirmed the theoretical results in all cases that we considered.

We are currently working on the convergence of the data-driven time–frequency analysis method for non-periodic signals. Our extensive numerical results seem to indicate that our method also converges for non-periodic signals. The theoretical analysis for this problem is more challenging. We will report the result in a subsequent paper.

Acknowledgments

This work was in part supported by the AFOSR MURI grant FA9550-09-1-0613, a DOE grant DE-FG02-06ER25727, and NSF Grants DMS-1159138 and DMS-1318377. The research of Dr. Z. Shi was in part supported by a NSFC Grant 11201257.

Appendix A. Error of the amplitude functions

Suppose

$$f(t) = f_0(t) + f_1(t)\cos\theta \tag{147}$$

is the signal we want to decompose. Let $a^m = f_1 \cos\Delta\theta^m$, $b^m = f_1 \sin\Delta\theta^m$, then we have

$$f = f_0 + a^m \cos\theta^m - b^m \sin\theta^m. \tag{148}$$

Let $L^m = \frac{\theta^m(T)-\theta^m(0)}{2\pi}$ and $\bar{\theta}^m = \theta^m/(2\pi L^m)$. Then we can rewrite f as follows:

$$f = f_0 + a^m \cos 2\pi L^m \bar{\theta}^m - b^m \sin 2\pi L^m \bar{\theta}^m. \tag{149}$$

Define the Fourier transform in the $\bar{\theta}$-space as:

$$\hat{f}_{\theta^m} = \int_0^1 f(t) e^{-i2\pi k \bar{\theta}^m}\, d\bar{\theta}^m. \tag{150}$$

Applying Fourier transform to both sides of (149), we have

$$\hat{f}_{\theta^m}(k) = \hat{f}_{0,\theta^m}(k) + \frac{1}{2}\big(\hat{a}^m_{\theta^m}\big(k+L^m\big) + \hat{a}^m_{\theta^m}\big(k-L^m\big)\big) - \frac{i}{2}\big(\hat{b}^m_{\theta^m}\big(k+L^m\big) - \hat{b}^m_{\theta^m}\big(k-L^m\big)\big). \tag{151}$$

Then, we get

$$\begin{aligned}
\hat{a}^m_{\theta^m}(k) - i\hat{b}^m_{\theta^m}(k) &= 2\hat{f}_{\theta^m}\big(k-L^m\big) - 2\hat{f}_{0,\theta^m}\big(k-L^m\big) - \hat{a}^m_{\theta^m}\big(k-2L^m\big) - i\hat{b}^m_{\theta^m}\big(k-2L^m\big),\\
\hat{a}^m_{\theta^m}(k) + i\hat{b}^m_{\theta^m}(k) &= 2\hat{f}_{\theta^m}\big(k+L^m\big) - 2\hat{f}_{0,\theta^m}\big(k+L^m\big) - \hat{a}^m_{\theta^m}\big(k+2L^m\big) + i\hat{b}^m_{\theta^m}\big(k+2L^m\big).
\end{aligned}$$

It is easy to solve for $\hat{a}^m_{\theta^m}$ and $\hat{b}^m_{\theta^m}$ to obtain:

$$\begin{aligned}
\hat{a}^m_{\theta^m}(k) = {} & \hat{f}_{\theta^m}\big(k+L^m\big) + \hat{f}_{\theta^m}\big(k-L^m\big) - \Big[\hat{f}_{0,\theta^m}\big(k+L^m\big) + \hat{f}_{0,\theta^m}\big(k-L^m\big)\\
& + \frac{1}{2}\big(\hat{a}^m_{\theta^m}\big(k+2L^m\big) + \hat{a}^m_{\theta^m}\big(k-2L^m\big)\big) - \frac{i}{2}\big(\hat{b}^m_{\theta^m}\big(k+2L^m\big) - \hat{b}^m_{\theta^m}\big(k-2L^m\big)\big)\Big],
\end{aligned} \tag{152}$$

$$\begin{aligned}
\hat{b}^m_{\theta^m}(k) = {} & -i\big(\hat{f}_{\theta^m}\big(k+L^m\big) - \hat{f}_{\theta^m}\big(k-L^m\big)\big) + i\Big[\hat{f}_{0,\theta^m}\big(k+L^m\big) - \hat{f}_{0,\theta^m}\big(k-L^m\big)\\
& + \frac{1}{2}\big(\hat{a}^m_{\theta^m}\big(k+2L^m\big) - \hat{a}^m_{\theta^m}\big(k-2L^m\big)\big) - \frac{i}{2}\big(\hat{b}^m_{\theta^m}\big(k+2L^m\big) + \hat{b}^m_{\theta^m}\big(k-2L^m\big)\big)\Big].
\end{aligned} \tag{153}$$

In our algorithm, $\mathcal{F}_{\theta^m}(\tilde{a}^m)$ and $\mathcal{F}_{\theta^m}(\tilde{b}^m)$ are approximated in the following way:

$$\hat{\tilde{a}}^m_{\theta^m}(k) = \begin{cases} \hat{f}_{\theta^m}(k+L^m) + \hat{f}_{\theta^m}(k-L^m), & -L^m/2 \leqslant k \leqslant L^m/2,\\ 0, & \text{otherwise}, \end{cases} \tag{154}$$

$$\hat{\tilde{b}}^m_{\theta^m}(k) = \begin{cases} -i(\hat{f}_{\theta^m}(k+L^m) - \hat{f}_{\theta^m}(k-L^m)), & -L^m/2 \leqslant k \leqslant L^m/2,\\ 0, & \text{otherwise}. \end{cases} \tag{155}$$

We then get the error of the approximation in the spectral space:

$$\widehat{\Delta a}^m_{\theta^m}(k) = \begin{cases} -[\hat{f}_{0,\theta^m}(k+L^m) + \hat{f}_{0,\theta^m}(k-L^m) + \frac{1}{2}(\hat{a}^m_{\theta^m}(k+2L^m) + \hat{a}^m_{\theta^m}(k-2L^m)) \\ \quad - \frac{i}{2}(\hat{b}^m_{\theta^m}(k+2L^m) - \hat{b}^m_{\theta^m}(k-2L^m))], & |k| \leqslant L^m/2,\\ \hat{a}^m_{\theta^m}(k), & |k| > L^m/2, \end{cases}$$

$$\widehat{\Delta b}^m_{\theta^m}(k) = \begin{cases} i[\hat{f}_{0,\theta^m}(k+L^m) - \hat{f}_{0,\theta^m}(k-L^m) + \frac{1}{2}(\hat{a}^m_{\theta^m}(k+2L^m) - \hat{a}^m_{\theta^m}(k-2L^m)) \\ \quad - \frac{i}{2}(\hat{b}^m_{\theta^m}(k+2L^m) + \hat{b}^m_{\theta^m}(k-2L^m))], & |k| \leqslant L^m/2,\\ \hat{b}^m_{\theta^m}(k), & |k| > L^m/2. \end{cases}$$

Thus, we have the following inequality for the l^1 norm of the error in the spectral space:

$$\begin{aligned}|\Delta a^m| &\leqslant \|\widehat{\Delta a}^m_{\theta^m}\|_1\\ &\leqslant 2\sum_{\frac{L^m}{2}<k<\frac{3}{2}L^m}|\hat{f}_{0,\theta^m}(k)| + \sum_{\frac{3}{2}L^m<k<\frac{5}{2}L^m}\left(|\hat{a}^m_{\theta^m}(k)| + |\hat{b}^m_{\theta^m}(k)|\right) + \sum_{|k|>\frac{L^m}{2}}|\hat{a}^m_{\theta^m}(k)|. \end{aligned} \tag{156}$$

Similarly, we get

$$|\Delta b^m| \leqslant 2\sum_{\frac{L^m}{2}<k<\frac{3}{2}L^m}|\hat{f}_{0,\theta^m}(k)| + \sum_{\frac{3}{2}L^m<k<\frac{5}{2}L^m}\left(|\hat{a}^m_{\theta^m}(k)| + |\hat{b}^m_{\theta^m}(k)|\right) + \sum_{|k|>\frac{L^m}{2}}|\hat{b}^m_{\theta^m}(k)|. \tag{157}$$

In the above derivation, we assume that the Fourier transform of f in θ^m-space can be calculated exactly. If only approximate Fourier transform is available, denoted as $\hat{\hat{f}}_{\theta^m}$, such as the signal with poor samples as we discussed in Section 3, there would be an extra term in the estimates of Δa^m and Δb^m,

$$\begin{aligned}|\Delta a^m| \leqslant{}& 2\sum_{\frac{L^m}{2}<k<\frac{3}{2}L^m}|\hat{f}_{0,\theta^m}(k)| + \sum_{\frac{3}{2}L^m<k<\frac{5}{2}L^m}\left(|\hat{a}^m_{\theta^m}(k)| + |\hat{b}^m_{\theta^m}(k)|\right)\\ &+ \sum_{|k|>\frac{L^m}{2}}|\hat{a}^m_{\theta^m}(k)| + 2\sum_{\frac{L^m}{2}<k<\frac{3}{2}L^m}|\hat{f}_{\theta^m}(k) - \hat{\hat{f}}_{\theta^m}(k)|, \end{aligned} \tag{158}$$

$$\begin{aligned}|\Delta b^m| \leqslant{}& 2\sum_{\frac{L^m}{2}<k<\frac{3}{2}L^m}|\hat{f}_{0,\theta^m}(k)| + \sum_{\frac{3}{2}L^m<k<\frac{5}{2}L^m}\left(|\hat{a}^m_{\theta^m}(k)| + |\hat{b}^m_{\theta^m}(k)|\right)\\ &+ \sum_{|k|>\frac{L^m}{2}}|\hat{b}^m_{\theta^m}(k)| + 2\sum_{\frac{L^m}{2}<k<\frac{3}{2}L^m}|\hat{f}_{\theta^m}(k) - \hat{\hat{f}}_{\theta^m}(k)|. \end{aligned} \tag{159}$$

Appendix B. Proof of Lemma 2.2

Proof. By a direct calculation, it is easy to verify that

$$s_L(t) = \int_0^1 s(\tau)g_L(t-\tau)\,d\tau, \tag{160}$$

where g_L is a periodic function over $[0,1]$ given by

$$g_L(t) = \mathcal{F}^{-1}\left[\left(\chi(1+k/L) + \chi(1-k/L)\right)\right] = \sum_{|k|<L/2,k\in\mathbb{Z}}\left(e^{i2\pi(k+L)t} + e^{i2\pi(k-L)t}\right), \quad \forall t \in \mathbb{R}. \tag{161}$$

Then we have

$$|s_L(t)| \leqslant \|s\|_\infty \int_0^1 |g_L(\tau-t)|\,d\tau = \|s\|_\infty \int_{-1/2}^{1/2}|g_L(t)|\,dt, \tag{162}$$

where we have used the fact that g_L is periodic over $[0,1]$. Define

$$G_L(t) = \mathcal{F}_{\mathbb{R}}^{-1}\left[\left(\chi(1+\omega/L) + \chi(1-\omega/L)\right)\right] = \int_{-3L/2}^{-L/2} e^{i2\pi\omega t}\,d\omega + \int_{L/2}^{3L/2} e^{i2\pi\omega t}\,d\omega, \quad \forall t \in \mathbb{R}, \tag{163}$$

where $\mathcal{F}_{\mathbb{R}}$ is the Fourier transform over the whole real axis $\mathbb{R}$.

Utilizing the definition of g_L and the definition of G_L and the relation that

$$\int_{k-1/2}^{k+1/2} e^{i2\pi\omega t}\,d\omega = \frac{\sin\pi t}{\pi t}e^{i2\pi kt}, \quad \forall k\in\mathbb{Z}, \tag{164}$$

it is easy to show that

$$g_L(t) = \frac{\pi t}{\sin\pi t}G_L(t), \quad \forall t\in\mathbb{R}. \tag{165}$$

This leads to the following estimate

$$\int_{-1/2}^{1/2} \left|g_L(t)\right|dt \leqslant \max_{t\in[-1/2,1/2]}\left|\frac{\pi t}{\sin\pi t}\right| \int_{-1/2}^{1/2}\left|G_L(t)\right|dt \leqslant \frac{\pi}{2}\int_{-\infty}^{+\infty}\left|G_L(t)\right|dt. \tag{166}$$

Notice that $G_L(t) = LG_1(Lt)$ which implies that

$$\int_{-\infty}^{+\infty}\left|G_L(t)\right|dt = \int_{-\infty}^{+\infty} L\left|G_1(Lt)\right|dt = \int_{-\infty}^{+\infty}\left|G_1(t)\right|dt. \tag{167}$$

Then the lemma can be proved by setting $\Gamma_s = \frac{\pi}{2}\|G_1\|_1$. □

Appendix C. Estimates of $\hat{f}_{0,\theta^m}(\omega)$, $\hat{a}^m_{\theta^m}(\omega)$ and $\hat{b}^m_{\theta^m}(\omega)$ in Theorem 2.1

We first estimate f_0. We proceed as follows:

$$\begin{aligned}
\left|\hat{f}_{0,\theta^m}(\omega)\right| &= \left|\int_0^1 f_0(t)e^{-i2\pi\omega\bar{\theta}^m}\,d\bar{\theta}^m\right| \\
&= \left|\int_0^1 \sum_{|k|\leqslant M_1}\hat{f}_{0,\theta}(k)e^{i2\pi(k\bar{\theta}-\omega\bar{\theta}^m)}\,d\bar{\theta}^m\right| \\
&= \left|\sum_{|k|\leqslant M_1}\hat{f}_{0,\theta}(k)\int_0^1 e^{i2\pi(k\bar{\theta}-\omega\bar{\theta}^m)}\,d\bar{\theta}^m\right| \\
&= \left|\sum_{|k|\leqslant M_1}\hat{f}_{0,\theta}(k)\int_0^1 e^{i2\pi(\alpha k-\omega)\bar{\theta}^m}e^{ik\Delta\theta^m/L}\,d\bar{\theta}^m\right|,
\end{aligned} \tag{168}$$

where $\alpha = L^m/L$. In the last equality, we have used the fact that $\theta = 2\pi L\bar{\theta}$, $\theta^m = 2\pi L^m\bar{\theta}^m$ and $\theta = \theta^m + \Delta\theta^m$.

Using Lemma 2.1, we obtain for any $|\omega| > L/2$ that

$$\left|\hat{f}_{0,\theta^m}(\omega)\right| \leqslant \left|\sum_{|k|\leqslant M_1}\hat{f}_{0,\theta}(k)\int_0^1 e^{i2\pi(\alpha k-\omega)\bar{\theta}^m}e^{ik\Delta\theta^m/L}\,d\bar{\theta}^m\right|$$

$$\leqslant C_0 \sum_{|k|\leqslant M_1} \frac{QM_0^n}{|\omega-\alpha k|^n} \sum_{j=1}^n \left|\frac{k}{L}\right|^j (2\pi M_0)^{-j} \left\| \mathcal{F}_{\theta^m}\left[(\Delta\theta^m)'\right]\right\|_1^j$$

$$\leqslant 2C_0 Q\left(\frac{|\omega|}{2}\right)^{-n} M_0^n M_1 \sum_{j=1}^n (M_1\gamma/L)^j, \tag{169}$$

where $C_0 = \max_{|k|\leqslant M_1}(|\hat{f}_{0,\theta}(k)|, |\hat{f}_{1,\theta}(k)|)$ and

$$Q = \frac{P(z,n)}{(\min(\bar{\theta}^m)')^n}, \qquad z = \frac{\|\mathcal{F}[(\bar{\theta}^m)']\|_1}{\min(\bar{\theta}^m)'}, \qquad \gamma = \frac{\|\mathcal{F}[(\Delta\theta^m)']\|_1}{2\pi M_0}. \tag{170}$$

In the above derivation, we need to assume that $L \geqslant 4M_1$ such that $|\omega-\alpha k| \geqslant |\omega|/2$ for all $|\omega| \geqslant L/2$ and $|k| \leqslant M_1$.

If we further assume that $\gamma \leqslant 1/4$, we have

$$\left|\hat{f}_{0,\theta^m}(\omega)\right| \leqslant C_0 Q\left(\frac{|\omega|}{2}\right)^{-n} M_0^n M_1\gamma. \tag{171}$$

Next, we estimate $\hat{a}_{\theta^m}^m$. The method of analysis is similar to the previous one, however the derivation is a little more complicated. We proceed as follows:

$$\left|\hat{a}_{\theta^m}^m(\omega)\right| = \left|\int_0^1 f_1(t)\cos\Delta\theta^m(t) e^{-i2\pi\omega\bar{\theta}^m}\, d\bar{\theta}^m\right|$$

$$\leqslant \frac{1}{2}\left|\int_0^1 \sum_{|k|\leqslant M_1} \hat{f}_{1,\theta}(k) e^{i2\pi k\bar{\theta}}\left(e^{i\Delta\theta} + e^{-i\Delta\theta}\right) e^{-i2\pi\omega\bar{\theta}^m}\, d\bar{\theta}^m\right|$$

$$\leqslant \frac{1}{2}\left|\sum_{|k|\leqslant M_1} \hat{f}_{1,\theta}(k) \int_0^1 e^{i2\pi(\alpha k-\omega)\bar{\theta}^m} e^{i(k+L)\Delta\theta/L}\, d\bar{\theta}^m\right|$$

$$+ \frac{1}{2}\left|\sum_{|k|\leqslant M_1} \hat{f}_{1,\theta}(k) \int_0^1 e^{i2\pi(\alpha k-\omega)\bar{\theta}^m} e^{i(k-L)\Delta\theta/L}\, d\bar{\theta}^m\right|. \tag{172}$$

For the first term in the above inequality, we have that for any $|\omega| > L/2$,

$$\left|\sum_{|k|\leqslant M_1} f_{1,\theta}(k) \int_0^1 e^{i2\pi(\alpha k-\omega)\bar{\theta}^m} e^{i(k+L)\Delta\theta/L}\, d\bar{\theta}^m\right|$$

$$\leqslant C_0 Q \sum_{|k|\leqslant M_1} \frac{M_0^n}{|\omega-\alpha k|^n} \sum_{j=1}^n \left|1+\frac{k}{L}\right|^j \gamma^j$$

$$\leqslant C_0 Q\left(\frac{|\omega|}{2}\right)^{-n} M_0^n \sum_{j=1}^n 2^{j-1}\gamma^j \sum_{|k|\leqslant M_1}\left(1+\left|\frac{k}{L}\right|^j\right)$$

$$\leqslant 4C_0 Q\left(\frac{|\omega|}{2}\right)^{-n} M_0^n (2M_1+1)\gamma. \tag{173}$$

Here we assume that $L \geqslant 4M_1$, $\gamma \leqslant 1/4$. The definition of Q and γ can be found in (43).

For the second term in (172), we can get the same bound for $|\omega| \geqslant L/2$,

$$\left|\sum_{|k|\leqslant M_1} f_{1,\theta}(k) \int_0^1 e^{i2\pi(\alpha k-\omega)\bar{\theta}^m} e^{i(k-L)\Delta\theta/L}\, d\bar{\theta}^m\right| \leqslant 4C_0Q\left(\frac{|\omega|}{2}\right)^{-n} M_0^n(2M_1+1)\gamma. \tag{174}$$

By combining (172), (173) and (174), we obtain a complete control of $\widehat{\bar{a}}$,

$$\left|\hat{a}_{\theta^m}^m(\omega)\right| \leqslant 4C_0Q\left(\frac{|\omega|}{2}\right)^{-n} M_0^n(2M_1+1)\gamma, \quad \forall|\omega| \geqslant L/2. \tag{175}$$

Similarly, we can estimate $\hat{b}_{\theta^m}^m$ by the same upper bound,

$$\left|\hat{b}_{\theta^m}^m(\omega)\right| \leqslant 4C_0Q\left(\frac{|\omega|}{2}\right)^{-n} M_0^n(2M_1+1)\gamma, \quad \forall|\omega| \geqslant L/2. \tag{176}$$

Appendix D. Proof of Lemma 3.1

Proof. Since $e^{i2\pi k\phi}$ is a periodic function over $[0,1]$, it can be represented by Fourier series:

$$e^{i2\pi k\phi(t)} = \sum_{l=-\infty}^{+\infty} d_l e^{i2\pi lt}, \quad t \in [0,1], \tag{177}$$

where $d_l = \int_0^1 e^{i2\pi k\phi(t)} e^{-i2\pi lt}\, dt$. By assumption, we have $\phi'(t) \in V_{M_0}$. Thus, we get

$$\phi'(t) = \sum_{j=-M_0}^{M_0} c_j e^{i2\pi jt}, \quad t \in [0,1], \tag{178}$$

where $c_j = \int_0^1 \bar{\theta}'(t) e^{-i2\pi jt}\, dt$.

Further, we have

$$\begin{aligned}
\frac{1}{L}\sum_{m=0}^{L-1} \phi'(t_m) e^{i2\pi k\phi(t_m)} &= \frac{1}{L}\sum_{m=0}^{L-1}\sum_{j=-M_0}^{M_0}\sum_{l=-\infty}^{+\infty} c_j\, d_l e^{i2\pi(l+j)t_m} \\
&= \frac{1}{L}\sum_{j=-M_0}^{M_0}\sum_{l=-\infty}^{+\infty} c_j\, d_l \sum_{m=0}^{L-1} e^{i2\pi(l+j)m/L} \\
&= \sum_{j=-M_0}^{M_0}\sum_{p\in\mathbb{Z}} c_j\, d_{pL-j} \\
&= \sum_{j=-M_0}^{M_0} c_j\, d_{-j} + \sum_{j=-M_0}^{M_0}\sum_{p\in\mathbb{Z},p\neq 0} c_j\, d_{pL-j} \\
&= \int_0^1 \bar{\theta}'(t) e^{i2\pi k\phi(t)}\, dt + \sum_{j=-M_0}^{M_0}\sum_{p\in\mathbb{Z},p\neq 0} c_j\, d_{pL-j} \\
&= \sum_{j=-M_0}^{M_0}\sum_{p\in\mathbb{Z},p\neq 0} c_j\, d_{pL-j}.
\end{aligned} \tag{179}$$

Using integration by parts, we have

$$\begin{aligned} |d_l| &= \left| \int_0^1 e^{i2\pi k\phi} e^{-i2\pi lt}\, dt \right| \\ &= \frac{1}{|l|^n} \left| \int_0^1 \left(\frac{d^n}{dt^n} e^{i2\pi k\phi} \right) e^{-i2\pi lt}\, dt \right| \\ &\leqslant \frac{1}{|l|^n} \int_0^1 \left| \left(\frac{d^n}{dt^n} e^{i2\pi k\phi} \right) \right| dt \\ &\leqslant \frac{1}{|l|^n} \max_t \left| \left(\frac{d^n}{dt^n} e^{i2\pi k\phi} \right) \right|. \end{aligned} \tag{180}$$

Using the inequality (26) in the proof of Lemma 2.1, and by a direct calculation, we can show that for any $n > 0$, there exists $C(n) > 0$ such that

$$\begin{aligned} \max_t \left| \left(\frac{d^n}{dt^n} e^{i2\pi k\phi} \right) \right| &\leqslant C(n) \sum_{j=1}^{n} |k|^j M_0^{n-j} \big\| \widehat{\phi'} \big\|_1^j \\ &= C(n)|k| M_0^{n-1} \big\| \widehat{\phi'} \big\|_1 \frac{\frac{|k|^n}{M_0^n} \|\widehat{\phi'}\|_1^n - 1}{\frac{|k|}{M_0} \|\widehat{\phi'}\|_1 - 1} \\ &\leqslant \begin{cases} 2C(n)|k|^n \|\widehat{\phi'}\|_1^n, & \frac{|k|}{M_0} \|\widehat{\phi'}\|_1 > 2, \\ 2C(n)(2M_0)^n, & \frac{|k|}{M_0} \|\widehat{\phi'}\|_1 \leqslant 2. \end{cases} \end{aligned} \tag{181}$$

As a result, we obtain

$$|d_l| \leqslant \begin{cases} 2C(n) \left| \frac{k\|\widehat{\phi'}\|_1}{l} \right|^n, & |k| \|\widehat{\phi'}\|_1 > 2M_0, \\ 2C(n) \left| \frac{2M_0}{l} \right|^n, & |k| \|\widehat{\phi'}\|_1 \leqslant 2M_0. \end{cases} \tag{182}$$

Finally, we derive the following estimate

$$\begin{aligned} \left| \sum_{j=-M_0}^{M_0} \sum_{p\in\mathbb{Z}, p\neq 0} c_j d_{pL-j} \right| &\leqslant \sum_{p\in\mathbb{Z}, p\neq 0} \sum_{j=-M_0}^{M_0} |c_j| |d_{pL-j}| \\ &\leqslant 2 \sum_{j=-M_0}^{M_0} |c_j| \sum_{p=1}^{+\infty} \max_j |d_{pL-j}| \\ &\leqslant 4C(n) \big\| \widehat{\phi'} \big\|_1 \sum_{p=1}^{+\infty} \max\left(\left| \frac{k\|\widehat{\phi'}\|_1}{pL - M_0} \right|^n, \left| \frac{2M_0}{pL - M_0} \right|^n \right) \\ &\leqslant 4C(n) \big\| \widehat{\phi'} \big\|_1 \max\left(\left| \frac{k\|\widehat{\phi'}\|_1}{L} \right|^n, \left| \frac{2M_0}{L} \right|^n \right) \sum_{p=1}^{+\infty} (p - M_0/L)^{-n} \\ &\leqslant 4(1 - M_0/L)^{-n+1} \frac{C(n)}{n-1} \big\| \widehat{\phi'} \big\|_1 \max\left(\left| \frac{k\|\widehat{\phi'}\|_1}{L} \right|^n, \left| \frac{2M_0}{L} \right|^n \right). \end{aligned} \tag{183}$$

This completes the proof of the lemma. □

References

[1] B. Boashash, Time–Frequency Signal Analysis: Methods and Applications, Longman–Cheshire/John Wiley Halsted Press, Melbourne/New York, 1992.
[2] A.M. Bruckstein, D.L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Rev. 51 (2009) 34–81.
[3] E. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 52 (12) (2006) 5406–5425.
[4] E. Candès, T. Tao, Near optimal signal recovery from random projections: Universal encoding strategies?, IEEE Trans. Inform. Theory 52 (12) (2006) 5406–5425.
[5] E. Candes, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal recovery from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2006) 489–509.
[6] E. Candes, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math. 59 (2006) 1207–1223.
[7] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Conf. Ser. in Appl. Math., vol. 61, SIAM Publications, 1992.
[8] D.L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (2006) 1289–1306.
[9] P. Flandrin, Time–Frequency/Time–Scale Analysis, Academic Press, San Diego, CA, 1999.
[10] D. Gabor, Theory of communication, J. IEEE 93 (1946) 426–457.
[11] T.Y. Hou, Z. Shi, Data-drive time–frequency analysis, Appl. Comput. Harmon. Anal. 35 (2013) 284–308.
[12] N.E. Huang, et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 454 (1998) 903–995.
[13] D.L. Jomes, T.W. Parks, A high resolution data-adaptive time–frequency representation, IEEE Trans. Acoust. Speech Signal Process. 38 (1990) 2127–2135.
[14] P.J. Loughlin, B. Tracer, On the amplitude – and frequency-modulation decomposition of signals, J. Acoust. Soc. Amer. 100 (1996) 1594–1601.
[15] B.C. Lovell, R.C. Williamson, B. Boashash, The relationship between instantaneous frequency and time–frequency representations, IEEE Trans. Signal Process. 41 (1993) 1458–1461.
[16] S. Mallat, Z. Zhang, Matching pursuit with time–frequency dictionaries, IEEE Trans. Signal Process. 41 (1993) 3397–3415.
[17] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, Academic Press, 2009.
[18] W.K. Meville, Wave modulation and breakdown, J. Fluid Mech. 128 (1983) 489–506.
[19] S. Olhede, A.T. Walden, The Hilbert spectrum via wavelet projections, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 460 (2004) 955–975.
[20] B. Picinbono, On instantaneous amplitude and phase signals, IEEE Trans. Signal Process. 45 (1997) 552–560.
[21] S. Qian, D. Chen, Joint Time–Frequency Analysis: Methods and Applications, Prentice Hall, 1996.
[22] S.O. Rice, Mathematical analysis of random noise, Bell Syst. Tech. J. 23 (1944) 282–310.
[23] J. Shekel, Instantaneous frequency, Proc. IRE 41 (1953) 548.
[24] J. Tropp, A. Gilbert, Signal recovery from random measurements via Orthogonal Matching Pursuit, IEEE Trans. Inform. Theory 53 (2007) 4655–4666.
[25] B. Van der Pol, The fundamental principles of frequency modulation, Proc. IEEE 93 (1946) 153–158.
[26] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, Adv. Adapt. Data Anal. 1 (2009) 1–41.