On the Uniqueness of Sparse Time-Frequency Representation of Multiscale Data

Chunguang Liu * Zuoqiang Shi † Thomas Y. Hou ‡

January 21, 2015

Abstract

In this paper, we analyze the uniqueness of the sparse time frequency decomposition and investigate the efficiency of the nonlinear matching pursuit method. Under the assumption of scale separation, we show that the sparse time frequency decomposition is unique up to an error that is determined by the scale separation property of the signal. We further show that the unique decomposition can be obtained approximately by the sparse time frequency decomposition using nonlinear matching pursuit.

Keywords. sparse time frequency decomposition; scale separation; nonlinear matching pursuit.

1 Introduction

Due to the fast development of information science, data plays more and more important role in our life. Correspondingly, how to efficiently extract useful information from massive amount of data attracts lots of attentions in scientific research. Among various of data analysis methods, time frequency analysis is an important branch. The goal of time frequency analysis is to extract information of frequencies and corresponding amplitudes from the measurement of signals. In many applications, frequencies encode important information of the underlying physical mechanism.

In order to conduct time frequency analysis, many methods have been developed, for instance, windowed Fourier transform, wavelet transform [5, 15], Wigner-Ville distribution [11], etc. In recent years, an adaptive time frequency analysis method, the Empirical Mode Decomposition (EMD) method [10, 21] was developed. This method provides an efficient adaptive method to extract frequency information from nonlinear and nonstationary data. Since its inception, the EMD method has found many applications. Due to its empirical nature, the EMD method still lacks a rigorous mathematical foundation. Recently, a number of attempts have been made to provide a mathematical foundation for this method and several methods have been proposed, see e.g. the synchrosqueezed wavelet transform [6], the Empirical wavelet transform [8], the variational mode decomposition [12].

*Department of Mathematics, Jinan University, Guangzhou, China, 510632. Email: tcgliu@jnu.edu.cn.
†Mathematical Sciences Center, Tsinghua University, Beijing, China, 100084. Email: zqshi@math.tsinghua.edu.cn.
‡Applied and Comput. Math, MC 9-94, Caltech, Pasadena, CA 91125. Email: hou@cms.caltech.edu.
In the last few years, inspired by EMD method and compressive sensing \textsuperscript{[2, 3, 7]}, Hou and Shi proposed a novel time frequency analysis method based on the sparsest representation of multiscale data \textsuperscript{[18]}. In this method, the signal is decomposed to several components

\[ f(t) = \sum_{j=1}^{M} a_j(t) \cos \theta_j(t) + r(t), \quad t \in \mathbb{R}, \quad (1.1) \]

where \(a_j(t), \theta_j(t)\) are smooth functions, \(\theta_j'(t) > 0, j = 1, \ldots, M\) \(M\) is an integer that is given \textit{a priori}, and \(r(t)\) is a small residual. We assume that \(a_j(t)\) and \(\theta_j'\) are less oscillatory than \(\cos \theta_j(t)\).

We call \(a_j(t) \cos \theta_j(t)\) as the Intrinsic Mode Functions (IMFs) \textsuperscript{[10]}. After the decomposition is obtained, the instantaneous frequencies \(\omega_j(t)\) are defined as

\[ \omega_j(t) = \theta_j'(t), \quad (1.2) \]

and the amplitude is \(a_j(t)\).

One main difficulty in computing the decomposition (1.1) is that the decomposition is not unique. To pick up the “best” decomposition among all feasible ones, Hou and Shi proposed to decompose the signal by looking for the sparsest decomposition by solving the following nonlinear optimization problem:

\[
\begin{align*}
\text{Minimize} & \quad M \\
\text{Subject to:} & \quad f = \sum_{k=1}^{M} a_k \cos \theta_k, \quad a_k \cos \theta_k \in D,
\end{align*}
\]

where \(D\) is the dictionary consist of all IMFs (see \textsuperscript{[18]} for its precise definition).

The idea of looking for the sparsest representation over the time frequency dictionary has been exploited a lot in the signal processing community \textsuperscript{[4, 16]}. Comparing with existing methods, the novelty of this method is that the time frequency dictionary being used is much larger. The choice of this dictionary makes this method fully adaptive to the signal. On the other hand, one needs to solve a nonlinear optimization problem which is more difficult to solve.

From another point of view, the problem (1.3) can be seen as a dictionary learning problem \textsuperscript{[19]}. In (1.3), the basis elements to decompose the signal are determined by the phase functions \(\theta_k, k = 1, \ldots, M\), which are not known \textit{a priori}. Then both the optimal coefficients and the optimal basis need to be determined as part of the nonlinear optimization problem. The dictionary learning problem has been studied a lot in signal processing. Many efficient methods have been developed, see e.g. \textsuperscript{[1, 9, 13, 14, 17]}. The main difficulty in our problem is that there is usually only one signal rather than a collection of signals which form a training set in a typical dictionary learning problem. Because we have only one signal, we cannot use those methods in dictionary learning to solve our problem directly.

To approximately solve (1.3), an efficient algorithm based on matching pursuit and Fast Fourier transform has been proposed to solve the above nonlinear optimization problem. In a subsequent paper \textsuperscript{[20]}, the authors proved the convergence of their nonlinear matching pursuit algorithm for periodic data that satisfy certain scale separation property. It has been shown that this method is effective and efficient in many problems. But from the theoretical point of view, one important question remains open. That is the uniqueness of the solution of the optimization problem (1.3). The goal of this paper is to fill this gap.
In this paper, we will show that under the assumption of scale separation, the solution of optimization problem (1.3) is unique up to an error determined by the scale separation property. More specifically, we will prove the following theorem,

**Theorem 1.1.** Let \( f(t) \) be a function satisfying the scale-separation property with separation factor \( \epsilon \ll 1 \) and frequency ratio \( d \) as defined in Definition 3.1.

\[
f(t) = \sum_{k=1}^{M} a_k(t) \cos \theta_k(t) + r(t), \quad a_k \cos \theta_k \in U_\epsilon, \quad a_k = O(1), \quad r = O(\epsilon).
\]

Then \( (a_k, \theta_k)_{1 \leq k \leq M} \) is a solution of the optimization problem (1.3) and it is unique up to the error \( \epsilon \), i.e. if \( (\tilde{a}_k, \tilde{\theta}_k)_{1 \leq k \leq \tilde{M}} \) is a solution of (1.3), then \( \tilde{M} = M \) and

\[
|a_k(t) - \tilde{a}_k(t)| = O(\epsilon), \quad \left| \frac{\theta_k(t) - \tilde{\theta}_k(t)}{\theta'_k(t)} \right| = O(\epsilon), \quad \forall t, \quad k = 1, \ldots, M.
\]

This theorem is proved by careful studying the wavelet transform of each IMF. The details of the proof can be found in Section 3.

In addition to the proof of uniqueness, we also show that this solution can be given approximately by nonlinear matching pursuit.

**Theorem 1.2.** Let \( f(t) \) be a function satisfying the scale-separation property with separation factor \( \epsilon \) and frequency ratio \( d \) as defined in Definition 3.1.

\[
f(t) = \sum_{k=1}^{M} a_k(t) \cos \theta_k(t) + r(t), \quad a_k \cos \theta_k \in U_\epsilon, \quad a_k = O(1), \quad r = O(\epsilon).
\]

Suppose there exists \( \alpha \in [1, d) \) and \( l \in \{1, \ldots, M\} \) such that

\[
\alpha^{-1} \theta'_l(t) \leq \theta'(t) \leq \alpha \theta'_l(t), \quad \forall t \in [0, 1]. \tag{1.4}
\]

If

\[
p(a, \theta) \leq p(a_l, \theta_l), \tag{1.5}
\]

where \( p(a, \theta) := \|f(t) - a(t) \cos \theta(t)\|_{L^2}^2 \) is given in (P2) in Section 4, then

\[
\frac{\|a \cos \theta - a_l \cos \theta_l\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}} = O(\sqrt{\epsilon}) \tag{1.6}
\]

This theorem shows that matching pursuit is an efficient method to solve the optimization problem (1.3) if the signal satisfies the scale separation property.

We remark that there has been some recent progress in developing a mathematical framework for an EMD like method using synchrosqueezed wavelet transforms by Daubechies, Lu and Wu [6]. This method is based on continuous wavelet transform and does not decompose its IMFs directly. So the question of uniqueness of the synchrosqueezed wavelet transform is not the same as that we consider here in this paper.

The rest of the paper is organized as follows. In Section 2, we set up the framework of sparse time frequency analysis. The uniqueness is analyzed in Section 3. In Section 4, Theorem 1.2 is proved. Some concluding remarks are made in Section 5.
2 Sparse Decomposition over Time-Frequency Dictionary

A typical time-frequency analysis method consists of two parts: a large dictionary of time-frequency functions that are used to represent the signal and a decomposition method to decompose the signal over the dictionary.

A dictionary is a collection of functions that are regarded as available waves. Fourier Dictionaries, Time-Scale Dictionaries, Time-Frequency Dictionaries, and Collection of Intrinsic Mode Function are mostly used (see [18] and references therein).

In this section, we construct a dictionary of waves in the form of \(a(t)\cos \theta(t)\), where \(a\) is the envelope function and \(\theta(t)\) is the phase function. For the problem that we consider, the oscillation of the mode \(a(t)\cos \theta(t)\) is mainly determined by the value of phase \(\theta(t)\) itself, not the oscillation of the envelope \(a(t)\) nor that of the instantaneous frequency \(\theta'(t)\). In other words, we aim at finding a dictionary of the form:

\[
\mathcal{D} := \{a(\cdot)\cos \theta(\cdot) : a > 0, \theta' > 0; a, \theta' \text{ are much smoother than } \cos \theta(\cdot)\}.
\] (2.1)

To precisely describe the terminology “smoother”, we introduce the following definition about scale-separation.

**Definition 2.1** (scale-separation). One function \(f(t) = a(t)\cos \theta(t)\) is said to satisfy a scale-separation property with a separation factor \(\epsilon > 0\), if \(a(t)\) and \(\theta(t)\) satisfy the following conditions:

\[
a(t) \in C^1(\mathbb{R}), \quad \theta \in C^2(\mathbb{R}), \quad \inf_{\theta' \in \mathbb{R}} \theta'(t) > 0,
\]

\[
\frac{\max \theta'}{\min \theta'} < 2, \quad \left| \frac{a'(t)}{\theta'(t)} \right| \leq \epsilon, \quad \left| \frac{\theta''(t)}{(\theta'(t))^2} \right| \leq \epsilon, \quad \forall t \in \mathbb{R}.
\]

**Remark 2.1.** In the previous theoretical study of time-frequency analysis, usually, the condition \(\frac{\theta''}{\theta'} \leq \epsilon\) is assumed, see e.g. [6]. Here, we replace this assumption by \(\frac{\max \theta'}{\min \theta'} < 2\) and \(\left| \frac{\theta''(t)}{(\theta'(t))^2} \right| \leq \epsilon\).

In the problems we consider, \(\theta'\) is large usually. Then, the assumption we use here is much weaker than the typical assumption, \(\frac{\theta''}{\theta'} \leq \epsilon\).

Now we can give a mathematical definition of the dictionary we will use: for a positive number \(\epsilon \ll 1\), denoted by

\[
\mathcal{D}_\epsilon := \{a(\cdot)\cos \theta(\cdot) : (a, \theta) \in U_\epsilon\},
\] (2.2)

where

\[
U_\epsilon := \left\{(a, \theta) : a > 0, \theta' > 0; \left| \frac{a'}{\theta'} \right| \leq \epsilon, \left| \frac{\theta''}{\theta'} \right| \leq c_0, \left| \frac{\theta''}{(\theta')^2} \right| \leq \epsilon \right\}.
\] (2.3)

Based on this dictionary, we try to decompose the given signal \(f\) by looking for the sparsest decomposition over \(\mathcal{D}\), i.e. solving the following optimization problem:

Minimize \(M\)

subject to \(|f(t) - \sum_{k=1}^{M} a_k(t) \cos \theta_k(t)| \leq \epsilon_0, \quad (P_0)\)

\(a_k(t) \cos \theta_k(t) \in \mathcal{D}, \quad k = 1, 2, \cdots, M.\)
Here $\epsilon_0$ is a given threshold of the accuracy of the decomposition. Typically, $\epsilon_0$ is set according to the amplitude of noise. Borrowing the terminology of EMD method (see [10, 21]), we say that $a_k \cos \theta_k$ is an Intrinsic Mode Function (IMF).

To solve $\{P_0\}$, we use the following algorithm based on matching pursuit.

- $r_0 = f, \quad k = 1$.

Step 1: Solve the following nonlinear least-square problem:

$$
(a_k, \theta_k) \in \text{Argmin}_{a,\theta} \|r_{k-1} - a \cos \theta\|_2^2
$$

Subject to: $a, \theta \in U_{\epsilon}$. (2.4)

Step 2: Update the residual

$$
r_k = f - \sum_{j=1}^{k} a_j \cos \theta_j.
$$

(2.5)

Step 3: If $\|r_k\|_2 < \epsilon_0$, stop. Otherwise, set $k = k + 1$ and go to Step 1.

The performance of this algorithm will be analyzed in Section 4 under the assumption of scale separation.

### 3 Uniqueness of $\{P_0\}$ for well-separated signal

In this section, we assume the signal function, $f(t)$, is periodic over $[0, 1)$ and is smooth enough. For this kind of signals, the existence of the solution of $\{P_0\}$ is easy to see. Since $f$ is periodic and smooth, it can be well approximated by truncated Fourier series, i.e. there exists $M > 0$, such that

$$
|f(t) - \sum_{k=0}^{M} c_k \cos(2\pi kt + \phi_k)| \leq \epsilon_0, \quad \forall t \in [0, 1],
$$

(3.1)

where $c_k, \phi_k$ are constants. Obviously, $c_k \cos(2\pi kt + \phi_k) \in U_{\epsilon}$, then the truncated Fourier series gives a feasible decomposition. Each feasible decomposition gives a positive integer, by collecting all these positive integers together, we get a set $A$. Then we know that $A$ is nonempty and has a lower bound. Let $M_0 = \inf A$. Since $A$ consists of positive integers, then $M_0 = \inf A$ can be achieved. Then the existence of the solution of $\{P_0\}$ is proved.

But the uniqueness is much more complicated. First of all, we can construct a signal $f$, such that the solution of $\{P_0\}$ is not unique. This fact suggests that to get uniqueness, we have to enforce some extra conditions on the signal $f$. In this section, we consider the well-separated signal which is defined as follows,

**Definition 3.1** (Well-separated signal). A periodic signal $f : [0, 1] \rightarrow \mathbb{R}$ is said to be well-separated with separation factor $\epsilon$ and frequency ratio $d$ if it can be written as

$$
f(t) = \sum_{k=1}^{M} a_k(t) \cos \theta_k(t) + r(t)
$$

(3.2)

where all $f_k(t) = a_k(t) \cos \theta_k(t)$ satisfies the scale-separation property with separation factor $\epsilon$, $r(t) = O(\epsilon_0)$ and their phase function $\theta_k$ satisfies

$$
\theta_k'(t) \geq d \theta_{k-1}'(t), \quad \forall t \in [0, 1].
$$

(3.3)
and \( d > 1, \ d - 1 = O(1) \).

We can prove that if \( f \) is well-separated signal with separation factor \( \epsilon \) and frequency ratio \( d > 1 \) and \( \epsilon \ll 1 \), then the solution of \( \{P_0\} \) is unique up to \( \epsilon \) and \( \epsilon_0 \).

To simplify the notation, in the rest of this paper, we assume \( \epsilon_0 \) and \( \epsilon \) have the same order and denote both of them by \( \epsilon \).

### 3.1 Main Result on the Uniqueness

The main theorem is stated as follows:

**Theorem 3.1.** Let \( f(t) \) be a function satisfying the scale-separation property with separation factor \( \epsilon \) and frequency ratio \( d \) as defined in Definition 3.1:

\[
 f(t) = \sum_{k=1}^{M} a_k(t) \cos \theta_k(t) + r(t), \quad a_k \cos \theta_k \in U_\epsilon, \quad a_k = O(1), \quad r = O(\epsilon).
\]

Then \( (a_k, \theta_k)_{1 \leq k \leq M} \) is a solution of the optimization problem \( \{P_0\} \) and it is unique up to the error \( \epsilon \), i.e. if \( (\tilde{a}_k, \tilde{\theta}_k)_{1 \leq k \leq \tilde{M}} \) is a solution of \( \{P_0\} \), then \( \tilde{M} = M \) and

\[
 |a_k(t) - \tilde{a}_k(t)| = O(\epsilon), \quad \frac{\left|\theta_k(t) - \tilde{\theta}_k(t)\right|}{\theta'_k(t)} = O(\epsilon), \quad \forall t, \ k = 1, \cdots, M. \tag{3.4}
\]

Using this theorem, we can see that the solution is almost unique. The ambiguity is very small if the signal satisfies the scale separation.

The theorem is proved by carefully studying the wavelet transform of each IMF. We can show in Lemma 3.1 that by properly choosing an appropriate wavelet function, the continuous wavelet transform (CWT) of each IMF is confined in a narrow band. Then the uniqueness can be obtained by comparing the continuous wavelet transform of different decompositions.

In order to complete the proof, first, we need to estimate the width of CWT of each IMF. This estimation is given in the following lemma.

**Lemma 3.1.** Let \( \psi \) be a wavelet function such that its Fourier Transform \( \hat{\psi} \in C^4 \) and has compact support. Suppose \( (a, \theta) \in U_\epsilon \), \( a, \theta' \) are periodic over \([0, 1]\), and

\[
 \frac{\max \theta'}{\min \theta'} < 2.
\]

Then we have

\[
 W(a \text{e}^{-i\theta})(t, \omega) = \frac{1}{\sqrt{\omega}} \int_{\mathbb{R}} a(\tau) \text{e}^{-i\theta(\tau)} \psi\left(\frac{\tau - t}{\omega}\right) d\tau = \sqrt{\omega} a(t) \text{e}^{-i\theta(t)} \hat{\psi}(\omega \theta'(t)) + C \sqrt{\omega} (I_1 + I_2 + I_3)\epsilon
\]

where \( C > 0 \) is a constant and

\[
 I_1 = \int_{\mathbb{R}} |\psi(\tau)| d\tau, \quad I_2 = \int_{\mathbb{R}} |\tau \psi'(\tau)| d\tau, \quad I_3 = \int_{\mathbb{R}} |\tau^2 \psi''(\tau)| d\tau.
\]

The proof of this lemma can be found in Appendix A. And now, we are ready to give the proof of Theorem 3.1.

**Proof.** of Theorem 3.1
The details of the proof may be a bit tedious but the idea is very clear. First, we assume there are two decompositions:

\[ f(t) = \sum_{k=1}^{M} a_k(t) \cos \theta_k(t) + O(\epsilon) = \sum_{k=1}^{M} \tilde{a}_k(t) \cos \tilde{\theta}_k(t) + O(\epsilon). \]  

(3.5)

Using Lemma 3.1 we have

\[ \omega^{-1/2} \mathcal{W}(f)(t, \omega) = \sum_{k=1}^{M} a_k(t) e^{-i\theta_k(t)} \tilde{\psi}(\omega \theta'_k(t)) + O(\epsilon) = \sum_{k=1}^{\tilde{M}} \tilde{a}_k(t) e^{-i\tilde{\theta}_k(t)} \tilde{\psi}(\omega \tilde{\theta}'_k(t)) + O(\epsilon). \]  

(3.6)

Using these equalities, we can show that (3.4) is true.

First, we pick up a specific wavelet function \( \psi \). We require that its Fourier Transform \( \hat{\psi} \in C^4 \) has support in \([1 - \Delta, 1 + \Delta]\) with \( 0 < \Delta < \frac{K}{\sqrt{d+1}} \) and \( \hat{\psi}(1) = 1 \) is the maximum of \( |\hat{\psi}| \). In this proof, we choose \( \hat{\psi} \) to be a fifth order B-spline function after proper scaling and translation. For any \( \theta \in C^1 \), define

\[ U_{\theta} = \left\{(t, \omega) \in \mathbb{R}^2 : |\hat{\psi}(\omega \theta'(t))| > \epsilon \right\}, \quad U_{\theta}(t) = \left\{\omega \in \mathbb{R} : |\hat{\psi}(\omega \theta'(t))| > \epsilon \right\}. \]

Now, fix a time \( t = t_0 \in [0, 1] \), for any \( l \in \{1, \cdots, M\} \), let \( \omega_{l,0} = \frac{1}{\theta'_l(t_0)} \). Using (3.6), we have

\[ \omega^{-1/2} \mathcal{W}(f)(t_0, \omega_{l,0}) = \sum_{k=1}^{M} a_k(t_0) e^{-i\theta_k(t_0)} \hat{\psi}(\theta'_k(t_0)/\theta'_l(t_0)) + O(\epsilon) \]

\[ = \sum_{k=1}^{\tilde{M}} \tilde{a}_k(t_0) e^{-i\tilde{\theta}_k(t_0)} \hat{\psi}(\tilde{\theta}'_k(t_0)/\theta'_l(t_0)) + O(\epsilon). \]  

(3.8)

Since \( \theta'_k(t_0)/\theta'_l(t_0) \leq 1/d < 1 - \Delta \) or \( \theta'_k(t_0)/\theta'_l(t_0) \geq d > 1 + \Delta \) for any \( k \neq l \), in the first summation of above equation, we have only one term left,

\[ a_l(t_0) e^{-i\theta_l(t_0)} \hat{\psi}(1) = \sum_{k=1}^{\tilde{M}} \tilde{a}_k(t_0) e^{-i\tilde{\theta}_k(t_0)} \hat{\psi}(\tilde{\theta}'_k(t_0)/\theta'_l(t_0)) + O(\epsilon). \]  

(3.9)

Then, there exists at least a \( I(l,t_0) \in \{1, \cdots, \tilde{M}\} \), such that

\[ \left| \hat{\psi}(\tilde{\theta}'_{I(l,t_0)}(t_0)/\theta'_l(t_0)) \right| > 0, \]

(3.10)

which means that \( 1 - \Delta < \tilde{\theta}'_{I(l,t_0)}(t_0)/\theta'_l(t_0) < 1 + \Delta \). Using the assumption that \( f \) is well-separated with frequency ratio \( d \) and \( 0 < \Delta < \frac{K}{\sqrt{d+1}} \), for any \( k \neq l \), we obtain

\[ \frac{\theta'_k(t_0)}{\theta'_l(t_0)} \geq d, \quad \text{or} \quad \frac{\theta'_l(t_0)}{\theta'_l(t_0)} \leq \frac{1}{d}. \]  

(3.11)

This gives that

\[ \frac{\tilde{\theta}'_{I(l,t_0)}(t_0)}{\theta'_l(t_0)} = \frac{\tilde{\theta}'_{I(l,t_0)}(t_0)}{\theta'_l(t_0)} \cdot \frac{\theta'_l(t_0)}{\theta'_l(t_0)} \geq d(1 - \Delta) > 1 + \Delta, \]

(3.12)
On the other hand, since \( f_\theta \) without loss of generality, we assume that

\[
\text{Now, let } \eta, \text{ constant over } [0,1], \text{ and for any } t \in [0,1], \text{ i.e. }
\]

\[
\text{Denote } I^{-1}(k,t) = I^{-1}(k,0), \forall t \in [0,1], k = 1, \cdots, M.
\]

(3.17)

Otherwise, suppose there exists \( t_0 \in [0,1] \), such that \( I^{-1}(k,t_0) \neq I^{-1}(k,0) \). Let \( A = \{0 \leq t \leq t_0 : I^{-1}(k,t) = I^{-1}(k,0)\} \) and \( \xi = \sup A \). Then for any \( \eta > 0 \), there exist \( t_1, t_2 \in [0,1] \), such that

\[
t_1 < \xi < t_2, \quad |t_2 - t_1| < \eta, \quad I^{-1}(k,0) = I^{-1}(k,t_1) \neq I^{-1}(k,t_2).
\]

(3.18)

Then, by the definition of the map \( I(k,t) \), we have

\[
1 - \Delta < \frac{\tilde{\theta}'_k(t_1)}{\theta'_k(t_1)} < 1 + \Delta, \quad 1 - \Delta < \frac{\tilde{\theta}'_k(t_2)}{\theta'_k(t_2)} < 1 + \Delta.
\]

Without loss of generality, we assume that \( \theta'_{k_2} > \theta'_{k_1} \). Then, we have

\[
\tilde{\theta}'_k(t_1) < (1 + \Delta)\theta'_k(t_1), \quad \tilde{\theta}'_k(t_2) > (1 - \Delta)\theta'_k(t_1).
\]

Now, let \( \eta \to 0 \), have \( t_1, t_2 \to \xi \), which gives

\[
\tilde{\theta}'_k(\xi) \leq (1 + \Delta)\theta'_k(\xi), \quad \tilde{\theta}'_k(\xi) \geq (1 - \Delta)\theta'_k(\xi).
\]

On the other hand, since \( f \) is well-separated, we know that

\[
\frac{\theta'_k(t)}{\theta'_{k_2}(t)} \leq 1/d.
\]
Then, we have

\[ \tilde{\theta}_k^i(\xi) \leq (1 + \Delta)\theta_{k_1}^i(\xi), \quad \tilde{\theta}_k^i(\xi) \geq d(1 - \Delta)\theta_{k_1}^i(\xi) > (1 + \Delta)\theta_{k_2}^i(\xi), \]

which is a contradiction. This means that \( I^{-1}(k, \cdot) \) is a constant over \([0, 1]\) and we can assume

\[ I^{-1}(k, t) = k, \quad \forall t \in [0, 1], \ k = 1, \ldots, M. \quad (3.19) \]

Now we have that

\[ 1 - \Delta < \frac{\tilde{\theta}_k(t)}{\theta_k^i(t)} < 1 + \Delta, \quad \forall t \in [0, 1], \ k = 1, \ldots, M. \quad (3.20) \]

Using the assumption that the signal \( f \) is well-separated with ratio \( d \) and the choice of \( \psi \), we have that for any \( k, \ l = 1, \ldots, M, \ k \neq l \)

\[ \hat{\psi}(\omega \theta_k^i(t)) = \hat{\psi}(\omega \theta_l^i(t)) = 0, \quad \forall (t, \omega) \in U_{\theta_k}, \quad (3.21) \]

\[ \hat{\psi}(\omega \theta_k^i(t)) = \hat{\psi}(\omega \theta_l^i(t)) = 0, \quad \forall (t, \omega) \in U_{\hat{\theta}_k}. \quad (3.22) \]

On the other hand, using Lemma 5.1 we have

\[ \omega^{-1/2}W(f)(t, \omega) = \sum_{k=1}^{M} a_k(t)e^{-i\theta_k(t)}\hat{\psi}(\omega \theta_k^i(t)) + O(\epsilon) = \sum_{k=1}^{M} \tilde{a}_k(t)e^{-i\tilde{\theta}_k(t)}\hat{\psi}(\omega \tilde{\theta}_k(t)) + O(\epsilon). \quad (3.23) \]

Using the three relations \( (3.21), (3.22) \), and \( (3.23) \), we have

\[ \left| a_k(t)e^{-i\theta_k(t)}\hat{\psi}(\omega \theta_k^i(t)) - \tilde{a}_k(t)e^{-i\tilde{\theta}_k(t)}\hat{\psi}(\omega \tilde{\theta}_k(t)) \right| = O(\epsilon), \quad \forall (t, \omega) \in U_{\theta_k} \cup U_{\hat{\theta}_k}, \quad (3.24) \]

which implies that

\[ \left| a_k(t)\hat{\psi}(\omega \theta_k^i(t)) - \tilde{a}_k(t)\hat{\psi}(\omega \tilde{\theta}_k(t)) \right| = O(\epsilon), \quad \forall (t, \omega) \in U_{\theta_k} \cup U_{\hat{\theta}_k}. \quad (3.25) \]

Next, we will prove that \( |a_k(t) - \tilde{a}_k(t)| = O(\epsilon) \) and \( |\theta_k^i(t) - \tilde{\theta}_k(t)| = O(\epsilon) \) from \( (3.25) \). First, we consider the envelopes \( a_k \) and \( \tilde{a}_k \).

If \( a_k(t) > \tilde{a}_k(t) \), we choose \( \omega = 1/\theta_k^i(t) \) to get

\[ a_k(t) \left| \hat{\psi}(1) - \tilde{a}_k(t) \left| \hat{\psi}(\tilde{\theta}_k^i(t)/\theta_k^i(t)) \right| \right| = O(\epsilon). \quad (3.26) \]

Since \( a_k(t) > \tilde{a}_k(t) \), we have

\[ 0 \leq \left| \hat{\psi}(1) (a_k(t) - \tilde{a}_k(t)) \right| \leq a_k(t) \left| \hat{\psi}(1) - \tilde{a}_k(t) \left| \hat{\psi}(\tilde{\theta}_k^i(t)/\theta_k^i(t)) \right| \right| = O(\epsilon). \quad (3.27) \]

This proves that

\[ a_k(t) - \tilde{a}_k(t) = O(\epsilon) \quad (3.28) \]

If \( a_k(t) < \tilde{a}_k(t) \), we take \( \omega = 1/\theta_k^i(t) \). By following a similar argument, we can prove

\[ \tilde{a}_k(t) - a_k(t) = O(\epsilon). \quad (3.29) \]

Combining these two cases, we obtain

\[ |a_k(t) - \tilde{a}_k(t)| = O(\epsilon). \quad (3.30) \]
Substituting the above relation to (3.25), we get
\[ |\hat{\psi}(\omega\theta_k'(t))| - |\hat{\psi}(\omega\theta_k'(t))| = O(\epsilon), \quad \forall(t, \omega) \in U_{\theta_k} \bigcup U_{\theta_k}. \] (3.31)

For any \( t \in [0, 1] \), let \( \omega = (1 - \Delta/2)/\theta_k'(t) \), then we have
\[ |\hat{\psi}(1 - \Delta/2) - |\hat{\psi}(1 - \Delta/2)\theta_k'(t)/\theta_k'(t)| = O(\epsilon). \] (3.32)

Since \( \hat{\psi}(\omega) \) is a fifth order B-Spline function and \( \epsilon \ll 1 \), it is easy to see that there exists a constant \( C > 0 \), such that
\[ \left| (1 - \Delta/2) \left( \theta_k'(t)/\theta_k'(t) - 1 \right) \right| \leq C\epsilon. \] (3.33)

Since \( \hat{\psi}(\omega) \) is a fifth order B-Spline function, \( \hat{\psi}(\omega) \) is symmetric with respect to \( \omega = 1 \). Thus we have \( \hat{\psi}(1 - \Delta/2) = \hat{\psi}(1 + \Delta/2) \). Then, there is also another possibility:
\[ \left| (1 - \Delta/2)\theta_k'(t)/\theta_k'(t) - (1 + \Delta/2) \right| \leq C\epsilon. \] (3.34)

If (3.33) holds, then we get
\[ \frac{|\theta_k'(t) - \hat{\theta}_k'(t)|}{\theta_k'(t)} = O(\epsilon). \] (3.35)

If (3.34) holds, then we have
\[ \frac{\hat{\theta}_k'(t)}{\theta_k'(t)} \geq \frac{1 + \Delta/2 - C\epsilon}{1 - \Delta/2} \geq 1 + \Delta, \] (3.36)

where we have used the assumption that \( \epsilon \ll 1 \).

Then, let \( \omega = 1/\theta_k'(t) \) in (3.31), we have
\[ 1 = \left| \hat{\psi}(1) - |\hat{\psi}(\theta_k'(t)/\theta_k'(t))| \right| = O(\epsilon). \] (3.37)

This argument shows that (3.34) cannot be true. This completes the proof. \( \square \)

### 3.2 Discussion on Signals with close frequencies

In this section, we will give a brief discussion on the signal with close frequencies, i.e. the frequency ratio \( d \rightarrow 1 \) in the definition of well-separated signal.

In the proof of Theorem 3.1, the frequency ratio \( d \) seems to be arbitrary, as long as it is larger than 1. Actually, the distance between \( d \) and 1 can not be too small. The gap is determined by the separation factor \( \epsilon \). First, it is easy to get that the integral \( I_1, I_2, I_3 \) in Lemma 3.1 satisfy
\[ I_1 = O(\Delta^{-1}), \quad I_2 = O(\Delta^{-2}), \quad I_3 = O(\Delta^{-3}). \] (3.38)

When \( d \) approach 1, \( \Delta = \frac{d-1}{\sqrt{d+1}} \) becomes smaller, then \( I_1, I_2, I_3 \) become larger. When \( I_1, I_2, I_3 \) is as large as \( 1/\epsilon \), Lemma 3.1 breaks down, which in turn leads to the break down of the proof of Theorem 3.1.

On the other hand, the frequency ratio is defined pointwisely. As long as there exists a point, such that the frequency ratio at this point is away from 1 comparing with \( \epsilon \), the argument before 3.10 of the proof of Theorem 3.1 still applies, which means that the decomposition given in 3.2 is a solution of the optimization problem (3.6). But in this case, the uniqueness is not
guaranteed. As mentioned before, we can construct an example such that the solution of \((P_0)\) is not unique.

\[
f(t) = \cos \theta_1(t) + \cos \theta_2(t), \quad t \in [0, 1],
\]

\[
\theta_1(t) = 6\pi kt + k\pi, \quad \theta_2(t) = 8\pi kt + k\sin 2\pi t,
\]

where \(k\) is a positive integer. We can choose \(k\) large enough, such that \(\cos \theta_1, \cos \theta_2 \in U_\epsilon\). At \(t = 0, \theta'_2(0)/\theta'_1(0) = 4/3\). From the discussion above, we know that \((3.39)\) gives a solution of \((P_0)\). On the other hand, it is easy to check that the following decomposition is also a solution of \((P_0)\).

\[
f(t) = \cos \phi_1(t) + \cos \phi_2(t), \quad t \in [0, 1],
\]

\[
\phi_1(t) = \begin{cases} 
6\pi kt + k\pi, & t \in [0, 1/2], \\
8\pi kt + k\sin 2\pi t, & t \in (1/2, 1]. 
\end{cases}
\]

\[
\phi_2(t) = \begin{cases} 
8\pi kt + k\sin 2\pi t, & t \in [0, 1/2], \\
6\pi kt + k\pi, & t \in (1/2, 1]. 
\end{cases}
\]

This example shows that when the frequencies intersect, the solution of \((P_0)\) may not be unique. In this case, we need to impose some extra constraints to obtain uniqueness of the decomposition.

One natural idea is to pick up the solution according to the regularity of \(a(t)\) and \(\theta'(t)\). The solution with regular amplitude and frequency is favorable. One method based on this idea is proposed in [19] to decompose signals that do not have well-separated IMFs. In that method, the regularity is related with the sparsity over Fourier (or Wavelet) dictionary. In the above example, this method would prefer the decomposition \((3.39)\), since in this decomposition, both of the amplitude and the frequencies are very sparse over the Fourier dictionary.

4 Optimal Solutions to \(P_2\) Problem

Using the result in the previous section, we know that for signals which are well-separated, the solution of \((P_0)\) is unique up to the separation factor \(\epsilon\). In this section, we will show that the algorithm based on matching pursuit we proposed in Section 2 could find this unique solution up to the separation factor \(\epsilon\).

In the algorithm based on matching pursuit, each IMF is given by solving the following nonlinear least-square problem:

Minimize \(p(a, \theta) := \|f(t) - a(t) \cos \theta(t)\|_{L^2}^2\)

subject to \((a, \theta) \in U_\epsilon\).

\((P_2)\)

What we are most interested in is under what conditions each IMF \(a_k(t) \cos \theta_k(t)\) of \(f(t)\) could be a local (approximate) optimizer of the above problem. The answer is summarized in the following theorem:

**Theorem 4.1.** Let \(f(t)\) be a function satisfying the scale-separation property with separation factor \(\epsilon\) and frequency ratio \(d\) as defined in Definition 3.1.

\[
f(t) = \sum_{k=1}^{M} a_k(t) \cos \theta_k(t) + r(t), \quad a_k \cos \theta_k \in U_\epsilon, \quad a_k = O(1), \quad r = O(\epsilon).
\]
Suppose there exists $\alpha \in [1, d)$ and $l \in \{1, \cdots, M\}$ such that

$$\alpha^{-1} \theta_l'(t) \leq \theta_l'(t) \leq \alpha \theta_l'(t), \quad \forall t \in [0, 1]. \quad (4.1)$$

If

$$p(a, \theta) \leq p(a_l, \theta_l), \quad (4.2)$$

where $p(a, \theta)$ is given in \cite{[12]}, then we have

$$\frac{\|a \cos \theta - a_l \cos \theta_l\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}} = O(\sqrt{\epsilon}). \quad (4.3)$$

Proof. First we know

$$0 \geq p(a, \theta) - p(a_l, \theta_l)$$

$$= \|f(t) - a(t) \cos \theta(t)\|_{L^2}^2 - \|f(t) - a_l(t) \cos \theta_l(t)\|_{L^2}^2$$

$$= \left\| \sum_{k \neq l} a_k \cos \theta_k + r(t) + a_l(t) \cos \theta_l(t) - a(t) \cos \theta(t) \right\|_{L^2}^2 - \left\| \sum_{k \neq l} a_k \cos \theta_k + r(t) \right\|_{L^2}^2 \quad (4.4)$$

$$= \|a_l \cos \theta_l - a \cos \theta\|_{L^2}^2 + 2 \langle a_l \cos \theta_l - a \cos \theta, \sum_{k \neq l} a_k \cos \theta_k + r(t) \rangle,$$

where the first equality follows from the definition of $p(a, \theta)$ in \cite{[12]}. In the rest of the proof, we try to control the second term of the above inequality.

It is easy to verify that

$$\left| \langle a_l \cos \theta_l, \sum_{k \neq l} a_k \cos \theta_k \rangle \right| \leq \sum_{k \neq l} \mu_{k,l} \|a_l \cos \theta_l\|_{L^2} \|a_k \cos \theta_k\|_{L^2}$$

$$= \delta_1 \|a_l \cos \theta_l\|^2, \quad (4.5)$$

where

$$\mu_{k,l} = \frac{|\langle a_l \cos \theta_l, a_k \cos \theta_k \rangle|}{\|a_l \cos \theta_l\|_{L^2} \|a_k \cos \theta_k\|_{L^2}}, \quad \delta_1 = \sum_{k \neq l} \mu_{k,l} \frac{\|a_k \cos \theta_k\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}}. \quad (4.6)$$

Similarly, we have

$$\left| \langle a \cos \theta, \sum_{k \neq l} a_k \cos \theta_k + r(t) \rangle \right|$$

$$\leq \delta_2 \|a \cos \theta\|_{L^2} \|a_l \cos \theta_l\|_{L^2}$$

$$\leq \delta_2 \|a \cos \theta - a_l \cos \theta_l\|_{L^2} \cdot \|a_l \cos \theta_l\|_{L^2} + \delta_2 \|a_l \cos \theta_l\|^2,$$

with

$$\delta_2 = \sum_{k \neq l} \mu_{k,l,\alpha} \frac{\|a_k \cos \theta_k\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}} + \frac{\|r(t)\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}}, \quad (4.7)$$

$$\mu_{k,l,\alpha} = \frac{|\langle a \cos \theta, a_k \cos \theta_k \rangle|}{\|a \cos \theta\|_{L^2} \|a_k \cos \theta_k\|_{L^2}}. \quad (4.8)$$
Thus it follows from (4.4), (4.5) and (4.7) that
\[ 0 > \| a_1 \cos \theta_l - a \cos \theta \|^2_{L^2} - 2 \delta_2 \| a \cos \theta - a_1 \cos \theta_l \|_{L^2} \cdot \| a_1 \cos \theta_l \|_{L^2} - 2(\delta_1 + \delta_2) \| a_1 \cos \theta_l \|^2_{L^2}, \]
(4.9)
which implies
\[ \frac{\| a \cos \theta - a_1 \cos \theta_l \|_{L^2}}{\| a_1 \cos \theta_l \|_{L^2}} = \delta_2 + \sqrt{\delta_2^2 + 2(\delta_1 + \delta_2)}. \]
(4.10)
Here, \( \mu_{k,l} \) and \( \mu_{k,l,\alpha} \) are just the coherences between \( a_1 \cos \theta_l, a \cos \theta \) and \( a_k \cos \theta_k \). Under the assumption of scale separation and the assumption that different IMF's are well separated, the behavior of \( a_1 \cos \theta_l, a \cos \theta \) and \( a_k \cos \theta_k \) are close to that of the standard Fourier basis. Then, it is natural to expect that the coherences, \( \mu_{k,l} \) and \( \mu_{k,l,\alpha} \), are small. Actually, the smallness of \( \mu_{k,l} \) and \( \mu_{k,l,\alpha} \) is given in Corollary 4.1. In particular, the estimate (4.18) from Corollary 4.1 shows that for all \( k \neq l \) we have
\[ \mu_{k,l} = O(\epsilon), \quad \mu_{k,l,\alpha} = O(\epsilon). \]
(4.11)
Together with the assumption that \( r(t) = O(\epsilon) \), we get
\[ \delta_1 = O(\epsilon), \quad \delta_2 = O(\epsilon). \]
(4.12)
It follows from (4.10) that
\[ \frac{\| a \cos \theta - a_1 \cos \theta_l \|_{L^2}}{\| a_1 \cos \theta_l \|_{L^2}} = O(\sqrt{\epsilon}). \]
(4.13)
This completes the proof.

In the proof of the above theorem, we have used the estimate (4.11) for \( \mu_{k,l} \) and \( \mu_{k,l,\alpha} \). This estimate can be derived by the following lemma:

**Lemma 4.1.** Let \( (a, \theta) \in U_\epsilon \) be such that \( a \cos \theta \) has period 1. Then we have
\[ \| a(t) \cos \theta(t) \|^2_{L^2} = \left( \frac{1}{2} + O(\epsilon) \right) \| a(t) \|_{L^2}^2. \]
(4.14)
Furthermore, if there is another pair \( (\bar{a}, \bar{\theta}) \in U_\epsilon \) being periodic over \([0, 1]\) such that
\[ \beta := \min_{t \in [0,1]} \sqrt{\theta'(t)} > 1, \]
(4.15)
then we have
\[ \langle a \cos \theta, \bar{a} \cos \bar{\theta} \rangle < C \epsilon \left( 1 + \frac{1}{(1 - \beta^{-1})^2} \right) \int_0^1 a(t) \bar{a}(t) dt, \]
(4.16)
where \( C > 0 \) is a constant.

The proof of this lemma can be found in Appendix B. Then (4.11) is just a direct corollary of the above lemma.
Corollary 4.1. Let \((a_k, \theta_k), \ k = 1, \cdots, M\) are well-separated with frequency ratio \(d\) and separation factor \(\epsilon\) as defined in Definition 3.1. Let \((a, \theta) \in U_\epsilon\) and there exists \(\alpha \in [1, d)\) and \(l \in \{1, \cdots, M\}\) such that
\[
\alpha^{-1} \theta_k'(t) \leq \theta'(t) \leq \alpha \theta'_l(t), \quad \forall t \in [0, 1].
\] (4.17)

Then for any \(1 \leq k \neq l \leq M\), we have
\[
\mu_{k,l} := \frac{|\langle a_k \cos \theta_k, a_l \cos \theta_l \rangle|}{\|a_l \cos \theta_l\|_{L^2} \|a_k \cos \theta_k\|_{L^2}} = O(\epsilon),
\]
\[
\mu_{k,l,\alpha} := \frac{|\langle a_k \cos \theta_k, a \cos \theta \rangle|}{\|a \cos \theta\|_{L^2} \|a_k \cos \theta_k\|_{L^2}} = O(\epsilon). (4.18)
\]

Proof. For any \(1 \leq k \neq l \leq M\), suppose \(\theta'_k > \theta'_l\). Since \((a_k, \theta_k), \ k = 1, \cdots, M\) are well-separated, we have
\[
\min_{t \in [0,1]} \frac{\theta'_k}{\theta'_l} > d^{k-l}. (4.19)
\]

Using Lemma 4.1 it is easy to check that
\[
\mu_{k,l} = \frac{|\langle a_k \cos \theta_k, a_l \cos \theta_l \rangle|}{\langle a_k, a_l \rangle} \cdot \frac{\langle a_k, a_l \rangle}{\|a_l \cos \theta_l\|_{L^2} \|a_k \cos \theta_k\|_{L^2}} \leq \|a_k \cos \theta_k\|_{L^2} \|a_l \cos \theta_l\|_{L^2} \leq \epsilon \left(\frac{1}{2} - \frac{3}{4} \epsilon\right)^{-1} \left(1 + \frac{1}{1 - d^{-|l-k|}}\right).
\]

For any \(1 \leq k \neq l \leq M\), since \((a_k, \theta_k), \ k = 1, \cdots, M\) are well-separated and
\[
\alpha^{-1} \theta_k'(t) \leq \theta'(t) \leq \alpha \theta'_l(t), \quad \forall t \in [0, 1]. (4.20)
\]
if \(\theta'_k < \theta'_l\), we know that
\[
\frac{\theta'_{k}}{\theta'_{k}} \geq \alpha^{-1} d^{-k}. \quad (4.21)
\]
And if \(\theta'_k > \theta'_l\), we have
\[
\frac{\theta'_{k}}{\theta'_{l}} \geq \alpha^{-1} d^{k-l}. \quad (4.22)
\]
Then It follows from Lemma 4.1 that, for \(k \neq l\),
\[
\mu_{k,l,\alpha} \leq \frac{|\langle a_k \cos \theta_k, a \cos \theta \rangle|}{\langle a_k, a \rangle} \cdot \frac{\|a_k\|_{L^2}}{\|a_k \cos \theta_k\|_{L^2}} \cdot \frac{\|a\|_{L^2}}{\|a \cos \theta\|_{L^2}} \leq \epsilon \left(\frac{1}{2} - \frac{3}{4} \epsilon\right)^{-1} \left(1 + \frac{1}{1 - \alpha d^{-|l-k|}}\right).
\]

\[
\square
\]

Remark 4.1. We remark that in Corollary 4.1 and Theorem 4.1 the condition (4.1) could be replaced by
\[
\frac{\theta'(t)}{\theta'_{l-1}(t)} \geq \frac{d}{\alpha}, \quad \frac{\theta'_{l+1}(t)}{\theta'_l(t)} \geq \frac{d}{\alpha}, \quad \forall t \in [0, 1]. (4.23)
\]
If we put a stronger assumption on the separation ratio \( d \) such that \( d > 4 \), then we obtain the following theorem about the global minimizer of the \( P_2 \) problem.

**Theorem 4.2.** Let \( f(t) \) be a function satisfying the scale-separation property with separation factor \( \epsilon \) and frequency ratio \( d \) as defined in Definition 3.1:

\[
f(t) = \sum_{k=1}^{M} a_k(t) \cos \theta_k(t) + r(t), \quad a_k \cos \theta_k \in U, \quad a_k = O(1), \quad r = O(\epsilon).
\]

Suppose further that \( d > 4 \), and that there exists \( l \in [1,K] \) such that

\[
\|a_k \cos \theta_k\| < \|a_l \cos \theta_l\|, \quad \forall k \in [1,K], \quad k \neq l.
\]  

(4.24)

If

\[
p(a, \theta) \leq p(a_l, \theta_l),
\]

where \( p(a, \theta) \) is given in (4.25), then we have

\[
\frac{\|a \cos \theta - a_l \cos \theta_l\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}} = O(\sqrt{\epsilon})
\]

(4.26)

**Proof.** First, we claim that, for each \( k \in [1,K] \),

\[
p(a_k, \theta_k) = \|f\|^2_{L^2} - \|a_k \cos \theta_k\|^2_{L^2} + O(\epsilon).
\]

(4.27)

Granting this, it follows from (4.24) that

\[
p(a_k, \theta_k) > p(a_l, \theta_l) \quad \forall k \neq l.
\]

(4.28)

To show (4.27), we notice that

\[
p(a_k, \theta_k) = \|f - a_k \cos \theta_k\|^2_{L^2}
\]

\[
= \|f\|^2_{L^2} + \|a_k \cos \theta_k\|^2_{L^2} - 2 \langle a_k \cos \theta_k, \sum_{k' \neq k} a_{k'} \cos \theta_{k'} + r(t), a_k \cos \theta_k \rangle
\]

(4.29)

By Lemma 4.1 we get

\[
\langle a_{k'}, \cos \theta_{k'}, a_k \cos \theta_k \rangle = O(\epsilon)
\]

(4.30)

for each \( k' \neq k \). By the assumption in Definition 3.1 we have \( r(t) = O(\epsilon) \), which implies

\[
\langle r(t), a_k \cos \theta_k \rangle = O(\epsilon).
\]

(4.31)

Substituting (4.30) and (4.31) into (4.29) proves (4.27).

Let \( \alpha := \frac{\sqrt{\epsilon}}{2} \) (so \( \alpha > 1 \)) and let

\[
\alpha_k := \begin{cases} 
\max_{t \in [0,1]} \frac{\theta_k(t)}{\theta_l(t)}, & k = 1 \\
\max_{t \in [0,1]} \frac{\theta_k(t)}{\theta_l(t)}, & k = K \\
\max \left\{ \max_{t \in [0,1]} \frac{\theta_k(t)}{\theta_l(t)}, \max_{t \in [0,1]} \frac{\theta_k(t)}{\theta_l(t)} \right\}, & 1 < k < K
\end{cases}
\]

(4.32)

\[
\alpha_{\min} := \min_{1 \leq k \leq K} \{\alpha_k\}.
\]

(4.33)

Let \( k_0 \) be such that \( \alpha_{k_0} = \alpha_{\min} \). We Claim that
(1) If \( k_0 > 1 \),
\[
\frac{\theta'(t)}{\theta'_{k_0-1}(t)} \geq \alpha, \quad \forall t \in [0, 1].
\] (4.34)

(2) If \( k_0 < K \),
\[
\frac{\theta'_{k_0+1}(t)}{\theta'(t)} \geq \alpha, \quad \forall t \in [0, 1].
\] (4.35)

To prove item (1), let us suppose on the contrary that there exists \( t_1 \in [0, 1] \) such that
\[
\frac{\theta'(t_1)}{\theta'_{k_0-1}(t_1)} < \alpha.
\] (4.36)

Then for any \( t \in [0, 1] \), we have
\[
\frac{\theta'(t)}{\theta'_{k_0-1}(t)} = \frac{\theta'(t)}{\theta'(t_1)} \cdot \frac{\theta'(t_1)}{\theta'_{k_0-1}(t_1)} \cdot \frac{\theta'_{k_0-1}(t_1)}{\theta'_{k_0-1}(t)} < 4\alpha = 2\sqrt{d}. \] (4.37)

On the other hand, (4.36) also implies that
\[
\frac{\theta'_{k_0}(t_1)}{\theta'(t_1)} = \frac{\theta'_{k_0}(t_1)}{\theta'_{k_0-1}(t_1)} \left( \frac{\theta'(t_1)}{\theta'_{k_0-1}(t_1)} \right)^{-1} > \frac{d}{\alpha} = 2\sqrt{d}. \] (4.38)

Using (4.37), (4.38), and the following estimates
\[
\frac{\theta'_{k_0}(t_1)}{\theta'(t_1)} \leq \frac{1}{d} \cdot \frac{\theta'_{k_0}(t_1)}{\theta'_{k_0-1}(t_1)} < \frac{\theta'_{k_0}(t_1)}{\theta'(t_1)},
\] (4.39)

we get
\[
\alpha_{k_0-1} < \max_{t \in [0, 1]} \frac{\theta'_{k_0}(t)}{\theta'(t)} \leq \alpha_{k_0}. \] (4.40)

This contradicts \( \alpha_{k_0} = \alpha_{\min} \). So item (1) is satisfied. Similarly we can show that item (2) is true.

It follows from (4.25) and (4.28) that
\[
p(a, \theta) \leq p(a_{k_0}, \theta_{k_0}). \] (4.41)

Using Theorem 4.1 by replacing condition (4.1) by (4.23) in Remark 4.1, we obtain that
\[
\|a \cos \theta - a_{k_0} \cos \theta_{k_0}\|_{L^2} = O(\sqrt{\epsilon}). \] (4.42)

This implies
\[
\|a \cos \theta - a_{k_0} \cos \theta_{k_0}\|_{L^2} = O(\sqrt{\epsilon}). \] (4.43)

Therefore we obtain
\[
|p(a, \theta) - p(a_{k_0}, \theta_{k_0})| = \frac{1}{d} \left| \frac{\theta'_{k_0}(t_1)}{\theta'(t_1)} \right| \left| a_{k_0} \cos \theta_{k_0} - a \cos \theta \sum_{k \neq k_0} a_k \cos \theta_k + r(t) \right| \] (4.44)

where the first equality is deduced using an argument similar to the equalities in (4.4), and the second one follows from (4.43). So by (4.24), we see that \( k_0 = l \) and \( \alpha_k > \alpha_l \) for all \( k \neq l \). Thus (4.12) implies (4.26). This completes the proof. \( \square \)
5 Concluding Remarks

In this paper, we discussed the uniqueness of the decomposition obtained by sparse time frequency decomposition. We proved that under the assumption of scale separation, the decomposition is almost unique up to an error associate with the scale separation. Moreover, we also show that under the same assumption, matching pursuit could give this sparse decomposition. The results in this paper establish a solid foundation for the sparse time frequency decomposition.

In our future work, we would like to relax the assumption of scale separation. In many problems, the decompositions seem to be unique although the scale separation is not satisfied. We plan to perform further theoretical study and get some guidance to decompose signals with poor scale separation.

Acknowledgments. This work was supported by NSF FRG Grant DMS-1159138, DMS-1318377, an AFOSR MURI Grant FA9550-09-1-0613 and a DOE grant DE-FG02-06ER25727. The research of Dr. Z. Shi was also in part supported by a NSFC Grant 11201257. The research of Dr. C.G. Liu was also supported by a NSFC Grant 11371173.

Appendix A: Proof of Lemma 3.1

Proof. of Lemma 3.1

First, we have that
\[ \hat{\psi}(\omega \theta'(t)) = \int_{\mathbb{R}} e^{-i\omega \theta'(t)z} \psi(z) dz = \frac{1}{\omega} \int_{\mathbb{R}} e^{-i\theta'(t)z} \psi \left( \frac{z}{\omega} \right) dz. \] (5.1)

Using this relation, we obtain
\[ \frac{1}{\sqrt{\omega}} \int_{\mathbb{R}} a(\tau)e^{-i\theta(\tau)} \psi \left( \frac{\tau - t}{\omega} \right) d\tau - |\omega|^{-1/2} \int_{\mathbb{R}} a(t)e^{-i\theta(t)} \hat{\psi}(\omega \theta'(t)) \]
\[ = \frac{1}{\sqrt{\omega}} \left[ \int_{\mathbb{R}} (a(\tau) - a(t)) e^{-i\theta(\tau)} \psi \left( \frac{\tau - t}{\omega} \right) d\tau + a(t) \int_{\mathbb{R}} e^{-i\theta(\tau)} - e^{-i\theta(t) + \theta'(t)(\tau - t))} \psi \left( \frac{\tau - t}{\omega} \right) d\tau \right]. \] (5.2)

For the first term, we have
\[ |\omega|^{-1/2} \int_{\mathbb{R}} (a(\tau) - a(t)) e^{-i\theta(\tau)} \psi \left( \frac{\tau - t}{\omega} \right) d\tau \]
\[ = |\omega|^{-1/2} \int_{\mathbb{R}} h(\tau, t)e^{-i\theta(\tau)} d\tau \]
\[ = -i|\omega|^{-1/2} \int_{\mathbb{R}} \left( \frac{h(\tau, t)}{\theta'(\tau)} \right)' e^{-i\theta(\tau)} d\tau, \]

where
\[ h(\tau, t) = (a(\tau) - a(t)) \psi \left( \frac{\tau - t}{\omega} \right). \] (5.3)

Direct calculation gives that
\[ \left( \frac{h(\tau, t)}{\theta'(\tau)} \right)' = \frac{a'(\tau)}{\theta'(\tau)} \psi \left( \frac{\tau - t}{\omega} \right) + \frac{a(\tau) - a(t)}{\omega} \psi \left( \frac{\tau - t}{\omega} \right) \frac{h(\tau, t)\theta''(\tau)}{(\theta'(\tau))^2} \] (5.4)
Using the assumption that \((a, \theta) \in U_\epsilon\), we have
\[
\frac{a'(\tau)}{\theta'(\tau)} \leq \epsilon, \quad \frac{\theta''(\tau)}{(\theta'(\tau))^2} \leq \epsilon,
\]
and
\[
\left| \frac{a(\tau) - a(t)}{\theta'(\tau)} \right| = \left| \frac{a'(t_\tau)(\tau - t)}{\theta'(\tau)} \right| = \left| \frac{\theta'(t_\tau)}{\theta'(\tau)} \cdot \frac{a'(t_\tau)(\tau - t)}{\theta'(t_\tau)} \right| \leq 2\epsilon|\tau - t|,
\]
where \(t_\tau\) is a point between \(t\) and \(\tau\). Then, we obtain
\[
|\omega|^{-1/2} \left| \int_R (a(\tau) - a(t)) e^{-i\theta(\tau)} \psi \left( \frac{\tau - t}{\omega} \right) d\tau \right| \leq C\epsilon|\omega|^{1/2}(I_1 + I_2),
\]
where \(C\) is a constant and
\[
I_1 = \int_R |\psi(\tau)| d\tau, \quad I_2 = \int_R |\tau \psi'(\tau)| d\tau.
\]
Now, we turn to bound the second term of (5.2). We have
\[
|\omega|^{-1/2} \int_R \left( e^{-i\theta(\tau)} - e^{-i\theta(t) + \theta'(t)(\tau-t)} \right) \psi \left( \frac{\tau - t}{\omega} \right) d\tau
\]
\[
= |\omega|^{-1/2} \int_R g(\tau, t) e^{-i\theta(\tau)} d\tau
\]
\[
= -i|\omega|^{-1/2} \int_R \left( \frac{g(\tau, t)}{\theta'(\tau)} \right)' e^{-i\theta(\tau)} d\tau,
\]
where
\[
g(\tau, t) = (1 - e^{i\Delta \theta}) \psi \left( \frac{\tau - t}{\omega} \right)
\]
and \(\Delta \theta = \theta(\tau) - \theta(t) - \theta'(t)(\tau - t)\).

Direct calculations show that
\[
\left( \frac{g(\tau, t)}{\theta'(\tau)} \right)' = -i \frac{(\theta'(\tau) - \theta'(t))}{\theta'(\tau)} e^{i\Delta \theta(\tau, t)} \psi \left( \frac{\tau - t}{\omega} \right) + \frac{1}{\theta'(\tau)} (1 - e^{i\Delta \theta(\tau, t)}) \frac{1}{\omega} \psi' \left( \frac{\tau - t}{\omega} \right)
\]
\[
- \frac{g(\tau, t) \theta''(\tau)}{(\theta'(\tau))^2}.
\]
Among the three terms, the third one is easiest to bound,
\[
\left| \left( \frac{g(\tau, t) \theta''(\tau)}{(\theta'(\tau))^2} \right) \right| \leq 2\epsilon \left| \psi \left( \frac{\tau - t}{\omega} \right) \right|.
\]
Then, we turn to estimate the second term. Let \(g_2(\tau, t) = \frac{1}{\theta'(\tau)} (1 - e^{i\Delta \theta(\tau, t)}) \frac{1}{\omega} \psi' \left( \frac{\tau - t}{\omega} \right)\). By using the integration by part again, we get
\[
\int_R g_2(\tau, t) e^{-i\theta(\tau)} d\tau = -i \int_R \left( \frac{g_2(\tau, t)}{\theta'(\tau)} \right)' e^{-i\theta(\tau)} d\tau,
\]
and
\[
\left( \frac{g_2(\tau, t)}{\theta'(\tau)} \right)' = - \frac{\theta'(\tau)(1 - e^{i\Delta \theta(\tau, t)})}{\omega(\theta'(\tau))^3} \psi' \left( \frac{\tau - t}{\omega} \right) - \frac{\theta'(\tau) - \theta'(t)}{\omega(\theta'(\tau))^2} e^{i\Delta \theta(\tau, t)} \psi' \left( \frac{\tau - t}{\omega} \right)
\]
\[
+ \frac{1}{\omega^2(\theta'(\tau))^2} \psi'' \left( \frac{\tau - t}{\omega} \right) \left( \frac{g_2(\tau, t) \theta''(\tau)}{(\theta'(\tau))^2} \right).
\]
We need some preparations to bound (5.13). First, we bound $|\theta'(t) - \theta'(\tau)|$ by the following

$$
|\theta'(\tau) - \theta'(t)| = |\theta''(t_\tau)(t-\tau)| \leq \epsilon|\theta'(t_\tau)|^2|\tau - t| \leq 4\epsilon|\theta'(\tau)|^2|\tau - t|,
$$

(5.14)

where $t_\tau$ is a point between $t$ and $\tau$.

We now estimate $|1 - e^{i\Delta_\theta(t,\tau)}|$.

$$
|1 - e^{i\Delta\theta(t,\tau)}| \leq |\Delta\theta(t,\tau)| \leq |\theta'(t^*) - \theta'(t)||t - \tau| \leq 4\epsilon|\theta'(\tau)|^2|t - \tau|^2,
$$

(5.15)

where $t^*$ is a number between $t$ and $\tau$. And also

$$
|1 - e^{i\Delta\theta(t,\tau)}| \leq |\Delta\theta(t,\tau)| \leq |\theta'(t^*) - \theta'(t)||t - \tau| \leq \theta'(\tau)|t - \tau|.
$$

(5.16)

Then, (5.13) can be bounded as follows,

$$
2\frac{\theta''(\tau)}{\omega(\theta'(\tau))^3}e^{i\Delta\theta(t,\tau)} \left|\frac{\tau - t}{\omega}\right| \leq 2\epsilon \left|\frac{\tau - t}{\omega}\right| \psi' \left(\frac{\tau - t}{\omega}\right),
$$

(5.17)

$$
\left|\frac{\theta'(\tau) - \theta'(t)}{\omega(\theta'(\tau))^2} e^{i\Delta\theta(t,\tau)} \psi' \left(\frac{\tau - t}{\omega}\right)\right| \leq 4\epsilon \left|\frac{\tau - t}{\omega}\right| \psi' \left(\frac{\tau - t}{\omega}\right),
$$

(5.18)

and

$$
\left|\frac{1 - e^{i\Delta\theta(t,\tau)}}{\omega^2(\theta'(\tau))^2} \psi'' \left(\frac{\tau - t}{\omega}\right)\right| \leq 4\epsilon \left|\frac{\tau - t}{\omega}\right|^2 \psi'' \left(\frac{\tau - t}{\omega}\right).
$$

(5.19)

By combining these inequalities, we get

$$
\left|\int_{\mathbb{R}} g_2(\tau, t)e^{-i\theta(\tau)} d\tau\right| \leq C\omega(I_2 + I_3)\epsilon,
$$

(5.20)

where

$$
I_3 = \int_{\mathbb{R}} |\tau^2\psi''(\tau)| d\tau.
$$

(5.21)

The estimation of the first term of (5.10) is most involved. Denote

$$
h_m(t, \tau) = -\frac{i(\theta'(\tau) - \theta'(t))^m}{\theta'(\tau)^m} e^{i\Delta\theta(t,\tau)} \psi \left(\frac{\tau - t}{\omega}\right),
$$

(5.22)

By integration by part, we get

$$
\int_{\mathbb{R}} h_m(t, \tau)e^{-i\theta(\tau)} d\tau = -i \int_{\mathbb{R}} \left(\frac{h_m(t, \tau)}{\theta'(\tau)}\right)' e^{-i\theta(\tau)} d\tau,
$$

(5.23)

and

$$
\left(\frac{h_m(t, \tau)}{\theta'(\tau)}\right)' = -\frac{in\theta''(\tau)(\theta'(\tau) - \theta'(t))^m-1}{\theta'(\tau)^{m+1}} e^{i\Delta\theta(t,\tau)} \psi \left(\frac{\tau - t}{\omega}\right) + i(m + 1)\theta''(\tau)(\theta'(\tau) - \theta'(t))^m \frac{e^{i\Delta\theta(t,\tau)} \psi \left(\frac{\tau - t}{\omega}\right)}{\theta'(\tau)^{m+2}} + \frac{(\theta'(\tau) - \theta'(t))^{m+1}}{\theta'(\tau)^{m+1}} \frac{e^{i\Delta\theta(t,\tau)} \psi \left(\frac{\tau - t}{\omega}\right)}{\theta'(\tau)^{m+1}} - \frac{i(\theta'(\tau) - \theta'(t))^m}{\omega(\theta'(\tau))^m+1} e^{i\Delta\theta(t,\tau)} \psi' \left(\frac{\tau - t}{\omega}\right).
$$

(5.24)
Let $\sigma = \max_{\tau \in [0,1]} \frac{\theta'(\tau) - \theta'(t)}{\theta'(\tau)}$. Using the assumption $\frac{\max \theta'}{\min \theta'} < 2$, we have that $\sigma < 1$. Then, (5.24) gives

$$H_m \leq (m+1)\epsilon \omega \sigma^m I_1 + m \epsilon \omega \sigma^{m-1} I_1 + C \epsilon \omega \sigma^{m-1} I_2 + H_{m+1},$$

(5.25)

where

$$H_m = \left| \int_R h_m(\tau,t)e^{-i\theta(\tau)}d\tau \right| \leq \sigma^m \omega I_1 \to 0, \quad \text{as } m \to \infty.$$  

(5.26)

The above recursion relation gives the upper bound of the first term of (5.20),

$$H_1 \leq \epsilon \omega I_1 \sum_{m=1}^{\infty} \left((m+1)\sigma^m + m\sigma^{m-1}\right) + C \epsilon \omega I_2 \sum_{m=1}^{\infty} \sigma^{m-1} \leq C \epsilon \omega (I_1 + I_2).$$

(5.27)

Then the proof is completed by combining (5.21), (5.24), (5.24) and (5.27).

Appendix B: Proof of Lemma 4.1

To prove Lemma 4.1 we need the following technical lemma.

**Lemma 5.1.** Let $\epsilon \in (0,1)$ and $g(t)$ be a positive, continuous, and piecewise $C^1$ function on $[c,c+2n\pi]$, where $n$ is an integer. Suppose

$$\left| \frac{g'(t)}{g(t)} \right| < \epsilon, \quad \forall t \in [c,d].$$

(5.28)

Then

$$\int_c^{c+2n\pi} g(t) \cos t dt < C \epsilon \int_c^{c+2n\pi} g(t) dt,$$

(5.29)

where $C > 0$ is a constant.

**Proof.** For each $m \in \{0, 1, 2, \cdots, n\}$, let $t_m = c + 2m\pi$. We have

$$\left| \int_c^{c+2n\pi} g(t) \cos t dt \right| = \left| \sum_{m=1}^{n} \int_{t_{m-1}}^{t_m} g(t) \cos t dt \right| = \left| \sum_{m=1}^{n} \int_{t_{m-1}}^{t_m} [g(t) - g(t_{m-1})] \cos t dt \right|$$

$$= \sum_{m=1}^{n} \int_{t_{m-1}}^{t_m} \left( \int_{t_{m-1}}^{t} g'(s) ds \right) \cos t dt$$

$$\leq \sum_{m=1}^{n} \int_{t_{m-1}}^{t_m} \left( \int_{t_{m-1}}^{t} |g'(s)| ds \right) dt$$

$$\leq \sum_{m=1}^{n} \int_{t_{m-1}}^{t_m} \left( \int_{t_{m-1}}^{t} \epsilon g(s) ds \right) dt$$

$$= \sum_{m=1}^{n} 2\pi \epsilon \left( \int_{t_{m-1}}^{t_m} g(s) ds \right)$$

$$= 2\pi \epsilon \int_c^{c+2n\pi} g(s) ds.$$

We complete the proof.
From the above proof, we can take $C = 2\pi$ in (5.29). But a more refined estimation shows that $C = 1$ is enough.

Now, we can give the proof of Lemma 4.1

Proof. of Lemma 4.1

First, we have

$$\|a(t) \cos \theta(t)\|_{L^2}^2 = \frac{1}{2} \|a(t)\|_{L^2}^2 + \frac{1}{2} \int_0^1 a^2(t) \cos 2\theta(t) dt. \quad (5.30)$$

Let $s = 2\theta(t)$. Then we obtain

$$\int_0^1 a^2(t) \cos 2\theta(t) dt = \frac{1}{2} \int_{2\theta(0)}^{2\theta(1)} g(s) \cos s ds, \quad (5.31)$$

where $t(s) := \theta^{-1} (\frac{s}{2})$ and

$$g(s) := \frac{a^2(t(s))}{\theta'(t(s))}. \quad (5.32)$$

So the derivative of $g$ is

$$g'(s) = \frac{a(t(s))a'(t(s))}{\theta'(t(s))^2} - \frac{a^2(t(s)) \cdot \theta''(t(s))}{2[\theta'(t(s))]^3}. \quad (5.33)$$

Hence

$$\left| \frac{g'(s)}{g(s)} \right| = \left| \frac{a'(t(s))}{a(t(s)) \cdot \theta'(t(s))} - \frac{\theta''(t(s))}{2[\theta'(t(s))]^2} \right| < \frac{3}{2} \epsilon, \quad \forall s \in [2\theta(0), 2\theta(1)]. \quad (5.34)$$

Using Lemma 5.1 we obtain

$$\left| \int_0^1 a^2(t) \cos 2\theta(t) dt \right| < \frac{3C}{4} \epsilon \int_{2\theta(0)}^{2\theta(1)} g(s) ds \quad (5.35)$$

$$= \frac{3C}{2} \pi \epsilon \int_0^1 a^2(t) dt.$$

The above estimate and (5.30) imply (4.14).

To prove (4.10), we represent the inner product in this inequality as

$$\langle a \cos \theta, a \cos \tilde{\theta} \rangle = \frac{1}{2} \left[ \int_0^1 a(t) \bar{a}(t) \cos(\tilde{\theta}(t) + \theta(t)) dt + \int_0^1 a(t) \bar{a}(t) \cos(\tilde{\theta}(t) - \theta(t)) dt \right]. \quad (5.36)$$

Let $s = \theta(t) - \tilde{\theta}(t)$. We obtain

$$\int_0^1 a(t) \bar{a}(t) \cos(\tilde{\theta}(t) - \theta(t)) dt = \int_{\tilde{\theta}(0) - \theta(0)}^{\tilde{\theta}(1) - \theta(1)} g(s) \cos s ds, \quad (5.37)$$

where $t(s) = (\tilde{\theta} - \theta)^{-1}(s)$ and

$$g(s) = \frac{a(t(s))\bar{a}(t(s))}{\theta'(t(s)) - \theta'(t(s))}. \quad (5.38)$$

Thus, we obtain

$$\frac{d}{ds} g(s) = \frac{(\tilde{\theta}' - \theta')(a \bar{a} + a' \bar{a}) - a\bar{a}(\tilde{\theta}'' - \theta'')}{(\tilde{\theta}'(t(s)) - \theta'(t(s)))^3} \quad (5.39)$$

21
and
\[ \left| \frac{g'(s)}{g(s)} \right| < \epsilon \frac{2}{(1 - \beta - 1)^2}, \quad \forall s \in [\tilde{\theta}(0) - \theta(0), \tilde{\theta}(1) - \theta(1)]. \tag{5.40} \]

Using Lemma 5.1 again, we get
\[ \left| \int_0^1 a(t)\bar{a}(t) \cos(\tilde{\theta}(t) - \theta(t)) \, dt \right| < \epsilon \frac{2C}{(1 - \beta - 1)^2} \int_0^1 a(t)\bar{a}(t) \, dt. \tag{5.41} \]

Similarly, we can show
\[ \left| \int_0^1 a(t)\bar{a}(t) \cos(\tilde{\theta}(t) + \theta(t)) \, dt \right| < 2C\epsilon \int_0^1 a(t)\bar{a}(t) \, dt. \tag{5.42} \]

Thus (4.16) follows by combining (5.36), (5.41) and (5.42).

References

[1] M. Aharon, M. Elad, and A. M. Bruckstein. The k-svd: An algorithm for designing of over-complete dictionaries for sparse representations. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

[2] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal recovery from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52:489–509, 2006.

[3] E. Candes and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies. *IEEE Trans. Inform. Theory*, 52:5406–5425, 2006.

[4] S. Chen, D. L. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20:33–61, 1998.

[5] I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series on Applied Mathematics, Vol. 61, SIAM Publications, 1992.

[6] I. Daubechies, J. Lu, and H. Wu. Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. *Appl. Comp. Harmonic Anal.*, 30:243–261, 2011.

[7] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52:1289–1306, 2006.

[8] K. Dragomiretskiy and D. Zosso. Varitational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3):531–544, 2014.

[9] K. Engan, K. Skretting, and J. Husøy. Family of iterative ls-based dictionary learning algorithms, ils-dla, for sparse signal representation. *Digital Signal Processing*, 17(1):32–49, 2007.

[10] N. E. Huang et al. The empirical mode decomposition and the hilbert spectrum for non-linear and non-stationary time series analysis. *Proc. R. Soc. Lond. A*, 454:903–995, 1998.

[11] P. Flandrin. *Time-Frequency/Time-Scale Analysis*. Academic Press, 1999.
[12] J. Gilles. Empirical wavelet transform. *IEEE Transactions on Signal Processing*, 61:3999–4010, 2013.

[13] M. Lewicki and T. Sejnowski. Learning overcomplete representations. *Neural Computation*, 12(2):337–365, 2000.

[14] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2009a.

[15] S. Mallat. *A wavelet tour of signal processing: the Sparse way*. Academic Press, 2009.

[16] Z. Zhang S. G. Mallat. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 12:3397–3415, 1993.

[17] K. Skretting and K. Engan. Recursive least squares dictionary learning algorithm. *IEEE Transactions on Signal Processing*, 58(4):2121–2131, 2010.

[18] T. Y. Hou and Z. Shi. Data-driven time-frequency analysis. *Appl. Comput. Harmon. Anal.*, 35:284–308, 2013.

[19] T. Y. Hou and Z. Shi. Sparse time-frequency decomposition by dictionary learning. *arXiv cs.IT 1311.1163*, 2014.

[20] T. Y. Hou and Z. Shi and P. Tavallali. Convergence of a data-driven time-frequency analysis method. *Appl. Comput. Harmon. Anal.*, 37(2):235–270, 2014.

[21] Z. Wu and N. E. Huang. Ensemble empirical mode decomposition: a noise assisted data analysis method. *Advances in Adaptive Data Analysis*, 1:1–41, 2008.