

# Sparse time-frequency representation of nonlinear and nonstationary data

*Dedicated to Professor Shi Zhong-Ci on the Occasion of his 80th Birthday*

HOU Thomas Yizhao<sup>1,\*</sup> & SHI ZuoQiang<sup>2</sup>

<sup>1</sup>*Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, USA;*

<sup>2</sup>*Mathematical Sciences Center, Tsinghua University, Beijing 100084, China*

*Email: hou@cms.caltech.edu, zqshi@math.tsinghua.edu.cn*

Received May 16, 2013; accepted September 10, 2013; published online October 17, 2013

**Abstract** Adaptive data analysis provides an important tool in extracting hidden physical information from multiscale data that arise from various applications. In this paper, we review two data-driven time-frequency analysis methods that we introduced recently to study trend and instantaneous frequency of nonlinear and nonstationary data. These methods are inspired by the empirical mode decomposition method (EMD) and the recently developed compressed (compressive) sensing theory. 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  $a$  is assumed to be less oscillatory than  $\cos(\theta(t))$  and  $\theta' \geq 0$ . This problem can be formulated as a nonlinear  $l^0$  optimization problem. We have proposed two methods to solve this nonlinear optimization problem. The first one is based on nonlinear basis pursuit and the second one is based on nonlinear matching pursuit. Convergence analysis has been carried out for the nonlinear matching pursuit method. Some numerical experiments are given to demonstrate the effectiveness of the proposed methods.

**Keywords** sparse representation, time-frequency analysis, data-driven

**MSC(2010)** 65N21, 94A12

**Citation:** Hou T Y, Shi Z Q. Sparse time-frequency representation of nonlinear and nonstationary data. *Sci China Math*, 2013, 56, doi: 10.1007/s11425-013-4733-7

## 1 Introduction

Nowadays we must process a massive amount of data in our daily life and in our scientific research. How to extract hidden pattern or physical information from these data has become essential in our scientific discovery. This calls for the need to develop a truly adaptive data analysis method. Traditional data analysis methods based on pre-determined basis provide an effective tool to process linear and stationary data. When applying these methods to analyze nonlinear and nonstationary data, they tend to generate some artificial harmonics. These limitations have been alleviated to some extent with the introduction of time-frequency analysis which represents a signal with a joint function of both time and frequency. Wavelet analysis provides an excellent tool for time-frequency analysis by introducing multi-scales to characterize signals, see [10, 23, 26, 31].

The concept of instantaneous frequency is an important development in time-frequency analysis. It can be traced back to Van der Pol [42], who introduced the so-called analytic signal method (AS) that

\*Corresponding author

uses the Hilbert transform to determine instantaneous frequency of a signal. Until very recently, this method works mostly for monocomponent signals in which the number of zero-crossings is equal to the number of local extrema [2]. There were other attempts to define instantaneous frequency such as the zero-crossing method [28,37,39] and the Wigner-Ville distribution method [2,13,24,25,34,36]. The zero-crossing method does not work well for signals with multiple components and is sensitive to noise. The methods based on the Wigner-Ville distribution suffer the interference between different components.

The empirical mode decomposition method (EMD) introduced by Huang et al. in [22] represents an important progress in adaptive data analysis. It provides a truly adaptive method to decompose a signal into a collection of intrinsic mode functions (IMFs) that give well-behaved Hilbert transforms. Application of the analytic signal method to each IMF gives physically meaningful instantaneous frequency and its time-frequency representation. Since its first introduction in 1998, the EMD method has found many applications, especially in the study of trend, detrend, and the variability of nonlinear and non-stationary time series, see, e.g., [45].

Inspired by the EMD method and the recently developed compressed (compressive) sensing theory, we have recently proposed two data-driven time-frequency analysis methods [18,19]. Our data-driven time-frequency analysis method is based on looking for the sparsest representation of a multiscale signal over certain multiscale basis. The multiscale basis is adapted to the signal instead of the determined *a priori*. This explains the term “data-driven”. In this sense, our method is very different from the compressed (compressive) sensing problem in which the basis is assumed to be known a priori. In our method, we reformulate the problem as a nonlinear optimization. We find the sparse basis and the decomposition simultaneously by looking for the sparsest decomposition among the largest possible dictionary consisting of intrinsic mode functions. The adoption of this data-driven basis and the search for the sparsest decomposition over a highly redundant basis make our time-frequency analysis method fully adaptive to the signal.

Our nonlinear optimization problem is formulated as follows:

$$\min_{(a_k)_{1 \leq k \leq M}, (\theta_k)_{1 \leq k \leq M}} M, \quad \text{subject to } f = \sum_{k=1}^M a_k \cos \theta_k, \quad a_k \cos \theta_k \in \mathcal{D}, \quad (1.1)$$

where  $\mathcal{D}$  is the dictionary we use to decompose the signal which will be defined later in the paper. When the signal is polluted by noise, the equality in the above constraint is relaxed to be an inequality depending on the noise level.

The above optimization problem can be viewed as a nonlinear version of the  $l^0$  minimization problem, which is known to be very challenging to solve. Inspired by the compressed (compressive) sensing theory [3,4,12,16], we have proposed two approaches to solve this nonlinear optimization problem. The first one is based on nonlinear basis pursuit using a  $TV^3$  norm. The second one is based on  $l^1$ -regularized nonlinear matching pursuit. The first approach has a close connection to the EMD method in the sense that the  $TV^3$  minimization produces local mean and the envelope function that are essentially cubic spline polynomials. The nonlinear matching pursuit has the advantage of being robust to noise and can be implemented very efficiently by fast Fourier transform (FFT). Moreover, for periodic data that satisfy certain scale separation conditions, we have proved the convergence of our nonlinear matching pursuit method. When the signal satisfies certain nonlinear sparsity conditions, our method can give exact recovery of the original signal [21]. When the sparsity condition is satisfied only approximately, we prove that our method recovers the IMFs and their instantaneous frequencies accurately [21].

We have performed extensive numerical experiments to test the robustness and the accuracy of our data-driven time-frequency analysis method for both synthetic data and some real data. We have also tested our method for sparsely sampled data or incomplete data and applied our method to study some real world data. In all cases, we have shown that the nonlinear matching pursuit method can decompose a multiscale signal into a sparse collection of intrinsic mode functions. In one of our recent applications to study climate data, we have discovered a new nearly annual cycle that has not been revealed previously [40]. Application of a simplified version of our method to study the pressure waveform of a heart patient has led to a new non-intrusive and highly effective medical index to diagnose heart diseases [32]. We have

also compared the performance of our method with that of EMD or EEMD (which stands for ensemble empirical mode decomposition) [44]. For data without noise, our methods give results comparable with those obtained by the EMD method. For noisy data, our method seems to give more accurate recovery of the instantaneous frequency and IMFs than EMD and EEMD.

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 et al. [11]. We have performed some preliminary numerical experiments to compare the synchrosqueezed wavelet approach with our nonlinear matching pursuit method. For data with good scale separation property, the two methods give complementary results. We note that the synchrosqueezed wavelet method does not extract IMFs directly. Some post processing is required to obtain IMFs.

The remaining of the paper is organized as follows. In Section 2, we give a brief review of some existing data analysis methods such as matching pursuit, basis pursuit and the EMD method. In Section 3, the first adaptive data analysis method based on nonlinear basis pursuit is introduced. We review the second data-driven time-frequency analysis method based on nonlinear matching pursuit in Section 4. In Section 5, we summarize our convergence results for the nonlinear matching pursuit method. Several numerical results are given in Section 6 to demonstrate the effectiveness of our method and to confirm our theoretical results. Some concluding remarks are made in Section 7.

## 2 Brief review of the existing sparse decomposition methods

There has been a growing interest in the signal processing literature on the development of the sparse signal representations over a redundant dictionary. In particular, matching pursuit [27] and basis pursuit [8] have attracted a lot of attention in recent years due to the development of compressed (compressive) sensing. These methods have two essential features, a dictionary to decompose the signal and a decomposition method to select the sparsest decomposition.

### 2.1 Dictionaries

Finding a good dictionary is essential in obtaining a sparse representation for a family of signals. Roughly speaking, a dictionary is a collection of parameterized waveforms  $\mathcal{D} = \{\phi_\gamma\}_{\gamma \in \Gamma}$ . Here we review a few dictionaries that have been used widely in the literature.

#### 2.1.1 A Fourier dictionary

The Fourier dictionary is perhaps the most well-known dictionary that has been used in many applications. It consists of a collection of sinusoidal waveforms with  $\gamma = (\omega, \nu)$ , where  $\omega \in [0, 2\pi]$  is an angular frequency variable and  $\nu \in \{0, 1\}$  indicates the phase type: sine or cosine. Specifically, they consist of the following two families,

$$\phi_{\omega,0}(t) = \cos(\omega t), \quad \phi_{\omega,1}(t) = \sin(\omega t), \quad \forall t \in \mathbb{R}. \quad (2.1)$$

For a standard Fourier dictionary,  $\omega$  runs through the set of all cosines with Fourier frequencies  $\omega_k = 2k\pi/n$ ,  $k = 0, 1, \dots, n/2$ , and all sines with Fourier frequencies  $\omega_k = 2k\pi/n$ ,  $k = 1, \dots, n/2 - 1$ , where  $n$  is the number of sample points. Sometimes, we also use an overcomplete Fourier dictionary by sampling the frequencies more finely. Let  $l > 1$ . We may choose  $\omega_k = 2k\pi/(ln)$ ,  $k = 0, 1, \dots, ln/2$  for cosines and  $\omega_k = 2k\pi/(ln)$ ,  $k = 1, \dots, ln/2 - 1$  for sines. This gives rise to an  $l$ -fold overcomplete system. As we will see later, we will use this kind of overcomplete Fourier dictionary to decompose non-periodic data.

#### 2.1.2 A wavelet dictionary

A wavelet dictionary is another widely used dictionary. It consists of a collection of translations and dilations of the basic mother wavelet  $\psi$  and a scaling function  $\varphi$ . Specifically, the waveforms are given as follows:

$$\phi_{a,b,0}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \quad \phi_{a,b,1}(t) = \frac{1}{\sqrt{a}}\varphi\left(\frac{t-b}{a}\right), \quad \forall t \in \mathbb{R}. \quad (2.2)$$

In this dictionary, the index  $\gamma = (a, b, \nu)$ , where  $a \in (0, \infty)$  is a scale variable,  $b \in \mathbb{Z}$  indicates location and  $\nu \in \{0, 1\}$  indicates gender. In a standard wavelet dictionary, we typically have  $a_j = 2^j/n$ ,  $j = j_0, \dots, \log_2(n) - 1$  which give dyadic scales,  $b_{j,k} = ka_j$ ,  $k = 0, \dots, 2^j - 1$  with  $j_0$  being the coarse scale. One can also define an overcomplete wavelet dictionary by sampling the locations more finely.

### 2.1.3 A time-frequency dictionary

One of the oldest time-frequency dictionaries is the Gabor dictionary [14]. In this dictionary, we take  $\gamma = (\omega, \tau, \theta, \delta)$ , where  $\omega \in [0, \pi)$  is frequency,  $\tau$  is a location,  $\theta$  is a phase, and  $\delta$  is the duration. We define the waveforms as follows:

$$\phi_\gamma(t) = \exp\left(-\frac{(t-\tau)^2}{\delta^2}\right) \cos(\omega(t-\tau) + \theta), \quad \forall t \in \mathbb{R}. \quad (2.3)$$

The frequencies of these waveforms are centered around  $\omega$  and essentially vanish far away from  $\tau$ .

### 2.1.4 An EMD dictionary

The EMD method [22] does not use a dictionary. But we can also define a dictionary as the collection of all intrinsic mode functions (IMFs). The IMFs are defined by enforcing the following two conditions:

1. The number of the extrema and the number of the zero crossings of the function must be equal or differ at most by one;
2. At any point of the function, the average of the upper envelope and the lower envelope defined by the local extrema should be zero (symmetric with respect to zero).

This dictionary is perhaps the largest one among the dictionaries listed here. In fact, any oscillatory sinusoidal wave multiplied by a smooth envelope function satisfies the definition of IMFs. Many commonly used dictionaries are included in this dictionary. For example, all the elements of a Fourier dictionary defined in (2.1) (standard or overcomplete) and the Gabor dictionary are IMFs. Some wavelets, such as the Morlet wavelet, also satisfy the conditions of IMFs.

## 2.2 Decomposition methods

Once we have chosen our desired dictionary, we need to select an effective decomposition method to construct the sparse representation of a signal. Here we review a few decomposition methods that can be used for this purpose. In recent years, there have been a lot of research activities in looking for the sparsest representation of a signal over a redundant dictionary, see, e.g., [4, 7, 8, 12, 27]. Specifically, we look for a decomposition of a signal  $f$  over a given dictionary  $\mathcal{D} = \{\phi_\gamma\}_{\gamma \in \Gamma}$  as follows:

$$f = \sum_{k=1}^M \alpha_{\gamma_k} \phi_{\gamma_k} + R^M, \quad (2.4)$$

with the smallest  $M$ , where  $R^M$  is the residual. How sparse we can represent a signal depends on the choice of the dictionary that we use to decompose the signal. In general, a more redundant dictionary tends to give better sparsity of the decomposition. However, if the dictionary is not a basis, the decomposition is not unique. We need to give a criterion to select the “best” decomposition among all the possible choices.

### 2.2.1 Matching pursuit

Matching pursuit introduced by Mallat and Zhang in [27] is a general decomposition method to obtain a sparse decomposition of a signal. Matching pursuit builds up a sequence of sparse approximations step by step. At stage  $k$ , the method searches for an atom that gives the best match to the residual and adds it to the current approximation. When stopped after a few steps, one obtains an approximate sparse representation of the form (2.4) using only a few atoms. A similar algorithm was proposed for the Gabor dictionary by Qian and Chen [35].

When the dictionary is orthogonal, the method works perfectly. To improve the performance of matching pursuit, several algorithms have been proposed, including the orthogonal matching pursuit (OMP)

which was analyzed by Tropp and Gilbert [41], compressive sampling matching pursuit by Needell and Tropp [29], regularized orthogonal matching pursuit (ROMP) by Needell and Vershynin [30], etc.

### 2.2.2 Basis pursuit

Basis pursuit introduced by Chen et al. [8] is another important class of decomposition methods. Suppose we have a discrete dictionary of  $p$  waveforms and we collect all these waveforms as columns of an  $n$  by  $p$  matrix  $\Phi$ . We can reformulate the decomposition problem (2.4) as follows:

$$\mathbf{s} = \Phi\alpha, \quad (2.5)$$

where  $\alpha = (\alpha_\gamma)$  is the vector of coefficients in (2.4). The main idea of basis pursuit is to find a sparse representation of the signal by minimizing the discrete  $l^1$  norm of the coefficients. Thus, the decomposition is obtained by solving the following optimization problem,

$$\min \|\alpha\|_{l^1}, \quad \text{subject to } \Phi\alpha = \mathbf{s}. \quad (2.6)$$

An important property of basis pursuit is that it can recover the exact solution of the original  $l^0$  minimization problem under some sparsity condition on the data [7, 12]. Basis pursuit has been applied to a variety of applications. Although basis pursuit reduces an  $l^0$  minimization problem to a convex  $l^1$  minimization problem, the computational cost of solving the  $l^1$  minimization is more expensive than that of the least-square problem in matching pursuit. We remark that several powerful algorithms have been introduced to speed up the  $l^1$  minimization problem, such as the split Bregman method [15] and proximal algorithms [1, 9]. The split Bregman method is also known as the augmented Lagrangian technique and can be derived from Douglas-Rachford algorithm [38] or PPXA+ [33].

### 2.2.3 The EMD decomposition via a sifting process

Although the EMD method was not designed as a sparse decomposition method, it can be used for this purpose. Like matching pursuit, the EMD method decomposes a signal to its IMFs sequentially. The basic idea of EMD is to remove the local median from a signal by using a sifting process. For a given signal  $f$ , a cubic spline polynomial is used to interpolate all the local maxima (minima) to obtain an upper (lower) envelope. One then averages the upper and lower envelopes to obtain an approximate median,  $m(t)$ . If  $f(t) - m(t)$  satisfies the two conditions in the definition of an EMD dictionary, then we accept  $f(t) - m(t)$  as the first IMF. If not, we treat  $f(t) - m(t)$  as a new signal and construct a new candidate for the IMF by using the same procedure described above. This sifting process continues until we obtain a satisfactory IMF, which we denote as  $f_n(t)$ . To find the remaining IMFs, we treat  $f(t) - f_n(t)$  as a new signal, and apply the same procedure to generate the second IMF,  $f_{n-1}(t)$ . This procedure continues until  $f_0(t)$  is either monotone or contains at most one extreme, see [22] for more details of the sifting process.

## 3 Data-driven time-frequency analysis

Inspired by the EMD method and the recently developed compressed sensing theory, we proposed two data-driven time-frequency analysis methods in [18, 19]. The first method uses a nonlinear basis pursuit approach, while the second one uses a nonlinear matching pursuit approach. Both methods are based on finding the sparsest decomposition of a signal by solving a nonlinear optimization problem. The nonlinear matching pursuit has the advantage of being robust to noise. It can be implemented very fast via fast Fourier transform. In this section, we will review both methods in some details.

### 3.1 A nonlinear basis pursuit approach

We first describe our nonlinear basis pursuit approach. As we mentioned before, we need to construct a large dictionary that can be used to obtain a sparse decomposition of the signal. The dictionary consists

of a collection of waveforms of the form  $a(t) \cos(\theta(t))$  with  $a(t)$  being less oscillatory than  $\cos(\theta(t))$  in some sense. To quantify this statement, we introduce a  $TV^3$  norm to measure the degree of oscillation. We say that  $a(t)$  is less oscillatory than  $b(t)$  if  $TV^3(a) \leq \lambda TV^3(b)$  for some small positive parameter  $\lambda$ , where the  $TV^n$  norm is defined as follows:

$$TV^n(g) = \int_a^b |g^{(n+1)}(x)| dx, \quad (3.1)$$

where  $g^{(n+1)}(x)$  is the  $(n+1)$ -th derivative of  $g$ . This leads to the definition of the following dictionary:

$$\mathcal{D} = \{a(t) \cos \theta(t) : \theta'(t) \geq 0, a(t) \text{ is less oscillatory than } \cos \theta(t)\}. \quad (3.2)$$

In some sense, the above dictionary can be seen as a collection of all possible IMFs, which makes our method as adaptive as the EMD method.

The choice of the dictionary  $\mathcal{D}$  is inspired by the definition of the intrinsic mode functions (IMFs) in the EMD method. The element of dictionary  $\mathcal{D}$  is a signal with amplitude modulation (AM) and frequency modulation (FM). Here  $a(t)$  plays the role of amplitude modulation. It is supposed to be less oscillatory than the corresponding  $\cos \theta(t)$  from the physical consideration. From a mathematical perspective, if this restriction is violated, we may get a trivial decomposition, such as  $a(t) = f(t)$ ,  $\theta(t) = 0$ . We remark that  $\theta'(t)$  is the instantaneous frequency of the signal. The physical meaning of frequency implies that it must be positive. Otherwise, we may set  $a(t) = \max_t |f(t)|$  and  $\theta(t) = \arccos(f(t)/a(t))$ . Obviously, we cannot obtain any useful information from this decomposition.

Since the dictionary  $\mathcal{D}$  is highly redundant, the decomposition over this dictionary is not unique. We need a criterion to select the "best" one among all possible decompositions. We assume that the data we consider have an intrinsic sparse structure in the time-frequency plane in some nonlinear and nonstationary basis. However, we do not know this basis a priori and we need to derive (or learn) this basis from the data. Based on this consideration, we adopt sparsity as our criterion to choose the best decomposition. This criterion yields the following nonlinear optimization problem:

$$(P_0) \quad \min M, \quad \text{subject to } f(t) = \sum_{k=1}^M a_k(t) \cos \theta_k(t), \quad a_k(t) \cos \theta_k(t) \in \mathcal{D}, \quad k = 1, \dots, M. \quad (3.3)$$

After this optimization problem is solved, we get a very clear time-frequency representation:

$$\text{Instantaneous frequency: } \omega_k(t) = \theta'_k(t), \quad \text{Amplitude: } a_k(t). \quad (3.4)$$

The nonlinear optimization  $(P_0)$  stated above is too difficult to solve numerically. We have introduced a recursive scheme to solve the nonlinear optimization problem  $(P_0)$  approximately.

First of all, we observe that after extracting the highest frequency IMF,  $a_1(t) \cos \theta_1(t)$ , the local median would become much less oscillatory than the original signal,  $f(t)$ . Based on this observation, we propose the following alternative method to solve the original nonlinear optimization problem: looking for  $a_1(t) \cos \theta_1(t) \in \mathcal{D}$  that gives the least oscillatory local median,  $a_0(t) = f(t) - a_1(t) \cos \theta_1(t)$ . This idea yields the following  $TV^3$ -based optimization problem:

$$(P) \quad \min TV^3(a_0) + TV^3(a_1), \quad \text{subject to } a_0(t) + a_1(t) \cos \theta(t) = f(t), \quad \theta'(t) \geq 0. \quad (3.5)$$

We note that minimizing the third order total variation of a function  $g$  tends to produce a piecewise constant approximation to the third order derivative of  $g$ . Thus our  $TV^3$ -based minimization tends to produce a cubic spline approximation for  $a_0$  and  $a_1$ . In this sense, our method shares some property similar to that of the EMD method.

In order to solve the above optimization problem, we use following Newton type of iteration:

**Initialization:**  $r^0 = f$ ,  $k = 0$ .

**While**  $\|r^k\|_2 \geq \epsilon_0$

- Give the Initial value of the phase function  $\theta_k^0$ , set  $n = 1$ .

- **While**  $\|\theta_k^n - \theta_k^{n-1}\|_2 \geq \epsilon_1$ 
  - Update  $a_0^n, a_1^n, b_1^n$  by solving the following linear optimization problem:

$$\min TV^3(a_0^n) + TV^3(a_1^n) + TV^3(b_1^n), \tag{3.6}$$

$$\text{subject to } a_0^n + a_1^n \cos \theta_k^{n-1}(t) + b_1^n \sin \theta_k^{n-1}(t) = r^k(t). \tag{3.7}$$

- Update the phase function  $\theta$ :

$$\theta_k^n = \theta_k^{n-1} - \mu \arctan \left( \frac{b_1^n}{a_1^n} \right), \tag{3.8}$$

where  $\mu \in [0, 1]$  is chosen to enforce that  $\theta^n$  is an increasing function:

$$\mu = \max \left\{ \alpha \in [0, 1] : \frac{d}{dt} \left( \theta_k^{n-1} + \alpha \arctan \left( \frac{b_1^n}{a_1^n} \right) \right) \geq 0 \right\}. \tag{3.9}$$

- **End While**
- Set  $\theta_k = \theta_k^n, a_{1,k} = a_1^n, r^{k+1} = a_0^n$ . Also set  $k = k + 1$ .

**End While**

### 3.2 A nonlinear matching pursuit approach

In our nonlinear matching approach, we choose our dictionary as follows:

$$\mathcal{D} = \{a(t) \cos \theta(t) : a(t) \text{ and } \theta'(t) \text{ are less oscillatory than } \cos \theta(t), \theta'(t) \geq 0, \forall t \in \mathbb{R}\}. \tag{3.10}$$

Let  $V(\theta, \lambda)$  be the collection of all the functions that are less oscillatory than  $\cos \theta(t)$ . Unlike our nonlinear basis pursuit approach, we quantify the degree of oscillation by constructing  $V(\theta, \lambda)$  as an overcomplete Fourier basis given below:

$$V(\theta, \lambda) = \text{span} \left\{ 1, \left( \cos \left( \frac{k\theta}{2L_\theta} \right) \right)_{1 \leq k \leq 2\lambda L_\theta}, \left( \sin \left( \frac{k\theta}{2L_\theta} \right) \right)_{1 \leq k \leq 2\lambda L_\theta} \right\}, \tag{3.11}$$

where  $L_\theta = \lfloor \frac{\theta(1) - \theta(0)}{2\pi} \rfloor$ ,  $\lfloor \mu \rfloor$  is the largest integer less than  $\mu$ , and the parameter  $\lambda \leq 1/2$  is to control the degree of oscillation of  $V(\theta, \lambda)$ . In our computations, we typically choose  $\lambda \leq 1/2$ . The dictionary  $\mathcal{D}$  can be written as

$$\mathcal{D} = \{a(t) \cos \theta(t) : a(t) \in V(\theta, \lambda), \theta'(t) \in V(\theta, \lambda), \text{ and } \theta'(t) \geq 0, \forall t \in \mathbb{R}\}. \tag{3.12}$$

The dictionary  $\mathcal{D}$  used in this approach is essentially similar to the dictionary we used in the previous section. The difference is that we explicitly construct a linear space  $V(\theta)$  which is dependent on the phase function  $\theta$ , to quantify the degree of the oscillation. From a mathematical perspective, the dictionary given in (3.2) is not well defined, since the term “less oscillatory” does not have a rigorous definition. To remove this ambiguity, we give the terminology “less oscillatory” a rigorous definition by constructing the space  $V(\theta)$  and defining that “ $a(t)$  is less oscillatory than  $\cos \theta(t)$ ” if and only if  $a(t) \in V(\theta)$ .

The space  $V(\theta)$  consists of the functions which are less oscillatory than  $\cos \theta(t)$ . The idea to construct  $V(\theta)$  is based on two observations:

- (1)  $\cos \lambda t, \sin \lambda t$  are less oscillatory than  $\cos t$  as long as  $\lambda < 1$ ;
- (2)  $\theta'(t) > 0$  which implies that  $\theta(t)$  can be used as a coordinate.

In our definition, we require that  $\lambda \leq 1/2$ . This parameter can be tuned in different problems to control the degree of oscillation of  $V(\theta)$ .

$V(\theta)$  may also be constructed by the wavelet transform instead of the Fourier modes used above. For more details, we refer to [20].

As before, we adopt sparsity as our criterion to choose the best decomposition among all possible decompositions. This criterion yields the following nonlinear optimization problem:

$$(P) \quad \min_{(a_k)_{1 \leq k \leq M}, (\theta_k)_{1 \leq k \leq M}} M, \quad \text{subject to } f = \sum_{k=1}^M a_k \cos \theta_k, \quad a_k \cos \theta_k \in \mathcal{D}, \quad k = 1, \dots, M, \tag{3.13}$$

or

$$(P_\delta) \quad \min_{(a_k)_{1 \leq k \leq M}, (\theta_k)_{1 \leq k \leq M}} M, \quad \text{subject to} \quad \left\| f - \sum_{k=1}^M a_k \cos \theta_k \right\|_{l^2} \leq \delta, \quad a_k \cos \theta_k \in \mathcal{D}, \quad k = 1, \dots, M, \quad (3.14)$$

if the signal has noise with noise level  $\delta$ .

The above optimization problem can be seen as a nonlinear  $l^0$  minimization problem. Matching pursuit has been shown to be a powerful method to solve the  $l^0$  minimization problem. We have proposed a nonlinear matching pursuit method to solve this optimization problem. Since we use the overcomplete Fourier basis to construct  $V(\theta, \lambda)$ , the above nonlinear least square problem may be ill-conditioned. Moreover, the simple least square method would introduce severe interference among different IMFs. In order to stabilize the above optimization problem and remove the interference, we proposed to add an  $l^1$  term to regularize the nonlinear least square problem. This would give us the following algorithm based on the  $l^1$ -regularized nonlinear least square:

- $r_0 = f, k = 1$ .

**Step 1.** Solve the following  $l^1$ -regularized nonlinear least-square problem ( $P_2$ ):

$$(P_2) \quad (a_k, \theta_k) \in \underset{a, \theta}{\text{Argmin}} \quad \gamma \|\widehat{a}\|_{l^1} + \|r_{k-1} - a \cos \theta\|_{l^2}^2, \quad (3.15)$$

subject to  $a \in V(\theta, \lambda), \theta' \geq 0, \forall t \in \mathbb{R}$ ,

where  $\gamma > 0$  is a regularization parameter and  $\widehat{a}$  is the representation of  $a$  in the overcomplete Fourier basis previously detailed in (3.11).

**Step 2.** Update the residual

$$r_k = f - \sum_{j=1}^k a_j \cos \theta_j. \quad (3.16)$$

**Step 3.** If  $\|r_k\|_{l^2} < \epsilon_0$ , stop. Otherwise, set  $k = k + 1$  and go to Step 1.

If signals are periodic, we can use the standard Fourier basis to construct  $V(\theta, \lambda)$  instead of the overcomplete Fourier basis. The  $l^1$  regularization term is not needed (i.e., we can set  $\gamma = 0$ ) since the standard Fourier basis are orthogonal to each other. In the next section, we will use this property to further simplify the above algorithm for periodic signals.

### 3.2.1 An $l^1$ -regularized nonlinear least square solver

Our  $l^1$ -regularized nonlinear least square problem is non-convex since the basis is not known a priori. We need to find the basis and the decomposition simultaneously. In the following, we propose a Gauss-Newton type method to solve the  $l^1$ -regularized nonlinear least square problem. In our iterative method, we gradually enlarge the space  $V(\theta, \eta)$  to update  $\theta'$  by increasing  $\eta$  during the iterations. Note that  $V(\theta, \eta = 0)$  for the first iteration consists only constants while  $V(\theta, \eta = \lambda)$  during the final iteration is the space in which our nonlinear optimization is defined. This procedure makes our method very robust and is insensitive to our initial condition. Throughout our computations, we choose  $\lambda = 1/2$  and the increment  $\Delta\eta = \lambda/20$ .

- $\theta_k^0 = \theta_0, \eta = 0$ .

**Step 1.** Solve the following  $l^1$ -regularized least-square problem:

$$(P_{2,l^2}) \quad (a_k^{n+1}, b_k^{n+1}) \in \underset{a, b}{\text{Argmin}} \quad \gamma (\|\widehat{a}\|_{l^1} + \|\widehat{b}\|_{l^1}) + \|r_{k-1} - a \cos \theta_k^n - b \sin \theta_k^n\|_{l^2}^2,$$

subject to  $a \in V(\theta_k^n, \lambda), b \in V(\theta_k^n, \lambda)$ ,

where  $\widehat{a}, \widehat{b}$  are the representations of  $a, b$  in the  $V(\theta_k^n, \lambda)$  space.

**Step 2.** Update  $\theta_k^n$ :



$$\Delta\theta' = P_{V(\theta_k^n; \eta)} \left( \frac{d}{dt} \left( \arctan \left( \frac{b_k^{n+1}}{a_k^{n+1}} \right) \right) \right), \quad \Delta\theta = \int_0^t \Delta\theta'(s) ds, \quad \theta_k^{n+1} = \theta_k^n - \beta \Delta\theta, \quad (3.17)$$

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

$$\beta = \max \left\{ \alpha \in [0, 1] : \frac{d}{dt} (\theta_k^n - \alpha \Delta\theta) \geq 0 \right\}, \quad (3.18)$$

where  $P_{V(\theta_k^n; \eta)}$  is the projection operator to the space  $V(\theta_k^n; \eta)$  and  $V(\theta_k^n; \eta)$  is the space defined in (3.11).

**Step 3.** If  $\|\theta_k^{n+1} - \theta_k^n\|_2 > \epsilon_0$ , set  $n = n + 1$  and go to Step 1. Otherwise, go to Step 4.

**Step 4.** If  $\eta \geq \lambda$ , stop. Otherwise, set  $\eta = \eta + \Delta\eta$  and go to Step 1.  $\lambda$  is the parameter we choose in (3.11).

### 3.2.2 A fast algorithm for periodic data

In the iterative algorithm described in the previous subsection, we need to solve an  $l^1$ -regularized least square problem in each step. This is the most expensive part of the algorithm especially when the number of the data points is large. In this subsection, we introduce a method based on fast Fourier transform (FFT) for periodic data.

An important advantage of our algorithm for periodic data is that we can use a standard Fourier basis to construct the  $V(\theta, \lambda)$  space instead of the overcomplete Fourier basis given in (3.11).

$$V_p(\theta, \lambda) = \text{span} \left\{ 1, \left( \cos \left( \frac{k\theta}{L_\theta} \right) \right)_{1 \leq k \leq \lambda L_\theta}, \left( \sin \left( \frac{k\theta}{L_\theta} \right) \right)_{1 \leq k \leq \lambda L_\theta} \right\}, \quad (3.19)$$

where  $\lambda \leq 1/2$  is a parameter to control the degree of oscillation of functions in  $V_p(\theta, \lambda)$  and  $L_\theta = (\theta(T) - \theta(0))/2\pi$  is a positive integer. We use the subscript  $p$  to denote this space is for the periodic signal. Since the standard Fourier basis is an orthogonal basis, the  $l^1$ -regularized term is not necessary in our nonlinear optimization. As a result, the nonlinear least-square problem is reduced to the following optimization problem (by setting  $\gamma = 0$ ):

$$\min_{a, b} \|r_k - a \cos \theta_k^n - b \sin \theta_k^n\|_{l^2}^2, \quad \text{subject to } a, b \in V_p(\theta_k^n, \lambda). \quad (3.20)$$

Notice that in the iterative process, the derivative of the phase function  $\theta_k^n$  is always monotonically increasing. Thus, we can use  $\theta_k^n$  as a new coordinate. In this new coordinate,  $\cos \theta_k^n$ ,  $\sin \theta_k^n$  and the bases of  $V_p(\theta_k^n, \lambda)$  are simple Fourier modes, then the least-square problem can be solved by using the fast Fourier transform. For more details, we refer to our paper [19].

There are two important advantages of this nonlinear matching pursuit approach. The first one is that this method is very stable to noise perturbation. The second one is that it can be implemented very efficiently. For periodic data, our method can be solved by FFT which gives a complexity of order  $O(N \log N)$ , where  $N$  is the number of data sample points that we use to represent the signal. The low computational cost and the robustness to noise perturbation make this method very effective in many applications. Moreover, for data that satisfy certain scale separation conditions, we prove that our method recovers the IMFs and their instantaneous frequencies accurately.

## 4 Convergence of nonlinear matching pursuit for periodic data

In a recent paper by Hou et al. [21], we have proved convergence and stability of the nonlinear matching algorithms proposed in the previous section for periodic data. We summarize our main findings in this section.

### 4.1 Convergence analysis for well-resolved signals

We first focus on well-resolved signals. By well-resolved signals, we mean that these signals are measured over a uniform grid and can be interpolated to any grid with very little loss of accuracy. 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  $[0, 1]$ . We consider a periodic signal  $f(t)$  that has the following decomposition:

$$f(t) = f_0(t) + f_1(t) \cos \theta(t), \quad f_1(t) > 0, \quad \theta'(t) > 0, \quad t \in [0, 1], \quad (4.1)$$

where  $f_0, f_1$  and  $\theta$  are the exact local mean, the envelope and the phase function that we want to recover from the signal, respectively.

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

$$\hat{f}_{0,\theta}(k) = \int_0^1 f_0 e^{-i2\pi k \bar{\theta}} d\bar{\theta}, \quad \hat{f}_{1,\theta}(k) = \int_0^1 f_1 e^{-i2\pi k \bar{\theta}} d\bar{\theta}. \quad (4.2)$$

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

Now we can state the theorem as follows:

**Theorem 4.1.** *Assume that the instantaneous frequency  $\theta'$  is  $M_0$ -sparse over the Fourier basis in the physical space, i.e.,*

$$\theta' \in V_{M_0} = \text{span}\{e^{i2k\pi t/T}, k = -M_0, \dots, 1, \dots, M_0\}. \quad (4.3)$$

Furthermore, we assume that the local mean  $f_0$  and the envelope  $f_1$  are  $M_1$ -sparse over the Fourier basis in the  $\bar{\theta}$ -space, i.e.,

$$\hat{f}_{0,\theta}(k) = \hat{f}_{1,\theta}(k) = 0, \quad \forall |k| > M_1. \quad (4.4)$$

If the initial guess of  $\theta^0$  satisfies

$$\|\mathcal{F}((\theta^0 - \theta)')\|_1 \leq \pi M_0/2, \quad (4.5)$$

then there exists  $\eta_0 > 0$  such that

$$\|\mathcal{F}((\theta^{m+1} - \theta)')\|_1 \leq \frac{1}{2} \|\mathcal{F}((\theta^m - \theta)')\|_1, \quad (4.6)$$

provided that  $L \geq \eta_0$ .

**Remark 4.2.** We remark that 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. Theorem 4.1 shows that our data-driven time-frequency analysis method has the capability to recover the exact instantaneous frequency for a larger range of signals.

If the signal does not have an exact sparsity structure in the  $\theta$ -space as required by Theorem 4.1, our method cannot reproduce the exact decomposition. But our analysis 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 4.3.** *Assume that the instantaneous frequency  $\theta'$ , has a sparse representation, i.e., there exists  $M_0$ , such that*

$$\theta'(t) \in V_{M_0} = \text{span}\{e^{i2k\pi t/T}, k = -M_0, \dots, 1, \dots, M_0\}, \quad (4.7)$$

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

$$|\hat{f}_{0,\theta}(k)| \leq C_0 |k|^{-p}, \quad |\hat{f}_{1,\theta}(k)| \leq C_0 |k|^{-p}. \quad (4.8)$$

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

$$\|\mathcal{F}((\theta^0 - \theta)')\|_1 \leq \pi M_0/2, \tag{4.9}$$

then we have

$$\|\mathcal{F}((\theta^{m+1} - \theta)')\|_1 \leq \Gamma_0(L/4)^{-p+2} + \frac{1}{2}\|\mathcal{F}((\theta^m - \theta)')\|_1, \tag{4.10}$$

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

This theorem shows that our iterative method will converge to the exact solution up to the truncation error determined by the scale separation property. We further consider a more general case: the instantaneous frequency is also approximately sparse instead of exactly sparse as we assume in Theorems 4.1 and 4.3. In this case, we can prove that the iterative algorithm also converges to an approximate result.

### 4.2 Convergence analysis for sparsely sampled signals

In this subsection, we consider a more challenging case, the sample points are too few to resolve the signal. In this case, the algorithm presented in last section does not apply directly. The reason is that the Fourier transform in the  $\theta^m$ -space,  $\mathcal{F}_{\theta^m}(\cdot)$ , cannot be computed accurately by the FFT-based method. Thanks to the recent development 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 an  $l^1$  minimization problem to generate the Fourier coefficients in the  $\theta^m$ -space.

Suppose the sample points  $t_j, j = 1, \dots, N_s$  are selected at random from a set of uniform grid  $l/N_f, l = 0, \dots, N_f - 1$ , we solve the following  $l_1$  minimization problem to get the Fourier transform of the signal  $f$  in the  $\theta^m$ -coordinate:

$$\min \|x\|_1, \quad \text{subject to } \Phi_{\theta^m} \cdot x = \tilde{f}, \tag{4.11}$$

where  $\tilde{f} = D_{\theta^m} \cdot f$  and  $D_{\theta^m} = \text{diag}(\sqrt{\frac{(\bar{\theta}^m)'}{N_f}})$  is an  $N_s \times N_s$  diagonal matrix,  $\Phi_{\theta^m} = D_{\theta^m} \cdot \Psi_{\theta^m}$  and  $\Psi_{\theta^m}$  is an  $N_s \times N_b$  matrix which is obtained by selecting  $N_s$  rows from an  $N_f \times N_b$  matrix  $U_{\theta^m}$ ,  $N_b$  is the number of Fourier modes used in the algorithm which has to be determined a priori,  $U_{\theta^m}$  is a matrix whose columns are Fourier modes in  $\theta^m$ -space. More specifically,  $(j, k)$ -th entry of  $U_{\theta^m}$  is

$$U_{\theta^m}(j, k) = e^{i2\pi k \bar{\theta}^m(t_j)}, \quad j = 1, \dots, N_s, \quad k = -N_b/2 + 1, \dots, N_b/2,$$

$\bar{\theta}^m = \frac{\theta^m - \theta^m(0)}{\theta^m(T) - \theta^m(0)}$  is the normalized phase function.

Notice that the columns of the matrix  $U_{\theta^m}$  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  $\theta$  over which the signal has the sparsest representation.

**Theorem 4.4.** Under the same assumption as in Theorem 4.1, there exist  $\eta_0 > 0, \eta_1 > 0$ , such that

$$\|\mathcal{F}((\theta^{m+1} - \theta)')\|_1 \leq \frac{1}{2}\|\mathcal{F}((\theta^m - \theta)')\|_1, \tag{4.12}$$

provided that  $L \geq \eta_0$  and  $S \geq \eta_1$ , where  $S$  is 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 [6], which is the smallest number such that  $(1 - \delta_S)\|c\|_{l_2}^2 \leq \|A_T c\|_{l_2}^2 \leq (1 + \delta_S)\|c\|_{l_2}^2$ , for all subsets  $T$  with  $|T| \leq S$  and coefficients sequences  $(c_j)_{j \in T}$ .

In Theorem 4.4, 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  $\delta_S$ , it is easy to see that  $\delta_{3S} \leq \delta_{4S}$ . 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 [7]. 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  $U = (U_{j,k})$  and  $U_{j,k} = \frac{1}{\sqrt{N}}e^{i2\pi jk/N}, j, k = 1, \dots, N$ , then the condition  $\delta_S(\Phi) < 1/2$  is satisfied with an overwhelming probability provided that  $S \leq C \frac{M}{(\log N)^6}$ , where  $C$  is a constant.

In our formulation, the matrix  $\Phi_{\theta^m}$  also consists of  $N_s$  rows of an  $N_f \times N_b$  matrix  $U_{\theta^m}$ . The main difference is that the matrix  $U_{\theta^m}$  is not a standard Fourier matrix. Instead it is a Fourier matrix in the  $\theta^m$ -space which makes it non-orthonormal. As a result, we cannot apply the result of Candes and Tao in [7] directly. Fortunately, we have been able to prove a generalized result which can be applied to matrix  $U_{\theta^m}$ . This generalized result basically shows that if the columns of  $U_{\theta^m}$  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  $U_{\theta^m}$  for  $\theta^m \in V_{M_0}$ . Using this result, we can show that the condition  $\nu_0 = \max_{k,j} |U_{\theta^m}^* U_{\theta^m} - I)_{k,j}| \leq \frac{1}{16N_b}$  is satisfied as long as  $N_f \geq C \|\mathcal{F}((\bar{\theta}^m)')\|_1 N_b$ , where  $C$  is a constant determined by  $N_b$ . Our analysis shows that if the sample points are selected at random, in each step, with probability  $1 - \delta$ , we can get the right answer.

## 5 Numerical results

In this section, we present several numerical results to demonstrate the effectiveness of our data-driven time-frequency analysis methods. We mainly focus on the nonlinear matching pursuit method. Numerical examples for the nonlinear basis pursuit method can be found in [18]. First we will present numerical results for the FFT-based algorithms for periodic data or data with a good scale separation property. We also present numerical results for non-periodic signals to demonstrate the effectiveness of our  $l^1$ -regularized least square algorithm. Finally, we present some numerical results for periodic data that validate our convergence results presented in the previous section.

### 5.1 Numerical results for the FFT-based algorithms

In this subsection, we present a number of numerical experiments to demonstrate the accuracy and robustness of our FFT-based algorithms. Throughout this section, we denote  $X(t)$  as white noise with zero mean and variance  $\sigma^2 = 1$ . The signal-to-noise ratio (SNR, measured in dB) is defined by

$$\text{SNR}[\text{dB}] = 10 \log_{10} \left( \frac{\text{var} f}{\sigma^2} \right). \quad (5.1)$$

We will apply our data-driven time-frequency analysis method to several different signals with increasing level of difficulty.

**Example 1.** First, we consider a signal that consists of three IMFs [19]. The signal is given by the following analytic formula:

$$f(t) = \frac{1}{1.5 + \cos(2\pi t)} \cos(60\pi t + 15 \sin(2\pi t)) + \frac{1}{1.5 + \sin(2\pi t)} \cos(160\pi t + \sin(16\pi t)) \\ + (2 + \cos(8\pi t)) \cos(140\pi(t+1)^2). \quad (5.2)$$

In Figure 1, we study the accuracy of the instantaneous frequencies obtained by our method with the exact instantaneous frequencies. The upper row corresponds to the signal without noise. As we can see, it is hard to tell any hidden structure from this signal even without noise. Our method recovers the three components of the instantaneous frequencies (blue) that match the exact instantaneous frequencies (red) very well. In the case when noise is added to the original signal, we cannot recognize any hidden pattern from the polluted signal. It is remarkable that our method could still recover the three components of the instantaneous frequencies with accuracy comparable to the noise level, see the bottom row of Figure 1.

**Example 2** (Length-of-day data). Next, we apply our method to the length-of-day data [19], see Figure 2. The data we adopt here was produced by Gross [17], covering the period from January 20, 1962 to January 6, 2001, for a total of 14,232 days (approximate 39 years). Figure 3 displays the first 5 IMFs extracted by the FFT-based method. These IMFs are ordered by their frequencies from high to low. Each component is enforced to be an IMF by the construction of our dictionary. As a result, we do not need to do shifting or post-processing as was done in the EMD or the EEMD method [44]. The IMFs that we obtained match qualitatively those obtained by EEMD with post-processing [44]. Furthermore,

we note that the results obtained by our method do not suffer from the mode mixing phenomenon that is presented in the EMD decomposition.

We observe that each IMF we obtained has a clear physical interpretation. For example, the period of  $C_1$  is around 14 days, corresponding to the semi-monthly tides. The period of  $C_2$  is about 28 days, corresponding to the monthly tides. Similarly, the period of  $C_4$  is about half a year, corresponding to the semi-annual cycle and  $C_5$  corresponds to the annual cycle.

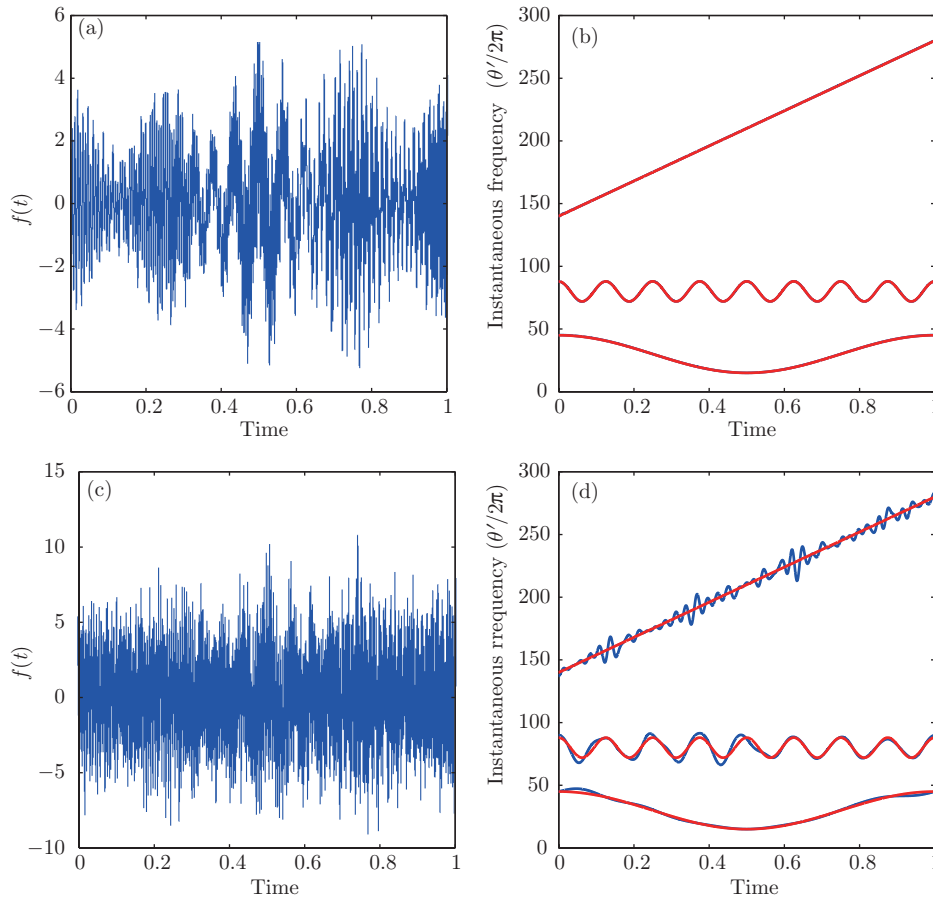
### 5.2 Numerical results for the $l^1$ -regularized nonlinear matching pursuit

The nonlinear matching pursuit method based on the fast Fourier transform does not work well for non-periodic data or data with poor scale separation. For this kind of data, the FFT-based method tends to produce some oscillations near the boundary due to the use of the standard Fourier basis. This so-called “end effect” is also present in the EMD method and other data analysis methods. To remove this “end effect” error, we need to use the  $l^1$ -regularized nonlinear matching pursued described in Subsection 3.2 with  $V(\theta, \lambda)$  being the overcomplete Fourier basis defined in (3.11).

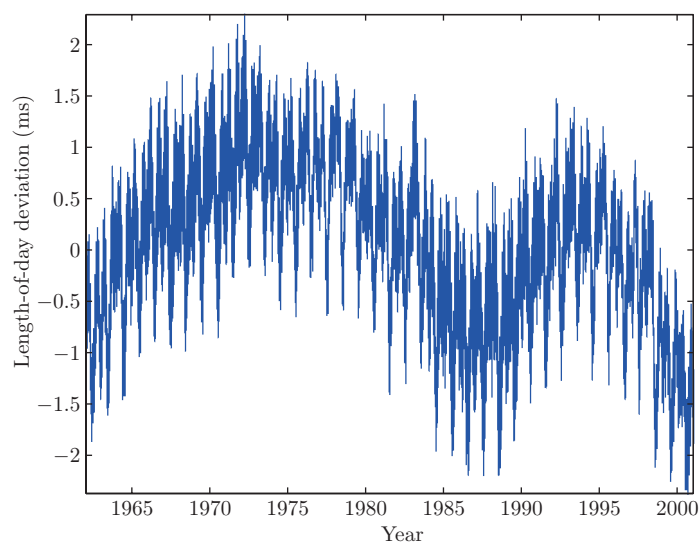
To test the performance of our  $l^1$ -regularized nonlinear matching pursuit, we apply our method to the following data [19]:

$$\theta_1 = 20\pi(t + 1)^2 + 1, \quad \theta_2 = 161.4\pi t + 4(1 - t)^2 \sin(16\pi t),$$

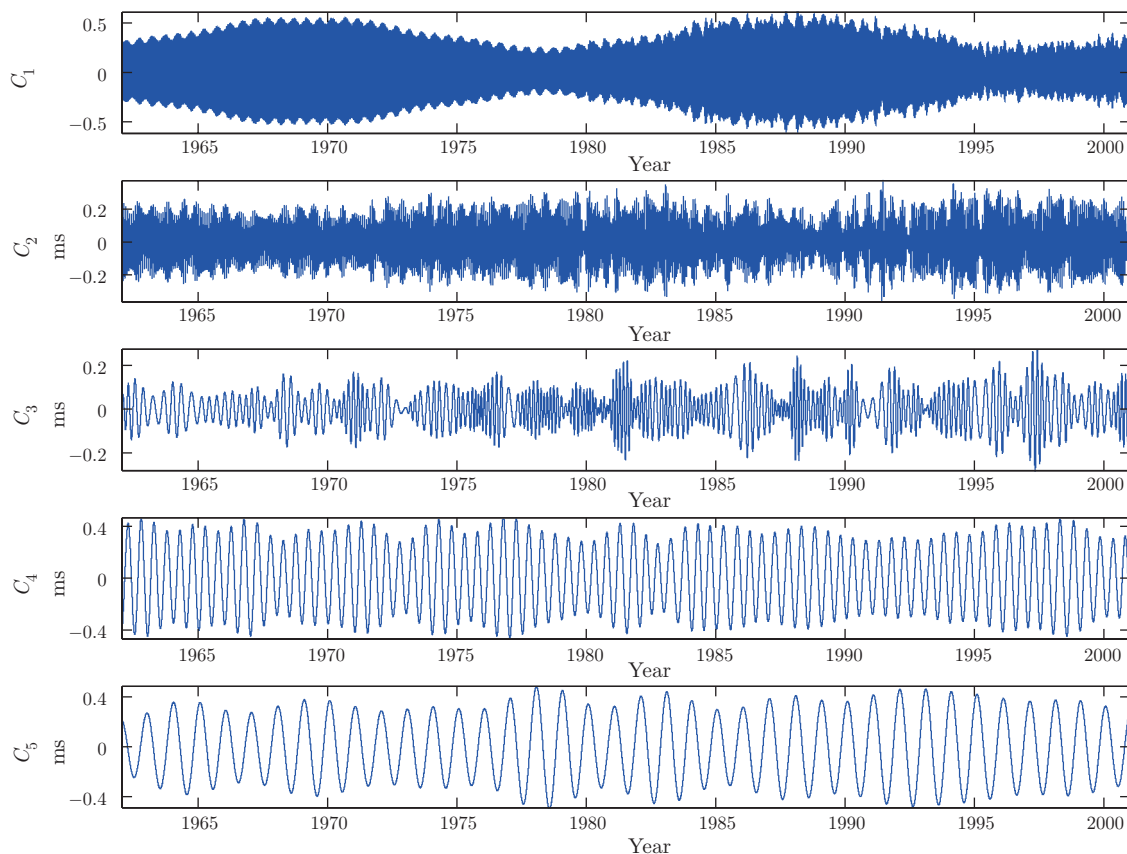
$$f(t) = \frac{1}{1.5 + \sin(1.5\pi t)} + (2t + 1) \cos \theta_1 + (2 - t)^2 \cos \theta_2.$$



**Figure 1** Upper row: (a) The signal defined in (5.2) without noise; (b) Instantaneous frequencies. Red: exact frequencies; blue: numerical results. Lower row: the same as the upper row except that white noise  $2X(t)$  was added to the original signal, the corresponding SNR is  $-0.8$  dB



**Figure 2** The daily length-of-day data from January 20, 1962 to January 6, 2001



**Figure 3** The first 5 IMFs with highest frequencies given by our FFT-based method

In this numerical example, the  $l^1$ -regularizing parameter  $\gamma$  is chosen to be 1. From Figure 4, we observe that the  $l^1$ -regularized nonlinear matching pursuit seems to produce considerably smaller error (see the blue curve) near the boundary compared with that produced by the FFT-based algorithm (see the black curve).

This example shows that the  $l^1$ -regularized nonlinear matching pursuit can be used to decompose non-periodic data with reasonable accuracy. On the other hand, the computational cost of the  $l^1$ -regularized

nonlinear matching pursuit is considerably higher than that of the FFT-based algorithm due to the extra cost of solving an  $l^1$ -regularized least square problem in each iteration. There are several ways to speed up the  $l^1$  optimization. A more efficient  $l^1$  solver could reduce the computational cost significantly. One effective way to reduce the computational cost is to use a hybrid method by applying the FFT-based algorithm in the majority of the interior domain and using the  $l^1$  minimization only near the boundary of the signal. This would lead to considerable saving.

### 5.3 Numerical validation of convergence results

In this subsection, we will perform several numerical experiments to confirm our theoretical results presented in the previous section.

**Example 3** (Exact recovery for a well-resolved signal). The first example is a well-resolved periodic signal [21]. In this example, the mean and the envelope have a sparse Fourier representation in the  $\theta$ -space and the instantaneous frequency has a sparse Fourier spectrum in the physical space. The signal we use is generated by the following formulae:

$$\begin{aligned} \theta &= 20\pi t + 2 \cos 2\pi t + 2 \sin 4\pi t, & \bar{\theta} &= \theta/10, \\ a_0 &= 2 + \cos \bar{\theta} + 2 \sin 2\bar{\theta} + \cos 3\bar{\theta}, & a_1 &= 3 + \cos \bar{\theta} + \sin 3\bar{\theta}, & f &= a_0 + a_1 \cos \theta. \end{aligned}$$

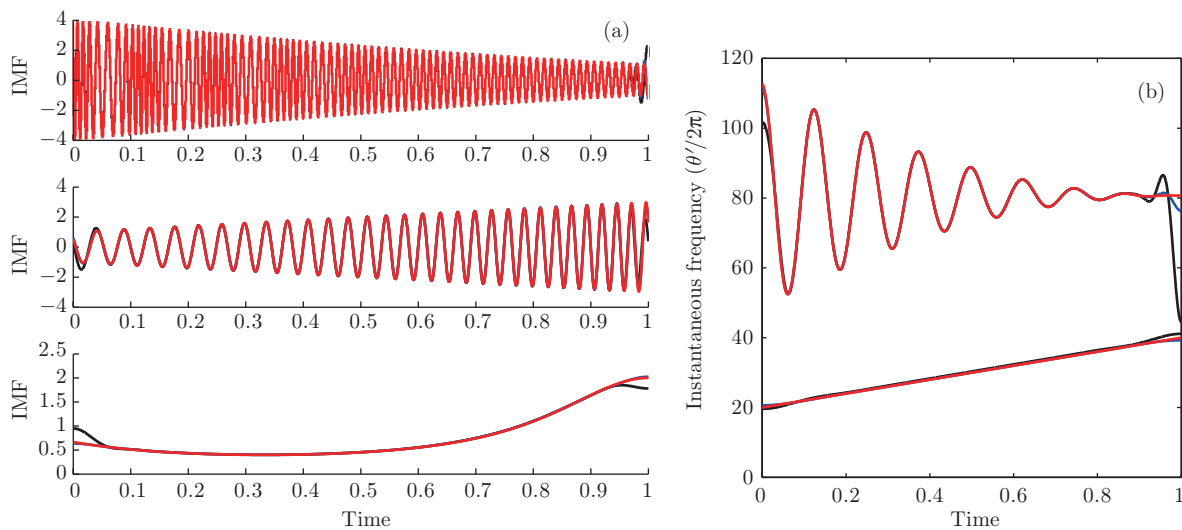
This signal is sampled over a uniform mesh of 256 points. There are about 12 samples in each period of the signal on average.

The numerical results are shown in Figure 5. In Figure 5, 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 4.1.

**Example 4** (Exact recovery for a signal with sparse random samples). The second example is designed to confirm the result of Theorem 4.4 [21]. 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 in (5.3),

$$\begin{aligned} \theta &= 200\pi t - 10 \cos 2\pi t - 2 \sin 4\pi t, & \bar{\theta} &= \theta/(100), \\ a_0 &= \cos \bar{\theta}, & a_1 &= 3 + \cos \bar{\theta} + \sin 2\bar{\theta}, & f &= a_0 + a_1 \cos \theta. \end{aligned}$$

The number of sample points is equal to 120. These sample points are selected at random over 4096 uniformly distributed points. On average, there are about 1.2 points in each period of the signal. We



**Figure 4** IMF (a) and instantaneous frequency (b) of the signal in (5.3) obtained from different methods. Red: exact; blue:  $l^1$  regularized nonlinear matching pursuit; black: FFT-based algorithm.

test 100 independent samples and our algorithm is able to recover the signal for 97 samples, which gives 97% success rate. Figure 6 gives one of the successful samples.

The right panel of Figure 6 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. By increasing the accuracy in solving the  $l^1$  optimization problem, we can increase the accuracy for both the IMF and the phase function, which confirms our convergence result. However, by imposing a higher accuracy in solving our optimization problem, we also increase the computational cost as a consequence. From practical viewpoint, controlling the error below  $10^{-2}$  seems to be sufficient.

## 6 Concluding remarks

In this paper, we have reviewed two recently introduced data-driven time-frequency analysis methods for analyzing nonlinear and nonstationary data. These methods are based on looking for a sparse representation over a highly redundant time-frequency dictionary. The adaptivity of the decomposition is obtained by looking for the sparsest representation of a signal in the time-frequency domain from a largest possible dictionary that consists of all possible candidates for intrinsic mode functions (IMFs). Solving this nonlinear optimization problem is in general very difficult. We have proposed two methods to solve this nonlinear optimization problem. The first method is based on nonlinear basis pursuit by minimizing the  $TV^3$  norm of the local mean and the envelope function. The second method is based on nonlinear matching pursuit by generalizing matching pursuit for solving the  $l^0$  optimization problem. One important advantage of our nonlinear matching pursuit method is it can be implemented very efficiently. Moreover, this approach is very stable to noise. It can be applied to sparsely sampled data or even incomplete data. Our data-driven methods can be used to extract some hidden pattern or physical information from the data while traditional data analysis methods tend to introduce many artificial harmonics or suffer from interference of different components. Applications of our data-driven time-frequency analysis methods to some real world data have led to some new scientific discoveries [32, 40].

We have also carried out convergence analysis for the nonlinear matching pursuit method. In the case when the signal satisfies certain scale separation conditions, we have shown that our iterative algorithm converges and gives exact recovery of the IMF and the instantaneous frequency. When the scale separation condition is not satisfied exactly, we have shown that our method still gives an approximate decomposition with the accuracy determined by the scale separation factor of the signal. This remains true even for

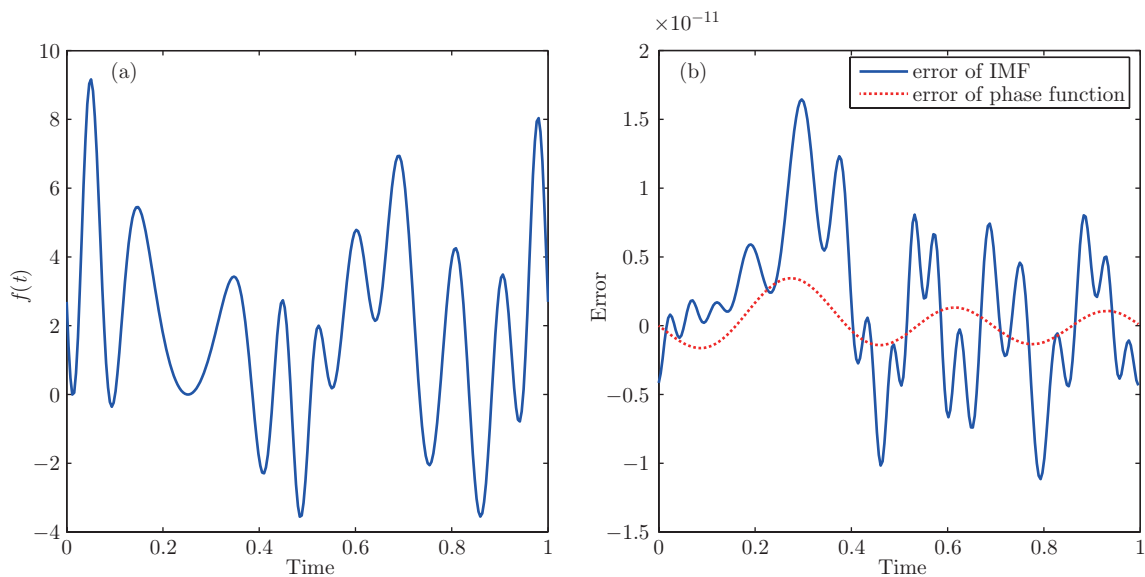
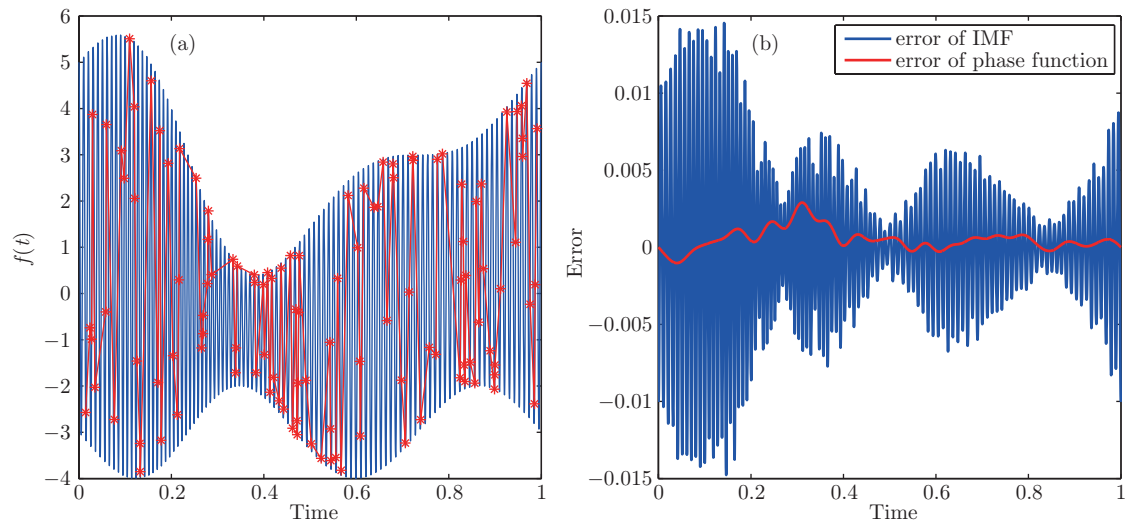


Figure 5 (a) Original signal; (b) Error of the IMF and the phase function





**Figure 6** (a) Original signal and the sample points; (b) Error of the IMF and phase function

sparsely sampled data as long as we generate the sampling points at random.

There are some remaining issues to be studied in the future. Currently, our method requires the data to satisfy certain scale separation property. In real world applications, the physical data may not satisfy the scale separation condition required by our convergence analysis. It is desirable to design an improved algorithm that can handle data with poor scale separation property and reduce the “end effect” of the data. Another important problem is to decompose data with intra-wave frequency modulation. This type of data is known to be very challenging. Naive application of traditional data analysis methods tends to introduce many harmonics. Recently, we have made some progress in decomposing this type of data. A key idea is to use a more general shape function [43] instead of using  $\cos(\theta)$  in our dictionary. We will further propose a new method to look for the appropriate shape function that gives rise to the sparsest decomposition of the signal. This work will be reported in our future publication.

**Acknowledgements** This work was supported by Air Force Office of Scientific Research, Multidisciplinary University Research Initiative, USA (Grant No. FA9550-09-1-0613), Department of Energy of USA (Grant No. DE-FG02-06ER25727), Natural Science Foundation of USA (Grant No. DMS-0908546) and National Natural Science Foundation of China (Grant No. 11201257). The authors would like to thank Professors Norden E. Huang and Zhaohua Wu for many stimulating discussions on EMD/EEMD and topics related to the research presented here. The authors would also like to thank Professors Ingrid Daubechies, Stanley Osher, and Zuwei Shen for their interest in this work and for a number of valuable discussions.

## References

- 1 Bach F R, Jenatton R, Mairal J, et al. Optimization with sparsity-inducing penalties. *Found Trends Mach Learn*, 2012, 4: 1–106
- 2 Boashash B. *Time-Frequency Signal Analysis: Methods and Applications*. Melbourne/New York: Longman-Cheshire/John Wiley, 1992
- 3 Bruckstein A M, Donoho D L, Elad M. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev*, 2009, 51: 34–81
- 4 Candès E, Romberg J, Tao T. Robust uncertainty principles: Exact signal recovery from highly incomplete frequency information. *IEEE Trans Inform Theory*, 2006, 52: 489–509
- 5 Candès E, Romberg J, Tao T. Stable signal recovery from incomplete and inaccurate measurements. *Comm Pure Appl Math*, 2006, 59: 1207–1223
- 6 Candès E, Tao T. Decoding by linear programming. *IEEE Trans Inform Theory*, 2005, 51: 4203–4215
- 7 Candès E, Tao T. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans Inform Theory*, 2006, 52: 5406–5425
- 8 Chen S, Donoho D, Saunders M. Atomic decomposition by basis pursuit. *SIAM J Sci Comput*, 1998, 20: 33–61
- 9 Combettes P L, Pesquet J C. Proximal splitting methods in signal processing. In: Bauschke H H, Burachik R,

- Combettes P L, et al., eds. Fixed-point Algorithms for Inverse Problems in Science and Engineering. New York: Springer-Verlag, 2010, 185–212
- 10 Daubechies I. Ten Lectures on Wavelets. CBMS-NSF Regional Conference Series on Applied Mathematics, vol. 61. Philadelphia: SIAM, 1992
  - 11 Daubechies I, Lu J, Wu H. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Appl Comput Harmon Anal*, 2011 30: 243–261
  - 12 Donoho D L. Compressed sensing. *IEEE Trans Inform Theory*, 2006, 52: 1289–1306
  - 13 Flandrin P. Time-Frequency/Time-Scale Analysis. San Diego, CA: Academic Press, 1999
  - 14 Gabor D. Theory of communication. *J Inst Elect Eng*, 1946, 93: 429–457
  - 15 Goldstein T, Osher S. The split Bregman method for  $L_1$ -regularized problems. *SIAM J Imaging Sci*, 2009, 2: 323–343
  - 16 Gribonval R, Nielsen M. Sparse representations in unions of bases. *IEEE Trans Inform Theory*, 2003, 49: 3320–3325
  - 17 Gross R S. Combinations of Earth Orientation Measurements: SPACE2000, COMB2000, and POLE2000. JPL Publication 01-2. Pasadena, CA: Jet Propulsion Laboratory, 2001
  - 18 Hou T Y, Shi Z. Adaptive data analysis via sparse time-frequency representation. *Adv Adapt Data Anal*, 2011, 3: 1–28
  - 19 Hou T Y, Shi Z. Data-drive time-frequency analysis. *Appl Comput Harmon Anal*, 2013, 35: 284–308
  - 20 Hou T Y, Shi Z. Sparse time-frequency decomposition by adaptive basis pursuit. Preprint
  - 21 Hou T Y, Shi Z, Tavallali P. Convergence of a data-driven time-frequency analysis method. *Appl Comput Harmon Anal*, submitted, 2013
  - 22 Huang N E, Shen Z, Long S R, 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*, 1998, 454: 903–995
  - 23 James D L, Parks T W. A high resolution data-adaptive time-frequency representation. *IEEE Trans Acoust Speech Signal Process*, 1990, 38: 2127–2135
  - 24 Loughlin P J, Tracer B. On the amplitude- and frequency-modulation decomposition of signals. *J Acoust Soc Amer*, 1996, 100: 1594–1601
  - 25 Lovell B C, Williamson R C, Boashash B. The relationship between instantaneous frequency and time-frequency representations. *IEEE Trans Signal Process*, 1993, 41: 1458–1461
  - 26 Mallat S. A Wavelet Tour of Signal Processing: the Sparse Way. New York: Academic Press, 2009
  - 27 Mallat S, Zhang Z. Matching pursuit with time-frequency dictionaries. *IEEE Trans Signal Process*, 1993, 41: 3397–3415
  - 28 Meville W K. Wave modulation and breakdown. *J Fluid Mech*, 1983, 128: 489–506
  - 29 Needell D, Tropp J. CoSaMP: Iterative signal recovery from noisy samples. *Appl Comput Harmon Anal*, 2008, 26: 301–321
  - 30 Needell D, Vershynin R. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *J Found Comput Math*, 2009, 9: 317–334
  - 31 Olhede S, Walden A T. The Hilbert spectrum via wavelet projections. *Proc R Soc Lond Ser A Math Phys Eng Sci*, 2004, 460: 955–975
  - 32 Pahlevan N M, Tavallali P, Petrasek D, et al. Systems science and a novel insight into cardiovascular physiology. *Science*, submitted, 2013
  - 33 Pesquet J C, Pustelnik N. A parallel inertial proximal optimization method. *Pacific J Optim*, 2012, 8: 273–305
  - 34 Picinbono B. On instantaneous amplitude and phase signals. *IEEE Trans Signal Process*, 1997, 45: 552–560
  - 35 Qian S, Chen D. Signal representation using adaptive normalized Gaussian functions. *Signal Processing*, 1994, 36: 1–11
  - 36 Qian S, Chen D. Joint Time-Frequency Analysis: Methods and Applications. New York: Prentice Hall, 1996
  - 37 Rice S O. Mathematical analysis of random noise. *Bell Syst Tech J*, 1944, 23: 282–310
  - 38 Setzer S, Steidl G, Teuber T. Deblurring Poissonian images by split Bregman techniques. *J Visual Commun Image Rep*, 2010, 21: 193–199
  - 39 Shekel J. Instantaneous frequency. *Proc IRE*, 1953, 41: 548–548
  - 40 Shi Y, Li K F, Yung Y L, et al. A decadal microwave record of tropical air temperature from AMSU-A/Aqua observations. *Climate Dynamics*, in press, 2013, doi: 10.1007/s00382-013-1696-x
  - 41 Tropp J, Gilbert A. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans Inform Theory*, 2007, 53: 4655–4666
  - 42 Van der Pol B. The fundamental principles of frequency modulation. *Proc Inst Elect Eng*, 1946, 93: 153–158
  - 43 Wu H T. Instantaneous frequency and wave shape functions (I). *Appl Comput Harmon Anal*, 2012, in press, doi: 10.1016/j.acha.2012.08.008
  - 44 Wu Z, Huang N E. Ensemble Empirical Mode Decomposition: A noise-assisted data analysis method. *Adv Adapt Data Anal*, 2009, 1: 1–41
  - 45 Wu Z, Huang N E, Long S R, et al. Trend, detrend, and the variability of nonlinear and non-stationary time series. *Proc Natl Acad Sci USA*, 2007, 104: 14889–14894