SCIENCE CHINA
Mathematics

• ARTICLES •

doi: 10.1007/s11425-016-9088-9

A two-level method for sparse time-frequency representation of multiscale data

Dedicated to Professor LI TaTsien on the Occasion of His 80th Birthday

LIU ChunGuang[1], SHI ZuoQiang[2] & HOU Thomas Yizhao[3,*]

[1]*Department of Mathematics, Jinan University, Guangzhou* 510632, China;
[2]*Yau Mathematical Sciences Center, Tsinghua University, Beijing* 100084, *China;*
[3]*Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA* 91125, *USA*

Email: tcgliu@jnu.edu.cn, zqshi@mail.tsinghua.edu.cn, hou@cms.caltech.edu

Received October 13, 2016; accepted April 20, 2017

Abstract Based on the recently developed data-driven time-frequency analysis (Hou and Shi, 2013), we propose a two-level method to look for the sparse time-frequency decomposition of multiscale data. In the two-level method, we first run a local algorithm to get a good approximation of the instantaneous frequency. We then pass this instantaneous frequency to the global algorithm to get an accurate global intrinsic mode function (IMF) and instantaneous frequency. The two-level method alleviates the difficulty of the mode mixing to some extent. We also present a method to reduce the end effects.

Keywords sparse representation, time-frequency analysis, matching pursuit, two-level method, end effects

MSC(2010) 94A12, 65N21

Citation: Liu C G, Shi Z Q, Hou T Y. A two-level method for sparse time-frequency representation of multiscale data. Sci China Math, 2017, 60, doi: 10.1007/s11425-016-9088-9

1 Introduction

Developing a truly adaptive data analysis method is important for our understanding of many natural phenomena. Traditional data analysis methods such as the Fourier transform or windowed Fourier transform often use pre-determined basis. Although they are very effective in representing linear and stationary data, applications of these methods to nonlinear and nonstationary data tend to give many unphysical harmonic modes. Time-frequency analysis is a more effective way to analyze data. It represents a signal with a joint function of both time and frequency (see [9]). There have been several powerful wavelet-based time-frequency analysis techniques (see [5, 22, 26, 30]). However, these methods do not remove the artificial harmonics completely.

Another important approach in the time-frequency analysis is to study instantaneous frequency of a signal. Some of the pioneering work in this area was due to Van der Pol [36] and Gabor [10]. They introduced the so-called analytic signal (AS) method that uses the Hilbert transform to define instantaneous frequency of a signal. However, this method works mostly for monocomponent signals, and

* Corresponding author

© Science China Press and Springer-Verlag Berlin Heidelberg 2017

requires that the number of zero-crossings in the signal is equal to the number of local extrema of the signal (see [1]). The zero-crossing method (see [28,33,34]) and the Wigner-Ville distribution method (see [1,9,24,25,31,32]) have been also introduced to define instantaneous frequency. However, the zero-crossing method cannot apply to signals with multiple components and is sensitive to noise. The method based on the Wigner-Ville distribution is known to suffer from the interference between different components.

The empirical mode decomposition (EMD) method introduced by Huang [18] is a truly adaptive data analysis method for multi-component signals. The EMD method decomposes a signal into a collection of intrinsic mode functions (IMFs) sequentially. The main idea is the removal of the local median from a signal by using a sifting process. The local median is approximated by the averaging the upper envelope and the lower envelope of the signal using cubic spline. The sifting process stops when the following two conditions are satisfied: (i) The number of the extrema and the number of the zero crossings of the function is equal or differ at most by one; (ii) the upper envelope and the lower envelope should be symmetric with respect to zero. Once one IMF is obtained, the Hilbert transform can be applied to get the instantaneous frequency.

The EMD method has found many applications (see [20, 38, 39]). One important property of these IMFs is that they contain physically meaningful information such as trend and instantaneous frequency, and give a physically meaningful Hilbert spectral representation. One drawback of the EMD method is its sensitivity to noise since it approximates the upper and lower envelopes of a signal by cubic spline based on local extrema of a signal. Clearly, local extrema of a signal are very sensitive to noise perturbation. To alleviate this difficulty, an ensemble EMD method (EEMD) was proposed to make it more stable to noise perturbation (see [37]).

Despite the considerable success of the EMD method in various applications, there is a lack of theoretical foundation of this method. Recently, there have been several attempts to establish a mathematical foundation for the EMD method (see [6, 7, 13–17, 19]). An interesting observation of the EMD method is that the signal is decomposed into the sum of only a small number of IMFs. In some sense, the EMD method can be considered as a method that gives a sparse representation over the dictionary consisting of all IMFs. Based on this observation, Hou and Shi [14] proposed a data-driven time-frequency analysis method by looking for the sparsest decomposition over all IMFs. This idea gives rise to an optimization problem

$$\begin{aligned} &\min_{(a_k)_{1\leqslant k\leqslant M},(\theta_k)_{1\leqslant k\leqslant M}} M \\ &\text{subject to} \quad f(t)=\sum_{k=1}^{M} a_k(t)\cos\theta_k(t), \quad 0\leqslant t\leqslant T, \quad a_k\cos\theta_k\in\mathcal{D}. \end{aligned} \tag{1.1}$$

When the signal is polluted by noise, the equality in the above constraint is relaxed to be an inequality depending on the noise level. Here, $\mathcal{D}$ is the dictionary consisting of all IMFs, which is defined as

$$\mathcal{D}=\{a(t)\cos\theta(t):\theta'(t)\geqslant 0,\ a(t)\in V(\theta)\},$$

$V(\theta)$ is the collection of all the functions that are less oscillatory than $\cos\theta(t)$, i.e.,

$$V(\theta)=\operatorname{span}\left\{\varphi\left(\frac{\xi\theta(t)}{2\pi\lambda_\varphi}-m\right):m=n_1,n_1+1,\ldots,n_2\right\}, \tag{1.2}$$

where $\xi\leqslant 1/2$ is a fixed parameter that controls the regularity of the basis functions in V. The condition $\xi\leqslant 1/2$ roughly corresponds to requiring that the derivative of the basis functions in this dictionary is less than $1/2$ of the derivation of θ itself. This is how we enforce that the envelope $a(t)$ is smoother than the phase function $\theta(t)$. Here, φ is a preselected scaling function and λ_φ is the center frequency of φ, and the integers $n_1=n_1(\varphi,\theta)$ and $n_2=n_2(\varphi,\theta)$ are chosen to make sure that the energy of each

$$\varphi\left(\frac{\xi s}{2\pi\lambda_\varphi}-m\right)$$

that lies inside the domain $[\theta(0), \theta(T)]$ is not very small. In our computation, we select all

$$\varphi\left(\frac{\xi s}{2\pi\lambda_\varphi} - m\right)$$

whose domain has nonempty intersection with $[\theta(0), \theta(T)]$. This choice of the dictionary generalizes the overcomplete Fourier basis used in [14].

In some sense, this is a nonlinear version of the l_0 minimization problem. Based on the studies in the compressed (compressive) sensing (see [3,4,8,11]), Hou and Shi [14] proposed a data-driven time-frequency analysis method. The 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 being determined *a priori*. This explains the term "data-driven". In [14], an efficient algorithm based on nonlinear matching pursuit and fast Fourier transform has been proposed to solve the above nonlinear optimization problem. Under the assumption of the scale separation, the convergence of the algorithm for periodic data has been analyzed in [17].

The uniqueness of the optimization problem (1.1) is analyzed in [23] under the assumption of the scale separation which is defined as follows.

Definition 1.1 (Scale-separation)**.** A 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:

$$\begin{aligned} &a(t) \in C^1(\mathbb{R}), \quad \theta \in C^2(\mathbb{R}), \quad \inf_{t\in\mathbb{R}} \theta'(t) > 0, \\ &\frac{\sup_{t\in\mathbb{R}} \theta'(t)}{\inf_{t\in\mathbb{R}} \theta'(t)} = M' < +\infty, \quad \left|\frac{a'(t)}{\theta'(t)}\right| \leqslant \epsilon, \quad \left|\frac{\theta''(t)}{(\theta'(t))^2}\right| \leqslant \epsilon, \quad \forall\, t \in \mathbb{R}. \end{aligned}$$

Before we proceed, we give some explanation on the above definition. First of all, the instantaneous frequency $\theta'(t)$ must be positive to have a physical meaning. This is why we impose that $\theta'(t) > 0$ over the time domain that we consider. Secondly, if the instantaneous frequency $\theta'(t)$ has unbounded variation over the time domain that we consider, then it increases significantly the chance of mode mixing. As a result, we would lose uniqueness. Therefore, it is necessary to assume that the instantaneous frequency $\theta'(t)$ has a bounded variation over the time domain. The condition that $|a'(t)| \leqslant \epsilon\theta'(t)$ implies that the envelope a is relatively smooth compared with the phase function $\theta(t)$, which is one of the essential requirements of scale separation in the sense that the envelope fluctuation is smaller that of the phase function. The last assumption on $\theta(t)$ is to ensure that $\theta(t)$ has some regularity in itself. This condition rules out the case when we have strong frequency modulation, which is a source of uniqueness in the data-driven time-frequency decomposition. Although the above definition is formulated on the entire domain, it can be relaxed to each local time interval over the entire domain time that we consider.

It is proved that if the signal is well separated as defined in Definition 1.2, the solution of the optimization problem (1.1) is unique up to an error associated with the scale separation factor ϵ and the noise level ϵ_0.

Definition 1.2 (Well-separated signal)**.** A signal $f : \mathbb{R} \to \mathbb{R}$ is said to be well-separated with separation factor ϵ and frequency ratio d if it can be written as

$$f(t) = \sum_{k=1}^{M} a_k(t)\cos\theta_k(t) + r(t), \tag{1.3}$$

where each

$$f_k(t) = a_k(t)\cos\theta_k(t)$$

satisfies the scale-separation property with separation factor ϵ, $r(t) = O(\epsilon_0)$ and their phase function θ_k satisfies

$$\theta_k'(t) \geqslant d\theta_{k-1}'(t), \quad \forall\, t \in \mathbb{R}. \tag{1.4}$$

and $d > 1$, $d - 1 = O(1)$.

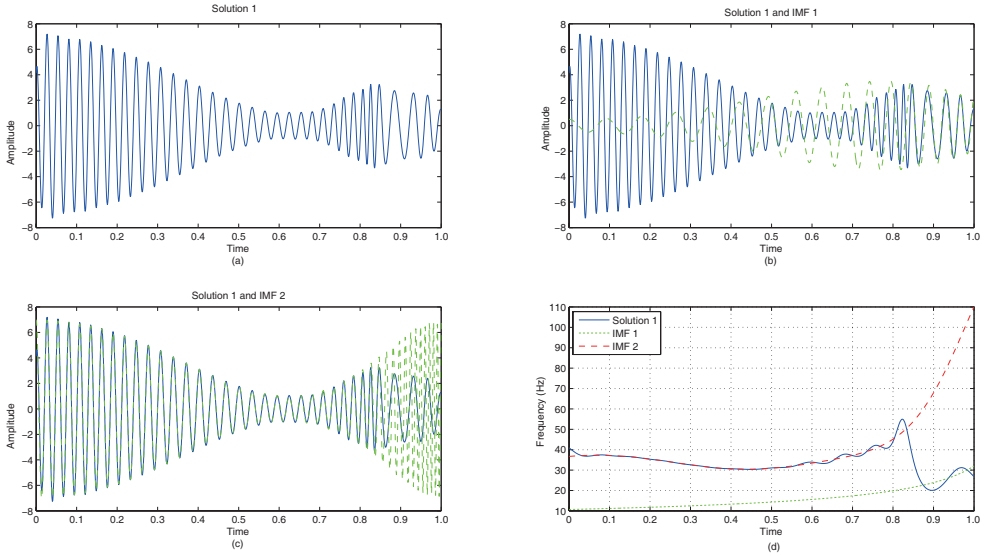

Figure 1 Mode mixing when applying the data-driven time-frequency analysis to the signal given in (1.5). (a) The first IMF given by the data-driven time-frequency analysis. (b) The computed IMF and the first true IMF $f_1(t) = a_1(t)\cos\theta_1(t)$. (c) The computed IMF and the second true IMF $f_2(t) = a_2(t)\cos\theta_2(t)$. (d) The computed instantaneous frequency (solid line) versus the two true instantaneous frequencies (dashed lines)

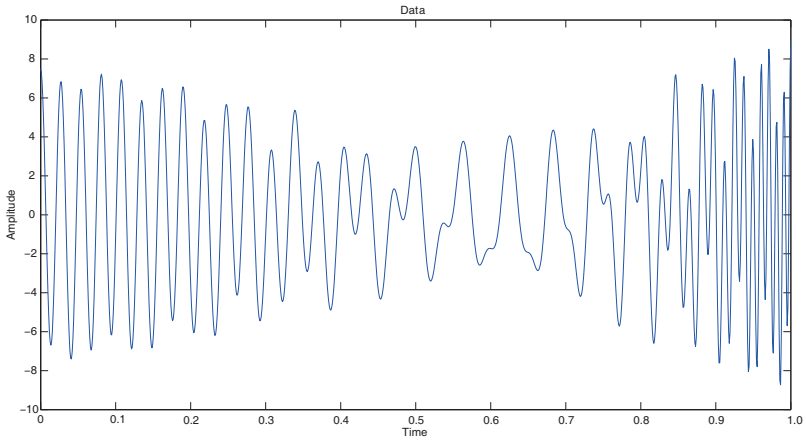

Figure 2 The signal given in (1.5)

In the above definition, in addition to requiring that each component satisfies the scale separation conditions stated in Definition 1.1, we also require that the instantaneous frequency $\theta_k'(t)$ in the k-th component is strictly greater than that of the $(k-1)$-th component by a constant factor d, i.e., strictly greater than one. This rules out the possibility that the two adjacent components have comparable instantaneous frequencies, which could easily lead to mode mixing, thus violates the uniqueness of the decomposition.

This series of papers has established a solid framework for data-driven time-frequency analysis, from theory to algorithms. However, there are still several issues unresolved. In this paper, we consider two of them: initialization of the iterative algorithm and the end effect.

Initialization (Mode mixing)**.** The data-driven time-frequency analysis is formulated as a global optimization problem. An iterative algorithm has been proposed to solve this nonconvex optimization problem. To start the iteration, we need a good initial guess of the instantaneous frequency. In our previous work, we use the Fourier transform to get the initial instantaneous frequency, which is a constant.

In the real world applications, the instantaneous frequency of the signal may have large variation. The constant initialization may introduce mode mixing in the decomposition. Figure 1 shows one example of this kind.

In Figure 2, the signal is given as follows:

$$f(t) = a_1(t)\cos\theta_1(t) + a_2(t)\cos\theta_2(t), \quad t \in [0,1], \tag{1.5}$$

where

$$t = \frac{3}{2}s - \frac{1}{2}s^2$$

and

$$\begin{aligned} a_1 &= 2 - 1.5\cos(1.5\pi s),\\ a_2 &= 4 + 3\cos(2\pi s),\\ \theta_1 &= 32\pi s,\\ \theta_2 &= 84\pi s + 10.4\sin(2\pi s) + 1.32\sin(4\pi s). \end{aligned}$$

In this example, the variation of the instantaneous frequencies is very large, as shown in Figure 1. However, it still satisfies the scale separation assumption and is also well separated, as demonstrated in (1.6), i.e.,

$$\begin{aligned} \max_{t\in[0,1]} \left| \frac{a_1'(t)}{a_1(t)\theta_1'(t)} \right| &\approx 0.053,\\ \max_{t\in[0,1]} \left| \frac{\theta_1''(t)}{[\theta_1'(t)]^2} \right| &\approx 0.02,\\ \max_{t\in[0,1]} \left| \frac{a_2'(t)}{a_2(t)\theta_2'(t)} \right| &\approx 0.033,\\ \max_{t\in[0,1]} \left| \frac{\theta_2''(t)}{[\theta_2'(t)]^2} \right| &\approx 0.0124. \end{aligned} \tag{1.6}$$

Based on the theoretical analysis in [23], we have a unique decomposition up to a small error depending on the scale separation factors given above. On the other hand, the optimization problem we want to solve is nonlinear and nonconvex. If we start the iteration with a poor initial instantaneous frequency, the algorithm may converge to a local minimum. As shown in Figure 1, from a constant initial instantaneous frequency given by the Fourier transform, the algorithm gives an IMF with mode mixing. In the interval $[0, 0.8]$, the mode that we compute matches the second IMF and in the rest of the interval, it matches the first IMF well. Two IMFs mix with each other in the decomposition. This phenomenon is called mode mixing and is also observed in the EMD method. Intermittent signals and the perturbation of noise are two of the common reasons that cause mode mixing. Moreover, for a signal in which the frequencies of the IMFs change significantly, mode mixing is often observed in the EMD or the EEMD method since the decomposition is dyadic. The mode mixing could also appear in the data-driven time-frequency analysis if the initial guess is not chosen properly.

In this paper, we propose a two-level method to tackle the problem of the initialization. The main idea is to generate the initial instantaneous frequency locally by cutting the signal to small pieces using a window function and passing the result to the original iterative algorithm as the initial guess to refine the result. Under the assumption that the instantaneous frequency is smooth, we can approximate the instantaneous frequency by a constant locally with reasonable accuracy. In this small interval, we can use the previous constant initialization and compute the decomposition. Then, we move the window to both sides of this local interval and compute the decomposition in the new interval. To connect the local decomposition to a global one, we keep the adjacent local intervals overlap with each other. After we move the interval to cover the entire time interval, we get a global decomposition. We then use this decomposition to obtain the initial guess of the instantaneous frequency to run the iterative algorithm globally. The detailed description of the two-level algorithm can be found in Section 3.

End effect. In this paper, we also propose a method to alleviate the end effect of the decomposition. The end effect is due to the finite time range of the signal. Reducing the end effect is important in the prediction of the signal beyond the given time interval. It is very difficult, in some sense impossible, to remove the end effect completely. In this paper, we introduce a method to alleviate the end effect under the scale separation assumption. The main idea is to extend the signal a little bit while preserving its intrinsic smoothness in the instantaneous frequency and the envelope. The extension algorithm will be given in Section 4.

The rest of the paper is organized as follows. In Section 2, we review the data-driven time-frequency analysis briefly for the sake of the completeness of the paper. The two-level algorithm is given in Section 3. The end effect is discussed in Section 4. The complete two-level method is summarized in Section 5. Several numerical examples are shown in Section 6. Finally, some concluding remarks are given in Section 7.

2 A brief review

In this section, we review the global iterative algorithm introduced by Hou and Shi [14]. This global algorithm was inspired by the well-known matching pursuit method. The optimization problem (1.1) can be seen as a nonlinear l^0 minimization problem. Matching pursuit and its variants have been shown to be a powerful method to solve the l^0 minimization problem (see [27, 29, 35]). Based on match pursuit, Hou and Shi [14] proposed a nonlinear matching pursuit method to solve the optimization problem (1.1).

Algorithm (NMP).

Given data: The function $f(t)$.

Task: Approximate the solution to (1.1) by matching pursuit.

Initialization: Let $k = 0$ and the initial residual $r^0(t) = f(t)$.

Main iteration:

Step 1. Increase k by 1. Solve the constrained nonlinear least-square problem

$$\begin{aligned} &\min_{a_k,\theta_k} \|r^{k-1}(t) - a_k(t)\cos\theta_k(t)\|_{L^2}^2 \\ &\text{subject to} \quad a_k \in V(\theta_k), \quad \theta_k' \geqslant 0. \end{aligned} \tag{P_2}$$

Step 2. If $\|a_k \cos\theta_k\|_{L^2} < \varepsilon_0$, decrease k by 1, STOP. Otherwise, update residual

$$r^k(t) = r^{k-1}(t) - a_k(t)\cos\theta_k(t). \tag{2.1}$$

Step 3. If $\|r^k\|_{L^2} < \varepsilon_0$, STOP. Otherwise, go back to Step 1.

Output: The approximate decomposition of f is

$$f(t) = \sum_{j=1}^{k} a_j(t)\cos\theta_j(t) + r^k(t). \tag{2.2}$$

Here, $V(\theta)$ (defined by (1.2)) is the collection of all the functions that are less oscillatory than $\cos\theta(t)$.

A key difficulty in solving the above nonlinear optimization problem is to solve Problem (P_2) in Step 1 of the above iteration. In [14], Hou and Shi proposed to use the following Gauss-Newton type algorithm, which is based on successive linearization of the nonlinear optimization problem and updating the solution using a Gauss-Newton type algorithm. For each linearized optimization problem, we can solve the linearized optimization problem by using matching pursuit. The algorithm is described below.

Algorithm ($\mathbf{P_2}$).

Initialization: $\theta_k^{(0)} = \theta_0$ (How to select θ_0 will be discussed later).

Iteration:

Step 1. Solve the following linear least-square problem:

$$\begin{aligned} &\text{Minimize} \quad \|r^{k-1}(t) - a_k^{(n+1)}(t)\cos\theta_k^{(n)}(t) - b_k^{(n+1)}(t)\sin\theta_k^{(n)}(t)\|_{L^2}^2 \\ &\text{subject to} \quad a_k^{(n+1)}, b_k^{(n+1)} \in V(\theta_k^{(n)}). \end{aligned} \tag{2.3}$$

Step 2. Update $\theta_k^{(n)}$,

$$\theta_k^{(n+1)} = \theta_k^{(n)} - \lambda \arctan\left(\frac{b_k^{(n+1)}}{a_k^{(n+1)}}\right), \tag{2.4}$$

where

$$\lambda = \sup\left\{\beta \in [0,1] : \frac{d}{dt}\left(\theta_k^{(n)} - 2\beta \arctan\left(\frac{b_k^{(n+1)}}{a_k^{(n+1)}}\right)\right) > 0\right\}. \tag{2.5}$$

This step is necessary to enforce the condition that the instantaneous frequency is positive throughout the iteration.

Step 3. If $\|\theta_k^{(n+1)} - \theta_k^{(n)}\| < \varepsilon_0$, STOP. Otherwise, replace n by $n+1$, go to Step 1.

Output:

$$\begin{aligned} &\text{Phase function}: \quad \theta_k = \theta_k^{(n+1)}, \\ &\text{IMF}: \quad I_k = a_k^{n+1}\cos\theta_k^{n+1}. \end{aligned}$$

In the iterative process, 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, $\cos\theta_k^{(n)}$ and $\sin\theta_k^{(n)}$ are simple Fourier modes, then the least-square problem can be solved by using the Fast Fourier Transform. For more detail of the algorithm, we refer to [14].

This algorithm is very efficient and stable to noise perturbation. One problem left is the initial guess θ_k^0 in Algorithm(P_2). In general, we set $\theta_k^{(0)}$ to be a linear function such that $(\theta_k^{(0)})'$ is the frequency at which $|\widehat{r}^{k-1}|$ has the largest value, i.e.,

$$\theta_k^{(0)}(t) = \lambda_0 t, \tag{2.6}$$

where

$$\lambda_0 = \arg\max_{\omega>0} |\widehat{r}^{k-1}(\omega)|.$$

In practice, the instantaneous frequency of a signal may have large variation. So, a constant initial frequency in general is not a good choice. It may introduce mode mixing as shown in the previous section. In the subsequent section, we will present a local algorithm to get a good initial instantaneous frequency.

3 A local algorithm

As we mentioned before, using a constant as an initial guess for the instantaneous frequency is clearly not a good choice for real world applications since the instantaneous frequency may vary a lot over the time interval that we consider. However, under the assumption that the instantaneous frequency is a smooth function of time locally (over several periods), it can be well approximated by a constant locally. If we divide the signal to a number of small time intervals, it would make sense to use a constant initial frequency over each local time interval. On each local time interval, we can run the algorithm introduced in the previous section with a constant initial frequency and get the local instantaneous frequency. To construct a global frequency from the local instantaneous frequencies, we make the local time intervals overlap with the adjacent time intervals. This is the main idea of our local algorithm.

Next, we will give the detailed implementation of the local algorithm. We start from one IMF. First, we could obtain one IMF using Algorithm (P_2) with constant frequency globally. Denote the corresponding

envelope and phase function as a_{global} and θ_{global}. As we mentioned before, this solution may not be correct, but it still gives us some useful information about the IMF. Then, we cut the first piece, denoted as $[t_{0,l}, t_{0,r}]$ as follows. The center is the point at which a_{global} reaches its maximum, i.e.,

$$t_{0,c} = \frac{t_{0,l} + t_{0,r}}{2} = \arg\max_t a_{\text{global}}(t).$$

Over this subinterval, we have approximately N_p periods, i.e.,

$$t_{0,r} - t_{0,l} = \frac{2\pi N_p}{\theta'_{\text{global}}(t_{0,c})}. \tag{3.1}$$

It is a challenge to select a suitable N_p when we have no prior information of the IMFs. Obviously, too large N_p may cause mode mixing. On the other hand, N_p cannot be too small either. Since we aim to get the high frequency information of the IMFs, it is necessary to assume that $N_p \geqslant 1$. However, taking $N_p = 1$ or 2 is often not very effective in practice. First of all, as we will show in the rest of this section (see (3.2) and (3.3)), we use the information of an IMF on a subinterval to define the next subinterval, so the number of periods of the IMF we are pursuing may change slightly among different subintervals. When we take N_p too small, this IMF may not have at least one period on some subinterval. Therefore, we would not get the correct instantaneous frequency of this IMF. Secondly, if the N_p is too small, the number of subintervals would become quite large. So the small N_p implies that the computation would become more expensive. Finally, if we take into account the interference from the other IMFs and noise, a small N_p could lead to a big error on the instantaneous frequency of the IMF, and increase the possibility of mode mixing. In fact, when we apply the two-level method with $N_p = 2$ on the signal defined in (1.5), we find that mode mixing still occurs (although this time the different parts of the real IMFs are connected more smoothly than the parts of Solution 1 in Figure 1, the mode mixing leads to an extra IMF and a big residual). From our numerical results, taking $N_p \geqslant 3$ is necessary to alleviate the difficulties mentioned above. If we have more knowledge about the IMFs, a bigger N_p could be better. In this paper, we choose $N_p = 6$.

The key to the success of our two-level algorithm is to construct a sequence of overlapping local subintervals and extend the locally constructed IMFs from one subinterval to its neighboring subintervals successively so that we can glue these locally constructed IMFs to form the accurate global IMFs. In order to make the construction successful, it is essential to determine the appropriate amount of overlap among these local subintervals. The amount of overlap will determine the accuracy of the approximation of the global IMFs from the local IMFs and the efficiency of our two-level algorithm. We obtain a nearly optimal amount of overlap of the subintervals by solving a local weighted optimization problem and using the fact that different IMFs are nearly orthogonal to each other. To maximize the accuracy of our construction, we begin our construction of local IMFs from a local subinterval that gives the best approximation to the global IMFs. Below, we will give a detailed description of our local algorithm and the rationale for the way we determine the amount of overlap among different subintervals.

We begin with a subinterval $[t_{0,l}, t_{0,r}]$. We first apply Algorithm (P_2) on the signal restricted to this interval with the constant initial frequency $\theta'_{\text{global}}(t_{0,c})$. Then we get an IMF on $[t_{0,l}, t_{0,r}]$, denoted by $g_0^{\text{local}}(t) = a_0^{\text{local}}(t)\cos\theta_0^{\text{local}}(t)$. Now, we move the interval to extend the local IMF. Denote the new interval as $[t_{1,l}, t_{1,r}]$. The left boundary is determined as follows:

$$t_{1,l} = \arg \min_{\substack{t\in[t_{0,l},t_{0,r}],\\ \theta_0^{\text{local}}(t)/\pi\in\mathbb{Z}}} |t - [t_{0,l} + \nu(t_{0,r} - t_{0,l})]| \tag{3.2}$$

with $\nu = 1/3$. The parameter ν is used to control the size of the overlap between $[t_{0,l}, t_{0,r}]$ and $[t_{1,l}, t_{1,r}]$. The requirement that $\theta_0^{\text{local}}(t)/\pi \in \mathbb{Z}$ is designed to alleviate the end effect. The right boundary $t_{1,r}$ is given as

$$t_{1,r} = t_{1,l} + 2\pi N_p \cdot \frac{[t_{0,r} - \nu(t_{0,r} - t_{0,l})] - [t_{0,l} + \nu(t_{0,r} - t_{0,l})]}{\theta[t_{0,r} - \nu(t_{0,r} - t_{0,l})] - \theta[t_{0,l} + \nu(t_{0,r} - t_{0,l})]}. \tag{3.3}$$

Based on this choice, $t_{1,r}-t_{1,l}\approx t_{0,r}-t_{0,l}$ (noticing (3.1) and that $\frac{[t_{0,r}-\nu(t_{0,r}-t_{0,l})]-[t_{0,l}+\nu(t_{0,r}-t_{0,l})]}{\theta[t_{0,r}-\nu(t_{0,r}-t_{0,l})]-\theta[t_{0,l}+\nu(t_{0,r}-t_{0,l})]}\approx\frac{1}{\theta'_{\text{global}}(t_{0,c})}$), and $[t_{1,l},t_{1,r}]$ has approximately N_p periods.

We denote $t'_0=t_{0,r}-\nu(t_{0,r}-t_{0,l})$ and solve the following weighted problem to extend IMF to $[t_{1,l},t_{1,r}]$:

$$\begin{aligned}&(a_1^{\text{local}},\theta_1^{\text{local}})=\arg\min_{a,\theta}\|f-a\cos\theta\|^2_{L^2_{[t_{1,l},t_{1,r}]}}+w_1\|g_0^{\text{local}}-a\cos\theta\|^2_{L^2_{[t_{1,l},t'_0]}},\\&\text{subject to}\quad a\in V(\theta),\quad \theta'\geqslant 0.\end{aligned}\tag{3.4}$$

Here, f is the original signal, and w_1 is the wight to insure that the IMF on $[t_{1,l},t_{1,r}]$ and

$$g_1^{\text{local}}=a_1^{\text{local}}\cos\theta_1^{\text{local}}\approx g_0^{\text{local}}$$

on $[t_{1,l},t'_0]\subset[t_{0,l},t_{0,r}]$ such that g_1^{local} is a natural extension of g_0^{local} on $[t_{1,l},t_{1,r}]$. Solving the above optimization problem, we are able to extend g to a larger interval. Similarly, we could extend g to the left from $[t_{0,l},t_{0,r}]$.

To determine the value of w_1 in (3.4), we suppose that the local IMF g_0^{local} gives a good approximation to the IMF $I_k=a_k\cos\theta_k$ on the interval $[t_{0,l},t_{0,r}]$. Since $[t_{1,l},t_{1,r}]$ is a small interval, we could suppose that all the assumptions of [23, Theorem 3.4] are fulfilled for f on $[t_{1,l},t_{1,r}]$. Then if $w_1=0$ in (3.4), some (a_m,θ_m) should be the approximate global optimal solution to (3.4), where m might be equal to k or other $l\neq k$, depending on which IMF has the largest energy. To avoid mode mixing, w_1 should be selected to make sure that $m=k$, i.e., for any other IMF $I_l=a_l\cos\theta_l$ $(l\neq k)$, the following inequality should be satisfied:

$$\|f-I_l\|^2_{L^2_{[t_{1,l},t_{1,r}]}}+w_1\|g_0^{\text{local}}-I_l\|^2_{L^2_{[t_{1,l},t'_0]}}>\|f-I_k\|^2_{L^2_{[t_{1,l},t_{1,r}]}}+w_1\|g_0^{\text{local}}-I_k\|^2_{L^2_{[t_{1,l},t'_0]}}.\tag{3.5}$$

Noticing that $g_0^{\text{local}}\approx I_k$ on $[t_{0,l},t_{0,r}]$, $t_{1,r}-t_{1,l}\approx t_{0,r}-t_{0,l}$, and that all the functions concerned do not change very much on $[t_{0,l},t_{1,r}]$, we have the following approximation:

$$\begin{aligned}&\|f-I_l\|^2_{L^2_{[t_{1,l},t_{1,r}]}}\approx\|f-I_l\|^2_{L^2_{[t_{0,l},t_{0,r}]}},\\&\|g_0^{\text{local}}-I_l\|^2_{L^2_{[t_{1,l},t'_0]}}\approx\frac{t'_0-t_{1,l}}{t_{0,r}-t_{0,l}}\|g_0^{\text{local}}-I_l\|^2_{L^2_{[t_{0,l},t_{0,r}]}},\\&\|f-I_k\|^2_{L^2_{[t_{1,l},t_{1,r}]}}\approx\|f-g_0^{\text{local}}\|^2_{L^2_{[t_{0,l},t_{0,r}]}},\\&\|g_0^{\text{local}}-I_k\|^2_{L^2_{[t_{1,l},t'_0]}}\approx 0.\end{aligned}\tag{3.6}$$

So (3.5) is roughly equivalent to (noticing that $\frac{t'_0-t_{1,l}}{t_{0,r}-t_{0,l}}\approx 1-2\nu$)

$$w_1(1-2\nu)\|g_0^{\text{local}}-I_l\|^2_{L^2_{[t_{0,l},t_{0,r}]}}>\|f-g_0^{\text{local}}\|^2_{L^2_{[t_{0,l},t_{0,r}]}}-\|f-I_l\|^2_{L^2_{[t_{0,l},t_{0,r}]}}.\tag{3.7}$$

To obtain the above inequality, it is necessary and sufficient to set

$$w_1>\frac{\|f-g_0^{\text{local}}\|^2_{L^2_{[t_{0,l},t_{0,r}]}}-\|f-I_l\|^2_{L^2_{[t_{0,l},t_{0,r}]}}}{(1-2\nu)\|g_0^{\text{local}}-I_l\|^2_{L^2_{[t_{0,l},t_{0,r}]}}}.\tag{3.8}$$

However, the right-hand side could not be computed, since we have no detailed information of I_l. The only thing we know is that all the IMFs of f are mutually nearly orthogonal (and g_0^{local} is nearly orthogonal to I_l for $g_0^{\text{local}}\approx I_k$), which implies that

$$\|f-I_l\|^2_{L^2_{[t_{0,l},t_{0,r}]}}\approx\|I_k\|^2_{L^2_{[t_{0,l},t_{0,r}]}}+\sum_{m\neq k,l}\|I_m\|^2_{L^2_{[t_{0,l},t_{0,r}]}}\geqslant\|I_k\|^2_{L^2_{[t_{0,l},t_{0,r}]}}\approx\|g_0^{\text{local}}\|^2_{L^2_{[t_{0,l},t_{0,r}]}},\tag{3.9}$$

$$\|g_0^{\text{local}}-I_l\|^2_{L^2_{[t_{0,l},t_{0,r}]}}\approx\|g_0^{\text{local}}\|^2_{L^2_{[t_{0,l},t_{0,r}]}}+\|I_l\|^2_{L^2_{[t_{0,l},t_{0,r}]}}>\|g_0^{\text{local}}\|^2_{L^2_{[t_{0,l},t_{0,r}]}}.\tag{3.10}$$

Thus, (3.8) holds and the optimization condition (3.5) should be satisfied if we set

$$w_1 = \max\left\{0, \frac{\|f - g_0^{\text{local}}\|^2_{L^2_{[t_{0,l},t_{0,r}]}} - \|g_0^{\text{local}}\|^2_{L^2_{[t_{0,l},t_{0,r}]}}}{(1-2\nu)\|g_0^{\text{local}}\|^2_{L^2_{[t_{0,l},t_{0,r}]}}}\right\}. \tag{3.11}$$

Repeat above process until the subintervals cover the whole interval, we get an IMF over the whole interval. Using the local algorithm, we could avoid the mode mixing. However, the error is relatively large. So, we pass the result of the local algorithm to the global algorithm in Section 2 to refine the result. This gives the two-level algorithm in Section 5.

4 End effects

In this section, we discuss the end effects. End effects always emerge in data analysis due to the finite time span of the real signal. Under the assumption that the envelope and the phase function are smooth over the whole time span, we propose a numerical method to alleviate the end effects based on smoothly extending the signal beyond the boundary.

First, we give the extension algorithm for a single IMF. Suppose the time span of the IMF is $[0, T]$, we use the following algorithm to extend the IMF smoothly to $[-T/2, 3T/2]$.

Algorithm (A).

Task: For an approximation (a, θ) to some pair (a_k, θ_k) of $f(t)$, diminish the error between (a, θ) and (a_k, θ_k) near the end points.

Main iteration:

Step 1. Extend $a(t), \theta(t)$ to $\tilde{a}(t), \tilde{\theta}(t)$ on the interval $[-\frac{T}{2}, \frac{3T}{2}]$ as follows:

$$\tilde{\theta}''(t) = \begin{cases} \theta''(t), & t \in [0, T], \\ 0.5\left[1 + \cos\left(\dfrac{t-T}{2^{-n_0}T}\right)\right] \cdot \theta''(T), & t \in [T, (1+2^{-n_0})T], \\ 0.5\left[1 + \cos\left(\dfrac{t}{2^{-n_0}T}\right)\right] \cdot \theta''(0), & t \in [-2^{-n_0}T, 0], \\ 0, & \text{elsewhere}, \end{cases} \tag{4.1}$$

and

$$\tilde{a}'(t) = \begin{cases} a'(t), & t \in [0, T], \\ 0.5\left[1 + \cos\left(\dfrac{t-T}{2^{-n_0}T}\right)\right] \cdot a'(T), & t \in [T, (1+2^{-n_0})T], \\ 0.5\left[1 + \cos\left(\dfrac{t}{2^{-n_0}T}\right)\right] \cdot a'(0), & t \in [-2^{-n_0}T, 0], \\ 0, & \text{elsewhere}, \end{cases} \tag{4.2}$$

where $n_0 \geqslant 2$ is chosen such that, for some fixed $\delta \in (0.5, 1)$,

$$\begin{aligned} &\delta \min_{[0,T]} \theta'(\cdot) < \tilde{\theta}'(t) < \delta^{-1} \max_{[0,T]} \theta'(\cdot), \\ &\delta \min_{[0,T]} a(\cdot) < \tilde{a}(t) < \delta^{-1} \max_{[0,T]} a(\cdot), \end{aligned} \quad \forall\, t \in \left[-\frac{T}{2}, \frac{3T}{2}\right]. \tag{4.3}$$

Let

$$\tilde{f}(t) = \begin{cases} f(t), & t \in [0, T], \\ \tilde{a}(t)\cos\tilde{\theta}(t), & t \in \left[-\dfrac{T}{2}, 0\right] \cup \left[T, \dfrac{3T}{2}\right]. \end{cases} \tag{4.4}$$

Step 2. Apply Algorithm (P_2) to $\tilde{f}(t)$ with the initial guess $\tilde{\theta}$. Denote by $b(t)$ and $\vartheta(t)$ the output.

Step 3. Replace (a,θ) by (b,ϑ).

Output: New (a,θ), which is supposed to have smaller error, especially near the end points.

Algorithm (A) is effective for a single IMF, but may not work well when there are two or more IMFs. For the signal with multiple IMFs and noise,

$$f(t)=a_0(t)+\sum_{k=1}^{K}a_k(t)\cos\theta_k(t)+\text{noise},\quad \forall\, t\in[0,T], \tag{4.5}$$

where $I_k=\{a_k(t)\cos\theta_k(t)\}$ are IMFs satisfying the scale separation property, and $a_0(t)$ is the trend, we carry out a sweeping iteration to alleviate the end effects. Suppose that we already have a decomposition of f,

$$f(t)=b_0(t)+\sum_{k=1}^{K}b_k(t)\cos\vartheta_k(t)+r(t),\quad \forall\, t\in[0,T], \tag{4.6}$$

where b_0, b_k and ϑ_k are approximations to a_0, a_k and θ_k, respectively, and $r(t)$ is the sum of noise and errors, we diminish the end effects for each IMF iteratively as follows:

Algorithm (End effects).

Task: Modify $b_0, b_1, \vartheta_1, \ldots, b_K, \vartheta_K$ given in (4.6).

Parameter: A predesigned positive integer L, and

$$\eta=\min\left\{\frac{\vartheta_k(T)-\vartheta_k(0)}{2\pi}: k=1,2,\ldots,K\right\},$$

the number of periods of the IMF with lowest frequency on the entire time interval.

Main iteration:

Step 1. $l=1$;

Step 2 (De-trend step, see Section 5)**.** Solve the optimization problem (5.1) with f replaced by $b_0(t)+r(t)$ and $n=2\lfloor\eta\rfloor$. Replace $b_0(t)$ and $r(t)$ by the solution and the remaining data, respectively.

Step 3. For $k=1:K$ {Modify (b_k,ϑ_k) using Algorithm (A) with f replaced by $b_k\cos\vartheta_k+r$. Replace r by the remaining data.}

Step 4. $l=l+1$. If $l=L$, STOP; otherwise, go back to Step 2.

In our computations, the number of iterations $L=3$, unless we indicate the value of L.

5 A two-level method

Combining the local algorithm and the global algorithm, we get a two-level algorithm. Notice that the local algorithm only works for IMFs with oscillations. If the signal has a large trend, it may introduce some trouble for the two-level method. First, we give a de-trend method based on Fourier extension (see [2, 21] and the references therein). Since a trend, denoted as a_0, is often a monotonic function, we regard a_0 as the restriction of $\tilde{a}_0$ on $[0,T]$, where $\tilde{a}_0$ is a periodic function with period $4T$, and its graph is symmetric with respect to the vertical line $t=-\frac{T}{2}$. We are going to solve the following optimization problem to get an approximation of the trend:

$$\begin{aligned}&a_0=\arg\min_h\|h-f\|_{L^2_{[0,T]}}\\&\text{subject to}\quad h\in G_n:=\operatorname{span}\left\{1,\cos\frac{\pi}{2T}k\left(t+\frac{T}{2}\right),k=1,2,\ldots,n\right\}.\end{aligned} \tag{5.1}$$

Combining all the algorithms together, we get the following two-level method:

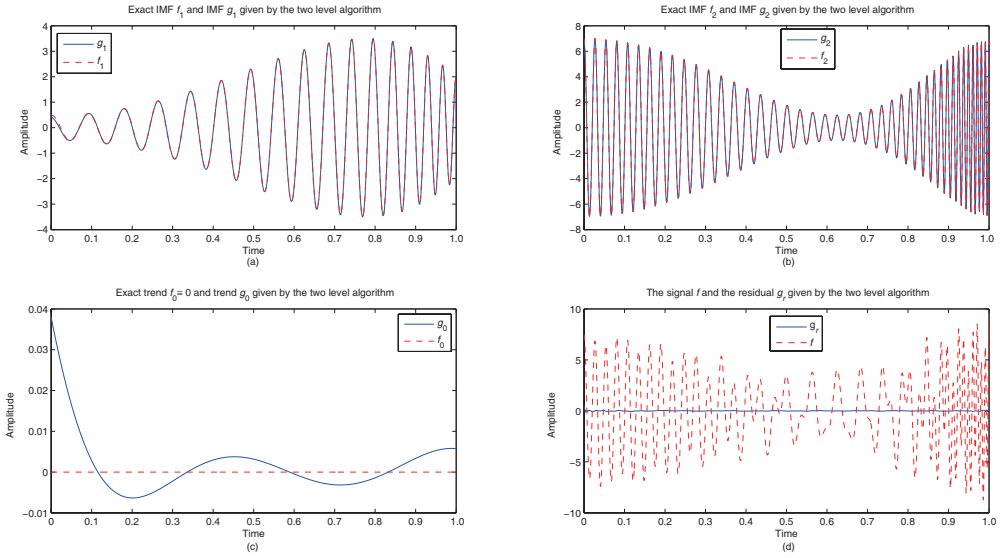

Figure 3 IMFs given by the two-level method for the signal in (1.5). (a) First IMF; (b) second IMF; (c) trend; (d) residual

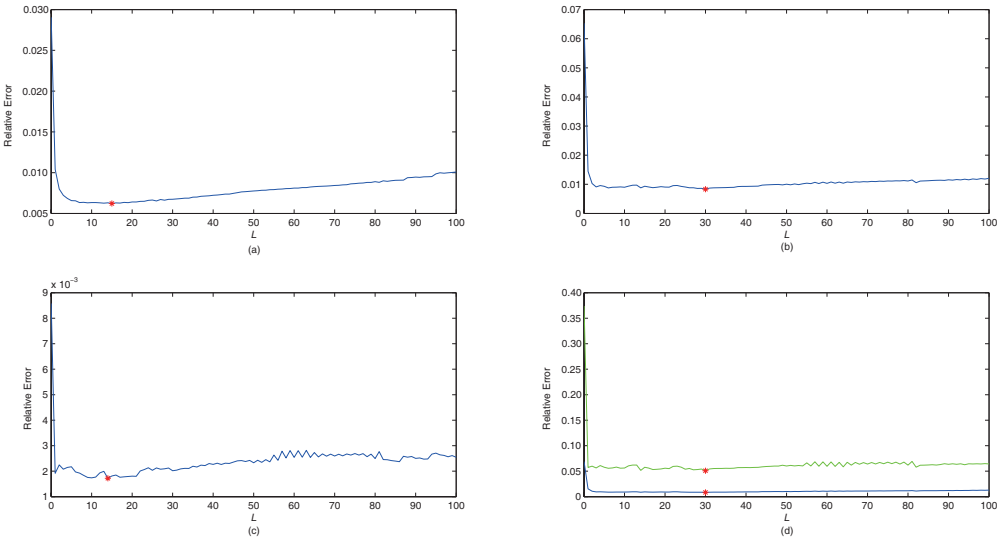

Figure 4 (a) $\frac{\|g_1-f_1\|_{L^2}}{\|f\|_{L^2}}$. The minimum 0.0062 is attained at $L=15$. (b) $\frac{\|g_2-f_2\|_{L^2}}{\|f\|_{L^2}}$. The minimum 0.0083 is attained at $L=30$. (c) $\frac{\|g_0-f_0\|_{L^2}}{\|f\|_{L^2}}$. The minimum 0.0017 is attained at $L=14$. (d) $\frac{\max|g_r|}{\max|f|}$ (upper line) and $\frac{\|g_r\|_{L^2}}{\|f\|_{L^2}}$ (lower line). The minima 0.0510 and 0.0085 are attained at $L=30$ simultaneously

Algorithm (Two-level).

Task: Approximate the modes of signal $f(t)$ given in (4.6) by the two-level method.

Parameters: The threshold ε_0, the scaling function φ, positive integer N_p and L.

Initialization: Let $k=0$ and the initial residual $r^0(t)=f(t)$. Let $a_0(t)\equiv 0$.

Main iteration:

Step 1. Let $\lambda_0=\arg\max_{\omega>0}|\widehat{r^k}(\omega)|$.

Step 2. If $\frac{\lambda_0\cdot T}{2\pi}\leqslant\frac{N_p}{2}$, solve the Fourier extension problem (5.1) with f replaced by r^k and $n=2N_p$. Add the solution to $a_0(t)$, subtract the solution from r^k. Go back to Step 1.

Step 3. Increase k by 1. Solve the constrained nonlinear least-square problem (P_2) about (a_k, θ_k).

Step 4. If $\|a_k \cos\theta_k\|_{L^2} < \varepsilon_0$, decrease k by 1, STOP.

Step 5. Use (a_k, θ_k) to determine the initial subinterval, and search the IMF piecewisely. Replace (a_k, θ_k) by the new envelope and phase function.

Step 6. Modify θ_k to keep θ_k' smooth. Solve Problem (P_2) again with initial guess $\theta_0 = \theta_k$. Replace (a_k, θ_k) by the new solution. Update the residual

$$r^{k+1}(t) = r^k(t) - a_k(t)\cos\theta_k(t). \tag{5.2}$$

Step 7. If $\|r^k\|_{L^2} < \varepsilon_0$, STOP. Otherwise, go back to Step 1.

Weaken the end effects: Let $K = k$ and $r(t) = r^k(t)$. Apply Algorithm (End Effects) to $\{a_0, a_1, \theta_1, \ldots, a_K, \theta_K\}$.

Output: The approximate decomposition of f is

$$f(t) = a_0(t) + \sum_{k=1}^{K} a_k(t)\cos\theta_k(t) + r(t), \tag{5.3}$$

where a_0 is the trend, $r(t)$ is the noise, and $\{a_k \cos\theta_k\}$ are IMFs.

Figure 3 shows the IMFs g_0, g_1 and g_2 given by the two-level method (with $L = 30$ in Algorithm (End effects)) for the signal given in (1.5). Comparing with the results shown in Figure 1, the mode mixing is completely removed and we get almost perfect IMFs up to the boundary. In addition, we give the relative errors with g_1, g_2, g_0 and g_r obtained by two-level method with L ranges from 0 to 100 (see Figure 4). We see that the first several iterations make the greatest improvement.

6 Numerical results

In this section, we present some numerical results to demonstrate the performance of the two-level algorithm.

Example 6.1. First, we consider a synthetic signal, which is generated as follows:

$$f(t) = b_0(t) + b_1(t)\cos\theta_1(t) + b_2(t)\cos\theta_2(t) + b_3(t)\cos\theta_3(t), \quad t \in [0, 1], \tag{6.1}$$

where

$$t = \frac{3}{2}s - \frac{1}{2}s^2$$

and

$$\begin{aligned}
&a_1 = a_3 = 2 - 1.5\cos(1.5\pi s),\\
&a_2 = 4 + 3\cos(2\pi s),\\
&\phi_1 = 32\pi s,\\
&\phi_2 = 84\pi s + 10.4\sin(2\pi s) + 1.32\sin(4\pi s) + 1,\\
&\phi_3 = 160\pi s + 20\sin(2\pi s) + 1.2\sin(4\pi s) + 2,
\end{aligned}$$

and

$$b_0 = \cos(2t + 4), \quad b_i = a_i(0.987t + 0.01), \quad \theta_i = \phi_i(0.987t + 0.01), \quad i = 1, 2, 3.$$

The signal $f(t)$ is sampled over 1,024 uniform grid points in $[0, 1]$.

The IMFs computed by the two-level algorithm are shown in Figure 5. The IMFs given by the two-level algorithm match the exact IMFs very well, while the IMFs obtained by the EMD method have mode mixing as shown in Figure 6.

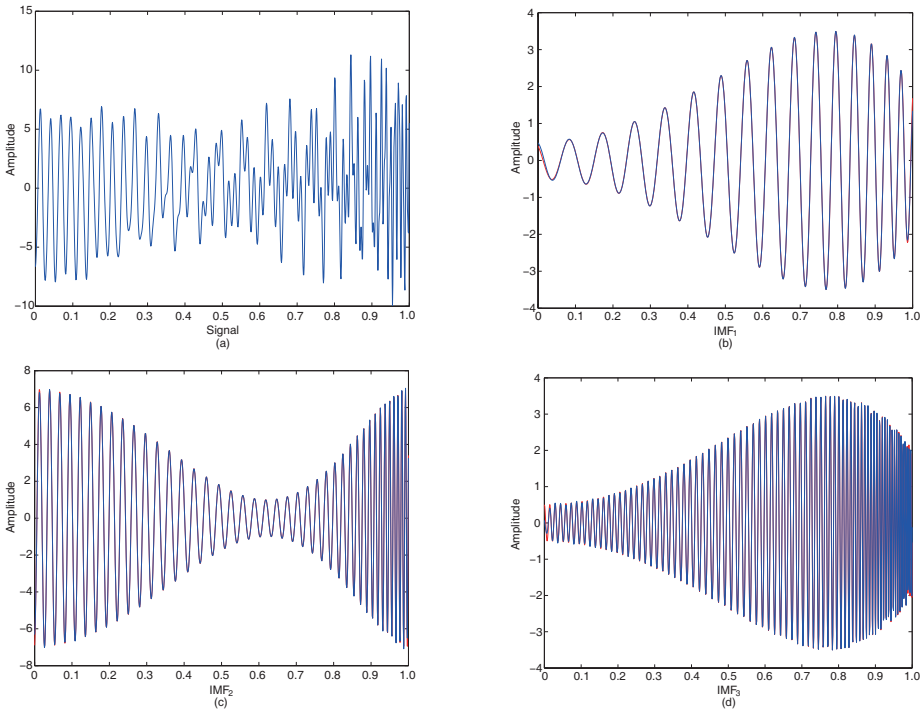

Figure 5 Original signal and IMFs in Example 6.1. Dash lines: exact IMFs; solid lines: IMFs obtained by the two-level algorithm

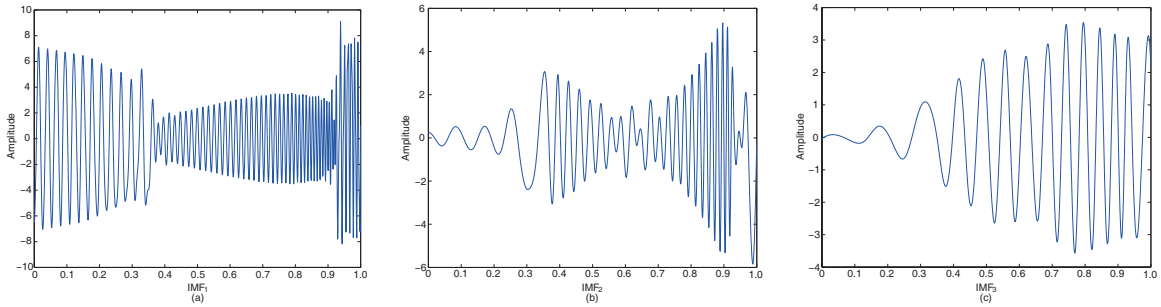

Figure 6 IMFs given by EMD in Example 6.1

We also test the two-level algorithm for signals with noise. In this numerical example, the noisy signal is $f(t)+0.5X(t)$, where $f(t)$ is given in (6.1) and $X(t)$ is the standard white noise with standard deviation $\sigma = 1$. The results are given in Figure 7. As we can see, even with noise, the two-level algorithm still gives IMFs with reasonable accuracy. In this case, the EMD method suffers from severe mode mixing as shown in Figure 8. The EMD method gives a total of 6 IMFs. Each IMF has severe mode mixing among noise and different IMFs. We can hardly see the pattern of the exact IMFs.

At the end of this example, we consider the prediction of the trend. As shown before in this example, the two-level method gives very good estimates of the IMFs up to the boundary, which means that the two-level method is capable of handling the end effect. In some sense, the alleviation of the end effect is related to the prediction of the signal. Actually, from the two-level method, we can get the prediction of the IMFs over a few periods beyond the boundary. For highly oscillatory IMFs, this is very small time interval, may not be very significant in the application. However, for the trend, we can give prediction over a relatively long interval. To predict trend is very important in some applications, for example, global warming. In the following test, we use the signal in $[0, 0.9]$ to predict the trend in $[0.9, 1]$. Figure 9 shows the results of the prediction. Recall that L is the number of iterations in the two-level method. $L = 0$

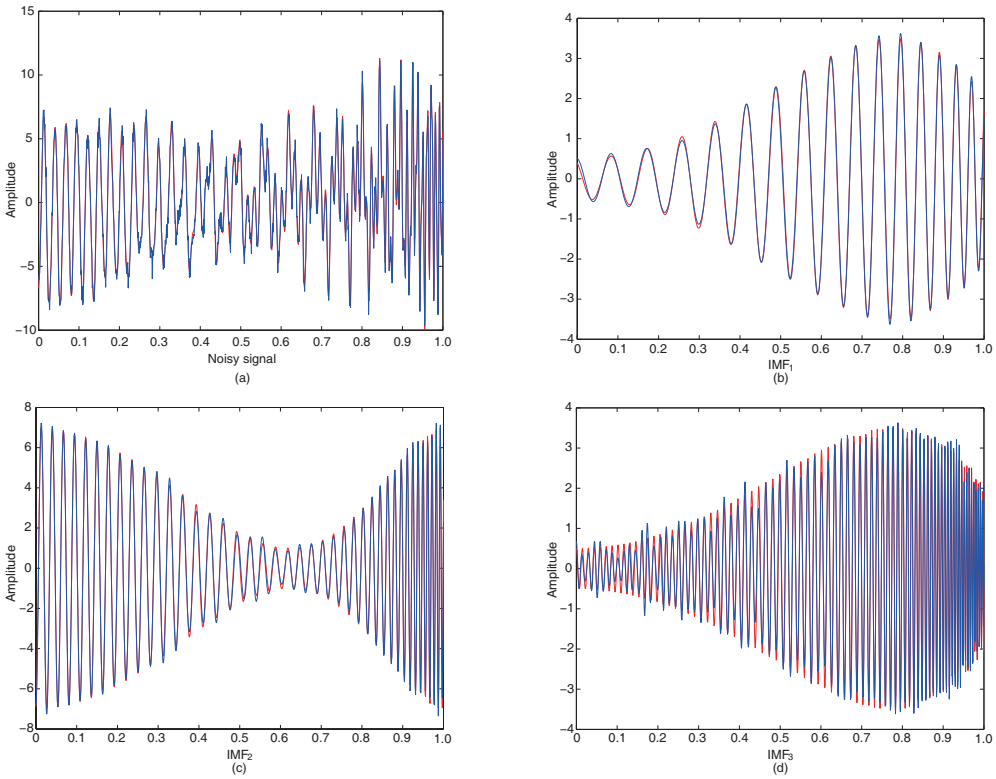

Figure 7 A noisy signal and IMFs obtained by the two-level algorithm in Example 6.1. Dash lines: exact IMF; solid lines: IMF given by the two-level algorithm

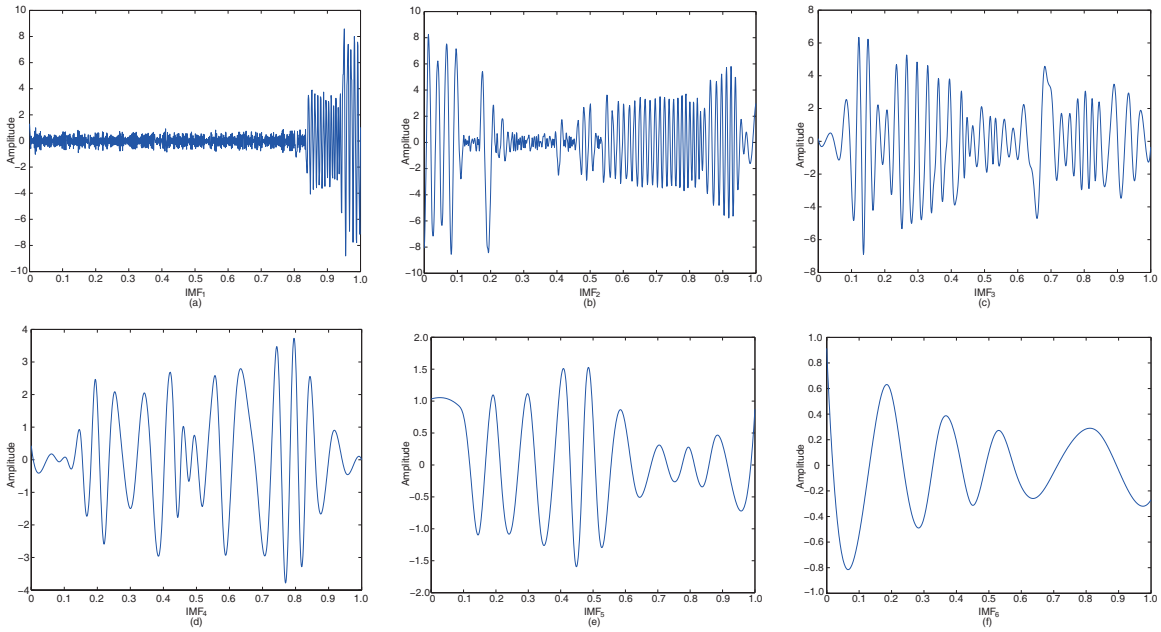

Figure 8 IMFs given by EMD method for a noisy signal given in Example 6.1

means no two-level iteration. We can see that, in this case, the end effect is very large and the prediction is very poor. After 3 times two-level iteration, the end effect is reduced significantly and the prediction also becomes much better. If we further run the two-level iteration to 60 times, the predictions become even better. At the same time, the computational cost also becomes higher. Based on our experience, $L = 3$ seems to be a good balance between accuracy and efficiency.

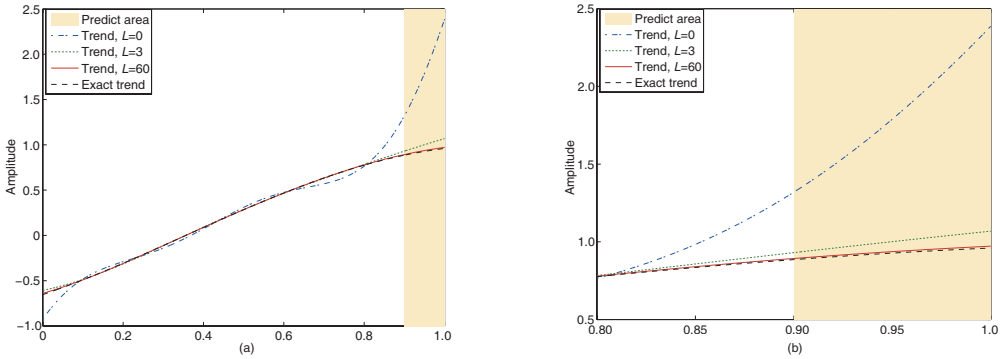

Figure 9 Prediction of trend using 90% data. (a) Entire interval $[0, 1]$. (b) Zoom in on $[0.8, 1]$

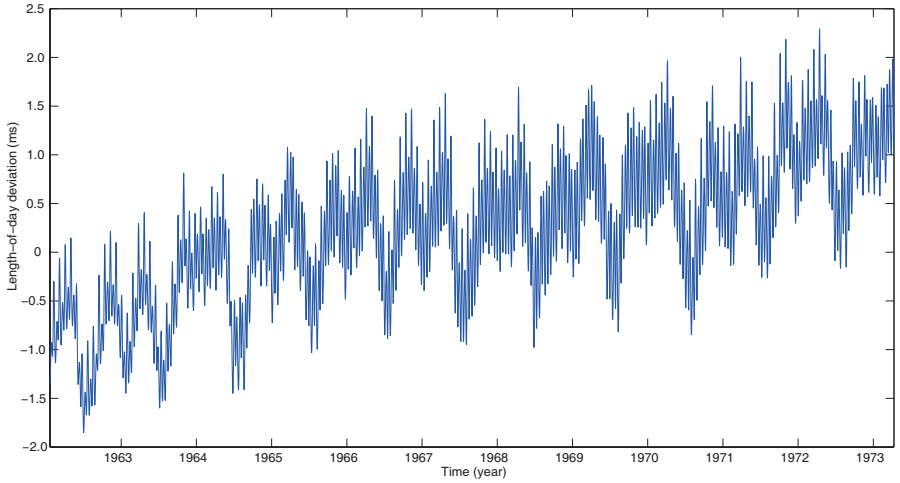

Figure 10 Length-of-day data from January 20, 1962 to April 8, 1973

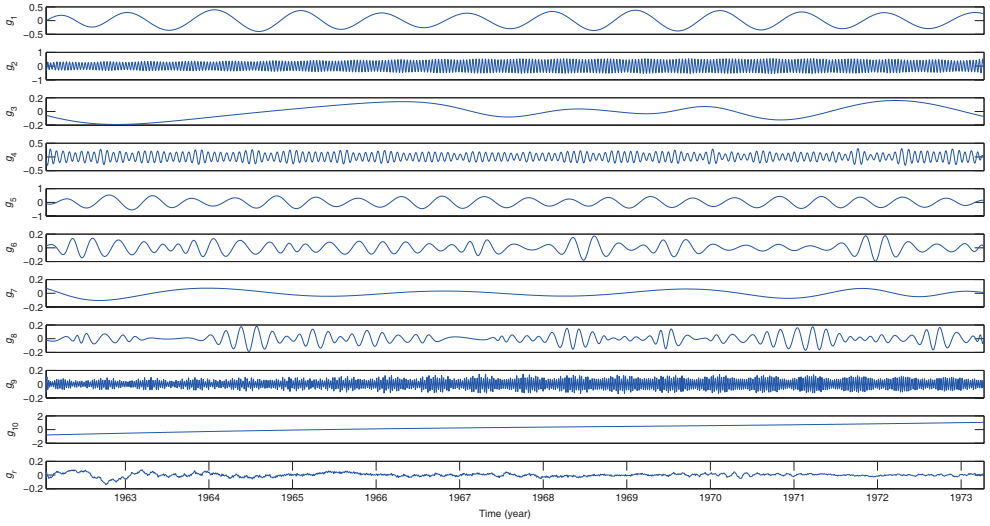

Figure 11 The 10 modes and the residual g_r obtained by the two-level method on the Length-of-day data

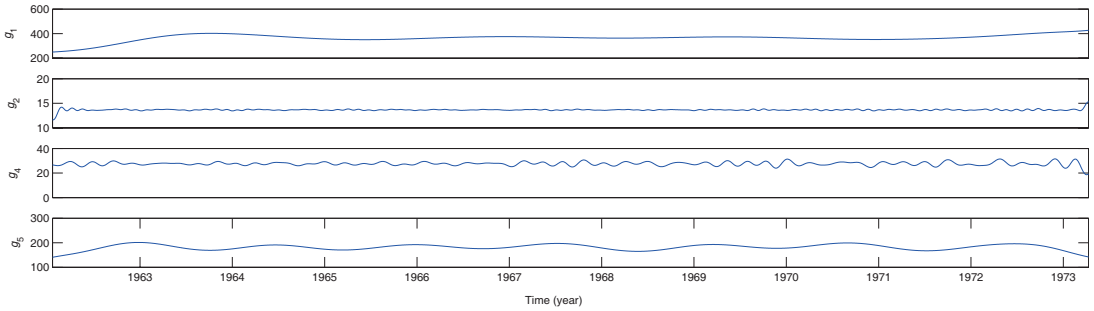

Figure 12 Instantaneous periods of g_1, g_2, g_4 and g_5. The average periods of these IMFs are approximately $362.3, 13.7, 27.5$ and 181.1, respectively

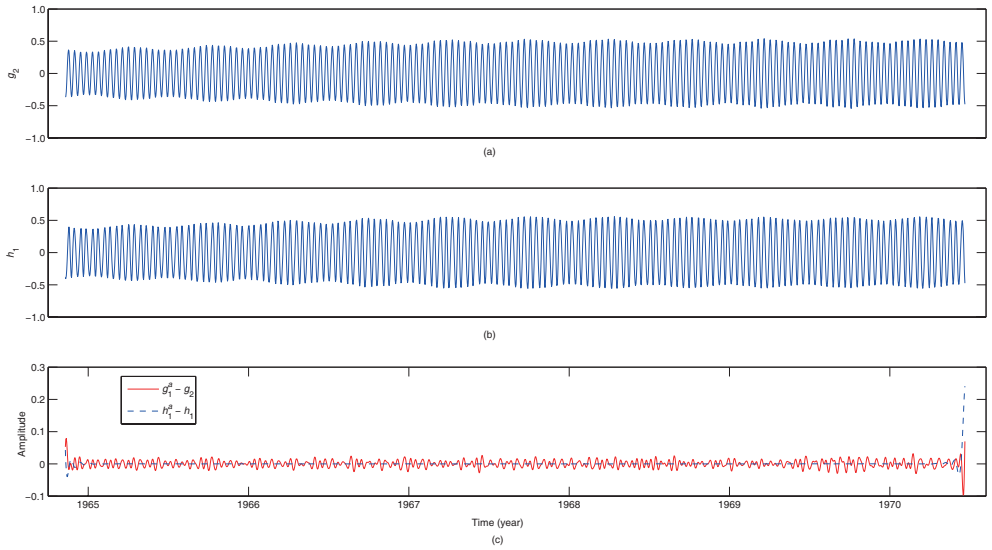

Figure 13 (a) g_2; (b) h_1; (c) solid line: $g_1^a - g_2$, dash line: $h_1^a - h_1$

Example 6.2 (Real data: Length-of-day data)**.** In this example, we apply the two-level method to real data: Length-of-day data, which was produced by Gross [12]. This set of data covers the time interval from January 20, 1962 to January 6, 2001, totally 14,232 days, nearly 39 years. Using this data, we perform the following two tasks:

Task I. To check if the periods of the IMFs obtained by the two-level method have relations with natural meteorological cycle. Here, we only deal with the data of the first 4,097 days, from January 20, 1962 to April 8, 1973 (see Figure 10).

Task II. To check if our Algorithm (End effects) is more effective than the EMD method on real data. Since we do not know what the true IMFs are, we restrict the IMFs obtained in Task I (and the IMFs obtained by the EMD method on the same data) to their central time interval (from November 9, 1964 to June 19, 1970, i.e., from the 1,025th day to the 3,073th day) and regard these restricted IMFs as nearly true IMFs. Then we apply the two-level method and the EMD method to the Length-of-day data in the same central time interval and compare the end effect of the new IMFs.

First of all, applying the two-level method to the data in Task I, we obtain 10 IMFs (see Figure 11). In the 10 IMFs, the first 4 IMFs (g_{10}, g_2, g_5, g_1) contain over 92.6% of the energy of the original signal. In addition, for the residual g_r, we have $\frac{\|g_r\|_{L^2}}{\|f\|_{L^2}} \approx 0.0323$, $\frac{\max|g_r|}{\max|f|} \approx 0.0619$. Figure 12 shows that g_2, g_5, g_1 and g_4 (the IMF with 5th largest energy) could be related to some natural phenomenon.

Secondly, to test how well the two-level method handles the end effect, we apply the two-level method to the interior part of the length-of-day data described in Task II, and obtain 8 IMFs, denoted by

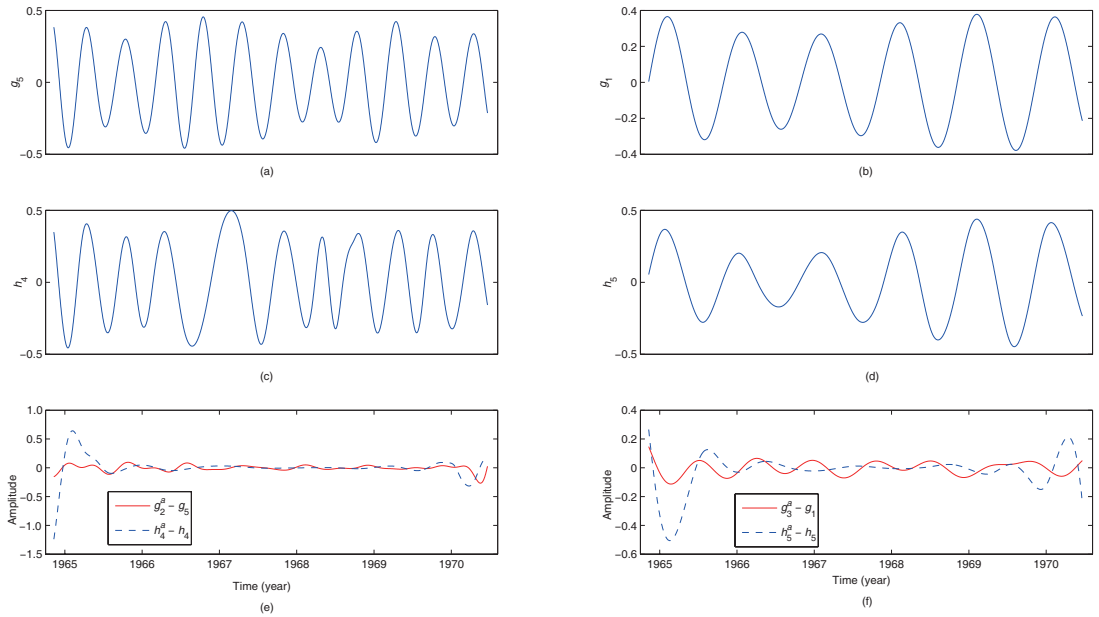

Figure 14 (a) g_5; (b) g_1; (c) h_4; (d) h_5; (e) solid line: $g_2^a - g_5$, dash line: $h_4^a - h_4$; (f) solid line: $g_3^a - g_1$, dash line: $h_5^a - h_5$

g_k^a $(k = 1, \ldots, 8)$. Except for the trend, the three IMFs with largest energies, g_1^a, g_2^a and g_3^a, have similar frequencies with g_2, g_5 and g_1, respectively, and on the time interval from November 9, 1964 to June 19, 1970, we have

$$\frac{\|g_1^a - g_2\|_{L^2}}{\|g_2\|_{L^2}} \approx 0.0346, \quad \frac{\|g_2^a - g_5\|_{L^2}}{\|g_5\|_{L^2}} \approx 0.2046, \quad \frac{\|g_3^a - g_1\|_{L^2}}{\|g_1\|_{L^2}} \approx 0.1906.$$

As a comparison, we also perform the same test for the EMD method. We first apply the EMD method to the length-of-day data from the first 4,097 days and from the 1,025th day to the 3,073th day, respectively, which give IMFs h_k $(k = 1, \ldots, 13)$ and h_k^a $(k = 1, \ldots, 12)$, respectively. The IMFs h_1, h_4 and h_5 have the largest energies except for the trend, and their frequencies are similar to those of g_2, g_5 and g_1, respectively. Their corresponding parts in $\{h_k^a\}$ are h_1^a, h_4^a and h_5^a. On the time interval from November 9, 1964 to June 19, 1970, we have

$$\frac{\|h_1^a - h_1\|_{L^2}}{\|h_1\|_{L^2}} \approx 0.0300, \quad \frac{\|h_4^a - h_4\|_{L^2}}{\|h_4\|_{L^2}} \approx 0.6268, \quad \frac{\|h_5^a - h_5\|_{L^2}}{\|h_5\|_{L^2}} \approx 0.5309.$$

Moreover, we see that the end effects of g_1^a, g_2^a and g_3^a are smaller than h_1^a, h_4^a and h_5^a, respectively. Here, we use g_2, g_5, g_1 and h_1, h_4, h_5 as the nearly true IMFs (see Figures 13 and 14).

7 Concluding remarks

In this paper, we introduced a two-level method based on the data-driven time-frequency analysis (see [14, 15]) to find the sparse time frequency decomposition. This method alleviates mode mixing in the data-driven time frequency analysis for certain signals that satisfy our scale separation assumption. The method seems to work well even when the frequencies of the modes vary significantly over the entire time interval. The main idea of the two-level method is to find all pieces of an IMF in overlapped subintervals and connect them into a global IMF. Then, we use this IMF as an initial guess to find a more accurate solution in the whole time domain.

Furthermore, we also constructed an extend-and-optimize algorithm to weaken the end effect. After obtaining all modes in the whole time domain, for each IMF, we add the residual to the selected IMF, and extend it to a larger time domain smoothly, then we use the data-driven time-frequency analysis to

decompose this signal again. By doing so, the end effect seems to be reduced effectively for the signals that we consider here.

There are some remaining issues to be resolved in the future. One of them is to generalize the two-level method for real world data that do not satisfy our scale separation assumption. This is an important issue that we plan to resolve in our future study.

Acknowledgements This work was supported by National Science Foundation of USA (Grants Nos. DMS-1318377 and DMS-1613861) and National Natural Science Foundation of China (Grant Nos. 11371220, 11671005, 11371173, 11301222 and 11526096).

References

1 Boashash B. Time-Frequency Signal Analysis: Methods and Applications. Melbourne-New York: Longman-Cheshire/John Wiley Halsted Press, 1992
2 Boyd J P. A comparison of numerical algorithms for Fourier extension of the first, second, and third kinds. J Comput Phys, 2002, 178: 118–160
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 Daubechies I. Ten Lectures on Wavelets. CBMS-NSF Regional Conference Series on Applied Mathematics, vol. 61. Philadelphia: SIAM, 1992
6 Daubechies I, Lu J, Wu H. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. Appl Comput Harmon Anal, 2011, 30: 243–261
7 Daubechies I, Wang Y, Wu H. ConceFT: Concentration of frequency and time via a multitapered synchrosqueezed transform. Philos Trans A Math Phys Eng Sci, 2016, 374: 20150193
8 Donoho D L. Compressed sensing. IEEE Trans Inform Theory, 2006, 52: 1289–1306
9 Flandrin P. Time-Frequency/Time-Scale Analysis. San Diego: Academic Press, 1999
10 Gabor D. Theory of communication. J IEEE, 1946, 93: 426–457
11 Gribonval R, Nielsen M. Sparse representations in unions of bases. IEEE Trans Inform Theory, 2003, 49: 3320–3325
12 Gross R S. Combinations of Earth Orientation Measurements: SPACE2000, COMB2000, and POLE2000. JPL Publication 1-2. Pasadena: Jet Propulsion Laboratory, 2000
13 Hou T Y, Shi Z. Adaptive data analysis via sparse time-frequency representation. Adv Adapt Data Anal, 2011, 3: 1–28
14 Hou T Y, Shi Z. Data-drive time-frequency analysis. Appl Comput Harmonic Anal, 2013, 35: 284–308
15 Hou T Y, Shi Z. Sparse time-frequency representation of nonlinear and nonstationary data. Sci China Math, 2013, 56: 2489–2506
16 Hou T Y, Shi Z. Sparse time-frequency decomposition by dictionary adaptation. Philos Trans A Math Phys Eng Sci, 2016, 374: 20150194
17 Hou T Y, Shi Z, Tavallali P. Convergence of a data-driven time-frequency analysis method. Appl Comput Harmonic Anal, 2014, 37: 235–270
18 Huang N E. 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
19 Huang N E, Daubechies I, Hou TY. Adaptive data analysis: Theory and applications. Philos Trans A Math Phys Eng Sci, 2016, 374: 20150207
20 Huang N E, Wu Z. A review on Hilbert-Huang Transform: The method and its applications on geophysical studies. Rev Geophys, 2008, 46: RG2006
21 Huybrechs D. On the Fourier extension of non periodic functions. SIAM J Numer Anal, 2010, 47: 4326–4355
22 Jomes D L, Parks T W. A high resolution data-adaptive time-frequency representation. IEEE Trans Acoust Speech Signal Process, 1990, 38: 2127–2135
23 Liu C G, Shi Z Q, Hou T Y. On the uniqueness of sparse time-frequency representation of multiscale data. Multiscale Model Simul, 2015, 13: 790–811
24 Loughlin P J, Tracer B. On the amplitude- and frequency-modulation decomposition of signals. J Acoust Soc Amer, 1996, 10: 1594–1601
25 Lovell B C, Williamson R C, Boashash B. The relationship between instantaneous frequency and time-frequency representations. J Acoust Soc Amer, 1993, 41: 1458–1461
26 Mallat S. A Wavelet Tour of Signal Processing: The Sparse Way. San Diego: 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 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
31 Picinbono B. On instantaneous amplitude and phase signals. IEEE Trans Signal Process, 1997, 45: 552–560
32 Qian S, Chen D. Joint Time-Frequency Analysis: Methods and Applications. Upper Saddle River: Prentice Hall, 1996
33 Rice S O. Mathematical analysis of random noise. Bell Syst Tech J, 1944, 23: 282–310
34 Shekel J. Instantaneous frequency. Proc IRE, 1953, 41: 548
35 Tropp J, Gilbert A. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans Inform Theory, 2007, 53: 4655–4666
36 Van der Pol B. The fundamental principles of frequency modulation. Proc IEEE, 1946, 93: 153–158
37 Wu Z, Huang N E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. Adv Adapt Data Anal, 2009, 1: 1–41
38 Wu Z, Huang N E, Chen X. The multi-dimensional ensemble empirical mode decomposition method. Adv Adapt Data Anal, 2009, 1: 339–372
39 Wu Z, Huang N E, Long S R, et al. On the trend, detrending, and variability of nonlinear and nonstationary time series. Proc Natl Acad Sci USA, 2007, 104: 14889–14894