

A Two-Level Method for Sparse Time-Frequency Representation of Multiscale Data

Dedicated to Professor Li Da-Qian on the Occasion of his 80th Birthday

Chunguang Liu¹, Zuoqiang Shi² & Thomas Y. Hou^{3,*}

¹*Department of Mathematics, Jinan University, Guangzhou, 510632, China;*

²*Yau Mathematical Sciences Center, Tsinghua University, Beijing, 100084, China;*

³*Applied and Comput. Math, MC 9-94, Caltech, Pasadena, CA, 91125, USA;*

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

Received ; accepted

Abstract Based on the recently developed data-driven time-frequency analysis [16], 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:

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 [11]. There have been several powerful wavelet-based time-frequency analysis techniques [5, 23, 27, 31]. 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 [8] and Gabor [12]. 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 requires that the number of zero-crossings in the signal is equal to the number of local extrema of the signal [1]. The zero-crossing method [29, 34, 35] and the Wigner-Ville distribution method [1, 11, 25, 26, 32, 33] have been also introduced to define instantaneous frequency. However, the zero-crossing method cannot apply

*Corresponding author

to signals with multiple components and is sensitive to noise. The methods based on the Wigner-Ville distribution is known to suffer from the interference between different components.

The Empirical Mode Decomposition (EMD) method introduced by Norden E. Huang et. al. [10] 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 e.g. [21, 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 [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 [6, 7, 15–20]. 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 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:

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

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

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

$V(\theta)$ is the collection of all the functions that are less oscillatory than $\cos \theta(t)$:

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

where $\xi \leq 1/2$ is a fixed parameter, 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 $\varphi(\frac{\xi s}{2\pi \lambda_\varphi} - m)$ that lies inside the domain $[\theta(0), \theta(T)]$ is not very small. In our computation, we select all $\varphi(\frac{\xi s}{2\pi \lambda_\varphi} - m)$ whose domain has nonempty intersection with $[\theta(0), \theta(T)]$.

In some sense, this is a nonlinear version of the l_0 minimization problem. Based on the studies in the compressed (compressive) sensing [3, 4, 9, 13], Hou and Shi proposed a data-driven time-frequency analysis method [16]. 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 [16], an efficient algorithm based on 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 [19].

The uniqueness of the optimization problem (1.1) is analyzed in [24] 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:

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

$$\frac{\sup_{t \in \mathbb{R}} \theta'(t)}{\inf_{t \in \mathbb{R}} \theta'(t)} = M' < +\infty, \quad \left| \frac{a'(t)}{\theta'(t)} \right| \leq \epsilon, \quad \left| \frac{\theta''(t)}{(\theta'(t))^2} \right| \leq \epsilon, \quad \forall t \in \mathbb{R}.$$

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 $f : \mathbb{R} \rightarrow \mathbb{R}$ 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

$$\theta'_k(t) \geq d\theta'_{k-1}(t), \quad \forall t \in \mathbb{R}. \tag{1.4}$$

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

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 real world applications, the instantaneous frequency of the signal may have large variation. The constant initialization may introduce mode mixing in the decomposition. Fig. 2 shows one example of this kind.

In Fig. 1, 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 Fig. 2. However, it still satisfies the scale separation assumption and is also well separated, as calculated in (1.6).

$$\begin{aligned} \max_{t \in [0,1]} \left| \frac{a'_1(t)}{a_1(t)\theta'_1(t)} \right| &\approx 0.053, & \max_{t \in [0,1]} \left| \frac{\theta''_1(t)}{[\theta'_1(t)]^2} \right| &\approx 0.02, \\ \max_{t \in [0,1]} \left| \frac{a'_2(t)}{a_2(t)\theta'_2(t)} \right| &\approx 0.033, & \max_{t \in [0,1]} \left| \frac{\theta''_2(t)}{[\theta'_2(t)]^2} \right| &\approx 0.0124. \end{aligned} \tag{1.6}$$

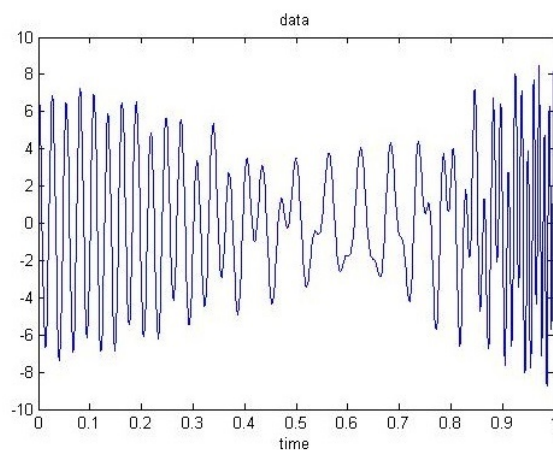


Figure 1 The signal given in (1.5).

Based on the theoretical analysis in [24], we have an 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 Fig. 2, 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. Besides, for a signal in which the frequencies of the IMFs change largely, mode mixing often happens in the EMD method for the decomposition is dyadic, and may happen in the data-driven time-frequency analysis if the initial guess is not good.

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. Notice that under the assumption that the instantaneous frequency is smooth, locally it can be well approximated by a constant function. 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 give 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 instantaneous frequency and 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

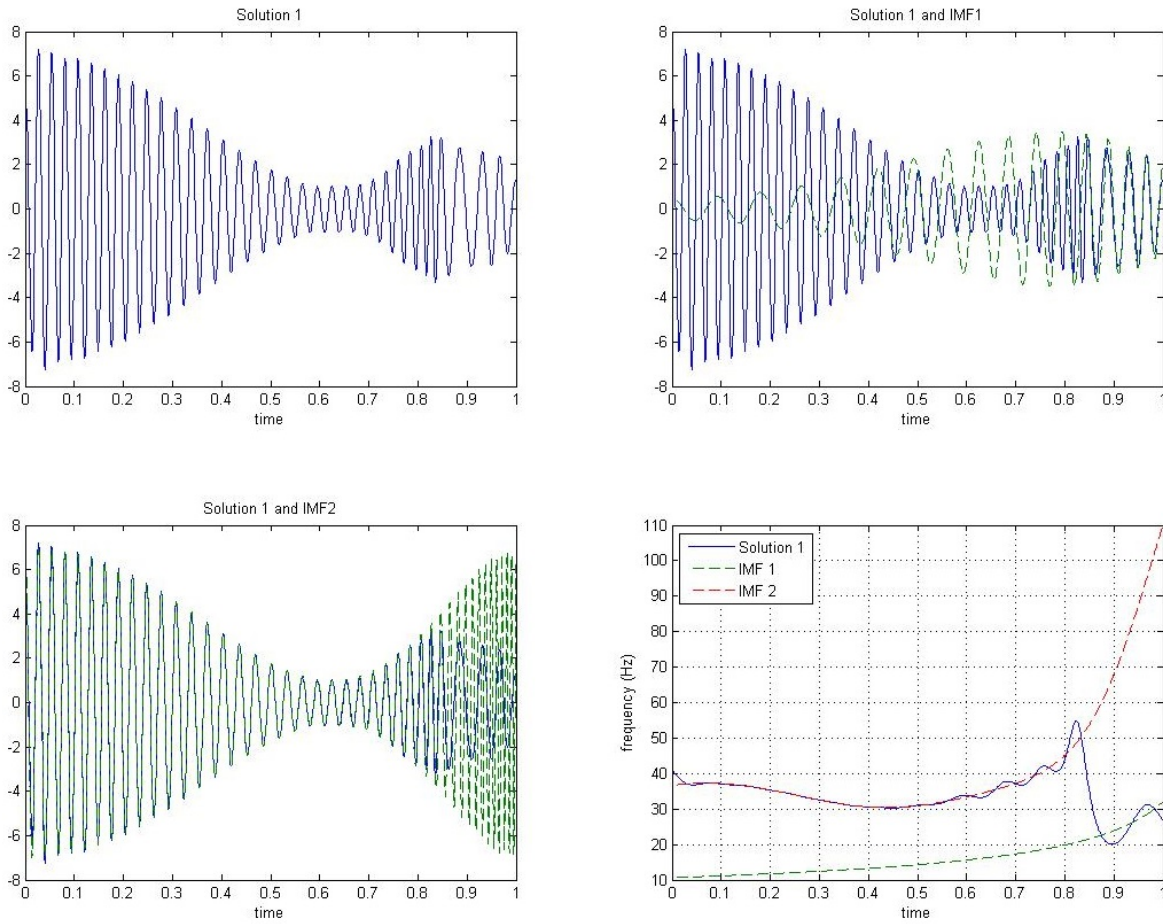


Figure 2. Mode mixing when applying the data-driven time-frequency analysis to the signal given in (1.5). Top left: The first IMF given by the data-driven time-frequency analysis. Top right: The computed IMF and the first true IMF $f_1(t) = a_1(t) \cos \theta_1(t)$. Bottom left: The computed IMF and the second true IMF $f_2(t) = a_2(t) \cos \theta_2(t)$. Bottom right: The computed instantaneous frequency (blue solid line) versus the two true instantaneous frequencies (dashed lines).

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 in [16]. 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 [28, 30, 36]. Based on match pursuit, Hou and Shi 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:

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

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

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

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

Output: The approximate decomposition of f is

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

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

To solve the problem (P_2) in Step 1 of the above iteration, we use the following Gauss-Newton type algorithm:

Algorithm (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

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

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

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

where

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

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

Output:

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

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, $\cos \theta_k^{(n)}$, $\sin \theta_k^{(n)}$ 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 [16].

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.

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

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

The real instantaneous frequency may have large variation. So, the constant initial frequency 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

Using a constant as an initial guess for the frequency is clearly not a good choice for real world applications since the 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.

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

The length is approximately N_p periods, i.e.

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

It is a challenge to select a suitable N_p when we have no information of the IMFs. Obviously, too large N_p may cause mode mixing. On the other hand, N_p cannot be too small. Since we aim to get the oscillation property of the IMFs, it is necessary to assume that $N_p \geq 1$. However, taking $N_p = 1$ or 2 is seldom effective in practice. First, as we 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 periods on some subinterval. Therefore we would not get the correct oscillation of it. Second, the smaller the N_p is, the bigger the number of subintervals is. So the small N_p means the long computing time. At last, considering the interference from the other IMFs and noise, too small N_p would cause 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), mode mixing also occur (although this time the different parts of the real IMFs are connected more smoothly than the parts of Solution 1 in Fig. 2, the mode mixing cause an extra solution and a big residual). By numerical results, taking $N_p \geq 3$ is necessary to avoid/reduce the above shortages. If we have more knowledge about the IMFs, a bigger N_p could be better. In this paper, we choose $N_p = 6$.

On $[t_{0,l}, t_{0,r}]$, we apply Algorithm (P_2) on the signal restricted to this interval with the constant initial frequency $\theta'_{global}(t_{0,c})$. Then we get an IMF on $[t_{0,l}, t_{0,r}]$, denoted by $g_0^{local}(t) = a_0^{local}(t) \cos \theta_0^{local}(t)$. 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:

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

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 $\theta_0^{local}(t)/\pi \in \mathbb{Z}$ is designed to alleviate the end effect. The right boundary $t_{1,r}$ is given as

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

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

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

Here f is the original signal, w_1 is the wight to insure that the IMF on $[t_{1,l}, t_{1,r}]$, $g_1^{local} = a_1^{local} \cos \theta_1^{local} \approx g_0^{local}$ 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 [24, Theorem 3.4] are fulfilled for f on $[t_{1,l}, t_{1,r}]$. Then if $w_1 = 0$ in the problem (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 $m = k$, that is, for any other IMF $I_l = a_l \cos \theta_l$ ($l \neq k$), the

following inequality should be satisfied:

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

Noticing that $g_0^{local} \approx I_k$ 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:

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

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

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

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

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

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 $g_0^{local} \approx I_k$), which implies that

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

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

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

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

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 give 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:

$$\tilde{\theta}''(t) = \begin{cases} \theta''(t) & t \in [0, T] \\ 0.5 [1 + \cos(\frac{t-T}{2^{-n_0}T})] \cdot \theta''(T) & t \in [T, (1+2^{-n_0})T] \\ 0.5 [1 + \cos(\frac{t}{2^{-n_0}T})] \cdot \theta''(0) & t \in [-2^{-n_0}T, 0] \\ 0 & \text{else where,} \end{cases} \quad (4.1)$$

and

$$\tilde{a}'(t) = \begin{cases} a'(t) & t \in [0, T] \\ 0.5 [1 + \cos(\frac{t-T}{2^{-n_0}T})] \cdot a'(T) & t \in [T, (1+2^{-n_0})T] \\ 0.5 [1 + \cos(\frac{t}{2^{-n_0}T})] \cdot a'(0) & t \in [-2^{-n_0}T, 0] \\ 0 & \text{else where,} \end{cases} \quad (4.2)$$

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

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

Let

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

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 single IMF, but may not work well when there are two or more IMFs. For the signal with multiple IMFs and noise,

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

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 ,

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

where b_0, b_k, ϑ_k are approximations to a_0, a_k, 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 $b_0, b_1, \vartheta_1, \dots, b_K, \vartheta_K$ given in (4.6).

Parameter: A predesigned positive integer L , and $\eta = \min \left\{ \frac{\vartheta_k(T) - \vartheta_k(0)}{2\pi} : k = 1, 2, \dots, K \right\}$, 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 $b_0(t), r(t)$ 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 [2,22] and 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 has symmetry about the vertical line $t = -\frac{T}{2}$. We are going to solve the following optimization problem to get an approximation of the trend:

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

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

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

Step 2. If $\frac{\lambda_0 \cdot T}{2\pi} \leq \frac{N_p}{2}$, 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 $K = k$, $r(t) = r^k(t)$. Apply Algorithm Algorithm (End Effects) to $\{a_0, a_1, \theta_1, \dots, a_K, \theta_K\}$.

Output: The approximate decomposition of f is

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

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

Fig. 3 shows the IMFs g_0, g_1, g_2 given by the two-level method (with $L = 67$ in Algorithm (End Effects)) for the signal given in (1.5). Comparing with the results shown in Fig. 2, 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, g_r obtained by two-level method with L ranges from 0 to 100 (see Fig. 4). We see that the first several iterations make the greatest improvement.

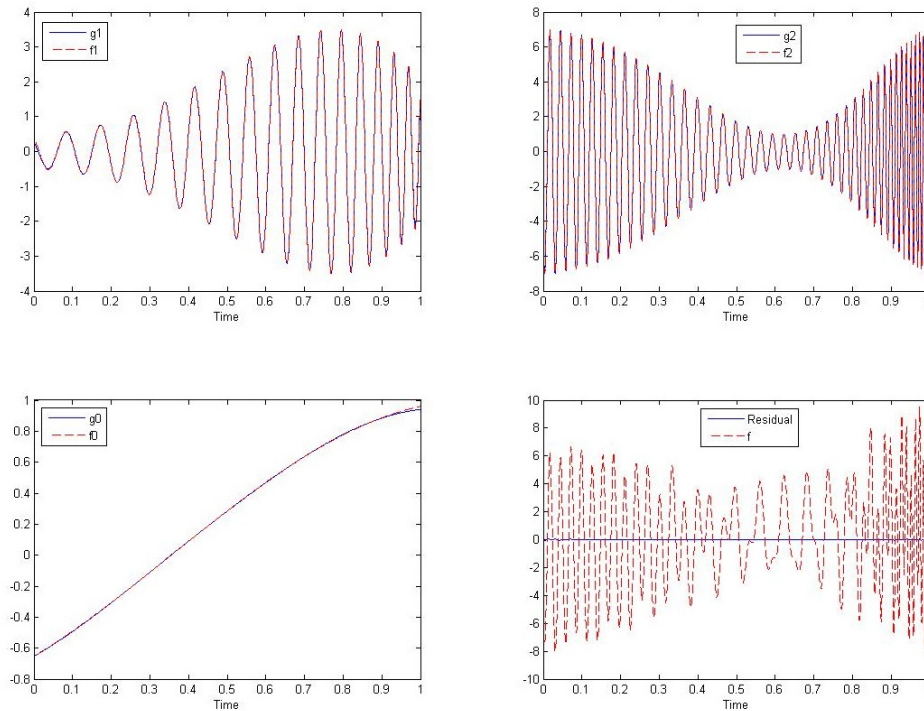


Figure 3. IMFs given by the two-level method for the signal in (1.5). Top left: first IMF; Top right: second IMF; Bottom left: trend. Bottom right: residual.

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:

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

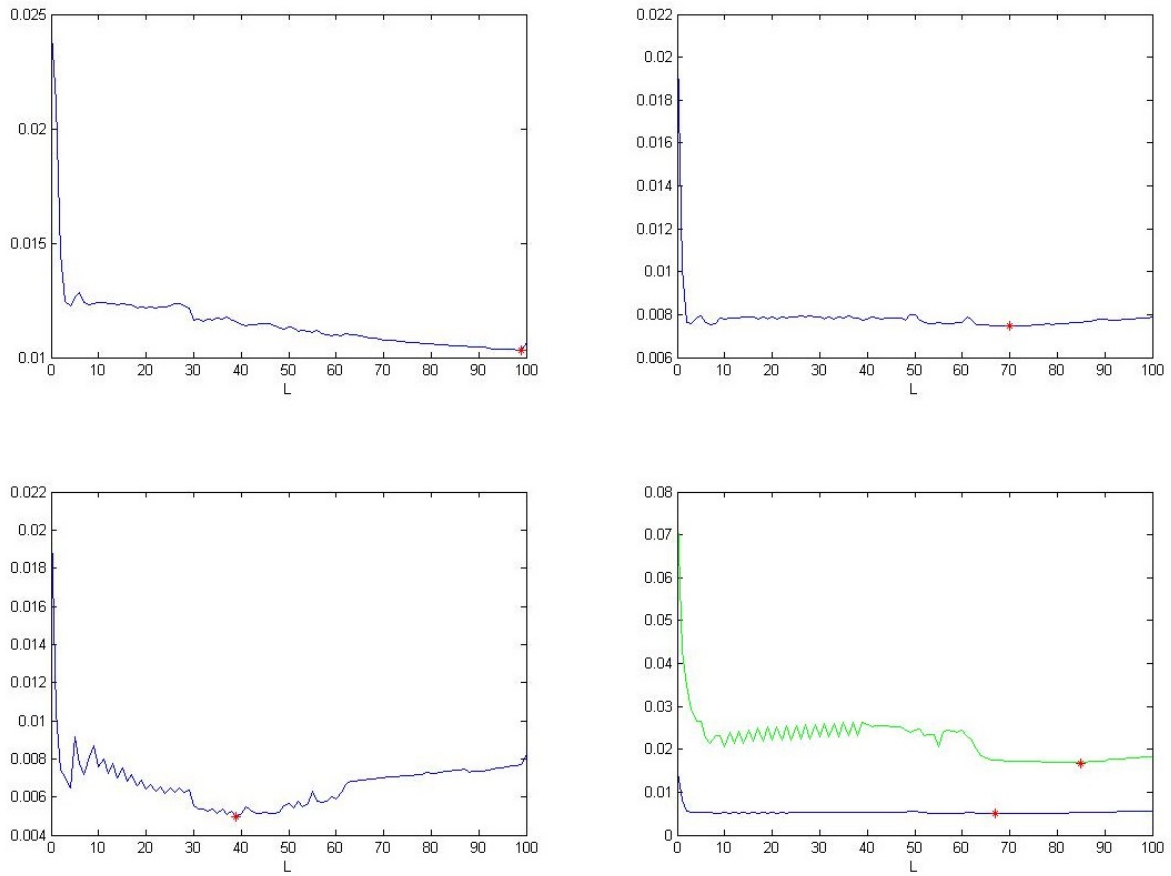


Figure 4. Topleft: $\frac{\|g_1-f_1\|_{L^2}}{\|f_1\|_{L^2}}$. The minimum 0.0103 is attained at $L = 99$. Topright: $\frac{\|g_2-f_2\|_{L^2}}{\|f_2\|_{L^2}}$. The minimum 0.0075 is attained at $L = 70$. Bottomleft: $\frac{\|g_0-f_0\|_{L^2}}{\|f_0\|_{L^2}}$. The minimum 0.0049 is attained at $L = 39$. Bottomright: $\frac{\max|g_r|}{\max|f|}$ (upper line) and $\frac{\|g_r\|_{L^2}}{\|f\|_{L^2}}$ (lower line). The minima 0.0168 and 0.0050 are attained at $L = 85$ and $L = 67$ respectively.

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

$$a_1 = a_3 = 2 - 1.5 \cos(1.5\pi s),$$

$$a_2 = 4 + 3 \cos(2\pi s)$$

$$\phi_1 = 32\pi s,$$

$$\phi_2 = 84\pi s + 10.4 \sin(2\pi s) + 1.32 \sin(4\pi s) + 1$$

$$\phi_3 = 160\pi s + 20 \sin(2\pi s) + 1.2 \sin(4\pi s) + 2,$$

and $b_0 = \cos(2t + 4)$,

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

The signal $f(t)$ is sampled over 1024 uniform grid points in $[0, 1]$.

The IMFs computed by the two-level algorithm are shown in Fig. 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 Fig. 6.

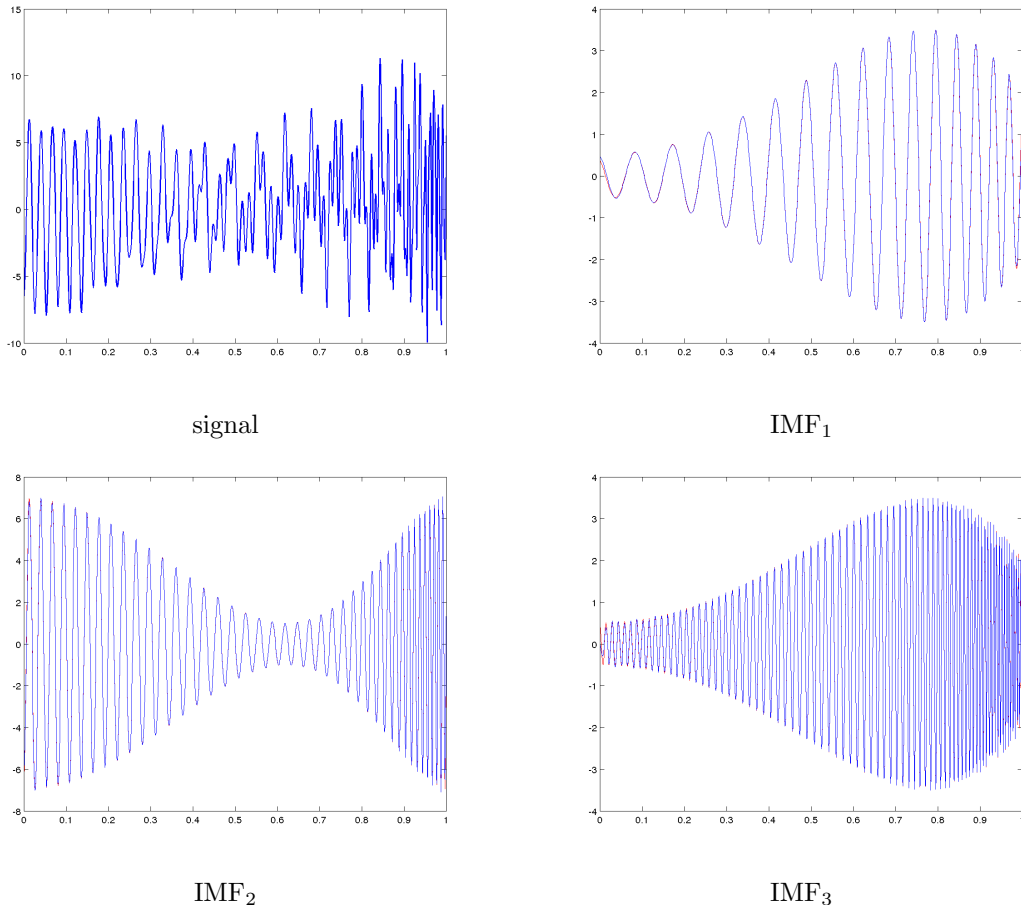


Figure 5. Original signal and IMFs in Example 6.1. Red dash lines: exact IMFs; blue solid lines: IMFs obtained by the two-level algorithm.

We also test the two-level algorithm for the signal 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 Fig. 7. As we can see, even with noise, the two-level algorithm still give IMFs with reasonable accuracy. In this case, the EMD method suffers from severe mode mixing as shown in Fig. 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.

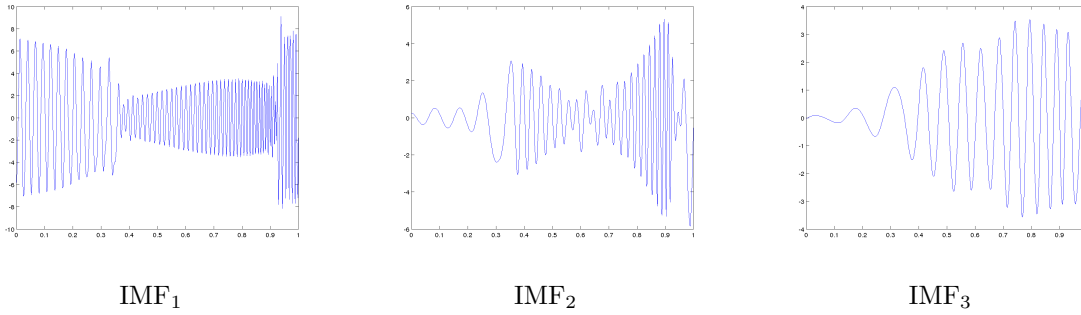


Figure 6 IMFs given by EMD in Example 6.1.

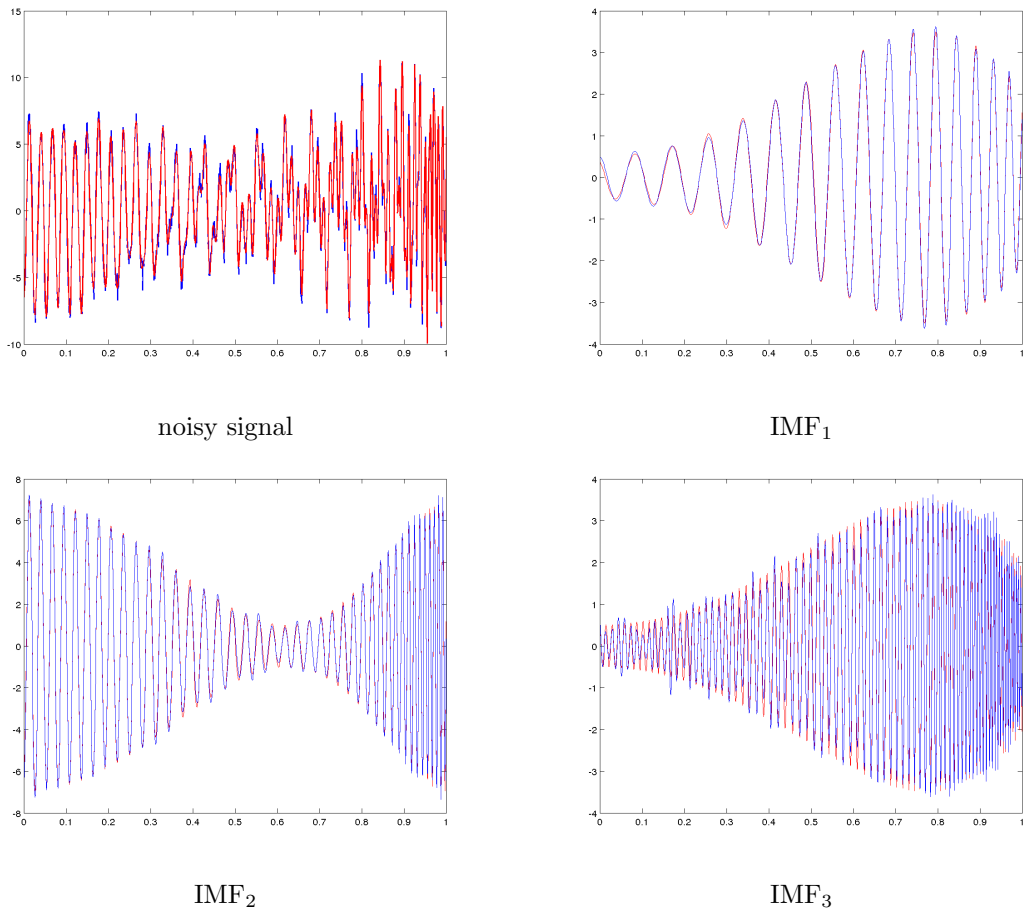


Figure 7. A Noisy signal and IMFs obtained by the two-level algorithm in Example 6.1. Red dash lines: exact IMF; blue solid lines: IMF given by the two-level algorithm.

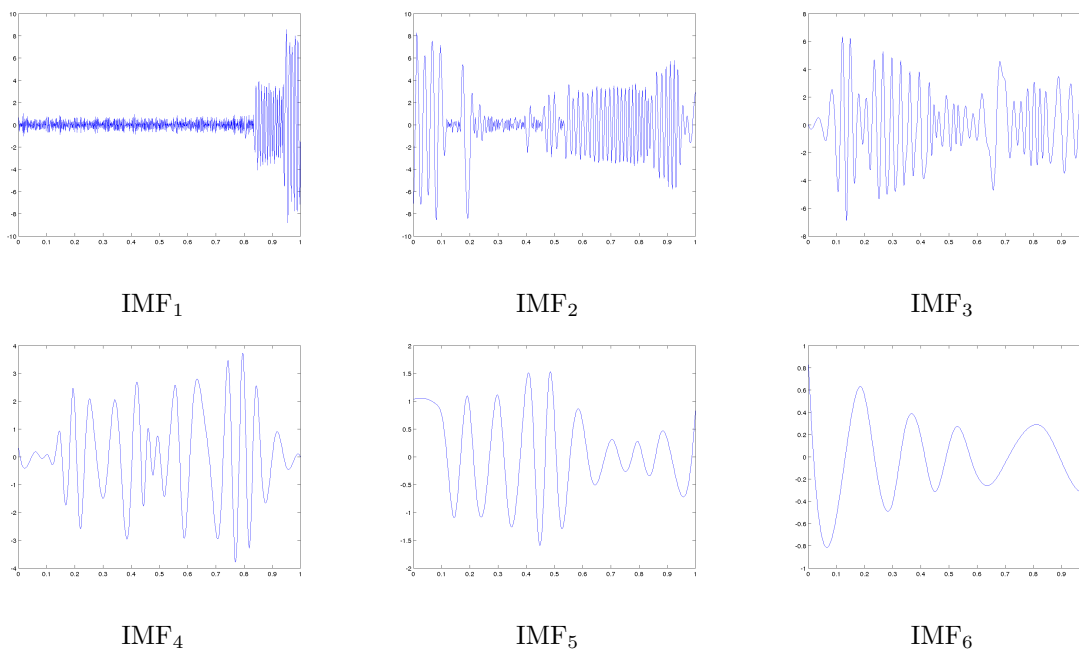


Figure 8 IMFs given by EMD method for a noisy signal given in Example 6.1.

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 to handle the end effect. In some sense, end effect alleviation is related to the prediction of the signal. Actually, from the two-level method, we can get the prediction of the IMFs 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 instance, global warming. In the following test, we use the signal in $[0, 0.9]$ to predict the trend in $[0.9, 1]$. Fig. 9 shows the results of the prediction. Recall that L is the number of iterations in the two-level method. $L = 0$ means no two-level iteration. We can see that, in this case, the end effect is very large and the prediction is very bad. 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 becomes even better. At the same time, the computational cost also become higher. Based on our experience, $L = 3$ seems to be a good balance between accuracy and efficiency.

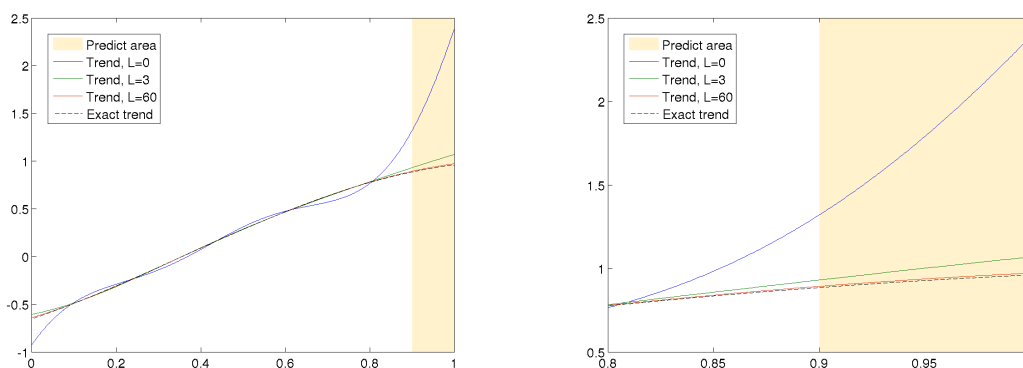


Figure 9 Prediction of trend using 90% data. Left: entire interval $[0, 1]$; right: zoom in on $[0.8, 1]$

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 [14]. This set of data covers the time interval from Jan. 20, 1962 to Jan. 6, 2001, totally 14,232 days, nearly 39 years. Using this data, we will do 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 4097 days, from Jan. 20, 1962 to Apr. 08, 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 Nov. 09, 1964 to Jun. 19, 1970, i.e., the 1025th day to the 3073th day) and regard these restricted IMFs as nearly true IMFs. Then we applying 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, applying the two-level method to the data in Task I, we obtain 10 IMFs (see Fig. 11). In the 10 IMFs, the first 4 IMFs (g_{10} - the trend, g_2, g_5, g_1) contain over 92.6% of the energy of the original signal. And for the residual g_r , we have

$$\frac{\|g_r\|_{L^2}}{\|f\|_{L^2}} \approx 0.0323, \quad \frac{\max |g_r|}{\max |f|} \approx 0.0619.$$

And Fig. 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.

Second, to test how well the two-level method handle 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 g_k^a , ($k = 1, \dots, 8$). Except for the trend, the three IMFs with largest energies, g_1^a, g_2^a, g_3^a , have similar frequencies with g_2, g_5, g_1 respectively, and on the time interval from Nov. 09, 1964 to Jun. 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 the Length-of-Day Data from the first 4097 days and from the 1025th day to the 3073th day respectively, which give IMFs $h_k(k = 1, \dots, 13)$ and $h_k^a(k = 1, \dots, 12)$, respectively. The IMFs h_1, h_4, h_5 have the largest energies except for the trend, and their frequencies are similar to those of g_2, g_5, g_1 , respectively. Their corresponding parts in $\{h_k^a\}$ are h_1^a, h_4^a, h_5^a . On the time interval from Nov. 09, 1964 to Jun. 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, g_3^a are smaller than h_1^a, h_4^a, 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 Fig. 13 and 14).

7 Concluding Remarks

In this paper, we introduced a two-level method based on the data-driven time-frequency analysis [16, 17] 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

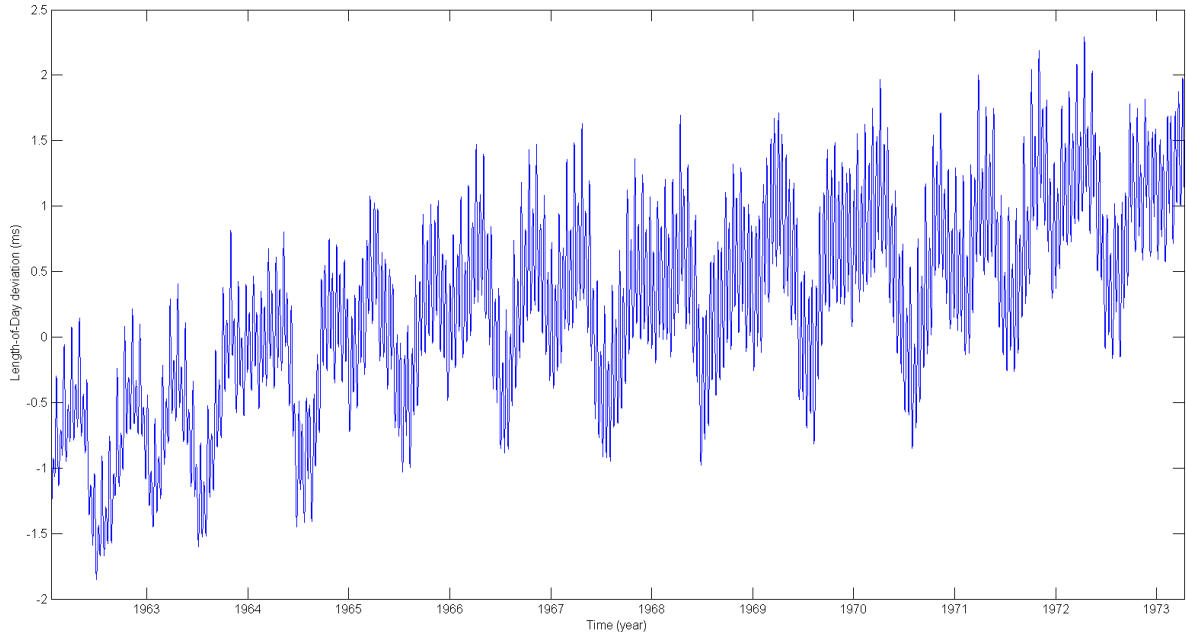


Figure 10 Length-of-Day data form Jan. 20, 1962 to Apr. 8, 1973.

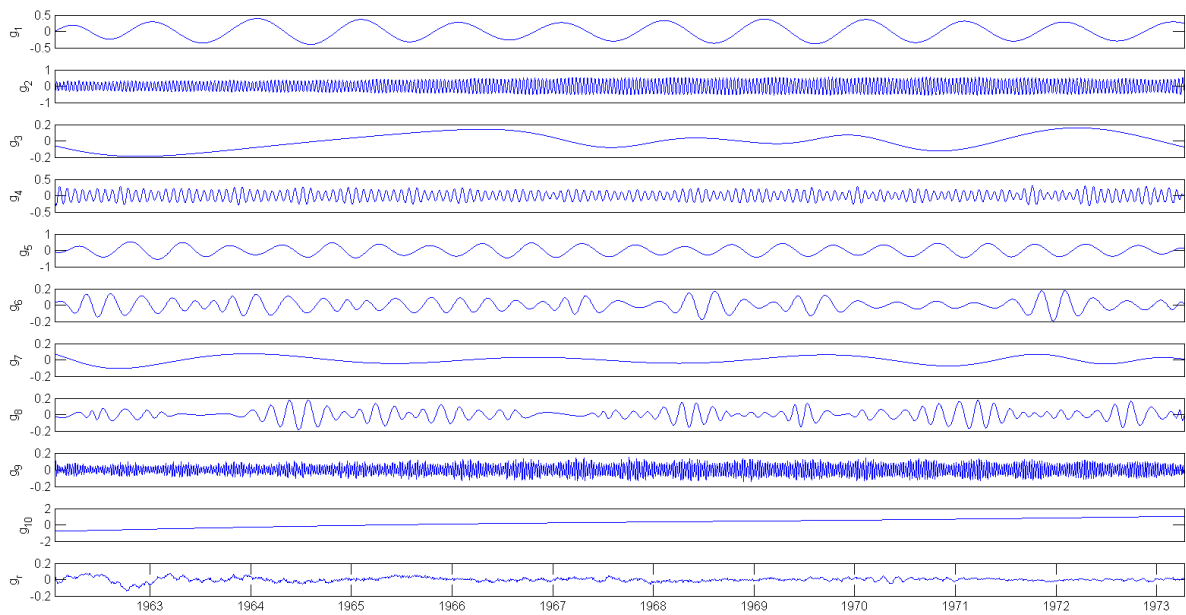


Figure 11 The 10 modes and the residual g_r obtained by the two-level method on the Length-of-Day data.

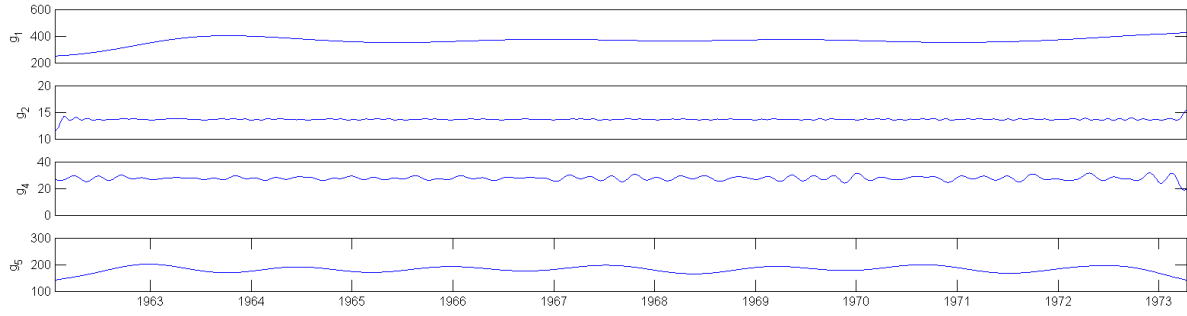


Figure 12. Instantaneous periods of g_1, g_2, g_4, g_5 . The average periods of these IMFs are approximately 362.3, 13.7, 27.5, 181.1 respectively.

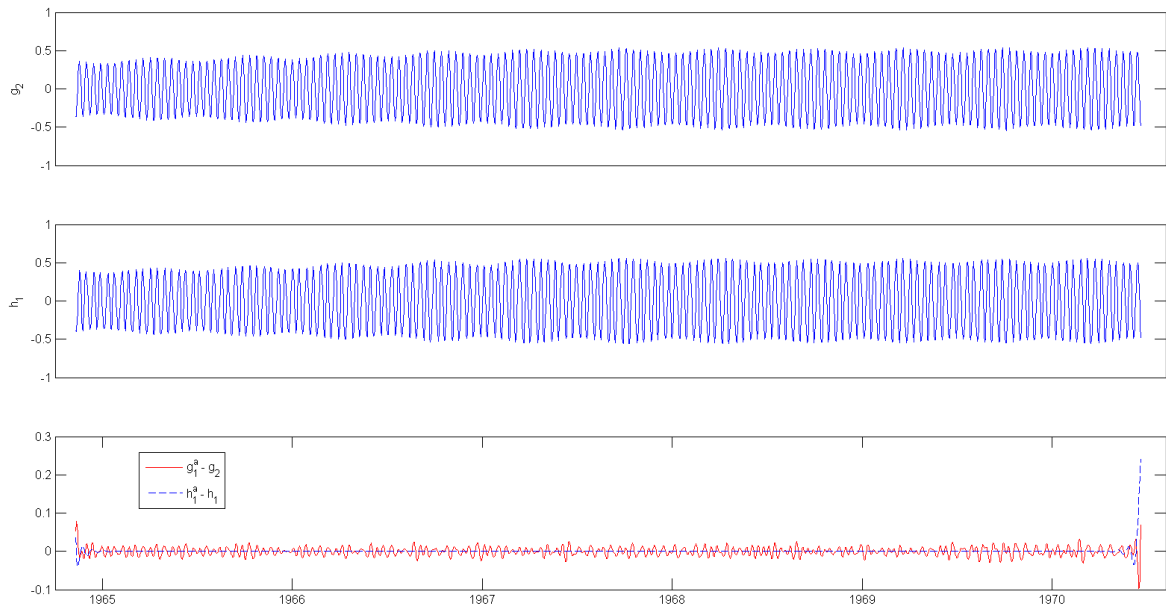


Figure 13 Top: g_2 ; middle: h_1 ; bottom: red solid line: $g_1^a - g_2$, blue dash line: $h_1^a - h_1$.

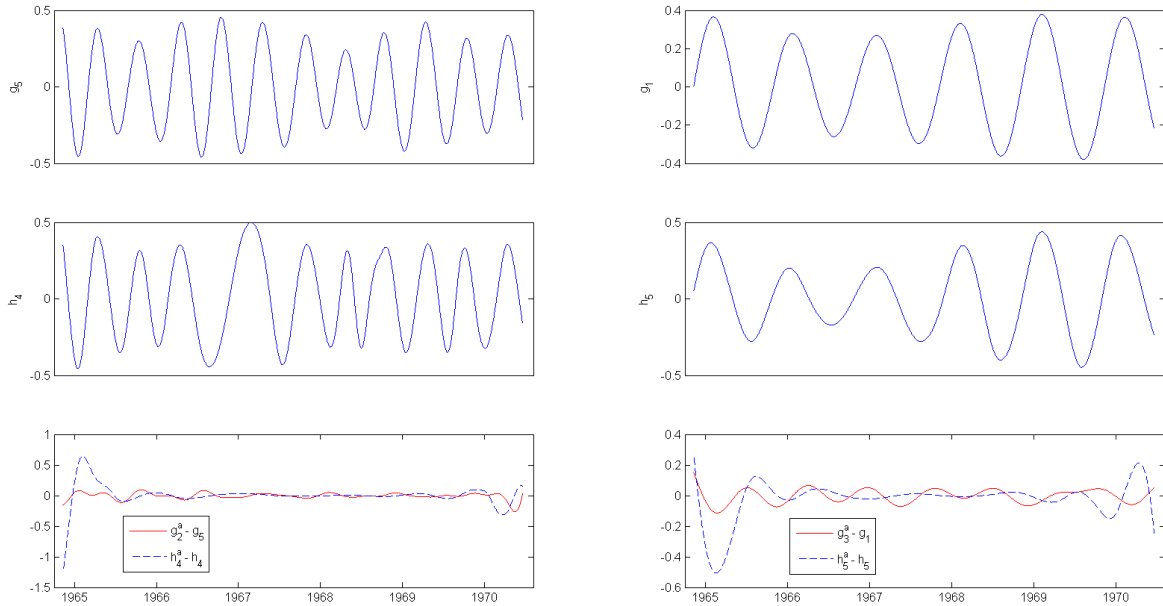


Figure 14. Top left: g_5 ; top right: g_1 ; middle left: h_4 ; middle right: h_5 ; bottom left: red solid line: $g_2^a - g_5$, blue dash line: $h_4^a - h_4$; bottom right: red solid line: $g_3^a - g_1$, blue dash line: $h_5^a - h_5$.

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.

Acknowledgments. This research of Dr. T. Y. Hou was supported in part by NSF FRG Grant DMS-1318377. The research of Dr. Z. Shi was supported by a NSFC Grant 11371220 and 11671005. The research of Dr. C. Liu was supported by a NSFC Grant 11371173, 11301222 and 11526096.

References

- 1 B. Boashash. *Time-Frequency Signal Analysis: Methods and Applications*. Longman-Cheshire, Melbourne and John Wiley Halsted Press, New York, 1992.
- 2 J. P. Boyd. A comparison of numerical algorithms for fourier extension of the first, second, and third kinds. *Journal of Computational Physics*, 178:118–160, 2002.
- 3 A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51:34–81, 2009.
- 4 E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal recovery from highly incomplete frequency information. *IEEE Trans. on Information Theory*, 52:489–509, 2006.
- 5 I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series on Applied Mathematics, Vol. 61, SIAM Publications, 1992.
- 6 I. Daubechies, J. Lu, and H. Wu. Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. *Appl. Comp. Harmonic Anal.*, 30:243–261, 2011.
- 7 I. Daubechies, Y. Wang, and H. Wu. Conceft: concentration of frequency and time via a multitapered synchrosqueezed

- transform. *Philos. Trans. A*, 374:(DOI: 10.1098/rsta.2015.0193), 2016.
- 8 B. Van der Pol. The fundamental principles of frequency modulation. *Proc. IEE*, 93:153–158, 1946.
 - 9 D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52:1289–1306, 2006.
 - 10 N. E. Huang et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A*, 454:903–995, 1998.
 - 11 P. Flandrin. *Time-Frequency/Time-Scale Analysis*. Academic Press, San Diego, CA, 1999.
 - 12 D. Gabor. Theory of communication. *J. IEE.*, 93:426–457, 1946.
 - 13 R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49:3320–3325, 2003.
 - 14 R. S. Gross. Combinations of earth orientation measurements: Space2000, comb2000, and pole2000. jpl publication 01-2. jet propulsion laboratory, pasadena, ca. 01-2.
 - 15 T. Y. Hou and Z. Shi. Adaptive data analysis via sparse time-frequency representation. *Advances in Adaptive Data Analysis*, 3:1–28, 2011.
 - 16 T. Y. Hou and Z. Shi. Data-drive time-frequency analysis. *Applied and Comput. Harmonic Analysis*, 35:284–308, 2013.
 - 17 T. Y. Hou and Z. Shi. Sparse time-frequency representation of nonlinear and nonstationary data. *Science China Mathematics*, 56:2489–2506, 2013.
 - 18 T. Y. Hou and Z. Shi. Sparse time-frequency decomposition by dictionary adaptation. *Philosophical Transactions A*, 374:20150194, 2016.
 - 19 T. Y. Hou, Z. Shi, and P. Tavallali. Convergence of a data-driven time-frequency analysis method. *Applied and Comput. Harmonic Analysis*, 37:235–270, 2014.
 - 20 N. E. Huang, I. Daubechies, and T. Y. Hou. Adaptive data analysis: theory and applications. *Philos. Trans. A*, 374:(DOI: 10.1098/rsta.2015.0207), 2016.
 - 21 N. E. Huang and Z. Wu. A review on hilbert-huang transform: the method and its applications on geophysical studies. *Rev. Geophys.*, 46(2):RG2006, doi:10.1029/2007RG000228, 2008.
 - 22 D. Huybrechs. On the fourier extension of non periodic functions. *SIAM J. Numer. Anal.*, 47:4326–4355, 2010.
 - 23 D. L. Jomes and T. W. Parks. A high resolution data-adaptive time-frequency representation. *IEEE Trans. Acoust. Speech Signal Process*, 38:2127–2135, 1990.
 - 24 C. G. Liu, Z. Q. Shi, and T. Y. Hou. On the uniqueness of sparse time-frequency representation of multiscale data. *Multiscale Model. Simul.*, 13:790–811, 2015.
 - 25 P. J. Loughlin and B. Tracer. On the amplitude - and frequency-modulation decomposition of signals. *J. Acoust. Soc. Am.*, 100:1594–1601, 1996.
 - 26 B. C. Lovell, R. C. Williamson, and B. Boashash. The relationship between instantaneous frequency and time-frequency representations. *J. Acoust. Soc. Am.*, 41:1458–1461, 1993.
 - 27 S. Mallat. *A wavelet tour of signal processing: the Sparse way*. Academic Press, 2009.
 - 28 S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Trans. Signal Process*, 41:3397–3415, 1993.
 - 29 W. K. Meville. Wave modulation and breakdown. *J. Fluid Mech.*, 128:489–506, 1983.
 - 30 D. Needell and J. Tropp. Cosamp: Iterative signal recovery from noisy samples. *Appl. Comput. Harmon. Anal.*, 26:301–321, 2008.
 - 31 S. Olhede and A. T. Walden. The hilbert spectrum via wavelet projections. *Proc. Roy. Soc. London A*, 460:955–975, 2004.
 - 32 B. Picinbono. On instantaneous amplitude and phase signals. *IEEE Trans. Signal Process*, 45:552–560, 1997.
 - 33 S. Qian and D. Chen. *Joint Time-Frequency Analysis: Methods and Applications*. Prentice Hall, 1996.
 - 34 S. O. Rice. Mathematical analysis of random noise. *Bell Syst. Tech. J.*, 23:282–310, 1944.
 - 35 J. Shekel. Instantaneous frequency. *Proc. IRE*, 41:548, 1953.
 - 36 J. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, 53:4655–4666, 2007.
 - 37 Z. Wu and N. E. Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1:1–41, 2009.
 - 38 Z. Wu, N. E. Huang, and X. Chen. The multi-dimensional ensemble empirical mode decomposition method. *Advances in Adaptive Data Analysis*, 1(3):339–372, 2009.
 - 39 Z. Wu, N. E. Huang, S. R. Long, and C. K. Peng. On the trend, detrending, and variability of nonlinear and nonstationary time series. *PNAS.*, 104(38):14889–14894, 2007.