Advances in Adaptive Data Analysis
Vol. 3, Nos. 1 & 2 (2011) 1–28

DOI: 10.1142/S1793536911000647

ADAPTIVE DATA ANALYSIS VIA SPARSE TIME-FREQUENCY REPRESENTATION

THOMAS Y. HOU* and ZUOQIANG SHI†

Applied and Computational Mathematics,
Caltech, Pasadena, CA 91125, USA
**hou@cms.caltech.edu*
†*shi@cms.caltech.edu*

We introduce a new adaptive method for analyzing nonlinear and nonstationary data. This method is inspired by the empirical mode decomposition (EMD) method and the recently developed compressed sensing theory. The main idea is to look for the sparsest representation of multiscale data within the largest possible dictionary consisting of intrinsic mode functions of the form $\{a(t)\cos(\theta(t))\}$, where $a \geq 0$ is assumed to be smoother than $\cos(\theta(t))$ and θ is a piecewise smooth increasing function. We formulate this problem as a nonlinear L^1 optimization problem. Further, we propose an iterative algorithm to solve this nonlinear optimization problem recursively. We also introduce an adaptive filter method to decompose data with noise. Numerical examples are given to demonstrate the robustness of our method and comparison is made with the EMD method. One advantage of performing such a decomposition is to preserve some intrinsic physical property of the signal, such as trend and instantaneous frequency. Our method shares many important properties of the original EMD method. Because our method is based on a solid mathematical formulation, its performance does not depend on numerical parameters such as the number of shifting or stop criterion, which seem to have a major effect on the original EMD method. Our method is also less sensitive to noise perturbation and the end effect compared with the original EMD method.

Keywords: Time-frequency analysis; instantaneous frequency; empirical mode decomposition; sparse representation of signal; L^1 minimization.

1. Introduction

Data are one of the most important links that we have with the physical world. Developing effective data analysis methods is an important path through which we can understand the underlying processes of natural phenomena. So far, most data analysis methods use a pre-determined basis to process data. These methods often assume linearity and stationarity of data. Time-frequency analysis has been developed to overcome the limitations of the traditional techniques by representing a signal with a joint function of both time and frequency. Time-frequency analysis provides a revealing picture in the time-frequency domain and can be applied to study nonstationary and nonlinear signals. The recent advances of wavelet analysis opened a new path for time-frequency analysis. A significant breakthrough of

wavelet analysis was the use of multiscales to characterize signals. This technique has led to the development of several wavelet-based time-frequency analysis techniques [Daubechies (1992); Jomes and Parks (1990); Mallat (2009)].

In real-world experimental and theoretical studies, we often deal with signals that have ever-changing frequency. A typical example is chirp signals used by bats as well as in radar. To better understand the physical mechanisms hidden in data, one needs to develop effective methods that can handle the nonstationarity and nonlinearity of the data. Such methods should be adaptive to the nature of the data, which requires the use of an adaptive basis that is not determined *a priori*. Instead, the basis should be derived from the data.

An important development in 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 (1946) and Gabor (1946), who introduced the so-called analytic signal (AS) method which uses the Hilbert transform to determine instantaneous frequency of a signal. This AS method has received a lot of attention and is one of the most popular ways to define instantaneous frequency. Until very recently, this method works mostly for monocomponent signals in which the number of zero-crossings is equal to the number of local extrema [Boashash (1992)]. There were other attempts to define instantaneous frequency such as the zero-crossing method [Meville (1983); Rice (1944); Shekel (1953)] and the Wigner–Ville distribution method [Boashash (1992); Lovell *et al.* (1993); Qian and Chen (1996); Flandrin (1999); Loughlin and Tracer (1996); Picinbono (1997)]. However, most of these methods are rather restrictive. More substantial progress has been made only recently with the introduction of the empirical mode distribution (EMD) method [Huang *et al.* (1998)] and the Hilbert spectral representation based on the wavelet projection [Olhede and Walden (2004)].

The main idea of EMD is to first compute the local median of a signal via a shifting procedure and then subtract the local median from the signal before applying the AS method to define its instantaneous frequency. The EMD method provides a powerful tool to decompose a signal into a collection of intrinsic mode functions (IMFs) that allow well behaved Hilbert transforms for computation of physically meaningful time-frequency representation. In spite of its considerable success, there is still a lack of mathematical understanding of the EMD method such as its convergence property and dependence on the number of shifting, the stopping criteria, and its stability to noise perturbation. We remark that there has been some recent progress in developing a mathematical framework for an EMD like method using synchrosqueezed wavelet transforms by Daubechiesa *et al.* (2011), see also the paper entitled "One or Two Frequencies? The Synchrosqueezing Answers" by Wu, Flandrin, and Daubechies in this same special issue of AADA as our paper. This is a very interesting line of work. For the examples they consider, their method produces excellent results.

Inspired by the EMD method and the recently developed compressed sensing theory [Candes and Tao (2006); Candes *et al.* (2006a); Donoho (2006)], we propose a new adaptive data analysis method. This method has a beautiful mathematical

structure and is fully adaptive to the data. It can be seen as a nonlinear version of compressed sensing and provides a mathematical foundation of the EMD method.

Our adaptive data analysis method is motivated by the observation that the multiscale data have an intrinsic sparse structure in the time-frequency plane, although its representation in the physical domain could be rather complicated. The challenge is that such sparsity structure is valid only for certain multiscale basis, which is adapted to the data and is unknown *a priori*. Thus, one of the main challenges is to find such nonlinear multiscale basis under which the multiscale data have a sparse representation. This is very different from the compressed sensing problem because the basis under which the data have a sparse representation is assumed to be known *a priori*. Traditionally, the adaptive basis is derived by learning the data. This approach requires a large number of data samples that share the similar physical property. This does not apply to our problem since we deal with only a single signal. In our approach, the adaptivity is achieved by adopting the largest possible dictionary. Then, the decomposition has enough freedom to choose the basis from this dictionary that provides the best match to the data. The trade-off is that the decomposition is not unique. We need to exploit the intrinsic sparse structure of the data to select the best one among all the possible decompositions.

Our method consists of two steps. First, we construct a highly redundant dictionary:

$$\mathcal{D} = \{a(t)\cos\theta(t) : \theta'(t) \geq 0,\ a(t) \text{ is smoother than } \cos\theta(t)\}\,. \tag{1}$$

Then, the signal is decomposed over this dictionary by looking for the sparest decomposition. The sparest decomposition can be obtained by solving a nonlinear optimization problem:

$$\begin{aligned} P: \quad & \text{Minimize} \quad M \\ & \text{Subject to:} \quad f(t) = \sum_{k=1}^{M} a_k(t)\cos\theta_k(t), \quad a_k(t)\cos\theta_k(t) \in \mathcal{D}, \quad k = 1, \ldots, M. \end{aligned} \tag{2}$$

This optimization problem can be seen as a nonlinear version of the l_0 minimization problem which has been studied extensively in the compressed sensing literature [Bruckstein *et al.* (2009)]. By generalizing the numerical method to solve the l_0 minimization problem, we propose an iterative algorithm to solve the above nonlinear optimization problem.

The above nonlinear optimization problem is very challenging. We introduce a recursive iterative scheme to solve this nonlinear optimization problem. Like the EMD method, we would like to first decompose a signal into its local median, $a_0(x)$, and its fluctuation (IMF), $a_1(x)\cos(\theta_1(x))$, with $a_1\cos(\theta_1) \in \mathcal{D}$. One way to impose sparsity of the decomposition, i.e., minimizing M, is to find the smoothest possible local median, a_0, so that $a_1\cos(\theta_1) \in \mathcal{D}$. If we can find the smoothest possible a_0 to decompose $f = a_0 + a_1\cos(\theta_1)$, then it is reasonable to expect that

we can decompose the local median, a_0, into smallest number of IMFs $a_j \cos(\theta_j)$ $(j = 2, \ldots, M)$ over $\mathcal{D}$. This would give rise to a recursive iterative scheme to find a sparse decomposition of f into its IMFs over $\mathcal{D}$.

One way to measure smoothness of a_0 is to use the total variation norm, which is defined as the L^1 norm of its first derivative. Another reason for using an L^1 norm is that L^1 minimization tends to give a sparse representation of data [Bruckstein *et al.* (2009); Candes and Tao (2006); Candes *et al.* (2006a); Candes *et al.* (2006b); Donoho (2006)]. However, it is well known that minimizing the total variation of a signal tends to produce a piecewise constant function, the so-called staircase effect. In order to preserve some high-order information (e.g., curvature) of a signal, we propose to use the third-order total variation, which is defined as the total variation of the third derivative of a function, to measure smoothness. This produces a much better result. Incidentally, the third-order total variation tends to favor cubic spline interpolations of a_0 and a_1. As a result, our method can reproduce some of the best results obtained by the EMD method in many cases.

One drawback of using a high-order total variation norm in our optimization problem is that it is more sensitive to noise. This is also the case for the original EMD method since a_0 and a_1 are approximated by interpolating the local extrema of the signal by using cubic splines. To overcome this sensitivity to noise, we develop a nonlinear adaptive filter and couple it with our iterative optimization solver. This considerably reduces the sensitivity of our method to noise. We also make a special effort to reduce the approximation error near the two end-points of a signal, the so-called end-point effect. For our method, this amounts to finding a good initial guess for the phase function θ for our iterative scheme.

An important issue in the implementation of our method is to use an effective L^1 minimization solver since we need to solve a discrete L^1 minimization problem within each nonlinear iteration. Like compressed sensing, the performance of our method depends on the efficient implementation of the L^1 minimization. We have applied both the interior point method for large-scale L^1-regularized least square method developed recently by Kim *et al.* (2007) and the split Bregman iteration developed by Goldstein and Osher (2009). For large-scale data, we find that the split Bregman iteration is more efficient than the L^1-regularized least square method.

We perform extensive numerical experiments to test the convergence and the accuracy of our method for both synthetic data and some real data. Our results show that the L^1-based nonlinear optimization can indeed decompose a multiscale signal into a sparse collection of IMF. For those data that satisfy certain scale separation condition, our method can recover the IMFs and their instantaneous frequencies accurately. We also compare our method with the original EMD method. In most cases, we find that our method gives results that are either comparable to or more superior than those obtained by the EMD method. In comparison with the EMD method, our method has the advantage of being insensitive to numerical parameters such as the number of shifting or the stopping criterion. These parameters seem to

have a major effect on the performance of the EMD method. Moreover, our method has a better stability property for noisy data than the EMD method. For the data we consider here, our method seems to provide better accuracy in approximating the instantaneous frequency of noisy data than the recently developed EEMD method [Wu and Huang (2005, 2009)].

The remaining part of the paper is organized as follows. We review the concept of instantaneous frequency and the EMD method in Sec. 2. In Sec. 3, we introduce our L^1-based nonlinear optimization method, its numerical algorithm, and provide some details of its implementation issues. In Sec. 4, we illustrate the convergence property of our method for various nonlinear, nonstationary data. In Sec. 5, we introduce our adaptive filter to decompose noisy data and demonstrate the robustness of the modified nonlinear optimization method which uses the adaptive filter within each iteration. Some concluding remarks are made in Sec. 6.

2. A Brief Review of the AS Method and the EMD Method

In this section, we give a brief review of the AS method and the EMD method. The EMD method was motivated by the AS method to some extent and our method is in turn inspired by the EMD method. Thus, it is natural for us to first understand the main ideas behind these two methods.

2.1. *The AS method*

The concept of instantaneous frequency has been used in adaptive signal analysis for many years. Some of the pioneering work in this area was due to Van der Pol (1946) and Gabor (1946), who introduced the so-called AS method to determine instantaneous frequency of a signal. Gabor's approach is summarized as follows: given a signal $x(t)$, we define its imaginary part through the Hilbert transform, i.e., $y(t) = H(x)(t)$. Then, we can express the original signal as the real part of an AS, $z(t)$

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)},$$

where $a(t) = \sqrt{x^2(t) + y^2(t)}$ and $\theta(t) = \tan^{-1}\frac{y(t)}{x(t)}$. The instantaneous frequency is then defined as $\omega(t) = \frac{d}{dt}\theta(t)$. This AS method has received a lot of attention and is one of the most popular ways to define instantaneous frequency. Until very recently, this method works mostly for monocomponent signals in which the number of zero crossings is equal to the number of local extrema [Boashash (1992)].

There are several difficulties in applying the AS method to extract the instantaneous frequency. First of all, not all the data are monocomponent. One has to remove the local median or local trend before one applies the Hilbert transform to the signal. Even though decomposing the data into a collection of monocomponent functions is now available by wavelet decomposition [Olhede and Walden (2004)]

or the EMD method [Huang *et al.* (1998)], there are other difficulties. One of the most serious ones is that the AS method implicitly assumes that

$$H(a(t)\cos(\theta(t))) = a(t)H(\cos(\theta(t)). \tag{3}$$

This is in general not valid unless the Fourier spectra of the envelope $a(t)$ and the carrier $\cos(\theta(t))$ are nonoverlapping as pointed out by Bedrosian (1963). This imposes a much sharper condition on the data: the data have to be not only monocomponent, but also narrow band. A more fundamental difficulty is that even if $a(t) = 1$, we know that $H(\cos(\theta(t))) = \sin(\theta(t))$ is not true for arbitrary function $\theta(t)$ as pointed out by Nuttall (1966) unless $\theta(t)$ is linear. This difficulty has been ignored by most investigators using the Hilbert transform to compute instantaneous frequency [Huang *et al.* (2009)].

2.2. *The EMD method*

The EMD method [Huang *et al.* (1998)] is an adaptive, temporally local data analysis method. We refer to [Huang *et al.* (2009, 1999); Wu and Huang (2005, 2009); Wu *et al.* (2007)] for more detailed discussions on EMD and its latest developments. The main idea of EMD is to first subtract the local median of a signal $x(t)$ before applying the AS method to define its instantaneous frequency. The EMD method provides an approximation to the local median via a shifting procedure. Specifically, the EMD method uses a cubic spline polynomial to interpolate all the local maxima of $x(t)$ to obtain an upper envelope, and a cubic spline to interpolate all the local minima to obtain a lower envelope, then average the upper and lower envelopes to obtain an approximate median $m_1(t)$. One then decides whether or not to accept the obtained $m_1(t)$ as our local median depending on whether $c_1(t) = x(t) - m_1(t)$ satisfies the following two conditions: (1) there must be one zero crossing between two local extrema and the number of zero crossings and the number of extrema must be equal and (2) c_1 is "symmetric" with respect to zero. If $x(t) - m_1(t)$ does not satisfy these conditions, one can treat $x(t) - m_1(t)$ as a new signal and repeat the same procedure until a satisfactory c_1 is found, which is defined as an IMF. This is called the shifting procedure.

Currently, the EMD method can be justified only under certain very restrictive assumptions that are seldom satisfied by practical data. The performance also depends sensitively on the number of shifting and the stopping criteria. The EMD method is also known to be very sensitive to noisy data. The recently introduced EEMD [Wu and Huang (2005, 2009)] has addressed some of these issues, but some essential difficulties remain.

3. Adaptive Data Analysis Based on the Sparsest Time-Frequency Representation of Signals

Our adaptive data analysis method is based on finding the sparsest decomposition of a signal by solving a nonlinear optimization problem. First, we need to construct

a large dictionary which can be used to obtain a sparse decomposition of a signal. In principle, the larger the dictionary is, the more adaptive (or sparser) the decomposition is. In this paper, we define the redundant dictionary as follows

$$\mathcal{D} = \{a(t)\cos\theta(t)\colon \theta'(t) \geq 0,\ a(t) \text{ is smoother than } \cos\theta(t)\}\,. \tag{4}$$

In some sense, the above dictionary can be seen as a collection of all possible IMFs, which makes our method as adaptive as the EMD method. Since the dictionary $\mathcal{D}$ is highly redundant, the decomposition over this dictionary is not unique. We need a criterion to pick up the "best" one. We observe that the multiscale data have an intrinsic sparse structure in the time-frequency plane, although its representation in the physical domain could be rather complicated. Based on this observation, we adopt sparsity as our criterion to choose the best decomposition. This criterion yields the following nonlinear optimization problem

$$\begin{aligned} P_0: \quad & \text{Minimize} \qquad M \\ & \text{Subject to:} \quad f(t) = \sum_{k=1}^{M} a_k(t)\cos\theta_k(t), \quad a_k(t)\cos\theta_k(t) \in \mathcal{D},\ \ k = 1, \ldots, M. \end{aligned} \tag{5}$$

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

$$\text{Instantaneous frequency:} \quad \omega_k(t) = \theta_k'(t), \quad \text{Amplitude: } a_k(t). \tag{6}$$

We remark that the EMD method typically decomposes a signal into a few IMFs, which provides a sparse decomposition of the signal implicitly.

3.1. *Adaptive decomposition based on a third-order total variation*

The nonlinear optimization (P_0) stated above is too difficult to solve numerically. In this section, we propose a recursive scheme to solve the nonlinear optimization problem (P_0) approximately.

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

$$\begin{aligned} & \text{Find the smoothest} \quad a_0(t) \\ & \text{Subject to:} \quad a_0(t) + a_1(t)\cos\theta_1(t) = f(t), \\ & \qquad\qquad \theta_1'(t) \geq 0, \quad a_1(t) \text{ is smoother than } \cos\theta_1(t). \end{aligned} \tag{7}$$

However, we need to give a quantitative measurement of smoothness in the above optimization problem (7) before we can solve it numerically.

A possible way to measure smoothness of a function is to minimize its total variation:

$$TV(g) = \int_a^b |g'(x)|dx. \tag{8}$$

The total variation norm has been used widely in shock capturing and PDE-based imaging analysis. On the other hand, it is also well known that minimizing the total variation would generate the "stair case." The stair case on the local median, a_0, introduces artificial high frequency information into the signal. To enforce a higher-order regularity of the local median, we propose to use a high-order total variation to measure smoothness. For this reason, we define the nth-order total variation as follows

$$TV^n(g) = \int_a^b |g^{(n+1)}(x)|dx, \tag{9}$$

where $g^{(n+1)}(x)$ is the $(n+1)$th derivative of g.

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

Now, the $TV^{(3)}$-based optimization problem (5) can be written in the following form

$$\begin{aligned} &\text{Minimize} && TV^3(a_0) \\ &\text{Subject to:} && a_0(t) + a_1(t)\cos\theta(t) = f(t), \\ &&& \theta'(t) \geq 0, a_1 \text{ is smoother than } \cos\theta(t). \end{aligned} \tag{10}$$

On the other hand, we need to enforce the condition that a_1 is smoother than $\cos(\theta_1)$. This can be done by using a Lagrangian multiplier approach. We choose the Lagrangian multiplier parameter to be one and reformulate the above optimization problem into the following form

$$\begin{aligned} (P)\quad &\text{Minimize} && TV^3(a_0) + TV^3(a_1), \\ &\text{Subject to:} && a_0(t) + a_1(t)\cos\theta(t) = f(t), \ \ \theta'(t) \geq 0. \end{aligned} \tag{11}$$

In the next section, we propose an iterative algorithm to solve this nonlinear optimization problem.

3.2. *An iterative algorithm*

In this section, we introduce a Newton type of iteration method to solve the nonlinear optimization problem (11) proposed in the previous section.

Initialization: $\theta^0 = \theta_0$.

Main iteration:

Step 1: Update a_0^n, a_1^n, and b_1^n by solving the following linear optimization problem:

$$\text{Minimize} \quad TV^3(a_0^n) + TV^3(a_1^n) + TV^3(b_1^n), \tag{12}$$

$$\text{Subject to:} \quad a_0^n + a_1^n \cos\theta^{n-1}(t) + b_1^n \sin\theta^{n-1}(t) = f(t). \tag{13}$$

Step 2: Update the phase function θ:

$$\theta^n = \theta^{n-1} - \mu \arctan\left(\frac{b_1^n}{a_1^n}\right), \tag{14}$$

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

$$\mu = \max\left\{\alpha \in [0,1] : \frac{d}{dt}\left(\theta_k^{n-1} + \alpha \arctan\left(\frac{b_1^n}{a_1^n}\right)\right) \geq 0\right\}. \tag{15}$$

Step 3: If $\|\theta^n - \theta^{n-1}\|_2 \leq \epsilon_0$, stop. Otherwise, go to Step 1.

3.3. *Normalization to obtain an initial guess for θ*

It remains to find a good initial guess for $\theta(t)$ to start our iterative algorithm to solve for the nonlinear optimization problem. In this section, we introduce a normalization operator to obtain a good initial guess for $\theta(t)$ from $f(t)$.

First of all, it is easy to prove the following proposition.

Proposition 3.1. *If $g(t), t \in [a,b]$ is continuous, and satisfies the following conditions*:

(1) $|g(t)| \leq 1$, $\forall\, t \in [a,b]$;
(2) *All local maximums of g are equal to* 1;
(3) *All local minimums of g are equal to* -1.

then there exists $\theta(t)$ such that $\theta'(t) \geq 0$ and $g(t) = \cos\theta(t)$.

We now introduce the normalization operator. Suppose z_i, z_{i+1} $(z_i < z_{i+1})$ are two adjacent extrema of $f(t)$ with $f(z_i)$ being the local minimum and $f(z_{i+1})$ the local maximum. Then, $f(t)$ can be normalized to satisfy the condition in Proposition 3.1 by using the following normalization operator

$$\Phi_c[f](t) = \frac{2f(t) - (f(z_{i+1}) + f(z_i))}{f(z_{i+1}) - f(z_i)}, \quad t \in [z_i, z_{i+1}]. \tag{16}$$

It is easy to check that for any continuous function $f(t)$, the normalized function $\Phi_c[f]$ satisfies the three conditions in Proposition 3.1. So, there exists a phase function $\theta_0(t)$ such that $\theta_0'(t) \geq 0$ and

$$\Phi_c[f](t) = \cos\theta_0(t). \tag{17}$$

This phase function θ_0 can be used as the initial guess of the nonlinear iteration method that we propose in the previous section.

We can prove that for the signals that satisfy the following scale-separation property, the normalization operator, Φ_c, defined above provide a good approximation to the exact phase function.

Assumption 3.1. (*Scale-separation*). *The decomposition* $f(x) = a_0(x) + a_1(x)\cos\theta_1(x)$ *is said to satisfy the scale-separation property if* a_0, a_1 *are smoother than* $\cos(\theta_1(x))$.

Throughout this paper, we say that $g_1(x)$ is smoother than $g_2(x)$ if the amplitude of the first-order derivative of g_1 is much smaller than that of g_2.

Proposition 3.2. *Let* $f = a_0 + a_1\cos\theta$. *Assume that* $x_j, 0 \le j \le m_1$ *are the local maxima of* f, *and* $z_j, 0 \le j \le m_2$ *the local minima. Define*

$$\epsilon = \max_{1\le j\le m_1}\left(\frac{|a_0'(x_j)| + |a_1'(x_j)|}{|a_1(x_j)||\theta'(x_j)|}\right) + \max_{1\le j\le m_2}\left(\frac{|a_0'(z_j)| + |a_1'(z_j)|}{|a_1(z_j)||\theta'(z_j)|}\right).$$

If ϵ *is small, which implies implicitly that* a_0, a_1, *and* $\cos(\theta)$ *satisfy the scale-separation condition, then we have*

$$\cos\theta(x_j) \approx 1 + O(\epsilon^2), \quad 0 \le j \le m_1 \tag{18}$$

$$\cos\theta(z_j) \approx -1 + O(\epsilon^2), \quad 0 \le j \le m_2. \tag{19}$$

Proof. Since $f'(x_j) = 0$, we have

$$f'(x_j) = a_0'(x_j) + a_1'(x_j)\cos\theta(x_j) - a_1(x_j)\theta'(x_j)\sin\theta(x_j) = 0. \tag{20}$$

Solving for $\sin\theta(x_j)$ from the above equation yields

$$\sin\theta(x_j) = \frac{a_0'(x_j) + a_1'(x_j)\cos\theta(x_j)}{a_1(x_j)\theta'(x_j)}. \tag{21}$$

Thus, we have

$$\begin{aligned}|\sin\theta(x_j)| &= \left|\frac{a_0'(x_j) + a_1'(x_j)\cos\theta(x_j)}{a_1(x_j)\theta'(x_j)}\right| \\ &\le \frac{|a_0'(x_j)| + |a_1'(x_j)|}{|a_1(x_j)||\theta'(x_j)|} \le \epsilon.\end{aligned} \tag{22}$$

Notice that

$$|\cos\theta(x_j) - 1| = |1 - \sqrt{1 - (\sin\theta(x_j))^2}| = O(|\sin\theta(x_j)|^2) = O(\epsilon^2). \tag{23}$$

Similarly, we have

$$|\cos\theta(z_j) + 1| = O(\epsilon^2). \tag{24}$$

□

The above analysis shows that the local extrema of the signal f give us the approximation to the upper envelop $a_0 + a_1$ and the lower envelope $a_0 - a_1$. Once

the extrema are identified, the upper and lower envelopes can be approximated by interpolating local maxima and minima. To improve the accuracy of our approximation to the upper and lower envelopes, we may use a high-order interpolation method, such as the cubic spline method.

Let us denote by $U(t)$ and $L(t)$ the cubic spline interpolation of the upper and lower envelopes, respectively. We may define a high-order normalization operator, Φ_s, as follows

$$\Phi_s[f](t) = \frac{2f(t) - (U(t) + L(t))}{U(t) - L(t)}. \tag{25}$$

Clearly, the accuracy of the normalization operator is based on the scale-separation property of the signal and the accuracy of the interpolation used to construct the envelopes. If the error introduced by the interpolation is smaller than the error introduced by the lack of scale separation, then the accuracy of the normalization cannot be improved by using a high-order interpolation. For this reason, we do not try to use an interpolation polynomial with order higher than the cubic spline.

One drawback in using the cubic spline interpolation to approximate $U(t)$ and $L(t)$ is that $\Phi_s[f]$ may not satisfy the conditions in Proposition 3.1. If it is the case, the lower-order normalization operator, Φ_c, is applied to $\Phi_s[f]$. Thus, the final normalization operator is defined as the composition of Φ_c and Φ_s:

$$\Phi[f] = \Phi_c \circ \Phi_s[f]. \tag{26}$$

The initial guess of θ is obtained by taking arccos:

$$\theta_0(t) = \arccos(\Phi[f]). \tag{27}$$

3.4. *Implementation*

In this section, we provide further details for the implementation of our iterative algorithm for solving the nonlinear optimization problem. First of all, we observe that in each step of the iterative algorithm, one third-order total variation minimization problem needs to be solved. This third-order total variation minimization problem can be written as an L_1 minimization problem, which has been well studied in the compressed sensing literature.

Suppose the signal is uniformly sampled at $t_i,\ i = 1, 2, \ldots, N$. Then, the third-order total variation minimization problem can be reformulated as follows

$$\min \quad \|\Phi \mathbf{x}\|_1, \quad \text{subject to:} \quad \mathbf{A}\mathbf{x} = \mathbf{f}. \tag{28}$$

where

$$\Phi = \left[\mathbf{D}^4, \mathbf{D}^4, \mathbf{D}^4\right], \quad x = \begin{bmatrix} a_0^n \\ a_1^n \\ b_1^n \end{bmatrix}, \quad \mathbf{A} = \left[\mathbf{I},\ \mathrm{diag}(\cos\theta^{n-1}),\ \mathrm{diag}(\sin\theta^{n-1})\right]$$

and $\mathbf{D}^4 \in \mathbb{R}^{(N-4)\times N}$ is the matrix, which is obtained by discretizing the fourth-order derivative by a finite difference method

$$\mathbf{D}^4 = \begin{bmatrix} 1 & -4 & 6 & -4 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 & -4 & 6 & -4 & 1 \end{bmatrix} \tag{29}$$

$I \in \mathbb{R}^{N\times N}$ is the identity matrix, and $\mathrm{diag}(\cos\theta^{n-1})$ and $\mathrm{diag}(\sin\theta^{n-1})$ are diagonal matrices:

$$\mathrm{diag}(\cos\theta^{n-1}) = \begin{bmatrix} \cos\theta^{n-1}(t_1) & 0 & \cdots & \cdots & 0 \\ 0 & \cos\theta^{n-1}(t_2) & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \cos\theta^{n-1}(t_{N-1}) & 0 \\ 0 & \cdots & \cdots & 0 & \cos\theta^{n-1}(t_N) \end{bmatrix} \tag{30}$$

$$\mathrm{diag}(\sin\theta^{n-1}) = \begin{bmatrix} \sin\theta^{n-1}(t_1) & 0 & \cdots & \cdots & 0 \\ 0 & \sin\theta^{n-1}(t_2) & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \sin\theta^{n-1}(t_{N-1}) & 0 \\ 0 & \cdots & \cdots & 0 & \sin\theta^{n-1}(t_N) \end{bmatrix}. \tag{31}$$

To solve the above L^1 minimization problem, we can use either the interior point method for large-scale L^1-regularized least square method developed recently by Kim *et al.* (2007) or the split Bregman iteration developed by Goldstein and Osher (2009). For large-scale data, we find that the split Bregman iteration is more efficient than the L^1-regularized least square method.

Another issue we need to consider in the implementation is the end effect. In the normalization process, we need to estimate the upper and lower envelopes of the signal at the two boundary points of the time domain. We have tried different methods to reduce this end effect, including the method proposed by [Wu and Huang (2009)].

4. Numerical Results

In this section, we perform a number of numerical experiments to test the convergence and accuracy of the proposed adaptive data analysis method for a number of examples involving a variety of multiscale data.

4.1. *Synthetic data*

We first apply our method to several synthetic data. The advantage of using the synthetic data is that we know what is the exact sparse decomposition that we try

to recover using our method. For real data, we do not have the luxury to know what is the correct decomposition. We can only use the underlying physical property of a signal as a guidance whether our decomposition captures the hidden physical property of the signal.

Example 1. First, we test our method for a simple nonstationary function given by

$$f(t) = 6t + \cos(8\pi t) + 0.5\cos(40\pi t). \tag{32}$$

The original signal is shown in Fig. 1, and the results are shown in Figs. 2 and 3. As one can see, we have recovered the three components of the original signal accurately. The first IMF corresponds to the highest-frequency component, $0.5\cos(40\pi t)$, and the second IMF corresponds to $\cos(8\pi t)$. The last component represents the trend, $6t$. Figure 2 also gives the comparison between the results obtained by our method and those obtained by the EMD method. In most part of the domain, the IMFs and trend obtained from these two methods agree very well. Near the two boundary end points, the performance of our method is slightly better than that of the EMD method.

The boundary effect is clearer in the result of instantaneous frequencies (Fig. 3). The instantaneous frequency obtained by the EMD method (and EEMD) is computed by the open source MATLAB program ifndq.m, which is available from the Web site http://rcada.ncu.edu.tw/research1_clip_program.htm. The instantaneous frequencies obtained by our method are relatively close to the exact instantaneous

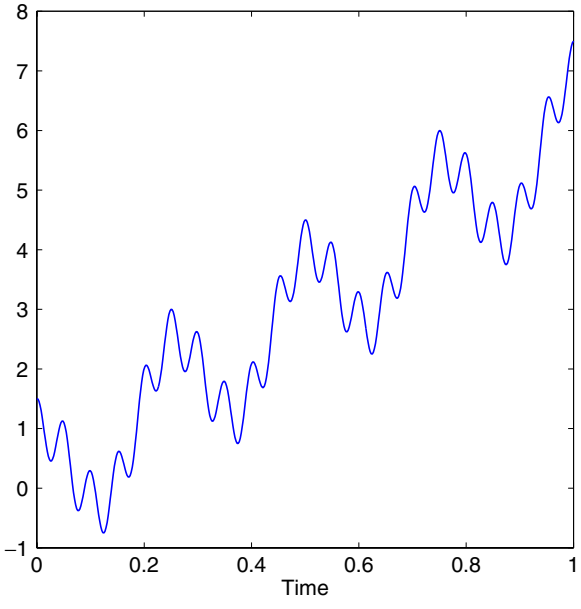


Fig. 1. Original data in Example 1.

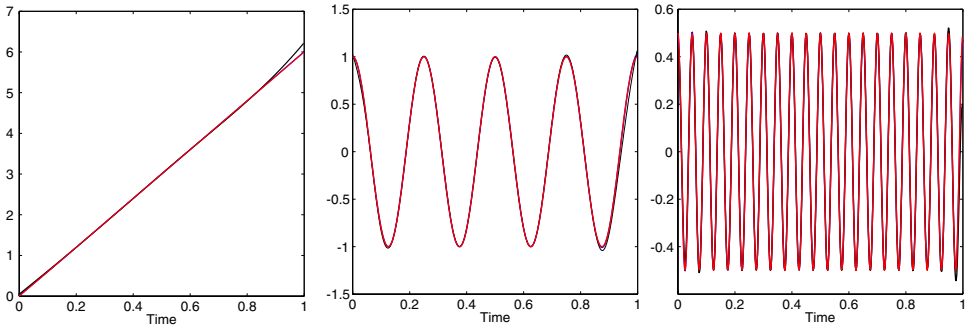


Fig. 2. IMFs and trend in Example 1. Red: analytical results; Blue: our method; and Black: EMD method.

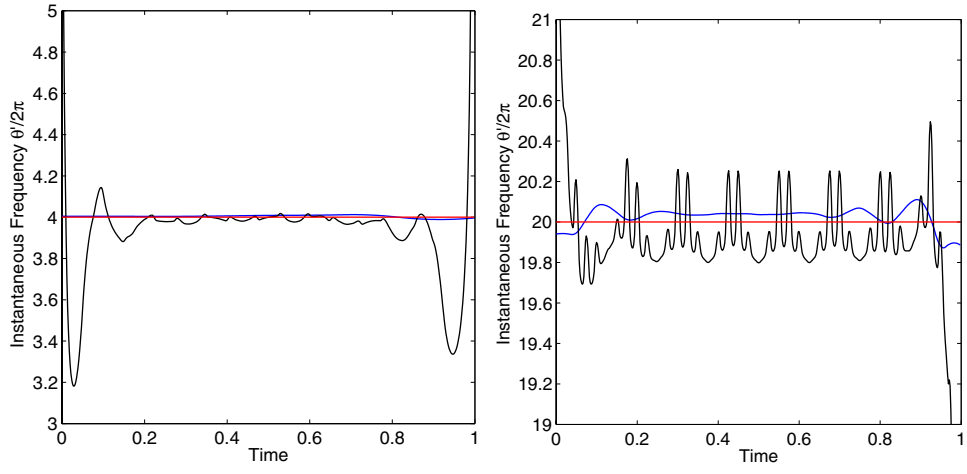


Fig. 3. Instantaneous frequency in Example 1. Red: analytical results; Blue: our method; and Black: EMD method.

frequencies. On the other hand, the results obtained by the EMD method tend to produce many small high frequency oscillations.

Example 2. The second example we consider is a little more complicated. It is a superposition of a signal with a discontinuous instantaneous frequency, a chirp signal, and a quadratic trend:

$$\theta(t) = \begin{cases} 60\pi t, & 0 \le t \le 0.5 \\ 80\pi t - 15\pi, & 0.5 < t \le 1. \end{cases} \tag{33}$$

$$f(t) = 6t^2 + \cos(10\pi t + 10\pi t^2) + \cos\theta(t).$$

The original signal is shown in Fig. 4 and the results are shown in Figs. 5 and 6, respectively. The three components are recovered very well by our method. We plot the instantaneous frequencies in Fig. 6. Both the EMD method and our method

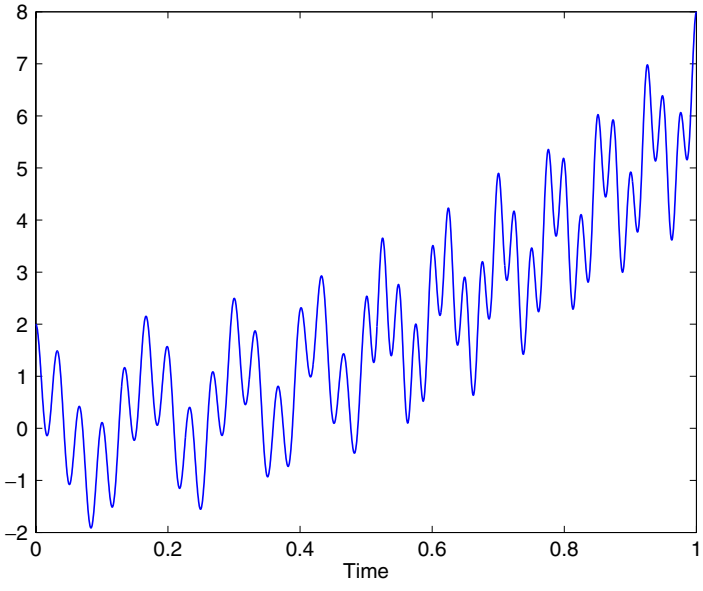


Fig. 4. Original data in Example 2.

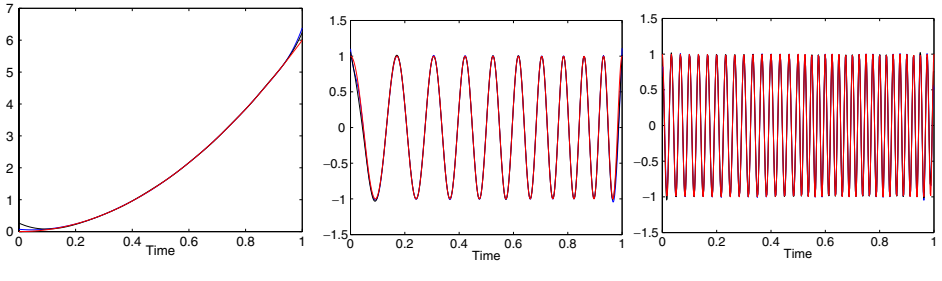


Fig. 5. IMFs and trend in Example 2. Red: analytical results; Blue: our method; and Black: EMD method.

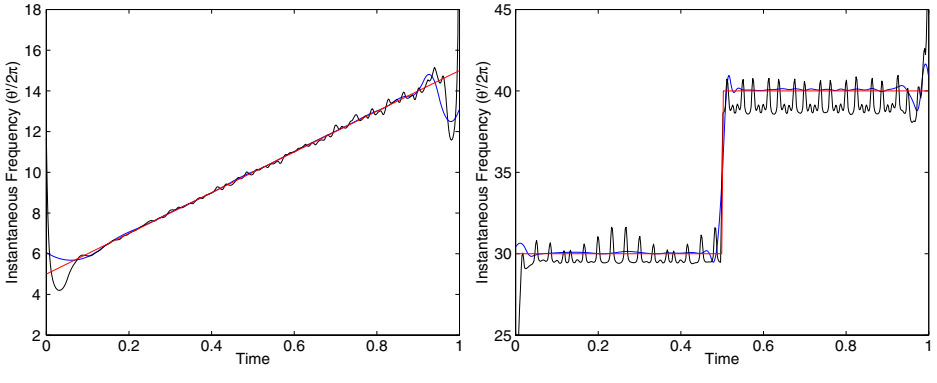


Fig. 6. Instantaneous frequency in Example 2. Red: analytical results; Blue: our method; and Black: EMD method.

capture accurately the position of the jump discontinuity of the instantaneous frequency of the first IMF (see the right plot). Again, we observe that the instantaneous frequency given by the EMD method has a large number of small oscillations while our method gives a much smoother result. The instantaneous frequency function for the chirp signal is almost perfect except near the end points due to the end effect.

Example 3. In this example, we try to decompose a signal that has intrawave frequency modulation, which is given as follows

$$f(t) = \frac{1}{1.2 + \cos(2\pi t)} + \frac{1}{1.5 + \sin(2\pi t)} \cos(32\pi t + 0.2\cos(64\pi t)). \tag{34}$$

This is very similar to the data obtained as the solution of the Duffing equation, which was first considered by Huang *et al.* (1998). This signal is challenging because the instantaneous frequency itself has very high frequency modulation (Fig. 7). In fact, the instantaneous frequency, $\theta'(t)$ is more oscillatory than $\cos(\theta(t))$ itself. This makes it extremely challenging for the nonlinear optimization problem.

We demonstrate that our method still applies to this challenging case with reasonable accuracy. As we can see from Fig. 8, our method can still extract the instantaneous frequency very well. The key for the success is that we need to obtain a good initial guess for θ in our iterative method. In the case with no noise pollution, we can come up with a relatively accurate initial guess for θ by using our normalization operator. However, the problem would become much harder if the

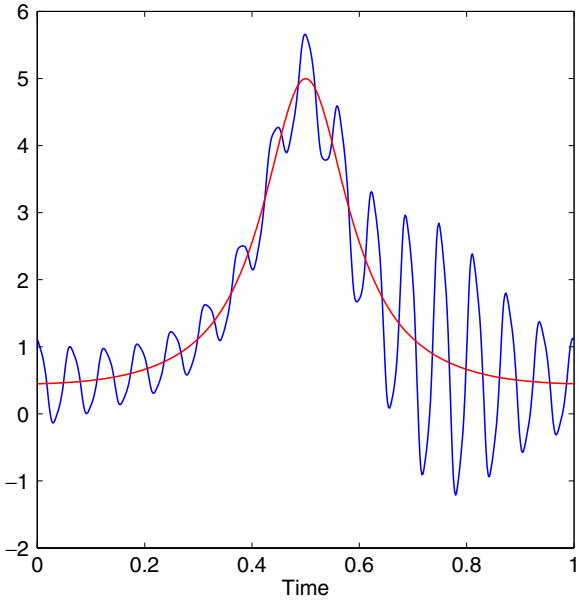


Fig. 7. Original data and local median of Example 3.

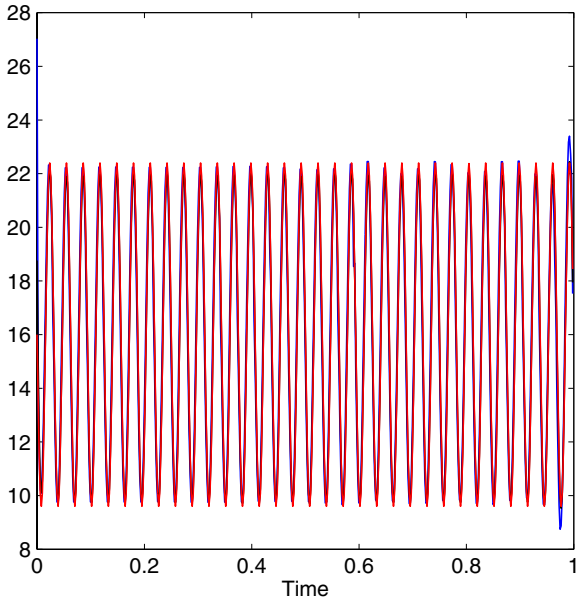


Fig. 8. Instantaneous frequency of the signal in Example 3. Red: analytical results; Blue: our method; and Black: EMD method.

signal is polluted with noise since it would be very hard to separate the physical instantaneous frequency, which contains very high-frequency information, from the high-frequency contribution due to noise.

4.2. *Real data*

In the previous section, we use several synthetic data to test the convergence and accuracy of our method. In this section, we apply our method to decompose two sets of real data. The first one we consider is the length of the day (LOD) data. Here, we only consider a segment of the LOD data in 700 consecutive days. This example has been used by a number of researchers as a prototype test case because it contains a number of hidden physical scales in this signal. In Fig. 9, we plot the original signal. The IMFs obtained by our method are shown in Fig. 10. Just like the EMD method, our method can recover almost all the physically meaningful IMFs. For example, the first component C_1 captures the semi monthly tides while the second component C_2 represents the monthly tides. Similarly, C_3 captures the semi annual cycle and C_4 captures the annual cycle.

The second example we consider is the annual global surface temperature (GST) data from 1880 to 2009. The original signal is shown in Fig. 11. The historical record shows clear evidence of a warming trend over the last century. We are interested in understanding the warming rates, the causes of warming, and an accurate approximation of the trend.

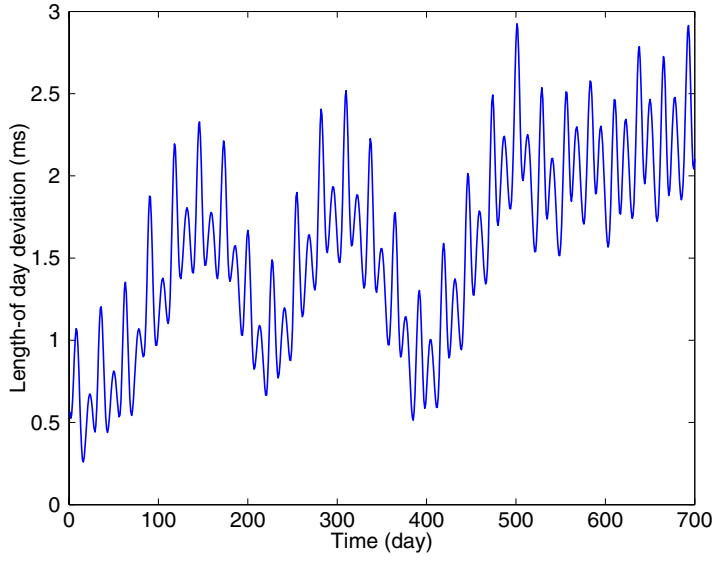


Fig. 9. LOD in 700 days.

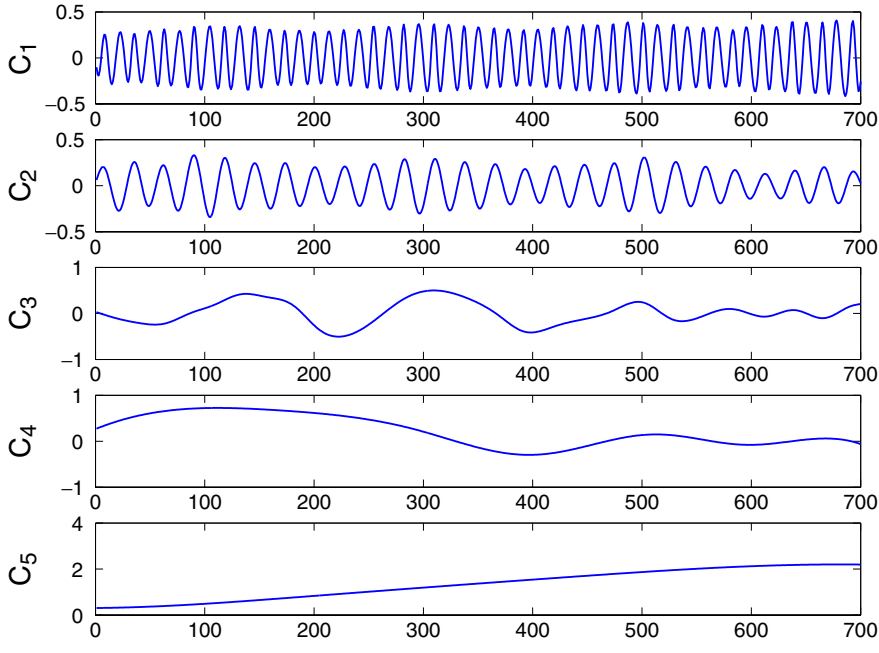


Fig. 10. The IMFs decomposed from LOD data.

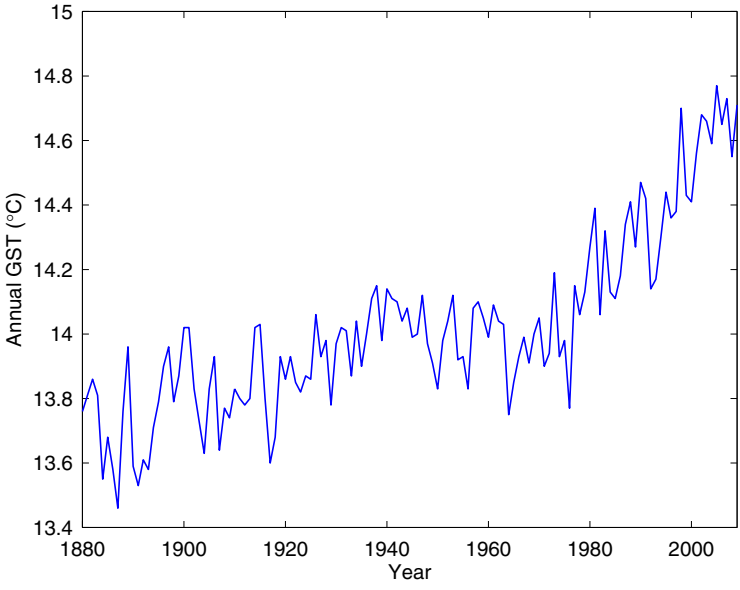


Fig. 11. Annual GST from 1880 to 2009.

Figure 12 shows the IMFs obtained by our method. Various trends, including the linear trend, the overall adaptive trend (the residual component C_6), and the multidecadal trend (the sum of C_5 and C_6), are plotted in Fig. 13. Here, the overall adaptive trend is the trend derived by using our method over the whole data span, and the multidecadal trend is the remainder after IMFs of periods shorter than multidecades are removed from the GST, which can be regarded as the union of the trends derived from consecutive multidecadal sections of GST. In Fig. 13, a comparison among the different fittings also is illustrated. The intrinsically determined overall adaptive trend shows that the rate of global warming is slower than the prediction given by the linear fitting. The multidecadal trend is most interesting. It catches essentially the meaningful variability and change associated with the annual GST. The multidecadal trend seems to suggest that the GST may reach a potential local maximum at the present time. If this were true, it may imply that the rate of global warming could slow down in the near future. On the other hand, the trend obtained by linear regression over the data from 1980 to 2009 tends to over estimate the rate of global warming.

We remark that the above analysis is preliminary. More careful analysis of other data that could contribute to global warming needs to be carried out before a definite conclusion can be reached. It is also essential to compare the results obtained by various other data analysis methods. What we demonstrate here is to show that our method has the potential to be applied to such an important dataset. The conclusion from analyzing this type of geophysical data may have a significant impact on our environment and society.

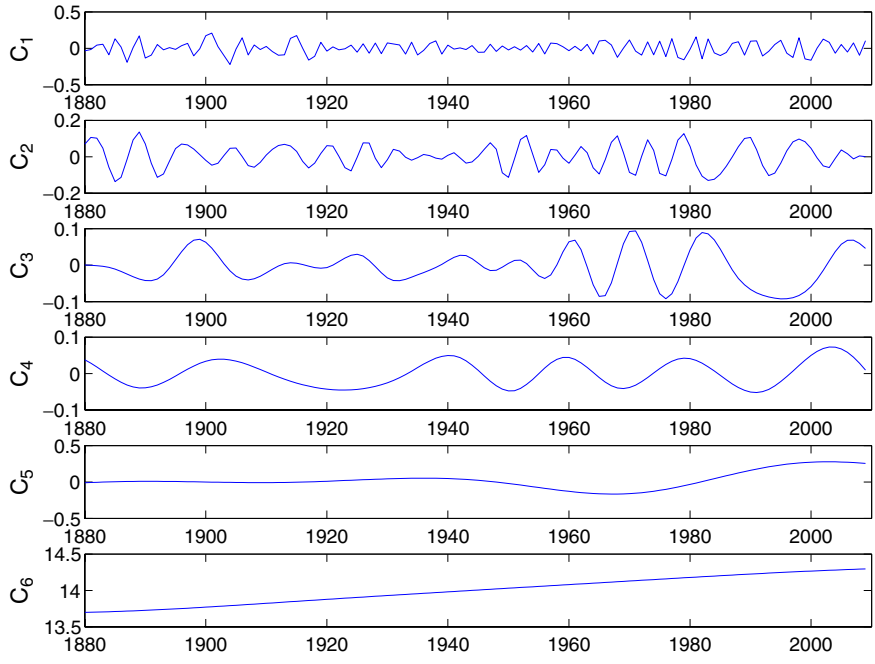


Fig. 12. IMFs decomposed from GST data.

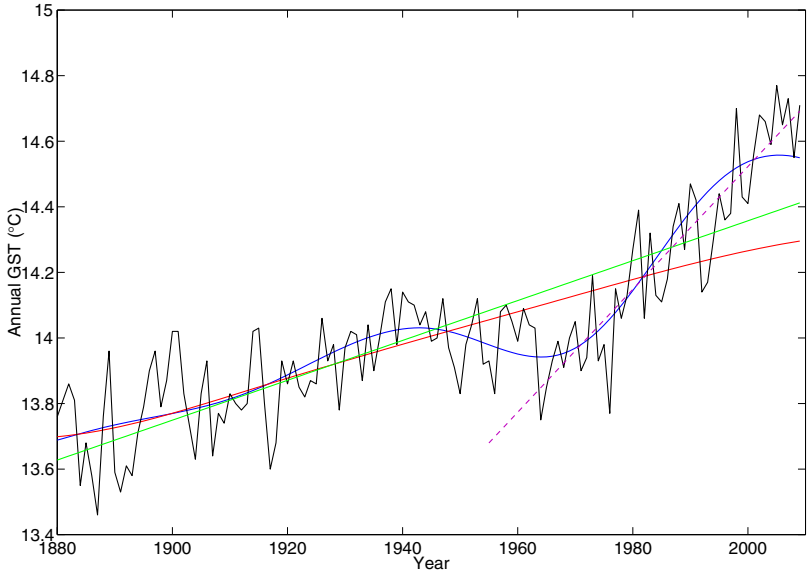


Fig. 13. The trends of the GST data. Red: trend obtained from our method (last IMF); Blue: last second trend (summation of last and last second IMF); Green: trend obtained by linear regression; and Purple dash: trend obtained by linear regression over the data from 1980 to 2009.

5. An Adaptive Filter for Noisy Data

As we demonstrated in the previous section, our $TV^{(3)}$-based nonlinear optimization method gives results that are either better than or comparable to those obtained by the EMD method. However, the use of the normalization process to obtain the initial guess also makes our method sensitive to numerical noise. This is because our normalization process depends on accurate approximations of the local extrema. This step shares some similarity with the EMD method, which also suffers from the same type of numerical instability for noisy data. To overcome this numerical instability to noise perturbation, we introduce an adaptive filter in our method to make it more stable to noise.

One way to deal with noisy data is to first remove the noise from the data, then apply our method to the denoised data. We observe that if we are interested in computing the instantaneous frequency associated with the phase function $\theta(t)$, we may treat those components with frequencies higher than $\cos\theta(t)$ as noise. Based on this observation, we can design an adaptive low-pass filter based on the current phase function θ to remove the high-frequency components. This filter does not harm the physical instantaneous frequency of interest at the level given by the phase function θ. More specifically, if $\theta(t)$ is a linear function, then we can choose the Meyer scaling function as a low-pass filter (Fig. 14):

$$\widehat{\chi}(k) = \begin{cases} 1, & |k| \le 1 \\ \dfrac{1}{2}\left(1 - \cos\left(\pi k\right)\right), & 1 < |k| < 2 \\ 0, & |k| > 2. \end{cases} \tag{35}$$

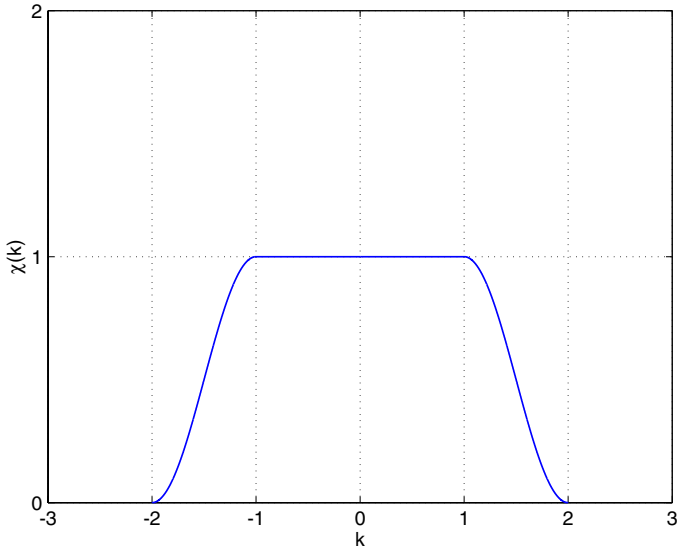


Fig. 14. The low-pass filter given in Eq. (35).

where $\widehat{\chi}$ is the Fourier transform of χ and k is the wave number. For any phase function θ, notice that in the iterative process, the phase function $\theta(t)$ is always monotonically increasing. Then, we can use $\theta(t)$ as a new coordinate. In this new coordinate, $\theta(t)$ is a simply linear function of θ. Thus, we can use the low-pass filter given in Eq. (35) in the θ-coordinate. Now, we have an adaptive filter strategy based on the phase function $\theta(t)$ as follows

Step 1. Interpolating $f(t)$ to the uniform mesh of θ-coordinate to get $f_\theta(\theta_j)$:

$$f_\theta(\theta_j) = \text{Interpolate } (\theta(t_i), f(t_i), \theta_j), \tag{36}$$

where $\theta_j,\ j = 1, \ldots, N$ are uniformly distributed in θ-coordinate.

Step 2. Applying the low-pass filter on the Fourier transform of f_θ

$$\bar{f}_\theta = \mathcal{F}^{-1}[\widehat{f}_\theta(k)\widehat{\chi}(k)], \tag{37}$$

where the low-pass filter is given in Eq. (35).

Step 3. Transforming $\bar{f}_\theta$ back to the t- coordinate to get the data after filtering:

$$\bar{f}(t_i) = \text{Interpolate } (\theta_j, \bar{f}_\theta(\theta_j), t_i), \tag{38}$$

Combining this adaptive filter strategy with the previous third-order total variation minimizing method, we obtain the following generalized iterative algorithm

Initialization: $\theta^0 = \theta_0$.
Main iteration:

Step 1. Interpolating $f(t)$ to the uniform mesh of θ^{n-1} coordinate to get $f_\theta^{n-1}(\theta_j^{n-1})$:

$$f_\theta^{n-1}(\theta_j^{n-1}) = \text{Interpolate } (\theta^{n-1}(t_i), f(t_i), \theta_j^{n-1}), \tag{39}$$

where $\theta_j^{n-1},\ j = 1, \ldots, N$ are uniformly distributed in θ^{n-1} coordinate.

Step 2. Applying the low-pass filter on the Fourier transform of f_θ^{n-1}

$$\bar{f}_\theta^{n-1} = \mathcal{F}^{-1}[\widehat{f}_\theta^{n-1}(k)\widehat{\chi}(k)], \tag{40}$$

where the low-pass filter is given in Eq. (35).

Step 3. Transforming $\bar{f}_\theta^{n-1}$ back to the t-coordinate to get the data after filtering:

$$\bar{f}^{n-1}(t_i) = \text{Interpolate } (\theta_j^{n-1}, \bar{f}_\theta^{n-1}(\theta_j^{n-1}), t_i), \tag{41}$$

Step 4. Update a_0^n, a_1^n, and b_1^n by solving the following linear optimization problem

$$\text{Minimize} \quad TV^3(a_0^n) + TV^3(a_1^n) + TV^3(b_1^n), \tag{42}$$

$$\text{Subject to:} \quad a_0^n + a_1^n \cos\theta^{n-1}(t) + b_1^n \sin\theta^{n-1}(t) = \bar{f}^{n-1}(t). \tag{43}$$

Step 5. Update the phase function θ:

$$\theta^n = \theta^{n-1} - \mu \arctan\left(\frac{b_1^n}{a_1^n}\right), \tag{44}$$

where $\mu \in [0,1]$ is chosen to enforce that θ^n is increasing

$$\mu = \max\left\{\alpha \in [0,1] : \frac{d}{dt}\left(\theta_k^{n-1} + \alpha \arctan\left(\frac{b_1^n}{a_1^n}\right)\right) \geq 0\right\}. \tag{45}$$

Step 6. If $\|\theta^n - \theta^{n-1}\|_2 \leq \epsilon_0$, stop. Otherwise, go to Step 1.

In the above algorithm, we need to perform the Fourier transform to apply the adaptive filter. In principle, this operation is only valid for periodic data. For a non-periodic signal, we extend the signal by a mirror reflection and treat the extended signal as a periodic signal. To get a smoother extension, we find that it is better to extend the signal from a point of local maximum or local minimum since its first derivative vanishes there.

For noisy data, we cannot obtain a good initial guess using the normalization operator. In our computations, we use some traditional time-frequency analysis methods, including the Fourier transform, to generate our initial guess. In the following numerical examples, the initial guess is obtained by using the Fourier transform. By estimating the wave number by which the high-frequency components are centered around, we can obtain a reasonably good initial guess for θ. The initial guess for θ obtained in this way is a linear function. We can see in the following numerical examples that even using these relatively rough initial guesses for θ, our algorithm still converges to the right answer with accuracy determined by the noise level.

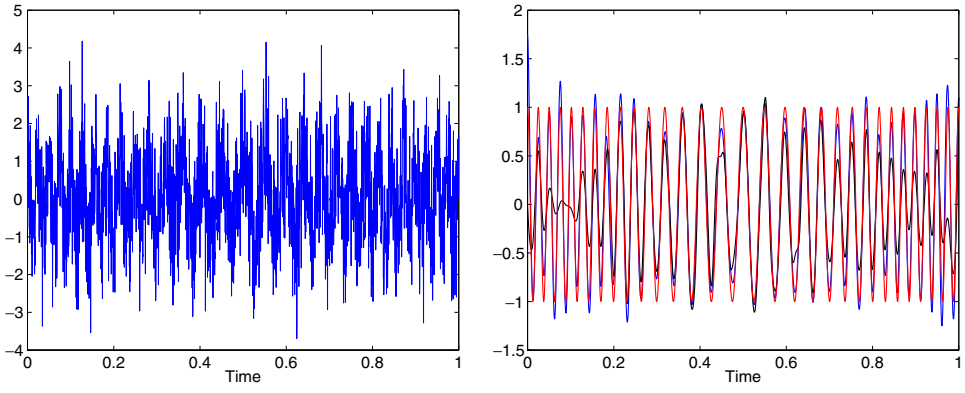
Next, we give several numerical examples to demonstrate how the above algorithm performs. Let $X(t)$ be a white noise with zero mean and variance $\sigma^2 = 1$.

Example 1. In the first example, the signal is a superposition of a single IMF and the Gaussian noise $X(t)$:

$$f(t) = \cos(60\pi t + 10\sin(2\pi t)) + X(t). \tag{46}$$

In this example, the initial guess for θ is $60\pi t$. Figure 15 shows the IMFs of the above signal obtained by different methods. Although the noise level is pretty high (the amplitude of the noise is $O(1)$), our nonlinear optimization method coupled with the adaptive filter can still decompose the instantaneous frequency and the corresponding IMF with reasonable accuracy (Figs. 15 and 16). Despite the large noise level, the estimation of the instantaneous frequency is still quite accurate. The amplitude of the IMF is less accurate, but the error is within the noise level.

As a comparison, we also show the IMF obtained by EEMD method in Fig. 15. In the EEMD approach, the number of ensemble average is chosen to be 200 and the standard deviation of the added noise is 0.2. In each ensemble average, the number of shifting is set to be equal to 8. The IMF obtained by the EEMD method is shown in Fig. 15. Among different components of IMFs, we select those components that are closest to the exact IMF in L_2 norm. As shown in Fig. 15, the IMF decomposed by EEMD fails to capture the phase of the exact IMF in some region. As a consequence, the accuracy of the instantaneous frequency is relatively poor.



(a) (b)

Fig. 15. (a) Noised data and (b) IMFs obtained by our method (blue), EEMD (black), and exact one (red).

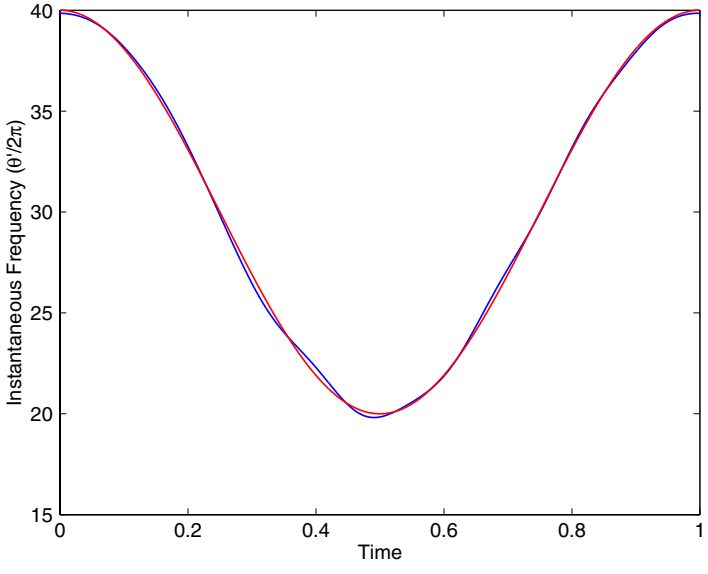


Fig. 16. The instantaneous frequency. Red: analytical results and Blue: our method.

Example 2. The second example is a little more complicated. The signal is generated by adding the noise, $X(t)$, to the signal given in Example 2 in the previous section. More precisely, the signal is generated as follows

$$\theta(t) = \begin{cases} 60\pi t, & 0 \le t \le 0.5 \\ 80\pi t - 15\pi, & 0.5 < t \le 1. \end{cases} \tag{47}$$

$$f(t) = 6t^2 + \cos(10\pi t + 10\pi t^2) + \cos\theta(t) + X(t).$$

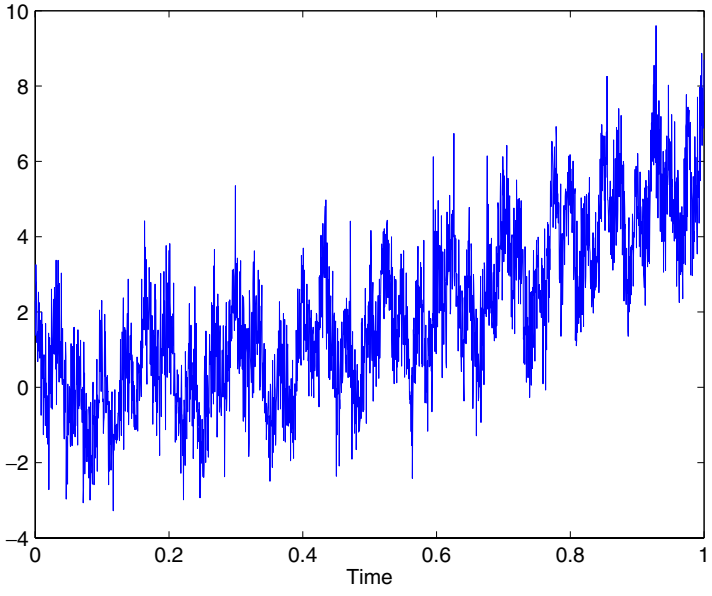
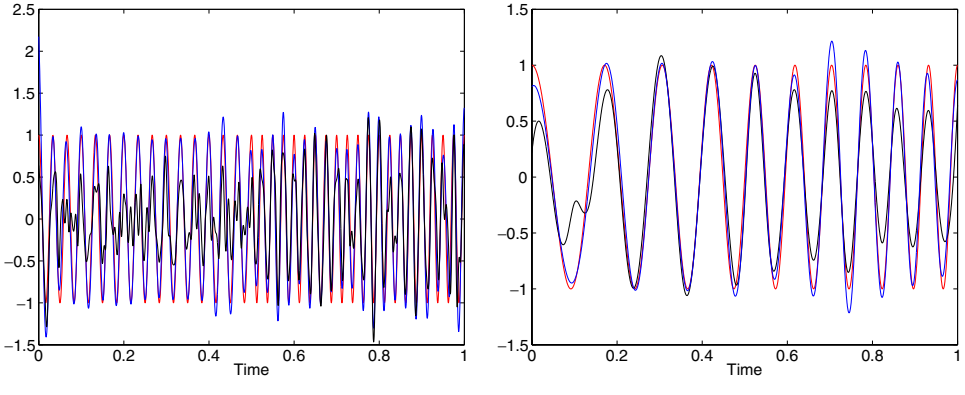


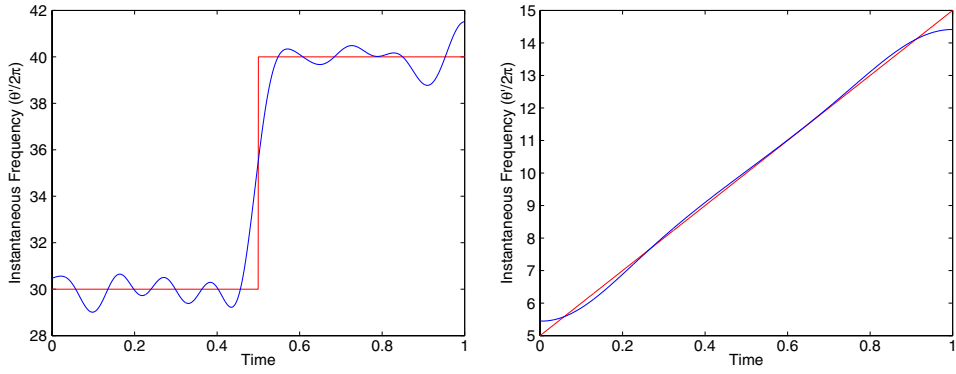
Fig. 17. The noised data.

The noisy signal is shown in Fig. 17. The initial guesses are $20\pi t$ and $80\pi t$ respectively for these two IMFs. The IMF and the instantaneous frequencies are plotted in Figs. 18 and 19, respectively. By comparing with the decomposition of the same data without noise in Fig. 6, we find that we do not capture the discontinuity of the instantaneous frequency for the noisy data as sharp as the data without noise. But if we take into consideration of the large noise level, the accuracy is still quite



(a) (b)

Fig. 18. (a) First IMF with jump frequency and (b) Second IMF with chirp frequency. Blue: our method; Black: EEMD; and Red: exact.



(a) (b)

Fig. 19. Instantaneous frequency of IMFs. (a) Jump instantaneous frequency of the first IMF and (b) Chirp instantaneous frequency of the second IMF. Blue: our method and Red: exact.

reasonable (Fig. 19). If we compare the IMFs obtained by our method with those obtained by EEMD, our result is actually pretty good (Fig. 18).

6. Concluding Remarks

In this paper, we developed a new adaptive data analysis method based on an L^1-based nonlinear optimization method. Adaptivity of our decomposition is obtained by looking for the sparsest representation of signals in the time-frequency domain from a largest possible dictionary that consists of all possible candidates for IMFs. Solving this nonlinear optimization problem is, in general, very difficult. We proposed an iterative algorithm and combined it with an efficient solver for L^1-minimization (the split Bregman method). Further, we introduced an adaptive filter and combined our iterative algorithm with this adaptive filter. The combined algorithm is more stable to noisy data. Numerical examples for both synthetic and noisy data show that our method can provide a sparse decomposition of nonlinear and nonstationary data without compromising the hidden physical property of the signal. We also compared the performance of our method with the EMD method and showed that our method gives results that are either comparable to or more superior to those obtained by the EMD method.

We have also carried out some preliminary convergence analysis for the nonlinear optimization method proposed in this paper. In the case when the signal has the form, $f(t) = a_0(t) + a_1(t)\cos(\theta(t))$, we can show that if a_0, a_1, and θ have a sparse representation in some given basis, then our iterative algorithm would converge to the correct decomposition if some additional condition which measures the mutual coherence of the iterative matrix is satisfied. The detail of this convergence analysis will be reported elsewhere in the future.

There are still some limitations with the method presented here. One of the more serious difficulties is the use of the $TV^{(3)}$ norm in our nonlinear optimization

method. This makes our method more sensitive to noise. Although we proposed an adaptive filter to alleviate this difficulty, our method still suffers some numerical instability when the noise level is large. In a subsequent paper, we will introduce another approach which is based on the matching pursuit approach. The nonlinear optimization method based on the matching pursuit approach can be implemented much faster than the current method and has a complexity of order $O(N \log N)$, where N is the number of data sample points that we use to represent the signal. When the signal satisfies the scale-separation property, it is also possible to obtain a sparse decomposition of the signal using the undersampled data. The most important advantage of this approach is that it is very stable to noise perturbation. This enables us to apply it to a wider range of real data. Numerical experiments seem to suggest that this new approach offers more superior performance than the EEMD method. The paper with the results of this new approach will appear in another journal [Hou and Shi (2011)].

Acknowledgments

This work was in part supported by the NSF grant DMS-0908546. We would like to thank Professors Norden E. Huang and Zhaohua Wu for many stimulating discussions on EMD/EEMD and topics related to the research presented here. Professor Hou would like to express his gratitude to the National Central University (NCU) for their support and hospitality during his visits to NCU in the past two years.

References

Bedrosian, E. (1963). A product theorem for Hilbert transforms. *Proc. IEEE.*, **51**: 868–869.

Boashash, B. (1992) *Time Frequency Signal Analysis Methods and Applications*, Longman-Cheshire, Melbourne and John Wiley Halsted Press, New York.

Bruckstein, A. M., Donoho, D. L. and Elad, M. (2009). From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.*, **51**: 34–81.

Candès, E. and Tao, T. (2006). Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory*, **52**(12): 5406–5425.

Candes, E. Romberg, J. and Tao, T. (2006a). Robust uncertainty principles: Exact signal recovery from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, **52**: 489–509.

Candes, E., Romberg, J. and Tao, T. (2006b). Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.*, **59**: 1207–1223.

Cohen, L. (1995). *Time-Frequency Analysis*, Prentice Hall, Englewood Cliffs, NJ.

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

Daubechiesa, I. Lu, J. and Wu, H.-T. (2011). Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Applied and Computational Harmonic Analysis*, **30**(2): 243–261.

Donoho, D. L. (2006). Compressed sensing. *IEEE Trans. Inf. Theory*, **52**: 1289–1306.

Flandrin, P. (1999). *Time-Frequency/Time-Scale Analysis*, Academic Press, San Diego, CA.

Gabor, D. (1946). Theory of communication. *J. IEE.,* **93**: 426–457.

Goldstein, T. and Osher, S. (2009). The split Bregman method for L_1-regularized problems. *SIAM J. Imaging Sci.* **2**(2): 323–343.

Hou, T. Y. and Shi, Z. (2011). Sparse time-frequency representation of multiscale data by nonlinear matching pursuit. *SIAM Multiscale Model. Sim.*, submitted.

Huang, N. E. *et al.* (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Londn. A*, **454**: 903–995.

Huang, N. E., Shen, Z. and Long, S. R. (1999). A new view of nonlinear water waves — The Hilbert spectrum. *Ann. Rev. Fluid Mech.*, **31**: 417–457.

Huang, N. E., Wu, Z., Long, S. R., Arnold, K. C., Chen, X. and Blank, K. (2009). On instantaneous frequency. *Adv. Adapt. Data Anal.* **1**(2): 177–229.

Jomes, D. L. and Parks, T. W. (1990). A high resolution data-adaptive time-frequency representation. *IEEE Trans. Acoust. Speech Signal Process.*, **38**: 2127–2135.

Kim, S. J., Koh, K., Lustig, M., Boyd, S. and Gorinevsky, D. (2007). An interior-point method for large-scale L^1-regularized least squares. *IEEE J. Sel. Top. Signal Process.*, **1**(4): 606–617.

Loughlin, P. J. and Tracer, B. (1996). On the amplitude — and frequency-modulation decomposition of signals. *J. Acoust. Soc. Am.*, **100**: 1594–1601.

Lovell, B. C., Williamson, R. C. and Boashash, B. (1993). The relationship between instantaneous frequency and time frequency representations. *IEEE Trans. Signal Process.*, **41**(3): 1458–1461.

Mallat, S. (2009). *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, San Diego, CA.

Meville, W. K. (1983). Wave modulation and breakdown. *J. Fluid Mech.*, **128**: 489–506.

Nuttall, A. H. (1966). On the quadrature approximation to the Hilbert transform of modulated signals. *Proc. IEEE.*, **54**: 1458–1459.

Olhede, S. and Walden, A. T. (2004). The Hilbert spectrum via wavelet projections. *Proc. Roy. Soc. Lond. A*, **460**: 955–975.

Picinbono, B. (1997). On instantaneous amplitude and phase signals. *IEEE Trans. Signal Process.*, **45**: 552–560.

Qian, S. and Chen, D. (1996). *Joint Time-Frequency Analysis: Methods and Applications*, Prentice Hall, Englewood Cliffs, NJ.

Rice, S. O. (1944). Mathematical analysis of random noise. *Bell Syst. Tech. J.*, **23**: 282–310.

Shekel, J. (1953). Instantaneous frequency. *Proc. IRE*, **41**: 548–548.

Van der Pol, B. (1946). The fundamental principles of frequency modulation. *Proc. IEEE*, **93**: 153–158.

Wu, Z. and Huang, N. E. (2005). *Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method*, COLA Technical Report 193. ftp://grads.iges.org/pub/ctr/ctr_193.pdf.

Wu, Z., Huang, N. E., Long, S. R. and Peng, C. K. (2007). Trend, detrend, and the variability of nonlinear and non-stationary time series. *Proc. Natl. Acad. Sci. USA.*, **104**: 14889–14894. doi: 10.1073/pnas.0701020104.

Wu, Z. and Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **1**(1): 1–41.