Optimal averages for nonlinear signal decompositions - another alternative for empirical mode decomposition*

Feng Zhou[1,3], Lijun Yang[2], Hao-min Zhou[3] and Lihua Yang[1†]

[1]School of Mathematics and Computational Science, Sun Yat-sen University, 510275, China

[2]School of Mathematics and Information Sciences, Henan University, 475004, China

[3]School of Mathematics, Georgia Institute of Technology, Atlanta, GA, 30332, USA

Abstract

The empirical mode decomposition (EMD) is an algorithm pioneered by N. Huang et. al. as an alternative technique to the traditional Fourier and wavelet methods for analyzing nonlinear and non-stationary signals. It aims at decomposing a signal, via an iterative sifting procedure, into several intrinsic mode functions (IMFs), and each of the IMFs has better behaved instantaneous frequency analysis. This paper presents an alternative approach for EMD. The main idea is to replace the average of upper and lower envelopes in the sifting procedure of EMD by a local average obtained by variational optimization framework. Therefore, an IMF can be produced by simply subtracting the average from the signal without iteration. Our numerical examples illustrate that the resulting decomposition is convergent and robust against noise.

*This research was partially supported by NSF Faculty Early Career Development (CAREER) Award DMS-0645266, DMS-1042998, and DMS-1419027, ONR Award (N000141310408), NSFC (Nos. 11371017, 91130009), RFDP (No. 20130171110016), and the "Computational Science Innovative Research Team" program and Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University.

†The corresponding author. E-mail: mcsylh@mail.sysu.edu.cn, Tel: (8620)84115508

Keywords. empirical mode decomposition, Hilbert-Huang transform, time-frequency analysis, intrinsic mode function, optimization.

1 Introduction

It is well known that nonlinear and non-stationary signal analysis is important and difficult. Historically, Fourier transform has provided a general method for analyzing signals, and has achieved unprecedented success for signals generated by linear and stationary processes. For nonlinear and non-stationary signals, there are many methods too. For examples, windowed Fourier transform, wavelet transform and the Wigner-Ville distribution are designed for non-stationary but linear signals. A number of nonlinear time series analysis methods, almost all of them are based on Fourier analysis, have been studied for nonlinear but stationary systems [16, 24, 20]. Despite some remarkable performances achieved by those methods in various applications, analyzing nonlinear and non-stationary signals still remains challenging.

The empirical mode decomposition (EMD), first proposed by Huang et. al. in 1998, is a highly adaptive scheme serving as a powerful complement to the existing approaches [11] . The main idea is that any complicated data set can be decomposed into a finite, and often small, number of intrinsic mode functions (IMFs) via the so-called sifting procedure, where an IMF is a function that satisfies the following two conditions quoted directly from [11]:

(1) In the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one;

(2) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

In recent years, the EMD has received considerable attentions in terms of interpretations ([5, 3,

23, 18]) and applications in many disciplines such as ocean science [10], biomedicine ([14, 21, 32]), speech signal processing [28], image processing [1], pattern recognition ([19, 30, 33]) and many more.

Despite its remarkable success, it is still lacking of mathematical understanding of the EMD method, such as its convergence property, dependence on the number of sifting, the stopping criteria, and its stability to noise perturbations. Those shortcomings are partly due to the highly adaptive nature of the sifting procedure as well as the ad hoc nature of using cubic splines or B-splines [3]. This sparkles many studies to provide mathematically sounding alternatives to the EMD method. For examples, in [4], Daubechies et. al. proposed a framework, named synchrosqueesed wavelet transform, that combines multiresolution analysis, amplitude modulated-frequency modulated (AM-FM) signals, and shape functions to achieve signal decompositions. Hou et. al. [7] proposed a variational approach, based on AM-FM functions and higher order total variation norms to achieve data dependent sparse signal decompositions. Lin et. al. [17] designed adaptive iterative filters to compute the local average, which is then used to replace the average of upper and lower envelopes in the sifting procedure in the EMD method. In [8, 29, 15], Peng et. al. developed a local linear operator-based method, which decomposes a signal into summation of local narrow band signals. And Pustelnik et. al. [9, 27] proposed a non-smooth convex optimization approach to calculate multi-components of the signal so that each one can be considered as an IMF. In their optimization model, a constrain is used to enforce near orthogonality among the components. In addition, many groups have conducted researches to improve the EMD method [22, 25].

In this paper, we present a new mathematical framework serving as another alternative, which not only has closer ties to the original EMD method, i.e., its components can be guaranteed as IMFs, but also can resolve some aforementioned mathematical challenges faced by the original EMD method. Our approach is motivated by a recent theoretical study [31] showing that condition (2) in

Huang's original IMF descriptions implies condition (1) provided that in the sifting procedure, the upper and lower envelopes are obtained by spline interpolations of the local maxima and minima respectively. Inspired by this result, and the fact that the sifting procedure is used to generate IMFs with zero local averages, we propose to use *local* variational optimization models to find local averages directly, and then the local averages are subtracted from the signal to generate IMFs. In other words, we use local optimization models to replace the sifting procedure in the EMD method.

In the design of our new framework, we take several factors into consideration. First, we want our method to be flexible, so we do not use any pre-defined functions, such as AM-FM functions or wavelets as the bases for the IMFs. This is a feature shared by the original EMD method and the non-smooth optimization approach taken by Pustelnik et. al [9, 27]. On the other hand, many existing alternatives, such as the methods studied in[4, 7, 8, 29, 15], use pre-defined functions as bases for IMFs. Second, we would like to design our method adaptively and locally, so we partition the region into subregions separated by local extreme points and then solve local optimization problems for the local averages. This ensures that information from regions far apart does not influence each other. And we like to stress that this is a distinguished feature in our framework in response to the nature of nonlinear and non-stationary signals. To the best of our knowledge, in other existing optimization based decompositions, such as the ones in [7, 8, 29, 15, 9, 27], the objective functions are globally defined. Third, considering the possibility that the intrinsic components in a signal may not be orthogonal to each other, we do not use any conditions to achieve near orthogonality in the decomposition. This is different from some existing methods in the literatures, such as the ones proposed in [9, 27], in which near orthogonality is enforced. However, numerical examples indicate that our approach exhibits data dependent adaptivity in producing orthogonal components, meaning it generates near orthogonal decompositions if the intrinsic components are near orthogonal, otherwise it does not necessarily decompose orthogonal

components. Finally, we also want to design our strategy to be robust against noise perturbations. For this reason, we use both total variation (TV) and high order TV norms, all defined on the subregions to avoid global influence, to account for the noise perturbations as well as possible discontinuities in the signal. On the contrary, only higher order TV norms are used in the existing methods. In computation, we employ the recent developed split Bregmen iteration [34, 6, 2] to compute the solutions of the local optimization problems efficiently. Our theoretical study and numerical experiments show that the proposed method can achieve the desirable features with convergent, efficient, and robust against noise properties.

The rest of this paper is organized as follows. We introduce our optimization model for obtaining the local average, and provide some details of its implementation issues in Section 2. In Section 3, we present an alternative approach for the EMD method in data decomposition. In addition, we apply discontinuous signal and monocomponent data to test the performance of our approach. We also demonstrate its ability in data decomposition, which cover simulative, real and noisy data. We present our conclusion in Section 4.

2 The novel local average based on optimization for data decomposition

A key idea in the EMD method is to find a locally determined curve (or average) so that the difference between the signal and average is symmetric as described by the two conditions of IMFs given in [11]. In the original EMD algorithm, the curve is obtained by the average of upper and lower envelopes calculated by cubic splines through the extreme points. However, overshoots and undershoots are common, it can generate new extrema, shift or exaggerate the existing ones. The produced difference usually does not satisfy the symmetry conditions. Hence, the EMD algorithm

use repeated sifting procedure to achieve the goal. Inspired by this idea, we start our investigation by finding another local average procedure to replace the sifting process.

Let x be a signal , the usual average of x defined on $[a, b]$ is given by the following formula

$$\bar{x} = \frac{1}{b-a}\int_a^b x(t)dt, \tag{2.1}$$

or a local weighted average which will not be discussed in this paper. But for the data x covering serval oscillations on $[a, b]$, the average by (2.1) can not achieve our purpose mentioned above. To vividly illustrate it, we consider a toy example

$$x(t) = 6t + \cos(8\pi t),\ t \in [0, 1]. \tag{2.2}$$

The left of Figure 1 gives its profile, the average computed by (2.1), and the ideal average we are looking for. From it, the red line, i.e., the average computed by (2.1) clearly can not be the average we want. This is because x contains several varying vibrations, but the red line does not follow the featured oscillations. Inspired by these, to compute the average of any oscillating signal, we should partition it into several monotonic parts in advance. However, directly applying (2.1) on the monotonic intervals creates the staircase problem, as the red curve depicted in the right of Figure 1, various jump points arise on it. Motivated by those observations, the average we are looking for not only should be adaptivity and locality, but also smoothness.

Let $S[a, b]$ be the space of all the continuous and piecewise monotonic functions. That is, for each $x \in S[a, b]$, the domain $[a, b]$ can be partitioned into $a = a_1^x < a_2^x < \cdots < a_{m^x+1}^x = b$ such that x is monotonic on each $B_i^x = [a_i^x, a_{i+1}^x]$ for $i = 1, \cdots, m^x$, where the superscripts of x express the dependence of the partition on x. We denote by $\mathcal{M} : S[a, b] \to S[a, b]$ the average operator we look for. Then, for each $x \in S[a, b]$, we would like $\mathcal{M}x$ to be a local, adaptive and smooth function achieving the following goals:

goal 1: $\int_{B_i^x}(x - \mathcal{M}x)dt = 0$ for $i = 1, \ldots, m^x$;

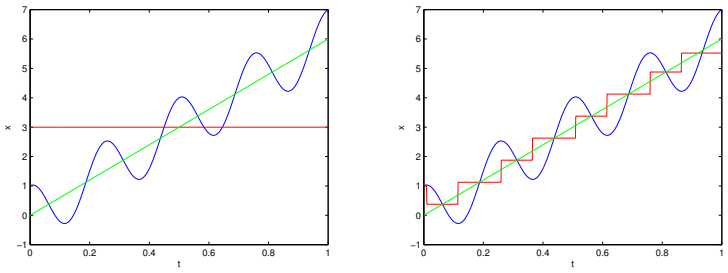

Figure 1: Left, blue: x, red: the average of x obtained from (2.1), green: the average we look for; Right, blue: x, red: the average given by utilizing (2.1) for each monotonic part of x, green: the average we look for.

goal 2: the average of $x-\mathcal{M}x$ should be zero, i.e., $\mathcal{M}(x-\mathcal{M}x)=0$. Here it needs to be emphasized that the average operator of x is different from that of $x-\mathcal{M}x$;

goal 3: $x-\mathcal{M}x$ is monotone on each B_i^x, $i=1,\ldots,m^x$.

It must be pointed out that the operator $\mathcal{M}$ is data dependent. That is, different data correspond to different average operators. Notice that, the first goal is to make $\mathcal{M}x$ be a local average; and the second one aims at ensuring $x-\mathcal{M}x$ to be symmetric with respect to zero; the last goal is to guarantee there is no new extrema appeared in $x-\mathcal{M}x$. We select optimization model to achieve the goals, which is given in the following subsection.

2.1 Adaptive and local average based on TV and TV$^\tau$ norms

For a function $g(t), t\in[a,b]$, a common way to measure its smoothness is through the τ-order TV norm

$$\mathrm{TV}^\tau(g)=\int_a^b |g^{(\tau+1)}(t)|dt, \qquad \tau\in\mathbb{N}\cup\{0\}. \tag{2.3}$$

However, in order to ensure that the information from regions far apart does not influence each other, we confine the TV$^\tau$ norm to characterize the smoothness separately on each partitioned monotone subinterval instead of the whole domain.

For any $x \in S[a,b]$, $\{B_i^x = [a_i^x, a_{i+1}^x]\}_{i=1}^{m^x}$ denotes its local monotone subintervals. Let T_x be the set of all the functions defined on $[a,b]$ satisfying the following conditions:

(i) $\int_{B_i^x}(x-g)dt = 0$, $i = 1, \ldots, m^x$;

(ii) $x - g$ is monotone on B_i^x, $i = 1, \ldots, m^x$;

(iii) $g(a_i^x) = \frac{1}{|B_{i-1}^x|+|B_i^x|}\int_{B_{i-1}^x \cup B_i^x} x(t)dt$, $i = 1, \ldots, m^x + 1$, where $B_0^x := B_{m^x+1}^x := \emptyset$.

We define the average $\mathcal{M}x$ as the solution of the following optimization problem:

$$\min_{g \in T_x} \ \alpha_1 \mathrm{TV}^0(g) + \alpha_2 \mathrm{TV}^\tau(g), \tag{2.4}$$

where $\tau \in \mathbb{N}^+$, α_1, α_2 are two positive penalty parameters satisfying $\alpha_1 + \alpha_2 = 1$.

Notice the condition (iii) in T_x, it makes the unknown function g be piecewise separable in problem (2.4). Hence, minimizing these total variation norms of x is intrinsically imposed on each B_i^x independently. Furthermore, taking both the 0-order and τ-order total variation norms in the objective are resulted from the following reasons. On one hand, the 0-order total variation norm is to account for those nonlinear and non-stationary data containing jump points. And on the other hand, using the τ-order total variation norm is to ensure the smoothness of the average. Hence, to make the average smooth, the parameter α_2 takes a larger value than α_1. When the signals are discontinuous, we use comparable values for both α_1 and α_2.

According to the conditions in T_x, it is easy to find that $\mathcal{M}x$ satisfies both the goal 1 and goal 3. Also, we will prove that it meets the goal 2 through the following theorem.

Theorem 1. *Let $x \in S[a,b]$ be the signal to be analyzed. If the average $\mathcal{M}x$ is generated from the optimization problem (2.4), then there holds $\mathcal{M}(x - \mathcal{M}x) = 0$.*

Proof. Let $B_i^x = [a_i^x, a_{i+1}^x]$, $i = 1, \ldots, m^x$, be the intervals on which x is monotone. Then $[a,b] = \cup_{i=1}^{m^x} B_i^x$. Consider the optimization problem (2.4), for any $g \in T_x$, let $x_1 := x - g$. Denote

$B_i^{x_1} = [a_i^{x_1}, a_{i+1}^{x_1}]$, $i = 1, \ldots, m^{x_1}$, as the intervals on which x_1 is monotone. Since x_1 is monotone on each B_i^x, it is followed that each $B_i^{x_1}$ is a union of some B_i^x, that is, $B_i^{x_1} = \cup_{j\in\Lambda_i} B_j^x$, where Λ_i is a subset of $\{1, \ldots, m^x\}$. Thus,

$$\int_{B_i^{x_1}} x_1(t)dt = \sum_{j\in\Lambda_i} \int_{B_j^x} (x(t) - g(t))dt = 0, \quad i = 1, \ldots, m^{x_1}.$$

It is followed immediately that

$$\int_{B_{i-1}^{x_1}\cup B_i^{x_1}} x_1(t)dt = 0, \quad i = 1, \ldots, m^{x_1} + 1,$$

where $B_0^{x_1} = B_{m^{x_1}+1}^{x_1} := \emptyset$. At last, x_1 is naturally monotonic on each $B_i^{x_1}$. These imply that $0 \in T_{x_1}$.

Let $g_1 \in T_{x_1}$ be a solution of (2.4). Then

$$\alpha_1 \mathrm{TV}^0(g_1) + \alpha_2 \mathrm{TV}^\tau(g_1) \le \alpha_1 \mathrm{TV}^0(0) + \alpha_2 \mathrm{TV}^\tau(0) = 0. \tag{2.5}$$

Notice that α_1, α_2 are positive, it concludes that g_1 equals a constant on $[a, b]$. Using the condition $g_1(a) = \int_{B_1^{x_1}} x_1(t)dt = 0$, we have that $g_1(t) = 0, \forall t \in [a, b]$. Hence, it yields $\mathcal{M}(x - \mathcal{M}x) = 0$. □

For Theorem 1, it indicates that our average can make the residue which is the average subtracted from the original signal, be symmetric with zero. That is, it meets the condition (2) of IMF in our average sense. Furthermore, similar to the analysis in [31], we have the following conclusion.

Corollary 1. *Let $x \in S[a, b]$, and T_x be the set defined in problem (2.4). Then for any continuous $g \in T_x$, the function $x - g$ has a single zero crossing on each of its strictly monotone interval.*

Proof. Let $B_i^x = [a_i^x, a_{i+1}^x]$, $i = 1, \ldots, m^x$, be the intervals on which x is monotone and a_i^x, $i = 1, \ldots, m^x$, be extreme points of x satisfying $a_i^x \neq a_{i+1}^x$ for $i = 1, \ldots, m^x$. Due to $g \in T_x$, we have

i) $x - g$ is monotonic on B_i^x and $\int_{B_i^x}(x - g)dt = 0$, $i = 1, \ldots, m^x$;

ii) $(x - g)|_{t=a_i^x} = \frac{1}{|B_{i-1}^x|+|B_i^x|} \int_{B_{i-1}^x\cup B_i^x} [x(a_i^x) - x(t)]dt$, $i = 1, \ldots, m^x + 1$.

Without loss of generality, suppose that x is increasing on B_i^x. Then $x(a_i^x) - x(t) \leq 0,\ \forall t \in B_{i-1}^x \cup B_i^x$, and $x(a_i^x) - x(a_{i+1}^x) < 0$. It is immediately followed that $(x-g)|_{t=a_i^x} < 0$. Similarly we have $(x-g)|_{t=a_{i+1}^x} > 0$. Since $x-g$ is monotone on B_i^x, it is clear to find that $x-g$ is increasing on B_i^x as x. Hence, $x-g$ has the same monotonicity with x.

Moreover, suppose that B_i^x is a strictly increasing interval of $x-g$. Since $x-g$ is continuous and satisfies $(x-g)|_{t=a_i^x} < 0$ and $(x-g)|_{t=a_{i+1}^x} > 0$, we clearly get it has a single zero crossing on B_i^x. The conclusion is achieved. □

Remark 1. *In the condition (iii) of T_x, there are many ways to define the values of $g(a_i^x), i = 1, \ldots, m^x + 1$, such as*

(1) $g(a_i^x) = \frac{x(a_{i-1}^x)+2x(a_i^x)+x(a_{i+1}^x)}{4}$, *where* $x(a_0^x) := 0$ *and* $x(a_{m^x+2}^x) := 0$;

(2) $g(a_i^x) = \sum_j \frac{1}{2^{|j-i+1|}\cdot|B_j^x|} \int_{B_j^x} x(t)dt$;

(3) $g(a_i^x) = \frac{1}{|B_{i-1}^x|+|B_i^x|} \int_{B_{i-1}^x \cup B_i^x} x(t)dt$.

We adopt the method (3) in this paper. This is because it has the following advantages over the others listed above. First, in contrast to (1), it fully uses the information on monotonic intervals of x rather than just the values on the three extreme points, this makes our method much better in grasping the trend of x. Second, the points used in (3) are just referred to the local monotonic intervals, which is different to the method in (2) from which the values are defined by the weighted average of the global information of x, this makes our method better in avoiding the global influence and robust to the noise perturbations.

In the following subsection, we propose the detail implementation for obtaining the numerical solution of the minimization problem (2.4).

2.2 Equivalent formula and numerical implementation

To compute the solution of the model (2.4), the major challenge is to make it satisfies the condition (ii) in T_x. We introduce the following proposition, which can not only give an equivalent description of this condition, and also provide an executable formula to be incorporated in computation.

Proposition 1. *If f is the solution of the following optimization problem*

$$\begin{aligned} &\min_{f\in S[a,b]\bigcap C^1[a,b]} TV^0(f) \\ &s.t.\ f(a)=\beta_1, f(b)=\beta_2, \end{aligned} \tag{2.6}$$

where β_1, β_2 are two given constants, then f is monotonic on $[a,b]$.

Proof. It is easy to see that any monotonic function in $C^1[a,b]$ is a solution of (**??**) and the minimal value of the objective functional is $|\beta_2-\beta_1|$. Since

$$TV^0(f)=\int_a^b |f'(t)|dt \geq \left|\int_a^b f'(t)dt\right| = |f(b)-f(a)| = |\beta_2-\beta_1| = \mathrm{TV}^0(f),$$

we conclude that $f'(t)=\lambda|f'(t)|,\ \forall t\in[a,b]$, where λ is the sign of $\beta_2-\beta_1$. Thus f is monotonic on $[a,b]$. □

Following Proposition 1, the condition (ii) in T_x can be expressed into a minimization problem. Thus, T_x can be equivalently reformed as follows:

(i) $\int_{B_i^x}(x-g)dt=0$ for $i=1,\ldots,m^x$;

(ii) and g is the solution of the minimization problem

$$\begin{aligned} &\min_g\ \mathrm{TV}^0(x-g) \\ &\text{s.t. } g(a_i^x)=\tfrac{1}{|B_{i-1}^x|+|B_i^x|}\int_{B_{i-1}^x\cup B_i^x} x(t)dt,\ i=1,\ldots,m^x+1. \end{aligned}$$

And then, similar to the penalty method in classical optimization theory [**?**], we approximate (2.4) as the following formula:

$$\begin{aligned} &\min_g \alpha \mathrm{TV}^0(x-g) + \alpha_1 \mathrm{TV}^0(g) + \alpha_2 \mathrm{TV}^\tau(g), \\ &\text{s.t.} \int_{B_i^x} (x-g)dt = 0 \text{ for } i = 1, \ldots, m^x; \\ &\quad g(a_i^x) = \tfrac{1}{|B_{i-1}^x|+|B_i^x|} \int_{B_{i-1}^x \cup B_i^x} x(t)dt, \ i = 1, \ldots, m^x + 1, \end{aligned} \tag{2.7}$$

where α_1, α_2 and B_i^x, $i = 0, 1, \ldots, m^x + 1$ are the same as those mentioned in (2.4), and α is a large enough positive parameter to control the monotonicity of $x - g$ on each B_i^x. This minimization problem can be discretized as an L_1 minimization problem, which has been well studied in the compressed sensing field [34, 6, 2].

Suppose the signal is uniformly sampled at t_i, $i = 1, 2, \ldots, n$, and $B_i^x = [t_{j_i}, t_{j_{i+1}}], i = 1, 2, \ldots m^x$. Then, the minimization problem (2.7) can be discretized as the following form:

$$\begin{aligned} &\min_{g \in \mathbb{R}^n} \ \alpha \|D^1(x-g)\|_1 + \alpha_1 \|D^1 g\|_1 + \alpha_2 \|D^{\tau+1} g\|_1 \\ &\text{s.t. } A(x-g) = 0 \\ &\quad Cg = b, \end{aligned} \tag{2.8}$$

where $A = (a_{il}) \in \mathbb{R}^{m^x \times n}$ denotes the piecewise integral operator, whose elements are defined by $a_{il} = 1$ when $t_l \in B_i^x$, and $a_{il} = 0$ otherwise, $i = 1, \ldots, m^x$; $b \in \mathbb{R}^{m^x+1}$, whose elements are defined as $b(i) = \frac{1}{|B_{i-1}^x|+|B_i^x|} \int_{B_{i-1}^x \cup B_i^x} x(t)dt$, $i = 1, \ldots, m^x + 1$; $C = (c_{il}) \in \mathbb{R}^{(m^x+1)\times n}$, where $c_{il} = 1$ when $t_l = t_{j_i}$, otherwise $c_{il} = 0$, $i = 1, \ldots, m^x + 1$; and $D^1 \in \mathbb{R}^{(n-1)\times n}$ and $D^{\tau+1} \in \mathbb{R}^{(n-\tau-1)\times n}$ are the 1-order and $(\tau+1)$-order differential matrixes respectively. Moreover, we utilize the split Bregman iterative algorithm discussed recently by Goldstein and Osher [34, 6, 2], which performs well for large-scale data, to solve the above L_1 optimization problem. The procedure is summarized in Algorithm 1.

The computational cost of Algorithm 1 contains the computation of g^k, d_1^k and d_2^k in the split Bregman iteration. Let n be the length of the signal. The g^k can be computed by using the

Algorithm 1 (Optimal Average)

1. initialization: $d_i^0 = 0,\ (i = 1, 2),\ u_i^0 = 0, \mu_i > 0, (i = 1, 2, 3, 4),\ g^0 = 0.$

2. while not converge, do

   - $u_1^{k+1} = u_1^k + A(x - g^k),\ u_2^{k+1} = u_2^k + b - Cg^k;$
   - $u_3^{k+1} = u_3^k + D^1 g^k - d_1^k,\ u_4^{k+1} = u_4^k + D^{\tau+1} g^k - d_2^k;$
   - $g^{k+1} \leftarrow \min\{\frac{\mu_1}{2}\|Ag - u_1^{k+1}\|^2 + \frac{\mu_2}{2}\|Cg - u_2^{k+1}\|^2 + \frac{\mu_3}{2}\|d_1^k - D^1 g - u_3^{k+1}\|^2 + \frac{\mu_4}{2}\|d_2^k - D^{\tau+1} g - u_4^{k+1}\|^2\};$
   - $d_1^{k+1} \leftarrow \min\{\alpha\|D^1 x - d_1\|_1 + \alpha_1\|d_1\|_1 + \frac{\mu_3}{2}\|d_1 - D^1 g^{k+1} - u_3^{k+1}\|^2\};$
   - $d_2^{k+1} \leftarrow \min\{\alpha_2\|d_2\|_1 + \frac{\mu_4}{2}\|d_2 - D^{\tau+1} g^{k+1} - u_4^{k+1}\|^2\}.$

3. end while.

existing optimization technique with computational complexity $O(n \log n)$ and d_1^k, d_2^k can be solved by the soft thresholding procedure respectively, which need only the operations of vector product and scalar contraction with a complexity $2O(n)$. If L iterations are implemented, then the total computational complexity is $L(O(n \log n) + 2O(n)) = L \cdot O(n \log n)$.

In the experiments in this paper, we take $\alpha = 500, \alpha_1 = 0.01, \alpha_2 = 0.99$ in problem (2.8) in general. We use weights $\alpha = 500, \alpha_1 = 0.5, \alpha_2 = 0.5$ only for discontinuous signals, such as the one in Example 1. Although the weights are given in advance and independent of the original signals, we found that the decomposition results are insensitive to the choices of the parameters in our experiments. For instance, we tried many different combinations of the parameters taking values in the ranges $\alpha \in [100, 1000]$, $\alpha_1 \in [0.3, 0.7]$ and $\alpha_2 = 1 - \alpha_1$ in Example 1, the results are similar to the ones that we depict in Figure 4. The same case is true for Example 2 when we take different parameter values $\alpha \in [100, 1000]$, $\alpha_1 \in [0.01, 0.2]$ and $\alpha_2 = 1 - \alpha_1$ respectively.

On the other hand, we put no restriction on the parameter τ except for $\tau \in \mathbb{N}^+$ in (2.4).

However, different choices of τ value give different results. For example, consider a signal defined by,

$$x(t) = \begin{cases} \cos(4\pi t) & t \in [0, 2] \\ 6(t-2) + \cos(8\pi t) & t \in (2, 4], \end{cases}$$

Figure 2 shows the averages obtained from (2.8) by taking $\tau = 1, 2, 3, 4$ respectively. As one can see, the first two averages are quite close to the desirable one. But when use $\tau = 3$ or 4, some small waves appear in the averages, and this is not ideal for our purpose. Hence, choosing a reasonable τ is important to the average. If the τ is too small, the average will lack of regularity. But if it is too large, it will bring in undulation effects as well. In this paper, we use $\tau = 2$, which seems to be the robustest among all the cases.

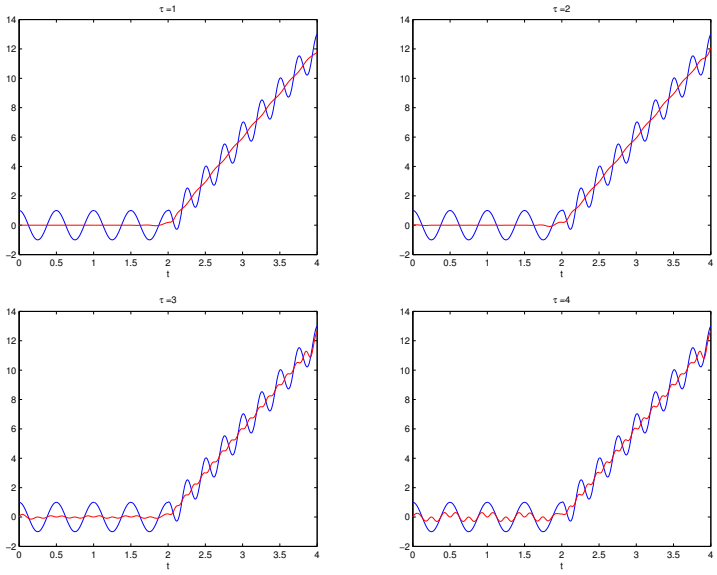

Figure 2: The averages generated from our approach with different τ. From left to right and top to bottom, the parameter τ is taken as $1, 2, 3, 4$ respectively.

Let us look back to the toy example $x(t) = 6t + \cos(8\pi t), t \in [0, 1]$ mentioned in (2.2). The red curve in Figure 3 shows the average obtained from our approach. Compare it with the two plots depicted in Figure 1, the average here are remarkably improved, and very close to the one we want

shown in the green line.

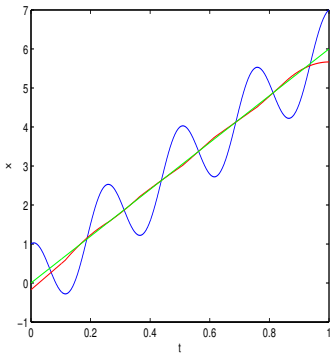

Figure 3: Blue: x, red: the average of x obtained from our approach, green: the average we are looking for.

3 An alternative algorithm to generate IMFs

Since our average is based on optimization model, we denote it as OA for simplicity. In this part, we primarily focus on using OA to perform an alternative algorithm for EMD. It is presented as follows.

3.1 OA-based data decomposition algorithm

Given a signal $x(t), t \in [a, b]$, let $s = x$, then the first IMF c_1 can be extracted by the following iteration.

1) Identify all the extrema of s, and then obtain the partition $\{B_i^s := [a_i^s, a_{i+1}^s]\}_{i=1}^{m^s}$ based on these extreme points.

2) Compute the average $\mathcal{M}s$ by solving the minimization problem (2.8). Then $c_1 = s - \mathcal{M}s$ is the first IMF.

The residue $r_1 = x - c_1$ is treated as the original signal, i.e., let $s = x - c_1$. And we apply it to the same iteration as described above again, the other IMFs can be calculated unless the residue is a

monotonic function. Finally, we obtain

$$x = \sum_{i=1}^{n} c_i + r_n. \tag{3.1}$$

3.2 Numerical experiments

In this section, we demonstrate the performance of the new alternative approach of EMD through a series of examples, which cover synthetic and real data. Prior to these, we first examine the performance of our average (OA) approach by comparing it with other approaches introduced in [11, 22, 25]. Since the averages presented in [11, 22, 25] are respectively based on cubic spline envelopes (CSE), a segment power-function based envelopes (SPFE) and monotone piecewise cubic interpolation based envelopes (IMCE). We denote them as CSA, SPFA and IMCA respectively in our experiments for simplicity, .

Example 1: We first consider the signal

$$x(t) = \begin{cases} \cos(12\pi t) & t \in [0,1] \cup [2,3] \\ 0.5 + 0.3\cos(8\pi t) & t \in (1,2), \end{cases} \tag{3.2}$$

which is composited by three intermittent signals. Since $\cos(12\pi t)$ and $0.3\cos(8\pi t)$ are monocomponent signals, the ideal average of x should be

$$\tilde{m}(t) = \begin{cases} 0 & t \in [0,1] \cup [2,3] \\ 0.5 & t \in (1,2). \end{cases}$$

Due to the discontinuity of x, to find an desirable average is challenging. This is reflected in Figure 4. The left side is the profile of the signal x, the middle figure depicts the averages obtained by the OA (blue line), the CSA (red), the IMCA (green dash) and the SPFA (black dash) respectively. Moreover, we plot the errors between the averages and the ideal one $\tilde{m}(t)$ on the right side. It is clear that the OA is the closest to $\tilde{m}(t)$. Meanwhile, we can use the following root mean square

(RMS) [26] to measure the error between an average $m(t)$ and the ideal one $\tilde{m}(t)$:

$$\mathrm{RMS}(m,\tilde{m}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|m(t_i)-\tilde{m}(t_i)|^2}. \tag{3.3}$$

Generally speaking, the smaller the RMS, the better the average is. Back to this example, Table 1 lists the RMSs of these averages with $\tilde{m}(t)$. It is also seen that the proposed OA is the best among all the cases.

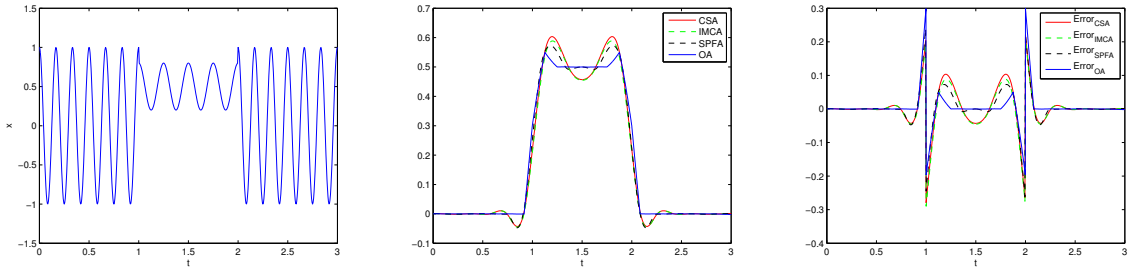

Figure 4: Left, the signal x. Middle, red solid: the CSA, green dash: the IMCA, black dash: the SPFA, blue solid: the OA. Right, red solid: CSA $-\tilde{m}$, green dash: IMCA $-\tilde{m}$, black dash: SPFA $-\tilde{m}$, blue solid: OA $-\tilde{m}$.

Table 1

	Methods			
	OA	IMCA	CSA	SPFA
RMSs	**0.0504**	0.0586	0.0595	0.0523

Example 2: In this example, we consider a monocomponent signal $x(t) = (2t+\cos t^2)\cos(2\pi t + 0.2\cos t)$, $t \in [0, 10]$. As we know, the ideal average $\tilde{m}(t)$ of any monocomponent signal should be zero. This actually can not be attained by the existing approaches. Figure 5 shows the profile of the signal and its average including the OA, the CSA, the IMCA and the SPFA. Clearly, we can also see that the OA is the closest to $\tilde{m}(t)$ from the right of Figure 5. Meanwhile, Table 2 shows

the RMSs of the averages. These all indicate the OA is best among them.

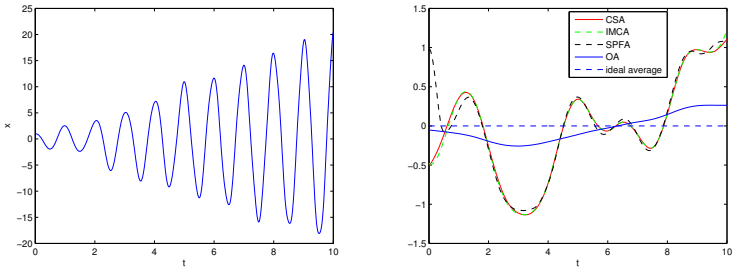

Figure 5: Left, the signal x. Right, red solid: the CSA, green dash: the IMCA, black dash: the SPFA, blue solid: the OA, blue dash: the ideal average $\tilde{m}$.

Table 2

	Methods			
	OA	IMCA	CSA	SPFA
RMSs	**0.1715**	0.6168	0.6094	0.6210

In the following examples, we test our optimal average method on several signals that may be non-stationary, or polluted by noise, or from the real world applications. We also compare our method with some existing approaches, such as the M-LFBF algorithm introduced in [9], and emsamble EMD (EEMD) proposed in[12, 13]. In addition, we calculate the normalized inner product of two components c_1 and c_2 to measure the angle between them, i.e.

$$\rho(c_1, c_2) = \frac{|\langle c_1, c_2 \rangle|}{\|c_1\|_2 \|c_2\|_2}.$$

It is clear that the smaller values of $\rho(c_1, c_2)$ indicate the components are near orthogonal. We want to remind that M-LFBF enforces the near orthogonality in the optimization problem specifically, while we do not in our optimal average algorithms.

Example 3: We consider the signal composed by a couple of orthogonal components $x(t) =$

$\cos(2\pi t) + \sin(8\pi t)$. The left of Figure 6 plots the original and the results obtained from our approach. We use the M-LFBF algorithm to decompose the same signal, and its results are shown in the right of Figure 6. The normalized inner products between the IMFs obtained from our approach and the M-LFBF respectively are computed and listed in Table 3. The numbers show that the two IMFs obtained by our method are also near orthogonal.

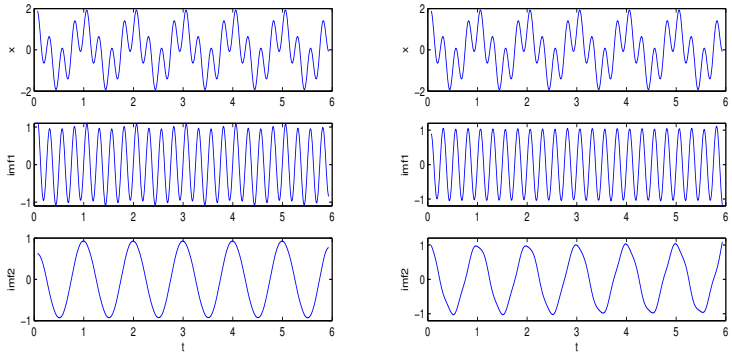

Figure 6: The original and IMFs in Example 3. Left: our approach; Right: the M-LFBF method.

Table 3: Performance of the orthogonality in Example 3.

ρ value	Our approach		ρ value	M-LFBF	
	imf1	imf2		imf1	imf2
imf1	1	0.0661	imf1	1	0.0354
imf2	0.0661	1	imf2	0.0354	1

Example 4: We test a non-stationary signal $x(t) = \cos(t) + \cos(t^2 + t + \cos(t))$, which is formed by two non-orthogonal components. Using our approach yields essentially a perfect decomposition with two IMFs and a residue that is close to zero, see the left of Figure 7. On the right of Figure 7, we depict the results from the M-LFBF algorithm. Table 4 lists the normalized inner products between the components obtained by these two methods respectively. This example demonstrates

that it is not ideal to enforce the orthogonality in the decompositions if the signal consists of non-orthogonal intrinsic components.

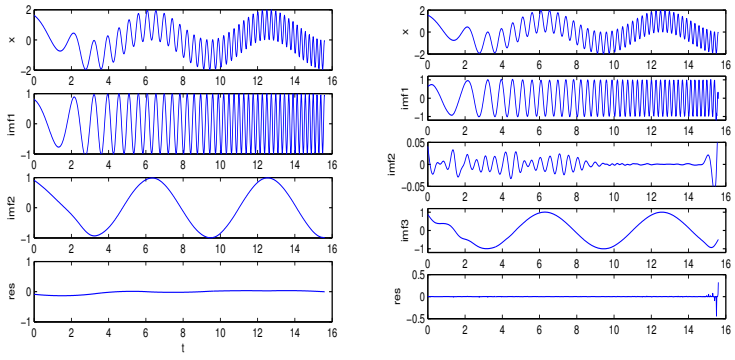

Figure 7: The original signal, IMFs and the residue of Example 4. Left: our approach; Right: the M-LFBF method.

Table 4: Performance of the orthogonality in Example 4.

ρ value	Our approach			ρ value	M-LFBF			
	imf1	imf2	res		imf1	imf2	imf3	res
imf1	1	0.0316	0.0195	imf1	1	0.0375	0.0058	0.0212
imf2	0.0316	1	0.0665	imf2	0.0375	1	0.0318	0.2565
res	0.0195	0.0665	1	imf3	0.0058	0.0318	1	0.0055
				res	0.0212	0.2565	0.0055	1

Example 5: In this example, we test our approach on signal polluted by Gaussian noise. Here the signal is given by

$$x(t) = 0.6t + \sin(t) + \sin(3t) + \epsilon(t), \tag{3.4}$$

where $\epsilon(t)$ is a Gaussian noise. Our method generates four IMFs and a trend, which are depicted in the top of Figure 8. The three plots on the left correspond to the polluted signal and noise

components respectively, the other three components on the right are very close to the intrinsic components of x. We also use the EEMD [1] technique [12, 13] to decompose the signal and show its results in the bottom of Figure 8. Table 5 lists the normalized inner products between each two components of our method, it also shows our method can generate near-orthogonal components in this case.

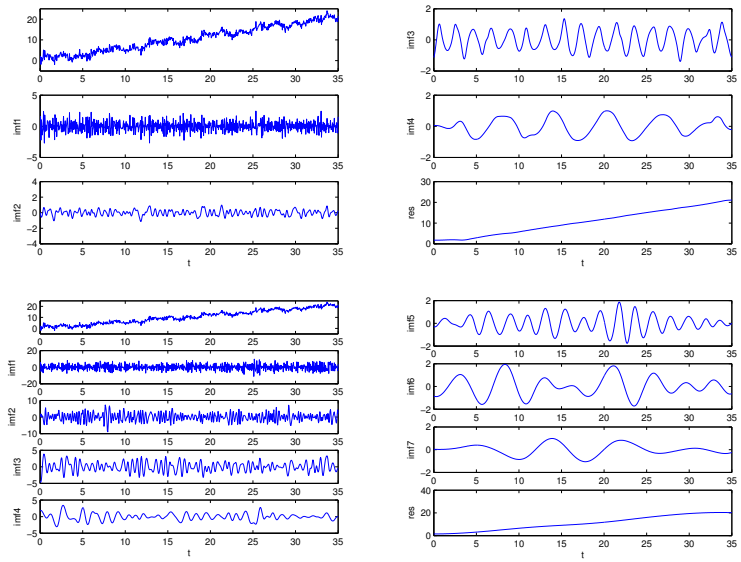

Figure 8: Top: the figures are corresponding to the original signal and its components obtained from our approach in Example 5. Bottom: these are the original signal and its components obtained from EEMD.

Example 6: In the previous examples, we use several synthetic signals to test our approach. Here, we consider the length of the day (LOD) data, which covers 700 consecutive days starting from 1978. The original signal and the IMFs derived from our approach are plotted in Figure 9. Like the EMD approach, our method can recover almost all the physically meaningful IMFs. Clearly, in Figure 9, the first IMF captures the semi monthly cycles while the second IMF indicates the monthly oscillations. Similarly, the third IMF captures the semi annual cycle and the last IMF captures the annual tendency. Table 6 indicates the angles between the components obtained by

[1] We use the code made by M. E. Torres et. al., it is available at http://perso.ens-lyon.fr/patrick.flandrin/emd.html.

our method.

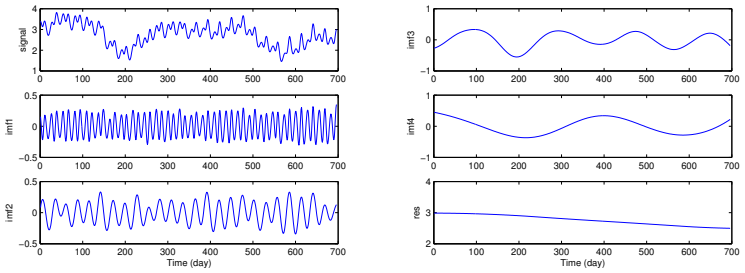

Figure 9: The IMFs and trend of LOD data in Example 6.

4 Conclusion

The EMD has found many applications in a variety of problems covering signal processing, pattern recognition, biomedicine engineering and ocean science since its advent. Although it often proves remarkably effective, a mathematical foundation is virtually nonexistent. The difficulty is partly due to the average based on cubic spline. In view of this, we present a novel adaptive and local average based on local optimization to overcome this problem. The improvements in our approach

Table 5: Performance of the orthogonality in Example 5.

ρ value	IMFs and residue				
	imf1	imf2	imf3	imf4	res
imf1	1	0.1789	0.0070	0.0040	0.0073
imf2	0.1789	1	0.3219	0.0408	6.1486×10^{-4}
imf3	0.0070	0.3219	1	0.1078	0.0016
imf4	0.0040	0.0408	0.1078	1	0.0466
res	0.0073	6.1486×10^{-4}	0.0016	0.0466	1

can be summarized as follows. First, our approach does not use any pre-defined functions, such as AM-FM functions or wavelets as the basis for IMFs, this makes our approach share similar flexibilities with the original EMD method. And second, the objective and the constraint conditions are defined on the monotonic intervals of the signal, this can avoid the global influence and then makes our approach robust to the noise perturbations. We also proved that the IMFs obtained by our alternative approa ch satisfy the two requirements proposed by Huang et. el. in the original EMD method. Moreover, the existing theory on total variation minimization provides convergence guarantee for our iterations. In addition, the experiments show very encouraging results, and we hope that this alternative approach can lead to more stable performance in applications, which will be considered in our future research studies.

References

[1] S. Boyd, and L. Vandenberghe. Convex optimization. *Cambridge university press*, 2009.

[2] N. Bi, Q. Sun, D. Huang, Z. Yang, and J. Huang. Robust image watermarking based on

Table 6: Performance of the orthogonality in Example 6.

ρ value	IMFs and residue				
	imf1	imf2	imf3	imf4	res
imf1	1	0.0391	0.0175	0.0059	0.0110
imf2	0.0391	1	0.0687	0.0339	0.0103
imf3	0.0175	0.0687	1	0.3264	0.0037
imf4	0.0059	0.0339	0.3264	1	0.0486
res	0.0110	0.0103	0.0037	0.0486	1

multiband wavelets and empirical mode decomposition. *Image Processing, IEEE Transactions on*, 16(8):1956–1966, 2007.

[3] J.-F. Cai, S. Osher, and Z. Shen. Split bregman methods and frame based image restoration. *Multiscale Modeling and Simulation*, 8(2):337–369, 2009.

[4] Q. Chen, N. E. Hunag, S. Riemenschneider, and Y. Xu. A b-spline approach for empirical mode decompositions. *Advances in Computational Mathematics*, 24(1-4):171–195, 2006.

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

[6] P. Flandrin, G. Rilling, and P. Goncalves. Empirical mode decomposition as a filter bank. *Signal Processing Letters, IEEE*, 11(2):112–114, 2004.

[7] T. Goldstein and S. Osher. The split bregman method for l_1-regularized problems. *SIAM Journal on Imaging Science*, 2(2):323–343, 2009.

[8] T. Y. Hou and Z. Shi. Adaptive data analysis via sparse time-frequency representation. *Advances in Adaptive Data Analysis*, 3(01n02):1–28, 2011.

[9] S. L. Peng and W.-L. Hwang. Adaptive signal decomposition based on local narrow band signals. *Signal Processing, IEEE Transactions on*, 56(7):2269–2676, 2008.

[10] N. Pustelnik, P. Borgnat, and P. Flandrin. A multicomponent proximal algorithm for Empirical Mode Decomposition. *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*,1880–1884, 2012.

[11] N. E. Huang, Z. Shen, and S. R. Long. A new view of nonlinear water waves: The hilbert spectrum 1. *Annual Review of Fluid Mechanics*, 31(1):417–457, 1999.

[12] N. E. Huang, Z. Shen, and S. R. Long et. al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998.

[13] Z. H. Wu, and N. E. Huang Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, 01(1):1–41, 2009.

[14] M. E.Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin A complete ensemble empirical mode decomposition with adaptive noise. *IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-11*, 4144–4147, 2011.

[15] W. Huang, Z. Shen, N. E. Huang, and Y. C. Fung. Engineering analysis of biological variables: an example of blood pressure over 1 day. *Proceedings of the National Academy of Sciences*, 95(9):4816–4821, 1998.

[16] X. Hu, S. L. Peng, and W.-L. Hwang. Multicomponent AM-FM signal separation and demodulation with null space pursuit *Signal, Image and Video Processing*, 7(6):1093–1102, 2013.

[17] H. Kantz and T. Schreiber. *Nonlinear time series analysis*, volume 7. Cambridge University Press, 2004.

[18] L. Lin, Y. Wang, and H.-M. Zhou. Iterative filtering as an alternative algorithm for empirical mode decomposition. *Advances in Adaptive Data Analysis*, 1(04):543–560, 2009.

[19] S. Meignen and V. Perrier. A new formulation for empirical mode decomposition based on constrained optimization. *Signal Processing Letters, IEEE*, 14(12):932–935, 2007.

[20] J. C. Nunes, S. Guyot, and E. Deléchelle. Texture analysis based on local analysis of the bidimensional empirical mode decomposition. *Machine Vision and Applications*, 16(3):177–188, 2005.

[21] U. Parlitz. Nonlinear time-series analysis. In *Nonlinear Modeling*, pages 209–239. Springer, 1998.

[22] S. C. Phillips, R. J. Gledhill, J. W. Essex, and C. M. Edge. Application of the hilbert-huang transform to the analysis of molecular dynamics simulations. *The Journal of Physical Chemistry A*, 107(24):4869–4876, 2003.

[23] S. R. Qin and Y. M. Zhong. A new envelope algorithm of hilbert–huang transform. *Mechanical Systems and Signal Processing*, 20(8):1941–1952, 2006.

[24] G. Rilling and P. Flandrin. One or two frequencies? the empirical mode decomposition answers. *Signal Processing, IEEE Transactions on*, 56(1):85–95, 2008.

[25] H. Tong. *Nonlinear time series analysis*. Wiley Online Library, 2007.

[26] L. Yang, Z. Yang, L. Yang, and P. Zhang. An improved envelope algorithm for eliminating undershoots. *Digital Signal Processing*, 23(1):401–411, 2013.

[27] L. Yang, Z. Yang, F. Zhou, and L. Yang. A novel envelope model based on convex constrained optimization. *Digital Signal Processing*, 2014.

[28] N. Pustelnik, P. Borgnat and F. Patrick. Empirical mode decomposition revisited by multi-component non-smooth convex optimization. *Signal Processing*, 102(1):313–331, 2014.

[29] Z. Yang, D. Huang, and L. Yang. A novel pitch period detection algorithm based on hilbert-huang transform. In *Advances in Biometric Person Authentication*, pages 586–593. Springer, 2005.

[30] S. L. Peng, and W.-L. Hwang. Null space pursuit: An operator-based approach to adaptive signal separation In *Signal Processing, IEEE Transactions on*, 58(5): 2475–2483. 2010.

[31] Z. Yang, D. Qi, and L. Yang. Signal period analysis based on hilbert-huang transform and its application to texture analysis. In *Image and Graphics, 2004. Proceedings. Third International Conference on*, pages 430–433. IEEE, 2004.

[32] Z. Yang and L. Yang. A new definition of the intrinsic mode function. In *Proceedings of World Academy of Science, Engineering and Technology*, volume 60, pages 822–825. Citeseer, 2009.

[33] Z. Yang, L. Yang, and D. Qi. Detection of spindles in sleep eegs using a novel algorithm based on the hilbert-huang transform. In *Wavelet Analysis and Applications*, pages 543–559. Springer, 2007.

[34] Z. Yang, L. Yang, D. Qi, and C. Y. Suen. An emd-based recognition method for chinese fonts and styles. *Pattern Recognition Letters*, 27(14):1692–1701, 2006.

[35] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for l_1-minimization with applications to compressed sensing. *SIAM Journal on Imaging Sciences*, 1(1):143–168, 2008.