EMD ALGORITHM BASED ON BANDWIDTH AND THE APPLICATION ON ONE ECONOMIC DATA ANALYSIS

Xie Qiwei¹ 2, Xuan Bo³, Li Jianping², Xu Weixuan², Han Hua³

1. School of Management, University of Science and Technology of China
   Hefei 230026, China
2. Institute of Policy and Management, Chinese Academy of Sciences
   Beijing 100080, China
3. Institute of Automation, Chinese Academy of Sciences
   Beijing 100080, China
   (e-mail: qw_xie@yahoo.com.cn)

ABSTRACT

This paper proposes a key point to improve the empirical mode decomposition algorithm. Recent works have demonstrated that EMD has remarkable effect in many applications, but EMD also has many problems. By analyzing the simulated and actual signals, it is confirmed that the IMFs get by the bandwidth criterion can approach the real components and reflect the intrinsic information of the analyzed signal. The paper not only updates the bandwidth criterion of the empirical mode decomposition algorithm, but also uses it on one data of the electric consumption data to extract some periodic rules. Based on the decomposition result, we obtain three conclusions which can instruct electricity production.

1. INTRODUCTION

Empirical mode decomposition, introduced by N.E.Huang et al. [1] in 1998, is a method for decomposing complex, multicomponent signal into several elementary intrinsic mode functions (IMFs). Differing from Fourier and wavelet analysis which have predefined basis, EMD only uses scale and frequency characters of the original signal. EMD is a local, fully data-driven and self-adaptive analysis approach. Moreover, the combination of EMD and the associated Hilbert spectral analysis can offer a powerful method for time frequency analysis.

Although it has often proved effective in many applications [1, 2], EMD method has many drawbacks. Firstly, it lacks of mathematical base which can represent EMD method naturally. Secondly, it is a pity that the criteria considered so far [1, 3, 4, 5] are all constraints on the amplitude and have nothing to do with the frequency or phase of the IMF. As a result, IMF obtained based on those criteria will have dramatically different frequencies and cause scale mixing problem.

This paper improves the EMD algorithm based on the two aspects and uses the updated EMD algorithm to analyse a time sequence about electricity consumption.

2. STOP CRITERION IMPROVEMENT

2.1 Emd Algorithm

Huang [1] proposed a class of functions designated as intrinsic mode functions to make their instantaneous frequencies have meaningful interpretation. Based on IMFs definition, sifting process can get a component approximation to the real component of the analyzed signal. The IMF and the sifting process are defined as follows. We will present EMD briefly, and details are available in [1, 6].

A function is an IMF if it satisfies two conditions: 1) in the whole data set, the number of extrema and the number of zeros crossings must either equal or differ at most by one; and 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.

EMD uses sifting process to extract IMFs from the analyzed signal. Sifting process can be summarized as follows.

1) Find all local minima and maxima of \( r(t) \)
2) Get the upper envelop \( e_{max}(t) \) by interpolating between maxima. Similarly get the lower envelop \( e_{min}(t) \) with minima
3) Compute the mean envelop \( m(t) = (e_{max}(t) + e_{min}(t))/2 \)
4) Let \( i = i + 1 \) and define the proto-mode function (PMF) as \( p_i(t) = r(t) - m(t) \), and let \( r(t) = p_i(t) \)
5) Repeat steps 1 – 4 on PMF \( p_i(t) \) until it is an IMF, then record the IMF \( IMF_i(t) = p_i(t) \)
6) Let \( r(t) = r(t) - IMF_i(t) \), if the extremum point number of \( r(t) \) is larger than three, let \( k = k + 1 \), \( i = 0 \), and go to step 1; otherwise finish the sifting process.

2.2 IMF Criterion

Given a real valued signal \( x(t) \), using the Hilbert transform and analytical signal theory [7] we can get a complex valued signal whose real part equals to \( x(t) \). For this reason, we use complex valued signal directly hereafter in this section. Given a normalized (energy is 1) signal \( z(t) = a(t)e^{j\varphi(t)} \), the instantaneous frequency of \( z(t) \) is defined as \( \varphi'(t) \). A signal has meaningful instantaneous frequency if it is a monocomponent signal or the sum of two unequal strength tones. For lack of exact definition of the monocomponent signal was given by now, Huang [1] used the two IMF condition to replace monocomponent. To guarantee the IMF components retain enough physical sense, Huang [1] limited the size of the standard deviation (SD). SD is computed from two con-
secutive PMFs as

\[ SD = \sum_{i=0}^{T} \frac{|p_{r-1}(t) - p_{i}(t)|^2}{p_{r-1}(t)}. \] (1)

The IMF conditions can get good component from the analyzed signal sometimes, but the instantaneous frequencies of IMFs usually are meaningless because the IMFs are not monocomponent. Then SD criterion from (1) is inappropriately to distern one component which is mixed with other scale signal. As an improvement to the SD criterion, Rilling et al. [4] brought forward a 3-threshold criterion. The shortcoming of the 3-thresholds criterion is that the thresholds do not adapt to the signal.

Recently, Cheng [3] put forward the energy difference tracking method based on the assumption that the residue and IMFs are orthogonal mutually. However, the IMFs of nonlinear and nonstationary system are not orthogonal. Even for linear and stationary system, the real components of the discrete signal we get will not be orthogonal as a result of environment noise. For these reasons, the energy difference tracking method can’t guarantee that the IMFs are the best approximation of the real components.

Damerval [5] proposed a criterion based on the number of iterations and the number of IMFs for bidimensional EMD. This criterion saves computational cost and has little boundary effect in sifting process. The number of iterations and IMFs should be selected carefully. Too few sifting steps can not eliminate the riding waves and IMFs obtained will dissatisfaction the two IMF conditions.

None of the above criterion uses frequency and phase information of the analyzed signal. This paper will put forward local narrow band signal to substitute monocomponent signal. A signal \( z(t) = a(t)e^{i\phi(t)} \) narrowband if \( a(t) \) is a bandlimited signal and the highest frequency of \( a(t) \) is far less than \( \phi(t) \). If any little segment of a signal is narrow banded, then we call the signal local narrowband. Bandwidth criterion which is related the essence of the monocomponent is proposed in [6]. Suppose \( \int a^2(t)dt = 1 \) and \( B \) is bandwidth of \( z(t) \), the bandwidth has two terms:

\[ B^2 = B^2_s + B^2_I \] (2)

where

\[ B^2_s = \int \left[ \frac{a'(t)}{a(t)} \right]^2 a^2(t)dt \] (3)

\[ B^2_I = \int [\phi'(t) - \langle \omega \rangle]^2 a^2(t)dt \] (4)

\[ \langle \omega \rangle = \int \phi'(t)a^2(t)dt. \] (5)

Equation (2) implies that bandwidth \( B \) has two terms: \( B_s \) and \( B_I \). We call \( B_I \) frequency bandwidth and \( B_s \) amplitude bandwidth. \( B_s \) results from the changes of \( a(t) \) and only associates with amplitude modulating. \( B_I \) results from the changes of instantaneous frequency and reflects the consistency of instantaneous frequency at all time extend. The smaller \( B_I \), the closer the scale characteristics at different times are, and the slighter the scale mixing problem is. So we can take \( B_I \) as a stop criterion for EMD sifting process.

2.3 Instantaneous Frequency estimation

There are many algorithms to compute instantaneous frequency, such as phase differencing of the analytic signal, Teager-Kaiser operator, counting the zeros-crossings (Z-Cs), and adaptive estimation methods based on the least mean square (LMS) algorithm etc. Since IF should be unrelated with amplitude, we choose zeros-crossings points to estimate the frequency of the analytic signal. For a sinusoidal signal, the frequency is given by the inverse of the period, or alternatively by half the inverse of the interval between zeros-crossings, i.e.,

\[ f = \frac{1}{2T_e} \] (6)

or

\[ f = \frac{Z}{2} \] (7)

where \( T_e \) is the interval between zero crossings, \( 2T_e \) is the period, \( f \) is the frequency, and \( Z \) is the zero-crossing rate. To reduce the variance of the zero-crossings estimate, an estimator is defined by

\[ f = \frac{Z(n)}{2} \] (8)

\[ Z(n) = \sum_{m=-M}^{M} [\text{sgn}[s(m)] - \text{sgn}[s(m-1)]] h(m-n) \] (9)

where \( M \) is the length of computation window and

\[ \text{sgn}[s(n)] = \begin{cases} 1, & \text{for } s(n) \geq 0 \\ -1, & \text{for } s(n) < 0 \end{cases} \]

and where

\[ h(n) = \begin{cases} 1/2M, & \text{for } 0 \leq n \leq M-1 \\ 0, & \text{otherwise}. \end{cases} \]

The details are available in [6]. We denote the variance of \( f \) is \( \sigma_f^2 \). Then the smaller the \( \sigma_f^2 \), the closer the frequencies at different sample points are. If we get a minimum for \( \sigma_f^2 \) during sifting process, we consider that current signal is an IMF.

3. OUR SIFTING PROCESS

Given a time-series \( f(t) \), \( t = 1, \ldots, N \), threshold value \( T \), let \( r(t) = f(t) \), we can modified the sifting process in EMD as follows.

1) Find all local minima and maxima of \( r(t) \)
2) Get the upper envelop \( \epsilon_{\text{max}}(t) \) by interpolating between maxima. Similarly get the lower envelop \( \epsilon_{\text{min}}(t) \) with minima
3) Compute the mean envelop as an approximation to the local average, \( m(t) = (\epsilon_{\text{max}}(t) + \epsilon_{\text{min}}(t))/2 \)
4) Extracted the proto-mode function (PMF): \( p_i(t) = r(t) - m(t) \), and let \( r(t) = p_i(t) \)
5) Repeat steps 1 – 4 on PMF \( p_i(t) \), stop criterion is presented by next three subsidiary process:
   a) Use 3-thresholds(\( \theta_1, \theta_2, \theta_3 \)) criterion to get PMF that almost satisfies the two conditions of IMF
   b) Continue the sifting process until we find the minimum for \( \sigma_{\text{PMF}} \).
c) Take the final PMF as an IMF, then the IMF has small frequency bandwidth mild scale mixing problem. Then record the IMF imfd(t) = p(t)
6) Let r(t) = r(t) − imfd(t), if the extremum point number of r(t) is larger than three, let k = k + 1, i = 0, and go to step 1; otherwise finish the sifting process.

4. ECONOMIC DATA ANALYSIS

We use the economic data (Poland everyday electricity consumption from 1990 to 1994 [8, 9]) to test our algorithm. This time-series data are shown in Fig. 1.

![Figure 1: The electricity load values of Poland from 1990’s.](image)

Business and economic time series frequently exhibit seasonality-period fluctuations that recur with about the same intensity each year. Economists are far more interested in the delicate patterns of the fluctuations which are superimposed upon the trend only in order to see these pattern more clearly. They consider economic time series can be represented as:

\[ Z(t) = S(t) + T(t) \]  \hspace{1cm} (10)

where \( S(t) \) and \( T(t) \) are unobservable seasonal and trend component, respectively. So the objective of economists is to find a decomposition

\[ Z(t) = m(t) + q(t) \] \hspace{1cm} (11)

\[ m(t) = E\{T(t)|Z(t)\} \] \hspace{1cm} (12)

\[ q(t) = E\{S(t)|Y(t)\} \] \hspace{1cm} (13)

They assume that each of the components follows an ARMA model,

\[ T(t) = \theta(L)\varepsilon(t) \] \hspace{1cm} (14)

\[ S(t) = \frac{\theta(L)}{P(L)\phi(L)}\varepsilon(t) \] \hspace{1cm} (15)

where \( L \) is the back-shift operator such that \( LS(t) = S(t-1) \), \( \mu(t), \alpha(L), \phi(L) \) are the ARMA operators, and \( P(L) = 1 + L + \cdots + L^{n-1} \). \( \varepsilon(t) \) are two mutually independent white noise with standard deviation \( \sigma_\varepsilon^2 \) and \( \sigma_\varepsilon^2 \), respectively. Based on this assumption, they get

\[ \beta_T(L) = \lambda \psi^{-1}(L) \frac{\alpha(L)\mu(L)\phi(L)}{\psi^{-1}(L)} \] \hspace{1cm} (17)

\[ s.t. m(t) = \beta_T Z(t), q(t) = \beta_R Z(t) \] \hspace{1cm} (18)

The details are available in [10][11]. They have two problems: 1) The model is not fully consider the intrinsic problem of the signal. 2) How to select the parameter for the model. Whereas we have presented the Bandwidth EMD algorithm, we consider that it is a good instrument to decompose the signal to many components which is more physical significance. Then we will demonstrate decomposed result of the Fig.1 and analyze the implication of the IMFs.

![Figure 2: The 8th IMF with 3-threshold criterion [4] and its’ Hilbert spectrum.](image)

![Figure 3: The 6th IMF of Damerval criterion and its’ Hilbert spectrum.](image)

If we don’t use the bandwidth stop criterion, we will get the results which are not perfect. We will show the compared result in Fig.2-5, Fig.2-5 are not used bandwidth stop criterion and Fig.6 is used. The right part of the figures is Hilbert spectrum of the IMF which is the seasonal component corresponding to the cyclic part that the period is one year. The thresholds used in this papers are \( \alpha_1 = 0.01, \theta_1 = 0.01, \) and \( \theta_2 = 0.1 \) for 3-threshold criterion(3-threshold criterion embody the classical EMD method, so we do not compare our result with the classical EMD method); the number of iterations is 10 for Damerval criterion. From the Fig 2-5, we provide that band criterion reflects more essential content and the result confirms seasonal factor.

It is shown that our sifting process can improve the result. We will analyze the electricity consumption with our
algorithm, and obtain some periodic rules of the electricity consumption. The decomposition result is shown in Fig.6, which is \{imf_i(t), i = 1, 2,..., 11\}, and imf_{11}(t) is the trend item.

Define the energy \(E(\cdot)\) and energy ratio \(E_r(\cdot)\) to investigate the decomposition items.

\[
E(imf_j) = \sum_{t=1}^{1400} imf_j^2(t) \quad (19)
\]

\[
E_r(imf_j) = \frac{\sqrt{E(imf_j)}}{\sqrt{\sum_{j=1}^{10} E(imf_j)}} \quad (20)
\]

We use energy defined in (19) and energy ration defined in (20) to select appropriate IMFs.

According to \(E\) and \(E_r\), table I shows that the actually items are \(h_1(t), h_2(t), h_0(t)\) and \(h_{11}(t)\). The reconstruction of \(x(t)\) with these components is

\[
f_1(t) = h_1(t) + h_2(t) + h_0(t) + h_{11}(t) \quad (21)
\]

We use absolute error ratio and similarity ratio to measure difference between reconstructed function \(f_1\) and the original signal \(f\). They are presented by:

\[
\sqrt{\frac{\sum(f - \bar{f})(f_1 - \bar{f}_1)}{\sqrt{E(f)E(f_1)}}} = 0.0483 \quad \text{(absolute error ratio)} \quad (22)
\]

\[
\frac{\sum(f - \bar{f})(f_1 - \bar{f}_1)}{\sqrt{E(f)E(f_1)}} = 0.9587 \quad \text{(similarity ratio)} \quad (23)
\]

where \(\bar{f}\) and \(\bar{f}_1\) are the mean of \(f\) and \(f_1\). Similarity ratio measure the correlation coefficient of the reconstructed function \(f_1\) and the original function \(f\). The reconstructed function \(f_1\) is shown in Fig. 7.

Fig. 6 illustrates the trend item \(imf_{11}(t)\). \(imf_{11}(t)\) is a slow-moving descendent trend which is difficult to detect in the original function \(x(t)\).

\(imf_9(t)\) is local-narrowed function. We investigate the zeroscrossings and extreme points in Tab2. The interval of zeroscrossings is 182.8 and the interval of extreme points is 180.85. If we consider the interval of zeroscrossings and extreme points, we can get that \(imf_9(t)\) is regarded as a year periodic time-series.

For \(imf_2(t)\) is a high-frequency component, extreme points number is very large. We count the interval extreme points’ distribution to express the periodicity. The interval distribution of the maximal points is shown in Fig.8. We can conclude that \(imf_2(t)\) is a week periodic series.

Similarly, the interval distribution of \(imf_1(t)\) is shown in Fig.8, and we conclude that it is half-week periodic series. From Fig.1 we gain that the electricity consumption presents the periodic rules, but these rules are mixed and contained noise. If apply our sifting algorithm, we can separate the periodic rules. So we obtain three key conclusions from above:

I The consumption of the electricity is slow descended from 1990 to 1994 which is appeared from the trend item \(imf_{11}\);

II The consumption of electricity expresses periodic property which is respectively half-week, one week and one year;

III These three kinds of motions of period is obviously stronger than any other motions, so we should pay attention to these motions of period when we are confront with electricity coordination.

This kind of periodic analysis is helpful to assign national electricity power. The result shows when carry on maintain, when should let full burden of equipments revolve and at which time we should pay attention to.

5. CONCLUSION

EMD is a generally nonlinear, non-stationary, data-driven, adaptive algorithm. We advance a new criterion to aim at the disadvantages of EMD. We take the experiment with electricity data and get some good results. This we only involve a few mathematical foundation, we should do some work on this aspect next research. Except that there are some problems which we should pay attention to, such as extremum
6. ACKNOWLEDGEMENT

This research is supported by the National Natural Science Foundation of China under Grant No.70531040. The authors would like to thank the three anonymous referees for their many very helpful comments and suggestions. We would also like to thank Professor Peng Silong for his suggestions.

REFERENCES


