International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:3, No:12, 2009

A new definition of the intrinsic mode function

Zhihua Yang, and Lihua Yang

Abstract—This paper makes a detailed analysis regarding the definition of the intrinsic mode function and proves that Condition 1 of the intrinsic mode function can really be deduced from Condition 2. Finally, an improved definition of the intrinsic mode function is given.

Keywords—Empirical Mode Decomposition (EMD), Hilbert-Huang transform(HHT), Intrinsic Mode Function(IMF).

I. Introduction

Both time analysis and frequency analysis are the basic signal processing methods. Some fundamental physical quantities such as the field, pressure, and voltage, themselves change in time, so they are called "time waveforms" or "signals". The time analysis, which investigates the variation of a signal with respect to time, is fundamental because a signal itself is a time waveform. However, to probe deeper, the study of different representations of a signal is often useful. This study is implemented by expanding a signal into a complete set of functions. From a mathematical point of view, there are infinite ways to expand a signal. What makes a particular representation important is that the characteristics of the signal are understood better in that representation.

Besides time, the second most important representation is frequency. The signal analysis based on frequency is called "frequency analysis". As a classic example of frequency analysis, the Fourier analysis has played an important role in stationary signal analysis and has been successful in many applications since it was proposed in 1807 [1]. Although the Fourier analysis is valid under extremely general conditions, there are some crucial restrictions of the Fourier spectral analysis: the system must be linear and the data must be strictly periodic or stationary, otherwise the resulting spectrum will make little physical sense. These restrictions suggest that some more strict conditions will be necessary to analyze a non-stationary signal.

Over the years, scientists have tried to find some available, adaptive and effective methods to process and analyze nonlinear and non-stationary data. Some methods have been found such as the spectrogram, the short-time Fourier transform, the Wigner-Ville distribution, the evolutionary spectrum, the wavelet transform, the empirical orthogonal function expansion and other miscellaneous methods [1], [2]. However, almost all of them depend on the Fourier analysis. A key point of these methods is that all of them try to modify the global representation of the Fourier analysis into a local one, which means that some intrinsic difficulties are inevitable. Hence, only a few of them perform really well unless in some special applications. Until now, wavelet analysis is still one of the best technologies for non-stationary signal analysis. It is often powerful, especially when the frequencies of a signal vary progressively. However, it can just be regarded as an extension of the Fourier analysis, because it also needs to expand a signal under a specified basis [2]. Once the selected basis does not match with the signal itself very well, the results are often unreliable.

The key point of developing adaptive and effective methods is the intrinsic and adaptive representations for the oscillatory modes of nonlinear and non-stationary signals. After considerable explorations, researchers have gradually realized that a complex signal should consist of some simple signals, each of which involves only one oscillatory mode at any time instance. These simple signals are called "mono-component signal" [1]. On the other hand, a superposition of some mono-component signals can form a complex signal. A real signal is often a complex one. Based on this model, Boashash has given a detailed discussion about the instantaneous frequencies of a signal and their corresponding time-frequency distributions [3]. However, up until now, it is still hard to accurately explain the significance of having only one oscillatory mode in any time location. Thus, there is no clear and accepted definition of how to judge whether or not a signal is a mono-component one.

Some researchers have suggested that the time-frequency distribution of a given signal should be defined first. Once the time-frequency distribution has been obtained, it will be easy to determine whether or not a signal is a mono-component one [4]. However, there are still almost insurmountable difficulties to find a logical time-frequency distribution.

A new mono-component signal model, which is called "Intrinsic Mode Function (IMF)", was proposed by Huang et. al in 1998 [5]. Meanwhile, a new algorithm entitled "Empirical Mode Decomposition (EMD)" [5] was developed to adaptively decompose a signal into a number of IMFs. With the Hilbert transform, the IMFs yield instantaneous frequencies as functions of time that give sharp identifications of imbedded structures. The final presentation is an energy-frequency-time distribution, designated as the Hilbert spectrum. Being different from the Fourier decomposition and the wavelet decomposition, EMD has no specified "basis". Its "basis" is adaptively produced depending on the signal itself, which makes not only decomposition efficiency very high but also makes localization of the Hilbert spectrum both on frequency and time much sharper and most important of all, makes much physical sense. Because of its excellence, EMD has been utilized and studied widely by researchers and experts in signal processing and other related fields [6], [7],

Zhihua Yang is with Information Science and technology School, Guangdong University of Business Studies, Guangzhou 510320, P. R. China.
Lihua Yang is with School of Mathematics and Computing Science, Sun Yat-sen University, Guangzhou 510275, P. R. China. Corresponding author. Email: mcsylh@mail.sysu.edu.cn, Tel: (8620)84115508, Fax: (8620)84111696.
This work was supported by NSFC (Nos.60873088,10631080).

[8], [9], [10]. Its applications have spread from earthquake research [11], to ocean science [12], fault diagnosis [13], signal denoising [14], image processing [15], [16], biomedical signal processing [17], speech signal analysis [18], pattern recognition [19] and so on.

Both conditions of the IMF have tried to restrict an IMF by involving only one oscillatory mode in any time location and by making the oscillations symmetric with respect to the time axis. The similar function of the two conditions has driven us to consider their relativity. After an acute analysis, we have proven that Condition 1 of the IMF can really be deduced from Condition 2. Finally, an improved definition of the IMF is given.

The rest of the paper is organized as follows: Section 2 contains the analysis of the definition of the intrinsic mode function. Section 3 plays a core role, in which some key results are proven and an improved definition of the intrinsic mode function is given. Finally, Section 4 contains the conclusion of this paper.

II. Analysis of the IMF Definition

The original objective of EMD was to identify the intrinsic oscillatory modes in each time location from a signal, one by one. With EMD, any complicated signal can be decomposed into a finite number of simple signals, each of which includes only one oscillatory mode in any time location. These extracted simple signals actually serve as approximations of so-called mono-component signals. However, it is difficult to tell what is an intrinsic oscillatory mode of a signal in a time location. This problem looks simple, but is really difficult. Intuitively, there are two ways to identify an intrinsic oscillatory mode: by the time lapse between the successive alternations of local maxima and minima such as $A \rightarrow B \rightarrow C$ as shown in figure 1; and by the time lapse between the successive zero crossings such as $D \rightarrow E \rightarrow F$ as shown in the same figure [23].

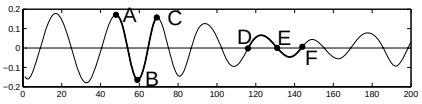

Fig. 1. A sketch map of intrinsic oscillatory mode.

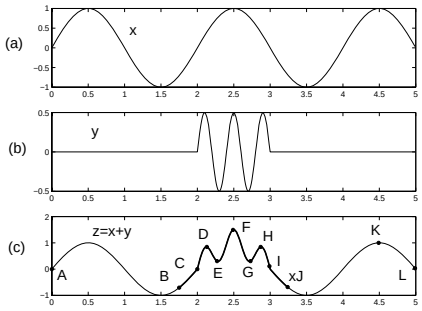

Fig. 2. A signal involving two oscillatory modes in some time locations.

In the literature [5], the first definition has been adopted, namely by the time lapse between the successive alternations of local maxima and minima, because it does not only give a much finer resolution of the oscillatory modes, but it can also be applied to signals with a non-zero mean, either all positive or all negative values, without zero crossings. However, an issue immediately arising is that an intrinsic oscillatory mode must not be a mono-component one according to this definition. For example, the signal z as shown in figure 2(c) is the superposition of the signal x and y as shown in figures 2(a) and 2(b), respectively. According to the definition, only one oscillatory mode is involved during each time lapse between $D \rightarrow F$ and $F \rightarrow H$. However, according to what we view intuitively, the real physical signal includes two oscillatory modes in some time locations, including these two time lapses. Therefore, the zero mean condition is considered to certainly get rid of this issue. Furthermore, the authors in [5] have presented the concept of IMF as follows:

Definition 2.1: An Intrinsic Mode Function (IMF) is a function that satisfies two conditions: (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; 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.

III. Improvement of the IMF Definition

The above definition refers to the concept of "envelope", however, there has never been any good and accepted definition of "envelope" up to this date. Intuitively, an upper/lower envelope of a signal should be a smooth line, which is always over/under and as near as possible to the signal itself. The literature [5] connects all the local maxima/minima of a signal by a cubic spline line to produce the upper/lower envelope. However, the possible results are the super-envelopes and/or under-envelopes. That is to say, the upper/lower envelope may be under/over the signal itself. An example of under-envelope is shown in figure 3, in which the dotted line is the cubic spline envelope of the signal plot on the solid line. In spite of some shortcomings, the envelope based on the cubic spline is still dominant in the EMD due to its simplicity and excellent performance for most signals.

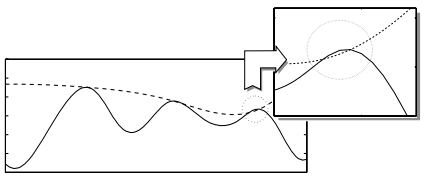

Fig. 3. An example of "under envelope" by the cubic spline interpolation.

We found that Condition 1 of IMF has been really involved in Condition 2 under the basic meaning of "envelope". Without a loss of generality, we consider only such kinds of signals which are continual and limitedly oscillatory in a limited interval (a, b). That is to say, the interval (a, b) can be divided into a finite number of small intervals. The signals are monotone on every small interval. We denote the space which is composed by all such signals as $S(a, b)$.

Definition 3.1: Let $f \in S(a,b)$, then $x_0 \in (a,b)$ is called an intrinsic maximum point of f, if it satisfies one of the two following conditions:

- Existing $\delta > 0$ to make $f(x_0) > f(x)(\forall x \in (x_0-\delta, x_0+\delta)\backslash x_0)$(as shown in figure 4, in which ξ is an intrinsic maximum point);
- Existing $\epsilon, \delta > 0$ to make:

$$\begin{cases} f(x) \text{ a constant during } (x_0-\delta, x_0+\delta); \\ f(x) \text{ a monotonic increasing function during} \\ \quad (x_0-\delta-\epsilon, x_0-\delta); \\ f(x) \text{ a monotonic decreasing function during} \\ \quad (x_0+\delta, x_0+\delta+\epsilon); \end{cases}$$

(such as η as shown in figure 4). Similarly, the intrinsic minimum point can be defined. The intrinsic maximum and minimum points are uniformly called the "intrinsic extreme points".

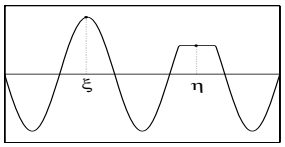

Fig. 4. Two kinds of the intrinsic maximum points.

The following theorem suggests that the intrinsic extreme points of a signal $f \in S(a,b)$ have to appear alternately:

Theorem 3.2: Let α, β be two intrinsic maximum/minimum points of $f \in S(a,b)$ satisfying $\alpha < \beta$ without a loss of generality, therefore there must be at least one intrinsic minimum/maximum point(s) in the interval (α, β); if there is no other intrinsic maximum/minimum point in the interval, then the number of intrinsic minimum/maximum points must be one.

Proof: Let ξ be a local minimum point of f in the interval $[\alpha, \beta]$. (a) If ξ is a strict local minimum point of f, then it is certainly an intrinsic minimum point; (b) If ξ is not a strict local minimum point of f, because f is decreased in a small enough left neighbor field of ξ and is increased in a small enough right neighbor field of ξ, it has to be non-strictly monotonic in at least one neighbor field. Let it be non-strictly monotonic in the right neighbor field without a loss of generality, therefore there must be some $\delta > 0$ which make f a constant in the intervals $[\xi, \xi+\delta]$. Let $[\alpha_1, \beta_1] \subset [\alpha, \beta]$ be the largest interval to ensure f is constantly equal to $f(\xi)$ and let $x_0 = \frac{1}{2}(\alpha_1 + \beta_1)$, therefore x_0 must be an intrinsic minimum point of f. Similarly, it can be proven that there must be some intrinsic minimum point(s) between two intrinsic maximum points.

If both ξ and η are intrinsic minimum points of f in the interval (α, β) and $\xi < \eta$, then there must be an intrinsic maximum point(s) between ξ and η. That is paradoxical with the condition of no other intrinsic maximum points in the interval (α, β), therefore the number of intrinsic minimum points between (α, β) has to be one. ■

An envelope should be a new signal as defined by the original signal, so the relation of a signal and its envelope can be viewed as an operator. Envelopes can be classified into upper and lower ones. Let T_u be an upper envelope operator, denoted by T for simplicity, then the upper envelope of the signal f can be written as Tf, thereby the lower envelope operator can be defined by:

$$T_l f := -T_u(-f) = -T(-f).$$

Without a loss of generality, let T satisfy the conditions below:

(i) $Tf \geq -T(-f)$, namely an upper envelope is always above a lower envelope;

(ii) $\forall f \in S(a,b)$, let ξ be an intrinsic minimum point of f, where we have $T_u f(\xi) - f(\xi) > f(\xi) - T_l f(\xi)$, namely:

$$Tf(\xi) - f(\xi) > f(\xi) + T(-f)(\xi).$$

That is to say that the difference between the upper envelope and the minimum is larger than that between the minimum and the lower envelope as shown in figure 5.

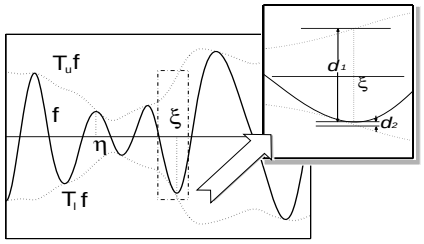

Fig. 5. The intuitive meaning of condition $T_u f(\xi) - f(\xi) > f(\xi) - T_l f(\xi)$. In the figure, $d_1 = T_u f(\xi) - f(\xi)$, $d_2 = f(\xi) - T_l f(\xi)$.

If η is any intrinsic maximum point of f, it is certainly an intrinsic minimum point of $-f$, therefore we have:

$$T(-f)(\eta) + f(\eta) > -f(\eta) + T(f)(\eta),$$

where, namely:

$$Tf(\eta) - f(\eta) < f(\eta) + T(-f)(\eta).$$

So for any intrinsic minimum point ξ and intrinsic maximum point η, we have:

$$f(\xi) < \frac{1}{2}[Tf(\xi) - T(-f)(\xi)],$$

$$f(\eta) > \frac{1}{2}[Tf(\eta) - T(-f)(\eta)].$$

Therefore, we can instantly reach a conclusion as follows:

Theorem 3.3: Let $f \in S(a,b)$ satisfy $T_u f(t) + T_l f(t) = 0\ \forall t \in (a,b)$, therefore for any intrinsic minimum point of f, ξ and intrinsic maximum point η, we have $f(\xi) < 0, f(\eta) > 0$. Therefore, the intrinsic extreme points and the crossing zero points will have to appear alternately as follows:

$$\cdots \rightarrow \text{MaxP} \rightarrow \text{ZP} \rightarrow \text{MinP} \rightarrow \text{ZP} \rightarrow \text{MaxP} \rightarrow \cdots$$

MaxP denotes the intrinsic maximum point, ZP denotes the crossing zero point and MinP denotes the intrinsic minimum point in the above expression.

Theorem 3.3 indicates that for signal f, if only the mean value of the upper and lower envelopes is equal to zero, then

the extreme points and crossing zero points will have to appear alternately. That means Condition 1 in Definition 2.1 can be deduced from Condition 2. Therefore, a more refined definition is given as follows:

Definition 3.4: An Intrinsic Mode Function (IMF) is a function that satisfies the condition that at any time instant, the mean value of the upper envelope as defined by the local maxima and the lower envelope as defined by the local minima is zero.

IV. Conclusion

This paper makes an acute analysis of the definition of the intrinsic mode function and proves that Condition 1 of the intrinsic mode function can really be deduced from Condition 2. Finally, an improved definition of the intrinsic mode function is given.

References

[1] L. Cohen, *Time-frequency analysis: theory and applications* Prentice-Hall, Inc., Upper saddle River, NJ, 1995.
[2] S. Mallat. *Wavelet tour of signal processing.* Academic Press, San Diego, USA, 1999.
[3] B. Boashash. *Estimating and interpreting the instantaneous frequency of a signal: Part I Fundamentals.* Proc. IEEE 80, 417-430, 1992.
[4] L. Cohen. *Time-Frequency distributions–A review.* Proc. IEEE, 77:941-981, 1989.
[5] 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, A(454):903–995, 1998.
[6] P. Flandrin, G. Rilling, and P. Goncalves. *Empirical mode decomposition as a filter bank.* IEEE Signal Processing Letters, 11(2): 112–114, 2004.
[7] G. Rilling, and P. Flandrin. *One or two frequencies? The empirical mode decomposition answers.* IEEE Trans. Signal Processing, 56(1): 85–95, 2007.
[8] S. Meignen, and V. Perrier. *A new formulation for empirical mode decomposition based on constrained optimization.* IEEE Signal Processing Letters, 14(12): 932–935, 2007.
[9] Z. H. Yang, D. X. Qi, and L. H. Yang. *Signal period analysis based on Hilbert-Huang transform and its application to texture analysis.* Proceedings of Third International Conference on Image and Graphics, Hong Kong, China, 430-433, 2004.
[10] A. Boudraa, and J. Cexus. *EMD-Based signal filtering.* IEEE Trans. Instrumentation and Measurement, 56(6): 2196–2202, 2007.
[11] C. H. Loh, T. C. Wu, and N. E. Huang. *Application of emd+hht method to identify near-fault ground motion characteristics and structural responses.* BSSA, Special Issue of Chi-Chi Earthquake, 91(5):1339–1357, 2001.
[12] N. E. Huang, Z. Shen, and S. R. Long. *A new view of nonlinear water waves: the Hilbert spectrum.* Annu. Rev. Fluid Mech, 31:417-57, 1999.
[13] B. Liu, S. Riemenschneider, and Y. Xu. *Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum.* Mechanical Systems and Signal Processing, 20(3): 718-734, 2006.
[14] D. F. Chen, and X. L. Wu. *Recovery of signal from transient scattered response contaminated by Gaussian white noise based on EMD method.* Chinese Journal of Electronics, 32(3): 496-498, 2004.
[15] N. Bi, Q. Y. Sun, D. R. Huang, Z. H. Yang, and J. W. Huang. *Robust image watermarking based on multiband wavelets and empirical mode decomposition.* IEEE Trans. Image Processing, 16(8): 1956–1966, 2007.
[16] Z. X. Liu, and S. L. Peng. *Directional EMD and its application to texture segmentation.* Science in China(F), 35(2):113-123, 2005.
[17] Z. H. Yang, L. H. Yang, and D. X. Qi. *Detection of spindles in sleep EEGs using a novel algorithm based on the Hilbert-Huang transform.* Applied and Numerical Harmonic Analysis, 543-559, 2006.
[18] M. K. Molla, and K. Hirose. *Single-Mixture audio source separation by subspace decomposition of Hilbert spectrum.* IEEE Trans. Audio, Speech, and Language Processing, 15(3): 893–900, 2007.
[19] Z. H. Yang, D. X. Qi, and L. H. Yang. *Chinese font recognition based on EMD.* Pattern Recognition Letters, 27(14):1692-1701, 2006.
[20] Y. J. Deng, W. Wang, C. C. Qian, and D. J. Dai. *Boundary-processing-technique in EMD method and Hilbert transform.* Chinese Science Bulletin, 46(3):954–961, 2001.
[21] E. Delechelle, J. Lemoine, and O. Niang. *Empirical mode decomposition: An analytical approach for sifting process.* IEEE Signal Processing Letters, 12: 764–767, 2005.
[22] Q. H. Chen, N. E. Huang, D. Riemenschneider, and Y. S. Xu. *A B-spline approach for empirical mode decomposition.* Advances in Computational Mathematics, 24: 171–195, 2006.
[23] P. G. Drazin. *Nonlinear systems.* Cambridge University Press, Cambridge, 1992.