Mechanical Systems and Signal Processing 40 (2013) 136–153

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis

Jinde Zheng [a,b,*], Junsheng Cheng [a,b], Yu Yang [a,b]

[a] State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha 410082, PR China
[b] College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, PR China

ARTICLE INFO

Article history:
Received 24 November 2012
Received in revised form
26 March 2013
Accepted 15 April 2013
Available online 12 May 2013

Keywords:
Generalized empirical mode decomposition
EMD
Non-stationary signal
Intrinsic time-scale decomposition
Fault diagnosis
Empirical envelope demodulation

ABSTRACT

As an adaptive time-frequency-energy representation analysis method, empirical mode decomposition (EMD) has the attractive feature of robustness in the presence of nonlinear and non-stationary data. It is evident that an appropriate definition of baseline (or called mean curve) of data plays a crucial role in EMD scheme. By defining several baselines, an adaptive data-driven analysis approach called generalized empirical mode decomposition (GEMD) is proposed in this paper. In the GEMD method, different baselines are firstly defined and separately subtracted from the original data, and then different pre-generated intrinsic mode functions (pre-GIMFs) are obtained. The GIMF component is defined as the optimal pre-GIMF among the obtained ones with the smallest rate of frequency bandwidth to center frequency. Next, the GIMF is subtracted from the original data and a residue is obtained, which is further regarded as the original data to repeat the sifting process until a constant or monotonic residue is derived. Since the GIMF in each frequency-band is the best among different pre-GIMFs derived from EMD and other EMD like methods, the GEMD results are best as well. Besides, a demodulating method called empirical envelope demodulation (EED) is introduced and employed to analyze the GIMFs in time-frequency domain. Furthermore, GEMD and EED are contrasted with the original Hilbert–Huang Transform (HHT) by analyzing simulation and rolling bearing vibration signals. The analysis results indicate that the proposed method consisting of GEMD and EED is superior to the original HHT at least in restraining the boundary effect, gaining a better frequency resolution and more accurate components and time frequency distribution.

1. Introduction

Hilbert–Huang Transform (HHT), which was introduced by Dr. Norden Huang in the mid-1990s, has been proved an effective and powerful approach for adaptive local time-frequency analysis [1,2]. The HHT mainly is composed of the empirical mode decomposition (EMD) and Hilbert spectral analysis (HSA) [3,4]. Unlike other traditional time-frequency analysis methods, such as the Fourier Transform and various wavelet decomposition methods, EMD method does not use a priori determined basis functions. Originally in the pioneer literature of EMD, cubic splines are used to define the upper and lower envelopes (respectively, fitted through the local maxima and minima of the time series) in each sift. And the baseline

* Corresponding author at: State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha 410082, PR China. Tel.: +86 731 88664008; fax: +86 731 88711911.
E-mail addresses: lqdlzheng@126.com, jdzheng1986@gmail.com (J. Zheng).

0888-3270/$ - see front matter
http://dx.doi.org/10.1016/j.ymssp.2013.04.005

is formed of the average of the two splines, which then was subtracted from the time series being sifted. Undoubtedly, the mean curve plays a crucial role in EMD, on which many improvements have been made in recent years. Typically, instead of fitting splines separately to the upper and lower local extrema, Chen et al. [5] combined the extrema and took the moving average of a linear combination of cubic B-splines through the local extrema and the moving average is considered as the baseline, then the EMD algorithm continues as usual. In 2008, Pegram et al. [6,7] proposed another modification to the EMD algorithm, whereby rational splines replaced cubic splines in the extrema envelope-fitting procedure. In 2007, Frei and Osorio [8] proposed a substitute for EMD called intrinsic time-scale decomposition (ITD) that constructs the baseline by a linear transform of the original data itself directly and then the baseline is subtracted from the signal to obtain a residual signal and next steps are similar to EMD. In this way, the over- and under-shooting in the upper and lower envelopes using cubic splines can be removed from the decomposition. Besides, there are also other literatures focusing on modification of the baseline that have been proved valid [9]. Hence, it is significant to define a reasonable baseline for the EMD algorithm in order to get a better decomposition results. Disappointingly, however, since different approaches are suited to deal with different signals that have different characteristics, such as AM, FM, or AM–FM signals, rather than all signals, a modification on the baseline of the EMD does not always mean an improvement on the original EMD method, namely, the modified approach, taking ITD for an example, may get a better analysis result for some special data, while for other data it may not.

Inspired by EMD, in this paper a comprehensive time-frequency–energy analysis method called generalized empirical mode function (GEMD) is proposed by integrating the advantages of improvements on several typical EMD-based approaches. Each GIMF component of GEMD in each rank is selected from different pre-GIMFs, which are derived from different methods with difference in construction of baseline. Since the GIMF in each rank is the best under a determined optimization criterion, the GEMD result will be superior to EMD and its modified methods.

Besides, after having derived the GIMFs, a demodulation method is needed to get the local time-frequency information of the initial data. The common demodulating methods include Hilbert transform (HT), Energy operator demodulation, iterated Hilbert transform and generalized demodulation, etc. HT has been most used for demodulation in recent years, especially after the HHT was proposed. However, HT is limited by Bedrosian and Nuttal theorems [10,11] and has a serious boundary effect, and negative frequency will arise if the theorems cannot be met. Iterated Hilbert transform (IHT) is a new method for analyzing multi-component AM–FM signals proposed by Gianfelici et al. [12]. Compared with HT, IHT has higher demodulation accuracy and lower computational complexity; however, there are limitations in the direct estimation of instantaneous frequencies via the phase signals of the previously obtained model [13]. Although energy operator demodulation [14,15] can obtain more reliable and more precise values than HT, it also has the drawbacks of being of a less range of application and only suitable for the data with a slow varying instantaneous amplitude and frequency. Generalized demodulation as a new demodulation method also has the limitation of selecting a suitable phase function [16].

In this paper another demodulating method named empirical envelope demodulation (EED) proposed by Cheng in [17] is employed for obtaining the detailed time-frequency–energy distribution of a given signal; further, the Hilbert–Huang spectrum and marginal spectrum of the original data can be acquired as well.

The rest of this paper is organized as follows. Section 2 reviews the EMD algorithm briefly, and then several methods for defining the baseline are also introduced. In addition, the optimal GIMF selection criterion is given as well; in Section 3 the GEMD method is proposed firstly and then the EED method is introduced briefly; in Section 4 comparison between the proposed method consisting of GEMD and EMD and the original HHT is made by analyzing simulation signals; in Section 5 the proposed method is applied to analyze the experimental data of rolling element bearing with fault; and the conclusions and discussions are given in the final section.

2. Review of EMD and definitions of baseline

2.1. Review of EMD

As mentioned above, EMD is an adaptive approach to remove oscillations (IMFs) successively through repeated subtraction of the baselines. For a given signal $x(t)$ $(t>0)$, the EMD algorithm comprised of the following steps [1–3,6,7]:

(1) Set $r_0(t)=x(t)$ and set $i=1$.
(2) Identify all the extrema in $r_{i-1}(t)$ and connect the sequential local maxima (respective minima) using cubic spline to derive the upper (respective lower) envelop $\max_{i-1}(t)$ (and $\min_{i-1}(t)$ correspondingly).
(3) Derive the baseline, $m_{i-1}(t)$, by averaging the upper and lower envelopes, namely

$$m_{i-1}(t)=\frac{[\max_{i-1}(t)+\min_{i-1}(t)]}{2} \tag{1}$$

(4) Extract the temporary local oscillation $IMF_i(t)=r_{i-1}(t)-m_{i-1}(t)$.
(5) If the mean of $IMF_i(t)$ is not zero, repeat steps (1)–(3) on the temporary local oscillation by setting $r_{i-1}(t)=IMF_i(t)$ and iterating until the mean of $IMF_i(t)$ is zero. Then, $IMF_i(t)$ is treated as an IMF, noted as $I_i(t)$.
(6) Compute the residue: $r_i(t)=r_{i-1}(t)-I_i(t)$.