Mechanical Systems and Signal Processing 114 (2019) 189–211

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Symplectic geometry mode decomposition and its application to rotating machinery compound fault diagnosis

Haiyang Pan [a,b], Yu Yang [a,*], Xin Li [a], Jinde Zheng [b], Junsheng Cheng [a]

[a] State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha 410082, PR China
[b] School of Mechanical Engineering, Anhui University of Technology, Ma'anshan 243032, PR China

ARTICLE INFO

Article history:
Received 21 July 2017
Received in revised form 23 March 2018
Accepted 10 May 2018

Keywords:
Symplectic geometry mode decomposition
Nonlinear system signal
Rotating machinery compound fault
Fault diagnosis

ABSTRACT

Various existed time-series decomposition methods, including wavelet transform, ensemble empirical mode decomposition (EEMD), local characteristic-scale decomposition (LCD), singular spectrum analysis (SSA), etc., have some defects for nonlinear system signal analysis. When the signal is more complex, especially noisy signal, the component signal is forced to decompose into several incomplete components by LCD and SSA. In addition, the wavelet transform and EEMD need user-defined parameters, and they are very sensitive to the parameters. Therefore, a new signal decomposition algorithm, symplectic geometry mode decomposition (SGMD), is proposed in this paper to decompose a time series into a set of independent mode components. SGMD uses the symplectic geometry similarity transformation to solve the eigenvalues of the Hamiltonian matrix and reconstruct the single component signals with its corresponding eigenvectors. Meanwhile, SGMD can efficiently reconstruct the existed modes and remove the noise without any user-defined parameters. The essence of this method is that signal decomposition is converted into symplectic geometry transformation problem, and the signal is decomposed into a set of symplectic geometry components (SGCs). The analysis results of simulation signals and experimental signals indicate that the proposed time-series decomposition approach can decompose the analyzed signals accurately and effectively.

1. Introduction

In recent years, time series analysis has played an important role in engineering application. A time series can be decomposed into a series of simple component series, and by the judgment, reconstruction and identification of the useful component series, the most essential information regarding the evolution behaviours of an originating system can be obtained [1]. In fact, the measured signal as a typical time series data usually consists of different modes which have complex time domain waveforms and time-varying amplitude and frequency. Meanwhile, the oscillatory signals contain a large amount of valuable system state information about the originating system, therefore, system signals need further study. However, in order to obtain all the characteristics of a particular component, it is necessary to separate the component from the signal. For a reliable signal decomposition method, it can efficiently extract the existed modes, separate them from each other and remove the noise signal. Thus, how best to achieve above objectives is a long-standing problem, and many different methods have been proposed [2]. Unfortunately, almost all of the existed approaches have at least one of the following flaws:

* Corresponding author.
E-mail address: yangyu@hnu.edu.cn (Y. Yang).

https://doi.org/10.1016/j.ymssp.2018.05.019
0888-3270/

(1) The method needs user-defined parameters, and it is very sensitive to the parameters. Besides, the parameters cannot be chosen adaptively.
(2) The method cannot decompose the signal effectively with noise in the signal, i.e. it does not have noise robustness.
(3) If there is a complex (non-sinusoidal) waveform, the method will decompose it into several inaccurate components.

Wavelet transform can decompose the time series into different frequency bandwidth or scale components of various resolutions [3,4], and provide time-domain and frequency-domain localization information of time series at the same time, so wavelet transform has the characteristics of multi-scale and "mathematical microscopic", which makes the wavelet analysis identify the mutation components in the time series [5,6]. In recent years, wavelet analysis is more and more widely used in mechanical signal processing [7]. However, the wavelet analysis is an adjustable window Fourier transform essentially. Due to the limited length of the wavelet basis function, wavelet transform will produce energy leakage, so it is difficult to analyze the signal in the time-domain and frequency-domain [8]. On the other hand, once the wavelet basis and the decomposition scale are chosen, the result is a fixed frequency band signal which is only related to the sampling frequency instead of the signal itself. Thus, the limited length of wavelet transform can cause unwanted leakage and decompose a complete component into several inaccurate components (drawback 1 and 3) [9].

Empirical mode decomposition (EMD) method can decompose an amplitude modulation and frequency modulation signal into a series of intrinsic mode functions (IMFs) adaptively, and the frequency component of each intrinsic mode function is related to the sampling frequency as well as the self-changing signal itself [10,11]. In addition, the method is used to decompose the signal without user-defined parameters and based on the time scale characteristics of the data itself. This is different from the wavelet transform method based on the wavelet basis function essentially. Because of above characteristics, in theory, EMD method can be applied to decompose any type of signal. In addition, EMD method has obvious advantages in dealing with nonlinear and no-stationary signal, and is very suitable for the analysis of nonlinear and no-stationary signal [12]. However, Cubic splines of EMD are employed to fit the lower and upper envelopes in each sift, which can lead to owe envelope, over-envelope and mode mixing (drawback 2 and 3) [13].

EMD is sensitive to noise, thus Wu and Huang have proposed an improved method of EMD called ensemble empirical mode decomposition (EEMD) to overcome above shortcomings of EMD [14]. The idea of EEMD is to get the final modes by adding individual subjective independent realizations of the Gaussian white noise signal to the original signal and applying EMD repeatedly. The EEMD decomposition principle is as follows: when the additional Gaussian white noise is evenly distributed over the entire time-frequency space, the time-frequency space is composed of different scale components divided by the filters. When the signal is coupled with a uniformly distributed Gaussian white noise, the signal areas of different scales are automatically mapped to the appropriate scale associated with the Gaussian white noise. Of course, each independent test can produce strong noisy result, and it is because that each additional noise component contains the signal and additional Gaussian white noise. Since the noise is different in each individual test, the noise will be eliminated by the ensemble-averages of sufficient tests. Therefore, EEMD can weaken the effect of noise on the decomposition of the original signal by adding white noise [15]. Even so, EEMD does not have good noise-robustness for strong noisy signal. In addition, it contains two non-adaptive parameters (drawback 1) which control the standard deviation of the additional noise and have a certain effect on the results [16,17].

Aiming at the shortages of above methods, Frei et al. have proposed a new adaptive analysis method called intrinsic time-scale decomposition (ITD) [18], which can decompose the non-stationary signal into a sum of several proper rotation components (PRCs) that are independent of one another adaptively. Compared with the EMD and EEMD, ITD has obvious advantages in end effect and computational efficiency [19]. In addition, the problems of end effect, over envelope, mode mixing, and IMF criterion in EMD are overcome [20]. The ITD method has been applied to the field of mechanical fault diagnosis in view of the fact that the ITD method is able to analyze non-stationary and nonlinear signal [21]. But this method fails to elaborate the physical significance of the algorithm itself and the PRC. Moreover, the baseline signal is obtained by using linear transformation to vibration signal, which results in distortion in the instantaneous amplitude and frequency of PRC [22].

Due to some problems existing in the ITD method, its further application is restricted and some improved ITD methods have been proposed. The envelope signals are got by using cubic spline instead of linear transformation and the local characteristic-scale decomposition (LCD) is put forward to obtain intrinsic scale components (ISCs) in the literature [19]. The cubic B-spline interpolation is used to obtain smooth single component envelope signal in the literature [23]. The problems of burr and distortion in the instantaneous amplitude and frequency are solved in the two improved ITD methods. Cubic spline interpolation method has the characteristics of good convergence and higher smoothness, but mode mixing would be produced in the cubic spline interpolation, especially in the strong noisy signals (drawback 2 and 3).

Inspired by the idea of the literature [23], the Hermite interpolation would be applied to the ITD method. The cubic Hermite interpolation with a characteristic of shape preservation is most widely used, and it is especially suitable for the analysis of signal that has strong non-stationary characteristics [24]. But the cubic Hermite interpolation is the deterministic interpolation and the shape of curve cannot be changed when condition is determined. In order to control the shape of curve surface flexibility and improve the approximation effect in determined condition, the rational Hermite interpolation is put forward and applied to the EMD method in the literature [25]. The rational Hermite interpolation can control the shape of the interpolation curve by the parameter, which can further improve the fitting accuracy of approximation and decompose the signal accurately. Many types of rational Hermite basic function have been presented, but the rational Hermite basic function with parameters is the best in the various types of rational Hermite basic function, which not only has simple structure but