Causes and solutions of overshoot and undershoot and end swing in Hilbert-Huang transform*

LONG Si-sheng (龙思胜) ZHANG Tie-bao (张铁宝) LONG Feng (龙 锋)
Earthquake Administration of Sichuan Province, Chengdu 610041, China

Abstract
There are overshoot and undershoot phenomenon and end swing phenomenon in the cubic spline fitting in Hilbert-Huang transform. The two problems influence data quality of the empirical mode decomposition seriously. The cubic spline fitting has been analysed, and the causes of producing the overshoot and undershoot phenomenon and the end swing phenomenon have been pointed out in this paper. Two new methods of cubic spline fitting and sine spline fitting and the new technique of handling the end points of the original data curve can completely remove the overshoot and undershoot phenomenon and the end swing phenomenon on the condition of unchanging original data, and have the advantages of the continuous fitting functions and its continuous one order derivative, the simple and convenient calculations, the small calculation amount and the easy work on it.

Key words: Hilbert-Huang transform; cubic spline fitting; sine spline fitting; overshoot and undershoot; end swings
CLC number: 315.01 Document code: A

Introduction
Hilbert-Huang transform (HHT) is a great break in processing nonlinear and non-stationary data and can be successfully used in many science domains. There are mainly two parts in this method. The first part is to decompose the original data into several intrinsic mode functions (IMF) with the empirical mode decomposition (EMD). IMF components are derived from the original data directly according to the local characteristics in the data under some rules, so that IMF are posteriori, adaptive, complete and almost orthogonal and can be used in the research of nonlinear and non-stationary data. Under the decomposition conditions, these IMFs are the intrinsic and imbedded oscillatory modes in the data and the wave trains approaching to harmonic wave with the intrawave amplitude and frequency modulations, for this reason, which have fine Hilbert transform property and the physical information of the data. The second part is to construct the energy-frequency-time distributions (Hilbert spectrum) with the Hilbert transform to IMFs, which can reveal the time and frequency position of each event for the detailed research of the process. (Huang et al, 1998, 1999, 2001).

To achieve empirical mode decomposition in HHT, the cubic spline fitting are used in con-
structing upper and lower envelopes. This fitting is one of the most applied fitting methods. The cubic spline fitting is classified into three types: intrinsic boundary, natural boundary and periodical boundary (GUAN and CHEN, 2002), and there is also the viewpoint on five types of this methods (LI and QI, 1979). But there are the overshoot and undershoot and the end swing problems in the cubic spline fitting method used by Huang et al, which may influence the results of EMD seriously (Huang et al, 1998). None have researched the overshoot and undershoot problem up to now. Some researchers of China have discussed the end swing problem, but they have not explained the cause of the end swing phenomenon, and the solution methods pointed by them are not ideal for the end swing problem because these methods will change the original data before handling the original data (SHI and LUO, 2003; LUO and SHI, 2003; GAI and ZHANG, 2002).

The causation of the overshoot and undershoot phenomenon has been analysed and one new convenient method of cubic spline fitting with the advantages of a small amount of calculation has been given out which can clean up the overshoot and undershoot phenomenon in this paper. And the causation of the end swing phenomenon has been analysed and a new technique has been also given out which can remove the end swing without any change in the original data. The combination of the method and the technique in the empirical mode decomposition will greatly improve the data processing quality and establish a reliable basis for EMD.

1 Cubic spline fitting method

1.1 Mathematical meaning of the overshoot and undershoot

No strict definition of the overshoot and undershoot has been given in Huang's papers. According to Huang's meaning about the overshoot and undershoot and our realization in the data processing practice, we think the overshoot and undershoot phenomenon is that some of the fitted numbers are bigger than the bigger one or smaller than the smaller one of the couple of the two adjacent data points in the closed subinterval which is closed by the two data points, i.e. some of the fitted numbers go up or down beyond the range of the two adjacent data points in the vertical directions. In order to remove the overshoot and undershoot phenomenon from the fitted curve, it is the requirement that all fitted numbers are not bigger than the bigger one or smaller than the smaller one of the couple of the two adjacent data points in the closed subinterval which is closed by the two data points.

1.2 The requirement of EMD for fitting function

According to Huang's meaning about EMD and our realization in the data processing practice, it is the aim of EMD to decompose the data curve into several curves of the different components with the narrower frequency subintervals which are not overlap each other. Each of these components is "purer" enough to be handling easily and to include the partial physical meaning in the original data set. And the data set can be directly obtained from all the components by adding all the components each other.

The major local extreme value points are the ones that represent the major characters of the data curve. The important tools for EMD are the upper and lower envelopes and their average value curve, which are fitted to these major local extreme points. Thus, the disturbance of the overshoot and undershoot and the end swing must be removed from the fitted curves if we do not want to reduce the handled data quality after fitting. On the other way, the requirement of the degree of smoothing of the fitted curves can be reduced lightly. Therefore it is the requirement of the fitting function that the fitting function and its first-order derived function should be continuous at every nodal point, the first-order derived function should equal to zero at every nodal point, noth-