Available online at www.sciencedirect.com
ScienceDirect
Mechanical Systems and Signal Processing 22 (2008) 509–519
Mechanical Systems and Signal Processing
www.elsevier.com/locate/jnlabr/ymssp

Theoretical analysis and comparison of the Hilbert transform decomposition methods

Michael Feldman*

Faculty of Mechanical Engineering, Israel Institute of Technology, Technion, Haifa 32000, Israel

Received 3 May 2007; received in revised form 28 August 2007; accepted 5 September 2007
Available online 5 October 2007

Abstract

This article considers the empirical mode decomposition pioneered by Huang together with the Hilbert vibration decomposition method. Both methods are intended for extracting simple components using the varying instantaneous frequency and amplitude from multicomponent non-stationary signals. The common properties of and the differences between the two methods are taken into account.

Keywords: Hilbert transform; Signal decomposition; Envelope; Instantaneous frequency

1. Introduction

Recently, Huang et al. [1] proposed the empirical mode decomposition (EMD) method to extract monocomponent and symmetric components, known as intrinsic mode functions (IMF), from nonlinear and non-stationary signals. The term "empirical" chosen by the authors emphasizes the empirical essence of the proposed identification of the IMF by their characteristic time scales in the initial complicated data. During the last decade, serious mathematic works [2–5] have been dedicated to detailed analyses of the local EMD method. However, a simple but important theoretical question remains: why is spline fitting of local extrema able to generate the simplest components?

Some years later, a different technique, called the Hilbert vibration decomposition (HVD) method, dedicated to the same problem of decomposition of non-stationary wideband vibration, was developed in [6]. The global HVD method is based on the Hilbert transform (HT) presentation of the instantaneous frequency (IF) and does not involve spline fitting.

In the present paper we will attempt to analyze and compare the above-mentioned methods by investigating and understanding their general principles and limitation, without discussion of the corresponding concrete signal processing procedures and algorithms.

*Tel.: +972 4 8292106; fax: +972 4 8228931.
E-mail address: MFeldman@technion.ac.il

0888-3270/$ - see front matter
doi:10.1016/j.ymssp.2007.09.013

2. Signal envelope and instantaneous frequency

The interpretation of an envelope, defined as the absolute value (modulus or magnitude) of a complex (typically analytic) representation of a signal, has been the subject of investigation for years [7]. When dealing with general modulated signals, it is often convenient to define the analytic signal $X(t) = x(t) + j\tilde{x}(t)$, where $\tilde{x}(t)$ is related to $x(t)$ by the HT. According to analytic signal theory, a real vibration process $x(t)$ measured by, for example, a transducer, is only one of the possible projections (the real part) of some analytic signal $X(t)$. Then, the second or quadrature projection of the same signal (the imaginary part $\tilde{x}(t)$) will be conjugated according to the HT. The analytic signal is represented geometrically in the form of a phasor rotating in a complex plane. Using the traditional representation of the analytic signal in its trigonometric or exponential form, $X(t) = |X(t)|[\cos\varphi(t) + i\sin\varphi(t)] = A(t)e^{i\varphi(t)}$, one can determine its instantaneous amplitude (envelope, magnitude)

$$A(t) = |X(t)| = \sqrt{x^2(t) + \tilde{x}^2(t)} = e^{\mathrm{Re}[\ln X(t)]}, \tag{1}$$

and its instantaneous phase $\psi(t) = \arctan\big(\tilde{x}(t)/x(t)\big) = \mathrm{Im}[\ln X(t)]$, where $\tilde{x}(t)$ is the HT of $x(t)$ that can be written as the convolution integral of $x(t)$ with $1/\pi t$ as $\tilde{x}(t) = x(t) * 1/\pi t$. The initial signal and its envelope have common tangents at points of contact, but the signal never crosses the envelope. The plus sign of the root square Eq. (1) corresponds to the upper envelope and the minus sign corresponds to the opposite sign lower envelope, so they are always in antiphase relation. The envelope function contains important information about the energy of the signal.

The instantaneous phase is a function of time; therefore, the first derivative of the instantaneous phase is also a single value function of time [7,8]

$$\omega(t) = \dot{\psi}(t) = \frac{x(t)\dot{\tilde{x}}(t) - \dot{x}(t)\tilde{x}(t)}{A^2(t)} = \mathrm{Im}\left[\frac{\dot{X}(t)}{X(t)}\right]. \tag{2}$$

It is called the instantaneous angular frequency, and it plays the most important role in an analytic signal. The IF $\omega(t)$ measures the rate of rotation in the complex plane. Naturally, for a simple monoharmonic signal, the envelope and the IF are constant, and the phase angle increases linearly with time. In the general case, the IF of the signal is a varying function of time. Moreover, the IF in some cases may change sign in some time intervals. This corresponds to the change of rotation of the phasor from the counterclockwise to the clockwise direction. The IF always has physical meaning and is nothing more than just the varying speed (rate) of the phasor rotation in polar axes. In other words, whether it is positive or negative is always meaningful, like the positive or negative instantaneous speed of a particle under a Brownian type motion along a real line. The IF sign switch corresponds to a stopping and reverse rotation of the phasor. A signed value of the IF indicates both the rate and the direction of rotation. In the time domain, the negative IF corresponds to the appearance of a complex riding cycle (complicated cycle of alternating signal).

2.1. Average values and frequency bandwidth

The instantaneous amplitude and frequency of complicated vibration signals are nonconstant; they vary in time. While the IF is a positive function, the signal itself has the same numbers of zero crossings and extrema. When the IF has a negative value, the signal has one or multiple extrema between successive zero crossing. The mean value of the envelope takes the form $\bar{A} = \int_{-\infty}^{\infty} A(t)\,dt$. The mean value of time derivative squared of the envelope $\overline{\dot{A}^2(t)} = \int_{-\infty}^{\infty} \dot{A}^2(t)\,dt$ determines the level of the envelope variation ($\bar{A} = 1$). The mean value of the IF $\bar{\omega} = \int_{-\infty}^{\infty} \omega(t)A^2(t)\,dt = (m_1/m_0)$, equal to the first normalized moment of the signal spectrum, is called the central frequency (here m_i is the ith moment of the spectrum) [7]. The mean value of the modulus of the IF given by $\overline{|\omega|} = \int_{-\infty}^{\infty} |\omega(t)|A^2(t)\,dt = (m_2/m_0)^{1/2}$ is equal to the number of the signal zero crossing.

The mean value of the IF squared $\overline{\omega^2} = \int_{-\infty}^{\infty} \omega^2(t)A^2(t)\,dt = (m_2/m_0) - \overline{\dot{A}^2(t)}$ determines the level of the IF variation. By summing up the IF variation around the mean value and the envelope variations, we obtained