一种改善 EMD端点效应的新方法

王 婷[1], 杨莘元[1], 李冰冰[2]
(1. 哈尔滨工程大学 信息与通信工程学院,黑龙江 哈尔滨 150001;
2. 承德石油高等专科学校学校办公室,河北 承德 067000)

摘 要:针对经验模态分解 (Empirical Mode Decomposition, EMD)过程中可能产生的端点效应问题,提出一种基于最近相似距离的端点延拓方法. 首先给出了匹配距离和波形相似系数的概念,然后按照一定配准和筛选准则,在原信号内部找到与信号端点处变化趋势最为相似的子波来对端点处的信号进行延拓. 该方法兼顾了信号的内在特性以及端点处的变化趋势,实现延拓数据与原信号交界处的光滑过渡. 将其分别应用于模拟信号中分析,仿真结果表明该方法可有效地抑制 EMD端点效应问题.

关键词:经验模态分解; 端点效应; 端点延拓; 波形相似
中图分类号: TN911. 7 **文献标志码**: A **文章编号**: 1007- 2683 (2009) 05- 0023- 04

A New Method for End Effect of EMD

WANG Ting[1], *YANG Shen-yuan*[1], *LI Bing-bing*[2]
(1. School of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China;
2. School of Administrative Office, Chengde Petroleum College, Chengde 067000, China)

Abstract: A new method of data extending based on minimum similarity distance is proposed in order to solve the problem of end effect in empirical mode decomposition (EMD). Firstly, the conception of waveform similarity coefficient and matching distance are given, and then based on the related matching rules, the extending waveform are chosen from the inner signal, which has the most minimum similarity distance compared with the ends. The method considers both the inner characteristics of signal and variation tendency of endpoint, and realizes the smooth transition between extending data and original signal. This new method is applied to analog signal analysis, and simulation results show that it is effective to restrain the end effects of EMD method.

Key words: EMD; end effect; end extending; waveform similarity

0 引 言

HHT(Hilbert-Huang Transform)方法是一种新的分析非线性、非平稳信号的时频方法,它包含两个过程:首先对信号进行非线性的自适应分解,称为经验模态分解 (Empirical Mode Decomposition, EMD),它把复杂信号分解成有限个本征模态函数 (Intrinsic Mode Function, IMF)之和,然后对各个 IMF进行 Hilbert变换去研究信号的时频能量分布. EMD分解是 HHT变换的核心部分,它的一个重要步骤是构造信号的上下包络线,以得到信号的瞬时平均.

收稿日期: 2009 - 03 - 23
基金项目: 总装预研基金 (51401040703CB0102)
作者简介: 王 婷 (1981—),女,博士研究生, E-mail: wt_0506@126. com;
杨莘元 (1944—),男,教授,博士生导师.

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Huang[1]等提出以信号的极大值点和极小值点拟合三次样条曲线的方式来构造上下包络线. 由于信号在端点处往往并非极值点,因此三次样条曲线容易在数据序列的两端出现发散现象,并且这种发散的结果会伴随着各个 MF的筛选过程逐渐向内“污染 整个数据序列而使所得结果严重失真,这就是 EMD方法的端点效应. 端点效应问题已经成为 EMD方法的一个瓶颈. 目前,已经提出的一些抑制端点效应的方法,包括极值延拓法[2]、神经网络预测法[3]、AR模型预测[4]、多项式拟合法[5]和支持矢量回归机[6]等. 这些方法在一定程度上都有抑制端点效应的作用.

本文在详细分析了现有改进方法的基础上,提出了一种最近相似距离的延拓方法. 首先给出了信号匹配距离和波形相似系数的计算公式,然后按照一定的筛选和匹配准则,在原信号内部找到与信号端点处最佳匹配的波段进行延拓,从而实现了信号端点处的波形沿着信号的自然趋势延拓.

1 EMD方法原理

经验模态分解是将信号分解为不同尺度特征的本征模态函数,这些本征模态函数满足如下两个条件:

1)其极值点和过零点的个数相等或者最多相差一个;

2)在任意点,由数据的极大值和极小值构成的包络线的平均值为零.

其筛选信号的过程如下[7]:确定信号 $x(t)$的所有局部极大值与局部极小值点,将其用三次样条函数拟合成原始信号的上下包络线;求出上下包络线的均值 $m(t)$,记:

$$h(t) = x(t) - m(t) \tag{1}$$

将 $h(t)$视为新的 $x(t)$,重复以上操作,直到 $h(t)$满足上述 MF条件,得到一阶 MF,记为

$$c_1 = h(t) \tag{2}$$

分离 MF,得到准余量函数:

$$r(t) = x(t) - c_1 \tag{3}$$

将 $r(t)$视为新的分解信号,重复以上过程可依次得到 $c_2, c_3, \cdots, c_n$ 或 $r(t)$满足给定的终止条件. 最终 $x(t)$被分解为

$$x(t) = \sum_{i=1}^{n} c_i + r(t) \tag{4}$$

对每阶 MF作 Hilbert变换,得

$$H[c(t)] = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{c(\tau)}{t-\tau} \mathrm{d}\tau \tag{5}$$

式中: P 代表柯西主值. 据此可求得其瞬时幅度 $a(t)$,瞬时相位 $\theta(t)$和瞬时频率 $f(t)$,即

$$a(t) = (c^2(t) + H[c(t)]^2)^{1/2} \tag{6}$$

$$\theta(t) = \arctan\left(\frac{H[c(t)]}{c(t)}\right) \tag{7}$$

$$f(t) = \frac{1}{2\pi} \frac{\mathrm{d}\theta(t)}{\mathrm{d}t} \tag{8}$$

信号 $x(t)$可表示为

$$x(t) = \mathrm{Re} \sum_{i=1}^{n} a_i(t) \exp\left[\mathrm{j} \int f(t) \mathrm{d}t\right] \tag{9}$$

2 基于最近相似距离的端点延拓处理

EMD方法自推出以来已经成功地应用到许多非线性研究领域. 但是在应用 EMD方法时存在着较为严重的端点问题,这个问题是在构造信号的包络线时产生的. 由于在确定信号的上、下包络线时,采用了三次样条函数的方法,而三次样条差值的特点是必须在极大值和极小值数据集两端增加极大值和极小值. 但是,因为原始信号的端点可能不是极值点,所以必须对其进行合理的延拓. 可以说数据延拓的质量直接决定了抑制的效果,进而影响 EMD分解的质量.

2.1 概念引入

定义两点之间的匹配距离为

$$D_i = x(t_1) - y(t_2) \tag{10}$$

式中: $x(t_1)$为信号 $x(t)$上的一点, $y(t_2)$为信号 $y(t)$上的一点.

定义两段波形之间的匹配距离为

$$D_{i\mathrm{sum}} = \sum_{i=1}^{n} \left| x_1(t_1 + i\Delta t) - y(t_2 + i\Delta t) \right| \tag{11}$$

式中: $\{x(t_1 + \Delta t), x(t_1 + 2\Delta t), \cdots, x(t_1 + n\Delta t)\}$为信号 $x(t)$上的局部信号; $\{y(t_2 + \Delta t), y(t_2 + 2\Delta t), \cdots, y(t_2 + n\Delta t)\}$为信号 $y(t)$上的局部信号; n为局部信号长度. 由式 (11)可知,两段曲线匹配距离越小则两段曲线越相近,反之差异越大.

定义两段波形之间的平均距离为

$$D_{\mathrm{mean}} = \frac{D_{i\mathrm{sum}}}{n}$$

定义两段波形之间的波形相似系数:

以 D_{mean}为参照计算两段信号对应各点的匹配距离与 D_{mean}的距离,该距离的累加和取平均值即是

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

波形相似系数:

$$R_i = \frac{\sum_{i=1}^{n} | x(t_1 + i\Delta t) - y(t_1 + i\Delta t) - D_{mean} |}{n} \tag{12}$$

由式 (12)可知,两段曲线波形相似系数越小则两段曲线越相似,反之差异越大.

2.2　最近相似距离延拓数据

端点延拓的关键就是要确定原始信号在端点处的变化趋势.原始信号的发展趋势不仅会表现在端点处,在信号的内部也必将存在.如果在信号的内部找到一段子波其变化趋势与端点处非常近似的话,那么就可以用这段子波的前后数据来延拓信号端点处的那段波形,这样的延拓方式最大限度地保护了原始信号的变化特征,从而能有效地抑制端点效应的产生.

匹配距离定量给出了两段信号在幅度上的差异;波形相似系数定性强调了两段波形特征的相似性而忽略了其幅度差异.为了更好的延拓信号端点处的波形,使延拓数据最为符合信号的自然走向,本文综合考虑了以上两个因素,信号的延拓包括左右两端,下面以左端点为例,对本文提出的算法具体介绍.设原始信号为 $x(t)$:

1) 从左到右标记信号 $x(t)$的所有极大值点所在下标的集合为 $\{M_{max} | M_0, M_1, \cdots, M_k\}$,所有极小值点集合为:$\{N_{min} | N_0, N_1, \cdots, N_k\}$;并记信号 $x(t)$左端点为 X_0;从 X_0 到 N_0 的这段波形记为 ω_0,设其长度为 N,其中从 X_0 到 M_0 的长度记为 L(这里假设先有极大值后有极小值),则有 $L < N$.

2) 令此段波形 ω_0 以 M_0 为参考点在极大值集合 M_{max}中向右移动,所移之处 M_0 均与 M_i 重合,分别计算该段相同长度的波形 ω_i 和 ω_0 的匹配距离 D_{isum}和波形相似系数 R_i.

3) 定义最近相似距离 $E_i = D_{isum} + R_i$,取 E_i最小时所对应的波形为最近匹配波形,记此时 M_0 对应的极大值为 M_p,则 X_0 对应的点的坐标应为 $X_p = M_p - L$.

4) 自匹配波 X_p 的前一点数据开始,将实际波形数据依次向左延拓到 X_0 前,延拓点数根据需要选择,若信号中 X_p 前数据点个数小于需要延拓的点数,可反复延拓此段波形.

5) 同理可对信号右端点进行延拓.

3　仿真分析

为了检验算法的有效性,对由下式给出的模拟信号进行经验模态分解,接着再对其分解结果进行希尔伯特变换,得到 Hilbert谱,并将其与极值延拓法分解结果的 Hilbert谱进行比较.

$$x(t) = \sin(2\pi 25t) + 0.5\cos(2\pi 10t) + 0.9\cos(2\pi 5t)$$

其中:$0 < t \leq 1$ s,采样频率为 600 Hz,数据点数为 $N = 600$.两种方法均是在相同 EMD分解条件下进行分解的,最终得到的 Hilbert时频谱如图 1、图 2所示.

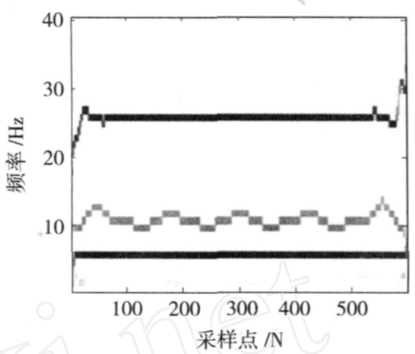

图 1　极值延拓法的 Hilbert 时频谱

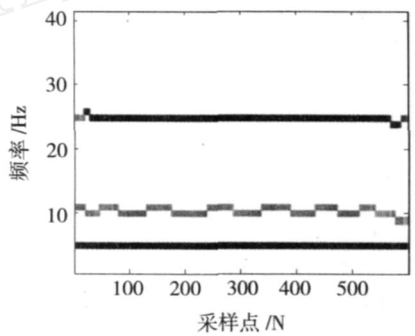

图 2　最近相似距离延拓 Hilbert 时频谱

以图可见,基于极值延拓法的 Hilbert时频谱在谱线的左右两端,瞬时频率均出现了明显的摆动,能量有泄露.而本文提出的基于最近相似距离延拓的 Hilbert时频谱上几乎消除了这种摆动现象.由此可见,该方法对信号的两端的延拓有效地抑制了 EMD分解在端点处的发散问题,大大提高了频率分辨率.

为了进一步验证算法的有效性,对钻井振动信号的数据进行了分析,采样频率为 4 096 Hz,数据点数为 $N = 200$. EMD分解结果如图 3所示.

图 3(a)为极值延拓后的 EMD分解结果图,图 3(b)为本文方法延拓后的 EMD分解结果图.比较分解结果,不难看出,虽然在分解的高频分量上差异不是很明显,但是经极值延拓后 EMD分解得到的低频分量,已经出现了发散现象,而采用基于最近相似距

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

离的数据延拓方法后,其得到的分解结果较为理想,这表明端点效应得到了有效的抑制.

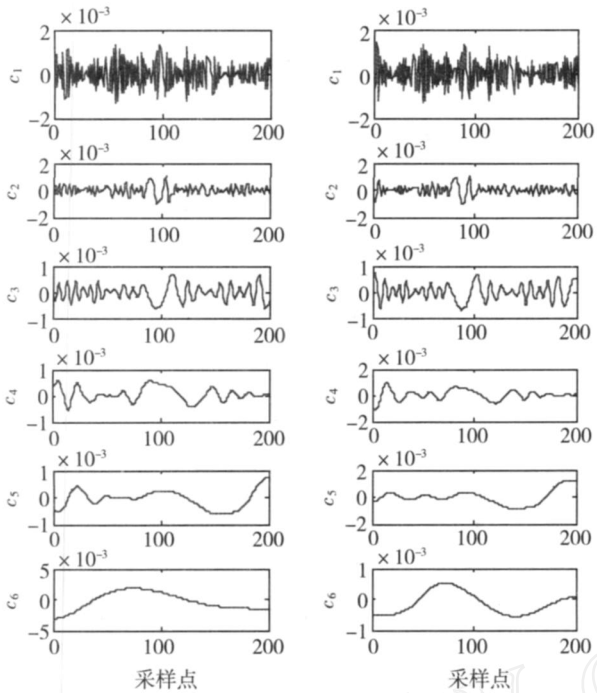

图 3 钻井振动信号的 EMD分解

4 结 语

本文提出的基于最近相似距离的延拓方法充分兼顾了原始信号中的极值点及非极值点的波形数据,同时给定了匹配距离和波形相似系数概念,综合考虑了波形和幅值两方面因素,使最终获得的匹配段更符合数据自然走向,实现了延拓数据与原信号交界处的光滑过渡.通过仿真分析,表明了该方法能有效地抑制 EMD分解中两个端点处信号的发散,从而提高对时频图两个端点处谱图的分辨率.

参 考 文 献:

[1] HUANG N E. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis [J]. Proceeding of the Royal Society of London, 1998, 454(12): 903 - 995.

[2] HUANG Daji. Practical Implementation of Hilbert-Huang Transform Algorithm [J]. Acta Oceanologica Sinica, 2003, 22(1): 1 - 14.

[3] DENG Yongjun, WANG Wei, QIAN Chengchun, et al. Boundary Processing Technique in EMD Method and Hilbert Transform [J]. Chinese Science Bulletin, 2001, 46(11): 954 - 960.

[4] 杨建文,贾民平. 希尔伯特 - 黄谱的端点效应分析及处理方法研究 [J]. 振动工程学报, 2006, 19(2): 283 - 288.

[5] 刘慧婷,张 昱,程家兴. 基于多项式拟合算法的 EMD端点问题的处理 [J]. 计算机工程与应用, 2004, 16: 83 - 86.

[6] 程军圣,于德介,杨 宇. 基于支持矢量回归机的 Hilbert-Huang变换端点效应问题的处理方法 [J]. 机械工程学报, 2006, 42(4): 23 - 31.

[7] 李雪耀,黄永平,张汝波. 一种基于支持向量回归机的经验模态分解方法 [J]. 哈尔滨工程大学学报, 2007, 28(7): 779 - 784.

[8] KACHA A, GRENEZ F, BENMAHAMMED K. Time-frequency Analysis and Iinstantaneous Frequency Estimation Using Two - sided Linear Prediction [J]. Signal Processing, 2005, 85(3): 491 - 503.

[9] 宋立新,王 祁,王玉静,等. 具有间断事件检测和分离的经验模态分解方法 [J]. 哈尔滨工程大学学报, 2007, 28(2): 178 - 182.

(编辑:温泽宇)

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net