

EMD 信号分析方法端点问题的处理

黎洪生, 吴小娟, 葛 源

(武汉理工大学 自动化学院, 湖北 武汉 430070)

摘要: 经验模态分解(EMD)是一种新的处理非线性、非平稳的数据分析方法,但是在利用样条插值获得上下包络过程中存在着棘手的端点问题。在解决该问题已有的添加极值点算法的基础上,提出了通过添加极值点和对称延拓相结合的方法抑制端点问题的思路和策略。针对一个仿真振动信号,对比分析了直接以数据端点作为极值点、多项式拟合算法、神经网络延拓算法、极值点与对称延拓相结合 4 种算法的效果,结果显示了所提出方法能有效地抑制端点效应。

关键词: 经验模态分解; 端点问题; 三次样条

中图分类号: TP 274

文献标识码: A

文章编号: 1006-6047(2005)09-0047-03

目前,对非平稳信号分析处理的主要方法有^[1]: 短时傅里叶变换、时频分析理论和小波变换。一种新的处理非线性、非平稳的数据处理方法——经验模态分解方法(EMD)^[2,3]与传统的数据处理方法完全不同,它认为每组数据可以分解为有限固有模态函数(IMF)之和,分解出的各个 IMF 分量突出了数据的局部特征,对其分析可以更准确把握原数据的特征信息。但是 EMD 分解中,由于计算局部极大值与局部极小值定义的包络均值时,采用了三次样条插值,而信号的 2 个端点可能不是局部极值点,因此,样条插值在信号的 2 个端点处会出现大的摆动,这种摆动可能会传播到信号数据的中间段并破坏整个数据特性。针对这一问题,文献[4]提出了通过添加极值点抑制,并指出该方法的前提是新的极大值和极小值数据集的最大间距必须大于等于原始信号长度;文献[5]提出了多项式拟合算法,通过拟合函数求出端点的近似值,由于该拟合函数是由极少数据确定,所以对最后的分解结果有一定的影响;文献[6]提出了神经网络算法,先确定网络模型的权重向量和偏移量,再由此模型对原始数据进行左右延拓,该方法对大多数信号数据,都能很好地抑制,但该算法的最大不足就是速度太慢。本文在以上算法的基础上,提出了添加极值点和对称延拓^[7]相结合的算法。

1 EMD 方法

EMD 方法从本质上讲是对一个信号进行平稳化处理,其结果是将信号中不同尺度的波动或趋势逐级分解,产生一系列具有不同特征尺度的数据序列,每个序列成为 1 个固有模态函数 IMF。IMF 必须满足以下条件:在一段数据中极值点数目等于数据过零点数目或者最多相差为 1;在任何 1 点由极大值点和极小值点定义的包络的平均值为零,运用 IMF 可以把任何信号 $x(t)$ 按如下步骤分解。

a. 确定信号所有的局部极值点,然后用三次样条线将所有的局部极大值点连接起来形成上包络线,再用三次样条线将所有的局部极小值点连接起来形成下包络线,上下包络线应该包络所有的数据点。

b. 上下包络线的平均值记为 m_1 ,得 $x(t)-m_1=h_1$ 。

c. 理想情况下,若 h_1 为 1 个 IMF,则 h_1 就表示 $x(t)$ 的第 1 个分量;若 h_1 不满足 IMF 的条件,把 h_1 作为原始数据,重复上述步骤,得到上下包络线的平均值 m_{11} ,再判断 $h_{11}=h_1-m_{11}$ 是否满足 IMF 的条件,如不满足,则重复循环 k 次,得到 $h_{1(k-1)}-m_{1k}=h_{1k}$,使得 h_{1k} 满足 IMF 的条件。记 c_1 为信号 $x(t)$ 中第 1 个满足 IMF 的分量。把 c_1 从 $x(t)$ 中分离,得到 $r_1=x(t)-c_1$,将 r_1 作为原始数据,重复以上的过程,得到 $x(t)$ 中第 2 个满足 IMF 的分量 c_2 ,重复循环 n 次,得到信号 $x(t)$ 的 n 个满足 IMF 的分量。这样就有

$$\begin{cases} r_1 - c_2 = r_2 \\ \vdots \\ r_{n-1} - c_n = r_n \end{cases}$$

当 r_n 成为 1 个单调函数不能再从中提取满足 IMF 条件的分量时,循环结束。由上可以得到 $x(t) = \sum_{j=1}^n c_j + r_n$ 。因此对任何 1 个信号 $x(t)$,可以将其分解为 n 个基本模式分量和 1 个残量 r_n 之和,其中,分量 $c_1, c_2, \dots, c_n$ 包含了信号从高到低不同频率段的成分,而 r_n 则表示了信号 $x(t)$ 的中心趋势。

2 端点问题

EMD 分解中的关键一步就是采用三次样条求上下包络的平均值,三次样条曲线具有光滑的一次微分和连续的二次微分特点。由于所分析信号的有限长度,信号的两端点不能确定是极值,因此,在进行三次样条插值时,必然使得信号的上下包络在信号的两端附近严重扭曲。在信号的高频分量,由于时间尺度小,极值间的距离小,端部的边缘效应仅局限在信号

两端很小的部分。但对于低频分量,时间尺度大,极值间的距离大,端部的边缘效应就传播到信号的内部,特别是原始信号数据集比较短时,会严重影响 EMD 分解的质量,使得分解出的 IMF 分量无实际物理意义。

3 对称延拓与添加极值点相结合

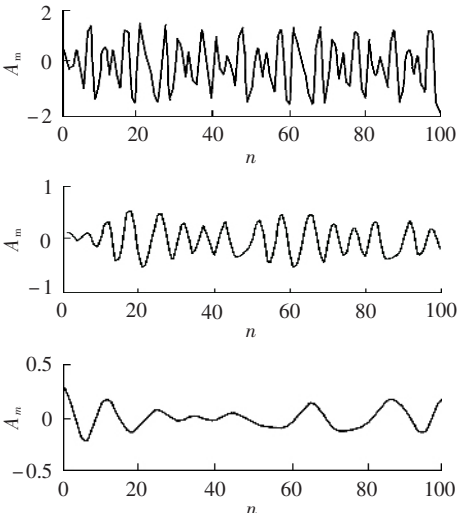
基于三次样条插值的特点,当原始信号的两端点不是极值点时,如果能够根据端点以内极值点序列的规律得到该序列在端点处的近似值,则可以防止对极值点进行样条插值得到的包络线出现极大的摆动。取出原极值点序列最左端 1/3 的极值点,根据该数据的间距均值和左端点的幅值,定出左端点需增加的极值点位置和幅值。同理,可定出右端点需增加的极值点位置和幅值,当所构成的新的极大值和极小值数据集的最大间距小于原始信号时,以近似的左端点处增加的极值点为起始点,向左进行数据对称延拓;以近似的右端点处增加的极值点为起始点,向右进行数据对称延拓。将得到的新的数据作为一个整体分解 EMD,需要指出的是所取的结果只是中间部分。虽然该方法只是求出左右端的近似值,对近似极值点进行对称延拓,但是如文献[5]所说,延拓的目的不是为了给出准确的原序列以外的数据,而是提供一种条件,使得包络完全由端点以内的数据确定。

4 几种处理方法的比较

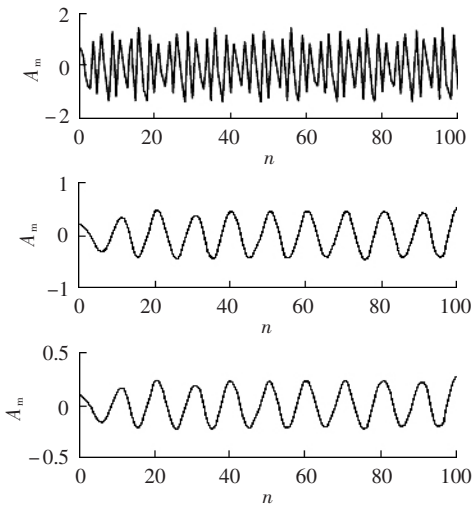
发生机械故障时,故障信息往往以调制形式存在于振动和噪声中,调制信号反映了故障的特征信息^[8]。下面仿真一个机械振动信号,该信号是由基频为 30 Hz,调制频率 15 Hz 的调频调幅信号和频率为 120 Hz 正弦信号叠加,表达式为

$$x(t)=[1+0.2\sin(2\pi7.5t)]\cos[2\pi30t+0.5\sin(2\pi15t)]+\sin(2\pi120t)$$

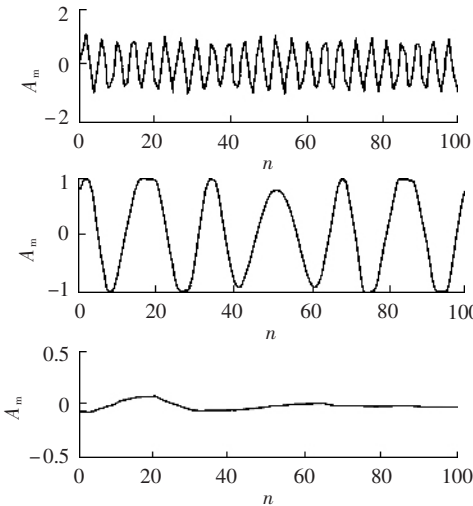
4 种算法对振动信号分解后的结果如图 1 所示。



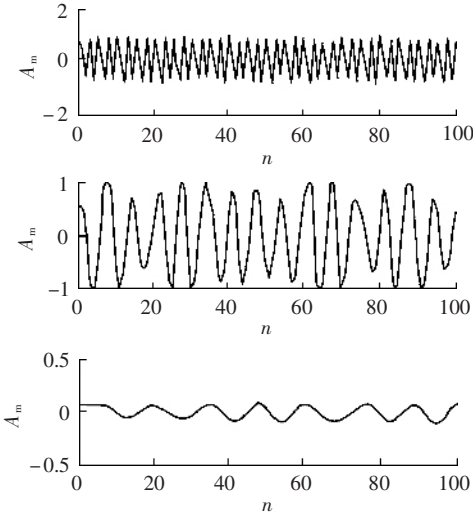
(a) 以数据端点为极值点的 EMD 分解结果



(b) 基于多项式拟合的 EMD 分解结果



(c) 神经网络延拓算法的 EMD 分解结果



(d) 极值点与对称延拓相结合后的 EMD 分解结果

图 1 4 种算法对振动信号分解后的结果

Fig.1 The decomposition results of vibration signal by four algorithms

图 1 是利用以数据端点作为极值点、多项式拟合、神经网络延拓、极值点与对称延拓相接合 4 种算法对该振动信号分解后的结果(图中 A_m 为幅值; n 为采样点数)。

从图 1(a)的分解结果可知,对于该仿真的振动信号,采用直接以数据端点作为极值点,虽然很好地抑制了解得到的 IMF 的端部特征,但是它实质是强制扭曲了信号的端部特征,因此严重污染了信号。从图 1(b)的分解结果可知,在多项式拟合算法中,由于拟合函数是利用最小二乘法得到,因此用来作为拟合的数据点数不能太多,其次多项式次数的取定,也是导致拟合函数能不能正确地反映原始数据序列在端点处的变化趋势。从图 1(c)的分解结果可知,神经网络延拓算法较好地抑制了端点效应,但由于在每次循环中要利用该方法对原始数据序列的两端进行数据延拓,每次若延拓 N 个点就要循环 N 次,因此,循环次数太多造成了运算速度太慢。如图 1(d)所示,极值点与对称延拓相结合的方法对该振动信号的分析结果十分理想,所以不论从理论分析还是实验分析,都可以看出该算法是一种较好的方法。

5 结论

从对 4 种算法的比较中,可以看出用数据集的端点作为极值点的方式抑制端部翘曲非常有效,但要注意低频的分解泄漏,产生不存在的 IMF 模式分量。用多项式拟合的方法抑制端点,由于其数据点不能取的过多以及多项式的次数的不确定性,使得最后的分解结果不能较好地反映原始信号的特征。神经网络延拓算法的最大不足就是运算速度太慢。添加极值点与对称延拓相结合的方法针对本文中的仿真信号效果较好,但是该方法中极值点的选取对分解结果影响较大,因此未来需要完善之处,在于根据所提出的基本思想,提出一个自适应的较好预测端点附加极值的方法。

参考文献:

- [1] 张贤达,保铮. 非平稳信号分析与处理[M]. 北京:国防工业出版社,1998.
- [2] HUANG N E. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. *Proc. Rsoc Lond A*, 1998, (454): 903-995.

- [3] HUANG N E, SHEN Z. A new view of non-linear water waves: The Hilbert spectrum[J]. *Annu Rev Fluid Mech*, 1999, (31): 417-457.
- [4] 陈忠,郑时雄. EMD 信号分析方法边缘效应的分析[J]. 数据采集与处理, 2003, 18(3): 114-118.
CHEN Zhong, ZHENG Shi-xiong. Analysis on end Effects of EMD method[J]. *Journal of Data Acquisition & Processing*, 2003, 18(3): 114-118.
- [5] 刘慧婷,张旻,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): 84-86.
LIU Hui-ting, ZHANG Min, CHENG Jia-xing. Dealing with the end issue of EMD based on polynomial fitting algorithm[J]. *Computer Engineering and Applications*, 2004, (16): 84-86.
- [6] 邓拥军,王伟,钱成春. EMD 方法及 Hilbert 变换中边界问题的处理[J]. 科学通报, 2001, 46(3): 257-263.
DENG Yong-jun, WANG Wei, QIAN Cheng-chun. EMD and Hilbert dealing in the problem of borderline [J]. *Science Bulletin*, 2001, 46(3): 257-263.
- [7] ZHAO Jin-ping. Mirror extending and circular spline function for empirical mode decomposition method [J]. *Journal of Zhejiang University*, 2000, 54(3): 247-252.
- [8] 杨世锡,胡劲松,吴昭同,等. 旋转机械振动信号基于 EMD 的希尔伯特变换和小波变换时频分析比较 [J]. 中国电机工程学报, 2003, 23(6): 102-107.
YANG Shi-xi, HU Jin-song, WU Zhao-tong, et al. The comparison of vibration signal's time-frequency analysis between EMD-based HT and WT method in rotating machinery [J]. *Proceedings of the CSEE*, 2003, 23(6): 102-107.

(责任编辑:汪仪珍)

作者简介:

黎洪生(1961-),男,湖北武汉人,教授,博士研究生导师,主要研究基于网络的远程控制与智能诊断理论、方法和技术,基于网络的远程维护技术、嵌入式控制与分布式控制、智能控制等;

吴小娟(1981-),女,湖北红安人,硕士研究生,研究方向为远程故障诊断(E-mail: wuxiaojuan810201@163.com);

葛源(1981-),男,新疆博湖人,硕士研究生,研究方向为计算机控制与信息系统集成。

Dealing with end issue of EMD method

LI Hong-sheng, WU Xiao-juan, GE Yuan

(Wuhan University of Technology, Wuhan 430070, China)

Abstract: EMD(Empirical Mode Decomposition) is a way to process the nonlinear or non-smooth data. But the end issue appears when using spline interpolation to get two envelopes of the data. A new method combining 'symmetrical continuation' with 'adding extremum' is proposed for it. A simulated vibration signal is decomposed by four methods for comparison: 'original end data as one extremum', 'polynomial fitting algorithm', 'neural network continuation algorithm', and 'combination of consecutive extremes and symmetrical continuation'. Results show that the proposed method restrains the end effects effectively.

Key words: empirical mode decomposition; end issue; cubic spline