

基于 EMD-R/S 分析的太赫兹光谱降噪

魏博熠 罗鉴鹏 张立臣

(广东工业大学计算机学院 广东 广州 510006)

摘要 针对太赫兹时域光谱系统由于延时线的重合抖动、采样抖动等产生的噪声,提出使用经验模态分解-R/S 分析方法对太赫兹光谱信号进行降噪。采集太赫兹时域光谱系统的时域信号,根据 EMD 算法将信号分解成本征模态函数(IMF);使用 R/S 分析法分别计算各个 IMF 的 Hurst 指数。根据设定的阈值判断是否各个 IMF 是否存在均值回复的情况。如果 IMF 存在均值回复的现象,则使用原始信号与 IMF 信号作差,所得信号即为降噪后的时域信号。实验结果表明,与小波降噪算法相比,EMD-R/S 分析算法能够有效地对太赫兹时域光谱信号降噪,能够有效还原太赫兹光谱信号特征。

关键词 太赫兹时域光谱 经验模态分解(EMD) R/S 分析 降噪

中图分类号 TP301 **文献标志码** A **DOI**:10.3969/j.issn.1000-386x.2022.03.011

TERAHERTZ SINGAL DENOISING METHOD BASED ON EMD-R/S ANALYSIS

Wei Boyi Luo Jianpeng Zhang Lichen

(School of Computers, Guangdong University of Technology, Guangzhou 510006, Guangdong, China)

Abstract In the terahertz time-domain spectroscopy system, due to the coincidence jitter of the delay line and the noise generated by sampling jitter, the empirical mode decomposition-R/S analysis method is proposed to denoise the terahertz spectral signal. The time domain signal of the terahertz time-domain spectroscopy system was acquired, and the signal was decomposed into a series of intrinsic mode components (IMF) according to the EMD algorithm. The Hurst index of each IMF was calculated using R/S analysis. It was determined whether the respective IMFs had an average reply according to the set threshold. If the IMF had the phenomenon of mean recovery, the original signal was used to make a difference with the IMF signal, and the resulting signal was the time domain signal after noise reduction. The experimental results show that compared with the wavelet denoising algorithm, the EMD-R/S analysis algorithm can effectively denoise the terahertz time-domain spectral signal and effectively reduce the terahertz spectral signal characteristics.

Keywords Terahertz time-domain spectroscopy Empirical mode decomposition R/S analysis Denosing

0 引言

太赫兹时域光谱(Terahertz Time Domain Spectrum)可以展示物质分子级别的化学键相互作用而引起的能级跃迁变化,反映了物质丰富的物理化学等信息,通过太赫兹光导天线、太赫兹探测器等器件获取太赫兹时域光谱信号,从而获得被测物质的信息。太赫

兹时域光谱系统发射一个太赫兹脉冲,这个太赫兹脉冲一般包含半个到几个振荡周期。太赫兹探测器直接测量太赫兹波的电场随时间的变化的信息。这个时域信息不但包含了光强,还测量了相位信息。此外,在太赫兹频段范围内,对应着有机物分子的偶极子转动和振动跃迁以及大分子间的骨架振动等^[1-2]。根据这个特性,太赫兹光谱技术在毒品检测^[3-5]、食品质量检测^[6-9]等领域有广泛应用。近年来,在太赫兹时域光

谱系统的基础上,进一步研究太赫兹光谱的数据处理方法,对将太赫兹技术推向更广阔的应用领域有重大意义。

使用太赫兹时域光谱系统测量物质,不可避免会混杂一些噪声,对含噪信号进行降噪处理很有必要。太赫兹光谱的噪声可以分为背景噪声和太赫兹波所携带的噪声。这两类噪声表现为白噪声和模糊噪声的混合形式。目前对于太赫兹时域光谱信号的降噪处理主要是采用小波变换的方法^[10-12]。小波变换的方法对于周期性的平稳信号比较有效,而由于非平稳信号包含尖峰或者突变信息,小波变换处理非平稳信息有时效果欠佳。而且,小波变换处理信号需要选择合适的小波基,对分解层数的选择也需要依靠先验经验,因此这种数据处理方法存在一定的局限性。经验模态分解(Empirical Mode Decomposition, EMD)是 Huang 等^[13]在 1998 年提出的处理非线性和非平稳时间序列信号的一种方法。与已有的方法相比,EMD 直观、直接并且能自适应不同的信号,因为 EMD 分解算法基于信号数据产生基函数。EMD 基于一个简单的假设:任何数据都包含了震荡的不同简单背部模态,所有不同的模态数据叠加最终得到复杂的数据。对时序信号采用 EMD 分解,结合信号物理背景,可以得到信号清晰的物理意义。

对太赫兹光谱信号进行经验模态分解后,可以得到若干个固有模态函数(Intrinsic Mode Function, IMF)分量,每个分量含有不同的频率成分。本文使用 Hurst 指数识别出 IMF 分量中的白噪声分量,将其从原始信号中剔除,从而获得较为准确的信号,方便后续处理。基于 R/S 分析法的 Hurst 指数是英国水文学家 H. E. Hurst 建立的用于判断时间序列数据遵从随机游走还是有偏的随机游走过过程的指标。统计学上一般认为当 Hurst 指数等于 0.5 时,该信号接近于高斯白噪声信号。根据太赫兹噪声的特点,采用 Hurst 指数识别噪声会是一个合适的选择。

本研究在使用 EMD 降噪的研究基础上,提出了一种基于 EMD 和 DFA 相结合的太赫兹光谱降噪方法。该方法对含噪的太赫兹光谱信号进行 EMD 分解,得到多个频率信号分量。通过引入 Hurst 指数判断各个 IMF 分量的回复性,将判断为噪声的信号从原始信号中剔除,获得降噪后的太赫兹光谱信号。

1 EMD 的理论基础

EMD 算法的目的是将信号分解成一系列的 IMF,

再通过希尔伯特变换获得信号的时频关系。经验模态分解作为一种新的数据分析方法,在处理非平稳、非线性数据上具有明显的优势,特别适合处理太赫兹时域信号。在 EMD 算法中,IMF 具有以下的定义:

(1) 整个数据中,极值点和过零点数目要么相等,要么最多相差 1。

(2) 任意一点,由局部极大值定义的包络和由局部极小值定义的包络的均值等于 0。

与简单的谐波函数相比,IMF 没有简单谐波函数中的恒定振幅和频率,具有以时间为函数变化的振幅和频率。

以太赫兹时域光谱信号为例,EMD 算法的基本计算过程如下:

Step1 计算太赫兹时域光谱信号 $x(t)$ 的所有极大值点和极小值点,用三次样条函数连接所有的极大值点,所得曲线构成上包络线 $u(t)$;用三次样条函数连接所有的极小值点,所得曲线构成下包络线 $v(t)$ 。

Step2 计算上下包络线的均值,计算公式为:

$$m(t) = \frac{u(t) + v(t)}{2} \quad (1)$$

EMD 算法的成分分量定义为:

$$h(t) = x(t) - m(t) \quad (2)$$

Step3 判断成分分量 $h(t)$ 是否满足 IMF 的定义。如果 $h(t)$ 满足 IMF 的定义,则 $h(t)$ 是第一个筛选出来的 IMF 分量,记为 $c_1(t)$ 。如果 $h(t)$ 不满足 IMF 的定义,则将 $h(t)$ 作为原始数据,重复上述两步,直至计算得到第一个 IMF 分量,该分量同样记为 $c_1(t)$ 。

Step4 将第一个 IMF 分量从原始信号中剔除,得到剩余信号,计算公式如下:

$$r_1(t) = h(t) - c_1(t) \quad (3)$$

将 $r_1(t)$ 作为新的待分析信号,重复上述 3 个步骤,这样就得到第二个 IMF 分量 $c_2(t)$ 。将第二个 IMF 分量从 $r_1(t)$ 中剔除,得到新的余项 $r_2(t) = r_1(t) - c_2(t)$ 。不断重复上述步骤,直至满足 EMD 算法的停止准则。EMD 算法的停止准则使用了 Cauchy 收敛性准则,这种检测要求相邻两个抽取操作的归一化平方差足够小,具体定义为:

$$SD_k = \frac{\sum_{t=0}^T |h_{k-1}(t) - h_k(t)|^2}{\sum_{t=0}^T h_{k-1}(t)^2} \quad (4)$$

式中: T 表示信号长度。当 SD_k 小于设定的阈值时,分

解结束。

2 R/S 分析法

R/S 分析法是一种用于处理时间序列的统计量的方法,其对于时间序列具备一定的预测功能,同时在分形理论中有着重要的作用。R/S 分析法的基本思想是对一段时间序列数据进行不同尺度的分解,研究其在不同的时间范围内的统计性变化规律,然后将小范围内的统计变化规律进一步延伸至大范围时间,得到整段时间序列数据的统计学规律。当然,也有学者将大范围时间变化规律用于小范围时间预测方面。

太赫兹时域光谱信号是一段辐射强度随时间变化的数据,这个信号数据可以表示成 $x(t_1), x(t_2), \dots, x(t_n)$ 的形式,也即为“时间序列”。臧元章等^[14]指出,太赫兹时域光谱仪的幅值噪声来源为延时线的重合抖动、采样抖动以及飞秒激光抖动,是白噪声与 $1/f$ 噪声的混合。这种混合分形噪声是非平稳过程,理论上并不适合采用傅里叶变换、小波变换等谱分析方法,而 EMD 算法针对非平稳信号,可以将特征信号与噪声分离。此时,采用 R/S 分析法即可判断哪些是信号,哪些是噪声,然后重构成高信噪比的信号。R/S 分析法的实现步骤如下:

Step1 对使用 EMD 算法分解后长度为 N 的 IMF 分量 $x(t)$,以长度 n 均分这个 IMF 分量,得到 m 个相邻的子区间。对这 m 个相邻的子区间,可以表示为 $I_k, k=1, 2, \dots, m$ 。其中 $N = m \times n$ 。任一个子区间中的元素可以表示为 $N(k, n)$,即表示任一个元素在第 k 个子区间,在该区间的第 n 位。可以得到任一子区间的均值为:

$$E_k = \frac{1}{n} \sum_{t=1}^n x(t) \quad (5)$$

Step2 计算任一子区间均值的累积离差:

$$X_k = \sum_{i=1}^u (N(k, i) - E_k) \quad (6)$$

其中, $u=1, 2, \dots, n$ 。

Step3 计算极差:

$$R_{I_k} = \max(X_k) - \min(X_k) \quad (7)$$

Step4 计算子区间的标准差:

$$S_{I_k} = \sqrt{\frac{\sum_{t=1}^n (N(k, t) - E_k)^2}{n}} \quad (8)$$

Step5 对每一个 R_{I_k} 由对应的 S_{I_k} 进行标准化,得到 R_{I_k}/S_{I_k} 。

Step6 n 从 4 开始计算,重复上述 5 个步骤,直至 $n = N$,即可得到序列 $[R/S]_n, n=4, 5, \dots, N$ 。Hurst 指数可以表示为:

$$(R/S)_n = (c \times n)^H \quad (9)$$

式中: H 表示 Hurst 指数; c 为常数。

对式(11)两边取自然对数,可得:

$$\ln \frac{R}{S} = H \ln c + H \ln n \quad (10)$$

Step7 对 $\ln \frac{R}{S}$ 与 $\ln n$ 做线性回归计算,所得斜率即为 Hurst 指数。

Hurst 指数是 R/S 分析的有效统计量。在分形理论中, $H=0.5$ 表明时间序列可以用随机游走来描述; $0 \leq H < 0.5$ 表明该时间序列是 $1/f$ 噪声,存在均值回复的特性; $H > 0.5$ 表示时间序列存在长期记忆性。

3 基于 EMD-R/S 分析的光谱降噪实验

黄瑞瑞等^[15]进行了乳糖的太赫兹光谱和红外光谱的特性研究。本文采用乳糖作为实验材料进行研究,将本文的实验数据与文献中的实验结果进行对比,本文测量的乳糖在频域相似位置出现了吸收峰,证明测量过程是准确的。以厚度为 1.3 mm 的乳糖压片的太赫兹透射谱为例,测量时温度为 23 °C,湿度为 3.4%。测量得到的太赫兹时域信号如图 1 所示。

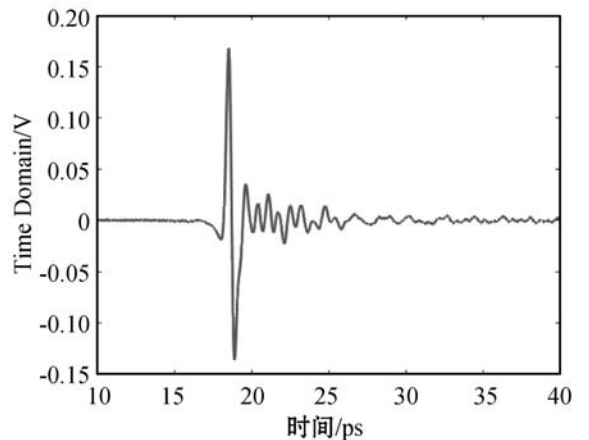


图1 乳糖压片太赫兹时域光谱

对乳糖压片的太赫兹时域光谱信号采用 EMD 算法分解得到如图 2 所示的结果。从分解结果中看到,IMF2 分量与原始信号形状相似,保留了原始信号的大部分信息。其余分量代表了噪声等不同的物理

信息。

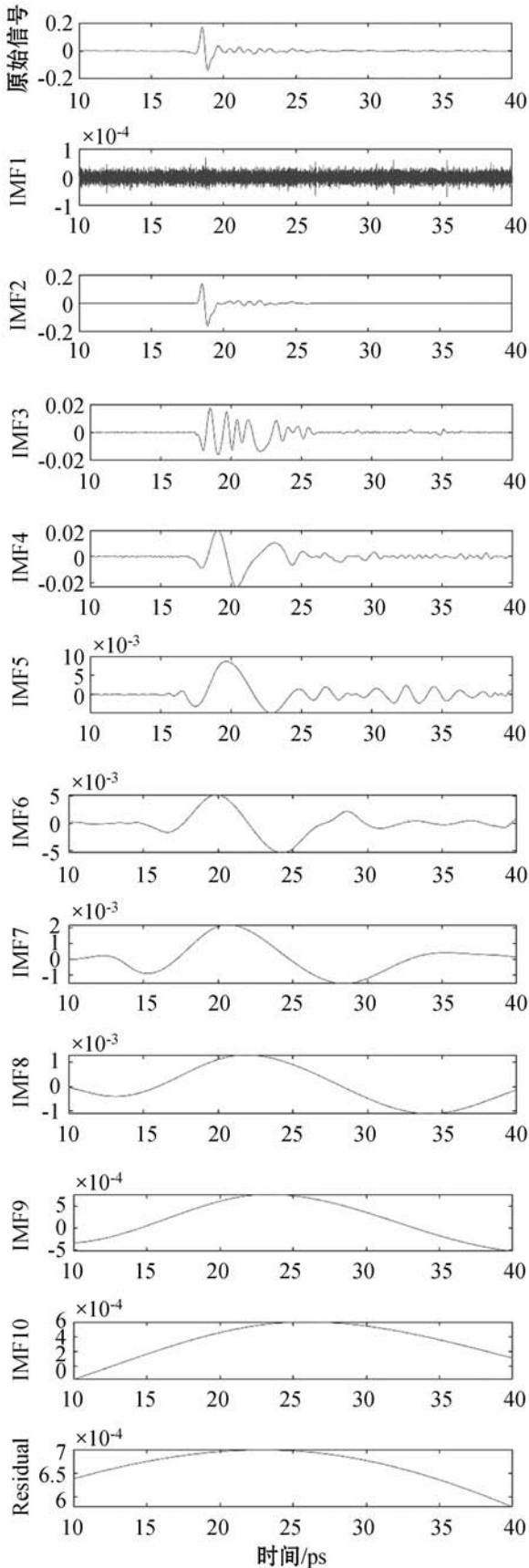


图2 EMD 分解结果

对 EMD 算法分解得到的 IMF 分量逐个使用 R/S 分析法计算 Hurst 指数。各个 IMF 分量的 Hurst 指数

分别为 0.42、0.69、0.76、0.66、0.84、0.79、0.91、0.96、0.99、0.99。从 Hurst 指数计算结果,可以得到 IMF1 分量是信号中的噪声。将噪声信号从原始信号中剔除,可以得到重建信号。为了方便对比,本文将重建信号向下平移了 0.1 个单位,效果如图 3 所示。重建后的信号跟原始信号相比更加平滑,毛刺更少。

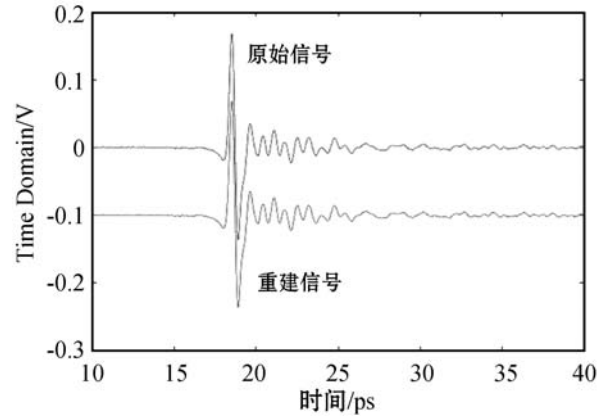


图3 原始时域信号与重建时域信号对比

对物质的太赫兹信号分析常常不局限于时域信号,不少学者会将时域信号转换到频域进行分析。实验分别对原始信号与重建信号做傅里叶变换。为了方便观察,重建频域信号进行了向下平移,所得结果如图 4 所示。与原始频域信号相比,重建频域信号更平滑。

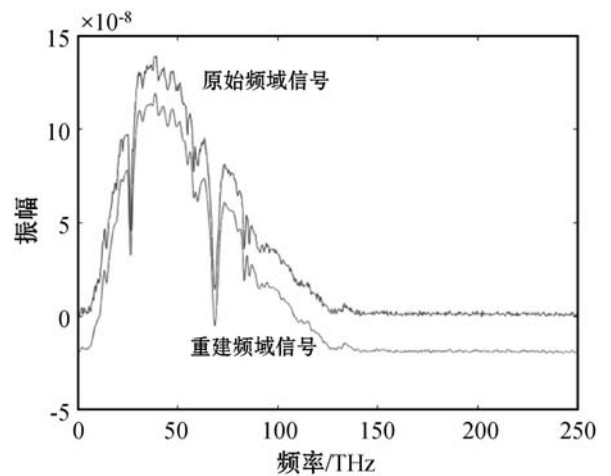


图4 原始频域信号与重建频域信号对比

4 算法性能评价

实验采用小波软阈值降噪、小波硬阈值降噪和本文提出的方法分别对 20 个不同厚度的乳糖压片进行了太赫兹光谱测量,分别计算原始时域信号和重建时域信号前后的相关系数、信噪比和峰值信噪比三个指标。本文测量 20 个实验室制造的乳糖压片,厚度范围

为0.8~1.6 mm,每个乳糖压片表面光滑。乳糖压片的厚度使用电子游标卡尺测量,对同一个乳糖压片测量5次取其平均值作为最终测量结果,测量误差为0.02 mm。小波去噪常用的方法是小波阈值去噪法,这是本文实验采用的对比方法。小波阈值去噪法的原理是将信号通过小波分解后,有用的信号小波系数幅值大,能量大,小波系数数量偏少;噪声信号小波系数幅值小,能量小,小波系数多且分布均匀。小波阈值去噪方法可细分为软阈值和硬阈值两种。李鹏鹏^[16]使用小波分析对太赫兹信号分析得出小波分解的小波基函数、分解层数和阈值选择对去噪效果有影响,其中的最优选择需要依靠经验。本文对乳糖的太赫兹光谱小波分解后,发现使用db2小波基、分解层数为两层时降噪效果比较好。因此,小波降噪法采用的小波基是db2函数,分解层数为两层。小波降噪法采用的软阈值通过无偏风险估计阈值计算得到,硬阈值采用固定阈值公式结合经验分析取得。固定阈值公式为:

$$\lambda = \sqrt{2 \log_{10}(N)} \quad (11)$$

式中: N 表示输入信号的长度。实验过程中,本文使用单一控制法控制各种变量,即尽量保证实验环境不变,测量不同厚度的乳糖压片。不同降噪算法的平均效果对比如表1所示。

表1 不同降噪算法的平均效果

方法	相关系数	信噪比/dB	峰值信噪比/dB
小波软阈值降噪	0.99	52.68	77.50
小波硬阈值降噪	0.99	54.79	79.60
EMD-R/S分析	0.99	56.23	81.04

由表1可以看出,本文提出的方法在信噪比和峰值信噪比的指标上优于小波降噪方法。主要的原因是,小波降噪依靠人工经验选取阈值,阈值选取的效果直接影响降噪算法的效果。而本文提出的方法物理意义清晰,可操作性强,能够很好地保留信号原有的特征信息。

5 结 语

本文提出了采用EMD-R/S分析的方法去除太赫兹时域光谱中的噪声。该方法利用EMD算法自适应分解太赫兹时域光谱信号,然后利用Hurst指数对分

解后的IMF分量进行噪声判断,最后将噪声从原始信号中剔除。该方法属于数据驱动的方法,克服了小波降噪方法依靠人工经验决定降噪效果的弊端,物理意义清晰,可操作性强。这对后期的太赫兹时域光谱应用推广有一定的参考价值。

参 考 文 献

- [1] 刘桂锋,赵红卫,葛敏,等. 太赫兹时域光谱技术在危险品检测方面的应用[J]. 光谱学与光谱分析,2008(5):966-969.
- [2] 唐忠锋,林海涛,陈晓伟,等. 基于太赫兹光谱的氨基酸检测[J]. 光谱学与光谱分析. 2009,29(9):2351-2356.
- [3] 和挺,沈京玲. 太赫兹光谱技术在毒品检测中的应用研究[J]. 光谱学与光谱分析. 2013,33(9):2348-2353.
- [4] 赵树森,陈思嘉,沈京玲. 用支持向量机识别毒品的太赫兹吸收光谱[J]. 中国激光. 2009,36(3):752-757.
- [5] 沈京玲,李宁,逯美红,等. 太赫兹电磁波在毒品检测中的应用[J]. 光学技术. 2006(5):747-749.
- [6] 陈阳,谭佐军,谢静,等. 食品质量与安全中的THz技术应用进展[J]. 食品工业科技. 2014,35(14):49-55.
- [7] 宝日玛,赵昆,滕学明,等. 地沟油的太赫兹波段光谱特性研究[J]. 中国油脂. 2013,38(4):61-65.
- [8] 杨晨,田璐,赵昆. 食用香料的太赫兹时域光谱[J]. 光子学报. 2012,41(5):627-630.
- [9] 张曼,蔡禾,沈京玲. 食品添加剂特丁基对苯二酚的太赫兹光谱及其检测分析[J]. 光谱学与光谱分析. 2011,31(7):1809-1813.
- [10] 徐伟,张帅,王克家. 拉曼光谱预处理中几种小波去噪方法的分析[J]. 应用科技. 2009,36(11):27-31.
- [11] 姚宝岱,王高. 基于小波变换的爆炸物太赫兹光谱降噪分析[J]. 光电技术应用. 2011,26(1):63-65.
- [12] 陈龙旺,孟阔,张岩. 小波变换在太赫兹时域光谱分析中的应用[J]. 光谱学与光谱分析. 2009,29(5):1168-1171.
- [13] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. The Royal Society of London A:Mathematical,Physical and Engineering Sciences, 1998,454:903-995.
- [14] 臧元章,邹德. 太赫兹时域光谱仪中幅值噪声的分析[J]. 太赫兹科学与电子信息学报. 2018,16(2):201-204.
- [15] 黄瑞瑞,赵国忠,刘影,等. 4种乳糖的太赫兹光谱和红外光谱特性研究[J]. 光科学报,2015,35(S2):228-234.
- [16] 李鹏鹏. 基于小波分析的太赫兹光谱信号分析研究[D]. 郑州:河南工业大学,2018.