

文章编号 1004-924X(2006)06-1088-05

应用小波变换实现光谱的噪声去除和基线校正

方勇华,孔超,兰天鸽,熊伟,董大明,李大成
(中国科学院安徽光学精密机械研究所,安徽合肥 230031)

摘要:为消除实测光谱信号中的噪声和基线干扰,给出了一种基于小波变换实现两者同时去除及其参数选择的新方法。该方法通过对光谱信号在小波域内的低频段小波系数置零来实现基线校正,通过对较高频段小波系数阈值处理来实现噪声去除;并利用纯光谱和常见基线、噪声的仿真信号,通过兼顾重构信号整体逼近和特征峰处局部逼近的评估系数 η 来实现小波基、分解层数、阈值估计方式等参数的选取。仿真实验表明,仿真信号采用 sym5、db5 或 db9 等小波基进行 5 次分解,然后低频成分置 0 及所有高频成分利用单层 Heursure 阈值估计算法进行硬阈值处理较为合适。进一步的实验表明,该方法对实测光谱中噪声和基线的消除是行之有效的。

关键词:红外光谱;噪声去除;基线校正;时段;频段

中图分类号:TP911.7 **文献标识码:**A

Denoising and baseline correction of spectrum by wavelet transform

FANG Yong-hua, KONG Chao, LAN Tian-ge, XIONG Wei, DONG Da-ming, LI Da-cheng

(Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, china)

Abstract: A denoising and baseline correction method based on wavelet transform and their parameters choosing method were developed to remove noise and correct baseline in the spectrum effectively. Baseline wander was removed by setting approximation coefficients to zero, and noise was removed by the method of threshold detail coefficients of higher-frequency period in wavelet filed, and with simulated signals of pure spectrum and familiar baseline and noise, the parameters such as wavelet basis, decomposed level, threshold selection rule and so on were chosen by the evaluation coefficient η which considered approaches of integral signal and every characteristic peak simultaneously. The simulated experiment shows that it is more suitable that the simulated signal is decomposed at level 5 on the wavelet basis of sym5, db5, db9 and so on, and then the approximation coefficients are set to zero and threshold detail coefficients by hard threshold with single layer Heursure threshold selection rule. Further experiment shows that this method provides a novel and to correct effective way to remove noise and to correct baseline in the real spectrum.

Key words: infrared spectrum; denoising; baseline correction; period of time; period of frequency

1 引言

近 10 年来,红外光谱鉴别技术在很多领域得到了广泛的应用,包括环境监测、食品科学、制药工业等领域。然而,实测光谱信号常常伴随着高频噪声和低频基线干扰,两者是影响光谱鉴别的主要因素。因此,对实测的光谱信号进行预处理来消除噪声,抑制基线干扰有利于降低光谱鉴别的虚警率。由于光谱的有用信号主要是表现为尖峰和突起的非平稳信号,基于频域的傅里叶变换滤波法存在保护信号的局部性(高频部分)和抑制噪声之间的矛盾,它无法将信号的高频部分和由噪声引起的高频干扰加以区分。

而小波变换(wavelet transform, WT)是近年发展起来的一种基于时频域的信号分析工具,它具有良好的时频局部性、选基灵活性和去相关性等优点,可用于光谱信号的噪声滤波和基线校正等。本文从小波变换的基本理论及小波系数的物理意义探寻入手,提出一种同时消除高频噪声和低频基线干扰的方法。

2 快速离散小波变换及小波系数的直观描述

Mallat 提出的离散小波变换的快速算法为:

$$f(t) = \sum_{k \in z} c_{J,k} \varphi_{J,k}(t) + \sum_{j=1}^J \sum_{k \in z} d_{j,k} \psi_{j,k}(t)$$

小波域是一种时频域,其时间和频率分辨率都不为 0,不能像时域和频域那样称某一时刻和

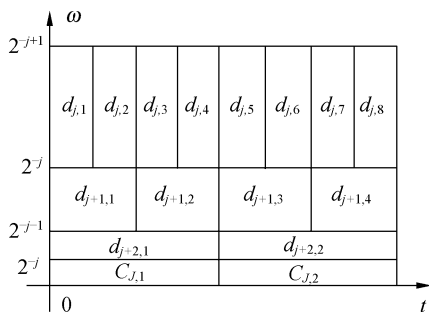


图 1 离散小波变换的时频平面图

Fig. 1 The time-frequency plane of discrete wavelet transform

某一频率。小波基函数将时间轴和频率轴分成若干段,分别称为时段和频段。

$d_{j,k}$ 为 $f(t)$ 在第 j 个频段上的第 k 个时段的分量 ($1 \leq j \leq J$),即:高频部分;而 $c_{J,k}$ 为 $f(t)$ 在第 $J+1$ 个频段上的第 k 个时段的分量,即:低频部分。时段、频段及小波系数在时频平面上的分布如图 1 所示。

3 小波域内的噪声去除和基线校正的基本原理及影响因素分析

小波变换用于基线校正和噪声去除的原理框图如图 2 所示。

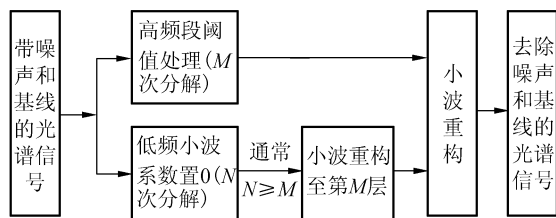


图 2 小波变换用于基线校正和噪声去除的原理框图
Fig. 2 Schematic description of baseline correction and denoising with wavelet transform

基线干扰主要集中在低频段的小波系数中,而光谱信号通常分布在较高频段的小波系数上。两者的区别为分布的频段不同,用低频段小波系数置零的办法来实现分离。在小波域内的同一频段上,由于小波变换的时频局部性,光谱中的尖峰被压缩在少数时段上,对应的小波系数幅值较大;而噪声的小波系数分布在所有时段上,小波系数幅值较小。两者的区别为同一频段不同时段上小波系数幅值的不同,用阈值法来实现分离。为消除伪吉布斯效应,Donoho 提出了平移不变法。

基线校正主要就是信号的分解与重构与低频小波系数置零,则其参数选取主要为小波基和分解层数的选取。而噪声去除主要包括信号的分解与重构和阈值处理两个过程,则其参数选取主要有信号分解时小波基和分解层数的选取以及阈值处理时阈值算法和量化方法的选取。

4 小波域内噪声去除和基线校正的实验

4.1 实验数据集

(1) 光谱信号: 为能对各个参数的效果进行评估, 最好采用不含干扰的纯光谱信号, 然而, 通过实测的办法是很难获得的, 本文采用仿真的纯光谱信号作为标准信号。

(2) 噪声信号: 严格地说, 应先考察光谱中噪声的种类再对噪声仿真, 本文认为高斯白噪声。

(3) 基线信号: 通过 $y = a_n X^n + a_{n-1} X^{n-1} + \dots + a_2 x^2 + a_1 x + a_0$ 来仿真得到, 如图 3 所示。

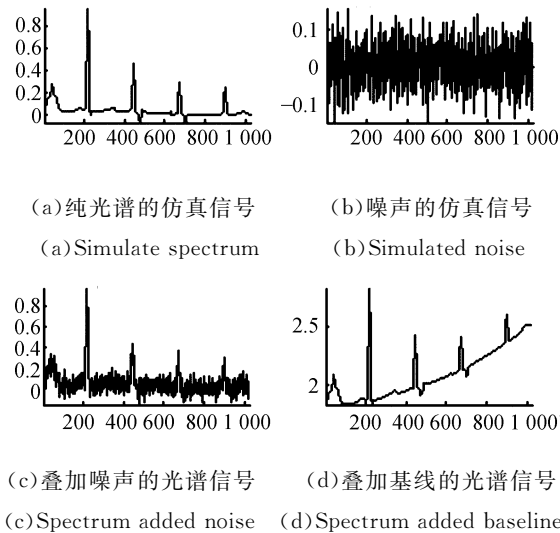


图 3 所有的仿真信号

Fig. 3 All the simulated noise

4.2 光谱信号处理的评估标准

光谱信号处理的目标: 保证重构信号与纯谱整体上的逼近(噪声残留尽量少)及特征峰处的局部逼近(有用信号损失少)。但两者并不统一, 本文提出了兼顾两者的评估系数 $\eta = \text{RMESR} \times \text{RMESRP}$, RMESR 和 RMESRP 分别为重构信号与纯谱在整体上和特征峰附近的相对均方根误差。

4.3 噪声消除和基线校正的参数评估与选择

噪声去除主要是对高频段的处理, 基线校正主要是对低频段的处理, 两者互不影响。

(1) 光谱信号长度 L 的选取

当信号的长度有限时, 噪声的随机性表现为偶然性, 噪声的偶然性会使信号难以恢复。

表 1 不同长度的信号多次去噪后的平均评估系数

Tab. 1 Mean evaluation coefficients of multi-denoising with different lengths

信号长度 L	平均评估系数 η
128	7.1216
256	3.5147
512	2.153
1024	0.7224
2048	0.3965
4096	0.1961

从表 1 可看出, 信号的长度每增长一倍, 去噪效果就改善一倍, 信号的抗噪性能提高了一倍, 这为光谱信号长度的选取提供了理论基础。考虑到计算量, 取 $L=1024$ 较为合理。

(2) 噪声功率估计算法及阈值量化算法的确定

本文讨论的噪声为高斯白噪声, 采用单层噪声功率估计 (sln) 较为合理。当噪声为有色噪声时, 采用多层噪声功率估计 (mln) 会较为有效。另外在光谱鉴别时, 峰值的保持是很重要的, 硬阈值法可以很好地保持信号的尖峰, 则采用硬阈值法。

(3) 噪声消除和基线校正中的分解层数 N 的选取

噪声消除时, N 越大噪声去除越多。基线校正时, N 过大则基线会保留, 过小则信号被削弱。从表 2 可以看出, 当 $N \geq 5$ 时, 去噪效果已经十分稳定了, 取 $N=5$ 。

从表 3 可以看出, 分解层数为 5~8 时都会得到很好的去基线效果, 取 $N=5$ 。

(4) 噪声消除和基线校正中的小波基的选取

噪声消除和基线校正采用相同的小波基, 这样两者的小波分解和重构可以同时进行。合适的小波基能够使信号能量分布在少数的几个小波系数上。

表 2 不同分解层数 N 多次去噪的平均评估系数

Tab. 2 Mean evaluation coefficients of multi-denoising with different decomposed level N

分解的层数 N	平均评估系数 η
1	3.2985
2	1.7224
3	0.9355
4	0.8615
5	0.8245
6	0.8257
7	0.8255
8	0.8256
9	0.8246
10	0.8244

表 3 不同分解层数 N 取基线后的平均评估系数

Tab. 3 Mean evaluation coefficients of baseline correction with different decomposed level N

分解层数 N	平均评估系数 η
1	0.3147
2	0.3128
3	0.2937
4	0.2247
5	0.0276
6	0.0189
7	0.0039
8	0.0024
9	176.00
10	486.8

表 4 不同小波基采用小波阈值法和时移不变法多次去噪的平均评估系数

Tab. 4 Mean evaluation coefficients of multi-denoising with different wavelets using methods of WaveShrink and Translation-invariant

	sym1	sym2	sym3	sym4	sym5	sym6	sym7	sym8	db1	db2	db3
平移不变法	0.7886	0.5798	0.6497	0.6802	0.693	0.7094	0.7151	0.7465	0.7886	0.5798	0.6497
	db4	db5	db6	db7	db8	db9	coif1	coif2	coif3	coif4	coif5
	0.7119	0.7729	0.7801	0.8115	0.8568	0.8077	0.5707	0.6788	0.7052	0.7351	0.7424
	sym1	sym2	sym3	sym4	sym5	sym6	sym7	sym8	db1	db2	db3
小波阈值法	1.6818	1.1908	1.2369	1.5807	1.0551	1.322	1.0871	1.0935	1.6818	1.1908	1.2369
	db4	db5	db6	db7	db8	db9	coif1	coif2	coif3	coif4	coif5
	1.2101	1.0314	1.0645	1.0604	1.3026	1.011	1.537	1.2705	1.0815	1.1858	1.2346

从表 4 可看出,平移不变法的去噪效果优于小波阈值法,并且受小波基的影响较小。当选用时移不变法时, sym2 (db2)、sym3 (db3)、sym4、coif1 都有很好的去噪效果。当选用小波阈值法时,sym5、db5、db6、db7、db9、coif3 有较好的去噪效果。

(5) 阈值估计方法的确定

Minimaxi 和 Rigrsure 的阈值估计比较保守,噪声较小常采用该估计方式,会有效保持信号中的有用部分。而 Heursure 和 Sqtwolog 会更有效的去除噪声。

表 5 不同的阈值估计方式多次去噪的平均评估系数

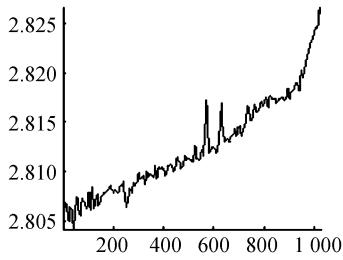
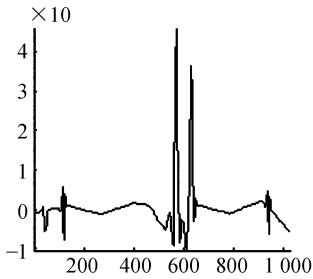
Tab. 5 Mean evaluation coefficients of different threshold evaluation methods

	sym5	sym7	sym8	db5	db6	db7	db9	coif3
Heursure	0.7703	0.7510	0.8043	0.7846	0.8520	0.8478	0.8631	0.7276
Sqtwolog	0.9735	1.0061	1.0106	1.1580	1.2160	1.1848	1.2129	0.9614
Rigrsure	1.2197	1.5581	1.5634	1.072	1.2565	1.2665	1.0855	1.7947
Minimaxi	1.4149	1.4825	1.4570	1.208	1.1152	1.1867	1.1122	1.4925

从表 5 中可以看出,Heursure 和 Sqtwolog 均有较好的去噪效果。

4.4 实测 NH_3 信号的噪声去除和基线校正

小波域内噪声去除和基线校正实验结果如图 6 所示。图 6(a)为 NH_3 的实测光谱信号;图 6(b)为去除噪声和基线校正后的光谱信号。

(a) NH_3 的实测信号(a) Detected spectrum of NH_3 with noise and baseline

(b) 去除噪声和基线后的实测光谱信号

(b) Detected spectrum of NH_3 after denoising and baseline correcton图 6 实测 NH_3 信号噪声去除和基线校正前后的对比Fig. 6 Comparison of the spectrum of NH_3 before and after denoising and baseline correction

5 结 论

本文利用时频平面法对小波系数的物理意义以及时段和频段的概念给出了直观的描述。在此基础上,根据基线主要集中在低频段的小波系数上,信号主要集中在高频段的小波系数上,采用将低频段小波系数置零的办法实现基线校正;根据噪声所在时段的小波系数通常较小,采用小波阈值法或平移不变法来实现噪声去除。

为了给以上过程选出合适的参数,本文首先利用纯光谱信号和常见噪声、基线的仿真信号,来获得带噪声光谱信号及带基线光谱信号,然后对它们进行去基线和去噪处理,通过兼顾信号整体逼近与特征峰处部分逼近的评估系数 η 对去噪和去基线效果进行评估来选出合适的参数。实验表明,仿真信号采用 sym5、db5 或 db9 等小波基进行 5 次分解,然后低频成分置 0 及所有高频成分利用单层 Heursure 阈值估计算法进行硬阈值处理可以较好地抑制噪声和基线干扰。进一步的实验表明,该方法对于实测光谱中噪声和基线的消除是行之有效的。该方法可以推广到类似非平稳信号的去基线和降噪处理。

参考文献:

- [1] 孙延奎. 小波分析与应用[M]. 北京:机械工业出版社,2005.
SUN Y K. *Analysis and application of wavelet* [M]. Beijing: China Machine Press, 2005: 217-244. (in Chinese)
- [2] 谢杰成,张大力,徐文立. 一种小波去噪方法的几点改进[J]. 清华大学学报(自然科学版),2002,42(9):1269-1272.
XIE J CH, ZHANG D L, XU W L. Several improvements for a wavelet denoising method[J]. *Tsinghua Univ. (Sci & Tech)*, 2002, 42 (9): 1269-1272. (in Chinese)
- [3] DONOHO D L, JOHNSTONE I M. Ideal spatial adaptation via wavelet shrinkage [J]. *Biometrika*, 1994, 81: 425-455.
- [4] DONOHO D L, JOHNSTONE I M. Wavelet shrinkage: Asymptopia [J]. *Journal of royal statistics society series (B)*, 1995, 57: 301-369.
- [5] MALLAT S. A theory for multiresolution signal decomposition: the wavelet representation [J]. *IEEE Trans PA-MI*, 1989, 11(7): 674-693.
- [6] DONOHO D L. De-Noising by soft threshold [J]. *IEEE Transactions on Information Theory*, 1995, 141(3): 613-627.
- [7] 赵洁红,孙文军,孟忠,等. CCD 遥感图像的小波分解特性及编码压缩[J]. 光学 精密工程, 2004, 12(2): 205-210.
ZHAO J H, SUN W J, MENG ZH, et al. Wavelet transform characteristics and compression coding of remote sensing images [J]. *Optics and Precision Engineering*, 2004, 12 (2): 205-210. (in Chinese)
- [8] 孙辉. 基于小波变换的图像数据融合方法[J]. 光学 精密工程, 2000, 8(6): 551-553.
SUN H. Data fusion for image based on wavelet transform [J]. *Optics and Precision Engineering*, 2000, 8 (6): 551-553. (in Chinese)

作者简介:方勇华(1966—),男,研究员,博士生导师,主要从事红外高光谱信息处理及应用技术研究。E-main: yhfang@aiofm. ac. cn