DOI: 10.3724/SP. J. 1096.2011.00911

## 基于多尺度小波变换的红外光谱谱峰识别算法

### 蔡 涛<sup>\*</sup> 王先培 杜双育 阳 婕

(武汉大学故障诊断与系统集成实验室,武汉 430079)

摘 要 传统的谱峰检测方法一般分为3个步骤:谱线平滑、基线校正和谱峰识别。现有的基于小波变换的 峰值检测方法能较好地将基线校正和谱峰识别两个步骤融为一步。在此基础之上,本研究将谱线平滑也很好 地融入到小波变换的峰值检测算法中,使整个峰值检测算法成为一个整体。在峰值提取时,原始谱图直接处 理,不再是处理加工过的谱图,减小了谱峰检测结果出错的可能性。另外,对基于小波变换的谱峰检测算法中 模极大值检测算法存在不确定阈值的问题进行了修改,使得基于小波变换的谱峰检测算法更为完善。

关键词 小波变换; 红外光谱; 谱峰检测

#### 1 引 言

传统的谱峰检测算法一般分为 3 个步骤: 谱线平滑、基线校正与谱峰识别。围绕这 3 个步骤,已提 出大量算法。如采用移动平均法进行谱线平滑<sup>[1]</sup>; 采用 Kaiser 滤波器进行谱线平滑<sup>[2,3]</sup>。这两种方法 均是通过将原始信号与平滑函数直接相乘达到平滑谱线的目的,方法简单,但信号失真较大。采用小波 变换法平滑谱线<sup>[4,5]</sup> 由于小波变换系数存在较大的冗余度,所以在实现谱线平滑的同时,能较好地减少信 息失真。在基线校准方面,文献[6]求出原始谱图信号的一阶导数,利用信号的单调性实现校正。文献[3] 采用线性插值法实现校正。文献[7]则采用小波变换实现基线校正,方法简单有效。在谱峰检测方面,基 于信噪比<sup>[2,3,6,7]</sup>、局部极大值<sup>[3,7,8]</sup>、斜率<sup>[9]</sup>、峰面积<sup>[4]</sup>、峰宽<sup>[10]</sup>、光谱变量降维<sup>[11]</sup>和谱峰模型<sup>[12]</sup>的方法都 有报道。但这些算法的成效均在很大程度上依赖于前期的谱线平滑与基线校准。Baggerly等<sup>[13]</sup>通过实验 证实了不同的谱线平滑与基线校准算法在很大程度上决定传统谱峰检测法的结果。文献[7,14]提出了基 于小波变换的峰值检测方法,并在该算法中很好地融合了基于小波变换的基线校准。但该算法在光谱平 滑方面融合的还不够,算法中模极大值的搜索算法存在重复搜索的问题。文献[15]对比了现有诸多谱峰 检测法,得出小波变换法的综合性能是最佳的。本研究参考文献[7,14],并加以改进。将基于小波变换的 谱线平滑也融合到谱峰检测过程中,并在两个不同阶段分别实现一次谱线平滑,使谱峰检测的精确度提 高。改进了模极大值搜索算法,消除了原算法需要设定搜索阀值而带来不确定因素的隐患,并且减小了原 算法的复杂度。通过对 SF。气体红外光谱谱图的分析处理,证实了算法的优越性。

#### 2 谱图的小波变换原理和特征

2.1 基线校准

在进行光谱分析前,有必要对光谱信号进行基线校准。传统的峰值提取方法是将基线校准作为单独步骤。采用小波变换提取峰值则避免了此步骤。假设 *F*(*t*)为原始光谱谱图,*B*(*t*)为基线函数,则谱图中的有用信号函数 *F*(*t*)可表示为:

$$F'(t) = F(t) - B(t)$$
(1)

如果对原始谱图进行小波变换 即:

$$Wf(a \ b) = [F(t) \cdot \Psi(a \ b) dt + [B(t) \cdot \Psi(a \ b) dt$$
(2)

一般情况下 基线是单调的 ,而且变换缓慢 ,所以基线函数 B(t) 可以表示为:

2010-08-19 收稿; 2011-01-21 接受

本文系国家自然科学基金项目(No. 50677047)资助

\* E-mail: cai-tao@ hotmail. com

其中 B(t) 为过零点的斜坡函数 ,C 为 常数。而小波母函数在支撑域的积分 为 0。由此可知 ,式(2) 中后一部分结果 为 0。因此 ,含基线的原始光谱谱图在 进行小波变换后 ,其小波系数中已不包 含基线信息。所以采用小波变换法进 行 谱峰检测时 ,无需进行基线校准。 图 1 为小波变换对基线移除的效果图。 2.2 信号滤波

李氏指数是数学上用来表征函数 局部特征的一种度量,它反映了曲线的 光滑程度。即函数在某一点的李氏指 数表征了该点的奇异性大小,指数越 大,该点的光滑度越高;指数越小,该点 光滑度越低。如果将该理论引入到小 波变换中<sup>[16]</sup>就会发现噪声对应的李氏 指数一般为负数,而有用信号的李氏指 数则为正数。对应到小波系数中就是 噪声产生的模极大值会随着尺度的变 大而减小,有用信号产生的模极大值会



图 1 小波变换对基线的移除: f(t) 为不含基线的原始信号  $W_f$  为尺 度为 6 时 其对应的小波系数; g(t) 为含有基线的原始信号;  $W_g$  为尺 度为 6 时 其对应的小波系数

Fig. 1 Baseline remove: f(t) is original signal ,  $W_f$  is wavelet coefficients of f(t) when the scale is 6, g(t) is the original contained baseline signal ,  $W_g$  is wavelet coefficients of g(t) when scale is 6

随着尺度的变大而变大。利用该规律 即可较好地识别真实信号和具有较高似真度的噪声信号。 2.3 谱峰对应的小波变换特征

Marr 小波正比于高斯函数的二阶导数。令  $\theta(t)$  为高斯函数。  $\Psi(t)$  为 Marr 小波母函数。

$$\Psi(t) \propto \frac{\mathrm{d}^2 \theta(t)}{\mathrm{d}t^2} \tag{4}$$

从另一个角度解释小波变换,即将原信号 f(t) 与伸缩小波卷积得到(此处暂不考虑时间平移)。以  $\Psi(t)$  为小波母函数 则对信号 f(t) 在尺度 a 的小波变换可写成:

$$Wf(t) = f(t) \cdot \Psi(\frac{1}{a}) \propto \left(f \cdot a^2 \frac{\mathrm{d}^2 \theta_a}{\mathrm{d}t^2}\right)(t) = a^2 \frac{\mathrm{d}^2(f \cdot \theta_a)(t)}{\mathrm{d}t^2}$$
(5)

由于 *f* • *θ<sub>a</sub>*(*t*) 可看作是高斯函数在尺度 a 下对信号 *f*(*t*) 进行平滑的结果。而小波变换则 *wf*(*t*) 可 看作信号 *f*(*t*) 在尺度 a 下由高斯函数平滑后再取二阶导数的结果。高斯函数平滑能够去除部分噪声, 而信号 *f*(*t*) 的二阶导数则能凸显出信号的位置与形态的显著变化。图 2 为采用 Marr 小波作为母函数对 信号进行小波运算的结果。在信号发生突变处 A 尺度的小波系数也有相应的模极值。而且相邻尺度的 模极值连起来如同一条山脊 ,并收敛于对应的谱峰位置。此处提到的模极值包括一个局部模极大值和紧 邻区域内左右各一个局部模极小值。利用该性质 ,本研究提出了新的谱峰检测法。文献 [14] 做了类似工 作 ,但所用的小波母函数为 Ridger 小波 Ridger 小波是奇函数 ,无法通过小波变换实现基线校准。

#### 3 多尺度小波变换的计算过程

据小波变换后各尺度小波系数之间的相互关系检测谱峰。具体步骤如下:

(i)对光谱信号进行小波变换,小波母函数选择 Marr 小波,分解尺度为32。但每隔一个尺度计算一次系数。(ii)搜索系数矩阵中的局部模极值。设小波系数矩阵为 *m*×*n* 阶矩阵,其中 *m* 对应分解尺度, *n* 对应光谱分辨率。*coef*(*i j*)表示系数矩阵中*i*行*j*列的系数。则小波系数矩阵中的模极大值应满足以下准则: (6)

 $\begin{aligned} |\operatorname{coef}(i \ j)| &\ge |\operatorname{coef}(i \ j-2)| \\ |\operatorname{coef}(i \ j)| &\ge |\operatorname{coef}(i \ j-1)| \\ |\operatorname{coef}(i \ j)| &\ge |\operatorname{coef}(i \ j+2)| \end{aligned}$ 

 $|coef(ij)| \ge |coef(ij+1)|$ 式(6)定义系数局部模极大值为大于或 等于该极大值前2个系数,且大于或等于 该极大值后2个系数,在寻找小波系数 模极大值的同时,实现去噪。将式(6)中 的大于等于准则改为小于等于准则,可 以寻找出系数矩阵中的局部模极小值。 (iii)寻找由各尺度局部模极大值构成的 山脊。采用文献[7]算法构建模极值序 列,包括一组模极大值序列和两组模极 小值序列。每组模极值序列在每个尺 度只取一个值。且相邻两个尺度的模 极值都应在一个允许的区间  $[n_k - 2^{k-1}, n_k + 2^{k-1}]^{[17]}$ 内寻找,其中  $n_k$ 是尺度 k的小波系数模极大位置。具体过程如

Amplitude(%) 0.5 Intensity 0  $-0.5^{L}_{0}$ |2| 4 6 8 10 0.5 coef (a=2) 0 -0.5 L 2 4 6 10 8 0.5 ſΛ 0 -0.5 L 4 6 10 2 8 0.5 coef (a=6) 0  $-0.5^{L}_{0}$ |2| 4 6 | 8 | 10 0.5 coef (a=8) 0 -0.5 2 4 6 8 10



t/s

Fig. 2 Coefficients matrix of original signal with Marr wavelet. There is a ridge along wave peak which remarked with dotted line

下: 将最大尺度上的第一个模极大值作为山脊的起始值,并记录该极大值对应的时间。为防止下面几个 尺度连续在相应的区域内未出现对应模极大值,定义了空缺参数g,初始值为0。当下个尺度无对应极

大值时 g 值加 1; 若 g > 3, g 则值归 0,判定 此山脊不成立,即该处无谱峰;否则该处可能 对应谱峰。(iv)重复以上步骤,搜索整个谱图, 显示出搜索到的所有谱峰。

#### 4 结果与讨论

选择两种已有的谱峰检测算法与本算法 进行比较。谱图选取实验室在进行 SF<sub>6</sub>气体绝 缘性研究时获取的 SF<sub>6</sub>气体红外光谱谱图,如 图 3 所示,验证了算法的有效性。

图 4 为各种算法峰值提取的效果图。其 中 A 为原始信号 ,D 为采用本算法得到的峰信









# 图 4 A: 原始谱图; B: 基于局部极大值算法的提取结果; C: 基于小波变换算法的提取结果; D: 本算法的提取结果

Fig. 4 A: Original spectrum , B: Result with Local Maximum method , C: Result with wavelet transform method , D: Result with this method

息。B为文献[6]中基于局部极大值提取算法 C为文献[7,14]中的基于小波变换的峰值提取法。

基于局部极大值的峰值提取算法是利用谱图的一阶导数提取峰值,该算法在控制噪声造成的干扰 信息时存在困难,需要依赖前期的谱图去噪与基线校正;并且在平顶谱峰处易造成连续辨认,如图 4B 的 a 处。基于文献[7,14]的小波变换算法能较好处理平顶谱峰,处理干扰信息的效果也优于基于局部极大值 的算法,但该算法不能完全消除噪声影响。由图 4B 和 4C 中的 b 处可见,两图均将峰臂上的噪声信号识 别为谱峰,而图 4D 没有。本算法将基线校准、去噪和峰值提取融为一体,降低了峰值提取算法的复杂 度。由于小波系数模极大值位于信号变化处产生,所以能较好地处理平顶谱峰,如图 4D 中的 a 处。

#### References

- 1 Li X, Gentleman R, Shi Q. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. NewYork: New York Springer. 2005: 91 ~ 109
- 2 Mantini D, Petrucci F, Pieragostino D, Delboccio P. BMC Bioinformatics 2007, 8(3): 101
- 3 Yasui Y , Pepe M , Thompson M L. Biostatistics , 2003 , 4(3): 449 ~463
- 4 Bellew M , Coram M , Fitzgibbon M. Bioinformatics , 2006 , 22(15): 1902 ~ 1909
- 5 Karpievitch Y V, Hill E G, Smolka A J. Bioinformatics, 2007, 23(2): 264~265
- 6 Coombes K R , Tsavachidis S , Morris J S. Proteomics , 2005 , 5(16) : 4107 ~ 4117
- 7 Du P , Kibbe W A , Lin S M. Bioinformatics , 2006 , 22(17): 2059 ~ 2065
- 8 Smith C A , Want E J , Maille G O. Anal. Chem. 2006 , 78(3): 779 ~ 787
- 9 Coombes K R , Fritsche H A , Clarke C. Clinical Chemistry , 2003 , 49(10): 1615 ~ 1623
- 10 Katajamaa M , Miettinen J , Oresic M. Bioinformatics , 2006 , 22(5): 634 ~ 636
- 11 JIN Xiang-Jun, ZHANG Yong, XIE Yun-Fei, CONG Qian, ZHAO Bing(金向军,张勇,谢云飞,丛茜,赵冰). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2009, 29(3): 656~660
- 12 Lange E, Gropl C, Reinert K. In Pac Symp Biocomput Maui., Hawaii, USA, 1996: 243 ~ 254
- 13 Baggerly K A, Morris J S, Coombers K R. Bioinformatics, 2004, 20(5): 777~785
- 14 Wee A, Grayden D B, Zhu Y G. Electrophoresis, 2008, 29(20): 4214 ~ 4225
- 15 Yang C, He Z Y, Yu W C. BMC Bioinformatics, 2009, 10: 4
- 16 Mallat S , Zhong S. IEEE Transactions on Pattern Analysis and Machine Intelligence , 1992 , 14(7): 710 ~ 732
- 17 Tao Wei-Liang, Wang Xian-Pei, LIU Yan, YUAN Lei(陶维亮,王先培,刘艳,袁磊). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2009, 29(5): 1241~1245

## Peak Detection for Infrared Spectrum Based on Continuous Wavelet Transform

CAI $\operatorname{Tao}^*\,$  , WANG Xian-Pei , DU Shuang-Yu , YANG Jie

(Laboratory of System Integrated and Faults Diagnosis, Wuhan University, Wuhan 430079)

Abstract In spectrum analysis, peak detection is an essential step for subsequent analysis. Traditionally, peak detection procedure is divided into three consequent parts: smoothing, baseline correction and peak finding. The existing peak detection method based on continuous wavelet transform can combine the baseline correction and peak finding into one part. It simplified the traditional peak detection procedure, but this method still has two parts. A method based on continuous wavelet transform which finishes the three parts at a time was proposed in this study. The baseline's function of original signal is monotone and linear, so after wavelet transform, there is no information of baseline in the wavelet coefficients. What we need do is deal with the coefficients. First, remove the noise in the coefficients based on Liapunov Exponent. Then, find the ridge mentioned in this study. The position of ridge is the peak's position. The proposed method further simplifies the peak detection procedure.

Keywords Continuous wavelet transform; Infrared spectrum; Peak detection

(Received 19 August 2010; accepted 21 January 2011)