

# 基于 Voigt 函数拟合的拉曼光谱谱峰判别方法

刘铭晖<sup>1,2</sup>, 董作人<sup>1,3\*</sup>, 辛国锋<sup>1</sup>, 孙延光<sup>1</sup>, 戴艳<sup>1,2</sup>, 瞿荣辉<sup>1</sup>, 魏芳<sup>1</sup>

<sup>1</sup>中国科学院上海光学精密机械研究所, 中国科学院空间激光信息传输与探测技术重点实验室, 上海 201800;

<sup>2</sup>中国科学院大学, 北京 100049;

<sup>3</sup>南京简智仪器设备有限公司, 南京 210038

**摘要** 提出了一种新的拉曼光谱谱峰判别算法, 采用 Voigt 函数作为拟合函数, 对小波变换脊线寻峰后得到的谱峰进行拟合处理, 将拟合前后的谱峰相似度作为判别谱峰的依据, 可有效识别与噪声幅值相近的小幅值拉曼信号。通过与已有的两种谱峰判别方法进行对比, 在误判率为 5% 的情况下, 该算法准确率提高了 60%; 在准确率高于 90% 的情况下, 该算法误判率降低了 10%。

**关键词** 光谱学; 拉曼光谱; Voigt 函数拟合; 拉曼谱峰判别; L-M 算法

中图分类号 O433

文献标识码 A

doi: 10.3788/CJL201744.0511003

## Discrimination Method of Raman Spectral Peaks Based on Voigt Function Fitting

Liu Minghui<sup>1,2</sup>, Dong Zuoren<sup>1,3</sup>, Xin Guofeng<sup>1</sup>, Sun Yanguang<sup>1</sup>, Dai Yan<sup>1,2</sup>,  
Qu Ronghui<sup>1</sup>, Wei Fang<sup>1</sup>

<sup>1</sup>Key Laboratory of Space Laser Communication and Detection Technology, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China;

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China;

<sup>3</sup>Nanjing S & S Instruments Co. Ltd, Nanjing 210038, China

**Abstract** A new discrimination algorithm of Raman spectral peaks is proposed. With Voigt function as the fitting function, the spectral peaks obtained after the ridge peak detection based on the wavelet transformation are fitted. The similarity of spectral peaks before and after fitting is taken as the basis for discrimination of spectral peaks. The Raman signal with amplitude comparable to that of noise is effectively discriminated. Compared with those of the existing two discrimination methods of spectral peaks, the true positive rate of the proposed method improves by 60% when the false data rate is 5% and the false data rate of the proposed method decreases by 10% when the true positive rate is higher than 90%.

**Key words** spectroscopy; Raman spectrum; Voigt function fitting; Raman spectral peak discrimination; L-M algorithm

**OCIS codes** 300.6450; 200.4560

## 1 引言

拉曼光谱分析是一种通过测量由分子的振动、转动引起的散射的光谱来研究分子结构的分析方法, 其具有快速、无损、无需样品制备等优点, 被广泛应用于爆炸物违禁品检测、考古和矿石鉴定、生物医学等领域<sup>[1-4]</sup>。通过检测拉曼光谱的峰位、峰宽以及谱峰高度等信息可以对待测物质进行成分的定性和定量分析。

收稿日期: 2016-12-23; 收到修改稿日期: 2017-02-21

基金项目: 上海市自然科学基金(14ZR1445100)

作者简介: 刘铭晖(1991—), 男, 博士研究生, 主要从事拉曼光谱检测方面的研究。E-mail: mhliu@siom.ac.cn

导师简介: 瞿荣辉(1972—), 男, 研究员, 博士生导师, 主要从事光电子器件及光通信方面的研究。

E-mail: rhqu@siom.ac.cn

\* 通信联系人。E-mail: zrdong@siom.ac.cn

因此,谱峰识别的效果会直接影响分析的准确度。光谱信号噪声来源的多样性,如荧光噪声、尖峰噪声、背景噪声,导致存在较高的误判率,尤其是较难实现对小幅值拉曼谱峰信号的判别。

目前,已有很多种寻峰算法,如幅值法、导数法、匹配拟合法、小波变换法<sup>[5]</sup>。幅值法、导数法和匹配拟合法在谱峰判别中,大多以局部谱峰信噪比作为判别依据,这种方法在信噪比高的情况下有明显的效果,而在小幅度峰值信号的情况下,容易出现误判。在运用寻峰算法之前,需要对光谱数据进行去除基线和平滑处理,有时会出现漏峰的情况。Du 等<sup>[6]</sup>提出一种基于连续小波变换的脊线寻峰法以识别质谱的谱峰,采用墨西哥帽小波作为小波变换的母函数,可以在噪声中有效提取质谱峰;通过连续小波多尺度变换,将谱峰数据转换到小波变换空间,利用小波变换空间中的尺度系数和幅度信息作为识别谱峰的依据,从而降低了误判率,提高了准确率。Yang 等<sup>[7]</sup>对比了这些谱峰寻峰算法,认为基于连续小波变换的脊线寻峰能够实现多尺度平滑,可有效去除伪峰,具有最佳的寻峰性能。Zhang 等<sup>[5]</sup>改进了谱峰判别算法,以拉曼谱峰在小波空间的极大值作为判别依据。在准确率为 40%的情况下,误判率由接近 25%降为 5%左右;但是当准确率达到最大值 86%时,误判率超过 80%。可见,该判别条件还有待进一步改进。

拉曼光谱的谱峰本质上是洛伦兹线型,由于仪器影响或样本特性,测量到的拉曼光谱通常是洛伦兹函数与高斯函数的卷积,即为 Voigt 函数<sup>[8]</sup>。目前分析光谱数据时,通常只是简单地把光谱数据看作输入向量,不考虑光谱数据本身的波形结构,也不考虑波形结构与样本待测参数之间固有的物理联系。区别于传统方法,李津蓉等<sup>[9]</sup>提出一种基于 Voigt 峰的快速光谱解析方法,该方法采用具有足够精度的 Voigt 函数的近似公式对光谱数据进行解析,解决了波数漂移、谱峰展宽等因素对测量影响的问题。

目前,基于小波变换的脊线寻峰算法大多采用墨西哥帽函数作为母函数,墨西哥帽函数是高斯函数的二阶导数,有利于识别高斯形状的信号峰,但与其他谱峰判别算法一样,判别具有 Voigt 波形结构的拉曼谱峰时,有可能会将高斯形状的基线起伏误认为谱峰,而将具有 Voigt 波形结构的小信号谱峰判别为噪声信号。

本文采用小波变换的脊线寻峰算法,寻找到光谱中的所有谱峰;采用 Voigt 函数对每一个谱峰进行拟合,将拟合前后的谱峰相似度作为判别谱峰的依据,对所有谱峰进行甄别筛选。对拉曼数据库(RDRS)的数据分析,验证了该方法的有效性。

## 2 原理与方法

### 2.1 基于 Voigt 函数拟合的谱峰判别算法原理

光谱谱峰解析通常采用高斯函数在最小二乘算法意义下进行谱峰拟合<sup>[10]</sup>,但实际拉曼光谱的谱峰形状为 Voigt 函数,这时用高斯函数拟合拉曼光谱数据会出现较大的偏差,而采用 Voigt 函数可以实现误差更小的拟合,如图 1 所示。

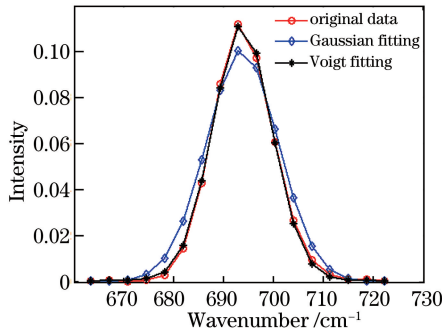


图 1 高斯函数和 Voigt 函数的拟合光谱

Fig. 1 Fitted spectra based on Gaussian function and Voigt function

拉曼光谱的谱峰本质上是洛伦兹线型<sup>[11]</sup>,由于仪器影响或样本特性,测量到的拉曼光谱通常是洛伦兹函数与高斯函数的卷积,即为 Voigt 函数。

标准的高斯函数数学表达式为

$$G(\nu) = \frac{1}{\gamma_G} \left( \frac{\ln 2}{\pi} \right)^{1/2} \exp \left[ - \left( \frac{\nu - \omega}{\gamma_G} \right)^2 \ln 2 \right], \quad (1)$$

式中  $\nu$  为波数,  $\gamma_G$  为高斯线宽,  $\omega$  表示谱峰所对应的波数。Voigt 函数没有解析解, 只能利用数值方法对其进行描述。Wertheim 等<sup>[12]</sup> 提出了 Voigt 函数的近似表示形式:

$$V(\nu, \alpha, \omega, \gamma, \theta) = \theta \alpha \exp \left[ - \frac{4(\nu - \omega)^2 \ln 2}{\gamma^2} \right] + (1 - \theta) \alpha \frac{\gamma^2}{(\nu - \omega)^2 + \gamma^2}, \quad (2)$$

式中  $\alpha$  为谱峰高度;  $\gamma$  为谱峰的半峰全宽;  $\theta$  为高斯-洛伦兹系数,  $0 \leq \theta \leq 1$ 。采用近似 Voigt 函数可以对测量到的拉曼光谱信号进行较准确的解析, 拟合误差与 Voigt 线型面积的比值可小于 0.001%<sup>[9]</sup>。

设  $\alpha$  为拟合后 Voigt 函数的峰高,  $R_{MSE}$  为拟合后的谱峰数据与原始数据间的均方根误差, 定义谱峰的判别参数表达式为

$$S_{Voigt} = \alpha / R_{MSE}。 \quad (3)$$

判别参数  $S_{Voigt}$  的大小不仅取决于谱峰的峰高信息, 还取决于谱峰信号与 Voigt 函数的相似程度。因此, 对于高频噪声或其他因素产生的非 Voigt 线型的伪峰, 其  $R_{MSE}$  较大, 在相同谱峰高度  $\alpha$  的情况下, 判别参数  $S_{Voigt}$  的值也会减小, 从而可以有效地去除伪峰, 提高判别率。

### 2.2 基于 Voigt 函数拟合的拉曼光谱谱峰判别算法计算步骤

基于 Voigt 函数拟合的拉曼光谱谱峰判别参数的计算方法如下:

1) 载入原始光谱数据, 通过广义 Whittaker 平滑算法对原始光谱信号进行平滑处理<sup>[13-14]</sup>, 对平滑后的光谱信号进行强度归一化。可通过调节正则参数来控制平滑, 减少噪声对后续寻峰的影响。

2) 采用基于 Zhang 等<sup>[5]</sup> 描述的小波变换谱峰识别算法, 选取墨西哥帽小波作为连续小波变换母函数, 对平滑后的数据进行小波变换脊线寻峰, 获得某个峰的  $\omega$ 、该峰脊线上连续小波变换后最大值所对应的尺度  $\sigma$ , 及谱峰的半峰全宽  $\gamma$ 。

3) 采用 Voigt 函数, 通过 L-M (Levenberg-Marquardt) 非线性最小二乘算法, 对以  $\omega$  为中心、 $\sigma$  范围内的原始光谱数据进行拟合。

4) 当(3)式计算出的谱峰判别参数  $S_{Voigt}$  小于设定阈值时, 判定为伪峰; 大于设定阈值时, 判定为真峰。

在步骤 3) 中, 为提高计算速度, 在拟合 Voigt 函数之前需要对函数的 4 个参数 ( $\omega, \alpha, \gamma$  及  $\theta$ ) 进行预估。初始参数值与真实值相差越小, 所需要的迭代次数越少, 计算量就越小, 拟合速度就越快。高斯-洛伦兹系数设定参数初始值为 0.5。

图 2 所示为碳酸钡矿拉曼光谱在不同波数下的谱峰及对应的 Voigt 函数拟合光谱。由图可以看出, 利用拉曼谱峰信号可实现对 Voigt 函数较为准确的拟合, 拟合误差较小; 散粒噪声峰虽然信号强度较大, 但是拟合误差较大。因此, Voigt 线型的小幅值拉曼信号谱峰可以得到较高的判别参数值, 这与其他谱峰判别算法的判别依据与谱峰幅值具有强相关性不同。

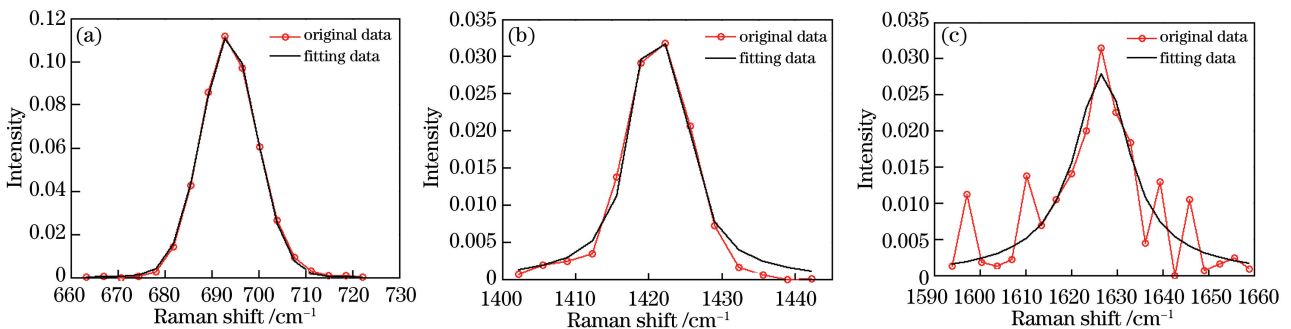


图 2 碳酸钡矿拉曼光谱在不同波数下的谱峰及对应的 Voigt 函数拟合光谱。

(a) 693  $\text{cm}^{-1}$ ; (b) 1423  $\text{cm}^{-1}$ ; (c) 1625  $\text{cm}^{-1}$

Fig. 2 Spectral peaks of witherite Raman spectra at different wavenumbers and corresponding fitted spectra based on Voigt function. (a) 693  $\text{cm}^{-1}$ ; (b) 1423  $\text{cm}^{-1}$ ; (c) 1625  $\text{cm}^{-1}$

Du 等<sup>[6]</sup>提出的质谱小波变换(MSW)算法定义谱峰信号为谱峰脊线上连续小波变换的最大值,谱峰噪声为谱峰附近数据在尺度 1 下的小波变换的绝对值的 95%分位数,信噪比  $R_{MSW}$  即为谱峰信号与噪声的比值。Zhang 等<sup>[5]</sup>提出的多尺度谱峰检测(MSPD)算法定义谱峰阈值  $R_{MSPD}$  为谱峰在小波变换脊线上的最大值。当谱峰阈值大于某一设定值时,判定所识别的谱峰为真峰;否则,判定所识别的谱峰为伪峰。

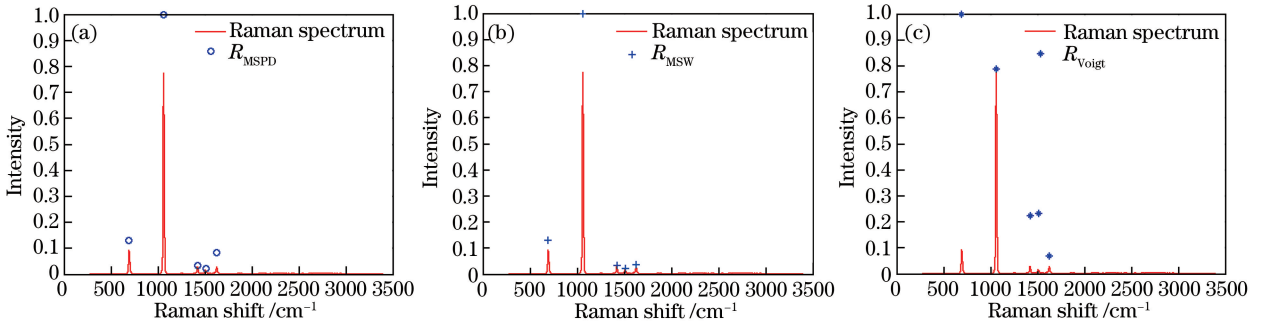


图 3 碳酸钡矿拉曼光谱不同谱峰判别参数。(a)信噪比  $R_{MSPD}$ ; (b)阈值  $R_{MSW}$ ; (c)判别参数  $S_{Voigt}$

Fig. 3 Different spectral peak discrimination parameters of witherite Raman spectrum. (a) Signal-to-noise ratio  $R_{MSPD}$ ; (b) threshold  $R_{MSW}$ ; (c) discrimination parameter  $S_{Voigt}$

单峰判别效果对比图如图 3 所示,对计算结果均采用了强度归一化处理。从图 3(a)、(b)中可以看出,针对已识别出的拉曼谱峰,通过 MSW 与 MSPD 算法计算出的谱峰信噪比或阈值与谱峰的峰值高度具有强相关性,因此很难通过设定合适的阈值对  $1423\text{ cm}^{-1}$  处小信号拉曼谱峰进行有效的判别。阈值设定过高时,无法对小信号拉曼谱峰进行判别,造成漏判;阈值设定过低时,会将  $1625\text{ cm}^{-1}$  处由高频噪声造成的尖峰信号判别为拉曼信号,造成误判。

从图 3(c)可以看出,基于 Voigt 函数拟合的谱峰判别算法采用的谱峰判别参数  $S_{Voigt}$  不仅考虑到峰值高度,还通过对谱峰数据进行 Voigt 函数拟合,衡量所识别谱峰与 Voigt 函数的相似性。这是基于拉曼光谱信号的物理本质对谱峰进行判别。该谱峰判别参数不仅与谱峰高度有关,还取决于谱峰与 Voigt 函数的相似性。谱峰高度越高,对谱峰数据进行 Voigt 函数拟合后的均方根误差越小,计算出的谱峰判别参数值就越高。

因此,小信号拉曼谱峰由于拟合误差小,可以计算得到较高的判别参数值;同时,还能有效地去除高频噪声或其他因素产生的非 Voigt 线型的信号尖峰,提高谱峰判别的准确性。

### 3 实验介绍与结果分析

#### 3.1 拉曼光谱谱峰判别实验

为了验证所提算法的效果,将谱峰判别算法与谱峰信噪比  $R_{MSW}$  判别方法及小波变换阈值  $R_{MSPD}$  判别方法进行比较。采用的原始拉曼光谱来源于网站(<http://rdrs.uaic.ro/>)上罗马尼亚拉曼光谱(RDRS)数据库内的矿石拉曼光谱。

可以通过误判率 ( $F_{DR}$ )和准确率 ( $T_{PR}$ )来衡量判别算法的准确性<sup>[9]</sup>:

$$\begin{cases} T_{PR} = \frac{T_P}{T_P + F_N} = \frac{T_P}{P} \\ F_{DR} = \frac{F_P}{F_P + T_P} \end{cases}, \quad (4)$$

式中  $T_P$  是拉曼光谱中识别到的真实峰的个数,  $F_N$  是未被识别到的真实峰的个数,  $P$  是光谱中真实峰的总数,  $F_P$  是将伪峰误判为真峰的个数。当对某一种谱峰判别算法的有效性进行评估时,在  $T_{PR}$  值相同的情况下,  $F_{DR}$  值小说明其具有较好的判别性能。

随机选取 RDRS 数据库中的 20 张拉曼光谱图,应用同一种小波变换脊线寻峰算法对谱图进行谱峰识别,随后应用三种谱峰判别算法对所识别的谱峰进行判别。MSPD 信噪比判别算法的阈值由 0.001 增加到

5;MSW 算法的阈值由 0.01 增加到 15;基于 Voigt 函数拟合的谱峰判别算法的阈值由 0.01 增加到 15。通过计算并统计三种谱峰判别算法在不同阈值下的  $F_{DR}$  与  $T_{PR}$  值,并绘制相应的  $F_{DR}$ - $T_{PR}$  曲线,如图 4 所示。

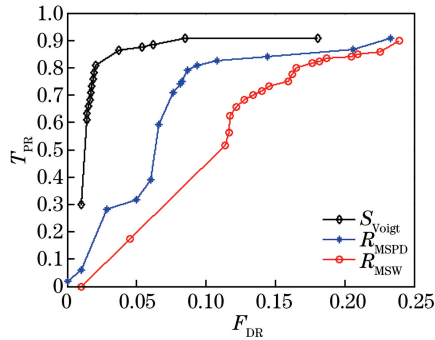


图 4 三种谱峰判别方法的  $F_{DR}$ - $T_{PR}$  对比图

Fig. 4  $F_{DR}$ - $T_{PR}$  comparison of three spectral peak discrimination methods

### 3.2 结果与分析

在拉曼光谱谱峰判别中,小信号的谱峰判别一直是一大难题。由于小信号的拉曼谱峰与由高频噪声或其他原因引起的尖峰的强度相当,很难从谱峰高度上对其进行有效的判别,这严重降低了谱峰的判别率。

从图 4 可以看出,对于同一种寻峰方法识别出的谱峰,采用三种判别算法进行谱峰判别时,在  $T_{PR}$  值相同的情况下,Voigt 函数拟合判别算法相较于另外两种算法始终有更小的  $F_{DR}$  值。当  $T_{PR}$  值达到 90% 时,误判率  $F_{DR}$  仅为 8%,远小于其他两种算法的 21% 和 23%。同时,在误判率  $F_{DR}$  值为 5% 的情况下,准确率达到 89%,远大于其他两种算法的 30% 和 20%。说明相较于另外两种算法,Voigt 函数拟合判别算法在准确率  $T_{PR}$  值相同的情况下,能够实现更小的误判率;同时,在谱峰判别的过程中,能够通过拟合获取谱峰的峰宽等信息。

## 4 结 论

为了实现更加准确的拉曼谱峰判别,提出了针对拉曼谱峰的 Voigt 函数拟合判别算法。应用 L-M 算法对拉曼光谱峰值数据进行近似 Voigt 函数拟合,计算得到谱峰的判别参数。通过与设定阈值进行比较,判别所识别的信号峰是否为真实的拉曼峰。实际拉曼光谱的应用实验结果表明,基于 Voigt 函数拟合的谱峰判别算法能够更加有效地判别小信号拉曼谱峰,去除高频噪声或其他因素产生的非 Voigt 线型的尖峰信号,大大降低了所识别拉曼谱峰的误判率。

### 参 考 文 献

- [1] Moore D S, Scharff R J. Portable Raman explosives detection[J]. Analytical and Bioanalytical Chemistry, 2009, 393(6-7): 1571-1578.
- [2] Li Xia, Huang Meizhen, Song Biao, *et al.* Nondestructive discrimination of Shoushan stone and Laos stone based on Raman spectroscopy and multivariate methods[J]. Acta Optica Sinica, 2016, 36(12): 1230001.  
李 夏, 黄梅珍, 宋 彪, 等. 拉曼光谱结合多变量统计方法在无损区分寿山石和老挝石中的应用[J]. 光学学报, 2016, 36(12): 1230001.
- [3] Culka J, Hyrs J, Jehlicka J. Gem and mineral identification using mineral identification using GL Gem Raman and comparison with other portable instruments[J]. Applied Physics A, 2016, 122(11): 959.
- [4] Xu Bin, Lin Manman, Yao Huilu, *et al.* Measurement of hemoglobin concentration of single red blood cell using Raman spectroscopy[J]. Chinese J Lasers, 2016, 43(01): 0115003.  
徐 斌, 林漫漫, 姚辉璐, 等. 拉曼光谱技术测量单个红细胞的血红蛋白浓度[J]. 中国激光, 2016, 43(01): 0115003.
- [5] Zhang Z M, Xia T, Ying P, *et al.* Multiscale peak detection in wavelet space[J]. Analyst, 2015, 140(23): 7955-7964.
- [6] Pan D, Kibbe W A, Lin S M. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching[J]. Bioinformatics, 2006, 22(17): 2059-2065.

- [7] Yang C, He Z Y, Yu W C. Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis[J]. BMC Bioinformatics, 2009, 10(1): 4.
- [8] Thorne A, Litzen U, Johansson S. Spectrophysics: Principles and applications[M]. Berlin: Springer, 1999.
- [9] Li Jinrong. Mathematical representation of Raman spectra and applications in quantitative analysis[D]. Hangzhou: Zhejiang University, 2013: 21-30.  
李津蓉. 拉曼光谱的数学解析及其在定量分析中的应用[D]. 杭州: 浙江大学, 2013: 21-30.
- [10] Feng Fei, Wang Fubei, Xie Fei, *et al.* Implementation of spectral overlap resolution based on wavelet transforms and Gaussian fitting[J]. Acta Photonica Sinica, 2015, 44(06): 0630001.  
冯 飞, 王府北, 谢 非, 等. 小波变换与高斯拟合在光谱重叠峰解析中的应用[J]. 光子学报, 2015, 44(06): 0630001.
- [11] Prince E, Stalick J K. Accuracy in powder diffraction II[M]. Washington: NIST Special Publication, 1992.
- [12] Wertheim G K, Butler M A, West K W, *et al.* Determination of the Gaussian and Lorentzian content of experimental line shapes[J]. Review of Scientific Instruments, 1974, 45(11): 1369-1371.
- [13] Eilers P H C. A perfect smoother[J]. Analytical Chemistry, 2003, 75(14): 3631-3636.
- [14] Yang Guiyan, Li Lu, Chen He, *et al.* Baseline correction method for Raman spectra based on generalized Whittaker smoother[J]. Chinese J Lasers, 2015, 42(09): 0915003.  
杨桂燕, 李 路, 陈 和, 等. 基于广义 Whittaker 平滑器的拉曼光谱基线校正方法[J]. 中国激光, 2015, 42(09): 0915003.