第38卷，第1期
2018年1月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 38, No. 1, pp93-98
January, 2018

基于匹配追踪的拉曼光谱信号重构算法

王 昕*，何 浩，范贤光，汤 明

厦门大学航空航天学院，福建 厦门 361005

摘 要 拉曼光谱技术是一种高灵敏度、无损伤、振动分子光谱技术，在医药、生物、分析化学等诸多领域有着重要的作用。然而，由于拉曼散射强度低，实际测得的拉曼信号容易被噪声所污染。特别是在较短的曝光时间，收集到的拉曼光谱的信噪比很低。因此，提出了一种基于匹配追踪算法的信号重构方法，用于提取低信噪比的拉曼信号。该方法首先通过阈值循环迭代的方法在平均谱上找出特征峰的位置、估计峰的区间。根据峰的位置区间等信息，用高斯密度函数生成字典。在噪声谱上，根据特征峰位置和区间，将其区分为有信号区间和无信号区间，在有信号区间上利用匹配追踪算法重构被噪声所掩盖的拉曼信号。该算法不仅能够很好的逼近掩盖在噪声中的拉曼信号，且在重构信号的过程中也会对基线进行扣除，无须作基线校正处理。在仿真和实验中对该算法与常规算法进行了比较，结果证明，该算法在低信噪比条件下能够较好的恢复拉曼信号。该算法不同于传统光谱去噪算法，能同时对拉曼光谱进行了基线扣除以及噪声的处理，且能取得较为理想的结果，不需要使用不同的算法对基线和噪声分别处理。其次，在算法上我们创造性地将匹配追踪算法用于拉曼光谱信号的稀疏逼近求解。

关键词 低信噪比；拉曼光谱；匹配追踪；信号重构

中图分类号：O657.3 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2018)01-0093-06

引 言

拉曼光谱技术是一种高灵敏度无损振动分子光谱技术，在医药、生物、分析化学等诸多领域有着重要的作用[1]。但是受困于拉曼散射强度低，实际测得的拉曼信号总是被荧光、射线、CCD暗电流等噪声所掩盖。为了克服噪声的干扰，除了通过提升硬件性能或增加曝光时间外，还可以利用信号处理的办法来提高光谱质量。近年来，国内外学者提出了各种算法用于对拉曼光谱进行噪声去除处理，如：S-G滤波、小波变换(wavelet transform)、经验模态分解EMD(empirical mode decomposition)等。1964年Savitzky等提出S-G滤波，用于数据的平滑处理[2]。其本质是一种基于时域的滑动窗口平滑算法。将长度为2M+1窗口的沿着时间序列滑动，在每个区间上利用最小二乘法拟合出与原始信号误差最小的多项式表示，利用卷积法求解多项式系数。将中间点的横坐标代入该多项式后得到的数据值即为平滑后该点的值。如此遍历整个信号区间即可实现数据的平滑。小波变换阈值去噪是一种时频分析算法，对含噪声的信号进行小波分解，低频部分主要被分解到较高的尺度上，高频部分主要被分解到较低的尺度上[3-4]。一般认为信号主要是低频部分，噪声集中在高频部分。因此利用阈值法剔除部分较低尺度的系数，然后进行小波逆变换即可得到去噪后的信号。EMD算法将信号分解为一系列频率不同的本征模态函数IMF(intrinsic mode function)以及一个趋势项，将这些IMF以及趋势项求和即可获得有用的拉曼信号[4-5]。

而在很多实验中，我们为了观察动态过程，需要采用较短的扫描时间，此时将获得信噪比较低的信号，而现有的算法对这样的信号处理效果并不理想。因此，本文提出了一种适用于低信噪比条件下的信号处理办法。该算法首先在平均谱上标定出特征峰的区间和位置，并据此利用高斯密度函数生成原子库。其次，根据上一步所获得的信息，将噪声谱分为有信号和无信号的区间。在有信号的波段利用匹配追踪算法提取拉曼信号。最后再重构出整个光谱。该算法不仅可以的得到信噪比良好的光谱，而且在重构过程中对基线进行了校正。仿真和实验中，我们分别对比了该算法和小波软阈值、小波硬阈值、EMD、S-G滤波等算法的去噪效果。

收稿日期：2016-11-02，修订日期：2017-05-21

基金项目：国家自然科学基金项目(21503171)资助

作者简介：王 昕，1984年生，厦门大学航空航天学院仪器与电气系副教授 e-mail：xinwang@xmu.edu.cn *通讯联系人

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

1 理 论

1.1 阈值循环迭代法寻峰

传统寻找信号峰值方法，一般都通过求导来实现。但由于噪声的影响，求导法很容易获得许多假峰。文献[6-7]中介绍了利用连续小波变换(CWT)的办法得到小波系数，再通过脊线搜索的办法确定峰位置的算法。但实际使用的效果不佳，甚至对纯净的信号进行处理时，峰位置也会存在偏差。

本文中使用阈值循环迭代的办法交替寻找信号中的极小、极大值。在寻找极小值点时，比较当前点值与该点之前到上一次寻找到的极值点之间所有点的最小值之差。如果该值大于设定的阈值即判断该点的前一个序列点为一个局部极小值点，可用式(1)来表示

$$x(k+1)-\min\{x(p),\cdots,x(k)\}>\Delta \tag{1}$$

其中，$x(i)$为信号序列；$x(p)$为上一个极值点或者起始点，若该式成立，即判定 $x(k)$为一个极小值点。Δ 为预先设定的迭代阈值。

寻找极大值点时，比较当前点值与该点之前到上一次寻找到的极值点之间所有点的最大值之差。可用式(2)来表示

$$\max\{x(p),\cdots,x(k)\}-x(k+1)>\Delta \tag{2}$$

式(2)成立，则判定 $x(k)$为一个极大值点。调整合理的迭代阈值可以准确的找出信号中所有感兴趣的极小和极大值点。

对于拉曼信号而言，所有的极大值点位置即被确认为拉曼特征峰的位置。每个峰的区间宽度的估计可以用式(3)来表示

$$W_i=2\times(P_i-V_i) \tag{3}$$

式(3)中，W_i 为第 i 个峰区间的宽度，P_i 为第 i 个峰的位置，V_i 为该峰位置之前相邻的一个极小值的位置。

1.2 匹配追踪算法

匹配追踪(matching pursuit)是一种适用于信号稀疏逼近的贪婪迭代算法，广泛应用在图像、语音处理，压缩感知等领域[8-9]。所谓稀疏逼近，指在通过在过完备字典内选择尽可能少的原子的线性组合来表示原信号。其收敛法则有稀疏度约束和误差约束法则。对于给定的字典 D，含噪信号 y 的稀疏重构可由式(4)来描述[10-11]

$$\gamma=\arg\min_{\gamma}\|\gamma\|_0 \text{ s.t } \|y-D\gamma\|_2^2\leqslant\varepsilon \tag{4}$$

其中，γ 为信号 y 在字典 D 上的稀疏表示；ε 为收敛误差；$D\gamma$ 为重构信号。

设给定的字典 $D=[d_1,d_2,\cdots,d_m]$包含 m 个原子，当信号 y 为长度 n 的向量时，原子为长度为 n 的单位列向量。设 $\gamma\in R^m$ 为稀疏系数向量，θ 为迭代残余部分。每次迭代更新后的稀疏系数向量 γ 以及相应的残余部分 θ 均加右上标来表示。则用匹配追踪算法重构含噪信号 y 的具体步骤如下：

(1) 设定初始稀疏系数向量 $\gamma^0=[0,0,\cdots,0]^{\mathrm{T}}\in R^m$，$\theta^0=y$，信号 y 可由式(5)表示

$$y=D\gamma^0+\theta^0 \tag{5}$$

(2) 计算第 $p-1$ 次迭代残余部分 θ^{p-1} 与字典 D 中各原子的内积的最大值，并将所得系数加入稀疏系数向量。见式(6)

$$\alpha_i^p=\max_{i\in M}(\theta^{p-1},d_i) \tag{6}$$

其中 i 为所选原子在字典 D 中的索引。

(3) 用更新的稀疏系数向量计算第 P 次迭代的稀疏逼近值 $y^p=D\gamma^p$，计算迭代残余 θ^p，见式(7)

$$\theta^p=\theta^{p-1}-\alpha_i^p\times d_i \tag{7}$$

(4) 根据稀疏度约束条件 $p\leqslant K$ 或误差约束条件($\|\theta^p\|_2^2\leqslant\varepsilon$)，判断是否停止迭代。其中 K 为设定的最大稀疏度，ε 为设定的最大残差。若未达到终止条件，则转到第 2 步，继续搜索匹配的原子。

(5) 迭代终止，获得含噪信号 y 基于字典 D 的稀疏逼近：$\hat{y}=D\gamma$。

1.3 基于匹配追踪的拉曼光谱重构方法

本文提出的算法首先通过循环迭代寻峰的办法在平均谱上找出特征峰的位置和区间。据此，可以将噪声谱分割为有信号的区间和无信号的区间。如式(8)所示

$$\text{Noise}=\sum_i S_i+\sum_j N_j \tag{8}$$

其中，Noise 为噪声信号，S_j 为有信号区间，N_j 为无信号区间。对于无信号区间，其值为 0；对于有信号区间，先用多项式拟合的办法对其进行基线的扣除，然后使用匹配追踪的办法来寻求真实拉曼信号的稀疏逼近。重构后的信号可由式(9)表示。

$$\hat{\text{Noise}}=\sum_i \hat{s}_i+\sum_j 0_j \tag{9}$$

重构信号的准确性在一定程度上取决于字典 D 的设计。字典的来源一般有两个[11]：提前设定和基于学习算法生成字典。本文中字典是根据特征峰区间以及位置信息来设计的。单个拉曼峰的分布可以近似认为正态分布[12]，因此本文中采用高斯密度函数来设计原子，如式(10)所示

$$d(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{10}$$

其中 μ 为峰位置 σ 为峰宽因子。

至此，可以将利用匹配追踪算法重构拉曼光谱信号的具体步骤描述如下：

(1) 利用阈值循环迭代的方法在平均谱上找出特征峰的位置和区间等信息。

(2) 利用特征峰的位置和区间等信息将待处理的噪声谱划分区间。

(3) 对每个有信号区间光谱分别进行多项式基线校正，将光谱无信号区间均置 0。

(4) 利用匹配追踪算法，寻求每个有信号区间光谱的稀疏逼近解。

(5) 合成重构后的拉曼光谱信号。

2 仿 真

本文中使用高斯函数的线性叠加来仿真纯净的拉曼光谱信号，并且使用余弦函数作为基函数来仿真真实平均谱中的基线背景，见式(11)

$$\text{Baseline}(x)=20\cos\left(\frac{x}{1\,000}\right) \tag{11}$$

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

为模拟较弱的真实平均谱信号，文中使用了较低的信号强度，单个峰的信号强度均低于 100。仿真信号的相关参数如表 1 所示。

图 1(a)为仿真的纯净拉曼信号，图 1(b)为加入基线背景的仿真平均谱信号，图 1(c)为加入 30 db 高斯白噪声后的仿真低信噪比拉曼信号，图 1(d)为使用基于匹配追踪算法的重构信号。

表 2 所示为，使用循环阈值算法对图 1(b)中的仿真平均拉曼光谱的寻峰结果。

表 1　仿真信号参数表

Table 1　Parameters of simulated pure Raman signal

Index	Peak position	Amplitude	Peak width factor
1	200	20	10
2	500	25	12
3	640	20	20
4	900	80	12
5	1 300	30	12
6	1 600	25	10

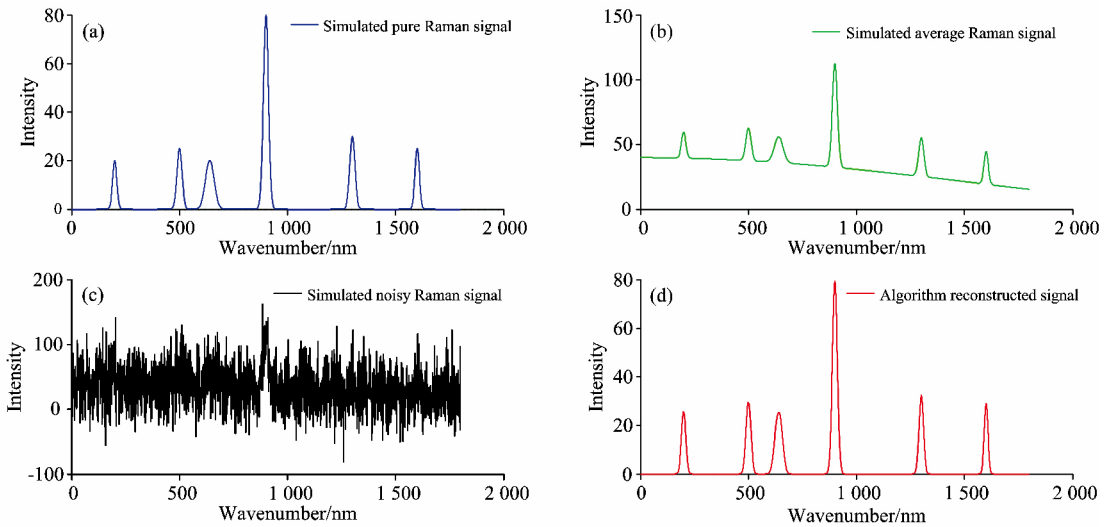

图 1　(a)仿真信号；(b) 仿真平均拉曼信号；(c) 仿真带噪拉曼信号；(d) 去噪后信号

Fig. 1　(a) Simulated pure Raman signal; (b) Simulated average Rman signal; (c) Simulated noisy Raman signal; (d) Denoised signal by the proposed algorithm

表 2　本方法获得的谱峰信息

Table 2　Peak information obtained by the proposed algorithm

Index	Peak position	Peak range
1	200	160～240
2	500	456～544
3	640	572～708
4	900	854～946
5	1 300	1 258～1 342
6	1 600	1 565～1 635

结合表 1 和表 2，以及图 1(a)和(d)，可以看出，由本文提出的算法对低信噪比含噪仿真信号去噪处理后的拉曼光谱信号峰位置以及峰区间均十分准确。结合图 1(a)和(d)来看：重构信号的前 3 个峰的强度略有误差，但均在可接受误差范围内；后 3 个峰强度均与图 1 (a) 较完美契合。为了比较本文算法与常规算法的去噪效果，本文分别使用了小波软阈值、小波硬阈值、S-G 滤波以及 EMD 四种算法对图 1(c)中的含噪声拉曼信号进行了处理，结果见图 2。

图 2(a)为使用小波软阈值处理图 1(c)中的噪声信号的结果，图 2(b)为使用小波硬阈值去噪后的结果，图 2(c)为使用 S-G 滤波去噪后的结果，图 2(d)为使用 EMD 算法去噪后的结果。结合图 1 和图 2 可以看出，以上 4 种算法对于信噪比较低的弱拉曼信号均不能达到很好的去噪结果，处理后的拉曼光谱仍然存在很大噪声。

为了测试算法的去噪能力，本文分别对图 1(b)中信号加入 5～60 db 高斯噪声。然后使用本文算法以及上述 4 种常规算法对噪声谱进行处理，并分别计算了去噪后的信号的均方根误差(RMSE) 以及其信噪比(SNR)。RMSE 与 SNR 的计算公式分别如式(12)、式(13)所示

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{s}(i)-s(i))^2} \tag{12}$$

$$\mathrm{SNR}=10\times\log_{10}\frac{\frac{1}{N}\sum_{i=1}^{N}(\hat{s}(i))^2}{\frac{1}{N}\sum_{i=1}^{N}(\hat{s}(i)-s(i))^2} \tag{13}$$

其中 $\hat{s}$ 为去噪后的信号，s 为原始信号。计算结果最终形成如图 3 所示各算法去噪处理后对应的噪声-均方根误差对应曲线，以及图 4 所示各种算法去噪处理后对应的噪声-信噪比曲线。

在加入 30 db 噪声后，使用本文算法处理后的光谱信噪比可以达到9. 74 db，而使用常规算法去噪后的光谱信噪比

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

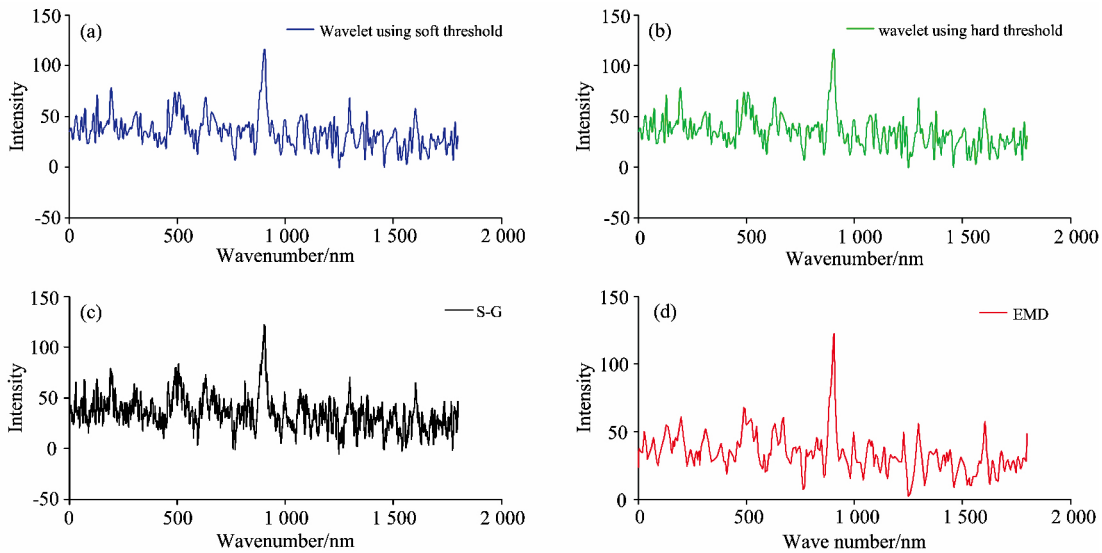

图 2 (a)小波软阈值法获得的去噪信号；(b) 小波硬阈值法获得的去噪信号；(c) S-G 滤波法获得的去噪信号；(d) EMD 获得的去噪信号

Fig. 2 (a) Denoised signal obtained by wavelet using soft threshold; (b) Denoised signal obtained by wavelet using hard threshold; (c) Denoised signal obtained by S-G filter; (d) Denoised signal obtained by EMD

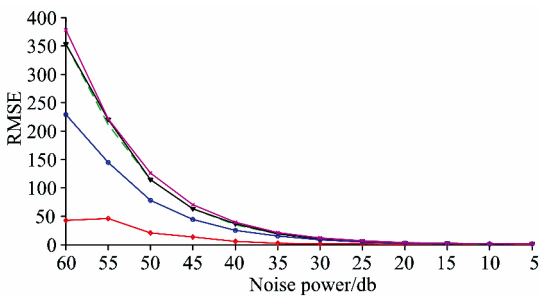

图 3 不同方法的 RMSE-Noise 曲线

Fig. 3 RMSE-Noise curves of denoised signal with different algorithms

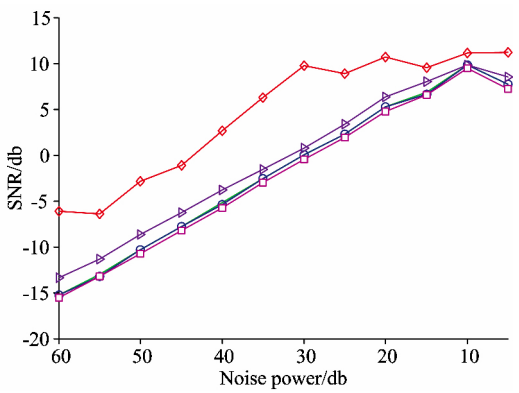

图 4 不同方法的 SNR-Noise 曲线

Fig. 4 SNR-Noise curves of denoised signal with different algorithms

则下降到 0 db 附近，表明此时算法失效。结合图 3 和图 4 可以看出，基于匹配追踪的拉曼光谱重构算法具有更好的去噪能力，EMD 算法次之，小波变换软阈值和硬阈值算法性能相差无几，S-G 滤波去噪能力最差。

3 实验部分

3.1 材料及仪器

本文实验仪器为 nanophoton 公司生产的第三代显微拉曼成像系统 Raman-11，该仪器具有较好的拉曼成像性能。本文实验样品为头孢呋辛酯片粉末，采用线激光作为激发光，每次扫描曝光时间为 0.5 s。

3.2 结果分析

由于曝光时间短，激发光能量低以及荧光背景等原因，单个拉曼光谱的噪声很大。为验证算法的性能，分别使用了本文提出的算法以及常规算法对实测的头孢呋辛酯片光谱进行了去噪处理，实验结果如图 5 和图 6 所示。

图 5(a)所示为头孢呋辛酯的平均拉曼光谱信号。图 5(b)所示为选取的扫描区域内某一点的光谱信号，拉曼光谱信号几乎均被噪声淹没。图 5(c)所示为使用本文算法重构后的信号。参照图 5(a)所示平均谱信号可以看出，重构后的信号不论在谱峰位置上还是峰区间上，均与平均谱信号较为吻合，且没有基线的影响无须进一步处理。表明算法较好的提取了高噪声背景中的弱拉曼信号。

图 6(a)所示为采用小波软阈值对图 5(b)所示的含噪光谱信号去噪后的结果；图 6(b)所示为采用小波硬阈值去噪后的结果；图 6(c)所示为采用 S-G 滤波去噪后的结果；图 6(d)所示为采用 EMD 去噪后的结果。对比图 5(a)所示平均谱信号，可以发现，以上四种常规算法在处理低信噪比的拉曼光谱信号时均不理想。最后，为了衡量计算速度，在相同的软硬件环境下(软件：Matlab R2014b，硬件：Intel Core M-5Y10C，

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

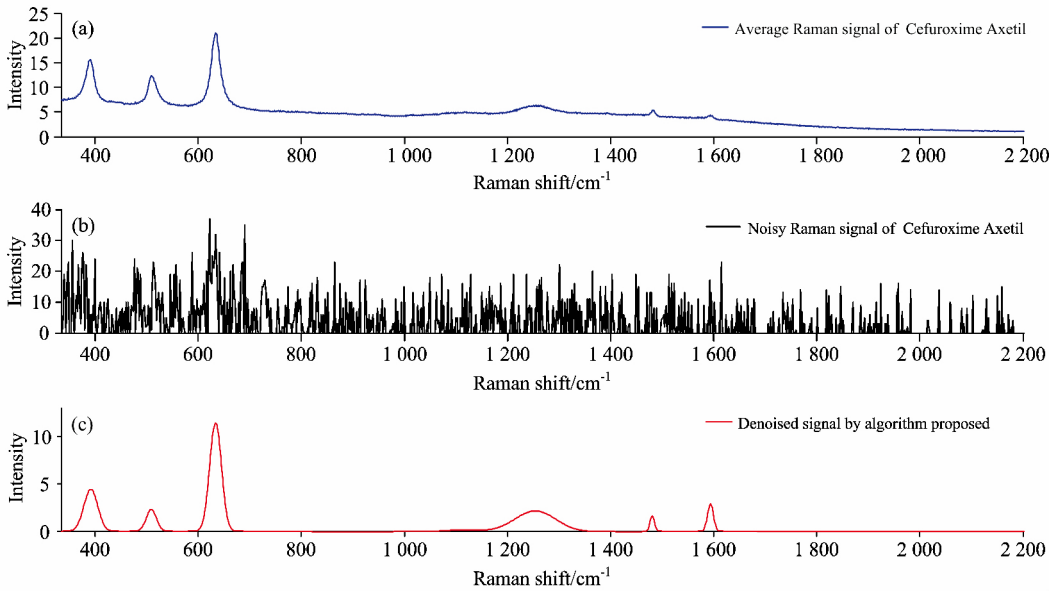

图 5　(a) 头孢的平均拉曼信号；(b) 头孢的带噪拉曼信号；(c) 本方法获得的去噪信号

Fig. 5　(a) Average Raman signal of cefuroxime axetil; (b) Noisy Raman signal of cefuroxime axetil; (c) Denoised signal of the Noisy signal by algorithm proposed

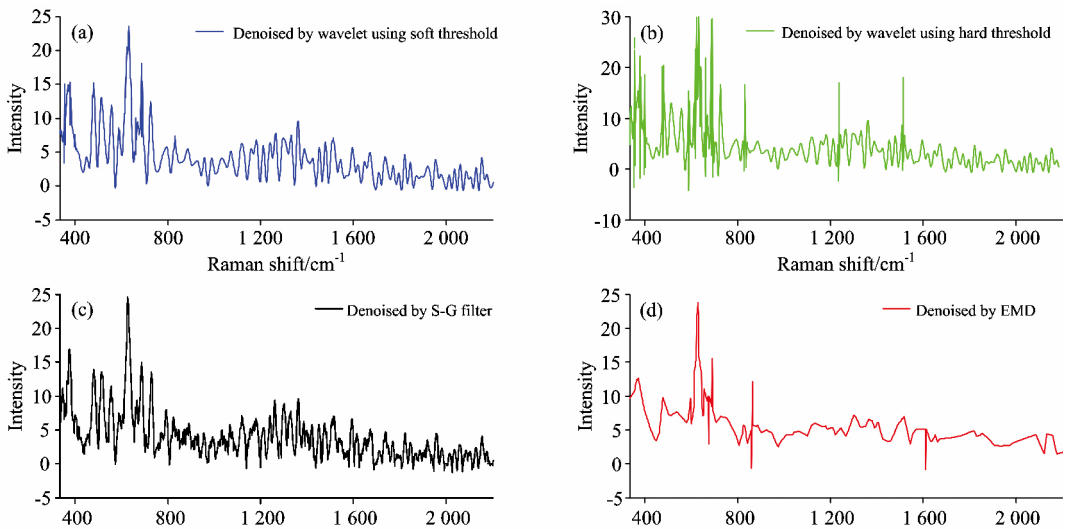

图 6　头孢的拉曼信号

(a)：小波软阈值法的去噪结果；(b)：小波硬阈值法的去噪结果；(c)：S-G 滤波法的去噪结果；(d)：EMD 算法去噪结果

Fig. 6　Raman signal of cefuroxime axetil

(a): Denoised by wavelet using soft threshold; (b): Denoised by wavelet using hard threshold; (c): Denoised by S-G filter; (d): Denoised by EMD

表 3　不同方法的运行时间

Table 3　Processing times in the experiment of different algorithms

Algorithm	Processing time/s
S-G	0.001
Wavelet using soft threshold	0.012
Wavelet using hard threshold	0.010
EMD	0.017
Algorithm proposed	0.034

4GB RAM (DDR3L)，给出了以上各种算法处理头孢呋辛酯片拉曼光谱计算时间，如表 3 所示。由于本文算法同时对光谱的基线和噪声进行了处理，需要相对较长的处理时间，但仍保持在 0.05s 以内，完全可以满足实际用户需求。

4　结　论

提出了一种基于匹配追踪的拉曼光谱的信号重构算法，用于低信噪比的拉曼光谱的信号处理。本文通过仿真与实验

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

吩该算法的性能进行了验证，并且与常规算法进行了比较。事实说明，使用基于匹配追踪的拉曼光谱重构算法处理的低信噪比拉曼光谱信号不仅在特征峰位置还是在峰区间上都较为准确且峰强度合理，同时没有基线的影响无须进一步处理；而常规算法在处理低信噪比拉曼信号时均不理想。该算法为处理极低信噪比的拉曼光谱信号提供了一个潜在的强有力的工具。

References

[1] HU Xiao-hong, ZHOU Jin-chi(**胡晓红，周金池**). Analytical Instrumentation(**分析仪器**), 2011, 6: 1.
[2] Daniel Suescún-Díaz, Héctor F Bonilla-Londoño, et al. Journal of Nuclear Science and Technology, 2016, 53(7): 944.
[3] Galvao R K H, et al. Analytica Chimica Acta, 2007, 581: 159.
[4] Zimon M J, et al. J. Comput. Phys., 2016, 312(15): 380.
[5] Chandra S, Hayshibe M, Thondyath A. Biomedical Signal Processing and Control, 2017, 31: 339.
[6] Zhang Zhimin, et al. J. Raman Spectrosc., 2010, 41: 659.
[7] Wang Xin, et al. Meas. Sci. Thchnol., 2015, 26: 115503.
[8] Liu Qianshun, Bai Jian, Yu Feihong. Applied Optics, 2014, 53(32): 7796.
[9] LIU Ya-xin, et al(**刘亚新，等**). Journal of Electronics & Information Technology(**电子与信息学报**), 2010, 32(11): 2713.
[10] Wright J, et al. IEEE Tansactions on Pattern Analysis and Machine Intelligentce, 2009, 31(2): 210.
[11] Rubinstein R, Peleg T, Elad M. IEEE Tansactions on Signal Processing, 2013, 61(3): 661.
[12] JIANG Cheng-zhi, SUN Qiang, LIU Ying, et al(**姜承志，孙　强，刘　英，等**). Acta Optica Sinca, 2014, 34(6).

Signal Processing Method for Raman Spectra Based on Matching Pursuit

WANG Xin*, HE Hao, FAN Xian-guang, TANG Ming
School of Aerospace Engineering, Xiamen University, Xiamen 361005, China

Abstract Raman spectroscopy, as a high sensitive, non-invasive vibrational molecular spectrocopy technique, plays a significant role in many fields such as pharmaceutical, biology and analytical chemistry etc. However, due to the weak Raman scattering intensity, the measured Raman signal is always contaminated by noise. Especially in the short exposure time, the SNR (signal to noise ratio) of collected Raman spectra is extremely low. Therefore, this paper proposed a signal reconstruction method based on matching pursuit algorithm, which is used to extract Raman signals from the low SNR Raman spectra. The method first finds the position of the characteristic peak on the average spectrum by threshold iterative method, and estimates the interval of the peak according to the location of the peak and peak interval, with a Gaussian density function to generate a dictionary. In the noise spectrum, according to the position and interval of the characteristic peak, it is divided into the signal interval and the non-signal interval. On the signal interval, the matching pursuit algorithm is used to reconstruct the Raman signal covered by noise. The algorithm not only can primely approximate the Raman signal which is covered in the noise, but also deducts the baseline in the procession of reconstructing the signal, and does not need any baseline correction further. The performances of the proposed algorithm and conventional algorithms were compared. The results show that the proposed algorithm can recover the Raman signals in the condition of low SNR. Different with the conventional de-noise algorithms, algorithm of this paper process the baselines and the random noises in Raman signals simultaneously, and the results have been proved good. So there is no need to use different algorithms to process the baselines and noises separately. Furthermore, in the aspect of algorithm, we creatively applied the matching pursuit algorithm to solve the sparse approximation of Raman signals.

Keywords Low SNR; Raman spectra; Matching pursuit; Signal reconstruction

(Received Nov. 2, 2016; accepted May 21, 2017)

* Corresponding author

?1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net