第 44 卷 第 3 期/2024 年 2 月/光学学报 研究论文

光学学报

改进对称零面积变换寻峰算法在拉曼光谱中的应用

王海，黄宁*，何泽，王鹏，袁靖茜

四川大学原子核科学技术研究所辐射物理及技术教育部重点实验室，四川 成都 610064

摘要 拉曼光谱法是一种高效、无损的化学信息获取分析方法。拉曼光谱中的特征峰位包含了物质的化学信息。对称零面积变换寻峰法是一种常用的寻峰方法，但在寻峰前需要输入有关谱线的各项参数，如窗宽、洛伦兹函数半峰全宽、高斯函数半峰全宽等等。对于不同的拉曼光谱，需要输入的这些参数可能不同。如果输入参数与当前拉曼光谱不符合，寻得的峰位可能不准确。本文对对称零面积法进行改进，将拉曼光谱谱峰的半峰全宽归一化，减少了需要输入的各项参数，并结合 Whittaker Smoother 去噪算法和非对称加权惩罚最小二乘（arPLS）去基线算法，形成 WALPSZ 寻峰算法。该算法提高了对不同分辨率的拉曼光谱寻峰的准确性和普适性。将该算法应用于 Raman Open Database 和实际测量的光谱数据的寻峰中，并将获取的峰位与相关文献的数据进行对比，验证了其可靠性和对不同拉曼光谱数据的适用性。

关键词 光谱学；拉曼光谱；自动寻峰；对称零面积变换

中图分类号 O433 **文献标志码** A **DOI**：10.3788/AOS231562

1 引 言

拉曼光谱法是一种高效、无损的化学信息获取分析方法。由于不需要破坏样品，拉曼光谱法被广泛应用于医学、材料、生物和考古等领域[1-5]。当光与物质相互作用时，光子可能被吸收或散射。当光子的能量对应于分子基态与激发态之间的能级差时，光子可能会被分子吸收；也有可能与分子相互作用时被散射出去。前后光子的能量差与分子振动能级差相对应，可以通过分析这些信息来获取分子振动能级[6]。这些信息就像是特定分子的“指纹”，具有独特性。通过拉曼位移获取拉曼峰可获得样品的化学组成；通过拉曼峰相对强度可获得物质相对含量。

在分析拉曼光谱时，需要对原始光谱进行预处理，包括去噪声、去基线和寻峰。之后，将寻得的峰位与数据库进行对比即可获取待测样品的化学组成。噪声可能会导致拉曼谱寻得假峰，影响拉曼分析的准确率，因此去噪声是处理拉曼光谱数据不可或缺的一步。拉曼分析某些样品时，可能会产生荧光背景，这种荧光有时比拉曼散射强几个数量级。关于扣除拉曼光谱背景（去基线）的方法近年来依然在不断发展[7-8]，非对称最小二乘拟合法由于可以自由调节拟合出来的基线偏移量，获得最优的光谱基线校正结果，在拉曼光谱去基线处理中被广泛使用[9]。经过去噪声和去基线的处理，可以通过寻峰获得光谱中的化学信息。常见的寻峰方法包括高斯乘积函数找峰法[10]、导数法[11]、协方差找峰法、连续小波变换法[12]和对称零面积变换法[13]。对称零面积变换法在自动寻峰中应用广泛，因为它具有弱峰识别、重峰识别和抑制高基底能力的优势。

拉曼开放数据库（Raman Open Database，ROD）是由 SOLSA H2020 项目开发的拉曼开放数据库[14]。里面有超过 1000 个高质量的拉曼光谱原始数据，这些数据来自各种拉曼光谱仪，使用了各种激发光源。想直接使用这些数据比较困难，如果对其中的原始光谱数据进行预处理以及寻峰算法得到相应的峰位以及峰强度的信息后，就能更好、更方便地使用它。对称零面积变换法虽然在自动寻峰上具有优势，也能获取谱峰对应的强度信息，但该寻峰算法需要输入与光谱数据有关的各项参数，如窗宽、洛伦兹函数半峰全宽、高斯函数半峰全宽等，因此对称零面积变换法在处理不同分辨率的拉曼光谱时，其普适性相对有限。本文对对称零面积变换法进行了一定的修正和改进，减少与光谱数据相关参数的输入，以适应不同谱峰宽度的数据。

收稿日期：2023-09-18；**修回日期**：2023-11-02；**录用日期**：2023-11-13；**网络首发日期**：2023-11-23

基金项目：四川省重大科技专项（2020ZDZX0004）

通信作者：*huang_ning@scu.edu.cn

2 算法与流程

2.1 基于 Whittaker Smoother 的去噪声算法

拉曼光谱的噪声主要是高斯噪声。可以使用常规滤波方法，如傅里叶变换频域滤波和小波变换[15]，去除这种高斯噪声。但是这些方法本质上是在频域上对噪声进行去除，在光谱数据较多时，这些方法会非常耗时。Eilers[16]提出了一种被称为“Whittaker Smoother”的高效、快速、简单的去噪声方法。该算法可以很快进行谱光滑，而且不会造成峰的偏移。此算法核心是在数据的失真度与粗糙程度取得一个平衡。当需要将一个有噪声的谱线 y 拟合到相对没有噪声的谱线 z 时，需要考虑两个因素：一是数据的失真度；二是数据的粗糙程度。需要在保证数据尽量不失真的情况下降低数据的噪声。然而当 z 越平滑时，它就越偏离 y，失真度会增加。谱线的粗糙程度可以通过 $\Delta z_i = z_i - z_{i-1}$ 来反映，失真程度可以通过 $y_i - z_i$ 来反映。设粗糙度 $R = \sum_i (\Delta z_i)^2$，谱线的失真度 $S = \sum_i (y_i - z_i)^2$，两个目标的平衡组合是 $Q = S + \lambda R$。其中 λ 是平衡因子，λ 越大，R 对 Q 的影响越大，当 Q 为最小值时，获得光滑后的谱线 z，如下式所示：

$$z = \arg\min_z \left[\lambda \cdot \sum_i (\Delta z_i)^2 + \sum_i (y_i - z_i)^2 \right]。 \tag{1}$$

为方便计算，引入矩阵的算法，向量 $\boldsymbol{Z}$ 为谱线 z 的数据，向量 $\boldsymbol{Y}$ 为原始谱 y 的数据。则公式描述为

$$\boldsymbol{Z} = \arg\min_{\boldsymbol{Z}} \left(\lambda \cdot |\boldsymbol{DZ}|^2 + |\boldsymbol{Y} - \boldsymbol{Z}|^2 \right), \tag{2}$$

式中，$\boldsymbol{D}$ 是差分矩阵，假设矩阵维度 $m=3$，则 $\boldsymbol{D}$ 为

$$\boldsymbol{D} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}。 \tag{3}$$

故设 $\boldsymbol{Q} = \lambda \cdot |\boldsymbol{DZ}|^2 + |\boldsymbol{Y} - \boldsymbol{Z}|^2$，则当 $\boldsymbol{Q}$ 最小时有

$$\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{Z}^{\mathrm{T}}} = 2\lambda \cdot \boldsymbol{D}^{\mathrm{T}}\boldsymbol{DZ} - 2(\boldsymbol{Y} - \boldsymbol{Z}) = 0, \tag{4}$$

$$\boldsymbol{Y} = (\lambda \cdot \boldsymbol{D}^{\mathrm{T}}\boldsymbol{D} + \boldsymbol{E})\boldsymbol{Z}。 \tag{5}$$

图 1 为在 ROD 中方沸石（Analcime）的原始拉曼光谱数据中加入高斯白噪声之后，使用 Whittaker Smoother 得到的去噪声效果（$\lambda = 4$）。可以看出 Whittaker Smoother 在保持峰位不变的情况下，有效抑制了噪声。

2.2 基于非对称加权惩罚最小二乘（arPLS）去基线算法

非对称最小二乘（AsLS）去基线算法是扣除拉曼光谱荧光背景的一个常用算法。基于上述的去噪声算法进行改进，得到非对称最小二乘去基线算法的核心[17]。引入权值因子 w_i，则粗糙度与失真度的平衡组合 Q 为

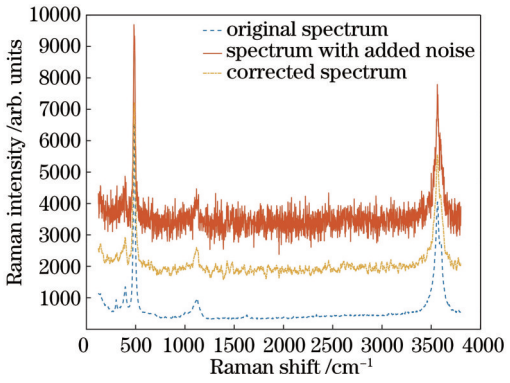

图 1 对加入了高斯白噪声的方沸石拉曼光谱数据使用 Whittaker Smoother 去噪

Fig. 1 Using Whittaker Smoother for denoising Raman spectroscopy data of Analcime with Gaussian white noise

$$Q = \lambda \cdot \sum_i (\Delta z_i)^2 + \sum_i w_i \cdot (y_i - z_i)^2, \tag{6}$$

式中：当 $y_i > z_i$ 时，w_i 为 p；当 $y_i \leqslant z_i$ 时，w_i 为 1-p。其中 p 为不对称修正权值。设 $\boldsymbol{W}$ 为 $\boldsymbol{w}$ 的对角矩阵，则式（6）可以改为

$$\boldsymbol{Q} = \lambda \cdot \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{DZ} + (\boldsymbol{Y} - \boldsymbol{Z})^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{Z})。 \tag{7}$$

通过 $\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{Z}^{\mathrm{T}}} = 0$ 获得其最小值，得到：

$$\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{Z}^{\mathrm{T}}} = 2\lambda \cdot \boldsymbol{D}^{\mathrm{T}}\boldsymbol{DZ} - 2\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{Z}) = 0, \tag{8}$$

$$(\boldsymbol{W} + \lambda \boldsymbol{D}^{\mathrm{T}}\boldsymbol{D})\boldsymbol{Z} = \boldsymbol{WY}。 \tag{9}$$

由于 $\boldsymbol{W}$ 与 $\boldsymbol{Y}$ 和 $\boldsymbol{Z}$ 相关，无法直接求出 $\boldsymbol{Z}$。需要给定初始的 λ 和 p 之后将 $\boldsymbol{W}$ 设为单位矩阵。然后通过式（9）算出 $\boldsymbol{Z}$，并使用新的 $\boldsymbol{W}$ 重复此过程，直到 $\boldsymbol{W}$ 不再变化或超过迭代限制次数。Baek[18]提出了自动修正权值的方法，无需设置 p，并解决了无峰时基线被高估以及有峰时基线可能被低估的问题。这种方法被称为 arPLS（asymmetrically reweighted penalized least squares）。Baek 等[18]使用了一种部分均衡的加权方案。在没有峰值的基线区域，假设噪声在基线以下和基线以上的分布是相等的，如下式所示：

$$w_i = \begin{cases} \dfrac{1}{1 + \exp\left\{ \dfrac{2\left[d_i - (-m + 2\sigma) \right]}{\sigma} \right\}} & y_i > z_i \\ 1 & y_i \leqslant z_i \end{cases}, \tag{10}$$

式中：$d_i = y_i - z_i$；m 和 σ 分别为期望和标准差。将式（10）代入式（6）～（9）即可得出基线。图 2 为对 ROD 中方沸石（Analcime）的原始拉曼光谱数据使用 arPLS 去基线后的结果，可以看出 arPLS 有效抑制了基线。

2.3 基于 Vogit 函数的少参数对称零面积寻峰算法

对称零面积变换法的基本思想是用面积为零的对称函数与光谱数据进行卷积变换，除了存在峰的地方以外，其他线性基底的卷积变换将为零，这种面积为零

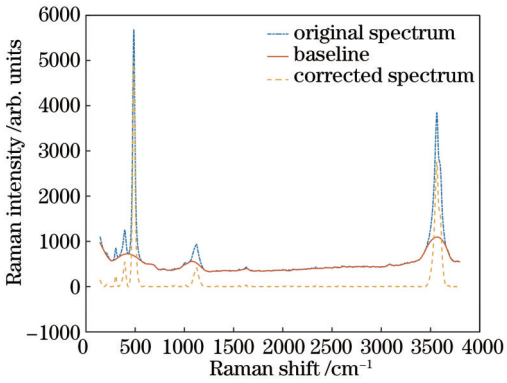

图 2　对方沸石拉曼光谱数据使用 arPLS 去基线

Fig. 2　Using arPLS to deduct baseline for Raman spectral data of Analcime

的对称函数一般被称为“窗函数”，其数学表达式为

$$\begin{cases} y_i' = \sum_{j=-m}^{m} C_j y_{i+j} \\ \sum_{j=-m}^{m} C_j = 0 \\ C_j = C_{-j} \end{cases}, \tag{11}$$

式中：y_i'为变化后的谱数据；y_i为原始谱数据；C_j为对称零面积变换函数，其一般为类峰形函数，是特定函数与常数 d 之差，如下式所示：

$$C_j = G_j - d, \tag{12}$$

$$d = \frac{1}{W}\sum_{j=-m}^{m} G_j, \tag{13}$$

式中：$W=2m+1$为窗宽；G_j一般为特定线型的函数。

在振动光谱学中，许多振动光谱的线轮廓本质上是洛伦兹线型。但由于样品本身的特性以及光谱仪设备存在的统计涨落等影响，光谱会进行一定程度的展宽，展宽因素接近高斯函数。因此拉曼光谱的线型既不是高斯线型，也不是洛伦兹线型。相反，它是洛伦兹函数与高斯函数的卷积，被称为 Voigt 函数[13]，其表达式为

$$G_{\mathrm{voigt}}(i) = \frac{kH_{\mathrm{L}}}{4j^2 + {H_{\mathrm{L}}}^2}\frac{2}{\pi} + (1-k)\frac{\sqrt{4\ln 2}}{\sqrt{\pi}\,H_{\mathrm{G}}} \cdot \exp\left[-4\ln 2\left(\frac{j}{H_{\mathrm{G}}}\right)^2 \right], \tag{14}$$

式中：H_{L}为洛伦兹函数的半峰全宽；H_{G}为高斯函数的半峰全宽；k为洛伦兹函数所占的比重。将式(14)代入式(11)～(13)便可求出 y_i'。当卷积变换后的谱线与它的标准偏差之比出现正极值，并且此极值超过一定数值 f 时，就认为该位置是峰。卷积变换后的谱线与它的标准偏差之比通常用 $S_{\mathrm{S}i}$ 来表示，其表达式为

$$S_{\mathrm{S}i} = \frac{y_i'}{\Delta y_i'} = \frac{\sum_{j=-m}^{m} C_j y_{i+j}}{\sum_{j=-m}^{m} {C_j}^2 y_{i+j}} > f。 \tag{15}$$

对于不同的拉曼光谱，都有各自最佳的 W、H_{L}、H_{G}，以便在对称零面积中准确地寻峰。W 越大，对统计涨落的干扰越不明显。然而，如果 W 太大，则无法识别重叠峰，一般取 $2H_{\mathrm{G}}$ 最合适。当采用不同分辨率的拉曼光谱时，最佳的 W、H_{L}、H_{G} 会有所不同。如果使用不适合的 H_{L}、H_{G}，则会导致寻峰时发生偏移、寻不到峰位或一峰多寻的情况，如图 3 所示。

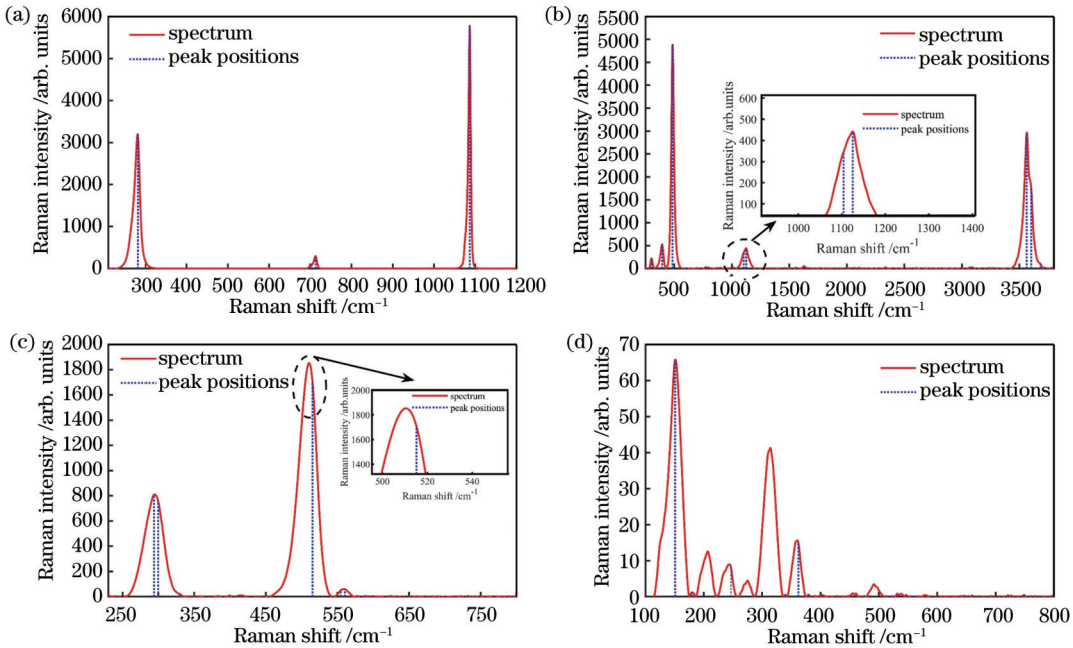

图 3　对 ROD 中拉曼光谱的寻峰效果($H_{\mathrm{L}}=8$, $H_{\mathrm{G}}=12$, $W=19$)。(a) 方解石；(b) 方沸石；(c) 水锑铅矿；(d) 板钛矿

Fig. 3　Peak-seeking effect of Raman spectroscopy in ROD ($H_{\mathrm{L}}=8$, $H_{\mathrm{G}}=12$, $W=19$). (a) Calcite; (b) Analcime; (c) Bindheimite; (d) Brookite

从图 3 中可以看出，在 $H_L=8$，$H_G=12$，$W=19$ 的情况下，对方解石的寻峰效果很好，对方沸石的寻峰效果较好，但出现一个峰寻得两峰位的情况。对水锑铅矿的寻峰出现了明显的偏移和一峰多寻，对板钛矿的寻峰出现了明显的峰没有寻到的情况。这主要是由于 ROD 中不同拉曼光谱分辨率可能不同，因此不同拉曼光谱最适合使用不同的 W、H_L、H_G 值。本文通过将半峰全宽归一化来确定 W、H_L、H_G，通过更普遍的谱线半峰全宽初始值 H 来推算其他参数，这样就统一了高分辨率的拉曼光谱与低分辨率的拉曼光谱的寻峰计算。本文将这种寻峰方式命名为少参数对称零面积（LPSZ）寻峰方法。

$$\begin{cases} H_G = H_L \times 1.5 \\ H_L = \dfrac{H}{\Delta R_S} \\ W = 2\,\mathrm{floor}(H_G) + 1 \end{cases}, \tag{16}$$

式中：ΔR_S 为谱数据拉曼位移的差分；floor( ) 为向下取整函数；H 为半峰全宽初始值，用于对半峰全宽进行初步的归一化，H 取 9.0 左右效果最佳。由于该算法对谱峰强度的信息考虑还不够充分，有时会出现一个很低的峰却有较大的 S_{Si}，而强度较高的峰的 S_{Si} 却很小，如图 4 所示。

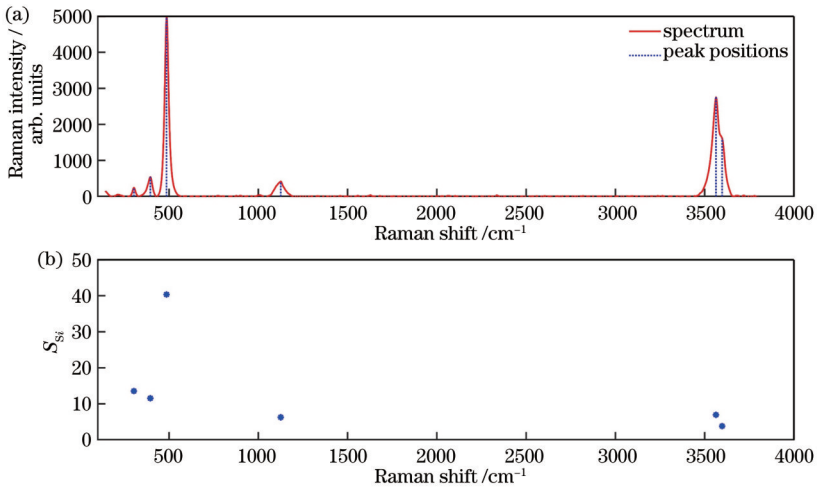

图 4　对方沸石拉曼光谱使用对称零面积法的寻峰效果。(a) 寻得的峰位；(b) 峰位对应的 S_{Si}

Fig. 4　Peak-seeking effect using symmetric zero-area method for Raman spectroscopy of Analcime. (a) Peaks of seeking; (b) S_{Si} of peaks

因此引入 S_{core} 参数对 S_{Si} 和谱峰强度进行平衡从而对峰进行评价：

$$S_{core}(i) = \frac{I_i}{I_{max}} \times p + (100 - p) \times \frac{S_{Si}}{SS_{max}}, \tag{17}$$

式中：I_i 为光谱峰强度；p 为光谱峰强度因子，它越大谱峰强度的考虑越多，其最高为 100。在求出所有寻得峰位的 $S_{core}(i)$ 后，$S_{core}(i)$ 最高为 100，会发现很多 $S_{core}(i)$ 很低的假峰和弱峰，因此可以通过清除 $S_{core}(i)$ 较小的峰来保留强度中等偏低以上的峰。

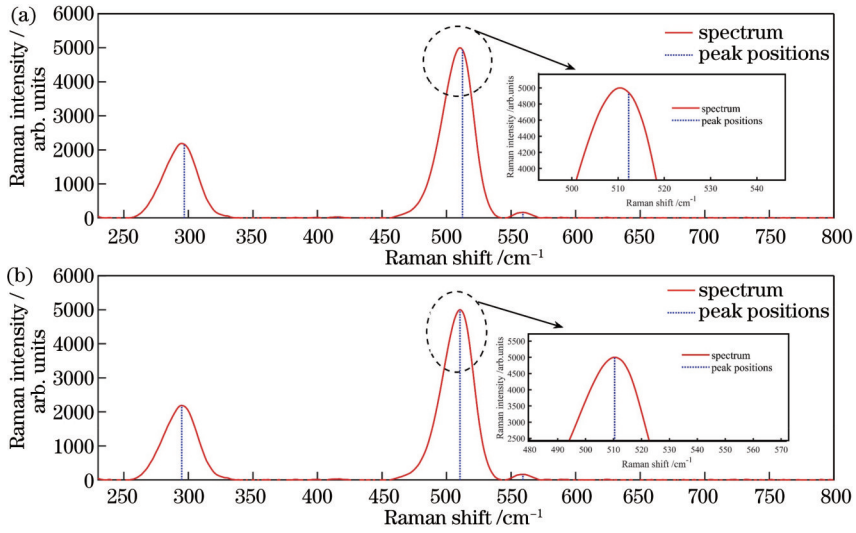

图 5　对 ROD 中水锑铅矿的寻峰修正。(a) 寻峰修正前；(b) 寻峰修正后

Fig. 5　Peak-seeking correction of Bindheimite in ROD. (a) Before peak-seeking correction; (b) after peak-seeking correction

由于通过归一化获得的参数 W、H_L、H_G 比较粗糙，这些参数可能距离完美的参数有一定偏差。修正之后寻得的峰位可能有一些存在些许的偏移[如图 5(a)所示]，需要再对峰位的偏移进行修正。第一次修正已经去除了绝大部分的假峰和一峰多寻的情况，并且将寻得的峰位的偏移控制在一个较低的水平。对第一次修正后的峰的前后几个点(前后至少各一个点)的数据进行查询，判断是否存在更大的强度，用强度最强的位置代替峰位，处理后如图 5(b)所示。

2.4　WALPSZ 寻峰算法的应用

将 Whittaker Smoother 去噪算法、arPLS 去基线算法、LPSZ 寻峰算法相结合，形成 WALPSZ 寻峰算法，其具体流程图如图 6 所示。y^s 是原始拉曼光谱数据 y 去噪声后的拉曼光谱数据。去噪声后初始化权值因子 w 为全 1 数组，通过求解式(9)获取基线 z 后根据与 y^s 的大小关系可以获得新的权值因子 w，重复这一步骤直到权值因子的变化很小时候就能得到最终的基线；y^{sz} 是去除噪声和背景的拉曼光谱数据。在归一化窗宽

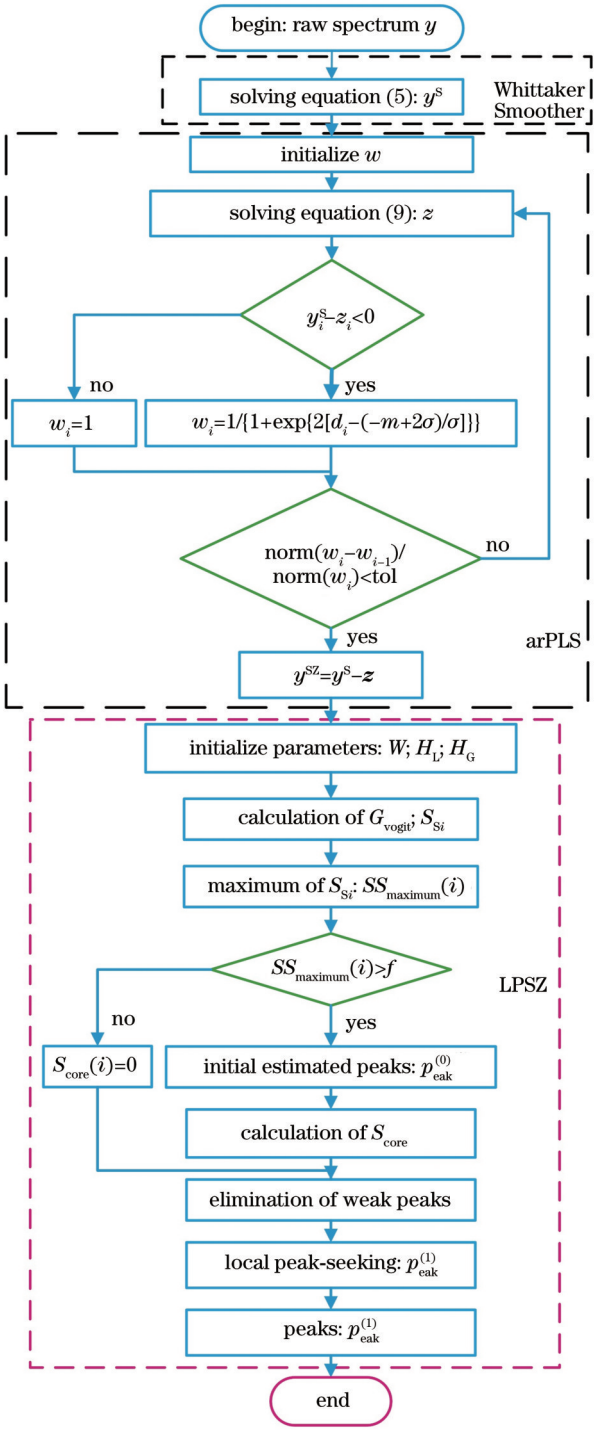

图 6　WALPSZ 寻峰算法的流程图

Fig. 6　Flow chart of WALPSZ peak-seeking algorithm

W、洛伦兹函数半峰全宽H_L和高斯函数半峰全宽H_G这些参数之后使用对称零面积法寻峰获得初步的峰位的数组$p_{eak}^{(0)}$。再根据S_{core}对$p_{eak}^{(0)}$进行调整，去除弱峰和局部寻峰后得出真正的峰位数组$p_{eak}^{(1)}$。

目前开放的拉曼数据库，如ROD、RRUFF等含有许多实验原始谱数据，但是没有提供特征峰位，不能直接通过峰位对比来获取物质的化学信息。WALPSZ寻峰算法具有普适性，使用该寻峰算法可以将ROD中的原始拉曼光谱数据转化成峰位和对应峰强的信息，方便使用。使用WALPSZ寻峰算法可以对数据库中大量原始拉曼光谱批量寻峰，并选出其$S_{core}(i)$最强的四个峰位。然后将物质名称、物质组成、包含元素、最强的四个峰位、峰位对应的强度等关键信息存入表中，形成一个丰富、方便的拉曼数据库，如表1所示，对应的部分寻峰效果如图7所示。在分析拉曼光谱时，可以直接通过对比拉曼光谱特征峰位来分析物质的化学信息。

表1　对ROD批量寻峰得到的部分峰位数据

Table1　Part of peaks obtained by batch peak-seeking of ROD

Name	Formula	Peak 1	Peak 2	Peak 3	Peak 4	S1	S2	S3	S4
Almandine	Al2Ca0.12Fe2.58Mg0.27O12Si3	917.43	355.28	558.73	505.22	10.0	4.7	3.5	2.9
Leiteite	As2O4Zn	457.06	217.93	148.51	601.70	9.8	8.3	3.9	3.4
Leightonite	CaCu0.68H4KO9S2	1003.07	461.17	625.57	—	10.0	5.0	1.6	—
Zwieselite	FFe2O4P	978.32	424.85	1068.96	607.09	10.0	2.7	2.2	1.9
Hanksite	C2ClKNa22O42S9	989.74	1080.38	630.57	1114.13	10.0	3.2	1.5	1.3
Lepidocrocite	FeHO2	247.58	372.93	214.31	306.40	10.0	4.0	2.3	1.7

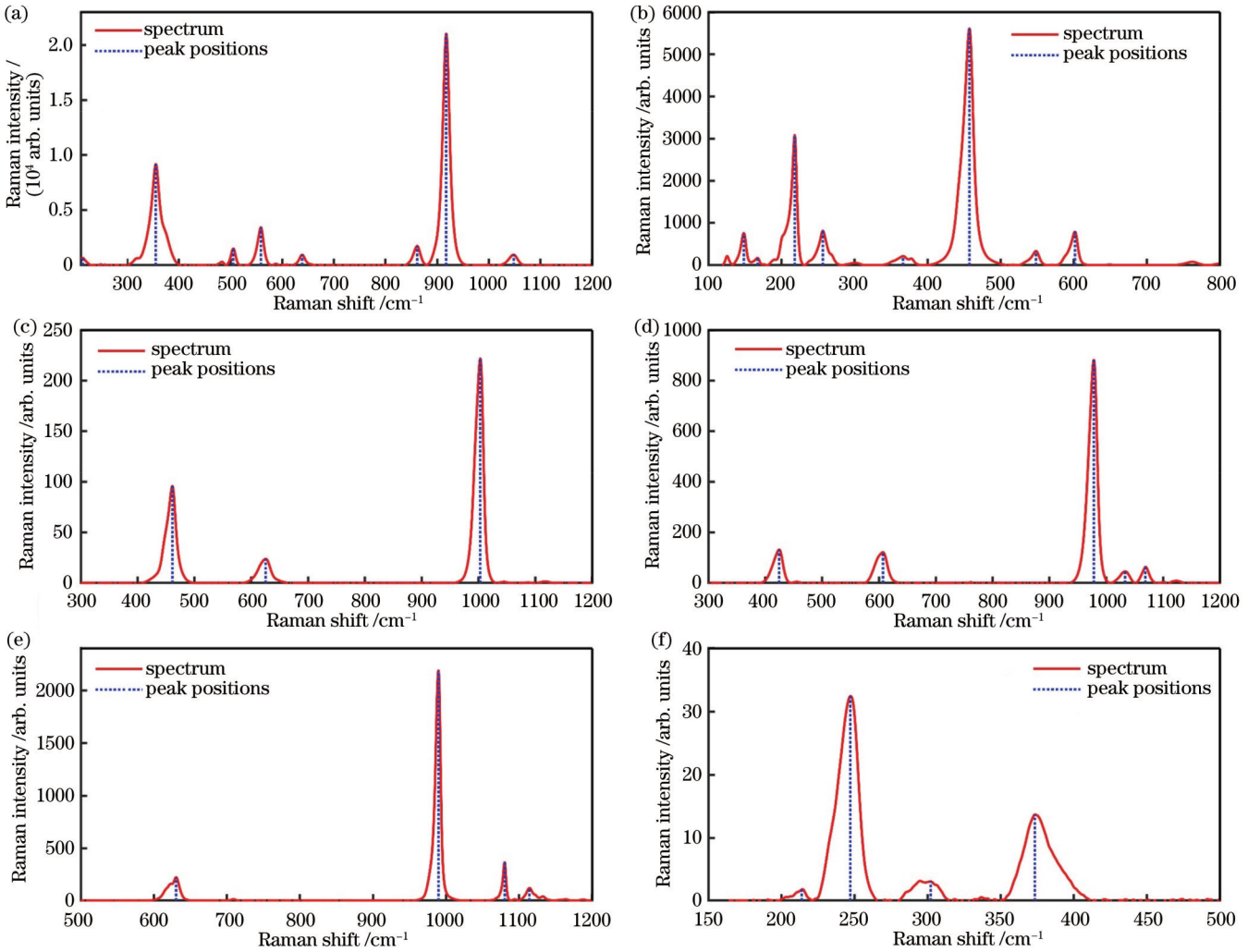

图7　对ROD批量寻峰得到的部分峰位数据的寻峰效果图。(a) 铁铝榴石；(b) 亚砷锌石；(c) 雷顿石；(d) 铁磷灰石；(e) 碳酸芒硝；(f) 纤铁矿

Fig. 7　Peak-seeking effects of partial peaks obtained by batch peak-seeking of ROD. (a) Almandine; (b) Leiteite; (c) Leightonite; (d) Zwieselite; (e) Hanksite; (f) Lepidocrocite

3 结果与讨论

3.1 WALPSZ 寻峰算法在拉曼开放数据库的寻峰效果

使用 WALPSZ 寻峰算法对图 3 中的拉曼光谱再次进行寻峰处理，结果如图 8 所示。通过对比可以发现，使用 WALPSZ 寻峰算法对方解石依然保持良好的寻峰效果[图 8(a)]；对方沸石在 1000～1500 cm^{-1} 处一峰多寻的情况消失[图 8(b)]；对水锑铅矿寻峰偏移的情况得到修正[图 8(c)]；对板钛矿未寻得峰的情况进行修正[图 8(d)]。因此修正后，对这些光谱数据的寻峰都达到了很好的效果，可以看出，此算法相较于传统的算法降低了物质和分辨率变化对寻峰的影响。

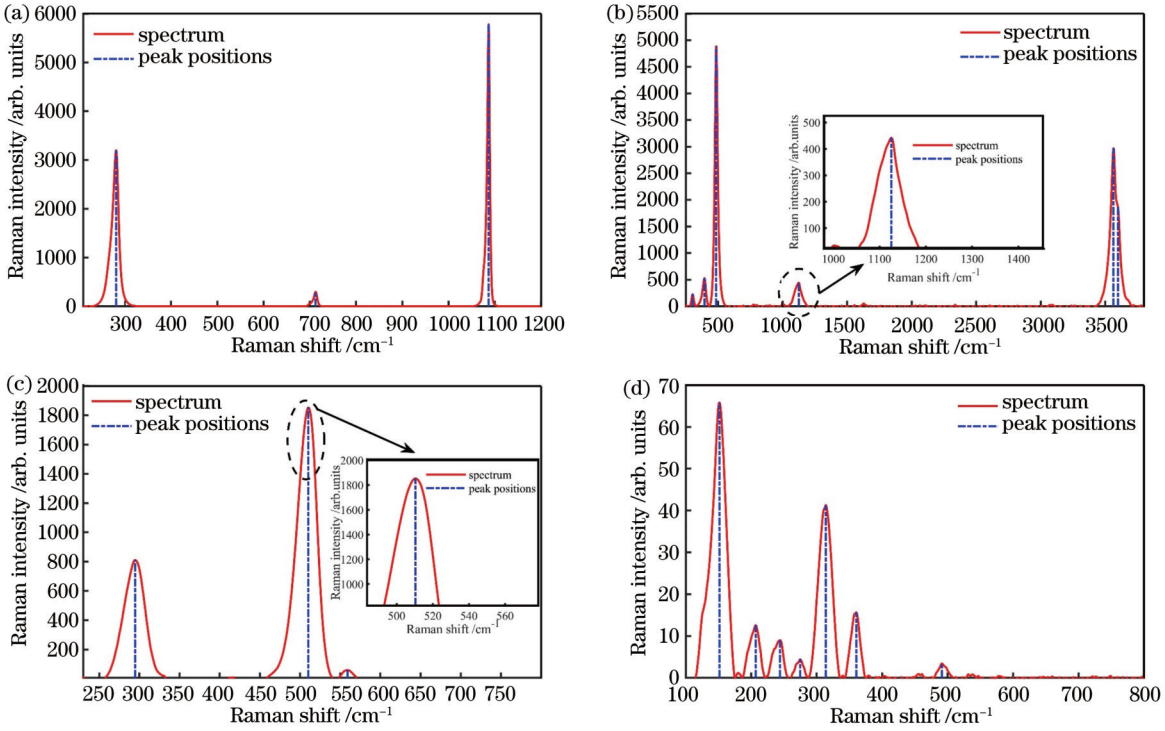

图 8 WALPSZ 寻峰算法的寻峰效果。(a) 方解石；(b) 方沸石；(c) 水锑铅矿；(d) 板钛矿

Fig. 8 Peak-seeking effect of WALPSZ peak-seeking algorithm. (a) Calcite; (b) Analcime; (c) Bindheimite; (d) Brookite

3.2 算法在实测光谱中的应用

为了进一步验证 WALPSZ 寻峰算法的适用性和准确性，以及探究该算法是否能在实际测量的拉曼光谱中依然适用，分别准备了无水硫酸钙样品、黄铁矿、莫桑石。这些样品使用法国 HORIBA 公司的激光拉曼光谱仪(LabRAM HR)进行测量，该光谱仪的焦长为 800 mm，光谱重复性≤±0.2 cm^{-1}，光谱分辨率≤±1.95 cm^{-1}，激光波长使用 785 nm。对实测样品的拉曼光谱分析如图 9 所示，从图中可见，一些原始光谱数据具有一定的背景和噪声。通过使用 Whittaker Smoother 去噪算法和 arPLS 去基线算法对实测拉曼光谱数据进行处理后得到的光谱数据如图 9(b)、9(d)和 9(f)所示，处理后的拉曼光谱数据已经没有明显的噪声和背景。

使用 WALPSZ 寻峰算法对从 ROD 和 RRUFF 数据库中抽取的无水硫酸钙、黄铁矿、莫桑石的拉曼光谱以及实际测量的拉曼光谱进行寻峰。如图 10 所示，WALPSZ 寻峰算法能准确地寻得 ROD 中无水硫酸钙[图 10(a)]、黄铁矿[图 10(b)]、莫桑石[图 10(c)]的拉曼光谱峰位，未发生峰位的偏移。对 RRUFF 数据库中样品的拉曼光谱寻峰效果如图 11 所示，RRUFF 数据库中的部分原始拉曼光谱的质量要低于 ROD 中的原始拉曼光谱，如无水硫酸钙的原始拉曼光谱有较高的背景和一定的噪声[图 11(a)]，但仍然可以对这些谱线准确地寻峰。使用 WALPSZ 寻峰算法对实际测量样品的拉曼光谱寻峰效果如图 12 所示，明显可以观察到，实测拉曼光谱的峰位也未发生偏移并且能与该算法对 ROD 和 RRUFF 数据库中拉曼光谱寻得的峰位相对应。

为了进一步验证 WALPSZ 寻峰算法的可靠性，将该算法寻得的数据库中的拉曼光谱峰位和实际测量的拉曼光谱峰位与实际峰位进行比较，如表 2 所示。可以看出，WALPSZ 寻峰算法在 ROD、RRUFF 数据库和实际测量拉曼光谱中寻得的峰位能相互对应，且与相关文献中的数据也能相互对应。

4 结 论

在分析拉曼光谱时，通过获取峰的拉曼位移与数

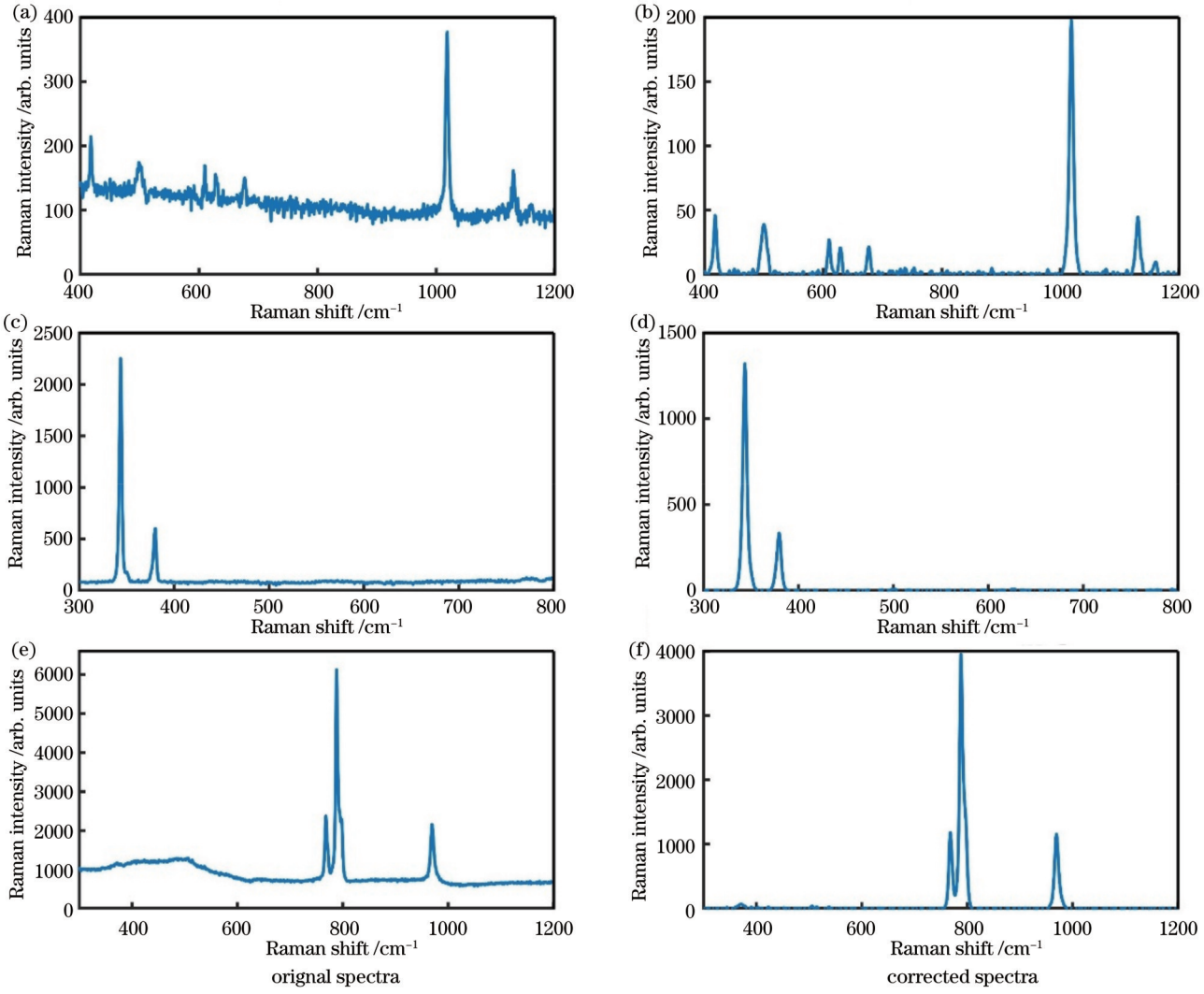

图 9 对实测样品拉曼光谱数据扣除噪声和背景。(a)、(b) 无水硫酸钙;(c)、(d) 黄铁矿;(e)、(f) 莫桑石

Fig. 9 Noise and background are deducted from Raman spectral data of measured sample. (a), (b) Anhydrite; (c), (d) Pyrite; (e), (f) Moissanite

表 2 四种来源的拉曼光谱峰位比较

Table 2 Comparison of Raman spectral peaks from four sources

Name	Peaks from ROD / cm^{-1}	Peaks from RRUFF /cm^{-1}	Peaks from measured samples / cm^{-1}	Actual peaks[19-21] /cm^{-1}
Anhydrite	1018.2	1017.1	1018.2	1017
	1130.9	1129.0	1130.9	1130
	500.2	498.8	500.2	500
	417.7	416.8	417.8	417
Pyrite	344.1	343.4	343.5	343
	380.3	379.9	379.6	379
Moissanite	789.1	788.7	788.2	789
	767.5	767.0	768.3	760-800
	968.5	965.7	969.5	965-972

据库对比可以识别样品的化学组成。然而,一些开放的拉曼数据库(如 ROD)仅有拉曼光谱原始数据,需要对其进行批量寻峰以获得特征信息。对称零面积法在自动寻峰方面具有优势,但处理 ROD 中各种分辨率的拉曼光谱时通用性不足。本文通过将对称零面积法寻峰进行改进,减少了输入参数后,再将其与 Whittaker

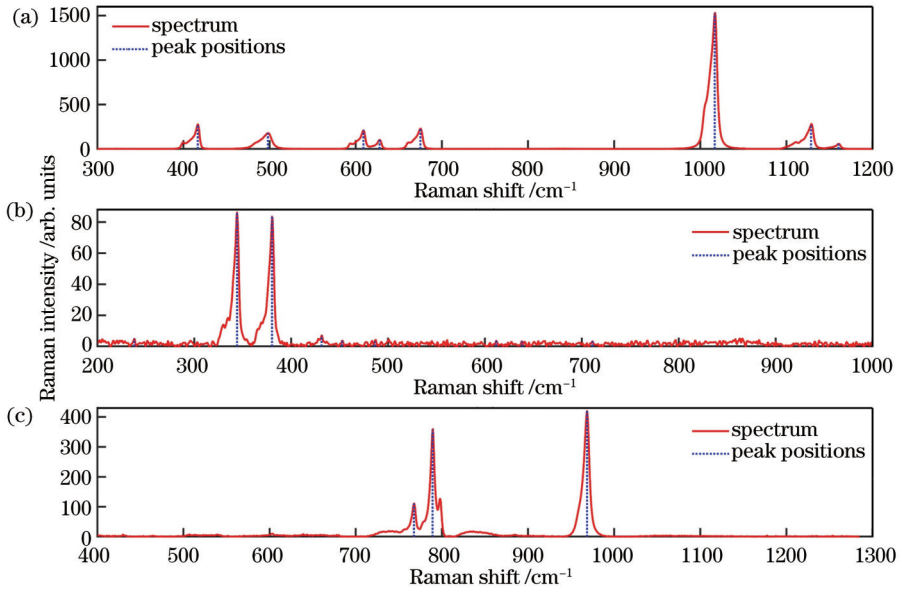

图 10　WALPSZ 寻峰算法对三种待测样品 ROD 中拉曼光谱的寻峰效果。(a) 无水硫酸钙;(b) 黄铁矿;(c) 莫桑石

Fig. 10　Effect of Raman spectroscopy peak-seeking of WALPSZ peak-seeking algorithm in ROD on three samples to be measured. (a) Anhydrite; (b) Pyrite; (c) Moissanite

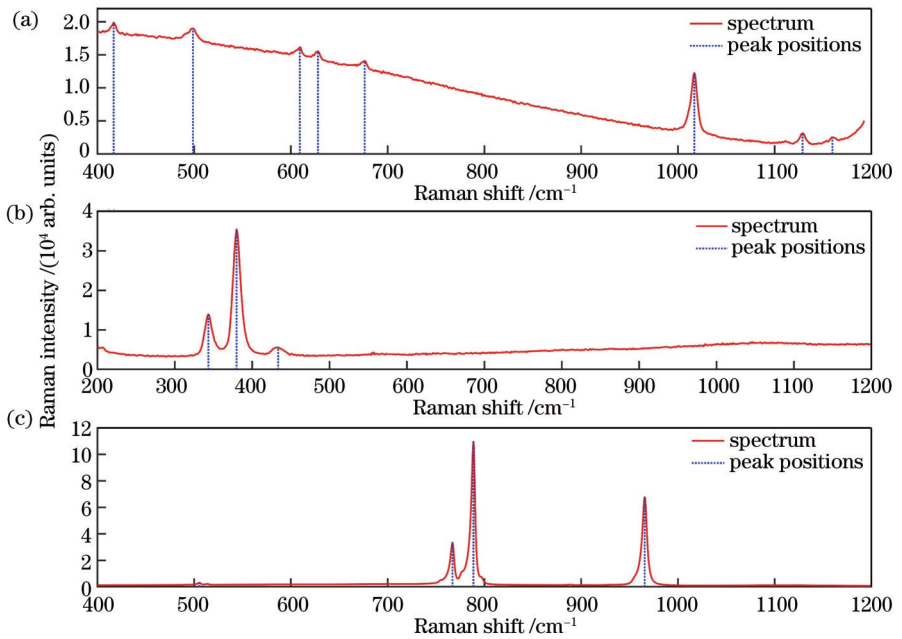

图 11　WALPSZ 寻峰算法对三种待测样品 RRUFF 中拉曼光谱的寻峰效果。(a) 无水硫酸钙;(b) 黄铁矿;(c) 莫桑石

Fig. 11　Effect of Raman spectroscopy peak-seeking of WALPSZ peak-seeking algorithm in RRUFF on three samples to be measured. (a) Anhydrite; (b) Pyrite; (c) Moissanite

Smoother 去噪算法、arPLS 去基线算法结合，形成 WAPLSZ 寻峰算法。WAPLSZ 寻峰算法相较于传统的对称零面积变换寻峰算法增加了其普适性，减少了参数的输入，使该算法可以对 ROD 中的光谱数据自动批量寻峰。

本文使用 WALPSZ 寻峰算法获取了 ROD 与 RRUFF 数据库中无水硫酸钙、黄铁矿、莫桑石的拉曼光谱的峰位。也通过该算法获取了实际测量上述样品得到的拉曼光谱数据的峰位，并将这些与相关文献中的峰位进行对比。结果表明，WALPSZ 寻峰算法对实测拉曼光谱数据，以及 ROD 中原始数据的自动寻峰是有效的，获取峰位可以相对应，且都与文献中记录的数据一致。这验证了 WALPSZ 寻峰算法对拉曼原始数据自动寻峰的可靠性与准确性，可以将 ROD 中自动寻峰的峰位建库，并和文献记录的数据对应，从而分析实测拉曼光谱的化学信息。

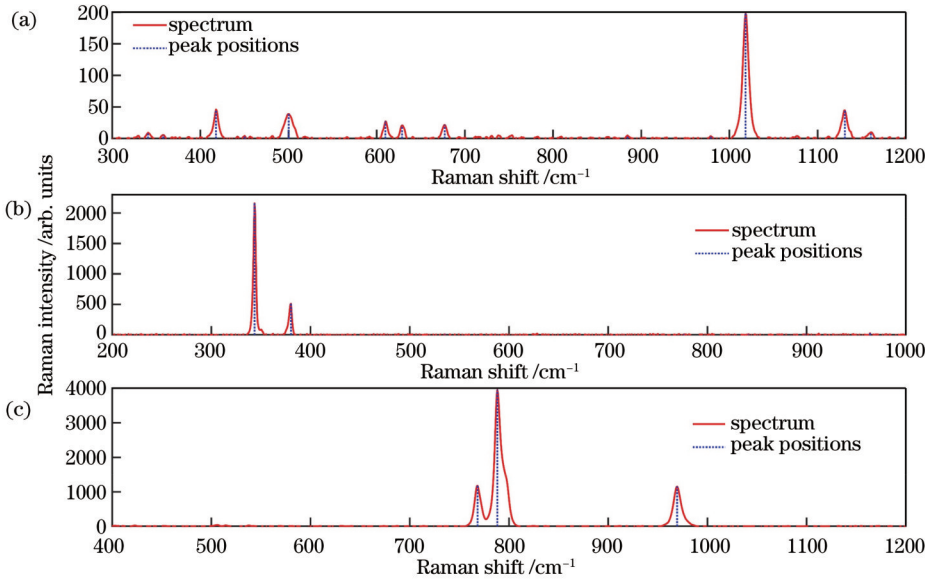

图 12 WALPSZ 寻峰算法对三种待测样品实测拉曼光谱的寻峰效果。(a) 无水硫酸钙;(b) 黄铁矿;(c) 莫桑石

Fig. 12 Peak-seeking effect of actual Raman spectroscopy of WALPSZ peak-seeking algorithm on three samples to be measured. (a) Anhydrite; (b) Pyrite; (c) Moissanite

参 考 文 献

[1] 张灿, 张洁, 朱永. 槽型波导耦合纳米结构增强拉曼光谱[J]. 光学学报, 2020, 40(3): 0313001.
Zhang C, Zhang J, Zhu Y. Slot-waveguide coupled nanostructure enhanced Raman spectroscopy[J]. Acta Optica Sinica, 2020, 40(3): 0313001.

[2] 覃宗定, 许雪棠, 张枝芝, 等. 基于拉曼光谱的硝酸甘油对活体血液作用的实时分析[J]. 光学学报, 2014, 34(1): 0130001.
Qin Z D, Xu X T, Zhang Z Z, et al. Real-time analysis of blood in vivo injected with nitroglycerin using Raman spectroscopy[J]. Acta Optica Sinica, 2014, 34(1): 0130001.

[3] Colantonio C, Clivet L, Laval E, et al. Integration of multispectral imaging, XRF mapping and Raman analysis for noninvasive study of illustrated manuscripts: the case study of fifteenth century "Humay meets the Princess Humayun" Persian masterpiece from Louvre Museum[J]. The European Physical Journal Plus, 2021, 136(9): 958.

[4] Ru C L, Wen W, Zhong Y. Raman spectroscopy for on-line monitoring of botanical extraction process using convolutional neural network with background subtraction[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2023, 284: 121494.

[5] 黄祖芳, 李玉玲, 杜生荣, 等. 基于拉曼光谱技术的精子评筛研究进展[J]. 中国激光, 2023, 50(15): 1507202.
Huang Z F, Li Y L, Du S R, et al. Recent Progress in sperm evaluation and screening based on Raman spectroscopy[J]. Chinese Journal of Lasers, 2023, 50(15): 1507202.

[6] Smith E, Dent G. Modern Raman spectroscopy: a practical approach[M]. 2nd ed. Singapore: John Wiley & Sons, 2019.

[7] 司赶上, 刘家祥, 李振钢, 等. 基于形态学与多项式拟合的紫外拉曼荧光背景扣除算法[J]. 光学学报, 2022, 42(22): 2230001.
Si G S, Liu J X, Li Z G, et al. Fluorescence background subtraction algorithm of UV Raman fluorescence based on morphology and polynomial fitting[J]. Acta Optica Sinica, 2022, 42(22): 2230001.

[8] 姚泽楷, 蔡耀仪, 李诗文, 等. 基于平滑样条曲线结合离散状态转移算法的拉曼光谱基线校正方法[J]. 中国激光, 2022, 49(18): 1811001.
Yao Z K, Cai Y Y, Li S W, et al. baseline correction for Raman spectroscopy using cubic spline smoothing combined with discrete state transformation algorithm[J]. Chinese Journal of Lasers, 2022, 49(18): 1811001.

[9] 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(9): 0915003.
Yang G Y, Li L, Chen H, et al. Baseline correction method for Raman spectra based on generalized Whittaker smoother[J]. Chinese Journal of Lasers, 2015, 42(9): 0915003.

[10] 吴和喜, 袁新宇, 刘庆成, 等. 一种γ谱弱峰优化检测方案研究[J]. 原子能科学技术, 2012, 46(9): 1142-1146.
Wu H X, Yuan X Y, Liu Q C, et al. Optimal scheme for detecting weak peak in γ-ray spectrum[J]. Atomic Energy Science and Technology, 2012, 46(9): 1142-1146.

[11] 汪雪元, 何剑锋, 刘琳, 等. 小波变换导数法 X 射线荧光光谱自适应寻峰研究[J]. 光谱学与光谱分析, 2020, 40(12): 3930-3935.
Wang X Y, He J F, Liu L, et al. Research on adaptive peak detection of X-ray fluorescence spectrum with wavelet transform and derivative method[J]. Spectroscopy and Spectral Analysis, 2020, 40(12): 3930-3935.

[12] Zhang Z M, Tong X, Peng Y, et al. Multiscale peak detection in wavelet space[J]. Analyst, 2015, 140(23): 7955-7964.

[13] 毕云峰, 李颖, 郑荣儿. LIBS/Raman 光谱对称零面积变换自动寻峰方法研究[J]. 光谱学与光谱分析, 2013, 33(2): 438-443.
Bi Y F, Li Y, Zheng R E. The symmetric zero-area conversion adptive peak-seeking method research for LIBS/Raman spectra [J]. Spectroscopy and Spectral Analysis, 2013, 33(2): 438-443.

[14] El Mendili Y, Vaitkus A, Merkys A, et al. Raman Open Database: first interconnected Raman-X-ray diffraction open-access resource for material identification[J]. Journal of Applied Crystallography, 2019, 52(3): 618-625.

[15] Yin S, Wang W. Denoising lidar signal by combining wavelet improved threshold with wavelet domain spatial filtering[J]. Chinese Optics Letters, 2006, 4(12): 694-696.

[16] Eilers P H C. A perfect smoother[J]. Analytical Chemistry, 2003, 75(14): 3631-3636.

[17] Eilers P H C. Parametric time warping[J]. Analytical Chemistry, 2004, 76(2): 404-411.

[18] Baek S J, Park A, Ahn Y J, et al. Baseline correction using asymmetrically reweighted penalized least squares smoothing[J].

The Analyst, 2015, 140(1): 250-257.
[19] Dobrzhinetskaya L, Mukhin P, Wang Q, et al. Moissanite (SiC) with metal-silicide and silicon inclusions from tuff of Israel: Raman spectroscopy and electron microscope studies[J]. Lithos, 2018, 310/311: 355-368.
[20] Muñoz E C, Gosetti F, Ballabio D, et al. Characterization of pyrite weathering products by Raman hyperspectral imaging and chemometrics techniques[J]. Microchemical Journal, 2023, 190: 108655.
[21] Prieto-Taboada N, Gómez-Laserna O, Martínez-Arkarazo I, et al. Raman spectra of the different phases in the $CaSO_4$-H_2O system[J]. Analytical Chemistry, 2014, 86(20): 10131-10137.

Application of Improved Symmetric Zero-Area Conversion Peak-Seeking Algorithm in Raman Spectroscopy

Wang Hai, Huang Ning*, He Ze, Wang Peng, Yuan Jingxi

Key Laboratory of Radiation Physics and Technology, Ministry of Education, Institute of Nuclear Science and Technology, Sichuan University, Chengdu 610064, *Sichuan, China*

Abstract

Objective Raman spectroscopy is an efficient and non-destructive analytical method for obtaining chemical information. The characteristic peaks in a Raman spectrum contain chemical information about the substance. The symmetric zero-area conversion is a commonly employed peak-seeking method. However, before peak seeking, various parameters related to the spectral line should be input, such as window width, Lorentz function half-width, and Gaussian function half-width. For different Raman spectra, these parameters to be input may be different, and if the input parameters do not match the current Raman spectrum, the obtained peak positions may be inaccurate. Currently, some open Raman databases only contain raw Raman spectral data without corresponding peak information. Preprocessing the raw spectral data and obtaining the corresponding peak positions and intensities by peak-seeking algorithms lead to better and more convenient utilization. Although the symmetric zero-area conversion method has advantages in automatic peak seeking and can obtain the intensity information corresponding to the spectral peaks, this peak-seeking algorithm requires various parameters related to the spectral data, such as window width, Lorentz function half-width, and Gaussian function half-width. Therefore, the universality of the symmetric zero-area conversion method is relatively limited during processing different Raman spectra in the database. We propose an improved symmetric zero-area method to reduce the input of parameters related to spectral data and adapt it to data with different spectral resolutions. We hope that this algorithm can automatically search peaks in batches for many raw Raman spectral data in the Raman database to generate a more concise and convenient database.

Methods This algorithm improves the peak-seeking algorithm of symmetric zero-area conversion by combining noise reduction and baseline removal algorithms. First, the Whittaker Smoother algorithm is employed to remove noise from the raw Raman spectrum, which can quickly and easily remove noise without producing peak position shifts. Then, the asymmetrically weighted penalized least squares (arPLS) algorithm is utilized to remove the spectrum baseline. Next, we improve the symmetric zero-area method by normalizing the half-width of the Raman spectrum peaks, thus reducing the number of required input parameters and suppressing peak-seeking offsets. After peak seeking, the found peak positions are further corrected to reduce offsets and accurately locate peaks. Finally, the WALPSZ peak-seeking algorithm is formed by combining the Whittaker Smoother and arPLS. Additionally, the algorithm is leveraged to automatically search for peaks in ROD′s raw Raman spectral data and adopted for experimental Raman spectral analysis of Anhydrite, Pyrite, and Moissanite. The obtained peak positions are compared with the literature′s data to verify their reliability and universality for different Raman spectral data.

Results and Discussions First, the traditional symmetric zero-area conversion method and the WALPSZ algorithm are applied to analyze the peak seeking of ROD′s Calcite, Analcime, Bindheimite, and Brookite original spectral data. When utilizing the traditional symmetric zero-area peak-seeking algorithm with fixed parameters, it has the best peak-seeking effect on Calcite [Fig. 3(a)] and a better peak-seeking effect on Analcime, but there is a situation where a peak is searched twice at 1000–1500 cm^{-1} [Fig. 3(b)]. The peak seeking of Bindheimite shows an obvious peak-seeking offset and a situation where one peak is searched twice [Fig. 3(c)]. The peak seeking of Brookite exhibits a clear missing peak case [Fig. 3(d)]. By employing the WALPSZ peak-seeking algorithm, it maintains a sound peak-seeking effect on Calcite and solves the above inaccurate peak-seeking problems when facing other Raman spectra, which indicates that the WALPSZ

peak-seeking algorithm has better universality. To further verify the universality and accuracy of the WALPSZ peak-seeking algorithm and explore whether the algorithm can still be applied in actual measured Raman spectra, Anhydrite, Pyrite, and Moissanite are prepared for Raman spectral measurement, and the WALPSZ peak-seeking algorithm is adopted for peak-seeking analysis (Fig. 12). The found peaks are compared with those found by the WALPSZ peak-seeking algorithm in the original spectral data of these three samples in ROD and RRUFF and literature data, and we find that these peaks can correspond to each other (Table 2).

Conclusions The symmetric zero-area conversion method is improved by reducing the input parameters and then is combined with the Whittaker Smoother and arPLS baseline removal algorithm to form the WAPLSZ peak-seeking algorithm, which enhances its universality. The WAPLSZ peak-seeking algorithm is compared with the traditional symmetric zero-area conversion method and the peak-seeking results of other original Raman spectra of ROD by the WAPLSZ peak-seeking algorithm. The results show that reducing the input parameters makes this algorithm capable of automatically batch searching for spectral data in open Raman databases. Meanwhile, we employ the WALPSZ peak-seeking algorithm to obtain the peak positions of Anhydrite, Pyrite, and Moissanite in ROD and RRUFF′s Raman spectra, obtain the peak positions of the measured Raman spectra of these samples by this algorithm, and compare them with the peak positions in literature. The results reveal that the WALPSZ peak-seeking algorithm is effective for automatically searching for peaks in measured Raman spectral data and original data in ROD and that the obtained peak positions can correspond to each other and are consistent with the data recorded in the literature. Then, the reliability and accuracy of the WALPSZ peak-seeking algorithm are verified for automatically searching for peaks in Raman original data. Finally, this algorithm can help establish a database of automatically searched peak positions in ROD and correspond to data recorded in literature to analyze chemical information from measured Raman spectra.

Key words spectroscopy; Raman spectroscopy; automatic peak seeking; symmetric zero-area conversion