

文章编号:1000-582X(2012)05-076-06

光谱重叠峰的曲线拟合解析策略与实现

胡耀垓¹, 张晓星², 赵正予¹, 冯 波²

(1. 武汉大学 电子信息学院, 湖北 武汉 430079; 2. 重庆大学 高电压与电工新技术教育部重点实验室, 重庆 400044)

摘 要:针对光谱分析中的重叠峰问题,提出了一种基于曲线拟合的光谱重叠峰解析算法,讨论了曲线拟合的参数调整和表达非唯一性等问题,提出了最小峰-峰间距(MRI)、可分离最小峰-峰间距(MSI)和拟合误差曲线的三个特征区间等概念;在峰位的确定中,引入由宽逐严的排除策略。含噪模拟光谱重叠峰解析结果表明,所提出的算法是有效的。

关键词:光谱重叠峰;曲线拟合;重叠峰解析;高斯分布

中图分类号:TM930.12+6

文献标志码:A

Strategy and implementation of resolve overlapping spectra based on curve fitting

HU Yao-gai¹, ZHANG Xiao-xing², ZHAO Zheng-yu¹, FENG Bo²

(1. College of Electronic Information, Wuhan University, Wuhan 430079, P. R. China; 2. Key Laboratory of High Voltage and Electrical New Technology, Ministry of Education, Chongqing University, Chongqing 400044, P. R. China)

Abstract: Aiming at the overlapping peaks in spectrum analysis, a novel method of curve fitting based on Gaussian function is presented to resolve the overlapping peaks. The theory of curve fitting is introduced firstly, and then an algorithm is proposed based on both the minimum separable peak-peak interval and the curve-fitting error. In the peak positioning, a gradually strict strategy is introduced to exclude the fake peaks. The resolution of several kinds of overlapping peaks with computer-simulated noise has been performed and discussed in details. The calculated results indicate that the peak positions can be extracted effectively, even in the case of serious overlapped, and clearly show the effectiveness of the proposed method.

Key words: overlapping peaks; curve fitting; resolving; Gaussian distribution

光谱信号中不同谱峰的峰位相近且峰宽较大,可致谱带之间发生严重重叠,从而给光谱分析带来不利影响。如欲对重叠的光谱有较为全面、准确的认识,对光谱进行重叠峰解析是十分必要的。一直以来,重叠峰解析都是光谱分析领域广受关注和亟

待解决的难题。目前常用的重叠峰解析技术可以分为数值方法和非数值方法两类。非数值方法包括提高仪器的分辨率、改进实验方法和改善实验条件等;数值方法则是借助某些数学或统计学方法,把通过化学方法和仪器未能完全分离的复合量测信号分解

收稿日期:2011-11-20

基金项目:国家自然科学基金资助项目(60701020);中央高校基本科研业务费专项资金资助项目(3103001);输配电装备及系统安全与新技术国家重点实验室研究项目(2007DA10512708406)

作者简介:胡耀垓(1972-),男,武汉大学副教授,主要从事光谱激励信号检测及特征信息提取、电离层物理及检测技术的研究,(E-mail)yaogaihu@whu.edu.cn。

成几个单独组分的信号,从而获取重叠谱中每个组分的相关信息,主要包括曲线拟合、Kalman 滤波、因子分析、Fourier 自去卷积和小波变换等方法^[1]。曲线拟合作为一种有力的研究手段^[2],是根据最小二乘法原理将原始谱图解析为一个或一个以上单峰之和,并且使拟合谱图与实测谱图之间误差平方和达到最小。在重叠含噪光谱解析中,确定峰位最为关键^[3]。Maddams 等^[4-5]曾应用二阶和四阶导数确定重叠峰的数量和位置,Fourier 自去卷积也可用于拟合曲线参数的估计^[6-7],基于自然计算的峰检测也有报道^[8,9]。此外,应用小波技术解析重叠峰也是一个相当活跃的领域^[3,10-14]。以上重叠峰解析数值方法都有其各自的局限性,到目前为止,尚无一种被普遍接受的、通用的有效手段。

笔者详细讨论了一种基于曲线拟合的光谱重叠峰解析算法,包括参数的调整和表达非唯一性等,提出基于拟合误差曲线确定合适的可分离最小峰-峰间距。在峰位的确定中,引入由宽逐严的排除策略。最后为了验证算法的有效性,采用提出的算法对含噪模拟重叠光谱进行了解析。

1 基本原理

1.1 基于高斯型模拟函数的曲线拟合

曲线拟合通常采用一些已知函数的组合、统计方法、经验公式和测定的峰型来假定曲线拟合的数学模型^[15]。光谱谱峰可用高斯函数、双高斯函数、指数衰减修正高斯函数、洛伦兹函数或者高斯函数和洛伦兹函数的组合函数进行模拟。笔者采用高斯函数作为对单个谱峰模拟,其他形式模拟函数的相应算法推导可采用完全类似的思想。

用高斯型函数模拟的单个谱峰形如:

$$g_{\mu,\sigma_{\mu}}(\lambda) = \frac{k_{\mu}}{\sqrt{2\pi}\sigma_{\mu}} e^{-\frac{(\lambda-\mu)^2}{2\sigma_{\mu}^2}} = A_{\mu} e^{-\frac{(\lambda-\mu)^2}{2\sigma_{\mu}^2}}, \quad (1)$$

式中: $g_{\mu,\sigma_{\mu}}(\lambda)$ 为关于光谱频率 λ (或波数 ν)的函数; μ 、 A_{μ} 和 σ_{μ} 分别为谱线的峰位、峰强和峰宽参数。

在连续情况下,可以认为 $\forall \lambda \in [a, b]$,皆存在单峰 $g_{\mu,\sigma_{\mu}}(\lambda)$ 。不存在单峰 $g_{\mu,\sigma_{\mu}}(\lambda)$ 之处,峰强 $A_{\mu}=0$ 。

基于实际光谱 $f^*(\lambda)$ 由若干个单峰谱带 $g_{\mu,\sigma_{\mu}}(\lambda)$ 相互叠加所形成的假设,拟合谱线函数为

$$f(\lambda) = \int_a^b g_{\mu,\sigma_{\mu}}(\lambda) d\mu = \int_a^b A_{\mu} e^{-\frac{(\lambda-\mu)^2}{2\sigma_{\mu}^2}} d\mu. \quad (2)$$

实际谱图 $f^*(\lambda)$ 的最优拟合函数 $f(\lambda)$ 所对应的误差平方和 $E = \frac{1}{2} \int_a^b [f(\lambda) - f^*(\lambda)]^2 d\lambda$ 为所有拟合函数中最小者。

离散情况下,区间 $[a, b]$ 内任一点以整数 i 或 j ($i, j=1, 2, \dots, N$)表示,以 i 点为中心的高斯函数记作:

$$g_{i,\sigma_i}(j) = \frac{k_i}{\sqrt{2\pi}\sigma_i} e^{-\frac{(j-i)^2}{2\sigma_i^2}} = A_i e^{-\frac{(j-i)^2}{2\sigma_i^2}}, \quad (3)$$

拟合谱图可表示为

$$f(j) = \sum_{i=1}^N g_{i,\sigma_i}(j) = \sum_{i=1}^N A_i e^{-\frac{(j-i)^2}{2\sigma_i^2}}, \quad (i, j = 1, 2, \dots, N). \quad (4)$$

$f(j)$ 与 $f^*(j)$ 的误差平方和函数定义为

$$E = \frac{1}{2} \sum_{j=1}^N [f(j) - f^*(j)]^2. \quad (5)$$

寻找最小二乘法意义下的最优拟合函数 $f(j)$,实质上是确定函数 E 的极小值问题。由式(4)和式(5)可知, E 是关于 (A_i, σ_i) ($i=1, 2, \dots, N$)的多元非线性函数,用解析法直接求解极值存在困难,但可以通过数值迭代法进行。数值迭代法分为两类:一类是导数法,该方法以目标函数导数的负方向作为下一次迭代方向;另一类是直接法,该方法一般是在目标函数的导数难以得到的情况下,直接利用目标函数而不利用其导数来确定下一次迭代方向。一般而言,前者的计算效率比后者高。这里目标函数 E 的导数存在并较易得到,因此笔者采用导数法。类似的思想可参见BP人工神经网络的网络参数迭代调整。

对 E 关于 A_i, σ_i ($i=1, 2, \dots, N$)分别求偏导,得到:

$$\frac{\partial E}{\partial A_i} = \sum_{j=1}^N [f(j) - f^*(j)] e^{-\frac{(j-i)^2}{2\sigma_i^2}}. \quad (6)$$

$$\frac{\partial E}{\partial \sigma_i} = \sum_{j=1}^N [f(j) - f^*(j)] e^{-\frac{(j-i)^2}{2\sigma_i^2}} \frac{A_i(j-i)^2}{\sigma_i^3}. \quad (7)$$

为了保证迭代收敛的稳定性,引入惯性因子 α 后的调整公式如下:

$$\Delta A_i(t+1) = -p_A \frac{\partial E}{\partial A_i} + \alpha_A \Delta A_i(t), \quad (8)$$

$$\Delta \sigma_i(t+1) = -p_{\sigma} \frac{\partial E}{\partial \sigma_i} + \alpha_{\sigma} \Delta \sigma_i(t), \quad (9)$$

其中: $\Delta A_i(t)$ 和 $\Delta A_i(t+1)$ 分别为 A_i 第 t 次和 $t+1$ 次迭代调整量, p_A 和 α_A 分别为其步长和惯性因子; $\Delta \sigma_i(t)$ 和 $\Delta \sigma_i(t+1)$ 分别为 σ_i 第 t 次和 $t+1$ 次迭代调整量, p_{σ} 和 α_{σ} 分别为其步长和惯性因子。

1.2 曲线拟合的非唯一性

在离散情况下,对任一实际光谱 $f^*(j)$ ($j=1, 2, \dots, N$),存在唯一反映真实情况的拟合光谱。此外,从纯粹数学的角度,总存在如下的拟合:

$$f(j) = \sum_{i=1}^N g_{i,\sigma}(j) = \sum_{i=1}^N f^*(i) e^{-\frac{(j-i)^2}{2\sigma^2}} \quad (10)$$

其中, $\sigma \rightarrow 0$ 且 $\sigma \neq 0$ 。 $f(j)$ 是由 N 个峰高为 $f^*(i)$, 峰宽“几乎”为 0 的一组处处存在的高斯函数 $g_{i,\sigma}$ ($i=1, 2, \dots, N$) 对原始光谱 $f^*(j)$ 进行的拟合。

$$\text{在 } \sigma \rightarrow 0 \text{ 且 } \sigma \neq 0 \text{ 条件下, } e^{-\frac{(j-i)^2}{2\sigma^2}} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases},$$

容易验证: $f(j) = f^*(j)$ 。

由此可见, 在一定的误差限内, 同一光谱曲线可由不同的一组高斯峰拟合, 即其拟合(表达)并不是唯一的。尽管如此, 在一定的先验知识或约束条件下, 有效地进行曲线拟合仍是可能的。

定义最小峰-峰间距(minimum real peak-peak interval, MRI) $R = \min(\lambda_{i+1} - \lambda_i)_{i=1, 2, \dots, n-1}$ (其中 λ_i 是第 i 个峰位, n 为总的峰个数) 是所有相邻峰位间距中最小者。最小最大峰强比记为 $\text{Thr} = \frac{A_{\min}}{A_{\max}}$, 其中 A_{\min} 和 A_{\max} 分别为谱图中最小和最大峰强。

为了能对重叠谱图进行正确的分峰拟合, 设定的可分离最小峰-峰间距(minimum separable peak-peak interval, MSD) r 必须小于或等于实际的最小峰-峰间距 R 。但 r 过小, 也会导致出现伪峰, 使谱峰的个数多于实际的个数, 从而也得不到正确的解析结果。因此, 分峰的关键之一是确定合适的可分离最小峰-峰间距 r 。结合最小峰-峰间距 R 和最小最大峰强比 Thr 这两个先验知识, 可以用于确定谱峰峰位, 以得到符合实际的分峰拟合。在下面的模拟光谱分峰试验中, 给出了在不同最小峰-峰间距条件下得到的不同解析结果。

2 调整策略和算法

在曲线拟合中, 作为优化的初始条件, $(A_i, \sigma_i)_{i=1, 2, \dots, N}$ 、 α 、 p 等初始值的确定和模型选择对迭代收敛速度和拟合结果的影响显而易见。对 $(A_i, \sigma_i)_{i=1, 2, \dots, N}$ 的初始值赋 $(0, 1)$ 间的随机数, 所有数值实验没有出现无法收敛的情况, 显示这样的选择是可行的。其次是调整步长 p 和惯性因子 α 大小的选择, 步长 p 过大, 虽有较快的收敛速度, 但也会导致收敛过程出现震荡, 可能很难或无法得到稳定的全局最优解; 过小的学习率又会因收敛过慢而增加迭代次数, 且跳出局部极值的能力很弱。因此在实际应用中需选择合适的 p 值, 试验显示, 选取 $p = 0.1 \sim 0.6$ 和 $\alpha = 0.03 \sim 0.3$ 能得到较好的收敛效果。

笔者提出的基于曲线拟合的重叠峰解析算法大

致过程如下:

1) 估计曲线拟合的迭代次数。

2) 计算 $E(r)$:

① 指定可分离最小峰-峰间距 r ;

② 确定峰位;

③ 确定峰强和峰宽;

④ 计算拟合误差 $E(r)$ 。

3) 确定合适的分峰区间:

① 相同初始条件下, 令 $r = 1, 2, \dots, M$, 由 2) 分别求得所有 $E(r)$;

② 由拟合误差曲线图 $E(r)$ 确定合适的分峰区间 SI, 并以此区间进行谱图的分峰拟合。

定义最大允许相对误差系数为:

$$k_e = \frac{\sum_{j=1}^N [f(j) - f^*(j)]^2}{\sum_{j=1}^N [f^*(j)]^2} = \frac{E_{\max}}{\sum_{j=1}^N [f^*(j)]^2}$$

其中, E_{\max} 为最大允许误差平方和。

基本迭代分两个阶段: 1) 确定峰位; 2) 优化峰强和峰宽。

2.1 估计迭代次数 T_{\max}

为避免设定迭代次数 T_{\max} 的盲目性, 对某一类型光谱进行解析前, 依如下步骤大致确定:

1) 初始化 $(A_i, \sigma_i)_{i=1, 2, \dots, N}$ 为 $(0, 1)$ 间的随机数, 步长 $p = 0.1$, 惯性因子 $\alpha = 0.3$, $\Delta A_i(1) = 0$, $\Delta \sigma_i(1) = 0$ ($i=1, 2, \dots, N$), $t=1$, 指定 E_{\max} 。

2) 由式(4)和式(5)得到拟合谱线和误差平方和函数 E 的值。

3) 判断 E 值是否小于 E_{\max} , 如果 $E \leq E_{\max}$, 则转步骤 6)。

4) 由式(6)和式(7)计算得到 E 关于 A_i, σ_i ($i=1, 2, \dots, N$) 的偏导。

5) 由式(8)和式(9)求 $\Delta A_i(t+1)$ 和 $\Delta \sigma_i(t+1)$ ($i=1, 2, \dots, N$), 并得到调整后的 A_i 和 σ_i 。 $t=t+1$, 返回步骤 2)。

6) $T_{\max} = t$, 结束。

以上算法是对原始光谱的“连续”拟合, 即无论峰强大小, 假设在所有点处都可能存在独立峰。这样得到的拟合光谱峰强大致与原始光谱曲线一致, 并且峰宽很窄。尽管这样的曲线拟合能很快收敛, 且误差能达到很小, 但并不是所希望的真实拟合。这一步仅对确定迭代次数 T_{\max} 提供参考。

2.2 确定峰位

在进行曲线拟合前, 需要首先确定谱峰的数量及峰位, 以避免出现无法收敛或得到虚假结果。设

定可分离最小峰-峰间距 r , 本阶段叠代次数 $T_{\max 1} = 0.4 \times T_{\max}$ 。流程如下:

1) 初始化 $(A_i, \sigma_i)_{i=1,2,\dots,N}$ 为 $(0, 1)$ 间的随机数, 步长 $p = 0.1$, 惯性因子 $\alpha = 0.3$, $\Delta A_i(1) = 0$, $\Delta \sigma_i(1) = 0$ ($i=1, 2, \dots, N$), $t=1$ 。

2) 由式(4)和式(5)得到拟合谱线和误差平方和函数 E 的值。

3) 迭代次数 t 是否超过 $T_{\max 1}$, 如果 $t > T_{\max 1}$, 则转步骤 8)。

4) 由式(6)和式(7)计算得到 E 关于 A_i, σ_i ($i=1, 2, \dots, N$) 的偏导。

5) 由式(8)和式(9)求 $\Delta A_i(t+1)$ 和 $\Delta \sigma_i(t+1)$ ($i=1, 2, \dots, N$), 并得到调整后的 A_i 和 σ_i 。

6) 由 A_i ($i=1, 2, \dots, N$) 确定 i 处是否为峰。记 $\text{Max}A_i$ 为以 i 为中心, $r \times e^{-k_1(1-\frac{t}{T_{\max 1}})}$ 为半径内的最大峰强值。如果 $A_i \geq \text{Max}A_i \times e^{-k_2(1-\frac{t}{T_{\max 1}})}$, 则 A_i 值保持不变; 否则如果 $A_i < \text{Max}A_i \times e^{-k_2(1-\frac{t}{T_{\max 1}})}$, 置 $A_i = 0$, 即认为 i 处无峰。

7) $t=t+1$, 返回步骤 2)。

8) 结束本阶段迭代。

在步骤 6) 中, 由峰强 A_i 的大小来确定 i 点是否为峰位。逐步加宽比较半径 $r \times e^{-k_1(1-\frac{t}{T_{\max 1}})}$ 和峰强阈值 $\text{Max}A_i \times e^{-k_2(1-\frac{t}{T_{\max 1}})}$, 经过该阶段最后一次迭代后, 在 A_i ($i=1, 2, \dots, N$) 中保留的峰位(峰强非零)最小峰-峰间距不小于 r 。由于采用形如 $K_k(t) = e^{-k(1-\frac{t}{T})}$ 的函数产生系数, 因此并非简单以峰强极大值处为峰位, 而是逐步扩大比较半径和峰强阈值, 直至迭代到第 $T_{\max 1}$ 次时, 才最终设定大部分峰强为 0。这样, 可以给足够的调整时间以确定非极大值峰位点。

2.3 确定峰强和峰宽

设定本阶段的迭代次数为 $T_{\max 2} = T_{\max} - T_{\max 1} = 0.6 \times T_{\max}$ 。

1) 经过第一阶段峰位确定后, 由式(4)和式(5)得到拟合谱线, 并计算误差平方和函数 E 的大小。

2) 判断 E 值和迭代次数 t 是否小于设定的最大误差值 E_{\max} 或者最大迭代次数 $T_{\max 2}$, 如果 $E \leq E_{\max}$ 或 $t > T_{\max 2}$, 则转步骤 7)。

3) 由式(6)和式(7)计算得到 E 关于 A_i, σ_i ($i=1, 2, \dots, N$) 的偏导。

4) 由式(8)和式(9)求 $\Delta A_i(t+1)$ 和 $\Delta \sigma_i(t+1)$ ($i=1, 2, \dots, N$), 并得到调整后的 A_i 和 σ_i 。

5) $\text{Max}A$ 为 A_i ($i=1, 2, \dots, N$) 中最大者, 如果

$A_i < \text{Thr} \times \text{Max}A$, 则 $A_i = 0$ 。

6) $t=t+1$, 返回步骤 2)。

7) 结束本阶段迭代。

经过该阶段最后一次迭代后, 不仅在 A_i ($i=1, 2, \dots, N$) 中保留的最小峰-峰间距不小于 r , 且对峰强和峰宽经过调整, 使得其拟合误差最小。

2.4 基于拟合误差曲线确定可分离最小峰-峰间距

记 $E(r)$ 为可分离最小峰-峰间距等于 r 时的拟合误差。在相同初始化情况下, 由不同 r 得到的拟合误差 $E(r)$ 绘制的图称为拟合误差曲线。由于噪声的存在, 理想的分峰拟合也无法使其拟合误差为 0。

r 的大小决定了拟合谱峰数。选取的 r 过大 ($r > R$) 会导致拟合谱峰少于实际谱峰数, 丢失(遗漏)部分谱峰, 且使得 $E(r)$ 相对较大, 无法收敛得到满意的允许误差, 该区域称为过大区 (too large interval, TLI); 设定的 r 过小, 会出现伪峰, 伪峰不仅无损反而有利于曲线拟合, 因此可能达到理想分峰都不可能达到的拟合误差, 该区域称为过小区 (too small interval, TSI)。可见过小的拟合误差并非可取, 可能暗示有伪峰的出现。当 r 取值介于过小区和过大区之间时, 没有伪峰出现也无谱峰遗漏, 该区域称为适中区 (suitable interval, SI)。此时的拟合误差可能会等于或稍大于过小区的拟合误差, 在向过大区过渡时则会因漏峰而出现 $E(r)$ 的陡增。即 $E(r)$ 曲线中出现大斜率(陡变)处, 可能暗示有峰的丢失, 这可用于确定适中区和过大区的临界 r 值。但并非所有 $E(r)$ 图必须有这 3 个区域, 如果实际谱图只有一个峰, 则不会出现过大区; 反之, 如果实际谱图峰-峰间距很小, 则可能没有过小区; 但无论如何, 总存在适中区, 且一般可结合一些先验知识加以确定。不同初始情况下得到的拟合误差曲线 $E(r)$ 可能并不完全相同, 相应的 3 个区间(过小区、适中区和过大区)也可能会有所变化。

由于很难判断和保证单个 r 值下的曲线拟合是否正确, 因此要通过拟合误差曲线 $E(r)$ 确定 r 的适中区以正确分峰。实验表明, 同一初始条件下得到的拟合误差曲线 $E(r)$ 确定的适中区, 都有正确的分峰。

3 重叠光谱解析试验

3.1 模拟光谱

为了验证以上重叠峰解析算法的有效性, 笔者构造了图 1 所示由 6 个高斯峰和随机噪声叠加而成

的模拟光谱 $f(j)(j=1,2,\dots,300)$:

$$f(j) = 5e^{-\frac{(j-50)^2}{2 \times 3^2}} + 2e^{-\frac{(j-100)^2}{2 \times 1^2}} + 2e^{-\frac{(j-150)^2}{2 \times 6^2}} + 2e^{-\frac{(j-160)^2}{2 \times 6^2}} + e^{-\frac{(j-200)^2}{2 \times 8^2}} + 1.3e^{-\frac{(j-220)^2}{2 \times 6^2}} + 0.1 \times \text{Noise}(j)。$$

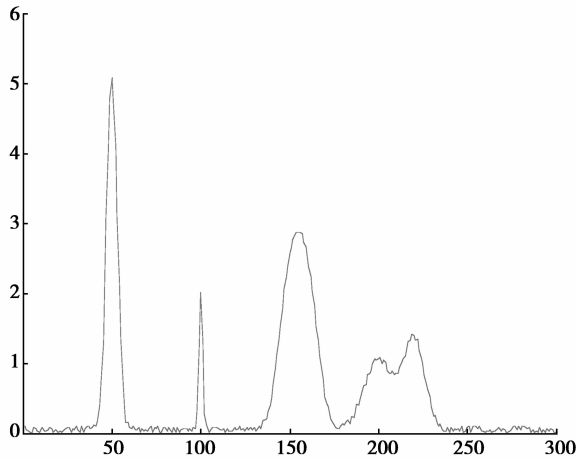


图 1 包含 6 个 Gaussian 峰和随机噪声叠加而成的模拟光谱

3.2 结果与讨论

采用前述算法,部分参数设定如下:衰减因子 $k_1=1, k_2=15$;最大最小峰强比 $\text{Thr}=0.01$;最大相对误差系数 $k_c=0.001$ 。

3.2.1 确定 T_{\max}

迭代误差曲线如图 2 所示。

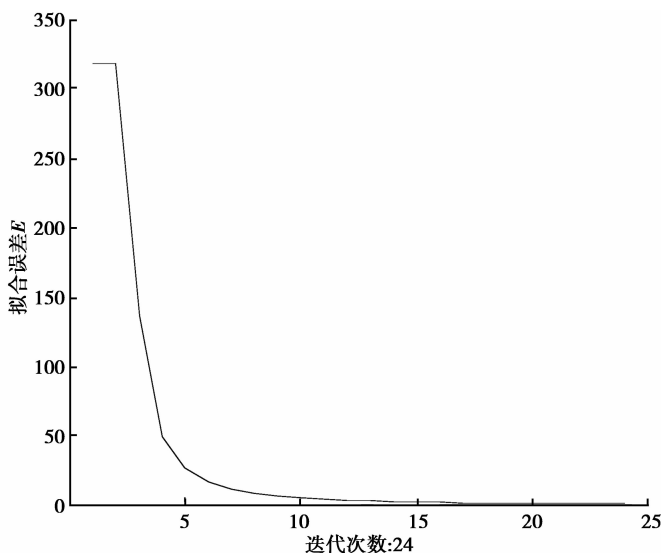


图 2 迭代误差曲线

当迭代次数为 24 次时,可得到满足误差要求的“连续”拟合曲线。给出一定余量后,设定 $T_{\max}=100$ 。

3.2.2 拟合误差曲线 $E(r)$ 及分峰

相同的初始化条件下,取 $r=1,2,\dots,20$,得到 $E(r)$ 曲线,如图 3 所示。

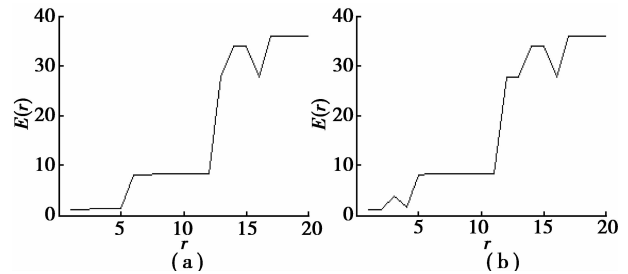


图 3 不同初始条件下得到的两组拟合误差 $E(r)$ 曲线图

不同初始化情况下,得到的 $E(r)$ 曲线及相应的 3 个区间确实稍有不同。 $r=12$ 在图 3(a) 中属于适中区,能正确分峰;而在图 3(b) 中则属于过大会,出现漏峰。 $r=5$ 也类似,在图 3(a) 中属于过小区,出现伪峰;在图 3(b) 中则属于适中区,能正确分峰。因此,无法仅由 r 值判断分峰正确与否,而应该结合 $E(r)$ 曲线看其是否属于适中区。根据以上讨论的 $E(r)$ 曲线特征确定图 3(a) 的 3 个区间:过小区($r < 6$)、适中区($6 \leq r \leq 12$)和过大会($r > 12$)。其中过小区($r < 6$)都出现伪峰,过大会($r > 12$)都出现漏峰,仅适中区($6 \leq r \leq 12$)能正确分峰。图 4 所示为 $r=8$ 的分峰结果,表 1 为曲线拟合结果(峰位、峰强和峰宽)与真实值的对比。

表 1 曲线拟合分峰结果(峰位、峰强和峰宽)与真实值对比表

峰位 μ		峰强 A		峰宽 σ	
理想值	拟合值	理想值	拟合值	理想值	拟合值
50	50	5	4.745 6	3	3.040 2
100	100	2	2.072 9	1	1.012 0
150	149	2	1.970 6	6	5.780 2
160	160	2	2.224 7	6	6.054 9
200	200	1	1.068 8	8	6.888 2
220	219	1.3	1.379 0	6	6.404 9

$E(8)=8.239 9 > 0$,原因是原始光谱信号中含有随机噪声,不可能得到 $E=0$ 的拟合结果。从分峰结果看,没有重叠的峰 1 和 2 的峰位、峰强和峰宽的误差最小,有部分重叠的次之,完全重叠的误差最大。在峰位、峰强和峰宽 3 个指标中以峰位的定位精度最高,峰强次之。

如图4(c)所示在迭代次数接近100处(第一阶段迭代终止时刻)出现误差陡升,原因是函数 $K_k(t) = e^{-k(1-\frac{t}{r_{\max}})}$ 的作用,使迭代次数 t 越接近100处,更多的非峰位处 A 值强制置0,从而导致误差的增大。这种误差的陡升属正常现象。从100开始,进入第二阶段的迭代,误差又逐渐减小。大量试验表明,没有出现误差发散或振荡等不能收敛的情况,说明算法具有很好的稳定性和可靠性。

图5为 $r=5$ 和 $r=20$ 的分峰结果。

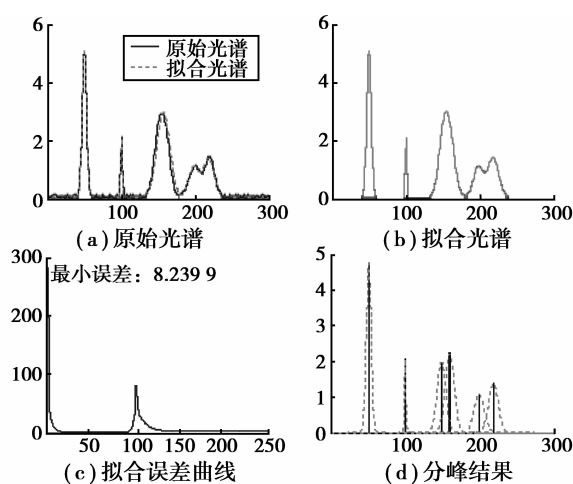


图4 $r=8$ 的分峰结果

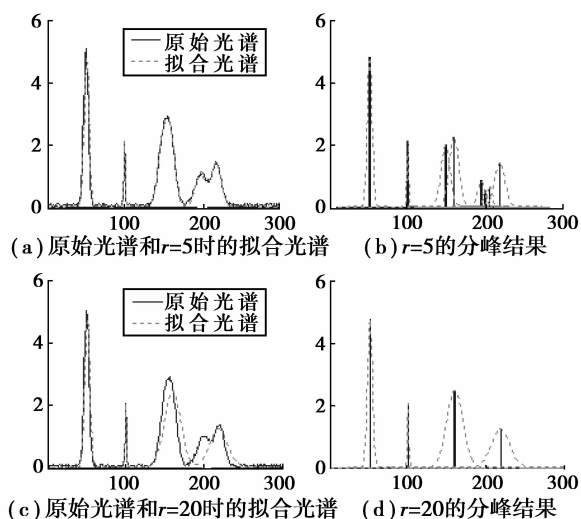


图5 $r=5$ 和 $r=20$ 的分峰结果

其中, $E(5) = 1.4304$, $E(20) = 35.839$ 。因原谱图中实际最小峰-峰间距 $R = 160 - 150 = 10$ (第3、4两峰间距),故 $r=5$ 或 $r=8$ 都能很好检测出所有谱峰峰位;可分离最小峰-峰间距 $r=20$ 过大,无法对第3、4和5、6峰进行分峰; $r=5$ 过小,导致出现多个伪峰。

3.2.3 分离度的影响

模拟光谱信号的第3、4峰严重重叠,不再能简单地以局部极大值为峰位。能否准确对该类重叠峰进行解析,可以作为各种分峰算法有效性的衡量标准。笔者采用文献[16]定义的分离度 S 衡量相邻峰的重叠程度, S 越小,表明两峰重叠越严重。大量实验表明,笔者算法能准确对 $S > 0.2$ 的重叠峰加以解析。

4 结论

笔者给出一种基于曲线拟合的光谱重叠峰解析算法,详细讨论了迭代次数、峰位、峰强和峰宽等参数调整和表达非唯一性等问题,提出可分离最小峰-峰间距和拟合误差曲线的3个特征区间(过小区、适中区和过大区)等概念,并基于此实现重叠峰解析。含噪模拟光谱解析结果显示,笔者提出的算法具有通用性强、不过分依赖于分析样品成分、含量等先验知识的优点,可对具有较小分离度的重叠峰进行有效解析,有望应用于包括色谱和光谱在内的各类谱图的重叠峰解析中。

参考文献:

- [1] 张秀琦, 刘辉, 郑建斌, 等. 信号处理技术在重叠化学信号解析中的应用[J]. 化学进展, 2002, 14(3): 174-181.
ZHANG XIU-QI, LIU HUI, ZHENG JIAN-BIN, et al. The application of signal processing in resolving overlapping chemical bands[J]. Progress in Chemistry, 2002, 14(3): 174-181.
- [2] HU Y G, LI W, HU J M. Resolving overlapped spectra with curve fitting [J]. Spectrochimica Acta Part A, 2005, 62(1/3): 16-21.
- [3] KOMATSUBARA M, NAMAZU T, NAGAI Y, et al. Raman spectrum curve fitting for estimating surface stress distribution in single-crystal silicon microstructure [J]. Japanese Journal of Applied Physics, 2009(48): 211-217.
- [4] 张晓星, 姚尧, 唐炬, 等. 导电微粒局部放电下 SF_6 分解组分数谱信号的曲线拟合分峰[J]. 电工技术学报, 2010, 25(7): 179-185.
ZHANG XIAO-XING, YAO YAO, TANG JU, et al. Separating overlapped chromatogram signals of SF_6 decomposed products under PD of conductive particles based on curve-fitting [J]. Transactions of China Electrotechnical Society, 2010, 25(7): 179-185.
- [5] BRELSTAFF G, BICEGO M, CULEDDU N, et al.

- Bag of peaks: interpretation of NMR spectrometry [J]. *Bioinformatics*, 2009, 25(2): 258-264.
- [6] CHUNG K S, CHOE H S, LEE J I, et al. A computer program for the deconvolution of thermoluminescence glow curves [J]. *Radiation Protection Dosimetry*, 2005, 115(1/4): 345-349.
- [7] 何星星, 贾建, 何秀丽, 等. 用于重叠峰分辨的卡尔曼滤波法[J]. *传感器与微系统*, 2006, 25(12): 26-31.
HE XING-XING, JIA JIAN, HE XIU-LI, et al. A method for resolving overlapped spectra based on Kalman filter [J]. *Transducer and Microsystem Technologies*, 2006, 25(12): 26-31.
- [8] 时有明, 刘刚, 周湘潭, 等. 基于曲线拟合的形态相似鹅膏菌的傅里叶变换红外光谱研究[J]. *分析化学*, 2008, 36(8): 1105-1108.
SHI YOU-MING, LIU GANG, ZHOU XIANG-PING, et al. Studies on similar amanita mushrooms by fourier transform infrared spectroscopy based on curve-fitting analysis [J]. *Chinese Journal of Analytical Chemistry*, 2008, 36(8): 1105-1108.
- [9] ZEMBATY Z. Discussion on: "Kalman filtering for neural prediction of response spectra from mining tremors" [J]. *Computers and Structures*, 2009, 87(13/14): 948-949.
- [10] 罗立强, 詹秀春. 应用遗传算法拟合偏振 X 射线荧光重叠谱[J]. *光谱学与光谱分析*, 2008, 28(3): 704-706.
LUO LI-QIANG, ZHAN XIU-CHUN. Resolution of overlapped spectra in polarization X-Ray fluorescence spectrometry by genetic algorithm [J]. *Spectroscopy and Spectral Analysis*, 2008, 28(3): 704-706.
- [11] ZHANG X Q, ZHENG J B, GAO H. Curve fitting using wavelet transform for resolving simulated overlapped spectra [J]. *Analytica Chimica Acta*, 2001 (443): 117-125.
- [12] NIE L, WU S G, WANG J W, et al. Continuous wavelet transform and its application to resolving and quantifying the overlapped voltammetric peaks [J]. *Analytica Chimica Acta*, 2001(450): 185-192.
- [13] RAMAKRISHNAN S, SELVAN S. Image texture classification using wavelet based curve fitting and probabilistic neural network [J]. *International Journal of Imaging Systems and Technology*, 2007, 17(4): 266-275.
- [14] 王佩佩, 宋晓峰, 杨平. 利用基于小波特征提取的网络模型解析色谱重叠峰[J]. *计算机与应用化学*, 2007, 24(5): 673-677.
WANG PEI-PEI, SONG XIAO-FENG, YANG PING. The application of network model based on wavelet feature extraction in resolution of overlapped chromatographic peaks [J]. *Computers and Applied Chemistry*, 2007, 24(5): 673-677.
- [15] GUTIERREZ J M, GUTES A, CESPEDES F, et al. Wavelet neural networks to resolve the overlapping signal in the voltammetric determination of phenolic compounds [J]. *Talanta*, 2008, 76(2): 373-381.
- [16] 张磊, 江海河, 储焰南, 等. 小波变换在重叠化学信号解析中的应用[J]. *理化检验-化学分册*, 2008, 44(3): 212-215.
ZHANG LEI, JIANG HAI-HE, CHU YAN-NAN, et al. Application of wavelet transformation to the resolution of overlapped chemical signals[J]. *Physical Testing and Chemical Analysis Part B: Chemical Analysis*, 2008, 44(3): 212-215.

(编辑 王维朗)