

## 拟牛顿法在谱线分峰中的应用研究

蒋夕平<sup>1</sup>, 吴凤凰<sup>1</sup>, 修连存<sup>2</sup>

(1. 南京农业大学 理学院, 江苏 南京 210095; 2. 南京地质矿产研究所, 江苏 南京 210016)

**摘 要:** 应用拟牛顿法对离散数据谱线进行分峰处理, 从而更精确地分析被测物质组分及组分含量. 应用最小二乘法对离散数据曲线进行曲线拟合, 用 C# 创建数值计算类, 并通过拟牛顿法求解非线性方程组, 实现对离散数据曲线的分峰处理, 研究非线性方程组变量迭代初始值误差和迭代参数值对迭代计算结果的影响. 通过对离散数据曲线进行分峰处理, 得到了独立的子峰 Lorentzian 函数及其峰高、半高宽、峰位, 并得到了非线性方程组变量迭代初始值误差和迭代参数值对迭代计算结果的影响规律. 结果表明, 对被测物质离散数据谱线作分峰处理, 可更准确计算被测物质组分及组分含量, 有较大实际运用价值.

**关键词:** 曲线拟合; 分峰; 拟牛顿法; 迭代; Lorentzian 函数

**中图分类号:** TP319

**文献标志码:** A

### 0 引言

现代科学研究积累了大量测量数据, 其中许多是波谱数据, 如高光谱遥感图谱、岩芯反射光谱、分析化学中的色谱和光谱、X 射线谱和核磁共振谱、作物反射光谱等. 由于被测物质组分多样, 不同组分谱线的谱峰可能比较接近而且谱峰的宽度较大, 致使不同组分谱线的谱峰之间发生重叠, 有些谱峰不完全重叠, 各谱峰尚有峰尖可辨, 有些谱峰近于完全重叠, 各谱峰的峰尖无法辨识. 由于谱线的谱峰重叠, 给被测物质组分的定性和定量分析造成了很大困难, 要对谱峰重叠的谱线有全面准确的认识, 对谱线进行分峰处理是十分必要的.

谱线分峰是利用计算机处理数字化的离散数据, 用合适的子峰函数通过最小二乘法, 来逼近实际谱线的重叠谱峰<sup>[1]</sup>. 以往谱线分峰多采用傅立叶变换去卷积方法<sup>[2]</sup>、导数分峰法<sup>[3]</sup>等, 目前谱线分峰的数值计算方法主要有曲线拟合法<sup>[4]</sup>、小波分析法<sup>[5]</sup>和人工神经网络法<sup>[6]</sup>等.

国内外已有谱线分峰研究报道, 其中有些将谱线分峰技术应用于高光谱遥感图谱处理<sup>[7]</sup>, 有些应用于色谱、光谱、电化学曲线处理<sup>[8]</sup>, 也有研究人员将谱线分峰技术用于研究作物反射光谱,

监测与诊断作物生长状况<sup>[9]</sup>, 这些研究应用的领域及研究侧重点不同, 使用的方法也有较大差异. 本研究应用最小二乘法对离散测量数据进行曲线拟合, 通过拟牛顿法求解非线性方程组得到各个独立的子峰函数, 由于非线性方程组不能直接求解, 只能通过迭代法求出变量的近似解, 为了求得变量近似解, 必须给出各变量迭代初始值, 并选择合适的迭代参数, 变量迭代初始值误差和迭代参数值对迭代计算是否收敛以及收敛速度有重要影响, 本研究对此进行了深入探讨.

### 1 研究方法

本研究处理的谱线谱峰比较尖锐, 由波谱技术可知应采用 Lorentzian 函数作为拟合子峰函数, 某个子峰函数可表示为

$$L_i(x_k) = \frac{A_i}{1 + \left(\frac{x_k - P_i}{B_i}\right)^2} \quad (1)$$

式中:  $k$  为谱线某个采样点,  $i = 1, 2, \dots, n$ , 共有  $n$  个子峰曲线, 谱线在某一采样点的测量值  $R(x_k)$  可以表示为

$$R(x_k) \approx L(x_k) = \sum_{i=1}^n L_i(x_k) \quad (2)$$

每个子峰函数都有 3 个参数: 峰高  $A_i$ 、半高宽

收稿日期: 2011-05-08; 修订日期: 2011-08-13

基金项目: 国土资源部机载高分辨率矿物成像光谱仪研制(1212010916030)

作者简介: 蒋夕平(1963-), 男, 江苏金坛人, 南京农业大学讲师, 硕士, 主要从事计算机仿真与数据建模研究, E-mail: jiangxp@njau.edu.cn.

$B_i$  和峰位  $P_i$ . 因此  $L(x_k)$  共有  $3n$  个参数, 对  $R(x_k)$  与  $L(x_k)$  差平方求和得到

$$S = \sum_{k=1}^m [R(x_k) - L(x_k)]^2 \quad (3)$$

式中:  $m$  为谱线采样点数, 要得到最佳的拟合曲线, 必须使  $S$  对每个参数的偏微分等于 0, 也就是

$$\frac{\partial S}{\partial A_i} = 0, (i = 1, 2, \dots, n); \frac{\partial S}{\partial B_i} = 0, (i = 1, 2, \dots, n); \frac{\partial S}{\partial P_i} = 0, (i = 1, 2, \dots, n), \text{ 计算并化简后得}$$

$$\sum_{k=1}^m [R(x_k) - L(x_k)] \cdot \frac{1}{1 + (\frac{x_k - P_i}{B_i})^2} = 0,$$

$$(i = 1, 2, \dots, n) \quad (4)$$

$$\sum_{k=1}^m [R(x_k) - L(x_k)] \cdot \frac{1}{[1 + (\frac{x_k - P_i}{B_i})^2]^2} \cdot$$

$$\frac{x_k - P_i}{B_i} = 0, (i = 1, 2, \dots, n) \quad (5)$$

$$\sum_{k=1}^m [R(x_k) - L(x_k)] \cdot \frac{1}{[1 + (\frac{x_k - P_i}{B_i})^2]^2} \cdot$$

$$\frac{(x_k - P_i)^2}{B_i^3} = 0, (i = 1, 2, \dots, n) \quad (6)$$

以上非线性方程组不能直接求解, 只能用迭代法求出变量的近似解. 将(4)、(5)、(6)非线性方程组简记为  $f_i(x_1, x_2, \dots, x_{3n}) = 0, (i = 1, 2, \dots, 3n)$  或  $f_i(X) = 0, (i = 1, 2, \dots, 3n)$ , 其中  $X = (x_1, x_2, \dots, x_{3n})^T$ , 下面是拟牛顿法求解非线性方程组的主要步骤:

(1) 取变量初始值,  $X = (x_1, x_2, \dots, x_{3n})^T$ ;

(2) 计算非线性方程组  $f_i(X)$  的值, 将计算结果保存在一维矩阵  $B(i)$  中,  $i = 1, 2, \dots, 3n$ ;

(3) 若满足  $\max_{1 \leq i \leq 3n} |B(i)| < \varepsilon$ , 其中  $\varepsilon$  为迭代计算精度, 则非线性方程组解为  $X = (x_1, x_2, \dots, x_{3n})^T$ , 迭代计算结束, 否则继续;

(4) 计算  $f_i(X_j)$ , 并将计算结果保存在两维矩阵  $A(i, j)$  中, 其中  $i, j = 1, 2, \dots, 3n$ ,  $X_j = (x_1, \dots, x_{j-1}, x_j + h, x_{j+1}, \dots, x_{3n})^T$ ,  $h$  为变量增量;

(5) 解线性方程组  $A \cdot Z = B$ , 得到解  $Z = (z_1, z_2, \dots, z_{3n})^T$ , 计算  $\beta = 1 - \sum_{j=1}^{3n} z_j$ ;

(6) 计算变量新迭代值  $x_i = x_i - \frac{hz_i}{\beta}, (i = 1, 2, \dots, 3n)$ ;

(7) 改变变量增量  $h = h \cdot t$ , 其中  $0 < t < 1$  为控制  $h$  大小的参数;

(8) 重复(2)以后的各步骤, 直到求得非线性方程组解或达到了设定的迭代次数.

## 2 实现技术

本研究在 Visual Studio 2008 中实现数值计算以及计算结果输出、显示, 编程语言为 C#. 要实现谱线分峰, 首先采用最小二乘法对离散数据曲线进行曲线拟合, 再应用拟牛顿法求解非线性方程组, 定义非线性方程组类 NonlinearEquationClass, 通过该类的 GetValuesetNewton 方法求解非线性方程组<sup>[10]</sup>, 该方法的返回值为数组, 数组元素都是 double 类型, 如果求解成功, 数组元素就是非线性方程组的变量近似解, 否则数组元素都为 0. GetValuesetNewton 方法输入参数有 iNumEquation (方程数目, 也就是变量数目)、x (变量迭代初始值)、t (变量增量调整因子)、h (变量增量)、iNumber (最大迭代次数)、dAccuracy (迭代控制精度), 输出参数 sInfo 包含了迭代计算信息, 如实际迭代计算次数、迭代计算时间等.

在 GetValuesetNewton 方法中, 调用了自身类的 Func 方法, 用于计算非线性方程组各个方程的值, 另外还调用了矩阵类 MatrixClass 的 GetData 方法和线性方程组类 LinearEquationClass 的 GetRootsetGauss 方法, GetData 方法用于根据矩阵的大小创建数组, 并将数组元素与矩阵元素关联, GetRootsetGauss 方法则根据线性方程组的系数矩阵、常数矩阵求解线性方程组.

## 3 计算结果

计算中使用的离散数据为岩芯吸收光谱, 波长范围为 1 300 ~ 1 600 nm, 采样间隔为 1 nm, 谱线分峰主要通过 NonlinearEquationClass 类的 GetValuesetNewton 方法完成, 根据给定的子峰数量, 对该方法的输入参数赋值, 设定最大迭代次数、变量增量、变量增量调整因子、迭代控制精度 eps, 程序进行谱线分峰计算, 谱线分峰计算结果见表 1, 离散数据曲线、拟合曲线及子峰 Lorentzian 曲线如图 1 所示.

图 1 中, 曲线 I 为离散数据曲线, 曲线 II 为拟合曲线, 即为 4 个子峰曲线的叠加曲线, 曲线 I、II 基本重合, 曲线 III、IV、V、VI 为分峰处理得到的 4 个子峰曲线, 它们的峰高、半高宽、峰位数值见表 1. GetValuesetNewton 方法的输出参数值表明, 这次迭代计算次数为 9 次, 迭代计算时间为 93 ms, 谱线分峰计算所用的计算机为普通 PC 机,

CPU 为 Intel Pentium III Xeon,主频 2.53 GHz,内存 2 G. 不论离散数据曲线是否含重叠谱峰,都能很好进行分峰(见图 1,2),该方法速度快、精度高,在大数据量或数据实时处理场合实用价值较大.

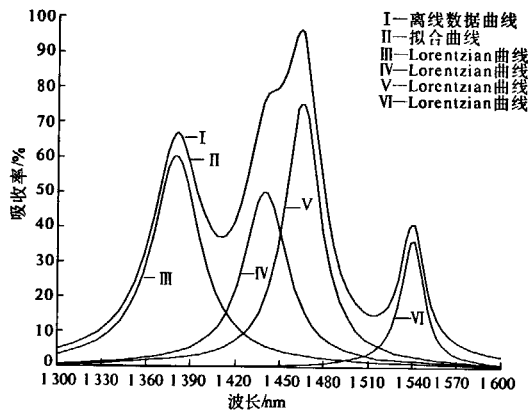


图1 含重叠峰的岩芯吸收光谱曲线拟合及分峰结果  
Fig.1 Result of fitted peaks for absorbed spectra of core with overlapped peaks

4 结果讨论

4.1 变量迭代初始值误差对迭代计算的影响  
用表 1 中非线性方程组变量近似解作为变量迭代初始值,改变某个变量迭代初始值,研究该变量迭代初始值误差对迭代计算的影响.

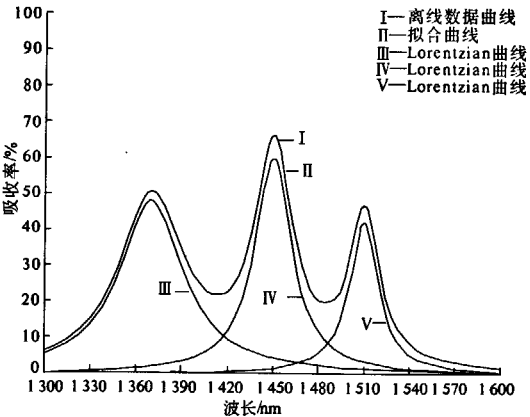


图2 不含重叠峰的光谱曲线拟合结果  
Fig.2 Result of fitted peaks for absorbed spectra without overlapped peaks

表 1 谱线分峰计算结果

Tab.1 Result of peak separation for the spectra			
序号	峰高	半高宽/nm	峰位/nm
1	60.155	19.965	1 379.9
2	50.164	17.847	1 439.9
3	75.227	14.931	1 465
4	36.023	10.024	1 540

4.1.1 峰高的影响  
实验表明,峰高允许的迭代初始值误差范围较大,可以大于  $\pm 100\%$ ,因此选取峰高迭代初始值比较容易,即使误差较大,仍然可以求得非线性方程组的解.

4.1.2 半高宽的影响  
半高宽迭代初始值误差不能太大,一般在 +3 到 -10 范围,否则迭代计算不收敛,即无法求得非线性方程组的解.从实验数据可以看到,半高宽的迭代初始值允许误差范围相对于变量近似解不对称,在小于变量近似解的一边允许误差范围较大,因此半高宽迭代初始值应该取较小的值,这样更容易求得非线性方程组的解.

4.1.3 峰位的影响  
峰位迭代初始值误差对迭代结果影响也比较大,误差允许范围在 +12 到 -12 之间,峰位迭代初始值误差允许范围相对于变量近似解基本对称.  
实际应用中,峰高迭代初始值允许的误差范围较大,根据离散测量数据比较容易确定峰位,因此峰高、峰位的迭代初始值比较容易得到,相对而言确定半高宽的迭代初始值比较困难,一般首先估算出它的迭代初始值,如果计算不收敛,则减小半高宽迭代初始值继续迭代计算.

4.2 迭代参数值对迭代计算的影响  
拟牛顿法求解非线性方程组的迭代参数有最大迭代次数  $n$ 、变量增量  $h$ 、变量增量调整因子  $t$ 、迭代控制精度  $eps$ ,改变迭代参数值并进行迭代计算,表 2 是实验结果.

表 2 迭代参数值对迭代计算的影响

Tab.2 The value of parameters effect on the iterated calculation							
序号	最大迭代次数 $n$	变量迭代初值增量 $h$	变量增量调整因子 $t$	迭代控制精度 $eps$	实际迭代次数	非线性方程组求解时间/ms	是否有解
1	100	0.1	0.1	0.000 000 1	8	78	是
2	100	0.3	0.3	0.000 000 1	8	78	是
3	100	0.6	0.6	0.000 000 1	12	125	是
4	100	0.8	0.8	0.000 000 1	100	984	否

从实验数据看到,  $h, t$  应取适当的值, 一般在 0.1 ~ 0.3, 如果它们取值不合理, 将增加实际迭代计算次数, 从而影响迭代计算速度, 甚至无解.  $eps$  的取值根据实际需要决定,  $eps$  取得越小, 求得的非线性方程组解精度越高, 但是迭代计算时间较长, 如果  $eps$  太小, 也可能得不到非线性方程组的解. 最大迭代次数  $n$  不要太大, 实验发现, 在能够求得非线性方程组的解时, 实际的迭代次数一般不超过 30 次.

## 5 结论

通过 C# 编程, 创建各种数值计算类用于曲线拟合、分峰, 再应用拟牛顿法求解非线性方程组, 得到了独立的子峰 Lorentzian 函数, 通过研究变量初始值误差和迭代参数值对迭代计算结果的影响规律, 帮助选择适当的变量迭代初始值和迭代参数值, 从而方便进行曲线分峰, 该研究对高光谱遥感图谱分析、矿物光谱分析、化学谱线分析及其他领域的波谱处理都有一定的参考价值.

## 参考文献:

- [1] 张箭, 李宏平, 戴明. 计算机分峰方法的设计和实现[J]. 河南科学, 1992, 10(4): 360 - 365.
- [2] PARENTE M, MAKAREWICZ H, BISHOP J L. Decomposition of mineral absorption bands using nonlinear least squares curve fitting: Application to martian-meteorites and CRISM data[J]. Planetary and Space Science, 2011, 59(5/6): 423 - 442.
- [3] LI Yuan-lu, TANG Hui-qiang, CHEN Hai-xiu. Fractional-order derivative spectroscopy for resolving simulated overlapped Lorentzian peaks[J]. Chemometrics and Intelligent Laboratory Systems, 2011, 107(1): 83 - 89.
- [4] HU Yao-gai, LIU Jin-yang, LI Wei. Resolution of overlapped spectra by curve-fitting[J]. Analytica Chimica Acta, 2005, 538(1/2): 383 - 389.
- [5] ZHANG Xiu-qi, ZHENG Jian-bin, GAO Hong. Curve fitting using wavelet transform for resolving simulated overlapped spectra[J]. Analytica Chimica Acta, 2001, 443(1): 117 - 125.
- [6] ZHANG Ya-xiong, LI Hua, HOU Ai-xia, et al. Artificial neural networks based on principal component analysis input selection for quantification in overlapped capillary electrophoresis peaks[J]. Chemometrics and Intelligent Laboratory Systems, 2006, 82(1/2): 165 - 175.
- [7] 刘俊. 高光谱图像混合像元分类技术研究[D]. 北京: 中国地质大学信息工程学院计算机科学与技术教研室, 2010.
- [8] 张晓星, 姚尧, 唐炬, 等. 导电微粒局部放电下 SF6 分解组分色谱信号的曲线拟合分峰[J]. 电工技术学报, 2010, 25(7): 179 - 185.
- [9] 陆伟. 棉花病害的高光谱特性分析及其田间药效试验的研究[D]. 江苏: 江苏大学农业工程研究院, 2007.
- [10] 周长发. C#数值计算算法编程[M]. 北京: 电子工业出版社, 2007.

## Application of Quasi-Newton Method for Separating Overlapping Peaks of Spectra

JIANG Xi-ping<sup>1</sup>, WU Feng-huang<sup>1</sup>, XIU Lian-cun<sup>2</sup>

(1. College of Science, Nanjing Agricultural University, Nanjing 210095, China; 2. Nanjing Institute of Geology and Mineral Resources, Nanjing 210016, China)

**Abstract:** To analyze the components and their contents within the tested material more precisely, quasi-Newton method can be applied to separate overlapping peaks of spectra plotted with discrete data. Based on the curve fitted with discrete data using the least squares method, the independent peaks in overlapping peaks can be identified through solving non-linear systems via quasi-Newton method implemented with C#. The effect of initial value error and iteration parameters for the iterative method in solving non-linear systems will also be studied. After identifying independent peaks in overlapping peaks of spectra plotted with discrete data, the information of independent peaks can be analyzed, such as Lorentzian function, the height, the half-width at half-height, and the position of a peak. Some effect rules of initial value error and iteration parameters for the iterative method in solving non-linear systems have also been found. Separating overlapping peaks of spectra plotted with discrete data, the components and their contents within the tested material will be calculated more precisely, which has great application value.

**Key words:** curve fitting; separate overlapping peak; Quasi-Newton Method; iteration; Lorentzian function