

## 激光诱导击穿光谱数据特征自动提取方法研究

刘立拓, 刘建国, 赵南京\*, 鲁翠萍, 陈东, 石焕, 王春龙, 张玉钧, 刘文清

中国科学院安徽光学精密机械研究所, 中国科学院环境光学与技术重点实验室, 安徽 合肥 230031

**摘要** 根据激光诱导击穿光谱(LIBS)展宽主要为非线性 Lorentz 函数模型, 利用 Levenberg-Marquardt(L-M)算法对该理论模型中的待定参数进行优化估值, 扣除了光谱信号中的连续背景, 校正和还原了峰位和峰强, 并对拟合数据点自动提取方法进行了分析; 以 MatrixVB 与 VB 混合编程相结合, 实现了基于 L-M 算法的 LIBS 光谱数据自动拟合与特征参数提取, 与手动操作的常用数据分析软件 Origin7.5 的拟合结果进行了比较, 验证了该方法的稳定性与可靠性。

**关键词** 激光诱导击穿光谱; 特征提取; Levenberg-Marquardt 算法; 非线性拟合

中图分类号: O433.5 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2011)12-3285-04

### 引言

激光诱导击穿光谱技术(LIBS)是当前探测物质成分信息的一项新技术, 它是利用激光激发物质产生等离子体, 再利用适当的探测系统得到等离子体中含有物质成分的光谱信号, 经过数据处理提取物质成分信息。与传统的物质成分分析方法相比, LIBS 技术具有实时、快速、原位检测等特点。随着激光技术的飞速发展, LIBS 技术正逐渐走向成熟, 并在各个领域开展了广泛的研究, 正发展成为一种极具应用前景的物质成分检测技术<sup>[1-4]</sup>。

LIBS 技术的关键是对光谱信息的准确提取, 并建立光谱强度与物质浓度之间的关系。通过光谱仪采集到的光谱数据为离散数字信号, 是有用信号与噪声的叠加。LIBS 光谱存在连续的背景辐射, 为了准确的提取光谱信息, 必须扣除背景与噪声的影响, 校准峰位、还原峰强。等离子体形成过程中大量粒子之间相互碰撞, 谱线展宽主要为 Lorentz 型<sup>[5,6]</sup>。可以通过对离散的 LIBS 光谱数据进行 Lorentz 线型拟合来消除背景与噪声的影响, 提取相对较为“干净”的光谱信息。含有非线性参数的 Lorentz 函数拟合可以采用非线性最小二乘法。Gauss-Newton 法是一种对非线性参数进行最小二乘拟合的方法, 但该方法对初值的选取要求应充分接近最小二乘估计值, 否则迭代过程不容易收敛, 甚至发散。为了放宽

对初值的限制, 已提出不少改进方法, 其中 L-M 是较好和比较常用的一种, 大量实验分析说明, L-M 算法是在 Gauss-Newton 基础之上通过引入阻尼因子, 在迭代过程中动态调整阻尼因子的大小, 进而使迭代收敛性较好, 收敛速度比较快<sup>[8-10]</sup>。另外, 为了实现对选定的一条谱线自动拟合, 需要自动提取拟合数据点。考虑到由于分光系统的分辨率及环境温度的影响, 实验数据中波长跟标准谱线数据库中的波长存在偏差, 分光系统不同, 偏差也不同。又由于外界环境的影响, 光谱仪往往存在波长漂移, 漂移的大小及方向不但随外界环境变化, 而且还随波长变化而变化。在拟合数据点提取过程中, 如何让程序像人眼观察一样, 无论存在波长漂移还是噪声或其他峰的干扰都能准确的判断哪些是要提取的拟合数据点是一项很关键和重要的工作, 尤其是在光谱仪分辨率及采样间隔不甚理想的情况下, 拟合数据点代表了更多的信息, 其选取的是否准确变得更为关键。必须结合实验仪器的具体性能参数来做相应的处理。

利用 VB6.0 可以进行丰富友好的界面设计, 但其数学运算能力尤其是对矩阵运算及图形处理能力较差; MATLAB 虽具有强大的数值计算能力, 但其界面设计能力较差, 且不能脱离其环境运行; 在 VB 中加载 MatrixVB, 既保留了 MATLAB 的矩阵运算及图形处理的强大功能, 又集成了 VB 界面友好的特点, 本文的研究是以 MatrixVB 与 VB 混合编程相结合来实现 L-M 算法的迭代计算过程。

收稿日期: 2011-01-15, 修订日期: 2011-06-05

基金项目: 国家自然科学基金项目(60908018), 中国科学院知识创新工程重大项目(kzcx-lyw-06-01), 安徽省科技计划项目(083H3b1581)和所长基金项目(073H371141)资助

作者简介: 刘立拓, 1982 年生, 中国科学院安徽光学精密机械研究所博士研究生 e-mail: ltuoliu@aiofm.ac.cn

\* 通讯联系人 e-mail: njzhao@aiofm.ac.cn

### 1 实验装置

实验系统如图 1 所示, 采用 Nd:YAG 波长 1 064 nm 激光器, 脉宽为 6 ns, 脉冲重复频率为 1~20 Hz, 脉冲能量 50 mJ; 光谱仪(Ocean optics HR2000+)测量波长范围: 300~500 nm, 分辨率 0.1 nm, 两个聚焦镜均为石英透镜, 样品台由步进电机控制其转动。首先通过光谱仪的控制软件控制光谱仪向激光器发送 TTL 信号使其出光, 激光经过全反射镜垂直入射到聚焦镜, 再经聚焦镜聚焦到样品上, 产生的激光等离子体通过另外一个聚焦镜收集耦合到光谱仪光纤输入端, 最后通过光谱仪将探测到的离散的等离子体数字信号输入到上位机进行处理。

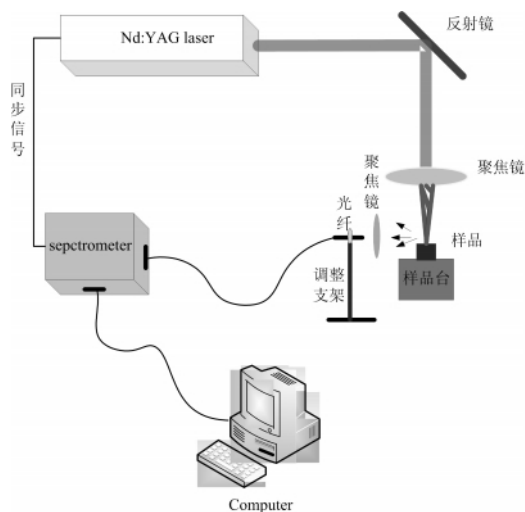


Fig 1 Experimental device

### 2 L-M 算法

L-M 算法是解决非线性拟合问题的较常用的一种方法, 其原理与过程如下: 设要进行拟合的非线性函数为:  $y = f(x; c)$ , 其中  $c$  为参数向量  $c = (c_1, c_2, \dots, c_m)^T$ ,  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ , 为拟合数据点, 根据最小二乘原理, 应对  $\chi^2 = \sum_{i=1}^n [y_i - f(x_i; c)]^2$  求极值, 直接求解  $\frac{\partial \chi^2}{\partial c_j} = 0$ , ( $j = 1, 2, \dots, m$ ) 很困难。需要对函数  $y = f(x; c)$  进行泰勒展开略去高次项得到

$$y = f(x; c) = f(x; c^{(0)}) + \left(\frac{\partial f}{\partial c_1}\right)_{c=c^{(0)}} \delta_1^{(1)} + \left(\frac{\partial f}{\partial c_2}\right)_{c=c^{(0)}} \delta_2^{(1)} + \dots + \left(\frac{\partial f}{\partial c_m}\right)_{c=c^{(0)}} \delta_m^{(1)} \quad (1)$$

其中  $c^{(0)} = (c_1^{(0)}, c_2^{(0)}, \dots, c_m^{(0)})^T$  为给定的参数初始值<sup>[10]</sup>。 $\delta_j^{(i)}$  ( $j = 1, 2, \dots, m$ ) 为参数增量,  $i$  为迭代次数。这样非线性拟合问题就转化为线性拟合, 增量用矩阵表示

$$\delta^{(i)} = (F^T F + D)^{-1} F^T (Y - Y_{(i-1)}) \quad i = 1, 2, \dots \quad (2)$$

$$Y_{(i-1)} = \begin{pmatrix} f(x_1; c^{(i-1)}) \\ f(x_2; c^{(i-1)}) \\ \vdots \\ f(x_n; c^{(i-1)}) \end{pmatrix} \quad F = \begin{pmatrix} f_{11} & f_{12} & \dots & f_{1m} \\ f_{21} & f_{22} & \dots & f_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ f_{n1} & f_{n2} & \dots & f_{nm} \end{pmatrix}$$

上式为 Gauss-Newton 迭代, L-M 算法在式(2)的基础上加上一个阻尼因子  $D$ , 如下式所示

$$\delta^{(i)} = (F^T F + D)^{-1} F^T (Y - Y_{(i-1)}) \quad (3)$$

$D$  为  $m$  阶对角矩阵。为了方便  $D$  取值, 设  $A = F^T F$ , 应将  $m$  阶方阵  $A$  标准化, 即将方阵  $A$  中元素  $a_{ij}$  用  $\frac{a_{ij}}{a_{ii} a_{jj}}$  代替, 为此, 构造对角矩阵  $B$ , 当  $i \neq j$  时,  $b_{ij} = 0$ , 当  $i = j$  时,  $b_{ij} = \frac{1}{\sqrt{a_{ij}}}$ <sup>[10]</sup>。标准化后矩阵  $\tilde{A} = BAB$ , 则  $A = B^{-1} \tilde{A} B^{-1}$ , 这样式(3)化为

$$\delta^{(i)} = (B^{-1}(\tilde{A} + D)B^{-1})^{-1} F^T (Y - Y_{(i-1)}) \quad (4)$$

拟合与各数据点之间的参差平方和为

$$\chi_i^2 = (Y - Y_{i-1} - F\delta^{(i)})^T (Y - Y_{i-1} - F\delta^{(i)}) \quad (5)$$

迭代过程如下:

- (1) 首先给出参数初值  $c^{(0)}$  和一足够小常数  $\epsilon$ , 并计算  $\chi_0^2$ , 对角阵  $D$  初始赋值为 0.001;
- (2) 根据(4)式计算参数增量  $\delta^{(1)}$ ;
- (3) 比较  $\chi_i^2 = \sum_{i=1}^n [y_i - f(x_i; c^{(0)} + \delta^{(1)})]^2$  与  $\chi_0^2$  的大小, 如果  $\chi_i^2 = \sum_{i=1}^n [y_i - f(x_i; c^{(0)} + \delta^{(1)})]^2 < \chi_0^2$ , 则令  $c^{(1)} = c^{(0)} + \delta^{(1)}$ , 若  $|\delta^{(1)}| < \epsilon$ , 则迭代截止, 输出结果。否则阻尼因子  $D$  缩小 10 倍, 转到步骤 2;

- (4) 如果  $\chi_i^2 = \sum_{i=1}^n [y_i - f(x_i; c^{(0)} + \delta^{(1)})]^2 > \chi_0^2$ , 则  $D$  扩大 10 倍, 回到步骤 1<sup>[11]</sup>。

从实验分析可知, L-M 算法收敛性较好,  $D$  越大, 迭代越易收敛。在收敛的情况下,  $D$  越小收敛速度越快。因此将  $D$  变化的取值, 这样可以达到最佳拟合效果。

### 3 拟合数据点自动提取

实验对山东掖县焦家黄棕土壤标准样品进行了分析, 重复测量 20 次累加平均, 光谱仪积分时间为 1 ms, 波长漂移范围约在 0.2 nm 以内, 根据实验分析可知, 其谱线展宽约为 0.15 nm。经过大量实验测量和研究表明, 选择拟合峰附近的一条强度比较大的主元素(内标元素)的谱峰作为波长漂移的参考, 可以有效的解决波长漂移随时空变化的问题。下面以 Fe; 421.618 nm 谱线拟合数据点提取为例, 如图 2 所示。利用临近谱线范围内强度较大的内标元素 Ca; 422.673 nm 作为波长漂移参考, 从图 2(a)中可知, 实验峰位位于标准波长中心 422.673 nm 的左边, 即: 小范围内测量谱线出现了“蓝移”, 因此可以判定 Fe 的谱线峰位在其标准波长中心 421.618 nm 的左边; 然后找到左边最近邻的最高实验数据点, 图 2(a)中“1”点, 以点“1”为中心向两边取点直到数据点强度大小顺序变化位置, 图 2(b)为需要拟合数据点的提取结果。用此方法对土壤中六种重金属进行了验证, 每种重金属选取五种不同浓度的样本, 每种样本重复测量 20 次, 结果准确率优于 98%, 因此该方法能够较准确的提取拟合数据点。

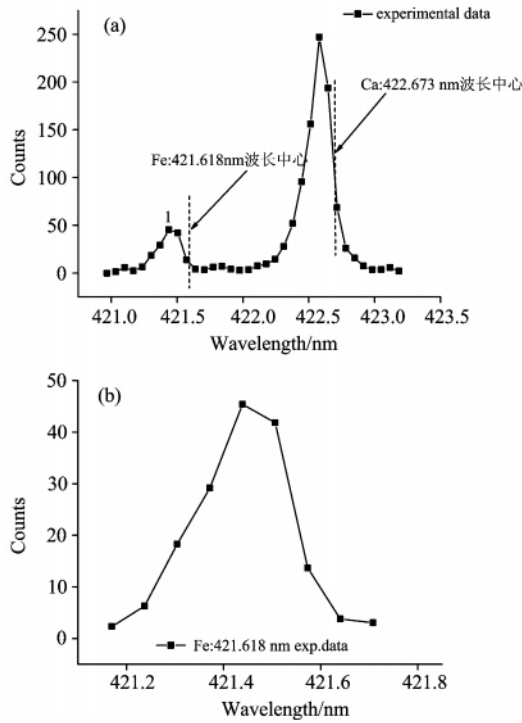


Fig 2 The progress and results of automatically extracted data for fitting Fe: 421.618 nm

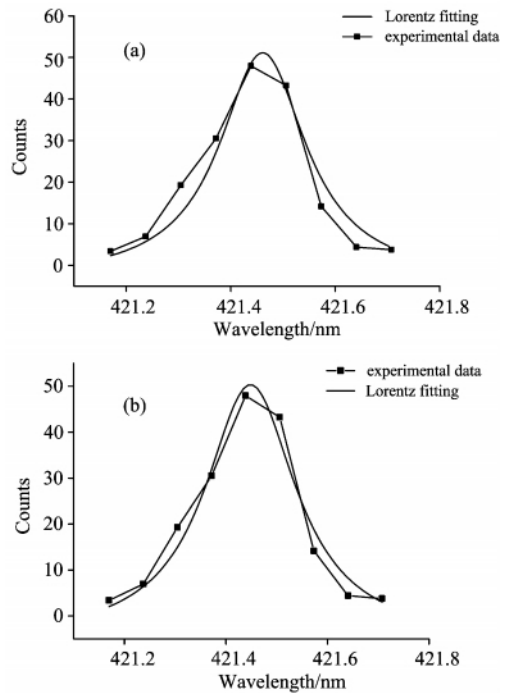


Fig 3 Comparison between fitting results

(a): Lorentz function fitting using L-M for Fe: 421.618 nm in soil;  
 (b): Lorentz function fitting using Origin7.5 for Fe: 421.618 nm in soil

## 4 数据拟合与结果分析

### 4.1 参数初值的选取

洛伦兹函数可表示为

$$I = \frac{2A\omega}{4\pi(x_i - \lambda)^2 + \pi\omega^2} + d \quad (6)$$

在拟合过程中  $A$ ,  $\omega$ ,  $\lambda$  和  $d$  分别为待定参数, 其中  $\omega$  为谱线半宽,  $\lambda$  为中心波长,  $d$  为背景基线。对于这四个参数初值的选取, 要依赖于前面所讲到的拟合数据点提取的方法, 对于  $A$  的初始值当  $x_i = \lambda$  时

$$A = \frac{\pi\omega(I-d)}{2} \quad (7)$$

式(7)中  $d$  的初始值估计为拟合数据点的左右两端点的平均值,  $I$  为数据点最高值与次高值的平均值。因此  $A$  的初始值可由(7)式确定。对于  $\lambda$  的初值选取可表示为

$$\lambda = \frac{2\max_1 + \max_2}{3} \quad (8)$$

式中  $\max_1$  与  $\max_2$  分别为拟合数据点中最大强度处波长与次大强度处波长。根据大量实验经验的总结,  $\omega$  的初值应取 0.15。

### 4.2 结果分析

实验对含 Fe 量为 1.79% 的标准土壤样品中 Fe: 421.618 nm 谱线进行了测量, 并进行了本文研究方法的测试, 自动提取了光谱特征参数信息, 并扣除了背景, 对峰位和峰强进行了校正和还原, 与 Origin7.5 数据软件手动分析结果进行了比较, 自动提取与手动分析结果分别如图 3 所示。

表 1 给出了 L-M 算法与 Origin7.5 拟合各参数的结果。

Table 1 The results of spectral feature parameters extraction

参数	L-M	Origin7.5
$\lambda$	421.461	421.449
$\omega$	0.202	0.219
$d$	2.549	0.393
$A$	17.365	19.140
$r$	0.989	0.963

从表 1 中可以看出, L-M 算法拟合结果中的各项参数均优于 Origin7.5 拟合结果, 验证了 L-M 算法的可靠性。同时, 实验对土壤中 Cu: 327.396 nm, Zn: 472.215 nm, Pb: 405.781 nm, Cd: 508.582 nm, Ni: 341.476 nm 等元素的 LIBS 光谱数据也分别进行了 L-M 算法拟合分析和特征提取, 拟合的相关系数均达到 0.98 以上。也验证了 L-M 算法的稳定性和可靠性; 整个迭代过程耗时约为 10 ms, 能够满足数据自动分析的需求。

## 5 结论

在 LIBS 光谱数据背景扣除、峰位和峰强校正还原及数据特征自动提取的过程中, 本研究与常用的数据分析软件 Origin7.5 手动拟合结果相比较, 不仅快速实现了光谱数据的自动拟合与特征参数的提取, 而且获得了更好的拟合结

果。在拟合数据点自动提取过程中, 需要考虑测量光谱波长漂移的问题, 以内标元素作为参考可以有效的解决波长变化对数据自动选取与拟合所带来的影响。本研究结果为进一步深入研究激光诱导击穿光谱技术用于土壤重金属多元素的快速、定量分析提供了基础。

## References

- [ 1 ] Christopher B Stipe, Brian D Hensley, Jeffrey L Boersema, et al. Applied Spectroscopy, 2010, 62(2): 154.
- [ 2 ] Tognoni E, Cristoforetti G, Legnaioli S. Spectrochim. Acta B, 2007, 62(12): 1287.
- [ 3 ] Kemal E Eseller, Markandey M Tripathi, Yueh Fang-Yu, et al. Applied Optics, 2010, 49(13): C21.
- [ 4 ] Rosalie A Multari, David A Cremers, Joanne M Dupre, et al. Applied Spectroscopy, 2010, 64(7): 750.
- [ 5 ] Writing Group of Emission Spectral Analysis(发射光谱分析编写组). Emission Spectral Analysis(发射光谱分析). Beijing: Metallurgical Industry Press(北京: 冶金工业出版社), 1977. 22.
- [ 6 ] David A Cremers, Leon J Radziemski. Handbook of Laser-Induced Breakdown Spectroscopy. Cambridge University, 2006. 25.
- [ 7 ] TAN Dong-lian, XIAO Ru-cheng(谭冬莲, 肖汝诚). Journal of Traffic and Transportation Engineering(交通运输工程学报), 2005, 5(3): 57.
- [ 8 ] ZHANG Hua-ren, LI Wei-guo(张华仁, 李维国). Journal on Numerical Methods and Computer Application(数值计算与计算机应用), 2010, 30(3): 187.
- [ 9 ] XU Meng, GAO Zi-you. Science China, Technological Science, 2010, 53(12): 3270.
- [ 10 ] LI Yao-qing(李耀清). Data Process of Experiment(实验的数据处理). Hefei: Press of University of Science and Technology of China(合肥: 中国科学技术大学出版社), 2003. 144.
- [ 11 ] ZHANG Hong-yan, GENG Zheng(张鸿燕, 耿征). Computer Engineering and Applications(计算机工程与应用), 2009, 45(19): 5.

## Study on the Automatic Extraction Method of Spectral Data Features in Laser Induced Breakdown Spectroscopy

LIU Li-tuo, LIU Jian-guo, ZHAO Nan-jing\*, LU Cui-ping, CHEN Dong, SHI Huan, WANG Chun-long, ZHANG Yu-jun, LIU Wen-qing

Key Laboratory of Environment Optics&Technology, Institute of Anhui Optics Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China

**Abstract** According to that the spectral line broadening of laser induced breakdown spectroscopy (LIBS) was lorentz nonlinear function model, using Levenberg-Marquardt(L-M) algorithm to optimally estimate the undetermined parameter in the function, the continuum background radiation and peak position and peak intensity were corrected, and the method of data points automatic extraction for fitting was analysed. Automatic fitting and feature parameter extraction for spectral data of LIBS based on L-M algorithm that was hybrid programming with matrixVB and VB were accomplished, and the result was compared with the origin7. 5 data analysis software which was manual and commonly used and verified its stability and reliability.

**Keywords** Laser induced breakdown spectroscopy; Features extraction; Algorithm of Levenberg-Marquardt; Nonlinear fitting

(Received Jan. 15, 2011; accepted Jun. 5, 2011)

\* Corresponding author