

基于峰值拟合算法的光谱分析

朱 军^{1,2}, 刘文清², 刘建国², 徐 亮²(1. 安徽大学 电子科学与技术学院, 安徽 合肥 230039;
2. 中国科学院 安徽光学精密机械研究所, 安徽 合肥 230031)

摘要:傅里叶变换光谱学已经用于环境气体的定量分析中,为了提供高灵敏、连续实时的在线测量,分析中需要根据实测光谱选择光谱拟合算法。文中将峰值拟合算法应用于气体红外光谱分析中,以实际测量得到的气体透过率光谱为拟合对象,将其与数据库中的气体分子模型光谱进行拟合,在算法中通过迭代使残差的加权平方和最小,获得了气体浓度等重要光谱拟合结果。算法实现结果表明该方法不仅精确可靠而且具有简单快速的特点。

关键词:峰值拟合;数据采集;光谱分析

中图分类号: TN219

文献标识码: A

文章编号: 1005-3751(2006)04-0125-02

Spectra Analysis Based on Peak Fitting Algorithm

ZHU Jun^{1,2}, LIU Wen-qing², LIU Jian-guo², XU Liang²(1. Institute of Electronics Science & Technology, Anhui University, Hefei 230039, China;
2. Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China)

Abstract: Fourier transform infrared spectroscopy has been employed in environmental gas quantitative analysis. In order to provide high sensitivity and continuous real-time on-line measurements, some key techniques such as fitting algorithms should be chosen to analyze the measured spectra. Peak fitting algorithm is applied to gaseous infrared spectra analysis in the paper. The measured gas transmittance spectra are taken as fitted object. They are fitted with model spectra in molecular database. The weighted sum square of residuum is required to be the least one in the iteration of algorithm. The important results such as gas concentration are acquired by fitting. The implementation of the algorithm has shown that the method is not only exact and reliable but also simple and rapid.

Key words: peak fitting; data collection; spectra analysis

0 引言

傅里叶变换(FT)技术是信号处理和波谱解析的有力手段,利用傅里叶变换可以从数据中提取更多的有用信息,即以傅里叶级数拟合原光谱曲线,用较少项的级数就可获得与原光谱良好的近似,从而使所得谱线平滑,消除了部分噪声。因此傅里叶变换红外(FTIR)能有效地应用于各种样品的高精度快速分析^[1,2]。

文中通过光谱仪得到了高分辨率分子光谱,然后采用峰值拟合算法确定其光谱学参数以及所感兴趣的物质信息,拟合过程基于已知的实验条件对各谱线的位置、强度、宽度和其它一些参数拟合。在测量的光谱和用参数计算的光谱之间不断进行迭代直到两者的方差和最小为止,则结束迭代。考虑到光谱仪产生的光谱失真,在拟合中需要确定连续性和仪器线型等参数的影响。拟合光谱的方法

快速、可再生而且方便,采用峰值拟合方法提供了再现性的精度高于 1%。

1 干涉图及光谱的获取

在傅里叶变换红外(FTIR)光谱仪中(如图 1 所示),仪器内的迈克尔逊干涉仪动镜的匀速运动把待分析光变成干涉光,探测器上探测到干涉图,它是分析光的干涉强度随光程差变化的函数,即干涉强度随时间变化的函数。干涉图理论上是时间域的信号,但 FTIR 光谱仪测定干涉图时不可能对光程差进行连续采样,它只能进行时间域的间隔采样,采样光程间隔可以用氦氖激光余弦信号的半波长作为基准,激光是波长精确的单色光,用它作基准可以保证采样的波数精度达 0.01cm^{-1} 。He-Ne 激光干涉图是 FTIR 光谱仪的参比干涉图,如果在每个 He-Ne 激光干涉图的过零点处采样,相当于每隔 $0.316496\mu\text{m}$ 取一个数据点;测量 $400\sim 4000\text{cm}^{-1}$ 范围的中红外区光谱时,采样间隔可取隔一个过零点。

为了使采集的数据能正确代表真实的干涉图,除需正确选定采样间隔外,还需通过光学滤波器滤去频率过高或过低的辐射,使它们不能被探测器所检测,以免干扰所得

收稿日期:2005-07-19

基金项目:国家自然科学基金资助项目(10274080);安徽省高校优秀中青年骨干教师项目

作者简介:朱 军(1968-),女,安徽合肥人,副教授,博士研究生,研究方向为光学信号处理与光谱分析。

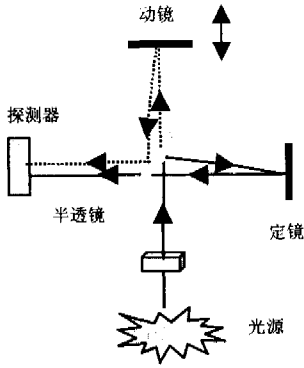


图 1 迈克尔逊干涉仪原理图

峰形。仪器内部的 DSP 电路通过多次数值积分的傅里叶变换,把采集到的干涉图数据变换成光谱图^[3]。红外光谱的测量常采用两种形式,即透过率谱和吸光度谱,两者的关系如式(1)所示:

$$A(\nu) = \ln[1/\tau(\nu)] \quad (1)$$

其中,透过率谱 $\tau(\nu)$ 可以直观地表示出样品对红外光的吸收情况;吸光度谱 $A(\nu)$ 中的吸光度值在一定范围内与样品的厚度以及样品的浓度成正比。测量透过率光谱时,首先分别求取样品和背景的单光束光谱,然后以样品的单光束光谱除以背景的单光束光谱即可得样品的透过率谱,这个求比谱的过程本质上是求取以背景为基准的差谱的过程。它可以消除测定时的各种干扰,包括部分仪器噪声,环境中的二氧化碳、水分等的干扰。

2 峰值拟合算法

文中测量了待测气体的透过率谱并对其进行峰值拟合。峰值拟合可以采用高斯、洛伦兹、Voigt 等峰值函数。给定函数 $y(x; \mathbf{a})$, 其中 \mathbf{a} 是 $N \times M$ 的参数矩阵, N 为观测值个数, M 为峰值拟合参数个数, y_i 为实验观测值, $y(x_i; \mathbf{a})$ 为峰值拟合的理论值。如果观测值 y_i 的标准误差为 σ_i , 则观测点数据的残差加权平方和称为拟合的 χ^2 量。

$$\chi^2 = \sum_{i=1}^N \left[\frac{y_i - y(x_i; \mathbf{a})}{\sigma_i} \right]^2 \quad (2)$$

将上述拟合的优度因子 χ^2 量作为目标函数, 寻求使 χ^2 量最小的参数值 \mathbf{a} 作为参数的估计值。令式(3)的一阶导数为零可以得到 χ^2 的最小值, 参数值 \mathbf{a} 可以通过式(4)的一阶导数方程组解出。

$$\frac{\partial \chi^2}{\partial a_k} = -2 \sum_{i=1}^N \left[\frac{y_i - y(x_i; \mathbf{a})}{\sigma_i^2} \right] \frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \equiv -2\beta_k \quad (3)$$

$k = 1, 2, \dots, M$

当二阶导数为零时, χ^2 的一阶导数得到最小值, χ^2 的二阶偏导如下:

$$\frac{\partial^2 \chi^2}{\partial a_k \partial a_l} = 2 \sum_{i=1}^N \frac{1}{\sigma_i^2} \left\{ \frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \frac{\partial y(x_i; \mathbf{a})}{\partial a_l} - [y_i - y(x_i; \mathbf{a})] \frac{\partial^2 y_i(x_i; \mathbf{a})}{\partial a_k \partial a_l} \right\} \equiv 2\alpha_{kl} \quad (4)$$

在实际的光谱数据处理过程中, 如果已经获得较为满意的结果, 这时可以不考虑式(4)的二阶导数项, 这样可

以简化算法处理。在算法中定义比例因子 λ 和矩阵 α' :

$$\alpha'_{jj} = \alpha_{jj}(1 + \lambda) \quad \alpha'_{jk} = \alpha_{jk} (j \neq k) \quad (5)$$

3 算法实现与光谱分析结果

通过图 2 所示的算法流程完成矩阵 α 的计算以达到减小 χ^2 的目的。如图 2 所示, 当 χ^2 的值小于给定值 ϵ 时结束计算, 即可以得到光谱峰值拟合的结果。

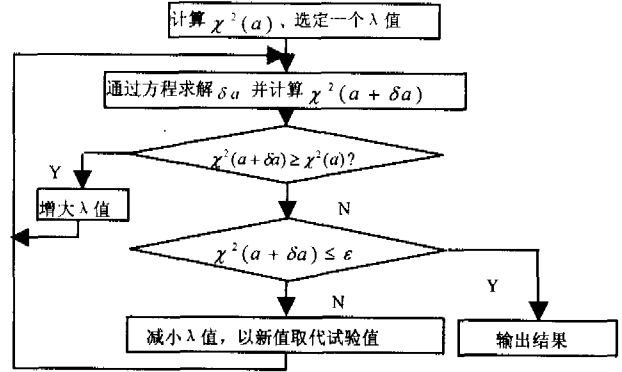


图 2 峰值拟合算法流程图

为了确定峰值采用 Savitsky - Golay 算法得到实测样品光谱曲线的二阶导数^[4,5], 算法采用最小二乘 k 次多项式线性回归拟合, 为了平滑数据至少需要 $k + 1$ 个光谱数据点。在各点对拟合的多项式求导, 计算采用矩阵形式进行。假定: d 表示求导的阶数, k 是多项式的次数, s 表示通过多项式拟合的点数, $m = (s - 1)/2$; P 表示 s 维数组, 数组值分别为 $(-m, \dots, 0, \dots, +m)$; Θ 表示 k 维数组; Y 表示实测光谱数据矩阵。在光谱拟合过程中, 实测光谱数据矩阵 Y 以及通过多项式拟合产生的中间矩阵 X 的关系:

$$Y = X\Theta$$

$$S = (Y - X\Theta)^T(Y - X\Theta) \quad (6)$$

将 S 对 Θ 求一阶导数并令其为零可以得到:

$$\Theta = (X^T X)^{-1} X^T Y = TY$$

$$T = (X^T X)^{-1} X^T \quad (7)$$

通过实验得到了 CO 气体的透过率光谱, 如图 3(a) 所示, 光谱的全波段范围选择在 $2000\text{cm}^{-1} \sim 2250\text{cm}^{-1}$ 。通过峰值拟合程序将气体透过率光谱与通过分子光谱数据库计算的光谱进行拟合, 全波段光谱的拟合结果如图 3(a) 所示, 局部波段的光谱拟合结果如图 3(b) 所示, 结果表明峰值拟合结果的残差非常小, 测量光谱与拟合光谱得到很好的吻合。

在全波段光谱拟合中的数据点为 519 个, 运行峰值拟合程序时, 用户需要设置一些初始参数: 如切趾码、分辨率、有效的切趾以及气压、温度和待测气体浓度的估计值等参数。最终经过 13 次迭代, 得到 CO 的浓度反演结果为 $4988.7 \times 10^{-6} (+/- 3.638E+001)$, 均方根误差为 0.7782。在反演另一未知浓度的 CO 气体时, 实测、拟合以及残留的透过率光谱迭代进行了 9 次, 得到 CO 气体的浓度为 $2540.6 \times 10^{-6} (+/- 9.938E+000)$, 均方根误差为 0.3601。

(下转第 129 页)

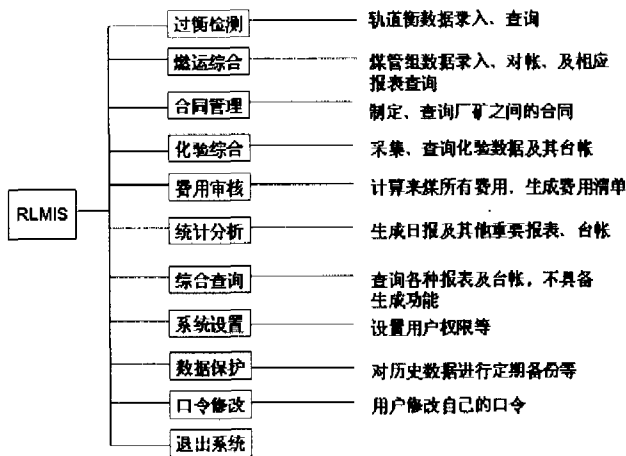


图2 系统结构图

3.4 系统测试维护

系统各模块组合成完整的系统后,是否能正常工作,是否达到系统设计的要求,要求进行调试工作。初步完成的原型系统存在着很大的漏洞和不完整性,但呈现在用户面前的已经是一个可以运行的系统,用户可以很方便地进行交流。经过用户一段时间的使用,对目前系统的反馈已经指出了设计上的缺陷和错误,例如,根据各自电厂的管理需要、工作习惯,对查询窗口检索条件的要求,打印格式的要求等提出自己的想法。开发人员就需要对系统进行进一步的分析和改进。软件工程中的螺旋模型就是对原型系统的不断改进,最终使原型系统最终解决目标系统,形成用户所需要的系统,这也是螺旋模型的精髓所在。

4 结束语

在火电厂燃料管理信息系统开发中,遵照软件体系结构方法,结合了软件工程中的螺旋模型,取得了比较好的效果。软件工程方法的应用,不仅使整个管理系统实现规范化,而且提高了软件的编制效率,在实现中体现了现代软件工程的思想。软件的编程实现了模块化、组件化,提高了软件代码的重用性。软件体系结构的规范方法大大提高了系统的工作效率,使软件的质量得到提高,后期的维护工作量也减少了,系统更加稳定。

参考文献:

- [1] 王珏,吴军.面向对象的软件工程在MIS开发中的应用[J].南京化工大学学报,2000,22(4):25-28.
- [2] 傅谦,张申生,曹键.面向软件工程的工作流管理系统[J].上海交通大学学报,2002,36(9):1351-1354.
- [3] 余金山.软件工程中的几个热点问题[J].华侨大学学报,2004,25(1):5-9.
- [4] 邸剑,李新叶,宋雨.面向对象软件工程方法分析[J].微机发展,2002,12(1):32-35.
- [5] 崔伟东,田捷.软件工程方法的演化[J].工业控制计算机,2002,15(6):34-37.
- [6] 马沪东,王明海.软件工程方法在飞行仿真系统开发中的应用[J].上海航天,2002(4):22-25.
- [7] 郝欣.软件工程在电厂DCS改造中的应用[J].东北电力技术,2003(3):42-44.

(上接第126页)

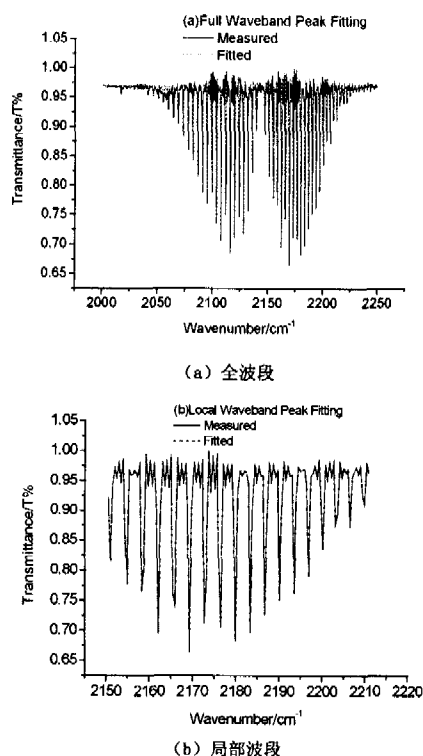


图3 峰值拟合后的光谱分析结果

4 结论

实验结果表明文中采用的算法能有效地获得气体的浓度信息,测量光谱与拟合光谱的残差非常小,结果准确。文中采用二阶导数法很好地确定了数据中峰的数目和位置,另外,还可以采用其它方法验证峰的位置和数目,如傅里叶自去卷积以及最大似然方法。峰值拟合算法的收敛速度较快,随着增加更多的峰,几乎总是能够得到一个更好的拟合。为了避免算法有时出现不收敛,要求实验人员为算法提供更多的信息,例如固定或限制某些已知或者能够准确估计的参数的范围。

参考文献:

- [1] 刘文清,崔志成,董凤忠.环境污染监测的光学和光谱学技术[J].光电子技术与信息,2002,15(5):1-12.
- [2] 朱军,刘文清,张保华.ASCM算法在FTIR监测系统中的应用[J].微机发展,2005,15(4):46-47.
- [3] 吴瑾光.近代傅里叶变换红外光谱技术及应用[M].北京:科学文献出版社,1994.
- [4] 张韵华,奚梅成,陈长松.数值计算方法和算法[M].北京:科学出版社,2000.
- [5] 王正明,易东云.测量数据建模与参数估计[M].长沙:国防科技大学出版社,1996.