第26卷,第10期　　光 谱 学 与 光 谱 分 析　　Vol. 26, No. 10, pp1850-1853
2006年10月　　Spectroscopy and Spectral Analysis　　October, 2006

基于非均匀B样条曲线的红外数据的精确拟合及校正

王先培[1], 张爱菊[2], 李少雄[1*]

1. 武汉大学系统集成与故障诊断实验室, 湖北 武汉　430072
2. 中南财经政法大学信息学院, 湖北 武汉　430064

摘　要　红外谱图数据的精确拟合是建立标准谱图的重要组成部分, 是光谱仪器正确进行定性和定量分析的基础。文章采用非均匀B样条对红外谱图数据进行曲线拟合, 通过采用基于最大范数距离的B样条拟合算法, 不仅解决了大容量红外谱图的精确拟合问题, 同时对谱图峰值精确定位也提供了方便。文章介绍了非均匀B样条曲线, 以及二阶(三阶)非均匀B样条曲线拟合的实现, 重点描述了红外谱图数据拟合精度的控制方法。

主题词　红外光谱分析仪; B样条; 红外谱图
中图分类号: TP89　　文献标识码: A　　文章编号: 1000-0593(2006)10-1850-04

引　言

对红外光谱分析仪器来说, 红外谱图中曲线的峰的位置、峰高和峰面积是定性和定量分析的重要依据。根据得到的离散数据进行拟合红外谱图曲线, 是研制专用红外光谱分析仪器的一个重要环节, 因为它直接影响着曲线峰的位置、峰高和峰面积的精确计算, 也直接影响着红外光谱仪器的准确性和精度。

一般情况下, 红外光谱曲线是多峰值的, 且差商变化很大, 因此, 对离散采样数据拟合的要求较高。目前, 根据离散数据绘制谱图主要有插值法、曲线拟合法[1]以及无理模型法[2]。插值法要求函数曲线经过插值点。主要有拉格朗日插值法、牛顿多项式插值法、埃米特插值法和样条插值法等。插值对多值曲线的拟合存在着许多不足, 如存在龙格现象、插值精度不够、无局部性、不能处理大斜率问题、不能处理数据点分配不均的曲线等[1,3]。无理模型方法是新近发展起来的基于神经网络和遗传算法等无理模型的曲线拟合法, 主要是用几何方法或神经网络的拓扑结构来确定红外数据间的关系。但在实际使用时, 由于得不到统一的曲线表达式, 无理模型方法将使红外谱图的定量分析变得更加复杂。曲线拟合法是用规则的曲线来近似表述红外数据的特性, 不要求函数严格的经过每个插值点, 但要求函数在插值点处与插值点的拟合误差矢量按某种范数达到最小值。曲线拟合法能部分解决测量的误差。主要方法有多项式最小二乘法、指数最小二乘法、三角函数拟合法和样条函数拟合法等。

综合各种因素, 在开发研制一个电力专用红外光谱分析仪器时, 采用了曲线拟合法中的非均匀B样条曲线来拟合红外谱图数据, 这种方法有很多优点, 如: 局部支柱性、凸组合性质、拟合曲线的线性变换条件下的几何不变性、节点处良好的连续性等[1,3]。实验证明, 该方法比较好地实现了红外谱图数据的精确拟合。采用控制顶点的直线类推插补的方法解决了峰值隐藏现象, 为峰的位置、峰高和峰面积的准确计算打下基础。本文首先介绍非均匀B样条曲线, 以及二阶(三阶)非均匀B样条曲线拟合的实现, 然后重点描述红外谱图数据拟合精度的控制方法。

1　非均匀B样条曲线

B样条(B-spline)广泛应用于当前的CAD/CAM系统中, 已经成为几何造型的核心方法。根据deBoor和Cox提出的B样条的递推定义[1], 若给定一组数据$\{V_i(x_i, y_i)\}$, 其中V_i为多边形的顶点, x_i, y_i为顶点坐标, $i = 1, 2, \cdots, n$, 则B样条表示为:

$$B_{i,0}(u) = \begin{cases} 1, & t_i \leqslant u \leqslant t_{i+1} \\ 0, & \text{其他} \end{cases} \tag{1}$$

$$B_{i,k}(u) = \frac{u - t_i}{t_{i+k} - t_i} B_{i,k-1}(u) + \frac{t_{i+k+1} - u}{t_{i+k+1} - t_{i+1}} B_{i+1,k-1}(u), \quad k > 0 \tag{2}$$

收稿日期: 2005-07-13, 修订日期: 2005-11-28
基金项目: 国家科技攻关项目(2004BA210A02)和湖北省科技攻关项目(2003AA101E05)资助
作者简介: 王先培, 1963年生, 武汉大学电子信息学院教授

这里，约定 0/0= 0。

式中 k 表示 B 样条的幂次，t_i 为由顶点得到的节点矢量，u 为引入的节点矢量自变量，下标 i 为 B 样条的序号。上式表明任意 k 次 B 样条可由 2 个相邻的 $k-1$ 次 B 样条的线性组合构成。根据该定义可以推导任意阶次的 B 样条曲线，第 i 段 k 次样条曲线的表达式为:

$$P_i(u) = [B_{i-k,k} \quad B_{i-k+1} \quad \cdots \quad B_{i,k}]\begin{bmatrix} V_i \\ V_{i+1} \\ \vdots \\ V_{i+k}\end{bmatrix}, \quad u \in [t_{i+1}, t_i] \tag{3}$$

其中，$B_{j,k}$ 为第 j 条 k 次 B 样条的有关段，作为构造 B 样条曲线的基函数，j 为 B 样条的序号，$j = i-k,\ i-k+1,\ \cdots,\ i$；$V_l$ 为第 l 个顶点，l 为顶点序号，$l = i,\ i+1,\ \cdots,\ i+k$。由 (3) 式说明第 i 段 k 次样条曲线由 $k+1$ 段 $B_{j,k}$ 和 $k+1$ 个顶点 V_l 组成。

2　二次(三阶)非均匀 B 样条曲线拟合的实现

采样和数据处理后的红外数据是一串离散的非重合数据 $\{V_i(x_i, y_i)\}$，$i = 0, 1, \cdots, n$，其中 V_i 为多边形的顶点(即为给定的数据点)，x_i，y_i 为顶点坐标。且设红外数据拟合曲线在连接处达到三阶连续。

2.1　节点矢量的计算

若给定一组数据 $\{V_i(x_i, y_i)\}$，$i = 0, 1, \cdots n$。在重端点的条件下，二次(三阶)非均匀 B 样条首、末端通过多边形的首、末顶点 V_0 和 V_n，中间各结点分别处于各对应边的中部位置。设节点矢量为(t_0，t_1，t_2，t_3，…，t_i，…，t_{n+2}，t_{n+3}，t_{n+4}，t_{n+5})，其中 $t_0 = t_1 = t_2 = 0$，$t_{n+3} = t_{n+4} = t_{n+5} = L$ 为扩充的矢量，节点矢量 t_i 的计算如下:

$$t_i = \sum_{j=0}^{i-3} l_j + l_{i-2}/2, \quad i = 3, 4, \cdots, n+2 \tag{4}$$

这里，L 为顶点边长的总和 $L = \sum_{i=0}^{n-1} l_i$，l_i 为计算顶点组成的边长:

$$l_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}, \quad i = 0, 1, \cdots, n-1 \tag{5}$$

2.2　任意一段二次(三阶)非均匀 B 样条的计算

对任意的相邻四节点(t_{i-1}，t_i，t_{i+1}，t_{i+2}，t_{i+3})，可以确定一条二次(三阶) B 样条(由 3 段组成)的计算公式，其表达式为:

$$B_{(0,2)}(u) = \frac{(u - t_i)^2}{(t_{i+1} - t_i)(t_{i+2} - t_i)}, \quad u \in [t_i, t_{i+1}] \tag{6}$$

$$B_{(1,2)}(u) = \frac{(u - t_i)^2}{(t_{i+1} - t_i)(t_{i+2} - t_i)} - \frac{(t_{i+3} - t_i)(u - t_{i+1})^2}{(t_{i+1} - t_i)(t_{i+1} - t_{i+2})(t_{i+1} - t_{i+3})}, \quad u \in [t_{i+1}, t_{i+2}] \tag{7}$$

$$B_{(2,2)}(u) = \frac{(t_{i+3} - u)^2}{(t_{i+3} - t_{i+1})(t_{i+3} - t_{i+2})}, \quad u \in [t_{i+2}, t_{i+3}] \tag{8}$$

公式(6)，(7)，(8)确定了一条完整的 B 样条曲线。这个 B 样条曲线由 4 个节点(t_{i-1}，t_i，t_{i+1}，t_{i+2}，t_{i+3})组成，u 为节点矢量参数。

由公式(3)，二次(三阶)非均匀 B 样条的第 i 段参数方程表示为:

$$P_i(t) = [B_{i-2,2} \quad B_{i-1,2} \quad B_{i,2}]\begin{bmatrix} V_i \\ V_{i+1} \\ V_{i+2}\end{bmatrix}, \quad u \in [t_i, t_{i+1}] \tag{9}$$

这里，$B_{i-2,2}$ 为第 $i-2$ 条 B 样条的 $B_{2,2}$ 段，$B_{i-1,2}$ 为第 $i-1$ 条 B 样条的 $B_{1,2}$ 段，$B_{i,2}$ 为第 i 条 B 样条的 $B_{0,2}$ 段。引入参数 t，令 $t = u - t_i$，将其代入(6)，(7)，(8)式，则有:

$$B_{(i-2,2)}(t) = \frac{(t_{i+1} - t_i - t)^2}{(t_{i+1} - t_{i-1})(t_{i+1} - t_i)}, \quad t \in [0, t_{i+1} - t_i] \tag{10}$$

$$B_{(i-1,2)}(t) = \frac{(t - t_i - t_{i-1})^2}{(t_i - t_{i-1})(t_{i+1} - t_{i-1})} - \frac{(t_{i+2} - t_{i-1})(t + t_i - t_i)^2}{(t_i - t_{i-1})(t_i - t_{i+1})(t_i - t_{i+2})}, \quad t \in [0, t_{i+1} - t_i] \tag{11}$$

$$B_{(i,2)}(t) = \frac{t^2}{(t_{i+1} - t_i)(t_{i+2} - t_i)}, \quad t \in [0, t_{i+1} - t_i] \tag{12}$$

将 V_i 顶点坐标和节点矢量(t_{i-1}，t_i，t_{i+1}，t_{i+2}，t_{i+3})分别代入到(9)，(10)，(11)，(12)式中，就可得到 B 样条的第 i 段曲线参数方程。

以 t 为变量，将坐标 x_i，y_i 分别代入方程，得到 B 样条曲线上各点坐标参数 $x_i(t)$，$y_i(t)$，即为第 i 段 B 样条曲线上点矢量，这条曲线实际上是从 $(V_i + V_{i+1}/2)$ 到 $(V_{i+1} + V_{i+2})/2$ 的一段 B 样条曲线。

将各个 B 样条连接起来，就得到二次(三阶)非均匀多项式 B 样条，表达式为:

$$P(t) = [B_{0,2} \quad \cdots \quad B_{i-2,2} \quad B_{i-1,2} \quad B_{i,2} \quad \cdots \quad B_{n,2}]\begin{bmatrix} V_0 \\ \vdots \\ V_i \\ V_{i+1} \\ V_{i+2} \\ \vdots \\ V_n\end{bmatrix}, \quad t \in [0, t_{i+1} - t_i] \tag{13}$$

3　二次(三阶)非均匀 B 样条曲线拟合精度的控制

由于采样频率的限制，根据采样数据，利用上面介绍的拟合算法得到的红外谱图曲线一般存在微小的误差。例如，若采样频率较高，用 B 样条拟合的谱图曲线在整体上看去和谱图数据直接直线连接的很接近，但局部放大后发现，在某些有峰值或谷值的波段，曲线的拟合效果并不是很好，取某一个谱图局部放大如图 1 和图 2 所示。

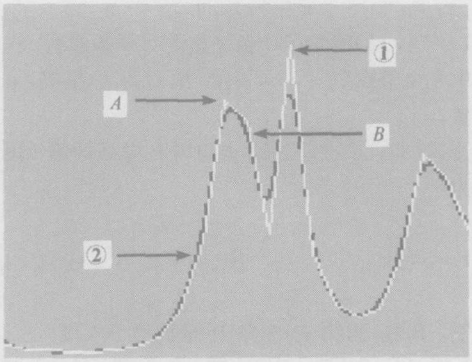

Fig. 1 Part of spectrum

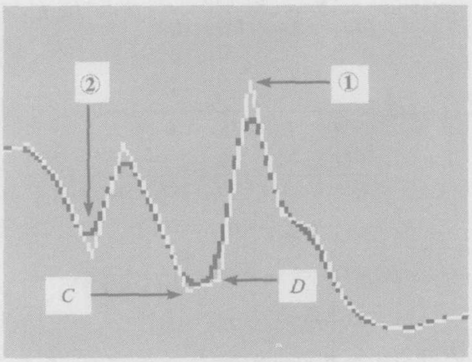

Fig. 2 Part of spectrum

图 1 和图 2 中，A，B，C，D 分别为红外数据点，曲线①为数据点直线连接而成的曲线，由白色线所示，曲线②为采用二次(三阶)非均匀多项式 B 样条拟合而成的曲线，由黑色线所示。拟合后的谱图局部放大显示：二次(三阶)非均匀多项式 B 样条的拟合方法，除谱图上的峰值和谷值处以及个别差商变化大的地方外，都能达到很好的拟合效果。另外，在数据点 A 和 B、C 和 D 之间本应有峰的地方，因为红外仪器的采样分辨率问题，出现真实峰值隐藏的现象，这种现象会随着红外仪器分辨率降低越加严重，这成为峰值定位难的重要原因，对红外仪器定性和定量分析产生严重的影响，因此必须对生成的 B 样条进行校正处理。

控制顶点的数目多少决定 B 样条曲线对数据点的整体逼近程度[1,4]，根据顶点使用的情况，校正的方法也不尽相同，经典的先用 Q 权因子来调整拟合顶点，再采用最小二乘来拟合曲线，以此来不断的提高拟合精度[1,3,5]，文献[6]则采用移动控制点的概念来度量曲线的逼近程度。为了得到高精度，本文摒除经典的基于 B 样条的最小二乘曲线逼近算法[4,5,7]，采用最大范数距离算法来实现[4,5]。

设 V_i 为顶点向量，R_i 为由 B 样条 $P(\boldsymbol{u})$ 决定的拟合曲线上的点向量，拟合后顶点与经过该点拟合曲线段的最大范数距离为：

$$d_{\max} = \max_{0 \leqslant i \leqslant n} \| V_i - R_i \| \tag{14}$$

根据文献资料采用二次曲线拟合[8]的方法比直线插补法能更加精确的解决 B 样条曲线在峰值处的拟合精度，但这种算法计算量大，这里我们采用直线类推插补法，目的是找出在峰值隐藏处峰值点的具体坐标 $V(x, y)$。本算法是为专用红外仪器设计的，其计算量小，方便快捷，精度可以满足要求。在峰值隐藏 V_i 和 V_{i+1} 处前后各取 5 个点 $V_i(x_i, y_i)$，计算点之间的差商 $p_j = \frac{y_{j+1} - y_j}{x_{j+1} - x_j}$，$j = 1, \cdots, 12$ 由差商分别类推 V 和 V_i，V 和 V_{i+1} 之间的直线斜率 K_i 和 K_{i+1}，根据类推的两个斜率和顶点 V_i，V_{i+1} 建立 2 个直线方程，由此求出他们交点即为类推的峰值点。插入峰值点后，曲线逼近的方法如下：设给定误差界为 E，则首先按得到的数据进行拟合，计算每个节点区间的最大范数距离，若最大范数距离大于 E，则在该节点区间新增加一个顶点，然后，重新拟合，一直到满足误差要求为止。用 JAVA 编程重新实现上谱图的拟合，该谱图经放大后图 1 和图 2 部分的拟合情况如下图 3 和图 4 所示。

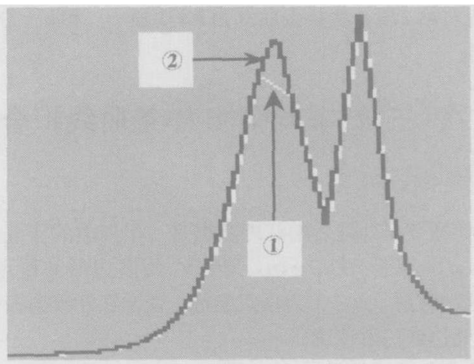

Fig. 3 After processing

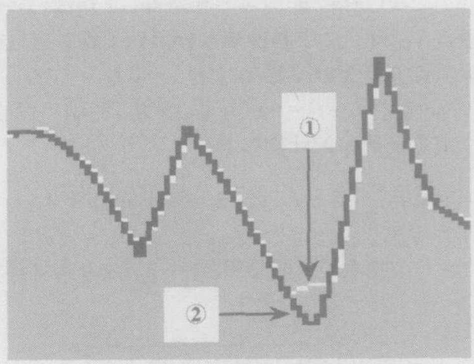

Fig. 4 After processing

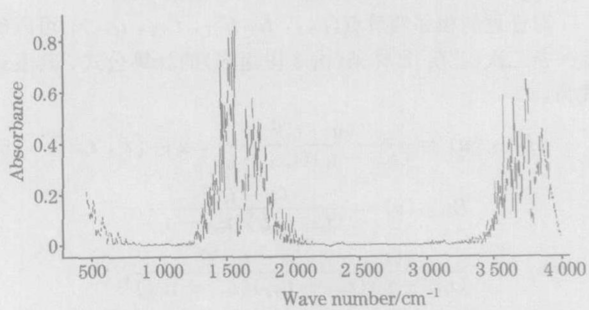

Fig. 5 The water spectrum fitted by B-spline

图 3 和图 4 中, 曲线①为原谱图数据点直线连接而成的曲线, 曲线②为采用调整后的 B 样条拟合曲线。图 3 和图 1, 图 4 和图 2 分别比较可以看到, 峰值隐藏的现象得到很好的解决, 而且在峰值处拟合达到了一定的精度。水的红外数据 B 样条拟合后的谱图如图 5 所示。

4 结 论

非均匀多项式 B 样条的谱图拟合曲线具有很多优点, 如: 多项式再生性, 单调性和凸性等等; 另外, 借助于调整顶点可以灵活的修改曲线的形状; 该算法稳定, 拟合的曲线通过精度的控制能很好的反应谱图特征。通过非均匀多项式 B 样条的拟合方法, 可以得到 1 个光滑的具有统一表达式的原始谱图, 这便于提高后续的谱图解析方法的准确性和精度, 这些解析方法如: 基于高斯函数的、双高斯函数的、指数衰减修正高斯(EMG)函数的、洛伦兹(Lorentz)函数的或高斯函数和洛伦兹函数的组合函数的等等, 它们依赖原始谱图来确定峰的位置和数量, 因此, 基于非均匀多项式 B 样条的谱图数据的精确拟合有利于提高红外分析的准确性和精度。另外采用基于最大范数距离控制精度的方法, 避免了最小二乘算法中解高维方程组的计算量, 方便谱图的快速形成与显示, 在便携式的红外分析仪器中具有重要的使用价值。

参 考 文 献

[1] SHI Fa-zhong(施法中). CAGD and NURBS(计算机辅助几何设计与非均匀有理 B 样条). Beijing: Higher Education Press(北京: 高等教育出版社), 2001.

[2] QIAO Li-shan, WANG Yu-lan, ZENG Jin-guang(乔立山, 王玉兰, 曾锦光). Journal of Chengdu University of Technology (成都理工大学学报), 2004, 31(2): 91.

[3] ZHU Xin-xiong (朱心雄). Free Curves and Surfaces Modeling Technology(自由曲线曲面造型技术). Beijing: Science Press(北京: 科学出版社), 2001.

[4] Piegl L, Tiller W. The NURBS Book. Berlin Heidelberg: Springer-Verlag, First Edition: 1995, Second Edition: 1997.

[5] Camara Amara, Han Xuli. Mathematical Theory and Applications, 2002, 22(1): 40.

[6] FENG Guo-xin, ZHANG Guo-xiong, XIE Ze-xiao, et al(冯国馨, 张国雄, 解则晓, 等). Journal of Tianjin University (天津大学学报), 2001, 34(3): 285.

[7] Robinson M P, Clegg J. IEEE Transactions on Electromagnetic Compatibility, 2005, 47(2): 399.

[8] Cheng F, Wang X, Barsky B A. Computers and Mathematics with Application, 2001, 41: 39.

Accurate Curve Fitting and Revising of Infrared Data Based on NUBSC

WANG Xian-pei[1], ZHANG Ai-ju[2], LI Shao-xiong[1*]

1. Laboratory of System Integration and Faults Diagnostics Science of Wuhan University, Wuhan 430072, China

2. Info-School of Zhongnan University of Economics and Law, Wuhan 430064, China

Abstract Accurate curve fitting of infrared spectrum data is important for building standard spectrum, and is the basic requirement for qualitative and quantitative spectral analysis. In the present paper, B-Spline was used to accurately fit the infrared spectrum data. Based on the max norm distances of B-Spline curve fitting arithmetic, this method not only resolved the curve fitting problem of infrared data, but also provided a convenient way to orient peak values of spectrum. The authors introduced the B-Spline, realized curve fitting of quadratic non-uniform B-Spline curves (for short: NUBSC), and described the precision control of curve fitting of infrared spectrum data.

Keywords Infrared spectrometers; B-Spline; Infrared spectrum

(Received Jul. 13, 2005; accepted Nov. 28, 2005)

* Corresponding author