

(12) 发明专利申请

(10) 申请公布号 CN 102360502 A

(43) 申请公布日 2012. 02. 22

(21) 申请号 201110263690. 6

(22) 申请日 2011. 09. 07

(71) 申请人 中国科学院武汉物理与数学研究所  
地址 430071 湖北省武汉市武昌小洪山

(72) 发明人 刘朝阳 鲍庆嘉 陈方 冯继文  
叶朝辉

(74) 专利代理机构 武汉宇晨专利事务所 42001  
代理人 王敏锋

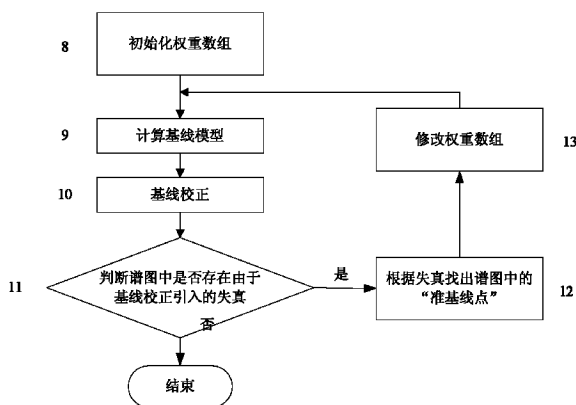
(51) Int. Cl.  
G06T 7/00(2006. 01)  
G01S 7/48(2006. 01)

权利要求书 2 页 说明书 6 页 附图 4 页

(54) 发明名称  
一种自动基线校正方法

(57) 摘要

本发明公开了一种自动基线校正方法,包括以下步骤:将原始数据经过变换得到谱图数据;利用连续小波变换计算谱图的数值导数;分别利用滑动窗口法和迭代阈值法识别谱图中的窄峰信号和宽峰信号;对宽峰信号进行轮廓拟合识别出宽峰信号的边缘;计算基线模型;根据基线模型对谱图数据进行校正;查找校正后的谱图数据中的畸变的信号,将畸变最严重的点作为准基线点并重新计算基线模型,直至没有畸变信号,得到最终的基线模型和最终的校正后的谱图数据。本发明与现有技术相比可以消除由于基线失真而造成的噪声计算不准确;准确的识别出谱图中的窄峰信号和宽峰信号;校正基线失真较大的谱图;避免在复杂谱图中由于基线校正引入的谱峰失真。



1. 一种自动基线校正方法,其特征在于,包括以下步骤:

步骤 1、利用傅里叶变换和相位校正处理采集到的原始数据,得到谱图数据;

步骤 2、利用连续小波变换对谱图数据进行计算,得到谱图的数值导数;

步骤 3、利用滑动窗口法和谱图的数值导数识别谱图中的窄峰信号,得到窄峰信号区间;

步骤 4、利用迭代阈值法识别谱图数据中的宽峰信号,得到部分宽峰信号区间;

步骤 5、根据部分宽峰信号区间,利用轮廓拟合方法拟合出宽峰信号的轮廓;

步骤 6、将谱图数据中大于步骤 4 中宽峰信号的轮廓最大值 3% 的点都作为宽峰信号点,从而得到完整宽峰信号区间;

步骤 7、将谱图数据减去窄峰信号区间和完整宽峰信号区间得到谱图的基线区间;

步骤 8、利用步骤 7 的得到的基线区间初始化权重数组,若谱图数据点位于信号区间,则权重数组值初始化为 0,若谱图数据点位于基线区间,则权重数组值初始化为 1;

步骤 9、利用权重数组和滤波算法计算基线模型;

步骤 10、利用步骤 9 中的基线模型对步骤 1 所述的谱图数据进行校正,得到校正后的谱图数据;

步骤 11、逐个判断校正后的谱图数据中的各个谱峰,若正峰信号中出现负数部分则认为该谱峰产生了畸变,若负峰信号中出现了正数部分,则认为该谱峰产生畸变,如果谱峰产生畸变则执行步骤 12,如果谱峰没有产生畸变则退出计算得到最终基线模型和校正后的谱图数据;

步骤 12、将步骤 11 中谱峰产生畸变最严重的点设置为准基线点;

步骤 13、将步骤 12 中的准基线点对应的权重数组设置为 1 并返回步骤 9 重新计算基线模型。

2. 根据权利要求 1 所述的一种自动基线校正方法,其特征在于,所述的步骤 3 中利用滑动窗口法包括以下步骤:

步骤 3.1、根据步骤 2 中得到的谱图数据导数计算噪声水平  $\sigma_{\text{noise}}$ ;

步骤 3.2、设定阈值为  $n \times \sigma_{\text{noise}}$ ,其中  $n$  为参数值;

步骤 3.3、比较滑动窗口的高度与步骤 3.2 中确定的阈值的大小,如果滑动窗口的高度大于阈值,则滑动窗口的中心点位于窄峰信号区间;如果滑动窗口的高度小于等于阈值,则滑动窗口的中心点位于基线区间。

3. 根据权利要求 1 所述的一种自动基线校正方法,其特征在于,所述的步骤 4 中迭代阈值法包括以下步骤:

步骤 4.1、利用谱图数据计算出迭代阈值,迭代阈值基于公式  $\text{MEAN}+3*\text{SDEV}$ , MEAN 是谱图数据的平均值, SDEV 是谱图数据的标准差;

步骤 4.2、将谱图数据与迭代阈值进行比较,若谱图数据存在比迭代阈值大的数据点,则进行步骤 4.3;若谱图数据中所有的点都小于迭代阈值,则停止迭代,得到所有的宽峰信号,从而得到部分宽峰信号区间;

步骤 4.3、将步骤 4.2 中得到的比迭代阈值大的谱图数据作为宽峰信号点,比迭代阈值小的谱图数据作为步骤 4.1 中的谱图数据并返回步骤 4.1。

4. 根据权利要求 2 所述的一种自动基线校正方法,其特征在于,所述的步骤 3.2 中参数

值  $n$  为 3, 所述的步骤 3.3 中滑动窗口的长度为整个谱图数据的宽度的 0.2%。

5. 根据权利要求 1 所述的一种自动基线校正方法, 其特征在于, 所述的步骤 5 中的轮廓拟合是基于最小化惩罚函数  $P$  :

$$P = \sum_{\Omega} P(\Omega)$$

其中 : 当  $E(\Omega) - S(\Omega) \geq 0$ ,  $P(\Omega) = (E(\Omega) - S(\Omega))^2$ ; 当  $E(\Omega) - S(\Omega) \leq 0$ ,  $P(\Omega) = (f \times (E(\Omega) - S(\Omega)))^2$ ,  $E(\Omega)$  是需要拟合的宽峰信号,  $S(\Omega)$  是谱峰信号,  $f$  是轮廓调整参数,  $\Omega$  为谱图频率坐标。

6. 根据权利要求 1 所述的一种自动基线校正方法, 其特征在于, 所述的步骤 9 中的滤波算法基于最小化自定义目标函数 :

$$F(m) = \sum_{i=1}^N w(S(m-1)_i) \times (S(m-1)_i - \tau(m)_i)^2 + \lambda \sum_{i=2}^{N-1} \{[\tau(m)_{i+1} - \tau(m)_i] - [\tau(m)_i - \tau(m)_{i-1}]\}^2$$

式中 :  $m$  是指第  $m$  次迭代,  $i = 1, 2, \dots, N$ ,  $N$  表示数据长度,  $\tau(m)$  表示第  $m$  次迭代所需计算的基线模型数据,  $S(m-1)_i$  是第  $m-1$  次迭代后基线校正的谱图数据,  $w(S(m-1)_i)$  是基于前一次校正后谱图计算得到的谱图权重数组, 当  $S(m-1)_i$  是信号点时,  $w(S(m-1)_i) = 0$ ; 当  $S(m-1)_i$  是噪声点时,  $w(S(m-1)_i) = 1$ 。

## 一种自动基线校正方法

### 技术领域

[0001] 本发明涉及谱图分析领域,更具体涉及一种自动基线校正方法。适用于核磁共振谱,色谱,拉曼光谱等各种谱图的分析 and 后期数据处理。

### 背景技术

[0002] 基线失真会对谱图解析以及谱图定量分析造成很大影响,消除基线失真是谱图后期数据处理的一个必要步骤。目前应用广泛的基线校正算法都为频域基线校正方法。其原理是在谱图数据的基础上构建出一条平滑基线模型,然后由原始的谱图数据减去这个基线模型从而得到没有失真的谱图数据。

[0003] 一般的频域基线校正方法是在识别出来的基线点上构建基线模型。相比于没有基线识别的基线校正算法,有基线识别的算法能够处理谱图中基线失真较大和既存在正峰又存在负峰的情况,所以其更适于各种分析谱图。1991年Dietrich提出了在谱图数值导数的基础上利用重复阈值迭代的算法来识别基线。Cobas利用连续小波变换代替数值导数计算从而改进了这种算法。2000年Golotvin提出了另一种基线识别方法,我们在此称之为滑动窗口法。上述的基线识别方法都存在各自不同的优缺点,如在谱图数值导数的基础上进行基线识别,会降低信噪比以及会使谱图中的宽峰信号消失;采用连续小波变换改进后可以解决信噪比降低的问题,但需要人为的调整参数才能识别宽峰信号;此外在实际应用的过程中我们发现利用迭代阈值法对谱峰边缘的识别不够准确,而滑动窗口法对谱图中的宽峰信号的识别不够准确甚至不能识别出来。

[0004] 在识别出来的基线的基础上构建合适的基线模型是基线校正中的第二步,也是最为关键的步骤。目前存在的各种模型构建方法可以分为两类:第一类是利用特定的函数来拟合识别出来的基线点,如多项式拟合,三次样条插值等;第二类是在识别出来的基线上利用平滑算法如Whittaker滤波等构建基线模型。利用特定的函数来拟合基线模型并不能适应于很多基线失真较大的谱图,因为基线失真可能是任意形状并不满足特定的函数形状。Whittaker滤波是一种比较有效的平滑算法,但是其在处理复杂数据的时候可能会使谱峰信号发生畸变。

### 发明内容

[0005] 本发明的目的是针对现有的技术存在的上述问题,提供了一种自动基线校正方法,该方法数据处理速度快,可以消除由于基线失真而造成的噪声计算不准确,且避免了基线校正引入的谱峰畸变。

[0006] 为了实现上述目的,本发明采用以下技术方案:

[0007] 一种自动基线校正方法,包括以下步骤:

[0008] 步骤1、利用傅里叶变换和相位校正处理采集到的原始数据,得到谱图数据;

[0009] 步骤2、利用连续小波变换对谱图数据进行计算,得到谱图的数值导数;

[0010] 步骤3、利用滑动窗口法和谱图的数值导数识别谱图中的窄峰信号,得到窄峰信号

区间；

[0011] 步骤 4、利用迭代阈值法识别谱图数据中的宽峰信号，得到部分宽峰信号区间；

[0012] 步骤 5、根据部分宽峰信号区间，利用轮廓拟合方法拟合出宽峰信号的轮廓；

[0013] 步骤 6、将谱图数据中大于步骤 4 中宽峰信号的轮廓最大值 3% 的点都作为宽峰信号点，从而得到完整宽峰信号区间；

[0014] 步骤 7、将谱图数据减去窄峰信号区间和完整宽峰信号区间得到谱图的基线区间；

[0015] 步骤 8、利用步骤 7 的得到的基线区间初始化权重数组  $w(S(m-1)_i)$ ，若谱图数据点位于信号区间，则权重数组值初始化为 0，若谱图数据点位于基线区间，则权重数组值初始化为 1；

[0016] 步骤 9、利用权重数组和滤波算法计算基线模型；

[0017] 步骤 10、利用步骤 9 中的基线模型对步骤 1 所述的谱图数据进行校正，得到校正后的谱图数据；

[0018] 步骤 11、逐个判断校正后的谱图数据中的各个谱峰，若正峰信号中出现负数部分则认为该谱峰产生了畸变，若负峰信号中出现了正数部分，则认为该谱峰产生畸变，如果谱峰产生畸变则执行步骤 12，如果谱峰没有产生畸变则退出计算得到最终基线模型和校正后的谱图数据；

[0019] 步骤 12、将步骤 11 中谱峰产生畸变最严重的点设置为准基线点；

[0020] 步骤 13、将步骤 12 中的准基线点对应的权重数组设置为 1 并返回步骤 9 重新计算基线模型。

[0021] 如上所述的步骤 3 中利用滑动窗口法包括以下步骤：

[0022] 步骤 3.1、根据步骤 2 中得到的谱图数据导数计算噪声水平  $\sigma_{\text{noise}}$ ；

[0023] 步骤 3.2、设定阈值为  $n \times \sigma_{\text{noise}}$ ，其中  $n$  为参数值；

[0024] 步骤 3.3、比较滑动窗口的高度与步骤 3.2 中确定的阈值的大小，如果滑动窗口的高度大于阈值，则滑动窗口的中心点位于窄峰信号区间；如果滑动窗口的高度小于等于阈值，则滑动窗口的中心点位于基线区间。

[0025] 如上所述的步骤 4 中迭代阈值法包括以下步骤：

[0026] 步骤 4.1、利用谱图数据计算出迭代阈值，迭代阈值基于公式  $MEAN+3*SDEV$ ， $MEAN$  是谱图数据的平均值， $SDEV$  是谱图数据的标准差；

[0027] 步骤 4.2、将谱图数据与迭代阈值进行比较，若谱图数据存在比迭代阈值大的数据点，则进行步骤 4.3；若谱图数据中所有的点都小于迭代阈值，则停止迭代，得到所有的宽峰信号，从而得到部分宽峰信号区间；

[0028] 步骤 4.3、将步骤 4.2 中得到的比迭代阈值大的谱图数据作为宽峰信号点，比迭代阈值小的谱图数据作为步骤 4.1 中的谱图数据并返回步骤 4.1。

[0029] 如上所述的步骤 3.2 中参数值  $n$  为 3，所述的步骤 3.3 中滑动窗口的长度为整个谱图数据的宽度的 0.2%。

[0030] 如上所述的步骤 5 中的轮廓拟合是基于最小化惩罚函数  $P$ ：

$$[0031] \quad P = \sum_{\Omega} P(\Omega)$$

[0032] 其中:当  $E(\Omega) - S(\Omega) \geq 0$ ,  $P(\Omega) = (E(\Omega) - S(\Omega))^2$ ; 当  $E(\Omega) - S(\Omega) \leq 0$ ,  $P(\Omega) = (f \times (E(\Omega) - S(\Omega)))^2$ ,  $E(\Omega)$  是需要拟合的宽峰信号,  $S(\Omega)$  是谱峰信号,  $f$  是轮廓调整参数,  $\Omega$  为谱图频率坐标。

[0033] 如上所述的步骤 9 中的滤波算法基于最小化自定义目标函数:

$$[0034] \quad F(m) = \sum_{i=1}^N w(S(m-1)_i) \times (S(m-1)_i - \tau(m)_i)^2 +$$

$$[0035] \quad \lambda \sum_{i=2}^{N-1} \{[\tau(m)_{i+1} - \tau(m)_i] - [\tau(m)_i - \tau(m)_{i-1}]\}^2$$

[0036] 式中:  $m$  是指第  $m$  次迭代,  $i = 1, 2, \dots, N$ ,  $N$  表示谱图数据长度,  $\tau(m)$  表示第  $m$  次迭代所需计算的基线模型数据,  $S(m-1)_i$  是第  $m-1$  次迭代后基线校正的谱图数据,  $w(S(m-1)_i)$  是基于前一次校正后谱图计算得到的谱图权重数组, 当  $S(m-1)_i$  是信号点时,  $w(S(m-1)_i) = 0$ ; 当  $S(m-1)_i$  是噪声点时,  $w(S(m-1)_i) = 1$ 。

[0037] 目标函数  $F(m)$  的第一项要求每次迭代计算出来的基线模型必须忠诚于上一次校正后谱图  $S(m-1)_i$  的基线点。第二项要求每次计算的基线模型数据必须足够平滑。

[0038] 如上所述的一种自动基线校正方法, 主要结合了连续小波变换, 迭代阈值法和滑动窗口法进行基线识别, 并对其中各自不足之处进行了合理的改进。利用连续小波变化计算谱图的数值导数, 目的是消除由于基线失真而造成的噪声计算不准确, 然后在谱图数据导数的基础上利用滑动窗口法进行基线识别。滑动窗口算法的基本原理是比较滑动窗口的高度和噪声水平, 如果滑动窗口的高度大于根据噪声水平设定的阈值, 则认为窗口的中心点为窄峰信号点, 否则为基线点。阈值设定为噪声水平乘以一个系数值 ( $n \sigma_{\text{noise}}$ ), 一般情况下滑动窗口长度取为整个谱宽的 0.2%, 阈值计算的系数值  $n$  取 3 可满足要求。

[0039] 滑动窗口方法可以准确的识别出谱图中的窄峰信号, 而针对宽峰信号则采用了迭代阈值法结合宽峰轮廓拟合进行识别。迭代阈值算法首先利用所有的数据计算出一个阈值 ( $\text{MEAN} + 3 * \text{SDEV}$ , 这里  $\text{MEAN}$  指的是数据的平均值,  $\text{SDEV}$  指的是数据的标准差), 然后利用比上一次迭代阈值小的数据点重新计算新的迭代阈值, 直到剩下的点都小于迭代阈值停止迭代。最后剩下比最后迭代阈值小的点为基线点。普通的迭代阈值算法对谱峰尾部的识别不够准确, 尤其是对宽峰信号的尾部。为此对其进行了一定的改进, 采用的方法是利用拟合出的谱峰轮廓来确定基线区间。首先利用迭代阈值识别出来的数据拟合出谱峰轮廓, 然后利用谱峰轮廓来确定基线区间。谱峰轮廓拟合的函数采用了 Tsallis 曲线, 这种曲线的形状由参数  $q$  决定 ( $q > 1$ ), 当  $q$  值趋近 1 时就是 Gaussian 曲线, 当  $q = 2$  时为罗伦茨曲线。轮廓拟合的方法采用了最小化惩罚函数  $P$  的方法。

[0040] 如上所述的步骤 9 中, 在构建基线模型的时候不仅利用了 1 ~ 5 步骤识别出来的基线点, 而且考虑在基线过程中可能引起的谱峰畸变, 通过不断的识别“准基线点”来避免这种畸变。

[0041] 与现有技术相比, 本发明的优点和有益效果是:

[0042] 1、可以消除由于基线失真而造成的噪声计算不准确, 噪声计算的不准确将直接影响基线识别的结果, 从而造成错误的基线校正结果。

[0043] 2、可以准确的识别出谱图中的窄峰信号和宽峰信号。尤其是对宽峰的识别, 普通的算法只能识别出部分宽峰信号或者不能识别出宽峰信号, 从而造成基线校正后的谱图宽

峰信号变小或者消失。如图 3 所示中利用现有的两种基线识别方法校正后,谱图中的宽峰信号面积缩小了 64% 或者直接被消除,而本发明利用的基线识别方法保持了原有的峰面积的 98%。

[0044] 3、可以校正基线失真较大的谱图。

[0045] 4、可以避免在复杂谱图中由于基线校正引入的的谱峰失真。在如图 4 所示中,利用现有的方法校正后的谱图都产生了一定程度的畸变,这种畸变直接影响谱图的积分面积,现有的各种基线校正方法对该谱图部分信号积分面积造成了 10%~26% 的误差,而采用新方法进行基线校正后积分面积值保持在 98%。

#### 附图说明

[0046] 图 1 是一种自动基线校正方法中基线识别方法流程图。

[0047] 图 2 是一种自动基线校正方法中基线模型构建方法流程图。

[0048] 图 3 是基线校正方法中采用不同的基线识别方法效果比较示意图。

[0049] 其中 (a) 原始的拟合 NMR 谱图数据 ;(b) 连续小波变换结合滑动窗口法进行基线识别 ;(c) 在 b 基础上加入普通迭代阈值法进行宽峰识别 ;(d) 基线识别采用了本文提出的方法。

[0050] 图 4 是基线模型构建时采用不同的方法进行基线校正结果对比示意图。

[0051] 其中 (a) 原始的代谢组学 NMR 谱图数据 ;(b) 利用多项式拟合方法进行基线校正 ;(c) 利用 Whittaker 滤波进行基线校正 ;(d) 利用 Hodrick-Prescott 滤波进行基线校正 ;(e) 利用本发明中采用的基线校正方法。

#### 具体实施方式

[0052] 下面结合附图,对本发明的技术方案作进一步的详细描述。

[0053] 实施例 1 :

[0054] 如图 1,图 2 所示,一种自动基线校正方法,包括以下步骤 :

[0055] 步骤 1(输入谱图数据 1)、利用傅里叶变换和相位校正处理采集到的原始数据,得到谱图数据 ;

[0056] 步骤 2(进行连续小波变换 2)、利用连续小波变换对谱图数据进行计算,得到谱图的数值导数 ;

[0057] 步骤 3(利用滑动窗口识别窄峰信号 3)、利用滑动窗口法和谱图的数值导数识别谱图中的窄峰信号,得到窄峰信号区间 ;

[0058] 步骤 4(利用迭代阈值法识别宽峰信号 4)、利用迭代阈值法识别谱图数据中的宽峰信号,得到部分宽峰信号区间 ;

[0059] 步骤 5(对宽峰信号进行轮廓拟合 5)、根据部分宽峰信号区间,利用轮廓拟合方法拟合出宽峰信号的轮廓 ;

[0060] 步骤 6(计算完整宽峰信号区间 6)、将谱图数据中大于步骤 4 中宽峰信号的轮廓最大值 3% 的点都作为宽峰信号点,从而得到完整宽峰信号区间 ;

[0061] 步骤 7(得到谱图的基线区间 7)、将谱图数据减去窄峰信号区间和完整宽峰信号区间得到谱图的基线区间 ;

[0062] 步骤 8(初始化权重数组 8)、利用步骤 7 的得到的基线区间初始化权重数组,若谱图数据点位于信号区间,则权重数组值初始化为 0,若谱图数据点位于基线区间,则权重数组值初始化为 1;

[0063] 步骤 9(计算基线模型 9)、利用权重数组和滤波算法计算基线模型;

[0064] 步骤 10(基线校正 10)、利用步骤 9 中的基线模型对步骤 1 所述的谱图数据进行校正,得到校正后的谱图数据;

[0065] 步骤 11(判断谱图中是否存在由于基线校正引入的失真 11)、逐个判断校正后的谱图数据中的各个谱峰,若正峰信号中出现负数部分则认为该谱峰产生了畸变,若负峰信号中出现了正数部分,则认为该谱峰产生畸变,如果谱峰产生畸变则执行步骤 12,如果谱峰没有产生畸变则退出计算得到最终基线模型和校正后的谱图数据;

[0066] 步骤 12(根据失真找出“准基线点” 12)、将步骤 11 中谱峰产生畸变最严重的点设置为准基线点;

[0067] 步骤 13(修改权重数组 13)、将步骤 12 中的准基线点对应的权重数组设置为 1 并返回步骤 9 重新计算基线模型。

[0068] 步骤 3 中利用滑动窗口法包括以下步骤:

[0069] 步骤 3.1、根据步骤 2 中得到的谱图数据导数计算噪声水平  $\sigma_{\text{noise}}$ ;

[0070] 步骤 3.2、设定阈值为  $n \times \sigma_{\text{noise}}$ ,其中  $n$  为参数值;

[0071] 步骤 3.3、比较滑动窗口的高度与步骤 3.2 中确定的阈值的大小,如果滑动窗口的高度大于阈值,则滑动窗口的中心点位于窄峰信号区间;如果滑动窗口的高度小于等于阈值,则滑动窗口的中心点位于基线区间。

[0072] 步骤 4 中迭代阈值法包括以下步骤:

[0073] 步骤 4.1、利用谱图数据计算出迭代阈值,迭代阈值基于公式  $\text{MEAN}+3*\text{SDEV}$ ,MEAN 是谱图数据的平均值,SDEV 是谱图数据的标准差;

[0074] 步骤 4.2、将谱图数据与迭代阈值进行比较,若谱图数据存在比迭代阈值大的数据点,则进行步骤 4.3;若谱图数据中所有的点都小于迭代阈值,则停止迭代,得到所有的宽峰信号,从而得到部分宽峰信号区间;

[0075] 步骤 4.3、将步骤 4.2 中得到的比迭代阈值大的谱图数据作为宽峰信号点,比迭代阈值小的谱图数据作为步骤 4.1 中的谱图数据并返回步骤 4.1。

[0076] 步骤 3.2 中参数值  $n$  为 3,所述的步骤 3.3 中滑动窗口的长度为整个谱图数据的宽度的 0.2%。

[0077] 步骤 5 中的轮廓拟合是基于最小化惩罚函数  $P$ :

$$[0078] \quad P = \sum_{\Omega} P(\Omega)$$

[0079] 其中:当  $E(\Omega)-S(\Omega) \geq 0, P(\Omega) = (E(\Omega)-S(\Omega))^2$ ;当  $E(\Omega)-S(\Omega) \leq 0, P(\Omega) = (f \times (E(\Omega)-S(\Omega)))^2$ ,  $E(\Omega)$  是需要拟合的宽峰信号,  $S(\Omega)$  是谱峰信号,  $f$  是轮廓调整参数,  $\Omega$  为谱图频率坐标。

[0080] 步骤 9 中的滤波算法基于最小化自定义目标函数:

$$[0081] \quad F(m) = \sum_{i=1}^N w(S(m-1)_i) \times (S(m-1)_i - \tau(m)_i)^2 +$$



$$[0082] \quad \lambda \sum_{i=2}^{N-1} \{ \tau(m)_{i+1} - \tau(m)_i \} - [ \tau(m)_i - \tau(m)_{i-1} ] \}^2$$

[0083] 式中： $\tau(m)$  表示第  $m$  次迭代所需计算的基线模型数据， $S(m-1)_i$  是第  $m-1$  次迭代后基线校正的谱图数据， $m$  是指第  $m$  次迭代， $i = 1, 2, \dots, N$ ， $N$  表示数据长度， $w(S(m-1)_i)$  是基于前一次校正后谱图计算得到的谱图权重数组，当  $S(m-1)_i$  是信号点时， $w(S(m-1)_i) = 0$ ；当  $S(m-1)_i$  是噪声点时， $w(S(m-1)_i) = 1$ 。

[0084] 如图 3 所示，其中 (a) 原始的拟合 NMR 谱图数据；(b) 连续小波变换结合滑动窗口法进行基线识别；(c) 在 b 基础上加入普通迭代阈值法进行宽峰识别；(d) 基线识别采用了本文提出的方法。

[0085] 如图 4 所示，其中 (a) 原始的代谢组学 NMR 谱图数据；(b) 利用多项式拟合方法进行基线校正；(c) 利用 Whittaker 滤波进行基线校正；(d) 利用 Hodrick-Prescott 滤波进行基线校正；(e) 利用本发明中采用的基线校正方法。

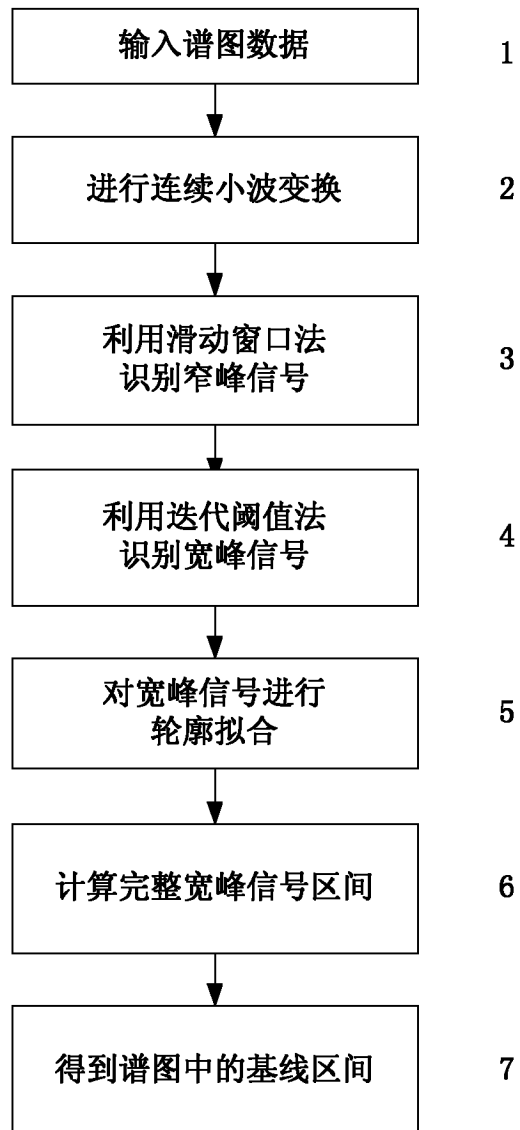


图 1

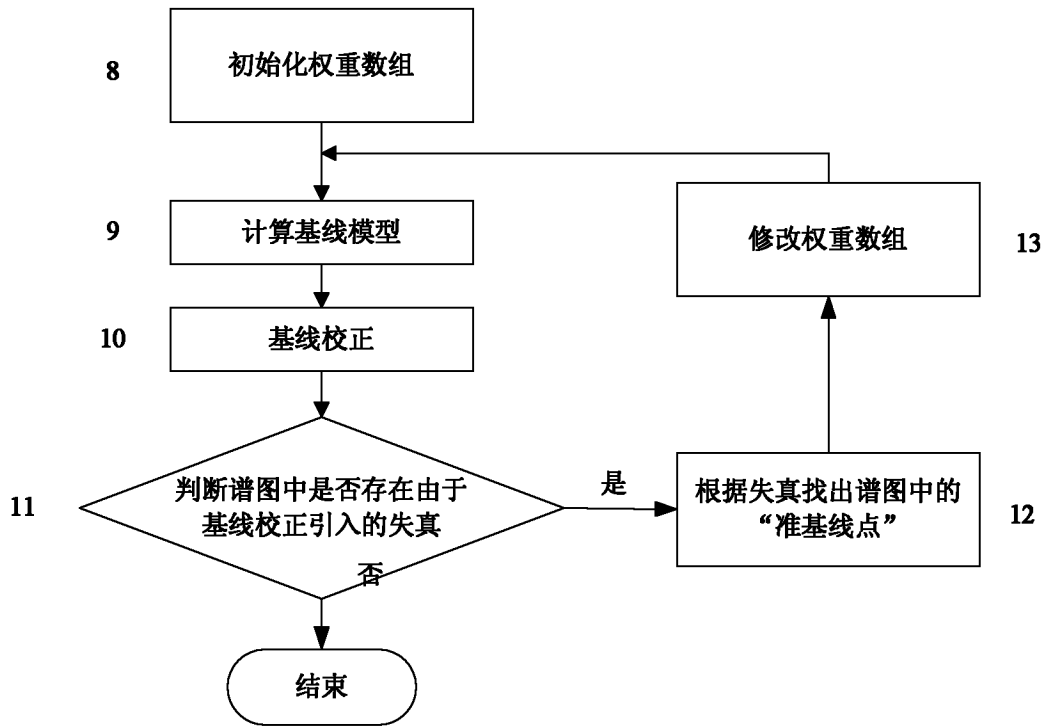


图 2

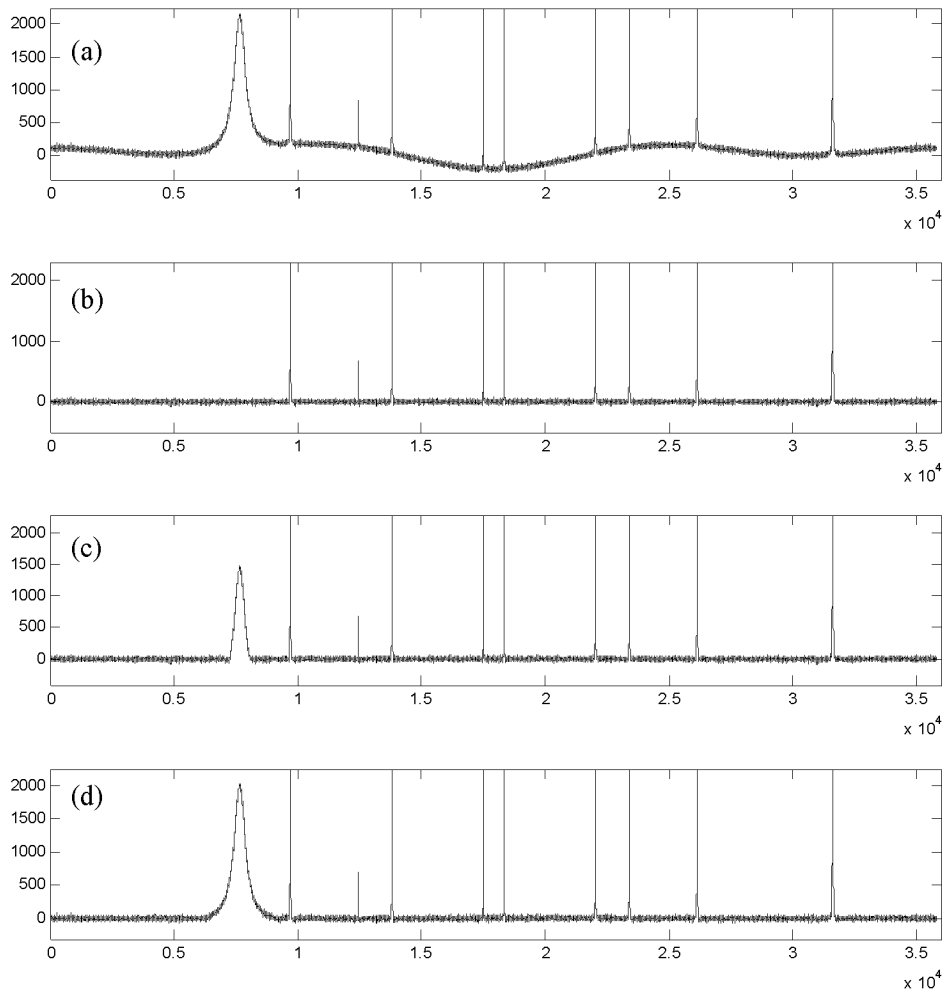


图 3

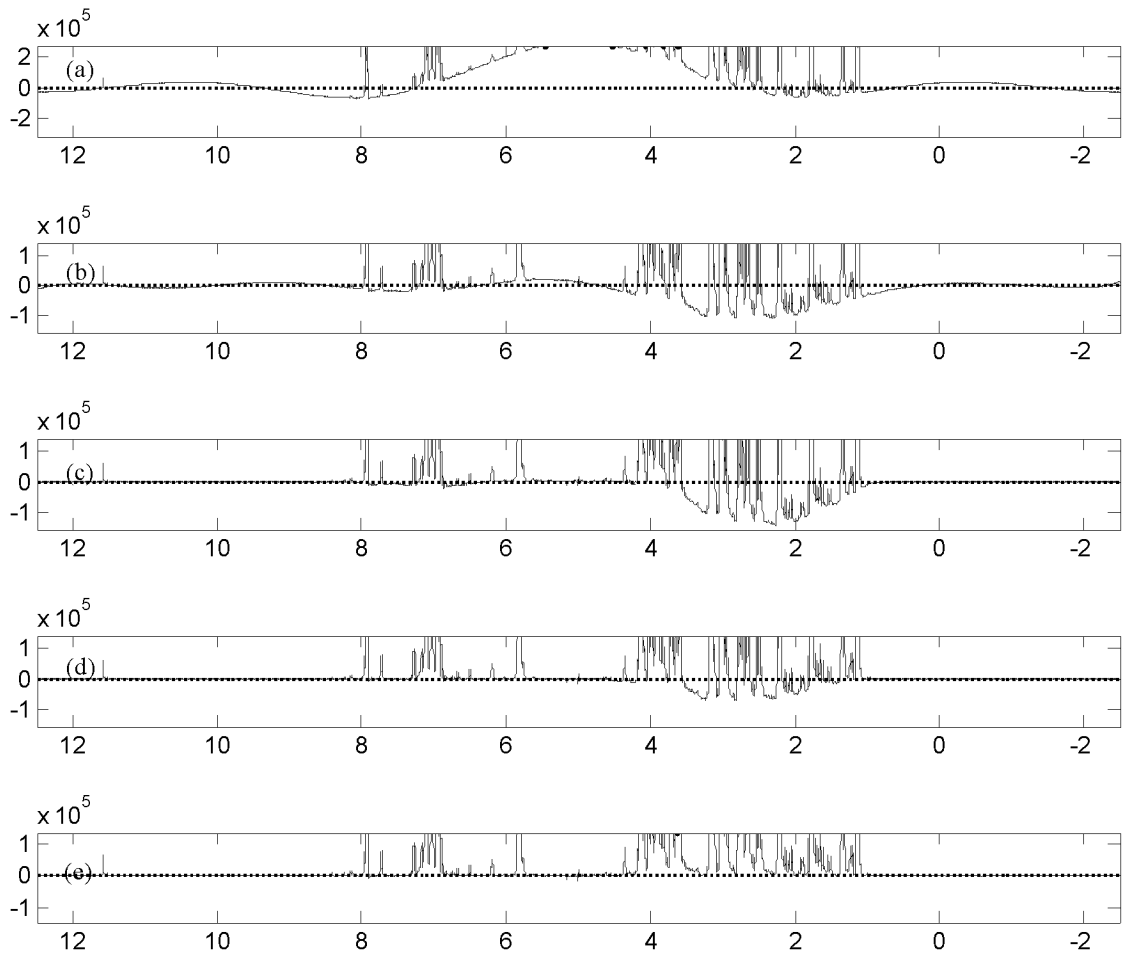


图 4