# 基于小波变换的质谱基线校正算法研究

罗勇<sup>\*</sup> 于佳佳 周旭 薛兵 金永星 唐朝阳 (上海裕达实业有限公司上海 200240)

# Baseline Correction Algorithm for Mass Spectrometry Based on Wavelet Transform

LUO Yong<sup>\*</sup>, YU Jiajia, ZHOU Xu, XUE Bing, JIN Yongxing, TANG Chaoyang (Shanghai Yuda Industrial Co., Ltd, Shanghai 200240, China)

Abstract The mass spectrometer is an accurate measurement instrument widely used in life sciences, food safety, environmental monitoring, industrial analysis, state security and other fields. Data processing is the critical link that affects the analysis results of mass spectrometry. In order to reduce the baseline drift during data acquisition, a baseline correction algorithm based on wavelet transform is proposed. Firstly, carry out single-layer wavelet decomposition to the original mass spectral signal several times, perform reconstruction to single-layer wavelet at the same time, calculate the signal-to-noise ratio of each single-layer, and acquire the mass spectral signal after noise reduction through the signal-to-noise ratio comparison method. Secondly, carry out single-layer wavelet decomposition to this signal several times to obtain wavelet details and the wavelet approximated frequency of each layer, divide the two frequencies to get the ratio and compare the ratio values of each layer, select the layer with the largest value as the best layer of wavelet decomposition. Finally, reset the wavelet approximated coefficient under the best layer of wavelet decomposition to zero and carry out wavelet reconstruction to obtain the mass spectral signal after baseline correction. Through the verification of experiments, this algorithm can precisely get the best layer of wavelet decomposition. Based on the retaining of real mass spectral signal as much as possible, remove the distribution of low frequency mass spectral baseline and high frequency noises; the baseline correction is sufficient and good results are achieved.

Keywords Mass spectrometer, Baseline correction, Wavelet decomposition, Wavelet reconstruction

摘要 质谱仪是广泛应用于生命科学、食品安全、环境监测、工业分析、国家安全等领域的精密测量仪器。数据处理是 影响质谱分析结果的关键环节,为了降低质谱仪数据采集过程中基线漂移,文章提出了一种基于小波变换的质谱基线校正算 法。首先对原始质谱信号多次进行单层小波分解,同时进行单层小波重构,计算得每层的信噪比并通过信噪比比对法获得降 噪后的质谱信号。接着对该信号进行多次单层小波分解,得到每层小波细节和小波近似的频率,将两个频率相除得到比值, 比较每层的比值大小并选择最大比值所在层作为小波分解的最佳层数,最后将小波分解最佳层数下的小波近似系数置零后 进行小波重构,获得基线校正后的质谱信号。通过实验验证发现,该算法可准确得出小波分解的最佳层数,在大幅保留真实 质谱信号的基础上去除低频质谱基线以及高频噪声的干扰,质谱基线校正充分,实际应用效果明显。

关键词 质谱仪 基线校正 小波分解 小波重构 中图分类号: O657.63 文献标识码: A doi: 10.13922/j.cnki.cjvst.202308010

质谱仪的原理是将被分析物离子化后,根据 不同质荷比的离子碎片在质量分析器中的不同表 现实现离子分离,进而测得不同离子数目最终得 到质谱图。质谱仪具有定性定量分析能力强、检 出限低的特性,近年来在各大领域应用越来越广, 检测需求不断提升,同时对质谱仪的性能提出了 更高的要求<sup>[1]</sup>。在实际应用场景中,仪器元器件老 化、温漂、内部高频高压等电控模块干扰和质谱 仪实际工作环境干扰等因素导致测得的质谱数据 受基线漂移的影响,进而影响质谱分析结果的准 确性和精确度。因此,基线校正是质谱仪数据处 理中的关键部分,得到了越来越多的关注和研究。 人们已经提出了大量基线校正方法来识别和降低 基线干扰,传统基线校正算法主要包括多项式拟 合法、移动平均法、小波变换法等<sup>[2]</sup>。许多学者在 这些算法基础上提出了新的思路,Morháč等<sup>[3]</sup>提出 了一种非线性迭代去除峰值方法,该方法计算速 度快但只能处理简单的基线。Gan等<sup>[4]</sup>提出了一 种具有自动阈值的迭代多项式拟合方法,但也存 在与其他多项式拟合方法<sup>[5]</sup>类似的问题,当多项 式的阶数选取不当时,基线校正效果并不理想。 还有一些研究尝试采用小波包变换来拟合基线<sup>[6]</sup>, 但是由于小波分析中难以避免的小波基选择、分 解层数选择、阈值及阈值函数选择等复杂问题,他 们也无法将基线在小波变换域中与质谱真实信号 准确分离。为此,本文提出一种可定量的小波分 解最佳层数计算方法,基于此方法进行小波变换 实现质谱信号基线校正。

#### 1 质谱数据采集原理

本文所用仪器为上海裕达实业有限公司研发的数字离子阱质谱仪 ASV MS5201。该质谱仪主要由进样系统、离子源、聚焦传输系统、离子阱质量分析器、离子检测系统、仪器电控组件和数据分析系统等部分组成,真空系统由涡旋泵和分子泵组成。硬件连接框图如图1所示。



图1 质谱仪硬件连接框图



其中,仪器电控组件由主控板、高频高压模块、 偶极激发模块和直流放大模块等组成。主控板基 于可编程逻辑阵列技术,用于精确的控制系统时序、 提供电压信号以及与其他模块进行信息交互。高 频高压模块为离子阱极片提供离子捕获波形,偶极 激发模块将主控板产生的偶极激发信号和高频高 压模块产生的高频高压波形进行耦合来驱动离子 阱。直流放大模块输出的直流电压驱动透镜系统 极片及离子阱的前端盖和后端盖。探测器产生的 电流信号经过小信号放大板的滤波、电流-电压转 换、放大后传输到主控板的高速 AD 采集模块<sup>[7]</sup>,经 上位机软件处理后得到质谱图。

理想状态下,该质谱信号不含低频项。对此理 想信号实施快速傅里叶变换,即可获得理想的质谱 分析数据。然而实际应用中采集得到的质谱信号 会难以避免的含有低频基线信号,导致得到的实际 信号与理想信号偏差较大,迫切需要对现实质谱数 据实施基线校正,从而得到理想质谱分析结果。

## 2 小波理论概述

#### 2.1 小波变换原理

小波变换是一种基于傅里叶变换的时频域信号分析工具。由于小波变换具有强大的多分辨率 分析能力和有效去噪能力,常被用于处理质谱数据 中的质谱信号和基线信号。某些条件下,将原始函 数进行缩放或平移即可得到一个函数族:

$$\varphi_{a,b} = \frac{1}{\sqrt{|a|}} \varphi\left(\frac{t-b}{a}\right) (a,b \in R, a \neq 0)$$
(1)

式中, a表示缩放变量, b表示平移变量,  $\varphi(t)$ 表示小 波基。

若用W<sub>f</sub>(m,n)表示信号函数f(t)的离散小波变换,则有:

$$\begin{split} W_f(m,n) &= < f(t), \varphi_{m,n}(t) > \\ &= a_0^{-m/2} \int_{-\infty}^{+\infty} f(t) \varphi(a_0^{-m}t - nb_0) \mathrm{d}t \end{split} (2)$$

式中, $a_0 > 1$ , $m \in Z$ , $b_0 \in R$ , $n \in Z_o$  m和n均为离散量。

#### 2.2 小波分解

被分析信号的时频局部化分析过程中,它的频 率窗和时间窗都能不断变化,具有较强的自适应性。 图 2 为被分析信号 X 的小波分解完整过程。



Fig. 2 Wavelet decomposition

上图中当原始信号 X 执行一层小波分解时,可 以得到 cA<sub>1</sub> 及 cD<sub>1</sub>,分别表示小波近似值及小波细 节值;继续将 cA<sub>1</sub>分解到第二层时,可以得到 cA<sub>2</sub> 和 cD<sub>2</sub>。经过若干层小波分解,可得到更多层的小 波近似值及小波细节值。其中,小波近似值可代表 信号 X 的低频部分,小波细节值代表信号 X 的高频 部分。随着分解层数的增加,信号频率变得更低, 小波近似值变得更小。然而由于离散信号的点数 有限,小波分解最大层数 N 必须满足信号离散点数 大于等于 2<sup>N</sup> 的条件。在质谱分析过程中,我们发现 有用信号主要集中在高频区间,而干扰信号通常分 布于低频部分,因此在处理时将所有低频部分的小 波系数置零即可去除质谱信号中的基线干扰。

#### 2.3 常用小波基函数介绍

小波变换中使用的小波基函数具有不唯一性, 小波基的选择是小波分析的关键环节之一。以下 为几种常见的小波基函数。

(1)Harr小波

Harr 小波的定义如下:

$$\Psi_{H}(\mathbf{x}) = \begin{cases} 1 & 0 \le x \le 1/2 \\ -1 & 1/2 \le x \le 1 \\ 0 & \pm \mathbf{t} \end{aligned} (3)$$

$$\int_{-\infty}^{\infty} \Psi(x)\Psi(x-n)dx = 0 \qquad (4)$$

Harr 小波是 Harr A 提出的一个具有紧支撑的 正交小波函数,同时也是最简单的一个函数<sup>[8]</sup>。 (2)Daubechies (dbN) 小波

Daubechies(dbN)小波函数是由 Ingrid Daubechies 在对尺度取 2 的整数幂条件下的小波变换深入研究提出的小波函数<sup>[9]</sup>。Daubechies 小波可以简写为 dbN,其中 N 为小波的阶数。dbN 也具备紧支撑正交特性,但是, dbN(N $\neq$ 1)不具备对称性的特征, 也无显示表达式。

(3)SymletA(symN)小波

SymletA 小波函数是 db 小波函数的一种改进, 具有近似对称性、紧支撑性及双正交性。SymletA 函数一般表示为 symN, 其中 *N*=2,3,4,5,6,7,8<sup>[9]</sup>。

#### (4)Coiflet (coifN) 小波系

Coiflet 小波基具有紧支撑正交特性,是由 Daubechies 构造的小波函数,函数有 coifN (N=1, 2, 3, 4, 5) 这一系列, Coiflet 小波基比 dbN 小波基对称 性更好,而且 coifN 与 db3N、sym3N 具有相同的支 撑长度,与 db2N、sym2N 具有相同的消失矩数目<sup>[9]</sup>。

小波基函数的主要属性有正交性、紧支撑性、 对称性、正则性及高阶消失矩等,上述四种小波基 函数的属性如表1所示。

表 1 小波基函数属性 Tab. 1 Wavelet basis function properties

			-	_
	Harr	dbN	symN	coifN
正交性	是	是	是	是
双正交性	是	是	是	是
紧支撑性	是	是	是	是
支撑长度	1	2 <i>N</i> -1	2 <i>N</i> -1	6 <i>N</i> -1
对称性	对称	不对称	近似对称	近似对称
正则性	否	否	否	否
消失矩	1	N	N	2N

实际小波分析中小波基的选择并没有固定的 准则,通常按照被分析信号的特征以及小波基函数 本身的属性来进行选取<sup>[10]</sup>。为了使分解后的能量更 加集中,本文选择消失矩最大(2*N*)的 coif*N* 小波基 作为本文小波变换的基函数。

#### 2.4 小波分解层数

质谱信号小波分解过程中使用的分解层数是 影响质谱基线校正效果的又一重要因素。若分解 层数选取过大,将无法达到基线校正效果;若分解 层数选取过小,将一并去除有用质谱信号,造成质 谱信号失真,并且无法使用小波重构获取真实的质 谱信息。以往分解层数大多是依据经验值而来,无 法达到最佳的基线校正效果,由此本文提出了可准确计算最佳分解层数的方法,基于此层数进行小波 分解与重构实现质谱基线校正。

## 3 算法步骤及实验结果

本文提出的基于小波变换的质谱基线校正算 法流程如图 3 所示,主要步骤如下:





步骤 1:选用 coifN 小波基作为小波变换的基函数,对原始质谱信号 S 单层小波分解,重构小波近似 值和小波细节值,计算重构质谱信号的信噪比记为 SNR<sub>1</sub>。

步骤 2: 将原始质谱信号 S 小波分解至第二层, 重构分解后得到的小波近似值和小波细节值, 计算 重构的质谱信号的信噪比记为 SNR<sub>2</sub>。将 SNR<sub>1</sub>与 SNR<sub>2</sub>进行比较,若 SNR<sub>1</sub>比 SNR<sub>2</sub>大,则进入下一步 骤,否则继续实施三层小波分解,求得重构信号信 噪比 SNR<sub>3</sub>,与 SNR<sub>2</sub>进行比较,若 SNR<sub>2</sub>比 SNR<sub>3</sub>大, 则进入下一步骤,否则继续实施四层小波分解,直 至进入下一步骤。

步骤 3: 若第 *i* 次分解时发现最大信噪比 SNR<sub>i</sub>, 停止流程, *i* 即为质谱信号小波最佳分解层数。然 后对原始质谱信号 S 实施 *i* 层小波分解并重构,得 到除噪后的质谱信号 T。

步骤 4: 若质谱信号 T 的离散点数为 m,则小波 分解最大层数 N 与 m 间存在关系:  $m \ge 2^{N}$ 。

步骤 5: 将除噪后的质谱信号 T 完全分解, 计算 每层分解的小波近似值cA<sub>n</sub>(n = 1,2,…N)和小波细节 值cD<sub>n</sub>(n = 1,2,…N)的能量。将两个频率之和Fd<sub>n</sub>作 为小波细节频率, 即:

$$Fd_n = \sum_{i=1}^{ld_n} cD_n(i)^2 + \sum_{i=1}^{la_n} cA_n(i)^2$$
 (5)

式中,  $ld_n$ 代表 $cD_n$ 的离散值数量,  $la_n$ 代表 $cA_n$ 的离散 值数量<sup>[8]</sup>。

步骤 6: 将上一步中小波细节和小波近似能量 差值的绝对值Fa,作为小波近似频率,即:

$$Fa_n = \left| \sum_{i=1}^{ld_n} cD_n(i)^2 - \sum_{i=1}^{la_n} cA_n(i)^2 \right|$$
 (6)

步骤 8: 对质谱信号 T 实施 j 层小波分解, 将分 解结果中的低频小波系数置零后进行小波重构, 即 可获得基线校正后的质谱信号 M, 至此算法流程 完成。

步骤 8: 对质谱信号 T 实施 j 层小波分解, 将分 解结果中的低频小波系数置零后进行小波重构, 即 可获得基线校正后的质谱信号 M, 至此算法流程 完成。

本算法实验验证采用 ASV MS5201 型数字离 子阱质谱仪检测冻干机腔室中的水氧含量时得到 的质谱图作为数据源。在不同的冻干机腔室中前 后进行了三组实验分别计算出最佳分解层数为 8、 10、6,将去噪后的质谱信号小波分解后把小波近似 系数置零,重构信号得到基线校正后的质谱信号。 为了比较本文算法与常用基线校正算法的校正效 果,基于 C++实现了各基线校正算法,计算了基线 校正后质谱信号的信噪比(*S/N*)、均方根误差

(RMSE)、峰值位置(pos)、峰值大小(V-max)等参 数作为基线校正效果、质谱图平滑性、保留信号特 征等的评价指标,计算各算法运行耗时(time)、所需 额外空间(space)作为算法实时性及时空间复杂度 的评价指标,得到各组结果如表 2-4 所示,基线校正 信号对比图如图 4-6 所示。

Tab. 2Baseline correction results of each algorithm (Group 1)								
	<i>S/N</i> (增量)	RMSE	pos	V-max(增量)	time/ms	space/Mb		
原始信号	856.67	/	32.08	3273	/	/		
多项式拟合法	1207.35(40.94%)	20.81	32.08	3127(-10.27%)	108	0.2		
移动平均法	1613.12(88.30%)	127.62	32.04	2142(-34.56%)	169	0.3		
小波包变换法	2424.95(183.07%)	23.94	32.08	3101(-5.26%)	880	0.8		
本文算法	2200.44(156.86%)	23.31	32.08	3110(-4.98%)	401	0.5		

#### 表 2 各算法基线校正结果(组1)

表 3 各算法基线校正结果(组 2)

Tal	<b>)</b> . 3	3	Base	line	correction	resu	lts of	eacl	1 al	lgorit	hm (	G	ìroup	2	)
-----	--------------	---	------	------	------------	------	--------	------	------	--------	------	---	-------	---	---

	S/N(增量)	RMSE	pos	V-max(增量)	time/ms	space/Mb
原始信号	901.12	/	17.36	4165	/	/
多项式拟合法	1083.60(20.25%)	42.12	17.32	3776(-9.36%)	203	0.3
移动平均法	1063.32(18.0%)	208.82	17.52	2598(-37.62%)	218	0.4
小波包变换法	2189.72(143%)	108.37	17.36	3303(-20.69%)	997	0.9
本文算法	2582.07(186.54%)	38.27	17.36	4000(-3.96%)	504	0.6

表 4 各算法基线校正结果(组 3)

	<i>S/N</i> (增量)	RMSE	pos	V-max(增量)	time/ms	space/Mb
原始信号	856.67	/	17.96	2082	/	/
多项式拟合法	1207.35(40.94%)	20.81	17.80	1888(-9.32%)	91	0.4
移动平均法	1613.12(88.30%)	127.62	17.90	1299(-37.61%)	104	0.4
小波包变换法	2424.95(183.07%)	23.94	17.94	1651(-20.70%)	750	0.7
本文算法	2200.44(156.86%)	23.31	17.96	2010(-3.46%)	380	0.6

由表 2-4 及图 4-6 可以看出,多项式拟合法计 算速度最快、所需额外空间最小,但是峰型与原始 信号区别不大, 信噪比提高幅度最低, 均方根误差 最小,峰值位置基本不变,峰值大小降低不多,基线 校正效果不明显;移动平均法计算速度较快、所需 额外空间较小,得到较平滑的质谱信号,同时信噪 比有了一定幅度提高,峰值位置基本不变,但是均 方根误差最大,最高峰强大幅降低,峰型被明显削 平,说明该算法会降低质谱分辨率,不宜单独使用 该算法基线校正:小波包变换法信噪比提高幅度大, 均方根误差较小,峰值位置基本不变,最高峰强降 低幅度较小,能在去除基线干扰的同时保留质谱峰 信号,但是该算法无法保证每次都能得到最优值, 算法稳定性较差,所用时间、空间大;本文算法得到 的处理结果表明该算法也能在很好地去除基线干 扰的同时保留质谱峰信号,信噪比提高幅度最大, 均方根误差最小,峰值位置保持不变,最高峰强降 低幅度最小同时能兼顾所用时间、空间大小。三组 实验结果较为一致,说明本算法稳定性好。综上所 述,与常用的基线校正算法相比,本文提出的基线 校正算法基线校正效果良好,能在很好地去除基线 干扰的同时保留有用信号,又能兼顾稳定性及时空 间复杂度,适合质谱仪的实际应用场景。

## 4 结束语

质谱仪以其极低的检出限,超强的定性定量分 析能力,近年来在各大领域应用越来越广。随着科 技的不断发展,检测需求不断提升,它将发挥更加 重要的作用。质谱数据处理是获取真实质谱信号 的关键步骤,基线校正是其中的重要环节。本文提



图4 基线校正结果(组1)。(a) 多项式拟合法,(b) 移动平均法,(c) 小波包变换法,(d) 本文算法





图5 基线校正结果(组2)。(a) 多项式拟合法,(b) 移动平均法,(c) 小波包变换法,(d) 本文算法

Fig. 5 Baseline correction results (Group 2). (a) Polynomial fitting, (b) moving average, (c) wavelet packet transform, (d) algorithm in this article



图6 基线校正结果(组 3)。(a) 多项式拟合法, (b) 移动平均法, (c) 小波包变换法, (d) 本文算法 Fig. 6 Baseline correction results (Group 3). (a) Polynomial fitting, (b) moving average, (c) wavelet packet transform, (d) algorithm in this article

出了可定量计算小波分解最佳层数的方法,基于此 方法进行小波变换以实现质谱信号基线校正。通 过使用自研质谱仪进行实验验证,结果表明与其 他常用质谱基线校正算法对比,本文提出的算法在 基线校正效果上有明显优势,同时具有良好的稳定 性和较低的时间、空间复杂度,充分说明了本算 法基线校正的可行性。可以看出,本文提出的基 于小波变换的基线校正算法具有一定的实际应用 意义。

#### 参考文献

- [1] Xu Z, Wang Y. Mass spectrometer[M]. Beijing: China Machine Press, 1995 徐贞林, 汪曣. 质谱仪器 [M]. 北京: 机械工业出版社, 1995(in chinese)
- [2] Kristian H L, Trygve A, Bjørn-Helge M. Optimal choice of baseline correction for multivariate calibration of spectra[J]. Applied spectroscopy, 2010, 64(9):1007-1016 (克里斯蒂安·霍夫德·利兰, 特里格夫·阿尔莫伊, 比约恩-赫尔格·梅维克.光谱多元校正中基线校正的最佳选择[J].应用光谱学, 2010, 64(9): 1007-1016(in chinese))
- [3] Morháč M, Kliman J, Matoušek V, et al. Background elimination methods for multidimensional coincidence *γ*ray spectra[J]. Nuclear Instruments & Methods in Physics

Research, 1997, 401(1): 113-132 (莫尔哈克, 克里曼, 马图谢克, 等. 多维复合 y 射线光谱的背景消除方法 [J]. 物理研究中的核仪器与方法, 1997, 401(1): 113-132(in chinese))

- [4] Gan F, Ruan G, Mo J. Baseline correction by improved iterative polynomial fitting with automatic threshold[J]. Chemometrics & Intelligent Laboratory Systems, 2006, 82(1-2): 59-65 (甘峰, 阮贵华, 莫金垣. 基于自动阈值的多项式拟合基线校正 [J]. 化学计量与智能实验系统, 2006, 82(1-2): 59-65(in chinese))
- [5] Baek S J, Park A, Shen A, et al. A background elimination method based on linear programming for raman spectra[J]. Journal of Raman Spectroscopy, 2011, 42(11): 1987–1993 (白柱俊, 派克艾伦, 沈爱国, 等. 一种基于线性规划的拉曼光谱背景消除方法 [J]. 拉曼光谱, 2011, 42(11): 1987–1993(in chinese))
- [6] Shi J, Wang J. A new method of adaptive filtering and wavelet transform to filter baseline shift[J]. Journal of Information Technology Research, 2018, 11(3): 119–134 (史健婷, 王建才. 一种基于自适应滤波和小波的基线 漂移校正的新型方法 [J]. 信息技术研究, 2018, 11(3): 119–134(in chinese))
- [7] Xue B. Research on digital ion trap mass spectrometer

and novel ion excitation methods[D]. Shanghai University, 2016 (薛兵. 数字离子阱质谱仪及新型离子激发方 法的研究 [D]. 上海大学, 2016 (in chinese))

- [8] Li Q, Xu Y, Li X, et al. Research on an improved wavelet baseline correction method[J]. Photoelectron Laser, 2009(5): 4 (李庆波, 徐玉坡, 李响,等. 一种改进的小波 基线校正方法的研究 [J]. 光电子·激光, 2009(5):4(in chinese))
- [9] Shen Z. Proficient in matlab signal processing[M]. Bei-

jing: Tsinghua University Press, 2015(沈再阳. 精通 MATLAB 信号处理 [M]. 北京:清华大学出版社, 2015 (in chinese))

[10] Feng A, Qiu J, Zheng Y, et al. Simulation analysis of mind evolutionary algorithm optimizing wavelet packet denoising[J]. Computer Simulation, 2020, 37(7):6 冯安 安, 岳建海, 郑义,等. 思维进化算法优化小波包去噪仿 真分析 [J]. 计算机仿真, 2020, 37(7):6(in chinese)