

文章编号:1002-2082(2019)03-0461-07

# 基于优化迭代算法的细胞荧光光谱解析

都 月, 孟晓辰, 祝连庆

(北京信息科技大学 光电测试技术北京市重点实验室, 北京 100192)

**摘要:** 使用光谱测量方法进行细胞多色荧光分析时, 发射光谱会产生部分光谱重叠, 为定性和定量分析造成了一定的困难。为此, 提出基于优化迭代算法的细胞荧光光谱解析算法, 建立重叠峰模型并确定单峰顶点; 根据每次构造峰面积的大小, 重新确定构造峰的构造方式, 最终得到模拟峰的顶点及面积信息。利用该算法对高斯函数叠加形成的重叠峰进行解析, 并与常规方法进行对比, 结果表明优化迭代算法解析误差稳定在 0.15% 以内; 加入随机噪声后, 解析误差可稳定在 0.85% 以内, 均优于另外两种算法。此外, 计算了该算法下的迭代效率, 结果表明该算法较常规方法提高了 32.2%。

**关键词:** 生物医学; 光谱重叠峰; 优化迭代; 高斯模型

中图分类号: TN201; TH789

文献标志码: A

DOI: 10.5768/JAO201940.0303001

## Novel algorithm for analysis of cellular fluorescence spectrum based on optimization iteration

DU Yue, MENG Xiaochen, ZHU Lianqing

(Beijing Key Laboratory for Optoelectronic Measurement Technology, Beijing Information Science and  
Technology University, Beijing 100192, China)

**Abstract:** When using the spectral measurement method for multi-color fluorescence analysis of cells, the emission spectrum can produce partial spectral overlap, which poses certain difficulties for qualitative and quantitative analysis. Aiming at the problem, a novel algorithm was developed to effectively decompose the overlapped spectral peaks based on optimization iteration. First, the separation model of overlapped peak was built up to obtain the peak vertex of overlapping peaks. And then, according to the size of each constructed peak area, the structure of the constructed peak was re-determined, and finally the vertex and area information of the simulated peak was obtained. Moreover, the algorithm was used to analyze the overlapping peaks formed by the superposition of Gaussian functions, and compared with the conventional methods. The results show that the analytical error of the optimized iterative algorithm is stable within 0.15%, while it can be stabilized within 0.85% after adding random noise, which are all better than the other two algorithms. In addition, the iterative efficiency under the algorithm was calculated, and the results show that the algorithm is 32.2% higher than the conventional method.

**Key words:** biomedicine; spectral overlapping peaks; optimization iteration; Gaussian model

收稿日期: 2018-08-13; 修回日期: 2019-01-10

基金项目: 国家自然科学基金(61605010); 教育部“长江学者和创新团队”发展计划(IRT\_16R07)

作者简介: 都月(1996—), 女, 硕士研究生, 主要从事生物医学检测技术及仪器研究。E-mail: 925334037@qq.com

通讯作者: 孟晓辰(1985—), 女, 博士, 硕士生导师, 主要从事生物医学检测技术与仪器、光电精密测试技术研究。

E-mail: mengxc@bistu.edu.cn

## 引言

光谱测量方法具有高效、快速、环保、选择性好、测量精度高等优点,应用领域非常广泛。近年来,为了更加精细地分析细胞特性而进行多色法荧光测量,即使用多种荧光染料标记微粒,并且使用光学扫描显微镜来检测各个荧光染料被激发后所发射的荧光。然而目前所使用的荧光染料的激发或发射光谱较宽,发射光谱范围会有一定的干扰重叠,形成重叠峰,这是定性和定量分析的一大障碍<sup>[1]</sup>。因此,必须提出有效的理论和方法来分离光谱信号重叠峰。

针对此问题,国内外进行了深入的研究。早期,Kauppinen 等<sup>[2]</sup>提出傅里叶变换的自去卷积技术,被广泛地用到光谱信号的解析,但较难选取合适的去卷积函数,且过程中噪音会呈指数增加,在提高分辨率的同时也降低了信噪比,易产生负旁瓣效应。随着现代信息处理方法的发展,冯飞等<sup>[4]</sup>利用小波变换对光谱信号进行多层次分解,结合高斯函数进行光谱拟合,但结果受小波母函数及分解层数的影响较大,且拟合具有较大的随机性;陶维亮等<sup>[5]</sup>提出将光谱信号转到小波时频域后进行分形维数计算,并利用人工神经网络法进行独立峰提取,可以较高精度分离重叠峰,但效率不高且算法较为复杂。此外,人们发现可以借鉴分析化学中的数学工具来研究光谱信号,为解析光谱重叠提供了新的方式。严彩娟等<sup>[6]</sup>通过单峰平移缩放对原始峰型进行修正,利用纵向迭代算法实现重叠峰分离;叶国阳等<sup>[7]</sup>在传统垂线法和切线法的基础上进行改进,提出交点垂线法及比例分配法,提高了峰面积的计算精度,但误差相对较大;朱强等<sup>[8]</sup>基于函数二分原理,重新规定了迭代方向,但结果受原始峰型的影响较大。上述方法虽对分析化学信号的重叠峰的解析取得了一定的运用效果,但仍具有一定的局限性。因此,为进一步有效分离细胞荧光重叠峰,本文提出一种优化迭代算法,既保证了迭代效率,又提高了分峰的准确性和面积的计算精度。

## 1 优化迭代算法实现

### 1.1 重叠峰数学模型

多色分析时细胞样本携带多种荧光染料,可以同时被相同波长的激发光源激发,得到不同的发射光谱。由于目前的荧光染料发射出的荧光具有比较宽的光谱范围,即使每种荧光的发射峰不同,发射光谱范围仍会有部分的光谱重叠<sup>[1]</sup>。但光谱并不是离散的谱线,而是表现为 Gaussian、Lorentz 线型等形式

的结构相近的谱峰<sup>[4]</sup>。因此,为方便研究,可假设构成重叠峰的单峰都是相似的。

以高斯模型  $ae^{-(x-b)^2/2c^2}$  作为各个单峰数学模型,其中: $a$  为高斯曲线峰值; $b$  为其对应的横坐标; $c$  为标准差(高斯 RMS 宽度)。实际情况中,三峰以上重叠的现象较少,所以本文仅对两峰及三峰重叠的情形进行分析。

两峰重叠时,建立对应模拟曲线图如图 1 所示。定义重叠峰为  $F$ ,构成重叠峰的单峰分别为  $A$  和  $B$ ,可用下式来表示两峰重叠时重叠峰的构成情况:

$$\begin{cases} A = a_1 e^{-(x-b_1)^2/2c_1^2} \\ B = a_2 e^{-(x-b_2)^2/2c_2^2} \\ F = A + B \end{cases} \quad (1)$$

(1)式即为双峰重叠时重叠峰的数学模型。如图 1,其中:横坐标表示模拟光谱范围;纵坐标表示荧光强度。

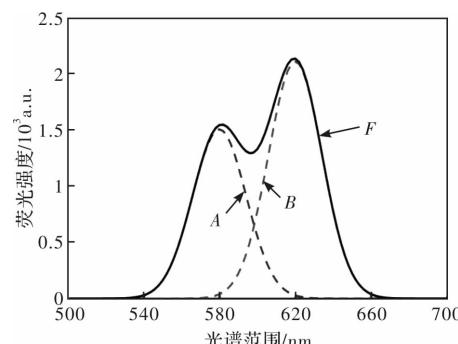


图 1 双峰重叠时重叠峰模型

Fig. 1 Overlapping peaks model of two peaks superimposed

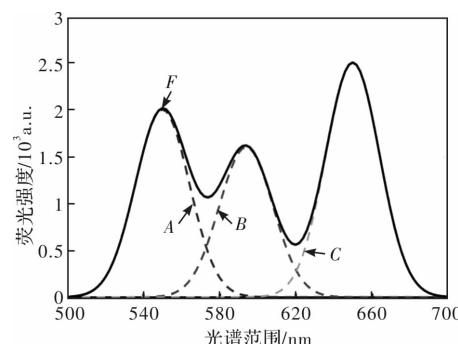


图 2 三峰重叠时重叠峰模型

Fig. 2 Overlapping peaks model of three peaks superimposed

同理,三峰重叠时,建立对应模拟曲线图如图 2 所示,同时满足(2)式:

$$\begin{cases} A = a_1 e^{-(x-b_1)^2/2c_1^2} \\ B = a_2 e^{-(x-b_2)^2/2c_2^2} \\ C = a_3 e^{-(x-b_3)^2/2c_3^2} \\ F = A + B + C \end{cases} \quad (2)$$

(2)式即为三峰重叠时重叠峰的数学模型。其中A、B、C为构成重叠峰F的三个单峰。

## 1.2 优化迭代法的算法步骤

为便于描述和理解,本节仅讨论双峰重叠时的重叠峰分峰步骤。根据构成重叠峰的单峰的相似性,重叠峰的前沿和后沿的重叠程度很小,主要重叠部分大部分分布在除前后沿以外的中间部分。故可利用前后沿对中间重叠部分进行修正,通过迭代实现对单峰的逼近<sup>[6]</sup>。确定每次的构造峰顶点是本文提出算法的关键,具体步骤如下:

1) 寻找重叠峰F的两个顶点,并以此作为第1次构造峰的顶点。重叠峰的前沿即为第1次左构造峰 $g_{11}$ 的前沿,根据顶点求出纵向修正系数 $K_1$ ,对重叠峰后沿使用系数 $K_1$ 进行修正,得到 $g_{11}$ 的后沿,即构造出第1次左构造峰 $g_{11}$ ;用重叠峰F的面积减去左构造峰 $g_{11}$ 的面积,即得到第1次右模拟峰 $f_{12}$ 。同理,可得出第1次右构造峰 $g_{12}$ 和右模拟峰 $f_{12}$ ,如图3所示。

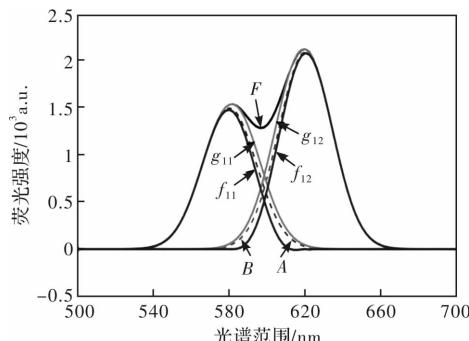


图3 第1次构造峰和模拟峰

Fig. 3 Primary tectonic and simulated peaks

2) 在此基础上计算出模拟峰 $f_{11}$ 、 $f_{12}$ 的顶点,作为第2次构造峰的顶点。求出纵向修正系数 $K_2$ ,对 $f_{11}$ 的前沿使用系数 $K_2$ 进行修正,得出第2次右构造峰 $g_{22}$ 的前沿;同时 $f_{12}$ 的后沿即为 $g_{22}$ 的后沿,即构造出第2次右构造峰 $g_{12}$ 。同理可得出第2次左构造峰 $g_{11}$ 。最后用重叠峰F分别减去左右构造峰,得到第2次模拟峰 $f_{21}$ 和 $f_{22}$ 。前两次得出的模拟峰如图4所示。

3) 按上述步骤继续修正,得到第3次构造峰 $g_{31}$ 和 $g_{32}$ 。第3次模拟峰构造方式如下:如果满足 $g_{31}+g_{32} < F$ ,则 $f_{31}=F-g_{32}$ , $f_{32}=F-g_{31}$ ,即模拟峰由重叠峰分别减去对应的构造峰得到;如果满足 $g_{31}+g_{32} > F$ ,则 $f_{31}=(g_{31}+g_{21})/2$ , $f_{32}=(g_{32}+g_{22})/2$ ,即模拟峰取本次构造峰和上次构造峰的中值;如果满足 $g_{31}+g_{32} = F$ ,则迭代停止, $g_{31}$ 和 $g_{32}$ 即为所要求的单峰。

4) 重复3)中的步骤,当 $f_{n1}+f_{n2}$ 无限接近F时,迭代停止。 $f_{n1}$ 、 $f_{n2}$ 即为所分离出的单峰。

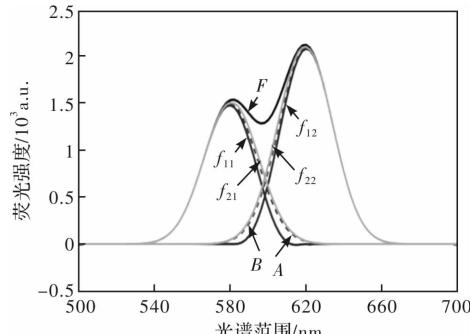


图4 第1、2次模拟峰

Fig. 4 The first & second simulated peaks

## 2 算法评价

在重叠峰分离的过程中,分离结果会受单峰的峰高比和原始重叠峰的分离度影响。本文分别模拟了相同峰高比下不同分离度和相同分离度下不同峰高比的重叠峰模型,然后分别算出原始重叠峰在优化迭代算法、原纵向迭代算法及平分迭代算法下的解析误差,进一步对比和验证优化迭代法的分离性能。在分离之前,首先对以下参数进行定义。

1) 峰高比:

$$r_H = \frac{H_A}{H_B} \quad (3)$$

其中 $H_A$ 和 $H_B$ 分别为两单峰顶点高度。

2) 分离度:

$$R = \frac{2(\mu_2 - \mu_1)}{(\omega_1 + \omega_2)} \quad (4)$$

其中: $\mu_1$ 和 $\mu_2$ 分别为两单峰位置; $\omega_1$ 和 $\omega_2$ 分别为单峰峰宽。R越大,则表明两峰分离越好。通常说当R=1.0时,分离度可达98%。

3) 解析误差:

$$\omega = \frac{S' - S}{S} \quad (5)$$

其中:S'为计算峰面积;S为真实峰面积。

4) 迭代效率比:

$$E = \left| \frac{\omega_2 - \omega_1}{\omega_1} \right| \quad (6)$$

通常,迭代效率用达到一定计算精度所需的迭代次数来表示。本文定义迭代效率比为相同迭代次数下,进行对比的两种方法的相对解析误差,用来表明迭代效率的提高程度。其中: $\omega_1$ 为进行对比的第1种方法达到该迭代次数时的解析误差; $\omega_2$ 为进行对比的第2种方法达到该迭代次数时的解析误差。迭代效率比越大,说明后种方法的迭

代效率相对前种方法越高。

首先分析了优化迭代算法分离重叠峰的解析误差。为进一步精确地讨论优化迭代法的性能,本次分析使用的重叠峰数学模型均由两高斯峰的叠加而得到。调整高斯函数的各个参数,首先分别计算了在分离度为 0.621 8 的情况下,峰高比分别为 1:1、1:2、1:3、2:1 和 3:1 时不同形状重叠峰的解析误差,及在峰高比为 2:3 的情况下,分离度分别为 0.466 3、0.559 6、0.632 8、0.777 2 和 0.808 3 时不同形状重叠峰的解析误差,分别迭代 6 次,并与传统

纵向迭代法及平分迭代法进行了比较。计算结果如表 1、2 所示,并作出对应柱状图如图 5、6 所示。

结果表明,分离度一定时,随着峰高比的增加,传统纵向迭代法的误差较大且会随之增加,平分迭代法的误差稳定在 1% 之内,优化迭代法的误差相对较小,且维持在 0.15% 之内;峰高比一定时,随着分离度的减小,传统纵向迭代法的误差较大且也会随之增加,平分迭代法的误差稳定在 0.5% 之内,优化迭代法的误差相对较小,且维持在 0.1% 之内。

表 1 分离度为 0.621 8 时的解析结果

Table 1 Analysis results at resolution of 0.612 8

峰高比	实际值		纵向迭代法		平分迭代法		优化迭代法	
	面积 s/a.u.	面积 s/a.u.	误差 r/%	面积 s/a.u.	误差 r/%	面积 s/a.u.	误差 r/%	
1:1	1 063.47	1 053.01	-0.98	1 069.72	0.59	1 063.85	0.04	
1:2	1 595.21	1 653.38	3.65	1 584.50	-0.67	1 593.35	-0.12	
1:3	2 126.94	1 988.56	-6.51	2 110.83	-0.76	2 128.99	0.10	
2:1	1 595.21	1 551.17	-2.76	1 587.55	-0.48	1 594.19	-0.06	
3:1	2 126.94	2 028.38	-4.63	2 142.92	0.75	2 129.29	0.11	

表 2 峰高比为 2:3 时的解析结果

Table 2 Analysis results at peak height ratio of 2:3

分离度	实际值		纵向迭代法		平分迭代法		优化迭代法	
	面积 s/a.u.	面积 s/a.u.	误差 r/%	面积 s/a.u.	误差 r/%	面积 s/a.u.	误差 r/%	
0.466 3	1 772.45	1 832.20	3.37	1 763.71	-0.49	1 774.19	0.10	
0.559 6	1 772.45	1 819.76	2.67	1 766.39	-0.34	1 771.31	-0.06	
0.632 8	1 772.45	1 804.41	1.80	1 776.55	0.23	1 772.94	0.03	
0.777 2	1 772.45	1 776.73	0.24	1 771.59	-0.05	1 772.90	0.03	
0.808 3	1 772.45	1 774.94	0.14	1 771.29	-0.07	1 772.01	-0.02	

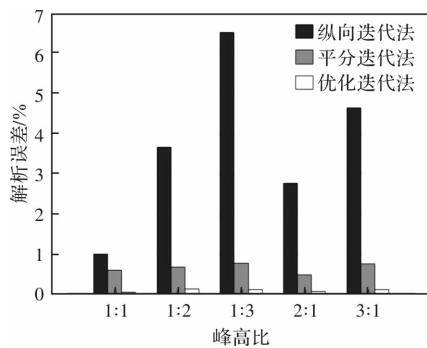


图 5 分离度为 0.621 8 时的解析结果

Fig. 5 Analysis results at resolution of 0.612 8

其次分析了优化迭代算法分离重叠峰的迭代效率。考虑到实际操作时会遇到三峰重叠现象,本次分析使用的重叠峰数学模型为三个高斯峰的叠加。其中:峰 A、B、C 的峰高比为 4:5:6;峰 A、峰 B 分离度为 0.621 8,峰 B、峰 C 分离度为

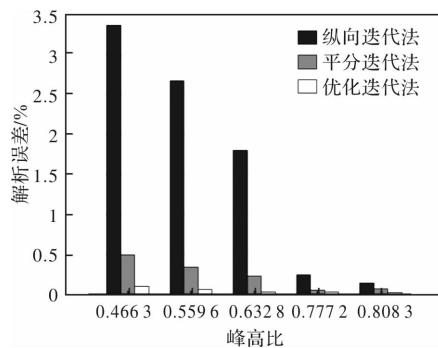


图 6 峰高比为 2:3 时的解析结果

Fig. 6 Analysis results at peak height ratio of 2:3

0.466 3,峰 A、峰 C 分离度为 1.088 1。分别使用优化迭代法、纵向迭代法、平分迭代法分离该重叠峰,根据所得的误差大小来比较三者的迭代效率。三峰重叠时的解析结果如表 3 所示。由表 3 可知,由于在第 1、2 次迭代时,三个方法得出构造峰的方

式相同,故有相同的迭代效率;在进行第3至第6次迭代时,使用优化迭代法时明显效率更高;迭代第6次时,迭代效率相对于纵向迭代法和平分迭代法分别提高了88.46%和66.67%。

表3 三峰重叠时的解析结果

Table 3 Analysis results at three peaks overlap

迭代次数	纵向迭代法	平分迭代法	优化迭代法
	误差 r/%	误差 r/%	误差 r/%
1	9.46	9.46	9.46
2	-4.33	-4.33	-4.33
3	2.72	1.92	0.92
4	1.97	1.16	-0.82
5	1.17	-0.42	0.19
6	-0.78	-0.27	-0.09

此外,还对含有随机噪声时分离重叠峰的迭代效率进行了分析。为方便实验,本次分析使用的重叠峰数学模型为峰高比为3:4、分离度为0.6218的两个高斯峰的叠加,解析结果如表4所示。由表4可知,迭代第6次时,迭代效率相对于纵向迭代法和平分迭代法分别提高了57.73%和32.20%。

表4 加随机噪声时的解析结果

Table 4 Analysis results with random noise

迭代次数	纵向迭代法	平分迭代法	优化迭代法
	误差 r/%	误差 r/%	误差 r/%
1	-11.43	-11.43	-11.43
2	6.99	6.99	6.99
3	-4.20	5.90	-4.20
4	-3.68	4.75	2.63
5	3.01	-3.24	-1.18
6	1.94	-1.21	0.82

表5 细胞荧光光谱解析结果

Table 5 Analysis results of cell fluorescence spectrum

迭代次数	实际值		纵向迭代法		平分迭代法		优化迭代法	
	面积 s/a.u.	面积 s/a.u.	误差 r/%	面积 s/a.u.	误差 r/%	面积 s/a.u.	误差 r/%	面积 s/a.u.
1	8 081.61	6 447.51	-20.22	6 447.51	-20.22	6 447.51	-20.22	
2	8 081.61	6 832.20	-15.46	6 832.20	-15.46	6 832.20	-15.46	
3	8 081.61	9 052.48	12.01	9 052.48	12.01	8 712.89	7.81	
4	8 081.61	7 349.02	-9.06	7 639.45	-5.47	7 824.17	-3.19	
5	8 081.61	7 533.98	-6.78	8 302.82	2.74	8 203.94	1.51	
6	8 081.61	7 773.98	-3.81	8 198.92	1.45	8 052.17	-0.36	

## 4 结论

针对荧光光谱峰重叠现象,本文结合并改进了传统纵向迭代法和平分迭代法,重新确定构峰的构造方式,提出优化迭代重叠峰分峰新方法。为评估该算法精度,通过模拟不同峰高比和

## 3 实验及结果分析

为进一步分析算法性能,使用纵向迭代法、平分迭代法和优化迭代法对实际光谱进行了分离及计算。将淋巴冻干粉与血清混合1 000 μL作为待测细胞,分别加入60 μL异硫氰酸荧光素(FITC)染剂、60 μL藻红蛋白(PE)染剂及30 μL藻红蛋白-得克萨斯红(ECD)染剂将其进行染色。将染色后的细胞充分拌匀后避光静置15 min,并将其放置在共聚焦培养皿上,用共聚焦荧光显微镜观察细胞染色情况,并对其测量到的细胞荧光光谱数据进行曲线拟合。拟合结果如图7所示。

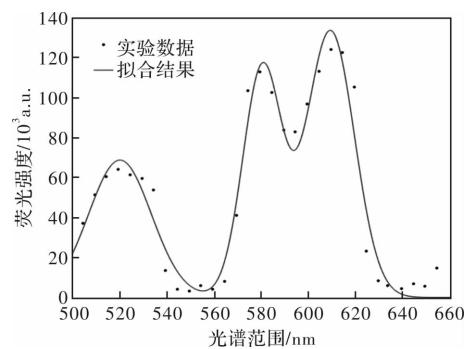


图7 细胞荧光光谱拟合图

Fig. 7 Fitting results of cell fluorescence spectrum

对数据进行高斯拟合后,分别用纵向迭代、平分迭代和优化迭代法进行解析的结果如表5所示。从表中可知,迭代6次后,优化迭代法的解析误差为0.36%,优于纵向迭代法和平分迭代法。并且,此时迭代效率相对于其他方法分别提高了90.55%和75.17%。故使用优化迭代法的分离结果更接近实际情况,进而进一步验证了该算法精度。

分离度的重叠峰,计算其面积的真实值与解析结果,并同纵向迭代法及平分迭代法进行了比较。结果表明优化迭代算法解析误差稳定在0.15%以内;加入随机噪声后,解析误差可稳定在0.85%以内,均优于另外两种算法。此外分析了该算法的迭代效率,其中三峰重叠情况下,迭代效率在原有

基础上提高了 66.67%;引入随机噪声时,迭代效率在原有基础上提高了 32.20%。对实际细胞荧光光谱中的重叠峰进行了解析,在相同迭代次数下,优化迭代法的解析误差为 0.36%,优于纵向迭代法和平分迭代法;且迭代效率分别提高了 90.55% 和 75.17%。此实验进一步验证了优化迭代算法的精度,故可使用该算法对重叠光谱准确进行定性和定量分析。

### 参考文献:

- [1] SHANG Xuemei, DONG Mingli, PAN Zhikang, et al. Identification of the parameters of spectral compensation for flow cytometer based on the recursive least squares algorithm[J]. Computer Simulation, 2015, 32(5): 252-255, 289.
- [2] KAUPPINEN J K, MOFFATT D J, MANTSCH H H, et al. Fourier self-deconvolution:a method for resolving intrinsically overlapped bands [J]. Applied Spectroscopy, 1981, 35(3): 271-276.
- [3] HUANG Hongquan, DING Weicheng, GONG Dishen, et al. Decomposition of X-ray fluorescence overlapping peaks based on statistical and genetic algorithms [J]. Spectroscopy and Spectral Analysis, 2015, 35(8): 2320-2323.
- [4] FENG Fei, WANG Fubei, XIE Fei, et al. Implementation of spectral overlap resolution based on wavelet transforms and gaussian fitting[J]. ACTA PHOTONICA SINICA, 2015, 44(6): 0630001-6.
- [5] TAO Weiliang, LIU Yan, WANG Xianpei, et al. Implementation of overlapping peak separation algorithm for absorption spectra by fractal dimension analysis in time-frequency domain [J], Spectroscopy and Spectral Analysis, 2017, 37(12): 3664-3669.
- [6] YAN Caijuan. Resolution of overlapping chromatographs by vertically iterative transition and modified vertically iterative transition[D]. Tianjin: Tianjin University, 2008.
- [7] YE Guoyang, XU Kejun. New decomposition peak methods of overlapping chromatographic peaks based on similarity principle[J]. Chinese Journal of Scientific Instrument, 2015, 36(2): 439-445.
- [8] ZHU Qiang, ZHANG Rong, YU Jiancheng. Novel algorithm for decomposition of overlapped chromatographic peaks based on function split iteration[J]. Chinese Journal of Vacuum Science and Technology, 2017, 37(5): 544-547.
- [9] SONG Lei, ZHENG Baozhong, ZHANG Ying, et al. Advanced EM algorithm based on Gaussian mixture model[J]. Journal of Applied Optics, 2013, 34(6): 985-989.
- [10] LIN Zhaopei, LI Yu, WU Huiwen. Analysis of overlapping chromatographic peaks based on quadratic differential and wavelet transform[J]. Journal of East China University of Science and Technology (Natural Science Edition), 2014, 40(1): 91-95.
- [11] LIU H, PANG Z, FAN G. Translation modification iteration for resolution and quantification of overlapping chromatographic peaks[J]. Chromatographia, 2016, 79: 1543-1552.
- [12] WANG Yue, MENG Xiaochen. Fluorescence lifetime representation method based on impulse delay estimation[J]. Journal of Applied Optics, 2018, 39(2): 207-211.
- [13] ZHU Zhongliang. Overlapping peaks resolve based on RBF neural network[D]. Hefei: Anhui Universi-

- ty, 2014.
- 朱仲良. 基于 RBF 神经网络的重叠峰解析[D]. 合肥: 安徽大学, 2014.
- [14] LI Lili. Research on GA overlapped peak analysis algorithm based on previous experience[D]. Hefei: Anhui University, 2018.
- 李莉莉. 基于前期经验的 GA 重叠峰解析算法研究 [D]. 合肥: 安徽大学, 2018.
- [15] CHEN Rui, ZHOU Lin. Measurement of combustion product composition by laser induced fluorescence spectroscopy[J]. Journal of Applied Optics, 2006, 27(5): 455-459.
- 陈锐, 周霖. 燃烧产物组成激光诱导荧光光谱的测量 [J]. 应用光学, 2006, 27(5): 455-459.