

基于改进迭代多项式拟合的红外光谱基线校正方法

宁志强^{1,2}, 刘家祥², 吴越², 陶孟琪², 方勇华^{1,2*}

¹中国科学技术大学环境科学与光电技术学院, 安徽 合肥 230026;

²中国科学院安徽光学精密机械研究所, 中国科学院环境光学与技术重点实验室, 安徽 合肥 230031

摘要 针对傅里叶变换红外光谱仪检测污染气体时红外光谱信号出现的基线漂移现象, 提出一种采用改进迭代多项式拟合的红外光谱基线校正方法, 该方法可以自动识别光谱信号的特征峰, 从光谱中截断完整的特征峰后, 采用双边阈值对红外光谱迭代得到最优的拟合基线。利用仿真光谱以及三氯乙烷、氨气的实测光谱对该方法进行验证, 并将其与常见的 3 种多项式拟合方法进行对比, 结果表明: 采用所提方法对光谱进行校正后, 特征峰更明显, 基线更平缓。仿真和实测数据表明, 所提方法提高了现有多项式拟合基线校正方法的准确性, 是一种具有良好性能的基线校正方法。

关键词 光谱学; 基线校正; 最小二乘法; 导数光谱

中图分类号 O433

文献标志码 A

doi: 10.3788/LOP57.033001

Infrared Spectrum Baseline Correction Method Based on Improved Iterative Polynomial Fitting

Ning Zhiqiang^{1,2}, Liu Jiexiang², Wu Yue², Tao Mengqi², Fang Yonghua^{1,2*}

¹*School of Environment Science and Optoelectronic Technology, University of Science and Technology of China, Hefei, Anhui 230026, China;*

²*Key Laboratory of Environmental Optics and Technology, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei, Anhui 230031, China*

Abstract The baseline of an infrared spectrum will drift when a Fourier transform infrared spectrometer is used to detect pollution gases. An infrared spectral baseline correction method based on improved iterative polynomial fitting is proposed in this paper. The method automatically identifies characteristic spectral peaks and subsequently cuts off the characteristic peaks from the spectrum. Then, it uses the bilateral threshold to iteratively fit the infrared spectrum to obtain the best-fitting baseline. The proposed method was verified using the simulated and measured spectra of trichloroethane and ammonia, and the results were compared with those obtained by the three common polynomial fitting methods. The results show that the characteristic peaks of the corrected spectra are more obvious and the baseline is more gradual after baseline correction using proposed method. The simulation and measured data show that our baseline correction method improves the accuracy of the existing polynomial fitting baseline correction method and exhibits good performance.

Key words spectroscopy; baseline correction; least square method; derivative spectrum

OCIS codes 300.6340; 070.4560

1 引言

傅里叶变换红外 (FTIR) 遥测技术是检测污染气体的常用方法之一^[1-4]。仪器响应、背景变化等因素使得 FTIR 光谱仪测得的光谱中除了含有实际信号的波形外, 还存在缓慢变化的趋势项, 这种现象称

为基线漂移, 它会对后续光谱的定量、定性分析造成极大影响。因此, 对光谱进行基线校正是红外光谱信号处理的关键步骤。

对光谱进行基线校正的方法有微分法^[5]、小波变换法^[6]、形态学方法^[7]和多项式拟合法 (PF)^[8]等。其中: 微分法可在一定程度上校正基线, 但同时

收稿日期: 2019-07-11; 修回日期: 2019-07-19; 录用日期: 2019-08-12

* E-mail: yhfang@aiofm.ac.cn

也会放大噪声,甚至会使光谱发生畸变;小波变换法可以有效去除基线漂移,但需要手动选取小波参数,而不同的小波参数会对基线拟合结果产生很大影响;形态学方法通过对光谱进行形态学操作来截断基线,因不同气体的红外光谱特征峰的宽度不同,因而选择合适的结构元非常困难;多项式拟合法利用最小二乘法对光谱进行基线拟合,其缺点是在拟合基线的同时会受到特征峰的影响,不能完整地拟合真实的基线。而迭代多项式拟合则是不断地对特征峰进行截断,逐次拟合出基线。传统的迭代多项式拟合(IPF)方法^[9]依次对多项式拟合后的结果与原光谱相同波数处的值进行比较,取二者中较小的值作为新的待拟合光谱,反复利用此方法对光谱信号进行拟合,直至前后两次拟合的基线基本重合,此时的拟合基线即为原光谱的最佳拟合基线。

Liu等^[10]将多项式拟合后的结果与原光谱进行比较,将光谱高于拟合基线一个标准差的点判定为特征峰,将特征峰截断后即可得到新的待拟合光谱;王昕等^[11]在文献[10]研究的基础上,提出了基线漂移模型(BDM)方法,该方法增强了对特征峰尾部的识别,在采用文献[10]所提方法找到特征峰后,在特征峰周围寻找大于拟合基线的数据点,并将其标记为特征峰,然后剔除所有被标记为特征峰的点,就可得到新的待拟合光谱。

以上迭代方法虽然在一定程度上消除了多项式拟合的缺点,但均未从缓变基线中剔除完整的特征峰,因此无法对光谱基线进行更精确的拟合。针对迭代多项式拟合中存在的不足,本文提出了一种改进迭代多项式的拟合方法,该方法可以有效去除整个特征峰,得到更精确的基线拟合结果。采用仿真光谱和实测光谱对该方法进行验证,并将该方法与常见的3种多项式拟合方法去除基线漂移的结果进行对比。

2 基本原理

2.1 多项式拟合

对 M 个点的光谱数据 S 进行 n 阶多项式拟合的表达式为

$$y(x, a_0, a_1, \dots, a_n) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n, \quad (1)$$

式中: a_i ($i=0, 1, 2, \dots, n$) 为多项式的系数; x 为光谱的波数; y 为拟合出的光谱值。

原光谱数据中各点到拟合数据各点的距离的平方和为

$$L^2 = \sum_{i=1}^M [S(i) - (a_0 + a_1x_i + a_2x_i^2 + \dots + a_nx_i^n)]^2, \quad (2)$$

式中: S 为原光谱数据; M 为光谱的总通道数。

为了使拟合出的光谱与原光谱的偏差最小,对(2)式右边分别求 a_i 的偏导数,令偏导数为 0,就可解出系数矩阵 a ,从而得到最小二乘拟合曲线 y 。

2.2 迭代多项式拟合

迭代多项式拟合是去除红外光谱基线的一种常用方法,该方法在多项式拟合的基础上,通过不断地截断特征峰,来逐次拟合逼近原始基线。迭代多项式拟合的步骤如下:

- 1) 将待拟合的 M 个点的光谱数据 S 表示为 y_0 ;
- 2) 利用最小二乘法对 y_0 进行 n 阶多项式拟合,得到拟合结果 y_n ;
- 3) 比较 y_0 和 y_n 中各点的值,利用各种方法截断 y_0 中的特征峰,对更新后的 y_0 进行拟合,得到新的 y_n ;
- 4) 将更新后的 y_n 与上一次的 y_n 进行比较,判断是否达到重合标准,若未达到,则转到步骤 3),如果基本重合,则将此时的 y_n 作为基线的最佳拟合结果;
- 5) 将 y_n 从待拟合光谱数据 S 中扣除即可得到基线校正后的光谱信号。

在迭代多项式拟合中,截断待拟合光谱 y_0 的特征峰,得到新 y_0 的方法有很多种。其中:IPF 方法选取光谱与拟合基线的最小值作为新光谱;文献[10]方法将大于基线一个阈值的光谱特征峰剔除后的光谱作为新光谱,阈值 P 的计算公式为

$$P = \left\{ \frac{\sum_{i=1}^M [y_0(i) - y_n(i)]^2}{M - n - 1} \right\}^{\frac{1}{2}}. \quad (3)$$

BDM 方法在文献[10]方法的基础上将大于基线的特征峰剔除后得到新光谱。这些方法均在一定程度上对特征峰进行了去除,但特征峰对拟合结果的影响导致真正的特征峰尾部在基线以下,因此应当考虑基线以下特征峰的影响,以得到更真实的迭代拟合基线。

2.3 改进迭代多项式拟合

针对常见的迭代多项式拟合不能完全消除特征峰的影响,进而导致基线不能被准确拟合的情况,提出了一种改进迭代多项式的拟合方法。该方法采用 4 阶多项式模型:

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4. \quad (4)$$

具体的拟合步骤如下：

1) 利用 Savitzky-Golay 滤波器^[12-14]对待拟合光谱信号 S 进行去噪处理得到 y_0 ，对 y_0 进行多项式拟合，得到基线 y_n ，求 y_0 的一阶导数 y_{ds} ，并对 y_{ds} 进行去噪处理；

2) 比较 y_0 和 y_n 的大小，若 $y_0 - y_n > P$ (P 为阈值)，则将此处的 y_0 标记为特征峰，将特征峰附近的 y_0 与 y_n 的左右交点分别记为 L_{left} 、 L_{right} ；

3) 令 L_{left} 和 $L_{left} - 1$ 处光谱的一阶导数的乘积为 D_l ，若 $D_l > 0$ ，令 $L_{left} - 1$ 为新的 L_{left} ，重新执行步骤 3)，直到 $D_l < 0$ ，说明此时找到了完整特征峰的左边界；

4) 令 L_{right} 和 $L_{right} + 1$ 处光谱的一阶导数的乘积为 D_r ，若 $D_r > 0$ ，令 $L_{right} + 1$ 为新的 L_{right} ，重新执行步骤 4) 直到 $D_r < 0$ ，说明此时找到了完整特征峰

的右边界；

5) 将 L_{left} 和 L_{right} 间的 y_0 截断，其他处的 y_0 值保持不变，构造新的 y_0 ；

6) 对新的 y_0 进行多项式拟合，得到新的基线 y_n ，比较前后两次基线是否相同，若相同，则将 y_0 视为原信号 S 的最优拟合基线，基线校正后的信号 $S' = S - y_0$ ，程序结束，若不相同，则转到步骤 7)；

7) 将 y_0 和 y_n 取差后求绝对值，将绝对值大于阈值 P 的 y_0 截断，得到新的 y_0 ，转到步骤 6)。

判断前后两次基线是否相同的条件式为

$$\sum_{i=1}^M \left| \frac{y'_n(i) - y_n(i)}{y_n(i)} \right| < \epsilon, \quad (5)$$

式中： y'_n 为最新拟合的基线； y_n 为上一次拟合的基线； ϵ 取 0.01。

改进迭代多项式拟合的流程图如图 1 所示。

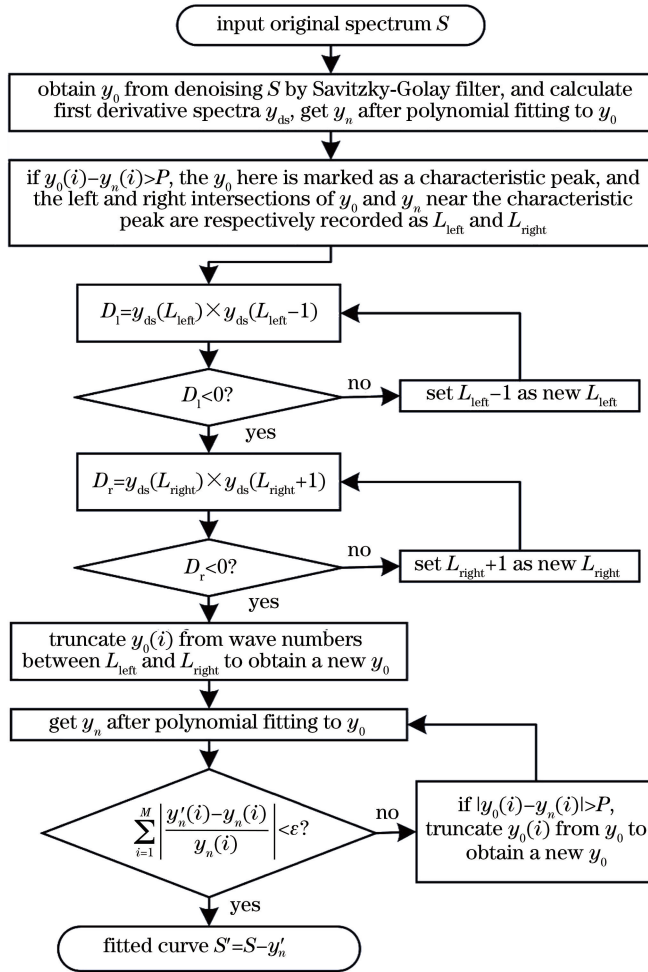


图 1 改进迭代多项式拟合方法的流程图

Fig. 1 Flow chart of improved iterative polynomial fitting method

3 实验结果分析

3.1 仿真光谱数据的实验结果分析

仿真光谱由 3 部分组成:特征峰、低频基线和随机噪声。特征峰和基线由高斯函数产生,公式为

$$f(x) = a \exp\left[-\frac{(x-b)^2}{2c^2}\right], \quad (6)$$

式中: a 为高斯函数的峰高; b 为峰尖位置的横坐标; c 为标准差,表征峰的宽度。

仿真光谱的具体参数如表 1 所示,其中,序号 1、2 表示特征峰,序号 3 表示低频基线。信号中加入了标准差约为最低特征峰高 3% 的随机噪声。

表 1 特征峰和基线的仿真参数

Table 1 Simulation parameters of characteristic peak and baseline

Type	a	b	c
1	1	35	5
2	2	100	10
3	1	140	50

仿真光谱和所提方法首次截断波峰后的光谱如图 2 所示。

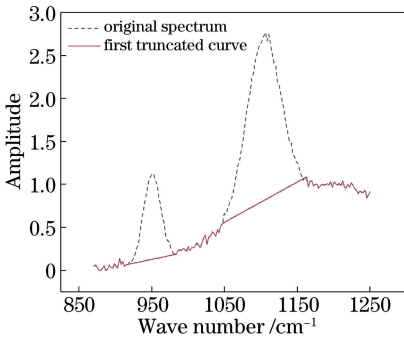


图 2 仿真光谱以及采用所提方法首次截断波峰后的光谱
Fig. 2 Simulation spectrum and the spectrum obtained with the proposed method for the first truncation of the peak

仿真光谱以及采用 PF 方法、IPF 方法、BDM 方法、所提方法得到的拟合基线如图 3 所示,去除基线后的结果如图 4 所示。

可以看出:PF 方法没有完全拟合出基线,基线校正后的光谱仍存在缓变的基线;IPF 方法对第 1 个峰的基线拟合不足,对第 2 个峰的基线拟合过高,导致基线去除后的信号的第 1 个峰偏高,第 2 个峰偏低;BDM 拟合存在同样的问题,但在 1150 ~ 1250 cm^{-1} 范围内的拟合效果略好于传统方法;所提方法的拟合效果最好,不仅得到了准确的特征峰,而且没有造成其他波段的光谱变形。

方均根误差(RMSE, α_{RMSE})是衡量估计值与实际

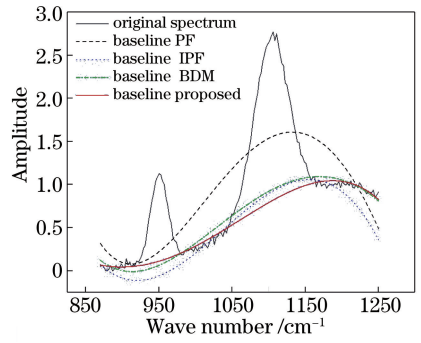


图 3 仿真光谱以及采用不同方法得到的拟合基线
Fig. 3 Simulated spectrum and fitted baseline obtained with different methods

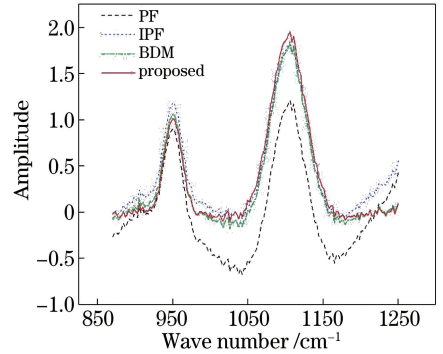


图 4 采用不同方法扣除基线后的光谱

Fig. 4 Spectra after deducting the baseline

值之间偏差的重要指标^[15-17],考虑到仿真光谱的基线是已知的,因此可采用 RMSE 作为评价标准对上述 4 种方法的性能进行评估。RMSE 的计算公式为

$$\alpha_{\text{RMSE}} = \sqrt{\frac{1}{M'} \sum_{i=1}^{M'} [y(i) - y(i)^*]^2}, \quad (7)$$

式中: $y(i)$ 为真实值; $y(i)^*$ 为估计值; M' 为样本个数。

上述 4 种拟合方法扣除基线后的光谱与不含基线的原光谱之间的 RMSE 如表 2 所示。

表 2 扣除基线后的光谱与不含基线的原光谱之间的 RMSE
Table 2 RMSEs between the baseline-corrected spectrum and the original spectrum without baseline

Method	RMSE
PF	0.5081
IPF	0.1694
BDM	0.1258
Proposed	0.0539

由表 2 可以看出:PF 方法得到的基线校正光谱的 RMSE 最大;IPF 方法和 BDM 方法的 RMSE 较小;所提方法的 RMSE 最小,分别比 PF 方法、IPF 方法和 BDM 方法得到的 RMSE 减小约 90%、70% 和 60%。

多项式拟合时阶次的选择对基线校正结果的影响较大,因此分别选择不同的最高阶次的多项式模型对上述4种方法进行比较。不同阶次下各种方法的RMSE如图5所示。

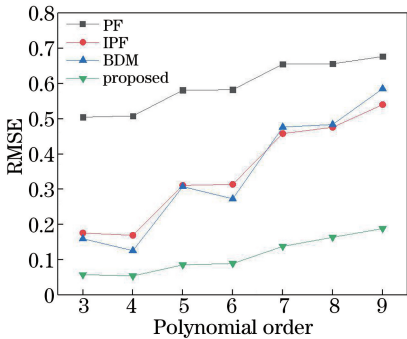


图5 在不同多项式阶次下,不同方法的RMSE
Fig. 5 RMSEs of different methods at different polynomial orders

从图5中可以看出:PF方法的RMSE最大,IPF和BDM方法的RMSE比PF方法低;所提方法在不同阶次的多项式拟合中均能得到较小的RMSE,其在9阶拟合时的RMSE最大,但仍然比其他3种方法低。分别计算得到了PF、IPF、BDM和所提方法在不同多项式阶次下的RMSE的均值,结果如表3所示。

表3 不同方法在不同多项式阶次下的RMSE的均值
Table 3 Mean RMSE of different methods at different polynomial orders

Method	Mean RMSE
PF	0.5951
IPF	0.3492
BDM	0.3446
Proposed	0.1108

由表3可知,所提方法的RMSE的均值比IP、IPF和BDM方法分别减小了约80%、70%和65%,表明该方法具有较高的稳定性。

3.2 实测光谱数据的实验结果分析

利用三氯乙烷和氨气的实测光谱对所提出的方法进行验证。为了便于数据处理,预先对光谱进行归一化处理。三氯乙烷的原光谱以及采用不同方法拟合得到的基线如图6所示,基线校正后的光谱如图7所示,氨气的原光谱和拟合得到的基线如图8所示,基线校正后的光谱如图9所示。

由图6~9可以看出:PF方法对三氯乙烷和氨气光谱特征峰附近的基线拟合得过高,基线校正后的光谱的峰被削低;IPF方法对三氯乙烷特征峰附近的基线拟合得过高,对非特征峰区域的基线拟合

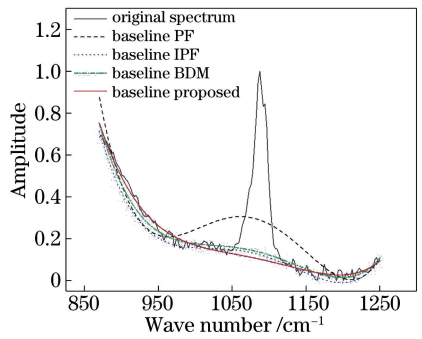


图6 三氯乙烷的原光谱和采用不同方法拟合得到的基线
Fig. 6 Original spectrum of trichloroethane and its fitted baselines obtained with different methods

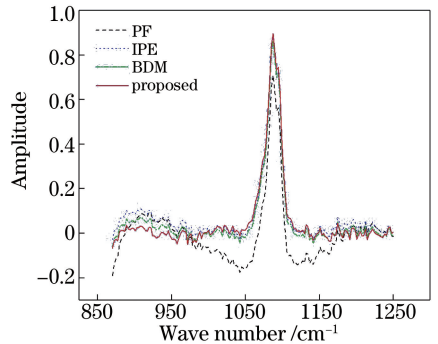


图7 采用不同方法对三氯乙烷基线进行校正后的光谱
Fig. 7 Spectra of trichloroethane after baseline correction with different methods

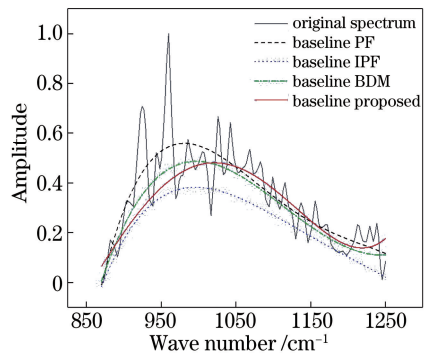


图8 氨气的原光谱以及采用不同方法拟合得到的基线
Fig. 8 Original spectrum of ammonia and its fitted baselines obtained with different methods

得过低,在削弱特征峰的同时保留了部分基线,对氨气非特征峰区域的基线拟合得过低,拟合后的信号仍存在基线成分;BDM方法对基线的拟合效果较好,但在三氯乙烷和氨气光谱特征峰附近的基线略高,扣除基线后的特征峰略有下降;所提方法对基线拟合的效果最好,对三氯乙烷和氨气光谱特征峰附近的基线实现了准确拟合,基线校正后的特征峰最明显,同时,对两种气体非特征峰区域的基线没有过高或过低拟合,校正后的基线最平缓。

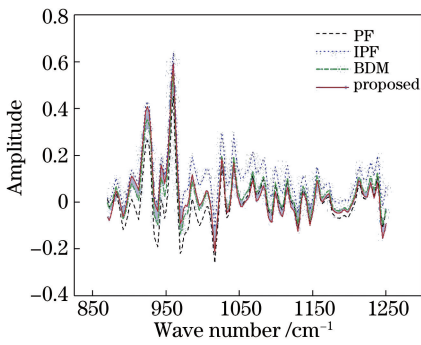


图9 采用不同方法对氨气基线进行校正后的光谱

Fig. 9 Spectra of ammonia after baseline correction with different methods

为了比较不同方法对实测光谱信号拟合的效果,删除校正后光谱特征峰区域的数据,然后求取信号方差。方差越小,说明校正后信号的基线越平缓,拟合效果越好。三氯乙烷的主要特征峰在 $1050 \sim 1130 \text{ cm}^{-1}$ 范围内,氨气的主要特征峰在 $910 \sim 970 \text{ cm}^{-1}$ 范围内,去除特征峰后信号的方差如表4所示。

表4 去除特征峰后信号的方差

Table 4 Variance of signal after removing characteristic peak

Method	Variance / 10^{-4}	
	Trichloroethane	Ammonia
PF	48.2	40
IPF	11.2	48
BDM	7.7	40
Proposed	4.5	39

由表4可以看出,与其他3种方法相比,所提方法在三氯乙烷和氨气实测数据中得到的方差最小,说明该方法得到的基线最平缓,波动最小,拟合效果最好。

综上所述,所提方法可以有效去除红外光谱的缓变基线,为后续定量分析带来了极大方便。

4 结 论

本文提出了一种采用改进迭代多项式拟合的红外光谱基线校正方法,并利用仿真光谱和实测光谱验证了该方法的准确性。结果表明,对于仿真光谱和实测光谱,所提方法都可以得到较好的基线校正结果。在仿真实验中,与PF、IPF、BDM等方法相比,所提方法的RMSE均值分别低了80%、70%、65%左右。在实测数据实验中,采用不同的方法对三氯乙烷和氨气基线漂移的红外光谱信号进行校正,结果表明,与其他3种方法相比,所提方法校正

后的光谱的特征峰最明显,非特征峰区域的方差最小,说明该方法得到的基线最平缓,拟合效果最好。综上所述,所提方法在气体红外光谱基线校正中取得了良好的效果,对红外光谱气体的检测具有重要意义。

参 考 文 献

- [1] Li Z Y, Tan R Q, Huang W, et al. Methane pressure detection based on Fourier transform infrared spectroscopy[J]. Chinese Journal of Lasers, 2017, 44(3): 0301006.
李志永, 谭荣清, 黄伟, 等. 傅里叶变换红外光谱技术测量甲烷气压的实验研究[J]. 中国激光, 2017, 44(3): 0301006.
- [2] Flores-Jardines E, Schäfer K, Black J, et al. Remote sensing of aircraft exhaust temperature and composition by passive Fourier transform infrared (FTIR) [J]. Proceedings of SPIE, 2007, 6745: 67451Q.
- [3] Wadsworth W, Williams D J. Mobile remote sensing FT-IR for plume detection[J]. Proceedings of SPIE, 2006, 6302: 630201.
- [4] Zheng W. Studying of air quality monitoring system based on FTIR [D]. Tianjin: Tianjin University, 2006: 6-9.
郑伟. 基于FTIR的环境气体监测系统研究[D]. 天津: 天津大学, 2006: 6-9.
- [5] Leger M N, Ryder A G. Comparison of derivative preprocessing and automated polynomial baseline correction method for classification and quantification of narcotics in solid mixtures [J]. Applied Spectroscopy, 2006, 60(2): 182-193.
- [6] Fang Y H, Kong C, Lan T G, et al. Denoising and baseline correction of spectrum by wavelet transform [J]. Optics and Precision Engineering, 2006, 14(6): 1088-1092.
方勇华, 孔超, 兰天鸽, 等. 应用小波变换实现光谱的噪声去除和基线校正[J]. 光学精密工程, 2006, 14(6): 1088-1092.
- [7] Chen Y L, Dai L K. An automated baseline correction method based on iterative morphological operations[J]. Applied Spectroscopy, 2018, 72(5): 731-739.
- [8] Feng X W, Zhu Z L, Shen M J, et al. The method of baseline drift correction of Raman spectrum based on polynomial fitting [J]. Computers & Applied Chemistry, 2009, 26(6): 759-762.
- [9] Lieber C A, Mahadevan-Jansen A. Automated

- method for subtraction of fluorescence from biological Raman spectra[J]. *Applied Spectroscopy*, 2003, 57(11): 1363-1367.
- [10] Liu J, Koenig J L. A new baseline correction algorithm using objective criteria [J]. *Applied Spectroscopy*, 1987, 41(3): 447-449.
- [11] Wang X, Lü S L, Li Y, et al. Automatic baseline correction of gas spectra based on baseline drift model [J]. *Spectroscopy and Spectral Analysis*, 2018, 38(12): 3946-3951.
王昕, 吕世龙, 李岩, 等. 基于基线漂移模型的气体光谱自动基线校正[J]. *光谱学与光谱分析*, 2018, 38(12): 3946-3951.
- [12] Lei L P. Curve smooth denoising based on Savitzky-Golay algorithm [J]. *Computer and Information Technology*, 2014, 22(5): 30-31.
雷林平. 基于 Savitzky-Golay 算法的曲线平滑去噪[J]. *电脑与信息技术*, 2014, 22(5): 30-31.
- [13] Fang X Q, Peng Y K, Li Y Y, et al. Rapid and quantitative detection method of sodium benzoate in carbonated beverage based on surface-enhanced Raman spectroscopy[J]. *Acta Optica Sinica*, 2017, 37(9): 0930001.
房晓倩, 彭彦昆, 李永玉, 等. 基于表面增强拉曼光谱快速定量检测碳酸饮料中苯甲酸钠的方法[J]. *光学学报*, 2017, 37(9): 0930001.
- [14] Cai T J, Tang H. Review of least square fitting principle of Savitzky-Golay smooth filter[J]. *Digital Communication*, 2011, 38(1): 63-68, 82.
蔡天净, 唐瀚. Savitzky-Golay 平滑滤波器的最小二乘拟合原理综述[J]. *数字通信*, 2011, 38(1): 63-68, 82.
- [15] Chai T, Draxler R R. Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature [J]. *Geoscientific Model Development*, 2014, 7(3): 1247-1250.
- [16] Yang Y R, Dai Y. Evaluation of the effect on eye aberration on retinal imaging quality based on the root mean square error and correlation coefficient [J]. *Acta Optica Sinica*, 2017, 37(3): 0333001.
杨彦荣, 戴云. 基于均方根误差和相关系数评价人眼像差对视网膜像质的影响[J]. *光学学报*, 2017, 37(3): 0333001.
- [17] Lin P, Chen Y M, He Y. Feasibility of using Fourier transform infrared spectroscopy combined with chemometrics for determining the total flavonoid contents in rape bee pollen[C] // 2010 International Conference on Artificial Intelligence and Computational Intelligence, October 23-24, 2010, Sanya, China. New York: IEEE, 2010: 350-354.