

引用格式: YAO Cheng-bin, JIA Yun-wei, WU Jiang-bo, *et al.* Baseline Fitting Algorithm of Spectral Signal Region[J]. *Acta Photonica Sinica*, 2020, 49(12):1230003

姚城斌,贾云伟,吴江波,等. 光谱信号区域基线拟合算法[J]. 光子学报, 2020, 49(12):1230003

光谱信号区域基线拟合算法

姚城斌, 贾云伟, 吴江波, 王坤, 郝晨翔

(天津理工大学 天津市先进机电系统设计与智能控制重点实验室, 天津 300384)

摘 要:为消除基线畸变对光谱信号后续处理产生的不利影响,提出了一种光谱信号区域基线拟合算法.首先,利用渐变惯性和吸引力的思想,针对具有曲率突变的基线提出了GradSuck拟合算法,该算法弥补了多项式拟合算法在信号区域周围基线曲率发生突变时拟合基线不准确的不足;然后将这种算法与分段二次多项式拟合算法进行结合,提出了更具有普适性的信号区域基线拟合算法,同时将该算法与不同的基线拟合算法做了对比.实验结果表明,信号区域基线拟合算法在不同的基线类型和不同的信噪比下均具有较高的准确性和稳定性,其总体拟合精度的相对误差仅为分段二次多项式拟合的47.0%,为AirPLS拟合的35.6%,为Wavelet拟合的20%,且由于只对光谱信号区域进行基线拟合,因此具有较好的实时性.

关键词:信号区域;基线拟合;基线畸变;二次多项式拟合;分段拟合;曲率突变

中图分类号:O657.33

文献标识码:A

doi:10.3788/gzxb20204912.1230003

Baseline Fitting Algorithm of Spectral Signal Region

YAO Cheng-bin, JIA Yun-wei, WU Jiang-bo, WANG Kun, HAO Chen-xiang

(Key Laboratory of Advanced Mechatronics System Design and Intelligent Control of Tianjin, Tianjin University of Technology, Tianjin 300384, China)

Abstract: A baseline fitting algorithm for spectral signal region is proposed in order to eliminate the adverse effects of baseline distortion in spectral processing. First, a GradSuck fitting algorithm is proposed for baseline with abrupt curvature by using the idea of gradient inertia force and suction, which cover the shortage of the piecewise quadratic polynomial fitting algorithm when the baseline curvature around the signal area changes suddenly. Then this algorithm is combined with the piecewise quadratic polynomial fitting algorithm to propose a more general fitting algorithm for spectral signal region. At the same time, the proposed algorithm has been compared with different baselines fitting algorithms. Experiments under various baseline types with different SNR show that the baseline fitting algorithm of signal region has reliable accuracy and stability, and it can extract the spectral baseline better than other algorithms. The relative error of its whole fitting accuracy is only 47.0% of the quadratic polynomial fitting, 35.6% of the AirPLS fitting, and 20% of the wavelet fitting. And it also has high real-time performance because of only fitting the baseline of the signal region.

Key words: Signal region; Baseline fitting; Baseline distortion; Quadratic polynomial; Piecewise fitting; Abrupt curvature

OCIS Codes: 300.6170; 300.6450; 070.4790

基金项目:国家自然科学基金(No.61201081),天津市自然科学基金(No.18JCYBJC19300)

第一作者:姚城斌(1995—),男,硕士研究生,主要研究方向为机器人导航. Email: 1261396678@qq.com

导 师:贾云伟(1979—),女,副教授,博士,主要研究方向为传感与数据处理、机器人导航. Email: yunweijia@163.com

收稿日期:2020-07-09;录用日期:2020-09-24

<http://www.photon.ac.cn>

0 引言

光谱可以用来从样品中提取信息,例如物质的化学结构、物理结构和溶液的浓度等信息.然而,这些信息的提取经常受基线畸变的影响,从而导致信号区域和信号强度检测不准(信号区域和强度分别代表样品的结构信息和浓度信息)^[1-2].为此,在信号提取时,常常需要对基线进行拟合并加以剔除,以消除基线畸变的影响.常用的基线拟合算法有线性插值^[3-4]、多项式拟合^[5-6]、迭代平滑^[7]、小波分解^[8-10]、惠塔克平滑算法(Whittaker Smoother algorithm, WS)^[11-12]及其改进算法^[13]、最小二乘法^[14]及其改进算法等^[15].由于基线干扰因素复杂、多变,很难用模型对其进行整体描述,所以整体提取的基线精度很难得到保证,即使采用没有基线模型的迭代算法,如自适应迭代加权惩罚最小二乘(Adaptive iterative re-weighted Penalized Least Squares, AirPLS),由于其平滑性参数和忠于原始数据参数之间的相互制约,其基线的提取精度也并不理想,从而导致信号后续的定性和定量检测受到影响.再如小波分解,其精度常常受小波基和分解次数的影响,而且小波基和分解次数通常根据个人经验来进行选取.基线拟合的目的是为了准确地提取信号,因此只要准确提取信号附近较短的一段基线即可.如果只在信号区域进行基线拟合,一方面由于拟合线段变短,利于提高拟合基线的精度;另一方面,不需要对所有数据进行处理,可以减少计算量、提高实时性.当拟合线段较短时,一般仅用二次多项式拟合即可取得较好的效果.然而当信号区域基线曲率变化较大或者出现阶跃时,即使是仅对信号区域基线进行二次多项式拟合,也难以取得较好的结果,因此本文首先针对信号区域基线曲率变化较大或者出现阶跃时的情况提出了(Gradual Suction)拟合算法,然后将该算法与分段二次多项式拟合进行结合,最终提出了一种具有普适性的信号区域基线拟合算法(Quadratic polynomial fuse Gradual Suction, QGS).

1 拟合算法

1.1 针对基线曲率突变的 GradSuck 拟合算法

GradSuck 为基于渐变(gradual)惯性力和吸引力(suction)的拟合方法.其中惯性力用于保证拟合基线的光滑性,吸引力则保证拟合基线的一致性,根据信号区域左右两侧的拟合数据与原基线的平均差,采用不同的方法进行 GradSuck 拟合得到最终的基线 B_{GS} ,其步骤如下:

- 1)用二次多项式分别拟合信号区域左侧(X_L, Y_L)和右侧(X_R, Y_R)数据,得到两条拟合基线 B_L 和 B_R .
- 2)计算左右两侧的惯性力 I_L, I_R 及惯性系数 k_1 ,表达式为

$$\begin{cases} I_L(i) = B_L(x_L + i) - B_L(x_L + i - 1) \\ I_R(i) = B_R(x_w - i + 1) - B_R(x_w - i) \\ k_1(i) = 1 - k_s(i) \end{cases} \quad (1)$$

式中, x_L, x_w 为信号区域左右端点的索引值, i 为光谱数据点索引值, S 和 k_s 分别表示引力和引力系数.

- 3)由式(2)和(3)分别得到拟合基线左右端点的初始值.

$$B_{GS}(x_L) = B_L(x_L - 1) + I_L(0)k_1(0) + S(0)k_s(0) \quad (2)$$

$$B_{GS}(x_w) = B_R(x_w + 1) - I_R(0)k_1(0) - S(0)k_s(0) \quad (3)$$

$S(0)$ 和 $k_s(0)$ 的计算如式(4)~(5)所示, R 为信号区域半宽, W 为信号宽度.

$$S(0) = \frac{B_R(x_w + 1) - B_L(x_L - 1)}{W + 1} \quad (4)$$

$$k_s(0) = \begin{cases} \frac{1}{W} & \text{One - side GradSuck fitting} \\ \frac{1}{R} & \text{Two - sides GradSuck fitting} \end{cases} \quad (5)$$

4)得到其余的拟合数据.如果两侧拟合基线误差的均方值类似,则采用 B_L, B_R 进行双侧 GradSuck 拟合,如式(6)所示;如果右侧的均方差大于左侧均方差的2倍,则只采用 B_L 进行单侧 GradSuck 拟合,如式(7)所示;如果左侧的均方差大于右侧均方差的2倍,则只采用 B_R 进行单侧 GradSuck 拟合,如式(8)所示.

$$\begin{cases} B_{GS}(x_L + i) = B_{GS}(X_L + i - 1) + I_L(i)k_1(i) + S(i)k_s(i) \\ B_{GS}(x_w - i) = B_{GS}(X_w - i + 1) - I_R(i)k_1(i) - S(i)k_s(i) \\ S(i) = \frac{B_{GS}(x_w - i + 1) - B_{GS}(x_L + i - 1)}{W + 1 - 2i} \end{cases} \quad i \in [1, (R - 1)] \quad (6)$$

$$\begin{cases} B_{GS}(x_L + i) = B_{GS}(X_L + i - 1) + I_L(i)k_1(i) + S(i)k_s(i) \\ S(i) = \frac{B_{GS}(x_w + 1) - B_{GS}(x_L + i - 1)}{W + 1 - i} \\ k_s(i) = \frac{i + 1}{W} \end{cases} \quad i \in [1, (W - 1)] \quad (7)$$

$$\begin{cases} B_{GS}(x_w - i) = B_{GS}(X_w - i + 1) - I_R(i)k_1(i) - S(i)k_s(i) \\ S(i) = \frac{B_{GS}(x_w - i + 1) - B_{GS}(x_L - 1)}{W + 1 - i} \\ k_s(i) = \frac{i + 1}{W} \end{cases} \quad i \in [1, (W - 1)] \quad (8)$$

1.2 针对任意基线的 QGS 拟合算法

GradSuck 拟合算法主要针对信号区域基线曲率突变或者出现阶跃时的情况,若信号区域基线曲率变化较为平缓,则其效果与分段二次多项式拟合相比并不具有优势.为取得最佳效果,本文针对任意情况下的基线拟合,将分段二次多项式拟合方法与 GradSuck 拟合方法结合,提出了 QGS 拟合算法.首先用二次多项式拟合信号区域左右两侧 $(X_L \cup X_R, Y_L \cup Y_R)$ 的数据,得到拟合基线 B_Q ,然后将 B_Q 与 B_L 、 B_R 进行比较,得到信号区域合适的拟合基线,如式(9)所示,分别计算 B_L 、 B_R 与 B_Q 在相应基线点处的差值,如果存在差值大于基线部分噪声均方差值的点,并且信号区域左右两侧基线数据点个数大于信号处数据点个数,则选择 GradSuck 拟合方法,反之选择二次多项式拟合,得到信号区域 (X_s, Y_s) 最终的拟合基线 B .拟合示意图如图 1.

$$B = \begin{cases} B_{GS} & \left\{ \begin{array}{l} \exists |B_Q(i_L) - B_L(i_L)| > N \text{ or } \exists |B_Q(i_R) - B_R(i_R)| > N \\ N_L > N_s \text{ and } N_R > N_s \\ B_{QS} \text{ Others} \end{array} \right. \end{cases} \quad (9)$$

式中, $i_L \in X_L, i_R \in X_R, N$ 为基线部分的噪声均方差, N_L, N_R 为信号区域左右侧基线数据点个数, N_s 为信号处数据点个数, B_{QS} 为 B_Q 在信号区域的取值.

2 实验及分析

为了分析本文所提算法在不同曲率下的性能,分别在 5 种典型基线: Gauss(高斯函数)、Exponential(指数函数)、sin(正弦函数)、Triangle(三角函数)、Step(阶跃函数)的不同位置上添加两个信号产生仿真光谱.首先,在有噪声情况下将 GradSuck 拟合与二次多项式拟合结果进行比较,对 GradSuck 算法的有效性进行验证.其次,分别按无噪声、有噪声两种情况分析 QGS 算法的性能,并将 QGS 算法与不同算法在不同基线畸变类型、不同曲率下的基线拟合效果进行对比,分析 QGS 算法的准确性和稳定性;然后变换不同的信噪比,研究 QGS 算法与其他拟合算法的拟合结果,分析信噪比对于 QGS 算法的影响.最后,使用实测光谱数据验证 QGS 算法的有效性.

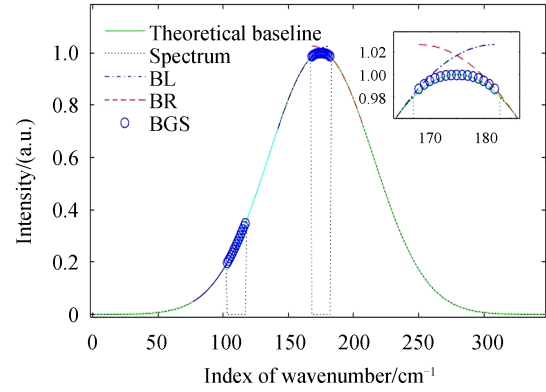
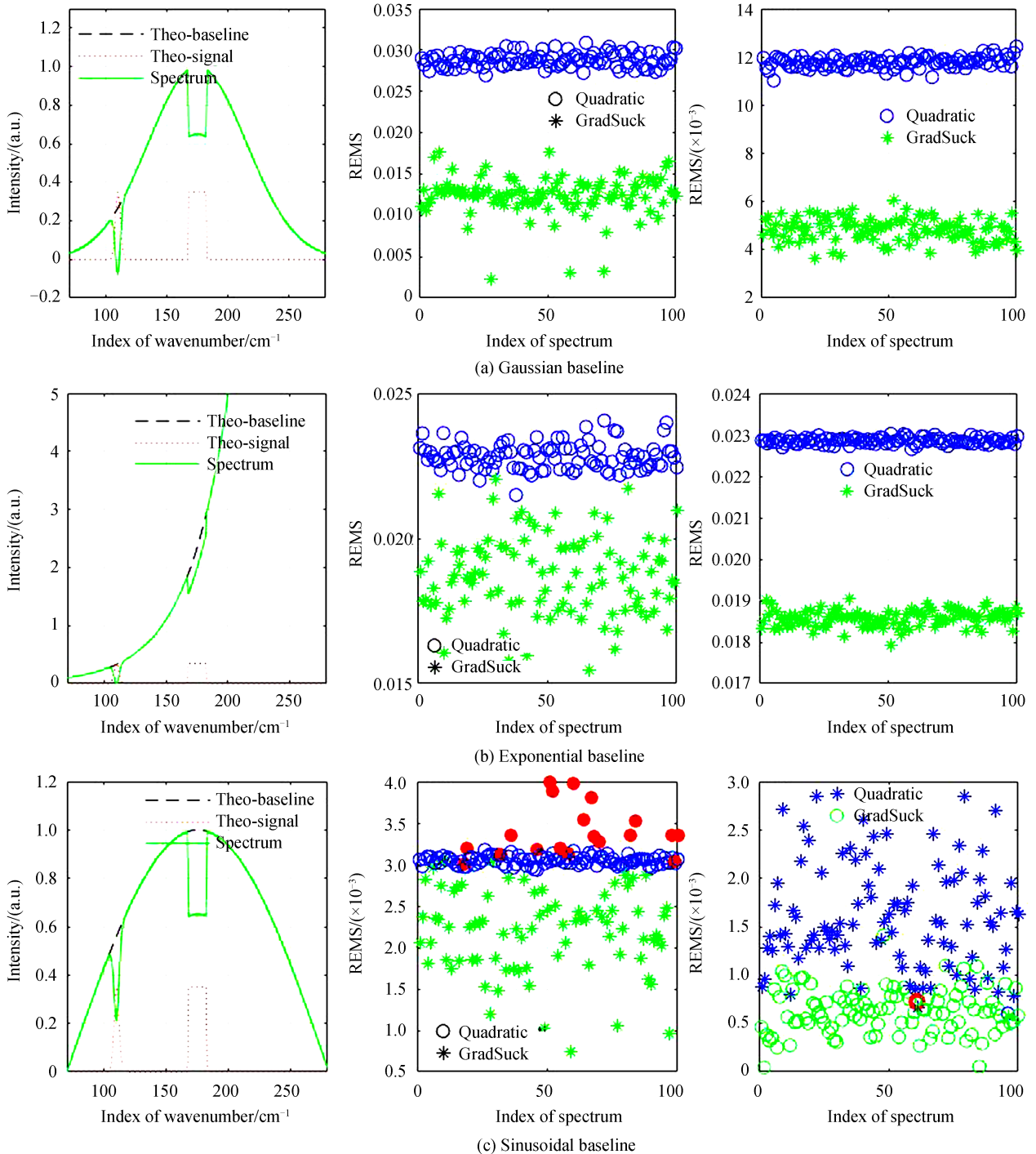


图1 信号区域拟合示意图

Fig.1 Schematic diagram of signal area fitting

2.1 GradSuck算法有效性分析

图2为SNR=50时,对不同的基线类型分别进行100次染噪及基线拟合的结果,其中图2(a)~(e)的中间子图为信号1所对应区域基线拟合结果的均方根误差统计,右侧子图为信号2所对应区域基线拟合结果的均方根误差统计;圆圈代表二次多项式拟合结果的均方根误差;*代表GradSuck拟合结果的均方根误差;由图2(a)、(b)、(c)的中间子图和(d)、(e)的右侧子图可知,若信号区域附近基线曲率突变或者出现阶跃时,Gradsuck拟合效果较分段二次多项式拟合效果好.若信号区域附近基线符合二次多项式或者基线曲率变化不大,如图2(c)的右侧子图和(d)、(e)的中间子图所示,则分段二次多项式拟合误差较小.由此可见在信号区域附近基线曲率突变或者出现阶跃时Gradsuck拟合算法确实有效.



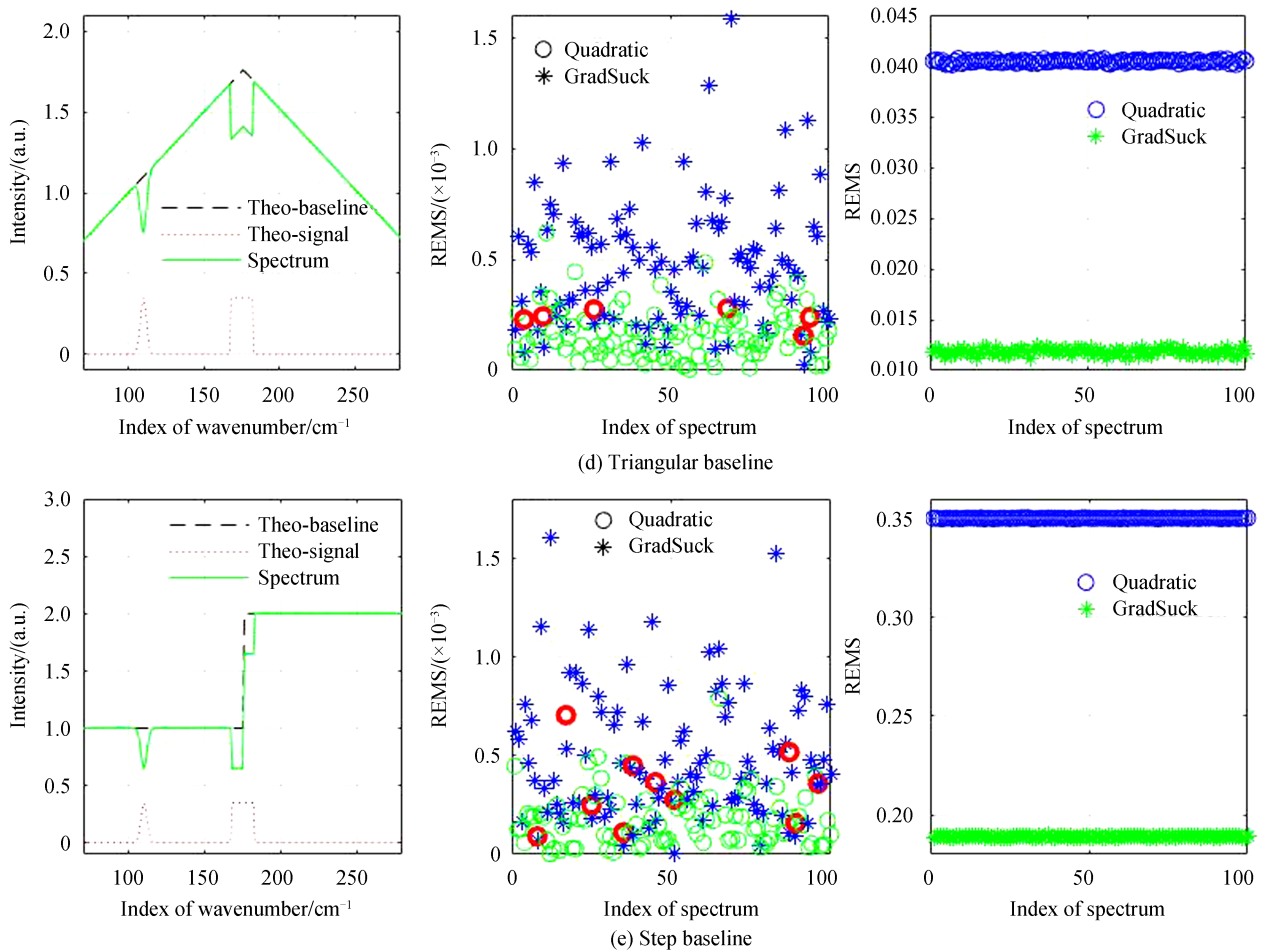


图2 SNR=50时 GradSuck 拟合和二次多项式拟合基线比较

Fig.2 Baseline-comparison of GradSuck fitting and quadratic polynomial fitting when SNR=50

2.2 QGS算法有效性分析

图2也可以体现QGS算法的拟合效果,将图中的误差标记为绿色和红色,其中绿色代表QGS正确选中拟合方式,即选中的拟合方式比另一种拟合方式误差小;红色代表QGS错误选中拟合方式,即选中的拟合方式比另一种拟合方式误差大;当QGS错误选中拟合方式,错误点的线宽加粗.可见,QGS错误选中拟合方式的概率非常低,因此其可以取长补短,获得更好的拟合效果.为进一步体现QGS算法的性能,在不同基线类型、不同信噪比下将QGS算法的拟合结果与经典的分段二次多项式拟合算法、AirPLS算法、Wavelet算法的拟合结果进行对比分析.

2.2.1 无噪声时QGS算法与其它算法的比较

从图3(a)中可以直观地看出在无噪声、基线存在阶跃时QGS算法拟合的基线和理论基线都很接近.二次多项式拟合和AirPLS拟合算法拟合效果在信号区域基线有阶跃时不如QGS算法,而Wavelet算法拟合效果较差.为了更进一步对不同算法的拟合结果进行误差分析,将不同算法在5种基线畸变类型下拟合的实验结果进行了误差统计,得到图3(b)、(c).在图3(b)中,每种拟合算法拟合结果都有10个从小到大排列的点,这10个点为同一种拟合算法分别在5种基线畸变类型下对两个信号区域进行基线拟合所得结果的相对误差的均值,该值的大小表征了不同算法拟合结果的准确度,其值越接近0,表明该算法的准确度越高.图3(c)与图3(b)类似,不同的是每个点的含义为相对误差的均方根值,该值的大小表征了不同算法拟合基线结果的离散度,其值越小,算法的离散度越低.可以看出QGS算法整体的准确度高、离散度低,较分段二次多项式拟合算法、AirPLS拟合算法及Wavelet拟合算法具有优势.所以,QGS拟合算法在无噪声时有效.

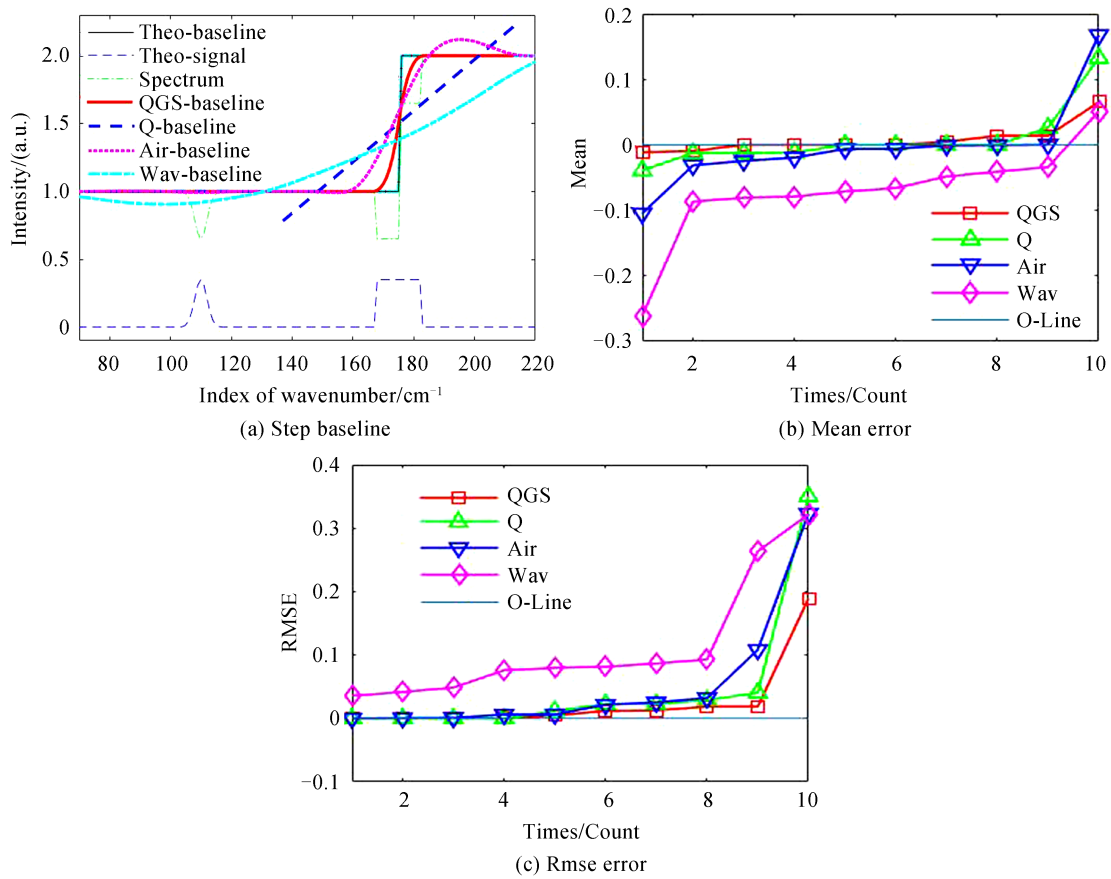
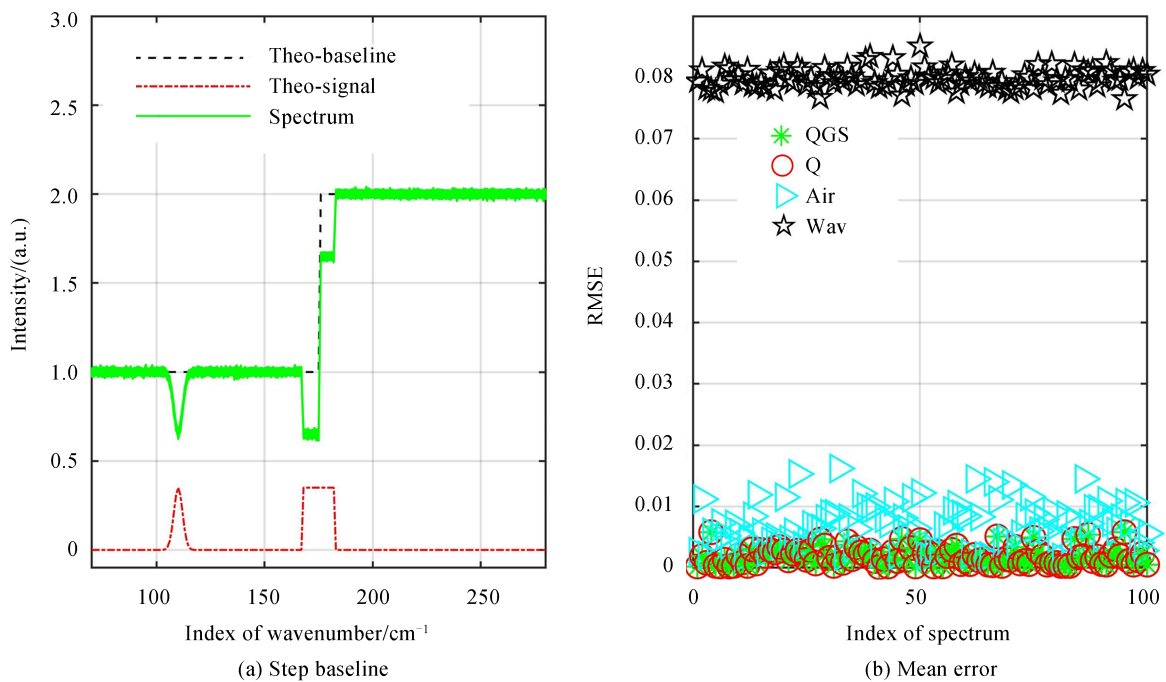


图3 无噪声时各算法拟合基线比较

Fig.3 Baseline-comparison of various algorithms fitting without noise

2.2.2 有噪声时QGS算法与其他算法的比较

图4直观显示了SNR=30时不同算法对基线畸变为step型时在不同曲率处进行拟合的100次实验结果.图4(b)为不同算法拟合信号1区域基线的结果,图4(c)为拟合信号2区域基线的结果.从图4(b)可以看



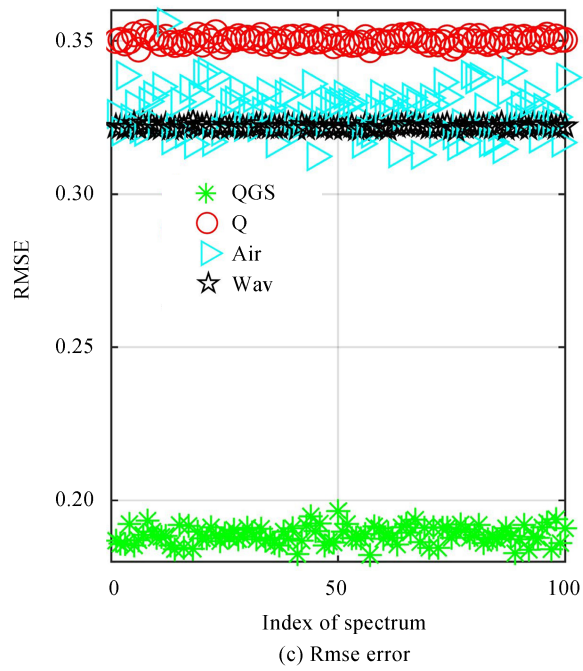
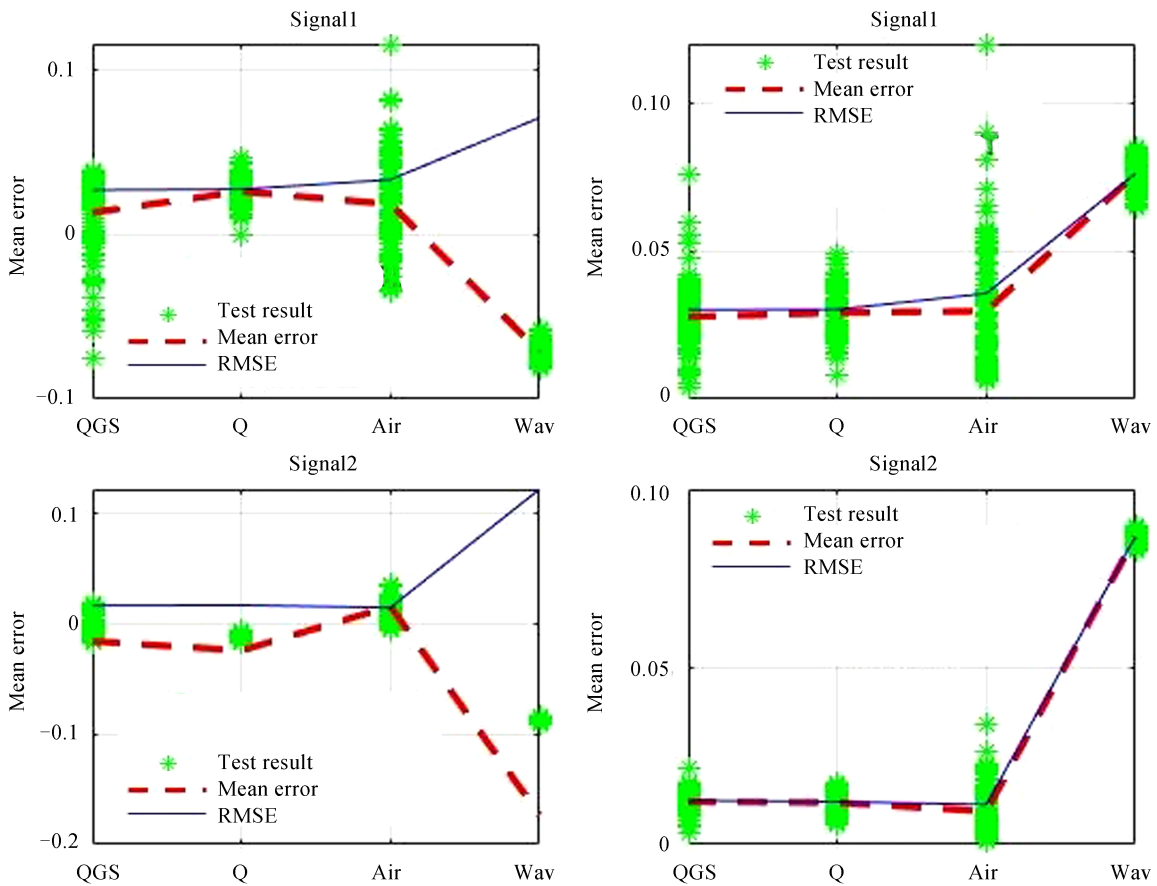


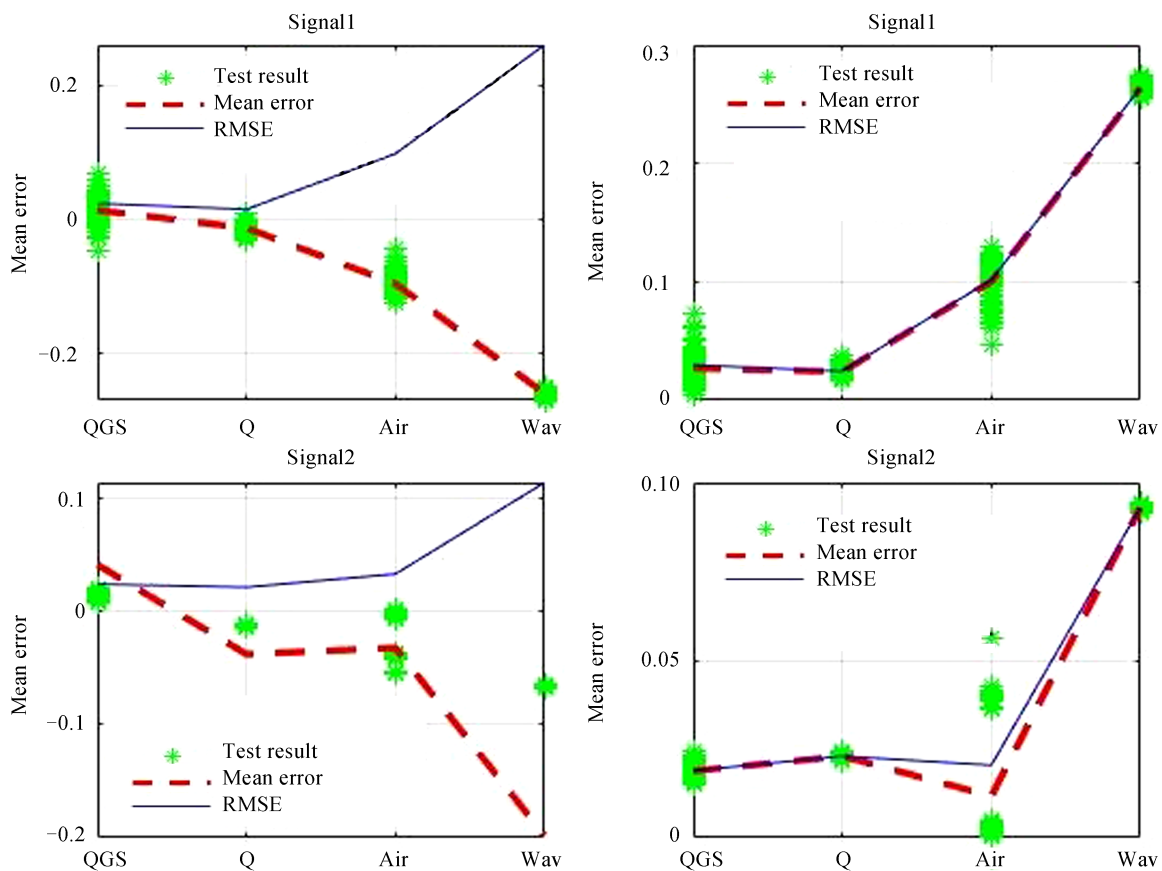
图4 SNR=30时step基线拟合比较

Fig.4 Comparison of baseline fitting of step when SNR=30

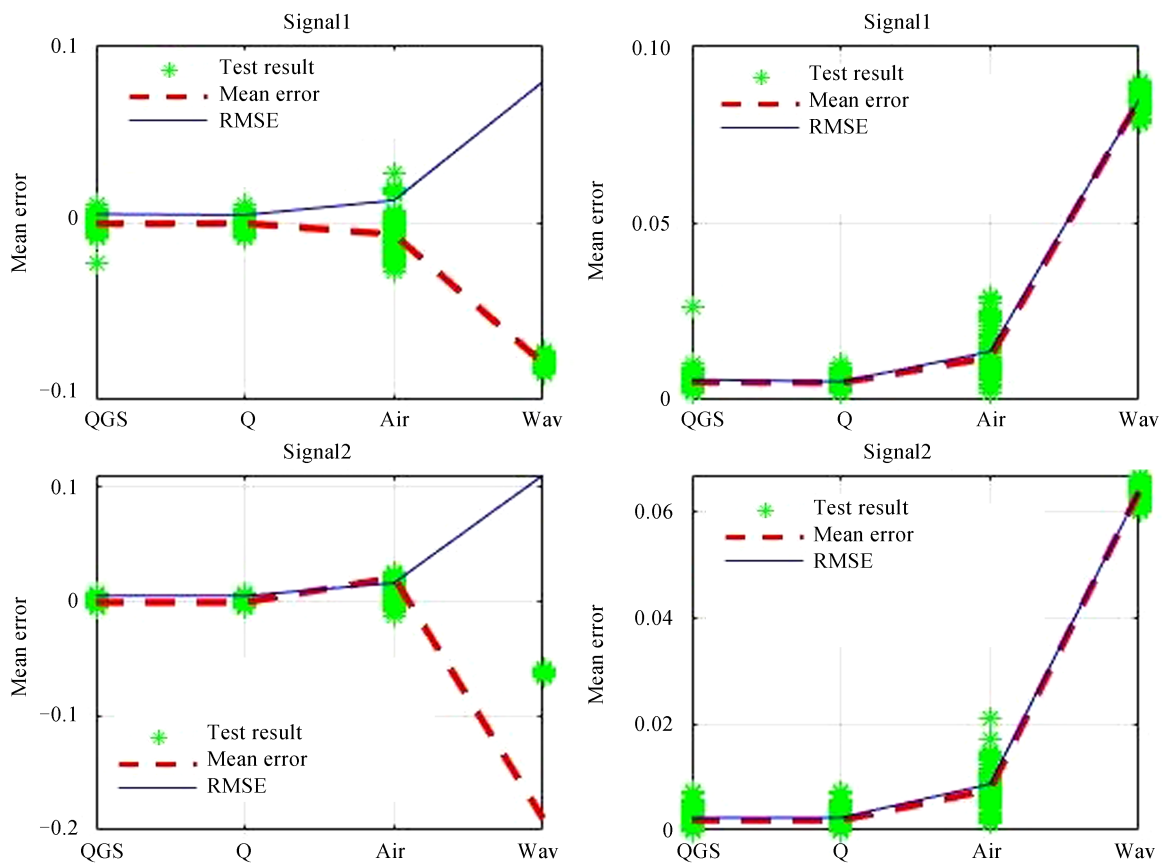
出当信号区域基线曲率变化不大时, QGS拟合结果与二次多项式拟合结果很相近. 从图4(c)中可以看出当信号区域基线出现阶跃时, QGS算法拟合结果较其他算法都更准确、稳定, 为了进一步分析不同算法在其他基线畸变类型下的拟合结果, 对不同算法在其他基线畸变下的拟合实验结果误差进行统计得到图5(a)~(d).



(a) The baseline distortion is Gaussian



(b) The baseline distortion is exponential



(c) The baseline distortion is sin

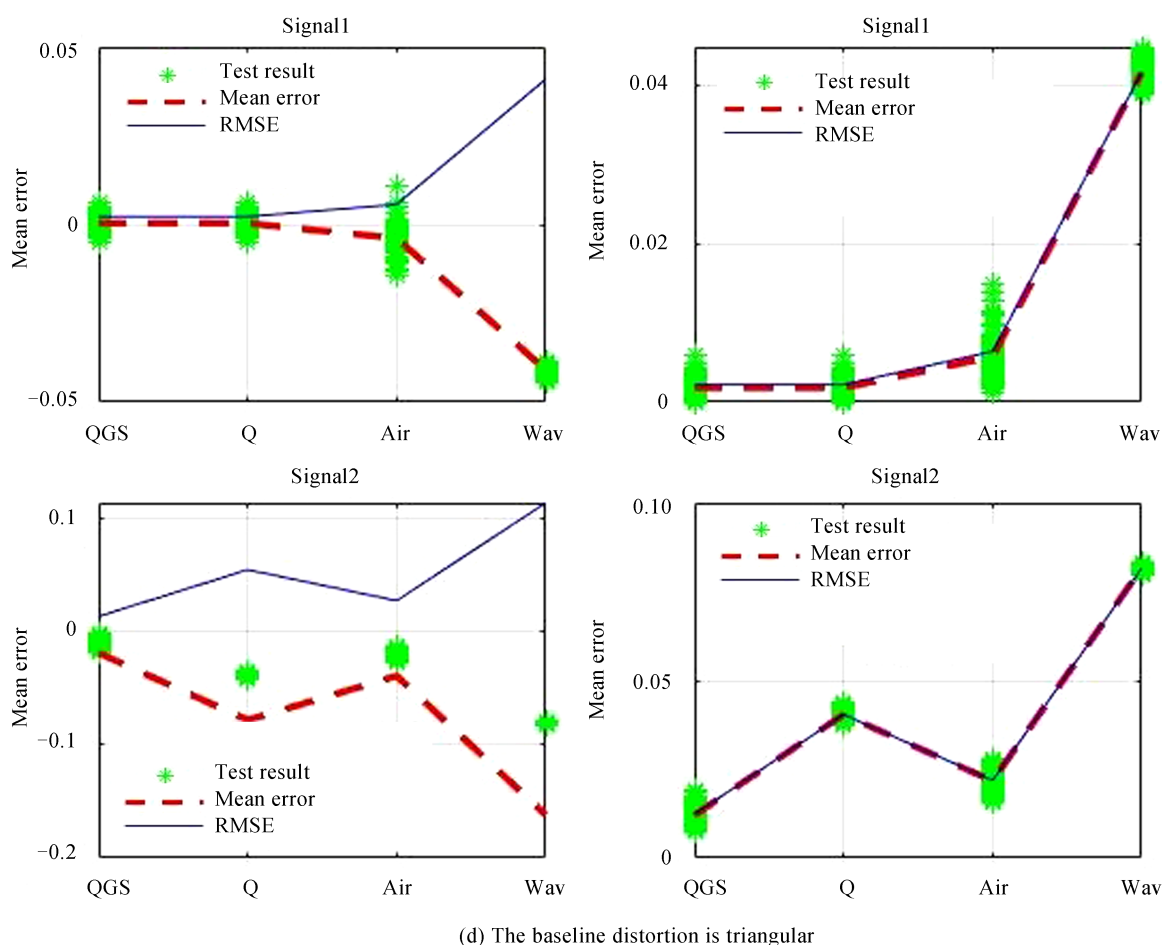


图5 SNR=30时100次实验的基线拟合结果
Fig.5 Baseline fitting results of 100 experiments when SNR=30

图5(a)中,左列子图中的每一个星花“*”代表一次拟合时信号区域拟合基线上各点的相对误差的均值,该值的大小表征了该信号区域该次基线拟合的准确度.为使结果更具有代表性,在相同条件下,不同算法各进行100次实验,每种算法的结果标记为100个星花“*”,再统计这100次实验数据的相对误差均值的均值和相对误差均值的均方根值,并分别用虚线和实线表示,其中虚线的值表征了该信号区域拟合基线结果在该算法下的准确度;实线的值表征了该信号区域拟合基线结果在该算法下准确度的稳定性.其右列子图中的每一个星花“*”代表一次拟合时信号区域拟合基线上各点的均方根误差值,该值的大小表征了该信号区域该次实验误差的离散度.与左列子图类似,同样进行100次实验,并统计其均值和均方根值,分别标记为虚线和实线,所不同的是虚线表征了该信号区域拟合基线结果在该算法下的离散度;实线表征了该信号区域拟合基线结果在该算法下离散度的稳定性.图5(b)~(d)各子图的意义和图5(a)相同,不同的是图5(b)~(d)是在不同的基线类型下完成的实验.

从图5(a)~(d)所有信号1及图5(a)、(c)信号2的拟合情况可以看出:当信号区域基线曲率变化不大时,QGS拟合和分段二次多项式拟合基线的准确度基本相同且其准确度的稳定性与离散度都较好.但当信号区域曲率变化较大或出现阶跃时,如图5(d)信号2的拟合情况,分段二次多项式拟合的准确度就远远不如QGS拟合.AirPLS拟合仅在图5(a)、(d)中与QGS拟合的结果相近,而Wavelet拟合的稳定性和离散度都很差.从图5(a)~(d)与图4整体上来看,无论基线是何种畸变,拟合的基线曲率是否突变,是否存在阶跃,本文所提出的QGS基线拟合算法都有较高的准确度、稳定性及较低的离散度.

2.2.3 不同噪声下QGS算法与其他算法的比较

为了进一步分析算法性能,变换不同的信噪比各进行100次实验,对所有实验结果进行统计得到图6(a)~(d).图6所有子图中红色实线表示QGS算法拟合,绿色点划线表示二次多项式拟合,蓝色实线表示AirPLS

拟合. 圆圈表示高斯型基线, 星号表示指数型基线, 三角型表示 sin 型基线, 菱形表示三角型基线, 正方形表示阶跃型基线.

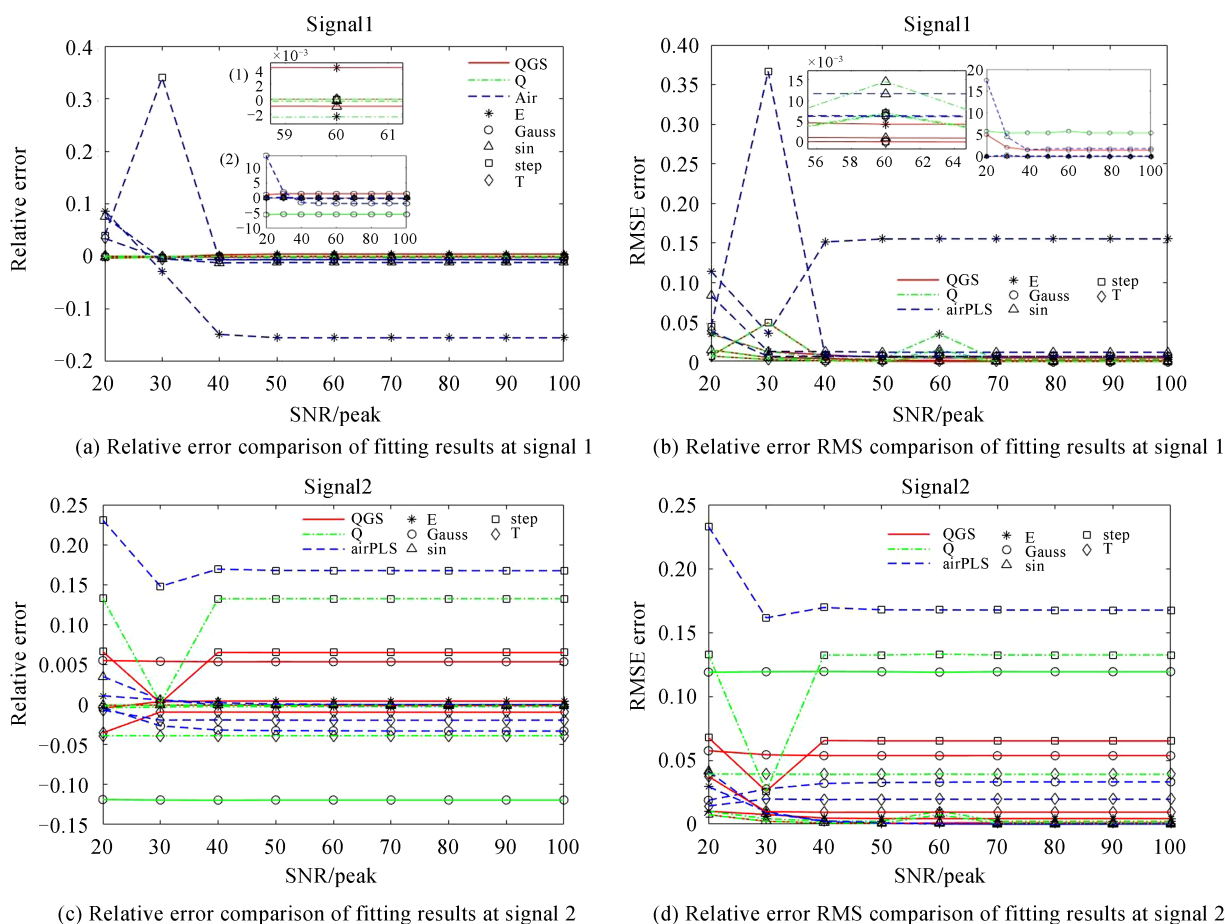


图6 不同信噪比、不同基线下不同算法的拟合结果

Fig 6 Fitting results of different algorithms with different noise ratios and different baselines

图6(a)、(b)中各点值为不同算法在不同基线下100次实验信号1处拟合结果相对误差均值的均值与均方根值,图6(c)、(d)中各点值为不同算法在不同基线下100次实验信号2处拟合结果相对误差均值的均值与均方根值,横坐标均为信噪比.图6(a)、(b)中子图(1)为SNR=60时的局部放大图,子图(2)为在高斯畸变下的误差结果缩放图.由于Wavelet拟合的相对误差较大,只比较QGS拟合、二次多项式拟合以及AirPLS拟合.从图6(a)、(c)可以看出,与其它算法相比,QGS算法对信号区域基线拟合结果的相对误差均值的均值都较小,说明QGS算法有较高的准确度;从图6(b)、(d)可以看出QGS算法的相对误差均值的均方根也明显较小,说明其也有较好的稳定性.分段二次多项式拟合算法在基线畸变类型为阶跃型、三角型、高斯型时信号2所对应区域基线拟合结果较QGS算法拟合结果很差.经分析,造成这一现象原因是阶跃型、三角型及高斯型基线在信号2区域附近基线不符合二次多项式且基线曲率变化较大.AirPLS算法在基线畸变为高斯型、三角型时拟合结果的准确度、稳定性和QGS算法相近.与其它算法相比,在不同信噪比下,QGS算法拟合基线的准确度、稳定性均较好,且基本不受基线畸变类型和基线曲率的影响.因此,QGS算法在拟合基线时具有良好的综合性能.

表1为统计的各算法分别在9种信噪比,5种基线类型下对两个信号区域进行基线拟合所得结果的相对

表1 各算法相对误差统计

Table1 The statistical relative error of each algorithm

	QGS	Q	Air	Wav
The mean of total relative error	1.20%	2.54%	3.36%	5.96%

误差均值的绝对值的均值,可以看出 QGS 拟合总相对误差仅为二次多项式拟合的 47.0%,为 AirPLS 拟合的 35.6%,为 Wavelet 拟合的 20%,表明 QGS 算法性能优异.

2.3 实测光谱数据 QGS 拟合实验

2.3.1 背景光谱数据拟合实验

背景光谱的获取方法为:EDF 提供 ASE 宽带光源,其增益范围为 1 525~1 565 nm;F-P 可调谐光滤波器用于选择透射波长;光电探测器为 InGaAs PIN 探测器,其工作波长范围为 1 000~1 650 nm.采集卡选用 NI 公司的 NI-USB-6251,其输入端口采集光电探测器检测到的光电压值,输出端口输出控制电压以驱动可调谐光滤波器实现透射波长扫描,通过输入输出端口的同步实现透射波长与该波长光功率值之间的对应.将采样得到的背景光谱数据作为实验的理论基线,如图 7(a)所示,随机将理论基线的一部分数据视作信号加以剔除,信号部分的横坐标中心通过随机产生,各信号的宽度也在 11~31 之间随机产生,然后用不同的算法分别得到信号部分的拟合基线,并计算各自的拟合误差.为了证明算法的可靠性,共随机产生 100 组信号,进行 100 次拟合实验,统计各次拟合结果的相对误差的均方根值,如图 7(b)所示.从图 7(b)中可以看出 QGS 算法拟合结果的相对误差均方值都接近于 0,有很好的稳定性与准确度,略优于二次多项式拟合与 AirPLS 拟合明显优 Wavelet 拟合.

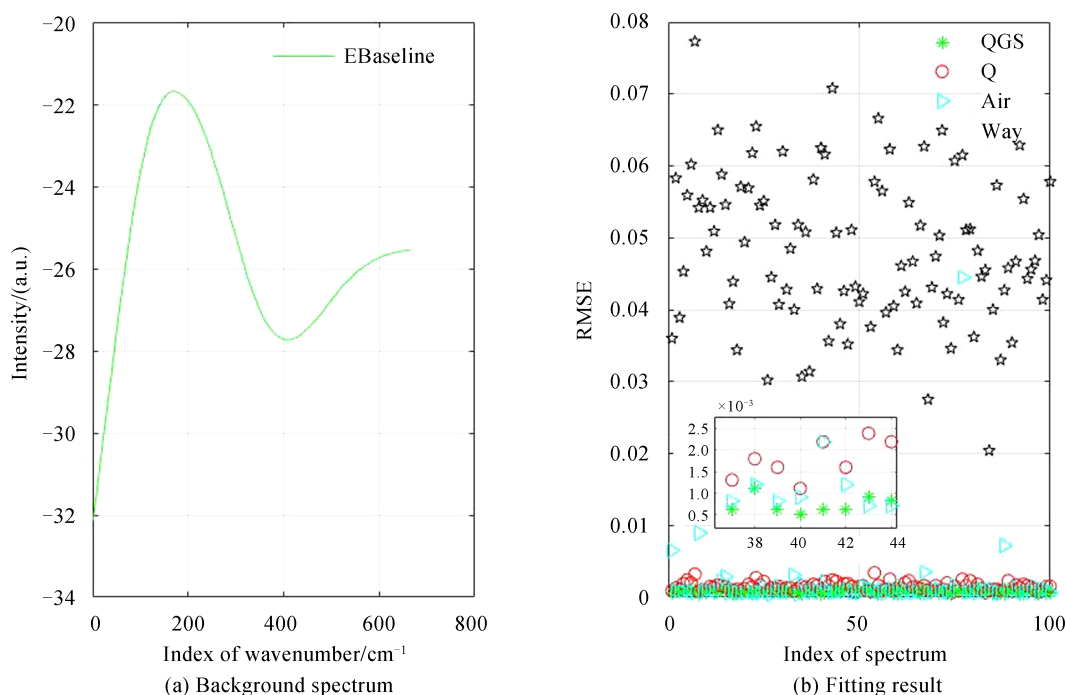


图 7 背景光谱实验数据基线拟合比较

Fig.7 Comparison of baseline fitting of background spectrum experimental data

2.3.2 色谱数据拟合实验

图 8(a)为真实色谱数据基线拟合结果图,色谱数据源自文献^[16],由于本文只研究基线拟合,尽量避免结果受信号提取的影响,因此信号区域采用手动标出.从图 8(a)可以看出此谱的信号左右两边数据存在阶跃现象,尽管如此,QGS 方法也成功地校正了基线,得到了比分段二次多项式拟合更合理的结果,明显优于 AirPLS 拟合和 Wavelet 拟合,更加忠实于原始数据,基线也更加平滑.图 8(b)为信号区域校正色谱数据图,从图中可以看出,由 QGS 算法校正出来信号出现负信号的数据更少,信号边缘更接近于 0,更符合对于信号提取的要求,体现了 QGS 算法的有效性.其中 AirPLS 和 Wavelet 算法效果太差是因为为了让评价更客观,全文采用了统一的参数(本文在仿真阶段对各算法参数进行调整使之达到最优之后将参数固定,后续变换光谱后参数不再更改).如果根据不同的实测数据修改 AirPLS 和 Wavelet 的参数,这两个算法也可以取得很好的效果,只是 AirPLS 和 Wavelet 参数的优化需要经验,而且修改参数会降低算法的通用性,后续 Raman 的

处理结果原因相同.

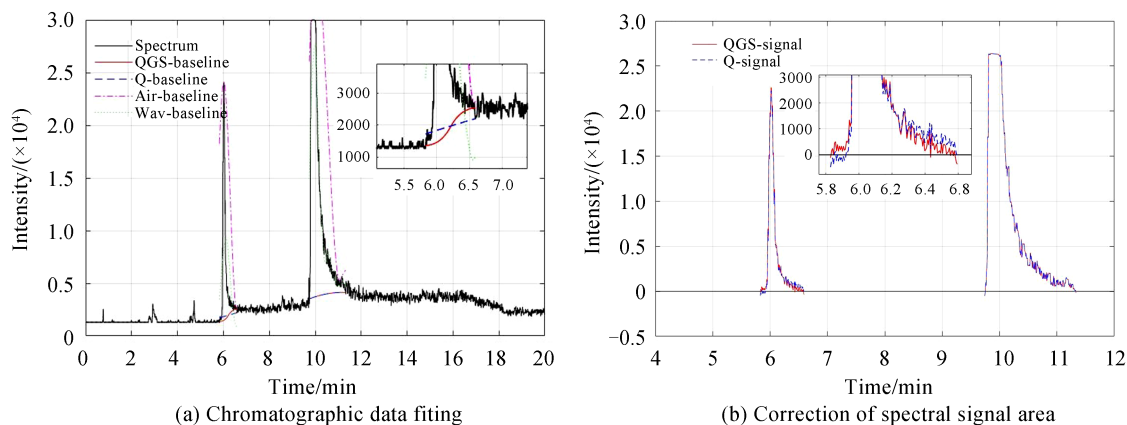


图8 真实色谱数据基线拟合比较

Fig.8 Comparison of baseline fitting of real chromatographic data

2.3.3 Raman 光谱真实数据拟合实验

图9的真实拉曼光谱数据是从文献^[17]中获得,同图8类似信号区域采用手动标出.从图9可知,QGS算法的拟合结果都接近于理论基线,明显优于AirPLS拟合和Wavelet拟合,在 $1\ 327\sim 1\ 429\ \text{cm}^{-1}$ 段起伏不明显处,各算法拟合的基线均与理论基线有较大偏差,这可能是由于两边基线区域过短,拟合数据点过少造成,也可能是因为该区域信号幅值较小导致相对误差较大(理论基线通过原始光谱与专家校正光谱相减得到,同样的校正误差,当信号幅值较小时,理论基线的相对误差会较大),但是与AirPLS拟合和Wavelet拟合相比,偏差还是处于可接受范围.由以上不同光谱数据拟合实验可知,QGS算法可以适用于多种光谱信号区域的基线拟合,若光谱信号区域宽度小于等于信号两边基线区域宽度,则QGS算法的效果最好.若信号区域跨度大,那么QGS算法将与分段二次多项式拟合相同.针对极端情况下如阶跃或曲率突变,QGS算法较其它算法拟合的基线更加忠实于原始基线,也更加平滑.综上,QGS拟合算法对各类真实光谱仍然有效.

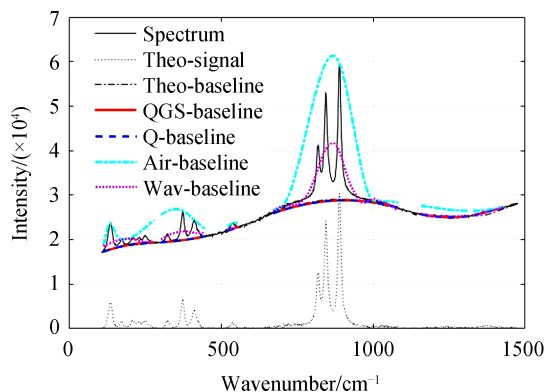


图9 Admite 光谱数据基线拟合比较

Fig.9 Comparison of baseline fitting of Admite spectrum

3 结论

针对分段二次多项式不能很好地处理信号区域基线曲率突变或阶跃情况提出了GradSuck拟合算法.在此基础上,为了让算法普适性更强提出分段二次多项式拟合结合GradSuck拟合的基线拟合算法——QGS算法.并利用仿真光谱和真实光谱验证了该方法的准确性,并与其他常用方法进行了比较.结果表明,与分段二次多项式拟合、AirPLS拟合等方法相比,QGS效果均具有较高的准确度与稳定性;同时,QGS算法与AirPLS等整体拟合算法相比,减少了计算量,提高了实时性.因此,本文所提出的方法,能够较好地实现光谱的基线拟合,可以作为一种有效的光谱基线拟合算法.

参考文献

- [1] JIA Yun-wei, SUN Si-yue, YANG Lu, *et al.* Salient space detection algorithm for signal extraction from contaminated and distorted spectrum[J]. *Analyst*, 2018, **143**(11): 2656-2664.
- [2] TANG Xiao-jun, WANG Jin, ZHANG Lei, *et al.* Spectral baseline correction by piecewise dividing in fourier transform infrared gas analysis[J]. *Spectroscopy and Spectral Analysis*, 2013, **33**(2): 334-339.
- [3] XIAN F, CORILO Y E, HENDRICKSON C L, *et al.* Baseline correction of absorption-mode Fourier transform ion cyclotron resonance mass spectra[J]. *International Journal of Mass Spectrometry*, 2012, **325**: 67-72.
- [4] ZHAO Man, GUO Yi-xin, HE Yu-qing, *et al.* Baseline correction of UV Raman spectrum based on improved piecewise linear fitting[J]. *Spectroscopy and Spectral Analysis*, 2020, **40**(6): 1862-1868.
赵曼, 郭一新, 何玉青, 等. 改进的紫外拉曼光谱分段线性拟合基线校正方法[J]. *光谱学与光谱分析*, 2020, **40**(6): 1862-1868.
- [5] HU Hai-bing, JING Bai, GUO Xia, *et al.* Improved baseline correction method based on polynomial fitting for Raman spectroscopy[J]. *Photonic Sensors*, 2018, **8**(4): 332-340.
- [6] WANG Xin, LU Shi-long, LI Yan, *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.
- [7] SHEN Xian-chun, XU Liang, YE Shu-bin, *et al.* Automatic baseline correction method for the open-path Fourier transform infrared spectra by using simple iterative averaging[J]. *Optics Express*, 2018, **26**(10): A609-A614.
- [8] YE Song, GAN Yong-yin, XIONG Wei, *et al.* Baseline correction of spatial heterodyne spectrometer using wavelet transform[J]. *Infrared and Laser Engineering*, 2016, **45**(11): 209-213.
叶松, 甘永莹, 熊伟, 等. 采用小波变换的空间外差光谱仪基线校正[J]. *红外与激光工程*, 2016, **45**(11): 209-213.
- [9] HU Hong-ping, ZHANG Lin-mei, YAN Hui-chao, *et al.* Denoising and baseline drift removal method of MEMS hydrophone signal based on VMD and wavelet threshold processing[J]. *Ieee Access*, 2019, **7**: 59913-59922.
- [10] FENG Fei, WANG Fu-bei, XIE Fei, *et al.* Implementation of spectral overlap resolution based on wavelet transforms and gaussian fitting[J]. *Acta Photonica Sinica*, 2015, **44**(6): 113-118.
冯飞, 王府北, 谢非, 等. 小波变换与高斯拟合在光谱重叠峰解析中的应用[J]. *光子学报*, 2015, **44**(6): 113-118.
- [11] SAWALL M, VON H E, MOOG A, *et al.* Multi-objective optimization for an automated and simultaneous phase and baseline correction of NMR spectral data[J]. *Journal of Magnetic Resonance*, 2018, **289**: 132-141.
- [12] LAU SK, WINLOVE P, MOGER J, *et al.* A Bayesian Whittaker-Henderson smoother for general-purpose and sample-based spectral baseline estimation and peak extraction[J]. *Journal of Raman Spectroscopy*, 2012, **43**(9): 1299-1305.
- [13] BAO Qing-jia, FENG Ji-wen, CHEN Fang, *et al.* A new automatic baseline correction method based on iterative method [J]. *Journal of Magnetic Resonance*, 2012, **218**: 35-43.
- [14] ZHANG Feng, TANG Xiao-jun, TONG Ang-xin, *et al.* Baseline correction for infrared spectra using adaptive smoothness parameter penalized least squares method[J]. *Spectroscopy Letters*, 2020, **53**(3): 222-233.
- [15] ZHANG Zhi-min, CHEN Shan, LIANG Yi-zeng. Baseline correction using adaptive iteratively reweighted penalized least squares[J]. *Analyst*, 2010, **135**(5): 1138-1146.
- [16] FU Hai-yan, LI He-dong, YU Yong-jie, *et al.* Simple automatic strategy for background drift correction in chromatographic data analysis[J]. *Journal of Chromatography A*, 2016, **1449**: 89-99.
- [17] LIU Jun-tao, SUN Jian-yang, HUANG Xiu-zhen, *et al.* Goldindec: a novel algorithm for raman spectrum baseline correction[J]. *Applied Spectroscopy*, 2015, **69**: 834-842.