

## Research Article

# Collaborative Penalized Least Squares for Background Correction of Multiple Raman Spectra

Long Chen ,<sup>1</sup> Yingwen Wu,<sup>1</sup> Tianjun Li,<sup>1</sup> and Zhuo Chen<sup>2</sup>

<sup>1</sup>Faculty of Science and Technology, University of Macau, E11 Avenida da Universidade, Taipa, Macau

<sup>2</sup>Chemistry and Chemical Engineering, College of Biology, Hunan University, Changsha 410082, China

Correspondence should be addressed to Long Chen; longchen@umac.mo

Received 21 April 2018; Revised 8 July 2018; Accepted 26 July 2018; Published 29 August 2018

Academic Editor: Małgorzata Jakubowska

Copyright © 2018 Long Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Although Raman spectroscopy has been widely used as a noninvasive analytical tool in various applications, backgrounds in Raman spectra impair its performance in quantitative analysis. Many algorithms have been proposed to separately correct the background spectrum by spectrum. However, in real applications, there are commonly multiple spectra collected from the close locations of a sample or from the same analyte with different concentrations. These spectra are strongly correlated and provide valuable information for more robust background correction. Herein, we propose two new strategies to remove background for a set of related spectra collaboratively. Based on weighted penalized least squares, the new approaches will use the fused weights from multiple spectra or the weights from the average spectrum to estimate the background of each spectrum in the set. Background correction results from both simulated and real experimental data demonstrate that the proposed collaborative approaches outperform traditional algorithms which process spectra individually.

## 1. Introduction

Raman spectroscopy, which provides valuable chemical and physical information of studied samples, is widely used as an analytical tool for many applications like material identification, chemical detection, and biomedical analysis [1–3]. Peaks in Raman spectra are the fingerprints of the analyte, and the corresponding peak heights or peak areas have strong correlations with the concentration of the analyte. However, spectral interferences have a strong negative effect on the measurement of peaks, and this in the long run hinders the performance of Raman spectroscopy-based quantitative analysis [4]. Representative interferences for Raman spectra include backgrounds mainly caused by instrument fluctuations and fluorescent substances. The noises of the instrument and the occasional spikes caused by cosmic rays also deteriorate the quality of Raman spectra. As a result, some preprocessing steps should be conducted to handle the interferences in the Raman spectra. In this paper, we mainly focus on the background correction problem.

Numeric background correction algorithms have been proposed in the past decades for Raman and other spectra. For example, the wavelet transform is used as a powerful tool for background removal by decomposing the Raman signals in the frequency domain [5–7]. Because the performance of wavelet-based approaches is greatly affected by the selection of base wavelets and scales, the adaptive wavelet transform was used in [8] to obtain a multiresolution decomposition of a Raman spectrum. The low-frequency background and high-frequency noise were removed thereafter.

The iterative smoothing algorithms also play an important role in background estimation because the background is usually characterized by its smooth variation. The general procedure of this approach is continually smoothing the spectrum until the background is obtained. Many well-known smoothing filters and their enhancements have been widely used in previous studies for the purpose of iteratively removing peaks and deriving backgrounds in the spectra [9–12]. The drawback of such smoothing algorithms is the difficulty in automating their iterations, although some endeavors have been made on this issue [10, 11].

Curve fitting is to fit the background with appropriate points in the spectrum by some fidelity or loss functions like the least squares [13–15]. Such a kind of selection-then-approximation approach is very similar to the manual background estimation procedure. The simple implementation and short running time of curve fitting have made it one popular background correction method used in real applications. Different curves such as Bezier curves [16], splines [13], and polynomial functions [14, 17] have been used to fit the background. On the contrary, without specifying the curve shape of the background, the penalized least squares- (PLS-) based algorithms attempt to automatically estimate the background by a direct approximation of the spectrum with a penalization on the roughness of the approximated curve [15, 18]. In the PLS-based algorithms, one critical issue is the weight setting for different points in the spectrum. These weights are used to indicate the contribution of corresponding points to the final curve construction. Many methods have been proposed to this end [19, 20], and some automatic setting techniques are also suggested [21, 22].

Some comparisons on different baseline correction approaches have been conducted [23], and the optimal choice of background removal for the statistical analysis of spectra has been explored [24]. However, by far, there is no single automatic method that can well handle all the spectra universally and be regarded as the best. Recently, more new baseline correction algorithms like the ones based on sparse representation [25] and neural networks [26] have been proposed.

In practical applications of Raman spectroscopy, multiple measurements of a given analyte are normal practices. A set of strongly correlated spectra is then obtained although they may be generated under different environmental conditions and sampling protocols. In the quantitative Raman analysis like the mixture analysis and multivariate calibration, except multiple measurements, multiple spectra derived from the same mixture with different analyte concentrations are also correlated. Due to the varying backgrounds and random noises in the set of related spectra, the clear information like the peak locations in one spectrum may not be significant in another spectrum. How do we collaboratively use the valuable information in a set of related spectra for the purpose of spectrum preprocessing? Foist et al. first noticed this problem and proposed a method to denoise multidimensional spectral data collaboratively [27]. For background correction, few approaches have been proposed by utilizing the common characteristics shared in a set of related spectra [6, 28, 29]. For instance, the multiple spectra baseline correction (MSBC) algorithm designed in [28] assumed that the pairwise differences between the background removed spectra are small and inserted a regularization for this prior to the asymmetric least squares.

In this paper, based on PLS, we propose a new approach focusing on collaborative background correction for a set of related spectra. Specifically, our main contribution is to design two ensemble strategies to embed the weight information of PLS from multiple spectra to boost each

spectrum's background correction. For PLS, the weight of each point in a spectrum denotes the contribution of the point to the final background estimation. In the first scheme, we directly use the weights derived from the average spectrum (by averaging all the related spectra) to calculate the background of each spectrum. The second scheme applies the average of weights obtained from each spectrum using traditional PLS-based approaches. By using the two schemes, our new approach utilizes the strong correlations among multiple spectra and suppresses the effect of noise and signal variation in different spectra.

To illustrate the advantage of collaboratively calculating the weights for PLS algorithms when handling several related spectra, we combine the adaptive iteratively reweighted penalized least squares algorithm (airPLS) [19] and the morphological weighted penalized least squares algorithm (MPLS) [20] with the proposed weight ensemble strategies. In the experiments on the synthetic and real Raman spectra, these enhanced PLS approaches show accurate and robust background removal capability.

## 2. Theory

**2.1. PLS for Background Correction.** The signal smoothing problem was first proposed by Whittaker in 1922 [30]. The pioneering works on applying PLS for baseline correction were conducted by Eilers et al. more than 10 years ago [31, 32]. The rationale behind PLS is to approximate the observed data by balancing the conflicts between the fidelity to original data and the roughness of fitting data.

Assume that  $y$  is a vector of the Raman spectrum and  $z$  is the fitting vector; both of them are with the length of  $N$  elements. The fitted  $z$  should keep the fidelity to  $y$  as well as the roughness of the fitted vector.  $F$  denotes the fidelity to the Raman spectrum  $y$ , which can be expressed as the sum of squares of differences between  $y$  and  $z$ :

$$F = \sum_{i=1}^N (y_i - z_i)^2. \quad (1)$$

$R$  denotes the roughness of the fitting vector  $z$ , which can be expressed as the sum of squares of differences between each element of  $z$  and its neighbors:

$$R = \sum_{i=2}^N (z_i - z_{i-1})^2, \quad (2)$$

where the square of first differences penalty is adopted in (2) to simplify the presentation. In other cases, it is also a natural way to quantify the roughness by the square of higher-order differences.

The following equation is adopted to measure the balanced combination of fidelity and roughness:

$$Q = F + \lambda R, \quad (3)$$

where  $\lambda$  is a user adjustable parameter that balances the fidelity and roughness. Larger  $\lambda$  favours a smoother fitted vector.

In order to apply the PLS to estimate background, a weight vector  $w$  was introduced for fidelity; its element  $w_i$  can be regarded as a weight that depicts the reliability of point  $i$  as a part of background. Then,  $F$  is changed to

$$F = \sum_{i=1}^N w_i (y_i - z_i)^2. \quad (4)$$

To solve the minimization problem of (3), we get a linear system by equating the partial derivatives of  $Q$  to zero ( $\partial Q/\partial z = 0$ ), and the matrix form of the obtained linear system is as follows:

$$(\text{diag}(w) + \lambda D^T D)z = \text{diag}(w)y, \quad (5)$$

where  $\text{diag}(w)$  is a diagonal matrix with  $w$  on its diagonal and  $D$  is the derivative of an identity matrix. Finally, we solve the fitting vector as follows:

$$z = (\text{diag}(w) + \lambda D^T D)^{-1} \text{diag}(w)y. \quad (6)$$

There have been some proposed methods for the weight calculation in PLS. To control the smoothness of the fitted vector iteratively, the airPLS method [19] calculates the weight vector  $w$  in an adaptive way.  $w_i$  in each iteration  $t$  is obtained as follows:

$$w_i = \begin{cases} 0, & y_i \geq z_i^{t-1}, \\ e^{t(y_i - z_i^{t-1})/|d^t|}, & y_i < z_i^{t-1}. \end{cases} \quad (7)$$

The vector  $d^t$  consists of negative elements obtained from the subtraction between  $y$  and  $z^{t-1}$  in the  $t$ th iteration step. The fitted vector  $z^{t-1}$  in the previous  $t-1$  iteration step is a candidate of the baseline. If the value of the signal  $y_i$  is greater than the candidate, it can be seen as a part of the peak, of which the weight is set to zero. If not, the weight is calculated as (7). When the iteration count reaches the maximum or when the following termination criterion is satisfied, the iteration will stop and the final weight vector is used for PLS to generate the background:

$$|d^t| < 0.001 \times |y|. \quad (8)$$

Unlike airPLS which adjusts the weights adaptively, the MPLS method [20] directly calculates the weight vector by applying the mathematical morphology operations on the spectrum to remove the peaks and generate a rough background firstly. The morphology operation involves an object spectrum  $y$  and a plane structuring element  $E$ . The transformation is an opening operation which consists of dilation and erosion. To refine the background, the local minimum points between peak areas are selected as meaningful background points with weight 1, and the remaining points are set with a weight of 0. The weighted PLS is then applied to get the final background.

**2.2. Collaborative Weighted Penalized Least Squares for Multiple Spectra.** As discussed in Introduction, in practical applications of Raman spectroscopy, we may collect strongly related spectra that are from either the same kind of material or the solution with different proportional concentrations.

In these cases, we can collaboratively estimate each spectrum's background by comprehensively considering the valuable information shared in the whole set of spectra. Specifically, for the weighted PLS-based approaches, we design two schemes to utilize the global information in the set of spectra for the weight calculation.

The simplest information fusion approach for a set of highly related spectra is to average them. More formally, given a set of  $m$  spectra  $y^1, y^2, \dots, y^m$ , we calculate the average spectrum as follows:

$$y^{\text{avg}} = \sum_{i=1}^m y^i. \quad (9)$$

By averaging, the effect of noise on spectra is suppressed, and the average spectrum can be regarded an informative representation of the set of related spectra. With the average spectrum, we can apply some traditional weighed PLS-based approaches to calculate the weights of different points that denote the reliabilities of their reliabilities as some parts of background. Then, the obtained weight vector  $w^{\text{avg}}$  is used for each single spectrum's background removal. For a more accurate fusion, we may set weights for different spectra in the summation of (9). For example, the high-quality spectrum with a higher signal-noise ratio may take a higher weight to contribute more to the final average spectrum. But in our experiments, we find the simple summation produces good results as well.

The second scheme to fuse the information from multiple spectra is to ensemble the weight vectors of all the spectra. For the spectrum  $y^i$ , we first use some traditional weighted PLS-based approaches to calculate the weight vector  $w^i$  for it. Then, the weight vectors for all the related spectra are combined into one as follows:

$$w^{\text{com}} = \sum_{i=1}^m w^i, \quad (10)$$

where the combined weight vector  $w^{\text{com}}$  will be used as the final weight vector in weighted PLS-based background estimation of each spectrum.

To illustrate the application of the two schemes proposed above, we improved the airPLS and MPLS methods by modifying their final weight vectors. In these two PLS-based background estimation methods, no matter the weight vector for points in the spectrum is adaptively adjusted in the iteration (the airPLS case) or the weight vector is directly obtained from morphology operations, the final step to estimate the background is a linear regression step that solves (5) by using (6) with the determined weights.

Given a set of  $m$  spectra  $y^1, y^2, \dots, y^m$ , we have the following 4 enhanced PLS approaches to estimate each spectrum's background.

**2.2.1. Average Spectrum-Based airPLS (AS airPLS).** In this method, we first derive the average spectrum  $y^{\text{avg}}$  from (9). Then, the airPLS is applied over the  $y^{\text{avg}}$ . But here, we only record the weight vector derived at the last iteration of airPLS using (7). This weight vector is denoted as  $w^{\text{avg}}$ . Now,

the background of  $y^i$  is calculated by using (6), in which the weight vector  $w$  is  $w^{\text{avg}}$  and the spectrum  $y$  is  $y^i$ .

**2.2.2. Combined Weight-Based airPLS (CW airPLS).** This method first applies airPLS for each spectrum. We only record the final weight vector for each spectrum obtained in the last iteration of airPLS. Then, the combined weight vector is calculated by using (10), and we use this weight vector to derive each spectrum's background by using (6).

**2.2.3. Average Spectrum-Based MPLS (AS MPLS).** This method is very similar to AS airPLS. The only difference is that  $w^{\text{avg}}$  is calculated by applying MPLS on the average spectrum  $y^{\text{avg}}$ .

**2.2.4. Combined Weight-Based MPLS (CW MPLS).** This method is similar to CW airPLS. The difference is that the initial weight vector for each spectrum is obtained by MPLS instead of airPLS.

Because the weight vector is also extensively used in other weighted PLS-based background removal approaches [33–35], including some recently proposed enhancements of airPLS [18], the average spectrum and combined weight-based schemes proposed in this paper can be easily adopted by these approaches to collaboratively process a set of related spectra.

### 3. Experimental

To verify the performance of collaborative approaches proposed in this paper on background removal of a set of related Raman spectra, we compare them with traditional approaches on simulated spectra and real Raman spectra.

**3.1. Simulated Data.** For the simulated data, each spectrum  $s^i$  in the set  $S$  consists of the pure signal  $p^i$ , background  $b^i$ , and noise  $n^i$ :

$$s^i = p^i + b^i + n^i. \quad (11)$$

As our methods are to process multiple correlated spectra, the simulated pure signal  $p^i$  is a mixture of three pure spectra illustrated in Figure 1(a) that represent three chemical components. The concentrations of different components are randomly drawn from 0% to 100% in the mixture. By doing this, the simulated pure signals show variations of the peaks. Figure 1(b) depicts some simulated pure signals with different concentrations of components. In addition, the background  $b^i$  in (11) is generated by exponential, polynomial, sigmoid, or sine curves with a random amplitude. Finally, the Gaussian white noise  $n^i$  is added to the summation of the background and pure signal. Altogether, we randomly generate 30 simulated spectra for each type of backgrounds, as shown in Figures 1(c)–1(f). Their corresponding backgrounds are plotted.

We compare the performance of our improved methods (AS airPLS, AS MPLS, CW airPLS, and CW MPLS) with one of the original airPLS and MPLS methods. Unlike the collaborative AS- and CW-based methods, the airPLS and MPLS process the spectra in the simulated data separately. In addition, the multiple spectra baseline correction (MSBC) algorithm proposed in [28] is also included in comparison.

In these simulated data, as the ground truth backgrounds are known, we can evaluate the performance of each method by calculating the mean squared error (MSE) or the relative error (RE) between the real background  $b_i$  and the estimated background  $\tilde{b}_i$ :

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (b_i - \tilde{b}_i)^2, \quad (12)$$

$$\text{Relative error} = \frac{1}{N} \sum_{i=1}^N \frac{|b_i - \tilde{b}_i|}{b_i} \%. \quad (13)$$

According to (12) or (13), for each spectrum in the 30 simulated spectra, we can get one MSE or RE value. For the multispectra, we compare the average MSE or average RE of the 30 simulated spectra. The smaller value of MSE or RE denotes the estimated background is more similar to the ground truth. Therefore, the smaller MSE or RE indicates the better background correction method.

As shown in Table 1, our AS- and CW-based methods can obtain much better RE than the traditional approaches. The bar plot in Figure 2 visualizes the results in Table 1. From these results, we can see the multiple spectra baseline correction (MSBC) algorithm performs similar to or a little better than the approaches relying on separate spectrum processing. But its performance is worse than the collaborative approaches proposed in this paper.

To have a better comparison, we also plot several representative pairs of ground truth backgrounds (the simulated sine curves) and the estimated ones obtained by different methods in Figure 3. The zoomed-in inserts from  $600\text{ cm}^{-1}$  to  $800\text{ cm}^{-1}$  are provided in Figure 3 for better visualization. It is pretty obvious that the estimated backgrounds obtained from our enhanced approaches (the red and green lines) are much closer to the real background (the black line).

Because the simulated data are spectra of some mixtures of three chemical components, we can conduct a regression analysis on the preprocessed spectra to predict the concentration values of different components. Here, the principal component regression (PCR) is applied on 29 simulated spectra, and the concentrations of three components in the remainder spectrum are predicted thereafter. We do this 30 times by choosing each simulated spectrum as the testing one once. The root mean squared error (RMSE) is used as the prediction error to evaluate different methods' performance. As shown in Table 2, the spectra preprocessed by the enhanced PLS show higher prediction accuracy when compared with other approaches.

For all the six PLS methods, the smoothness parameter  $\lambda$  should be adjusted to get a good estimation of

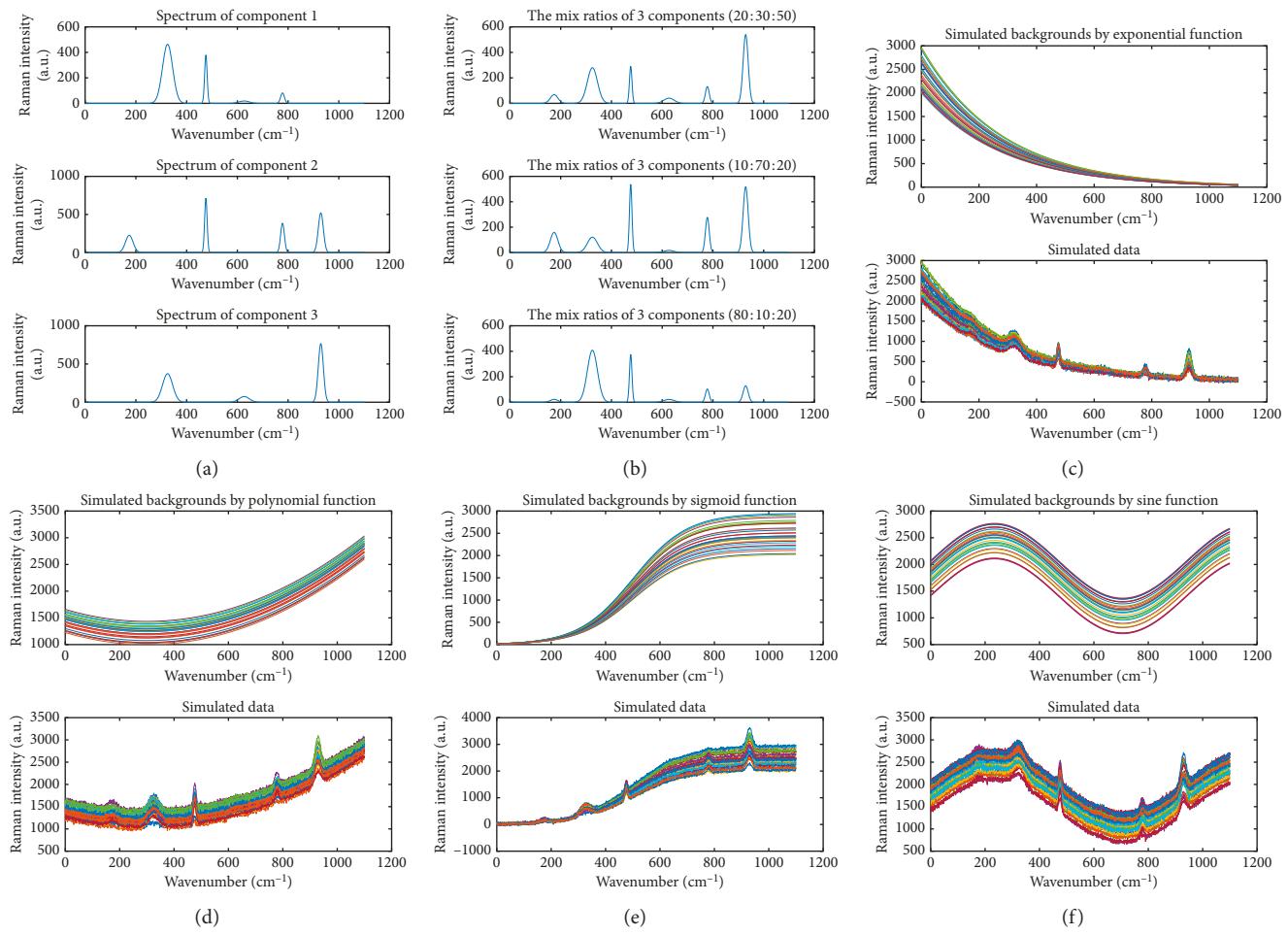


FIGURE 1: Simulated data. (a) Spectra of three chemical components. (b) Three sampling mixture signals with different concentrations of components. (c-f) The simulated spectra and their backgrounds.

TABLE 1: Relative error for different backgrounds.

	Exponential (%)	Polynomial (%)	Sigmoid (%)	Sine (%)
airPLS	21.3	1.9	24.1	1.5
MPLS	19.8	2.2	26.3	1.3
MSBC	19.6	2.1	26.3	2.1
AS airPLS	3.7	0.3	5.3	0.3
CW airPLS	1.5	0.2	2.1	0.3
AS MPLS	3.6	0.3	4.6	0.6
CW MPLS	3.5	0.5	7.0	0.8

background. If  $\lambda$  is too small, the estimated background would fit the peaks too much. However, if  $\lambda$  is too large, the estimated background would be too smooth to catch the fluctuation of backgrounds. So, it is necessary to delve further into the value of  $\lambda$ . As we have mentioned before, we compare the performance of different methods by calculating the MSE between real and estimated backgrounds. Here, we design the experiments with various  $\lambda$  and compare the MSE so as to find the optimum  $\lambda$ .  $\lambda$  is varied in the log scale as is recommended in Eilers' paper [31]. Therefore, we test with  $\lambda$  values from  $10^2$  to  $10^8$  and

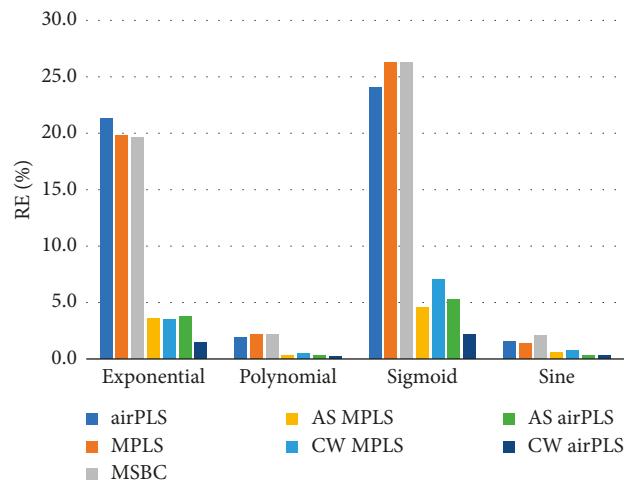


FIGURE 2: Relative error of different methods on different backgrounds.

discard the  $\lambda$  value with bad estimated background at first glance. Through comprehensive consideration, we let  $\lambda$  to change from  $10^4$  to  $10^7$  and record the MSE value for each

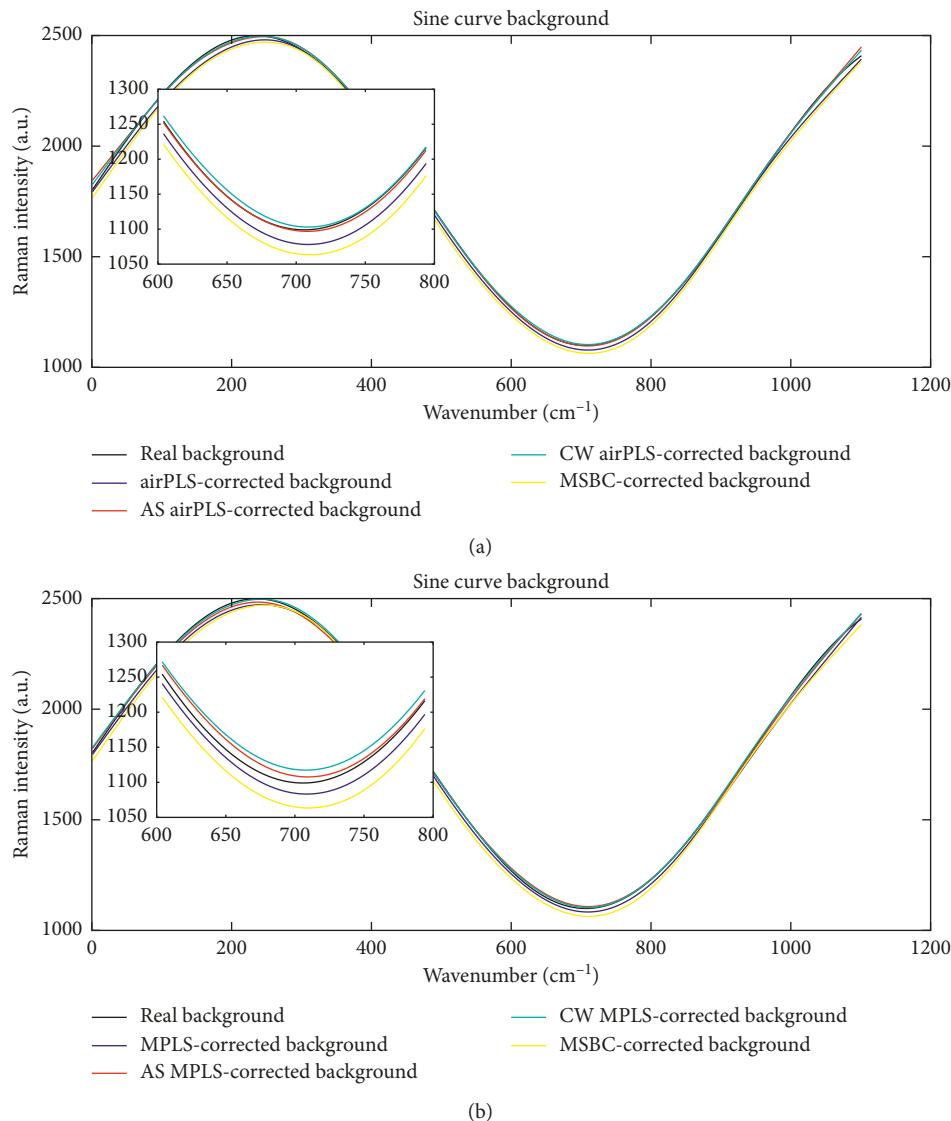


FIGURE 3: (a, b) Comparison of real backgrounds and estimated backgrounds obtained by different methods.

TABLE 2: RMSE of PCR analysis on preprocessed spectra.

	Exponential	Polynomial	Sigmoid	Sine
airPLS	0.0216	0.0202	0.0206	0.0208
MPLS	0.0226	0.0214	0.0224	0.0208
MSBC	0.0249	0.0248	0.0248	0.0249
AS airPLS	0.0181	0.0185	0.0185	0.0185
CW airPLS	0.0219	0.0188	0.0189	0.0189
AS MPLS	0.0186	0.0180	0.0190	0.0188
CW MPLS	0.0203	0.0196	0.0199	0.0193

method. The MSE results of each method with various  $\lambda$  are displayed in Figure 4. Overall, our collaborative methods can always obtain lower MSE compared with original methods, which means that these methods can get a better estimation of background. The optimal results appear around  $\lambda = 10^{5.2}$ , and similar performance can be obtained between  $10^{4.8}$  and  $10^{5.8}$ . It seems that our methods

are not too sensitive to  $\lambda$ . What it means is that our methods are relatively robust to the choice of  $\lambda$ .

**3.2. Real Raman Spectra.** Graphene-isolated-Au-nanocrystal (GIAN) is a unique and stable nanostructure and has been utilized for different biomedical applications, such as sensitive Raman imaging, drug loading for chemotherapy, and photothermal therapy [36]. Herein, we utilized the GIAN to verify the efficiency of proposed background correction approaches. The Raman spectra of GIANs were collected via a static scan in the region of  $200\text{--}2200\text{ cm}^{-1}$ . The spectrum was saved every 0.01 second, and 45 numbers of spectra signals were collected as shown in Figure 5. Because the set of spectra is collection from the same sample in a short time, although the varying backgrounds make the spectra show different intensities, the background-removed spectra should show the same intensity ideally. As shown in Figure 6,

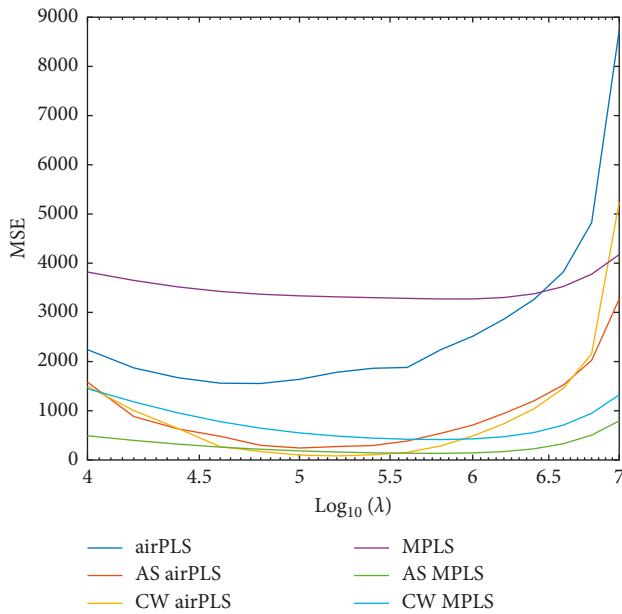
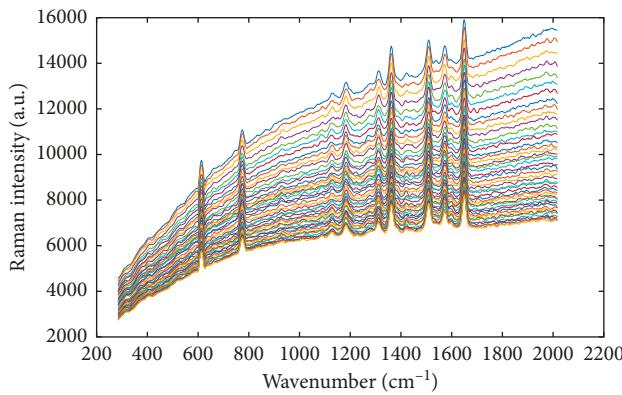
FIGURE 4: MSE of each PLS method for different  $\lambda$ .

FIGURE 5: A set of Raman spectra for GIANs collected in a short time.

the processed spectra by AS- and CW-enhanced PLS methods show more homogeneous features of spectra. This is especially observable in the zoomed region from  $1500\text{ cm}^{-1}$  to  $1600\text{ cm}^{-1}$ .

To quantitatively measure the consistency of background-removed spectra, the average intensity variance of all the processed spectra is calculated. A smaller variance means a closer relation of the background-removed spectra, which indicates a relative better background correction result. From Table 3, we see the variances of the AS and CW methods are smaller than those of the original methods, which implies the better performance of collaborative methods.

As a commonly used approach to quantify the background correction [19, 20, 35], principal component analysis (PCA) is also conducted on the matrix of spectra before and after background removal. In PCA, the score of a data point along with a principal component (PC) is the distance from the origin to the data point's projection on this PC. Figure 7 shows the score plot of the original spectra according to the

first two principal components. The PCA score plots of background-corrected spectra are shown in Figure 8. For the real Raman data used in this experiment, the discrepancy in spectra is mainly caused by the baseline interference. The effective removal of baselines should increase similarity in processed spectra, which is reflected by a tighter PCA cluster [16]. In Figure 8, the convex hulls of data are also highlighted in score plots to illustrate the compactness of PCA clusters. Clearly, the enhanced PLS methods show tighter convex hulls, and this confirms their advantages over classical approaches.

#### 4. Conclusion

Considering the valuable information in the whole set of related spectra, we propose to remove their backgrounds collaboratively. Based on penalized least squares, we improve the background correction methods for a set of correlated Raman spectra by designing two collaborative weighting

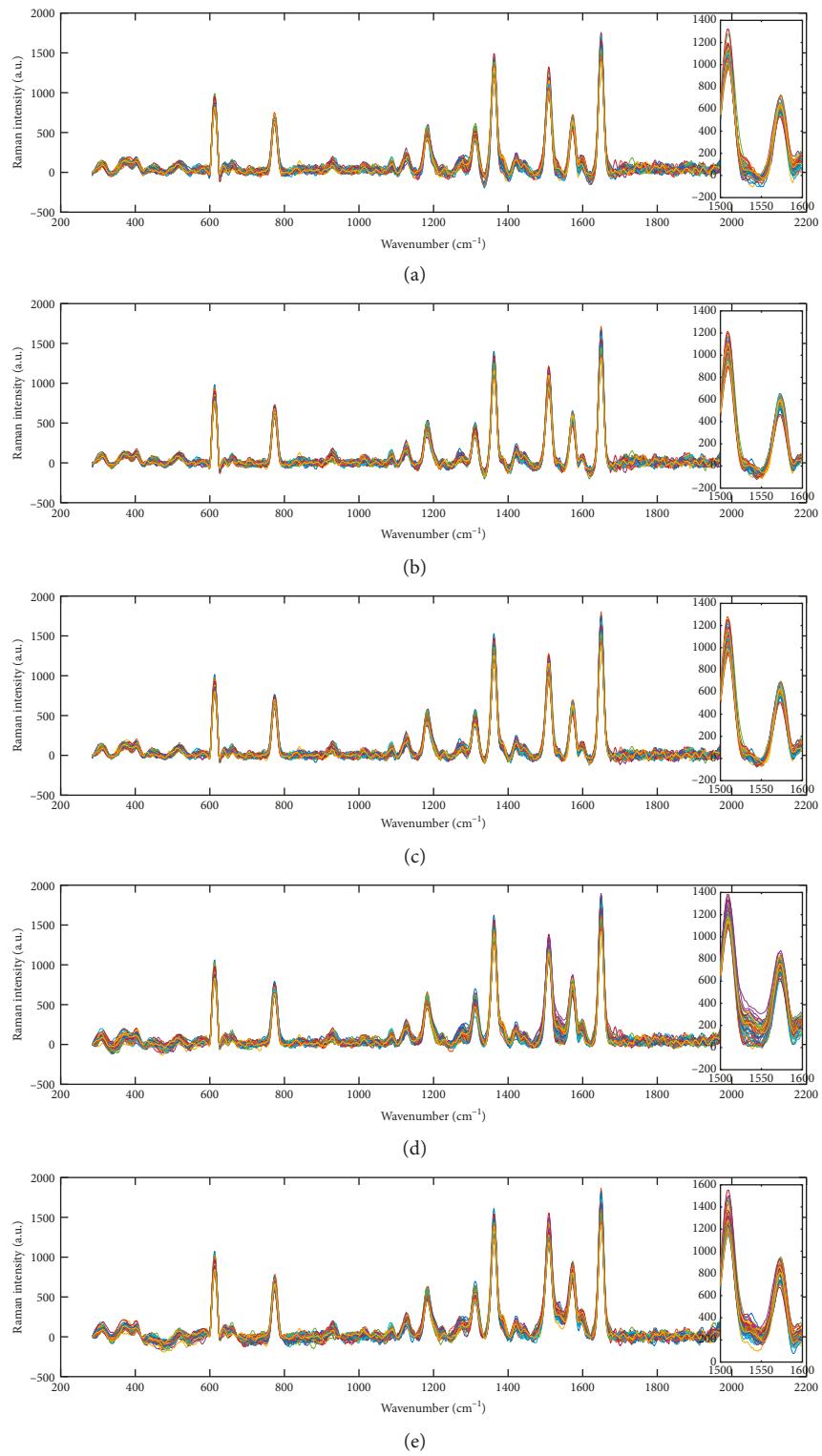


FIGURE 6: Continued.

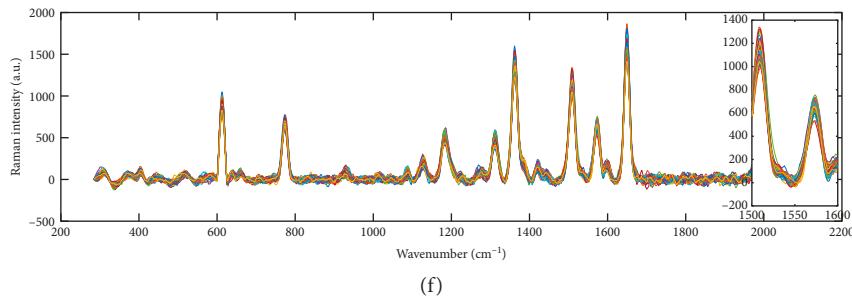


FIGURE 6: Spectra of GIAN with background removed by airPLS (a), AS airPLS (b), CW airPLS (c), MPLS (d), AS MPLS (e), and CW MPLS (f).

TABLE 3: Average intensity variance of background-corrected GIAN data.

Method	Variance
airPLS	1101.02
AS airPLS	<b>880.67</b>
CW airPLS	914.54
MPLS	1635.78
AS MPLS	1530.35
CW MPLS	<b>988.65</b>

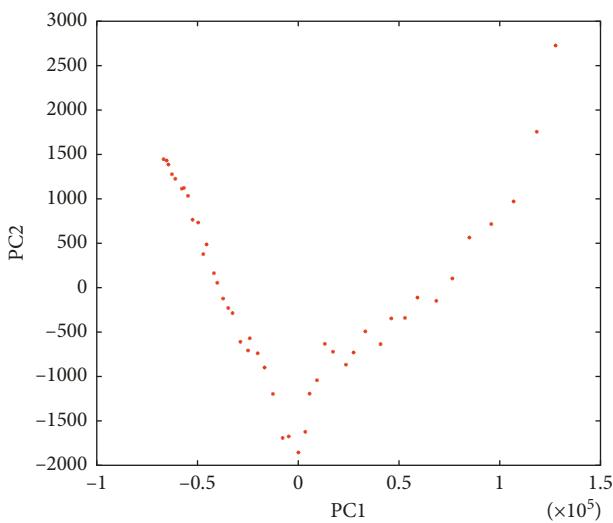


FIGURE 7: Score plot for original spectra.

schemes for background estimations. The average spectrum (AS) method fuses all the considered spectra into an average one to calculate the weights or contributions of different points to the background. The combined weight (CW) method averages the weights derived from all spectra for background estimation. To illustrate the performance of such collaborative approaches on background correction problems, we apply the AS and CW versions of airPLS and MPLS to process simulated and real Raman spectra. The results demonstrate the collaborative approaches are much better than traditional approaches that process spectra individually.

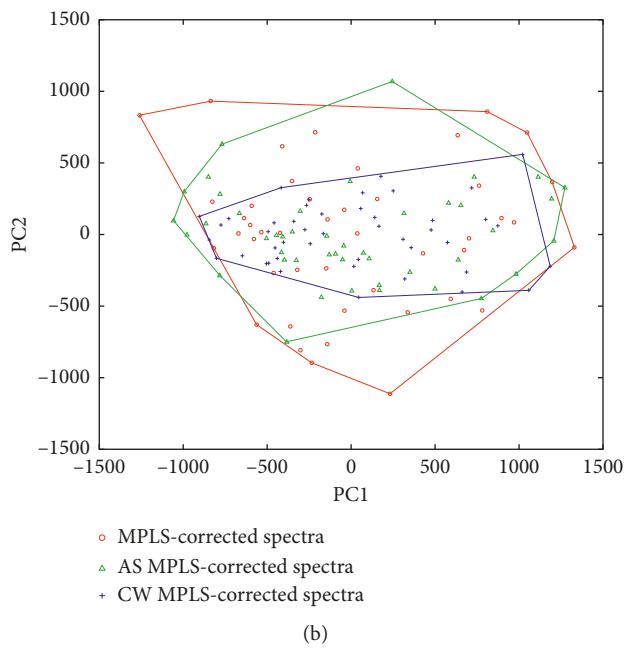
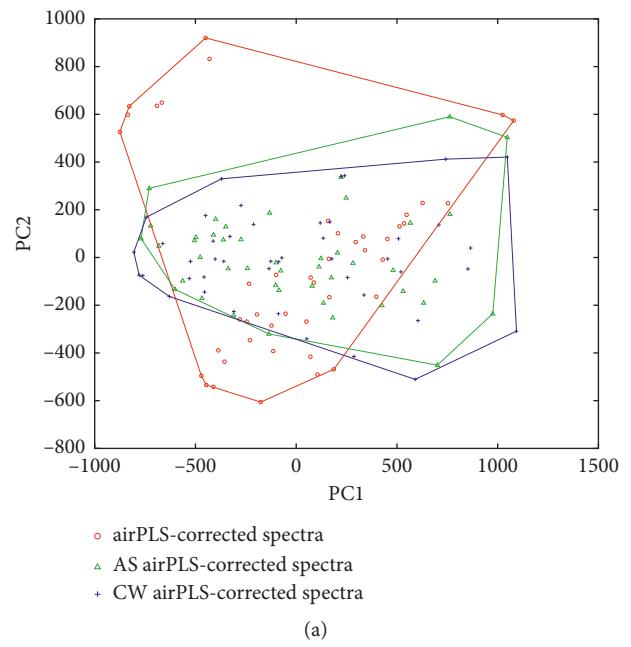


FIGURE 8: (a, b) Score plots for corrected spectra.

## Data Availability

The Raman spectrum data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was financially supported by the Science and Technology Development Fund of Macao S.A.R (067/2014/A and 097/2015/A3), the National Key Basic Research Program of China (no. 2013CB932702), the Research Fund for the Program on National Key Scientific Instruments and Equipment Development of China (no. 2011YQ0301241402), the National Natural Science Foundation of China (nos. 21522501, 21521063, and 61673405), and the Hunan Innovation and Entrepreneurship Program.

## Supplementary Materials

The source code for the proposed methods in this study can be accessed at <http://www.cis.umac.mo/cybernetics/backRemoveCode.zip>. Please download and upzip the file and check the readme.pdf for more information. (*Supplementary Materials*)

## References

- [1] S. S. R. Dasary, P. C. Ray, A. K. Singh, T. Arbneshi, H. Yu, and D. Senapati, "A surface enhanced Raman scattering probe for highly selective and ultra sensitive detection of iodide in water and salt samples," *Analyst*, vol. 138, no. 4, pp. 1195–1203, 2013.
- [2] X. Wu, C. Xu, R. A. Tripp, Y.-W. Huang, and Y. Zhao, "Detection and differentiation of foodborne pathogenic bacteria in mung bean sprouts using field deployable label-free SERS devices," *Analyst*, vol. 138, no. 10, pp. 3005–3012, 2013.
- [3] Y. Zou, L. Chen, Z. Song et al., "Stable and unique graphitic Raman internal standard nanocapsules for surface-enhanced Raman spectroscopy quantitative analysis," *Nano Research*, vol. 9, no. 5, pp. 1418–1425, 2016.
- [4] W. Ilewicz, M. Kowalcuk, M. Niegabitowski, D. Buchczik, and G. Adam, "Comparison of baseline estimation algorithms for chromatographic signals," in *Proceedings of 2015 20th International Conference on Methods and Models in Automation and Robotics (MMAR)*, pp. 925–930, IEEE, Międzyzdroje, Poland, August 2015.
- [5] J. Li, L. Choo-Smith, Z. Tang, and M. G. Sowa, "Background removal from polarized Raman spectra of tooth enamel using the wavelet transform," *Journal of Raman Spectroscopy*, vol. 42, no. 4, pp. 580–585, 2011.
- [6] C. G. Bertinetto and T. Vuorinen, "Automatic baseline recognition for the correction of large sets of spectra using continuous wavelet transform and iterative fitting," *Applied Spectroscopy*, vol. 68, no. 2, pp. 155–164, 2014.
- [7] F. Zhao, J. Wang, and A. Wang, "An improved spectral background subtraction method based on wavelet energy," *Applied Spectroscopy*, vol. 70, no. 12, pp. 1994–2004, 2016.
- [8] D. Chen, Z. Chen, and E. Grant, "Adaptive wavelet transform suppresses background and noise for quantitative analysis by Raman spectrometry," *Analytical and Bioanalytical Chemistry*, vol. 400, no. 2, pp. 625–634, 2011.
- [9] H. Ren, Z. Chen, X. Zhang et al., "Rapid and quantitative determination of S-adenosyl-L-methionine in the fermentation process by surface-enhanced Raman scattering," *Journal of Analytical Methods in Chemistry*, vol. 2016, Article ID 4910630, 6 pages, 2016.
- [10] K. Chen, H. Wei, H. Zhang, T. Wu, and Y. Li, "A Raman peak recognition method based automated fluorescence subtraction algorithm for retrieval of Raman spectra of highly fluorescent samples," *Analytical Methods*, vol. 7, no. 6, pp. 2770–2778, 2015.
- [11] K. Chen, H. Zhang, H. Wei, and Y. Li, "Improved Savitzky–Golay-method-based fluorescence subtraction algorithm for rapid recovery of Raman spectra," *Applied Optics*, vol. 53, no. 24, pp. 5559–5569, 2014.
- [12] H. G. Schulze, R. B. Foist, K. Okuda, A. Ivanov, and R. F. B. Turner, "A model-free, fully automated baseline-removal method for Raman spectra," *Applied Spectroscopy*, vol. 65, no. 1, pp. 75–84, 2011.
- [13] S. He, S. Fang, X. Liu et al., "Investigation of a genetic algorithm based cubic spline smoothing for baseline correction of Raman spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 152, pp. 1–9, 2016.
- [14] C. Gallo, V. Capozzi, M. Lasalvia, and G. Perna, "An algorithm for estimation of background signal of Raman spectra from biological cell samples using polynomial functions of different degrees," *Vibrational Spectroscopy*, vol. 83, pp. 132–137, 2016.
- [15] L. Martin, A. Barcaru, M. J. Sjerps, and G. Vivó-Truyols, "Leveraging probabilistic peak detection to estimate baseline drift in complex chromatographic samples," *Journal of Chromatography A*, vol. 1431, pp. 122–130, 2016.
- [16] Y. Liu, X. Zhou, and Y. Yu, "A concise iterative method using the Bezier technique for baseline construction," *Analyst*, vol. 140, no. 23, pp. 7984–7996, 2015.
- [17] J. Liu, J. Sun, X. Huang, G. Li, and B. Liu, "Goldindec: a novel algorithm for Raman spectrum baseline correction," *Applied Spectroscopy*, vol. 69, no. 7, pp. 834–842, 2015.
- [18] S.-J. Baek, A. Park, Y.-J. Ahn, and J. Choo, "Baseline correction using asymmetrically reweighted penalized least squares smoothing," *Analyst*, vol. 140, no. 1, pp. 250–257, 2015.
- [19] Z.-M. Zhang, S. Chen, and Y.-Z. Liang, "Baseline correction using adaptive iteratively reweighted penalized least squares," *Analyst*, vol. 135, no. 5, pp. 1138–1146, 2010.
- [20] L. Zhong, D.-J. Zhan, J.-J. Wang et al., "Morphological weighted penalized least squares for background correction," *Analyst*, vol. 138, no. 16, pp. 4483–4492, 2013.
- [21] A. T. Weakley, P. R. Griffiths, and D. E. Aston, "Automatic baseline subtraction of vibrational spectra using minima identification and discrimination via adaptive, least-squares thresholding," *Applied Spectroscopy*, vol. 66, no. 5, pp. 519–529, 2012.
- [22] L. G. Johnsen, T. Skov, U. Houlberg, and R. Bro, "An automated method for baseline correction, peak finding and peak grouping in chromatographic data," *Analyst*, vol. 138, no. 12, pp. 3502–3511, 2013.

- [23] Ł. Komsta, "Comparison of several methods of chromatographic baseline removal with a new approach based on quantile regression," *Chromatographia*, vol. 73, no. 7-8, pp. 721–731, 2011.
- [24] K. H. Liland, T. Almøy, and B.-H. Mevik, "Optimal choice of baseline correction for multivariate calibration of spectra," *Applied Spectroscopy*, vol. 64, no. 9, pp. 1007–1016, 2010.
- [25] X. Ning, I. W. Selesnick, and L. Duval, "Chromatogram baseline estimation and denoising using sparsity (beads)," *Chemometrics and Intelligent Laboratory Systems*, vol. 139, pp. 156–167, 2014.
- [26] A. Mani-Varnosfaderani, A. Kanginejad, K. Gilany, and A. Valadkhani, "Estimating complicated baselines in analytical signals using the iterative training of Bayesian regularized artificial neural networks," *Analytica Chimica Acta*, vol. 940, pp. 56–64, 2016.
- [27] R. B. Foist, H. G. Schulze, A. Jirasek, A. Ivanov, and R. F. B. Turner, "A matrix-based two-dimensional regularization algorithm for signal-to-noise ratio enhancement of multidimensional spectral data," *Applied Spectroscopy*, vol. 64, no. 11, pp. 1209–1219, 2010.
- [28] J. Peng, S. Peng, A. Jiang, J. Wei, C. Li, and J. Tan, "Asymmetric least squares for multiple spectra baseline correction," *Analytica Chimica Acta*, vol. 683, no. 1, pp. 63–68, 2010.
- [29] Y. Wu, Q. Gao, and Y. Zhang, "A robust baseline elimination method based on community information," *Digital Signal Processing*, vol. 40, pp. 53–62, 2015.
- [30] E. T. Whittaker, "On a new method of graduation," *Proceedings of the Edinburgh Mathematical Society*, vol. 41, pp. 63–75, 1922.
- [31] P. H. C. Eilers, "A perfect smoother," *Analytical Chemistry*, vol. 75, no. 14, pp. 3631–3636, 2003.
- [32] P. H. C. Eilers and H. F. M. Boelens, "Baseline correction with asymmetric least squares smoothing," Leiden University Medical Centre Report, vol. 1, no. 1, Leiden University Medical Centre, Leiden, Netherlands, 2005.
- [33] H. Ruan and L. K. Dai, "Automated background subtraction algorithm for Raman spectra based on iterative weighted least squares," *Asian Journal of Chemistry*, vol. 23, no. 12, p. 5229, 2011.
- [34] P. J. Cadusch, M. M. Hlaing, S. A. Wade, S. L. McArthur, and P. R. Stoddart, "Improved methods for fluorescence background subtraction from Raman spectra," *Journal of Raman Spectroscopy*, vol. 44, no. 11, pp. 1587–1595, 2013.
- [35] Z.-M. Zhang, S. Chen, Y.-Z. Liang et al., "An intelligent background-correction algorithm for highly fluorescent samples in Raman spectroscopy," *Journal of Raman Spectroscopy*, vol. 41, no. 6, pp. 659–669, 2010.
- [36] X. Bian, Z.-L. Song, Y. Qian et al., "Fabrication of graphene-isolated-Au-nanocrystal nanostructures for multimodal cell imaging and photothermal-enhanced chemotherapy," *Scientific Reports*, vol. 4, no. 1, 2014.

