Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 161 (2016) 58–63

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

An auto-adaptive background subtraction method for Raman spectra

Yi Xie [a], Lidong Yang [a], Xilong Sun [a], Dewen Wu [a], Qizhen Chen [b], Yongming Zeng [b], Guokun Liu [c,*]

[a] Fujian Key Laboratory of Sensing and Computing for Smart City, School of Information Science and Engineering, Xiamen University, China
[b] College of Chemistry and Chemical Engineering, Xiamen University, China
[c] State Key Laboratory of Marine Environmental Science, College of the Environment and Ecology, Xiamen University, Xiamen, Fujian 361102, China

ARTICLE INFO

Article history:
Received 25 November 2015
Received in revised form 6 February 2016
Accepted 22 February 2016
Available online 24 February 2016

Keywords:
Raman spectrum
Background subtraction
Auto-adaptive

ABSTRACT

Background subtraction is a crucial step in the preprocessing of Raman spectrum. Usually, parameter manipulating of the background subtraction method is necessary for the efficient removal of the background, which makes the quality of the spectrum empirically dependent. In order to avoid artificial bias, we proposed an auto-adaptive background subtraction method without parameter adjustment. The main procedure is: (1) select the local minima of spectrum while preserving major peaks, (2) apply an interpolation scheme to estimate background, (3) and design an iteration scheme to improve the adaptability of background subtraction. Both simulated data and Raman spectra have been used to evaluate the proposed method. By comparing the backgrounds obtained from three widely applied methods: the polynomial, the Baek's and the airPLS, the auto-adaptive method meets the demand of practical applications in terms of efficiency and accuracy.

1. Introduction

By providing fingerprint information, Raman spectroscopy has been widely applied in material characterization and identification [1]. However, the Raman signal of a target may be obscured or swamped by various background from the fluorescence of surrounding medium, contaminator [2,3] or the target itself. Under this condition, background subtraction is inevitable to obtain a reliable Raman spectrum for further analysis.

Various strategies have been applied to subtract the background from Raman spectrum. The polynomial method [4], one of the mostly applied methods in commercial Raman software, estimates the background by a strategy of polynomial fitting based on least squares. The degree of polynomial fitting is manually adjusted according to the profile of Raman spectrum [5], and the fitting quality is improved by applying the IasLS method [6] during the procedure of polynomial fitting. Wavelet transform plays a core role in some methods of background subtraction [7–9]. These methods can estimate background when the mother wavelet and the decomposed level of wavelet are determined appropriately [10]. Although the above mentioned methods remove the background from Raman spectrum in an accurate and efficient way, their performances are significantly affected by parameter manipulating. Such inconveniences hinder in some extent the wide application of Raman technique. For example, as a non-specialist user of handheld Raman spectrometers, one prefers to directly obtain reliable data without any complicated data analysis.

In order to meet this demand, the methods based on least squares, such as the airPLS method [11], were proposed to reduce the requirements of users' experiences [12]. By introducing an iteration scheme to adaptively adjust the weight vector for background estimation, the airPLS only needs one adjustable parameter of λ, which is mostly compensated by the iteration scheme. Therefore, the background subtraction obtained by the airPLS is less sensitive to parameter values, in comparison to the above mentioned methods. Furthermore, Bake et al. [13] proposed an automatic method without parameter adjustment, which subtracted background when the spectrum peaks could be quickly detected using the derivation operation. However, the Bake's method lost its accuracy when Raman peaks are overlapped.

By incorporating the merits of the Baek's and airPLS methods, we propose an auto-adaptive background subtraction method free of parameter manipulating. Here the three successively smoothed derivative of raw data is regarded as the derivative of background. The iteration scheme is used to promote the adaptability and overcome the issue of peak overlapping.

We use both simulated data and Raman spectra (from different targets and different conditions of the same target) to evaluate the auto-adaptive method of background subtraction which is compared with the other three reported methods: the polynomial, the Baek's and the airPLS. The proposed method displays higher accuracy than the polynomial and the Baek's, and is slightly less than the airPLS method with its optimal parameter. The experimental results have demonstrated that the auto-adaptive method works well in subtracting background under different conditions, and provides comparable performance to the airPLS, one of the best methods reported for background subtraction.

* Corresponding author.
E-mail address: guokunliu@xmu.edu.cn (G. Liu).

http://dx.doi.org/10.1016/j.saa.2016.02.016
1386-1425/

2. Materials and methods

In general, the intersections of the raw data and its background curve are the local minima of the spectrum without background, shown in Fig. 1(a). Therefore, we can detect the locations of local minima of a spectrum, and appropriately fill data among these minima to estimate the background. The auto-adaptive method performs four main operations to obtain the background: smoothing operation, local minimum detection, interpolation among local minima and iterative procedure.

2.1. Smoothing operation

Smoothing operation is necessary and runs through the proposed method for background subtraction. The Savitzky–Golay filter [14] is selected because it performs a generalized moving average filter which can smoothen out the Raman signal without significantly destroying original characteristics of raw data. The input of raw data is expressed as one vector $x[j], j = 1, \ldots, L_x$, where L_x is the vector length. Using the Savitzky–Golay filter, the smoothed vector $\boldsymbol{y} = smooth(\boldsymbol{x}, 2N + 1)$, is calculated according to Eq. (1) with the span of $2N + 1$. In this smoothing function, the coefficients of filter w_t are derived by the unweight linear least squares fit using a polynomial whose degree is *2*.

$$\begin{aligned} y[j] &= \sum_{t=-N}^{N} w_t x[j-t] \\ &= w_{-N} x[j-N] + \ldots + w_N x[j+N] \end{aligned} \tag{1}$$

2.2. Local minimum detection

Searching the local minima of a Raman spectrum is a critical step for estimating background. In this paper, the curve of raw data is denoted as $s(q)$ where q is the unit of wavenumber. Shown in Eq. (2), $s(q)$ is the sum of Gaussian peaks $g(q)$ and the background $b(q)$, and then the derivative of $s(q)$ is expressed as Eq. (3). Therefore, $g(q)$ can be regarded as the spectrum data without background, whose derivative can be calculated by Eq. (4).

$$s(q) = g(q) + b(q) \tag{2}$$

$$\frac{ds(q)}{dq} = \frac{dg(q)}{dq} + \frac{db(q)}{dq} \tag{3}$$

$$\boldsymbol{dg} \triangleq \frac{dg(q)}{dq} = \frac{ds(q)}{dq} - \frac{db(q)}{dq} \tag{4}$$

Here, $\frac{ds(q)}{dq}$ can be approximated by Eq. (5) since that the raw data of spectrum is discrete and denoted as a vector $\boldsymbol{s}$, where l is the index of sample which corresponds to q. In order to reduce the influence of noise, we apply the smoothing operation twice, before and after $diff(\boldsymbol{s})$, shown in Eq. (6). $\frac{db(q)}{dq}$ can be calculated as Eq. (7) since that the first derivative of background is similar to three successively smoothed derivative of raw data (denoted as $\boldsymbol{s}3\boldsymbol{ds}$) [13]. The spans of Savitzky–Golay filter L_n and L_b are set as appropriate values.

$$\boldsymbol{ds} \triangleq \frac{ds(q)}{dq} \approx \mathrm{diff}(\boldsymbol{s}) = s[l] - s[l-1] \tag{5}$$

$$\boldsymbol{ds} = smooth(diff(smooth(\boldsymbol{s}, L_n)), L_n) \tag{6}$$

$$\frac{db(q)}{dq} \triangleq \boldsymbol{s}3\boldsymbol{ds} = smooth(smooth(smooth(\boldsymbol{ds}, L_b), L_b), L_b) \tag{7}$$

Therefore, the first derivative $\boldsymbol{dg}$ is obtained on the basis of Eqs. (4), (6) and (7). Shown in Fig. 1(b), the local minima of $g(q)$ exist where the sign of $\boldsymbol{dg}$ changes from negative to positive. Then

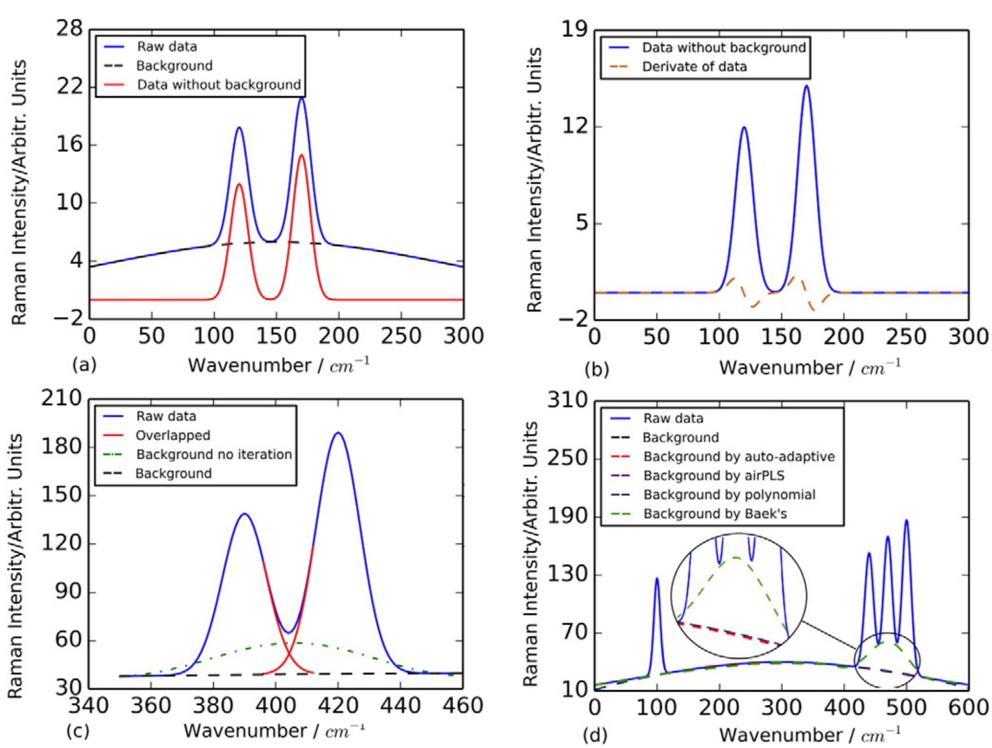

Fig. 1. (a) The local minima of simulated data. (b) Simulated data and its first derivative. (c) Overlap peaks in simulated data. (d) Estimated background by four methods.