Hindawi Publishing Corporation Spectroscopy: An International Journal Volume 27 (2012), Issue 5-6, Pages 311–317 doi:10.1155/2012/530791

Background Estimation of Biomedical Raman Spectra Using a Geometric Approach

Nikolaos Kourkoumelis,¹ Alexandros Polymeros,² and Margaret Tzaphlidou¹

¹Department of Medical Physics, Medical School, University of Ioannina, Ioannina 45110, Greece ²Department of Physics, University of Ioannina, Ioannina 45110, Greece

Correspondence should be addressed to Nikolaos Kourkoumelis, nkourkou@cc.uoi.gr

Copyright © 2012 Nikolaos Kourkoumelis 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.

Abstract. Raman spectroscopy grows into an essential tool for biomedical applications. Nevertheless, the weak Raman signal associated mainly with biological samples is often obscured by a broad background signal due to the intrinsic fluorescence of the organic molecules present, making further analysis unfeasible. A computational geometry method based on the definition of convex hull is described to estimate the background from Raman spectra of samples with biological interest. The method is semiautomated requiring sample-dependent user intervention. It does not depend, however, on curve fitting, requires no information about background distribution or source, and keeps the original spectral data intact.

Keywords: Raman spectroscopy, background estimation, computational geometry

1. Introduction

Raman spectroscopy has been extensively applied in recent years in a variety of biological research ranging from the in situ tissue diagnosis to the analysis of subcellular components. Being a vibrational spectroscopic technique based on inelastic scattering, Raman spectroscopy provides rich molecular information about the chemical composition of samples and exhibits high sensitivity to minute biochemical changes. Furthermore, it is attractive for biomedical studies since it is intrinsically nonintrusive and does not require external labels. The positions and relative intensities of the Raman bands are the basic spectral characteristics for exploring the structure and the function of several biological molecules. This interpretation, however, is often hindered by the broad background signal mostly due to fluorescence from organic molecules and contaminants. The intensity of fluorescent is usually much higher than the weak Raman signal in biological samples, and therefore the subtraction of background is an essential process to extract reliable analytical information from biomedical Raman data.

Apart from instrumental specific design approaches, a number of computational methods have been proposed for background removal from Raman spectra. These methods include polynomial fitting [1–7], first- and second-order differentiation [8, 9], wavelet transformation [10–15], frequency-domain filtering [16], and principal component analysis (PCA) [17]. All of the above methods have certain strengths and drawbacks depending on the problem they are trying to deal with. For example, low-order



Figure 1: Simulated spectrum with curved background and random noise. (a) Fourier filtering, (b) convex optimization, (c) background estimation.

polynomial fitting is suitable for spectra with broad background but it is not effective for biological samples which feature Raman spectra with several adjacent, not readily obvious, peaks. Higher-order polynomials may be susceptible to data over fitting [2]. Differentiation may also distort peak shapes and therefore creates an inconsistent spectrum compared to the preprocessed one [1]. Wavelets analysis, which is the Fourier transform analog for localized functions, is a promising solution although the transformation of the signal into predetermined frequency bands may cause distortion in some part of the spectra [15].

In the present study, we describe a novel semiautomated background removal method which is based on the geometric definition of convex hull [18]. The effectiveness of the method is demonstrated through theoretical and experimental biomedical Raman spectra.

2. Theoretical Background

The signal, *S*, is assumed as a composition of a low frequency component (B(x), background) and the true information, P(x), so that $S = P(x) \times B(x)$. The background is the slow part of the composite signal and resides in the vicinity of the low frequencies range. With the application of low-pass filtering, we extract from the complex signal a rough estimate of the true background component. The first step works by applying a Fourier transform to the signal, that is, decreasing the high frequency components and inverse transforming the result. In this way, we have managed to break up the signal into a superposition



Figure 2: Raman spectrum of paracetamol. (a) Fourier filtering, (b) convex optimization, (c) background estimation.

of infinitely many sinusoids. Each sinusoid can be manipulated individually and then recombined to obtain an approximation to the original periodic function [19]. The second step is to decompose the signal to parts which have the characteristics to be convex sets. This is accomplished by taking regions from peaks to valleys, of the previously filtered signal, via a simple pattern search of a table consisting of 0 and 1 referring to the slope of the signal. A convex hulling minimization routine supplies the single optimal solution for all sets [18] and is able to extract the true background part of the region by introducing a new parameter "median." The latter is a line segment calculated from the statistical data average and by definition is constructed to divide the convex region into two parts. All points with values higher than the median are part of the upper part of convex hull and represent the peaks, while the remaining points represent the true background. The only remaining problem is the continuity of one convex region in respect with the previous or the next one (it is already continuous in the interior of its domain). The simplest approach is by defining user variable (joins) which controls the number of linking points of the lower part of the convex region which must be included in the final background array. The outcome captures every essential feature of the background component through a purely geometric semiautomated procedure. Due to its high point of reduction degree, the signal is suitable for subsequent polynomial interpolation, smoothing, and so forth.



Figure 3: Raman spectrum of PAT. (a) Fourier filtering, (b) convex optimization, (c) background estimation.

3. Materials and Methods

The algorithm was implemented in Mathematica software package (Wolfram Research). For signals sampled at discrete intervals, as in our case, Mathematica uses the discrete Fourier transform [20]. Raman spectra were chosen from literature for comparison purposes. Simulated data is identical to that from [15], while experimental data was acquired with permission from [15] and the hyperSpec project (http://hyperspec.r-forge.r-project.org/).

4. Results and Discussion

Simulated spectrum consisting of three Gaussian peaks with curved background and random noise is shown in Figure 1.

As previously discussed, the first step, (a), is the low-pass filtering, the second step, (b), is finding and optimizing the convex sets, and the last one, (c), is joining the convex sets in a continuous manner. In the case of simulated data, the performance of the algorithm is flawless. Figures 2–4 depict the experimental Raman spectra of paracetamol, prednisone acetate tablets (PATs), and chondrocytes in cartilage, respectively.



Figure 4: Raman spectrum of chondrocytes in cartilage. (a) Fourier filtering, (b) convex optimization, (c) background estimation.

It is evident that the more complicated the signal is, the more Fourier components are needed to approximate the experimental baseline curve. A rough approximation, however, is adequate even for complicated spectra with several bands (Figure 2). In all cases, the background is clearly defined and the signal which does not belong to peak areas is efficiently diminished. Since the Fourier transformation is not applied for smoothing but for extracting the geometric characteristics, the signal retains all its original features avoiding distortions. Nevertheless, in some spectra with low S/N ratio, this may result in negative peaks in the background estimation procedure (circle in Figure 3(c)) due to the calculation methodology of the "median" which does not take into consideration the local slope of the signal but the whole one. A fitting procedure of the data within each convex region will immediately remove such artifacts. However, we did not introduce this computationally intensive improvement because (i) negative peaks appeared only once in our test cases and (ii) we tried to keep the method simple and purely geometric.

5. Conclusions

A computational geometry method for the estimation of the Raman background signal of highly fluorescent samples has been described in this study. Background subtraction was achieved in all cases while the peaks were preserved. The proposed algorithm is semiautomated and requires user input for two variables which define the degree of the Fourier series approximation and the connection of the convex sets. The method is valid for all signals which are convex, that is, one-directional, and, as such, it can be possibly applied to other spectroscopic techniques as well as X-ray powder diffractograms. Preliminary results confirm its wide applicability across diverse spectroscopic data.

Acknowledgments

The authors thank Dr. Zhi-Min Zhang and Dr. Claudia Beleites for providing the raw data considered in this paper.

References

- [1] C. A. Lieber and A. Mahadevan-Jansen, "Automated method for subtraction of fluorescence from biological Raman spectra," *Applied Spectroscopy*, vol. 57, no. 11, pp. 1363–1367, 2003.
- [2] B. D. Beier and A. J. Berger, "Method for automated background subtraction from Raman spectra containing known contaminants," *Analyst*, vol. 134, no. 6, pp. 1198–1202, 2009.
- [3] A. Jirasek, G. Schulze, M. M. L. Yu, M. W. Blades, and R. F. B. Turner, "Accuracy and precision of manual baseline determination," *Applied Spectroscopy*, vol. 58, no. 12, pp. 1488–1499, 2004.
- [4] V. Mazet, C. Carteret, D. Brie, J. Idier, and B. Humbert, "Background removal from spectra by designing and minimising a non-quadratic cost function," *Chemometrics and Intelligent Laboratory Systems*, vol. 76, no. 2, pp. 121–133, 2005.
- [5] M. N. Leger and A. G. Ryder, "Comparison of derivative preprocessing and automated polynomial baseline correction method for classification and quantification of narcotics in solid mixtures," *Applied Spectroscopy*, vol. 60, no. 2, pp. 182–193, 2006.
- [6] J. Zhao, H. Lui, D. I. Mclean, and H. Zeng, "Automated autofluorescence background subtraction algorithm for biomedical raman spectroscopy," *Applied Spectroscopy*, vol. 61, no. 11, pp. 1225– 1232, 2007.
- [7] T. J. Vickers, R. E. Wambles, and C. K. Mann, "Curve fitting and linearity: data processing in Raman spectroscopy," *Applied Spectroscopy*, vol. 55, no. 4, pp. 389–393, 2001.
- [8] D. Zhang and D. Ben-Amotz, "Enhanced chemical classification of raman images in the presence of strong fluorescence interference," *Applied Spectroscopy*, vol. 54, no. 9, pp. 1379–1383, 2000.
- [9] A. O'Grady, A. C. Dennis, D. Denvir, J. J. McGarvey, and S. E. J. Bell, "Quantitative Raman spectroscopy of highly fluorescent samples using pseudosecond derivatives and multivariate analysis," *Analytical Chemistry*, vol. 73, no. 9, pp. 2058–2065, 2001.
- [10] T. T. Cai, D. Zhang, and D. Ben-Amotz, "Enhanced chemical classification of Raman images using multiresolution wavelet transformation," *Applied Spectroscopy*, vol. 55, no. 9, pp. 1124– 1130, 2001.
- [11] C. Camerlingo, F. Zenone, G. M. Gaeta, R. Riccio, and M. Lepore, "Wavelet data processing of micro-Raman spectra of biological samples," *Measurement Science and Technology*, vol. 17, no. 2, pp. 298–303, 2006.
- [12] Y. Hu, T. Jiang, A. Shen, W. Li, X. Wang, and J. Hu, "A background elimination method based on wavelet transform for Raman spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 85, no. 1, pp. 94–101, 2007.
- [13] P. M. Ramos and I. Ruisánchez, "Noise and background removal in Raman spectra of ancient pigments using wavelet transform," *Journal of Raman Spectroscopy*, vol. 36, no. 9, pp. 848–856, 2005.

- [14] J. Li, L. P. 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.
- [15] 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.
- [16] P. A. Mosier-Boss, S. H. Lieberman, and R. Newbery, "Fluorescence rejection in Ramanspectroscopy by shifted-Spectra, edge-detection, and FFT filtering techniques," *Applied Spectroscopy*, vol. 49, pp. 630–638, 1995.
- [17] T. Hasegawa, J. Nishijo, and J. Umemura, "Separation of Raman spectra from fluorescence emission background by principal component analysis," *Chemical Physics Letters*, vol. 317, no. 6, pp. 642–646, 2000.
- [18] E. W. Weisstein, Convex Hull, MathWorld-A Wolfram Web Resource, http://mathworld.wolfram .com/ConvexHull.html.
- [19] E. W. Weisstein, Fourier Series, MathWorld-A Wolfram Web Resource, http://mathworld.wolfram .com/FourierSeries.html/ConvexHull.html.
- [20] E. W. Weisstein, Discrete Fourier Transform, MathWorld-A Wolfram Web Resource, http://mathworld.wolfram.com/DiscreteFourierTransform.html.



International Journal of Medicinal Chemistry



Organic Chemistry International





International Journal of Analytical Chemistry



Advances in Physical Chemistry



Research International

Catalysts



