Denoising of spectra by adaptive multiwindow model fitting.

S. Charonov

Horiba France SAS, 455 Avenue Eugène Avinée, 59120 Loos and France
E-mail : serguei.charonov@horiba.com

Abstract: A method for noise reduction of spectra based on fitting a multi-window model is presented. The spectrum is modeled as the sum of the polynomial background and Lorentzian peaks. This model applies to all points in the spectrum and to all window sizes. An iterative algorithm is used for fitting. Based on the initial data, the background calculated by the direct least squares method is subtracted. Positive data values are inverted using the 1/x function and the same procedure is used to fit the Lorentzian peaks. The weighted sum of all windows fit containing the point to be processed is used as the result. The weighting factors are calculated by evaluating the quality of the fit. The performance of the presented method is compared with the Savitsky-Golay method and the wavelet noise reduction method. The proposed approach provides good noise reduction performance without using user-entered parameters.

Keywords: Denoising; Savitzky-Golay filter; polynomial fitting

---

Introduction

Many methods have been proposed for noise reduction of spectroscopic data. One of the most popular is the Savitsky-Golay (SG) filter (1) with many modifications (2-4). This filter is based on fitting a polynomial in a moving window. The degree of the polynomial and the size of the window determine the quality of the smoothing. Wavelet smoothing (5) is also widely used, but it requires the choice of many parameters Many other methods have been invented, such as Wiener estimation (6), vector casting (7), Artificial Neural Network approach (8) etc.

The proposed method uses the approach presented in the author's previous work (9). The main difference is the use of a two-component model versus a one-polynomial one. This model matches the nature of the data better, resulting in a better quality fit.

The performance of the proposed method has been compared with the SG, wavelets denoising methods and with one component model on the simulated and real data. Method is simple for implementation and does not required any preparation steps as learning, spike removing, baseline correction and do not need input parameters. It provides good results for the different noise level.

Algorithms

The fitting algorithm is iterative. For each iteration, the following sequence is executed

$$\begin{aligned} S_n &= S_{n-1} + B_{n-1} \\ B_n &= \mathrm{Poly}(S_n) \\ S_n &= S_n - B_n \end{aligned} \qquad (1)$$

$$S_n = S_n + L_{n-1}$$
$$L_n = 1/\mathrm{Poly}(1/S_n \text{ for all } S_n > Thr)$$
$$S_n = S_n - L_n$$

Where S is the original spectrum, B is the background, L is the Lorentz peak, Poly() is the polynomial approximation. Using only points above the Thr threshold will reduce noise impact. Figure 1 shows the first 4 iterations of this algorithm. In this work, the algorithm stops after a fixed number of iterations equal to 5.

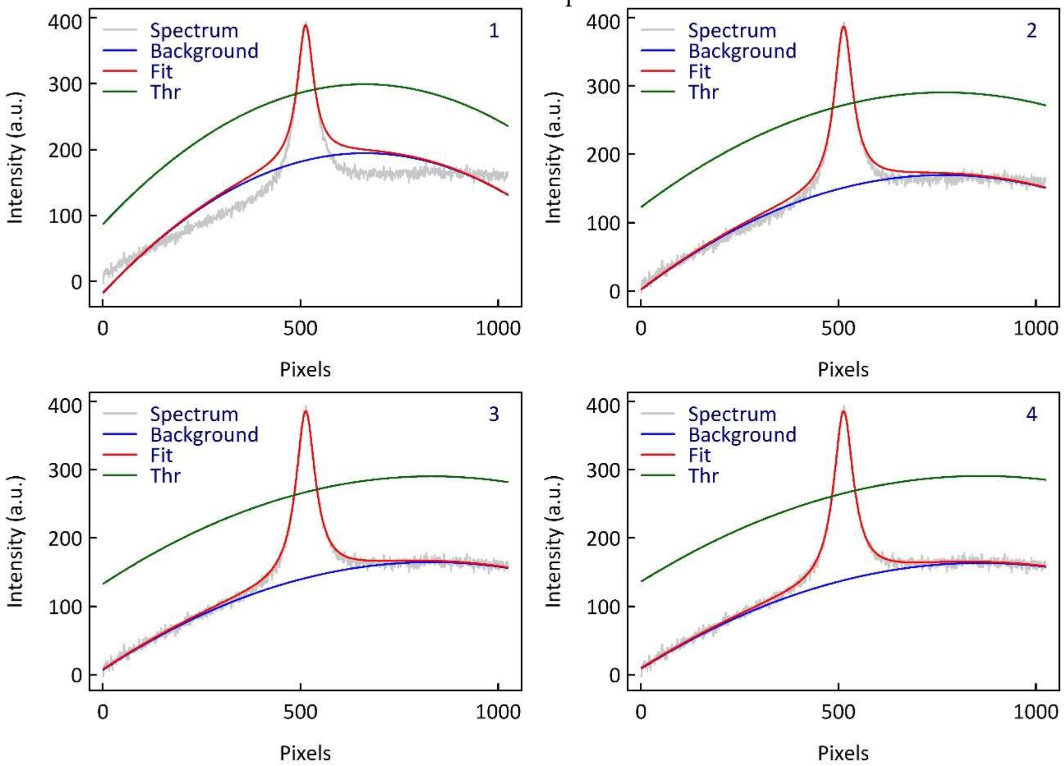

Figure 1. Fitting algorithm iterations.

The denoising algorithm includes the following steps. For each window size and for each window position:

A two-component fitting procedure is applied.

A weighting coefficient is calculated based on the quality of the fit.

For each point in the input data, the filter value is computed as the weighted sum of the fits of all windows that include the given point.

$$Result = \frac{\sum F_{ij} \cdot W_{ij}}{\sum W_{ij}} \quad (2)$$

where i and j are the position and size of the window. Window size starts at polynomial order+2. Figure 2 shows this process

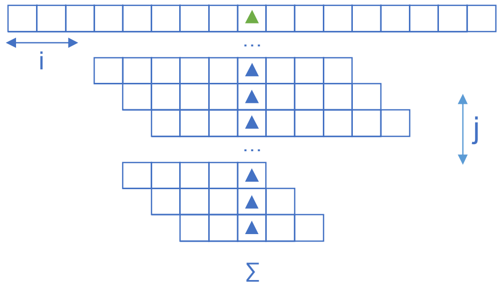

Figure 2. Filtering result (upper triangle) as a weighted sum of fitted values (other triangles) of windows including the processed point.

The weighting coefficients are calculated by evaluating the quality of the fit. Compared to previous work, this article uses the following function.

For each point of the input data and for all windows, the center of which is this point, the following condition is checked:

$$|Fit\text{ - }Data| < NT \tag{3}$$

where NT is noise threshold. Windows are scanned in decreasing order from maximum to minimum size. The process stops at the first window that satisfies condition. The weighting coefficients are defined as

$$W_{ij} = \begin{cases} S_{ij}, if\ j \geq jmax \\ 0, if\ j < jmax \end{cases} \tag{4}$$

Where S is inverted sigmoid function

$$S = \frac{1}{1+e^{Diff}} \tag{5}$$

$$Diff = \frac{Max\left((Fit-Data)^2\right)}{NT^2} \tag{6}$$

The following procedure is used to calculate the noise parameters of the weighting function. The SG algorithm with a small window is applied to the data. The absolute difference values between the original and filtered data are sorted in ascending order. The mean value of the initial part of the sequence is calculated. This value, which is proportional to the noise variance (10), is used to characterize noise.

$$NC = \frac{\sum Sort(|Data- \quad (Window,Order)|,Part)}{DataSize\cdot Part} \tag{7}$$

where *Sort (x, y)* - sort the x values in ascending order and keep only y part of them and *SG (x, y)* - the SG filter with window size x and polynomial order y. In this work, the following values are used: *Window* = 5, *Order* = 2, *Part* = 0.25. From model experiments, it can be concluded that the noise value is stable over a wide range of data and does not depend on peaks and baseline types. Using only a fraction of the points minimizes the influence of spike noise and other artifacts. The noise parameters for the weighting functions obtained from experiments with simulated spectral data are $NT = 21\cdot NC$.

Results and Discussion

Simulated and real data were used to validate the proposed method. Simulation data includes two types of samples. The first sample is based on a random set of Lorentz peaks with a width of 1 to 10 points. The baseline is modeled by the sum of 3 Gaussian/Lorentz functions with a width of 100 to 300 points. The second sample is one of the standard artificial signal testing functions. Figure 3 shows this data. Gaussian noise is used to contaminate data. The signal-to-noise ratio SNR in decibels is calculated using the formula

$$SNR = 10log_{10}\left(\frac{\sum Signal^2}{\sum Noise^2}\right) \quad (8)$$

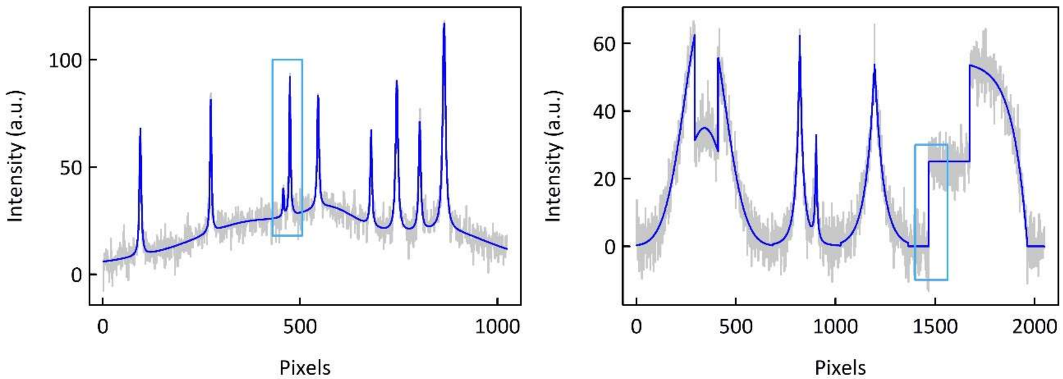

Figure 3. Simulated spectrum(left) and artificial signal(right). Gray color is used for noisy data. SNR is 15 dB. The rectangle marks the area shown in Figure 7.

The proposed method, the SG filter and the wavelet filter were used to process data with SNR levels from 0 to 30 dB. The window size in the SG method varied from 5 to 51, with polynomial order 2. The MatLab application was used for the wavelet filter (Sym4 type and Empirical Bayes method) and the decomposition level was from 1 to 9. The window size of the SG filter and the wavelet filter decomposition level were optimized for achieving the best SNR result. Figure 4 shows the SNR of the cleaned data as a function of the SNR of the noisy data. Figures 5 illustrate the filtering result for an SNR of 15 dB. The sizes of the SG window are 11 and 29, and the wavelet decomposition level is 5 and 6. Figures 6 show the processing of the selected data areas marked in Figure 3. The results show the best performance of the proposed approach, especially at SNR above 10 dB.

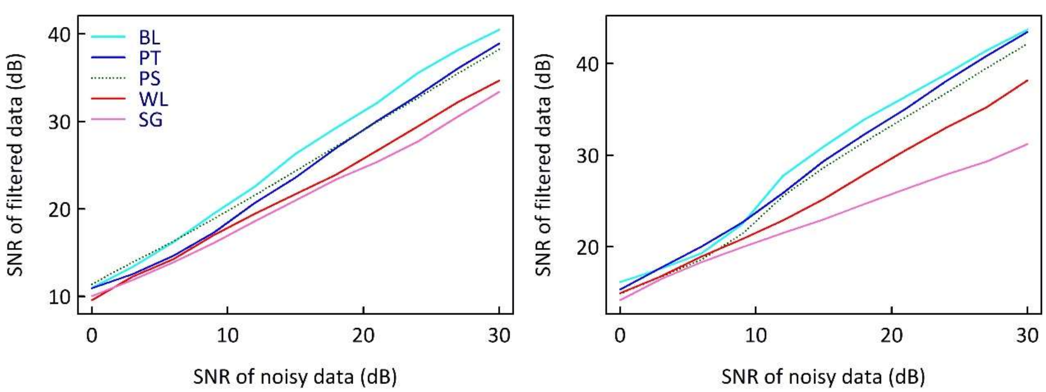

Figure 4. SNR of filtered data as a function of SNR of noisy data for the simulated spectrum(left) and the artificial signal(right). The order of the items in the legend corresponds to the order of the curves. BL - the proposed method, used background and Lorentzian peak fit, PT and PS - the single component methods with threshold and sigmoid evaluation functions(9), WL - wavelet filter, SG - Savitsky-Golay filter.

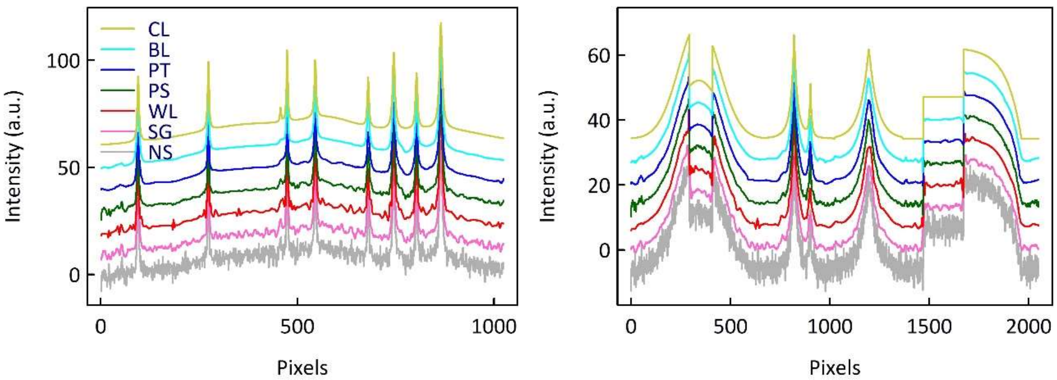

Figure 5. Filtering result for data with 15 dB noise level. CL - original data, BL - the proposed method, PT and PS - the single component methods with threshold and sigmoid estimation functions, WL - wavelet filter, SG - Savitsky-Golay filter, NS - noisy data. The order of the items in the legend corresponds to the order of the curves. Graphs are shifted only for better visibility.

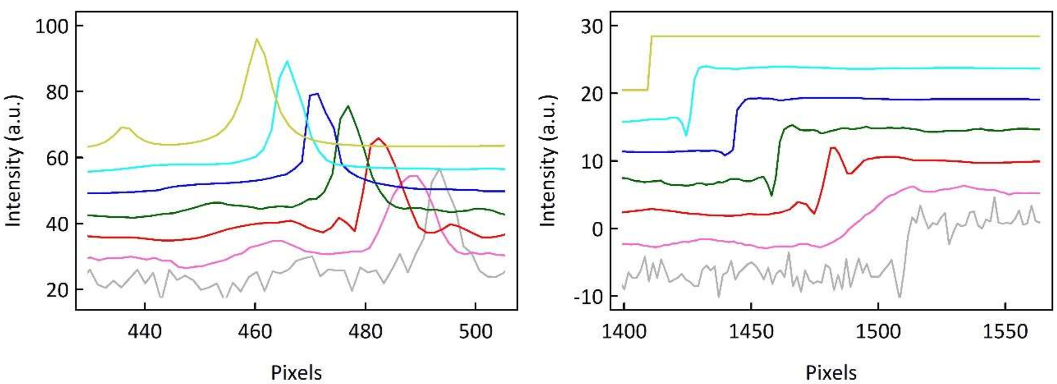

Figure 6. The result of filtering data with a noise level of 15 dB for the zones selected in Figure 3. The legends are the same as in Figure 5.

The result shows the best performance of the proposed approach.

Raman spectra of mineral samples (11) were used to test noise reduction techniques on real data. Figure 7 shows the processing of 2 spectra with different noise levels. The SG and wavelet parameters are the same as for the previous 15 dB simulated data. The results demonstrate a good visual quality of the data, processed by the proposed method.

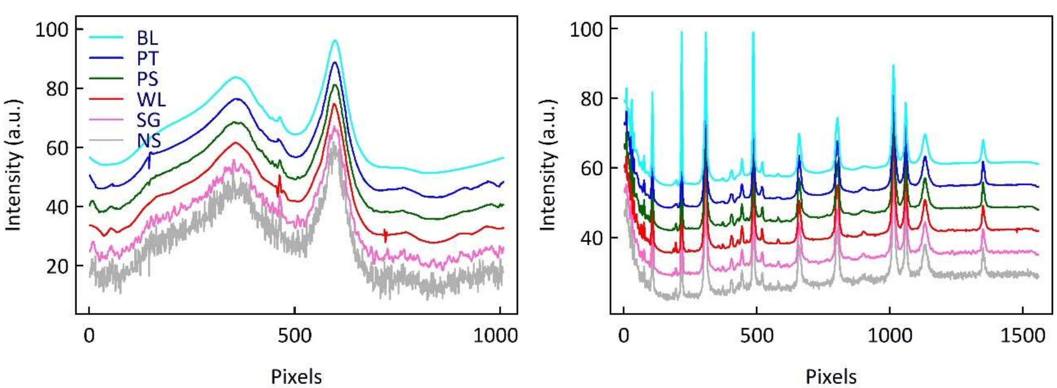

Figure 7. Processing of Raman spectra.

Conclusion

The proposed method shows good results on simulated and real data. It does not require user-entered parameters. Processing can be applied directly without preparation steps such as baseline correction, learning, spike removal, etc. The algorithms are very simple to implement and can be optimized in various ways. It can also be used to process 2D and 3D images and other multidimensional data.

Supplemental Material: The online version of the method implementation is available at (12). Datasets can be downloaded from (13).

References

1. A. Savitzky, M.J.E. Golay. "Smoothing and differentiation of data with simplified least squares procedures". Anal. Chem. 1964. 36: 1627–1639. doi: 10.1021/ac60214a047
2. M Browne, N Mayer, TRH Cutmore. "A multiscale polynomial filter for adaptive smoothing". Digital Signal Processing. 2007. 17: 69–75. doi:10.1016/j.dsp.2006.01.006
3. M Niedźwiecki, M Ciołek. "Fully Adaptive Savitzky-Golay Type Smoothers". 27th European Signal Processing Conference (EUSIPCO), 2019. doi:10.23919/EUSIPCO.2019.8902652
4. J. Dombi, A. Dineva. "Adaptive Multi-round Smoothing Based on the Savitzky-Golay Filter". In: V. Balas, L. Jain L, M. Balas, editors. Soft Computing Applications. SOFA 2016. Advances in Intelligent Systems and Computing, vol 633. Springer, Cham. doi:10.1007/978-3-319-62521-8_38
5. X. Yang, L. Yuee, D. Zhizhen, L. Yang, "A Novel Pre-Processing Algorithm Based on the Wavelet Transform for Raman Spectrum," Appl. Spectrosc. 2018. 72:1752-1763. : doi:10.1177/0003702818789695
6. Y Bai, Q Liu. "Denoising Raman spectra by Wiener estimation with a numerical calibration dataset". Biomed. Opt. Express. 2020. 11(1): 200-214. doi:10.1364/BOE.11.000200.
7. M.Gebrekidan, C. Knipfer, A. Braeue "Vector casting for noise reduction". J Raman Spectrosc. 2020. 51:731–743. doi:10.1002/jrs.5835
8. L. Pan, C. Daengngam, P. Pipitsunthonsan, P. Zhang, A. Booranawong, M. Chongcheawchamnan, "Noise Reduction Technique for Raman Spectrum using Deep Learning Network", arXiv:2009.04067, 2020.
9. Charonov, S. Denoising of Spectra by Adaptive Multiwindow Polynomial Fitting. Preprints 2021, 2021080183
10. Charonov, S. Estimation of Gaussian Noise in Spectra by the Selective Polynomial Fit. Preprints 2021, 2021080098
11. Handbook of Raman Spectra for geology. http://www.geologie-lyon.fr/Raman/
12. Online version of an adaptive multiwindow model filter. https://spectralmultiplatform.blogspot.com/p/math3denoiser.html
13. Data sets. https://doi.org/10.6084/m9.figshare.16350654