Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021

Denoising of spectra by adaptive multiwindow polynomial fitting

S. Charonov, Horiba France SAS, 455 Avenue Eugène Avinée, 59120 Loos, France

E-mail: serguei.charonov@horiba.com

Abstract

A method for noise reduction of spectra based on the adaptive application of the Savitsky-Golay polynomial filter is presented. A polynomial approximation is calculated at all points of the spectrum and for all window sizes. The weighted sum of all polynomials containing the point to be processed is used as the result. The weighting factors are calculated by evaluating the quality of the fit. This paper proposes two evaluation functions. 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 performance of the proposed method has been compared with the SG and wavelets denoising methods 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 denoising algorithm includes the following steps. For each window size and for each window position:

A polynomial is fit by a least-squares.

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 polynomials of all windows that include the given point.

© 2021 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021

$$Result = \frac{\sum P_{ij} \cdot W_{ij}}{\sum W_{ij}} \quad (1)$$

where i and j are the position and size of the window. Window size starts at polynomial order+2. Figure 1 shows this process.

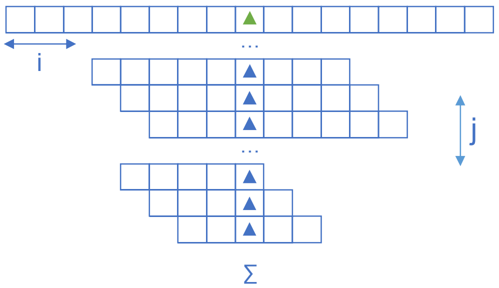


Figure 1. Filtering result (upper triangle) as a weighted sum of polynomial values (other triangles) of windows including the processed point.

The weighting coefficients are calculated by evaluating the quality of the fit. In this paper, two evaluation functions are used.

The first function is a threshold type. For each point of the input data and for all odd windows, the center of which is this point, the following conditions are checked:

$$|Fit - Data| < NExt \quad (2)$$

and

$$|Z| < Thr \quad (3)$$

where

$$Z_k = Z_{k-1} + Signt(Fit\text{-}Data,\ NInt),\ Z_0 = 0 \quad (4)$$

$$Signt(x,y) = \begin{cases} -1, if\ x < 0\ and\ |x| > y \\ 0, if\ |x| \le y \\ 1, if\ x > 0\ and\ |x| > y \end{cases} \quad (5)$$

NInt and NExt are internal and external noise thresholds. Windows are scanned in decreasing order from maximum to minimum. The process stops at the first window that satisfies both conditions. The weighting coefficients are defined as

$$W_{ij} = \begin{cases} 1, if\ j = jmax \\ 0, if\ j \ne jmax \end{cases} \quad (6)$$

Figure 2 illustrates this calculation and shows the use of the second condition to preserve fine data details, usually associated with small peaks.

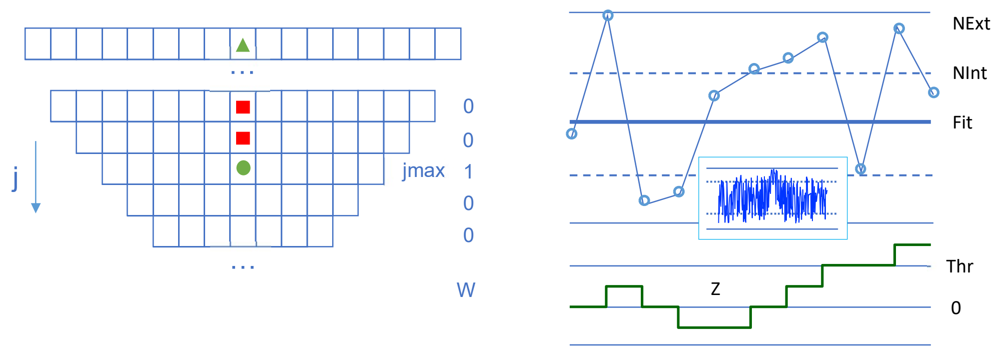


Figure 2. Calculation of the weight coefficients(left). The process stops at the jmax window. W indicates coefficients values. Calculation of the Z function(right). The central rectangle shows a small peak that satisfies the first condition, but not the second.

The second evaluation function uses the maximum absolute difference between the fitted curve and the original curve as a criterion. This value divided by the noise factor is used as an input for the inverted sigmoid function

$$W = \frac{1}{1+e^{Diff}} \tag{6}$$

$$Diff = \frac{Max\left((Fit-Data)^2\right)}{NF^2} \tag{7}$$

The following procedure is used to calculate the noise parameters of the weighting function (9). 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, is used to characterize noise.

$$NC = \frac{\sum Sort(|Data-SG(Window,Order)|,Part)}{DataSize \cdot Part} \tag{8}$$

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 $NF = 12.5\ NC$, $NExt = 26\ NC$, $NInt = 0.3\ NExt$ and $Thr = 4$.

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 Gaussian/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

$$SNR \ = 10log_{10}\left(\frac{\sum Signal^2}{\sum Noise^2}\right) \tag{9}$$

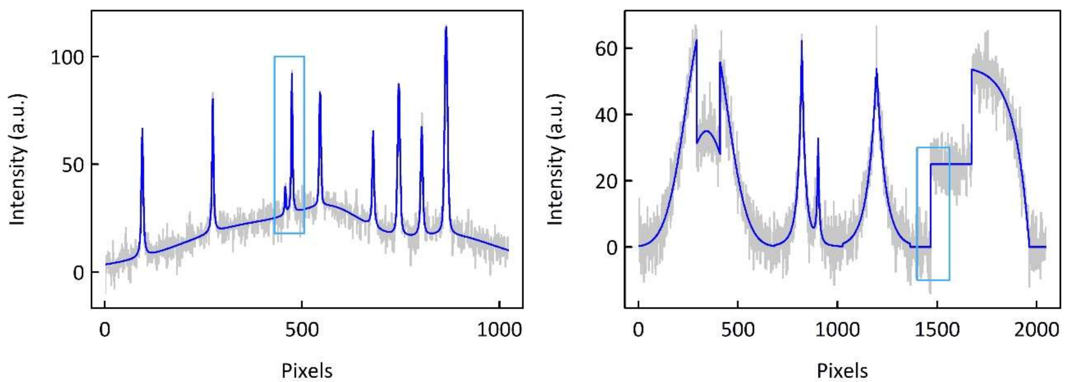


Figure 3. Simulated spectrum(left) and artificial signal(right). Gray color is used for noisy data. SNR is 15 dB. The rectangle marks the area shown in Figure 7.

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.

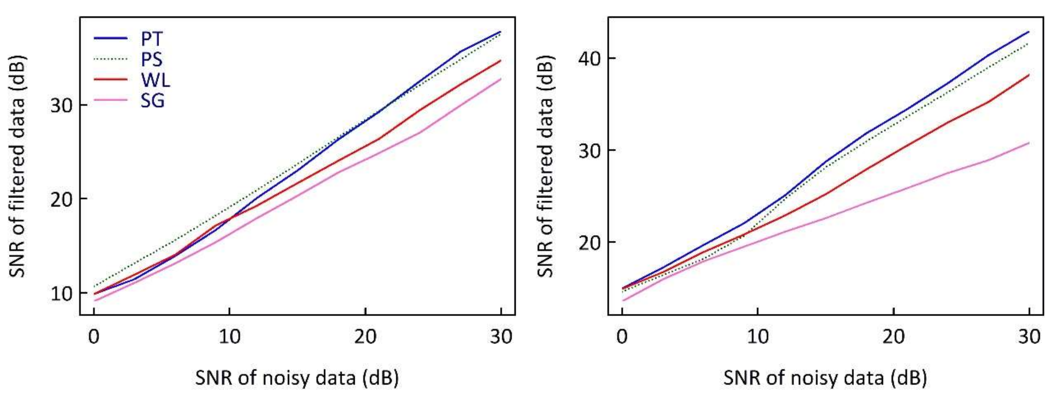


Figure 4. SNR of filtered data as a function of SNR of noisy data for the simulated spectrum(left) and the artificial signal(right). The order of the items in the legend corresponds

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021

to the order of the curves. PT and PS - the proposed method with threshold and sigmoid evaluation functions, WL - wavelet filter, SG - Savitsky-Golay filter.

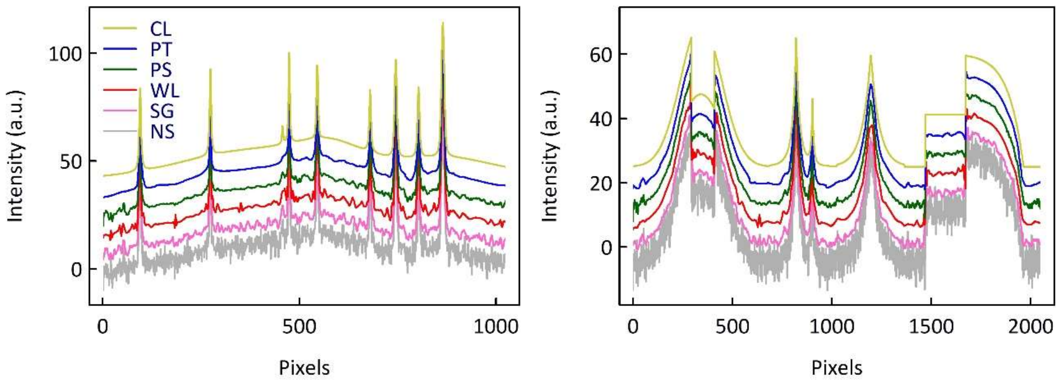


Figure 5. Filtering result for data with 15 dB noise level. CL - original data, PT and PS - proposed method with threshold and sigmoid estimation functions, WL - wavelet filter, SG - Savitsky-Golay filter, NS - noisy data. The order of the items in the legend corresponds to the order of the curves. Graphs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

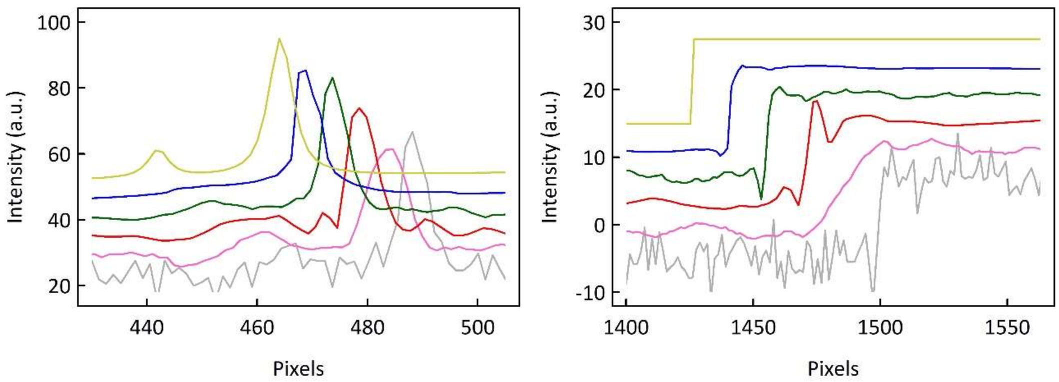
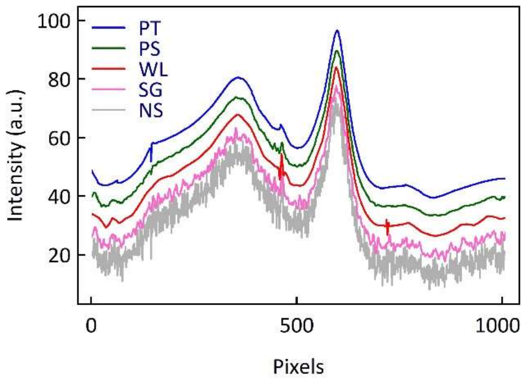


Figure 6. The result of filtering data with a noise level of 15 dB for the zones selected in Figure 3. The legends are the same as in Figure 5.

The result shows the best performance of the proposed approach.

Raman spectra of mineral samples (10) 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 8 illustrates the difference between the proposed evaluation functions. The threshold function provides better smoothness, while the sigmoid function keeps better small details.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021



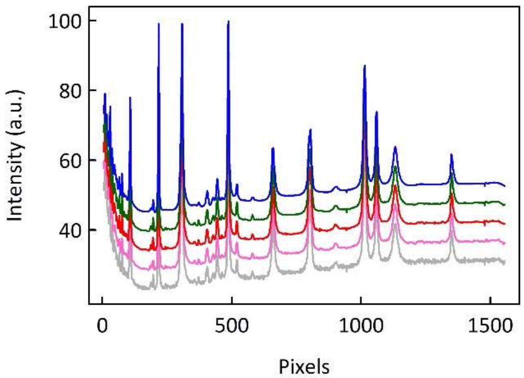


Figure 7. Processing of Raman spectra.

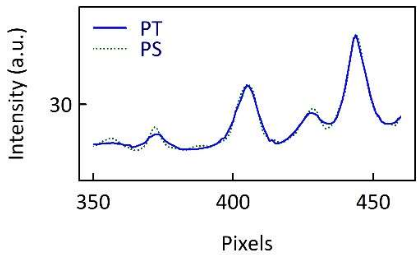


Figure 8. Comparison of evaluation functions. The threshold function provides better smoothness, while the sigmoid function preserves better small details.

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 (11). Datasets can be downloaded from (12).

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 August 2021

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. Estimation of Gaussian Noise in Spectra by the Selective Polynomial Fit. Preprints 2021, 2021080098

10. Handbook of Raman Spectra for geology. http://www.geologie-lyon.fr/Raman/

11. Online version of an adaptive polynomial filter. https://spectralmultiplatform.blogspot.com/p/mathsmooth.html

12. Data sets. https://doi.org/10.6084/m9.figshare.14899329