The analysis of the different filtering algorithm effects on the fluorescence spectrum data processing

Haiqiang Wu[1, a] , Jie Sun[2,b *] , Longyang Zheng[3,c]

The key laboratory of film electronic and communication device，Electronic and information engineering college, Tianjin University of Technology, Tianjin300384, China

[a]963646625@qq.com,[b]sun_jie_mail@126.com,[c]918593814@qq.com

Key words: data processing algorithm; digital filtering; spectral data reducing.

Abstract: Filtering is a main method to eliminate or reduce the noise and is also a basic step to analyzing spectral data accurately in spectral data processing. In this paper, four methods of average value filtering, median filtering, cubical smoothing algorithm with five-point approximation filtering and Savitzky - Golay filtering are used for the noise reducing in the spectral data processing, and results are analyzed and compared.

Introduction

During spectral measurements, interference inside and outside of the instrument will result in random noise on measured spectral data with different extents. The noise can show up as jitter, burr, or as a sharp pulse on the normal signal curve, and it will influence the spectra characteristics analysis and also increase the error of subsequent analysis and detection [1].

The noise can be reduced by applying digital processing to measured data, and it is a method to improve signal-to-noise ratio and named as the pretreatment method in spectral analysis. Currently, typical spectrum noise processing methods include the average value filtering, the median filtering, the cubical smoothing algorithm with five-point approximation filtering and Savitzky - Golay filtering [2], etc. In this paper, we analyze digital processing results with the above four methods. Our analyses show that the best template filter can indeed be obtained by setting the parameters of the processing method based on the size of error and the stability of processing results [3].

Filter methods of spectral analysis

The average value filtering method. The average value filtering is a linear approach, and its basic principle is to represent measured value by an average of multiple point values within a neighboring domain S. For a M pixel neighborhood window, the mathematical expression for the process template is

$$\bar{g}(i) = \frac{1}{M}\sum_{j\in S}^{M} g(j) \tag{1}$$

where g(i) is the measurement at point i with S as its neighborhood domain, and M the total number of points within S.

This method is obvious for noise reduction, but over smoothing with a large M can lead to a gentler spectral curve slope. Therefore, M should be carefully controlled in removing noise

interference to avoid spectral details blurring [4].

The median filtering method. The median filtering method is based on the theory of order statistics in nonlinear signal processing technology which can effectively restrain noise. The basic principle of the median filter is that the digital data point values are approximated by the median of nearby data point values, which can eliminate isolated noise points. The expression of this method is

$$V_{out} = med\{a_1, a_2, a_3, \ldots, a_i, a_{i+1}, \ldots, a_n\} \tag{2}$$

where $a_1, a_2, a_3, \ldots, a_i, a_{i+1}, \ldots, a_n$ are the values within the neighborhood of a_i, and V_{out} is the median of n point window.

The median filter is very effective for processing pulse random noise, such as salt noise or pepper noise [5].

The cubical smoothing algorithm with five-point approximation filtering method. The cubical smoothing algorithm with five-point approximation filtering method uses the polynomial least square approximation to construct smooth filtering of sampling points. The formula is

$$\sum_{i=-n}^{n} Y_j t_i^k = \sum_{j=0}^{m} a_j \sum_{i=-n}^{n} t_i^{k+1} \tag{3}$$

Take n=2 (5 nodes) and m=3 in (3), solve for a_0, a_1, a_i, a_{it}, substitute them back into equation (3), and choose t=0，+1，-1，+2，-2，we get the cubical smoothing algorithm with five-point approximation formula

$$\bar{y}_{i-2} = (69y_{i-2} + 4y_{i-1} - 6y_i + 4y_{i+1} - y_{i+2})/70 \tag{4}$$

$$\bar{y}_{i-1} = (2y_{i-2} + 27y_{i-1} + 12y_i - 8y_{i+1} + 2y_{i+2})/35 \tag{5}$$

$$\bar{y}_i = (-3y_{i-2} + 12y_{i-1} + 17y_i + 12y_{i+1} - 3y_{i+2})/35 \tag{6}$$

$$\bar{y}_{i+1} = (2y_{i-2} - 8y_{i-1} + 12y_i + 27y_{i+1} + 2y_{i+2})/35 \tag{7}$$

$$\bar{y}_{i+2} = (-y_{i-2} + 4y_{i-1} - 6y_i + 4y_{i+1} + 69y_{i+2})/70 \tag{8}$$

where $\bar{Y}_i$ is the improved value of Y_i

The cubical smoothing algorithm with five-point approximation filter has the feature of eliminating interference components and keeping the original curve characteristics unchanged [6].

Savitzky-Golay filtering method. The main idea of Savitzky - Golay (S - G) filtering algorithm is to take a fixed number of points around x_i to fit a polynomial g(x) which is used to calculate the data point value g_i at x_i ($g_i \in$ (G)) [7] as the following

$$g_i = \sum_{k=0}^{M} b_k \left[\frac{x - x_k}{\Delta x} \right] k \tag{9}$$

where g(x) is a Mth order polynomial with uniform spacing $x_{i+1} - x_i = \Delta x$. Let n_1 be the number of

points to the left side of x_i, and n_r be that to the right side, coefficients b_k in (9) are derived by minmizing the following expression

$$\min \sum_{i=1}^{i+n_r} [p_i(x_j) - y_j]^2 \tag{10}$$

The simulation experiment and analysis

Analysis of the effect on the simple signal. To contrast the characteristics of these four filter algorithms, a sine function y (x) = sin (x) is used as the ideal signal component. The total signal is a superposition of the ideal signal and a Gaussian white noise N(0, 0.005) with mean value 0 and mean square error 0.005. The average value filter, the median filter, the cubical smoothing algorithm with five-point approximation filter and Savitzky - Golay filtering are used to process the signal. The output signal curves are described respectively in the following.

Figure 1 shows signal curve and filtering results with the four methods. The comparison of the four methods is shown in figure 2.

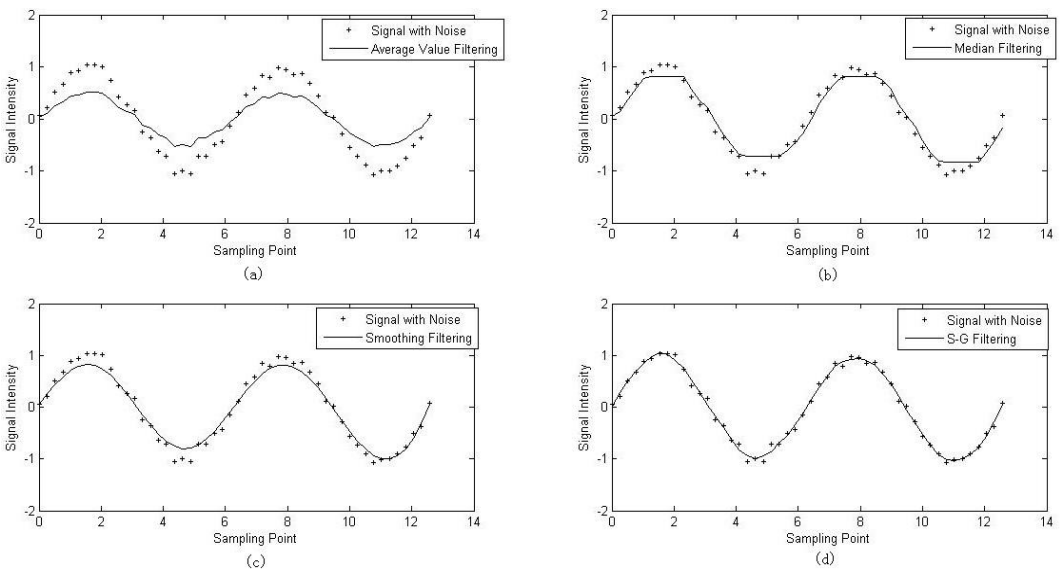

Figure 1. Filtering result with the four methods
(a) The average value filtering, (b) The median filtering, (c) The cubical smoothing algorithm with five-point approximation filtering, (d) Savitzky - Golay filtering

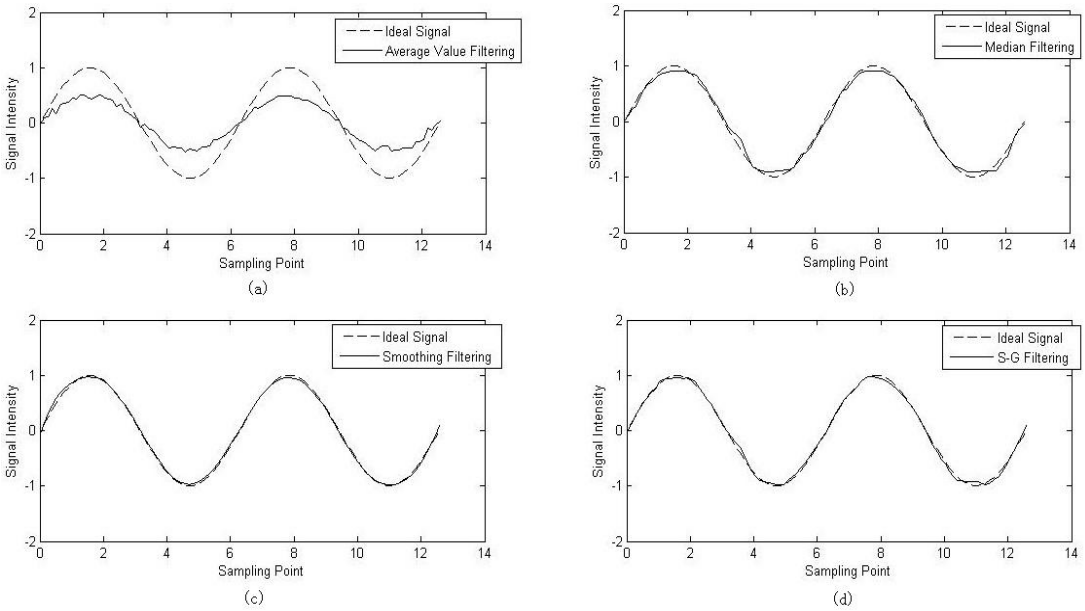

Figure 2. Comparison of the different filters
(a) The average value filtering, (b) The median filtering, (c) The cubical smoothing algorithm with five-point approximation filtering, (d) Savitzky - Golay filtering

Figures 2(a)-(d) are the results processed by the average value filtering, the median filtering, the cubical smoothing algorithm with five-point approximation filtering and Savitzky - Golay filtering respectively. From the curves we can see that the average value filtering effect is the worst. The cubical smoothing algorithm with five-point approximation filtering and the Savitzky - Golay filtering are better, in which the interference component is eliminated and the original curve characteristics unchanged. It should be emphasized that the cubical smoothing algorithm with five-point approximation filtering is more complex than average value filtering and median filtering, and also computational cost is higher with Savitzky - Golay filtering.

Statistical error analysis. Table 1 lists filtered data values of 11 sampling points in the range of $(0, 2\pi)$. $y(x) = \sin(x)$ is the ideal signal, $f(x) = y(x) +$ Gaussian noise N (0, 0.005) the total signal, $y_1(x)$, $y_2(x)$, $y_3(x)$ and $y_4(x)$ represent respectively data processing values by the average value filtering, the median filtering, the cubical smoothing algorithm with five-point approximation filtering and Savitzky - Golay filtering.

Table 1. The value contrast of the different filtering methods

x	0.000	0.571	1.205	1.840	2.475	3.110	3.745	4.379	5.014	5.649	6.283
y(x)	0.000	0.540	0.934	0.963	0.618	0.032	-0.567	-0.945	-0.955	-0.593	0.000
f(x)	0.086	0.543	0.855	0.816	0.612	0.128	-0.596	-0.862	-0.867	-0.580	0.069
$y_1(x)$	0.043	0.271	0.427	0.408	0.306	0.064	-0.298	-0.431	-0.433	-0.290	0.035
$y_2(x)$	0.052	0.567	0.984	0.927	0.636	0.042	-0.590	-1.000	-0.950	-0.567	0.045
$y_3(x)$	0.005	0.525	0.919	0.957	0.639	0.079	-0.523	-0.948	-0.947	-0.570	-0.024
$y_4(x)$	0.074	0.537	0.973	0.955	0.611	0.018	-0.538	-0.958	-0.942	-0.566	0.062

Table 2 shows errors correspondingly, with absolute error $\Delta x_n = |y_n(x) - y(x)|$, mean measurement error $\bar{x} = \frac{1}{M}\sum_{t=1}^{M}|y_n(x) - y(x)|$, and the root mean square error $\delta = \sqrt{\frac{\sum_{t=1}^{M}[y_n(x)-y(x)]^2}{M}}$, where n = 1, 2, 3, 4 representing results of the four methods.

Table 2. The statistics errors of the four filtering methods

x	0.000	0.571	1.206	1.841	2.475	3.110	3.745	4.379	5.014	5.649	6.283	$\bar{x}$	δ
Δx_1	0.043	0.269	0.507	0.555	0.312	0.033	0.269	0.514	0.522	0.303	0.035	0.306	0.362
Δx_2	0.053	0.027	0.050	0.036	0.018	0.010	0.023	0.055	0.005	0.026	0.045	0.032	0.036
Δx_3	0.005	0.015	0.014	0.006	0.021	0.047	0.044	0.003	0.008	0.023	0.024	0.019	0.024
Δx_4	0.074	0.004	0.040	0.008	0.007	0.014	0.029	0.013	0.013	0.027	0.062	0.026	0.035

We can see from table 2 that the cubical smoothing algorithm with five-point approximation filtering has the smallest errors, and errors associated with Savitzky - Golay filtering and median filtering are also small. The smaller the error value and the smaller the error of the discrete degree, the more close to the original signal.

Figure 3 plots the same errors from Table 2 for the four filtering algorithms. We can see that the smallest errors are from cubical smoothing algorithm with five-point approximation filtering and the Savitzky-Golay filtering.

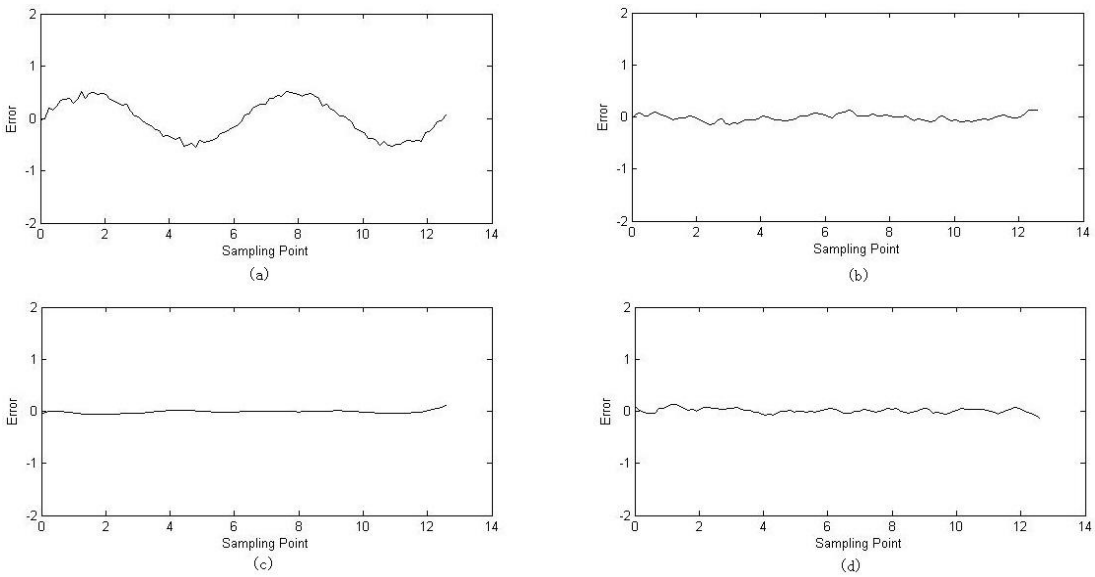

Figure3. Comparison of the absolute errors
(a) The average value filtering, (b) The median filtering, (c) The cubical smoothing algorithm with five-point approximation filtering, (d) Savitzky - Golay filtering

Conclusion

In this paper, four methods of the average value filtering, the median, the cubical smoothing algorithm with five-point approximation and Savitzky - Golay are used to study signal processing by varying noise intensity and signal frequency and amplitude and by using signals with multiple inputs. Our comparative study shows that Savitzky - Golay method is the best one both in theory and in practice for fluorescence spectrum data processing.

Reference

[1]Shaohui Yu.Research on Denoising Algorithms of Fluorescence Spectroscopy. Journal of Hefei Normal University,2013,31(3):4-6.

[2]Yaqin Tang.A Review of Several Image Smoothing and Denoising Methods [J].Journal of Southwest Agricultural University,2009,31(11):125-128.

[3]Chapter3:Image Processing, Matrox MIL Help,Matrox Electronic Systems Ltd，2005.

[4]Jianhua Wu, Chisheng Li , Weixing Zhou.Comparison between the Noise suppressing Abilities of the Median Filter and the Average Filters.Journal of Nanchang University, 1998, (1):32-35.

[5]Limei Liu.Research of median filter technology.Journal of Yunnan Normal University,2004,24(1):23-27.

[6]Wei Wu,Biao Chen,Jiangfeng Wu,Kui Huang.Study on reverse deduction of reservoir inflow based on cubical smoothing algorithm with five-point approximation. Water Resources and Hydropower Engineering, 2013,44(12):100-102.

[7]Chinrungrueng C. Combining Savitzky-Golay filters and median filters for reducing speckle noise in SAR images[C]. IEEE International Conference on Systems, Man & Cybernetics. IEEE, 2003:690 - 696.