

Moving Average and Savitzki-Golay Smoothing Filters Using Mathcad

José Luis Guiñón, Emma Ortega, José García-Antón, Valentín Pérez-Herranz
Departamento de Ingeniería Química y Nuclear.
Universidad Politécnica de Valencia.
P.O. Box 22012. E-46071 Valencia (Spain)
jlguinon@iqn.upv.es

Abstract - All measurement processes cause certain amount of random variations in the signal; this phenomenon is called noise. The procedure to reduce or smooth the noise of a measured signal is commonly known as filtering. In this paper we present the use of Mathcad software for the implementation and analysis of the moving average and Savitzky-Golay filters. The moving average filter is the simplest digital filter to understand and use. As its name suggests, this filter operates by averaging a number of points in a recursive fashion. In spite of its simplicity, the moving average filter is effective for time domain encoded signals. The Savitzky-Golay filter is based on the least squares polynomial fitting across a moving window within the data in the time domain. Generally, a high order polynomial (n=4) allows a high level of smoothing without attenuation of data features. By contrast, the Savitzky-Golay filtering method is better than averaging because it tends to preserve data features such as peak height and width, which are usually attenuated by the moving average filter. By means of the Mathcad software, moving average and Savitzky-Golay filters were successfully applied to the smoothing of photochemical and electrochemical reactor data.

Index Terms - Mathcad, Computer Applications, Moving Average, Savitzki-Golay, Filters, Noise.

INTRODUCTION

All measurement processes cause certain amount of random variations in the signal; this phenomenon is called noise. The procedure to reduce or smooth the noise of a measured signal is commonly known as filtering. There are many different filter designs, but the most common ones are based on averaging.

The moving average filter is the simplest digital filter to understand and use [1-4]. In spite of its simplicity, the moving average filter is optimal for a common task: reducing random noise while retaining a sharp step response. This makes it the most suitable type of filter for time domain encoded signals. However, the moving average filter can produce significant distortions and reduce the signal intensity, with the subsequent loss of adjacent small signals. This effect increases with increasing filter bandwidth.

The Savitzky-Golay filter is based on the least squares polynomial fitting across a moving window within the data in the time domain [5]. Typically, a long polynomial, and a moderate order polynomial allows a high level of smoothing without attenuation of data features. By contrast, the Savitzky-Golay filtering method is better than averaging because it tends to preserve features of the data such as peak height and width, which are usually attenuated by the moving average filter.

Different mathematical algorithms, generally based on Excel macro and Matlab, have been developed to implement the moving average and Savitzky-Golay filters [2, 6-8]. Among the different symbolic solvers, Mathcad is one of the most widely used. The ease with which Mathcad lets the user type equations and text, nearly exactly as we would write them down paper, and its capability of generating complex graphs, make Mathcad be a powerful and flexible tool to develop a variety of instructional designed programs to promote knowledge.

The use of Mathcad in the curriculum of chemical engineering has been discussed in previous papers [9-12]. In this manuscript we report the use of the Mathcad software for the implementation and analysis of the moving average and Savitzky-Golay filters. Finally, Mathcad was successfully implemented to the smoothing of photochemical and electrochemical reactor data.

FUNDAMENTALS

As its name suggests, the moving average filter operates by averaging a number of points from the input signal to generate each point in the output signal. This procedure is repeated so that a moving window of M points is used to calculate the average of the data set. The algorithm takes the mathematical expression:

$$y_i := \begin{cases} \frac{\sum_{j=0}^{M-1} (s_{i+j})}{M} & \text{if } i > 0 \wedge i < N - (M - 1) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where (s_{i+j}) is the input signal, y_i is the output signal, M is the number of points in the average and N the total number of data points.

As an alternative, the group of points of the input signal can be chosen symmetrically around the output point. In that case M has to be an odd number. In equation form, this is written:

$$y_{S_i} := \begin{cases} \frac{\sum_{j=\frac{-(M-1)}{2}}^{\frac{(M-1)}{2}} (s_{i+j})}{M} & \text{if } i > \frac{(M-1)}{2} \wedge i < N - \frac{(M+1)}{2} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The Savitzky-Golay filtering can be considered as a generalized moving average. The underlying function within the moving window is not a constant (whose estimate is the average of M points), but coefficients resulting from an unweighted linear least squares fit using a polynomial of a given degree, n . The Savitzky-Golay filter of a waveform can be calculated by:

$$g_m := \sum_{k=-n_L}^{n_R} c_{k+n_L} \cdot s_{m+k} \quad (3)$$

where s_{m+k} is the input signal, g_m is the output signal. Here n_L is the number of points used to the left of a data point m , while n_R is the number used to the right. The Savitzky-Golay coefficients, c_{k+n_L} , are given by:

$$c_{M+n_L} := \left[\left((A^T \cdot A)^{-1} \cdot (A^T) \right) \right]_{0, M+n_L} \quad (4)$$

where:

$$M := -n_L \dots n_R \quad (5)$$

$$m := n_L \dots N - n_R \quad (6)$$

$$j := 0 \dots n \quad (7)$$

$$A_{M+n_L, j} := M^j \quad (8)$$

RESULTS AND DISCUSSION

Figure 1 shows the random noise and the input signal formed by two Gaussians, a major one centred at 0.20 s, while the minor one is centred at 0.24 s. Figure 2 shows the output signals obtained with 25-point moving average filters symmetrical and directional respectively. Both filters produce a sharp noise reduction, however, the directional filter with the points on only one side, causes a relative shift between the input and output signals. The effect of the number of points on the average is shown in Figure 3. Clearly, as the number of points in the filter increases, noise signal becomes lower; however, there are significant distortions of the signal by the smoothing filter, which increases with increasing filter bandwidth. This phenomenon is expected as it inevitably involves averaging part of the baseline with the rising portion of the minor peak, leading to its disappearance. Figure 4 shows the linear variation in the

plot of signal-to noise-ratio, S/N , vs the square-root of the number of points in the average, $M^{1/2}$, where S is the maximum value of the signal and N the standard deviation of the baseline.

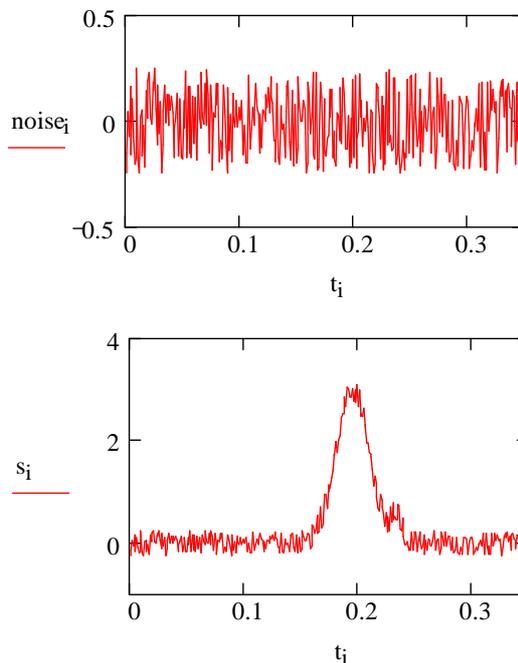


Figure 1.- Top: Random noise. Bottom: Simulation of a signal with two Gaussian features and noisy baseline.

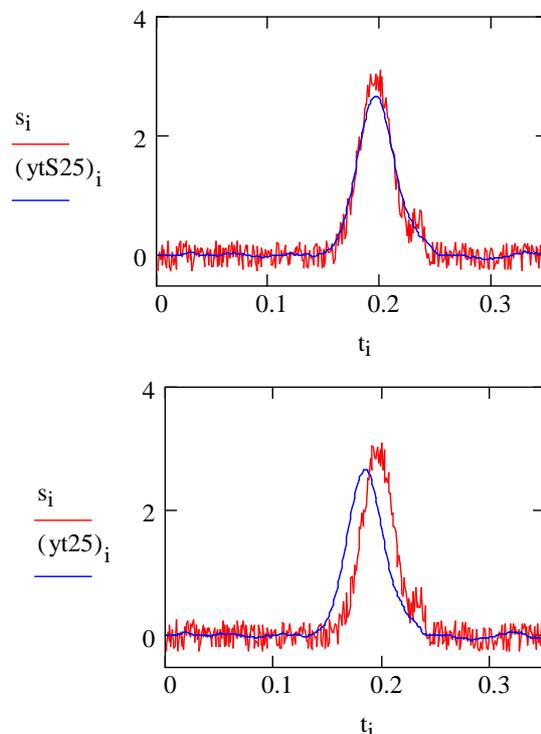


Figure 2.- Effect of the 25-point moving average filter. Top: Symmetrical filter. Bottom: Directional filter.

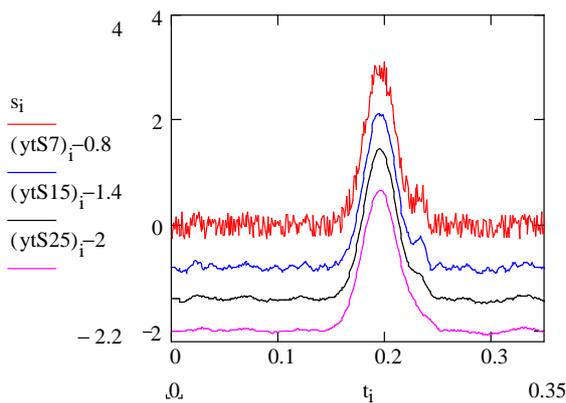


Figure 3.- Noisy Signal and the same signal filtered with 7, 15 and 25 point symmetrical moving average filters, respectively. Signals are offset for visibility.

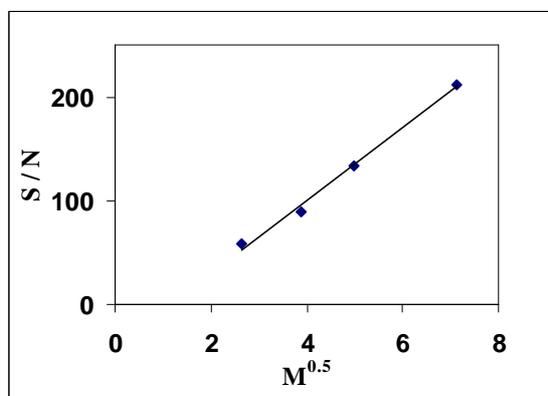


Figure 4.- Plot of signal-to-noise ratio vs square-root of the number of points in the average.

The coefficients of Savitzky-Golay filters for orders $n=2$ and $n=4$, with several choices of n_L and n_R are shown in Table I. The central column is the coefficient applied to the data s_m for obtaining the smoothed g_m . Coefficients to the left are applied to earlier data and to the right to later data. The coefficient values always add the unity. Note that the coefficients always have a central positive lobe and with smaller outlying corrections of both positive and negative.

Figure 5 illustrates the effect of a 25 point Savitzky-Golay filter (order 2) on the same input signal as shown in Figure 1. Clearly, the baseline noise amplitude is reduced by

the Savitzky-Golay filter much more than the moving average (Figure 2). On the other hand, Figure 5 demonstrates that there is little distortion in the signals, with possible observation of minor peaks.

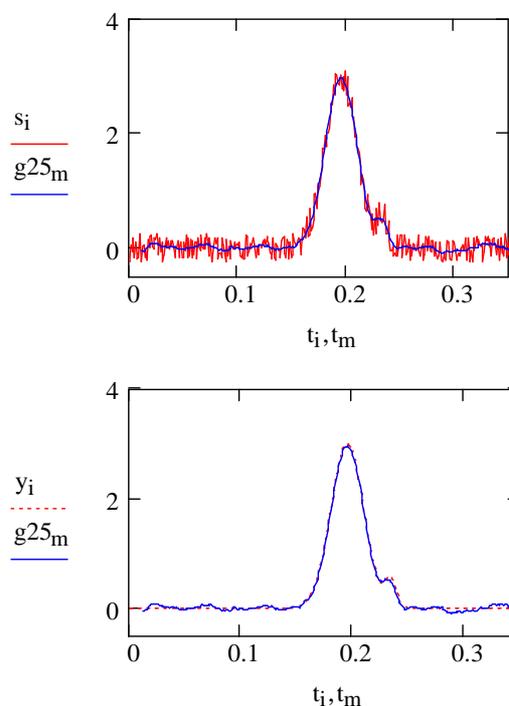


Figure 5.- Top: Effect of 25-point Savitzky-Golay filter on the noisy signal (s_i). Bottom: Comparison of the signal without noise (y_i) and the filtered signal (g_{25}).

Moving average and Savitzky-Golay filters were successfully applied to the smoothing of photochemical and electrochemical reactor data. Figure 6 (top) shows the noisy signal (energy vs wavelength, λ) obtained in a photochemical reactor equipped with a hydrogen or xenon lamp. The effect of the 25-point Savitzky-Golay ($n=2$) and 25-point moving average filters is shown in Figure 6 (bottom). Savitzky-Golay filter produces very little distortion in the signal in comparison to the moving average filter.

Table I.- Savitzky-Golay Smoothing Filters

n	nL=nR	C _{M+nL} : Savitzky-Golay Coefficients										
2	3			-0.095	0.143	0.286	0.333	0.286	0.143	-0.095		
2	4		-0.091	0.061	0.169	0.234	0.255	0.234	0.169	0.061	-0.091	
2	5	-0.084	0.021	0.103	0.161	0.196	0.207	0.196	0.161	0.103	0.021	-0.084
4	3			0.022	-0.013	0.325	0.567	-0.13	0.022			
4	4		0.035	-0.128	0.07	0.315	0.417	0.315	0.07	-0.128	0.035	
4	5	0.042	-0.105	-0.023	0.14	0.28	0.333	0.28	0.14	-0.023	-0.105	0.042

The use of the Savitzky-Golay filter is consequently much easier to justify because the signal-to-noise ratio is improved by a factor of around 20 relative to the non filtered input signal, whereas for the moving average filter this ratio is approximately 10.

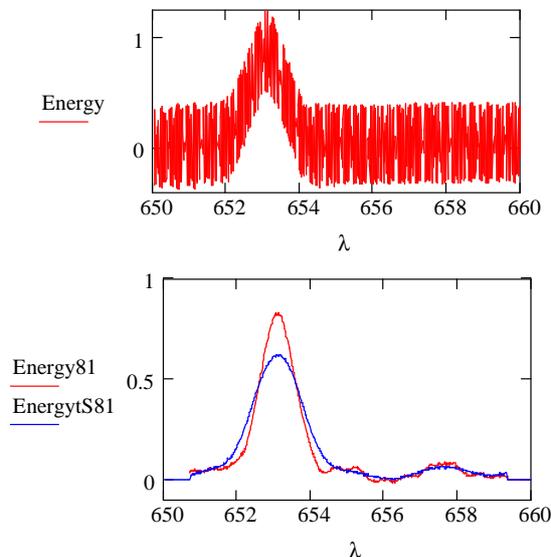


Figure 6.- Top: Noisy signal in a photochemical reactor. Bottom: Relative effect of the 81-point Savitzky-Golay and moving average filters on the noisy signal.

Figure 7 shows the potential-time response to applying a pulse of current in an electrochemical reactor. In this case, the use of 201-point Savitzky-Golay ($n=1$) filter, is not so excellent as the above example; however, it suffices enough to retrieve the signal with little noise.

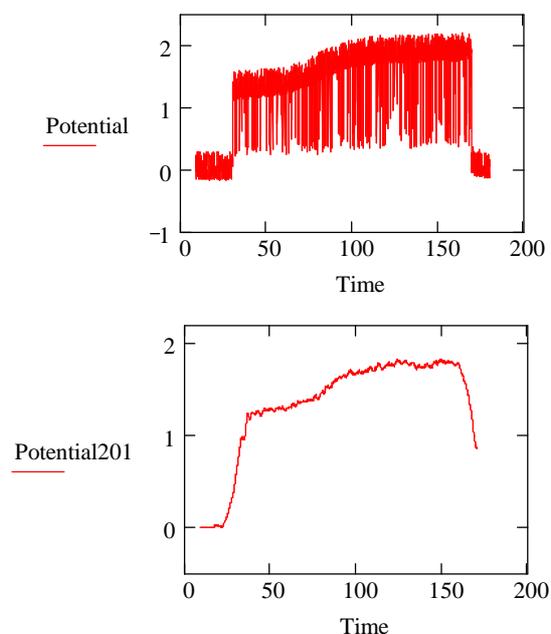


Figure 7.- Top: Potential-time response in a electrochemical reactor. Bottom: Effect of 201-point Savitzky-Golay filter. Order $n=1$.

CONCLUSIONS

This paper describes the use of the Mathcad software for the implementation and analysis of the moving average and Savitzky-Golay filters. Mathcad was successfully implemented to the smoothing of photochemical and electrochemical reactor data.

ACKNOWLEDGMENT

We wish to express our gratitude to the Universidad Politécnica de Valencia, and to Dr Asunción Jaime for her translation assistance.

REFERENCES

- [1] Donnelly, D. "The fast Fourier transform for experimentalist, part V: Filters" *Computing in Science&Engineering*, 8 (1), p. 92,(2006).
- [2] De Levie, R. "Advanced Excel for Scientific Data Analysis" Oxford University Press, Oxford, (2004).
- [3] Tham, M.T. "Dealing with measurement noise. Moving average filter" *Chemical Engineering and Advanced Materials*, University of Newcastle upon Tyne. (1998). www.lorien.ncl.ac.uk/ming/filter/filmav.htm
- [4] DATAQ Instruments. "A closer look at the advanced CODAS moving average algorithm". www.dataq.com/applicat/articles/an14.htm
- [5] Savitzky, A., Golay, M.J.E. "Smoothing and differentiation of data by simplified least squares procedures" *Analytical Chemistry*, 36(2), p.1627, (1964).
- [6] Borchers, H.W. "Savitzky-Golay smoothing an R implementation" in <http://tostoy.newcastle.edu.au/R/help/04/02/0385.html>
- [7] Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. "Numerical recipes in C: The Art of Scientific Computing" Cambridge University Press, Cambridge, (1992).
- [8] Press, W.H., Teukolsky, S.A. "Savitzky-Golay smoothing filters" *Computers in Physics*, Nov/Dec, p. 669, (1990).
- [9] Guiñón, J.L., Igual, A., Pérez-Herranz, V., García-Antón, J. "Simulación del análisis frecuencial de señales y del derrame espectral mediante el programa Mathcad" *Ingeniería Química*, 39, (445), p.144, (2007).
- [10] Guiñón, J.L., Ortega, E., García-Antón, J., Pérez-Herranz, V. "Regresión no lineal aplicada a la cinética de un reactor discontinuo: Análisis comparativo mediante dos programas informáticos" *Ingeniería Química*, 38, (441), p.170, (2006).
- [11] Ortega, E., Guiñón, J.L., García-Antón, J., Pérez-Herranz, V. "Nonlinear regression of a second-order kinetics using Mathcad: Ethyl acetate saponification" Abstracts of Papers 230th ACS National Meeting: Computer 45. Washington, DC, United States, Aug. 28, (2005).
- [12] Guiñón, J.L., García-Antón, J., Ortega, E., Pérez-Herranz, V. "Simulation of a lock-in amplifier by means of the Mathcad symbolic software solver" *Afinidad*, 60(506), p. 396, (2003).