Savitzky–Golay smoothing and differentiation filter for even number data

Jianwen Luo, Kui Ying, Jing Bai*

Department of Biomedical Engineering, Tsinghua University, Beijing 100084, PR China

Received 19 April 2004; received in revised form 22 February 2005

Abstract

The Savitzky–Golay smoothing and differentiation filter optimally fits a set of data points to a polynomial in the least-squares sense. The Savitzky–Golay filter has been developed and generalized well in the literatures. However, the data subset is subject to an odd number \( (2m + 1) \). In this communication, the Savitzky–Golay filter is extended for even number data. Simulations are performed to validate the feasibility of such an approach. And some corresponding properties are discussed.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Savitzky–Golay; Smoothing; Differentiation; Least squares; Polynomial

1. Introduction

In various fields such as signal processing, imaging processing, analytical chemistry and spectroscopic analysis, smoothing and differentiation is important and necessary. In the classical paper written by Savitzky and Golay [1], which has been cited more than 3800 times according to ISI Web of Science, one kind of digital filter for smoothing and differentiation was developed. In their approach, each successive subset of \( 2m + 1 \) points is fitted by a polynomial of degree \( p \) \((p \leq 2m)\) in the least-squares sense. The \( d \)th \((0 \leq d \leq p)\) differentiation (zeroth differentiation = smoothing) of the original data at the midpoint is obtained by performing the differentiation on the fitted polynomial rather than on the original data. Finally, the running least-squares polynomial fitting can be simply and automatically performed by convolving the entire input data with a digital filter of length \( 2m + 1 \). The convolution coefficients can be obtained for all data points, all polynomial degrees and all differentiation orders but with only an odd number of data sets [2,3].

As pointed out by Mark and Workman [4], one limitation is that the existing Savitzky–Golay filter...
is subject to using an odd number of data subset. However, the data subset can also have an even number of data \(2m\), i.e., the SG filter can have an even length. This extension will be of interest to some readers, e.g., spectroscopists [4,5] and engineers [6,7]. In this communication, a generalized SG smoothing and differentiation filter is presented to calculate the convolution coefficients for even number data using a matrix form [3]. Several closed-form solutions are given. The feasibility of such an approach is validated by computer simulations, and some properties are discussed.

2. Theory

Referring to Ref. [3], assume that \(2m+1\) data points are positioned symmetrically about the origin (i.e. the midpoint)

\[
x = [x_{-m}, x_{-m+1}, \ldots, x_m]^T.
\]

The matrix \(G_{(2m+1)\times(2p+1)}\) contains the convolution coefficients of the SG filter for different differentiation orders at the midpoint, given by [3]

\[
G = S(S^T S)^{-1} = [g_0, g_1, \ldots, g_p].
\]

where \(S\) is the \(2m+1\)-by-\(2p+1\) basic matrix, as

\[
S = [s_0, s_1, \ldots, s_p].
\]

\[
s_i = [(-m)', (-m+1)', \ldots, m']^T \quad (0 \leq i \leq p).
\]

The convolution coefficients for the \(d\)th differentiation are given by [3]

\[
h_{p,d} = d!g_d = d![g_d(-m), g_d(-m+1), \ldots, g_d(m)]^T,
\]

where \(g_d\) is the \(d\)th column vector of the matrix \(G\) while \(g_d(n)\) is the \(n\)th element of the vector \(g_d\).

Now we consider an even number \((2m, 2m-1 > p)\) of data sets. The index of the data samples ranges from \(-m+1\) to \(m\). The data samples can be equivalently regarded as being symmetrical about the origin, by shifting the origin one half. Therefore, the input signals are represented by

\[
x = [x_{-m+1/2}, \ldots, x_{-1/2}, x_{1/2}, \ldots, x_{m-1/2}]^T.
\]

Consequently, the column vectors of the basic matrix \(S_{(2m)\times(2p+1)}\) are represented by

\[
s_i = [(-m + 1/2)', (-m + 3/2)', \ldots, (m - 1/2)']^T \quad (0 \leq i \leq p).
\]

### Table 1

Convolution coefficients of Savitzky–Golay filter for even number data

<table>
<thead>
<tr>
<th>Differentiation order ((d))</th>
<th>Polynomial degree ((p))</th>
<th>Convolution coefficients (h_{p,d}(k)) ((k = -m + 1, -m + 2, \ldots, m))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 or 1</td>
<td>(\frac{1}{2m})</td>
</tr>
<tr>
<td>0</td>
<td>2 or 3</td>
<td>(\frac{3[12m^2 - 7 - 20(k - 1/2)^2]}{32(m+1)m(m-1)})</td>
</tr>
<tr>
<td>0</td>
<td>4 or 5</td>
<td>(\frac{15[240m^4 - 920m^2 + 407 - 1120m(k - 1/2)^2 + 1960(k - 1/2)^2 + 1008(k - 1/2)^4]}{2048(m+2)(m+1)m(m-1)(m-2)})</td>
</tr>
<tr>
<td>1</td>
<td>1 or 2</td>
<td>(\frac{6(k - 1/2)}{2m+1)m(m-1)})</td>
</tr>
<tr>
<td>1</td>
<td>3 or 4</td>
<td>(\frac{5(k - 1/2)[240m^4 - 360m^2 + 155 - 336m^2(k - 1/2)^2 + 196(k - 1/2)^2]}{8(2m+3)(2m+1)m(m-1)(2m-1)(2m-3)})</td>
</tr>
<tr>
<td>2</td>
<td>2 or 3</td>
<td>(\frac{15[-4m^2 + 1 + 12(k - 1/2)^2]}{4(2m+1)m(m-1)(m-2)(m-1)})</td>
</tr>
</tbody>
</table>
The convolution coefficients corresponding to Eq. (5) are given by
\[ h_{p,d} = d! g_{d} = d! [g_{d}(-m + 1), \ldots, g_{d}(m)]^T. \] (8)

The coefficients of the SG filter are then obtained by directly substituting Eq. (7) into Eq. (2) (where the matrix \( G \) is \( 2m \)-by-\( p + 1 \) now) and Eq. (8). Several closed-form solutions are given in Table 1. Here only the SG filter at the midpoint or the center of symmetry (1/2) is considered. In practice, the SG filter at any position can be obtained. However, the SG filter at the midpoint has the advantage of symmetry (or anti-symmetry) and linear phase property.

3. Simulations

Simulations in MATLAB 6.5 (The MathWorks Inc., Natick, MA) are used to validate the feasibility of the approach presented previously. The underlying function of the data is assumed to be a Gaussian function. Adding uncorrelated Gaussian white noise, we obtained the noisy data. The maximum and the full-width at half maximum (FWHM) are set at 1.0 and 5.0, respectively. The additional noise is zero-mean and has a standard deviation of 0.05. The sampling interval is set to be \( \frac{1}{20} \) FWHM, i.e. 20 data points per FWHM.

The smoothed data and the smoothed differentiation are given by the convolution of the noisy data and the coefficients of the SG smoothed and differentiation filter, respectively. The parameters of the SG filter are taken as \( m = 10, p = 3 \) and \( d = 0, 1, 2 \), respectively. The filter coefficients are calculated from the closed-form solutions and listed in Table 2. The corresponding frequencies (transfer functions) are shown in Figs. 1–3.

The results are presented in Figs. 4–6, respectively. As can be seen, the SG filter presented in this communication is valid to calculate the

![Fig. 1. The frequency response of the Savitzky–Golay smoothing filter (m = 10, p = 3, d = 0).](image)
smoothed value and smoothed differentiation for even number of noisy data.

4. Discussions

According to Eq. (2), we have

\[ S^T G = I, \]  \hspace{1cm} (9)

where \( I \) is the \( p+1 \)-by-\( p+1 \) identity matrix.

Expand the column-wise of the left-hand side and the right-hand side, respectively, and substitute Eq. (5). We can get

\[ \sum_{k=-m}^{m} k^i h_{p,d}(k) = d! \delta(i-d), \hspace{0.5cm} i = 0, 1, \ldots, d, \]  \hspace{1cm} (10)

where \( \delta(n) \) represents the Dirac delta function.
Similarly, the coefficient restriction for even number data is given by

$$X_m k = \frac{C_0}{m+1} k / C_1 h_p d(i) = d! \delta(i - d),$$

$$i = 0, 1, \ldots, d.$$  

(11)

The coefficient restriction given by Eqs. (10) and (11) is related to the flatness of the frequency response at $\omega = 0$, or to the moment conservation of the input signal or signal differentiation [3,8].

Assuming that the input signal is contaminated by independent Gaussian white noise (or a weaker condition of uncorrelated and homoscedastic noise), the noise amplification factor of a digital filter is given by the sum of the squares of the filter impulse response [3,9]. In statistics, the Gauss–Markov theorem [10] states that the best linear unbiased estimators of the coefficients for a linear model with independent Gaussian white noise or errors (or a weaker condition mentioned previously) are the least-squares estimators, which are equivalent to the SG filter. Therefore, for an even filter length, the SG filter presented in this work is an optimal filter in the sense of minimizing the noise amplification factor, but its coefficients are subject to additional constraints given by Eq. (11) or to the aforementioned flatness constraint of the frequency response at $\omega = 0$.

Since the origin has been shifted to the imaginary halfway position in the implementation, the SG filter for even number data will result in a time delay of half a sampling interval. With the compensation of a noninteger time delay, the SG filter can have better results. For example, the two-point backward differentiation [11] that can be regarded as the SG filter for even number data in the case where $m = 1$, $p = 1$, $d = 1$ is more precise to calculate the differentiation at a point halfway between the data points, rather than at either one of the two points.

An ideal full-pass differentiation filter has a gain that increases with frequency ($H(e^{j\omega}) = (\omega)^d$); therefore it greatly amplifies high-frequency noises. In practice, one would readily choose a low-pass differentiation filter rather than a full-pass one. The SG differentiation filter can be regarded as a kind of low-pass filters. According to the symmetry or antisymmetry of the coefficients [1,3], the SG filters of odd length for even and odd order differentiations belong to type I and type III filter [12], respectively. Similarly, the SG filters of even length for even and odd order differentiations belong to type II and type IV filter [12], respectively. It is known that the frequency responses at $\omega = \pi$ of the type II and type III filter are zero [12]. Therefore, for the even order differentiation, the SG filter of even length has a better characteristic of attenuation at the high frequency of $\omega = \pi$ than that of odd length (see Figs. 1 and 3). It may be helpful to reduce some high-frequency noises when they are present. For the odd length SG smoothing filter, Hamming presented a modified least-squares method (i.e., a modified SG filter) to restrict the frequency response at $\omega = \pi$ to be zero [9]. The use of the even length SG smoothing filter would be an alternative method. For the odd order differentiation, the SG filter of even length has a worse characteristic of attenuation at the high frequency than that of even length (see Fig. 2), and hence should be used with caution. The performance evaluation and the comparison of the SG filter of even length and that of odd length, both in theory and in applications, remains to be investigated.

Fig. 6. The ideal 2nd differentiation (—) and the smoothed 2nd differentiation (--) using the Savitzky–Golay differentiation filter ($m = 10$, $p = 3$, $d = 2$).
Acknowledgements

This work is supported in part by the National Natural Science Foundation of China (60171039, 30470466).

References