Available online at www.sciencedirect.com
ScienceDirect
Digital Signal Processing 17 (2007) 69–75
Digital Signal Processing
www.elsevier.com/locate/dsp

A multiscale polynomial filter for adaptive smoothing

M. Browne [a,*], N. Mayer [b], T.R.H. Cutmore [c]

[a] *GCCM, Griffith University, PMB50, Gold Coast Mail Centre, QLD 9726, Australia*
[b] *Osaka University, Japan*
[c] *School of Applied Psychology, Griffith University, Australia*

Available online 20 February 2006

Abstract

The effectiveness of Savitzky–Golay type symmetric polynomial smoothers is known to be strongly dependent on the window size. Many authors note that selection of the appropriate window size is essential for achieving the correct trade-off between noise reduction and avoiding the introduction of bias. However, it is often overlooked that, in the case of non-stationary signals, the optimal window size will vary with the dynamics of the signal. A multiresolution approach is outlined, along with criteria for varying window size with respect to translation, based on evaluation of the residuals of the smoothed data in the local region. Adaptive window polynomial smoothing is shown to be superior to fixed window smoothing for a test signal at various signal-to-noise ratios.

Keywords: Savitzky–Golay; Adaptive; Polynomial; Filtering

1. Introduction

In a now classic paper, Savitzky and Golay [1] defined a FIR digital filter as the output of a least-squares polynomial fit to the data in a local window. Although the frequency response of the unweighted Savitzky–Golay (SG) filter is undoubtedly poor in terms of straight-forward frequency separation [2], the interpretation of the output as the 'automatic' realization of a polynomial least-squares fitting procedure has made the SG filter a popular choice in spectroscopy [3], voltammetry [4], and other fields such as biomedical monitoring [2]. The SG coefficients for arbitrary order (length) and polynomial degree may be derived analytically [5] or from a table [6], and the filter may implemented efficiently as a convolution operation. Improved implementations of the SG filter (e.g., as specified in [2]) now allow for non-uniformly spaced data [7], and for non-symmetric polynomial estimation [8,9]: causal or anti-causal SG filters may therefore be applied.

More recently, researchers have focused on adaptive and multiscale extensions of the SG approach [4,10,11]. This work is motivated by the recognition that appropriate selection order and degree parameters of SG filter is problematic. Indeed, the implementation example provided in a much referenced volume [12] demonstrates the principle of trade-off between noise-reduction and signal preservation. For given degree, larger-order filters effectively suppress noise, at the expense of distorting smaller scale signal features, and generally features that are poorly approximated by a

* Corresponding author.
E-mail address: m.browne@griffith.edu.au (M. Browne).

1051-2004/$ – see front matter
doi:10.1016/j.dsp.2006.01.006

polynomial over a particular spatial scale [13]. Smaller-order filters preserve smaller scale filter features well, with the disadvantage of being less effective at suppressing noise [12]. For fixed order, a similar argument is made for varying the degree of the polynomial [4].

It is clear that whether adapting order or degree to suit the data at hand, the ideal degree of model responsiveness for a particular local region is dependent on both the complexity of the (unknown) underlying signal and the relative amplitude of the noise. A reasonable approach is to increase the model responsiveness until just sufficient to capture the dynamics of the signal. Obviously if insufficient model flexibility is allowed, the model will underfit, whilst 'extra' flexibility will result in some fitting to the noise, serving to decrease the goal of noise-suppression. The problem is a challenging one because the optimal level of model flexibility is generally not constant; as in the example in [12], the signal is often non-stationary—some local regions are characterized by slow changes in level, while the signal changes more rapidly in other regions.

Recent work [4,10] has focused on adapting the degree of the polynomial filter to the local characteristics of the signal whilst keeping the order constant. The polynomial degree is progressively increased until the fit quality (as measured by the residual sum-square) displays statistically insignificant improvement. Another report [11] focusing on a data with relatively abrupt level shift, proposes adapting the SG filter length in order to optimally balance the competing goals of reducing bias and uncertainty. The statistical assumptions regarding the known measurement uncertainty are tested, and the filter length reduced until there is relative confidence that approximation difference is accounted for by uncertainty from noise, rather than approximation bias.

In the current work, a framework is developed which utilizes the SG filter as a multiscale transform and tests the statistical properties of the residuals in a region that is itself scale dependent. The tests of residuals amount to a test for bias in the estimate; the largest filter size that does not violate the test of residuals is used to produce final filtered output.

2. Theory

There exist a number of alternative methods for computation of a least-squares polynomial filter, of varying complexity and efficiency [14]. Gram polynomials, for instance, represent a flexible and efficient approach [15]. Since the current application is quite straight-forward; evenly spaced data points, a query point in the filter centre, and smoothing, rather than derivative approximation, an intuitive formalism [12] based on solving a system of polynomial equations is presented. The smoothing SG filter coefficients $\phi[n]$ operates via convolution on noisy data $y[i]$ with parameters polynomial degree M and filter length $2N+1$:

$$\hat{y}[i] = \mathbf{y} \otimes \mathbf{\Phi} = \sum_{n=-N}^{N} y[i-n]\phi[n]. \tag{1}$$

It is possible to compute least-square instantaneous fit via simple convolution with the filter $\mathbf{\Phi} = \phi[n]$ because the least-squares polynomial fit estimate is only required for the time point $i = [i-n]|n=0$. For a symmetrical filter, the output should correspond to to a least-squares approximation of $y[i]$ fitted on the data in local region $\{y[i-n]\}_{-N}^{N}$, using a polynomial of degree M, $a_0 + a_1 n + \cdots + a_M n^M = a_0$ for $n = 0$. Given a Vandermonde-type matrix $\mathbf{S}$ whose entries

$$\mathbf{S} = S[n,m] = n^m, \quad n = \{-N,\ldots,N\},\ m = \{0,\ldots,M\} \tag{2}$$

are treated as the sampled polynomial basis functions $\{1, n, \ldots, n^M\}$, then the vector of $a[m]$'s as a function of $y[n]$'s is

$$\mathbf{a} = a[m] = (\mathbf{S}^T \cdot \mathbf{S})^{-1} \cdot (\mathbf{S}^T \cdot \mathbf{y}) \tag{3}$$

and this least squares approximation simplifies to

$$\phi[n] = \sum_{m=0}^{M} \left\{(\mathbf{S}^T \cdot \mathbf{S})^{-1}\right\}[0,m] n^m, \tag{4}$$

where the indexing $[0,m]$ implies that only a single row to the inverse matrix (at $n=0$) is required for computation. Since the higher order polynomial components disappear when $n=0$, the filter coefficients $\phi[n] = a_0$, and do not depend on the input data.

It is straight-forward to extend the standard SG filter, with fixed filter length N, in order to create a multiscale polynomial transform

$$\hat{y}[i,k] = \sum_{n=-k}^{k} y[i-n]\phi^k[n] \tag{5}$$

with the set of variable length symmetric Savitzky–Golay filters $\{\phi^k\}$ of length $2k+1$, noting that for filter sizes $2k+1 \leqslant M$, the polynomial approximation is perfect, and no filtering takes place.

This paper is concerned with the issue of generating an filter that, given input $y[i]$ and fixed degree M, adapts the filter size with the goal of minimizing the total estimation error, which is considered to be a combination of both model bias $b[i,k]$ and model estimation error $e[i,k]$. Assuming an underlying signal $x[i]$ and zero-mean uncorrelated white noise $n[i]$; $y[i] = x[i] + n[i]$ and given the output of a degree M, $f^M[i,k] = x_i + b_{ik} + e_{ik}$, the ideal filter length k at translation i is one that minimizes

$$\mathsf{E}|\epsilon[i,k]| = \mathsf{E}|b[i,k]| + \mathsf{E}|e[i,k]|. \tag{6}$$

In the context of polynomial curve fitting, model bias arises when the polynomial order of the local signal curve $x[i-n]$, $\{n=-k,\ldots,k\}$ is greater than M and the estimate $\hat{y}[i,k]$ consequently under-fits the signal. The expected model error magnitude $\mathsf{E}|e[i,k]|$ due to noise $n[i]$, from classical statistical theory will decrease in proportion to $1/\sqrt{2k+1}$.

The expected magnitude of the model bias $\mathsf{E}|b[i,k]| = 0$ for $2k+1 = M$ and *generally* increases with k in proportion to the magnitude of higher signal moments encompassed within the window. However, this increase in bias is not necessarily monotonic, due to spurious concurrence of under-fit models with $x[i]$. Figures 1b and 1c display the absolute residuals of the multiscale SG filter (or model bias, since the input is noise free)

$$|R[i,k]| = |\hat{y}[i,k] - x[i]| \equiv \mathsf{E}|b[i,k]|. \tag{7}$$

on a noise free Gaussian (Fig. 1a).

In Fig. 1b, which displays the residuals of a multiscale SG filter with $M=0$, bias generally increases in the region of the Gaussian as filter length k increases. Of interest are the regions around $i = 8.5, 11.5$, where bias is low irrespective on filter length. At these points, despite generally poor polynomial approximation over $n = \{-k,\ldots,k\}$ (a simple constant average in this case, since $M=0$), very little bias occurs at the query point $n=0$, since the higher order components 'balance out' the fit on either side of the symmetrical filter. Put in a more technical way, the SG filter preserves exactly moments up to $M+1$ and thus the contours (Figs. 1b and 1c) reflect the absolute $(M+2)$th derivative of the input signal.

The implication of this phenomena is that even for noise free signals, the residuals $R[i,k]$ are an imperfect estimate of true model fit; Eq. (7) holds only in the narrow definition of model bias measured not over all query points $n = \{-k,\ldots,k\}$ but only for $n=0$. This is a natural consequence of the fundamental computational advantage of the SG filter, which computes the approximating polynomial only for a single point.

Returning to the aim of estimating the filter lengths k that minimize Eq. (6) for all i, a 'comprehensive' approach would be to minimize a cost function with respect to k, such as

$$\chi^*[i,k] = \frac{1}{(2k+1)} \sum_{n=-k}^{k} \left(\hat{y}[i,k]^n - y[i+n]\right)^2 + \frac{\sigma^2}{\sqrt{2k-M+1}}, \tag{8}$$

where $\hat{y}[i,k]^n$ is the polynomial approximation at query point n,[1] and σ^2 is an estimate of the noise variance. Note that in Eq. (8), $y[i+n]$ represents the input data at translation i and offset n. For particular translation i and scale k (filter length), the set of values $\hat{y}[i,k]^n|n = \{-k,\ldots,k\}$ is the polynomial approximation of the windowed input data $y[i+n]|n = \{-k,\ldots,k\}$. This approach is unappealing since, as well as requiring an estimate of σ^2, it requires calculation of all query points n, and avoiding this computation is a principle attractive feature of the SG filter.

[1] As opposed to the standard use of the Savitzky–Golay filter (i.e., $\hat{y}[i,k]$) which only calculates an approximation at the instantaneous query at $n=0$. Thus, calculation of $\hat{y}[i,k]^n$ involves a significant increase in computational expense.

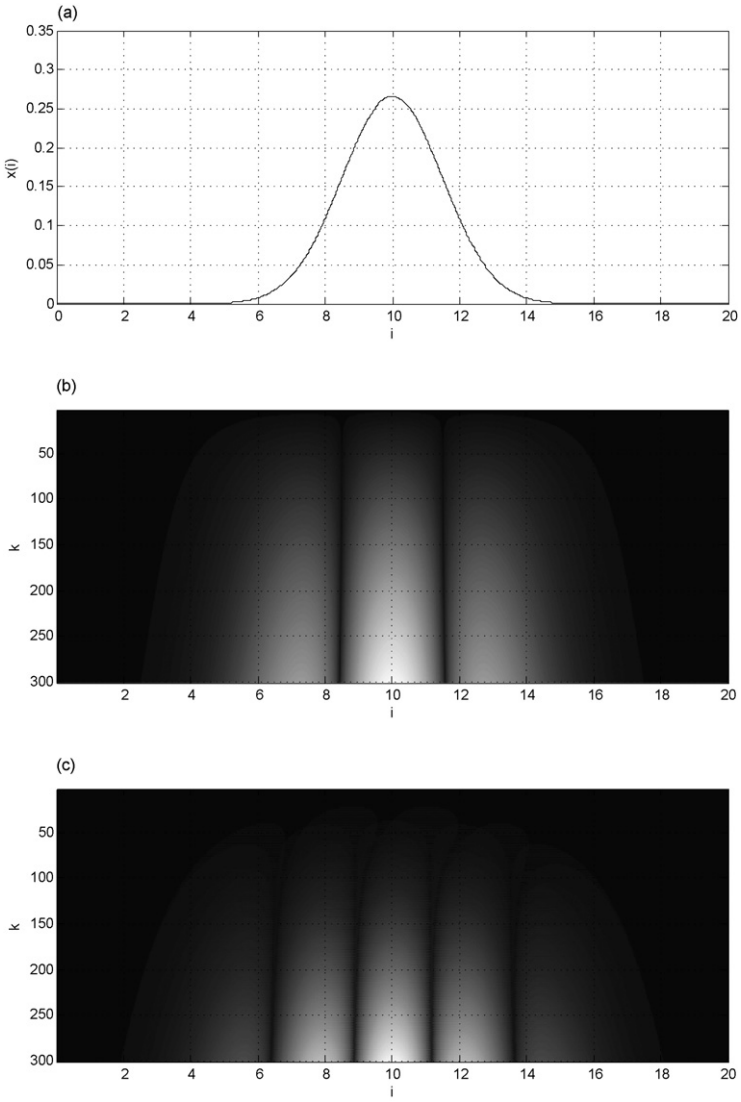

Fig. 1. Residuals of multiscale filtering of noise free Gaussian (a). Residual of SG filters with order 0 (b) and order 2 (c) polynomials are shown. Lighter values indicate larger residuals.

The approach proposed here utilizes the fact that for given k, there is overlap between nearby SG estimates; specifically, since

$$R[i,k] = \hat{y}[i,k] - y[i] = f\big(\{y[i-k],\ldots,y[i+k]\}\big), \tag{9}$$

then the input point $y[i]$ affects the set of residuals $\{R[(i-k),k],\ldots,R[(i+k),k]\}$. Characteristics of the input signal that generate bias around $y[i]$ is therefore reflected not only at the residual $R[i,k]$, but at its neighbors as well. Since these neighboring residuals are available 'for free,' it is worthwhile to establish whether a function of the form

$$\psi[i,k] = f\big(\{R[(i-k),k],\ldots,R[(i+k),k]\}\big), \tag{10}$$

which is designed to indicate adequacy of model $\hat{y}[i,k]$, may prove a useful basis for adaptively determining filter lengths k.

There are many options in choosing the function f (in Eq. (10)), and the ideal choice would be expected to depend on the known characteristics of the noise and signal for the particular application in question. For example, f be some form of statistical test-of-residuals, yielding an estimate of model adequacy. The reasoning is as follows: if the local

residuals (in Eq. (10)) are significantly different from those expected from the noise model (not having a mean zero, for instance), then it may be assumed that signal components are incorporated in the residuals, the polynomial model is under-fit, the denoised signal subject to bias and k must be reduced. However, this not being the case, a larger k is to preferred in order to reduce the variance of the estimate. This is because shorter filter lengths k represent more freedom or adaptivity in the estimate. An unnecessarily adaptive model is more susceptible to noise. Thus the rule "use the largest filter length possible that does not introduce bias" should give a reasonably good approximation of the ideal filter length. This approach is satisfying from a scientific point of view, as the signal as whole will be estimated using the rule of 'simplest sufficient model.'

The criterion for determining whether bias exists that is implemented here relies on the assumption that the noise $n[i]$ is an uncorrelated process: there is no time dependence between subsequent measurements of $n[i]$. This implies that if there is an correlation between subsequent residuals $R[i,k]$ of the polynomial approximation $\hat{y}[i,k]$ in the local region $\{i-k,\ldots,i+k\}$, then $R[i,k]$ may be suspected of comprising non-noise (i.e., signal) energy. Thus, correlated residuals are an indication that there exists unapproximated signal structure, and therefore that the filter length is too large. Although higher order correlations may be tested for, our simulations indicate that a first-order correlation coefficient is sufficient for real-world signals. For brevity, we shall avoid the subscripts i,k by first defining lagged local residuals in the region around i, as described in Eq. (10) with $\mathbf{r}^0 = r[h] = \{R[(i-k),k],\ldots,R[(i+k-1),k]\}$ and $\mathbf{r}^1 = r[h] = \{R[(i-k+1),k],\ldots,R[(i+k),k]\}$. The first-order autocorrelation is calculated by substituting Pearson's correlation:

$$\psi[i,k] = g(\mathbf{r}^0,\mathbf{r}^1) = \frac{\sum(\mathbf{r}^0-\bar{\mathbf{r}}^0)(\mathbf{r}^1-\bar{\mathbf{r}}^1)}{\sqrt{\sum(\mathbf{r}^0-\bar{\mathbf{r}}^0)^2\sum(\mathbf{r}^1-\bar{\mathbf{r}}^1)^2}} \tag{11}$$

with $\bar{\mathbf{r}}^n$ denoting the mean of the nth lagged local residual vector.

Defining the subset $\{\psi[i,q]\}|\psi[i,q] \leqslant \psi^{\text{crit}}$ for all $\{q=1,\ldots,Q\}$ and fixed i, the optimal filter size is defined as $k^*[i] = Q-s$, where s is some small positive integer, and $\psi^{\text{crit}} = 0.3$, an established criterion r for this measure. The adaptive window SG filter (AWSGF) output $\hat{y}^*[i]$ is calculated from Eq. (5) as $\hat{y}^*[i] = \hat{y}[i,k^*[i]]$. It may be imagined that at each translation, the data is filtered using a gradually increasing filter, until the point at which first-order residual correlations become significant. The actual filter used is of length s less than the filter that produces significant residual correlations. Taking this approach produces a smoother and more reliable estimate of the optimal $k^*[i]$ than using a smaller ψ^{crit}, as for moderate sample sizes, correlation coefficients below 0.3 are somewhat unreliable. The following experiment used $s=10$.

Finally, we emphasize that above we have utilized a particular assumption regarding the noise process (no first-order correlation) and measured this using a particular statistical tool (Pearson's r). Depending on the application, a number of criterion may applied the task of differentiating between noise versus signal plus noise in the local residuals.

3. Results

The performance of the AWSGF was compared with that of the SGF using a synthetic signal with additive noise at varying signal-to-noise ratios (SNRs: 64, 32, 16). The synthetic signal was constructed from a sum of Gaussians of various dilations, sampled at 1024 points, with the general aim of constructing a signal typical for chemical and spectroscopic applications. Second-order Savitzky–Golay filters were used in each filter case. There is no standard way to select the filter length of the SGF, and it was desirable to discount the possibility of obtaining suboptimal SGF performance due to inappropriate choice of filter length. Thus the SGF output at all filter lengths was compared with the signal, and the best performing filter (between $k=11$ and 16, depending on SNR) was chosen as the SGF output. The AWSGF operated as usual, having access to only the noisy input signal. Since an ideal, rather than typical estimate of SGF performance was obtained, the result was a more rigorous test of the AWSGF. The performance of both filtering methods was assessed 10 times over each input noise condition using the SNR improvement (SNRI):

$$\text{SNRI} = 20\left(\log_{10}\frac{\sum(x[i]-\hat{y}^*[i])^2}{\sum(n[i])^2} - \log_{10}\frac{\sum(n[i])^2}{\sum(x[i])^2}\right). \tag{12}$$

Figure 2 illustrates the operation of the AWSGF on the noisy synthetic signal.

The mean SNRI performance and standard deviations are shown below in Table 1.

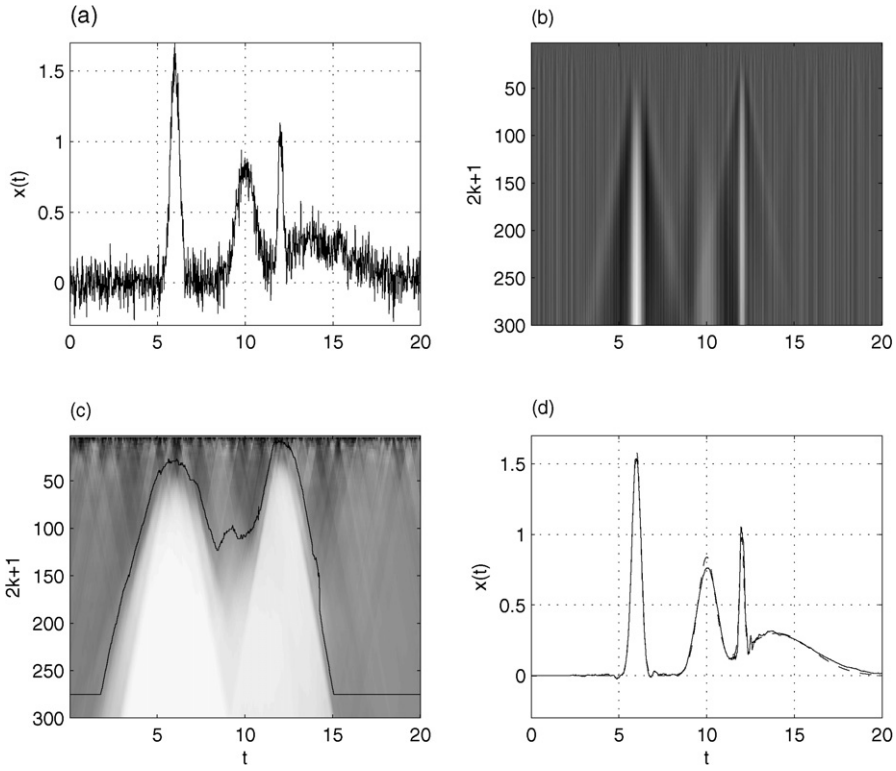

Fig. 2. AWSGF operation on synthetic noisy data. Displayed is the input signal (a), the matrix of residuals $R[i,k]$ (b), and the model adequacy first-order correlation matrix $\psi[i,k]$ (c) along with a line denoting $k^*[i]$. The AWSGF output $\hat{y}^*[i] = \hat{y}[i,k_i^*]$ (solid line) along with the noise-free input signal (dashed line) is shown in (d).

Table 1
Means and standard deviations (in italics) comparing the SNR improvement performance of standard and adaptive window Savitzky–Golay filters

	Initial SNR			
	64	32	16	
SGF	43.01	49.44	50.35	
	2.24	*2.10*	*3.42*	
AWSGF	56.30	58.83	50.46	
	4.50	*6.59*	*6.37*	$N = 10$

The results of a two-way analysis of variance (ANOVA) indicated that the strongest effect was the overall better performance of the AWSGF compared to that of SGF ($F = 41.28,\ p < 0.0001$). However, this effect was somewhat inconsistent over various SNRs, as was indicated by a significant interaction between SNR and filter type. Table 1 also shows that the performance of the AWSGF was more variable than the SGF.

4. Discussion

The present study compared an adaptive filter length approach with the conventional fixed length Savitzky–Golay filtering, in a context where the SGF would be expected to perform optimally. The results, while not overwhelming, indicate that adapting the length of the filter can produce better SNRI even in a context where the SGF already does very well.

Figure 2 illustrates a typical run of the AWSGF, and visual inspection of AWSGF behavior over a large number of runs indicated that first-order correlations are a reliable indicator of bias in a polynomial approximation. As Fig. 2c shows, the AWSGF used very long filter lengths at points where the input signal is characterized by slow changes, and

much shorter filter lengths in the regions of fast dynamics; changes in filter length were quite smooth and consistent. This resulted in very good noise reduction in flat regions, whilst introducing relatively little bias near the higher dynamic regions (Fig. 2d). However, close inspection of this figure shows that a certain degree of bias is still present in the AWSGF output.

From Table 1, it is clear than both the SGF and AWSGF produced a strong improvement in the SNR. Although the AWSGF did perform significantly better, the performance improvement may not have been as consistently marked due to: (a) the previously mentioned 'ideal' filter length determination for the SGF, (b) the strong performance of both systems on this particular input signal. Tests performed on other forms of input signals (such as those incorporating discontinuities), in which a filter with any fixed length will invariably produce a large degree of bias/small degree of SNRI, show a very strong improvement in the AWSGF over the SGF. Finally, the results of the current study indicate that high degrees of additive noise appeared to make it more difficult to detect first-order correlations.

The primary goal of the present study is to introduce and test a framework for adapting the length of the SGF in response to local data characteristics. It is not recommended that applied researchers replace the SGF by implementing the AWSGF in its current form. As with any data-adaptive approach, the behavior of AWSGF needs to be tested in a wide range of situations. Further work also needs to concentrate on efficient computational implementation, refinement of the model adequacy measure, and a number of advanced issues, such as non-central, non-symmetric filtering. Despite the computational and conceptual complexities introduced by an adaptive filter of this type, it clearly provide an effective approach for balancing the competing goals of bias and noise reduction, as they change over the local regions of the signal.

References

[1] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data with simplified least squares procedures, Anal. Chem. 36 (1964) 1627–1639.
[2] J.W. Luo, K. Ying, P. He, J. Bai, Properties of Savitzky–Golay digital differentiators, Digital Signal Process. 15 (2) (2004) 122–136.
[3] B.C.H. Turton, A novel variant of the Savitzky–Golay filter for spectroscopic applications, Meas. Sci. Technol. 3 (1992) 858–863.
[4] M. Jakubowska, W.W. Kubiak, Adaptive-degree polynomial filter for voltammetric signals, Anal. Chim. Acta 512 (2004) 241–250.
[5] H.H. Madden, Comments of Savitzky–Golay convolution method for least squares fit smoothing and differentiation of digital data, Anal. Chem. 50 (1978) 1383–1386.
[6] J. Steiner, Y. Termonia, J. Deltour, Comments on smoothing and differentiation of data by simplified least squares procedure, Anal. Chem. 44 (1972) 1906–1909.
[7] P.A. Gorry, General least-squares smoothing and differentiation of nonuniformly spaced data by the convolution method, Anal. Chem. 63 (1991) 534–536.
[8] R.A. Leach, C.A. Carter, J.M. Harris, Least-squares polynomial filters for initial point and slope estimation, Anal. Chem. 56 (1984) 2304–2307.
[9] P.A. Baedecker, Comments on least-square polynomial filters for initial point and slope estimation, Anal. Chem. 57 (1985) 1477–1479.
[10] P. Barak, Smoothing and differentiation by an adaptive-degree polynomial filter, Anal. Chem. 67 (1995) 2758–2762.
[11] I. Kawrakow, On the de-noising of Monte Carlo calculated dose distributions, Phys. Med. Biol. 47 (2002) 3087–3103.
[12] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C: The Art of Scientific Computing, second ed., Cambridge Univ. Press, Cambridge, UK, 1992.
[13] U. Bromba, H. Zlegler, Application hints for Savitzky–Golay digital smoothing filters, Anal. Chem. 53 (1981) 1583–1586.
[14] P.O. Persson, G. Strang, Smoothing by Savitzky–Golay and legendre filters, Comm. Comp. Finance 13 (2003) 301–316.
[15] P.A. Gorry, General least-squares smoothing and differentiation by the convolution (Savitzky–Golay) method, Anal. Chem. 62 (1990) 570–573.

Matthew Browne received the Ph.D. degree in psychophysiological signal processing from Griffith University, in 2001. He has undertaken post-doctoral work in autonomous robotics and image processing in Japan, Germany, and Australia. He has interest in biologically inspired approaches to intelligent computing and statistical signal processing. Most recently, he has been pursuing research in environmental informatics at Griffith University and the CSIRO.