



Time-domain analysis of the Savitzky–Golay filters

Quan Quan^{a,b,*}, Kai-Yuan Cai^a

^a Department of Automatic Control, Beihang University, Beijing 100191, China

^b State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China

ARTICLE INFO

Article history:

Available online 25 November 2011

Keywords:

Savitzky–Golay
Smoothing
Differentiation
Time domain

ABSTRACT

This paper proposes a time-domain method to analyze the estimation performance of Savitzky–Golay (SG) filters, which includes estimation performance of SG smoothing filters and SG differentiation filters. An estimation error bound is given and its qualitative properties are analyzed. Statistical properties of the estimation error in the presence of random noise are analyzed as well.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

The Savitzky–Golay (SG) filter was first proposed by Savitzky and Golay [1]. Since then, various modified SG filters have been proposed to improve smoothing and differentiation performance or to satisfy different requirements [2–9]. The properties of the SG smoothing filters have been well reviewed in [10–15]. However, as pointed out by Luo et al. in [16], the properties of first-order SG differentiation filters have not been extensively reported. The same is true of second-order SG differentiation filters. In [17, 18], an alternative way was offered to understand the properties of SG filters including the smoothing, first- and second-order differentiation filters. These motivate us to analyze the properties of SG differentiation filters of order determined by context (0th-order differentiation = smoothing). The digital differentiation filters are usually designed in the frequency domain [19–21] to satisfy various frequency-domain performance specifications. However, time-domain performance measures, such as the estimation error bound, are often difficult to obtain. This motivates us to conduct a time-domain analysis for SG differentiation filters.

In this paper, the SG differentiation filters are formulated first. In this formulation, the estimation error is decomposed into the approximation error and the uncertainty caused by random noise. Then some preliminary results are introduced. Among them, Lemma 2 provides conditions under which we can reconstruct the real signal perfectly from the sampling data by using the best approximation. With the help of a Taylor expansion and Lemma 2, we

obtain the approximation error bound in explicit form. Using this result, we further obtain an estimation error bound as the sum of the approximation error bound and the bound on uncertainty caused by random noise. Qualitative properties of the estimation error bound are analyzed as well. Finally, we analyze statistical properties of the estimation error caused by random noise. Here, we do not focus on the significance of SG differentiation filters or the calculation method of the filter coefficients. Instead, we focus on the analysis of the SG differentiation filters' properties, especially the relationships between the estimation error bound and various parameters, such as the filter length and the sampling period. Several results may be straightforward and already known in the existing literature. However, certain other results are interesting and believed to be new. More importantly, our approach offers an alternative way to understand the properties of widely-used SG differentiation filters in a unified framework. Specifically, the major contributions are: (i) the estimation performance of the SG differentiation filters is analyzed; (ii) a time-domain method is developed to analyze the estimation performance; (iii) a sufficient condition is proposed for perfect reconstruction from the sampling data using the best approximation.

Notation used in this paper is as follows. \mathbb{R}^n is Euclidean space of dimension n . \mathbb{N} is the set of nonnegative integers. $\mathbb{C}_q^n[a, b]$ denotes the space of continuous n -dimensional vector functions on $[a, b]$ which are q th-order differentiable. $\mathbf{x}^{(q)}(t) \triangleq \left[\frac{d^q x_1(t)}{dt^q} \quad \frac{d^q x_2(t)}{dt^q} \quad \dots \quad \frac{d^q x_n(t)}{dt^q} \right]^T \in \mathbb{R}^n$, $q \in \mathbb{N}$, where $\mathbf{x}(t) \triangleq [x_1(t) \ x_2(t) \ \dots \ x_n(t)]^T \in \mathbb{R}^n$. $\mathbf{A}_{i,j}$ denotes the element in the i th row and j th column of matrix \mathbf{A} . \mathbf{I}_n is the identity matrix with dimension n . $\text{tr}(\mathbf{A})$ represents the trace of matrix \mathbf{A} . $\|\cdot\|$ is defined as Euclidean norm or a matrix norm induced by Euclidean norm.

* Corresponding author at: Department of Automatic Control, Beihang University, Beijing 100191, China.

E-mail address: qq_buaa@asee.buaa.edu.cn (Q. Quan).

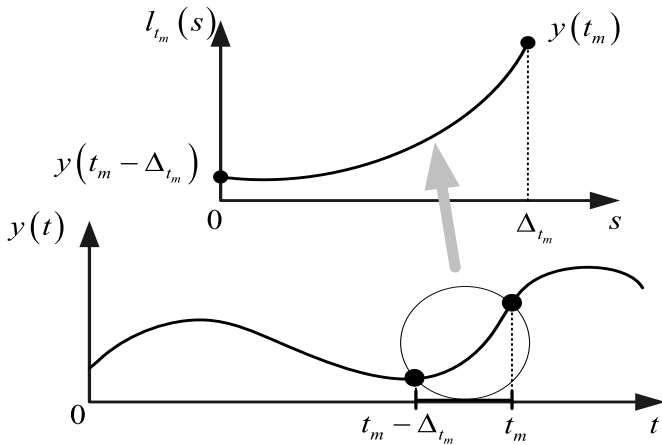


Fig. 1. The relationship between $l_{t_m}(s)$ and $y(t)$.

2. Problem formulation

2.1. n th-order SG differentiation filters formulation

Suppose $y(t) \in \mathbb{C}_{q_y}^1[0, \infty)$ is a noise-free signal to be considered. In practice, we can only observe the noisy signal $y_\xi(t)$ of the form

$$y_\xi(t) = y(t) + \xi(t)$$

where $\xi(t)$ is an unknown disturbance. By sampling $y_\xi(t)$ with sampling period T , the data are usually obtained at the sampling times t_k and denoted by $y_\xi(t_k)$, $k \in \mathbb{N}$. For given $q \in \mathbb{N}$, the estimate of $y^{(q)}(t)$ at $t \leq t_m$ is given by q th-order SG differentiation filters when the noisy sampling data $y_\xi(t_k)$, $t_k \leq t_m$ are measurable, where $q \leq q_y$.

The idea of SG differentiation filters is to utilize a polynomial to fit a segment of data points, then to differentiate the polynomial analytically up to the q th order, where $q \in \mathbb{N}$. In order to make the idea clear, the SG differentiation filters will be formulated next.

For clarity, we define a function $l_{t_m}(s)$ on the segment $[0, \Delta_{t_m}]$ as follows:

$$l_{t_m}(s) = y(t_m - \Delta_{t_m} + s) \quad (1)$$

where $t_m \geq \Delta_{t_m} > 0$, $m \in \mathbb{N}$. This is shown in Fig. 1.

Obviously,

$$l_{t_m}^{(q)}(s) = y^{(q)}(t_m - \Delta_{t_m} + s) \quad (2)$$

where $q \leq q_y$ and $q \in \mathbb{N}$. The segment $[0, \Delta_{t_m}]$ contains $N_{t_m} + 1$ sampling points and the corresponding data points are denoted by $\mathbf{l}_{t_m} + \boldsymbol{\xi}_{t_m}$, where

$$\mathbf{l}_{t_m} = [l_{t_m}(s_0) \quad l_{t_m}(s_1) \quad \cdots \quad l_{t_m}(s_{N_{t_m}})]^T \in \mathbb{R}^{N_{t_m}+1}, \quad (3)$$

$$\boldsymbol{\xi}_{t_m} = [\xi_{t_m}(s_0) \quad \xi_{t_m}(s_1) \quad \cdots \quad \xi_{t_m}(s_{N_{t_m}})]^T \in \mathbb{R}^{N_{t_m}+1}, \quad (4)$$

$s_k = iT$, $\Delta_{t_m} = N_{t_m}T$ and $i = 0, 1, \dots, N_{t_m} \in \mathbb{N}$. Without loss of generality, we only consider the estimation performance of SG differentiation filters for $y^{(q)}(t)$ at $t_m - \Delta_{t_m} \leq t \leq t_m$ when $\mathbf{l}_{t_m} + \boldsymbol{\xi}_{t_m}$ is known. For convenience, the subscript of Δ_{t_m} , N_{t_m} , \mathbf{l}_{t_m} , $\boldsymbol{\xi}_{t_m}$, l_{t_m} will be dropped in the following, i.e., $\Delta = \Delta_{t_m}$, $N = N_{t_m}$, $\mathbf{l} = \mathbf{l}_{t_m}$, $\boldsymbol{\xi} = \boldsymbol{\xi}_{t_m}$, $l = l_{t_m}$.

Based on the notation above, a detailed mathematical expression for SG differentiation filters is derived as follows:

(1) the function $l(s)$ on $[0, \Delta]$ will be reconstructed to be $\bar{\boldsymbol{\theta}}_\xi^T \mathbf{L}(s)$ by using the noisy data points $\mathbf{l} + \boldsymbol{\xi}$, where $\mathbf{L}(s) \triangleq [L_0(s) \quad L_1(s) \quad \cdots \quad L_{p-1}(s)]^T \in \mathbb{R}^p$, $L_k(s) = s^k$, $k = 0, 1, \dots, p-1$ and $\boldsymbol{\theta}_\xi \in \mathbb{R}^p$;

(2) the parameter $\bar{\boldsymbol{\theta}}_\xi$ is derived by solving the following optimization problem

$$\min_{\boldsymbol{\theta}} \sum_{i=0}^N [l(s_i) + \xi(s_i) - \boldsymbol{\theta}^T \mathbf{L}(s_i)]^2,$$

namely,

$$\min_{\boldsymbol{\theta}} \|(\mathbf{l} + \boldsymbol{\xi}) - \boldsymbol{\Psi} \boldsymbol{\theta}\|^2 \quad (5)$$

where $\boldsymbol{\Psi} = [\mathbf{L}(s_0) \quad \mathbf{L}(s_1) \quad \cdots \quad \mathbf{L}(s_N)]^T \in \mathbb{R}^{(N+1) \times p}$;

(3) then $l^{(q)}(s)$ will be estimated by $\bar{\boldsymbol{\theta}}_\xi^T \mathbf{L}^{(q)}(s)$.

If $\boldsymbol{\Psi}^T \boldsymbol{\Psi} \in \mathbb{R}^{p \times p}$ is nonsingular, then the optimal solution $\bar{\boldsymbol{\theta}}_\xi$ of (5) can be obtained explicitly as follows [22]:

$$\begin{aligned} \bar{\boldsymbol{\theta}}_\xi &= (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T (\mathbf{l} + \boldsymbol{\xi}) \\ &= \bar{\boldsymbol{\theta}} + (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \boldsymbol{\xi} \end{aligned}$$

where $\bar{\boldsymbol{\theta}} = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{l}$. In fact, $\bar{\boldsymbol{\theta}}$ can be defined as

$$\bar{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{l} - \boldsymbol{\Psi} \boldsymbol{\theta}\|. \quad (6)$$

At time t_m , for given \bar{s} , the estimate of $y^{(q)}(t_m - \Delta + \bar{s})$, denoted by $\hat{y}_\xi^{(q)}(t_m - \Delta + \bar{s}|t_m)$, can be expressed as

$$\hat{y}_\xi^{(q)}(t_m - \Delta + \bar{s}|t_m) = \bar{\boldsymbol{\theta}}_\xi^T \mathbf{L}^{(q)}(\bar{s}).$$

Substituting $\bar{\boldsymbol{\theta}}_\xi = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T (\mathbf{l} + \boldsymbol{\xi})$ into the equation above results in

$$\hat{y}_\xi^{(q)}(t_m - \Delta + \bar{s}|t_m) = \mathbf{h}^{(q)}(\bar{s})^T (\mathbf{l} + \boldsymbol{\xi}) \quad (7)$$

where

$$\mathbf{h}(s) \triangleq \boldsymbol{\Psi} (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \mathbf{L}(s) \in \mathbb{R}^{N+1}. \quad (8)$$

In particular, the estimate of the q th-order derivative at the midpoint, namely $\bar{s} = \frac{\Delta}{2}$, is

$$\hat{y}_\xi^{(q)}\left(t_m - \frac{\Delta}{2} \middle| t_m\right) = \mathbf{h}^{(q)}\left(\frac{\Delta}{2}\right)^T \mathbf{Y}_\xi(t_m) \quad (9)$$

where $\mathbf{Y}_\xi(t_m) = [y_\xi(t_m - \Delta) \quad y_\xi(t_m - s_{N-1}) \quad \cdots \quad y_\xi(t_m)]^T = \mathbf{l} + \boldsymbol{\xi}$.

The midpoint estimate $\hat{y}_\xi^{(q)}(t_m - \frac{\Delta}{2}|t_m)$ can be further represented by a transfer function as follows:

$$\hat{y}_\xi^{(q)}(z) = H_q(z) y_\xi(z)$$

where $H_q(z) = \sum_{k=0}^N h_{q,k}(\frac{\Delta}{2}) z^{-N+k}$ and $\mathbf{h}^{(q)}(\frac{\Delta}{2}) = [h_{q,0}(\frac{\Delta}{2}) \quad \cdots \quad h_{q,N}(\frac{\Delta}{2})]^T$. $H_q(z)$ is a FIR digital filter with tap coefficients $h_{q,k}(\frac{\Delta}{2})$, $k = 0, 1, \dots, N$.

Remark 1. The analysis above also allows $N + 1$ to be even. In fact, it does not restrict the parity of N . If $N + 1$ is even, then $\mathbf{h}(s)$ is still defined as in (8).

2.2. Objective

At time t_m , for an $\bar{s} \in [0, \Delta]$, the estimation error can be written as

$$e_{\xi,q}(t_m - \Delta + \bar{s}|t_m) \triangleq y^{(q)}(t_m - \Delta + \bar{s}) - \hat{y}_\xi^{(q)}(t_m - \Delta + \bar{s}|t_m).$$

Substituting (2) and (7) into the equation above yields

$$e_{\xi,q}(t_m - \Delta + \bar{s}|t_m) = \underbrace{l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top \mathbf{l}}_{\text{approximation error}} - \underbrace{\mathbf{h}^{(q)}(\bar{s})^\top \boldsymbol{\xi}}_{\text{caused by noise}}. \quad (10)$$

In the following sections, we focus on the bound and statistical properties of $e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)$. As seen in (10), the estimation error is decomposed into the approximation error $l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top \mathbf{l}$ and an uncertainty caused by random noise, i.e., $\mathbf{h}^{(q)}(\bar{s})^\top \boldsymbol{\xi}$. This will help to analyze the estimation error explicitly.

3. Preliminary results

To begin with, the following preliminary results are needed; the proofs of Lemmas 1–5 are in Appendix A.

Lemma 1. *If and only if $N + 1 \geq p$, then $\Psi^\top \Psi \in \mathbb{R}^{p \times p}$ is nonsingular.*

Lemma 2. *If $l(s) \in \Pi_L^{p-1}[0, \Delta] \triangleq \text{span}(L_0(s), L_1(s), \dots, L_{p-1}(s))$, $s \in [0, \Delta]$ and $N + 1 \geq p$, then $l(s) - (\Psi(\Psi^\top \Psi)^{-1} \mathbf{L}(s))^\top \mathbf{l} \equiv 0$, for all $s \in [0, \Delta]$.*

Remark 2. It is well known that the same sampling data may be obtained from two different continuous signals. This means that important time variations between sampling instants may be missed if the sampling data are used to reconstruct the real signal. Lemma 2 implies that if $l(s) \in \Pi_L^{p-1}[0, \Delta]$ and the number of sampling data in the segment $[0, \Delta]$ is larger than the polynomial degree, then the polynomial $l(s)$ can be reconstructed perfectly from the sampling data by using the best approximation.

Lemma 3. *If Δ is a positive constant, then $\lim_{T \rightarrow 0} (T \Psi^\top \Psi)^{-1} = \mathbf{G}^{-1}$, where $\mathbf{G}_{i,j} = \int_0^\Delta s^{i+j-2} ds$.*

Lemma 4. *$(\Psi^\top \Psi)^{-1} \Psi^\top \mathbf{w} = \mathbf{v}$, where $\mathbf{v} = [1 \ 0 \ \dots \ 0]^\top \in \mathbb{R}^p$, $\mathbf{w} = [1 \ 1 \ \dots \ 1]^\top \in \mathbb{R}^{N+1}$.*

Lemma 5. *$\lim_{N \rightarrow \infty} N^{2q} \|\bar{\Psi}(\bar{\Psi}^\top \bar{\Psi})^{-1} \mathbf{L}^{(q)}(\gamma N)\|^2 = 0$, where $\gamma \in [0, 1]$, $\bar{\Psi} = [\mathbf{L}(0) \ \mathbf{L}(1) \ \dots \ \mathbf{L}(N)]^\top \in \mathbb{R}^{(N+1) \times p}$.*

4. Estimation error analysis

In this section, we first give a bound on $e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)$. Based on it, the qualitative relationships between the estimation error bound and various parameters, such as Δ, N and T , are further analyzed. Then, statistical properties of $e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)$ are derived.

4.1. Estimation error bound

First, we give a bound on $e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)$ in the special case where $\xi(t) \equiv 0$, i.e., $y_\xi(t_k) = y(t_k)$. In this special case, $e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)$ is only the approximation error.

Theorem 1. *Suppose (i) $l(s) \in \mathbb{C}_{q_y}^1[0, \Delta]$ and $l_{\max,p} = \max_{s \in [0, \Delta]} \|l^{(p)}(s)\|$; (ii) $N + 1 \geq p$; (iii) $q < p \leq q_y$. Then for $\bar{s} \in [0, \Delta]$, we have*

$$\|l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top \mathbf{l}\| \leq B(\mathbf{L}, p, q, \Delta, T, l_{\max,p}, \bar{s})$$

where

$$B(\mathbf{L}, p, q, \Delta, T, l_{\max,p}, \bar{s}) \triangleq \frac{l_{\max,p}}{p!} \sqrt{T(\Delta + T)} \|\Psi(T \Psi^\top \Psi)^{-1} \mathbf{L}^{(q)}(\bar{s})\| \max_{s \in [0, \Delta]} \|s - \bar{s}\|^p.$$

Proof. See Appendix A.6. \square

With $\xi(t) \equiv 0$, the estimation error is just the approximation error. However, in practice, the estimation error is affected not only by the approximation error but also by random noise. This effect will be investigated in Theorem 2.

Theorem 2. *Suppose that the conditions of Theorem 1 are satisfied and $\max_{0 \leq k \leq N} \|\xi(s_k)\| \leq \delta$. Then*

$$\begin{aligned} & \|l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top (\mathbf{l} + \boldsymbol{\xi})\| \\ & \leq B(\mathbf{L}, p, q, \Delta, T, l_{\max,p}, \bar{s}) + B_\xi(\mathbf{L}, p, q, \Delta, T, \delta, \bar{s}) \end{aligned} \quad (11)$$

where

$$B_\xi(\mathbf{L}, p, q, \Delta, T, \delta, \bar{s}) = \delta \sqrt{T(\Delta + T)} \|\Psi(T \Psi^\top \Psi)^{-1} \mathbf{L}^{(q)}(\bar{s})\|. \quad (12)$$

Proof. See Appendix A.7. \square

With Theorem 2 in hand, we have

Corollary 1. *Suppose (i) $y(t) \in \mathbb{C}_{q_y}^1[0, \infty)$, $y_{\max,p} = \max_{t \in [0, \infty)} \|y^{(p)}(t)\|$ and $\sup_{t_k} \|\xi(t_k)\| \leq \delta$; (ii) $N + 1 \geq p$; (iii) $q < p \leq q_y$. Then for $\bar{s} \in [0, \Delta]$, we have $\|e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)\| \leq B(\mathbf{L}, p, q, \Delta, T, y_{\max,p}, \bar{s}) + B_\xi(\mathbf{L}, p, q, \Delta, T, \delta, \bar{s})$.*

The estimation error bound is a function of the variables $T, \Delta, N, p, q, \bar{s}, y_{\max,p}$ and δ . In a specific case, the estimation error may reach the upper bound. In order to further clarify the relationships between the estimation error bound and various parameters Δ, N, T , two qualitative properties of the estimation error bound are given next.

Property 1. *Suppose that the conditions of Corollary 1 are satisfied. If N remains constant and T tends to zero, i.e., Δ tends to zero, then $\|e_{\xi,0}(t_m - \Delta + \bar{s}|t_m)\|$ is bounded and the upper bound on $\|e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)\|$ is unbounded when $q \geq 1$.*

Proof. If $\bar{s} = \gamma NT$, $\gamma \in [0, 1]$, then the relationship between $\mathbf{L}^{(q)}(\bar{s})$ and $\mathbf{L}^{(q)}(\gamma N)$ is

$$\mathbf{L}^{(q)}(\bar{s}) = T^{-q} \mathbf{U} \mathbf{L}^{(q)}(\gamma N) \quad (13)$$

furthermore

$$\Psi = \bar{\Psi} \mathbf{U} \quad (14)$$

where $\mathbf{U} = \text{diag}(1, T, \dots, T^{p-1}) \in \mathbb{R}^{p \times p}$ and $\bar{\Psi}$ is defined in Lemma 5.

Using (13) and (14), we have

$$\|\Psi(\Psi^\top \Psi)^{-1} \mathbf{L}^{(q)}(\bar{s})\| = T^{-q} \|\bar{\Psi}(\bar{\Psi}^\top \bar{\Psi})^{-1} \mathbf{L}^{(q)}(\gamma N)\|. \quad (15)$$

Note that $\min_{\bar{s} \in [0, \Delta]} \max_{s \in [0, \Delta]} \|s - \bar{s}\|^p = (\frac{\Delta}{2})^p$, $\max_{\bar{s} \in [0, \Delta]} \max_{s \in [0, \Delta]} \|s - \bar{s}\|^p = \Delta^p$ and $\Delta = NT$. Then the bound on the approximation error is

$$\frac{1}{2^p} M_1 T^{p-q} \leq B(\mathbf{L}, p, q, \Delta, T, y_{\max,p}, \bar{s}) \leq M_1 T^{p-q} \quad (16)$$

where

$$M_1 = \frac{y_{\max,p}}{p!} \sqrt{N+1} \|\bar{\Psi}(\bar{\Psi}^\top \bar{\Psi})^{-1} \mathbf{L}^{(q)}(\gamma N)\| N^p. \quad (17)$$

Since $p - q > 0$ and M_1 is a constant as N remains constant, the approximation error bound $B(\mathbf{L}, p, q, \Delta, T, y_{\max,p}, \bar{s})$ tends to zero

as T tends to zero. By using (15) and $\Delta = NT$, the bound on the uncertainty (12) caused by random noise is

$$B_{\xi}(\mathbf{L}, p, q, \Delta, T, \delta, \bar{s}) = M_2 T^{-q} \quad (18)$$

where

$$M_2 = \delta \sqrt{N+1} \|\bar{\Psi}(\bar{\Psi}^T \bar{\Psi})^{-1} \mathbf{L}^{(q)}(\gamma N)\| \quad (19)$$

is a constant when N remains constant. Therefore, if N remains constant, then $M_2 T^{-q}$ will tend to infinity as T tends to zero except in the case $q = 0$. Therefore, if the conditions of Corollary 1 are satisfied, N remains constant and T tends to zero, then $\|e_{\xi,0}(t_m - \Delta + \bar{s}|t_m)\|$ is bounded and the upper bound on $\|e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)\|$ is unbounded when $q \geq 1$. \square

Remark 3. Property 1 shows that the approximation error will decrease as Δ decreases (refer to Eq. (16)). However, when random noise exists in the sampling data and N remains constant, the estimation error may be very large when Δ is small, and is more sensitive as the derivative order increases (refer to Eq. (18)). Therefore, in practice, if N is a small value, we cannot achieve good estimation performance using a small sampling period T . Since the estimation error is bounded by

$$\|e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)\| \leq M_1 T^{p-q} + M_2 T^{-q}$$

we can conclude that if $q = 0$ and N remains constant, then the smallest sample time T should be chosen. On the other hand, if $q \geq 1$ and N remains constant, then the optimal sampling period is

$$T^* = \arg \min_T (M_1 T^{p-q} + M_2 T^{-q}) = \left(\frac{q M_2}{(p-q) M_1} \right)^{\frac{1}{p}}$$

Substituting (17) and (19) into the above equation results in

$$T^* = \left(\frac{q \delta p!}{(p-q) l_{\max,p}} \right)^{\frac{1}{p}} \frac{1}{N}$$

From the equation above, we can observe that the optimal sampling period T^* decreases as N increases, and vice versa. Generally, if N can be chosen arbitrarily, then the lowest upper bound is always associated with the smallest sample time.

Property 2. Suppose the conditions of Corollary 1 are satisfied. If Δ remains constant and T tends to zero, i.e., N tends to infinity, then

$$\begin{aligned} & \lim_{T \rightarrow 0} \|e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)\| \\ & \leq \left(\delta + \frac{y_{\max,p}}{p!} \Delta^p \right) \sqrt{\Delta} \sqrt{\mathbf{L}^{(q)}(\bar{s})^T \mathbf{G}^{-1} \mathbf{L}^{(q)}(\bar{s})}. \end{aligned}$$

Proof. See Appendix A.8. \square

Remark 4. If Δ remains constant and T tends to zero, then the estimation error is bounded. The bound on the uncertainty caused by random noise, i.e., $B_{\xi}(\mathbf{L}, p, q, \Delta, T, \delta, \bar{s})$, may result in the estimation error bound being conservative. The reason is as follows. In the proof of Theorem 2, the first inequality of (30) (see Appendix A.7) has used

$$\|\mathbf{L}^{(q)}(\bar{s})^T (\Psi^T \Psi)^{-1} \Psi^T \xi\| \leq \|\Psi(\Psi^T \Psi)^{-1} \mathbf{L}^{(q)}(\bar{s})\| \|\xi\|. \quad (20)$$

When N is large enough, ξ will then show random characteristics so that ξ and $\Psi(\Psi^T \Psi)^{-1} \mathbf{L}^{(q)}(\bar{s})$ are linearly independent. The possibility of both sides of (20) being equal is very small, especially when N is large. Therefore, the inequality is strict in most cases. Generally, we may expect that

$$\|\mathbf{L}^{(q)}(\bar{s})^T (\Psi^T \Psi)^{-1} \Psi^T \xi\| \ll \|\Psi(\Psi^T \Psi)^{-1} \mathbf{L}^{(q)}(\bar{s})\| \|\xi\|.$$

Therefore, the upper bound of estimation error is conservative when N is large.

4.2. Statistical properties of the estimation error

According to Remark 4, statistical variables are proposed to describe the estimation error. This is more appropriate when N is large.

Theorem 3. Suppose the random noise $\xi(t_k)$ is an uncorrelated process, moreover, for all $k \in \mathbb{N}$, the expected value $E(\xi(t_k)) = \mu$ and the variance $D(\xi(t_k)) = \sigma^2$. Then

$$\begin{aligned} E(e_{\xi,0}(t_m - \Delta + \bar{s}|t_m)) &= e_0(t_m - \Delta + \bar{s}|t_m) - \mu, \\ E(e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)) &= e_q(t_m - \Delta + \bar{s}|t_m), \quad q \geq 1, \\ D(e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)) &= \|\Psi(\Psi^T \Psi)^{-1} \mathbf{L}^{(q)}(\bar{s})\|^2 \sigma^2 \end{aligned}$$

where $e_q(t_m - \Delta + \bar{s}|t_m)$ represents the approximation error, i.e., $e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)$ with $\xi(t) \equiv 0$.

Proof. See Appendix A.9. \square

Remark 5. If the expected value of the random noise is a constant, then the expected value of the estimation error of smoothing is equal to the approximation error minus the constant, and the expected value of the estimation error of the q th-order derivative ($q \geq 1$) is equal to the approximation error.

The following Property 3 shows the qualitative relationships between the variance of the estimation error and the variables T, Δ, N .

Property 3. Suppose the conditions of Corollary 1 and Theorem 3 are satisfied. (i) If Δ remains constant and T tends to zero, then

$$\lim_{T \rightarrow 0} D(e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)) = 0;$$

(ii) if T remains constant and N tends to infinity, then

$$\lim_{N \rightarrow \infty} D(e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)) = 0.$$

Proof. See Appendix A.10. \square

Remark 6. The variance of the estimation error tends to zero as the number of sampling data used to fit the polynomial, i.e., the filter length, tends to infinity. In other words, the uncertainty caused by random noise will be reduced as the length of the filter, i.e. $N + 1$, increases. However, when the sampling period is fixed, a large number of sampling data results in a large value of Δ which in turn gives a large approximation error.

Fig. 2 shows the behavior of expected value bound and variance of estimation error as N increases, when $y_{\max,p} = 1, p = 2, T = 0.01$ and $q = 0$. Hence, N should be selected appropriately to achieve a tradeoff between the approximation error and the uncertainty caused by random noise.

5. Conclusions

The estimation performance of the n th-order SG differentiation filters is analyzed in the time domain. We obtain the following

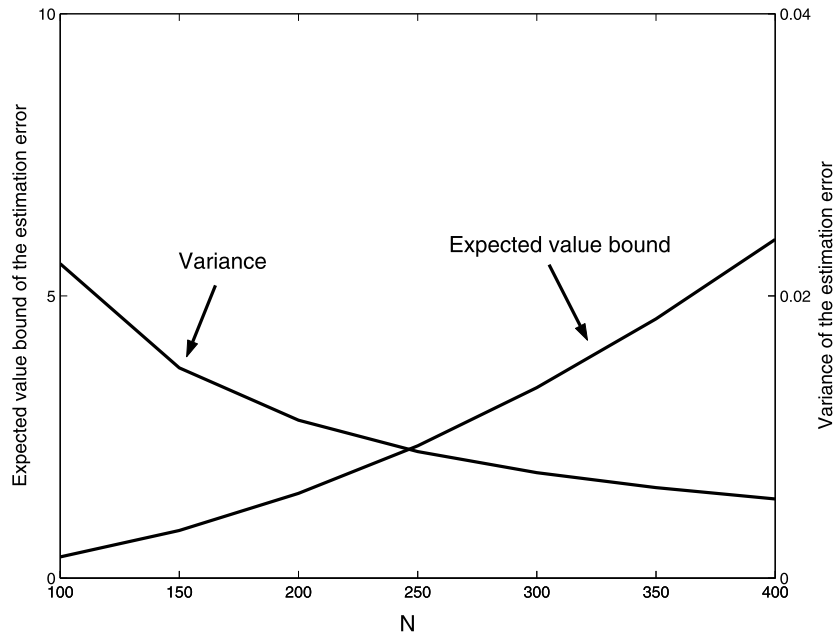


Fig. 2. The behavior of expected value bound and variance of estimation error as N increases.

conclusions: (i) when the filter length remains constant, the approximation error will decrease as the sampling period decreases; (ii) when random noise exists in the sampling data and the filter length remains constant, the estimation error may be very large when the sampling period is small, and is more sensitive as the derivative order increases; (iii) the optimal sampling period satisfies $T^* \propto (\frac{qp!}{(p-q)})^{\frac{1}{p}} \frac{1}{N}$; generally, if N can be chosen arbitrarily, then the lowest upper bound is always associated with the smallest sample time; (iv) if the time interval used for the estimate remains constant and the sampling period tends to zero, then the estimation error of the q th-order differentiation is bounded ($q \geq 0$); (v) if the expected value of the random noise is constant, then the expected value of the estimation error of the q th-order derivative is equal to the approximation error except in the case $q = 0$: the expected value of the estimation error of smoothing is equal to the approximation error minus the constant; (vi) the variance of the estimation error will be reduced as the filter length increases.

Acknowledgments

This work was supported by the 973 Program (2010CB327904), the National Natural Science Foundation of China (61104012), and “Weishi” Young Teachers Talent Cultivation Foundation of BUAA (YWF-11-03-Q-013).

Appendix A

A.1. Proof of Lemma 1

Obviously, when $N + 1 < p$, matrix Ψ is not of full column rank. When $N + 1 = p$, matrix Ψ is a Vandermonde matrix. Since s_0, s_1, \dots, s_N are different values, we have $\det(\Psi) = \prod_{0 \leq k_1 < k_2 \leq N} (s_{k_2} - s_{k_1}) \neq 0$. Then Ψ is of full column rank. Consequently, Ψ is still of full column rank when $N + 1 > p$. Since $\Psi^T \Psi$ is nonsingular $\Leftrightarrow \Psi$ is of full column rank, hence $\Psi^T \Psi \in \mathbb{R}^{p \times p}$ is nonsingular if and only if $N + 1 \geq p$ holds.

A.2. Proof of Lemma 2

First, we let $\bar{\theta} = \arg \min_{\theta} (\max_{s \in [0, \Delta]} \|l(s) - \bar{\theta}^T L(s)\|)$. Since $l(s) \in \Pi_L^{p-1}[0, \Delta]$, we have $\max_{s \in [0, \Delta]} \|l(s) - \bar{\theta}^T L(s)\| \equiv 0$ by the definition of $\bar{\theta}$. Then $\|l(s_k) - \bar{\theta}^T L(s_k)\| = 0, k = 0, 1, \dots, N$, i.e., $\|l - \Psi \bar{\theta}\| = 0$. By the definition of $\bar{\theta}$ in (6), we obtain $0 \leq \|l - \Psi \bar{\theta}\| \leq \|l - \Psi \bar{\theta}\| = 0$, thus $\|l - \Psi \bar{\theta}\| = 0$. Next proves $\bar{\theta} = \bar{\theta}$. Since $l - \Psi \bar{\theta} = l - \Psi \bar{\theta} = 0$, we obtain $\Psi(\bar{\theta} - \bar{\theta}) = 0$. Note that matrix Ψ is of full column rank when $N + 1 \geq p$ by Lemma 1, hence $\bar{\theta} = \bar{\theta}$. This gives

$$\begin{aligned} l(s) - \bar{\theta}^T L(s) &= l(s) - \bar{\theta}^T L(s) \\ &= l(s) - (\Psi(\Psi^T \Psi)^{-1} L(s))^T l \equiv 0, \end{aligned}$$

for all $s \in [0, \Delta]$.

A.3. Proof of Lemma 3

Since $s_k = kT$, every element of matrix $T \Psi^T \Psi$ is

$$(T \Psi^T \Psi)_{i,j} = \sum_{k=0}^{\lfloor \frac{\Delta}{T} \rfloor} L_{i-1}(kT) L_{j-1}(kT) T$$

where $\lfloor \frac{\Delta}{T} \rfloor$ rounds $\frac{\Delta}{T}$ toward zero and $L_k(s) = s^k, k = 0, 1, \dots, p - 1$.

Since Δ remains constant, hence the following equation holds as T tends to zero,

$$\begin{aligned} \lim_{T \rightarrow 0} (T \Psi^T \Psi)_{i,j} &= \lim_{T \rightarrow 0} \left(\sum_{k=0}^{\lfloor \frac{\Delta}{T} \rfloor} L_{i-1}(kT) L_{j-1}(kT) T \right) \\ &= \mathbf{G}_{i,j} \end{aligned}$$

where $\mathbf{G}_{i,j} = \int_0^\Delta L_{i-1}(s) L_{j-1}(s) ds = \int_0^\Delta s^{i+j-2} ds$. Since $\int_0^\Delta L_{i-1}(s) L_{j-1}(s) ds$ is an inner product of L_{i-1} and L_{j-1} , hence $\mathbf{G} \in \mathbb{R}^{p \times p}$, called the Gram matrix of $L_0(s), L_1(s), \dots, L_{p-1}(s)$, is nonsingular by [22, p. 56]. Therefore, we have $\lim_{T \rightarrow 0} (T \Psi^T \Psi)^{-1} = \mathbf{G}^{-1}$.

A.4. Proof of Lemma 4

Noticing the definition of Ψ , we have

$$\Psi^T \Psi \mathbf{v} = \left(\sum_{k=0}^N \mathbf{L}(s_k) \mathbf{L}(s_k)^T \right) \mathbf{v}.$$

Since $\mathbf{L}(s_k)^T \mathbf{v} = 1$ for $k = 0, \dots, N$, we have

$$\begin{aligned} \Psi^T \Psi \mathbf{v} &= \sum_{k=0}^N \mathbf{L}(s_k) (\mathbf{L}(s_k)^T \mathbf{v}) \\ &= \sum_{k=0}^N \mathbf{L}(s_k) = \Psi^T \mathbf{w} \end{aligned}$$

hence $(\Psi^T \Psi)^{-1} \Psi^T \mathbf{w} = \mathbf{v}$.

A.5. Proof of Lemma 5

Matrix $\tilde{\Psi}$ is rewritten as $\tilde{\Psi} = [\boldsymbol{\rho}(1) \ \boldsymbol{\rho}(2) \ \dots \ \boldsymbol{\rho}(p)]$, where $\boldsymbol{\rho}(k) = [0^{k-1} \ 1^{k-1} \ \dots \ N^{k-1}]^T \in \mathbb{R}^{N+1}$, $k = 0, \dots, N$. When N is large enough, every element of $\tilde{\Psi}^T \tilde{\Psi}$ is

$$\begin{aligned} (\tilde{\Psi}^T \tilde{\Psi})_{i,j} &= \boldsymbol{\rho}(i)^T \boldsymbol{\rho}(j) = \sum_{k=0}^N k^{i+j-2} \\ &= N^{i+j-1} \sum_{k=0}^N \left(\frac{k}{N} \right)^{i+j-2} \frac{1}{N} \\ &= N^{i+j-1} \mathbf{Q}_{i,j} \end{aligned}$$

where $\mathbf{Q}_{i,j} = (\tilde{\mathbf{G}})_{i,j} + o(N^{-1})$, $o(\cdot)$ presents the higher-order infinitesimal and $(\tilde{\mathbf{G}})_{i,j} = \int_0^1 s^{i+j-2} ds$. Therefore

$$\tilde{\Psi}^T \tilde{\Psi} = \mathbf{F} \mathbf{Q} \mathbf{F}$$

where $\mathbf{Q} \in \mathbb{R}^{p \times p}$ and $\mathbf{F} = \text{diag}([N^{\frac{2 \cdot 1 - 1}{2}} \ N^{\frac{2 \cdot 2 - 1}{2}} \ \dots \ N^{\frac{2 \cdot p - 1}{2}}]^T)$.

Since $\tilde{\mathbf{G}} \in \mathbb{R}^{p \times p}$, called the Gram matrix of $1, s, \dots, s^{p-1}$, is nonsingular by [22, p. 56], we have $(\mathbf{Q}^{-1})_{i,j} = (\tilde{\mathbf{G}}^{-1})_{i,j} + o(N^{-1})$. Consequently, every element of $(\tilde{\Psi}^T \tilde{\Psi})^{-1}$ is

$$((\tilde{\Psi}^T \tilde{\Psi})^{-1})_{i,j} = N^{-i-j+1} o(1). \quad (21)$$

Moreover,

$$(\mathbf{L}^{(q)}(\gamma N))_{k,1} = \begin{cases} 0 & k - 1 - q < 0, \\ N^{k-1-q} o(1) & k - 1 - q \geq 0. \end{cases} \quad (22)$$

Using (21) and (22), we obtain

$$\mathbf{L}^{(q)}(\gamma N)^T (\tilde{\Psi}^T \tilde{\Psi})^{-1} \mathbf{L}^{(q)}(\gamma N) = N^{-1-2q} o(1). \quad (23)$$

Since

$$\|\tilde{\Psi} (\tilde{\Psi}^T \tilde{\Psi})^{-1} \mathbf{L}^{(q)}(\gamma N)\|^2 = \mathbf{L}^{(q)}(\gamma N)^T (\tilde{\Psi}^T \tilde{\Psi})^{-1} \mathbf{L}^{(q)}(\gamma N)$$

substituting (23) into the inequality above yields

$$\|\tilde{\Psi} (\tilde{\Psi}^T \tilde{\Psi})^{-1} \mathbf{L}^{(q)}(\gamma N)\|^2 \leq N^{-1-2q} o(1).$$

Therefore $\lim_{N \rightarrow \infty} N^{2q} \|\tilde{\Psi} (\tilde{\Psi}^T \tilde{\Psi})^{-1} \mathbf{L}^{(q)}(\gamma N)\|^2 = 0$.

A.6. Proof of Theorem 1

The Taylor expansion of $l(s)$ about the point $\bar{s} \in [0, \Delta]$ is

$$l(s) = g(s) + r(\lambda, s) \quad (24)$$

where

$$g(s) = \sum_{k=0}^{p-1} \frac{l^{(k)}(\bar{s})}{k!} (s - \bar{s})^k, \quad r(\lambda, s) = \frac{l^{(p)}(\bar{s} + \lambda(s - \bar{s}))}{p!} (s - \bar{s})^p,$$

$$\lambda \in [0, 1], \quad s \in [0, \Delta].$$

According to (24), \mathbf{l} in (3) can be written as

$$\mathbf{l} = \mathbf{g} + \mathbf{r}(\lambda) \quad (25)$$

where

$$\begin{aligned} \mathbf{g} &= [g(s_0) \ g(s_1) \ \dots \ g(s_N)]^T \in \mathbb{R}^{N+1}, \\ \mathbf{r}(\lambda) &= [r(\lambda_0, s_0) \ r(\lambda_1, s_1) \ \dots \ r(\lambda_N, s_N)]^T \in \mathbb{R}^{N+1}, \\ \lambda &= [\lambda_0 \ \lambda_1 \ \dots \ \lambda_N]^T \in \mathbb{R}^{N+1}, \quad \lambda_0, \lambda_1, \dots, \lambda_N \in [0, 1]. \end{aligned}$$

Since $N + 1 \geq p$ and $g(s) \in \Pi_L^{p-1}[0, \Delta]$, we have

$$g(s) - (\Psi (\Psi^T \Psi)^{-1} \mathbf{L}(s))^T \mathbf{g} = 0, \quad \text{for all } s \in [0, \Delta]$$

by Lemma 2. Thus

$$\begin{aligned} l(s) - \mathbf{h}(s)^T \mathbf{l} &= g(s) + r(\lambda, s) - (\Psi (\Psi^T \Psi)^{-1} \mathbf{L}(s))^T (\mathbf{g} + \mathbf{r}(\lambda)) \\ &= r(\lambda, s) + \mathbf{r}(\lambda)^T \Psi (\Psi^T \Psi)^{-1} \mathbf{L}(s) \end{aligned} \quad (26)$$

where (8) and (25) are utilized. Note the definition of $r(\lambda, s)$, we obtain $r^{(q)}(\lambda, \bar{s}) \equiv 0$ for all $\lambda \in [0, 1]$, where $r^{(q)}(\lambda, \bar{s}) = \frac{d^q r(\lambda, s)}{ds^q} |_{s=\bar{s}}$ and $q < p$. Therefore for the $\bar{s} \in [0, \Delta]$, Eq. (26) becomes

$$\begin{aligned} l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^T \mathbf{l} &= r^{(q)}(\lambda, \bar{s}) + \mathbf{r}(\lambda)^T \Psi (\Psi^T \Psi)^{-1} \mathbf{L}^{(q)}(\bar{s}) \\ &= \mathbf{r}(\lambda)^T \Psi (\Psi^T \Psi)^{-1} \mathbf{L}^{(q)}(\bar{s}) \end{aligned}$$

where $q < p$. Taking norm $\|\cdot\|$ on both sides of the above equation yields

$$\begin{aligned} \|l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^T \mathbf{l}\| &\leq \left(\max_{\lambda_0, \lambda_1, \dots, \lambda_N \in [0, 1]} \|\mathbf{r}(\lambda)\| \right) \|\Psi (\Psi^T \Psi)^{-1} \mathbf{L}^{(q)}(\bar{s})\|. \end{aligned} \quad (27)$$

Since $l_{\max, p} = \max_{s \in [0, \Delta]} \|l^{(p)}(s)\|$, hence

$$\begin{aligned} \|r(\lambda, \bar{s})\| &= \left\| \frac{l^{(p)}(\bar{s} + \lambda(s - \bar{s}))}{p!} (s - \bar{s})^p \right\| \\ &\leq \frac{l_{\max, p}}{p!} \max_{s \in [0, \Delta]} \|s - \bar{s}\|^p. \end{aligned}$$

Consequently

$$\begin{aligned} \max_{\lambda \in [0, 1]} \|\mathbf{r}(\lambda)\| &\leq \sqrt{N+1} \max_{\lambda \in [0, 1]} \|r(\lambda, \bar{s})\| \\ &\leq \sqrt{N+1} \frac{l_{\max, p}}{p!} \max_{s \in [0, \Delta]} \|s - \bar{s}\|^p. \end{aligned} \quad (28)$$

Substituting (28) into the inequality (27) yields

$$\begin{aligned} \|l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^T \mathbf{l}\| &\leq \frac{l_{\max, p}}{p!} \sqrt{N+1} \|\Psi (\Psi^T \Psi)^{-1} \mathbf{L}^{(q)}(\bar{s})\| \max_{s \in [0, \Delta]} \|s - \bar{s}\|^p \\ &= B(\mathbf{L}, p, q, \Delta, T, l_{\max, p}, \bar{s}), \end{aligned}$$

where $\Delta = NT$ has been utilized. \square

A.7. Proof of Theorem 2

For an $\bar{s} \in [0, \Delta]$, using (8), we have

$$l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top (\mathbf{I} + \boldsymbol{\xi}) = l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top \mathbf{I} - \mathbf{L}^{(q)}(\bar{s})^\top (\boldsymbol{\Psi}^\top \boldsymbol{\Psi}) \boldsymbol{\Psi}^\top \boldsymbol{\xi}. \quad (29)$$

Since $\max_{0 \leq k \leq N} \|\xi(s_k)\| \leq \delta$, hence $\|\boldsymbol{\xi}\| \leq \delta\sqrt{N+1}$. Taking norm $\|\cdot\|$ on both sides of Eq. (29) yields

$$\begin{aligned} & \|l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top (\mathbf{I} + \boldsymbol{\xi})\| \\ & \leq \|l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top \mathbf{I}\| + \|\boldsymbol{\Psi}(\boldsymbol{\Psi}^\top \boldsymbol{\Psi})^{-1} \mathbf{L}^{(q)}(\bar{s})\| \|\boldsymbol{\xi}\| \\ & \leq \|l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top \mathbf{I}\| + B_\xi(\mathbf{L}, p, q, \Delta, T, \delta, \bar{s}). \end{aligned} \quad (30)$$

Therefore, (11) is satisfied by Theorem 1.

A.8. Proof of Property 2

Since Δ remains constant and N tends to infinity, hence

$$\begin{aligned} & \lim_{T \rightarrow 0} \sqrt{T(\Delta + T)} \|\boldsymbol{\Psi}(T\boldsymbol{\Psi}^\top \boldsymbol{\Psi})^{-1} \mathbf{L}^{(q)}(\bar{s})\| \\ & = \lim_{T \rightarrow 0} \sqrt{\Delta + T} \sqrt{\mathbf{L}^{(q)}(\bar{s})^\top (T\boldsymbol{\Psi}^\top \boldsymbol{\Psi})^{-1} \mathbf{L}^{(q)}(\bar{s})} \\ & = \sqrt{\Delta} \sqrt{\mathbf{L}^{(q)}(\bar{s})^\top \mathbf{G}^{-1} \mathbf{L}^{(q)}(\bar{s})} \end{aligned}$$

by Lemma 3. Consequently,

$$\begin{aligned} & \lim_{T \rightarrow 0} B(\mathbf{L}, p, q, \Delta, T, y_{\max, p}, \bar{s}) \\ & \leq \frac{y_{\max, p}}{p!} \Delta^p \sqrt{\Delta} \sqrt{\mathbf{L}^{(q)}(\bar{s})^\top \mathbf{G}^{-1} \mathbf{L}^{(q)}(\bar{s})} \end{aligned}$$

and

$$\lim_{T \rightarrow 0} B_\xi(\mathbf{L}, p, q, \Delta, T, \delta, \bar{s}) \leq \delta \sqrt{\Delta} \sqrt{\mathbf{L}^{(q)}(\bar{s})^\top \mathbf{G}^{-1} \mathbf{L}^{(q)}(\bar{s})}.$$

Combining the two inequalities above concludes the proof.

A.9. Proof of Theorem 3

In light of (29), since $E(\boldsymbol{\xi}) = \mu \mathbf{w}$, where $\mathbf{w} = [1 \ 1 \ \dots \ 1]^\top \in \mathbb{R}^{N+1}$, then

$$\begin{aligned} & E(l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top (\mathbf{I} + \boldsymbol{\xi})) \\ & = l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top \mathbf{I} - \mathbf{L}^{(q)}(\bar{s})^\top (\boldsymbol{\Psi}^\top \boldsymbol{\Psi}) \boldsymbol{\Psi}^\top E(\boldsymbol{\xi}) \\ & = l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top \mathbf{I} - \mu \mathbf{L}^{(q)}(\bar{s})^\top (\boldsymbol{\Psi}^\top \boldsymbol{\Psi}) \boldsymbol{\Psi}^\top \mathbf{w}. \end{aligned}$$

Since $(\boldsymbol{\Psi}^\top \boldsymbol{\Psi}) \boldsymbol{\Psi}^\top \mathbf{w} = \mathbf{v}$ by Lemma 4, the equation above becomes

$$E(l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top (\mathbf{I} + \boldsymbol{\xi})) = l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top \mathbf{I} - \mu \mathbf{L}^{(q)}(\bar{s})^\top \mathbf{v}$$

where $\mathbf{v} = [1 \ 0 \ \dots \ 0]^\top \in \mathbb{R}^p$.

Note that $\mathbf{L}^{(0)}(\bar{s})^\top \mathbf{v} = 1$ and $\mathbf{L}^{(q)}(\bar{s})^\top \mathbf{v} = 0, q \geq 1$, hence

$$\begin{aligned} E(e_{\xi,0}(t_m - \Delta + \bar{s}|t_m)) & = E(l(\bar{s}) - \mathbf{h}(\bar{s})^\top (\mathbf{I} + \boldsymbol{\xi})) \\ & = e_0(t_m - \Delta + \bar{s}|t_m) - \mu, \\ E(e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)) & = E(l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top (\mathbf{I} + \boldsymbol{\xi})) \\ & = e_q(t_m - \Delta + \bar{s}|t_m), \quad q \geq 1. \end{aligned}$$

Since the random noise $\xi(t_k)$ is an uncorrelated process, the variance of $e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)$ is

$$\begin{aligned} D(e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)) & = D(l^{(q)}(\bar{s}) - \mathbf{h}^{(q)}(\bar{s})^\top (\mathbf{I} + \boldsymbol{\xi})) \\ & = D(\mathbf{L}^{(q)}(\bar{s})^\top (\boldsymbol{\Psi}^\top \boldsymbol{\Psi}) \boldsymbol{\Psi}^\top \boldsymbol{\xi}) \\ & = \|\boldsymbol{\Psi}(\boldsymbol{\Psi}^\top \boldsymbol{\Psi})^{-1} \mathbf{L}^{(q)}(\bar{s})\|^2 \sigma^2. \end{aligned}$$

A.10. Proof of Property 3

If Δ remains constant and T tends to zero, i.e., N tends to infinity, then

$$\begin{aligned} & \lim_{T \rightarrow 0} D(e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)) \\ & = \lim_{T \rightarrow 0} T^2 \|\boldsymbol{\Psi}(T\boldsymbol{\Psi}^\top \boldsymbol{\Psi})^{-1} \mathbf{L}^{(q)}(\bar{s})\|^2 \sigma^2 \\ & = \lim_{T \rightarrow 0} T \mathbf{L}^{(q)}(\bar{s})^\top (T\boldsymbol{\Psi}^\top \boldsymbol{\Psi})^{-1} \mathbf{L}^{(q)}(\bar{s}) \sigma^2 \\ & = \sigma^2 \lim_{T \rightarrow 0} T \lim_{T \rightarrow 0} \mathbf{L}^{(q)}(\bar{s})^\top (T\boldsymbol{\Psi}^\top \boldsymbol{\Psi})^{-1} \mathbf{L}^{(q)}(\bar{s}). \end{aligned}$$

By Lemma 3, we obtain

$$\begin{aligned} \lim_{T \rightarrow 0} D(e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)) & = \sigma^2 \mathbf{L}^{(q)}(\bar{s})^\top \mathbf{G}^{-1} \mathbf{L}^{(q)}(\bar{s}) \sigma^2 \lim_{T \rightarrow 0} T \\ & = 0. \end{aligned}$$

Recalling (15), we have

$$D(e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)) = \sigma^2 T^{-q} \|\bar{\boldsymbol{\Psi}}(\bar{\boldsymbol{\Psi}}^\top \bar{\boldsymbol{\Psi}})^{-1} \mathbf{L}^{(q)}(\gamma N)\|.$$

If T remains constant and N tends to infinity, then

$$\lim_{N \rightarrow \infty} D(e_{\xi,q}(t_m - \Delta + \bar{s}|t_m)) = 0$$

by Lemma 5.

References

- [1] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* 36 (1964) 1627–1639.
- [2] K.L. Ratzlaff, J.T. Johnson, Computation of two-dimensional polynomial least-squares convolution smoothing integers, *Anal. Chem.* 61 (1989) 1303–1305.
- [3] P.A. Gorry, General least-squares smoothing and differentiation by the convolution (Savitzky–Golay) method, *Anal. Chem.* 62 (1990) 570–573.
- [4] J.E. Kuo, H. Wang, S. Pickup, Multidimensional least-squares smoothing using orthogonal polynomials, *Anal. Chem.* 63 (1991) 630–635.
- [5] P. Barak, Smoothing and differentiation by an adaptive-degree polynomial filter, *Anal. Chem.* 67 (1995) 2758–2762.
- [6] J. Luo, K. Ying, J. Bai, Savitzky–Golay smoothing and differentiation filter for even number data, *Signal Process.* 85 (7) (2005) 1429–1434.
- [7] M. Browne, N. Mayer, T.R.H. Cutmore, A multiscale polynomial filter for adaptive smoothing, *Digital Signal Process.* 17 (2007) 69–75.
- [8] G. Viv-Truyols, P. Schoenmakers, Automatic selection of optimal Savitzky–Golay smoothing, *Anal. Chem.* 78 (2007) 4598–4608.
- [9] H. Hassanpour, A time–frequency approach for noise reduction, *Digital Signal Process.* 18 (2008) 728–738.
- [10] H.W. Schüssler, P. Steffen, Some advanced topics in filter design, in: J.S. Lim, A.V. Oppenheim (Eds.), *Advanced Topics in Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1988.
- [11] C.G. Enke, T.A. Nieman, Signal-to-noise ratio enhancement by least-squares polynomial smoothing, *Anal. Chem.* 48 (1976) A705–A712.
- [12] G.R. Phillips, J.M. Harris, Predicting the statistical properties of least-squares polynomial filters, *Anal. Chem.* 62 (1990) 2749–2752.
- [13] H. Ziegler, Properties of digital smoothing polynomial (DISPO) filters, *Appl. Spectr.* 35 (1981) 88–92.
- [14] T.H. Edwards, P.D. Willson, Digital least-squares smoothing of spectra, *Appl. Spectr.* 28 (1974) 541–545.
- [15] P.D. Willson, T.H. Edwards, Sampling and smoothing of spectra, *Appl. Spectr. Rev.* 12 (1976) 1–81.
- [16] J. Luo, K. Ying, P. He, J. Bai, Properties of Savitzky–Golay digital differentiators, *Digital Signal Process.* 15 (2005) 122–136.
- [17] K. Kitamura, K. Hozumi, Effect of the degree of polynomials in the Savitzky–Golay method for calculation of second-derivative spectra, *Anal. Chim. Acta* 172 (1985) 111–118.
- [18] K. Kitamura, K. Hozumi, Effect of Savitzky–Golay smoothing on second-derivative spectra, *Anal. Chim. Acta* 201 (1987) 301–304.
- [19] S. Sunder, R.P. Ramachandran, Least-squares design of higher order nonrecursive differentiators, *IEEE Trans. Signal Process.* 42 (1994) 956–961.
- [20] G. Mollova, R. Unbehauen, Analytical design of higher-order differentiators using least-squares technique, *Electron. Lett.* 37 (2001) 1098–1099.
- [21] Y.-D. Jou, Neural least-squares design of higher order digital differentiators, *IEEE Signal Process. Lett.* 12 (2005) 9–12.
- [22] D.G. Luenberger, *Optimization by Vector Space Methods*, Wiley, New York, 1968.



Quan Quan is an assistant professor at the School of Automation Science and Electrical Engineering, Beihang University (Beijing University of Aeronautics and Astronautics) since 2010. He received his B.S. degree in 2004 and Ph.D. degree in 2010, both from Beihang University. He was a research fellow at National University of Singapore from 7/2011 to 9/2011. His research interests cover flight control and vision based navigation and control.



Kai-Yuan Cai is a Cheung Kong Scholar (Chair Professor), jointly appointed by the Ministry of Education of China and the Li Ka Shing Foundation of Hong Kong in 1999. He has been a full professor at Beihang University (Beijing University of Aeronautics and Astronautics) since 1995. He was born in

April 1965 and entered Beihang University as an undergraduate student in 1980. He received his B.S. degree in 1984, M.S. degree in 1987, and Ph.D. degree in 1991, all from Beihang University. He was a research fellow at City University, London, and a visiting scholar at Purdue University (USA). He was also a Visiting Professorial Fellow with the University of Wollongong, Australia. Dr. Cai has published many research papers in international journals and is the author of three books: *Software Defect and Operational Profile Modeling* (Kluwer, Boston, 1998); *Introduction to Fuzzy Reliability* (Kluwer, Boston, 1996); *Elements of Software Reliability Engineering* (Tshinghua University Press, Beijing, 1995, in Chinese). He serves on the editorial board of the international journal *Fuzzy Sets and Systems*. He also served as guest editor for *Fuzzy Sets and Systems* (1996), the *International Journal of Software Engineering and Knowledge Engineering* (2006), the *Journal of Systems and Software* (2006), and *IEEE Transactions on Reliability* (2011). His main research interests include software reliability and testing, reliable flight control, and software cybernetics.