Signal Processing 142 (2018) 1–10

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Identification for Hammerstein nonlinear ARMAX systems based on multi-innovation fractional order stochastic gradient☆

Songsong Cheng, Yiheng Wei, Dian Sheng, Yuquan Chen, Yong Wang*

Department of Automation, University of Science and Technology of China, Hefei, 230027, PR China

ARTICLE INFO

Article history:
Received 19 January 2017
Revised 28 May 2017
Accepted 20 June 2017
Available online 30 June 2017

Keywords:
Fractional order calculus
Hammerstein nonlinear ARMAX systems
MIFOSG
Identification
Convergence

ABSTRACT

A multi-innovation fractional order stochastic gradient (MIFOSG) algorithm, which involves a variable initial value scheme, is investigated to identify the Hammerstein nonlinear ARMAX systems in this paper. Firstly, according to an improved fractional order gradient method, the MIFOSG algorithm is proposed. Furthermore, according to the martingale convergence theorem, the convergence analysis of the proposed algorithm is developed. In addition, for the purpose of improving the convergence performance, a forgetting factor on step size and a variable gradient order are introduced. Given a sufficiently large number of independent runs, the effectiveness of the proposed algorithm is demonstrated in two numerical examples finally.

1. Introduction

As we known, many systems with nonlinear dynamics such as chemical systems [1], signal processing [2] and mechanical systems [3], etc., can be modeled as Hammerstein nonlinear ARMAX systems [4]. The identification of Hammerstein nonlinear ARMAX systems has drawn lots of scholars' attention. A recursive least-squares method and an iterative least-squares method are proposed in [4]. Also, on the basis of gradient search principle, the recursive stochastic gradient method and iterative stochastic gradient method are designed in [5].

However, for the method based on conventional stochastic gradient, the convergence rate is not satisfying enough since only current knowledge of the systems is utilized. In order to obtain a better convergence rate, Ding and Chen creatively proposed a notable new method based on MISG and analyzed its performance in [6]. Because of the superiority of the MISG in convergence rate, the method and its variants have been utilized to identify CARMA systems [7], Hammerstein nonlinear CARMA systems [5], Box-Jenkins systems [8] and Hammerstein nonlinear CARAR systems [9], etc., and obtained more extraordinary convergent performance than conventional stochastic gradient method.

As pointed out in [10,11], fractional order calculus has attracted many scholars' attention in system control [12–14] and signal processing [15,16], etc, because of the non–locality, namely, the next state of a system not only depends on its current state but also on its historical states starting from the initial time. Some crucial pioneering work by applying fractional order calculus into signal processing can be found in [17,18] and some fundamental theories and applications are presented in [19,20]. In the complex domain, Tseng et al. designed fractional order derivative constrained 1-D and 2-D FIR filters in [21]. By utilizing Riemann–Liouville integral, Wang et al. presented fractional zeros phase filter in [22]. In [23], Liu et al. developed a signal reconstruction scheme in fractional Fourier domain. In [24], Pu et al. proposed a factional order gradient method to seek the fractional extreme value. However, because of the non-locality of the fractional order calculus, the calculated fractional extreme value is sensitive to the initial value and different from the exact extreme value in general. For the purpose of ensuring that the algorithm will converge to the exact extreme value, a fractional order LMS algorithm with variable initial value scheme is designed in [16]. Raja et al. developed another type of fractional order LMS algorithm and applied it into channel equalisation [25,26], active noise rejection [27,28] and system identification [15,29,30] etc., and achieved good performance. Almost all of the existing results show that introduction of fractional order calculus can help obtain a better performance than the conventional counterpart. Therefore, the superiority can be also reflected by combination of multi-innovation and fractional order calculus.

Motivated by the discussions above in this paper, an improved fractional order gradient method with variable initial value is developed for a quadratic function. Based on the improved fractional order gradient method, a fractional order stochastic gradient (FOSG) algorithm is designed to identify the Hammerstein nonlin-

☆ Fully documented templates are available in the elsarticle package on CTAN.

* Corresponding author.
E-mail address: yongwang@ustc.edu.cn (Y. Wang).

http://dx.doi.org/10.1016/j.sigpro.2017.06.025
0165-1684/© 2017 Published by Elsevier B.V.

ear ARMAX systems. Then the developed algorithm is extended to the multi-innovation case and MIOSG algorithm is obtained. The convergence performance of MIOSG algorithm is analyzed in the light of the martingale convergence theorem.

The remainder of this paper is organized as follows. Section 2 provides some preliminaries of fractional order calculus and develops an improved fractional order gradient method. The MIFOSG algorithm and its performance analysis are presented in Section 3. In Section 4, two numerical examples are provided to illustrate the effectiveness of the proposed method. Conclusions are given in Section 5.

Notations The ∞ denotes the positive infinite ($+\infty$). $\lambda_{\min}\{\cdot\}$ ($\lambda_{\max}\{\cdot\}$) is the minimal (maximum) eigenvalue of a given matrix. z^{-1} means a unit backward shift operator, i.e., $z^{-1}x(n)=x(n-1)$, $\boldsymbol{a}\cdot\boldsymbol{b}$ means two vectors multiplied element by element. The norm of a given matrix X is calculated by $||X||_2=\mathrm{Tr}[XX^{\mathrm{T}}]$. $\lim_{n\to\infty}\frac{g_1(n)}{g_2(n)}=0$ can be shorted as $g_1(n)=o(g_2(n))$. For $g_2(n)\geq 0$, $g_1(n)=O(g_2(n))$ stands for that $\exists\delta>0$ to make $|g_1(n)|\leq\delta g_2(n)$.

2. Preliminaries

2.1. Fractional order calculus

There are three popular definitions of fractional order derivatives, namely, Grünwald–Letnikov derivative, Caputo derivative and Riemann–Liouville derivative. In this paper, we process the fractional order calculus based on the Riemann–Liouville definition, which can be expressed in the following derivative.

Definition 1. The αth derivative of a given function $f(t)$ is defined as

$$ {}_{t_0}\mathscr{D}_t^{\alpha} f(t)=\frac{1}{\Gamma(n-\alpha)}\frac{\mathrm{d}^n}{\mathrm{d}t^n}\int_{t_0}^{t}(t-\tau)^{n-\alpha-1}f(\tau)\mathrm{d}\tau, \tag{1}$$

where $n-1\leq\alpha<n$, and $\Gamma(n-\alpha)=\int_0^{\infty}t^{n-\alpha-1}\mathrm{e}^{-t}\mathrm{d}t$ is the so-called Gamma function.

Based on the Riemann–Liouville derivative of $f(t)$, the following lemma can be introduced.

Lemma 1 (see [16]). *For a given quadratic function $f(t)=(t-a)^2-(t_0-a)^2$, a corresponding fractional order gradient method with variable initial value $t_0=t(n-1)$ can be expressed as*

$$\begin{aligned} t(n+1) &\\ &= t(n)-\mu\,{}_{t(n-1)}\mathscr{D}_{t(n)}^{\alpha}f(t(n))\\ &= t(n)-\mu\frac{f^{(1)}(t(n-1))}{\Gamma(2-\alpha)}[t(n)-t(n-1)]^{1-\alpha}\\ &\quad -\mu\frac{f^{(2)}(t(n-1))}{\Gamma(3-\alpha)}[t(n)-t(n-1)]^{2-\alpha}, \end{aligned}\tag{2}$$

which will converge to the exact extreme value of $f(t)$, if the algorithm in (2) is convergent.

As discussed in [16], $t(n)$ is usually a slowly varying sequences, namely, $|t(n)-t(n-1)|\ll 1$. Then, ${}_{t(n-1)}\mathscr{D}_{t(n)}^{\alpha}f(t(n))$ is determined by $\frac{f^{(1)}(t(n-1))}{\Gamma(2-\alpha)}\times[t(n)-t(n-1)]^{1-\alpha}$. Besides, it is possible that $t(n)-t(n-1)<0$, which will lead that $[t(n)-t(n-1)]^{1-\alpha}$ is a complex number. Therefore, it is necessary to modify $[t(n)-t(n-1)]^{1-\alpha}$ as $|t(n)-t(n-1)|^{1-\alpha}$. Furthermore, when $1<\alpha<2$, $t(n)=t(n-1)$ will lead to $[t(n)-t(n-1)]^{1-\alpha}=\infty$. Thereby, $[|t(n)-t(n-1)|]^{1-\alpha}$ should be modified as $[|t(n)-t(n-1)|+\epsilon]^{1-\alpha}$, where ϵ is a small number. Consequently, for the quadratic function, the fractional order gradient method in (2) can be improved as follows

$$ t(n)=t(n-1)-\mu\frac{f^{(1)}(t(n-2))}{\Gamma(2-\alpha)}[|t(n-1)-t(n-2)|+\epsilon]^{1-\alpha}. \tag{3}$$

2.2. Hammerstein nonlinear ARMAX system

A Hammerstein nonlinear ARMAX system is governed by the following equation [29,31]

$$ A(z)y(n)=B(z)\bar{u}(n)+D(z)v(n) \tag{4}$$

where $u(n)$, $y(n)$ and $v(n)$ are the system input, system output and white Gaussian noise with zero mean and σ^2 variance, respectively. In detail, $A(z)=1+a_1z^{-1}+\cdots+a_{l_a}z^{-l_a}$, $B(z)=b_1z^{-1}+\cdots+b_{l_b}z^{-l_b}$, $D(z)=1+d_1z^{-1}+\cdots+d_{l_d}z^{-l_d}$ and $\bar{u}(n)=\boldsymbol{h}^{\mathrm{T}}\boldsymbol{f}(u(n))$, where $\boldsymbol{f}(u(n))\triangleq[f_1(u(n)),\cdots,f_{l_h}(u(n))]^{\mathrm{T}}$ is a vector whose elements are known nonlinear functions of $u(n)$, $\boldsymbol{a}\triangleq[a_1,\cdots,a_{l_a}]^{\mathrm{T}}$, $\boldsymbol{b}\triangleq[b_1,\cdots,b_{l_b}]^{\mathrm{T}}$, $\boldsymbol{h}\triangleq[h_1,\cdots,h_{l_h}]^{\mathrm{T}}$, $\boldsymbol{d}\triangleq[d_1,\cdots,d_{l_d}]^{\mathrm{T}}$ are unknown parameter vectors of the system in (4) to be identified.

According to (4), the system output can be expressed as

$$ y(n)=\boldsymbol{\phi}^{\mathrm{T}}(n)\boldsymbol{\theta}+v(n), \tag{5}$$

where $\boldsymbol{\theta}=[\boldsymbol{a}^{\mathrm{T}},\boldsymbol{h}^{\mathrm{T}}\otimes\boldsymbol{b}^{\mathrm{T}},\boldsymbol{d}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbb{R}^{l}$ and $\boldsymbol{\phi}(n)=[\boldsymbol{\phi}_y^{\mathrm{T}},\boldsymbol{\phi}_{f_1}^{\mathrm{T}},\ldots,\boldsymbol{\phi}_{f_{l_h}}^{\mathrm{T}},\boldsymbol{\phi}_v^{\mathrm{T}}]^{\mathrm{T}}\in\mathbb{R}^{l}$ with $l=l_a+l_bl_h+l_d$, $\boldsymbol{\phi}_y=[y(n-1),\ldots,y(n-l_a)]^{\mathrm{T}}\in\mathbb{R}^{l_a}$, $\boldsymbol{\phi}_{f_i}=[f_i(u(n-1)),\ldots,f_i(u(n-l_b))]^{\mathrm{T}}\in\mathbb{R}^{l_b}$, $i=1,\ldots,l_h$, $\boldsymbol{\phi}_v=[v(n-1),\cdots,v(n-l_d)]^{\mathrm{T}}\in\mathbb{R}^{l_d}$.

Remark 1. As discussed in [4,5], we can set $h_1=1$. Then the estimates $\hat{\boldsymbol{a}}$, $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{d}}$ of $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{d}$ can be directly obtained from $[\hat{\theta}_1,\cdots,\hat{\theta}_{l_a}]$, $[\hat{\theta}_{l_a+1},\cdots,\hat{\theta}_{l_a+l_b}]$ and $[\hat{\theta}_{l_a+l_hl_b+1},\cdots,\hat{\theta}_{l_a+l_hl_b+l_d}]$, where $\hat{\theta}_i$ is the ith element of $\hat{\theta}$ and $\hat{\theta}$ means the estimated value of θ. $\hat{\boldsymbol{h}}$ can be obtained by

$$ \hat{\boldsymbol{h}}=\frac{1}{l_b}\sum_{i=1}^{l_b}\left[1,\frac{\hat{\theta}_{l_a+l_b+i}}{\hat{\theta}_{l_a+1}},\cdots,\frac{\hat{\theta}_{l_a+l_hl_b+i}}{\hat{\theta}_{l_a+l_b}}\right]^{\mathrm{T}}. \tag{6}$$

Therefore, one of the main objectives of this paper is to design the MIFOSG algorithm to identify $\boldsymbol{\theta}$.

3. Main results

3.1. The FOSG algorithm

For the purpose of identifying the parameters of system (4), the FOSG algorithm with variable initial value $\hat{\boldsymbol{\theta}}(n-2)$ will be presented in this section. Consider $\hat{\boldsymbol{\theta}}(n)$ as the estimated value of $\boldsymbol{\theta}$ and define the criterion function as

$$ J(n)\triangleq\frac{1}{2}[y(n)-\boldsymbol{\phi}^{\mathrm{T}}(n)\boldsymbol{\theta}]^2. \tag{7}$$

Based on the improved fractional order gradient method in (3), $\boldsymbol{\theta}$ can be identified by

$$\begin{aligned} \hat{\boldsymbol{\theta}}(n) &= \hat{\boldsymbol{\theta}}(n-1)+\mu[y(n)-\boldsymbol{\phi}^{\mathrm{T}}(n)\hat{\boldsymbol{\theta}}(n-1)]\\ &\quad\times\boldsymbol{\phi}(n)\cdot\frac{[|\hat{\boldsymbol{\theta}}(n-1)-\hat{\boldsymbol{\theta}}(n-2)|+\epsilon]^{1-\alpha}}{\Gamma(2-\alpha)}\\ &= \hat{\boldsymbol{\theta}}(n-1)+\mu[y(n)-\boldsymbol{\phi}^{\mathrm{T}}(n)\hat{\boldsymbol{\theta}}(n-1)]\frac{\Xi(\hat{\boldsymbol{\theta}},\alpha,n)}{\Gamma(2-\alpha)}\boldsymbol{\phi}(n), \end{aligned}\tag{8}$$

where $\Xi(\hat{\boldsymbol{\theta}},\alpha,n)=\mathrm{diag}\{[|\hat{\theta}_1(n-1)-\hat{\theta}_1(n-2)|+\epsilon]^{1-\alpha},[|\hat{\theta}_2(n-1)-\hat{\theta}_2(n-2)|+\epsilon]^{1-\alpha},\cdots,[|\hat{\theta}_l(n-1)-\hat{\theta}_l(n-2)|+\epsilon]^{1-\alpha}\}$.

However, as discussed in [4,5], since $\boldsymbol{\phi}(n)$ contains unknown $v(n-i)$, $i=1,\cdots,l_d$, the designed algorithm cannot be realized in practice. Hereby, in this paper, $\boldsymbol{\phi}(n)$ in (8) is replaced by $\hat{\boldsymbol{\phi}}(n)=[\boldsymbol{\phi}_y^{\mathrm{T}}(n),\boldsymbol{\phi}_{f_1}^{\mathrm{T}}(n),\cdots,\boldsymbol{\phi}_{f_{l_h}}^{\mathrm{T}}(n),\hat{\boldsymbol{\phi}}_v^{\mathrm{T}}(n)]^{\mathrm{T}}$, where